



Article

Cloud Transform Algorithm Based Model for Hydrological Variable Frequency Analysis

Xia Bai ¹, Juliang Jin ^{1,2}, Shaowei Ning ^{1,2}, Chengguo Wu ^{1,2,*}, Yuliang Zhou ^{1,2}, Libing Zhang ^{1,2} and Yi Cui ^{1,2}

¹ School of Civil Engineering, Hefei University of Technology, Hefei 230009, Anhui, China; 2021800032@hfut.edu.cn (X.B.); jinjl66@hfut.edu.cn (J.J.); ning@hfut.edu.cn (S.N.); zhouyuliang@hfut.edu.cn (Y.Z.); zhangyuliang@mail.hfut.edu.cn (L.Z.); ycui@hfut.edu.cn (Y.C.)

² Institute of Water Resources and Environmental Systems Engineering, Hefei University of Technology, Hefei 230009, Anhui, China

* Correspondence: wucguo@knights.ucf.edu

Abstract: Hydrological variable frequency analysis is a fundamental task for water resource management and water conservancy project design. Given the deficiencies of higher distribution features for the upper tail section of hydrological variable frequency curves and the corresponding safer resulting design of water conservancy projects utilizing the empirical frequency formula and Pearson type III function-based curve fitting method, the normal cloud transform algorithm-based model for hydrological variable frequency analysis was proposed through estimation of the sample empirical frequency by the normal cloud transform algorithm, and determining the cumulative probability distribution curve by overlapping calculation of multiple conceptual cloud distribution patterns, which is also the primary innovation of the paper. Its application result in northern Anhui province, China indicated that the varying trend of the cumulative probability distribution curve of annual precipitation derived from the proposed approach was basically consistent with the result obtained through the traditional empirical frequency formula. Furthermore, the upper tail section of the annual precipitation frequency curve derived from the cloud transform algorithm varied below the calculation result utilizing the traditional empirical frequency formula, which indicated that the annual precipitation frequency calculation result utilizing the cloud transform algorithm was more optimal compared to the results obtained by the traditional empirical frequency formula. Therefore, the proposed cloud transform algorithm-based model was reliable and effective for hydrological variable frequency analysis, which can be further applied in the related research field of hydrological process analysis.

Keywords: hydrological variable frequency analysis; cloud transformation algorithm; empirical frequency; Pearson type III frequency curve; annual precipitation; northern Anhui province



Citation: Bai, X.; Jin, J.; Ning, S.; Wu, C.; Zhou, Y.; Zhang, L.; Cui, Y. Cloud Transform Algorithm Based Model for Hydrological Variable Frequency Analysis. *Remote Sens.* **2021**, *13*, 3586. <https://doi.org/10.3390/rs13183586>

Academic Editors: Pingping Luo, Xindong Wei, Kanhua Yu, Bin Guo and Joshua Viers

Received: 7 August 2021

Accepted: 3 September 2021

Published: 9 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The distribution pattern of hydrological variables, which frequently consists of multiple certainty elements and stochastic factors, is the comprehensive outcome of numerous influencing factors, including meteorological climate, natural geography, human activities, etc. [1,2]. Thus, the distribution characteristic exploration of hydrological variables is known as an effective approach to reveal the response mechanism of various influencing factors on hydrological process evolution [3,4]. Meanwhile, as an important tool to describe the natural characteristics of hydrological process evolution from spatiotemporal variation perspectives based on hydrology, statistics, meteorology, and mathematical principles, hydrological variable frequency analysis primarily concentrates on the exploration of statistical rules of hydrological events and relationship description between hydrological variable distribution and corresponding designed frequency of water conservancy projects [3], and the related research results can provide a reasonable and scientific decision-making basis for the construction of water conservancy projects [5,6].

Overall, hydrological variable frequency analysis consists of research that includes sample selection and data verification, calculation of sample empirical frequency, selection of frequency distribution pattern, parameter estimation of frequency distribution function, verification of frequency distribution function, determination of designed results of the hydrological variable and corresponding uncertainty analysis, etc. [7–9], and the calculation of sample empirical frequency and determination of frequency distribution function are the key issues of hydrological variable frequency analysis. Generally, According to the Regulation for calculating design flood of water resources and hydropower projects in China, the hydrological variable frequency curve can be derived through the Pearson III function-based curve-fitting method. Specifically, (1) in terms of the determination of hydrological frequency distribution function, the actual frequency distribution function of a hydrological variable so far still cannot be obtained from the perspective of the hydrological evolution mechanism [7,10]. In practice, based on the selection of a certain type of frequency distribution function and parameter estimation through historical observed data series, the frequency distribution function of hydrological variables can be derived according to the fitting results of the empirical and theoretical frequency of observed historical samples when the curve-fitting error is acceptable, which essentially is an approximate determination approach of the hydrological variable frequency distribution function [7,11]. In particular, the curve-fitting error for the upper and lower tail sections (i.e., external extending parts) of the frequency curve cannot be neglected, because the top tail section of the hydrological frequency curve (corresponding to low occurring possibility) is always related to the occurrence of extreme hydro-meteorological events. (2) As for the calculation of sample empirical frequency of hydrology variables, the estimation results of sample empirical frequency would be different depending on the selection of frequency distribution function [4,11]. Therefore, the sample empirical frequency would be the primary factor that affects the frequency-curve-fitting error, except for the frequency distribution function and curve-fitting principles [12]. Since Hazen first proposed the empirical frequency formula in 1914, many types of calculation formulas of empirical frequency have been reported around the world [7,12]. Afterwards, a great number of works have been carried out focusing on the improvement of the empirical frequency calculation formula of hydrological variables; for instance, V.P. Singh proposed the parameter estimation method of frequency distribution curve, according to the principle of maximum entropy [13,14]. P. Xie constructed the calculation framework of hydrological frequency analysis for inconsistent annual runoff series affected by frequent human activities and climate change [4]. Y. Li summarized the latest research progress of hydrological frequency analysis and application prospects of the probability weighted method, layered Archimedes copulas function, and Bayesian theory in the field of hydrological frequency analysis [15]. M. Zhang proposed the maximum entropy-principle-based method to determine the residual probability distribution function of an annual runoff stochastic model, and recommended the simulation procedures of a pure random series of maximum entropy distribution through rejection method [10]. K. Qin and G. Y. Zhang obtained a frequency distribution curve of stochastic variables through vast scattered historical samples, and then testified the rationality of the frequency-curve-fitting results of stochastic variables [16,17].

To sum up, it could be revealed from the previous studies that the difference of empirical frequency estimation results calculated by different approaches is not obvious for the middle part of the frequency fitting curve, but is quite evident for both sides (especially the top tail section), which will undoubtedly lead to an inaccurate designing outcome of hydrological variables. Therefore, the normal cloud transform algorithm, which primarily concentrates on the quantitative description of the cognitive process of qualitative fuzzy concept by human intelligence, was innovatively introduced in the hydrological frequency analysis field in this study. Specifically, based on the obtained probability density curve of hydrological series through frequency histogram, firstly, the initial probability density curve was transformed into multiple normal cloud distribution patterns with different granularities; then, the probability density curve of hydrological variables was derived

through overlaying and merging calculation of a different normal cloud distribution; and, finally, the framework of the normal cloud transform algorithm-based model (CTAM) for hydrological variable frequency analysis was proposed, which was further applied in the frequency-curve-fitting calculation of the annual precipitation variable from different hydrological stations in the northern Anhui province, China. The application results indicated that the frequency-curve-fitting result of annual precipitation in northern Anhui province based on the proposed CTAM model was basically consistent with the calculation result obtained through the traditional empirical frequency formula-based curve-fitting method (EFCF), and the design value of annual precipitation based on the CTAM method was more optimal and reasonable than the corresponding result derived from the EFCF method, especially for low cumulative frequency scenarios, which could provide rigorous basis for the validation of the availability and effectiveness of the proposed CTAM method.

2. Overview of Study Area and Data Collection

The study area is located in the northern Anhui province, which resides on the northern part of the mainstream Huaihe River, with a total area of 42,200 km². As shown in Figure 1, it is also located on the south borders of Shandong, Henan, and Jiangsu provinces, which are within the hinterland of the central economic zone of China. The drought-affected area in this region is approximately 53% of the total provincial arable land area [18–20]. Meanwhile, the northern Anhui province lies in the transitional zone of the subtropical zone and warm temperate zone, with obvious transitional climate characteristics. The southeast wind prevails in spring and summer, and the climate is warm and humid; meanwhile, the northwest wind prevails in autumn and winter, and the climate is cold and dry. The annual average precipitation in this region is about 850 mm. The seasonal allocation and temporal distribution of precipitation vary significantly due to the influence of global climate change, which leads to the alternation of cold and warm seasons and the occurrence of drought disasters more frequently as well in the northern Anhui province.

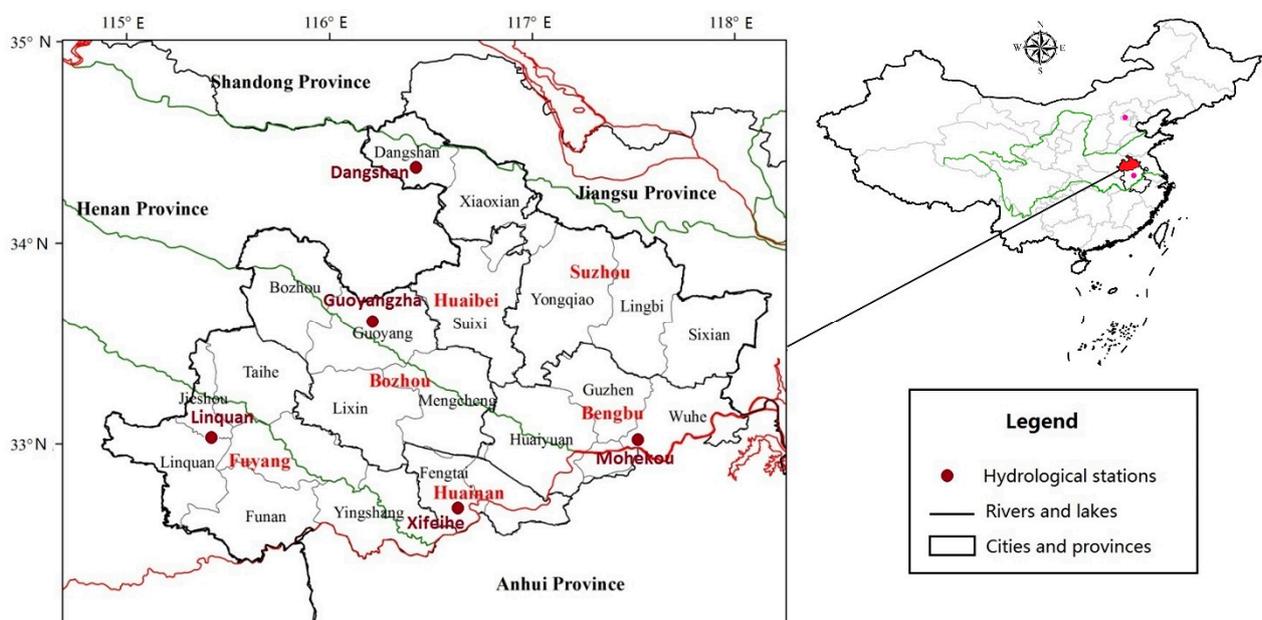


Figure 1. Distribution of regions and observation sites in northern Anhui province, China.

In addition, the historical annual precipitation data series over the period from 1954 to 2010 of the five hydrological stations in the northern Anhui province (Linquan, Guoyangzha, Dangshan, Xifeihe, and Mohekou) was downloaded from National Meteorological Information Center, which was also employed to further testify the feasibility

and effectiveness of the proposed cloud transform algorithm-based model (CTAM) for hydrological variable frequency analysis.

3. Methodologies

3.1. Cloud Model and Cloud Transform Algorithm

The cloud model, proposed by Chinese scholar D.Y. Li in 1995, is an effective mathematical cognitive tool for the exploration of uncertain transformation mechanisms from a qualitative concept to its quantitative expression, utilizing fuzzy mathematics and probability statistics theories [21]. The cloud model can combine the randomness of a variable sample with the fuzziness of concept recognition, and then realize the natural transformation from a qualitative concept to its quantitative expression [21]. The definition of cloud model can be explained as follows [21,22].

Supposing U to be a universal set denoted by precise sample data and C to be the qualitative concept related to U , if quantitative value x ($x \in U$) is a random sample of the concept C and satisfies $x \sim N(Ex, En'^2)$, $En' \sim N(En, He^2)$, then the membership degree of quantitative value x belonging to concept C is defined as the certainty degree, as follows:

$$\mu(x) = e^{-\frac{(x-Ex)^2}{2En'^2}} \quad (1)$$

where the distribution of variable x in universal set U is defined as normal cloud, which can be denoted as $C(Ex, En, He)$, and each data point $(x, \mu(x))$ is called cloud drop. Meanwhile, the distribution characteristics of the cloud model are described by three cloud parameters, average Ex is the distribution center of qualitative concept C , which can best represent the properties of the concept C , entropy En is used to represent the fuzziness of concept C , reflecting the range of sample values accepted by concept C , and hyper-entropy He represents the uncertainty of entropy En , revealing the concentration of samples under certain deviation and making cloud drops uniformly distributed on both sides [21,23,24].

As a significant approach of probability statistics, the innovation of Gauss transformation is dividing the probability density function of a random variable into a multiple Gauss distribution, which could provide realization of ideas for the division and overlay of the multi-granularity concept [24]. By statistical transformation using mathematical principles, the cloud transform algorithm can be applied to represent the distribution features of scattered sample series through several cloud distribution patterns [25–27]. It is always necessary to describe the qualitative concept from different perspectives; human intelligence could realize logical reasoning from different granularities, and promote the improvement of concept cognition level from lower granularity to higher granularity [24,28,29]. Therefore, the cloud transform process can be essentially deemed as gradual recognition and description of the concept by human intelligence from the actual sample distribution universe [23]. It has been theoretically verified that the probability density function of a hydrological variable can be divided into several normal distributions [7,23,30]. Therefore, the probability density function of a hydrological variable could be described with the combination of multiple normal cloud distribution, denoted as $C_i(Ex_i, En_i, He_i)$, and each cloud distribution represents a certain discrete qualitative concept [21,24]. Thus, the process of the cloud transform algorithm can be represented as follows:

$$g(x) = \sum_{i=1}^n (a_i \cdot f_i(x)) + \delta(x) \quad (2)$$

where $g(x)$ denotes the probability density function of variable x , $f_i(x)$ represents the expectation function of scattered normal cloud distribution, corresponding to the qualitative concept of different granularities, i and n denote the serial and total number of scattered normal cloud distribution patterns, a_i represents the amplitude coefficient of i th scattered cloud distribution, and $\delta(x)$ is the fitting error function, which can be obtained through iterative calculation using the characteristic parameters of scattered cloud distribution [23,24,26].

From the perspective of data mining, the cloud transform calculation can be considered as a gradual extraction process of a more refined concept from the actual sample distribution of a certain attribution of the rough concept. Furthermore, for a certain hydrological variable, the contribution of samples corresponding to a higher peak value of the probability density curve of qualitative concept is more significant than that of the samples corresponding to a lower peak value. Therefore, firstly, the first local maximum value of the probability density curve was considered as the qualitative concept center (which is the average of the normal cloud model). Secondly, samples corresponding to the qualitative concept center were excluded from the original probability density curve, and the next local maximum value could be obtained from the probability density curve derived from the rest of the samples, which could be deemed as the center of the next qualitative concept. Thirdly, the above steps were repeated until the occurring probability of the rest of the samples decreased below a certain threshold value, and, finally the exploration process of the cloud transform algorithm for qualitative concept recognition from the original sample distribution was accomplished [16,17,23]. In the next section of the manuscript, taking the cumulative probability distribution frequency curve fitting of annual precipitation as an example, the calculation process of the cloud transform algorithm is more explicitly explained.

3.2. Calculation Procedures of the CTAM Method for Hydrological Variable Frequency Analysis

Considering the deficiencies of the Pearson type III function-based approach for hydrological frequency analysis and disagreement of different estimation formulas of sample empirical frequency, the CTAM model for annual precipitation frequency-curve-fitting analysis was recommended in this manuscript. More specifically, the realization procedures of the CTAM model for the cumulated distribution frequency-curve-fitting of annual precipitation could be explained as follows [17,23]:

Step 1: Determination of the probability density function $g(x)$ of annual precipitation. Firstly, the observed historical annual precipitation series was sorted from highest to lowest, and the actual distribution interval could be divided into several subintervals with constant step ΔP . Then, the average probability density of each subinterval was obtained through frequency histogram of annual precipitation series, which varied more and more closely to the median probability density when the sample capacity increased and calculation step decreased gradually. Finally, the probability density frequency curve of annual precipitation, denoted as $g(x)$, could be obtained using the median series of different subintervals as the X-axis and corresponding average probability density as the Y-axis.

Step 2: Determination of scattered conceptual cloud distribution of annual precipitation based on the cloud transform algorithm. Firstly, the peak of the probability density frequency curve of annual precipitation $g(x)$ was defined as the center of current scattered conceptual cloud, and the central samples corresponding to the peak value as the average Ex_i (i was the number of conceptual cloud, $i = 1 \sim n$) of the current conceptual cloud. Then, starting from the average Ex_i of the first conceptual cloud distribution, the amplitude coefficient a_i of sample variation between the first maximum and closest local minimum on both sides could be determined, and the sample distribution belonging to the current conceptual cloud could be obtained as well when the amplitude coefficient a_i was greater than a certain threshold value σ_1 , otherwise, the local minimum value was updated until the amplitude coefficient a_i was greater than threshold value σ_1 . Finally, the normal backward cloud algorithm without certainty degree could be applied to determine the entropy En_i and hyper-entropy He_i for each scattered conceptual cloud [21,24], by which the corresponding expectation curve could be considered as the probability density function $f_i(x)$ of each conceptual cloud [17,24,31].

Step 4: Overlapping calculation of the normal conceptual cloud for annual precipitation probability density function. As indicated from step 1 to step 3, it was evident that numerous computations and adjustments were needed during the determination of a different scattered conceptual cloud and its corresponding probability density function. In

addition, through the adjustment of entropy En_i and hyper-entropy He_i for each scattered conceptual cloud, the probability density function curve $g(x)$ of annual precipitation could be obtained when the curve-fitting error of overlapping calculation was lower than a certain threshold value σ_2 ; the overlapping calculation formula of the conceptual cloud could be denoted as follows [24]:

$$g(x) = \sum_{i=1}^n a_i \cdot f_i(x) \quad (3)$$

Step 5: Cumulative probability distribution function-curve-fitting analysis of annual precipitation. Theoretically, the integral function form of the probability density function $g(x)$ of annual precipitation is the cumulative probability distribution function. In practice, the cumulative probability distribution function of annual precipitation could be determined according to the frequency-curve-fitting method based on the Pearson type III function through calculation of the empirical frequency of reselected samples through the integral function of probability density function $g(x)$.

Among the above steps, it was crucial for the adjustment of the entropy value of the different conceptual cloud, which could be accomplished by applying a heuristic method [17,32,33]. More explicitly, based on the determination for the average Ex_i of the current conceptual cloud, the first local minimum value was selected from both the right and left side of median average Ex_i , and then the amplitude coefficient between the central sample corresponding to the median average Ex_i and the sample relating to the selected local minimum was also calculated. If the amplitude coefficient was greater than a certain threshold σ_1 , then the variation range, denoted as Ex_i , responding to the local minimum value that was closer to the median average value Ex_i could be considered as the distribution range of current cloud drops, and the samples varying within the interval of $[Ex_i - Ex_i', Ex_i + Ex_i']$ could be considered as the research universe of the current conceptual cloud, and the related entropy En_i and hyper-entropy He_i could be determined using the normal backward cloud algorithm without certainty degree [16,17]. Meanwhile, the parameters of entropy En_i and hyper-entropy He_i also need to be adjusted when the curve-fitting error of the overlapping calculation for the different conceptual cloud was greater than a certain threshold value σ , and the detailed adjusting principles of entropy can be referred from [16,17,21].

4. Results and Discussion

In this section, the proposed CTAM method was further applied in hydrological frequency-curve-fitting analysis of the annual precipitation variable of the five hydrological stations in the northern Anhui province. To be more specific, taking the Guoyangzha station as an example, firstly, the varying range of annual precipitation from 1954 to 2010 was divided into 15 subintervals, with the calculating step of 80 mm, and the histogram of annual precipitation variation was drawn and the average probability density corresponding to each subinterval was determined by counting the number of years with the annual precipitation varying within each subinterval. Then, the probability density curve of historical annual precipitation series was obtained using the median of each subinterval as the X-axis and the average probability density corresponding to each subinterval as the Y-axis. Finally, the derived initial probability density curve of annual precipitation was divided into five normal cloud distribution patterns with different granularities utilizing the cloud transform algorithm. Similarly, the varying ranges of annual precipitation from 1954 to 2010 in Dangshan, Linquan, Mohekou, and Xifeihezha stations were divided into 13, 12, 14, and 12 subintervals, with the calculating steps of 80 mm, 115 mm, 80 mm, and 85 mm, respectively, and the derived probability density curves of annual precipitation corresponding to different stations were divided into 5, 6, 7, and 8 normal cloud distribution patterns. The characteristic parameters and amplitude coefficients of the different cloud distribution patterns in different hydrological stations in northern Anhui province are given in Table 1.

Table 1. Cloud characteristics of normal conceptual clouds in northern Anhui province, China.

Station	Cloud Number	<i>Ex</i>	<i>En</i>	<i>He</i>	Amplitude Coefficient
Guoyangzha	1	678.24	97.43	208.10	0.001974
	2	912.28	36.28	24.26	0.001316
	3	998.40	41.40	29.37	0.001096
	4	1086.9	37.41	16.89	0.000439
	5	1248.85	82.27	104.11	0.000439
Dangshan	1	765.34	57.60	74.49	0.002632
	2	598.24	68.56	138.09	0.001754
	3	923.13	42.09	23.37	0.001316
	4	1017.65	25.32	21.65	0.000439
	5	1243.70	49.07	52.59	0.000439
Linquan	1	920.53	65.66	126.72	0.002288
	2	700.09	73.71	170.34	0.001831
	3	1171.83	76.48	92.27	0.00061
	4	1393.40	34.09	36.54	0.000305
	5	1833.45	59.34	31.02	0.000305
	6	1454.6	79.71	41.67	0.000153
Mohekou	1	958.61	80.27	168.55	0.002516
	2	746.8	29.95	60.96	0.001509
	3	666.98	32.24	52.53	0.001258
	4	1134.87	25.68	11.97	0.000755
	5	513.9	53.53	57.37	0.000503
	6	1251.8	38.77	29.14	0.000503
	7	1407.55	44.75	47.96	0.000503
Xifeihezha	1	877.89	79.58	107.92	0.001858
	2	697.08	37.69	53.75	0.001651
	3	1037.62	81.76	66.61	0.001238
	4	611.53	46.33	67.99	0.000826
	5	455.07	29.47	18.54	0.000619
	6	1219.90	26.57	13.89	0.000413
	7	1362.5	1.13	0.59	0.000413
	8	1549.45	29.64	15.49	0.000413

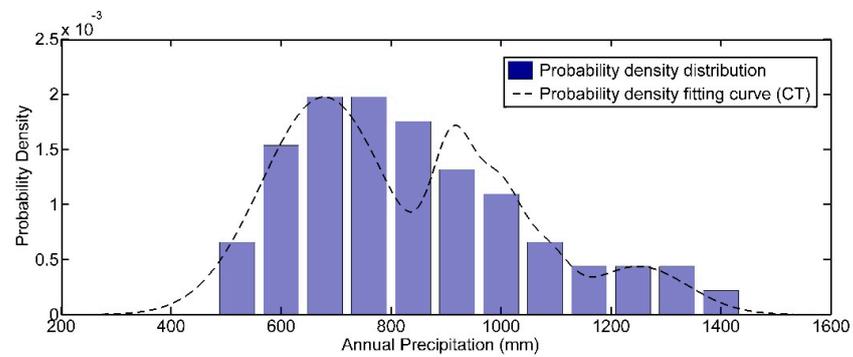
Secondly, the normal cloud-distribution-based conceptual overlaying principle was employed to conduct the conceptual merging calculation of the normal cloud distribution, and then the probability density function curve of annual precipitation for different stations was derived (as shown in Figure 2), by which the ultimate empirical frequency of historical annual precipitation data series for different hydrological stations in northern Anhui province was determined through integral computation. In addition, the correlation coefficient of empirical frequency of historical annual precipitation data series from 1954 to 2011 of the five hydrological stations in northern Anhui province, derived from the traditional empirical frequency formula and cloud transform algorithm, respectively, was above 0.98, which indicated that the empirical frequency series of annual precipitation calculated by the two methods was consistent, and the cloud transform algorithm was also suitable for the determination of the empirical frequency of hydrological variables.

Finally, the Pearson type III function-based curve-fitting method was applied to determine the cumulative probability distribution frequency curve of annual precipitation, based on the historical empirical frequency determined by cloud transform algorithm from 1954 to 2010 of Guoyangzha, Dangshan, Linquan, Mohekou, and Xifeihezha hydrological stations. The statistical parameters of the cumulative probability distribution frequency curve of annual precipitation, determined by both the traditional empirical frequency formula and cloud transform algorithm, respectively, are shown in Table 2, and the trend of the cumulative probability distribution frequency curve of annual precipitation determined by two approaches is indicated in Figure 3 as well.

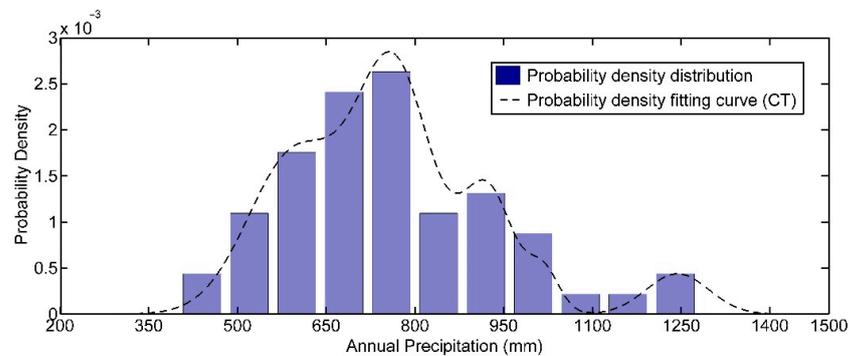
Table 2. Statistical parameters of the annual precipitation frequency curve in northern Anhui province, China.

Station	Name	Statistical Parameters of Curve Fitting		
		Ex	Cv	Cs
Guoyangzha	EFCF	838.83	0.27	1.05
	CTAM	838.83	0.33	1.08
Dangshan	EFCF	760.88	0.25	0.71
	CTAM	760.88	0.22	0.59
Linquan	EFCF	927.10	0.31	1.24
	CTAM	927.10	0.28	1.11
Mohekou	EFCF	927.08	0.24	0.49
	CTAM	927.08	0.24	0.48
Xifeihezha	EFCF	888.37	0.27	0.54
	CTAM	888.37	0.25	0.61

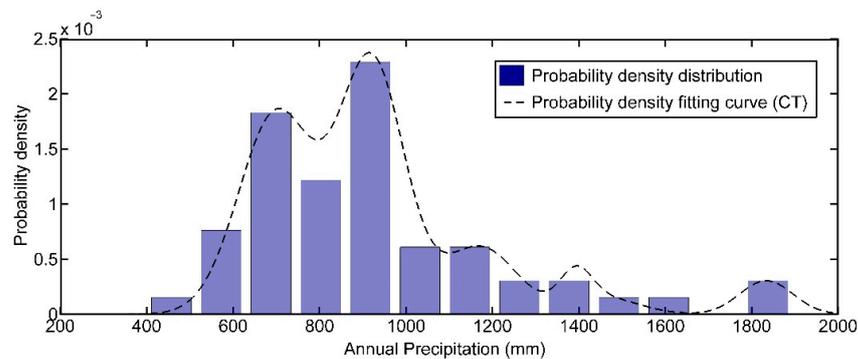
Note: EFCF and CTAM denote frequency curve fitting methods based on empirical frequency and cloud transformation algorithm, respectively.



(a)



(b)



(c)

Figure 2. Cont.

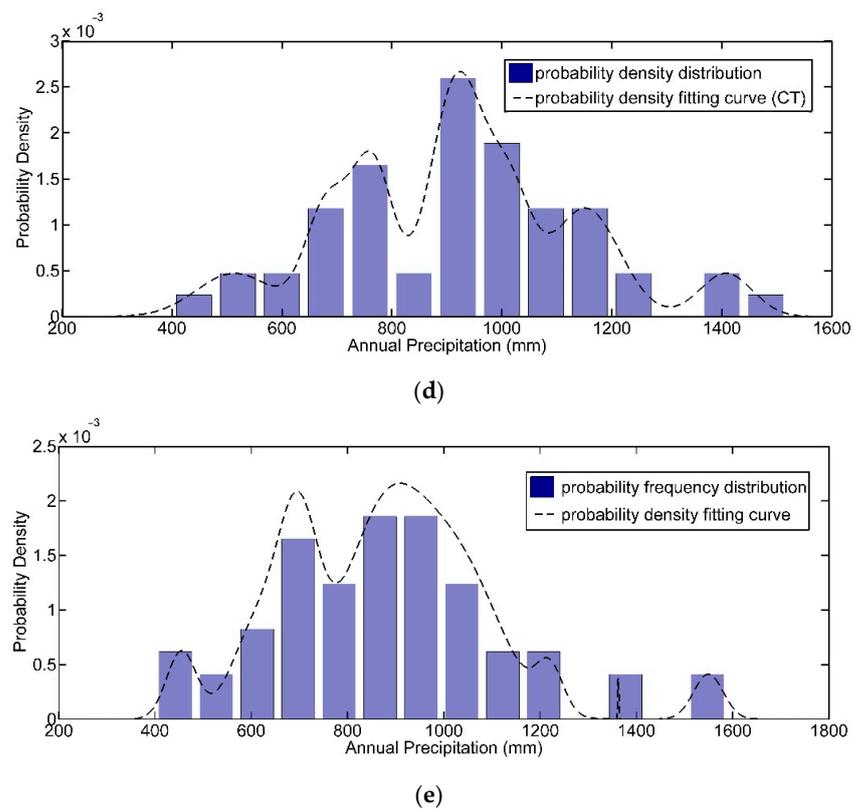


Figure 2. Distribution of probability density fitting curve of annual precipitation in north Anhui province, China. (a) Guoyangzha Station; (b) Dangshan Station; (c) Linquan Station; (d) Mohekou Station; and (e) Xifeihezha Station.

It can be revealed from Table 2 and Figure 3 that (1) the varying trend of the cumulative probability distribution frequency curve of annual precipitation derived from the traditional empirical frequency formula and cloud transform algorithm, respectively, had better consistency for most of the five hydrological stations in northern Anhui province. (2) Except for the Guoyangzha station, the upper tail section (i.e., when the cumulative frequency was lower than 0.2) of the cumulative probability distribution frequency curve of annual precipitation derived from the cloud transform algorithm varied below the corresponding frequency curve derived from the traditional empirical frequency formula, which was opposite to the varying trend of the lower tail section (i.e., when the cumulative frequency was higher than 0.9) of the cumulative probability distribution frequency curve of annual precipitation, and the fitting value of annual precipitation determined by the two approaches when the cumulative frequency varied within the interval of (0.2, 0.9) was almost uniform, which indicated that the CTAM model for hydrological frequency analysis was capable of avoiding the shortage that the upper tail section of the frequency curve and the corresponding designed value of hydrological series were higher when the cumulative frequency varied lower. Therefore, it was evident that the hydrological frequency analysis result utilizing the peak value principle-based cloud transform algorithm was more reasonable, especially for the two varying tails of the cumulative probability distribution frequency curve, and further discussion for the application feasibility of the cloud transform algorithm in hydrological frequency analysis was carried out in terms of the error of frequency-curve-fitting calculation and frequency analysis of typical returning period in the next section of the manuscript.

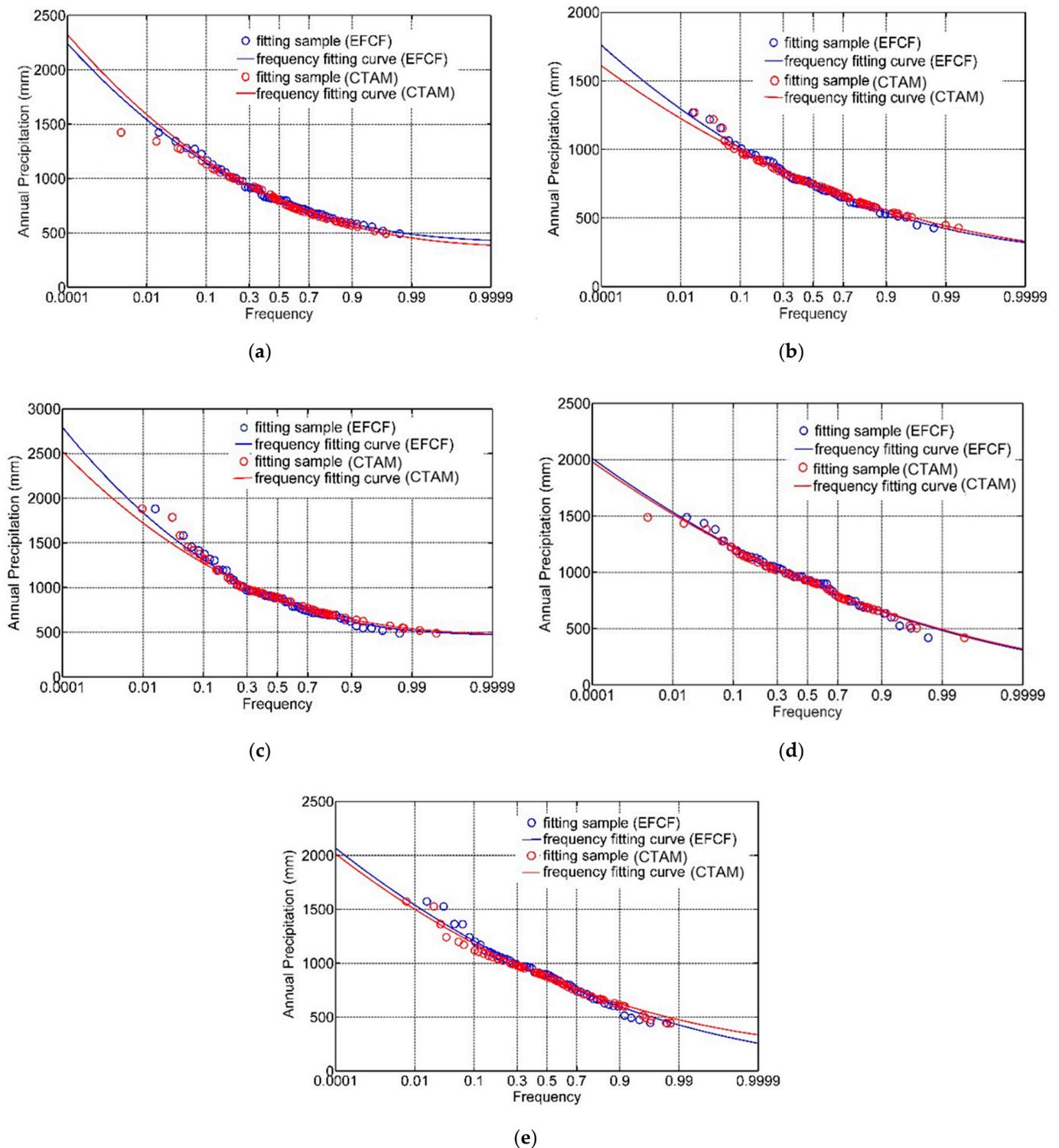


Figure 3. Frequency distribution fitting curve of annual precipitation in northern Anhui province, China. (a) Guoyangzha Station; (b) Dangshan Station; (c) Linquan Station; (d) Mohekou Station; and (e) Xifeihezha Station.

4.1. Fitting Error Analysis of Historical Samples

To further testify the feasibility for the application of peak value theory in the field of hydrological frequency analysis, the index of cumulated relative fitting error (CRFE) [11,30], which reveals the cumulated error values of frequency curve fitting analysis, was employed to comparatively analyze the differences of cumulative probability distribution calculation

results of the annual precipitation series of the five hydrological stations in northern Anhui province utilizing the traditional empirical frequency formula and CTAM method, respectively, and the CRFE index value and corresponding varying trend of the frequency-curve-fitting calculation of annual precipitation in different hydrological stations utilizing the two approaches are indicated in Table 3.

Table 3. Cumulated relative fitting error (CRFE) of annual precipitation in northern Anhui province, China.

Name	Guoyangzha	Dangshan	Linqan	Mohekou	Xifeihezha
EFCF	0.960642	0.992969	0.972185	0.980417	0.980378
CTAM	0.954421	0.978128	0.968479	0.972718	0.972126

It can be revealed from Table 3 that, (1) for the frequency-curve-fitting calculation employing the peak value theory-based cloud transform algorithm, the CRFE index value of cumulative probability distribution frequency analysis of annual precipitation in Dangshan station was the highest, at 0.978128, the CRFE index value of the curve-fitting calculation in Guoyangzha station was the lowest, at 0.954421, and the varying trend of the CRFE index utilizing the cloud transform algorithm was consistent with the corresponding trend of the CRFE index employing the traditional empirical frequency formula. (2) The average of CRFE index for cumulative probability distribution frequency analysis of annual precipitation for different stations in northern Anhui province employing the cloud transform algorithm and traditional empirical frequency formula were 0.969176 and 0.977318, respectively, and the CRFE index of cumulative probability distribution frequency analysis of annual precipitation utilizing the cloud transform algorithm was lower than the corresponding CRFE index for different stations in northern Anhui province. (3) The fitting error and varying trend of cumulative probability distribution frequency analysis of annual precipitation utilizing the two approaches were nearly consistent, which preliminarily verified the feasibility and reasonability for the application of the cloud transform algorithm in the field of hydrological frequency analysis.

4.2. Fitting Error Analysis under Typical Returning Period

The frequency fitting calculation of hydrological variables under a typical returning period scenario (especially for the rare probability condition) is a fundamental task for water conservancy project planning and development [7]. To verify the adaptation of the cloud transform algorithm in the application of the hydrological frequency analysis field, the fitting value of annual precipitation under the typical returning period of the five hydrological stations in northern Anhui province was determined utilizing the cloud transform algorithm and traditional empirical frequency formula, respectively, and the index of relative fitting error (RFE) was also applied to reveal the deviation result of calculated values of annual precipitation through the CTAM method from the corresponding values obtained through the EFCF approach [7], as given in Table 4.

It can be concluded from Table 4 that, (1) except for the Guoyangzha station, the fitting value of annual precipitation determined by the cloud transform algorithm was lower than the corresponding annual precipitation fitting value determined by the empirical frequency formula when the typical returning period was lower than 20%. Meanwhile, the lower the typical returning period of annual precipitation, the higher the deviation of the annual precipitation fitting results determined by cloud transform algorithm and traditional empirical frequency formula, respectively, which was consistent with the varying characteristics of the cumulative probability distribution frequency curve of annual precipitation derived from the cloud transform algorithm and traditional empirical frequency formula, respectively. (2) The deviation of annual precipitation fitting results in Linqan station calculated by the two approaches was the most obvious, and the average RFE for the annual precipitation fitting calculation for different typical returning periods was -9.33% in Linqan station, while the deviation of annual precipitation fitting results

in Mohekou station calculated by the two approaches was the lowest, and the average RFE for the annual precipitation fitting calculation for different typical returning periods was -0.88% in Mohekou station. (3) Therefore, it was evident that the fitting values of annual precipitation of typical returning period conditions determined by the cloud transform algorithm were lower than the corresponding fitting results through the traditional empirical frequency formula, which indicated that the cloud transform algorithm could be applied as a supplementary tool for annual precipitation fitting calculations, especially under rare probability levels, to avoid the shortage of calculation results through the traditional empirical frequency formula. (4) In addition, it also could be seen that the lower the cumulative probability level, the more evident the uncertainty of the annual precipitation fitting calculation, which revealed that there existed significant uncertainty for the top extending tail section of the hydrological frequency fitting curve as a result of the insufficient historical hydrological series, especially for rare hydrological events data, and this was also the key factor leading to the lower accuracy of hydrological frequency fitting calculation.

Table 4. Fitting result of annual precipitation for different returning periods in northern Anhui province, China.

Station	Name	Frequency (%)									
		0.02	0.05	0.10	0.20	0.5	1	2	5	10	20
Guoyangzha	EFCF	2140	2004	1900	1794	1652	1541	1428	1271	1145	1010
	CTAM	2219	2076	1967	1856	1705	1589	1469	1303	1169	1024
	RFE/%	3.70	3.60	3.52	3.41	3.24	3.08	2.86	2.48	2.06	1.43
Dangshan	EFCF	1694	1605	1537	1466	1370	1295	1216	1104	1012	910
	CTAM	1556	1483	1427	1368	1289	1225	1159	1064	985	896
	RFE/%	-8.15	-7.61	-7.16	-6.67	-5.96	-5.35	-4.67	-3.62	-2.66	-1.47
Linquan	EFCF	2653	2466	2324	2180	1987	1838	1685	1477	1312	1136
	CTAM	2405	2250	2131	2010	1847	1721	1592	1414	1272	1118
	RFE/%	-9.33	-8.78	-8.32	-7.80	-7.01	-6.32	-5.52	-4.25	-3.05	-1.54
Mohekou	EFCF	1945	1855	1785	1713	1612	1533	1448	1328	1226	1110
	CTAM	1921	1833	1765	1694	1597	1519	1437	1319	1220	1107
	RFE/%	-1.27	-1.20	-1.14	-1.07	-0.97	-0.89	-0.80	-0.64	-0.50	-0.32
Xifeihezha	EFCF	1993	1894	1816	1737	1627	1539	1447	1316	1206	1081
	CTAM	1941	1844	1769	1691	1585	1501	1412	1287	1183	1065
	RFE/%	-2.60	-2.61	-2.61	-2.60	-2.56	-2.51	-2.42	-2.21	-1.95	-1.50

Note: EFCF and CTAM denote annual precipitation probability empirical frequency fitting method and cloud transformation fitting method, respectively.

4.3. Variation Characteristics Analysis of Annual Precipitation Series

To further explore the spatial distribution characteristics of annual precipitation in the northern Anhui province through the cumulative probability distribution frequency curve of annual precipitation derived from the cloud transform algorithm, firstly, if referring to the classification standard of hydrological year type based on river flow frequency [24], the historical hydrological years of different stations in the northern Anhui province could be divided into five categories based on the variation of annual precipitation frequency p , including wet year ($p < 12.5\%$), partial wet year ($12.5\% < p \leq 37.5\%$), normal year ($37.5\% < p \leq 62.5\%$), partial dry year ($62.5\% < p \leq 87.5\%$), and dry year ($p > 87.5\%$). Then, the estimated value of annual precipitation corresponding to the frequency threshold for a certain hydrological year type could be determined for different stations in northern Anhui province through the cumulative probability distribution frequency curve of annual precipitation derived from the cloud transform algorithm. Finally, the statistical result for the variation of annual precipitation of different stations in northern Anhui province could be obtained, as shown in Table 5.

Table 5. Statistical dividing standard of types of hydrological year based on annual precipitation in northern Anhui province.

Station	Wet Year	Partial Wet Year	Normal Year	Partial Dry Year	Dry Year
Guoyangzha	$p > 1123$	$875 < p \leq 1123$	$726 < p \leq 875$	$581 < p \leq 726$	$p < 581$
Dangshan	$p > 958$	$799 < p \leq 958$	$693 < p \leq 799$	$574 < p \leq 693$	$p < 557$
Linquan	$p > 1223$	$962 < p \leq 1223$	$808 < p \leq 962$	$661 < p \leq 808$	$p < 661$
Mohekou	$p > 1185$	$981 < p \leq 1185$	$841 < p \leq 981$	$680 < p \leq 841$	$p < 580$
Xifeihezha	$p > 1146$	$937 < p \leq 1146$	$798 < p \leq 937$	$644 < p \leq 798$	$p < 644$

Based on the dividing standard of types of hydrological year indicated in Table 5, it could be revealed that, (1) through 1954 to 2010, the proportion of wet years of Guoyangzha, Dangshan, Linquan, Mohekou, and Xifeihezha stations were 12.28%, 14.04%, 14.04%, 11.32%, and 12.28%, respectively, and the proportion of dry years of the different stations were 7.02%, 12.28%, 14.04%, 13.21%, and 15.79%, respectively. Thus, the difference for the proportion of dry years of different stations in northern Anhui province was obvious, with the maximum of 15.79% in Xifeihezha station and minimum of 7.02% in Guoyangzha station, which indicated that the proportion of dry years in Huainan city, represented by Xifeihezha station, was the highest in northern Anhui province. (2) It has been mentioned that the northern Anhui province lies in the transitional zone of the subtropical zone and warm temperate zone, with obvious transitional climate characteristics, which has resulted in the situation of severe drought hazard evolution in the northern area as compared to the southern area of the study region. In addition, the minimum annual precipitation of a wet year was 958 mm, and the maximum annual precipitation of a dry year was 557 mm in Dangshan hydrological station, located in the northernmost Anhui province, which was consistent with the actual hydrological situation of serious water resource shortage and severe drought loss in historical years in Suzhou city, represented by Dangshan station.

In conclusion, a new hydrological frequency analysis approach based on the cloud transform algorithm was proposed, and its feasibility and reasonability were also testified in terms of hydrological frequency curve fitting calculation in historical years and hydrological variable estimation calculation for typical returning periods, and the application performance of the proposed CTAM approach is affected by many factors, such as the stationarity or variation trend, tail characteristics for the distribution of hydrological variables, the higher regularity for the normal distribution dividing patterns of the hydrological variables, and the high calculation accuracy for the frequency curve fitting analysis. In a word, the proposed CTAM approach can effectively improve the shortages of higher estimation results of hydrological variables for the rare lower frequency scenario utilizing the traditional empirical frequency formula, which can be considered as an effective attempt to extend the application field of cloud theory in hydrological variable frequency analysis.

5. Conclusions

The primary innovation of this study is to introduce the normal cloud transform algorithm in hydrological variable frequency analysis, by which the sample empirical frequency for frequency curve fitting calculation can be determined instead of utilizing the traditional empirical frequency formula. Through theoretical exploration for the establishment of the CTAM model for hydrological frequency analysis and its application research in northern Anhui province, the primary conclusions can be summarized as follows:

- (1) The varying trend of the annual precipitation frequency curve derived from CTAM and EFCF methods had better consistency for most of the hydrological stations in northern Anhui province, and the upper tail section of the annual precipitation frequency curve derived from the CTAM model varied below the corresponding frequency curve derived from the EFCF approach, which indicated that the CTAM model was capable of avoiding the shortage of higher distribution of the upper tail section and the larger designed value of hydrological variable leading by the traditional empirical frequency formula.

- (2) The annual precipitation frequency analysis result for the five stations utilizing the CTAM model was basically consistent with the corresponding frequency analysis result using the traditional empirical frequency formula, the average of the CRFE index for annual precipitation frequency analysis of different stations employing the CTAM model and EFCF were 0.969176 and 0.977318, respectively, and the fitting error and its varying trend of annual precipitation frequency analysis were nearly consistent.
- (3) In terms of spatial distribution of annual precipitation, the proportion of dry years in Huainan city, represented by Xifeihezha station, was the highest in northern Anhui province, and the minimum in wet year and maximum in dry year of annual precipitation were 958 mm and 557 mm, respectively, in Dangshan station, located in the northernmost Anhui province, which was consistent with the actual hydrological situation of serious water resource shortages and severe drought loss in historical years in Suzhou city, represented by Dangshan station.

To summarize, the CTAM model for hydrological frequency analysis is reliable and reasonable, which was testified as an effective attempt to extend the application field of cloud model theory. However, the key for the application of the cloud transform algorithm is how to derive the cloud characteristic parameters from observed sample distribution and realize the overlapping calculation of multiple cloud distribution. Furthermore, too many challenges still exist for the wide application of cloud model theory in hydrological process research, such as the distinct description for the distribution pattern of hydrological variables using cloud model parameters, the scientific explanation for the theoretical mechanism of the conceptual cloud transform process, etc.

Author Contributions: Conceptualization, J.J. and Y.C.; methodology, C.W., S.N. and X.B.; validation, C.W. and Y.Z.; formal analysis, X.B.; investigation, S.N. and L.Z.; data curation, S.N.; writing—original draft preparation, C.W. and X.B.; writing—review and editing, X.B.; supervision, C.W. and J.J.; funding acquisition, S.N. and Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (Grant No. 2017YFC1502405), the National Natural Science Foundation of China (Grant No. 51779067 and 51709071), and Fundamental Research Funds for the Central Universities of China (Grant No. JZ2020HGQA0202).

Data Availability Statement: The data presented in this study are available on reasonable from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Luo, P.P.; Mu, D.R.; Xue, H.; Thanh, N.; Kha, D.; Takara, K.; Nover, D.; Schladow, G. Flood inundation assessment for the Hanoi Central Area, Vietnam under historical and extreme rainfall conditions. *Sci. Rep.* **2018**, *8*, 1–6. [[CrossRef](#)] [[PubMed](#)]
2. Huo, A.D.; Peng, J.B.; Cheng, Y.X.; Luo, P.P.; Zhao, Z.X.; Zheng, C.L. Hydrological analysis of loess plateau highland control schemes in Dongzhi Plateau. *Front. Earth Sci.* **2020**, *8*, 1–14. [[CrossRef](#)]
3. Lyu, J.Q.; Mo, S.H.; Luo, P.P.; Zhou, M.M.; Shen, B.; Nover, D. A quantitative assessment of hydrological responses to climate change and human activities at spatiotemporal within a typical catchment on the Loess Plateau, China. *Quat. Int.* **2019**, *527*, 1–11. [[CrossRef](#)]
4. Xie, P.; Chen, G.C.; Xia, J. Hydrological frequency calculation principle of inconsistent annual runoff series under Changing environments. *Eng. J. Wuhan Univ.* **2005**, *38*, 6–15.
5. Zhu, Y.H.; Luo, P.P.; Zhang, S.; Sun, B. Spatiotemporal analysis of hydrological variations and their impacts on vegetation in semiarid areas from multiple satellite data. *Remote Sens.* **2020**, *12*, 4177. [[CrossRef](#)]
6. Hong, X.J. *Hydrological Drought Indices and Frequency Analysis Methods and Their Applications*; Wuhan University: Wuhan, China, 2017.
7. Song, S.B. Research challenges and suggestions of hydrological frequency calculation. *J. Water Resour. Archit. Eng.* **2019**, *17*, 12–18.
8. Jin, G.Y. Empirical frequency in Hydrological analysis. *Hydrology* **1994**, *1*, 1–9.
9. Jin, G.Y. A review of hydrologic frequency analysis. *Adv. Water Sci.* **1999**, *10*, 319–327.
10. Zhang, M.; Jin, J.L.; Wang, G.Q.; Zhou, R.J. Drought frequency analysis using stochastic simulation with maximum entropy model. *J. Hydroelectr. Eng.* **2013**, *32*, 101–106.

11. Song, S.B.; Kang, Y.; Song, X.Y.; Wang, X.J.; Jin, J.L. *Frequency Calculation Principles and Their Application of Single Variable Hydrological Sequence*; Science Press: Beijing, China, 2018.
12. Guo, S.L.; Ye, S.Z. Discussion of empirical frequency in hydrology calculation. *J. Wuhan Univ. Hydraul. Electr. Eng.* **1992**, *25*, 38–45.
13. Singh, V.P. *Entropy-Based Parameters Estimation in Hydrology*; Kluwer Academic Publishers: London, UK, 1998.
14. Huang, Z.P.; Sa, D.Y.; Huang, C.X.; Ma, J.J. Study on formulas for calculation of empirical frequency in curve-fitting method. *Adv. Sci. Technol. Water Resour.* **2002**, *5*, 69.
15. Li, Y. *Research on New Theory and Application of Hydrologic Frequency Analysis*; Xi'an, Northwest A&F University: Xianyang, China, 2013.
16. Qin, K.; Wang, P. New method of curve fitting based on cloud transform. *Comput. Eng. Appl.* **2008**, *44*, 56–74.
17. Zhang, G.Y. *Study on Prediction Interval Estimation and Generation Scheduling Method for Power System with Large-Scale Wind Power Integration*; Huazhong University of Science & Technology: Wuhan, China, 2015.
18. Tian, H.; Li, C.; Zhang, S.Y. The climate change in the Yangtze-Huaihe river valley over the past 50 years. *Period. Ocean. Univ. China* **2005**, *35*, 539–544.
19. Qi, H. Analysis of precipitation changes in Huaibei Plain over last 50 years. *Chin. J. Agrometeorol.* **2009**, *30*, 138–142.
20. Jin, J.L.; Song, Z.Z.; Jiang, S.M.; Zhou, Y.L.; Zhang, M. Spatiotemporal distribution of potential evapotranspiration in Huaibei Plain based on cloud model. *South–North Water Transf. Water Sci. Technol.* **2017**, *15*, 25232.
21. Li, D.Y.; Liu, C.Y. Study on the Universality of the Normal Cloud Model. *Engineering Science. Eng. Sci.* **2004**, *6*, 28–34.
22. Yang, J.; Wang, G.Y.; Pang, Z.L. Acceleration mechanism of cloud transformation based on density peaks. *J. Chin. Comput. Syst.* **2018**, *39*, 1299–1304.
23. Meng, H.; Wang, S.L.; Li, D.Y. Concept extraction and concept hierarchy construction based on cloud transformation. *J. Jilin Univ.* **2010**, *40*, 782–787.
24. Li, D.Y.; Du, Y. *Uncertainty Artificial Intelligence*; National Defense Industry Press: Beijing, China, 2014.
25. Du, Y.; Li, D.Y. Concept partition based on cloud and its application to mining association rules. *J. Softw.* **2001**, *12*, 196–203.
26. Ren, J. Linguistic-stochastic multi-criterion decision-making method based on cloud model. *Comput. Integr. Manuf. Syst.* **2012**, *18*, 2792–2797.
27. Hu, B.Q.; Wu, Y.G.; Cheng, T. Research on the Selection of Typical Hydrologic Year with the Cloud Model. *China Rural. Water Hydropower* **2016**, *5*, 97–100.
28. Keyantash, J.A.; Dracup, J.A. An aggregate drought index-assessing drought severity based on fluctuations in the hydrologic cycle and surface water storage. *Water Resour. Res.* **2004**, *40*, 1–13. [[CrossRef](#)]
29. Mu, D.R.; Luo, P.P.; Lyu, J.Q.; Zhou, M.M.; Huo, A.D.; Duan, W.L.; Nover, D.; He, B.; Zhao, X.L. Impact of temporal rainfall patterns on flash floods in Hue City, Vietnam. *J. Flood Risk Manag.* **2020**, *14*, e12668. [[CrossRef](#)]
30. Hao, L.; Zhang, X.Y.; Liu, S. Risk assessment to China's agricultural drought disaster in county unit. *Nat. Hazards* **2012**, *61*, 785–801. [[CrossRef](#)]
31. Zhang, Y.G.; Li, Y.L. Application of Cloud Model in Logging Curve Hierarchical. *Comput. Digit. Eng.* **2014**, *42*, 1613–1616.
32. Zhou, Z.T.; Guo, B.; Xing, W.X.; Zhou, J.; Xu, Y. Comprehensive evaluation of latest GPM era IMERG and GSMaP precipitation products over mainland china. *Atmos. Res.* **2020**, *246*, 1–15. [[CrossRef](#)]
33. Guo, B.; Bian, Y.; Zhang, M.; Su, Y.; Wang, X.X.; Zhang, B.; Wang, Y.; Chen, Q.J.; Wu, Y.R.; Luo, P.P. Estimating socio-economic parameters via machine learning methods using LuoJia1-01 nighttime light remotely sensed images at multiple scales of China in 2018. *IEEE Access* **2021**, *99*, 1–15.