PROJECTION FOR FUTURE CHANGE OF CONFIDENCE INTERVAL AND PREDICTION INTERVAL OF EXTREME RAINFALLS BASED ON BAYESIAN METHOD USING A LARGE ENSEMBLE DATA SET

KEITA SHIMIZU
Graduate School of Science and Engineering, Chuo University, 1-13-27, Kasuga, Bunkyo-ku, Tokyo, 112-8851, Japan
a13.8nx5@g.chuo-u.ac.jp

TADASHI YAMADA
Faculty of Science and Engineering, Chuo University, 1-13-27, Kasuga, Bunkyo-ku, Tokyo, 112-8851, Japan
yamada@civil.chuo-u.ac.jp

TOMOHITO YAMADA
Faculty of Engineering, Hokkaido University, N13-W8, Kita-ku, Sapporo, Hokkaido 060-8628, Japan
tomohito@eng.hokudai.ac.jp

ABSTRACT

The main purpose of hydrological frequency analysis is to estimate hydrological quantity with design return period. In traditional hydrological frequency analysis, we estimate *T*-year hydrological quantity by using observed data which were accumulated for several decades. However, observed data of extremes which were accumulated so far are so limited ones that the design hydrological quantity includes uncertainty to a large extent. Moreover, one of the difficulties caused by the shortage of extreme value data is not to predict record heavy rainfall which deviate greatly from an adopted probability distribution used for river planning. So, it is often impossible to evaluate record heavy rainfalls and these rainfalls are treated as unexpected. Therefore, we propose a new hydrological frequency analysis introducing confidence interval and prediction interval based on probability limit method test. By introducing this confidence interval into hydrological frequency analysis, uncertainty of design hydrological quantity can be quantified. Also, by introducing this prediction interval, it can be possible to predict the scale and occurrence risk of catastrophic heavy rainfalls. In addition, another difficulty of hydrological frequency analysis is how to manage non-stationarity of rainfall which is caused by climate change. To manage non-stationarity of rainfall, using Bayesian statistics is effective. In this research, Bayesian update of confidence interval and prediction interval derived from past observed data was conducted by using a large ensemble database. In this research, a theoretical framework of hydrological frequency analysis introducing confidence interval and prediction interval and update method of these intervals using Bayesian statistics are shown.

Keywords: Probability limit method test, Confidence interval, Prediction interval, Future projection database, Climate change, Uncertainty

1. INTRODUCTION

In recent years, the flood risk evaluation method using a large ensemble database was proposed to manage the intensification of heavy rainfall. A large ensemble database has super lots of calculated results for past and future climate. This database can be interpreted as rainfall data which we could have experienced in the past and will experience in the future. In traditional flood control management, the design rainfall is decided by using an only sample of observed data which exists in our history. On the other hand, by using this database, uncertainty of design rainfall (the range that adopted probability distribution can take) can be estimated as a form of confidence interval. In traditional hydrological frequency analysis, confidence interval was often expressed by numerical methods such as Jack knife method or bootstrap method which are based on an assumption that probable rainfall follows parametric probability distributions. Many of these methods use an assumption of normality based on central limit theorem to estimated statistics. However, it can be difficult to treat a distribution of probable rainfall as normal distribution. Because, it is also difficult to assume that normal distribution to each statistic such as probable rainfall under limited extreme rainfall data we can use in the present. Considering these situations, we constructed a new hydrological frequency analysis based on confidence interval which does not need any assumption as possible by using probability limit method test (Moriguti, 1995). Also, there is another concept of prediction interval besides confidence interval. Prediction interval is defined as range which observed data in future time exists. Considering this definition, prediction interval includes distribution of random variable which expresses observed data in future time with a given confidence coefficient.

By introducing confidence interval and prediction interval, the risk in flood control management can be expressed due to discuss how far rainfall within confidence interval and prediction interval is considered. However, this analysis based on confidence interval and prediction interval still uses an assumption of stationarity, so that cannot consider non-stationarity caused by climate change. Stationarity of rainfall is assumed in traditional hydrological frequency analysis. Stationarity means that the probability characteristic of rainfall does not change depending on proceeding of time. Therefore, in general, probability distribution which dominates rainfall system is not changed in traditional analysis. On the other hand, confidence interval and prediction interval which reflect the effect of climate change can be constructed due to that incorporating information of future projection data and updating these intervals by using the Bayesian method. In this paper, the method to derive confidence interval and prediction interval under the situation when global warming proceeds by incorporating climate projection data into extreme value distribution derived from past observed extreme rainfall data which are available so far is also shown.

2. FORMULATION OF CONFIDENCE INTERVAL AND PREDICTION INTERVAL

In this chapter, the method of statistical estimation and prediction are shown. Also, the concept of confidence interval and prediction interval are illustrated.

2.1 The method of statistical estimation

Statistical estimation is to estimate characteristic values of population based on an available sample $X=\{X_1, X_2,\ldots, X_n\}$. As statistical estimation methods, there are point estimation and interval estimation. Point estimation is a method to assign parameter θ an estimated statistic $\hat{\theta}$ which was derived from an available sample. On the other hand, interval estimation is a method to derive an interval constructed of lower confidence limit value $L_{C.I.}(X)$ and upper confidence limit value $U_{C.I.}(X)$ based on an available statistical sample, and verify that this interval includes true value of parameter with a probability of more than (1-β). Here, β ($0<\beta<1$) is defined as a confidence coefficient. An interval $[L_{C.I.}(X), U_{C.I.}(X)]$ is defined as 100p% confidence interval for parameter θ. Next, definition of 100p% confidence interval concerning parameter θ is shown in Eq.(1).

$$P\left(L_{C.I.}\left(X\right)\leq\theta\leq U_{C.I.}\left(X\right)\right)\geq 1-\beta \tag{1}$$

Where, (1-β) is a confidence coefficient.

2.2 The method of statistical prediction

Statistical prediction is to predict an unknown random value Y which occurs in the future. Here, Y is random variable which expresses observed data in future time. In hydrology, this method is used for prediction of extreme rainfall Y in the future time by using the past observed data X. As statistical prediction methods, there are point prediction and interval prediction. Point prediction is to predict future data Y by assigning one predicted value $\hat{Y}$ derived from a sample X to an unknown random variable Y. On the other hand, interval prediction is to predict by a form that future data exists with an arbitrary probability within interval constructed of lower prediction limit value $L_{P.I.}(X)$ and upper prediction limit value $U_{P.I.}(X)$, which are derived from sample X. Prediction interval for an unknown random variable Y in future time is shown in Eq.(3).

$$P\left(L_{P.I.}\left(X\right)\leq Y\leq U_{P.I.}\left(X\right)\right)\geq 1-\beta \tag{3}$$

Here, interval $[L_{P.I.}(X), U_{P.I.}(X)]$ verify that Y is included in this interval with a probability of more than (1-β) is called as 100(1-β) % prediction interval. When estimating the range which adopted probability distribution derived from sample $\boldsymbol{X}$ can take, confidence interval should be used. On the other hand, when prediction for future extremes itself is needed, prediction interval should be used.

3. OUTLINE OF PROBABILITY LIMIT METHOD TEST

Probability limit method test is the hypothesis theory which has strong power of test at the tail of an assumed probability distribution. This test was proposed by Moriguti (1995) as theory to improve the weak power of test of Kolmogorov-Smirnov test (Kolmogorov, 1933) at the tail of an assumed probability distribution. Here, the power of test is expressed as the adopted region which can estimate the range observed data can take. The adopted region of probability limit method test narrows at the tail of an assumed probability distribution. This means that extreme which corresponds to design level can be estimated as an adopted region with high accuracy. "The power of test" and "confidence interval and prediction interval" relates each other. In general, the greater power of test the adopted hypothesis test theory has, more precise confidence interval and prediction interval based on the adopted hypothesis theory becomes. Therefore, in this research, confidence interval and prediction interval based on probability limit method test which have high accuracy at the tail of distribution are constructed to evaluate uncertainty of hydrological statistics. Figure 1 shows the construction process of an adopted region corresponding to 5% two-sided probability. From this diagram, it can be seen that adopted regions $[z_L(i), z_U(i)]$ for ith cumulative probability are converted to adopted regions $[\chi_X(z_L(i)), \chi_X(z_U(i))]$ for ith order annual maximum total rainfall $X_{(i)}$ itself through the representing function of an assumed probability distribution. It is a remarkable point for probability limit method test to construct an adopted region for observed order statistics $X_{(i)}$ by using information that order statistics from standard uniform distribution follow beta distribution which parameter is (i, n-i+1).

The confidence limit line can be constructed by the same function form to the adopted distribution to limit values of an adopted regions. Also, the prediction limit line can be constructed by the probability distribution fitted with limit values of an adopted regions (Shimizu et al,2018).

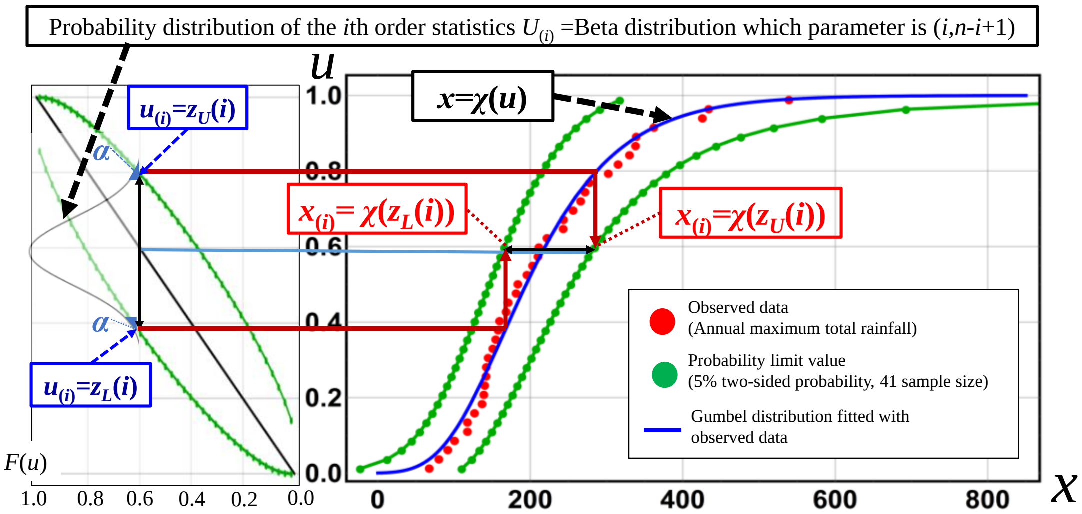

Figure 1. Construction process of probability limit line of probability limit method test for an assumed probability distribution in the case of 5% two-side probability and 41 sample size (Here, an adopted region is constructed by lower rejection limit values and upper rejection limit values. Rejection limit values of probability limit method test is defined as probability limit values. Also, lower probability limit value is described as $\chi_X(z_L(i))$ and upper probability limit value is described as $\chi_X(z_U(i))$ in an assumed probability distribution.)

4. BAYESIAN BASED PROJECTION METHOD TO ESTIMATE CONFIDENCE INTERVAL AND PREDICTION INTERVAL

To manage the non-stationarity of rainfall caused by climate change, Bayesian theory is so effective one. Because it can construct the conditional probability distribution which can synthesize projection data and observed data. Here, the conditional probability distribution means future change of our targeted extremes variable (annual maximum rainfall) under the scenario which assumes higher global average temperature caused by global warming. In this research, metropolis method, the one of Markov Chain Monte Carlo methods is adopted for Bayesian update to let our recognition against extreme rainfall be that under the proceeding of global warming. In this research, the regional experiment of database for Policy Decision making for Future climate change (d4PDF) as the future projection data and it was substituted into the likelihood function of metropolis method algorithm for Bayesian update of Gumbel distribution fitted with 44 observed annual maximum daily rainfalls in Ushizuma observatory in Abe river water system, Japan. The regional experiment of d4PDF is created by regional climate model, and calculated data of horizontal resolution of 20 km (Mizuta et al., 2016). Also, this regional experiment is constructed of ensemble data calculated by regional climate model which horizontal resolution is 20km. More precisely, this regional experiment is constructed of the past experiment which targets for past 60 years from 1951 to 2010 and the +4K experiment which assumes a condition of increasing of global average temperature by +4K from pre-industrial revolution and targets for future 60 years from 2051 to 2110. Also, the past experiment (60 years × 50 members = 3,000 years) has 50 ensemble members added to perturbation of sea ice condition, sea surface temperature and different initial conditions. The +4K experiment (60 years × 6 sea surface patterns × 15 members = 5,400 years) has 6 patterns of sea surface temperature pattern and 15 ensemble members added to perturbation of these patterns of sea surface temperature.

Figure 1 shows the analytical data (black points), Gumbel distribution fitted with observed data (a blue solid line), 95% confidence interval of this Gumbel distribution in past climate (a blue range), Bayesian updated Gumbel distribution and 95% confidence interval of this Bayesian updated Gumbel distribution. From overlapped range of both intervals in past and future climate, there is a probability exists that probable rainfall in past climate could occur even in future climate. Also, confidence interval in future climate takes wider range than that of past climate, and it is recognized that probable rainfall will increase in future climate. Moreover, considering the return period of 100-year which is often used for design level in Japan, 100-year annual maximum daily rainfall in future climate (493.3mm) is an about 1.33 times greater value than that of past climate (370.9mm). Also, 95% upper confidence limit value in future climate (621.4mm) is an about 1.35 times greater value than that of past climate (460.3mm). By quantifying the future change ratio of design level rainfall like this, it can be expected that update of confidence interval suggests feasibility of consideration on probable rainfall under climate change accompanied with global warming. It is expected that update of confidence interval by assimilating future projection data to an adopted probability distribution used for river planning contributes a consideration on the estimation for probable rainfall under climate change.

Figure 1 shows observed data (black points), Gumbel distribution fitted with observed data (a blue line), 95% prediction interval based on probability limit method test in the past climate (a blue range), Bayesian updated Gumbel distribution from which is derived by applying MCMC method to Gumbel distribution fitted with observed data (a red solid line) and 95% prediction interval based on probability limit method test in the future climate (a red range). From overlapped range of both prediction intervals, it can be recognized that annual maximum daily rainfall in future climate could occur even in the past climate with some probability. Also, prediction interval in the future climate takes wider range than that of past climate, it shows occurrence risk of catastrophic rainfall increases in the future climate. Moreover, considering the return period of 100-year which is often used for design level in Japan, 95% upper prediction limit value in future climate (883.8mm) is an about 1.39 times greater value than that of past climate (638.0mm). By update of prediction interval, it can be possible to evaluate the scale and occurrence risk of catastrophic rainfall under the time of proceeding of global warming.

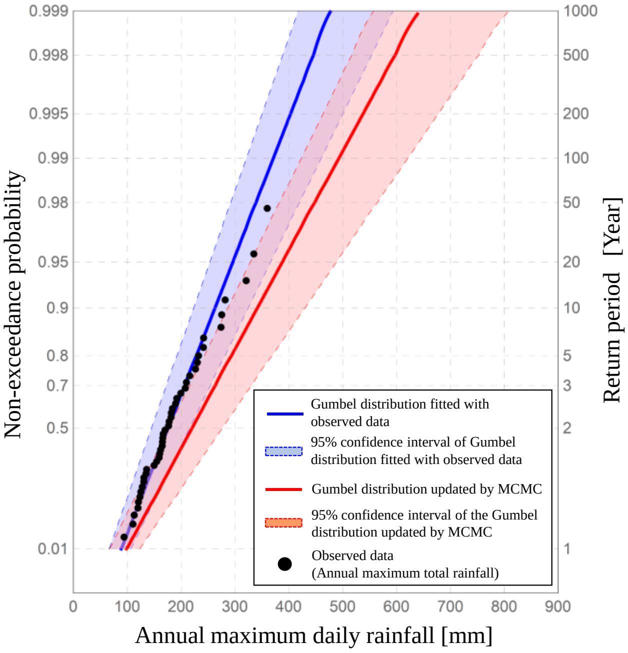

Figure 2. Update of confidence interval using Bayesian method which assimilates ensemble climate projection database

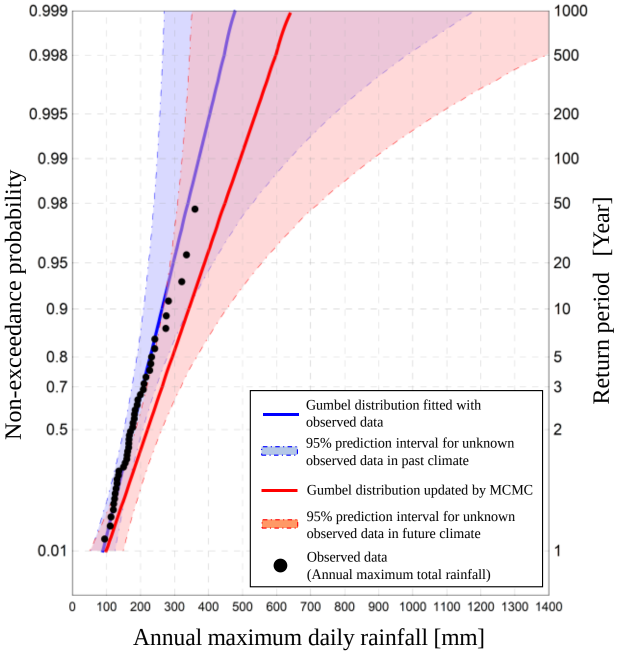

Figure 3. Update of prediction interval using Bayesian method which assimilates ensemble climate projection database

5. CONCLUSIONS

Confidence interval quantifies the degree of instability of the adopted probability distribution, and the risk in river planning can be expressed by considering the range of confidence interval. Prediction interval quantifies the range future observed data could take, and it can be possible to estimate the occurrence risk and scale of catastrophic heavy rainfall. In this paper, the construction method of confidence interval and prediction interval based on probability limit method test is shown, and its actual application is illustrated. Also, in this research, it was shown that a harmonic relation holds between confidence interval based on physical Monte Carlo method which uses ensemble climate projection database and confidence interval based on a mathematical theory, probability limit method test. This result suggests scientific verification for an introduction of concept of intervals concerning probable rainfall in hydrological frequency analysis.

REFERENCES

Sigeiti Moriguti (1995). Testing Hypothesis on Probability Representing Function – Kolmogorov-Smirnov Test Reconsidered -, pp.233-244, Japanese Journal of Statistics and Data Science, Vol.25, 1995.

Kolmogorov, A. (1933). Sulla Determinazione Empiri-ca di una Legge di Distribuzione. Inst. Ital. At-tuari, Giorn, 4, 83-91.

Keita SHIMIZU, Tadashi YAMADA and Tomohito YAMADA (2019). Uncertainty Evaluation of Hydrological Frequency Analysis Based on Confidence Interval and Prediction Interval, Advances in River Engineering, Vol.25, pp.13-18.

Ryo Mizuta, Akihiko Murata, and Masayoshi Ishii (2016). Over 5,000 Years of Ensemble Future Climate Simulations by 60-km Global and 20-km Regional Atmospheric Models, Bull.Am. Meteorol. Soc., pp.1383-1393.