Rapid Communication

Preliminary research on the relationship between long-range correlations and predictability*

Zhang Zhi-Sen(张志森)[a)b)], Gong Zhi-Qiang(龚志强)[c)], Zhi Rong(支 蓉)[c)], Feng Guo-Lin(封国林)[a)b)c)†], and Hu Jing-Guo(胡经国)[a)‡]

[a)] *College of Physics Science and Technology, Yangzhou University, Yangzhou* 225002, *China*

[b)] *Key Laboratory of Regional Climate-Environment for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing* 100029, *China*

[c)] *National Climate Center, China Meteorological Administration, Beijing* 100081, *China*

(Received 5 August 2010; revised manuscript received 4 September 2010)

By establishing the Markov model for a long-range correlated time series (LRCS) and analysing its evolutionary characteristics, this paper defines a physical effective correlation length (ECL) τ, which reflects the predictability of the LRCS. It also finds that the ECL has a better power law relation with the long-range correlated exponent γ of the LRCS: $\tau = K\exp(-\gamma/0.3) + Y$, $(0 < \gamma < 1)$ – the predictability of the LRCS decays exponentially with the increase of γ. It is then applied to a daily maximum temperature series (DMTS) recorded at 740 stations in China between the years 1960–2005 and calculates the ECL of the DMTS. The results show the remarkable regional distributive feature that the ECL is about 10–14 days in west, northwest and northern China, and about 5–10 days in east, southeast and southern China. Namely, the predictability of the DMTS is higher in central-west China than in east and southeast China. In addition, the ECL is reduced by 1–8 days in most areas of China after subtracting the seasonal oscillation signal of the DMTS from its original DMTS; however, it is only slightly altered when the decadal linear trend is removed from the original DMTS. Therefore, it is shown that seasonal oscillation is a significant component of daily maximum temperature evolution and may provide a basis for predicting daily maximum temperatures. Seasonal oscillation is also significant for guiding general weather predictions, as well as seasonal weather predictions.

Keywords: long-range correlation, information entropy, effective correlation length, predictability

PACS: 92.60.Wc **DOI:** 10.1088/1674-1056/20/1/019201

1. Introduction

Most traditional time series models are based on the assumption that the events in the series are independent of each other.[1] However, recent research achievements in nonlinear science indicate that many natural records exhibit the persistence of time series in their records, i.e. long-range correlations found in meteorological data, DNA sequences, heart rate dynamics, hydrographic data and financial data,[2−10] thus suggesting that it is a universal law in nature.

Most current studies on the long-range correlativity of climatic series[11−14] quantitatively investigate the value of the long-range correlated exponent (LRCE) using the autocorrelation method developed early on in the field or the recently established detrended fluctuation analysis (DFA), and then provide qualitative explanations of the factors influencing the magnitude of the exponent. However, there are quite a few papers which deal with its practical application, such as papers exploring the relationship between the predictability of climatic systems or extreme weather-climate events and their long-range correlativity.[15−18] In recent years, studies have made progress in the predictability of climate systems:[19−34] Chou[19] used a cell-to-cell mapping theory to obtain the global predictability limit of a system; Mu,[20] Mu and Wang[21] highlighted the concepts of nonlinear singular vectors and values, and sought the effect of the nonlinear on predictability; Wang[22] investigated the predictability of the East Asian summer climate from two aspects of the persistence in monthly mean wind field anomalies and the global circulation anomalous features associated with the inter-annual

*Project supported by the National Natural Science Foundation of China (Grant Nos. 40930952, 40875040, and 41005043), the Special Project for Public Welfare Enterprises (Grant No. GYHY200806005) and the National Science/Technology Support Program of China (Grant Nos. 2007BAC29B01 and 2009BAC51B04).

†Corresponding author. E-mail: fenggl@cma.gov.cn

‡E-mail: jghu@yzu.edu.cn

© 2011 Chinese Physical Society and IOP Publishing Ltd

anomalies of the East Asian monsoon; Li[23] and Li *et al.*[24,25] demonstrated the dependence of the model's predictability on the mechanism and the model itself, also using an optimal numerical integration method, Li[23], Li *et al.*,[25] Ding and Li[26] studied spatially distributive characteristics in the predictability of circulation systems and gave its 3-D (dimensional) diagram. Some researchers[35−38] based on facts or characteristics have discussed the predictability of the climate. Wang *et al.*[11] examined the long-range correlations of the Lorenz system using the DFA method and found that the system possesses such typical long-range correlativity that the stronger the correlativity of the Lorenz system, the more predictable the system. Nevertheless, how to quantitatively describe the predictability of time series in terms of long-range correlativity is still one of the urgent scientific problems that need to be solved in nonlinear time series research.

We discuss predictability in this paper based on the information dissipation in the climate data, and emphasize the relationship between long-range correlations and predictability; this is different from the perspective of global[18−21] or local[22,35] climates in the past. We briefly describe the DFA and Markov model of time series, and then define the effective correlation length of the time series in Section 2; we perform a series of numerical experiments which produces an artificial long-range correlated series and establishes the power-law relationship between effective correlation length (ECL) and LRCE based on the analysis results of the artificial series in Section 3; and we then apply the daily maximum temperature series (DMTS) data to investigate the spatial distribution of the predictability of daily maximum temperature over China in Section 4, and give conclusions and some indications from the results of this paper in Section 5.

2. Research methods

2.1. Long-range correlations and the DFA method

Persistence of time series $\{x_i, i = 1, N\}$ is usually expressed with the autocorrelation function $C(h)$:

$$C(h) = \frac{1}{(N-h)\sigma^2} \times \sum_{i=1}^{N-h} (x_i - \overline{x})(x_{i+h} - \overline{x}) \propto h^{-\gamma}, \quad (1)$$

where $\bar{x}$ is the mean, σ the standard deviation and γ is the LRCE of $\{x_i\}$. If there is no persistence in the time series $\{x_i\}$, then it is not long-range correlated and its autocorrelation function is $C(h) = 0$; and if there is persistence in the series within the time scale h, when h is less than a threshold value h_c, the autocorrelation function $C(h) > 0$, after h exceeds h_c, the persistence disappears, $C(h) = 0$. But most observational time series are non-stationary, that is, where there is the superposition of different time scales or various noises. In 1994, Peng *et al.*[39] presented the DFA as a tool for detecting the long-range correlations in time series. DFA is able to treat the series with the statistical heterogeneity that arises from the non-stationary data and avoids the spurious detection of apparent long-range correlations by effectively removing the components of trends of various orders from the time series; DFA is able to detect the long-range correlativity of non-stationary time series. Therefore, the scaling exponent obtained by using the DFA is more reliable than that obtained using traditional methods;[40,41] the DFA method is currently applied in diverse scientific fields and is briefly introduced below.

For a time series $\{x_i, i = 1, N\}$, the profile y_n of the series is first calculated,

$$y_n = \sum_{i=1}^{n} \Delta x_i, \quad (2)$$

then, the profile series $\{y_i, i = 1, N\}$ is divided into N/h non-overlapping subseries with the same length of h. In an n-order DFA process, the local trend for each profile subseries of h length is fitted with an n-order polynomial, and subtracting those fitted local trends from the profile series $\{y_i, i = 1, N\}$ yields its fluctuation series. Finally, the root-mean-square fluctuation function $F_n(h)$ of the profile series $\{y_i, i = 1, N\}$ is obtained by averaging the fluctuation functions over all of the subseries:

$$F_n(h) = \left\langle \frac{1}{N_h} \sum_{i=1}^{N_h} y_h^2(i) \right\rangle^{1/2}. \quad (3)$$

If there is a scaling relation between $F_n(h)$ and window length h, then

$$F_n(h) \propto h^{\alpha}, \quad (4)$$

where α is the scaling exponent, which can be obtained by the linear fitting in the bi-logarithmic plot of $F_n(h)$ versus h. Different values of α may indicate different properties of the original time series:

(1) the $\alpha = 0.5$ suggests that the series is a random series, i.e. an independent stochastic process with $c(h) = 0$;

(2) the $0 < \alpha < 0.5$ indicates that the time series is only short-range correlated (short-range memory);

(3) the $0.5 < \alpha < 1$ means that values of the series are not independent of each other, i.e. the series is long-range correlated and cases where a rising (declining) trend before time t is present generally indicate also a rising (declining) trend after t, whose strengthening behaviour depends on the extent of $\alpha > 0.5$;

(4) the $\alpha = 1$ indicates the series representing a $1/f$ process; and

(5) the $\alpha = 1.5$ suggests that the series represents Brownian movement.

According to the stochastic walk theory, for a given window length h, the root-mean-square fluctuation function, $F_n(h)$, of the profile series is related to the autocorrelation function $C(h)$ of the original series. The relationship between γ and α can be written in the following form:

$$\gamma = 2 - 2\alpha. \tag{5}$$

Based on the research results of existing papers,[39,42,43] quadric DFA (DFA2) is used to analyse and test the series of daily maximum temperature in this paper.

2.2. Effective correlation length

For a Markov chain of discrete time, comprised of finite or infinite states, $e_1, e_2, e_3, e_4, \ldots, e_i, \ldots, e_k$, let $x_n = x(t_n)$ be the state at time $t = t_n(t_n = nT)$, $x_m \to x_{m+1} \to \ldots \to x_n \to \ldots$ represents the evolution process of the system for $n > m > 0$.

Let

$$p_i(m) = P\{x_m = e_i\} \tag{6}$$

denote the probability that the system is in state e_i at $t = t_m$, and

$$p_{ij}(m, n) = P\{x_n = e_j | x_m = e_i\}, \tag{7}$$

the probability that the system is in state e_i at $t = t_m$ and reaches state e_j at $t = t_n$ (it is independent of whatever state the system was before $t = t_m$). If $p_{ij}(m, n)$ only depends on the step number of the walk, then this Markov chain is a so-called time-continuous Markov chain. In this case, the transition probability is stationary, i.e.

$$P\{x_{m+n} = e_j | x_m = e_i\} \equiv p_{ij}(n) = p_{ij}^{(n)}, \tag{8}$$

which denotes the conditional probability for the system reaching state e_j after n step walk from state e_i. The one-step transition probability is

$$p_{ij} = P\{x_{n+1} = e_j | x_n = e_i\}, \tag{9}$$

then, the one step transition matrix for the homogeneous chain is

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1j} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2j} & \cdots & p_{2k} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ p_{i1} & p_{i2} & \cdots & p_{ij} & \vdots & p_{ik} \\ \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{kj} & \cdots & p_{kk} \end{bmatrix}, \tag{10}$$

and $\sum\limits_j p_{ij} = \sum\limits_j p\{x_{n+1} = e_j | x_n = e_i\} = 1$.

After the coarse graining and the transition matrix construction of a series $\{x_i\}$, the information entropy S for the series can be calculated according to[44]

$$S = \sum_{i=1}^{k} \sum_{j=1}^{k} f(i, j) \ln p(i, j), \tag{11}$$

where $f(i, j)$ is the joint distribution probability of $(x_t = i, x_{t+1} = j)(i = 1, 2, \ldots, k,\ j = 1, 2, \ldots, k)$. The information entropy S indicates the chaotic extent of the series and its quantity of effective information: a smaller S suggests that the series is more stationary, contains more effective information and therefore requires less additional information for prediction of the series; a larger S indicates that the series is more chaotic, has less effective information and therefore more additional information is needed for its prediction; and the largest S means that the series is a completely random series and hence unpredictable.

Theoretically, long-range correlativity is a global characteristic of a time series; but viewed from the angle of series prediction, effective information about the system states of a series beyond a certain time scale is dissipated out, thus they are meaningless for prediction. In this paper, we define the effective correlation length of a long-range correlated time series (LRCS) as the time-lag length at which effective information about the system states of the series is dissipated out. If the ECL of a series is longer, then it is more predictable, and vice versa. Therefore, the ECL in this paper is taken as a physical indicator of the predictability of time series.

3. Numerical experiments

To find the qualitative relationship between long-range correlations and predictability of time series, we have performed a series of numerical experiments. The procedures of numerical experiments and calculations are briefly described below (also see Fig. 1).

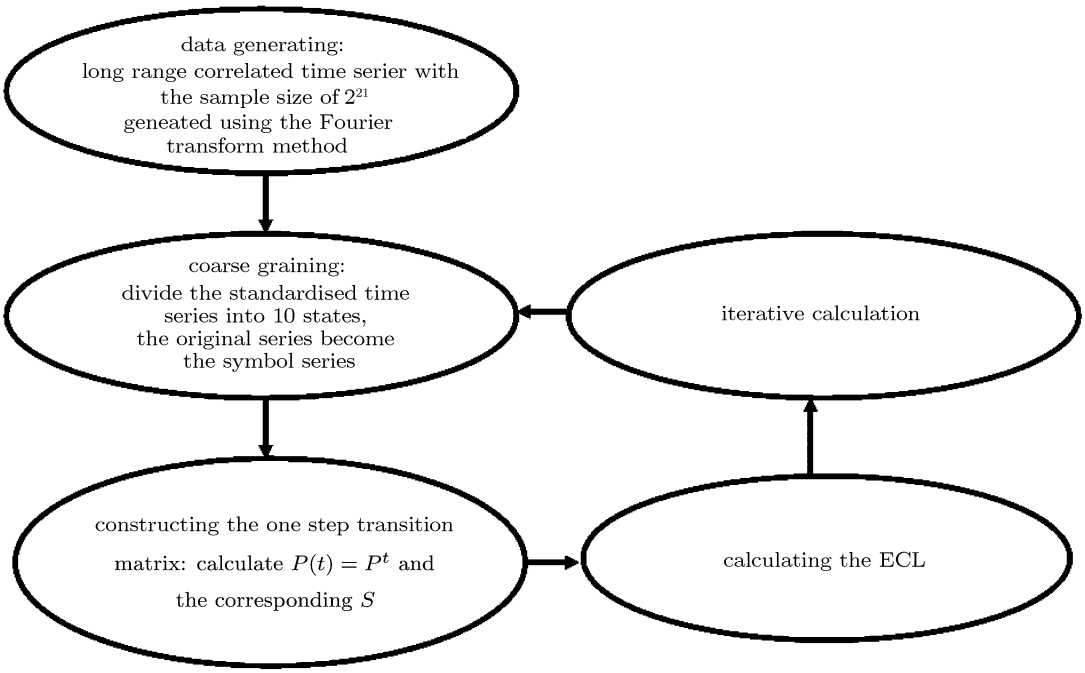


Fig. 1. Flow chart for numerical experiments.

The Fourier transform technique[45] was used to generate nine LRCS of the sample size of 2^{21} with the correlation exponent $\gamma = 0.1, 0.2, 0.3, \ldots, 0.9$, respectively; each series was first standardised and then converted into a symbol series by replacing the concrete magnitude of each data point of the series with one of 1, 2, ..., 10 according to the size of the magnitude, with 1 representing the least and 10 representing the largest magnitude state of the series. After counting the occurrence time number $F(i,j)$ of $(x_t = i, x_{t+1} = j)$ $(i = 1, 2, \ldots, 10, j = 1, 2, \ldots, 10)$ in the symbol series, the conditional probability $p(i,j) = F(i,j)/\sum_{j=1}^{10} F(i,j)$ was calculated, and then the one-step transition matrix constructed. We calculated $P(t) = P^t, (t = 1, 2, 3, \ldots, 30)$ and the corresponding information entropy S, performed the fit of S as a function of t in the $S - \lg(t)$ diagram. We defined τ as the ECL of the LRCS when $\mathrm{d}S/\mathrm{d}t|_{t=\tau} = 0.01S_0$ (τ is a real number, and S_0 is the largest entropy), according to the physical meaning of information entropy and computational practice.

Figures 2(a)–2(d) show the evolution of information entropy in the LRCS with $\gamma = 0.1$, 0.3, 0.5 and 0.7 with system evolution (i.e. increase in time lag); their ECLs are $\tau = 5.6$, 3.7, 2.7, 2.2, respectively. It can be seen from the figures that along with system evolution, the information entropies of the four series gradually all converge to 1, which means that the effective information series contained at current time is gradually dissipated out, and the additional information needed for prediction of the state of the systems is correspondingly increased; when the effectiveness of information becomes retrievable, the evolutionary time (time lag) is defined as the ECL (τ) of the series. If the evolutionary time step lengths are the same as in this paper, then the smaller the γ of a series is (the stronger the long-range correlativity of the series), the smaller the information entropy of the series (the more effective information the series possesses), and the slower the effective information decays in the system evolution (the larger ECL the series has). Therefore, the series possessing stronger long-range correlativity is more predictable. We performed the fit of the evolutionary curves of S with $\lg(t)$ in the $S - t$ diagram (Fig. 2) with the following polynomial:

$$S(t) = A[\lg(t)]^4 + B[\lg(t)]^3 + C[\lg(t)]^2 + D[\lg(t)] + E, \quad (0 < t \leq 30) \tag{12}$$

with the fitting coefficients, fitting errors and ECLs (τ) listed in Table 1. The fitting errors for the nine series with various values of γ are all of the order 10^{-4}, indicating that Eq. (12) better characterises the evolutionary characters of the information entropy of the series. According to the above definition of ECL, a value for the series with different values of γ, τ was also determined. Obviously, along with the increase of γ, τ exponentially declines. We investigate the quantitative relation between γ and τ below.

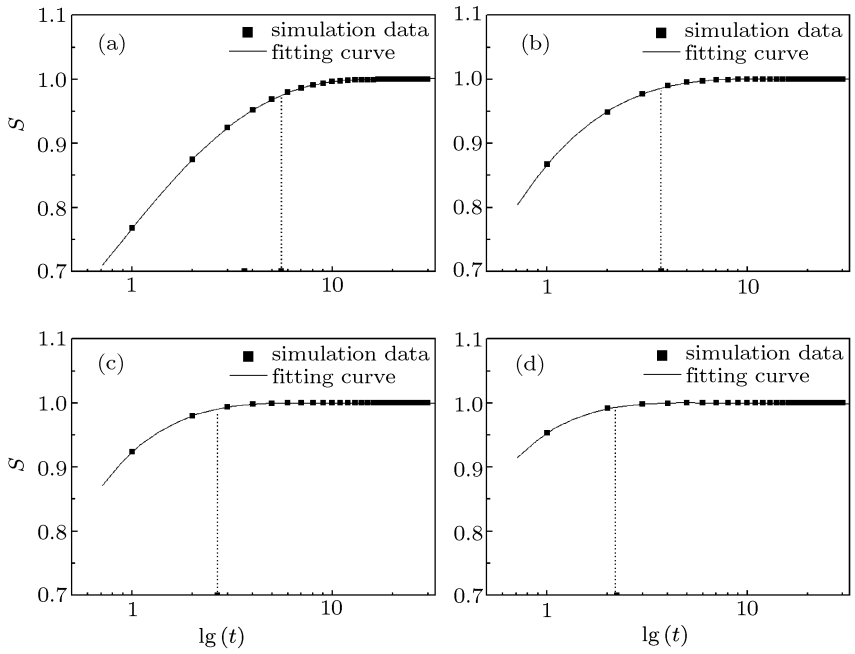


Fig. 2. Information entropy of the time series as a function of evolutionary step numbers with (a) $\gamma = 0.1$, (b) $\gamma = 0.3$, (c) $\gamma = 0.5$ and (d) $\gamma = 0.7$.

Table 1. Fitting coefficients and ECLs calculated from the nine artificial LRCSs generated by numerical experiments using the Fourier transform procedure[23] with γ= 0.1, 0.2, ..., 0.9, respectively.

γ	A	B	C	D	E	fitting errors	τ
0.1	0.08	–0.19	–0.05	0.39	0.78	3.97×10^{-4}	5.6
0.2	0.04	–0.03	–0.22	0.39	0.82	7.10×10^{-4}	4.5
0.3	–0.01	0.13	–0.35	0.37	0.87	6.45×10^{-4}	3.7
0.4	–0.04	0.21	0.40	0.33	0.90	3.86×10^{-4}	3.1
0.5	–0.05	0.24	–0.40	0.29	0.92	1.88×10^{-4}	2.7
0.6	–0.05	0.23	–0.36	0.24	0.94	1.72×10^{-4}	2.4
0.7	–0.05	0.21	–0.32	0.20	0.95	2.37×10^{-4}	2.2
0.8	–0.05	0.19	–0.28	0.17	0.96	2.79×10^{-4}	2.1
0.9	–0.04	0.18	–0.23	0.14	0.97	2.91×10^{-4}	1.9

The exponential decay of τ with increasing γ in Fig. 3 suggests the strong impact of long-range correlativity on the ECL of the series, that is, the impact of long-range correlativity on the predictabilities of the series is strong. The relation between τ and γ can be fitted as Eq. (13) at an accuracy of 0.999 (i.e. the fitting error is 8.17×10^{-4}),

$$\tau = K \exp\left(-\frac{\gamma}{0.3}\right) + Y, (0 < \gamma < 1), \qquad (13)$$

where K is a fitting coefficient, and Y is a parameter determined by the characteristics of the data and frequently specified as the mean value from multiple experiments. In the case of the nine artificial series in this paper, $K = 5.50$, $Y = 1.66$.

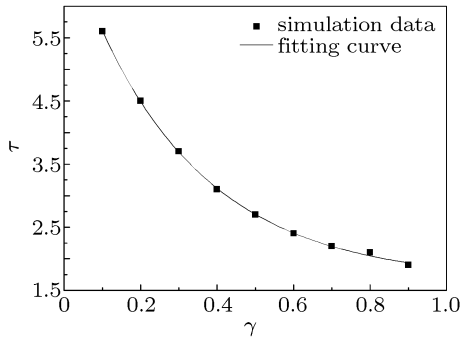


Fig. 3. The τ (ECL) as a function of γ (LRCE) for simulation and fitting.

4. Spatial distribution of the ECL for daily maximum temperatures

We applied the procedures for calculating the correlation exponent and length of the long-range correlated series in Section 3 to the daily maximum temperature data at the 325 stations, selected from China's 740 stations according to the continuity requirement of records, in the period from 1960 to 2005. Following the flow chart in Fig.1, we first calculated the information entropy of the DMTS at each station and plotted an $S-t$ diagram as Fig. 2, then determined the ECL of the series with Eq. (12) according to the condition $\mathrm{d}S/\mathrm{d}t|_{t=\tau} = 0.01S_0$, for instance, the ECL calculated for the Beijing DMTS $\tau_0 = 10.82$. Figure 4 shows the spatial distribution of the ECLs calculated at 325 stations throughout China.

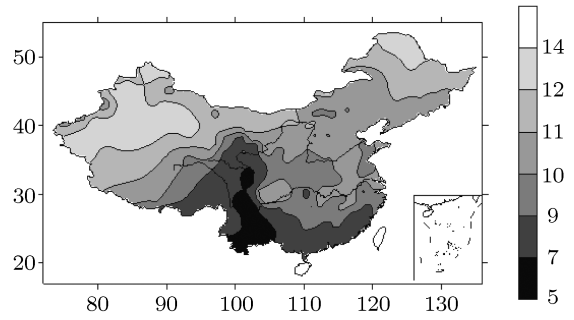


Fig. 4. Distribution of the ECLs of DMTS between 1960 and 2005 throughout China.

The spatial distribution of the ECL of DMTS (Fig. 4) displays two basic characters: firstly, the ECL was 10–14 days in Northwest China, North China, and Northeast China, 7–10 days in East China, South China, the southeast coastal region, and Sichuan Province, and 5–7 days in Yunnan Province; secondly, the ECLs were all less than 14 days, which further verifies the upper limit of two-weeks for the predictability of synoptic systems presented by Lorenz.[46,47] The ECL was smaller in the southeast side of the Qinghai–Xizang plateau and South China, but relatively larger in Northwest China and Northeast China. On the whole it was larger in northern China, but smaller in southern China, thus exhibiting a distinct zonal distribution. Therefore, daily maximum temperature is more predictable in western, northwestern and northern China than in eastern, southeastern and southern China. The reasons for such a distribution of ECL may be due to the impacts of the persistence of synoptic systems and the terrain of the Qinghai–Xizang plateau. The maximum temperature at a site is affected by various factors at the site, such as precipitation, cloud cover, wind speed and moisture content, etc. The region where the ECL of the temperature series was smaller, the east side of Qinghai–Xizang plateau, was influenced by warm/moist flows from the Indian Ocean and the warm Western Pacific Ocean. Moisture was abundant and synoptic processes changed rapidly there and these two factors affected the daily maximum temperature the most. Owing to the topographic effect of the plateau, weather on the east side was mainly governed by synoptic systems in the south branch of the westerlies, which greatly reduced the correlativity of daily maximum temperatures. In contrast, correlativity was strong in the desert regions of Northwest China where the moisture content in the air was much lower, and the weather frequently fine and relatively slow changing.

Figure 5(a) shows the distribution of the LRCE of the DMTS over China. The zonally distributive character of the LRCE in Fig. 5(a) is distinctive: the LRCE was smaller in northern China, but larger in southern China; thus the long-range correlativity of the DMTS is stronger in the north but weaker in the south, indicating that the DMTS is more predictable in northern China than in southern China. This is basically in agreement with the conclusion of the zonal distribution of ECL obtained from Fig. 4. To determine the K and Y of Eq. (13) for the DMTS in the years 1960–2005, we selected the ECL in Fig. 4 and the LRCE in Fig. 5(a) for some of the 325 stations, and performed numerical fit experiments by using Eq. (13), yielding fitting coefficients of Eq. (13) $K = 49.82$, and $Y = 2.66$, i.e.

$$\tau = 49.82 \exp\left(-\frac{\gamma}{0.3}\right) + 2.66, \quad (0 < \gamma < 1), \tag{14}$$

which was then used to derive the ECL in Fig. 5(b) from the LRCE in Fig. 5(a). The Person correlation coefficient was between Fig. 5(b) and Fig. 4 is 0.98. Therefore, both the magnitude order and spatial pattern of the ECL in Fig. 5(b) and Fig. 4 are in better accordance with each other, thus verifying that Eq. (13) is reliable and also that the ECL and LRCE of the DMTS better satisfy the corresponding quantitative exponential relation (14).

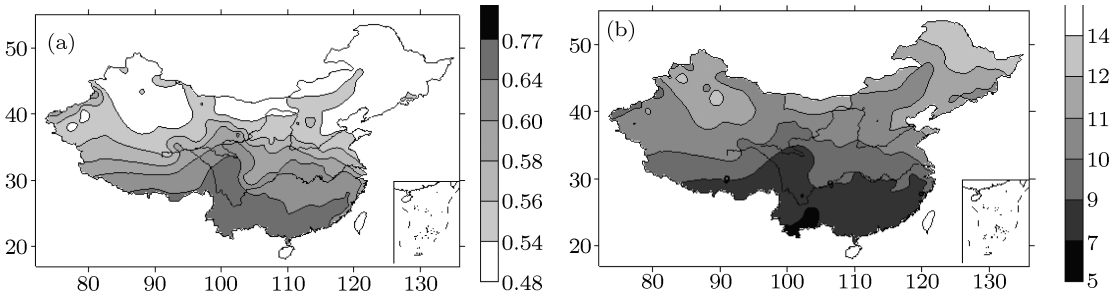


Fig. 5. Distributions of the LRCE (γ) and ECL (τ) for the DMTS over China in the years 1960–2005. The LRCE (a) was calculated from Eqs. (4) and (5), and the ECL (b) was obtained from the Eq. (14).

5. Factors affecting ECL

To further analyse factors affecting the ECL of DMTS, the following computations were performed for each station. The LRCE for a station (Beijing for example, $\gamma_1 = 0.53$) was first calculated from the DMTS (Fig. 6(a)) by using Eq. (5), and then the ECL ($\tau_1 = 11.25$) was obtained from $\gamma_1 = 0.53$ by using Eq. (14); the means of daily maximum temperature for each season over the years 1960–2005 were calculated, and their seasonal average was subtracted from the original DMTS to produce the seasonal oscillation-filtered DMTS of Beijing station (Fig. 6(b)), which was used to calculate the new LRCE ($\gamma_2 = 0.97$) and ECL ($\tau_2 = 4.62$); this procedure was repeated again except by subtracting the decadal linear trend from the original DTMS to calculate the LRCE ($\gamma_3 = 0.53$) and ECL ($\tau_3 = 11.25$) for the decadal linear trend-detrend DTMS (Fig. 6(c)). The spatial distributions of the calculated LRCE and ECL for all stations are given in Figs. 7(a), 7(c), 7(e), and 7(f), respectively.

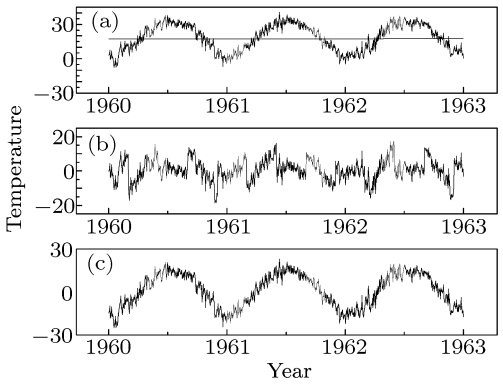


Fig. 6. Original observation data of DMTS (a), black line denoting linear trend with the slop of 6×10^{-5}, the distribution of DMTS with seasonal oscillation (b) and decadal linear trend (c) filtered at Beijing station, respectively.

The ECL for the seasonal oscillation-filtered DMTS is shown in Fig. 7(b), where it can be seen that the ECL was 4–5 days in Northwest China, Northeast China, East China and South China, and 5–7 days in central China and North China; its difference from the original DMTS (Fig. 4) is shown in Fig. 7(c), which inicates that the ECL was remarkably reduced and the reduction was 5–8 days in northern China, 3–5 days in central China, and 1–3 days in South China, Southwest China and Sichuan, etc. Therefore, the impact of the seasonal oscillation on the ECL gradually increased from south to north, which reflects the weather fact that seasonality is stronger in the mid-latitudes than in lower latitudes. The LRCE in Fig. 7(a) ranged from 0.7 to 1.0, and its difference from that for the original DMTS ranged from 0.3 to 0.5 in most areas of China, which indicates that after subtracting the seasonal oscillation signal from the DMTS, the ECL and LRCE changed a great deal, leading to larger changes in the predictability of the DMTS. Therefore, seasonal oscillation is an important factor which affects the long-range correlativity of the DMTS and is significant for 60–15 days extended range forecasts. In extended range forecasts, one should adhere to the concept of combining dynamic prediction and statistical prediction, and enhance the simulation of climate state. After filtering the seasonal oscillation signals, the ECL for the filtered DMTS is 5–6 days in most areas of China, which suggests that except due to seasonal changes, the variations of daily maximum temperature at a station are mainly affected by synoptic systems passing through, whose time scale is about 5–6 days as generally recognised by meteorologists.

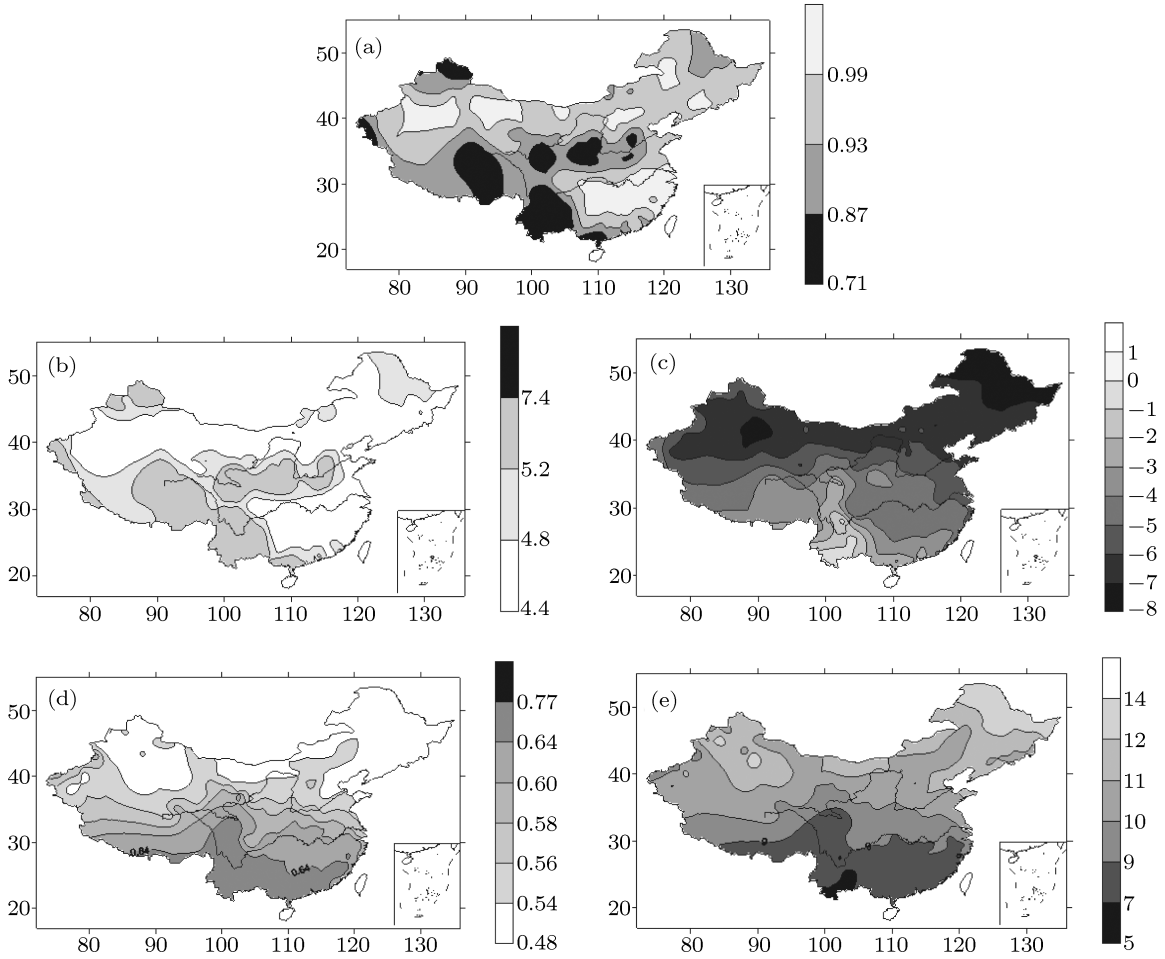


Fig. 7. Distributions of the γ and τ of seasonal oscillation-filtered DMTS (a), (b) and their deviations of τ from it of the original series (c), and of decadal linear trend-filtered DMTS (d), (e) in China.

Figures 7(d) and 7(e) show the distributions of the LRCE and ECL of the decadal linear trend-filtered DMTS over China, respectively. Comparing Fig. 7(e) with Fig. 4 we see that changes of the ECL are very small whether in its magnitude or in its spatial pattern. To corroborate this result, we constructed three artificial LRCS with $\gamma = 0.51$, 0.61, and 0.69 using the Fourier transform method, and then added different linear trends to the three series, respectively, and subsequently computed the ECL for the three linear trend-added artificial time series by using the DFA2, with the results plotted in Fig. 8. It can be observed from Fig. 8 that the effect of different linear trends on the calculated ECL of the three series are all smaller, this further confirms that the result of Figs. 7(d) and 7(f) is reliable. The reason for this may be that the DFA method is a de-trended algorithm, the trend term of the series has been first filtered out in the computation process; therefore, the linear trend term has little effect on the long-range correlativity or ECL of the series, and therefore little effect on the predictability of the series.

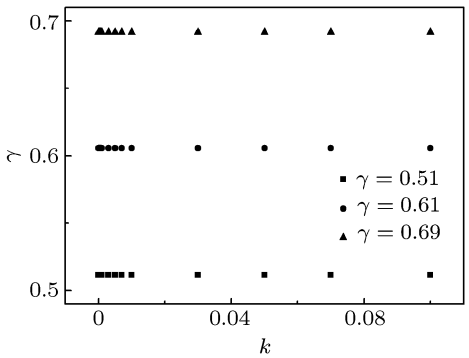


Fig. 8. The LRCE (γ) as a function of the linear trend term (slope k) added to the DMTS.

6. Conclusions

In this paper, we have established the Markov model for LRCS, analysed their evolutionary characteristics and then defined a physical parameter, ECL, of the series associated with the predictability of the series in terms of information entropy. We found that ECL (τ) has a better power law relation with the LRCE (γ): $\tau = K\exp(-\gamma/0.3) + Y$, $(0 < \gamma < 1)$, the coefficients K and Y are related to the background of LRCS and their dynamical characteristics. To investigate the predictability of the daily maximum temperature in China, we have further applied the power law relation and the Markov model to the DMTS at 740 stations for the period of years 1960–2005 and calculated the ECL of the DMTS. The results showed remarkable regional distributive features. The ECL is: about 10–14 days in Northwest China, North China and Northeast China; about 7–10 days in East China, South China, the southeast coastal region and Sichuan; and 5–7 days in Yunnan. That is, the ECL is longer in western, northwestern and northern China than in the east, southeast and south, suggesting that the predictability of DMTS is higher in the first three regions than in the second three regions. In addition, factors affecting the ECL of the DMTS have also been investigated, and the results show that seasonal oscillations of the DMTS had a significant effect on its ECL, while the decadal linear trend only had a small impact on its ECL. After subtracting the seasonal oscillation signals from the DMTS, the ECL was reduced by 5–8 days in northern China, 3–5 days in central China and 1–3 days in South China, Southwest China and Sichuan; indicating that the seasonal oscillation is of significance for extended-range forecasts. Therefore, ECL is a useful factor for determining the predictability of LRCS, but depends on the LRCE γ of LRCS and the background data such as seasonal oscillations and the overall trends etc. of LRCS. The results also show us that, in extended range predictions, one should adhere to the practice of combining dynamic prediction with statistical prediction as well as climate state simulation .

References

[1] Gumbel E J 1958 *Statistics of Extremes* (New York: Columbia University Press) p36

[2] Bunde E K, Bunde A, Havlin S and Goldreich Y 1996 *Physica* A **231** 393

[3] Bunde E K, Bunde A, Havlin S, Roman H E, Goldreich Y and Schellnhuber H J 1998 *Phys. Rev. Lett.* **81** 729

[4] Talkner P and Weber R O 2000 *Phys. Rev.* E **62** 150

[5] Eichner J F, Bunde E K, Bunde A, Havlin S and Schellnhuber H J 2003 *Phys. Rev.* E **68** 046133

[6] Gong Z Q, Feng G L, Wan S Q and Li J P 2006 *Acta Phys. Sin.* **55** 477 (in Chinese)

[7] Feng G L, Dong W J and Gong Z Q 2006 *The Non-linear Space-time Distribution Theory and Methods of Observational Data* (Beijing: China Meteorological Press) (in Chinese) p100

[8] Gong Z Q, Zou M W, Gao X Q and Dong W J 2005 *Acta Phys. Sin.* **54** 3947 (in Chinese)

[9] Feng G L and Dong W J 2003 *Chin. Phys.* **12** 1076

[10] Feng G L, Gao X Q, Dong W J and Li J P 2008 *Chaos, Solitons and Fractals* **37** 487

[11] Wang Q G, Zhi R and Zhang Z P 2008 *Acta Phys. Sin.* **57** 5343 (in Chinese)

[12] Feng G L, Wang Q G, Hou W, Gong Z Q and Zhi R 2009 *Acta Phys. Sin.* **58** 2853 (in Chinese)

[13] Zhang D Q and Qian Z H 2008 *Acta Phys. Sin.* **57** 4634 (in Chinese)

[14] Zhang D Q, Feng G L and Hu J G 2008 *Chin. Phys.* B **17** 736

[15] Eichner J F, Kantelhardt J W, Bunde A and Havlin S 2006 *Phys. Rev.* E **73** 016130

[16] Santhanam M S and Kantz H 2008 *Phys. Rev.* E **78** 051113

[17] Feng G L, Dong W J, Jia X J and Cao H X 2002 *Acta Phys. Sin.* **51** 1181 (in Chinese)

[18] Feng G L, Dai X G, Wang A H and Chou J F 2001 *Acta Phys. Sin.* **50** 606 (in Chinese)

[19] Chou J F 1989 *Adv. Atoms. Sci.* **6** 335

[20] Mu M 2000 *Science in China* (D) **43** 375

[21] Mu M and Wang J C 2001 *Science in China* (D) **44** 1128

[22] Wang H J 2005 *Chin. J. Atmos. Sci.* **29** 64 (in Chinese)

[23] Li J P 2000 *Bulletin of Chinese Academy of Science* **15** 428 (in Chinese)

[24] Li J P, Zeng Q C and Chou J F 2000 *Science in China* (E) **43** 449 (in Chinese)

[25] Li J P, Zeng Q C and Chou J F 2001 *Science in China* (E) **44** 55 (in Chinese)

[26] Ding R Q and Li J P 2008 *Acta Phys. Sin.* **57** 7494 (in Chinese)

[27] Shi N, Chen L W and Xia D D 2002 *Adv. Atmos. Sci.* **19** 6

[28] Feng G L and Dong W J 2003 *Acta Phys. Sin.* **52** 2347 (in Chinese)

[29] Dai X G, Wang P and Chou J F 2004 *Prog. Nat. Sci.* **14** 73

[30] Zou M W, Feng G L and Gao X Q 2006 *Chin. Phys* **15** 1384

[31] Hou W, Yang P and Feng G L 2008 *Acta Phys. Sin.* **57** 3932 (in Chinese)

[32] Gong Z Q, Wang X J, Zhi R and Feng G L 2009 *Acta Phys. Sin.* **58** 4342 (in Chinese)

[33] Yang P, Hou W and Feng G L 2008 *Acta Phys. Sin.* **57** 5333 (in Chinese)

[34] Gong Z Q, Zhou L, Zhi R and Feng G L 2008 *Acta Phys. Sin.* **57** 5351(in Chinese)
[35] Duan A M, Mao J Y and Wu G X 2004 *Plateau Meteorology* **23** 19 (in Chinese)
[36] Dai X G, Fu C B and Wang P 2005 *Chin. Phys.* **14** 850
[37] He W P, Feng G L, Dong W J and Li J P 2006 *Acta Phys. Sin.* **55** 969 (in Chinese)
[38] Zhang L, Zhang D Q and Feng G L 2010 *Acta Phys. Sin.* **59** 5896 (in Chinese)
[39] Peng C K, Buldyrev S V and Havlin S 1994 *Phys. Rev.* E **49** 1685
[40] Buldyrev S V, Goldberger A L and Havlin S 1995 *Phys. Rev.* E **51** 5084
[41] Gilberto E P, Jose A R and Alejandro V 2006 *Annals of Nuclear Energy* **33** 1309
[42] Lee J M, Kim I Y, Park K S and Kim S I 2002 *Comp. Biod. Medic.* **32** 37
[43] Kantelhardt J W, Bunde E K, Rego H A and Havlin A 2001 *Physica* A **295** 441
[44] Zhang X W and Ma L 1991 *Entropy Meteorology* (Beijing: China Meteorological Press) (in Chinese) p46
[45] Makse H A, Havlin S, Schwartz M and Stanley H E 1995 *Phys. Rev.* E **53** 5445
[46] Lorenz E N 1963 *J. Atmos.* **20** 130
[47] Lorenz E N 1976 *Quart. Res.* **6** 495