Advances in Economics, Business and Management Research (AEBMR), volume 37
International Conference on Transformations and Innovations in Management (ICTIM-17)

TAIL DEPENDENCE ANALYSIS OF EXTREME FOG AND HAZE WEATHER OCCURRING IN BEIJING AND TIANJIN BASED ON COPULA FUNCTION AND EVT

Yubo Ma

Business School, University of Shanghai for Science and Technology, P.R.China

Guangkuo Gao

Business School, University of Shanghai for Science and Technology, P.R.China

Abstract

In order to quantitatively verify that there is a strong correlation between the occurrence of extreme haze in Beijing and Tianjin. First, we estimated marginal distribution functions of the occurring of extreme fog and haze weather in two cities according to the POT model. Then, conditional probability formula of upper tail was derived through combining two marginal distribution functions and Copula function. Furthermore, specific probability was calculated through conditional probability formula of upper tail, while PM2.5 concentrations of two cities is no less than given concentrations. The results showed that when the average daily PM2.5 concentrations in adjacent places increased and reached a certain threshold, the probability of extreme fog and haze weather showed an increasing trend. In the last, this paper gave some policy suggestions about promoting the integration of the prevention and treatment of air pollution in an area and reaching the level of joint prevention and coordinate governance.

Key words: Copula Function; EVT; POT Model; Tail Dependence Analysis

JEL code: C650

1. Introduction

In recent years, fog and haze weather has been a normalization which badly endanger the mental and physical health of human beings. In addition, every walk of life has been effected in different extent by extreme fog and haze weather. This effect is stronger in those area where fog and haze happens at a high frequency such as Beijing and Tianjin. Under the circumstance of correspondingly steady pollutant source, there are some exterior factors effecting in fog and haze weather, such as temperature, wind direction, wind power, landform and so on. While Tobler's 'first law of geography' (1970) point that 'all attribute values on a geographic surface are related to each other, but closer values are more strongly related than are more distant ones.'

In order to prove the existence of correlation between the adjacent regions in extreme fog and haze weather. This paper selected Beijing and Tianjin's daily PM2.5 concentrations from Jan 1st, 2016 to Oct 21st, 2016 as observational data and made a scatter plots (see Fig. 1). The unit of concentrations is $\mu g/m^3$. It's clearly that the variation regularity of Beijing and Tianjin's PM2.5 concentrations show a strong correlation. On the other hand, the correlation between the adjacent regions in extreme fog and haze weather has an important reference value to transportation, health, agriculture and tourism of two cities. Hence, it's of great significance to research this correlation.

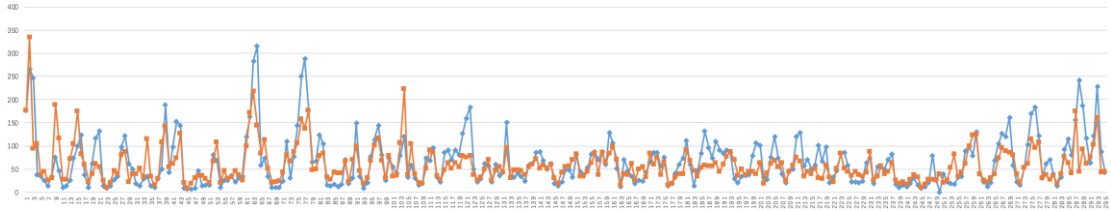

Figure 1. Scatter plots of variation of daily PM2.5 concentrations in 295 days, blue(Beijing), orange(Tianjin)

With the deepening attention of the pollution of fog and haze weather, relevant researches have been increasing. Further researches were carried out from the cause of formation, influence and solution by scholars. However, researches about the correlation of extreme fog and haze weather occurring in adjacent regions were rare. Lou Z. (2015) put forward that flowing air and adjacent regions form the synchronization of the pollution of fog and haze weather in this region. Ma L.M. et al. (2014) carried out a research about the correlation of fog and haze pollution in adjacent regions through measuring and calculating global Moran's I index. Masiol *et al.* (2012) did a comparative study about the diffusion of PM2.5 in the local and long-distance sources based on PAHs and wind direction data. And the consequence is that there existed a strong correlation in local area. In conclusion, domestic and overseas scholars researched spatial correlation of fog and haze in two ways. One is chemical process. The other is Moran's I index.

In contrast to general linear dependence, the correlation of extreme fog and haze weather occurring in two cities is particularly complex. Copula function can describe this complicated correlation between variances well. And tail dependence is always used to commendably depict the degree of correlation of variances when extreme issues happen. Meanwhile, POT model and Generalized Pareto distribution in extreme value theory are adopted to estimate the marginal distribution functions of happening of extreme fog and haze weather. Therefore, combining with copula function and EVT, this paper studied the tail dependence of extreme fog and haze weather occurring in two cities.

The primary air pollutants of Beijing and Tianjin is PM2.5. In order to facilitate our research, this paper only adopted PM2.5 concentrations to reflect the severity order of fog and haze pollution.

The rest of this paper is organized as follows. Section 2 presents a brief overview of POT model and copula function and derives the conditional probability formula of upper tail. Section 3

carries out an empirical research based on sample data. Section 4 provides the result of tail dependence analysis. Section 5 draws the conclusion and gives some advices combining with existing relevant policies and regulations.

2. Methodologies

2.1. POT model in EVT

For the sake of estimating the marginal distribution functions, this paper adopted POT (Peaks Over Threshold) model in EVT (Extreme Value Theory) and built a model based on observational data which exceeded threshold. The conditional threshold exceedance distribution function $F_U(y)$ is given by:

$$F_U(y) = P(x - u \le y | x > u) = \frac{F(u+y) - F(u)}{1 - F(u)} \quad (1)$$

Thus,

$$F(x) = F_U(y)\big(1 - F(u)\big) + F(u) \quad (2)$$

$$F(u) = \frac{n - N_u}{n} \quad (3)$$

$$F(x) = 1 - \frac{N_u}{n}(1 - F_U(y)) \quad (4)$$

where the distribution of x is $F(x)$. u is threshold and $y = x - u$ is exceedance threshold. n is observational sample capacity while N_u is the quantities of those observational sample which is greater than threshold.

Pickands-Balkema-de Haan Theorem (1974) indicated that, while u is large enough, $F_U(y)$ is approximately equal to GPD (Generalized Pareto distribution) $G_{\xi,\beta}(y)$:

$$F_U(y) \approx G_{\xi,\beta}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\beta} y\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - e^{-\frac{y}{\beta}}, & \xi = 0 \end{cases} \quad (5)$$

where ξ is the shape parameter and β is the scale parameter.

In order to estimate ξ, β and improve the accuracy of the model, the evaluation of u in POT model is extremely important. This paper estimated u by using the Hill plot and goodness of fit test.

After determining the best threshold u, MLE (maximum likelihood estimation) was adopted to estimate ξ and β. First of all, $g_{\xi,\beta}(y)$ is worked out, which is the probability density function of $G_{\xi,\beta}(y)$. Secondly, the logarithmic likelihood function of $g_{\xi,\beta}(y)$ is derived:

$$\ln L(\xi,\beta|y_i) = \begin{cases} -n\ln\beta - (1+\frac{1}{\xi})\sum_{i=1}^{n}\ln(1+\frac{\xi}{\beta}y_i), & \xi \neq 0 \\ -n\ln\beta - \frac{1}{\beta}\sum_{i=1}^{n} y_i, & \xi = 0 \end{cases} \quad (6)$$

Then, I calculated the maximum of Eq.(6) and obtained the estimated value of $\hat{\xi}$ and $\hat{\beta}$.

Substituting $\hat{\xi}$ and $\hat{\beta}$ into Eq.(4), the estimated marginal distribution function can be shown as:

$$\hat{F}(x) = \begin{cases} 1 - \frac{N_u}{n}(1+\frac{\hat{\xi}}{\hat{\beta}}(x-u))^{-\frac{1}{\hat{\xi}}}, & \xi \neq 0 \\ 1 - \frac{N_u}{n}e^{-\frac{x-u}{\hat{\beta}}}, & \xi = 0 \end{cases} \quad (7)$$

2.2. Copula function and tail dependence

Copula function can link different one-dimensional marginal distribution functions into multi-dimensional joint distribution function. And this procedure can be implemented through Sklar's Theorem. Two-dimensional Sklar's Theorem can be expressed as:

$$F_{XY}(x,y) = C_{XY}\big(F_X(x), F_Y(y)\big) \quad (8)$$

Combining with Two-dimensional Sklar's Theorem, conditional probability of upper tail can be given as:

$$P(X > x|Y > y) = \frac{P(X > x, Y > y)}{P(Y > y)} = \frac{1 - F_X(x) - F_Y(y) + C_{XY}\big(F_X(x), F_Y(y)\big)}{1 - F_Y(y)} \quad (9)$$

Conditional probability of upper tail was applied to describe the correlation of two cities suffering from extreme fog and haze weather. It can be interpreted as the probability of the occurrence of extreme fog and haze weather in X when Y is also suffered from extreme fog and haze weather. It's not hard to see from Eq.(9), this conditional probability can be constructed by two marginal distribution functions $F_X(x)$, $F_Y(y)$ and two-dimensional Copula function.

Two marginal distribution functions can be derived through EVT. And Copula function which was used to 'link' those two marginal distribution functions can be confirmed through parameter estimation and goodness of fit test.

Common Copula functions include Elliptic Copula and Archimedean Copula. This paper adopted three Copula functions in Archimedean Copula family including Gumbel Copula, Clayton Copula and Frank Copula.

Table 1. Three common Copula functions in Archimedean Copula family.

Family	Function
Gumbel Copula	$C(u,v) = \exp\{-\left[(-\ln u)^{\theta} + (-\ln v)^{\theta}\right]^{1/\theta}\}$
Clayton Copula	$C(u,v) = \left[u^{-\theta} + v^{-\theta} - 1\right]^{-1/\theta}$
Frank Copula	$C(u,v) = -\theta^{-1}\ln[1 + (e^{-\theta u} - 1)(e^{-\theta v} - 1)/(e^{-\theta} - 1)]$

In order to further calculate the value of θ, we need to estimate unknown parameter. Common parameter estimations in Copula function include maximum likelihood estimation, two-stage least squares estimation, non-parameter estimation and so on. This paper used maximum likelihood estimation to estimate θ. Then K-S test was used to test three Copula functions and choose a Copula function which is best fitted.

Finally, substituting two estimated marginal distribution functions and estimated Copula function into Eq.(9), then we can derive the conditional probability formula of upper tail.

3. Empirical research

3.1. Estimation of two marginal distribution functions

Figure. 2 are the empirical distribution functions of Beijing and Tianjin based on sample data. This two scatter plots are individually approximate to a line. So this two threshold exceedance samples are approximately subject to Generalized Pareto distribution.

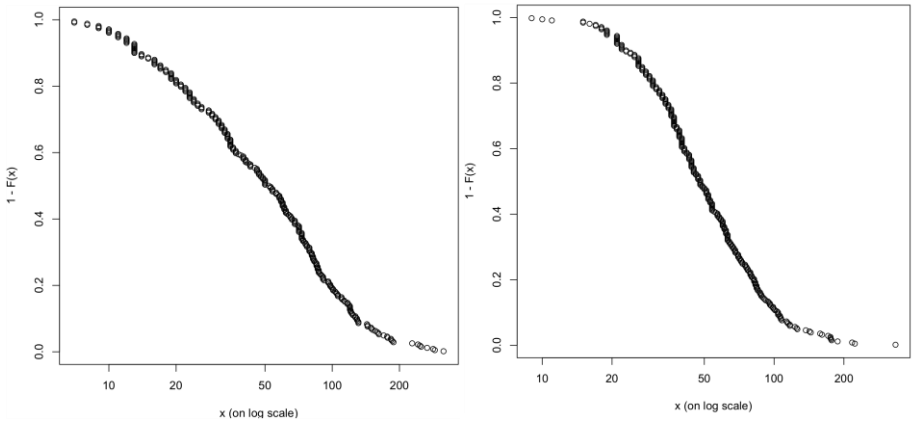

Figure 2. Empirical distribution functions

Left side(Beijing), right side(Tianjin)

Threshold u and unknown parameter ξ and β should be estimated with the purpose of deriving the marginal distribution functions. This paper used the Hill plot to tentatively determine the range of u. And then K-S test was adopted to select the best threshold.

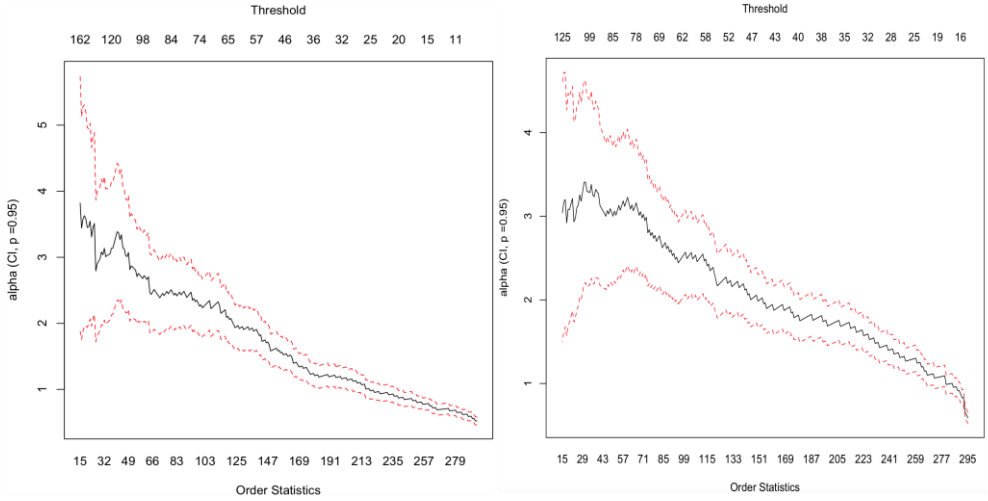

Figure 3. Hill plot

Left side(Beijing), right side(Tianjin)

Through this two hill plot, we can see that the line of Beijing is asymptotically stable when the threshold is 120, while the line of Tianjin is asymptotically stable when the threshold is 112. Therefore, we chose three thresholds of Beijing $u_1 = 118, u_2 = 119, u_3 = 120$ and three thresholds of Tianjin $u_1' = 111, u_2' = 112, u_3' = 113$ to compare. Tables 2 and 3 are the comparisons of threshold through K-S test.

Table 2. Comparison of threshold (Beijing)

Threshold	Exceedance	$\hat{\xi}$	$\hat{\beta}$	P-value
$u_1 = 118$	42	0.1935342	40.1452663	0.2828
$u_2 = 119$	41	0.1980454	40.0941415	0.2831
$u_3 = 120$	37	0.00541327	53.75815094	0.2845

Table 3. Comparison of threshold (Tianjin)

Threshold	Exceedance	$\widehat{\xi'}$	$\widehat{\beta'}$	P-value
$u_1' = 111$	22	0.006970807	49.026711796	0.2943
$u_2' = 112$	22	0.03110596	46.86615502	0.2943
$u_3' = 113$	21	0.002626109	49.494326869	0.2955

From Table 2, we know that the best threshold of Beijing is $u = 119$, and the corresponding parameters of GPD function are $\hat{\xi} = 0.1980454$, $\hat{\beta} = 40.0941415$.

Thus, when $x > 119$, the estimated marginal distribution function of Beijing is:

$$\hat{F}(x) = 1 - 0.13898305\left(1 + \frac{0.1980454}{40.0941415}(x - 119)\right)^{-\frac{1}{0.1980454}} \quad (10)$$

Similarly, the best threshold of Tianjin is $u' = 113$, and the corresponding parameters of GPD function are $\widehat{\xi'} = 0.002626109$, $\widehat{\beta'} = 49.494326869$.

Hence, when $y > 113$, the estimated marginal distribution function of Tianjin is:

$$\hat{F}(y) = 1 - 0.07118644\left(1 + \frac{0.002626109}{49.494326869}(y - 113)\right)^{-\frac{1}{0.002626109}} \quad (11)$$

Finally, we made two fitting plot based on each estimated parameters. Fig.4 show the goodness of fit.

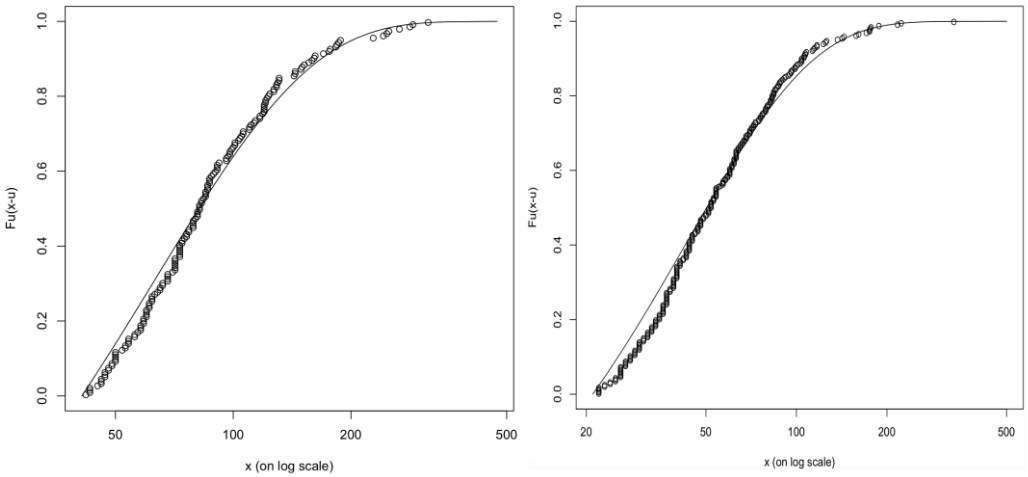

Figure 4. Goodness of fit

Left side(Beijing), right side(Tianjin)

3.2. Estimation of Copula function

In order to find the best fitted Copula function from Gumbel Copula, Clayton Copula and Frank Copula, we need to estimate the parameter θ. This paper adopted maximum likelihood estimation method to calculate θ. The consequence was listed in Table 4.

Table 4. Estimated θ in three Copula functions.

Copula function	Gumbel Copula	Clayton Copula	Frank Copula
Estimated θ	2.2818	1.9518	7.7463

After testing these three Copula functions by using K-S test, we got Gumbel Copula which is the best fitted function. Combining Gumbel Copula with two marginal distribution functions, Copula function can be given by:

$$\hat{C}_{XY}\left(\hat{F}_X(x), \hat{F}_Y(y)\right) = \exp\left\{-\left[\left(-\ln\left(\hat{F}_X(x)\right)\right)^{2.2818} + \left(-\ln\left(\hat{F}_Y(y)\right)\right)^{2.2818}\right]^{\frac{1}{2.2818}}\right\} \quad (12)$$

Substituting Eqs.(10), (11) and (12) into Eq.(9), conditional probability formula of upper tail can be written as:

$$P(X > x|Y > y) = \frac{1 - \hat{F}_X(x) - \hat{F}_Y(y) + \hat{C}_{XY}\left(\hat{F}_X(x), \hat{F}_Y(y)\right)}{1 - \hat{F}_Y(y)} \quad (13)$$

4. Results

After deriving estimated conditional probability formula of upper tail, this paper chose several groups of concentrations which are above the threshold to calculate the probability. And calculated conditional probability of upper tail in particular cases are listed in Table 5.

Table 5. Conditional probability of upper tail in particular cases.

(x, y)	$\hat{C}_{XY}\left(\hat{F}_X(x), \hat{F}_Y(y)\right)$	$P(X > x \mid Y > y)$
(120,120)	0.8564164389744	0.8702948970176
(150,150)	0.9269929057687	0.8423875696751
(200,200)	0.9726179840288	0.8397465266343
(250,250)	0.9882235341748	0.8693523096330
(300,300)	0.9943236578285	0.9085941386377
(500,500)	0.9993362300675	0.9913843170766

Notes: $P(X > x \mid Y > y)$ represents the conditional probability that when the daily PM2.5 concentrations of Tianjin exceeds $y\ \mu g/m^3$, the daily PM2.5 concentrations of Beijing also exceeds $x\ \mu g/m^3$.

It's clear to see from Table 5, the probability of Beijing's daily PM2.5 concentrations exceeding $120\mu g/m^3$ is 87.0294% when the daily PM2.5 concentrations of Tianjin exceeds $120\mu g/m^3$. Similarly, when the daily PM2.5 concentrations of Tianjin respectively exceeds $150\mu g/m^3$, $200\mu g/m^3$, $250\mu g/m^3$, $300\mu g/m^3$ and $500\mu g/m^3$, the conditional probabilities of Beijing's daily PM2.5 concentrations exceeding $150\mu g/m^3$, $200\mu g/m^3$, $250\mu g/m^3$, $300\mu g/m^3$ and $500\mu g/m^3$ are 84.2388%, 83.9747%, 86.9352%, 90.8594% and 99.1384%.

On the basis of *People's Republic of China ambient air quality standards*, ambient air is slightly polluted when daily average PM2.5 concentrations exceeds $75\mu g/m^3$. Daily average PM2.5 concentrations exceeding $115\ \mu g/m^3$ belongs to moderate pollution, while exceeding $200\ \mu g/m^3$ belongs to severe pollution. The probability of Beijing's daily PM2.5 concentrations exceeding $200\mu g/m^3$ is 83.9747% when the daily PM2.5 concentrations of Tianjin exceeds $200\mu g/m^3$. Besides, the higher daily PM2.5 concentrations are, the stronger correlation of two cities suffering from extreme fog and haze weather appears to. Thus, it can be seen that the correlation of extreme fog and haze weather occurring in two cities reaches a very strong degree. One city's daily PM2.5 concentrations has great impacts on another.

5. Conclusions

According to *law on the prevention of air pollution of People's Republic of China* latest revised version Chapter Ⅴ Article 86, 'The State establishes joint prevention and control mechanism of regional air pollution, makes overall plans and coordinates the prevention works of regional air pollution. The administrative department of environmental protection under the State Council shall delimit national key regions of air pollution prevention and be subject to the approval of the State Council on the basis of major function oriented zoning, the regional quality of the atmospheric environment and the regulation of atmospheric pollution transport and diffusion.', it shows that the State principally adopted joint prevention and control mechanism to prevent from atmospheric pollution. The State hopes to realize the integration of reginal atmospheric pollution prevention by joint monitoring and management in an area. Thus, it can effectively minimize the various impacts of extreme fog and haze weather in an area.

Combing with the results of empirical research (mentioned in Section 4), there exist high correlation between two cities suffering from extreme fog and haze weather. And this high correlation will make the impacts of adjacent region suffering from extreme fog and haze weather appear to a consistency. With the purpose of minimizing these impacts, we should promote the integration of the prevention and treatment of air pollution in an area. Meanwhile, joint prevention and coordinate governance should be taken into account. For the sake of promoting the integration of the prevention and treatment of air pollution in an area, this paper gives some advices:

- Establishing a supervision and management department in the key fog and haze pollution prevention area.

  This department should make overall plans and coordinate relevant departments based on the characteristics of their region. Practical joint prevention and control mechanism should be established. And it should effectively coordinate the communication of each other.

- Further improving the laws and regulations of atmospheric pollution prevention.

  Making practical laws and regulations is a must. We should break through the restraints of traditional administration system and realize cross-area cooperation to face the impacts brought from fog and haze pollution together.

- Setting up a perfect, unified and efficient big data information platform in an area.

  Each districts should upload real-time relevant pollutants concentrations, wind direction, wind power and other relevant data from every monitoring stations at every moment to the platform. Then, all these structural and non-structural data will be disposed and analyzed through the big data information platform. Finally, analytical consequences will be immediately fed back to each districts and relevant departments. Therefore, departments can adopt specific measure to govern and prevent from atmospheric pollution.

Acknowledgement

This work was supported by National Social Science Fund of China (Grant No. 15BTJ017), Climb Plan Project of Humanities and Social Science of USST (Grant No. SK17PA01) and Municipal and University's Training Programs of Innovation and Entrepreneurship for Undergraduates (Grant No. SH2017064、XJ2017104).

References

A.A. Balkema, L. De Haan. (1974) Residual life time at great age. *The Annals of Probability*, vol.2, no.5, pp.792-804.

Lou, Z.Y. (2015) The intergovernmental cooperation on tackling haze in the Beijing-Tianjin-Hebei region. *Huazhong university of Science and Technology*.

Ma, L.M., Zhang, X. (2014) The spatial effect of China's haze pollution and the impact from economic change and energy structure. *China industrial economics,* no.4, pp.19-31.

Masiol, M., et al. (2012) GC-MS analyses and chemometric processing to discriminate the local and long-distance sources of PAHs associated to atmospheric PM2.5. *Environmental Science and Pollution Research International*, vol.19, no.8, pp.3142-3151.

Song, S.B., *et al.* (2012) Copulas function and its applications in hydrology. Beijing: Science Press.

Teng, S.W., Ran, G.H. (2011) Estimation on VAR of agricultural calamity based on POT-GPD model of loss distribution. *Statistical research*, vol.28, no.7, pp.79-83.

Tobler, W. (1970) A computer movie simulating urban growth in the detroit region. *Economic Geography*, vol.46, no.2, pp.234-240.

Wang, L.F. (2012) Copula distribution estimation algorithm. Beijing: China Machine Press.

Wu, J.H., Wang, X.J., Zhang, Y. (2014). The choice of the Copula function in the correlation analysis. *Statistical research*, vol.31, no.10, pp.99-107.

Zhang, L.Z., Hu, X. (2014) Comparison of parametric and semiparametric estimation methods for Copula. *Statistical research*, vol.31, no.2, pp.91-95.

GB 3095-2012 People's Republic of China ambient air quality standards.

Ministry of environmental protection of the People's Republic of China. *Law on the prevention of air pollution of People's Republic of China*.2015-08-29.