

Article

Third-Order Polynomial Normal Transform Applied to Multivariate Hydrologic Extremes

Yeou-Koung Tung ^{1,*}, Lingwan You ¹ and Chulsang Yoo ² 

¹ Disaster Prevention and Water Environment Research Center, National Chiao Tung University, Hsinchu 300, Taiwan; lucie751111@gmail.com

² Department of Civil, Environment, and Architecture, Korea University, Seoul 02841, Korea; envchul@korea.ac.kr

* Correspondence: yk2013tung@gmail.com

Received: 26 December 2018; Accepted: 4 March 2019; Published: 8 March 2019



Abstract: Hydro-infrastructure systems (e.g., flood control dams, stormwater detention basins, and seawalls) are designed to protect the public against the adverse impacts of various hydrologic extremes (e.g., floods, droughts, and storm surges). In their design and safety evaluation, the characteristics of concerned hydrologic extremes affecting the hydrosystem performance often are described by several interrelated random variables—not just one—that need to be considered simultaneously. These multiple random variables, in practical problems, have a mixture of non-normal distributions of which the joint distribution function is difficult to establish. To tackle problems involving multivariate non-normal variables, one frequently adopted approach is to transform non-normal variables from their original domain to multivariate normal space under which a large wealth of established theories can be utilized. This study presents a framework for practical normal transform based on the third-order polynomial in the context of a multivariate setting. Especially, the study focuses on multivariate third-order polynomial normal transform (TPNT) with explicit consideration of sampling errors in sample L-moments and correlation coefficients. For illustration, the modeling framework is applied to establish an at-site rainfall intensity–duration–frequency (IDF) relationship. Annual maximum rainfall data analyzed contain seven durations (1–72 h) with 27 years of useable records. Numerical application shows that the proposed modeling framework can produce reasonable rainfall IDF relationships by simultaneously treating several correlated rainfall data series and is a viable tool in dealing with multivariate data with a mixture of non-normal distributions.

Keywords: polynomial normal transform; multivariate modeling; sampling errors; non-normality; extreme rainfall analysis

1. Introduction

In hydrosystem design, performance evaluation, and simulation, the problems often involve multiple random variables that are correlated with a mixture of non-normal marginal distributions. Under this condition, it is generally difficult, if not impossible, to establish an analytical joint probability distribution for these variables. In comparison with univariate distributions, there are relatively few analytical multivariate distribution functions under special combinations of parametric marginal distributions, and most of them are of the same type, which can be found in [1,2]. Examples of using analytical multivariate distributions in hydrology are bivariate Gamma distribution [3] and bivariate generalized extreme distribution [4]. Their use is somewhat limited to many practical problems because of different marginal distributions.

Due to the difficulty in establishing a truly multivariate joint distribution model for problems involving mixtures of several correlated, non-normal variables, approximated approaches, such as copula or normal transform, are often used by preserving marginal distributions or moments, including the correlation features among the variables. However, one should realize that, unlike using a true multivariate joint distribution function, preservation of the marginal distributions and dependence structure represents the retention of partial information of the concerned multivariate random variables in the analysis [5].

The concept of copula is one type of approximated multivariate approaches that has recently received tremendous attention and applications by researchers in various disciplines, including in hydrology [6]. Some examples of applying copula in multivariate hydrologic modeling can be found in analyzing floods [7,8], droughts [9–12], dam safety [13], and extreme rainfalls [14,15]. Most of the copula-based applications deal with bivariate problems and some trivariate problems under some restrictive conditions on correlation structures [8]. Applications of copula to higher dimension multivariate problems are rare primarily because there are only a few copula families that are rather restrictive in describing the dependence structure. Recently, the introduction of vine copulas has shown the advantage of overcoming the limitation of currently used copulas in multivariate analysis [16–19]. A copula-based approach is parametric by nature in that analytical marginal distribution models for the involved variables are specified.

Alternatively, another viable scheme in treating multivariate problems involving correlated non-normal random variables is to apply a NORTA (normal-to-anything) algorithm [20]. By a NORTA algorithm, normal transformation of an individual non-normal variable is made by preserving its marginal probability content in the normal variable domain as $\Phi(z) = F_x(x)$ with $\Phi(\cdot)$ and $F_x(\cdot)$, respectively, being the cumulative distribution functions (CDFs) of the standard normal variable Z and the original variable X . In addition, a relationship must be established to allow the determination of an equivalent correlation coefficient, ρ_{z_j, z_k} , of a pair of normal transformed variables, Z_j and Z_k , from the correlation coefficient, ρ_{x_j, x_k} , of the corresponding random variables, X_j and X_k , in the original space. Once the correlation matrix of standard normal variables Z 's, $\{\rho_{z_j, z_k}\}$, is obtained from that of the non-normal variables X 's, $\{\rho_{x_j, x_k}\}$, appropriate orthogonal transformation can be implemented to transform the original correlated variables into uncorrelated standard normal space for analysis.

The determination of ρ_{z_j, z_k} from ρ_{x_j, x_k} is made through the Nataf transform [21], which requires solving an implicit non-linear equation in the form of a double integration involving marginal distributions of a pair of random variables, X_j and X_k , under consideration:

$$\rho_{x_j, x_k} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{x_j - \mu_j}{\sigma_j} \right) \left(\frac{x_k - \mu_k}{\sigma_k} \right) \phi_{jk}(z_j, z_k | \rho_{z_j, z_k}) dz_j dz_k \quad (1)$$

where $x_j = F_j^{-1}[\Phi(z_j)]$, and $\phi_{jk}(\cdot) =$ bivariate standard normal joint probability density function (PDF). Lebrun and Dutfoy [22] provide an insightful analysis of Nataf transform and uncover that it is a special modeling of dependence structure using Gaussian copula. To facilitate practical engineering applications, a set of empirical equations for 10 commonly used distribution functions has been established to relate ρ_{z_j, z_k} to ρ_{x_j, x_k} and their distribution properties [23]. Such empirical relations were applied to reliability analysis of engineering systems [5,24]. Later, computationally more efficient methods based on root finding and linear search [25], the false position method [26], and the artificial neural network method [27] were proposed to solve Equation (1) for ρ_{z_j, z_k} from the known ρ_{x_j, x_k} and marginal PDFs of X_j and X_k .

The above mentioned schemes (i.e., copula, NORTA, and Nataf transform) all require the stipulation of marginal PDFs. The stipulation of a distribution function implies knowing the complete statistical information of the random variable, including its moments of all orders. This ideal situation is attainable only when one has a large amount of data, which generally is not the case in practice.

Therefore, to relax the information requirement without having to specify the distribution functions, third-order polynomial normal transform (TPNT) can be used. By TPNT, each individual non-normal random variable is related to a 3rd-order polynomial function of the corresponding standard normal variable [28]. The polynomial coefficients are determined by matching the statistical moments or quantiles of the individual random variables. The multivariate version of TPNT was first proposed by Vale and Maurelli [29] to simultaneously consider statistical moments and correlation coefficients. A multivariate TPNT procedure has been applied to different fields including, but not limited to, Monte Carlo simulation for generating multivariate random variates [24,30–33], wind power modeling [34], load computation in power network planning [35], and reliability analysis [36,37].

It should be noted that the great majority of multivariate TPNT applications are done under the assumption of known marginal statistical moments (i.e., product-moments and L-moments) and correlation coefficients. However, in real-life hydrologic applications, the amount of available data generally is not sufficiently large to reliably ascertain the true marginal probability distribution functions, statistical moments, and correlation coefficients. Therefore, the sample statistical moments and correlation coefficients used could be subject to sampling errors. In this study, a procedure is proposed to (1) optimally estimate multivariate TPNT coefficients by explicitly incorporating sampling errors associated with the sample moments and correlation coefficients, and (2) comply with a one-to-one monotonicity increasing relation between quantiles of the original and normal transformed variables. The procedure is illustrated by analyzing annual maximum rainfall data series involving seven different durations to establish at-site rainfall intensity–duration–frequency (IDF) and depth–duration–frequency (DDF) relationships.

2. Methods

2.1. Third-Order Polynomial Normal Transform (TPNT)

2.1.1. Univariate TPNT

By TPNT, a univariate non-normal random variable, X , is approximated by the standard normal variable, Z , in the form of a 3rd-order polynomial functional relation as [28]

$$X = \text{TPNT}(Z \mid a_0, a_1, a_2, a_3) = a_0 + a_1Z + a_2Z^2 + a_3Z^3 \tag{2}$$

where $\text{TPNT}(Z \mid a_0, a_1, a_2, a_3)$ denotes the 3rd-order polynomial transform with $a_0, a_1, a_2,$ and a_3 being the transformation coefficients. The TPNT coefficients can be determined by several methods of varying mathematical complexity. By preserving the first four product-moments, the TPNT coefficients are related to the first four product moments of the standardized variable, $X' = (X - \mu_x)/\sigma_x$, as [28]

$$0 = a_0 + a_2 \tag{3}$$

$$1 = a_1^2 + 6a_1a_3 + 2a_2^2 + 15a_3^2 \tag{4}$$

$$\gamma_x = 2a_2(a_1^2 + 24a_1a_3 + 105a_3^2 + 2) \tag{5}$$

$$\kappa_x = 3 + 24[a_1a_3 + a_2^2(1 + a_1^2 + 28a_1a_3) + a_3^2(12 + 48a_1a_3 + 141a_2^2 + 225a_3^2)] \tag{6}$$

in which γ_x = skew coefficient; κ_x = kurtosis of the original random variable X . Alternatively, the TPNT coefficients in Equation (2) can also be related to the first four L-moments as [38]

$$a_0 + a_2 = \lambda_1 \tag{7}$$

$$0.5642a_1 + 1.4104a_3 = \lambda_2 \tag{8}$$

$$0.5513a_2 \lambda_3 \tag{9}$$

$$0.0692a_1 + 0.8078a_3 = \lambda_4 \quad (10)$$

in which λ_m = the m th order L-moment [39] of the original non-normal random variable, X . Other than the above two moment-matching methods, TPNT coefficients can also be determined by the quantile-based least square method and the Fisher–Cornish asymptotic expansion (FC) method [40]. Chen and Tung [41] investigated the performance of different methods in determining the TPNT coefficients with regard to their accuracy and robustness in capturing the probabilistic features of the random variable X under the condition that the population distribution is known. It was found that, among the various methods for estimating TPNT coefficients, the L-moment based method is computational simplistic and can yield a satisfactory performance under a wide range of distribution conditions. The product-moment method can also yield a satisfactory normal transformation provided that accurate estimations of skew coefficient and kurtosis in Equations (3)–(6) can be made. However, when the statistical moments are to be estimated from finite data, the sample L-moments have been proven to be more stable and robust than those of product-moments [42], especially when the sample size is not large.

By referring to Equations (3)–(6), one also realizes that determining TPNT coefficients based on the product-moments requires solving a system of non-linear equations. It is expected that solving Equations (3)–(6) would be more difficult than solving L-moments based on Equations (7)–(10), which is linear. Sometimes, the solution to the system of non-linear equation may not be attainable. According to Equations (7)–(10), TPNT coefficients can be easily obtained in terms of L-moments as

$$a_0 = \lambda_1 - 1.8138\lambda_3 \quad (11)$$

$$a_1 = 2.2552\lambda_2 - 3.9376\lambda_4 \quad (12)$$

$$a_2 = 1.8138\lambda_3 \quad (13)$$

$$a_3 = -0.1931\lambda_2 + 1.5751\lambda_4 \quad (14)$$

In the transformation process, it is necessary to preserve probability content in both original space and standard normal space, i.e., $F_X(x_p) = \Phi(z_p) = p$. This implies that quantiles of the two variables should satisfy the following relationship:

$$x_p = a_0 + a_1z_p + a_2z_p^2 + a_3z_p^3 \quad (15)$$

where x_p and z_p = p th-order quantiles of random variable X and standard normal random variable, Z , respectively, that is, $x_p = F^{-1}(p)$ and $z_p = \Phi^{-1}(p)$. Furthermore, inherently embedded in Equation (15) is a requirement of one-to-one monotonically increasing relations between x_p and z_p . This, then, requires that TPNT coefficients must comply with the following conditions:

$$a_3 > 0 \text{ and } a_2^2 - 3a_1a_3 < 0. \quad (16)$$

It should be noted that the TPNT coefficients obtained from solving Equations (3)–(6), Equations (7)–(10), or other methods mentioned above do not guarantee the compliance of the monotonicity condition stipulated in Equation (16). This is especially a major concern when sample statistics are used in determining TPNT coefficients.

2.1.2. Multivariate TPNT

The TPNT coefficients can be determined by preserving the statistical moments of individual random variables. Specifically, L-moments are used herein to determine the multivariate TPNT coefficients due to simple, linear functional relationships between the TPNT coefficients and the L-moments as shown in Equations (11)–(14). Furthermore, sample L-moments have several desirable

sampling properties over the product-moments as proven by Hosking [42]. In the context of fitting the first four L-moments of a total of N correlated variables, Equations (7)–(10) can be re-written as

$$a_{0j} + a_{2j} = \lambda_{1j} \tag{17}$$

$$0.5642a_{1j} + 1.4104a_{3j} = \lambda_{2j} \tag{18}$$

$$0.5513a_{2j} \lambda_{3j} \tag{19}$$

$$0.0692a_{1j} + 0.8078a_{3j} = \lambda_{4j} \tag{20}$$

in which λ_{mj} = the m th order L-moment of the j th random variable X_j for $j = 1, 2, \dots, N$.

In addition to preserving marginal statistical moments of involved variables, multivariate TPNT must also simultaneously preserve the statistical dependence between random variables in the transformation. The correlation coefficient of any two correlated random variables, X_j and X_k , is imbedded in their 2nd-order cross-product moment of which Vale and Maurelli [29] had shown the explicit expressions in terms of TPNT coefficients as

$$CP_{j,k}(a_j, a_k; \rho_{z_j, z_k}) = E[X_j X_k] = \mu_j \mu_k - \rho_{x_j, x_k} \sigma_j \sigma_k \tag{21}$$

$$= (6a_{3j}a_{3k}) \rho_{z_j, z_k}^3 + (2a_{2j}a_{2k}) \rho_{z_j, z_k}^2 + [(a_{1j} + 3a_{3j})(a_{1k} + 3a_{3k})] \rho_{z_j, z_k} + [(a_{0j} + a_{2j})(a_{0k} + a_{2k})]$$

in which ρ_{x_j, x_k} , ρ_{z_j, z_k} = correlation coefficient of random variables (X_j, X_k) and its equivalent (Z_j, Z_k) in normal space; μ_j and σ_j = mean and standard deviation of random variable X_j , respectively. The correlation coefficient in the original scale, ρ_{x_j, x_k} , is related to its counterpart in the normal space, ρ_{z_j, z_k} , in a 3rd-order polynomial relationship through TPNT coefficients.

Upon the determination of TPNT coefficients for the two concerned random variables, the correlation coefficient in the normal space, ρ_{z_j, z_k} , corresponding to that in the original space, ρ_{x_j, x_k} , can be obtained by finding the real root of Equation (21). The mathematical relations between the two correlation coefficients are [20]

$$\rho_{x_j, x_k} \times \rho_{z_j, z_k} > 0; \quad |\rho_{x_j, x_k}| \leq |\rho_{z_j, z_k}| \tag{22}$$

Equation (21) is used repeatedly to solve for ρ_{z_j, z_k} for all pairs of correlated random variables to establish the correlation matrix in multivariate normal space.

2.2. Optimization Framework for Determining Multivariate TPNT Coefficients

2.2.1. Objective Function

To determine the multivariate TPNT coefficients that best preserve the known values of L-moments, the least-square criterion is used in the study by which the objective function can be expressed as

$$Minimize \sum_{m=1}^4 \sum_{j=1}^N \delta_{mj}^2 \tag{23}$$

where δ_{mj} = a decision variable defining the deviation between the m th-order TPNT-based L-moments computed by the left-hand side of Equations (17)–(20) and the known values, λ_{mj} , of the j th random variable, X_j . Of course, other forms of objective function, such as minimizing the sum of absolute deviations, can be used.

2.2.2. Constraints

Several constraints are essential to make sure that multivariate TPNT coefficients obtained are able to preserve the known statistical features and mathematical relationships of concerned random variables.

(a) Preservation of L-moments for the individual variable X_j :

The deviation δ_{mj} in the objective function defining the degree of preserving the known values of the first-four L-moments of individual variable X_j can be written, according to Equations (17)–(20), as

$$a_{0j} + a_{2j} + \delta_{1j} = \lambda_{1j} \tag{24}$$

$$0.5642a_{1j} + 1.4104a_{3j} + \delta_{2j} = \lambda_{2j} \tag{25}$$

$$0.5513a_{2j} + \delta_{3j} = \lambda_{3j} \tag{26}$$

$$0.0692a_{1j} + 0.8078a_{3j} + \delta_{4j} = \lambda_{4j}. \tag{27}$$

Note that the value of δ_{mj} is unrestricted-in-sign, meaning that its value can be negative, zero, and positive, depending on the relative magnitudes of TPNT-based L-moments and those of the known values.

In reality, statistical properties of a random variable are estimated from a finite number of sample data. Consequently, sample L-moments of random variables, X_j , are subject to uncertainty. In practice, two approaches are used to estimate sample L-moments: plotting position-based estimators and unbiased estimators. This study adopts the latter by which the first four unbiased estimators of L-moments, $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$, can be computed respectively as [42]

$$\ell_1 = \binom{n}{1}^{-1} \sum_{i=1}^n x_{i:n} \tag{28}$$

$$\ell_2 = \frac{1}{2!} \binom{n}{2}^{-1} \sum_{i=1}^n \left\{ \binom{i-1}{1} - \binom{n-i}{1} \right\} x_{i:n} \tag{29}$$

$$\ell_3 = \frac{1}{3!} \binom{n}{3}^{-1} \sum_{i=1}^n \left\{ \binom{i-1}{2} - 2 \binom{i-1}{1} \binom{n-i}{1} + \binom{n-i}{2} \right\} x_{i:n} \tag{30}$$

$$\ell_4 = \frac{1}{4!} \binom{n}{4}^{-1} \sum_{i=1}^n \left\{ \binom{i-1}{3} - 3 \binom{i-1}{2} \binom{n-i}{1} + 3 \binom{i-1}{1} \binom{n-i}{2} - \binom{n-i}{3} \right\} x_{i:n} \tag{31}$$

in which ℓ_m = sample estimator of the m th-order L-moment, λ_m ; $x_{i:n}$ = the i th ranked sample (ascending order) in a data of size n .

Suppose that the sampling distributions of sample L-moments are derived or approximated. Proper bounds can then be incorporated into Equations (24)–(27) for determining the suitable and probabilistically plausible TPNT coefficients for all N random variables X_1, X_2, \dots, X_N . Assuming that the lower and upper bounds of the L-moments can be determined from their corresponding sampling distributions, constraint Equations (24)–(27) then can be modified as

$$\ell_{1j}^{(L)} \leq a_{0j} + a_{2j} + \delta_{1j} \leq \ell_{1j}^{(U)} \tag{32}$$

$$\ell_{2j}^{(L)} \leq 0.5642a_{1j} + 1.4104a_{3j} + \delta_{2j} \leq \ell_{2j}^{(U)} \tag{33}$$

$$\ell_{3j}^{(L)} \leq 0.5513a_{2j} + \delta_{3j} \leq \ell_{3j}^{(U)} \tag{34}$$

$$\ell_{4j}^{(L)} \leq 0.0692a_{1j} + 0.8078a_{3j} + \delta_{4j} \leq \ell_{4j}^{(U)} \tag{35}$$

for $j = 1, 2, \dots, N$. In Equations (32)–(35), $\ell_{mj}^{(U)}$ and $\ell_{mj}^{(L)}$ are, respectively, the upper and lower bounds containing the unknown population m th-order L-moment of random variable X_j , λ_{mj} . The derivation of bounds for unknown population L-moments is described in Section 2.3.1.

(b) Preservation of the monotonic probability–quantile relationship for individual variable X_j :

$$a_{3j} > 0; a_{2j}^2 - 3a_{1j}a_{2j} < 0, \text{ for } j = 1, 2, \dots, N \tag{36}$$

(c) Preservation of the correlation between all pairs of different variables, X_j and X_k :

Based on Equation (21), any pair of two correlated random variables X_j and X_k must satisfy the following equation.

$$CP_{j,k}(a_j, a_k; \rho_{z_j, z_k}) - (\mu_j \mu_k - \rho_{x_j, x_k} \sigma_j \sigma_k) = 0, \text{ for all variable pairs } j \neq k \tag{37}$$

where $CP_{j,k}(a_j, a_j; \rho_{z_j, z_k}) = E(X_j, X_k) =$ cross-product moment of variables, X_j and X_k , defined in Equation (21), which are functions of the corresponding TPNT coefficients $a_j = (a_{0j}, a_{1j}, a_{2j}, a_{3j})$ and $a_k = (a_{0k}, a_{1k}, a_{2k}, a_{3k})$; ρ_{x_j, x_k} and ρ_{z_j, z_k} = correlation coefficients of random variables, X_j and X_k , and their normal equivalents, Z_j and Z_k , respectively; μ_j, σ_j = mean and standard deviation of random variable X_j , respectively.

Similarly, constraint Equation (37) on correlation coefficients can be modified as

$$r_{x_j, x_k}^{(L)} \leq \frac{CP_{j,k}(a_j, a_k; r_{z_j, z_k}) - m_j(a_j)m_k(a_k)}{s_j(a_j) \times s_k(a_k)} \leq r_{x_j, x_k}^{(U)}, \text{ for all variable pairs } j \neq k \tag{38}$$

in which $r_{x_j, x_k}^{(L)}, r_{x_j, x_k}^{(U)}$ = lower and upper bounds, respectively, of the unknown population correlation coefficient, ρ_{x_j, x_k} , between the random variables X_j and X_k (see Section 2.3.2); r_{z_j, z_k} = equivalent correlation coefficient of the random variables in the standard normal domain; $m_j(a_j), s_j(a_j)$ = TPNT-based estimation of mean and standard deviation of random variables X_j which can be computed according to Equations (3) and (4) as

$$m_j(a_j) = a_{0j} + a_{2j} \tag{39}$$

$$s_j^2(a_j) = a_{1j}^2 + 6a_{1j}a_{3j} + 2a_{2j}^2 + 15a_{3j}^2. \tag{40}$$

Equation (38) can alternatively be expressed as

$$CP_{j,k}(a_j, a_k; r_{z_j, z_k}) - m_j(a_j) m_k(a_k) - r_{x_j, x_k}^{(U)} s_j(a_j) s_k(a_k) \leq 0 \tag{41}$$

$$CP_{j,k}(a_j, a_k; r_{z_j, z_k}) - m_j(a_j) m_k(a_k) - r_{x_j, x_k}^{(L)} s_j(a_j) s_k(a_k) \geq 0 \tag{42}$$

In summary, by considering sampling errors of sample L-moments and correlation coefficients, the optimization model to determine the most plausible TPNT coefficients for establishing multivariate relationships can be summarized as follows:

The objective function is expressed in Equation (23) or its variations, which is subject to the following constraints:

- Equations (32)–(35) for preserving plausible L-moments ($8 \times N$ constraints);
- Equation (36) for complying with a probability–quantile monotonic relationship ($2 \times N$ constraints);
- Equations (41) and (42) for preserving plausible correlation coefficient ($N \times (N - 1)$ constraints); and
- unrestrictive-in-sign of polynomial coefficients $(a_{0j}, a_{1j}, a_{2j}, a_{3j})$ and deviations δ_{mj} .

2.3. Determination of Bounds for L-Moments and Correlation Coefficients

2.3.1. Bounds for L-Moments

To determine the bounds for L-moments, the sampling distributions corresponding to the sample L-moments are needed. For independent random samples of size n from a distribution function $F_x(x)$ having the m th-order population L-moment λ_m , Hosking [39] showed that the statistic $n^{1/2}(\ell_m - \lambda_m)$, with ℓ_m being the sample L-moment of order m , is unbiased, having a sampling distribution asymptotically converge to the normal distribution with the mean zero and variance Λ_{mm} . Therefore, the variance of the m th-order sample L-moment, ℓ_m , has the variance of $\sigma^2(\ell_m) = \Lambda_{mm}/n$. For the first four orders of sample L-moment, the value of Λ_{mm} can be computed by

$$\Lambda_{11} = \iint_{x < y} (y - x)^2 du dv \tag{43}$$

$$\Lambda_{22} = \iint_{x < y} [(3 - 4v)(1 - 4u)](y - x)^2 du dv \tag{44}$$

$$\Lambda_{33} = \iint_{x < y} [(7 - 24v + 18v^2)(1 - 12u + 18u^2)](y - x)^2 du dv \tag{45}$$

$$\Lambda_{44} = \iint_{x < y} [(-13 + 84v - 150v^2 + 80v^3)(-1 + 24u - 90u^2 + 80u^3)](y - x)^2 du dv \tag{46}$$

in which $u = F_x(x)$ and $v = F_x(y)$. To estimate the values of Λ_{mm} based on the ranked sample observations, the double integration stated in Equations (43)–(46) can be carried out numerically as

$$\hat{\Lambda}_{11} = \frac{1}{n(n-1)} \sum_{i=1}^{n-1} \sum_{k=i+1}^n [x_{(k)} - x_{(i)}]^2 \tag{47}$$

$$\hat{\Lambda}_{22} = \frac{1}{n(n-1)} \sum_{i=1}^{n-1} \sum_{k=i+1}^n \left\{ (3 - 4p_{k:n})(1 - 4p_{i:n}) [x_{(k)} - x_{(i)}]^2 \right\} \tag{48}$$

$$\hat{\Lambda}_{33} = \frac{1}{n(n-1)} \sum_{i=1}^{n-1} \sum_{k=i+1}^n \left\{ (7 - 24p_{k:n} + 18p_{k:n}^2)(1 - 12p_{i:n} + 18p_{i:n}^2) [x_{(k)} - x_{(i)}]^2 \right\} \tag{49}$$

$$\hat{\Lambda}_{44} = \frac{1}{n(n-1)} \sum_{i=1}^{n-1} \sum_{k=i+1}^n \left\{ (-13 + 84p_{k:n} - 150p_{k:n}^2 + 80p_{k:n}^3)(-1 + 24p_{i:n} - 90p_{i:n}^2 + 80p_{i:n}^3) [x_{(k)} - x_{(i)}]^2 \right\} \tag{50}$$

in which $x_{(i)}$ = the i th ranked sample in ascending order, i.e., $x_{(1)} < x_{(2)} < \dots < x_{(i)} < \dots < x_{(k)} < \dots < x_{(n)}$; $p_{i:n}$ = estimated cumulative probability for the i th ranked sample, i.e., $\Pr[X \leq x_{(i)}]$, by using the well-known Weibull plotting position formula, $p_{i:n} = i/(n + 1)$. Makkonen [43] has shown that the Weibull plotting position formula [44] provides the best estimate for the underlying non-exceedance probability. The superiority of the Weibull formula gets more pronounced with a decreasing sample size. By adopting the normality distribution assumption, the α -confidence interval for the unknown population λ_k can be obtained as

$$\left(\ell_m^{(L)}, \ell_m^{(U)} \right) = \left(\ell_m - z_{\alpha/2} \sqrt{\frac{\Lambda_{mm}}{n}}, \ell_m + z_{\alpha/2} \sqrt{\frac{\Lambda_{mm}}{n}} \right) \tag{51}$$

in which $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$, a standard normal quantile with an exceedance probability of $\alpha/2$, with $\Phi^{-1}(\cdot)$ being the inverse standard normal CDF.

2.3.2. Bounds for Correlation Coefficients

To quantify the lower and upper bounds of a correlation coefficient, Fisher transform is often used by which the sampling distribution of the inverse hyperbolic tangent function of sample correlation approximately follows a normal distribution as [45,46]

$$\tanh^{-1}(r) = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right) \sim \text{Normal}\left(\frac{1}{2} \ln\left(\frac{1+\rho}{1-\rho}\right), \frac{1}{n-3}\right) \tag{52}$$

in which r, ρ = sample and population correlation coefficients, respectively; n = number of sample pairs. With a specified confidence level α , the corresponding lower and upper bounds for the unknown population coefficient ρ can be obtained as

$$\left[r^{(L)}, r^{(U)}\right] = \left[\tanh\left(\tanh^{-1}(r) - \frac{z_{\alpha/2}}{\sqrt{n-3}}\right), \tanh\left(\tanh^{-1}(r) + \frac{z_{\alpha/2}}{\sqrt{n-3}}\right)\right] \tag{53}$$

where the hyperbolic tangent function is defined as $\tanh(\theta) = (e^{2\theta} - 1) / (e^{2\theta} + 1)$.

2.4. Solution Algorithm

A recursive procedure is proposed to solve the above optimization models for determining multivariate TPNT coefficients. The procedure consists of four steps of initialization, optimization, validation, and updating. Solution algorithm for determining multivariate TPNT coefficients considering sampling errors of L-moments and correlation coefficients is detailed below and outlined in Figure 1.

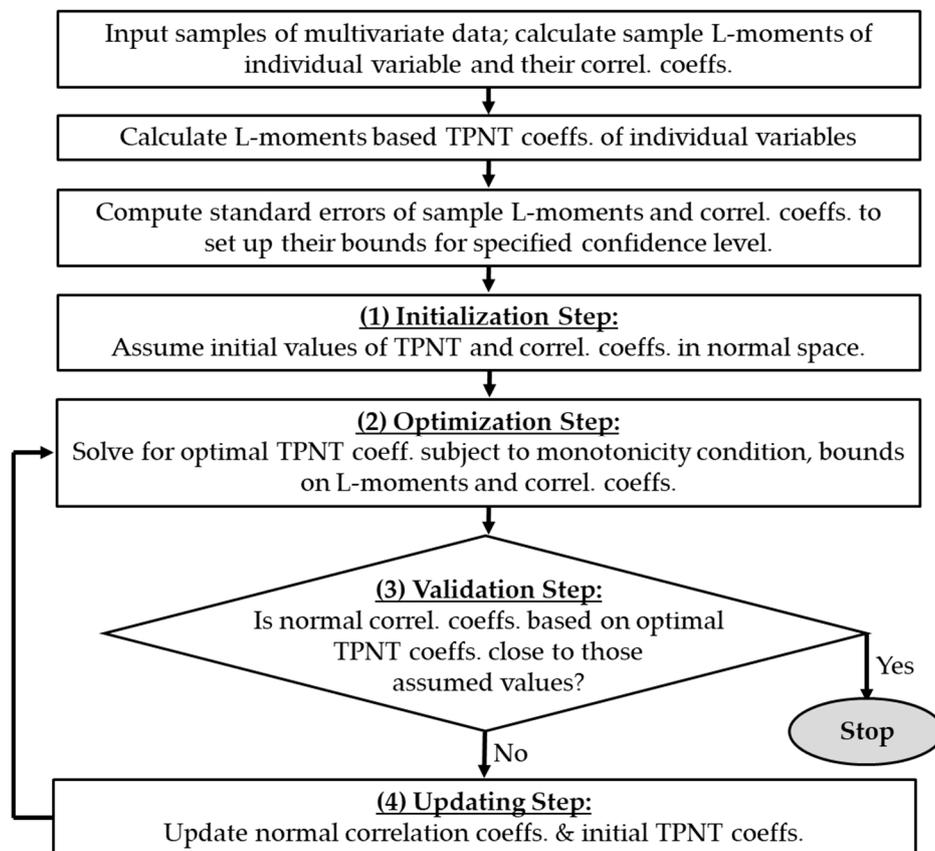


Figure 1. Flow diagram showing the procedure of multivariate TPNT modeling considering sampling errors of sample statistics.

Step (1): Initialization—Since the problem involves a nonlinear optimization model, an initial solution would be needed. One straightforward and sound initial solutions for the TPNT coefficients, $a_j^{(0)}$, are those provided by Equations (11)–(14) for all variables $j = 1, 2, \dots, N$. The initial TPNT coefficients obtained this way will automatically satisfy the constraint Equations (24)–(27). However, they do not necessarily comply with the monotonicity constraint Equation (36). As for the initial normal correlation coefficients, rather than arbitrarily choosing a set of initial correlation coefficients, let $\{\rho_{z_j, z_k}^{(0)}\} = \{r_{x_j, x_k}\}$ in which r_{x_j, x_k} is the sample correlation coefficient between random variables X_j and X_k . Alternatively, obtain a feasible set of initial $\{\rho_{z_j, z_k}^{(0)}\}$ by solving the 3rd-order polynomial function of ρ_{z_j, z_k} in Equation (21) according to the initially assumed TPNT coefficients, $a_j^{(0)}$.

Step (2): Optimization—Based on the initially adopted TPNT coefficients, $a_j^{(0)}$, and the normal transformed correlation coefficients $\{\rho_{z_j, z_k}^{(0)}\}$, solve the optimization model with objective function Equation (23) and constraint Equations (32)–(35), (36), and (41)–(42) for the optimal TPNT coefficients $a_j^* = (a_{0j}^*, a_{1j}^*, a_{2j}^*, a_{3j}^*)$ for random variable X_j , $j = 1, 2, \dots, N$.

Step (3): Validation—From the optimal feasible TPNT coefficients $a_j^* = (a_{0j}^*, a_{1j}^*, a_{2j}^*, a_{3j}^*)$ for $j = 1, 2, \dots, N$ obtained from Step (2), determine the equivalent normal variates corresponding to the sample data by solving the 3rd-order polynomial function:

$$a_{0j}^* + a_{1j}^* z_{j,i} + a_{2j}^* z_{j,i}^2 + a_{3j}^* z_{j,i}^3 = x_{j,i}, \text{ for all } j = 1 \sim N; i = 1 \sim n \quad (54)$$

where $z_{j,i}$ = unknown normal variate corresponding to the i th observation of the j th random variable $x_{j,i}$ under the optimal set of TNPT coefficients $a_j^* = (a_{0j}^*, a_{1j}^*, a_{2j}^*, a_{3j}^*)$. From the normal-transformed data series of two different variables, $z_j = (z_{j,1}, z_{j,2}, \dots, z_{j,n})$ and $z_k = (z_{k,1}, z_{k,2}, \dots, z_{k,n})$, the corresponding correlation coefficient, ρ_{z_j, z_k}^* , in the normal space is calculated.

Step (4): Updating—Compare the discrepancies between the initialized $\rho_{z_j, z_k}^{(0)}$ and validated ρ_{z_j, z_k}^* for all different pairs of concerned random variables. If the discrepancy in any pair of durations is judged to be significant, update the initial normal correlation as $\rho_{z_j, z_k}^{(0)} = \rho_{z_j, z_k}^*$ and TPNT coefficients $a_j^{(0)} = a_j^*$, and the process from Steps (2)–(4) is repeated. Otherwise, the optimal solutions are obtained and the iteration stops.

With regard to the optimization step presented in Step (2), the sequential quadratic programming (SQP) algorithm is implemented [47]. The SQP tackles a nonlinear optimization problem by successively finding the approximated optimum solution to the quadratic programming (QP) representation of the original problem. The approximated solution is improved iteratively by solving the QP problem. Boggs and Tolle [48] elaborated some useful properties of the SQP algorithm. The subroutine “sqp.m” in Matlab is used in this study to solve the optimization model.

3. Numerical Example

In this section, at-site rainfall intensity–duration–frequency (IDF) and depth–duration–frequency (DDF) relations are established to demonstrate the proposed multivariate TPNT method and examine its general performance. Rainfall IDF relations are widely used in the planning, design, and management of hydrosystem infrastructures, such as stormwater sewer systems and detention basins [49,50]. Such relations at a given location involves at-site frequency analysis of annual maximum rainfall intensity (or depth) data of several selected durations. The conventional approach in rainfall frequency analysis chooses a proper parametric probability distribution model to individually fit the observed annual maximum rainfall data of different durations. The choice of a distribution model for the rainfall intensity–frequency relations is largely statistical without much physical justification [51].

By the conventional approach, resulting rainfall intensity–frequency curves of different durations could sometimes intersect within the probability range of practical application. The crossover

phenomenon often occurs when data record length is relatively short. According to the physical reality, rainfall intensity–frequency curves of different durations should not crossover or intersect. Porras and Porras [52] attributed the occurrence of crossover of rainfall IDF curves to short record data of questionable representation in which a significant amount of sampling errors existed in the estimated rainfall quantiles by frequency analysis. One other plausible reason for the possible crossover of IDF curves is that frequency analysis of rainfall data is performed separately for each duration without considering the inter-correlations that are intrinsically embedded in rainfall data of different durations. Haktanir [53] earlier pointed out that rainfall frequency analysis of different durations in the process of establishing IDF relationships should not be performed independently of each other, but did not propose a mechanism to handle the correlation directly. Recently, Gräler et al. [54] applied D-vine copula, along with the generalized extreme value distributions, to derive rainfall IDF relationships based on rainfalls of five durations. You and Tung [55], under the TPNT framework, developed a constrained least square model to simultaneously considering rainfall data of seven durations for establishing at-site rainfall IDF relations. However, their model does not explicitly take into account the correlation among rainfall data of different durations.

The multivariate TPNT-based model presented above was applied to establish at-site rainfall IDF relations using annual maximum hourly rainfall data of various durations at a raingauge in Zhongli City of Taoyuan County, Taiwan. Annual maximum rainfall intensity data cover the record period of 1988–2015, but the year 1992 was excluded from the analysis due to long periods of registers with technical issues. Hence, only 27-year data ($n = 27$) with seven ($N = 7$) durations (i.e., 1, 2, 6, 12, 24, 48, and 72 h) are used in this illustration (see data in Table 1). The sample values of the mean, standard deviation, and first-four L-moments of rainfall data of different durations are tabulated in Table 2. Furthermore, the standard error values corresponding to the first four sample L-moments, according to Equations (47)–(50), are listed in Table 3. The sample correlation coefficients of all rainfall intensity pairs of different durations in the original and normal-transformed domains are shown in Tables 3 and 4, respectively. Based on the information given in Tables 2 and 4, one is able to define the lower and upper bounds for the L-moments and correlation coefficients according to the desired confidence level, α , by Equations (51) and (53), respectively. Table 5 lists the values of correlation coefficients in normal-transformed space, r_{z_j, z_k} , provided by the solution to constraint Equations (41) and (42) in the optimization model.

Under different constraint types and confidence levels for the L-moments and correlation coefficients, the corresponding optimal multivariate TPNT coefficients can vary. With the confidence level of $\alpha = 90\%$ for both L-moments and correlation coefficients, Table 6a–d list the optimal TPNT coefficients under four different constraint types, including “LM” for L-moments by Equations (32)–(35), “Mono” for monotonicity by Equation (36), “Corr” for correlation by Equations (41) and (42), and “NC” for no-crossover by Equation (56). Once the optimal TPNT coefficients associated with each rainfall duration are obtained from solving the multivariate TPNT model, the rainfall IDF relations, according to Equation (15), can be established as

$$i_{t,T}^* = a_{0,t}^* + a_{1,t}^* z_T + a_{2,t}^* z_T^2 + a_{3,t}^* z_T^3 \tag{55}$$

where $i_{t,T}^*$ = estimated t -h, T -year rainfall intensity; $a_{0,t}^*$, $a_{1,t}^*$, $a_{2,t}^*$, and $a_{3,t}^*$ = optimum TPNT coefficients corresponding to rainfall of duration t (h); z_T is the standard normal quantile corresponding to return period T -year having an annual exceedance probability of $1 - \Phi(z_T) = 1/T$.

Table 1. Annual maximum rainfall intensities (mm/h) at Zhongli Station, Taiwan.

Year	1 h	2 h	6 h	12 h	24 h	48 h	72 h
1988	48.0	41.0	18.1	9.0	4.6	2.9	2.2
1989	83.5	69.5	38.2	19.9	10.0	6.1	4.1
1990	89.0	51.3	25.1	14.0	8.1	5.7	3.8
1991	52.5	36.3	13.1	6.8	4.4	3.9	2.9
1993	73.0	50.5	25.8	15.5	8.1	4.3	3.0
1994	44.5	31.0	19.1	11.7	8.7	5.0	3.3
1995	87.5	59.0	27.1	13.7	7.0	4.7	3.5
1996	80.5	40.3	14.0	10.3	8.0	4.6	3.1
1997	35.5	22.8	16.4	11.0	6.3	3.7	2.7
1998	70.5	35.3	21.2	11.1	7.8	4.5	3.0
1999	58.5	29.3	12.2	7.7	5.5	2.9	2.0
2000	43.5	25.8	19.8	14.0	11.2	8.2	5.8
2001	49.5	43.5	28.0	21.5	14.3	12.4	8.7
2002	35.0	28.3	12.4	6.9	5.0	2.7	1.8
2003	30.5	20.5	9.2	7.1	4.7	2.6	1.7
2004	51.5	35.3	19.4	13.3	9.6	6.1	4.1
2005	38.0	30.3	14.8	9.3	7.6	4.6	3.7
2006	48.5	32.3	19.7	13.8	7.7	5.3	4.6
2007	65.5	60.5	24.9	12.6	8.4	6.5	5.3
2008	42.0	32.3	13.5	9.5	7.1	5.5	4.8
2009	33.0	29.3	14.9	10.7	6.2	3.2	2.6
2010	47.5	35.0	13.6	9.3	7.0	3.6	2.4
2011	71.5	48.8	21.0	12.6	7.1	3.7	2.6
2012	72.0	47.8	43.4	34.3	17.6	8.8	5.9
2013	52.0	46.0	25.8	14.5	7.6	6.7	4.5
2014	41.0	26.3	15.0	11.3	7.9	4.9	3.6
2015	35.5	25.5	15.6	10.3	5.6	3.2	2.1

Table 2. Sample moments (in mm/h) of rainfall intensity data.

Moments	1 h	2 h	6 h	12 h	24 h	48 h	72 h
$\hat{\mu}$	54.80	38.26	20.04	12.65	7.89	5.05	3.63
$\hat{\sigma}$	17.8	12.4	7.9	5.6	2.9	2.2	1.5
l_1	54.80	38.26	20.04	12.65	7.89	5.05	3.63
l_2	10.21	7.02	4.25	2.67	1.46	1.13	0.83
l_3	1.673	1.393	1.110	0.879	0.389	0.317	0.200
l_4	0.322	0.715	0.706	0.863	0.460	0.246	0.148

Table 3. Standard errors (in mm/h) of sample L-moments of rainfall intensity data.

Std. Error	1 h	2 h	6 h	12 h	24 h	48 h	72 h
$se(l_1)$	3.423	2.386	1.519	1.071	0.554	0.415	0.296
$se(l_2)$	1.110	0.952	0.750	0.682	0.324	0.226	0.148
$se(l_3)$	1.005	0.660	0.447	0.389	0.159	0.127	0.089
$se(l_4)$	0.713	0.404	0.209	0.182	0.088	0.066	0.050

Table 4. Sample correlation coefficients between rainfall intensity of different durations.

Duration	1 h	2 h	6 h	12 h	24 h	48 h	72 h
1 h	1.000						
2 h	0.826	1.000					
6 h	0.624	0.760	1.000				
12 h	0.421	0.496	0.910	1.000			
24 h	0.306	0.335	0.762	0.915	1.000		
48 h	0.210	0.338	0.635	0.732	0.869	1.000	
72 h	0.161	0.324	0.576	0.671	0.812	0.976	1.000

Table 5. Correlation coefficients of normal-transformed rainfall intensity of different durations.

Duration	1 h	2 h	6 h	12 h	24 h	48 h	72 h
1 h	1.000						
2 h	0.849	1.000					
6 h	0.654	0.768	1.000				
12 h	0.511	0.569	0.918	1.000			
24 h	0.430	0.440	0.736	0.870	1.000		
48 h	0.396	0.462	0.713	0.796	0.903	1.000	
72 h	0.323	0.438	0.670	0.755	0.854	0.979	1.000

Table 6. Multivariate TPNT coefficients obtained under different constraints with $\alpha = 0.90$.

(a) Constraints: L-moments only (LM)							
TPNT Coefficients	1 h	2 h	6 h	12 h	24 h	48 h	72 h
a_0	51.76	35.73	18.03	11.06	7.19	4.47	3.27
a_1	21.76	13.02	6.80	2.62	1.47	1.58	1.29
a_2	3.03	2.53	2.01	1.59	0.70	0.58	0.36
a_3	−1.465	−0.230	0.292	0.843	0.444	0.170	0.072
$a_3 > 0$	−1.46 *	−0.23 *	0.29	0.84	0.44	0.17	0.07
$a_2^2 - 3a_1a_2 < 0$	104.8 *	15.4 *	−1.90	−4.09	−1.46	−0.47	−0.15
(b) Constraints: L-moments and Monotonicity (LM/Mono)							
	1 h	2 h	6 h	12 h	24 h	48 h	72 h
a_0	51.99	35.81	18.03	11.06	7.19	4.47	3.27
a_1	17.52	11.98	6.80	2.62	1.47	1.58	1.29
a_2	2.80	2.45	2.01	1.59	0.70	0.58	0.36
a_3	0.150	0.167	0.292	0.843	0.444	0.170	0.072
$a_3 > 0$	0.15	0.17	0.29	0.84	0.44	0.17	0.07
$a_2^2 - 3a_1a_2 < 0$	−0.01	−0.01	−1.9	−4.1	−1.5	−0.5	−0.1
(c) Constraints: L-moments, Monotonicity, and Correlation (LM/Mono/Corr)							
	1 h	2 h	6 h	12 h	24 h	48 h	72 h
a_0	51.76	35.81	18.03	11.06	7.19	4.47	3.27
a_1	17.14	11.98	6.80	2.62	1.47	1.58	1.29
a_2	3.03	2.45	2.01	1.59	0.70	0.58	0.36
a_3	0.382	0.167	0.292	0.843	0.444	0.170	0.072
$a_3 > 0$	0.38	0.17	0.29	0.84	0.44	0.17	0.07
$a_2^2 - 3a_1a_2 < 0$	−10.4	−0.01	−1.9	−4.1	−1.5	−0.5	−0.1
(d) Constraints: L-moments, Monotonicity, Correlation, and No Crossover (LM/Mono/Corr/NC)							
	1 h	2 h	6 h	12 h	24 h	48 h	72 h
a_0	51.99	35.81	18.01	11.06	7.19	4.48	3.26
a_1	17.52	11.98	6.80	2.62	1.47	1.61	1.25
a_2	2.80	2.45	2.03	1.59	0.71	0.57	0.37
a_3	0.150	0.167	0.292	0.843	0.444	0.159	0.089
$a_3 > 0$	0.15	0.17	0.29	0.84	0.44	0.16	0.09
$a_2^2 - 3a_1a_2 < 0$	−0.01	−0.01	−1.8	−4.1	−1.5	−0.4	−0.2

Note: * indicates a violation of monotonicity condition.

4. Results and Discussions

By varying the value of z_T for different return periods in Equation (55), in conjunction with the optimal TPNT coefficients listed in Table 6a–d, one can establish IDF curves as shown in Figures 2 and 3. Part (a) of Table 6 and Figures 2–4 (denoted by “LM”) shows the results from considering only the bounding constraints of L-moments, Equations (32)–(35). In fact, the optimal TPNT coefficients

corresponding to each duration can be obtained separately from the exact solutions using sample L-moments in Equations (11)–(14). Note that the TPNT coefficients obtained from each rainfall duration at this stage do not necessarily comply with a one-to-one monotonic increasing relation of rainfall quantile and probability. This can be clearly seen in Table 6a for 1 and 2 h rainfalls for which the two monotonicity constraints are violated (shown by *). Part (b) (denoted by “LM/Mono”) shows the results by considering both L-moment constraints, Equations (32)–(35), and the monotonicity constraints, Equation (36), for each rainfall duration. In this case, both results presented in Parts (a) and (b) in Table 6 and Figures 2–4 can be obtained separately by treating rainfall data of different durations without considering their inter-correlations. Results in Part (c), denoted by “LM/Mono/Corr,” were obtained by incorporating correlation constraints of rainfall data with different durations, Equations (41) and (42), in determining the multivariate TPNT coefficients.

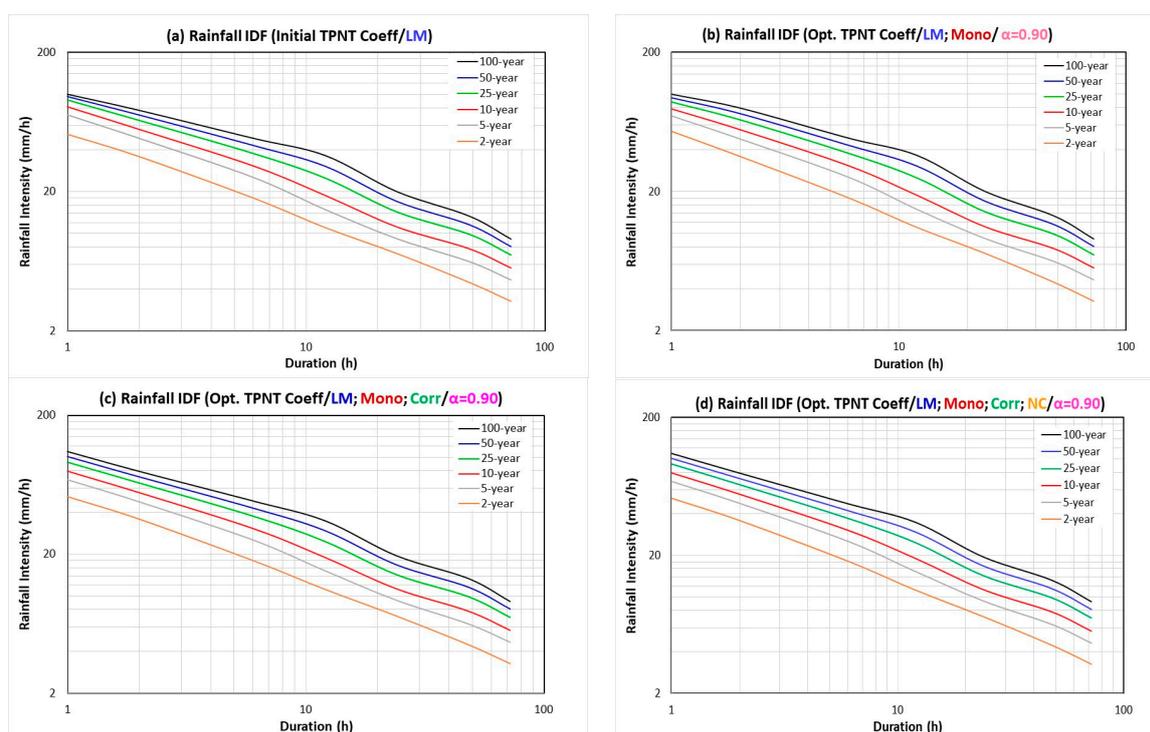


Figure 2. Multivariate TPNT modeling of rainfall intensity–duration relationships of varying return periods under different constraints which consider: (a) L-moments only; (b) L-moments and monotonicity; (c) L-moments, monotonicity and correlation; (d) L-moments, monotonicity, correlation and no crossover.

To show the degree of goodness-of-fit of normal transformed rainfall data by the proposed multivariate TPNT procedure, a normal probability plot of 24 h rainfall data (after normal transformation) with the fitted line and 95% confidence band are shown in Figure 5 as an example. The goodness-of-fit test shown in Figure 5 was achieved by the Anderson–Darling test [56] by which the test statistic is 0.535 with a p -value of 0.155. Figure 5 represents the worst case among the seven durations considered. The range of p -value varies from 0.155 (for 72 h) to 0.933 (for 2 h), which are higher than the generally adopted significance level of 0.05. This indicates that the normal transform by the proposed multivariate TPNT procedure is quite adequate.

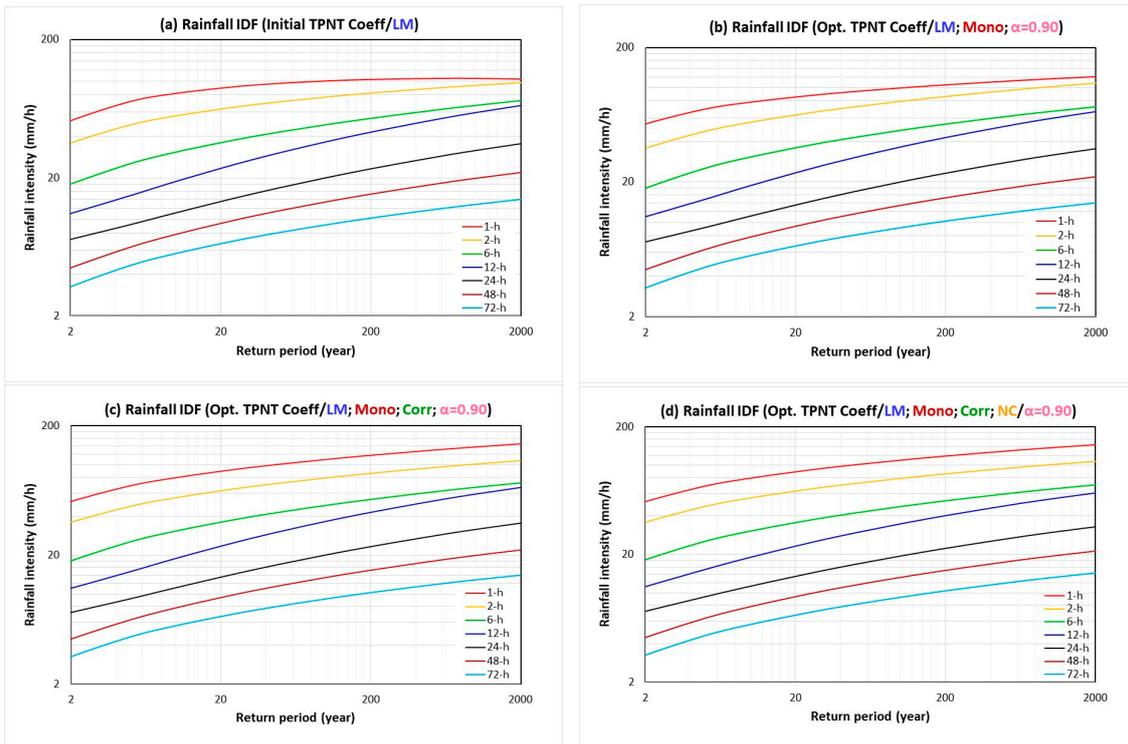


Figure 3. Multivariate TPNT modeling of rainfall intensity–frequency relationships of varying durations under different constraints which consider: (a) L-moments only; (b) L-moments and monotonicity; (c) L-moments, monotonicity and correlation; (d) L-moments, monotonicity, correlation and no crossover.

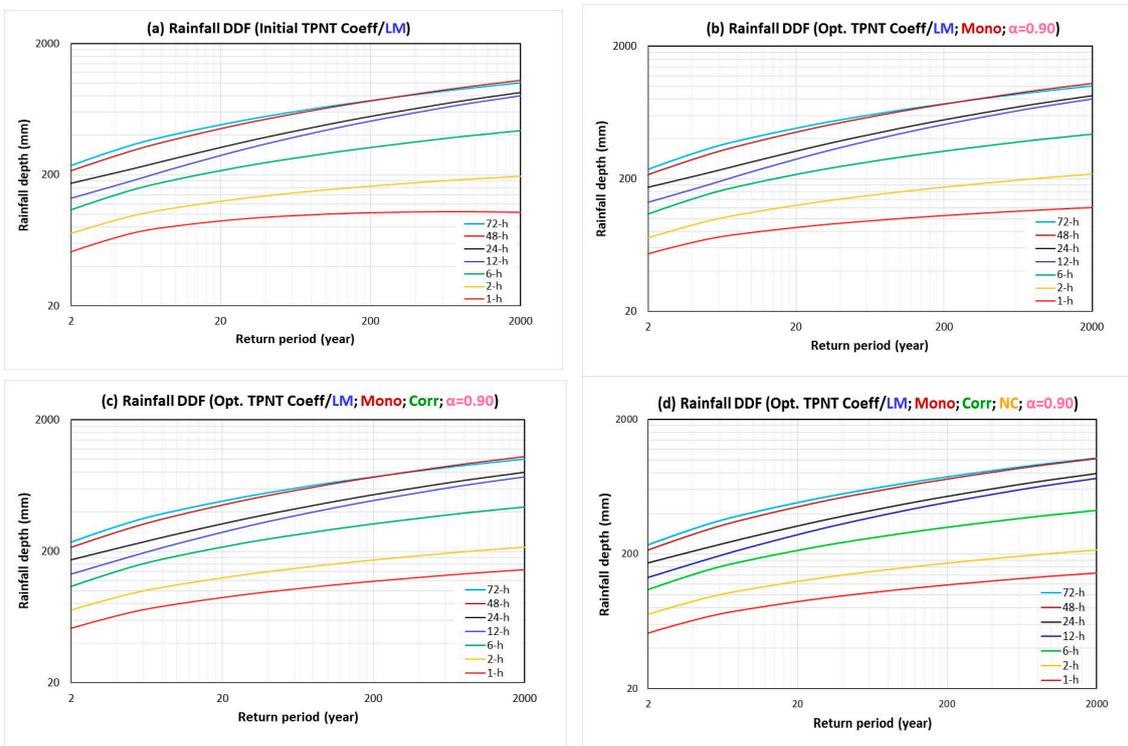


Figure 4. Multivariate TPNT modeling of rainfall depth–frequency relationships of varying durations under different constraints which consider: (a) L-moments only; (b) L-moments and monotonicity; (c) L-moments, monotonicity and correlation; (d) L-moments, monotonicity, correlation and no crossover.

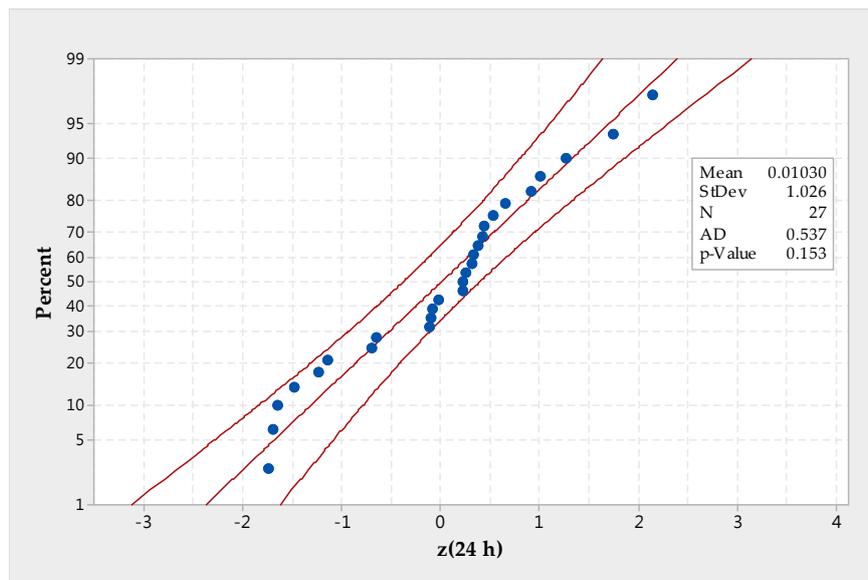


Figure 5. Probability plot of normalized 24 h rainfall data.

Note that the solution obtained up to this stage does not necessarily comply with the physical reality that rainfall intensity (depth) of a given return period is a decreasing (an increasing) function of duration. In other words, rainfall intensity/depth–frequency curves of different durations should not intersect or crossover each other. However, in the process of establishing rainfall IDF/DDF relationships, one does not know in advance if any two resulting two curves would intersect before the statistical model is developed. Therefore, a special set of intersections avoidance constraints are imposed in establishing the IDF curves:

$$\left(a_{0,t_j} - a_{0,t_k}\right) + \left(a_{1,t_j} - a_{1,t_k}\right)z_{T_*} + \left(a_{2,t_j} - a_{2,t_k}\right)z_{T_*}^2 + \left(a_{3,t_j} - a_{3,t_k}\right)z_{T_*}^3 > 0, \text{ for durations } t_j < t_k \quad (56)$$

where T_* = upper limit of selected rainfall return period below which no crossover of IDF curves is permitted to occur; z_{T_*} = standard normal quantile obtainable from $\Phi^{-1}\left(1 - \frac{1}{T_*}\right)$. Hence, additional $N - 1$ no-crossover (NC) constraints are included in the optimization model to solve for multivariate TPNT coefficients. Part (d) results (denoted by “LM/Mono/Corr/NC”) show the rainfall IDF relations considering the NC constraints.

Figure 2 shows the rainfall intensity–duration curves corresponding to various frequencies. For this particular data set, by only preserving sample L-moments, Figure 2a reveals two unusual features for those curves when return period is high (say, ≥ 100 years). They are (1) curves that tend to converge together for rainfall duration in the vicinity of 1 h and (2) the relatively pronounced undulation of curves for medium and long duration. These features are indications of possible anomalies that should not appear in a reasonable rainfall IDF relation. The convergence of rainfall intensity–duration curves in Figure 2a, shown in a different form in intensity–frequency relation as Figure 3a, reveal that the 1 h curve (in red) clearly does not satisfy the monotonicity condition according to Equation (36), which requires a rainfall intensity quantile value to increase continuously with a return period (see also Table 4a). In fact, the 2 h intensity–frequency curve (in gold) also mildly violates the monotonicity requirement as the curve starts to bend down for high return periods. The violation of the monotonicity condition can also be observed in the form of the depth–frequency curve for a 1 h duration (see Figure 4a). In this circumstance, the non-monotonicity of the 1 h rainfall intensity–frequency relation produces a crossover with the 2 h curve shown in Figure 3a.

Interestingly, Figure 3a also reveals that 6 and 12 h rainfall intensity–frequency curves have a strong tendency to intersect as rainfall frequency increases. This tendency to intersect could be attributed to a relatively large undulation of intensity–duration curves in the range of 6–12 h when

rainfall frequency increases (see Figure 2a). From 6 to 12 h, the gradient of intensity–duration curves flatten out for larger return periods. The empirical results show some evidence of improvement (in terms of a decrease in undulation, for large frequencies) when more constraints are considered. However, the improvement is not significantly enough to remove undulation. In practical engineering applications, the undulation of rainfall IDF curves such as those shown in Figure 2a–d is removed by fitting the estimated rainfall intensity–duration data by an empirical IDF model, such as Sherman’s equation [57].

It is clear that, by considering the monotonicity constraint, Equation (36), the crossover tendency of intensity–duration curves (see Figure 2b) in the vicinity of 1–2 h disappears (see also Table 6b), as does that of the 1 h and 2 h intensity–frequency curves in Figure 3b. Correspondingly, the concave down appearance of the 1 h depth–frequency curve and, to a lesser extent, the 2 h curve is corrected (see Figure 4b).

Notice that joint consideration of complying with L-moments and the monotonicity condition does not truly take into account the inter-correlations of rainfall intensity or depth with different durations. The appearance of undulation in the rainfall intensity–duration curves for medium and long durations (≥ 6 h), which satisfy the monotonicity condition, is not affected. Hence, the crossover tendency of 6 and 12 h intensity–frequency curves (see Figure 3b) and the actual intersection of 48 and 72 h depth–frequency curves (see Figure 4b) remain unchanged.

With further consideration of inter-correlations of rainfalls of different durations, Equations (41) and (42), the resulting rainfall IDF and DDF curves are shown in Part (c) of Figures 2–4. Figure 2c shows that the rainfall intensity–duration curves in the range of short duration for a high return period completely remove the crossover tendency. Both Figures 3c and 4c show that rainfall intensity–frequency and depth–frequency curves for 1 and 2 h are parallel to each other. Still, the 48 and 72 h rainfall depth–frequency curves intersect (see Figure 4c).

For illustration, this application artificially select $T_* = 5000$ -year in Equation (56) as the limiting frequency below which rainfall depth–frequency or intensity–frequency curves of any two durations are not allowed to intersect. The obvious results of imposing no-crossover constraint is that the 48 h rainfall depth–frequency curve in Figure 4d would not intersect with the 72 h curve.

As for the effect of confidence level, numerical results indicate that a feasible solution for TPNT coefficients may not exist when the confidence levels for the unknown true L-moments and correlation coefficients are too low. This is expected because the width of confidence interval shrinks toward the sample L-moments and correlation coefficients as the confidence level reduces. At a certain confidence level, the corresponding width of the confidence band might be too restrictive for the optimization model to find feasible TPNT coefficients that simultaneously satisfy the monotonicity constraints. How low the limiting confidence level is depends on the problem. In this numerical example, the limiting confidence level is about 70%, below which no feasible solution can be found for multivariate TPNT coefficients. On the other hand, a reasonable confidence interval allows one to obtain a suitable set of TPNT coefficients to approximate multivariate relations.

5. Summary and Conclusions

Statistical modeling and data analysis in hydrosystems engineering often encounter multiple correlated random variables following non-normal distributions. Due to the difficulty in establishing a full joint probability density function for the involved variables, most of the methods tackling multivariate problems preserve the marginal statistical properties (e.g., distributions or moments) of individual variables and their correlation structures. In this study, focus is placed on the third-order polynomial transform (TPNT) procedure, which relies on the preservation of marginal L-moments and correlations among variables. In particular, a general framework is presented to optimally determine multivariate TPNT coefficients incorporating the constraints that (1) preserve the statistical L-moments and correlations with explicit consideration of their associated sampling errors; (2) comply with a one-to-one monotonicity increasing relation between quantiles of the original and normal

transformed variables. Other than the above basic constraints required to hold the statistical and mathematical validity of the TPNT method, additional constraints that are relevant to the problem at hand can be incorporated into the modeling framework. In the illustrative example of establishing rainfall intensity–duration–frequency (IDF) relations, the no-intersection constraints for rainfall depth–frequency curves of different durations, Equation (56), are introduced in the model formulation to ensure that the resulting IDF relationships comply with the physical reality. The proposed method not only solves for the suitable multivariate TPNT coefficients that satisfy the monotonicity condition for individual variables, but also produces the correlation coefficients between random variables in the normal space. At this stage, the proposed multivariate TPNT procedure has not gone through a formal mathematical testing for its performance under different scenarios of multivariate distributions, correlation structures, and sample sizes. However, the procedure is based on a good logic with sound statistical and mathematical theory. The results from the empirical application to establish at-site rainfall IDF relationships appear to be quite reasonable.

Author Contributions: Conceptualization: Y.-K.T. and L.W.Y.; formal analysis: Y.-K.T., L.W.Y., and C.S.Y.; investigation: Y.-K.T., L.W.Y., and C.S.Y.; writing-original draft preparation: Y.-K.T. and L.W.Y.; writing-review & editing: Y.-K.T., L.W.Y., and C.S.Y.; project administration: Y.-K.T. and C.S.Y.; funding acquisition: Y.-K.T. and C.S.Y.

Funding: This study was primarily supported by the Joint Cooperative Research Program managed by the National Research Foundation of Korea (2016K2A9A1A06922023) and the Ministry of Science & Technology of Taiwan (MOST 105–2923-E-009-004-MY2). Additional funding was received from the General Research Program of the Ministry of Science & Technology of Taiwan (106-2221-E-009 -067).

Acknowledgments: The authors wish to express their gratitude to the two anonymous reviewers for their thorough review and constructive comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kotz, S.; Balakrishnan, N.; Johnson, N.L. Models and Applications. In *Continuous Multivariate Distributions*, 2nd ed.; Wiley and Sons Inc.: New York, NY, USA, 2005; Volume 1.
2. Hutchinson, T.P.; Lai, C.D. *Continuous Bivariate Distributions—Emphasizing Applications*; Rumsby Scientific Publishing: Adelaide, South Australia, 1990.
3. Yue, S. A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrol. Process.* **2001**, *15*, 1033–1045. [[CrossRef](#)]
4. Nadarajah, S.; Shiao, J.T. Analysis of extreme flood events for the Pachang River, Taiwan. *Water Resour. Manag.* **2005**, *19*, 363–374. [[CrossRef](#)]
5. Der Kiureghian, A.; Liu, P.L. Structural reliability under incomplete probability information. *J. Eng. Mech.* **1986**, *112*, 85–104. [[CrossRef](#)]
6. Genest, C.; Chebana, F. Chapter 30: Copula Modeling in Hydrologic Frequency Analysis. In *Handbook of Applied Hydrology*, 2nd ed.; Singh, V.P., Ed.; McGraw-Hill Book Company: New York, NY, USA, 2017.
7. Favre, A.C.; Adlouni, S.E.; Perreault, L.; Thiemonge, N.; Bobee, B. Multivariate hydrological frequency analysis using copulas. *Water Resour. Res.* **2004**, *40*, W01101. [[CrossRef](#)]
8. Ganguli, P.; Reddy, M.J. Probabilistic assessment of flood risks using trivariate copulas. *Theor. Appl. Climatol.* **2013**, *111*, 341–360. [[CrossRef](#)]
9. Chen, L.; Singh, V.P.; Guo, S.L.; Mishra, A.K.; Guo, J. Drought analysis using copulas. *J. Hydrol. Eng.* **2013**, *18*, 797–808. [[CrossRef](#)]
10. Khedun, C.P.; Chowdhary, H.; Mishra, A.K.; Giardino, J.R.; Singh, V.P. Water deficit duration and severity analysis based on runoff derived from Noah land surface model. *J. Hydrol. Eng.* **2013**, *18*, 817–833. [[CrossRef](#)]
11. Sadri, S.; Burn, D.H. Copula-based pooled frequency analysis of droughts in the Canadian Prairies. *J. Hydrol. Eng.* **2014**, *19*, 277–289. [[CrossRef](#)]
12. Tosunoglu, F.; Kisi, O. Joint modelling of annual maximum drought severity and corresponding duration. *J. Hydrol.* **2016**, *543*, 406–422. [[CrossRef](#)]
13. Requena, A.I.; Mediero, L.; Garrote, L. A bivariate return period based on copulas for hydrologic dam design: Accounting for reservoir routing in risk estimation. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 3023–3038. [[CrossRef](#)]

14. Li, C.; Singh, V.P.; Mishra, A.K. A bivariate mixed distribution with a heavy tailed component and its application to single site daily rainfall simulation. *Water Resour. Res.* **2013**, *49*, 767–789. [[CrossRef](#)]
15. Jun, C.; Qin, X.S.; Gan, T.Y.; Tung, Y.K.; De Michele, C. Bivariate frequency analysis of rainfall intensity and duration for urban stormwater infrastructure design. *J. Hydrol.* **2017**, *553*, 374–383. [[CrossRef](#)]
16. Aas, K.; Czado, C.; Frigessi, A.; Bakken, H. Pair-copula constructions of multiple dependence. *Insur. Math. Econ.* **2009**, *44*, 182–198. [[CrossRef](#)]
17. Shafaei, M.; Fard, A.F.; Dinpashoh, Y.; Mirabbasi, R.; Michele, D.C. Modeling flood event characteristics using d-vine structures. *Theor. Appl. Climatol.* **2017**, *130*, 713–724. [[CrossRef](#)]
18. Vernieuwe, H.; Vandenberghe, S.; De Baets, B.; Verhoest, N.E.C. A continuous rainfall model based on vine copulas. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 2685–2699. [[CrossRef](#)]
19. Tosunoglu, F.; Singh, V.P. Multivariate Modeling of Annual Instantaneous Maximum Flows Using Copulas. *J. Hydrol. Eng.* **2018**, *23*, 04018003. [[CrossRef](#)]
20. Qing, X. Generating correlated random vector involving discrete variables. *Commun. Stat. Theory Methods* **2017**, *46*, 1594–1605.
21. Nataf, A. Determination des distributions dont les marges sont donnees. *Comptes Rendus l'Academie Sciences* **1962**, *225*, 42–43.
22. Lebrun, R.; Dutfoy, A. An innovating analysis of the Nataf transformation from the copula viewpoint. *Probab. Eng. Mech.* **2009**, *24*, 312–320. [[CrossRef](#)]
23. Liu, P.L.; Der Kiureghian, A. Multivariate distribution models with prescribed marginals and covariances. *Probab. Eng. Mech.* **1986**, *1*, 105–112. [[CrossRef](#)]
24. Chang, C.H.; Tung, Y.K.; Yang, J.C. Monte Carlo simulation for correlated variables with marginal distributions. *J. Hydraul. Eng.* **1994**, *120*, 313–331. [[CrossRef](#)]
25. Chen, H.F. Initialization for NORTA: Generation of random vectors with specified marginal and correlations. *INFORMS J. Comput.* **2001**, *13*, 312–331. [[CrossRef](#)]
26. Li, H.S.; Lu, Z.Z.; Yuan, X.K. Nataf transformation based point estimate method. *Chin. Sci. Bull.* **2008**, *53*, 2586–2592. [[CrossRef](#)]
27. Niaki, S.T.A.; Abbasi, B. Generating correlation matrices for normal random vectors in NORTA algorithm using artificial neural networks. *J. Uncertain Syst.* **2008**, *2*, 192–201.
28. Fleishman, A.L. A method for simulating non-normal distributions. *Psychometrika* **1978**, *43*, 521–532. [[CrossRef](#)]
29. Vale, C.D.; Maurelli, V.A. Simulating multivariate non-normal distributions. *Psychometrika* **1983**, *48*, 465–471. [[CrossRef](#)]
30. Headick, T.C.; Sanwilowsky, S.S. Simulating correlated multivariate nonnormal distributions: Extending the Fleishman power method. *Psychometrika* **1999**, *64*, 25–35. [[CrossRef](#)]
31. Chen, X.Y.; Tung, Y.K. Applications of TPNT in multivariate Monte Carlo simulation. In *Water Resources Planning and Management*; EWRI/ASCE: Philadelphia, PA, USA, 2003; pp. 23–26.
32. Hodis, F.A. Simulating Univariate and Multivariate Nonnormal Distributions Based on a System of Power Method Distributions. Ph.D. Thesis, Southern Illinois University, Carbondale, IL, USA, 2008.
33. Demirtas, H.; Hedeker, D.; Mermelstein, R.J. Simulation of massive public health data by power polynomials. *Stat Med.* **2012**, *31*, 3337–3346. [[CrossRef](#)] [[PubMed](#)]
34. Yang, H.; Zou, B. The point estimate method using third order polynomial normal transformation technique to solve probabilistic power flow with correlated wind source and load. In Proceedings of the Asia-Pacific Power and Energy Engineering Conference, Shanghai, China, 27–29 March 2012; pp. 1–4.
35. Cai, D.; Shi, D.; Chen, J. Probabilistic load flow computation with polynomial normal transformation and Latin hypercube sampling. *IET Gen. Trans. Distrib.* **2013**, *7*, 474–482. [[CrossRef](#)]
36. Chen, X.Y. Investigating Third-Order Polynomial Normal Transform and Its Applications to Uncertainty and Reliability Analyses. Master's Thesis, Hong Kong University of Science and Technology, Hong Kong, China, 2002.
37. Zhao, Y.G.; Lu, Z.H. Fourth moment standardization for structural reliability assessment. *J. Struct. Eng.* **2007**, *133*, 916–924. [[CrossRef](#)]
38. Tung, Y.K. Polynomial normal transformation in uncertainty analysis. *Appl. Probab. Stat.* **1999**, *1*, 167–174.
39. Hosking, J.R.M. *The Theory of Probability Weighted Moments*; IBM Research Report, RC12210; IBM: Yorktown Heights, NY, USA, 1986.

40. Fisher, R.A.; Cornish, E.A. The percentile points of distribution having known cumulants. *Technometrics* **1960**, *2*, 209–225. [[CrossRef](#)]
41. Chen, X.Y.; Tung, Y.K. Investigation of polynomial normal transform. *Struct. Saf.* **2003**, *25*, 423–445. [[CrossRef](#)]
42. Hosking, J.R.M. L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *J. R. Stat. Soc. Ser. B Method* **1990**, *52*, 105–124. [[CrossRef](#)]
43. Makkonen, L.; Pajari, M.; Tikanmaki, M. Discussion on “Plotting positions for fitting distributions and extreme value analysis”. *Can. J. Civil Eng.* **2013**, *40*, 130–139. [[CrossRef](#)]
44. Weibull, W. A statistical theory of the strength of materials. *R. Swed. Inst. Eng. Res. Proc.* **1939**, *151*, 1–45.
45. Fisher, R.A. On the ‘probable error’ of a coefficient of correlation deduced from a small sample. *Metron*. **1921**, *1*, 1–32.
46. Kennedy, J.B.; Neville, A.M. *Basic Statistical Methods for Engineers and Scientists*, 3rd ed.; Happer and Row Publishing: New York, NY, USA, 1986.
47. Wilson, R.B. A simplicial Method for Convex Programming. Ph.D. Thesis, Harvard University, Boston, MA, USA, 1963.
48. Boggs, P.T.; Tolle, J.W. Sequential quadratic programming. *Acta Numer.* **1995**, *4*, 1–51. [[CrossRef](#)]
49. Akan, A.O.; Houghtalen, R.J. *Urban Hydrology, Hydraulics, and Stormwater Quality*; Wiley: Hoboken, NJ, USA, 2003.
50. Sun, S.A.; Djordjevic, S.; Khu, S.T. Decision making in flood risk based storm sewer network design. *Water Sci Technol.* **2010**, *64*, 247–254. [[CrossRef](#)]
51. Singh, V.P.; Strupczewski, W.G. On the status of flood frequency analysis. *Hydrol Process.* **2002**, *16*, 3737–3740. [[CrossRef](#)]
52. Porras, P.J.S.; Porras, P.J., Jr. New perspective on rainfall frequency curves. *J. Hydrol. Eng.* **2001**, *6*, 82–85. [[CrossRef](#)]
53. Haktanir, T. Divergence criteria in extreme rainfall series frequency analyses. *Hydrol. Sci. J.* **2003**, *48*, 917–937. [[CrossRef](#)]
54. Gräler, B.; Fischer, S.; Schumann, A. Joint modeling of annual maximum precipitation across different duration levels. In *EGU General Assembly Conference Abstracts*; Discussion Paper SFB 823; EGU: Munich, Germany, 2016.
55. You, L.; Tung, Y.K. Derivation of rainfall IDF relations by third-order polynomial normal transform. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 2309–2324. [[CrossRef](#)]
56. D’Agostino, R.B.; Stephens, M.A. *Goodness-of-Fit Techniques*; Marcel Dekker: New York, NY, USA, 1986.
57. Sherman, C.W. Frequency and intensity of excessive rainfalls at Boston, Massachusetts. *Transaction* **1931**, *95*, 951–960.

