INCORPORATION OF WEIBULL DISTRIBUTION IN *L*-MOMENTS METHOD FOR REGIONAL FREQUENCY ANALYSIS OF PEAKS-OVER-THRESHOLD WAVE HEIGHTS

Yoshimi Goda[1], Masanobu Kudaka[1], and Hiroyasu Kawai[2]

The *L*-moments of the Weibull distribution are derived and incorporated in the regional frequency analysis of peaks-over-threshold significant wave heights at eleven stations along the eastern coast of Japan Sea. The effective duration of wave measurements varies from 18.0 to 37.2 years with the mean rate of 10.4 to 15.1 events per year. The eleven stations are divided into three regions to assure homogeneity of the data. Both the Weibull and Generalized Pareto (GPA) distributions fit well to the observed data. The 100-year wave height varied from 8.2 to 11.2 m by the Weibull and 7.6 to 10.3 m by the GPA. The GPA distribution is not recommended for determination of design waves for these stations because it has an inherent upper limit and a tendency of under-prediction.

Keywords: extreme wave analysis, distribution function, Weibull distribution, GPA distribution, L-moments, regional frequency analysis

INTRODUCTION

Extreme wave analysis is the first step in coastal structure design by providing the basis of selecting design wave heights. A variety of methodology has been proposed for extreme analysis of environmental data such as flood discharge, storm wind speed, storm wave height, etc. An extreme data set is prepared by either the annual maximum (AM) data or the partial duration series data. The latter is also called the peaks-over-threshold (POT) data, because a peak value per storm over a preset threshold is employed to constitute the extreme data set.

There are many methods for distribution fitting to the data set such as the method of moments, probability-weighted moment method, least squares method, maximum likelihood method, *L*-moments method, etc., among which the *L*-moments method is the newest one proposed by Hosking (1990). It was developed from the probability-weighted moment method by Greenwood et al. (1978) and has the merits of easy application and straightforward estimation of distribution parameters. The FORTRAN program "lmoments" by Hosking can be downloaded at http://lib.stat.cmu.edu.

Hosking and Wallis (1997) have promoted the methodology of regional frequency analysis using the *L*-moments method for estimation of some common distributions to be fitted to multiple data sets within a homogeneous region. Application of regional frequency analysis to extreme waves has been reported by van Gelder et al. (2000) and Ma et al. (2008). Van Gelder et al. analyzed the wave data at nine stations off the Dutch coast, while Ma et al. examined the buoy data of the National Data Buoy Center at seven stations off the California coast of U.S.A. Waves off the Dutch coast were fitted with the Generalized Pareto (GPA) distribution, while waves off the California coast was fitted with the Pearson type III distribution.

Both studies did not include the Weibull distribution because Hosking (1990) did not derive the *L*-moments of the Weibull distribution and the formulas for parameter estimation. However, the Weibull is one of the most favored distributions in coastal and ocean engineering since Petruaskas and Aagaard (1971) have worked out the unbiased plotting position formulas. A number of extreme wave data sets have been fitted to the Weibull distribution. Its omission from the candidate distributions certainly causes discontinuity in the practice of extreme wave analysis. To prevent such discontinuity, derivation of the *L*-moments of Weibull distribution is made and the regional frequency analysis of extreme waves at the Japan Sea coast with large data sets is described in the present paper.

L-MOMENTS OF WEIBULL DISTRIBUTION AND PARAMETER ESTIMATE

The following functional form is given for the Weibull distribution to be employed in the present paper:

$$F(x) = 1 - \exp[-\{(x-B)/A\}^{k}, \qquad B \leq x < \infty \tag{1}$$

where $F(x)$ denotes the cumulative distribution of data x, and A, B, and k are the scale, location, and shape parameter, respectively. The peaks-over-threshold (POT) storm significant wave heights

---

[1] ECOH CORPORATION, 2-6-4 Kita-Ueno, Taito-Ku, Tokyo 110-0014, Japan.
[2] Port and Airport Research Institute, 3-1-1 Nagase, Yokosuka 239-0826, Japan

constitute the data x. The quantile of the Weibull distribution for a given non-exceedance probability F is derived as the inverse function of Eq. (1) as below.

$$x_F = B + A[-\ln(1-F)]^{1/k} \tag{2}$$

In the regional frequency analysis, the first to fourth L-moments denoted by λ_1 to λ_4 are derived from a given distribution function, and the mutual ratios of $\tau = \lambda_2/\lambda_1$, $\tau_3 = \lambda_3/\lambda_2$, and $\tau_4 = \lambda_4/\lambda_2$ are calculated. These ratios are called the L-CV (coefficient of L-variation), L-skewness, and L-kurtosis, respectively. Please refer to **Appendix A** for the definition of L-moments. The first L-moment represents the mean of the distribution function.

The L-moments and their ratios for the Weibull distribution are given by

$$\left.\begin{aligned} &\lambda_1 = B + A\Gamma(1+1/k), \quad \lambda_2 = A(1-2^{-1/k})\Gamma(1+1/k), \\ &\tau_3 = 3 - 2(1-3^{-1/k})/(1-2^{-1/k}), \\ &\tau_4 = \{5(1-4^{-1/k}) - 10(1-3^{-1/k}) + 6(1-2^{-1/k})\}/(1-2^{-1/k}) \end{aligned}\right\} \tag{3}$$

where $\Gamma(\cdot)$ denotes the Gamma function. Van Gelder (2000) lists a set of formulas of the L-moments for the Weibull distribution, but his L-skewness (τ_3) is given the sign opposite to Eq. (3) mistakenly. For a sample of extreme data, the L-moments are evaluated by algebraic calculations. The estimated values may deviate from Eq. (3) owing to sample variability. Hosking and Wallis (1997) use the symbols of t, t_3, and t_4 for the sample values for their differentiation from the population values of τ, τ_3, and τ_4. Nevertheless, the latter symbols are used in the present paper for the sake of simplicity.

Once the L-moment ratios are evaluated from a sample, the parameters of the Weibull distribution are directly estimated by the following formulas:

$$k = 285.3\,\lambda_3^6 - 658.6\,\lambda_3^5 + 622.8\,\lambda_3^4 - 317.2\,\lambda_3^3 + 98.52\,\lambda_3^2 - 21.256\,\lambda_3 + 3.516 \tag{4}$$

$$A = \frac{\lambda_2}{(1-2^{-1/k})\Gamma(1+1/k)}, \quad B = \lambda_1 - A\Gamma(1+1/k) \tag{5}$$

Equation (4) is empirically obtained by fitting a polynomial function to the relation between τ_3 and k; the fitting error is less than 0.3% for the range of $0.6 < k < 3.0$.

CHARACTERIZATION OF DISTRIBUTION FUNCTIONS

The present paper employs the Generalized Extreme Value (GEV), the Generalized Pareto (GPA), and the Weibull distributions as the candidates of data fitting. The GEV distribution is a synthesis of the Fisher-Tippett type I, II, and III distributions, while the GPA distribution is a non-asymptotic distribution function. These functions are well discussed in statistical textbooks such as Coles (2001). Their functional forms are given below.

GEV:
$$F(x) = \begin{cases} \exp[-\{1 - k(x-B)/A\}^{1/k}], & k > 0: \ -\infty \le x < B + A/k \\ & k < 0: \ B - A/k < x < \infty \\ \exp[-\exp\{-(x-B)/A\}], & k = 0: \ -\infty < x < \infty \end{cases} \tag{6}$$

GPA:
$$F(x) = \begin{cases} 1 - [1 - k(x-B)/A]^{1/k} & 1 \ge k > 0: \ B \le x < B + A/k \\ & k < 0: \ B \le x < \infty \\ 1 - \exp\{-(x-B)/A\}, & k = 0: \ B \le x < \infty \end{cases} \tag{7}$$

When the shape parameter k of GEV is positive, it becomes the FT-III and has the upper limit at $B + A/k$. When $k = 0$, it is the FT-I or Gumbel distribution. When $k < 0$, it becomes the FT-II or Frechet distribution and has the lower limit at $B - A/k$. The GPA distribution becomes the exponential distribution for $k = 0$. It has the lower limit at B and the upper limit at $B + A/k$ for $k > 0$. The L-moments and the formulas for parameter estimation are well described in the book by Hosking and Wallis (1997) and are not listed here.

These distribution functions are characterized with the mutual relationship between L-kurtosis and L-skewness as shown in Fig. 1, which includes the Pearson type III distribution. The GEV and GPA distributions are plotted with differentiations for $k > 0$, $k = 0$, and $k < 0$. Positions of the Weibull distribution with typical k-values are shown with square symbols, while positions of the GEV with

typical negative *k*-values are marked with diamond symbols. Because the behavior of the Pearson type III distribution is not much different from the Weibull or GPA distribution, it is not included in the candidate distributions for extreme wave data; it is rather unfamiliar in coastal engineering applications.
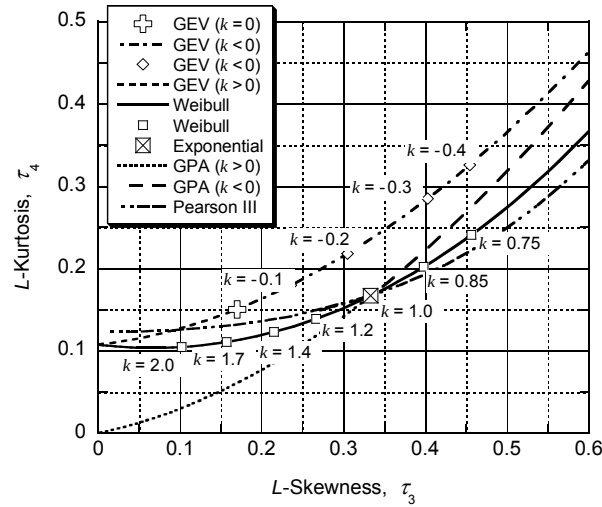

Figure 1. Relationship between *L*-kurtosis and *L*-skewness of several extreme distribution functions.

PEAKS-OVER-THRESHOLD WAVE DATA FOR ANALYSIS

The Port-administrating agencies under the Ministry of Land, Infrastructure, Transport and Tourism of Japan have been operating a nation-wide wave observation network called NOWPHAS (Nationwide Ocean Wave information network for Port and HAbourS) since 1970 (See Goda et al, 2000). Currently it has 61 seabed-mounted wave sensors and 11 GPS buoys, most of which are sending wave data to the Port and Airport Research Institute continuously for 24 hours daily.

The wave data are statistically and spectrally analyzed for the duration of 20 minutes every two hours. Some stations have the wave record for nearly 40 years. Figure 2 shows the locations of major NOWPHAS stations. The data at eleven stations along the eastern coast of Japan Sea are analyzed for the regional frequency analysis. Table 1 lists the station names, locations, and other characteristics. The threshold level H_c was selected to yield the mean occurrence rate in the range of 10 to 15 per year. Several levels of the threshold heights were tried, but changes in estimated return wave heights were slight.

All the POT data were normalized by being divided by the mean value of each data set. The first *L*-moment λ_1 and *L*-moment ratios of 11 station data have been calculated as listed in Table 2. The first *L*-moment λ_1 is calculated before normalization and is equal to the mean of the POT data. The discordancy measure D_i is calculated for six stations from Rumoi to Niigata and five stations from Wajima to Hamada separately. According to the criterion by Hosking and Wallis, the wave data at these stations are judged as homogeneous.

Table 1: Wave stations and major information of wave data analyzed in the present study.

Id	Name	Lat.	Long.	Depth h (m)	Starting year	Effec. duration (year)	Nos. of data	H_c (m)	$(H_{1/3})_{max}$ (m)
a	Rumoi	43°52′	141°28′	49.8	1970	37.2	554	3.5	7.83
b	Setana	43°15′	141°17′	52.9	1980	26.3	330	4.0	9.43
c	Fukaura	40°40′	139°55′	51.0	1979	28.0	408	4.0	10.36
d	Akita	39°44′	140°00′	29.4	1981	22.8	314	4.0	8.74
e	Sakata	39°01′	139°47′	45.9	1970	34.4	518	4.0	10.65
f	Niigata	38°00′	139°08′	34.5	1989	18.4	252	3.5	8.48
g	Wajima	37°25′	136°54′	52.0	1979	28.7	298	4.0	7.73
h	Kanazawa	36°37′	136°34′	21.1	1970	32.3	387	4.0	8.14
i	Fukui	36°10′	136°04′	36.7	1980	18.0	248	3.5	7.79
j	Tottori	35°33′	134°10′	30.9	1979	26.2	320	3.5	7.54
k	Hamada	34°54′	132°02′	50.1	1974	29.9	400	3.5	7.93

Table 2: *L*-moment ratios and discordance measure.

| Station name	Nos. of data, N	λ_1 (m)	L-CV	L-skew-ness	L-kurt-osis	D_i
Rumoi	554	4.440	0.0961	0.2484	0.1136	1.41
Setana	330	4.989	0.0921	0.2766	0.1316	0.97
Fukaura	408	5.097	0.1002	0.2792	0.1377	0.85
Akita	314	5.202	0.1060	0.2676	0.1252	0.23
Sakata	518	5.249	0.1093	0.2595	0.1258	0.93
Niigata	252	4.437	0.1054	0.2995	0.1242	1.61
Wajima	298	4.934	0.0819	0.2533	0.1207	0.64
Kanazawa	387	5.104	0.0956	0.2248	0.0933	0.83
Fukui	248	4.594	0.1111	0.2587	0.1149	1.08
Tottori	320	4.346	0.0892	0.2974	0.1612	1.19
Hamada	400	4.354	0.0912	0.2786	0.1259	1.27
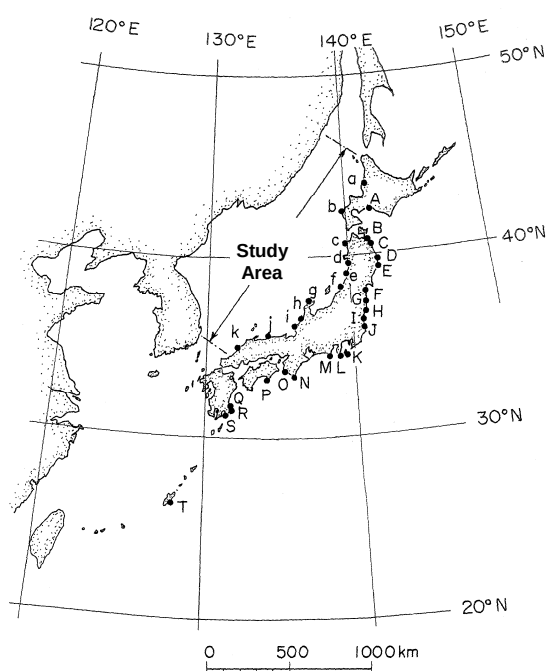
Figure 2. Location map of major NOWPHAS wave measurement stations.
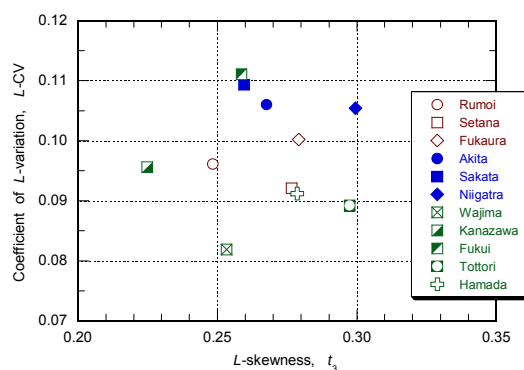
**Figure 3. *L*-CV vs. *L*-skewness**
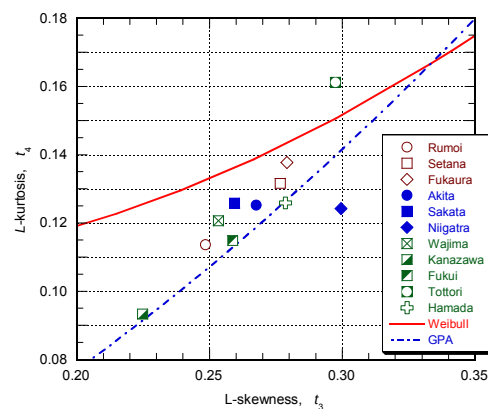
Figure 4. *L*-kurtosis vs. *L*-skewness

The procedure of regional frequency analysis is summarized in **Appendix B.** As demonstrated in Figs. 3 and 4, the *L*-moment ratios are spread in relatively narrow ranges. Theoretical curves of τ_4 versus τ_3 for the Weibull and GPA distributions are also shown in Fig. 4, suggesting the both distribution would fit to the wave data.

REGIONAL DIVISION AND HETEROGENUITY MEASURES

The regional frequency analysis requires homogeneity of the data within a region. Hosking and Wallis (1997) propose to check the homogeneity by using the following heterogeneity measure *H*:

$$H = \frac{(V - \mu_V)}{\sigma_V} \tag{8}$$

where *V* is the quantity defined by the following with μ_V and σ_V being the mean and standard deviation of *V* to be evaluated by numerical simulation, respectively:

$$V = \left\{ \frac{\sum_{n_i}^{N} n_i \left(t^{(i)} - t^R \right)^2}{\sum_{n_i}^{N} n_i} \right\}^{1/2} \tag{9}$$

where *N* is the number of stations, n_i is the number of data at the *i*-th station, $t^{(i)}$ is a quantity related to the *L*-moment ratio of the *i*-th station, and t^R is the regional mean of the quantity estimated by fitting a Kappa distribution having four parameters to the regional data.

The heterogeneity measure is checked with three quantities. The first measure *H*(1) concerns with the spread of *L*-CV value within a region. The second measure *H*(2) examines the distance of a data set from the center of gravity on the scatter diagram of *L*-CV and τ_3. The third measure *H*(3) calculates the distance of a data set from the center of gravity on the scatter diagram of τ_4 and τ_3. Hosking and Wallis (1997) states that the region can be judged as homogeneous when $H \le 1$ and may be heterogeneous if $H > 2$. Several regional divisions of 11 stations were tried as listed in Table 3. The three *L*-moment ratios are the regional weighted means with the weight of the number of data. According to the result of Table 3, the regions A, B, and C are regarded as heterogeneous as far as *H*(1) is concerned and the regions D, E, F are judged as homogeneous. The heterogeneity measures *H*(2) and *H*(3) remains below 2 for all the regions A to F.

Table 3: Regional mean *L*-moment ratios and heterogeneity measures.

Region	Stations	Nos.	*L*-CV	*L*-skew	*L*-kurt	*H* (1)	*H* (2)	*H* (3)
A	a to k	11	0.0980	0.2656	0.1243	5.27	0.46	-0.70
B	a to f	6	0.1014	0.2680	0.1256	2.34	-0.31	-1.25
C	g to k	5	0.0931	0.2621	0.1225	4.49	1.27	0.69
D	a, b, c	3	0.0964	0.2653	0.1258	0.27	0.06	-0.19
E	d, e, f	3	0.1074	0.2712	0.1252	-0.93	-0.30	-0.83
F	g, h, j, k	4	0.0900	0.2627	0.1238	1.35	1.43	1.27

FITTING OF REGIONAL DISTRIBUTION FUNCTIONS

Hosking and Wallis (1997) have proposed to judge the degree of goodness-of-fitting of a distribution by using the following quantity *Z*:
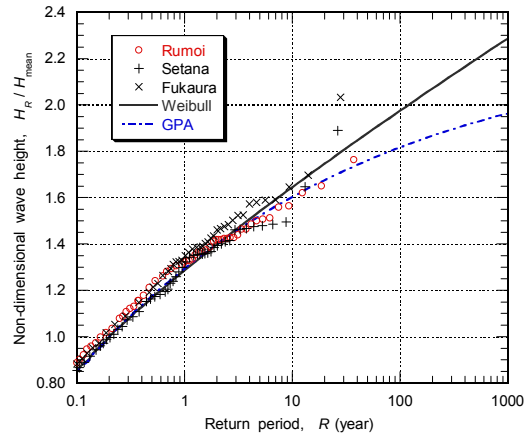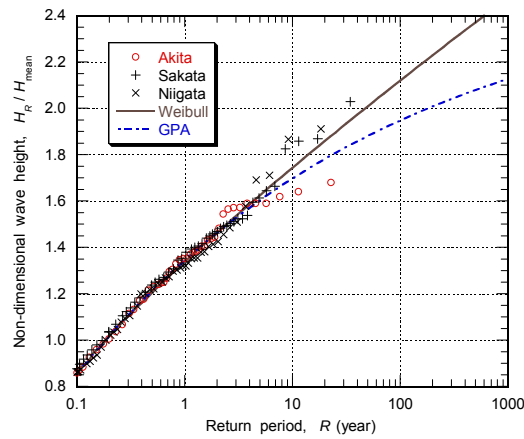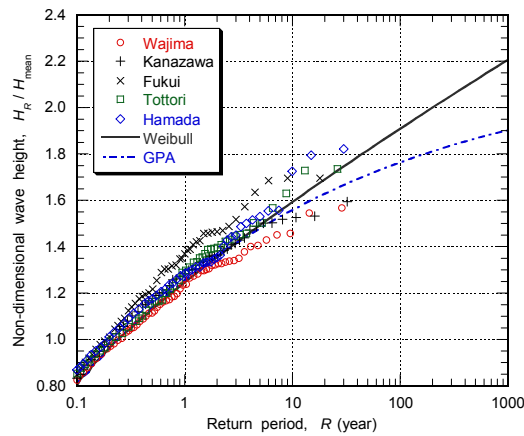
$$Z^{DIST} = (t_4^R - \tau_4^{DIST}) / \sigma_4 \tag{10}$$

where t_4^R is the regional mean of *L*-kurtosis, τ_4^{DIST} is the value of *L*-kurtosis estimated from the regional *L*-skewness for a particular distribution, and σ_4 is the regional standard deviation of *L*-kurtosis for the Kappa distribution estimated by numerical simulation. They have been given an acceptance criterion of $| Z^{DIST} | \le 1.64$ in consideration of the 90% confidence interval of the normal distribution.

The GEV distribution was found to have the *Z* value in excess of 6 for the regions A to F, and thus it was unaccepted as the regional distribution. The GPA distribution had the *Z* value from –0.37 to –1.12 and was accepted as the regional distribution. The Weibull distribution had the *Z* value of 1.32 to 1.61 for the regions C to F and was also accepted. Table 4 lists the parameter values of the regional Weibull and GPA distributions for the regions C to F.

Table 4: Parameter values of regional Weibull and GPA distribution.

| Region	Weibull distribution			GPA distribution		
	k	A	B	k	A	B
D	1.207	0.2348	0.7794	0.1612	0.2418	0.7917
E	1.187	0.2574	0.7571	0.1467	0.2644	0.7694
F	1.216	0.2209	0.7929	0.1678	0.2277	0.8050
C	1.219	0.2291	0.7854	0.1693	0.2364	0.7980

**Figure 5. Fitting of Weibull and GPA distribution functions for the region D.**

Figure 6. Fitting of Weibull and GPA distribution functions for the region E.


Figure 7. Fitting of Weibull and GPA distribution functions for the region C.

Figures 5 to 7 show the comparison of non-dimensional wave heights with the estimated return wave heights for the range of return period from 0.1 to 1000 years. The return period of the m-th descending order wave height is assigned with the non-exceedance probability of $F_m = 1 - m/n_i$ and $R_m = 1/[\lambda_i(1 - F_m)]$, where n_i is the number of data at the i-th station and λ_i is the mean rate, because the L-moments of the wave data have been computed using the unbiased plotting position as recommended by Hosking and Wallis (1997). The return period of the largest data is equal to the effective duration.

As seen in Figs. 5 to 7, the return wave heights estimated by the GPA distribution are smaller than those by the Weibull distribution for the range of return period longer than about 3 years. Individual wave data are scattered around the fitted distributions, but the fitted distributions are regarded as representing the population distribution at each region. The wave data at Fukui shown in Fig. 7 exhibit some deviation from other stations and the deviation is a cause of a high heterogeneity measure $H(1)$ in Table 3; the distributions fitted to other four stations in the region C are nearly the same as those shown in Fig. 7.

RETURN WAVE HEIGHTS OF INDIVIDUAL STATIONS

The distribution functions for individual stations are obtained by multiplying the non-dimensional scale and location parameters in Table 4 with the mean wave height of respective stations while retaining the same shape parameter. Examples of return wave height estimates together with observed heights are shown in Figs. 8 to 13.

The curves marked with "region" are the estimates of wave heights using the regional distribution, while those with "single" are the estimates using the distribution fitted to individual station data. In the case of Fukui in Fig. 12, individually fitted Weibull and GPA distributions differ much from the regional ones, reflecting the deviation of Fukui data from other stations shown in Fig. 7.
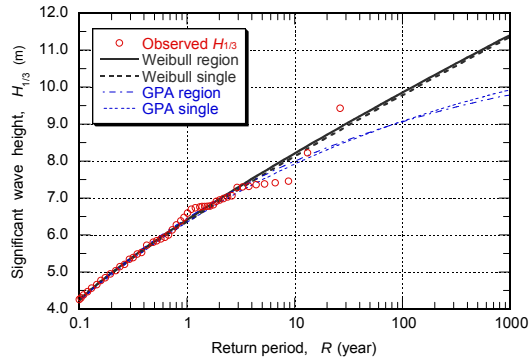

Figure 8. Return wave height of Setana
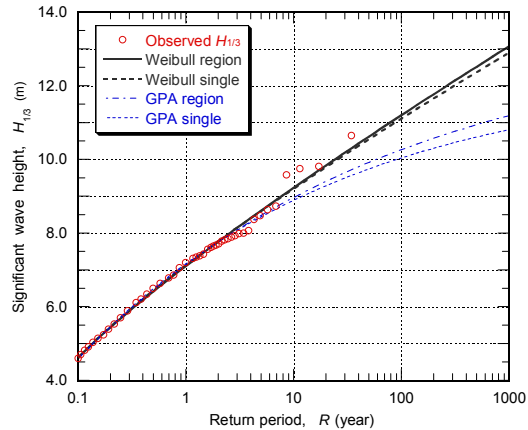

Figure 9. Return wave height of Sakata
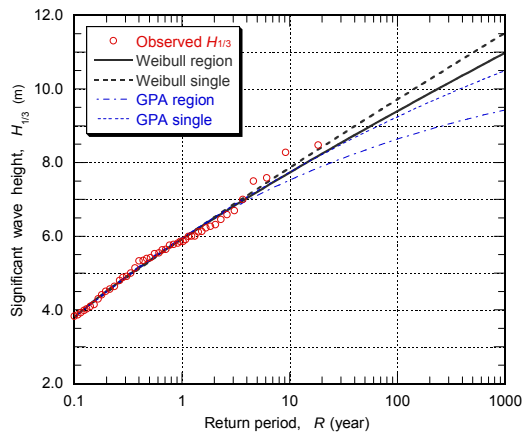
Figure 10. Return wave height of Niigata
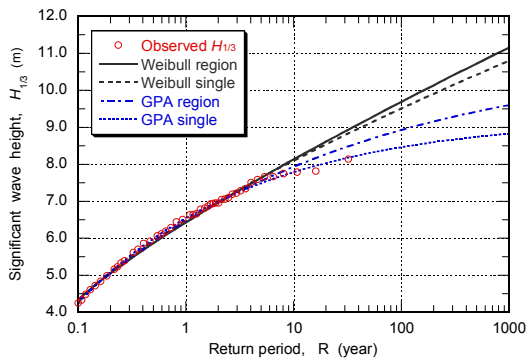
**Figure 11. Return wave height of Kanazaw**
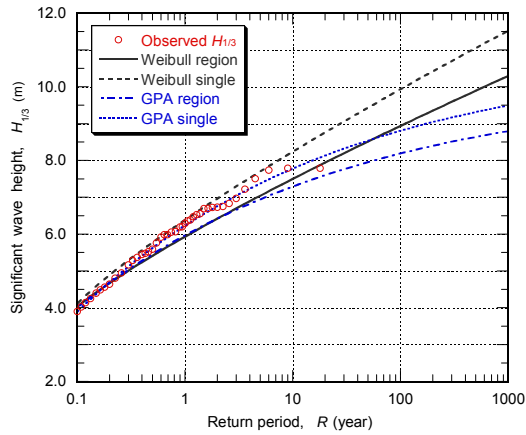
Figure 12. Return wave height of Fukui
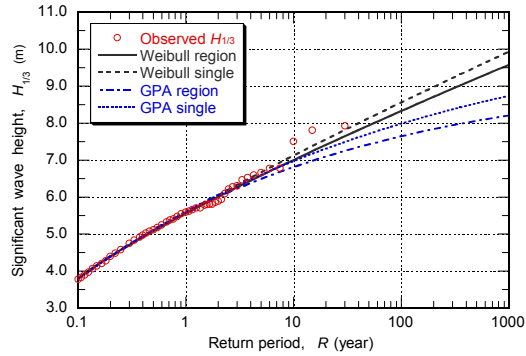
Figure 13. Return wave height of Hamada

Choice of the Weibull or the GPA distribution for the coastal stations along the Japan Sea requires some engineering judgment, because the both distributions fit well to observed wave data. However, the GPA distribution predicts a smaller return height for a long return period than the Weibull. Table 5 is a comparison of 100-year significant wave heights estimated with the Weibull and GPA distributions as well as the maximum significant wave height observed during 18 to 37 effective years (see **DISCUSSION** for LSQ H_{100}).

Table 5: Comparison of estimated 100-year wave height with maximum observed height.

Region	Station name	Weibull H_{100} (m)	GPA H_{100} (m)	$(H_{1/3})_{max}$ (m)	LSQ H_{100} (m)
D	Rumoi	8.88	8.12	7.83	8.66
D	Setana	9.86	9.06	9.43	9.52
D	Fukaura	10.17	9.32	10.36	10.24
E	Akita	11.03	10.13	8.74	10.10
E	Sakata	11.20	10.27	10.65	10.98
E	Niigata	9.40	8.64	8.48	9.07
F	Wajima	9.28	8.58	7.73	8.76
F	Kanazawa	9.69	8.93	8.14	9.67
C	Fukui	8.94	8.19	7.79	9.45
F	Tottori	8.26	7.61	7.54	8.17
F	Hamada	8.33	7.65	7.93	8.25

As exhibited in Table 5, the 100-year wave height estimated by the GPA distribution is less than the maximum observation height at four stations. It would not be wise to employ the GPA distribution for the viewpoint of safe structural design.

DISCUSSIONS

Previously, Goda et al. (2000) made another approach in search of regional population distributions of extreme wave heights around Japan with the least squares method. The rejection criteria called REC and DOL were applied to POT wave data at multiple stations in a region, and the distribution function least rejected was recommended as the population distribution. For the eleven stations around the Japan Sea coast, same as the present ones, the Weibull distribution with the shape parameter of $k = 1.4$ was judged to represent the population. The data duration was 10 years less than the present ones.

The estimated 100-year wave heights by the least squares method (LSQ) with the present data set are listed at the rightmost column of Table 5. They are generally smaller than the Weibull estimates with $k \approx 1.2$ but greater than the GPA estimate. Smaller prediction by LSQ may be caused by fitting of a larger shape parameter of $k = 1.4$, but occurrence of large storm waves in the last ten years must have contributed to the increase of 100-year wave heights in the present analysis.

In the present paper, confidence intervals for the estimated return wave heights have not been worked out yet. Hosking and Wallis (1997) recommend use of the Monte Carlo simulation technique for estimation of the confidence interval once a best-fitting distribution function is selected. This is one of the tasks which are intended to be pursued in the near future by the authors.

CONCLUSIONS

Major findings of the present study are summarized below.

1) Long wave observation records at the eastern coast of Japan Sea enabled a detailed examination of extreme statistics by means of the regional frequency analysis method with the sample size of 248 to 554 for the effective duration of 18.0 to 37.2 years.
2) The *L*-moments of the Weibull distribution and parameter estimation formulas have been derived and successfully applied to the POT data.
3) Eleven stations around the east coast of Japan Sea were grouped into three homogeneous regions.
4) Characteristics of regional distributions were established and the shape, scale, and location parameters were estimated for three regions.
5) Return wave heights at eleven stations were estimated by converting the regional, normalized distribution function into the dimensional distribution functions at respective stations.
6) Both the Weibull and GPA distributions fitted well to the wave data analyzed, but the latter is not recommended for engineering applications because of its tendency to predict a low wave height for a long return period.

APPENDIX A: Definition of *L*-moments and *L*-moment Ratios

For a random variate *X*, a cumulative distribution function is defined as the probability that *X* is equal or less than some specified value *x*, *i.e.*

$$F(x) = \Pr[X \leq x] \tag{A.1}$$

On the other hand, the value *x* for a given probability *P* is called as the quantile function *x(P)*, *i.e.*

$$F[x(P)] = P, \qquad x = F^{-1}(P) \tag{A.2}$$

The probability-weighted moment β_r is defined with the quantile function as follows:

$$\beta_r = \int_0^1 x(P) P^r dP \tag{A.3}$$

Hosking (1990) defined the *L*-moments using the moment β_r as in the following:

$$\lambda_1 = \beta_0 \tag{A.4}$$

$$\lambda_2 = 2\beta_1 - \beta_0 \tag{A.5}$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \alpha_0 \tag{A.6}$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \tag{A.7}$$

He further defined the *L*-moment ratios as below.

L-CV (coefficient of L-variation): $\qquad \tau = \lambda_2 / \lambda_1 \tag{A.8}$

L-skewness: $\qquad \tau_3 = \lambda_3 / \lambda_2 \tag{A.9}$

L-kurtosis: $\qquad \tau_4 = \lambda_4 / \lambda_2 \tag{A.10}$

For a sample of extreme data of the size *n*, the data are first sorted in the ascending order of $x_1 \leq x_2 \leq \ldots \leq x_n$. Then the unbiased estimator of β_r is calculated with the following formulas:

$$b_0 = \frac{1}{n} \sum_{j=1}^{n} x_j \tag{A.11}$$

$$b_1 = \frac{1}{n} \sum_{j=2}^{n} \frac{(j-1)}{(n-1)} x_j \qquad (A.12)$$

$$b_2 = \frac{1}{n} \sum_{j=3}^{n} \frac{(j-1)(j-2)}{(n-1)(n-2)} x_j \qquad (A.13)$$

and in general,

$$b_r = \frac{1}{n} \sum_{j=r}^{n} \frac{(j-1)(j-2)\ldots(j-r)}{(n-1)(n-2)\ldots(n-r)} x_j \qquad (A.14)$$

The sample *L*-moments are calculated with the above estimators as follows:

$$\ell_1 = b_0 \qquad (A.15)$$

$$\ell_2 = 2b_1 - b_0 \qquad (A.16)$$

$$\ell_3 = 6b_2 - 6b_1 + b_0 \qquad (A.17)$$

$$\ell_4 = 20b_3 - 30b_2 + 12b_1 - b_0 \qquad (A.18)$$

The sample *L*-moment ratios are given by

Sample *L*-CV (coefficient of L-variation): $t = l_2 / l_1$ (A.19)

Sample *L*-skewness: $t_3 = l_3 / l_2$ (A.20)

Sample *L*-kurtosis: $t_4 = b_4 / b_2$ (A.21)

An example of calculation of sample *L*-moments is given in Table A.1 for annual maximum wind speeds for ten years cited in Hosking and Wallis (1997). The first column is the order number, the second column is the annual maximum wind speed, and the third column is the sorted wind speed in the ascending order.

Table A.1. Example of calculation of sample *L*-moments.

j	x	$b_0 = x_j$	b_1	b_2	b_3
1	45	37			
2	47	42	4.666667		
3	65	44	9.777778	1.222222	
4	50	45	15.00000	3.750000	0.535714
5	56	47	20.88889	7.833333	2.238095
6	55	50	27.77778	13.88889	5.952381
7	37	53	35.33333	22.08333	12.61905
8	53	55	42.77778	32.08333	22.91667
9	44	56	49.77778	43.55556	37.33333
10	42	65	65.00000	65.00000	65.00000
		49.4	27.1	18.94167	14.65952

From the calculation result in Table A.1, the sample *L*-moments and their ratios are obtained as follows:

$$\ell_1 = 49.4, \quad \ell_2 = 4.8, \quad \ell_3 = 0.450, \quad \ell_4 = 0.7405,$$
$$t = 0.0972, \quad t_3 = 0.0938, \quad t_4 = 0.1543$$

If the sample is fitted to the Weibull distribution, the sample *L*-skewness of $t_3 = 0.0938$ yields an estimate of $k = 2.17$ by means of the empirical formulas of Eq. (4) and $A = 19.8$ and $B = 31.8$ by Eq. (5).

APPENDIX B: Procedure of Regional Frequency Analysis with the *L*-moments Method

Suppose that extreme statistical data such as POT wave heights are obtained at *N* stations with the sample size being n_i for the *i*-th station. According to Hosking and Wallis (1997), the procedure of regional frequency analysis is as follows:

1) Normalize the wave data of each station by dividing them by its mean value so that the data become non-dimensional.
2) Calculate the *L*-moments and of *L*-moment ratios of normalized data at each station. The first *L*-moment l_1 is 1 because of normalization.
3) Group the stations into a number of regions by means of the site characteristics such as geographical features.
4) Calculate discordancy measure D_i for the stations within a region, and check if no station is discordant with other stations.
5) Calculate the heterogeneity measure by Eqs. (8) and (9), and check if each region is homogeneous.
6) Calculate the regional *L*-moment ratios for each region as the weighted means using the sample size as the weight.
7) Estimate the parameters of several candidate distributions using the regional *L*-moment ratios.
8) Select the best-fitting regional distribution function by referring to the value of Z^{DIST} by Eq. (10).
9) Construct the distribution function of each station by using the dimensional scale and location parameters which are obtained as the product of regional non-dimensional parameters and the mean value of each station. The shape parameter *k* is set as equal to the regional shape parameter.
10) Estimate the return extreme value corresponding to a given return period in combination of the mean occurrence rate.

REFERENCES

Coles, S. 2001. *An Introduction to Statistical Modeling of Extreme Values*, Springer, 208p.

Goda, Y., Konagaya, O., Takeshita, N., Hitomi, H., and T. Nagai. 2000. Population distribution of extreme wave heights estimated through regional analysis, *Coastal Engineering 2000* (*Proc. 26th ICCE*, Sydney), ASCE, Sydney, 1078-1091.

Greenwood, J A., J. M. Landwehr, N. C. Matalas, and J. R. Wallis. 1978. Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form, *Water Resources Res.*, Vol. 15, No. 5, pp. 1049-1064.

Hosking, J. R. M. 1990. *L*-moments: Analysis and estimation of distributions using linear combinations of order statistics, *J. Roy. Statistical Soc.*, *Series B*, **52**, pp. 105-24.

Hosking, J. R. M. and J. R. Wallis. 1997. *Regional Frequency Analysis*, Cambridge Univ. Press, 224p.

Ma, Q.-S., Li, Y.-B., and J. Li .2006. Regional frequency analysis of siginicant wave heights based on *L*-moments, *China Ocean Engineering*, **20**(1), pp. 85-98.

Petruaskas, C. and P. M. Aagaard. 1971. Extrapolation of historical storm data for estimating design wave heights, *J. Soc. Petroleum Engrg.*, **11**, pp. 23-27.

van Gelder, P. H. A. J. M. 2000. *Statistical Methods for the Risk-Based Design of Civil Structures*, Ph.D. thesis Delft University of Technology, 249p.

van Gelder, P. H. A. J. M., J. De Ronde, N. W. Neykov, and P. Neytchev. 2000. Regional frequency analysis of extreme wave heights: trading space for time, *Coastal Engineering 2000* (*Proc. 26th ICCE*, Sydney), ASCE, pp. 1099-1112.