Information Entropy of a Rainfall Network in China

Jiguo Zhang, Huimin Wang, and Vijay P. Singh

Abstract In this paper, we use the directional information transmission index which is based on information entropy to investigate the spatial nonuniformity of rainfall in the Huaihe River basin in China, and propose methods of information transfer function, information distance and information area which are employed to determine the area where a rain gauge is capable of representing. The analysis also involves transformation of the rainfall field from discrete state to continuous state.

Keywords Distribution nonuniformity · Information entropy · Information field · Information transmission · Rainfall network · Uncertainty

1 Introduction

Entropy, as a measure of information and uncertainty, has been employed for evaluation and design of rainfall network as well as for expansion of networks. Caselton and Husain (1980) used the maximum information transmission to select

J. Zhang (✉)
Business School, Hohai University at Changzhou, Changzhou 213022, People's Republic of China
and
State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, People's Republic of China
e-mail: zhangjg@hhuc.edu.cn

H. Wang
State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, People's Republic of China

V.P. Singh
Department of Biological and Agricultural Engineering, Texas A&M University, College Station, TX 77843, USA

Y. Zhou and D.D. Wu (eds.), *Modeling Risk Management for Resources and Environment in China*, Computational Risk Management,
DOI 10.1007/978-3-642-18387-4_2,

stations in a hydromeric network. Husain (1989) applied an entropy-based methodology to (a) selecting the optimum number of stations from a dense network, and (b) expanding a network using data from an existing sparse network by interpolation of information and identification of zones with minimum hydrological information. Harmancioglu and Yevjeich (1987) used entropy to measure the transmission of information among gauging stations on a river. Krstanovic and Singh (1992a, b) developed an entropy-based approach for space and time evaluation of rainfall networks in Louisian. Employing a measure of information flow, called directional information transfer index (DITI), between gauging stations in a network, Yang and Burn (1994) presented an entropy-based methodology for design of data collection networks. DITI is both a measure of the information transmission capacity and an indicator of the dependence of station pairs. The concept of DITI can be applied to regionalization of networks.

Mutual information and DITI can be applied to describe the degree of relationship between stations in a rainfall network. DITI between station pairs and area pairs can be used to show the spatial distribution of rainfall in a river basin from one direction to the other. Similarly, a graph of DITI and distance can show the difference in information transmission from one direction to the other, e.g. from east (E) to west (W) or from south (S) to north (N).

Huaihe River basin is located in east of China and between Yangtse River basin and Yellow River basin from which the 99 gauging stations, lying between N 31°–35° and E 112°–118°, upper Benbu Station (No. 86, E 117°23′, N 32°56′), as shown in Fig. 1, are selected by this paper. The objective of this paper is to investigate the spatial distribution nonuniform of rainfall in the area.

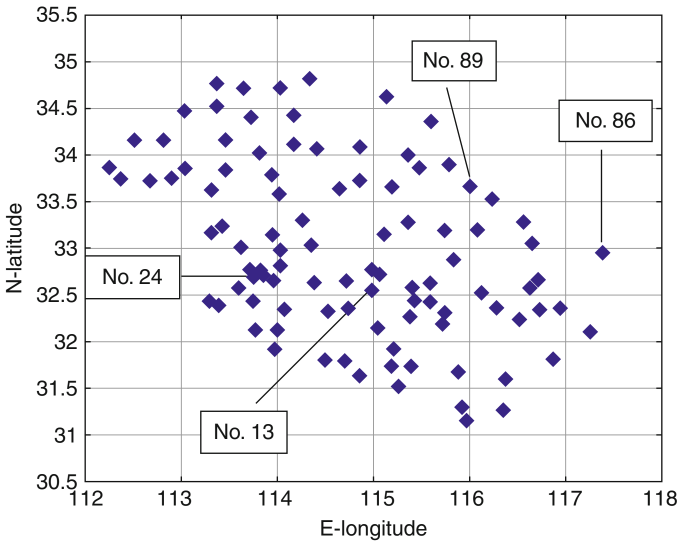

Fig. 1 Sketch map on 99 stations and 4 special stations

2 Preliminary Concepts

The Shannon information entropy is defined as:

$$H(X) = H(p_i, p_2 \cdots, p_n) = -\sum_{i=1}^{n} p_i \log p_i. \tag{1}$$

where X is a discrete random variable, p_i is the probability that X assumes a value $X = x_i$, and n is the number of values (sample size) that X can take on. Equation (1) defines entropy in real time or space.

To measure the transmission of information and indicate the dependence between two stations X and Y, a directional information transfer index (*DITI*) is defined as:

$$DITI_{XY} = I(X,Y) \Big/ H(Y) = \frac{H(Y) - H(Y/X)}{H(Y)}. \tag{2}$$

where $DITI_{XY}$ is *DITI* of X about station Y, $H(Y/X)$ is the entropy of Y conditioned on X, and $I(X,Y)$ is the mutual information defined as:

$$I(X,Y) = H(Y) - H(Y/X) = H(X) - H(X/Y) = I(Y,X). \tag{3}$$

Equation (3) is a measure of information transmission. In $DITI_{XY}$, X is called the basic point (basic station) and Y the auxillary point (auxillary station). $DITI_{XY}$ varies from 0 to 1; a large value indicates large capability of information transmission from X to Y. The concept of *DITI* can be extended to any number of stations and measures inherent dependence among stations of a set as:

$$SDITI = \sum_{i=1}^{m} \sum_{j=1, j \neq i}^{m} DITI_{ij}. \tag{4}$$

where m is the number of stations in the set.

3 Data Analysis

Using (1), which eliminates random uncertainty of the annual rainfall series and shows spatial characteristics of precipitation distribution, the entropy values of 99 stations were calculated. Using (3), the mutual information values between station pairs were obtained.

Generally, the majority of Huaihe River basin is influenced by plum rains from July to September because of the southeast monsoon. The plum rains starts in Yangtse River basin at first, and then Huaihe River basin, finally moves to Yellow

River basin. For the sake of observing this phenomenon and demonstrating that *DITI* is a correlative index between station pairs, the area mentioned was partitioned according to latitude as well as longitude.

It is obvious that *DITI* was associated with distance. Thus, when information transmission is combined, the distance must be accounted for at the same time. Two stations, No. 24 and No. 89 (see Fig. 1), are selected to show the information transfer graph with distance. There are two reasons for choosing these stations. One is that these two stations are much different in location, weather and topography, for station No. 24 is located in the mountainous area and station No. 89 in the northeast plain area. The other is that they are located in the border of the region. Therefore, if a direction is extended from the basic station along E-longitude or N-latitude, then more auxiliary stations can be located in order to analyze the relationship between *DITI* and distance.

Now let stations No. 24 and No. 89 (see Fig. 1) be basic stations, some auxiliary station points are chosen separately.

①Station No. 24 as a basic station, 7 auxiliary stations to be chosen from S to N
②Station No. 24 as a basic station, 8 auxiliary stations to be chosen from W to E
③Station No. 89 as a basic station, 7 auxiliary stations to be chosen from E to W
④Station No. 89 as a basic station, 7 auxiliary stations to be chosen from N to S

The relation between DITIs and distances is illustrated separately in Figs. 2 and 3.

These figures display the following properties: (1) *DITI* decreases with increasing distance; (2) *DITI* decreases faster in the S-N direction than in the E-W direction. In other words, the information transmission is greater in the E-W direction than in the S-N direction; and (3) the longer the distance is, the smaller the *DITI* is. This shows that the rainfall field of the Huaihe River basin spreads out

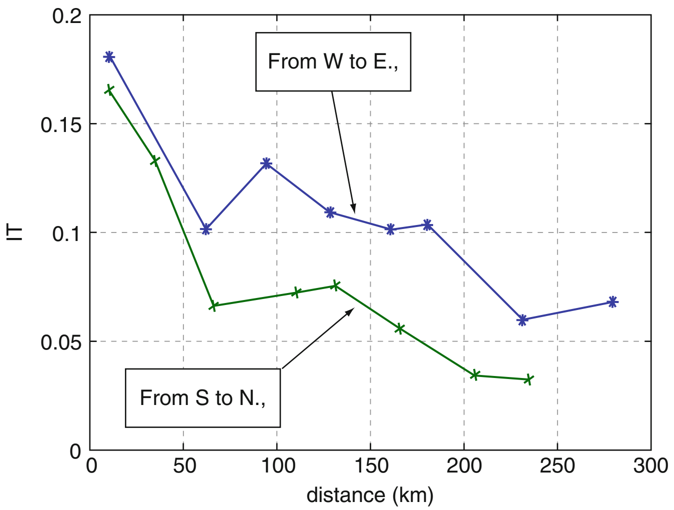

Fig. 2 Information transfer graph of basic station No. 24

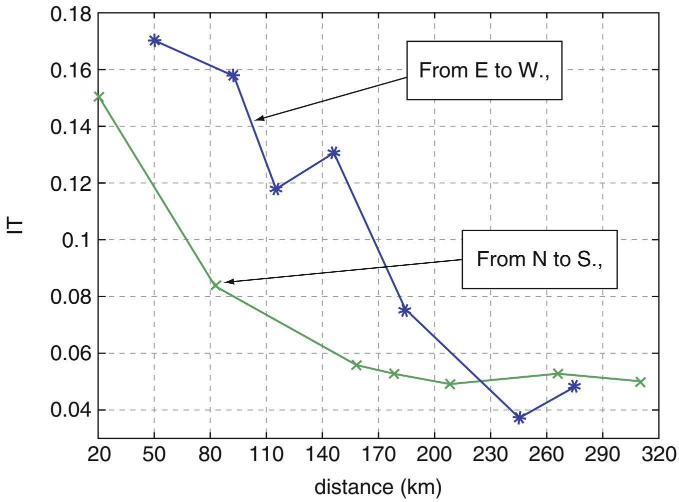

Fig. 3 Information transfer graph of basic station No. 89

in the E-W direction and moves from south to north. Thus, *DITI* reflects the relationship of a gauging station to its nearby region, and its value reduces with increasing distance. Based on a large number of stations studied, it was observed that this characteristic not only exists for each station in the Huaihe River basin, but also the functional curves have the same properties. This phenomenon will be further researched in the following section.

4 Theoretical Models

The use of entropy requires defining the spatial structure of rainfall by a multivariate probability distribution. In general, a continuous distribution function is assumed (Husain 1989) or a non-parametric distribution is applied (Yang and Burn 1994). This study extends the discrete field to the continuous field. In this case, each point in the area is a gauging station. In the continuous field, an information transfer function model is derived, and two information indexes, information distance and information field which measure the capability of a station with respect to its nearby regions, are defined. Then, an algorithm in the frequency domain is employed following the principle of maximum entropy.

4.1 Information Transfer Function

The rainfall field is characterized by a finite number of stations and is called a discrete field. It may, however, be more convenient to extend the discrete field to the continuous field for developing a theoretical model for information transmission. In this case, every point in an area on a spatial coordinate surface is considered a rainfall gauging station. Therefore, the transmission of information of a given station successively changes and be in progress in any of the directions.

Let X be a basic point in the rainfall field and Y an auxillary station being in a certain direction. Let s_{XY} be the geometric distance between X and Y. Then the information transfer function (*ITF*) of X about Y can be defined as:

$$ITF_X = \frac{1}{(1+as)^b}. \tag{5}$$

where $a, b > 0$ are parameters.

When a, b are smaller in (5), ITF_X is greater, suggesting greater information from X in a given direction. Although different river basins have probably different information transmission function, two general properties can be easily summarized:

1. $ITF_X(0) = 1 \qquad ITF_X \to 0 (s \to \infty)$
2. ITF_X is a differential function, and $ITI = \dfrac{d(ITF_X)}{ds} \le 0$

where *ITI* is the abbreviation for information transfer intensity.

Thus, using a nonlinear regression analysis, the simulated curves representing 4 broken lines in Figs. 2 and 3 can be expressed analytically as:

$$\text{Station No.24} \quad \text{from W to E } ITF = 1/(1+2.784\text{s})^{0.471} \quad \text{R} = 0.80. \tag{6}$$

$$\text{from S to N } ITF = 1/(1+0.126\text{s})^{1.289} \quad \text{R} = 0.94. \tag{7}$$

$$\text{Station No.89} \quad \text{from E to W } ITF = 1/(1+0.092\text{s})^{1.074} \quad \text{R} = 0.94. \tag{8}$$

$$\text{from N to S } ITF = 1/(1+1.181\text{s})^{0.633} \quad \text{R} = 0.98. \tag{9}$$

Equations (6), (7), (8) and (9) are called information transfer functions which are represented with curves as in Figs. 4 and 5 (*IT* is *ITF* in following figures).

4.2 Information Distance

Let X be a station and N the length from X to the border of a basin along a given direction, then the information distance (ID_X) of X in this direction can be defined as:

$$ID_X = \left(2\int_0^N sITF_X ds\right)^{\frac{1}{2}} = \frac{1}{a^2}\left[\frac{(1+aN)^{2-b}-1}{2-b} - \frac{(1+aN)^{1-b}-1}{1-b}\right]. \tag{10}$$

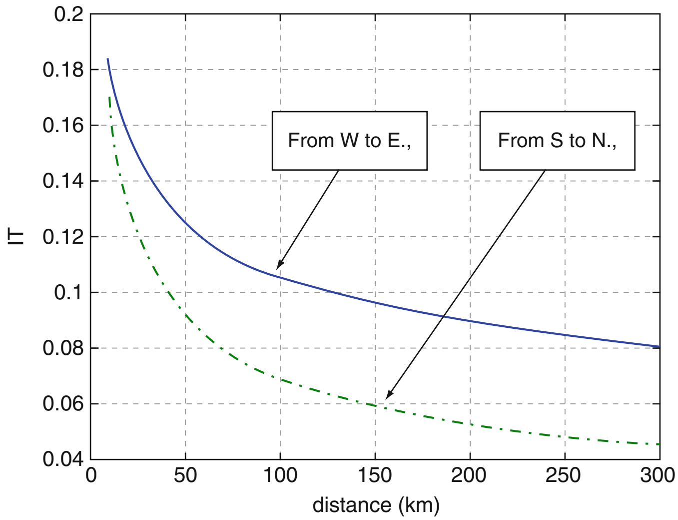

Fig. 4 Sketch map of ITF on No. 24 in W-E and S-N

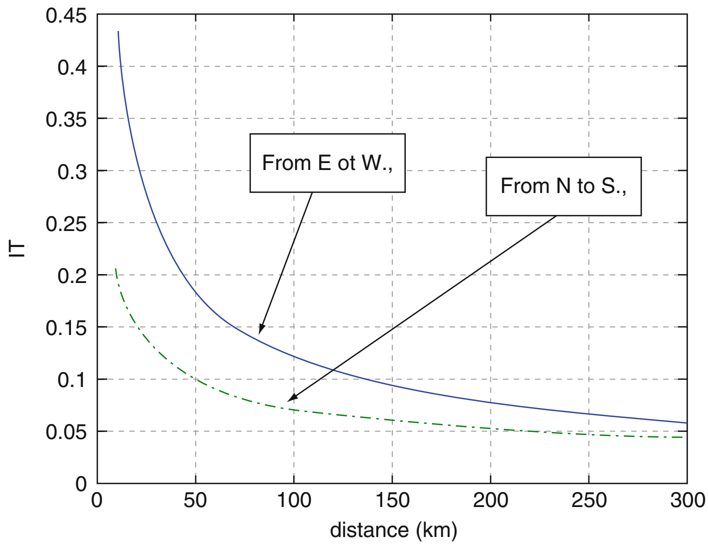

Fig. 5 Sketch map of ITF on No. 89 in E-W and N-S

In light of the principle of fuzzy mathematics, ITF_X may be considered a possibility distribution, that means the membership function of station Y belongs to the basic station X. Therefore, ID_X measures the average information transfer of X in the above-mentioned direction in a statistical sense. For example, if $ITF_X = 1$

Table 1 ID's of stations No. 24 and 89 with distances (km)

Dis.	W→E (No. 24)	E→W (No. 89)	S→N (No. 24)	N→S (No. 89)
50	8.47	16.84	7.47	8.83
100	14.16	24.23	11.38	13.13
150	19.16	29.60	14.57	16.52
200	23.77	33.96	17.35	19.44
250	28.11	37.68	19.87	22.04
300	32.24	40.97	22.20	24.41

then $ID_X = \mathrm{N}$, which means the information may be entirely transferred to the basin boundary; on the contrary, if, except $\mathrm{N} = 0$, $ITF_X = 0$, then $ID_X = 0$, in this case, station X does not transfer any information outward. Thus, ID_X defined is reasonable.

Consider the information distance of stations No. 3 and 89. The geometric distances of station 3 to the border from west to east and from south to north separately are 325 and 270 km. And about station 89, the distances separately are 305 km (from E to W) and 335 km (from N to S). Then, the *ID*'s of stations No. 24 and 89 with distances separately were graphed in Table 1.

From Table 1, *ID* in S-N is even larger than that in E-W within about 65 km. Thus, on average information transfer capacity, there is no obvious difference between E-W and S-N no more than 70 km. But it is also seen that increasing velocity of *ID* in S-N is more slowly than that in E-W. It is shown that time and intensity of precipitation are very identical in the zone of N 32°–33°, but not in the zone of E 113°–114°.

4.3 Information Area

Both *DITI* and *ID* are associated with some direction. Geometrically, infinite directions can be drawn from a station, and there can be positive connections between stations in any of these directions. Therefore, there is an *ITF* in each direction for a ststion, so an information field is defined. For a station X in the rainfall field, there exists information transmission in all directions. Then, we call the area affected by station X as the information field (abbreviated as *IFIELD*) of X. It is equivalent to a set consisting of the whole ITF_X of X. To mathematically formulate *IFIELD*, an appropriate coordinate system needs to be established. Let the geometric location of X be a pole and one direction (e.g., selecting the direction from east to west) be a polar axis; thus, polar coordinates are obtained. Let θ be the polar angle. Then, one can consider parameters a and b in ITF relating to θ, that is, $a = a(\theta)$, $b = b(\theta)$, and $ITF = ITF(\theta)$. Therefore, information field of X is defined as:

$$IFIELD_X = \{ITF_X(\theta),\ 0 \leq \theta \leq 2\pi\}. \tag{11}$$

Suppose that *ITF* is a continuous function of s and θ. Then, we designate:

$$IA_X = \frac{1}{2}\int_0^{2\pi} (ID_x)^2 d\theta. \tag{12}$$

as the information area (*IA*) of *X*, which measures the total region of influence of *X* with respect to its nearby region.

In practice, all *ITFs* of *X* can not be obtained due to the limitation on the number of stations. One can obtain only finite *ITFs*. For illustration, station No. 13 (E 114°59′ N 32°33′) (see Fig. 1) is considered, because it is located almost in center of the area. Beginning with $\theta = 0$, and at a set interval of $\pi / 4$, 8 *ITFs* are computed and given as:

$$\mathrm{ITF} = 1/(1 + 0.578\mathrm{s})^{0.542} \quad \mathrm{ITF} = 1/(1 + 1016.441\mathrm{s})^{0.212}$$

$$\mathrm{ITF} = 1/(1 + 2.501\mathrm{s})^{0.500} \quad \mathrm{ITF} = 1/(1 + 0.874\mathrm{s})^{0.593}$$

$$\mathrm{ITF} = 1/(1 + 6.454\mathrm{s})^{0.407} \quad \mathrm{ITF} = 1/(1 + 0.914)^{0.532}$$

$$\mathrm{ITF} = 1/(1 + 0.052)^{1.602} \quad \mathrm{ITF} = 1/(1 + 0.477\mathrm{s})^{0.689}$$

There are 8 distance values from station No. 13 (see Fig. 1) to border of the area along 8 different directions. The minimum of them is 145 (km) which is considered as a integral upper limit while computing *ID*. It is seen that *IA* is 5761 (km^2) if distance is 90 (km), and *IA* is 13256 (km^2) if distance is 145 (km).

5 Conclusion

The following conclusions are drawn from this study: (1) The Information transfer graph shows a decline in Information of a station in the east-west as well as in the south-north direction; (2) The cluster graphs show spatial nonuniformity of the rainfall field; (3) In section of Data Analysis, the values of SDITI of station-sets separately from N 32°–33° and E 115°–116° are relatively larger, the reason is, the most of stations chosen come from the region where the distribution of precipitation is uniformity, thus, relative *DITI* values are larger than others; (4) *ID* is an index used to measure average information transmission capacity of a station in the given direction; and *IA* is an influence indicator of a station to its nearby region in information.

Acknowledgments This work is supported by the Public-interest Industry Project of Ministry of Water Resources of China (No. 200801027).

References

Caselton WF, Husain T (1980) Hydrologic network information transmission. J Water Resour Plan Manag 106:503–520

Harmancioglu N, Yevjevich V (1987) Transfer hydrologic information among river points. J Hydrol 191:103–118

Husain T (1989) Hydrologic uncertainty measure and network design. Water Resour Bull 25(3):527–534

Krstanovic PF, Singh VP (1992a) Evaluation of rainfall network using entropy: I. theoretical development. Water Resour Manag 6:279–293

Krstanovic PF, Singh VP (1992b) Evaluation of rainfall network using entropy: II. application. Water Resour Manag 6:295–314

Yang Y, Burn DH (1994) An entropy approach to data collection network design. J Hydrol 157:307–324

http://www.springer.com/978-3-642-18386-7

Modeling Risk Management for Resources and
Environment in China
Wu, D.D.; Zhou, Y. (Eds.)
2011, XIV, 578 p., Hardcover
ISBN: 978-3-642-18386-7