Article ID: 1006-8775(2001) 02-0163-12

PROGNOSTICATION AND INVESTIGATION OF EL NIÑO EVENT PROBABILITY

LI Ke (栗　珂) , LIU Yao-wu (刘耀武), YANG Wen-feng (杨文峰)，XU Xiao-hong (徐小红), ZHENG Xiao-hua (郑小华)

(*Shanxi Provincial Institute of Meteorological Sciences, Xi'an,* 710015 *China*)

ABSTRACT: Based on the El Nino event data sequence from 1854 to 1993, the nature of sequences was determined by using statistical normal and independent tests, etc. With the Markov random process and first order auto-regression predictive model, we set up the prognostication mode and give the time limit of the occurrence of next El Nino event, which probably occurs around 2002.The occurring probability for 2001 is 44 %, and it is 61 % for 2002.

Key words: El Niño event; probability prognostication; Markov process

CLC number: P732 **Document code:** A

1 INTRODUCTION

Being an important reflection of large-scale interactions between the oceans and atmosphere in the eastern tropical Pacific, the El Niño event is a strong signal for climate change. It inflicts heavy effects on the weather and climate across the globe, especially in the northern and northwestern China, which are exposed to drought or severe drought. It is therefore essential to make detailed study of the prediction of El Niño events.

Since the 1970's, studying the El Niño event has been on the rise in international community of academics[1, 2] and its prediction has been on the agenda. Owing to the complexity of the issue itself, the puzzle of physical mechanism for the formation of El Niño phenomena has not been completely solved. For that matter, Mark[3] pointed out that in principle it was impossible to predict the El Niño well ahead of time and it was probably determined by the inherent characteristics of the coupled air-sea system. It is due to this reason that existing air-sea coupled models have run into great difficulty, only operational on an experimental basis.

As early as in the 1980's, Barnett[4] used the statistic method to work on the prediction of the El Niño phenomena. Zhu et al[5]. attempted to forecast it using a grey model. Wang et al.[6] used the phase of solar spots to forecast the El Niño event. Few of these works, however, has any relevance to the prediction of probability for the formation of the event. According to Murphy[7], meteorological forecasting is in nature anything but definite and the study on chaotic dynamics has now expanded, with a clear trend, to stochastic systems[8]. In view of the development, the current work presents probabilities expected to occur with a forthcoming El Niño event by exploiting the probability diagnostic approach used in the geophysical field for earthquake generation[9] and studying the El Niño event based on related dataset sequence and Markov stochastic processes.

Received date: 2000-04-25; **revised date:** 2001-08-02

Foundation item: Research project of meteorological science and technology in China (96-908-05-03)

Biography: LI Ke (1944 –), male, native from Xi'an City Shanxi Province, associate professor, mainly undertaking the study of climatic prediction and application.

2 DATA SEQUENCE

2.1 *Source of data*

Reference [1] made an investigation of the SST in the region 0° ~ 15°S, 90°W ~ 180°W with the discovery that there were a total of 32 El Niño events from 1854 to 1993 (Tab.1).

Tab. 1 The El Nino events and its time intervals between 1854 and 1993

ElNinoYr	1855	1858	1864	1868	1877	1880	1884	1888	1891	1896	1899
t_i		3	6	4	9	3	4	4	3	5	3
X_i		0.477	0.778	0.602	0.954	0.477	0.602	0.602	0.477	0.699	0.477
ElNinoYr	1902	1904	1911	1913	1918	1925	1930	1939	1940	1944	1951
t_i	3	2	7	2	5	7	5	9	1	4	7
X_i	0.477	0.301	0.845	0.301	0.699	0.845	0.699	0.954	0.000	0.602	0.845
ElNinoYr	1953	1957	1963	1965	1968	1972	1976	1982	1986	1991	
t_i	2	4	6	2	3	4	4	6	4	5	
X_i	0.301	0.602	0.778	0.301	0.477	0.602	0.602	0.778	0.602	0.699	

For an El Niño event, the standard definition goes as:

(1) Peak value is as high as 10°C in the positive anomaly of SST and the positive anomaly lasts more than 12 months persistently;

(2) A year with peaks of positive anomaly of SST is defined to be an El Niño year;

(3) If the positive SST anomaly lasts over 6 months in the year before or after the El Niño year, it is defined to be a successive year of the same El Niño event;

(4) When the SST anomaly varies in a manner like the El Niño event but with the positive peak anomaly less than 1.0°C or the positive anomaly lasts for less than 12 months, then a weak El Niño event is defined to be prevalent.

The relevant data of El Niño events are cited from reference [1].

2.2 *Pre-processing of data*

Based on the sequence of El Niño events, a sequence of time interval $\{t_i\}$ is obtained for the generation of El Niño events. To improve the linearity and stability of the sequence, a logarithmic transformation is conducted for the sequence $\{t_i\}$ so that

$$X_i = \lg t_i$$

From the sequence $\{X_i\}$ (Tab.1), information can be extracted that is useful for prediction.

The definition of time intervals are as follows. The interval starting from the last El Niño event till the present is set as t_0 and that between the present and next El Niño event is set as t_1. It is obvious that the interval between any two El Niño events is $t = t_0 + t_1$.

3 TESTING OF SEQUENCE

3.1 *Normality test*

A test of normality is conducted to determine whether the sequence $\{X_i\}$ observes the distribution of normal distribution, which is easier for processing and derivation in statistical study.

Set the size of sample as n in the sequence $\{X_i\}$, the mean and standard deviation of X_i (estimated without departure) are respectively

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \,; \qquad\qquad S_1 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

$\overline{X}$ and S_1^2 are used to replace the overall mean m and variance o^2 respectively.

It is assumed that $\{X_i\}$ observes the normal distribution. It is divided into k groups and the frequency observed with each of the groups is n_i, the theoretic probability for it is $P_i(i=1, 2, \cdots, k)$. Then from the table of normal distribution, we have[11]

$$P_i = \Phi\left(\frac{X_{i+1} - \overline{X}}{S_1}\right) - \Phi\left(\frac{X_i - \overline{X}}{S_1}\right)$$

The theoretic frequency $m_i = nP_i$ for a group. The normal distribution is used to have a fitting of X_i. When n is sufficiently large, the Pearson criterion

$$c^2 = \sum_{i=1}^{k}\frac{\left(n_i - nP_i\right)^2}{nP_i} \tag{1}$$

is used to differentiate the goodness of fitting. This statistic approximately observes the distribution of c^2 at the time. As the normal distribution of fitting has three linear conditions: $m = x^2, s^2 = S_1^2, \sum m_i = \sum n_i$, the degree of freedom $f = k-3$. If $c^2 \langle c_a^2$ using Eq.(1), then it is suggested that there is not much difference between n_i and m_i and $\{X_i\}$ observes the normal distribution; if $c^2 \rangle c_a^2$, then $\{X_i\}$ does not observe it.

3.2 *Independence test*

For the sequence $\{X_i\}$, some of the autocorrelation among the X_i are more significant than the other. It necessitates a test of independence for $\{X_i\}$. Then patterns of prediction models are determined based on the test result.

In the context of maximum ratio of similarity, the autocorrelation coefficient of the sequence is computed by the expression of

$$R_k = \frac{2\sum\left(X_{i-k} - \overline{X}_k\right)\left(X_i - \overline{X}_0\right)}{\sum\left(X_{i-k}\overline{X}_k\right)^2 + \sum\left(X_i - \overline{X}_0\right)^2} \tag{2}$$

where

$$\overline{X}_0 = \frac{1}{n-k}\sum_{i=k+1}^{n} X_i\ ; \qquad \overline{X}_k = \frac{1}{n-k}\sum_{i=1}^{n-k} X_i\ ;$$

k is the number of orders in the autocorrelation, usually taken as k = 1, 2, 3, 4.

The R_k, derived with Eq.(2), is compared with $\left|R_k\right|_2$ that is with a confidence of a [9]. If $R_k < \left|R_k\right|_2$, the correlation is considered insignificant and $\{X_i\}$ an independent sequence; if $R_k \geq \left|R_k\right|_2$, the correlation is thought to be significant and $\{X_i\}$ a dependent sequence.

3.3 *Tendency test*

For the sequence $\{X_i\}$, a tendency test has to be done to determine whether there is a definitive principal value. Generally, only the tendency for straight lines is considered for simplicity.

Set the equation of straight lines as

$$X_i^* = a + bU_i \tag{3}$$

in which $U_i = i - (n+1)/2$. Using the least square method, we have

$$\hat{a} = \frac{1}{n}\sum_{i=1}^{n} X_i\ ; \qquad \hat{b} = \sum_{i=1}^{n} X_i U_i \Big/ \sum_{i=1}^{n} U_i^2 \tag{4}$$

It is assumed that there is no tendency for straight lines in the sequence $\{X_i\}$, then $\hat{b} = 0$. Consequently, the value of expectation $E\left(\hat{b}\right) = 0$; The variance $V\left(\hat{b}\right) = s^2 \Big/ \sum U_i^2$. As a result, the statistic

$$\frac{\hat{b} - 0}{\sqrt{s^2 \Big/ \sum U_i^2}} = \hat{b}\sqrt{\sum U_i^2} \Big/ s$$

observes the normal distribution of

$$\hat{b} = \sqrt{\sum U_i^2} \Big/ s \sim N(0, 1)$$

The standard deviation of the sample is used to conduct an unbiased estimate of

$$S_1^2 = \frac{\sum\left(X_i - \overline{X}\right)^2 - \hat{b}\sum X_i U_i}{n-2},$$

which substitutes s^2 to obtain that

$$t = \frac{\hat{b}\sqrt{(n-2)\sum U_i^2}}{\sqrt{\sum\left(X_i - \overline{X}\right)^2 - \hat{b}\sum X_i U_i}} \tag{5}$$

in which the degree of freedom is $f = n - 2$.

By table look-up, we obtain t_a[10]. If $t < t_a$ is determined, using Eq.(5), the sequence $\{X_i\}$ is considered to have insignificant tendency for straight lines; if $t > t_a$, the sequence is considered to have significant tendency for straight lines.

3.4 *Test results of the sequence*

(1) Test of normality: For the sequence $\{X_i\}$, the sample size n = 31. We get $\bar{x} = 0.5933$ and $S_1 = 0.2187$. Divide X_i into 5 groups and derive all parameters for the normality test following procedures in 3.1. The results are listed in Tab.2.

Tab. 2 The normality test of the sequence $\{X_i\}$

$X_i \sim X_{i-1}$	$t_i = (X_{i-1} - \bar{X})/S$	$\Phi(t_i)$	p_i	np_i	n_i	$(n_i - nP_i)^2$	$(n_i - nP_i)^2 / nP_i$
~0.301	-1.374	0.0853	0.0853	2.64	5	5.570	2.110
0.301~0.481	−0.533	0.2981	0.2128	6.60	6	0.360	0.055
0.481~0.661	0.308	0.6217	0.3236	10.03	8	5.290	0.529
0.661~0.841	1.150	0.8749	0.2532	7.85	7	0.723	0.092
0.841~		1.0000	0.1251	3.88	5	1.254	0.323
Σ			1.0000	31	31	13.197	3.109

Based on Tab.2 and Eq.(1), the Pearson criterion $x^2 = 3.109$ is obtained. Following $f = k - 3 = 2$ and a =0.05, $x_z^2 = 5.991$ is known by table look-up[11] for the distribution of x^2. As $x^2 < x_a^2$, there is not much difference between m_i and n_i. It is then known that the sequence $\{X_i\}$ observes the distribution of normality.

(2) Test of independence: For the sequence $\{X_i\}$, the autocorrelation coefficient R_1 = –0.3160, R_2 = –0.0464, R_3 = 0.1469, R_4 = –0.2690 are obtained following the method in 3.2 and using Eq.(2). When the significance level $2a$ = 0.1 is assumed, $f = n - k - 1$, the critical value R_k[9] is $\left|R_k\right|_{2a}$, which is respectively shown as $\left|R_1\right|_{2a}$ = 0.301, $\left|R_2\right|_{2a}$ = 0.306, $\left|R_3\right|_{2a}$ = 0.315 and $\left|R_4\right|_{2a}$ = 0.317. It is obvious that $\left|R_1\right| > \left|R_1\right|_{2a}$. It is then established that the sequence $\{X_i\}$ has a significant tendency for straight lines.

(3) Test of tendency: With the method in 3.3, $\overline{X}$ = 0.5933, S_1 = 0.2187 are obtained. $\hat{b}$ = –0.0004 and $|t|$ = 0.1042 are obtained using Eq.(4) and Eq.(5), respectively. Following $f = n - 2$ = 29, a = 0.05, t_a = 2.045 is known by table look-up. As $|t| < t_a$, it is known that the sequence

$\{X_i\}$ has an insignificant tendency for straight lines.

With the statistic tests shown above, it is now clear that the sequence $\{X_i\}$ observes the normality distribution with significant first order autocorrelation and insignificant tendency for straight lines. Then, the sequence $\{X_i\}$ is considered a stable Markov stochastic process.

4 PREDICTIVE MODELS

4.1 *Basic ideas of probability prediction*

For predictions of probability in this particular case, information useful for the prediction is extracted from the sequence of the El Niño event only. The prediction is conducted by giving a probability for a designated period of the event to occur or a time interval between the present and succeeding events.

Wherefore, the distribution function of $\boldsymbol{\tau}$ is first derived from the known sequence $\{X_i\}$ by

$$F(t)=\int_0^t f(t)\mathrm{d}t$$

Specifically, $f(t)$ is the function of probability density. Then, for the time after t_0, the probability for an El Niño event to take place is that

$$\int_{t_0}^{\infty} f(t)\mathrm{d}t = 1-F(t_0)$$

Thus, within the period of t_1, the conditional probability for an El Niño event to take place is expressed by

$$G(t_1)=\int_0^{t_1} g(t_1)\mathrm{d}t_1=\frac{F(t_0+t_1)-F(t_0)}{1-F(t_0)}$$

in which $g(t_1)$ is the conditional probability density of t_1.

A key question here is that the sequence $\{X_i\}$ must conform to the normal distribution. As long as it fits in the requirement, the distribution function of X, $\Phi(X)$, can be derived, which is then used to obtain $F(t)$. From $F(t)$ and t_0, $G(t_1)$ is known. This is the basic idea of probability prediction.

4.2 *Markov process prediction model*

4.2.1 AUTO-REGRESSION EQUATION

The (n - 1) groups of data are used to fit an auto-regression equation of

$$X_i=a_0+a_1\left(X_{i-1}-\overline{X}_1\right)$$

where $i = 2, 3, \cdots, n$; $\overline{X}_1 = \dfrac{1}{n-1}\sum_{i=2}^{n} X_{i-1}$.

Applying the least square method, it is known that

$$\hat{a}_0 = \overline{X}_0 = \frac{1}{n-1}\sum_{i=2}^{n} X_i \; ; \; \hat{a}_1 = \frac{\sum_{i=2}^{n}\left(X_{i=1} - \overline{X}_1\right)\left(X_i - \overline{X}_0\right)}{\sum_{i=2}^{n}\left(X_{i-1} - \overline{X}_1\right)^2} \tag{6}$$

Then, a first-order auto-regression equation

$$\hat{X}_{n+1} = \hat{a}_0 + \hat{a}_1\left(X_n - \overline{X}_1\right) \tag{7}$$

is obtained. To determine whether Eq.(7) has significant effect of regression, a test of F has to be done. The statistic of

$$F = \frac{U/m}{Q/(n-m-1)} \tag{8}$$

is used where n is the size of sample, m is the number of factors, $f_1 = m,\ f_2 = n - m - 1$, U and Q are the sum of regression square and sum of residual difference square, as in

$$U = \sum_{i=1}^{n-1}\left(\hat{X}_i - \overline{X}_1\right)^2, \qquad Q = \sum_{i=1}^{n-1}\left(X_i - \hat{X}_1\right)^2 \tag{9}$$

F_a is looked up in the table[10]. If $F \geq F_a$ using Eq.(8), then Eq.(7) has significant effect of regression; if $F < F_a$, then the regression is considered insignificant.

4.2.2 PREDICTIVE MODEL OF PROBABILITY

If $\hat{X}_{n+1}$ is predicted based on X_n with Eq.(7), the auto-regression equation, the mathematical expectation $E\left(X_{n+1} - \hat{X}_{n+1}\right)$ =0 is satisfied for the prediction error of $X_{n+1} - \hat{X}_{n+1} = X_{n+1} - \overline{X}_0 - \hat{a}_1\left(X_n - \overline{X}_1\right)$. As the new X_{n+1} is independent from $\overline{X}_0$ and $\hat{a}_1$ and their variance are separately

$$\begin{cases} V\left(\overline{X}_0\right) = s^{2}/(n+1) \\ V\left(\hat{a}_1\right) = s^{2}\sum\left(X_{i+1} - \overline{X}_1\right)^2 \\ V\left(\hat{X}_{n+1}\right) = s^{2}\left(\dfrac{1}{n-1} + \dfrac{\left(X_n - \overline{X}_1\right)^2}{\sum\left(X_{i-1} - \overline{X}_1\right)^2}\right) \end{cases}, \tag{10}$$

Therefore, from Eq.(10) we get the variance of prediction error

$$V\left(X_{n+1}-\overline{X}_{n+1}\right)=s^{2}\left(1+\frac{1}{n-1}+\frac{\left(X_{i}-\overline{X}_{1}\right)^{2}}{\sum\left(X_{i-1}-\overline{X}_{1}\right)^{2}}\right)$$

As a result,

$$\left(X_{n+1}-\hat{X}_{n+1}\right)\sim N\left(0,\ s\sqrt{1+\frac{1}{n-1}+\frac{\left(X_{i}-\overline{X}_{1}\right)^{2}}{\sum\left(X_{i-1}-\overline{X}_{1}\right)^{2}}}\right)$$

Substituting the variance of sample

$$S_{1}^{2}=\frac{\sum\left(X_{i}-\overline{X}_{0}\right)-\hat{a}_{1}\sum\left(X_{n-1}-\overline{X}_{1}\right)\left(X_{i}-\overline{X}_{0}\right)}{n-3}$$

for s^{2}, we obtain that

$$t=\frac{X_{n+1}-\overline{X}_{0}-\hat{a}_{1}\left(X_{n}-\overline{X}_{1}\right)}{\sqrt{\frac{S_{00}-\hat{a}_{1}S_{01}}{n-3}\left(n+\frac{\left(X_{n}-\overline{X}_{1}\right)^{2}}{S_{11}}\right)}},\tag{11}$$

that observes the distribution of t with the freedom degree of $f = n$ -3. Specifically,

$$\begin{cases}S_{00}=\frac{1}{n-1}\sum\left(X_{i-1}-\overline{X}_{1}\right)^{2}\\ S_{01}=\frac{1}{n-1}\sum\left(X_{i-1}-\overline{X}_{1}\right)\left(X_{i}-\overline{X}_{0}\right)\\ S_{11}=\frac{1}{n-1}\sum\left(X_{i-1}-\overline{X}_{1}\right)^{2}\end{cases},\tag{12}$$

The distribution function of the statistic t is expressed by

$$F(t)=\int_{-\infty}T_{n-3}(t)\mathrm{d}(t)$$

in which $T_{n-3}(t)$ is the function of probability density in the t distribution. Thus, the probability can be predicted by

$$\Phi\left(X_{n+1}\right)=F\left[t\left(X_{n+1}\right)\right]$$

in which $t\left(X_{n+1}\right)$ is the general form of Eq.(11). It is therefore possible that

$$F\left(t_{n+1}\right)=\Phi\left(X_{n+1}\right)=F\left[t\left(X_{n+1}\right)\right]$$

or,

$$F(t_{n+1}) = P[t \le t(X_{n+1})] \tag{13}$$

In this way, the predicted probability for an El Niño event to take place is given by Eq.(13) for a specific forthcoming time interval of t_{n+1}.

Under the condition that there are not any El Niño events within t_0, the conditional probability for the event to take place with the elapse of t_1 is expressed below, based on computations using Eq.(13).

$$G(t_1) = \frac{F(t_0 + t_1) - F(t_0)}{1 - F(t_0)} \tag{14}$$

In the context of the given probability *P*, as

$$\int_{-\infty}^{(P)} T_{n-3}(t)\mathrm{d}t = P, \quad f = n-3,$$

we can have *t*(*P*) by looking up the *t* table[11]. Then from Eq.(11), we can have $\hat{X}_{n+1}$ and its corresponding time interval $\hat{t}_{n+1}$ in

$$\begin{cases} \hat{X}_{n+1} = \overline{X}_0 + \hat{a}_1(X_n - \overline{X}_1) + t(P)\sqrt{\dfrac{S_{00} - \hat{a}_1 S_{01}}{n-3}\left(n + \dfrac{(X_n - \hat{X}_1)^2}{S_{11}}\right)} \\ \hat{t}_{n+1} = 10^{X_{n+1}} \end{cases} \tag{15}$$

Thus, given a particular probability *P*, the time limit is derived for the next possible El Niño event to occur, based on Eq.(15).

5 RESULTS OF PREDICTION

5.1 *Predictions of the generation probability*

Employing the computational method in 4.2.1, we obtain that $\hat{a}_0 = \overline{X}_0 = 0.5993$, $\hat{a}_1 = -0.3176$; $\overline{X}_1 = 0.5919$. Then from Eq.(7), we have

$$\hat{X}_{n+1} = 0.5993 - 0.3176(X_n - 0.5919) \tag{16}$$

The *F* test is performed for Eq.(16). *F* = 0.315 is known using Eq.(8), following $f_1 = 1$, $f_2 = n - m - 1 = 29$. Looking up the *F* distribution table[10], we obtain that $F_a = 2.89$ ($a = 0.1$). As $F > F_a$, Eq.(16) has significant regression with 90% confidence level.

Since the 1997 El Niño event, there has not been any in 1998. To predict the probability for it to happen in year 2000, $F(t)$ and $G(t_1)$ are derived using $t_{n+1} = 3$, $t_0 = 1$, $t_1 = 2$.

Following the predictive model of probability in 4.2.2, we have $S_{00} = 0.0453$, $S_{01} = -0.0144$, $S_{11} = 0.0454$ using Eq.(12); Eq.(11) is then used to obtain $t_{n-3} = -0.7436$. Following

$f = n - 3 = 28$, we know by looking up the t distribution table[10] that the distribution function of t_{n-3} is $P(t \le t_{n-3}) = 0.227$. As a result, from Eq.(13) $F(t_{n+1}) = F(3) = 0.227$ is obtained. The El Niño event to take place in 2000 is predicted to be 0.227.

Likewise, $F(t_0) = F(1) = 0.001$ is known. For the last step, Eq.(14) is used to obtain $G(t_1) = G(2) = 0.226$. In other words, under the condition that there was not any El Niño event in 1998 with the elapse of 2 years, the probability for it to take place in 2000 is 0.226.

Following similar procedures, the probability for the El Niño event to happen in 1999 ~ 2006 is predicted and the result is presented in Tab.3.

Tab. 3 The forecasts of probability with the El Niño event from 1999 to 2006 ($t_0 = 1$)

year	1999	2000	2001	2002	2003	2004	2005	2006
r	2	3	4	5	6	7	8	9
$\hat{F}(r)$	0.061	0.227	0.441	0.618	0.742	0.841	0.894	0.933
r_1	1	2	3	4	5	6	7	8
$\hat{G}(r_1)$	0.060	0.226	0.440	0.617	0.741	0.840	0.893	0.932

It is known from Tab.3 that prior to year 2000, the El Niño event would take place with a probability of $\hat{P} \le$ 0.227; $\hat{P} \ge$ 0.441 from 2001 onwards. In addition, the El Niño data from 1855 to 1976 are used to set up a model to conduct retrospective tests of generation probability of the El Niño event in 1982, 1986 and 1991. The results show P at 0.691, 0.4101, and 0.520, respectively. It is obvious that the El Niño event will not take place unless P is at least 0.401. It is based on it that a fresh El Niño event is predicted to occur at the earliest in 2001 or later.

5.2 *Predictions of the generation time*

Given a specific probability $P(\%) = 10$, 20, 30, 40, 50, 60, 70, 80, 90, Eq.(15) is used to determine the time intervals for the next El Niño event $\hat{t}_{n+1}$ and the results are shown in Tab.4.

Tab. 4 The forecasts of time intervals of El Niño event occurrence given the probability

P (%)	10	20	30	40	50	60	70	80	90
$\hat{X}_{n+1}$	0.355	0.453	0.522	0.581	0.635	0.690	0.748	0.818	0.915
$\hat{t}_{n+1}$	2.3	2.8	3.3	3.8	4.3	4.9	5.6	6.6	8.2
year	2000	2000	2001	2001	2002	2002	2003	2004	2006

It is known from Tab.4 that given the probability $P \le 40$, $\hat{t}_{n+1} \le 3.8$ (year), i.e. the El Niño event will happen at a probability P no more than 40% before 2001; when P = 50%, $\hat{t}_{n+1}$ = 4.3, i.e. the probability for an El Niño event to take place has reached 50% from 2002 onwards. Based on the analysis, it is possible that another El Niño event will appear around the year 2002.

Summarizing the forecast results above, we know that the El Niño event will again show itself around 2002, the probability being 44% for 2001 and 61% for 2002.

Being one of the important factors affecting the global climatic change, the El Niño event plays a complicated role in the climate of China. It is shown as inhomogeneous distribution of precipitation from the south of China to the north in the summer and autumn, resulting in floods in the south but drought in the north of the nation[11]. The emphasis should be on the prevention of a major flash flood in the valley of Changjiang River and a major drought in the northern (especially northwestern) China, around the year 2002.

6 CONCLUDING REMARKS

a. A sequence $\{t_i\}$ is set up based on the sequence of the El Niño event. Transformed logarithmically, it changed to a sequence $\{X_i\}$, which is improved in terms of the linearity and stability that information for forecasting is made easier to extract.

b. The nature of the sequence $\{X_i\}$ is isolated through a number of statistical tests. The result shows that $\{X_i\}$ observes the normal distribution with significant first-order auto-correlation but insignificant tendency for straight lines. It can be taken as a stable Markov stochastic process.

c. In view of the nature of $\{X_i\}$, one learns that the selection of a predictive model employing a stable Markov process is reasonable and forecasting of its probability on basis of it is feasible.

d. According to the prediction, a new round of El Niño event will take place around 2002. It is expected that it will come by a probability of 44% in 2001 and 61% in 2002. Precautions should be strengthened to fight against both dry and wet climate over this period.

e. The current work collects forecasting information from the sequence $\{t_i\}$, without any discussion of the physical mechanism for forming the El Niño event. It is a well-known shortage of statistical prediction. It is then known that the probability predictions are only preliminary, which are to be tested in practice for improvement.

Acknowledgements: Mr. CAO Chao-xiong, who works at the Guangzhou Institute of Tropical and Oceanic Meteorology, has translated the paper into English.

REFERENCES:

[1] LI Zhi-gang, QIAN Zheng-an. The El Niño phenomena and climate anomalies [J]. *Atmospheric Information*, 1984, **14** (4): 1-12.
[2] CHAO Ji-ping. More study on mechanisms forming the El Niño and predictions [J]. *Atmospheric Science Research and Application*, 1999, (1): 111-113.
[3] MARK A C. Experimental forecasts of El Niño [J]. *Meteorological Science and Technology*, 1987, (3): 42-47.
[4] BARNETT T P. Prediction of the El Niño of 1982-1983 [J]. *Monthly Weather Review*, 1984, **112** (7): 1403-1407.
[5] ZHU Zheng-xin, CAO Hong-xing. Use of grey model for predicting the El Niño [J]. *Journal of Tropical Meteorology*, 1994, **11**: 361-365.
[6] LI Bing. Chinese scientists predicting next El Niño event [N]. *Chinese Meteorology News*, 1998-05-18(3).
[7] MURPHY A, KATG R W. Probability, statistics and decision making in the atmospheric sciences [M].

Boulder and London: Westview Press, 1985. 36-58.
[8] GUO Bing-rong, JIANG Jian-min, CHOU Ji-fan, et al. Non-linear Characteristics of Climate Systems and the Prediction theory [M]. Beijing: Meteorological Press, 1996. 133-136.
[9] XU Zhong-ji, WEI Gong-yi, SONG Liang-yu, et al. Probability of the time of earthquake occurrence (I) [J]. *Acta Geophysica Sinica*, 1974, **17** (1): 51-71.
[10] Mathematics Institute of Chinese Academy of Sciences, Conventional Table of Mathematics and Physics [M]. Beijing: Science Press, 1974. 7-14.
[11] ZHAO Zhen-guo. Impact of El Niño events on atmospheric circulations in the Northern Hemisphere and precipitation in China [J]. *Scientia Atmospherica Sinica*, 1996, **20** (4): 422-428.