Predicting extreme rainfall over eastern Asia by using complex networks*

He Su-Hong(何苏红)[a], Feng Tai-Chen(封泰晨)[b], Gong Yan-Chun(龚艳春)[a], Huang Yan-Hua(黄雁华)[a], Wu Cheng-Guo(吴成国)[a], and Gong Zhi-Qiang(龚志强)[c,d]†

[a] *Institute of Science, PLA University of Science, Nanjing* 211101, *China*
[b] *Atmosphere and Science College, Lanzhou University, Lanzhou* 730000, *China*
[c] *Laboratory for Climate Monitoring and Diagnosing, National Climate Center, China Meteorological Administration, Beijing* 100081, *China*
[d] *Laboratory for Climate Diagnosis of National Climate Center, China Meteorological Administration, Beijing* 100081, *China*

(Received 16 August 2013; revised manuscript received 26 November 2013; published online 25 March 2014)

A climate network of extreme rainfall over eastern Asia is constructed for the period of 1971–2000, employing the tools of complex networks and a measure of nonlinear correlation called event synchronization (ES). Using this network, we predict the extreme rainfall for several cases without delay and with n-day delay ($1 \leq n \leq 10$). The prediction accuracy can reach 58% without delay, 21% with 1-day delay, and 12% with n-day delay ($2 \leq n \leq 10$). The results reveal that the prediction accuracy is low in years of a weak east Asia summer monsoon (EASM) or 1 year later and high in years of a strong EASM or 1 year later. Furthermore, the prediction accuracy is higher due to the many more links that represent correlations between different grid points and a higher extreme rainfall rate during strong EASM years.

Keywords: prediction, extreme rainfall, synchronization, complex networks

PACS: 92.60.Wc **DOI:** 10.1088/1674-1056/23/5/059202

1. Introduction

Extreme rainfall events have a significant impact on the environment and human activities. In recent years, extreme events over eastern Asia have happened frequently, resulting in severe flood throughout the region. As weather and climate extremes can have devastating effects on the society and the environment,[1] it is of immense value to the society and the economy to deeply and clearly understand the dynamical structure of extremes and to forecast them well in advance.

Extreme rainfall over eastern Asia results from the interaction of several complex atmospheric processes evolving at many different spatial and temporal scales, e.g., the low latitude east Asia summer monsoon (EASM), the Indian summer monsoon, and the mid-latitude blocking system.[2–5] Hence, extreme rainfall shows complex spatial and temporal patterns. Most recent studies on extreme rainfall events over eastern Asia mainly focused on the extreme value, its corresponding temporal change characteristics, and the dynamical analysis of single events. Few studies have paid attention to the prediction of extreme rainfall over eastern Asia. At present, the prediction accuracy of extreme rainfall is still low, thus it is vital to develop a climate model to point out where skillful forecasts can be made due to the changes in atmospheric composition.[6–8]

During the last decade, the study of networks has become a major research topic in many scientific areas, e.g., neural networks in biology, the World Wide Web in computer science, and citation networks in the social sciences.[9–11] Recently, the network perspective was also applied to climate research, revealing spatio–temporal structures of climatic variation and the mechanisms of climate dynamics at both global and local scales.[12–18,20–30] In most of those works, uni-directional networks were constructed by employing the Pearson correlation. In this study, we construct a directed network using a new method called event synchronization.[3,20,31] This directed network can measure not only the strength of synchronization between two grid points but also their delay behavior. Extreme rainfall has temporal and spatial memories; the state of this moment will influence the state of the next moment; and a rain event occurring at one grid point will make any other grid points vulnerable to such an event too. A similar approach was recently chosen by Malik *et al.*, who researched the temporal evolution of extreme rainfall patterns and spatial intricacies over south Asia.[3] This paper predicts extreme rainfall over eastern Asia for several cases without delay and with n-day delay. Furthermore, we present the relationship between the EASM and the prediction accuracy.

2. Data

The data used here are daily gridded rainfall data from 1951 to 2007 developed as a part of the project Asian rainfall highly resolved observational data integration towards the

*Project supported by the National Natural Science Foundation of China (Grant No. 41205040), the National Basic Research Program of China (Grant Nos. 2012CB955203 and 2012CB955902), and the Special Scientific Research Project for Public Interest, China (Grant No. GYHY201306021).

†Corresponding author. E-mail: gzq0929@126.com

© 2014 Chinese Physical Society and IOP Publishing Ltd

evaluation of water resources (APHRODITE).[26] It is available from the website at http://www.chikyu.ac.jp/precip. We have extracted the data from 1971 to 2000 for the eastern Asian region with a latitude–longitude resolution of 1°×1°. The range of latitude is between 10° N and 55° N, and that of longitude is between 100° E and 135° E, as a result, the total number of grid points is 1575.

3. Methodology

3.1. Event synchronization

Event synchronization (ES) can be employed as a nonlinear correlation to measure the synchronization strength and the delay direction of rainfall events between two different grid points. ES also serves as the basis for constructing complex networks. ES has been previously introduced by Quiroga *et al.*[30] and modified by Malik N *et al.*[3,20]

At each grid point during the three summer months (JJA), we consider the events above the 90th percentile of the time series (30 years of data × 92 days per season × 10% = 276 events at each grid point). The 90th percentile threshold gives a unique net amount of rainfall per day as the threshold for each grid point, which is shown in Fig. 1. It is clear that the coastal regions have larger amounts of rainfall than the inland regions.

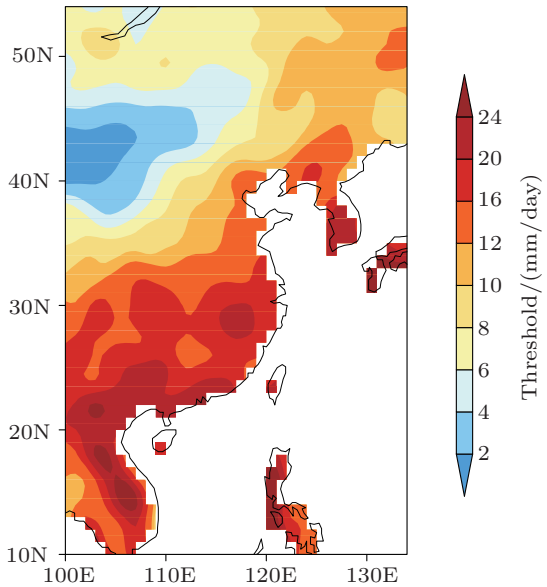


Fig. 1. (color online) Daily rainfall thresholds calculated from 30 years of data (1971-2000).

Let us suppose that events above the threshold occur at time t_l^i at grid point i and at time t_m^j at grid point j ($l = 1,2,\ldots,s_i$, $m = 1,2,\ldots,s_j$), where s_i and s_j are the total numbers of such events occurring at grid points i and j, respectively. According to Malik *et al.*,[3] we allow a time lag $\pm\tau_{lm}^{ij}$ between two "synchronous" events, which is defined as

$$\tau_{lm}^{ij} = \min\left\{t_{l+1}^i - t_l^i, t_l^i - t_{l-1}^i, t_{m+1}^j - t_m^j, t_m^j - t_{m-1}^j\right\}/2. \quad (1)$$

We count the number of events at i after an event appears at j, denoted as

$$c(i|j) = \sum_{l=1}^{s_i}\sum_{m=1}^{s_j} J_{lm}, \quad (2)$$

with

$$J_{lm} = \begin{cases} 1, & \text{if } 0 < t_l^i - t_m^j < \tau_{lm}^{ij}, \\ 1/2, & \text{if } t_l^i = t_m^j, \\ 0, & \text{else.} \end{cases} \quad (3)$$

Similarly, we can obtain $c(j|i)$. Then we define the symmetrical quantity $\boldsymbol{Q}$ and the asymmetrical quantity $\boldsymbol{q}$

$$Q_{ij} = \frac{c(i|j) + c(j|i)}{\sqrt{s_i s_j}}, \quad (4)$$

$$q_{ij} = \frac{c(i|j) - c(j|i)}{\sqrt{s_i s_j}}. \quad (5)$$

The Q_{ij} measures the strength of synchronization between grid points i and j, and q_{ij} measures their delay behavior.[31] They are normalized to $0 \le Q_{ij} \le 1$ and $-1 \le q_{ij} \le 1$, where $Q_{ij} = 1$ implies the complete synchronization between grid points i and j, and $q_{ij} = 1$ implies that an event at grid point i always precedes an event at grid point j.[31]

3.2. Constructing adjacency matrix

We set a threshold to get rid of redundant information and extract the most significant correlations. If the strength is above this threshold, significant links between different grid points exist. Only the most significant correlations are analyzed: we suppose that only 5% of the total grid points are connected and calculate a threshold θ on Q_{ij} in this paper.[3] Using θ, we convert $\boldsymbol{Q}$ into a binary matrix called the adjacency matrix $\boldsymbol{A}^Q$

$$A_{ij}^Q = \begin{cases} 1, & \text{if } Q_{ij} > \theta_{ij}, \\ 0, & \text{else.} \end{cases} \quad (6)$$

Similarly, we can convert q_{ij} into a binary matrix called the adjacency matrix $\boldsymbol{A}^q$

$$A_{ij}^q = \begin{cases} 1, & \text{if } q_{ij} > \left|\theta_{ij}^q\right|, \\ 0, & \text{else.} \end{cases} \quad (7)$$

Here $\boldsymbol{A}^Q$ is a symmetric matrix, while $\boldsymbol{A}^q$ is an asymmetric matrix. A further adjacency matrix $\boldsymbol{A}^{Qq}$ can be constructed as

$$A_{ij}^{Qq} = A_{ij}^q A_{ij}^Q. \quad (8)$$

The $\boldsymbol{A}^{Qq}$ is also an asymmetric matrix presenting the information of strength and direction of the link.[3]

4. Results and discussion

4.1. Predicting extreme rainfall

Let us assume that an event above the threshold occurs at grid point i, and this event will make any other grid point j vulnerable to such an event as well. The vulnerability must depend on the distance from i to j and the link of these two grid points. Apart from the link strength, the link directionality is also important, only those grid points that have incoming links from i being vulnerable to such an event. From the adjacency matrix $\boldsymbol{A}^{Qq}$, we can obtain information about strength and direction of the link.[3] The vulnerability of grid point j can be calculated to be

$$\rho_j = \frac{\sum_{i\in V(t)} A_{ij}^{Qq}/L_{ij}}{\sum_{i=1}^{N} A_{ij}^{Qq}/L_{ij}}, \tag{9}$$

where L_{ij} is the distance from i to j, which can be calculated using the formula for the spherical Earth projected onto a plane, i.e.,

$$L_{ij} = R\sqrt{(\delta\phi_{ij})^2 + (\cos(\phi_{\rm m})\delta\varphi_{ij})^2}, \tag{10}$$

where $\delta\phi_{ij}$ and $\delta\varphi_{ij}$ are differences in latitude and longitude in radians between grid points i and j, $\phi_{\rm m}$ is the mean of the latitudes of i and j, and R is the radius of the Earth.[3] In formula (9), $V(t)$ is the set of grid points where such rain events happen at time t, and N is the number of all grid points.

We separate the data into two parts: the first 15 years and the second 15 years. We determine a threshold ρ_0 on ρ_j from the first half of the data and use the threshold to predict rain events for the second half. We set a pre-assigned value for ρ_j to calculate the prediction accuracy and ensure the threshold ρ_0. If ρ_j is greater than the threshold ρ_0, an event occurs at grid j, or else no event occurs at grid j. We define two ratios to judge the quality of the prediction. First, we have $\varepsilon_{\rm p}/E_{\rm p}$, which is the ratio of the number of correctly predicted events ($\varepsilon_{\rm p}$) to the total number of predicted events ($E_{\rm p}$). Second, we have $E_{\rm p}/E_{\rm T}$, which is the ratio of the total number of predicted events ($E_{\rm p}$) to the total number of events ($E_{\rm T}$). The value of ρ_0 has been chosen such that $\varepsilon_{\rm p}/E_{\rm p} \approx E_{\rm p}/E_{\rm T}$ with the best accuracy in the first half of the data. We use the symbol ε for the accuracy of prediction and it is given by $\varepsilon = \varepsilon_{\rm p}/E_{\rm p}$, where $\varepsilon_{\rm p}/E_{\rm p} \approx E_{\rm p}/E_{\rm T}$.

We analyze several cases without delay and with n-day delay ($1 \le n \le 10$). The delay of n days is introduced by changing the set $V(t)$ to $V(t-n)$ in Eq. (9). Figure 2 shows the relation between ρ_0 and the number of delay days. With the increase of the number of delay days, ρ_0 first rockets and then remains fundamentally unchanged after the number of delay days reaches 2. Figure 3 indicates that the prediction accuracy ε first plummets with the increase of the number of delay days and then also remains fundamentally unchanged after the number of delay days reaches 2. As shown in Figs. 2 and 3, $\rho_0 \approx 0.42$, $\varepsilon \approx 0.59$, i.e., 59% of the events are correctly predicted when we use no time delay, $\rho_0 \approx 0.59$, $\varepsilon \approx 0.21$, i.e., 21% of the events are correctly predicted when we use a 1-day delay, and $\rho_0 \approx 0.68$, $\varepsilon \approx 0.12$, i.e., only 12% of the events are correctly predicted when we use n-day delay ($2 \le n \le 10$).

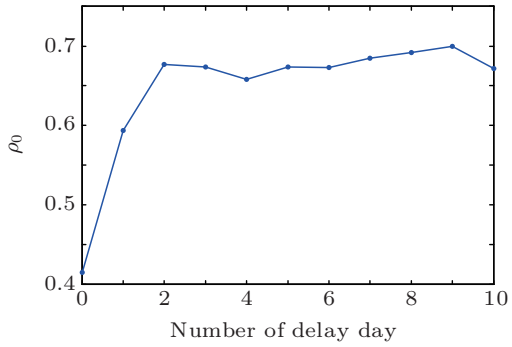


Fig. 2. (color online) The ρ_0 as a function of the number of delay days.

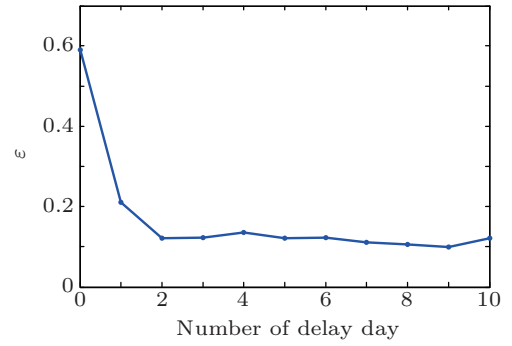


Fig. 3. (color online) The ε as a function of the number of delay days.

Figure 4 presents the spatial distribution of the prediction accuracy ($\varepsilon = \varepsilon_{\rm p}/E_{\rm p}$) for the last half of the data. The spatial distribution of the prediction accuracy without delay is shown in Fig. 4(a). We observe that for most of the subcontinents, the accuracy is above 60% and even reaches up to 90% in some places. In Fig. 4(b), we find that for some subcontinents, the accuracy with 1-day delay is above 30%. The spatial distribution of the prediction accuracy with 2-day delay is shown in Fig. 4(c). In this case, the accuracy for most of the subcontinents is under 20%. The calculation indicates that in other cases ($2 < n \le 10$), the accuracy for most of the subcontinents is also under 20%. The prediction without delay can be considered as a test when $t \to 0$, which has no practical use for the prediction of extreme rainfall events. We are interested in the situations with n-day delay ($n \ge 1$), and we find that over some places in the north and some places in the south, the prediction accuracy is high, as shown in Figs. 4(b) and 4(c).

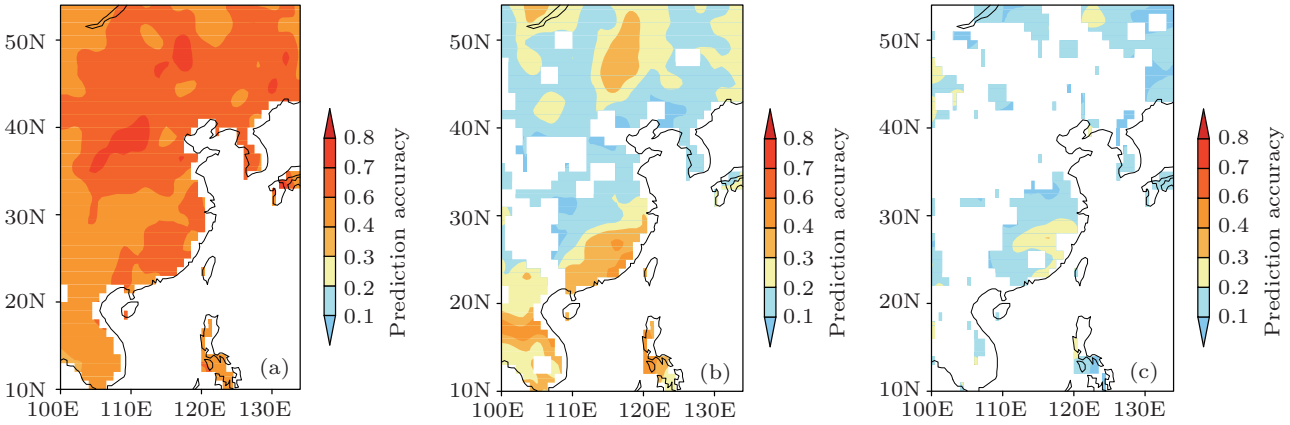


Fig. 4. (color online) Spatial distributions of the prediction accuracy ε for the last half of the data: (a) without delay, (b) with 1-day delay, and (c) with 2-day delay.

4.2. Possible influence index for the prediction accuracy

With the northward expansion of the EASM, the monsoon brings much rainfall, especially to east China, Japan, and the Korean peninsula. The EASM has a large impact on the summer rainfall over eastern Asia, and the location of the monsoon controls the location of extreme rainfall.[2,32–35] It has been shown very recently that a dynamical coupling exists between the EASM and the rainfall. This coupling has been of wide interest among climate researchers,[36] and we intend to identify the influence of the EASM on the rainfall field.

Figure 5 presents the temporal evolution of ε per year for the entire region, Fig. 5(a) is with 1-day delay, and Fig. 5(b) is with 2-day delay. The cross correlation coefficient of them is 0.6, which is rather larger than the value for statistical confidence level 99.9%. To identify the influence of the EASM on the rainfall field, we make a scatter plot between the prediction accuracy ε and the monsoon index, as shown in Fig. 6. The monsoon information is available from the website at http://ljp.lasg.ac.cn/dct/page/1, and the calculation method of the monsoon index was proposed by Shi *et al.*[37] The 10 years with the highest monsoon index are called strong EASM years, the 10 years with the lowest monsoon index are called weak EASM years, and the other 10 years are called ordinary EASM years. Similarly, according to the prediction accuracy, we separate 30 years into 3 parts: 10 years with high prediction accuracy, 10 years with low prediction accuracy, and 10 years with ordinary prediction accuracy. In order to achieve the desired results, in most years we select the monsoon index and the prediction accuracy in the same year. However, in some other years, we select the prediction accuracy with 1-year delay. From Fig. 6, we can observe in weak EASM years or 1 year later, the accuracy is low, and in the strong EASM years or 1 year later, the accuracy is high. This trend is indicated by the solid straight line, the slope of which is 0.09 in Fig. 6(a), and 0.04 in Fig. 6(b). And this relationship is especially obvious with 1-day delay. However we also observe several exceptions to this rule: the accuracy is high in some weak EASM years, and low in some strong EASM years. It is possible that extreme rainfall is affected not only by the EASM, but also by other complicated factors such as ENSO, typhoon, and so on.

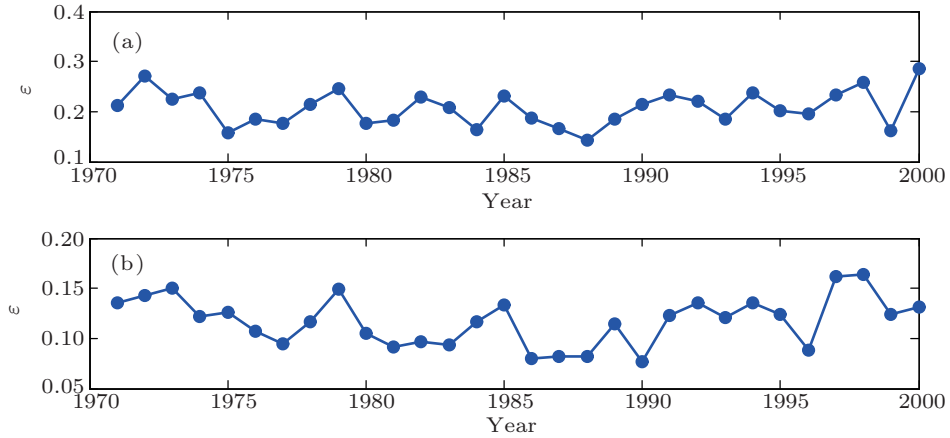


Fig. 5. (color online) Temporal evolution of the prediction accuracy ε per year for the entire region: (a) with 1-day delay, and (b) with 2-day delay.

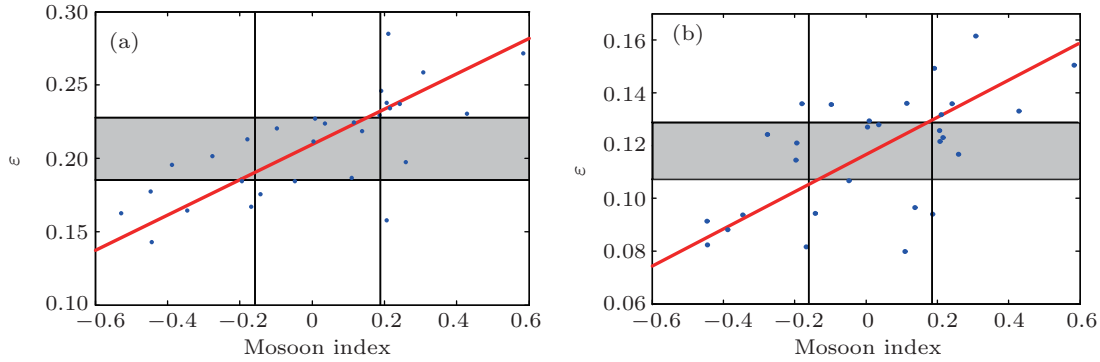


Fig. 6. (color online) Scatter plot between the prediction accuracy ε and the monsoon index (normalized): (a) with 1-day delay, and (b) with 2-day delay.

4.3. Possible influence mechanism

Given the above, we can conclude that the accuracy is low during weak EASM years and high during strong EASM years. In the next step, we attempt to present the possible influence mechanism. Figure 7 shows the total number of links at each grid point as delineated by the fraction of the total global area to which that point is connected.[14] Figure 7(a) shows the result of strong EASM years, Fig. 7(b) shows the result of ordinary EASM years, Fig. 7(c) shows the result of weak EASM years, and Fig. 7(d) shows the fractional difference between strong EASM years and weak EASM years. The conclusion from Fig. 7 is that with the weakening of the monsoon, the network loses many links.

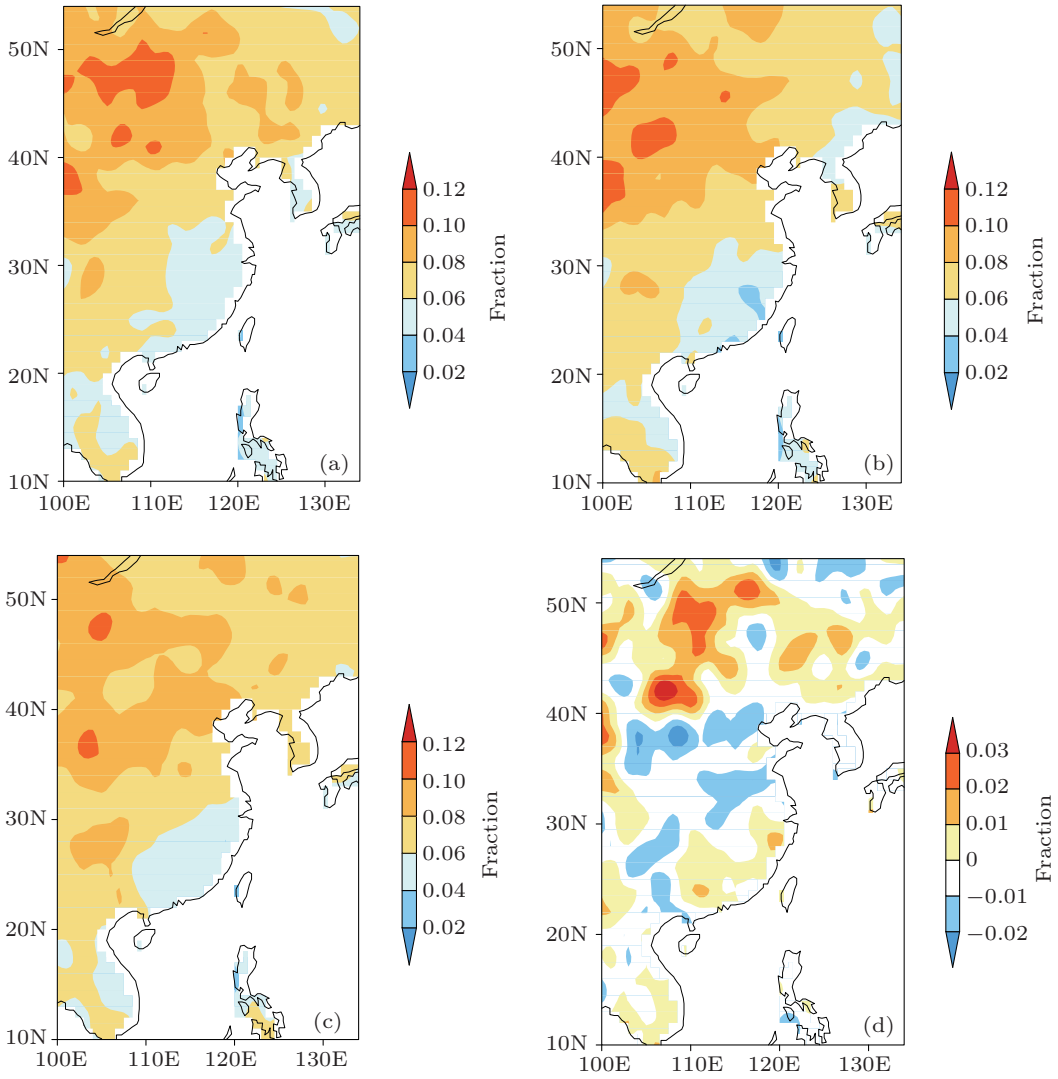


Fig. 7. (color online) Total number of links at each grid point as delineated by the fraction of the total global area to which that point is connected: (a) strong EASM years, (b) ordinary EASM years, (c) weak EASM years, and (d) fractional difference between strong EASM years and weak EASM years.

Figure 8 shows a scatter plot between the extreme rainfall rate and the monsoon index (normalized). The extreme rainfall rate is defined as

$$\alpha=\frac{1}{N}\sum_{i}^{N}\omega_i, \tag{11}$$

where N is the total number of grid points, and ω_i is the average number of extreme rainfall events occurring at grid point i each day. Similarly, in most years, we select the monsoon index and the extreme rainfall rate in the same year, but in some other years, we select the extreme rainfall rate with 1-year delay. The red solid straight line is a fitted curve, the slope of which is 0.006. As Fig. 8 indicates, during strong EASM years, the rate of extreme rainfall is greater than that during weak EASM years.

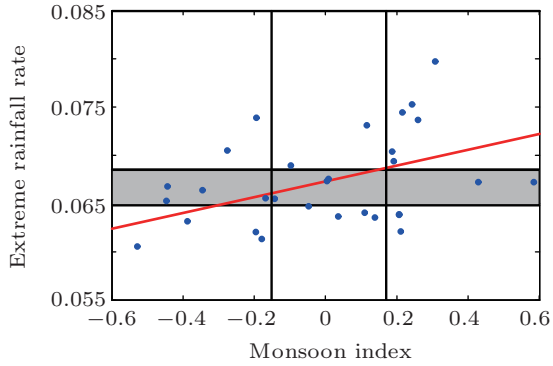


Fig. 8. (color online) Scatter plot of the extreme rainfall rate versus the monsoon index (normalized).

We speculate that, on average, the EASM acts as an extra affecting factor in the system, improving the synchronization of the network.[38] Therefore during strong EASM years, the network has many more links. If the network has many more links and a higher extreme rainfall rate, the prediction accuracy is higher because of more valuable information (see Eq. (9)). It can be seen that over some places in the north and some places in the south, the prediction accuracy is high with 1-day delay and with 2-day delay as shown in Figs. 4(b) and 4(c), respectively. From the above discussion, it can be concluded that the reasons for high prediction accuracy in these two regions are different. The high prediction accuracy over some places in the north results from the greater number of network links (see Fig. 7). However, the high prediction accuracy over some places in the south results from the higher extreme rainfall rate.

5. Conclusion

By constructing a climate network employing nonlinear correlation, we predict the extreme rainfall over eastern Asia. It is well known that the EASM and the extreme rainfall are intricately related, and we further find that this relationship has an effect on the prediction of the extreme rainfall field, a fact not previously known. Our results indicate that the accuracy has been low in weak EASM years or 1 year later and high in strong EASM years or 1 year later. Many more links and a higher extreme rainfall rate during the strong EASM years produce the higher prediction accuracy. Recognizing and understanding these effects may result in improvements in extreme rainfall prediction.

References

[1] Alexander L V and Zhang X 2006 *J. Geophys. Res.* **111** D05109
[2] Ding Y H and Johnny C L C 2005 *Meteorol. Atmos. Phys.* **89** 117
[3] Malik N, Bookhagen B, Marwan N and Kurths J 2012 *Clim. Dyn.* **39** 971
[4] Li J P and Chou J F 1997 *Acta Meteor. Sin.* **11** 57
[5] Li J P, Zeng Q C and Chou J F 2000 *Sci. China. Ser. E* **30** 550 (in Chinese)
[6] Shi Y N 1984 *Journal of Nanjing University* **20** 796 (in Chinese)
[7] Yang P C and Zhou X J 2005 *Acta. Meteor. Sin.* **63** 556 (in Chinese)
[8] Li J P and Chou J F 1997 *Science in China B* **27** 89 (in Chinese)
[9] Watts D J and Strogatz S H 1998 *Nature* **393** 440
[10] Albert R and Barabasi A L 2002 *Rev. Mod. Phys.* **74** 47
[11] Newman M E J 2003 *SIAM Rev.* **45** 167
[12] Tsonis A A and Roebber P J 2004 *Phys. A* **333** 497
[13] Tsonis A A, Swanson K L and Roebber P J 2006 *Bull. Am. Meteorol. Soc.* **87** 585
[14] Tsonis A A and Swanson K L 2008 *Phys. Rev. Lett.* **100** 228502
[15] Feng G L, Dong W J, Gong Z Q, Hou W, Wan S Q and Zhi R 2006 *Nonlinear Theories and Methods on Spatial-temporal Distribution of the Observational Data* (Beijing: Metrological press) p. 86 (in Chinese)
[16] Feng G L, Gao X Q, Dong W J and Li J P 2008 *Chaos Soliton. Fract.* **37** 487
[17] Feng G L, Gong Z Q, Zhi R and Zhang D Q 2008 *Chin. Phys. B* **17** 2745
[18] Feng G L, Dong W J and Jia X J 2003 *Chin. Phys.* **13** 413
[19] Swanson K L and Tsonis A A 2009 *Geophysical Research Letters* **36** L06711
[20] Malik N, Marwan N and Kurths J 2010 *Nonlin. Processes Geophys.* **17** 371
[21] Wang X J, Gong Z Q, Zhou L and Zhi R 2009 *Acta Phys. Sin.* **58** 6651 (in Chinese)
[22] Wang X J, Feng A X, Zhi R, He W P and Gong Z Q 2012 *Chin. Phys. B* **21** 029201
[23] Gong Z Q, Zhou L, Zhi R and Feng G L 2008 *Acta Phys. Sin.* **57** 5351 (in Chinese)
[24] Gong Z Q, Wang X J, Zhi R and Feng A X 2011 *Chin. Phys. B* **20** 079201
[25] Feng A X, Gong Z Q, Wang Q Q and Feng G L 2012 *Theor. Appl. Climatol.* **109** 635
[26] Donges J F, Zou Y, Marwan N and Kurths J 2009 *Eur. Phys. J. Special Topics* **174** 157
[27] Donges J F, Zou Y, Marwan N and Kurths J 2009 *Europhys. Lett.* **87** 48007
[28] Donges J F, Schultz H C H, Marwan N, Zou Y and Kurths J 2011 *Eur. Phys. J. B* **84** 635
[29] Dai X G, Wang P and Chou J F 2004 *Progress Natural Science* **14** 984
[30] Yatagai A, Arakawa O, Kamiguchi K and Kawamoto H 2009 *SOLA* **5** 137
[31] Quiroga R Q, Kreuz T and Grassberger P 2002 *Phys. Rev. E* **66** 041904
[32] Zhu K Z 1934 *Chin. Geogr. Soc.* **1** 1 (in Chinese)
[33] Tu C W and Huang S S 1944 *Acta Meteor Sin.* **18** 1 (in Chinese)
[34] Wu R and Wang B 2001 *Clim. Dyn.* **17** 277
[35] Wang B and Lin H 2002 *J. Climate.* **15** 386
[36] kummar K K, Rajagopalan B, Hoerling M, Bates G and Cane M 2006 *Science* **314** 115
[37] Shi N, Zhu Q G and Wu B G1996 *Scientia Atmospherica Sinica* **20** 575 (in Chinese)
[38] He S H, Huang H B, Zhang X, Liu Z X, Xu D S and Shen C K 2006 *Phys. Rev. E* **74** 057203