Vietnam Journal of Earth Sciences, 46(1), 12-33, https://doi.org/10.15625/2615-9783/19302

Vietnam Journal of Earth Sciences
https://vjs.ac.vn/index.php/jse

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

Mohamad K. Najib[1], Sri Nurdiati[1*], Ardhasena Sopaheluwakan[2]

[1]*Division of Computational Mathematics, Department of Mathematics, IPB University, Bogor 16680, Indonesia*
[2]*Center for Applied Climate Services, Agency for Meteorology, Climatology, and Geophysics, Jakarta 10720, Indonesia*

Received 07 March 2023; Received in revised form 09 July 2023; Accepted 30 October 2023

ABSTRACT

Hotspots in Kalimantan are significantly correlated with local and global climatic conditions. These hotspots have been represented in previous explorations using copula-based mean regression technique. However, this study focused on advancing hotspots model through the use of copula-based quantile regression. Probabilistic method was also introduced to depict the characteristics of hotspots in Kalimantan. To achieve this objective, the technique of the inference of functions for margins was applied. Several copula functions, including Gumbel, Clayton, Frank, Joe, Galambos, BB1, BB6, BB7, and BB8, were meticulously chosen. The selection of the most suitable copula was based on the results of the Anderson-Darling and Cramer-von Mises hypothesis tests. The results showed that the combination of quantile and mean regression yielded satisfactory results. Moreover, an uncertainty range was established by assessing the outermost quantile, which aided the assessment of the reliability of estimated hotspots. Probabilistic model introduced a fresh viewpoint to modeling process. Instead of forecasting an exact value, model estimated the probability of hotspots occurrences based on specific climatic conditions. Among the three scenarios examined, precipitation-based model showed an average accuracy of 89.7%, while dry spells-based outperformed the value with a score of 90.3%. After evaluating the results from both regression and probabilistic model, dry spells-based method outperformed precipitation-based. On the other hand, precipitation-based performed better in capturing certain minor details compared to dry spells-based model.

Keywords: Copula-based quantile regression, El Nino-Southern Oscillation (ENSO), hotspots, probabilistic model, rotated copula.

1. Introduction

Indonesia is known to have experienced continuous forest and land fires for the space of three decades (Thoha et al., 2023). These fires are climatological disasters that contribute to climate change, primarily due to the emission of greenhouse gases resulting from biomass burning (Enríquez-de-Salamanca, 2020). The majority of the extensive forest fires in Indonesia took place on peatlands. The country boasts the largest expanse of tropical peatlands in the world and

*Corresponding author, Email: nurdiati@apps.ipb.ac.id

ranks fourth globally in terms of entire peatland coverage, which extends across Kalimantan, Sumatra, and Papua (Page et al., 2011). In 1997 and 2015, severe fire incidents occurred in tropical peatland regions across Southeast Asia, including Indonesia. These events led to the emission of a significant amount of CO_2 into the atmosphere, ranging from 0.8 to 9.43 Gt, during a single fire season. This emission is equivalent to approximately 30% of the total global fossil fuel emissions observed in 2020 (Horton et al., 2022).

Repeated forest and land fires lead to substantial post-fire expenses in Indonesia. The country experienced losses of approximately IDR 221 trillion and IDR 75 trillion during the extreme fires of 2015 and 2019, respectively (Thoha et al., 2023). Beyond causing economic losses, these fires also exert a negative influence on public health due to air pollution (Marlier et al., 2019) and disrupt the balance of flora and fauna ecosystems (Harrison et al., 2009). Furthermore, fires pose significant threats to various aspects of society, including the physical environment, economy, agriculture, and social structure. Peatland fires trigger social and livelihood changes, intensifying the vulnerability of local populations through recurrent damage to infrastructure and agriculture. This situation raises tensions and necessitates a shift from subsistence livelihoods to innovative practices that transform degraded peatlands into economically productive landscapes (Medrilzam et al., 2014; Goldstein, 2020; Lounela, 2021).

The phenomenon of ocean-atmosphere interaction known as El-Nino and Southern Oscillation (ENSO) profoundly impacts climatic conditions, including those in Indonesia, specifically in the Pacific Ocean (Nicholls, 1984; Amirudin et al., 2020). The warm phase of ENSO, commonly referred to as El Nino, results in cooler sea temperatures around maritime regions of the country. Consequently, evaporation rates decrease and condensation is hindered. Reduced cloud cover leads to diminished precipitation on the Indonesian mainland and an increased frequency of dry days (Philander, 1983; Iskandar et al., 2019). This conditions culminates in reduced water availability across several parts of the nation, including Kalimantan, where vegetation becomes parched, and the risk of forest and land fires escalates (Salafsky, 1994; Jim 1999; Khan et al., 2020). Previous study shows that the El Nino phase has a significant influence on forest and land fires in Kalimantan, the island facing the most severe forest fires (Nurdiati et al., 2022a). For example, the well-known forest fires in 1997 and 2015 occurred concurrently with extreme El Nino conditions (Huijnen et al., 2016; Fanin and van der Werf, 2017). This shows the need for model capable of establishing the connection between ENSO phase variations and climate indicators, such as precipitation levels and the frequency of rainless days (dry spells). It should be noted that these indicators are closely associated with forest fires in Kalimantan (Nurdiati et al., 2021).

Numerous study analysts have developed model to predict forest fire indicators (including hotspots and burnt areas) using various methods, such as artificial neural networks (Nikonovas et al., 2022; Mezbahuddin et al., 2023), Bayesian model (Ardiyani et al., 2023; Charizanos and Demirhan, 2023; Koh et al., 2023), machine learning (Grari et al., 2022; Shao et al., 2022), polynomial and generalized logistic functions (Nurdiati et al., 2022c), and copula regression (Najib et al., 2022b). Copula regression model computes the mean value of conditional distribution (known as copula mean regression) of hotspots with respect to climate indicators (e.g., total precipitation and dry

spells) in the three distinct ENSO phases. By computing the mean value of conditional distribution, a higher hotspots value significantly contributes to the average. As a result, an alternative method is essential for calculating the median value or other quantile values of conditional distribution of hotspots generated through copula function. This method is often referred to as copula-based quantile regression (El Adlouni, 2018; Li et al., 2021; Zahiri et al., 2022).

Copula function in multivariate analysis offers several key advantages (1) flexibility in selecting arbitrary marginal distribution functions and their dependence structure, (2) scalability to involve more than two variables, and (3) capability to separately analyze the marginal distribution function and its dependence structure (Salvadori et al., 2005; Tahroudi et al., 2020). Although copula theory is not novel, the volume of literature discussing the functions in diverse contexts has significantly increased in the past decade (Wahl et al., 2012; Tootoonchi et al., 2022). Copula-based quantile regression uses copula function to unveil relationships between variables, thereby constructing joint probabilities between them. Quantile regression, which is an extension of linear regression, facilitates regression across the entire time series of data, accommodating values exceeding the chosen quantile. Consequently, when high (low) quantiles are under consideration, its regression allows exploration of the upper (lower) tail of the probability distribution function (Treppiedi et al., 2021). One advantage of quantile regression compared to ordinary least squares is the heightened robustness against outliers in the response measurements (Koenker, 2005). Copula-based quantile regression offers flexibility for general time-to-event datasets (El Adlouni, 2018; Pan and Joe, 2022) and finds application in drought and flood monitoring (Zhang et al., 2023).

Numerous journals explore the application of copula quantile regression model across various scientific domains. Abdallah et al. (2022) advocated copula-based quantile regression as a promising alternative technique for estimating daily ETo under hyper-arid climate conditions worldwide. Zhang et al. (2023) used polarimetric decomposition and copula quantile regression to probabilistically estimate surface soil moisture (SSM). In the study, copula quantile regression established an uncertainty range for the SSM estimate, facilitating an assessment of its reliability. Wu et al. (2022) proposed an agricultural drought prediction model grounded in D-vine copula quantile regression, taking into account diverse factors related to agricultural drought mechanisms in China.

This study aims to model hotspots in Kalimantan using copula-based mean and quantile regression, ultimately predicting the likelihood of its occurrence under specific climatic conditions. The organization of this article includes, Section 2 providing a detailed overview of the study area and dataset. Section 3 elucidates the methodologies consisting of copula functions, parameter estimation, copula-based joint probability, regression, quantile regression, and probabilistic model. The discussion and presentation of results are encapsulated in Section 4, while Section 5 concludes the article. The results are expected to significantly contribute to the development of an early warning system for forest fires in Indonesia.

2. Study area and datasets

Indonesia is the largest tropical peatland in the world, spanning a total area of 13.43 million hectares across three primary islands, namely Sumatera, Kalimantan, and Papua. This study specifically centers on Kalimantan, including 33.8% of peatlands, distributed

among five provinces, including West, East, Central, South, and North Kalimantan (Yuwati et al., 2021). In the 2019 fire incident, Central and West Kalimantan were known as the provinces with the highest count of hotspots, closely trailed by Jambi, Riau, and South Sumatra provinces on Sumatra Island (Thoha et al., 2023).

Hotspots constitute a common method to rapidly monitor forest fires across expansive regions. These hotspots denote areas with relatively elevated surface temperatures compared to their surroundings, established based on specific thresholds observed through remote sensing satellites (Coutts et al., 2016; Vatresia et al., 2022). Hotspots are identified by detecting instances of forest and land fires within certain pixel dimensions. Detection occurs when the satellite directly observes hotspots under relatively cloud-free conditions, using a designated algorithm (Nainggolan et al., 2020). Precipitation and dry spells (defined as days with daily precipitation below one millimeter) serve as predictors within three distinct ENSO conditions (La Nina, neutral, and El Nino). Table 1 shows the source and summary of the data.

Table 1. Source and summary of data

Datasets	Spatial resolution	Periods	Source
Precipitation	0.25	2001–2020	Monthly CMORPH-CRT (https://ftp.cpc.ncep.noaa.gov/precip/PORT/SEMDP/CMORPH_CRT/DATA/)
Dry spells	0.25	2001–2020	Processed using daily CMORPH-CRT (https://ftp.cpc.ncep.noaa.gov/precip/PORT/SEMDP/CMORPH_CRT/DATA/)
Hotspots	0.25	2001–2020	Agency for Meteorology, Climatology, and Geophysics (BMKG) Indonesia
ENSO index	–	2001–2020	Climate Prediction Center, NCEP, NOAA processed using ERSSTv5 data (https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php)

The used data has been processed within fire-prone regions of Kalimantan (Najib et al., 2021). Kalimantan predominantly experiences two seasonal rainfall patterns: equatorial and monsoonal. Using clustering, hotspots data in this region is categorized into clusters, revealing areas susceptible to forest fires. Many of these regions lie in central, western, and southern Kalimantan, which adhere to a monsoonal rainfall pattern. In these selected zones, data is aggregated to capture general characteristics of rainfall, dry spells, and hotspots occurrences in fire-prone Kalimantan regions. Dependency analysis was then conducted on the monthly hotspots data, revealing that the two-month average of total precipitation and the three-month cumulative dry spells exerted the most substantial influence on monthly hotspots. Concurrently, ONI data divides the monthly dataset based on three ENSO phases, including La Nina (ONI $\leq$ -0.5), El Nino (ONI $\geq$ 0.5), and neutral for other instances. The data are grouped into four categories, consisting of monthly hotspots counts in Kalimantan, a two-month average of total precipitation, a three-month cumulative of dry spells, and the Oceanic Nino Index (ONI) as ENSO conditions indicator. All data were collected between July and November from 2001 to 2020.

3. Methods

3.1. Copula function

The multivariate normal distribution function served as the most basic method for

crafting a joint distribution among two or more variables. However, this method relied on assumptions that were rarely encountered in actual data, particularly climate-related data. A crucial assumption was the requirement for every linear combination of individual component variables to adhere to a normal distribution (Yan, 2006; Danaher and Smith, 2011). Consequently, flexibility for the marginal variables was lacking.

Copula function presented an alternative avenue to construct a joint distribution in cases where the normality assumption was unmet (Sartika et al., 2019; Masseran and Hussain, 2020; Pambabay-Calero et al., 2021). When faced with non-normally distributed marginal variables, copula function facilitated the creation of a joint distribution (Vuolo, 2017). Remarkably, this construction remained feasible even when each variable exhibited a distinct distribution (Salvadori and De Michele, 2007; Tahroudi et al., 2022). Copula, functioning as a connector, constructed a multivariate distribution function through its univariate marginal distribution function (Schölzel and Friederichs, 2008; Amini et al., 2022). Another perspective defined copula function as a multivariate distribution function, where the marginal function was uniformly distributed over the interval [0,1] (Nelsen, 2006). In the bivariate context, copula function $C: I^2 \rightarrow I$, with $I \in [0,1]$, mapped the univariate marginal distribution functions of variables X and Y denoted as F_X and F_Y, respectively to the multivariate distribution function F_{XY}, illustrated as follows (Sklar 1959):

$$F_{XY}(X,Y) = C\big(F_X(X), F_Y(Y)\big) \quad (1)$$

where X represents the climate indicators (e.g., precipitation or dry spells) and Y represents hotspots data (Najib et al., 2022a).

In contrast to the multivariate normal distribution, copula function could tailor the joint shape of the distribution based on the presence or absence of tail dependencies between variables. Copula exhibited greater flexibility than the multivariate normal distribution because each copula function captured distinct tail dependencies (Babić et al., 2019; Li et al., 2020). For example, the Gumbel Copula characterized upper tail dependencies, while the Clayton Copula handled lower tail dependencies.

3.2. Parameter estimation

This study used a two-step methodology to estimate copula parameters, specifically the Inference of Function for Margins (IFM) (Joe, 1997; 2005). The first step involved estimating the univariate marginal distribution for each variable pair. Various univariate distribution functions were fitted to the data, involving Extreme Value, Logistic, Weibull, Normal, Log-normal, Log-logistic, Gamma, Inverse-Gaussian, and Generalized Extreme Value for climate indicators, and Negative Binomial for hotspots data. The suitable distribution function was chosen through the Anderson-Darling hypothesis test (Anderson, 2011).

The second step of the IFM method included estimating copula parameters using transformed variables in line with the corresponding marginal distributions, ensuring the transformed variables adhered to a uniform distribution within the [0,1] interval. This transformation, often referred to as the probability integral transformation (Yan 2007; Brechmann, 2014), underpinned the analysis. This study used diverse copula functions, including Gumbel, Clayton, Frank, Joe, Galambos, BB1, BB6, BB7, and BB8. Determining the most suitable copula function involved the Cramer-von Mises (CvM) hypothesis test (Berg, 2009) and used criteria such as RMSE and AIC for final selection. Comprehensive details regarding the steps for fitting copula parameters were outlined in Najib et al. (2022b).

3.3. Copula-based joint probability

If F_X and F_Y were continuous cumulative distribution functions (CDFs), the unique copula function C (Eq. 1) could be expressed as follows:

$$C(u,v) = \int_0^u \int_0^v c(u,v)\, du\, dv \qquad (2)$$

where u and v represented the transformed random variables, expressed as $u = F_X(x)$ and $v = F_Y(y)$ (Schölzel and Friederichs 2008). The function c was known as copula density function or dependence function (Mirabbasi et al., 2012; Ly et al., 2019; Mukhopadhyay and Parzen, 2020), as it revealed the dependencies between random variables. This dependency function had originally been introduced by Hoeffding in 1940 and was defined as the quotient of the joint density and the product of its marginal densities. A fundamental implication of Sklar theorem (1959) was that any joint probability between two variables could be represented as the product of the marginal probability density and copula density, given by:

$$f_{XY}(x,y) = f_X(x) \cdot f_Y(y) \cdot c\big(F_X(x), F_Y(y)\big) \qquad (3)$$

The joint probability distribution of X_1 and X_2 could be uniquely determined when F_1 and F_2 were continuous CDFs. This method could be extended to incorporate a combination of continuous and discrete random variables. If F_X and F_Y entailed both continuous and discrete CDFs, then copula would exclusively ascertain the joint probability distribution across the scope of $F_X \times F_Y$ (Pleis, 2018).

3.4. Copula-based regression

Copula-based regression (also known as copula regression) offered an alternative to the conventional regression method, providing the significant advantage of independently modeling the marginal distribution and the dependency structure (Joe, 2014; Li et al., 2022). This method removed restrictions on the distribution function of the data used (Thevaraja and Rahman, 2019). Copula regression leveraged conditional probability function obtained from copula function and then estimated the expected value (mean) of conditional probability function as follows:

$$E(Y|X = x) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x)\, dy \qquad (4)$$

where

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = c\big(F_X(x), F_Y(y)\big) \cdot F_Y(y) \qquad (5)$$

is conditional probability function of a variable Y given a set of variables X. Copula regression is also called copula mean regression.

3.5. Copula-based quantile regression

Quantile regression extended the classical regression method by incorporating comprehensive information about conditional distribution of the response variable (Koenker and Hallock 2001; Herawati 2020). In quantile regression, the focus was on approximating conditional functions of the response variable Y given a set of variables X. The method used conditional distribution function of the response variable Y, which was given by:

$$F_{Y|X}(Y < y|X = x) = \int_0^y f_{Y|X}(y|x)\, dx \qquad (6)$$

The function $f_{Y|X}$ represented conditional probability function, while $F_{Y|X}$ represented conditional distribution function. The q-th quantile of conditional distribution function $F_{Y|X}$ was calculated as follows

$$\hat{y} = F_{Y|X}^{-1}(q) \qquad (7)$$

where $q \in [0,1]$. The value of q served as the hyper-parameter or tuning parameter that needed to be determined. In this study, the chosen q was the one that minimized the root mean squared error (RMSE) between $\hat{y}$ and y, given by:

$$E_q(\hat{y}, y) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y} - y)^2} \qquad (8)$$

where n represented the size of the data.

3.6. Copula-based probabilistic model

Probabilistic model referred to a method used to determine the likelihood of a random variable taking on specific values or sets of values. Instead of focusing on estimating exact values, probabilistic model aimed to assess the probability of random events. This model used probability theory to account for uncertainties in the data.

To estimate the probability of hotspots being lower than a specific value, notated by $F_{Y|X}(Y < y|X = x)$, the formula presented in Eq. 6 could be used. Similarly, the formula for determining the probability of hotspots exceeding a specific value $F_{Y|X}(Y \geq y|X = x)$ was provided as

$$F_{Y|X}(Y \geq y|X = x) = 1 - F_{Y|X}(Y < y|X = x) = 1 - \int_0^y f_{Y|X}(y|x)\,dx \quad (9)$$

This formula was referred to as conditional survival function. The flowchart of this study was shown in Fig. 1.

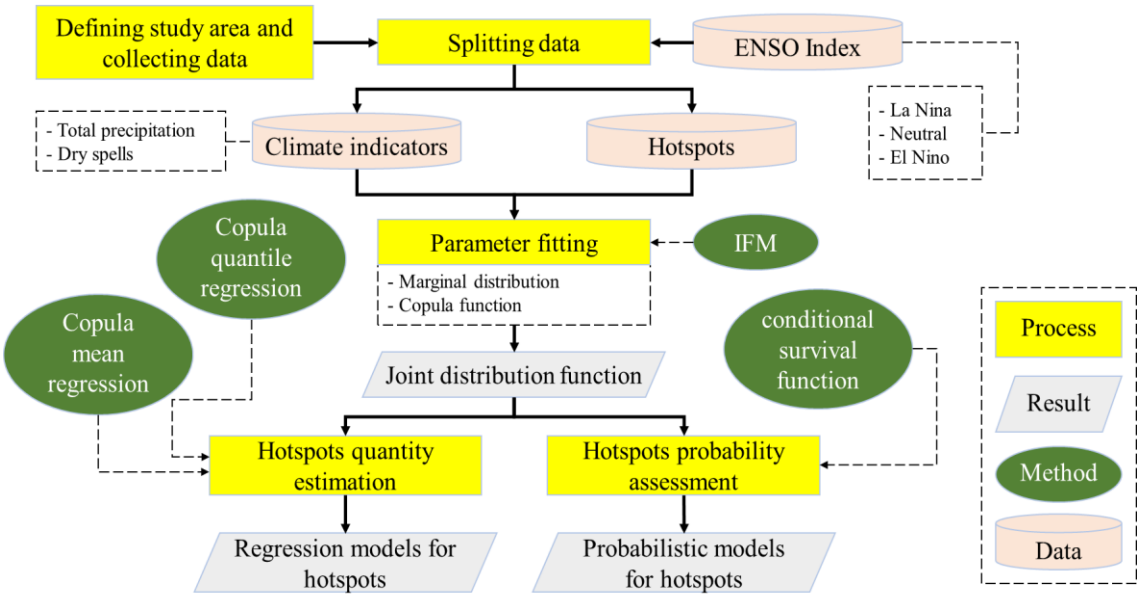

Figure 1. Flowchart of this study

4. Results and Discussion

4.1. Data description

A brief description of the data used in this study was shown in Fig. 2. A total of 100 data points were used, comprising 51 points during normal ENSO conditions, as well as 21 and 28 data points under La Nina and El Nino conditions, respectively. As shown in Fig. 2b, precipitation distribution was lower during El Nino conditions and higher during La Nina conditions. However, dry spells distribution shown in Fig. 2c and the count of hotspots in Fig. 2d, were more substantial during El Nino conditions.

Figure 2e showed the relationship between precipitation and the number of dry spells, indicating a negative correlation. This implied that minimal precipitation led to an increase in hotspots in Kalimantan. The correlation between these variables was relatively strong, and according to Pearson linear correlation coefficient, the most pronounced correlation occurred during La Nina ENSO conditions, followed by El Nino, and Normal conditions. Furthermore, alternative correlation

coefficients such as Spearman and Kendall, specifically designed to capture non-linear relationships, showed the strongest correlation during El Nino, followed by La Nina and Normal phases. Figure 2g showed a positive correlation between dry spells and the count of hotspots, suggesting that longer dry spells corresponded to a higher number of hotspots in Kalimantan. Similar to the correlation observed between precipitation and hotspots, the link between dry spells and hotspots was quite robust. Strong correlations between variables provided a solid foundation for an analysis using copula method.

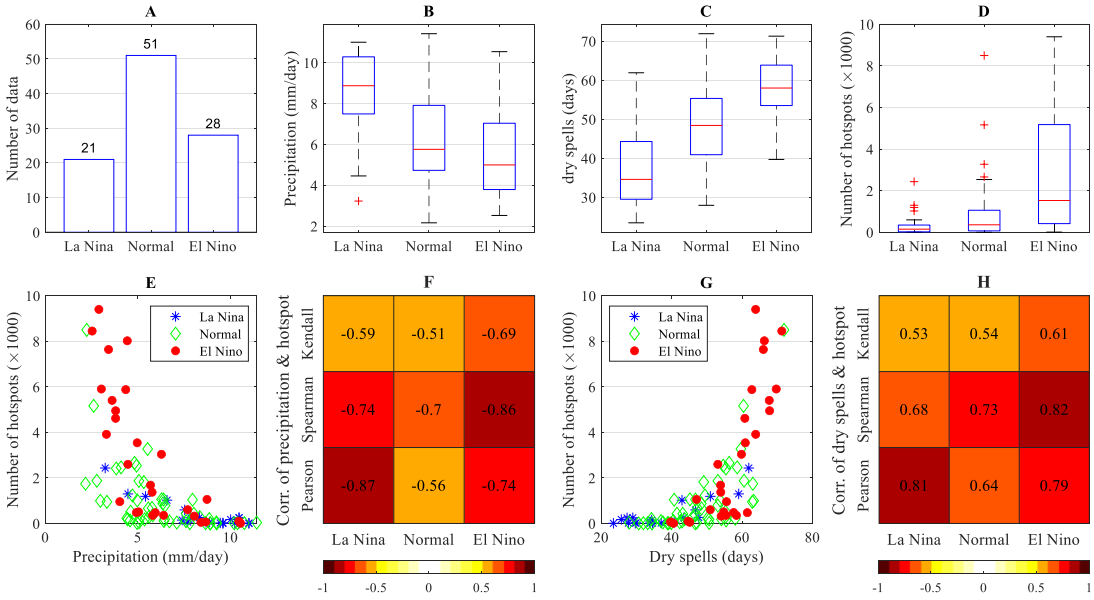

Figure 2. Description of the data: a) the number of data, b-d) distribution of precipitation, dry spells, and the number of hotspots, e-f) scatter plot and correlation between precipitation and the number of hotspots, and g-h) scatter plot and correlation between dry spells and the number of hotspots

4.2. Parameter estimation and goodness-of-fit test of marginal distributions

The parameters of the marginal distribution were estimated through the maximum likelihood method, and the outcomes were shown in Table 2. The suitability of attaching these distributions to the data namely, precipitation, dry spells, and the number of hotspots was assessed using the Anderson-Darling test.

Table 2 showed the most suitable marginal distribution for data under distinct conditions. The numbers alongside distribution names represented the Anderson-Darling statistics, while the values in parentheses indicated their corresponding p-values. With a significance level of 5%, all selected distributions passed the Anderson-Darling hypothesis test, evident from p-values exceeding 5%.

Table 2. Fitting results of marginal distributions with Anderson-Darling statistics and its p-value

Datasets	La Nina	Normal	El Nino
Precipitation	Extreme Value 0.3256 (0.9169)	Gamma 0.3361 (0.9084)	Log-normal 0.2394 (0.9757)
Dry spells	Gen. Extreme Value 0.1966 (0.9915)	Normal 0.1523 (0.9985)	Gen. Extreme Value 0.1794 (0.9951)
Hotspots	Negative Binomial 0.5286 (0.7158)	Negative Binomial 0.1987 (0.9908)	Negative Binomial 0.4793 (0.7666)

For precipitation data, the most fitting distributions varied across different ENSO conditions. Extreme value distribution was suitable during La Nina, while Gamma and Log-Normal distributions was consistent with normal and El Nino conditions, respectively. However, the generalized extreme value distribution was the best match for dry spells during La Nina and El Nino phases. Under normal ENSO conditions, the Normal distribution served as the appropriate marginal for dry spells data. Lastly, the negative binomial distribution was deemed suitable for modeling the count of hotspots.

4.3. Parameter estimation and goodness-of-fit test of copula functions

Different copula functions exhibited unique requirements regarding the correlation of their variables. For example, the Gumbel and Clayton copula demanded pairs of random variables with positive correlations, while the Frank copula could be applied to random variables displaying both positive and negative correlations. According to Fig. 2, the correlation between precipitation and hotspots was negative, but the correlation between dry spells and hotspots was positive. Most copula functions mentioned in the methods section were suited exclusively for pairs of variables with positive correlations. Consequently, implementing the rotation technique for copula function became essential to handle pairs of variables with negative dependencies as well (Tan et al., 2022; Wang et al., 2022).

Parameters of copula function were estimated using the maximum likelihood method, and the outcomes were shown in Table 3. Furthermore, the Cramer-von Mises test was used to evaluate the appropriateness of copula function for the data. The numerical values following copula names represented the Cramer-von Mises statistics, with the figures within parentheses indicating the corresponding p-values. With a significance level of 5%, all selected copula functions successfully cleared the Cramer-von Mises hypothesis test, as evidenced by p-values surpassing 5%.

Counterclockwise rotations of 90, 180, and 270 degrees were applied to copula function (Kosmidis and Karlis, 2016). Mathematical formulas for these rotations were provided by Brechmann and Schepsmeier (2013). Using a rotation of 90 or 270 degrees enabled modeling of negative dependencies, which the standard non-rotated version could not address. The results in Table 3 for precipitation and hotspots showed that the suitable copula functions during La Nina and El Nino were the Joe and BB7 copula with a 90-degree rotation, respectively. Meanwhile, the Galambos copula with a rotation of 270 degrees arose as the most appropriate choice for normal ENSO conditions.

Table 3. Fitting results of copula functions with Cramer-von Mises statistics and its p-value

Datasets	La Nina	Normal	El Nino
Precipitation - Hotspots	Joe-90° 0.04358 (0.6918)	Galambos-270° 0.03027 (0.7087)	BB7-90° 0.03334 (0.7197)
Dry spells - Hotspots	Clayton-180° 0.06694 (0.5973)	BB7 0.02393 (0.7367)	Galambos-180° 0.02546 (0.7558)

Copula function subjected to a 180-degree rotation was referred to as the survival copula function. The survival copula function exhibited characteristics contrary to those of the standard copula function (Liu et al., 2018; Nurdiati et al., 2022b). For example, the survival Clayton copula was well-suited for modeling dry spells and hotspots in La Nina. Given the lower tail dependence of the Clayton copula, its survival version showed upper tail dependence. Moreover, the most suitable copula functions under normal ENSO and El Nino conditions were the standard BB7 and survival Galambos copula, respectively.

4.4. Copula-based quantile regression

The bivariate joint distribution between the two variables was obtained based on the selected bivariate copula function under specific ENSO conditions. With detailed information about ENSO conditions and the monthly amounts of precipitation or the count of dry spells, it was possible to estimate the number of hotspots using the joint distribution. This estimation method was referred to as copula-based quantile regression.

The choice of quantile value for estimation was determined by seeking the value that yielded the best RMSE and R^2 values between the estimated and actual numbers of hotspots. This process, known as hyperparameter tuning, involved the use of quantile value (q), also called the hyperparameter or tuning parameter. Various quantile values (multiples of 0.02) were experimented with for hotspots estimation, and the results for RMSE and R^2 were shown in Fig. 3. Variations in pairs of variables and ENSO conditions led to differences in the optimal quantile value points selected. For conditional probability given a specific amount of precipitation (Fig. 3a-c), the optimal quantile values varied for different conditions: 0.32, 0.66, and 0.56 for La Nina, normal, and El Nino phases of ENSO conditions, respectively. Meanwhile, for conditional probability given a specific count of dry spells (Fig. 3d-f), the optimal quantile values were around 0.6 for each ENSO conditions, specifically 0.62, 0.58, and 0.56 during La Nina, normal, and El Nino phases, respectively.

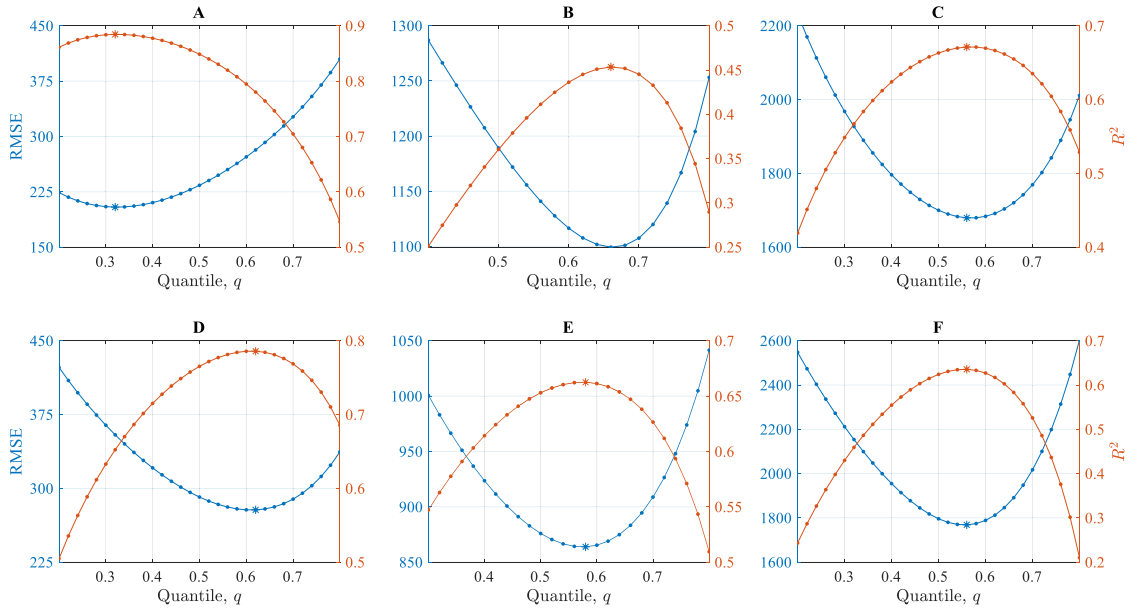

Figure 3. Hyperparameter tuning for quantile value of conditions probability given a specific value of a-c) precipitation and d-f) dry spells during La Nina, Normal, and El Nino, respectively

Apart from estimating the number of hotspots using quantile value of hotspots conditional probability, the number of hotspots was also estimated using the expected value (mean) of hotspots conditional probability (Eq. 4). Table 4 showed a comparison of RMSE and R^2 values for copula-based mean regression and quantile regression. Since none of the methods dominated the other, this study used a combination of mean and quantile regression for the estimation.

In the La Nina phase, quantile regression showed superiority over mean regression due to lower RMSE and higher R^2 values. Mean regression was observed to be more suitable during El Nino conditions, showing better fit. Under normal ENSO conditions, quantile regression was better suited for estimating hotspots given specific amounts of precipi-tation. On the other hand, mean regression was more fitting for hotspots estimation given a specific count of dry spells. In Table 4, bold numbers indicated the chosen method for hotspots estimation under each condition.

Table 4. Comparison of RMSE and R^2 for copula mean and quantile regression for each condition

Pair-Variables	Phases	Quantile Regression			Mean Regression	
		Quantile	RMSE	R^2	RMSE	R^2
Precipitation-Hotspot	La Nina	0.32	**204**	**88.44%**	223	84.87%
	Normal	0.66	**1099**	**45.35%**	1118	43.43%
	El Nino	0.56	1680	67.11%	**1653**	**68.13%**
Dry spells-Hotspot	La Nina	0.62	**278**	**78.58%**	297	75.48%
	Normal	0.58	864	66.25%	**855**	**66.92%**
	El Nino	0.56	1768	63.54%	**1732**	**65.00%**

During La Nina and El Nino, goodness-of-fit for hotspots estimation based on specific amounts of precipitation yielded superior results compared to using specific counts of dry spells, as indicated by better RMSE and R^2 values. In these conditions, hotspots estimation proved reasonably accurate, with R2 values exceeding 60%. Conversely, under normal ENSO conditions, estimating hotspots based on specific amounts of precipitation resulted in lower accuracy, with an R^2 value of less than 50%. More favorable outcomes were observed when using the count of dry spells as conditions, leading to values exceeding 60%. Therefore, opting for dry spells as conditions for hotspots probability allowed for good accuracy across all conditions.

4.5. Regression model of hotspots with lower and upper bound

One of the advantages of quantile regression was its ability to calculate quantile values used as the lower and upper bounds of a deterministic model, ensuring all data points fell within these bounds. Figure 4 showed the hyperparameter tuning process for quantile values used as lower and upper bounds. Quantile value for the lower bound was selected based on the highest quantile value involving all other greater data points. However, quantile value for the upper bound was chosen based on the smallest quantile value covering all other smaller data points.

Figure 4a showed the results of tuning the lower bound hyperparameter for model under La Nina, Normal, and El Nino conditions, using a specific amount of precipitation. The highest quantile values covering all data points during La Nina, Normal, and El Nino were 8.2%, 2.0%, and 7.6%, respectively. Any quantile value exceeding these values contained at least one data point smaller than the predicted lower bound, signifying data lying outside the lower bound. Meanwhile, Fig. 4b showed the tuning outcomes for the upper bound of model associated with a specific amount of precipitation under La Nina, Normal, and El Nino conditions. The application of quantile values for the lower and upper bounds offered greater flexibility compared to a 95% confidence interval. Even within El Nino conditions, when using a 95% confidence interval, there still existed at least one point beyond the upper bound interval, specifically at the 97.5% quantile. However, by extending the upper bound to 99.0%, all data points fell within the bound as seen in Fig. 4b. Figs. 4c and 4d also showed the results of

tuning the lower and upper bounds for model concerning a specific number of dry spells under La Nina, Normal, and El Nino conditions.

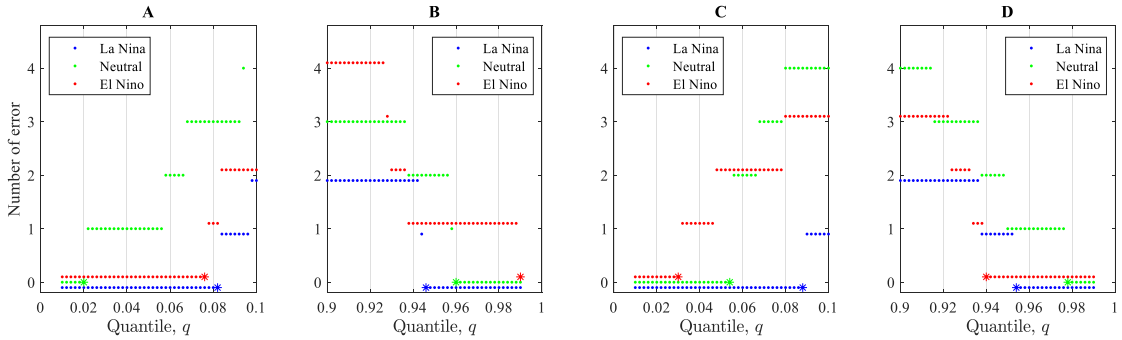

Figure 4. Hyperparameter tuning process for quantile values used as the (a) lower and (b) upper bounds of model given a specific amount of precipitation, and the (c) lower and (d) upper bounds of model given a specific number of dry spells

After the derivation of regression line for hotspots using either mean or quantile regression methods, along with the lower and upper bounds of regression line, Fig. 5 showed regression outcomes for each condition. The first row revealed regression outcomes when the specific amount of precipitation was known, yielding R^2 values of 88.44%, 45.35%, and 68.13% under La Nina, Normal, and El Nino conditions, respectively. Meanwhile, if the specific count of dry spells was known, the second row showed R^2 values of 78.58%, 66.92%, and 65.00% under La Nina, Normal, and El Nino conditions, respectively.

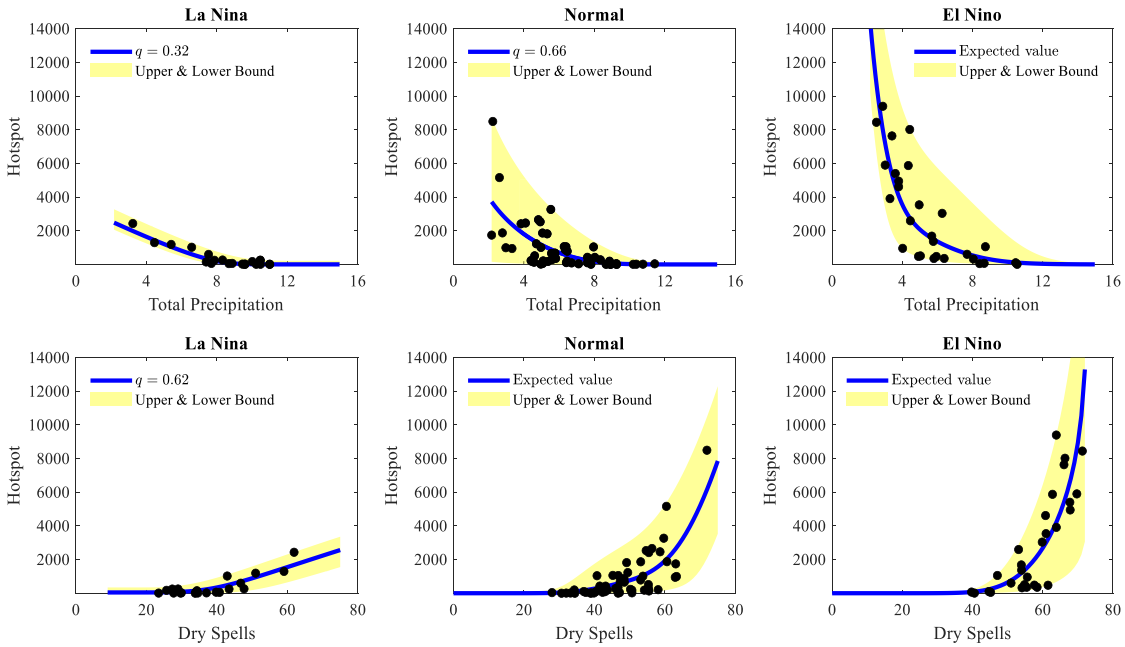

Figure 5. Copula mean and quantile regression results for hotspots with lower and upper bounds given a specific amount of precipitation (first row) and a specific number of dry spells (second row) for each ENSO conditions

Based on Fig. 5, changes in climate indicators influenced by La Nina, Normal, and El Nino conditions became evident when a specific count of dry spells was known. Despite the interval being almost the same during La Nina and normal conditions around 20 to 60 days the interval for dry spells shifted significantly to about 40 to 70 days during El Nino conditions. While an increase in hotspots was observable during El Nino, the alteration in precipitation amounts was relatively subtle, remaining within the range of 2 to 12 mm/day. These results reinforced the notion that dry spells provided better characterization of hotspots compared to precipitation amounts.

The time series of copula regression model shown in Fig. 6 was obtained by evaluating all data points within regression model in Fig. 5. Although regression outcomes did not drastically differ from those obtained through copula mean regression alone, significant enhancement was discernible in the resulting uncertainty range. The use of quantile regression in the method improved the previously modeled lower and upper bounds, which were based on copula mean regression (Najib et al., 2022b). In the earlier model, the 95% confidence interval failed to include certain points. By modifying the lower and upper bounds using suitable quantile values of conditional probability, model developed in this study (Fig. 6) effectively covered all points along the boundary of regression model. In regression model for hotspots, given a specific amount of precipitation (referred to as precipitation-based model), the R^2 and RMSE values were 69.55% and 1179, respectively. Meanwhile, in regression model for hotspots, with a specific number of dry spells known as dry spells-based model, the R^2 and RMSE were 73.07% and 1109, respectively.

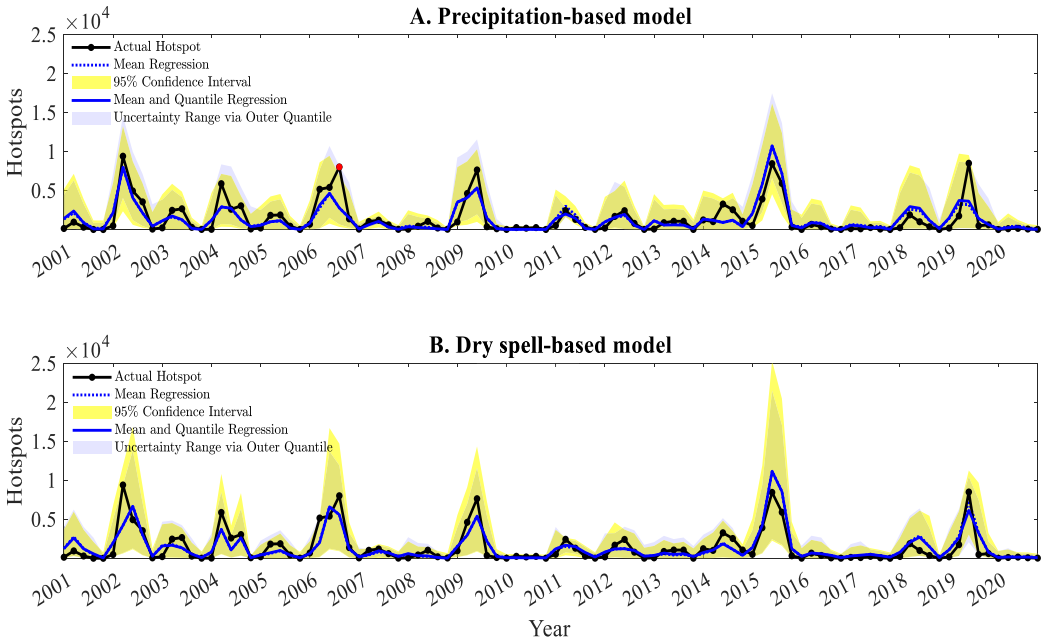

Figure 6. Time series of copula mean and quantile regression model for hotspots given the specific amount of precipitation (top) and the specific number of dry spells (bottom) for each ENSO conditions

4.6. Probabilistic model of hotspots

This section discussed an alternative representation model achievable through copula-based conditional probability. Instead of extracting regression line from the derived conditional probability, probabilistic model

could be fashioned by estimating the probability of hotspots occurrences being below or above a specific value (Equations 6 and 9). Figure 7 showed a visualization of hotspots probabilities for values below a specific threshold, for each ENSO conditions, given the knowledge of either the amount of precipitation (top row) or the count of dry spells (bottom row).

For example, if it was established that the average amount of precipitation over 2 months was 4 mm/day, the probability of hotspots occurrence being below 1000 under La Nina conditions was 3.1% as indicated by the blue dot in Fig. 7. This suggested that the probability of hotspots occurrence being above 1000 was 96.9%. This implied that even during La Nina, if precipitation was low (less than 4 mm), the probability of hotspots occurrence exceeding 1000 was significantly high. However, the probability of hotspots surpassing 2500 was very low, at only 3.6%, according to the black dot. Consequently, the peak probability for hotspots, assuming an average precipitation of 4 mm/day over 2 months during La Nina, lay between 1000 and 2500 hotspots. Even when precipitation was less than 4 mm, the probability of hotspots count exceeding 5000 was practically non-existent during La Nina conditions, as depicted by the absence of contour levels for 5000 hotspots at the top left corner in Fig. 7.

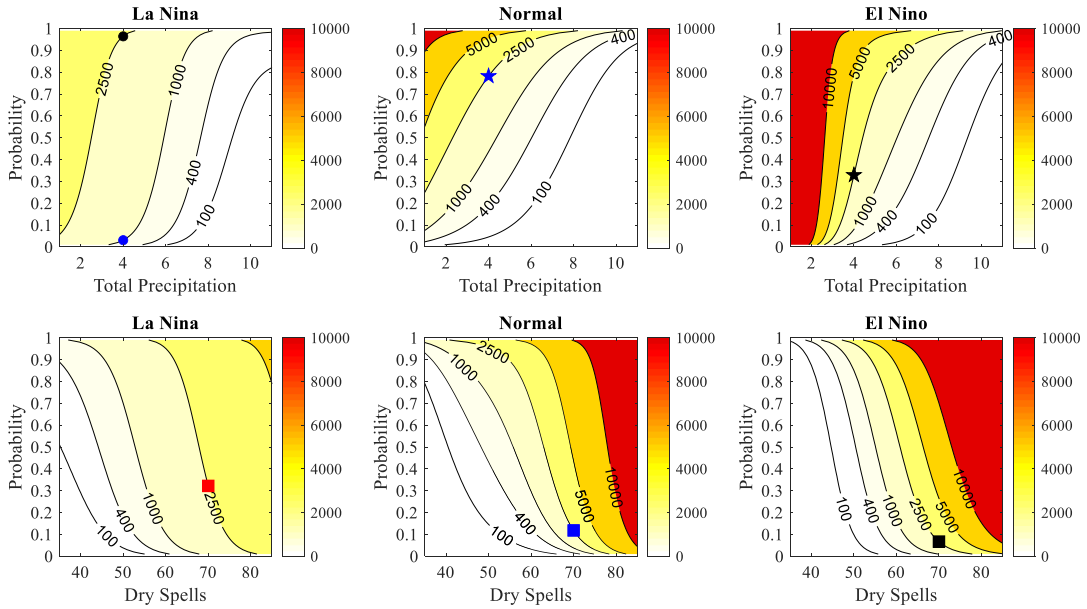

Figure 7. Probabilistic model for hotspots given a specific amount of precipitation (first row) and a specific number of dry spells (second row) for each ENSO conditions

Regarding Normal and El Nino conditions, when the average precipitation for 2 months was 4 mm/day, the probability of hotspots occurrence exceeding 2500 rose to 21.8% and 67.1%. This was denoted by the blue and black stars, respectively, demonstrating an increase compared to La Nina conditions. In a situation where the average precipitation is less than 4 mm/day over 2 months, the probability of hotspots occurrence falling below 2500 hotspots per month decreased, or conversely, the likelihood of hotspots count surpassing 2500 increased. When compared to La Nina conditions, where contour levels for probabilities exceeding 5000 hotspots were absent, these contours were present under ENSO Normal and El Nino conditions. Moreover, even contours for probabilities

surpassing 10,000 hotspots existed in the red area under Normal and El Nino conditions.

During ENSO Normal conditions, the probability of hotspots occurrence exceeding 10,000 was recorded when the average precipitation over 2 months was 3 mm/day. Meanwhile, the probability of more than 10,000 hotspots materializing under El Nino conditions became significant when the average precipitation fell below 2 mm/day over 2 months. This illustrated that El Nino conditions could trigger more severe hotspots despite less severe average precipitation compared to Normal or La Nina conditions. Additionally, the peak probability of hotspots occurrence exceeding 10000 under ENSO Normal conditions was 10.6%, occurring when the average precipitation over 2 months was 1 mm/day. Under El Nino conditions, the probability of hotspots occurrence surpassing 10000 reached 100% when the average precipitation was below 2 mm/day over 2 months.

A similar interpretation could be applied to probabilistic model using the number of dry spells (dry spells-based model). If it was known that the average number of dry spells within 3 months was 70 days, the probability of hotspots occurrences exceeding 2500 hotspots was 67.8%, 88.2%, and 93.3% for La Nina, Normal, and El Nino conditions, indicated by red, blue, and black squares in Fig. 7, respectively. Based on these probabilities, there was a substantial likelihood of hotspots number surpassing 2500 for each ENSO conditions when the average amount of dry spells was 70 days or more over 3 months. As opposed to precipitation-based model, dry spells-based model allowed for the occurrence of hotspots exceeding 5000 even in La Nina conditions, albeit with a low probability. This likelihood started evolving when the average number of dry spells exceeded 78 days over 3 months. However, the probability of hotspots exceeding 10,000 still did not appear in La Nina conditions.

Under ENSO Normal and El Nino conditions, the probability of hotspots occurrence exceeding 10,000 hotspots arose when the average number of dry spells was 70 and 59 days over 3 months, respectively. As previously observed, El Nino conditions triggered more severe hotspots despite the average number of dry spells being less intense than the Normal or La Nina conditions. Fortunately, this model managed to capture more extreme hotspots events under ENSO Normal conditions than precipitation-based model did. This was evident from the highest probability of hotspots occurrence exceeding 10,000 hotspots, reaching 95.5% when dry spells spanned 85 days over 3 months. Additionally, during El Nino conditions, the peak probability of hotspots occurrence surpassed 10,000 hotspots, reaching 94.1%.

In dry spells-based model, the propagation of the probability of hotspots occurrence exceeding 10,000 appeared to be more gradual compared to model based on given precipitation under El Nino conditions. Even slight changes in precipitation led to significant shifts in hotspots probability exceeding 10,000. On the other hand, probabilistic model based on the average number of dry spells exhibited more resilience and insensitivity.

The time series of probabilistic model in Fig. 8 showed the probability of having more than 1,000, 5,000, and 10,000 hotspots in Kalimantan. This probability was derived by evaluating all data points within model shown in Fig. 7. This study was classified into three scenarios, namely the classification of hotspots in Kalimantan as exceeding 1,000, 5,000, and 10,000 hotspots or not. Precipitation-based model yielded the highest probability of hotspots occurrence exceeding 10,000 hotspots in September 2015, at

63.41%. The actual number of hotspots in that month was 8,448, hence, the obtained probability was excessively high, surpassing the 50% threshold. Precipitation-based model achieved an unrealistically low probability for September 2019, a month with 8,497 hotspots. The probability of hotspots occurrence exceeding 5,000 hotspots in that month was a mere 19.51%.

It should be noted that dry spells-based model provided improved results compared to others. The probability of hotspots occurrence exceeding 10,000 hotspots in September 2015 was 38.31%. Although this value was relatively high, it remained below 50%. The probability of hotspots occurrence exceeding 5,000 hotspots in September 2019 reached 72.10%, surpassing the 50% threshold. Additionally, dry spells-based model better captured the pattern observed in 2004, which was characterized by two peaks of hotspots. The results from dry spells-based more accurately reflected the mentioned hotspots occurrences compared to precipitation-based model.

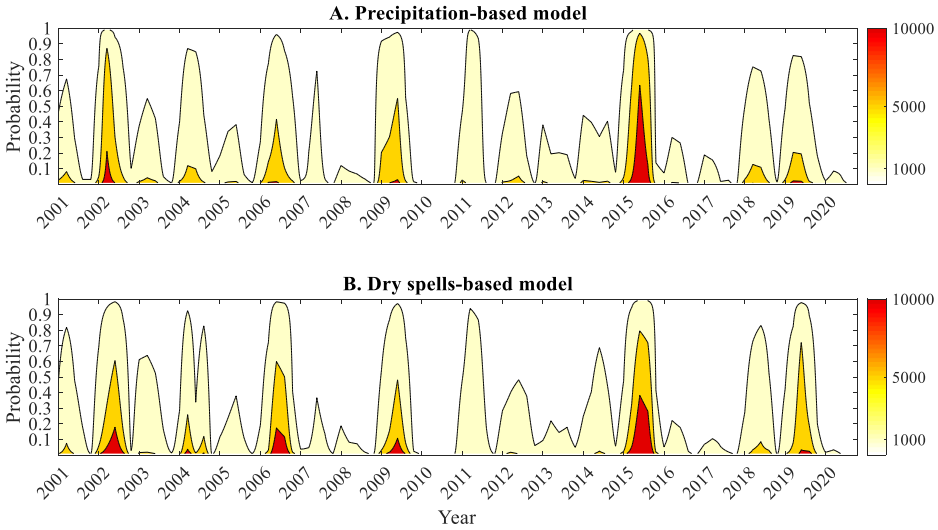

Figure 8. Time series of probabilistic model for hotspots given the specific amount of precipitation (top) and the specific number of dry spells (bottom) for each ENSO conditions

Table 5 showed the accuracy of probabilistic model based on copula function across the three classification scenarios. Accuracy was computed by binary assessment of the generated probability values within probabilistic model. When the probability value exceeded 50%, model predicted that hotspots would exceed a specific threshold value (in this case, 1,000, 5,000, and 10,000 hotspots). Accuracy was determined by summing the true positives and true negatives and dividing by the total number of data. While both models exhibited the same accuracy at the 1,000 hotspots threshold, dry spells-based model outperformed precipitation-based in the other two cases. On average, precipitation-based model achieved an accuracy of 89.7% across the three classification scenarios, and dry spells-based reached an accuracy of 90.3%.

Table 5. Accuracy of probabilistic model in the three classification cases

Class	Precipitation-based model	Dry spells-based model
More than 1000 or not	76.0%	76.0%
More than 5000 or not	94.0%	95.0%
More than 10000 or not	99.0%	100.0%
Average	**89.7%**	**90.3%**

The pronounced misclassification originating from precipitation-based model

was very pronounced even though the entire accuracy percentages indicated close values between precipitation-based and dry spells-based model. This was exemplified by the misclassification observed in the estimated hotspots probabilities for 2015 and 2019, as previously explained. It was concluded that dry spells-based model outperformed precipitation-based model, even though some minor details were better captured by precipitation-based model than dry spells-based.

4.7. Discussions

This article explored a novel method to modeling hotspots in Kalimantan through the integration of copula regression and copula quantile regression. As proposed by Hoang (2018) in his dissertation, copula regression (both quantile and mean regression) showed greater robustness compared to non-parametric regression trees and traditional linear regression methods. Koenker (2005) also elucidated that quantile regression offered enhanced resilience when handling outliers in the response variable. However, this study found that quantile and mean regression yielded nearly identical outcomes as seen in Table 4. Quantile regression proved more suitable for estimating hotspots numbers in wet conditions (La Nina), whereas mean regression was more appropriate for dry conditions (El Nino).

One significant advantage of copula regression lay in its capacity to establish an uncertainty range (Zhang et al., 2023). Accessing the outermost quantile of conditional distribution function generated through copula function, granted flexibility in determining the lower and upper bounds of the uncertainty interval for the resulting hotspots estimation (Fig. 4). This interval served as a means to evaluate the reliability of hotspots estimates derived from copula quantile regression.

Beyond the use of quantile regression, this study introduced an alternative method, i.e., probabilistic model of hotspots through copula-based joint distributions. A similar study endeavor proposing probabilistic strategy was presented by Nikonovas et al. (2022), which used a multilayer perceptron to formulate an artificial neural network model for estimating hotspots probabilities surpassing a certain threshold. As opposed to Nikonovas et al. (2022), who estimated hotspots probabilities for each grid cell across Indonesia, including Kalimantan, this study provided estimations of hotspots probabilities in a more general sense. The current study specifically calculates the likelihood of the total number of hotspots in Kalimantan as a whole for a given month. This method eliminated the need to consider various location-specific factors, such as land management practices, policy decisions, and fire suppression efforts (Page and Hooijer, 2016; Tacconi, 2016). However, it was entirely feasible to conduct further investigation using copula function to predict hotspots probabilities for individual grid cells.

There remained scenarios where precipitation could complement the information provided solely by dry spells as predictors, although the results showed that precipitation-based model did not outperform dry spells-based model. This implied that the potential application of multidimensional copula functions (Liu and Li, 2022; Wu et al., 2022; Zhao et al., 2022) was anticipated to be developed for enhancing this probabilistic prediction model in the future. Incorporating the temporal variability of both global and local climatic conditions could also be more effectively modeled using a time-varying copula (Lee and Kim, 2021; Maposa et al., 2021; Xu et al., 2021). Consequently, the prospects for refining and expanding upon this model remained intriguing avenues for further exploration.

5. Conclusions

In conclusion, this article explored hotspots modeling in Kalimantan based on ENSO conditions using two types of predictors, including the amount of precipitation and the number of dry spells. Model construction used copula function, with two methods being used, namely copula regression and probabilistic modeling.

The constructed copula model included regression created through copula-based mean and quantile regression, yielding hotspots predictions. The main advantage of copula-based quantile regression model was that the lower and upper bounds of regression model could cover all the data. This inclusivity arose from the use of quantile, which was consistent with the data obtained. However, the significant contribution of this article lay in probabilistic model formed based on copula-based conditions distribution function. Through this method, the estimation of extreme hotspots event probabilities in Kalimantan became feasible, diverging from obtaining a definite predictive value from regression model. Among the three scenarios examined, precipitation-based model attained an average accuracy of 89.7%, while dry spells-based was 90.3%. Despite the proximity of total accuracy percentages between precipitation-based and dry spells-based model, the pronounced misclassification stemming from precipitation-based method was distinguished. Based on the outcomes derived from regression and probabilistic model, it can be concluded that dry spells-based model surpassed precipitation-based model. On the other hand, certain minor details were more effectively captured by precipitation-based model than by dry spells-based model.

References

Abdallah M., Mohammadi B., Modathir M.A., Omer A., Cheraghalizadeh M., Eldow M.E.E., Duan Z., 2022. Reference evapotranspiration estimation in hyper-arid regions via D-vine copula based-quantile regression and comparison with empirical approaches and machine learning models. Journal of Hydrology: Regional Studies, 44. Doi: 10.1016/j.ejrh.2022.101259.

Amini S., Bidaki R.Z., Mirabbasi R., Shafaei M., 2022. Flood risk analysis based on nested copula structure in Armand Basin, Iran. Acta Geophysica, 70(3), 1385–1399. Doi: 10.1007/s11600-022-00766-y.

Amirudin A.A., Salimun E., Tangang F., Juneng L., Zuhairi M., 2020. Differential influences of teleconnections from the Indian and Pacific oceans on rainfall variability in Southeast Asia. Atmosphere, 11(9), 886. Doi: 10.3390/ATMOS11090886.

Anderson T.W., 2011. Anderson-Darling Tests of Goodness-of-Fit. In: International Encyclopedia of Statistical Science. Berlin, Heidelberg: Springer, 52–54. Doi: 10.1007/978-3-642-04898-2_118.

Ardiyani E., Nurdiati S., Sopaheluwakan A., Septiawan P., Najib M.K., 2023. Probabilistic Hotspot Prediction Model Based on Bayesian Inference Using Precipitation, Relative Dry Spells, ENSO and IOD. Atmosphere, 14(2), 286.

Babić S., Ley C., Veredas D., 2019. Comparison and classification of flexible distributions for multivariate skew and heavy-tailed data. Symmetry, 11(10). Doi: 10.3390/sym11101216.

Berg D., 2009. Copula goodness-of-fit testing: An overview and power comparison. European Journal of Finance, 15(7-8), 675–701. Doi: 10.1080/13518470802697428.

Brechmann E.C., 2014. Hierarchical Kendall copulas: Properties and inference. Canadian Journal of Statistics, 42(1), 78-108. Doi: 10.1002/cjs.11204.

Brechmann E.C., Schepsmeier U., 2013. Modeling dependence with C- and D-vine copulas: The R package CDVine. Journal of Statistical Software, 52(3), 1–27. Doi: 10.18637/jss.v052.i03.

Charizanos G., Demirhan H., 2023. Bayesian prediction of wildfire event probability using normalized difference vegetation index data from an Australian forest. Ecological Informatics, 73. Doi: 10.1016/j.ecoinf.2022.101899.

Coutts A.M., Harris R.J., Phan T., Livesley S.J., Williams N.S.G., Tapper N.J., 2016. Thermal

infrared remote sensing of urban heat: Hotspots, vegetation, and an assessment of techniques for use in urban planning. Remote Sensing of Environment, 186, 637–651. Doi: 10.1016/j.rse.2016.09.007.

Danaher P.J., Smith M.S., 2011. Modeling multivariate distributions using copulas: Applications in marketing. Marketing Science, 30(1), 4–21. Doi: 10.1287/mksc.1090.0491.

El Adlouni S., 2018. Quantile regression C-vine copula model for spatial extremes. Natural Hazards, 94(1), 299–317. Doi: 10.1007/s11069-018-3389-6.

Enríquez-de-Salamanca Á., 2020. Contribution to climate change of forest fires in Spain: Emissions and loss of sequestration. Journal of Sustainable Forestry, 39(4), 417–431.
Doi: 10.1080/10549811.2019.1673779.

Fanin T., van der Werf G., 2017. Precipitation-fire linkages in Indonesia (1997-2015). Biogeosciences, 14(18), 3995–4008. Doi: 10.5194/bg-14-3995-2017.

Goldstein J.E., 2020. The Volumetric Political Forest: Territory, Satellite Fire Mapping, and Indonesia's Burning Peatland. Antipode, 52(4), 1060–1082. Doi: 10.1111/anti.12576.

Grari M., Idrissi I., Boukabous M., Moussaoui O., Azizi M., Moussaoui M., 2022. Early wildfire detection using machine learning model deployed in the fog/edge layers of IoT. Indonesian Journal of Electrical Engineering and Computer Science, 27(2), 1062–1073. Doi: 10.11591/ijeecs.v27.i2.pp1062-1073.

Harrison M.E., Page S.E., Limin S.H., 2009. The global impact of Indonesian forest fires. Biologist, 56(3), 156–163.

Herawati N., 2020. The Effectiveness of Quantile Regression in Dealing with Potential Outliers. Barekeng: Jurnal Ilmu Matematika dan Terapan, 14(2), 301–308.

Hoang Q., 2018. Copula Regression and Robustness. Dissertation, Texas Tech University.

Hoeffding W., 1940. Massstabinvariante korrelationstheorie. Schriften des Mathematischen Seminars und des Instituts fu¨r Angewandte Mathematik der Universita¨t Berlin, 5, 179–233.

Horton A.J., Lehtinen J., Kummu M., 2022. Targeted land management strategies could halve peatland fire occurrences in Central Kalimantan, Indonesia. Communications Earth and Environment, 3(1).
Doi: 10.1038/s43247-022-00534-2.

Huijnen V., et al., 2016. Fire carbon emissions over maritime southeast Asia in 2015 largest since 1997. Scientific Reports 6. Doi: 10.1038/srep26886.

Iskandar I., Lestrai D.O., Nur M., 2019. Impact of El Niño and El Niño Modoki Events on Indonesian Rainfall. Makara Journal of Science, 23(4), 217–222. Doi: 10.7454/mss.v23i4.11517.

Jim C.Y., 1999. The forest fires in Indonesia 1997-1998: Possible causes and pervasive consequences. Geography, 84(3), 251–260.

Joe H., 1997. Multivariate models and multivariate dependence concepts. London: CRC Press.

Joe H., 2005. Asymptotic efficiency of the two-stage estimation method for copula-based models. Journal of Multivariate Analysis, 94(2), 401–419. Doi: 10.1016/j.jmva.2004.06.003.

Joe H., 2014. Dependence modeling with copulas. CRC Press. Doi: 10.1201/b17116.

Khan M.F., et al. 2020. El Niño driven haze over the Southern Malaysian Peninsula and Borneo. Science of the Total Environment, 730. Doi: 10.1016/j.scitotenv.2020.139091.

Koenker R., 2005. Quantile Regression. Cambridge: Cambridge University Press.

Koenker R., Hallock K.F., 2001. Quantile regression. Journal of Economic Perspectives, 15(4), 143–156. Doi: 10.1257/jep.15.4.143.

Koh J., Pimont F., Dupuy J.-L., Opitz T., 2023. Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. The Annals of Applied Statistics, 17(1). Doi: 10.1214/22-aoas1642.

Kosmidis I., Karlis D., 2016. Model-based clustering using copulas with applications. Statistics and Computing, 26(5), 1079-1099. Doi: 10.1007/s11222-015-9590-5.

Lee N., Kim J.M., 2021. Dynamic functional connectivity analysis based on time-varying partial correlation with a copula-DCC-GARCH model. Neuroscience Research, 169, 27–39. Doi: 10.1016/j.neures.2020.06.006.

Li H., Huang G., Li Y., Sun J., Gao P., 2021. A c-vine copula-based quantile regression method for

streamflow forecasting in xiangxi river basin, China. Sustainability (Switzerland), 13(9). Doi: 10.3390/su13094627.

Li Z., Beirlant J., Yang L., 2022. A new class of copula regression models for modelling multivariate heavy-tailed data. Insurance: Mathematics and Economics, 104, 243–261. Doi: 10.1016/j.insmatheco.2022.02.002.

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. Doi: 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. Doi: 10.1155/2018/5486185.

Liu S., Li S., 2022. Multi-model D-vine copula regression model with vine copula-based dependence description. Computers and Chemical Engineering, 161. Doi: 10.1016/j.compchemeng.2022.107788.

Lounela A.K., 2021. Shifting Valuations of Sociality and the Riverine Environment in Central Kalimantan, Indonesia. Anthropological Forum, 31(1), 34–48. Doi: 10.1080/00664677.2021.1875197.

lvadori G., De Michele C., Kottegoda N.T., Rosso R., 2005. Extremes in Nature: An approach using Copulas. Linz, Austria: Springer Science & Business Media.

Ly S., Pho K.H., Ly S., Wong W.K., 2019. Determining distribution for the product of random variables by using copulas. Risks, 7(1), 23. Doi: 10.3390/risks7010023.

Maposa D., Seimela A.M., Sigauke C., Cochran J.J., 2021. Modelling temperature extremes in the Limpopo province: bivariate time-varying threshold excess approach. Natural Hazards, 107(3), 2227–2246. Doi: 10.1007/s11069-021-04608-w.

Marlier M.E., et al., 2019. Fires, Smoke Exposure, and Public Health: An Integrative Framework to Maximize Health Benefits From Peatland Restoration. GeoHealth, 3(7), 178–189. Doi: 10.1029/2019GH000191.

Masseran N., Hussain S.I., 2020. Copula modelling on the dynamic dependence structure of multiple air pollutant variables. Mathematics, 8(11), 1–16. Doi: 10.3390/math8111910.

Medrilzam M., Dargusch P., Herbohn J., Smith C., 2014. The socio-ecological drivers of forest degradation in part of the tropical peatlands of Central Kalimantan, Indonesia. Forestry, 87(2), 335–345. Doi: 10.1093/forestry/cpt033.

Mezbahuddin S., Nikonovas T., Spessa A., Grant R.F., Imron M.A., Doerr S.H., Clay G.D., 2023. Accuracy of tropical peat and non-peat fire forecasts enhanced by simulating hydrology. Scientific Reports, 13(1). Doi: 10.1038/s41598-022-27075-0.

Mirabbasi R., Fakheri-Fard A., Dinpashoh Y., 2012. Bivariate drought frequency analysis using the copula method. Theoretical and Applied Climatology, 108(1-2), 191–206. Doi: 10.1007/s00704-011-0524-7.

Mukhopadhyay S., Parzen E., 2020. Nonparametric universal copula modeling. Applied Stochastic Models in Business and Industry, 36(1), 77-94. Doi: 10.1002/asmb.2503.

Nainggolan H.A., Veanti D.P.O., Akbar D., 2020. Utilisation of Nasa - Gfwed and firms satellite data in determining the probability of hotspots using the Fire Weather Index (Fwi) in Ogan Komering Ilir Regency, South Sumatra. International Journal of Remote Sensing and Earth Sciences (IJReSES), 17(1), 85. Doi: 10.30536/j.ijreses.2020.v17.a3202.

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(1). Doi: 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. Doi: 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. Doi: 10.1007/s11069-022-05346-3.

Nelsen R.B., 2006. An Introduction to Copulas. 2nd ed. New York: Springer Science & Business Media.

Nicholls N., 1984. The Southern Oscillation and Indonesian sea surface temperature. Monthly Weather Review, 112(3), 424–432. Doi: 10.1175/1520-0493(1984)112<0424:TSOAIS>2.0.CO;2.

Nikonovas T., Spessa A., Doerr S.H., Clay G.D., Mezbahuddin S., 2022. ProbFire: A probabilistic fire early warning system for Indonesia. Natural Hazards and Earth System Sciences, 22(2), 303–322. Doi: 10.5194/nhess-22-303-2022.

Nurdiati S., et al., 2022a. The impact of El Niño southern oscillation and Indian Ocean Dipole on the burned area in Indonesia. Terrestrial, Atmospheric and Oceanic Sciences, 33(15). Doi: 10.1007/S44195-022-00016-0.

Nurdiati S., Najib M.K., Thalib A.S., 2022b. 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. Doi: 10.21163/GT_2022.172.10.

Nurdiati S., Sopaheluwakan A., Septiawan P., 2021. Spatial and temporal analysis of El Niño impact on land and forest fire in Kalimantan and Sumatra. Agromet, 35(1), 1–10. Doi: 10.29244/j.agromet.35.1.1-10.

Nurdiati S., Sopaheluwakan A., Julianto M.T., Septiawan P., Rohimahastuti F., 2022c. Modelling and analysis impact of El Nino and IOD to land and forest fire using polynomial and generalized logistic function: cases study in South Sumatra and Kalimantan, Indonesia. Modeling Earth Systems and Environment, 8(3), 3341–3356. Doi: 10.1007/s40808-021-01303-4.

Page S.E., Hooijer A., 2016. In the line of fire: The peatlands of Southeast Asia. Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1696). Doi: 10.1098/rstb.2015.0176.

Page S.E., Rieley J.O., Banks C.J., 2011. Global and regional importance of the tropical peatland carbon pool. Global Change Biology, 17(2), 798–818. Doi: 10.1111/j.1365-2486.2010.02279.x.

Pambabay-Calero J., Bauz-Olvera S., Nieto-Librero A., Sánchez-García A., Galindo-Villardón P., 2021. Hierarchical modeling for diagnostic test accuracy using multivariate probability distribution functions. Mathematics, 9(11). Doi: 10.3390/math9111310.

Pan S., Joe H., 2022. Predicting times to event based on vine copula models. Computational Statistics and Data Analysis, 175, 107546. Doi: 10.1016/j.csda.2022.107546.

Philander S.G.H., 1983. El Nino southern oscillation phenomena. Nature, 302(5906), 295–301.

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

Salafsky N., 1994. Drought in the rain forest: Effects of the 1991 El Niño-Southern Oscillation event on a rural economy in West Kalimantan, Indonesia. Climatic Change, 27(4), 373–396. Doi: 10.1007/BF01096268.

Salvadori G., De Michele C., 2007. On the Use of Copulas in Hydrology: Theory and Practice. Journal of Hydrologic Engineering, 12(4), 369–380. Doi: 10.1061/(asce)1084-0699(2007)12:4(369).

Sartika Q.R., Widiharih T., Mukid M.A., 2019. Value at Risk in Stock Portfolio using t-Copula: Case study of PT. Indofood Sukses Makmur, Tbk. and Bank Mandiri (Persero), Tbk. Media Statistika, 12(2), 175. Doi: 10.14710/medstat.12.2.175-187.

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. Doi: 10.5194/npg-15-761-2008.

Shao Y., Feng Z., Sun L., Yang X., Li Y., Xu B., Chen Y., 2022. Mapping China's Forest Fire Risks with Machine Learning. Forests, 13(6). Doi: 10.3390/f13060856.

Sklar A., 1959. Fonctions de Répartition à n Dimensions et Leurs Marges. Publications de L'Institut de Statistique de L'Université de Paris, 8, 229–231.

Tacconi L., 2016. Preventing fires and haze in Southeast Asia. Nature Climate Change, 6(640–643). Doi: 10.1038/nclimate3008.

Tahroudi M.N., Ramezani Y., de Michele C., Mirabbasi R., 2022. Multivariate analysis of rainfall and its deficiency signatures using vine copulas. International Journal of Climatology, 42(4), 2005-2018. Doi: 10.1002/joc.7349.

Tahroudi M.N., Ramezani Y., De Michele C., Mirabbasi R., 2020. Analyzing the conditional behavior of rainfall deficiency and groundwater level deficiency

signatures by using copula functions. Hydrology Research, 51(6), 1332–1348. Doi: 10.2166/nh.2020.036.

Tan S.R., Li C., Yeap X.W., 2022. A time-varying copula approach for constructing a daily financial systemic stress index. North American Journal of Economics and Finance, 63, 101821. Doi: 10.1016/j.najef.2022.101821.

Thevaraja M., Rahman M., 2019. Regression analysis based on Copula theory - by using Gaussian family Copula. International Journal of Statistics and Reliability Engineering, 6(1), 24–28.

Thoha A.S., Istima N., Daulay I.A., Hulu D.L.N., Budi S., Ulfa M., Mardiyadi Z., 2023. Spatial distribution of 2019 forest and land fires in Indonesia. Journal of Physics: Conference Series, 2421(1), 012035. Doi: 10.1088/1742-6596/2421/1/012035.

Tootoonchi F., Sadegh M., Haerter J.O., Räty O., Grabs T., Teutschbein C., 2022. Copulas for hydroclimatic analysis: A practice-oriented overview. Wiley Interdisciplinary Reviews: Water, 9(2). Doi: 10.1002/wat2.1579.

Treppiedi D., Cipolla G., Francipane A., Noto L.V., 2021. Detecting precipitation trend using a multiscale approach based on quantile regression over a Mediterranean area. International Journal of Climatology, 41(13), 5938–5955. Doi: 10.1002/joc.7161.

Vatresia A., Rais R.R., Utama F.P., Oktarianti W., 2022. Mining fire hotspots over Nusa Tenggara and Bali Islands. Indonesian Journal of Forestry Research, 9(1), 73–85. Doi: 10.20886/ijfr.2022.9.1.73-85.