Producing Outdoor Design Conditions Based on the Dependency between Meteorological Elements: Copula Approach

Zhichao Jiao, Craig Farnham, Jihui Yuan, Kazuo Emura

***Abstract*—**It is common to use the outdoor design weather data to select the air-conditioning capacity in the building design stage. The meteorological elements of outdoor design weather data are usually selected based on their excess frequency separately while the dependency between the elements is not well considered. It means that the simultaneous occurrence probability of these elements is smaller than the original excess frequency which may cause an overestimation of selecting air-conditioning capacity. Therefore, the copula approach which can capture the dependency between multivariate data was used to model the joint distributions of the meteorological elements, like air temperature and global solar radiation. We suggest a method based on the specific simultaneous occurrence probability of these two elements of selecting more credible outdoor design conditions. The hourly weather data at 12 noon from 2001 to 2010 in Tokyo, Japan are used to analyze the dependency structure and joint distribution, the Gaussian copula represents the dependence of data best. According to calculating the air temperature and global solar radiation in specific simultaneous occurrence probability and the common exceeding, the results show that both the air temperature and global solar radiation based on simultaneous occurrence probability are lower than these based on the conventional method in the same probability.

***Keywords*—**Copula approach, Design weather database, energy conservation, HVAC.

I. INTRODUCTION

THIS study analyzed the marginal distribution functions and dependence structure of air temperature and global solar radiation using the copula method based on Tokyo hourly data from 2001-2010 by combining the marginal distribution functions and dependence structure into joint distribution functions. In order to improve the accuracy and reliability of outdoor design weather data for buildings, we propose to use the simultaneous occurrence probability based on the joint distribution instead of independent exceedance probabilities based on the joint distribution.

In the architectural design phase, in order to determine the appropriate building structure and capacity of air-conditioning equipment, we need the appropriate design weather data for design that can represent the local climate characteristics. In general, meteorological data for design purpose include the elements such as air temperature, absolute humidity, solar radiation, precipitation and wind speed and direction etc.

The air temperature, absolute humidity and solar radiation are the most important elements to determine the appropriate air conditioning capacity. Theoretically a copula can build a multivariate joint distribution model. As an initial attempt, we modeled a bivariate joint distribution model based on the dependence between global solar radiation and air temperature in this study. According to the Society of Heating, Air-Conditioning and Sanitary Engineers of Japan (SHASE) [1], the usual approach in air-conditioning system design involves the computation of peak design load at a specific hour of a design day using indoor and outdoor design conditions.

To prevent the air conditioning system from being over-designed due to the extreme conditions, the Technical Activities Committee (TAC) [2] of American Society of Heating, Refrigerating and Air-conditioning Engineers (ASHRAE) first used the concept of exceeding probability to select the design temperature. The TAC method is calculated by the concept of exceeding probability from a cumulative curve based on the outdoor hourly temperature data. It takes some specific excess frequency rate from the cumulative curve that is calculated by the hourly data for a year as the design conditions. The excess frequency rate is below a chosen percent of the curve; by 0.4%, 1.0%, or 2.0%. Similarly, the TAC method based on percentage frequency is also used in the UK [3] to generate design weather data. Besides the air temperature, design weather data also include precipitation and wind conditions, etc. In Japan, the TAC method is used to select other design conditions separately, like the global solar radiation and absolute humidity, etc. [4]. However, none of the above selection methods has considered the correlation between different meteorological elements. For example, as shown in Fig. 1, the 1% excess frequency rate of temperature and global solar radiation (Fig. 1 (a) red line) are chosen as the design conditions independently. Since temperature and solar insolation are not perfectly positively correlated, the probability of their simultaneous occurrence (Fig. 1 (b) red line) will be much less than 1% in practice, while the blue line at Fig. 1 (b) means the 1% simultaneous occurrence probability. Therefore, if multiple meteorological elements of design conditions are selected separately to directly combine by using the TAC method, it will result in over-designed air-conditioning capacity [5].

The purpose of this study is to generate design weather data for design by calculating air temperature and global solar

Zhichao Jiao*, Craig Farnham, and Kazuo Emura* are with the Housing and Environmental Design, Graduate School of Human Life Science, Osaka City University, Osaka, Japan (e-mail: jiaozhichao1995@ gmail.com; emura@osaka-cu.ac.jp).

Jihui Yuan is with the Architecture and Civil Engineering Department, Toyohashi University of Technology, 1-1 Hibarigaoka, Tempaku-cho, Toyohashi, Aichi 441-8580, Japan.

Open Science Index, Architectural and Environmental Engineering Vol:16, No:12, 2022 publications.waset.org/10012818.pdf

radiation for different simultaneous occurrence probabilities, such as 0.4%, 1.0%, or 2.0%, setting the probability of occurrence of each element as equal. Thus, we need the joint distribution function of air temperature and global solar radiation.

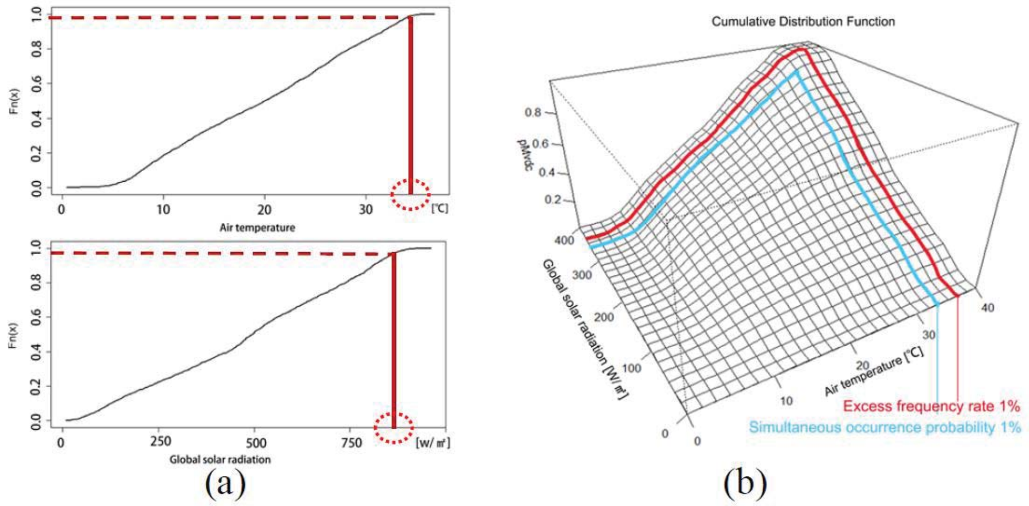


Fig. 1 (a) 1% excess frequency rate of temperature and global solar radiation, (b) simultaneous occurrence probability of air temperature and global solar radiation based on joint distribution

Traditional methods for building multivariate models require all the marginal distributions are from the same family, which does not fit in this study. Based on the above considerations, the copula approach, which allows us model the margins individually, is selected to satisfy the joint distribution and calculate the simultaneous occurrence probability in this study [6].

II. Methodology

A. Copula Method

Copula is a method that can be used to build multidimensional models. Sklar [7] first presented the central theorem of copula theory which is known as Sklar's Theorem. The copula method divides the multidimensional model into two parts: the marginal distributions of each univariate variables and the dependence structure between variables. Assume that $F(x_1, \ldots, x_d)$ is a multidimensional model based on a d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)^T$, $d = 1, \ldots, d$. $F_d(x_d)$ is the corresponding marginal distribution of X_d, $d = 1, \ldots, d$. Sklar's Theorem is given as:

$$F(x_1,\ldots,x_d) = C(F_1(x_1),\ldots,F_d(x_d)) \tag{1}$$

where C is a copula function which can represent the dependence structure between variables as mentioned previously. More specifically, the copula C is a multivariate cumulative distribution function, and the marginal probability distribution of each variable is a uniform distribution. Thus, we can consider that the marginal distribution has no effect on the dependence between the variables. This also shows that a copula can represent the dependence structure between variables.

B. Probability and Quantile Transformations

In order to determine the corresponding dependence between the variables, we need to transform the raw data into the copula format, which is the cumulative distribution function where the marginal function is a uniform distribution. For this purpose, since a cumulative distribution function is a uniform distribution, the probability integral transform was used to get the uniform cumulative distribution function of each variable. Then we can satisfy the copula function based on the empirical multivariate cumulative distribution function calculated by the marginal uniform distribution.

As shown in Fig. 2, it is assumed that the Fig. 2 (a) is the data of global solar radiation and air temperature, the copula points which is also known as the empirical copula. Data transformed by probability integral transform are shown on the Fig. 2 (b), we can find the most suitable two-dimensional theoretical copula function C based on data transformed copula points, then the theoretical copula function C is transformed to left by inverse probability integral transform [8], yielding the joint distribution (cumulative distribution) of air temperature and global solar radiation. In practice, since the formula calculation is very complex, we simulate a large number of samples based on copula C and distribution functions $F_1(x_1)$ and $F_2(x_2)$ to calculate the approximate joint distribution.

C. Model Estimation and Selection Criteria

The Maximum Likelihood Estimator method [9] was used to estimate the parameters of the copula in this study. The log-likelihood function is written as:

$$\theta_n = \arg\sup_{\theta\in\Theta} \sum_{i=1}^{n} \log c(F_1(x_{1,i}),\ldots,F_d(x_{d,i})) \tag{2}$$

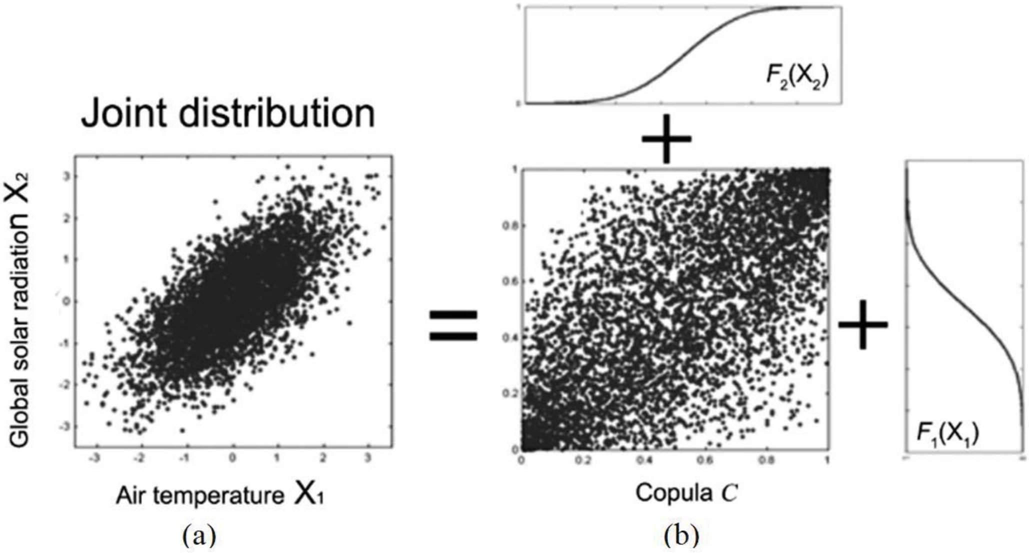


Fig. 2 Illustration of the copula theory

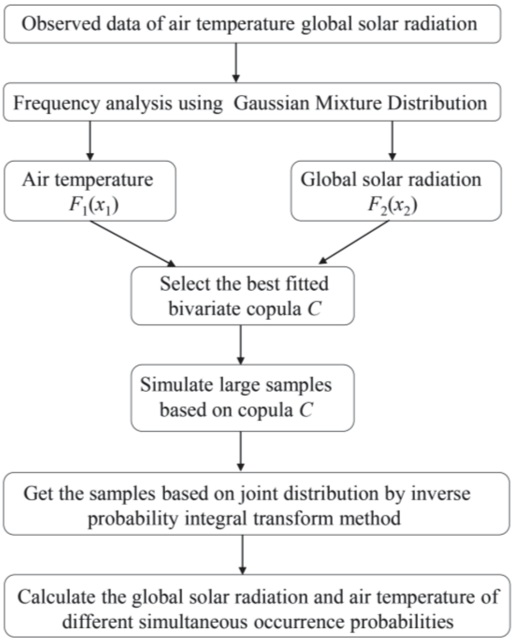


Fig. 3 Illustration of research process

where the θ_n is the parameter space and c are the density of copula C.

Two main selection criteria are used in this study, one is graphical diagnostics that can compare the characteristics of empirical and theoretical models, the other is the commonly used Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) [10], [11]. The AIC can be written as:

$$AIC = -2\log L + 2k \tag{3}$$

where k is the number of estimated parameters and log$L$ is the likelihood function.

The BIC can be written as:

$$BIC = -2\log L + k\log(n) \tag{4}$$

where n is the number of data points. Briefly, the models with lower AIC and BIC are generally preferred.

D. Research Process

The research process is shown as follows,

1) Fit the distributions of air temperature $F_1(x_1)$ and global solar radiation $F_2(x_2)$ respectively.
2) Search the best-fitted theoretical copula C based on the empirical copula that transformed by the probability integral transform method.
3) Simulate a large number (n = 10000) of copula samples based on copula C and transform them by inverse probability integral transform.
4) Calculate the air temperature and global solar radiation of different simultaneous occurrence probabilities based on the joint distribution.

Fig. 3 shows the illustration of research process.

III. Application of Copula

A. Location and Data

The air temperature and global solar radiation of Expanded AMeDAS [12] hourly weather data at 12 noon from 2001 to 2010 in Tokyo, Japan are used in this study. The scatter plot of data is shown in Fig. 4.

The marginal distribution is required to satisfy the copula model. Therefore, we satisfy the best-fitted marginal distribution of air temperature and global solar radiation in this section. According to the histogram of these two variables as shown in Fig. 4, it is obvious that the general distribution families are not suitable. Therefore, we select the Gaussian mixture distribution (GMD) to fit the marginal distributions.

The Gaussian mixture model [13] is a probabilistic model for representing normally distributed subpopulations within an overall population. For a Gaussian mixture model with K

components, the k^{th} component has a mean of μ_k and variance of σ_k for the univariate case. The mixture component weights are defined as φ_k with the constraint that sum of φ_k equal 1 so that the total probability distribution normalizes to 1. The Gaussian mixture model can be expressed as:

$$p(x) = \sum_{k=1}^{K} \varphi_k N\left(x \middle| \mu_k, \sigma_k^{\ 2}\right) \tag{5}$$

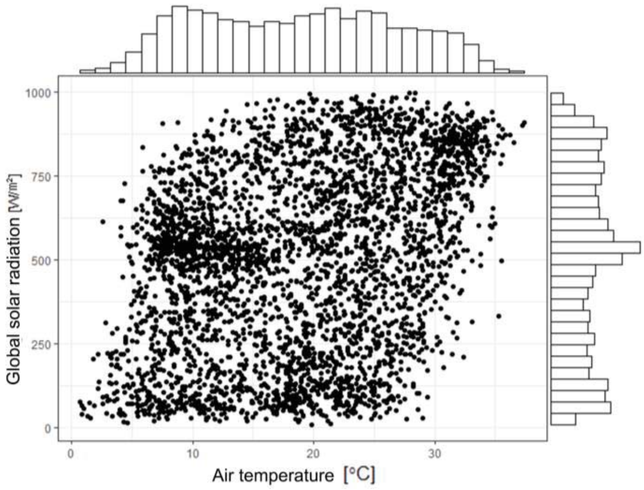


Fig. 4 Scatter plot of data

We also used the Maximum Likelihood Estimator method to estimate the parameters of GMD. The log-likelihood function is:

$$l(\theta) = \sum_{i=1}^{n} \log \sum_{k=1}^{K} \varphi_k N\left(x \middle| \mu_k, \sigma_k^{\ 2}\right) \tag{6}$$

where θ are the unknown parameters of the GMD.

In this study, we used five Gaussian mixture models to analyze the frequency of air temperature and global solar radiation. The components number of the Gaussian mixture models from 1 to 5 are denoted as N_K, K = 1, …, 5. Figs. 5 and 6 show the graphical diagnostics of the fitted result. It was shown that when the number of components is greater than three, the GMDs can fit the observed data well.

The Q−Q and P−P plots which show the comparison of characteristics of empirical and theoretical models also support this conclusion [14], [15]. The AIC and BIC values of each fitted model are shown in Table I. The 5 components Gaussian mixture models (N5) of global solar radiation and 4 components Gaussian mixture models (N4) of air temperature that show the lowest AIC and BIC values are selected as the best-fitted models. The parameters of the models are shown in Table II.

TABLE I
AIC AND BIC VALUES OF WEATHER ELEMENTS

Variables	GMD	Fitting Statistic Results	
		AIC	BIC
Air temperature	N	25684.98	25697.39
	N2	25208.78	25239.80
	N3	25109.00	25158.62
	N4	**25092.82**	**25161.05**
	N5	25096.22	25183.06
Solar radiation	N	51203.76	51216.17
	N2	50529.92	50560.94
	N3	50242.50	50292.12
	N4	50088.58	50156.81
	N5	**50056.50**	**50143.34**

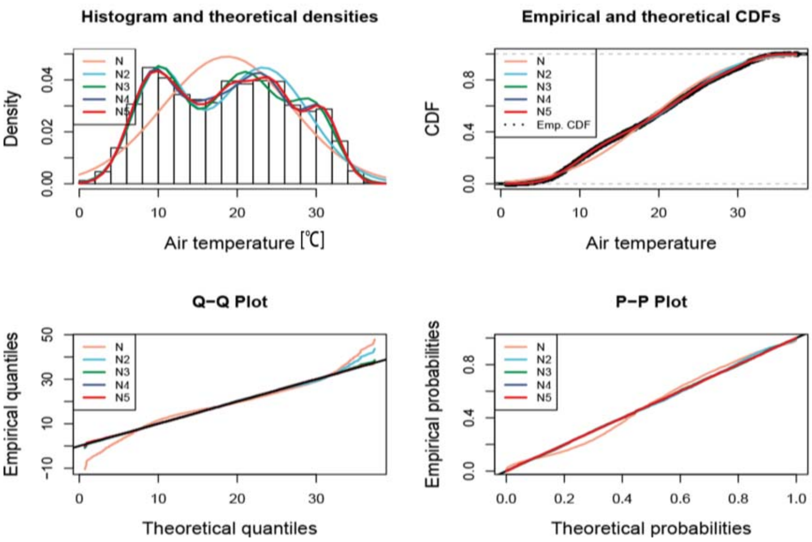


Fig. 5 Visualization of candidate distribution fitting on air temperature

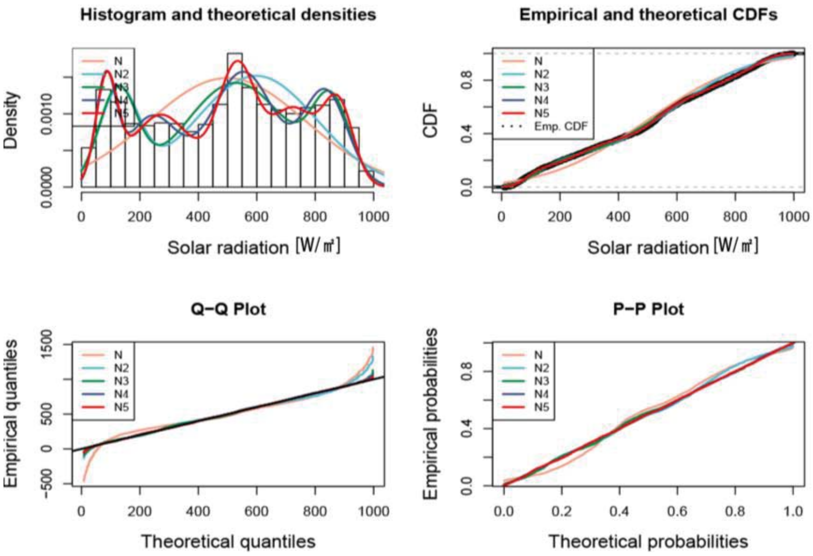


Fig. 6 Visualization of candidate distribution fitting on global solar radiation

TABLE II
PARAMETER OF SELECTED MODELS

	Air temperature(N4) [°C]			Global solar radiation(N5) [W/m²]		
i	μ_i	σ_i	φ_i	μ_i	σ_i	φ_i
1	9.19	2.87	0.29	86.10	36.42	0.13
2	16.28	3.46	0.23	272.54	103.85	0.26
3	23.43	3.36	0.33	532.55	65.51	0.27
4	30.70	2.29	0.15	725.98	79.55	0.21
5				879.68	51.71	0.14

TABLE III
DETAIL INFORMATION OF THE CANDIDATE COPULAS

Copula	Fitting Statistic Results		Parameter	Kendall's tau
	AIC	BIC		Empirical = 0.22
Gaussian	-496.32	-490.12	$\theta = 0.36$	0.23
Student t	-463.87	-451.46	$\theta = 0.35$ $df = 30$	0.23
Survival Clayton	-439.5	-433.29	$\theta = 0.45$	0.18
Gumbel	-409.79	-403.59	$\theta = 1.25$	0.20
Frank	-389.89	-383.69	$\theta = 1.98$	0.21
Joe	-346.5	-340.29	$\theta = 1.35$	0.16
Survival BB1	-466.12	-453.71	$\theta = 0.36$ $\delta = 1.06$	0.20
BB6	-407.67	-395.27	$\theta = 1.00$ $\delta = 1.25$	0.20
Survival BB7	-466.04	-453.64	$\theta = 1.08$ $\delta = 0.41$	0.20
BB8	-472.51	-460.1	$\theta = 1.9$ $\delta = 0.86$	0.21

B. Copula

The next step is to select the best fitted bivariate copula C for joint modelling of air temperature and global solar radiation. For this purpose, we considered various bivariate copula families for modelling the dependence structure between pairs. These bivariate copulas are Gaussian, Student's t, Gumbel, Frank, Joe, BB1 (Clayton–Gumbel), BB6 (Joe–Gumbel), BB7 (Joe–Clayton), BB8 (Joe–Frank), survival Clayton, survival BB1, survival BB6, survival BB7, and survival BB8. The R package VineCopula [16] was used for modelling bivariate copulas. Based on the results of the AIC and BIC values as shown in Table III, the best fitted copula for air temperature and global solar radiation is a Gaussian copula ($\theta = 0.36$).

C. Modeling Joint-Distribution

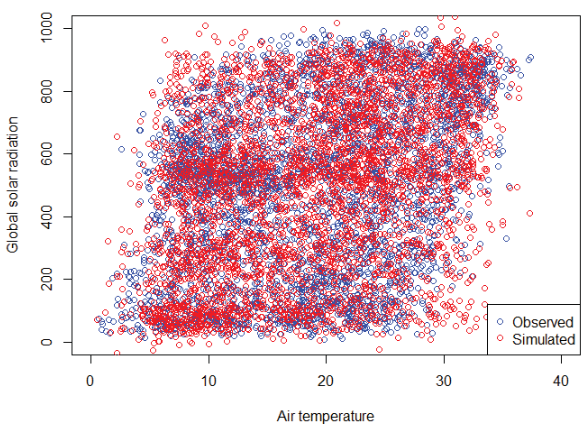
The Gaussian copula can be written as:

$$C_P^{Ga}\left(u_1, u_2\right) = \Phi_2\left(\Phi^{-1}\left(u_1\right), \Phi^{-1}\left(u_2\right)\right) \tag{7}$$

where the Φ_2 is the bivariate normal cumulative distribution function with linear correlation coefficient and Φ^{-1} is the inverse function of the standard normal cumulative distribution function. Moreover, u_1 and u_2 are two marginal distribution functions as shown in Table II.

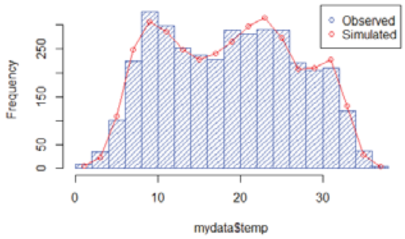
D. Evaluation of Model Performance

To evaluate the model performance, the same number of copula points were generated by a random number generator based on the selected Gaussian copula. These copula points were transformed to simulated weather data by the inverse probability integral transform method based on the selected Gaussian mixture distribution.



(a)

(b)



(c)

Fig. 7 Comparison of observed and simulated data: (a) scatter plot of data, (b) histogram plot of air temperature, (c) histogram plot of solar radiation

Fig. 7 shows the comparison of simulated and observed data. According to the histogram plot of air temperature and solar radiation, the marginal distributions of the simulated and observed data are almost identical. On the other hand, the scatter plot also shows a good fit result. Also, the Kendall's tau of the simulated data is 0.23, while it is 0.22 for the observed data. This indicates that the marginal distributions and dependence structure are well fitted.

IV. Result

Table IV shows the design air temperature and generated by TAC and global solar radiation simultaneous occurrence probability (SOP) methods in different exceeding probability. As shown in Table IV, the design air temperature and global solar radiation generated by the empirical SOP method are slightly smaller than those generated by theoretical SOP method. Nonetheless, the deviation can be ignored, and the results can still be considered consistent. The difference of global solar radiation between 0.4% empirical SOP and theoretical SOP is larger than the difference in other SOP. It seems that there is a need for further examination of the distribution of extreme global solar radiation. Furthermore, as shown in Table IV, for the same exceeding probability, the design air temperature generated by the SOP method is about 2.5 °C lower, and the global solar radiation generated is approximately 50 W/m^2 lower, while compared to TAC method. This difference of design conditions for air-conditioning design is not small. Moreover, the design air temperature and global solar radiation based on the TAC method in exceeding probability 5% are approximately almost equal to those based on the SOP method in exceeding probability 0.1%. That is to say, if we use the original TAC method with 1% excess frequency rate to select the outdoor design conditions, it may lead to over-designed air conditioning capacity.

TABLE IV
DESIGN WEATHER DATA BASED ON TAC AND SOP METHOD IN DIFFERENT EXCEEDING PROBABILITIES

Exceeding probability	Air temperature (°C)			Global solar radiation (W/m^2)		
	TAC method	Empirical SOP	Theoretical SOP	TAC method	Empirical SOP	Theoretical SOP
0.4%	35.0	32.1	33.11	975.00	911.11	928.51
1.0%	34.1	31.3	31.77	955.55	891.67	899.63
2.0%	33.2	30.6	30.61	938.89	869.44	873.03
2.5%	33.0	30.3	30.13	930.56	861.11	858.54
5.0%	31.8	28.8	28.33	902.78	825.00	811.54

V. Conclusion

In this study, we selected more realistic and suitable external design conditions based on the probability of simultaneous occurrence of air temperature and global solar radiation. For the purpose, we modeled the joint distribution of these two elements on 12-noon based on 10-year period data of Tokyo by copula approach. According to the comparison of 0.4%, 1.0%, 2.0%, 2.5% and 5% exceeding probability based on observed data and joint distribution, it is shown that the design conditions selected by original TAC method are needlessly stricter than the SOP method. It means that the SOP of air temperature and global solar radiation is less than their exceeding probability that may lead to over-design. Therefore, the results of this study suggest that using the simultaneous occurrence probabilities can generate more suitable outdoor design conditions.

References

[1] SHASE. *Calculation Method of Design Thermal Peak Load*, Tokyo: The Society of Heating, Air-Conditioning and Sanitary Engineers of Japan, 1989 (in Japanese).
[2] ASHRAE. *Handbook-2017 Fundamentals*, Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers, 2017.
[3] CIBSE. *Environmental design: CIBSE guide A*. London: Chartered Institution of Building Services E, 2006.
[4] Takeda H. Tokyo weather data for air-conditioning: Outdoor design conditions for heating and cooling loads by the TAC method. *Energy and Buildings* 1990; 15: 263-269.
[5] Yuan J, Emura K, Farnham C, et al. The creation of weather data for AC design using two weather indices for Osaka. *Energy and Buildings* 2017; 134: 248-258.
[6] Genest C and Favre AC. Everything You Always Wanted to Know About Copula Modeling but Were Afraid to Ask, *Journal of Hydrological Engineering* 2007; 12: 347–368.
[7] Nelsen RB. *An Introduction to Copulas*, New York: Springer, 2010.
[8] Gilchrist W. *Statistical Modelling with Quantile Functions*. Boca Raton: Chapman and Hall/CRC, 2019.
[9] Choroś B, Ibragimov R and Permiakova E. Copula Estimation. In: Jaworski P, Durante F, Härdle WK, et al.(eds) *Copula Theory and Its*

Applications. Berlin Heidelberg: Springer, 2010.

[10] Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. In: Parzen E, Tanabe K and Kitagawa G (eds) *Selected Papers of Hirotugu Akaike*. New York: Springer, 1998, pp.199-213.

[11] Huard D, Évin G and Favre AC. Bayesian copula selection. *Computational Statistics & Data Analysis* 2006; 51: 809-822.

[12] Akasaka H, et al. *Expanded AMeDAS Weather Data*. Tokyo: Architectural Institute of Japan/Maruzen, 2003.

[13] Everitt BS and Hand DJ. Mixtures of normal distributions. *Finite Mixture Distributions*. Dordrecht: Springer Netherlands, 1981, pp.25-57.

[14] Tsai DM and Yang CH. A quantile–quantile plot based pattern matching for defect detection. *Pattern Recognition Letters* 2005; 26: 1948-1962.

[15] Holmgren EB. The P-P Plot as a Method for Comparing Treatment Effects. *Journal of the American Statistical Association* 1995; 90: 360-365.

[16] Nagler T, Schepsmeier U, Stoeber J, et al. VineCopula: Statistical inference of vine copulas. R package version 2.3.0, 2019.