



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Mixed Geometric Truncated Poisson Model for Sequences of Wet Days

¹Sayang Mohd Deni and ²Abdul Aziz Jemain

¹Centre for Statistical Studies, Faculty of Information Technology and Quantitative Sciences,
Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia

²School of Mathematical Sciences, Faculty of Science and Technology,
Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

Abstract: Present study is aimed to propose the mixture of geometric distribution with the truncated Poisson model (MGTPD) as the alternative probability model to describe the distribution of wet (dry) spells in daily rainfall events. In order to compare the performance of this new model with the other five existing probability models in fitting the distribution of wet spells, daily rainfall data from five stations over Peninsular Malaysia for the period of 1975 to 2004 was considered. In determining the most successfully fitted and the best fitting model to represent the observed distribution of wet spells at each station, a Chi-square goodness-of-fit test was used. The results demonstrated that all the data sets were found to successfully fit the new proposed model, the MGTPD. Moreover, this model was also found to produce a better fit than the existing mixed geometric with Poisson model (MGPD) in describing the distribution of wet spells over the five selected stations.

Key words: Wet spells, probability model, mixed geometric truncated Poisson model, best fitted model

INTRODUCTION

Analyzing daily rainfall data is increasing and becoming important in order to provide a more reliable prediction in climatic events. Modeling daily rainfall occurrence is one of the important aspects in analyzing rainfall data. The development of rainfall occurrence models is continuously being explored by a number of researchers in the field since the 20th century. In selecting the most successful model to describe the distribution of the rainfall event, the model with the less number of parameters is preferred. However, there have been cases where the model with more parameters did not show a significant fit. Therefore, it is necessary to develop a more appropriate model which can be used for various applications including the water resource management in the hydrological and agricultural sectors.

Several probability models such as geometric distribution (GD), compound geometric distribution (CGD), geometric log series distribution (GMLS), log series distribution, modified log series distribution and truncated negative binomial have been applied for the distribution of dry (wet) spells. These probability models had been applied by a number of researchers in their respective areas of study (Deni *et al.*, 2008; Deni and Jemain, 2008; Tolika and Maheras, 2005; Anagnostopoulou *et al.*, 2003; Wilks, 1999; Chapman,

1997). The improvement on the probability models by considering the mixture distributions is continuously being explored by a number of researchers in the field. The mixture of the two geometric models (MGD) proposed by Racsco *et al.* (1991) fitted well to the distribution of dry spells in Hungary. They found that the duration of wet spells could be approached by a single geometric distribution, while for long dry spells the mixed distributions were more appropriate. In addition, Deni and Jemain (2008) reported that the MGD was also found to successfully fit the sequence of dry days over Peninsular Malaysia during two different periods, 1940s to 1976 and 1977 to 2004. Apart from that, the mixture of geometric and Poisson distribution (MGPD) which was first introduced by Dobi-Wantuch *et al.* (2000) successfully fitted the distribution of wet (dry) spells for the two stations in Hungary. However, the findings of Deni and Jemain (2008) indicated that MGPD was found not to successfully fit the distribution of dry spells over Peninsular Malaysia.

In order to further enhance the development of the geometric probability models, this present study is aimed to propose a mixture of geometric with truncated Poisson distributions (MGTPD) as the alternative probability model to describe the occurrence in daily rainfall events. The daily rainfall data from five selected stations over Peninsular Malaysia for the period of 1975 to 2004 will be

used to demonstrate the performance of the MGTPD and the existing models. In selecting the best fitting probability model to represent the distribution of wet spells at this station, a Chi-square goodness-of-fit test is used.

MATERIALS AND METHODS

The study area and data: Peninsular Malaysia lies entirely in the equatorial zone which is situated in the northern latitude between 1 and 6°N and the eastern longitude from 100 to 103°E. There are two types of monsoons that influence the climate of the country, namely, the Southwest monsoon (May to August) and the Northeast monsoon (November to February). Table 1 displays the geographical coordinates of the five selected rainfall stations for the period of 1975 to 2004 in Peninsular Malaysia. In addition, Fig. 1 shows the location of the selected rainfall stations that will be used in the analysis. The homogeneity of the data series was checked using four types of homogeneity tests as recommended by Wijngaard *et al.* (2003). The standard normal homogeneity test, the Buishand range test, the Pettit test and the Von Neumann ratio test. The results showed that the annual mean of the wet spells of the data series was homogeneous. The data used in this present study can be considered good quality data as three out of the five stations had less than 4% missing values throughout the 30-year period. The missing values in the data series for the period of 1975 to 2004 were estimated using various types of weighting methods such as the inverse distance, the normal ratio and the correlation between the target and the neighboring stations (Suhaila *et al.*, 2008; Teegavarapu and Chandramouli, 2005; Sullivan and Unwin, 2003; Eischeid *et al.*, 2000).

Probability models for wet spells: A wet day is defined as a day with a rainfall amount of at least 0.1 mm. A wet (dry) spells is a period of consecutive days of exactly, say x,

wet (dry) days immediately preceded and followed by a dry (wet) day. The minimum length of a wet spells is taken as one day which means a single wet day. Six probability models will be fitted to the observed length (days) of wet spells to determine the best fitted probability model in representing the sequences of wet days at the stations. One out of the six probability models tested is the new model proposed in this present study. According to Lana and Burgueno (1998) there are two advantages in applying the Poisson models to very long dry episodes, e.g., 22 to 50 days. Firstly, the ability of this model in quantifying the probability and return periods concerning a long episode and secondly, the probability of the drought event could be quantified for a repeated fixed number of times during a fixed number of years. Based on the modification on the Poisson model, which is also known as the truncated Poisson distribution and also the success of the one parameter GD in representing the sequences of wet (dry) days, thus, the new probability model MGTPD is proposed as an alternative probability model for daily rainfall occurrence in this present study. This new probability model is developed to provide a statistical description of the observation which is meant for various purposes not only for data generation but also for modeling climatic events in order to describe the physical explanation of the rainfall occurrence.

In fitting a particular probability model to the observed distribution of wet spells, the parameters are estimated by using either the methods of maximum likelihood, moment or factorial moment. Table 2 shows the probability models applied including the probability functions and also the method of estimating the parameters for the six models tested. The parameter of the GD is estimated using the maximum likelihood method. However, for mathematical convenience due to the complexity of the two other distributions CGD and GMLS, the factorial moment and moment method will be applied respectively when estimating the parameters. Moreover,

Table 1: The geographical coordinates, percentage of missing data and the main characteristics of wet spells length for each of the five selected rainfall stations over Peninsular Malaysia for the period of 1975-2004

	Rainfall stations				
	A. Pedu	Ipoh	J. Bahru	Ldg. Boh	Sitiawan
Geographical coordinates					
Latitude (°N)	6.23	4.57	1.47	4.45	4.22
Longitude (°E)	100.77	101.10	103.75	101.42	100.70
Percentage of missing data	3.83	0.00	3.12	0.28	0.00
Main characteristics					
Mean	2.73	2.81	2.45	3.44	2.48
Median	1.39	1.47	1.28	1.73	1.20
Standard deviation	2.49	2.55	2.34	3.43	2.24
Coefficient of variation	0.91	0.91	0.95	1.00	0.90
Maximum length of wet spells	19.00	22.00	48.00	26.00	26.00
Total No. of wet days	4627	5913	5009	6004	5160

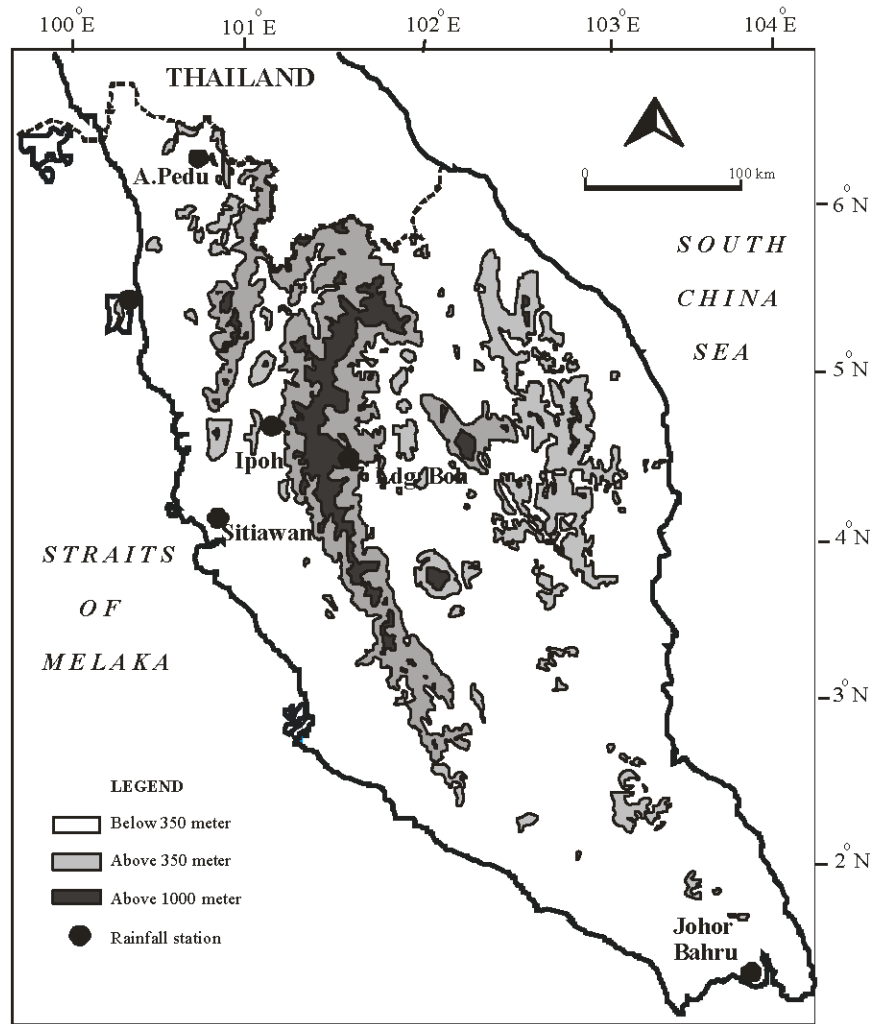


Fig. 1: The physical map showing the five selected rainfall stations in Peninsular Malaysia

Table 2: The list of probability models, probability function and the method of estimation used for fitting the distribution of wet spells, $x = 1, 2, \dots$

Probability models	Probability function	Method of estimation
GD	$P(x) = p(1-p)^{x-1}$	Maximum likelihood
CGD	$P(0) = b(a-1)^{-1}$ $P(x) = \frac{a+x-2}{a+b+x-1} P(x-1)$	Factorial moment
GMLS	$P(x) = (A_1 + \frac{A_2}{x}) p^x$	Moment
MGD	$P(x) = Wp_1(1-p_1)^{x-1} + (1-W)p_2(1-p_2)^{x-1}$	Maximum likelihood
MGPD	$P(x) = Wp(1-p)^{x-1} + (1-W) \frac{\lambda^{x-1}}{(x-1)!} e^{-\lambda}$	Maximum likelihood
MGTPD	$P(x) = Wp(1-p)^{x-1} + (1-W) \frac{\lambda^x e^{-\lambda}}{x!(1-e^{-\lambda})}$	Maximum likelihood

for the other three probability models MGD, MGPD and MGTPD, the parameters of these mixture probability models are estimated using the Quasi-Newton method

with the maximum likelihood estimation. The parameter of and for each probability model applied ranges from 0 to 1, while W is the weight factor, where the sum of W and is

Table 3: The parameter estimates for each model in fitting the distribution of wet spells for the five selected rainfall stations over Peninsular Malaysia

Station	GD		CGD			GMLS		MGD			MGPD			MGTPD		
	p	a	b	p	A1	A2	W	p ₁	p ₂	W	p	λ	W	p	λ	
A. Pedu	0.37	12.85	8.43	0.70	433.31	555.79	0.54	0.28	0.57	0.95	0.42	8.79	0.95	0.43	9.55	
Ipoh	0.36	14.55	9.05	0.71	457.34	794.68	0.10	1.00	0.33	0.97	0.39	10.42	0.73	0.30	0.83	
Ldg. Boh	0.29	14.59	6.98	0.77	284.54	499.76	0.60	0.22	0.55	0.96	0.34	13.12	0.96	0.34	13.87	
J. Bahru	0.41	6.91	5.75	na	na	na	0.02	0.12	0.43	1.00	0.41	33.41	0.60	0.33	1.04	
Sitiawan	0.40	9.47	7.41	0.70	335.73	1097.19	0.47	0.63	0.31	1.00	0.40	40.84	0.95	0.46	8.56	

unity. Table 3 shows the parameters estimation for each probability model tested in the study. In some cases, the GMLS has a disadvantage in estimating the parameter, p, which is obtained by solving the quadratic equation. Due to some circumstances, the real value of p cannot be produced as the imaginary term exists in the quadratic function.

In order to determine the most successful and the best fitting probability model to represent the distribution of wet spells at each of the five selected rainfall stations, the chi-square goodness-of-fit test is considered. The test will be computed based on the differences between the observed and expected frequencies of various lengths of wet spells (Snedecor and Cochran, 1989). The classes should be grouped for adjacent spells, whenever frequencies are less than 5. The Pearson’s chi-square goodness-of-fit test is shown:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

where, O_i and E_i are the observed and expected frequencies of wet spells and n is the number of classes. The chi-square test of 5% level of significance with the degree of freedom, $v = n-t-1$, where, t is the number of parameters estimated from each probability model, will be considered. Note that a higher probability associated to the lower chi-square statistics value produces a better fit.

RESULTS AND DISCUSSION

Main characteristics of the length of wet days: Table 1 shows the usual statistical parameters of the distribution of wet days at each of the five selected rainfall stations in Peninsular Malaysia. These include the mean, median, standard deviation, coefficient of variation, maximum spell length and the total number of wet days. It is shown that almost all the characteristics of wet spells in the Ldg. Boh station was observed to be slightly higher than in the other stations. This may be due to the geographical aspect of the Ldg. Boh station which is located in the highlands, as the highlands are the wettest areas in the Peninsula. The mean, median and the standard deviation of the duration of wet spells that varied from 2 to 3 days

were observed in all the stations except in the Ldg. Boh station. Over the study period, the J. Bahru station, which is located in the southern Peninsula, experienced the longest wet spells length, with a maximum of 48 days. Meanwhile, the A. Pedu station, which is situated in the northwestern area of the Peninsula, had the shortest of the maximum length of wet spells duration which was 19 days. Moreover, the results indicated that the highest total number of wet days of 6004 days was observed at the Ldg. Boh station.

The most successful and the best fitting probability model for wet spells:

The most successful model as shown in Table 4 is based on the highest number of stations that successfully fit the distribution of wet spells to a particular probability model at 5% level of significance. In determining the most successful model to describe the distribution of wet spells, the model with the less number of parameters is always a main interest in this field of study. However, not all of the one or two parameter models can be fitted to the distribution of interest. For example, the results in Table 4 indicated that the wet spells data at the A.Pedu station failed to fit both the one and two parameter models. In this situation, the model with the higher number of parameters is recommended. The results revealed that all the three parameter models were found to successfully fit the wet spells at the A.Pedu station except GMLS. It is observed in Table 4 that the new proposed model MGTPD was the only model that successfully fitted the distribution of wet spells in all the five selected stations. Additionally, MGTPD was also found to successfully fit the data in the J. Bahru and Sitiawan stations which were not able to fit with the existing probability model, MGPD. Moreover, the results of this present study have proven that the performance of GMLS, which is a three-parameter model, could not be compared with the rest of the models including MGTPD, in the J. Bahru station due to the failure in producing the parameter, p.

In the present study, the best fitting model was based on the largest probability associated to the lower chi square values as shown in Table 4 (in bold) for the distribution of wet spells at each of the five selected rainfall stations over Peninsular Malaysia. The findings of

Table 4: The chi square goodness-of-fit test of various types of distributions for the wet spells in each rainfall station over Peninsular Malaysia

Station		No. of parameters					
		1	2	3			
		GD	CGD	GMLS	MGD	MGPD	MGTPD
A. Pedu	Chi-sq.	65.25	23.91	22.94	17.13	16.09	15.93
	Prob.	0.00	0.01	0.01	0.07	0.10	0.10
	DF	10	10	10	10	10	10
Ipoh	Chi-sq.	50.94	8.50	12.69	25.53	12.99	8.83
	Prob.	0.00	0.67	0.24	0.00	0.29	0.55
	DF	11	11	10	10	11	10
J. Bahru	Chi-sq.	21.82	12.43	na	8.62	31.07	8.59
	Prob.	0.01	0.26	na	0.38	0.00	0.48
	DF	9	10	na	8	7	9
Ldg. Boh	Chi-sq.	95.36	19.54	12.68	7.24	24.25	22.43
	Prob.	0.00	0.11	0.47	0.89	0.06	0.10
	DF	13	13	13	13	15	15
Sitiawan	Chi-sq.	62.45	14.10	14.44	6.70	62.45	17.11
	Prob.	0.00	0.17	0.11	0.67	0.00	0.07
	DF	10	10	9	9	8	10

The shaded areas and bold face indicate the successful and best fitted models at each rainfall station respectively. na; No solution exists for the GMLS because of the imaginary term of the parameter p, $p < 0$; Successfully fitted at 5% level of significance

Table 5: The main features of the various types of probability models fitted to the length of wet spells at each of the five selected rainfall stations for the period of 1975 to 2004

Station	Probability models	Mean spell (days)	Standard deviation (days)	Median spell (days)	Skewness	Length of spells for different percentiles			Fraction of spells with length (days)				
						0.25	0.75	0.95	1	2	3	≥6	≥10
A. Pedu	Observed	2.73	2.49	1.39	0.60	2.88	7.78	0.30	0.42	0.22	0.14	0.11	0.03
	GD	2.72	2.14	1.57	0.68	3.03	6.58	0.28	0.37	0.23	0.15	0.10	0.02
	CGD	2.71	2.36	1.45	0.63	2.92	6.92	0.28	0.40	0.23	0.14	0.10	0.02
	GMLS	2.77	2.46	1.42	0.61	3.04	7.32	0.33	0.41	0.21	0.13	0.12	0.03
	MGD	2.72	2.46	1.39	0.60	2.90	7.30	0.32	0.41	0.22	0.13	0.11	0.03
	MGPD	2.72	2.45	1.42	0.62	2.86	7.57	0.29	0.40	0.23	0.13	0.11	0.03
	MGTPD	2.72	2.45	1.41	0.62	2.86	7.57	0.29	0.40	0.23	0.13	0.10	0.03
Ipoh	Observed	2.81	2.55	1.47	0.63	3.01	7.37	0.29	0.40	0.22	0.13	0.12	0.03
	GD	2.80	2.22	1.63	0.70	3.17	6.81	0.25	0.36	0.23	0.15	0.11	0.02
	CGD	2.80	2.47	1.51	0.65	3.02	7.21	0.27	0.38	0.23	0.14	0.11	0.03
	GMLS	2.78	2.52	1.39	0.60	3.03	7.48	0.35	0.42	0.20	0.12	0.12	0.03
	MGD	2.79	2.35	1.51	0.63	3.19	7.10	0.31	0.40	0.20	0.13	0.12	0.02
	MGPD	2.80	2.48	1.52	0.66	2.98	7.33	0.26	0.38	0.23	0.14	0.11	0.03
	MGTPD	2.79	2.46	1.46	0.63	3.01	7.43	0.30	0.40	0.23	0.13	0.12	0.03
J. Bahru	Observed	2.45	1.28	2.34	0.58	2.59	5.98	0.30	0.43	0.24	0.13	0.07	0.01
	GD	2.45	1.38	1.88	0.61	2.71	5.77	0.26	0.41	0.24	0.14	0.07	0.01
	CGD	2.43	1.20	2.29	0.55	2.53	6.18	0.34	0.45	0.23	0.12	0.08	0.02
	GMLS	na	na	na	na	na	na	na	na	na	na	na	na
	MGD	2.42	1.32	1.99	0.59	2.61	5.73	0.28	0.42	0.24	0.14	0.07	0.01
	MGPD	2.40	1.35	1.81	0.60	2.65	5.64	0.27	0.42	0.24	0.14	0.07	0.01
	MGTPD	2.44	1.29	2.05	0.59	2.55	6.14	0.28	0.43	0.25	0.13	0.08	0.02
Ldg. Boh	Observed	3.44	3.43	1.73	0.70	3.84	10.10	0.34	0.36	0.20	0.12	0.18	0.06
	GD	3.43	2.85	2.02	0.86	4.03	8.71	0.27	0.29	0.21	0.15	0.18	0.04
	CGD	3.39	3.20	1.84	0.77	3.82	9.22	0.30	0.32	0.21	0.14	0.17	0.05
	GMLS	3.46	3.30	1.78	0.71	4.00	9.80	0.35	0.35	0.19	0.12	0.19	0.06
	MGD	3.42	3.36	1.73	0.71	3.85	9.91	0.35	0.35	0.20	0.13	0.18	0.06
	MGPD	3.43	3.32	1.83	0.77	3.76	10.06	0.29	0.32	0.21	0.14	0.16	0.06
	MGTPD	3.44	3.33	1.83	0.77	3.76	10.17	0.30	0.32	0.21	0.14	0.17	0.06
Sitiawan	Observed	2.48	2.24	1.20	0.55	2.57	6.50	0.36	0.45	0.23	0.11	0.09	0.02
	GD	2.48	1.90	1.40	0.62	2.73	5.84	0.26	0.40	0.24	0.14	0.08	0.01
	CGD	2.46	2.14	1.26	0.57	2.60	6.14	0.32	0.44	0.23	0.13	0.08	0.02
	GMLS	2.45	2.20	1.11	0.52	2.58	6.50	0.43	0.48	0.21	0.11	0.09	0.02
	MGD	2.47	2.19	1.19	0.55	2.59	6.51	0.37	0.46	0.22	0.12	0.09	0.02
	MGPD	2.48	1.90	1.40	0.62	2.73	5.84	0.26	0.40	0.24	0.14	0.08	0.01
	MGTPD	2.48	2.17	1.25	0.57	2.57	6.60	0.32	0.44	0.24	0.13	0.09	0.02

na; No solution exists for the GMLS because of the imaginary term of the parameter p

this present study indicated that the new proposed model, MGTPD was found not only to successfully fit all the data

sets, but this model showed the best fit to the wet spells data at the J. Bahru and A. Pedu stations. The results

indicated that the feature of the main characteristics of the theoretical frequency of wet spells produced by MGTPD at each of the five selected rainfall stations were slightly underestimated or overestimated. The overall findings however, indicated that this new proposed model, MGTPD, was better than the existing probability model, MGPD (Table 5).

CONCLUSION

The development of the rainfall occurrence model will benefit many areas not only for data generation purposes but also in managing water resources for the hydrological and agricultural planning sectors. With a slight modification on the Poisson model which is also known as the truncated Poisson distribution and also the success of the one parameter GD in representing the sequences of wet (dry) days, the new probability model, MGTPD, is proposed as an alternative probability model for daily rainfall occurrence in this present study. The findings indicated that the MGTPD was superior to the existing probability model, MGPD, in describing the distribution of wet spells at each of the five selected rainfall stations for the period of 1975 to 2004. It is shown that MGTPD successfully fitted all the data sets used in this present study. Moreover, two out of the five selected rainfall stations were found not to successfully fit the existing MGPD but these data sets were found to successfully fit the MGTPD. The new probability model proposed in this present study is the alternative probability model which was developed to enhance the development in rainfall modeling particularly on the sequences of wet (dry) days. This model can be applied not only to the whole of Malaysia, but everywhere throughout the world, by considering different monsoon seasons and different threshold values.

ACKNOWLEDGMENTS

The authors are indebted to the staff of the Drainage and Irrigation Department and Malaysian Meteorological Department for providing the daily rainfall data for this study. They also acknowledge their sincere appreciation to reviewers for their valuable comments and suggestions. This research was funded by the Universiti Kebangsaan Malaysia (UKM-GUP-PI-08-34-318).

REFERENCES

Anagnostopoulou, C.H.R., P. Maheras, T. Karacostas and M. Vafiadis, 2003. Spatial and temporal analysis of dry spells in Greece. *Theor. Applied Climatol.*, 74: 77-91.

- Chapman, T.G., 1997. Stochastic models for daily rainfall in the Western Pacific. *Math. Comput. Simulation*, 43: 351-358.
- Deni, S.M. and A.A. Jemain, 2008. Probability models for dry spells in Peninsular Malaysia. *Asia-Pacific J. Atmos. Sci.*, 44: 37-47.
- Deni, S.M., A.A. Jemain and K. Ibrahim, 2008. The spatial distribution of wet and dry spells over Peninsular Malaysia. *Theor. Applied Climatol.*, 10.1007/s00704-007-0355-8.
- Dobi-Wantuch, I., J. Mika and L. Szeidl, 2000. Modeling Wet and Dry Spells with Mixture Distributions. *Meteorol. Atmos. Phys.*, 73: 245-256.
- Eischeid, J.K., P.A. Pasteris, H.F. Diaz, M.S. Plantico and N.J. Lott, 2000. Creating a serially complete, national daily time series of temperature and precipitation for the Western United States. *J. Applied Meteor.*, 39: 1580-1591.
- Lana, X. and A. Burgueno, 1998. Probabilities of repeated long dry episodes based on the Poisson distribution. An example for Catalonia (NE Spain). *Theor. Applied Climatol.*, 60: 111-120.
- Racsko, P., L. Szeidl and M. Semenov, 1991. A serial approach to local stochastic weather models. *Ecol. Model*, 57: 27-41.
- Snedecor, G.W. and W.G. Cochran, 1989. *Statistical Methods*. 8th Edn. Iowa State University Press, Ames/IO, ISBN: 0-8138-1561-6, 0813815614, pp: 503.
- Suhaila, J., S.M. Deni and A.A. Jemain, 2008. Revised spatial weighting methods for estimation of missing rainfall data. *Asia-Pacific J. Atmos. Sci.*, 44: 93-104.
- Sullivan, D.O. and D.J. Unwin, 2003. *Geographic Information Analysis*. 2nd Edn., Wiley, Hoboken, NJ, ISBN: 0471211761 .
- Teegavarapu, R.S.V. and V. Chandramouli, 2005. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *J. Hydrol.*, 312: 191-206.
- Tolika, K. and P. Maheras, 2005. Spatial and temporal characteristics of wet spells in Greece. *Theor. Applied Climatol.*, 81: 71-85.
- Wijngaard, J.B., A.M.G. Klien Tank and G.P. Konnen, 2003. Homogeneity of 20th Century European daily temperature and precipitation series. *Int. J. Climatol.*, 23: 679-692.
- Wilks, D.S., 1999. Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agric. For. Meteorol.*, 93: 153-169.