# Effects of Higher Moments on Distributions of Annual Maximum Wind Speeds

**Jun Kanda and Yuan-Lung Lo**
Professor, Graduate School of Frontier Science,
The University of Tokyo
5-1-5 Kashiwa Campus, Kashiwa-shi, Chiba, 277-8563, Japan

**Keiko Ohnishi**
3-22-4-207, Kitaotuka, Toshima-ku, Tokyo, Japan

## Abstract

The relation between higher order moments of 10-minute mean wind speeds and the probability distribution of annual maximum wind speeds have been examined by utilizing observed wind data in Japan. It is found that the four moment parameters determine the tail characteristics of extreme value distributions and these moment parameters vary yearly and regionally. The variations of skewness and kurtosis play significant roles in particular. When the mean and the standard deviation of four moments of parent distribution are appropriately estimated, the simulation of annual maximum wind speeds provides a good correspondence to the distribution of observed annual maximum wind speeds. The possibility of good estimation of annual maximum wind speeds may be expected from a short period of data, such as 5 years when the variations of four moments are properly estimated as a representative one for long-term statistics.

## 1. Introduction

In order to determine the appropriate design wind speed for structures, it is important and essential to estimate the probability distribution of annual maximum wind speeds. For the sustainable society such importance is highly appreciated, since the lower estimates may cause frequent failures and the higher estimates will cause over-design which leads to the waste of materials and energy. Therefore the tail characteristics of extreme value distributions for natural hazards have to be examined to provide convenient mathematical models for engineering decision making regarding the long term policy for the structural safety in a society.

The maximum value of 10-minute mean wind speeds is only one value among 52,560 data per year. However it seems reasonable to assume that the statistical nature of 10-minute mean wind speeds provides sufficient information corresponding to the characteristics of annual maximum wind speeds. The annual variation of statistical parameters is also generally expected and may suggest the commonly assumed hypothesis of identical distribution for the extreme value theory is not satisfied.

When a sufficient number of data is available, the statistics of extremes may provide good models by simply applying the extreme value distribution theory. However generally the period of records is not long enough and the number of observation sites is not sufficient for a specific area. The parent distributions for wind speeds are examined to find out the relation between the four moments of parent distribution and the probability distribution of annual maximum wind speeds. It has been reported that the four moments of parent distribution vary to suggest the non-identical nature and that the coefficients of variations of the third and fourth moments contribute to the tail characteristics of annual maximum distribution significantly (Kanda and Choi, 2008). Then it is expected that once the standard deviations of higher moments of parent distribution are appropriately estimated, the design wind speed corresponding to a return period, i.e., the inverse of annual probability of exceedance, can be estimated reasonably well from relatively short periods of data.

Higher moments of parent distributions of wind speeds at observation sites in Japan are examined and simulation studies are carried out to confirm that once the mean of four moments and the standard deviation of four moments represent the probabilistic nature of wind speed distributions, the probability distributions of annual maximum wind speeds can be estimated. The tail characteristics of annual maximum wind speeds vary rather significantly depending on regions. Then a procedure for estimating the annual maximum wind speeds from relatively short period data at several sites is attempted to examine its appropriateness in practices.

## 2. Four moments of parent distribution of 10-minute mean wind speeds in Japan

Moment parameters of probability distributions are properly estimated when the number of data is sufficient. However the data caused by natural phenomena, such as the 10-minute mean wind speeds, may not be identically distributed and mutually independent. Necessary number of data is examined to have consistent estimates of moment parameters for 10-minute mean wind speeds. The 10-year period of data (from 1998 to 2007) of the meteorological sites in Tokyo and Nagoya are provided as two common cases to investigate the degree of the scatter due to the insufficiency of data. Regional differences between these two sites are not discussed here while the yearly variations of four moments will be mentioned later.

The equivalent number of independent random variables for the 10-minute mean wind speeds in one year has been examined and found as 30,000 (Choi and Kanda, 2003). The number needed to estimate moments, however, may be much less than the equivalent independent number. The variation of estimates due to every 1 hour data (the number of data is 8760), every 3 hour data (2920), every 6 hour data (1460), every 12 hour data (730) and every 24 hour data (365) are compared with those obtained from full data (52560) in Figure 1. The first 10-minute mean wind speed within one hour is indicated as the every 1 hour data and so is as the every 3 hour data, every 6 hour data, every 12 hour data and every 24 hour data. In both sites, the scatter becomes significant for every 6 hour data, although the correlation is not generally so bad, while a very good correspondence can be found for every 3 hour data, every 1 hour data and full data. It is interesting even for a higher moment, such as the skewness and kurtosis, one sixth of data (every 1 hour data) or sometimes one $18^{th}$ of data (every 3 hour data) are sufficient for the estimation of moment parameters. Estimated moments due to longer than 3 hour interval may not be regarded as good estimates.

Then the distributions of these four moments in 155 observation sites in Japan are examined. In all sites, every 3 hour data in 42 years (from 1961 to 2002) is utilized. Since the moment parameters vary yearly, the mean of four moments and the standard deviation of four moments are plotted as histograms in Figure 2. The mean of mean varies from 0.7 m/s to 7.5 m/s. The mean of standard deviation varies from 0.4 m/s to 4.4 m/s. The mean of skewness varies from 0.4 to 2.1 and the mean of unbiased kurtosis varies from 1.9 to 8.6. The standard deviation for these four moments also varies, as from 0.1 m/s to 2.3 m/s, from 0.1 m/s to 1.2 m/s, 0.1 to 0.7 and 0.3 to 6.7 for the mean, standard deviation, skewness and unbiased kurtosis respectively. 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 and the identical hypothesis would not be applied.

The correlation between the skewness and the kurtosis is examined for 8 sites respectively in Figure 3. The correlation seems to be very high and indicates that the non-Gaussian nature may be represented either by the skewness or the kurtosis (Kanda and Choi, 2008). 8 selected sites are Wakkanai, Hachinohe, Niigata, Tokyo, Nagoya, Osaka, Hagi and Kagoshima whose geographical locations are shown in Figure 4.

## 3. Simulation of annual maximum wind speeds by the polynomial translation method

Effects of the moment parameters of parent distribution on the probabilistic distributions of annual maximum wind speeds are examined by a parametric study. The flow of simulation procedure is shown in Figure 5. Since the non-identical nature is clearly observed in Figure 2, 100 sets (corresponding to 100 years) of four moments are generated with the mean of four moments and the standard deviation of four moments which are parametrically varied referring to the histogram of regional distribution. As shown in Figure 6, the mean and the standard deviation of four moments are regarded as variables for the parametric study and one of these eight parameters is varied as regional mean plus N times regional standard deviation, where N is varied within meaningful ranges according to Figure 2, while other 7 parameter values are fixed to the regional mean.

By using these mean and standard deviation of four moments, which are regarded as yearly mean and yearly standard deviation, the i-th set of four moments for one year parent distribution can be generated by the following equations.

$$\mu_i = E(\mu) + R_{1,i}\sigma(\mu) \tag{1}$$

$$\sigma_i = E(\sigma) + R_{2,i}\sigma(\sigma) \tag{2}$$

$$\gamma_{1i} = E(\gamma_1) + R_{3,i}\sigma(\gamma_1) \tag{3}$$

$$\gamma_{2i} = E(\gamma_2) + R_{4,i}\sigma(\gamma_2) \tag{4}$$

where $R_{1,i}$, $R_{2,i}$, $R_{3,i}$, and $R_{4,i}$ are random numbers generated from the standard normal distribution for i-th set of four moments; $E(\mu)$, $E(\sigma)$, $E(\gamma_1)$, and $E(\gamma_2)$ represent the mean of four moments; $\sigma(\mu)$, $\sigma(\sigma)$, $\sigma(\gamma_1)$, and $\sigma(\gamma_2)$ represent the standard deviation of four moments;

Then for a distribution of each set of the four moments, 30,000 samples are generated to find the maximum which corresponds to the annual maximum according to the polynomial translation method (Choi and Kanda 2003).

A random variable $y$, which is given a set of four moments, is written in a polynomial form with respect to a standard normal random variable $x$ as

$$y = a + bx + cx^2 + dx^3 \tag{5}$$

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

$$E(y) = a + c = 0 \qquad (6)$$

$$Var(y) = b^2 + 6bd + 2c^2 + 15d^2 = 1 \qquad (7)$$

$$\gamma_1(y) = 2c(b^2 + 24bd + 105d^2 + 2) \qquad (8)$$

$$\gamma_2(y) = 24\{bd + c^2(1 + b^2 + 28bd) + d^2(12 + 48bd + 141c^2 + 225d^2)\} \qquad (9)$$

where $\gamma_1, \gamma_2$ are the skewness and the unbiased kurtosis.

By the procedure 100 annual maxima as $y$ are generated and plotted in a descending order on the Gumbel paper. Then by repeating this procedure for 11 times, 11 sets of 100 annual maxima are generated. Distributions are rather scattered, but the median estimate is regarded as the best estimate which is approximated by the Gumbel distribution described as,

$$F(v) = \exp\left\{\exp\left\{-\left(\frac{v-l}{s}\right)\right\}\right\} \qquad (10)$$

where $v$ is the annual maxima; $l$ is the location parameter and $s$ is the scale parameter. Both location and scale parameters can be estimated by the moment method for one set of 100 annual maxima.

The effects of variations of the mean of four moments and the standard deviation of four moments on the estimated location parameters and scale parameters are shown in Figure 6. The abscissa shows the variation of each parameter in terms of the standard deviation of regional variation from the regional mean. When the mean of the mean and the standard deviation increases, the location parameter increases. When the standard deviation of the standard deviation and the kurtosis increases, the scale parameter increases. The correlation between the skewness and the kurtosis is assumed independent here only to investigate the effect of each parameter on the Gumbel parameter estimation without considering the high correlations shown in Figure 3.

## 4. Comparison of simulated maxima with statistical estimates and discussions

Every 3 hour data of 8 sites are selected to examine the agreement between the simulated estimates and annual maximum wind speeds. The simulation is carried out in the same manner as described in the previous section but the mean of four moments and the standard deviation of four moments are those obtained from the statistical data of each individual site. The period of data is 42 year, i.e. from 1961 till 2002. The skewness and kurtosis are treated as fully dependent as their correlation is very high for any sites as observed in Figure 3.

The simulation results for estimation of annual maximum wind speeds are shown in Figure 7 together with observed annual maximum wind speeds. The firm line with hollow circles indicates the median estimate of 11 sets of simulations. The agreements are generally fairly good except for the case of Kagoshima which shows a concave tendency on the Gumbel probability paper. The variation of four moment parameters is significant as shown in Figure 2. The non-identical nature clarified by the variation is represented by the simulation. Although the extreme value distribution such as the Gumbel distribution of equation (10) is derived with the identical hypothesis, statistical estimates based on the observed data rather consistent to the Gumbel form except for the simulated results of Kagoshima.

The variation of four moments also varies with the time. Examples are shown for these 8 sites in Figure 8. Exceptionally high kurtosis values are observed at Kagoshima in 1996. The tendency of variation of Kagoshima is quite similar to that of Kumamoto. Some observation sites within a small distance may have similar tendencies. But this is not always the case. When the mean of four moments and the standard deviation of four moments are properly estimated, the simulation is expected to provide a good estimate of the annual maximum distribution.

Then 5-year period of data for 8 sites are used to estimate the mean of four moments and the standard deviation of four moments. The simulation is proceeded to examine the agreement with observed statistics. It is interesting to see the correspondence with the Figure 7. Cases for (a)Wakkanai, (b)Hachinohe, (c)Niigata and (f)Hagi show good agreements except for one or two cases of simulations based on 5-year period of data. Cases for (e)Tokyo and (g)Osaka show some underestimations for the range of longer return period with some cases of good agreements. Cases for (d)Nagoya and (h)Kagoshima show large scatters for the agreement with observed data. In the case of the 1$^{st}$ 5-year period simulation at Nagoya and in the cases of the 1$^{st}$, 2$^{nd}$, 6$^{th}$, and 8$^{th}$ at Kagoshima, the median estimate shows the concave tendency in the tail of the distribution on the Gumbel probability paper and this tendency seems to be related to a high variation of higher moments in the 5-year period. In order to summarize the agreements of fitting for estimations based on 5-year period statistics, Table 1 is shown together with the standard deviation of skewness variation. In general when the variation of skewness is significant in the 5-year period of data, the agreement is expected to be fairly good, although some overestimated cases are found when the standard deviation of skewness is large. When the moment parameters vary insignificantly, i.e. the standard deviation is small, simulated results tend to be underestimated. Then attention should be paid when making estimations for the annual maximum wind speed based on a short period data while the standard deviations of higher moments in this short period are small or extremely large.

Further examination is made to observe the effects of insufficient data on the simulation results of annual maxima by utilizing 5-year period of data (from 2001 to 2005) at Nagoya and Tokyo. Simulation results are shown in Figure 9 together with observed annual maximum wind speeds from 1961 to 2005. It is obvious that insufficient data (every 6 hour data or longer) may lead to improper estimates of moment parameters or even to underestimation in simulation results of annual maxima.

When a construction project is planned, wind speed measurements at the site are always encouraged. It is expected to take more than 5 to 10 years for a large project before the beginning of construction. Even when only a short period data is available, such information as the four moments of 10-minute mean wind speeds in certain period of years are considered to be meaningful for the estimation of annual maximum wind speed distribution. Even in cases when the direct estimation of the tail of extreme value distribution from the simulation based on a short period data may not be possible, it is useful as additional information to a macro-scale wind hazard map such as shown in AIJ load recommendations(A.I.J. 1996) for the determination of design wind speed.

## 5. Conclusions

The mean, standard deviation, skewness and kurtosis for 10-minute mean wind speeds in Japan were examined. The yearly variations of these four moment parameters are generally significant and the non-identical nature of the parent distribution is confirmed. The regional variation of these moments was also examined. The estimation of these four moments is fairly stable for reduced number of data. Every 3 hour data provides reasonably accurate estimation for four moments of the parent distribution. When the variation of four moment parameters are considered for the simulation of annual maximum wind speeds by the polynomial translation method, good agreements were observed by the median estimate for the annual maximum wind speed distribution and the statistical annual maxima. The utilization of short period data for such as 5 years for a successful simulation has been examined and the possibility was exemplified for various sites.

## References


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

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

Choi, H. and Kanda, J. (2005a), Extreme value distribution of annual maximum wind speeds as I.N.I.D. random variables, *Extreme Value Theory and Applications* (2), The Institute of Statistics Mathematics, Japan, pp59-77 (in Japanese)

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

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

Kanda, J. and Choi, H. (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*


Table 1

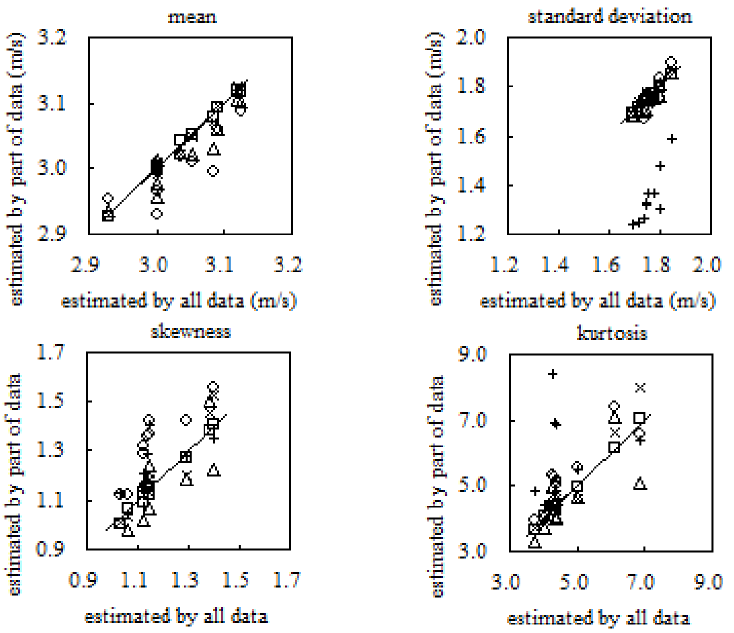
Summarized judgments on fitting agreement with standard deviation of skewness variation for 5-year period of data

| | Wakkanai | | Hachinohe | | Niigata | | Nagoya | |
|---|---|---|---|---|---|---|---|---|
| | s.d. | fitting | s.d. | fitting | s.d. | fitting | s.d. | fitting |
| 1961-1965 | 0.09 | good | 0.16 | + | 0.08 | good | 0.18 | + |
| 1966-1970 | 0.11 | good | 0.12 | + | 0.10 | good | 0.08 | - |
| 1971-1975 | 0.16 | + | 0.10 | good | 0.11 | good | 0.08 | - |
| 1976-1980 | 0.18 | good | 0.24 | good | 0.20 | good | 0.08 | - |
| 1981-1985 | 0.12 | good | 0.08 | good | 0.07 | good | 0.08 | - |
| 1986-1990 | 0.23 | + | 0.18 | good | 0.23 | good | 0.05 | - |
| 1991-1995 | 0.12 | good | 0.15 | + | 0.06 | good | 0.03 | - |
| 1996-2000 | 0.11 | good | 0.16 | good | 0.05 | good | 0.15 | good |

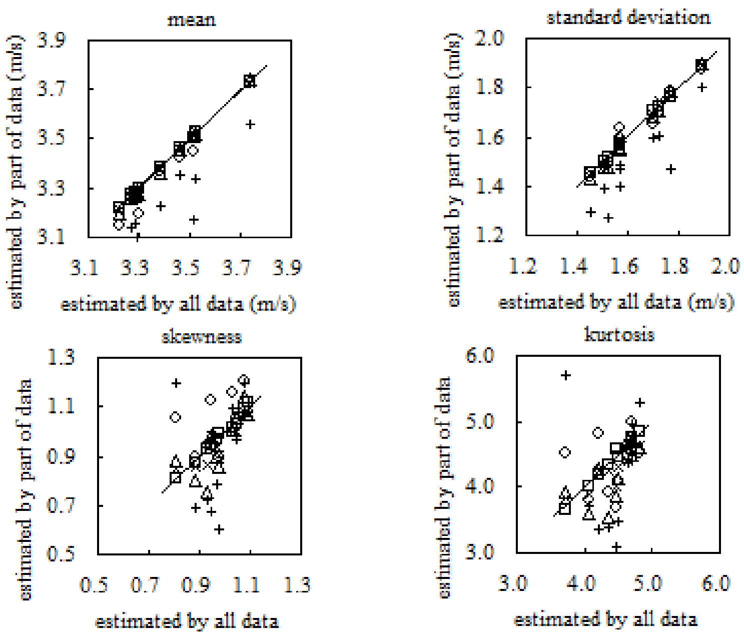
| | Tokyo | | Hagi | | Osaka | | Kagoshima | |
|---|---|---|---|---|---|---|---|---|
| | s.d. | fitting | s.d. | fitting | s.d. | fitting | s.d. | fitting |
| 1961-1965 | 0.09 | good | 0.11 | + | 0.10 | - | 0.49 | + |
| 1966-1970 | 0.13 | good | 0.12 | + | 0.10 | good | 0.53 | + |
| 1971-1975 | 0.14 | good | 0.16 | good | 0.16 | good | 0.39 | good |
| 1976-1980 | 0.36 | good | 0.21 | good | 0.04 | - | 0.23 | - |
| 1981-1985 | 0.12 | good | 0.08 | good | 0.09 | - | 0.08 | - |
| 1986-1990 | 0.05 | - | 0.23 | good | 0.12 | - | 0.42 | + |
| 1991-1995 | 0.07 | - | 0.16 | good | 0.09 | good | 0.20 | good |
| 1996-2000 | 0.10 | good | 0.20 | good | 0.16 | good | 0.63 | + |

※ s.d. : standard deviation of skewness    good : good fitting.
+ : significant overestimation    – : significant underestimation

—— estimated by all data  □ every 1 hour  × every 3 hour  △ every 6 hour  ◇ every 12 hour  + every 24 hour




(a) Nagoya




(b) Tokyo

Figure 1: Comparison of estimated four moments with reduced number of data
(The legend "estimated by all data" is a 45 degree line representing every 10 minute data.)

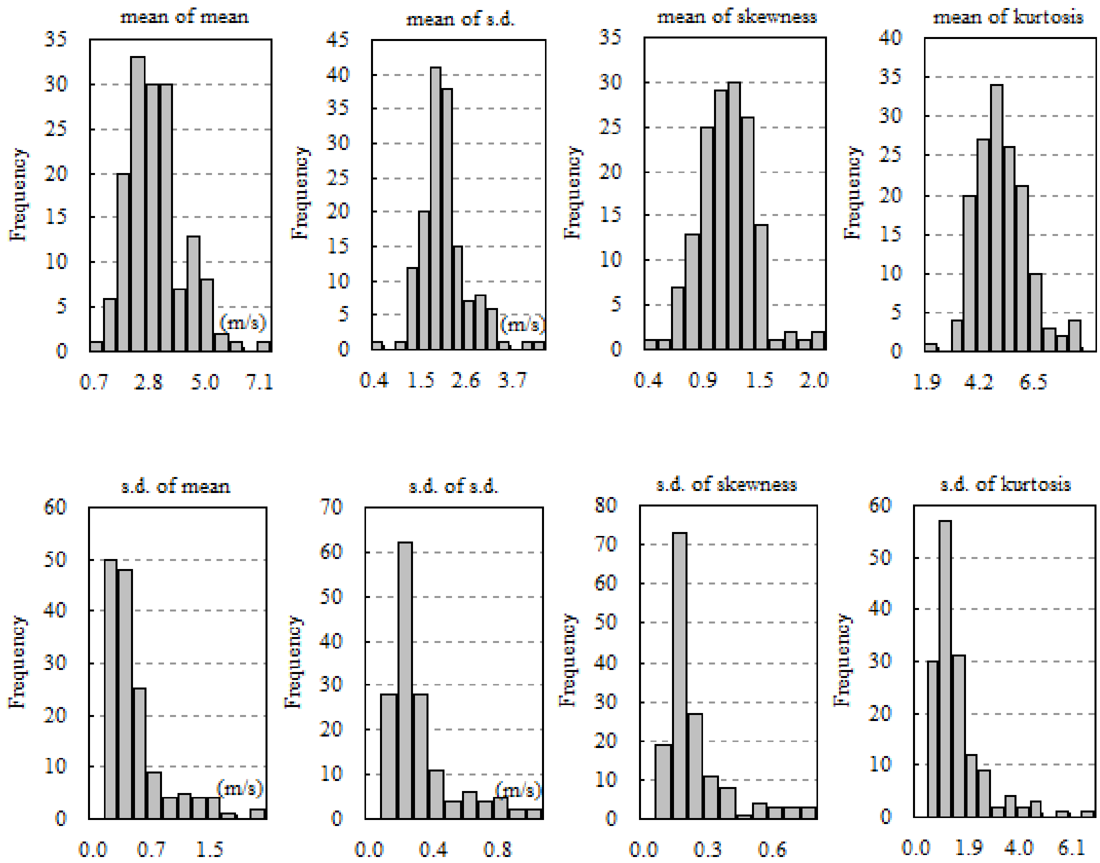



Figure 2: Histograms for the mean and the standard deviation for four moments of 10-minute mean wind speeds for 155 sites in Japan (s.d.: standard deviation)

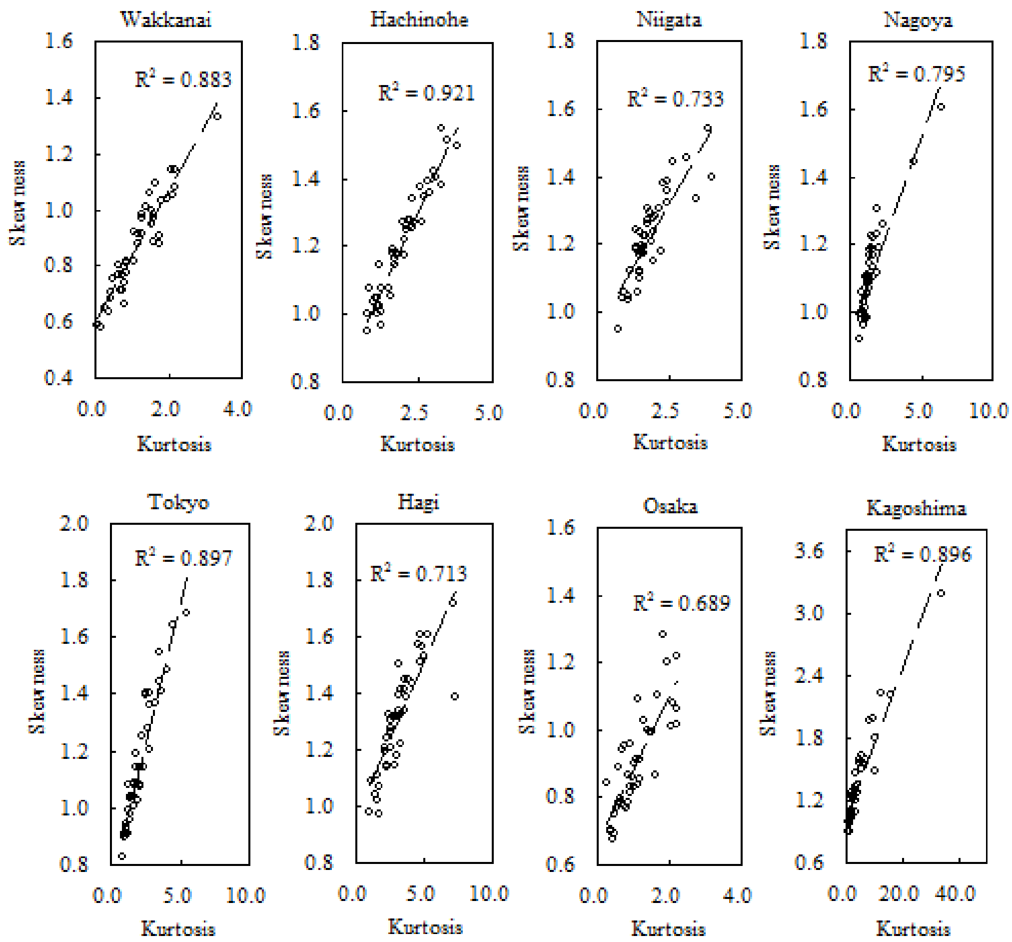



Figure 3: Correlation between the skewness and kurtosis for 8 sites


Figure 4: Geographical locations for 8 sites in Japan
(Wakkanai, Hachinohe, Niigata, Nagoya, Tokyo, Hagi, Osaka, Kagoshima)

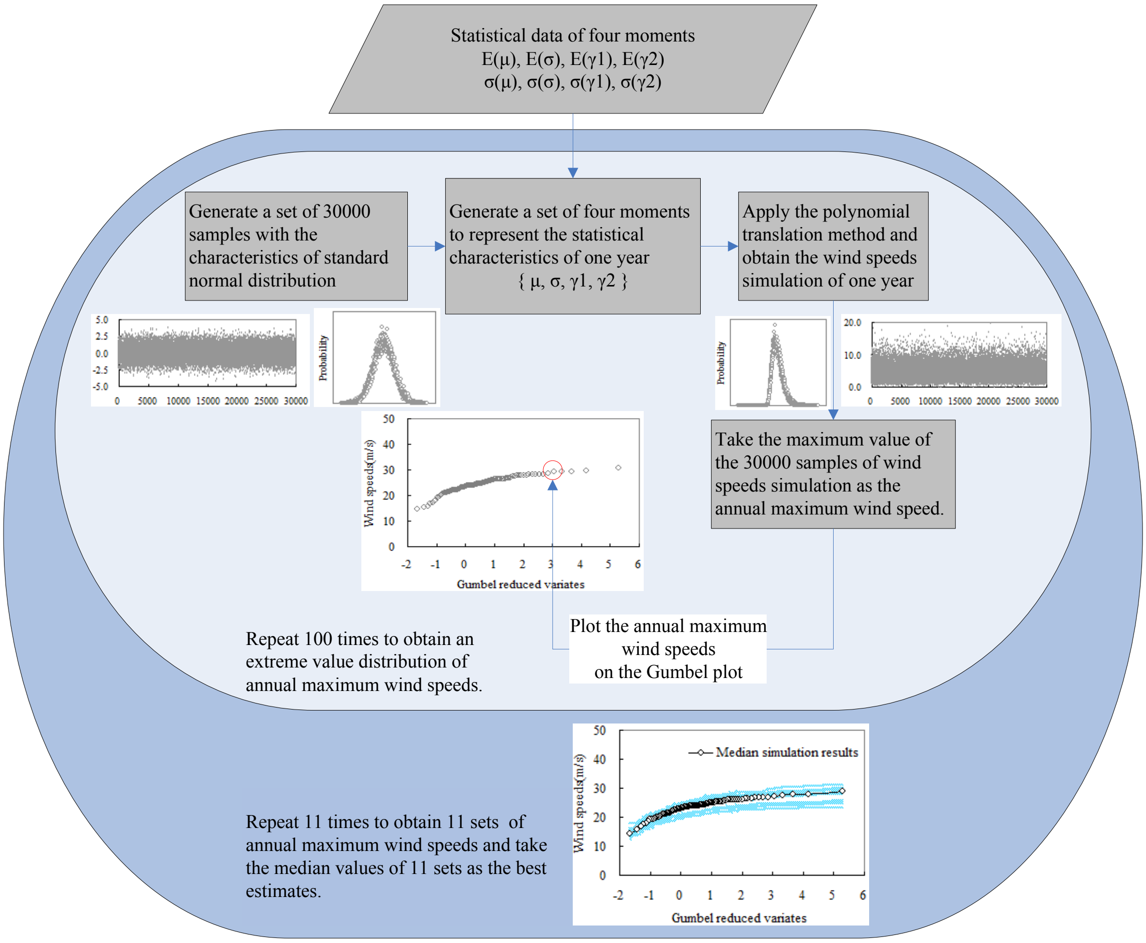



Figure 5: Flow diagram for simulation process

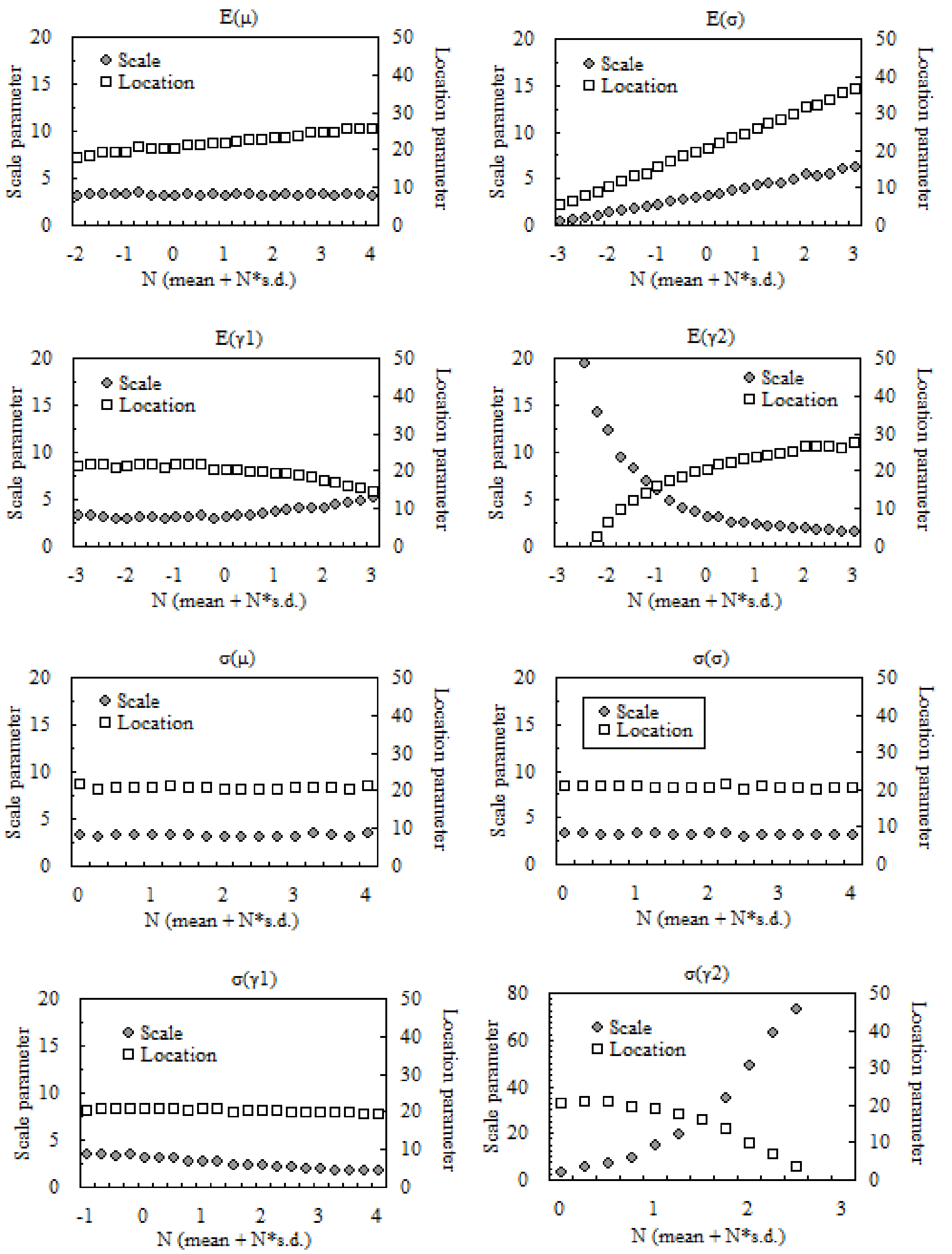



Figure 6: Effects of the mean and the standard deviation of four moments on the Gumbel parameters

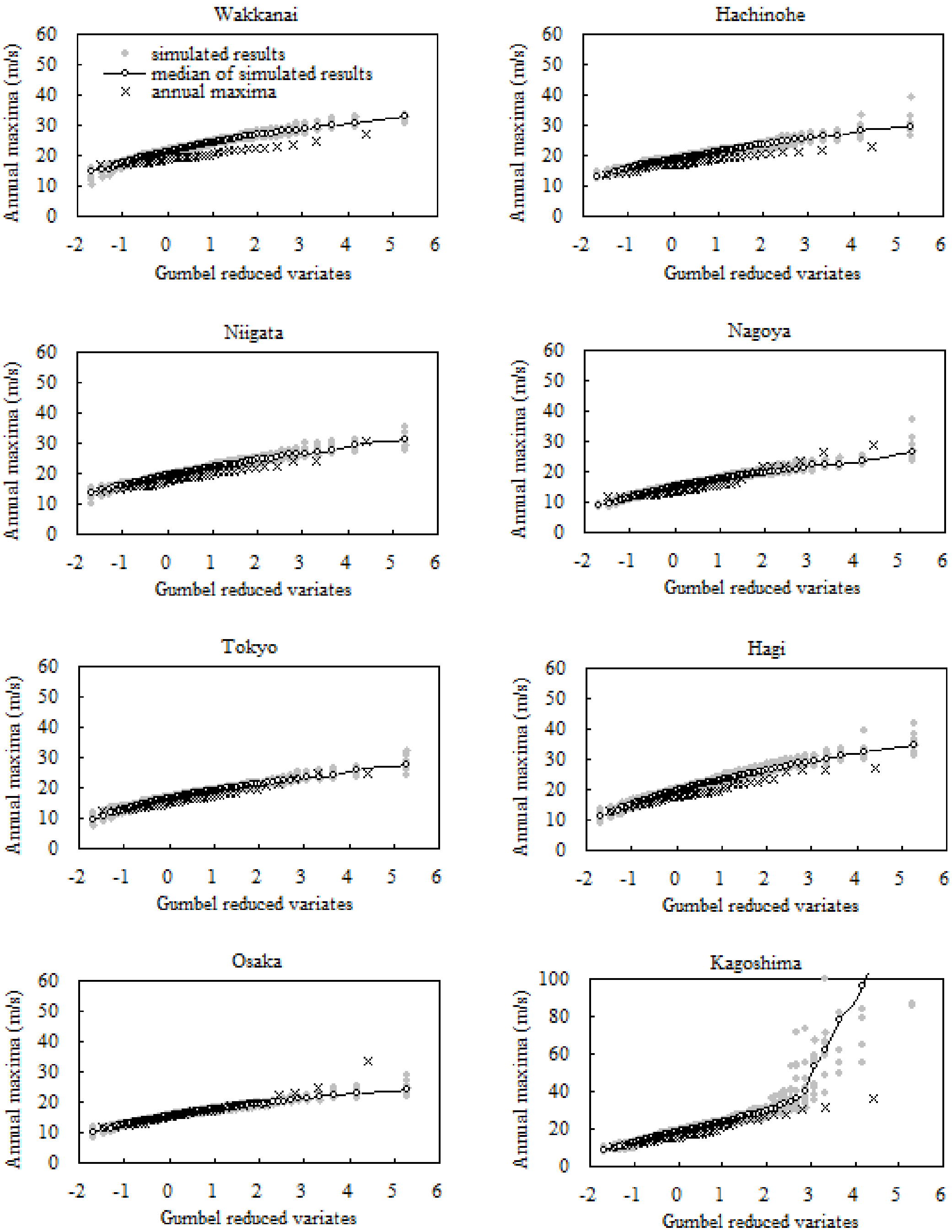



Figure 7: Simulated results of annual maxima based on every 3 hour data (1961-2002)

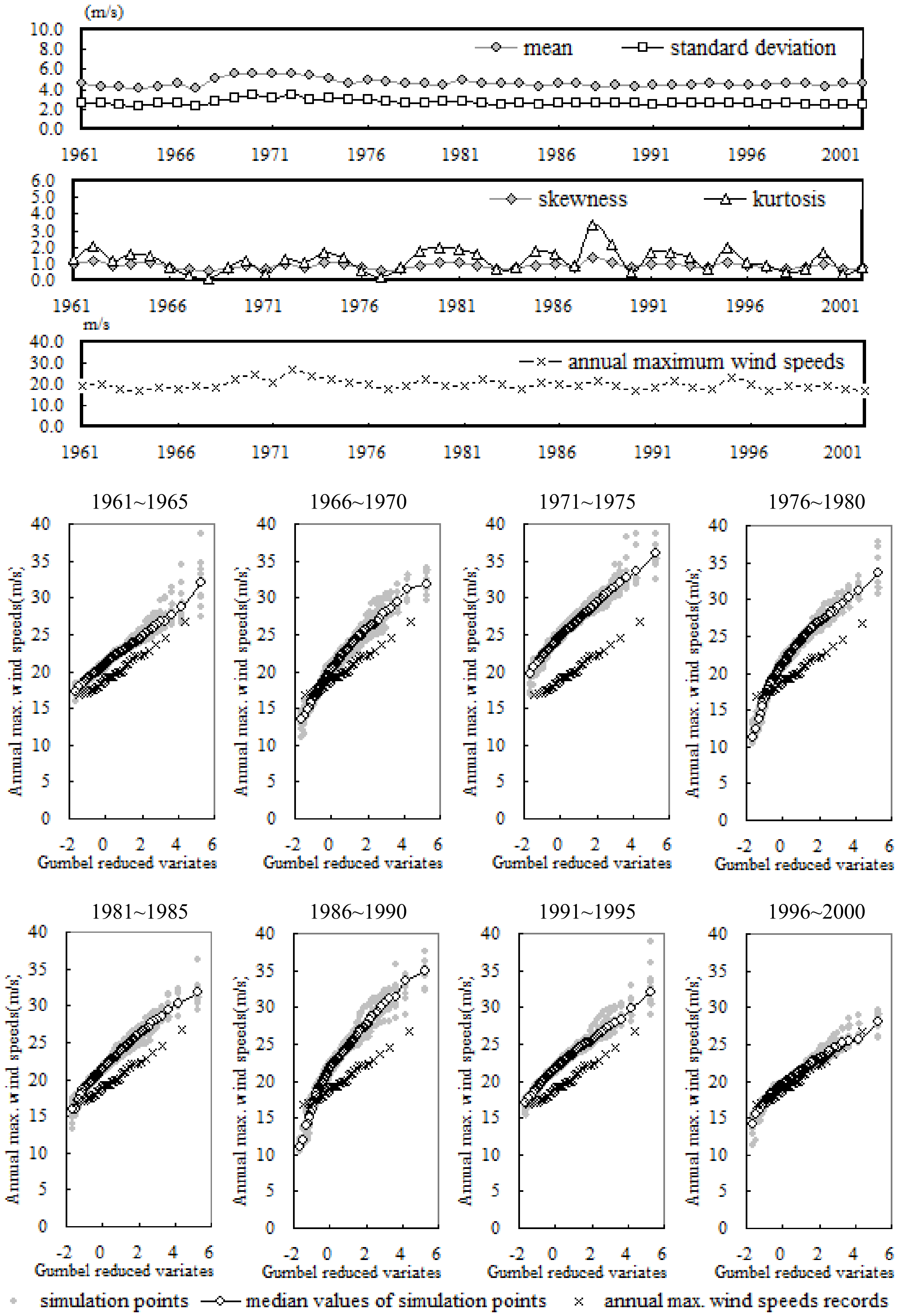



Figure 8(a) Yearly variations of four moments and annual maximum wind speeds;
Simulation results based on 5-year period of data at Wakkanai

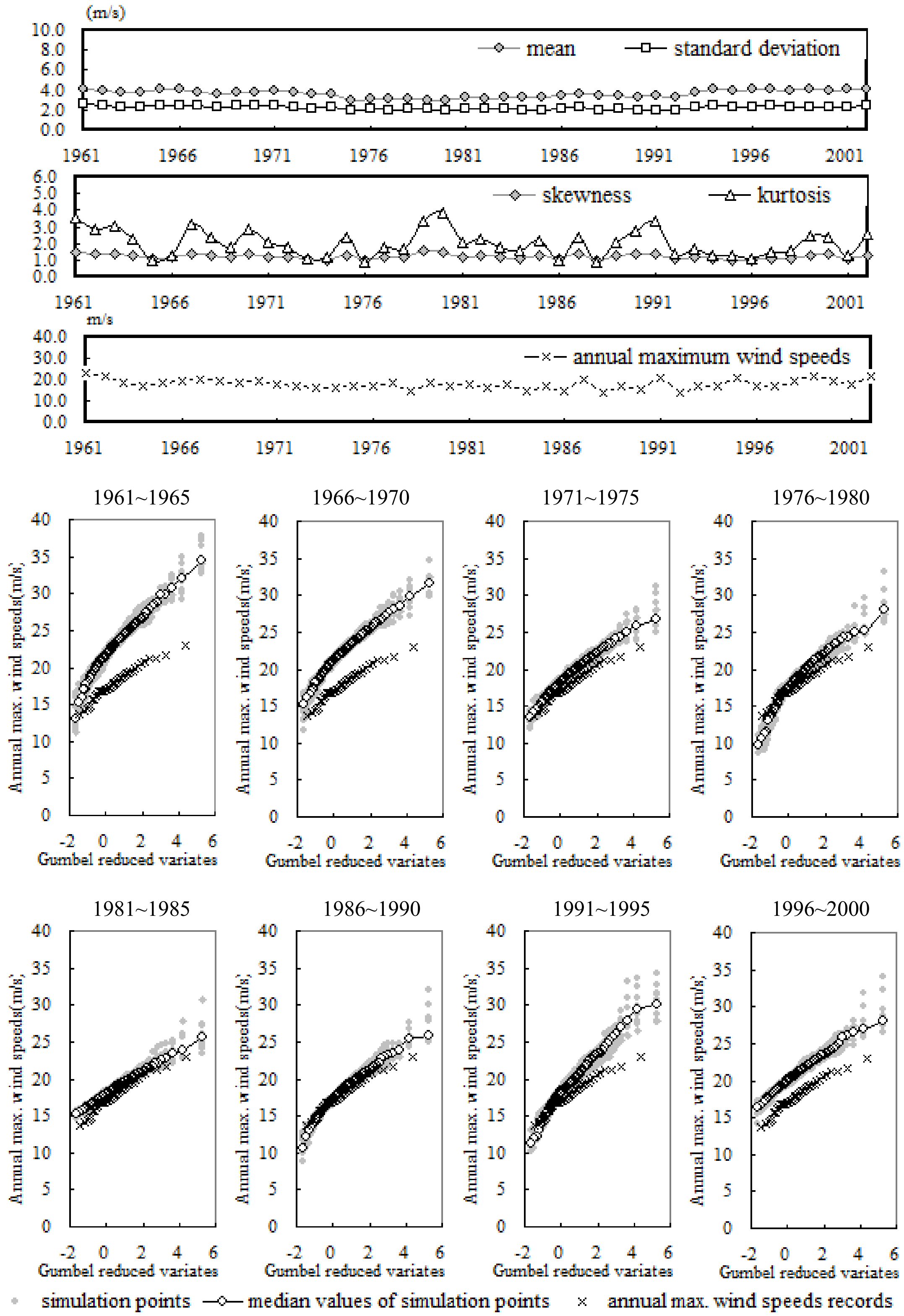



Figure 8(b) Yearly variations of four moments and annual maximum wind speeds; Simulation results based on 5-year period of data at Hachinohe

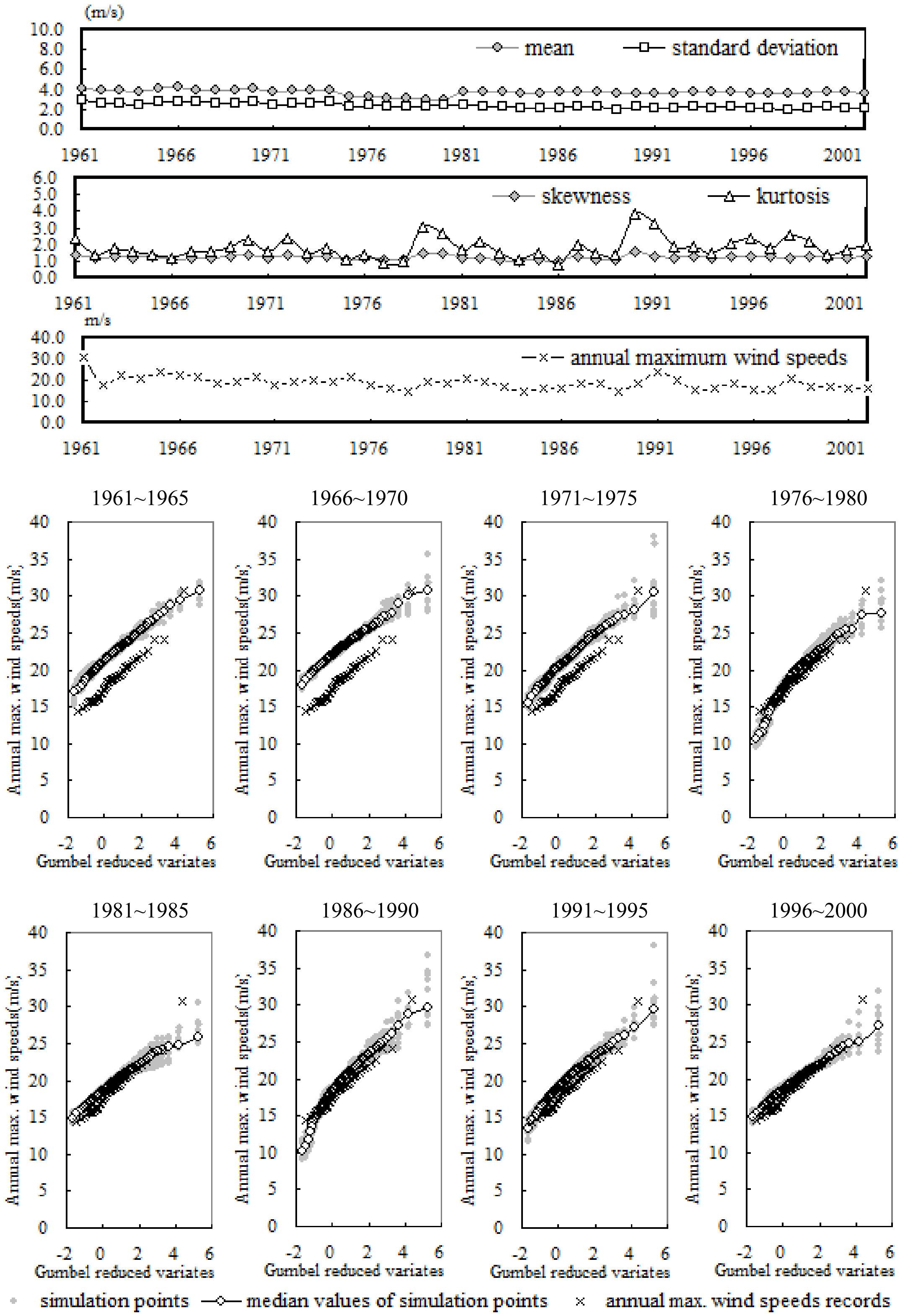



Figure 8(c) Yearly variations of four moments and annual maximum wind speeds;
Simulation results based on 5-year period of data at Niigata

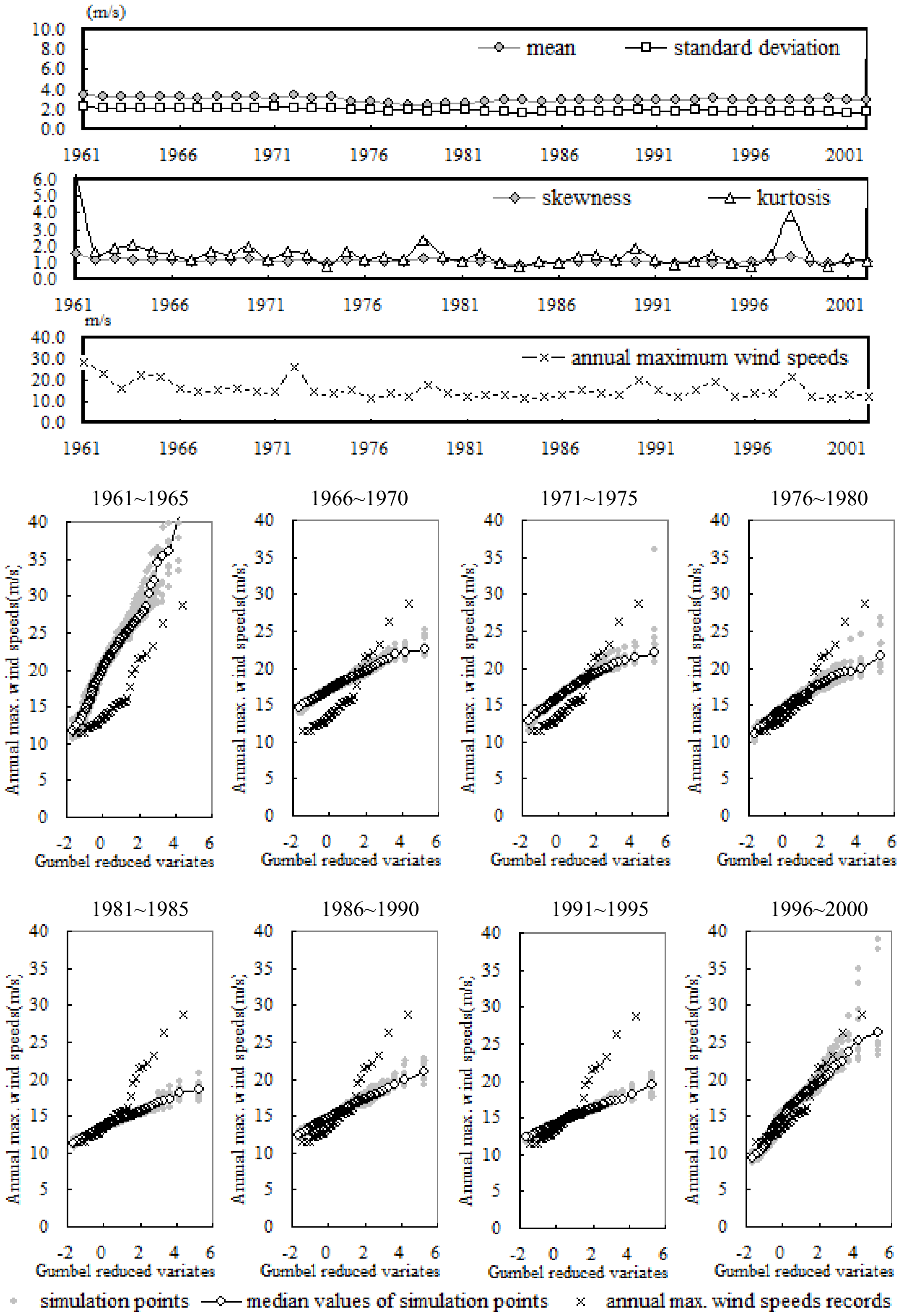



Figure 8(d) Yearly variations of four moments and annual maximum wind speeds;
Simulation results based on 5-year period of data at Nagoya

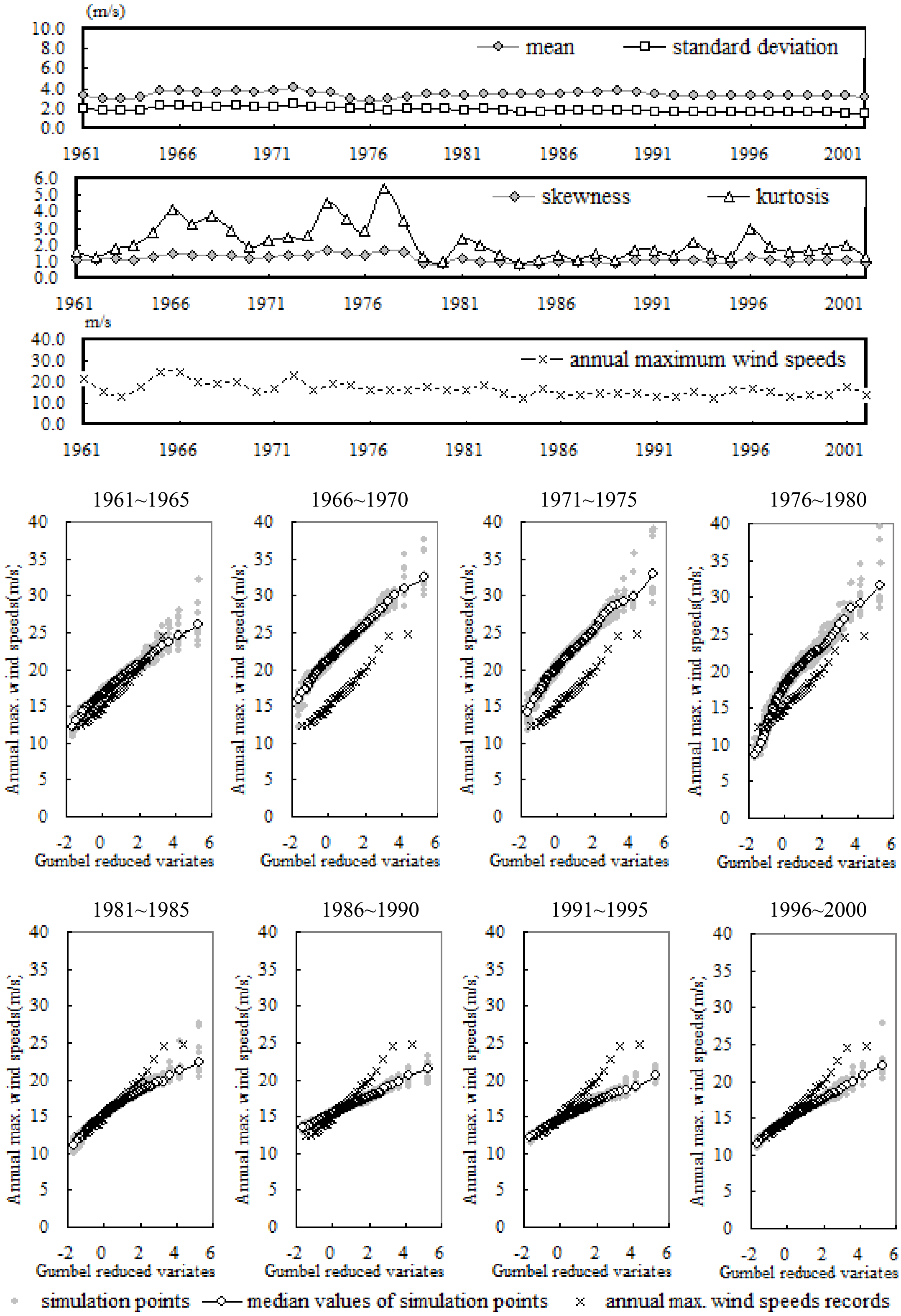



Figure 8(e) Yearly variations of four moments and annual maximum wind speeds;
Simulation results based on 5-year period of data at Tokyo

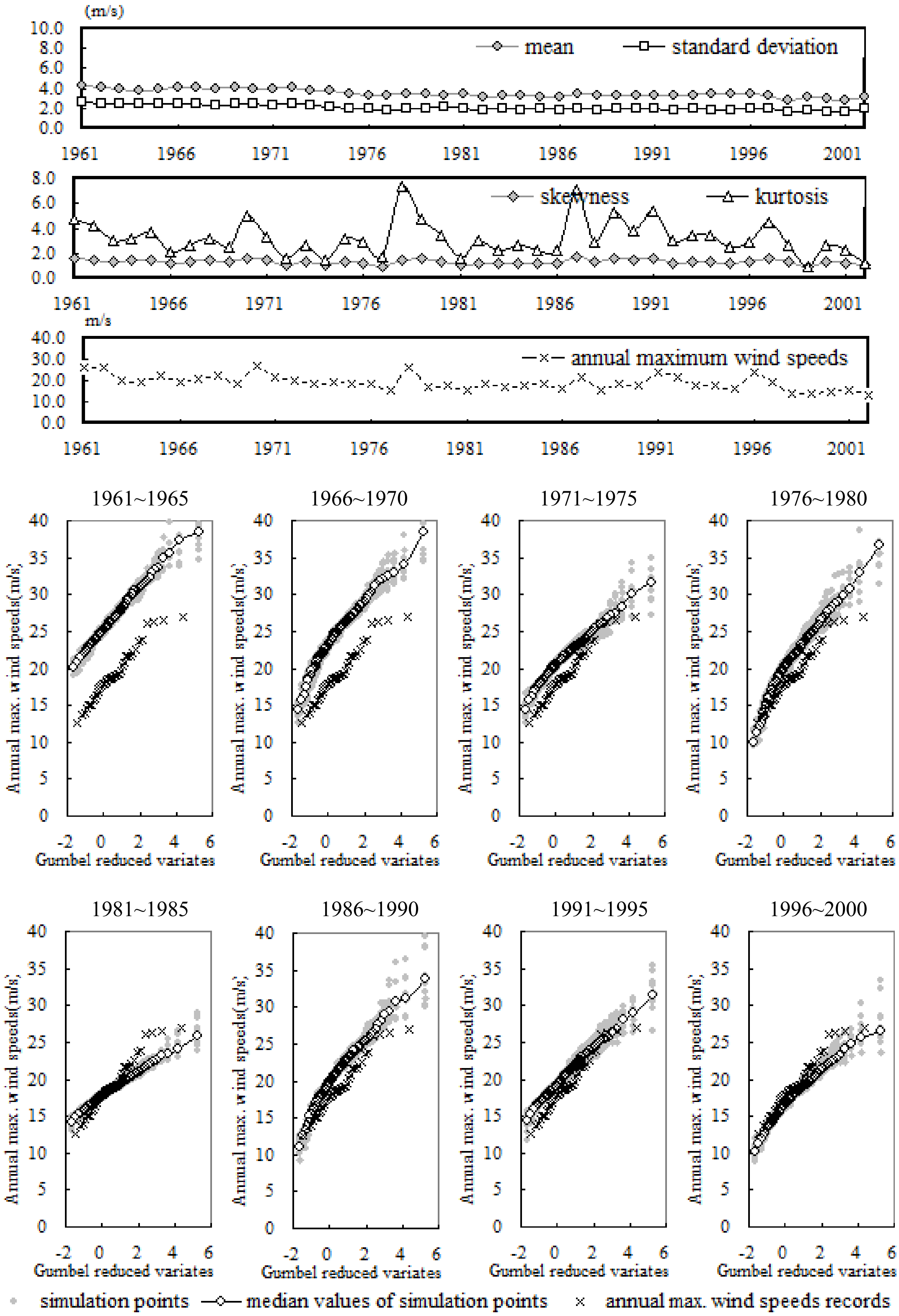



Figure 8(f) Yearly variations of four moments and annual maximum wind speeds; Simulation results based on 5-year period of data at Hagi

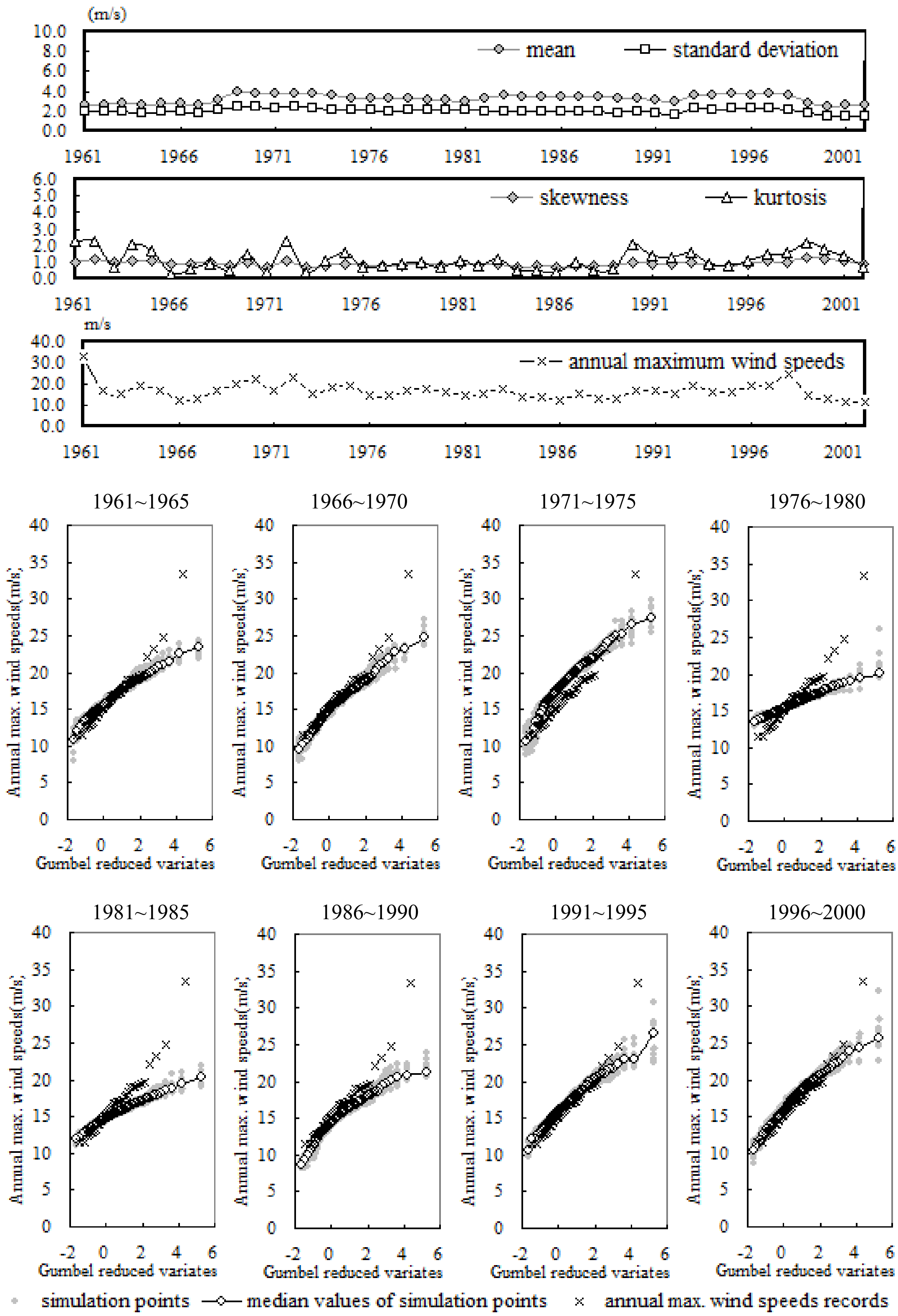



Figure 8(g) Yearly variations of four moments and annual maximum wind speeds;
Simulation results based on 5-year period of data at Osaka

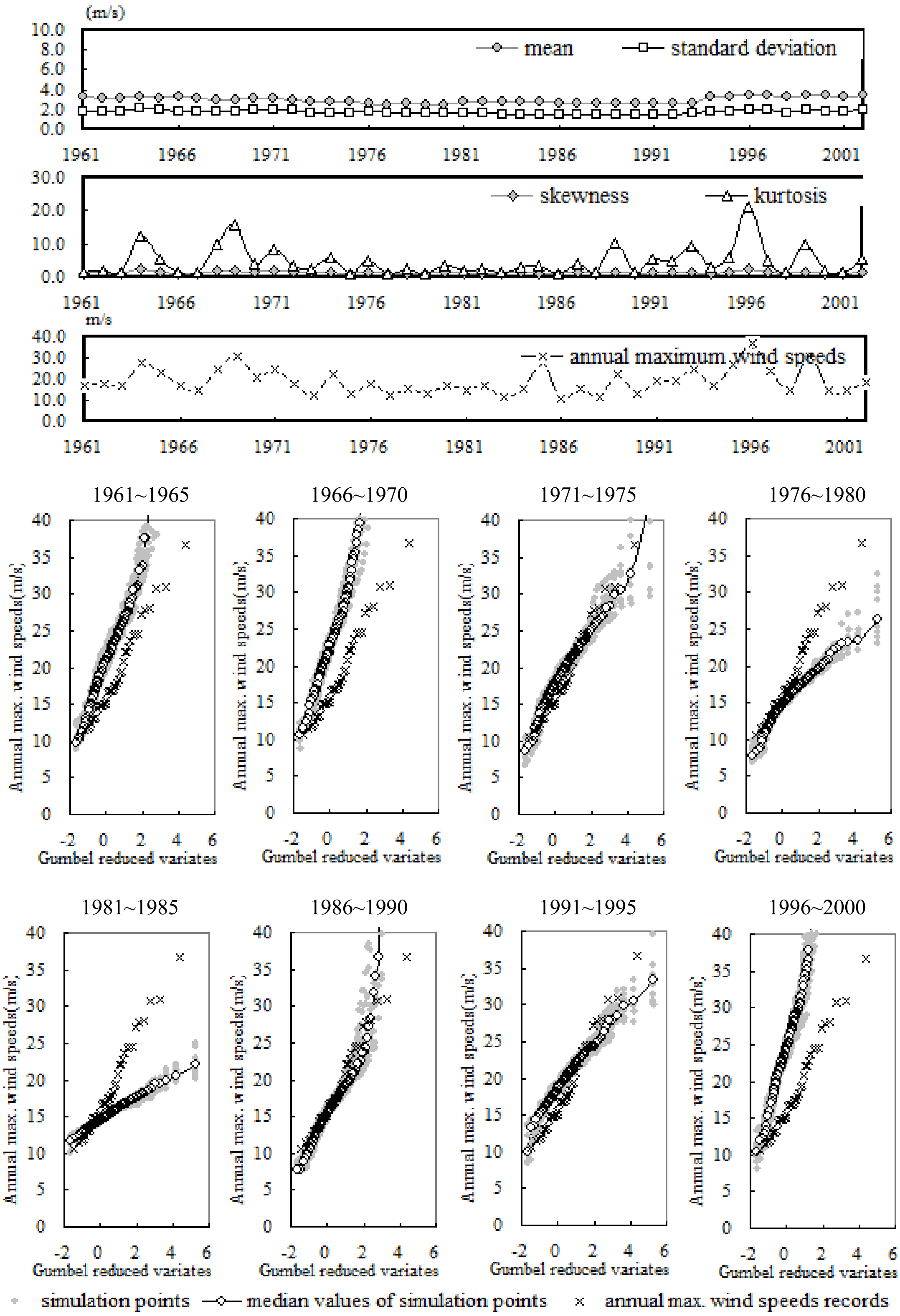
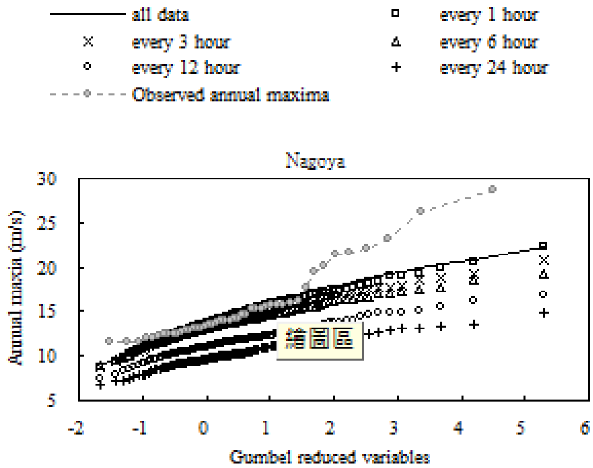
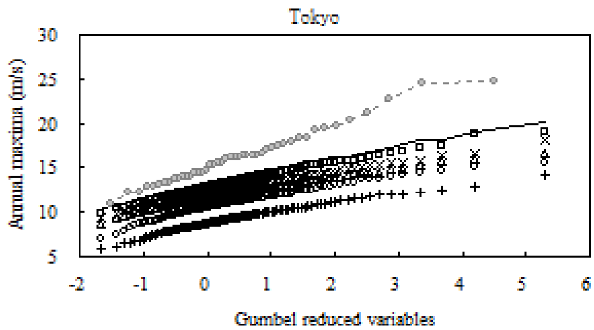



Figure 8(h) Yearly variations of four moments and annual maximum wind speeds; Simulation results based on 5-year period of data at Kagoshima




(a) Nagoya




(b) Tokyo

Figure 9: Median estimates of simulated results based on 5-year period of data from 2001 to 2005 and annual maxima from 1961 to 2005