CHAPTER 3

Parametric Versus Non-parametric Approach in Copula Modelling

Wan Zawiah Wan Zin, Izzat Fakhruddin Kamaruzaman, Shariffah Suhaila Syed Jamaludin, and Adina Najwa Kamarudin

3.1 INTRODUCTION

Copulas have gained much attention in recent years especially in the fields of statistics, quantitative finance, civil engineering, hydrology as well as medicine due to its ability to describe dependence structure among random variables. Several copula families are available to describe the relationship between the variables of interest. Hence, selecting the most suitable family to be used according to the nature of dependency is crucial. If properly fitted, an empirical copula will be able to automatically incorporate this dependence nature, and subsequently allows the modelling of dependence structures.

In the construction of a copula model, the marginal distributions of the variables of interest are assumed known. This type of parametric marginal models is frequently applied in many studies related to copula modelling. For instance, studies on modelling the dependency of rainfall characteristics using parametric models have been applied by Wan Zin et al. (2010), Chen et al. (2015), Daneshkhah et al. (2016), Tosunoglu and Kisi (2016).

Alternatively, non-parametric copula may be applied when modelling the dependency between variables. This type of copula assumes that an estimate of underlying dependency structure of the data is unconstrained by any a-priori model on how the data interacts with each other. In parametric copula, the parameters of the marginal distributions of the variables of interest are assumed as the best fit to the data and are able to retain the copula's functional form when fitted to any copula model. The functional form of empirical copula, on the other hand, is not limited to the parameters based on data, thus making it more flexible to capture any correlation pattern.

In this chapter, a non-parametric approach based on kernel density estimator is used to generate the copula model. This alternative approach was selected to circumvent the inherent arbitrariness of having to identify suitable distributions in parametric frequency analysis. The Standard Precipitation Index (SPI) is used as an application example. Farahmand and AghaKouchak (2015) introduced Standardized Drought Analysis Toolbox (SDAT) which provides a nonparametric framework of drought indicators and showed the deficiency of the traditional parametric SPI. Ghamghami et al. (2017) applied a nonparametric method based on Kernel Density Estimator (KDE) for calculating the SPI at Tehran which, based on their results, provides more accurate results pertaining to the frequency of occurrences of extreme drought events. Other studies on nonparametric SPI can be found at Cancelliere et al. (2006), Kim et al. (2006) and Abdul Rauf and Zeephongsekul (2014).

The kernel approach is a conceptually simple method similar to a smoothed histogram in nature. In kernel estimation, the most critical parameter is the one that monitors the level of smoothing. There are various types of kernel functions that can be used. Statistical literature suggests that the choice of smoothing parameter is relatively more crucial than the kernel choice. Faucher et al. (2002) discovered no major differences between four types of kernels known as Biweight, Gaussian, Epanechnicov and EV1. Thus, this chapter focused on the Gaussian kernel as it has been widely practiced in hydrology (Adamowski, 2000; Sharma, 2000; Kim & Heo, 2002; Mosthaf & Bardossy, 2017).

3.2 COPULA METHODS

The derivation for SPI was introduced by Mckee et al. (1993) which is used to represent the drought condition of an area based on rainfall characteristics. In the process of SPI derivation, rainfall amount at specified scales is used. From the generated SPI values, rainfall characteristics such as duration and severity can be obtained. Researchers are often interested in these rainfall characteristics when studying about extreme events. Thus, copula models are usually applied to represent the relationship between these rainfall characteristics. Parametric marginal distributions of the characteristics are used in the copula modelling hence it is crucial that the most accurate distribution is identified for each characteristic of interest.

In this chapter, a kernel smoothing method is used to obtain estimation of density function for fitting the rainfall time series data. The selection of a bandwidth parameter is essential when applying the kernel smoothing technique. The cumulative density function, $F(y)$ of the observed rainfall event can be estimated via optimal bandwidth parameter. The SPI will then be calculated based on the relationship:

$$\mathrm{SPI} = \Phi^{-1}\big(F(y)\big) \tag{3.1}$$