MODELING FIRE HOTSPOTS IN KALIMANTAN, INDONESIA USING NESTED 3-COPULA REGRESSION BASED ON PRECIPITATION AND DRY DAYS DURING DIFFERENT ENSO PHASES

***Teduh Wulandari MAS'OED*[1], *Sri NURDIATI*[1], *Ardhasena SOPAHELUWAKAN*[2], *Mohamad Khoirun NAJIB*[1], *Ayudya SALSABILA*[1]**

DOI: 10.21163/GT_2024.192.21

ABSTRACT

The persistent forest and land fires in Indonesia have piqued the interest of numerous social groups. Due to the huge number of losses caused by fires, fire prediction is an important part of fire prevention. This study focused on Kalimantan, which is one of the major contributors to fires in Indonesia. This article focuses on modeling the relationship between total precipitation, the number of dry days, and hotspots in Kalimantan, Indonesia for each ENSO phase using a nested 3-copula approach. Using the selected copula structure, the number of hotspots was estimated using nested 3-copula regression with two predictors. Copula regression offers more robustness to outliers and non-normality in the data compared to traditional regression techniques. The results reveal that the regression model based on ENSO phases has an RMSE of 1204 hotspots per month and can explain up to 70% of the variance in hotspots. These results outperform models without ENSO phases, highlighting the importance of ENSO phases in simulating hotspots in Kalimantan. From the regression plane, the ENSO phase has a small impact on the hotspots at low levels. When it comes to high or intense hotspots, the ENSO phase is very important. El Nino is the most dangerous phase for extreme hotspots, while La Nina is the least dangerous. The findings of this study can help researchers better understand the influence and dependence of local and global climate conditions on hotspots in Kalimantan, which can be evolved into an early warning model for forest fires in Indonesia in the future.

Key-words: *Copula regression, Dry spells, Hierarchical copula, High-dimension, Multivariate copula, Uncertainty, Wildfire.*

1. INTRODUCTION

Indonesia is known as a group of islands separated by vast oceans. Its location around the equator flanked by two continents (Asia and Australia) and two oceans (Pacific and Indian) makes the Indonesian region vulnerable to climate variability. The climate in Indonesia is strongly influenced by natural phenomena in the surrounding sea, such as the El Nino Southern Oscillation (ENSO) (Firmansyah et al., 2022; Kurniadi et al., 2021). Most of the natural disasters that have occurred in Indonesia are closely related to the ENSO phase, such as droughts (Lestari et al., 2018), floods (Rodysill et al., 2019), and forest fires (Zahra et al., 2023).

ENSO is a sea surface condition in the Pacific Ocean region that experiences an increase or decrease in sea surface temperature resulting in a shift in the seasons in Indonesia. There are three phases of ENSO: neutral, El Nino and La Nina phases. El Nino refers to warming the ocean surface, or above-average sea surface temperatures (SSTs), in the central and eastern tropical Pacific Ocean. It represents the warm phase of the ENSO cycle. On the contrary, La Nina refers to the cooling of the central and eastern tropical Pacific Ocean surface and is the cold phase of the ENSO cycle (Wahiduzzaman et al., 2022). In Indonesia, El Nino phase correlates to less rainfall and drought, while La Nina phase correlates to rainfall increase and floods (Nugroho, 2022). Several unforgettable forest fire events in Indonesia occurred during a very strong El Nino, such as in 1997 and 2015 (Fanin &

[1]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, Indonesia. Teduhma@apps.ipb.ac.id (TWM), corresponding author*: nurdiati@apps.ipb.ac.id (SN), mohknajib@gmail.com (MKN), 13072001ayudya@apps.ipb.ac.id (ASo)*

[2]*Center for Applied Climate Information Services, Agency for Meteorology, Climatology, and Geophysics, Jakarta, Indonesia, ardhasena@bmkg.go.id (ASa)*

Van Der Werf, 2017). These forest fires impact various sectors from rampant deforestation (Adrianto et al., 2020) and alarming declines in air quality (Rahman et al., 2024) to the tragic loss of flora and fauna (Harrison et al., 2016; Wasis et al., 2018) and the health problems faced by affected residents (Hein et al., 2022; Uda et al., 2019). This phenomenon underscores the far-reaching consequences of forest fires, impacting every facet of ecosystems and human life.

Forest fires are a recurring environmental issue that has significant local, regional, and global implications. Forest and peatland fires during the 2015 El Nino drought were among the worst in Southeast Asia, contributing to carbon emissions across the region, with the haze causing an air pollution crisis affecting millions of people (Lee et al., 2017). Peatland fires in Indonesia also result in long-term health impacts causing premature mortality due to chronic respiratory, cardiovascular and lung cancer (Uda et al., 2019) . As a result, forest and peatland fires have become an issue that has attracted the attention of the government, communities and researchers.

Several researchers have explored the intricate relationship between global and local climate factors and their correlation with hotspots as indicators of forest and land fires. Yananto & Dewi (2016) highlighted how El Niño events significantly increased hotspots in Sumatra and Kalimantan, particularly between July and October 2015, when reduced rainfall coincided with a surge in fires. Aflahah et al. (2018) further examined forest fire indicators in Kalimantan, using multiple linear regression analysis to show the strong interconnection among visibility data, the number of hotspots, and temperature, underscoring their substantial influence on fire incidents. Nurdiati et al., (2022b) investigated the effects of ENSO and IOD conditions on the distribution of dry days and total precipitation in southern Sumatra, concluding that these factors significantly impacted the dry season but not the rainy season. Najib et al. (2022b) proposed a fire risk model for Kalimantan using copula regression, demonstrating that dry spells served as a better climatic predictor for fire risks than total precipitation. Collectively, these studies illuminate the complex dynamics of climate and fire activity, offering vital insights for effective forest fire management.

Copula regression is a statistical modeling technique that combines two important concepts: copulas and regression analysis. Copulas are mathematical functions used to describe the dependence structure between random variables, while regression analysis is a method for modeling the relationship between a dependent variable and one or more independent variables. Copula regression brings these two concepts together to provide a flexible and powerful approach for modeling complex dependencies and relationships in data (Czado et al., 2022; Kolev & Paiva, 2009).

One of the advantages of copula regression is its flexibility in modeling dependencies. Copula regression is often more robust to outliers and non-normality in the data compared to traditional regression techniques. Unlike traditional regression models that assume normality and linearity, copula regression can capture a wide range of dependency patterns, such as tail dependencies, asymmetry, and non-monotonic relationships. Copula regression has applications in various fields including environmental science (Najib et al., 2023), finance (Pan et al., 2023), medicine (Gayawan et al., 2023), and engineering (Ma et al., 2022).

Previous studies have shown that apart from decreasing rainfall due to the El Nino phase, a high number of days without rain (or dry days) is an important factor triggering forest and peatland fires, especially in Kalimantan, Indonesia. In Najib et al., (2022a), the two climatic factors are used separately to model hotspots using bivariate copulas. Both of these factors have their respective advantages in predicting hotspots in Kalimantan, although in general the number of dry days outperforms as a predictor of hotspots. Therefore, it is a challenge to model hotspots using both climatic factors using the copula trivariate. Multivariate copulas can be constructed based on bivariate copulas. Two commonly used approaches are nested copula (Serinaldi & Grimaldi, 2007) and vine copula (Heredia-Zavoni & Montes-Iturrizaga, 2019).

Nested and vine copulas are advanced techniques used to model complex multivariate dependencies, each of which has advantages and disadvantages. Here, we are interested in studying nested copula which has advantages, including a simple structure and less computational complexity (Saad et al., 2015), especially when dealing with a relatively small number of variables. Meanwhile, vine copulas can be advantageous when dealing with high-dimensional data, as they provide a

structured approach to managing the computational complexity of estimating multivariate dependencies.

Based on that description, this article focuses on modeling the relationship between total precipitation, number of dry days, and hotspots in Kalimantan, Indonesia for each ENSO phase using a nested 3-copula approach. There are several bivariate copula functions used in this article, both 1-parameter and 2-parameter copula. Using the selected copula structure, the number of hotspots was estimated using copula regression with two predictors using the Riemann sum discretization approach (Jha & Danjuma, 2020).

The article is organized as follows. Datasets are detailed in Section 2. Section 3 presents the methods, including copula functions, parameter estimation, and nested 3-copula regression. The results and discussion are reported in Section 4. Finally, Section 5 concludes this article.

2. STUDY AREA AND DATASETS

This research focuses on Kalimantan, the Indonesian section of the island of Borneo, the world's third-largest island. The island of Borneo is shared by three countries: Indonesia, Malaysia, and Brunei. Kalimantan is the Indonesian portion of Borneo, accounting for roughly three-quarters of the island's total area. West, Central, South, East, and North Kalimantan are the five provinces of Kalimantan (**Fig. 1a**). Each province has its unique culture, topography, and governance system.

Forest fires have been a reoccurring environmental and ecological hazard in Kalimantan, as well as other parts of Southeast Asia. These fires are frequently caused by a combination of factors and have far-reaching consequences for the region, including human activities, deforestation, infrastructure development, land management policies and climate changes. This study focuses on the climate factors that play a significant role in causing forest fires in Kalimantan, as they influence the environmental conditions that can lead to fire ignition and propagation, such as total precipitation and the number of dry days.

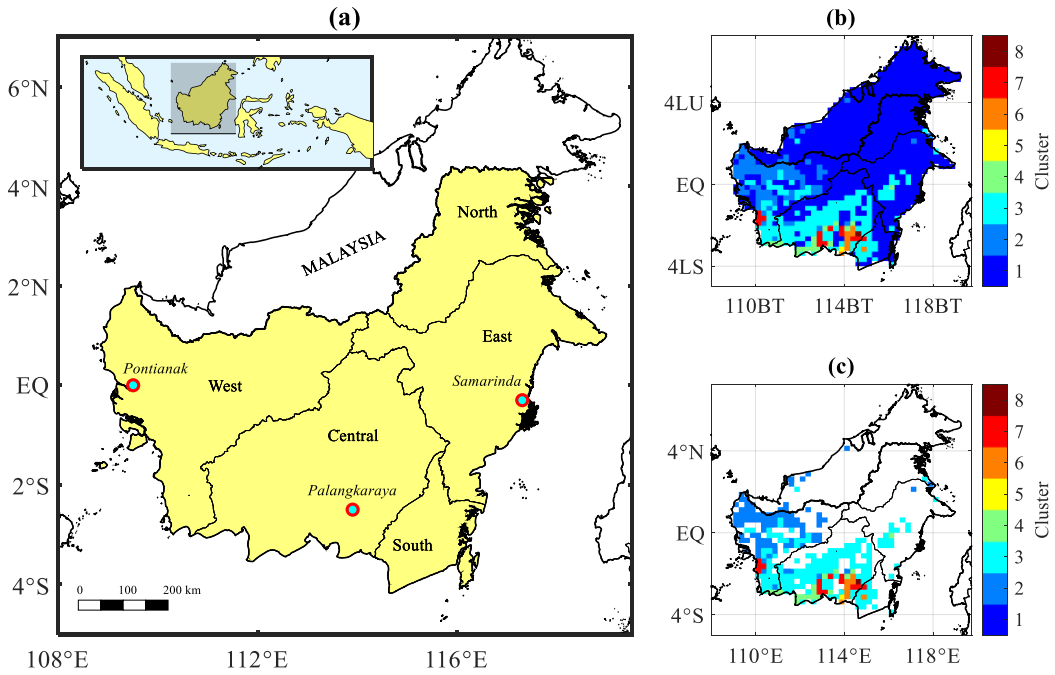

Fig. 1. (a) Map of Kalimantan, (b) clusters of hotspots in Kalimantan, and (c) selected areas for research.

This study makes use of numerous datasets. The information includes hotspots, total precipitation, the number of dry days (also known as dry spells), and the ENSO index (Nino 3.4). Hotspot data covers the Asia-Pacific region with a spatial resolution of $0.25^\circ \times 0.25^\circ$. The data was obtained from the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG) in the 2001-2020 period. Precipitation and number of dry days were taken and derived from the CMORPH-CRT product. The acronym CMORPH stands for "CPC (Climate Prediction Center) MORPHing technique", while CRT stands for "Calibrated Rainfall Technique" (Xie et al., 2019). By incorporating satellite data and calibrating the estimates with ground-based measurements (Bruster-Flores et al., 2019), CMORPH-CRT provides a more comprehensive and accurate picture of precipitation patterns on a global scale. CMORPH-CRT is adjusted through matching the probability density functions (PDF) of daily CMORPH-RAW against that for the CPC unified daily gauge analysis at each month over land (Xie et al., 2017). The spatial resolution for the precipitation data used is $0.25^\circ \times 0.25^\circ$, while the temporal resolution used is the monthly resolution of precipitation and number of dry days.

According to Septiawan et al. (2019), forest fire patterns are generally divided into two characteristics, namely Sumatra and Kalimantan fires. In terms of temporal characteristics, Septiawan et al. (2019) also stated that fires in Sumatra have two characteristics (6-month and 12-month periods), while fires in Kalimantan generally have an annual period. Compared to Sumatra which is more affected by the Indian Ocean Dipole phenomenon, forest fires in Kalimantan are more affected by the El Nino-Southern Oscillation phenomenon (Nurdiati et al., 2022a). Thus, the general characteristics of fires in Kalimantan can be said to be the same. However, not all areas but only a few points in Kalimantan are affected by forest fires. Therefore, a classification needs to be carried out so that only areas that have a significant influence on forest fires in Kalimantan are considered.

The data used is the result of extraction from previous research by Najib et al. (2021) which classifies forest fires in Kalimantan into several clusters (**Fig 1b**). In general, cluster 1 is an area with very low fire incidence, has relatively high land topography (mountains) and has quite high rainfall. Thus, the area in cluster 1 can be ignored because it can interfere with the general characteristics of forest fires in Kalimantan. Only cluster 1 was removed from the study area due to its classification as a low fire-prone region, characterized by a maximum hotspot occurrence of only about 3.43 hotspots per grid point, i.e., in 2002. The selection criteria focused on hotspot frequency and geographic coverage, allowing for a more targeted analysis of higher-risk areas, while higher-level clusters exhibit more hotspots and reflect conditions conducive to increased fire risks. The datasets are then aggregated into fire-prone areas in Kalimantan, i.e., grid points with significant concentrations of hotspots (**Fig. 1c**) to generate general characteristics of the data. Moreover, data were taken in months with high hotspots: July to November (**Fig. 2**).

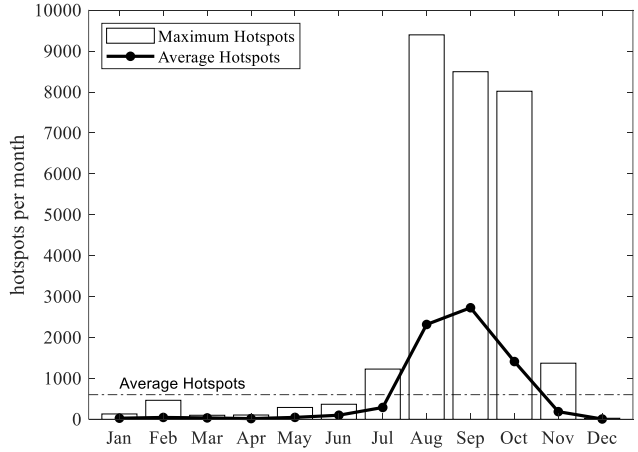

Fig. 2. Average and maximum monthly hotspots in 2001-2020 in fire-prone areas in Kalimantan.

According to Najib et al., (2021), the total precipitation that gives the strongest relationship is the two-month average, meaning that if the hotspots are in September, then the intended total precipitation is the average of the precipitation in August and September. If hotspots are detected in October, the total precipitation used for analysis would typically be calculated as a two-month average, encompassing the precipitation from September and October, and so on. Meanwhile, the number of dry days that is most correlated with hotspots is the three-monthly number of dry days, which is averaged over all fire-prone areas in Kalimantan. These data are used as predictors for hotspots in this study which will later be referred to as total precipitation and number of dry days.

3. METHODS

3.1. Copula Function

A copula function is a mathematical concept used in probability theory and statistics to describe the dependency structure between multiple random variables. Copula theory's basic concept is to separate the modeling of marginal distributions from the modeling of joint distributions, allowing for more flexible and complete representations of dependence patterns. Therefore, copulas provide a flexible way to model various types of dependencies between random variables, including linear, nonlinear, positive, negative, and tail dependencies. This flexibility allows them to capture complex relationships that may not be easily represented by traditional multivariate distributions. Traditional multivariate statistical techniques often assume normal distributions for variables, but real-world data frequently deviates from this assumption. Mathematically, a copula is defined as follows.

Definition 1. An n-dimensional copula (or n-copula) is a function C from $\mathbf{I}^n \rightarrow \mathbf{I}$ with the following properties:

1. For every $\mathbf{u} \in \mathbf{I}^n$, $C(\mathbf{u}) = 0$ if at least one coordinate of $\mathbf{u}$ is 0, and if all coordinates of $\mathbf{u}$ are 1 except u_k, then $C(\mathbf{u}) = u_k$
2. For every $\mathbf{a}, \mathbf{b} \in \mathbf{I}^n$ such that $\mathbf{a} \leq \mathbf{b}$, $V_C([\mathbf{a}, \mathbf{b}]) \geq 0$, where V_C is C-measure of a set.

Sklar's theorem is the core of the copula theory (Nelsen, 2006). Based on Sklar's theorem, copula is referred to as a function that links the multivariate joint cumulative distribution function to the corresponding univariate marginal cumulative distribution functions (Li et al., 2019).

Theorem 1. *Let F be a joint distribution function for a set of n continuous random variables $X_1, X_2, \dots, X_n$ each with marginal distribution function $F_1, F_2, \dots, F_n$. According to Sklar's theorem* (1959), *there exists a copula function C such that for any $x_1, x_2, \dots, x_n \in \mathbb{R}$:*

$$F(x_1, x_2, \dots, x_n) = C(u_1, u_2, \dots, u_n) \tag{1}$$

where $u_i = F_i(x_i)$ for $i = 1,2, \dots, n$ and is called n-copula.
Proof: See (Nelsen, 2006) □

In other words, the joint distribution function F can be expressed in term of the marginal distribution functions $F_1, F_2, \dots, F_n$ and a copula function C, which characterizes the dependence structure between the variables. If marginal distributions are continuous, the copula function is unique. However, this assumption can easily extend to a mixture of continuous and discrete variables (Schölzel & Friederichs, 2008). A copula function C is uniquely determined on $Ran\ F_1 \times Ran\ F_2 \times \cdots \times Ran\ F_n$, if not all marginal distributions are continuous.

3.1.1. Nested Copula

In the context of multivariate dependence modeling, an exchangeable copula is a copula function that reflects the exchangeability property. This is a simple and elegant approach to constructing high dimensional copulas (Zhang & Singh, 2019b). It is assumed that the variable dependence structure remains constant regardless of the order in which the variables are analyzed. Exchangeable copulas

are frequently symmetric, which means they do not distinguish between variables based on their involvement in the dependent structure. This symmetry is an unavoidable result of the exchangeability property, so exchangeable copulas are also often called symmetric copulas. Therefore, they may not capture more complex or heterogeneous dependence structures that exist in real-world data, although exchangeable copulas offer simplification in modeling high-dimensional dependencies.

Following (Joe, 1997), (Whelan, 2004), and (Serinaldi & Grimaldi, 2007), n-copula can be written in a form called "fully nested" or "asymmetric" which is obtained from a generalization of 2-copula, since it takes into account the non-exchangeability of the variables. The nested copula approach outperforms the exchangeable copula approach significantly (Aas & Berg, 2009). Asymmetries, allowing for more realistic dependencies, are obtained by plugging in Archimedean copulas into each other (Segers & Uyttendaele, 2014). There are $n-1$ bivariate copula functions for n-dimensional random variables modeled with fully nested copula, resulting in dependence structure with partial exchangeability. The fully nested copula structure is constructed with the following procedures (based on the degree of dependence between the pair variables):

1. As the first two variables (1 and 2), select the variables with the highest degree of dependence (rank-based).
2. Using variables 1 and 2, estimate the copula.
3. Evaluate the degree of dependence (rank-based) between empirical copula from step 2 with the remaining variables.
4. Select variable 3, which has the maximum degree of dependence (rank-based) with the copula constructed using variables 1 and 2.
5. Continue the process until the last variable is considered.

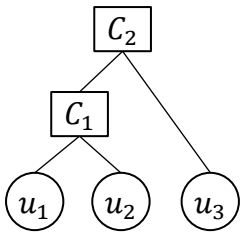

Fig. 3. Three-dimensional nested copula structure.

Figure 3 presents an example of a three-dimensional nested copula structure. Bivariate copulas are the foundation of nested copulas. Since $n = 3$, the nested 3-copula equation is given by

$$C(u_1, u_2, u_3) = C_2(C_1(u_1, u_2), u_3) \tag{2}$$

Figure 3 shows that two bivariate copulas are required to describe the dependence for three-dimensional random variables using nested copulas, as shown below. First, u_1 and u_2 are coupled by copula C_1, then the resulting variable is linked with u_3 by copula C_2. In general, the first two variables are coupled by a 2-copula, then the resulting copula is coupled with another variable by a second copula, and so on. The nested copula model makes it possible to construct joint distributions with different degrees of positive dependence within different bivariate margins (McNeil, 2008).

In the context of this research, u_1 represents the transformed cumulative distribution function (CDF) values of total precipitation, while u_2 corresponds to the transformed CDF values of the number of dry days. Additionally, u_3 signifies the transformed CDF values of hotspots. The function C_1 is the copula function that describes the dependence between the climate factors (total precipitation and number of dry days), while C_2 is the copula function that links these climate factors to the hotspots. This nested copula structure allows for a comprehensive analysis of the relationships and dependencies among the various variables, facilitating a deeper understanding of how climate factors influence hotspot occurrences.

3.2. Parameter Estimation and Hypothesis Tests

There are various types of copula functions, each with its own characteristics, properties and suitability for different dependencies. Different copulas differ in describing the dependence structures (Li et al., 2020). We use many types of copula functions, both 1-parameter and 2-parameter, including Gaussian, t-student, Clayton, Gumbel, Frank, Joe, Galambos, BB1, BB6, BB7, and BB8 copulas. Meanwhile, there are also many types of marginal univariate distribution that are used, including normal, lognormal, inverse gaussian, extreme value, generalized extreme value, logistic, loglogistic, exponential, gamma, and Weibull distributions for climate factors and negative binomial distribution for hotspots.

We use a 2-step method to estimate the copula parameters called the inference of function for margins (IFM), which estimates the parameters of the marginal distribution first before estimating the copula parameters (Joe, 1997). In summary, the process for constructing a nested 3-copula is as follows.

1. Estimate F_1, F_2, and F_3.
2. Estimate C_1 using u_1 and u_2.
3. Estimate C_2 using u_3 and $C_2(u_2, u_1)$.

Several statistics were used such as the Anderson Darling hypothesis test to select the most fit marginal distribution, while the copula function was selected using the Akaike Information Criterion (AIC) and tested based on the Cramer-von Mises hypothesis test. For more details, the parameter estimation process can be seen in Najib et al., (2022b).

The Anderson-Darling test departs with the null hypothesis that the data comes from a population with the selected distribution, where the test statistic is given by

$$A^2 = -\left(\sum_{t=1}^{N} \frac{2t-1}{N}\left[\ln F(x_t) + \ln\left(1 - F(x_{n+1-t})\right)\right]\right) - N \tag{3}$$

over the ordered sample values $x_1 \leq x_2 \leq \cdots \leq x_N$ (Anderson & Darling, 1954). Meanwhile, the Cramer-von Mises hypothesis test was performed to test the selected theoretical copula, with the null hypothesis that the data comes from the population with the selected theoretical copula, where the test statistic is given by

$$\hat{T} = \sum_{t=1}^{N}\left[C\left(F_1(x_1^t), F_2(x_2^t)\right) - \tilde{C}(x_1^t, x_2^t)\right]^2 \tag{4}$$

where $\tilde{C}$ is the empirical frequency (copula) estimated using the Gringorten formula (1963):

$$\tilde{C}(x_1^t, x_2^t) = \frac{\#(X_1 \leq x_1^t, X_2 \leq x_2^t) - 0.44}{N + 0.12} \tag{5}$$

where N is the sample size of data. The p-value is estimated, then with a significance level of 5%, if the p-value is greater than 5%, then the test fails to reject the null hypothesis.

3.3. Nested 3-Copula Regression Model

From Equations 1 and 2, it can be written that

$$F(x_1, x_2, x_3) = C_2(C_1(u_1, u_2), u_3) \tag{6}$$

Since $u_i = F_i(x_i)$ for $i = 1,2,3$, a joint probability density function f can be obtained by deriving both sides with respect to x_1, x_2, x_3, so that

$$f(x_1, x_2, x_3) = c_2(C_1(u_1, u_2), u_3) \cdot c_1(u_1, u_2) \cdot f_1(x_1) \cdot f_2(x_1) \cdot f_3(x_3) \tag{7}$$

where c_1 and c_2 are density functions of C_1 and C_2, respectively. A copula density c of a copula function C is given by its derivative with respect to each of its marginals:

$$c(u_j, u_k) = \frac{\partial^2}{\partial u_j \partial u_k} C(u_j, u_k) \tag{8}$$

where $u_j = F_j(x_j)$ and $u_k = F_k(x_k)$ for $j \neq k$. If F_j and F_k are continuous CDFs, then copula will uniquely determine the joint probability distribution of X_j and X_k. If F_j and F_k are a mixture of discrete and continuous CDFs, then copula will only uniquely determine the joint probability distribution of X_j and X_k over range $F_j \times$ range F_k (Pleis, 2018).

Let X_3 be the response variable while X_1 and X_2 are the explanatory variables, then the conditional probability density function of x_3 given x_1 and x_3 is defined by

$$f(x_3|x_1, x_2) = \frac{f(x_1, x_2, x_3)}{f(x_1, x_2)} = c_2(C_1(u_1, u_2), u_3) \cdot f_3(x_3) \tag{9}$$

due to $f(x_1, x_2) = c_1(u_1, u_2) \cdot f_1(x_1) \cdot f_2(x_1)$.

If we wish to predict the value of x_3, then we might take the expected value of the conditional density (Eq. 6), which is so-called conditional expectation. The conditional expectation value gives the minimum mean square error in the prediction for x_3, so it is also called the minimum-mean-square-error predictor. Using Equation 6, the conditional expectation value of x_3 given x_1 and x_2 is defined by

$$E(x_3|x_1, x_2) = \int_{-\infty}^{\infty} x_3 \cdot f(x_3|x_1, x_2)\, dx_3 = \int_{-\infty}^{\infty} c_2(C_1(u_1, u_2), u_3) \cdot f_3(x_3) \cdot x_3\, dx_3 \tag{10}$$

Since nested copulas are used to construct the conditional density, we call this formula a nested copula regression. Copula regression is often more robust to outliers and non-normality in the data compared to traditional regression techniques. It can handle data with heavy tails and non-standard distributions more effectively.

For computational convenience, we use the Riemann sum approach to estimate the value of the integral in Equation 7:

$$E(x_3|x_1, x_2) \approx \sum_{i=1}^{p} c_2(C_1(u_1, u_2), u_3^{(i)} \cdot f_3(x_3^{(i)}) \cdot x_3^{(i)} \cdot \Delta x_3^{(i)} \tag{11}$$

where p represents the number of partitions used (Jha & Danjuma, 2020).

3.4. Performance Metrics

We use several metrics to measure the performance of the resulting regression models, including root mean squared error (RMSE) and explained variance score (EVS). The RMSE is defined as:

$$\text{RMSE}(y, \hat{y}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{12}$$

where y is the actual data and $\hat{y}$ is the predicted data. Meanwhile, the explained variance score is estimated by:

$$\text{EVS}(y, \hat{y}) = 1 - \frac{var(y_i - \hat{y}_i)}{var(y_i)} \tag{13}$$

Scores close to 1.0 are highly desired, indicating better squares of standard deviations of errors (Oyedele et al., 2023).

4. RESULTS AND DISCUSSION

Here, we use scenario 1 without observing the ENSO effect and scenario 2 by observing the ENSO effect. This study analyzed the ENSO data monthly, allowing for a detailed examination of its influence on precipitation and dry days throughout the year. The ENSO phases were then classified into neutral, El Niño (ENSO > 0.5), and La Niña (ENSO < −0.5) categories. Scenario 1 uses all data without splitting (100 rows), while scenario 2 split data based on the ENSO phase: neutral (51 rows), El Nino (23 rows), and La Nina (26 rows). This section presents the selected marginal distribution, nested copula construction, and nested copula regression results.

Table 1.

Kendall-tau correlation between variables.

Datasets	$X_1 - X_2$	$X_1 - X_3$	$X_2 - X_3$
No Split	**0.7119**	0.5972	0.6186
El Nino	0.6614	**0.6878**	0.6138
Neutral	**0.7396**	0.5069	0.5367
La Nina	**0.6857**	0.5871	0.5298

Suppose X_1, X_2, and X_3 are negative of total precipitation, number of dry days, and hotspots, respectively. We choose to use the negative of total precipitation as X_1 so that the two predictors (X_1 and X_2) have a positive dependence on hotspots. **Table 1** shows the Kendall-tau correlation between variables. Except in El Nino conditions, the pair of variables X_1 and X_2 produces the strongest Kendall-tau correlation. For equality, the copula structure chooses X_1 and X_2 to be coupled as the first pair of variables. Aside from that, the presence of correlation is a prerequisite for copula-based modeling. Because the correlation value is quite strong, copula modeling can be performed.

4.1. Marginal Distribution of Variables

Estimating the marginal distribution of all variables is the first procedure for copula modeling. For instance, **Table 2** shows the statistics of distribution fitting results for the number of dry days in scenario 1.

Table 2.

Statistics of Distribution Fitting Results for the Number Of Dry Days in Scenario 1.

Distribution	**Anderson-Darling**		**AIC**
	Statistics	***p-value***	
Generalized Extreme Value	0.27261	0.95711	776.38
Weibull	0.3138	0.92727	775.88
Normal	0.43073	0.81744	777.94
Logistic	0.58951	0.65754	784.07
Extreme Value	0.63162	0.6182	783.24
Gamma	0.85789	0.4408	781.98
Log-logistic	1.0708	0.322	789.67
Lognormal	1.2084	0.26427	786.42
Inverse Gaussian	1.2687	0.24271	786.42
Exponential	26.397	6e-06	978.94

A total of ten distributions were tested for the fitting of the marginal distributions of each variable. The results in **Table 2** indicate that the Generalized Extreme Value (GEV) distribution provided the most significant fit for the number of dry days in scenario 1, as evidenced by its high p-value of 0.95711 and the lowest Akaike Information Criterion (AIC) value of 776.38 among all distributions tested. These statistics suggest that the GEV distribution effectively captures the characteristics of the data, making it the optimal choice for modeling the number of dry days in this study. In contrast, other distributions, such as the Exponential distribution, exhibited poor fit with a low p-value (6e-06), indicating a significant deviation from the observed data. This selection process highlights the

importance of evaluating multiple distribution models to identify the most appropriate one for the given dataset.

The fitting process was subsequently carried out for each variable in both scenario 1 and scenario 2. This involved applying the same methodology of estimating marginal distributions and assessing the goodness-of-fit for various distributions across all relevant variables. By conducting this comprehensive fitting analysis, the study aimed to ensure that the selected distributions accurately represent the underlying characteristics of the data for both scenarios, facilitating a more robust copula modeling approach for understanding the relationships between the climate factors and hotspot occurrences.

Table 3 presents the marginal distributions that were determined to be the best fit for the data across different scenarios, with the p-values from the Anderson-Darling hypothesis test provided in brackets. These p-values serve as an indicator of how well each distribution fits the observed data, with a focus on testing the null hypothesis (H_0) that the data follows the specified distribution. In this case, all marginal distributions exhibited p-values greater than 0.05, which indicates that there is insufficient evidence to reject the null hypothesis. This result suggests that the observed data does not significantly deviate from the fitted distributions, implying that these distributions are suitable for accurately representing the characteristics of the data.

Consequently, since the p-values indicate a good fit, all identified marginal distributions can be confidently used in subsequent analyses and modeling processes. This is a crucial step in the copula modeling framework, as ensuring that the marginal distributions are appropriately fitted provides a solid foundation for understanding the dependencies and interactions between the variables in the study. The ability to utilize these distributions in further processing enhances the robustness of the research findings and the reliability of the conclusions drawn regarding the relationships among the climate factors and hotspot occurrences.

Table 3.

Fitting Results of Marginal Distributions with Anderson-Darling Test.

Datasets	X_1	X_2	X_3
No split	Generalized Extreme Value $k = -0.475,$ $\sigma = 78.826, \mu = -214.47$ AD p-value = 0.5129	Generalized Extreme Value $k = -0.381,$ $\sigma = 12.013, \mu = 45.105$ AD p-value = 0.9571	Negative Binomial $R = 0.42177,$ $P = 0.00030405$ AD p-value = 0.6772
Neutral	Generalized Extreme Value $k = -0.447,$ $\sigma = 72.164, \mu = -204.56$ AD p-value = 0.9113	Normal $\mu = 48.189,$ $\sigma = 9.6703$ AD p-value = 0.9985	Negative Binomial $R = 0.44483,$ $P = 0.00046452$ AD p-value = 0.9908
El Nino	Generalized Extreme Value $k = -0.721,$ $\sigma = 74.768, \mu = -176.82$ AD p-value = 0.9956	Generalized Extreme Value $k = -0.532,$ $\sigma = 9.2313, \mu = 55.484$ AD p-value = 0.9951	Negative Binomial $R = 0.64896,$ $P = 0.00022228$ AD p-value = 0.7666
La Nina	Generalized Extreme Value $k = 0.241,$ $\sigma = 40.913, \mu = -288.80$ AD p-value = 0.9668	Generalized Extreme Value $k = 0.0745,$ $\sigma = 7.8011, \mu = 32.976$ AD p-value = 0.9915	Negative Binomial $R = 0.4662,$ $P = 0.0012023$ AD p-value = 0.7158

The results show that generalized extreme value distribution is the dominant distribution for variables X_1 and X_2. Generalized extreme value distribution is a probability distribution used to model the extreme values of random variables, that includes three types of extreme value distributions as a special case: the Gumbel, Fréchet, and Weibull distributions (Boudrissa et al., 2017; Jenkinson, 1955). Since X_1 refers to negative values of total precipitation, the smaller the position parameter (μ) of the generalized extreme value distribution the wetter the climate conditions. Based on that, El Nino causes drier precipitation in Kalimantan, while La Nina causes wetter precipitation than the neutral phase. Moreover, El Nino causes more dry days while La Nina causes fewer dry days than the neutral phase based on the fittest distribution of X_2.

For X_3, the negative binomial distribution is chosen because of its data type which contains many zero values (Greene, 1994). In the context of data analysis with many zero values, the parameter P in the negative binomial distribution refers to the probability of a zero value (excess zeros component). The results show that El Nino has a smaller probability of zero hotspots than the neutral phase, which means that El Nino is more likely to have high hotspots. Conversely, La Nina has a greater probability of zero hotspots than the neutral phase, so it is safer from high hotspots. A comparison of the position parameters and the probability of a zero-value parameters is presented in **Figure 4**.

Using selected distributions, each variable is estimated with the cumulative distribution function (CDF) which is known as the probability transformation. Probability transformation refers to the process of converting probabilities from one distribution to another using a specific mathematical function. Here, we transform the selected distributed original data into uniformly distributed using the CDF value. The variables resulting from this transformation are denoted U_1, U_2, and U_3 which are used for the copula parameter fitting process.

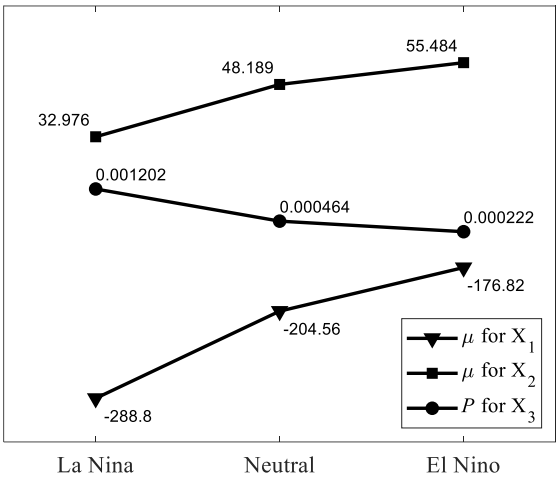

Fig. 4. Comparison of the position parameters and the probability of a zero-value parameters.

4.2. Selected Copulas and Its Parameters

The parameters of each copula function used are estimated using U_1, U_2, and U_3, then the fittest copulas are selected based on Akaike's Information Criterion (AIC). **Table 4** shows the fittest copula functions and their parameters for each condition, as well as the p-values of the Cramer-von Mises hypothesis test. Based on these p-values, all copulas have a p-value > 0.05, meaning that there is not enough evidence to reject H_0 (data comes from the selected copula). Thus, all these copulas can be used for further processing. Gaussian is the most preferred copula compared to other copula functions. The other copula functions selected are the copula functions of Galambos-180°, Gumbel, and Joe.

Table 4.

Fitting Results of Copula Functions With Cramer-Von Mises Test.

Datasets	C_1	C_2	**CvM (p-value)**
No split	Gaussian, $\rho = 0.89$	Gaussian, $\rho = 0.834$	0.059 ($p = 0.570$)
Neutral	Gaussian, $\rho = 0.9067$	Gaussian, $\rho = 0.7744$	0.040 ($p = 0.640$)
El Nino	Gaussian, $\rho = 0.8905$	Galambos-180°, $\theta = 1.7879$	0.028 ($p = 0.689$)
La Nina	Gumbel, $\theta = 3.4330$	Joe, $\theta = 3.1067$	0.028 ($p = 0.690$)

In the data without splits and in the neutral phase, the inner C_1 and outer C_2 copulas choose Gaussian as the fittest copula. A Gaussian copula can be understood as a member of the elliptical copula family. Elliptical copulas are a class of copula functions that encompass a broader range of

dependence structures beyond just the Gaussian (normal) distribution. By obtaining the Gaussian as the fittest copula, this means that there is no significant tail correlation in the constructed model. This is natural because during normal phases, climatic factors (conditions) and forest fires usually have no upper or lower extremes. However, the parameters of the Gaussian copula control the correlation or dependence between variables. Thus, due to the high value of ρ, a strong dependency is seen in the relationship between climatic factors as well as between climatic factors and hotspots.

For the El Nino phase, the Gaussian copula was chosen as the inner copula and Galambos-180° as the outer copula. The term 180° indicates that there is a rotation of the Galambos copula by 180°, also known as the survival Galambos copula (Liu et al., 2018). The Galambos copula is a specific type of copula that is often used to model extreme value dependencies. It is well-suited for capturing positive tail dependence, which means that extreme values of one variable are likely to be accompanied by extreme values of another variable.

For the La Nina phase, the Gumbel copula was chosen as the inner copula and Joe as the outer copula. Like the Galambos copula, the Gumbel copula is often used to model joint extreme events (Budiarti et al., 2018). Meanwhile, the Joe copula, is a family of copulas that generalizes the Gumbel and Clayton copulas. It allows for a smooth transition between these two copulas, providing flexibility in capturing different types of dependence structures.

4.3. Mean Regression

Using Eq. 8, the mean regression for hotspots is estimated based on the corresponding climate factor values. If the regression with one predictor is visualized using the regression line, then in the case of a two-predictor regression, the visualization uses the regression plane.

We plot the regression plane in a two-dimensional plane using a contour plot, while bubble plots are used to plot the actual hotspots. A contour plot, also known as a level plot or isoline plot, is a graphical representation used to visualize three-dimensional data on a two-dimensional surface. It is commonly used to show the variations and patterns in data that have two independent variables (represented on the x and y axes) and a dependent variable (represented through contour lines or color gradients). Here, the more yellow the color, the higher the estimated number of hotspots. Meanwhile, a bubble plot extends the concept of a scatter plot by introducing a third dimension using varying sizes of markers, usually represented as circles (or bubbles). This allows us to represent three variables in a two-dimensional space. We also provide color accents to emphasize the high and low number of hotspots that occur. Here, the larger the circle and the yellower the color, the higher the number of actual hotspots. **Figure 5** shows the regression plane for hotspots on the data without splitting based on the ENSO phases.

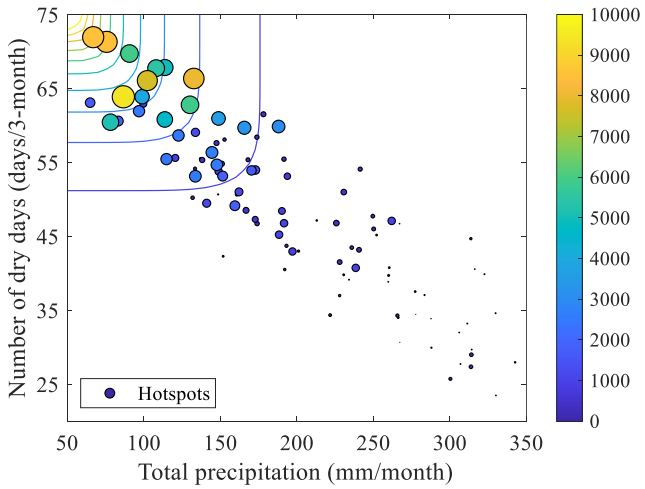

Fig. 5. Regression plane for hotspots on the data without splitting based on the ENSO phases.

Based on **Figure 5**, if the total precipitation value is smaller and the number of dry days is higher, then the number of hotspots is increasing. From the regression plane, hotspots start to appear more than 1000 hotspots in a month when the average 2-month total precipitation is less than 175 mm/month, and the number of dry days is more than 51 days in three months. The regression plane shows color gradations like the actual hotspots values indicated by the colors in the circles, indicating that the regression gives good results for estimating the number of hotspots. This is reinforced by its performance metrics which show an RMSE of 1340 hotspots and an EVS of 63.17%. This shows that the estimated hotspots can explain the variance of the actual hotspots by 63.17%. These results have not considered the ENSO phase in the modeling. Regression planes for hotspots based on ENSO phases are visualized in **Figure 6**.

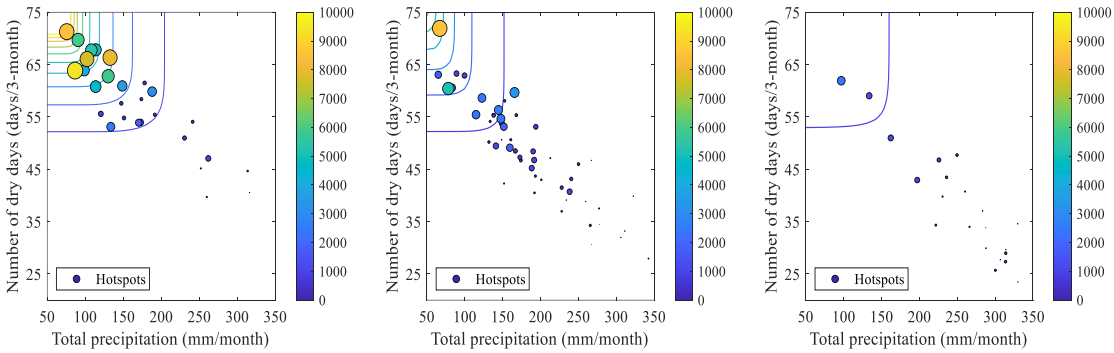

Fig. 6. Regression planes for hotspots based on ENSO phases: (a) El Nino, (b) neutral and (c) La Nina.

Figure 6 presents the regression results for hotspots in the El Nino, neutral and La Nina phases. It is interesting to note that the borders for hotspots above or below 1,000 hotspots/month emerge almost on the same line, i.e., when total precipitation is 150-200 mm/month and the number of dry days is approximately 50 days/3 months, based on the regression plane. This suggests that the ENSO phase has no impact on low hotspot situations with fewer than 1,000 hotspots/month. The significant difference can be noticed in hotspots above 1,000 hotspots/month. The regression plane, which only has one contour level at 1,000 hotspots/month, demonstrates that the La Nina phase will not touch the value of 2,000 hotspots/month. Meanwhile, in the neutral phase, this allows hotspots of up to more than 5,000 hotspots/month, though there was one event where hotspots reached up to 8,000 hotspots/month. This event occurred in 2019 during a very strong positive IOD phenomena, despite ENSO being under neutral conditions (Iskandar et al., 2022). More extreme hotspots appear during the El Nino phase. In this phase, there is a possibility that hotspots can reach more than 10,000 hotspots in a month when total precipitation is less than 80 mm/month, and the number of dry days is more than 70 days in 3 months.

Table 5 shows the performance metrics of regression model based on ENSO phases. The RMSE value describes that the El Nino phase has the greatest and the La Nina phase has the lowest. This number is proportional to the number of monthly hotspots that occur in each phase. Because the number of hotspots is so significant during the El Nino phase, the regression model's RMSE is also very high. Likewise, for the La Nina phase, the number of hotspots is relatively small.

Table 5.
Performance metrics of regression model based on ENSO conditions.

Phase	RMSE	EVS
El Nino	1789	66.74%
Neutral	1025	54.35%
La Nina	295	77.43%

Based on the EVS value, the regression model in the neutral phase is the lowest: 54.35%. This means that the regression model can explain the variance of the original data by 54.35%. This also implies that the neutral phase has high variance when compared to other phases. In this neutral phase, hotspots can be very high or very low with quite high extreme values as happened in 2019. Meanwhile, the La Nina phase produces the highest EVS values, which corresponds to the low variance in this phase. In this phase, monthly hotspots tend to be low and there have been no extreme hotspot events in this phase.

Finally, the regression model separated by ENSO phases yields an RMSE of 1204 hotspots and an EVS of 70.01%. This is an improvement compared to the previous model which did not differentiate based on ENSO phases, which only produced an EVS of 63.17%. This highlights the importance of ENSO phases in modeling hotspots in Kalimantan. In comparison to previous research by Najib et al. (2022b), the RMSE results are relatively close; that study, which also classified ENSO phases, reported an RMSE of 1189 hotspots when using total precipitation as a single predictor, while using the number of dry days as a predictor yielded an RMSE of 1110 hotspots. Therefore, the results obtained in this study remain highly acceptable when compared to those findings. However, there is still potential for further research to enhance the accuracy of this copula regression model, aiming to achieve an RMSE that approaches or surpasses that of the single-variable predictor copula regression models. One potential avenue for improvement is the implementation of more advanced structures, such as vine copulas, which could better capture the dependencies among the variables and enhance model performance.

5. CONCLUSIONS

This study focuses on modeling hotspots based on total precipitation and the number of dry days using nested 3-copula regression. Nested copulas offer a simple way to construct high-dimensional multivariate distributions. By estimating conditional probability values, the number of monthly hotspots in Kalimantan can be estimated. Even though there are many advantages offered by nested copulas, there are several notes that need to be considered in regression using copulas. One of them is the dependency between the predictors. In section 4 it was revealed that we used negative data from total precipitation rather than the original data. This is because if you use total precipitation data, it will result in the relationship between total precipitation and the number of dry days being negative. This means that the relationship between the predictor data and the response variable is different. Thus, cancellation of the response variable occurs when a joint distribution between the predictor data is formed. In other words, if the predictor has a negative relationship, then the relationship between the joint distribution of the predictor and the response variable will be low. This results in predictions being inaccurate. Therefore, the relationship between predictor data and response data is very important in nested copula-based regression models. This conclusion is supported by McNeil (2008) which states that the nested copula model allows to construct joint distributions with different levels of *positive dependence* in different bivariate margins.

Regression results for hotspots using nested copula regression show satisfactory performance where overall the model based on ENSO phases can explain the variance of hotspot data up to 70%. From the regression plane that is formed, the ENSO phase does not really affect the hotspots at low levels. The ENSO phase is very influential when talking about high or extreme hotspots, where El Nino is the phase that has the greatest opportunity for extreme hotspots to occur, even up to more than 10,000 hotspots per month. Meanwhile, La Nina is the safest phase for extreme hotspots. From the performance metrics, it can be concluded that the ENSO phase has a big influence on modeling hotspots in Kalimantan based on total precipitation and the number of dry days.

Although the nested copula approach significantly improves the symmetric copula approach, it is still insufficient to capture all possible interrelationships among n-dimensional random variables. Based on the multivariate density decomposition (Joe, 1997), other approaches such as Pair-Copula Construction (PCC) allow for the free specification of copulas that are hierarchical in nature. There are two main types of PCCs, canonical (C)-vines and drawable (D)-vines copula (Zhang & Singh,

2019a). This study can be developed further by applying a similar regression concept, with a copula model construction process using the vine copula approach. However, it should be noted that the construction of the vine copula model is more complicated than the nested copula model in the high-dimensional case.

ACKNOWLEDGEMENT

The authors want to take this opportunity to acknowledge their gratitude to the Department of Mathematics at IPB University, as well as the Meteorological, Climatological, and Geophysical Agency, for their constant encouragement and help during this study.

Code Availability

We use MATLAB for all computational processes conducted in this study. The codes can be downloaded from https://mkhoirun-najiboi.github.io/mycopula/.

Conflict of Interest

The authors declare that they have no conflict of interest for this research.

REFERENCES

Aas, K., & Berg, D. (2009). Models for construction of multivariate dependence - a comparison study. *European Journal of Finance*, *15*(7–8), 639–659. https://doi.org/10.1080/13518470802588767

Adrianto, H. A., Spracklen, D. V., Arnold, S. R., Sitanggang, I. S., & Syaufina, L. (2020). Forest and land fires are mainly associated with deforestation in Riau Province, Indonesia. *Remote Sensing*, *12*(1). https://doi.org/10.3390/RS12010003

Aflahah, E., Hidayati, R., Hidayat, R., & Alfahmi, F. (2018). Pendugaan hotspot sebagai indikator kebakaran hutan di Kalimantan berdasarkan faktor iklim. *Journal of Natural Resources and Environmental Management*, *9*(2), 405–418. http://dx.http//journal.ipb.ac.id/index.php/jpsl

Anderson, T. W., & Darling, D. A. (1954). A Test of Goodness of Fit. *Journal of the American Statistical Association*, *49*(268), 765–769. https://doi.org/10.1080/01621459.1954.10501232

Boudrissa, N., Cheraitia, H., & Halimi, L. (2017). Modelling maximum daily yearly rainfall in northern Algeria using generalized extreme value distributions from 1936 to 2009. *Meteorological Applications*, *24*(1), 114–119. https://doi.org/10.1002/met.1610

Bruster-Flores, J. L., Ortiz-Gómez, R., Ferriño-Fierro, A. L., Guerra-Cobián, V. H., Burgos-Flores, D., & Lizárraga-Mendiola, L. G. (2019). Evaluation of precipitation estimates CMORPH-CRT on regions of Mexico with different climates. *Water (Switzerland)*, *11*(8). https://doi.org/10.3390/w11081722

Budiarti, R., Wigena, A. H., Purnaba, I. G. P., & Achsani, N. A. (2018). Modelling the Dependence Structure of Financial Assets: A Bivariate Extreme Data Study. *IOP Conference Series: Earth and Environmental Science*, *187*(1). https://doi.org/10.1088/1755-1315/187/1/012003

Czado, C., Bax, K., Sahin, Ö., Nagler, T., Min, A., & Paterlini, S. (2022). Vine Copula Based Dependence Modeling in Sustainable Finance. *The Journal of Finance and Data Science*, *8*, 309–330. https://doi.org/10.1016/j.jfds.2022.11.003

Fanin, T., & Van Der Werf, G. R. (2017). Precipitation-fire linkages in Indonesia (1997-2015). *Biogeosciences*, *14*(18), 3995–4008. https://doi.org/10.5194/bg-14-3995-2017

Firmansyah, A. J., Nurjani, E., & Sekaranom, A. B. (2022). Effects of the El Niño-Southern Oscillation (ENSO) on rainfall anomalies in Central Java, Indonesia. *Arabian Journal of Geosciences*, *15*(24). https://doi.org/10.1007/s12517-022-11016-2

Gayawan, E., Egbon, O. A., & Adegboye, O. (2023). Copula based trivariate spatial modeling of childhood illnesses in Western African countries. *Spatial and Spatio-Temporal Epidemiology*, *46*. https://doi.org/10.1016/j.sste.2023.100591

Greene, J. (1994). Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models. *NYU Working Paper No. EC-94-10*, *9*, 265–265. http://ssrn.com/abstract=1293115

Gringorten, I. I. (1963). A plotting rule for extreme probability paper. *Journal of Geophysical Research*, *68*(3), 813–814. https://doi.org/10.1029/jz068i003p00813

Harrison, M. E., Capilla, B. R., Thornton, S. A., Cattau, M. E., & Page, S. E. (2016). Impacts of the 2015 fire season on peat-swamp forest biodiversity in Indonesian Borneo. *15Th International Peat Congress*, *2006*(August 2006), 713–717.

Hein, L., Spadaro, J. V., Ostro, B., Hammer, M., Sumarga, E., Salmayenti, R., Boer, R., Tata, H., Atmoko, D., & Castañeda, J. P. (2022). The health impacts of Indonesian peatland fires. *Environmental Health: A Global Access Science Source*, *21*(1). https://doi.org/10.1186/s12940-022-00872-w

Heredia-Zavoni, E., & Montes-Iturrizaga, R. (2019). Modeling directional environmental contours using three dimensional vine copulas. *Ocean Engineering*, *187*. https://doi.org/10.1016/j.oceaneng.2019.06.007

Iskandar, I., Lestari, D. O., Saputra, A. D., Setiawan, R. Y., Wirasatriya, A., Susanto, R. D., Mardiansyah, W., Irfan, M., Rozirwan, Setiawan, J. D., & Kunarso. (2022). Extreme Positive Indian Ocean Dipole in 2019 and Its Impact on Indonesia. *Sustainability (Switzerland)*, *14*(22). https://doi.org/10.3390/su142215155

Jenkinson, A. F. (1955). The Frequency Distribution of the Annual Maximum (or Minimum) of Meteorological Elements. *Quarterly Journal of the Royal Meteorological Society*, *81*, 158–171.

Jha, B. K., & Danjuma, Y. J. (2020). Unsteady Dean flow formation in an annulus with partial slippage: A riemann-sum approximation approach. *Results in Engineering*, *5*. https://doi.org/10.1016/j.rineng.2019.100078

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall. https://doi.org/10.1201/9780367803896

Kolev, N., & Paiva, D. (2009). Copula-based regression models: A survey. *Journal of Statistical Planning and Inference*, *139*(11), 3847–3856. https://doi.org/10.1016/j.jspi.2009.05.023

Kurniadi, A., Weller, E., Min, S. K., & Seong, M. G. (2021). Independent ENSO and IOD impacts on rainfall extremes over Indonesia. *International Journal of Climatology*, *41*(6), 3640–3656. https://doi.org/10.1002/joc.7040

Lee, B. P. Y. H., Davies, Z. G., & Struebig, M. J. (2017). Smoke pollution disrupted biodiversity during the 2015 El Nino fires in Southeast Asia. *Environmental Research Letters*, *12*(9). https://doi.org/10.1088/1748-9326/aa87ed

Lestari, D. O., Sutriyono, E., Sabaruddin, & Iskandar, I. (2018). Severe Drought Event in Indonesia Following 2015/16 El Niño/positive Indian Dipole Events. *Journal of Physics: Conference Series*, *1011*(1). https://doi.org/10.1088/1742-6596/1011/1/012040

Li, H.-N., Zheng, X.-W., & Li, C. (2019). Copula-Based Joint Distribution Analysis of Wind Speed and Direction. *Journal of Engineering Mechanics*, *145*(5). https://doi.org/10.1061/(asce)em.1943-7889.0001600

Li, Z., Shao, Q., Tian, Q., & Zhang, L. (2020). Copula-based drought severity-area-frequency curve and its uncertainty, a case study of Heihe River basin, China. *Hydrology Research*, *51*(5), 867–881. https://doi.org/10.2166/nh.2020.173

Liu, J., Sirikanchanarak, D., Sriboonchitta, S., & Xie, J. (2018). Analysis of Household Consumption Behavior and Indebted Self-Selection Effects: Case Study of Thailand. *Mathematical Problems in Engineering*, *2018*. https://doi.org/10.1155/2018/5486185

Ma, M., Wang, X., Liu, N., Song, S., & Wang, S. (2022). Nested Copula Model for Overall Seismic Vulnerability Analysis of Multispan Bridges. *Shock and Vibration*, *2022*. https://doi.org/10.1155/2022/3001933

McNeil, A. J. (2008). Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation*, *78*(6), 567–581. https://doi.org/10.1080/00949650701255834

Najib, M. K., Nurdiati, S., & Sopaheluwakan, A. (2021). Quantifying the joint distribution of drought indicators in Borneo fire-prone area. *IOP Conference Series: Earth and Environmental Science*, *880*. https://doi.org/10.1088/1755-1315/880/1/012002

Najib, M. K., Nurdiati, S., & Sopaheluwakan, A. (2022a). Copula-based joint distribution analysis of the ENSO effect on the drought indicators over Borneo fire-prone areas. *Modeling Earth Systems and Environment*, *8*(2), 2817–2826. https://doi.org/10.1007/s40808-021-01267-5

Najib, M. K., Nurdiati, S., & Sopaheluwakan, A. (2022b). Multivariate fire risk models using copula regression in Kalimantan, Indonesia. *Natural Hazards*, *113*(2), 1263–1283. https://doi.org/10.1007/s11069-022-05346-3

Najib, M. K., Nurdiati, S., & Sopaheluwakan, A. (2023). Prediction of hotspots pattern in Kalimantan using copula-based quantile regression and probabilistic model: a study of precipitation and dry spells across varied ENSO conditions. *Vietnam Journal of Earth Sciences, Early Access*. https://doi.org/10.15625/2615-9783/19302

Nelsen, R. B. (2006). *An introduction to copulas*. Springer.

Nugroho, A. R. (2022). *Teleconnections of El Niño–Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) to streamflow in Java, Indonesia* [Dissertation, Gifu University]. http://hdl.handle.net/20.500.12099/88935

Nurdiati, S., Bukhari, F., Julianto, M. T., Sopaheluwakan, A., Aprilia, M., Fajar, I., Septiawan, P., & Najib, M. K. (2022). The impact of El Niño southern oscillation and Indian Ocean Dipole on the burned area in Indonesia. *Terrestrial, Atmospheric and Oceanic Sciences*, *33*(15). https://doi.org/10.1007/S44195-022-00016-0

Nurdiati, S., Najib, M. K., & Thalib, A. S. (2022). Joint Distribution And Coincidence Probability Of The Number Of Dry Days And The Total Amount Of Precipitation In Southern Sumatra Fire-Prone Area. *Geographia Technica*, *17*(2), 107–118. https://doi.org/10.21163/GT_2022.172.10

Oyedele, A. A., Ajayi, A. O., Oyedele, L. O., Bello, S. A., & Jimoh, K. O. (2023). Performance evaluation of deep learning and boosted trees for cryptocurrency closing price prediction. *Expert Systems with Applications*, *213*. https://doi.org/10.1016/j.eswa.2022.119233

Pan, S., Joe, H., & Li, G. (2023). Conditional Inferences Based on Vine Copulas with Applications to Credit Spread Data of Corporate Bonds. *Journal of Financial Econometrics*, *21*(3), 714–741. https://doi.org/10.1093/jjfinec/nbab016

Pleis, J. R. (2018). *Mixtures of discrete and continuous variables: Considerations for dimension reduction* [Dissertation]. University of Pittsburgh.

Rahman, R. A., White, B., & Ma, C. (2024). The effect of growth, deforestation, forest fires, and volcanoes on Indonesian regional air quality. *Journal of Cleaner Production*, *457*. https://doi.org/10.1016/j.jclepro.2024.142311

Rodysill, J. R., Russell, J. M., Vuille, M., Dee, S., Lunghino, B., & Bijaksana, S. (2019). La Niña-driven flooding in the Indo-Pacific warm pool during the past millennium. *Quaternary Science Reviews*, *225*. https://doi.org/10.1016/j.quascirev.2019.106020

Saad, C., El Adlouni, S., St-Hilaire, A., & Gachon, P. (2015). A nested multivariate copula approach to hydrometeorological simulations of spring floods: the case of the Richelieu River (Québec, Canada) record flood. *Stochastic Environmental Research and Risk Assessment*, *29*(1), 275–294. https://doi.org/10.1007/s00477-014-0971-7

Schölzel, C., & Friederichs, P. (2008). Multivariate non-normally distributed random variables in climate research - Introduction to the copula approach. *Nonlinear Processes in Geophysics*, *15*(5), 761–772. https://doi.org/10.5194/npg-15-761-2008

Segers, J., & Uyttendaele, N. (2014). Nonparametric estimation of the tree structure of a nested Archimedean copula. *Computational Statistics and Data Analysis*, *72*, 190–204. https://doi.org/10.1016/j.csda.2013.10.028

Septiawan, P., Nurdiati, S., & Sopaheluwakan, A. (2019). Numerical Analysis using Empirical Orthogonal Function Based on Multivariate Singular Value Decomposition on Indonesian Forest Fire Signal. *IOP Conference Series: Earth and Environmental Science*, *303*. https://doi.org/10.1088/1755-1315/303/1/012053

Serinaldi, F., & Grimaldi, S. (2007). Fully Nested 3-Copula: Procedure and Application on Hydrological Data. *Journal of Hydrologic Engineering*, *12*(4), 420–430. https://doi.org/10.1061/(asce)1084-0699(2007)12:4(420)

Sklar, M. (1959). Fonctions de Répartition àn Dimensions et Leurs Marges. *Publications de L'Institut de Statistique de L'Université de Paris*, *8*, 229–231.

Uda, S. K., Hein, L., & Atmoko, D. (2019). Assessing the health impacts of peatland fires: a case study for Central Kalimantan, Indonesia. *Environmental Science and Pollution Research*, *26*(30), 31315–31327. https://doi.org/10.1007/s11356-019-06264-x

Wahiduzzaman, M., Cheung, K., Tang, S., & Luo, J. J. (2022). Influence of El Niño–Southern Oscillation on the long-term record of floods over Bangladesh. *Theoretical and Applied Climatology*, *147*(1–2), 173–184. https://doi.org/10.1007/s00704-021-03814-7

Wasis, B., Winata, B., & Marpaung, D. R. (2018). Impact of land and forest fire on soil fauna diversity in several land cover in Jambi Province, Indonesia. *Biodiversitas*, *19*(2), 660–666. https://doi.org/10.13057/biodiv/d190249

Whelan, N. (2004). Sampling from Archimedean copulas. *Quantitative Finance*, *4*(3), 339–352. https://doi.org/10.1088/1469-7688/4/3/009

Xie, P., Joyce, R., Wu, S., Yoo, S.-H., Yarosh, Y., Sun, F., & Lin, R. (2019). NOAA Climate Data Record (CDR) of CPC Morphing Technique (CMORPH) High Resolution Global Precipitation Estimates, Version 1. In *NOAA National Centers for Environmental Information*. NOAA National Centers for Environmental Information. https://doi.org/10.25921/w9va-q159

Xie, P., Joyce, R., Wu, S., Yoo, S. H., Yarosh, Y., Sun, F., & Lin, R. (2017). Reprocessed, bias-corrected CMORPH global high-resolution precipitation estimates from 1998. *Journal of Hydrometeorology*, *18*(6), 1617–1641. https://doi.org/10.1175/JHM-D-16-0168.1

Yananto, A., & Dewi, S. (2016). Analisis Kejadian El Nino Tahun 2015 dan Pengaruhnya Terhadap Peningkatan Titik Api di Wilayah Sumatera dan Kalimantan. *Jurnal Sains \& Teknologi Modifikasi Cuaca*, *17*(1), 11. https://doi.org/10.29122/jstmc.v17i1.544

Zahra, R. A., Nurjani, E., & Sekaranom, A. B. (2023). The Analysis of Fire Hotspot Distribution in Kalimantan and Its Relationship with ENSO Phases. *Quaestiones Geographicae*, *42*(1), 75–86. https://doi.org/10.14746/quageo-2023-0006

Zhang, L., & Singh, V. P. (2019a). Asymmetric Copulas: High Dimension. In *Copulas and their Applications in Water Resources Engineering* (pp. 172–241). Cambridge University Press. https://doi.org/10.1017/9781108565103.006

Zhang, L., & Singh, V. P. (2019b). Symmetric Archimedean Copulas. In *Copulas and their Applications in Water Resources Engineering* (pp. 123–171). Cambridge University Press. https://doi.org/10.1017/9781108565103