Infectious Disease Modelling 10 (2025) 60–74

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

Flexible regression model for predicting the dissemination of *Candidatus* Liberibacter asiaticus under variable climatic conditions☆,☆☆

Julio Cezar Souza Vasconcelos [a, b, *], Silvio Aparecido Lopes [a], Juan Camilo Cifuentes Arenas [a], Maria Fátima das Graças Fernandes da Silva [c]

[a] *Departamento de Pesquisa e Desenvolvimento, Fundo de Defesa da Citricultura (Fundecitrus), Av. Dr. Adhemar Pereira de Barros, 201 - Vila Melhado, Araraquara, 14807-040, SP, Brazil*

[b] *Instituto de Ciência e Tecnologia da Universidade Federal de São Paulo, Rodovia Presidente Dutra km 145. Jardim Diamante, São José dos Campos, 12223-201, SP, Brazil*

[c] *Departamento de Química, Universidade Federal de São Carlos, Rodovia Washington Luis s/n km 235, São Carlos, 13565-905, SP, Brazil*

ARTICLE INFO

Article history:
Received 21 June 2024
Received in revised form 10 September 2024
Accepted 10 September 2024
Available online 13 September 2024
Handling Editor: Dr Daihai He

Keywords:
Citrus cultivation
Cubic smoothing splines
Environmental factors
Huanglongbing
Random-effects regression models

ABSTRACT

Greening, or Huanglongbing (HLB), poses a severe threat to global citrus cultivation, affecting various citrus species and compromising fruit production. Primarily transmitted by psyllids during phloem feeding, the bacterium *Candidatus* Liberibacter induces detrimental symptoms, including leaf yellowing and reduced fruit quality. Given the limitations of conventional control strategies, the search for innovative approaches, such as resistant genotypes and early diagnostic methods, becomes essential for the sustainability of citrus cultivation. The development of predictive models, such as the one proposed in this study, is essential as it enables the estimation of the bacterium's concentration and the vulnerability of healthy plants to infection, which will be instrumental in determining the risk of HLB. This study proposes a prediction model utilizing environmental factors, including temperature, humidity, and precipitation, which play a decisive role in greening epidemiology, influencing the complex interaction among the pathogen, vector, and host plant. In the proposed modeling, it addresses non-linear relationships through cubic smoothing splines applications and tackles imbalanced categorical predictor variables, requiring the use of a random-effects regression model, incorporating a random intercept to account for variability across different groups and mitigate the risk of biased predictions. The model's ability to predict HLB incidence under varying climatic conditions provides a significant contribution to disease management, offering a strategic tool for early intervention and potentially reducing the spread of HLB. Using climatological and environmental data, the research aims to develop a predictive model, assessing the influence of these variables on the spread of *Candidatus* Liberibacter asiaticus, essential for effective disease management. The proposed flexible model demonstrates robust predictions for both training and test data, identifying climatological and environmental predictors influencing the

☆ This document is the result of the research project funded by the Fundo de Defesa da Citricultura (Fundecitrus) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

☆☆ This study develops a predictive model for the dissemination of *Candidatus* Liberibacter asiaticus, utilizing environmental factors such as temperature, humidity, and precipitation, and addresses non-linear relationships through cubic smoothing splines and random-effects regression models.

* Corresponding author. Departamento de Pesquisa e Desenvolvimento, Fundo de Defesa da Citricultura (Fundecitrus), Av. Dr. Adhemar Pereira de Barros, 201 - Vila Melhado, Araraquara, 14807-040, SP, Brazil.

E-mail addresses: juliocezarvasconcelos@hotmail.com (J.C.S. Vasconcelos), silvio.lopes@fundecitrus.com.br (S.A. Lopes), juan.arenas@fundecitrus.com.br (J.C. Cifuentes Arenas), dmfs@ufscar.br (M.F.G.F. Silva).

Peer review under responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.idm.2024.09.005

dissemination of *Candidatus* Liberibacter asiaticus, the vascular bacterium associated with Huanglongbing (HLB) or greening.

1. Introduction

Phytobacteriosis, known as citrus greening or Huanglongbing (HLB), represents a significant threat to global citriculture, affecting various citrus species and compromising citrus fruit production. It is mainly transmitted by psyllids during feeding in the phloem of plants, and the bacterium *Candidatus* Liberibacter causes detrimental symptoms, such as leaf yellowing and reduced fruit quality. Given the limitations of conventional control strategies, researchers are increasingly focusing on innovative approaches, including resistant genotypes and early diagnostic methods, which are important for citriculture sustainability.

Several authors have conducted research on phytobacteriosis in recent years. For example, Cifuentes-Arenas et al. (2018) investigated the impact of citrus shoot ontogeny on the development of the psyllid species Diaphorina citri, developing a six-stage scale. Their findings suggested that considering shoot ontogeny and flow density could optimize psyllid control strategies, helping to determine multiplication risks in citrus orchards. Bassanezi et al. (2019) conducted a comprehensive analysis aiming to update the characteristics of HLB epidemics, urging farmers and phytosanitary defenders to adopt quarantine measures to prevent HLB entry into unaffected regions. The study emphasizes the importance of regional control strategies. Graham et al. (2020) highlighted the lack of effective HLB control in the USA due to limited tools against the vector Diaphorina citri and the transmission of *Candidatus* Liberibacter asiaticus, advocating for the removal of infected trees and presenting experiences from Florida, Texas, and California.

Moreover, Raiol-Junior et al. (2021) investigated the lateral movement of *Candidatus* Liberibacter asiaticus (Las) in citrus and concluded that the bacterium primarily moves through the phloem, especially in fully girdled plants. Their study also highlighted the importance of protecting new shoots to minimize pathogen transmission, as starch accumulation occurred above the girdling site, possibly due to changes in phloem sap flow. Cifuentes-Arenas et al. (2022) conducted a five-year study on 'Femminello' Sicilian lemons to assess the impact of severe Asian HLB forms, noting that symptoms increased over time, resulting in significant production reductions and an increase in budding. The study called for the removal of diseased trees and preventive control of the Asian citrus psyllid. Most recently, Dutt et al. (2023) evaluated 'Valencia' orange trees grafted with grapefruit rootstock and Swingle interstock, revealing that HLB-tolerant grapefruit interstocks can promote vigorous tree growth and tolerance to susceptible canopies. Ojo et al. (2024) developed an automated, drill-free trunk injection system to treat Huanglongbing (HLB) in citrus trees, improving application efficiency and reducing environmental impact. Chuang et al. (2024) evaluated individual protective covers (IPCs) and found they effectively excluded Asian citrus psyllids and prevented huanglongbing in young citrus trees, though they increased issues with sooty mold and spider mites, reflective mulch and kaolin treatments did not outperform the grower standard.

With mathematical and statistical modeling, researchers seek effective control strategies for this global threat. Taylor et al. (2019) developed methods to estimate the fundamental thermal niche of pests and pathogens, applying these to HLB, and produced global maps indicating areas at higher risk of disease based on temperature. Zhang et al. (2020) developed a mathematical model analyzing HLB transmission between citrus trees and the Asian citrus psyllid, showing that eliminating infected ACP is essential for controlling HLB spread. They found that optimal control strategies were more effective and economical than constant control strategies. Ajene et al. (2020) used species distribution models to predict the current and future distribution of citrus greening pathogens Las in Africa and Laf globally, emphasizing the need for proactive monitoring. Heit et al. (2020) estimated the potential distribution of HLB in South America using climate models, highlighting the importance of integrating results for decision-making. Luo et al. (2021) formulated a disease model considering insecticide-resistant vectors, exploring optimal control strategies. Alves et al. (2022) analyzed HLB presence and absence data in Minas Gerais, Brazil, over 13 years, identifying temperature as inversely related to disease prevalence. Tang et al. (2023) developed an HLB transmission model, calculating the basic reproduction number (R_0) and proposing an optimal control model to minimize control costs.

Environmental factors also play a significant role in understanding greening dynamics. Temperature, relative humidity, and precipitation affect disease spread and epidemiology. Temperature, in particular, is critical for greening incidence and development, as warmer conditions favor psyllid reproduction and accelerate pathogen life cycles. In high temperatures, disease symptoms may intensify, affecting crop metabolism. Relative humidity influences vector ecology, aiding psyllid survival and reproduction, while precipitation affects psyllid population dynamics and disease spread.

In statistical modeling, nonlinear relationships between covariates and response variables present challenges for traditional linear models. To address this complexity, spline smoothing has emerged as a useful strategy for capturing these interactions accurately. Santos et al., (2024) employed a heteroscedastic semiparametric regression model using the Gamma distribution (GA), this model integrated vegetation indices, accumulated precipitation, treatments applied during two years of field experiments, including dosages of P_{2}O_{5} fertilizer and liquid inoculant {\it Bacillus} strains,

among other covariates, to predict sugarcane yield. Another common modeling issue is an imbalanced categorical predictor variable. To handle this, regression models with random effects on the intercept are often employed, allowing the intercept to vary randomly across groups. Vasconcelos, Ortega, Vila, and Cancho (2023) used an extended partially linear regression model to account for bimodal and correlated data. Vasconcelos, Ortega, Kluge, et al. (2023) applied a flexible partially linear regression model with random effects to study lychee fruit respiratory rates under varying temperature levels.

In addition to statistical approaches, machine learning models have also been widely employed in the study of HLB. Deng et al. (2019) developed a non-destructive hyperspectral reflectance method to detect HLB in Shatangju tangerines, using machine learning algorithms to identify optimal bands for early detection. This method was effective in reducing dimensionality and demonstrated the potential of reflectance data for early disease detection. Lan et al. (2020) used UAV-mounted multispectral cameras to detect HLB in large citrus orchards, finding that the combination of vegetative index features and original values yielded high accuracy in disease detection. Finally, Weng et al. (2021) employed machine learning models to study the effects of HLB on citrus leaves, identifying disruptions in photosynthesis and distinguishing HLB symptoms from magnesium deficiencies. Menezes et al. (2023) used hyperspectral imaging and convolutional neural networks to automatically detect HLB disease in citrus trees, achieving 92.86% classification accuracy. This method offers an efficient alternative to traditional, labor-intensive detection techniques. Yang et al. (2024) developed 1D-CNN models for detecting HLB disease in citrus using micro-FTIR spectroscopy, achieving 98.65% accuracy in the lab and 91.21% in the field, based on callose accumulation.

In this context, this study aims to develop a predictive model using climatological and environmental data to evaluate the influence of predictor variables on the spread of *Candidatus* Liberibacter asiaticus. Understanding the role of these factors is vital for developing effective disease management strategies. By integrating climatological, environmental, epidemiological, and agronomic data, this research seeks to enhance disease spread anticipation, improve preventive measures, and optimize HLB control in various cultivation conditions.

The remainder of the article is organized as follows. Section 2 outlines the methods and materials used in the study, including the field experiments conducted to monitor the bacterium *Candidatus* Liberibacter asiaticus, the relevant climatic variables for its epidemiology, and the proposed statistical model, Normal partially linear regression with random effects, for analyzing the collected data. Section 3 details the application of the proposed model, illustrating how it captures the relationship between climatic variables and the quantity of *Candidatus* Liberibacter asiaticus. This section emphasizes the model's accuracy and effectiveness in forecasting and interpreting the data. Finally, Section 4 presents the concluding remarks.

2. Material and methods

2.1. Field trials

The experiments discussed here refer to the repeated measures of log Las (*Candidatus* Liberibacter asiaticus) cell g^{-1} from orange trees at 10 locations in the southeastern region of Brazil (Table 1).

2.2. Climatic variables relevant to Candidatus Liberibacter asiaticus

The meteorological variables considered were temperature, relative humidity, and accumulated precipitation due to their known influence on the ecological dynamics of the pathogen, vector, and citrus plants (Carvalho et al., 2021; Lopes et al., 2017; Zorzenon et al., 2021). This methodological approach aims to provide a more precise and comprehensive understanding of the environmental factors underlying the epidemiology of this bacterium in specific ecosystems. Additionally, the Köppen-Geiger climate classification was incorporated as a “climate type” variable, which condenses temperature and precipitation seasonality patterns and can serve as a generalizing factor to apply the final model to other regions. In this system, the classification is based on specific criteria for each study location, using distinct codes such as Aw, Cfa, and Cwa to represent, respectively, humid tropical climates, humid subtropical climates, and humid subtropical climates with dry winters (Alvares et al., 2013).

Table 1
Locations where the samples were taken for monitoring Las titers in citrus trees.

City	State	Latitude	Longitude
Agudos	São Paulo	−22.4691	−48.9895
Altair	São Paulo	−20.5204	−49.0534
Analândia	São Paulo	−22.1267	−47.6637
Colina	São Paulo	−20.7162	−48.5457
Comendador Gomes	Minas Gerais	−19.7319	−49.0654
Frutal	Minas Gerais	−20.0197	−48.9194
Guaimbê	São Paulo	−21.9108	−49.8975
Guapiaçu	São Paulo	−20.7958	−49.2206
Itaí	São Paulo	−23.414	−49.0927
Rincão	São Paulo	−21.5875	−48.0679

2.3. Model structuring

The Normal, or Gaussian, distribution is widely used in statistical modeling due to its symmetry around the mean. This bell-shaped distribution is essential in various fields, providing a solid understanding of variability and probability in random processes.

Consider $f(y)$ as a probability density function (PDF), defined by:

$$f(y|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] \qquad (1)$$

for $-\infty < y < \infty$, $-\infty < \mu < \infty$, and $\sigma > 0$. The parameters μ and σ represent the mean and standard deviation, respectively.

2.3.1. Normal regression with random effect

In research with unbalanced repeated measures, the inclusion of a random effect in the intercept is essential for dealing with unexplained variations and heterogeneity between experimental units. This method becomes essential, especially in covariates with unbalanced factors, ensuring a robust approach in estimating fixed effects. However, the validity of inferences depends on checking the model assumptions and interpreting considering its specific properties. We adopt the following structure as a guide to ensure accurate and reliable analysis.

Let a sample be partitioned into N groups and let Y_{ij} denote the response variable of the j-th object in the i-th group, where $i = 1, \ldots, N$, and $j = 1, \ldots, n_i$. Note that there are n_i repeated measures within each group.

Consider the scenario where the response variable is associated with fixed covariates $\mathbf{x}_{ij} = (x_{ij1}, \ldots, x_{ijp})^\top$, and all individuals within the same group share common random effects (denoted as w_i). Assuming that each individual's response includes a common random effect and these random effects are unobserved random variables, the model for the parameter μ_{ij} can be expressed as

$$\mu_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + w_i, \qquad (2)$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ represents the vector of unknown parameters, and w_i denotes the random effects for the ith group. Each group in Equation (2) possesses a random effect w_i, which is represented by independent and identically distributed random variables. These random effects are assumed to be independent of the response variable.

When incorporating a random effect into the regression model (2), certain assumptions are made.

- Conditional on w_i, the observations y_{ij} are normally distributed: $y_{ij}|w_i \sim \text{Normal}(\mu_{ij}, \sigma)$, with the conditional probability density function (pdf) given by

$$f(y_{ij}|w_i) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(y_{ij}-\mu_{ij})^2}{2\sigma^2}\right]. \qquad (3)$$

- The random effects w_i are normally distributed: $w_i \sim \text{Normal}(0, \sigma_a^2)$, with the pdf expressed as

$$g(w_i;\sigma_a) = \frac{1}{\sqrt{2\pi}\sigma_a}\exp\left(-\frac{1}{2}\frac{w_i^2}{\sigma_a^2}\right), \qquad -\infty < w_i < \infty. \qquad (4) < x > < \Big/ x >$$

The variances of w_i are given by $\text{Var}(w_i) = \sigma_a^2$.

- **Likelihood Function**

Considering the response variable vector $\mathbf{Y}_i = (\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{in_i})^T$ associated with each group i, the conditional probability function to the random effects for the observations of group i is expressed as follows:

$$L_i(y_{ij}|w_i) = \prod_{j=1}^{n_i} f(y_{ij}|w_i), \qquad (5)$$

where $f(y_{ij}|w_i)$ represents the probability density function defined in Equation (3). Assuming that, conditional on the random intercept w_i, the responses Y_{ij} are independent, group i contributes with:

$$\int L_i(y_{ij}|w_i)\, g(w_i;\sigma_a) dw_i,$$

where $g(\cdot)$ represents the probability density function of the random effects as defined in equation (4), and $L_i(y_{ij}|w_i)$ is expressed by (5).

Since independence is assumed between the groups N, the marginal likelihood function for the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma, \sigma_a)^\top$ reduces to

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f(y_{ij}|w_i)\, g(w_i, \sigma_a) dw_i. \qquad (6)$$

It follows from a dataset of $n = n_1 + \cdots + n_N$ observations, (y_{11}, x_{11}), …, (y_{1n_1}, x_{1n_1}), …, (y_{N_1}, x_{N_1}), …, (y_{Nn_N}, x_{Nn_N}), considering y_{ij} as the response variable and x_{ij} as the vector of predictor variables associated with object j of group i, the logarithm of the marginal likelihood (6) is expressed as:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log\left\{ \int_{-\infty}^{+\infty} \prod_{j=1}^{n_i} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y_{ij}-\mu_{ij})^2}{2\sigma^2} \right] \left[\frac{1}{\sigma_a\sqrt{2\pi}} \exp\left(-\frac{w_i^2}{2\sigma_a^2} \right) \right] dw_i \right\}, \qquad (7) <x> < \Big/ x>$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma, \sigma_a)^T$ is the parameter vector.

The maximum likelihood estimation (MLE) $(\widehat{\boldsymbol{\theta}})$ for the set of unknown parameters can be obtained by maximizing the log-likelihood (7). We use the `gamlss` (Stasinopoulos & Rigby, 2008) library in the R software to calculate the estimate $\widehat{\boldsymbol{\theta}}$.

2.3.2. Normal partially linear regression with random effects

Consider $\mathbf{t}$ as the set of predictor variable with a nonlinear effect on the response variable, where $i = 1, \ldots, N$ and $j = 1, \ldots, n_i$. In this section, we examine the expanded systematic component as follows:

$$\mu_{ij} = x_{ij}^T \boldsymbol{\beta} + w_i + h(t_{ij}), \qquad (8) <x> < \Big/ x>$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ represents the set of unknown parameters, $h(\cdot)$ denotes an arbitrarily unknown smoothing function associated with t_{ij}, and it is emphasized that t_{ij} is a predictor variable that has a nonlinear effect on y_{ij}.

To formulate the penalized log-likelihood, it is essential to specify the smoothing function. In this study, we adopt the cs$(\cdot)$ function for smoothing, as described by Stasinopoulos et al. (2017), which is based on a cubic spline smoothing function, notably *smooth. spline*$(\cdot)$. For a deeper understanding of cubic spline smoothing, refer to Hastie and Tibshirani (1990), Green and Silverman (1993), and Wood (2017).

Let (y_{11}, x_{11}), …, (y_{1n_1}, x_{1n_1}), …, (y_{N1}, x_{N1}), …, (y_{Nn_N}, x_{Nn_N}) be a set of observations, where y_{ij} denotes the response variable and x_{ij} is the vector of explanatory variables. The expression for the penalized log-likelihood for the vector $(\boldsymbol{\theta}^T, \mathbf{h}^T)^T$ of model (8) is:

$$l_p(\boldsymbol{\theta}, \mathbf{h}) = l(\boldsymbol{\theta}) + \frac{\lambda}{2} \mathbf{h}^T \mathbf{K} \mathbf{h}, \qquad (9) <x> < \Big/ x>$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma, \sigma_a)^T$ represents the parameter vector related to the parametric part, $l(\boldsymbol{\theta})$ is defined in equation (7), $\lambda > 0$ represents a scale parameter to control the quality of the fitted curve, $\mathbf{h} = (h(t_1), \ldots, h(t_q))^T$ is the set of smoothed functions, where q indicates the number of distinct and ordered observations of the non-parametrically controlled explanatory variable. The matrix $\mathbf{K} = \mathbf{Q}\mathbf{V}^{-1}\mathbf{Q}^T$ is a positive definite matrix of size $q \times q$, where $\mathbf{Q}$ is a matrix of order $q \times (q-2)$ and $\mathbf{V}$ is a matrix of order $(q-2) \times (q-2)$. For additional information, refer to Green and Silverman (1993). Thus, $\boldsymbol{\theta}$ and $\mathbf{h}$ are calculated by optimizing Equation (9).

2.4. Statistical model of observational information

2.4.1. Objective of the climatological influence experiment

The objective of the experiment was to verify if climatological variables have any influence on the titer of *Candidatus* Liberibacter asiaticus (log Las cell g^{-1}) (y_{ij}).

The values of the predictor variables were obtained during the 10 days prior to the collection of log Las cell g^{-1}, as follows.

- x_{ij1}: number of occurrences of temperatures lower than 18 °C (Celsius);
- x_{ij2}: number of occurrences of temperatures between 18 °C (inclusive) and less than 32 °C (exclusive);
- x_{ij3}: number of occurrences of temperatures greater than or equal to 32 °C (inclusive);
- x_{ij4}: number of occurrences of relative humidity less than or equal to 30%;
- x_{ij5}: number of occurrences of relative humidity greater than or equal to 80%;
- x_{ij6}: accumulated precipitation. Note that the measurements $j = 1, \ldots, n_i$ of the orange trees are associated with the i-th climatic classification i, $i = 1, 2, 3$.

2.4.2. Regression analysis with systematic component

The systematic component of μ_{ij} for the Normal partially linear regression with random effects is:

$$\mu_{ij} = w_i + \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij3} + \beta_4 x_{ij4} + \beta_5 x_{ij5} + h(x_{ij6}) \qquad (10) <x> < \Big/ x>$$

where w_i denotes the random effects, and $h(\cdot)$ is an arbitrarily unknown smooth function associated with the predictor variable x_{ij6} (for $j = 1, \ldots, n_i$ and $i = 1, 2, 3$).

To assess the model performance, the dataset was divided into two different subsets (70% training set and 30% test set). Thus, the proposed model was trained using the training set and subsequently tested using the test set to verify the model's adequacy in predicting the Las cell g^{-1} titre. The total sample size was 118, representing the number of unique observations. The counts within each climatic classification were aggregated over a 10-day period and include all recorded occurrences, not individual samples.

3. Results and discussion

3.1. Descriptive analysis of the complete dataset

The basic statistics for the complete dataset of the response variable log Las cell g^{-1}(y_{ij}), representing the quantity of *Candidatus* Liberibacter asiaticus, indicate an average of 2.0800 units g^{-1} with a symmetrical distribution, indicated by the proximity between the mean (2.0800) and median (2.0980). However, the presence of minimum values (0.0000) and a negative kurtosis (−0.5068) reveal variability and a flatter distribution. This analysis is fundamental to understanding the variability in the concentration of *Candidatus* Liberibacter asiaticus in samples, providing information about its presence, distribution, and behavior.

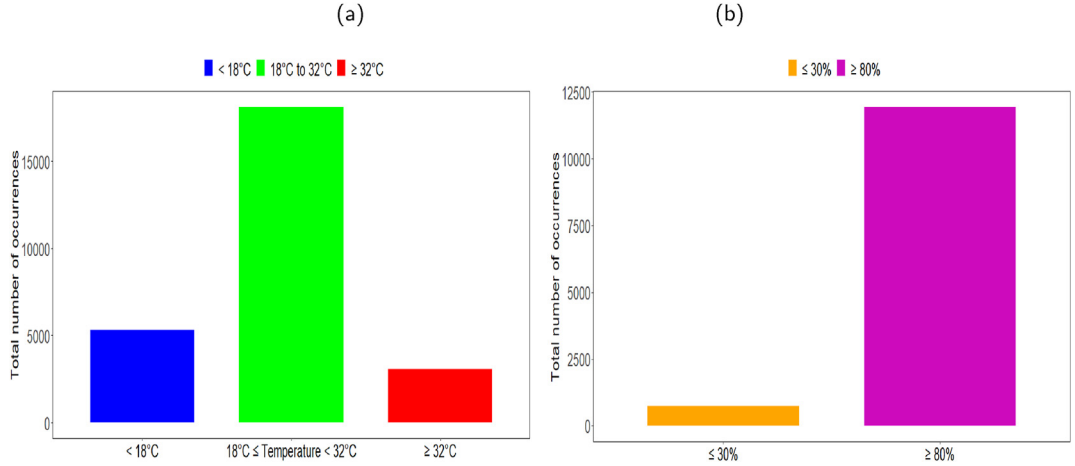

Fig. 1. Bar graph for occurrences of temperatures (a) and relative humidities (b) in different ranges.

The bar graph representing the occurrences of temperatures and relative humidities in different ranges can offer significant information in an exploratory analysis. Fig. 1(a) displays the graphical representation revealing the distribution of temperatures in the 10 days prior to sample collection for the variable "log Las cell g^{-1}" in the southeastern region of Brazil.

It was observed that temperatures below 18 °C occurred approximately 5000 times, indicating a significant incidence of lower temperatures in this period. The range from 18 °C to 32 °C recorded over 15,000 occurrences, evidencing a high frequency of moderate temperature cases. In contrast, temperatures equal to or above 32 °C were observed about 2500 times, pointing to less frequent but still present heat events in the region. Fig. 1(b) bar graph reveals distinct patterns of relative humidity also in the 10 days prior to sample collection for the variable log Las cell g^{-1} in the southeastern region of Brazil. It was observed that in approximately 1000 instances, Relative Humidity was $\leq$30%, indicating low humidity conditions during this period. Conversely, in over 12,500 instances, Relative Humidity was $\geq$80%, suggesting a substantial frequency of high humidity in this time interval.

It is important to note that while this study focused on extreme humidity conditions ($\leq$30% and $\geq$80%), future research will include a more detailed analysis of the intermediate humidity ranges. This will provide a more comprehensive understanding of how varying levels of relative humidity influence the behavior and distribution of Candidatus Liberibacter asiaticus.

Both analyses provide a preliminary view of the relationship between climatic variables (temperature and relative humidity) in the southeastern region of Brazil. The importance of considering these environmental factors in understanding the dynamics of *Candidatus* Liberibacter asiaticus population in citrus plants is highlighted, offering essential perspectives for in-depth climatological studies and the development of management strategies in affected areas.

The descriptive analysis of values over 10 days of the response variable *Candidatus* Liberibacter asiaticus, expressed as log Las cell g^{-1}, in relation to the Köppen-Geiger climatic classification in locations in the southeastern region of Brazil was performed. The climatic classification variable is a factor with three levels: Aw, Cfa, and Cwa, each with different mean and standard deviation (SD) values (Aw: mean = 1.3829, SD = 0.8834; Cfa: mean = 2.8922, SD = 0.9806; Cwa: mean = 2.2227, SD = 1.0137).

It is important to note that these groups have different sample sizes: 52, 40, and 26, respectively. The total sample size is 118, representing the number of unique observations. The number of occurrences refers to the sum of counts within the defined temperature and humidity intervals across all collection points, not the number of individual samples. Each data line reflects aggregated observations over a 10-day period for different locations and climatic classifications. Thus, the total of 25,000 occurrences does not correspond to the number of individual samples but rather the sum of counts recorded within these intervals. Bar graphs are used to represent these distributions to clearly visualize the frequency of occurrences in each interval, aiding in data analysis and interpretation.

From a biological point of view, it is observed that the different levels of climatic classification may reflect different environmental conditions in which the orange trees are inserted.

These conditions, such as temperature, humidity, and climatic classifications, can directly influence the presence and expression of the response variable. Therefore, understanding how variability in climatic classification relates to biological response is paramount for a deeper interpretation of the results.

Considering this biological perspective and the disparity in sample sizes, it becomes even more relevant to evaluate a regression model with random effect on the intercept for the climatic classification variable. This statistical approach, in addition to incorporating the underlying biological complexity, allows for robust handling of sample inequality and incorporates variability between different orange trees. This more realistic and robust modeling contributes to more reliable results and a precise interpretation of the relationships between the response variable and the climatic classification, enriching the biological understanding of the studied phenomenon.

The motivation for employing a partially linear model in this study lies in the clear evidence of a nonlinear relationship between the variable of interest, log Las cell g^{-1}, and the accumulated precipitation collected in the 10 days prior to the collection of the response variable (Fig. 2). The complexity of this relationship suggests that a strictly linear model may not be adequate to capture the full variation in the data. Opting for a partially linear model, a particular case of the semiparametric model, allows for flexibility in modeling, as it admits the presence of both linear and nonlinear components simultaneously. Thus, the partially linear model offers a more comprehensive approach to representing the intricate nature of the association between the variables, enabling a more precise and refined interpretation of the underlying patterns in the data.

3.2. Evaluation and modeling on the training data

3.2.1. Detailed view of the training data

In the more specific analysis of the training data for the response variable log Las cell $g^{-1}(y_{ij})$, an average of 2.193 units g^{-1} is observed. The distribution of the training data maintains an evident symmetry, highlighted by the proximity between mean (2.193) and median (2.205). However, aspects such as the presence of minimum values (0.0000) and a negative kurtosis (−0.5627) indicate variability and a slightly flatter distribution.

In the analysis of the training data in the southeastern region of Brazil, similar patterns to the complete data were identified, as shown in Fig. 3(a). The occurrence of temperatures below 18 °C was approximately 4000 times, between 18 °C

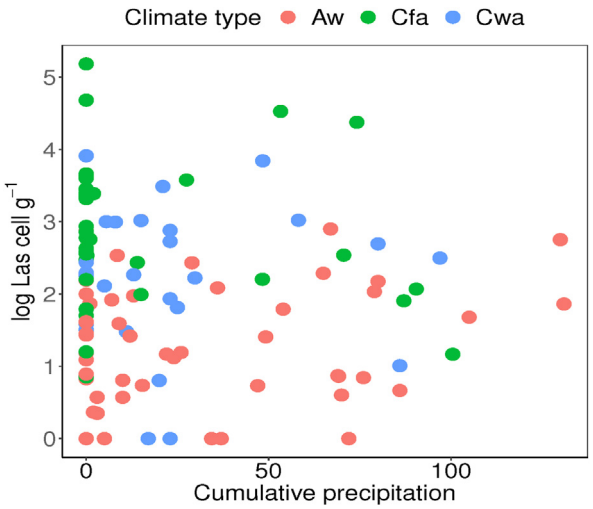

Fig. 2. Scatter plot between log Las cell $g^{-1}(y_{ij})$ and accumulated precipitation (x_{ij6}) in 10 days.

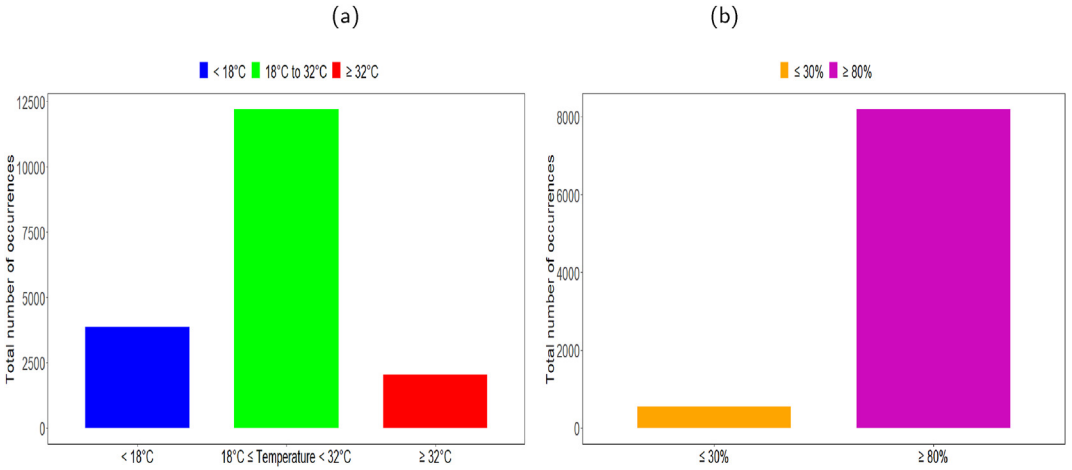

Fig. 3. Occurrences of temperatures (a) and relative humidities (b) in the last 10 days for log Las cell g^{-1} in the southeastern region of Brazil (training data).

and 32 °C about 12,500 times, and temperatures above 32 °C about 2500 times. Regarding relative humidity, Fig. 3(b) shows that low humidities (≤30%) were recorded in about 500 occasions, while high humidities (≥80%) occurred in over 8000 instances. These results indicate consistency in the relationships between climatic variables and the concentration of *Candidatus* Liberibacter asiaticus.

Considering the preliminary analysis of the training data, we observed patterns similar to the complete data regarding the Köppen-Geiger climatic classification in the southeastern region of Brazil. The climatic groups (Aw, Cfa, and Cwa) exhibited distinct means and standard deviations, indicating variations in the biological response of *Candidatus* Liberibacter asiaticus. The Aw group, with 38 observations, had a mean of 1.4352 and standard deviation of 0.9437, while the Cfa group, with 28 observations, recorded a mean of 2.9657 and standard deviation of 0.8807. The Cwa group, composed of 15 observations, showed a mean of 2.6708 and standard deviation of 0.6195. Although similar patterns were observed in the training data compared to the complete data, it is important to highlight specific nuances of the climatic groups.

Focusing on the training data, we observed preliminary evidence of the non-linear relationship between the variable of interest, log Las cell g^{-1}, and the accumulated precipitation in the 10 days prior to the collection of the response variable (Fig. 4). These indications suggest that as the accumulated precipitation increases, there is a tendency for a reduction in log Las. This initial observation reinforces the motivation for using a partially linear model, as the relationship between the variables seems to involve both linear and non-linear components simultaneously.

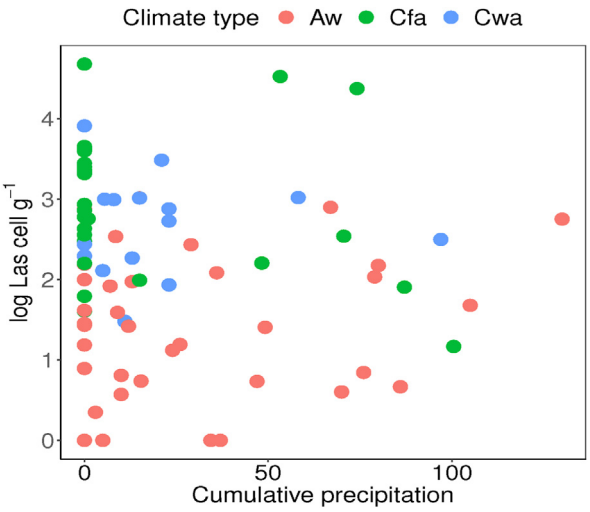

Fig. 4. Scatter plot between log Las cell $g^{-1}(y_{ij})$ and accumulated precipitation (x_{ij6}) during the 10 days prior to collection (training dataset).

3.2.2. Tests for multicollinearity, normality, and homogeneity of variances

Before delving into the analyses, we subjected the dataset to a meticulous evaluation. Initially, we conducted formal tests to assess multicollinearity among the predictor variables, a condition whose presence could compromise the interpretation of the results. Concurrently, we performed thorough tests for normality and homogeneity of variances of the residuals, aiming to validate the fundamental assumptions of the statistical models under consideration.

The multicollinearity tests indicated that the explanatory variables are not related, allowing for the inclusion of all variables in the model. The residuals follow a normal distribution, enabling the use of a partially linear regression model with a random intercept effect. Homogeneity of variances of the residuals was observed, eliminating the need to model the dispersion parameter. These results ensure the robustness of the analyses and the validity of the inferences based on the results.

3.2.3. Modeling results - training data

Understanding the parameters $\boldsymbol{\beta}$ associated with the predictor variables is essential for interpreting the impact of these variables on the response variable y_{ij} (log Las cell g^{-1}) in the model's context. It is worth noting the statistical significance of the predictor variables x_{ij3} (temperature $\geq$32 °C) and x_{ij5} (relative humidity $\geq$80%) (Table 2). An in-depth analysis of these parameters, β_3 and β_5, provides valuable understanding of the substantial impact of these significant variables on the variability of the response variable.

3.2.3.1. Interpretation of β_3. The negative estimate of β_3 suggests that an increase in temperature equal to or greater than 32 °C is associated with a reduction in the levels of the bacterium *Candidatus* Liberibacter asiaticus (log Las cell g^{-1}), indicating a negative influence of x_{ij3} on y_{ij}. The specific value of the coefficient $\beta_3 = -0.0159$ indicates that for each one-unit increase in x_{ij3} (measured in occurrences of temperatures equal to or greater than 32 °C), there is an average decrease of 0.0159 in y_{ij} (log Las cells g^{-1}), highlighting the sensitivity of the bacterium *Candidatus* Liberibacter asiaticus to high temperature variations.

3.2.3.2. Interpretation of β_5. The positive coefficient β_5 suggests that high relative humidities ($\geq$80%) are associated with an increase in the levels of the bacterium *Candidatus* Liberibacter asiaticus (log Las cell g^{-1}). The estimate of the parameter $\widehat{\beta_5} =$

Table 2
Estimates of the parameters, Standard Error (SE), and p-value of the partially linear Normal regression model with random effect adjusted for the training data. The notation (*) denotes the statistical significance of the variables, indicating p-value < 0.05.

Effects	Parameter	Estimate	SE	Value-p
Intercept	β_0	1.7759	0.2893	$<$0.0001
x_{ij1}	β_1	−0.0014	0.0026	0.5869
x_{ij2}	β_2	0.0012	0.0024	0.6237
x_{ij3}	β_3	−0.0159	0.0053	0.0035*
x_{ij4}	β_4	0.0051	0.0088	0.5658
x_{ij5}	β_5	0.0095	0.0029	0.0021*
	$\log(\sigma)$	−0.4341	0.0786	$<$0.0001
	σ_a	0.3956		

0.0095 indicates that for each one-unit increase in x_{ij5} (measured in occurrences of relative humidity equal to or greater than 80%), there is an average increase of 0.0095 in y_{ij} (log Las cells g^{-1}). Therefore, each additional occurrence of relative humidity equal to or above 80% is associated with a corresponding increase in the levels of the bacterium *Candidatus* Liberibacter asiaticus, highlighting that although x_{ij5} has a positive effect on y_{ij}, this increase in the bacterium's levels is undesirable.

Additionally, Table 2 displays the estimate $\sigma_a = 0.3956$ associated with the random effects on the intercept for different climatic classifications (Aw, Cfa, Cwa). The estimate σ_a represents the standard deviation of the random effects on the intercept, indicating the magnitude of the unobserved variation in the mean cell densities of *Candidatus* Liberibacter asiaticus (log Las cell g^{-1}) among the climatic groups. This estimate highlights considerable variability in bacterial means, unexplained by the fixed explanatory variables in the model. Interpretatively, the substantial influence of climatic classification on the average variability of *Candidatus* Liberibacter asiaticus (log Las cell g^{-1}) is evident. The continuation of this variability underscores the practical relevance of unobserved climatic influences on the dynamics of bacterial load, even when considering already known explanatory variables.

As previously mentioned, the predictor variable x_{ij6} (cumulative precipitation) did not show a linear trend with the response variable. Thus, a spline smoothing function was assigned, a statistical technique that offers flexibility in modeling complex relationships between variables. Unlike point estimates, the smoothed curve allows capturing nonlinear patterns, adapting locally to the data, and providing an intuitive visual interpretation through smoothed curves. This approach is particularly valuable when the relationship between variables cannot be adequately represented by simple linear functions. Therefore, due to the lack of a direct interpretation for the numerical estimate associated with x_{ij6}, its interpretation was performed via a graph. In Fig. 5, when examining the smoothed curve in the graph, where the shaded region represents the 95% confidence band for the adjusted smoothed curve, specific nonlinear variations in the relationship between the bacterial load of *Candidatus* Liberibacter asiaticus and x_{ij6} (accumulated rainfall) are noted, with a general decrease in the bacterial load. This decrease is generally evident in the accumulated rainfall ranges between 0 and 30 mm (approximately) and 60–100 mm (approximately). This overall analysis suggests a consistent downward trend in the presence of the bacterium in response to these specific rainfall conditions, indicating a significant influence of climatic conditions on the dynamics of the bacterial load.

In Table 3, the indicators R^2 (coefficient of determination), MAE (Mean Absolute Error), RMSE (Root Mean Square Error), and %RMSE (percentage Root Mean Square Error) that analyze the performance of the partially linear regression model with random effect on the intercept demonstrate its effectiveness. The analysis reveals that the proposed model was able to explain 70% of the variation in the data, as indicated by R^2 (Nagelkerke et al., 1991). Furthermore, both MAE and RMSE, which translate the average error of the predictive model in relation to the original data (observed values), exhibit reduced values. Following the considerations of Li et al. (2013), the accuracy of the model can be categorized as excellent for %RMSE $<$ 10%, good for 10% $<$ %RMSE $<$ 20%, fair for 20% $<$ %RMSE $<$ 30%, and unsatisfactory for %RMSE $\geq$30%. The adjusted model demonstrates a fair accuracy (Table 3), as evidenced by the following indicators: MAE = 0.53, RMSE = 0.65, and %RMSE = 0.29%. These results indicate that the model's projections consistently exhibit low mean and squared errors compared to the actual values, demonstrating an adequate ability of the model to adapt to the observed data.

Fig. 6 presents a scatter plot contrasting the predicted values (green points) by the proposed model with the observed values (black points) in the training data. The notable proximity between these sets, evidenced by the striking overlap of points, strengthens the consistency and accuracy of the predictions in relation to the real data. This congruence is supported by the coefficient of determination (R^2), as discussed earlier.

Fig. 7 displays a worm plot (Buuren & Fredriks, 2001) of quantile residuals to assess the model's adequacy. The absence of vertical or horizontal trends, as well as the lack of concentrated points within the elliptical regions, suggests that the partially

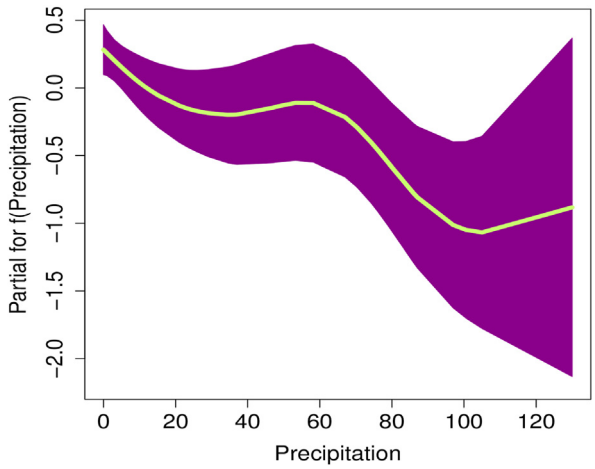

Fig. 5. Smoothed curve analysis of the impact of rainfall on the load of *Candidatus* Liberibacter asiaticus: training data.

Table 3
Evaluation metrics of the partially linear regression model with random effect on the intercept for the training data.

Statistical measures	Performance of the models
R^2	70%
MAE	0.53
RMSE	0.65
% RMSE	0.29%

linear regression model with random intercept is appropriately fitted to the training data. This uniform and seemingly random distribution of residuals enhances confidence in the model's ability to capture and explain the underlying complexities of the data, thus validating its robustness and effectiveness in representing the studied phenomenon.

3.3. Generalization evaluation on test data

3.3.1. Detailed view of test data

When analyzing the test data for the response variable log Las cell $g^{-1}(y_{ij})$, we observe an average of 1.8311 units g^{-1}. The distribution of the test data also maintains an evident symmetry, highlighted by the proximity between the mean (1.8311) and median (1.7899). Aspects such as the presence of minimum values (0.0000) and a positive kurtosis (0.17838) indicate variability and a slightly more concentrated distribution. These additional nuances offer a more refined perspective on the concentration of *Candidatus* Liberibacter asiaticus in the test data, complementing the analyses performed on the complete dataset and the training data.

In the analysis of the test data in the southeastern region of Brazil, we observe in Fig. 8 the occurrence of temperatures below 18 °C in about 1500 instances, between 18 °C and 32 °C in approximately 6000 occurrences, and temperatures equal to or higher than 32 °C in about 1000 instances. Regarding relative humidity, low humidities ($\leq$30%) were recorded in about 4000 instances, while high humidities ($\geq$80%) occurred in over 8000 instances. These results indicate consistent patterns in the relationships between climatic variables and the concentration of *Candidatus* Liberibacter asiaticus in the test data. Compared to previous analyses of the complete dataset and training data in the southeastern region of Brazil, the patterns identified in the test data show consistency in the relationships between climatic variables and the concentration of the response variable, reinforcing the relevance of these factors in understanding the phenomenon under study.

The detailed analysis of the test data for the *Candidatus* Liberibacter asiaticus response variable, considering the Köppen-Geiger climate classification in the southeastern region of Brazil, revealed specific patterns in the climatic groups (Aw, Cfa, and Cwa), showing variations in biological response. The results demonstrate distinct means and standard deviations for each group, enriching the understanding of the relationship between climate classification and the concentration of *Candidatus* Liberibacter asiaticus in orange trees. This analysis complements the perspectives of the complete and training datasets, providing a comprehensive view of the climatic influence on this biological dynamic.

During the analysis of the test data, we identified patterns suggestive of a nonlinear relationship between the variable of interest, log Las cell g^{-1}, and the accumulated precipitation in the 10 days prior to collection. Fig. 9 visualizes these patterns, indicating that as the accumulated precipitation increases, there is a tendency for a reduction in log Las. This observation highlights the complexity in the relationship between the variables and reinforces the choice of a partially linear model, following the same approach adopted in the analysis of the complete and training data. This strategy allows for the incorporation of linear and nonlinear components simultaneously, seeking to more accurately capture the patterns in the interactions between log Las cell g^{-1} and accumulated precipitation in the test data.

3.3.2. Modeling results - test data

The tests confirmed the absence of multicollinearity, the normality of the residuals, and the homogeneity of variances, providing a solid basis for analysis. The modeling results for the test data were consistent with those of the training data, highlighting the influence of the predictor variables on the dynamics of the bacterial load of *Candidatus* Liberibacter asiaticus. This consistency reinforces the robustness of the model and the validity of the conclusions.

In Table 4, the performance indicators of the partially linear regression model applied to the test data reveal reliable results that make the model more robust from the adopted approach. The coefficient of determination R^2 reaches a remarkable level of 71%, highlighting the model's remarkable ability to explain the variation in the response variable. This result not only validates the robustness of the model structure but also highlights its distinctive ability to capture patterns underlying the test data.

When analyzing the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE), essential precision indicators, we find exceptionally low values. The MAE, with an average absolute error of only 0.48, reveals exceptional precision in modeling projections relative to the actual data. Similarly, the RMSE, with a value of 0.66, denotes remarkable precision in assessing the standard deviation of model errors. These metrics not only point to remarkable precision but also highlight the model's unique ability to make highly accurate projections. The %RMSE, a key measure reflecting the percentage of the root

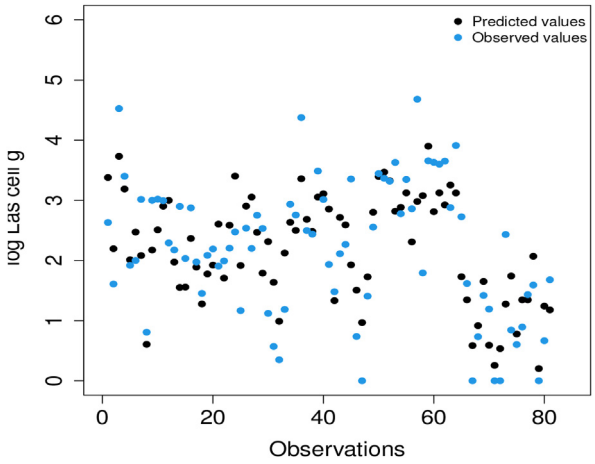

Fig. 6. Scatter plot: comparison between predicted and observed values for the training dataset.

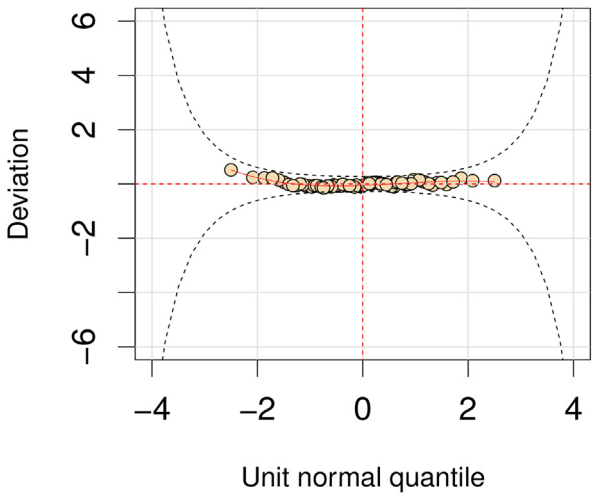

Fig. 7. Worm plot of quantile residuals: assessment of the quality of fit of the partially linear regression model with random intercept for the training data.

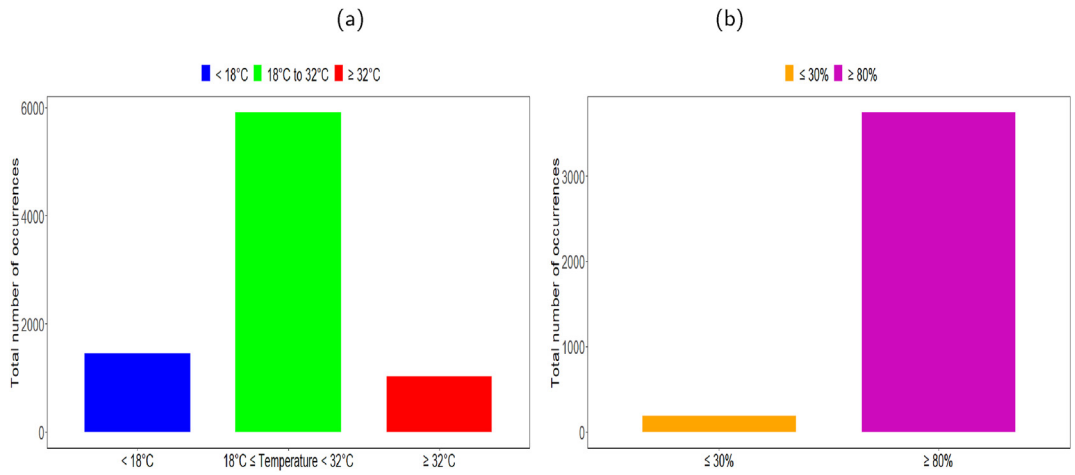

Fig. 8. Bar chart for occurrences of temperatures (a) and relative humidities (b) in different ranges in the test data in the southeastern region of Brazil.

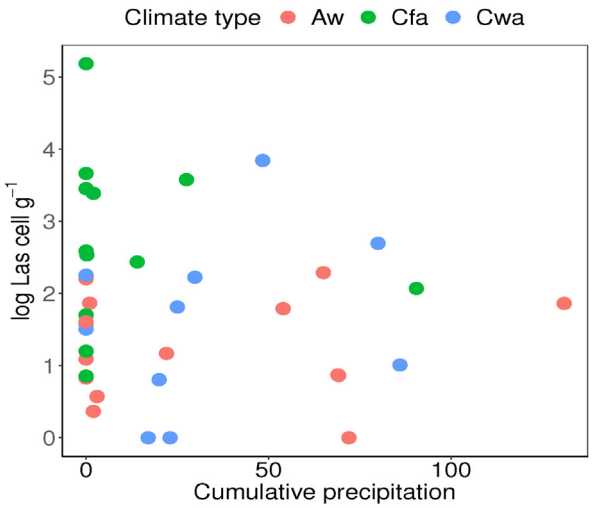

Fig. 9. Relationship between log Las cell g^{-1} and accumulated precipitation in 10 days: test dataset.

Table 4
Metrics for evaluating the partially linear regression model with random intercept effect for test data.

Statistical measures	Performance of the models
R^2	71%
MAE	0.48
RMSE	0.66
% RMSE	0.36%

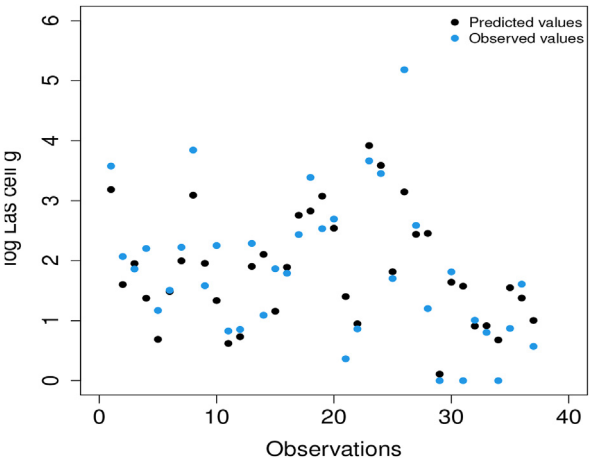

Fig. 10. Scatter plot: comparison between predicted and observed values for the test dataset.

mean square error relative to the mean of the response variable, reveals extraordinarily satisfactory performance, with a value of 0.36%. This result, falling into the “excellent” category according to Li et al. (2013), highlights that the model's projections have consistently negligible mean and squared errors compared to the actual values in the test data.

Fig. 10 highlights the effectiveness of the partially linear regression model with random intercept effect by displaying a scatter plot between the predicted values (green dots) and the observed values (black dots) in the test data. The notable proximity between these points, supported by the high coefficient of determination (R^2), indicates consistency and accuracy in the predictions compared to the actual data.

Fig. 11 displays a worm plot (Buuren & Fredriks, 2001) of the quantile residuals, a useful tool for assessing the adequacy of the model to the test data. The absence of vertical or horizontal trends, as well as the uniform dispersion of the points, suggests that the partially linear regression model with random intercept is well-fitted to the test data. The apparently

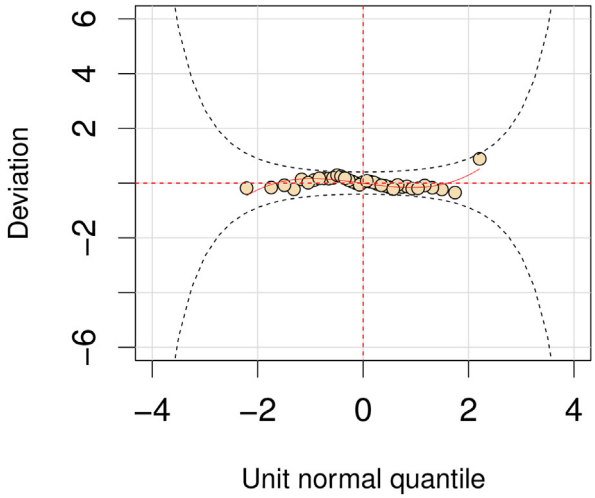

Fig. 11. Worm plot of quantile residuals: evaluation of the fit quality of the partially linear regression model with random intercepts for test data.

random distribution of the residuals enhances confidence in the model's ability to capture and explain the underlying complexities of the data.

These robust results solidify confidence in the predictive ability of the model, demonstrating its exceptional accuracy and adaptability to the test data.

4. Conclusions

The Normal partially linear regression model with random effects has proven effective in predicting *Candidatus* Liberibacter asiaticus bacterial load, explaining around 70% of the variability. This model effectively incorporates climatic variables, such as temperatures above 32 °C and relative humidities above 80%, highlighting their significant impact on bacterial presence. The use of spline smoothing addressed the nonlinear relationship between precipitation and bacterial load, while random intercept accounted for unobserved variability among different climatic groups. The model's approach, considering the Köppen-Geiger climate classification, underscores the importance of including unobserved climatic influences for accurate predictions. The results suggest that future research should expand to diverse climatic regions and contexts, including other major orange-producing countries such as China, the European Union, Mexico, Egypt, the USA, and Turkey, as well as additional regions within Brazil, to enhance model accuracy and generalizability. This model could aid in developing targeted disease control strategies based on specific weather conditions. In summary, addressing the model's limitations and exploring further research are fundamental for advancing our understanding and management of *Candidatus* Liberibacter asiaticus.

Disclosure statement

The authors declare that there are no conflicts of interest.

CRediT authorship contribution statement

Julio Cezar Souza Vasconcelos: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Silvio Aparecido Lopes:** Supervision, Methodology. **Juan Camilo Cifuentes Arenas:** Writing – review & editing, Methodology. **Maria Fátima das Graças Fernandes da Silva:** Project administration.

Declaration of competing interest

I, Julio Cezar Souza Vasconcelos, hereby submit the manuscript entitled “Flexible regression model for predicting the dissemination of *Candidatus* Liberibacter Asiaticus under variable climatic conditions" for consideration for publication in Infectious Disease Modelling. This work has not been published previously and is not under consideration for publication in another journal. All authors have contributed significantly to the work and have agreed to its submission.

We declare that there are no conflicts of interest related to this work. None of the authors have financial or personal relationships that could inappropriately influence the results or interpretations presented in this manuscript.

We appreciate your consideration of our work for publication in Infectious Disease Modelling.

Acknowledgments

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) under project 465357/2014-8 – Instituto Nacional de Ciências e Tecnologia para Controle Biorracional de Inseto Pragas e Fitopatógenos, and Process 133057/2023-2, as part of the INCT 2014 – CHAMADA PÚBLICA MCTI/CNPQ/CAPES/FAPS Nº 16/2014 – PROGRAMA INCT. Additional support was provided by the Fundo de Defesa da Citricultura – Fundecitrus, BRAZIL.

References

Ajene, I. J., Khamis, F., van Asch, B., Pietersen, G., Rasowo, B. A., Ekesi, S., & Mohammed, S. (2020). Habitat suitability and distribution potential of liberibacter species ("candidatus liberibacter asiaticus" and "candidatus liberibacter africanus") associated with citrus greening disease. *Diversity and Distributions, 26*, 575–588.

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J.d. M., Sparovek, G., et al. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift, 22*, 711–728.

Alves, K. S., Rothmann, L. A., & Del Ponte, E. M. (2022). Linking climate variables to large-scale spatial pattern and risk of citrus huanglongbing: A hierarchical bayesian modeling approach. *Phytopathology, 112*, 189–196.

Bassanezi, R. B., Lopes, S. A., Júnior, J. B., Spósito, M. B., Yamamoto, P. T., de Miranda, M. P., do Carmo Teixeira, D., & Wulff, N. A. (2019). Epidemiologia do huanglongbing e suas implicações para o manejo da doença. *Citrus Research & Technology, 31*, 11–23.

Buuren, S.v., & Fredriks, M. (2001). Worm plot: A simple diagnostic device for modelling growth reference curves. *Statistics in Medicine, 20*, 1259–1277.

Carvalho, E. V., Cifuentes-Arenas, J. C., Raiol-Junior, L. L., Stuchi, E. S., Girardi, E. A., & Lopes, S. A. (2021). Modeling seasonal flushing and shoot growth on different citrus scion-rootstock combinations. *Scientia Horticulturae, 288*, Article 110358.

Chuang, A., Kadyampakeni, D., Liesenfelt, T., Vincent, C., Dewdney, M., & Diepenbrock, L. (2024). Comparison of tools to support healthy young citrus plantings in a region with endemic huanglongbing, clas, and asian citrus psyllid (diaphorina citri). *Crop Protection, 185*, Article 106871.

Cifuentes-Arenas, J. C., de Goes, A., de Miranda, M. P., Beattie, G. A. C., & Lopes, S. A. (2018). Citrus flush shoot ontogeny modulates biotic potential of diaphorina citri. *PLoS One, 13*, Article e0190563.

Cifuentes-Arenas, J. C., de Oliveira, H. T., Raiol-Júnior, L. L., de Carvalho, E. V., Kharfan, D., Creste, A. L., Gastaminza, G., Salas, H., Bassanezi, R. B., Ayres, A. J., et al. (2022). Impacts of huanglongbing on fruit yield and quality and on flushing dynamics of Sicilian lemon trees. *Frontiers in Plant Science, 13*, Article 1005557.

Deng, X., Huang, Z., Zheng, Z., Lan, Y., & Dai, F. (2019). Field detection and classification of citrus huanglongbing based on hyperspectral reflectance. *Computers and Electronics in Agriculture, 167*, Article 105006.

Dutt, M., Mahmoud, L. M., & Grosser, J. W. (2023). Field performance of 'valencia'sweet orange trees grafted onto pummelo interstocks and swingle citrumelo rootstocks under huanglongbing (hlb) endemic conditions. *Horticulturae, 9*, 719.

Graham, J., Gottwald, T., & Setamou, M. (2020). Status of huanglongbing (hlb) outbreaks in Florida, California and Texas. *Tropical Plant Pathology, 45*, 265–278.

Green, P. J., & Silverman, B. W. (1993). *Nonparametric regression and generalized linear models: A roughness penalty approach*. Crc Press.

Hastie, T., & Tibshirani, R. (1990). *Generalized additive models london chapman and hall*. Inc.

Heit, G., Sione, W. F., & Aceñolaza, P. G. (2020). A multiple model approach for future potential distribution of hlb: South America case study. *The International Journal of Food and Agricultural Economics, 8*, 179–192.

Lan, Y., Huang, Z., Deng, X., Zhu, Z., Huang, H., Zheng, Z., Lian, B., Zeng, G., & Tong, Z. (2020). Comparison of machine learning methods for citrus greening detection on uav multispectral images. *Computers and Electronics in Agriculture, 171*, Article 105234.

Li, M. F., Tang, X. P., Wu, W., & Liu, H. B. (2013). General models for estimating daily global solar radiation for different solar radiation zones in mainland China. *Energy Conversion and Management, 70*, 139–148.

Lopes, S. A., Luiz, F. Q., Oliveira, H. T., Cifuentes-Arenas, J. C., & Raiol-Junior, L. L. (2017). Seasonal variation of 'candidatus liberibacter asiaticus' titers in new shoots of citrus in distinct environments. *Plant Disease, 101*, 583–590.

Luo, Y., Zhang, F., Liu, Y., & Gao, S. (2021). Analysis and optimal control of a huanglongbing mathematical model with resistant vector. *Infectious Disease Modelling, 6*, 782–804.

Menezes, J., Dharmalingam, R., & Shivakumara, P. (2023). Hlb disease detection in omani lime trees using hyperspectral imaging based techniques. In *International conference on recent trends in image processing and pattern recognition* (pp. 67–81). Springer.

Nagelkerke, N. J., et al. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*, 691–692.

Ojo, I., Ampatzidis, Y., Neto, A.d. O. C., & Batuman, O. (2024). Development of an automated needle-based trunk injection system for hlb-affected citrus trees. *Biosystems Engineering, 240*, 90–99.

Raiol-Junior, L. L., Cifuentes-Arenas, J. C., de Carvalho, E. V., Girardi, E. A., & Lopes, S. A. (2021). Evidence that 'candidatus liberibacter asiaticus' moves predominantly toward new tissue growth in citrus plants. *Plant Disease, 105*, 34–42.

Santos, D. P., Soares, A., de Medeiros, G., Christofoletti, D., Arantes, C. S., Vasconcelos, J. C., … Cançado, G. M. (2024). Evaluation of sugarcane yield response to a phosphate-solubilizing microbial inoculant: Using an aerial imagery-based model. *Sugar Tech, 26*(1), 143–159.

Stasinopoulos, D. M., & Rigby, R. A. (2008). Generalized additive models for location scale and shape (gamlss) in r. *Journal of Statistical Software, 23*, 1–46.

Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V., & De Bastiani, F. (2017). *Flexible regression and smoothing: Using GAMLSS in R*. CRC Press.

Tang, S., Gao, S., Zhang, F., & Liu, Y. (2023). Role of vector resistance and grafting infection in huanglongbing control models. *Infectious Disease Modelling, 8*, 491–513.

Taylor, R. A., Ryan, S. J., Lippi, C. A., Hall, D. G., Narouei-Khandan, H. A., Rohr, J. R., & Johnson, L. R. (2019). Predicting the fundamental thermal niche of crop pests and diseases in a changing world: A case study on citrus greening. *Journal of Applied Ecology, 56*, 2057–2068.

Vasconcelos, J. C., Ortega, E. M., Kluge, R. A., Cordeiro, G. M., & Vila, R. (2023a). A flexible partially linear regression with random effects for bimodal data with application in postharvest. *Communications in Statistics - Simulation and Computation*, 1–22.

Vasconcelos, J. C., Ortega, E. M., Vila, R., & Cancho, V. G. (2023b). An extension of the partially linear rice regression model for bimodal and correlated data. *Brazilian Journal of Probability and Statistics, 37*, 177–194.

Weng, H., Liu, Y., Captoline, I., Li, X., Ye, D., & Wu, R. (2021). Citrus huanglongbing detection based on polyphasic chlorophyll a fluorescence coupled with machine learning and model transfer in two citrus cultivars. *Computers and Electronics in Agriculture, 187*, Article 106289.

Wood, S. N. (2017). *Generalized additive models: An introduction with R*. CRC press.

Yang, B., Yang, Z., Xu, Y., Cheng, W., Zhong, F., Ye, D., & Weng, H. (2024). A 1d-cnn model for the early detection of citrus huanglongbing disease in the sieve plate of phloem tissue using micro-ftir. *Chemometrics and Intelligent Laboratory Systems*, Article 105202.

Zhang, F., Qiu, Z., Zhong, B., Feng, T., & Huang, A. (2020). Modeling citrus huanglongbing transmission within an orchard and its optimal control. *Mathematical Biosciences and Engineering, 17*, 2048–2069.

Zorzenon, F. P., Tomaseto, A. F., Daugherty, M. P., Lopes, J. R., & Miranda, M. P. (2021). Factors associated with diaphorina citri immigration into commercial citrus orchards in são paulo state, Brazil. *Journal of Applied Entomology, 145*, 326–335.