

Multivariate Fire Risk Models Using Sample-Size Reduction and Copula Regression in Kalimantan, Indonesia

Mohamad Khoirun Najib^{1*} · Sri Nurdiati¹ · Ardhasena Sopaheluwakan²

✉ Mohamad Khoirun Najib (corresponding author)
mkhoirun_najib@apps.ipb.ac.id
ORCID: 0000-0002-4372-4661

¹ Department of Mathematics, IPB University, Bogor 16680, Indonesia

² Center for Applied Climate Services, Agency for Meteorology, Climatology, and Geophysics, Jakarta 10720, Indonesia

Abstract

Forest fires have become a national issue every year and get serious attention from the government and researchers, especially in Kalimantan. The copula-based joint distribution can construct a fire risk model to improve the early warning system of forest fires. This study aims to model and analyze the copula-based joint distribution between climate conditions and hotspots in Kalimantan. We constructed the bivariate joint distributions between climate conditions, either total precipitation or dry spells, and hotspots with sample size reduced by ENSO conditions, i.e., La Nina, normal, and El Nino. From the joint distribution, fire risk models are calculated using conditional probability and copula regression. The results show that the relationship between climate conditions and hotspots in La Nina and normal ENSO conditions have an upper tail dependence but no lower tail dependence. Meanwhile, the relationship has both upper and lower tail dependences during El Nino. There is an outlier in normal ENSO conditions with more hotspots than normally, i.e., in September 2019. The probability is very low during normal ENSO conditions, i.e., less than 2%. The only relatively high probability is during El Nino, i.e., more than 10%. Moreover, the copula regression models show that the model given specific dry spells is better than that given specific total precipitation as climate condition. The copula regression for hotspots given specific total precipitation and ENSO conditions has the RMSE value of 1339 hotspots and the R^2 value of 60.70%. Meanwhile, the copula regression for hotspots given specific dry spells and ENSO conditions has the RMSE value of 1185 hotspots and the R^2 value of 69.21%.

Keywords Copula · Dry spells · ENSO · Hotspot · Precipitation · Wildfire

1 Introduction

Indonesia has a vast forest area. The Ministry of Environment and Forestry of Indonesia (KLHK) said that the forest area reached 94.1 million hectares or about 50.1% of the total land area in 2019 (KLHK 2020). However, forest areas in Indonesia have experienced severe degradation in both quantity and quality in the last few decades (Miettinen et al. 2011). In 2000-2009, Indonesia lost 1.4 million ha of natural forest/year. In the following period (2009-2013), the area lost natural forest decreased to 1.1 million ha/year but increased again in 2013-2017 to 1.4 million ha/year (FWI 2020). One of the causes of deforestation is forest fires activity (Austin et al. 2019).

Forest fires have become a national issue every year and get serious attention from the government and researchers, especially in Kalimantan (Aflahah et al. 2019). Generally, forest fires are caused by human and natural factors. Natural factors such as El Nino can reduce rainfall intensity and prolong the dry season so that plants become more dehydrated and more prone to fire (Field et al. 2016; Nikonovas et al. 2020). Severe forest fires in 1997 and 2015 coincided with an extreme El Nino episode (Fanin and van der Werf 2016; Nikonovas et al. 2020).

Many experts have conducted research related to forest fires with various underlying problems, such as deforestation, declining air quality (Ryan et al. 2021), death of flora and fauna (Krebs et al. 2019), to health problems from affected residents (Silveira et al. 2021). Their research uses various climate indicators such as temperature (Xu et al. 2021), soil conditions (Sachdeva et al. 2018), precipitation (Brogan et al. 2019), wind (Daşdemir et al. 2021), El Nino-Southern Oscillation (Cardil et al. 2021), and others. The methods used also vary, ranging from regression (Zhao and Xu 2021), classification (Sulova and Arsanjani 2021), clustering (Viana-Soto et al. 2020), neural networks (Larsen et al. 2021), to copula (Xi et al. 2020).

A fire risk model is essential to improve the early warning system of forest and land fires in Indonesia. A fire risk model is a conditional probability of a fire size given specific climate indicators (Madadgar et al. 2020). The concept of joint distribution is needed to develop a conditional probability model. One approach that is often used to calculate the joint distribution is copula functions. The copula is a function that links the joint distribution and its marginal distributions (Schölzel and Friederichs 2008). The copula is often applied in finance, but there is very little application in climatology, especially forest fires analysis. Usually, fire size uses hotspots to quickly monitor forest fires in large areas (LAPAN 2016). Meanwhile, we know that forest fires in Kalimantan are very sensitive to El Nino conditions (Nurdiati et al. 2021a). Moreover, we can measure the drought severity using drought indicators, such as total precipitation and dry spells (Marinović et al. 2021; Thoithi et al. 2021).

This study aims to model and analyze the copula-based joint distribution between climate conditions and hotspots in Kalimantan, Indonesia. We constructed the bivariate joint distributions between climate conditions, either total precipitation or dry spells, and hotspots with sample size reduced by ENSO conditions, i.e., La Nina, normal, and El Nino. From the joint distribution, fire risk models are calculated using conditional probability and copula regression. The results are expected to increase understanding of the effect of ENSO on the joint distribution between drought indicators and hotspots in Kalimantan. Moreover, the fire risk model can be developed to predict future hotspots and improve the early warning system for forest fires in Indonesia.

2 Study area and data

The study area was taken in the Kalimantan fire-prone area, Indonesia (Najib et al. 2021a). This fire-prone area is obtained using k-mean clustering by eliminating the cluster with minimum hotspots. Hotspot data has a spatial resolution of $0.25^\circ \times 0.25^\circ$ derived from MODIS sensors of the Terra and Aqua satellites in 2001-2020 (Ardiansyah et al. 2017). Meanwhile, total precipitation and dry spells are derived from CMORPH CRT datasets (Sun et al. 2016). We defined dry spells as the number of days with total precipitation less than 1 mm/day. Based on previous research, we use an average of 2 months of total precipitation (TP) and a total of 3 months of dry spells (DS) because it has the strongest association with hotspots in the Kalimantan fire-prone area (Najib et al. 2021a). We only observe months with a significant number of hotspots, i.e., July-November. The time series is obtained by taking the average values of hotspots and drought indicators in the Kalimantan fire-prone area (Fig. 1). Moreover, Oceanic Nino Index (ONI) is used for classifying ENSO conditions. There are three classifications for ENSO condition, i.e., El Nino if ONI is more than 0.5, La Nina if ONI is less than -0.5, and normal condition if otherwise.

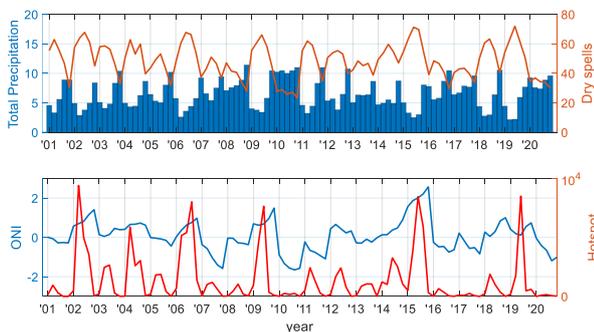


Fig. 1 Line plot of processed variables.

3 Method

3.1 Copula function

The copula is a function that links the joint distribution and its marginal distributions (Schölzel and Friederichs 2008). For the bivariate case, suppose that random variables X_1 and X_2 has marginal distributions F_1 dan F_2 , respectively. Therefore, the joint distribution between X_1 and X_2 is defined as a function of its marginal distributions, given by

$$F_X(x_1, x_2) = C[F_1(x_1), F_2(x_2)] \quad (1)$$

where $C: [0,1]^2 \rightarrow [0,1]$ is a copula function (Sklar 1959; Laux et al. 2011). Sklar then said that “every joint probability density f_X is also writable by the product of its marginal probability densities f_i and the copula density c ”, given by

$$f_X(x_1, x_2) = f_1(x_1) \cdot f_2(x_2) \cdot c[F_1(x_1), F_2(x_2)] \quad (2)$$

where

$$c[F_1(x_1), F_2(x_2)] = C''[F_1(x_1), F_2(x_2)] \quad (3)$$

We use different copula functions because each will further describe the dependency structure, such as one-parameter copulas and two-parameter copulas (Table 1). One-parameter copulas include Clayton, Frank, Gumbel, Joe, and Galambos copula. Meanwhile, two-parameter copulas include BB1, BB6, BB7, and BB8 copula.

Table 1 The copula function and parameter range

Code	Copula name	Copula function (C)	Parameter range
CL	Clayton	$(u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}$	$\theta > 0$
FR	Frank	$-\frac{1}{\theta} \ln \left[1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{(e^{-1} - 1)} \right]$	$\theta \neq 0$
GU	Gumbel	$\exp \left[- \left\{ [-\ln(u_1)]^\theta + [-\ln(u_2)]^\theta \right\}^{1/\theta} \right]$	$\theta \geq 1$
JO	Joe	$1 - \left\{ 1 - \left[1 - (1 - u_1)^\theta \right] \left[1 - (1 - u_2)^\theta \right] \right\}^{1/\theta}$	$\theta \geq 1$
GA	Galambos	$u_1 u_2 \exp \left[\left\{ [-\ln(u_1)]^{-\theta} + [-\ln(u_2)]^{-\theta} \right\}^{-1/\theta} \right]$	$\theta \geq 0$
BB1	Clayton-Gumbel	$\left\{ 1 + \left[(u_1^{-\theta} - 1)^\delta + (u_2^{-\theta} - 1)^\delta \right]^{1/\delta} \right\}^{-1/\theta}$	$\theta > 0, \delta \geq 1$
BB6	Joe-Gumbel	$1 - \left\{ 1 - \exp \left[- \left(\left\{ -\ln[1 - (1 - u_1)^\theta] \right\}^\delta + \left\{ -\ln[1 - (1 - u_2)^\theta] \right\}^\delta \right)^{1/\delta} \right] \right\}^{1/\theta}$	$\theta \geq 1, \delta \geq 1$
BB7	Joe-Clayton	$1 - \left(1 - \left\{ \left[1 - (1 - u_1)^\theta \right]^\delta + \left[1 - (1 - u_2)^\theta \right]^\delta - 1 \right\}^{-1/\delta} \right)^{1/\theta}$	$\theta \geq 1, \delta > 0$
BB8	Joe-Frank	$\frac{1}{\theta} \left(1 - \left\{ 1 - \frac{\left[1 - (1 - \delta u_1)^\theta \right] \left[1 - (1 - \delta u_2)^\theta \right]}{1 - (1 - \delta)^\theta} \right\}^{1/\delta} \right)$	$\theta \geq 1, 0 < \delta \leq 1$

3.2 Parameter estimation

The inference of functions for margins (IFM) method is the most common method used to estimate the copula parameters (Bouyé et al. 2000; Wei et al. 2020). This method is a two-step method. The first step is to assess the parameters of the marginal distribution of each variable. The marginal distributions are estimated by maximizing the log of the likelihood function, given by

$$\hat{\alpha}_i = \arg \max \sum_{t=1}^N \ln f_i(x_i^t; \alpha_i) \quad (4)$$

where $\hat{\alpha}$ is the estimated parameter of α and f_i is the probability density function of X_i for $i = 1, 2$. For drought indicators, we use generalized extreme value (GEV) and lognormal (LN) distributions, while negative binomial (NBIN) distribution is employed to fit the distribution of hotspot data. For simplicity, we use the negative of total precipitation to estimate the marginal distribution and copula parameter. Therefore, the dryness of drought indicators and the worse condition of hotspots are on the distribution joint's upper tail (right-top). However, we return the original data for visualization and analysis (Najib et al. 2021b). The probability density functions of each distribution are given by (Hilbe 2011; Pobočíková et al. 2017; Farooq et al. 2018; Sylvi et al. 2018; Baran et al. 2021)

$$f_{GEV}(x | k, \mu, \sigma) = \frac{1}{\sigma} \exp \left[-(1+kz)^{-\frac{1}{k}} \right] (1+kz)^{-1-\frac{1}{k}} \quad (5)$$

where $z = (x - \mu)/\sigma$,

$$f_{LN}(x | \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp \left[\frac{-(\ln x - \mu)^2}{2\sigma^2} \right], x > 0 \quad (6)$$

$$P(X = x; r, p) = \frac{\Gamma(x+r)}{x!\Gamma(r)} p^r (1-p)^x \quad (7)$$

The second step of the IFM method is to estimate the copula parameter by maximizing the log of the likelihood function of copula density, given by

$$\hat{\theta} = \arg \max \sum_{t=1}^N \ln c \left[F_1(x_1^t; \hat{\alpha}_1), F_2(x_2^t; \hat{\alpha}_2); \theta \right] \quad (8)$$

where $\hat{\theta}$ is the estimated parameter of θ and c is the copula density function. The fittest copula function is selected based on several statistical tests, such as

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (P_E^t - P_T^t)^2} \quad (9)$$

and

$$AIC = 2k - 2 \ln L \quad (10)$$

where k is the number of parameters, L is the likelihood function, P_T and P_E are the theoretical and empirical frequency, respectively. The theoretical frequency is obtained from the model of copula distribution. Meanwhile, the empirical frequency is calculated using Gringorten's formula (1963), given by

$$P_E^t = \frac{\#(X_1 \leq x_1^t, X_2 \leq x_2^t) - 0.44}{N + 0.12} \quad (11)$$

where N is the sample size of data.

3.3 Joint probability and return period

We can estimate the rarity of extreme or outlier events using the joint probability and return period. The joint probability of two conditions that are simultaneously exceeding a specific threshold ($X > x$ and $Y > y$) is defined by (Afshar et al. 2020)

$$\begin{aligned} P(X > x \cap Y > y) \\ = 1 - F_X(x) - F_Y(y) + C[F_X(x), F_Y(y)] \end{aligned} \quad (12)$$

Moreover, the joint return period (in months) is defined by $1/P$ (Zscheischler and Fischer 2020). However, this joint probability and return period are uncertain depending on the sample size (Link et al. 2020; Najib et al. 2021b). We can calculate the 95% confidence interval of the joint probability (Link et al. 2020), which is $[P - \varepsilon, P + \varepsilon]$, where

$$\varepsilon = 2\sqrt{\frac{P(1-P)}{N}} \quad (13)$$

3.4 Multivariate fire risk model

Madadgar et al. (2020) said, “the fire risk model is defined as the conditional probability of fire size given a specific climate condition (e.g., temperature or precipitation amount).” The conditional probability density function of fire size (Y) given a specific climate condition (X) is defined as (Madadgar and Moradkhani 2014)

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = f_Y(y) \cdot c[F_X(x), F_Y(y)] \quad (14)$$

where $f_{XY}(x,y)$ is a joint probability function of X and Y (Eq. 2), $c[F_X(x), F_Y(y)]$ is a copula density function derived from $C[F_X(x), F_Y(y)]$ (Eq. 3), $f_X(x)$ and $F_Y(y)$ are the probability density function of the marginal distributions of X and Y . From this probability, we can calculate the conditional probability $F_{Y|X}(Y > y|X = x)$ from the area under the curve $f_{Y|X}(y|x)$ in the domain of $Y > y$, given by

$$F_{Y|X}(Y > y|X = x) = \int_y^{\infty} f_{Y|X}(y|x) dx \quad (15)$$

where x can be a specific or interval climate condition, and we called it a conditional survival function.

We use drought indicators (either total precipitation or dry spells) and El Nino-Southern Oscillation (ENSO) as the climate conditions. Meanwhile, we use hotspots data as fire size. The ENSO conditions are not used to calculate the joint distribution, but we use them to reduce the sample size of data (Najib et al. 2021b). Therefore, we obtained three joint distributions between climate conditions and fire size for one model. Because we use two options for drought indicators, then we will discuss two models, i.e., total precipitation and dry spells as the climate conditions.

3.5 Copula regression

From the conditional probability (Eq. 9), we also can obtain the expected value of Y given a specific climate condition $X = x$ as (Noh et al. 2013)

$$\begin{aligned} E(Y|X = x) &= \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x) dy \\ &= \int_{-\infty}^{\infty} y \cdot f_Y(y) \cdot c[F_X(x), F_Y(y)] dy \end{aligned} \quad (16)$$

This kind of model is also known as copula-based or copula regression models (Kolev and Paiva 2009; Masarotto and Varin 2017; Cooke et al. 2020). We can estimate the expected value (Eq. 10) using Riemann’s sum approach (Jha and

Danjuma 2020). By partitioning the domain of Y into n finite number of sub-intervals, we can estimate the expected value of Y given a specific value of $X = x$ as

$$E(Y | X = x) = \sum_{i=1}^n y_i \cdot f_{Y|X}(y | x) \cdot \Delta y_i \quad (17)$$

where Δy_i is the length of the sub-interval $[y_{i-1}, y_i]$ (Nurdiati et al. 2021b). Meanwhile, we can calculate the 95% confidence interval $[y_a, y_b]$, by (Boubakar et al. 2018)

$$F_{Y|X}(Y < y_a | X = x) = F_{Y|X}(Y > y_b | X = x) = 2.5\% \quad (18)$$

4 Results and discussion

4.1 The copula parameter

First, we discuss the marginal distributions fitting process (Tabel 2). The statistical results of the marginal distributions show that the fittest distribution of total precipitation and dry spells passed the Anderson-Darling test with a 5% significant level (shown by p -value > 0.05) so that the marginal distributions can be employed to estimate the copula parameter. Meanwhile, we still use the negative binomial distribution to deal with zero-inflated hotspot data (Zhang and Yi 2020; Kang et al. 2021).

Table 2 The marginal distributions fitting results

ENSO Condition	Total Prec.		Dry Spells		Hotspots Dist
	Dist	p -value	Dist	p -value	
La Nina	GEV	0.9668	LN	0.9871	NBIN
Normal	GEV	0.9113	GEV	0.9983	NBIN
El Nino	GEV	0.9956	GEV	0.9951	NBIN

Using these marginal distributions, we estimate the copula parameters of TP – H (total precipitation – hotspots) and DS – H (dry spells – hotspots) with different ENSO conditions (Eq. 8). Interestingly, we got the same fittest copula for TP – H and DS – H, i.e., Joe, Galambos, and BB1 copulas during La Nina, normal, and El Nino conditions, respectively. They showed that both TP-H and DS-H had the same tail dependence structure characteristic. The Joe and Galambos copulas have an upper tail dependence but no lower tail dependence, while the BB1 copula has both lower and upper tail dependences (Joe 1997). The upper tail dependence coefficients for Joe and Galambos copulas (Joe 1997) are defined by

$$\lambda_U^{JO} = 2 - 2^{1/\theta} \quad (19)$$

$$\lambda_U^{GA} = 2^{-1/\theta} \quad (20)$$

Therefore, the upper tail dependence coefficients for TP – H and DS – H during La Nina are 0.75 and 0.72, respectively, and during normal ENSO conditions are 0.55 and 0.61, respectively. Moreover, the upper and lower tail dependence coefficients for BB1 copula (Joe et al. 2010) are defined by

$$\lambda_U^{BB1} = 2 - 2^{1/\delta} \quad (21)$$

$$\lambda_L^{BB1} = 2^{-1/(\delta\theta)} \quad (22)$$

Therefore, the upper and lower tail dependence coefficients for TP – H during El Nino are 0.52 and 0.67. Meanwhile, the upper and lower tail dependence coefficients for TP – H during El Nino are 0.47 and 0.70.

Table 3 The copula parameter fitting results

Data	Copula	θ	δ	RMSE	AIC
TP-H La Nina	Joe	3.0707	-	0.0482	-20.01
TP-H Normal	Galambos	1.1693	-	0.0316	-32.67
TP-H El Nino	BB1	0.9679	1.7809	0.0341	-31.48
DS-H La Nina	Joe	2.7993	-	0.0576	-18.62
DS-H Normal	Galambos	1.3859	-	0.0258	-41.18
DS-H El Nino	BB1	1.1862	1.6411	0.0308	-31.63

TP: Total precipitation, DS: Dry spells, H: Hotspots

4.2 The joint probability density function

Using the marginal distributions and copula parameters (Eq. 2), we construct the joint probability density functions between total precipitation and hotspots (Fig. 2) also between dry spells and hotspots (Fig. 3) during different ENSO conditions. Both figures show that ENSO conditions significantly affect the joint probability density function between drought indicators and hotspots in Kalimantan. El Nino will increase the chances of extreme hotspots in Kalimantan, while La Nina will reduce the chances of extreme hotspots. However, there is an outlier during normal ENSO conditions, i.e., in September 2019, where there are 8497 hotspots in a month. This 2019 hotspot event is infrequent to happen. We can calculate the rarity from the joint probability (Eq. 12 and 13) between total precipitation and hotspot in September 2019, i.e., 0.0034 with a 95% confidence interval $[0, 0.0196]$. In other words, the joint return period is 296 months with a 95% confidence interval $[51, \infty]$ if all months are assumed in normal ENSO conditions. The same approach can be applied using the joint distribution between dry spells and hotspots, and we found that the joint return period is 399 months with a 95% confidence interval $[61, \infty]$ if all months are in normal ENSO conditions. The joint return periods show the rarity of extreme hotspots events during normal ENSO conditions.

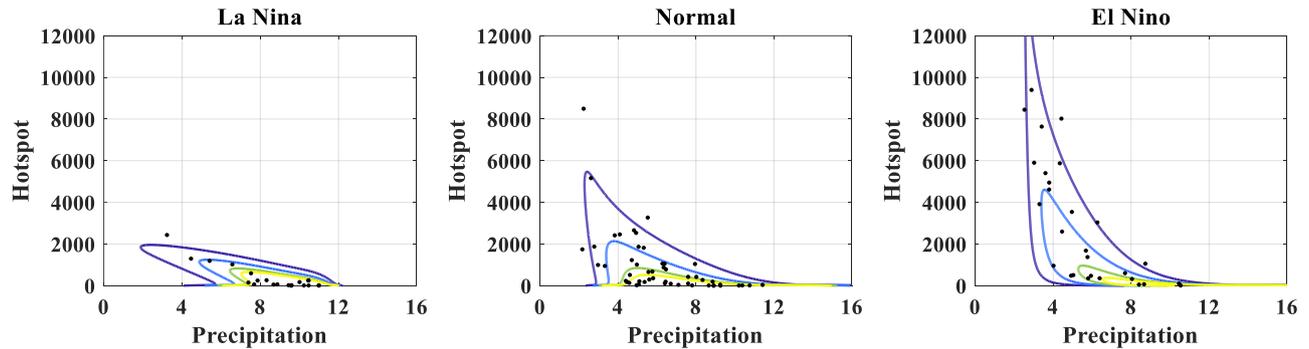


Fig. 2 The joint probability density functions between total precipitation and hotspots

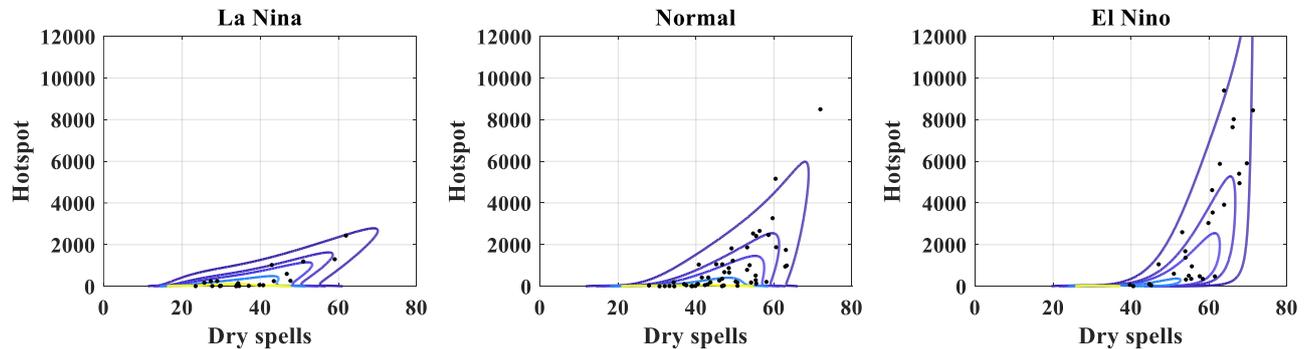


Fig. 3 The joint probability density functions between dry spells and hotspots

4.3 The conditional probability

To get more detail about the effect of ENSO conditions on extreme hotspots events, we calculated the conditional survival functions (Eq. 15) for total precipitation and hotspots (Fig. 4) and dry spells and hotspot (Fig.5). We use the dryness of climate conditions as a given condition, i.e., lower tercile for total precipitation ($TP < 5.2201$ mm/day) and upper tercile for dry spells ($DS > 54$ days).

The conditional survival function given the dry condition of total precipitation (Fig. 4) found that extreme hotspots' probability depends on ENSO conditions. We divide the analysis into three conditions, i.e., more than 90% quantile, more than 95% quantile, and more than 2019 hotspots event, the latest extreme wildfire in Kalimantan. We obtained the conditional probability of more than 90% of hotspots quantile: 2.26%, 12.80%, and 45.92% during La Nina, normal, and El Nino, respectively. El Nino increased the conditional probability more than triple normal conditions, while La Nina significantly suppressed the probability of extreme hotspots in Kalimantan. In more extreme conditions (95% quantile), the probability decreases to 0.22% during La Nina, while during normal conditions, the probability is still a bit high, i.e., 4.91%. However, for very extreme hotspots events (more than in 2019), the condition probability during La Nina and normal conditions is very small, i.e., 0.006% and 1.1%. This probability indicates that extreme hotspot events in September 2019 or more severe are infrequent during normal ENSO conditions even can be impossible during La Nina. Otherwise, the conditional probability is still relatively high during El Nino, i.e., 14.4% or about 1/7 months.

The same pattern also appears in the conditional survival function given the dry condition of dry spells (Fig. 5). The conditional probability for extreme hotspots is very small during La Nina. Meanwhile, there is still a bit high probability for 90% and 95% quantiles during normal ENSO conditions, but the probability of extremely high hotspots events (more than in 2019) is very small, i.e., 1.51%. However, the conditional probability is still high during El Nino, i.e., 11.25%.

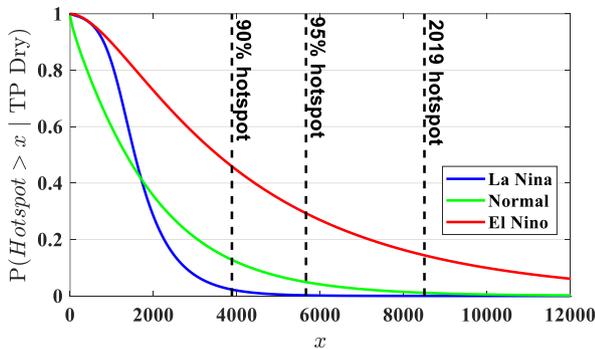


Fig. 4 The conditional survival function of hotspots given the dry condition of total precipitation

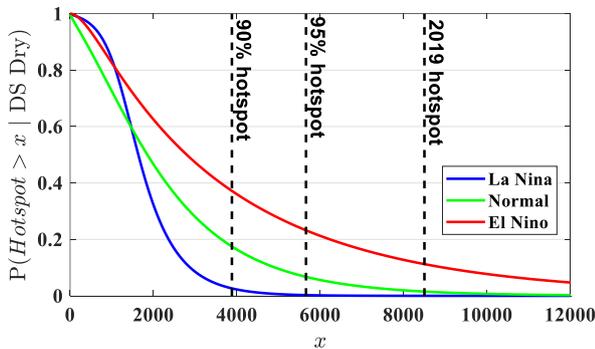


Fig. 5 The conditional survival function of hotspots given the dry condition of dry spells

4.4 The copula regression models

Using the joint probability density function (Fig. 2 and 3), we calculate the expected value for hotspots (Eq. 16) with a 95% confidence interval (Eq. 18) given a specific climate condition (precipitation and dry spells) for each ENSO condition. Given specific total precipitation, the result of copula regression (Fig. 6) shows that almost all data points are inside the 95% confidence interval, except one point in normal ENSO conditions. Furthermore, RMSE values of the expected value from the original data are 271, 1062, and 2073 hotspots for La Nina, normal, and El Nino conditions, respectively. Meanwhile, the coefficient of determinations (R^2) are 79.65, 48.93, and 49.90% for La Nina, normal, and El Nino conditions. From these results, the fittest copula regression given total precipitation is in La Nina conditions. Moreover, although the RMSE value during El Nino was much higher than normal conditions, the copula regression model's R^2 during El Nino better than normal conditions. Other than that, at least all data points during El Nino are inside the 95% confidence interval.

Meanwhile, all data points are inside the 95% confidence interval on the result of copula regression given specific dry spells conditions. The statistical results show that the RMSE values are 262, 855, and 1.907 hotspots; also the R^2 values are 80.98, 66.92, and 57.59% for La Nina, normal, and El Nino conditions. The highest error of the copula regression model is during La Nina conditions, while the lowest is during El Nino, but the R^2 value is still high, i.e., 57.59%. The RMSE and R^2 values of copula regression given specific dry spells much better than given specific total precipitation during all ENSO conditions.

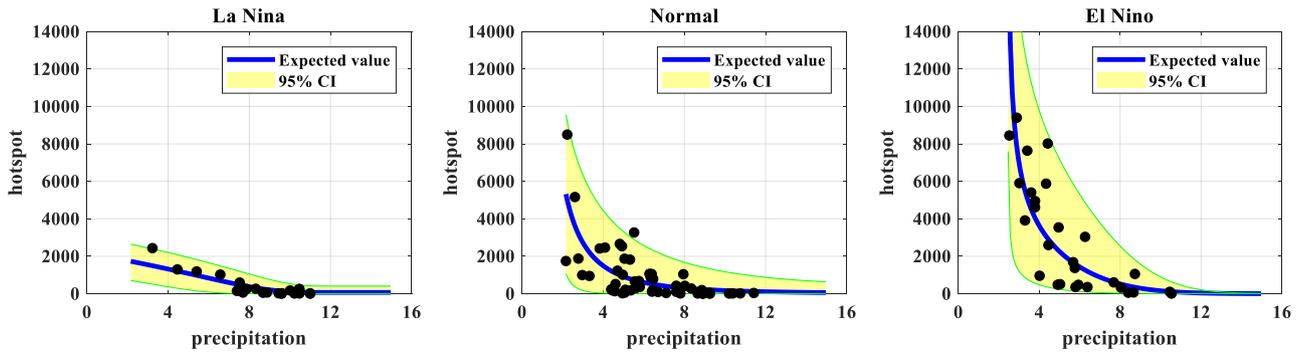


Fig. 6 The copula regression results: the expected value for hotspots with a 95% confidence interval given specific precipitation for each ENSO condition.

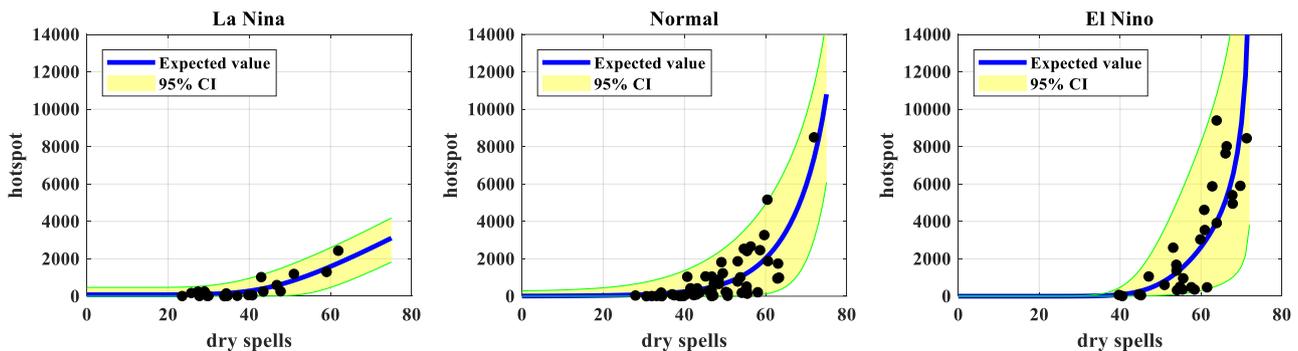


Fig. 7 The copula regression results: the expected value for hotspots with a 95% confidence interval given specific dry spells for each ENSO condition.

By evaluating all data points into the copula regression model based on specific climate dan ENSO conditions, we obtained the time series of the copula regression model with the 95% confidence interval (Fig. 8 and 9). If given specific precipitation, the time series result (Fig. 8) shows that the number of hotspots in August 2014 is outside the 95% confidence interval of the copula regression model as previously discussed. Several years have also not shown satisfactory results, especially in 2004, 2006, 2014, 2015, and 2019. The expected value in 2014 does not follow the actual hotspot pattern (up,

down, and up). In 2006, 2015, and 2019, the expected values were underestimated from the actual hotspots. Even in 2014, the actual hotspots were not reached by the confidence interval. Contrary, the expected values were wildly overestimated in 2015, almost twice the actual hotspots. However, the expected value for hotspots in 2002 is satisfactory. Overall, the copula regression for hotspots given specific total precipitation and ENSO conditions have the RMSE value of 1339 hotspots and the R^2 value of 60.70%.

Meanwhile, the time series of the copula regression model for hotspots given specific dry spells and ENSO conditions (Fig. 9) shows the expected value and 95% confidence interval are satisfactory. The expected values given specific dry spells fit better than given total precipitation (Fig. 8) in 2004, 2006, 2014, 2015, and 2019. In 2004, the expected values followed the pattern of the actual hotspots number. The copula regression model was not too underestimated in 2006 and 2014; also not too overestimated in 2015. Even better, all actual hotspots are within the 95% confidence interval, including 2014. The expected value was even very fit in 2019. However, the expected value in 2002 is not as good as the previous model. Overall, the copula regression for hotspots given specific dry spells and ENSO conditions have the RMSE value of 1185 hotspots and the R^2 value of 69.21%. These statistical results are better than the previous model.

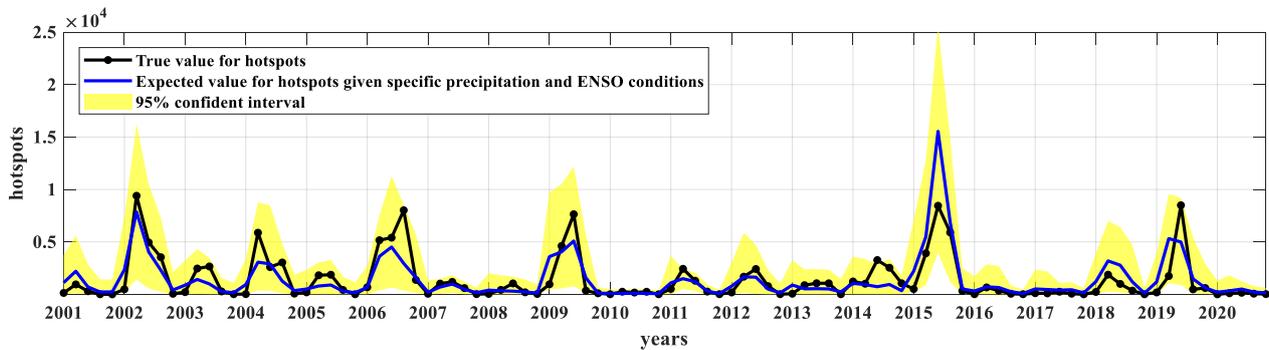


Fig. 8 The time series of the copula regression model for hotspots given specific total precipitation and ENSO conditions

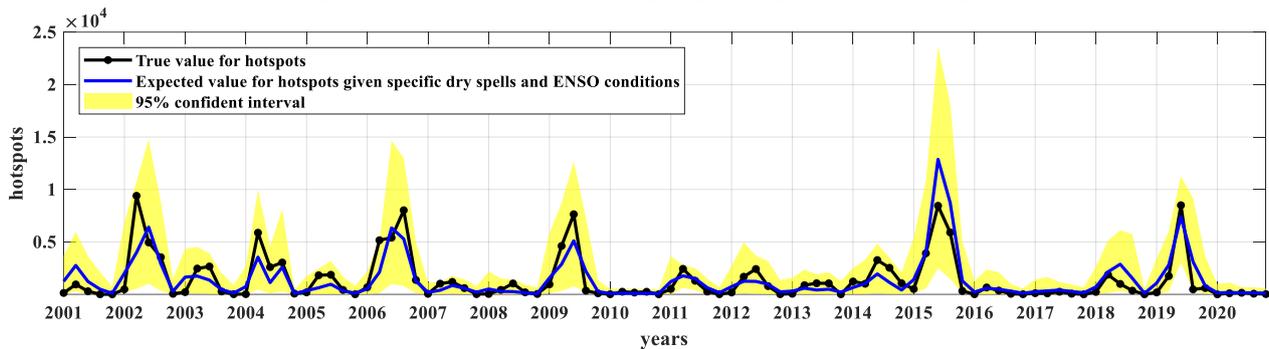


Fig. 9 The time series of the copula regression model for hotspots given specific total precipitation and ENSO conditions

5 Summary and Conclusion

This study focused on modeling and analyzing the copula-based joint distribution between climate conditions and hotspots with different ENSO conditions, i.e., La Nina, normal, and El Nino. We calculated the expected value of hotspots given specific climate and ENSO conditions with a 95% confidence interval using these joint distributions. This study partially uses two indicators as climate conditions, i.e., precipitation and dry spells, so we obtained two models based on the indicator we used.

The copula parameters fitting results show that the copula functions of total precipitation – hotspots and dry spells – hotspots are the same with different values of parameters, i.e., Joe, Galambos, and BB1 copula for La Nina, normal, and El Nino condition, respectively. We know from these copulas that the relationship between climate conditions and hotspots in La Nina and normal ENSO conditions have an upper tail dependence but no lower tail dependence. Meanwhile, during El Nino, the relationship has both upper and lower tail dependences.

From the joint probability function and conditional survival function, there is an outlier in normal ENSO conditions that has more hotspots than normally, i.e., in September 2019. This anomaly is very rare to happen, shown by the joint return period of total precipitation – hotspots, i.e., 296 months with a 95% confidence interval $[51, \infty]$ months and the joint return period of dry spells – hotspots, i.e., 366 months with a 95% confidence interval $[61, \infty]$ months. The conditional survival function given the dry condition of climate indicators reinforced this statement. The probability of hotspots more severe than in 2019 is very low during normal ENSO conditions, i.e., less than 2%. The only ENSO condition with relatively high probability is during El Nino, i.e., greater than 10%. Moreover, the probability of hotspots being more severe than in 2019 during La Nina is very low, almost near zero.

The copula regression models show that the model given specific dry spells is better than that given specific total precipitation as climate condition, shown by better RMSE and R^2 values. The copula regression for hotspots given specific total precipitation and ENSO conditions have the RMSE value of 1339 hotspots and the R^2 value of 60.70%. Meanwhile, the copula regression for hotspots given specific dry spells and ENSO conditions have the RMSE value of 1185 hotspots and the R^2 value of 69.21%. Overall, given specific dry spells, the copula regression model is more fit every year, especially in 2004, 2006, 2014, 2015, and 2019. Otherwise, the copula regression model given specific total precipitation is more fit only in 2002.

This study is limited to bivariate joint distribution. Therefore, the obtained model contains only one climate condition, either total precipitation or dry spells, with ENSO condition as a sample size reducer. The suggestion for further research from this study is to use a trivariate copula or more to build a joint distribution so that the climate condition used is not only one between total precipitation and dry spells.

Declarations

Funding

There is no funding for this research

Conflicts of interest

There is no conflict of interest for this research.

Availability of data and material

The raw data of climate conditions provided by NOAA CPC Morphing Technique ("CMORPH") can be downloaded freely on https://ftp.cpc.ncep.noaa.gov/precip/PORT/SEMDP/CMORPH_CRT/DATA. Meanwhile, the raw data of hotspots provided by the Agency for Meteorology, Climatology, and Geophysics of Indonesia were processed from MODIS sensors of the Terra and Aqua satellites. This article's processed data (time series) are available from the corresponding author on reasonable request.

Code availability

The parameter fitting codes can be downloaded from <https://github.com/mkhoirun-najiboi/mycopula/>, and the specific codes used in this article are available from the corresponding author on reasonable request.

Ethics approval

This article does not contain any studies with human or animal participants.

Consent to Participate

We provide our full consent to participate in the process of the journal.

Consent to Publish

We provide our full consent to publish our article in the journal.

References

- Aflahah E, Hidayati R, Hidayat R (2019) Pendugaan hotspot sebagai indikator kebakaran hutan di Kalimantan berdasarkan faktor iklim. *J Pengelolaan Sumberd Alam dan Lingkung* 9(2):405–418. <https://doi.org/10.29244/jpsl.9.2.405-418>
- Afshar MH, Şorman AÜ, Tosunoğlu F, Bulut B, Yılmaz MT, Danandeh Mehr A (2020) Climate change impact assessment on mild and extreme drought events using copulas over Ankara, Turkey. *Theor Appl Climatol* 141(3–4):1045–1055. <https://doi.org/10.1007/s00704-020-03257-6>
- Ardiansyah M, Boer R, Situmorang AP (2017) Typology of land and forest fire in South Sumatra, Indonesia Based on Assessment of MODIS Data. *IOP Conf Ser Earth Environ Sci* 54(1):012058. <https://doi.org/10.1088/1755-1315/54/1/012058>
- Austin KG, Schwantes A, Gu Y, Kasibhatla PS (2019) What causes deforestation in Indonesia? *Environ Res Lett* 14(2). <https://doi.org/10.1088/1748-9326/aaf6db>
- Baran S, Szokol P, Szabó M (2021) Truncated generalized extreme value distribution-based ensemble model output statistics model for calibration of wind speed ensemble forecasts. *Environmetrics*. <https://doi.org/10.1002/env.2678>
- Boubakar T, Lassina D, Belco T, Abdou F (2018) The Shortest Confidence Interval for the Mean of a Normal Distribution. *Int J Stat Probab* 7(2):33. <https://doi.org/10.5539/ijsp.v7n2p33>
- Bouyé E, Durrleman V, Nikeghbali A, Riboulet G, Roncalli T (2000) Copulas for Finance - A Reading Guide and Some Applications. <https://dx.doi.org/10.2139/ssrn.1032533>
- Brogan D, Nelson P, MacDonald L (2019) Spatial and temporal patterns of sediment storage and erosion following a wildfire and extreme flood. *Spat temporal patterns sediment storage Eros Follow a wildfire Extrem flood* 7(2):1–48. <https://doi.org/10.5194/esurf-2018-98>
- Cardil A, Rodrigues M, Ramirez J, de-Miguel S, Silva CA, Mariani M, Ascoli D (2021) Coupled effects of climate teleconnections on drought, Santa Ana winds and wildfires in southern California. *Sci Total Environ* 765. <https://doi.org/10.1016/j.scitotenv.2020.142788>
- Cooke RM, Joe H, Chang B (2020) Vine copula regression for observational studies. *ASTA Adv Stat Anal* 104(2):141–167. <https://doi.org/10.1007/s10182-019-00353-5>
- Daşdemir İ, Aydın F, Ertuğrul M (2021) Factors Affecting the Behavior of Large Forest Fires in Turkey. *Environ Manage* 67(1):162–175. <https://doi.org/10.1007/s00267-020-01389-z>
- Fanin T, van der Werf G (2016) Precipitation-fire linkages in Indonesia (1997–2015). *Biogeosciences Discuss* :1–21. <https://doi.org/10.5194/bg-2016-443>
- Farooq M, Shafique M, Khattak MS (2018) Flood frequency analysis of river swat using Log Pearson type 3, Generalized Extreme Value, Normal, and Gumbel Max distribution methods. *Arab J Geosci* 11(9). <https://doi.org/10.1007/s12517-018-3553-z>
- Field RD, Van Der Werf GR, Fanin T, Fetzer EJ, Fuller R, Jethva H, Levy R, Livesey NJ, Luo M, Torres O, Worden HM (2016) Indonesian fire activity and smoke pollution in 2015 show persistent nonlinear sensitivity to El Niño-induced drought. *Proc Natl Acad Sci U S A* 113(33):9204–9209. <https://doi.org/10.1073/pnas.1524888113>
- FWI (2020) Within 75 Years of Independence, Indonesia has lost more than 75 times the size of Yogyakarta Province of its forest. <https://fwi.or.id/en/within-75-years-of-independence-indonesia-has-lost-more-than-75-times-the-size-of-yogyakarta-province-of-its-forest/>. Accessed 29 Aug 2021
- Gringorten II (1963) A plotting rule for extreme probability paper. *J Geophys Res* 68(3):813–814
- Hilbe JM (2011) *Negative binomial regression*. Cambridge University Press
- Jha BK, Danjuma YJ (2020) Unsteady Dean flow formation in an annulus with partial slippage: A riemann-sum approximation approach. *Results Eng* 5. <https://doi.org/10.1016/j.rineng.2019.100078>
- Joe H (1997) *Multivariate models and multivariate dependence concepts*. CRC Press, London
- Joe H, Li H, Nikoloulopoulos AK (2010) Tail dependence functions and vine copulas. *J Multivar Anal* 101(1):252–270. <https://doi.org/10.1016/j.jmva.2009.08.002>
- Kang KI, Kang K, Kim C (2021) Risk factors influencing cyberbullying perpetration among middle school students in korea: Analysis using the zero-inflated negative binomial regression model. *Int J Environ Res Public Health* 18(5):1–13. <https://doi.org/10.3390/ijerph18052224>
- KLHK (2020) Hutan dan Deforestasi Indonesia Tahun 2019. https://ppid.menlhk.go.id/siaran_pers/browse/2435. Accessed 30 Apr 2021
- Kolev N, Paiva D (2009) Copula-based regression models: A survey. *J Stat Plan Inference* 139(11):3847–3856. <https://doi.org/10.1016/j.jspi.2009.05.023>

- Krebs MA, Reeves MC, Baggett LS (2019) Predicting understory vegetation structure in selected western forests of the United States using FIA inventory data. *For Ecol Manage* 448:509–527. <https://doi.org/10.1016/j.foreco.2019.06.024>
- LAPAN (2016) Informasi Titik Panas (Hotspot) Kebakaran Hutan / Lahan. http://pusfatja.lapan.go.id/files/uploads_ebook/publikasi/Panduan_hotspot_2016_versi_draft_1_LAPAN.pdf. Accessed 21 Feb 2021
- Larsen A, Hanigan I, Reich BJ, Qin Y, Cope M, Morgan G, Rappold AG (2021) A deep learning approach to identify smoke plumes in satellite imagery in near-real time for health risk communication. *J Expo Sci Environ Epidemiol* 31(1):170–176. <https://doi.org/10.1038/s41370-020-0246-y>
- Laux P, Vogl S, Qiu W, Knoche HR, Kunstmann H (2011) Copula-based statistical refinement of precipitation in RCM simulations over complex terrain. *Hydrol Earth Syst Sci* 15(7):2401–2419. <https://doi.org/10.5194/hess-15-2401-2011>
- Link R, Wild TB, Snyder AC, Hejazi MI, Vernon CR (2020) 100 Years of Data Is Not Enough To Establish Reliable Drought Thresholds. *J Hydrol X* 7. <https://doi.org/10.1016/j.hydroa.2020.100052>
- Madadgar S, Moradkhani H (2014) Spatio-temporal drought forecasting within Bayesian networks. *J Hydrol* 512:134–146. <https://doi.org/10.1016/j.jhydrol.2014.02.039>
- Madadgar S, Sadegh M, Chiang F, Ragno E, AghaKouchak A (2020) Quantifying increased fire risk in California in response to different levels of warming and drying. *Stoch Environ Res Risk Assess* 34(12):2023–2031. <https://doi.org/10.1007/s00477-020-01885-y>
- Marinović I, Cindrić Kalin K, Güttler I, Pasarić Z (2021) Dry spells in Croatia: Observed climate change and climate projections. *Atmosphere (Basel)* 12(5). <https://doi.org/10.3390/atmos12050652>
- Masarotto G, Varin C (2017) Gaussian copula regression in R. *J Stat Softw* 77(1). <https://doi.org/10.18637/jss.v077.i08>
- Miettinen J, Shi C, Liew SC (2011) Deforestation rates in insular Southeast Asia between 2000 and 2010. *Glob Chang Biol* 17(7):2261–2270. <https://doi.org/10.1111/j.1365-2486.2011.02398.x>
- Najib MK, Nurdianti S, Sopaheluwakan A (2021a) Quantifying the Joint Distribution of Drought Indicators in Borneo Fire-Prone Area. In: *The 4th International Conference on Science & Technology Applications in Climate Change*
- Najib MK, Nurdianti S, Sopaheluwakan A (2021b) Copula-based joint distribution analysis of the ENSO effect on the drought indicators over Borneo fire-prone areas. *Model Earth Syst Environ*. <https://doi.org/10.1007/s40808-021-01267-5>
- Nikonovas T, Spessa A, Doerr SH, Clay GD, Mezbahuddin S (2020) Near-complete loss of fire-resistant primary tropical forest cover in Sumatra and Kalimantan. *Commun Earth Environ* 1(1). <https://doi.org/10.1038/s43247-020-00069-4>
- Noh H, El Ghouch A, Bouezmarni T (2013) Copula-based regression estimation and inference. *J Am Stat Assoc* 108(502):676–688. <https://doi.org/10.1080/01621459.2013.783842>
- Nurdianti S, Bukhari F, Julianto MT, Najib MK, Nazria N (2021a) Heterogeneous Correlation Map Between Estimated ENSO And IOD From ERA5 And Hotspot In Indonesia. *Jambura Geosci Rev* 3(2):65–72. <https://doi.org/10.34312/jgeosrev.v3i2.10443>
- Nurdianti S, Khatizah E, Najib MK, Fatmawati LL (2021b) El nino index prediction model using quantile mapping approach on sea surface temperature data. *Desimal J Mat* 4(1):79–92. <https://doi.org/10.24042/djm.v4i1.7595>
- Pobočková I, Sedláčková Z, Michalková M (2017) Application of Four Probability Distributions for Wind Speed Modeling. *Procedia Eng* 192:713–718. <https://doi.org/10.1016/j.proeng.2017.06.123>
- Ryan RG, Silver JD, Schofield R (2021) Air quality and health impact of 2019–20 Black Summer mega-fires and COVID-19 lockdown in Melbourne and Sydney, Australia. *Environ Pollut* 274. <https://doi.org/10.1016/j.envpol.2021.116498>
- Sachdeva S, Bhatia T, Verma AK (2018) GIS-based evolutionary optimized Gradient Boosted Decision Trees for forest fire susceptibility mapping. *Nat Hazards* 92(3):1399–1418. <https://doi.org/10.1007/s11069-018-3256-5>
- Schölzel C, Friederichs P (2008) Multivariate non-normally distributed random variables in climate research - Introduction to the copula approach. *Nonlinear Process Geophys* 15(5):761–772. <https://doi.org/10.5194/npg-15-761-2008>
- Silveira S, Kornbluh M, Withers MC, Grennan G, Ramanathan V, Mishra J (2021) Chronic mental health sequelae of climate change extremes: A case study of the deadliest Californian wildfire. *Int J Environ Res Public Health* 18(4):1–15. <https://doi.org/10.3390/ijerph18041487>
- Sklar M (1959) Fonctions de Répartition à Dimensions et Leurs Marges. *Publ L'Institut Stat L'Université Paris* 8:229–231
- Sulova A, Arsanjani JJ (2021) Exploratory analysis of driving force of wildfires in Australia: An application of machine learning within google earth engine. *Remote Sens* 13(1):1–23. <https://doi.org/10.3390/rs13010010>
- Sun R, Yuan H, Liu X, Jiang X (2016) Evaluation of the latest satellite-gauge precipitation products and their hydrologic applications over the Huaihe River basin. *J Hydrol* 536:302–319. <https://doi.org/10.1016/j.jhydrol.2016.02.054>
- Sylvi N, Ispriyanti D, Sugito S (2018) Penerapan Regresi Zero-Inflated Generalized Poisson dan Pengujian Autokorelasi Spasial Pada Kasus Penyakit Filariasis di Jawa Tengah. *J Stat Univ Muhammadiyah Semarang* 6(1):29–33
- Thoithi W, Blamey RC, Reason CJC (2021) Dry Spells, Wet Days, and Their Trends Across Southern Africa During the Summer Rainy Season. *Geophys Res Lett* 48(5). <https://doi.org/10.1029/2020GL091041>
- Viana-Soto A, Aguado I, Salas J, García M (2020) Identifying post-fire recovery trajectories and driving factors using Landsat time series in fire-prone Mediterranean pine forests. *Remote Sens* 12(9). <https://doi.org/10.3390/RS12091499>
- Wei X, Zhang H, Singh VP, Dang C, Shao S, Wu Y (2020) Coincidence probability of streamflow in water resources area, water receiving area and impacted area: Implications for water supply risk and potential impact of water transfer. *Hydrol Res* 51(5):1120–1135. <https://doi.org/10.2166/nh.2020.106>
- Xi DDZ, Dean CB, Taylor SW (2020) Modeling the duration and size of extended attack wildfires as dependent outcomes. *Environmetrics* 31(5). <https://doi.org/10.1002/env.2619>
- Xu Z, Liu D, Yan L (2021) Temperature-based fire frequency analysis using machine learning: A case of Changsha, China. *Clim Risk Manag* 31. <https://doi.org/10.1016/j.crm.2021.100276>

- Zhang X, Yi N (2020) Fast zero-inflated negative binomial mixed modeling approach for analyzing longitudinal metagenomics data. *Bioinformatics* 36(8):2345–2351. <https://doi.org/10.1093/bioinformatics/btz973>
- Zhao S, Xu Y (2021) Exploring the dynamic Spatio-temporal correlations between pm2.5 emissions from different sources and urban expansion in Beijing-Tianjin-Hebei region. *Int J Environ Res Public Health* 18(2):1–19. <https://doi.org/10.3390/ijerph18020608>
- Zscheischler J, Fischer EM (2020) The record-breaking compound hot and dry 2018 growing season in Germany. *Weather Clim Extrem* 29. <https://doi.org/10.1016/j.wace.2020.100270>