pISSN 2597-4602
eISSN 2597-4637

# Non-Gaussian Analysis of Herbarium Specimen Damage to Optimize Specimen Collection Management

Aris Yaman [a, 1, *], Yulia Aris Kartika [a, 2], Ariani Indrawati [a, 3], Zaenal Akbar [a, 4], Lindung P. Manik [b, 5], Wita Wardani [c, 6], Tutie Djarwaningsih [c, 7], Taufik Mahendra [d, 8], Dadan R. Saleh [a, 9]

*[a] Research Center for Computing, National Research and Innovation Agency,*
*Kawasan Cisitu Bandung, Jl. Sangkuriang, Dago, Coblong, Bandung, Jawa Barat 40135, Indonesia*

*[b] Research Center for Data and Information Sciences, National Research and Innovation Agency,*
*Kawasan Cisitu Bandung, Jl. Sangkuriang, Dago, Coblong, Bandung, Jawa Barat 40135, Indonesia*

*[c] Research Center for Biosystematics and Evolution, National Research and Innovation Agency,*
*Cibinong Science Center, Jl. Raya Jakarta-Bogor, Pakansari, Cibinong, Bogor, Jawa Barat 16915, Indonesia*

*[d] Directorate of Scientific Collection Management, National Research and Innovation Agency,*
*Gedung InaCC, Cibinong Science Center, Jl. Raya Jakarta-Bogor, Cibinong, Bogor, Jawa Barat 16915, Indonesia*

*[1] aris.yaman@brin.go.id *; [2] yulia.aris.kartika@gmail.com; [3] indrawati.ariani@gmail.com, [4] zaenal.akbar@gmail.com,*
*[5] lindung.manik@gmail.com, [6] wt.wardani@gmail.com, [7] tutie_teresia@yahoo.com,*
*[8] taufikmahendra337@gmail.com, [9] dadan.rs@gmail.com*
** corresponding author*

ARTICLE INFO

*Article history:*
Received 16 November 2021
Revised 17 March 2022
Accepted 4 June 2022
Published online 7 November 2022

*Keywords:*
Counted data
Damage Analysis
Herbarium Specimen
NBR
Poisson Regression

ABSTRACT

Damage to specimen collections occurs in practically every herbarium across the world. Hence, some precautions must be taken, such as investigating the factors that cause specimen damage in their collections and evaluating their herbarium collection handling and usage policy. However, manual investigation of the causes of herbarium collection damage requires a lot of effort and time. Only a few studies have attempted to investigate the causes of herbarium collection damage. So far, the non-gaussian approach to detecting the causes of damage to herbarium specimens has not been studied before. This study attempted to explore the effect of species type, time, location, storage, and remounting status on the level of damage to herbarium specimens, especially those in the genus *Excoecaria*. Gaussian modeling is not good enough to model the counted data phenomenon (the amount of damage to herbarium specimens). Negative binomial regression (NBR) provides a better model when compared to generalized Poisson regression and ordinary Gaussian regression approaches. NBR detects non-uniformity in the storage process, causing damage to herbarium specimens. Natural damage to herbarium specimens is caused by differences in species and the origin of specimens.

## I. Introduction

The herbarium is no longer a place to store preserved and classified plant samples. Moreover, the herbarium has become an important supporting facility that provides valuable information on preserved flora specimens collections for many uses, especially in biodiversity. Extinct, uncommon, endemic, and common plant species are preserved in herbarium collections to serve as a reference for future study. Herbarium collections are widely used in a remarkable number of ways: to identify and discover species [1][2], to study specific biological events in the past [3][4], to understand ecological interactions [5][6], to learn about the benefits of flora such as for medication [7][8], to investigate biomolecular based on DNA [9][10], and many more uses of herbarium collections. A herbarium has to protect the herbarium specimens against loss or damage. They must provide a safe and secure environment for all specimen collections and guarantee that the collection's condition is well maintained and done according to conservation standards. However, unfortunately, pests, poor storage conditions, irresponsible handling, and other factors have significantly harmed the herbarium

https://doi.org/10.17977/um018v5i12022p1-16

collection over the years. Damage to the herbarium collection can be seen in Figure 1. These circumstances may cause bias in herbarium specimen data and uncertainty in decision-making and study outcomes.

Herbarium Bogoriense (BO) is the largest herbarium center in Southeast Asia and one of the top three in the world. This herbarium collection comprises a comprehensive collection of flowering plants, gymnosperms, ferns and lycophytes, mosses, liverworts, fungi, and many more. Nearly one million specimens from the Malesian region (Indonesia, Singapore, Malaysia, Brunei Darussalam, Timor-Leste, Papua New Guinea, and Philippines) obtained through field expeditions and gifts or exchanges between herbariums around the world [11]. The herbarium specimens, both dry and wet collections, are stored and arranged in the space provided by the curator. Collections are classified according to their respective taxons. The collection is placed separately from the collection of monocots and dicots. Arrangement of collections alphabetically by family, genus, species, and sites. Specimen sheets using acid-free paper, species folders, and genus maps. The placement of type specimens is separated from the general collection [11]. BO, one of the main reference centers for research on tropical plant taxonomy, ecology, ethnobiology, physiology, morphogenetics, and phytochemistry in the Malesian region, must ensure that all its collections are always of good quality and minimize the possibility of damage.

Keeping the herbarium collection in good condition throughout the process, from specimen collecting to storage, was challenging for the curator. In some cases, the herbarium sheet itself represents the plant, as all the plants may be lost in that place. So, protecting the sheets from fungal and insect pests is an important step. After the collection has been preserved, it should be checked regularly to ensure that the plants are healthy and free of insects or excessive dampness. Insects have the potential to destroy herbarium collections. Insects will inevitably attack the species, even with the most meticulous care and the best equipment. The curators also routinely check [12][13] the specimens to see if any specimens are damaged, especially damage caused by fungi or insects. Although preventive measures have been taken to eliminate insects and fungi that could damage the specimens, the curators still found some damaged specimens. The specimens most damaged by insects or fungi were from the genus *Excoecaria*. So, they took the initiative to investigate the factors that cause the specimens' damage in their collections. Several studies have investigated the damage. Meineke used digital herbarium specimens to study long-term insect-plant interactions [14]. For phenological research, Pearson uses machine learning on digital herbarium specimens [15].

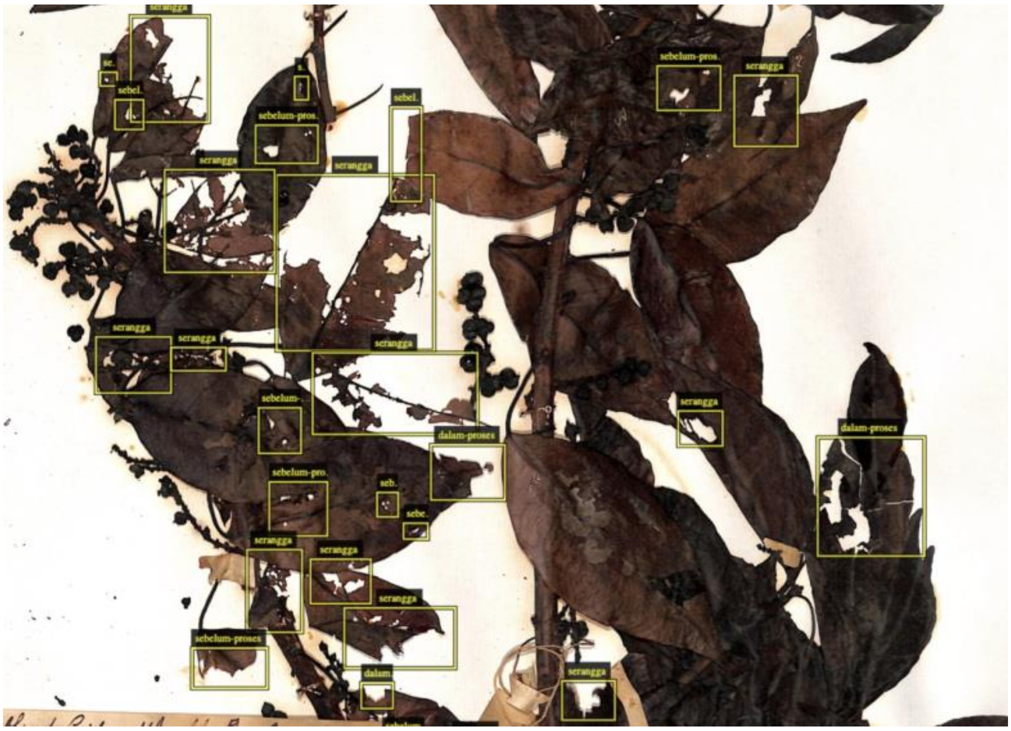


Fig. 1. Herbarium collection damage caused by natural damage, mounting or remounting process, and insect

It is a vital strategy to review and evaluate the policy of their herbarium collection handling and usage. However, manual investigation of the causes of herbarium collection damage requires a lot of effort and time. Only a few studies have attempted to investigate the causes of herbarium collection damage. Many metadata-based studies have been carried out before. Studies have been conducted to discover time series patterns and specimen distributions of genetic changes in a specimen. Studies link herbarium specimen metadata to climate change patterns [16][17][18]. On the other hand, this study looks at how labels on herbarium specimen metadata affect the damage to herbarium specimens.

The curator assesses specimen damage. If the specimen is damaged, the curator will mark the damaged area in the photo and offer details on the source of the damage. The damage marker box size varies and depends on the specimen's damage. One specimen sheet can have several flaws from various sources. Herbarium specimens are damaged in three ways: before processing (BP), in-processing (IP), and caused by insects. The first category includes damage that occurred before collection (i.e., damage caused by natural forces in nature). The second category includes damage that occurred during the collection or remounting of herbarium specimens (in-process collecting damage). Insect damage is the last type of damage that can occur to herbarium specimens.

Damage identification in a herbarium specimen is based on the number of damaged spots and the source of damage (BP, IP, or insect). Thus, the study's response variable is counted data. So linear regression cannot be used to model the phenomena in this investigation. The Generalized Linear Model (GLM) can model data with non-linear characteristics. GLM modeling requires three essential components: random, systematic, and link functions [19]. Non-linear regression with counted data is achievable using Generalized Poisson and Negative Binomial Regression [20]. Generalized Poisson Regression (GPR) is suitable for modeling with counted data [20]. The Generalized Poisson distribution is used to distribute the response variables in the GPR model (GPD). This GPD can model overdispersion and underdispersion well [20][21]. Negative Binomial Regression can also be used to model counted data. The negative binomial distribution is a Poisson-Gamma mixed function. It can accommodate overdispersion in Poisson regression because it does not require equidispersion [20][22].

## II. Methods

The stages of analysis in this study are depicted in Figure 2. The first step is the herbarium damage quantification specimen. At this stage, we annotate each type of damage per herbarium specimen. In the second stage, we will evaluate whether the three types of damage are multivariate phenomena (identification through the correlation value of each pair of types of damage). Multivariate modeling will be carried out if there is a significant correlation between each pair of types of damage. Otherwise, univariate modeling will be carried out. The next stage is modeling with non-Gaussian regression. At this stage, modeling two types of non-Gaussian regression (NBR and Poisson) is carried out. As a comparison, Gaussian regression modeling is still being carried out. In the last stage, we will evaluate the model based on the results obtained from the previous stage. AIC parameters are used to evaluate the best type of modeling.

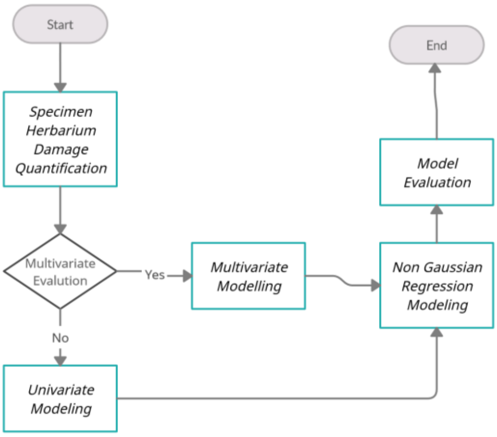


Fig. 2. Research analysis flow chart

### *A. Specimen Overview*

Recently, the scientific curator of BO reported that his collection was damaged. Several genera were damaged, such as *Antidesma*, *Baccaurea*, *Breynia*, *Excoecaria*, etc. However, the most damage occurred in the genus *Excoecaria*. In that genus, curators found 2,146 defects in 175 *Excoecaria* specimens. It includes damage from nature, damage from mounting or remounting, and damage caused by insects, all types of damage that can happen.

*Excoecaria* is a genus of plants in the Euphorbiaceae family [23]. *Excoecaria* is derived from the Latin word excaeco, which means "to blind," and refers to the sap of some plants that can induce temporary blindness [24]. *Excoecaria* is Shrubs or trees with milky latex, glabrous, monoecious, or dioecious. Leaves alternate with two glands at the petiole-lamina junction. Inflorescences have a spike or raceme with flowers clustered in the axils of bracts; female inflorescences are shorter than males. Perianth segments 2 or 3. Stamens 2 or 3, filaments basally fused. Ovary 2-or 3-locular, solitary ovule in each loculus; style 3, linear, free. 3-lobed capsules. The milky latex irritates the skin and can cause injury and blindness if applied to the eye. Distribution and frequency of occurrence: 40 species worldwide, from tropical Africa to Malaysia and Australia [25]. There can be only one cause of damage on a specimen, but there can also be more than one source of damage. Examples of specimens that suffered damage caused by a single source of damage can be seen in Figure 1.

### *B. Specimen Herbarium Damage Quantification*

We quantified herbivory on a few genus *Excoecaria* specimens collected in Indonesia, New Guinea, Malaysia, and the Philippines and preserved within the Herbarium Bogoriense. We chose the genus *Excoecaria* because specimens from the genus *Excoecaria* were the most damaged in the Herbarium Bogoriense.

The curator assesses specimen damage. If there is damage to the specimen, the curator will put a checkmark on the damaged part in the specimen photo and provide information on the source of the damage. The size of the damaged marker box is not uniform. The size of the damaged marker box depends on the size of the damaged part of the specimen. One specimen sheet can consist of one or more defects with different sources of damage. The causes of damage to herbarium specimens are classified into three categories. First, damage that occurred before the specimen collection process or damage caused by natural factors in nature (natural damage). The second damage cause was identified as damage caused during the collection process or remounting herbarium specimens (in the process of collecting damage). The last cause or source of damage is herbarium specimen damage caused by insects at the specimen storage location (damage by insects).

Differentiating between pre-collection and post-collection herbivory on herbarium specimens is a challenge. Pre-collection herbivory on the leaves of some plant species can be distinguished by the presence of a thin and darkening contour around the damaged area. It means the plant was still alive when the herbivory killed the cells in a specific area [6]. If localized cell death does not occur surrounding the injured area, post-collection herbivory or storage-related damage is assumed [26]. We discovered leaf damage morphology in *Excoecaria* was similar before and after collection, so we used the same method to distinguish pre-collected herbivores and used the curator's opinion to differentiate pre- and post-collection damage.

The specimen damage due to the mounting or remounting process is usually indicated by the presence of an envelope attached to the specimen sheet. The envelope helps accommodate broken stems or torn leaf pieces. Process-damaged leaves are often seen at the leaf tips or margins, not on the inside of the leaves. One of the causes of leaf damage during the process is leaf folds during the drying process, which causes the leaf shape to become imperfect. In addition, the leaves and stems are ripped or broken during the transfer procedure from the old specimen paper to the new specimen paper because of their fragility.

### *C. Statistical Analysis*

This study was divided into three causes of damage to herbarium specimens (as a response variable). First, damage that occurred before the specimen collection process or damage caused by natural factors in nature (natural damage/BP). The second damage cause was identified as damage caused during the collection process or remounting herbarium specimens (in collecting damage/IP).

The last cause or source of damage is herbarium specimen damage caused by insects at the specimen storage location (damage by insects).

Systematic identification of damage in a herbarium specimen is based on the number of damage spots along with identifying the source of damage (BP, IP, and caused by insect). Based on this, the response variable in the study is the counted data. The Kolmogorov-Smirnov test was applied to assess distribution fit inferentially [27]. So, it cannot use the usual linear regression approach to model the phenomena in this study. The Generalized Linear Model (GLM) approach can model data whose parameters are not linear. Modeling with GLM requires three main components: a random component, a systematic component, and a link function [19]. There are at least two non-linear regression approaches with the counted data in response: Generalized Poisson Regression and Negative Binomial Regression [20].

Generalized Poisson Regression (GPR) has been proven to be good in modeling the response variable in the form of counted data [20]. As the name implies, the response variables in the GPR model are distributed according to the Generalized Poisson distribution (GPD). This GPD is good at modeling overdispersion and under-dispersion data conditions [20][21]. Another approach to modeling the counted data is Negative Binomial Regression. In this study, the negative binomial distribution is a mixed function between Poisson-Gamma. The gamma distribution can accommodate overdispersion in Poisson regression because it does not assume equi-dispersion conditions in its application [20][22].

This study attempted to explore the effect of species type, time, location, storage, and remounting status on the level of damage to herbarium specimens (especially those in the genus *Excoecaria*). In all models, the response was the total number of spots with BP, IP, and caused by insect damage to herbarium specimens (HS). The models were defined as:

Number of spots damage before collecting process (BP):

$$logit(BP) = \alpha + \beta_1(Species\ of\ HS) + \beta_2\ (Ages\ of\ HS) + \beta_3(Origin\ of\ HS) \tag{1}$$

Number of spots damage caused by collecting process (IP):

$$\begin{aligned} logit\ (IP) = a + \beta_1(Species\ of\ HS) + \beta_2\ (ages\ of\ HS) + \beta_3\ (Origin\ of\ HS) + \\ \beta_5\ (Storage\ Location\ of\ HS) + BP + Insect \end{aligned} \tag{2}$$

Number of spots damage caused by insect at storage collection (Insect):

$$\begin{aligned} logit\ (Insect) = a + \beta_1\ (Species\ of\ HS) + \beta_2\ (ages\ of\ HS) + \beta_3\ (Origin\ of\ HS) + \\ \beta_4\ (Remounting\ Status\ of\ HS) +\ \beta_5\ (Storage\ Location\ of\ HS) + \\ BP + IP \end{aligned} \tag{3}$$

As shown in the above equation, there are three models of the level of damage to herbarium specimens. The first model, logit (BP) is a function of variable a, intercept, species type (categorical variable), age of collection (numeric variable), and origin of species (categorical variable). The second model, logit (IP)/level of damage due to the collection/remounting process, is a function of variables a, intercept, species type (categorical variable), age of collection (numeric variable), the origin of species (categorical variable), collection storage location, number of damage caused before collection (BP), and number of damage caused by insects in storage collection (categorical variables). Precisely for this second model, the samples used in the modeling are herbarium specimens that have undergone a remounting process. The third model, logit (insect), is a function of variables $\alpha$ and intercept, species type (categorical variable), age of collection (numeric variable), origin of species (categorical variable), collection storage location (categorical variable), remounting status, number of damaged before the collecting process (BP), and the number of damaged insects in the storage collection (insect).

This study observed four species belonging to the genus *Excoecaria*, namely: *Excoecaria agallocha*, *Excoecaria cochinchinensis*, *Excoecaria humilis*, and *Excoecaria oppositifolia*. The origin of the specimens in the study was spread across nine locations, including Borneo, Celebes, Java, Kawasan_II, Malaypen, Moluccas, New Guinea, the Philippines, and Sumatra. Meanwhile, there are nine different collection storage locations in the focus of this research. The explanatory variable for remounting status is a variable that states whether a specimen has experienced remounting or not before.

This study's explanatory variables are descriptions or labels (metadata) in a herbarium specimen. The data cleansing stage produced as many as 175 herbarium data specimens (which could be further analyzed). Furthermore, this study's entire sample of specimens will be modeled into three models described previously. A pre-analysis was conducted to see the relationship pattern between the response variables (BP, IP, and insects). If there is a significant correlation between them, it is necessary to do multivariate modeling. On the other hand, if there is no significant correlation between the response variables, it is sufficient to do univariate modeling (partial modeling for each response variable).

After assessing the closeness of the relationship between the response variables, the stages of statistical analysis are modeled with several modeling schemes, including modeling based on GPR or Negative Binomial Regression. As a comparison, modeling based on simple multiple linear regression is also carried out. The AIC (Akaike Information Criteria) parameter is used to assess which model best models the phenomena in this study. The lower the AIC value, the better the resulting model for modeling the phenomena contained in the study [22].

After obtaining the best model based on the lowest AIC value, the next stage tests to see which explanatory variables significantly affect the built model. This study applies a partial F test to see which explanatory variables significantly impact the model. The partial F test is a test that compares the full model (a model with all explanatory variables) with a partial model (a model without one of the explanatory variables, which will be tested). The logic is built to see the change in goodness models if one of the explanatory variables is omitted [28]. However, the Wald test was used for categorical variables to see which level of the categorical variables had the most significant impact on the damage to herbarium specimens [29].

## III. Results and Discussion

### *A. Exploratory Data Analysis*

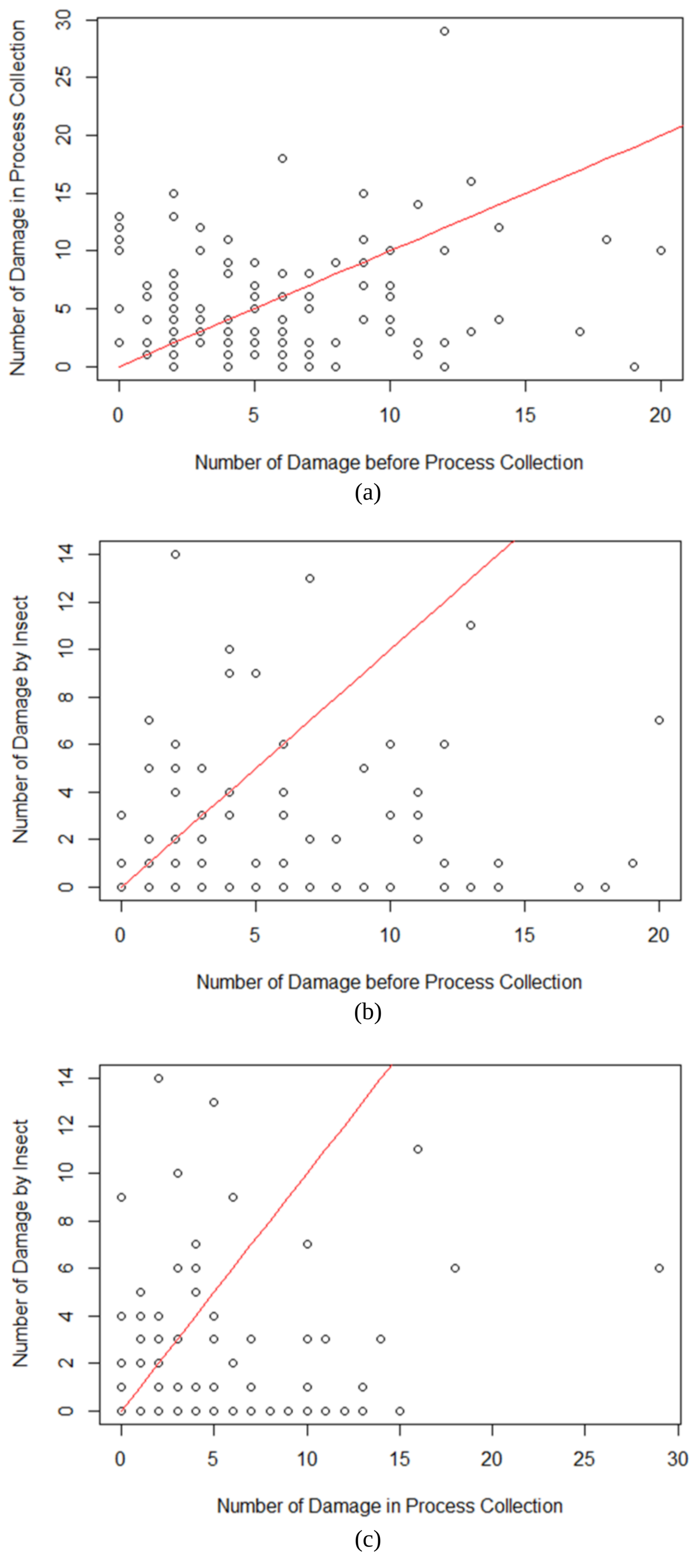
In this study, the causes of damage were divided into three categories: firstly, the cause of damage is natural processes that occur while the specimen is still in nature (natural damage/before the collecting process). Secondly, the damage caused during the specimen collection process (in-process damage), and the third was the damage to herbarium specimens caused by insects at the collection storage location (preservation damage by insects). In order to determine the modeling procedure later, the first step is to evaluate the correlations among the various causes of damage. This evaluation is intended to determine whether there is a correlation between the sources of damage. When there is a significant relationship between response variables, it is better to carry out a multivariate analysis procedure. On the other hand, if there is no correlation between the response variables (the source of the damage to the specimen), then partial modeling (univariate analysis) is carried out.

Table 1 shows the correlation between sources causing damage to herbarium specimens, with a P-value exceeding α (5%), which indicates no significant correlation between the response variables. It indicates that there is no significant correlation between the response variables. So, a partial analysis procedure (univariate analysis) was applied in this study.

Figure 3 shows a comparison plot of the number of damage events for each pair of sources causing damage to herbarium specimens: (a) between before process (natural damage) and in-process damage; (b) between natural damage and preservation damage by insects; and (c) between in-process and preservation damage by insects. The picture shows the number of damage points on the herbarium specimens. Due to the collection process, the distribution pattern of damage points on herbarium specimens looks the same as the distribution pattern of collection damage points due to

Table 1. Correlation between response variables (source of damage)

| Correlation/P-Value | Num_Damage_BP | Num_Damage_IP |
|---|---|---|
| Num_Damage_IP | 0.146 | - |
| | 0.104 | - |
| Num_Damage_Insect | 0.069 | 0.069 |
| | 0.440 | 0.441 |



(a)

(b)

(c)

Fig. 3. Scatterplot for each pair cause damage to the herbarium specimen: (a) natural damage & in-process damage, (b) natural damage & damage by insect, (c) in-process damage & damage by insect

natural factors (natural damage). Insects in the storage of herbarium collections caused less damage to the collections than the other two causes.Specimen Distribution

Collection dates for Excoecaria specimens in the Herbarium Bogoriense (BO) span 154 years, from 1866 through 2020. Most collections (36; 27; and 24 out of 175) were collected from Java Island, Sumatra Island, and Moluccas, and only five samples came from Malaysia-Peninsula. The low sample size for the Malaysia-Peninsula region could have caused a bias caused by the non-representation of the region [26].

*Excoecaria agallocha* is the most abundant species in the collection to be analyzed in this study (121 out of 175). Meanwhile, there were only four samples of *Excoecaria oppositifolia*. Because these specimens were given to the herbarium Bogoriense by other researchers, there is a low number of specimens from specific species and places. Figure 4 shows that the existing data are not normally distributed.

Figure 5 shows the distribution of damage for each of the analyzed species. Figure 5a shows that the highest level of damage before the collection process occurred in *Excoecaria cochinchinensis*. However, we cannot conclude that this species was the most severely damaged before the collection process. In the box plot, there are slices of the same amount of damage as *Excoecaria agallocha* and *Excoecaria humilis*. In contrast to the pattern of damage caused by the remounting process (Figure 5b), it is seen that tremendous damage occurred in *Excoecaria oppositifolia*. The way that tends to be homogeneous occurs in the damage caused by insects in the collection storage area (Figure 5c). Visually, for each species, the level of damage tends to be the same. These visual findings need to be clarified inferentially. It is to obtain valid conclusions.

Visually exploring whether differences in specimen origin affect the damage to herbarium specimens is shown in Figure 6, which shows no significant differences between the origin of the specimen and the degree of damage (Figure 6a and Figure 6c). Different things can be seen in Figure 6b, it can be seen that specimens from Malaysia-peninsula have the highest level of damage compared to other specimens from the origin. Similar to the species variable, the specimen origin variable needs to be tested for inference to see a valid level of significance for the damage level of the specimen. Other variables also need to be clarified regarding their influence on specimen damage at the modeling stage.

### *B. Model Fitting*

The normality distribution test for each damage cause is a critical process that must be performed to select the suitable model for analysis. Because the distribution of damage occurrences for all causes of specimen collecting damage is not normally distributed, as shown in Figure 4, Poisson or Negative Binomial models can be utilized in this investigation. Table 2 shows the Kolmogorov Smirnov distribution fittest results for those models.

The P-Value on the Negative Binomial exceeded 5% for all sources of damage, indicating that the Negative Binomial is the best model to study the factors that cause specimen collection damage. The AIC value comparison between Multiple Linear Regression, Generalized Poisson Regression, and Negative Binomial Regression confirms it. The Negative Binomial Regression approach obtains the AIC optimal score (the last one), as shown in Table 3.

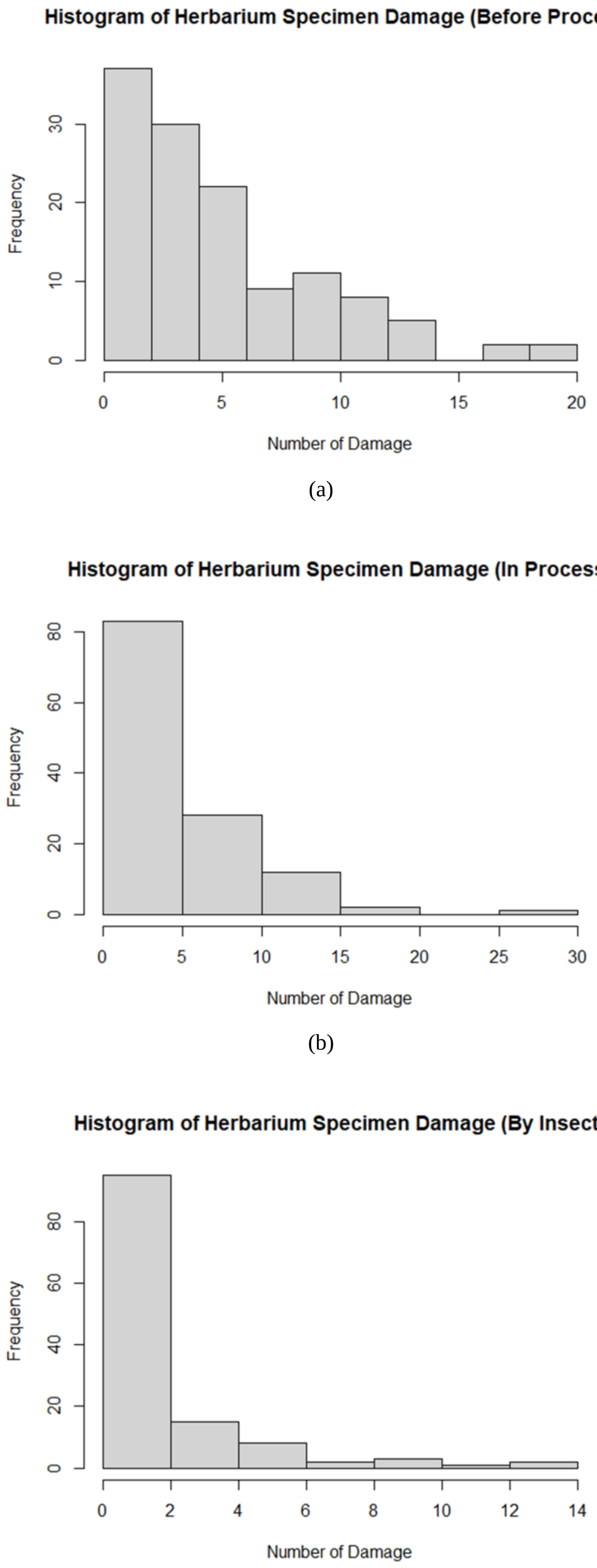
Table 2. Goodness-of-fit test for distribution

| Response Variable | P-Value | | |
|---|---|---|---|
| | Normality Test | Poisson Test | Negative Binomial Test |
| Num. of Damage before Process | 0.0032 | 4.536627e-24 | 0.363 |
| Num. of Damage in Process Collection | 0.0006 | 5.398994e-42 | 0.248 |
| Num. of Damage by Insect | 1.829e-09 | 2.897935e-36 | 0.437 |

*Alternative hypothesis: data not distributed as a test (P-value > α, accept alternative hypothesis)*

Table 3. Goodness-of-fit Model (AIC)

| Response Variable | AIC | | |
|---|---|---|---|
| | Multiple Linear Regression | Generalized Poisson Regression | Negative Binomial Regression |
| Num. of Damage before Process | 728.87 | 768.85 | 681.24 |
| Num. of Damage in Process Collection | 752.10 | 789.58 | 678.58 |
| Num. of Damage by Insect | 623.68 | 537.20 | 426.52 |



(a)

(b)

(c)

Fig. 4. Histogram of each response variable: (a) natural damage, (b) in-process damage, (c) damage by insect

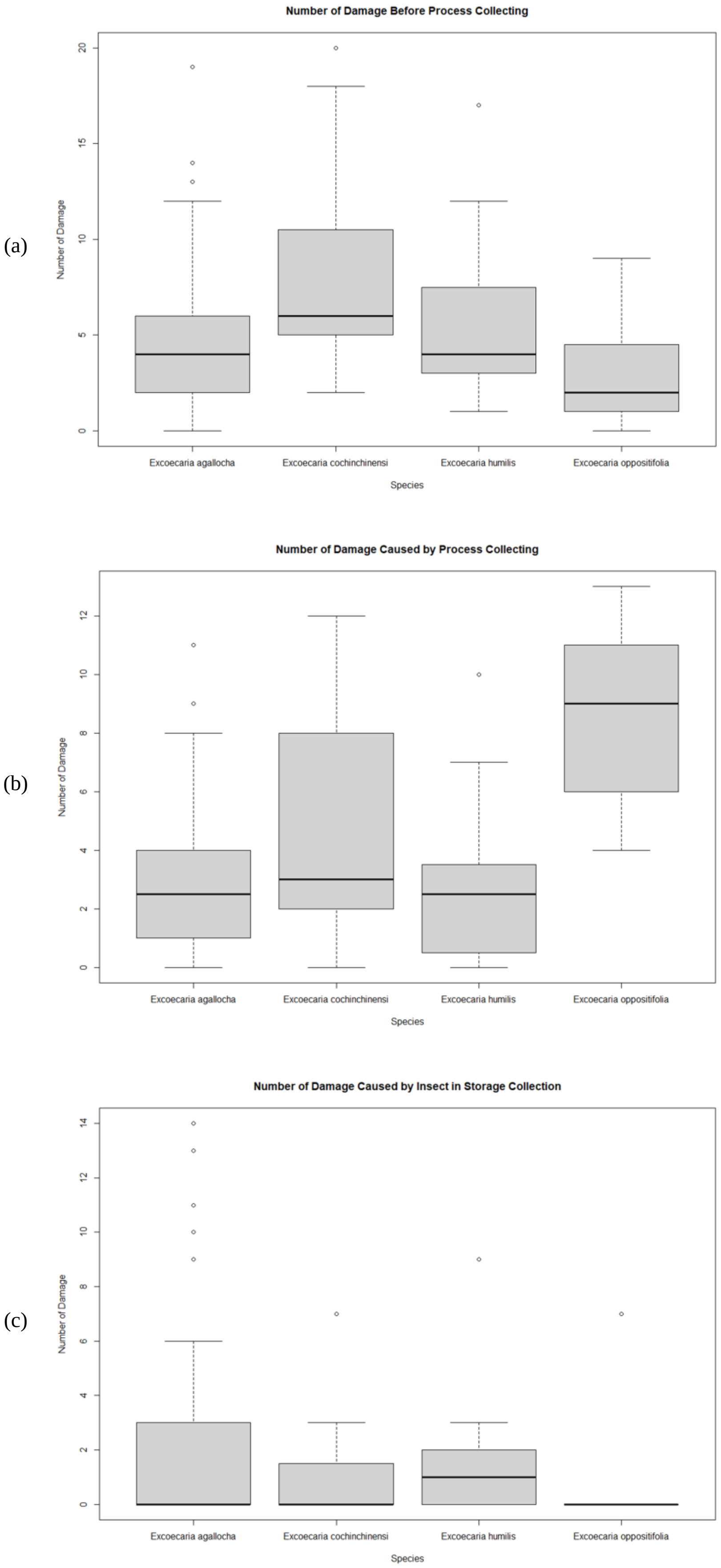


Fig. 5. Distribution of damage for each species: (a) natural damage, (b) in-process damage, (c) damage by insect

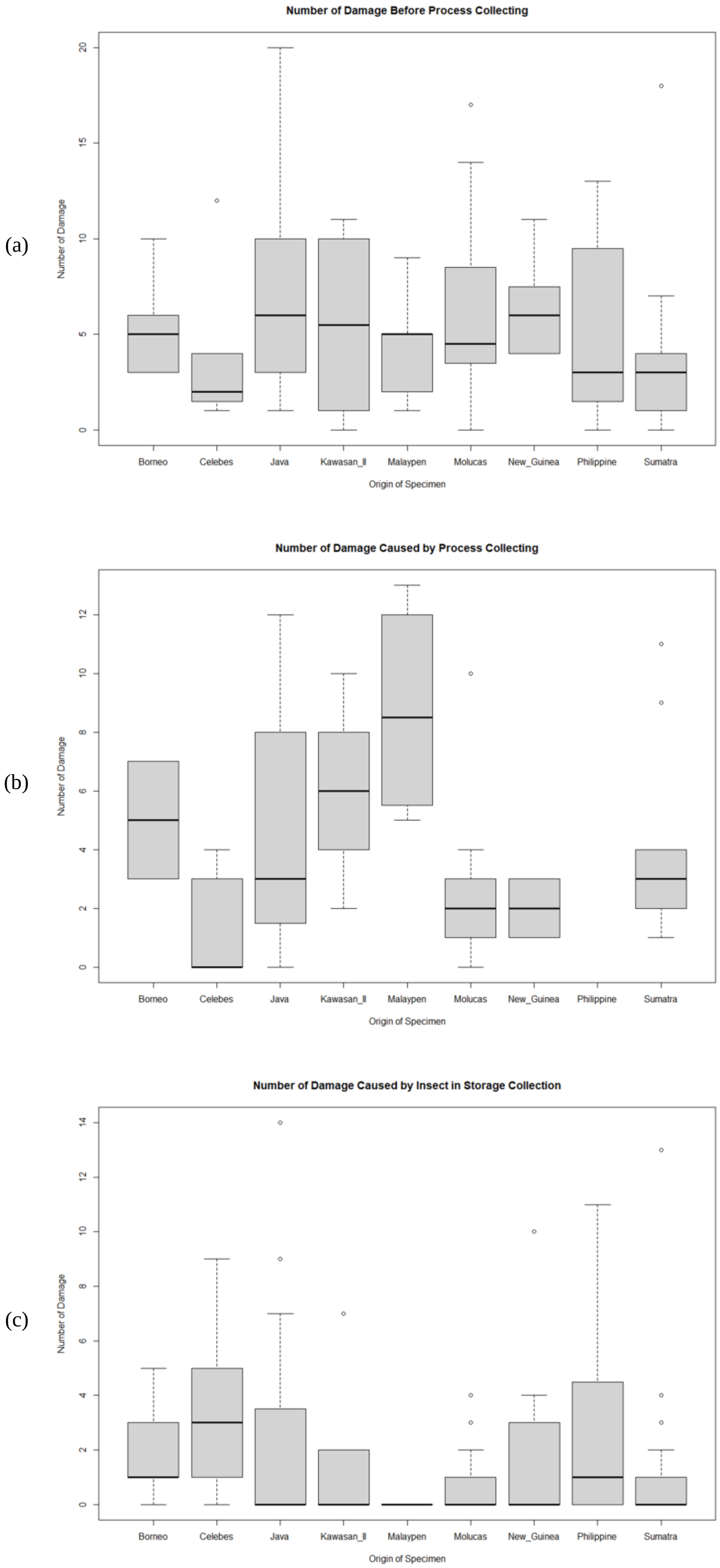


Fig. 6. Distribution of damage for each origin of specimen: (a) natural damage, (b) in-process damage, (c) damage by insect

### *C. Statistical Modeling*

The modeling in Table 4 (partial F-Test) shows that the explanatory variables of specimen origin and species significantly affect the level of specimen damage before the collection process. The Wald test was carried out as shown in Table 5. This test was to see which group significantly affected the level of specimen damage before the collection process on each explanatory variable. In addition, this test also shows the direction of influence of each explanatory variable.

*Excoecaria cochinchinensi* is a species that significantly affects the damage to herbarium specimens (BP). A positive value in the estimated coefficient of this variable indicates that this species has a higher vulnerability to damage than the other three species. Natural damage was more common in the specimens of *E. cochinchinensis* than in the other three species.

Table 6 shows that the difference in storage places significantly affects the damage during the remounting process. It indicates that different storage locations can affect the level of specimen damage due to this technical factor (remounting). No_PH7 has a higher level of damage due to remounting than other storage areas (see Table 7).

Modeling with the response variable of the level of damage due to insects at the storage location shows that only the explanatory variables of the storage area and the level of natural damage have a significant effect (Table 8).

The Wald test in Table 9 shows the direction of the influence of the variable level of natural damage and the specimen's storage place. It is seen that the more damaged the specimen is due to natural factors, the higher the level of damage due to insects in the storage location. Meanwhile, locations No_PH10 and No_PH15 significantly adversely affected the level of specimen damage due to insects. It means that both storage areas have a lower level of damage than other storage

Table 4. Partial F-test effect for predictor variable (response variable: natural damage before collecting process/BP)

| | **Model** | **theta Resid.** | **df** | **2 x log-lik.** | **Test** | **df** | **LR stat.** | **Pr(Chi)** |
|---|---|---|---|---|---|---|---|---|
| 1 | age_specimen + species | 2.59 | 121 | -667.05 | | | | |
| 2 | Origin_Spec + species | 3.02 | 114 | -653.29 | | | | |
| 3 | Origin_Spec + age_specimen | 2.67 | 116 | -664.22 | | | | |
| 4 | Origin_Spec + age_specimen + species | 3.02 | 113 | -653.24 | 1 vs 4 | 8 | 13.81 | 0.09 |
| | | | | | 2 vs 4 | 1 | 0.06 | 0.81 |
| | | | | | 3 vs 4 | 3 | 10.98 | 0.01 |

*Alternative hypothesis: have significant effect, (P-value/Pr(Chi) < α, accept alternative hypothesis)*

Table 5. Wald test for response variable: natural damage before collecting process

| **Coefficients** | **Estimate** | **Std. Error** | **z value** | **Pr(>\|z\|)** | |
|---|---|---|---|---|---|
| (Intercept) | 1.416783 | 0.398122 | 3.559 | 0.000373 | *** |
| Origin_SpecCelebes | -0.325470 | 0.422238 | -0.771 | 0.440806 | |
| Origin_SpecJava | 0.277354 | 0.347138 | 0.799 | 0.424306 | |
| Origin_SpecKawasan_II | 0.149753 | 0.521106 | 0.287 | 0.773826 | |
| Origin_SpecMalaypen | 0.505655 | 0.617376 | 0.819 | 0.412764 | |
| Origin_SpecMolucas | 0.188239 | 0.333341 | 0.565 | 0.572276 | |
| Origin_SpecNew_Guinea | 0.371863 | 0.415613 | 0.895 | 0.370929 | |
| Origin_SpecPhilipphine | 0.150636 | 0.410971 | 0.367 | 0.713964 | |
| Origin_SpecSumatra | -0.360010 | 0.345490 | -1.042 | 0.297398 | |
| age_specimen | 0.000612 | 0.002635 | 0.232 | 0.816367 | |
| speciesExcoecaria cochinchinensi | 0.409421 | 0.196950 | 2.079 | 0.037635 | * |
| speciesExcoecaria humilis | 0.367186 | 0.247606 | 1.483 | 0.138090 | |
| speciesExcoecaria oppositifolia | -0.716610 | 0.529615 | -1.353 | 0.176028 | |

---
*Signif. codes: 0’***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘’ 1*

Table 6. Partial F-Test effect for predictor variable (response variable: damage caused by remounting process/IP)

| | Model | theta Resid. | df | 2 x log-lik. | Test | df | LR stat. | Pr(Chi) |
|---|---|---|---|---|---|---|---|---|
| 1 | age_specimen + No_PH + species + Num_Damage_BP + Num_Damage_Insect | 3.68 | 53 | -301.92 | | | | |
| 2 | Origin_Spec + No_PH + species + Num_Damage_BP + Num_Damage_Insect | 4.74 | 48 | -290.73 | | | | |
| 3 | Origin_Spec + age_specimen + species + Num_Damage_BP + Num_Damage_Insect | 3.52 | 51 | -303.27 | | | | |
| 4 | Origin_Spec + age_specimen + No_PH + Num_Damage_BP + Num_Damage_Insect | 4.53 | 48 | -291.27 | | | | |
| 5 | Origin_Spec + age_specimen + No_PH + species + Num_Damage_Insect | 4.41 | 48 | -292.07 | | | | |
| 6 | Origin_Spec + age_specimen + No_PH + species + Num_Damage_BP | 4.91 | 48 | -289.80 | | | | |
| 7 | Origin_Spec + age_specimen + No_PH + species + Num_Damage_BP + Num_ Damage_Insect | 4.96 | 47 | -298.77 | 1 vs 7 | 6 | 12.15 | 0.059 |
| | | | | | 2 vs 7 | 1 | 0.96 | 0.328 |
| | | | | | 3 vs 7 | 4 | 13.50 | 0.009 |
| | | | | | 4 vs 7 | 1 | 1.50 | 0.220 |
| | | | | | 5 vs 7 | 1 | 2.30 | 0.129 |
| | | | | | 6 vs 7 | 1 | 0.03 | 0.852 |

Table 7. Wald test for response variable: damage caused by remounting process

| Coefficients: | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 38.71 | 47450000 | 0 | 1 | |
| Origin_SpecCelebes | -38.35 | 47450000 | 0 | 1 | |
| Origin_SpecJava | -38.52 | 47450000 | 0 | 1 | |
| Origin_SpecKawasan_II | -1.311 | 1.029 | -1.274 | 0.2026 | |
| Origin_SpecMalaypen | -1.046 | 1.097 | -0.954 | 0.3402 | |
| Origin_SpecMolucas | -1.491 | 0.7259 | -2.053 | 0.04 | * |
| Origin_SpecNew_Guinea | -1.562 | 1.002 | -1.559 | 0.119 | |
| Origin_SpecSumatra | -1.005 | 0.748 | -1.343 | 0.1792 | |
| age_specimen | 0.004875 | 0.004903 | 0.994 | 0.3201 | |
| No_PHPH8 | 1.188 | 0.4783 | 2.483 | 0.013 | * |
| No_PHPH12 | -37.37 | 47450000 | 0 | 1 | |
| No_PHPH13 | -37.04 | 47450000 | 0 | 1 | |
| No_PHPH15 | 0.7252 | 0.4267 | 1.7 | 0.0892 | |
| No_PHPH16 | -36.83 | 47450000 | 0 | 1 | |
| No_PHPH17 | -35.89 | 47450000 | 0 | 1 | |
| speciesExcoecaria humilis | -0.7847 | 0.641 | -1.224 | 0.2209 | |
| Num_Damage_BP | 0.03377 | 0.02226 | 1.517 | 0.1292 | |
| Num_Damage_Insect | -0.008157 | 0.0422 | -0.193 | 0.8467 | |

---
*Signif. codes: 0'***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1*

areas.

### D. Discussion

*Excoecaria cochinchinensi* is a species that significantly affects the damage to herbarium specimens (BP). This species has the highest level of damage before the collection process compared to other species (the highest level of natural damage). The specimen's origin also significantly determines the level of susceptibility to damage to the specimen before undergoing the collection process. So, specimens from such locations as analyzed and *Excoecaria cochinchinensis* need to be treated more intensely in the following collection process.

Table 8. Partial F-test effect for predictor variable (response variable: preservation damage by insect)

| | Model | theta Resid. | df | 2 x log-lik. | Test | df | LR stat. | Pr(Chi) |
|---|---|---|---|---|---|---|---|---|
| 1 | age_specimen + Stat_remounting + No_PH + species + Num_Damage_BP + Num_Damage_IP | 0.49 | 111 | -389.76 | | | | |
| 2 | Origin_Spec + Stat_remounting + No_PH + species + Num_Damage_BP + Num_Damage_IP | 0.58 | 105 | -379.52 | | | | |
| 3 | Origin_Spec + age_specimen + No_PH + species + Num_Damage_BP + Num_Damage_IP | 0.58 | 105 | -379.31 | | | | |
| 4 | Origin_Spec + age_specimen + Stat_remounting + species + Num_Damage_BP + Num_Damage_IP | 0.46 | 110 | -393.18 | | | | |
| 5 | Origin_Spec + age_specimen + Stat_remounting + No_PH + Num_Damage_BP + Num_Damage_IP | 0.57 | 105 | -381.10 | | | | |
| 6 | Origin_Spec + age_specimen + Stat_remounting + No_PH + species + Num_Damage_IP | 0.55 | 105 | -383.67 | | | | |
| 7 | Origin_Spec + age_specimen + Stat_remounting + No_PH + species + Num_Damage_BP | 0.58 | 105 | -379.12 | | | | |
| 8 | Origin_Spec + age_specimen + Stat_remounting + No_PH + species + Num_Damage_BP + Num_Damage_IP | 0.59 | 104 | -379.09 | 1 vs 8 | 7 | 10.67 | 0.154 |
| | | | | | 2 vs 8 | 1 | 0.44 | 0.509 |
| | | | | | 3 vs 8 | 1 | 0.23 | 0.633 |
| | | | | | 4 vs 8 | 6 | 14.09 | 0.029 |
| | | | | | 5 vs 8 | 1 | 2.02 | 0.155 |
| | | | | | 6 vs 8 | 1 | 4.58 | 0.032 |
| | | | | | 7 vs 8 | 1 | 0.04 | 0.845 |

Table 9. Wald test for response variable : preservation damage by insect

| Coefficients: | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | 6.14E-01 | 9.51E-01 | 0.646 | 0.518 | |
| Origin_SpecCelebes | 6.40E-01 | 8.68E-01 | 0.737 | 0.461 | |
| Origin_SpecJava | 1.02E+00 | 1.03E+00 | 0.993 | 0.321 | |
| Origin_SpecKawasan_II | -3.57E+00 | 2.73E+00 | -1.308 | 0.191 | |
| Origin_SpecMalaypen | -4.12E+01 | 3.00E+07 | 0 | 1.000 | |
| Origin_SpecMolucas | -2.42E+00 | 1.75E+00 | -1.383 | 0.167 | |
| Origin_SpecNew_Guinea | -1.93E+00 | 1.88E+00 | -1.027 | 0.305 | |
| Origin_SpecPhilippine | -2.70E+00 | 1.98E+00 | -1.367 | 0.172 | |
| Origin_SpecSumatra | -3.24E+00 | 1.85E+00 | -1.756 | 0.079 | |
| age_specimen | -5.58E-03 | 6.68E-03 | -0.836 | 0.403 | |
| Stat_remountingwith_remounting | 2.34E-01 | 4.46E-01 | 0.524 | 0.600 | |
| No_PHPH8 | -1.53E+00 | 9.33E-01 | -1.639 | 0.101 | |
| No_PHPH9 | -6.21E-01 | 9.65E-01 | -0.643 | 0.520 | |
| No_PHPH10 | -2.76E+00 | 1.25E+00 | -2.21 | 0.027 | * |
| No_PHPH12 | 7.37E-01 | 1.84E+00 | 0.402 | 0.688 | |
| No_PHPH13 | 2.98E+00 | 1.92E+00 | 1.554 | 0.120 | |
| No_PHPH15 | -2.55E+00 | 7.63E-01 | -3.341 | 0.001 | *** |
| No_PHPH16 | 1.95E+00 | 1.96E+00 | 0.995 | 0.320 | |
| No_PHPH17 | 3.75E+00 | 3.22E+00 | 1.163 | 0.245 | |
| speciesExcoecaria humilis | -2.42E+00 | 1.73E+00 | -1.404 | 0.160 | |
| Num_Damage_BP | 9.64E-02 | 4.00E-02 | 2.411 | 0.016 | * |
| Num_Damage_IP | 7.55E-03 | 3.50E-02 | 0.216 | 0.829 | |

---

Signif. codes: 0’***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘’ 1

The damage caused by the remounting process on herbarium specimens is primarily due to the specimen storage area. There is a difference in the quality of the specimen storage area. It indicates the existence of non-uniformity in the management of storage media. Meanwhile, the damage caused by insects at the collection storage location is caused by the factors where the specimen is stored and the specimen level of damage before the collection process (natural damage before the collecting process). Storage areas appear to affect the rate of insect damage significantly. It indicates clearly that due to poor quality in certain storage places, in other words, the need for standardized specimen management. In addition, it can be seen that if specimens found before the collection process were damaged, they are more likely to be damaged by insects when stored.

## IV. Conclusion

This study attempted to explore the effect of species type, time, location, storage, and remounting status on the level of damage to herbarium specimens (especially those in the genus *Excoecaria*). The response was the total number of spots with BP, IP, and Insect Damage Herbarium specimens (HS) with Negative Binomial Regression (NBR), Poisson regression, and ordinary Gaussian regression approaches. The experiment shows that the typical distribution-based regression modeling approach was not practical enough in modeling the damage phenomenon in herbarium specimens. The method based on the distribution of the enumerated data (amount of damage to herbarium specimens), predominantly Negative Binomial Regression, can better model the phenomenon of damage to herbarium specimens compared to GPR modeling and ordinary Gaussian regression models.

Based on Negative Binomial Regression modeling, it was detected that there was a non-uniformity in the storage process. The storage location factor significantly positively affects damage to herbarium specimens (caused by insects and the remounting process). The procedure for storing herbarium specimens needs to be standardized. Meanwhile, damage due to natural factors is caused by factors of different types of species. BO management needs to be concerned and handle the *Excoecaria cochinchinensis* species.

This research is limited to modeling *Excoecaria cochinchinensis* species. It will probably have an impact on different species. New species will be added in the future to make the results obtained more general than the existing model.

## Declarations

### *Author contribution*

All authors contributed equally as the main contributor of this paper. All authors read and approved the final paper.

### *Funding statement*

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### *Conflict of interest*

The authors declare no known conflict of financial interest or personal relationships that could have appeared to influence the work reported in this paper.

### *Additional information*

Reprints and permission information are available at http://journal2.um.ac.id/index.php/keds.

Publisher's Note: Department of Electrical Engineering - Universitas Negeri Malang remains neutral with regard to jurisdictional claims and institutional affiliations.

## References

[1] N. O. Kin, N. Y. Demchenko, and S. N. Ryabtsov, "Rare plants of the Voronezh region in ecosystems of Khrenovsky pine forest," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 817, no. 1, 2021.

[2] A. A. Pinto, J. J. C. Mont, D. E. M. Jiménez, A. G. Noriega, J. J. Barrios, and A. C. McCormick, "Characterization of riparian tree communities along a river basin in the pacific slope of guatemala," *Forests*, vol. 12, no. 7, pp. 1–12, 2021.

[3] T. K. Miller, A. S. Gallinat, L. C. Smith, and R. B. Primack, "Comparing fruiting phenology across two historical datasets: Thoreau's observations and herbarium specimens," *Ann. Bot.*, vol. 128, no. 2, pp. 159–170, 2021.

[4] A. G. Auffret, "Historical floras reflect broad shifts in flowering phenology in response to a warming climate," *Ecosphere*, vol. 12, no. 7, 2021.

[5] L. A. Jenny, L. R. Shapiro, C. C. Davis, J. Davies, N. E. Pierce, and E. K. Meineke, "Herbarium specimens reveal herbivory patterns across the genus Cucurbita," *Bioarvix*, 2021.

[6] E. K. Meineke, A. T. Classen, N. J. Sanders, and T. Jonathan Davies, "Herbarium specimens reveal increasing herbivory over the past century," *J. Ecol.*, vol. 107, no. 1, pp. 105–117, 2019.

[7] W. Milliken, B. E. Walker, M. J. R. Howes, F. Forest, and E. Nic Lughadha, "Plants used traditionally as antimalarials in Latin America: Mining the tree of life for potential new medicines," *J. Ethnopharmacol.*, vol. 279, no. March, 2021.

[8] P. Maher *et al.*, "The Value of Herbarium Collections to the Discovery of Novel Treatments for Alzheimer's Disease, a Case Made With the Genus Eriodictyon," *Front. Pharmacol.*, vol. 11, no. March, 2020.

[9] S. Acha, A. Linan, J. MacDougal, and C. Edwards, "The evolutionary history of vines in a neotropical biodiversity hotspot: Phylogenomics and biogeography of a large passion flower clade (Passiflora section Decaloba)," *Mol. Phylogenet. Evol.*, vol. 164, no. July, p. 107260, 2021.

[10] N. Forin, A. Vizzini, F. Fainelli, E. Ercole, and B. Baldan, "Taxonomic re-examination of nine rosellinia types (Ascomycota, xylariales) stored in the saccardo mycological collection," *Microorganisms*, vol. 9, no. 3, 2021.

[11] D. Girmansyah, Y. Santika, Rugayah, and J. S. Rahajoe, *Index Herbariorum Indonesianum*. 2018.

[12] V. Bestandssituation, "Bärlappe in Thüringen –Verbreitung und Bestandssituation," *Landschaftspfl. und Naturschutz Thüringen*, vol. 52, no. 2, pp. 51–54, 2015.

[13] A. Güntsch, W. Berendsohn, and P. Mergen, "The BioCASE Project - a Biological Collections Access Service for Europe," *Ferrantia*, vol. 51, no. June 2014, pp. 103–108, 2007.

[14] E. K. Meineke, C. Tomasi, S. Yuan, and K. M. Pryer, "Applying machine learning to investigate long-term insect–plant interactions preserved on digitized herbarium specimens," *Appl. Plant Sci.*, vol. 8, no. 6, pp. 1–11, 2020.

[15] K. D. Pearson *et al.*, "Machine learning using digitized herbarium specimens to advance phenological research," *Bioscience*, vol. 70, no. 7, pp. 610–620, 2020.

[16] I. Koh *et al.*, "Modeling the status, trends, and impacts of wild bee abundance in the United States," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 1, pp. 140–145, 2016.

[17] C. Meyer, P. Weigelt, and H. Kreft, "Multidimensional biases, gaps and uncertainties in global plant occurrence information," *Ecol. Lett.*, vol. 19, no. 8, pp. 992–1006, 2016.

[18] M. A. Jamieson, A. L. Carper, C. J. Wilson, V. L. Scott, and J. Gibbs, "Geographic biases in bee research limits understanding of species distribution and response to anthropogenic disturbance," *Front. Ecol. Evol.*, vol. 7, no. JUN, pp. 1–8, 2019.

[19] A. Agresti, C. Franklin, and B. Klingenberg, *The art and science of learning from data*, Fourth. Ed., vol. 53, no. 95. New York: Pearson, 2012.

[20] P. C. Consul and F. Famoye, "Generalized poisson regression model," *Communications in Statistics - Theory and Methods*, vol. 21, no 1. pp. 89–109, 1992.

[21] A. Zeileis, C. Kleiber, and S. Jackman, "Regression models for count data in R," *J. Stat. Softw.*, vol. 27, no. 8, pp. 1–25, 2008.

[22] J. M. Hilbe, "Modeling count data," *model. Count Data*, no. 3, pp. 1–294, 2014.

[23] C. von Linnei, *Systema naturae*, Editio Dec. Impensis Direct. Laurentii Salvii, 1759.

[24] Flora Fauna Web, "Excoecaria cochinchinensis Lour..," Singapore National Parks, 2019. https://www.nparks.gov.sg/florafaunaweb/flora/2/0/2010 (accessed Jan. 28, 2021).

[25] T. A. James and G. J. Harden, "Genus Excoecaria," NEW SOUTH WALES FLORA ONLINE, 2021. https://plantnet.rbgsyd.nsw.gov.au/cgi-bin/NSWfl.pl?page=nswfl&lvl=gn&name=Excoecaria.

[26] E. K. Meineke and B. H. Daru, "Bias assessments to expand research harnessing biological collections," *Trends Ecol. Evol.*, pp. 1–12, 2021.

[27] A. Hazra, "An Exact Kolmogorov–Smirnov Test for the Negative Binomial Distribution with Unknown Probability of Success," *Res. Rev. J. Stat.*, vol. 2, no. 1, pp. 1–13, 2013.

[28] M. Jamshidian, R. I. Jennrich, and W. Liu, "A study of partial F tests for multiple linear regression models," *Comput. Stat. Data Anal.*, vol. 51, no. 12, pp. 6269–6284, 2007.

[29] C. M. Woods, L. Cai, and M. Wang, "The Langer-Improved Wald Test for DIF Testing With Multiple Groups: Evaluation and Comparison to Two-Group IRT," *Educ. Psychol. Meas.*, vol. 73, no. 3, pp. 532–547, 2013.