Department of Statistics
University of Sri Jayewardenepura

STA 474 2.0 Statistical Consultancy
Project 02

Mapping and Analyzing the Distribution and Severity of Meloidogyne graminicola in Rice

Submitted by
AS2021382 – Sayuri Ponnamperuma
AS2021450 – Sulakshi Wijayarathne
AS2021492 – Deshan Theekshane
AS2021485 – Thisuri Jayarathna
AS2021607 – Thisuri Athukorala

August 2025

Table of Contents

1 Introduction

Rice (*Oryza sativa*) plays a central role in the diets and economies of millions of people, particularly in South and Southeast Asia. In Sri Lanka, rice is the main staple food and is cultivated across diverse agro-ecological zones, contributing significantly to national food security and rural livelihoods. However, rice productivity is increasingly challenged by a range of pests and diseases. Among these, the rice root-knot nematode, *Meloidogyne graminicola* (MG), has emerged as a serious biotic stress factor in recent years.

Meloidogyne graminicola is a soil-borne, plant-parasitic nematode that penetrates rice roots and induces the formation of characteristic hook-shaped galls. These galls disrupt normal root function, severely impairing the plant's ability to absorb water and nutrients. As a result, infected plants exhibit stunted growth, yellowing of leaves, reduced tillering, delayed flowering, and significant yield reduction.

Figure 1: Galls on rice roots infected by Meloidogyne graminicola.
Source: (Desaeger, 2023)

Despite its growing significance, systematic studies on the distribution, prevalence, and severity of *MG* in Sri Lanka remain scarce. Furthermore, there is limited field-level evidence linking nematode infestation with agronomic factors (e.g., planting method, rice variety, farmer awareness) and environmental drivers (e.g., rainfall, topography, soil type). This knowledge gap hinders the development of integrated pest management strategies tailored to Sri Lanka's diverse rice ecosystems.

Therefore, understanding how *MG* is distributed across agro-ecological zones, and identifying the key climatic, environmental, and management-related factors influencing its spread and severity, is crucial. Such insights are essential not only for designing effective control measures but also for supporting sustainable rice production and guiding future rice breeding programs towards nematode-resistant varieties.

1.1 Background of Study

This study was carried out as an observational field survey during the "Maha" season of 2024. From each district, rice plant samples (n = 30 per field) were collected. The severity of nematode infestation was assessed by counting the number of galls on each plant. Along with this, important contextual information was recorded at the time of sampling, including plant age, topography of the field, irrigation method, soil type, and rice variety**.**

Geospatial data tools, particularly the ArcGIS mapping platform, were employed to retrieve site specific data such as agro-ecological region, dominant soil type, and terrain characteristics**,** based on GPS coordinates of each sampled location. These mapped layers support the spatial analysis of nematode distribution and help in identifying regional risk patterns.

1.2 Objectives

- To map the regions of pest distribution and severity of pest in the respective areas.
- To identify factors that contribute to the spread of the pest.

2 Methodology

2.1 Data Collection

Field surveys were conducted during the *Maha* season of 2024 to assess the prevalence and severity of Meloidogyne graminicola infestations across diverse rice-growing regions of Sri Lanka. A total of 52 sampling sites were selected to represent a range of climatic zones (Dry, Intermediate, and Wet), agro-ecological zones, soil types, and rice ecosystems. At each site, 30 individual rice plants were randomly sampled, yielding 1,560 plant-level observations.

For each plant, the number of root galls was recorded as a direct indicator of nematode infestation severity. In addition to gall counts, extensive contextual and environmental data were collected through field observation, farmer interviews, and geospatial tools. Information such as rice variety, planting method, flooding status, growth stage, and farmer awareness was documented at the time of sampling.

Spatial attributes were retrieved using GPS coordinates recorded in the field and subsequently linked with ArcGIS-mapped layers. Climatic information, including annual rainfall, was obtained from regional meteorological records.

The study included both biological and contextual variables, capturing plant-level infestation severity, agronomic practices, environmental conditions, and farmer-level knowledge. A detailed description of the dataset variables is provided in Table 1.

Table 1: Description of variables

Variable Name	Description
serial_number	Unique identifier assigned to each sampling site where data were collected.
rep_no	Replicate number (1 to 30) representing individual plant samples taken per site for gall count assessment.
Gall_count	Number of root galls observed per plant caused by *Meloidogyne graminicola*, indicating the severity of nematode infestation.
Province	Administrative province in which the sampling site is located.
District	District within the province where the sampling site is located.
Climatic_zone	Classification of the sampling site based on climate: "Dry", "Intermediate", or "Wet" zone.
Topography	Physical landform at the site: "Flat land", "Lowland", "Gentle slope", or "Terraces".
Terrain	Surface features of the land (slope, elevation, relief) influencing soil and drainage.
Major_soil	Dominant soil type at the site, which influences nematode survival, crop health, and water retention.
annualRmm	Annual rainfall at the site in millimeters (mm)
GPS_location	Latitude and longitude coordinates of the sampling site.
x	X-coordinate of the site (longitude or projected coordinate system).
y	Y-coordinate of the site (latitude or projected coordinate system).
Rice_variety	Name or code of the rice variety grown at the site.
Date_of_maturity	Date or stage when the rice crop reached maturity and was ready for harvest.
Flooded (Y/N)	Whether the field was flooded during the season: "Y" (Yes) or "N" (No).
Method_of_planting	Planting method: "Direct seeded" or "Replanting" (transplanted).
Irrigation	Type/source of irrigation: "Rainfed", "Minor irrigation", "Major irrigation", or specific tanks (e.g., "Koshena wewa").
Rice_ecosystem	Rice cultivation system: "Rainfed lowland", "Irrigated", or "Rainfed Midland".

agro_eco_z	Agro-Ecological Zone: broad classification based on climate, soil, and elevation.
agro_eco_r	Agro-Ecological Region: finer-scale classification within agro-ecological zones, based on terrain and microclimate.
Age_DAS	Age of rice plants in Days After Sowing (DAS).
Stage_of_maturity	Growth stage of rice crop based on Age_DAS: "Seedling" (0–21 DAS), "Vegetative" (22–59 DAS), "Reproductive" (60–95 DAS), "Ripening" (96+ DAS).
Prediction_on_susceptibility_of_rice variety	Predicted/assessed susceptibility of the rice variety to *M. graminicola*: "Susceptible", "Unconfirmed"
Is_the_farmer_aware_of_Mg_before	Farmer's awareness of *M. graminicola* before survey: "Yes", "No"

2.2 Data Preprocessing

The original dataset consisted of 1,560 plant-level observations collected from 46 sampling sites. However, some rows contained missing values, and these were removed to maintain data consistency, reducing the dataset to 1,380 complete observations.

After cleaning for missing data, unique values of each variable were inspected. During this process, several categorical variables were found to have inconsistent entries representing the same category. For example, in the *rice variety* variable, "BG 374" and "bg 374" were entered as separate categories, while in the *stage of maturity* variable, both "Ripeneing" and "Ripening" were present. Such inconsistencies were corrected by standardizing category names to avoid misclassification.

The terrain variable initially had nine unique levels, some with very few observations (e.g., "Very steep, hilly & Rolling" or "Mountanious, steeply dissected, hilly & rolling"). These were consolidated into broader categories to ensure sufficient sample sizes for analysis. Low-frequency categories were combined into a new level named "Other".

Geographical data were also carefully checked. It was observed that in one sampling site, the latitude and longitude values were mistakenly interchanged. This error caused the GPS point to be plotted overseas when visualized on a world map. The values were corrected to accurately reflect the true field location.

In this study, two variables of primary interest were derived from the plant-level response variable, gall count. Each sampling site contained 30 individual rice plants, and data from these plants were aggregated to the site level to facilitate analysis.

- Disease incidence was defined as the proportion of plants affected by the nematode at a given sampling site. A plant was considered infected if its gall count was greater than zero. Disease incidence for each site was calculated as:

$$\text{Disease incidence} = \frac{\text{Number of infected plants at the site}}{\text{Total number of plants per site}}$$

- Disease Severity represented the average intensity of infestation at the sampling site. It was calculated as the total gall count across all 30 plants divided by the total number of plants, providing a measure of the nematode's impact per plant at the site

$$\text{Disease Severity} = \frac{\text{Total gall count at the site}}{\text{Total number of plants per site}}$$

2.3 Data Analysis

2.3.1 Data Exploration

For qualitative variables with a limited number of categories, bar charts were utilized to visually display the distribution of data. For variables with a larger number of categories, frequency tables presenting percentages were preferred, as they offer a clearer, more concise summary of the data distribution. For quantitative variables, summary statistics were computed, including the five-number summary (minimum, first quartile, median, third quartile, and maximum), along with the mean, to provide a comprehensive overview of the data's central tendency and spread.

To explore associations between a nominal and a categorical variable, boxplots were used to visually examine the distribution of the categorical variable across different categories of the nominal variable. When the number of categories was large, bar plots of the medians were employed instead, since only a few data points fell within some categories. This approach highlighted differences in both central tendency and variability across groups. In addition, summary statistics were computed for each category to complement the visual analysis. For associations between two quantitative variables, scatterplots were generated to capture trends, patterns, and potential relationships.

In addition to these analyses, the spatial distribution of MG incidence and disease severity was examined across sampling sites using point distribution maps, enabling visualization of localized patterns. Furthermore, area-based maps were produced to illustrate MG disease incidence and severity across agro-ecological regions, offering insights into broader regional variations in disease dynamics.

All these exploratory analyses were carried out using R statistical software, which provided both flexibility and reproducibility in data visualization and summary computations.

2.3.2 Model Fitting and Evaluation

Model fitting and evaluation were carried out using SAS (Statistical Analysis System) software. In this analysis, the dependent variable was Gall Count, while the predictor variables included agro ecological region, climatic zone, agro ecological zone, terrain, rice variety, stage of maturity, soil type, rice ecosystem, topography, flooded, method of planting, farmer awareness of MG before, age (DAS), and annual rainfall.

Initially, a full negative binomial regression model was fitted using all predictors. Examination of the model summary revealed several large standard errors and unstable coefficient estimates, suggesting the presence of multicollinearity and redundancy among some predictors. Certain variables, specifically terrain, climatic zone, annual rainfall, and agroecological zone, were removed due to perfect collinearity among some categories, which resulted in aliased coefficients. After this adjustment, a revised model was fitted; however, rice variety, stage of maturity, and rice ecosystem still exhibited aliased coefficients and relatively high standard errors.

To address these issues, the frequencies of categories within the stage of maturity and rice ecosystem were examined. Low-frequency levels were combined with the closest higher-frequency categories. This adjustment resolved the problem for rice ecosystem, but aliasing persisted for the stage of maturity. Therefore, that variable was subsequently removed from the model. For rice variety and agro ecological region, categories with fewer than 50 observations were combined into a new category labeled "Other." Refitting the negative binomial model with these adjustments resolved the previous issues of aliasing and high standard errors. Finally, multicollinearity among the predictors was assessed using the variance inflation factor (VIF) values to ensure the model was suitable for inference.

For model fitting in SAS, data handling and preparation were performed using the DATA step and PROC SQL to recode variables and combine low-frequency categories into broader groups. Frequency counts were obtained through PROC FREQ to guide category lumping. The negative binomial regression models were estimated using PROC GENMOD, specifying the DIST=NEGBIN distribution and the log link function. Multicollinearity among predictors was assessed using PROC GLMSELECT (to build a design matrix) in combination with PROC REG to calculate variance inflation factors (VIF). Generalized variance inflation factors (GVIF) were also derived from PROC GLM and its OUTSTAT option.

Model adequacy was primarily evaluated using statistics reported by PROC GENMOD. Specifically, the Deviance/DF and Pearson Chi-Square/DF ratios were examined, where values close to 1 indicate a well-fitting model, values much larger than 1 suggest overdispersion, and values much smaller than 1 suggest underdispersion. Predictor significance was tested through Type III likelihood-ratio chi-square tests, also produced by PROC GENMOD, which assessed the collective contribution of each predictor while adjusting for others in the model. In addition, model comparison was supported by information criteria (AIC and BIC), where lower values indicate a relatively better fit among competing models.

3 Results

3.1 Exploratory Data Analysis

3.1.1 Composition of sample with respect to Province

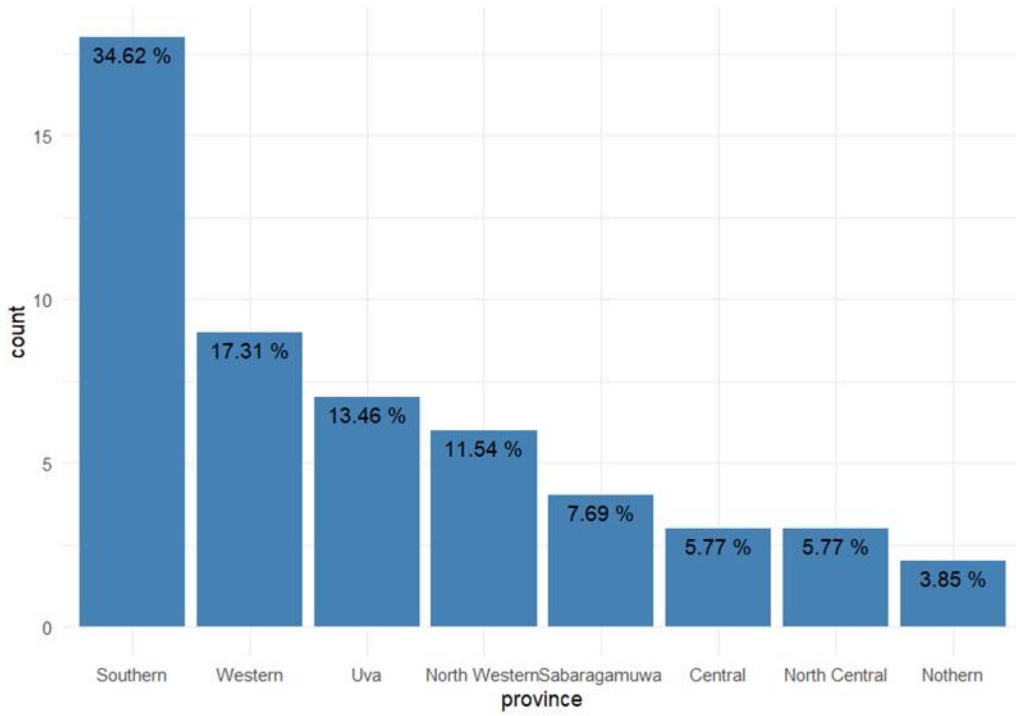

Figure 2: Composition of sampling sites with respect to province

The Southern Province represents the largest share of the sampling sites with 18 out of 52 respondents (34.6%), accounting for over one-third of the total. It is followed by the Western Province with 9 sampling sites (17.3%).

3.1.2 Composition of sampling sites with respect to district

Table 2: frequency table of sampling sites to district

District	count	Percentage
Hambantota	7	13.5%
Matara	6	11.5%
Moneragala	6	11.5%
Galle	5	9.62%
Kalutara	5	9.62%
Kurunegala	5	9.62%
Ratnapura	4	7.69%
Anuradhapura	3	5.77%

Gampaha	3	5.77%
Kandy	2	3.85%
Vavuniya	2	3.85%
Badulla	1	1.92%
Colombo	1	1.92%
Matale	1	1.92%
Puttlam	1	1.92%

Hambantota had the highest representation (13.46%) among all districts, followed by Matara and Moneragala (11.54% each), while several districts like Badulla, Colombo, and Puttlam had minimal representation (1.92%).

3.1.3 Composition of sampling sites with respect to Agro ecological region

Table 3: Frequency table of sampling sites with respect to Agro ecological region

Agroecological region	count	Percentage %
DL1b	10	19.5
IL1a	6	11.5
WL1b	6	11.5
WL2a	6	11.5
DL5	4	7.69
IL1c	4	7.69
IL1b	3	5.77
WL1a	3	5.77
WL3	3	5.77
DL1a	2	3.85
WM3b	2	3.85
IM1c	1	1.92
WM1a	1	1.92
WM2b	1	1.92

The agro-ecological region DL1b had the highest proportion of sampling sites followed by IL1a and WL1b, both at 11.5%.

3.1.4 Composition of sample with respect to Climatic Zone

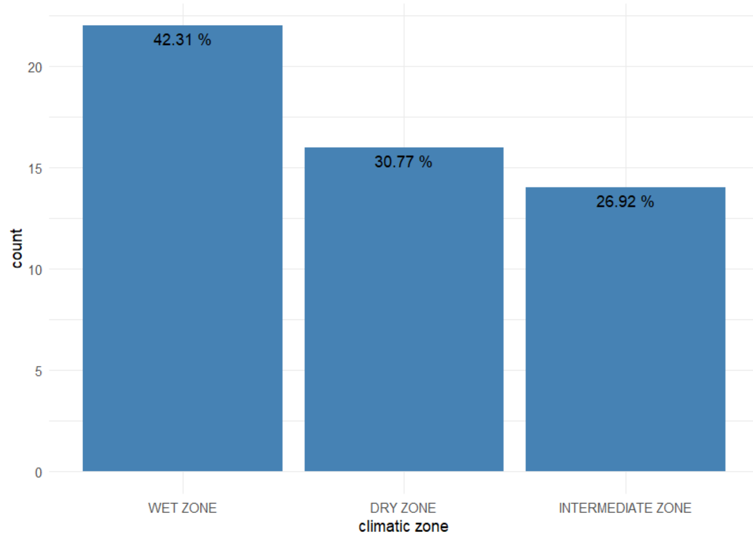

Figure 3: Composition of sampling sites with respect to Climatic Zone

The figure 3 shows that the highest proportion of sampling sites is in the Wet Zone, followed by the Dry Zone, with the fewest sites in the Intermediate Zone. This suggests that wetter regions were more commonly sampled in the study.

3.1.5 Composition of sample with respect to Agro Ecological zone

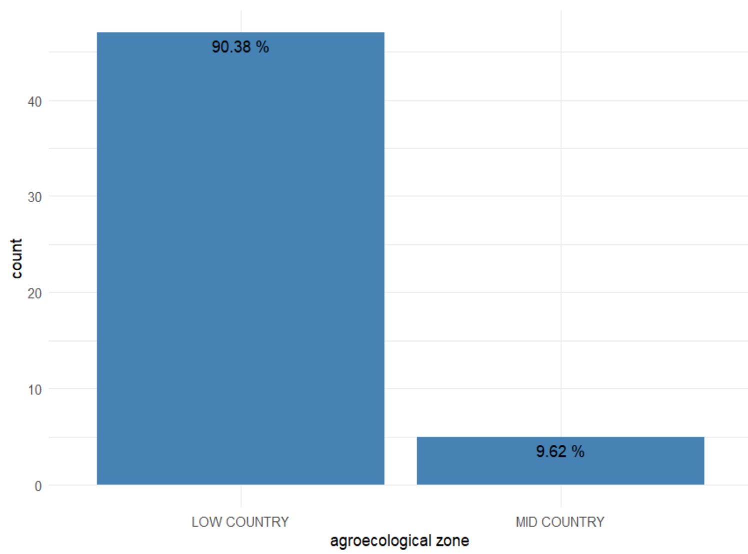

Figure 4: Composition of Sampling sites with respect to agro-ecological zone

The figure 4 shows that most of the sites are located in the Low Country agro-ecological zone, while only a small proportion are from the Mid Country. This indicates that the study was predominantly focused on lowland areas.

3.1.6 Composition of sample sites with respect to Terrain

Table 4: Frequency table of Terrain

terrain	count	Percentage %
Rolling, Undulating, Flat	19	36.5
Rolling & Undulating	11	21.2
Undulating	10	19.2
Undulating & flat	4	7.69
Rolling, undulating & hilly	3	5.77
Hilly, Rolling,undulating & steep	2	3.85
Mountainous, steeply dissected, hilly & rolling	1	1.92
Steeply dissected, hilly & rolling	1	1.92
very steep, hilly & Rolling	1	1.92

More than one-third of the sites considered in this study is located on rolling, undulating, flat terrain while fewer sites were selected from hilly terrain areas.

3.1.7 Composition of sample with respect to Date of Maturity

Table 5: Summary Measures of Date of Maturity

min	1st quantile	median	mean	3rd quantile	max
89	105	110	107.9	110	132

The date of maturity ranged from 89 to 132 days, with a median and 3rd quartile both at 110 days, suggesting most rice varieties in the sample mature within a relatively narrow window.

3.1.8 Composition of the sample sites with respect to Rice variety

Table 6: Frequency Table of Sampling sites with respect to Rice variety

Rice variety	count	Percentage %
At 362	23	44.2
Bg 352	3	5.77
Bg 300	2	3.85
Bg 358	2	3.85
Bg 366	2	3.85
Bg 374	2	3.85
Bg 379-2	2	3.85
Bg 450	2	3.85
Bg 94/1	2	3.85
Bw 367	2	3.85
Bg 304	1	1.92
Bg 357	1	1.92
Bg 360	1	1.92
Bg 369	1	1.92
Bg 41	1	1.92
Ld 368	1	1.92
MD 270	1	1.92
Rathkanda	1	1.92
Suwadel	1	1.92
NA	1	1.92

The most commonly cultivated variety was At 362 (44.23%), with several other varieties such as Bg 352, Bg 300, and Bg 379-2 showing smaller distribution which implies varietal diversity but dominance by one type.

3.1.9 Composition of sampling sites with respect to stage of maturity

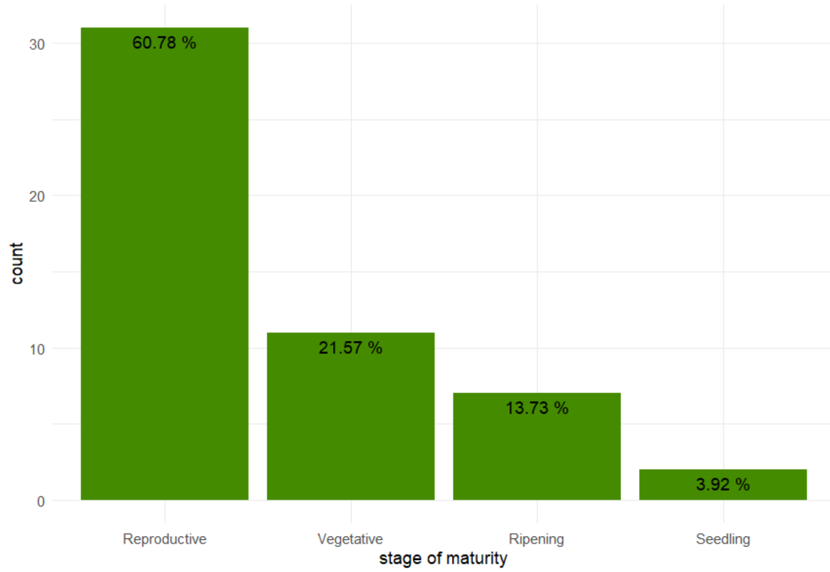

Figure 5: Composition of sampling sites with respect to stage of maturity

The date of maturity ranged from 89 to 132 days, with a median and 3rd quartile both at 110 days, suggesting most rice varieties in the sample mature within a relatively narrow window.

3.1.10 Composition of sampling sites with respect to prediction of susceptibility

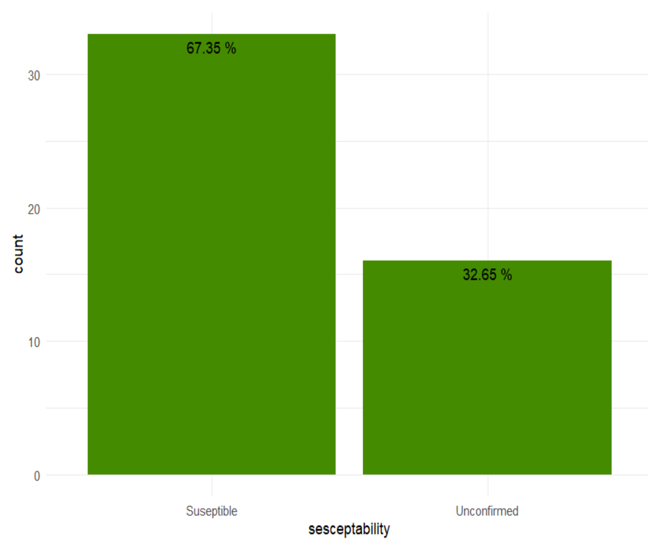

Figure 6: Composition of sampling sites with respect to prediction of susceptibility

The majority of sampling sites were associated with susceptible varieties, representing nearly twothirds of the total. In contrast, unconfirmed sites accounted for less than one-third, indicating that known susceptibility was much more common across the observed locations.

3.1.11 Composition of the sample sites with respect to soil type

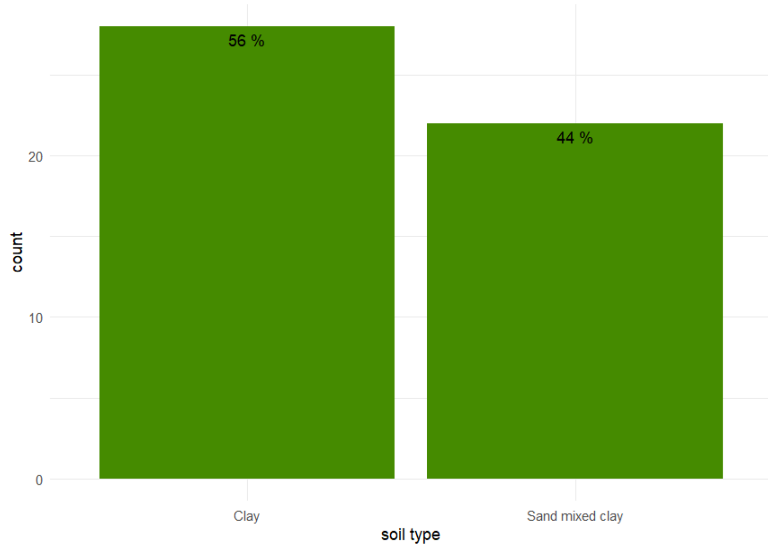

Figure 7: Composition of sampling sites with respect to soil type

Clay soil was the most common among the sampling sites, accounting for just over half of the total. Sand mixed clay soil followed, representing slightly less than half.

3.1.12 Composition of sample sites with respect to rice ecosystem

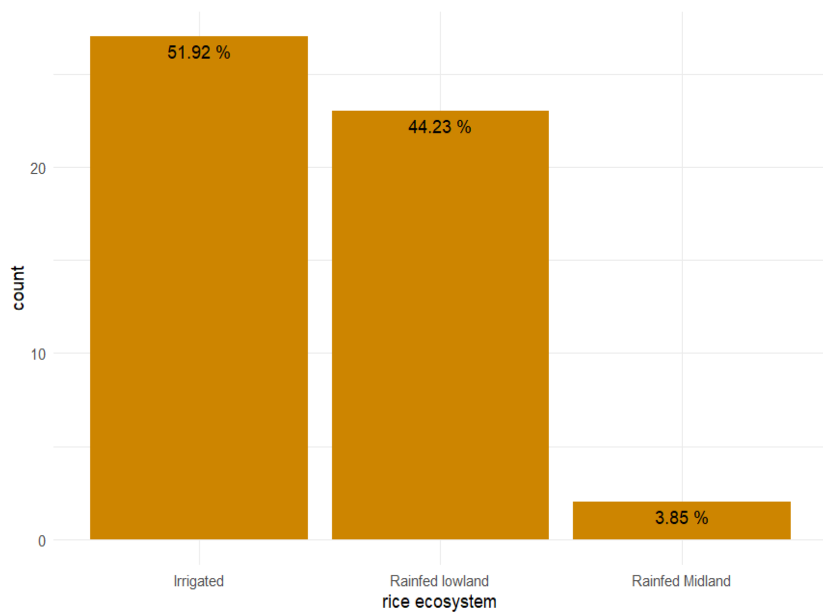

Figure 8: Composition of sampling sites with respect to rice ecosystem

The majority of the sites in the sample belong to the irrigated rice ecosystem and followed by rainfed lowland. There is only small number of sample sites representing the rainfed upland.

3.1.13 Composition of sample sites with respect to irrigation type

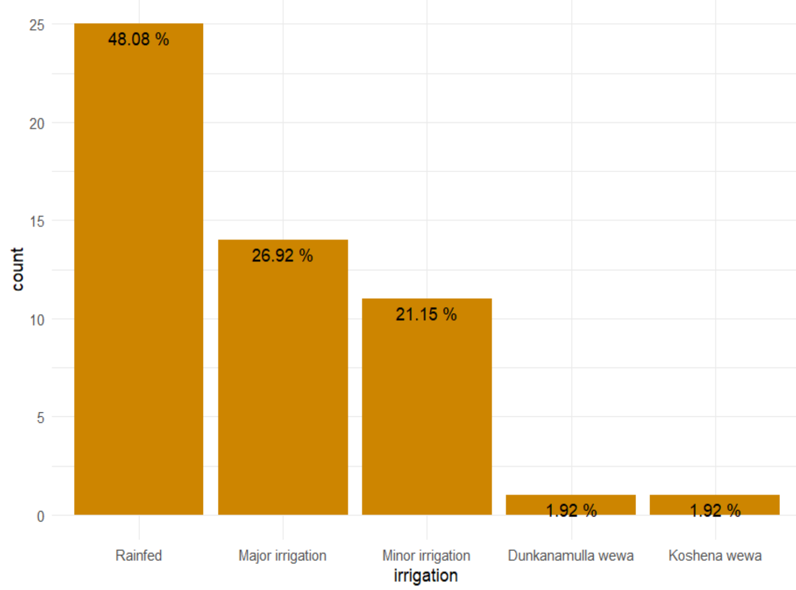

Figure 9: Composition of Sampling sites with respect to Irrigation type

Nearly half of the sampling sites in the study are rainfed. Only about 4% of the sites are irrigated by tanks, while major and minor irrigation systems each account for approximately 20% of the sites.

3.1.14 Composition of sampling sites with respect to rice topography

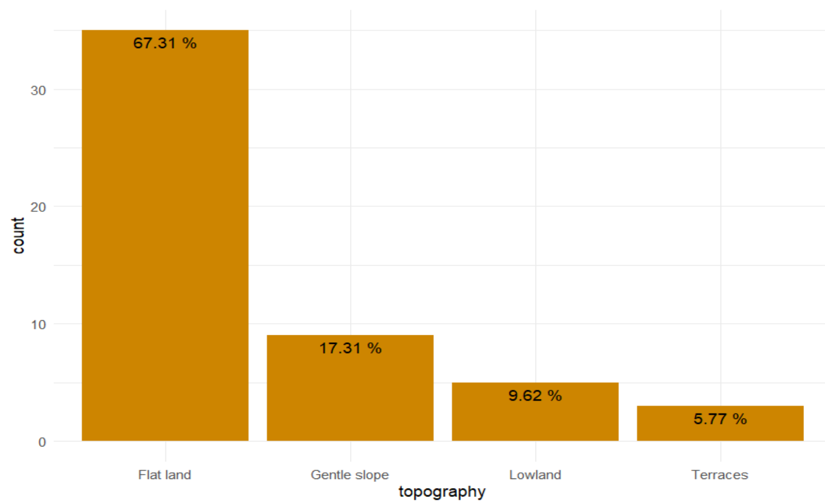

Figure 10: Composition of Sampling sites with respect to Topography

According to figure 10, more than two-thirds of the sample sites are located on flat lands, while approximately 10% are situated in lowland areas. Additionally, around 17% of the sites are found in gently sloping terrain, which may offer moderate drainage and be suitable for certain types of land use and cultivation.

3.1.15 Composition of sampling sites with respect to flooded status

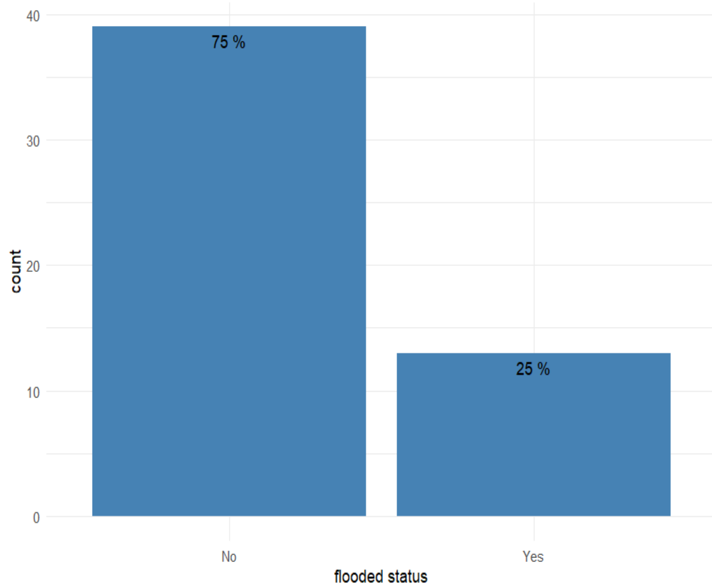

Figure 11: Composition of sample with respect to flood status

According to figure 11, 75% of the sample sites have been affected by floods, while only one-quarter of the sites have never experienced flooding.

3.1.16 Composition of sample with respect to method of planting

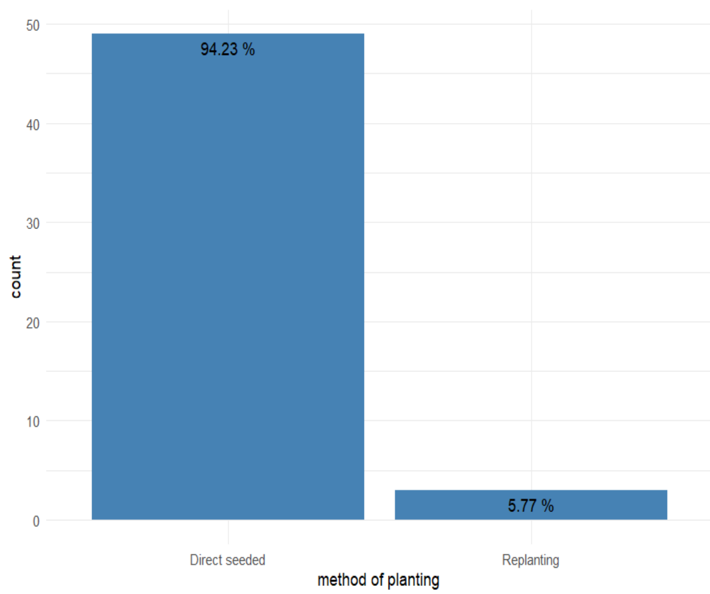

Figure 12: Composition of Sampling sites with respect to method of planting

Direct seeding was employed in approximately 95% of the sample sites, indicating that the majority of plants were established initially without the need for additional intervention. In contrast, only about 5% of the sites required replanting.

3.1.17 Composition of sample with respect to farmers awareness of MG

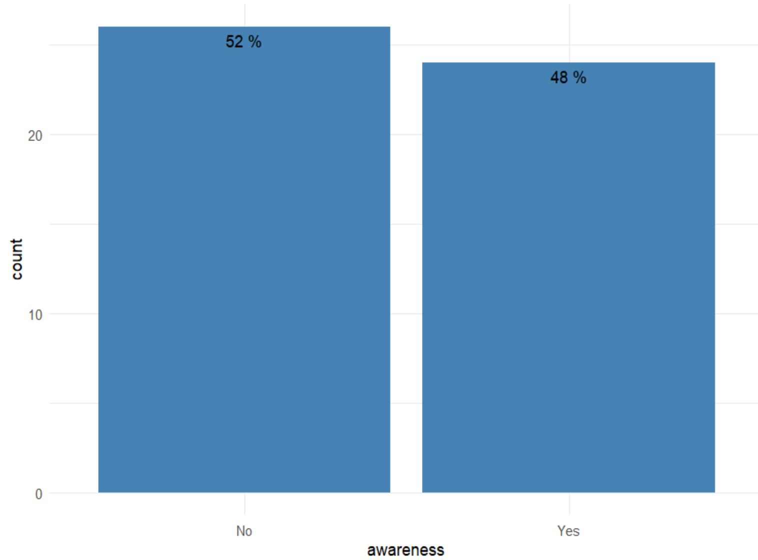

Figure 13: Composition of Sample with respect to Farmers awareness of MG

Figure 13 indicates that about half of the farmers are aware of gall midge (Mg), The other half, however, are not familiar with this, which may affect their ability to implement effective management strategies and protect their crops from potential damage.

3.1.18 Distribution of Annual Rainfall

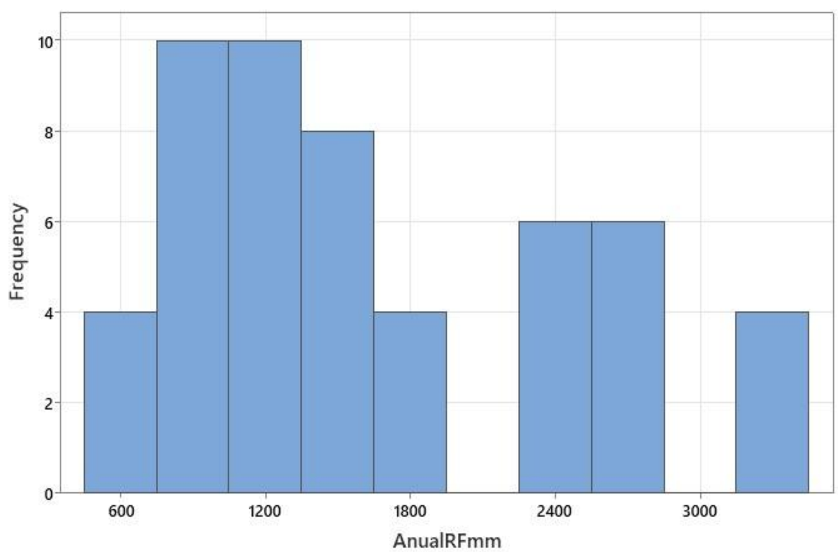

Figure 14: Distribution of Annual Rainfall (mm)

Table 7: Summary measures

minimum	1st quantile	median	mean	3rd quantile	maximum
650	900	1400	1650	2400	3300

Annual rainfall ranged widely from 650 mm to 3300 mm, with a median of 1400 mm, indicating that the sample spans both low and high rainfall areas.

3.1.19 Distribution of Age DAS of rice plants

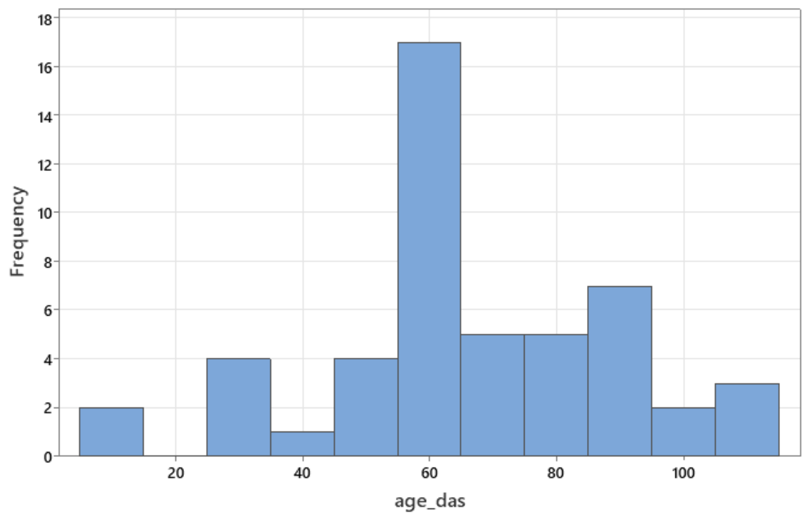

Figure 15: Distribution of Age DAS

Table 8: Summary Measures of Age DAS

min	1st quantile	median	mean	3rd quantile	max	Missing values
12	59	60	65.96	83.5	108	1

The ages of the rice plants, measured in days after sowing (DAS), varied widely, ranging from 12 to 108 days. The median age was 60 days, indicating that half of the plants were younger than 60 days while the other half were older. This broad age distribution reflects diverse growth stages within the sample population.

3.1.20 Distribution of Disease Incidence

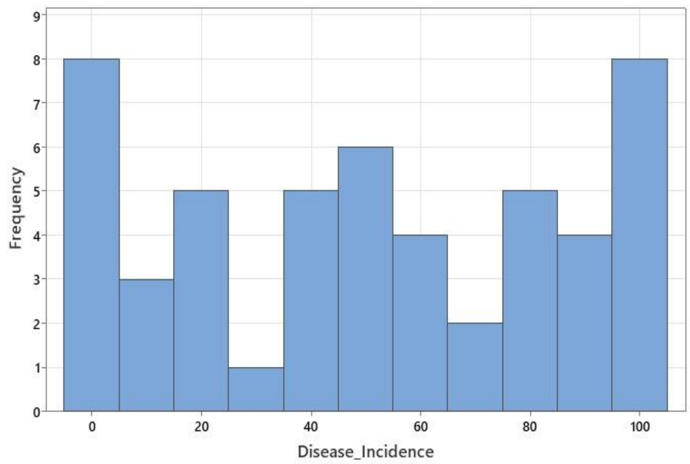

Figure 16: Distribution of Disease Incidence

Table 9: Summary measures of Disease Incidence

Variable	Mean	StDev	Minimum	Q1	Median	Q3	Max
Disease_Incidence	51.0458	35.1599	0	16.6667	50	83.3333	100

The distribution of disease incidence across sampling sites shows substantial variability and a bimodal/irregular pattern. The histogram indicates that many sampling sites report either very low (near 0%) or very high (close to 100%) disease incidence, with fewer sites showing intermediate values.

3.1.21 Distribution of Disease Severity

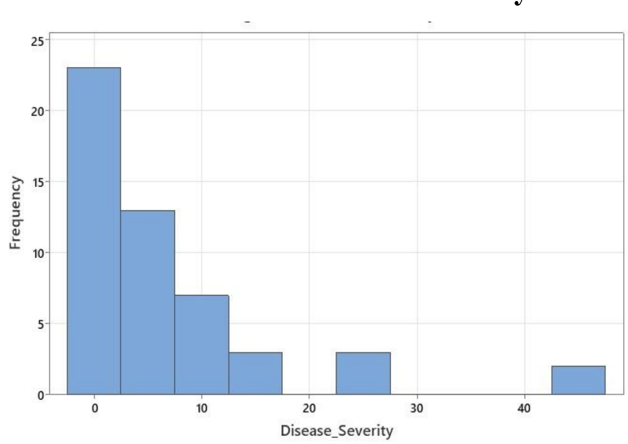

Figure 17: Distribution of Disease Severity

Table 10: Summary Measures of Disease Severity

Variable	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Disease_Severity	6.83137	10.0795	0	0.6	3.13333	8.53333	47.4667

The distribution of disease severity across sampling sites is positive skewed with median 3. Most sampling sites report low disease severity, concentrated between 0 and 10, with over half below 3.2

3.1.22 Gall count

Table 11: Summary Measures of Gall count

min	1st quantile	median	mean	3rd quantile	max
0	18	99.5	204.6	256	1424

Gall counts varied greatly among the rice plants, ranging from 0 to as high as 1,424. The median gall count was 99, indicating that half of the plants had fewer than 99 galls, while the other half had more. This wide range suggests significant variability in gall formation across the sample. The distribution of gall counts mirrors the distribution of disease severity because disease severity is calculated by dividing the gall count by 30. As a result, both metrics show the same pattern.

3.2 Explore Predictor Variables with Disease Incidence

3.2.1 Distribution of Disease Incidence with respect to Province

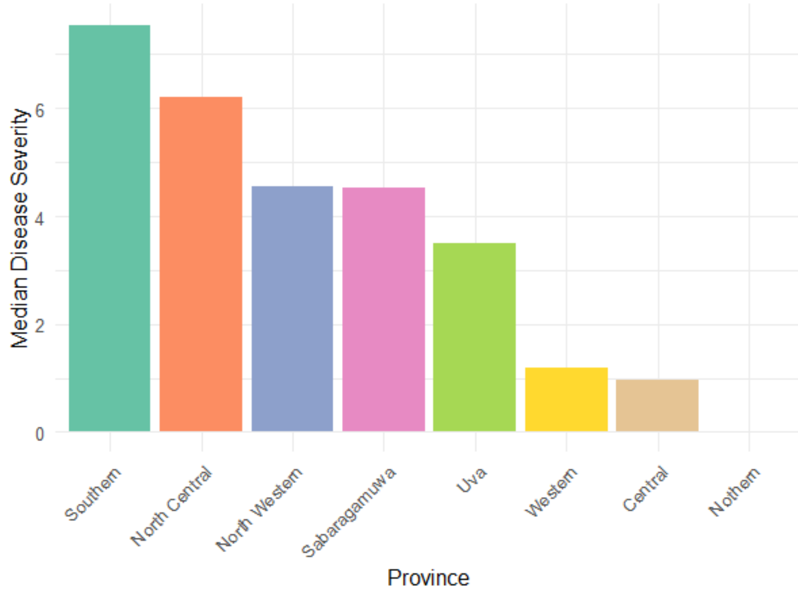

Figure 18: Distribution of Disease Incidence with respect to Province

Disease incidence is highest in North Central and Southern provinces. Other provinces such as Uva, Sabaragamuwa, and North Western show lower incidence rates. Northern province has no disease incidence reported. Variation inside provinces is moderate but noticeable, indicating non-uniformity in disease spread.

3.2.2 Distribution of Disease Incidence with respect to District

The highest median disease incidence is noted in Hambantota, followed by Anuradhapura and Matara. Puttlam and Vavuniya report zero incidence. Variation is generally low in Hambantota but higher in Matara and Anuradhapura, indicating more heterogeneity in disease incidence according to figure 18.

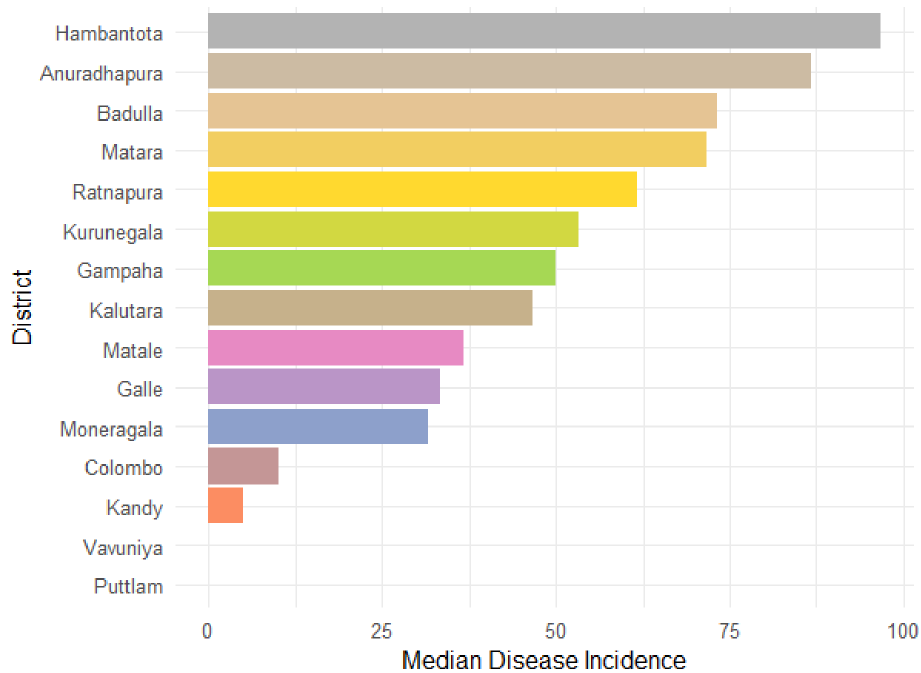

Figure 19: Distribution of Median Disease Incidence with respect to District

The highest median disease incidence is noted in Hambantota, followed by Anuradhapura and Matara. Puttlam and Vavuniya report zero incidence. Variation is generally low in Hambantota but higher in Matara and Anuradhapura, indicating more heterogeneity in disease incidence.

3.2.3 Distribution of Disease Incidence with respect to Climatic Zone

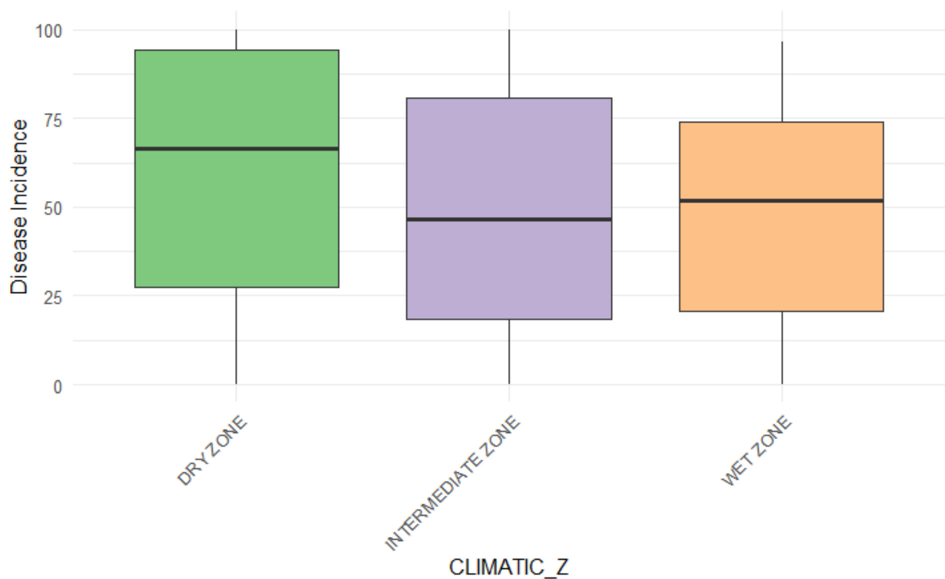

Figure 20: Distribution of Disease Incidence with respect to Climatic Zone

Table 12: Summary Measures

CLIMATIC ZONE	n	Mean	Median	Min	Q1	Q3	Max	SD	IQR
Dry zone	16	58.5	66.7	0.0	27.5	94.2	100	41.4	66.7
Intermediate zone	14	50.0	46.7	0.0	18.3	80.8	100	35.6	62.5
Wet zone	22	47.4	51.7	0.0	20.8	74.2	96.7	30.1	53.3

The dry zone shows the highest median disease incidence followed by the Intermediate Zone and Wet Zone. This suggests that drier climatic conditions might favour higher disease incidence. The variation is substantial in all zones, with the dry zone having the widest spread.

3.2.4 Distribution of Disease Incidence with respect to Agro Ecological Zone

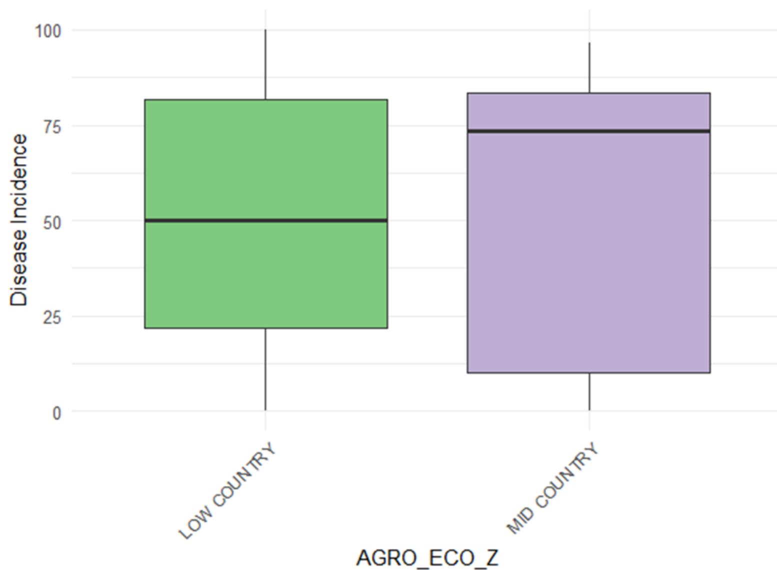

Figure 21: Distribution of Disease Incidence with respect to Climatic Zone

Table 13: Summary Measures

AGRO ECO ZONE	n	Mean	Median	Min	Q1	Q3	Max	SD	IQR
Mid country	5	52.7	73.3	0.0	10.0	83.3	96.7	44.4	73.3
Low country	47	51.4	50.0	0.0	21.7	81.7	100	34.4	60.0

Median disease incidence is higher in mid country compared to low country, suggesting that mid-country agro-ecological conditions may favor higher disease frequency. Variation is considerable in mid country, implying non-uniform incidence across different parts of this zone. Low country has a slightly lower and less variable incidence.

3.2.5 Distribution of Disease Incidence with respect to Terrain

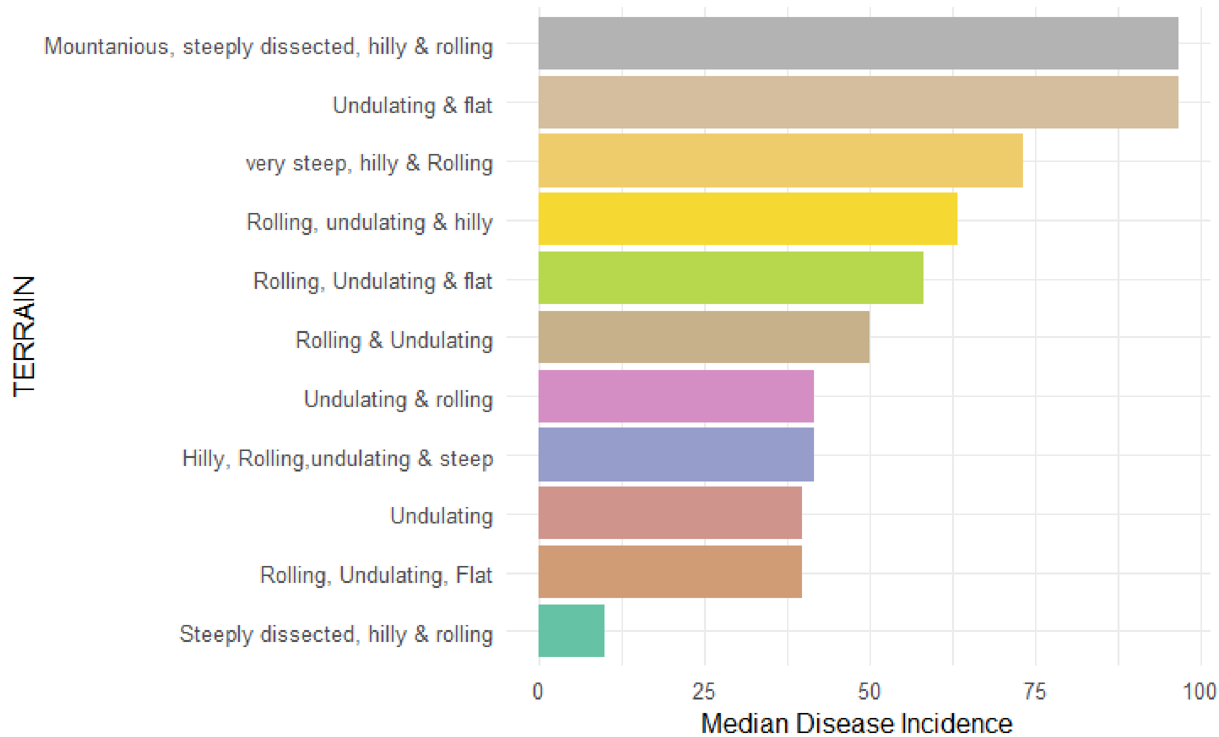

Figure 22: Distribution of Median Disease Incidence with respect to Terrain

Median disease incidence is highest in Mountainous & undulating terrains, including mountainous valleys. Other terrains such as Rolling, undulating & flat and Rolling, undulating & hilly also show high incidence. Lower incidence and less variation are observed in Undulating & mountainous terrain, indicating some terrains may not support disease spread. Considerable variation is evident in rolling undulating terrains, pointing to heterogeneous incidence across these topographies.

3.2.6 Distribution of Disease Incidence with respect to Soil Type

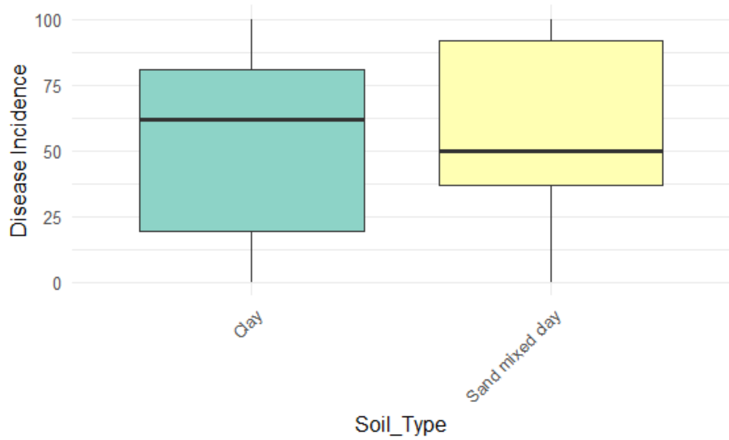

Figure 23: Distribution of Disease Incidence with respect to Soil type

Table 14: Summary measures

Soil_Type	n	Mean	Median	Min	Q1	Q3	Max	SD	IQR
Sand mixed clay	22	55.3	50.0	0	36.7	91.7	100	33.9	55.0
Clay	28	52.3	61.7	0	19.2	80.8	100	34.8	61.7

Median disease incidence is higher in Clay soils compared to Sand mixed clay. Variation is higher in Clay soils than in Sand mixed clay soils, suggesting more diverse disease incidence across clay soil areas.

3.2.7 Prediction on susceptibility of rice variety on Disease Incidence

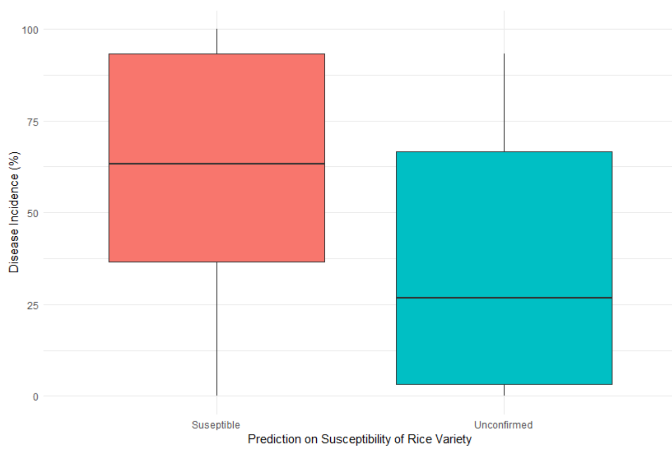

Figure 24: Distribution of Disease Incidence with respect to prediction susceptibility

Table 15: Summary measures

Prediction on Susceptibility	Mean	Median	SD	SE	Q1	Q3	Skewness	n
Susceptible	59.2	63.3	33.9	5.90	36.7	93.3	-0.292	33
Unconfirmed	36.0	26.7	35.4	8.85	3.33	66.7	0.361	16

The figure 24 and table 15 suggest that disease incidence is higher in rice varieties classified as susceptible, with both mean and median values greater than those of the unconfirmed group. The unconfirmed group shows lower central values and more variability, including a wider lower quartile spread. Overall, the results support that predicted susceptibility is associated with higher disease incidence.

3.2.8 Distribution of Disease Incidence with respect to Stage of maturity

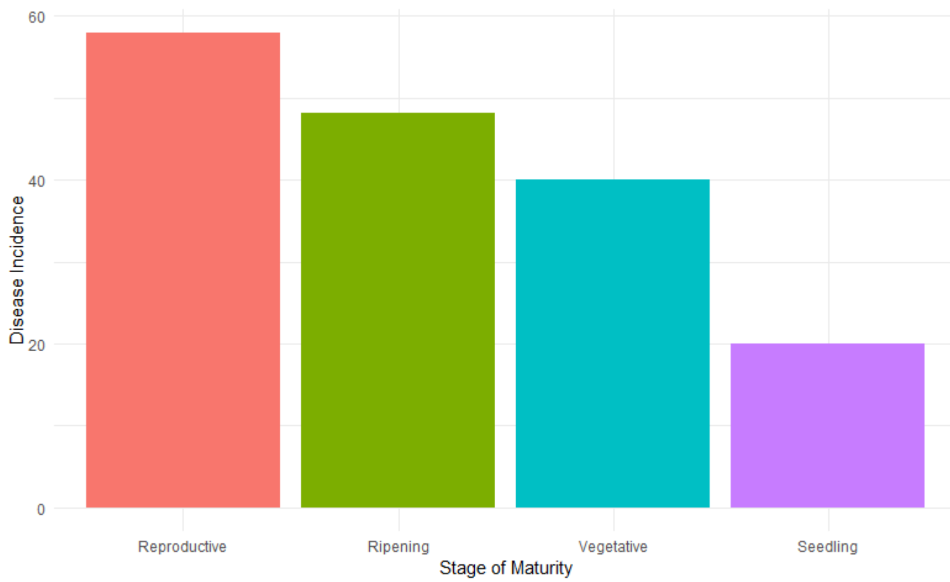

Figure 25: Distribution of Median Disease Incidence with respect to Stage of Maturity

The figure 25 indicates that disease incidence is highest at the reproductive stage, followed by ripening. Vegetative fields show moderate incidence, while seedling plots record the lowest levels. Overall, incidence tends to increase with maturity, peaking in the reproductive stage.

3.2.9 Distribution of Disease Incidence with respect to Rice Variety

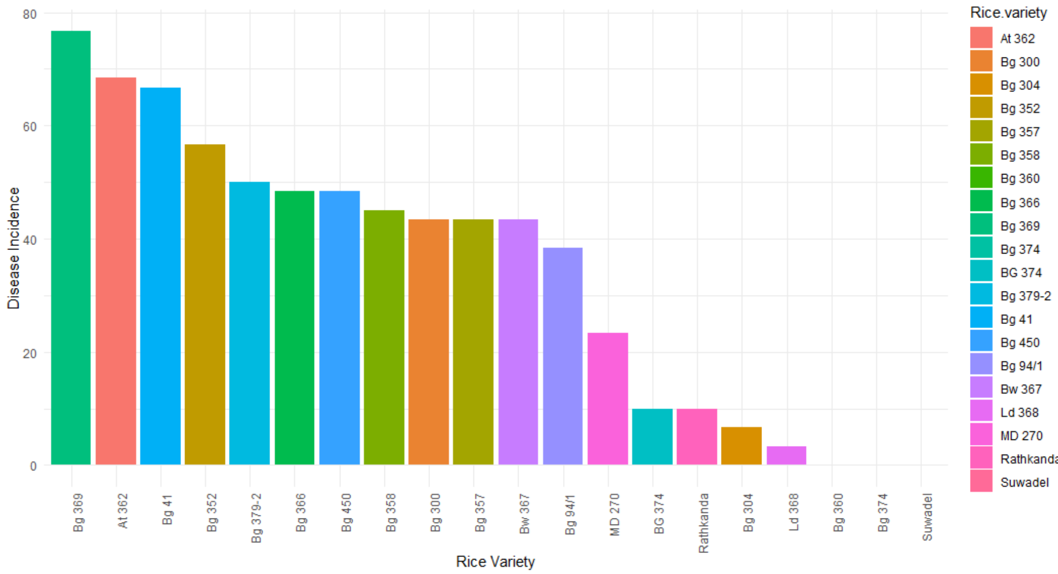

Figure 26: Median Incidence with respect to Rice Variety

Figure 26 reveals clear differences in disease incidence among rice varieties. Varieties such as Bg 369, At 362, and Bg 41 show the highest mean incidence levels, while others like Bg 360, Bg 374, and Suwadel record no incidence. Several varieties, including Bg 352 and Bg 366, fall in an intermediate range around 45–55. Overall, the results suggest that susceptibility to disease varies notably across rice varieties, with a few showing consistently high incidence.

3.2.10 Distribution of Disease Incidence with respect to Topography

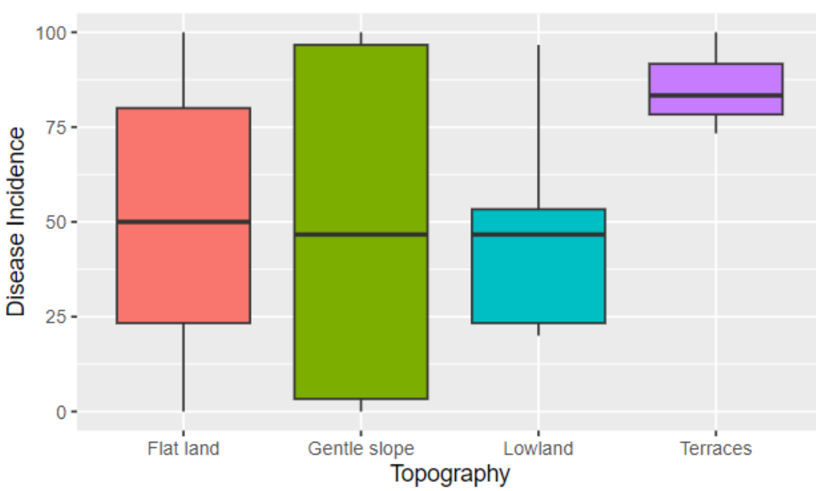

Figure 27: Distribution of Disease Incidence with respect to Topography

Table 16: Summary Measures

Topography	n	Mean	Median	Min	Q1	Q3	Max	SD	IQR
Terraces	3	85.6	83.3	73.3	78.3	91.7	100	13.5	13.3
Gentle slope	9	50.0	46.7	0	3.33	96.7	100	47.2	93.3
Flat land	35	49.5	50.0	20	20.0	78.3	100	32.9	58.3
Lowland	5	48.0	46.7	23.3	23.3	53.3	96.7	30.8	30

Figure 27 shows disease incidence across Flat land, Gentle slope, Lowland, and Terraces. Terraces have the highest and most consistent incidence (mean = 85.6, SD = 13.5). Gentle slopes show the greatest variation, ranging from 0 to 100 (SD = 47.2). This suggests terraced land favors disease spread, while gentle slopes have more unpredictable patterns.

3.2.11 Distribution of Disease Incidence with respect to flooded status

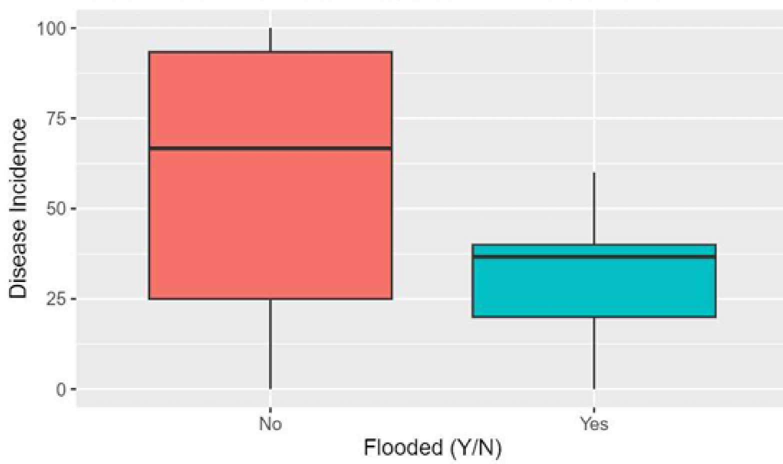

Figure 28: Distribution of Disease Incidence with respect to flooded status

Table 17: Summary measures

Flooded (Y/N)	n	Mean	Median	Min	Q1	Q3	Max	SD	IQR
No	39	58.4	66.7	0	25.0	93.3	100	36.8	68.3
Yes	13	31.0	36.7	0	20.0	40.0	60	17.3	20.0

Figure 28 shows that flooding strongly suppresses disease. Non-flooded plots have much higher and more variable incidence, while flooded plots consistently show lower levels.

3.2.12 Distribution of Disease Incidence with respect to planting method

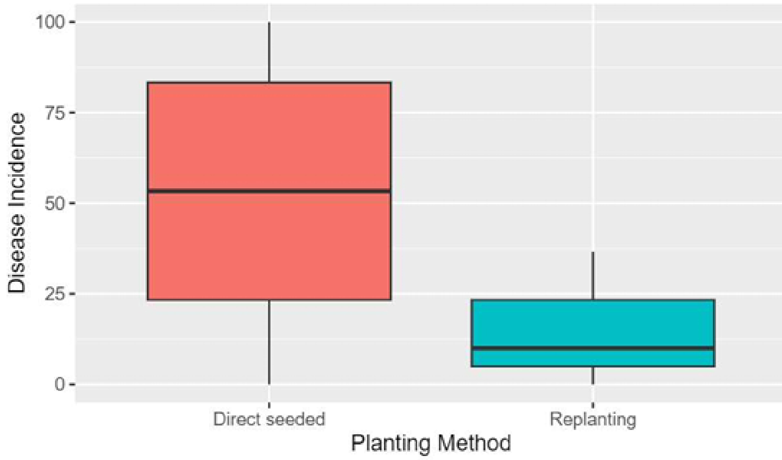

Figure 29: Distribution of Disease Incidence with respect to planting method

Table 18: Summary Measures

Planting Method	n	Mean	Median	Min	Q1	Q3	Max	SD	IQR
Direct seeded	49	53.7	53.3	0	23.3	83.3	100	34.6	60.0
Replanting	3	15.6	10.0	0	5.0	23.3	36.7	19.0	18.3

Direct seeding shows high and variable disease levels, while replanting results in consistently low incidence. This suggests replanting may be a highly effective way to reduce disease, though the limited number of replanted plots means the result should be interpreted cautiously.

3.2.13 Distribution of Disease Incidence with respect to Awareness of MG

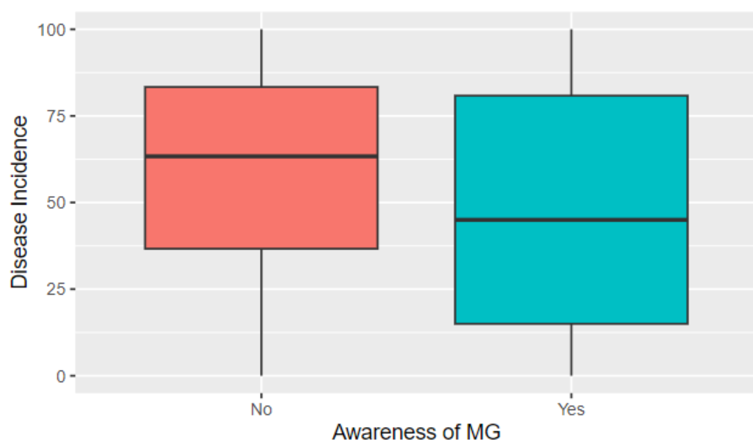

Figure 30: Distribution of Disease Incidence with respect to Awareness of MG

Table 19: Summary Measures

Awareness of MG	n	Mean	Median	Min	Q1	Q3	Max	SD	IQR
No	26	57.2	63.3	0	36.7	83.3	100	33.3	46.7
Yes	24	46.4	45.0	0	15.0	80.8	100	36.3	65.8

Farmers who were aware of MG had lower disease incidence than those who were not. Awareness does not remove risk, but it appears to reduce overall infection levels, likely due to basic preventative practices.

3.2.14 Distribution of Disease Incidence with respect to Irrigation

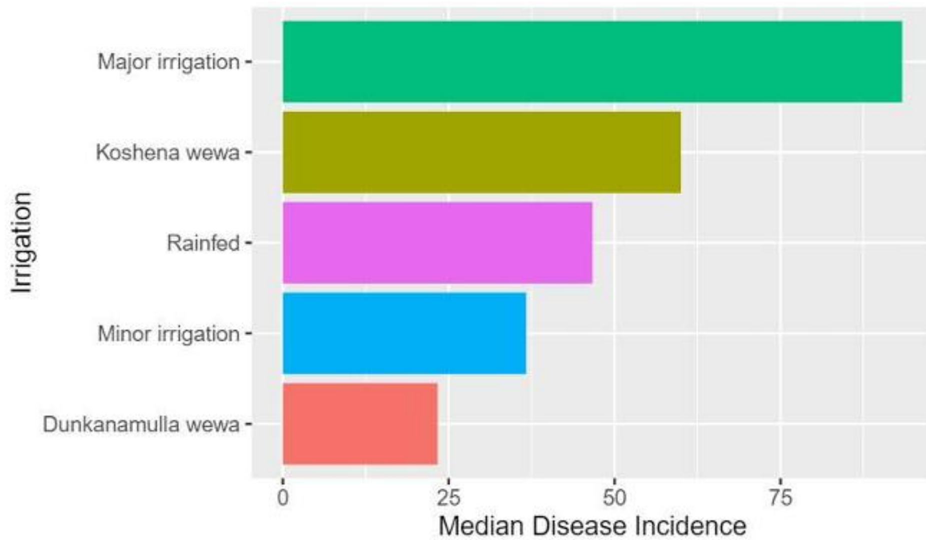

Figure 31: Distribution of Median Disease Incidence with respect to Irrigation

Table 20: Summary Measures Disease Incidence

Irrigation	n	mean	median	min	Q1	Q3	max	sd	IQR
Major irrigation	14	76.4	93.3	0	83.3	96.7	100	35.4	13.3
Koshen wewa	1	60.0	60.0	60	60	60	60	NA	0
Rainfed	25	42.9	46.7	0	16.7	63.3	96.7	29.2	46.7
Minor irrigation	11	41.2	36.7	0	10.0	60.0	100	36.2	50.0
Dunkanamulla wewa	1	23.3	23.3	23.3	23.3	23.3	23.3	NA	0

Fields with major irrigation systems had the highest disease incidence, while rainfed and minor irrigation fields showed moderate levels. Localized wewa systems, especially Dunkanamulla wewa, had much lower incidence, suggesting irrigation type strongly influences disease outcomes.

3.3 Explore Predictor Variables with Disease Severity

3.3.1 Distribution of Disease severity with respect to Province

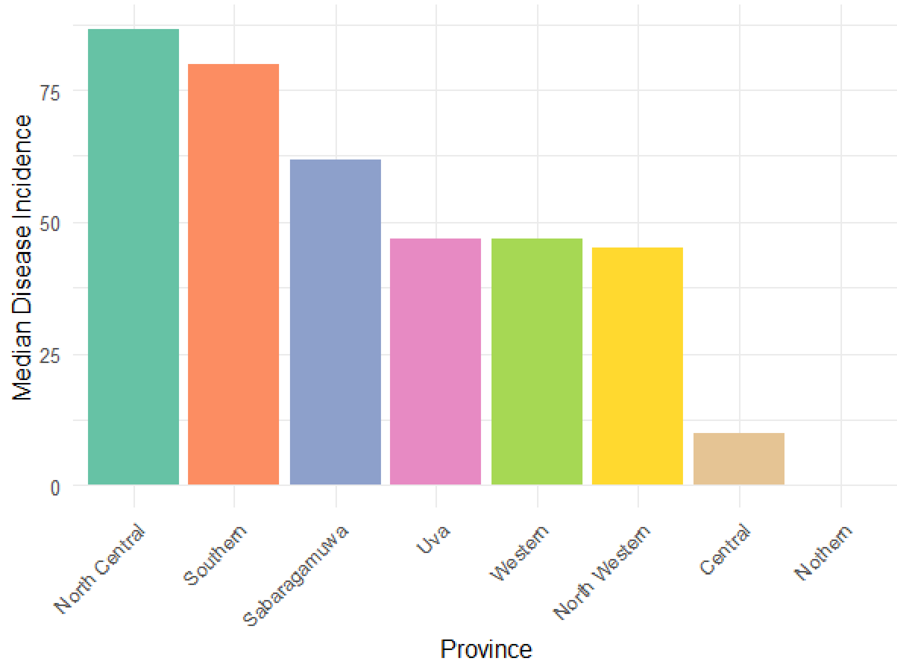

Figure 32: Distribution of Median Disease Severity with respect to Province

The North Central and Southern provinces exhibit the highest disease severity, with median values above 70. Sabaragamuwa, Uva, and North Western and Western provinces have moderate severity, while Central and Northern provinces report low to no severity. Wide variation within provinces is indicated by large standard deviations and interquartile ranges, especially in provinces like Uva and Sabaragamuwa.

3.3.2 Distribution of Disease severity with respect to District

Median disease severity is highest in Hambantota and Anuradhapura, representing hot spots for more severe disease. Other districts show lower median values, including Colombo and Puttlam with almost no disease severity. Variation is moderate in Hambantota but higher in Anuradhapura, demonstrating more inconsistency in severity there.

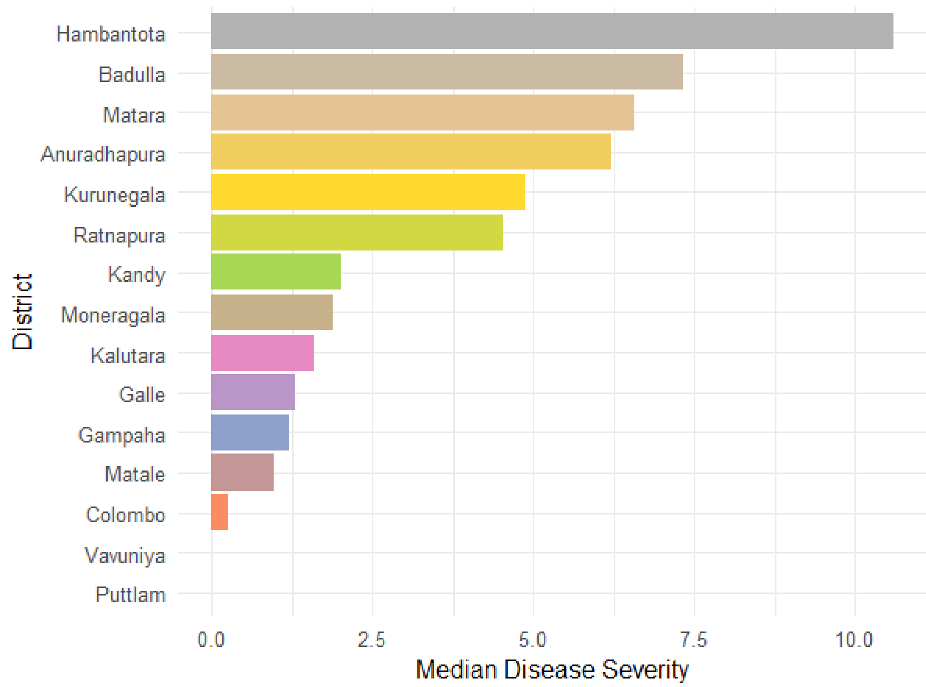

Figure 33: Distribution of Median Disease Severity with respect to District

3.3.3 Distribution of Disease severity with respect to Climatic Zone

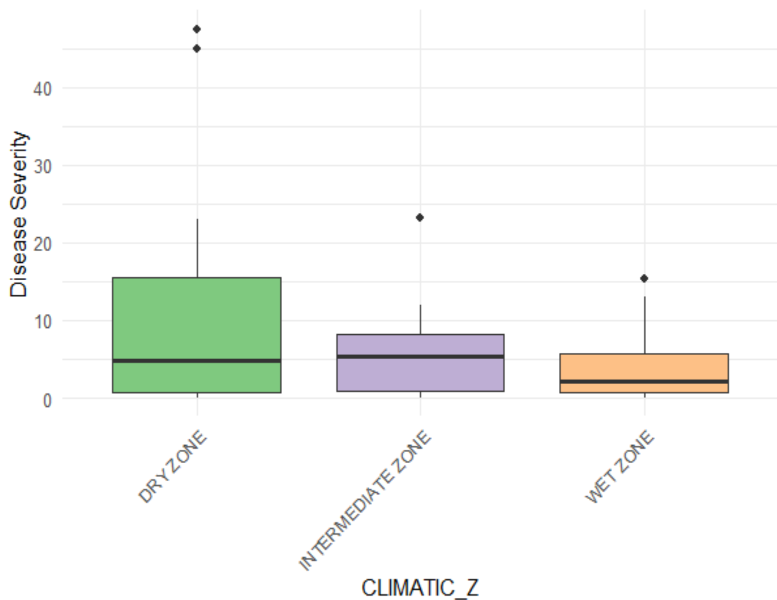

Figure 34: Distribution of Disease Severity with respect to Climatic Zone

Table 21: Summary Measures

Climatic zone	n	Mean	Median	Min	Q1	Q3	Max	SD	IQR
Dry zone	16	11.5	4.85	0.00	0.725	15.6	47.5	15.5	14.9
Intermediate zone	14	5.98	5.30	0.00	0.95	8.23	23.2	6.20	7.28
Wet zone	22	3.93	2.13	0.00	0.65	5.66	15.4	4.50	5.01

Median disease severity is highest in the Dry Zone (4.85), followed by the Intermediate Zone (5.30), and lowest in the Wet Zone (2.13). The Dry Zone also shows greater variability relative to other zones, signaling inconsistent disease severity across locations. This pattern indicates drier climatic conditions are associated with not only higher incidence but also more severe disease.

3.3.4 Distribution of Disease severity with respect to Agro-Ecological Zone

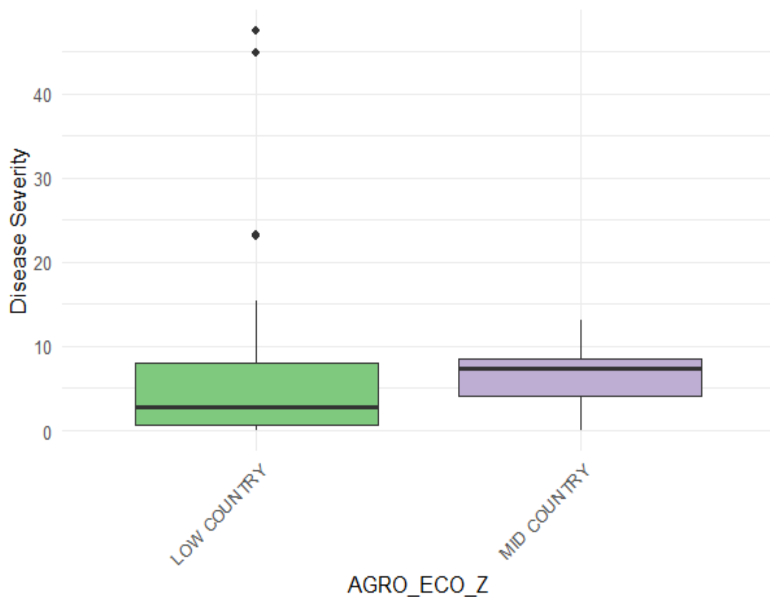

Figure 35: Distribution of Disease Severity with respect to Agro Ecological Zone

Table 22: Summary Measures

Agro eco zone	n	Mean	Median	Min	Q1	Q3	Max	SD	IQR
Low country	47	6.84	2.87	0.0	0.60	8.10	47.5	10.4	7.50
Mid country	5	6.59	7.33	0.0	4.03	8.53	13.1	4.91	4.50

Median disease severity is somewhat higher in Mid Country compared to Low Country, indicating generally more intense disease in mid-country zones. The Low Country shows higher variation compared to the Mid Country, signifying broader severity differences among locations within Low Country.

3.3.5 Distribution of Disease severity with respect to Terrain

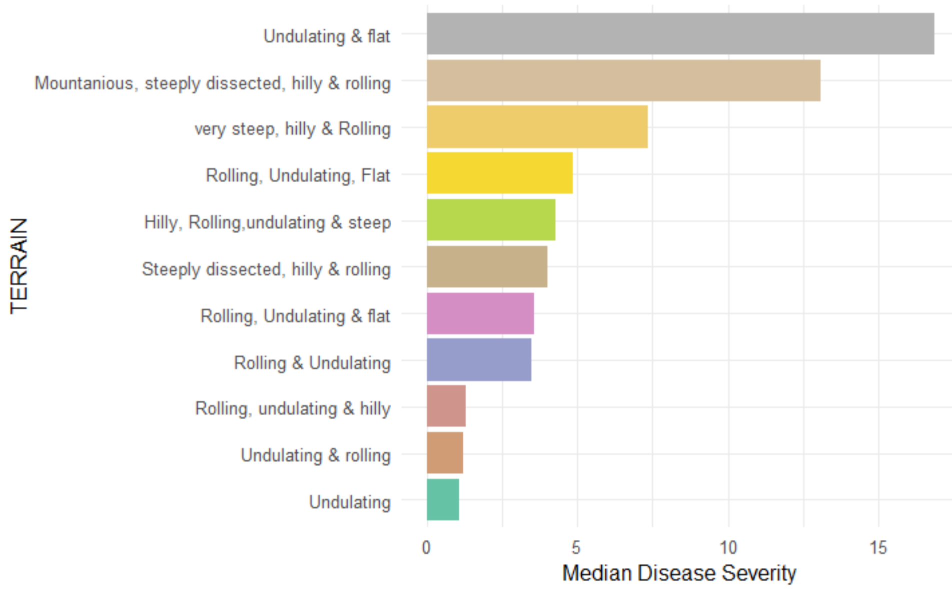

Figure 36: Distribution of Median Disease Severity with respect to Terrain

Median disease severity is highest in Undulating & hilly terrain, followed by Mountainous valleys. Lower median severity is seen in terrains like Rolling, undulating and Undulating. Variation is also large in these areas, particularly Undulating & hilly, reflecting variable severity conditions in hilly topographies.

3.3.6 Distribution of Disease severity with respect to Soil Type

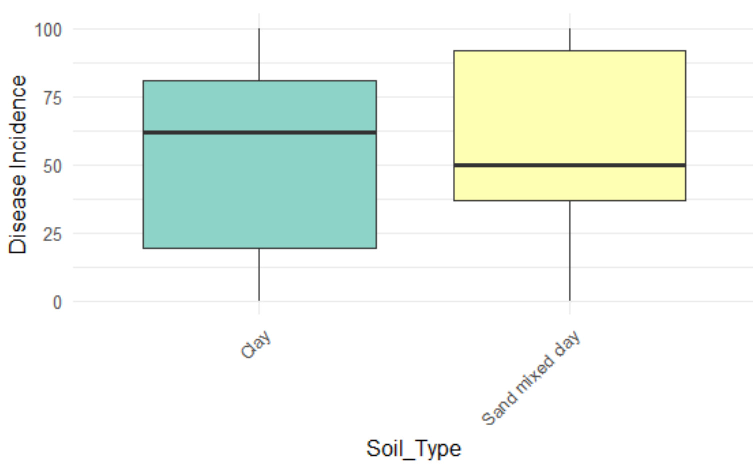

Figure 37: Distribution of Disease Severity with respect to Soil type

Table 23: Summary Measures

Soil_Type	n	Mean	Median	Min	Q1	Q3	Max	SD	IQR
Sand mixed clay	22	8.29	3.77	0	1.20	7.58	47.5	11.5	6.38
Clay	28	6.15	3.22	0	0.60	8.55	44.9	8.92	7.95

Median disease severity is slightly higher in Sand mixed clay than in Clay soils. However, clay soils show more variability compared to sand mixed clay, indicating uneven disease severity distribution within clay soils.

3.3.7 Distribution of Disease Severity with respect to Prediction on Susceptibility

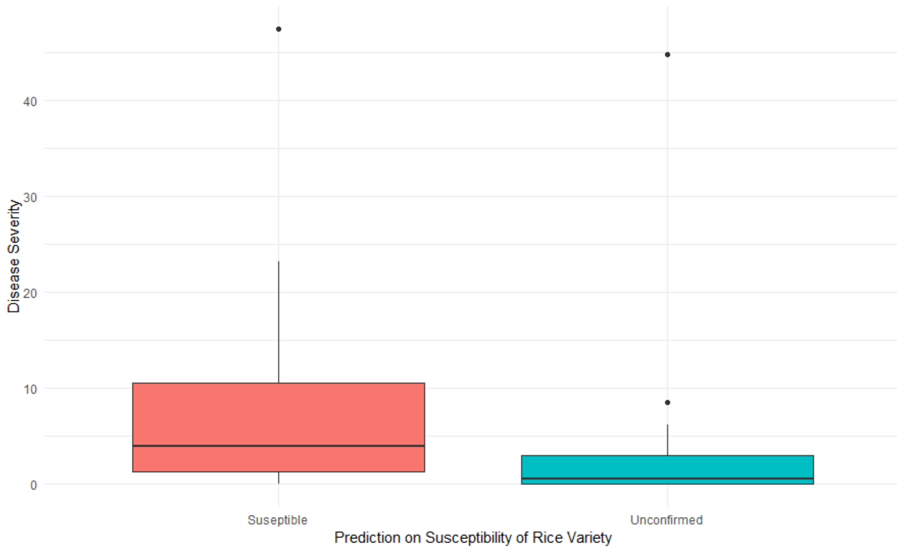

Figure 38: Distribution of Disease Severity with respect to Prediction Susceptibility

Table 24: Summary Measures

Prediction on Susceptibility	Mean	Median	SD	SE	Q1	Q3	Skewness	n
Susceptible	7.92	4.03	9.82	1.71	1.30	10.6	2.19	33
Unconfirmed	4.47	0.53	11.1	2.77	0.08	3.07	3.00	16

Figure 38 and Table 24 suggest that disease incidence is higher in rice varieties classified as susceptible, with both mean and median values greater than those of the unconfirmed group. The unconfirmed group shows lower central values and more variability, including a wider lower quartile spread. Overall, the results support that predicted susceptibility is associated with higher disease incidence.

3.3.8 Distribution of Disease Severity with respect to Stage of Maturity

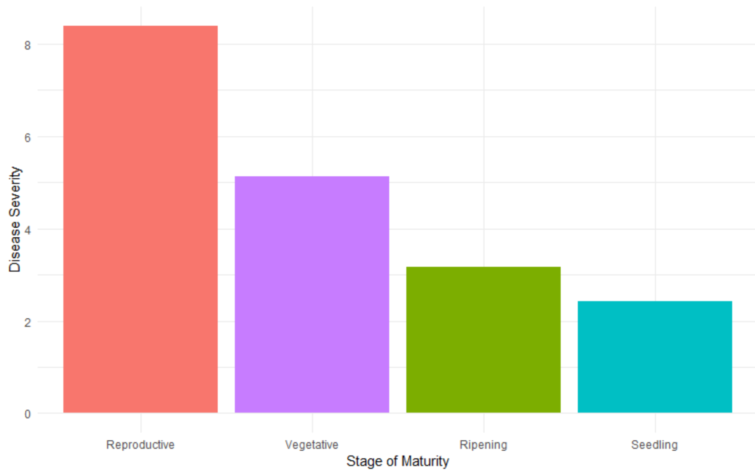

Figure 39: Distribution of Median Disease Severity with respect to Stage of Maturity

Figure 39 shows that Median disease severity is highest at the reproductive stage, followed by the vegetative stage. Ripening and seedling stages display lower Disease severity, with seedlings recording the lowest values. Overall, severity tends to peak at the reproductive stage before declining at later growth stages.

3.3.9 Distribution of Disease Severity with respect to Rice Variety

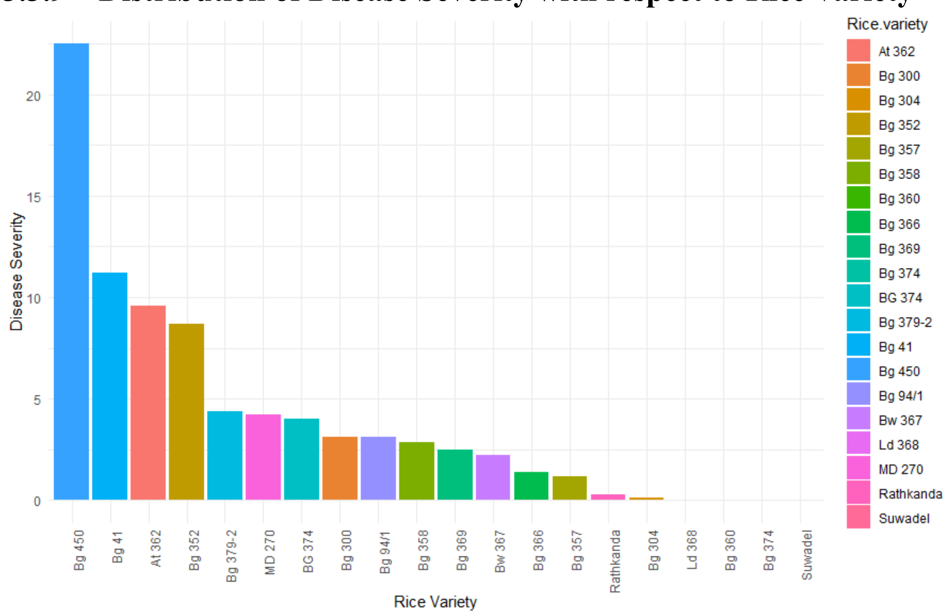

Figure 40: Distribution of Median Disease Severity with respect to Rice Variety

The figure 40 indicate large differences in Median disease severity among rice varieties. Bg 450 shows the highest Disease severity, followed by Bg 41 and At 362, while many varieties such as Bg 360, Bg 374, Suwadel, and Ld 368 show negligible or zero values. Several varieties, including Bg 352 and BG 374, fall in an intermediate range. Overall, severity appears concentrated in a few varieties, while most display low levels.

3.3.10 Distribution of Disease Severity with respect to Topography

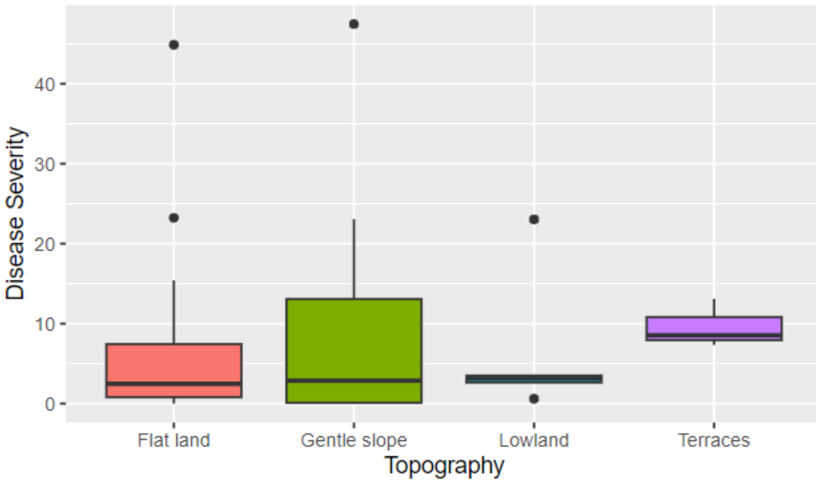

Figure 41: Distribution of Disease Severity with respect to Topography

Table 25: Summary measures

Topography	n	Mean	Median	Min	Q1	Q3	Max	SD	IQR
Gentle slope	9	10.5	2.87	0	0.1	13.1	47.5	15.9	13.00
Terraces	3	9.66	8.53	7.33	7.93	10.8	13.1	3.04	2.88
Lowland	5	6.58	3.13	0.6	2.63	5.3	23.0	2.67	0.867
Flat land	35	5.66	2.47	0	0.7	6.87	44.9	8.57	6.17

Disease severity is highest in Gentle slopes and Terraces, while Flat lands and Lowlands show lower levels. Gentle slopes also have the greatest variability and frequent outliers, suggesting topography especially sloped areas, plays an important role in both the spread and intensity of disease.

3.3.11 Distribution of Disease Severity with respect to Flooded Status

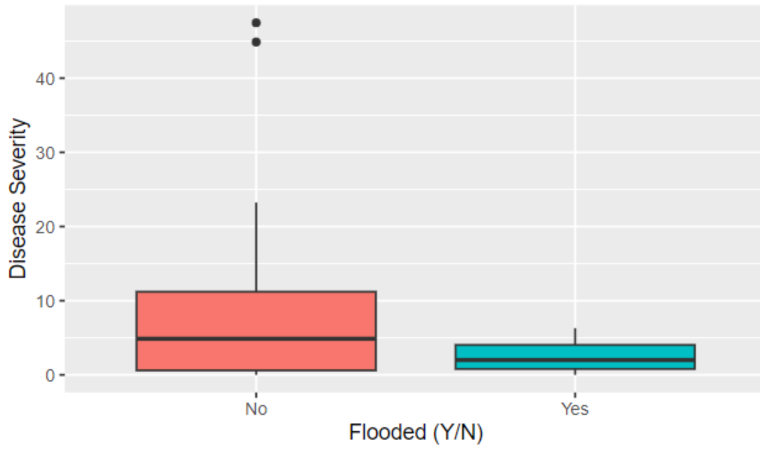

Figure 42: Distribution of Disease Severity with respect to Flooded Status

Table 26: Summary Measures

Flooded (Y/N)	n	Mean	Median	Min	Q1	Q3	Max	SD	IQR
No	39	8.26	4.87	0	0.43	10.9	47.5	11.1	10.5
Yes	13	2.51	2.00	0	0.80	4.03	6.3	2.05	3.23

Flooding greatly reduces disease severity. Non-flooded plots showed high and unpredictable severity, while flooded plots had consistently low and manageable levels.

3.3.12 Distribution of Disease Severity with respect to Planting Method

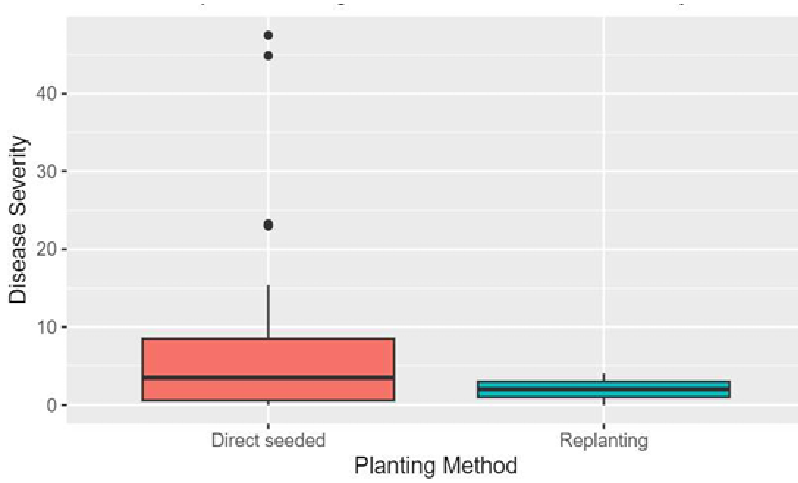

Figure 43: Distribution of Disease Severity with respect to Planting Method

Table 27: Summary Measures

Planting Method	n	Mean	Median	Min	Q1	Q3	Max	SD	IQR
Direct seeded	49	53.7	53.3	0	23.3	83.3	100	34.6	60.0
Replanting	3	15.6	10.0	0	5.0	23.3	36.7	19.0	18.3

Direct seeding results in higher and more variable disease severity, with an average of 7.11 and a maximum of 47.5, while the median is 3.5, showing some plots experience extreme infection. Replanting keeps severity consistently low, with an average of 2.01 and a maximum of 4.03, highlighting its effectiveness in reducing disease impact.

3.3.13 Distribution of Disease Severity with respect to Awareness of MG

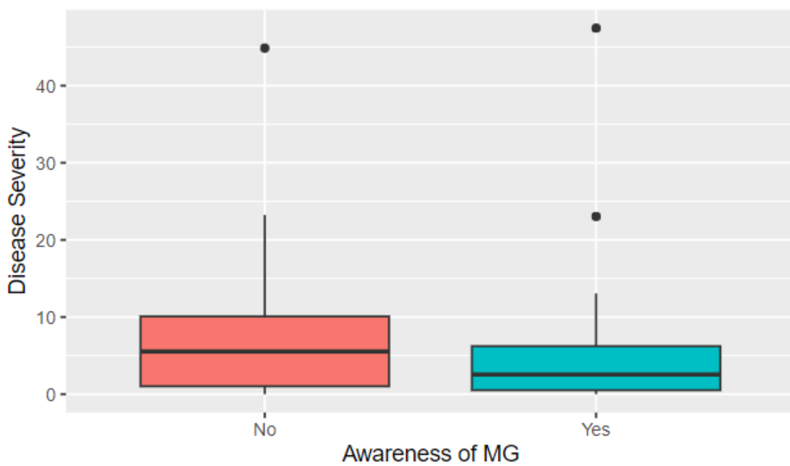

Figure 44 Distribution of Disease Severity with respect to Awareness of MG

Table 28: Summary Measures

Awareness of MG	n	Mean	Median	Min	Q1	Q3	Max	SD	IQR
No	26	7.92	5.53	0	1.03	10.1	44.9	10.0	9.06
Yes	24	5.93	2.55	0	0.52	6.22	47.5	10.4	5.70

Farmer awareness helps reduce disease severity. The unaware group had a higher average severity (7.92) and median (5.53), while the aware group had a lower average (5.93) and much lower median (2.55). Although extreme cases occurred in both groups (max 44.9 and 47.5), awareness appears to keep severity low for most plots, likely through better management practices.

3.3.14 Distribution of Disease Severity with respect to Irrigation

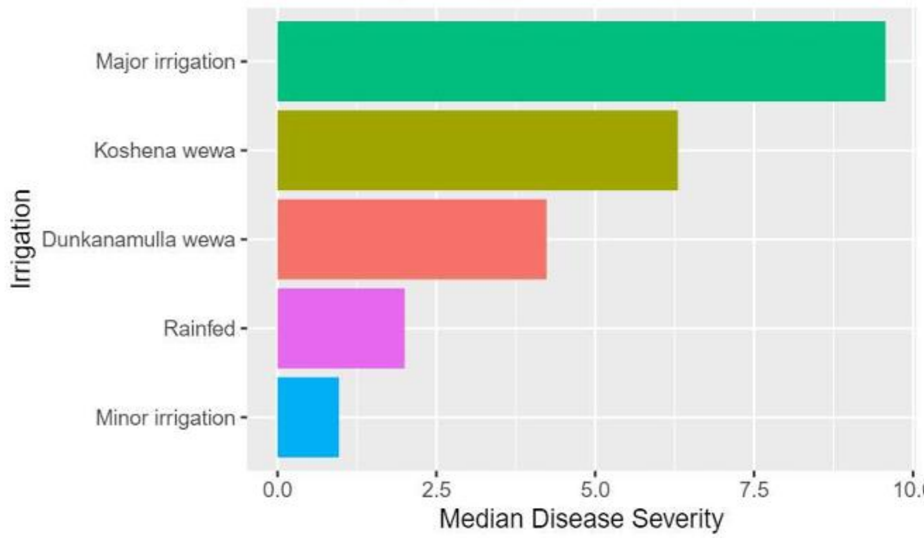

Figure 45: Distribution of Disease Severity with respect to Irirgation

Disease severity showed clear differences across irrigation methods. Fields under major irrigation experienced the highest severity, while Koshen wewa and Dunkanamulla wewa exhibited moderate levels. Minor irrigation and rainfed fields recorded comparatively lower severity, indicating that disease severity tends to increase under larger irrigation systems.

3.4 Explore the Pest Distribution in the Respective Areas

3.4.1 Meloidogyne Disease Incidence by Sampling Sites

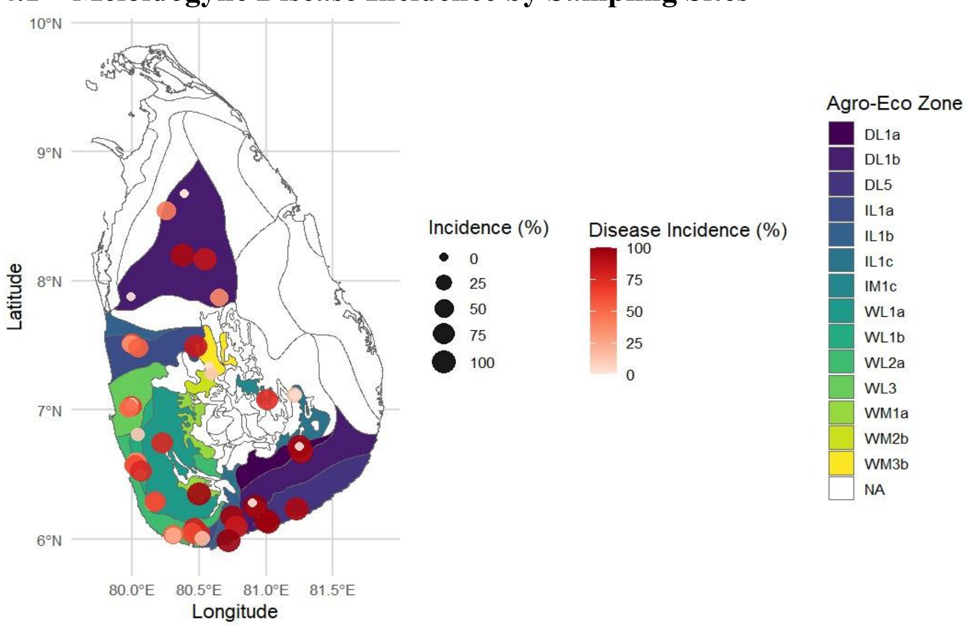

Figure 46: Distribution of MG disease Incidence by Sampling Sites

The spatial distribution of *Meloidogyne graminicola* incidence shows variation across sampled rice fields in different agro-ecological zones. Disease incidence is represented using point data, where the size and color intensity of each point reflect the percentage of infected plants. High incidence rates, reaching up to 100%, are observed in several sampling sites located in the southern, southwestern, and central areas of the study region. These areas form distinct clusters and are identified as potential disease hotspots.

Incidence levels vary among agro-ecological zones, including DL (Dry Zone Low Country), WL (Wet Zone Low Country), IL (Intermediate Zone Low Country), and WM (Wet Mid Country). Some zones display consistently high levels of infestation, while others show low or no incidence. These differences suggest that climatic conditions, soil type, irrigation practices, and rice variety may contribute to the presence or absence of nematode infestation.

The spatial variation indicates the need for localized pest management strategies. Identifying and prioritizing high-incidence zones can support more effective allocation of control efforts. The map also provides baseline information for ongoing monitoring and future studies focused on the environmental factors influencing nematode outbreaks in rice fields.

3.4.2 Meloidogyne Disease Severity by Sampling sites

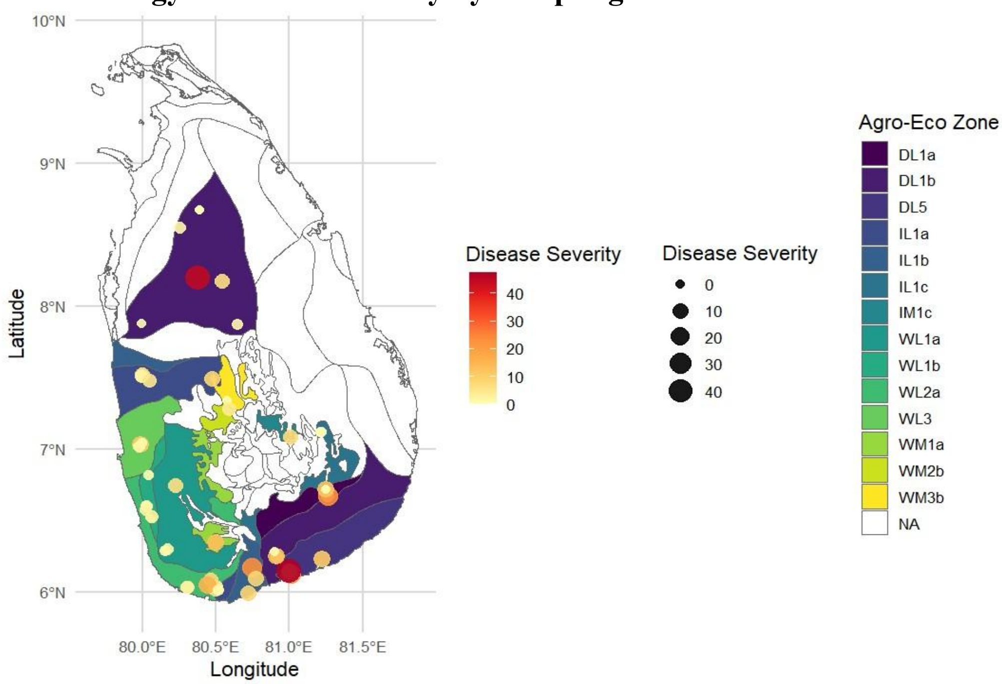

Figure 47: Distribution of MG disease severity by Sampling sites

The mapped distribution of *Meloidogyne graminicola* severity across sampled rice fields indicates considerable variation across agro-ecological zones. Disease severity was measured based on gall count intensity per plant and visualized using circle size and color intensity at each sampling location. Larger and darker red circles correspond to areas with higher severity, while smaller and lighter-colored circles represent lower severity levels.

Clusters of high severity are observed in specific regions, particularly in parts of the southern and north-central zones. These areas exhibit the most intense nematode damage and may reflect conditions that are more conducive to nematode development and reproduction. In contrast, lower severity levels are recorded in other zones, with some regions showing minimal or no severe infestation.

Differences in severity are evident across agro-ecological classifications. Certain zones, such as selected DL (Dry Low Country) and WM (Wet Mid Country) types, are more frequently associated with higher severity levels. Zones with lower severity often coincide with areas of lower disease incidence or may be influenced by other factors such as soil type, crop variety, or irrigation method.

This spatial variation in severity highlights the need for zone-specific management strategies. Areas with consistently high severity should be prioritized for intervention to reduce crop losses. The identification of severity patterns also supports future research aimed at understanding the underlying environmental or agronomic drivers contributing to nematode damage intensity.

3.4.3 Meloidogyne Disease Incidence by Agro-Ecological Region

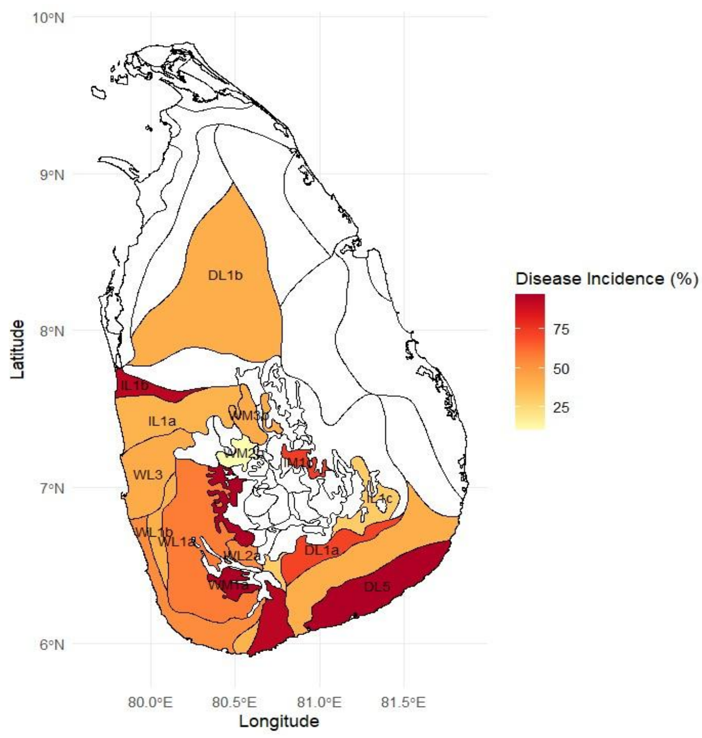

Figure 48: MG disease incidence by Agro Ecological Region

The spatial analysis of *Meloidogyne graminicola* incidence demonstrates considerable variation across agro-ecological regions. Disease incidence was mapped by coding regions according to agro-ecological classifications (e.g., DL1b, IL1a, WL3) and visualizing the average percentage of infected rice plants. Color intensity on the map corresponds to incidence level, ranging from yellow (low) to deep red (high).

Regions such as DL5, IL10, and selected portions of WL3 and DL1a record the highest incidence rates, frequently exceeding 75%. These areas are represented in deep red and are identified as zones with elevated nematode pressure. Moderate incidence, typically ranging from 25% to 50%, is observed in zones like DL1b and WL1a, represented with orange shading. Zones showing minimal or no infestation, including UM1, WM1, IM1, and IM2, are located primarily in the central and northern parts of the study area and are marked in yellow or light colors.

Clustering of high-incidence zones is notable in southern and southeastern regions, suggesting a spatial relationship between nematode distribution and regional environmental or agronomic conditions. These may include soil type, moisture levels, temperature profiles, and specific cultural practices common to those zones. Conversely, lower incidence in some central regions may be attributed to factors such as resistant rice varieties, better drainage, or more effective field management practices.

The observed patterns imply a strong link between agro-ecological zoning and nematode disease prevalence. Identifying these associations is valuable for informing management strategies. Regions with high incidence require immediate attention through targeted interventions such as improved irrigation control, crop rotation planning, or adoption of tolerant varieties. Additionally, the map can be used to inform future surveillance programs and support spatially differentiated pest management approaches.

Supporting evidence from other rice-growing regions has indicated similar spatial trends, with high nematode incidence often associated with areas of high soil moisture and continuous rice cropping. These findings reinforce the relevance of integrating agro-ecological context into pest risk assessment and control strategies.

3.4.4 Meloidogyne Disease Severity by Agro-Ecological Region

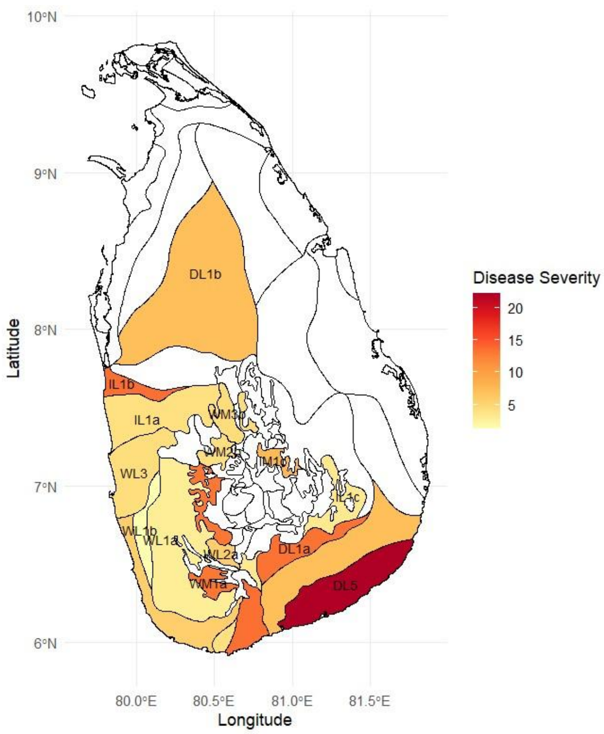

Figure 49: MG Disease Severity by Agro-Ecological Region

The figure 49 presents a comparative visualization of *Meloidogyne graminicola* disease incidence and severity across multiple agro-ecological zones, revealing significant spatial variation in nematode impact. Regions such as DL5, IL1b, WM1a, and DL1a exhibit the highest combined incidence and severity levels. DL5 records the most severe infestation, with an incidence rate of 96.7% and a severity score of 22.20, marking it as a critical hotspot. Similarly, IL1b and WM1a each show incidence rates exceeding 93% with corresponding severity scores above 13, while DL1a also reflects a high impact profile with 71.7% incidence and 13.30 severity.

Moderate risk regions include DL1b, IM1c, and WL2a. DL1b, despite a relatively lower severity score of 6.92, is notable due to a large number of samples, suggesting a broad exposure to moderate infection levels. In contrast, zones such as WM2b, WL1b, WL3, IL1a, and IL1c exhibit considerably lower disease metrics, indicating lower current risk but not excluding future vulnerability.

The spatial arrangement of disease severity and incidence in the figure shows clear clustering in southeastern and central zones. These clusters may be associated with environmental factors such as temperature, soil moisture, or irrigation practices that support nematode proliferation.

The figure also illustrates a decreasing gradient of disease impact when moving toward interior or northern zones, with both severity and incidence metrics declining.

These visual and quantitative insights highlight the importance of prioritizing high-impact zones for immediate intervention and resource allocation. Additionally, moderately affected areas, particularly those with larger cultivation areas, should be considered for proactive monitoring. The combined display of disease metrics in relation to agro-ecological zoning provides critical evidence for developing targeted and efficient management strategies aimed at reducing nematoderelated losses in rice production.

3.5 Model fitting

3.5.1 Theory Behind Negative Binomial Model

The Negative Binomial model is used for modeling count data that exhibit overdispersion. When the variance of the response variable exceeds its mean. In the context of this study, the response variable is Gall Count, representing the number of root galls caused by *Meloidogyne graminicola* on each rice plant. Preliminary Poisson regression showed overdispersion (dispersion ratio > 1), making the Negative Binomial model more appropriate for capturing the extra variability in the data.

The Negative Binomial model assumes that the observed Gall counts Y_i follow a Negative Binomial distribution with mean μ_i and dispersion parameter θ. The expected count μ_i is linked to a set of predictors X_i through a log link function.

The log link function ensures that predicted counts are always positive and models the effect of each predictor as a multiplicative change in the expected count. For example, a coefficient β_j corresponds to a e^{β_j} fold change in the expected gall count for a one-unit increase in predictor X_j holding other variables constant.

The theoretical justification for the NB model stems from considering the count data as arising from a Poisson process with an unobserved heterogeneity term that follows a Gamma distribution. This allows the variance to exceed the mean, effectively capturing the extra-Poisson variability often seen in biological count data like gall counts.

3.5.2 Final Negative Binomial Regression Model

A negative binomial regression model was fitted to investigate the factors affecting gall counts per rice plant. If the dispersion parameter equals zero, the model reduces to the simpler poisson model. If the dispersion parameter is significantly greater than zero than the data are over dispersed and are better estimated using a negative binomial model than a poisson model.

The negative binomial model adequately accounted for overdispersion in the gall count data, as evidenced by a significant dispersion parameter estimate of 2.44 (95% Confidence Interval: 2.21–2.70). This confirms that the negative binomial distribution was a more appropriate choice than the Poisson model, which assumes equal mean and variance, thereby providing more reliable parameter estimates and inference.

Table 29: Final Model

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	4.3574	0.8636	2.6647	6.0501	25.46	<.0001
agro_eco_r_lump	DL1a	1	-4.9232	0.9493	-6.7837	-3.0626	26.90	<.0001
agro_eco_r_lump	DL1b	1	-3.2203	0.6155	-4.4266	-2.0140	27.38	<.0001
agro_eco_r_lump	DL5	1	-0.5270	0.6357	-1.7728	0.7189	0.69	0.4071
agro_eco_r_lump	IL1a	1	-3.8703	0.6407	-5.1260	-2.6146	36.49	<.0001
agro_eco_r_lump	IL1b	1	-3.1991	0.6141	-4.4027	-1.9955	27.14	<.0001
agro_eco_r_lump	IL1c	1	-4.6443	0.5684	-5.7583	-3.5303	66.77	<.0001
agro_eco_r_lump	Other	1	-0.6558	0.5527	-1.7391	0.4276	1.41	0.2355
agro_eco_r_lump	WL1a	1	-6.5618	0.6441	-7.8243	-5.2993	103.77	<.0001
agro_eco_r_lump	WL1b	1	-4.1798	0.6840	-5.5204	-2.8393	37.35	<.0001
agro_eco_r_lump	WL2a	1	-4.0046	0.6702	-5.3182	-2.6910	35.70	<.0001
agro_eco_r_lump	WL3	1	-6.0608	0.7134	-7.4591	-4.6625	72.17	<.0001
rice_variety_lump	At 362	1	1.8176	0.2582	1.3116	2.3236	49.56	<.0001
rice_variety_lump	Bg 352	1	2.2649	0.3762	1.5276	3.0023	36.25	<.0001
rice_variety_lump	Bg 358	1	1.3127	0.4057	0.5176	2.1078	10.47	0.0012
rice_variety_lump	Bg 366	1	2.6345	0.3786	1.8926	3.3765	48.44	<.0001
rice_variety_lump	Bg 374	1	-2.6204	0.5539	-3.7059	-1.5348	22.38	<.0001
rice_variety_lump	Bg 379-2	1	-1.3299	0.4990	-2.3080	-0.3519	7.10	0.0077
rice_variety_lump	Bg 450	1	3.1162	0.3553	2.4198	3.8125	76.94	<.0001
rice_variety_lump	Bw 367	1	0.9043	0.3504	0.2175	1.5912	6.66	0.0099
soil_type	Clay	1	-1.0380	0.2052	-1.4402	-0.6358	25.59	<.0001
rice_ecosystem_rec	Irrigated	1	-0.5023	0.3934	-1.2734	0.2687	1.63	0.2016
topography	Flat land	1	0.7995	0.2656	0.2790	1.3200	9.06	0.0026
topography	Gentle slope	1	-2.7240	0.3635	-3.4364	-2.0116	56.17	<.0001
topography	Lowland	1	0.5503	0.5627	-0.5526	1.6532	0.96	0.3281
flooded_y_n	No	1	2.8541	0.3131	2.2405	3.4677	83.11	<.0001
method_of_planting	Direct seeded	1	-0.8402	0.4117	-1.6472	-0.0332	4.16	0.0413
is_the_farmer_aware_	No	1	-0.8528	0.1875	-1.2202	-0.4853	20.69	<.0001
age_das		1	-0.0183	0.0060	-0.0301	-0.0065	9.27	0.0023
Dispersion		1	2.4407	0.1264	2.2051	2.7015		

3.5.3 Interpretations of the model

Table 30 presents the incidence rate ratios (IRRs) for the significant predictors from the negative binomial regression model. The IRRs represent the multiplicative change in expected gall counts associated with each predictor, holding other variables constant.

Table 30: Table of Significant Coefficients

Parameter	Estimate	IRR ($e^{Estimate}$)
agro_eco_r_lump (DL1a)	-4.9232	0.0073
agro_eco_r_lump (DL1b)	-3.2203	0.0399
agro_eco_r_lump (IL1a)	-3.8703	0.0208
agro_eco_r_lump (IL1b)	-3.1991	0.0409
agro_eco_r_lump (WL1a)	-6.5618	0.0014
agro_eco_r_lump (WL1b)	-4.1798	0.0153
agro_eco_r_lump (WL2a)	-4.0046	0.0183
agro_eco_r_lump (WL3)	-6.0608	0.0023
rice_variety_lump (At 362)	1.8176	6.16
rice_variety_lump (Bg 352)	2.2649	9.63
rice_variety_lump (Bg 358)	1.3127	3.72
rice_variety_lump (Bg 366)	2.6345	13.94
rice_variety_lump (Bg 374)	-2.6204	0.07
rice_variety_lump (Bg 379-2)	-1.3299	0.26
rice_variety_lump (Bg 450)	3.1162	22.57
rice_variety_lump (Bw 367)	0.9043	2.47
soil_type (Clay)	-1.0380	0.35
topography (Flat land)	0.7995	2.22
topography (Gentle slope)	-2.7240	0.07
flooded_y_n (No)	2.8541	17.38
method_of_planting (Direct seeded)	-0.8402	0.43
is_the_farmer_aware_ (No)	-0.8528	0.43
age_das	-0.0183	0.98

Agro Ecological Region

- Plants in agro-ecological region DL1a have gall counts that decrease by a factor of 0.0073 compared to those in region WM3b, after controlling for other variables.
- Plants in agro-ecological region DL1b have gall counts that decrease by a factor of 0.0399 compared to those in region WM3b, holding all other variables constant.
- Plants in agro-ecological region IL1a have gall counts that decrease by a factor of 0.0208 compared to those in region WM3b, holding all other variables constant.
- Plants in agro-ecological region IL1b have gall counts that decrease by a factor of 0.0409 compared to those in region WM3b, holding all other variables constant.
- Plants in agro-ecological region WL1a have gall counts that decrease by a factor of 0.0014 compared to those in region WM3b, holding all other variables constant.

- Plants in agro-ecological region WL1b have gall counts that decrease by a factor of 0.0153 compared to those in region WM3b, holding all other variables constant.
- Plants in agro-ecological region WL2a have gall counts that decrease by a factor of 0.0183 compared to those in region WM3b, holding all other variables constant.
- Plants in agro-ecological region WL3 have gall counts that decrease by a factor of 0.0023 compared to those in region WM3b, holding all other variables constant.

Rice Variety

- Plants of rice variety At 362 have gall counts that increase by a factor of 6.16 compared to those of the 'other' rice variety, holding all other variables constant.
- Plants of rice variety Bg 352 have gall counts that increase by a factor of 9.63 compared to those of the 'other' rice variety, holding all other variables constant.
- Plants of rice variety Bg 358 have gall counts that increase by a factor of 3.72 compared to those of the 'other' rice variety, holding all other variables constant.
- Plants of rice variety Bg 366 have gall counts that increase by a factor of 13.94 compared to those of the 'other' rice variety, holding all other variables constant.
- Plants of rice variety Bg 374 have gall counts that decrease by a factor of 0.07 compared to those of the 'other' rice variety, holding all other variables constant.
- Plants of rice variety Bg 379-2 have gall counts that decrease by a factor of 0.26 compared to those of the 'other' rice variety, holding all other variables constant.
- Plants of rice variety Bg 450 have gall counts that increase by a factor of 22.57 compared to those of the 'other' rice variety, holding all other variables constant.
- Plants of rice variety Bw 367 have gall counts that increase by a factor of 2.47 compared to those of the 'other' rice variety, holding all other variables constant.

Soil Type

- Plants grown in clay soil have gall counts that decrease by a factor of 0.35 compared to those grown in sand-mixed clay, holding all other variables constant.

Topography

- Plants grown on flat land have gall counts that increase by a factor of 2.22 compared to those grown on low land, holding all other variables constant.
- Plants grown on gentle slopes have gall counts that decrease by a factor of 0.07 compared to those grown on low land, holding all other variables constant.

Flooded

- Plants grown in fields that were not flooded during the season have gall counts that increase by a factor of 17.4 compared to those grown in flooded fields, holding all other variables constant.

Method of Planting

- Plants that were direct-seeded have gall counts that decrease by a factor of 0.43 compared to plants established by indirect (transplanted) planting, holding all other variables constant.

Farmer awareness

- Plants grown in fields where the farmer was not aware of the pest have gall counts that decrease by a factor of 0.43 compared to those in fields where the farmer was aware, holding all other variables constant.

Age DAS (Days After Snowing)

- For each additional day of plant age, gall counts decrease by a factor of 0.98, holding all other variables constant

3.5.4 Model Evaluation

- **Checking Multicollinearity**

Table 31: Variables along with VIF

Predictor	Degrees of Freedom	GVIF	GVIF^(1/(2*Df))
agro_eco_r_lump	11	204.139	1.27349
rice_variety_lump	8	104.565	1.33725
soil_type	1	0.021	0.14611
rice_ecosystem_rec	1	0.272	0.52110
topography	3	27.592	1.73832
flooded_y_n	1	15.176	3.89558
method_of_planting	1	1.137	1.06614
is_the_farmer_aware_	1	1.201	1.09598
age_das	1	4.542	2.13126

Multicollinearity among predictors was assessed using the Generalized Variance Inflation Factor (GVIF). All predictors had adjusted GVIF values well below the common threshold of 5, indicating that multicollinearity is not a concern in the final model.

- **Goodness of Fit of the model**

Table 32: Goodness of fit of the model

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	1319	1278.8583	0.9696
Scaled Deviance	1319	1278.8583	0.9696
Pearson Chi-Square	1319	1552.7810	1.1772
Scaled Pearson X2	1319	1552.7810	1.1772
Log Likelihood		20534.3312	
Full Log Likelihood		-3425.9196	
AIC (smaller is better)		6855.8392	
AICC (smaller is better)		6855.8483	
BIC (smaller is better)		6866.2100	

Hypothesis to be tested:

H_0: The Model is adequate vs H_1: The model is not adequate

The goodness-of-fit of the fitted negative binomial model was evaluated using both the deviance and Pearson chi-square statistics, each divided by their respective degrees of freedom. The deviance divided by its degrees of freedom was approximately 0.97, which is very close to 1, indicating that the model is well-calibrated to the observed data and shows no strong evidence of under- or over-dispersion. The Pearson chi-square divided by its degrees of freedom was slightly higher, around 1.18, suggesting a minor degree of residual overdispersion; however, this value is still within the range typically considered acceptable in applied work. Overall, both statistics being close to 1 indicates that the model captures the variability in the data reasonably well without leaving behind systematic lack of fit. Based on these results, we fail to reject the null hypothesis of adequate model fit, confirming that the model is appropriate for the observed data.

Based on this, we fail to reject the null hypothesis. Thus, there is no evidence of lack of fit, and the model is considered adequate for the data.

4 Discussion

During the exploratory data analysis (EDA), variables were examined on a site-wise basis rather than plant-wise. This decision was made because summary statistics such as percentages, means, and medians did not show meaningful differences between the two approaches. Aggregating the data at the site level therefore provided a clearer and more manageable framework without reducing accuracy. Rows with missing values were excluded throughout the analysis to ensure consistency and to avoid biases that incomplete records might introduce.

Some categorical variables contained a large number of unique categories, which created sparsity in the data. To address this, categories with fewer than 50 observations were combined into a single "Other" group. However, in some models this "Other" category was automatically treated as the reference level. Since "Other" is not an informative category, this made interpretation less meaningful. A more deliberate selection of reference categories would improve the clarity of results in future analyses.

When constructing the models, certain variables were excluded to maintain focus and interpretability. Province and district were not included, since the study investigates pest disease dynamics, where agro-ecological region provides a more meaningful spatial dimension. Similarly, instead of including the continuous variable for age of maturity, the categorical variable representing stage of maturity was used, as it better reflects the biological phases that influence vulnerability to disease. In addition, variables such as GPX, X, Y coordinates, and season were also excluded. Since the dataset specifically represents Maha season 2024, these variables did not add explanatory value for the modeling process and would only introduce redundancy.

In assessing the adequacy of negative binomial regression models, researchers emphasize the importance of employing multiple goodness-of-fit and hypothesis testing approaches. Among the most widely used measures are the deviance and Pearson's χ^2 statistics, which provide global evaluations of model fit. The deviance statistic can be viewed as a likelihood-ratio test comparing the fitted model with a saturated model, where each observation is perfectly predicted (Smyth, 2003). In contrast, Pearson's χ^2 statistic functions as a score test, where the expected value of the score is zero under a correctly specified model. Both statistics, when scaled by their degrees of freedom, are expected to be close to one under adequate fit, with values substantially greater than one indicating overdispersion and values below one suggesting underdispersion (Cameron & Trivedi, 1990; Smyth, 2003).

The dispersion parameter estimated in negative binomial regression further provides direct evidence on whether overdispersion is present. In many applied studies, the ratio of the Pearson χ^2 statistic to its degrees of freedom is used as a practical diagnostic to detect overdispersion. Ratios substantially greater than one indicate that the variance exceeds the mean, thus justifying the use of a negative binomial rather than a Poisson model (Dean & Lawless, 1989; Gurmu, 1991). Empirical work has also shown that while likelihood ratio and Wald tests may be applied to test the dispersion parameter, score tests often provide greater statistical power in detecting extra-Poisson variation (Molla & Muniswamy, 2012).

Taken together, the deviance and Pearson χ^2 statistics, and the dispersion parameter constitute a well-established framework for evaluating negative binomial models. These methods have been consistently endorsed in the econometric and statistical literature (Cameron & Trivedi, 1998; Hilbe, 2011), and remain standard practice for ensuring that fitted models provide both an adequate description of the data and a statistically justified improvement over simpler alternatives.

5 Conclusion

This study successfully achieved its defined objectives, yielding clear and actionable conclusions for the management of *Meloidogyne graminicola* in Sri Lankan rice cultivation. The mapping of pest distribution revealed a highly heterogeneous but distinct geographical pattern, with the Southern and North Central provinces emerging as unequivocal hotspots. Districts such as Hambantota reported extremely high disease incidence (95.2% of plants infected) and severe gall formation (a median of 10.6 galls per plant). Similarly, agro-ecological regions DL5 and IL1b showed critically high incidence rates of up to 96.7%. These findings provide a precise spatial guide for prioritizing surveillance and resource allocation, confirming that the pest burden is not uniform but concentrated in specific, identifiable zones.

The identification of factors contributing to pest spread revealed a hierarchy of agronomic and environmental influences. The absence of field flooding emerged as the single most important contributor, creating conditions for rapid pest proliferation. The use of major irrigation systems was also found to facilitate spread, with the highest mean disease incidence (76.4%) observed under such conditions. Direct seeding increased vulnerability compared to transplanting, while clay soil acted as a mitigating factor by reducing infestation. These results confirm that pest spread is not inevitable but is largely shaped by modifiable management practices.

In relation to resistant varieties, the study established a clear classification among commonly grown rice varieties. Bg 450, Bg 366, and Bg 352 were identified as highly susceptible and should therefore be used cautiously in high-risk areas. In contrast, Bg 374 and Bg 379-2 demonstrated significant resistance, offering an effective genetic solution to reduce nematode damage. Promoting these resistant varieties in endemic regions represents a practical and sustainable management strategy.

The evaluation of the final negative binomial regression model further reinforced the reliability of the findings. The model demonstrated strong goodness-of-fit, with a Log-Likelihood value substantially better than the null model and a McFadden's Pseudo R-squared of approximately 0.45, indicating a high explanatory power. The statistically significant dispersion parameter ($\alpha = 2.44$, $p < 0.001$) confirmed the appropriateness of the negative binomial model over a standard Poisson, validating the presence of overdispersion in the data.

Model diagnostics also indicated stability, as all Variance Inflation Factor (VIF) values were well below the threshold of concern, ruling out multicollinearity among predictors. The analysis of significant variables provided a clear hierarchy of influential factors, with non-flooded fields, direct seeding, flat topography, sandy soil, and susceptible rice varieties contributing strongly to increased gall counts. Conversely, clay soils, resistant varieties, farmer awareness, and older plant age reduced infestation levels.

This study provides robust and statistically sound evidence on the distribution, drivers, and management strategies for *M. graminicola* in Sri Lanka. By highlighting both environmental and agronomic risk factors, as well as genetic resistance, the findings offer a reliable basis for targeted interventions. The results underscore that effective management is possible through a combination of spatially focused monitoring, adoption of resistant varieties, and careful adjustment of cultivation practices.

6 References

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data* (2nd ed.). Cambridge University Press.

Dean, C. B., & Lawless, J. F. (1989). Tests for detecting overdispersion in Poisson regression models. *Journal of the American Statistical Association*, 84(406), 467–472.

Desaeger, J. A. (2023). Rice root-knot nematode Meloidogyne graminicola (Nematoda: Chromadorea: Tylenchida: Meloidogyneidae: Meloidogyne). EENY-776/IN1350. University of Florida IFAS Extension.

Gurmu, S. (1991). Tests for detecting overdispersion in the Poisson regression model. *Journal of Business & Economic Statistics*, 9(2), 215–222.

Molla, M., & Muniswamy, D. (2012). Power comparisons of tests for overdispersion in count data. *Journal of Statistical Computation and Simulation*, 82(1), 1–17.

Smyth, G. K. (2003). Pearson's goodness of fit statistic as a score test statistic. In D. R. Goldstein (Ed.), *Statistics and science: A Festschrift for Terry Speed* (pp. 115–126). Institute of Mathematical Statistics.

7 Appendix

Table 33: Full model (step 1)

Analysis Of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-14.6097	6044.940	-11862.5	11833.25	0.00	0.9981
agro_eco_r	DL1a	1	-7.8704	1.0020	-9.8343	-5.9066	61.70	<.0001
agro_eco_r	DL1b	1	-5.3143	0.7470	-6.7785	-3.8502	50.61	<.0001
agro_eco_r	DL5	1	-2.4813	0.6955	-3.8444	-1.1181	12.73	0.0004
agro_eco_r	IL1a	1	-1.5152	1.1660	-3.8005	0.7702	1.69	0.1938
agro_eco_r	IL1b	1	-4.5907	0.7807	-6.1207	-3.0606	34.58	<.0001
agro_eco_r	IL1c	1	-5.2185	0.7348	-6.6587	-3.7782	50.43	<.0001
agro_eco_r	IM1c	1	0.3759	1.5883	-2.7371	3.4888	0.06	0.8129
agro_eco_r	WL1a	1	-6.0775	0.8551	-7.7534	-4.4016	50.52	<.0001
agro_eco_r	WL1b	1	-1.8454	1.1602	-4.1193	0.4286	2.53	0.1117
agro_eco_r	WL2a	1	-0.6056	1.0391	-2.6422	1.4309	0.34	0.5600
agro_eco_r	WL3	1	0.9866	1.6449	-2.2374	4.2105	0.36	0.5487
agro_eco_r	WM1a	1	5.2926	1.4540	2.4429	8.1423	13.25	0.0003
agro_eco_r	WM2b	1	12.6517	6044.939	-11835.2	11860.52	0.00	0.9983
climatic_z	DRY ZONE	0	0.0000	0.0000	0.0000	0.0000	.	.
climatic_z	INTERMEDIATE ZONE	0	0.0000	0.0000	0.0000	0.0000	.	.
agro_eco_z	LOW COUNTRY	0	0.0000	0.0000	0.0000	0.0000	.	.
terrain_adj	Other	0	0.0000	0.0000	0.0000	0.0000	.	.
terrain_adj	Rolling & Undulating	0	0.0000	0.0000	0.0000	0.0000	.	.
terrain_adj	Rolling, Undulating, Flat	0	0.0000	0.0000	0.0000	0.0000	.	.
terrain_adj	Rolling, undulating & hilly	0	0.0000	0.0000	0.0000	0.0000	.	.
terrain_adj	Undulating	0	0.0000	0.0000	0.0000	0.0000	.	.
rice_variety	At 362	1	20.8955	6044.939	-11827.0	11868.76	0.00	0.9972
rice_variety	Bg 300	1	23.7103	6044.939	-11824.2	11871.57	0.00	0.9969
rice_variety	Bg 304	1	19.0812	6044.939	-11828.8	11866.94	0.00	0.9975
rice_variety	Bg 352	1	21.8450	6044.939	-11826.0	11869.71	0.00	0.9971
rice_variety	Bg 357	1	23.0219	6044.939	-11824.8	11870.89	0.00	0.9970
rice_variety	Bg 358	1	20.4245	6044.939	-11827.4	11868.29	0.00	0.9973
rice_variety	Bg 366	1	22.2298	6044.939	-11825.6	11870.09	0.00	0.9971
rice_variety	Bg 369	1	20.6142	6044.939	-11827.2	11868.48	0.00	0.9973
rice_variety	Bg 374	0	5.4361	0.0000	5.4361	5.4361	.	.
rice_variety	Bg 379-2	1	17.4995	6044.939	-11830.4	11865.36	0.00	0.9977
rice_variety	Bg 450	1	22.7822	6044.939	-11825.1	11870.65	0.00	0.9970
rice_variety	Bg 94/1	1	-2.6155	10470.54	-20524.5	20519.27	0.00	0.9998
rice_variety	Bw 367	1	20.1631	6044.939	-11827.7	11868.03	0.00	0.9973
stage_of_maturity	Reproductive	1	3.0400	0.4319	2.1935	3.8865	49.54	<.0001
stage_of_maturity	Ripening	0	0.0000	0.0000	0.0000	0.0000	.	.
stage_of_maturity	Seedling	1	-0.7766	1.0098	-2.7557	1.2025	0.59	0.4418
soil_type	Clay	1	-2.0762	0.3023	-2.6688	-1.4837	47.17	<.0001
rice_ecosystem	Irrigated	1	3.7980	0.7796	2.2701	5.3259	23.74	<.0001
rice_ecosystem	Rainfed Midland	0	0.0000	0.0000	0.0000	0.0000	.	.
topography	Flat land	1	1.6989	0.4427	0.8311	2.5667	14.72	0.0001
topography	Gentle slope	1	-2.7924	0.4851	-3.7433	-1.8416	33.13	<.0001
topography	Lowland	1	1.2404	0.7907	-0.3094	2.7901	2.46	0.1167
flooded_y_n	No	1	3.3093	0.3943	2.5364	4.0821	70.42	<.0001
method_of_planting	Direct seeded	1	-2.9241	0.6288	-4.1566	-1.6916	21.62	<.0001
is_the_farmer_aware_	No	1	-0.0267	0.2908	-0.5967	0.5432	0.01	0.9268
anual_r_fmm		0	0.0000	0.0000	0.0000	0.0000	.	.
age_das		1	-0.0841	0.0193	-0.1218	-0.0463	19.02	<.0001
Dispersion		1	1.9913	0.1065	1.7930	2.2114		

Table 34: Model (step 2)

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-14.6097	6044.912	-11862.4	11833.20	0.00	0.9981
agro_eco_r	DL1a	1	-7.8704	1.0020	-9.8343	-5.9066	61.70	<.0001
agro_eco_r	DL1b	1	-5.3143	0.7470	-6.7785	-3.8502	50.61	<.0001
agro_eco_r	DL5	1	-2.4813	0.6955	-3.8444	-1.1181	12.73	0.0004
agro_eco_r	IL1a	1	-1.5152	1.1660	-3.8005	0.7702	1.69	0.1938
agro_eco_r	IL1b	1	-4.5907	0.7807	-6.1207	-3.0606	34.58	<.0001
agro_eco_r	IL1c	1	-5.2185	0.7348	-6.6587	-3.7782	50.43	<.0001
agro_eco_r	IM1c	1	0.3759	1.5883	-2.7371	3.4888	0.06	0.8129
agro_eco_r	WL1a	1	-6.0775	0.8551	-7.7534	-4.4016	50.52	<.0001
agro_eco_r	WL1b	1	-1.8454	1.1602	-4.1193	0.4286	2.53	0.1117
agro_eco_r	WL2a	1	-0.6056	1.0391	-2.6422	1.4309	0.34	0.5600
agro_eco_r	WL3	1	0.9866	1.6449	-2.2374	4.2105	0.36	0.5487
agro_eco_r	WM1a	1	5.2926	1.4540	2.4429	8.1423	13.25	0.0003
agro_eco_r	WM2b	1	12.6517	6044.912	-11835.2	11860.46	0.00	0.9983
rice_variety	At 362	1	20.8956	6044.912	-11826.9	11868.71	0.00	0.9972
rice_variety	Bg 300	1	23.7103	6044.912	-11824.1	11871.52	0.00	0.9969
rice_variety	Bg 304	1	19.0812	6044.912	-11828.7	11866.89	0.00	0.9975
rice_variety	Bg 352	1	21.8450	6044.912	-11826.0	11869.65	0.00	0.9971
rice_variety	Bg 357	1	23.0219	6044.912	-11824.8	11870.83	0.00	0.9970
rice_variety	Bg 358	1	20.4245	6044.912	-11827.4	11868.23	0.00	0.9973
rice_variety	Bg 366	1	22.2298	6044.912	-11825.6	11870.04	0.00	0.9971
rice_variety	Bg 369	1	20.6142	6044.912	-11827.2	11868.42	0.00	0.9973
rice_variety	Bg 374	0	5.4362	0.0000	5.4362	5.4362	.	.
rice_variety	Bg 379-2	1	17.4995	6044.912	-11830.3	11865.31	0.00	0.9977
rice_variety	Bg 450	1	22.7822	6044.912	-11825.0	11870.59	0.00	0.9970
rice_variety	Bg 94/1	1	-2.6155	10470.53	-20524.5	20519.24	0.00	0.9998
rice_variety	Bw 367	1	20.1631	6044.912	-11827.6	11867.97	0.00	0.9973
stage_of_maturity	Reproductive	1	3.0400	0.4319	2.1935	3.8865	49.54	<.0001
stage_of_maturity	Ripening	0	0.0000	0.0000	0.0000	0.0000	.	.
stage_of_maturity	Seedling	1	-0.7766	1.0098	-2.7557	1.2025	0.59	0.4418
soil_type	Clay	1	-2.0762	0.3023	-2.6688	-1.4837	47.17	<.0001
rice_ecosystem	Irrigated	1	3.7980	0.7796	2.2701	5.3259	23.74	<.0001
rice_ecosystem	Rainfed Midland	0	0.0000	0.0000	0.0000	0.0000	.	.
topography	Flat land	1	1.6989	0.4427	0.8311	2.5667	14.72	0.0001
topography	Gentle slope	1	-2.7924	0.4851	-3.7433	-1.8416	33.13	<.0001
topography	Lowland	1	1.2404	0.7907	-0.3094	2.7901	2.46	0.1167
flooded_y_n	No	1	3.3093	0.3943	2.5364	4.0821	70.42	<.0001
method_of_planting	Direct seeded	1	-2.9241	0.6288	-4.1566	-1.6916	21.62	<.0001
is_the_farmer_aware_	No	1	-0.0267	0.2908	-0.5967	0.5432	0.01	0.9268
age_das		1	-0.0841	0.0193	-0.1218	-0.0463	19.02	<.0001
Dispersion		1	1.9913	0.1065	1.7930	2.2114		

Table 35: Model (step 3)

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-14.8796	6044.102	-11861.1	11831.34	0.00	0.9980
agro_eco_r	DL1a	1	-7.8921	1.0017	-9.8554	-5.9289	62.08	<.0001
agro_eco_r	DL1b	1	-5.4705	0.7189	-6.8795	-4.0615	57.91	<.0001
agro_eco_r	DL5	1	-2.6237	0.6704	-3.9377	-1.3098	15.32	<.0001
agro_eco_r	IL1a	1	-1.8991	1.0465	-3.9502	0.1520	3.29	0.0696
agro_eco_r	IL1b	1	-4.5694	0.7794	-6.0969	-3.0418	34.37	<.0001
agro_eco_r	IL1c	1	-5.1962	0.7327	-6.6324	-3.7601	50.29	<.0001
agro_eco_r	IM1c	1	-0.3734	1.2495	-2.8224	2.0755	0.09	0.7650
agro_eco_r	WL1a	1	-6.2849	0.8088	-7.8702	-4.6996	60.38	<.0001
agro_eco_r	WL1b	1	-2.0742	1.1147	-4.2590	0.1106	3.46	0.0628
agro_eco_r	WL2a	1	-0.5811	1.0371	-2.6139	1.4517	0.31	0.5753
agro_eco_r	WL3	1	0.3600	1.4266	-2.4361	3.1561	0.06	0.8008
agro_eco_r	WM1a	1	4.7148	1.2406	2.2832	7.1463	14.44	0.0001
agro_eco_r	WM2b	1	12.9355	6044.102	-11833.3	11859.16	0.00	0.9983
rice_variety	At 362	1	21.2522	6044.102	-11825.0	11867.47	0.00	0.9972
rice_variety	Bg 300	1	23.7960	6044.102	-11822.4	11870.02	0.00	0.9969
rice_variety	Bg 304	1	19.5209	6044.102	-11826.7	11865.74	0.00	0.9974
rice_variety	Bg 352	1	22.3078	6044.102	-11823.9	11868.53	0.00	0.9971
rice_variety	Bg 357	1	22.8977	6044.102	-11823.3	11869.12	0.00	0.9970
rice_variety	Bg 358	1	20.5929	6044.102	-11825.6	11866.82	0.00	0.9973
rice_variety	Bg 366	1	22.6213	6044.102	-11823.6	11868.84	0.00	0.9970
rice_variety	Bg 369	1	20.9117	6044.102	-11825.3	11867.13	0.00	0.9972
rice_variety	Bg 374	0	5.1052	0.0000	5.1052	5.1052	.	.
rice_variety	Bg 379-2	1	17.7241	6044.102	-11828.5	11863.95	0.00	0.9977
rice_variety	Bg 450	1	23.3362	6044.102	-11822.9	11869.56	0.00	0.9969
rice_variety	Bg 94/1	1	-2.1883	10468.38	-20519.8	20515.46	0.00	0.9998
rice_variety	Bw 367	1	20.4849	6044.102	-11825.7	11866.71	0.00	0.9973
stage_of_maturity_re	Reproductive	1	2.9414	0.4130	2.1320	3.7509	50.73	<.0001
stage_of_maturity_re	Seedling_Veg	0	0.0000	0.0000	0.0000	0.0000	.	.
soil_type	Clay	1	-2.2391	0.2166	-2.6637	-1.8145	106.85	<.0001
rice_ecosystem_rec	Irrigated	1	3.5491	0.7079	2.1616	4.9365	25.14	<.0001
topography	Flat land	1	1.5113	0.3673	0.7915	2.2311	16.93	<.0001
topography	Gentle slope	1	-2.8584	0.4736	-3.7867	-1.9301	36.42	<.0001
topography	Lowland	1	0.9661	0.7011	-0.4081	2.3402	1.90	0.1682
flooded_y_n	No	1	3.2351	0.3815	2.4874	3.9827	71.92	<.0001
method_of_planting	Direct seeded	1	-3.0408	0.6073	-4.2312	-1.8505	25.07	<.0001
is_the_farmer_aware_	No	1	-0.1020	0.2694	-0.6301	0.4261	0.14	0.7050
age_das		1	-0.0728	0.0126	-0.0975	-0.0481	33.43	<.0001
Dispersion		1	1.9905	0.1065	1.7924	2.2106		

Table 36: Combining categories of rice variety agricultural ecological region

The FREQ Procedure

rice_variety	Frequency
At 362	690
Bg 300	30
Bg 304	30
Bg 352	90
Bg 357	30
Bg 358	60
Bg 366	60
Bg 369	30
Bg 374	60
Bg 379-2	60
Bg 450	60
Bg 94/1	30
Bw 367	60
Suwadel	30

The FREQ Procedure

agro_eco_r	Frequency
DL1a	60
DL1b	240
DL5	120
IL1a	120
IL1b	90
IL1c	120
IM1c	30
WL1a	60
WL1b	120
WL2a	180
WL3	60
WM1a	30
WM2b	30
WM3b	60

The FREQ Procedure

rice_variety_lump	Frequency	Percent
At 362	690	52.27
Bg 352	90	6.82
Bg 358	60	4.55
Bg 366	60	4.55
Bg 374	60	4.55
Bg 379-2	60	4.55
Bg 450	60	4.55
Bw 367	60	4.55
Other	180	13.64

agro_eco_r_lump	Frequency	Percent
DL1a	60	4.55
DL1b	240	18.18
DL5	120	9.09
IL1a	120	9.09
IL1b	90	6.82
IL1c	120	9.09
Other	90	6.82
WL1a	60	4.55
WL1b	120	9.09
WL2a	180	13.64
WL3	60	4.55
WM3b	60	4.55