Qualitative and quantitative uncertainties in regional rainfall frequency analysis[*]

Qian ZHU, Xiao XU, Chao GAO, Qi-hua RAN, Yue-ping XU[†‡]

(*Institute of Hydrology and Water Resources, Zhejiang University, Hangzhou 310058, China*)

[†]E-mail: yuepingxu@zju.edu.cn

Abstract: Uncertainty exists widely in hydrological analysis, and this makes the process of uncertainty assessment very important for making robust decisions. In this study, uncertainty sources in regional rainfall frequency analysis are identified for the first time. The numeral unite spread assessment pedigree (NUSAP) method is introduced and is first employed to quantify qualitative uncertainty in regional rainfall frequency analysis. A pedigree matrix is particularly designed for regional rainfall frequency analysis, by which the qualitative uncertainty can be quantified. Finally, the qualitative and quantitative uncertainties are combined in an uncertainty diagnostic diagram, which makes the uncertainty evaluation results more intuitive. From the integrated diagnostic diagram, it can be determined that the uncertainty caused by the precipitation data is the smallest, and the uncertainty from different grouping methods is the largest. For the downstream sub-region, a generalized extreme value (GEV) distribution is better than a generalized logistic (GLO) distribution; for the south sub-region, a Pearson type III (PE3) distribution is the better choice; and for the north sub-region, GEV is more appropriate.

Key words: Qualitative uncertainty, Uncertainty analysis, Numeral unite spread assessment pedigree (NUSAP) method, Regional rainfall frequency analysis, Pedigree matrix, Diagnostic diagram

doi:10.1631/jzus.A1400123   **Document code:** A   **CLC number:** TV125

1 Introduction

Regional frequency analysis is an important topic in hydrology and water resources. However, uncertainties existing in regional frequency analysis make this problem complicated. Statistical flood or rainfall estimates are affected by increasing uncertainty with decreasing frequency of occurrence because the quantiles of the probability distribution of the extreme flood flows or rainfall amounts are inferred from a data sample of relatively short length (Michele and Rosso, 2001). Regionalization procedures attempt to overcome the shortage of limited measurement data through increasing the sample size by substituting 'space' to augment 'time'. However, besides the data sources, a lot of other uncertainties exist in regional frequency analysis. There is uncertainty from different grouping methods to define the hydrological homogeneous regions. The choice of frequency distribution is also an uncertainty source for regional frequency analysis. Thus, evaluating the uncertainty in regional frequency analysis is important, particularly for robust engineering infrastructure design and management.

Mainstream methods, such as generalized likelihood uncertainty estimation (GLUE) (Jin *et al.*, 2010; Li *et al.*, 2010; Delsman *et al.*, 2013), Monte Caro simulation (Jeremiah *et al.*, 2011; Houska *et al.*, 2013), and the Bayesian approach (Parent and Bernier, 2003; Reis and Stedinger, 2005; Bouda *et al.*, 2012), are often used for uncertainty analysis. Key

dimensions of uncertainty in regional frequency analysis that need to be addressed are technical, methodological, and epistemological. Quantitative methods mentioned above address the technical dimension only (van der Sluijs *et al.*, 2005). Qualitative uncertainties, such as those originating from methods, are far less well studied. Therefore, how to quantify qualitative uncertainty remains a difficult task in hydrological analysis. Apart from that, how to assess the qualitative and quantitative uncertainties integrally is also a great challenge.

In this paper, to account for the qualitative uncertainty, the numeral unite spread assessment pedigree (NUSAP) method proposed by Funtowicz and Ravetz (1990) is introduced for the first time to evaluate both quantitative and qualitative uncertainties in regional rainfall frequency analysis (RFA). The NUSAP method is able to address aspects of data, methods, or model quality resulting from uncertainties that are hard to quantify, such as methodological and epistemological uncertainties, and that are not systematically taken into account in scientific studies. A pedigree matrix is particularly designed for regional rainfall frequency analysis, by which the qualitative uncertainty can be effectively quantified.

2 Methodology

The framework of this study is presented in Fig. 1. Uncertainty sources in regional frequency analysis are first defined. Among these, three sources, i.e., precipitation measurement error, different methods to identify homogeneous regions, and different frequency distributions, are selected to assess their impact on the quantitative and qualitative uncertainties on design rainfall with the NUSAP method. The pedigree matrix, particularly designed for regional frequency analysis, is used to quantify the qualitative uncertainties. The impact of measurement errors in precipitations (quantitative uncertainty) is analyzed by the Latin Hypercube simulation. These two kinds of uncertainties are then assessed integrally by a diagnostic diagram.

2.1 Regional rainfall frequency analysis and its uncertainties

Frequency analysis of extreme rainfall usually implies extrapolations well beyond the range of the

available at-site data (Norbiato *et al.*, 2007). Therefore, regional frequency analysis is used to provide an alternative for statistical analysis of these extremes. Regional frequency analysis has been a well-established method in hydrology for many years; an example is the index variable procedure of Dalrymple (1960). The index variable procedure is a convenient way of pooling summary statistics from different data sites, which is also employed in this study. Among five assumptions (Hosking and Wallis, 2005) of index variable procedure, the main one is that the sites in a homogeneous region have an identical frequency distribution apart from a site-specific scaling factor, which is the index variable. According to Hosking and Wallis (2005), the index is usually the mean or median of the site-specific data. In this study, the index variable is the mean of the annual maximum daily precipitation. The regional frequency analysis based on an index variable procedure involves four steps, which are screening of the data, identification of the homogenous regions, choice of a frequency distribution, and estimation of the frequency distribution parameters.
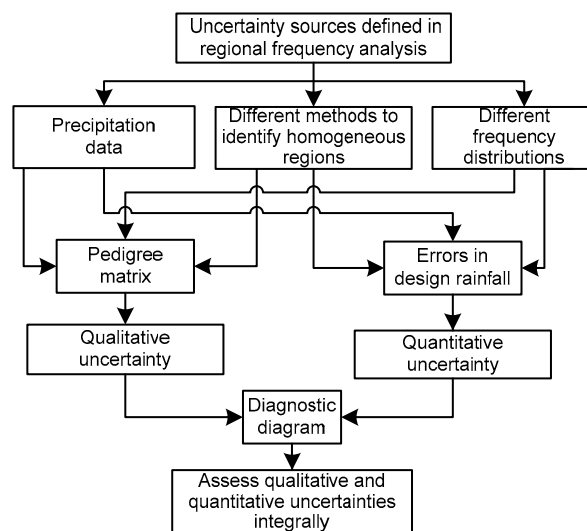

Fig. 1 Framework used in this study

The first step in the procedure is screening of the data, which is to test for incorrect data values, outliers, shifts, and trends. A discordancy measure (Hosking and Wallis, 2005) is used to identify those sites from a group of given sites that are grossly discordant with the group as a whole. The discordancy measure is a single statistic based on the difference between the

L-moment ratios of a site and the average *L*-moment ratios of a group of similar sites (Norbiato *et al.*, 2007).

The second step is identifying homogeneous regions, which is to form groups of stations with identical frequency distributions apart from a scale factor. A variety of methods have been proposed for forming groups of similar sites for use in regional frequency analysis. The geographical grouping method is used to define geographical regions which contain contiguous sites based on administrative areas or major physical grouping of sites (Hosking and Wallis, 2005). Cluster analysis is a standard method of statistical multivariate analysis, which has been successfully and widely used to identify homogeneous regions (Baeriswyl and Rebetez, 1997; Castellarin *et al.*, 2001; Lin and Chen, 2006; Ramachandra Rao and Srinivas, 2006). Recently, the self-organization feature map (SOM), a modern hydroinformatic tool, has been applied for clustering watersheds (Lin and Chen, 2006; Farsadnia *et al.*, 2014). In this study, a geographical grouping method and a direct cluster analysis method are chosen to analyze the uncertainty from different methods to identify homogeneous regions.

There are two methods commonly used to test homogeneity of the grouped regions. According to Ilorme and Griffis (2013), compared with the product-moment coefficient of variation (Wiltshire, 1986), *L*-moment ratios, such as *L*-CV and *L*-skewness, the first and third moment, respectively (Hosking and Wallis, 2005), are most commonly used in practice, and will be employed herein. The latter heterogeneity measure is called the *H*-statistic, which contains three measures. The first, *H*(1), is the standard deviation of the at-site *L*-CVs, which is used in this study and is used as *H* hereafter. The second, *H*(2), and the third, *H*(3), can also be used, but many studies (Castellarin *et al.*, 2001; Hosking and Wallis, 2005; Ilorme and Griffis, 2013) show that *H*-statistic based solely on the *L*-CV is the most effective at discriminating between homogeneous and heterogeneous regions. A region is considered 'acceptably homogeneous' if *H*<1, 'possibly heterogeneous' if 1≤*H*≤2, and 'definitely heterogeneous' if *H*>2 (Hosking and Wallis, 2005). The heterogeneity measure is calculated by

$$H = \frac{V - \mu_V}{\sigma_V},\qquad(1)$$

where *V* represents the at-site weighted standard deviation of the proposed region, which is defined as

$$V = \left\{ \frac{\sum_{i=1}^{N} n_i (L\text{-CV}^i - \overline{L\text{-CV}})^2}{\sum_{i=1}^{N} n_i} \right\}^{1/2},\qquad(2)$$

where n_i and $L\text{-CV}^i$ are the record length and *L*-CV of site *i*, respectively; *N* is the number of sites in the proposed region; and μ_V and σ_V are the mean and standard deviation values of *V*, respectively, computed for the simulated regions. A four-parameter kappa distribution is used to simulate a large number of homogeneous regions by a Monte Carlo method.

The third step is choosing a frequency distribution. The aim of this step is to find a distribution that will yield accurate quantile estimates for each region. Many goodness-of-fit techniques are available for this purpose. In this study, the *L*-moment ratio diagram (Hosking and Wallis, 2005) showing the relationship of *L*-kurtosis (the fourth moment) versus *L*-skewness is used to compare the five selected distributions obtained from at-site data and the corresponding regional data.

The fourth step is estimating the parameters of the frequency distributions. Compared with other estimation methods, such as the maximum likelihood method and the conventional product moments method, the *L*-moment method is less influenced by the effects of sampling variability and can yield more efficient parameter estimates (Norbiato *et al.*, 2007). Therefore, a regional *L*-moment algorithm is employed to estimate the parameters of the regional frequency distributions. Five models are considered: a generalized extreme value (GEV) distribution, a generalized Pareto (GPA) distribution, a generalized logistic (GLO) distribution, a three-parameter lognormal (LN3) distribution, and a Pearson type III (PE3) distribution. The five models are all three-parameter distributions, and their parameters are obtained by the three sample *L*-moments (Hosking and Wallis, 2005). The formulas for the parameter estimations can be referred to Hosking and Wallis (2005).

As mentioned above, uncertainty exists widely in regional frequency analysis. In this study, uncertainty sources in regional RFA are summarized in Table 1.

Three main sources are used to assess the impact of quantitative and qualitative uncertainties on regional frequency analysis by employing the NUSAP method. The three sources are measurement errors from precipitation data, different methods to identify homogeneous regions, and different frequency distributions.

2.2 NUSAP and the diagnostic diagram

The NUSAP method is a notional system originally proposed by Funtowicz and Ravet (1990), which aims to provide an analysis and diagnosis of uncertainty for complex policy problems. It captures both quantitative and qualitative dimensions of uncertainty and enables one to communicate those dimensions in a standardized and self-explanatory way.

NUSAP is a system designed to evaluate quality uncertainty. van der Sluijs *et al.* (2005) presented the details of the five qualifiers. The first three are the normal quantitative aspects of the analysis, and the last two are the qualitative part of the method. The first is numeral, which is usually an ordinary number. The second is unit, which will be a millimeter to describe the amount of precipitation in this study. The third is spread, which generalizes from the random error of experiments or the variance of statistics. Although spread is usually conveyed by a number (either ±, %, or 'factor of'), it is not an ordinary quantity, for its own inexactness is not the same sort as that of measurements. Assessment expresses qualitative judgments about the information, which is usually related with the pedigree matrix. The pedigree matrix makes a distinction among empirical, methodological, and statistical assessment criteria. To minimize arbitrariness and subjectivity in measuring strength, a pedigree matrix is used to code qualitative expert judgments for criterion into a discrete numeral scale for 0 (weak) to 4 (strong) with modes of each level on the scale. In this study, a pedigree matrix for regional frequency analysis is proposed in Table 2. In the matrix, the columns represent the statistical, empirical, and methodological assessment criteria, and within each column there are linguistic descriptions, normatively ranked in descending order.

NUSAP addresses two independent properties related to uncertainty in numbers, namely spread and strength. The two metrics can be combined in a diagnostic diagram mapping strength and criticality to

Table 1 Uncertainty sources of regional rainfall frequency analysis

RFA process	Uncertainty from hypothesis	Uncertainty from method	Uncertainty from data
Precipitation data	Strictly controlled process of measurement; calibrated measuring instruments	Differences in measurement errors from different measurement methods	Insufficient precipitation stations and precipitation data series; unavoidable factors, e.g., climate change
Identification of homogeneous regions	Independent precipitation stations; unrelated precipitation series	Prediction errors from different methods to define homogeneous regions; different methods to test homogeneity of the grouped regions	Measurement error from precipitation data; different methods to select precipitation data series
Choice of a frequency distribution	Robust models; samples excellently fitted to the selected distributions	Different frequency distributions, such as GEV, PE3, and GLO	Measurement errors from precipitation data; different methods to select precipitation data series
Parameter estimation	Precipitation data series fit to the distribution defined by the parameters	Different parameter estimation methods, such as L-moment and maximum likelihood method	Measurement errors from precipitation data; different methods to select precipitation data series
Index-variable method	Design rainfall calculated with the annual maximum precipitation series	Methods to calculate regional design rainfall other than index flood method	Uncertainty from selecting annual maximum precipitation; methods to deal with outliers
Error propagation	Completely randomly sampling	Different sampling methods, e.g., Markov chain Monte Carlo (MCMC) and Latin hypercube sampling (LHS)	Errors of precipitation (the propagation of quantity errors); uncertainty in defining values of pedigree matrix (the propagation of quality errors)

spread. The diagnostic diagram is based on the notion that neither spread alone nor strength alone is a sufficient measure for uncertainty. Spread expresses inexactness whereas strength expresses the methodological and epistemological limitations of the underling knowledge base. Spread and strength also refer to the quantitative and qualitative uncertainty, respectively. Mapping those two metrics in a diagnostic diagram reveals the best spot and helps in the setting of priorities for improvement.

3 Case study

3.1 Study area

The Qiantang River Basin is located in the Zhejiang Province that lies in eastern China between 117.62°E and 121.87°E, and 28.17°N and 30.48°N (Fig. 2). The drainage area is about 55 600 km^2, and the total length is about 589 km. The Qiantang River Basin is the largest and longest river system in Zhejiang Province, which passes through the provincial capital Hangzhou, before flowing into the East China Sea through the Hangzhou Bay. Because it is close to the sea, the area is subject to a subtropical monsoon climate. The mean annual precipitation is about 1200–2200 mm in various locations.

Daily precipitation data of 20 stations in the Qiantang River Basin are used in this study (Fig. 2). All the stations have more than 40 years of record length, and 17 stations among them have more than

50 years. Table 3 lists the information of the precipitation stations.

3.2 Uncertainty from precipitation data

The measurement errors and randomness of the precipitation result in uncertainties from the precipitation data. Measurement errors have a great effect on the results of regional frequency analysis, and uncertainty from the randomness of the precipitation data can be referred to Xu *et al.* (2011). Here only measurement errors from precipitation are of concern. There are three main sources of measurement errors in precipitation, which are measurement errors of wind, moisture, and evaporation (SL21-90, 1991). Measurement errors from these three sources are summarized in Table 4.
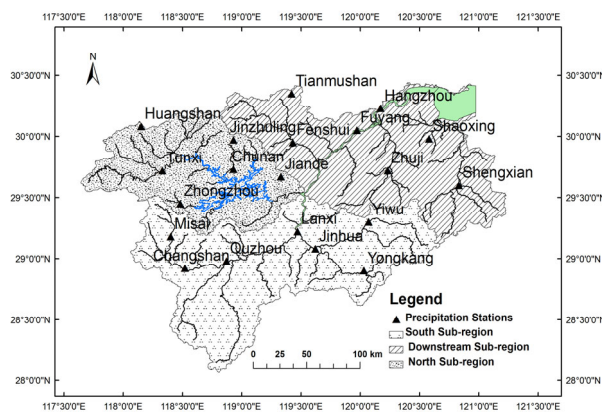

Fig. 2 Location of precipitation stations and homogeneous regions of the geographic grouping method

Table 2 Pedigree matrix for regional frequency analysis

Score	Statistical quality	Empirical quality	Methodological quality
4	Excellent fit to well-known statistical models (GEV, GLO, PE3, and so on)	Uniformly and reasonably distributed precipitation stations; strictly controlled measurement and calculation; enough direct measurements ($n \geq 50$)	RFA methods (e.g., methods to delineate homogeneous regions and methods to calculate design rainfall) in well-established discipline
3	Good fit to a reliable statistical model by most fitting tests, but not all	Well distributed precipitation stations; uncontrolled measurement and calculation; small sample direct measurements ($n < 50$)	Reliable and common method within established discipline or best available in immature discipline
2	Fitting tests not significant, model not clearly related to data, or model inferred from similar data	Sparsely and unreasonably distributed precipitation stations; uncontrolled measurement and calculation; indirect measurements	Acceptable methods but limited consensus on reliability
1	Distributions are chosen subjectively without using test fitting experiments	Educated guesses indirect approx.; rule of thumb estimates	Preliminary methods with unknown reliability
0	Unknown models	Crude speculation	No discernible rigor

n: record length

Table 3 Information of precipitation stations used in this study

Station	N	Longitude (°)	Latitude (°)	Station	N	Longitude (°)	Latitude (°)
Tianmushan	42	119.42	30.35	Jiande	38	119.33	29.67
Hangzhou	58	120.17	30.23	Misai	44	118.40	29.18
Jinhua	56	119.62	29.08	Fuyang	51	119.97	30.05
Shengxian	56	120.83	29.60	Lanxi	58	119.47	29.22
Quzhou	58	118.87	28.98	Shaoxing	57	120.58	29.98
Huangshan	53	120.75	28.57	Yiwu	61	120.07	29.30
Tunxi	56	118.33	29.72	Yongkang	70	120.03	28.90
Changshan	52	118.52	28.92	Zhongzhou	50	118.48	29.45
Chun'an	51	118.93	29.73	Jinzhuling	51	118.93	29.97
Fenshui	54	119.43	29.95	Zhuji	56	120.23	29.72

To analyze the impacts of the measurement errors from precipitation on regional frequency analysis, the most unfavorable circumstance, which is 14%, is chosen. The Latin hypercube sampling (LHS) method is used to propagate the measurement error into design rainfalls. The results in Tianmushan of the downstream sub-region are taken as an example, which are presented in Table 5. The GEV is chosen to calculate the design rainfalls, for it is the most appropriate frequency distribution for the downstream sub-region, which will be discussed in Section 3.4. The results show that measurement errors cannot be ignored in regional frequency analysis. For the 200-year return period, the errors propagated into the design rainfall reaches 13%.

3.3 Uncertainty from different methods to identify homogeneous regions

The Qiantang River Basin has two important sources which are Xin'anjiang located in the north of the region and Lanjiang located in the south of the region. Those two sources join with each other and become Fuchunjiang, which is located downstream. Therefore, the basin can be divided into three sub-regions with the geographical grouping method: the north sub-region which is the Xin'anjiang catchment; the south sub-region which is the Lanjiang catchment; and the downstream sub-region which is the Fuchunjiang catchment (Fig. 2). When using the cluster analysis method to delineate the homogeneous regions, four cluster factors are chosen, which are longitude, latitude, average annual precipitation, and *L*-skewness. The results are presented in Fig. 3. According to cluster analysis, the region is divided into four clusters. The heterogeneity measure *H* is reported in Table 6 for the sub-regions of the two

Table 4 Measurement errors of precipitation

Source	Measurement error
Wind	2%–10%
Moisture	0.05–0.3 mm (1%–2%)
Evaporation	1%–2%
Total	4%–14%

Table 5 Design rainfalls of Tianmushan station

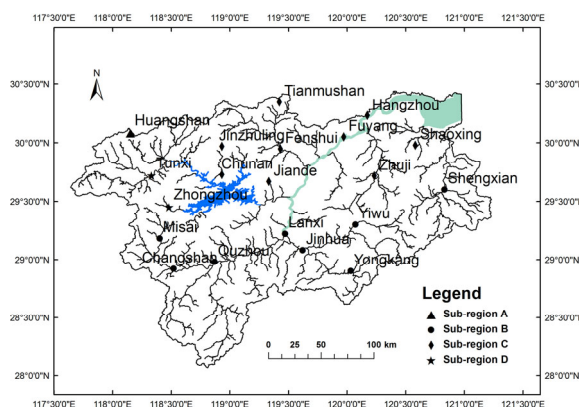| Return period (year)	Mean (mm)	Maximum (mm)	Minimum (mm)	Standard deviation (mm)	Error (%)
5	122.72	127.50	118.48	1.20	7.36
10	146.96	152.45	141.87	1.50	7.20
20	172.38	178.49	166.52	1.96	6.94
50	208.77	218.30	200.04	2.97	8.75
100	238.88	252.24	226.33	4.13	10.85
200	271.54	289.64	254.21	5.67	13.05

Fig. 3 Homogeneous regions of cluster analysis

grouping methods. Meanwhile, all the sub-regions pass the discordancy test. The results show that the grouping of four sub-regions is reasonable. Here, Chun'an station, Jiande station, and Jinzhuling station

are taken as examples to analyze the impact on design rainfalls caused by different grouping methods. The results in Fig. 4 show that design rainfalls calculated based on cluster analysis are smaller than those calculated based on the geographical grouping method in small return periods, but larger in large return periods.

Table 6 Heterogeneity measure *H* for the sub-regions of the two grouping methods

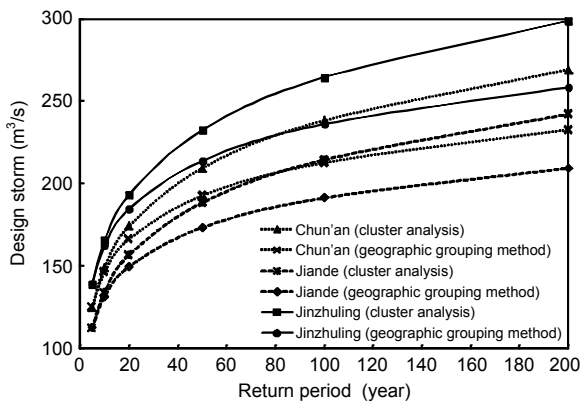| Sub-region	Geographical grouping method			Cluster analysis method			
	North	South	Down-stream	A	B	C	D
H	−0.992	0.5	−0.959	−	−1.323	−0.05	−1.62

Fig. 4 Design rainfalls of different grouping methods

To quantify the differences of design rainfalls based on the geographic grouping method and cluster analysis method, four ungauged hydrological sites are employed. The methodology used to calculate design rainfalls in those four stations is based on the deviation mean in the 'Atlas of storms in short duration for Zhejiang Province' (BOHZJ, 2003). The average relative error of design rainfalls calculated by the grouping method and by the deviation mean is considered the uncertainty from different methods to delineate homogenous regions. Calculated results show that the errors from the geographic grouping method and cluster analysis method are 19.05% and 21.23%, respectively.

3.4 Uncertainty from different frequency distributions

The downstream sub-region is presented herein to analyze the uncertainty from different frequency distributions. Fig. 5 illustrates the *L*-moment ratio diagram for downstream sub-region, where OLB is the overall lower bound of *L*-kurtosis as a function of *L*-skewness. It shows that GEV and GLO fit better than the other distributions. The design rainfalls calculated with these two frequency distributions are shown in Table 7. The results reveal that design rainfalls obtained by GLO are larger than those obtained by GEV, which shows that frequency distributions do cause uncertainties in design rainfalls particularly for large return periods.

To test the goodness of fit of these two distributions, the Kolmogorov-Smirnov test is used. The goodness of fit statistics between GEV and GLO is quantified. Table 8 presents the results. The errors are the relative differences between *L*-kurtosis based on observed data and *L*-kurtosis based on these two distributions. The results show that the GEV is the most appropriate distribution for the downstream sub-region. The errors are used as quantified uncertainty from different frequency distributions.

3.5 Integrated assessment of uncertainties

This section presents the assessment of uncertainties from precipitation, different grouping methods, and different frequency distributions. Here the diagnostic diagram is employed to combine quantitative and qualitative uncertainties from those three sources in the regional frequency analysis.

Measurement errors from precipitation are evaluated based on the geographic grouping method, thus the score of the statistical quality of measurement
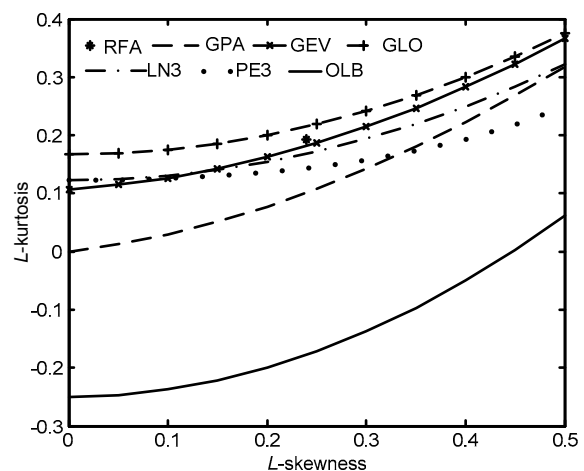

Fig. 5 *L*-moment ratio diagram for downstream area

errors is identical to that of the geographic grouping method, which is 3.8 (it will be discussed later). Precipitation stations in the Qiantang River Basin distribute reasonably. Measurement and calculations are well controlled. As mentioned above, all the stations have more than 40 years of record length, and 17 stations have more than 50 years of record length. But compared to other international studies on regional frequency analysis (Fowler and Kilsby, 2003; Saf, 2010; Zaman *et al.*, 2012), stations in the Qiantang River Basin distribute slightly dispersedly. Therefore, according to the pedigree matrix, the score of the empirical quality is set to 3.5. Among the methods employed to evaluate the impacts of measurement errors of precipitation on regional frequency analysis, LHS and index variable procedures are standard methods in well-established disciplines, and the geographic grouping method is considered reliable and common within established disciplines, which results in a score of methodological quality of 3.6.

Table 7 Design rainfalls of the Tianmushan station calculated by GEV and GLO

Distri-bution	Design rainfall (mm)					
	5 year	10 year	20 year	50 year	100 year	200 year
GEV	122.46	146.16	171.84	207.40	237.02	269.61
GLO	120.49	143.20	168.88	209.37	245.91	289.37

Table 8 Goodness of fit statistics between GEV and GLO

Distribution	*L*-skewness	*L*-kurtosis	Error (%)
GEV	0.2511	0.1881	7.02
GLO	0.2511	0.2190	8.26
Results based on observed data	0.2511	0.2023	–

As calculated previously, the errors from the geographic grouping method and cluster analysis are 19.05% and 21.23%, respectively. The qualitative uncertainty of the grouping methods is assessed according to the pedigree matrix. The selected frequency distributions fit well to the observations, so the score of the statistical quality of the two grouping methods is 3.8. The score of the empirical quality of grouping methods is identical to that of the measurement errors, which is 3.5. The geographic grouping method is commonly used, but slightly subjective; therefore, the score of the methodological quality is set to 2.8. The cluster analysis method is reliable and widely used, but the number of cluster factors is rather limited, so the score of the methodological quality is set to 3.0.

Every sub-region has its own fitted distributions. Sub-regions grouped by the geographic grouping method are taken as examples to analyze the uncertainty from different distributions. According to the ratio diagram of *L*-kurtosis versus *L*-skewness of frequency distributions, each two best fitted distributions are selected for each sub-region, which are then used to derive the differences caused by the distributions. The quantitative uncertainty is assessed by the fitted errors based on the Kolmogorov-Smirnov test, and the qualitative uncertainty is scored according to the pedigree matrix.

The results mentioned above are summarized in Table 9 for design rainfalls with a 200-year return period. The strength value is equal to the summation of the statistical quality score, empirical quality score, and methodological quality score, which are presented in the 'Pedigree score' in Table 9, respectively, divided by 12.

Table 9 Integrated uncertainty assessment

Number	Unit	Spread (%)	Pedigree score	Strength
Precipitation data	mm	13.05	(3.8, 3.5, 3.6)	0.91
GEV (downstream sub-region)	mm	7.02	(3.5, 3, 3.5)	0.83
GLO (downstream sub-region)	mm	8.26	(3, 3, 3.5)	0.79
PE3 (south sub-region)	mm	4.25	(3.8, 3, 3.5)	0.86
LN3 (south sub-region)	mm	15.49	(3, 3, 3.5)	0.79
GEV (north sub-region)	mm	3.21	(3.8, 3, 3.5)	0.86
LN3 (north sub-region)	mm	6.36	(3.6, 3, 3.5)	0.84
Geographic grouping method	mm	19.05	(3.8, 3, 2.8)	0.80
Cluster analysis	mm	21.23	(3.8, 3, 3.0)	0.82

Then, the diagnostic diagram (Fig. 6) can be used to illustrate the results intuitively. The horizontal axis is the spread. The smaller spread means less quantitative uncertainty. The vertical axis is the strength value. The smaller value means more qualitative uncertainty. Thus, the dot which is the closest to (0, 1) has both the least quantitative and qualitative uncertainties. Therefore, the conclusions can be drawn from Fig. 6 that among the three main uncertainty sources for regional frequency analysis in this case study, the uncertainty caused by the precipitation data is the smallest, but the uncertainty from the different grouping methods is the largest. For the downstream sub-region, GEV is better than GLO; for the south sub-region, PE3 is the better choice; for the north sub-region, GEV is more appropriate.
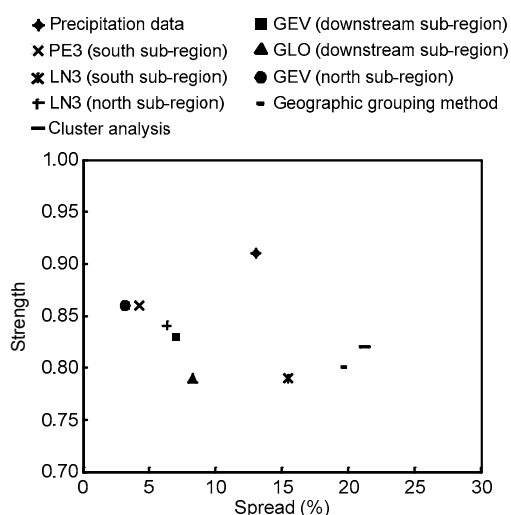

Fig. 6 Integrated diagnostic diagram of uncertainty

4 Conclusions

This study summarized uncertainty sources in regional RFA and designed a pedigree matrix particularly for regional frequency analysis. Based on the pedigree matrix, the qualitative uncertainty in regional frequency analysis was evaluated. Finally, the qualitative and quantitative uncertainties were combined in an integrated diagnostic diagram. In this study, the proposed NUSAP method proved to be effective in evaluating both the qualitative and quantitative uncertainties of the regional frequency analysis.

The diagnostic diagram is a helpful tool for decision-makers to have an overview of the quality and quantity of the data sources, methods, or models employed within the regional frequency analysis. To minimize the arbitrariness and subjectivity in measuring strength, the pedigree matrix is used to code the qualitative expert judgments. However, there are a few points which should be noted. Experts may have different judgments on which mode of each column of a pedigree matrix best represents the state of knowledge. Besides, they may have different judgments on the score set to the quality of the method or models used. Therefore, formal questionnaires on how to describe and define each column of a pedigree matrix and how to score the statistical, empirical, and methodological quality are proposed in the further study within worldwide experts in the field of uncertainty analysis, hydrology, and water resources.

Meanwhile, only limited data, methods, and models were used in this case study for illustration of the NUSAP method, which resulted in an underestimation of uncertainty in regional frequency analysis. It is highly recommended to make a systematic analysis of both qualitative and quantitative uncertainties in regional frequency analysis.

References

Baeriswyl, P.A., Rebetez, M., 1997. Regionalization of precipitation in Switzerland by means of principal component analysis. *Theoretical and Applied Climatology*, **58**(1-2):31-41. [doi:10.1007/BF00867430]

BOHZJ (Bureau of hydrology, Zhejiang Province), 2003. Atlas of Storms Statistical Parameters for Zhejiang Province Hangzhou, Zhejiang. BOHZJ, Hangzhou (in Chinese).

Bouda, M., Rousseau, A.N., Konan, B., *et al.*, 2012. Bayesian uncertainty analysis of the distributed hydrological model HYDROTEL. *Journal of Hydrologic Engineering*, **17**(9):1021-1032. [doi:10.1061/(ASCE)HE.1943-5584. 0000550]

Castellarin, A., Burn, D., Brath, A., 2001. Assessing the effectiveness of hydrological similarity measures for flood frequency analysis. *Journal of Hydrology*, **241**(3-4): 270-285. [doi:10.1016/S0022-1694(00)00383-8]

Delsman, J.R., Essink, G.H.P.O, Beven, K.J., *et al.*, 2013. Uncertainty estimation of end-member mixing using generalized likelihood uncertainty estimation (GLUE), applied in a lowland catchment. *Water Resources Research*, **49**(8):4792-4806. [doi:10.1002/wrcr.20341]

Farsadnia, F., Kamrood, M.R., Nia, A.M., *et al.*, 2014. Identification of homogeneous regions for regionalization of watersheds by two-level self-organizing feature maps. *Journal of Hydrology*, **509**:387-397. [doi:10.1016/j.jhydrol.2013.11.050]

Fowler, H., Kilsby, C., 2003. A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. *International Journal of Climatology*, **23**(11):1313-1334. [doi:10.1002/joc.943]

Funtowicz, S.O., Ravetz, J.R., 1990. Uncertainty and Quality in Science for Policy. Springer Science & Business Media. [doi:10.1007/978-94-009-0621-1]

Hosking, J.R.M., Wallis, J.R., 2005. Regional Frequency Analysis: an Approach Based on L-moments. Cambridge University Press, UK.

Houska, T., Multsch, S., Kraft, P., *et al.*, 2013. Monte Carlo based calibration and uncertainty analysis of a coupled plant growth and hydrological model. *Biogeosciences Discussions*, **10**(12):19509-19540. [doi:10.5194/bgd-10-19509-2013]

Ilorme, F., Griffis, V.W., 2013. A novel procedure for delineation of hydrologically homogeneous regions and the classification of ungauged sites for design flood estimation. *Journal of Hydrology*, **492**:151-162. [doi:10.1016/j.jhydrol.2013.03.045]

Jeremiah, E., Sisson, S., Marshall, L., *et al.*, 2011. Bayesian calibration and uncertainty analysis of hydrological models: a comparison of adaptive metropolis and sequential Monte Carlo samplers. *Water Resources Research*, **47**(7):W07547. [doi:10.1029/2010WR010217]

Jin, X., Xu, C.Y., Zhang, Q., *et al.*, 2010. Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. *Journal of Hydrology*, **383**(3-4):147-155. [doi:10.1016/j.jhydrol.2009.12.028]

Li, L., Xia, J., Xu, C.Y., *et al.*, 2010. Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models. *Journal of Hydrology*, **390**(3-4):210-221. [doi:10.1016/j.jhydrol.2010.06.044]

Lin, G.F., Chen, L.H., 2006. Identification of homogeneous regions for regional frequency analysis using the self-organizing map. *Journal of Hydrology*, **324**(1-4):1-9. [doi:10.1016/j.jhydrol.2005.09.009]

Michele, C.D., Rosso, R., 2001. Uncertainty assessment of regionalized flood frequency estimates. *Journal of Hydrologic Engineering*, **6**(6):453-459. [doi:10.1061/(ASCE)1084-0699(2001)6:6(453)]

Norbiato, D., Borga, M., Sangati, M., *et al.*, 2007. Regional frequency analysis of extreme precipitation in the eastern Italian Alps and the August 29, 2003 flash flood. *Journal of Hydrology*, **345**(3-4):149-166. [doi:10.1016/j.jhydrol.2007.07.009]

Parent, E., Bernier, J., 2003. Bayesian POT modeling for historical data. *Journal of Hydrology*, **274**(1-4):95-108. [doi:10.1016/S0022-1694(02)00396-7]

Ramachandra Rao, A., Srinivas, V., 2006. Regionalization of watersheds by hybrid-cluster analysis. *Journal of Hydrology*, **318**(1-4):37-56. [doi:10.1016/j.jhydrol.2005.06.003]

Reis, D.S.Jr., Stedinger, J.R., 2005. Bayesian MCMC flood frequency analysis with historical information. *Journal of Hydrology*, **313**(1-2):97-116. [doi:10.1016/j.jhydrol.2005.02.028]

Saf, B., 2010. Assessment of the effects of discordant sites on regional flood frequency analysis. *Journal of Hydrology*, **380**(3-4):362-375. [doi:10.1016/j.jhydrol.2009.11.011]

SL21-90, 1991. Precipitation Observation Criteria. The Ministry of Water Resources of the People's Republic of China (in Chinese).

van der Sluijs, J.P., Craye, M., Funtowicz, S., *et al.*, 2005. Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment: the NUSAP system. *Risk Analysis*, **25**(2):481-492. [doi:10.1111/j.1539-6924.2005.00604.x]

Wiltshire, S., 1986. Regional flood frequency analysis I: homogeneity statistics. *Hydrological Sciences Journal*, **31**(3):321-333. [doi:10.1080/02626668609491051]

Xu, Y., Xu, X., Huang, Y., *et al.*, 2011. Uncertainty assessment of flood frequency analysis in Hanjiang Basin (I): uncertainty in discharge measurement stage. *Journal of Basic Science and Engineering*, **19**(S1):177-183 (in Chinese). [doi:10.3969/j.issn.1005-0930.2011.s1.020]

Zaman, M.A., Rahman, A., Haddad, K., 2012. Regional flood frequency analysis in arid regions: a case study for Australia. *Journal of Hydrology*, **475**:74-83. [doi:10.1016/j.jhydrol.2012.08.054]

中文概要

题　目： 区域降雨频率分析中的质量和数量不确定性分析

目　的： 通过引进 NUSAP 方法量化区域降雨频率分析不确定性来源中质量方面的不确定性，并结合数量方面的不确定性，分析这些不确定性对降雨频率分析的影响，为水资源风险决策和水利工程设计等提供更好的指导。

创新点： 总结区域降雨频率分析中的不确定性来源，并在区域频率分析中引进 NUSAP 方法用以量化其质量不确定性，针对区域频率分析提出 Pedigree 矩阵。

方　法： 1. 选取区域频率分析中三个主要不确定性来源，即降雨测量不确定性、水文分区不确定性和分布线型的不确定性；2. 提出针对区域频率分析的评价依据 Pedigree 矩阵，量化区域频率分析中的质量不确定性；3. 将质量和数量两类不确定性结合在不确定性诊断图中，综合评估区域频率分析中的质量不确定和数量不确定性。

结　论： NUSAP 方法可以有效地量化区域降雨频率分析中的质量不确定性，并通过不确定性诊断图将质量不确定和数量不确定性很好地结合起来，为水资源风险决策和水利工程设计等提供了直观的方案。

关键词： 质量不确定性；不确定性分析；NUSAP 方法；区域频率分析；Pedigree 矩阵；不确定性诊断图