

# Modelling Rainfall Duration And Severity Using Copula

Pauline Mah Jin Wee<sup>1,3\*</sup> and Mahendran Shitan<sup>1,2</sup>

<sup>1</sup>Laboratory of Computational Statistics and Operations Research, Institute for Mathematical Research, Universiti Putra Malaysia.

<sup>2</sup>Department of Mathematics, Faculty of Science, Universiti Putra Malaysia.

<sup>3</sup>Department of Statistics, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Malaysia.

Corresponding Author (Email: paulinejmah@yahoo.com)

Received: 18 January 2013 / Revised: 17 May 2013

## ABSTRACT

*Malaysia experiences two monsoon periods which are usually accompanied by floods and such occurrences have been quite frequent over the past two decades. Muda river which is located in the Northern part of Malaysia has experienced major flood events which are linked to the rainfall duration and severity. Copula modelling has become popular in various fields like insurance, finance, geostatistics, hydrology, etc. In Malaysia, copula modelling of floods and rainfall have not been undertaken yet and therefore the objective of this study is to fit the copulas to the rainfall duration and severity data of a selected station in Malaysia. We found that both the rainfall duration and severity follow Gamma distributions and the Gumbel-Hougaard copula appears to be a suitable model for modelling rainfall duration and severity.*

**Keywords:** Standardised Precipitation Index, Copulas, Rainfall Severity, Rainfall duration

## 1. Introduction

The monsoon seasons in Malaysia [Ref. 10] are the southwest monsoon that occurs from June to September each year and the northeast monsoon stretches from November to March. During these periods, flood is a frequent occurrence and on some occasions it is costly and life threatening. The Stormwater Management and Road Tunnel (SMART) in the capital city of Kuala Lumpur is one of the mitigation projects constructed by the government to reduce the impact of flash floods during heavy downpour. Another mitigation project in Northern Malaysia (along the Muda river) is the Muda River Mitigation Plan [Ref. 3]. Muda river lies in the border that separates the States of Kedah and Penang. The area in the vicinity of Muda river has experienced numerous floods over the past 20 years and major flood events were recorded in the years 1988, 1995, 1998 and 2003. The occurrences of floods are

usually linked to the duration and severity of rainfall. Therefore, it is essential to model these variables so that good mitigation plans and proper water resources management can be carried out.

In Malaysia, Tekolla (2010) studied the rainfall and flood frequency analysis for Pahang river basin by applying the methods of Principal Component Analysis and Log-Pearson Type III distribution respectively. Shafie (2009) did a case study on floods of 2006 and 2007 in the State of Johor, Malaysia by analysing and discussing floods from the viewpoint of the maximum precipitation and their causes. Bustami et al. (2011) modelled the flood mitigation structures for Sarawak river sub-basin using infoworks River Simulation (RS) and the results showed that the most appropriate structure is by constructing retention ponds. Razi et al. (2010) estimated the flood of Johor river using Hydrologic Modeling System (HEC-HMS) and found that models developed by HEC-HMS are useful tools to predict flood levels. A study on flood frequency analysis of annual stream flow using L-Moments and TL-Moments was done by Ahmad et al. (2011) for the stations in the State of Negeri Sembilan, Malaysia.

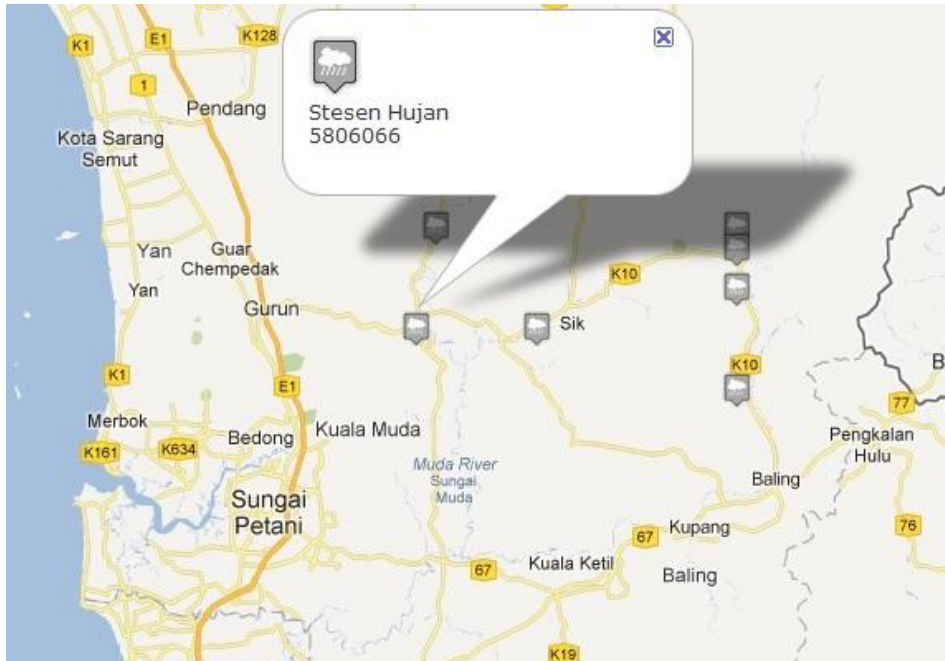
Copula Modelling has become more popular in recent times and has been used in various fields, for instance, the fields of insurance and finance [Ref. 1], geostatistics [Ref. 4], hydrology [Ref. 8] and drought studies [Ref. 15, 20, 21, 23]. There have been numerous uses of copulas in the analysis of floods. Hu et al. (2010) applied copula functions to construct a joint distribution for typhoon and plum rain (of the East Asian rainy season) while Grimaldi and Serinaldi (2006) applied asymmetric copula in multivariate flood frequency analysis. Wang et al. (2009) and Shiau et al. (2006) studied a copula-based flood frequency (COFF) analysis at the confluences of river systems and bivariate frequency analysis of floods using copula respectively. In a recent study, Rauf and Zeepongsekul (2011) modelled rainfall duration and severity of North-Eastern Victoria using copulas. In that study, the rainfall duration and severity were defined by the Standardised Precipitation Index (SPI).

Although flood and rainfall analyses in Malaysia have been done by many methods, copula modelling has not been undertaken. Therefore the objective of this study is to fit the copulas to the rainfall duration and severity data of a selected station in Malaysia based on the maximum loglikelihood function. In Section 2, a brief description of the SPI, bivariate copulas and parameter estimation are discussed. Results are presented in Section 3 and finally Section 4 contains the conclusions.

## **2. Methodology**

### **2.1 The Data Set**

In this study, we considered the Klinik Jeniang station (station no.: 5806066) that is situated on the upstream of the Muda river in the State of Kedah, Malaysia. A map of this station is shown in Figure 1 and its exact location is  $05^{\circ}48'N$  and  $100^{\circ}37'E$ . The station has a relatively high annual mean rainfall of 2325 millimeters (mm).



**Figure 1: The map of Klinik Jeniang station in Kedah, Malaysia (source: DID)**

The Department of Irrigation and Drainage (DID), Malaysia collects data at this station in three different ways. The method of data collections are (i) weekly collection using rain recorder, (ii) quarterly card collection using logger charts, and (iii) daily manual gauge readings. The rainfall data used in this study consist of 504 monthly precipitation values from the year 1970 to 2011, which was provided by the DID Malaysia.

## 2.2 Standardised Precipitation Index (SPI)

The Standardised Precipitation Index (SPI) was first developed by McKee et al. (1993). For details on the calculation of SPI, one can refer to Guttman (1998). However, for the sake of easy reference, we illustrate the computation of the 3-month SPI (SPI-3) by the following example:

### Example

The monthly precipitation values (January 1970 to December 2011) of the Klinik Jeniang station are shown in Table 1.

The details of the computation are as follows:

First, we obtain the 3-month cumulative precipitation values for each month. For example, the month of March 1970 is the sum of January, February and March 1970, similarly for April 1970 is the month of February, March and April 1970. However, no values are available for January and February 1970 as the precipitation values of November and December 1969 are not available. The values are tabulated as in Table 2.

**Table 1: Monthly precipitation values of Klinik Jeniang station from January 1970 to December 2011**

	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
1970	63.0	4.0	181.4	241.1	357.1	82.0	231.8	272.9	289.7	349.9	528.5	235.6
1971	21.5	96.1	151.1	59.1	166.7	262.7	128.5	320.0	296.9	314.6	169.9	197.9
1972	19.7	96.6	99.3	333.1	123.6	300.8	74.7	81.1	329.8	405.4	574.9	152.9
1973	13.8	53.2	23.8	288.8	222.6	120.4	187.2	250.5	259.7	331.5	165.2	288.8
1974	106.8	157.4	84.5	232.0	241.8	70.5	227.2	164.9	278.7	169.4	287.7	20.0
1975	181.0	242.5	143.5	227.5	216.4	54.5	310.0	99.0	427.5	229.0	172.5	297.0
1976	7.0	80.5	147.5	56.0	202.5	119.5	185.0	140.0	366.5	413.5	422.5	85.0
1977	94.5	27.5	35.0	76.0	394.5	150.0	132.5	402.0	404.5	582.0	85.5	132.0
1978	184.5	12.0	234.0	230.0	413.5	202.8	226.2	265.0	63.5	374.5	381.5	37.0
1979	20.5	46.5	61.0	446.0	214.5	228.0	326.0	149.5	288.0	349.0	381.0	3.0
1980	0.0	17.1	189.4	291.6	265.7	163.6	244.5	279.0	510.6	472.3	420.1	107.0
1981	50.5	101.5	55.0	323.5	510.0	182.5	156.0	91.0	290.5	234.0	148.5	26.0
1982	0.0	12.5	123.0	267.5	278.0	40.0	139.5	175.5	260.5	483.0	549.5	61.0
1983	30.5	9.5	2.5	178.5	394.0	142.4	108.6	88.5	89.0	98.5	30.0	34.0
1984	35.0	35.0	109.5	240.0	161.5	67.5	315.0	67.0	74.0	90.0	86.5	92.0
1985	26.0	203.0	274.0	163.5	274.5	107.0	127.0	184.5	203.0	311.0	468.5	73.0
1986	24.0	3.5	130.0	288.5	231.5	97.0	190.5	168.0	445.0	558.5	320.0	173.5
1987	0.0	11.5	281.5	110.0	215.5	240.0	126.5	330.5	411.0	654.5	269.0	128.5
1988	65.0	118.0	140.0	302.0	162.0	141.5	252.5	327.5	270.5	233.5	265.5	28.5
1989	12.5	38.5	117.0	359.0	248.0	136.5	219.5	149.0	347.0	367.0	55.5	86.0
1990	36.0	57.5	93.0	158.0	285.0	111.5	127.5	57.5	230.0	402.0	205.0	42.0
1991	6.5	80.5	208.5	253.0	529.0	148.0	385.0	250.0	246.0	324.5	79.5	60.0
1992	99.5	89.5	79.5	178.0	161.5	191.0	160.0	504.5	184.5	252.0	133.5	42.0
1993	10.0	40.0	65.5	177.0	230.0	168.0	488.5	271.5	201.0	251.5	192.0	159.0
1994	1.5	195.5	308.0	308.0	173.5	194.5	28.5	137.5	212.0	251.0	197.5	23.5
1995	0.0	52.5	277.5	218.0	95.0	165.5	198.0	264.5	378.0	251.0	334.5	120.0
1996	96.5	63.0	175.0	395.5	213.5	203.5	308.0	346.5	169.5	249.0	156.5	154.5
1997	14.5	297.0	47.0	270.5	119.5	134.0	155.5	314.5	339.5	584.5	232.5	115.5
1998	85.5	121.0	0.0	17.5	217.5	180.5	232.5	251.0	331.5	462.0	412.0	114.0
1999	111.5	135.5	352.5	237.5	211.3	97.2	291.0	432.9	375.7	717.5	298.6	136.4
2000	102.8	82.1	179.3	308.8	256.5	289.0	113.6	180.2	394.2	233.4	276.2	138.9
2001	248.8	48.5	242.8	353.6	264.5	199.5	155.9	142.1	207.6	388.2	195.5	56.0
2002	28.0	7.5	115.0	427.5	111.0	71.0	182.5	142.0	416.5	192.0	295.5	285.0
2003	33.0	56.0	0.0	0.0	0.0	0.0	0.0	213.5	282.0	843.5	275.0	89.5
2004	72.0	143.0	143.7	225.0	76.5	0.0	0.0	0.0	0.0	0.0	0.0	38.2
2005	1.0	30.4	202.5	241.5	95.4	74.5	101.0	83.5	205.5	350.0	224.0	268.5
2006	12.0	73.5	120.0	144.5	259.5	196.0	92.5	119.5	276.5	382.0	278.5	164.0
2007	108.0	174.0	111.5	201.5	62.5	110.5	267.0	433.0	301.5	359.5	200.5	162.0
2008	127.0	182.5	385.5	249.0	173.0	147.5	110.0	137.0	322.0	595.5	199.0	78.5
2009	48.5	55.0	327.0	396.0	276.5	29.5	265.5	484.0	169.0	220.5	290.5	15.5
2010	31.0	161.5	162.5	294.0	182.5	329.0	147.0	349.0	350.5	152.5	474.0	102.0
2011	38.0	100.0	604.5	305.0	176.5	129.0	21.0	362.5	188.5	247.5	254.0	183.0

**Table 2: Three months precipitation totals of Klinik Jeniang station from January 1970 to December 2011**

	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
1970	-	-	248.4	426.5	779.6	680.2	670.9	586.7	794.4	912.5	1168.1	1114.0
1971	785.6	353.2	268.7	306.3	376.9	488.5	557.9	711.2	745.4	931.5	781.4	682.4
1972	387.5	314.2	215.6	529.0	556.0	757.5	499.1	456.6	485.6	816.3	1310.1	1133.2
1973	741.6	219.9	90.8	365.8	535.2	631.8	530.2	558.1	697.4	841.7	756.4	785.5
1974	560.8	553.0	348.7	473.9	558.3	544.3	539.5	462.6	670.8	613.0	735.8	477.1
1975	488.7	443.5	567.0	613.5	587.4	498.4	580.9	463.5	836.5	755.5	829.0	698.5
1976	476.5	384.5	235.0	284.0	406.0	378.0	507.0	444.5	691.5	920.0	1202.5	921.0
1977	602.0	207.0	157.0	138.5	505.5	620.5	677.0	684.5	939.0	1388.5	1072.0	799.5
1978	402.0	328.5	430.5	476.0	877.5	846.3	842.5	694.0	554.7	703.0	819.5	793.0
1979	439.0	104.0	128.0	553.5	721.5	888.5	768.5	703.5	763.5	786.5	1018.0	733.0
1980	384.0	20.1	206.5	498.1	746.7	720.9	673.8	687.1	1034.1	1261.9	1403.0	999.4
1981	577.6	259.0	207.0	480.0	888.5	1016.0	848.5	429.5	537.5	615.5	673.0	408.5
1982	174.5	38.5	135.5	403.0	668.5	585.5	457.5	355.0	575.5	919.0	1293.0	1093.5
1983	641.0	101.0	42.5	190.5	575.0	714.9	645.0	339.5	286.1	276.0	217.5	162.5
1984	99.0	104.0	179.5	384.5	511.0	469.0	544.0	449.5	456.0	231.0	250.5	268.5
1985	204.5	321.0	503.0	640.5	712.0	545.0	508.5	418.5	514.5	698.5	982.5	852.5
1986	565.5	100.5	157.5	422.0	650.0	617.0	519.0	455.5	803.5	1171.5	1323.5	1052.0
1987	493.5	185.0	293.0	403.0	607.0	565.5	582.0	697.0	868.0	1396.0	1334.5	1052.0
1988	404.0	252.0	263.5	559.0	604.0	605.5	556.0	721.5	850.5	831.5	769.5	527.5
1989	306.5	79.5	168.0	514.5	724.0	743.5	604.0	505.0	715.5	863.0	769.5	508.5
1990	177.5	179.5	186.5	308.5	536.0	554.5	524.0	296.5	415.0	689.5	837.0	649.0
1991	253.5	129.0	295.5	542.0	990.5	930.0	1062.0	783.0	881.0	820.5	650.0	464.0
1992	239.0	249.0	268.5	347.0	419.0	530.5	512.5	855.5	849.0	941.0	570.0	427.5
1993	185.5	92.0	115.5	282.5	472.5	575.0	886.5	928.0	961.0	724.0	644.5	602.5
1994	352.5	356.0	505.0	811.5	789.5	676.0	396.5	360.5	378.0	600.5	660.5	472.0
1995	221.0	76.0	330.0	548.0	590.5	478.5	458.5	628.0	840.5	893.5	963.5	705.5
1996	551.0	279.5	334.5	633.5	784.0	812.5	725.0	858.0	824.0	765.0	575.0	560.0
1997	325.5	466.0	358.5	614.5	437.0	524.0	409.0	604.0	809.5	1238.5	1156.5	932.5
1998	433.5	322.0	206.5	138.5	235.0	415.5	630.5	664.0	815.0	1044.5	1205.5	988.0
1999	637.5	361.0	599.5	725.5	801.3	546.0	599.5	821.1	1099.6	1526.1	1391.8	1152.5
2000	537.8	321.3	364.2	570.2	744.6	854.3	659.1	582.8	688.0	807.8	903.8	648.5
2001	663.9	436.2	540.1	644.9	860.9	817.6	619.9	497.5	505.6	737.9	791.3	639.7
2002	279.5	91.5	150.5	550.0	653.5	609.5	364.5	395.5	741.0	750.5	904.0	772.5
2003	613.5	374.0	89.0	56.0	0.0	0.0	0.0	213.5	495.5	1339.0	1400.5	1208.0
2004	436.5	304.5	358.7	511.7	445.2	301.5	76.5	0.0	0.0	0.0	0.0	38.2
2005	39.2	69.6	233.9	474.4	539.4	411.4	270.9	259.0	390.0	639.0	779.5	842.5
2006	504.5	354.0	205.5	338.0	524.0	600.0	548.0	408.0	488.5	778.0	937.0	824.5
2007	550.5	446.0	393.5	487.0	375.5	374.5	440.0	810.5	1001.5	1094.0	861.5	722.0
2008	489.5	471.5	695.0	817.0	807.5	569.5	430.5	394.5	569.0	1054.5	1116.5	873.0
2009	326.0	182.0	430.5	778.0	999.5	702.0	571.5	779.0	918.5	873.5	679.5	526.0
2010	336.5	208.0	355.0	618.0	639.0	805.5	658.5	825.0	846.5	852.0	977.0	728.5
2011	614.0	240.0	742.5	1009.5	1086.0	610.5	326.5	512.5	572.0	798.5	690.0	684.5

1. Next, let  $\{y_{ij}\}$  be the entries of Table 2 where  $i = 1, 2, 3, \dots, 42$  and  $j = 1, 2, 3, \dots, 12$ . As zero's are omitted in the computation of the mean of  $j^{\text{th}}$  column, we define  $\bar{y}_j$  as:

$$\bar{y}_j = \begin{cases} \frac{1}{41} \sum_{i=2}^{42} y_{ij}, & j = 1 \text{ and } 2, \\ \frac{1}{42} \sum_{i=1}^{42} y_{ij}, & j = 3, 4 \text{ and } 12, \\ \frac{1}{41} \sum_{i=1}^{42} y_{ij}, & j = 5, 6, 7, 8, 9, 10 \text{ and } 11. \end{cases}$$

2.  $B_j$  is then computed as follows:

$$B_j = \begin{cases} \ln(\bar{y}_j) - \ln \frac{1}{41} \sum_{i=2}^{42} \ln(y_{ij}), & j = 1 \text{ and } 2, \\ \ln(\bar{y}_j) - \ln \frac{1}{42} \sum_{i=1}^{42} \ln(y_{ij}), & j = 3, 4 \text{ and } 12, \\ \ln(\bar{y}_j) - \ln \frac{1}{41} \sum_{\substack{i=1 \\ i \neq 34}}^{42} \ln(y_{ij}), & j = 5, 6 \text{ and } 7, \\ \ln(\bar{y}_j) - \ln \frac{1}{41} \sum_{\substack{i=1 \\ i \neq 35}}^{42} \ln(y_{ij}), & j = 8, 9, 10 \text{ and } 11. \end{cases}$$

3. We compute the estimates,  $\hat{\alpha}$  and  $\hat{\beta}$  of the Gamma distribution by the following equations.

$$\hat{\alpha}_j = \frac{1}{4B_j} \left( 1 + \sqrt{1 + \frac{4}{3}B_j} \right),$$

$$\hat{\beta}_j = \frac{\bar{y}_j}{\hat{\alpha}_j},$$

where  $j = 1, 2, 3, \dots, 12$ .

4. The cumulative Gamma distribution,  $G(y_{ij})$  is given by

$$G(y_{ij}) = \frac{1}{\hat{\beta}_j^{\hat{\alpha}_j} \Gamma(\hat{\alpha}_j)} \int_0^{y_{ij}} t^{\hat{\alpha}_j-1} e^{-\frac{t}{\hat{\alpha}_j}} dt.$$

As  $G(y_{ij})$  is defined for  $y_{ij} = 0$ , therefore  $i = 1, 2, 3, \dots, 42$  and

$j = 1, 2, 3, \dots, 12$ ,

$$\tilde{G}(y_{ij}) = \begin{cases} G(y_{ij}) & \text{if } y_{ij} \neq 0, \\ p + (1-p)G(y_{ij}) & \text{if } y_{ij} = 0, \end{cases}$$

where  $p$  is the probability of  $y_{ij} = 0$ .

5. The  $\tilde{G}(y_{ij})$  found are then transformed to a standard normal distribution with mean,  $\mu = 0$  and variance  $\sigma^2 = 1$ . These transformed values are the SPI-3 values which are displayed in Table 3.

**Table 3: The SPI-3 of Klinik Jeniang station from January 1970 to December 2011**

	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
1970	-	-	-0.146	-0.132	0.794	0.443	0.626	0.232	0.531	0.260	0.860	1.131
1971	1.537	0.766	-0.012	-0.790	-1.389	-0.767	0.094	0.824	0.312	0.323	-0.267	0.025
1972	-0.028	0.564	-0.381	0.334	-0.334	0.876	-0.208	-0.471	-1.001	-0.068	1.216	1.173
1973	1.394	-0.005	-1.619	-0.446	-0.451	0.156	-0.046	0.086	0.090	0.021	-0.350	0.324
1974	0.742	1.625	0.462	0.091	-0.321	-0.394	0.002	-0.436	-0.037	-0.840	-0.420	-0.675
1975	0.444	1.185	1.469	0.675	-0.162	-0.700	0.207	-0.431	0.712	-0.288	-0.113	0.074
1976	0.391	0.919	-0.239	-0.930	-1.217	-1.508	-0.166	-0.542	0.062	0.285	0.949	0.679
1977	0.901	-0.095	-0.872	-2.125	-0.622	-0.383	0.654	0.703	1.132	1.638	0.604	0.362
1978	0.044	0.640	0.877	0.101	1.226	1.342	1.338	0.747	-0.626	-0.485	-0.143	0.344
1979	0.221	-1.010	-1.164	0.436	0.521	1.552	1.044	0.790	0.394	-0.175	0.454	0.175
1980	-0.045	-2.625	-0.451	0.200	0.641	0.674	0.639	0.715	1.496	1.305	1.436	0.869
1981	0.808	0.247	-0.447	0.119	1.273	2.152	1.361	-0.631	-0.718	-0.830	-0.639	-0.956
1982	-1.410	-2.059	-1.084	-0.250	0.261	-0.129	-0.434	-1.090	-0.517	0.282	1.174	1.085
1983	1.046	-1.045	-2.493	-1.624	-0.229	0.641	0.510	-1.187	-1.897	-1.943	-1.967	-2.388
1984	-2.217	-1.010	-0.670	-0.346	-0.590	-0.900	0.024	-0.512	-1.164	-1.970	-1.947	-1.653
1985	-1.161	0.600	1.204	0.778	0.476	-0.390	-0.158	-0.697	-0.843	-0.502	0.353	0.504
1986	0.761	-1.051	-0.867	-0.155	0.167	0.066	-0.103	-0.477	0.570	1.054	1.248	0.992
1987	0.465	-0.258	0.141	-0.250	-0.057	-0.257	0.213	0.760	0.844	1.657	1.275	0.992
1988	0.054	0.204	-0.045	0.459	-0.073	-0.005	0.085	0.870	0.771	-0.015	-0.307	-0.486
1989	-0.469	-1.321	-0.771	0.272	0.533	0.800	0.318	-0.197	0.174	0.094	-0.307	-0.556
1990	-1.384	-0.300	-0.611	-0.776	-0.447	-0.328	-0.078	-1.454	-1.386	-0.537	-0.087	-0.078
1991	-0.804	-0.743	0.156	0.388	1.689	1.753	2.126	1.137	0.898	-0.053	-0.722	-0.727
1992	-0.904	0.185	-0.013	-0.550	-1.138	-0.485	-0.137	1.435	0.765	0.354	-1.017	-0.875
1993	-1.316	-1.155	-1.305	-0.940	-0.816	-0.196	1.506	1.717	1.218	-0.405	-0.742	-0.227
1994	-0.210	0.780	1.213	1.370	0.839	0.418	-0.786	-1.055	-1.573	-0.891	-0.684	-0.695
1995	-1.035	-1.371	0.358	0.413	-0.145	-0.835	-0.429	0.437	0.729	0.197	0.297	0.095
1996	0.703	0.370	0.383	0.751	0.814	1.168	0.862	1.445	0.659	-0.253	-0.999	-0.371
1997	-0.359	1.281	0.515	0.679	-1.030	-0.528	-0.712	0.319	0.597	1.241	0.830	0.708
1998	0.196	0.606	-0.451	-2.125	-1.940	-1.266	0.443	0.608	0.620	0.680	0.957	0.842
1999	1.033	1.084	1.596	0.805	0.893	-0.383	0.297	1.296	1.734	1.979	1.410	1.215
2000	0.650	0.602	0.545	0.504	0.631	1.382	0.574	0.213	0.045	-0.098	0.120	-0.079
2001	1.128	1.153	1.360	0.794	1.155	1.195	0.393	-0.239	-0.891	-0.353	-0.235	-0.107
2002	-0.635	-1.161	-0.934	0.422	0.185	0.020	-0.979	-0.837	0.292	-0.306	0.120	0.287
2003	0.945	0.868	-1.644	-3.346	-1.981	-1.981	-1.981	-1.856	-0.947	1.510	1.430	1.333
2004	0.210	0.511	0.516	0.259	-0.980	-1.867	-1.981	-1.981	-1.981	-1.981	-1.981	-4.083
2005	-3.317	-1.466	-0.247	0.094	-0.427	-1.293	-1.544	-1.665	-1.515	-0.736	-0.274	0.478
2006	0.512	0.770	-0.458	-0.601	-0.515	-0.039	0.045	-0.760	-0.985	-0.205	0.219	0.430
2007	0.701	1.196	0.696	0.151	-1.397	-1.529	-0.533	1.252	1.373	0.829	-0.010	0.143
2008	0.447	1.304	1.946	1.388	0.921	-0.231	-0.587	-0.843	-0.551	0.711	0.725	0.557
2009	-0.356	-0.281	0.877	1.261	1.725	0.568	0.161	1.120	1.050	0.130	-0.616	-0.492
2010	-0.297	-0.088	0.496	0.693	0.111	1.132	0.571	1.312	0.755	0.056	0.337	0.162
2011	0.947	0.128	2.109	1.963	2.055	0.026	-1.212	-0.156	-0.535	-0.131	-0.579	0.032

The SPI values are categorised as in Table 4 (Rauf and Zeepongsekul 2011)

**Table 4: Categories of SPI**

Values	Category
$SPI \geq 2.00$	extremely wet
$1.50 \leq SPI < 2.00$	severely wet
$1.00 \leq SPI < 1.50$	moderately wet
$-1.00 \leq SPI < 1.00$	near normal
$-1.50 < SPI \leq -1.00$	moderately dry
$-2.00 < SPI \leq -1.50$	severely dry
$SPI \leq -2.00$	extremely dry

For the purpose of this study, the rainfall duration,  $\tilde{D}$  is defined as the length of time where  $\text{SPI} \geq 1$ . For a given duration period,  $\tilde{D}$ , the corresponding rainfall severity,  $\tilde{S}$  is defined as the sum of SPI values  $\geq 1$  observed during that period.

### 2.3 Brief Discussion of Bivariate Copula

Sklar (1956) developed copula functions which are able to join two univariate marginal functions that are uniformly distributed on  $[0, 1]$  to form bivariate functions.

Sklar's theorem states that if  $F_{X,Y}$  is a joint distribution with marginal distributions given by  $F_X$  and  $F_Y$ , then there exists a copula  $C$  such that

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y)).$$

If  $f_X$ ,  $f_Y$  and  $c$  are the density functions of  $F_X$ ,  $F_Y$  and  $C$  respectively, then

$$f_{X,Y}(x, y) = c(F_X(x), F_Y(y))f_X(x)f_Y(y)$$

and

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}.$$

For a more comprehensive study of Copula, see Nelsen (2006).

In the study, we considered two one-parameter Archimedean copulas, namely, the Gumbel-Hougaard copula and Clayton copula as in a recent study by Rauf and Zeepongsekul (2011), they too considered these two copulas. The Gumbel-Hougaard copula is given by  $C_\theta(u_1, u_2) = e^{-[(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{\frac{1}{\theta}}}$ ,  $\theta \geq 1$  while the Clayton copula is given by  $C(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}$ ,  $\theta \geq 0$  [17, 23].

### 2.4 Parameter Estimation

Estimation of parameters can be done in various ways. The method that we employed in this study is the maximum likelihood method. By assuming that the rainfall duration follows an Exponential distribution and rainfall severity follows a Gamma distribution [Ref. 12, 15], the parameters of the two marginal distributions were estimated separately. If the estimated parameter of the Exponential distribution,  $\hat{\lambda}_{\tilde{d}}$  and the estimated parameters of the Gamma distribution  $\hat{\alpha}_{\tilde{s}}$  and  $\hat{\beta}_{\tilde{s}}$ , then  $\zeta = (\hat{\lambda}_{\tilde{d}}, \hat{\alpha}_{\tilde{s}}, \hat{\beta}_{\tilde{s}})^T$ . The parameter of the copula,  $\theta$  would then be estimated by the loglikelihood function [Ref. 22] given as

$$L(\eta) = \log c[F_{\tilde{D}}(\tilde{d}; \zeta), F_{\tilde{S}}(\tilde{s}; \zeta); \theta] + \log f_{\tilde{D}}(\tilde{d}; \zeta) + \log f_{\tilde{S}}(\tilde{s}; \zeta).$$

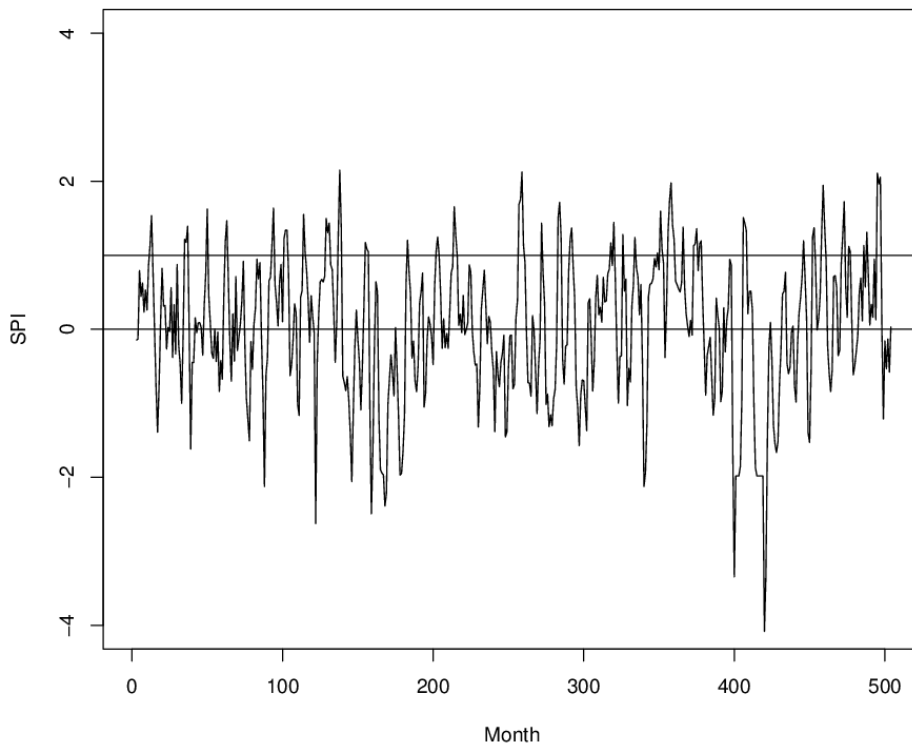
where  $\eta = (\zeta^T, \theta)^T$ .

The maximum loglikelihood estimator of  $\eta$  is  $\hat{\eta} = \text{argmax } L(\eta)$ .



### 3. Results

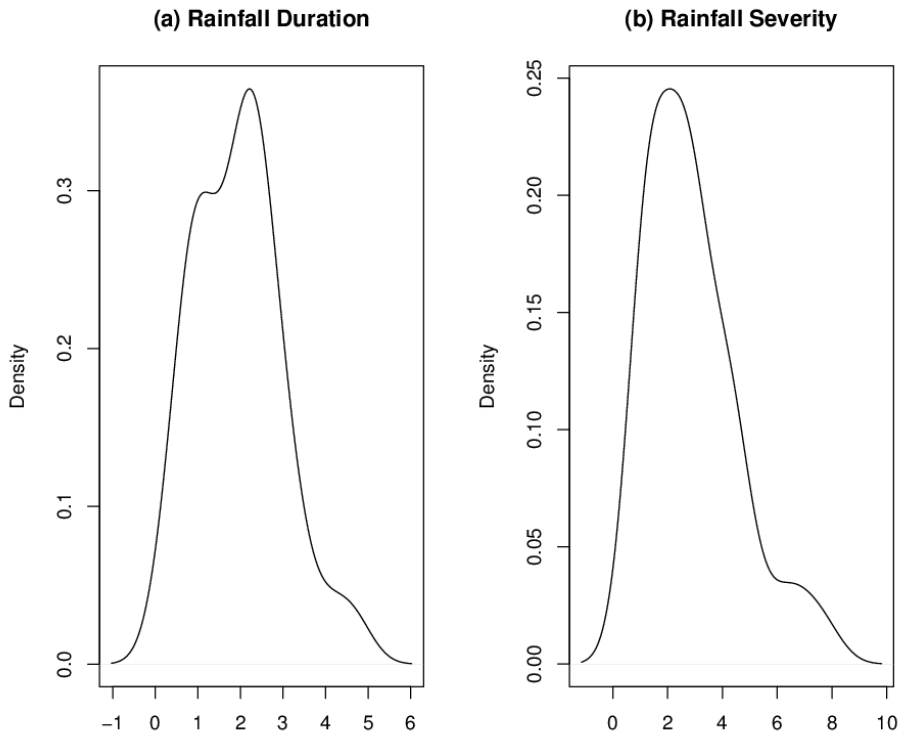
Based on these data, the 3-month SPI were computed and a plot of it is shown in Figure 2. Out of these SPI values, 36 of them have SPI values  $\geq 1.00$  which are the wet conditions. The summary statistics for the rainfall duration and severity are tabulated in Table 5 while the Kendall's tau of rainfall duration and severity is found to be 0.848. The density plots of rainfall duration and severity are shown in Figure 3.



**Figure 2: The SPI-3 plot of the Klinik Jeniang station from 1970 to 2011.**

**Table 5: Summary statistics of rainfall duration and severity.**

Statistics	Rainfall Duration	Rainfall Severity
Minimum	0.3175	1.033
1 <sup>st</sup> Quartile	1.0130	1.422
Median	2.0030	2.640
Mean	1.9480	2.896
3 <sup>rd</sup> Quartile	2.4630	3.814
Maximum	4.6770	7.634



**Figure 3: Density plots (a) Rainfall Duration and (b) Rainfall Severity**

It is usual to fit Exponential distribution for rainfall duration [Ref 12]. Following this, we also fitted an Exponential distribution ( $\hat{\lambda} = 0.513$ ). However, a chi-square goodness of fit test rejected the null hypothesis that the rainfall duration follows an Exponential distribution at the 5% level of significance (exact  $p$ -value = 0.001). Therefore we fitted a Gamma distribution ( $\hat{\alpha} = 3.176$ ,  $\hat{\beta} = 0.613$ ) for the rainfall duration. Then, a chi-square goodness of fit test was conducted and the null hypothesis that the rainfall duration is a Gamma distribution was not rejected. Similarly, rainfall severity also follows a Gamma distribution ( $\hat{\alpha} = 3.461$ ,  $\hat{\beta} = 0.836$ ). Hence, in our study, both the rainfall duration and severity follow Gamma distribution.

We then used the package “copula” in R [Ref. 22] to fit the joint distribution to the rainfall duration and rainfall severity with the copulas given in Section 2.3. The fitted parameter values of the copulas together with their loglikelihood functions are presented in Table 6. The maximised loglikelihood function is used as the criterion to choose the better joint distribution and Gumbel-Hougaard emerged as the better model.

**Table 6: The values of the estimated parameter and loglikelihood function.**

Copula	Parameter, $\theta$	Loglikelihood function
Gumbel-Hougaard	4.629	24.566
Clayton	1.822	2.845

The suitable joint distribution for the rainfall duration and rainfall severity for the Klinik Jeniang station is therefore given by:

$$C(u, v) = e^{-[(-\ln u_1)^{4.629} + (-\ln u_2)^{4.629}]^{\frac{1}{4.629}}},$$

where

$$u_1 = F_{\bar{D}}(\bar{d}) = \frac{1}{0.613^{3.176}\Gamma(3.176)} \int_0^{\bar{d}} x^{3.176-1} e^{-\frac{x}{0.613}} dx,$$

$$u_2 = F_{\bar{S}}(\bar{s}) = \frac{1}{0.836^{3.461}\Gamma(3.461)} \int_0^{\bar{s}} y^{3.461-1} e^{-\frac{y}{0.836}} dy.$$

#### 4. Conclusions

The objective of this study was to fit copulas for rainfall duration and severity for the selected station, Klinik Jeniang station as this station is located in the vicinity of the Muda river which has experienced numerous flood situations over the past two decades. We found that both rainfall duration and severity follow Gamma distribution and the Gumbel-Hougaard copula appears to be an appropriate model based on the loglikelihood function. This model will be useful for making probability statements regarding the return periods in the future. The authors are currently considering different lengths of SPI and the results will be reported in a future paper.

#### Acknowledgements

We would like to acknowledge the Department of Irrigation and Drainage (DID) Malaysia for providing the rainfall data which were used in this study.

#### References

1. Ahmad U. N., Sabri A., and Zakaria Z. A. (2011): Flood Frequency Analysis of Annual Maximum Stream Flows using L-Moments and TL-Moments. Applied Mathematical Sciences, 5: 243 – 253.
2. Bustami R., Bong C., Mah D., Hamzah A., and Patrick M. (2011): Modeling of Flood Mitigation Structures for Sarawak River Sub-basin using Infoworks River Simulation (RS). World Academy of Science, Engineering and Technology, 54: 14 – 18. <http://www.waset.org/journals/waset/v54/v54-3.pdf>. Accessed on November 10, 2011.

3. Department of Irrigation and Drainage, Malaysia (2012): [http://www.didkedah.gov.my/rtbsm/index.php?option=com\\_content&view=article&id=141&Itemid=175](http://www.didkedah.gov.my/rtbsm/index.php?option=com_content&view=article&id=141&Itemid=175). Accessed on April 10, 2012.
4. Frees E. W., and Valdez E. A. (1998): Understanding Relationships using Copulas. *North American Actuarial Journal*, 2 (1): 1 – 25, DOI: 10.1080/10920277.1998.10595667
5. Grimaldi S. and Serinaldi F. (2006): Asymmetric Copula in Multivariate Flood Frequency Analysis. *Advances in Water Resources*, 29: 1155 – 1167, DOI: 10.1016/j.advwatres.2005.09.005 , DOI: 10.1016/j.advwatres.2005.09.005
6. Guttman N. B. (1998): Comparing the Palmer Drought Index and the Standardized Precipitation Index. *Journal of the American Water Resources Association*, 34 (1): 113 – 121, DOI: 10.1111/j.1752-1688.1998.tb05964.x
7. Hu S. Y., Wang Z. Z., Wang Y. T., Wu H. Y., Jin J. L., Fend X. C., and Chen L. (2010): Encounter Probability Analysis of Typhoon and Plum Rain in the Tiahu Lake Basin. *Science China Technological Sciences*, 53 (12): 3331 – 3340, DOI: 10.1007/s11431-010-4158-2
8. Kazianka H., and Pilz J. (2009): Copula-based Geostatistical Modeling of Continuous and Discrete Data Including Covariates. *Stochastic Environmental Research and Risk Assessment*, 24: 661 – 673, DOI: 1.1017/s00477-009-0353-8.
9. McKee T. B., Doeskan N. J., and Kliest J. (1993): The Relationship of Drought Frequency and Duration to the Time Scale. In *Eighth Conference on Applied Climatology*, pages 179 – 184.
10. Meteorological Department, Malaysia (2012): [http://www.met.gov.my/index.php?option=com\\_content&task=view&id=75&Itemid=1089](http://www.met.gov.my/index.php?option=com_content&task=view&id=75&Itemid=1089). Accessed on June 22, 2012.
11. Nelsen R. B. (2006): *An Introduction to Copulas*. Springer Science+Business Media, Inc., second edition, ISBN-13: 9780387286594.
12. Rauf U. A. and Zeepongsekul P. (2011): Modelling Rainfall Severity and Duration in North-Eastern Victoria using Copulas. In *19<sup>th</sup> International Congress on Modelling and Simulations*, pages 3462 – 3468.
13. Razi M. A. M., Arriffin J., Tahir W., and Arish M. A. M. (2010): Flood Estimation Studies using Hydrologic Modeling System (HEC-HMS) for Johor River, Malaysia. *Journal of Applied Sciences*, 10 (11): 930 – 939, ISSN 1812-5654.
14. Shafie A. (2009): Extreme Flood Event: A Case Study on Floods of 2006 and 2007 in Johor, Malaysia, Technical Report.

15. Shiau J. T. (2006): Fitting Drought Duration and Severity with Two-Dimensional Copulas. *Water Resources Management*, 20: 795 – 815. DOI: 10.1007/s11269-005-9009-9.
16. Shiau J. T., Wang H. Y., and Tsai C. T. (2006): Bivariate Frequency Analysis of Floods using Copulas. *Journal of the American Water Resources Association*, 46 (6): 1549 – 1564, DOI: 10.1111/j.1752-1688.2006.tb06020.x
17. Sklar A. (1956): Fonctions De Répartition à n Dimensions Et Leurs Marges. *Publ. Inst. Statist. Univ. Paris*, 8: 229 – 231.
18. Tekolla A. W. (2010): Rainfall and Flood Frequency Analysis in Pahang River Basin, Malaysia. Master's thesis, Division of Water Resources, Lund University.
19. Wang C., Chang N. -B., and Yeh G. T. (2009): Copula-based Flood Frequency (COFF) Analysis at the Confluences of River Systems. *Hydrological Processes*, 23: 1471 – 1486, DOI: 10.1002/hyp.7273.
20. Wong G., Lambert M. F., and Melcalfe A. (2008): Trivariate Copulas for Characterisation of Droughts. *ANZIAMJ*, 49: C306 – C323.
21. Wong G., Lambert M. F., Leonard M., and Melcalfe A. (2010): Drought Analysis using Trivariate Copulas Conditional on Climate States. *Journal of Hydrologic Engineering*, 5 (2): 129 – 141, DOI: 10.1061/(ASHE)HE.1943-5584.0000169.
22. Yan J. (2007): Enjoy the Joy of Copula: With a Package Copula. *Journal of Statistical Software*, 21 (4). <http://www.jstatsoft.org>.
23. Zhang Q., Chen Y. D., Chen X., and Li J. (2011): Copula-based Analysis of Hydrological Extremes and Implications of Hydrological Behaviors in the Pearl River Basin, China. *Journal of Hydrologic Engineering*, 16 (7): 598 – 607, DOI: 10.1016/(ASHE)HE.1943-5584.0000350.
24. Zin W. Z. W., Ibrahim K., and Jemain A. (2010): Evaluating the Dry Conditions in Peninsular Malaysia using Bivariate Copula. *ANZIAMJ*, 51: C555 – C569.

