

## Research on Extreme Learning Algorithm and Its Application to Atmospheric Nonlinear Systems

Qunbo Huang<sup>1,2, a</sup>, Bainian Liu<sup>1, b</sup>, Weimin Zhang<sup>1</sup>, Jingzhe Sun<sup>1</sup>, Mengbin Zhu<sup>1</sup>, Shiwei Lin<sup>1</sup>, Weifeng Wang<sup>2</sup>

<sup>1</sup>Academy of Ocean Science and Engineering, National University of Defense Technology, Changsha, 410073, China

<sup>2</sup>Weather Center, PLA Air Force, 100843, Beijing, China

<sup>a</sup>email: hqb09@163.com, <sup>b</sup>email: liuxialei580@163.com

**Keywords:** Extreme learning machine; Nonlinear chaotic system; Lorenz63

**Abstract.** Extreme learning machine (ELM) algorithm, which becomes more and more popular in the area of artificial intelligence for the past few years, is faster than the traditional machine learning algorithms, especially than the single hidden layer feed-forward neural networks (SLFNs). However, ELM is merely commonly used in the field of computer science or other hot areas. This paper investigates the ability of ELM to emulate the atmospheric nonlinear systems. The performance of ELM on emulating the nonlinear chaotic system - Lorenz63 is analyzed. The results show that ELM can accurately and quickly simulate the Lorenz63 forecast field at different forecast length, and thus providing a new idea for solving kinds of atmospheric nonlinear equations.

### Introduction

The traditional neural networks (such as BP neural network) based on the gradient descent algorithms are widely used in training the multilayer feed-forward neural network for its good learning ability [1]. However, these traditional learning algorithms need to correct weight and threshold values with many iterations, and easily lead to time-consuming, slow speed, fall into local optimal solution and must adjust different parameters (learning rate, for instance) for different applications. The problems described above become major bottlenecks restricting the development of traditional neural networks. Therefore, a novel feed-forward neural network was proposed [2], namely Extreme learning machine algorithm, which can fast and accurately obtain global optimal solution with simple adjustment of network parameters, and thus has been the concerns and hot topics for many scholars in recent years [3,4]. Zong studied the performance of the one-against-all (OAA) and one-against-one (OAO) ELM for classification in multi-label face recognition applications [5]. A real-time watermarking scheme based on Regularized Extreme Learning Machine (RELM) was developed, which can extract the watermark from the watermarked image within very short time, the training speed always hundreds of times faster than BP neural network and Support Vector Machine (SVM) [6]. Moreover, Liu proposed a general learning framework, termed multiple kernel extreme learning machines (MK-ELM), which can provide a general framework for ELM to integrate multiple heterogeneous data sources for classification [7].

Numerical Weather Prediction (NWP) is essentially the numerical dynamic simulation of atmospheric motion by highly nonlinear differential equations, and how to quickly and accurately solve the equations has been an important work for meteorologists. However, equations in NWP always tend to be very complex, if there is a kind of new method for a new algorithm which would be used for NWP, for convenience, we usually choose a relatively simple model and can still represent the actual characteristics of atmospheric state. Lorenz63 is such a system, for instance, which was chosen by Cao to verify the performance of a new data assimilation method based on dual-number theory in numerical weather prediction, as in [8]. Chen also adopted following three methods: Augmented Ensemble Kalman Filter (AEnKF), Dual Ensemble Kalman Filter (DEnKF) and Simultaneous Optimization and Data Assimilation (SODA) to compare the effect of simultaneous states and parameters estimation based on Lorenz63 model [9]. Therefore this paper

employs the Lorenz63 system to do the numerical experiments.

In this paper, we explore the possibility of applying the ELM algorithm to simulate the atmospheric nonlinear chaotic system - Lorenz63. The structure of the paper is organized as follows. The following section introduces the calculation principle and flow of ELM algorithm. Then, we present the feature of Lorenz63 system and discuss the results of the numerical simulated experiments. Finally, conclusions and plans for future work are drawn.

## The Elm Alogrithm

The ELM algorithm randomly generates the continuous weight and threshold values between the input layer and the hidden layer without adjusting parameters in the training phase except setting the number neurons in hidden layer, and thereby can obtain the global optimal solution. It has the advantages of fast learning speed and good generalization performance compared with the traditional training methods. Figure1 shows the structure of a typical SLFN.

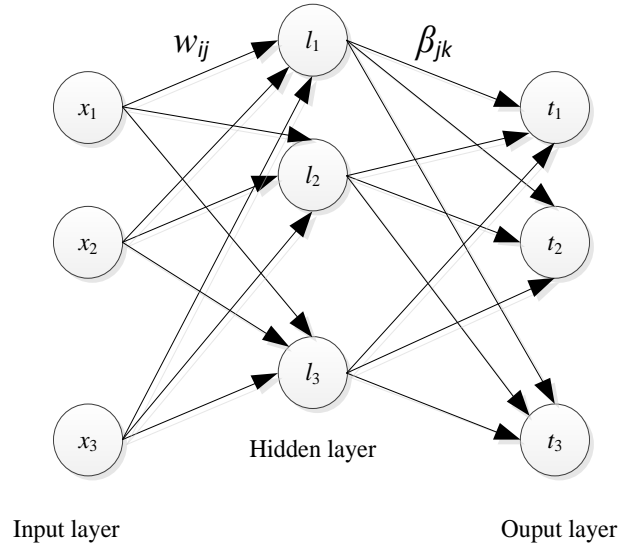


Fig.1. Typical structure of a SLFN

For  $N$  random samples  $(\mathbf{x}_i, \mathbf{t}_i)$ , where  $\mathbf{x} = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$ ,  $\mathbf{t} = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$ . The mathematical model of a SLFN with  $l$  nodes in hidden layer can be expressed as (1) and (2):

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T} \quad (1)$$

$$\sum_{i=1}^l \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + \mathbf{b}_i) = t_j, j = 1, \dots, N \quad (2)$$

Among them, the left side of the Eq. 1 corresponds to the left side of the Eq. 2.  $\mathbf{H}$  being the output matrix of the hidden layer,  $\boldsymbol{\beta}, \mathbf{T}, g(\mathbf{x})$ ,  $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$  and  $\mathbf{b}_i$  are output weights, expected output, activation function, input weight and bias respectively.

Minimizing the output error is the learning goal of a neural network, that is

$$\sum_{j=1}^l \|\mathbf{o}_j - \mathbf{t}_j\| = 0 \quad (3)$$

wherein  $\mathbf{o}_j$  is the forecast values generated from the neural network.

It is hard for us to adjust all the parameters in traditional neural network based on the gradient descent method to deal with different cases. However, for ELM algorithm, the output weight  $\mathbf{w}_i$  and bias  $\mathbf{b}_i$  can be given randomly and the output matrix  $\mathbf{H}$  is uniquely determined.

The algorithm procedure of ELM is described as follows:

step1: randomly generate the parameters used for hidden nodes  $(\mathbf{w}_i, \mathbf{b}_i), i = 1, K, l$ .

step2: calculate the output matrix  $\mathbf{H}$  of the hidden layer.

step3: calculate the output weight  $\beta$ , and the optimal solution of which is  $\hat{\beta} = \mathbf{H}^\dagger \mathbf{T}$ , where  $\mathbf{H}^\dagger = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$  is the Moore-Penrose generalized inverse matrix of  $\mathbf{H}$ , and it has been proved that its norm is minimum and unique.

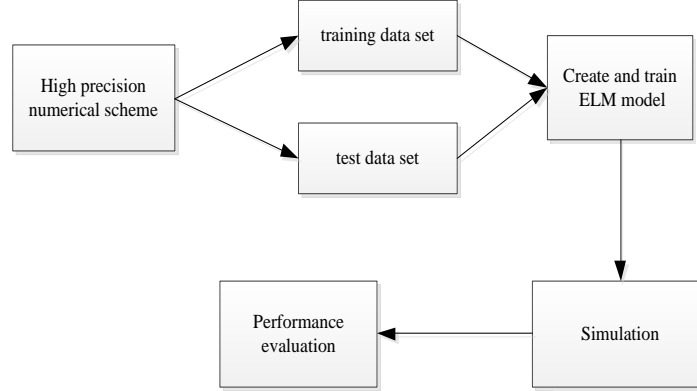


Fig.2. Flow of ELM

In summary, the flow of ELM algorithm can be expressed as follows (Fig. 2):

(1) Generate the training data set and test data set: in order to obtain better generalization performance of the simulated model, sufficient and better representative samples are required for ELM.

(2) Create and train the ELM model.

(3) ELM simulation test.

(4) Performance evaluation. The generalization ability of the ELM model can be evaluated by the means of the error (such as the root mean square error, RMSE) between the predicted and real values.

### Lorenz63 Nonlinear Chaotic System

In 1963, meteorologist E.N. Lorenz found a famous nonlinear differential system (called as Lorenz63) when he did the numerical calculation of dynamic models in weather prediction, and it was the first strange attractor observed in chaotic phenomenon [10]. It has been found that Lorenz equations can be used as accurate models of many real chaotic motion (such as turbulence, convection, oblique wave and so on), and thus it is brought to the attention of many scientists. Lorenz63 model equations are:

$$\begin{aligned}
 dx/dt &= \sigma(-x + y) \\
 dy/dt &= \gamma x - y - xz \\
 dz/dt &= xy - \beta z
 \end{aligned} \tag{4}$$

Where  $\sigma, \gamma, \beta$  represent the atmospheric Prandtl number, Rayleigh number and parameters associated with convective scale respectively [11].  $x$  can be considered as atmospheric convection intensity,  $y$  is the maximum temperature difference,  $z$  is the changing of atmospheric layers caused by convection motion. Although the model is simple, this non-linear system has a lot of similar dynamic features with the atmosphere. The Lorenz63 system is dissipated with the phase volume shrinking to zero along with the evolution of time, so that all trajectories in the phase space will ultimately be attracted to the point set where equals to zero, and the point set is called as attractor. The three parameters used by Lorenz are  $\sigma = 10, \gamma = 28, \beta = 8/3$ , and these equations are nonlinear differential equations without analytical solutions but numerical solutions. Moreover, the numerical solutions of these equations are very sensitive to the selected parameters and the initial conditions. Changed parameters and initial conditions may let the system to be stable or chaotic. For instance, any two motion trajectories which begin from two much closed points at the initial

moment would ultimately separate very far from each other. The strange attractors in the Lorenz system are ergodic, any motion trajectory almost all passes through the points on the strange attractor. The phenomenon described above is also known as “chaotic motion” which is highly similar to the characteristics of atmospheric turbulence motion. However, the atmosphere itself has chaotic characteristics, even if we employ the most accurate numerical models and the most exact initial conditions, we would still get two completely different weather forecast states after a certain time point began from the same model and initial conditions, and this is called as the predictability of the atmosphere. Therefore, Lorenz63 nonlinear chaotic system implicates some complicated weather changing mechanism, and the whole system presents an appearance shape of butterfly.

## Numerical Experiment

In the experiment, firstly integrate Eq. 4 forward 1500 time steps by using a fourth-order Runge-Kutta (R-K4) method. In order to express the research content more clearly, we assume that only a single station is selected to simulate, the time step length is  $\Delta t = 0.01$ , with initial background condition for the Lorenz system:  $\mathbf{u}^0 = (2, 3, 2.5) + \delta$ , where  $\delta$  is a Gaussian noise which makes the initial condition is closer to the real observation. For training data set, 900 data are considered, and 100 data are employed for validation. Our ultimate goal is to produce accurately evolving state of the Lorenz system by ELM algorithm according to the background field.

Subsequently, we select the Lorenz integral values at 300th, 500th, 600th, 800th and 1000th time steps as training and testing samples, and we name this experiment group as EX1. We also consider the validity of the ELM when the noise becomes larger (namely  $10 * \delta$ ) and call this experiment group as EX2.

Figure 3 shows the forecast errors of variables  $x$ ,  $y$  and  $z$  at 1000th time step between the ELM forecast filed and the truth in test data set in experiment group EX1. It is obvious that the errors are very small, and the results can also be confirmed by the distribution of the truth and the ELM generated forecast field in Fig. 4, where the forecast and true values are basically in the state of superposition for the 100 samples. These results present the accuracy of ELM algorithm used in Lorenz system is very high.

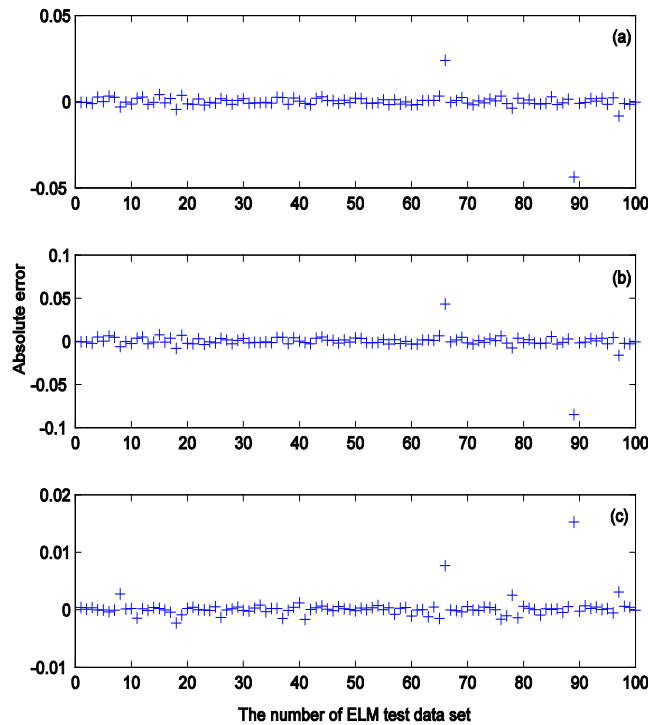


Fig.3. Forecast errors of variables  $x$ (a),  $y$ (b) and  $z$ (c) at 1000th time step between the ELM forecast filed and the truth in test data set in experiment group EX1

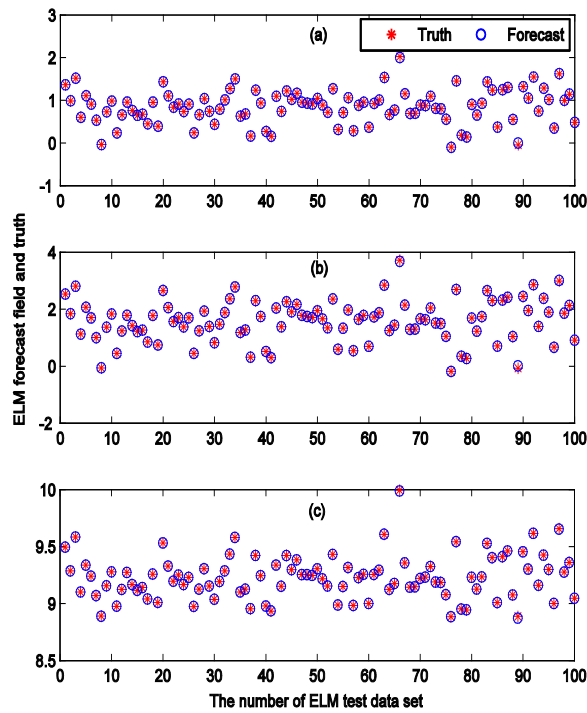


Fig.4. The distribution of the truth and the ELM forecast field of variables x(a),y(b) and z(c) in experiment group EX1

The running times of integrating 1000 steps in EX1 and EX2 are both shown in Table 1. The R-K4 method is nearly 2~4 times slower than ELM algorithm. However, with the evolution of time or the increase of data noise, the effect of ELM on Lorenz63 system will have a sharp decline, and there is a loss of accuracy after the 1000th time step, mainly because of the chaos in the Lorenz system, and thus leading to the unpredictability for a long period in highly nonlinear chaotic system. These conclusions are shown in the Table 2.

Table 1. The running times of integrating 1000 steps for the lorenz63 equations by using r-k4 and Elm

<b>Time(s)</b> <b>Experiment</b>	<b>R-K4</b>	<b>ELM</b>
EX1	0.04292	0.01874
EX2	0.04176	0.01739

Table 2. The root mean square error of forecast field are generated by ELM method at 300th, 500th, 600th, 800th and 1000th time steps

<b>RMS</b> <b>Experiment</b>	<b>300</b>	<b>500</b>	<b>600</b>	<b>800</b>	<b>1000</b>
EX1	1.7098e-09	4.5275e-09	2.1309e-09	5.5320e-07	1.9536e-05
EX2	5.3854e-07	1.0714e-06	8.9152e-05	0.0017	0.7396

Table 3. the correlation coefficients of variables x, y and z between forecast values and test data

<b>test</b> <b>simulated</b>	<i>x</i>	<i>y</i>	<i>z</i>
<i>x</i>	0.999897	0.999836	0.985152
<i>y</i>	0.999907	0.999886	0.984076
<i>z</i>	0.984562	0.983415	0.999980

As is known to all, correlation coefficient is a measure of the dependence between variables, which indicates the degree of the fitting of a model. Finally, we verify the correlation coefficients of variables  $x$ ,  $y$  and  $z$  between the forecast values and the test data. Table 3 shows the degree of fitting for each variable is very high, which also shows the strong correlations among the variables in Lorenz63 system. The above results demonstrate the superiority of ELM algorithm in aspects of simulating atmospheric nonlinear chaotic systems.

## Conclusion

This paper firstly introduces the ELM algorithm and it can be successfully used as a novel idea for solving the nonlinear differential equations in the area of atmosphere. The ELM technique was tested in Lorenz63 nonlinear chaotic system and presents a faster calculation speed and optimistic calculation accuracy in the numerical experiments except when the data noise is not really big. We will seek to apply the ELM algorithm in real atmospheric models in the future.

## Acknowledgement

Project supported by the Natural Science Foundation of China (Grant Nos. 41375113 and 41305101).

## References

- [1] T.S. Liu. The research and application on BP neural network improvement. Ph.D. dissertation, Northeast Agricultural University, Harbin, China, 2011.
- [2] G.B. Huang, Q.Y. Zhu, S. Cheekheong. Extreme learning machine: Theory and applications. *Neurocomputing*, 2006 70(1/2/3) 489-501.
- [3] J. Wang, H.Y. Bi. A New Extreme Learning Machine Optimized by PSO. *J. Zhengzhou Univ.(Nat. Sci. Ed.)*, 2013 45(1) 100-104.
- [4] P.D. Xie, X.W. Liu, J.P. Yin, Y.J. Wang. Absent extreme learning machine algorithm with application to packed executable identification. *Neural Computing and Applications*, 2016 27(1) 93-100.
- [5] W.W. Zong, G.B. Huang. Face recognition based on extreme learning machine. *Neurocomputing*, 2011 74(16) 2541-2551.
- [6] W.Y. Deng, L. Chen. Color image water marking using regularized extreme learning machine. *Neural Network World*, 2010 20(3) 317-330.
- [7] X.W. Liu, L. Wang, G.B. Huang, J. Zhang, J.P. Yin. Multiple Kernel extreme learning machine. *Neurocomputing*, 2015 149(A) 253-264.
- [8] X.Q. Cao, Q.B. Huang, B.N. Liu, M.B. Zhu and Y. Yu. A New Data Assimilation Method Based on Dual-number Theory. *Acta Phys. Sin.*, 2015 64(13) 1305021-13050212
- [9] W.J. Chen, C.L. Huang, H.F. Shen. Comparison of Methods for Simultaneous States and Parameters Estimation based on Lorenz-63 Model. *Remote Sensing Technology and Application*, 2015 30(4) 684-693.
- [10] E.N. Lorenz. Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 1963 20(2) 130-141.
- [11] J.D. Gao, J.F. Chou. The effects of the model sensitivity to initial condition upon the variational four-dimensional assimilation – The study based on Lorenz model. *Acta. Meteorologica Sinica.*, 1995 53(4) 471-479.