Application Range of Polynomial Translation Method and Its Modification of Estimation of Annual Maximum Wind Speeds

Jun Kanda and **Yuan-Lung Lo**
Graduate School of Frontier Science, the University of Tokyo

ABSTRACT:

Polynomial translation method for translating Gaussian random variables to non-Gaussian random variables has been proposed for more than half centuries. Application of this method has also been utilized to many aspects. However, the approximation of parameters of the polynomial form cannot provide good estimation when random variables are given large skewness and kurtosis. In this research, the approximation method used for the parameter estimation is investigated for wider ranges of given skewness and kurtosis, which are observed from the 10 minute mean wind speed samples of 155 sites from 1961 to 2002. The proposed simulation process of the estimation of annual maximum wind speeds is then modified based on the investigation results. For those extreme cases with large observed skewness or kurtosis, the modified simulation process can provide fairly good estimation results as normal cases.

1. INTRODUCTION

Edgeworth proposed the 3rd order polynomial form for the non-Gaussian random variables and the method of moment has been used for the parameter estimation of the polynomial form. Choi and Kanda (2003) discussed the development history of NST and also conducted the practicability of an approximation sheet used for the parameter estimation. However, the given skewness and kurtosis were only limited to small values. Once the given skewness and kurtosis are larger than certain ranges, the calculated skewness and kurtosis from the translated non-Gaussian random variables do not fit to the given ones. The approximation sheet mentioned by Choi and Kanda (2003) may not provide a wide range of practical application.

In this research, the ranges of the given skewness and kurtosis are further examined based on a testing procedure. The simulation process is then further modified based on the examination results. Data from 155 meteorological sites are utilized to the simulation process for the estimation of annual maximum wind speeds and the effect of the modification.

2. TEST OF APPLICATION RANGE OF POLYNOMIAL TRANSLATION METHOD

A non-Gaussian random variable, Y, which is given a set of four moments, is written in a polynomial form with respect to a standard Gaussian random variable, X as

$$\mathbf{Y} = \mathbf{a} + \mathbf{bX} + \mathbf{cX}^2 + \mathbf{dX}^3 \tag{1}$$

The coefficients of the polynomial form can be obtained from the following nonlinear equations,

$$\mathbf{E(Y)} = \mathbf{a} + \mathbf{c} = 0 \tag{2}$$

$$\mathbf{Var(Y)} = \mathbf{b}^2 + 6\mathbf{bd} + 2\mathbf{c}^2 + 15\mathbf{d}^2 = 1 \tag{3}$$

$$\gamma_1(\mathbf{Y}) = 2\mathbf{c}\left(\mathbf{b}^2 + 24\mathbf{bd} + 105\mathbf{d}^2 + 2\right) \tag{4}$$

$$\gamma_2(\mathbf{Y}) = 24\left\{\mathbf{bd} + \mathbf{c}^2\left(1 + \mathbf{b}^2 + 28\mathbf{bd}\right) + \mathbf{d}^2\left(12 + 48\mathbf{bd} + 141\mathbf{c}^2 + 225\mathbf{d}^2\right)\right\} \tag{5}$$

where γ_1 and γ_2 are given skewness and unbiased kurtosis. To solve these four nonlinear equations, an algorithm like least square method is needed. Table 1 was then proposed to provide the approximation of the parameters, a, b, c, d, in equation (1) ~ (5). By using this approximation sheet, Choi and Kanda (2003) showed good agreements of the approximation results.

Table 1 Coefficients of the polynomial form: $Y = a + bX + cX^2 + dX^3$

j	T_j	b_j		c_j		d_j	
		γ_2<1.5	γ_2>=1.5	γ_2<1.5	γ_2>=1.5	γ_2<1.5	γ_2>=1.5
1	1	1.0000	0.9698	0.0000	0.0012	0.0000	0.0112
2	γ_1	-0.0014	-0.0305	0.1668	0.1566	0.0007	0.0129
3	γ_2	-0.1238	-0.0765	0.0000	-0.0009	0.0412	0.0236
4	γ_1^2	0.1224	0.0558	0.0019	-0.0024	-0.0469	-0.0177
5	γ_2^2	0.0353	0.0054	0.0000	0.0002	-0.0131	-0.0018
6	γ_1^3	-0.0491	-0.0348	0.0653	0.0466	0.0258	0.0216
7	γ_2^3	-0.0085	-0.0002	0.0001	0.0000	0.0033	0.0001
8	$\gamma_1\gamma_2$	0.0027	0.0181	-0.0397	-0.0155	-0.0009	-0.0061
9	$\gamma_1^2\gamma_2$	-0.0768	-0.0130	0.0178	0.0236	0.0314	0.0087
10	$\gamma_1\gamma_2^2$	-0.0075	-0.0041	0.0183	0.0026	0.0021	0.0016
11	$\gamma_1^3\gamma_2$	0.0134	0.0029	-0.0068	0.0023	-0.0108	-0.0009
12	$\gamma_1\gamma_2^3$	0.0007	0.0003	-0.0018	-0.0002	0.0010	-0.0001
13	$\gamma_1^2\gamma_2^2$	-0.0101	-0.0002	-0.0071	-0.0029	0.0018	-0.0005
14	$\gamma_1^2\gamma_2^3$	0.0103	0.0001	0.0136	0.0001	0.0002	0.0000
15	$\gamma_1^3\gamma_2^2$	-0.0322	-0.0002	-0.0167	-0.0011	0.0165	-0.0001
16	$\gamma_1^3\gamma_2^3$	0.0127	0.0000	0.0207	0.0000	-0.0033	0.0000

Parameters: $\mathbf{b} = \sum_{j=1}^{16} \mathbf{T_j b_j}$ $\mathbf{c} = \sum_{j=1}^{16} \mathbf{T_j c_j}$ $\mathbf{d} = \sum_{j=1}^{16} \mathbf{T_j d_j}$

However, the ranges of the given skewness and the kurtosis of the observed data are from 0.0643 to 2.9687 and from -0.5239 to 21.6440 respectively, which are over the range examined by Choi and Kanda (2003). Therefore, to make sure the applicability of this approximation sheet, a testing flow is suggested as Figure 1.

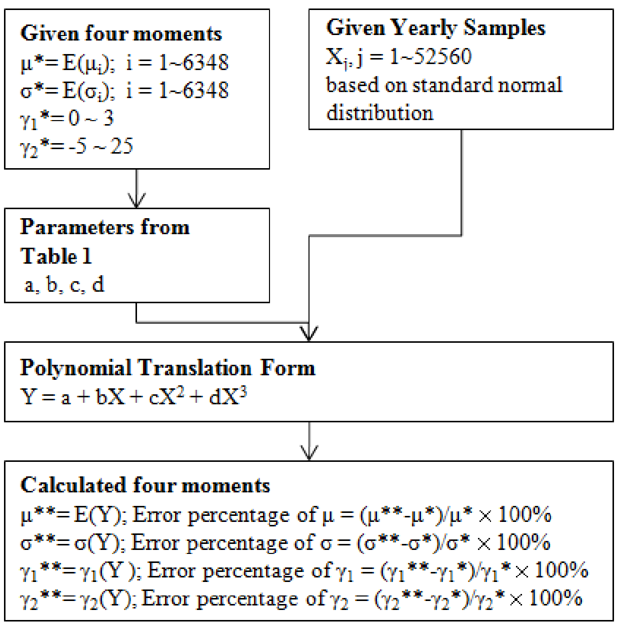


Figure 1 Testing flow of the parameter estimation based on Table 1

For the given four moments in Figure 1, given mean and standard deviation values are determined by the mean of all mean and standard deviation values at 155 sites from 1961 to 2002. The given mean value is fixed at 3.0960 and the given standard deviation value is fixed at2.0405. Given skewness varies from 0 to 3 with interval equals 0.01 and given kurtosis varies from -5 to 25 with interval equals 0.1. The number of total combinations of given skewness and kurtosis tested is 90,601 (301×301).

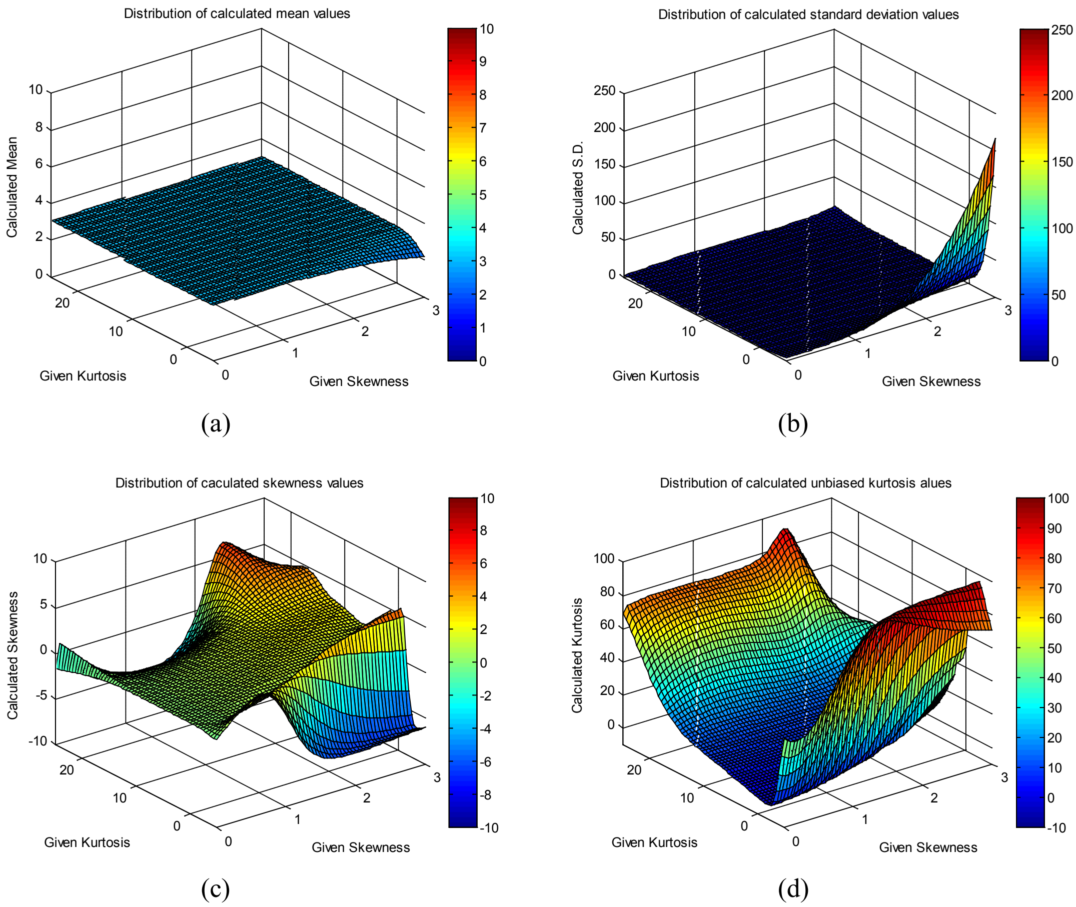


Figure 2 Distributions of calculated four moments based on Figure 1: (a) calculated mean; (b) calculated standard deviation; (c) calculated skewness; (d) calculated unbiased kurtosis.

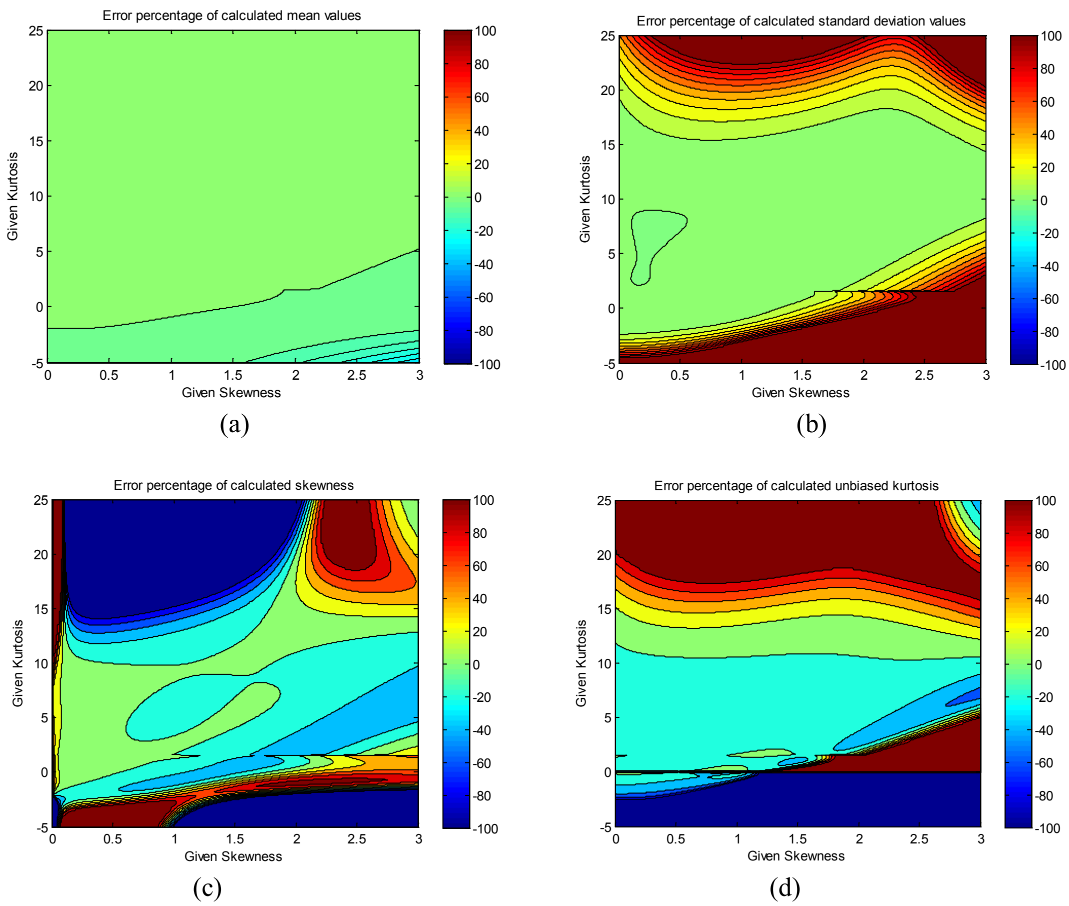


Figure 3 Contour of error percentages of calculated four moments: (a) calculated mean; (b) calculated standard deviation; (c) calculated skewness; (d) calculated unbiased kurtosis.

To visually inspect the agreements of these 90,601 combinations of given four moments, a skewness-kurtosis plane is demonstrated. Figure 2 shows the distributions of the calculated four moments on the given skewness-kurtosis plane. Figure 3 shows the contours of error percentages of calculated four moments. In Figure 3, when the error is more than 100% or less than -100%, the error is assigned 100% or -100% at most.

Generally speaking, two kinds of combinations can be observed to have significant error estimation of parameters: (1) large given skewness with negative given unbiased kurtosis; (2) large given unbiased kurtosis. For those given moment values near zero, error percentage is also significant because of the definition of error percentage calculation. Combining the observation results from Figure 3, a criterion of error percentage is needed when this approximation sheet is used for parameter estimation. Figure 4 shows observed skewness and kurtosis on the four contours corresponding to four criteria, 5%, 10%, 15%, and 20%. Within the red range of the contour, the calculated four moments are considered acceptable for the parameter estimation of polynomial form. Among these four contours, 20% criterion is used for the further investigation in this research.

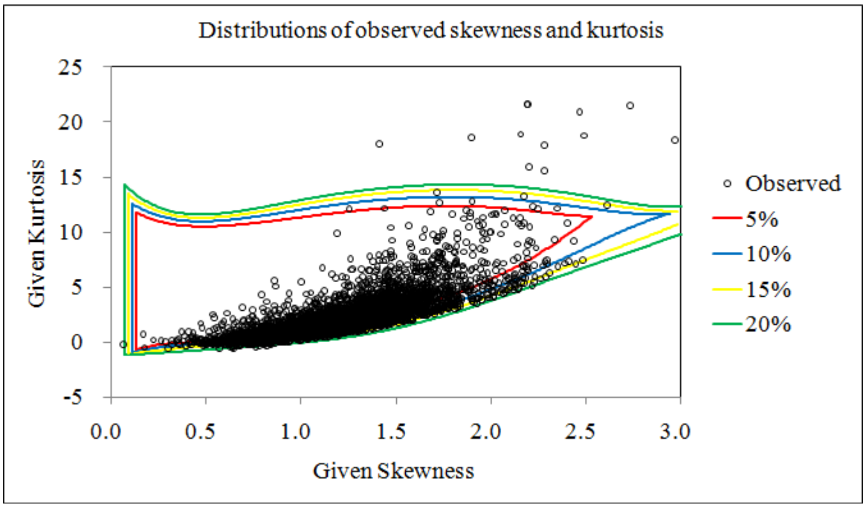


Figure 4 Different application ranges of calculated four moments based on different criteria:

3. BASIC CHARATERISTICS OF YEARLY AND REGIONALLY VARIATIONS OF FOUR MOMENTS

Basic characteristics of four moments observed from the 155 sites were examined by Kanda and Lo (2009) yearly and regionally. A linear relation between skewness and kurtosis is confirmed by calculating the correlation coefficients of 155 sites. The mean value of the 155 correlation coefficients is 0.9088 and the standard deviation is 0.0453, which indicates the high correlation exists in most sites (Table 2). Figure 5 shows the histograms of moment parameters of 155 sites. Mean and standard deviation of four moments are calculated to represent each site's characteristics. In some sites the standard deviation of four moments is close to zero which suggests the identical nature for parent distributions, but in most sites the standard deviation is significant so that the identical hypothesis would not be applied.

Probability distributions of four moments for 155 sites are also plotted. Figure 6 shows two examples of the probability distributions of observed skewness and kurtosis. Among 155 sites, 106 sites tend to have lognormal distribution for skewness while 108 sites for kurtosis.

Table 2 The mean and the standard deviation of 155 correlation coefficients between four moments

	μ-σ	μ-γ_1	μ-γ_2	σ-γ_1	σ-γ_2	γ_1-γ_2
mean of 155 coefficients	0.5894	-0.0853	-0.0581	0.1681	0.0846	0.9088
S.D. of 155 coefficients	0.3328	0.3007	0.2598	0.3164	0.2748	0.0453

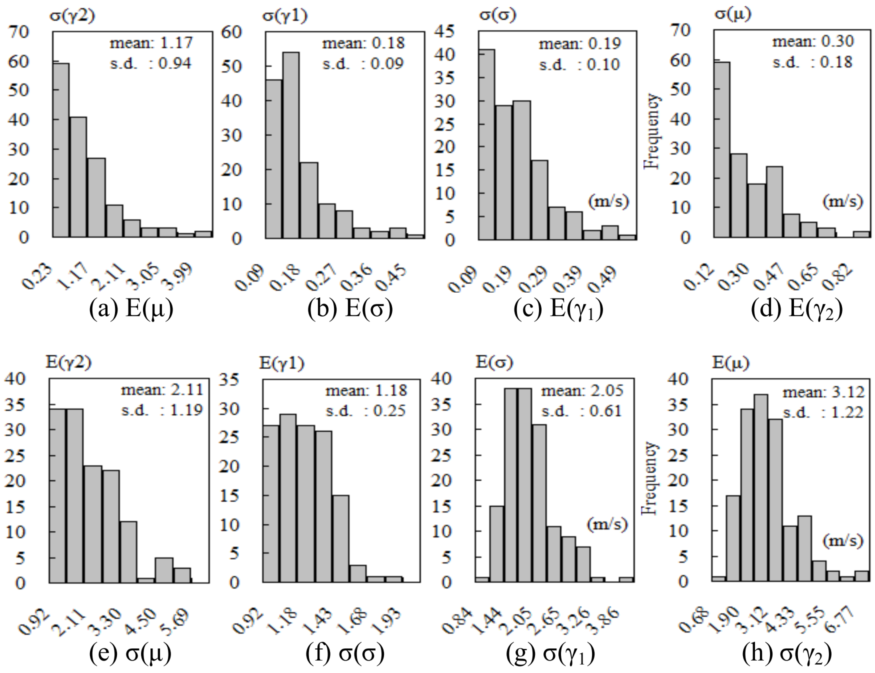


Figure 5 Histograms of moment parameters of 155 meteorological sites

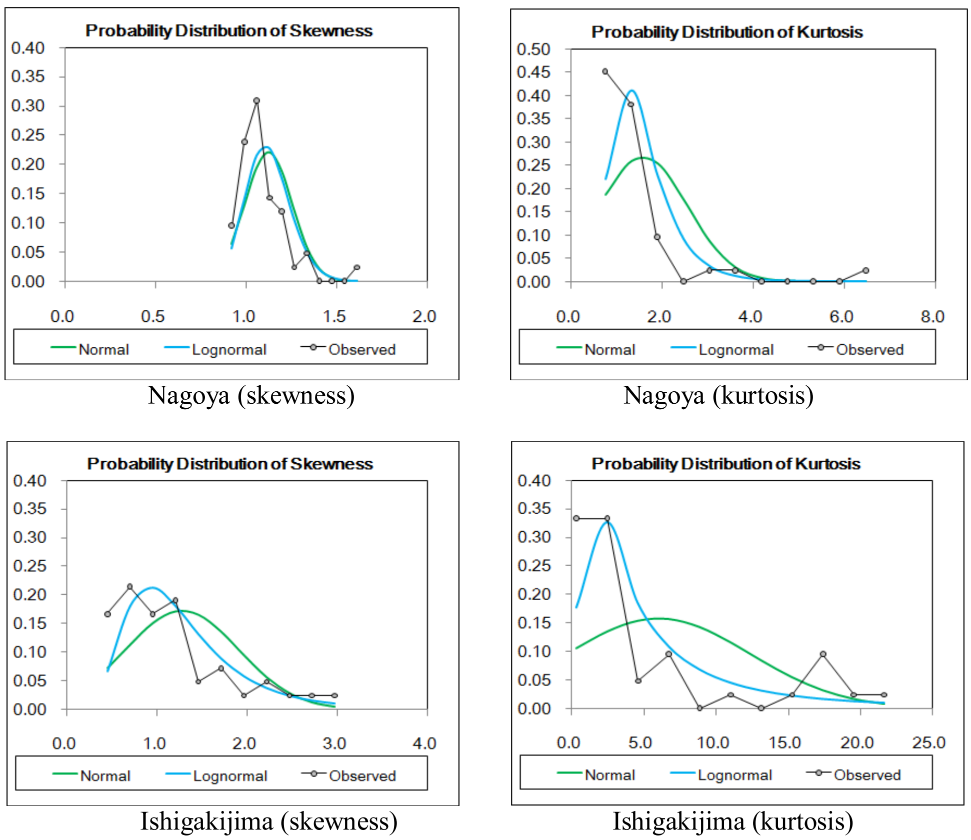


Figure 6 Probability distributions of skewness and kurtosis

4. SIMULATION PROCESS AND ITS MODIFICATION

The simulation process for the estimation of annual maximum wind speeds proposed by Kanda and Lo (2009) is applied to 155 sites. Figure 7 shows the simulation process.

Basic characteristics of four moments mentioned in the previous section are concerned in the simulation process. To test the agreement of estimation results, many conditions are considered in the simulation process. For the generation of given skewness and kurtosis in the second step, the given skewness and kurtosis can be independent or fully-correlated. Further, the assumption of the probability distributions of given skewness and kurtosis can be normal or lognormal.

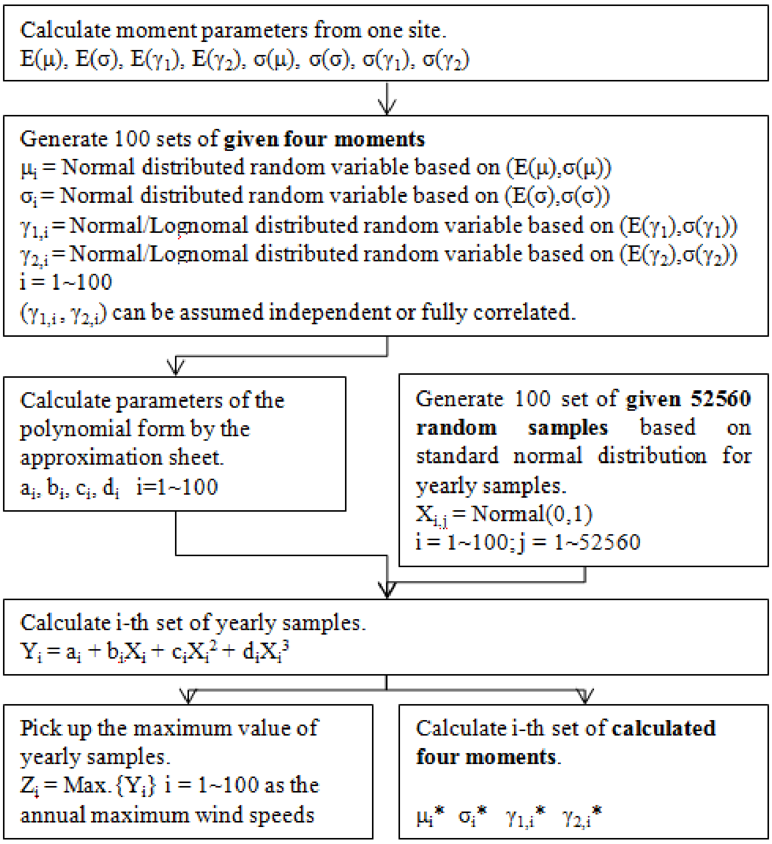


Figure 7 Simulation process of estimation of annual maximum wind speeds

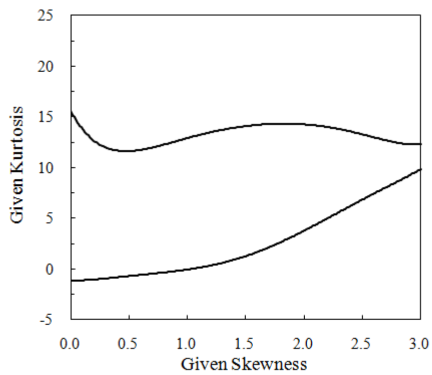


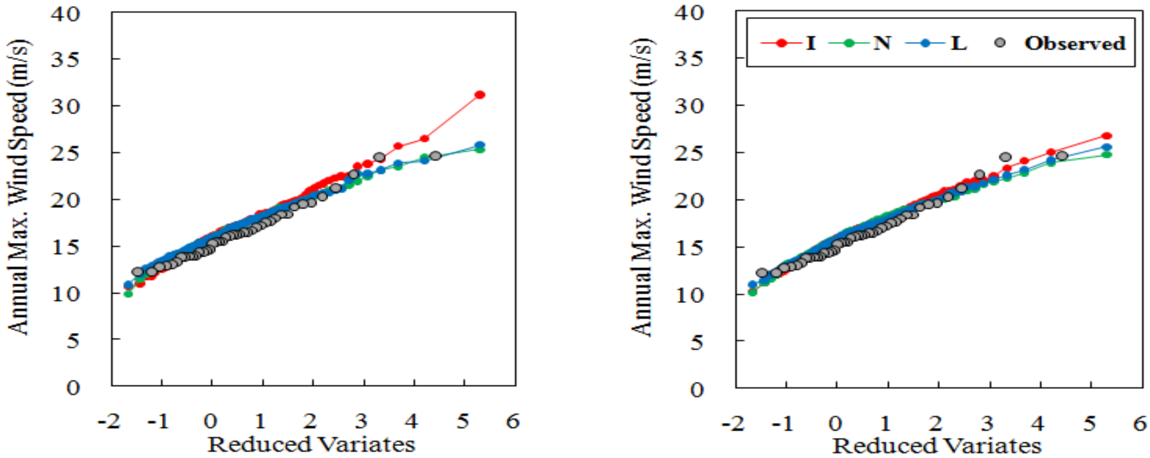
Figure 8 Upper and lower truncation boundaries

Since Figure 4 shows that there are many observed skewness and kurtosis outside the application range with 20% criterion, generation of given skewness and kurtosis in some sites may lead to significant error in estimation results. Therefore, modification of the extremely large skewness and kurtosis should be considered and added to the simulation process. Figure 8 shows the upper and lower truncation boundaries of 20% criterion. When the given kurtosis is generated, the value is checked to be inside the application range. Any generated given kurtosis outside the range is neglected in the simulation. For a more strict limitation, other criteria can be used. However, 20% is preliminarily taken as a trial in this research.

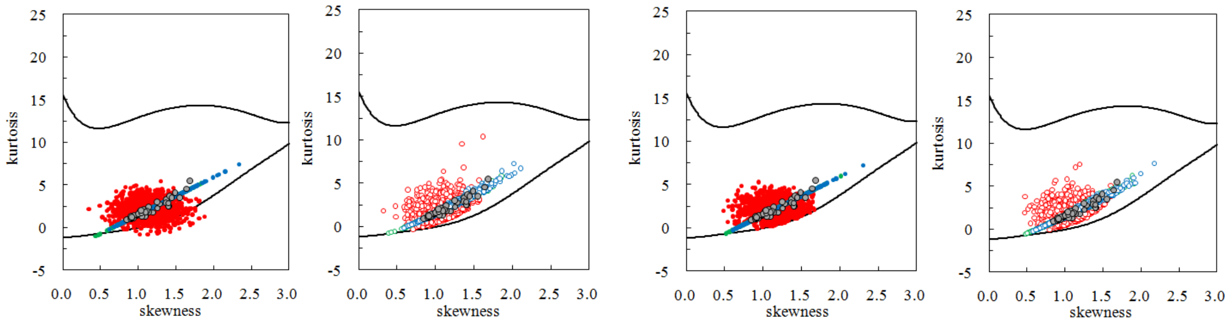
Simulation process shown in Figure 7 is repeated for 11 times and then the median values at every reduced variate are picked up as the best estimation results. Given skewness and kurtosis generated based on different assumptions are plotted on the 20% application range for 11 times. Calculated skewness and kurtosis from the simulated non-Gaussian yearly samples are also plotted for 11 times.

5. ESTIMATION RESULTS OF SEVRAL METEORILOGICAL SITES

Several of 155 sites are illustrated in Figure 9 ~ 11. Among these estimation results, Tokyo is considered as a normal case since there is no observed skewness and kurtosis outside the 20% application range. Other two sites, Kagoshima and Ishigakijima, are considered as extreme cases for lots of their observed skewness and kurtosis values are outside the range.



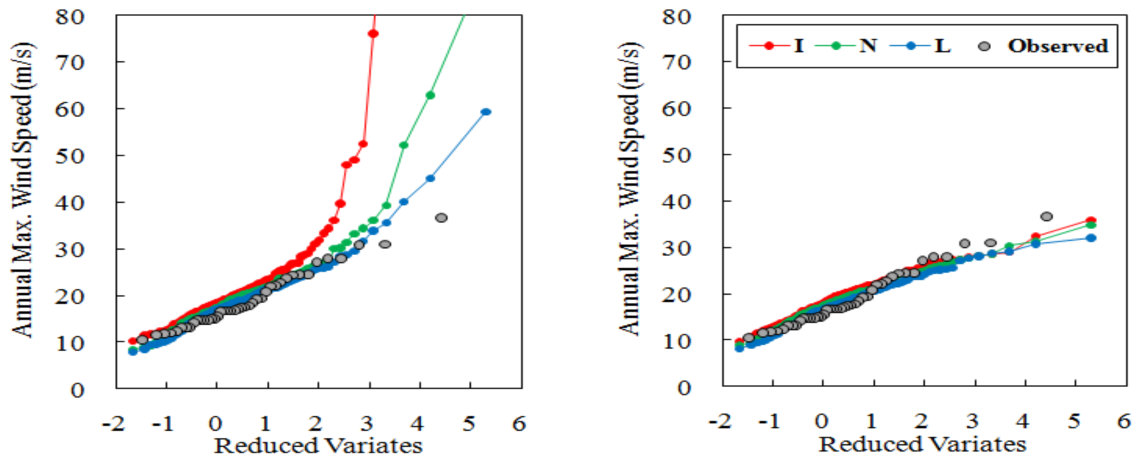
(a) Estimation results (not modified simulation) (b) Estimation results (modified simulation)



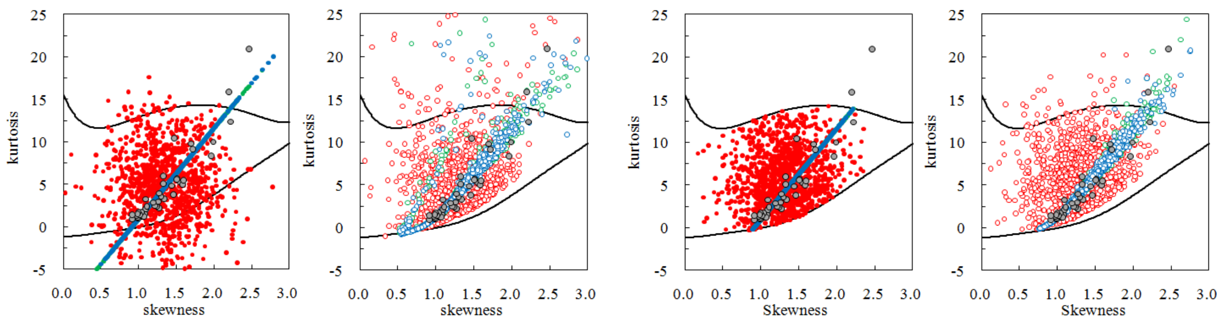
(c)-1 Given γ_1 and γ_2 (c)-2 Calculated γ_1 and γ_2
(not modified simulation)

(d)-1 Given γ_1 and γ_2 (d)-2 Calculated γ_1 and γ_2
(modified simulation)

Figure 9 Estimation results of Tokyo



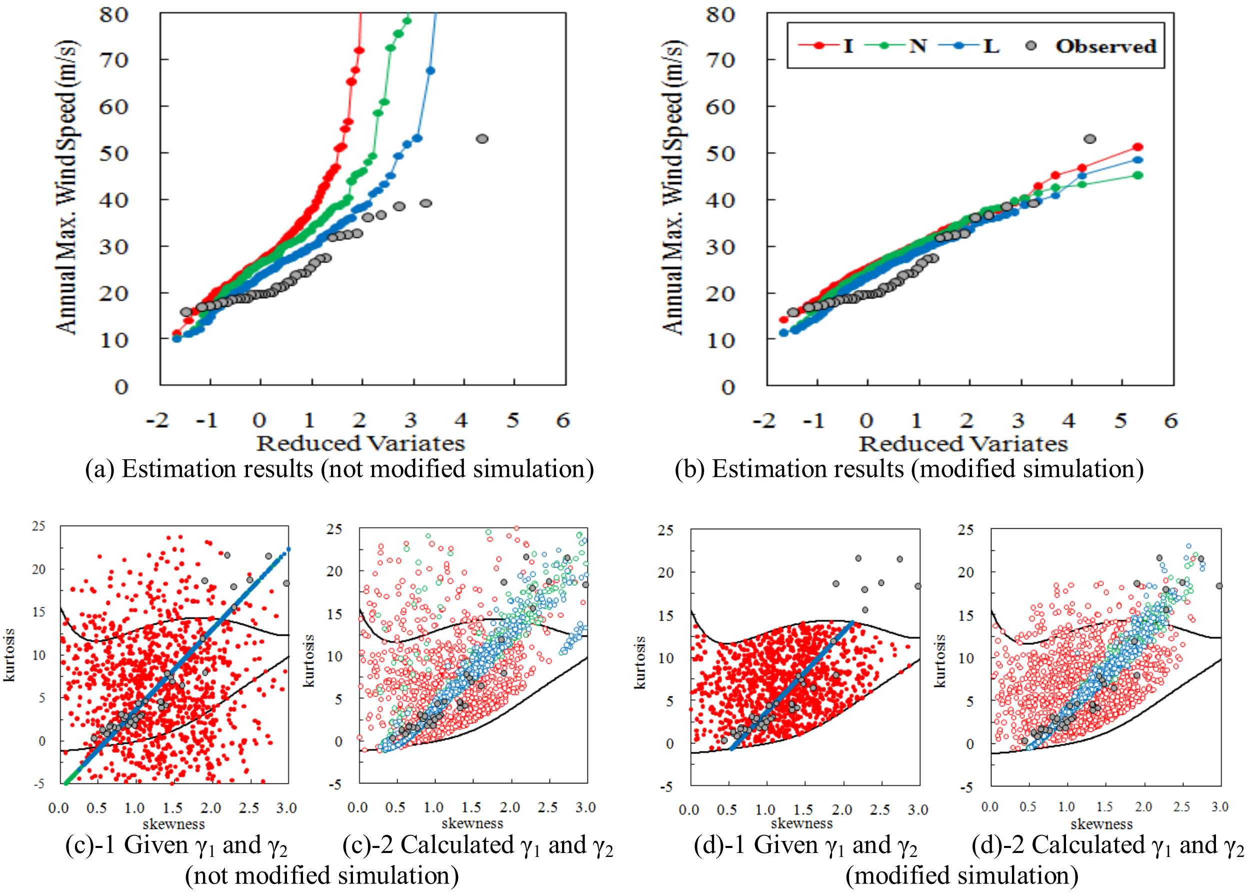
(a) Estimation results (not modified simulation) (b) Estimation results (modified simulation)



(c)-1 Given γ_1 and γ_2 (c)-2 Calculated γ_1 and γ_2
(not modified simulation)

(d)-1 Given γ_1 and γ_2 (d)-2 Calculated γ_1 and γ_2
(modified simulation)

Figure 10 Estimation results of Kagoshima



(a) Estimation results (not modified simulation) (b) Estimation results (modified simulation)

(c)-1 Given γ_1 and γ_2 (c)-2 Calculated γ_1 and γ_2
(not modified simulation)

(d)-1 Given γ_1 and γ_2 (d)-2 Calculated γ_1 and γ_2
(modified simulation)

Figure 11 Estimation results of Ishigakijima

Table 3 Moment parameters of Tokyo, Kagoshima, and Ishigakijima

	$E(\mu)$	$E(\sigma)$	$E(\gamma_1)$	$E(\gamma_2)$	$\sigma(\mu)$	$\sigma(\sigma)$	$\sigma(\gamma_1)$	$\sigma(\gamma_2)$
Tokyo	3.2724	1.8067	1.1509	2.1486	0.2412	0.1699	0.2225	1.0384
Kagoshima	2.8272	1.6471	1.3456	4.6173	0.3144	0.1501	0.3812	4.3350
Ishigakijima	4.5129	2.1467	1.2808	6.3152	0.3787	0.2263	0.6496	6.4729

In Figure 9 ~ 11, the generations of given skewness and kurtosis are assumed as three conditions: (1) **I**: independently normal distributed; (2) **N**: fully correlated and normal distributed; and (3) **L**: fully correlated and lognormal distributed. Table 3 lists the moment parameters of each site.

Generally from Figure 9~11, it is observed that fully-correlated skewness and kurtosis are generally better than independent ones. Further, a lognormal distribution for generation of skewness and kurtosis is better than a normal one in Kagoshima and Ishigakijima. It seems better to assume a lognormal distribution when the moment parameters, $E(\gamma_2)$ and $\sigma(\gamma_2)$, are relatively large. Comparing the truncated estimation results with not truncated ones, it is observed that the tails of estimation results are so suppressed that differences between condition **I**, **N**, and **L** are small. Given skewness and kurtosis in the truncated cases are limited inside the application range. Some of calculated skewness and kurtosis are outside the range because of 20% error acceptable.

Though it seems the simulation process with truncation in condition **I**, **N**, or **L** can always provide a fairly good agreement of estimation results, the tail part is sometimes under-estimated, especially in those extreme cases. The possibility of generating a large skewness and kurtosis based on large $E(\gamma_2)$ and $\sigma(\gamma_2)$ should not be ignored by truncation in the simulation process. A better method to estimate parameters of polynomial translation method combined with condition **L** may provide a universal simulation process for all the cases.

6. CONCLUSIONS

The approximation sheet proposed by Edgeworth was verified not suitable when the given skewness and kurtosis are extremely large. Observed skewness and kurtosis of 155 sites were plotted to see the general distribution on the application range of parameter estimation. Simulation process with truncation was introduced and several conditions were concerned based on basic characteristics observed four moments. For those sites with large observed skewness and kurtosis, a lognormal distribution of skewness and kurtosis is better than normal distribution. Simulation process with truncation may improve the agreement of the estimation results. However, truncation also causes under-estimation in those extreme cases. A more general application for the parameter estimation of the polynomial form should be investigated for a more universal practicality.

REFERENCE

1 A.I.J., 1996, 'AIJ Recommendations for Loads on Buildings', Architectural Institute of Japan.

2 H. Choi and J. Kanda, 2003, 'Translation method: a historical review and its application to simulation of non-Gaussian stationary process', *Wind and Structures*, 6(5), pp357-386.

3 H. Choi and J. Kanda, 2005, 'Extreme value distribution of annual maximum wind speeds as I.N.I.D. random variables', *Extreme Value Theory and Applications,* The Institute of Statistics Mathematics, Japan, pp59-77 (in Japanese).

4 H. Choi and J. Kanda, 2005, 'The cult of isolated statistics in extreme value analysis', *Proc. 6th Asia-Pacific Conference on Wind Engineering*, pp979-1004.

5 F.Y. Edgeworth, 1898, 'On the representation of statistics by mathematical formulae (Part I)', *Journal of the Royal Statistical Society*, 61(4), pp670-700.

6 H. Choi and J. Kanda, 2008, 'Tail characteristics of non-i.i.d. maximum wind speed model', *Proceedings of Asia Pacific Symposium on Structural Safety and its Applications*, Hong Kong.

7 J. Kanda and Yuan-Lung Lo, 2009, 'Annual maximum wind speed simulations based on moment parameters of parent distributions', *ISM Reports on Statistical Computing*, Institute of Statistical Mathematics.