Estimating probability of extreme rainfall over Japan using Extended Regional Frequency Analysis

Masato Sugi¹, Yukiko Imada¹, Toshiyuki Nakaegawa¹ and Kenji Kamiguchi²

¹Meteorological Research Institute, Japan
²Japan Meteorological Agency, Japan

Abstract:

A method of frequency analysis, Extended Regional Frequency Analysis (ERFA), is proposed for reliable estimates of extreme daily rainfall probabilities for a long return period from relatively short daily rainfall records. The method uses combined data in a wide meteorologically homogeneous region (e.g., all Japan) to ensure a large number (order of 10,000) of data to minimize the effects of statistical sampling error in the frequency analysis. We applied the ERFA to daily rainfall data observed over Japan and to a high resolution atmospheric model simulation data over the meteorologically homogeneous land region of Japan. We found very good agreement between the empirical probability distribution and theoretical distribution estimated by ERFA, suggesting that the method is promising. However, we have noted some problems regarding ERFA: selection of the distribution, selection of the region, and model bias. These problems, along with possible solutions, are discussed.

KEYWORDS regional frequency analysis; extreme value analysis; extreme rainfall; return period; return value

INTRODUCTION

There is a practical requirement for reliable estimates of the probability of extreme rainfall with a long return period (longer than 100 years) from relatively short (20–30 years) rainfall records. The number of observation sites having long records (longer than 100 years) are limited. Furthermore, considering climate change, statistical stationarity could be questionable in some long record. When we estimate the future probability of extreme rainfall from climate model simulations, we usually need to make estimates using a relatively short (20–30 years) record of rainfall data from the model simulations.

To estimate reliable extreme daily rainfall probabilities with a long return period from relatively short daily rainfall records, Regional Frequency Analysis (RFA, Hosking and Wallis, 1997) is often used. Toyama and Mizuno (2002) applied RFA to a 20 year record of daily rainfall data from 860 observation sites in the Automated Meteorological Data Acquisition System (AMeDAS), a rain gauge network over Japan, and estimated the daily rainfall with a return period of 2 to 1000 years at each observation site. Ishihara (2010) also applied RFA to a 21 year record of daily rainfall data from 1062 observation sites in the AMeDAS, and estimated 30 year return values of daily precipitation. He compared the results with the conventional estimates by frequency analysis at each site, and reported that the RFA estimate was more reliable than the conventional estimate. Chikamori and Nagai (2013) applied RFA to a 30 year record of daily rainfall data from 155 Japan Meteorological Agency observation sites and estimated 100 year return values of daily rainfall at each observation site. They also concluded that the RFA estimates were more reliable than conventional estimates obtained by frequency analysis at each site. In these RFA analyses, the Japan region was divided into 49, 61 and 8 regions, respectively, and the number of observation sites in one region ranged from 17 to 19 on average, with some regions having only a few sites. In some regions, the number of combined data may not be large enough to reliably estimate the probability of extreme daily rainfall with a long return period.

Koutsoyiannis (2004) reported that a reliable probability of extreme daily rainfall can be estimated by applying frequency analysis to combined data in a very large region. He performed frequency analysis for the combined 100–154 year daily rainfall data from 169 stations in the United States and Europe (the total number of the combined data was 18,065 station-year). He found a remarkable agreement between the empirical probability distribution and the theoretical distribution estimated by the frequency analysis. His results suggest that it is important to minimize the effects of sampling error by using a sufficiently large number (order of 10,000) of data for a reliable estimation of the probability of extreme rainfall.

Another method of frequency analysis called Station Year method (SY method, Linsley et al., 1949) also uses a large number of combined data in a wide region. Suzuki and Kikuchihara (1984) applied SY method to the annual maximum daily rainfall data at 137 stations in Japan (60 year record at each station on average and total 8,218 station-year combined data) to estimate the probability of extreme daily rainfall. They claimed that they could successfully estimate the extreme rainfall amount for return period up to 1,000 years. However, the estimated rainfall amount is likely to have a large sampling error, because the SY method esti-
mates the probability distribution at each station using the parameter estimated by the data at each station (see Text S1 for further detail). In this paper, we propose a method of frequency analysis to reliably estimate extreme daily rainfall probabilities with a long return period from relatively short daily rainfall records. In the proposed method, Extended Regional Frequency Analysis (ERFA), we perform frequency analysis for combined data in a much wider region than the conventional RFA to ensure a sufficiently large number of data for the frequency analysis. Koutrouviannis (2004) applied frequency analysis to the data normalized based on the sample mean of the annual maximum data at each station as in conventional RFA. Similarly, in ERFA we perform frequency analysis for the data normalized based on the sample mean of the annual maximum data at each site in the region. In RFA, a statistically homogeneous region is selected, so that all the data in the region follow a common probability distribution. However, the homogeneous region is selected based on the sample statistics at each station, and therefore, the results of the frequency analysis are considerably affected by sampling errors. On the other hand, in ERFA, to minimize the effects of sampling errors on the results, the frequency analysis is performed for a large region including a large number (order of 10,000) of data. A large, meteorologically homogeneous region, which is considered to be statistically homogeneous except for sampling error effects, was selected for the frequency analysis. The meteorologically homogeneity of a region can be determined based on the spatially smoothed sample statistics parameter fields (see Text S2 for the further detail of the meteorologically homogeneous region). It is expected that all data in a meteorologically homogeneous region will follow a common probability distribution as in RFA. In the present study, we apply ERFA to the daily rainfall data over Japan to examine the validity of the method.

METHOD AND DATA

The proposed method, ERFA, is the same as the conventional RFA (Hosking and Wallis, 1997) except for the following two points:

i) In RFA, before performing the frequency analysis, the region is divided into small statistically homogeneous sub-regions based on the sample statistics at each station. In ERFA, the frequency analysis is performed on a relatively wide meteorologically homogeneous region based on the smoothed fields of sample statistics to ensure a sufficiently large number of data for the frequency analysis to minimize the effects of sampling error.

ii) In RFA, the best fit distribution is selected for each sub-region from five candidate distributions. In ERFA, the best fit distribution is selected for the relatively wide region from the two theoretical distributions: Generalized Extreme Value (GEV) distribution (including Gumbel distribution as a special case) or Generalized Pareto (GPA) distribution (for peak over threshold data). The parameters of the theoretical distributions are calculated by L-moments.

In ERFA, as in RFA, the data in a region is normalized by the sample mean of the data at each site. We can expect the normalized data in a region to have a common distribution as in RFA, if the meteorologically homogeneous region is statistically homogeneous except for the effects of sampling errors. Then, as in RFA, the return value (quantile) of daily rainfall $R(T)$ for return period $T$ at site $i$ is written as,

$$R(T) = \mu_T r(T)$$

where $\mu_T$ is the sample mean of the annual maximum daily rainfall at site $i$ and $r(T)$ is the common return value for return period $T$ in the region. Note that the spatial distribution of the return value is solely represented by $\mu_T$, while the probability distribution is solely represented by $r(T)$.

In the present study, to demonstrate the validity and usefulness of ERFA, we first applied ERFA to the 28 year annual maximum daily rainfall data observed over Japan from 1979–2006 at 1001 stations in AMeDAS. The number of combined observations for the Japan region was 27,749 station-year (1001 stations × 28 years minus number of missing data). Then, we also applied ERFA to the 25 year annual maximum daily rainfall data from a present climate simulation (SPA, Super-highresolution Present climate simulation with model A) and a future climate simulation (SFA, Super-highresolution Future climate simulation with model A) using a 20-km mesh global atmospheric model, MRI-AGCM3.2 (Mizuta et al., 2012). The SPA experiment uses observed sea surface temperature (SST) and greenhouse gas and aerosol concentration for 1979–2003, while the SFA experiment uses projected SST and greenhouse gases and aerosol based on the IPCC RCP8.5 scenario for 2075–2099 (Mizuta et al., 2014). The number of combined simulation data was 26,750 grid-year (1070 grids × 25 years) for the Japan land region and 195,650 grid-year (91 × 86 grids × 25 years) for the Japan region including surrounding oceans (129.2°E–146.1°E, 29.9°N–45.8°N).

RESULTS

Some preliminary results on the application of ERFA to the annual maximum rainfall data over the land region of Japan, which was found to be a meteorologically homogeneous region, are shown in Figure 1. The method for finding the meteorologically homogeneous region is discussed in the next section. Here we show the results with a GPA distribution. We calculated GPA parameters using the upper 10% of all the normalized annual maximum rainfall data in the region to find the theoretical distribution that best fits to the empirical distribution in the range of our interest ($1000 > T > 10$). We consider the upper 10% data as an approximate peak over threshold data with the value at $T = 10$ as threshold. The selection of the best fit distribution is discussed in the next section. Figures 1a, 1c and 1e show the probability distribution of normalized daily rainfall for the AMeDAS observation data and the SPA and SFA simulation data, respectively. The empirical probability distribution (Cunnane plot) is shown by black dots and the GPA distribution estimated by ERFA is shown by red dots. We can see that the GPA distributions agree very well with the empirical distributions for all these data for the return period ranging from 10 years to 1000 years.

Figures 1b, 1d and 1f show the geographical distributions of the 100 year return values $R(100)$ of daily rainfall for the AMeDAS observation data, the SPA and SFA simulation data, respectively. In Figure 1b, the mean of the return val-
Selection of probability distribution

In Figure 1, we selected GPA as the best fit distribution for the normalized daily rainfall data over Japan. In Figure 2, we show how Gumbel, GEV and GPA distributions fit the empirical probability distribution for AMeDAS, SPA and SFA data. Generally, the Gumbel distribution significantly overestimates the return period for large normalized rainfalls, while both the GEV and GPA agree very well with the empirical distribution. For the SPA Japan region including the surrounding ocean regions (Figures 2h and 2i) and the SFA Japan land region (Figures 2k and 2l), GPA agrees a little better than GEV for data with large return periods. This is probably because we calculated parameters of GPA using only the upper 10% of all normalized rainfall data in the region to find the theoretical distribution that best fits to the empirical distribution in the range of our interest ($1000 > T > 10$).

Selection of region

It is important to note that the return values for a large return period are considerably different between the SPA...
Japan land region (Figures 2e and 2f) and SPA Japan region including the oceans (Figures 2h and 2i). The 100 year return value is 2.3 for SPA Japan land region, while it is 2.56 for SPA Japan region. This indicates that the probability distribution depends on the selection of region. These two regions seem to be statistically different. Figures 3a and 3b show the spatial distribution of L-CV and L-skewness calculated at each grid point of the SPA Japan region. We can see small scale random variations both in L-CV and L-skewness due to statistical sampling effects. In addition to these random variations, we can see some systematic differences over the land region and surrounding ocean region; both L-CV and L-skewness tend to be smaller over the land. This difference seems to be due to the difference in the meteorological characteristics of the land and ocean regions. Because of this systematic difference in L-CV and L-skewness between the two regions, the regional mean L-CV and L-skewness are different, and the probability distributions are different for the land region and the entire Japan region.

In RFA, frequency analysis is performed over the statistically homogeneous region based on the sample statistics at each station. If we select a homogeneous region depending on small scale random variation of parameters due to statistical sampling effect, the results of the frequency analysis also depend on the statistical sampling effect. To avoid the influence of the sampling effect, in ERFA we select a meteorologically homogeneous region depending on large scale variations of L-CV and L-skewness. A simple way to find such meteorologically homogeneous region is to remove the small scale random variations, which are considered to be sampling errors, in L-CV and L-skewness by spatial smoothing. Figures 3c and 3d show the L-CV and L-skewness smoothed by 600 km × 600 km running mean. The figures indicate that the land region of Japan is meteorologically homogeneous and appropriate for ERFA, while the whole Japan region including the surrounding oceans is not meteorologically homogeneous and not appropriate for ERFA.

It should be noted that the range of variation in the parameters over the land region is significantly reduced by spatial smoothing. They are 0.21 (0.11 to 0.32) for L-CV, 0.04 (0.212 to 0.252) for smoothed L-CV, 0.75 (−0.15 to 0.6) for L-skewness and 0.06 (0.192 to 0.253) for smoothed L-skewness. This suggests that the range of possible sampling error (uncertainty) in the estimation of the parameters is significantly reduced by smoothing, that is, the sampling errors are effectively removed by smoothing.

Figures 4a and 4b show the global distribution of L-CV and L-skewness for SPA data. Both L-CV and L-skewness are significantly larger over the subtropics and the eastern equatorial Pacific than the other regions. In these regions, the cause of the extreme rainfall seems to be different from

![Figure 3](image1.png)

**Figure 3.** Geographical distributions of (a) L-CV and (b) L-skewness calculated by annual maximum daily rainfall data at each grid point in SPA (present climate simulation) Japan region. (c) and (d) are same as (a) and (b) but smoothed by 600 km × 600 km running mean

![Figure 4](image2.png)

**Figure 4.** Same as Figure 3 but for global domain
the other regions. Figures 4c and 4d show the L-CV and L-skewness smoothed by 600 km × 600 km running mean. We can find a meteorologically homogeneous region by using these smoothed L-CV and L-skewness fields.

Model bias

We have noted that the spatial pattern of 100 year return values for AMeDAS (Figure 1b) and SPA Japan land region (Figure 1d) agree well, but there are some differences in detail between the two patterns. These differences are probably due to both model bias and statistical sampling effects. Because of the sampling effects, estimating bias at each grid point is not appropriate. The bias should be estimated based on the probability distribution of the sample mean of the annual maximum daily rainfall $\mu_i$. It is also not appropriate to remove the random pattern due to sampling effects by spatial smoothing, because the smoothing removes the topographic effect on the mean rainfall as well. For model data, the best way to remove the sampling effects is to take ensemble average using the ensemble simulation data. This method is not possible for the observation data, because we do not have ensemble observation data. Probably, the best estimate of the ensemble mean observation data is the bias corrected model ensemble mean data.

SUMMARY

We propose a method of frequency analysis, Extended Regional Frequency Analysis (ERFA), to reliably estimate extreme rainfall probabilities using a relatively short period of observed or simulated rainfall data. In ERFA, we perform frequency analysis in a wide region to ensure a large number of data to estimate probabilities of extreme rainfall with return period 100 years or longer.

We applied ERFA to some observed and simulated daily rainfall data over the land region of Japan that was found to be meteorologically homogeneous. The results of the frequency analysis showed that the estimated theoretical distributions agreed very well with the empirical distributions. It was found that the GPA distributions agreed best in the range of our interest ($1000 > T > 10$) when the parameters were estimated using the upper 10% of all normalized data for the region. It is suggested that a large meteorologically homogeneous region should be selected for ERFA, rather than a small statistically homogeneous region. We can find meteorologically homogeneous regions by spatial smoothing or ensemble averaging of L-CV and L-skewness fields, although we have not yet established the best way of identifying meteorologically homogeneous regions. The land region of Japan was found to be a meteorologically homogeneous region for extreme daily rainfall and appropriate for ERFA.

We have noted that there are small variations in the smoothed parameters (L-CV and L-skewness) in the meteorologically homogeneous land region of Japan. It is an important subject of future work to find the best estimate of the probability distribution at each location in a region considering small variations of the smoothed parameters.

In ERFA, the rainfall data is normalized by the sample mean of the annual maximum daily rainfall data at each station. We should note that the sample mean of the annual maximum daily rainfall data at each station may have a considerable sampling error. To find some better parameters for normalization which is less sensitive to sampling error is also an important subject of future work.

ACKNOWLEDGEMENT

This study was supported by Theme C of the SOUSEI program funded by the Ministry of Education, Culture, Sports, Science and Technology of Japan.

SUPPLEMENTS

Text S1. Station Year method
Text S2. Meteorologically homogeneous region
Text S3. Future changes in extreme rainfall region
Figure S1. Probability distribution of extreme rainfall over Japan estimated by Station Year method
Figure S2. Same as Figure S1 but for the six subregions of Japan
Figure S3. Future changes in 100 year return values of extreme rainfall over Japan
Table S1. Return values of normalized rainfall over Japan

REFERENCES

Chikamori H, Hagai A. 2013. Regional frequency analysis of daily rainfall observed at meteorological observatories in Japan. Transactions of The Japanese Society of Irrigation, Drainage and Rural Engineering 81: 439–451. DOI: 10.11408/jsirde.81.439 (in Japanese).

Hosking JRM, Wallis J. 1997. Regional Frequency Analysis: An approach based on L-moments. Cambridge University Press, Cambridge, UK; 224.

Ishihara K. 2010. Estimation of probable daily precipitation at AMeDAS stations by applying regional frequency analysis to each prefecture. Journal of Japan Society of Hydrology and Water Resources 23: 118–128. DOI: 10.3178/jjshwr.23.118 (in Japanese).

Koutsoyiannis D. 2004. Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. Hydrological Sciences Journal 49: 591–610. DOI: 10.1623/hysj.49.4.591.54424.

Linsley RK, Kohler MA, Paulhus JLH. 1949. Applied hydrology. McGraw-Hill, New York, USA; 689.

Mizuta R, Yoshimura H, Murakami H, Matsueda M, Endo H, Ose T, Kamiguchi K, Hosaka M, Sugii M, Yukimoto S, Kusunoki S, Kito A. 2012. Climate simulations using MRI-AGCM3.2 with 20-km grid. Journal of Meteorological Society of Japan 90A: 233–258. DOI: 10.2151/jmsj.2012-A12.

Mizuta R, Arakawa O, Ose T, Kusunoki S, Endo H, Kito A. 2014. Classification of CMIP5 future climate responses by the tropical sea surface temperature changes. Scientific Online Letters on the Atmosphere 10: 167–171. DOI: 10.2151/sola.2014-035.

Suzuki A, Kikuchi H. 1984. Method for estimating the return period of annual maximum daily precipitation considering about extreme values. Tenki 31: 179–189 (in Japanese).

Toyama N, Mizuno H. 2002. Daily and 1-hour precipitation for meteorologically homogeneous regions. Transactions of The Japanese Society of Irrigation, Drainage and Rural Engineering 81: 439–451. DOI: 10.11408/jsirde.81.439 (in Japanese).