

Modelling Dependence between Traffic Noise and Traffic Flow through An Entropy-Copula Method

K. Huang¹, L. M. Dai^{1,*}, M. Yao², Y. R. Fan³, X. M. Kong⁴

¹Industrial Systems Engineering, Faculty of Engineering and Applied Science, University of Regina, Regina, SK S4S 0A2, Canada

²Mechanical Engineering, Faculty of Engineering, University of Waterloo, Ontario N2L 3G1, Canada

³Environmental Systems Engineering, Faculty of Engineering and Applied Science, University of Regina, Regina, SK S4S 0A2, Canada

⁴MOE Key Laboratory of Regional Energy and Environmental Systems Optimization, North China Electric Power University, Beijing 102206, China

Received 26 September 2014; revised 26 March 2015; accepted 28 March 2015; published online 5 April 2015

ABSTRACT. In this study, an entropy-copula method is proposed for modelling dependence between traffic volume and traffic noise on the Trans-Canada Highway (#1 highway of Canada) in the City of Regina based on a series of field experiment measurements. The proposed entropy-copula method combines the maximum entropy and copula methods into a general framework. The marginal distributions of traffic volume and traffic noise are estimated through the principle of maximum entropy (POME) theory, and the joint probabilities are derived through the Gaussian and Student t copulas. The underlying assumptions of the coupled entropy-copula method are that: i) the entropy variables are mutually independent from each other, and ii) the marginal distributions of traffic flow and traffic noise are continuous. The proposed method is applied to two field experiment sites on the Trans-Canada Highway. Based on the K-S and A-D tests and RMSE value, the entropy method shows well performance in quantifying the probability distributions of traffic volume and traffic noise. Meanwhile, both the Gaussian and Student t copulas can well model the joint probability distributions of the traffic volume and traffic noise at the both experiment sites, which is demonstrated by the Cramér von Mises statistics and the RMSE value. Furthermore, the conditional CDFs of the traffic noise at the two experiment sites are derived based on the established copulas with respect to different traffic volume scenarios. These conditional CDFs indicate positive structures between traffic volume and traffic noise at the both experiment sites. The conditional PDFs of the traffic noise under different traffic flow scenarios are also generated, indicating the potential reduction effect of traffic noise due to the decrease of the traffic volume. This proposed approach can quantify the dependence between traffic flow and traffic noise, and reveal the inherent uncertain relationship between these two variables. Moreover, the obtained results can provide useful information for traffic noise reduction through traffic flow management.

Keywords: acoustic measurement, traffic noise, entropy, copula, multivariate analysis, uncertainty

1. Introduction

Urban noise pollution is causing increased health risks in human beings, such as hearing impairment, cardiovascular diseases, and child physical development. Among various noise emission sources in urban area, the noise levels associated with transport, have increased dramatically since the mid-twentieth century (UN, 2004). Road traffic noise is caused by the combination of rolling noise, consisting of the friction noise between road surface and vehicle tires, and the propulsion noise caused by exhaust systems or engines. The levels of traffic noise are influenced by many factors such as types of engines, exhaust systems, tires interacting with the road, weather and road conditions. Generally, these factors can be categorized into

four clusters: traffic, road, environmental and weather factors. Heavier traffic volumes, higher speed, and greater number of trucks are expected to increase the loudness of traffic noise (Abo-Qudais and Alhiary, 2005). Traffic volume is one of the most significant impact factors for traffic noise pressure, especially on a highway that the speed of vehicles is almost constant. Consequently, quantifying the dependence between traffic volume and traffic noise is required to explore the inherent interrelationship between these two variables. Such quantification can be helpful for decision makers to mitigate the traffic noise levels from a view of traffic volume management.

Temporal and spatial variations in vehicle volumes have a substantial impact on traffic noise emissions, especially on highways that the vehicle speed would not change dramatically. The traffic flow is one of the two most important variables considered in the development of traffic noise prediction model (Alves Filho, 1997). Increased traffic flow, in terms of vehicles per hour, causes an increase in the noise level, given constant traffic composition, speed and driving patterns (da Paz et al., 2005). Previously, amounts of research have been proposed to

* Corresponding author. Tel.: +1 306 5854498; fax: +1 306 5854855
E-mail address: liming.dai@uregina.ca (L. M. Dai).

investigate the impacts of vehicle volume on traffic noise emissions. For example, Mehdi et al. (2011) analyzed spatio-temporal patterns of road traffic noise pollution in Karachi, Pakistan. Agarwal and Swami (2011) proposed a comprehensive approach for the development of traffic noise prediction model for Jaipur city. Dai et al. (2013) developed a traffic noise prediction model on inland waterway of China using the FHWA. Engel et al. (2014) proposed statistical analysis of a combination of objective and subjective environmental noise data using factor analysis and multinomial logistic regression. However, most of previous research works mainly focused on traffic noise predictions, in which the traffic volume was considered as one of the main inputs. Also some statistical analyses between traffic volume and traffic noise were conducted through simple correlation coefficient methods (e.g. Pearson's r). For instance, through the analysis of variance (ANOVA) and the Tukey test, To et al. (2002) revealed that the total traffic flow and the number of heavy vehicles are the most significant factors for traffic noise. The study proposed by da Paz and Zannin (2010) showed that the traffic noise is highest correlated with the total vehicle flow, followed by the percentage of heavy vehicle and light vehicle flows. However, most previous studies just qualitatively indicate the impact of traffic flow on traffic noise. Few studies were reported to quantitatively investigate the inherent interrelationships between traffic volume and traffic noise, especially for quantifying how the traffic noise would distribute under various traffic volume conditions.

Copula and entropy methods have been widely used in many engineering applications such as streamflow simulation, drought frequency analysis, flooding forecasting, bearing strength prediction, and so on. For instance, Reddy and Ganguli (2011) applied copula method to derive the drought severity-duration-frequency curves. Sraj et al. (2014) used copula function for bivariate flood frequency risk analysis. Kong et al. (2015) proposed a maximum entropy-Gumbel-Hougaard copula method for simulation of monthly streamflow in Xiangxi River, China. However, no study was reported to use copula and entropy method for traffic noise impact analysis.

Consequently, this study aims to model dependence between traffic volume and traffic noise through an entropy-copula method. As an extension of pervious research, this study will innovatively introduce the copula method into traffic noise impacts analysis. From a view of statistics, this research work will explore the inherent interrelationship between traffic noise and traffic flow, and reveal the traffic noise mitigation effect through traffic flow reduction. The proposed entropy-copula method combines the maximum entropy and copula methods into a general framework, in which the marginal distributions of traffic volume and traffic noise are estimated through the principle of maximum entropy (POME) theory, and the joint probabilities (including joint cumulative distribution function (CDF) and joint probability density function (PDF)) are derived through the Gaussian and Student t copulas. The reasons we choose the coupled entropy-copula approach are that (i) the entropy method can approximate any probability function; (ii) the copula method can allow the marginal distribution and the joint distribution to be calculated in separate processes, and different

marginal distributions to be used together to generate the joint distribution. The proposed method would then be applied to model the dependence of traffic volume and traffic noise on the Trans-Canada Highway (#1 highway of Canada) in the City of Regina, based on a series of field experiment measurements.

2. Methodology

2.1. Entropy Theory

The concept of entropy was firstly introduced by Shannon (1948) to quantitatively measure the mean uncertainty associated with a probability distribution of a random variable (Chen et al., 2013). Jaynes (1957a, b, 1982) then developed the principle of maximum entropy (POME) for deriving a probability distribution with given constraints reflecting certain information. Shannon entropy is one of the widely used entropies for many engineering applications such as drought analysis, streamflow forecasting, and bearing strength prediction. (Singh, 1997; Huang 1998; Huang et al., 2006; Singh, 2011; Li, 2013).

The definitions of Shannon entropy can be based on both discrete and continuous random variables. For a discrete random variable $X = \{x_1, x_2, \dots, x_n\}$, the entropy for X can be defined as:

$$H(X) = -\sum_{i=1}^n p(x_i) \ln(p(x_i)) \quad (1)$$

where $p(x_i)$ is the probability for $X = x_i$.

If the random variable X is continuous and holds a probability density function (PDF) $f(x)$, the Shannon entropy $H(X)$ is defined as (Shannon, 1948):

$$H(X) = -\int_a^b f(x) \ln f(x) dx \quad (2)$$

where $H(X)$ is the entropy of X (also referred to as entropy function); x is a value of random variable X . a and b are the lower and upper limits of X , respectively.

The principle of maximum entropy (POME), formulated by Jaynes (1957a, b), is established to find the most suitable probability distribution subject to the given information. In the mathematical frame of POME, the given information is expressed as a set of constraints formed as expectations of function $h_i(x)$ (Papalexiou and Koutsoyiannis, 2012):

$$\int_a^b h_i(x) f(x) dx = E(h_i(x)) \quad (3)$$

The application of the principle of maximum entropy is based on the assumption that the entropy variables are mutually independent from each other (Jaynes, 1957a). In the estimation of PDFs for traffic flow and noise, it is necessary to preserve the statistical characteristics of the observed samples. Consequently, mean, standard deviation, skewness and kurtosis can be regarded as constraints for deriving the distributions of traffic flow and noise. In detail, Equation (3) can be expressed

ssed as:

$$C_0 = \int_a^b f(x)dx = 1 \quad (4)$$

$$C_i = \int_a^b h_i(x)f(x)dx = \overline{h_i(x)} \quad i = 1, 2, \dots, m \quad (5)$$

where x is traffic flow or noise value; $h_i(x)$ is a known function of random variable X , which can be specified as $h_1 = x$, $h_2 = x^2$, $h_3 = x^3$, and $h_4 = x^4$, for the proposed constraints; $\overline{h_i(x)}$ is the expectation of $h_i(x)$, which can be specified as $\{x, x^2, x^3, x^4\}$. Equation (4) states that the PDF must satisfy the cumulative probability theorem (Huang and Loucks, 2000; Li et al., 2010).

The maximization of entropy, based on specified constraints expressed as Equations (4) and (5), can be accomplished by using the method of Lagrange multipliers. The Lagrangian function L , subject to Equations (4) and (5), can be expressed as (Kapur and Kesavan, 1992; Li and Huang 2009):

$$L = -\int_a^b f(x) \ln f(x)dx - (\lambda_0 - 1) \left[\int_a^b f(x)dx - 1 \right] - \sum_{i=1}^m \lambda_i \left[\int_a^b f(x)h_i(x)dx - C_i \right] \quad (6)$$

where $\lambda_i (i = 1, 2, \dots, m)$ are the Lagrange multipliers. $f(x)$ is the PDF of random variable X ; $h_i(x)$ are known functions of X ; C_i are constraints of $f(x)$. Then $f(x)$ can be generated through maximizing L , leading to the derivative of L respect to $f(x)$ being zero (Li et al., 2008a,b; 2009; Lv et al., 2010):

$$\frac{\partial L}{\partial f} = -[1 + \ln f(x)] - (\lambda_0 - 1) - \sum_{i=1}^m \lambda_i h_i(x) = 0 \quad (7)$$

Hence, the resulting maximum-entropy-based (ME-based) PDF can be expressed as (Kapur and Kesavan, 1992):

$$f(x) = \exp[-\lambda_0 - \sum_{i=1}^m \lambda_i h_i(x)] \quad (8)$$

Substituting Equation (8) into Equation (4), the 0th Lagrange multiplier λ_0 can be obtained as:

$$\lambda_0 = \ln \left[\int_0^\infty \exp \left(- \sum_{i=1}^m \lambda_i h_i(x) \right) dx \right] \quad (9)$$

Consequently, the probability density function for X can be obtained as follows (Qin et al., 2007):

$$f(x) = \exp \left[- \ln \left(\int_0^\infty \exp \left(- \sum_{i=1}^m \lambda_i h_i(x) \right) dx \right) - \sum_{i=1}^m \lambda_i h_i(x) \right] \quad (10)$$

The PDF defined by Equation (10) can preserve the most im-

portant statistical moments. The cumulative distribution function (CDF) of the ME-based PDF can be expressed as:

$$E_x(x) = \int_a^x f(t)dt \quad (11)$$

In general, the analytical solutions for obtaining the Lagrange multipliers do not exist, and then the numerical solution method is required (Hao and Singh, 2011). As stated by Hao and Singh (2011), the Lagrange multipliers can be found through minimizing a convex function expressed as:

$$Z(\lambda) = \ln \left(\int_0^\infty \exp \left(- \sum_{i=1}^m \lambda_i h_i(x) \right) dx \right) + \sum_{i=1}^m \lambda_i h_i(x) \quad (12)$$

In this study, the Lagrange multipliers $\lambda_i (i = 1, 2, \dots, m)$ in Equation (12) can be determined using the Conjugate Gradient (CG) method. The CG method has played a special role in solving large scale nonlinear optimization problems with advantages of super linear convergence, simple recurrence formula and less calculation. The detailed process of CG method is presented as follows (Kong et al., 2015):

- (1) Starting from some initial value $\lambda_{(0)}$;
- (2) Let the initial value of search direction $d_{(0)}$:

$$d_{(0)} = -g_{(0)} = - \frac{\partial Z}{\partial \lambda_{(0)i}} \quad (i = 1, 2, \dots, m)$$

- (3) For step $k = 0, 1, \dots, n - 1$, calculate:

$$f(\lambda_{(k)} + \alpha_{(k)} d_{(k)}) = \min f(\lambda_{(k)} + \alpha d_{(k)})$$

$$\lambda_{(k+1)} = \lambda_{(k)} + \alpha_{(k)} d_{(k)}$$

$$g_{(k)} = \nabla Z(\lambda_{(k)})$$

$$\beta_{(k)} = \frac{g_{(k+1)}^T (g_{(k+1)} - g_{(k)})}{\|g_{(k)}\|_2^2}$$

$$d_{(k+1)} = -g_{(k+1)} + \beta_{(k)} d_{(k)}$$

The advantage of the entropy-based method is that no assumption is required about the marginal distribution of the historical data, and can be applied to non-normal data without transforming these data to normal (Hao and Singh, 2011). Consequently, this approach can be suitable for estimating the probability distributions without known forms. Furthermore, the probability distributions derived by POME is the least biased estimate possible on the given information and it is maximally non-committal with regard to missing information. (Papalexioiu and Koutsoyiannis, 2012). However, the POME may be computationally cumbersome when more statistics are to be preserved

and determination of more Lagrange multipliers is involved (Hao and Singh, 2011).

2.2. Copula Method

The word ‘copula’ was first employed in a mathematical or a statistical sense by Sklar (1959) in the theorem describing the functions that connect one-dimensional distributions functions to form multivariate distribution functions (Nelsen, 2006). After that, amounts of mathematical theories were developed on copulas for multivariate dependence analysis. In this study, some definitions and concepts on copulas to be used will be presented in this section. The interested reader is referred to Joe (1997) and Nelsen (2006) for more detailed description of copula function properties.

Copulas are joint cumulative distribution functions of n univariate standard uniform random variables. For example, a bivariate copula can be represented as:

$$C: [0, 1]^2 \rightarrow [0, 1] \tag{13}$$

It has to fulfil “boundary” and “increasing” conditions as follows:

- (1) $C(u, 1) = u, C(1, v) = v$ and $C(u, 0) = C(0, v) = 0,$
- (2) $C(u_1, u_2) + C(v_1, v_2) - C(u_1, v_2) - C(v_1, u_2) \geq 0$ if $u_1 \geq v_1, u_2 \geq v_2$ and $u_1, u_2, v_1, v_2 \in [0, 1].$

For n -dimensional distribution function F , it can be written as:

$$F(x_1, x_2, \dots, x_n) = C(F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)) \tag{14}$$

where $F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)$ are marginal distributions of random vector (X_1, X_2, \dots, X_n) . If these marginal distributions are continuous, a single copula function C exists, which can be written as (Sraj et al., 2014):

$$C(u_1, u_2, \dots, u_n) = F(F_{X_1}^{-1}(u_1), F_{X_2}^{-1}(u_2), \dots, F_{X_n}^{-1}(u_n)) \tag{15}$$

where $0 \leq u_1, u_2, \dots, u_n \leq 1.$

The Archimedean, elliptical and extreme value copulas are some widely applied classes of copula functions. In the present study, the Gaussian and Student t copulas, as two widely used elliptical copulas would be considered for the analysis due to its capability of modelling the asymmetric or unbalanced relationship and dependence structures. The analytical expressions of bivariate Gaussian and Student t copulas are elaborately derived based on some pioneering works (Kelly and Krzysztofowicz, 1997; Fang et al., 2002; Demarta and McNeil, 2005; ŹeŹula, 2009; Ma et al., 2013).

The Gaussian copula can be expressed as:

$$C(u_1, u_2; \Sigma) = \Phi_{\Sigma}(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) =$$

$$\int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{(2\pi)^2 |\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2} \mathbf{w}^T \Sigma^{-1} \mathbf{w}) d\mathbf{w} \tag{16}$$

where $\Phi_{\Sigma}(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$ denotes the bivariate standard normal distribution; $\Phi^{-1}(\cdot)$ means the inverse of the standard normal distribution of $\Phi(\cdot)$; Σ is the covariance matrix; $\mathbf{w} = [w_1, w_2]^T$ represents the corresponding integral variables. The density function of the Gaussian copula can be formulated as:

$$c(u_1, u_2; \Sigma) = \frac{\partial^2 C(u_1, u_2; \Sigma)}{\partial u_1 \partial u_2} = \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2} (\zeta^T \Sigma^{-1} \zeta - \zeta^T \zeta)) \tag{17}$$

where $\zeta = [\Phi^{-1}(u_1), \Phi^{-1}(u_2)]^T.$

The formulation of Student t copula can be expressed as:

$$C(u_1, u_2; \Sigma, \nu) = T_{\Sigma, \nu}(T_{\nu}^{-1}(u_1), T_{\nu}^{-1}(u_2)) \\ = \int_{-\infty}^{T_{\nu}^{-1}(u_1)} \int_{-\infty}^{T_{\nu}^{-1}(u_2)} \frac{\Gamma(\frac{\nu+3}{2})}{\Gamma(\frac{\nu}{2})(\pi\nu)^{\frac{3}{2}} |\Sigma|^{\frac{1}{2}}} (1 + \frac{\mathbf{w}^T \Sigma^{-1} \mathbf{w}}{\nu})^{\frac{\nu+3}{2}} d\mathbf{w} \tag{18}$$

with the density expressed as:

$$c(u_1, u_2; \Sigma, \nu) = \frac{\partial^2 C(u_1, u_2; \Sigma, \nu)}{\partial u_1 \partial u_2} = |\Sigma|^{\frac{1}{2}} \frac{\Gamma(\frac{\nu+3}{2})}{\Gamma(\frac{\nu}{2})} \frac{\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+1}{2})} \left[\frac{\nu+3}{2} \right]^2 \times \\ \frac{(1 + \frac{\zeta^T \Sigma^{-1} \zeta}{\nu})^{\frac{\nu+3}{2}}}{\nu} \prod_{i=1}^2 (1 + \frac{[T_{\nu}^{-1}(b_i)]^2}{\nu})^{-\frac{\nu+1}{2}} \tag{19}$$

where $T_{\Sigma, \nu}(T_{\nu}^{-1}(u_1), T_{\nu}^{-1}(u_2))$ denotes the bivariate Student t distribution, and $T_{\nu}^{-1}(\cdot)$ denotes the inverse of the Student t distribution, with ν being the degree of freedom of Student t distribution; $\Gamma(\cdot)$ denotes the gamma distribution; $b_i = T_{\nu}^{-1}(u_i), i = 1, 2$ and $\zeta = [T_{\nu}^{-1}(u_1), T_{\nu}^{-1}(u_2)]^T.$

The parameters of Gaussian and Student t copula can be estimated through the maximum pseudo-likelihood estimation method (Nadarajah and Kotz, 2005; Nadarajah, 2006; Shiau, 2006; Shiau et al., 2007; Reddy and Ganguli, 2011; Ma et al., 2013). For bivariate Gaussian copula, its maximum pseudo-likelihood function can be expressed as:

$$L(x_1, x_2; \Sigma) = \prod_{i=1}^n c(u_{1i}, u_{2i}; \Sigma_i) \tag{20}$$

where $c(\cdot)$ is the density function of Gaussian copula; i means the i th observation $(x_{1i}, x_{2i}); u_{1i} = F_{X_1}(x_{1i}), u_{2i} = F_{X_2}(x_{2i})$ denote the marginal probability distributions of considered variables; and Σ_i is the covariance matrix of (x_{1i}, x_{2i}) . Then the unknown

parameter in Gaussian copula Σ can be estimated through $\partial \ln L / \partial \Sigma = 0$. Thus we have:

$$\sum_{i=1}^n \frac{\partial}{\partial \Sigma} \{\ln [c(u_{1i}, u_{2i}; \Sigma_i)]\} = 0 \quad (21)$$

For the bivariate Student t copula, the maximum pseudo-likelihood function can be expressed as:

$$L(x_1, x_2; \Sigma, \nu) = \prod_{i=1}^n c(u_{1i}, u_{2i}; \Sigma_i, \nu) \quad (22)$$

Consequently, the unknown parameters Σ and ν can be obtained through considering $\partial \ln L / \partial \Sigma = 0$ and $\partial \ln L / \partial \nu = 0$, and thus we can have:

$$\begin{cases} \sum_{i=1}^n \frac{\partial}{\partial \Sigma} \{\ln [c(u_{1i}, u_{2i}; \Sigma_i, \nu)]\} = 0 \\ \sum_{i=1}^n \frac{\partial}{\partial \nu} \{\ln [c(u_{1i}, u_{2i}; \Sigma_i, \nu)]\} = 0 \end{cases} \quad (23)$$

If an appropriate copula function is established, the conditional joint distribution can then be obtained. Following Zhang and Singh (2006), the conditional distribution function of U_1 given $U_2 = u_2$ can be expressed as:

$$C_{U_1|U_2=u_2}(u_1) = C(U_1 \leq u_1 | U_2 = u_2) = \frac{\partial}{\partial u_2} C(u_1, u_2) | U_2 = u_2 \quad (24)$$

Similar conditional cumulative distribution for U_2 given $U_1 = u_1$ can be obtained. Moreover, the conditional cumulative distribution function of U_1 given $U_2 \leq u_2$ can be expressed as:

$$C_{U_1|U_2 \leq u_2}(u_1) = C(U_1 \leq u_1 | U_2 \leq u_2) = \frac{C(u_1, u_2)}{u_2} \quad (25)$$

Likewise, an equivalent formula for the conditional distribution function for U_2 given $U_1 \leq u_1$ can be obtained.

Moreover, the probability density function (pdf) of a copula function can be expressed as:

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} \quad (26)$$

and the joint pdf of the two random variables can be obtained as:

$$f(x_1, x_2) = \frac{\partial^2 C(u_1, u_2)}{\partial x_1 \partial x_2} = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} \frac{\partial u_1}{\partial x_1} \frac{\partial u_2}{\partial x_2} = f_{X_1}(x_1) f_{X_2}(x_2) \times c(u_1, u_2) \quad (27)$$

Consequently, the conditional pdf of X_1 , given the value of X_2 , can be formulated as:

$$f(x_1 | x_2) = \frac{f(x_1, x_2)}{f_{X_2}(x_2)} = f_{X_1}(x_1) c(u_1, u_2) \quad (28)$$

and the conditional pdf of X_2 , given the value of X_1 , can be expressed as:

$$f(x_2 | x_1) = \frac{f(x_1, x_2)}{f_{X_1}(x_1)} = f_{X_2}(x_2) c(u_1, u_2) \quad (29)$$

The application of copulas has grown rapidly in engineering and science fields, due to its superior capability of describing the dependence among multiple variables. The main advantage of copula approach is the independence between construction of joint distribution and choice of marginal distributions of individual variables. It can separate the estimation process of marginal and joint probability distributions. Consequently, multiple marginal distributions can be integrated into the joint probability through copula.

2.3. Goodness-of-fit Statistical Tests

After parameter estimation for both the marginal and joint distributions, the goodness-of-fit statistic tests would be performed to determine whether those estimated distributions are satisfied. The root mean square error (RMSE), the Kolmogorov-Smirnov (K-S) and the Anderson-Darling (A-D) goodness-of-fit tests would be employed to evaluate the performance of the marginal distributions obtained through the entropy method. And the Rosenblatt transformation (Rosenblatt, 1952) would be applied to investigate the performance of joint distributions in describing the dependency between traffic flow and traffic noise.

2.3.1. Goodness-of-fit Statistics for Marginal Distribution

In the process of evaluating the performance of marginal distribution obtained through the entropy method, the empirical nonexceedance probabilities would be obtained through the Gringorten plotting position formula (Gringorten, 1963), which is expressed as:

$$P(K \leq k) = \frac{k - 0.44}{N + 0.12} \quad (30)$$

where N stands for the sample size; k stands for the k^{th} smallest observation in the data set; and the data set was arranged in an increasing order.

The RMSE, the K-S test and the A-D test are used to evaluate the probability distributions of traffic flow and noise estimated through the entropy method. The RMSE can be expressed as (Willmott and Matsuura, 2005):

$$RMSE = \sqrt{\frac{\sum_{k=1}^K (x_k^{est} - x_k^{obs})^2}{K}} \quad (31)$$

where x_k^{est} denotes theoretical value from the fitted probability distribution; x_k^{obs} denotes the empirical probability obtained through Equation (30); K is the sample size.

The K-S test is a nonparametric probability distribution free test (Zhang and Singh, 2012). The statistic of K-S test quantifies the largest vertical difference between the estimated and empirical distributions (Massey, 1951; Razali and Wah, 2011). Given n increasing ordered data points, $x_{(.)}$, the K-S test statistic is defined as (Conover, 1999):

$$T = \sup_x |F^*(x) - F_n(x)| \quad (32)$$

where $F^*(x)$ means the estimated distribution, $F_n(x)$ denotes the empirical distribution, and ‘sup’ stands for supremum. The null hypothesis H_0 is: $F(x) = F^*(x)$ for all x from $-\infty$ to ∞ (The sample data follow the hypothesized distribution). If T exceeds the $1 - \alpha$ quantile, then we reject H_0 at the level of significance, α (Razali and Wah, 2011). The P-value for K-S test was approximated using Miller’s approximation (Zhang and Singh, 2012).

The A-D test is one of the most powerful empirical distribution function (EDF) tests (Arshad et al., 2003). It can examine whether the sample data is drawn from a specific distribution (Scholz, 1987; Farrel and Stewart, 2006). Given n increasing ordered data points, $x_{(.)}$, the A-D test statistic can be expressed as follows (Anderson and Darling, 1954; Arshad et al., 2003):

$$W_n^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) \{ \log F^*(x_i) + \log(1 - F^*(x_{n+1-i})) \} \quad (33)$$

where $F^*(x_i)$ stands for the cumulative distribution function of the hypothesized distribution. The null hypothesis H_0 is: The sample data follow the hypothesized distribution. If W_n^2 exceeds the $1 - \alpha$ quantile, then we reject H_0 at the level of significance, α (Razali and Wah, 2011). The P-value for A-D test was approximated using Monte Carlo simulation (Zhang and Singh, 2012).

2.3.2. Goodness-of-fit Statistics for Copula

The goodness-of-fit test based on Rosenblatt transformation will be employed in this study to evaluate the performance of copula based on the recommendation by Genest et al. (2009).

Definition (Genest et al., 2009): Rosenblatt’s probability integral transform of a bivariate copula C is the mapping $\mathfrak{R} : (0, 1) \rightarrow (0, 1)$ which to $(u, v) \in (0, 1)^2$ assigns another vector $\mathfrak{R}(u, v) = (e_1, e_2) \in (0, 1)$ with $e_1 = u$ and $e_2 = C(v|u)$.

A critical property of Rosenblatt’s probability integral transform is that the joint distribution of $\mathbf{v} = (u, v)$ is distributed as C , denoted as $\mathbf{v} \sim C$, if and only if the joint distribution of

$\mathfrak{R}(\mathbf{v})$ is the bivariate independence copula $C_{\perp}(e_1, e_2) = e_1 \times e_2$, denoted as $\mathfrak{R}(\mathbf{v}) \sim C_{\perp}$ (Genest et al., 2009).

Based on the property of Rosenblatt’s transform, the null hypothesis $H_0: \mathbf{v} \sim C \in C_0$ is equivalent to $H_0^*: \mathfrak{R}_{\theta}(\mathbf{v}) \sim C_{\perp}$ for some $\theta \in \mathcal{G}$, where $C_0 = \{C_{\theta} : \theta \in \mathcal{G}\}$ is the specific parametric family of copulas and \mathcal{G} is an open subset of R^p for some integer $p \geq 1$ (Genest et al., 2009).

The steps of Rosenblatt transformation are as flows:

- (1) Null hypothesis $H_0: \mathbf{v} \sim C \in C_0$.
- (2) Under the null hypothesis H_0 , the empirical distribution function:

$$D_n(\mathbf{v}) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\mathbf{E}_i \leq \mathbf{v}), \quad \mathbf{v} \in [0, 1] \quad (34)$$

where $\mathbf{E}_i = \mathfrak{R}_{\theta_i}(U_i)$ ($i = 1, 2, \dots, n$) are pseudo-observations from the independence copula C_{\perp} ; U_i ($i = 1, 2, \dots, n$) are pseudo-observations from the copula C ; $\mathbf{v} = (u, v)$ denotes the marginal distributions of random variables X and Y ; n is the sample size.

The Cramér von Mises statistic can be used to test the null hypothesis H_0 (Genest et al., 2009):

$$S_n^{(B)} = n \int_0^1 \{D_n(\mathbf{v}) - C_{\perp}(\mathbf{v})\}^2 d\mathbf{v} = \frac{n}{3^2} - \frac{1}{2} \sum_{i=1}^n (1 - E_{i1}^2)(1 - E_{i2}^2) + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (1 - E_{i1} \vee E_{j1})(1 - E_{i2} \vee E_{j2}) \quad (35)$$

where $a \vee b = \max(a, b)$. The corresponding P-value of the Cramér von Mises test statistic is:

$$P\text{-value}^{[Cramer-von Mises]} = \frac{1}{N} \sum_{k=1}^N \{1(S_{n,k}^{(B)*} - S_n^{(B)}) \geq 0\} \quad (36)$$

3. Case Study

3.1. Selection of the Measuring Sites

Regina is the capital city of the Province Saskatchewan, Canada, with a population of more than 200,000. The City of Regina can be reached by several highways including the Trans-Canada Highway (#1 highway of Canada) from both the west and east sides and four provincial highways from other directions. The city is served with the Ring Road, a high speed connection between Regina’s east and northwest that loops around the city’s east side (the west side of the loop is formed by a busy road named Lewvan Drive). Future plans will be conducted to construct another perimeter highway to encircle the city farther out. This type of highways with no traffic lights is one of the important sources of traffic noise in most cities in Canada.

In order to investigate the dependence between traffic flow and traffic noise pressure, two locations, named the South and North site, as shown in Figure 1, are selected as the measuring sites. These two sites are distributed along the Trans-Canada

Highway on which the speed limit is 100 km/h. More significantly, the sites are very close to each other with similar traffic and road construction conditions. The South site is located along Assiniboine Ave to Wascana Pkwy in the City, and the North site is on the opposite side of the South site, located from Wascana Pkwy to Assiniboine Ave. The field experiment was conducted two times, with each time lasting one week (May 24 to 31, and July 12 to 19).

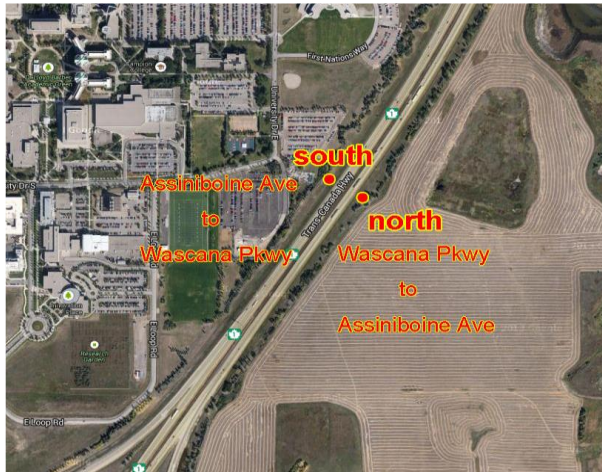


Figure 1. Selection of the measuring sites.

3.2. Data Collection

In the field experiment, the sound pressure levels are measured by utilizing the B&K Modular Precision Sound Analyzer 2260 in the experiment. Sound Analyzer 2260 supports one-channel measurements of environmental noise and noise at work, including real-time analyses of sound, and measurements of reverberation time. The resolution/accuracy for this instrument is 6.3 Hz to 20 kHz in 1/3-octaves of the frequency range covered and it can be extended to allow measurements of reverberation time in 1/1- or 1/3-octaves. To ensure the quality of the experimental data, the equipment is firmly fixed on the measuring sites with cement blocks as shown in Figure 2. In addition, a power box is set beside the equipment to provide electric power for a longtime required for the testing. As can be seen in Figure 2, all the equipment and accessories were set in a sealed case. The case was locked and fixed on the cement blocks to prevent any possible damage.

Traffic noise from a stream of vehicles varies over time in strength depending on many different factors such as number of vehicles passed by, speed, weather conditions in the tests and so on. The time averaged noise level is employed to convert the fluctuating values of a certain time interval into a simple mean value. In current research, a 1-hour time-averaged noise level (L_{Aeq1h}) is employed as temporal noise indicator for further analysis. The layout of the microphone adopted in the experiment is shown in Figure 3. The microphone is placed 1.5 m above the ground, and 7.5 m away from the center line of the highway, as can be seen in the figure. The data collected and analyzed are the One-Hour Time-Averaged Noise

Pressure Level in terms of equivalent continuous sound level L_{Aeq} (dB) and Traffic Flow Q is measured with the unit in vehicle/hour, indicating the passing motor vehicles in a unit time. In this research, the traffic flow at the testing spots along Trans-Canada Highway are collected and supplied by the City of Regina. Tables 1 and 2 present the statistical characteristics for field measurements of the traffic flow and traffic noise

Table 1. The Statistic Characteristic for the Traffic Flow and Traffic Noise at the North Site

		North		
NO.			Traffic Flow	Traffic Noise
1	Percentile	Minimum	104.00	41.10
		25%	591.00	54.60
		50%	1271.00	61.50
		75%	1631.00	63.40
		Maximum	2869.00	72.60
2	Range		2765.00	31.50
3	Mean		1175.58	59.41
4	Std		641.29	6.18
5	Skewness		0.01	-0.33
6	Kurtosis		-0.61	-0.35

Table 2. The Statistic Characteristic for the Traffic Flow and Traffic Noise at the South Site

		South		
NO.			Traffic Flow	Traffic Noise
1	Percentile	Minimum	87.00	40.30
		25%	625.00	49.60
		50%	1241.00	59.40
		75%	1699.00	63.70
		Maximum	2724.00	69.80
2	Range		2637.00	29.50
3	Mean		1195.50	56.80
4	Std		681.65	8.45
5	Skewness		0.11	-0.38
6	Kurtosis		-0.76	-1.18

4. Result Analysis and Discussion

4.1. Marginal Distribution Estimation through Entropy Method

Based on the principal of maximum entropy (POME) theory, the marginal probability density functions (PDFs) and cumulative probability functions (CDFs) can be expressed as Equations (10) and (11). Then the Conjugate Gradient (CG) method was employed to generate the unknown parameters in Equation (10), with the initial values of $\lambda_{(0)}$ being the mean, standard deviation (i.e. Std), skewness, and kurtosis. The generated marginal CDFs were compared with the empirical CDFs estimated from the Gringorten plotting position formula (Gringorten, 1963). Figure 4 shows the comparison between the theoretical probabilities, calculated through the ME-based PDF, and the empirical probabilities obtained through Gringorten plotting position formula at the North site. Figure 5 presents



Figure 2. Experimental instruments and setup.

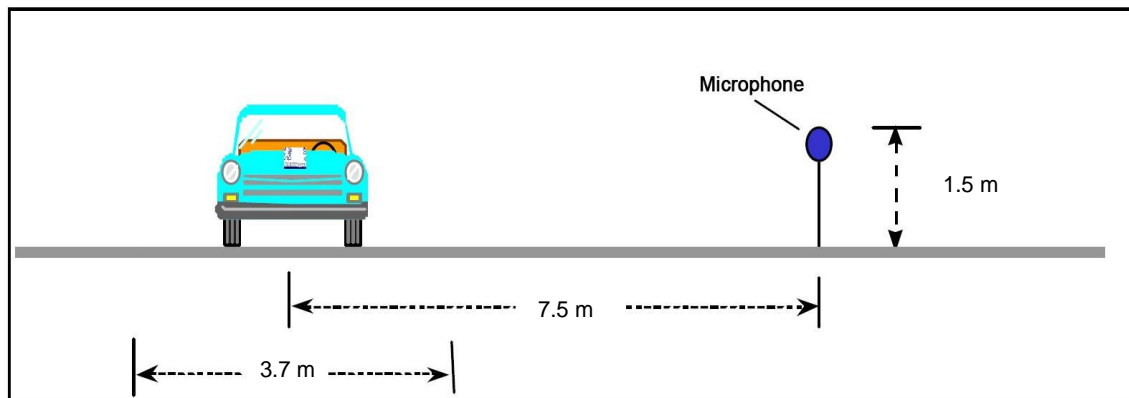


Figure 3. Layout of microphone adopted in the experiments.

the comparison of the theoretical probabilities and the empirical probabilities at the South site. They indicate that the ME-based PDFs can fit the empirical probabilities very well, but the ME-based PDFs for the traffic flows performed better than those for the traffic noise.

To further evaluate the performance of the entropy method in quantifying the probability distributions for the traffic volumes and traffic noise levels, the RMSE values, K-S test and A-D test are employed to reflect the effectiveness of the ME-based marginal distributions. As presented in Table 3, the P-values calculated from both K-S test and A-D test are higher than 0.05 for both traffic volumes and noise levels at the North and South sites. These results indicate that the null hypothesis cannot be rejected, and the marginal distributions obtained through the entropy method can appropriately quantify the distribution of the measured traffic volume and noise. Furthermore, the P-values of K-S and A-D test for traffic flow are higher

than those for traffic noise at the North Site, while the RMSE value for traffic flow is smaller than that for traffic noise. These indicate that the ME-based PDF performed better for traffic volume than that for traffic noise at the North site. Similar condition appears at the South site. Such results are consistent with the results showed in Figures 4 and 5.

4.2. Joint Probability Distributions for Traffic Flow and Traffic Noise through the Gaussian and Student *t* Copulas

Traffic volume is one of the most important factors in traffic noise emission. Before modelling the dependence between traffic volume and noise in the highway of the City of Regina, the correlations between traffic volume and traffic noise are investigated. The Pearson's, Spearman's and Kendall's correlation coefficients were employed in this study to analyze the inherent relationship between traffic volume and noise. As shown in Table 4, the traffic noise and traffic volume is highly

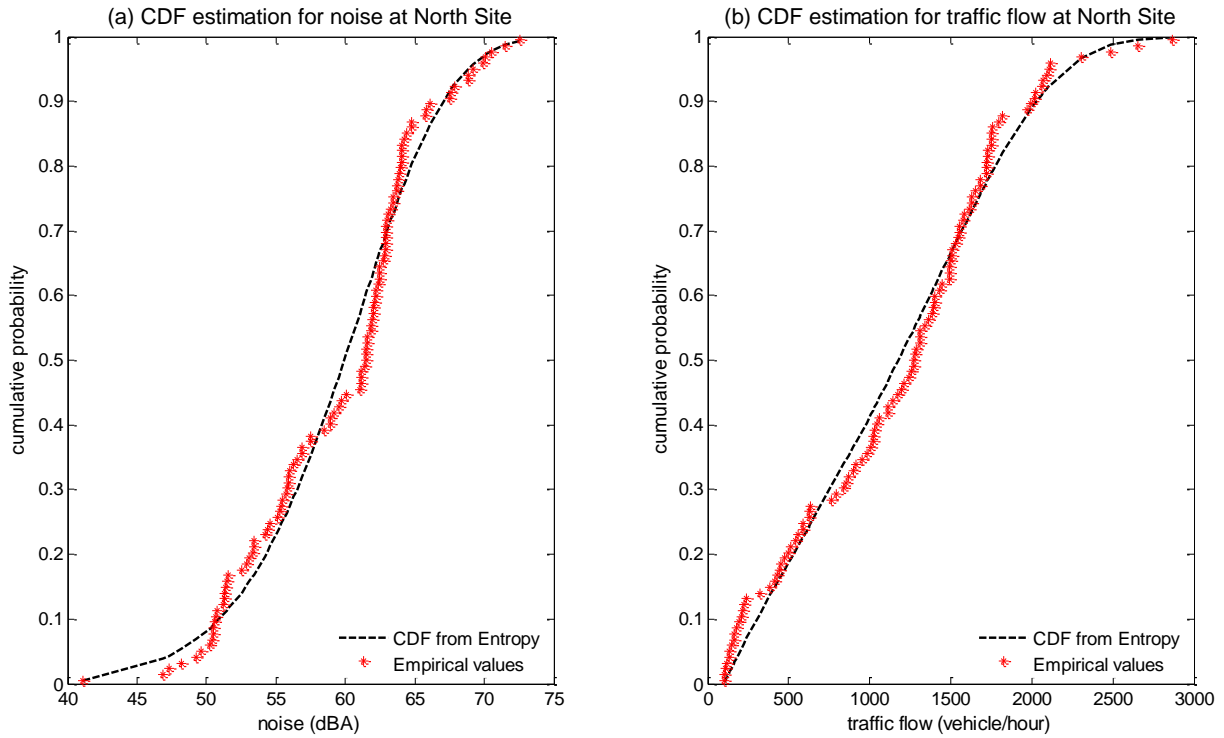


Figure 4. Comparison between the theoretical and empirical probabilities at the North site.

Table 3. The Goodness-of-fit for the ME-based Marginal Distributions

Methods	Index	K-S test		A-D test		RMSE
		T	P-value	W^2_n	P-value	
North	Traffic Flow	0.0678	0.3450	0.4879	0.2193	0.0341
	Noise	0.1075	0.0716	0.6389	0.0929	0.0477
South	Traffic Flow	0.0514	0.4834	0.3722	0.4157	0.0273
	Noise	0.0918	0.1034	0.5242	0.1787	0.0410

monotone association. Specifically, the correlation coefficients in Table 4 indicate a positive monotonic association between traffic noise and volume.

To further quantify the inherent interrelationship between the traffic volume and noise, the Gaussian copula and Student *t* copula are applied to model the dependence between these two variables. The parameters in the Gaussian and Student *t* copulas were estimated through the maximum pseudolikelihood estimation method. Figure 6 shows the joint CDF and PDF of traffic volume and noise, obtained through the Gaussian and Student *t* copulas, at the North site. Figure 7 presents the joint CDF and PDF of traffic volume and noise at the South site. As shown in Figure 6, the PDFs generated through the Gaussian and Student *t* copulas showed an apparently upper tail dependence between traffic noise and volume, indicating the fact that extreme traffic volume would usually result in extreme traffic noise level. A similar upper tail dependence appeared in the joint PDFs at the South site, as can be seen in Figure 7. In Figures 6 and 7, the values on the z axis are obtained through the joint PDF with respect to the standard values of

traffic volume and traffic flow, which vary within the unit interval (i.e. [0, 1]). This leads to extremely high values of pdf. Furthermore, to apparently reveal the probabilistic distribution characteristics of traffic flow and traffic noise, the values on the *x* and *y* axis are set as the original data.

To identify the performance of the Gaussian and Student *t* copulas in modelling the joint probability of traffic volume and traffic noise, comparisons between the theoretical and empirical joint cumulative probabilities were conducted, in which the empirical probabilities were calculated based on the bivariate Gringorten plotting position formula (Zhang and Singh, 2006), and the theoretical probabilities were obtained through the estimated Gaussian and Student *t* copulas Figure 7

Table 4. Values of Correlation Coefficients between Traffic Flow and Traffic Noise (95% confidence level)

Site.	Pearson's r	Kendall's tau	Spearman's rho
North	0.8159	0.6015	0.7774
South	0.8545	0.6634	0.8616

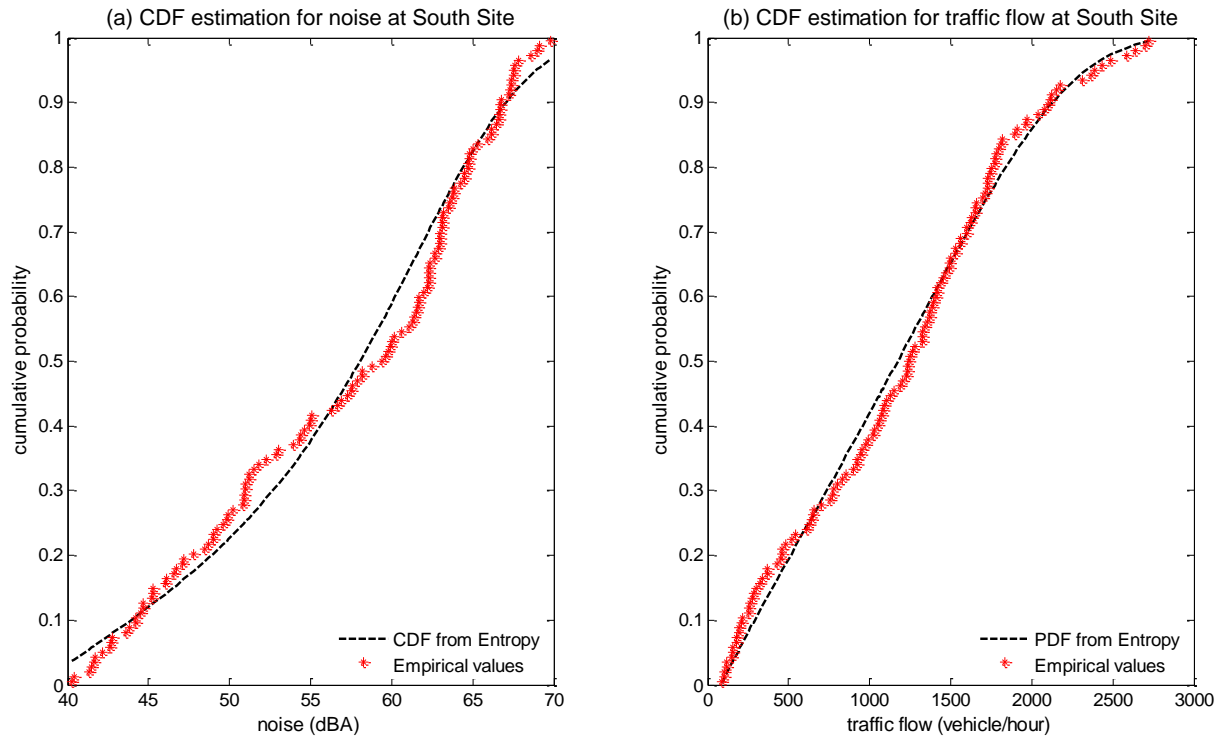


Figure 5. Comparison between the theoretical and empirical probabilities at the South site.

shows the comparison between the theoretical and empirical probabilities at the North site. It indicates that the Gaussian and Student *t* copulas produced good graphical fits to the empirical probabilities at the North site. Figure 8 presents the comparison between the theoretical and empirical probabilities at the South site. Similarly, the proposed Gaussian and Student *t* copulas performed well for plotting the joint probability of traffic volume and noise at the South site.

To further identify the performance of the Gaussian and Student *t* copulas, the root mean square error (RMSE), and the Rosenblatt transformation with Cramér von Mises statistic are employed to test the goodness of fit of sample data to the theoretical joint distribution obtained using copula functions. Table 5 presents the Cramér von Mises statistic associated with P-values and the RMSE values for joint distributions obtained through Gaussian and Student *t* copulas for traffic flow and noise. The results indicate that the Gaussian and Student *t* copulas produced satisfied performance in quantifying the joint probability distributions between traffic volume and noise, since all the P-values of the Cramér von Mises statistics are higher than the significant level (i.e. $\alpha = 0.5$). Furthermore, at both measurement sites, the differences between Gaussian and Student *t* copulas for traffic volume and noise are rarely small. For example, the RMSE value for the Gaussian and Student *t* copula at the North site is 0.0442 and 0.0438, respectively, while RMSE value for the Gaussian and Student *t* copula at the South site is 0.0362 and 0.0364, respectively. These results indicate that both the Gaussian and Student *t* copulas can be applied to represent the joint distribution of traffic volume and

noise.

4.3. Conditional Probability Distributions for Traffic Noise under Different Traffic Volume Scenarios

Based on the results presented in Table 5, both the Gaussian and Student *t* copulas can be applied to quantify the joint probability distributions of traffic volume and noise at the two sites. Consequently, the conditional cumulative probability functions (CDF) of traffic volume and traffic noise can be derived through the established copula functions. In current study, the traffic noise is the main concern under consideration due to its various negative effects on human health. Consequently, the conditional distribution of the traffic noise are only analyzed under various traffic volume scenarios. The traffic volume scenarios are designed based on the field experiment measurements, in which the 5, 25, 50, 75 and 95% percentile values are calculated based on experimental measurements.

In this research, the conditional CDFs of the traffic noise with the traffic volume less than some predetermined percen-

Table 5. The Goodness-of-fit for the Gaussian and Student *t* copulas

Site	Copulas	Cramér von Mises statistic		RMSE
		T	P-value	
North	Gaussian	53.8692	0.4145	0.0442
	Student <i>t</i>	53.6148	0.4345	0.0438
South	Gaussian	66.6018	0.8164	0.0362
	Student <i>t</i>	66.5498	0.8064	0.0364

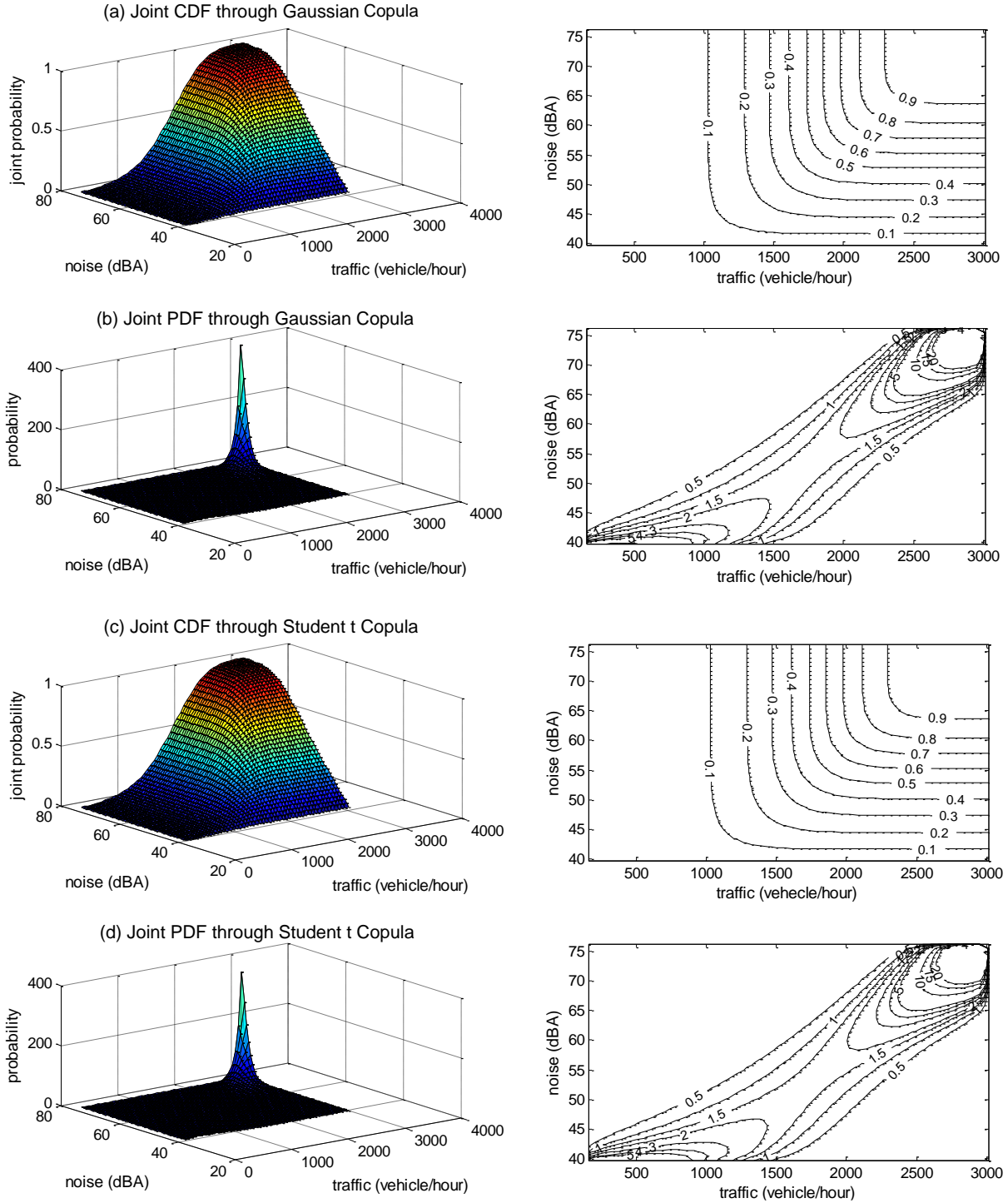


Figure 6. The joint CDF and PDF for traffic volume and noise at the North site.

tile values are produced based on Equation (25). This is because that compared with the traffic noise emission under consistent traffic volume, people pay more attention to the traffic noise under a traffic volume limit on the highway. Figure 10 shows the conditional CDFs of the traffic noise at the North site, which are generated through the established Gaussian and Student *t* copulas (shown in Figure 6). It can be seen that, the

values of conditional cumulative probabilities for traffic noise would decrease as the value of the traffic volume increases. This indicates a positive correlation structures between the traffic volume and traffic noise, which is consistent with the correlation coefficients presented in Table 4. Furthermore, as shown in Figure 10, the traffic noise would probably be no more than 30 dBA when the traffic volume is less than 5% of the

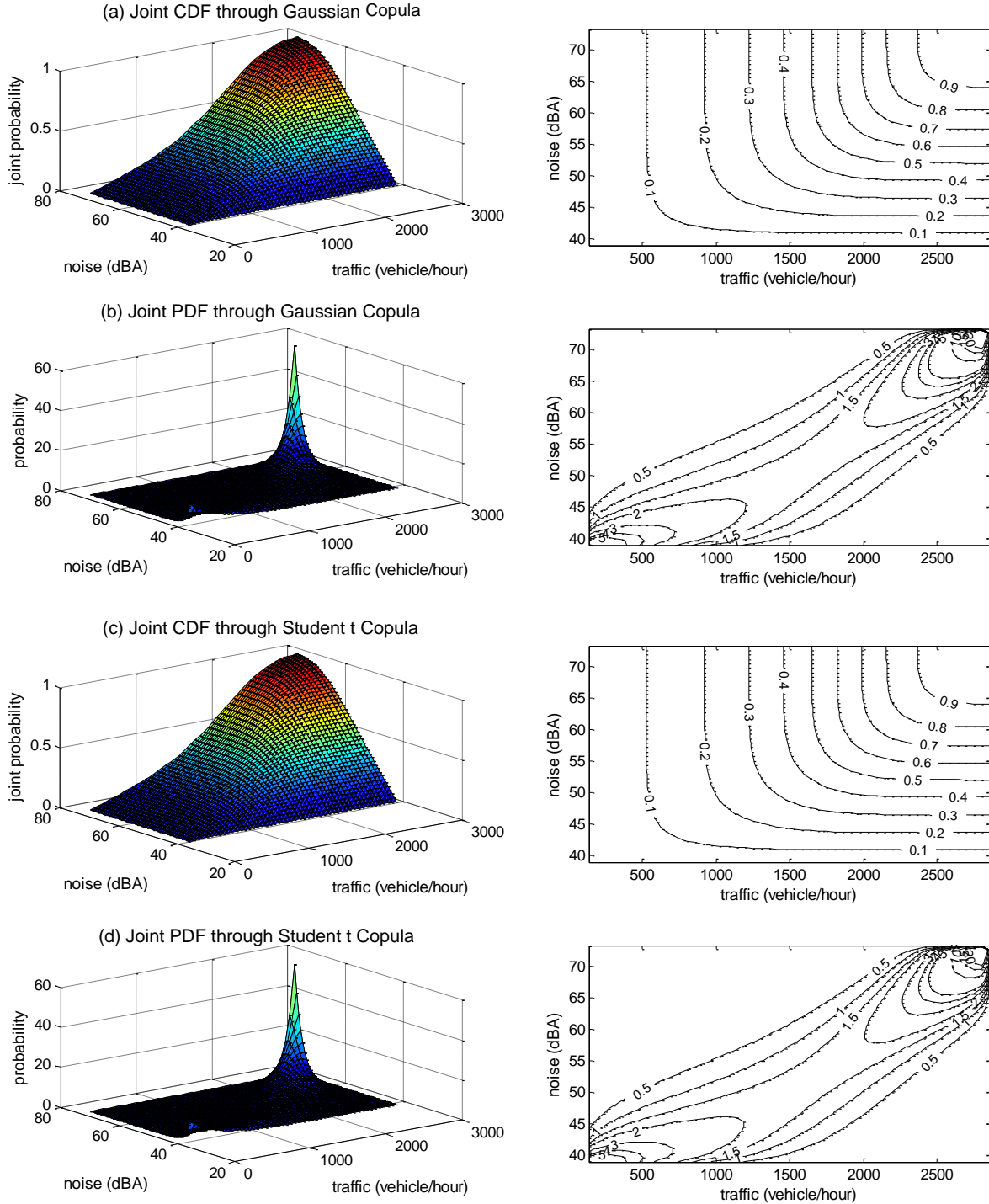


Figure 7. The joint CDF and PDF for traffic volume and noise at the South site.

measurement records. Similarly condition can also be found in the conditional CDF of traffic noise at the South site, as presented in Figure 11.

4.4. Conditional Probability Density Function for Traffic Noise under Different Traffic Volume Scenarios

In addition to derive the conditional cumulative distribu-

tion functions based on the Gaussian and Student *t* copulas for traffic noise under different traffic volume scenarios, the conditional probability density functions of traffic noise can also be generated based on Equations (26) - (29). In traffic noise control practices, the traffic volume would be a critical factor for traffic noise pressures. Given constant traffic composition, speed and driving patterns, some research work reported that a 50% reduction of the traffic volume results in a 3 dB reduc-

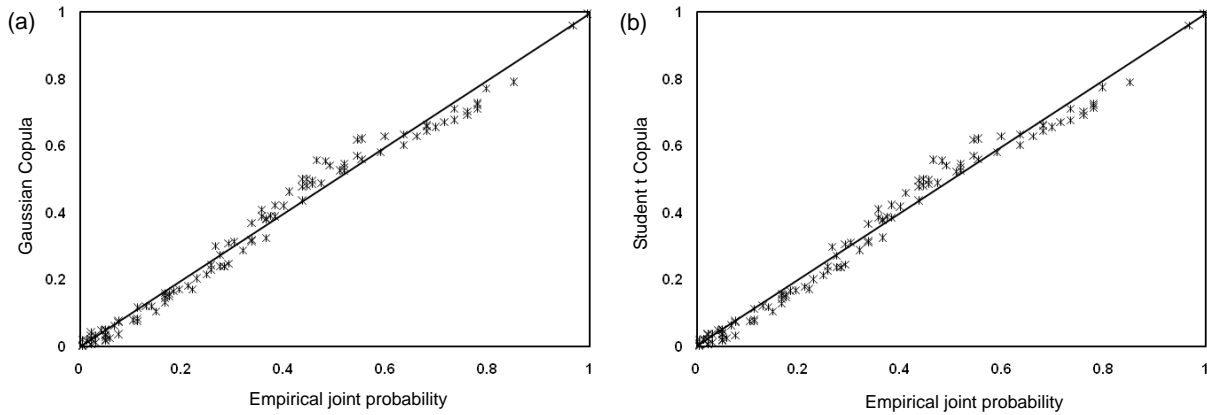


Figure 8. Comparison between the theoretical and empirical joint cumulative probabilities at the North site.

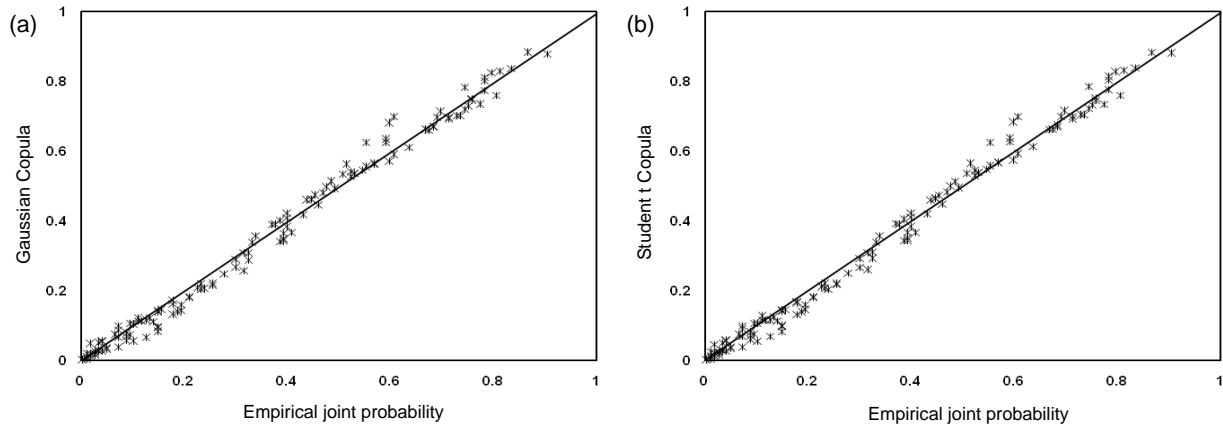


Figure 9. Comparison between the theoretical and empirical joint cumulative probabilities at the South site.

tion in noise level, regardless of the absolute number of vehicles (Ellegjerg, 2007). However, due to extensive uncertainties in road traffic systems (e.g. traffic speed, composition, and so on), the actual traffic noise reduction effect resulting from traffic volume reduction would also be uncertain. The probability density distribution of the traffic noise under different traffic volume scenarios can well reflect such uncertainty and be applied to explore the traffic noise reduction through traffic volume management.

Figure 12 shows the distributions of traffic noise pressure conditional on different traffic flow volumes at the North site. In this study, the traffic volumes scenarios are assume to be 5, 25, 50, 75 and 95% percentile of the field experimental measurements. Each curve represents a probability distribution function (PDF) of the traffic noise associated with certain traffic volume. Since both the Gaussian and Student *t* copulas can well quantify the joint probability of traffic noise and traffic flow, the two copula functions are all employed to derive the conditional PDF of the traffic noise under different traffic flow scenarios. According to the PDFs of traffic noise, as the traffic volume increases, the traffic noise would be expected to be higher as well. For example, if 1,631 (75% percentile of the measurements) vehicles pass by the North site in one hour, the

noise pressure is likely to be more intense than the traffic volume scenario of 1,271 vehicle/hour (50% percentile of the measurements). Moreover, when high traffic volume occurred, the distribution of the traffic noise is narrow around its mode. For instance, the PDF associated with the traffic volume of 591 vehicle/hour is wider than the PDF associated with the traffic volume of 1,271 vehicle/hour. This may result to a larger range of traffic noise given the 25% percentile traffic volume, as compared to the 50% percentile traffic volume.

Figure 13 shows the distributions of traffic noise conditional on different traffic volume scenarios at the South site. Similarly with the conditional PDFs at the North site. The conditional PDFs at the South site, obtained through the Gaussian and Student *t* copulas, indicate that the increase of traffic volume would also lead to increase of the traffic noise. This is due to the positive dependence between the traffic volume and traffic noise. Furthermore, different from the conditional PDFs at the North site, the distribution of the traffic noise is narrow around its mode at very high and very low traffic flow scenarios. As can be seen from Figure 13, the PDFs associated with 5 and 95% percentiles of the traffic volume is narrower than the PDFs associated with the 25, 50, 75% percentiles of the traffic volume.

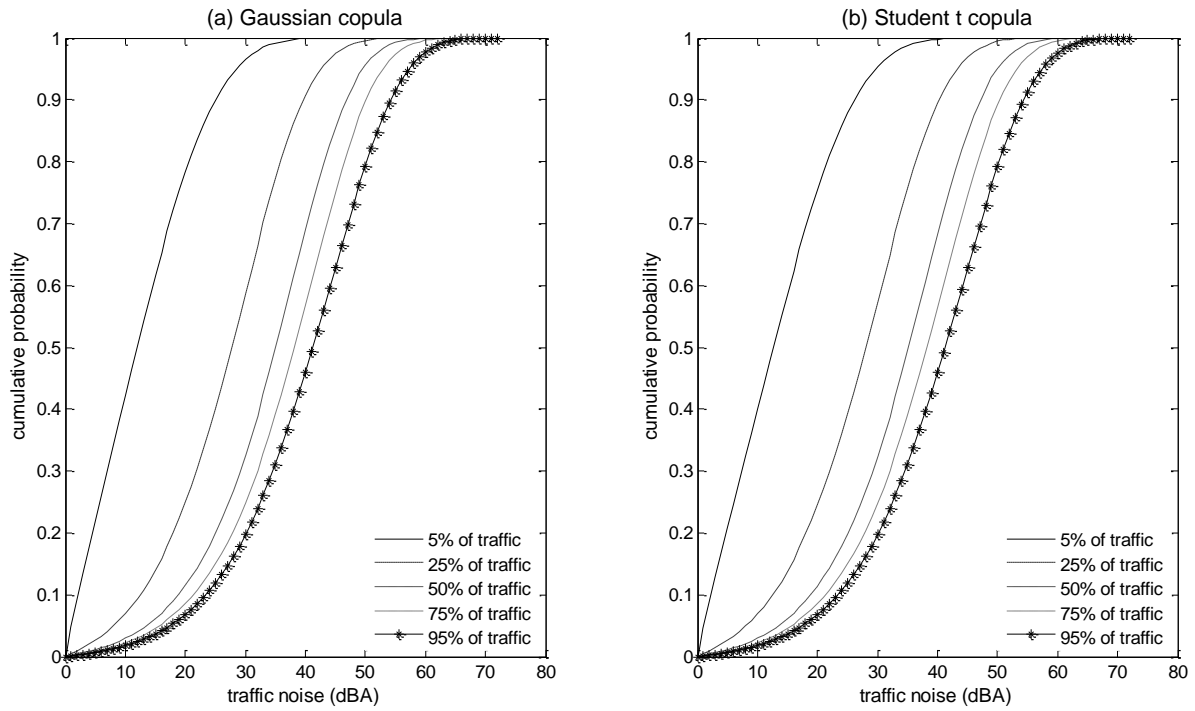


Figure 10. The conditional CDF of traffic noise under different traffic volume scenarios at the North site.

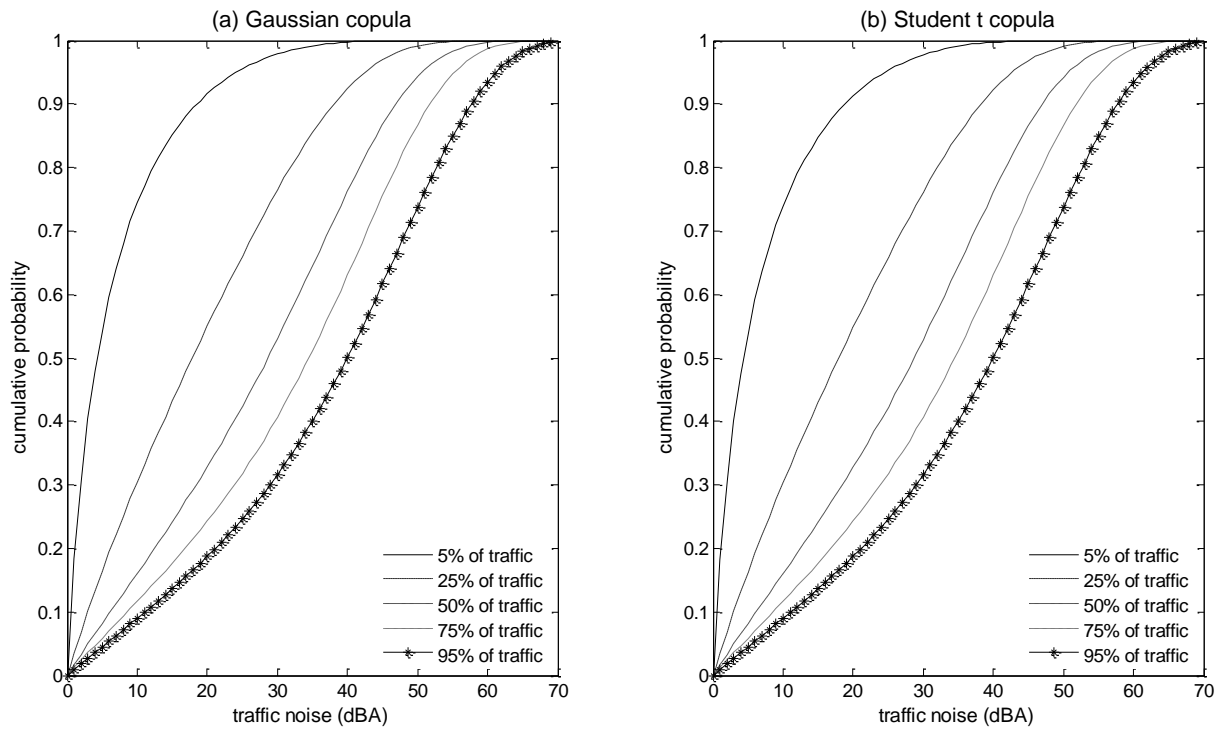


Figure 11. The conditional CDF of traffic noise under different traffic volume scenarios at the South site.

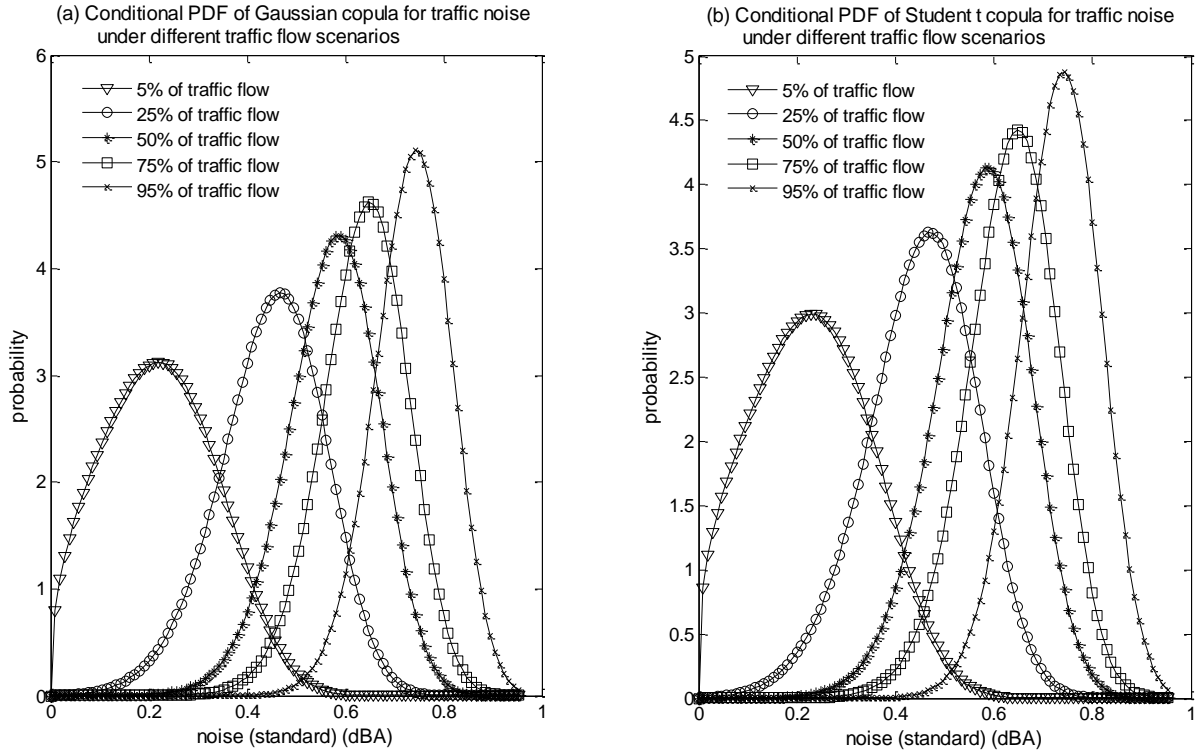


Figure 12. The conditional PDF of traffic noise under different traffic volume scenarios at the North site.

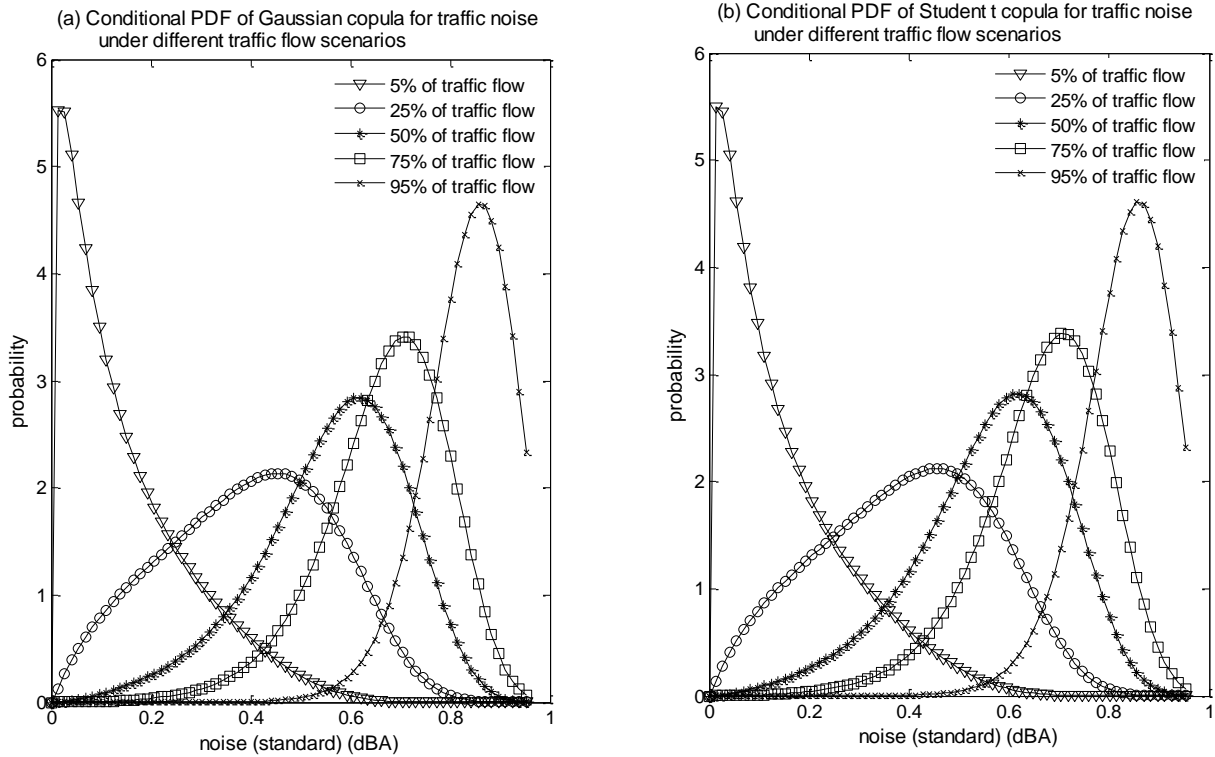


Figure 13. The conditional PDF of traffic noise under different traffic volume scenarios at the South site.

The conditional PDFs of the traffic noise under different traffic flow scenarios can indicate the potential reduction effect due to the decrease of the traffic volume. As can be seen from Figures 12 and 13, decrease of the traffic volume would likely reduce the traffic noise. However, at the same site, the reduction effect would be different as the difference in the traffic flow reduction. For example, the noise reduction effect seems to be increase as the decrease of the traffic flow in both North and South sites. Moreover, as shown in Figures 12 and 13, it seems that, as the decrease of the traffic flow, the noise reduction effect at the South site would be more significant than that at the North site.

4.5. Discussion

The interrelationship between traffic flow and traffic noise is complex and can hardly be well reflected through some simple statistical analysis. Moreover, extensive uncertainties existing in the traffic system will intensify the complexity of such an interrelationship. The proposed entropy-copula approach can model the dependence between traffic flow and traffic noise through a joint probability. Furthermore, the CDFs and PDFs of the traffic noise, conditional on different traffic flow scenarios, can be derived based on the copula function. Such conditional CDFs and PDFs can provide useful information for decision makers for traffic noise mitigation through traffic flow management.

Figures 10 ~ 13 shows the conditional CDFs and PDFs for the traffic noise at the measurement sites under different traffic flow scenarios. These conditional probability distributions can indicate the general probabilistic characteristics of traffic noise under different traffic flow control measures. For example, if the traffic flow reduces by 25%, then the conditional probabilistic distributions (CDFs and PDFs) can be obtained through the copula function. These conditional CDF and PDF can reveal the probabilistic distributional characteristics of the traffic noise, if the traffic flow is reduced by 25%.

Figures 10 and 12 show the conditional CDFs and PDFs of traffic noise at the North site, while Figures 11 and 13 indicates the conditional probabilistic distributions of traffic noise at the South site. Comparison of the conditional probabilistic distributions between the South and North sites reveal that traffic flow seems to play a more important role in traffic noise at South Site. This means that, if the traffic flow is reduced by the same amount, the noise at South Site would decrease more significantly. This is due to that the road at South Site has three lanes, and thus the same reduction rate of traffic flow would lead to more decrease in traffic density at South Site. Therefore, traffic noise mitigation through traffic flow control will be more effective at the South Site.

5. Conclusions

In this study, an entropy-copula method has been proposed for modelling dependence between traffic volume and traffic noise. In the proposed entropy-copula framework, the entropy method was applied to quantify the marginal probability dis-

tributions of traffic volume and noise, with the Lagrange multipliers in the ME-based PDF being obtained through the Conjugate Gradient (CG) method. Afterwards, the joint probability distributions (i.e. joint PDF and joint CDF) were established through the Gaussian and Student t copulas, with the unknown parameters in the copula being obtained through the maximum pseudo-likelihood estimation method. Various goodness-of-fit test methods were proposed to evaluate the performance of the entropy method in estimating the marginal probabilities, and the Gaussian and Student t copulas in modelling the joint probabilities of traffic volume and noise.

The proposed entropy-copula method was applied for modelling the dependence of the traffic volume and traffic noise on the Trans-Canada Highway (#1 Highway) in the City of Regina, based on a series of field experiments. The results showed that the entropy method could appropriately quantify the probability distributions of traffic volume and traffic noise, with low RMSE values and the P-values of K-S and A-D tests being larger than 0.5. Meanwhile, both the Gaussian and Student t copulas could well model the joint probability distributions for the traffic volume and traffic noise at the both experiment sites, with low RMSE values and the P-values of the Cramér von Mises statistics are higher than the significant level (i.e. $\alpha = 0.5$). Also, comparison between the theoretical joint probabilities, obtained by the copulas, and the empirical joint probabilities demonstrated the well performance of the copula methods in modeling the joint probabilities for traffic volume and traffic noise. Furthermore, the conditional CDFs of the traffic noise at the two experiment sites, were derived based on the established Gaussian and Student t copulas under different traffic volume scenarios. The obtained conditional CDFs indicated positive structures between traffic volume and traffic noise at the both experiment sites.

The traffic noise are attracting increasing attention from various researchers due to its various negative effect on humans. The impacts of traffic noise comes from amounts of aspects, including traffic system, weather conditions and road conditions. This study attempted to model the dependence of the traffic noise and its impact factor (i.e. traffic volume in this study) from a statistical view. However, since the traffic noise emission are influenced by many factors, multivariate copula methods are required to explore the impacts from multiple factors on the traffic noise emission. Moreover, one of the underlying assumptions of the entropy-copula approach is that the marginal distributions of traffic flow and traffic noise are continuous. Such an assumption may not be true for traffic flow, which requires further study to construct copulas with respect to discrete and continuous random variables.

Acknowledgements: This research was supported by the National Key Research and Development Plan (2016YFC0502800).

References

- Abo-Qudais, S., and Alhiary, A., (2005). Effect of traffic characteristics and road geometric parameters on developed traffic noise levels. *Can. Acoustics*, 33(1), 43-50.

- Agarwal, S., and Swami, B.L., (2011). Comprehensive approach for the development of traffic noise prediction model for Jaipur city. *Environ. Monit. Assess.*, 172(1-4), 113-120. <http://dx.doi.org/10.1007/s10661-010-1320-z>
- Alves Filho, J.M. (1997). *Influence of Traffic Composition on the Noise Generated on Highways*, Master's dissertation in Mechanical Engineering, Federal University of Santa Catarina.
- Anderson, T.W., and Darling, D.A. (1954). A test of goodness of fit. *J. Am. Stat. Assoc.*, 49(268), 765-769. <http://dx.doi.org/10.1080/01621459.1954.10501232>
- Arshad, M., Rasool, M.T., and Ahmad, M.I. (2003). Anderson Darling and modified Anderson Darling tests for generalized Pareto distribution. *Pakistan J. Appl. Sci.*, 3(2), 85-88. <http://dx.doi.org/10.3923/jas.2003.85.88>
- Billera, D., Schmidt, B., and Miller, W. (1997). *Stop the Whine! Narrow-Band Noise Level Measurements of Three Highway Pavements*, Transportation Research Record 1601, TRB, National Research Council, Washington, D.C., pp. 71-76, 1997.
- Conover, W.J. (1999). *Practical Nonparametric Statistics*, Third Edition, John Wiley & Sons, Inc. New York, pp. 428-433.
- Chen, L., Ye, L., Singh, V., Zhou, J., and Guo, S. (2013). Determination of input for artificial neural networks for flood forecasting using the copula entropy method. *J. Hydrol. Eng.*, 19(11). [http://dx.doi.org/10.1061/\(ASCE\)HE.1943-5584.0000932](http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000932)
- Dai, B.L., He, Y.L., Mu, F.H., Xu, N., and Wu, Z. (2013). Development of a traffic noise prediction model on inland waterway of China using the FHWA. *Sci. Total Environ.*, 482-483, 480-485. <http://dx.doi.org/10.1016/j.scitotenv.2013.06.019>
- da Paz, E.C., Ferreira, A.M.C., and Zannin, P.H.T. (2005). Comparative study of the perception of urban noise. *J. Public Health*, 39(3), 467-472.
- da Paz, E.C., and Zannin, P.H.T. (2010). Urban daytime traffic noise prediction models. *Environ. Monit. Assess.*, 163(1-4), 515-529. <http://dx.doi.org/10.1007/s10661-009-0854-4>
- Demarta, S., and McNeil, A.J. (2005). The t copula and related copulas. *Int. Stat. Rev.*, 73(1), 111-129. <http://dx.doi.org/10.1111/j.1751-5823.2005.tb00254.x>
- Ellebjerg, L. (2007). *Effectiveness and Benefits of Traffic Flow Measures on Noise Control*, report to the European Commission silence project.
- Engel, M.S., de Vasconcelos Segundo, E.H., and Zannin, P.H.T. (2014). Statistical analysis of a combination of objective and subjective environmental noise data using factor analysis and multinomial logistic regression. *Stochastic Environ. Res. Risk Assess.*, 28(2), 393-399. <http://dx.doi.org/10.1007/s00477-013-0759-1>
- Fang, H.B., Fang, K.T., and Kotz, S. (2002). The meta-elliptical distributions with given marginals. *J. Multivariate Anal.*, 82(1), 1-16. <http://dx.doi.org/10.1006/jmva.2001.2017>
- Farrel, P.J., and Stewart, K.R. (2006). Comprehensive study of tests for normality and symmetry: Extending the Spiegelhalter test. *J. Stat. Computation Simul.*, 76(9), 803-816. <http://dx.doi.org/10.1080/01629360500109023>
- Genest, C., R émillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insur. Math. Econ.*, 44(2), 199-213. <http://dx.doi.org/10.1016/j.insmatheco.2007.10.005>
- Gringorten, I.I. (1963). A plotting rule for extreme probability paper. *J. Geophys. Res.*, 68(3), 813-814. <http://dx.doi.org/10.1029/JZ068i003p00813>
- Hao, Z., and Singh, V.P. (2011). Single-site monthly streamflow simulation using entropy theory. *Water Resour. Res.*, 47(9), W09528. <http://dx.doi.org/10.1029/2010WR010208>
- Huang, G.H. (1998). A hybrid inexact-stochastic water management model. *Eur. J. Oper. Res.*, 107(1), 137-158. [https://doi.org/10.1016/S0377-2217\(97\)00144-6](https://doi.org/10.1016/S0377-2217(97)00144-6)
- Huang, G.H., and Loucks, D.P. (2000). An inexact two-stage stochastic programming model for water resources management under uncertainty. *Civ. Eng. Environ. Syst.*, 17(2), 95-118. <https://doi.org/10.1080/02630250008970277>
- Huang, G.H., Huang, Y.F., Wang, G.Q., and Xiao, H.N. (2006). Development of a forecasting system for supporting remediation design and process control based on NAPL-biodegradation simulation and stepwise-cluster analysis. *Water Resour. Res.*, 42(6). <https://doi.org/10.1029/2005WR004006>
- Jaynes, E.T. (1957a). Information theory and statistical mechanics. *Phys. Rev.*, 106(4), 620. <http://dx.doi.org/10.1103/PhysRev.106.620>
- Jaynes, E.T. (1957b). Information theory and statistical mechanics. II. *Phys. Rev.*, 108(2), 171. <http://dx.doi.org/10.1103/PhysRev.108.171>
- Jaynes, E.T. (1982). On the rationale of maximum-entropy methods. *Proc. IEEE*, 70(9), 939-952. <http://dx.doi.org/10.1109/PROC.1982.12425>
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*, New York: Chapman and Hall, 1997. <http://dx.doi.org/10.1201/b13150>
- Kapur, J.N., and Kesavan, H.K. (1992). *Entropy Optimization Principles with Applications*, Academic Press, New York.
- Kelly, K.S., and Krzysztofowicz, R. (1997). A bivariate meta-Gaussian density for use in hydrology. *Stochastic Hydrol. Hydraul.*, 11(1), 17-31. <http://dx.doi.org/10.1007/BF02428423>
- Kong, X.M., Huang, G.H., Fan, Y.R., and Li, Y.P. (2015). Maximum entropy-Gumbel-Hougaard copula method for simulation of monthly streamflow in Xiangxi River, China. *Stochastic Environ. Res. Risk Assess.*, 29(3), 833-846. <http://dx.doi.org/10.1007/s00477-014-0978-0>
- Li, H.S. (2013). Maximum entropy method for probabilistic bearing strength prediction of pin joints in composite laminates. *Composite Struct.*, 106, 626-634. <http://dx.doi.org/10.1016/j.compstruct.2013.05.040>
- Li, Y.P., Huang, G.H., Nie, S.L., and Liu, L. (2008a). Inexact multistage stochastic integer programming for water resources management under uncertainty. *J. Environ. Manage.*, 88(1), 93-107. <https://doi.org/10.1016/j.jenvman.2007.01.056>
- Li, Y.P., Huang, G.H., Yang, Z.F., and Nie, S.L. (2008b). IFMP: Interval-fuzzy multistage programming for water resources management under uncertainty. *Resour. Conserv. Recycling*, 52(5), 800-812. <https://doi.org/10.1016/j.resconrec.2007.11.007>
- Li, Y.P., and Huang, G.H. (2009). Fuzzy-stochastic-based violation analysis method for planning water resources management systems with uncertain information. *Inf. Sci.*, 179(24), 4261-4276. <https://doi.org/10.1016/j.ins.2009.09.001>
- Li, Y.P., Huang, G.H., Huang, Y.F., and Zhou, H.D. (2009). A multi-stage fuzzy-stochastic programming model for supporting sustainable water-resources allocation and management. *Environ. Model. Software*, 24(7), 786-797. <https://doi.org/10.1016/j.envsoft.2008.11.008>
- Li, Y.P., Huang, G.H., and Nie, S.L. (2010). Planning water resources management systems using a fuzzy-boundary interval-stochastic programming method. *Adv. Water Resour.*, 33(9), 1105-1117. <https://doi.org/10.1016/j.advwatres.2010.06.015>
- Lv, Y., Huang, G.H., Li, Y.P., Yang, Z.F., Liu, Y., and Cheng, G.H. (2010). Planning regional water resources system using an interval fuzzy bi-level programming method. *J. Environ. Inf.*, 16(2), 43-56. <https://doi.org/10.3808/jei.201000177>
- Ma, M.W., Song, S.B., Ren, L.L., Jiang, S.H., and Song, J.L. (2013). Multivariate drought characteristics using trivariate Gaussian and Student t copulas. *Hydrol. Process.*, 27(8), 1175-1190. <http://dx.doi.org/10.1002/hyp.8432>
- Massey, J.F. (1951). The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc.*, 46(253), 68-78. <http://dx.doi.org/10.1080/010801010292010WR010208>

- 1621459.1951.10500769
- Mehdi, M.R., Kim, M., Seong, J.C., and Arsalan, M.H. (2011). Spatio-temporal patterns of road traffic noise pollution in Karachi, Pakistan. *Environ. Int.*, 37(1), 97-104. <http://dx.doi.org/10.1016/j.envint.2010.08.003>
- Nadarajah, S. (2006). Fisher information for the elliptically symmetric Pearson distributions. *Appl. Math. Computation*, 178(2), 195-206. <http://dx.doi.org/10.1016/j.amc.2005.11.037>
- Nadarajah, S., and Kotz, S. (2005). Information matrices for some elliptically symmetric distribution. *SORT: Stat. Oper. Res. Trans.*, 29(1), 43-56.
- Nelsen, R.B. (2006). *An Introduction to Copulas*, Springer, New York, 2006.
- Papalexioiu, S.M., and Koutsoyiannis, D. (2012). Entropy based deviation of probability distributions: A case study to daily rainfall. *Adv. Water Resour.*, 45, 51-57. <http://dx.doi.org/10.1016/j.advwatres.2011.11.007>
- Qin, X.S., Huang, G.H., Zeng, G.M., Chakma, A., and Huang, Y.F. (2007) An interval-parameter fuzzy nonlinear optimization model for stream water quality management under uncertainty. *Eur. J. Oper. Res.*, 180(3), 1331-1357. <https://doi.org/10.1016/j.ejor.2006.03.053>
- Razali, N.M., and Wah, Y.B. (2011), Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *J. Stat. Model. Analytics*, 2(1), 21-33.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Stat.*, 23(3), 470-472. <http://dx.doi.org/10.1214/aoms/1177729394>
- Reddy, M.J., and Ganguli, P. (2011). Application of copulas for derivation of drought severity-duration-frequency curves. *Hydrol. Process.*, 26(11), 1672-1685. <http://dx.doi.org/10.1002/hyp.8287>
- Shannon, C.E. (1948). A mathematical theory of communications. *Bell Syst. Tech. J.*, 27(3), 379-423. <http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Scholz, F.W., and Stephens, M.A. (1987). K-sample Anderson-Darling tests. *J. Am. Stat. Assoc.*, 82(399), 918-924. <http://dx.doi.org/10.2307/2288805>
- Shiau, J.T. (2006). Fitting drought duration and severity with two dimensional copulas. *Water Resour. Manage.*, 20(5), 795-815. <http://dx.doi.org/10.1007/s11269-005-9008-9>
- Shiau, J.T., Feng, S., and Nadarajah, S. (2007). Assessment of hydrological droughts for the Yellow River, China, using copulas. *Hydrol. Process.*, 21(16), 2157-2163. <http://dx.doi.org/10.1002/hyp.6400>
- Singh, V.P. (1997). The use of entropy in hydrology and water resources. *Hydrol. Process.*, 11(6), 587-626. [http://dx.doi.org/10.1002/\(SICI\)1099-1085\(199705\)11:6<587::AID-HYP479>3.0.CO;2-P](http://dx.doi.org/10.1002/(SICI)1099-1085(199705)11:6<587::AID-HYP479>3.0.CO;2-P)
- Singh, V.P. (2011). Hydrologic synthesis using entropy theory: Review. *J. Hydrol. Eng.*, 16(5), 421-433. [http://dx.doi.org/10.1061/\(ASCE\)HE.1943-5584.0000332](http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000332)
- Sklar, A. (1959). Fonction de répartition à n dimensions et leurs marges, *Publications de L'Institut de Statistique de L'Université de Paris* 8, Paris, pp. 229-231.
- Sraj, M., Bezak, N., and Brilly, M. (2014). Bivariate flood frequency analysis using the copula function: A case study of the Litja station on the Sava River. *Hydrol. Process.*, 29(2), 225-238. <http://dx.doi.org/10.1002/hyp.10145>
- To, W.M., Ip, R.C.W., Lam, G.C.K., and Yau, C.T.H. (2002). A multiple regression model for urban traffic noise in Hong Kong. *J. Acoust. Soc. Am.*, 112(2), 551-556. <http://dx.doi.org/10.1121/1.1494803>
- United Nations (UN) (2004). *World Urbanization Prospects: The 2003 Revision*, New York, 2004.
- Willmott, C.J., and Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.*, 30(1), 79-82. <http://dx.doi.org/10.3354/cr030079>
- Zhang, L., and Singh, V.P. (2006). Bivariate flood frequency analysis using the copula method. *J. Hydrol. Eng.*, 11(2), 150-164. [http://dx.doi.org/10.1061/\(ASCE\)1084-0699\(2006\)11:2\(150\)](http://dx.doi.org/10.1061/(ASCE)1084-0699(2006)11:2(150))
- Zhang, L., and Singh, V.P. (2012). Bivariate rainfall and runoff analysis using entropy and copula theories. *Entropy*, 14(9), 1784-1812. <http://dx.doi.org/10.3390/e14091784>
- Žežula, I. (2009). On multivariate Gaussian copulas. *J. Stat. Plann. Inference*, 139(11), 3942-3946. <http://dx.doi.org/10.1016/j.jspi.2009.05.039>