Rice Blast forecasting using interval valued data at Coimbatore, India

ABSTRACT

Aims: The persistence of rice blast, caused by the fungus *Magnaporthe oryzae*, continues to pose a significant threat to rice production worldwide, impacting both yields and food security. The primary goal of this study is to apply interval-valued independent weather data to accurately model the dependent variable of percentage disease incidence.
Study design: In this paper, we present a detailed study on forecasting rice blast outbreaks through the application of Average method, Center method and Min Max method using interval valued weather data and percentage disease incidence.
Place and Duration of Study: The blast disease data include percent disease incidence (PDI) collected at the Paddy Breeding Station (PBS), Tamil Nadu Agricultural University, Coimbatore, from 2018 to 2021.And Weather variables includes the following: Maximum Temperature, Minimum Temperature, Relative humidity (morning), Relative humidity (evening) from 2018 to 2021.
Methodology: The available interval weather parameter data and disease incidence data are utilized to fit a regression model, specifically employing simple linear regression and multiple linear regression, in the R version 4.3.0.
Results: Upon analyzing various methods, it is evident that the variables of Minimum temperature exhibit a significant relationship with a high level of significance, indicating a significance level at $P \leq 0.001$.
Conclusion: Minimum temperature shows more contribution in disease incidence followed by relative humidity at evening.

Keywords: [Rice blast, Interval valued data, Average method, Center method, Min Max method}

1. INTRODUCTION

For millions of households worldwide, rice production is the main occupation and source of income. Among the various biotic stresses such as bacterial infections, viral diseases, nematode infestations, fungal attacks, and other organisms, rice blast stands out as a highly influential factor contributing to the reduction in rice yield. This devastating disease plays a pivotal role in limiting rice production. For the first time Rice blast was seen in Tanjore district of Tamil Nadu in 1918 [1].

Rice blast disease is the most serious rice diseases in rice-growing areas around the world. The fungus *Pyricularia oryzae Sacc*. is the causative agent, and the ideal stage is *Magnaporthe grisea Sacc*. Under favourable environmental conditions, the potential devastating pathogen *Magnaporthe grisea* can lead to significant yield losses [2]. Annually, rice blast causes a reduction in yield ranging from 10% to 30%.[3] It can infect rice at any stage of its life cycle. Early symptoms include white to grey (or brown) leaf spots or lesions on the leaf, followed by nodal rot and neck blast. Change in the climatic conditions can alter the infection rate of the disease. During cold temperatures and heavy moisture conditions the severity of the disease is high, because of dormant conidia (seeds or spores) do not germinate in direct sunshine [4]. A weather-based forecasting system could offer the desired prediction accuracy, as weather significantly influences the occurrence, proliferation, and dissemination of the rice blast fungus[5]. Depending on various environmental conditions, the length of the life cycle might vary. The fungus persists between growing seasons as infected plant debris or as spores (conidia) on the soil surface or in water, which can be disseminated by wind, water, or human activity. The disseminated conidia infect the new rice plants and continues its life cycle. Secondary infections can occur on leaves, stems, panicles, and grains.

Interval data is a kind of quantitative data where values are represented in lower and upper bound. $X = [X_{L,} X_U]$ were X_L=Lower bond X_U=Upper bond.The primary distinction between classic and symbolic data is that a classic data point has a single point in p-dimensional space as its value, whereas a symbolic data point has a hypercube (or hyperrectangle) in p-dimensional space as its value[6].

Interval-valued data appears naturally in numerous situations that involve expressing uncertainty, such as confidence intervals, capturing variability like minimum and maximum daily temperatures, and more. Interval-valued data can be viewed from multiple perspectives.[7]

When forecasting diseases, various meteorological factors are examined, including maximum temperature, minimum temperature, relative humidity in the morning, and relative humidity in the evening. These factors are carefully studied and incorporated into the forecasting models. Predicting disease status can be enhanced through the utilization of meteorological-based modeling, which offers appropriate tools for early forewarning of diseases.[8]

2. MATERIALS & METHODS

2.1. Data collection

This study incorporated disease data representing the percentage disease incidence (PDI) of blast attacks in paddy crop from 2018 to 2021 at the Paddy Breeding Station (PBS), Tamil Nadu Agricultural University, Coimbatore. The weather variables considered as predictor or explanatory variables consist of maximum temperature, minimum temperature, relative humidity (morning), and relative humidity (evening) for the same time period of 2018 to 2021. The daily weather variables are aggregated into standard meteorological weeks using three different methods: Average, Center, and Min-Max method.

2.2. Disease scale

The severity of the disease was monitored every 7 days before crop harvesting. Standard rating scale (0-9) of International Rice Research Institute (IRRI), Manila, Philippines (Anonymous 2002) is given in table 1. The symptoms of blast disease on rice leaves ranges from 0-9 scale are shown in Fig 1.

Table 1: Rice Blast Disease Scale: 0-9

Scale	Infected leaf area
0	There were no lesions seen.
1	Pinpoint sized small brown spots
2	little roundish to slightly elongated 1 -2 mm in diameter
3	Similar to scale 2, but on upper leaves as well
4	under 4%
5	4-10%
6	11-25%
7	26-50%
8	51-75 %
9	above 75%

Disease severity can be calculated by using percentage disease index [9]

$$\text{PDI} = \frac{\text{sumation of numerical ratings}}{\text{Number of leaf observed x highest rating}} \text{x } 100$$

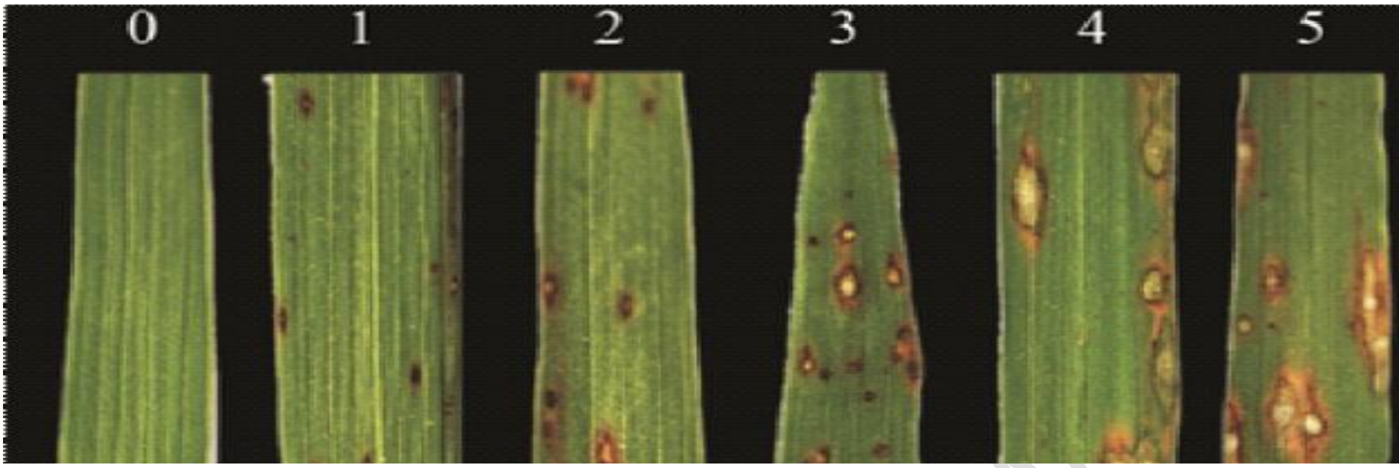


Fig.1 disease scale 0-9 for rice blast [0= There were no lesions seen, 1= Pin point sized small brown spots,2= little roundish to slightly elongated 1 -2 mm in diameter,3= similar to scale 2, but on upper leaves as well,4= under 4%,5=4-10%,6=11-25%,7=26-50%,8=51-75 %,9= above 75%

2.3. Statistical modelling

2.3.1. Classical linear Regression Methods

The classical linear regression model is a statistical approach for modelling the connection between one or more independent variables and a dependent variable. It shows the relationship between independent variables and the dependent variable.

Linear regression model represented as

$$\mathrm{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots\ldots\ldots \beta_P X_P + e \qquad \text{---- Eq 1}$$

Y is a response variable and X_1, X_2, X_3......X_4 are independent / input variable. β_0 is known as intercept.β_1,β_2….β_n are regression coefficients that indicate each independent variable's influence on the dependent variable. The error term or residual indicates the unexplained fluctuation in the dependent variable.

The linear model in matrix form represented as $\mathrm{Y} = \mathrm{X}\beta + \mathrm{e}$

The regression coefficient is calculated as $\hat{\beta} = (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{Y}$

Based on independent and predicted Y values $\hat{\mathrm{Y}} = \mathrm{X}\,\hat{\beta}$

2.3.2. Average method

We have a set of daily weather variables represented by $X_{11}, X_{12}, X_{13,\ldots\ldots\ldots} X_{ij}$.To aggregate these variables into standard meteorological weeks, we will calculate the average of every 7 data points.

Avg $[X_{11}, X_{12,\ldots\ldots\ldots} X_{17}]$

Avg = $\frac{X_{11}+X_{12}+X_{13}\ldots\ldots X_{17}}{7}$

2.3.3. Min Max method

In this method regression model fitted superably for lower & upper values of response & input variables as classical approach method.

$$X_{iL} = MIN[X_{11}, X_{12,\ldots\ldots\ldots} X_{17}]$$
$$X_{iU} = MAX[X_{11}, X_{12,\ldots\ldots\ldots} X_{17}]$$

2.3.4. Center method

In this center method, β parameters are computed using interval's midpoints. There are control variables X_1, X_2…… X_n each variable is in the form of interval data set such as $X_i = [X_{iL}, X_{iU}]$ were X_{iL} is the lower limit and X_{iU} is the upper limit. In case of response variable Y, which is also taken in interval form such that $Y_i = [Y_{iL}, Y_{iU}]$.

Let X_c is the matrix with interval's midpoint such that $\mathrm{X_C} = \frac{\mathrm{X_L}+\mathrm{X_U}}{2}$ and $\mathrm{Y_C} = \frac{\mathrm{Y_L}+\mathrm{Y_U}}{2}$ the midpoints of Y. The main aim of center method is to fit a linear regression of interval data of independent variable to the dependent variables in interval form [10].

$$\mathrm{Y_C} = X_C\beta + C$$
$$\hat{\beta} = (X_C^T \mathrm{X})(X_C^T Y)$$

$$Y_L = X_L\hat{\beta} \qquad Y_U = X_U\hat{\beta}$$

3. RESULTS AND DISCUSSIONS

With the help of percentage disease index (PDI) and weather variables three methods has been employed namely Average method, centre method and min max method. This analysis is performed using R version 4.3.0.

Table 2: Simple linear models for different methods

	Parameters	Model	R^2 %	F value	P Value
Average method	Maximum Temperature	Y= 9.335-0.104 Tmax ‘.’	00.51	00.40	0.52
	Minimum Temperature	Y= 29.865- 1.082 Tmin ‘***’	38.11	48.04	1.07 x 10^{-9}
	RH_m	Y= -26.482+ 0.381 RH_m ‘**’	09.46	08.15	0.006
	RH_e	Y= 13.539 -0.134 RHe ‘***’	19.74	19.19	3.64 x10^{-5}
Center method	Maximum Temperature	Y=8.342-0.073 Tmax	00.27	00.21	0.65
	Minimum Temperature	Y= 27.948-1.001 Tmin ‘***’	35.22	42.40	6.59 x10^{-9}
	RH_m	Y= -12.730+0.220 RH_m ‘*’	05.73	04.74	0.03
	RH_e	Y= 13.132-0.125 RHe ‘***’	18.93	18.22	5.4 x 10^{-5}
Min method	Maximum Temperature	Y=7.071-0.033 Tmax ‘ ’	00.08	00.06	0.81
	Minimum Temperature	Y= 21.137-0.742 Tmin ‘***’	32.86	38.17	2.74 x 10^{-8}
	RH_m	Y= -8.025+0.174 RH_m ‘**’	08.58	7.322	0.01
	RH_e	Y=13.118-0.154 RH_e ‘***’	23.85	24.43	4.32 x 10^{-6}
Max method	Maximum Temperature	Y=9.733-0.112 Tmax ‘.’	00.57	00.44	0.51
	Minimum Temperature	Y= 32.158-1.114 Tmin ‘***’	29.05	31.93	2.50 x 10^{-7}
	RH_m	Y=2.981+0.035 RH_m	00.09	00.07	0.79
	RH_e	Y= 11.441-0.080 RH_e ‘**’	11.78	10.42	0.002

0; ‘***’ Pvalue ≤ 0.001; ‘**’ Pvalue ≤ 0.01; ‘*’ Pvalue ≤ 0.05; ‘.’ Pvalue ≤ 0.1; ‘ ’ Pvalue ≤ 1
(RH_m=Relative humidity morning RH_e= Relative humidity evening)

Table 3: Comparision of Multiple regression model

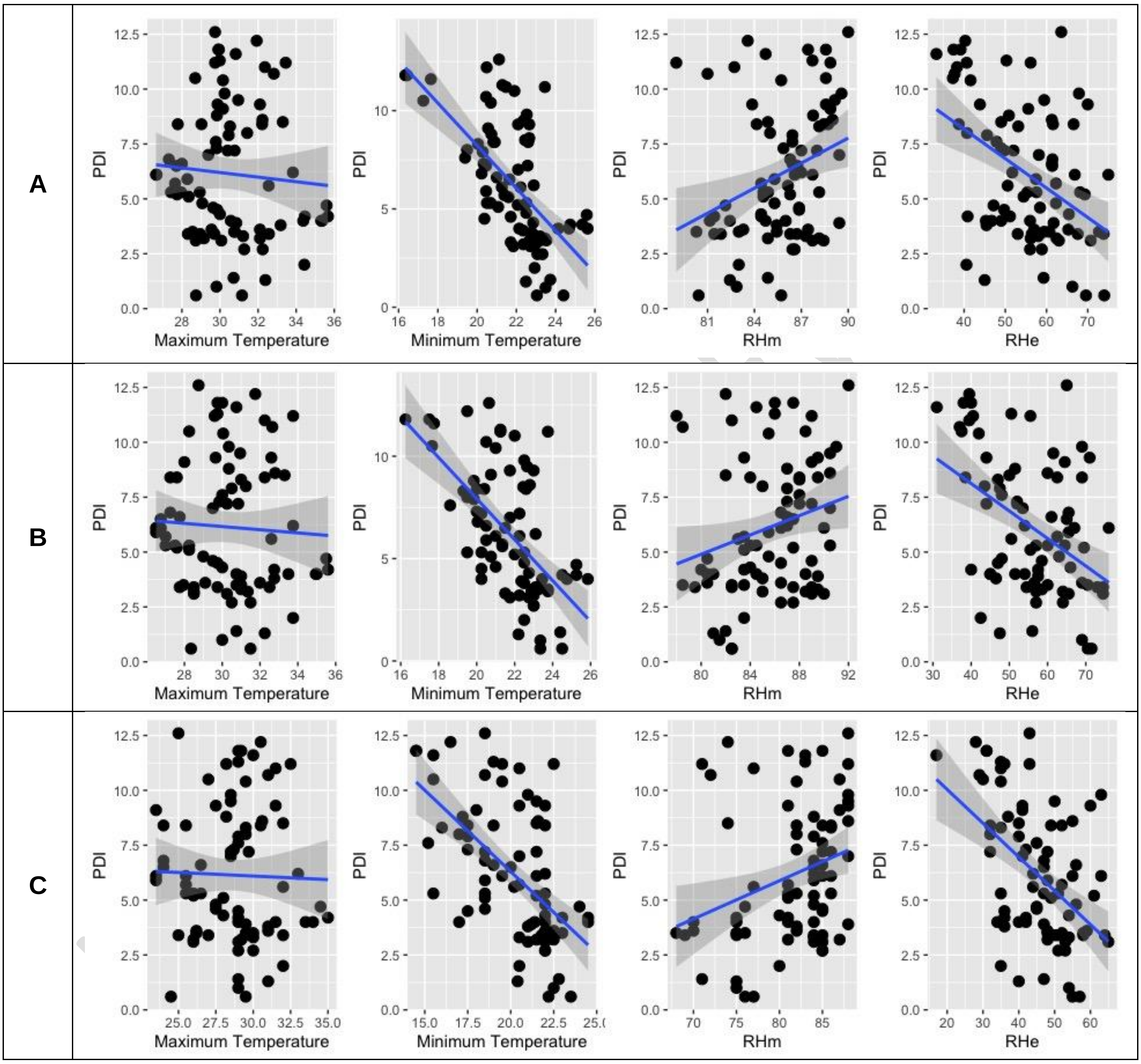
METHODS	Model	R^2 %	F value	P Value
Average Method	Y= 05.759-0.041X_1-0.681X_2‘.’+0.263X_3*-0.108X_4	46.66	16.4	1.08 X 10^{-9}
Center Method	Y= 13.529-0.045X_1-0.670X_2*+0.166X_3‘.’-0.100X_4‘.’	43.23	14.28	1.04 X 10^{-8}
Min-Max Method	Y_L = 10.918-0.035X_1-0.416X_2*+0.114X_3‘.’-0.104X_4*	40.19	12.6	6.85 X 10^{-8}
	Y_U =18.798+0.350X_1-1.315X_2***+0.105X_3-0.039X_4	38.61	11.79	1.75 X 10^{-7}

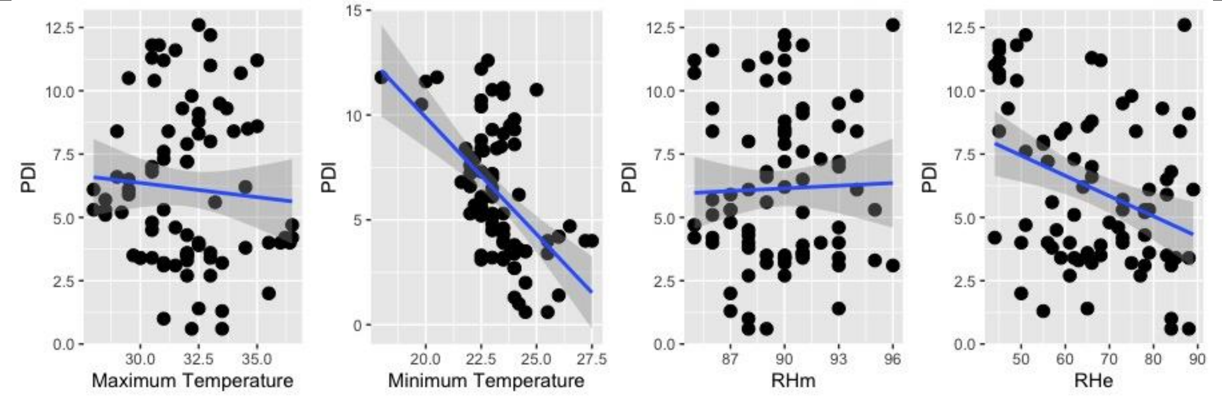
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

X_1= Maximum Temperature X_2= Minimum Temperature X_3=RH(morning) X_4=RH(evening)

Disease development were monitored weakly & recorded on scale from 0 to 9 as shown in Fig1[9].Daily weather data from 2018 to 2022 is collected from Agro Climate Research Center, Coimbatore. These daily weather data is converted to standard metrological weeks using different methods as listed in section 2.3. Linear regression is performed for disease incidence using interval data with different methods in R version 4.3.0. In all the methods, the maximum temperature, minimum temperature, and evening relative humidity exhibited negative relationship (Table 3,Fig 2). On the other hand, the morning relative humidity in the showed a positive correlation The minimum temperature shows strong statistical significance in all analyzed methods, with a p-value of less than or equal to 0.001.

Figure 2: Relationship between disease incidence (PDI) with Weather parameters



D	

A- Average method B-Center method C- Minimum method D-Maximum method
(RHm= Relative humidity morning , RHe =Relative humidity evening)

The line represents the best linear fit to the data points. Points near the line indicate a strong correlation, scattered points suggest a weak correlation. Average method in Fig 2A shows that the data points for the maximum temperature are spread out, indicating a lack of strong correlation with the fitted line. On the other hand, the data points for the minimum temperature are closely clustered around the fitted line, suggesting a significant and tight association between the variables.In Fig 2B, when using the center method, we observe that the data points representing the minimum temperature are closely clustered together, indicating a strong association between the variables. However, for the maximum temperature, the data points appear to be more scattered, suggesting a weaker correlation with the PDI. Additionally, relative humidity at morning and evening also shows a scattered pattern, indicating a less defined relationship with the other variables being analyzed.In fig 2C using Minimum method only relative humidity at morning shows positive correlation remaining weather variables exhibits a negative correlation with PDI.In Fig 2D, when applying the Maximum method, we observe that the data points for relative humidity at morning are widely scattered, and the calculated R-squared value is relatively low. This indicates a weak positive correlation between relative humidity at morning and diseaase incidence and the remaining weather variables shows negative correlation with disease incidence.

From the multiple regression model average method shows highest R^2 value is 46.66%, it means that 46.66% of the variance in the percentage disease index is explained by the independent variables, and the remaining 54% of the variance is still unexplained and it may be due to other factors not included in the analysis. Average method exhibits the best fit model (High R^2 values) followed by centre method (Table 3).

4. Conclusions

The average and centre methods can be used for forecasting the percentage disease incidence of blast for interval value weather data. Blast incidence in rice can be predicted using minimum temperature & evening relative humidity.

5. References

1. Pattanayak, S. and S. Das, *Outbreak of rice blast on the coastal region of South-Eastern India.* International Journal of Agriculture, Environment and Biotechnology, 2020. **13**(1): p. 59-70.
2. Ch'iu, J.-c. *Rice diseases; proceedings*. in *Tao tso ping hai chuan ti yen tao hui (1969: Taipei)*. 1971. Joint Commission on Rural Reconstruction.
3. Sakulkoo, W., et al., *A single fungal MAP kinase controls plant cell-to-cell invasion by the rice blast fungus.* Science, 2018. **359**(6382): p. 1399-1403.
4. Kirtphaiboon, S., et al., *Model of rice blast disease under tropical climate conditions.* Chaos, Solitons & Fractals, 2021. **143**: p. 110530.

5. Malicdem, A.R. and P.L. Fernandez, *Rice blast disease forecasting for northern Philippines.* WSEAS Trans. Inf. Sci. Appl, 2015. **12**: p. 120-129.
6. Garcia-Ascanio, C. and C. Maté, *Electric power demand forecasting using interval time series: A comparison between VAR and iMLP.* Energy policy, 2010. **38**(2): p. 715-725.
7. Maia, A.L.S., F.d.A. de Carvalho, and T.B. Ludermir, *Forecasting models for interval-valued time series.* Neurocomputing, 2008. **71**(16-18): p. 3344-3352.
8. Kumar, A., *Forewarning Models for Alternaria blight in mustard crop.* 2013.
9. SAHA, M., *Forecasting of rice blast disease severity in West Bengal, India based on PDI values and Cumulative logit model.*
10. Chacón, J.E. and O. Rodríguez, *Regression models for symbolic interval-valued variables.* Entropy, 2021. **23**(4): p. 429.