CHAPTER 6
Effect of Zero Measurements in Bivariate Rainfall Data

Shariffah Suhaila Syed Jamaludin and Kong Ching Yee

6.1 INTRODUCTION

Rainfall data consists of zero and non-zero rainfall values. Numerous rainfall analyses do not consider zero values in the data set, which could mean crucial information is lost. Therefore, to keep sufficient details, it is necessary to examine the impact of zero values in rainfall analysis before they are removed. This chapter introduces the bivariate mixed model, which combines continuous and discrete distributions to reveal the percentage of zero values in a rainfall analysis. The rainfall data from two rain gauge stations will be divided into three cases in this chapter. The first case will consider data from both stations with only positive (non-zero) values. The second case will examine data from either station with positive values, and the third case will consider data from both stations with all data values, including zeroes. The interstation correlation coefficients of the bivariate mixed distribution will also be illustrated. This chapter will use the rain gauge stations in the Johor region as a case study. The significance of zero values in rainfall analysis and the proper way to reconstruct the data set for the bivariate copula model are covered in great length in this chapter. By evaluating the significance of zero values, researchers in hydrology and climate change will decide whether it is important to incorporate zero measurements in their analysis.

6.2 REVIEW OF THE RELATED WORKS

The combination of a discrete and continuous distribution is typically difficult to comprehend in statistical models (Dunn, 2004). As a result, when analysing rainfall data, several studies overlooked zero values. On the other hand, drought, dry spells, and climatic factors analysis required zero values (Ha & Yoo, 2007; Deni & Jemain, 2009; Pakoksung & Takagi, 2017). Therefore, before zero values are ignored in any rainfall study, it is important to examine their effects. A mixed distribution was first used by Kedem et al. (1990) to analyse discrete and continuous data. A mixed distribution is one that has elements that are neither discrete nor continuous but a mixture of both.

Some authors (Fernandes et al., 2009; Kamarianakis et al., 2008; Velarde et al., 2004) proposed mixture models between Bernoulli and gamma to incorporate the zeros with a continuous dataset or log-normal distributions. Based on the idea of a mixed distribution, Shimizu (1993) proposed a mixed bivariate lognormal distribution to explain the relationship between rainfall at two rain gauge stations. The data utilised in the mixed bivariate lognormal distribution goes through four steps. In the first phase, only data with positive values (i.e., non-zero values) obtained at both stations were considered. The second and third phases only contained the positive results at each location, but the fourth phase included all the data. Ha and Yoo (2007) asserted that rainfall data were notorious for their spatial and temporal discontinuity and were also significantly biased. However, its impact on neighbor-to-neighbor interaction hasn't been previously examined. In order to investigate the relationship between severely biased and mixed random fields, a mixed bivariate lognormal distribution was adopted. Yoo and Ha (2007) expanded on their earlier work, which was centred on the effects of zero data on quantifying the rainfall field's spatial correlation function. Their study indicates that the only scenario where rainfall measurements at both stations are considered is valid for calculating the rainfall field.

6.3 STRUCTURE OF BIVARIATE RAINFALL DATA

The daily rainfall will be restructured into four categories which are $(0,0),(x^*,0),(0,y^*)$ and (x,y), where x^*, y^*, x and y are the positive values. These categories of data will then be used to obtain the parameters of the bivariate mixed lognormal distribution. Under the homogeneous assumption in time, a sample of size *N* is shown as

$\sum_{r=0}^{3} n_r = N, x_i^* > 0 \ (i=1,2,..,n_1), y_j^* > 0 \ (j=1,2,..,n_2)$ and $x_k y_k > 0 \ (k=1,2,...,n_3)$. The number $n_r\,(r=0,1,2,3)$ is a positive integer.

6.3.1 Bivariate Mixed Lognormal

A bivariate mixed lognormal distribution is used as the fitted distribution that best describes the characteristics of two rainfall stations.

Given a random vector (*X*, *Y*), a general bivariate distribution satisfies the condition as

$$
\begin{aligned}
&P(X=0,Y=0)=\delta_0 \\
&P(0<X\le x,Y=0)=\delta_1 F(x), && x>0 \\
&P(X=0,0<Y\le y)=\delta_2 G(y), && y>0 \\
&P(0<X\le x,0<Y\le y)=\delta_3 H(x,y), && x,y>0
\end{aligned}
\qquad (6.1)
$$