

Interdependence between the Water Levels of Kotte Canal and Kelani River using Copulas

K. Perera^{1*}, P. L. N. S. Sumathipala², and P. N. Wikramanayake³

¹Department of Engineering Mathematics, Faculty of Engineering,
Postgraduate Institute of Science, University of Peradeniya, Sri Lanka

²Postgraduate Institute of Science, University of Peradeniya, Sri
Lanka

³Department of Civil Engineering, The Open University of Sri Lanka,
Nawala, Sri Lanka

*Corresponding Author: E-mail: kanthip@pdn.ac.lk

Received: 20, May 2015/ Revised: 19, September 2015/ Accepted: 29, September 2015

ABSTRACT

The interdependence of water levels of the Colombo canal system and the Kelani River is important for flood management in the city of Colombo. This study aims at deriving the interdependence between the water levels of the Kotte canal and the Kelani River using the Copula method. The advantage of the Copula method is that it is not necessary to assume that the water levels are independent or normally distributed or have the same of marginal distribution. Since there was a dependency between the two water levels (Kendall's rank correlation (τ) p -value < 0.05) the Copula method was used to derive the joint distribution. By considering the minimum values of the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), the lognormal distribution was identified as the best marginal distribution for both water levels and the Clayton copula was identified as the best Copula. By combining the identified marginal distributions using the Clayton copula, the joint distribution was derived for the two water levels and the joint return periods were calculated. The joint distribution has a peak at (0.32, 0.34) implying that these are the most frequently occurring pair of water levels. Also the return periods for the river to be high when the canal water level is high were quite long which shows that drainage to the river is feasible. Moreover the return periods were lower when the river water level is high

and canal water level is low which can be used to design the gates and the flood bund to prevent flooding in the city.

Keywords: Kendall's rank correlation, Joint distribution, Copula, Marginal distributions, Joint Return Periods

1. Introduction

The Colombo metropolitan area is situated in the wet zone of Sri Lanka and on the Kelani river flood plain. The city of Colombo is vulnerable to flooding, as most of the area consists of low lying land. The city is drained by a system of canals that has outlets to both the river and the sea. The city is protected from flooding due to high water levels in the river by a flood bund and by gates on the outlets to the river, which prevent the flow of water from the river back into the canals. Therefore, the outlets to the sea are the only means of discharging flood water from the city when the river water level is high.

Illegal development within the canal reservation has caused difficulties in canal maintenance and reduced the capacity of the canals. As a result of filling of marshy lands that act as retention areas, the retention capacity of the Colombo basin is reduced and certain parts of the canal system are now overloaded and overflow frequently (Kamil, 2010). The blockage of drainage system also causes flooding in the city. Therefore floods are common in low lying areas of the city during periods of heavy rainfall. Although many actions have been taken to prevent flooding in the city of Colombo, the city still suffers from frequent floods. Pumping water from the canal system to the Kelani river has been proposed as a measure to reduce the flood levels. Pumping is needed if the river water level is higher than the water level in the canals.

In this study, we produce results which can be used to assess the feasibility of pumping water from the Colombo Basin to the Kelani river to reduce flooding in the city. The water levels of the Kotte canal and the Kelani River have been considered for this purpose. In general the feasibility of moving water from the canals to the river depends on the joint distribution of the water levels. The water levels are usually dependent, may or may not be correlated, do not follow the normal distribution and do not have the same type of marginal distributions.

This study aims to derive a joint distribution between the water levels of the Kelani river and the Kotte canal using the copula method. The advantage of the copula method is that no assumption is needed for the variables to be independent or normal or having the same type of the marginal distributions (Zhang and Singh, 2007). Also the copulas capture the features of dependence among the random variables (Dorey and Joubert, n.d; Canela and Collazo, 2006).

Zhang and Singh (2007) have used the copula method to derive the joint distributions of rainfall intensity and depth, rainfall intensity and duration, and depth and duration using hourly rainfall data from the Amite river basin in Louisiana, United States with three stations which are Clinton, Denham Springs and Liberty. They found that the Gumbel copula is the best copula for rainfall depth and duration at Liberty, the Ali-Mikhail-Haq copula is the best for rainfall intensity and depth for three sites and the Frank copula is the best copula for rainfall intensity and duration for all three sites and for rainfall depth and duration at Clinton and Denham Springs. The bivariate distributions were then employed to determine joint and conditional return periods (Zhang and Singh, 2007).

Ghosh (2010) has used the copula method to derive the bivariate joint distribution of monthly monsoon data in neighbouring metrological subdivisions in India. He has found the best copula as Gumbel- Haugard copula. The fitted copula is then applied to derive the bivariate distribution, joint return period and conditional distribution. This was done for the two metrological subdivisions in North India.

2. Methodology

2.1 Data Used

The daily average water levels (in meters) of the Kelani River and the Kotte canal from October 2001 to December 2009 were used for this study. The Kelani river data were obtained from the Irrigation Department and the canal data were obtained from the Sri Lanka Land Reclamation and Development Cooperation (SLLRDC). The average of the river water levels from the gauging stations at Nagalagam Street and Kelanimulla were chosen for this study as the proposed location for the discharge of pumped water is between these two stations.

2.2 Kendall's Rank Correlation

Coefficient of Kendall's Rank Correlation (τ) is a nonparametric measure of dependence of two or more variables. There is a mathematical relationship between the generating function $\varphi(t)$ and the Kendall's rank correlation (τ) for the bivariate Archimedean copulas (Nelson, 1999). The relationship is given by

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt \quad (1)$$

Where $\varphi(t)$ is the generating function of the copula. The advantage of this method is that the marginal distributions do not need to be specified (Manner, 2007). Kendall's τ can be calculated from,

$$\tau_n = \binom{n}{2}^{-1} \sum_{i < j} \text{sign}[(X_i - X_j)(Y_i - Y_j)] \tag{2}$$

where n is the number of observations; $\text{sign} = 1$ if $X_i \leq X_j$ and $Y_i \leq Y_j$, otherwise, $\text{sign} = -1$; $i, j = 1, 2, \dots, N$ and τ_n is the estimate of τ from observations (Zhang and Singh, 2007). X and Y denote the two random variables of water levels. In this study, the software 'R' is used for obtaining τ and copula parameter θ (Yan, 2007).

2.3 Concept of copula

Mathematically, a copula is a function which allows one to combine univariate distributions to joint distribution with a particular dependence structure (Dorey and Joubert, n.d.). For p uniform random variables U_1, U_2, \dots, U_p on the interval $[0, 1]$ the joint distribution function C is defined as

$$C(u_1, u_2, \dots, u_p) = Pr(U_1 \leq u_1, U_2 \leq u_2, \dots, U_p \leq u_p) \tag{3}$$

The function C is called a copula. In the bivariate case, the copulas are the joint distributions of two random variables.

In this report we consider, a bivariate copula of two unit uniform random variables U and V with $C(u, v) = Pr(U \leq u, V \leq v)$. Copula joins multivariate distributions to their corresponding marginal distribution functions (Frees and Valdez, 1997; Poulin *et al.*, 2007). Let a joint distribution function $H(x, y)$ have the marginal distributions $F_X(x)$ and $F_Y(y)$ which are cumulative distribution functions of X and Y . Then there exists a function such that

$H(x, y) = C(F_X(x), F_Y(y))$ Where C is the copula. To construct the copula, Let $U = F_X(x)$ and $V = F_Y(y)$. If $F_X(x)$ and $F_Y(y)$ are the marginal distributions then C can be defined by

$$C(u, v) = F(F_X^{-1}(u), F_Y^{-1}(v)) \quad ; 0 \leq u, v \leq 1 \tag{4}$$

This gives $C(F_X(x), F_Y(y)) = F(F_X^{-1}(F_X(x)), F_Y^{-1}(F_Y(y))) = H(x, y)$.

The function $C(u, v)$ is called the copula.

2.4 Lognormal Distribution

Lognormal distribution is the probability distribution of a random variable v , whose logarithm is normally distributed. Lognormal pdf is given by:

$$\ln(v; \phi, \lambda) = \frac{1}{v\phi\sqrt{2\pi}} \exp\left[\frac{-(\ln(v) - \lambda)^2}{2\phi^2}\right] \tag{5}$$

Lognormal cdf is written as

$$LN(v; \phi, \lambda) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left[\frac{\ln(v) - \lambda}{\phi \sqrt{2}} \right] \quad (6)$$

Where $\operatorname{erf}(v) = \frac{2}{\sqrt{\pi}} \int_0^v \exp(-t^2) dt$

Lognormal parameters λ and Φ estimated using maximum likelihood method are:

$$\lambda = \frac{1}{N} \sum_{i=1}^N \ln(v_i)$$

$$\phi^2 = \frac{1}{N} \sum_{i=1}^N [\ln(v_i) - \lambda]^2$$

2.5 The Joint Return Period

The joint return period $T_{X,Y}(x, y)$ of a bivariate random variable (X, Y) corresponding to a value (x, y) is given by;

$$T_{X,Y}(x, y) = \left(\frac{1}{1 - F_{XY}(x, y)} \right) \quad (7)$$

Where $F_{XY}(x, y) = P(X \leq x, Y \leq y)$ is the joint cumulative distribution function. The joint return period of an event expressed as that either X or Y or both values are exceeded (Zhang and Singh, 2007; Ghosh, 2010).

2.6 Tail Dependence

The tail dependence relates to the amount of dependence in the upper right quadrant tail or lower left quadrant tail of a bivariate distribution (Poulin *et al.*, 2007).

The upper tail dependence is given by

$$\lambda_U = \lim_{t \rightarrow 1^-} P\{F_X(x) > t | F_Y(y) > t\} \quad (8)$$

Similarly, The lower tail dependence is given by

$$\lambda_L = \lim_{t \rightarrow 0^+} P\{F_X(x) < t | F_Y(y) < t\} \quad (9)$$

3. Results and Discussion

The ‘R’ software was used for the entire analysis. The Kendall’s rank correlation (τ) of the water levels of the Kotte canal and the Kelani river is 0.353 (p-value < .01). Thus the two water levels are correlated. The distributions Gamma, Lognormal, Weibull and Exponential were used to fit the marginal distributions

to water levels. Parameters were estimated using maximum likelihood method. The AIC and BIC values of fitted distributions are shown in the Table 1.

Table 1: AIC and BIC values of different fitted probability distributions

Probability Distributions	Kotte Canal		Kelani River	
	AIC	BIC	AIC	BIC
Gamma	-2534.385	-2522.446	-491.014	-479.074
Exponential	1194.966	1200.936	2230.196	2236.165
Lognormal	-2834.006	-2822.066	-703.016	-691.076
Weibull	-1697.8	-1685.861	87.95	99.89

According to the Table 1 since the lognormal distribution has the minimum AIC and BIC values, the relatively best marginal distribution for both water levels is the lognormal distribution. The fitted lognormal distributions for the actual water levels of the Kotte canal and the Kelani river are shown in the Figure 1 and the Figure 2 respectively.

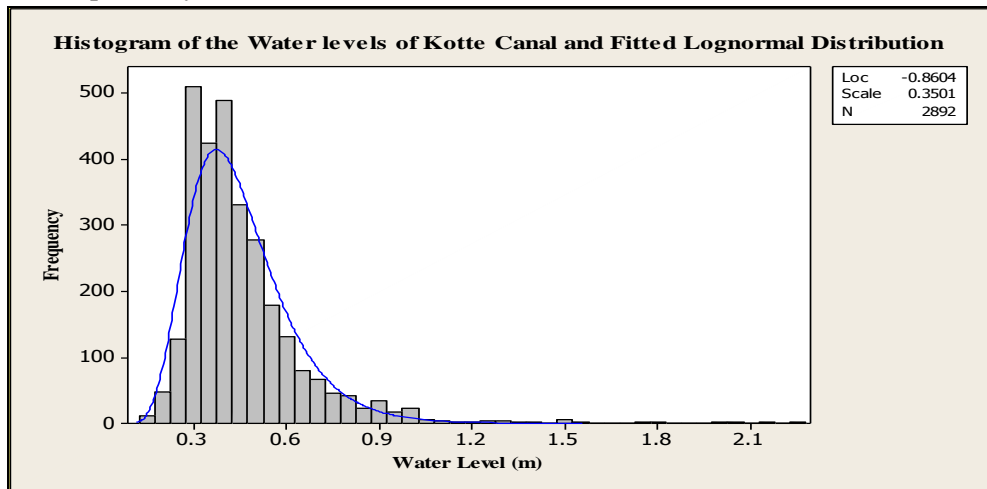


Figure 1: The histogram and the Fitted Lognormal Distribution for the Kotte canal

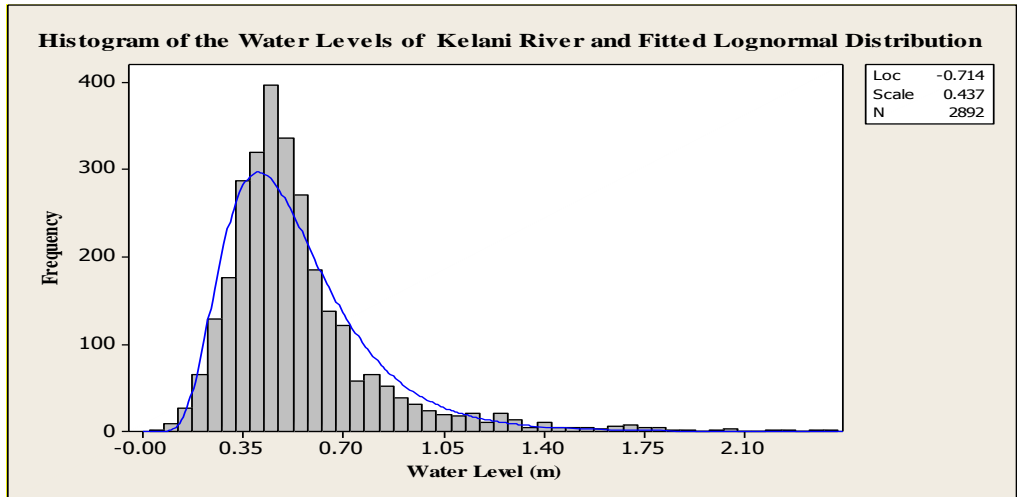


Figure 2: The histogram and the Fitted Lognormal Distribution for the Kelani river

According to the Kendall’s rank correlation τ , copula parameters (θ) were calculated. Using the estimated parameters, the copulas Gumbel, Frank, Clayton and Joe were fitted (See Appendix). The parameter values and the AIC and BIC values for different Copulas are given in Table 2.

Table 2: Estimated parameters, AIC and BIC values of the fitted copulas

Copula	Parameter (θ)	AIC value	BIC value
Gumbel	1.545	1271.79	1277.76
Frank	3.546	-957.54	-951.57
Clayton	1.091	-1038.19	-1032.22
Joe	1.99	4260.54	4266.51

The relatively best Copula was identified as the Clayton Copula by considering the minimum of the AIC and BIC values. The joint distribution of the two water levels was obtained by combining the lognormal distributions using the identified Clayton Copula. Parametric Estimators of lower and upper tail dependence coefficient for the selected copulas are given by the Table 3.

Table 3: Tail dependence coefficient for the copulas

Copula	λ_L	λ_U
Gumbel	0	$2 - 2^{\frac{1}{\theta}} = 0.433$
Frank	-	-
Clayton	$2^{\frac{-1}{\theta}} = 0.529$	0
Joe	0	$2 - 2^{\frac{1}{\theta}} = 0.588$

Since the selected Clayton copula has the lower tail dependence it can be seen that the two water levels will not be very high.

The fitted joint distribution is given by

$$H(x, y) = (F_X(x)^{-1.091} + F_Y(y)^{-1.091} - 1)^{\frac{-1}{1.091}}$$

Where $F_X(x)$ is the cumulative lognormal distribution of the water levels of the Kotte canal with the shape parameter -0.860 and the scale parameter 0.350. $F_Y(y)$ is the cumulative lognormal distribution of the water levels of the Kelani river with the shape parameter -0.714 and the scale parameter 0.437. The plot of the fitted joint distribution and contour plot are shown in the Figure 3.

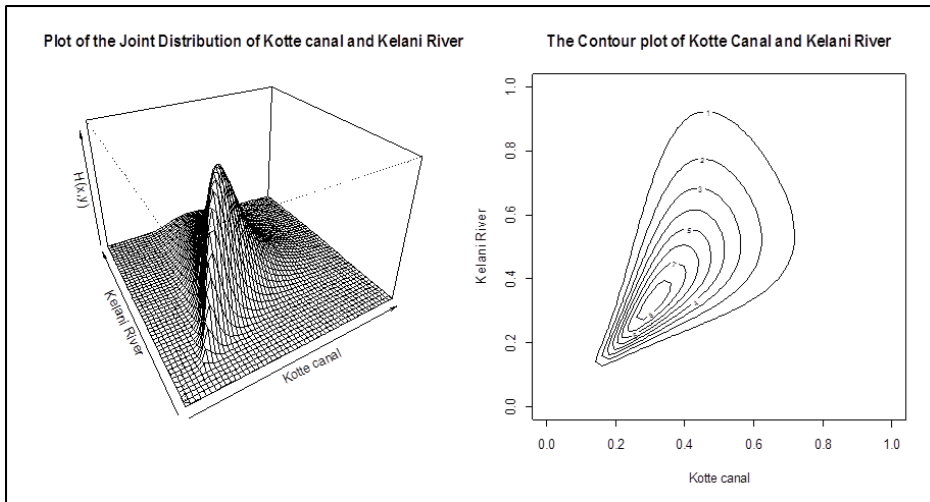


Figure 3: The plot of the fitted joint distribution and the Contour plot

It can be seen from the Figure 3 that the maximum water level occurred at the pair of water levels (0.32, 0.34) for Kotte canal and Kelani River respectively. Thus it is the most frequently occurred pair of water level.

3.1 Calculating the Return Periods

The maximums of the Kotte canal water levels and corresponding Kelani river water levels are considered for calculating the return periods. According to Table 4, the return periods are longer when the water levels of Kotte and Kelani are high. Hence the water can be discharged to river through the gates without pumping for a longer time period.

Table 4: The joint CDF and the Return Periods for the maximum values of Kotte canal

Kotte Canal (x)	Kelani River (y)	CDF $P(X \leq x, Y \leq y)$	Months
2.23	1.25	0.984	2
2.14	2.21	0.999	117
2.05	1.40	0.991	4
2.00	1.56	0.995	8
1.80	0.72	0.811	<1
1.75	1.65	0.997	12

The maximums of the Kelani river water levels and corresponding Kotte canal water levels were considered for calculating the return periods. The six pairs of water levels, joint CDF, and calculated return periods are shown in the Table 5.

Table 5: The joint CDF and the Return Periods for the maximum values of Kelani River

Kotte Canal (x)	Kelani River (y)	CDF $P(X \leq x, Y \leq y)$	Months
0.92	2.38	0.986	2
0.88	2.34	0.981	<2
0.90	2.26	0.984	2
1.52	2.26	0.999	92
2.14	2.21	0.999	117
0.20	2.06	0.016	<1

According to Table 4, the return periods are longer when the water levels of Kotte canal are high and the water level of Kelani river are high while it gives lower return period when the river water level is high and canal water level is

low. This means that if there were no gates water would flow from the river in to the canals and the result could be flooding in the city.

4. Conclusions

The concept of copula is used to derive joint distribution between two water levels and for calculating the return periods. The lognormal distribution was identified as the best marginal distribution for both water levels and among the Archimedean copulas, it was found that the Clayton copula is the best copula according to the minimum values of AIC and BIC. The joint density function was obtained by combining identified marginal distributions using the best fitted Clayton copula. The maximum of the fitted joint density $h(x, y)$ occurs at the pair of water levels (0.32, 0.34) for Kotte canal and Kelani river, respectively, which implies that it is the most frequently occurred pair of water levels.

The return periods are longer when the water levels of Kotte canal high and the water level of Kelani river are high while it gives lower return period when the river water level is high and canal water level is low. This means that if there were no gates water would flow from the river in to the canals and the result could be flooding in the city. The obtained results can be used for designing the gates and the flood bund to prevent flooding in the city

References

1. Canela, M. A. and Collazo, E. P. (2006). Modelling dependence in Latin American markets using copula functions -*Working Paper*.
2. Dorey, M. and Joubert, P., n.d. *Modelling Copulas: An Overview*.
3. Frees, E. W. and Valdez, A. E.(1997). Understanding Relationships using Copulas. *North American Actuarial Journal*, 2(1).
4. Ghosh, S. (2010). Modelling Bivariate rainfall distribution and generating bivariate correlated rainfall data in neighbouring metrological subdivisions using copula. *HYDROLOGICAL PROCESSES*, Volume 24, pp. 3558-3567.
<http://dx.doi.org/10.1002/hyp.7785>
5. Kamil, O.(2010). *Action taken from Colombo Municiple Council for reducing the risk of natural disasters in Colombo City*, Colombo: s.n.
6. Manner, H. (2007). Estimation and Model selection of Copulas with an Application to Exchange Rates. *METEOR*.

7. Nelson, R. B. (1999). *An Introduction to Copulas*. s.l.:Springer Science and Business Media.
<http://dx.doi.org/10.1007/978-1-4757-3076-0>
8. Poulin, A., Huard, D. and Fugin, A. C. (2007). Important of Tail Dependence in Bivariate Frequency Analysis. *Journal of Hydrologic Engineering*, Volume 12, pp. 394-403.
[http://dx.doi.org/10.1061/\(ASCE\)1084-0699\(2007\)12:4\(394\)](http://dx.doi.org/10.1061/(ASCE)1084-0699(2007)12:4(394))
9. Yan, J. (2007). Enjoy the Joy of Copulas: With a Package copula. *Journal of statistical Software*, 21(4).
<http://dx.doi.org/10.18637/jss.v021.i04>
10. Zhang, L. and Singh, V. (2007). Bivariate Rainfall frequency distributions using Archimedean copulas. *Journal of Hydrology*, Volume 332, pp. 93-109.
<http://dx.doi.org/10.1016/j.jhydrol.2006.06.033>

Appendix Copulas

U and V are the uniformly distributed random variables varying from 0 to 1, τ is the Kendall's tau rank correlation coefficient of the correlation between X and Y .

$U = F_X(x)$ = Cumulative Distribution Function of X

$V = F_Y(y)$ = Cumulative Distribution Function of Y

Gumbel- Haugaard Copula

$$C(u, v) = \exp(-[(-\log(u))^\theta + (-\log(v))^\theta]^{\frac{1}{\theta}}) ; \text{ Where } \theta \in [1, \infty]$$

$$H(x, y) = C(F_X(X), F_Y(Y))$$

$$= \exp(-[(-\log(F(X)))^\theta + (-\log(F(Y)))^\theta]^{\frac{1}{\theta}})$$

$\tau = \frac{\theta}{\theta+1}$; Where θ is copula parameter and τ is Kendall's tau. Gumbel family covers only positive dependency.

Frank Copula

The Frank Copula function is given by;

$$C(u, v) = -\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right) ; \text{ Where } \theta \in (0, \infty), \theta \neq 0$$

$$H(X, Y) = C(F_X(X), F_Y(Y))$$

$$= -\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta F_X(X)} - 1)(e^{-\theta F_Y(Y)} - 1)}{e^{-\theta} - 1}\right)$$

$$\tau = 1 - \frac{4[D_1(-\theta) - 1]}{\theta} \quad ; \text{ Where } D_1 \text{ is the first order Debye function } D_k$$

which is defined as $D_k(\theta) = D_k(\theta) \frac{k}{X^k} \int_0^\theta \frac{t^k}{e^t - 1} dt$; $\theta > 0$ and the Debye function D_k with the negative argument can be expressed as $D_k(-\theta) = D_k(\theta) + \frac{k\theta}{k+1}$;Where θ is copula parameter and τ is Kendall's tau. Frank family covers the positive and the negative dependency.

Clayton copula

The Clayton copula function is,

$$C(u, v) = \max[(u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}, 0] \quad ; \text{ Where } \theta \in [-1, \infty] \setminus \{0\}$$

$$\tau = \frac{\theta}{\theta + 2} \quad ; \text{ Where } \theta \text{ is copula parameter and } \tau \text{ is Kendall's tau. Clayton family covers only the positive dependency.}$$

Joe Copula

The Joe Copula function is given by;

$$C(u, v) = 1 - ((1 - u)^\theta + (1 - v)^\theta - (1 - u)^\theta (1 - v)^\theta)^{\frac{1}{\theta}} \quad ; \quad \text{Where } 1 \leq \theta < \infty$$

$$\tau = 1 + \frac{4}{\theta^2} \int_0^1 t \log(t) (1 - t)^{\frac{2(1-\theta)}{\theta}} dt \quad ; \text{ Where } \theta \text{ is copula parameter and } \tau \text{ is Kendall's tau. Joe copula covers only the positive dependency.}$$