

**UNIVERSIDADE ESTADUAL DO OESTE DO PARANÁ
CAMPUS DE CASCAVEL
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA AGRÍCOLA**

**AVALIAÇÃO DE SINISTROS AGRÍCOLAS VIA SENSORIAMENTO REMOTO ORBITAL E
APRENDIZADO DE MÁQUINA**

EDUARDO MATIAS RAMBO

Cascavel - Paraná - Brasil

Agosto 2020

EDUARDO MATIAS RAMBO

**AVALIAÇÃO DE SINISTROS AGRÍCOLAS VIA SENSORIAMENTO REMOTO ORBITAL E
APRENDIZADO DE MÁQUINA**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Agrícola para obtenção do título de mestre em Engenharia Agrícola, área de concentração Sistemas Biológicos e Agroindustriais, com a temática Geoprocessamento, Estatística Espacial e Agricultura de Precisão.

Orientador: Dr. Miguel Angel Uribe Opazo
Co-orientador: Dr. Jerry Adriani Johann

Cascavel - Paraná - Brasil

Agosto 2020

Ficha de identificação da obra elaborada através do Formulário de Geração Automática do Sistema de Bibliotecas da Unioeste.

Rambo, Eduardo Matias
AVALIAÇÃO DE SINISTROS AGRÍCOLAS VIA SENSORIAMENTO
REMOTO ORBITAL E APRENDIZADO DE MÁQUINA / Eduardo Matias
Rambo; orientador(a), Miguel Angel Uribe Opazo;
coorientador(a), Jerry Adriani Johann, 2020.
86 f.

Dissertação (mestrado), Universidade Estadual do Oeste
do Paraná, Campus de Cascavel, Centro de Ciências Exatas e
Tecnológicas, Programa de Pós-Graduação em Engenharia
Agrícola, 2020.

1. Seguro rural. 2. Perfil espectro-temporal. 3.
TIMESAT. 4. EVI. I. Opazo, Miguel Angel Uribe. II. Johann,
Jerry Adriani. III. Título.

EDUARDO MATIAS RAMBO

**AVALIAÇÃO DE SINISTROS AGRÍCOLAS VIA SENSORIAMENTO REMOTO ORBITAL E
APRENDIZADO DE MÁQUINA**

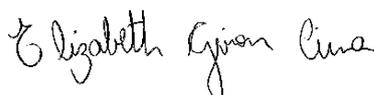
Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Mestre em Engenharia Agrícola, área de concentração Sistemas Biológicos e Agroindustriais, linha de pesquisa Geoprocessamento, Estatística Espacial e Agricultura de Precisão, APROVADO pela seguinte banca examinadora:



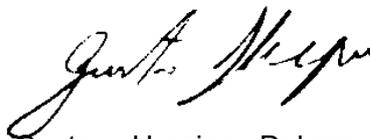
Orientador(a) - Miguel Angel Uribe Opazo
Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)



Jerry Adriani Johann
Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)



Elizabeth Giron Cima
Universidade Estadual do Oeste do Paraná - Campus de Toledo (UNIOESTE)



Gustavo Henrique Dalposso
Universidade Tecnológica Federal do Paraná (UTFPR)



Alfredo Richart
Pontifícia Universidade Católica do Paraná (PUCPR)

Cascavel, 20 de Agosto de 2020.

BIOGRAFIA RESUMIDA

Brasileiro, nascido em Romelândia-SC, no ano de 1987. É mestre em Engenharia Agrícola pelo Programa de Pós-Graduação em Engenharia Agrícola (PGEAGRI), na Universidade Estadual do Oeste do Paraná (UNIOESTE), na área de concentração de Sistemas Biológicos e Agroindustriais, linha de pesquisa em Geoprocessamento, Estatística Espacial e Agricultura de Precisão, com bolsa de estudos pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Graduado em Agronomia como bolsista do PROUNI pela Pontifícia Universidade Católica do Paraná (PUCPR). Foi voluntário de iniciação científica em 2012 nas áreas de pesquisa de desenvolvimento de fertilizantes químicos alternativos. Possui graduação a nível Técnico em Meio Ambiente no Centro Estadual de Educação Profissional de Cascavel (CEEP). Integra o grupo de pesquisadores do Núcleo de Pesquisa em Geotecnologias e Ciência de Dados (GeoScience), na área de sensoriamento remoto aplicada a agricultura, atuando principalmente nos seguintes temas: mapeamento e estimativa de área, determinação de datas do ciclo fenológico e fiscalização de operações de crédito e seguro rural. Atualmente é perito agrícola com atuação em parte do segmento do seguro rural com experiência em perícias agrícolas em casos de sinistros nas culturas de soja, milho, trigo, feijão e sorgo. Experiência em Agronomia, atendimento a clientes a campo, condução ensaios a campo e treinamento de corpo técnico, produtores e estudantes.

AGRADECIMENTOS

À Universidade Estadual do Oeste do Paraná – UNIOESTE e ao Programa de Pós-Graduação em Engenharia Agrícola – PGEAGRI, pela oportunidade oferecida para fazer o curso de mestrado;

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, pelo apoio financeiro;

Ao meu orientador Professor Dr. Miguel Angel Uribe Opazo e ao meu co-orientador Professor Dr. Jerry Adriani Johann, pelo apoio, dedicação, paciência e companheirismo;

Aos colegas e amigos do Laboratório de Estatística Aplicada – LEA, em especial a Willyan Becker, Thiago Ló, Laíza Cavalcante, Jonathan Richetti, Alex Paludo e Rodrigo Lorbiesk, pelo companheirismo na dissertação, no mestrado e na vida;

A toda minha família, em especial à minha querida esposa que suportou comigo o fardo das atividades acadêmicas e pessoais.

E a todos que de alguma forma contribuíram na realização deste trabalho.

Obrigado!

AVALIAÇÃO DE SINISTROS AGRÍCOLAS VIA SENSORIAMENTO REMOTO ORBITAL E APRENDIZADO DE MÁQUINA

RESUMO

O seguro rural é uma alternativa importante para tornar o setor agrícola financeiramente estável, mesmo com ocorrências de eventos naturais adversos. Trata-se de um contrato securitário no qual a parte contratada se obriga a indenizar a outra por um prejuízo eventual. Dessa forma, para o agricultor, é um mecanismo de transferência de uma despesa futura incerta e de valor elevado, por uma despesa antecipada de valor reduzido. Devido aos grandes valores envolvidos nessas transações, são necessários mecanismos para fiscalização da aplicação desses recursos. O sensoriamento remoto agrícola proporciona o acompanhamento amplo e sistemático das lavouras de forma contínua e com baixo custo, facilitando, assim, o processo de fiscalização e tomada de decisão sobre seguros rurais. Este trabalho se propõe a definir uma metodologia para a confirmação de sinistros agrícolas pela análise dos padrões sazonais de EVI/Landsat-8 e dados climáticos. Para isso, serão utilizadas informações de lavouras sinistradas (milho, soja e trigo) e lavouras sem a ocorrência de sinistro, com a finalidade de verificar o padrão existente entre esses parâmetros, indicando a ocorrência ou não de sinistros e qual é o tipo de sinistro incidente, comparando com os laudos de perícias reais cedidos por empresa parceira do estudo. Para as análises foram utilizados algoritmos de classificação supervisionada do *Support vector machine* (SVM), *Random forest* (RF) e *Decision tree* (DT). O classificador RF obteve melhor desempenho dentre os demais, pois classificou a ocorrência de sinistro com acurácias de 83, 96 e 81% para milho, soja e trigo, respectivamente. Para os tipos de sinistro nas culturas sinistradas, observou-se maiores acurácias para o SVM na cultura do milho (99%) e soja (90%). Para o trigo a maior acurácia ocorreu com o RF (86%). A metodologia apresentada se demonstrou eficaz no levantamento de informações confiáveis para a confirmação de ocorrências de sinistro no ramo de seguros rurais, sendo uma alternativa viável e de grande importância para a estabilidade das seguradoras, como auxiliar no procedimento de peritagem e fiscalização de ações relacionadas a esses segmentos agrícolas.

PALAVRAS-CHAVE: Seguro rural, perfil espectro-temporal, EVI, TIMESAT, *Support vector machine*, *Random forest*, *Decision tree*, Eficiência.

EVALUATION OF AGRICULTURAL CLAIMS VIA ORBITAL REMOTE SENSING AND MACHINE LEARNING

ABSTRACT

Agricultural insurance is an important alternative to convert the agricultural sector into a financially stable model, even when faced with adverse events and natural disasters. Agricultural insurance is a financial instrument to reduce risk related to natural disasters by establishing a future contract in which one party is obligated to compensate the damage loss to the other party by paying a premium. Hence, for the producer, it works as a way to substitute an uncertain future financial loss by a reduced, predictable investment. Due to the higher transaction values involved, it is necessary to develop methods to inspect farms and verify the claimed losses. Remote sensing has the potential to support the insurance industry by providing an alternative to crop monitoring in large scales and at a low cost, facilitating the processes of fiscalization and decision making regarding agricultural insurance. Thus, this research aims to develop a methodology to confirm if a claimed loss occurred by applying seasonal trend analysis in Landsat-8/EVI time-series combined with weather. For this effect, information about both affected and non-affected areas are employed, using real inspected farmers (sown with maize, soybean, and wheat), in order to verify the existing pattern between these parameters, indicating their occurrence and distinguishing the natural disaster, by comparing the judicial investigation data provided by the insurance company helping this research. For the analyses, three classifiers were applied: Decision Tree (DT), Random Forest (RF) and Support Vector Machine (SVM). RF classifier achieved 83, 96, and 81% for maize, soybean, and wheat, respectively, when determining whether a natural disaster did or did not occur. SVM classifier achieved 99 and 90% in maize and soybean, respectively, to detect the type of disaster, and RF achieved 86% for wheat at the same task. This methodology has proved to be efficient to confirm and detect a natural disaster, being a viable and important alternative solution for insurance companies to minimize their risks and increase their efficiency, helping in the process of insurance verification and fiscalization of actions related to such agricultural segments.

Keywords: Agricultural insurance, crop insurance, time-series, EVI, TIMESAT, Support vector machine, Random forest, Decision tree, Efficiency.

SUMÁRIO

LISTA DE TABELAS	9
LISTA DE FIGURAS	11
1 INTRODUÇÃO	13
2 OBJETIVOS	16
2.1 Objetivo geral	16
2.2 Objetivos específicos.....	16
3 REVISÃO BIBLIOGRÁFICA	17
3.1 Crédito rural.....	17
3.2 Seguro rural.....	18
3.2.1 Programa de garantia da atividade agropecuária (PROAGRO) e Programa de subvenção ao prêmio do seguro rural (PSR).....	20
3.2.2 Indicadores do porte do seguro rural no Brasil	21
3.2.3 Fiscalização de sinistros <i>in loco</i>	25
3.3 Sensoriamento remoto	27
3.3.1 Índices de vegetação.....	27
3.3.2 Landsat-8	29
3.4 TIMESAT	30
3.5 Mineração de dados	31
3.5.1 <i>Support vector machine</i> (SVM).....	33
3.5.2 <i>Random forest</i>	34
3.5.3 Árvore de decisão	35
4 MATERIAL E MÉTODOS	36
4.1 Aquisição de dados	37
4.2 Pré-processamento	43
4.3 Processamento	48
4.3.1 Comparações entre dados reais e estimações do TIMESAT	51
4.4 Validação	52
5 RESULTADOS E DISCUSSÃO	56
5.1 Resultado das comparações entre dados reais e estimações do TIMESAT	58
5.2 Previsão da ocorrência de sinistro.....	59
5.3 Previsão do tipo de sinistro ocorrido	67
5.4 Relação entre dados de satélites e laudos de sinistros	72
6 CONCLUSÕES	74
6.1 Trabalhos futuros	74
6.2 Contribuições esperadas.....	75
6.3 Considerações finais	75
REFERÊNCIAS	77
ANEXOS	84
Anexo A Modelo das páginas do laudo da seguradora	85
Anexo B Resultado do teste Mann-Whitney entre os dados de precipitação oriundos da plataforma Nasa Power, em relação a estações SIMEPAR	87

LISTA DE TABELAS

Tabela 1	Informações de saída do TIMESAT, referentes ao perfil espectro-temporal de IV das culturas	30
Tabela 2	Número de laudos agrícolas por ano da vistoria	37
Tabela 3	Eventos climáticos causadores dos sinistros nas culturas da soja, milho 2 ^a safra e trigo - Paraná.....	39
Tabela 4	Estatísticas relacionadas à área (hectare) das lavouras vetorizadas de milho 2 ^a safra e trigo, sinistradas e normais, neste banco de dados para o estado do Paraná.....	41
Tabela 5	Quantidade de dados resultantes da aquisição de imagens de interesse, separados por culturas sinistradas e normais	43
Tabela 6	Exemplo dos dados <i>seasonality</i> convertidos para números, datas e dias viáveis de análise	50
Tabela 7	Atributos selecionados no banco de dados para composição das análises	52
Tabela 8	Atributos climáticos selecionados no banco de dados para composição das análises	52
Tabela 9	Estatísticas de comparação em dias das datas de semeadura e colheita, entre os dados reais (laudos) com os estimados pela análise de padrões sazonais do TIMESAT.....	58
Tabela 10	Compilação das matrizes de confusão do conjunto de teste em porcentagem, para os algoritmos SVM, RF e DT aplicados aos dados das culturas de milho 2 ^a safra, soja e trigo.....	60
Tabela 11	Medidas de desempenho dos classificadores SVM, RF e DT, considerando a ocorrência de sinistro na cultura do milho.....	61
Tabela 12	Medidas de desempenho dos classificadores SVM, RF e DT, considerando a ocorrência de sinistro na cultura da soja.....	63
Tabela 13	Medidas de desempenho dos classificadores SVM, RF e DT, considerando a ocorrência de sinistro na cultura do trigo	65
Tabela 14	Resumo da acurácia global \pm desvio padrão para o conjunto de treinamento e teste dos algoritmos testados para ocorrência de sinistro nas culturas de milho, soja e trigo	67
Tabela 15	Medidas de desempenho dos classificadores SVM, RF e DT para os tipos de sinistros (chuva excessiva e seca) da cultura do milho.....	67
Tabela 16	Medidas de desempenho dos classificadores SVM, RF e DT para os tipos de sinistros (chuva excessiva, seca, granizo) da cultura da soja	69

Tabela 17 Medidas de desempenho dos classificadores SVM, RF e DT para os tipos de sinistros (seca, seca + geada) da cultura do trigo	70
Tabela 18 Resumo da acurácia global \pm desvio padrão para o conjunto de treinamento e teste dos algoritmos testados nos sinistros das culturas de milho, soja e trigo	71

LISTA DE FIGURAS

Figura 1	Valores totais de subvenção ao seguro rural no Brasil no período de 2006 a 2018.	22
Figura 2	Valores totais de subvenção ao seguro rural por estado no período de 2006 a 2018.	23
Figura 3	Valores totais de subvenção ao seguro rural e prêmios pagos pelos produtores no período de 2006 a 2018.....	23
Figura 4	Evolução da área segurada nas principais culturas subvencionadas no Brasil, no período de 2006 a 2018.....	24
Figura 5	Número de contratações (apólices) e número de contratantes (produtores) de seguro no período de 2006 a 2018.	25
Figura 6	Padrão de um perfil espectro-temporal de índice de vegetação EVI para uma cultura de verão.....	31
Figura 7	Processo de descoberta de conhecimento em banco de dados – KDD.....	32
Figura 8	Tarefas de mineração de dados.	32
Figura 9	Construção de hiperplanos de separação dos dados, de tal forma que maximizem as distâncias entre os elementos de cada grupo.	33
Figura 10	Representação genérica do modelo <i>Random forest</i> com suas diversas árvores.	34
Figura 11	Ilustração da hierarquia da árvore de decisão, no qual os nós representam um teste lógico, conectados por tomadas de decisão (nó/ramo), obtendo um resultado final (folha).....	35
Figura 12	Fluxograma geral da dissertação com as etapas de aquisição de dados, pré-processamento, processamento e validação.	36
Figura 13	Mapa de localização dos eventos causadores de sinistros nas culturas da soja, milho 2ª safra e trigo, neste banco de dados para o estado do Paraná.	38
Figura 14	Croqui elaborado no laudo final (esquerda), vetor da área na interface do Google <i>Earth</i> (direita).....	39
Figura 15	Mapa de localização das lavouras sinistradas de soja, milho 2ª safra e trigo, neste banco de dados para o estado do Paraná.	40
Figura 16	Mapa de localização das lavouras de soja, milho 2ª safra e trigo consideradas normais, neste banco de dados para o estado do Paraná.	41
Figura 17	Interface da plataforma Nasa <i>Power</i> , com destaque a janela de seleção de dados.	42

Figura 18	Sobreposição de <i>tiles</i> do satélite Landsat 8 e, em destaque, algumas lavouras de interesse na sobreposição das imagens.	44
Figura 19	Comparação do perfil de EVI submetido à interpolação linear e à interpolação quadrática.	45
Figura 20	Comparação do perfil de EVI original (a esquerda) e depois da interpolação linear dos dados (direita).	45
Figura 21	Comparativo entre a localização dos dados extraídos do NASA <i>Power</i> e estações meteorológicas do SIMEPAR.	46
Figura 22	Posição das variáveis climáticas no perfil temporal de EVI definidas para o estudo em um exemplo de lavoura de soja com sinistro.	47
Figura 23	Exemplo da disposição dos dados de entrada em arquivo <i>.txt</i>	48
Figura 24	Tela do <i>software</i> TIMESAT onde podem ser vistas as ferramentas que o <i>software</i> disponibiliza para o ajuste do perfil temporal.	49
Figura 25	Exemplo de Tabela <i>Seasonality.txt</i>	50
Figura 26	Valores máximos, mínimos e perfis médios de EVI em <i>pixels</i> puros (30x30m) para lavouras sinistradas.	56
Figura 27	Valores máximos, mínimos e perfis médios de EVI em <i>pixels</i> puros (30x30m) para lavouras normais.	57
Figura 28	DT resumida para os fatores relevantes na ocorrência de sinistro na cultura do milho.	62
Figura 29	DT resumida para os fatores relevantes na ocorrência de sinistro na cultura da soja.	64
Figura 30	DT resumida para os fatores relevantes na ocorrência de sinistro na cultura do trigo.	66
Figura 31	DT resumida para os fatores relevantes nos tipos de sinistro da cultura do milho.	68
Figura 32	DT resumida para os fatores relevantes nos tipos de sinistro da cultura da soja.	70
Figura 33	DT resumida para os fatores relevantes nos tipos de sinistro da cultura do trigo.	71

1 INTRODUÇÃO

O Brasil é um dos principais fornecedores mundiais de produtos agrícolas, atendendo a uma necessidade alimentícia cada vez maior. Para alimentar as esperadas 9 bilhões de pessoas que existirão em 2050, os produtores agrícolas precisam produzir tanto alimento nos próximos 50 anos quanto o que já produziram desde os estágios iniciais da civilização. Espera-se que o Brasil cubra uma grande parte do aumento previsto nessa produção de alimentos, devido à sua enorme disponibilidade de recursos naturais e sua grande dimensão territorial (LOPES; LOWERY; PEROBA, 2016).

A relevância desse segmento é observada ao longo do desenvolvimento brasileiro, alavancando o crescimento de regiões e de inovações tecnológicas, indústrias de insumo, distribuição e logística de produtos agrícolas, como também das unidades produtivas (CONTINI et al., 2012). O MAPA (2019a) projeta no agronegócio e no setor agroindustrial brasileiro uma perspectiva de crescimento até 2028 de 68,5 milhões de hectares cultivados com arroz, feijão, milho, soja e trigo. Em 2028/2029, estima-se que a produção brasileira de soja seja de 151,9 milhões de toneladas, o que corresponde, em relação à safra 2018/19, a um aumento de 32,9% da produção, 22,6% do consumo e 41,8% da exportação desse grão. Em relação à produção de milho, estima que seja de 114,5 milhões de toneladas, com aumento na produção de 20,2%, no consumo 19,7% e na exportação 33,4%. Quanto à produção de trigo, estima que seja de 7,2 milhões de toneladas, com aumento na produção de 31,6%, no consumo 14,6% e importação 0,9%.

Apesar das boas projeções para o segmento agrícola, o desempenho da atividade está vinculado a fatores exógenos à produção, como as condições edafoclimáticas, o ataque de pragas, a aquisição de insumos e a volatilidade de preços na comercialização de produtos agrícolas (BUAINAIN, VIEIRA, CURY, 2014). Por isso, a agricultura demanda o uso racional e eficiente do capital e trabalho e o emprego de tecnologias eficientes na redução de custos e riscos (RAMOS, 2009). Todavia, os fatores relacionados às condições climáticas ocasionam grandes incertezas para o setor, provocando perdas econômicas e afetando o bem-estar do indivíduo (HARWOOD et al., 1999).

Diversos pesquisadores apontam cenários agrícolas negativos, em função das mudanças climáticas. Burke e Emerick (2016) relatam o efeito das mudanças climáticas na agricultura americana e seus impactos no setor econômico. Zhang, Zhang e Chen (2017) contemplam as variáveis meteorológicas umidade e vento para elaborar um modelo, visando compreender cenário futuro e, considerando-se apenas o Brasil, tem-se uma expectativa de aumento de 2 a 3 °C da temperatura, em todas as regiões, de redução do período de chuvas e modificação dos períodos de plantio e dos métodos empregados.

Para contornar essas condições, é necessário implantar um manejo eficiente dos riscos envolvidos, prevendo a proteção das pessoas vinculadas às atividades, a viabilidade econômica do setor, oferecendo um ambiente interessante para investimentos (OECD, 2017). Uma das formas de reduzir esses riscos é a adoção do seguro agrícola, definido como um mecanismo de política agrícola que visa à proteção e à melhoria da atividade agropecuária, por meio da prevenção e redução de riscos, incentivo à diversificação da produção e estímulo a novas práticas de manejo (RAMOS, 2009).

A *Organization for Economic Co-operation and Development* (Organização para Cooperação e Desenvolvimento Econômico - OECD, 2015) recomenda a adoção de políticas públicas que compreendam os riscos da atividade agrícola, incluindo o treinamento de agricultores para lidarem com as adversidades, como também subsídios para o seguro agrícola. No Brasil, esse segmento é um mercado privado com participação estatal, conduzido por duas iniciativas principais: o Programa de Garantia da Atividade Agropecuária (PROAGRO) e o Programa de Subvenção ao Prêmio do Seguro Rural (PSR).

Ocasionalmente, o desempenho agrícola estadual está vinculado a um conjunto de expectativas que nem sempre se concretizam, como o clima favorável. A safra paranaense de verão 2011/2012, por exemplo, foi marcada por uma forte estiagem, o que ocasionou uma redução de 4,63 milhões de toneladas na colheita da soja (CONAB, 2012), impactando negativamente o setor. A ocorrência de sinistros na lavoura reforça a importância da contratação de apólices de seguro, pois os prejuízos decorrentes de um evento negativo seriam arcados pelas seguradoras.

Os procedimentos do seguro rural e os programas envolvidos mencionam uma perícia *in loco* na propriedade para avaliação da lavoura em caso de sinistros (FAEP, 2017; ALMEIDA, 2016; BUAINAIN; VIEIRA; CURY, 2014). Entretanto, tal metodologia é passível de fraudes. Cunha e Assad (2001) comentam que na história do PROAGRO já foram premiados casos duvidosos, como “altas taxas de sinistralidade, coberturas duvidosas (seca em lavoura irrigada, por exemplo)”. Além disso, a avaliação periódica *in loco* de lavouras financiadas ou seguradas torna-se impraticável, quando existem muitas propriedades rurais para serem vistoriadas no mesmo período.

Nesse contexto, as imagens de satélite de observação da superfície terrestre, aliadas às técnicas de geoprocessamento, podem auxiliar no acompanhamento sistemático e individualizado de lavouras em grandes áreas, além de definir áreas prioritárias a serem visitadas (MACEDO; RUDORFF, 2003).

O processamento dessas informações se encaixa como uma ferramenta periódica para a fiscalização de operações de crédito e seguro rural, por auxiliarem na avaliação do comprometimento das lavouras, possibilitando a visualização espacial das áreas, a identificação da cultura agrícola avaliada, a quantificação das áreas plantadas, incluindo estimativas de produção e produtividade durante o ciclo da cultura. Portanto, a utilização do

sensoriamento remoto poderá facilitar e agilizar o processo, melhorando a qualidade da informação obtida, permitindo a redução dos gastos da fiscalização, proporcionando resultados com maior antecedência, maior precisão e menor custo operacional que as técnicas tradicionais (JOHANN et al., 2012).

2 OBJETIVOS

2.1 Objetivo geral

Propor uma metodologia para avaliação de sinistros em culturas agrícolas por técnicas de sensoriamento remoto e algoritmos de aprendizado de máquina para auxiliar a fiscalização de seguro rural.

2.2 Objetivos específicos

- Construir e avaliar uma base de dados com as séries temporais, índices de vegetação EVI/Landsat-8 e dados meteorológicos (precipitação, temperatura média, máxima e mínima) para lavouras de milho, soja e trigo atingidas por sinistros e lavouras normais;
- Aplicar algoritmos de mineração de dados nas séries temporais do EVI/Landsat-8 e nos dados meteorológicos de lavouras normais e lavouras sinistradas, a fim de verificar a relação entre o sinistro e a base de dados construída;
- Avaliar a aplicabilidade da metodologia para confirmação de informações sobre lavouras sinistradas, descrevendo as características e limitações das ferramentas utilizadas, tendo como referência as exigências da regulamentação do Banco Central do Brasil nº 4.427/2015.

3 REVISÃO BIBLIOGRÁFICA

3.1 Crédito rural

O crédito rural tem por propósito fornecer crédito a juros baixos para que pequenos produtores rurais e cooperativas possam financiar os custos de produção e comercialização, implantado por meio da exigência legal de que os bancos dediquem uma parte de seus depósitos à vista a linhas de crédito rural (SANTANA; NASCIMENTO, 2012). Este custeio pode se destinar às despesas normais do ciclo produtivo de lavouras (perene ou semiperene), à extração de produtos vegetais até seu beneficiamento e armazenagem, como também para o setor agropecuário (BRASIL, 2018).

Todas as operações de custeio rural são susceptíveis à análise e fiscalização da aplicação dos recursos e são executadas a partir de levantamento de dados com trabalhos a campo, exigindo a mobilização de uma equipe para averiguação. Tal procedimento encarece o produto, além de demandar elevado tempo, o que muitas vezes inviabiliza a operação, muitas vezes, também, por se tratar de locais de difícil acesso (MARTIN, 2015).

Na fiscalização de crédito, é comum a emissão de laudos relatando a condição da lavoura em fases específicas durante o ciclo da cultura objeto do financiamento. Na maioria dos casos a própria cultura serve como uma garantia ao órgão financiador, motivo pelo qual, geralmente, as lavouras objeto de crédito estão automaticamente seguradas (FAEP, 2016).

Em contrapartida, o Banco Central do Brasil emitiu a Resolução Nº 4.427 de 25 de junho de 2015 (BRASIL, 2015, p. 1), a qual “autoriza a utilização do sensoriamento remoto para fins de fiscalização de operações de crédito rural e determina o registro das coordenadas geodésicas do empreendimento financiado por operações de crédito rural no Sistema de Operações do Crédito Rural e do PROAGRO (SICOR)”. O Artigo 1º da resolução, detalha que:

As coordenadas geodésicas (CG) do empreendimento financiado pelas operações de crédito de custeio agrícola e pelas operações de crédito de investimento (...) devem: a) ser informadas no orçamento, plano ou projeto relativo ao empreendimento a partir de 1º de julho de 2016, nas operações acima de R\$ 40.000,00 (quarenta mil reais) (...) compreender os pontos necessários à identificação do perímetro que define a gleba a ser cultivada ou, se for o caso, das duas ou mais glebas objeto da mesma operação de financiamento (BRASIL, 2015, p.1).

Este ponto é crucial, pois demonstra a necessidade de que os pontos sejam coerentes para a correta identificação da propriedade rural. Na sequência, é reafirmada a importância da área obtida através de sensoriamento, que deve ser compatível com a documentação,

ainda que não se tenha definida claramente a precisão e/ou a possibilidade de sobreposições entre propriedades.

Outro ponto fundamental nesta resolução, refere-se às condições necessárias para o emprego do sensoriamento remoto para fins de fiscalização. A resolução define a necessidade de uma sequência de imagens do empreendimento com resolução espacial inferior a 30 metros e resolução radiométrica mínima de 10 bits, que quantifique a área cultivada com erro máximo de 10%; identifique a cultura e seu desenvolvimento vegetativo em cada fase do cultivo; a obtenção de, no mínimo, três imagens, registrando as seguintes fases do cultivo: I – desenvolvimento vegetativo inicial ou, para culturas permanentes, obtenção de imagem em data apropriada para fins de fiscalização prévia; II – desenvolvimento vegetativo pleno; e III – estágio final de maturação. O item 5 limita a presença de nuvens nas imagens:

Para cada fase mencionada na alínea 'c' do item 4, as imagens utilizadas podem apresentar, no máximo, 10% da área do empreendimento coberta por nuvens, sombras ou outros fatores que comprometam a extração de informações, admitindo-se a sobreposição de imagens tomadas em momentos diferentes da mesma fase (BRASIL, 2015, p. 3).

Ressalta-se que o uso de sensoriamento remoto não desobriga os bancos em fazer o monitoramento amostral *in loco*, como afirma o item 8:

É obrigatória a vistoria local, por fiscais não relacionados com os trabalhos ou com as entidades contratadas para os serviços de sensoriamento, de 10% dos empreendimentos fiscalizados por sensoriamento remoto, selecionados aleatoriamente pela instituição financeira (BRASIL, 2015, p. 4).

A lei não deixa claro se, caso a seguradora não utilize sensoriamento remoto para fiscalização dos imóveis, esta terá que visitar todos os imóveis segurados.

Por fim, em virtude dos aspectos técnicos abordados na resolução Nº 4427, e a dificuldade de inspeção *in loco* nos estabelecimentos rurais, é necessário elaborar um protocolo para uso dos dados de sensoriamento remoto, que envolva a obtenção de informações, preferencialmente, por meio de satélites de uso gratuito, que sejam acessíveis e replicáveis aos técnicos envolvidos com tecnologias geodésicas, bem como com os responsáveis pelas instituições financeiras e seguradoras.

3.2 Seguro rural

O sucesso do empreendimento agrícola depende fundamentalmente de fatores que estão fora do controle do produtor rural. O clima é uma das principais variáveis de impacto na produção agrícola. Outros fatores adversos também podem reduzir drasticamente a

produtividade, influenciar na volatilidade dos preços e levar a quedas bruscas do faturamento. Para amenizar os efeitos de quebra de safra e demais riscos atrelados, tem-se o seguro rural, considerado como um dos mais importantes instrumentos de política agrícola, por permitir ao produtor proteger-se contra perdas decorrentes, principalmente, de fenômenos climáticos adversos (SUSEP, 2019).

Esse tipo de seguro tem por meta garantir ao agricultor a cobertura das culturas implantadas e tecnicamente conduzidas, garantindo indenização pelos prejuízos causados à lavoura decorrentes direta ou indiretamente de um determinado conjunto de eventos.

Basicamente, o seguro rural se constitui num contrato no qual mediante uma taxa (prêmio), a parte contratada (seguradoras) se obriga a indenizar a outra por um prejuízo eventual. Sendo assim, o seguro é um mecanismo pelo qual se transfere uma despesa futura e incerta (dano), de valor elevado, para uma despesa antecipada e certa, de valor relativamente menor (OZAKI; CAMPOS, 2017).

As condições e coberturas do seguro são definidas na apólice, em que as taxas são definidas conforme o risco. Outro fator preponderante na definição das taxas é a frequência com que o evento ocorre, por exemplo, em uma região onde seja comum a ocorrência de severas estiagens, as taxas de danos por seca serão mais elevadas do que em outra onde este evento é pouco relevante. Além disso, pesam também os mecanismos de prevenção disponíveis, como sistemas de irrigação por exemplo (FORNAZIER et al., 2014).

Ozaki e Campos (2017) relatam que a ausência de informações é um dos grandes entraves ao desenvolvimento deste mercado. Em geral, as seguradoras quantificam e precificam o risco com base nas informações individualizadas das unidades seguradas; veja-se o seguro de automóveis, em que as seguradoras traçam um perfil de risco com base nas características do veículo, local de uso e condutor principal. No seguro agrícola a ideia é a mesma, com o diferencial de que informações de produtividade real das propriedades rurais são praticamente inexistentes.

Nesse contexto, as seguradoras utilizam dados de produtividade por municípios, obtidos pelo Instituto Brasileiro de Geografia e Estatística (IBGE), implicando uma taxa de prêmio municipalizada, ou seja, todos os produtores do mesmo município pagam a mesma taxa de prêmio independente do seu nível tecnológico, pois as variações climáticas e de solo, entre outras, não seguem as subdivisões municipais. Ainda, a liberação dessas estimativas pelo IBGE é lenta, levando pelo menos dois anos para divulgação (OZAKI; CAMPOS, 2017).

Contudo, o seguro rural é mais abrangente, cobrindo não só a atividade agrícola, mas também a atividade pecuária, florestal, aquícola, patrimônio do produtor rural, seus produtos, o crédito para comercialização desses produtos, além do seguro de vida dos produtores (SUSEP, 2019).

O guia de seguro rural da FAEP (2016) descreve as formas de contratação das coberturas do seguro agrícola:

- a) Seguros Multirriscos: seguros nos quais são cobertos diversos riscos climáticos em uma única cobertura;
- b) Seguros de Riscos Nomeados: seguros nos quais os riscos cobertos são apresentados em coberturas distintas, havendo a possibilidade de contratar apenas as coberturas de maior interesse (FAEP, p. 10, 2016).

Ainda de acordo com a FAEP (2016), nos seguros agrícolas é importante também observar as seguintes variáveis:

- I) Produtividade Esperada: Trata-se da referência de potencial produtivo da lavoura segurada. Deverá corresponder tanto quanto possível à média histórica de produtividade da área a ser segurada tendo como referência dados do IBGE;
- II) Nível de Cobertura: Refere-se a um percentual de proteção garantido pela apólice aplicável à produtividade esperada ou faturamento esperado. Varia entre 50% e 80% conforme a seguradora e o produto agrícola;
- III) Riscos Excluídos: Variam de um seguro para outro, mas normalmente são excluídas, por exemplo, falha de manejo ou, as perdas causadas por pragas e doenças, plantio fora do período previsto pelo Zoneamento Agrícola de Risco Climático para uma determinada cultura;
- IV) Coberturas Adicionais: Alguns seguros oferecem a possibilidade de contratação de coberturas adicionais à cobertura básica, tais como cobertura de replantio (FAEP, p. 10, 2016).

3.2.1 Programa de garantia da atividade agropecuária (PROAGRO) e Programa de subvenção ao prêmio do seguro rural (PSR)

O Programa de Garantia da Atividade Agropecuária (PROAGRO), instituído pela Lei Nº 12.058, de 13 de outubro de 2009, visa atender aos pequenos e médios produtores e tem por objetivo exonerar o beneficiário do cumprimento de obrigações financeiras em operações de crédito rural de custeio, no caso de perdas das receitas em consequência de fenômenos naturais fortuitos (chuva excessiva, geada, granizo, seca, variação excessiva de temperatura, ventos fortes, ventos frios, doença ou praga sem método difundido de combate, controle ou profilaxia, técnica e economicamente exequíveis); indenizar os recursos próprios do beneficiário, utilizados em custeio rural, inclusive em empreendimento não financiado, no caso de perdas das receitas em consequência fenômenos naturais fortuitos; promover a utilização de tecnologia, obedecida a orientação preconizada pela pesquisa. Este programa atua de forma semelhante a um programa securitário (BNDS, 2011), embora atue de forma desvinculada das seguradoras e da venda de apólices (CUNHA: ASSAD, 2001).

O Programa é custeado por recursos alocados pela União e por recursos provenientes de taxas pagas pelos produtores ao aderirem ao programa, denominado de adicional. Além disso, o PROAGRO é administrado pelo Banco Central do Brasil e operado por seus agentes, representados pelas instituições financeiras autorizadas a operar em crédito rural, as quais contratam as operações de custeio e se encarregam de formalizar a adesão do mutuário ao programa, da cobrança do adicional, das análises dos processos e da decisão dos pedidos de cobertura, do encaminhamento dos recursos à Comissão Especial de Recursos (CER),

dos pagamentos e registros das despesas. A Secretaria Executiva do CER está ligada ao Ministério da Agricultura. O Banco Central do Brasil age com rigor para efetuar as indenizações em caso de ocorrência de sinistros, considerando fielmente as normas do Manual de Crédito Rural (BRASIL, 2018).

O Programa de Subvenção ao Prêmio do Seguro Rural (PSR) é um dos pilares da política agrícola brasileira, com o objetivo de reduzir o prêmio pago pelo segurado e elevar o valor recebido pela asseguradora, viabilizando assim o desenvolvimento desse mercado (BNDES, 2011). Em tese, esse programa pode atender qualquer produtor rural que tenha interesse em contratar quaisquer modalidades do seguro rural (seguro agrícola, aquícola, florestal, penhor rural, benfeitorias e produtos agropecuários, seguro de vida e do produtor rural, seguro de cédula do produtor rural). Diferente do PROAGRO, o PSR possui relação direta com as seguradoras licenciadas pelo governo, sendo regulado pela Superintendência de Seguros Privados (SUSEP). As companhias seguradoras habilitadas no PSR são: Aliança do Brasil, Allianz; Essor, Excelsior; Fairfax; Mapfre; Markel, Nobre; Porto seguro; Sancor e Swiss Re (MAPA, 2019c).

Diferentemente de outras modalidades de seguros, nos seguros agrícolas existe o risco de catástrofe climática. Nessa condição, um evento pode ocasionar perdas em muitas propriedades localizadas em uma mesma região. Esse fato, dentre outras características da agricultura, torna o seguro agrícola mais caro. Por tais razões, nos países em que o seguro agrícola já está a mais tempo consolidado, é comum a existência de programas de apoio dos governos para que os produtores contratem a proteção para as suas lavouras, pagando parte do custo do seguro. A subvenção é destinada ao produtor rural, mas a solicitação da subvenção é feita por meio da própria seguradora, que submete as apólices contratadas à apreciação do Ministério da Agricultura, Pecuária e Abastecimento (MAPA). Este, por sua vez, avalia a situação cadastral do produtor e, não havendo restrição e disponibilidade de recursos, concede o benefício liquidando parte do prêmio (FAEP, 2016).

Os percentuais variam de acordo com a cultura, tipo de cobertura e tipo de produto. No caso dos grãos, o percentual pode ser de 30%, 35% ou 40%, a depender da cobertura contratada na apólice. Para as demais atividades e modalidades o percentual é fixo em 35%. Existe ainda um limite financeiro anual para cada produtor, em termos de subvenção, na modalidade agrícola, esse limite financeiro é de R\$ 72.000,00, enquanto nas demais modalidades o limite é de R\$ 24.000,00 (pecuário, aquícola e de florestas) (MAPA, 2019b).

3.2.2 Indicadores do porte do seguro rural no Brasil

O Ministério da Agricultura, Pecuária e Abastecimento dispõe a série histórica das informações a respeito do seguro rural no Brasil através do portal Atlas do seguro rural (MAPA,

2019c). Além de inúmeras informações, no portal é possível observar a série histórica dos valores totais de subvenção no país desde 2006 (Figura 1), nela pode ser observado um grande crescimento até 2014, em que se encontra o maior valor (R\$ 689,10 milhões); após, nota-se um sensível decréscimo desses valores, provavelmente pela crise financeira que o país atravessou nesse período. Seguindo-se para os anos de 2015, 2016, 2017 e 2018, o montante da subvenção do prêmio do seguro rural foi de R\$ 276,93 milhões; R\$ 386,83 milhões; R\$ 365,22 milhões; R\$ 368,68 milhões, respectivamente, ou seja, a contratação seguro rural no Brasil tem se mantido relativamente constante desde então.

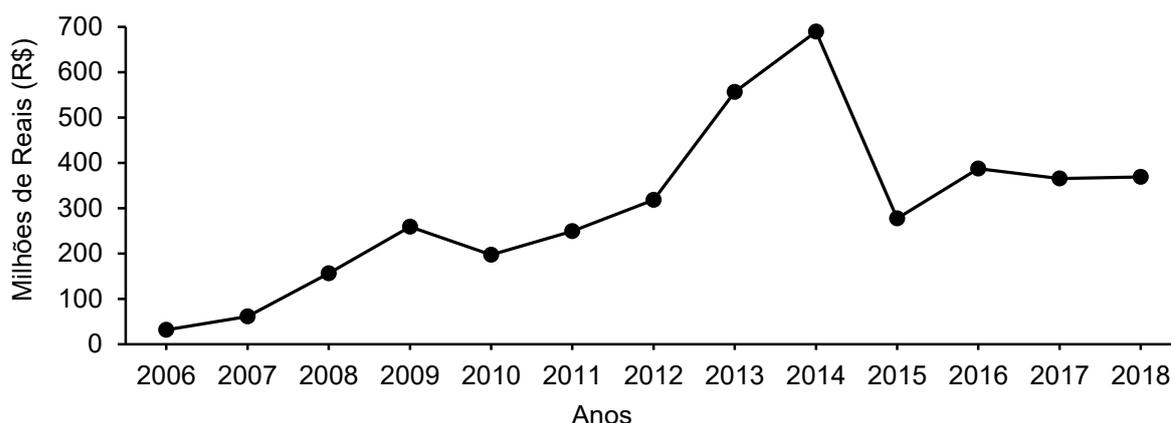


Figura 1 Valores totais de subvenção ao seguro rural no Brasil no período de 2006 a 2018.

Fonte: MAPA (2019c).

De 2015 a 2018, a demanda maior foi dos produtores de grãos, totalizando 75,53% (R\$ 1.055 bilhões) dos investimentos desses 4 anos, sendo a soja a cultura com maior valor de subvenção (R\$ 542,48 milhões); seguida do milho segunda safra com R\$ 263,20 milhões e trigo com R\$ 180,85 milhões. Nesse período, essas três culturas representaram 93,45% dos investimentos governamentais nesta categoria. Em seguida, a categoria “frutas” foi a que recebeu mais recursos, totalizando no mesmo período de 4 anos 18,97% da subvenção (R\$ 265,08 milhões), sendo maçã, uva e pêssego as atividades com maiores valores envolvidos: R\$ 134,50 milhões; R\$ 96,63 milhões e R\$ 12,95 milhões, respectivamente (MAPA, 2019c).

Observa-se que os valores são subvencionados entre todas as unidades da Federação (Figura 2). De um total de quase 4 bilhões de reais em subvenção no período de 2006 a 2018, o estado do Paraná ficou em primeiro lugar entre os estados com 29,64% desse valor, totalizando R\$ 1.160.201.637 milhão; seguido pelo Rio Grande do Sul com 22,24% (R\$ 870,72 milhões); São Paulo com 12,69% (R\$ 496,62 milhões); Santa Catarina com 8,74% (R\$ 342,28 milhões) e Mato Grosso do Sul com 8,66% (R\$ 339,12 milhões) (MAPA, 2019c).

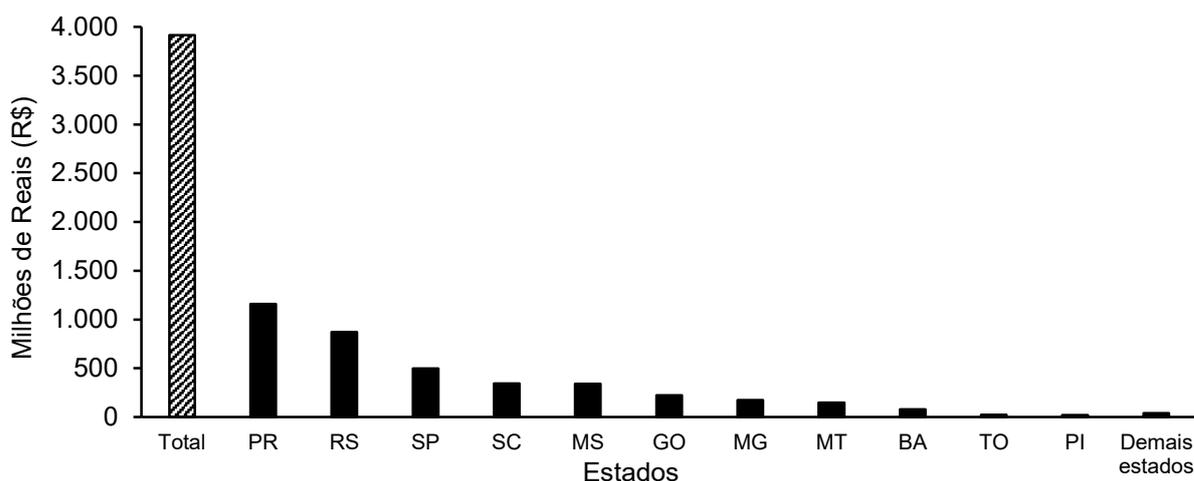


Figura 2 Valores totais de subvenção ao seguro rural por estado no período de 2006 a 2018.

Fonte: MAPA (2019c).

No ano de 2018, a adesão ao PSR mostra novamente que o estado do Paraná é a federação com maior subvenção rural, concentrando 30,26% do total, sendo a produção de soja a atividade com maior adesão ao seguro, em função do maior risco de sinistros na região, além da valorização da cultura (SILVA; TEIXEIRA; SANTOS, 2014). Notadamente, o estado se destaca na produção das *commodities*, pois o Paraná possui a maior produtividade (kg ha^{-1}) e a segunda maior área semeada (ha) e de produção (toneladas) de soja até a safra 2016/2017 (IBGE, 2017).

Na Figura 3 são apresentados os valores totais subvencionados pelo governo, comparativamente com os prêmios pagos pelos produtores, acumulados ano a ano, entre 2006 e 2018. Nesse período, o prêmio arrecadado foi de R\$ 3.766 bilhões, valor que foi muito próximo do total de aporte com recursos do governo (R\$ 3.914 bilhões) (MAPA, 2019c).

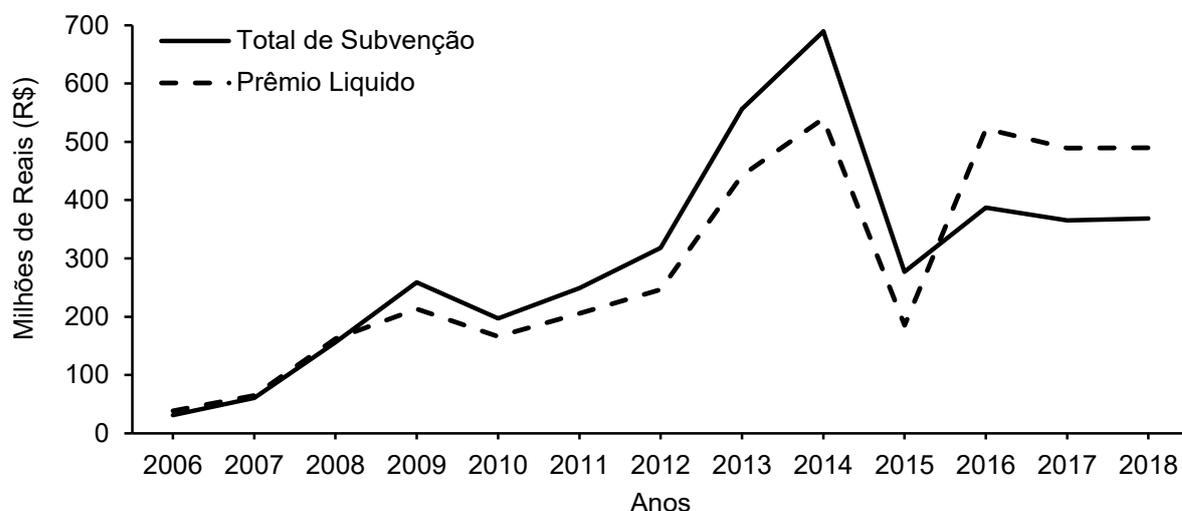


Figura 3 Valores totais de subvenção ao seguro rural e prêmios pagos pelos produtores no período de 2006 a 2018.

Fonte: MAPA (2019c).

Conforme os indicadores do seguro rural no Brasil (MAPA, 2019c), nota-se que, ano a ano, o prêmio líquido foi inferior aos valores submetidos à subvenção até 2015. Depois desse ano, ocorreu uma forte redução orçamentária, em que os valores da subvenção foram inferiores aos prêmios pagos pelos produtores contratantes de seguro. Isso sinaliza uma redução da capacidade governamental em arcar plenamente com despesas desse mercado. O fato de o valor do prêmio pago pelo produtor ter sido insuficiente para liquidar as indenizações indica que a subvenção concedida pelo governo foi importante para que as seguradoras tivessem capacidade de honrar o pagamento junto aos produtores. Ademais, esse fato sugere que a atividade agropecuária possui relativa exposição ao risco e que o seguro rural é importante para reduzir as consequências desses riscos.

O histórico de 2006 a 2018, com foco na área segurada da categoria grãos (Figura 4), demonstra que as culturas com maiores áreas foram a soja, o milho segunda safra e o trigo. As apólices de seguro rural negociadas nessas atividades garantiram a cobertura de 57,65 milhões de hectares, isto é, 86,12% da área total segurada, associados a uma importância segurada de R\$ 122.8 trilhões. Vale salientar que a partir de 2015, houve uma redução da área com milho 1ª safra, ao passo que o milho segunda safra ganhou cada vez mais espaço no campo e nas seguradoras (MAPA, 2019c).

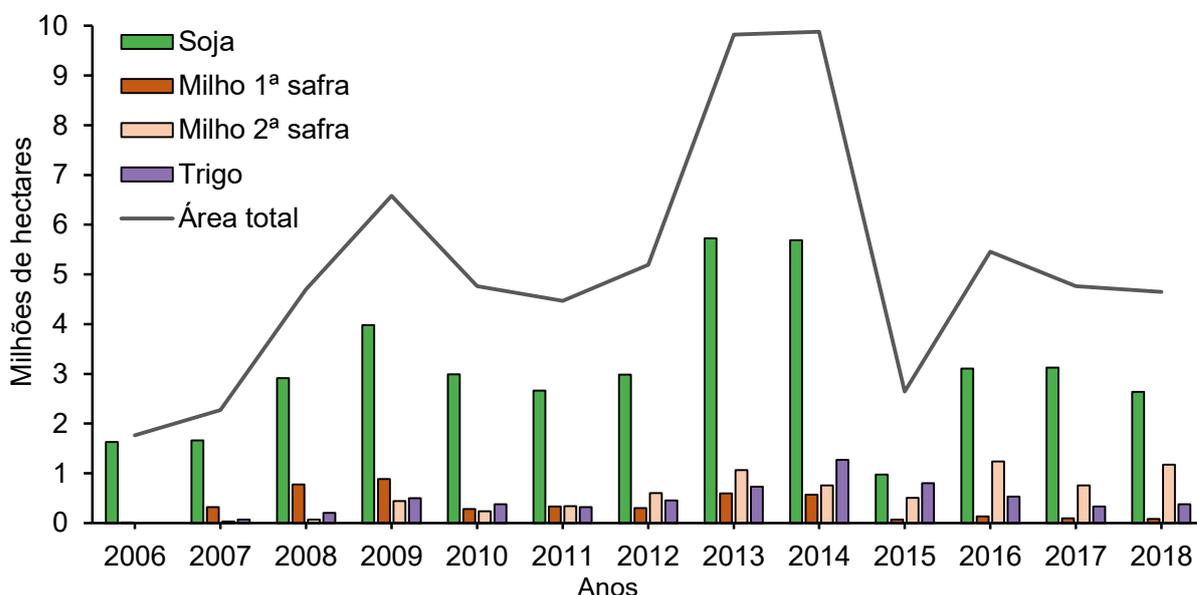


Figura 4 Evolução da área segurada nas principais culturas subvencionadas no Brasil, no período de 2006 a 2018.

Fonte: MAPA (2019c).

O histórico desse período mostra que 555.862 produtores brasileiros adquiriam seguro rural de alguma natureza, totalizando 821.941 apólices. Na Figura 5, observa-se que, em todos os anos da série histórica, ocorrem mais contratações de seguro do que o número de

produtores, em geral cada produtor faz uma ou duas apólices, sendo raros os casos com mais de cinco apólices por produtor (MAPA, 2019c).

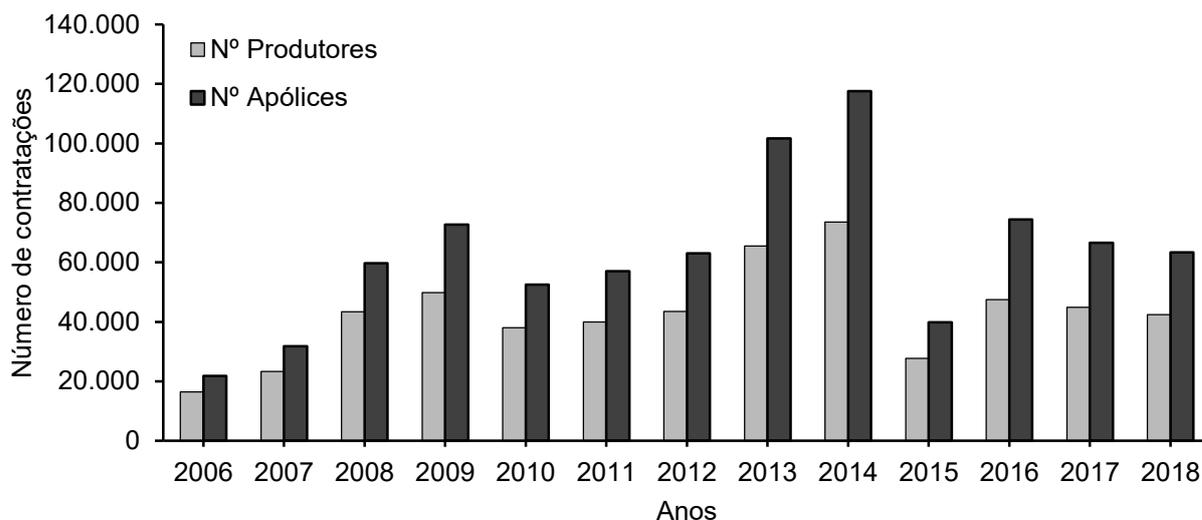


Figura 5 Número de contratações (apólices) e número de contratantes (produtores) de seguro no período de 2006 a 2018.

Fonte: MAPA (2019c).

3.2.3 Fiscalização de sinistros *in loco*

Uma vez implantada a lavoura, espera-se que a condução da safra ocorra dentro da normalidade. Porém, devido a fatores exógenos não controláveis, pode haver a necessidade de que o produtor tenha que ser indenizado em virtude da suspeita de que o rendimento não seja capaz de superar os custos do empreendimento. Para que o produtor possa ser atendido na sua solicitação de cobertura do sinistro ocorrido é preciso certificar-se do cumprimento das suas obrigações estabelecidas na apólice do seguro (MAPA, 2018).

As fiscalizações em caso de sinistro (chuva excessiva, geada, granizo, seca, variação excessiva de temperatura, ventos fortes e ventos frios) são realizadas respeitando-se o seguinte processo: tão logo o produtor identifique perdas relevantes sobre sua lavoura por evento climático coberto pelo seguro, deve comunicar ao seu agente de seguro o mais rapidamente possível para que possa ser providenciada a vistoria (MAPA, 2018).

Nesse momento, a seguradora exige que a área danificada não seja alterada/colhida antes da liberação do perito, caso contrário, a indenização poderá não ser paga (PORTAL TSS, 2015). Em até cinco dias úteis, a contar do recebimento da comunicação de perdas, o agente financeiro indicará um perito para vistoriar a lavoura e fazer a comprovação de perdas. Pode haver ou não uma vistoria preliminar com a cultura ainda “verde”, ou seja, antes da etapa final da colheita para a verificação dos sintomas do evento climático informado.

No local, o perito deve realizar a vistoria de danos com o croqui da área em mãos e qualquer divergência entre o croqui e área apresentada será relatada no laudo. Após a

identificação da área, o perito avalia aspectos como a cultivar (semente própria ou não), datas de plantio, tipo de solo, área e tempo de cultivo, estágio fenológico no momento do sinistro e da inspeção para verificar se o estágio fenológico condiz com a data de plantio informada e relacionar com o zoneamento agrícola. Dessa forma, subdivide-se a lavoura em glebas para conduzir uma amostragem (manual ou mecanizada), distribuída de forma a contemplar os diversos níveis de danos presentes na lavoura segurada, ou seja, área total, independente da ocorrência de sinistro (MAPA, 2018).

Conforme as normativas internas das seguradoras que regulamentam o processo do seguro, no procedimento de amostragem, devem ser desprezadas as bordaduras da gleba e a carga amostrada deve ser, preferencialmente, pesada, evitando-se determinar a produtividade utilizando como parâmetro a capacidade do tanque graneleiro. por exemplo. Após a amostragem, pesagem, classificação e respectivos abatimentos é possível determinar as prováveis perdas de produtividade de cada gleba e concluir o laudo na presença do produtor rural ou responsável designado por ele. O laudo emitido após esta amostragem é considerado o laudo final, liberando toda a área para ser colhida e a partir do laudo final inicia-se a tramitação para indenização ao produtor. se necessário.

Todo esse processo é oneroso e pode ser muito lento dependendo da situação e tamanho da propriedade rural, aliado ao fato de ocorrer uma alta demanda de serviço no mesmo período (colheita em uma determinada região). Outro fator que dificulta o cumprimento das vistorias de forma ágil é o número relativamente baixo de peritos vinculados às seguradoras. Além da questão técnica desses profissionais, a questão moral também é posta em xeque por alguns assegurados, quando estes oferecerem alguma vantagem ao perito para que este informe nos laudos rendimentos de produtividade inferiores aos reais, aumentando, assim, sua indenização (MAPA, 2019d). Embora comum, essa prática é ilegal, e muito difícil de ser rastreada.

Conforme já citado, em caso de expectativa de redução de rendimentos, o produtor rural pode acionar a seguradora. Entretanto, isso é muito subjetivo visto que a maioria dos produtores rurais não se baseiam em ferramentas de acompanhamentos mais detalhados como análises climáticas, análise de folhas etc. ou mesmo técnicas mais simples a nível de campo como a estimativa de produtividade de soja e milho (PIONEER, 2019a, 2019b). Tal situação pode ocasionar o acionamento desnecessário da seguradora, já que a produtividade seria pouco afetada. De qualquer modo, é direito do segurado acionar o seguro a qualquer momento e, embora os prazos para envio do vistoriador desde o acionamento, sejam os mesmos para todos os segurados, poderia haver de alguma forma a seleção de áreas prioritárias para visita, caso a adversidade climática seja de intensidade moderada.

3.3 Sensoriamento remoto

Conforme Meneses e Almeida (2012), o sensoriamento remoto pode ser descrito como o conjunto de técnicas de obtenção de informações e imagens dos objetos da superfície terrestre sem que haja um contato físico entre o sensor e o objeto. Dessa forma, essa concepção estabelece base para definir o sensoriamento em três preceitos: i) Não pode existir matéria no espaço entre o objeto e o sensor; ii) A informação do objeto é passível de ser transportada pelo espaço vazio; iii) O elo de comunicação entre o objeto e o sensor é a radiação eletromagnética.

Uma via de obtenção das características dos objetos se dá por meio de sensores orbitais que possuem características distintas, como: a resolução espacial, que define a escala de observação de um objeto na superfície; resolução temporal, definida pelo tempo de revisita de um determinado sensor a um local específico na superfície; resolução espectral, caracterizada pela quantidade de bandas (faixas do espectro eletromagnético) que o sensor é capaz de captar; resolução radiométrica, que permite a distribuição de *pixels* em diferentes níveis de cinza (MORAES, 2015).

No caso do sensoriamento remoto aplicado à agricultura, o alvo de maior interesse para os estudos é a planta. Trata-se de uma ferramenta que possibilita monitorar o crescimento e desenvolvimento das culturas, além de permitir detectar condições de stress que limitam a produtividade. Sendo assim, essa tecnologia permite obter informações espaciais e temporais detalhadas sobre o comportamento dos alvos nos diferentes ambientes (FERNANDES, 2009).

A radiação eletromagnética, ao interagir com o alvo terrestre, se divide em três componentes: a refletida, a absorvida e a transmitida. Cerca 50% da energia que chega até a planta é absorvida, os outros 50% restante são refletidos ou transmitidos, sendo que somente a refletância é utilizada pelo processo de sensoriamento remoto. A resposta espectral da refletância da vegetação refere-se à capacidade da planta em refletir a radiação eletromagnética nos diferentes comprimentos de onda do espectro (MOREIRA, 2005).

3.3.1 Índices de vegetação

Entre as principais contribuições dos dados do sensoriamento remoto para fins agrícolas, estão aquelas relacionadas ao monitoramento vegetativo e a estimativa de parâmetros biofísicos das culturas. Os Índices de Vegetação (IVs) desempenham um importante papel na caracterização do vigor vegetativo e na dinâmica temporal, visto que esses índices se correlacionam com diversos parâmetros da vegetação (área foliar, biomassa,

porcentagem de cobertura verde, produtividade e atividade fotossintética) (FORMAGGIO; SANCHES, 2017).

Os IVs são construídos por transformações lineares de bandas espectrais, que geralmente abrangem as regiões do vermelho e infravermelho próximo do espectro eletromagnético (EPIPHANIO *et al.*, 1996). O processamento digital de imagens fornece ferramentas para análise da imagem por meio de diferentes algoritmos para a construção desses índices, analisando as regiões de maior reflectância da vegetação e destacando as características de interesse na imagem (GANDHI *et al.*, 2015).

Logo, os IVs realçam informações sobre a fitomassa verdes contidas em dados de refletância espectral, visando reduzir efeitos indesejáveis do solo, atmosfera e variações da geometria Sol-Alvo-Sensor (FORMAGGIO; SANCHES, 2017). Porém, ressalta-se que esses índices possuem limitações em função da interação da radiação eletromagnética com a atmosfera, pois a turbidez atmosférica, geralmente, inibe medidas confiáveis de vegetação e pode retardar a detecção de um início de estresse em dosséis (HUETE, 1988).

Para o monitoramento remoto da atividade agrícola, é imprescindível uma análise temporal para registrar os estágios fenológicos da cultura (HOLBEN, 1986), ou seja, uma análise em múltiplas datas por imagens de satélite, para obtenção de informações do padrão espectro-temporal (perfil) do alvo (JOHANN *et al.*, 2016; BECKER *et al.*, 2017). O sensoriamento remoto possibilita, assim, o monitoramento em diferentes escalas (individual, regional e estadual) de culturas agrícolas, além da redução do custo pela distribuição gratuita das imagens.

3.3.1.1 *Enhanced vegetation index* (EVI)

Diversos trabalhos utilizam o índice de vegetação *Normalized difference vegetation index* (NDVI), proposto por Rouse *et al.* (1973) citados por Motta, Fontana e Weber (2003). Esse índice tem, em geral, associação direta com algumas variáveis agronômicas como altura de planta, índice de área foliar, matéria seca, grau de cobertura de solo e interceptação de radiação solar, entre outros, sendo possível, inclusive, obter avaliações qualitativas e quantitativas dos plantios em períodos anteriores à colheita.

Contudo, esse índice apresenta, como problema, a possibilidade de saturação, ou seja, mantém-se praticamente o mesmo valor de NDVI a partir de índices de área foliar por volta de 4 ou 5, o que o torna insensível ao aumento da biomassa, mesmo quando há aumento da densidade do dossel da cultura (GAMON *et al.*, 1995).

Nesse sentido, o índice de vegetação *Enhanced vegetation index* (EVI – Equação 1), ou índice de saturação realçado, foi desenvolvido como uma otimização destinada a realçar o sinal da vegetação, sendo mais sensível a variação da estrutura foliar, arquitetura do dossel

e a fisionomia da planta, buscando maior sensibilidade para locais com alta fitomassa, reduzindo a influência do solo e da atmosfera (HUETE et al., 1999).

Dessa forma, o EVI tem se mostrado adequado para o monitoramento agrícola, pois substitui, com vantagens, o NDVI, por ser menos suscetível à saturação e mais sensível à variação das características da planta (HUETE et al., 2002).

O índice de vegetação EVI se dá pela seguinte equação:

$$EVI = 2,5 \frac{NIR - RED}{1 + NIR + 6 RED - 7,5 BLUE} \quad (1)$$

em que:

NIR: Banda do infravermelho próximo;

RED: Banda do vermelho;

BLUE: Banda do azul.

3.3.2 Landsat-8

Diversos fatores relacionados à fiscalização de recursos agrícolas podem ser otimizados com técnicas de sensoriamento remoto. Essas técnicas proporcionam maior segurança em quantificar o nível de dano na cultura no momento do acionamento, auxiliando a aferição do perito a campo, na redução de fraudes e concedendo ao agricultor legitimidade ao seu direito (MACEDO; RUDORFF, 2003).

Para o estudo da viabilidade do uso de sensores remotos com a finalidade de fiscalização de seguros e créditos rurais, um dos satélites que pode ser utilizados é o Landsat-8 que conta com uma média resolução espacial e relativa capacidade de revista. Esse satélite representa a mais longa coleção de dados de sensoriamento terrestre de resolução moderada, baseados em espaço do mundo (MARTIN, 2015). Quatro décadas de imagens fornecem um recurso exclusivo para aqueles que trabalham em agricultura, geologia, silvicultura, planejamento regional, educação, mapeamento e pesquisa sobre mudanças globais (USGS, 2018).

O Landsat-8 possui órbita circular heliossíncrona descendente, com 98,02° de inclinação, a uma altitude de 705 km. Seu horário de imageamento é as 10:00 horas da manhã. Em relação às bandas do sensor, tem-se o pancromático P&B (banda 8) com uma resolução de 15 m. O multiespectral (bandas de 1 a 7 e 9) com uma resolução de 30 m. Tem-se também a banda termal (bandas 10 a 11) com uma resolução de 100 m. A precisão de localização tem uma margem de erro circular de até 12 m nas bandas de 1 a 9. A margem de erro circular é de até 41 m nas bandas 10 e 11. O aparelho tem uma frequência de revista de, aproximadamente, 16 dias dependendo da altitude do local com uma largura de faixa imageada de aproximadamente 185 km (USGS, 2018).

A disponibilização das imagens é gratuita e ininterrupta e os diversos produtos oferecidos proporcionam novas possibilidades para o monitoramento da atividade agrícola em escalas regionais.

3.4 TIMESAT

A extração de parâmetros do perfil espectro-temporal da vegetação é uma das etapas essenciais para compreensão da dinâmica temporal das culturas. A extração dos parâmetros pode ser realizada pelo *software* TIMESAT (EKLUNDH; JÖNSSON, 2015a, 2015b), que permite a aplicação de filtros de suavização *Savitzky-Golay*, (SAVITZKY; GOLAY, 1964), *Gaussian Asymmetric* (JÖNSSON; EKLUNDH, 2002), *Double Logistic* (BECK et al., 2006) e, com base em uma parametrização pré-estabelecida (JOHANN et al., 2016), faz-se a extração de 11 parâmetros sazonais do perfil espectro-temporal das culturas (Tabela 1).

Tabela 1 Informações de saída do TIMESAT, referentes ao perfil espectro-temporal de IV das culturas

Variável	Descrição	Sigla
<i>Start</i>	Data de semeadura	DS
<i>End</i>	Data da colheita	DC
<i>Length</i>	Ciclo da cultura	CI
<i>Base val.</i>	Valor do índice de vegetação de base	BASE
<i>Peak t.</i>	Data de máximo desenvolvimento vegetativo	DMDV
<i>Peak val.</i>	Valor do índice de vegetação na DMDV da cultura	MAX
<i>Ampl.</i>	Diferença entre o valor do índice de vegetação no pico vegetativo e o valor do índice de vegetação de base	AMP
<i>L. deriv.</i>	Taxa de crescimento do início do ciclo da cultura, calculada pela razão de diferença entre os níveis 20 e 80%, pelo tempo decorrido entre estas medidas	L-DER
<i>R. deriv.</i>	Taxa de decréscimo no fim do ciclo da cultura, calculada pela razão da diferença entre os níveis 80 e 20%, pelo tempo decorrido entre estas medidas	R-DER
<i>L. integral</i>	Área sobre o perfil de índice de vegetação até o eixo X do gráfico, desde a data de semeadura até a data da colheita	L-INT
<i>S. integral</i>	Área sobre o perfil de índice de vegetação até o valor do índice de vegetação de base	S-INT

Fonte: Adaptado de Eklundh e Jönsson (2015a).

O procedimento para extração de informações pode ser feito individualmente, para pequenos bancos de dados, ou automaticamente, para grandes bancos de dados. Para automatizar o procedimento, o TIMESAT configura um arquivo com as parametrizações gerais (*settings file*), o qual é aplicado para todo o conjunto de dados (EKLUNDH; JÖNSSON, 2015a, 2015b; SHAO et al., 2016).

Na Figura 6, são expressos visualmente nove dos onze parâmetros sazonais do perfil espectro temporal para uma cultura de verão. Nessa figura os dados se referem a imagens EVI/MODIS e o significado de cada sigla pode ser observado na Tabela 1.

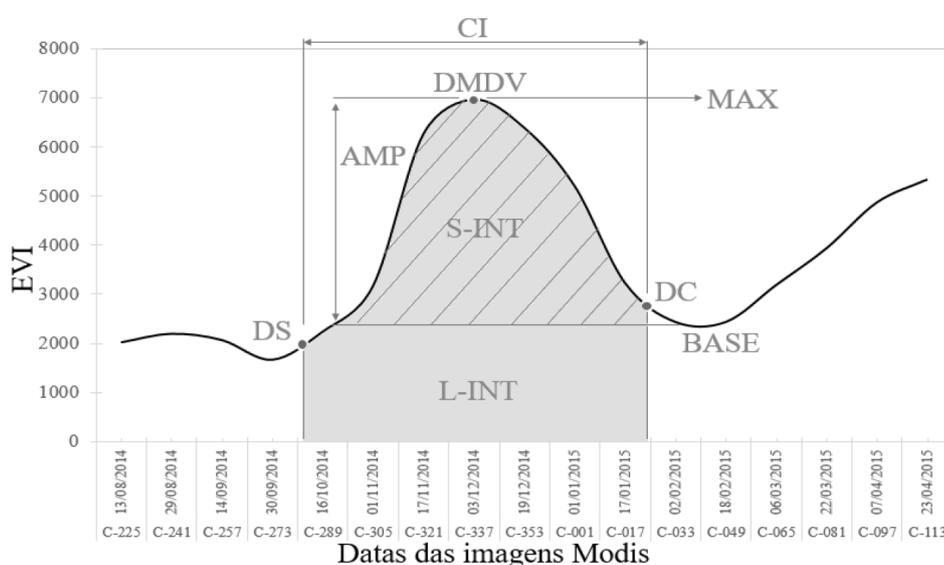


Figura 6 Padrão de um perfil espectro-temporal de índice de vegetação EVI para uma cultura de verão.

Fonte: Johann et al. (2016).

Estudos indicam o potencial desta ferramenta para a extração de informações sobre o padrão espectro-temporal das culturas e a utilização de tais informações nas mais variadas aplicações, desde a distinção de culturas à previsão de datas de colheita (JOHANN, 2012, GRZEGOZEWSKI et al., 2013, JOHANN et al., 2016). Mishra e Chaudhuri (2015) utilizaram o TIMESAT para a extração de informações para aplicação no que chamaram de técnica de análise de padrões sazonais (*Seasonal trend analysis – STA*), ou seja, um método que consiste em identificar os pontos inicial e final dos ciclos sazonais de cultivo e calcular métricas fenológicas a partir deles (JOHANN et al., 2016). Porém, até o momento, não se observaram estudos com o TIMESAT relacionados a sinistros agrícolas.

3.5 Mineração de dados

Uma linha de estudo que vem sendo recentemente abordada para superar os desafios na melhoria dos mapeamentos e classificações é a utilização de mineração de dados (GALVÃO; MARIN, 2009; SOUZA et al., 2010; NONATO; OLIVEIRA, 2013), que é a principal etapa do processo de descoberta de conhecimento em banco de dados (*Knowledge discovery from data – KDD*) e objetiva encontrar relações ou padrões ocultos em bases de dados,

convertendo dados brutos em informações úteis (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; TAN; STEINBACH; KUMAR, 2009).

Para Fayyad, Piatetsky-Shapiro e Smyth (1996), a descoberta do conhecimento é uma sequência de processos iterativos que seguem os seguintes passos: pré-processamento (limpeza, integração, seleção), transformação dos dados (extração, normalização), mineração (técnicas e métodos de descoberta de conhecimento), avaliação e interpretação (avaliação dos padrões obtidos e apresentação do conhecimento) (Figura 7).

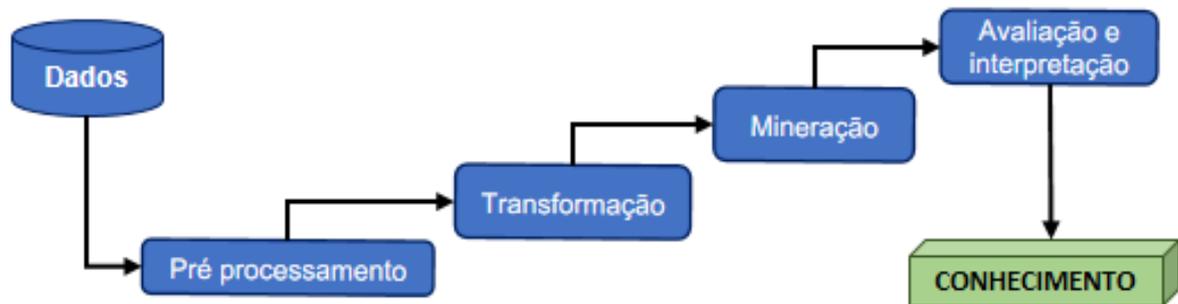


Figura 7 Processo de descoberta de conhecimento em banco de dados – KDD.

Fonte: Adaptado de Fayyad, Piatetsky-Shapiro e Smyth (1996).

As tarefas de mineração de dados são geralmente divididas em duas categorias principais: tarefas de predição e de descrição (Figura 8). As tarefas de predição objetivam prever o valor de um determinado atributo baseado nos valores de outros atributos e as tarefas descritivas objetivam derivar padrões (correlações, tendências, grupos, trajetórias e anomalias) que resumam os relacionamentos subjacentes nos dados (TAN; STEINBACH; KUMAR, 2009).

As tarefas de previsão se subdividem em: classificação, usada para variáveis discretas; regressão, usada para variáveis contínuas. As tarefas de descrição se subdividem em: associação (extração de padrões), agrupamento (agrupa conjuntos de dados similares) (TAN; STEINBACH; KUMAR, 2009) e sumarização (descrição compacta para um conjunto de dados) (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

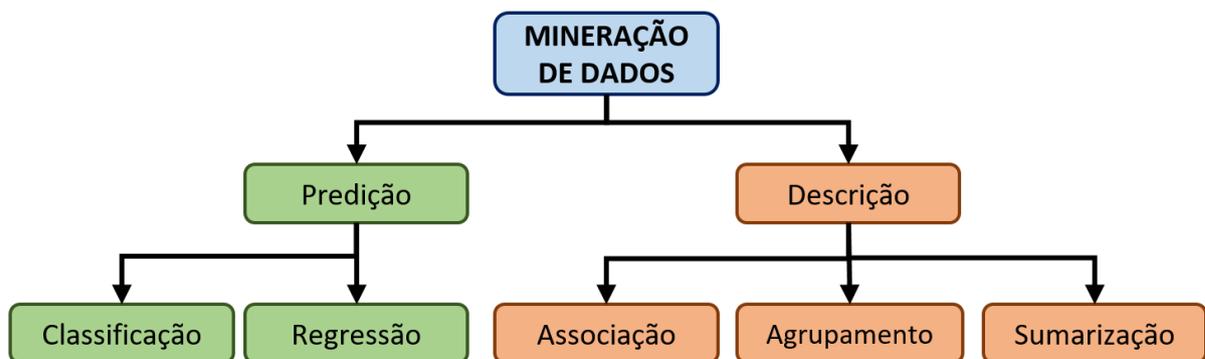


Figura 8 Tarefas de mineração de dados.

Fonte: Adaptado de Fayyad, Piatetsky-Shapiro e Smyth (1996) e Tan, Steinbach e Kumar (2009).

3.5.1 Support vector machine (SVM)

O *Support vector machine* (SVM), também conhecido como Máquina de suporte vetorial, foi proposto por Boser, Guyon e Vapnik (1992). O SVM consiste numa técnica computacional de aprendizado supervisionado de reconhecimento de padrão, cujo objetivo é classificar determinado conjunto de dados que são mapeados para um espaço de características multidimensionais, usando um conjunto de funções matemáticas chamadas *kernels*. Nessa abordagem, o limite de decisão no espaço de entrada é representado por um hiperplano em dimensão superior no espaço (VAPNIK et al., 1997).

A classificação SVM é baseada no princípio da separação ótima entre classes, tal que se as classes são separáveis, a solução é escolhida de forma a separar ao máximo as classes utilizando o conceito de planos de decisão (VAPNIK, 1995).

O objetivo do modelo do SVM é encontrar um hiperplano ótimo, no qual seja possível classificar corretamente todos os pontos de treinamento e, além disso, com a maior margem de separação possível para os vetores da classe -1 e da classe $+1$. O objetivo é obter um hiperplano com a melhor generalização e robustez, dado que existem muitas possibilidades de separação dos conjuntos. A (Figura 9) apresenta três possibilidades de separação dos conjuntos de dados. Sendo que nas Figuras (9a) e (9b) são apresentadas separações quaisquer, sem a necessidade da máxima distância entre o hiperplano separador e os dados. Já para a Figura (9c) é apresentado o gráfico do hiperplano separador ótimo.

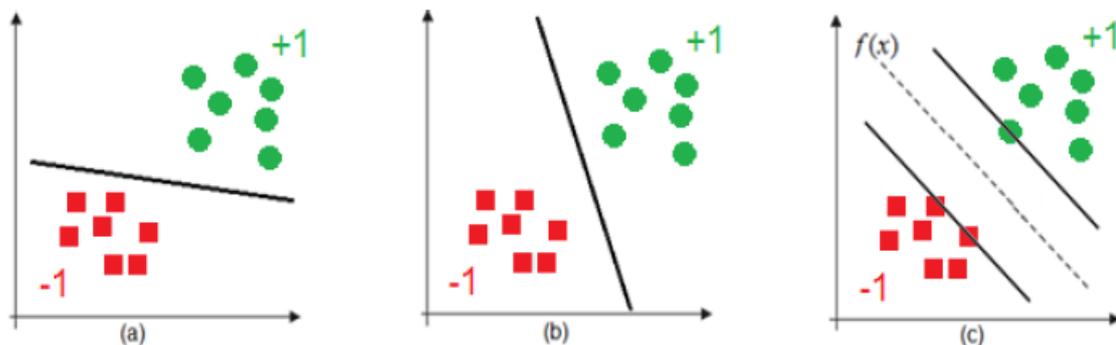


Figura 9 Construção de hiperplanos de separação dos dados, de tal forma que maximizem as distâncias entre os elementos de cada grupo.

Fonte: Abreu (2016).

O SVM possui grande abrangência de aplicações em diversas áreas, como finanças, biologia, medicina e logística, entre outras. Isso ocorre devido às suas vantagens de aplicação, como: bom desempenho de generalização, tratabilidade matemática, interpretação geométrica e a utilização para a exploração de dados não rotulados (SARADHI; KAMIK, MITRA, 2005).

3.5.2 *Random forest*

O algoritmo *Random forest* (RF) consiste num conjunto de árvores de decisão geradas dentro de um mesmo objeto, em que cada objeto (conjunto de árvores) passa por um mecanismo de votação (*bagging*) que elege a árvore mais votada (BREIMAN, 2001). Tal método, é uma combinação de preditores de árvores, de modo que cada árvore depende dos valores de um vetor aleatório amostrado independentemente e com a mesma distribuição para todas as árvores na floresta. O método RF é um algoritmo de aprendizado supervisionado que usa aprendizado por conjuntos para classificação e regressão (JAMES et al., 2014). Ele também pode ser utilizado como um método embutido de seleção de características (MENZE et al., 2009).

Para ranquear a importância das características, o RF considera duas informações: *Mean square error, MSE - %IncMSE* e *IncNodePurity*. O *%IncMSE* é calculado considerando os dados permutados. Para cada árvore o erro médio quadrático de previsão é registrado depois de permutar cada variável preditora. A diferença entre os nós é calculada sobre toda a árvore e normalizado pelo desvio padrão das diferenças. A importância da variável é diretamente proporcional a essa diferença. Já o *IncNodePurity* refere-se à diminuição total das impurezas no nó da variável em relação à média de toda a árvore.

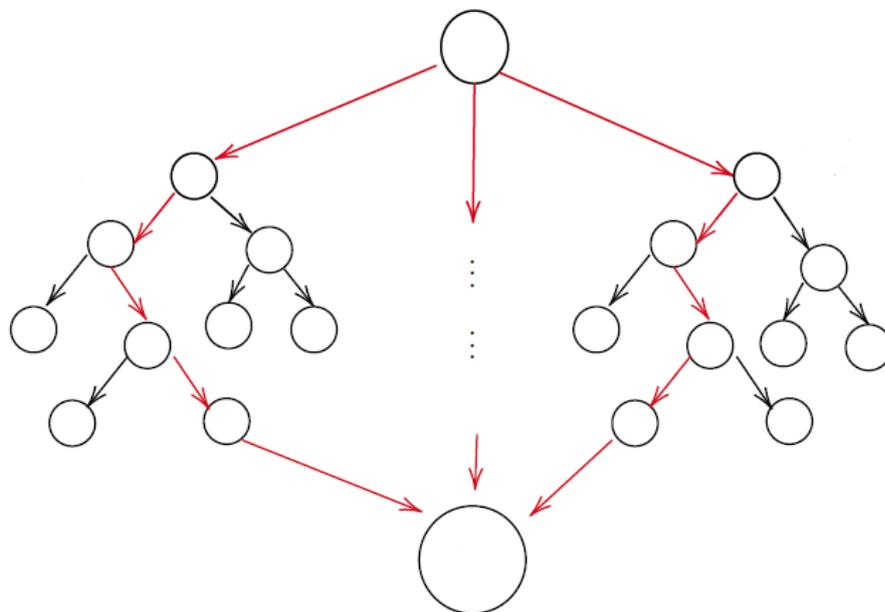


Figura 10 Representação genérica do modelo *Random forest* com suas diversas árvores.

Fonte: Adaptado de Breiman (2001).

3.5.3 Árvore de decisão

As árvores de decisão (também chamadas de árvores de classificação ou de regressão) do inglês *Decision tree* (DT), são técnicas de mineração de dados utilizadas para classificação e predição das amostras conhecidas através de aprendizado de máquina, ou seja, dados de treinamento (CRIVELENTI et al., 2009). Para tal, efetua-se a construção de padrões com os dados de treinamento em, a partir da árvore obtida, pode-se classificar novas amostras desconhecidas. Tais árvores consistem em uma hierarquia de nós internos e externos que são conectados por ramos. Cada nó representa uma decisão sobre uma variável através de teste lógico, que ramifica para o próximo nó descendente ou resultado final, chamada folha (Figura 11).

As árvores de decisão apresentam como vantagens a versatilidade e elevada legibilidade, permitindo identificar os fatores mais influentes, além de permitir a classificação de uma amostra desconhecida sem a necessidade da análise de todos os atributos (NONATO; OLIVEIRA, 2013; CRIVELENTI et al., 2009).

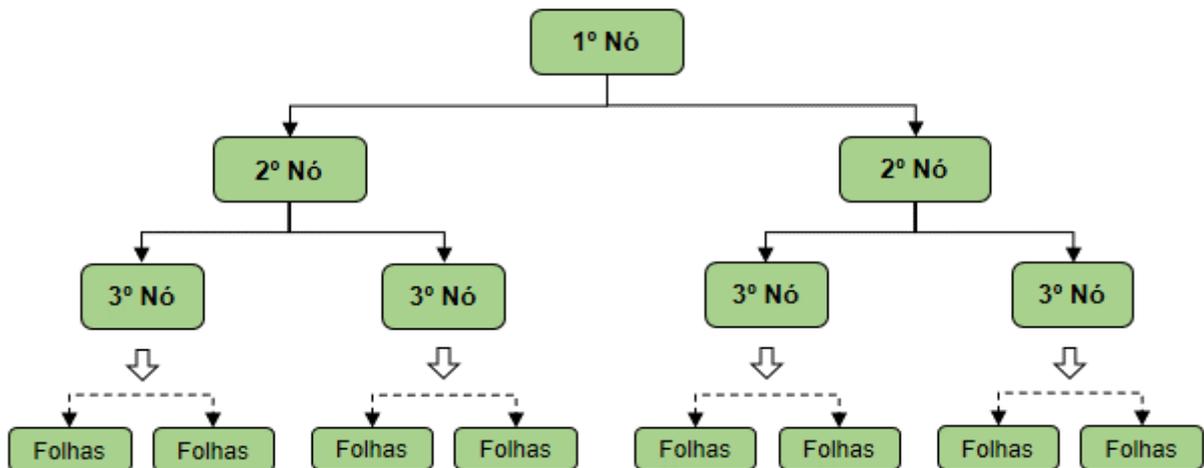


Figura 11 Ilustração da hierarquia da árvore de decisão, no qual os nós representam um teste lógico, conectados por tomadas de decisão (nó/ramo), obtendo um resultado final (folha).

Fonte: Autor (2020).

4 MATERIAL E MÉTODOS

A execução deste estudo foi dividida em quatro etapas: aquisição de dados, pré-processamento, processamento e validação, representadas no fluxograma (Figura 12).

A etapa inicial refere-se à construção do banco de dados, incluindo as subetapas aquisição de imagens de satélite, dados climatológicos, dados reais de sinistros agrícolas (laudos de perícias) e de áreas não sinistradas consideradas normais. Posteriormente, o pré-processamento contemplou a vetorização das áreas (descritas nos laudos e sem sinistro), filtragem de imagens por alvos agrícola, ajuste da granulometria dos dados climáticos, tabulação e, geração do perfil espectro-temporal das áreas. O processamento abrangeu a extração de métricas do perfil espectro-temporal e dos dados climáticos, categorização de perfis das áreas sinistradas e normais. Por fim, na etapa de validação, os dados obtidos foram submetidos ao processo de mineração de dados afim de verificar o padrão existente entre os parâmetros climáticos de EVI e os sinistros, visando indicar a ocorrência ou não de sinistros e, também, qual tipo de sinistro incidente, comparando com os laudos originais.

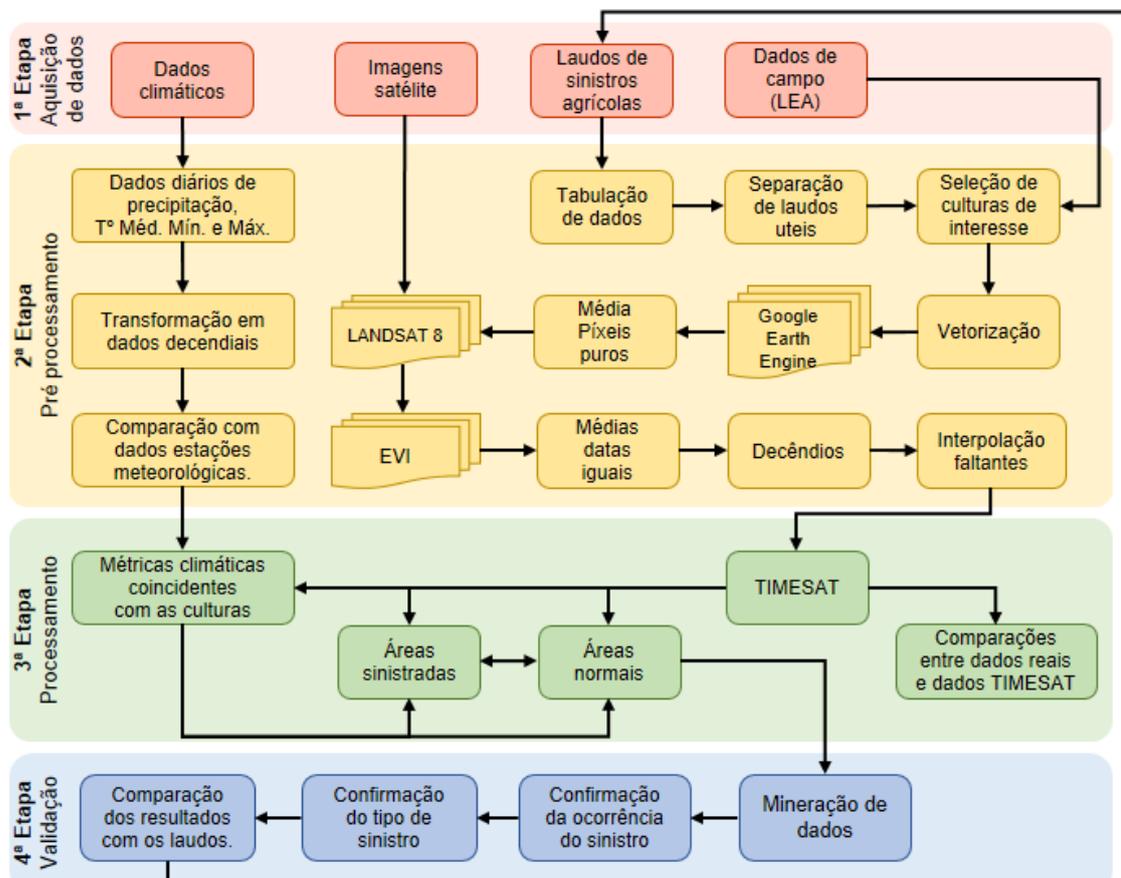


Figura 12 Fluxograma geral da dissertação com as etapas de aquisição de dados, pré-processamento, processamento e validação.

Fonte: Autor (2020).

4.1 Aquisição de dados

A etapa inicial consistiu na criação de um banco de dados formado por laudos confidenciais de perícias agrícolas, cedidos exclusivamente para este estudo por uma empresa parceira. Para isto, todas as referências aos envolvidos nos processos de sinistros analisados foram mantidas em sigilo. Essas vistorias foram realizadas entre os anos de 2016 a 2018, totalizando 575 laudos (Tabela 2). Os documentos, originalmente, preenchidos a mão, foram digitalizados e tabulados (Anexo A).

Tabela 2 Número de laudos agrícolas por ano da vistoria

Ano	Não amostrados	Informações repetidas	Laudos úteis	Total
2016	34	5	165	204
2017	50	41	122	213
2018	40	20	98	158
Total	124	66	385	575

Fonte: Autor (2020).

Inicialmente, foi realizada uma triagem dos laudos, removendo-se aqueles em que não foi executada a amostragem da área (colheita mecanizada ou manual). Essa decisão decorreu dos seguintes fatos: o segurado declinou do seguro por acreditar que a produtividade da área segurada seria satisfatória e superior a informada na apólice; a área segurada já foi colhida antes da perícia, impossibilitando qualquer análise de produtividade confiável. Visto que a produtividade atesta o quão prejudicial foram as intempéries climáticas na lavoura vistoriada, logo a ausência desta informação inviabiliza certas análises. Por esse motivo, foram descartados 124 laudos. Outro motivo de descarte foram laudos em duplicidade, ou seja, realização de mais de uma vistoria na propriedade, descartando-se 66 laudos entre os três anos. Assim, o número de laudos reduziu para 385.

Em função dos laudos impressos serem digitalizados para formato de imagem, a tabulação foi realizada individualmente no MS Excel® para posterior filtragem de informações. Cada laudo apresentou as seguintes variáveis: número do sinistro (para facilitar a rastreabilidade dos documentos); data da vistoria (data da colheita); município; estado; coordenadas da sede ou ponto dentro do talhão; altitude (m); cultura; tipo cultivo (sistema plantio direto ou preparo convencional do solo); cultura anterior; início sinistro, fim sinistro (datas de início ou término do evento climático em questão); estágio planta na ocasião do sinistro; evento causador do sinistro; plantas por metro linear; espaçamento (m); tipo de solo; zoneamento agrícola (início e fim); cultivar; data de semeadura, área cultivada (ha); produtividade (kg ha^{-1}) de cada talhão.

Os laudos contemplavam informações de dois estados: Paraná e Goiás, entretanto, no estado do Paraná foi observada uma variedade maior de eventos sinistrantes em

comparação ao estado de Goiás que, basicamente, concentrou-se em seca para milho 2ª safra. Logo, optou-se por analisar apenas os sinistros ocorridos no Paraná em função da maior diversidade de culturas pericidas, ou seja, soja, trigo e milho 2ª safra. Situado na região sul do Brasil (Figura 13), entre os paralelos 22°29'S e 26°43'S e os meridianos 48°2'W e 54°38'W, o estado do Paraná é constituído por 399 municípios, subdivididos em 10 mesorregiões. O clima no estado, devido ao seu grande tamanho territorial, enquadra-se em quatro tipos: Aw, Cfa, Cfb e Cwa (APARECIDO et al., 2016).

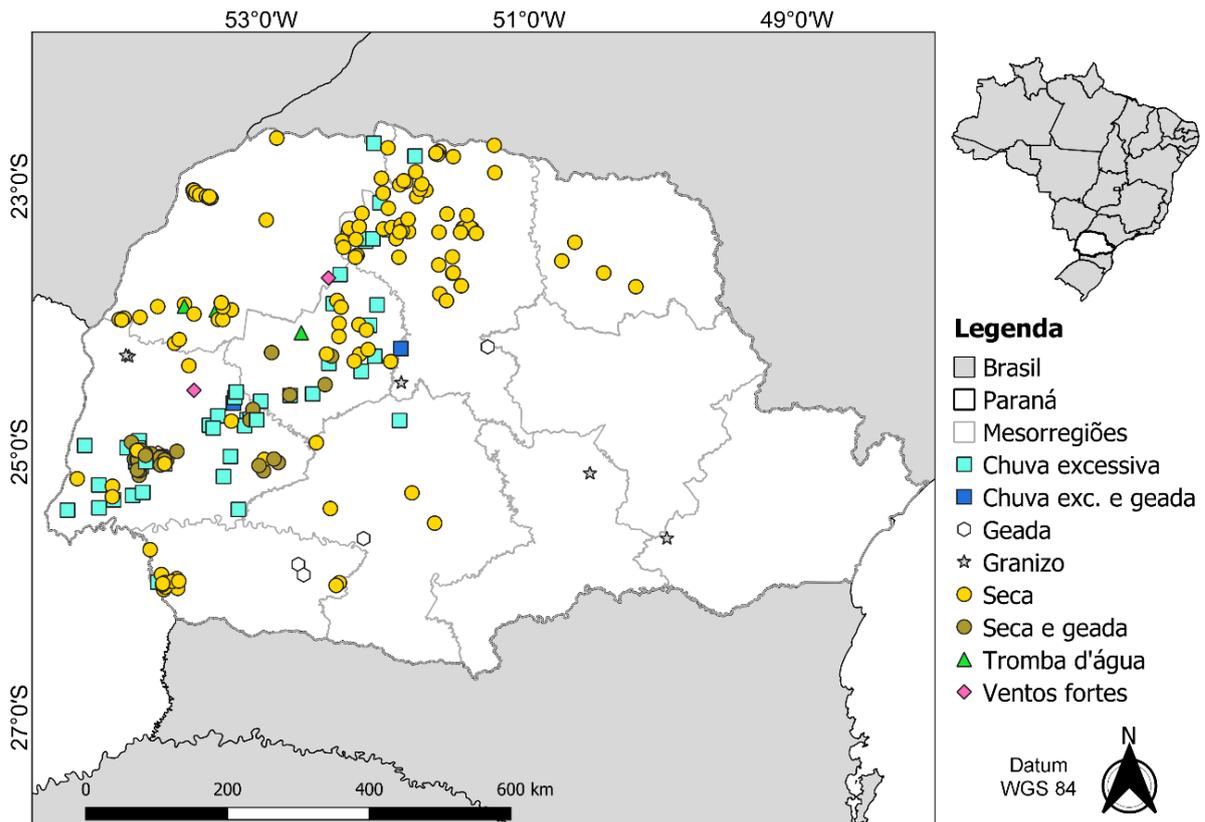


Figura 13 Mapa de localização dos eventos causadores de sinistros nas culturas da soja, milho 2ª safra e trigo, neste banco de dados para o estado do Paraná.

Fonte: Autor (2020).

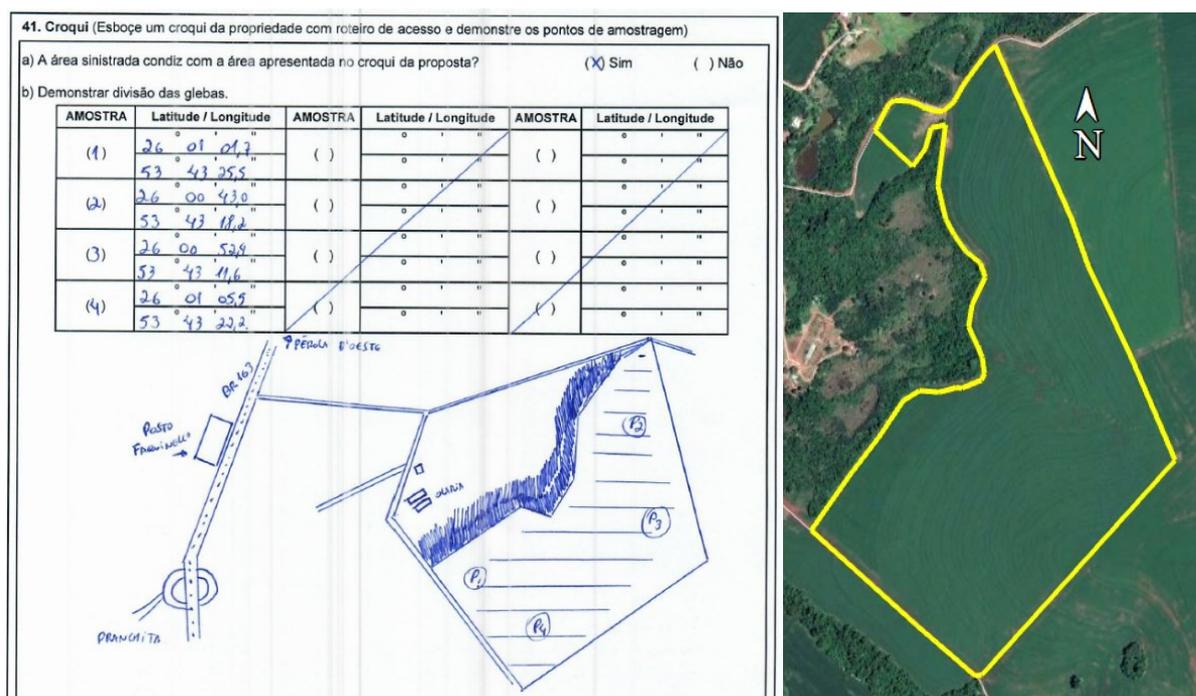
Os sinistros estão distribuídos por todo o Paraná, embora se concentrem mais nas mesorregiões: oeste, sudoeste, noroeste, centro ocidental e norte central (Figura 13). Na Tabela 3, é apresentado o número de eventos climáticos causadores de sinistro por cultura para os anos de 2016 a 2018. Vale ressaltar que em alguns laudos foram informadas as ocorrências de dois eventos climáticos distintos (chuva excessiva e geada, seca e geada). A informação original foi preservada a fim de se obter maior variedade de combinações entre sinistros e culturas.

Tabela 3 Eventos climáticos causadores dos sinistros nas culturas da soja, milho 2ª safra e trigo - Paraná

Evento \ Cultura	Milho		Soja		Trigo		Total
	2017	2018	2015/16	2016/17	2016	2017	
Chuva excessiva	37		15		1		53
Chuva excessiva + geada	2				1		3
Geada	1				4		5
Granizo			4		1		5
Seca		84	5		4	32	129
Seca + geada	1					34	36
Tromba d'água				3			3
Ventos fortes	1	1					2
Total	42	85	24	3	11	66	236

Fonte: Autor (2020).

Os talhões de cada perícia foram vetorizados com base nos croquis informados nos respectivos laudos no *software Google Earth* e transformados em vetores (*shapefile*) no *software QGIS 3.4.3*. Na Figura 14, observa-se um exemplo de croqui original e ao seu lado a área vetorizada na interface do *Google Earth*.

**Figura 14** Croqui elaborado no laudo final (esquerda), vetor da área na interface do *Google Earth* (direita).

Fonte: Autor (2020).

Na Figura 15, visualiza-se a localização das culturas sinistradas e analisadas neste banco de dados, bem como o ano de análise. Nota-se que a maior concentração de laudos está nas regiões oeste e norte do Paraná, sendo estes predominantemente compostos de lavouras de milho 2ª safra.

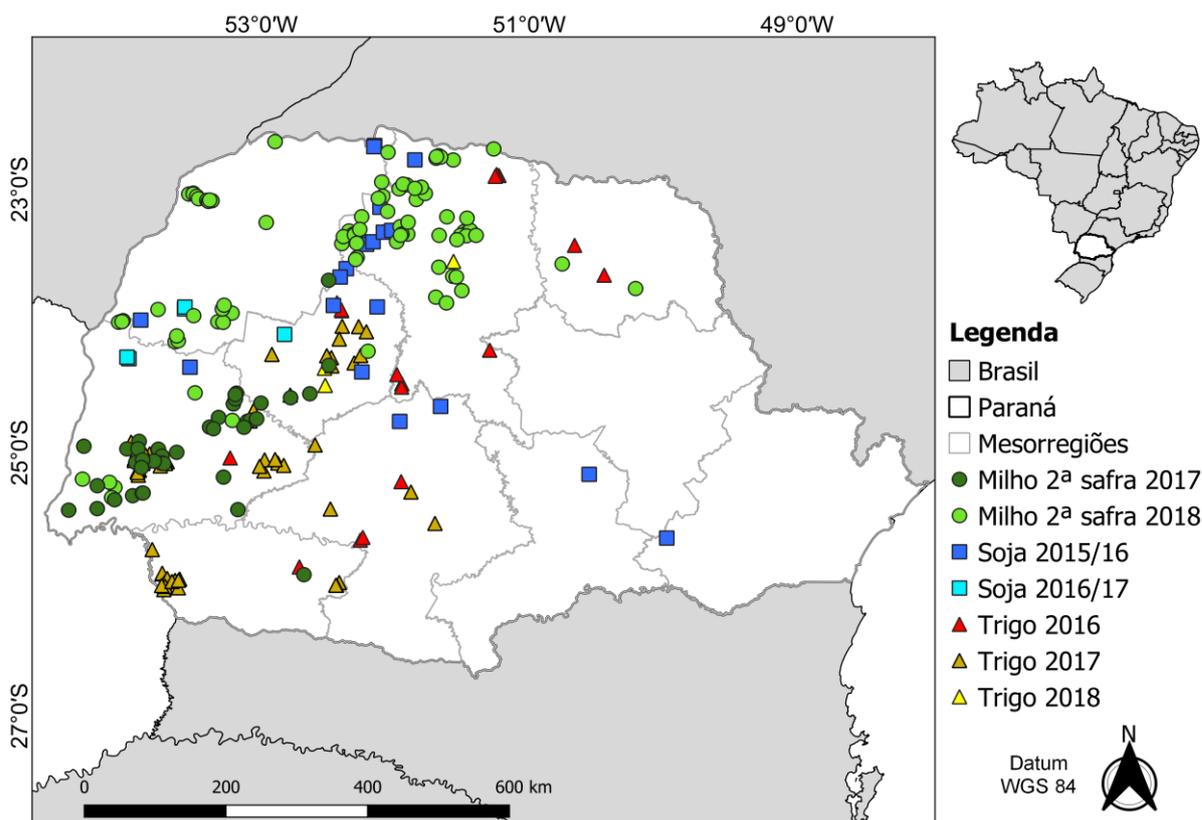


Figura 15 Mapa de localização das lavouras sinistradas de soja, milho 2ª safra e trigo, neste banco de dados para o estado do Paraná.

Fonte: Autor (2020).

Na Figura 16, está demonstrado o local de cada propriedade onde se têm informações de campo. As áreas foram consideradas normais para comparação com as propriedades sinistradas, porém a quantidade de informações dessas áreas é reduzida em comparação às informações obtidas nos laudos de sinistro supracitados, não sendo possível observar com fidedignidade, em muitos casos, as datas de plantio e colheita e os dados de densidade de plantas, cultura anterior, tipo de solo e produtividade, entre outros.

Vale ressaltar que o mapa descreve as lavouras consideradas normais em mapas distintos do Paraná na mesma figura, para evitar sobreposição de informações dos dados nos mesmos locais, com base nos relatos dos produtores em suas áreas em mais de uma safra.

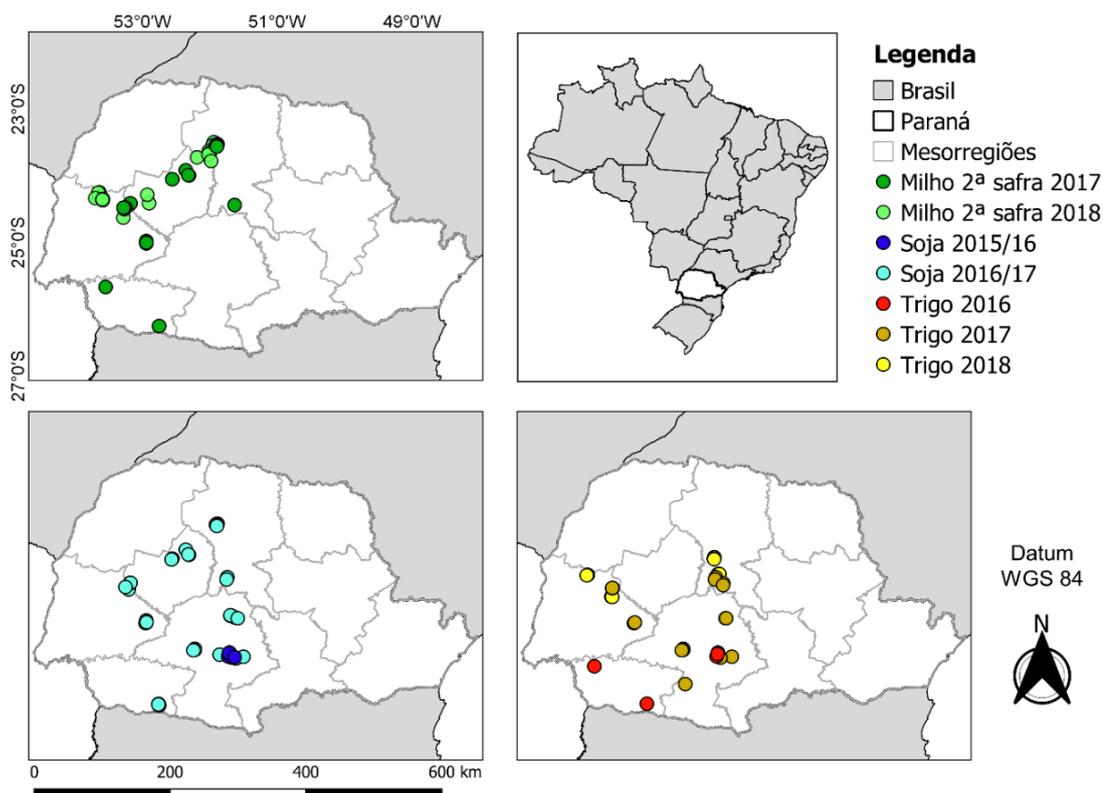


Figura 16 Mapa de localização das lavouras de soja, milho 2ª safra e trigo consideradas normais, neste banco de dados para o estado do Paraná.

Fonte: Autor (2020).

Portanto, ao fim do processo de vetorização, obtiveram-se as áreas das lavouras sinistradas e normais (Tabela 4), com 17.570,71 ha em 245 feições sinistradas e 8.365,84 ha em 138 feições normais, totalizando 25.936,55 ha analisados neste estudo.

Tabela 4 Estatísticas relacionadas à área (hectare) das lavouras vetorizadas de milho 2ª safra e trigo, sinistradas e normais, neste banco de dados para o estado do Paraná

	Cultura	Ano	N	Soma	\bar{X}	Md	Dp	CV%	Mín	Máx	1º Quartil	3º Quartil	IQR
Sinistrada	Milho	2017	42	2737,65	651,82	37,68	644,80	98,92	10,10	312,87	25,45	70,97	45,51
	Milho	2018	87	7415,84	863,78	69,19	644,15	74,57	7,36	358,82	35,07	115,92	80,85
	Soja	15/16	21	1326,65	631,74	4174,00	476,96	75,50	20,92	193,27	26,30	96,79	70,49
	Soja	16/17	5	297,19	594,37	72,21	29,95	5,04	11,81	89,42	37,98	85,78	47,80
	Trigo	2016	16	909,12	568,20	221,18	757,38	133,29	5,06	309,28	153,71	212,49	58,78
	Trigo	2017	69	4457,00	645,94	44,94	561,29	86,90	6,19	265,73	25,49	80,84	55,35
	Trigo	2018	5	427,27	854,53	52,24	826,84	96,76	8,39	228,50	13,65	124,48	110,83
Normal	Milho	2017	24	1039,33	433,06	37,95	282,01	65,12	9,08	119,21	22,24	535,80	313,37
	Milho	2018	19	1510,80	795,16	42,31	100,69	12,66	10,02	455,53	29,49	91,70	62,22
	Soja	15/16	5	692,30	138,46	119,88	48,04	34,70	81,58	216,81	106,88	167,15	60,27
	Soja	16/17	54	3382,55	626,40	34,80	114,70	18,31	4,93	821,14	17,90	57,79	39,89
	Trigo	2016	5	437,59	87,52	106,88	580,20	662,93	6,99	167,15	36,69	119,88	83,19
	Trigo	2017	19	601,81	316,74	20,80	273,64	86,39	4,93	115,20	12,90	37,92	25,03
	Trigo	2018	12	701,46	58,46	54,06	363,14	621,18	9,25	111,77	24,55	92,62	68,07

Notas: N: número de áreas; \bar{X} : média; Md: mediana; Dp: desvio padrão; CV%: coeficiente de variação; IQR: intervalo interquartilico.

Fonte: Autor (2020).

Para compreender os eventos climáticos causadores de sinistro buscou-se uma base de dados climáticos que fosse capaz de fornecer informações em escala e período adequado para cada lavoura. Nesse sentido, em função do potencial de utilização e facilidade de acesso, optou-se em utilizar a plataforma Nasa *Power* (*Prediction of worldwide energy resource*). Essa base de dados está focada na indústria e setores de pesquisa com parâmetros e formatos especificamente direcionados às necessidades de energia, edifícios, indústrias e agricultura.

A plataforma disponibiliza de forma gratuita dados de radiação solar em um *grid* global de 1 grau e dados meteorológicos em um *grid* global de 0,5 grau. Os dados solares são baseados em observações de satélite, enquanto os meteorológicos são extraídos a partir de modelos globais de assimilação. Os dados meteorológicos incluem pressão do ar, temperatura média/ mínima/ máxima do ar, umidade relativa, temperatura do ponto de orvalho, da superfície do solo, temperatura e velocidade do vento (CHANDLER et al., 2013).

Foram selecionadas as coordenadas pontuais coincidentes com as lavouras de interesse para extração dos dados, porém a plataforma também permite que seja processada uma área máxima de 4,5° x 4,5°, o que não se aplica neste estudo, pois o objetivo é a análise pontual das propriedades agrícolas. Os dados diários avaliados foram: precipitação, temperatura média, máxima e mínima a 2 m do solo. As datas inicial e final escolhidas foi de 01/01/2015 a 31/12/2018, contemplando todas as áreas de todas as culturas analisadas. As séries históricas foram extraídas em arquivos “.CSV” individuais, para posterior análise. Na Figura 17, visualiza-se a interface da plataforma e a janela de seleção de dados de interesse.

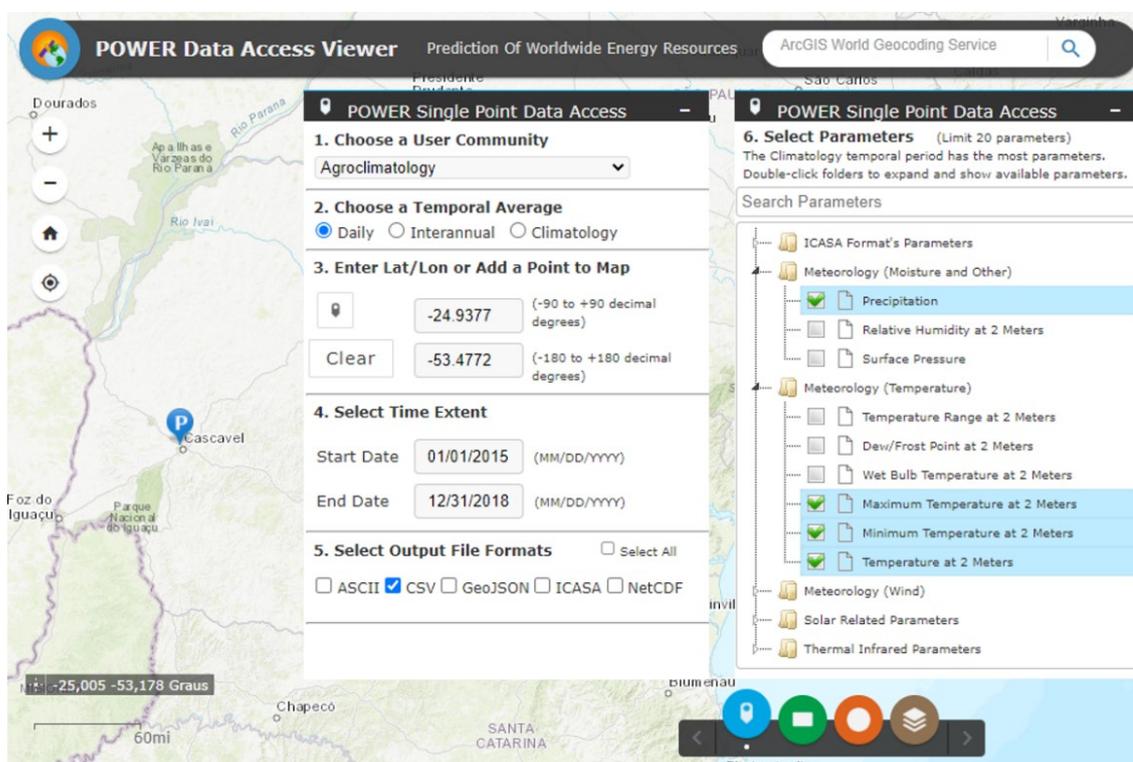


Figura 17 Interface da plataforma Nasa *Power*, com destaque a janela de seleção de dados.

Fonte: Autor (2020).

4.2 Pré-processamento

A aquisição das séries históricas de imagens do sensor Landsat-8, corrigidas para Topo de Atmosfera (*Top of Atmosphere* – TOA) foi realizada pela plataforma *Google Earth Engine* (GEE) por meio de uma rotina em *Javascript*. Para cada cultura e ano, foi selecionado o período que abrangia todo o desenvolvimento da cultura e calculado o índice de vegetação EVI. A partir dessas imagens, extraiu-se apenas a média dos *pixels* puros para cada dia de imageamento referentes a cada área analisada. Resultante desse processo, gerou-se aproximadamente 6 GB de dados contendo os valores de EVI dos períodos estipulados.

Na Tabela 5, informa-se o número de feições, *tiles* e imagens envolvidas nessa etapa.

Tabela 5 Quantidade de dados resultantes da aquisição de imagens de interesse, separados por culturas sinistradas e normais

Classe	Cultura	Ano	Imagens	Tiles	Feições	Total feições por cultura
Sinistrado	Milho	2017	109	6	42	129
	Milho	2018	178	9	87	
	Soja	2015/16	108	8	21	26
	Soja	2016/18	51	3	5	
	Trigo	2016	141	9	16	90
	Trigo	2017	99	5	69	
	Trigo	2018	45	3	5	
Normal	Milho	2017	76	6	24	43
	Milho	2018	86	6	19	
	Soja	2015/16	14	1	5	59
	Soja	2016/18	92	6	54	
	Trigo	2016	24	2	5	36
	Trigo	2017	40	3	19	
	Trigo	2018	42	3	12	
TOTAL			1105	70	383	383

Fonte: Autor (2020).

A extração do perfil-temporal de EVI das imagens das culturas foi executada na plataforma do *Google Colaboratory*, que é uma plataforma em nuvem, a qual disponibiliza uma máquina virtual para o processamento de dados através da linguagem de programação Python e é especialmente adequada para *machine learning* e análise de dados. Mais tecnicamente, o “Colab” é um serviço de *notebooks* hospedados do Jupyter que não requerem nenhuma configuração para usar e oferece acesso gratuito a recursos de computação.

Com uso de um *script* próprio, foram tabulados os valores de EVI para o MS Excel®, em que cada feição corresponde a uma linha e cada coluna corresponde a uma leitura de imageamento realizada pelo Landsat-8/TOA na data descrita.

Nesse processo, algumas colunas do arquivo representam uma mesma data, alterando apenas o número do *path/row* da leitura do satélite, e isso se dá pelo fato de que algumas feições de interesse estão numa região de divisa com sobreposição dos limites entre dois *tiles* como representada na Figura 18, dessa forma em um mesmo dia a feição do

shapefile pode apresentar mais que uma leitura de EVI, sendo necessário calcular a média das leituras para se obter uma leitura de EVI única para cada data.

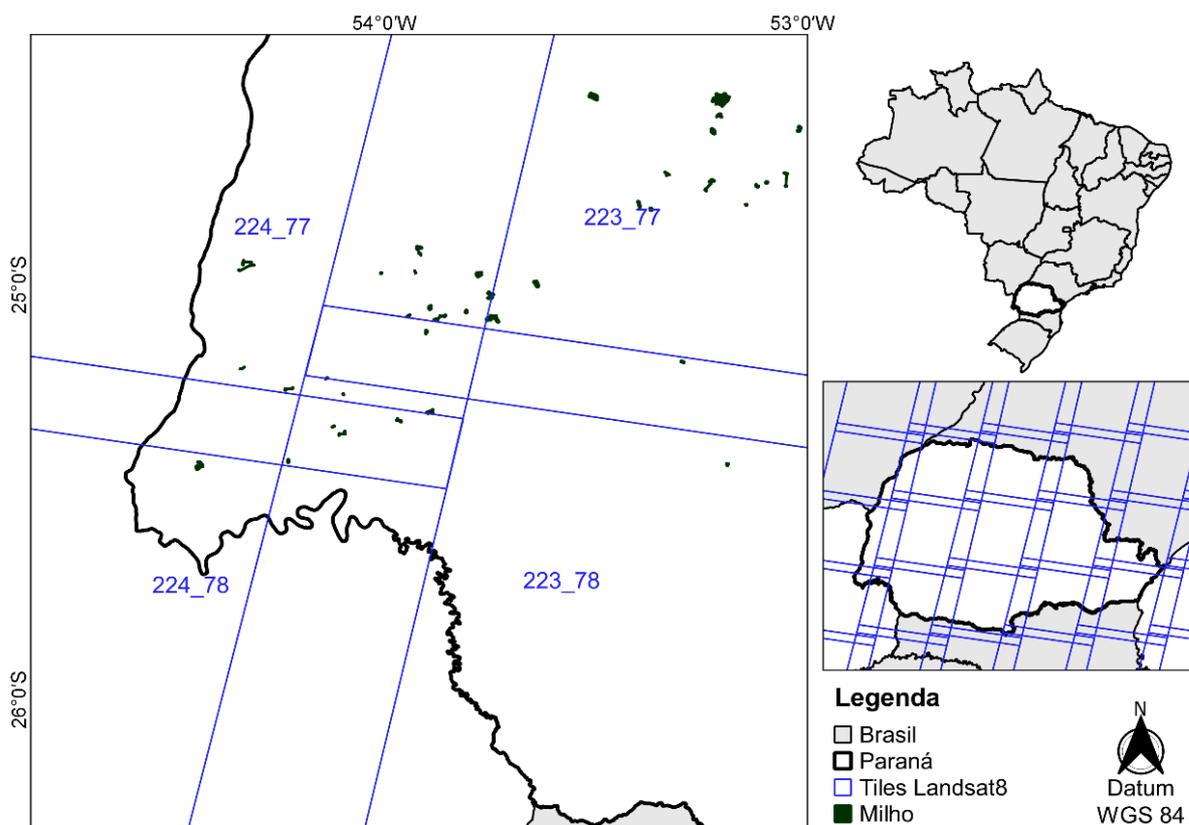


Figura 18 Sobreposição de *tiles* do satélite Landsat 8 e, em destaque, algumas lavouras de interesse na sobreposição das imagens.

Fonte: Autor (2020).

Após a unificação de datas iguais quando necessária, as leituras de EVI que tem a periodicidade de 16 dias foram convertidas para dados decendiais, de forma a facilitar a compreensão das informações. Todavia, essa transformação acarretou dados decendiais de satélite faltantes. Para contornar isso, foi executada uma interpolação linear dos dados de EVI em Python na plataforma *Google Colaboratory*.

Por meio de análises de sensibilidade de interpolações lineares e quadráticas, optou-se pela interpolação linear como ideal para este estudo, como exemplifica a Figura 19. A interpolação linear preserva os dados originais da série e não extrapola os dados faltantes na intenção de suavizar a curva, gerando leituras diferentes das originais, o que seria prejudicial na posterior composição das métricas no TIMESAT. Assim, como observado por Becker et al.; (2020) e Eklundh e Jönsson (2015b) e corroborado com dados de campo, quanto menor as distorções do perfil original em relação ao suavizado, maior a acurácia dos resultados.

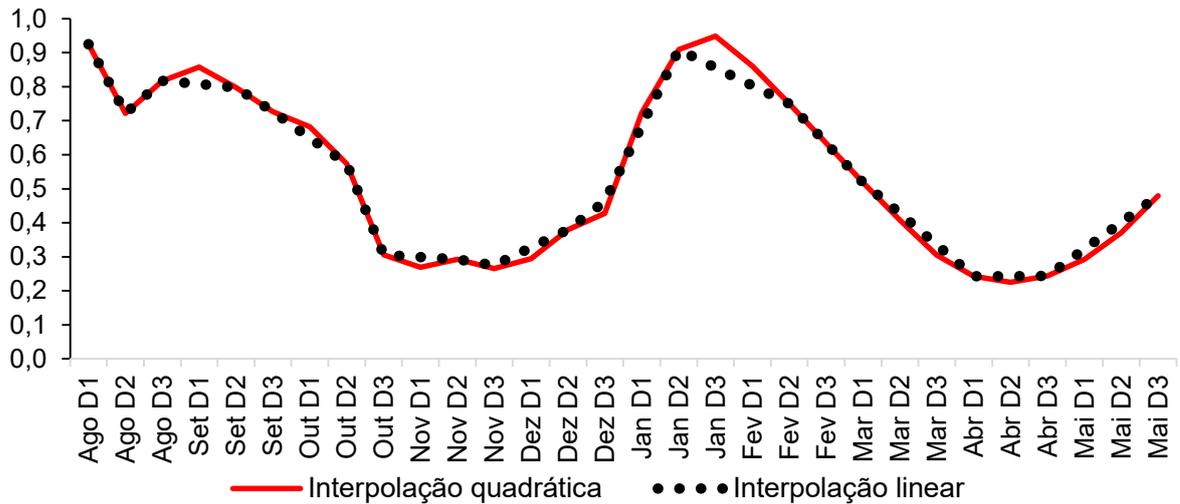


Figura 19 Comparação do perfil de EVI submetido à interpolação linear e à interpolação quadrática.

Fonte: Autor (2020).

A Figura 20 exemplifica esse processo em 30 decêndios selecionados, ilustrando os dados antes e depois da interpolação. Dados pré-existentes (14) não são alterados e apenas as lacunas vazias (16) são preenchidas com o valor médio entre os dados existentes próximos (anterior e posterior) aos faltantes para gerar a serie temporal de interesse completa.

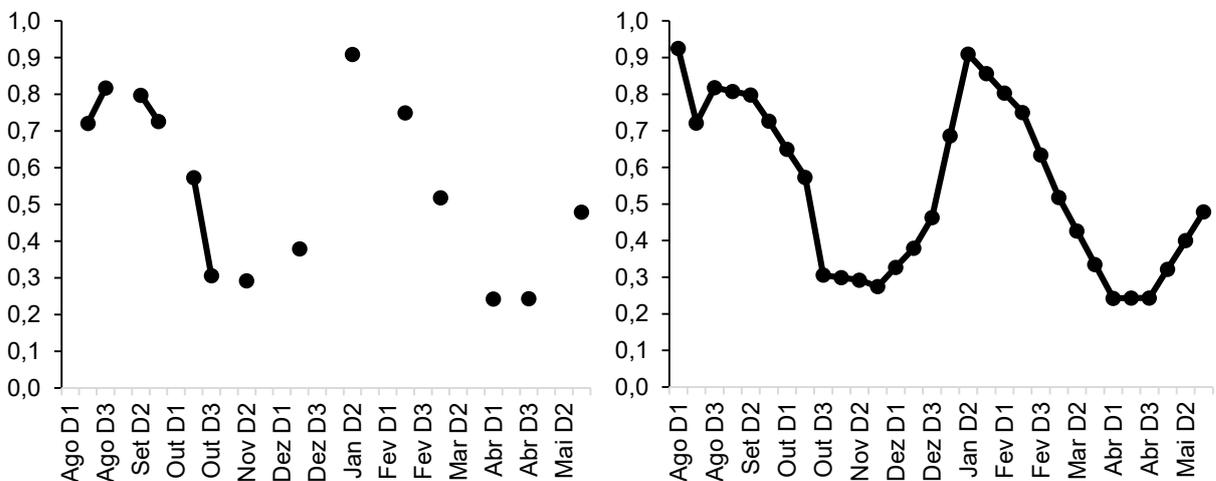


Figura 20 Comparação do perfil de EVI original (a esquerda) e depois da interpolação linear dos dados (direita).

Fonte: Autor (2020).

Os dados climáticos diários também foram convertidos em dados decendiais para igualar a granulometria das imagens de satélite, facilitando a legibilidade e processamento dos dados. Em seguida, para atestar a qualidade dos dados de precipitação, estes foram comparados com dados extraídos de estações meteorológicas em solo do Sistema de Tecnologia e Monitoramento Ambiental do Paraná (SIMEPAR, 2019).

O período analisado abrangeu todo o período do estudo: 2015 a 2018. Nessa comparação, utilizaram-se 20 estações meteorológicas localizadas em municípios próximos das lavouras de interesse (Figura 21). A análise não contemplou dados de temperatura, em função dessa variável ser mais estável espacialmente e somente apresentar variações em larga escala e não pontualmente como as chuvas.

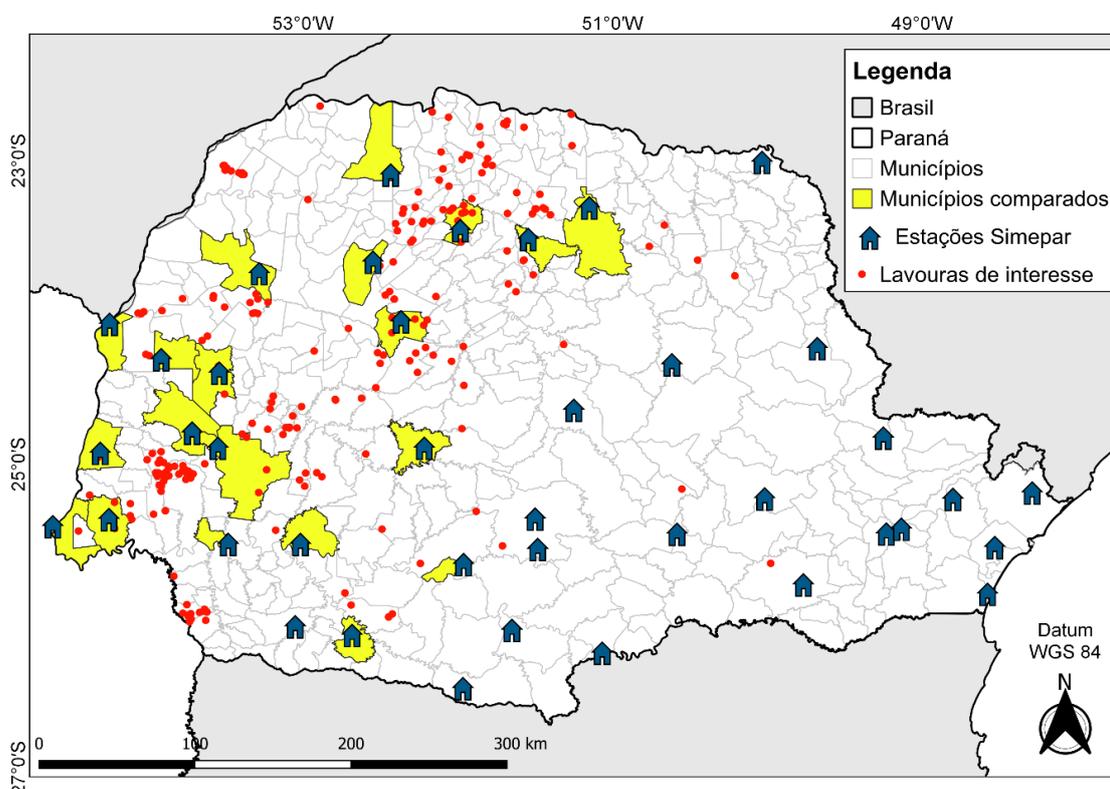


Figura 21 Comparativo entre a localização dos dados extraídos do NASA *Power* e estações meteorológicas do SIMEPAR.

Fonte: Autor (2020).

A comparação entre as bases de dados (NASA *Power* e SIMEPAR) foi executada pelo teste estatístico não-paramétrico de Mann-Whitney (HOLLANDER; WOLFE, 1973), para avaliar se as populações que originaram essas amostras podem ser consideradas semelhantes ou não. Esse procedimento foi executado no *software* Minitab, versão 16. O teste foi realizado com 5% de significância, considerando duas hipóteses: H_0 : os dados provêm de populações iguais; H_1 : as duas populações são diferentes.

Dessa forma, foram realizadas 80 comparações (20 municípios x 4 anos), onde inseriu-se as informações correspondentes para cada município e ano de ambas as bases de dados. Das 80 comparações, rejeitou-se H_1 em 78 casos (97,5% dos casos). Logo, conclui-se que os dados de ambas as bases de dados podem ser considerados iguais. O resumo do teste estatístico para os 20 municípios e quatro anos analisados está no Anexo B.

Após a confirmação da semelhança dos dados de precipitação da plataforma Nasa *Power* com dados reais, foram criadas diversas variáveis para proceder com a mineração de dados. Algumas dessas variáveis têm como referência o pico vegetativo apontado pelo perfil

temporal de EVI, bem como o ciclo total da cultura e outras variáveis que consideram o período de sinistro informado nos laudos. Assim, as lavouras consideradas normais não possuirão essas métricas relacionadas a sinistros. Sendo:

- Variáveis utilizando o pico vegetativo (EVI) como referência:

Da variável PRECIPITAÇÃO: soma durante o ciclo; valor no decêndio de pico; soma dos valores dos decêndios anterior até o posterior do pico (-1 a 1); soma dos valores dos dois decêndios anteriores até dois posteriores ao pico (-2 a 2);

Das variáveis TEMPERATURA MÉDIA, MÍNIMA E MÁXIMA: média durante o ciclo; média no decêndio do pico; média dos valores dos decêndios anterior até o posterior do pico (-1 a 1); média dos valores dos dois decêndios anteriores até dois posteriores ao pico (-2 a 2).

- Variáveis utilizando o período de sinistro como referência (quando ocorreu)

Da variável PRECIPITACAO: soma durante o periodo ou decêndio de ocorrência do sinistro; valor no decêndio central da ocorrência do sinistro (as duas variáveis anteriores podem apresentar o mesmo valor dependendo do tipo e duração do sinistro); soma dos valores dos decêndios anterior até o posterior à data central de ocorrência do sinistro (-1 a 1); soma dos valores dos dois decêndios anteriores até dois posteriores à data central de ocorrência do sinistro (-2 a 2);

Das variáveis TEMPERATURA MÉDIA, MÍNIMA E MÁXIMA: média durante o periodo ou decêndio de ocorrência do sinistro; média no decêndio central da ocorrência do sinistro; média dos valores dos decêndios anterior até o posterior à data central de ocorrência do sinistro (-1 a 1); média dos valores dos dois decêndios anteriores até dois posteriores à data central de ocorrência do sinistro (-2 a 2).

Na Figura 22 visualiza-se a posição das variáveis indicadas acima em um exemplo de lavoura de soja com sinistro.

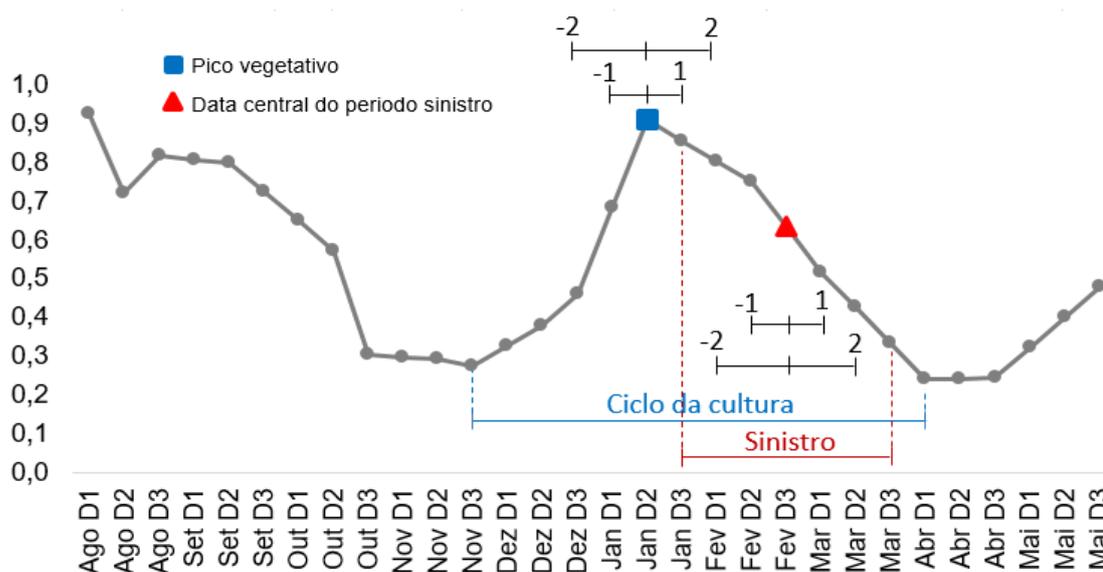


Figura 22 Posição das variáveis climáticas no perfil temporal de EVI definidas para o estudo em um exemplo de lavoura de soja com sinistro.

Fonte: Autor (2020).

4.3 Processamento

Após a geração das séries temporais de EVI completas, foi possível seguir com o processo de extração dos parâmetros de sazonalidade do perfil espectro-temporal de EVI das áreas agrícolas de interesse no *software* TIMESAT.

Inicialmente, os dados do perfil espectro-temporal de cada talhão foram salvos em formato “.txt”, conforme exemplo (Figura 23). Sendo que a primeira linha do arquivo contém a informação sobre o número de anos ou ciclos da série a analisar (*nyear*), o número de dados por ano ou ciclo (*nptperyear*) e o número total de series no arquivo (*nts*).

2	15	16																		
0.1833	0.2058	0.2283	0.3316	0.4349	0.2486	0.2487	0.2487	0.2488	0.2488	0.6441	0.7040	0.7639	0.8199	0.7388	0.6577	0.5695	0.4814	0.3932	0.3050	0.2168
0.1670	0.1889	0.2108	0.3598	0.4044	0.4490	0.5834	0.7177	0.8521												
0.3329	0.2908	0.2488	0.1671	0.1608	0.1545	0.1681	0.1817	0.1953	0.3519	0.5127	0.6734	0.6135	0.5536	0.4542	0.3385	0.3227	0.2908	0.2945	0.2739	0.2158
0.2099	0.2041	0.1907	0.1773	0.2841	0.3910	0.4979	0.6048	0.7117												
0.2982	0.2675	0.2368	0.1507	0.1520	0.1532	0.2293	0.3053	0.3814	0.4297	0.6488	0.8679	0.8515	0.8351	0.5463	0.4998	0.4533	0.3179	0.2913	0.2647	0.1947
0.1770	0.1594	0.1633	0.1671	0.3098	0.4524	0.5950	0.7376	0.8802												
0.3040	0.3142	0.3244	0.2230	0.2115	0.1999	0.2890	0.3780	0.4671	0.4999	0.5824	0.7449	0.7232	0.7015	0.4853	0.3748	0.3644	0.2342	0.2317	0.2291	0.1873
0.1779	0.1685	0.1686	0.1687	0.3098	0.4508	0.5919	0.7329	0.8740												
0.3493	0.2933	0.2373	0.2155	0.1936	0.1595	0.2043	0.2491	0.2939	0.3387	0.7457	0.7823	0.8189	0.8611	0.8682	0.8754	0.7541	0.6329	0.5117	0.3905	0.2693
0.1888	0.1850	0.1813	0.3793	0.3814	0.3835	0.5223	0.6611	0.7999												
0.5411	0.5158	0.4174	0.3933	0.4514	0.2278	0.2332	0.2418	0.3597	0.4510	0.5490	0.5412	0.5534	0.6568	0.6220	0.5450	0.4632	0.3282	0.2836	0.2605	0.3473
0.3637	0.3153	0.3285	0.3416	0.3665	0.3915	0.5332	0.6749	0.8166												
0.8361	0.6749	0.5137	0.2361	0.2426	0.2491	0.2107	0.1688	0.1855	0.4668	0.5730	0.6791	0.6728	0.6665	0.6370	0.5924	0.5479	0.4734	0.4153	0.3572	0.2211
0.1931	0.1651	0.1855	0.2058	0.3401	0.4744	0.6087	0.7430	0.8773												
0.1507	0.2602	0.3697	0.3008	0.1977	0.1991	0.1927	0.1863	0.1799	0.2418	0.3277	0.5055	0.6724	0.7290	0.7723	0.8544	0.8414	0.6435	0.4720	0.4088	0.2464
0.1767	0.2010	0.1515	0.3446	0.3389	0.2364	0.4208	0.7607	0.8260												
0.4633	0.4941	0.5249	0.5557	0.4559	0.3561	0.3451	0.3341	0.3231	0.4530	0.5525	0.6519	0.6745	0.6970	0.8217	0.7836	0.7455	0.5207	0.4059	0.3512	0.3206
0.3076	0.3447	0.4031	0.3615	0.3642	0.3470	0.3298	0.3127	0.2955												
0.5226	0.7474	0.3054	0.3017	0.3211	0.3291	0.3261	0.3230	0.2180	0.1595	0.1520	0.1601	0.2621	0.4507	0.6960	0.7170	0.7778	0.7740	0.7703	0.7665	0.5376
0.4643	0.3659	0.2912	0.3118	0.2936	0.3106	0.2470	0.1834	0.6398												
0.4507	0.5048	0.5589	0.6130	0.6055	0.5979	0.5838	0.5697	0.2556	0.2254	0.2677	0.2799	0.4051	0.4302	0.4212	0.4431	0.4413	0.4509	0.4553	0.4798	0.4081
0.4077	0.2572	0.3438	0.3804	0.4995	0.4367	0.3739	0.3112	0.2484												
0.3179	0.2990	0.2801	0.2612	0.1855	0.1498	0.1204	0.1311	0.1419	0.2024	0.3290	0.3557	0.4557	0.5557	0.5334	0.5285	0.5237	0.3467	0.2585	0.2704	0.2605
0.2627	0.2648	0.2194	0.1739	0.1931	0.2469	0.3006	0.3543	0.4080												
0.3548	0.3387	0.3226	0.3065	0.3884	0.4704	0.4720	0.4736	0.4751	0.2436	0.2158	0.2380	0.2499	0.3317	0.5469	0.6040	0.6610	0.6608	0.6606	0.6604	0.5165
0.4155	0.3144	0.2646	0.2148	0.2932	0.3124	0.3316	0.3508	0.3700												
0.5501	0.5002	0.4504	0.4005	0.4419	0.4832	0.5218	0.5605	0.5992	0.5308	0.4904	0.4501	0.2867	0.2232	0.3008	0.3379	0.3749	0.4492	0.5235	0.5978	0.5688
0.5320	0.5251	0.4717	0.3482	0.2859	0.3351	0.3844	0.4336	0.4828												
0.4683	0.4382	0.4081	0.3780	0.3608	0.3435	0.3914	0.4392	0.4870	0.4512	0.3677	0.2842	0.2933	0.3024	0.4181	0.4709	0.5238	0.5979	0.6720	0.7462	0.6952
0.6540	0.6127	0.4784	0.3441	0.2223	0.3070	0.3917	0.4763	0.5610												
0.5516	0.4396	0.3275	0.2155	0.2775	0.3396	0.3096	0.2796	0.2496	0.2726	0.1670	0.1815	0.2373	0.2930	0.4143	0.5698	0.6253	0.6781	0.7691	0.7800	0.6799
0.5525	0.4652	0.3819	0.2385	0.1869	0.2709	0.3548	0.4387	0.5227												

Figura 23 Exemplo da disposição dos dados de entrada em arquivo .txt.

Fonte: Autor (2020).

Para a análise e extração de dados de perfis espectro-temporais de índices de vegetação utiliza-se especificamente o módulo TIMESAT Graphical User Interface - TSM_GUI (Figura 24), entretanto, antes da extração dos dados, foi necessário o ajuste de alguns parâmetros, conforme detalhado abaixo.

a) Data Plotting:

Usuário escolhe o método de ajuste teórico (Gaussian, Logistic, Savitzky-Golay) que melhor representa o perfil espectro-temporal do EVI ao longo do ciclo da cultura.

b) Common setting:

- **Data range:** intervalo de dados, do mínimo (0) ao máximo (10.000);
- **Amplitude value:** amplitude variando de 0 a 1. Quanto maior o número estabelecido maior a remoção de *pixels* com fraca sazonalidade;

- *Spike method*: método de modificação de pico.
- c) *Class-Specific settings*:
 - *Seasonal par.*: definição do número de ciclos anuais: Se for 1 ciclo anual: 0 (ex.: cana-de-açúcar); se forem dois ciclos anuais: 1 (ex.: soja e milho safrinha);
 - *Adaptation strength*: intensidade de adaptação;
 - *Nº of envelope iterations*: interação de envelopes otimizando o pico vegetativo;
 - *(Fitting method) Window size*: quanto mais próximo de 1 mais ajustado o filtro selecionado fica a série original;
 - *Star/end of season*: ajuste da data de plantio e colheita para melhor ajuste do ciclo.

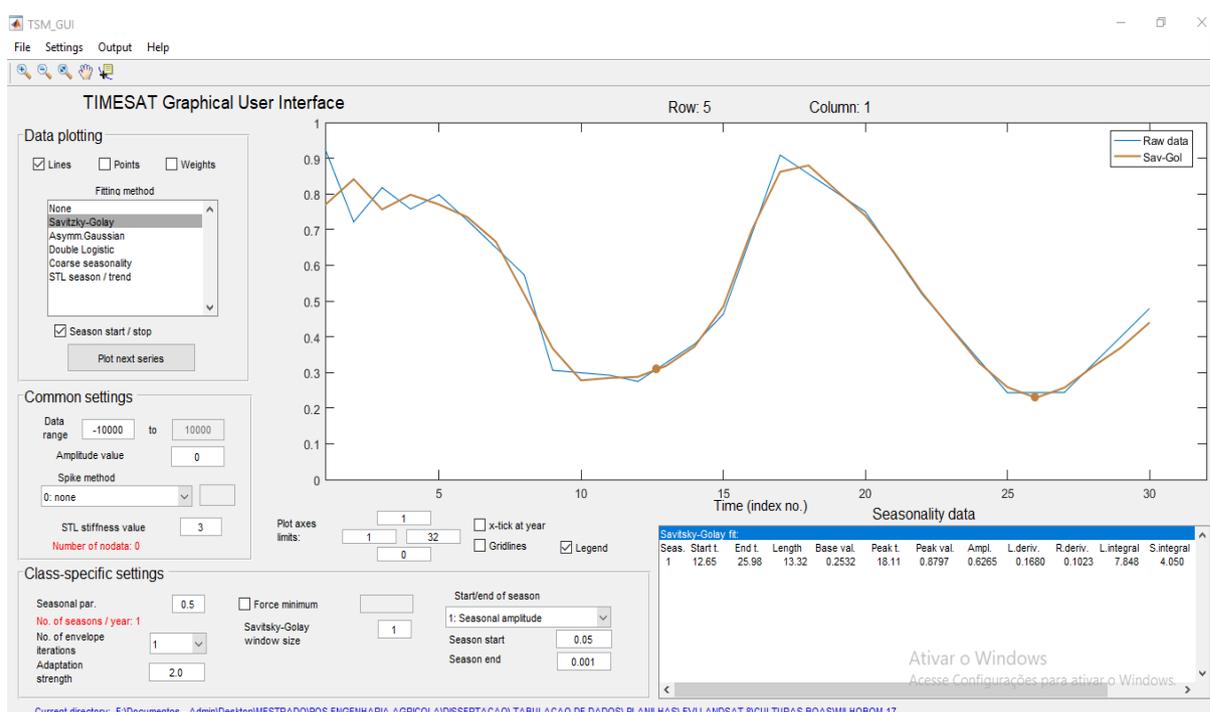


Figura 24 Tela do *software* TIMESAT onde podem ser vistas as ferramentas que o *software* disponibiliza para o ajuste do perfil temporal.

Fonte: Autor (2020).

Baseando-se nessas configurações iniciais, o *software* gera as variáveis abaixo:

Start t. (Start of season): Data de semeadura;

End t. (End of season): Data da colheita;

Length (Length of season): Ciclo da cultura;

Base Val (Base value): Valor do índice de vegetação médio;

Peak t. (Position of middle of season): Data do pico vegetativo da cultura;

Peak val. (Maximum of fettle data): Valor do índice de vegetação no pico da cultura;

Ampl. (Amplitude): Diferença entre o valor do índice de vegetação no pico vegetativo e o valor do índice de vegetação usado como base;

L. deriv. (Left derivative): Taxa de crescimento do início do ciclo da cultura. Calculada pela razão de diferença entre os níveis 20% e 80% pelo tempo decorrido entre estas medidas;

R. deriv. (Right derivative): Taxa de decréscimo no fim do ciclo da cultura. Calculada pela razão da diferença entre os níveis 80% e 20% pelo tempo decorrido entre estas medidas.

L. integral (Large Integral): Área sobre o perfil de índice de vegetação até o eixo X do gráfico, desde a data da semeadura até a data da colheita;

S. integral (Small Integral): Área sobre o perfil de índice de vegetação até o valor do índice de vegetação usado como base.

Os dados gerados pelo *software* são dispostos pelo programa na tabela *Seasonality.txt* (Figura 25), em forma numérica. Os valores relacionados às datas estão dispostos em números inteiros com base nas cenas do perfil temporal definido no arquivo de entrada.

Seas.	Start t.	End t.	Length	Base val.	Peak t.	Peak val.	Ampl.	Lderiv.	Rderiv.	Lintegral	Sintegral	Start val.	End val.
1	12.65	25.98	13.32	0.2532	18.11	0.8797	0.6265	0.1680	0.1023	7.848	4.050	0.3078	0.2294

Figura 25 Exemplo de Tabela *Seasonality.txt*.

Fonte: Autor (2020).

Os dados acima foram posteriormente convertidos em valores que contivessem unidades, para o processo de análise dos dados (Tabela 6).

Tabela 6 Exemplo dos dados *seasonality* convertidos para números, datas e dias viáveis de análise

Nomenclatura de saída	Dados de saída	Nomenclatura convertida	Dados convertidos
Start t.	12.65	Data plantio	29/11/2015
End t.	25.98	Data colheita	10/04/2016
Length	13.32	Ciclo em dias	133
Base val.	0.2532	I.V Base	0,2532
Peak t.	18.11	Data pico vegetativo	23/01/2016
Peak val.	0.8797	I.V Pico	0,8797
Ampl.	0.6265	Amplitude	0,6265
L. deriv.	0.1680	T. Cresc. 20 a 80%	0,168
R. deriv.	0.1023	T. Decresc. 80 a 20%	0,1023
L. integral	7.848	Área maior	7,848
S. integral	4.050	Área menor	4,05
Start val.	0.3078	I.V plantio	0,3078
End val.	0.2294	I.V colheita	0,2294

Fonte: Autor (2020).

O procedimento descrito no TIMESAT foi realizado para todas as lavouras sinistradas e, também, as consideradas normais deste banco de dados. Dessa forma, com os parâmetros de sazonalidade das áreas de interesse, foi possível filtrar as métricas climáticas coincidentes definidas anteriormente, que necessitavam informações do perfil temporal.

Todo o processo de agrupamento dos arquivos foi realizado em etapa única em código Python na interface *Google Colaboratory*. Compilando em tabelas separadas por cultura e ano, todas as informações contidas nos laudos, mais os parâmetros de sazonalidade do perfil

temporal, mais as métricas climáticas escolhidas, gerando o banco de dados para realização das análises propostas.

4.3.1 Comparações entre dados reais e estimações do TIMESAT

Após o processo de extração dos parâmetros de sazonalidade do perfil espectro-temporal de EVI das áreas agrícolas de interesse no *software* TIMESAT, foi realizada uma avaliação da variação entre as datas reais (da semeadura e colheita) contidas nos laudos, comparando-se com as datas estimadas pelo TIMESAT, das mesmas etapas do ciclo da culturas analisadas, com o objetivo de observar a o grau de acerto das estimações do *software*, em relação aos dados reais de campo deste estudo.

Esse tipo de comparação é valiosa, haja vista que, em todos os laudos de sinistro desse banco de dados, observam-se datas exatas de semeadura e colheita, o que, por muitas vezes, pode ser de difícil acesso para calibrações de informações de satélites.

Para tal, foram utilizadas algumas comparações estatísticas como:

- O erro médio ou *mean error* (ME);
- O erro médio absoluto ou *mean absolute error* (MAE);
- A raiz quadrada erro médio ou *root mean square error* (RMSE)
- O erro percentual absoluto médio ou *mean absolute percentage error* (MAPE).

Dada pelas seguintes equações.

$$ME = \frac{1}{N} \sum_{n=1}^N (obs^n - real^n) \quad (2)$$

$$MAE = \frac{1}{N} \sum_{n=1}^N |obs^n - real^n| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N (obs^n - real^n)^2} \quad (4)$$

$$MAPE = \frac{1}{N} \sum_{n=1}^N \left| \frac{obs - real}{real} \right| \times 100 \quad (5)$$

4.4 Validação

A partir das séries temporais de EVI juntamente dos dados agrometeorológicos de lavouras sinistradas e normais, os dados foram submetidos a algoritmos de mineração de dados, a fim de verificar a existência de padrões ocultos nas bases de dados. Para analisar as informações, nessa etapa, foram elaborados dois questionamentos:

- 1) Se é possível confirmar se houve sinistro ou não;
- 2) Se é possível confirmar qual tipo de sinistro ocorrido.

E para tentar respondê-los, foi necessário decidir quais atributos do banco de dados seriam utilizados nas análises, sendo elencados os seguintes parâmetros (Tabelas 7 e 8).

Tabela 7 Atributos selecionados no banco de dados para composição das análises

Atributo	Descrição
CAMPOSIN	(Informações que apenas as áreas sinistradas possuem) = Altitude, Densidade de plantas, Tipo solo, Produtividade (kg ha ⁻¹).
FENTRIGO	(Fenologia do trigo) = germinação, perfilhamento, alongamento, emborachamento, florescimento, antese, grão leitoso e grão em massa, maturação.
FEN	(Fenologia soja e milho) = Ve, V1, V2, V3, V4, V5, V6, V7, V8, V9, V10, V11, V12, R1, R2, R3, R4, R5, R6, R7, R8, R9.
EVI	(Parâmetros de sazonalidade gerados pelo TIMESAT referentes ao EVI) = I.V Base, I.V Pico, Amplitude, T. Cresc. 20 a 80%, T. Decresc. 80 a 20%, Área maior, Área menor, I.V plantio, I.V colheita.
CICLOSIN	(Duração do ciclo e duração do sinistro informados nos laudos) = Ciclo em dias laudo, Dias sinistro.
CICLOEVI	(Duração do ciclo em dias baseado nos parâmetros de EVI) = Ciclo em dias.

Fonte: Autor (2020).

Além disso, também foram usadas as métricas de precipitação e temperatura, mencionadas anteriormente, para o ciclo total da cultura e, também, apenas no período do sinistro ocorrido.

Tabela 8 Atributos climáticos selecionados no banco de dados para composição das análises

Atributo	Descrição
PREC	PREC Soma ciclo, PREC pico, Soma PREC pico (-1 a 1), Soma PREC pico (-2 a 2).
TMED	Média T_MED ciclo, T_MED pico, Média T_MED pico (-1 a 1), Média T_MED pico (-2 a 2).
TMIN	Média T_MIN ciclo, T_MIN pico, Média T_MIN pico (-1 a 1), Média T_MIN pico (-2 a 2).
TMAX	Média T_MAX ciclo, T_MAX pico, Média T_MAX pico (-1 a 1), Média T_MAX pico (-2 a 2).
PRECSIN	PREC soma sinistro, PREC sinistro, Soma PREC sinistro (-1 a 1), Soma PREC sinistro (-2 a 2);
TMEDSIN	Média T_MED sinistro, T_MED sinistro, Média T_MED sinistro (-1 a 1), Média T_MED sinistro (-2 a 2);
TMINSIN	Média T_MIN sinistro, T_MIN sinistro, Média T_MIN sinistro (-1 a 1), Média T_MIN sinistro (-2 a 2);
TMAXSIN	Média T_MAX sinistro, T_MAX sinistro, Média T_MAX sinistro (-1 a 1), Média T_MAX sinistro (-2 a 2).

Fonte: Autor (2020).

Todos procedimentos a seguir foram desenvolvidos em ambiente Python na plataforma *Google Colaboratory*.

Para o procedimento de confirmação de ocorrências de sinistros (questionamento 1), foi necessário unificar as informações comuns de ambas as bases de dados (lavouras normais e lavouras sinistros) de cada cultura e definir qual seria a coluna alvo em questão, sendo definido os *features* em comum: EVI; PREC; TMED; TMIN; TMAX e CICLOEVI. A coluna “evento causador” foi definida como alvo.

Com a base de dados construída e retificada, executou-se a classificação das áreas sinistradas e não-sinistradas utilizando três algoritmos: *Support vector machine* (SVM), Floresta aleatória (*Random forest* - RF) e Árvore de decisão (*Decision tree* - DT).

A partir dos parâmetros da matriz de confusão é possível derivar medidas de desempenho de um classificador nas classes negativa e positiva, independentemente. Entre essas medidas, destacam-se as taxas de verdadeiro positivo (VP), falso positivo (FP), verdadeiro negativo (VN) e falso negativo (FN), como exemplificadas na matriz de confusão genérica apresentada a seguir:

		PREVISTO	
		SIM	NÃO
REAL	SIM	Verdadeiro positivo (VP)	Falso negativo (FN)
	NÃO	Falso positivo (FP)	Verdadeiro negativo (VN)

Em que:

- Verdadeiros positivos (VP): são os valores positivos previstos corretamente, o que significa que o valor da classe real é sim e o valor da classe prevista também é sim.

- Verdadeiros negativos (VN): são os valores negativos previstos corretamente, o que significa que o valor da classe real é não e o valor da classe prevista também é não.

Nesses dois casos, se o valor da classe indicar alguma informação a classe prevista indicará a mesma coisa.

- Falsos positivos (FP): quando a classe real é não e a classe prevista é sim.

- Falsos negativos (FN) - quando a classe real é sim, mas a classe prevista em não.

Os falsos positivos e falsos negativos, ocorrem quando sua classe real contradiz a classe prevista.

Dessa forma, as métricas desenvolvidas para analisar o desempenho dos algoritmos e os resultados das predições foram: acurácia, precisão, *recall*, F1 score, em que:

Acurácia: corresponde à proporção de previsões corretas do classificador. Quanto mais próximo de 1, mais próximo VN e VP estará representando 100% de acerto. Porém, é necessário observar outros parâmetros para avaliar o desempenho individual da categoria.

$$Acurácia = \frac{VP+VN}{VN+VP+FP+FN} \quad (6)$$

Precisão: É a razão entre as observações positivas previstas corretamente e o total de observações positivas previstas. A alta precisão está relacionada à baixa taxa de falsos positivos. Neste caso, ter uma elevada precisão na categoria positiva significa acertar a ocorrência do sinistro.

$$Precisão = \frac{VP}{VP+FP} \quad (7)$$

Recall: “Rechamada”, “sensibilidade” ou taxa de verdadeiro positivo, é a razão entre as observações positivas previstas corretamente e todas as observações da classe real sim.

$$Recall = \frac{VP}{VP+FN} \quad (8)$$

F1 Score: É a média ponderada de precisão e *recall*. Portanto, essa pontuação considera tanto os falsos positivos quanto os falsos negativos. Geralmente é mais útil que precisão, especialmente se for de uma distribuição de classe desigual. A precisão funciona melhor se os falsos positivos e falsos negativos tiverem um custo semelhante.

$$F1\ Score = 2 \times \frac{(Recall \times Precisão)}{(Recall + Precisão)} \quad (9)$$

Todos os algoritmos precisam de parametrizações, visando à obtenção de maior acurácia. Assim, as parametrizações para o algoritmo SVM foram feitas de forma automática entre os valores de 'C' entre: 0.5, 1, 10, 100; 'gamma' entre: 0.1, 0.01, 0.001, 0.005 e 'kernel'='rbf'.

Para o algoritmo RF, a parametrização foi definida a partir de 50 árvores, de profundidade máxima de 10 nós cada. Para o algoritmo DT foi imposta a restrição: possuir profundidade máxima de 5 nós. Essas parametrizações foram impostas com intenção de que as árvores geradas por esses algoritmos (RF e DT) fossem pequenas e de simples interpretação.

O banco de dados originais possuía: para milho 129 sinistros e 43 lavouras normais; soja 26 sinistros e 59 normais e trigo 90 sinistros e 36 normais. O conjunto de treinamento e teste foi dividido entre 75 e 25%, respectivamente, do total do conjunto de dados para cada cultura. Cada algoritmo foi executado dez vezes para cada cultura e os resultados apresentados são a média dessas dez repetições.

Para compreender melhor o resultado do processo de classificação realizado, são dispostas as representações gráficas das DT (para o conjunto de treinamento), apontando os fatores mais influentes na ocorrência ou não do sinistro. Cada ramo da árvore de decisão termina no que é chamado de folha (que pode ser normal ou sinistro), que é entendida como uma regra de classificação (R1, R2, R3, e assim por diante). Assim, para cada folha ou regra, o número de acertos/ número de erros é apresentado entre colchetes.

Para verificar qual foi o sinistro nas culturas analisadas (questionamento 2), utilizou-se apenas o banco de dados que continham lavouras com sinistros. A variável alvo continuou sendo o evento causador, sendo possível utilizar todos os atributos (*features*) disponíveis. Porém, optou-se por utilizar informações obtidas exclusivamente via satélite, deixando de considerar *features* como CAMPOSIN, FENTRIGO, FEN, CICLOSIN, CICLOEVI que foram obtidos com o produtor rural ou apontadas pelo perito no ato da vistoria. Essa opção se justifica, pois minimiza a necessidade de informações prévias ou constatadas a campo para realizações de análises dessa natureza, ou seja, inferir sobre a ocorrência de sinistro apenas com dados de satélite.

Os atributos utilizados nesta etapa para compor as análises foram: EVI; PREC; TMED; TMIN; TMAX; PRECSIN; TMEDSIN; TMINSIN; TMAXSIN. Essa base de dados foi submetida aos mesmos algoritmos de classificação: SVM, RF e DT e avaliadas pelas mesmas métricas de acurácia, precisão, *recall* e *F1 score*.

Na definição da quantidade de amostras foi necessário escolher quais sinistros estariam em análise, em virtude do número de sinistros por cultura ser reduzido, como já demonstrado na Tabela 3. Para evitar problemas de classificação, os sinistros com poucas amostras em relação aos com mais representantes dentro de cada cultura foram descartados, restando apenas os seguintes dados: milho (chuva excessiva 37 e seca 86), soja (chuva excessiva 15, seca 5 e granizo 4) e trigo (seca 42 e seca + geada 38).

A definição de parâmetros, as restrições dos algoritmos e a visualização das árvores de decisão são as mesmas da análise anterior. Cada algoritmo foi executado dez vezes para cada cultura e os resultados apresentados são a média dessas dez repetições.

5 RESULTADOS E DISCUSSÃO

As Figuras 26 e 27 ilustram as composições dos valores máximos, mínimos e a média de EVI para as áreas de lavouras sinistradas e normais, após o processo de interpolação. De forma geral, o acréscimo de feições por cultura acarretou maior amplitude dos valores máximos e mínimos de EVI nos decêndios, em virtude da diferença das datas de plantio e ciclos da cultura.

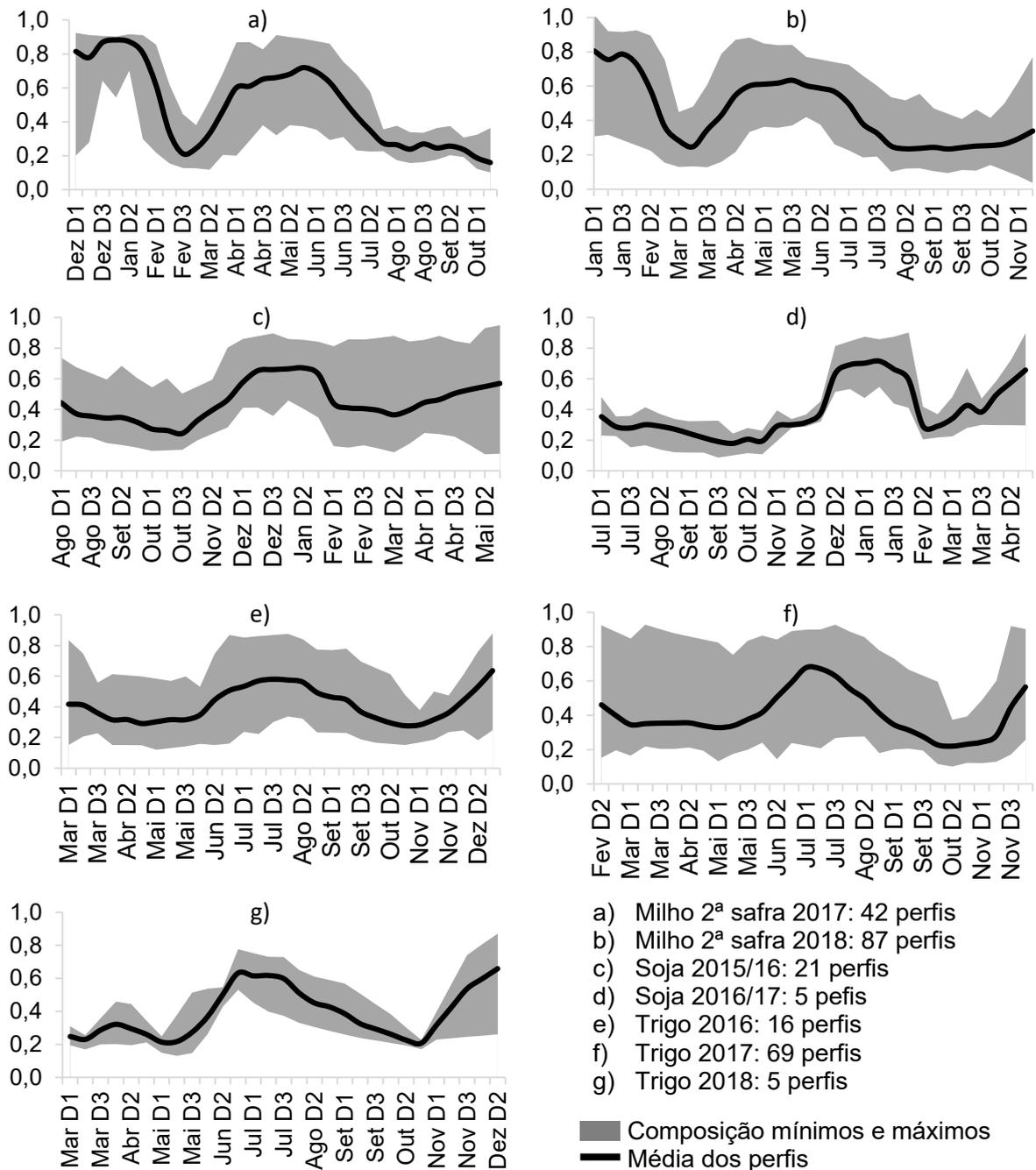


Figura 26 Valores máximos, mínimos e perfis médios de EVI em *pixels* puros (30x30m) para lavouras sinistradas.

Fonte: Autor (2020).

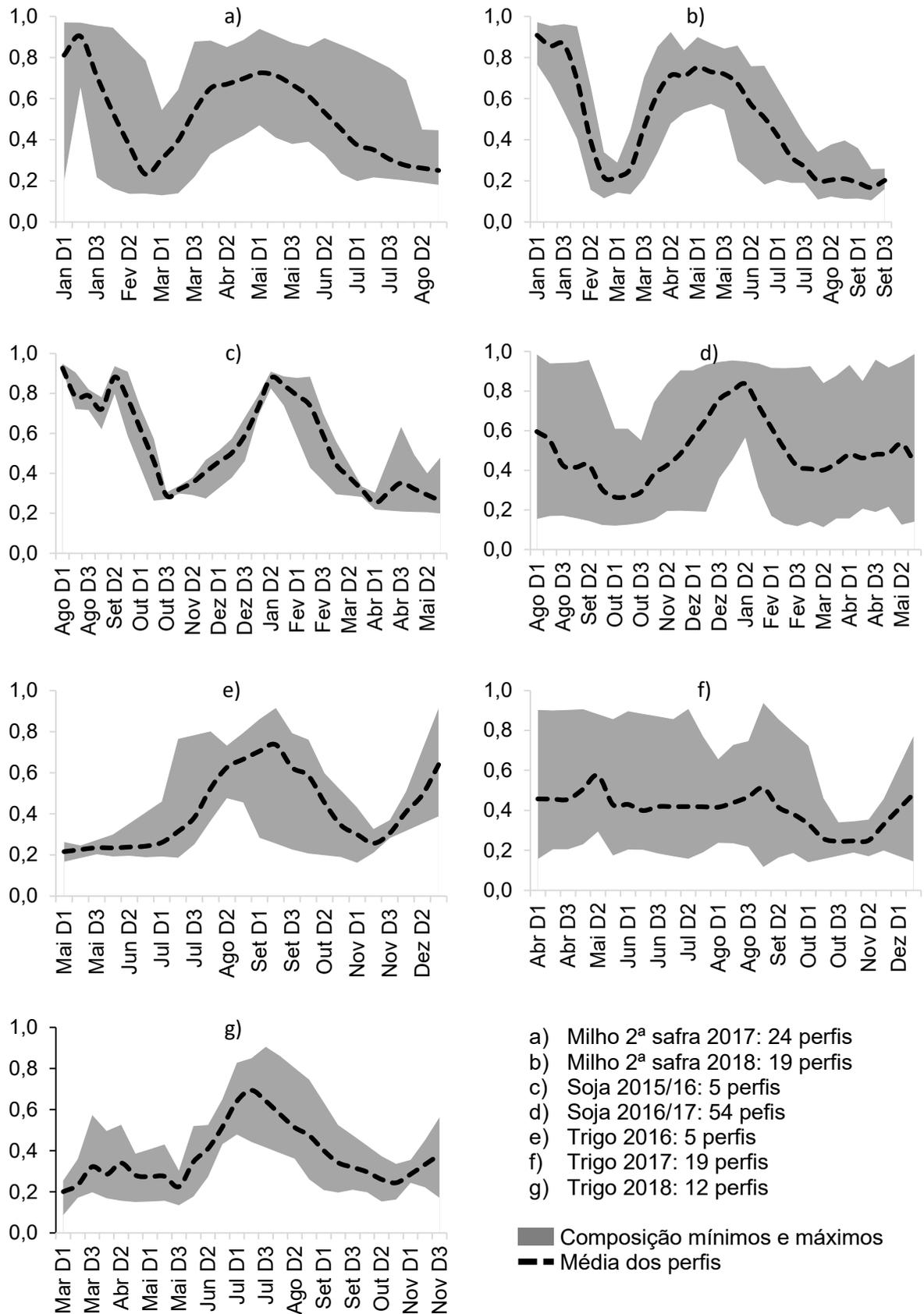


Figura 27 Valores máximos, mínimos e perfis médios de EVI em *pixels* puros (30x30m) para lavouras normais.

Fonte: Autor (2020).

5.1 Resultado das comparações entre dados reais e estimações do TIMESAT

Os resultados da comparação das datas reais de semeadura e colheita (contidas nos laudos) sugerem baixa variação entre os decêndios das datas reais e estimadas do perfil temporal EVI, conforme se constata na Tabela 9.

Tabela 9 Estatísticas de comparação em dias das datas de semeadura e colheita, entre os dados reais (laudos) com os estimados pela análise de padrões sazonais do TIMESAT

Estatística / Cultura		Milho 2017	Milho 2018	Soja 2015/16	Soja 2016/17	Trigo 2016	Trigo 2017	Trigo 2018
Semeadura	n	42	87	21	5	16	69	5
	ME	14	12	1	5	5	3	5
	MAE	14	13	4	5	8	7	7
	RMSE	17,00	15,70	5,94	5,84	10,02	10,23	8,71
	MAPE	0,032	0,030	0,008	0,011	0,019	0,017	0,017
Colheita	ME	-3	-2	-4	3	2	-1	7
	MAE	7	8	5	9	5	4	8
	RMSE	9,61	10,78	11,89	11,42	5,85	6,28	11,10
	MAPE	0,017	0,018	0,012	0,020	0,011	0,010	0,017

Notas: n: Número de amostras; **ME:** Erro médio; **MAE:** Erro médio absoluto; **RMSE:** Raiz quadrada erro médio; **MAPE:** Erro percentual absoluto médio.

Fonte: Autor (2020).

Ao se observar o erro médio (ME) e o erro médio absoluto (MAE) para a semeadura em todas as áreas estudadas, observamos que o maior valor, ou seja, a maior diferença entre a data real e estimada é de 14 dias (Milho, 2017). Considerando-se que está sendo utilizado um satélite de resolução temporal de 16 dias convertida para dados decendiais, pode-se concluir que o TIMESAT inferiu uma data com erro inferior a dois decêndios, o que pode ser considerado um bom resultado, pois, após a semeadura das culturas, a semente leva alguns dias para completar o processo de germinação e emergência (ROSSETTO et al., 1997; MAGALHAES; DURÃES, 2006) para, só então, iniciar atividade fotossintética e ser passível de ser captada via satélite (ALMEIDA et al., 2005).

Observando-se o ME e o MAE, para os dados de colheita, constataram menores erros, em relação aos dados de semeadura em todas as áreas estudadas. O que pode ser explicado pelo fato de que o ato da colheita provoca uma maior variação no perfil de EVI em menor espaço de tempo, pois, numa leitura do satélite a cultura ainda pode estar a campo, gerando leituras (mesmo que com EVI baixos) e, ao ser colhida, abruptamente não há mais cultura a gerar refletância (MARCHESAN et al., 2019; COUTO JUNIOR et al., 2011).

Observam-se alguns números negativos do ME na colheita (Milho 2017; Milho 2018; Soja 2015/16; Trigo 2017), devido à maioria das estimações das datas para essas culturas terem sido consideradas em datas anteriores às reais.

Na semeadura não se observa nenhum número negativo para ME, pois a maioria das estimações dessas datas ocorreram depois das datas reais.

Notou-se também, menores erros no ME e demais índices quanto menor é a quantidade de representantes (N) por grupo (cultura/ano) devido à menor variação de datas observadas, o que também se observa nas Figuras 26 e 27.

Em relação aos índices RMSE para os dados de semeadura, observaram-se maiores valores para a cultura de milho no ano de 2017 (17,00), corroborando o maior valor de ME e MAE observado nas semeaduras em todas as áreas.

Para os dados de colheita, observou-se maior valor de RMSE para a soja 2015/2016 (11,89). Embora esse grupo não possua os maiores valores de ME e MAE, o fato de um alto índice de RMSE se deve (não exclusivamente) ao fato de que no banco de dados uma data foi estimada com uma variação de 49 dias em relação à data real, o que faz com que este valor da diferença entre a real e observada ao quadrado (Equação 4) seja alto, conseqüentemente, elevando a média desse índice para a respectiva cultura.

Para o índice MAPE de semeadura, observou-se novamente maior valor para a cultura de milho em 2017 (0,032). Para a colheita, observou-se maior MAPE para soja 2016/17 (0,020).

Resultados semelhantes foram relatados por Becker et al. (2020), que observaram valores promissores, quando comparados com áreas em escala das propriedades: datas de plantio, de máximo desenvolvimento vegetativo e de colheita de fazendas monitoradas de soja e milho, em relação às inferidas pelo TIMESAT das mesmas áreas.

Metodologias dessa natureza demonstram a precisão que pode ser obtida e, principalmente, a escalabilidade e replicabilidade com o uso de imagens de satélite para obter dados do ciclo fenológico das culturas agrícolas.

5.2 Previsão da ocorrência de sinistro

As matrizes de confusão foram geradas a partir das 10 repetições de cada algoritmo. Na Tabela 10, apresenta-se o compilado com as médias das matrizes de confusão (em porcentagem) dos valores de cada posição para cada cultura e seus respectivos algoritmos no conjunto teste. As matrizes de confusão servem de base para comparação dos acertos e erros dos verdadeiros e falsos das classificações.

Tabela 10 Compilação das matrizes de confusão do conjunto de teste em porcentagem, para os algoritmos SVM, RF e DT aplicados aos dados das culturas de milho 2ª safra, soja e trigo

		SVM				RF				DT			
		PREVISTO				PREVISTO				PREVISTO			
		Sim		Não		Sim		Não		Sim		Não	
MILHO	REAL Sim	VP 11,40	FN 14,19	VP 12,09	FN 13,49	VP 16,28	FN 9,30						
	REAL Não	FP 1,86	VN 72,56	FP 3,26	VN 71,16	FP 11,40	VN 63,02						
SOJA	REAL Sim	VP 64,09	FN 4,09	VP 68,18	FN 0,00	VP 61,36	FN 6,82						
	REAL Não	FP 6,36	VN 25,45	FP 3,64	VN 28,18	FP 3,18	VN 28,64						
TRIGO	REAL Sim	VP 7,81	FN 20,31	VP 15,00	FN 13,13	VP 16,25	FN 11,88						
	REAL Não	FP 0,94	VN 70,94	FP 5,63	VN 66,25	FP 12,50	VN 59,38						

Notas: VP: Verdadeiros positivos; VN: Verdadeiros negativos; FP: Falsos positivos; FN: Falsos negativos.

Fonte: Autor (2020).

De forma geral, o algoritmo RF dá sinais de uma melhor capacidade de classificação, em comparação com o SVM e DT, principalmente, por apresentar menores valores em suas diagonais secundárias, tendo apresentado melhor desempenho em prever a ocorrência de sinistro (exceto para a soja que teve melhor precisão na situação inversa). A importância desse resultado consiste em errar menos em situações em que de fato o sinistro aconteceu, confirmando o seu evento com a maior grau de confiança possível.

Na Tabela 11, apresentam-se as métricas de desempenho: precisão, *recall*, F1 score e acurácia (\pm Desvio padrão), considerando a ocorrência de sinistro, para os três algoritmos escolhidos para a cultura do milho. Os resultados dessa e das demais tabelas apresentadas correspondem à média de 10 repetições.

Tabela 11 Medidas de desempenho dos classificadores SVM, RF e DT, considerando a ocorrência de sinistro na cultura do milho

	Algoritmo	Condição	Precisão	Recall	F1-score	Acurácia	Amostras
SVM	Treinamento	Normal	0,957	0,778	0,858		32
		Sinistrado	0,930	0,989	0,958		97
		Média	0,944	0,884	0,908	0,936 (\pm 0,018)	129
	Teste	Normal	0,883	0,443	0,583		11
		Sinistrado	0,838	0,976	0,900		32
		Média	0,861	0,710	0,742	0,839 (\pm 0,032)	43
RF	Treinamento	Normal	1,000	1,000	1,000		32
		Sinistrado	1,000	1,000	1,000		97
		Média	1,000	1,000	1,000	1,000 (\pm 0)	129
	Teste	Normal	0,793	0,472	0,573		11
		Sinistrado	0,843	0,958	0,896		32
		Média	0,818	0,715	0,735	0,832 (\pm 0,060)	43
DT	Treinamento	Normal	0,844	0,850	0,844		32
		Sinistrado	0,952	0,947	0,947		97
		Média	0,898	0,899	0,896	0,922 (\pm 0,012)	129
	Teste	Normal	0,613	0,639	0,614		11
		Sinistrado	0,873	0,846	0,858		32
		Média	0,743	0,743	0,736	0,793 (\pm 0,057)	43

Fonte: Autor (2020).

O algoritmo SVM teve uma acurácia global de 0,936 (\pm 0,018) para o treinamento e 0,839 \pm (0,032) para o conjunto teste. A classe positiva teve um alto grau de acerto, porém, ao se observar isoladamente o número de amostras classificadas como falsos negativos (14,19% - Tabela 10), percebe-se uma redução de predição do algoritmo, indicando que ele teve uma capacidade maior de identificar as lavouras sinistradas do que normais.

O algoritmo RF obteve para o conjunto de treinamento acurácia de máxima, porém, quando aplicado ao conjunto de teste, obteve uma acurácia global de 0,832 (\pm 0,060). Para este banco de dados, apresentou comportamento similar ao SVM, em relação aos valores falsos negativos (13,49% - Tabela 10) evidenciados pelos índices de *recall* e *F1 score*.

A acurácia da DT do conjunto teste para a cultura de milho foi de 0,793 (\pm 0,057), porém, diferentemente dos classificadores SVM e RF, acabou apresentando menores erros na classe de falsos positivos (11,40% - Tabela 10) e falsos negativos (9,30% - Tabela 10), por isso, têm-se valores de *recall* e *F1 score* para a classe normal maiores que os algoritmos anteriores. Entretanto, os valores de *recall* e *F1 score* para a classe sinistrada foi ligeiramente menor que os demais.

Embora esta etapa de análises trate de diversos fatores e sinistros envolvidos sobre cada cultura, a representação gráfica das DT apresenta quais foram os fatores relevantes para a ocorrência ou não de sinistro (Figura 28 a Figura 30).

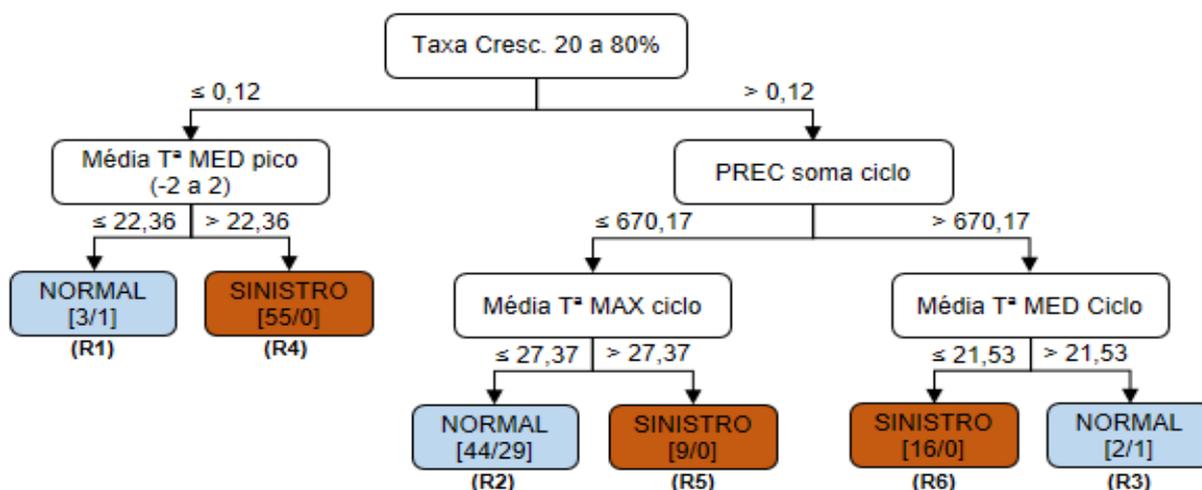


Figura 28 DT resumida para os fatores relevantes na ocorrência de sinistro na cultura do milho.

Fonte: Autor (2020).

Para a cultura do milho, a DT identificou com acurácia de 0,922 ($\pm 0,012$) cinco atributos como preditores necessários para a classificação de áreas normais ou com sinistro (Figura 28). Assim, foram geradas 6 regras para classificação, sendo três para as áreas normais e três para as sinistradas. Observa-se uma alta taxa de acerto na maioria das regras exceto para a R2 que apontou 29 erros de classificação das amostras. Basicamente, tem-se a incidência de dois tipos de sinistro mais representativos (chuva excessiva e seca). Assim, o fator mais relevante na decisão da árvore tem relação com a taxa de crescimento de 20 a 80%, que é um reflexo da sanidade da planta, pois se não falta água (PREC soma ciclo) e as temperaturas são ideais (Média Tª média pico -2 a 2), o crescimento da planta será normal.

Na Tabela 12, são apresentadas as métricas de desempenho para os três algoritmos escolhidos, considerando-se a ocorrência de sinistro na cultura da soja.

Tabela 12 Medidas de desempenho dos classificadores SVM, RF e DT, considerando a ocorrência de sinistro na cultura da soja

	Algoritmo	Condição	Precisão	Recall	F1-score	Acurácia	Amostras
SVM	Treinamento	Normal	0,970	0,994	0,982	0,971 (\pm 0,017)	44
		Sinistrado	0,984	0,920	0,949		19
		Média	0,977	0,957	0,966		63
	Teste	Normal	0,913	0,940	0,925	0,895 (\pm 0,029)	15
		Sinistrado	0,883	0,800	0,828		7
		Média	0,898	0,870	0,877		22
RF	Treinamento	Normal	1,000	1,000	1,000	1,000 (\pm 0)	44
		Sinistrado	1,000	1,000	1,000		19
		Média	1,000	1,000	1,000		63
	Teste	Normal	0,952	1,000	0,976	0,963 (\pm 0,034)	15
		Sinistrado	1,000	0,886	0,934		7
		Média	0,976	0,943	0,955		22
DT	Treinamento	Normal	0,964	0,957	0,961	0,946 (\pm 0,011)	44
		Sinistrado	0,905	0,914	0,909		19
		Média	0,935	0,936	0,935		63
	Teste	Normal	0,954	0,901	0,925	0,900 (\pm 0,039)	15
		Sinistrado	0,832	0,901	0,854		7
		Média	0,893	0,901	0,890		22

Fonte: Autor (2020).

O conjunto e dados de treinamento para o SVM teve acurácia de 0,971 (\pm 0,017), porém, para o conjunto de teste a acurácia foi de 0,895 (\pm 0,029). Os valores de Precisão, *recall* e F1 score foram maiores para os casos considerados normais do que sinistrados com 6,36% de falsos positivos e 4,09% de falsos negativos (Tabela 10), indicando com mais facilidade a classificação de uma lavoura normal do que uma sinistrada.

De uma acurácia máxima para os dados de treinamento para o RF, para o conjunto teste a acurácia teve uma redução para 0,963 (\pm 0,034). Nesse caso, o preditor teve uma alta taxa de acerto aos positivos verdadeiros (68,18% - Tabela 10), pois acertou todos os casos em que ocorreram sinistro e não foi capaz de prever corretamente apenas um caso, ou seja, 3,64% (Tabela 10) dos casos pois os classificaram como falsos positivos.

A DT aplicada à soja teve uma acurácia de 0,946 (\pm 0,011) para os dados de treinamento e quando aplicada ao conjunto de teste 0,900 (\pm 0,039). Obteve uma acurácia menor que o SVM, para os dados de treinamento, porém, quando aplicada ao conjunto teste, teve um melhor desempenho, com maior capacidade de classificação de casos normais. O que também pode ser observado nas matrizes de confusão em que a diagonal secundária (Tabela 10) da DT (FP 3,18% e FN 6,82%) é menor do que a do SVM (FP 6,36% e FN 4,09%), para a soja.

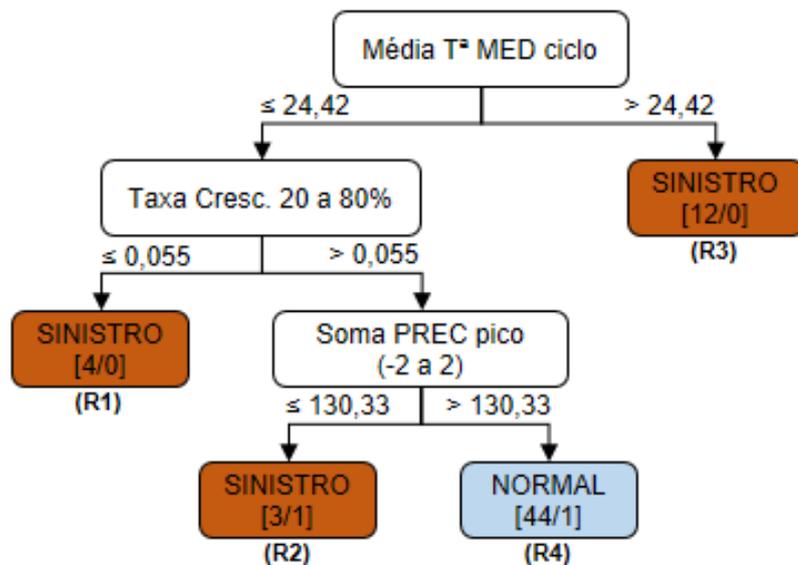


Figura 29 DT resumida para os fatores relevantes na ocorrência de sinistro na cultura da soja.

Fonte: Autor (2020).

A DT para cultura da soja (Figura 29) apontou com acurácia de 0,946 ($\pm 0,011$), três atributos como mais relevantes para classificação: A média da temperatura média durante todo o ciclo, relacionado também a taxa de crescimento e soma da precipitação nos decêndios próximos ao período de sinistro. Foram geradas três regras para áreas sinistradas e uma para normal, todas com alta taxa de acerto. Embora com menos regras, a classificação normal apresenta a maioria das amostras do grupo de treinamento. Em virtude de possuir ocorrências de chuva excessiva, seca e granizo, a temperatura aliada a taxa de crescimento além da precipitação exercem forte influência para ocorrência ou não do sinistro neste caso.

A Tabela 13 apresenta as métricas de desempenho para os três algoritmos escolhidos considerando a ocorrência de sinistro para a cultura do trigo.

Tabela 13 Medidas de desempenho dos classificadores SVM, RF e DT, considerando a ocorrência de sinistro na cultura do trigo

	Algoritmo	Condição	Precisão	Recall	F1-score	Acurácia	Amostras
SVM	Treinamento	Normal	1,000	0,949	0,972	0,985 (\pm 0,007)	27
		Sinistrado	0,981	1,000	0,991		67
		Média	0,991	0,975	0,982		94
	Teste	Normal	0,922	0,275	0,410	0,787 (\pm 0,030)	9
		Sinistrado	0,778	0,988	0,869		23
		Média	0,850	0,632	0,640		32
RF	Treinamento	Normal	1,000	1,000	1,000	1,000 (\pm 0)	27
		Sinistrado	1,000	1,000	1,000		67
		Média	1,000	1,000	1,000		94
	Teste	Normal	0,745	0,534	0,604	0,812 (\pm 0,059)	9
		Sinistrado	0,837	0,922	0,876		23
		Média	0,791	0,728	0,740		32
DT	Treinamento	Normal	0,838	0,876	0,853	0,914 (\pm 0,018)	27
		Sinistrado	0,949	0,930	0,939		67
		Média	0,894	0,903	0,896		94
	Teste	Normal	0,588	0,578	0,558	0,756 (\pm 0,059)	9
		Sinistrado	0,841	0,826	0,828		23
		Média	0,715	0,702	0,693		32

Fonte: Autor (2020).

O algoritmo SVM obteve uma acurácia global de 0,985 (\pm 0,007) para o conjunto de treinamento e 0,787 (\pm 0,030) quando submetido ao conjunto teste. Essa redução se deve à alta taxa de predição de valores falsos negativos (20,31% - Tabela 10), observados no *recall* de 0,275 e F1 *score* de 0,410 da categoria normal.

O RF mais uma vez teve acurácia de 100% para o conjunto de treinamento, porém no conjunto de teste foi de 0,812 (\pm 0,059). A diagonal secundária da matriz de confusão da RF (Tabela 10) demonstra que ocorreram classificações de valores falsos negativos e falsos positivos (13,13 e 5,63% respectivamente). Embora se tenha obtido acurácia superior ao SVM, ainda se pode observar comportamento semelhante em relação aos desempenhos na classificação da categoria sinistrados superiores que as normais.

A DT obteve uma acurácia de 0,914 (\pm 0,018) para o conjunto de treinamento e 0,756 (\pm 0,059) para o conjunto teste. É possível observar na diagonal secundária da sua matriz de confusão (Tabela 10) que foram classificados 11,88% como falsos negativos e 12,50% como falsos positivos. Erros que proporcionaram baixo desempenho da categoria normal em relação à sinistrada em todas as métricas utilizadas.

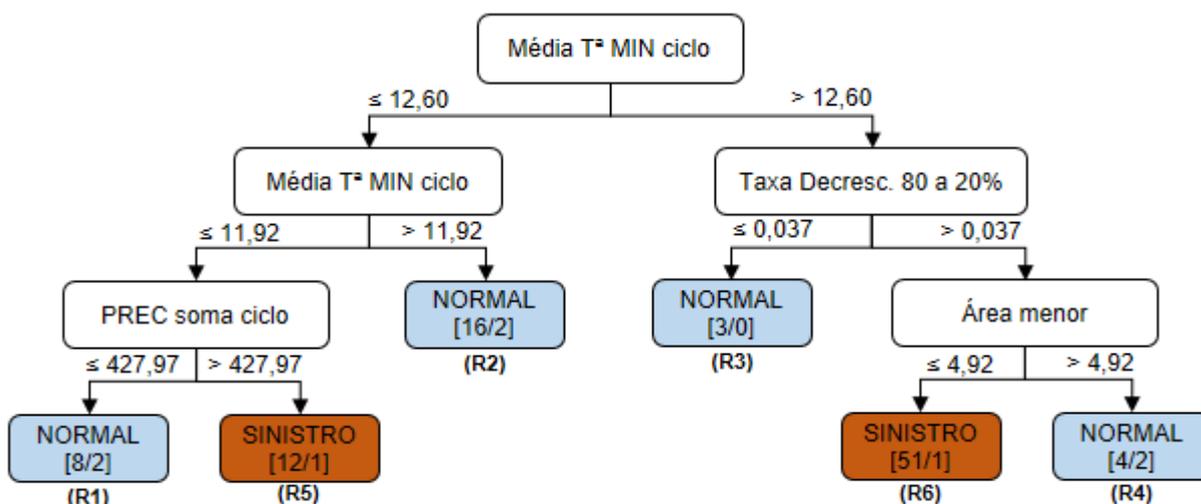


Figura 30 DT resumida para os fatores relevantes na ocorrência de sinistro na cultura do trigo.

Fonte: Autor (2020).

A DT para cultura do trigo (Figura 30) indicou com acurácia de 0,914 ($\pm 0,018$), quatro atributos como os preditores mais necessários para classificação. Foram geradas seis regras, das quais duas serviram para classificar áreas sinistradas, no entanto, são elas que contém a maioria das amostras do banco de dados. Os sinistros com número mais representativo no banco de dados para cultura do trigo são: seca e seca + geada. Embora a falta de água represente, em teoria, o fator principal de um evento de seca, neste caso, a árvore aponta a média da temperatura mínima durante todo o ciclo como fator mais relevante na ocorrência de sinistros.

O custo financeiro em prever que não aconteceria sinistro em uma situação em que a lavoura sofreu sinistro (prever normal quando era sinistrado) é maior do que no caso em que se prevê sinistro para uma lavoura que de fato não teve problemas (prever sinistro quando era normal), visto que no primeiro caso aconteceria uma saída de caixa (indenização) não esperada e, no segundo caso, haveria apenas uma alocação de recursos para outra finalidade.

A capacidade preditiva dos classificadores deve ser boa em ambas situações, pois numa situação de análise com suspeita de irregularidade a capacidade de inferir se houve ou não sinistro sobre uma lavoura, pode auxiliar a seguradora numa tomada de decisão fortuita.

Após a projeção de cada algoritmo, cabe a comparação geral de cada preditor em todas as culturas analisadas anteriormente. A Tabela 14 mostra o melhor desempenho preditivo do *Random forest*, quando comparado aos demais métodos no quesito acurácia global, exceto o SVM que foi levemente superior para o caso do milho. A maior precisão do RF na categoria positiva é interessante e mostra seu potencial para predição de ocorrência de sinistros no mercado agrícola. Vale ressaltar que os métodos foram treinados com um banco de dados de casos reais, porém pequeno e, seguramente, se o banco de dados fosse maior ou alimentado por uma empresa do ramo securitário no esquema de *data warehouse*

continuo, por exemplo, a melhoria da precisão dos classificadores seria constante e poderia servir para otimizar decisões de corporações.

Tabela 14 Resumo da acurácia global \pm desvio padrão para o conjunto de treinamento e teste dos algoritmos testados para ocorrência de sinistro nas culturas de milho, soja e trigo

Cultura	Conjunto	SVM	RF	DT
MILHO	Treinamento	0,936 (\pm 0,018)	1 (\pm 0)	0,922 (\pm 0,012)
	Teste	0,839 (\pm 0,032)	0,832 (\pm 0,060)	0,793 (\pm 0,057)
SOJA	Treinamento	0,971 (\pm 0,017)	1 (\pm 0)	0,946 (\pm 0,011)
	Teste	0,895 (\pm 0,029)	0,963 (\pm 0,034)	0,900 (\pm 0,039)
TRIGO	Treinamento	0,985 (\pm 0,007)	1 (\pm 0)	0,914 (\pm 0,018)
	Teste	0,787 (\pm 0,030)	0,812 (\pm 0,059)	0,756 (\pm 0,059)

Fonte: Autor (2020).

5.3 Previsão do tipo de sinistro ocorrido

Após a comparação dos bancos de dados de lavouras normais e sinistradas, procedeu-se com análises para a verificação dos tipos de sinistros específicos incidentes nas culturas. A Tabela 15 apresenta as métricas de desempenho: precisão, *recall*, *F1 score* e acurácia (\pm Desvio padrão), para os três algoritmos e tipos de sinistros da cultura do milho.

Tabela 15 Medidas de desempenho dos classificadores SVM, RF e DT para os tipos de sinistros (chuva excessiva e seca) da cultura do milho

	Algoritmo	Tipo de Sinistro	Precisão	Recall	F1-score	Acurácia	Amostras
SVM	Treinamento	Chuva excessiva	0,976	1,000	0,984	0,991 (\pm 0,004)	28
		Seca	1,000	0,984	0,992		64
		Média	0,988	0,992	0,988		92
	Teste	Chuva excessiva	0,990	0,989	0,989	0,993 (\pm 0,012)	9
		Seca	0,996	0,995	0,996		22
		Média	0,993	0,992	0,993		31
RF	Treinamento	Chuva excessiva	1	1	1	1 (\pm 0)	28
		Seca	1	1	1		64
		Média	1	1	1		92
	Teste	Chuva excessiva	1,000	0,967	0,982	0,990 (\pm 0,014)	9
		Seca	0,988	1,000	0,994		22
		Média	0,994	0,984	0,988		31
DT	Treinamento	Chuva excessiva	1	1	1	1 (\pm 0)	28
		Seca	1	1	1		64
		Média	1	1	1		92
	Teste	Chuva excessiva	0,989	0,912	0,947	0,971 (\pm 0,027)	9
		Seca	0,967	0,995	0,981		22
		Média	0,978	0,954	0,964		31

Fonte: Autor (2020).

De modo geral, os três classificadores utilizados obtiveram uma alta taxa de acurácia global para os dados da cultura do milho, a diferença do menor para o maior índice foi de apenas 0,022. Os algoritmos demonstraram uma alta taxa de acerto nos valores verdadeiros positivos e verdadeiros negativos, pois os sinistros que acometeram essas lavouras são em tese, opostos, pois onde há seca não pode haver chuva excessiva. Dessa forma, os classificadores facilmente identificam como fatores mais importantes na classificação dessas amostras os atributos relacionados à precipitação.

Na Figura 31, demonstra-se na representação gráfica da DT que com apenas duas regras o algoritmo foi capaz de prever todos os dados (acurácia 1 (± 0)).

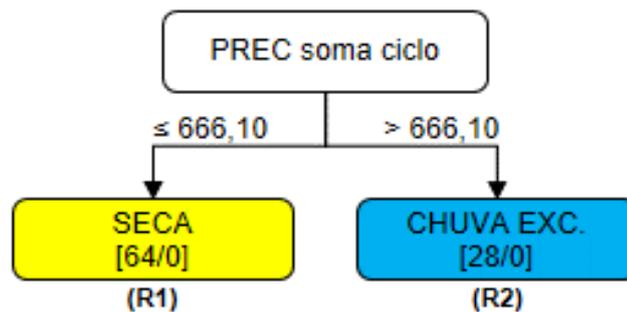


Figura 31 DT resumida para os fatores relevantes nos tipos de sinistro da cultura do milho.

Fonte: Autor (2020).

A Tabela 16 apresenta as métricas de desempenho, para os três algoritmos escolhidos para os tipos de sinistros da cultura da soja.

Tabela 16 Medidas de desempenho dos classificadores SVM, RF e DT para os tipos de sinistros (chuva excessiva, seca, granizo) da cultura da soja

	Algoritmo	Tipo de Sinistro	Precisão	Recall	F1-score	Acurácia	Amostras
SVM	Treinamento	Chuva excessiva	1	1	1		11
		Seca	1	1	1		4
		Granizo	1	1	1		3
		Média	1	1	1	1 (± 0)	18
	Teste	Chuva excessiva	1,000	0,900	0,926		4
		Seca	0,750	0,800	0,767		1
		Granizo	0,900	1,000	0,934		1
		Média	0,883	0,900	0,876	0,900 ($\pm 0,200$)	6
RF	Treinamento	Chuva excessiva	1	1	1		11
		Seca	1	1	1		4
		Granizo	1	1	1		3
		Média	1	1	1	1 (± 0)	18
	Teste	Chuva excessiva	0,900	0,950	0,912		4
		Seca	0,500	0,500	0,500		1
		Granizo	0,933	1,000	0,950		1
		Média	0,778	0,817	0,787	0,883 ($\pm 0,107$)	6
DT	Treinamento	Chuva excessiva	0,927	0,865	0,887		11
		Seca	0,563	0,900	0,686		4
		Granizo	0,180	0,300	0,225		3
		Média	0,557	0,688	0,599	0,778 ($\pm 0,043$)	18
	Teste	Chuva excessiva	0,892	0,625	0,722		4
		Seca	0,216	0,700	0,327		1
		Granizo	0,200	0,300	0,234		1
		Média	0,436	0,542	0,428	0,583 ($\pm 0,083$)	6

Fonte: Autor (2020).

O banco de dados de sinistros, para cultura da soja, contém o menor número de amostras dentre as culturas escolhidas. Para tentar minimizar os erros de classificação optou-se por utilizar três tipos de sinistros desse banco de dados: chuva excessiva, seca, granizo.

Basicamente os resultados das métricas: precisão, *recall* e F1 score são maiores para a categoria chuva excessiva no conjunto teste, nos três classificadores, mostrando a tendência dos algoritmos em classificar as amostras nessa categoria. Isso ocorre, pois são quatro amostras para chuva excessiva e apenas uma amostra cada para seca e granizo. Nesse caso, foi obtida uma maior variação de acurácia global entre os três classificadores, sendo o SVM com maior valor 0,900 ($\pm 0,200$) e DT com 0,583 ($\pm 0,083$).

A representação gráfica dessa DT (Figura 32) conta com acurácia de 0,778 ($\pm 0,043$) e mostra a soma da precipitação durante o ciclo e a média da temperatura média no pico do sinistro como fatores mais influentes na definição dos sinistros. Com apenas três regras, foi possível classificar as amostras e apenas na primeira regra ocorreram erros, nas demais regras se obteve total acerto.

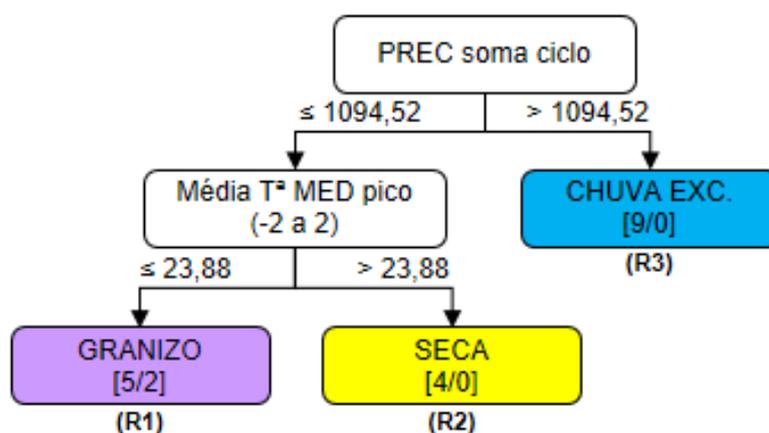


Figura 32 DT resumida para os fatores relevantes nos tipos de sinistro da cultura da soja.

Fonte: Autor (2020).

A Tabela 17 apresenta as métricas de desempenho, para os três algoritmos escolhidos para os tipos de sinistros da cultura do trigo.

Tabela 17 Medidas de desempenho dos classificadores SVM, RF e DT para os tipos de sinistros (seca, seca + geada) da cultura do trigo

	Algoritmo	Tipo de Sinistro	Precisão	Recall	F1-score	Acurácia	Amostras
SVM	Treinamento	Seca	1,000	0,979	0,986	0,988 (± 0,008)	31
		Seca + Geada	0,979	1,000	0,986		29
		Média	0,990	0,990	0,986		60
	Teste	Seca	0,760	0,962	0,845	0,810 (± 0,062)	11
		Seca + Geada	0,946	0,635	0,748		9
		Média	0,853	0,799	0,797		20
RF	Treinamento	Seca	1	1	1	1 (± 0)	31
		Seca + Geada	1	1	1		29
		Média	1	1	1		60
	Teste	Seca	0,873	0,914	0,880	0,865 (± 0,083)	11
		Seca + Geada	0,917	0,810	0,834		9
		Média	0,895	0,862	0,857		20
DT	Treinamento	Seca	0,902	0,951	0,923	0,915 (± 0,023)	31
		Seca + Geada	0,945	0,876	0,905		29
		Média	0,924	0,914	0,914		60
	Teste	Seca	0,816	0,831	0,812	0,805 (± 0,085)	11
		Seca + Geada	0,835	0,786	0,795		9
		Média	0,826	0,809	0,804		20

Fonte: Autor (2020).

Nessa análise, restaram dois sinistros: seca e seca seguida de geada. Embora sejam eventos parcialmente semelhantes, os algoritmos foram capazes de os diferenciar. Para o conjunto de treinamento, obtiveram-se acurácias (± desvio padrão) acima de 0,915 (± 0,023). Para o conjunto teste, os três algoritmos tiveram acurácias entre 0,805 (± 0,085) e 0,8650 (± 0,0838). De modo geral, os algoritmos classificaram com maior acurácia a categoria seca, o que pode ser observado nos parâmetros de *recall* e *F1 score* dessa categoria em relação à seca + geada, em todos os algoritmos testados.

A representação gráfica dessa DT (Figura 33) conta com acurácia de 0,915 ($\pm 0,023$) e indica a média da temperatura mínima nos decêndios próximos ao sinistro como fator mais influente na definição dos sinistros. A taxa de crescimento de 20 a 80% e o IV no pico vegetativo também são considerados relevantes nas decisões, pois têm relação com a baixa precipitação que ocasiona a seca, pois os índices de água no ciclo o IV e o crescimento são baixos. Com 5 regras, foi possível classificar as amostras e apenas as regras R2 e R3 contêm erros, nas demais regras se obteve total acerto.

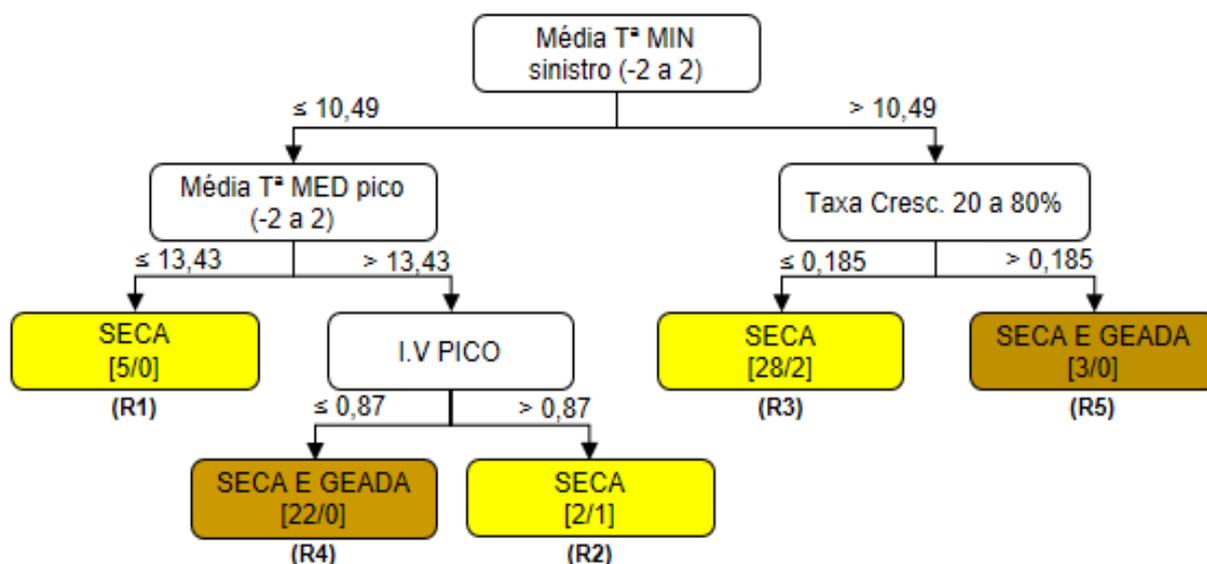


Figura 33 DT resumida para os fatores relevantes nos tipos de sinistro da cultura do trigo.

Fonte: Autor (2020).

A Tabela 18 mostra o compilado da acurácia global dos algoritmos testados em relação aos sinistros das culturas de interesse. Nela são observados maiores índices para o SVM na cultura do milho e soja. Para a cultura do trigo a maior acurácia foi do algoritmo RF. A DT em todos os casos apresentou os menores índices de acurácia global, porém é interessante utilizá-la, caso seja necessário facilitar a legibilidade dos fatores mais influentes nas classificações.

Tabela 18 Resumo da acurácia global \pm desvio padrão para o conjunto de treinamento e teste dos algoritmos testados nos sinistros das culturas de milho, soja e trigo

Cultura	Conjunto	SVM	RF	DT
MILHO	Treinamento	0,991 ($\pm 0,004$)	1 (± 0)	1 (± 0)
	Teste	0,993 ($\pm 0,012$)	0,990 ($\pm 0,014$)	0,971 ($\pm 0,027$)
SOJA	Treinamento	1 (± 0)	1 (± 0)	0,778 ($\pm 0,043$)
	Teste	0,900 ($\pm 0,200$)	0,883 ($\pm 0,107$)	0,583 ($\pm 0,083$)
TRIGO	Treinamento	0,988 ($\pm 0,008$)	1 (± 0)	0,915 ($\pm 0,023$)
	Teste	0,810 ($\pm 0,062$)	0,865 ($\pm 0,083$)	0,805 ($\pm 0,085$)

Fonte: Autor (2020).

5.4 Relação entre dados de satélites e laudos de sinistros

A fim de compreender a existência de uma relação entre o sinistro relatado e o perfil espectro-temporal das áreas, efetuou-se uma análise de ambas as informações com análise visual do perfil de EVI/Landsat-8 no decorrer do ciclo, para confirmar as anomalias prejudiciais informadas nos laudos. Entretanto, essa análise não permitiu concluir quais alterações do perfil estavam associadas aos eventos climáticos relatados.

De forma visual, os perfis podem ser considerados normais, pois não é possível identificar quedas bruscas em momentos pontuais ou redução do EVI em períodos maiores associados diretamente ao evento causador de sinistro, justificando a aplicação de técnicas como a mineração de dados, a fim de verificar a existência de padrões ocultos nessa base de dados.

Em relação aos sinistros ocorridos, foram observadas nos laudos das vistorias algumas especificidades dos eventos que valem ser apontadas:

Com relação ao período de chuva excessiva nas culturas do milho segunda safra, soja e trigo, geralmente, este foi curto e, em alguns casos, coincide com a queda natural do EVI devido à senescência das plantas. Quando mais longo, pode causar interferência nas leituras dos sensores devido à grande quantidade de nuvens e baixa atividade fotossintética. No caso do trigo, foi apontado como período crítico já no final do ciclo da cultura, interferindo mais na qualidade do grão que no desenvolvimento da planta e IV propriamente ditos.

Esperava-se que a queda do EVI em casos de geada fosse mais pronunciada imediatamente após a ocorrência deste evento, porém, não se notou tal comportamento nos perfis de EVI nos laudos observados. Pois, conforme Taiz e Zeiger (2004), as plantas quando danificadas por geada, apresentam redução na absorção da água, murchamento e crescimento lento; as folhas apresentam descoloração, lesões ou necroses, apresentando encharcamento e redução na fotossíntese.

Os casos de granizo vistoriados, quando ocorreram no meio do ciclo da cultura foram de baixa intensidade, não refletindo em alterações perceptíveis no EVI, mantendo a produtividade normal da cultura, ou ocorreram logo no início do ciclo do cultivo, quando ainda a entrelinha da cultura estava bem aparente, representando início do perfil espectro-temporal, pois ainda as plantas têm a reflectância menor que a do solo adjacente. Certamente o granizo pode causar injúrias na planta, porém quando detectado pouco tempo após o plantio, apenas de forma remota, é difícil atribuir exclusivamente ao granizo a redução de produtividade.

Em muitos casos, o período crítico de estiagem, acaba ocorrendo da metade para o final do ciclo das culturas, quando a planta naturalmente começa a fase de senescência e seu EVI naturalmente começa a decrescer, sendo difícil atribuir exclusivamente à seca, a queda do EVI. Por outro lado, em alguns casos quando o período de seca aliado a altas temperaturas se dá nas fases iniciais da planta, nota-se que o EVI em todo período permanece baixo,

evidenciando que a seca prejudicou o desenvolvimento da planta por completo até a colheita e não apenas de forma pontual. Esse evento climático, em especial, pode provocar a redução mais significativa de EVI, pois prejudica a plena instalação da cultura e impede que a planta se desenvolva naturalmente. Um dos sinais mais característicos é a redução do tamanho e atividade fotossintética.

6 CONCLUSÕES

Para confirmação de ocorrência ou não do sinistro, o método RF demonstrou desempenho levemente superior para o caso particular desse banco de dados, exibindo uma boa performance, especialmente na predição de ocorrência de sinistro nas culturas de milho, soja e trigo analisadas. As métricas de desempenho: precisão, *recall* e *F1 score* também foram superiores em relação aos demais.

Em relação à verificação dos sinistros incidentes nas culturas sinistradas, são observados maiores índices para o SVM na cultura do milho e soja. Para a cultura do trigo a maior acurácia foi do algoritmo RF.

Para se obter estes resultados foi envolvido baixo custo computacional visto que todas as etapas de mineração de dados foram realizadas com processamento em nuvem.

Para aprimoramento do estudo, seria necessária a contínua alimentação do banco de dados com novos casos de lavouras sinistradas e de lavouras consideradas normais ou, neste caso, sem sinistro e com baixa incidência de dano ou baixa redução na produtividade, o que não seria problema para as empresas seguradoras, pois a maioria dos laudos atuais são feitos de forma mais automática, registrada em aplicativos e *softwares* de GIS, evitando o trabalho moroso da transcrição de laudos impressos para formatos digitais, ou seja, alimentação precisa, valiosa e contínua de informações reais.

6.1 Trabalhos futuros

Como extensão desta pesquisa, podem ser testados outros satélites ou sensores com resoluções espaciais e temporais diferentes, bem como outras fontes de dados climáticos, outra forma de interpolação de dados faltantes e outros algoritmos de classificação isolados ou combinados, a fim de verificar o aumento de desempenho dos classificadores.

Uma forma de tornar o processo completo mais ágil, seria automatizar principalmente a etapa de análise de padrões sazonais no TIMESAT, visto que as demais etapas já possuem alguns *scripts* para a execução.

Vale salientar, que as análises deste trabalho foram realizadas após transcorrer o período completo do ciclo das culturas, sendo assim, um questionamento pertinente acerca da confirmação de eventos adversos nas culturas seguradas seria determinar qual o período mínimo de análise do ciclo para poder inferir alguma informação com grau de confiança

razoável, antes da vistoria final, trazendo, assim, um ganho de informação valioso para as seguradoras, pois, de fato, poderiam determinar áreas prioritárias para as visitas a campo.

6.2 Contribuições esperadas

A contribuição esperada de metodologias dessa natureza é a redução do processo de avaliação pericial e gastos relacionados, promovendo o acompanhamento das lavouras agrícolas, a redução das fraudes associadas à concessão de recursos oriundos das seguradoras e órgãos públicos, como também servir de garantia ao agricultor e legitimar seu direito. Por fim, com o uso de sensoriamento remoto dar-se-á mais segurança e fidedignidade dos resultados aos bancos, seguradoras e resseguradoras que atuam no crédito e seguro rural. Dessa forma, atuando com menor risco, é possível que se diminua o prêmio pago pelos produtores contratantes do seguro, também a acurácia desses dados servirá de auxílio a produtores e peritos nas tomadas de decisão.

6.3 Considerações finais

A integração de diferentes áreas do conhecimento para a resolução de problemas, como a área agrícola e a inteligência artificial é um desafio. Há poucos trabalhos que utilizam inteligência artificial no mercado segurador agrícola, visto que a maior concentração de estudos envolve o mercado segurador automotivo. Sendo assim, devido ao potencial dessas áreas, este trabalho se mostra relevante na atualidade.

No desenvolvimento desta pesquisa, pode-se constatar a possibilidade de aplicação do processo KDD na área agrícola, possibilitando a descoberta de padrões ocultos em bases de dados que geram informações valiosas e, neste caso, foi possível confirmar eventos sinistrantes em lavouras reais.

O material disponível para a realização do experimento foi considerado satisfatório no desenvolvimento do trabalho. A utilização das imagens Landsat-8 e as ferramentas *Google Earth Engine*, *Nasa Power*, *Minitab*, *TIMESAT* e *Google Colaboratory* com suas ferramentas de programação em Python, foi considerado eficiente, permitindo que se atingissem os objetivos propostos. Em especial, a metodologia para extração dos parâmetros de sazonalidade ou Análise de Padrões Sazonais realizada no *software* *TIMESAT*, permitiu a extração de informações valiosas do perfil temporal do EVI das três culturas analisadas, como

o pico vegetativo, as datas de semeadura, colheita e outras métricas. De modo geral, obteve-se uma baixa variação entre as datas reais e as estimadas, mesmo tendo o perfil temporal disposto em decêndios. Embora a extração dos dados de cada perfil seja exaustiva, esse tipo de análise é eficiente e permite sua aplicação para diversas finalidades e escalas.

Ainda, antes da etapa de mineração de dados, tem-se que os dados oriundos da plataforma Nasa *Power* são uma fonte confiável de informação e que ela pode ser utilizada em casos futuros, devido seu potencial de uso e facilidade de navegação.

Este estudo teve o objetivo de avançar na predição/confirmação de ocorrências de sinistro do ramo agrícola, sendo esta uma variável de grande importância para a estabilidade das seguradoras e de corporações envolvidas com os segmentos de seguro e crédito rural, auxiliando, assim, o procedimento de peritagem e fiscalização de ações relacionadas a esses segmentos agrícolas. Vale salientar que informações reais dessa natureza são de difícil obtenção em bancos ou seguradoras, pois se tratam de informações bancárias, portanto, sigilosas, impossíveis de reprodução externa mesmo que para estudos. Contudo, através de estreita relação com uma empresa de perícias agrícolas, foi possível obter o conteúdo necessário do banco de dados das vistorias realizadas de 2016 a 2018 por esta equipe, o que se mostrou excelente, pois, dessa forma, foi possível observar diversos casos de sinistros, culturas e regiões distintas.

REFERÊNCIAS

- ABREU, A. L. E. **Bootstrap e modelos de Support Vector Machine – SVM**. 2016. 117 f. Tese (Doutorado em Métodos Numéricos) - Universidade Federal do Paraná, Curitiba – PR, 2016.
- ALMEIDA, T. F. **Julgamento de seguro agrícola na cooperativa Sicredi**. 2016, 34 f. Trabalho de Conclusão de Curso (Graduação em Agronomia) - Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016.
- ALMEIDA, T. S.; FONTANA, D. C.; MARTORANO, L. G.; BERGAMASCHI, H. Índices de vegetação para a cultura da soja em diferentes condições hídricas e de sistema de manejo do solo. SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12., 2005, Goiânia - GO. **Anais...** Goiânia – GO: INPE, 2005. p. 17-24.
- APARECIDO, L. E. O. ROLIM, G. S.; RICHETTI, J.; SOUZA, P. S.; JOHANN, J. A. Köppen, Thornthwaite and Camargo climate classifications for climatic zoning in the State of Paraná, Brazil. **Ciência e Agrotecnologia**, Lavras – MG, v. 40, n. 4, p. 405–417, ago. 2016.
- BANCO NACIONAL DO DESENVOLVIMENTO ECONÔMICO E SOCIAL - BNDES. **Instrumentos de gestão do risco agrícola: o caso Brasil**. 2011. (Informativo Técnico SEAGRI. n. 1. 2011). Disponível em: <https://web.bndes.gov.br/bib/jspui/bitstream/1408/2440/1/InformativoSEAGRI_n.1%2c%20jan.2011._P_BD.pdf> Acesso em: 24 out. 2017.
- BECK, P. S. A.; ATZBERGER, C.; HOGDA, K. A.; JOHANSEN, B.; SKIDMORE, A. K. Improved monitoring of vegetation dynamics at very high latitudes: a new method using MODIS NDVI. **Remote Sensing of Environment**, v. 100, n. 3, p. 321–334, 2006.
- BECKER, W. R.; JOHANN, J. A.; RICHETTI, J.; MERCANTE, E.; ESQUERDO, J. C. D. M.; SILVA JUNIOR, C. A.; PALUDO, A. Agricultural soybean and corn calendar based on moderate resolution satellite images for southern Brazil. **Semina: Ciências Agrárias**, Londrina – PR, v. 41, n. 2, p. 1–9, 2020.
- BECKER, W. R.; JOHANN, J. A.; RICHETTI, J.; SILVA, L. C. A. Data mining techniques for separation of summer crop based on satellite images. **Engenharia Agrícola**, Jaboticabal – SP, v. 37, n. 4, p. 750–759, 2017.
- BOSE, B. E.; GUYON, I. M.; VAPNIK, V. N. A Training algorithm for optimal margin classifiers. In: ANNUAL WORKSHOP ON COMPUTACIONAL LEARNING, 5, 1992, Pittsburgh. **Proceedings...** ACM Press. Pittsburgh: Haussler D, jul 1992. p.144-152 .
- BRASIL - Banco Central do Brasil. **Resolução nº 4.427**, de 25 de junho de 2015. Autoriza a utilização do sensoriamento remoto para fins de fiscalização de operações de crédito rural e determina o registro das coordenadas geodésicas do empreendimento financiado por operações de crédito rural no Sistema de Operações do Crédito Rural e do Proagro (Sicor). Disponível em: <http://www.bcb.gov.br/pre/normativos/res/2015/pdf/res_4427_v1_O.pdf>. Acesso em: 10 out. 2017.
- BRASIL. Banco Central do Brasil. **Manual do crédito rural (MCR)**. 2018. Disponível em: <<http://www3.bcb.gov.br/mcr>> Acesso em: 12 out 2018.
- BREIMAN, L. Random forests. **Machine learning**, v. 45, n. 1, p. 5–32, 2001.

- BUAINAIN, A. M.; VIEIRA, P. A.; CURY, W. J. M. **Gestão do risco e seguro na agricultura brasileira**. 1. ed. Rio de Janeiro: Funenseg, 2014.
- BURKE, M.; EMERICK, K. Adaptation to climate change: Evidence from US agriculture. **American Economic Journal: Economic Policy**, v. 8, n. 3, p. 106–140, 2016.
- CHANDLER, W. S., STACKHOUSE JR, P. W.; HOELL, J. M.; WESTBERG, D. ZHANG, T. NASA Prediction of worldwide energy resource high resolution meteorology data for sustainable building design. **Proceedings of the Solar**. 2013. Baltimore, Maryland Conference (American Solar Energy Society) 2013, April 16-20.
- COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB – **Acompanhamento da safra brasileira de grãos**. 2012. Disponível em: <http://www.conab.gov.br/OlalaCMS/uploads/arquivos/12_05_10_08_49_52_boletim_maio_2012.pdf>. Acesso em: 08 julho 2017.
- CONTINI, E.; PENA JÚNIO, M. A. G.; SANTANA, C. A. M.; MARTHA JÚNIOR, G. Exportações: motor do agronegócio brasileiro. **Revista de Política Agrícola**, Brasília – DF, v. 21, n. 2, p. 88-102, 2012.
- COUTO JUNIOR, A. F. et al. Tratamento de ruídos e caracterização de fisionomias do cerrado utilizando séries temporais do sensor MODIS. **Revista Árvore**, Viçosa -MG, v. 35, n. 3, supl. 1, p. 699-705, jun. 2011. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-67622011000400014&lng=en&nrm=iso>. Acesso em: 15 set. 2020.
- CRIVELENTI, R.; COELHO, R. M.; ADAMI, S. F.; OLIVEIRA, S. R. M. Mineração de dados para inferência de relações solo-paisagem em mapeamentos digitais de solo. **Pesquisa Agropecuária Brasileira**, Brasília – DF, v. 44, n. 12, p. 1707–1715, dez. 2009.
- CUNHA, G. R.; ASSAD, E. D. An overview of the RBA special issue on crop zoning in Brazil. **Revista Brasileira de Agrometeorologia**, Passo Fundo – RS, v. 9, n. 3, p. 377–385, 2001.
- EKLUNDH, L.; JÖNSSON, P. **TIMESAT 3.2** with parallel processing. Software Manual, 2015a. 88 p.
- EKLUNDH, L.; JÖNSSON, P. TIMESAT: a software package for time-series processing and assessment of vegetation dynamics. In: KUENZER, C.; DECH, S.; WAGNER, W. (Eds.). **Remote Sensing Time Series**. Springer International Publishing, 2015b. p. 141–158.
- EPIPHANIO, J. C. N.; GLERIANI, J. M.; FORMAGGIO, A. R.; RUDORFF, B. F. T. Índices de vegetação no sensoriamento remoto da cultura do feijão. **Pesquisa Agropecuária Brasileira**, Brasília – DF, v. 31, n. 6, p. 445-454, jun. 1996.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. Knowledge discovery and data mining: towards a unifying framework. In: INTERNATIONAL CONFERENCE KNOWLEDGE DISCOVERY AND DATA MINING, 1996. Portland OR. **Proceedings...** 1996, p. 82–88.
- FEDERAÇÃO DA AGRICULTURA DO ESTADO DO PARANÁ - FAEP. **Cartilha sobre Seguros Rurais e PROAGRO**. 2017. Disponível em: <<http://www.sistemafaep.org.br/servico/guia-de-seguros-rurais-e-proagro>> Acesso em: 17 out. 2017.

FEDERAÇÃO DA AGRICULTURA DO ESTADO DO PARANÁ – FAEP. Federação Nacional de Seguros Gerais - FenSeg, Confederação da Agricultura e Pecuária do Brasil, CNA; Sistema Ocepar. **Guia seguros rurais e Proagro**. 2016. Disponível em: <https://www.cnabrazil.org.br/assets/arquivos/bibliotecas/cartilha_seguro_rural.pdf> Acesso em: 05 maio 2019.

FERNANDES, E. **Monitoramento da cultura de cana-de-açúcar no estado de São Paulo por meio de imagens SPOT *Vegetation* e dados meteorológicos**. 2009. 114 f. Dissertação (Mestrado em Engenharia Agrícola) - Universidade Estadual de Campinas, Campinas/SP, 2009.

FORMAGGIO, A. R.; SANCHES, I. D. **Sensoriamento remoto em agricultura**. São José dos Campos - SP: Oficina de Textos, 2017.

FORNAZIER, A.; SOUZA, P. M.; PONCIANO, N. J. A Importância do seguro rural na redução de riscos da agropecuária. **Revista de Estudos Sociais**, Cuiabá – MT, v. 14, n. 28, p. 39-52, aug. 2014. ISSN 2358-7024.

GALVÃO, N. D.; MARIN, H. D. F. Técnica de mineração de dados: uma revisão da literatura. **ACTA Paulista de Enfermagem**, São Paulo – SP, v. 22, n. 5, p. 686–690, 2009.

GAMON, J. A.; FIELD, C. B.; GOULDEN, M. L.; GRIFFIN, K. L.; HARTLEY, A. E.; JOEL, G.; PEÑUELAS, J.; VALENTINI, R. Relationships between NDVI, canopy structure, and photosynthesis in three Californian vegetation types. **Ecological Applications**, v. 5, p. 28-41, 1995.

GANDHI, G. M.; PARTHIBAN, S.; THUMMALU, N.; CHRISTY, A. Ndvi: Vegetation change detection using remote sensing and gis. A case study of vellore district. **Procedia Computer Science**, v. 57, p. 1199–1210, 2015.

GRZEGOZEWSKI, D.; SOUZA, C. H. W. MERCANTE, E.; URIBE OPAZO, M.; BECKER, W.; JOHANN, J. (2013). Uso de imagens de satélite para detecção de datas de semeadura, pico vegetativo e colheita das culturas do milho e soja na região Oeste do Paraná -Safrá 2011/2012. SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO – SBSR, 16. Foz do Iguaçu – PR. **Anais...** Foz do Iguaçu -PR: INPE, 2013. p. 1610-1617.

HARWOOD, J.; HEIFNER, R.; COBLE, K.; PERRY, J.; SOMWARU, A. Managing risk in farming: concepts, research and analysis. **Agricultural Economic Report**, n. 774, p. 83, 1999.

HOLBEN, B. N. Characteristics of maximum value composite images from temporal AVHRR data. **International Journal of Remote Sensing**, v. 7, n. 11, p. 1417-1435, 1986.

HOLLANDER M, WOLFE D. **Nonparametric statistical methods**. John Wiley and Sons. 1973.

HUETE, A. D. L.; MIURA, T.; RODRIGUEZ, E. P.; GAO, X.; FERREIRA, L. G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote Sensing of the Environment**. v.83, n.1-2, p.195-213, nov, 2002.

HUETE, A. R. A soil-adjusted vegetation index (SAVI). **Remote Sensing of Environment**, v. 25, n. 3, p. 295–309, 1988.

HUETE, A.; JUSTICE, C.; LEEUWEN, W. VAN. **Modis vegetation index (MOD 13): algorithm theoretical basis document**. Washington: National Aeronautics and Space Administration 1999.

INSTITUTO BRAILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Banco de dados agregados - **SIDRA - estimativas de safra**. 2017. Disponível em: <<https://sidra.ibge.gov.br/Tabela/1612>>. Acesso em: 13 abr. 2019.

JAMES, G; WITTEN, D; HASTIE, T; TIBSHIRANI R. **An introduction to statistical learning: with applications in R**. Springer Publishing Company, Incorporated, 2014.

JOHANN, J. A.; BECKER, W. R.; URIBE-OPAZO, M. A.; MERCANTE, E. Uso de imagens do sensor orbital modis na estimação de datas do ciclo de desenvolvimento da cultura da soja para o estado do Paraná – Brasil. **Engenharia Agrícola**, Jaboticabal – SP, v. 36, n. 1, p. 126–142, fev. 2016.

JOHANN, J. A.; ROCHA, J. V.; DUFT, D. G.; LAMPARELLI, R. A. C. Estimativa de áreas com culturas de verão no Paraná, por meio de imagens multitemporais EVI/Modis. **Pesquisa Agropecuária Brasileira**, Brasília – DF, v. 47, n. 9, p. 1295–1306, set. 2012.

JÖNSSON, P.; EKLUNDH, L. Seasonality extraction by function fitting to time-series of satellite sensor data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 40, n. 8, p. 1824–1832, 2002.

LOPES, D.; LOWERY, S.; PEROBA, T. L. C. Crédito rural no Brasil: desafios e oportunidades para a promoção da agropecuária sustentável. **Revista do BNDES**, Rio de Janeiro – RJ, 45, p. 155-196, jun. 2016. Disponível em: <<https://web.bndes.gov.br/bib/jspui/bitstream/1408/9518/1/5-%20Cr%C3%A9dito%20rural%20no%20Brasil%20desafios%20e%20oportunidades%20para%20a%20promo%C3%A7%C3%A3o%20da%20agropecu%C3%A1ria%20sustent%C3%A1vel.pdf>>. Acesso em: 22 maio 2019.

MACEDO, M. A.; RUDORFF, B. F. T. **Geotecnologias no seguro agrícola do milho safrinha**. Agricultura, São Paulo – SP, v. 50, n. 1, p. 43-52, 2003.

MAGALHAES, P. C.; DURÃES, F. O. M. **Fisiologia da produção de milho**. Sete Lagoas – MG: Embrapa Milho e Sorgo. 2006. (Circular Técnica 76).

MARCHESAN, R. S. P. et al. Séries temporais de índices de vegetação do sensor *modis* para detecção de limiares de desmatamento no bioma mata atlântica. SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 19, 2019, Santos. **Anais...** Santos – SP: SBSR, 2019.

MARTIN, L. **Crédito rural agora é mais simples e viável**. 2015. Disponível em: <<http://www.engesat.com.br/credito-rural-simples-viavel/>> Acesso em: 30 set. 2017.

MENESES, P. R; ALMEIDA, T. **Introdução ao processamento de imagens de Sensoriamento remoto**. UnB, CNPq, Brasília, 266p, 2012. Disponível em: <<http://www.cnpq.br/documents/10157/56b578c4-0fd5-4b9f-b82a-e9693e4f69d8>>. Acesso em: 30 maio 2019.

MENZE, B. H.; KELM, B. M.; MASUCH, R.; HIMMELREICH, U.; BACHERT, P.; PETRICH, W.; HAMPRECHT, F. A. A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data. **BMC bioinformatics**, v. 10, n. 213, n. p., 2009.

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO - MAPA. **Guia de seguros rurais**. 2018 Disponível em <<http://www.agricultura.gov.br/assuntos/riscos-seguro/seguro-rural/publicacoes-seguro-rural/guia-dos-seguros-rurais>>. Acesso em: 05 maio 2019.

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO - MAPA. Secretaria de Política Agrícola. **Projeções do agronegócio**. Brasil 2018/19 a 2028/29: projeções de longo prazo. Brasília - DF: MAPA/ACE. 2019a.

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO - MAPA. **Percentuais de subvenção ao seguro rural**. 2019b. Disponível em: <<http://www.agricultura.gov.br/assuntos/riscos-seguro/seguro-rural/limites-percentuais-de-subvencao>>. Acesso em: 05 maio 2019.

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO - MAPA. PSR – Programa de Subvenção ao Prêmio do Seguro Rural. **Atlas do Seguro Rural**. 2019c. Disponível em: <<http://indicadores.agricultura.gov.br/atlasdoseguro/index.htm>> Acesso em: 31 abr. 2019.

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO - MAPA. **Encontro sobre melhorias dos serviços de seguros rurais**. 2019d. Disponível em: <<https://www.gov.br/agricultura/pt-br/assuntos/noticias/mapa-discute-melhorias-dos-servicos-de-seguros-rurais-com-corretores-e-seguradoras-em-sao-paulo>>. Acesso em: 16 mar. 2020.

MISHRA, N. B.; CHAUDHURI, G. Spatio-temporal analysis of trends in seasonal vegetation productivity across Uttarakhand, Indian Himalayas, 2000–2014. **Applied Geography**, v. 56, p. 29–41, jan. 2015.

MORAES, E. C. **Fundamentos de sensoriamento remoto**. Apostila. Instituto Nacional de Pesquisas Espaciais (INPE/MCTI). 2015. Disponível em: <http://www.dsr.inpe.br/vcsr/files/capitulo_1.pdf>. Acesso em: 15 maio 2019.

MOREIRA, M. A. **Fundamentos do sensoriamento remoto e metodologias de aplicação**. Viçosa – MG: Livraria UFV, 2005.

NONATO, R. T.; OLIVEIRA, S. R. DE M. Técnicas de mineração de dados para identificação de áreas com cana-de-açúcar em imagens Landsat 5. **Engenharia Agrícola**, Jaboticabal – SP, v. 33, n. 6, p. 1268–1280, 2013.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT - OECD. **Agricultural Outlook 2015-2024: HIGHLIGHTS**. 2015. Disponível em: <<http://www.fao.org/3/a-i4738e.pdf>>. Acesso em: 03 ago. 2017.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT - OECD. **Risk management in agriculture: what role for governments?** 2017. Disponível em: <<http://www.oecd.org/agriculture/agricultural-policies/49003833.pdf>>. Acesso em: 10 set. 2017.

OZAKI, V. A.; CAMPOS, R. C.; Reduzindo a incerteza no mercado de seguros: uma abordagem via informações de sensoriamento remoto e atuária. **Revista Brasileira de Economia**, Rio de Janeiro, v. 71, n. 4, p. 489-514, dez. 2017.

PIONEER. **Como estimar a produtividade do milho**. Agronegócio em foco. 2019b. Disponível em: <<http://www.pioneersementes.com.br/blog/26/como-estimar-a-produtividade-do-milho>> Acesso em: 25 maio 2019.

PIONEER. **Estimando a produtividade na cultura da soja**. Agronegócio em foco. 2019a. Disponível em: <<http://www.pioneersementes.com.br/blog/46/estimando-a-produtividade-na-cultura-da-soja>> Acesso em: 25 maio 2019.

- PORTAL TSS - **Tudo sobre seguros**. Entenda o seguro rural. 2015. Disponível em: <<http://www.tudosobresseguros.org.br/portal/pagina.php?l=584>> Acesso em: 13 out. 2017.
- RAMOS, R. C. O seguro rural no Brasil: origem, evolução e proposições para aperfeiçoamento. **Informações Econômicas**, v. 39, n. 3, p. 16, 2009.
- ROSSETTO, C. A. V et al. Efeito da disponibilidade hídrica do substrato, da qualidade fisiológica e do teor de água inicial das sementes de soja no processo de germinação. **Science agric.** Piracicaba, v. 54, n. 1-2, p. 97-105, Jan. 1997.
- MOTTA, J. L. G.; FONTANA, D. C.; WEBER, E. Evolução temporal do NDVI/NOAA em áreas cobertas por *pixels* com proporções variáveis de soja. **Revista Brasileira de Agrometeorologia**, Santa Maria, v. 11, n. 2, p. 353-369, 2003.
- SANTANA, C. A. M.; NASCIMENTO, J. R. **Public policies and agricultural investment in Brazil**: final report. Brasília: Food and Agriculture Organization of the United Nations (FAO), 2012.
- SARADHI, V., KAMIK, H., MITRA, P. A Decomposition method for support vector clustering. INTERNATIONAL CONFERENCE ON INTELLIGENT SENSING AND INFORMATION PROCESSING – ICISIP, 2. 2005. Chennai, Índia. In: **Proceedings...** Chennai, Índia: IEEE,, 2005. p. 268-271.
- SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**. v. 36, n. 8, p. 1627-1639, 1964.
- SHAO, Y. et al. An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data. **Remote Sensing of Environment**, v. 174, p. 258-265, mar. 2016.
- SILVA, J. A.; TEIXEIRA, M. S. G.; SANTOS, V. G. Avaliação do programa de subvenção ao prêmio do seguro rural – 2005 a 2012. **Revista de Política Agrícola**, Brasília – DF, v. 23, n. 1, p. 105–118, 2014.
- SISTEMA DE TECNOLOGIA E MONITORAMENTO AMBIENTAL DO PARANÁ – SIMEPAR. **Dados das estações meteorológicas 2019**. Disponível em: http://www.simepar.br/prognozweb/simepar/dados_estacoes/25264916. Acesso em: 15 maio 2019
- SOUZA, Z. M. et al. Análise dos atributos do solo e da produtividade da cultura de cana-de-açúcar com o uso da geoestatística e árvore de decisão. **Ciência Rural**, Santa Maria – RS, v. 40, n. 4, p. 840–847, abr. 2010.
- SUPERINTENDÊNCIA DE SEGUROS PRIVADOS - SUSEP –. **O que é seguro rural?** 2019. Disponível em: <<http://www.susep.gov.br/menu/informacoes-ao-publico/planos-e-produtos/seguros/seguro-rural>> Acesso em: 01 maio 2019.
- TAIZ, L.; ZEIGER, E. **Fisiologia vegetal**. Porto Alegre/RS: Artmed, 2004. p. 449-484
- TAN, P. N.; STEINBACH, M.; KUMAR, V. **Introdução ao data mining**: mineração de dados. Rio de Janeiro: Ciência Moderna, 2009. 932 p.
- USGS, science for a changing world. **Landsat 8 | Landsat Missions**. 2018. Disponível em: <<https://www.usgs.gov/land-resources/nli/landsat>>. Acesso em: 23 abril 2019.
- VAPNIK, V. **The nature of statistical learning theory**. New York: Springer-Verlag. 1995.

VAPNIK, V.; DRUCKER, H.; BURGESS, C. J.; KAUFMAN, L.; SMOLA, A. **Support vector regression machines**. Advances in neural information processing systems. Morgan Kaufmann Publishers, p. 155–161, 1997.

ZHANG, P.; ZHANG, J.; CHEN, M. Economic impacts of climate change on agriculture: The importance of additional climatic variables other than temperature and precipitation. **Journal of Environmental Economics and Management**, v. 83, p. 8–31, 2017.

ANEXOS

38. Resumo da Apuração

Gleba	Cultivar	Data de Plantio	Plantio realizado conforme ZOAGRO?	Área (ha)	Produtividade (kg/ha)	Produção (kg)
		___/___/___	() Sim () Não			
		___/___/___	() Sim () Não			
		___/___/___	() Sim () Não			
		___/___/___	() Sim () Não			
		___/___/___	() Sim () Não			
		___/___/___	() Sim () Não			
		___/___/___	() Sim () Não			
		___/___/___	() Sim () Não			
		___/___/___	() Sim () Não			
		___/___/___	() Sim () Não			

39. Produção Total (Kg) - (Somatório de todas as produções informadas por gleba)

40. Observações (Comente sobre os aspectos relevantes da cultura)

a) Qual o número de propriedades existentes na proposta de seguro? Informe a área (ha) de cada uma delas.

b) Faça breves comentários sobre as condições da lavoura e evento causador do sinistro.

c) Comente sobre os aspectos fitossanitários, nutricionais, pragas e de infestação de plantas daninhas da lavoura.

d) Comentários adicionais

1ª Via - Seguradora 2ª Via - Segurado 3ª Via - Vistoriador
Assinatura do Segurado/Preposto: _____ Assinatura do Perito: _____

41. Croqui (Esboce um croqui da propriedade com roteiro de acesso e demonstre os pontos de amostragem)

a) A área sinistrada condiz com a área apresentada no croqui da proposta? () Sim () Não

b) Demonstrar divisão das glebas.

AMOSTRA	Latitude / Longitude	AMOSTRA	Latitude / Longitude	AMOSTRA	Latitude / Longitude
()	° ' "	()	° ' "	()	° ' "
()	° ' "	()	° ' "	()	° ' "
()	° ' "	()	° ' "	()	° ' "
()	° ' "	()	° ' "	()	° ' "

★ O segurado está ciente de que qualquer agravamento do sinistro deve ser comunicado a seguradora (Central de Atendimentos)? () Sim () Não

★ O segurado está ciente de que deve avisar a seguradora (Central de Atendimentos), com antecedência de 15 dias da data de colheita? () Sim () Não

42. O Perito autoriza nesta data a realização da colheita? () Sim () Não Se SIM, qual a área (ha) _____

43. O Segurado declara não possuir outro seguro para essa mesma propriedade? () Sim () Não

44. Nome do segurado ou preposto autorizado **45. Assinatura do segurado ou preposto** **46. CPF ou RG do segurado ou preposto**

47. Nome do perito **48. CREA Nº** **49. Assinatura do perito** **50. CPF ou RG do perito**

1ª Via - Seguradora 2ª Via - Segurado 3ª Via - Vistoriador
Assinatura do Segurado/Preposto: _____ Assinatura do Perito: _____

Anexo B Resultado do teste Mann-Whitney entre os dados de precipitação oriundos da plataforma Nasa Power, em relação a estações SIMEPAR

(continua)

Município	Ano	Estatística	N	Nasa	SIMEPAR	U	p-valor	Decisão
Apucarana	2015	Mediana	36	47,50	56,80	1264,0	0,5772	ns
Apucarana	2016	Mediana	36	45,12	47,10	1272,0	0,6402	ns
Apucarana	2017	Mediana	36	37,23	56,20	1274,0	0,6564	ns
Apucarana	2018	Mediana	13	25,63	50,20	173,0	0,9183	ns
Assis Chateaubriand	2015	Mediana	36	37,12	68,10	1145,0	0,0577	ns
Assis Chateaubriand	2016	Mediana	36	41,91	50,80	1270,0	0,6241	ns
Assis Chateaubriand	2017	Mediana	36	43,70	30,50	1422,0	0,2252	ns
Assis Chateaubriand	2018	Mediana	13	36,09	21,80	200,0	0,2168	ns
Campo Mourão	2015	Mediana	36	46,02	67,60	1238,0	0,3952	ns
Campo Mourão	2016	Mediana	36	38,98	37,80	1293,0	0,8174	ns
Campo Mourão	2017	Mediana	36	38,64	39,00	1336,0	0,8086	ns
Campo Mourão	2018	Mediana	15	32,66	43,60	223,0	0,7088	ns
Capitão Leônidas M.	2015	Mediana	36	29,47	36,40	1250,0	0,4745	ns
Capitão Leônidas M.	2016	Mediana	36	47,20	42,80	1382,0	0,4471	ns
Capitão Leônidas M.	2017	Mediana	36	44,27	42,10	1368,0	0,5468	ns
Capitão Leônidas M.	2018	Mediana	13	47,03	29,60	189,0	0,5049	ns
Cascavel	2015	Mediana	36	37,69	50,00	1255,5	0,5136	ns
Cascavel	2016	Mediana	36	43,22	51,60	1214,0	0,2624	ns
Cascavel	2017	Mediana	36	48,77	71,80	1210,0	0,2437	ns
Cascavel	2018	Mediana	15	44,13	46,60	219,0	0,5897	ns
Cianorte	2015	Mediana	36	39,93	56,10	1239,0	0,4014	ns
Cianorte	2016	Mediana	36	39,30	32,70	1313,5	1,00	ns
Cianorte	2017	Mediana	36	40,83	33,80	1404,0	0,3130	ns
Cianorte	2018	Mediana	15	29,21	36,20	229,0	0,9009	ns
Foz do Iguaçu	2015	Mediana	36	42,64	41,00	1315,0	0,9955	ns
Foz do Iguaçu	2016	Mediana	36	41,68	29,40	1365,0	0,5695	ns
Foz do Iguaçu	2017	Mediana	36	50,17	40,80	1371,0	0,5245	ns
Foz do Iguaçu	2018	Mediana	15	55,44	44,60	247,0	0,5614	ns
Guaíra	2015	Mediana	36	35,76	46,70	1280,0	0,7060	ns
Guaíra	2016	Mediana	36	39,62	42,30	1312,5	0,9910	ns
Guaíra	2017	Mediana	36	43,08	51,90	1269,0	0,6161	ns
Guaíra	2018	Mediana	13	33,28	32,20	192,0	0,4116	ns

(continuação)

Município	Ano	Estatística	N	Nasa	SIMEPAR	U	p-valor	Decisão
Londrina	2015	Mediana	36	42,06	40,70	1313,0	0,9955	ns
Londrina	2016	Mediana	36	39,70	33,60	1334,0	0,8261	ns
Londrina	2017	Mediana	36	38,69	34,40	1349,0	0,6975	ns
Londrina	2018	Mediana	15	28,02	24,40	242,0	0,7086	ns
Maringá	2015	Mediana	36	43,46	51,60	1257,0	0,5246	ns
Maringá	2016	Mediana	36	41,03	38,10	1342,0	0,7568	ns
Maringá	2017	Mediana	36	38,42	41,50	1325,5	0,9013	ns
Maringá	2018	Mediana	15	27,68	48,20	222,0	0,6782	ns
Palmital	2015	Mediana	36	37,64	43,60	1266,0	0,5927	ns
Palmital	2016	Mediana	36	40,17	60,60	1200,0	0,2012	ns
Palmital	2017	Mediana	36	41,58	45,70	1278,0	0,6893	ns
Palmital	2018	Mediana	15	45,87	45,80	232,0	1,00	ns
Palotina	2015	Mediana	36	35,66	39,70	1279,0	0,6976	ns
Palotina	2016	Mediana	36	41,91	27,70	1400,5	0,3328	ns
Palotina	2017	Mediana	36	43,7	38,20	1333,0	0,8348	ns
Palotina	2018	Mediana	13	36,09	39,60	182,0	0,7582	ns
Paranavaí	2015	Mediana	36	42,79	45,80	1262,5	0,5657	ns
Paranavaí	2016	Mediana	36	39,99	27,00	1390,5	0,3920	ns
Paranavaí	2017	Mediana	36	35,39	23,30	1374,0	0,5020	ns
Paranavaí	2018	Mediana	15	26,77	26,60	232,0	1,00	ns
Pato Branco	2015	Mediana	36	33,34	70,30	1120,0	0,0293	*
Pato Branco	2016	Mediana	36	52,45	49,80	1316,0	0,9865	ns
Pato Branco	2017	Mediana	36	48,36	42,80	1350,0	0,6892	ns
Pato Branco	2018	Mediana	15	36,11	14,00	261,0	0,2453	ns
Pinhão	2015	Mediana	36	38,25	55,60	1122,0	0,0310	*
Pinhão	2016	Mediana	36	43,45	46,40	1290,0	0,7913	ns
Pinhão	2017	Mediana	36	38,34	34,50	1351,0	0,6809	ns
Pinhão	2018	Mediana	15	41,51	24,20	253,0	0,4067	ns
Quedas do Guaçu	2015	Mediana	36	29,47	40,40	1208,0	0,2348	ns
Quedas do Guaçu	2016	Mediana	36	47,2	14,60	1480,0	0,0623	ns
Quedas do Guaçu	2017	Mediana	36	44,27	41,90	1348,0	0,4744	ns
Quedas do Guaçu	2018	Mediana	15	47,03	32,00	260,0	0,2622	ns
Santa Helena	2015	Mediana	36	36,63	52,70	1209,0	0,2392	ns
Santa Helena	2016	Mediana	36	41,5	33,30	1406,0	0,3017	ns
Santa Helena	2017	Mediana	36	46,51	34,70	1392,0	0,3819	ns
Santa Helena	2018	Mediana	15	42,88	39,00	244,0	0,6482	ns

(continuação)

Município	Ano	Estatística	N	Nasa	SIMEPAR	U	p-valor	Decisão
São Miguel Iguazu	2015	Mediana	36	36,2	27,50	1345,5	0,7270	ns
São Miguel Iguazu	2016	Mediana	36	46,35	28,90	1438,0	0,1642	ns
São Miguel Iguazu	2017	Mediana	36	45,57	38,30	1301,0	0,8880	ns
São Miguel Iguazu	2018	Mediana	15	49,79	42,60	246,0	0,5897	ns
Toledo	2015	Mediana	36	37,44	51,10	1221,5	0,3001	ns
Toledo	2016	Mediana	36	45,73	44,10	1286,0	0,7568	ns
Toledo	2017	Mediana	36	48,52	41,20	1337,5	0,7956	ns
Toledo	2018	Mediana	15	42,29	29,80	242,0	0,7089	ns
Umuarama	2015	Mediana	36	41,82	51,70	1241,5	0,4174	ns
Umuarama	2016	Mediana	36	43,11	32,40	1350,5	0,6851	ns
Umuarama	2017	Mediana	36	40,84	28,80	1370,0	0,5318	ns
Umuarama	2018	Mediana	15	29,17	61,00	213,0	0,4304	ns

Notas: **N:** número de amostras (decêndios); **U:** soma dos ranques da amostra; **p-valor:** valor significância calculado; **Decisão:** ns: Diferença não significativa; * significativa a 5% de probabilidade