MONALISA CRISTINA DE CÓL

WEATHER-BASED LOGISTIC REGRESSION MODELS FOR PREDICTING THE RISK OF WHEAT BLAST EPIDEMICS

Dissertation submitted to the Plant Pathology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

Adviser: Emerson Medeiros Del Ponte

VIÇOSA- MINAS GERAIS
2023

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T

C683w
2023

Cól, Monalisa Cristina de, 1999-

Weather-based logistic regressiom models for prediciting the risk of wheat blast epidemics / Monalisa Cristina de Cól. – Viçosa, MG, 2023.

1 dissertação eletrônica (46 f.): il. (algumas color.).

Texto em inglês.

Orientador: Emerson Medeiros Del Ponte.

Dissertação (mestrado) - Universidade Federal de Viçosa, Departamento de Fitopatologia, 2023.

Referências bibliográficas: f. 40-46.

DOI: https://doi.org/10.47328/ufvbbt.2023.529

Modo de acesso: World Wide Web.

1. Trigo - Doenças e pragas - Previsão. 2. *Magnaporthe oryzae*. 3. *Triticum aestivum*. I. Del Ponte, Emerson Medeiros, 1973-. II. Universidade Federal de Viçosa. Departamento de Fitopatologia. Programa de Pós-Graduação em Fitopatologia. III. Título.

CDD 22. ed. 633.1194

Bibliotecário(a) responsável: Alice Regina Pinto Pires CRB-6/2523

MONALISA CRISTINA DE CÓL

WEATHER-BASED LOGISTIC REGRESSION MODELS FOR PREDICTING THE RISK OF WHEAT BLAST EPIDEMICS

Dissertation submitted to the Plant Pathology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

APPROVED: July 27th, 2023.

Assent:

Documento assinado digitalmente
MONALISA CRISTINA DE COL
Data: 01/09/2023 10:59:15-0300
Verifique em https://validar.iti.gov.br

Monalisa Cristina de Cól
Author

Documento assinado digitalmente
EMERSON MEDEIROS DEL PONTE
Data: 01/09/2023 11:29:46-0300
Verifique em https://validar.iti.gov.br

Emerson Medeiros Del Ponte
Adviser

ACKNOWLEDGEMENTS

To God.

To my parents, Lisane and Ademir, for all the support in my decisions, for the encouragement, and for always being there when I needed.

To my partner, Ricardo Augusto Neves Forner, I am grateful for his company, for listening to me, for helping me gain perspective, and for believing in myself.

To my Advisor, Prof. Emerson Medeiros Del Ponte, for his invaluable guidance, support, patience, encouragement, unwavering commitment and teachings.

To the Federal University of Viçosa, for the opportunity to complete the graduate course.

To the Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES), for granting the scholarship.

To everyone who directly or indirectly contributed to my personal and professional development. Thank you very much!

ABSTRACT

DE CÓL, Monalisa, M.Sc., Universidade Federal de Viçosa, July, 2023. **Weather-based logistic regression models for predicting the risk of wheat blast epidemics**. Adviser: Emerson Medeiros Del Ponte.

Wheat blast, caused by *Pyricularia oryzae* Triticum lineage, is an important yield-limiting disease in the tropics of Brazil. This study aimed to develop models for predicting the within-season risk of wheat blast outbreaks. Data sets used in this study were obtained from field trials conducted in Patos de Minas (n = 103 epidemics) as well as in 10 other locations (n = 40 epidemics); the latter as part of the cooperative fungicide trial network. The trials were conducted across six states of Brazil over a nine-year period (2012-2020). A binary response variable was created based on disease incidence being ≥20% or <20%. Daily meteorological variables including minimum (Tmin), maximum (Tmax), and mean temperature (Tmean), relative humidity (RH) and precipitation (P) were obtained from the NASA POWER. The wheat heading date (WHD) was used to define four time windows, consisting of two intervals of seven days each, before and after the WHD. These windows combined with the weather variables resulted in 36 prediction variables (9 weather variables × 4 windows). Logistic regression models were fitted with selection of variables using LASSO followed by best subset selection. Four accuracy measures were computed and the model performance was evaluated using leave-one-out cross validation (LOOCV). The models were further run for six sites using a historical series of 23 years of weather data. The variables that best explained outbreak risk were RH, days with Tmean < 22ºC and Tmean x RH two weeks before WHD, Tmean x RH one week before WHD, and RH and Psum one week after WHD. The selected models included 2 to 5 predictors, with accuracies ranging from 0.80 to 0.85. Sensitivities ranged from 0.80 to 0.91, specificities from 0.72 to 0.86, and AUC values from 0.89 to 0.91. The accuracy values obtained for LOOCV ranged from 0.77 to 0.81. The model predictions generally agreed with most historical reports in the tropical region of Brazil. This study enhanced our understanding of the complex relationship between weather variables and wheat blast, contributing valuable insights for disease management.

Keywords: *Magnaporthe oryzae*. *Triticum aestivum*. Prediction models.

RESUMO

DE CÓL, Monalisa, M.Sc., Universidade Federal de Viçosa, julho de 2023. **Modelos de regressão logística baseados em dados meteorológicos para prever o risco de epidemias de brusone no trigo**. Orientador: Emerson Medeiros Del Ponte.

A brusone, causada por *Pyricularia oryzae* linhagem Triticum, é uma doença importante que limita a produção nas regiões tropicais do Brasil. Este estudo teve como objetivo desenvolver modelos para prever o risco de epidemias de brusone em trigo. Os dados foram obtidos de ensaios de campo conduzidos em Patos de Minas (n = 103 epidemias), e em outros 10 locais (n = 40); sendo este último como parte da rede de ensaios com fungicidas. Os ensaios foram conduzidos em seis estados do Brasil em um período de nove anos (2012-2020). Uma variável resposta binária foi criada com base na incidência da doença ser ≥20% ou <20%. Variáveis climáticas diárias, incluindo temperatura mínima (Tmin), temperatura máxima (Tmax), temperatura média (Tmean), umidade relativa (RH) e precipitação (P), foram obtidas no NASA POWER. A data de espigamento do trigo (DET) foi usada para definir quatro períodos, compostas por dois intervalos de sete dias cada, antes e depois do DET. Esses períodos, combinados com as variáveis do clima, resultaram em 36 preditores (9 variáveis climáticas × 4 períodos). Modelos de regressão logística foram ajustados com seleção de variáveis utilizando regressão LASSO seguida de seleção do melhor subconjunto. Quatro medidas de acurácia foram calculadas e o desempenho do modelo foi avaliado usando validação cruzada leave-one-out (LOOCV). Os modelos foram executados em seis locais para uma série histórica climática de 23 anos. As variáveis que melhor explicaram o risco de surto epidêmico foram RH, dias com Tmean < 22ºC e Tmean x RH duas semanas antes da DET, Tmean x RH uma semana antes da DET, e RH e soma de P uma semana após a DET. Os modelos selecionados incluíram de 2 a 5 preditores, apresentando acurácias de 0,80 a 0,85. As sensibilidades variaram de 0,80 a 0,91, as especificidades de 0,72 a 0,86 e os valores de AUC de 0,89 a 0,91. Os valores de acurácia da LOOCV variaram de 0,77 a 0,81. As predições do modelo geralmente estavam de acordo com os relatos históricos na região tropical. Este estudo aprimorou nossa compreensão da complexa relação entre variáveis meteorológicas e brusone do trigo, fornecendo informações valiosas para o manejo da doença.

Palavras-chave: *Magnaporthe oryzae*. *Triticum aestivum*. Modelos de predição.

SUMMARY

1 INTRODUCTION

In Brazil, wheat is mainly grown in the southern, subtropical, regions of the country, but its cultivation has considerably expanded beyond the subtropics and reached the midwestern region of the country. In fact, the Brazilian Cerrado is recognized as the agricultural frontier for the expansion of wheat cultivation (Pasinato et al., 2018). Nevertheless, the endemic presence of a fungal disease, wheat blast, has limited the expansion of the crop and the attainment of enhanced wheat yields in this particular region (Pasinato et al., 2018; Torres et al., 2022).

Wheat blast is caused by the ascomycetous fungus *Pyricularia oryzae* Triticum lineage (PoT) (synonym *Magnaporthe oryzae*) (Couch and Kohn, 2002). The fungus thrives at temperatures ranging between 25°C and 30°C, and relative humidity exceeding 90% (Alves et al., 2006; Cardoso et al., 2008; Mills et al., 2020). In such conditions, and if control measures (e.g. fungicide sprays) are not utilized, the disease can severely reduce crop yield by more than 80% (Coelho et al., 2016; Dos Santos et al., 2022a).

Disease models that rely on weather variables have been developed for predicting wheat blast infection risk (Cardoso et al., 2008; Cruz et al., 2016a; Fernandes et al., 2017; Fernandes et al., 2021). Such models have been useful for basing decisions on fungicide application, identifying areas at risk by utilizing historical data, and conducting risk analysis in regions where the pathogen has not yet been introduced (Fernandes et al., 2021; Cruz et al., 2016a; Mottaleb et al., 2018).

Mechanistic (concept-driven) or empirical (data-driven) modeling approaches can be used to develop prediction models for plant diseases (De Wolf et al., 2003; Fernandes et al., 2017; Cruz et al., 2016a; Shah et al., 2019; Dalla Lana et al., 2021). Both approaches have been used for modeling wheat blast. The empirically-derived model was developed based on observations of the disease (infection frequency) in a controlled environment (Cardoso et al., 2008). Mechanistic models combine different components (rules or equations) to predict key stages of the epidemic process such as inoculum build-up and favorability for infection, based on weather variables (Fernandes et al., 2021).

Data on wheat blast in Brazil have been collected since 2012 in coordinated fungicide trials across Brazil, as well as in Patos de Minas, Minas Gerais, in experimental plots to test the effect of sequential planting dates on disease intensity. These datasets of field observations provided us with an opportunity, for the first time, to empirically derive prediction models for wheat blast using regression modeling. Therefore, the objectives of this study were to (i) identify time windows associated with weather variables that best explain the occurrence of epidemic outbreaks; (ii) build weather-based prediction models for predicting the occurrence of wheat blast; (iii) evaluate model performance by running it retrospectively for a series of past years-heading dates in a few selected locations in Brazil.

2 LITERATURE REVIEW

2.1 Origin and spread of wheat blast

Wheat blast is caused by the fungus *Pyricularia oryzae* Triticum lineage (PoT) (synonym to *Magnaporthe oryzae*) (Couch and Kohn, 2002). The first epidemic outbreaks of wheat blast occurred in 1985 during the season, whereby various municipalities in the Brazilian state of Paraná were impacted (Igarashi et al., 1986). In the following years, wheat blast has been documented in several other Brazilian states and neighboring countries (Goulart et al., 1990; Picinini and Fernandes, 1990; Prabhu et al., 1992; Dos Anjos et al., 1996; Barea and Toledo, 1996; Viedma and Morel, 2002; Cabrera and Gutierrez, 2007; Ceresini et al., 2018).

Outside Brazil, wheat blast outbreaks resulting in significant crop yield losses were documented in Bolivia during 1996 and 1997 (Barea and Toledo, 1996; Kohli et al., 2011), in Paraguay in 2002 and 2005 (Viedma, 2005; Viedma et al., 2010), and more recently in Bangladesh in 2016 (Malaker et al., 2016) and Zambia in 2018 (Tembo et al., 2020).

2.2 Disease cycle and epidemiology

Wheat blast can manifest during both the vegetative and reproductive stages of the plant (Cruz et al., 2015; Martínez et al., 2018). In relation to the infection cycle, there is empirical evidence suggesting that wheat blast is a polycyclic disease (Gongora-Canul et al., 2020, Mills et al., 2020). The PoT fungus has the potential to infect multiple types of cereal crops, forage and grasses (Tosa and Chuma, 2014), which can act as alternative hosts and have a significant impact on the disease epidemiology. These plants can serve as a source of inoculum and as reservoirs for the pathogen's long-term survival (Fernandes et al., 2021). Although PoT has saprophytic activity, its survival is greatly affected by weather conditions. Therefore, the presence of alternative hosts becomes even more crucial in ensuring the pathogen's survival and spread (Ceresini et al., 2018; Pizolotto et al., 2019).

The spread of the wheat blast disease is attributed to both air-borne and seed-borne dispersal. While the former is responsible for local dissemination between fields, the latter facilitates long-distance dissemination, thereby explaining the rapid spread observed between different states and countries (Gomes et al., 2018; Singh et al., 2021).

During the infection of the plant, the fungal conidia adhere to the leaf surface and subsequently form germ tubes and appressoria, leading to colonization of the leaf tissue (Cruz et al., 2016b). Cruz et al. (2016b) conducted a study that demonstrated that the initial symptoms of the disease were evident 40 hours post-inoculation, while conidia were visible on the leaf surface after 120 hours, which is critical to the secondary cycle of the pathogen. Elevated production of conidia may be favored by 40 to 45% relative humidity and temperatures ranging between 24°C and 28°C (Cardoso et al., 2008; Kovaleski et al., 2020; Roy et al., 2020). For infection, a minimum wetting-period of 10 hours is necessary. The optimal conditions for infection are temperatures between 25 and 30ºC and a longer wetting-period ranging from 25 to 40 hours, which may result in disease intensity of more than 80% (Cardoso et al., 2008).

2.3 Symptoms and economical significance

Symptoms of wheat blast have been observed in all aerial plant parts, although the incidence normally has been reported on leaves and spikes (Cruz and Valent, 2017; Tembo et al., 2020; Gongora-Canul et al., 2020). According to previous studies, wheat blast foliar symptoms can be detected in the senescent leaves, particularly in the three basal leaves of wheat (Cruz et al., 2015). The expression of foliar symptoms may vary depending on the cultivar and strain. Necrotic lesions with a grayish center and brown margin, chlorotic lesions, necrotic diamond-shaped lesions, greenish necrotic lesions, and elongated brown lesions have been reported (Perelló et al., 2017). In addition to leaf symptoms, the presence of sporulation of these lesions was seen (Cruz et al., 2015), which might have an impact on the incidence and severity spike in the heading stage of wheat.

According to Perelló et al. (2017), symptoms of wheat blast on the spike can include ellipsoidal necrotic lesions with a brown edge and lighter center on the glumes, a graying rachis, and total or partial spike blanching. The author also reported that the affected spikes exhibited smaller and shriveled grains. The economic significance of a disease's impact on grain characteristics is evident, demonstrating the negative effects of wheat blast on 1000-grain weight and grain quality, thereby affecting crop yield (Martínez et al., 2018). In Brazil, Minas Gerais state, losses ranging from 62.8 to 80.1

% were reported (Coelho et al., 2016). Other experiments in Brazil have shown a range of incidence between 0% and 100% with high losses being reported (Dianese et al., 2021; Dos Santos et al., 2022a). Notably, when the incidence (percent of heads diseased) or severity (percent of the wheat head area affected) reached 25%, relative losses in the yield were estimated at 39.8% and 52.8% respectively, according to a recent modeling study (Dos Santos et al. 2022a).

2.4 Wheat blast management

Several measures can be used for combating wheat blast. These include crop rotation, use of certified pathogen-free seeds, fungicide seed treatment, delayed planting date, use of resistant cultivars and chemical control (Ceresini et al., 2018). In the Cerrado region of Brazil, delay in planting date (shifting from February-March to March-April) has shown to be effective to reduce the incidence of wheat blast by avoiding the more conducive environmental conditions during wheat heading and flowering. However, there is a risk of shortage of water for wheat growth which requires that late-planted fields are irrigated (Coelho et al., 2016).

Notwithstanding the development of genetically resistant cultivars, chemical control remains necessary for effective management of wheat blast, as the genetic resistance is frequently overcome by the emergence of new pathogen races rewrite better (Ceresini et al., 2018). In Brazil, multiple chemical products have undergone field testing for disease control, and the results suggest that two to three applications of fungicides may be necessary, being the first application during the heading stage (Santana et al., 2013, 2021; Cruz et al., 2019). The effectiveness of fungicides utilized in Brazil for managing wheat blast varies from 24.2% to 66.5%, depending on the geographic region and type of fungicide used. Furthermore, a 9-year analysis of fungicide efficacy revealed no decrease in efficacy over time (Ascari et al., 2021).

2.5 Overview of predictive models

Being able to predict the timing of disease outbreaks is an important application of plant disease epidemiology (Campbell and Madden, 1990). Mechanistic or empirical modeling approaches have frequently been employed to develop disease prediction models (De Wolf et al., 2003; Cruz et al., 2016a; Fernandes et al., 2017; Shah et al., 2019; Dalla Lana et al., 2021; González-Domínguez et al., 2023). Mechanistic models utilize submodels to describe the various substages of the disease cycle, often incorporating results from other researchers' studies on the pathogen biology. Empirical models frequently employ field experiment data to elucidate the correlation between weather patterns and disease intensity. Such models can be utilized to make inferences about the underlying biology of a system (De Wolf and Isard, 2007).

The disease development over space and time is reliant upon the interplay among the host plant, the pathogen, and the environment (Madden et al., 2007). As a result, weather models play a critical role in disease forecasting systems, as they enable the examination of disease dynamics over both spatial and temporal scales. The interaction between plant-pathogen is greatly affected by weather factors, particularly the interaction between temperature and moisture (González-Domínguez et al., 2023).

2.6 Wheat blast models

The pioneering study conducted by Cardoso et al. (2008) investigated the correlation between temperature, spike wetting-period, and wheat blast intensity. Their findings provided fundamental insights for developing a weather-based model for predicting probability of infection. However, it should be noted that the experiment was performed in a greenhouse setting. In their respective studies, Cruz et al. (2016a) and Fernandes et al. (2017) developed process-based models that accounted for the various stages of the epidemic process, including inoculum buildup, infection, survival, and dissemination, in order to investigate their impact on the occurrence of epidemics. These models utilized specific climatic variables as predictors for some stages of the disease cycle and demonstrated that these variables could be utilized for forecasting the occurrence of epidemics. However, due to limitations in the availability of disease data in the field and in various wheat-growing localities, as well as the need for further analysis of climate data, the models remain subject to challenges in terms of their testability. Fernandes et al. (2017) emphasized the importance of utilizing remotely estimated or simulated weather data for expanding spatial boundaries, refining prediction scales, and streamlining model development.

The dependence of wheat blast epidemics on seasonal weather patterns highlights the usefulness of prediction models in forecasting outbreak risks early in the season, making them a crucial tool for disease control (Kim and Choi, 2020; Fernandes et al., 2021). These models serve as the foundation for the development of warning systems that enable farmers to make informed decisions about fungicide application (Fernandes et al., 2021). Additionally, prediction models have the ability to identify areas at risk by utilizing historical data, and perform risk analyses in regions where a pathogen has not yet been introduced (Cruz et al., 2016a; Mottaleb et al., 2018).

3 MATERIAL AND METHODS

3.1 Data sources and disease variables

Data on wheat blast intensity, as well as crop heading dates used in our study were obtained from two sources where the data were made publicly available. The first source pertains to a cooperative fungicide trial network, and Ascari et al. (2021) studied the effect of fungicides on disease control at various sites in Brazil between 2012 and 2020. In their investigation, a susceptible variety (data not shown) specifically adapted to the respective site was utilized. The second source pertains to EPAMIG, and the study publicly examined the impact of wheat blast on wheat yields at a single location (Patos de Minas, MG, Brazil) from 2013 to 2019 (Dos Santos et al., 2022a). The experiments of the second study were originally designed to investigate the effect of planting date on disease intensity. Three Brazilian spring wheat cultivars susceptible to wheat blast, namely BRS 264, BRS 404, and MGS Brilhante, were utilized in the study. Each case in this study considered the average incidence value across the cultivars, given their similarity in susceptibility to wheat blast.

A total of 143 epidemics, considering the combination of location, year, and planting date, were conducted across eleven locations in six states of Brazil over a nine-year period (Table 1). In all experiments, the disease occurred naturally.

Table 1. Information of data sets of wheat blast disease cases conducted across different regions of Brazil used in this study.

State	Location	Climate$^{(a)}$	N. of years	Years	N. of cases	N. of outbreaks (Incidence>20%)
DF	Planaltina	Aw	8	2012-2014, 2016-2020	8	6
MG	Patos de Minas	Aw	7	2013-2019	103	50
MG	Uberaba	Aw	1	2015	1	1
MS	Dourados	Cwa	2	2012, 2013	2	2
MT	Campo Verde	Aw	3	2015, 2016, 2018	3	2

PR	Campo Mourão	Cfa	3	2017, 2018, 2020	5	0
PR	Guarapuava	Cfb	1	2019	1	1
PR	Londrina	Cfa	4	2012, 2014-2016	6	5
PR	Palmeira	Cfb	7	2014-2020	8	6
PR	Palotina	Cfa	2	2012, 2014	2	2
SP	Itaberá	Cfa	4	2012-2015	4	3

[a] According to Koppen (1918).

The incidence of wheat blast was quantified as the percentage of diseased heads in a sample of 100 heads obtained from the central row of each plot. Meanwhile, the conditional severity of the disease was measured as the percentage of the symptomatic area within each spike exhibiting symptoms. Both assessments were carried out visually by the same rater within the same location between 25 and 30 days after flowering (phenological stage 85 from the scale of Zadoks) (Zadoks et al., 1974). According to Dos Santos et al. (2022b), there exists a positive correlation between the severity and incidence (conditional severity x incidence) of wheat blast, meaning that severity can be predicted from incidence. According to the same author, when the incidence exceeds 20%, significant productivity losses are already observed. Because incidence can be estimated with greater precision than severity (Bock et al., 2022), this variable was used in this study to define the occurrence of an outbreak as a binary outcome - either outbreak if incidence equal or greater than a threshold 20% or non-outbreak if incidence was below.

3.2 General modeling approach

The study was conducted in a stepwise manner, involving three main steps: i) identification of the most significant variables that explained the disease occurrence, ii) selection of the logistic regression models, and iii) model evaluation. A diagram was used to depict the flow of these steps (Figure 1).

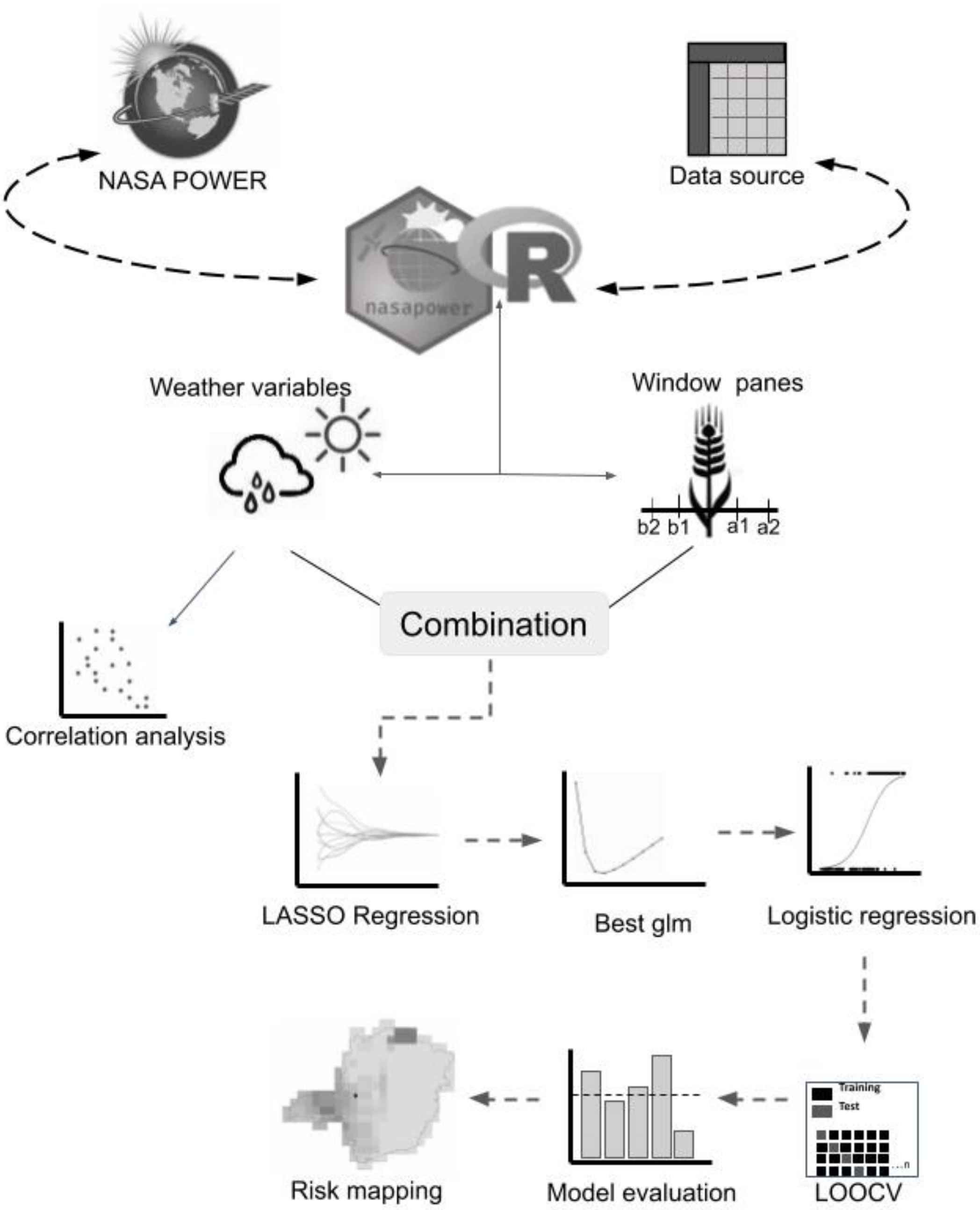

Figure 1. Diagram illustrating the stepwise approach used to develop and evaluate an empirically-derived prediction model for wheat blast.

3.3 Source of weather data and predictor variables

Daily meteorological variables were obtained from reanalysis data provided by the NASA Prediction Of Worldwide Energy Resources (POWER) project (https://power.larc.nasa.gov/) using the nasapower R package (Sparks et al., 2023). For each case, the package was used to retrieve the following meteorological variables based on their respective geographical coordinates (latitude and longitude): daily minimum, maximum, and mean temperature, as well as relative humidity and daily precipitation. These variables were obtained for each case.

The wheat heading date was used to define the time windows. The creation of time windows is usually an initial step in the development of empirically-derived weather-based models (Dalla Lana et al., 2021). In our case, two time windows were defined consisting of a seven-day size before (wb1: first week before heading and wb2: second week before) and two windows after the heading date (wa1: first week after heading and wa2: second week after), resulting in four windows. Thereafter, the windows were combined with the weather-based variables and each combination became a predictor variable to be considered in the analysis. Then, within each time window the number of days with mean temperature lower than 22ºC (cooler days), days with relative humidity greater than 90% (wet days) (Mills et al., 2020) and days with rainfall occurrence (wet days) was summarized. Additionally, the window-average of the minimum, maximum and mean temperature, precipitation and relative humidity was summarized (Table 2). After summarizing the weather variables, the interaction between weather variables was also analyzed and defined as new variables.

A total of 56 prediction variables (14 weather variables x 4 windows) were created. A correlation matrix (Pearson's correlation) using all weather variables to estimate the correlation coefficient (r) was constructed. To avoid extraordinarily large estimated standard errors and large estimated coefficient in the logistic model caused by multicollinearity (Hosmer et al. 2013), those weather variables with r ⩾ 0.95 were removed.

Table 2. Acronym and definition of variables used for developing logistic models to predict wheat blast outbreaks based on weather data.

Variable acronym	**Definition**
RH	Average daily RH (%)
RH90	Duration (days) RH ≥ 90%
TmeanMax	Average daily maximum T (ºC)
TmeanMin	Average daily minimum T (ºC)
Tmean	Average daily mean T (ºC)
ND.Tmean.S22	Duration (days) TMEAN < 22ºC

Rainsum	Sum of precipitation (mm)
Rain.ND.G0	Duration (days) of precipitation greater than zero
TminRH	TMEANMIN x RH
TminRainsum	TMEANMIN x RAINSUM
TminRH90	TMEANMIN x RH90
TmeanRH	TMEAN x RH
TmeanRainsum	TMEAN x RAINSUM
TmeanRH90	TMEAN x RH90

3.4 Selection of predictor variables for the model

The predictor variables with the highest relevance for model prediction were selected utilizing the Least Absolute Shrinkage and Selection Operator (LASSO) method. Lasso regression is a statistical method that shrinks the coefficient estimates towards zero, forcing some coefficient estimates to be exactly equal to zero when the tuning parameter λ is sufficiently large (James et al., 2013). Variables with coefficients of zero are eliminated minimizing the prediction error (Ranstam and Cook, 2018). To determine the optimal λ that results in the lowest cross-validated error, a K-fold cross-validation procedure was performed. This optimization was conducted utilizing the cv.glmnet function in the *glmnet* package (Friedman et al., 2010). As a result of LASSO regression, the number of predictor variables included and evaluated in the best subset selection is reduced by eliminating variables.

The predictor variables selected by LASSO were applied in the *bestglm* function of the *bestglm* package (McLeod et al., 2020) which uses the best subset selection method and provides better combinations of variables to apply in logistic regression models. The best subset selection method to select predictors variables considers all possible subsets of p predictors (2^p) and chooses the best model from all models that contain only some of the predictors. The process begins with model 0, which is the null model that includes no predictors, only the intercept and the sample mean for each observation. Then, model 1 is created, which is the best model with one predictor among all possible models with one predictor. The same process is repeated until reaching the maximum number p of predictors (James et al., 2013). This means that the number of models is equivalent to the number of predictors, ranging from 0 to p.

Among these models the Log-likelihood and Bayesian information criterion (BIC) were used to select the model with the best fit and to adjust the training error, respectively. Models with bigger Log-likelihood and lower BIC are preferable. Overall, using LASSO for variable selection and the Best subset selection to select the best combinations of variables based on Log-likelihood and BIC can help identify the most important predictors and construct a parsimonious and accurate logistic regression model for predicting the outcome variable (James et al., 2013).

3.5 Logistic regression models

The combinations of predictor variables selected by the best subset were applied in the logistic regression model. The logistic regression model can be written as:

$$ln\left(\frac{P(y\ =\ 1)}{1-\ P(y\ =\ 1)}\right)=\beta 0+\beta 1x1+\beta 2x2+\ldots+\beta pxp$$

in which β0 is the intercept, β1, β2, ..., βp are the coefficients associated with the predictor variables x1, x2, ..., xp, respectively. x1, x2, ..., xp are the predictor variables (weather variables). $P(y=1)$ is the probability of an epidemic outbreak occurring.

Each logistic regression model was evaluated to check for multicollinearity by using the variance inflation factor (vif) of *vif* function from the car package. The outbreak threshold was determined by calculating the sensitivity and specificity for each model and choosing the one that has the largest summation. This was analyzed using the function error.threshold.plot of package PresenceAbsence (Freeman and Moisen, 2008). Furthermore, the performance statistics, including accuracy, sensitivity, specificity, and Area Under the ROC Curve (AUC), were evaluated using the *presence.absence.accuracy* function. Lastly, the different logistic regression models were compared based on interpretability and accuracy measurements.

The leave-one-out cross-validation (LOOCV) technique is a robust method for assessing the predictive accuracy of a statistical model. This technique is particularly suitable for small datasets because it maximizes the use of available data (Hastie et al., 2009). In the context of this study, LOOCV was applied to assess the performance of a logistic regression model that was built to predict the occurrence of outbreaks. In the LOOCV process, the dataset was split into a training set and a testing set. This split was performed such that one single observation (or "epidemic") was used for testing, while all remaining observations (in this case, 142 epidemics) were used for model fitting or training. Once the logistic regression model was trained on the 142 epidemics, it was then applied to the single epidemic left out (the test set). The model's prediction for this epidemic was compared to the actual outcome to determine if it was correct. The accuracy of the model in each iteration was quantified as the proportion of correctly classified samples. If the model correctly predicted the outcome for the left-out epidemic, the accuracy for that iteration was 1; otherwise, it was 0. Since accuracy alone can sometimes be misleading (particularly when the classes are imbalanced), the misclassification error rate was also calculated. This was done by subtracting the accuracy from 1. So, if the model accurately predicted the outcome for the left-out epidemic, the misclassification error for that iteration was 0; otherwise, it was 1. After running the LOOCV process across all possible iterations (each time leaving out a different epidemic for testing), the overall accuracy of the model was calculated. This was done by taking the mean of the accuracies obtained from all iterations of the LOOCV, as described by James et al. (2013). This provided a single measure of the logistic regression model's performance in correctly predicting epidemics, taking into account all available data and considering each observation in turn as a test case.

3.6 Model comparison and evaluation

To further compare and evaluate the models, six distinct locations in Brazil with known absence/presence of wheat blast outbreaks were carefully selected. The chosen sites for evaluation included Uberaba, Madre de Deus, and Patos de Minas in Minas Gerais, Planaltina in Distrito Federal, Londrina in Paraná, and Passo Fundo in Rio Grande do Sul.

Weather data spanning from 2000 to 2022 was gathered for each location using NASA POWER. The specific 30-day periods considered for each location, corresponding to usual period of the heading date in the region, were as follows: April and May for the sites in Minas Gerais and Distrito Federal, mid-June to mid-July for Londrina, and September for Passo Fundo.

The frequency of outbreak occurrence for each model was calculated retrospectively for each location, considering 30 heading dates and spanning a 23-year period. The frequency was calculated based on the probability of an outbreak occurrence, taking into account the specific threshold set for each model. The probability of an outbreak occurrence, estimated using logistic regression models, can be represented as:

$$p = \frac{Exp\ (\beta 0+\beta 1x1+\beta 2x2+ \ldots+\beta pxp)}{(1+ exp(\beta 0+\beta 1x1+\beta 2x2+ \ldots+\beta pxp))},$$

where p represents the probability, β_0, β_1, β_2, ..., β_p are the coefficients of the predictor variables x_1, x_2, ..., x_p, and *exp* denotes the exponential function. The logistic function in the numerator transforms the linear combination of predictor variables into a probability value, while the denominator normalizes the probability value between 0 and 1.

This analysis was useful in evaluating and comparing the similarity of model performance and selecting a representative more realistic model among the majority, while also considering the model with the fewest predictor variables. The chosen single model was employed to evaluate the seasonal risk of wheat blast for each year within the site. A qualitative evaluation was implemented, which involved the consultation of relevant scientific literature and engagement with domain experts on each region.

Among the 23 years, one year with a high outbreak frequency, one year with frequency close to 0.5 and one year with a low outbreak frequency were selected for each site, and probability of outbreak occurrence for each heading date was assessed. Furthermore, a risk map of the state of Minas Gerais was generated to assess the probability of an outbreak of wheat blast during two potential heading periods within the months of April and May in one outbreak year and one non-outbreak year.

4 RESULTS

4.1 Data and weather variables

Out of the 143 cases, 65 were classified as non-outbreak and 78 were classified as outbreak, based on the criterion that a case was considered as outbreak when the incidence exceeded 20%. The variables TminRH, TminRainsum, TminRH90, TmeanRainsum, TmeanRH90 exhibited a strong correlation ($r \geq 0.95$) with other weather variables. Consequently, these variables were excluded from the analysis, further reducing the variables to nine. Consequently, a total of 36 predictor variables were utilized (4 windows x 9 predictor variables).

4.2 Selection of predictor variables for the model

Among the initial set of 36 predictor variables, the LASSO method identified nine variables whose coefficients remained non-zero, indicating their higher significance (Figure 2). Importantly, at least one weather variable was selected in each time window, with six variables selected in the pre-heading windows and three variables selected in the post-heading windows. The selected variables were: RH_wa1, RH_wb1, RH_wb2, ND.Tmean.S22_wb2, Rainsum_wa1, Rain.ND.G0_wb1, TmeanRH_wa2, TmeanRH_wb1 and TmeanRH_wb2. The RH variable was consistently selected across all windows when considering the TmeanRH variable. The TmeanRH was only not selected for the window the first week after the heading date. In contrast, the variables ND.Tmean.S22_wb2, Rainsum_wa1, and Rain.ND.G0_wb1 were specifically selected in individual windows.

The selection by best subset generated the top ten combinations of variables based on both Log-likelihood and the Bayesian Information Criterion (BIC). The method took into consideration nine variables, leading to combinations ranging from zero to nine predictor variables (Figure 3).

These particular criteria - Log-likelihood and BIC - are widely used for model selection as they provide a balanced evaluation of the model's goodness of fit and its complexity. The Log-likelihood metric, for instance, measures the extent to which a model is capable of predicting the observed data; a higher value of this metric is indicative of a better fit. In contrast, the BIC also incorporates the model's complexity into its evaluation, with lower BIC values being more desirable as they suggest a more parsimonious model (James et al., 2013). Upon analysis of the models, we observed a trend where an increase in the number of variables led to a rise in the Log-likelihood values. This trend continued until the model reached a saturation point where the addition of further variables did not significantly enhance the model's fit (as demonstrated in Figure 3A). This stabilization became evident upon the inclusion of five predictor variables, where the Log-likelihood values consistently hovered around -55.5. This suggests that beyond five predictor variables, the improvement in model fit may not be substantial, highlighting the potential to achieve a balance between model accuracy and complexity.

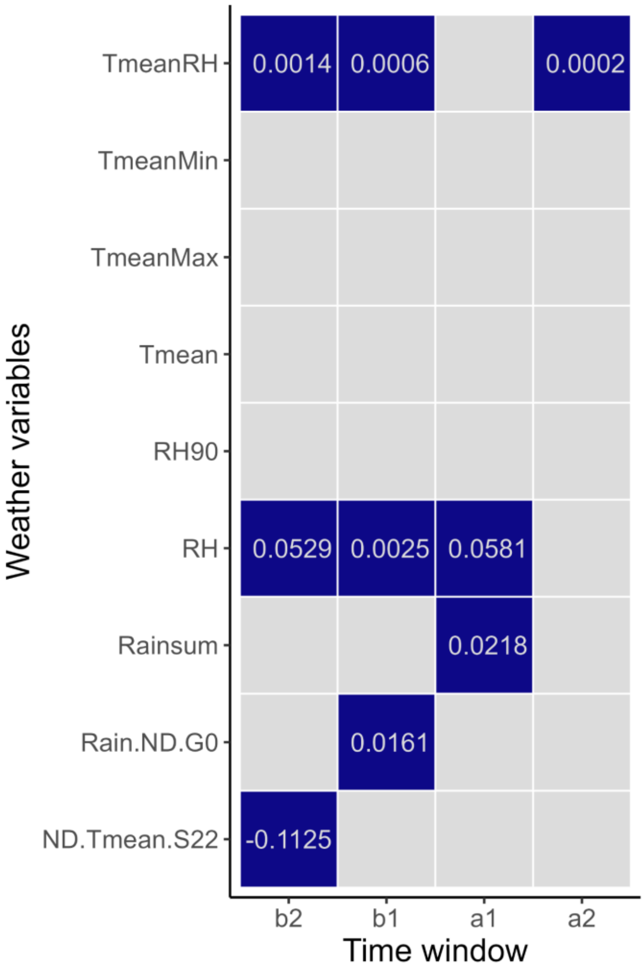

Figure 2. Selected (blue color) weather variables by the LASSO regression method in each time window. Predictor weather variables are defined in Table 2. b1, first week before the heading date; b2, second week before the heading date; a1, first week after the heading date; a2, second week after the heading date.

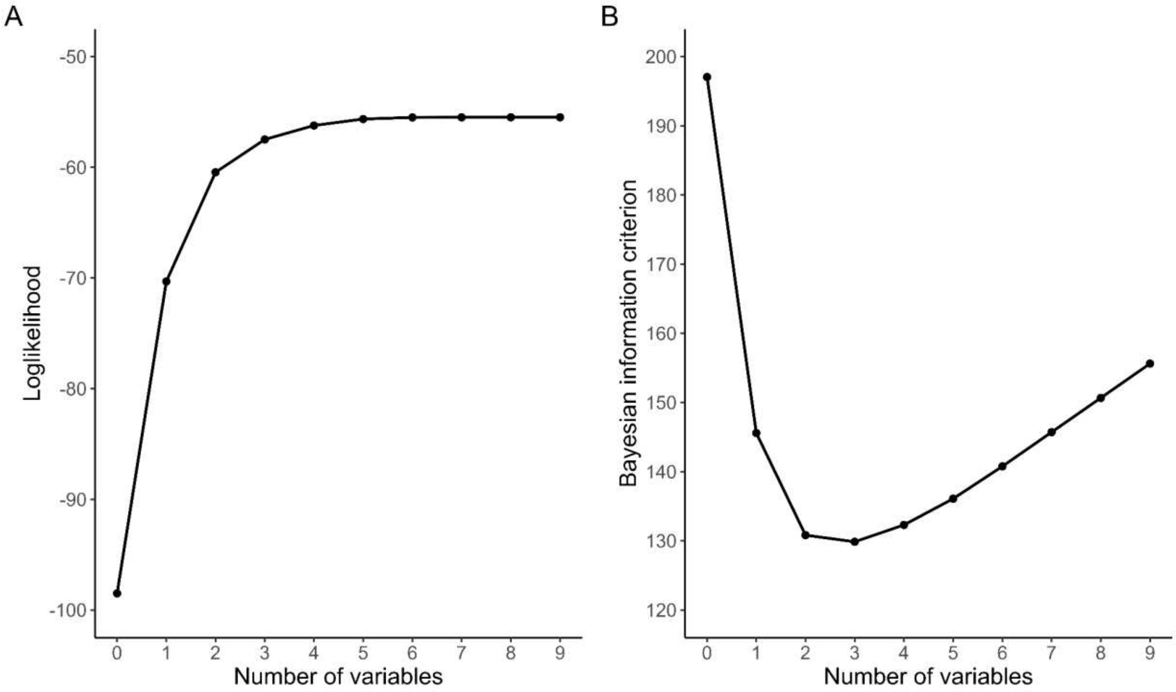

Figure 3. Relationship between the number of variables and model evaluation metrics: A comprehensive analysis of Log-likelihood (A) and Bayesian Information Criterion (B)

Examining the Bayesian Information Criterion (BIC), we found the optimal value when using three variables. However, it's important to note that there was a negligible difference when compared to the BIC value using two variables, as illustrated in Figure 3B. Interestingly, we observed an increase in BIC when more than three predictor variables were used. Despite this upward trend, using up to six variables still yielded better BIC values than when employing just a single variable. Taking into consideration both statistical metrics – the Log-likelihood and BIC – we narrowed our selection down to four distinct combinations of variables. These combinations ranged from two to five variables and were selected for application in logistic regression models. As a result of this selection process, three variables – TmeanRH_wa2, RH_wb1, and Rain.ND.G0_wb1 – were excluded from the initial pool of variables selected by the LASSO method. Upon evaluation by best subset, the combination containing three predictor variables was deemed the most effective. Figure 4 visually represents the variables included in the optimal combinations selected by best subset method, exhibiting their behavior in both outbreak and non-outbreak contexts.

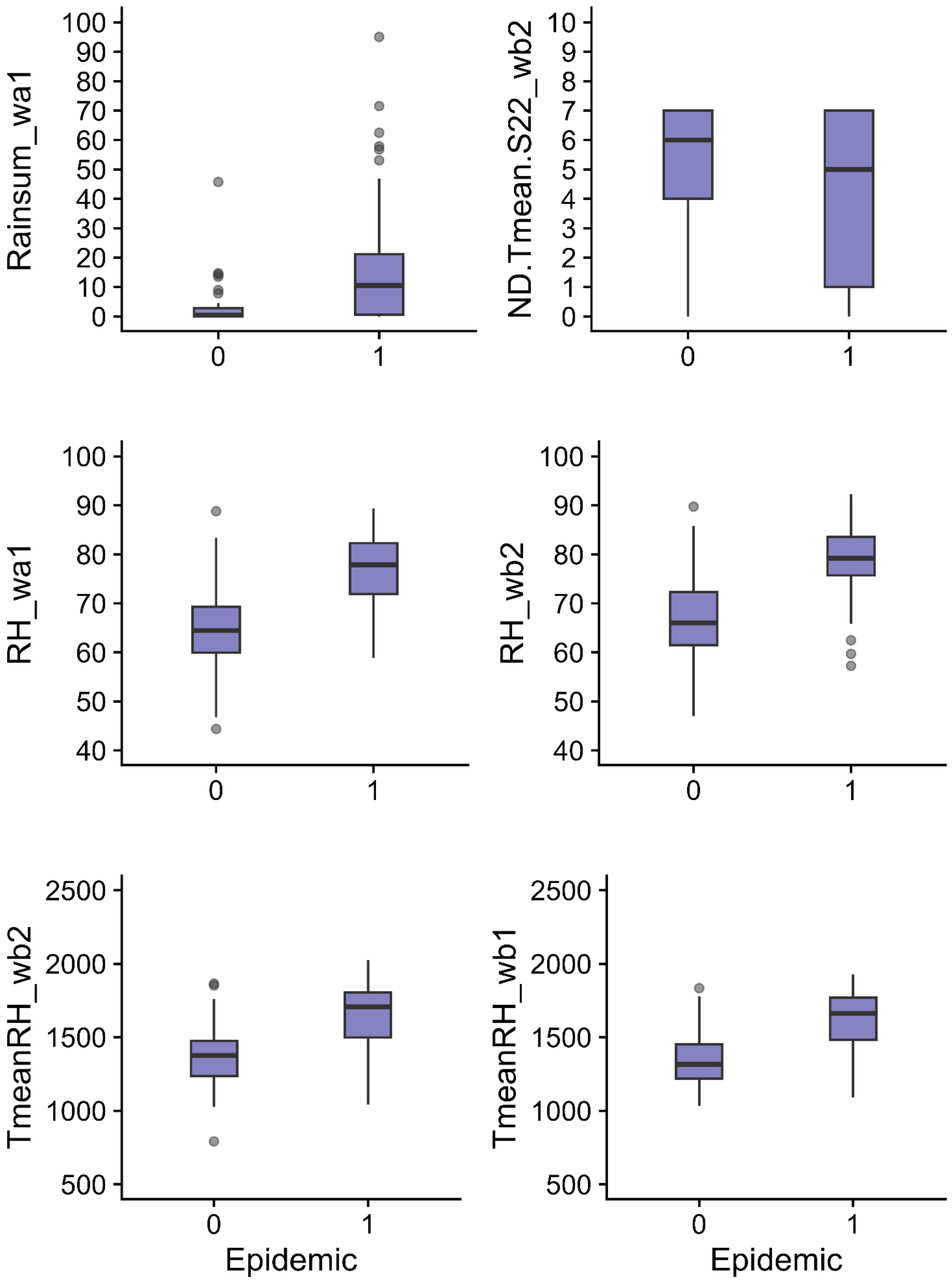

Figure 4. Box plots illustrating the distribution of the top six predictor variables of wheat blast disease in 143 cases classified as outbreak (1) and non-outbreak (0). Predictor variables are defined in Table 2.

4.3 Logistic regression models

The logistic regression models M1, M2, M3, and M4 were constructed using the selected best combinations of variables. There was no multicollinearity among the variables used in the different models, with VIF values all below 2.6. The weather variable relative humidity was the only one that was present in all models. The models and their statistics and performance are shown in Table 3.

The probability threshold for predicting the occurrence or non-occurrence of an outbreak, determined by maximizing the sum of specificity and sensitivity, was determined as 0.44 for M1, 0.68 for M2, 0.62 for M3, and 0.64 for M4 (Figure 5). When the threshold decreased, sensitivity increased, resulting in more cases classified as positive. Conversely, when the threshold increased, more cases were classified as negative, thereby increasing specificity and decreased sensitivity.

The accuracy of the models ranged from 80% (M2) to 85% (M4). The model that employed a greater number of variables exhibited higher accuracy. However, the model M1, which utilized only two variables, outperformed the model M2, which employed three variables. Additionally, the accuracy of M1 was comparable to that of M3, despite M3 including four variables.

In general, the sensitivity of the models exceeded the specificity, with the exception of model M1. The sensitivity values ranged from 0.80 (M1) to 0.91 (M2 and M4), while the specificity values ranged from 0.72 (M2) to 0.86 (M1). All models exhibited a robust discriminatory capacity, characterized by higher true positive rates compared to false positive rates, as indicated by the AUC values ranging from 0.89 to 0.91. Regarding the accuracy obtained from Leave-One-Out Cross-Validation (LOOCV), all models demonstrated good performance. M1 achieved an accuracy of 0.81, while M4 achieved a slightly lower accuracy of 0.80. Similarly, M2 and M3 exhibited accuracies of 0.76 and 0.77, respectively.

Table 3. Logistic regression models to predict wheat blast outbreaks, and their corresponding statistics and performance measures.

Model	Label(a)	Est. coef.(b)	P(c)	p*(d)	Acc(e)	Sens(f)	Spec(g)	AUC(h)	LOOCV(i)
M1	Intercept	-15.908	<0.001	0.44	0.83	0.80	0.86	0.89	0.81
	RH_wa1	0.137	<0.001						
	TmeanRH_wb2	0.004	<0.001						
M2	Intercept	-10.846	<0.001	0.68	0.8	0.91	0.72	0.9	0.76
	RH_wb2	0.164	<0.001						
	ND.Tmean.S22_wb2	-0.326	<0.001						
	Rainsum_wa1	0.069	<0.010						
M3	Intercept	-12.985	<0.001	0.62	0.83	0.86	0.8	0.9	0.77
	RH_wb2	0.141	<0.001						
	ND.Tmean.S22_wb2	-0.226	0.043						
	Rainsum_wa1	0.069	0.011						
	TmeanRH_wb1	0.002	0.114						
M4	Intercept	-13.633	<0.001	0.64	0.85	0.91	0.79	0.91	0.80
	RH_wa1	0.052	0.282						
	RH_wb2	0.114	0.007						
	ND.Tmean.S22_wb2	-0.264	0.026						
	Rainsum_wa1	0.053	0.077						
	TmeanRH_wb1	0.002	0.251						

[a] Predictor variables included in each model, along with the intercept; [b] Estimated coefficient for each predictor variable and intercept; [c] P-value, indicating the statistical significance of each variable in the model; [d] Threshold that maximizes the sum of sensitivity and specificity; [e] Accuracy, the overall accuracy of the model; [f] Sensitivity, also known as true positive of the model; [g] Specificity, the true negative rate of the model; [h] Area Under the Curve; [i] Accuracy of model obtained by Leave-One-Out Cross-Validation.

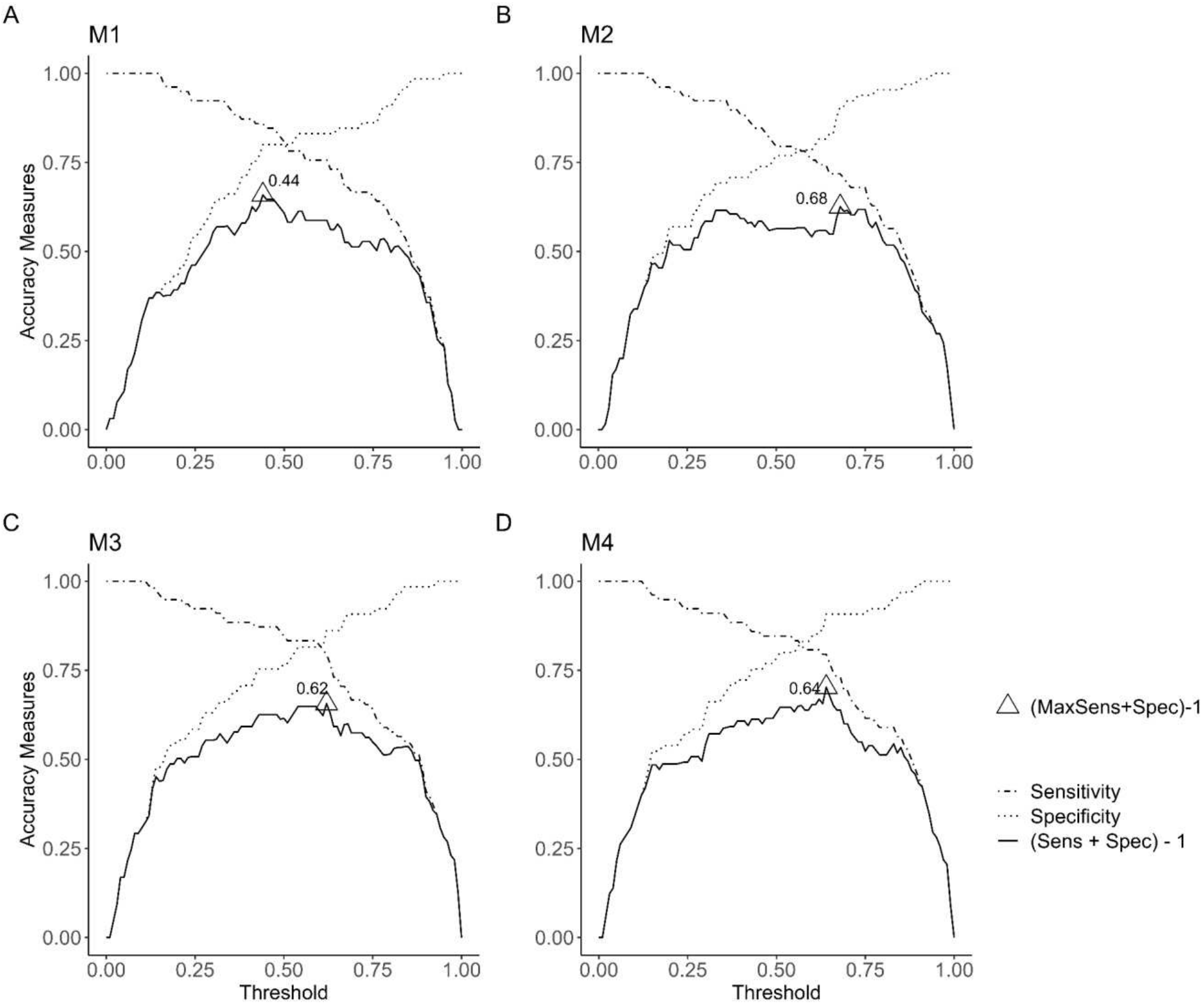

Figura 5. Threshold-based performance curves for sensitivity, specificity, and error rate for each model.

4.4 Model comparison and evaluation

Considering that the models exhibited comparable and similar performance metrics, a single model was further chosen for the purpose of retrospective prediction in each of the six locations. First, the outbreak frequency was examined for each model within each site to evaluate the comparability of the models in predicting outbreak occurrences (Figure 6).

The observed frequency of outbreak occurrences exhibited a consistent pattern across all sites (except for Passo Fundo). The M1 consistently displayed higher frequencies of outbreak occurrences compared to other models. The distributions and medians of the M3 and M4 exhibited a higher degree of similarity to each other compared to the M2. Consequently, among the models with the closest resemblance, the M3 model, which had fewer predictor variables, was selected as the preferred choice for further evaluation. The M3 applied in all years in each site reveals differences in risk index between the years, with years with high risk while others with risk lower than 0.5 (Figure 7).

In discussions with researchers specialized in the field, substantial agreement was observed between prediction model results and shared insights, with one notable exception being the inaccurate prediction for Passo Fundo. The years with high, median and lower outbreak frequencies at each site, when the prediction was according to the reports of the researchers, to investigate the probability of outbreak occurrence during the heading date of wheat were selected (Figure 8).

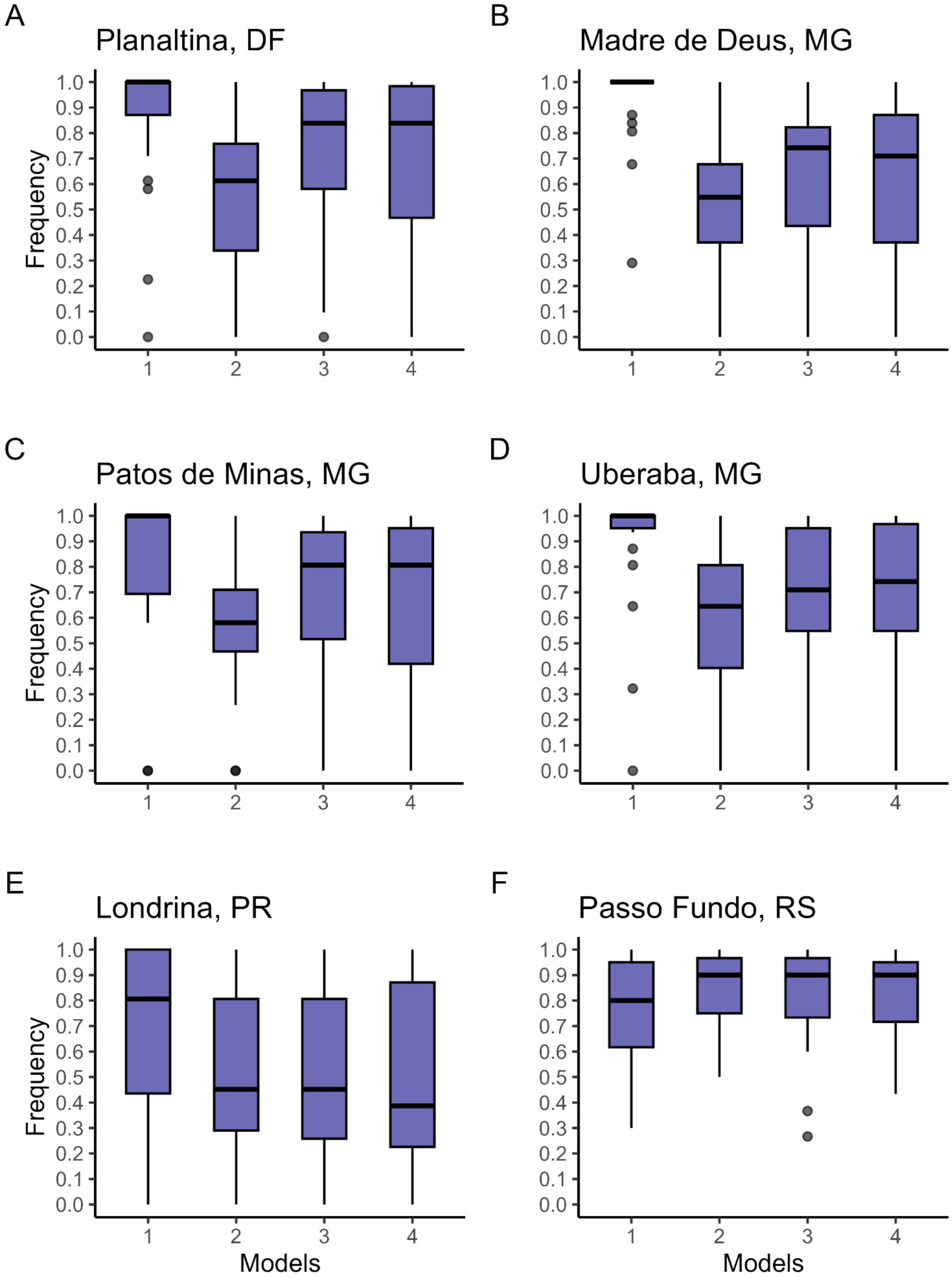

Figure 6. Box Plots of outbreak frequency across 23 years in April and May for sites A, B, C, and D, June to July for site E, and September for site F using four selected models.

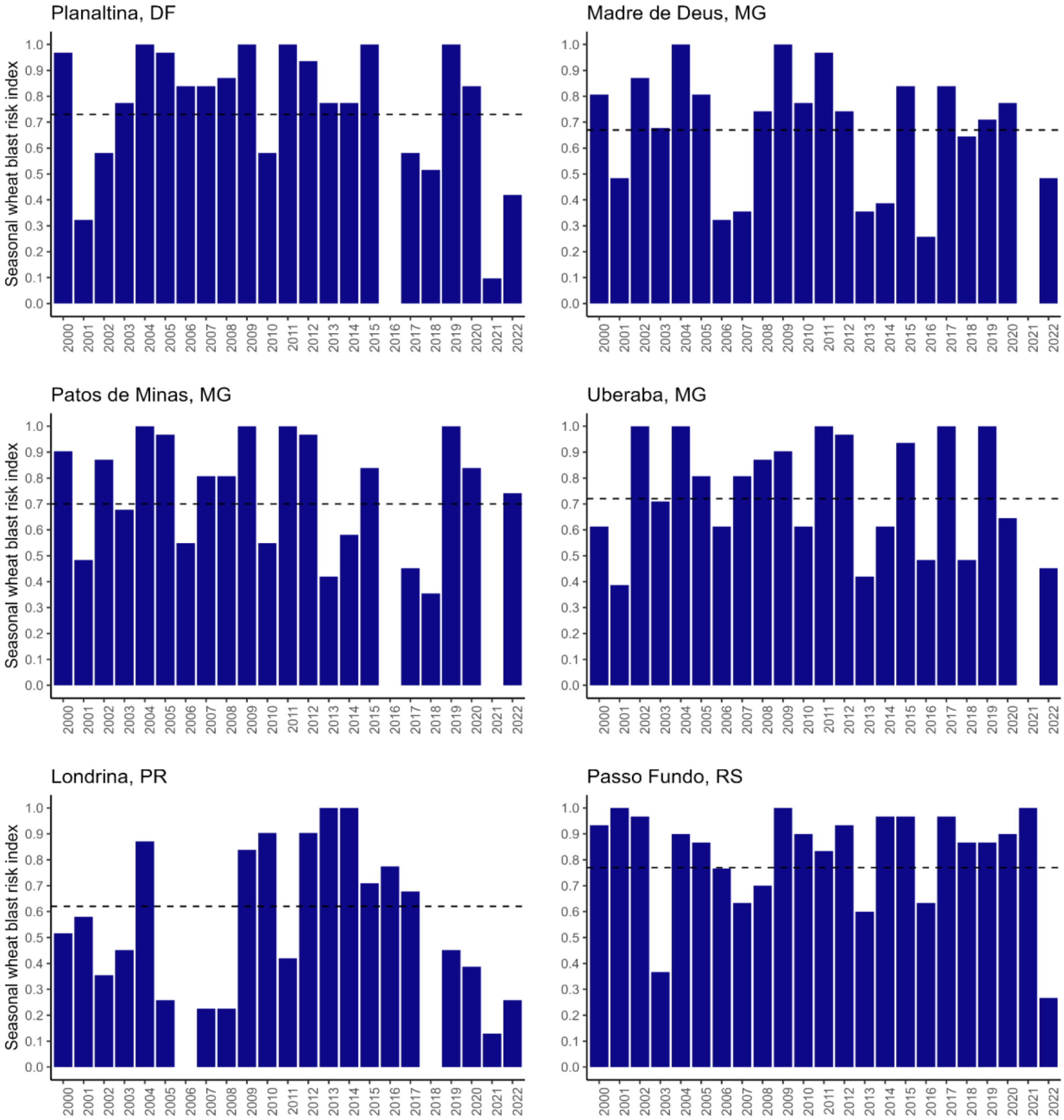

Figure 7. Seasonal wheat blast risk index in different sites of Brazil from 2000 to 2022. The dashed line indicates the general mean of the risk index in all years within each site.

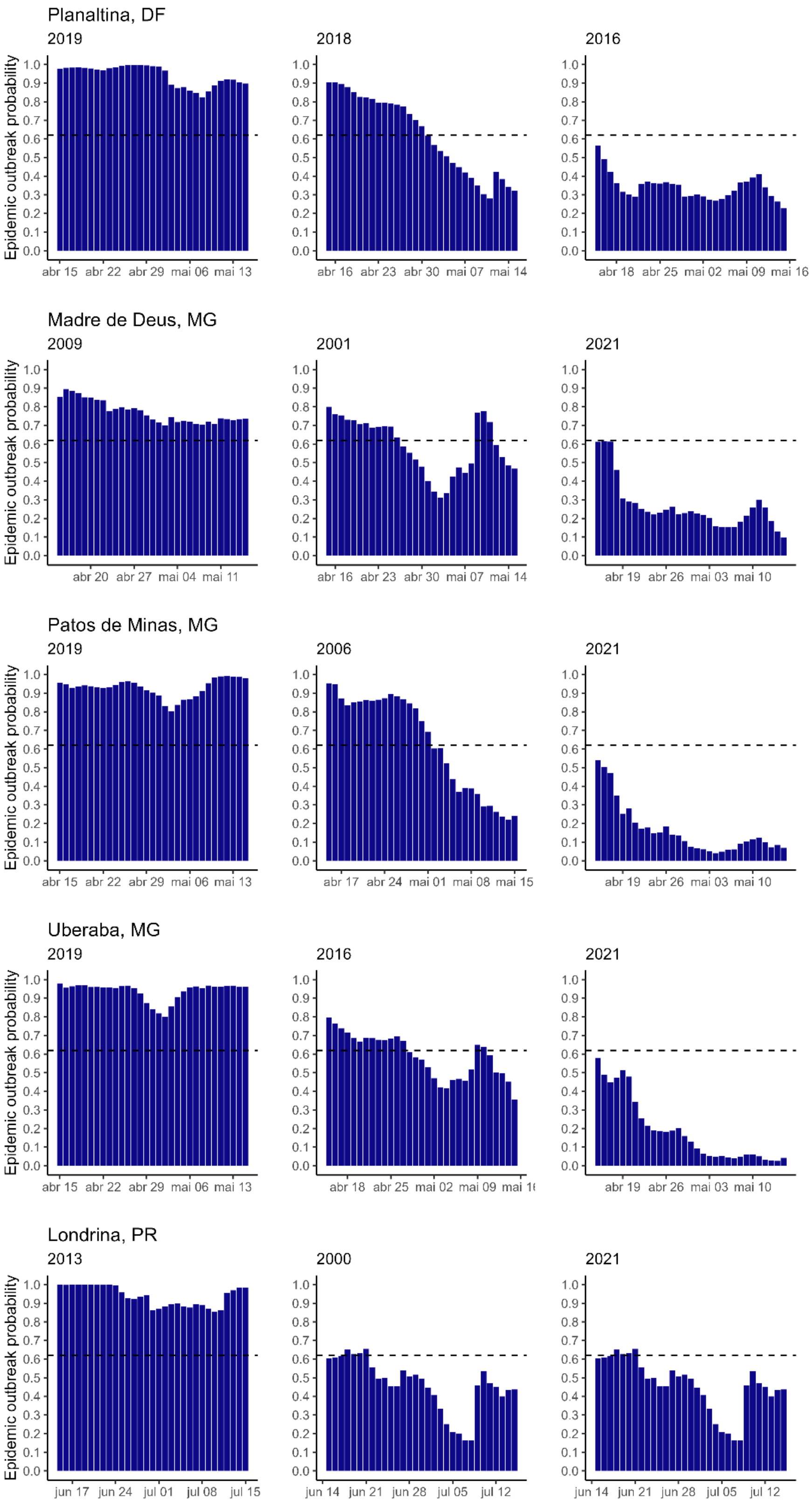

Figure 8. Probability of wheat blast outbreaks occurrence at a specific heading date of wheat for years selected with high, median and lower outbreak frequencies in different sites. The dashed line indicates the probability of outbreak threshold calculated by the maximum sum of sensitivity and specificity of M3.

In sites within the Cerrado biome, the results indicated that for years with high frequencies, regardless of the timing of wheat heading, which typically falls between April and May, the probability of outbreaks occurring in all locations consistently exceeded the threshold of 0.62. However, the risk probability exhibited a slight yet discernible decrease in early May across all selected sites. A remarkably similar pattern was observed for the chosen years with a disease occurrence frequency close to 0.5. Specifically, when the heading date of wheat occurred in April, there was a notable probability of an outbreak. However, when the heading date extended from the end of April onwards, the probability exhibited a significant decrease. However, in Madre de Deus (2001) and Uberaba (2016), an increase in the probability of outbreak occurrence was observed when the wheat heading date fell in the latter half of May. In the case of years selected with low frequencies of outbreak occurrence, it was observed that, in general, the probability of an outbreak was higher until approximately April 20th, without surpassing the designated outbreak threshold of 0.62. However, starting from April onwards, the probability of an outbreak occurring was found to be exceedingly low.

In Londrina, during the months of June to July (heading date period), in years with a high frequency of outbreaks, the probability of outbreaks consistently surpassed the threshold of 0.62, similar to the occurrences observed in Cerrado locations. In the year 2000, chosen as a representative year of the average frequency of outbreaks, there was a decrease in the occurrence of outbreaks towards the end of June and the beginning of July. However, around the 10th of July, there was an increase in the probability of outbreaks once again. A similar pattern was observed in the non-outbreaks year, although with a lower number of occurrences surpassing the threshold of 0.62, which indicates the occurrence of an outbreak.

Risk maps to Minas Gerais state depicting heading periods between April 15th and 30th, as well as May 1st and 15th, showed outcomes that aligned with empirical observations from the years 2019 and 2023 (Figure 9). The Triângulo Mineiro region, located in the westernmost portion of the state and considered the primary producer region, was specifically analyzed. The risk map indicated that in the outbreak year of 2019, even with heading in May, there remained a high probability of an outbreak, considering the cultivation of susceptible cultivars. However, in the non-outbreak year of 2023 (Dr. Paulo Kuhnem, personal communication), the probability of an outbreak decreased with delayed planting, which aligned with the reality of farmers planting later this year, mainly less susceptible cultivars.

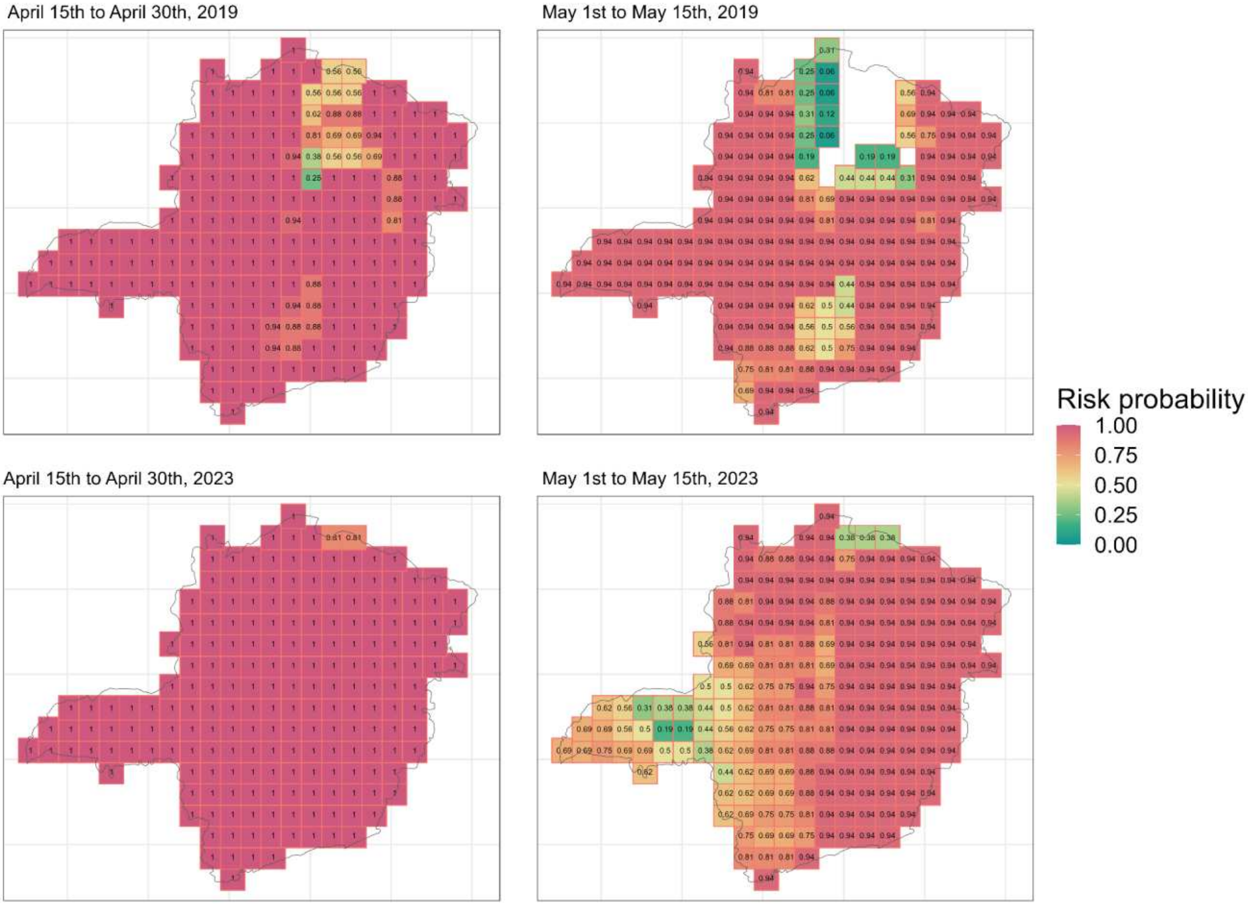

Figure 9. Risk map for wheat blast in two different heading periods for the state of Minas Gerais in an outbreak (2019) and non-outbreak year (2023). In the upper-right quadrant, there was an absence of weather data (white background) for certain points in the map.

5 DISCUSSION

This study represents a pioneering effort in wheat blast research, as this is the first model developed empirically using datasets of disease observed in the field. Weather variables in the vicinity of the heading date were identified as predictors of outbreak occurrence based on time window analysis, an approach that has been previously employed for modeling other diseases, particularly flowering diseases caused by Fusarium in cereals (De Wolf et al., 2003; Kriss et al., 2010; Dalla Lana et al., 2021).This study found that the key variables affecting the occurrence of wheat blast during the period preceding the heading date were relative humidity, the interaction between mean temperature and relative humidity, and the number of days, within a 7-day window, when the mean temperature remained below 22°C. In general, wetter and warmer conditions during the pre-heading period favored outbreak occurrence, which may be linked to conditions that favor inoculum production in the field (Fernandes et al., 2017). Similar pattern was suggested for Fusarium head blight of wheat, where empirical logistic models selected temperature and moisture-related variables 10 days prior to wheat flowering as predictors of outbreaks, as they would be related to perithecial inoculum production (De Wolf et al., 2003).

A previous study in a greenhouse setting showed that conidia production was higher under temperature of 28°C than under 23°C during conditions of increased relative humidity of the air (Alves and Fernandes, 2006). Moreover, Kovaleski et al. (2020), showed that the peak of conidia production occurred at a temperature range of 24 to 27°C, decreasing at lower and higher values (quadratic response). These findings are in line with our findings and may help explain the selection of moisture and temperature-related variables at periods preceding the heading stage, which may favor inoculum production. This is particularly relevant considering previous reports of auto-infection in wheat blast outbreaks (Cruz et al., 2015; Gongora-Canul et al., 2020).

In all models developed, the weather variables selected after the heading date were related only to moisture (sum of precipitation and mean relative humidity) specifically during the first week following the heading date. All models included at least one moisture-related variable after the heading date and temperature-related variables were not selected. The occurrence of blast outbreaks in wheat in Brazil has been linked to rainy periods during the heading stage (Fernandes et al., 2017). Although increased temperatures (>25 °C) have been identified as an important driver of wheat blast outbreaks (Goulart et al., 2007; Fernandes et al., 2017; Mills et al., 2020), they were not selected as a variable in our study. Cardoso et al. (2008) demonstrated that although the optimal temperature for infection is 25°C, the disease can still establish at cooler conditions (≤20°C) conditioned to extended period (> 30h) of moisture. Indeed, moisture conditions have been shown to significantly influence the potential infection of PoT (Montes et al., 2022) and PoL (*Lolium* lineage) (Mills et al., 2020). These authors further hypothesized that the significant role of relative humidity during the heading stage can be attributed to its influence on germination and infection process, as previously observed in the case of MoO (*Oryzae* lineage) (Suzuki, 1975).

Our "best" model (M3) provided realistic historical predictions of a "seasonal wheat blast index" for locations in the Brazilian Cerrado region and parts of South Brazil - e.g. Londrina. For Passo Fundo, however, inconsistencies were observed between the model's predictions and observations. Out of the 23 years analyzed, 17 presented a high risk of blast outbreak during the heading stage, even though blast is not a prevalent issue in these regions (Cruz et al., 2016; Fernandes et al., 2017). In Passo Fundo, the heading date of wheat crops is commonly observed in September, coinciding with a historical mean temperature (2007 to 2022) range of 13 to 20°C (INMET, 2023). This temperature range may limit disease development, as evidenced by Mills et al. (2020), who found that at 20°C there was no notable increase in disease intensity.

It is worth noting that none of the models utilized in this study included temperature as a predictor for post-heading disease, which may also explain the inaccurate predictions for this location, in addition to the limited representativeness of the data used in this study to build the model. Only two locations, Guarapuava and Palmeiras, were considered representative of the similar climatic conditions found in Passo Fundo, both with Cfb classification, indicating a temperate climate (Wrege et al., 2012; Aparecido et al., 2016). These two sites contributed only nine cases (6.29% of the total cases) used to develop the model. To achieve accurate predictions of wheat blast in temperate climates, it is crucial to incorporate a more extensive dataset that includes a sufficient number of representative experiments from this specific climatic region. However, the scarcity of data in this area is primarily due to the general low incidence of wheat blast in these regions. More importantly, the absence of wheat blast outbreaks in temperate conditions can be due to absence or low levels of primary inoculum, likely limited by the low temperatures observed in the winter months prior to the heading. These low temperatures act as limiting factors for inoculum build-up (Cruz et al., 2016a; Fernandes et al., 2017).

In Londrina, located in the northwestern state of Paraná, according to personal communication by the agronomist Dalvin Tochiaki Sato, there was no occurrence of outbreaks in the last five years (after 2018). Additionally, Fernandes et al. (2017) and Cruz et al. (2016a) reported that wheat blast affected about 80 to 100% of wheat area in 2004 and 2009 in northern Paraná, where Londrina is located. These reports are consistent with our predictions. Londrina climate is classified as subtropical, belonging to the Cfa climate type (Wrege et al., 2012). The dataset used in this study comprised a larger number of cases from locations with the same climate classification as Londrina compared to cases with the same classification as Passo Fundo. Notably, approximately 11.9% of the dataset exhibited a Cfa climate classification, which can have contributed to a more robust prediction outcome.

For the sites in the Brazilian Cerrado, the predictive model was evaluated by researchers Angelo Aparecido Barbosa Sussel (Planaltina), Vanoli Fronza (Uberaba and Madre de Deus), and Maurício Antonio Coelho Filho (Patos de Minas). Their assessments confirmed that the model's results align with reality. In addition, Júlio Albrecht, a researcher at Embrapa Cerrados, reported severe outbreaks of wheat blast in Planaltina during the years 2005 and 2009, consistent with the predictions made by the model. Additionally, a subsequent report in 2019 documented severe outbreaks of wheat blast in various locations within the Brazilian Cerrado, including Uberaba and Patos de Minas, corroborating the model's predictions (Embrapa, 2019).

In the Brazilian Cerrado region, the heading date of wheat typically occurs in April and May, but tends to concentrate in April due to risk of drought in May, which can adversely affect wheat production when grown in rainfed (Albuquerque and da Silva, 2008; Pereira et al., 2019). Our findings indicate that in years characterized by medium and low outbreak risk levels, the probability of outbreak occurrence in May was significantly lower compared to that in April, which corroborates previous findings (Coelho et al., 2016). This suggests that a decrease in rainfall (lowering moisture) during this specific month can decrease the occurrence of disease. However, it is important to note that such dry conditions also increase the vulnerability of crop production to drought-induced losses.

The scarcity of weather stations across Brazil's geographic expanse introduces potential data gaps, which can limit the availability of weather-related variables for agronomic studies (Duarte and Sentelhas, 2020). The lack of meteorological stations in close proximity to study areas represents a challenge in obtaining comprehensive weather data. Consequently, satellite-based meteorological data sources have emerged as a potential alternative for acquiring such information (Savary et al., 2012; Bebber et al., 2016; Alves et al., 2022). The NASA POWER platform (Prediction Of Worldwide Energy Resources), used in this study, serves as an interesting tool for acquiring climate data. However, it is important to point out that the spatial resolution of the platform is 0.5° x 0.5° latitude/longitude. This spatial resolution can introduce potential interference in the estimation process, as within a selected point, there may be considerable local weather variations. Additionally, the platform does not provide access to hourly data, which may be important for building or testing disease models (Chandler et al., 2004). However, some studies have been performed to evaluate the accuracy of NASA POWER in estimating weather data, and these studies have consistently produced encouraging results for most variables analyzed (Duarte and Sentelhas, 2020; Rodrigues and Braga, 2021).

In conclusion, this study confirms that weather plays a crucial role in the epidemiology of wheat blast. The identification of three windows, consisting of specific weather variables around the heading date, provided valuable insights into the dynamics of disease outbreaks especially in the tropics where temperature seemed to not be a limiting factor during infection. The evaluated model was found to be reliable for predicting wheat blast in the Brazilian Cerrado region and locations in Southern Brazil with a subtropical climate. Overall, this study enhances our understanding of the complex relationship between weather variables and wheat blast outbreak occurrence, contributing valuable insights for disease management. For example, in years characterized by high and medium levels of outbreak risk, which demonstrate a high probability of outbreaks occurring in the month of April, the use of fungicides is imperative to mitigate the detrimental effects of wheat blast on crop yield. Ascari et al. (2021) have previously suggested the potential requirement of three fungicide sprays in heading. Conversely, in years or periods of lower risk of outbreaks, a reduction in these applications might be feasible as well as the use of more susceptible cultivars. In fact, considering that the models were trained using susceptible cultivars, further refinements in the model might be possible should new data become available in the future.

REFERENCES

Albuquerque, A. C. S., da Silva, A. G. 2008. Agricultura tropical: quatro décadas de inovações tecnológicas, institucionais e políticas. Brasília, DF: Embrapa Informação Tecnológica.

Alves, K. J. P., Fernandes, J. M. C. 2006. Influência da temperatura e da umidade relativa do ar na esporulação de *Magnaporthe grisea* em trigo. Fitopatologia Brasileira. v. 31, p. 579-584.

Alves, K. S., Shah, D. A., Dillard, H. R., Del Ponte, E. M., Pethybridge, S. J. 2022. From reanalysis data to inference: a framework for linking environment to plant disease epidemics at the regional scale.

Aparecido, L. E. D. O., Rolim, G. D. S., Richetti, J., Souza, P. S. D., Johann, J. A. 2016. Köppen, Thornthwaite and Camargo climate classifications for climatic zoning in the State of Paraná, Brazil. Ciência e Agrotecnologia. v. 40, p. 405-417.

Ascari, J.P., Barro, J.P., Santana, F.M., Padua, J.M.V., Maciel, J.L.N., Lau, D., Torres, G.A.M., Sbalcheiro, C.C., Seixas, C.D.S., Goulart, A.C.P., Sussel, A.A.B., Schipanski, C.A., Chagas, D.F., Coelho, M.A.O., Montecelli, T.D.N., Amaral, D.R., Custodio, A.A.P., Moreira, L.S.O., Utiamada, C.M., Venâncio, W.S., Goussain, R.C.S., Alves, K.S., Del Ponte, E.M. 2021. Sequential Post-Heading Applications for Controlling Wheat Blast: A 9-Year Summary of Fungicide Performance in Brazil. Plant Disease. v. 105, n. 12, p. 4051-4059.

Barea, G., Toledo, J. 1996. Identificación y zonificación de Pyricularia o brusone (*Pyricularia oryzae*) en el cutivo de trigo en el departamento de Santa Cruz, Informe Tecnico Proyecto de Investigacion Trigo. Centro de Investigación Agrícola Tropical, Santa Cruz de la Sierra, Bolívia.

Bebber, D. P., Castillo, Á. D., Gurr, S. J. 2016. Modelling coffee leaf rust risk in Colombia with climate reanalysis data. Philosophical Transactions of the Royal Society B: Biological Sciences. v. 371, p. 20150458.

Bock, C.H., Pethybridge, S.J., Barbedo, J.G.A., Esker, P.D., Mahlein, A.-K., Del Ponte, E.M. 2022. A phytopathometry glossary for the twenty-first century: towards consistency and precision in intra- and inter-disciplinary dialogues. Tropical Plant Pathology. v. 47, p. 14–24.

Cabrera, M.G., Gutierrez, S.A. 2007. Primer registro de *Pyricularia oryzae* en cultivos de trigo del NE de Argentina, in: Jornada de Actualización En Enfermedades de Trigo. Instituto Fitotécnico de Santa Catalina, Lavallol, Buenos Aires, p. 18–180.

Campbell, C.L., Madden, L.V. 1990. Introduction to Plant Disease Epidemiology. Wiley, New York.

Cardoso, C.A. de A., Reis, E.M., Moreira, E.N. 2008. Development of a warning system for wheat blast caused by *Pyricularia grisea*. Summa Phytopathology. v. 34, p. 216–221.

Ceresini, P.C., Castroagudín, V.L., Rodrigues, F.Á., Rios, J.A., Eduardo Aucique-Pérez, C., Moreira, S.I., Alves, E., Croll, D., Maciel, J.L.N. 2018. Wheat blast:

Past, present, and future. Annu. Rev. Phytopathology. v. 56, p. 427–456.

Chandler, W.S., Whitlock, C.H., Stackhouse Jr., P.W. 2004. NASA climatological data for renewable energy assessment. Acesso em: 24 de junho de 2023. Disponível em: https://power.larc.nasa.gov/docs/methodology/.

Coelho, M.A. de O., Torres, G.A.M., Cecon, P.R., Santana, F.M. 2016. Sowing date reduces the incidence of wheat blast disease. Pesquisa Agropecuária Brasileira. v. 51, p.631–637.

Couch, B.C., Kohn, L.M. 2002. A multilocus gene genealogy concordant with host preference indicates segregation of a new species, *Magnaporthe oryzae*, from *M. grisea*. Mycologia. v. 94, p. 683–693.

Cruz, C.D., Kiyuna, J., Bockus, W.W., Todd, T.C., Stack, J.P., Valent, B. 2015. *Magnaporthe oryzae* conidia on basal wheat leaves as a potential source of wheat blast inoculum. Plant Pathology. v. 64, p. 1491–1498.

Cruz, C.D., Santana, F.M., Todd, T.C., Maciel, J.L.N., Kiyuna, J., Baldelomar, D.F., Cruz, A.P., Lau, D., Seixas, C.S., Goulart, A.C.P., Sussel, A.A., Schipanski, C.A., Chagas, D.F., Coelho, M., Montecelli, T.D.N., Utiamada, C., Custódio, A.P., Rivadeneira, M.G., Bockus, W.W., Valent, B. 2019. Multi-environment assessment of fungicide performance for managing wheat head blast (WHB) in Brazil and Bolivia. Tropical Plant Pathology. v. 44, p. 183–191.

Cruz, C.D., Valent, B., 2017. Wheat blast disease: danger on the move. Tropical Plant Pathology. v. 42, p. 210–222.

Cruz, C. D., Magarey, R.D., Christie, D.N., Fowler, G.A., Fernandes, J.M., Bockus, W.W., Valent, B., Stack, J.P. 2016a. Climate Suitability for *Magnaporthe oryzae* Triticum Pathotype in the United States. Plant Disease. v. 100, p. 1979–1987.

Cruz, M.F.A., Rios, J.A., Araujo, L., Ávila Rodrigues, F. 2016b. Infection process of *Pyricularia oryzae* on the leaves of wheat seedlings. Tropical Plant Pathology. v. 41, p. 123–127.

Cunha, G.R. da, Pasinato, A., Pimentel, M.B.M., Haas, J.C., Maluf, J.R.T., Pires, J.L.F., Dalmago, G.A., Santi, A. 2011. Regiões para trigo no Brasil: ensaios de VCU, zoneamento agrícola e época de semeadura. In: Pires, J.L.F.; Vargas, L.; Cunha, G.R. da. (Ed.). Trigo no Brasil: bases para produção competitiva e sustentável. Passo Fundo: Embrapa Trigo. p. 27-40.

Dalla Lana, F., Madden, L.V., Paul, P.A. 2021. Logistic models derived via LASSO methods for quantifying the risk of natural contamination of maize grain with deoxynivalenol. Phytopathology. v. 111, p. 2250–2267.

De Wolf, E.D., Isard, S.A. 2007. Disease Cycle Approach to Plant Disease Prediction. Phytopathology. v. 45, p. 203–220.

De Wolf, E.D., Madden, L.V., Lipps, P.E. 2003. Risk assessment models for wheat fusarium head blight epidemics based on within-season weather data. Phytopathology. v. 93, p. 428–435.

Dianese, A. D. C., Zacaroni, A. B., de Souza, B. C. P., da Silva Pagani, A. P., Pinheiro,

N. O., de Castro Gomes, E. M., Torres, G. A. M., Consoli, L., Café-Filho, A. C. 2021. Evaluation of wheat genotypes for field resistance to wheat blast caused by *Magnaporthe oryza*e pathotype Triticum (MoT) and correlation between yield loss and disease incidence in the Brazilian Cerrado. Euphytica. v. 217, p. 1-13.

Dos Anjos, J.R.N., Da Silva, D.B., D'Avila Charchar, M.J., Rodrigues, G.C. 1996. Occurrence of blast fungus (*Pyricularia grisea*) on wheat and rye in the savanna region of central Brazil. Pesquisa Agropecuária Brasileira. v. 31, p. 79–82.

Dos Santos, G. B., Coelho, M. A., Del Ponte, E. M. 2022a. Critical-point yield loss models based on incidence and severity of wheat head blast epidemics in the Brazilian Cerrado. European Journal of Plant Pathology. v. 165, n. 3, p. 421-431.

Dos Santos, G. B., Coelho, M. A., Del Ponte, E. M. 2022b. Incidence-severity relationships in non-treated and fungicide-treated wheat head blast epidemics in Brazil. European Journal of Plant Pathology. v. 163, n. 4, p. 1003-1010.

Duarte, Y. C., Sentelhas, P. C. 2020. NASA/POWER and DailyGridded weather datasets—how good they are for estimating maize yields in Brazil?. International Journal of Biometeorology. 64, 319-329.

Embrapa. 2019. Brusone em lavouras de trigo no Brasil Central - Safra 2019. Nota Técnica, Embrapa Trigo, Passo Fundo, RS, Brasil.

Fernandes, J.M., Del Ponte, E.M., Ascari, J.P., Krupnik, T.J., Pavan, W., Berton, T. 2021. Towards an early warning system for wheat blast: epidemiological basis and model development, In: Oliver, R. (Eds.), Burleigh Dodds Series in Agricultural Science. Burleigh Dodds Science Publishing. p. 623–642.

Fernandes, J.M.C., Nicolau, M., Pavan, W., Hölbig, C.A., Karrei, M., de Vargas, F., Bavaresco, J.L.B., Lazzaretti, A.T., Tsukahara, R.Y. 2017. A weather-based model for predicting early season inoculum build-up and spike infection by the wheat blast pathogen. Tropical plant pathology. v. 42, p. 230–237.

Freeman, E. A. and Moisen, G. 2008. PresenceAbsence: An R Package for Presence-Absence Model Analysis. Journal of Statistical Software. v. 23, n.11, p.1-31, <http://www.jstatsoft.org/v23/i11>.

Friedman, J., Hastie, T., Tibshirani, R. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software. v. 33, n. 1, p.1-22, <https://www.jstatsoft.org/v33/i01/>.

Gomes, D.P., Rocha, V.S., Rocha, J.R. do A.S. de C., Pereira, O.L., Souza, M.A. de. 2018. Potential of transmission of *Pyricularia graminis-tritici* from plant to seed and from seed to seedling in wheat genotypes with different degrees of blast resistance. Journal of Seed Science. v. 40, p. 16–24.

Gongora-Canul, C., Salgado, J.D., Singh, D., Cruz, A.P., Cotrozzi, L., Couture, J., Rivadeneira, M.G., Cruppe, G., Valent, B., Todd, T., Poland, J., Cruz, C.D. 2020. Temporal Dynamics of Wheat Blast Epidemics and Disease Measurements Using Multispectral Imagery. Phytopathology. v. 110, p. 393–405.

Gonzalez-Dominguez, E., Caffi, T., Rossi, V., Salotti, I., Fedele, G. 2023. Plant disease models and forecasting: changes in principles and applications over the last 50 years. Phytopathology. v. 113, n. 4, p. 678-693.

Goulart, A. C. P., Sousa, P. G., Urashima, A. S. 2007. Danos em trigo causados pela infecção de *Pyricularia grisea*. Summa Phytopathologica. v. 33, p. 358-363.

Goulart, A.C.P., Paiva, F., Mesquita, A.N. 1990. Occurrence and losses caused by wheat (*Triticum aestivum* L.) blast (*Pyricularia oryzae* Cav.) in the state of Mato Grosso do Sul, in 1988, in: Annual Wheat Newsletter. Department of Agronomy, Colorado State University, Fort Collins.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. 2009. The elements of statistical learning: data mining, inference, and prediction. New York: springer. v. 2, p. 1-758.

Hosmer, D.W., Lemeshow, S., Sturdivant, R.X. 2013. Applied logistic regression, Third edition. ed, Wiley series in probability and statistics. Wiley, Hoboken, New Jersey.

Igarashi, S., Utiamada, C.M., Igarashi, L.C., Kazuma, A.H., Lopes, R.S. 1986. Occurrence of *Pyricularia* sp. in wheat (*Triticum aestivum* L.) in the State of Paraná, Brazil. Fitopatologia Brasileira. v. 11, p. 351–352.

INMET - Instituto Nacional de Meteorologia. 2023. Estação Meteorológica de Observação de Superfície Automática, Brasília, DF, Brasil. Disponível em: http://www.inmet.gov.br/portal/index.php?r=estacoes/estacoesAutomaticas. Acesso em: 04 de julho de 2023.

James, G., Witten, D., Hastie, T., Tibshirani, R. 2013. An introduction to statistical learning. New York: springer. v. 112, p. 18.

Kim, K.-H., Choi, E.D. 2020. Retrospective Study on the Seasonal Forecast-Based Disease Intervention of the Wheat Blast Outbreaks in Bangladesh. Frontiers in Plant Science. v. 11, p. 570381.

Kohli, M.M., Mehta, Y.R., Guzman, E., De Viedma, L., Cubilla, L.E. 2011. *Pyricularia* blast – a threat to wheat cultivation. Czech J. Genetic Plant Breed. v. 47, p. S130–S134.

Köppen, W. 1918. Klassification der klimate nach temperatur, niederschlag und jahreslauf. Petermanns Geographische Mitteilungen, Gotha. v. 64, p. 193-203.

Kovaleski, M., Maciel, J.L.N., Dos Santos, G.B., Silva, A.N. da, Deuner, C.C. 2020. Conidia sporulation of *Pyricularia oryzae* in segments of wheat plants under six different temperatures. Ciência Rural. v. 50. n. 4, p. e20190573.

Kriss, A. B., Paul, P. A., & Madden, L. V. 2010. Relationship between yearly fluctuations in Fusarium head blight intensity and environmental variables: a window-pane analysis. Phytopathology. v. 100, n. 8, p. 784-797.

Madden, L.V., Hughes, G., van den Bosch, F. 2007. The Study of Plant Disease Epidemics, Epidemiology. APS Press, St. Paul.

Malaker, P.K., Barma, N.C.D., Tiwari, T.P., Collis, W.J., Duveiller, E., Singh, P.K., Joshi, A.K., Singh, R.P., Braun, H.-J., Peterson, G.L., Pedley, K.F., Farman, M.L., Valent, B. 2016. First Report of wheat blast caused by *Magnaporthe oryzae* pathotype Triticum in Bangladesh. Plant Disease. v. 100, p. 2330.

Martínez, S.I., Sanabria, A., Fleitas, M.C., Consolo, V.F., Perelló, A. 2018. Wheat blast: Aggressiveness of isolates of *Pyricularia oryzae* and effect on grain quality. Journal of King Saud University-Science. v. 31, n. 2, p. 150-157.

McLeod, A., Xu, C., Lai, Y. 2020. bestglm: Best Subset GLM and Regression Utilities. R package version 0.37.3, <https://CRAN.R-project.org/package=bestglm>.

Mills, K. B., Madden, L. V., Paul, P. A. 2020. Quantifying the effects of temperature and relative humidity on the development of wheat blast incited by the Lolium pathotype of *Magnaporthe oryzae*. Plant Disease. v. 104(10), p. 2622-2633.

Montes, C., Hussain, S. G., Krupnik, T. J. 2022. Variable climate suitability for wheat blast (*Magnaporthe oryzae* pathotype Triticum) in Asia: results from a continental-scale modeling approach. International Journal of Biometeorology. v. 66, p. 2237–2249.

Mottaleb, K.A., Singh, P.K., Sonder, K., Kruseman, G., Tiwari, T.P., Barma, N.C.D., Malaker, P.K., Braun, H.-J., Erenstein, O. 2018. Threat of wheat blast to South Asia's food security: An *ex-ante* analysis. PLOS ONE. v. 13, n. 5, p. e0197555.

Pasinato, A., Cunha, G.R. da, Fontana, D.C., Monteiro, J.E.B. de A., Nakai, A.M., Oliveira, A.F. de. 2018. Potential area and limitations for the expansion of rainfed wheat in the Cerrado biome of Central Brazil. Pesquisa Agropecuária Brasileira. v. 53, p. 779–790.

Pereira, J. F., Cunha, G. R., Moresco, E. R. 2019. Improved drought tolerance in wheat is required to unlock the production potential of the Brazilian Cerrado. Crop Breeding and Applied Biotechnology. v. 19, p. 217-225.

Perelló, A.E., Martinez, I., Sanabria, A., Altamirano, R., Sibole, J.V. 2017. Pathogenicity of isolates of *Magnaporthe* spp. from wheat and grasses infecting seedlings and mature wheat plants in Argentina. Plant Pathology. v. 66, p. 1149–1161.

Picinini, E. C., Fernandes, J. M. C. 1990. Occurrence of wheat blast (*Pyricularia oryzae*) in commercial fields in the state of Rio Grande do Sul, Brazil. Fitopatologia Brasileira (Brazil). v. 15, p. 83–84

Pizolotto, C.A., Maciel, J.L.N., Fernandes, J.M.C., Boller, W. 2019. Saprotrophic survival of *Magnaporthe oryzae* in infested wheat residues. European Journal of Plant Pathology. v. 153, p. 327–339.

Prabhu, A.S., Filippi, M.C., Castro, N., 1992. Pathogenic variation among isolates of *Pyricularia oryzae* affecting rice, wheat, and grasses in Brazil. International Journal of Pest Management. v. 38, p. 367–371.

Ranstam, J., Cook, J.A. 2018. LASSO regression. British Journal of Surgery. v. 105, p. 1348–1348.

Rodrigues, G. C., Braga, R. P. 2021. Evaluation of NASA POWER reanalysis products to estimate daily weather variables in a hot summer mediterranean climate. Agronomy. v. 11, p. 1207.

Roy, K.K., Muzahid-E-Rahman, Reza, M.A., Mustarin, K.E., Malaker, P.K., Barma, N.C.D., Hossain, I., He, X., Singh, P.K. 2020. First report of triticale blast caused by the fungus *Magnaporthe oryzae* pathotype Triticum in Bangladesh. Canadian Journal of Plant Pathology. v. 43, n. 2, p. 288-295.

Santana, F.M., Lau, D., Sbalcheiro, C.C., Sussel, A.A.B., Venancio, W.S., Pádua, J.M.V., Costa, A.A. 2021. Eficiência de fungicidas para controle de brusone de trigo: resultados dos ensaios cooperativos na safra 2020. Embrapa Trigo, Passo Fundo, RS. Brasil.

Santana, F.M., Maciel, J.N., Lau, D., Cargnin, A., Seixas, C.D.S., Bassoi, M.C., Goulart, A.C.P., Sussel, A.A.B., Schipanski, C.A., Montecelli, T.D.N., Chargas, J.H., Guizeline, J. 2013. Eficiência de fungicidas para o controle da brusone do trigo: resultados dos ensaios cooperativos - safra 2011. Embrapa Trigo, Passo Fundo, RS. Brasil.

Savary, S., Nelson, A., Willocquet, L., Pangga, I., & Aunario, J. 2012. Modeling and mapping potential epidemics of rice diseases globally. Crop Protection. v. 34, p. 6-17.

Shah, D.A., Paul, P.A., De Wolf, E.D., Madden, L.V. 2019. Predicting plant disease epidemics from functionally represented weather series. Philosophical Transactions of the Royal Society B. v. 374, n. 1775, p. 20180273.

Singh, P.K., Gahtyari, N.C., Roy, C., Roy, K.K., He, X., Tembo, B., Xu, K., Juliana, P., Sonder, K., Kabir, M.R., Chawade, A. 2021. Wheat blast: A disease spreading by intercontinental jumps and its management strategies. Frontiers in Plant Science. v. 12, p. 710707.

Sparks A. 2023. nasapower: NASA-POWER Data from R. R package version 4.0.10, <https://CRAN.R-project.org/package=nasapower>

Suzuki, H. 1975. Meteorological factors in the epidemiology of rice blast. Annual Review of Phytopathology. v. 13, p. 239-256.

Tembo, B., Mulenga, R.M., Sichilima, S., M'siska, K.K., Mwale, M., Chikoti, P.C., Singh, P.K., He, X., Pedley, K.F., Peterson, G.L., Singh, R.P., Braun, H.J. 2020. Detection and characterization of fungus (*Magnaporthe oryzae* pathotype Triticum) causing wheat blast disease on rain-fed grown wheat (*Triticum aestivum* L.) in Zambia. PLOS ONE. v. 15, p. 1–10.

Torres, G.A.M., Ferreira, J.R., Binneck, E., Maciel, J.L.N., Consoli, L. 2022. Blast disease and wheat production in Brazil. Pesquisa Agropecuária Brasileira. v. 57, p. e02487.

Tosa, Y., Chuma, I. 2014. Classification and parasitic specialization of blast fungi. Journal of General Plant Pathology. v. 80, p. 202–209.

Viedma L. Q., Kohli, M., Cubilla L., Cabrera G. 2010. Manejo integrado de mancha amarilla y la *Piricularia* en el cultivo de trigo en Paraguay. In Tercer Semin. *Nacional de Trigo:* "*Del Grano al Pan*", ed. M Kohli, LE Cubilla, G Cabrera, p.

31–42. Asuncion, Paraguai: Capeco/Inbio.

Viedma L. Q., Morel W. 2002. Añublo o *Piricularia* del Trigo (Wheat Blast). Díptico. Capitán Miranda, Itapúa, Parag.: MAG/DIA/CRIA.

Viedma, L. Q. 2005. Wheat blast occurrence in Paraguay. Phytopathology: S152.

Wrege, M. S., Steinmetz, S., Reisser Junior, C., de Almeida, I. R. 2012. Atlas climático da região sul do Brasil: estados do Paraná, Santa Catarina e Rio Grande do Sul. 2012. Pelotas: Embrapa Clima Temperado; Colombo: Embrapa Florestas.

Zadoks J. C., Chang T. T., Konzak C. F. 1974. A decimal code for the growth stages of cereals. Weed research. v. 14, p. 415–421.