Article ID: 1006-8775(2015) S1-0034-12

CORRELATION ANALYSIS OF THE WESTERN PACIFIC SUBTROPICAL HIGH AND EASTERN ASIAN SUMMER MONSOON SYSTEM BASED ON FUZZY SYSTEMS AND DYNAMICAL MODEL INVERSION FORECAST

HONG Mei (洪　梅)[1], ZHANG Ren (张　韧)[1], YAN Heng-qian (阎恒乾)[1], GE Jing-jing (葛晶晶)[2], YU Dan-dan (余丹丹)[3]

(1. Research Center of Ocean Environment Numerical Simulation, Institute of Meteorology and oceanography, PLA University of Science and Technology, Nanjing 211101 China; 2. Meteorological and Hydrological Center of Nanjing City, Nanjing 211101 China; 3. Troop No.61741 of PLA, Nanjing 211101 China)

Abstract: Based on the adaptive network fuzzy inference system (ANFIS), methods to filter out the noise of impact factors from the main signal are discussed. Focusing on the abnormal weather conditions in 2010, we use the delay-relevant method to analyze the five members of the summer monsoon system that had the largest effect on the subtropical high anomalies from the observational data. ANFIS is suitable for research and simulation of subtropical highs that are difficult to describe accurately with dynamics, allowing the effect of five factors on the subtropical high anomalies to be examined. Our results show that the Mascarene cold high, the Indian monsoon latent heat flux, and the South China Sea monsoon trough had the largest effect on the subtropical high anomalies. Diagnostic analysis, with genetic algorithms (GA) and dynamical reconstruction theory, reconstructed the nonlinear dynamical model of the subtropical high and its main factors objectively and accurately from the sequence of observations in 2010. Furthermore, a dynamically extended forecast experiment is performed. The forecasts for the subtropical high area index, the Mascarene cold high index, the Indian monsoon latent heat flux, and the South China Sea monsoon trough index all show a strong short-term effect over less than 25 days. The forecasting trend is accurate, and the error rate is no more than 7%. Our results provide new insight and methods for research on the association between the western Pacific subtropical high and the East Asian summer monsoon system, and for the prediction of the western Pacific subtropical high index.

Key words: dynamical model inversion; ANFIS; western Pacific subtropical high; East Asian summer monsoon system

CLC number: P732; P738 **Document code:** A

doi: 10.16555/j.1006-8775.2015.S1.004

1 INTRODUCTION

The western Pacific subtropical high (WPSH) is an important component of the East Asian summer monsoon (EASM) system (Miyasaka and Nakamura[1]) and its intensity and position show complex seasonal evolution. In summer, the WPSH reaches its northernmost position and strongly affects the rainfall over China and Japan[1]. Changes in the position and intensity of the WPSH also affect the rainy season, causing heavy rain, floods, and drought in China. Recently, persistent anomalies in the high have resulted in frequent severe meteorological disasters. For example, the catastrophic Yangtze River floods in August 1998 were induced by the abnormal southward movement of the subtropical high (SH): the continuous heavy rainfall that lasted for more than 30 days in the Huaihe River Basin in 2003 resulted from the abnormal westward motion of a strong SH swinging between northern and southern latitudes of 24°N; both the sustained high temperatures in Chongqing and eastern Sichuan in the summer of 2006 and heavy rains in Huaihe River Basin in July 2007 arose from the continuing westward movement of the seasonal SH; and 14 storms from May to July in 2010 in South China and Jiangnan region were caused by abnormal SH activity. Therefore, understanding the variability in the anomalies of WPSH is essential

Received 2014-08-20; **Revised** 2015-08-14; **Accepted** 2015-09-15

Foundation item: Young Scientists Fund of National Natural Science Fund of China (41005025/D0505); National Natural Science Fund of China (41075045); Natural Science Fund of Jiangsu Province, China (BK2011123)

Biography: HONG Mei, Ph.D., primarily undertaking research on ocean meteorology.

Corresponding author: HONG Mei, e-mail: flowerrainhm@126.com

for estimating the variation in the monsoon climate over East Asia.

The WPSH is a central topic in atmospheric sciences, and many East Asian scientists have studied WPSH extensively, owing to its dominance in East Asian climates. The seasonal variation of WPSH is closely related to the onset and withdrawal of the EASM system. The abrupt northward movement of WPSH from winter to summer is accompanied by sudden changes in the circulation patterns over the East Asian monsoon area (Tao et al.[2]). The distribution of summer rainfall in China is associated with seasonal variations in the intensity, structure, and location of WPSH (Tao and Wei[3]; Wang and Xue[4]). Zhang and Tao[5] reported that the secondary northward shift of the SH ridgeline is closely related to the secondary northward shift of the low-level west wind at the equator and the northward equatorial convection. Xu et al.[6, 7] argued that the development and strengthening of convection over Bengal led to the convection interruption near the South China Sea and West Pacific Ocean, and to the strengthening and westward extension of the SH western ridge. Changes in WPSH would influence the Mei-yu front by affecting the convergence of tropical water vapor with extratropical airflow (Ninomiya and Kobayashi[8]; Zhou and Yu[9]). The WPSH also plays a major role in modulating weather and climate in Korea and Japan (Kurihara[10]).

Understanding the laws governing the behavior of the WPSH has been a focus of much research. Ren et al.[11] showed that the dynamical and thermal mechanism of SH short-term variability was closely related to the abnormal activity of South Asian high pressure and the high-latitude circulation system. Zhang et al.[12, 13] analyzed the effect of thermal factors, such as solar radiation heating, monsoon rainfall, and convective condensation heating of the monsoon trough precipitation, on the shape and stability of the SH. The relationship between the position of SHs and monsoon disturbance in the South China Sea and Indian Ocean was determined (Zhang[14]). The dynamic mechanism of the teleconnection between the East and West Pacific SH was analyzed (Zhang et al.[15]; Grinsted and Moore[16]). However, studies are mainly of diagnostic analysis of the relationship between SHs and the EASM. Because the activity and aberrance of SHs is complex (Cao et al.[17]), the mechanisms leading to the development and maintenance of the WPSH are still disputed.

It is still difficult to determine which factors affect complex weather systems, such as SHs, the nonlinear relationship between the factors, and to what degree different factors affect the SH system. Fuzzy systems provide a simple, effective way to extract summer monsoon factors affecting the SH anomalies from limited observations (Zhang[14]), to refine and summarize qualitative rules and implicit mapping relationship from observations, and incorporate them into a quantitative control system and diagnostic prediction model. The adaptive network fuzzy inference system (ANFIS) has many characteristics, such as fault-tolerance, adaptive learning, and nonlinearity, which make it suitable for investigating phenomena, such as SHs, that are not accurately described by dynamics. Initially, we discuss the effect and contribution of key members of the EASM on the SH anomalies in 2010. Next, the three most significant factors are identified, which are the Mascarene cold high index (MH), the Indian monsoon latent heat flux (FLH), and the South China Sea monsoon trough (ST).

Physical mechanisms of complicated weather system activity and abnormalities can be described by dynamical models of meteorology elements or weather systems, which are obtained by inverting the time series of the observed data. To address the local convergence of errors and their calculation during the inversion, Zhang et al.[18] introduced genetic algorithms (GA) to improve the model. Zhang et al.[19] reported the inversion of the nonlinear dynamical forecast model of the SH index, and achieved good results. Because the SH is a complex system with many factors, using a single modeling factor restricts the model reasonability. Thus, the model factors should be carefully chosen and extracted objectively. The SH is a member of and interacts with the EASM system. To refine the dynamical model of the subtropical high area index (SI) other members of the monsoon system with a stronger correlation with SI can be introduced.

In this work, the SI is defined as a measurement of the scope and form of SHs. In Section 2, data from the last 30 years is reanalyzed and ANFIS is introduced. In Section 3, three summer monsoon factors that have a stronger correlation with SI are chosen with ANFIS. The dynamical model inversion of SI and its impact factors is conducted in Section 4 to overcome the problem of single model elements. A dynamical forecast experiment is performed in Section 5. Finally, we present a summary of our results in Section 6.

2 RESEARCH DATA AND METHODS

2.1 *Research data*

Daily reanalysis data from May to October from the last 30 years (1985–2014) are provided by the USA National Center for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR), and include: (1) horizontal wind field and geopotential height field at 850 and 200 hPa, geopotential height field at 500 hPa, and sea level

pressure field with a resolution of 2.5°×2.5°; (2) the sensible heat and convective precipitation rate in a Gaussian grid; and (3) the observed long wave radiation (OLR) data from NOAA satellites.

2.2 *ANFIS*

ANFIS can be trained and adapted to improve the approximation efficiency and reduce errors, whereas conventional fuzzy systems rely on experience to adjust the membership functions.

The fuzzy inference system is based on composite learning, with the after-transfer gradient descent method and the least squares method to identify the linear and nonlinear parameters, and contains a series of "IF ... THEN ..." rules, shown in Fig.1.

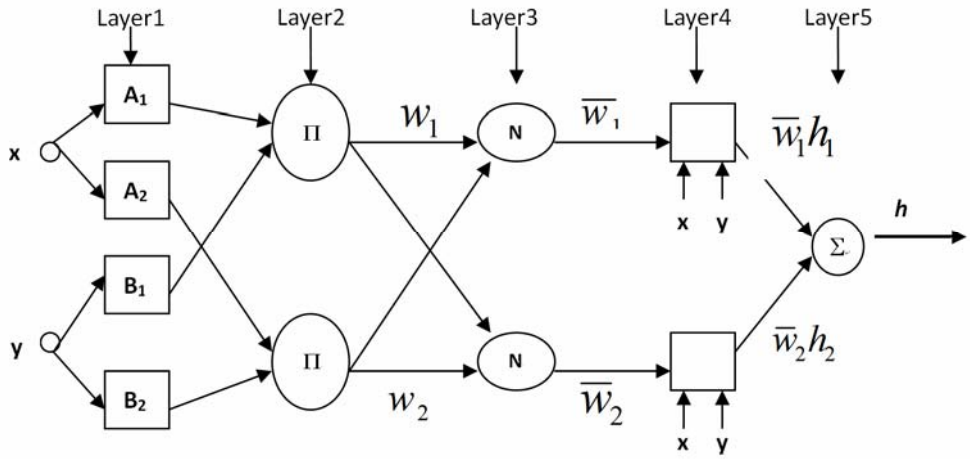

Figure 1. The structure diagram of ANFIS fuzzy system.

The appropriate membership functions are gradually deployed to satisfy the relationship between the fuzzy reasoning input and output. The main rules are as follows.

Rule 1: If x is A_1 and y is B_1, then $h_1 = p_1 x + q_1 y + r_1$.

Rule 2: If x is A_2 and y is B_2, then $h_1 = p_2 x + q_2 y + r_2$.

Here, A_i, B_i are the mapping values of membership functions; x, y are the assumptions and training input of the fuzzy inference system; $D = GP$ are fuzzy inference conclusions, where $i = 1, 2$. The weighted average method is used for non-fuzzy processes, and the fuzzy inference output is $h = \frac{w_1}{w_1 + w_2} h_1 + \frac{w_2}{w_1 + w_2} h_2$, where, w_i is the output weight of node *i*. The Takagi-Sugeno fuzzy inference system is:

$$h = \overline{w}_1 h_1 + \overline{w}_2 h_2 = (\overline{w}_1 x) p_1 + (\overline{w}_1 y) q_1 + (\overline{w}_1) r_1 + (\overline{w}_2 x) p_2 + (\overline{w}_2 y) q_2 + (\overline{w}_2) r_2 \quad (1)$$

Because the premise and reasoning parameters have been decoupled in the composite learning process and ANFIS is an adaptive network, its learning efficiency is higher than a neural network. Further fuzzy approximation and signal de-noising methods are described elsewhere (Zadeh[20]; Takagi and Sugron[21]).

3 DETECTION AND ANALYSIS OF FACTORS AFFECTING SH INTENSITY

3.1 *Subtropical high activity*

The WPSH activity, which is the dominant component of the EASM system, varies seasonally and is most frequent during the boreal summer. WPSH migrates northward in a stepwise manner characterized by two distinct northward jumps. In mid-June, WPSH jumps northward for the first time, and the Mei-yu season in the Yangtze River valley, Japan, and Korea begins. The second northward jump usually occurs in late July or sometimes in August. Then, WPSH shifts to its most northern position, signaling the end of the Mei-yu season in the Yangtze River valley, Japan, and Korea, and the start of the rainy season in north and northeastern China (He et al.[22]).

The abnormal activity of SHRL often results in subtropical circulation anomalies in East Asia and extreme weather events in China such as the SH double ridge in 1998 (Miyasaka and Nakamura[1]; Tao et al.[2]; Nitta[23]; Chen [24]). Thus, the SHRL position

and its abnormal activity often have a large effect on the summer weather in China.

3.2 *Correlation analysis and statistical significance test*

The correlation coefficient, r_{xy}, measures the relationship between any two variables. For the time series of two variables, x, y, with the same sample length, n, the correlation coefficient, r_{xy}, can be calculated with:

$$r_{xy} = \frac{\sum_{t=1}^{n}(x_t - \overline{x})(y_t - \overline{y})}{\sqrt{\sum_{t=1}^{n}(x_t - \overline{x})^2}\sqrt{\sum_{t=1}^{n}(y_t - \overline{y})^2}} \quad (2)$$

The statistical significance of r_{xy}, or simply r, can be evaluated by using the *t*-test. The constructed statistic t is expressed as:

$$t = \sqrt{n-2}\frac{r}{\sqrt{1-r^2}} \quad (3)$$

Statistic t follows the *t*-distribution with $n-2$ degrees of freedom. Given the significance level, α, we check the t-distribution table: if $t > t_\alpha$, the two variables are significantly correlated. For a fixed sample size, the critical value of the correlation coefficient, r_c, can be obtained by the criterion

$$r_c = \sqrt{\frac{t_\alpha^2}{n-2+t_\alpha^2}} \quad (4)$$

If $r > r_c$, the significance of the t test is accepted.

3.3 *Correlation analysis between members of EASM system and SI index*

The SI reveals the relationship between members of the Asian summer monsoon system and SH. The SI is defined by the Central Meteorological Observatory[25] as characterizing the range and intensity of the SH, where the geopotential grid points have an average height greater than 588 gpm at 500 hPa with resolution of 2.5°×2.5°, north of 10°N, within the range 110°E to 180°E. The higher the value of SI, the wider the range and the higher the intensity of the SH.

EASM has many members; there are many factors that affect SH, 21 of which are closely related to the WPSH (Xue et al.[26]; Yu et al.[27]). If all the factors were used for modeling, the equations would be too complex. Therefore, the correlation analysis method is used to determine the relationship between these factors and SI based on the average data from the last 30 years (1985–2014). Definitions of each factor are given in Xue et al.[26] and Yu et al.[27]. The values of these factors can be calculated from NECP/NCAR data, as in Section 2. The best correlation factors are identified for further study, which are as follows.

(1) The MH: Average grid points of sea level pressure in the [40°–60°E, 25°–35°S] region.

(2) The Somali low-level jet (SLLJ): The 850 hPa average longitude wind in the [40°–50°E, 5°–5°N] region.

(3) The South China Sea monsoon trough (ST): The average of OLR grid points located in the [110°–130°E, 7.5°–17.5°N] region.

(4) The Indian monsoon latent heat flux (FLH): The latent heat flux in the [80°–100°E, 12.5°–22.5°N] region.

(5) The monsoon circulation index at Bay of Bengal (J1V): The average grid point J1V = V850–V200 in the [80°–100°E, 0°–20°N] region.

The delay results associated with the SI are shown in Table 1. The correlation coefficient between the five factors and SI exceed 0.85. The MH in the southern hemisphere increases the SH early; this is a positive correlation and indicates a very close relationship, consistent with previous research[26]. The close relationships among the SLLJ, ST, FLH, J1V, and SI are also consistent with previous research (Yu et al.[27]; Wang and Cao[28]).

Table 1. The correlation analysis of five main factors and the subtropical high area index.

No.	five main factors	Correlation analysis (Time)
1	Mascarene cold high strength index (MH)	0.85 (8d)
2	Somali low-level jet (SLLJ)	0.90 (6d)
3	Indian monsoon latent heat flux (FLH)	0.87 (4d)
4	Tibetan high activity index (XZ)	-0.86 (-2d)
5	The monsoon circulation index at Bay of Bengal (J1V)	0.91 (2d)

3.4 *Mapping features of the fuzzy inference system between SI and EASM*

ANFIS uses the Sugeno fuzzy inference system with the 2^N rule to train the input data, where N is the dimension of input data (generally $N<7$). After fuzzy system training, the data is output as a fuzzy interference system matrix. In this paper, ANFIS consists of five inputs and one output. Model training and reasoning simulation use Matlab in the Fuzzy

Toolbox simulation environment. Based on the results of the analysis in section 3.3, the five input data sets are MH 8 days ahead, SLLJ 6 days ahead, ST 3 days ahead, FLH 2 days ahead, and J1V 1 day ahead, and the SI output data set is used for training. The training periods of SI are from May 1 to October 31 in the average 30-year data (total of 184 days).

After 250 training iterations, the specified error magnitude is reached (10^{-2}), to establish a fuzzy inference system and fuzzy mapping relationship between the five impact factors and SI. For simplicity in figures and calculations, MH, SLLJ, ST, FLH, J1V, and SI are taken as anomalies, and are divided by 1000, 10, 100, 10, 10, and 200, respectively. The fuzzy inference system is a multi-dimensional system. For convenience, 10 different three-dimensional cross-sections are analyzed and compared; the four most representative three-dimensional cross-sections are chosen for analysis.

Figure 2 is the mapping relationship between the outputs of MH 8 days ahead and SLLJ 6 days ahead, and the input of SI. Inputs 1 and 2 are SLLJ 6 days ahead and MH 8 days ahead, respectively, whereas the output is the SI lagging behind. When the MH high is 8 days ahead strengthens and shows outbreaks (positive anomaly), regardless of whether SLLJ 6 days ahead 6 days strengthens and shows outbreaks or weakens (positive or negative anomalies), SH strengthens and shows outbreaks (positive anomaly), as shown by points A and B. The MH 8 days ahead is weak (negative anomaly). Regardless of whether the SLLJ in 6 days is strong or weak (positive anomaly or negative anomaly), SH is weak (negative anomaly), as shown at points C and D. The effect of the MH on the SH is more significant than that of the SLLJ.

In contrast to Fig.2, Fig.3 shows that when the MH 8 days ahead strengthens and shows outbreaks (positive anomaly) and the ST 3 days ahead is weak (negative anomaly), the SH intensity changes little and does not increase the outbreak (around zero), as shown at point A. As the ST 3 days ahead increases from the smallest negative anomaly to 0 and then increases to the maximum positive anomaly, the corresponding SH strength increases from 0, suddenly decreases to 0, and then continues to increase until it reaches the maximum SH strength (maximum positive anomaly), as shown by points B and C. The MH 8 days in advance is weak (negative anomaly). If the corresponding the ST 3 days ahead is also weak (negative anomaly), SH is weak (negative anomaly; point D). However, if the ST 3 days ahead increases to outbreak (positive anomaly), which may offset the effect of the MH on SH, the SH strength returns to normal (point E). Therefore, the MH and ST have a significant effect on the SH strength, particularly when they interact.

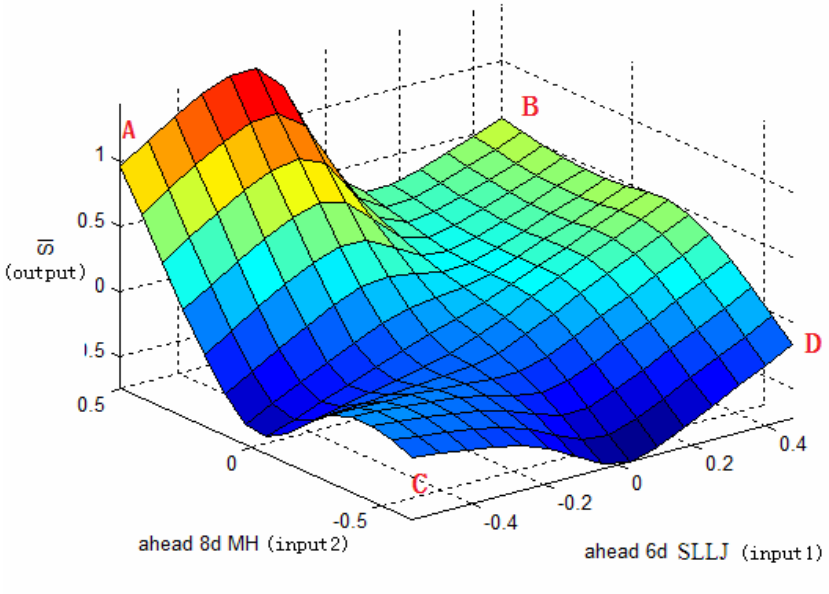

Figure 2. The fuzzy reasoning map of MH 8 days ahead (input2), SLLJ 6 days ahead (input1) and SI (output).

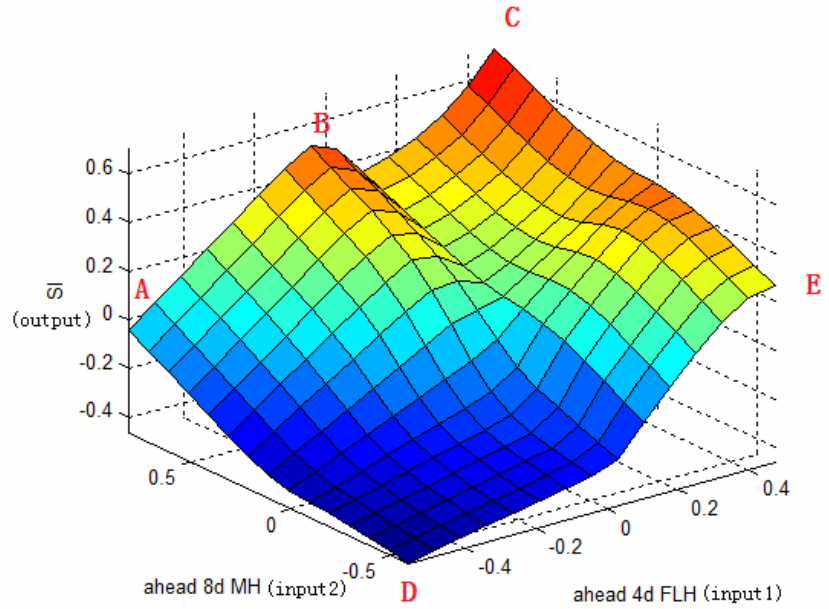

Figure 3. The fuzzy reasoning map of MH 8 days ahead (input2), FLH 4 days ahead (input1) and SI (output).

Similar to the SLLJ in Fig.2, in Fig.4, when the MH 8 days ahead is increased to outbreak (positive anomaly), regardless of whether J1V is 2 days ahead is strengthened or weakened (positive or negative anomaly), the SH will be increased to outbreak (positive anomaly), at points A and B. In contrast to Fig.2, the increase in SH changes little. If the MH 8 days ahead is weak (negative anomaly), regardless of whether the J1V 2 days ahead is strong or weak (positive or negative anomaly), the SH will be weak (negative anomalies), as shown in points C and D. In contrast to Fig.1, when J1V gradually increases, the SH intensity weakens. Because of the effect of the SH strength, the MH is more significant than the J1V.

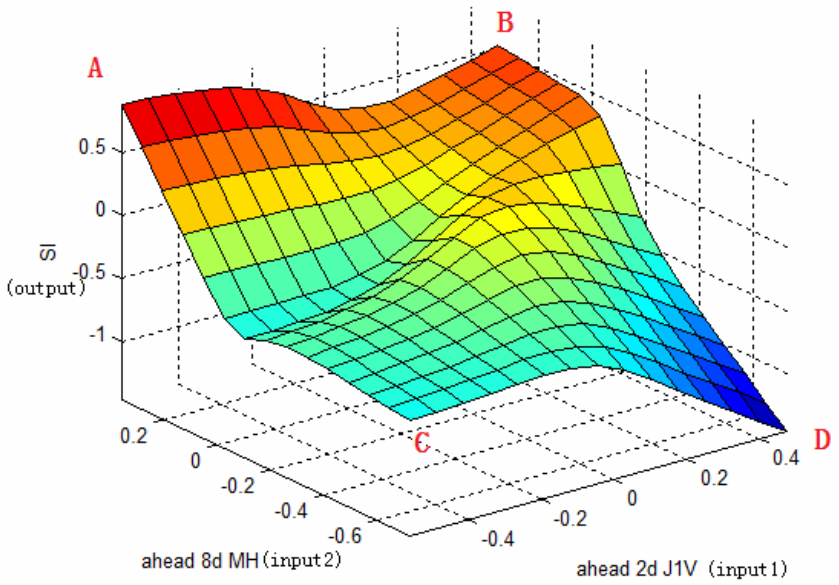

Figure 4. The fuzzy reasoning map of MH 8 days ahead (input2), J1V 2 days ahead (input1) and SI (output).

Figure 5 shows that there is a negative correlation between FLH and SH; therefore, when the MH 8 days ahead strengthens to outbreak (positive anomaly) and the corresponding FLH is controlled by the lowest pressure (maximum negative anomaly), the SH strengthens to outbreak (point A). If FLH is controlled by the highest pressure (positive anomaly), although the SH strength is high, the strengthened outbreak is not obvious (point B). If the MH 8 days ahead is weak, the FLH 2 days ahead increases from the smallest negative anomaly to 0, and then continues to increase to the maximum positive anomaly. The corresponding SH intensity anomalies is reduced from 0 to the minimum, and then continues to increase, as

shown by points C–F. When FLH 2 days ahead achieves the maximum positive value, the SH intensity anomalies may either return to 0, which means it returns to normal; or remain small, which means the SH remains weak. Thus, the effects of the MH and FLH on the SH intensity are significant, although MH shows a positive effect and FLH shows a negative effect. The other six three-dimensional cross-sections are similar to the previous four, and are not described here individually. Comparing all the fuzzy inference results, we can conclude that the effect of the MH, ST, and FLH on SH strength is more significant than the other two factors.

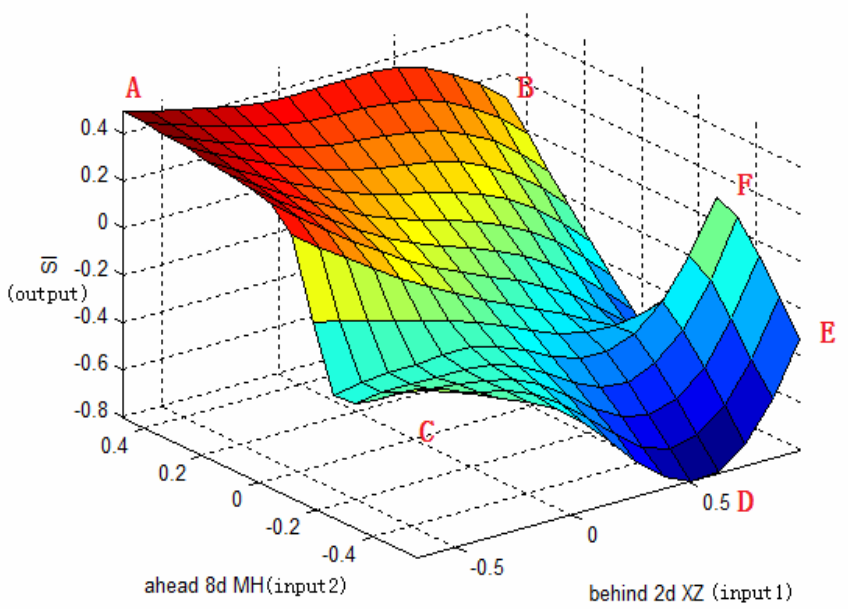

Figure 5. The fuzzy reasoning map of MH 8 days ahead (input2), XZ 2 days behind (input1) and SI.

The reasoning mapping feature largely reflects the main features of the relationship between SH and the five related factors in the summer monsoon system in 2010. Because this fuzzy inference system is entirely based on the time-series data collection of five related factors in the summer monsoon system and SH, it is objective and credible.

4 RECONSTRUCTION OF DYNAMICAL MODEL OF SI AND ITS IMPACT FACTORS

Takens[29] set out the principles for reconstructing dynamical systems from time series of observed data in phase space reconstruction theory. The study suggested that the evolution of any component in the system could be determined by other components that interacted with it, and the information about the components could be obtained from the evolution of the other correlated components. Therefore, it is possible to reconstruct a dynamical model of system evolution from a time series of observed data. We introduce the concept of dynamical system retrieval and model parameter optimization and use four time series for SHRL, MH, ST, and FLH to reconstruct the dynamical model based on our analysis of the SH and its impact factors in abnormal years.

4.1 *Reconstruction of the dynamical model*

Suppose that the physical law of a nonlinear system evolving with time can be expressed as:

$$\frac{dq_i}{dt} = f_i(q_1, q_2, ..., q_i, ..., q_N), i = 1, 2, ..., N \qquad (5)$$

where f_i is the generalized nonlinear function of $q_1, q_2, ..., q_i, ..., q_n$ and N is the number of state variables. N can generally be determined by the complexity of the dynamical system and measured by calculating its fractal dimensions. The difference form of Eq. 5 is:

$$\frac{q_i^{(j+1)\Delta t} - q_i^{(j-1)\Delta t}}{2\Delta t} = f_i(q_1^{j\Delta t}, q_2^{j\Delta t}, ..., q_i^{j\Delta t}, ..., q_N^{j\Delta t})$$
$$j = 2,3,.....M-1 \qquad (6)$$

where M is the length of the time series of the observed data. The model parameters and the system structure can be obtained by an inversion algorithm based on the observed data. $f_i(q_1^{j\Delta t}, q_2^{j\Delta t}, ..., q_i^{j\Delta t}, ..., q_N^{j\Delta t})$ is an unknown nonlinear function and we assume that $f_i(q_1^{j\Delta t}, q_2^{j\Delta t}, ..., q_i^{j\Delta t}, ..., q_N^{j\Delta t})$ contains two parts: G_{jk}, representing the expanding items containing variable q_i, and P_{ik} representing the corresponding parameters that are real numbers ($i = 1, 2, ...N$, $j = 1, 2, ...M$, $k = 1, 2, ..., K$). We assume that:

$$f_i(q_1, q_2, ..., q_n) = \sum_{k=1}^{K} G_{jk} P_{ik} \quad (7)$$

The matrix form of Eq. (7) is

$$D = GP \quad (8)$$

in which:

$$D = \begin{Bmatrix} d_1 \\ d_2 \\ ... \\ d_M \end{Bmatrix} = \begin{Bmatrix} \frac{q_i^{3\Delta t} - q_i^{\Delta t}}{2\Delta t} \\ \frac{q_i^{4\Delta t} - q_i^{2\Delta t}}{2\Delta t} \\ ... \\ \frac{q_i^{M\Delta t} - q_i^{(M-2)\Delta t}}{2\Delta t} \end{Bmatrix},$$

$$G = \begin{Bmatrix} G_{11}, G_{12}, ...G_{1K} \\ G_{21}, G_{22}, ...G_{2,K} \\ ... \\ G_{M1}, G_{M2}, ...G_{M,K} \end{Bmatrix}, \quad P = \begin{Bmatrix} P_{i1} \\ P_{i2} \\ ... \\ P_{iK} \end{Bmatrix}$$

Coefficients of the generalized unknown Eq.(8) can be identified through inverting the observed data. Given vector D, vector P can be solved to satisfy the above Eq.(8). The system is nonlinear with respect to q and linear with respect to P, assuming that P is unknown. Therefore, the classical least squares method can be used to estimate the equation, and the regular equation $G^T GP = G^T D$ can be derived by minimizing the residual sum of squares $S = (D - GP)^T (D - GP)$.

Because $G^T G$ is usually a singular matrix, its eigenvalues and eigenvectors can be solved easily. After removing eigenvalues and eigenvectors with values of 0, the remaining components are K numbers of $\lambda_1, \lambda_2, ..., \lambda_i$ that can make up a diagonal matrix, Λ_k, and the K numbers corresponding to the eigenvectors form the diagnostic matrix, U_L.

$P = HD$ can be solved with $V_L = \frac{GU_i}{\lambda_i}$ and $H = U_L \Lambda^{-1} V_L^T$, and parameter P can be obtained.

Based on this approach, coefficients of the nonlinear dynamical systems can be determined and the nonlinear dynamical equations of the observed data can be established.

4.2 *Dynamical model retrieval of SI and its impact factors by genetic algorithms*

Existing parameter estimation methods, such as the neighborhood search method and the least squares method, are mostly one-way search methods that must travel the entire parameter space, resulting in a low search efficiency. The limitation of the error gradient convergence and its dependence on the initial solution mean that the parameter estimation often locates the local optimum rather than the global optimum (Wang[30]). GA are now widely used for global optimization. GA are excellent for global searching and parallel computing; error convergence rates can be improved greatly by GA, thus GA are useful for optimizing parameters[30]. Therefore, we use GA to reconstruct the dynamical models and optimize the model parameters based on the SI, ST, MH, and FLH time series, T_1, T_2, T_3, T_4, respectively.

With the minimum square error $S = (D - GP)^T (D - GP)$ as the boundaries, the model parameter inversion method follows the principle described in Section 4.1 and seeks populations (solutions) and a parallel mode to search for the optimal parameters in the parameter space.

There are a many second-order linear equations in geophysical fluid dynamics. Equations that contain linear and quadratic terms, such as the Navier-Stokes equation, are used to model the main characteristics of the atmosphere and ocean. The following nonlinear second-order ordinary differential equations are used for the retrieval and reconstruction of the dynamical model of SH and its impact factors. In the retrieval model, the data is selected as follows from the average of 30 years of data: SI for May 1 to July 31; MH 8 days ahead of SH, from April 23 to July 23; ST 3 days in advance of SH, from April 28 to July 28; and FLH 2 days ahead of SH, from April 29 to July 29. The total length for the four time series is 92 days. They are used as the expectation data to optimize and retrieve model parameters.

$$\begin{aligned}
\frac{dT_1}{dt} &= a_1T_1 + a_2T_2 + a_3T_3 + a_4T_4 + a_5T_1^2 + a_6T_2^2 + a_7T_3^2 + a_8T_4^2 + a_9T_1T_2 + a_{10}T_1T_3 + a_{11}T_1T_4 + a_{12}T_2T_3 + a_{13}T_2T_4 + a_{14}T_3T_4 \\
\frac{dT_2}{dt} &= b_1T_1 + b_2T_2 + b_3T_3 + b_4T_4 + b_5T_1^2 + b_6T_2^2 + b_7T_3^2 + b_8T_4^2 + b_9T_1T_2 + b_{10}T_1T_3 + b_{11}T_1T_4 + b_{12}T_2T_3 + b_{13}T_2T_4 + b_{14}T_3T_4 \\
\frac{dT_3}{dt} &= c_1T_1 + c_2T_2 + c_3T_3 + c_4T_4 + c_5T_1^2 + c_6T_2^2 + c_7T_3^2 + c_8T_4^2 + c_9T_1T_2 + c_{10}T_1T_3 + c_{11}T_1T_4 + c_{12}T_2T_3 + c_{13}T_2T_4 + c_{14}T_3T_4 \\
\frac{dT_4}{dt} &= d_1T_1 + d_2T_2 + d_3T_3 + d_4T_4 + d_5T_1^2 + d_6T_2^2 + d_7T_3^2 + d_8T_4^2 + d_9T_1T_2 + d_{10}T_1T_3 + d_{11}T_1T_4 + d_{12}T_2T_3 + d_{13}T_2T_4 + d_{14}T_3T_4
\end{aligned} \quad (9)$$

Suppose that parameter matrix $P=[a_1,a_2,.....a_9;b_1,b_2,.....b_9;c_1,c_2,.....c_9]$ of the above Eq.(9) is the population. The objective function is the minimal residual sum of squares $S=(D-GP)^T(D-GP)$. The individual fitness value is $l_i=\frac{1}{S}$ and the total fitness value is $L=\sum_{i=1}^{n} l_i$. The idiographic steps include coding and creating the initial population; calculation of fitness; choosing the male individual; and crossover and variation. The calculation theory and detailed explanation have been described elsewhere[30]. In the calculation, the step length is one month. After the fourth cycle of the GA and optimization searching, the method rapidly converges on the target adaptive value and we can retrieve each optimized parameter of the dynamical equations.

After eliminating terms with a small dimension coefficient, we retrieve the nonlinear dynamical model of SI and its impact factors as:

$$\begin{cases} \dfrac{dT_1}{dt} = -18.4577T_1 + 0.0918T_2 - 8.2256\times10^{-5}T_2^2 + 1.5431\times10^{-2}T_1T_2 + 1.4522\times10^{-3}T_2T_3 \\ \dfrac{dT_2}{dt} = 235.1221T_1 - 2.5112T_2 - 11.4223T_3 + 12.3321T_4 + 4.0568\times10^{-3}T_2^2 - 2.2533T_1T_2 \\ +1.3355\times10^{-2}T_2T_3 \\ \dfrac{dT_3}{dt} = -12.2247T_1 + 3.1322T_3 - 4.6555T_4 + 4.2258\times10^{-2}T_1T_2 - 1.2554\times10^{-3}T_2T_3 + 3.041\times10^{-4}T_2T_4 \\ \dfrac{dT_4}{dt} = 4.1334T_2 + 5.8651T_4 + 4.0651T_1^2 + 3.1121T_3^2 + 2.0493T_1T_4 + 2.3368\times10^{-3}T_3T_4 \end{cases} \quad (10)$$

In a fitting test, the effects of the SI and FLH time series are strong, with correlation coefficients of 0.9023 and 0.8766, respectively. In contrast, those of the MH and ST time series are slightly weaker; although the correlation coefficients are 0.8321 and 0.8261, respectively, they are still greater than 0.8.

5 TESTING THE NONLINEAR MODEL PREDICTION OF SH AND ITS RELATED FACTORS

5.1 *Prediction test for 2010*

To test the effect of model prediction, we performed prediction tests for 2010, which was the year with the greatest WPSH anomalies (arrow, Fig.6). From May to October in 2010, SI is above the mean and it is at a 30-year peak. The anomaly in the WPSH strength resulted in unusual weather in China in 2010. The time series of SI, MH, FLH, and ST (August 1 to September 5 2010), which do not participate in the inversion of model, are chosen to verify the effect of model forecasts. SI on August 1 in 2010, MH (8 days ahead) on July 24 in 2010, FLH (2 days ahead) on July 28 in 2010, and ST (3 days ahead) on July 27 in 2010 are taken as the initial values to be substituted into the nonlinear dynamic model equations. Numerical integration is performed to obtain a total of 35 days of predicted SI results from August 1 to September 2010 (Fig.6a). The predicted results for the other three factors are shown in Fig. 6b–6d. The 15, 25, and 35-day forecast effects are shown.

Figure 6 (a) shows that the forecast effect of SI is still good. In the first 15 days, the forecast trend is accurate, with a correlation coefficient of 0.9682; the relative error between the forecast value and the true value is only 3.41%. For days 15–25, the forecast trend is similar to the first 15 days, and is accurate, with a correlation coefficient of 0.9088 and a small error of 4.99%. After nearly 25 days, the error begins to increase. For days 25–35, the forecast trend is still accurate; with the peaks and valleys forecasted out, the correlation coefficient is 0.8652, and forecast divergence is weak. The valley and peak numbers are forecasted to be slightly larger at day 29 and day 32, and the remaining values are forecasted more accurately. The error within 35 days is 9.35%. Fig.6b–6d shows that MH, FLH, and ST are similar to SI. The forecast trend within 25 days is good, with a correlation coefficient above 0.85; the errors between the forecast value and the true value are less than 7%. However, after 25 days, the divergence of these three factors increases more than SI, which is related to the divergence of the numerical integration later in the simulation; the errors also increase, reaching 10%–15%. In particular, the divergence of MH after 25 days in Fig.6b is larger, which may be caused by the higher value of SI compared with the other three indexes. The divergence of FLH after 25 days is small compared with the other two factors, which is consistent with the better fitting results of FLH.

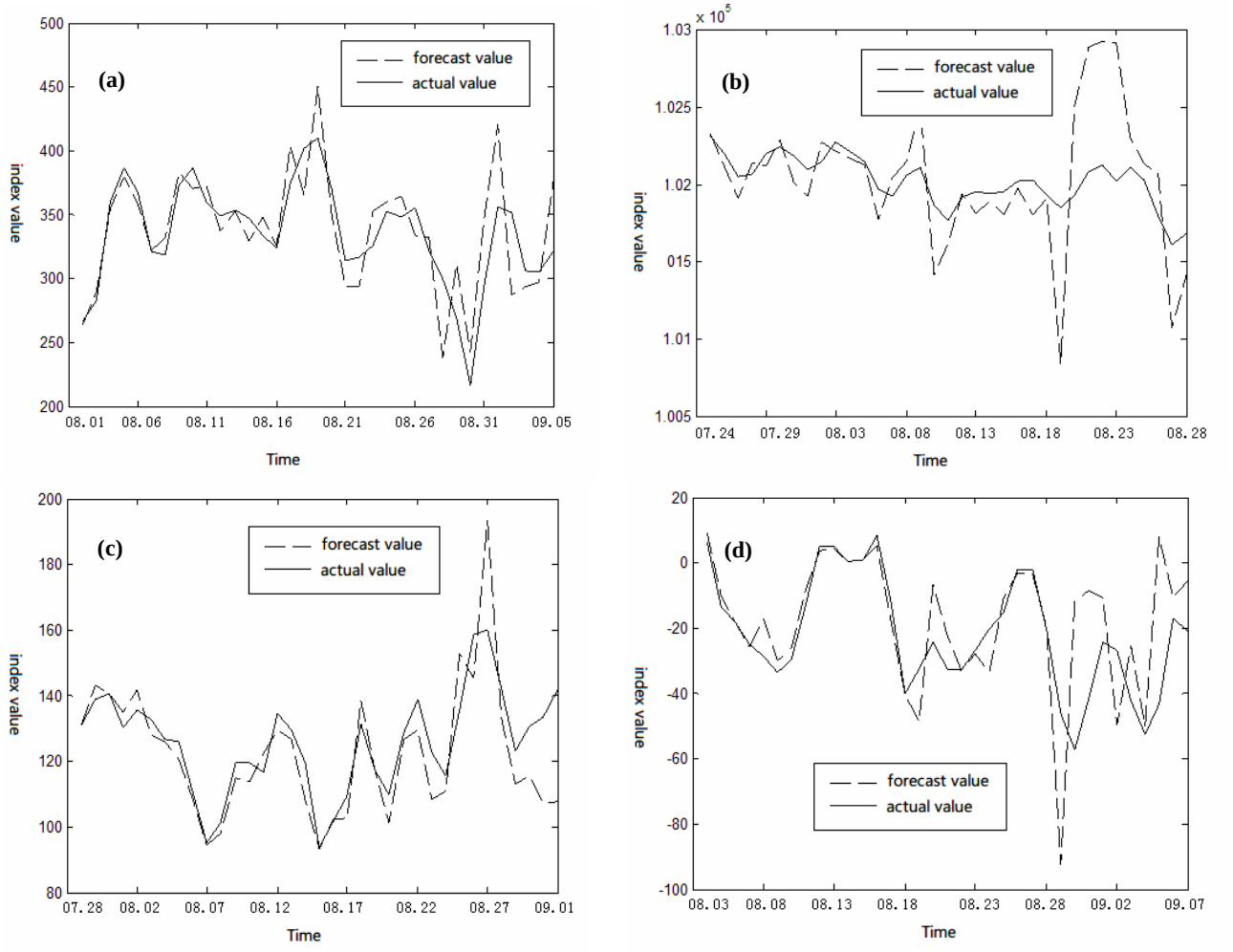

Figure 6. The 35 days forecast map of four indexes.

Figure 6 shows that although the long-term forecasts of SI, MH, FLH, and ST diverge, the short-term forecasting effects for fewer than 25 days are good. The errors for the indexes are less than 7%, which also demonstrates that the forecasts are good.

5.2 *Statistical analysis of the forecasting experiments*

To test the forecast performance of the inversion model, more experiments were performed. We chose another 4 years in which the SH intensity was abnormally strong (bigger SI) and 5 years in which the SH intensity was abnormally weak (smaller SI) to carry out integral forecast experiments for SI. We compared the forecast results for different time periods (short term, 1–15 days; medium term, 16–25 days; long term 26–35 days) with the observed data. The statistical test results are displayed in Table 2. The forecast results for the short and the medium term are good, whereas the results for the long term (>26 days) are acceptable. The results for MH, FLH, and ST are similar to those of SI.

Table 2. Correlation coefficients and root mean square errors between forecast value and real value of different events.

Forecast events	Statistical tests					
	Short term (1~15days)		medium term (16~25 days)		long term (26~35 days)	
	Correlation coefficient	Root mean square error	Correlation coefficient	Root mean square error	Correlation coefficient	Root mean square error
SI bigger event1(1998. 06.21 as initial values to forecast)	0.957	2.92%	0.812	4.51%	0.723	8.91%
SI bigger event2(2006. 07.18 as initial values to forecast)	0.936	2.88%	0.877	4.72%	0.776	7.98%
SI bigger event3(2003. 07.08 as initial values to forecast)	0.942	3.16%	0.881	3.97%	0.718	8.48%
SI bigger event4(1983. 08.05 as initial values to forecast)	0.958	3.40%	0.820	4.06%	0.729	7.90%

SI smaller event1(1984. 07.28 as initial values to forecast)	0.951	3.12%	0.815	3.43%	0.789	8.21%

6 CONCLUSIONS

Detection and analysis based on fuzzy system is different from the frequency structure-based filtering method. It uses fuzzy reasoning to identify the impact factors responsible for a specific contribution, and then filters out the interference. Therefore, this method can be used to analyze and detect the effects of different impact factors on abnormal changes in the atmosphere or ocean systems. We used the ANFIS method to determine the contribution and effect of the key members of EASM on the abnormal SH in 2010 and identify the three most significant factors: the MH, FLH, and ST.

As the evolution of EASM circulation and SH activity is complex, it is difficult to construct dynamic models accurately, so we used GA to invert a dynamical model of SI and three significant factors from 30 years of observed data, and then performed dynamical extended forecast experiments. Our results show that SI has the best forecast effect. The forecast trend is accurate and the error within 35 days is less than 15%. Although MH, FLH, and ST diverged in long-term forecasts (more than 25 days), the short-term forecasts were better (less than 25 days). Moreover, the basic error was less than 10% and showed a good trend in forecasting indexes. However, the trough and peak forecasts were not sufficient; we intend to address this in future work. Based on statistical analysis of the forecasting experiments, the operability and forecast period of the inversion model were significantly better than conventional statistical forecasting methods, such as neural networks (Liu et al.[31]; Zou et al.[32]). Our method provides a new model for the investigation and prediction of complex weather and climate systems, particularly where exact dynamical models cannot be obtained.

In addition, in our model forecast experiments, only the initial value of the dynamical equations is required, unlike neural networks and statistical regression forecasts that require many predictors. Furthermore, our forecasting model also provides forecasts for multiple periods, in contrast to statistical methods that require multiple forecasting models. Therefore, our method has many advantages over both statistical forecasting models and numerical prediction methods.

REFERENCES:

[1] MIYASAKAT T, NAKAMURA H. Structure and formation mechanisms of the Northern Hemisphere summertime subtropical highs [J]. J Climate, 2005, 18:5 046-5 065.

[2] TAO Shi-yan, NI Yun-qi, ZHAO Si-xiong. The study on formation mechanism and forecast of the 1998 summer rainstorm in China [M]. Beijing: China Meteorological Press, 2001 (in Chinese).

[3] TAO Shi-yan, WEI Jing. The westward and northward advance of the subtropical high over the west Pacific in summer [J]. J Appl Meteorol Sci, 2006, 17: 513-524 (in Chinese).

[4] WANG Hui-jun, XUE Fei. Interannual variability of somali jet and its influences on the inter-hemispheric water vapor transport and on the east Asian summer rainfall [J]. Chin J Geophys, 2003, 1: 230-239 (in Chinese).

[5] ZHANG Qin-yun, TAO Shi-yan. The northward jump and anomaly of The West Pacific subtropical high in summer [J]. Acta Meteorol Sinica, 1999, 57: 539-548 (in Chinese).

[6] XU Hai-ming, HE Jin-hai, ZHOU Bin. The possible mechanism of the Western Pacific subtropical high northward jump and western extension and the evolution characteristics of atmospheric circulation before the Meiyu onset [J]. J Appl Meteorol, 2001, 12(2): 150-158 (in Chinese).

[7]XU Xiao-lin, XU Hai-ming, SI Dong. Relationship between June sustained torrential rain in South China and anomalous convection over the bay of Bengal [J]. J Nanjing Inst Meteorol, 2007, 30: 463-471 (in Chinese).

[8] NINOMIYA K, KOBAYASHI C. Precipitation and moisture balance of the Asian summer monsoon in 1991. Part II: Moisture transport and moisture balance [J]. J Meteorol Soc Japan, 1999, 77: 77-99.

[9] ZHOU Tian-jun, YU Ren. Atmospheric water vapor transport associated with typical anomalous summer rainfall patterns in China [J]. J Geophys Res, 2005, 110, D08104, doi:10.1029/2004JD005413.

[10] KURIHARA K. A climatological study on the relationship between the Japanese summer weather and the subtropical high in the western northern Pacific [J]. Geophys Mag, 1989, 43: 45-104.

[11] REN Rong-cai, WU Guo-xiong. On the Short-term structure and formation of the subtropical anticyclone in the summer of 1998 [J]. Acta Meteorol Sinica, 2003, 61: 180-195.

[12] ZHANG Ren, SHI Han-sheng, YU Shi-hua. Study of the nonlinear stability problem of subtropical high in Western Pacific[J]. Chin J Atmos Sci, 1995, 19: 687-700 (in Chinese).

[13] ZHANG Ren, YU Zhi-hao. Numerical dynamical analyses of heat source forcing and restricting subtropical high activity [J]. Adv Atmos Sci, 2000, 17: 61-71.

[14] ZHANG Ren. The wavelet packet energy diagnosis of the South Asia Summer Monsoon and West Pacific subtropical high[J]. J Trop Meteorol, 2004, 20: 113-121 (in Chinese).

[15] ZHANG Ren, WANG Ji-guang, YU Zhi-hao. Teleconnection of Rossby inertial gravity waves and East - West Pacific subtropical high [J]. Appl Math Mechan, 2002, 23: 707-714.

[16] GRINSTED J C, MOORE S J. Application of the cross wavelet transform and wavelet coherence to geophysical time series [J]. Nonlinear Proc Geophys, 2004, 11: 561-566.

[17] CAO Jie, HUANG Rong-hui, XIE Ying-qi, TAO Yun. The study of the physical mechanism of West Pacific subtropical high evolution [J]. Sci China (Ser D), 2002, 08: 659-666 (in Chinese).

[18] ZHANG Ren, HONG Mei, SUN Zhao-bo, et al Non-linear

Dynamic Model Retrieval of Subtropical High Based on Empirical Orthogonal Function And Genetic Algorithm [J]. Appl Math Mechan, 2006, 27: 1 645-1 654.
[19] ZHANG Ren, HONG Mei, WANG Hui-zan, et al. Retrieval the non-linear dynamic forecast model of El Nino/La Nina index based on the genetic algorithm optimization [J]. Chin J Geophys, 2008, 51: 633-645 (in Chinese).
[20] ZADEH L A. Fuzzy Sets [J]. Inform Contr, 1965, 8: 338-353.
[21] TAKAGI T, SUGRON M. Fuzzy identification and its application to modeling and Control [J]. IEEE SMC, 1985, 15(1):116-132.
[22] HE Jin-hai, JU Jian-hua, WEN Zhi-ping, LV jun-mei, JIN Fang-hua, A review of recent advances in research on Asian monsoon in China [J]. Adv Atmos Sci, 2007, 24: 972-992.
[23] NITTA T. Convection activities in the tropical western Pacific and their impact on the Northern Hemisphere circulation [J]. J Meteorol Soc Jpn, 1987, 67: 375-383.
[24] CHEN Lie-ting. The Role of the Anomalous Snow Cover over the Qinghai-Xizang Plateau and ENSO in the Great Floods of 1998 in the Changjiang River Valley [J]. Chin J Atmos Sci, 2001, 25: 184-192 (in Chinese).
[25] THE CENTRAL METEOROLOGICAL STATION LONG-TERM FORECASTING GROUP. The technology experience of the Long term weather forecast (Appendix) [M] //Beijing: Central Meteorological Observatory, 1976 (in Chinese).
[26] XUE Feng, WANG Hui-jun, HE Jin-hai. The influence of Mask Lin high and Australia high interannual variation to the East Asian summer monsoon precipitation [J]. Chin Sci Bull, 2003, 3: 19-27 (in Chinese).
[27] YU Dan-dan, ZHANG Ren, HONG Mei. A characteristic correlation analysis between the Asia summer monsoon memberships and west Pacific subtropical high [J]. J Trop Meteorol, 2007, 13(1): 101-104.
[28] WANG Xiao-ping, CAO Li-ming. Genetic algorithm theory and its software implementation [M]// Xi'an: Xi'an Jiao Tong University Press, 2003 (in Chinese).
[29] TAKENS F. Detecting strange attractors in fluid turbulence [J]. Lecture Notes in Math, 1981, 898: 361-381.
[30] WANG Lin. Intelligent Optimization Algorithms and Its Application [M]. Beijing: Tsinghua University Press, 2001 (in Chinese).
[31] LIU Ken-feng, ZHANG Ren, HONG Mei, et al. BP neural network forecast model of subtropical high index based on hierarchy genetic algorithm optimization [J]. J Central South Univ (Nat Sci Edi), 2007, 38 (Suppl): 886-890 (in Chinese).
[32] ZOU Li-wei, ZHOU Tian-jun, WU Bo, et al. The interannual variability of summertime western Pacific subtropical high hindcasted by GAMIL CliPAS experiments [J]. Chin J Atmos Sci, 2009, 33(5): 959-970 (in Chinese).

Citation: HONG Mei, ZHANG Ren, YAN Heng-qian, et al. Correlation analysis of the western Pacific subtropical high and eastern Asian summer monsoon system based on fuzzy systems and dynamical model inversion forecast [J]. J Trop Meteorol, 2015, 21(S1): 34-45.