Journal of Quality Measurement and Analysis *JQMA* 21(1) 2025, 237-251
e-ISSN: 2600-8602 *http://www.ukm.my/jqma*
https://doi.org/10.17576/jqma.2101.2025.15

FORECASTING LOCATIONS OF FOREST FIRES IN INDONESIA THROUGH NONPARAMETRIC PREDICTIVE INFERENCE WITH PARAMETRIC COPULA: A CASE STUDY

(Meramal Lokasi Kebakaran Hutan di Indonesia Melalui Inferens Ramalan Bukan Parametrik dengan Kopula Parametrik: Kajian Kes)

AMIRAH HAZWANI ROSLIN, NORYANTI MUHAMMAD*, EVIZAL ABDUL KADIR, WARIH MAHARANI & HANITA DAUD

ABSTRACT

Wildfires caused major damage and incurred high restoration costs. Despite numerous predictive studies in this field, wildfire management still had uncertainties. The machine learning technique was popular on this topic, but it portrayed gaps of non-generalisable and inaccuracy possibilities. This study intended to apply nonparametric predictive inference (NPI) with a parametric copula to predict the next wildfire location using the coordinate parameters. The NPI quantifies the uncertainties via imprecise probabilities, $(\underline{P}, \overline{P})$, while the copula integration considers the spatial correlation by modelling the dependence structure between the past coordinates in predicting the next location. Unlike other methods, the NPI generates a set of bounded probabilities that provide confidence in the prediction result. This paper applied the proposed method to the Moderate Resolution Imaging Spectroradiometer satellite dataset for Indonesia (2020). Several wildfire hotspots in Sumatra and Kalimantan archipelago were focused on this study. It was evaluated via the differences $(\bar{d})$ within the $(\underline{P}, \overline{P})$ and showcased low values ($\bar{d} < 0.001$). The results show that NPI with parametric copula was highly accurate for both archipelagoes, highlighting its generalisability specifically for Indonesia. Each wildfire hotspot had a different optimal copula to predict the best future hotspot. Clayton and Gumbel copulae were the best to be integrated with NPI to predict the next wildfire location in Sumatra while Normal and Gumbel copulae for Kalimantan locations. In conclusion, the NPI is considered a reliable alternative for wildfire location prediction.

Keywords: copula; imprecise probability; Indonesia; nonparametric predictive inference; wildfire hotspot

ABSTRAK

Kebakaran hutan menyebabkan kerosakan serius yang menyumbang kepada kos pemulihan yang tinggi. Walaupun terdapat banyak kajian ramalan dalam bidang ini, pengurusan kebakaran hutan masih mempunyai ketidakpastian. Teknik pembelajaran mesin popular dalam topik ini, tetapi tetap mempunyai jurang kajian seperti tidak boleh digeneralisasikan dan ketidak tepatan. Kajian ini bertujuan untuk menggunakan inferens ramalan bukan parametrik (NPI) dengan kopula parametrik dalam meramalkan lokasi kebakaran hutan yang seterusnya berdasarkan parameter koordinat. NPI mengukur ketidakpastian melalui kebarangkalian yang tidak tepat, $(\underline{P}, \overline{P})$, manakala integrasi kopula mempertimbangkan korelasi keruangan melalui pemodelan struktur pergantungan antara koordinat lepas dalam meramalkan lokasi seterusnya. Tidak seperti kaedah lain, NPI menjana satu set kebarangkalian terhad yang memberikan keyakinan terhadap hasil ramalan. Kajian ini menggunakan kaedah cadangan ini kepada data Indonesia (2020) yang diperoleh daripada satelit Spektroradiometer Pengimejan Resolusi Sederhana. Kajian ini menumpukan beberapa lokasi titik panas kebakaran di kepulauan Sumatera dan Kalimantan sahaja. Analisis ini dinilai melalui perbezaan $(\bar{d})$ dalam $(\underline{P}, \overline{P})$ dan mempamerkan nilai yang rendah ($\bar{d} < 0.001$). Keputusan menunjukkan bahawa NPI dengan kopula parametrik adalah sangat tepat untuk kedua-dua kepulauan tersebut, menunjukkan keupayaan metod ini untuk digeneralisasikan, khususnya untuk keseluruhan Indonesia. Setiap titik panas kebakaran hutan mempunyai kopula optimum berbeza untuk meramalkan lokasi titik panas pada masa depan. Kopula *Clayton* dan *Gumbel* adalah yang terbaik untuk diintegrasikan dengan NPI bagi meramalkan lokasi kebakaran hutan seterusnya di Sumatera manakala kopula *Normal* dan *Gumbel* pula untuk lokasi di Kalimantan. Kesimpulannya, NPI boleh dianggap sebagai alternatif yang dipercayai untuk ramalan lokasi kebakaran.

Kata kunci: Indonesia; inferens ramalan bukan parametrik; kebarangkalian terhad; kopula; titik panas kebakaran hutan

1. Introduction

Annually, ample hectares (ha) of global forests are lost to forest fires. Indonesia ranked sixth worldwide for the highest annual tree cover loss due to forest fires from 2001 until 2022 (Global Forest Watch 2024). Forest fire or wildfire is among the most destructive natural disasters with multiple serious implications on the economy, ecology and society. It is a periodic disaster that happens constantly as compared to other disaster types (Gong *et al.* 2021). Commonly, wildfires are not a major concern in Southeast Asia, but Indonesia recorded a dreadful forest fire history which raises concerns (Chew *et al.* 2022; Negara *et al.* 2020). Based on Global Forest Watch (2024), Indonesia suffered a total of 26.6Mha of tree cover loss from wildfires for the same period. The most extreme wildfire record was in 2016 due to the El-Niño phenomenon which lost 30% of the Indonesian forests (Miettinen *et al.* 2017). It inflicted property destruction, reduced vegetation density and grave suffering for occupants leading to expensive restoration costs. However, small fire detection was rarely successful for early fire containment. This issue urged the government and firefighter departments to find small fire detection solutions before they become dangerous (Gong *et al.* 2021).

Mathematical and statistical concepts bring great advantages to forest fire management like predictive analytics and machine learning (ML). They can identify the hidden relationship within the complex environment which is usually present in the forest fire occurrences. Forest firefighter departments and policymakers started to shift to statistical methods as they are more dependable and valid in monitoring wildfire crises. Amongst, ML has been the most popular discussion among forest fire researchers in the past decades. Information extracted from the ML techniques is trusted since it can analyse large data and consider the variability simultaneously (Nguyen *et al.* 2018; Yu *et al.* 2018). There is an abundance of wildfire prediction past studies but still little was known about this nature. Generally, common ignition wildfire parameters were specified in these studies but showed diverse results when investigated in other regions.

Despite the ML flexibility, it has several flaws in parameter selection and their generalizability. Although ML models deliver highly accurate results, the suitable parameters selected in a certain region tend to be context-specific and cannot be replicated in other places (Chew *et al.* 2022). Different literature reviews yielded varied conclusions on the most influential parameters. Unique geographic and weather conditions in each study area called for a new ML framework design each time, creating redundancy and complicating future research efforts (He *et al.* 2022; Mohajane *et al.* 2021; Negara *et al.* 2020; Nguyen *et al.* 2018; Pham *et al.* 2020). Apart from that, current ML algorithms struggle with white noises that may influence forest fire prediction accuracy. For instance, weather conditions, infrared beam interference, and heat radiation can lead to misinformation in satellite data collection (Yu *et al.* 2018).

Jain *et al.* (2020) suggested including the spatial correlation between factors to predict the forest fire locations. This added consideration can enhance the prediction accuracy, but it is unsuitable for parametric methodologies like most ML models employed in forest fire studies due to their correlation assumption. Violation of the correlation assumption exposes the ML results to inaccuracy possibility. This situation was relevant to Indonesia's forest fire environment as claimed by Chew *et al.* (2022). Hence, nonparametric predictive inference (NPI) with a parametric copula was proposed to predict the next forest fire hotspot. It is a new methodology in forest fire study and is supposedly able to enhance wildfire prediction accuracy because it considers spatial correlation. The NPI will utilise new parameters of Indonesia's past coordinates of wildfire history to predict the next wildfire location. The proposed method would be deployed with the Indonesia (2020) dataset attained through the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite. Still, it is noteworthy that the NPI method has yet to find application in natural studies.

2. Literature Reviews

Table 1 compiled summaries of past literature reviews on wildfire prediction studies that used ML models along with the parameters. The bold parameter in Table 1 refers to the influential factor(s) in wildfire susceptibility prediction, except for He *et al.* (2022) who focus on fire spread and Safi and Bouroumi (2013) who predicted the burned area based on wildfire parameters. Many researchers in wildfire studies employed ML predictive models and a considerable number of them conducted comparison studies (Mohajane *et al.* 2021; Negara *et al.* 2020; Nguyen *et al.* 2018; Pham *et al.* 2020; Saha *et al.* 2023). Meanwhile, He *et al.* (2022), Bui *et al.* (2017) and Safi and Bouroumi (2013) studied on relationship between the fire parameters and fire ignition using only one ML model. Among notable predictive models commonly used were Support Vector Machine (SVM), Artificial Neural Network (ANN) algorithms, Multilayer Perceptron (MLP), Random Forest (RF) and Decision Tree (DT). They gain attention due to their ability to process large datasets efficiently and adapt to the data patterns to generate relevant outputs. Despite their promise, it is crucial to acknowledge the existence of certain challenges within ML models, as claimed by Chew *et al.* (2022) and Jain *et al.* (2020).

These papers covered several countries' case studies and different prediction parameters were considered. Even Bui *et al.* (2017), Nguyen *et al.* (2018) and Pham *et al.* (2020) did not consider the same parameters despite being in the same country. Similar to Mohajane *et al.* (2021) and Purnama *et al.* (2024). Among common parameters were the normalized difference vegetation index (NDVI), slope and wind speed. The listed models in Table 1 delivered highly accurate prediction results but have a few limitations. For instance, unanalysable uncertainties issues (Pham *et al.* 2020), ML models' inadequacy to detect the fire causes (He *et al.* 2022; Negara et al., 2020) and false accuracy (Saha *et al.* 2023). Although there was no specific issue mentioned in Mohajane *et al.* (2021), this study recommended employing the frequency ratio (FR) models in similar situations. Based on Tale 1, none of the wildfire spots studied by these researchers experienced the exact parameters combination with each other so it was unlikely for the five proposed hybrid FR models to be generalised in other places. Some of these parameters used by wildfire researchers represented spatial considerations that could be roughly categorized into geographic (slope, elevation, aspect), demographic (distance) and vegetation (NDVI) (Iban & Sekertekin, 2022).

Despite a majority of wildfire studies in Table 1 being spatial modelling of wildfire prediction, none of them considered the spatial input as a fire ignition parameter other than their specific demographic and weather conditions. Only Safi and Bouroumi (2013) utilized the latitude and longitude values as part of wildfire causes but it has been outdated for more than a decade. Another recent wildfire study that included the coordinate data was Ghibeche *et al.* (2024). However, the coordinate factor in this study was meant for the dashboard development to pinpoint the wildfire location and exclude it from the predictive model. This study was not focused on identifying the prominent wildfire cause like most. Meanwhile, Phelps and Woolford (2021) compared several ML models together and used coordinate values as predictors in some of them. The models with spatial information outperformed others with a minor margin. As for Chen *et al.* (2024), they were interested in the generalisability of fire occurrence prediction in China. Even though the improvement was small, Phelps and Woolford (2021) and Chen *et al.* (2024) highlighted the importance of spatial factors in wildfire prediction studies (Jain *et al.* 2020) and sought generalisability (Chew *et al.* 2022).

Table 1: Forest fire prediction literature reviews.

Author(s), Place	Models	Parameters	Result/Recommendation
Safi and Bouroumi (2013), Montesinho Natural Park, Portugal	ANN-MLP	Coordinate, Time, Moisture, Drought, Initial Spread Index, Temperature, Humidity, Wind, Rainfall	Had a low error rate but was recommended for less sensitive ANN models.
Bui *et al.* (2017), Lam Dong, Vietnam	Particle Swarm Optimized Neural Fuzzy (PSO-NF)	**NDVI**, **Distance**, Slope, Aspect, Elevation Temperature, Land Cover, Wind, Rainfall	PSO-NF model surpassed the benchmark models. Recommended new algorithms for predictive improvement.
Nguyen *et al.* (2018) Thuan Chau, Vietnam	SVM, RF, MLP-Network	Slope, Aspect, **Elevation**, Curvature, **NDVI**, Distance, Land Use, Temperature, **Humidity**, Rainfall	All models were reliable and accurate but better optimisation methods were recommended.
Negara *et al.* (2020), Riau, Indonesia	DT, Bayesian Network (BN)	Time, **Temperature**, Dew Point **Humidity**, **Wind**, Pressure, **Precipitation Average**	BN model outperformed the DT model in predicting forest fire areas but was still unable to identify the wildfire causes.
Pham *et al.* (2020), Nghe An, Vietnam	BN, Naïve Bayes (NB), DT, Multivariate Logistic Regression	Slope, **Elevation**, Aspect, River Density, **Land Cover**, **Temperature**, Drought Index, **Distance**	All models were robust Unable to analyse the uncertainties in prediction outcomes.
Mohajane *et al.* (2021), Tanger-T´etouan-Al Hoceima, Morocco	5 Hybrid FR Models	Slope, Elevation, Aspect, Land Cover, **NDVI**, **Rainfall**, Temperature, **Wind**, **Distance**	All models had excellent predictive performance.
Iban and Sekertekin (2022), Adana and Mersin, Türkiye	LR, SVM, Linear Discriminant Analysis (LDA), Ensemble RF	**Elevation**, Slope, Aspect, Humidity, Land Cover, NDVI, Temperature, Rainfall, Radiation, Wind, Distance	All models delivered a satisfactory result but below 90% accuracy. Most correlations between fire causes and ignition were low.
He *et al.* (2022), New South Wales, Australia	ML-RF	Temperature, Precipitation, Soil Moisture, **Wind**, Flammability, NDVI Slope	The fire causes were detectable but had weak correlations with fire ignition.
Saha *et al.* (2023), Ayodhya hill, India	RF, Multivariate Adaptive Regression Splines, DLNN	**Altitude**, **Slope**, Curvature, Aspect, Temperature, Humidity, Wind, VI, Land Cover, **Distance**	Not entirely accurate due to the forest fire complexity in a real situation.
Purnama *et al.* (2024), Mediterranean - Türkiye	DT, NB, RF, ANN, SVM	**Elevation**, Aspect, **Slope**, **Land Cover**, Precipitation, **Temperature**, Wind, Humidity, Moisture, Demographic	RF was the best model with satisfactory overall predictive performance with not more than 80% results.
Ghibeche *et al.* (2024), Northern Algeria	K-Nearest Neighbour, DT, RF	Coordinate, Time, Temperature, Humidity, Wind, Pressure, Precipitation	RF was the best model to predict wildfire occurrence.

Nonetheless, the NPI with a parametric copula is anticipated to improve the existing models. The proposed method in this study generated prediction based on the frequency of the past location for wildfire recurrence. Mohajane *et al.* (2021) focusing on the same concept of frequency data in the hybrid models had performance results of approximately 90% indicating its potential in predicting wildfire susceptibility. Khosravi and Nahavandi (2013) were the first to explore the NPI potential in the nature field. They combined a similar concept of lower and upper prediction intervals (PIs) constructed by a nonparametric method with the NN model and performed better against the individual NN model to predict the wind power. Zhang (2023) and Bazionis and Georgilakis (2021) outlined researches succession (Kavousi-Fard *et al.* 2015; Quan *et al.* 2013; Shi *et al.* 2017; Wu *et al.* 2018) of the nonparametric PIs methods ever since that elevated their robustness and uncertainty. Zhang (2023) also emphasised the PIs importance of providing the necessary information for decision-making and management by further proposing a new PIs method that aligned with Khosravi and Nahavandi (2013) result. Since wind study has variability and uncertainty challenges like wildfire study, NPI may have the same potential to deliver better performance than the ML models for forest fire fields because of the nonparametric nature (Chen *et al.* 2024; Wang *et al.* 2021).

Accordingly, the researchers sought to explore the NPI application in a new branch to assess its predictive capabilities. Several researchers have discussed the practicality of NPI due to its attractiveness of using minimal assumptions in its probability predictions (Coolen-Schrijner & Coolen 2007; Coolen *et al.* 2011). NPI is suitable for real-world applications as it is flexible to predict the probability of the future situation. Particularly, an ambiguous situation like in wildfire that is full of uncertainty (Chen *et al.* 2024). These situations are more fit for bounded prediction that offers an acceptable range with multiple possible predicted values (Coolen *et al.* 2011; Mahtani 2019).

3. Research Methodology

3.1. *Data preparation*

The NPI with parametric copula utilized Indonesia's (2020) dataset with 16,201 observations of past forest fire hotspots. The longitude (x) and latitude (y) parameters were anticipated to be generalised to other countries too since the coordinate values are unique and independent with distinct places. Since this paper only focused on Indonesia, the limit for x and y were determined beforehand as [94°E, 141°E] and [-13°, 8°N], respectively. The Indonesia (2020) dataset was too large and time-consuming to process. Thus, the dataset was divided into Sumatra, Kalimantan, Java, Lesser Sunda, Sulawesi, Maluku and Papua based on the Indonesian archipelagoes as illustrated in Figure 1.

3.2. *Nonparametric predictive inference (NPI) framework*

NPI is a frequentist method that relies on the frequency of specific observations in past data. It constructs an imprecise probability that includes lower ($\underline{P}$) and upper ($\overline{P}$) probabilities of the next likely outcome. This imprecise probability is presented in the form of confidence interval, ($\underline{P}, \overline{P}$). Unlike other analyses that produce exact outputs, NPI provides two sets of probabilities to form a bound of acceptable probabilities. The gap within the ($\underline{P}, \overline{P}$) quantifies the uncertainties of the predicted value, thus providing confidence. NPI assumes that future real-valued random observation will have direct conditional probabilities, given the observed values. It emphasises the conditional probabilities from the ranked of the existing bivariate data with the partially specified predictive probability distribution of Eq. (1) and Eq. (2) (Muhammad, 2016; Muhammad *et al.* 2024).

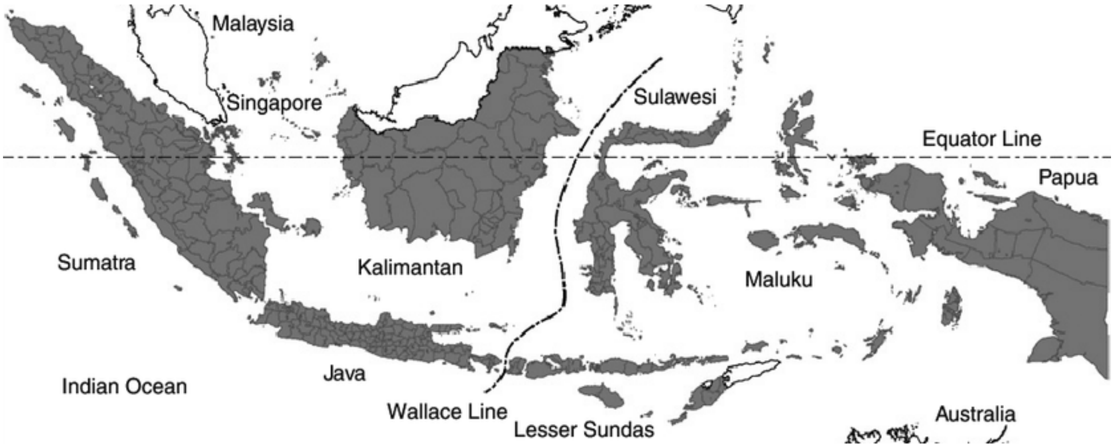


Figure 1: Indonesian archipelagoes map (Elyazar *et al.* 2011)

$$P\left(X_{n+1} \in \left(x_{i-1}, x_i\right)\right) = \frac{1}{n+1} \tag{1}$$

$$P\left(Y_{n+1} \in \left(y_{j-1}, y_j\right)\right) = \frac{1}{n+1} \tag{2}$$

for $i, j = 1, 2, \ldots, n+1$, $x_0, y_0 = -\infty$ and $x_{n+1}, y_{n+1} = \infty$, where, X_{n+1} and Y_{n+1} denote the future observation, x_i and y_j denote the ordered observations, and n denotes the total observations.

The parametric copula (θ) describes the dependence and inter-correlation between two parameters. This study concentrates on four parametric copula types that are commonly used to model the dependency of bivariate data. They are the Normal (θ_n), Clayton (θ_c), Frank (θ_f) and Gumbel (θ_g) copulae. Each has distinct features and proceeds differently with Kendall's tau (τ) values to come up with their respective cumulative distribution function (cdf). The maximum likelihood estimator (MLE) method estimated the copula parameters that would be integrated into the NPI framework. The copula element in the NPI contributed to the x and y correlation to improve the prediction results and consider the spatial correlation. This association is represented as a threshold (t) formula and may differ depending on the case study (Joe 2014; Muhammad 2016).

There are several ways to evaluate the NPI with parametric copula performance. One of them is the assessment of imprecision within the ($\underline{P}, \overline{P}$) where smaller imprecision indicates higher prediction accuracy. If the range between $\underline{P}$ and $\overline{P}$ is very close to each other, it can be inferred as the near-perfect prediction outcome. The imprecision refers to the difference ($\bar{d}$) between the $\underline{P}$ and $\overline{P}$ ($\bar{d} = \overline{P} - \underline{P}$). Based on the $\bar{d}$ value, the uncertainties in the forest fire likelihood of happening at the desired locations can be quantified. This quantification provides confidence in real-world applications to be relied on (Muhammad, 2016). Also, the comparison of the lowest $\bar{d}$ values among the copula types can identify the best one to incorporate with the NPI in predicting the next forest fire hotspot across Indonesia (Muhammad *et al.* 2024).

3.3. *Case study*

The study analysed Indonesia's (2020) dataset past coordinates to predict the next forest fire location. Indonesia's longitude (x) and latitude (y) are the exchangeable bivariate random quantities analysed in the NPI with the parametric copula. Firstly, the raw data were filtered based on the mentioned limits of the Indonesian subgroups earlier. Afterward, the (x, y) observations including the minimum and maximum limits of x and y were sorted separately in ascending order (x_i, y_j) depending on n of each bivariate parameter. The separately ordered observation is to ensure the dependence structure between x and y is properly modelled marginally. Referring to Eq. (1) and Eq. (2), the marginal NPI of the X_{n+1} and Y_{n+1} probabilities were computed.

Then, the (X_{n+1}, Y_{n+1}) were transformed into [0, 1] values as a connector to the dependence structure step later. This data transformation is denoted as $(\tilde{X}_{n+1}, \tilde{Y}_{n+1})$ and derived as $\left(\tilde{X}_{n+1} \in \left(\frac{i-1}{n+1}, \frac{i}{n+1}\right), \tilde{Y}_{n+1} \in \left(\frac{j-1}{n+1}, \frac{j}{n+1}\right)\right)$, for $i, j = 1, 2, \ldots, n+1$. Next, the copula parameters (θ_i) were estimated using the MLE method from the (x, y) pairs information and combined with the marginal NPI. A matrix decomposition of the cdf for the $(\tilde{X}_{n+1}, \tilde{Y}_{n+1})$ data was conducted to get the ranked probabilities (h_{ij}). Next, the summation of h_{ij} block followed Eq. (3) for lower probabilities and Eq. (4) for upper probabilities, corresponding to the unique t values to construct the $(\underline{P}, \overline{P})$ for the next forest fire hotspots prediction. The association of the copula between x and y or the threshold for this case study was $t = x_i + y_i$.

$$\underline{P}\left(T_{n+1} > t\right) = \sum_{i,j \in L_t} h_{ij}\left(\hat{\theta}\right) \tag{3}$$

for $L_t = \left\{(i, j) : x_{i-1} + y_{j-1} > t, \quad 1 \le i \le n+1, \quad 1 \le j \le n+1\right\}$

$$\overline{P}\left(T_{n+1} > t\right) = \sum_{i,j \in U_t} h_{ij}\left(\hat{\theta}\right) \tag{4}$$

for $U_t = \left\{(i, j) : x_i + y_j > t, \quad 1 \le i \le n+1, \quad 1 \le j \le n+1\right\}$ where, T_{n+1} denotes the future association threshold, L_t denotes the lower probability that exceeds the particular t, and U_t denotes the upper probability that exceeds the particular t.

4. Analysis and Discussion

Figure 2 illustrates the frequency scatterplot of wildfire occurrences across Indonesia in 2020. Every Indonesian archipelago has a forest fire history with a repetition frequency of approximately 700 times at certain same spots. Kalimantan is experiencing the most wildfire recurrence than others, especially in the West and East Kalimantan provinces. Besides that, Sumatra Riau and East Java have a high recurrence rate for 2020. Meanwhile, the least forest fire recurrence was Papua and Maluku archipelagoes except for the South Papua region. Even for them, the recurrence frequency was still around 20 times. Generally, the forest fire repetition in the same spot for Indonesia decreased when moving to the east.

Figure 2 proved that the repetition rate is unnecessarily equivalent to the severity of the forest fire disaster. A reference from Global Forest Watch (2024), the top four Indonesian provinces for the most tree cover loss from wildfires in 2020 were South Sumatra, Jambi, Central Kalimantan and West Kalimantan consecutively. Concurrently, each province lost as much as 14.5kha, 14.3kha, 11.7kha and 8.25kha of tree cover. Even so, South Sumatra, Jambi and Central Kalimantan had repeated events less than 20 times despite their highest statistics. These provinces were mostly covered in much darker tone markers and only a few spots of light green and yellow markers. Regardless, there was a situation like West Kalimantan that recorded high tree cover loss and high frequency of repetition simultaneously, which were filled with lighter tone markers.

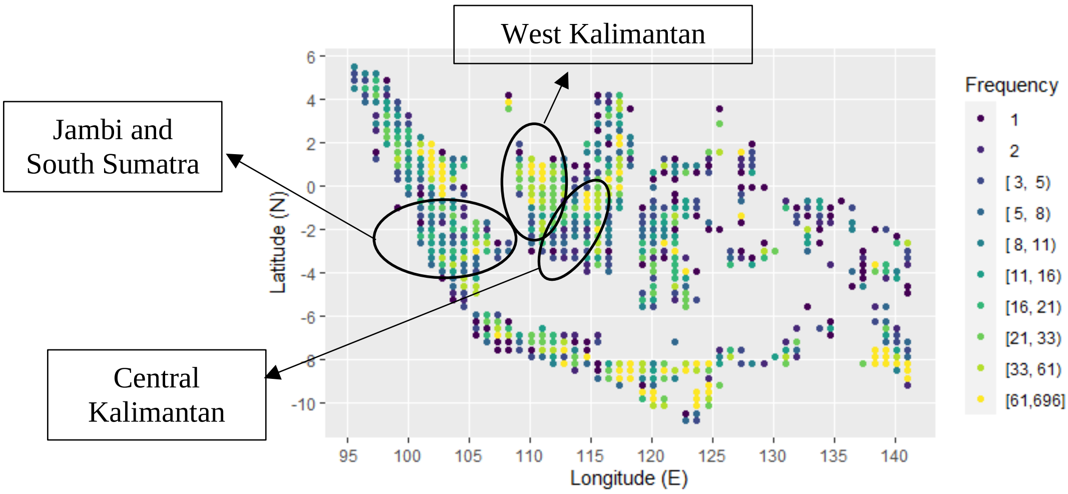


Figure 2: Frequency scatter plot of forest fire location in Indonesia for 2020

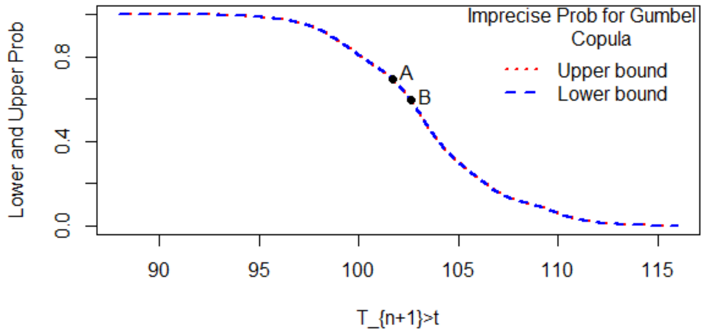
Only the Sumatra and Kalimantan groups were discussed in this paper due to their highest tree cover loss from the Indonesia wildfires in 2020. Also, the selection of the two groups was to analyse the semiparametric analysis generalisability ability in other regions. Although Sumatra and Kalimantan are both Indonesian archipelagoes, they are separate lands with different environments. Furthermore, the Kalimantan archipelago is shared with Malaysian Borneo and Brunei and could constitute a 'different country' due to its similarity of demographics, specifically with Malaysian Borneo. Also, other $(\underline{P}, \overline{P})_i$ for Sumatra and Kalimantan have the same structures, so only θ_g would be focused on this paper for easy explanation. θ_g was chosen due to its low $\bar{d}$ within the $\left(\underline{P}, \overline{P}\right)_g$ for the majority of both lands. Apart from θ_c, other θ_i values were different for Sumatra and Kalimantan as listed in Table 2. The Indonesia's (2020) dataset had a $\theta_c = 1.6246 \text{ x } 10^{-4}$. These values were integrated with the NPI framework to generate imprecise probabilities based on the appropriate groups.

Table 2: Copula parameter values for Indonesia groups

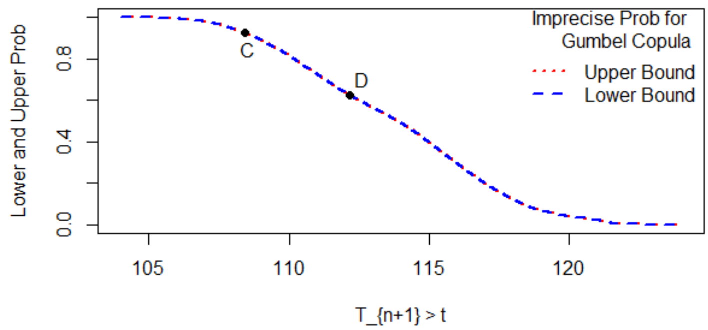
Archipelago	θ_n	θ_g	θ_f
Kalimantan	-0.0282	1.0649	0.1442
Sumatra	-0.4666	1.0002	-3.9542

4.1. *NPI: imprecise probability location prediction*

Figure 3 shows the $(\underline{P}, \overline{P})_g$ on the forest fire occurrences on the next observation appropriately with the t locations of Sumatra (Figure 3[a]) and Kalimantan (Figure 3[b]). It seemed that the $\underline{P}$ and $\overline{P}$ were overlapped with each other which depicted the small imprecision between the $(\underline{P}, \overline{P})_g$ for both groups. Subsequently, the NPI with parametric copula can generate a highly accurate imprecise probability when predicting the likelihood of the next forest fire occurrence in the Sumatra and Kalimantan archipelagoes. Additionally, the points outlined in Figure 3 signified the region representatives for South Sumatra, Jambi, Central Kalimantan and West Kalimantan provinces. From Figure 3(a), Point A was Ogan Komering Ilir (OKI) (t = 101.6826°) from South Sumatra province and Point B was Muaro Jambi ($t = 102.6092°$) from Jambi province. Accordingly, Point C and Point D from Figure 3(b) each depicted Ketapang, West Kalimantan Barat ($t = 108.4449°$) and Kapuas, Central Kalimantan ($t = 112.1460°$).



(a)



(b)

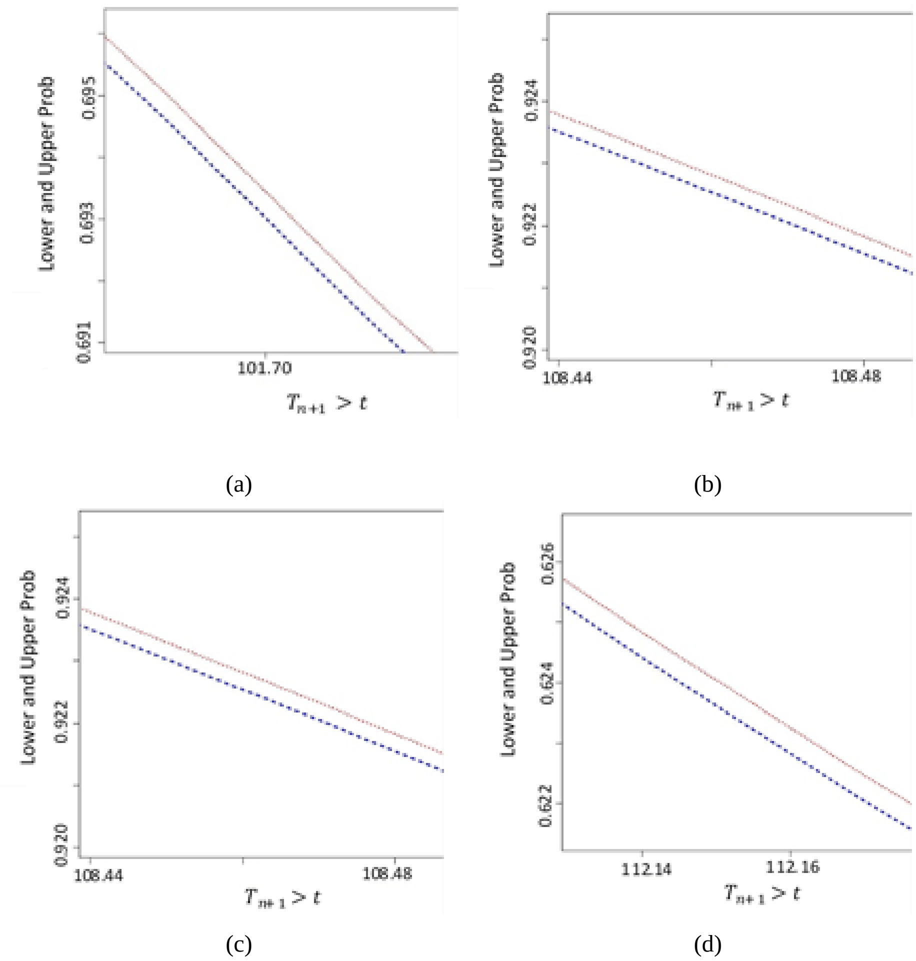
Figure 3: Imprecise probability of gumbel copula for Sumatra and Kalimantan

Table 3: Gumbel imprecise probability interpretations

Point	$(\underline{P}, \overline{P})_g$	Interpretations
Point A	(0.6946, 0.6950)	The next wildfire event is 69.46% to 69.5% will occur at Ogan Komering Ilir, South Sumatra
Point B	(0.5942, 0.5947)	The next wildfire event is 59.42% to 59.47% will occur at Muaro Jambi, Jambi
Point C	(0.9233, 0.9236)	The next wildfire event is 92.33% to 92.36% will occur at Ketapang, West Kalimantan
Point D	(0.6239, 0.6244)	The next wildfire event is 62.39% to 62.44% will occur at Kapuas, Central Kalimantan

Figure 4 is the $(\underline{P}, \overline{P})_g$ zoom-in for Point A, Point B, Point C and Point D respectively starting from Figure 4(a) to Figure 4(d). Based on Figure 4, there was a clear difference between the $\underline{P}$ and $\overline{P}$ for all points despite their overlay appearances in Figure . The close gap reflected the semiparametric analysis predictive performance for Sumatra and Kalimantan locations, which can be inferred as near-perfect predictions. Thus, Table 3 compiled the interpretations of $(\underline{P},\overline{P})_g$ for all points.

Out of all $(\underline{P},\overline{P})_g$, Point C had the highest values. This reflected the NPI properties which relied on frequency to generate the outcome. Since West Kalimantan experienced frequent recurrence of forest fire events in the same spots, its $(\underline{P},\overline{P})_g$ is anticipated to be higher than in other provinces. As evidenced by Table 3, the NPI frequentist properties generated a high $(\underline{P},\overline{P})_g$ for Point C but moderately lower $(\underline{P},\overline{P})_g$ for others due to their less forest fire repetition history at South Sumatra, Jambi and Central Kalimantan. Nevertheless, all points have minimal $\bar{d}$ values despite the different $(\underline{P},\overline{P})_g$ values.



(a) (b) (c) (d)

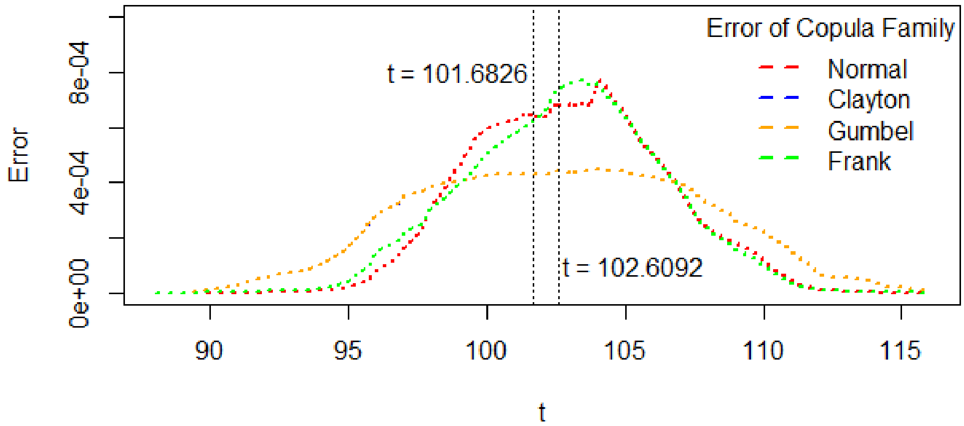
Figure 4: Imprecise probability zoom-in of gumbel copula for Sumatra and Kalimantan

4.2. *NPI: performance evaluation*

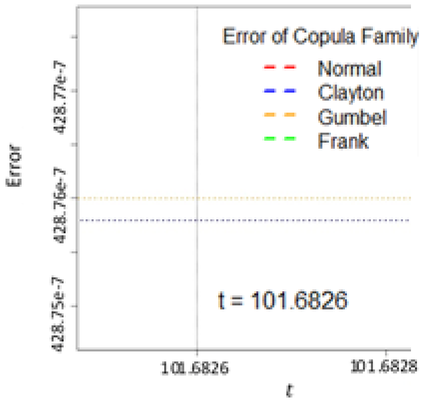
The near-gap as shown in Figure 4 were the $\bar{d}_g$ values and were quantified in Table 4 along with other $\bar{d}_i$ values for comparison. The bold font in Table 4 was denoted as the lowest value for each point. Hence, the optimal copula that gave the smallest imprecision when integrated with the NPI for Point B and Point D is θ_g, for Point A was θ_c and for Point C was θ_n. Figure 5 and Figure 6 supported these numerical results with the line graph of the $\bar{d}_i$ for all unique t locations. Graphical summaries provided additional information like the general copula behaviour with different imprecision values. Stemming from these, the t could be segmented into the relevant locations.

Table 4: Imprecise probability differences of copula family

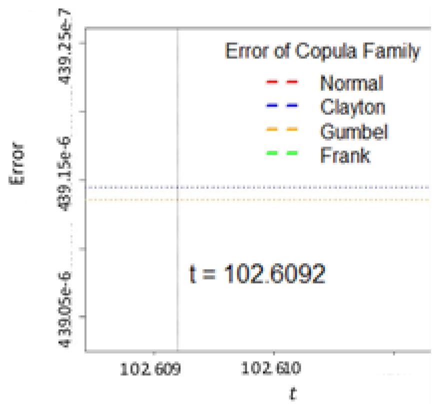
Imprecision ($\times 10^{-4}$)	$\bar{d}_n$	$\bar{d}_c$	$\bar{d}_g$	$\bar{d}_f$
Point A	6.4899	4.2876	4.2877	6.2493
Point B	6.7970	4.3914	4.3914	7.3748
Point C	2.5825	2.6310	2.7694	2.6829
Point D	4.5457	4.4585	4.2704	4.3881



(a)



(b)

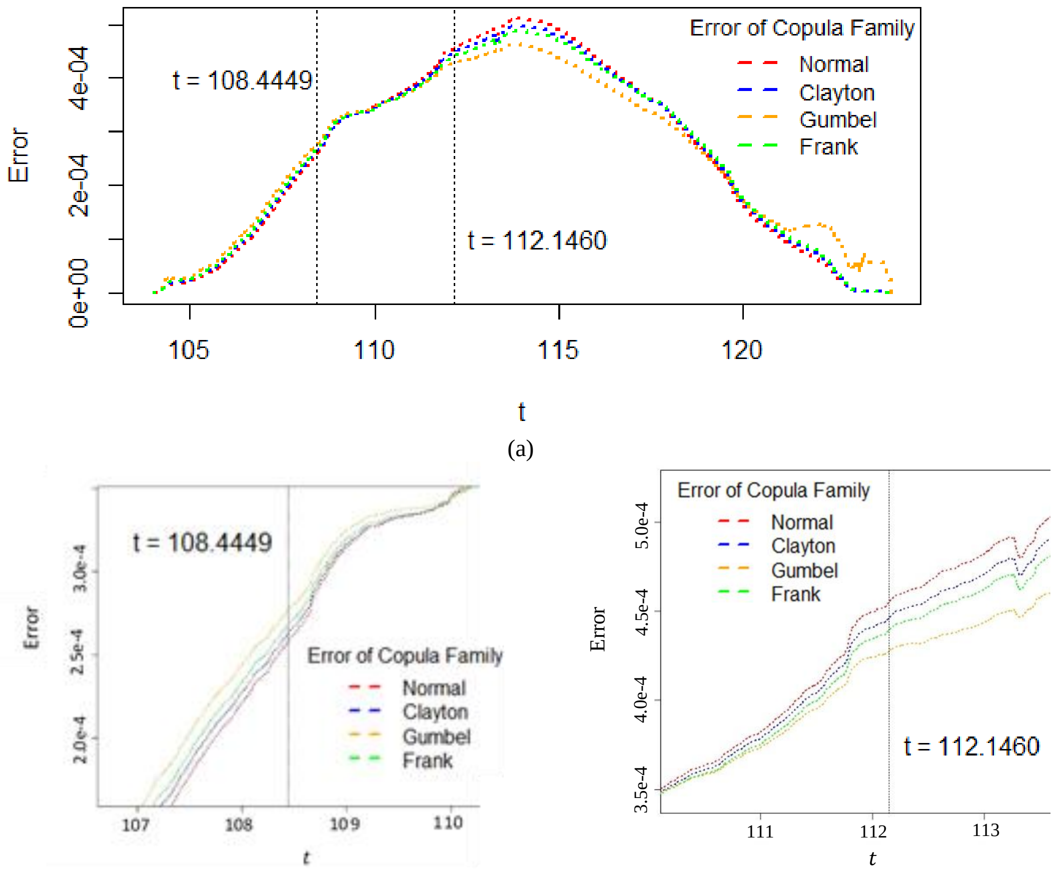


(c)

Figure 5: Line graph of imprecise probability differences for Sumatra

Figure 5(b) shows the zoom-in of $\bar{d}_i$ line graphs for Point A, while Figure 5(c) for Point B. All θ_i followed certain patterns in pairs and peaked around the same range of t. The line graph of $\bar{d}_n$ was aligned with $\bar{d}_f$ line graph, meantime, the $\bar{d}_g$ and $\bar{d}_c$ lines seem to coincide with each other in Figure 5(a). However, the $\bar{d}_g$ and $\bar{d}_c$ lines were distinct from each other as portrayed in the zoom-in. Their role as the optimal copula for Sumatra locations was alternated throughout the graphs but at Point A, $\bar{d}_c$ line appeared to be at the bottom while at Point B, $\bar{d}_g$ line was at the bottom, indicating their lowest values when compared to other $\bar{d}_i$.

Based on Figure 5(a), the highest $\bar{d}_i$ for each copula type was when the t values were between 99° to 104° and declined afterward. Accordingly, that range signified the majority of Sumatra land from Acheh to Lampung. As for the remaining Sumatra land was represented by t values of [105°, 106°). These areas are coastal regions which were the small islands between Riau and Malaysia.



(a)

(b)

(c)

Figure 6: Line graph of imprecise probability differences for Kalimantan

Figure 6(b) and Figure 6(c) were the zoom-in for $\bar{d}_i$ line graphs for Point C and Point D of the Kalimantan province. Unlike Sumatra, Figure 6(a) illustrated that all $\bar{d}_i$ lines follow closely to one another with minor deviation. Except for $\bar{d}_g$ line at the most right where it became slightly irregular and higher than the other $\bar{d}_i$ lines. This is due to the θ_g traits that possessed a strong right-tail dependence. It was expected that the h_{ij} was directly proportional to the dependence strength. A high h_{ij} tends to produce high $\bar{d}_g$. The zoom-in of Point C and Point D displayed the lowest $\bar{d}_i$ lines which were the $\bar{d}_n$ and $\bar{d}_g$ lines.

Figure 6(a) showed that the highest $\bar{d}_i$ for Kalimantan province when the t values were between 111° to 114°. Then, the $\bar{d}_i$ lines steadily decreased until the end. The highest $\bar{d}_i$ lines depicted the less mountainous areas across West, Central and South Kalimantan. Threshold values of [111°, 115°) are the West and Central Kalimantan regions that had high repeated wildfire events as shown in Figure. While $115° \leq t < 122°$ is the mountainous areas at East and North Kalimantan with occasional wildfire events. The western part of West Kalimantan, which consists of coastal areas and human settlements is $108° \leq t < 111°$.

5. Conclusion

In conclusion, this study suggested a new method for predicting forest fire hotspots. The increasing trends of wildfires in recent years have motivated the related parties to look for alternatives in forest fire detection. Even though the ML method proved to be useful in forest fire prediction, it still grappled with inaccuracy risk and was unable to generalise. Therefore, NPI with a parametric copula was proposed to work out these issues. Based on the outcome, the proposed method could generate a highly accurate $(\underline{P}, \overline{P})$ for all θ_i with minimal $\bar{d}$ (<0.001). The results indicated the near-perfect probability prediction of the next forest fire location for Sumatra land. Kalimantan archipelago had the same predictive performances as Sumatra indicating the NPI's ability to generalize its performance in other regions.

Besides, the t segmentations of Sumatra, [99°, 106°) and Kalimantan, [108°, 122°) were identified in this study along with information regarding the severity and recurrence of Indonesia forest fires. These insights aid forest firefighters in their decision-making process as references in developing crisis planning and appropriate policies based on their jurisdictional areas. They can minimize loss while maximizing their available resources efficiently in a shorter time and are more confident due to the small imprecision for forest fire hotspot prediction in both archipelagoes. Plus, the optimal θ_i were finalized in this paper to be integrated with the NPI framework to predict the best $(\underline{P}, \overline{P})_i$ for the next wildfire events at Sumatra (θ_c, θ_g) and Kalimantan (θ_n, θ_g) locations. Though located in the same archipelago, all focused hotspots had different optimal θ_i that enlighten the different wildfire trends across extensive Indonesian lands.

This study has a few limitations, but the critical one was the processing time in delivering the result. Due to the large data size, it took a few days to generate the $(\underline{P}, \overline{P})_i$, even after the Indonesia dataset had been segmented into subgroups. This situation is not ideal for daily prediction or in shorter time intervals. Nevertheless, the processing time can be reduced by further filtering the data appropriately with the related parties' jurisdiction areas, specific time intervals and specific θ_i type. This study only discussed two archipelagoes due to limited time and may be insufficient to understand more about the Indonesia forest fire trend. This detail can be investigated further in future research by implementing the proposed method on other archipelagoes. Moreover, NPI can be applied in other countries as well to inspect deeper on its replicability performance in various geographic and weather conditions. Hence, the NPI with parametric copula can be a reliable alternative for real-world circumstances, befitting the information presented in this paper.

Acknowledgments

The authors thank Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) for providing financial support under the International Matching Grant UMP RDU222706 (UIC221519) and the UMPSA facilities. The authors also would like to thank the reviewers for the valuable comments for improvements to this paper.

References

Bazionis I.K. & Georgilakis P.S. 2021. Review of deterministic and probabilistic wind power forecasting: Models, methods, and future research. *Electricity* **2**(1): 13-47.

Bui D.T., Bui Q.T., Nguyen Q.P., Pradhan B., Nampak H. & Trinh P.T. 2017. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. *Agricultural and Forest Meteorology* **233**: 32-44.

Chen Z., Zhang C., Li W., Gao L., Liu L., Fang L. & Zhang C. 2024. Fire danger forecasting using machine learning-based models and meteorological observation: A case study in Northeastern China. *Multimedia Tools and Applications* **83**(22): 61861-61881.

Chew Y.J., Ooi S.Y., Pang Y.H. & Wong K.S. 2022. A review of forest fire combating efforts, challenges and future directions in Peninsular Malaysia, Sabah, and Sarawak. *Forests* **13**(9): 1405.

Coolen-Schrijner P. & Coolen F.P.A. 2007. Nonparametric adaptive age replacement with a one-cycle criterion. *Reliability Engineering & System Safety* **92**(1): 74-84.

Coolen F.P., Troffaes M.C. & Augustin T. 2011. Imprecise probability. In Lovric M. (ed.). *International Encyclopedia of Statistical Science*: 645-648. Berlin: Springer Verlag.

Elyazar I.R., Hay S.I. & Baird J.K. 2011. Malaria distribution, prevalence, drug resistance and control in Indonesia. *Advances in Parasitology* **74**: 41-175.

Ghibeche Y., Sellam A., Nouri N., Khaldi A., Harrane A. & Ghibeche I. 2024. Machine learning for forest fire prediction: A case study in North Algeria. *Ingénierie des Systèmes d'Information* **29**(1): 337-346.

Global Forest Watch. 2024. *Global*. https://www.globalforestwatch.org/dashboards/global/ (1 February 2024).

Gong A., Li J. & Chen Y. 2021. A spatio-temporal brightness temperature prediction method for forest fire detection with MODIS data: A case study in San Diego. *Remote Sens* **13**(15): 2900.

He W., Shirowzhan S. & Pettit C.J. 2022. GIS and machine learning for analysing influencing factors of bushfires using 40-year spatio-temporal bushfire data. *ISPRS International Journal of Geo-Information* **11**(6): 336.

Iban M.C. & Sekertekin A. 2022. Machine learning based wildfire susceptibility mapping using remotely sensed fire data and GIS: A case study of Adana and Mersin provinces, Turkey. *Ecological Informatics* **69**: 101647.

Jain P., Coogan S.C., Subramanian S.G., Crowley M., Taylor S. & Flannigan M.D. 2020. A review of machine learning applications in wildfire science and management. *Environmental Reviews* **28**(4): 478-505.

Joe H. 2014. *Dependence Modeling with Copulas*. Florida, USA: CRC press.

Kavousi-Fard A., Khosravi A. & Nahavandi S. 2015. A new fuzzy-based combined prediction interval for wind power forecasting. *IEEE Transactions on Power Systems* **31**(1): 18-26.

Khosravi A. & Nahavandi S. 2013. Combined nonparametric prediction intervals for wind power generation. *IEEE Transactions on Sustainable Energy* **4**(4): 849-856.

Mahtani A. 2019. Imprecise probabilities. In Pettigrew R. & Weisberg J. (eds.). *The Open Handbook of Formal Epistemology*: 107–130. PhilPapers Foundation.

Miettinen J., Shi C. & Liew S.C. 2017. Fire distribution in Peninsular Malaysia, Sumatra and Borneo in 2015 with special emphasis on peatland fires. *Environmental Management* **60**: 747-757.

Mohajane M., Costache R., Karimi F., Pham Q.B., Essahlaoui A., Nguyen H., Laneve G. & Oudija F. 2021. Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area. *Ecological Indicators* **129**: 107869.

Muhammad N. 2016. Predictive inference with copulas for bivariate data. PhD Thesis. Durham University.

Muhammad N., Roslin A.H., Daud H., Kadir E.A. & Maharani W. 2024. Predicting forest fire spots using nonparametric predictive inference with parametric copula: Malaysia case study. *AIP Conference Proceedings* **3189**(1): 100003.

Negara B., Kurniawan R., Nazri M., Abdullah S., Saputra R. & Ismanto A. 2020. Riau forest fire prediction using supervised machine learning. *Journal of Physics: Conference Series* **1566**(1): 012002.

Nguyen N.T., Dang B.T.N., Pham X.C., Nguyen H.T., Bui H.T., Hoang N.D. & Tien B.D. 2018. Spatial pattern assessment of tropical forest fire danger at Thuan Chau area (Vietnam) using GIS-based advanced machine learning algorithms: A comparative study. *Ecological* informatics **46**: 74-85.

Pham B.T., Jaafari A., Avand M., Al-Ansari N., Dinh Du T., Yen H.P.H., Phong T.V., Nguyen D.H., Le H.V., Mafi-Gholami D. Prakash I, Thuy H.T. & Tuyen T.T. 2020. Performance evaluation of machine learning methods for forest fire modeling and prediction. *Symmetry* **12**(6): 1022.

Phelps N. & Woolford D.G. 2021. Comparing calibrated statistical and machine learning methods for wildland fire occurrence prediction: a case study of human-caused fires in Lac La Biche, Alberta, Canada. *International Journal of Wildland Fire* **30**(11): 850-870.

Purnama M.I., Jaya I.N.S., Syaufina L., Çoban H.O. & Raihan M. 2024. Predicting forest fire vulnerability using machine learning approaches in The Mediterranean Region: a case study of Türkiye. *IOP Conference Series: Earth and Environmental Science* **1315**(1): 012056.

Quan H., Srinivasan D. & Khosravi A. 2013. Short-term load and wind power forecasting using neural network-based prediction intervals. *IEEE Transactions on Neural Networks and Learning Systems* **25**(2): 303-315.

Safi Y. & Bouroumi A. 2013. Prediction of forest fires using artificial neural networks. *Applied Mathematical Sciences* **7**(6): 271-286.

Saha S., Bera B., Shit P.K., Bhattacharjee S. & Sengupta N. 2023. Prediction of forest fire susceptibility applying machine and deep learning algorithms for conservation priorities of forest resources. *Remote Sensing Applications: Society and Environment* **29:** 100917.

Shi Z., Liang H. & Dinavahi V. 2017. Direct interval forecast of uncertain wind power based on recurrent neural networks. *IEEE Transactions on Sustainable Energy* **9**(3): 1177-1187.

Wang Y., Zou R., Liu F., Zhang L. & Liu Q. 2021. A review of wind speed and wind power forecasting with deep neural networks. *Applied Energy* **304**: 117766.

Wu Y.K., Su P.E., Wu T.Y., Hong J.S. & Hassan M.Y. 2018. Probabilistic wind-power forecasting using weather ensemble models. *IEEE Transactions on Industry Applications* **54**(6): 5609-5620.

Yu M., Yang C. & Li Y. 2018. Big data in natural disaster management: A review. *Geosciences* **8**(5): 165.

Zhang X. 2023. Developing a hybrid probabilistic model for short-term wind speed forecasting. *Applied Intelligence* **53**(1): 728-745.

Centre for Mathematical Sciences,
Universiti Malaysia Pahang Al-Sultan Abdullah
Lbh Persiaran Tun Khalil Yaakob,
Kampung Melayu Gambang,
26300 Kuantan,
Pahang, MALAYSIA
E-mail: mirahazwa@gmail.com, noryanti@umpsa.edu.my[*]

Department of Informatics Engineering,
Universitas Islam Riau
Jl. Kaharuddin Nst No.113,
Simpang Tiga, Kec. Bukit Raya,
Kota Pekanbaru, Riau 28284,
INDONESIA
E-mail: evizal@eng.uir.ac.id

School of Computing,
Telkom University
Jl. Telekomunikasi No. 1,
Bandung Terusan Buahbatu - Bojongsoang, Sukapura,
Kec. Dayeuhkolot, Kabupaten Bandung,
Jawa Barat 40257,
INDONESIA
E-mail: wmaharani@telkomuniversity.ac.id

Department of Applied Mathematics,
Universiti Teknologi Petronas,
Persiaran UTP,
32610 Seri Iskandar,
Perak, MALAYSIA
E-mail: hanita_daud@utp.edu.my

Received: 14 May 2024
Accepted: 24 December 2024

[*]Corresponding author