

A Nonlinear Artificial Intelligence Ensemble Prediction Model for Typhoon Intensity

LONG JIN

Guangxi Research Institute of Meteorological Disasters Mitigation, Nanning, China

CAI YAO

Department of Atmospheric Science, School of Physics, Peking University, Beijing, China

XIAO-YAN HUANG

Guangxi Research Institute of Meteorological Disasters Mitigation, Nanning, China

(Manuscript received 30 May 2007, in final form 26 March 2008)

ABSTRACT

A new nonlinear artificial intelligence ensemble prediction (NAIEP) model has been developed for predicting typhoon intensity based on multiple neural networks with the same expected output and using an evolutionary genetic algorithm (GA). The model is validated with short-range forecasts of typhoon intensity in the South China Sea (SCS); results show that the NAIEP model is clearly better than the climatology and persistence (CLIPER) model for 24-h forecasts of typhoon intensity. Using identical predictors and sample cases, predictions of the genetic neural network (GNN) ensemble prediction (GNNEP) model are compared with the single-GNN prediction model, and it has been proven theoretically that the former is more accurate. Computation and analysis of the generalization capacity of GNNEP also demonstrate that the prediction of the ensemble model integrates predictions of its optimized ensemble members, so the generalization capacity of the ensemble prediction model is also enhanced. This model better addresses the “overfitting” problem that generally exists in the traditional neural network approach to practical weather prediction.

1. Introduction

Ensemble numerical prediction (ENP; see the appendix for a list of the key acronyms used in this paper) is a new technique that has been developed within the last decade (Stensrud et al. 2000; Du 2002; Scherrer et al. 2004). An ENP model, whether created with different physical process parameterization schemes or with different initial conditions from a Monte Carlo approach, formally consists of many different ensemble members; the effectiveness of ENP has been widely recognized (Nohara and Tanaka 2004; Zhou and Johnny 2006). At present, traditional mathematic modeling methods, such as multivariate analysis and time series analysis are widely used in statistical prediction and dynamical-

statistical prediction (Zhou and Huang 1997; Ding et al. 2002), in which the future state of a prediction object is forecasted using a statistical prediction equation (Zhou and Huang 1997; Ding et al. 2002).

With the development of the artificial intelligence technique, artificial neural networks (ANNs) have been applied successfully in many disciplines (Li et al. 2003; Has et al. 2004; Liu 2005). In the atmospheric sciences, many applications have been found in research areas such as short-range climate prediction, interpretation and application of numerical prediction products, air pollution prediction, and precipitation data processing of radar-satellite cloud pictures (Jin 2004; Ali 2004; Tapiador et al. 2004). However, despite its excellent performance in self-adaptive learning and nonlinear mapping, in-depth studies have shown that ANNs lack the guidance of a rigorous theoretical system in determining adequate network structure; the effect of the application mainly depends on personal experience. In particular, “overfitting” of the ANN method frequently

Corresponding author address: Long Jin, Guangxi Research Institute of Meteorological Disasters Mitigation, Nanning 530022, China.

E-mail: jinlong01@163.com

occurs in meteorological prediction modeling due to subjective determination of hidden nodes of the network, impeding its wide application (Jin 2005a,b).

A genetic algorithm (GA) is a global optimum algorithm based on natural selection and natural inheritance that has been widely used in the field of artificial intelligence techniques in recent years (Zheng et al. 2003; Guo et al. 2004). With GA, as in biological evolution, new quality populations are continuously generated by the genetic evolutionary operations of selection, crossover, and mutation (information exchange) among individuals of a genetic population. Therefore, GA is a population-searching algorithm independent of gradient information and is very effective at solving complex and nonlinear problems (Luo et al. 2004; Zhou et al. 2004). In 2005, Jin found that by optimizing the network structure and the connection weight of ANNs, genetic evolution is able to create a number of different neural network individuals (Jin et al. 2006). In this work, using key principles from ensemble prediction in numerical weather prediction (NWP), we try to construct a number of quality individual neural networks to build a new NAIEP model for South China Sea (SCS) typhoon intensity prediction.

2. Nonlinear meteorological ensemble prediction

In addition to NWP, statistical and dynamic-statistical prediction methods are still the predominant objective prediction modeling methods used in the atmospheric sciences. In traditional regression analysis, the prediction equation predicts the future state of a specific prediction object (predictand). In statistical ensemble prediction, however, several different prediction equations are set up for the same predictand using different methods, and the final deterministic prediction is obtained by integrating the results of these different prediction equations with equal or different weights (Liu et al. 2003). Until now, among the statistical weather prediction methods there has been no one analog to ensemble prediction in NWP. Ensemble prediction of NWP is motivated by the fact that NWP forecasts are sensitive both to small uncertainties in the initial conditions and model errors, so it is hard to further improve the accuracy of single model deterministic predictions. Ensemble prediction is therefore suggested in which a series of forecasts made by different initial value perturbations or different physical parameterization schemes are integrated to yield a resultant ensemble forecast. A number of ensemble prediction studies suggest that ensemble forecasts are more accurate than deterministic forecasts of a single model. This is because ensemble prediction focuses on nonlinear

weather processes, which are governed by the complicated comprehensive effects of interior dynamics and exterior environmental conditions. Present numerical prediction models still have difficulties in accurately describing the physical processes of the genesis and development of weather systems. This design of ensemble prediction has obviously improved the effectiveness of NWP, so ensemble prediction is widely used in NWP. Statistical weather prediction and NWP attempt to predict weather change, which is characteristic of obvious nonlinearity, transientness, and abrupt change. However, traditional regression analysis models for typhoon intensity prediction are constructed based on linear correlations between the predictand and predictors. This modeling basis for (dynamical) statistical weather prediction contains similar uncertainties to those found in the NWP process, so it is clear that commonly used statistical prediction, model output statistics, and the perfect prog contain similar uncertainties as do those in the single-model deterministic prediction of NWP. Here, we try to develop a prediction modeling theory that differs from traditional regression analysis and to design and construct a new nonlinear statistical ensemble prediction model to improve the traditional single statistical prediction equation model.

3. Principle behind and method for creating ensemble prediction individuals

To construct an NAIEP model, a number of individual neural networks are first created and then integrated to build an ensemble prediction model.

A GA is used to construct the members of the ensemble, and a three-layer back-propagation (BP) network is used as the basic model for the neural networks [for the detailed algorithm, see Jin et al. (2003)]; the major computational steps are summarized below:

- 1) Randomly generate the connection weights and thresholds from input layer to hidden layer and from hidden layer to output layer, and set the global convergence error, ε , of the model.
- 2) Perform supervised learning training of the network with learning matrix samples, calculate the error between the real input and expected output of the network, and tune the connection weight coefficients from input layer to hidden layer and from hidden layer to output layer using the learning algorithm of the error-inverse propagation of the BP network.
- 3) If the calculated output error of the model is greater than ε , return to step 2; otherwise, end the training and compute the prediction value using the connection weights, thresholds of the network, and predictors of the prediction samples.

The three-layer BP network model is the most widely used neural network model in various disciplines; evolutionary GA is used in this paper to create a limited number of individual analog BP neural networks. The GA was first presented by Professor J. Holland (Chen et al. 1996). Its fundamental principle is to emulate the adaptation seen in natural biological evolution processes for building an intelligent global optimization searching algorithm. It is simpler and better in the goodness of fit than are traditional optimization algorithms. To compute optimization, GA starts researching the population using fitness information, and does not tend to choose local optima. Even if the fitness function is discontinuous or noise impacted, it can probably find the global optimum solution.

In recent years, the GA has been applied to function optimization, machine learning, highway traffic, and industries (Su et al. 2005; Wang and Wang 2006; Wang et al. 2006). It provides an effective approach for optimizing the practicability, structure, learning rules, and connection weights of neural networks, and for concurrently optimizing the structure and connection weights. In the atmospheric sciences, the GA approach is used to estimate near-surface specific humidity from satellite-based observations (Singh et al. 2005), as well as nonlinear dynamic model retrieval of subtropical high characteristic indices (Zhang et al. 2006). However, our attempt to construct a meteorological NAIEP model based on multiple neural networks using an evolutionary GA is new.

Ensemble prediction modeling of genetic neural network (GNN) involves two major steps: creating the members of the neural network ensemble using the basic GA (Chen et al. 1996), and integrating the outputs of the members. The basic GA consists of selection (breeding), crossover (recombination), and mutation (abrupt change) operations. It performs these genetic operations through parameter coding and the continuous evolution of networks in order to determine the most probable global optimum solution. The following computational aspects must be considered when creating the members of the neural network ensemble.

a. Encoding

The nodes, connection weights, and thresholds for each layer of the three-layer BP network model are arranged in order in a code string, forming a chromosome (a genetic individual) using a mixed encoding of binary and real numbers. This code has two parts: the control code, a binary code of network structure mainly controlling the number of hidden nodes, and the real number code, representing the connection weights and thresholds. Each genetic individual is a potentially op-

timized individual and randomly generates an initial genetic population in the encoding space.

b. Computation of fitness

To compute fitness, we must decode the m genetic individuals of a genetic population into their hidden nodes and connection weights, input training samples, and compute the output of the hidden layer,

$$b_i = f\left(\sum_{h=1}^n a_{hi}v_{hi} + \theta_i\right), \quad i = 1, 2, \dots, p, \quad (1)$$

and the output of the network,

$$\bar{y}_i = f\left(\sum_{j=1}^p w_{ij}b_j + \gamma_j\right), \quad i = 1, 2, \dots, p, \quad (2)$$

where p is the total number of hidden nodes; v_{hi} and w_{ij} are the matrices of the connection weight coefficients from the input layer to the hidden layer and from the hidden layer to the output layer; θ_i and γ_j are the corresponding thresholds, respectively; and the transition function $f(x) = 1/(1 + e^{-x})$. We then calculate the global error of the network:

$$E = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}, \quad (3)$$

where n is the number of training samples. The fitness function, defined as

$$F(x) = \frac{1}{E} = \frac{1}{\sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}}, \quad (4)$$

can be used to calculate the fitness of each genetic individual.

c. Computation of evolutionary operations

The three genetic operations (selection, crossover, and mutation) are applied in the evolutionary operation to genetic populations depending on the fitness of the genetic individuals; roulette wheel selection is used in the selection operation. The fitness value $F_i(x)$ of each individual, as well as the sum of the fitness values of the genetic population, are first calculated using expression (4); the probability of each individual being selected is then computed using the following expression:

$$P_{is} = F_i(x) / \sum_{i=1}^m F_i(x). \quad (5)$$

Those genetic individuals with a higher fitness (i.e., higher quality) have a better chance of producing offspring. Multipoint crossover is used in the crossover operation; except for selected genetic individuals in the population, gene exchange is performed between the code strings of the remaining genetic individuals at multiple crossover points randomly determined using the probability of intersection P_c , producing new genetic individuals. In the mutation operation, allelic gene replacement between genetic individuals happens with probability P_m , forming new genetic individuals. If a neuron of a genetic individual is deleted by mutation, the corresponding weight coefficient code is set to zero, and if a neuron is added, its initial weight coefficient code is generated randomly.

A new generation of the genetic population is produced after each run of the three genetic operators, and the evolutionary operation repeats for a prescribed number of times (N). Afterward, the connection weights and hidden nodes of the m neural networks are determined by decoding each genetic individual; thus, the m ensemble members for ensemble prediction modeling are also obtained. In this paper, each ensemble member is assigned an equal weight in ensemble prediction modeling, so the sum of the prediction value of each neural network yields the ensemble prediction value of the GNN ensemble prediction (GNNEP) model.

4. Ensemble prediction experiments for typhoon intensity in July

We produced 24-h predictions of SCS typhoon intensity as an experiment of the nonlinear meteorological ensemble prediction modeling in this paper. For more than the past 10 years, studies on statistical and statistical-dynamic forecast models for typhoon tracks and intensity have been undertaken in China and overseas (DeMaria and Kaplan 1991; Lu et al. 1996); however, typhoon intensity predictions have not improved significantly over the climatology and persistence (CLIPER) prediction method (Meng et al. 2002). Knaff et al. (2003, 2005) and DeMaria et al. (2005) studied objective prediction methods for hurricane intensity in the Atlantic and the northeast and northwest Pacific basins, but objective prediction tools for the intensity of SCS typhoons are rare. This paper takes the intensity of SCS typhoons as the prediction object, and uses the CLIPER prediction of typhoon intensity as a base level of the forecasting skill to assess our new prediction method. CLIPER was chosen because it is objective (its prediction model does not contain tunable parameters), which makes it suitable for comparing prediction meth-

ods. Our results show that our new ensemble prediction method provides much more accurate predictions than does CLIPER; it is, therefore, a new tool as well as a successful prediction modeling method for objectively predicting typhoon intensity.

We used the GNNEP model (described in section 3) method for creating individuals in our ensemble prediction experiment. For objective comparison, the CLIPER prediction equation for typhoon intensity was first developed, and the same predictors that were selected in the CLIPER model were also used in the GNNEP prediction modeling. Furthermore, the data for the two models were similarly manipulated in the following independent sample experiments.

a. Data

Forty-six years of SCS typhoon data were taken from the "Typhoon Almanac" published by the China Meteorological Administration from 1960 through 2005. SCS typhoons in this paper refer to typhoons that formed in or moved into the sea area 10° – 23.5° N west of 123° E and lasted at least 48 h (required for CLIPER). The typhoon track was sampled every 12 h, starting from the first instant the typhoon moved into the sea area or from when a cyclone developed into a typhoon in the area. Sample typhoons from 1960 to 1989 were used in the prediction modeling, and those from 1990 to 2005 were used as independent samples in prediction testing. The data from 1960 to 1989 contain 56 relevant typhoons in July, with 330 corresponding typhoon samples for prediction modeling, and the data from 1990 to 2005 include 27 typhoons and 156 independent samples.

b. Predictors and the CLIPER prediction equation

The CLIPER prediction of typhoon intensity assumes that a future change in typhoon intensity is associated with the current intensity, location (latitude and longitude), and current rates of change. Thirty-one CLIPER predictors were selected for this paper for prediction modeling from the sample data of SCS typhoons (sample size = 330), and their individual correlation coefficients with the predictand (typhoon intensity) are statistically significant above the 0.05 confidence level, with the maximum being 0.53 and the minimum being 0.20 (Table 1). CLIPER is widely used as a standard statistical comparison in assessing the improvements that result from new objective prediction methods. In this paper CLIPER is used as a reference method to show whether or not the GNNEP model is better than the CLIPER model, which serves as a baseline of forecasting skill (Bessafi et al. 2002; Aberson

TABLE 1. Individual correlation coefficients between the predictor and predictands for CLIPER models in July ($n = 330$).

Persistence factor	Name of predictor	Correlation coef	Persistence factor	Name of predictor	Correlation coef
x_1	Initial lat ($^{\circ}\text{N}$)	-0.3383	x_{17}	Max surface wind at -12 h	-0.2418
x_2	Initial lon ($^{\circ}\text{E}$)	0.4294	x_{18}	Max surface wind at -24 h	0.3692
x_3	Initial min pressure at center	-0.4142	x_{19}	Zonal motion from -24 to -12 h	0.3125
x_4	Initial max surface wind	0.5312	x_{20}	Lat difference between 0 and -12 h	-0.2170
x_5	Zonal motion at -12 h	-0.2524	x_{21}	Wind speed difference between 0 and -12 h	-0.2458
x_6	Zonal motion at -24 h	-0.2477	x_{22}	Lat difference between 0 and -24 h	0.2256
x_7	Squared zonal motion at -12 h	0.2079	x_{23}	Lon difference between 0 and -24 h	-0.2413
x_8	Squared zonal motion at -24 h	0.2573	x_{24}	Lon at -6 h	0.4449
x_9	Zonal motion at -12 h \times one at -24 h	0.2477	x_{25}	Lon at -36 h	0.4504
x_{10}	Zonal \times meridional motion at -12 h	0.2592	x_{26}	Lat at -6 h	-0.3379
x_{11}	Lon at -12 h	-0.2097	x_{27}	Lat at -36 h	-0.3265
x_{12}	Lon at -24 h	0.4507	x_{28}	Max surface wind at -6 h	0.4562
x_{13}	Lat at -12 h	0.4439	x_{29}	Max surface wind at -36 h	0.3394
x_{14}	Lat at -24 h	-0.3340	x_{30}	-6 h min pressure at the center	-0.3371
x_{15}	-12 h min pressure at the center	-0.3120	x_{31}	-36 h min pressure at the center	-0.2643
x_{16}	-24 h min pressure at the center	-0.2898			

and Sampson 2003). The regression analysis was mostly adopted to build the prediction equation; it is objective, since it contains no tunable parameters. The only uncertainty is that in the face of a multitude of predictors, choosing a different F may yield different combinations of predictors and, therefore, different prediction equations with different prediction capabilities. For objective comparison and analysis, five stepwise regression prediction equations for typhoon intensity were developed corresponding to $F = 1.0$, $F = 2.0$, $F = 3.0$, $F = 4.0$, and $F = 5.0$, respectively:

$$\begin{aligned} \widehat{Y}_1 = & -61.1391 - 1.1503x_1 + 0.8654x_2 - 0.4489x_3 \\ & + 0.5608x_4 - 4.3676x_5 - 0.3955x_{10} \\ & + 0.1677x_{19} - 0.1202x_{20} + 36.5089x_{24} \\ & + 0.3985x_{30}, \\ (F = 1.0, \quad \sigma = 6.8553, \quad \text{and} \quad R = 0.7688), \end{aligned} \quad (6)$$

$$\begin{aligned} \widehat{Y}_2 = & -68.5671 - 1.1389x_1 + 0.8574x_2 - 0.4873x_3 \\ & + 0.5652x_4 - 4.3774x_5 - 0.4008x_{10} \\ & - 0.0736x_{20} + 36.5519x_{24} + 0.5651x_{30}, \\ (F = 2.0, \quad \sigma = 6.8563, \quad \text{and} \quad R = 0.7679), \end{aligned} \quad (7)$$

$$\begin{aligned} \widehat{Y}_3 = & -86.0145 - 1.3210x_1 + 0.9033x_2 - 0.4289x_3 \\ & + 0.5790x_4 - 0.1653x_5 - 0.3667x_{10} \\ & + 0.4482x_{30}, \\ (F = 3.0, \quad \sigma = 6.8921, \quad \text{and} \quad R = 0.7634), \end{aligned} \quad (8)$$

$$\begin{aligned} \widehat{Y}_4 = & -191.2276 - 1.3514x_1 + 0.8887x_2 + 0.7033x_4 \\ & - 0.1503x_5 - 0.3562x_{10} + 0.1251x_{19}, \\ (F = 4.0, \quad \sigma = 6.9962, \quad \text{and} \quad R = 0.7542), \quad \text{and} \end{aligned} \quad (9)$$

$$\begin{aligned} \widehat{Y}_5 = & -187.1563 - 1.4077x_1 + 0.8903x_2 + 0.7104x_4 \\ & - 0.1914x_{10} + 0.1208x_{19}, \\ (F = 5.0, \quad \sigma = 7.0343, \quad \text{and} \quad R = 0.7502), \end{aligned} \quad (10)$$

where σ denotes the residual standard deviation and R the multiple correlation coefficient.

The sample sizes of the above five CLIPER prediction equations are all 330, and the predictors in the prediction equations for $F = 1, 2, 3, 4$, and 5 are 10, 9, 7, 6, and 5, respectively. The five CLIPER prediction equations were verified using the 156 independent samples in 1990–2005; Fig. 1 shows the observed and predicted values of typhoon intensity from the five CLIPER equations. The statistics of the mean absolute prediction errors (Table 2) show that the prediction errors of Eqs. (6)–(8) ($F = 1, 2$, and 3), which contain 10, 9, and 7 predictors, respectively, are smaller (5.04 , 4.98 , and 4.99 m s^{-1}), while those of Eqs. (9) and (10) ($F = 4$ and 5) are relatively larger (5.28 and 5.41 m s^{-1}).

c. Nonlinear ensemble prediction of typhoon intensity

Five GNNEP models were constructed (see Fig. 2 for the procedures of prediction modeling) using the GNN

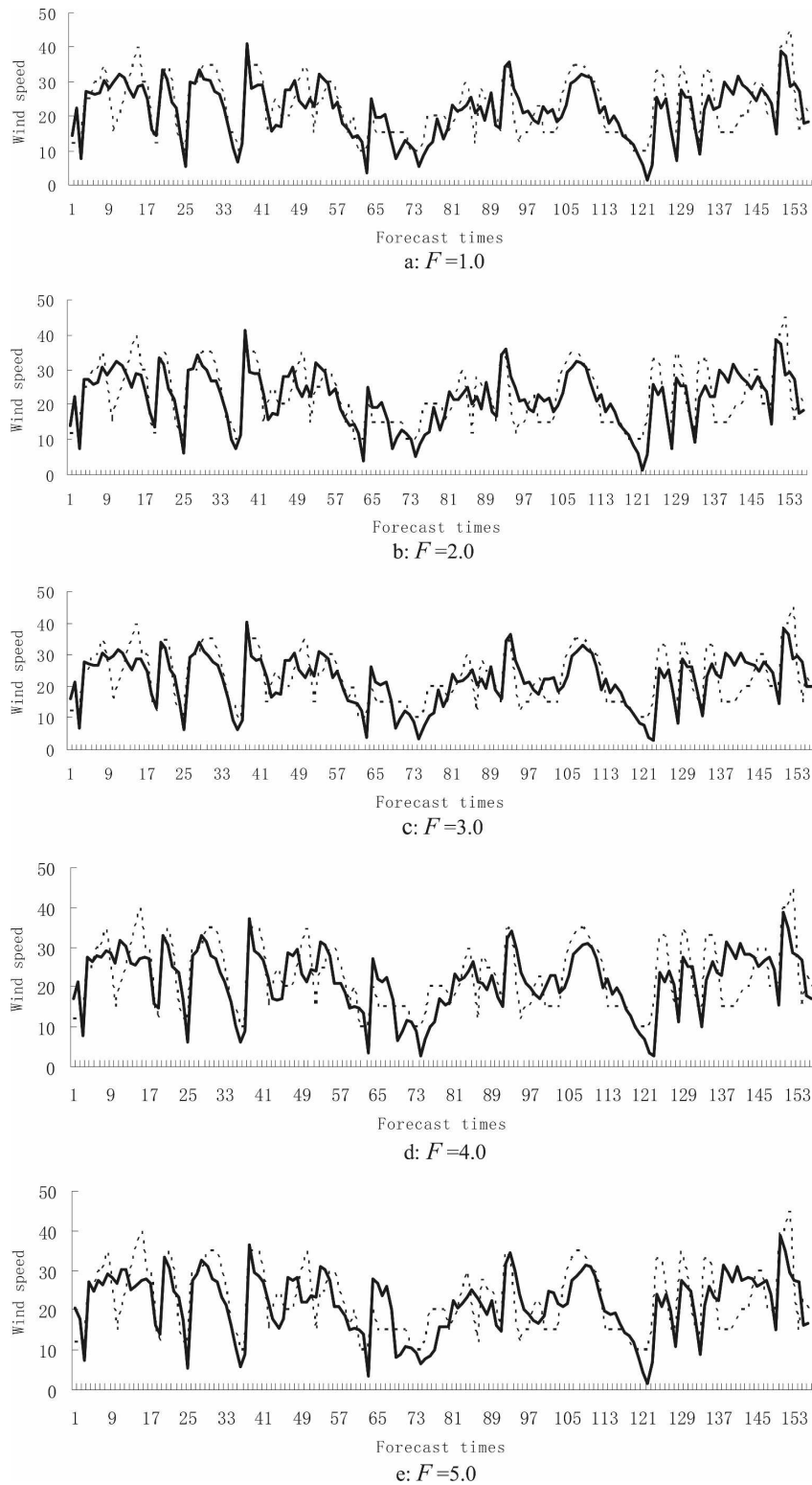


FIG. 1. Predicted (solid) values of the typhoon intensities of 156 independent samples by the five CLIPER prediction equations and the corresponding observed (dashed) values.

TABLE 2. Statistics of the fitting and prediction errors for the five CLIPER methods.

F (predictor No.)	Fitting error ($n = 330$)		Prediction error ($n = 156$)	
	Relative error (%)	Absolute error (m s^{-1})	Relative error (%)	Absolute error (m s^{-1})
1.0 (10)	27.6731	5.3767	26.5918	5.0484
2.0 (9)	27.6717	5.3772	26.1997	4.9892
3.0 (7)	27.8512	5.3861	26.6686	4.9926
4.0 (6)	28.3965	5.4595	27.6664	5.2841
5.0 (5)	28.9459	5.5462	28.6035	5.4170

ensemble method for creating individuals described in section 3, and the CLIPER predictors of Eqs. (6)–(10). In the evolutionary GA computation, we specified the size of the initial population and the number of evolu-

tionary generations to be 50. Roulette wheel selection was used in the selection operator and the multiple-point crossover with a crossover probability of 0.9 in the crossover operator; the threshold and crossover prob-

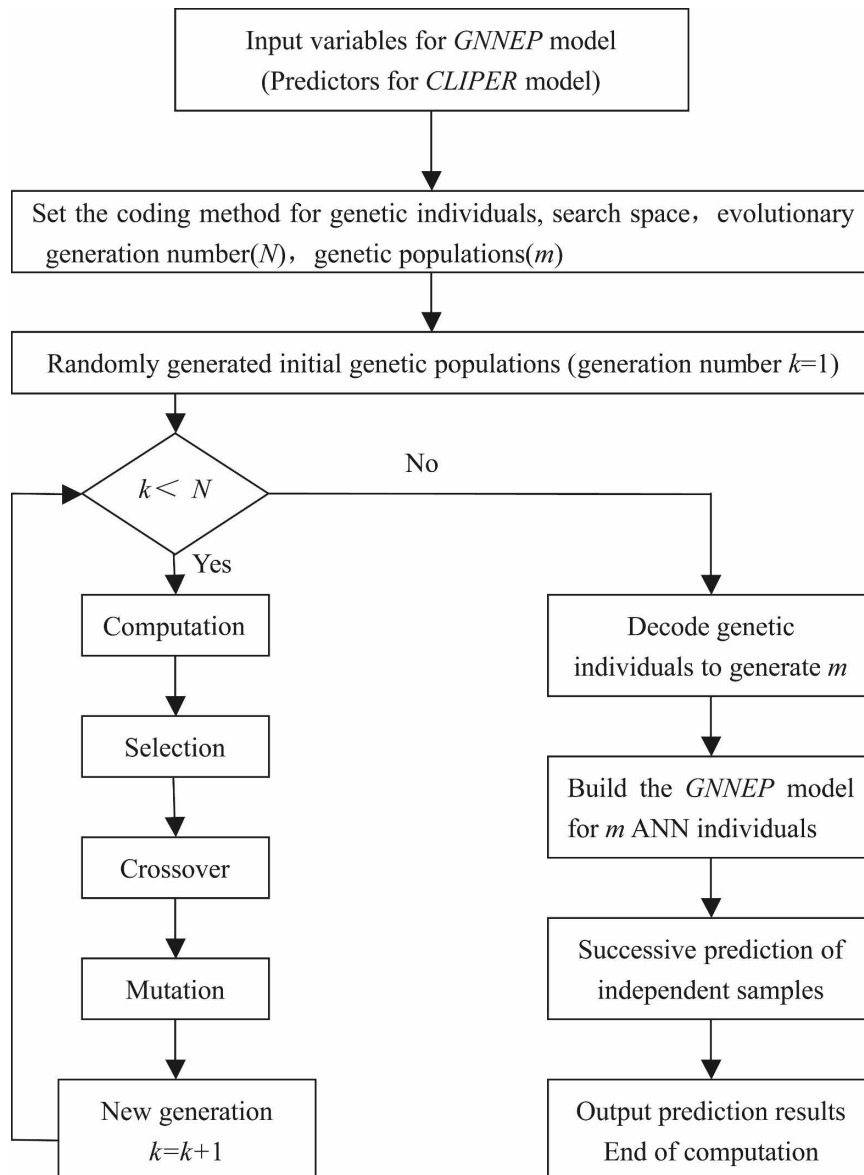


FIG. 2. Flowchart for GNNEP prediction modeling.

abilities of the weight coefficient were both 0.6, and the mutation algorithm had a mutation probability of 0.05. The input nodes of GNN were the number of predictors in the corresponding CLIPER equation, and the number of output nodes was one. The size of the search space for the hidden nodes of the GNN model was 0.5–1.5 times the number of the input nodes, with the solution space of the network connection weight having been specified as $[0, 1]$. The training time was 200, and the learning and momentum factors were both 0.5. After the evolutionary computation, the 50 genetic individuals were decoded to yield 50 neural network ensemble members, which were subsequently used to construct the resultant five GNNEP models using the ensemble average method. Likewise, prediction verification of the 156 independent samples from the years 1990–2005 was performed using each of the five GNNEP models. For each ensemble–CLIPER model, the first independent sample was predicted using the 330 modeling samples, the second using 331 modeling samples (330 modeling samples plus the first independent sample, which had become a known variable when the second sample was forecasted), and so on, until the last (156th) independent sample was predicted using 485 modeling samples. In all successive predictions, all parameters of the GA and GNN were kept unchanged, to ensure comparability between independent sample predictions and actual operational predictions. Figure 3 shows the predicted values of the typhoon intensities of 156 independent samples using the five GNNEP models constructed with different numbers (10, 9, 7, 6, and 5) of CLIPER predictors and the corresponding observed values. The statistics of the deviations of the predicted values from those observed (Table 3) show that the mean absolute prediction errors for the five GNNEP models are reduced by 20.3%, 22.7%, 21.7%, 24.6%, and 20.9%, respectively, in comparison to the corresponding CLIPER models with identical predictors. The prediction performance of the CLIPER and ANNEP models for the 156 samples was also assessed with grading errors. When $F = 1$, the prediction of a grading error $\leq 5 \text{ m s}^{-1}$ by the CLIPER (ANNEP) model occurs 84 (110) times, accounting for 53.8% (70.5%); therefore, the prediction occurrence of small grading errors with the CLIPER model is less than that for the ANNEP model; when $F = 2.0, 3.0, 4.0$, and 5.0 , the results are similar. Comparison and analysis of the statistics of the prediction errors in Tables 2 and 3 show that although the modeling samples, predictors, and independent samples of CLIPER and the ensemble models are identical, the prediction ability of GNNEP is clearly higher than that of the traditional CLIPER model for typhoon intensity.

5. Performance analysis of the ensemble prediction model

a. Contrast analysis of the ensemble model and ensemble members

The basic principle of the GNNEP model is that given the same expected output, the members (50 in this paper) of the model are first created using the evolutionary GA; then, their predictions are integrated with an equal weight to yield the resultant prediction of the ensemble model. To emphasize the difference in prediction performance between GNNEP and single-GNN prediction, five single-GNN prediction models were built using the predictors of the CLIPER prediction Eqs. (6)–(10) and the three-layer BP network model introduced in section 3. The input and output nodes of the network are the number of predictors and one, respectively, and the number of hidden nodes is determined by $Q = \sqrt{R + S^L} + a$ (Ali 2004), where R is the number of input nodes, S^L is the number of output nodes, and a is a constant. The learning factor and momentum factor are 0.5, and the training time is 200. The predictions of the five trained single-GNN prediction models were verified, respectively, using 156 independent samples from the years 1990 to 2005, and the predicted and observed values of the samples are shown in Fig. 4 with their relevant prediction errors in Table 4. Tables 2 and 3 clearly show that the mean prediction errors of the GNN prediction models are all smaller than those of the corresponding CLIPER models. This may be because the CLIPER prediction equation was set up using linear regression analysis, while the GNN method was set up using artificial intelligence nonlinear prediction modeling, which possesses strong nonlinear mapping and self-adaptation learning abilities and therefore better reflects the nonlinear evolutionary features of typhoon intensity. Furthermore, comparing the prediction performances of the five single-GNN models and the five GNNEP models (see Tables 3 and 4) clearly reveals that for identical predictors, modeling samples, and independent samples, the GNNEP models are all more accurate than the single-GNN models. The reason for this will be proven from a theoretical standpoint in the following subsection.

b. Theoretical derivation of the performance of the GNNEP model

The simple mean method, where the mean weight (λ_i ; $i = 1, 2, \dots, m$; $m = 50$ in this paper; $\lambda_i > 0$, and $\sum_{i=1}^m \lambda_i = 1$) is given for each GNN member, is used to integrate the predictions of the GNN ensemble members. If the model input for the i th GNN member is x_p , and its computed output is $f_i(x_p)$, then the output of the

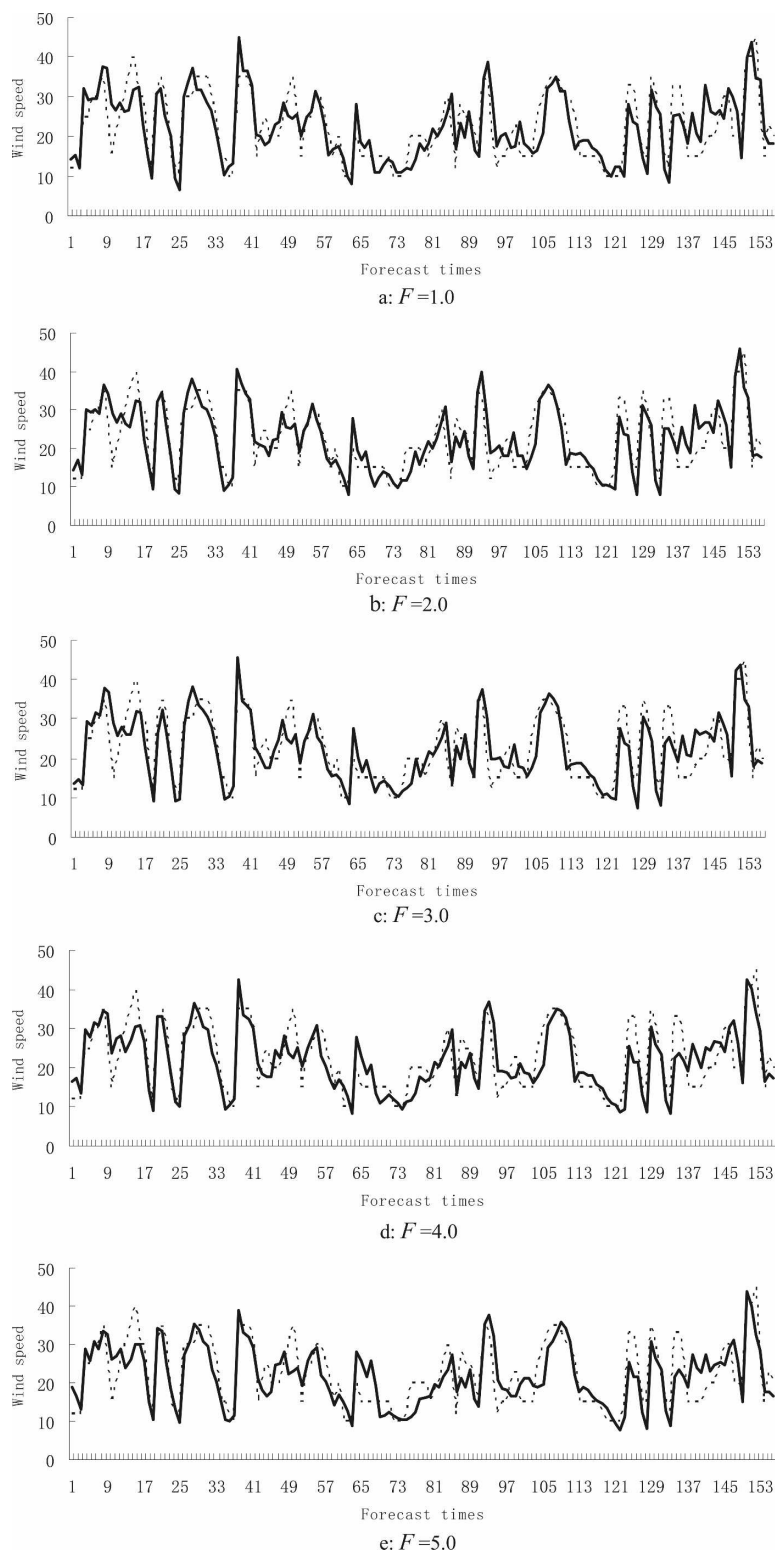


FIG. 3. Predicted (solid) values of the typhoon intensities of 156 independent samples by the ensemble model and the corresponding observed (dashed) values.

TABLE 3. Statistics of the fitting and prediction errors for the five ensemble prediction models.

<i>F</i> (factor No.)	Fitting error (<i>n</i> = 330)		Prediction error (<i>n</i> = 156)	
	Relative error (%)	Absolute error (m s ⁻¹)	Relative error (%)	Absolute error (m s ⁻¹)
1.0 (10)	18.6170	3.7443	20.2437	4.0221
2.0 (9)	19.6140	3.9590	19.4349	3.8559
3.0 (7)	19.3620	3.9060	19.3257	3.9074
4.0 (6)	21.9780	4.4205	19.7924	3.9835
5.0 (5)	24.0410	4.8309	20.8284	4.2866

ensemble prediction of the GNNEP model consisting of *m* GNN members is

$$\bar{f}(x_p) = \sum_{i=1}^m \lambda_i f_i(x_p), \quad (11)$$

and the corresponding ensemble prediction error is

$$E = \sum_{p=1}^N [y_p - \bar{f}_i(x_p)]^2. \quad (12)$$

Because the expected output of all GNN members is the same, the prediction error of some member of the ensemble model is

$$E_i = \sum_{p=1}^N [y_p - f_i(x_p)]^2, \quad (13)$$

and the weighted average of the prediction errors for all GNN members is

$$\bar{E} = \sum_{i=1}^m \lambda_i E_i. \quad (14)$$

When the model input is x_p , if the computed output of the *i*th GNN member is $f_i(x_p)$ and that of the ANN model is $\bar{f}_i(x_p)$, then the diversity between the *i*th member and the ensemble model is

$$S_i = \sum_{p=1}^m [f_i(x_p) - \bar{f}_i(x_p)]^2, \quad (15)$$

and the diversity of all members is

$$\bar{S} = \sum_{i=1}^m \lambda_i S_i. \quad (16)$$

Based on Eqs. (16) and (15), the relationship between the prediction error of the ensemble model and the mean prediction error of its members can be derived as follows.

Assuming that the model input $x \in R^m$ satisfies the distribution $p(x)$, then the following relation can be obtained using Eqs. (16) and (15):

$$\begin{aligned} \bar{S} &= \sum_{i=1}^m \lambda_i S_i = \sum_{i=1}^m \lambda_i \int dx p(x) [f_i(x) - \bar{f}(x)]^2 \\ &= \int dx p(x) \left\{ \sum_{i=1}^m \lambda_i [f_i(x) - \bar{f}(x)]^2 \right\} \\ &= \int dx p(x) \left\{ \sum_{i=1}^m \lambda_i [f_i(x) - d(x) + d(x) - \bar{f}(x)]^2 \right\} \\ &= \int dx p(x) \left\{ \sum_{i=1}^m \lambda_i [f_i(x) - d(x)]^2 + [d(x) - \bar{f}(x)]^2 \right. \\ &\quad \left. + 2 \sum_{i=1}^m \lambda_i [f_i(x) - d(x)][d(x) - \bar{f}(x)] \right\} \\ &= \int dx p(x) \left\{ \sum_{i=1}^m \lambda_i [f_i(x) - d(x)]^2 - [d(x) - \bar{f}(x)]^2 \right\} \\ &= \int dx p(x) \left\{ \sum_{i=1}^m \lambda_i [f_i(x) - d(x)]^2 \right. \\ &\quad \left. - \int dx p(x) [d(x) - \bar{f}(x)]^2 \right\} \\ &= \sum_{i=1}^m \lambda_i \int dx p(x) [f_i(x) - d(x)]^2 \\ &\quad - \int dx p(x) [d(x) - \bar{f}(x)]^2 \\ &= \sum_{i=1}^m \lambda_i E_i - E \\ &= \bar{E} - E \end{aligned}$$

therefore,

$$E = \bar{E} - \bar{S}. \quad (17)$$

Equation (17) clearly shows that the prediction error of GNNEP is associated with the mean prediction error and the diversity of all of the ensemble members. Since the diversity E_i of each member is nonnegative, the prediction accuracy of the ensemble model should be

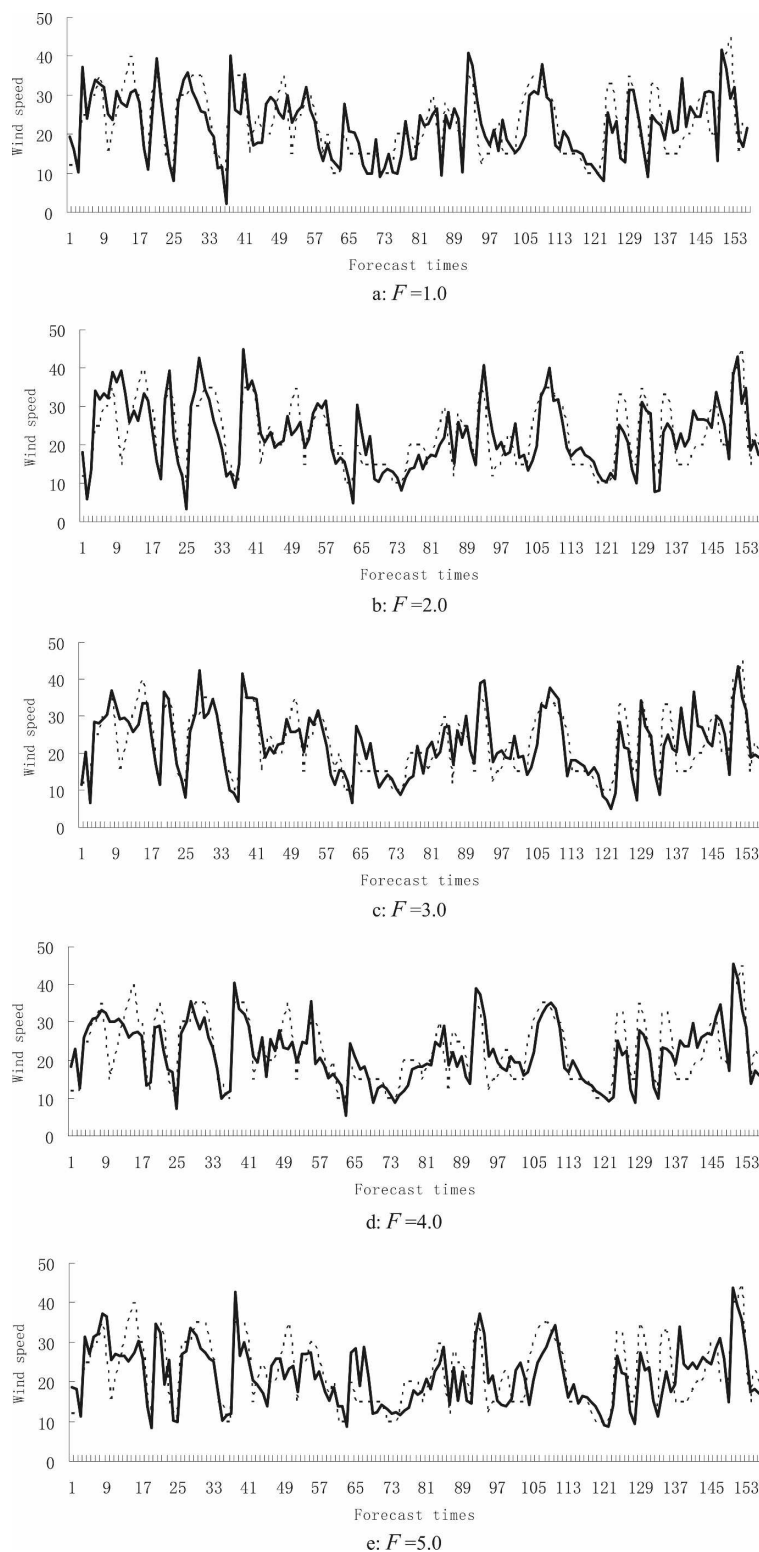


FIG. 4. Predicted (solid) values of the typhoon intensities of 156 independent samples by the single-ANN models and the corresponding observed (dashed) values.

TABLE 4. Statistics of the fitting and prediction errors for single-GNN models.

<i>F</i> (factor No.)	Fitting error ($n = 330$)		Prediction error ($n = 156$)	
	Relative error (%)	Absolute error (m s^{-1})	Relative error (%)	Absolute error (m s^{-1})
1.0 (10)	21.9410	4.3671	23.8894	4.5981
2.0 (9)	19.0530	3.8687	24.3264	4.4260
3.0 (7)	21.7960	4.4071	21.7012	4.3607
4.0 (6)	23.5340	4.7406	20.9480	4.2956
5.0 (5)	23.0360	4.7577	22.5289	4.7751

higher than the mean accuracy of all ensemble members, and Zheng et al. (2004) also reached the same conclusion after analyzing the generalization errors of neural networks. The results of the prediction verification of SCS typhoon intensity in section 4 confirm this.

c. Generalization capability of the ensemble prediction model

The structure of a single neural network prediction model is generally determined by personal experience due to lack of theoretical guidance (Jin 2004). In meteorological applications of the neural network method, it is difficult to optimize the structure of a network before the weather event occurs, which results in “overfitting” of the prediction model (Jin 2005b), thus reducing the prediction accuracy. When training a three-layer BP neural network using a training set of samples, after a certain time, the fitting accuracy of the network model to the training samples gradually increases, while the prediction error obviously grows with the increase in training time (Fig. 5). In the learning stage, there is no way to correctly determine the fitting accuracy at which the prediction model will have the best prediction capacity for unknown samples (future weather). Therefore, the “overfitting” of a neural network prediction model is an important indicator in assessing whether the model is of operational value. The GNNEP model put forth in this paper is, by its nature, an ensemble model consisting of multiple neural networks, so it is necessary to computationally analyze whether overfitting occurs in predicting with the new ensemble prediction model.

To determine whether overfitting occurs in GNNEP, the five ensemble models in Table 3 were trained with different training times (20, 50, 100, 200, 300, 400, 500, 600, 700, and 800 times; all other parameters remained unchanged), and the successive prediction tests of 156 independent samples for each ensemble model were performed. Table 5 shows the variations in the fitting accuracy of the ensemble models to the training samples and the related prediction accuracy for the independent samples. The results indicate that when the

training time increased from 20 to 100, the fitting and prediction errors of the five ensemble models changed only slightly, and that even when the training time increased incrementally by 100 from 100 to 800, the two errors still remained unchanged, indicating that there was no overfitting phenomenon in the prediction of the GNNEP models. The prediction modeling method is of practical significance, because the overfitting of any neural network prediction model necessarily deteriorates the prediction accuracy of the model to independent samples. Using the same predictors, modeling samples, and independent samples, the GNNEP model has a prediction accuracy that is not only higher than the CLIPER model, but also higher than the single-GNN model. Because the GNNEP model overcomes the overfitting obstacle, it is applicable to actual operational weather forecasts.

Using identical modeling samples, predictors, and independent samples, the prediction accuracy of July SCS typhoon intensities from the GNNEP models was compared with widely used traditional CLIPER models and single-GNN models; the improved prediction capability and generalization capability of the new prediction modeling have been theoretically proven to be reasonable in sections 4 and 5. To examine the stability of its prediction capability, similar prediction experiments

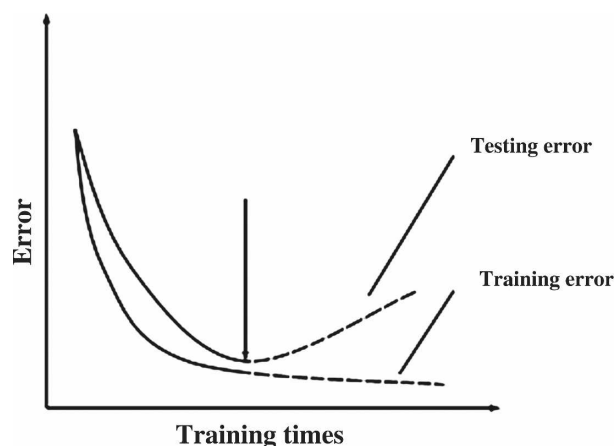


FIG. 5. Sketch map for the overfitting of a BP network mode.

TABLE 5. Statistics of the fitting (FE) and prediction errors (PE) (m s^{-1}) of the five GNNEP models for July.

Factors	Training times	20	50	100	200	300	400	500	600	700	800
10	FE	3.827	3.745	3.744	3.744	3.744	3.744	3.744	3.744	3.744	3.744
	PE	3.975	4.022	4.023	4.022	4.022	4.022	4.022	4.022	4.022	4.022
9	FE	4.003	3.961	3.959	3.959	3.959	3.959	3.959	3.959	3.959	3.959
	PE	3.847	3.856	3.856	3.855	3.855	3.855	3.855	3.855	3.855	3.855
7	FE	3.965	3.906	3.906	3.906	3.906	3.906	3.906	3.906	3.906	3.906
	PE	3.907	3.908	3.907	3.907	3.907	3.907	3.907	3.907	3.907	3.907
6	FE	4.429	4.420	4.420	4.420	4.420	4.420	4.420	4.420	4.420	4.420
	PE	3.978	3.983	3.983	3.983	3.983	3.983	3.983	3.983	3.983	3.983
5	FE	4.831	4.830	4.830	4.830	4.830	4.830	4.830	4.830	4.830	4.830
	PE	4.292	4.286	4.286	4.286	4.286	4.286	4.286	4.286	4.286	4.286

for August and September were also performed, and the results are similar to those for July. As space is limited, it is unnecessary to go into detail here. The analysis results from the aforementioned large sample size prediction experiments for July, August, and September SCS typhoon intensities clearly show that the nonlinear GNNEP model for typhoon intensity introduced here tends to have a higher prediction accuracy, a stable prediction capability, and robust generalization capability.

6. Summary

Ensemble prediction improves the effectiveness of NWP. In this work, we present a NAIEP modeling approach based on GA and ANN that emulates the ensemble prediction of NWP whose merits are as follow:

- 1) In the case of identical predictors, prediction modeling samples, and independent prediction samples, the prediction accuracy of our GNNEP model is clearly higher than that of the traditional CLIPER model (which uses regression analysis).
- 2) Likewise, under the same conditions, the GNNEP model is also more accurate than the single neural network prediction model, because the structure of its member network is determined by the optimized GA instead of by personal experience. Further computation and analysis indicate that there is no “overfitting” in the predictions of the new ensemble model, a phenomenon that generally exists in predictions by traditional single neural network methods. Therefore, GNNEP is directly applicable to actual operational weather prediction, thus providing statistical weather prediction with a new nonlinear intelligence prediction technique.
- 3) The relationship, derived in section 5b, of the prediction error of GNNEP model with the mean prediction error and diversity of its network members,

provides theoretical evidence for the computational result that GNNEP is more accurate than the single neural network prediction model.

Overall, this research suggests that the meteorological ensemble modeling approach of GNN opens up a vast range of possibilities for operational weather prediction. The theoretical derivation suggests that constructing a more highly diverse population of ensemble members will further improve the prediction ability of the GNNEP model. Therefore, construction of a highly diverse ensemble merits further exploration.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant 40675023) and the Science and Technology Commonwealth of China (Grant 2004DIB3J122).

APPENDIX

Key Acronyms and Abbreviations

ANN	Artificial neural network
BP	Back propagation
CLIPER	Climatology and persistence
ENP	Ensemble numerical prediction
FE	Fitting error
GA	Genetic algorithm
GNN	Genetic neural network
GNNEP	Genetic neural network ensemble prediction
NAIEP	Nonlinear artificial intelligence ensemble prediction
NWP	Numerical weather prediction
PE	Prediction error
SCS	South China Sea

REFERENCES

- Aberson, S. D., and C. R. Sampson, 2003: On the predictability of tropical cyclone tracks in the northwest Pacific basin. *Mon. Wea. Rev.*, **131**, 1491–1497.

- Ali, A. H., 2004: Application of neural network principal components to climate data. *J. Atmos. Oceanic Technol.*, **21**, 149–158.
- Bessafi, M., A. Lasserre-Bigorry, C. J. Neumann, F. Pignolet-Tardan, D. Payet, and M. Lee-Ching-Ken, 2002: Statistical prediction of tropical cyclone motion: An analog-CLIPER approach. *Wea. Forecasting*, **17**, 821–831.
- Chen, G.-L., X.-F. Wang, and Z.-Q. Zhuang, 1996: *Genetic Algorithm and Application* (in Chinese). Beijing Communication Press, 178 pp.
- DeMaria, M., and J. Kaplan, 1991: A statistical model for predicting tropical cyclone intensity change. Preprints, *19th Conf. on Hurricanes and Tropical Meteorology*, Miami, FL, Amer. Meteor. Soc., 521–526.
- , M. Mainelli, L. K. Shay, J. A. Knaff, and J. Kaplan, 2005: Further improvement to the Statistical Hurricane Intensity Prediction Scheme (SHIPS). *Wea. Forecasting*, **20**, 531–543.
- Ding, Y.-G., Z.-Q. Cheng, and B.-Y. Cheng, 2002: A prediction experiment by using the generalized canonical mixed regression model based on MSSA-SVD for ENSO (in Chinese). *Acta Meteor. Sin.*, **60**, 361–369.
- Du, J., 2002: Present situation and prospects of ensemble numerical prediction (in Chinese). *J. Appl. Meteor. Sci.*, **13**, 16–28.
- Guo, Z.-L., and Coauthors, 2004: Genetic algorithm-neural network-based economic losses assessment of seismic disaster (in Chinese). *J. Nat. Disasters*, **13** (6), 92–96.
- Has, B.-G., J.-W. Ma, Q.-Q. Li, X. Chen, and Q. Dai, 2004: Remote sensing data classification based on tolerant rough set and neural networks. *Sci. Chin.*, **34D**, 967–974.
- Jin, L., 2004: *Modeling Theory Method and Application for Weather Forecast Based on Neural Network* (in Chinese). China Meteorological Press, 218 pp.
- , 2005a: Short-term climate prediction model of neural network based on genetic algorithms (in Chinese). *J. Plateau Meteor.*, **24**, 981–987.
- , 2005b: Study on the overfitting of the artificial neural network forecast model (in Chinese). *Acta Meteor. Sin.*, **63**, 216–225.
- , Y. Luo, and Y.-H. Li, 2003: Study on mixed prediction model of artificial neural network for long-range weather (in Chinese). *J. Syst. Eng.*, **118**, 331–336.
- , C. Yao, and X.-Y. Huang, 2006: An improved method on meteorological prediction modeling using genetic algorithm and artificial neural network. *The Sixth World Conference on Intelligent Control and Automation*, Vol. 1, Dailan, China, IEEE, 31–35.
- Knaff, J. A., M. DeMaria, C. R. Sampson, and J. M. Gross, 2003: Statistical 5-day tropical cyclone intensity forecasts derived from climatology and persistence. *Wea. Forecasting*, **18**, 80–92.
- , C. R. Sampson, and M. DeMaria, 2005: An operational statistical typhoon intensity prediction scheme for the western North Pacific. *Wea. Forecasting*, **20**, 688–699.
- Li, A.-J., H.-J. Li, K.-Z. Li, and Z.-B. Gu, 2003: Modeling of CVI process in fabrication of carbon/carbon composites by an artificial neural network. *Sci. China*, **33E**, 209–216.
- Liu, M.-Q., 2005: Discrete-time delayed standard neural network model and its application. *Sci. China*, **33E**, 1031–1048.
- Liu, Y.-D., B. Wang, and Z.-M. Hou, 2003: Application of the optimum decision method to typhoon track forecasting (in Chinese). *J. Tropical Meteor.*, **19**, 219–224.
- Lu, C.-L., S.-H. Chen, and Y.-T. Zhu, 1996: Research and apply multiple dynamic interdependent model (MDIM) to predict typhoon track, intensity and wind-speed simultaneously (in Chinese). *Acta Meteor. Sin.*, **54**, 737–744.
- Luo, C.-S., Q. Zuo, and B. Li, 2004: Application of artificial neural network based on the genetic algorithm in predicting the root distribution of winter wheat. *Chin. J. Appl. Ecol.*, **15**, 354–356.
- Meng, Z.-Y., L.-S. Chen, and X.-D. Xu, 2002: Recent progress on tropical cyclone research in China. *Adv. Atmos. Sci.*, **19**, 103–110.
- Nohara, D., and H. L. Tanaka, 2004: Development of prediction model using ensemble forecast assimilation in nonlinear dynamical system. *J. Meteor. Soc. Japan*, **82**, 167–178.
- Scherrer, S. C., and Coauthors, 2004: Analysis of the spread-skill relations using the ECMWF Ensemble Prediction System over Europe. *Wea. Forecasting*, **19**, 552–565.
- Singh, R., P. C. Joshi, and C. M. Kishtawal, 2005: A new technique for estimation of surface latent heat fluxes using satellite-based observations. *Mon. Wea. Rev.*, **133**, 2692–2710.
- Stensrud, D. J., J.-W. Bao, and T. W. Thomas, 2000: Using initial condition and model physical perturbation in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077–2107.
- Su, J.-H., Q.-M. Dong, and P. Liu, 2005: Establishing the knowledge repository of rapidly solidified aging Cu–Cr–Zr alloy on the artificial neural-network. *Sci. Chin.*, **35E**, 1176–1185.
- Tapiador, F. J., C. Kidd, V. Levizzani, and F. S. Marzano, 2004: A neural networks-based fusion technique to estimate half-hourly rainfall estimates at 0.1° resolution from satellite passive microwave and infrared data. *J. Appl. Meteor.*, **43**, 576–594.
- Wang, H.-P., T.-J. Lu, and W.-H. Yu, 2006: Pressure forecasting of submarine-borne missile cover system based on neural network evolved by genetic algorithm (in Chinese). *J. Beijing Inst. Technol.*, **2006**, 23–26.
- Wang, Y., and H. Wang, 2006: Prediction of highway traffic based on BP-neural network and genetic algorithms (in Chinese). *Comp. Eng. Appl.*, **42**, 226–228.
- Zhang, R., and Coauthors, 2006: Non-linear dynamic model retrieval of subtropical high based on empirical orthogonal function and genetic algorithm. *Appl. Math. Mech.*, **27**, 1645–1653.
- Zheng, D.-L., R.-X. Liang, Y. Zhou, and Y. Wang, 2003: A chaos genetic algorithm for optimizing an artificial neural network of prediction silicon content in hot metal (in Chinese). *J. Univ. Sci. Technol. Beijing*, **10** (20), 68.
- Zheng, J.-J., Y.-S. Liu, Q.-X. Liu, and M. Sun, 2004: A dynamic integration approach for a neural network ensemble (in Chinese). *Comput. Eng.*, **30**, 49–51.
- Zhou, C.-H., and Coauthors, 2004: A novel approach for complexity modeling based on genetic-neural network training (in Chinese). *J. Tianjin Inst. Technol.*, **20** (3), 57–59.
- Zhou, J.-B., and J.-Y. Huang, 1997: Advances in statistical meteorology in China in recent years (in Chinese). *Acta Meteor. Sin.*, **55**, 297–305.
- Zhou, X.-Q., and C. L. Johnny, 2006: Ensemble forecasting of tropical cyclone motion using a baroclinic model. *Adv. Atmos. Sci.*, **23**, 342–354.

Copyright of *Monthly Weather Review* is the property of American Meteorological Society and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.