Evaluating the dry conditions in Peninsular Malaysia using bivariate copula

W. Z. Wan Zin[1]   K. Ibrahim[2]   A. A. Jemain[3]

Abstract

Prolonged dry condition has become a recurrent phenomenon in Peninsular Malaysia. The country's agricultural sector and water resources are under severe constraints from this condition. The dry conditions are examined using the Standard Precipitation Index based on the data of monthly rainfall for the period of 30 years from ten stations. The duration and severity of the dry conditions, as identified from the distribution of the values of Standard Precipitation Index, are assumed to follow the exponential and gamma distributions respectively. To study the potential risk due to the dry conditions, we model the joint distribution of these variables, upon which the return periods are estimated. The joint density function of the duration and severity are investigated by fitting six types of bivariate copula models. The Algorithmic Information Content measure identifies the most appropriate copula distribution. For the majority of stations under study, the Galambos distribution provides the best fit, and the Gumbel–Hougard distribution the second best.

Contents

1 Introduction

Drought has brought about negative impacts in terms of deterioration in the quality of life and lower crop production in Peninsular Malaysia. The crisis of prolonged dryness causes many problems, such as forcing the local authorities to conduct water-rationing in many places and plunging of revenues from crop production. Planning for a more efficient water planning system is essential in reducing the impact due to drought.

The two characteristics of drought considered are duration and severity, determined using the Standard Precipitation Index (spi). Since these two characteristics may be related, a joint distribution is used to describe their associations by making use of bivariate copula distribution. Section 2 explains spi and associated drought characteristics. Section 3 discusses copulas. Section 4 discusses application to data from ten rain-gauge stations in Peninsular Malaysia.

2   Drought index and definition

The Standard Precipitation Index (SPI), as introduced by McKee et al. [1] for defining meteorological drought, is a valuable tool for assessing the drought condition at a particular location. The calculation of SPI is based on several pre-defined timescales according to research objective and application. The commonly used time scales are 3-month, 6-month, 12-month and 24-month which correspond to the past 3, 6, 12 and 24 months of observed precipitation totals respectively. Smaller time scales may be appropriately used to reflect the soil moisture condition, information which is important for agricultural purposes. The information derived from larger time scales is useful, particularly for hydrological structure planning. Since SPI is standardised, it allows for the comparison of climate condition of areas governed by different hydrological regions (Bordi et al. [2]). Analysis is made on SPI of a one month time scale since autocorrelation is avoided in the data.

To calculate the SPI, standardisation divides the difference between a monthly rainfall total with the overall mean value for the particular month over the period of study with the standard deviation. The monthly rainfall amount for each particular month is arranged according to the respective month. The monthly data for each particular month is fitted with a suitable probability distribution. Following a suggestion by Edwards and McKee [3], the two parameter gamma distribution, which is often used to represent the distribution of precipitation amount, is fitted to the data. Next, the cumulative probabilities based on gamma distribution are then transformed to a standard normal random variable z with mean 0 and variance of 1 to produce the SPI values for each month in the data set. The SPI values obtained is then arranged in chronological order. The SPI values range from less than -2, indicating extremely dry, to greater than 2, for extremely wet. To determine the threshold levels for dry and wet conditions based on SPI values, SPI values of -1 and 1 are the threshold values for dry and wet conditions respectively.

Two drought characteristics derived from SPI values are investigated. Since a
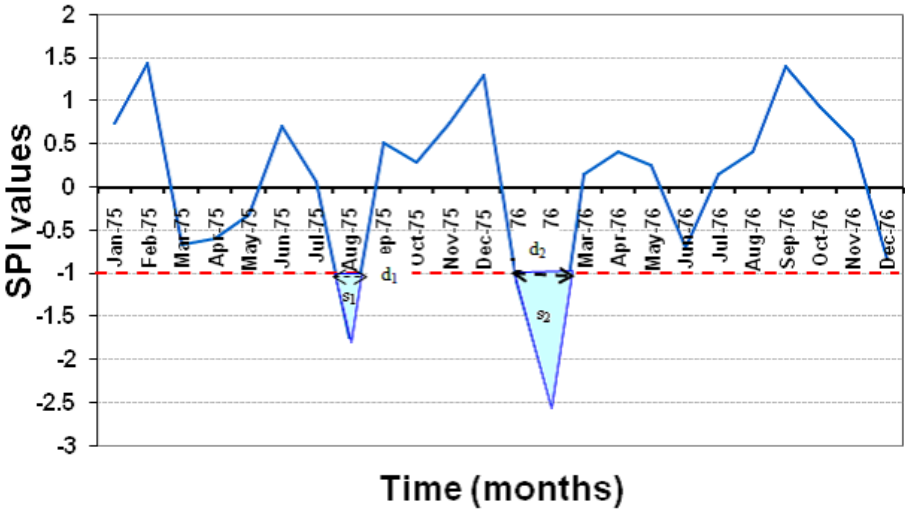
FIGURE 1: Definition sketch of drought characteristics.

drought event has been defined in the earlier paragraph as a period in which the SPI values are consistently less than or equal to -1, drought duration is defined as the number of months until the SPI value increases above the threshold of -1. Cumulative SPI values for a particular drought duration is used to measure the magnitude of a drought event and is called the drought severity. Figure 1 graphs SPI values against time to give a rough idea regarding the relationship between SPI values and drought characteristics.

From Figure 1, the dotted horizontal line at the SPI values equals to -1 indicates the threshold for dry events. The values of the ds indicate the duration between the start to the end of a dry period, that is, from the time the SPI value turns less than -1 until it becomes greater than -1. Another drought characteristic, severity which is the shaded area in Figure 1, is defined as the cumulative value of SPIs during a dry duration, that is $S = -\sum_{i=1}^{d} \text{SPI}_i$.

3 Bivariate copulas

The two drought characteristics, which are duration (D) and severity (S), are treated as random variables that are correlated, described by a joint bivariate cdf denoted by $F_{S,D}(s,d)$ and their relationship modelled using the copula function $F_{S,D}(s,d) = C(u,v)$ where u and v denote the marginal distribution functions $F_S(s)$ and $F_D(d)$ respectively. Several types of copula functions are fitted to the data for describing the association between D and S. Here, D is assumed to follow an exponential distribution with parameter λ, whereas S follows a gamma distribution with parameters α and β. The selected copula candidates are Ali–Mikhail–Haq (AMH), Clayton, Frank, Galambos, Gumbel–Hougard (GH) and Plackett distributions.

A goodness-of-fit test is employed to find the most suitable bivariate copula model that may describe the data. In order to do this, firstly, the copula parameter, denoted as θ for each copula candidate, is estimated using a maximum likelihood method. This method involves determining the value of θ which maximise the following likelihood function:

$$\log L(d,s;\lambda,\alpha,\beta,\theta) = \log L_C(F_D(d),F_S(s);\theta) + \log L_D(d;\lambda) + \log L_S(s;\alpha,\beta) \tag{1}$$

where $\log L_C(\cdot)$ is the log-likelihood for the copula function, whereas $\log L_D(.)$ and $\log L_S(.)$ are the log-likelihood functions for duration and severity respectively. The next step is to identify the most appropriate copula model to explain the relationship between duration and severity based on the Akaike Information Criterion (AIC), as developed by Akaike [4], which is AIC $= -2\log(\text{maximized likelihood for the model}) + 2(\text{number of fitted paramters})$. The smaller value of AIC indicates a better fitted model.

4 Application and study region

Peninsular Malaysia lies entirely in the equatorial zone, situated in the northern latitude between 1 and 6°N and the eastern longitude from 100 to 103°E. The climate of Peninsular Malaysia is influenced by the Southwest monsoon (sw), from May to August, and the Northeast monsoon (ne), from November until February. The latter monsoon brings about heavier rainfall to the peninsula, with the worst affected areas in the east and south. In between these two monsoons are the inter-monsoon periods, occurring in March–April and September–October, bringing intense convective rain to many areas in the peninsula. On the other hand, the driest period for the peninsula usually occurs during the Southwest monsoon with the northern part, on average, experiencing relatively long spells. Reports on potential water rationing are common during the dry seasons. Therefore, knowledge of the return period of the dry season may be able to assist in minimizing the impact of drought.

Monthly rainfall amounts from ten rain-gauge stations in Peninsular Malaysia are selected based on the completeness of the data and the length of records. Figure 2 shows the geographical locations of the stations.

The values of spi for each station are calculated and plotted against time (in months). An example of the plot for Subang Station is provided in Figure 3. The two horizontal lines at spi values of -2 and 2 indicate thresholds for extremely dry and wet events respectively. Although Malaysia experiences tropical climate, with wet condition throughout the year, the occurrences of extremely dry conditions at the majority of the stations are slightly more than the occurrences of extremely wet condition. Hence, this condition could imply that for extreme weather events, severe dry conditions occur more frequently as compared to extreme wet conditions.

Table 1 displays the Kendall τ correlation coefficient, which measures the strength of relationship between the two variables of interest. All stations exhibit positive correlation between S and D. As a copula is concerned with the dependence between variables, the value of τ is used to estimate θ.
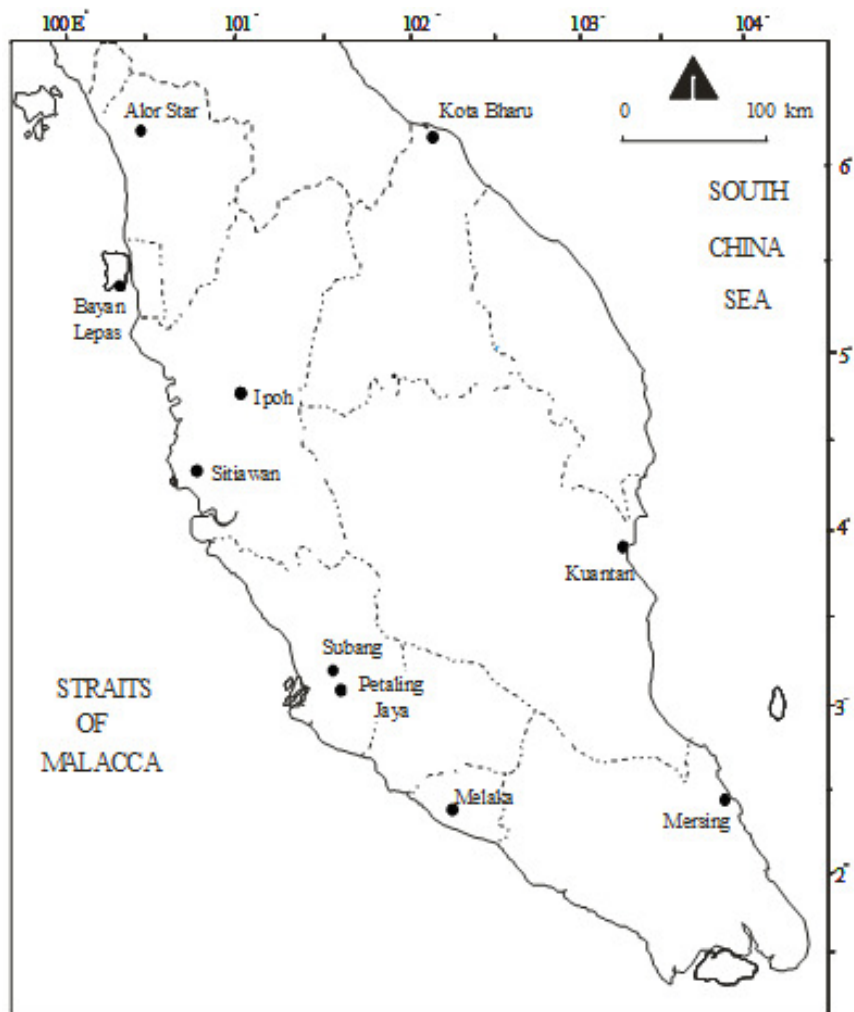
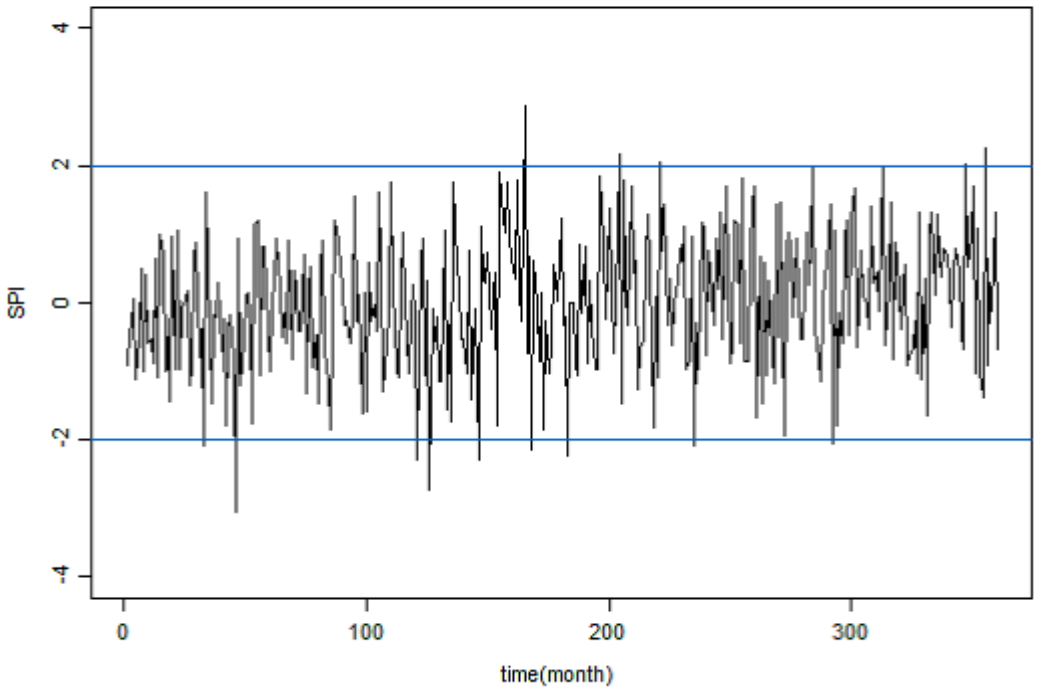FIGURE 2: Location of stations used in this study.

FIGURE 3: Time series plot of SPI values for Subang station during the 30 years period.

TABLE 1: Kendall τ correlation coefficient between S and D for all stations.

Station	τ
Melaka	0.588
Mersing	0.501
Petaling Jaya	0.422
Subang	0.304
Kuantan	0.561
Sitiawan	0.442
Ipoh	0.324
Bayan Lepas	0.621
Kota Bharu	0.593
Alor Star	0.528

Table 2 provides the exponential and gamma distributions that were fitted to the drought duration and severity respectively and results.

The six identified copula distributions are fitted to model the relationship between D and S for each station. The fitted model which yields the lowest AIC value is considered as the best distribution. As given in Table 3, when the AIC values for all the fitted models are compared, the Galambos is the best model for all stations and this is followed by GH.

Information from the analysis may be utilised for hydrological planning, particularly for drought management purpose. For example, if the drought duration has exceeded a certain number of days, it would be of interest to know what would be the probability of attaining a particular severity level? The probability of observing severity of at most s given that the duration is at least d, is

$$P\left(S \leqslant s \mid D \geqslant d\right) = \frac{F_S\left(s\right) - C\left(F_D\left(d\right), F_S\left(s\right)\right)}{1 - F_D\left(d\right)}. \tag{2}$$

In hydrological studies, the return period, defined as the average elapsed time between occurrences of an event with a certain magnitude or greater, is often

TABLE 2: Estimated parameters for describing the distribution of D and S at each station.

Variables	Duration D	Severity S	
Distribution	Exponential	Gamma	
Station	λ	α	γ
Melaka	0.831	4.770	0.398
Mersing	0.842	4.667	0.398
Petaling Jaya	0.912	10.060	0.170
Subang	0.948	9.926	0.160
Kuantan	0.778	3.989	0.497
Sitiawan	0.843	4.130	0.476
Ipoh	0.942	9.764	0.173
Bayan Lepas	0.746	3.440	0.608
Kota Bharu	0.800	5.215	0.362
Alor Star	0.830	5.706	0.331

used as a standard criterion for planning and management of water resources system [5, e.g.]. For example, it is informative to know when a particular drought condition would repeat itself. This information is important in ensuring, for example, adequate water supply is available when extreme events occur in a short return period. The return period, say T-year, for an event $S \geqslant s$ given $D \leqslant d$ is

$$T_{S|D}\left(s \mid d\right) = \frac{1}{1 - H_{S|D}\left(s \mid d\right)}$$

$$\text{where} \quad H_{S|D}\left(s \mid d\right) = \frac{H_{S,D}\left(s, d\right)}{F_D\left(d\right)} = \frac{C\left(u, v\right)}{v} . \tag{3}$$

The chance of observing a long drought is quite small. To illustrate this, plots for a station, namely Subang which is found to have a slightly higher chance in observing a longer drought than other stations, is provided in Figure 4. Figure 5 plots return periods (in years) for several levels of severity at different

TABLE 3: Copula parameters and AIC values for all distributions considered corresponding to each particular station. A * indicates the lowest AIC for the corresponding station.

| Stations		GH	Frank	Clayton	AMH	Plackett	Galambos
Melaka	θ	2.19	6.07	0.92	0.88	10.12	1.51
	AIC	−34.10	−23.93	−5.00	−7.71	−22.81	−33.43*
Mersing	θ	1.83	5.37	0.81	0.86	7.89	1.12
	AIC	−20.59	−17.56	−3.29	−6.00	−16.44	−20.91*
Petaling Jaya	θ	1.76	4.60	0.56	0.80	6.13	1.06
	AIC	−18.83	−11.53	−0.45	−3.10	−11.11	−19.34*
Subang	θ	1.43	3.50	0.34	0.72	4.15	0.71
	AIC	−7.012*	−4.25	1.13	−0.64	−4.03	−6.70
Kuantan	θ	2.16	6.08	0.92	0.88	10.31	1.47
	AIC	−32.48	−23.92	−4.97	−7.67	−23.09	−32.65*
Setiawan	θ	1.81	4.74	0.63	0.82	6.57	1.11
	AIC	−17.50	−10.46	−0.70	−2.95	−10.09	−17.68*
Ipoh	θ	1.50	3.67	0.41	0.73	4.42	0.80
	AIC	−8.82	−4.52	0.85	−0.70	−4.32	−9.26*
Bayan Lepas	θ	2.50	7.00	1.13	0.91	13.79	1.82
	AIC	−43.32	−30.68	−8.19	−10.05	−29.95	−43.46*
Kota Bharu	θ	2.34	6.53	1.02	0.89	11.86	1.66
	AIC	−35.26	−24.87	−5.72	−7.88	−24.05	−35.57*
Alor Star	θ	2.02	5.64	0.87	0.87	8.79	1.33
	AIC	−25.12	−18.10	−3.71	−6.12	−17.16	−25.42*
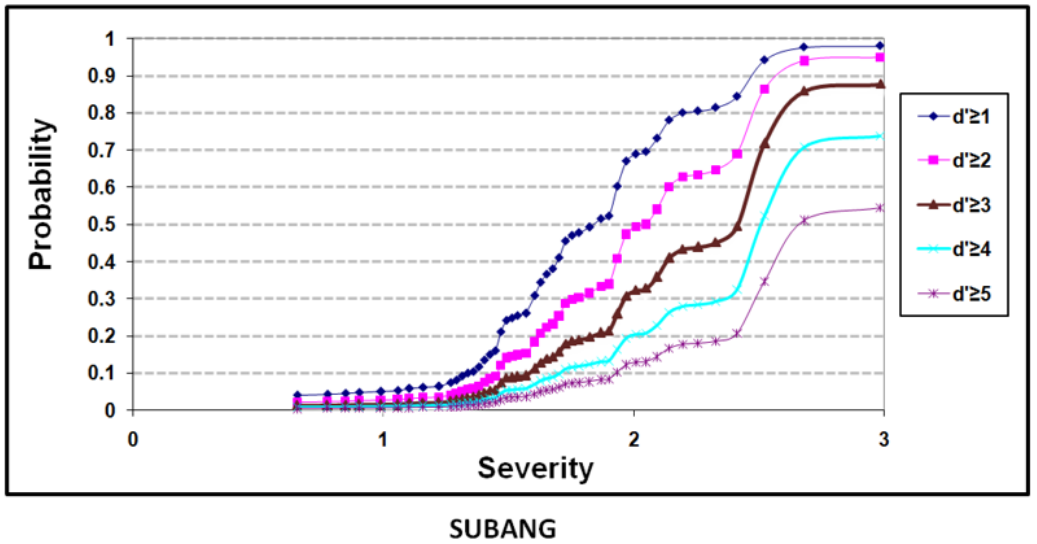
FIGURE 4: Plot of conditional probability for minimum drought severity for various drought durations.
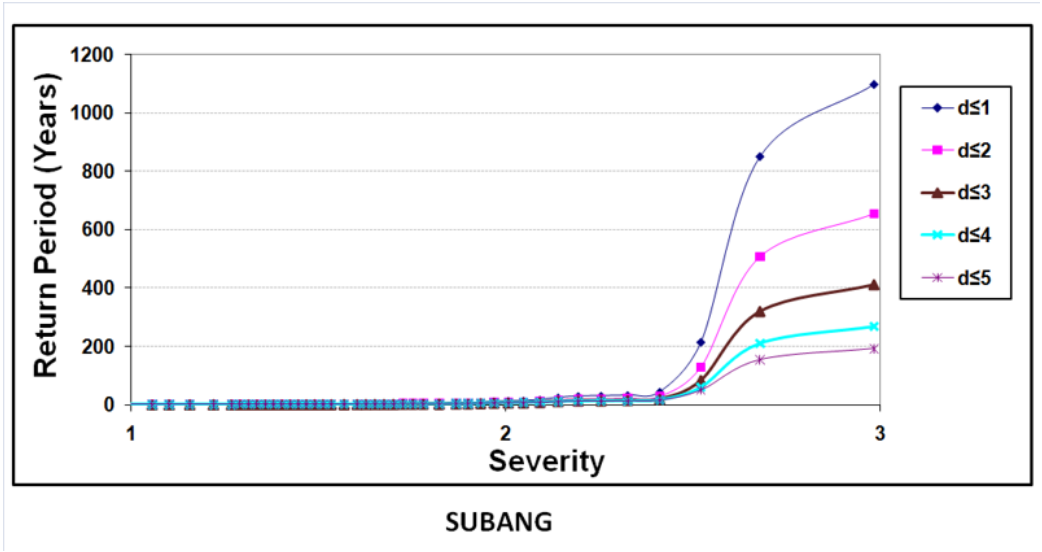
FIGURE 5: Conditional return period of drought severity for various drought durations

drought durations. In a similar manner, the plots for each station was drawn. For all stations, at higher severity levels, higher return period is observed when $D = 1$, implying that on average, the chances of occurrence of severe drought when drought duration is short is small. When the stations are compared with respect to the durations considered, stations Petaling Jaya, Subang and Ipoh are found to show relatively higher return periods for the same severity levels, implying that the dry conditions at these areas occur less frequently.

5   Conclusion

This study investigates the application of a climate index, namely Standard Precipitation Index (SPI) to evaluate the dry condition at ten stations in

Peninsular Malaysia. Time series plot of SPI at these ten stations show higher occurrences of extreme dry condition as opposed to extreme wet condition. This situation is unexpected since Peninsular Malaysia is in the tropics, receiving sunshine and rainfall throughout the year. Thus, this study highlights that extreme dry condition occur more frequently than extreme wet condition. Thus, emphasis should also be given on measures to reduce the impact of dry conditions, although the authorities usually put more emphasis on extreme wet condition.

Two drought characteristics derived from SPI values at each station, which are severity and duration, are used to analyse the dry conditions at the stations. Since these two variables are dependent, a bivariate joint distribution is employed to measure their association. The copula joint distribution of these two variables are analysed and AIC values used to determine the best fitted copula distributions for each station. Two copula distributions which come from the extreme value family, those of the Galambos and Gumbel–Hougard copula, are identified as the best fitted models for all stations in the study. In addition, the return period for each station given a particular duration are calculated. Results show that for any given duration, three stations which are Petaling Jaya, Subang and Ipoh, which are all located in the urban areas, are expected to experience higher return period compared to other stations when a high severity level is considered.

References

[1] T. B. McKee, N. J. Doesken and J. Kleist. The relationship of drought frequency and duration to time scales, *Proceedings of the 8th Conference on Applied Climatology*, 1993, 179–184. C557

[2] I. Bordi, K. Fraedrich, M. Petitta and A. Sutera. Extreme value analysis of wet and dry periods in Sicily. *Theoretical and Applied Climatology*, 2007, 87, 61–71. doi:10.1007/s00704-005-0195-3 C557

[3] D.C. Edwards and T. B. McKee. Characteristics of 20th century drought in the United States at multiple time scales. *Climatology Rep 97-2. Colorado State University Dept. of Atmospheric Science Fort Collins Colorado*, 1997, 155 pp. C557

[4] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control AC* 1974, 6, 71672. C559

[5] C. T. Haan. *Statistical Methods in Hydrology*. Iowa State University Press: Ames, IA. 1977. C564

Author addresses

1. **W. Z. Wan Zin**, School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor, MALAYSIA.
   mailto:w_zawiah@ukm.my

2. **K. Ibrahim**, School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor, MALAYSIA.

3. **A. A. Jemain**, School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor, MALAYSIA.