Modelling Nonlinear Bivariate Dependence Using the Boubaker Polynomials Copula: Application to Infiltration Rainfall Patterns in Saddine-1 (Makthar, Northern Tunisia)

Emna Gargouri-Ellouze, Noreen Sher Akbar and Sohail Nadeem

1Institut Supérieur des Etudes Technologiques de Radès, Tunisie (High Institute of Technological Studies of Rades), 2098, Tunisia
2Department of Mathematics, Quaid-i-Azam University 45320, Islamabad 44000, Pakistan

Abstract: In this paper, we apply the Boubaker Polynomials copula to a set of discrete random vectors possessing uniform margins. We further suggest a pragmatic way to fit the dependence structure of multivariate data to Boubaker Polynomials copula and empirical contingency tables. Finally, we discuss an application of the relationship between infiltration index (φ-index) and the average intensity of rainfall event in northern Tunisia (Saddine 1, Makthar). We study the modeling of dependence patterns of infiltration index (φ-index) and the average intensity (Imoy) of rainfall event via a multi-parametric linear regression framework equipped with the Boubaker polynomial expansion to the multi-dimensional copula and estimate it with empirical data of infiltration index (φ-index) and the average intensity of rainfall event (hydrodynamical data). Our result shows that it is a feasible approach to describe the dependence patterns with the Boubaker Polynomials copulas approximation.

Key words: Boubaker polynomial copula, infiltration index (φ-index), dependence patterns, intensity rainfall

INTRODUCTION

We find in Schweizer [1], Joe [2]; Frees et al. [3] Nelsen [4]; Roncalli [5], Denuit et al. [6], Genest et al. [7] and references herein, Copula theory.

A copula is a multivariate joint cumulative distribution defined on the p-dimensional unit cube \([0, 1]^p\) such that every marginal distribution is uniform on the interval \([0, 1]\). Consider a vector \(X = (X_1, X_2, \ldots, X_p)\) of \(p \geq 2\) continuous random variables, we call the function \(C\) a copula [8], if:

\[
\text{Pr}\{X_1 \leq x_1, X_2 \leq x_2, \ldots, X_p \leq x_p\} = C\left(F_1(x_1), F_2(x_2), \ldots, F_p(x_p)\right)
\]

such as: \(F_k(x_k) = \text{Pr}(X_k \leq x_k)\) is the marginal distribution of \(X_k\).

\(C\) is the cumulative distribution function of a vector \((U_1, U_2, \ldots, U_p)\) of dependent uniform random variables on the interval \([0,1]\):

\[
C\left(F_1(x_1), F_2(x_2), \ldots, F_p(x_p)\right) = \text{Pr}(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_p \leq x_p)
\]

The associated copula to \(X = (X_1, X_2, \ldots, X_p)\) is,

\[
C(u_1, u_2, \ldots, u_p) = F_1^{-1}(u_1), F_2^{-1}(u_2), \ldots, F_p^{-1}(u_p)
\]

\[
= \text{Pr}(U_1 \leq u_1, U_2 \leq u_2, \ldots, U_p \leq u_p)
\]

Specifically, \(C: [0,1]^p \rightarrow [0,1]\) is an \(p\)-dimensional copula if:

\[
C(u_1, u_2, \ldots, u_p) = 0
\]

and

\[
C(1, 1, \ldots, 1) = u; i = 1 \ldots p
\]

\(C\) is bounded and \(p\)-increasing.

The density function \(f\) of \(F\) is expressed by:

\[
f(x_1, x_2, \ldots, x_p) = \prod_{k=1}^{p} f_k(x_k) = \prod_{k=1}^{p} f_k(x_k) \times \prod_{k=1}^{p} f_k(x_k)
\]

\(f_k\) is density function of \(X_k\) and \(c(u_1, u_2, \ldots, u_p)\) is copula density defined by:
Parallel to this definition, let:

\[
c(u, u_2, \ldots, u_d) = \frac{\partial^d C(u, u_2, \ldots, u_d)}{\partial u \partial u_2 \cdot \ldots \partial u_d}
\]  \tag{6}

In addition, it is worth notice that each copula is bounded by the Frechet-Hoeffding bounds [9], so that:

\[
\max \{F_1(x_1) + F_2(x_2) + \ldots + F_d(x_d) - 1, 0\} \leq F(x_1, x_2, \ldots, x_d) \\
\leq \min \{F_1(x_1), F_2(x_2), \ldots, F_d(x_d)\}  \tag{7}
\]

There are several types of copula [4, 7, 10, 11]: non-parametric copula and parametric one (one or multi-parameters). Among copula belong to the first category, empiric copula, independence copula, periodic copula introduced by Alfonsi et al. [4] and Fréchet copula (« upper bound » maximum copula and « lower bound » minimum copula). In the second copula category we distinguish several families: Archimedean copula [1-8], meta-elliptical copula, in particular Gaussian copula, Cauchy copula and Student copula, exponential copula and contamination copula [10], survival copula [12], extreme value copula: Joe, Clayton, Gumbel copula, BB5 [7] and Galambos [13].

In hydrology, one often uses families of multivariate distributions which are extensions of univariate and suffer from several limitations and constraints, such as the marginal distributions which may belong to the same probability family [14, 15]. Thus to avoid these limitations, copulas models are used [7]. Indeed, copulas allow the description of the dependence structure between random variables without information on the marginal distributions and also the description of the multivariate distributions with any kind of marginal distributions [11]. It is worth notice that Archimedean copula and meta-elliptical copula families remain the most used.

In this paper we apply the Boubaker Polynomials copulas to a set of discrete random vectors retrieved from data of infiltration index (\(\phi\)-index) and the average intensity (\(I_{\text{moy}}\)) of rainfall event.

**BOUBAKER POLYNOMIALS COPULA**

For an integer \(d\), let \(U = (U_1, \ldots, U_d)\) be random vector whose marginal component \(U_{ik} = 1, k\) follows a discrete uniform distribution over \(T_{ik} = 1, d = \{0, 1, 2, \ldots, m-1\}\) with \(m\) an non-null integer and:

\[
p(U_1, \ldots, U_d) = P\left(\bigcap_{i=1}^d \{U_i = k_i\}\right)
\]

for all

\[
k_{i k_1, k_2, \ldots, k_d} \in \times_{i=1}^d T_i
\]  \tag{8}

Parallel to this definition, let:

\[
\bar{B}_m(x) = \frac{B_m(\beta x)}{\int_0 B_m(\beta x)dx}
\]  \tag{9}

where \(B_m\) are the Boubaker polynomials [16-25] and \(\beta_i\) is the \(i^{\text{th}}\) ordered Boubaker polynomial positive root [19-22].

Then:

\[
c_{\text{BP}}(u_1, \ldots, u_d) = \sum_{k_1=0}^{m-1} \ldots \sum_{k_d=0}^{m-1} p(k_1, \ldots, k_d) \prod_{i=1}^d m B_{m-\ldots}(\beta_i u_i)
\]

with:

\[
(u_1, \ldots, u_d) \in [0,1]^d  \tag{10}
\]

defines the density of a \(d\)-dimensional Boubaker Polynomials copula. Hence, \(c_{\text{BP}}\) is the Boubaker Polynomials copula density induced by \(U\).

The Boubaker Polynomials copula, as a function: \([0,1]^d \rightarrow [0,1]\), verifies the main properties:

- \(c_{\text{BP}}\) is grounded, i.e. for every \(U = (U_1, \ldots, U_d) \in [0,1]^d\), \(c_{\text{BP}}(U) = 0\) if the condition: \(\exists i/U_i = 0\) is ensured.
- \(c_{\text{BP}}\) is \(d\)-increasing, i.e. for every \(U = (U_1, \ldots, U_d) \in [0,1]^d\) and \(V = (V_1, \ldots, V_d) \in [0,1]^d\), the Copula-volume of the box \([U,V]\) is strictly positive.
- \(c_{\text{BP}}(1, \ldots, 1, U_{d+1}, \ldots, U_k) = U_k\).

**Hydrometric data:** In hydrology, the most difficult problem is how to determine the amount of effective rainfall to route. It is a nonlinear problem that involves a variety of hydrological processes, the heterogeneity of rainfall intensities, soil characteristics and antecedent conditions. One of the approximation infiltration process models is the Horton process supposing that runoff is generated by rainfall intensities that are greater than the soil infiltration capacity. The index infiltration method (\(\phi\)-index) represents the average value of infiltration capacity through the duration of rainfall. This method is still largely used for estimating effective rainfall and deducing flood volume for specific rainfall events.

Furthermore, the infiltration capacity and infiltration index are generally a function of the antecedent soil moisture condition. The method of Antecedent Precipitation Index (API) allows the actualization of infiltration index for the runoffs progress during the hydrological year.

Since the \(\phi\)-index is the average value of infiltration capacity \(f\) through the duration \((D_R)\) of rainfall (Eq. (11)), we propose to explore the relationship identified between \(\phi\)-index and rainfall intensity.
Table 1: Characteristics of studied events

<table>
<thead>
<tr>
<th>Event</th>
<th>P (mm)</th>
<th>D (min)</th>
<th>$I_{\max}$ (5mn) (mm/h)</th>
<th>$I_{\moy}$ (5mn) (mm/h)</th>
<th>V (m$^3$)</th>
<th>Qp (m$^3$/s)</th>
<th>tp (min)</th>
<th>tb (min)</th>
<th>$\phi$ (5mn) (mm/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20/5/92</td>
<td>106.0</td>
<td>116</td>
<td>260.0</td>
<td>55</td>
<td>33059.0</td>
<td>34.70</td>
<td>25</td>
<td>60</td>
<td>166.00</td>
</tr>
<tr>
<td>24/5/92</td>
<td>36.0</td>
<td>299</td>
<td>36.0</td>
<td>7</td>
<td>1509.0</td>
<td>0.60</td>
<td>25</td>
<td>85</td>
<td>26.20</td>
</tr>
<tr>
<td>14/9/92</td>
<td>26.0</td>
<td>27</td>
<td>84.0</td>
<td>58</td>
<td>10657.0</td>
<td>3.00</td>
<td>60</td>
<td>120</td>
<td>58.00</td>
</tr>
<tr>
<td>31/7/94</td>
<td>35.5</td>
<td>42</td>
<td>120.0</td>
<td>51</td>
<td>20843.0</td>
<td>19.80</td>
<td>30</td>
<td>60</td>
<td>73.00</td>
</tr>
<tr>
<td>8/1/95</td>
<td>12.0</td>
<td>138</td>
<td>10.0</td>
<td>5</td>
<td>400.0</td>
<td>0.10</td>
<td>60</td>
<td>190</td>
<td>8.98</td>
</tr>
<tr>
<td>8/6/95</td>
<td>14.0</td>
<td>28</td>
<td>58.8</td>
<td>30</td>
<td>1767.6</td>
<td>0.04</td>
<td>-</td>
<td>-</td>
<td>53.00</td>
</tr>
<tr>
<td>24/6/95</td>
<td>11.5</td>
<td>13</td>
<td>101.0</td>
<td>53</td>
<td>3980.0</td>
<td>0.60</td>
<td>50</td>
<td>470</td>
<td>87.80</td>
</tr>
<tr>
<td>24/8/95</td>
<td>12.5</td>
<td>12</td>
<td>102.0</td>
<td>63</td>
<td>41940.0</td>
<td>26.70</td>
<td>20</td>
<td>95</td>
<td>5.50</td>
</tr>
<tr>
<td>4/9/95</td>
<td>39.5</td>
<td>13</td>
<td>324.0</td>
<td>182</td>
<td>67200.0</td>
<td>85.60</td>
<td>15</td>
<td>29</td>
<td>162.00</td>
</tr>
<tr>
<td>4/9/95 bis</td>
<td>8.5</td>
<td>30</td>
<td>33.6</td>
<td>17</td>
<td>15164.0</td>
<td>2.10</td>
<td>50</td>
<td>200</td>
<td>10.60</td>
</tr>
<tr>
<td>16/9/95</td>
<td>7.5</td>
<td>14</td>
<td>56.0</td>
<td>32</td>
<td>16055.0</td>
<td>0.10</td>
<td>60</td>
<td>-</td>
<td>13.10</td>
</tr>
<tr>
<td>7/2/96</td>
<td>8.0</td>
<td>73</td>
<td>31.2</td>
<td>7</td>
<td>3152.0</td>
<td>0.40</td>
<td>5</td>
<td>360</td>
<td>21.30</td>
</tr>
<tr>
<td>15/8/96</td>
<td>39.5</td>
<td>44</td>
<td>115.0</td>
<td>54</td>
<td>1476.0</td>
<td>2.00</td>
<td>10</td>
<td>-</td>
<td>103.20</td>
</tr>
<tr>
<td>9/9/96</td>
<td>12.0</td>
<td>74</td>
<td>56.6</td>
<td>10</td>
<td>15573.0</td>
<td>10.40</td>
<td>25</td>
<td>35</td>
<td>13.10</td>
</tr>
<tr>
<td>9/9/96bis</td>
<td>13.0</td>
<td>53</td>
<td>28.8</td>
<td>15</td>
<td>15030.0</td>
<td>0.10</td>
<td>40</td>
<td>-</td>
<td>13.10</td>
</tr>
<tr>
<td>18/8/97</td>
<td>10.5</td>
<td>26</td>
<td>68.4</td>
<td>24</td>
<td>6338.0</td>
<td>2.60</td>
<td>45</td>
<td>225</td>
<td>48.40</td>
</tr>
<tr>
<td>21/9/97</td>
<td>17.5</td>
<td>21</td>
<td>118.8</td>
<td>50</td>
<td>26393.0</td>
<td>22.30</td>
<td>25</td>
<td>85</td>
<td>38.80</td>
</tr>
<tr>
<td>4/11/97</td>
<td>5.5</td>
<td>16</td>
<td>32.4</td>
<td>20</td>
<td>383.0</td>
<td>0.20</td>
<td>95</td>
<td>240</td>
<td>31.20</td>
</tr>
<tr>
<td>6/12/97</td>
<td>10.5</td>
<td>243</td>
<td>4.8</td>
<td>3</td>
<td>2336.0</td>
<td>0.20</td>
<td>135</td>
<td>335</td>
<td>3.75</td>
</tr>
<tr>
<td>5/8/99</td>
<td>27.0</td>
<td>36</td>
<td>99.6</td>
<td>45</td>
<td>35093.0</td>
<td>7.90</td>
<td>60</td>
<td>390</td>
<td>47.30</td>
</tr>
</tbody>
</table>

P: rainfall depth; D: rainfall duration; $I_{\max}$: rainfall maximum intensity; $I_{\moy}$: rainfall average intensity; V: runoff volume; Qp: peak discharge; tp: peak time; tb: base time; $\phi$: infiltration index

\[ \phi = \frac{1}{D} \int f(t) \, dt \quad (11) \]

Instead of considering the entire hyetograph, we rather focus on modeling the relationship between infiltration index and average rainfall intensity ($I_{\moy}$).

As mentioned above, infiltration is a function of a set of variables, therefore needs the modeling of their joint behavior. On the basis of works of Ellouze-Gargouri et al. [26], we suggest prospecting the nonlinear dependence between $\phi$-index and $I_{\moy}$, using a nonlinear multivariate approach.

The measurements come from a small catchment: Saddine 1. It is located adjacent to Makthar in Tunisia (Northern latitude 35°48'06" and Eastern longitude 9°04'09") in a mountainous zone, monitored by the DGACTA (Direction Générale des Aménagements et Conservation des Terres Agricoles) and the IRD (Institut de Recherche et Développement) from 1992 to 2000, within the framework of the HYDROMED project (Programme de recherche sur les lacs collinaires dans les zones semi-arides du pourtour méditerranéen). This catchment is controlled by a headwater dam whose filling was realized in 1992. We select a sample of 51 (among 55, the four remaining are for validation) observed rainfall hyetographs (I) with a reference time increment $\Delta t$ equal to 5 minutes; thus a set ($I_{\moy}$) has been constituted through the observation period. On other hand, the confrontation of rainfall hyetograph-runoff volume series allows establishing $\phi$-index set.

Besides the marginal distributions of each variable $\phi$-index and $I_{\moy}$ are fitted; the former is exponential distribution with a mean $\mu_\phi = 29.2$ mm/h and a position parameter of $\phi_0 = 3.2$ mm/h; and the latter $I_{\moy}$ is Lognormal distribution with mean $\mu_{I_{\moy}} = 2.8$ mm/h and standard deviation $\sigma_{I_{\moy}} = 0.98$ mm/h.

Data and application of the Boubaker Polynomials copula: In this work, the rank correlation coefficient Kendall’s tau ($\tau$) [2] is used for the characterization of dependence to measure the association between $\phi$-index and $I_{\moy}$ obtained from raw measurements (Table 1). The Kendall coefficient of rank correlation can be used for revealing dependence of two qualitative characteristics, provided that the elements of the sample can be ordered with respect to these characteristics. This coefficient, which measures non linear dependence, integrates the rank of observations rather than their value. Therefore the value itself is not so important than its rank among other values. In summary the more the Kendall’s $\tau$ is high the more the dependence is important.
A test of independence can be adopted for Kendall’s $\tau$, since under the null-hypothesis $H_0$, this statistic is close to normal with zero mean and variance ($n$ size of sample). As a result $H_0$ would be rejected at approximate level $\alpha$ if:

\[
|\tau| > z_{\alpha/2} \sqrt{\frac{2(2n+5)}{9n(n-1)}}
\]  \hspace{1cm} (12)

For $\alpha=5\%$, $z_{0.025} = 1.96$. Let $z^*$ represent the quantity:

\[
z^* = z_{\alpha/2} \sqrt{\frac{2(2n+5)}{9n(n-1)}}
\]  \hspace{1cm} (13)

Table 2 deals with $\tau$’s value and the corresponding statistic for the couple ($\phi$, $I_{moy}$). The analysis of Table shows that $H_0$ independence hypothesis is rejected. Consequently $\phi$ depends on $I_{moy}$.

On the basis of the methodology precedentely performed by Gargouri-Ellouze et al. [26], Gumbel copula has been adopted for ($\phi$, $I_{moy}$) with a parameter $a = 3$. Figure 1 shows the simulated and observed data. The Boubaker Polynomials copulas are applied through

the conjoint distribution ($\phi$, $I_{moy}$) according to Eq. (8-10). The observations are reconstituted; in addition the four excluded hydrological events (20/5/92, 24/5/92, 14/9/93 and 31/7/94) used for the validation have been also reconstituted. Consequently we may assume that this modeling has been validated.

**ANALYSIS AND DISCUSSION**

In the related example, the raw data (‘Observations’ in Fig. 1) exhibited a relative dispersion (Pearson coefficient $\approx 0.61$) in the binary system ($\phi$, $I_{moy}$). For relevant catchments located in Tunisia, an eventual dependence between these two variables has been investigated and the Boubaker polynomials copula has been used in order to smooth the empirical data using a test of goodness-of-fit.

In fact, the average infiltration rate during a rainfall event, given by the $\phi$-index, is supposed to be primarily a function of soil characteristics and antecedent moisture content. Only for comparatively small rainfall amounts, i.e., having intensities lesser than infiltration capacity, the $\phi$-index would also depend on rainfall characteristics. In view of this, it is meaningless and very difficult to obtain a sensible estimate of $\phi$-index by conditioning it only on the average rainfall intensity.

For example, for the actual soil type, $\phi$-index will vary tremendously for the same average rainfall intensity if the antecedent moisture content is very low or very high. Conversely, for very low or very high average
rainfall intensity, the \( \phi \)-index may come out to be of equal magnitude if the antecedent moisture was very low or very high respectively. This indicates that there is little value of conditioning \( \phi \)-index on rainfall.

In this context, a primal accurate sampling process has appropriately yielded a subset of validated measurements (‘Validation’ in Fig. 1). This latter was consequently subjected to the Boubaker Polynomials copula analysis (Continuous line in Fig. 1). The validated profile was in good agreement with the Boubaker Polynomials copula results along with those using the \(*\)-product defined by Darsow et al. [27].

**CONCLUSION**

Copulas are actually a major tool in modelling the dependence structure in hydrological, mass transfer and several other nonlinear systems applications. In this work, we studied a new object in multivariate analysis called the Boubaker Polynomials copula. We showed that, subject to regularity, any copula can be efficiently approximated by some Boubaker Polynomials copula. This copula representation should allow us to take advantage of the properties of the copula function whenever multivariate normality is not a good assumption and application to a standard problem of dependence patterns of infiltration index (\( \phi \)-index) and the average intensity (\( I_{\text{moy}} \)) of rainfall event in a given geographic zone could be successfully carried out. We also showed that modeling joint distributions using copulas relaxes the restriction of selecting marginals for hydrodynamic similar variables.

It is clear that a lot has been left out from this paper. The results can be extended to the right censored data using any centred unbiased Kaplan-Meier-like estimator along with i. e. Akaike information criterion and Bayesian information criteria [28]. A bandwidth choice in practice remains an open question and existing relevant methods like cross-validation or random sampling can be investigated.

**ACKNOWLEDGMENT**

The authors wish to thank the eminent Professor K. Boubaker for helpful comments and correction.

**REFERENCES**