JOURNAL OF STATISTICAL MODELING & ANALYTICS (JOSMA)
(ISSN: 2180-3102)
UNIVERSITI MALAYA

Modeling the Impact of Weather Variability on Paddy Yield in Malaysia using Copula Methods

Nurul Syuhada Samsudin[1] , Siti Rohani Mohd Nor[1*] , Shariffah Suhaila Syed Jamaludin[1] and Siti Mariam Norrulashikin[1]

[1]*Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia*

**Corresponding author: sitirohani@utm.my*

RESEARCH ARTICLE

Received 4 July 2025
Accepted 13 Jan 2025
Published 9 May 2026

Abstract

Paddy is an important crop in Malaysia, with the country producing more than 2.43 million tons of paddy annually, making it a key contributor to national food security and the economy. Rice, the staple food for most of the population, also serves as a primary source of income for many, particularly small-scale farmers. Despite its significance, extreme weather continues to negatively impact paddy yields and threatens future production. Traditional linear models often fail to capture the complex and nonlinear dependencies between weather extremes and agricultural output, resulting in limited predictive accuracy. To address this gap, this study aims to examine the effects of weather indices on paddy yield and rice production. It utilizes annual data from 1963 to 2022, focusing on temperature and rainfall as the primary indicators of weather that disrupts paddy cultivation. The copula method is employed to analyze the relationship between extreme weather and both paddy yield and rice production. This method is useful for measuring the dependency structure between the variables. The findings indicate that the Frank Copula provides the best fit among the tested copula families, producing the lowest Akaike Information Criterion (AIC) value, suggesting a significant dependence between weather extremes and paddy production. These results highlight the importance of accurately modelling weather and yield relationships to support food security planning in Malaysia.

Keywords: Copula Method, Weather Indicator, Food Security, Paddy

1. Introduction

Rice is one of the most vital staple crops globally, providing nourishment to nearly half of the world's population, especially across Asia. As reported by the Food and Agriculture Organization of the United Nations (FAO), the average annual rice consumption in Asia has been around 85 kg per person in recent years, significantly higher than Europe's average of just 6 kg per person. In Malaysia, agriculture plays a vital role as a primary source of food supply for the population. However, since the implementation of the Eighth Malaysian Plan in 2001, there has been a noticeable decline in focus on agricultural development. As a result, the country has become increasingly dependent on food imports. By 2015, Malaysia imported food worth RM45 billion, which represented approximately 20% of the national budget that year (Lee & Baharuddin, 2018), indicating a lack of self-sufficiency in food production.

In addition, the rapid pace of urbanization has led to the conversion of agricultural land into residential and commercial areas. In 2016, 75% of Malaysia's population was classified as urban, leaving only 25% in rural areas (The World Bank, 2016). Fruit trees and other crops have been cleared to make way for urban development, further reducing the nation's agricultural capacity. This continued loss of agricultural land poses a long-term risk to Malaysia's food security, especially as domestic crop production must compete with the expanding urban infrastructure.

Malaysia, on the other hand, is more vulnerable to climate variability as the country has experienced warm and inconsistent rainfall, particularly in the last two decades (Tang, 2019). Climate change is defined as long-term changes in temperature, rainfall, or precipitation patterns (NASA, 2024) and it has a direct impact on agriculture. As a traditionally agrarian country, Malaysia's agriculture sector is particularly susceptible to the effects of climate change. Therefore, Malaysia needs to assess the causes of climate change and implement effective strategies to mitigate its impact, to ensure the long-term sustainability of the agricultural sector in the future. Hence, the objective of this study is to evaluate the impacts of climate-related factors, particularly temperature and rainfall, on paddy and rice production in Malaysia and to provide insights that can support strategies for ensuring the sustainability of the agricultural sector.

Rainfall and temperature are key climate change factors that have been linked to crop yields. In regions that rely on rain-fed agriculture and experience stable temperatures, climate change plays a crucial role in determining whether paddy crops can grow successfully through to harvest. This also indicates the stability of paddy production, which is the staple food source in Malaysia. Understanding the relationship between crop yield and climate change is very important. To address this issue, the study employs a copula approach because copulas allow flexible modeling of the dependency structure between temperature, rainfall, and paddy yield, without requiring strict assumptions about their marginal distributions (Ribeiro et al., 2019). This makes copulas particularly suitable for capturing complex, nonlinear, and asymmetric dependencies that traditional correlation methods cannot adequately represent.

Copula modeling, also referred to as dependence modeling, is based on a core principle (Stamatatou et al., 2018). It separates a joint distribution function into two distinct components: one that represents the marginal (or univariate) behaviour and another that captures the dependence structure between variables. The copula function serves to link these marginal distributions while maintaining their individual properties and quantifying the dependence between them, without requiring a predefined joint distribution. Instead, the marginal distributions are transformed into uniform distributions through the probability integral transformation, after which their joint dependence is modelled using a copula function (Vuolo, 2017).

2. Materials and Methods

Copulas, introduced by Sklar in 1959 , are functions used to connect univariate marginal distributions to construct multivariate distribution functions. One key advantage of using copulas in multivariate distribution construction is their ability to simplify complex joint distribution modelling by focusing on the relationships and dependencies between correlated random variables, each with its own specific univariate marginal distribution (Shiau, 2006). Copulas can be classified into several families, including Archimedean and Elliptical copulas.

This study will use the Archimedean copula family, consisting of Clayton, Gumbel and Frank copulas and the Elliptical copula family, which are Student's and Gaussian copulas, to study the dependence between crop yield, namely paddy and rice, and climate variables, namely temperature and rainfall.

2.1 Archimedean Copulas

Copula with dimension $d \geq 2$ is called Archimedean if there is a generator φ hence for every $u = (u_1, \ldots, u_d) \in I^d$

$$C(u) = \varphi^{-1}.\left(\varphi(u_1)+..+\varphi(u_d)\right) \tag{1}$$

where φ is a generator function from Eq. (1) where $\varphi(0) = \infty$ and $\varphi(1) = 0$.

The Clayton copula in Eq. (2) is applied to get the positive dependency of a bivariate variable in the lower tail with $0 \leq \theta \leq \infty$. The formula follows by:

$$c(u_1, u_2) = ({u_1}^{-\theta} + {u_2}^{-\theta} - 1)^{-\frac{1}{\theta}} \tag{2}$$

The Gumbel copula parameter, given in Eq. (3) reflects the strength of positive dependence and indicates the degree of upper tail dependence, where $1 \leq \theta < \infty$. The Gumbel copula is expressed as follows,

$$c(u_1, u_2) = e^{-[(-ln u_1)^{\theta}+(-ln u_2)^{\theta}]^{\frac{1}{\theta}}} \tag{3}$$

Besides, the Frank copula, classified as a symmetric copula, captures positive dependence when the parameter lies between $(0, +\infty)$, negative dependence when the parameter is within $(-\infty, 0)$, and indicates independence when the parameter equals zero. Its formula is defined in Eq. (4) as

$$c(u_1, u_2) = -\frac{1}{\theta} ln\left[1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{(e^{-\theta}-1)}\right] \tag{4}$$

2.2 Elliptical Copulas

Elliptical copulas are derived from elliptical distributions of d dimensional random vectors (Dewick et al., 2023). In practice, time series data, such as crop yields and climate variables, may be stabilized using log-difference transformation to ensure approximate stationarity before applying copula models. Therefore, this study will utilize the elliptical copula family, specifically the Student's T copula and the Gaussian copula, to examine the relationship between climate change, crop yield and crop production.

The student's t copula is well-suited for modeling distributions with fat tails. It is defined as

$$C_{\Theta}(u_1, \ldots, u_n) = t_{v,\Sigma}[t_v^{-1}(u_1), \ldots, t_v^{-1}(u_n)] \tag{5}$$

where $\Theta = \{(\mathrm{v}, \Sigma): \mathrm{v} \in (1, \infty), \Sigma \in \mathcal{R}^{n \times n}$. Here, t_v^{-1} in Eq. (5) denotes the inverse of the univariate *t*-distribution with v degrees of freedom, and Σ represents the correlation matrix.

Meanwhile, the Gaussian copula given in Eq. (6) is commonly employed as a financial indicator to measure tail dependence. It is defined as

$$C(u_1, \ldots, u_n) = \Phi_{\delta}[\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)] \tag{6}$$

where Φ denotes the cumulative distribution function (CDF) of the standard normal distribution N (0,1) and δ represents the correlation coefficient between the random variables.

2.3 Marginal Distribution

Prior to modeling the dependence structure using copulas, appropriate univariate marginal distributions were fitted to each variable. This study has considered Normal, Gamma, and Log-Normal distributions to be included, which are commonly used for agricultural and climatic data.

The Normal distribution is appropriate for variables that are approximately symmetric and continuous. The formula for the normal distribution is described in Eq. (7) with $X \sim N(\mu, \sigma^2)$ is as follows,

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), -\infty < x < \infty \tag{7}$$

where μ is the mean and σ^2 is the variance.

The Gamma distribution is suitable for non-negative, skewed variables such as rainfall or other quantities that cannot take negative values. The formula of gamma in Eq. (8), where $X \sim Gamma\ (\alpha, \beta)$ is given by,

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x > 0 \tag{8}$$

where $\alpha > 0$ is the shape parameter, $\beta > 0$ is the rate parameter, and $\Gamma(\alpha)$ is a gamma function.

The log-normal distribution as stated in Eq. (9) is used for positively skewed variables where the logarithm of the variables is approximately normally distributed. The formula of $X \sim LogNormal\ (\mu, \sigma^2)$ is as follows,

$$f_X(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right), x > 0 \tag{9}$$

where μ and σ^2 are the mean and variance of $\ln X$.

The best-fitting marginal distribution for each variable is determined using the Akaike Information Criterion (AIC), which provides a balance between goodness-of-fit and model complexity. Selecting appropriate marginal distributions is essential because copulas separate the modeling of marginal behavior from the dependence structure between variables, ensuring that the copula accurately captures the relationships among variables.

3. Results and Discussion

In Malaysia, paddy has emerged as one of the key agricultural commodities, following palm oil and rubber (Firdaus et al., 2020; Dorairaj & Govender, 2023). It is cultivated twice a year, across two distinct cropping seasons. The main season, which runs from August to February during the humid period, typically operates without reliance on irrigation systems. Conversely, the off-season, from March to July during drier weather, requires irrigation support to sustain paddy growth.

In addition, paddy cultivation zones are strategically located throughout Peninsular Malaysia, Sabah, and Sarawak. Among the major rice-producing regions are the Muda Agricultural Development Authority (MADA), Kemubu Agricultural Development Authority (KADA), North Terengganu Integrated Agricultural Development (KETARA), Project Barat Laut Selangor (PBLS), along with the areas of Krian, Seberang Perak, Seberang Perai, Kemasin, Rompin, Kota Belud, and Batang Lupar. Each of these key zones encompasses more than 4,000 hectares and is supported by well-developed irrigation infrastructure to facilitate extensive rice production (Dorairaj and Govender, 2023).

Beyond these main granaries, the country also manages 74 secondary and 172 minor paddy-growing areas, which contribute an additional 28,441 hectares and 47,653 hectares of cultivated land, respectively (Rahmat et al., 2019). Nevertheless, Malaysia has the smallest total paddy cultivation area in the Southeast Asian region, with only 689,268 hectares devoted to rice farming (Firdaus et al., 2020). Of this, around two-thirds are situated in Peninsular Malaysia, while the remaining one-third lies in Sabah and Sarawak (Ramli et al., 2019).

Currently, an estimated 195,000 farmers are actively engaged in paddy cultivation and productivity enhancement efforts (Khazanah Research Institute, 2019). Figure 1 illustrates the annual mean data for temperature, rainfall, paddy, and rice production in Malaysia from 1963 to 2022. During this period, the mean annual temperature ranged from 25.4°C to 26.93°C, while annual rainfall varied between 2,408.61 mm and 3,587.67 mm. The paddy yield ranged from 766,168 to 2,741,404 kg, whereas rice yield ranged from 483,900 to 1,834,831 kg per hectare.

Table 1 shows the descriptive statistics on the annual average of temperature, annual total of rainfall, and annual paddy and rice production. The skewness coefficient of temperature is -0.05, which is nearly symmetrical. This shows that the temperature distribution is approximately normal, with very slight left (negative) skewness. The skewness coefficient for rainfall is 0.20, showing that positively skewed. For paddy and rice production, the skewness coefficient is -0.54 and -0.51, indicating moderate negative skewness. This means that both paddy and rice production distributions have longer left tails, reflecting the presence of fewer extremely low production values and a concentration of observations at relatively higher production levels.

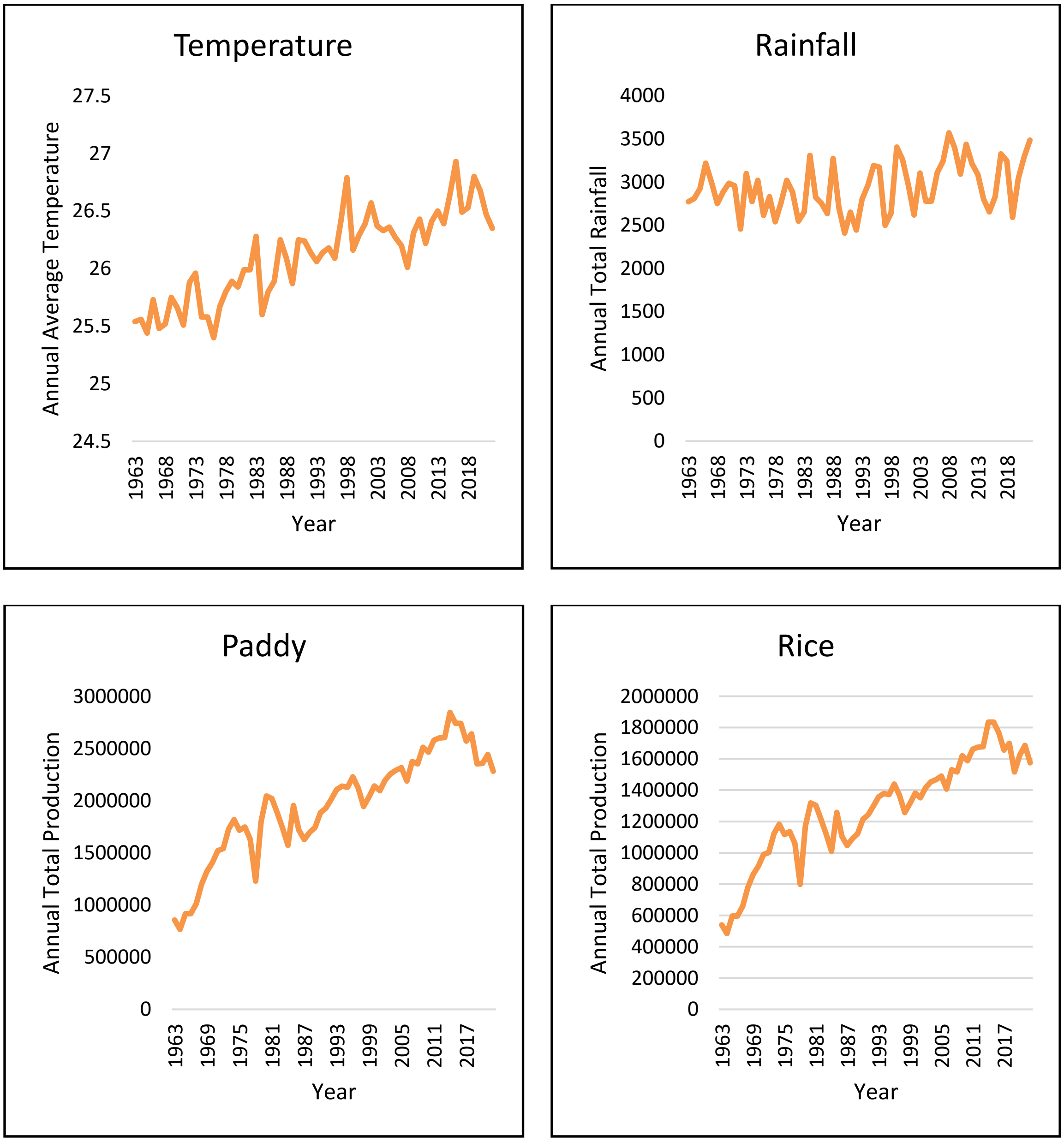

Figure 1. Annual average of temperature, annual total rainfall, annual paddy and annual rice production

Table 1. Descriptive statistics on annual temperature, precipitation, paddy and rice production

	Average Temperature	Total Rainfall	Total Paddy Production	Total Rice Production
Minimum	25.40	2408.61	766168	483900
Maximum	26.93	3568.67	2844983	1834831
Mean	26.10	2933.96	1959660	1271072
Median	26.15	2902.87	2028271	1308679
Variance	0.15	84608.16	2.54×10^{11}	1.09×10^{11}
Std. Deviation	0.38	290.87	503875.70	330742
Skewness	-0.05	0.20	-0.54	-0.51
Kurtosis	-0.79	-0.84	-0.17	-0.19

Table 2 presents the p-values of the correlation coefficients for the pairs average temperature and paddy production, average temperature and rice production, total annual rainfall and paddy production, and total annual rainfall and rice production, using Pearson, Spearman rank, and Kendall's tau methods. The p-values for the marginal distributions are all below 0.05, indicating that the null hypothesis of no correlation between the variables is rejected. Since significant correlations exist among the variables, this study proceeds to identify the best-fit marginal distribution for each variable, which are temperature, rainfall, paddy, and rice production.

Table 2. P-value of correlation coefficient among temperature, rainfall, paddy and rice production

	Temperature-Paddy	Temperature-Rice	Rainfall-Paddy	Rainfall-Rice
Pearson	0.0000	0.0000	0.0311	0.0256
Spearman rank	0.0000	0.0000	0.0102	0.0075
Kendall's tau	0.0000	0.0000	0.0159	0.0146

Table 3. Akaike information criterion for temperature, rainfall, paddy and rice production

Marginal distribution	Akaike Information Criterion (AIC)			
	Temperature	Rainfall	Paddy	Rice
Normal distribution	**57.90**	854.01	**1748.87**	**1698.36**
Gamma distribution	57.98	856.88	542972.70	368055.80
Log-Normal distribution	57.95	**853.24**	1763.55	1712.91

Table 3 shows the result of Akaike information criterion (AIC) for temperature, rainfall, paddy and rice by using normal distribution, gamma distribution and log-normal distribution. The smallest result of the AIC value indicates the best distribution of the tested variable. The result shows that temperature, paddy and rice have the lowest AIC value by using normal distribution, while rainfall has the lowest AIC value by using log-normal distribution. Based on the result from Table 3, not all variables are suitable to fit with the gamma distribution due to the highest AIC value. Choosing the marginal distribution with the lowest AIC is essential, as copula models require accurately fitted marginal distributions to effectively represent the dependence structure between variables.

Table 4 shows the fitted copula models of Gaussian, Student's t, Clayton, Gumbel and Frank copula for the dependence variables between temperature-paddy, temperature-rice, rainfall-paddy and rainfall-rice based on the parameter, log-likelihood and AIC value. The parameter (θ) measures the strength of dependence between variables and the highest value of log-likelihood shows that the copula model fits the dependence variables well. In addition, the lowest AIC value will indicate the appropriate copula model that suits the dependence variables. The dependence variables of temperature-paddy temperature-rice, rainfall-paddy and rainfall-rice show that the Frank copula has the highest log-likelihood value and the lowest AIC value. The result indicates that Frank Copula captures the full range of dependence, and it is symmetric in its dependence structure. This suggests that the relationship between paddy yield and rice production responds to climatic variations in a balanced way, where dependence is present but not skewed toward extreme events.

Table 4. Fitted copula models

Copula Model	Temperature-Paddy			Temperature-Rice		
	Parameter (θ)	Log-likelihood	AIC	Parameter (θ)	Log-likelihood	AIC
Gaussian	0.8	30.7	-59.5	0.8	31.1	-60.1
Student's t	0.8	30.7	-57.5	0.8	31.1	-58.1
Clayton	10.1	3.4	-18.1	3.5	9.4	-16.8
Gumbel	2.4	28.7	-55.3	2.4	29.2	-56.5
Frank	8.4	**32.1**	**-62.1**	8.4	**32.3**	**-62.6**
Copula Model	Rainfall-Paddy			Rainfall-Rice		
	Parameter (θ)	Log-likelihood	AIC	Parameter (θ)	Log-likelihood	AIC
Gaussian	0.3	2.4	-2.9	0.3	2.6	-3.2
Student's t	0.3	2.4	-0.9	0.3	2.6	-1.2
Clayton	0.5	-1.6	5.2	0.6	-1.5	5.0
Gumbel	1.3	3.0	-3.9	1.3	**3.3**	**-4.5**
Frank	1.8	**3.0**	**-4.0**	1.9	**3.3**	**-4.5**

4. Conclusion

Malaysia experienced some of its most severe floods in recent decades during 2021 and 2022. According to environmental anthropologist Serina Rahman, these events resulted in financial losses totaling RM6.1 billion. Such extreme weather events raise concerns about the country's preparedness, both logistically and economically, to cope with increasing climate-related risks, particularly in vulnerable regions. The agricultural sector was heavily affected, with the Department of Statistics Malaysia reporting that the rice paddy industry suffered the highest losses at RM81.4 million, followed by the livestock sector, which is RM36.9 million and the non-rice crops sector, which is RM33.2 million. These impacts highlight the vulnerability of national food security and the livelihoods of agricultural stakeholders, especially small-scale farmers who rely entirely on farming.

Statistical methods, such as regression and copula models, have been developed to explore the relationships between crop yield and weather variables. However, many existing models consider only a single variable at a time, overlooking the complex interdependencies among multiple variables. Ignoring these dependencies can reduce the model's ability to represent the true complexity of crop-climate relationships. Furthermore, traditional linear regression assumes that the effect of weather on crop yields is constant across all yield levels, which may fail to capture nonlinear responses to extreme climatic events.

Copula models provide a flexible approach to measuring dependencies among multiple variables. Therefore, this study focuses on examining the relationship between climate change and crop yield using copula methods. Specifically, it aims to assess whether climate variables significantly influence crop yield and production. Based on model comparisons using log-likelihood and AIC values, the results indicate that the Frank copula provides the best fit for the dependence structure between temperature, rainfall, paddy yield, and rice production. The Frank copula captures symmetric dependence across the full range of observations, reflecting a balanced relationship between climate variables and crop yield.

5. Acknowledgements

This research was supported by UTM Fundamental Research Grant with reference number of PY/2024/00951 and cost centre number Q.J130000.3854.23H62.

6. References

Dewick, P. R., Liu, S., Liu, Y., & Ma, T. (2023). Elliptical and skew-elliptical regression models and their applications to financial data analytics. *Journal of Risk and Financial Management, 16*(7), 310.

Dorairaj, D., & Govender, N. T. (2023). Rice and paddy industry in Malaysia: Governance and policies, research trends, technology adoption and resilience. *Frontiers in Sustainable Food Systems, 7*, 1093605.

Firdaus, R. R., Leong Tan, M., Rahmat, S. R., & Senevi Gunaratne, M. (2020). Paddy, rice and food security in Malaysia: A review of climate change impacts. *Cogent Social Sciences, 6*(1), 1818373.

Khazanah Research Institute. (2019). *The status of the paddy and rice industry in Malaysia*. Khazanah Research Institute.

Lee, W. C., & Baharuddin, A. H. (2018). Impacts of climate change on agriculture in Malaysia. In *The impact of climate change on our life: The questions of sustainability* (pp. 179–195). Springer.

NASA (2024). Climate change: How do we know? https://climate.nasa.gov/evidence/.

Ramli, N. S., Hassan, M. S., Man, N., Samah, B. A., Omar, S. Z., Rahman, N. A. A., ... & Shamsul, M. (2019). Seeking of agriculture information through mobile phone among paddy farmers in Selangor. *International Journal of Academic Research in Business and Social Sciences, 9*(6), 527–538.

Rahmat, F., Rahmat, S. T. Y., Rhian, I., & Semerdanta, P. (2019). The role of service quality and customer satisfaction: A case study for applications of Go-Food. *Russian Journal of Agricultural and Socio-Economic Sciences, 91*(7), 263–269.

Ribeiro, A. F., Russo, A., Gouveia, C. M., Páscoa, P., & Pires, C. A. (2019). Probabilistic modelling of the dependence between rainfed crops and drought hazard. *Natural Hazards and Earth System Sciences, 19*(12), 2795–2809.

Shiau, J. T. (2006). Fitting drought duration and severity with two-dimensional copulas. *Water Resources Management, 20*, 795–815.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. In *Annales de l'ISUP, 8* (3), 229-231.

Stamatatou, N., Vasiliades, L., & Loukas, A. (2018). Bivariate flood frequency analysis using copulas. In *Proceedings, 2*(11), 635.

Tang, K. H. D. (2019). Climate change in Malaysia: Trends, contributors, impacts, mitigation and adaptations. *Science of the Total Environment, 650*, 1858–1871.

Vuolo, M. (2017). Copula models for sociology: Measures of dependence and probabilities for joint distributions. *Sociological Methods & Research, 46*(3), 604–648.

World Bank. (2016). *Urban population (% of total population) — Malaysia*. https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS