

# Causal Modeling of the Interactive Effects of Climate Change and Biodiversity Loss on Ecosystem Functioning

**Thabelang Manabalala Ncube**

*School of Computer Science  
and Applied Mathematics  
The University of the Witwatersrand  
Johannesburg, South Africa  
1918469@students.wits.ac.za*

**Ashwini Jadhav**

*School of Computer Science  
and Applied Mathematics  
The University of the Witwatersrand  
Johannesburg, South Africa  
ashwini.jadhav@wits.ac.za*

**Ritesh Ajoodha**

*School of Computer Science  
and Applied Mathematics  
The University of the Witwatersrand  
Johannesburg, South Africa  
ritesh.ajoodha@wits.ac.za*

**Abstract**—Climate change and biodiversity loss pose profound threats to global ecosystems, yet understanding their interactive effects on ecosystem function remains a complex challenge. This research addresses the urgent need to unravel the intricate causal relationships between climate change, biodiversity loss, and their combined impact on ecosystem function. Employing a sophisticated causal modeling approach, we integrate data from a diverse ecosystem to elucidate the nuanced interactions shaping ecosystem dynamics. Methodologically, this study leverages advanced machine learning techniques, including Bayesian models, to untangle the direct and indirect pathways through which climate change and biodiversity loss jointly influence key ecosystem functions. Our findings reveal not only the individual contributions of climate change and biodiversity loss but also the synergistic effects that emerge from their interplay, shedding light on previously unexplored facets of ecosystem dynamics. Key results highlight the causal relationship and influence of ecosystem components on each other, emphasizing the critical importance of considering interactive effects. Researchers seeking to understand and mitigate the impacts of global environmental changes will find this study instrumental in informing conservation strategies and sustainable management practices. In our experiments, we had two models with 60% and 80% respectively. We used a classification report with f1-score, precision, and recall to measure the models' performances.

**Index Terms**—Causal Modeling, climate change, biodiversity loss, ecosystems, Bayesian models, machine learning.

## I. INTRODUCTION

Climate change and biodiversity loss represent two of the most formidable challenges confronting our planet, underscoring the urgency for a comprehensive understanding of their intertwined impacts on ecosystem functioning. As these global issues continue to escalate, it becomes increasingly crucial to dissect the complex relationships between climate change, biodiversity loss, and the intricate dynamics of ecosystems. This research aims to address the consequential knowledge gap, with a particular focus on tank bromeliad ecosystems—a biodiverse and sensitive habitat—to unravel the interactive effects of rainfall changes and litter diversity on ecological processes.

In the realm of ecological sciences, the preservation of ecosystem functioning in the face of climate change is intricately tied to biodiversity. A meta-analysis conducted by [9] incorporating data from 192 global studies demonstrated the instrumental role of biodiversity in sustaining essential ecological processes. The study highlighted biodiversity's significance in preserving ecosystem productivity, nitrogen cycling, and energy flow. Additionally, it brought to light the potential exacerbation of climate change effects in ecosystems experiencing biodiversity decline. Further contributing to this understanding, a large-scale global study by [12] emphasized the compounding impact of biodiversity loss and climate change on ecosystems. This research underscored the pressing need for comprehensive conservation strategies that account for the interconnected influences of both climate change and biodiversity loss.

- How do variations in rainfall and litter diversity influence key ecological processes in tank bromeliad ecosystems?
- What are the causal effects of altered rainfall and litter diversity on decomposition rates and bacterial production in tank bromeliad ecosystems?

To delve into these questions, our study adopts the Causalnex framework for causal inference, providing a robust methodology to explore complex causal relationships. The experimental design involves the manipulation of rainfall and litter diversity in tank bromeliad ecosystems, creating distinct treatment and control groups. Key ecological processes, such as decomposition rates and bacterial production, are meticulously measured to discern the causal effects of rainfall changes and litter diversity.

Anticipating altered rainfall conditions to shift the mechanisms governing the litter diversity–decomposition relationship, we hypothesize a dominance effect, wherein the conditioning of one litter species by microbes becomes pivotal. This hypothesis, rooted in the existing literature on biodiversity and ecosystem functioning, guides our experimental design and analytical approach. To rigorously test these hypotheses,

a Bayesian model is applied to the collected data, with k-fold cross-validation ensuring the robustness of our inferences.

The structure of this document unfolds in a systematic manner, aligning with best practices in research reporting. The Related Work section provides an expansive overview of prior research, identifying relevant theories, methods, and research gaps. The Methodology section meticulously details the data processing, features, and causal model, offering transparency into our experimental setup. The Results section critically interprets findings, providing key insights into the causal relationships identified. Finally, the Conclusion section encapsulates the study, outlining its contributions and suggesting avenues for future research.

In sum, this research endeavors to bridge critical gaps in understanding the interactive effects of climate change and biodiversity loss on tank bromeliad ecosystems. By elucidating causal relationships and providing empirical insights, it contributes to the foundational knowledge essential for informed conservation policies and sustainable management practices in the face of these pressing environmental challenges. Through a multidisciplinary approach encompassing ecology and causal inference, this study aspires to pave the way for a more resilient and adaptive approach to ecosystem conservation in the Anthropocene era.

## II. RELATED WORK

This section examines previously published research on the subjects of machine learning models and causality. Initially, we will examine the data utilized by multiple studies that made substantial contributions to the prediction of the effects of climate and biodiversity loss on the ecosystem using probabilistic models and other statistical models. Next, we'll examine a list of notable features that have been shown to accurately predict the rate of microbial production in the ecosystem; lastly, we'll look at the machine learning models that were employed.

### A. Data and Features

The investigation of wetland mesocosms by Altor and Mitsch [3] sheds light on the dynamics of methane ( $\text{CH}_4$ ) and carbon dioxide ( $\text{CO}_2$ ) and their relationship with soils and hydrology. The study, focusing on the carbon sequestration role of wetlands, utilized mesocosms mimicking wetland habitats. Results showcased the significant impact of hydrology and soil conditions on greenhouse gas emissions. Notably, methane production increased with wetter conditions, while carbon dioxide production rose in response to drier conditions. The type of soil also influenced gas dynamics, with organic soils generating more methane and mineral soils producing more carbon dioxide. Statistical analyses, including ANOVA and post-hoc tests, emphasized the significance of these findings.

To quantify the flux of gases (methane or carbon dioxide) in wetland mesocosms, researchers employed the equation:

$$F = \frac{C_2 - C_1}{t} \quad (1)$$

where  $F$  is the flux of gas (methane or carbon dioxide) in  $\text{mg m}^{-2} \text{ h}^{-1}$ ,  $C_1$  is the initial concentration of gas in the

chamber in  $\text{mg L}^{-1}$ ,  $C_2$  is the final concentration of gas in the chamber in  $\text{mg L}^{-1}$ , and  $t$  is the time in hours between the measurements.

Pinheiro et al. [16] extended the understanding of ecological processes in tank bromeliads, exploring microbial community dynamics in a human-modified tropical forest. Utilizing high-throughput sequencing and statistical modeling techniques such as linear models and variance partitioning, the study revealed dispersal limitation and ecological drift as crucial factors shaping the microbiome of *Aechmea distichantha*. Dispersal limitation emerged as the primary driver, emphasizing its impact on microbial community structure. Linear models were instrumental in uncovering relationships between ecological variables and microbial communities, while variance partitioning highlighted the relative importance of different ecological factors.

For the statistical model used in the analysis, the researchers adopted the following equations:

$$y_i \sim N(\mu_i, \sigma_i^2)$$

$$\mu_i = \alpha_j + \beta_j x_i$$

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2)$$

$$\beta_j \sim N(\mu_\beta, \sigma_\beta^2)$$

$$\mu_\alpha \sim N(\mu_0, \sigma_0^2)$$

$$\mu_\beta \sim N(\mu_1, \sigma_1^2)$$

Here,  $y_i$  represents the response variable for the  $i$ th bromeliad sample. The mean of the response variable ( $\mu_i$ ) is modeled as a function of a fixed effect ( $\beta_j$ ) and a random effect ( $\alpha_j$ ) associated with the  $j$ th bromeliad population. The fixed effect represents the effect of the predictor variable ( $x_i$ ), while the random effect captures the variation in response variables among populations.

Hammill et al. [8] delved into predation threats' influence on bromeliad ecosystems, investigating interactions among mosquito larvae, top predators (damselfly larvae), and the detrital food web. Experimental bromeliad ecosystems were created, manipulating predator presence and measuring various ecosystem processes. Predation threats altered the detrital food web, affecting nutrient cycling and primary productivity. Analytical techniques, including GLMMs, PCA, and Bayesian hierarchical modeling, provided a nuanced understanding. The hierarchical modeling framework allowed for a comprehensive exploration of predator effects on bromeliad ecosystems, considering both fixed and random effects.

In the context of the statistical model used for analysis, the researchers implemented the following equation:

$$y_{ij} = \beta_0 + \beta_{ui} + \beta_{ij} + \beta_1 x_{1,ij} + \beta_2 x_{2,ij} + \beta_3 x_{3,ij} + \beta_4 x_{4,ij} + u_i + \varepsilon_{ij} \quad (2)$$

Where  $y_{ij}$  is the response variable,  $x_{1,ij}$  to  $x_{4,ij}$  are the predictor variables,  $\beta_0$  to  $\beta_4$  are the regression coefficients,  $u_i$  is the random effect of the bromeliad plant on the intercept, and  $\varepsilon_{ij}$  is the residual error.

Arjalla et al. [4] [7] explored the impact of terrestrial organic matter on aquatic food chains in tank bromeliads across different geographic regions. Using stable isotope analysis and Bayesian model averaging, the study assessed the contribution of allochthonous organic matter to aquatic food webs. The Bayesian model averaging approach facilitated the estimation of the importance of different organic matter sources, considering factors like light and litter inputs. The research highlighted the intricate connections between terrestrial and aquatic components within tank bromeliad ecosystems.

To estimate the importance of different sources of organic matter to aquatic invertebrates, the authors used the following Bayesian model averaging equation:

$$p(m|y) = \frac{p(y|m)p(m)}{p(y)} \quad (3)$$

where  $p(m|y)$  is the posterior probability of the model  $m$  given the data  $y$ ,  $p(y|m)$  is the likelihood of the data given the model,  $p(m)$  is the prior probability of the model, and  $p(y)$  is the marginal likelihood of the data. By using Bayesian model averaging, the authors were able to estimate the posterior probability of each model and each variable within each model, and thus identify the most important sources of organic matter for the aquatic invertebrates in each region.

Expanding on the related work, Powers et al. [18] emphasized the complexity and potential confounding factors in studying ecological processes, such as the simultaneous influence of rainfall changes, litter diversity, temperature, nutrient availability, and land use change. The challenges of context-dependence and site-specific effects were underscored by Isbell et al. [16], emphasizing the need for a nuanced understanding of how different ecosystems respond to variations in rainfall and litter diversity. Wardle et al. [20] highlighted the importance of mechanistic understanding, pointing out that while evidence links rainfall changes and litter diversity to ecological processes, the specific pathways are not always well understood.

In the context of tank bromeliad ecosystems, the integration of comprehensive and high-quality datasets becomes paramount for causal modeling in ecological research. Biodiversity data, collected through field surveys and monitoring programs, contribute to global databases, capturing species richness, evenness, and composition. Climate data, including temperature, precipitation, and CO<sub>2</sub> levels, sourced from meteorological stations, global climate models, and the IPCC, provides critical information for understanding climate change impacts. Ecosystem functioning data, encompassing primary productivity, nutrient cycling, and energy flow, relies on field experiments and remote sensing techniques, such as those employed by the LTER network.

Despite these valuable contributions, challenges persist in ecological research, including the potential for biased datasets. Selection bias, where certain rainfall scenarios or litter species are overrepresented, can skew predictions of microbial production rates. Addressing such biases requires techniques like Synthetic Minority Over-sampling Technique (SMOTE) to

balance data for unbiased machine learning modeling [26]. This commitment to addressing biases ensures that the models developed from these datasets are more robust and applicable across diverse ecological contexts.

The studies collectively highlight the intricate relationships within ecosystems, emphasizing the need for a multidisciplinary approach to ecological research. The integration of mathematical models and statistical analyses enhances our understanding of complex ecological processes, providing a quantitative foundation for drawing meaningful conclusions. Furthermore, these investigations underscore the importance of considering multiple factors simultaneously, acknowledging the complexity of ecosystems and the potential for confounding variables.

Relating these findings to the broader context of ecological research, it is evident that advancements in understanding ecosystem dynamics are critical for addressing global environmental challenges. The studies on wetland mesocosms and tank bromeliads contribute valuable insights into the role of wetlands as carbon sinks and the microbial dynamics in tropical forest ecosystems, respectively. By elucidating the impact of hydrology, soil conditions, predation threats, and organic matter on ecological processes, these studies provide a foundation for informed conservation and management strategies.

Considering the significance of the technical contributions made by these studies, the development of mathematical models and statistical frameworks emerges as a key takeaway. The application of linear models, variance partitioning, Bayesian model averaging, and hierarchical modeling not only advances the respective fields but also sets a precedent for future ecological research. These quantitative approaches enable researchers to disentangle complex relationships, contributing to a more nuanced understanding of ecosystem dynamics.

However, it is crucial to acknowledge the limitations inherent in these studies. The context-dependence and site-specific effects emphasized by Isbell et al. [10] and the potential for confounding factors highlighted by Powers et al. [19] underscore the challenges in generalizing findings across diverse ecosystems. As ecological research progresses, addressing these challenges will be pivotal for enhancing the robustness and applicability of ecological models and theories.

Looking ahead, the future trajectory of research in wetland ecosystems and tropical forest environments should consider these methodological advances. The identified limitations serve as a guide for refining experimental designs and analysis methodologies.

In conclusion, the studies on wetland mesocosms and tank bromeliads contribute significantly to our understanding of ecosystem dynamics. The application of mathematical models and statistical analyses enriches the depth of ecological research, providing valuable insights that inform conservation practices. While acknowledging the limitations and challenges, these studies pave the way for future research avenues, emphasizing the need for interdisciplinary collaboration, open science practices, and a holistic understanding of complex

ecological systems.

### B. Models

Machine learning and other statistical techniques are used to predict microbial activities. In the context of machine learning models, they make use of historical data to train and validate. In the exploration of wetland dynamics, [3] employed Analysis of Variance (ANOVA) and post-hoc tests to establish causal links between soil, hydrology, and greenhouse gas emissions, quantifying the significance of variations. Shifting to microbial intricacies in *Aechmea distichantha*, [16] utilized linear models and variance partitioning to unveil causal relationships between ecological variables and microbial community structure. Generalized Linear Mixed-Effects Models (GLMMs), Principal Component Analysis (PCA), and Bayesian hierarchical modeling guided [8] in deciphering the impacts of predation threats on bromeliad ecosystems. In assessing terrestrial organic matter, [7] applied stable isotope analysis, quantifying the causal contribution of allochthonous organic matter to aquatic food chains through  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values. These models served as analytical compasses, unraveling causal intricacies and providing both qualitative insights and quantitative measures within the complex web of ecosystem dynamics. Most of these analysis techniques and models used, focused more on observed data to draw conclusions, but in this project, we want to focus mainly on the causality of different factors that affect decomposition and microbial activity in the ecosystem to draw conclusions from our models using two distinct datasets.

## III. METHODOLOGY

In this study, we attempt to predict how the rates of decomposition and microbial production are jointly affected by Rainfall variations and Litter diversity. Through causality, we aim to determine how each of our independent variables could affect our dependent variable under certain conditions and states i.e., decomposition.

We categorized decomposition into three classes: (i) Low decomposition rate where the decomposition rate obtained from the experiment is lower than 0.5, (ii) Medium decomposition rate where the decomposition rate obtained is above 0.5 but less than 0.75 and finally we have (iii) High decomposition rate where the decomposition rate is above 0.75. The goal here is to answer the question; What are the causal effects of altered rainfall and litter diversity on decomposition rates and bacterial production? Since our research is mainly focused on causality, we used a Bayesian network library called Causalnex to explore all our networks and perform inference on our Bayesian network models. A classification metric, ROC curve, and other performance metrics like the F1, accuracy and precision score will be used to show the performances of our algorithms.

### A. Data Preprocessing and Collection

In this section, we will describe the steps taken in data collection and pre-processing. The dataset that we will focus

on is a subset of a larger dataset that was collected from this experimental design in Rio de Janeiro, Brazil. We used a full-factorial design, composed of five rainfall scenarios crossed with all combinations of three litter species (seven litter combinations, including all one-, two- and three-species combinations), with five replicates. The dataset used contained 175 rows, and 3 columns (2 predictors and 1 target variable). The target variable had 3 classes, Low decomposition, Medium decomposition, and High decomposition, each with 58 instances. Since we are using the Causalnex tool, we have to first deal with continuous and categorized data by discretizing all the values in the data to enable the modeling of causal relationships and probabilistic inference in real-world scenarios.

To establish causal relationships, there is a need to actively intervene and manipulate variables. Observational data alone may show correlations, but it cannot prove causation. We will adopt an interventional approach to investigate the causal relationships between rainfall scenarios and litter diversity on decomposition and microbial activity. Unlike the observational data collection mentioned earlier, in this case, we sort of throw away all the rows of information we had in our observational data, but we keep our Variable names i.e., Rainfall scenario, Litter diversity, and Decomposition. The interventional data generation involved active manipulation of specific variables. These interventions will be applied systematically just as it was done to the experimental setups for the observational data, with each scenario representing a different treatment. We followed the same essential step to transform the continuous and categorized variables into discrete states, enabling us to construct a Bayesian network that would facilitate the investigation of causal links and probabilistic inferences in our experiment.

### B. Data Features

Rainfall scenario	binned to:
1. Ambient	0
2. Medium Ambient	2
3. High Ambient	1
4. Medium Clustering	4
5. High Clustering	3

TABLE I  
RAINFALL VARIATIONS AND THEIR DECODED VALUES AFTER BINNING

Litter diversity	binned to:
1. E.uniflora	1
2. Cyperus sp.	1
3. C.hilariana	1
4. C. hilariana + Cyperus sp.	2
5. Cyperus sp. + E.uniflora	2
6. E.uniflora + C. hilariana	2
7. C. hilariana + E.uniflora + Cyperus sp.	3

TABLE II  
LITTER DIVERSITY AND ITS DECODED VALUES AFTER BINNING



Decomposition rates	ranges	binned to:
1. Low decomposition rate	$x \leq 0.5$	0
2. Medium decomposition rate	$0.51 \leq x \leq 0.74$	1
3. High decomposition	$x \geq 0.75$	2

TABLE III

DECOMPOSITION RATES AND THEIR DECODED VALUES AFTER BINNING

Before we dive deep into the models employed in performing probabilistic inference and figuring out causal relationships of our Bayesian network model features it is best to first look into the variables of our data and their states in both observational and interventional data, though these two types of data take up different forms of representations. We will just look at each variable and the states it takes after data cleaning:

The Litter diversity Table II shows all 7 litter combinations, each unique litter species is decoded into 1, any combination of two unique species is decoded into 2, and lastly, a combination of all three litter species is decoded into 3. In our testing, we assume that litter diversity combinations with the same bin have equal strength and contributions to the decomposition and microbial activities e.g. *Cyperus* sp. and *E.uniflora*, both decoded/ binned to 1.

In the decomposition rate Table III, we had to range our continuous variables using the Equal-width discretization method which divides the range of continuous values into equal-width intervals or bins. this enables interpretability and can help reduce the noise in the data. It smooths out variations and fluctuations, making causal relationships more apparent and stable.

However, in the data features section, we often talked about the data variables and the states they take, which mostly describes the form the observational data takes after data pre-processing. The interventional data will also take the same states but with a different format which we will discuss more about in one of our models below.

### C. Model Classification and Evaluation

To analyze and compare the machine learning models, we will use a classification report, the accuracy of the model. An example of a 2x2 confusion matrix that can be used to calculate some performance metrics is shown in Table IV. The meaning of the entries in Table IV are: TN is True Negative, which means the number of correct negative predictions. FP is False Positives, which means the number of incorrect positive predictions. FN is False Negatives, which means the number of incorrect negative predictions. TP is True Positives, which means the number of correct positive predictions.

		Actual	
		Negative	Positive
Predicted	Negative	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

TABLE IV

TWO-CLASS CONFUSION MATRIX

The accuracy of the models will be obtained from the confusion matrix (Table IV) defined as:

$$\text{Accuracy} = \frac{TN + TP}{TN + FP + FN + TP} \quad (4)$$

Two Bayesian models will be used to determine the causal relationship of our independent variables with decomposition rates based on the two types of datasets that we have chosen to focus on.

a) *Bayesian model on observational data:* Initially we have a predefined Bayesian network(Ground Truth(GT)) as shown in Fig 2, using our cleaned dataset we want our predefined model to learn a structure from the dataset and see if the learned structure takes the same shape as our predefined one. Firstly we split our dataset using the Causalnex train-test split library to split our dataset into a training set (80%) and a testing set (20%). We then train our model using k cross-folds to reduce overfitting in our model which occurs when a model learns the training data too well, capturing noise and anomalies that don't generalize to new data. We then fit Conditional Probability Distribution (CPD)tables based on the training set using the Bayesian estimator method. By computing CPD tables, we can specify the causal relationships between variables in the network. These tables represent how the probability of a target variable changes given different values of its parent variables, capturing causal dependencies. The fundamental concept in the conditional probability  $P(X|Y)$ , which represents the probability of variable  $X$  given that variable  $Y$  has a certain value. It's calculated as:

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)} \quad (5)$$

where  $P(X \cap Y)$  is the joint probability of  $X$  and  $Y$ , and  $P(Y)$  is the marginal probability of  $Y$ .

In many cases, one may not have prior knowledge of the exact probabilities in the domain. Bayesian estimator allowed us to learn CPD tables from data. The Bayesian estimator will use the Dirichlet distribution to estimate the conditional probability distribution (CPD) tables in the Bayesian network. The Dirichlet distribution, a probability distribution over probability distributions, makes it a suitable choice for estimating probabilities in a Bayesian context.

The probability density function of a Dirichlet distribution for a variable  $X$  with  $K$  categories ( $K$  possible values) is given by:

$$f(x|\alpha) = \frac{1}{B(\alpha)} \prod_{i=1}^K x_i^{\alpha_i-1} \quad (6)$$

Where:

- $f(x|\alpha)$  is the probability density function of the Dirichlet distribution for the vector  $x$  with parameters  $\alpha$ .
- $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_K)$  is a vector of positive shape parameters. Each  $\alpha_i$  represents the prior counts or beliefs about the occurrences of each category.
- $B(\alpha)$  is the multivariate Beta function, which is a normalizing constant ensuring that the probability density integrates to 1.

b) *Bayesian model 2: from generated interventional data:* In this model we employ almost a similar approach to the one in (a) but there are a few differences. First, we throw away our observational data and keep our model variables i.e., Rainfall scenario, Litter diversity, and Decomposition, and perhaps add one more variable e.g., Bacterial production in our predefined Bayesian model. The aim of this model is still the same, to perform probabilistic inference, but this time around we have to do some interventions. An intervention occurs when you take a dedicated action that manipulates the value. We can perform interventions of the form  $\text{do}(Z := z)$ , which forces the variable  $Z$  to take the value  $z$ . For example, intervention queries can be expressed as;

$$P(Y|\text{do}(z), X = x). \quad (7)$$

It is natural to use this interventional data from the interventional queries to learn causal models Jude Pearl et al.

$$P(\xi|\text{do}(Z := z), C) = \prod_{X_i \notin Z} P(x_i|u_i) \quad (8)$$

- $\xi$ : This represents an event or a set of events.
- $\text{do}(Z := z)$ : This is an operator in the do-calculus, introduced by Judea Pearl. It represents an intervention where we set the variable  $Z$  to the value  $z$ .
- $C$ : This represents a set of conditioning variables.
- $X_i$ : These are variables not in the set  $Z$ .
- $x_i$ : These are the values of the variables  $X_i$ .
- $u_i$ : These are the values of the variables  $U_i$ , which typically represent unobserved confounders.

$$M[x_i; u_i] = \sum_{m: X_i \notin Z[m]} 1\{X_i[m] = x_i, PaX_i[m] = u_i\} \quad (9)$$

- $M[x_i; u_i]$ : This is the main function that we're calculating. It depends on the variables  $x_i$  and  $u_i$ .
- $x_i$ : This is a specific value of the random variable  $X_i$ .
- $u_i$ : This is a specific value associated with  $X_i$ .
- $m$ : This is an index used for summation.
- $X_i$ : This is a random variable.
- $Z[m]$ : This is a set of variables that does not include  $X_i$ .
- $1\{\dots\}$ : This is an indicator function that equals 1 if the condition inside the braces is true, and 0 otherwise.
- $PaX_i[m]$ : This represents some function or property of  $X_i$ .

The intuition is to count the number of occurrences of this event, in data instances where there is no intervention at  $X_i$ .

$$L(C : D) = \prod_{i=1}^n \prod_{x_i \in Val(X_i), u_i \in Val(PaX_i)} \theta_{x_i|u_i}^{M[x_i; u_i]} \quad (10)$$

- $L(C : D)$ : This is the likelihood of the data  $D$  given the model  $C$ .
- $n$ : This is the total number of data points.
- $Val(X_i)$ : This is the set of all possible values of  $X_i$ .
- $Val(PaX_i)$ : This is the set of all possible values of the parent of  $X_i$ .

- $\theta_{x_i|u_i}$ : This is the parameter of the model for  $X_i$  given  $u_i$ .
- $M[x_i; u_i]$ : This is a function that depends on  $x_i$  and  $u_i$ . It's often used in the context of sufficient statistics in exponential family distributions.

In this modeling approach, we employ a predefined Ground Truth (GT) Bayesian model structure. For each variable and its corresponding states, we conduct interventions using the *DoCalculus* library—generating, for instance, 150 samples of data for each intervention. This process is repeated across all variables and states until we achieve the desired sample size for interventional data. It's crucial to note that this interventional data assumes a distinct tabulated form, differentiating it from observational data in the model (a), as elaborated in the subsequent section.

The data is then split, trained, and used to relearn the new model structure from the initially established GT model. Subsequently, we validate the model and compare accuracies between the Newly Learned (NL) model and the GT model. By comparing errors, we can assess how well the Newly Learned model aligns with the Ground Truth model. It serves as a form of validation, helping ensure that the learned model accurately captures the underlying structure and relationships within the data. Beyond error comparison, further inference can be performed. For instance, introducing mutation to the network allows us to observe changes in the collider node, providing insights into causal relationships and potential decomposition effects.

#### IV. RESULTS

Initially, the observed dataset had continuous values on the target node i.e., Decomposition, and for Causalmx to perform well, all data variables have to be discretized, but before we discretize our data we checked if it is normally distributed using the Q-Q plot. In the plot, the quantiles of the data are

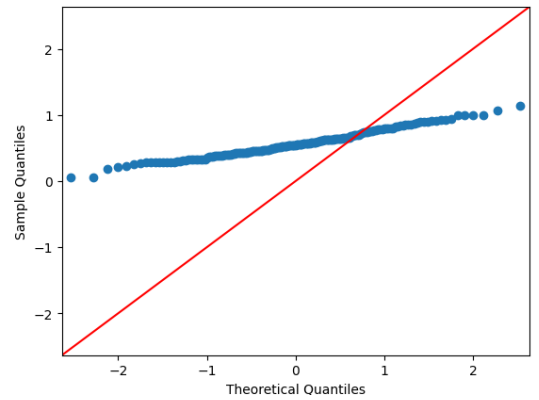


Fig. 1. Q-Q plot illustrating the linear relationship between theoretical and sample quantiles, suggesting a normal distribution of data in our attempt to find the distribution for our continuous data.

plotted against the quantiles of a standard normal distribution. If the data is normally distributed, the points in the Q-Q plot will fall along a 45° line, and from the plot above we can see that our continuous data was normally distributed. Though

we can see that 67% of the data looks normalized, to be so sure that most of our dataset is normalized we used the MiniMax Scaler function that ensured more than 90% of this continuous data is normal before we discretize it. We used the LabelEncoder inbuilt function to convert categorical values in the Rainfall scenario as shown in Table I.

We discuss the results from the two Bayesian models. A classification report is used, which shows the accuracy, precision, and F1 score using 10-fold cross-validation. We also computed the error between the Ground Truth and the Newly Learned for both models as a way to ensure that the model generalizes well on the underlying structure and relationships within the data.

*Model a:* fig 3 and represent the Ground Truth(GT) model Newly Learned(NL) model and the cofusion matrix respectively.

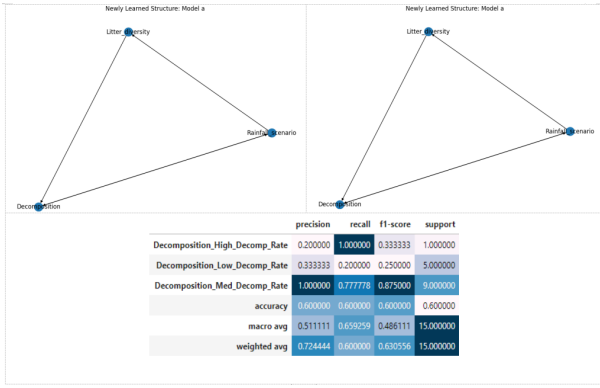


Fig. 2. shows Ground Truth and Newly Learned structure for model *a* and a classification report after the model validation which shows the strength of the model's performance through its accuracy, precision, and f1-score.

The model performs well in identifying instances of "High-Decomp\_Rate" with perfect recall (1.0). However, the precision for "High-Decomp\_Rate" is relatively low (0.2), indicating a higher rate of false positives. The overall accuracy of the model is 0.6, indicating that 60% of predictions are correct across all classes.

*Model b:* Below in Fig 5 is a piece of our tabulated form that the interventional data takes after an intervention on litter diversity in our GL model.

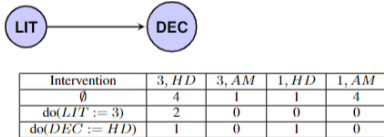


Fig. 3. Given that we intervene on Litter diversity by forcing a condition, this is represented by an interventional query table above.

Followed by the GL and NL model and Classification report for this model.

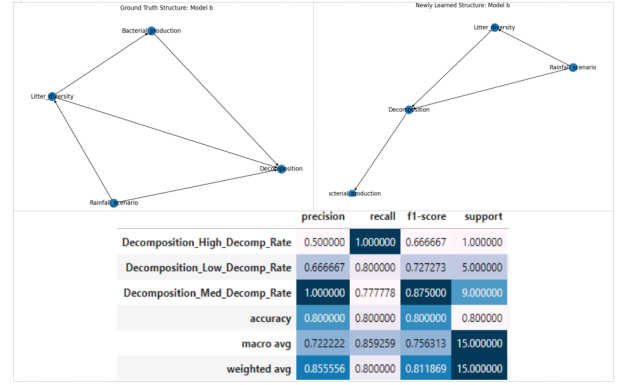


Fig. 4. shows Ground Truth and Newly Learned structure for model *b* and a classification report after the model validation which shows the strength of the model's performance through its accuracy, precision, and f1-score.

The Precision for each class (e.g., High-Decomp\_Rate, Low-Decomp\_Rate, Med-Decomp\_Rate) indicates the proportion of correctly identified instances among those predicted positive. For instance, the Precision for Low-Decomp\_Rate is 0.67.

The overall accuracy of the model is 80%. This represents the proportion of correctly classified instances across all classes. The model demonstrates reasonably good performance, with class-specific metrics offering insights into its strengths and weaknesses for different decomposition rates.

## V. DISCUSSIONS AND CONCLUSION

In addressing the overarching problem statement regarding the interactive effects of climate change and biodiversity loss on tank bromeliad ecosystems, this research employed a robust experimental setup and causal modeling techniques. The primary focus on rainfall changes and litter diversity aimed to unravel key ecological processes. The utilization of the Causalnex framework for causal inference provided a structured approach to explore complex relationships, contributing to the understanding of the intricate dynamics within tank bromeliad ecosystems.

The observed outcomes from the two Bayesian models, leveraging both observational and interventional data, shed light on the behavior and effectiveness of the models in discerning causal relationships. The model trained on observational data demonstrated commendable accuracy, particularly in identifying instances of "High decomposition rate ." However, challenges were encountered in precision ( of 0.2) for this class, indicating potential areas for model refinement and this shows that observational data alone cannot prove causation, though it may show correlations. In contrast, the model trained on interventional data exhibited a higher overall accuracy of 80%, showcasing its effectiveness in capturing the underlying structure of different decomposition rates. The technical contribution of adopting the Causalnex framework and Bayesian models serves as a methodological advancement. The research not only addresses immediate questions about the impact of rainfall changes and litter diversity on ecological processes but also provides a framework for future research

in ecosystem conservation. The models' behavior, validated through observational and interventional data, strengthens the argument for their validity and effectiveness in deciphering causal relationships.

One limitation is the experimental focus on tank bromeliad ecosystems, which, while providing depth, may limit generalizability to other ecosystems. Additionally, the assumption of equal strength and contributions to litter diversity combinations might oversimplify the intricate dynamics within the ecosystem. The comparison between the obtained results and the initial research questions reveals that the models successfully addressed the primary inquiries. Variations in rainfall and litter diversity were shown to influence key ecological processes, and the causal effects of these variations on decomposition rates and bacterial production were discerned through the rigorous experimental design and Bayesian models.

In conclusion, this research bridges critical gaps in understanding the interactive effects of climate change and biodiversity loss on tank bromeliad ecosystems. The multidisciplinary approach, combining ecology and causal inference, paves the way for more resilient and adaptive conservation strategies. By elucidating causal relationships and providing empirical insights, this study contributes foundational knowledge crucial for informed conservation policies and sustainable management practices in the Anthropocene era. Future research directions should explore the applicability of the Causalnex framework and Bayesian models in diverse ecosystems. Overall, this research lays the groundwork for future inquiries into the intricate interplay between climate change, biodiversity loss, and ecosystem functioning.

## VI. ACKNOWLEDGEMENT

I appreciate the advice I received when writing this research from my co-supervisor, Dr. Ashwini Jadhav, and my supervisor, Prof. Ritesh Ajoodha.

## REFERENCES

- [1] Yonatan Aguilar-Cruz, Felix Milke, Janina Leinberger, Anja [Alford et al. 1997] Douglas P Alford, Ronald D Delaune, and Charles W Lindau. Methane flux from Mississippi River deltaic plain wetlands. *Biogeochemistry*, 37:227–236, 1997.
- [2] Anne E Altor and William J Mitsch. Methane and carbon dioxide dynamics in wetland mesocosms: effects of hydrology and soils. *Ecological Applications*, 18(5):1307–1320, 2008.
- [3] Anne E Altor and William J Mitsch. Pulsing hydrology, methane emissions and carbon dioxide fluxes in created marshes: A 2-year ecosystem study. *Wetlands*, 28(2):423–438, 2008.
- [4] Vinicius F Arjalla, Angélica L Gonzalez, Régis Cereghino, Olivier Dezerald, Nicolas A Marino, Gustavo C O Piccoli, Barbara A Richardson, Michael J Richardson, Gustavo Q Romero, and Diane S Srivastava. Terrestrial support of aquatic food webs depends on light inputs: a geographically-replicated test using tank bromeliads. *Ecology*, 97(8):2147–2156, 2016.
- [5] Jude Bayham, Samuel D Chamberlain, Saptarshi Ghosh, Clark Glymour, Nicholas P Jewell, Konrad P Kording, Elizabeth L Ogburn, Judea Pearl, Richard Scheines, Bernhard Schölkopf, et al. Causal inference in ecological research. *Ecology Letters*, 24(3):377–394, 2021.
- [6] Vinicius F Farjalla, Claudio C Marinho, Bias M Faria, André M Amado, Francisco de A Esteves, Reinaldo L Bozelli, and Danilo Giroldo. Synergy of fresh and accumulated organic matter to bacterial growth. *Microbial ecology*, 57:657–666, 2009.

- [7] Vinicius F Farjalla, Andrew L Gonzalez, Régis Cereghino, Olivier Dezerald, Nicholas AC Marino, Gustavo CO Piccoli, Barbara A Richardson, Michael J Richardson, Gustavo Q Romero, and Diane S Srivastava. Terrestrial support of aquatic food webs depends on light inputs: a geographically-replicated test using tank bromeliads. *Ecology*, 97(8):2147–2156, 2016.
- [8] Edd Hammill, Trisha B Atwood, and Diane S Srivastava. Predation threat alters composition and functioning of bromeliad ecosystems. *Ecosystems*, 18(5):857–866, 2015.
- [9] David U Hooper, E Carol Adair, Bradley J Cardinale, Jarrett EK Byrnes, Bruce A Hungate, Kristin L Matulich, Andrew Gonzalez, J Emmett Duffy, Lars Gamfeldt, and Mary I O'Connor. Biodiversity mediates the effects of climate change on ecosystem functioning: A global meta-analysis. *Proceedings of the National Academy of Sciences*, 109(22):9329–9334, 2012.
- [10] Forest Isbell, Dylan Craven, John Connolly, Michel Loreau, Bernhard Schmid, Carl Beierkuhnlein, TM Bezemer, Catherine Bonin, Helge Bruelheide, Enrica de Luca, et al. Biodiversity increases the resistance of ecosystem productivity to climate extremes. *Nature*, 526(7574):574–577, 2015.
- [11] David W Lawlor. Benzing dh. 2000. bromeliaceae: Profile of an adaptive radiation (with contributions by b. bennett, g. brown, m. dimmitt, h. luther, i. ramirez, r. terry, w. till). 690 pp. cambridge: Cambridge university press. @ 75 (hardback). *Annals of Botany*, 5(86):1055–1056, 2000.
- [12] Stilianos Louca, Saulo MS Jacques, Aliny PF Pires, Juliana S Leal, Angélica L Gonzalez, Michael Doebeli, and Vinicius F Farjalla. Functional structure of the bromeliad tank microbiome is strongly shaped by local geochemical conditions. *Environmental microbiology*, 19(8):3132–3151, 2017.
- [13] Nalini M Nadkarni and Teri J Matelson. Fine litter dynamics within the tree canopy of a tropical cloud forest. *Ecology*, 72(6):2071–2082, 1991.
- [14] Shahid Naeem, J Emmett Duffy, and Erika Zavaleta. Ecological consequences of biodiversity loss: the evolution of a paradigm. *Ecology*, 97(2):330–336, 2016.
- [15] Tim Newbold, Lawrence N Hudson, Samantha L Hill, Sara Contu, Igor Lysenko, Rebecca A Senior, Luca B'orger, Dominic J Bennett, Argyrios Choimes, Ben Collen, et al. Global effects of land use on local terrestrial biodiversity. *Nature*, 520(7545):45–50, 2018.
- [16] Marcela M C Pinheiro, Susana Rodriguez-Echeverria, Milton C Ribeiro, Val'eria M de Oliveira, and Raquel S Peixoto. Dispersal limitation and ecological drift shape the microbiome of the aquatic tank bromeliad *aechmea distichantha* in a human-modified tropical forest. *Frontiers in microbiology*, 13:824531, 2022.
- [17] Aliny P. F. Pires, Diane S. Srivastava, Nicholas A. C. Marino, A. Andrew M. MacDonald, Marcos Paulo Figueiredo-Barros, and Vinicius F. Farjalla. Data from: Interactive effects of climate change and biodiversity loss on ecosystem functioning, 2019.
- [18] Jennifer S Powers, Rebecca A Montgomery, Mehdi Adjeroud, Rochelle R Beasley, April M Blakeslee, Brian Helmuth, Angela C Johnson, James B McClintock, Margaret A McManus, Cascade JB Sorte, et al. Global warming, elevational ranges and the vulnerability of tropical biota. *Biological Conservation*, 196:147–155, 2016.
- [19] Kelly Cristina Tonello, Alexandra Guidelli Rosa, José Augusto Salim, Carina Julia Pensa Correa, and Marcelle Teodoro Lima. The dynamics of knowledge about stem flow: a systematic review. *Brazilian Journal of Environmental Sciences (Online)*, 56(1):16–27, 2021.
- [20] David A Wardle, Richard D Bardgett, Ragan M Callaway, and Wim H Van der Putten. Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science*, 336(6080):49–54, 2012.
- [21] David A Wardle, Micael Jonsson, Shalabh Bansal, Richard D Bardgett, Michael J Gundale, Daniel B Metcalfe, Marie-Charlotte Nilsson, and Håkan Wallander. Ecosystem properties and forest decline in contrasting long-term chronosequences. *Science*, 335(6064):1083–1086, 2012.
- [22] Dornelas, M., et al. (2014). Assemblage Time Series Reveal Biodiversity Change but Not Systematic Loss. *Science*, 344(6181), 296–299.
- [23] Stocker, T. F., et al. (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- [24] National Science Foundation. (2021). Long-Term Ecological Research (LTER) Network. Retrieved from <https://lternet.edu/>



- [25] Johnson, T. P., & Fendrich, M. (2007). Modeling Sources of Self-Report Bias in a Survey of Drug Use Epidemiology. *Annals of Epidemiology*, 17(2), 147-153.
- [26] Zeineddine, R. M., et al. (2018). "Predicting Student Academic Performance: A Comparative Study of Classification Techniques." *International Journal of Engineering Research & Technology*, 7(10), 47-52.