Proceedings of the 2004 Winter Simulation Conference
R .G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, eds.

ADAPTIVE WAVELET NEURAL NETWORK FOR PREDICTION OF HOURLY NOX AND NO2 CONCENTRATIONS

Zhiguo Zhang
Ye San

P.O. Box 126, Harbin Institution of Technology,
Simulation Center, Harbin Institute of Technology
Harbin, 150001 CHINA

ABSTRACT

Adaptive neural network is a powerful tool for prediction of air pollution abatement scenarios. But it is often difficult to avoid overfit during the training of adaptive neural network. In this paper, based on the wavelet theory, a new algorithm is proposed to improve the generalization of adaptive neural network during on-line learning. The new algorithm trains adaptive wavelet neural network to model hourly NO_x and NO_2 concentrations of variance of emission sources. Results show that the new algorithm improves the generalization and the convergence velocity of adaptive wavelet neural network during on-line learning. The simulations also illustrate that adaptive wavelet neural network is capable of resolving variance of emission sources.

1 INTRODUCTION

Neural network can model air pollution with more advantages compared with general statistical methods because air dynamics encompasses multiple seasonality, long memory and heteroscedesticity (Alessandro Fassò, 2002a; 2002b). These advantages include: the more flexibility than general statistical methods (Marija Božnar1997a; 1997b; Abdul-Wahab et al. 2002), the ability to make efficient use of proxy data when the optimum predictor variables are unavailable and the better predictions than those given by other general models, e.g. multivariate regression models (Asha B. Chelani, 2002), statistical linear models (Gardner and Dorling1999) and even the deterministic modeling system (Jaakko, 2003). But, the main advantages of NN are that emission factors can be replaced by time of day inputs without any detrimental effects (Gardner and Dorling1998; 1999) and air pollution concentration can be predicted with time series and basic meteorological variables(Claudio, et al.2001; Kolehmainen et al.2001). This enables the models to be easily constructed, but also intensify the disadvantage of NN models, which are not applicable for evaluating air pollution abatement scenarios, especially variance of emission source (Garndner, et al., 1999; Jaakko, et al., 2003; Weizhen Lu, et al., 2003).

Adaptive NN is one solution to this problem (Wenjian Wang, et al., 2003), but it is often difficult to avoid overfit with the methods used in static NN, e.g. division of the data into several sets in order to, respectively, train and test NN (Marija, et al, 1997; Gardner, et al.1998, 1999; Claudio, et al.2001). RBF networks are often used as ANN in order to improve the learning efficiency. T.Poggio and F.Girosi (1990a, 1990b) analyze various networks architectures for their approximation abilities and point out that RBF networks possess the property of best approximation. These advantages are further strengthened with the introduction of wavelet into neural network (Qianhua Zhang, et al,1992; Jun Zhang, et al, 1995). The wavelet neural network (WNN) is considered as a kind of RBF networks(Jun Zhang, et al, 1995) and possesses more advantages than the general networks(Qinghua Zhang, et al,1997; Licheng Jiao, et al, 2001; Jian-xin Xu, et al, 2001), such as faster convergence, avoiding local minimum, easy decision and adaptation of structure, so it is soon used in online identification (Christophe, 1997; Andreas , et al, 1998 ; N. Sureshbabu , et al, 1999). Many works (Jian-xin Xu, et al, 2001; Christophe, et al, 1997; N. Sureshbabu , et al, 1999; Robert, et al, 1995; Jay A. Farrell, et al, 1996a, 1996b; Jinhua Xu ,2002) dedicate themselves to the adjustment of the weight and structure of WNN.

This work takes the adaptive wavelet neural network (AWNN) as a tool for prediction of NO_x and NO_2 concentrations, and focuses on avoiding overfit during the training of AWNN. This work is divided into two parts. The first part, consisting of section 2 and 3, proposes the new algorithm. Section 2 discusses the new algorithm of one dimension. In this section, three theorems are introduced after discussion of localization of the energy of WNN in frequency neighborhood. These theorems are dedicated to verify the uniqueness and convergence of the approximator whose support of Fourier transform is in $[-\frac{\pi}{T}, \frac{\pi}{T}]$. And then, the new algorithm is proposed based on the conclusions of three theorems. Section 3 extends the new algorithm to higher dimensions. Application of AWNN for prediction of NO_x and NO_2 is stated in the second part consisting of Section 4 and 5.

2 ANALYSIS OF WAVELET NETWORK

We will use standard notation throughout: $<\cdot,\cdot>$ represents the L_n^2 inner product on maps and Euclidean inner product on vectors, $\|\cdot\|$ represents the associated norms and the induced norms, and $|\cdot|$ represents the absolute value of any real number. On vector spaces, the symbol $+$ denotes a sum, $\dot{+}$ denotes a direct sum, $\otimes$ represents the tensor products on maps and Kronecker product on matrix, R^n is the vector space of all real n-tuples $x=(x_1,\cdots,x_n)$, and $Z=\{\cdots-2,-1,0,1,2\cdots\}$.

2.1 Brief Introduction of WNN

Practically every known wavelet can be constructed with a multiresolution ladder. In this work, we restrict attention to only those wavelet bases associated with an MRA (multiresolution analysis). An MRA consists of a sequence of successive approximation closed subspaces v_j with the following properties (1)—(4):

$$\cdots\subset V_{-1}\subset V_0\subset V_1\subset\cdots$$

$$f(x)\in V_0 \Rightarrow f(x-n)\in V_0 \quad n\in Z \tag{1}$$

$$f(x)\in V_j \Leftrightarrow f(2x)\in V_{j+1},\ j\in Z \tag{2}$$

$$\bigcap_{j\in Z} V_j=\{0\}, close_{L^2}\left(\bigcup_{j\in Z} V_j\right)=L^2(R). \tag{3}$$

There exists a function $\phi\in L^2$ (called the scaling function or father wavelet), with $\phi_{j,n}=2^{j/2}\phi(2^j x-n)$, such that $\{\phi_{0,n};n\in Z\}$ is a basis for V_0 . (4)

Whenever a collection of such subspaces exist, there exists a wavelet basis $\{\psi_{j,k},\ j,k\in Z\}$, which can be constructed explicitly from the scaling function (Mallat, 2003).

Indeed, we can write

$$V_{j+1}=V_j\dot{+}W_j, \tag{5}$$

where W_j is generated by the basis functions $\{\psi_{j,k},\ k\in Z\}$. It then follows that $L^2(R)=\cdots W_{-1}\dot{+}W_0\dot{+}W_1\dot{+}\cdots$. That is, the wavelet basis generates a decomposition of the L^2 space. This means that any function $f\in L^2(R)$ can be uniformly approximated using a wavelet series $f(x)=\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\langle f,\tilde{\psi}_{j,k}\rangle\psi_{j,k}(x)$,where $\tilde{\psi}_{j,k}(x)$ is dual wavelet of $\psi_{j,k}(x)$.

The above properties indicate that any function $f(x)\in L^2$ can be written as a unique linear combination of wavelets of different resolutions. That is, we can write $f(x)=\cdots+g_{-1}(x)+g_0(x)+g_1(x)+\cdots$, where $g_j(x)\in W_j$ are unique. Because $V_j=W_j\dot{+}W_{j-1}\dot{+}\ldots$ and spaces V_j defined in (2) ~ (5) can be generated by the father wavelet $\phi(x)\in L^2$, there exists

$$\tilde{f}(x)=\sum_{-\infty}^{\infty}a_{J,k}\phi_{J,k}(x)+\sum_{j\geq J}\sum_{-\infty}^{\infty}c_{jk}\psi_{jk} \tag{6}$$

such that $\|f(x)-\tilde{f}(x)\|\to 0$ with j converges to infinite. Christophe (1997) also shows that $\tilde{f}(x)=\sum_{-\infty}^{\infty}a_{J,k}\phi_{J,k}(x)+\sum_{j\geq J}\sum_{-\infty}^{\infty}c_{jk}\psi_{jk}$ is equal to

$$\tilde{f}(x)=\sum_k c_k\phi_{j,k}(x)=\Phi C. \tag{7}$$

(7) is just the approximator of WNN with the structure similar to that of Jun Zhang's work in 1995.

2.2 The Frequency Neighborhood of WNN

Wavelet is the function f whose energy is well localized in time and whose Fourier transform $\hat{f}$ has an energy concentrated in a small frequency neighborhood. So, for any given small number $\varepsilon_w>0$, there exist a_w and b_w such that

$$\begin{aligned}\int_{-\infty}^{\infty}|\hat{f}(w)|^2\,dw&=\int_{a_w}^{b_w}|\hat{f}(w)|^2\,dw+\varepsilon_{w1}+\varepsilon_{w2}\\&\leq\int_{a_w}^{b_w}|\hat{f}(w)|^2\,dw+\varepsilon_w,\end{aligned} \tag{8}$$

where $\varepsilon_{w1} = \int_{-\infty}^{a_w} |\hat{f}(w)|^2 dw$ and $\varepsilon_{w2} = \int_{b_w}^{+\infty} |\hat{f}(w)|^2 \, dw$.

Let

$$g_j(t) = \sum_k c_k \psi_{j,k}(x) \tag{9}$$

where $\psi(x)$ is a mother wavelet with $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x + k)$. Then

$$G_j(w) = \int_{-\infty}^{\infty} g_j(t)\exp(-iwt)dt = 2^{-j/2}\hat{\psi}(\tfrac{w}{2^j})\sum_k c_k \exp(\tfrac{kwi}{2^j})$$

and $\int_{-\infty}^{\infty} |g_j(t)|^2 dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |G(w)|^2 dw = \frac{1}{2\pi}(\int_{2^j a_w}^{2^j b_w} |G(w)|^2 dw$

$$+\int_{2^j b_w}^{+\infty} |2^{-j/2}\hat{\psi}(\tfrac{w}{2^j})|^2 \cdot |\sum_k c_k \exp(\tfrac{kwi}{2^j})|^2 \, dw + \int_{-\infty}^{2^j a_w} |2^{-j/2}\hat{\psi}(\tfrac{w}{2^j})|^2 \cdot |\sum_k c_k \exp(\tfrac{kwi}{2^j})|^2 \, dw) \le \tfrac{1}{2\pi}(\int_{2^j a_w}^{2^j b_w} |G(w)|^2 dw + (\sum_k |c_k|)^2(\varepsilon_{w1} + \varepsilon_{w2})). \tag{10}$$

(10) implies that the energy of $g_j(t)$ concentrates in the frequency neighborhood $[2^j a_w, 2^j b_w]$ if there exists constant M such that $(\sum_k |c_k|)^2 < M$ since ε_{w1} and ε_{w2} are small enough.

Since the father wavelet is an aggregation of wavelet at scales larger than 1 and $V_j = W_j \dot{+} W_{j-1} \dot{+} \cdots$ (Mallat,2003), the energy of WNN concentrates in the interval $\bigcup_{-\infty}^{j}[2^j a_w, 2^j b_w]$. Because of symmetric Fourier transform of real function, the energy of WNN concentrates in the frequency band $[-2^j b_w, 2^j b_w]$ when the real father wavelet is used as active function; therefore, we can think the Fourier transform of $\tilde{f}(x)$ satisfies $\mathrm{supp}(\tilde{f}(w)) \subseteq [-2^j b_w, 2^j b_w]$.

2.3 Prevention of Overfit

Let $y = \sum_k e_k^2$ be cost function. Then, WNN may overfit only if the weights of WNN converge to different points with cost function converging to minimum. This section verifies that overfit of WNN is avoided only if the input weights are chosen correctly.

Theorem 1 *Assume* $\sum |f(kT)| < +\infty$, *then there exists a unique function* $\tilde{f}(x)$ *with its Fourier transform* $\tilde{f}(w)$ *such that* $\mathrm{supp}(\tilde{f}(w)) \subseteq [-\frac{\pi}{T}, \frac{\pi}{T}]$ *and* $\tilde{f}(kT) = f(kT)$.

Proof Let

$$s(w) = \begin{cases} 1 & w \in [-\frac{\pi}{T}, \frac{\pi}{T}] \\ 0 & w \notin [-\frac{\pi}{T}, \frac{\pi}{T}] \end{cases} \tag{11}$$

There exist $\tilde{f}(w)$ and $\tilde{f}(x)$ such that $\tilde{f}(w) = T(\sum_k f(kT)\exp(-iwkT)) \cdot s(w)$ and $\tilde{f}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}(w) \exp(iwx)dw$ since $|\tilde{f}(w)| = T|\sum_k f(kT)\exp(-iwKT)| \cdot |s(w)| \le Ts(w)\sum_k |f(kT)| < +\infty$ and $|\tilde{f}(x)| = |\frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}(w) \exp(iwx)dw| < \frac{1}{2\pi}\int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} |\tilde{f}(w)| \, dw \le \sum_k |f(kT)| < +\infty$.

Since $\{\exp(iwkT), k \in Z\}$ is a set of orthogonal basis of $w \in [-\frac{\pi}{T}, \frac{\pi}{T}]$, then $\tilde{f}(lT) = \frac{T}{2\pi}\int_{-\frac{\pi}{T}}^{\frac{\pi}{T}}(\sum_k f(kT)\exp(-iwkT)) \exp(iwlT)dw = \frac{T}{2\pi}\int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} \sum_k f(kT)\exp(iw(lT - kT))dw = f(lT)$.

Assume there are two functions $\tilde{f}(x)$ and $\tilde{g}(x)$ such that $\tilde{f}(kT) = \tilde{g}(kT) = f(kT)$, $\mathrm{supp}(\tilde{f}(w)) \subseteq [-\frac{\pi}{T}, \frac{\pi}{T}]$, and $\mathrm{supp}(\tilde{g}(w)) \subseteq [-\frac{\pi}{T}, \frac{\pi}{T}]$.

Then, $f(kT) - g(kT) = \frac{1}{2\pi}\int_{-\frac{\pi}{T}}^{\frac{\pi}{T}}(f(w) - g(w)) \exp(iwkT) \, dw = 0$, so $f(w) = g(w)$. □

By theorem 1, there is a unique function $\hat{f}(x)$ such that $\mathrm{supp}(\hat{f}(w)) \subseteq [-\frac{\pi}{T}, \frac{\pi}{T}]$ and $\hat{f}(kT) = f(kT)$. suppose that $f(kT)$ is produced by $\hat{f}(x)$. Then $\hat{f}(x)$ can be approximated by WNN $\tilde{f}(x)$ for $\hat{f}(x) \in L^2(R)$. So, for any given small number $\varepsilon > 0$, there exists $\tilde{f}(x)$ such that $2\pi \|\tilde{f}(x) - \hat{f}(x)\|^2 = \|\tilde{f}(w) - \hat{f}(w)\|^2 < \varepsilon$. Since $\|\tilde{f}(w) - \hat{f}(w)\|^2 = \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} |\tilde{f}(w) - \hat{f}(w)|^2 \, dw + F_1 + F_2$

where $F_1=\int_{-\infty}^{-\frac{\pi}{T}}|\tilde{f}(w)|^2\,dw$, $F_2=\int_{\frac{\pi}{T}}^{+\infty}|\tilde{f}(w)|^2\,dw$. Then $0<(F_1+F_2)<\varepsilon-\int_{-\frac{\pi}{T}}^{\frac{\pi}{T}}|\tilde{f}(w)-\hat{f}(w)|^2\,dw$. Since $0\le\int_{-\frac{\pi}{T}}^{\frac{\pi}{T}}|\tilde{f}(w)-\hat{f}(w)|^2\,dw$, then $0<(F_1+F_2)<\varepsilon$, which means that the energy of $\tilde{f}(x)$ concentrates in $[-\frac{\pi}{T},\frac{\pi}{T}]$. So, the sample errors are small enough if $\mathrm{supp}(\tilde{f}(w))$ is limited to $[-\frac{\pi}{T},\frac{\pi}{T}]$. On the other hand, $\mathrm{supp}(\tilde{f}(w))\subseteq[-\frac{\pi}{T},\frac{\pi}{T}]$ ensures the uniqueness of $\tilde{f}(x)$ with $\tilde{f}(kT)=f(kT)$, which helps WNN avoid local minimum. So suppose that the input weights are chosen to limit the energy of approximator to $[-\frac{\pi}{T},\frac{\pi}{T}]$ prior to training since the frequency band $[-2^j b_w, 2^j b_w]$ of WNN is only decided by input weights.

The training of WNN with fixed input weights can be represented as follows:

$$e(k+1)=f-\Phi C(k+1), \tag{12}$$

$$C(k+1)=C(k)+Ae(k) , \tag{13}$$

where $\Phi=(\varphi_{i,k})_{m\times n}$, $\varphi_{i,k}=\phi_{j,l}(x_i)$, $k=1,\cdots,(\max(l)-\min(l)+1)$.

The course of training is just choice of matrix A to make $\sum_l e_l^2(k+1)<\sum_l e_l^2(k)$. The following theorems ensure the convergence of iteration(12), (13).

Theorem 2 *For any* n $\phi_{j,k}(x)$ *with* $\mathrm{supp}(\phi_{j,k}(w))\subseteq[-\frac{\pi}{T},\frac{\pi}{T}]$ *, let* $B=(b_{i,k})_{n\times n}$*, and* $b_{i,k}=\phi_{j,k}(x_i)$ *where* $|x_{i+1}-x_i|=T$ *and* $x_i\in\cup\mathrm{supp}(\phi_{j,k})$ *, i=1…n, then* $rank(B)=n$.

Proof Suppose $\{\phi_{j,n};n\in Z\}$ is a basis for $V_{j,n}$. Since $\phi_{j,k}(w)\subseteq[-\frac{\pi}{T},\frac{\pi}{T}]$, then by theorem 1, there is a unique set of coefficients $\{c_k\}$ such that

$$f(x_i)=\sum_k c_k\phi_{j,k}(x_i)\text{, i.e. } f=BC , \tag{14}$$

where $f(x)\in V_{j,n}$, $f=[f(x_1)\cdots f(x_n)]^T$ and $C=[c_1\cdots c_2]^T$. Since C is unique for $f(x)\in V_{j,n}$, then $rank(B)=n$. □

By theorem2, any n row vectors of $\Phi_{m\times n}$ are linear independent and, by theorem1, there exists at most one set of c with $\Phi C=f$ for $\mathrm{supp}(\tilde{f}(w))\subseteq[-\frac{\pi}{T},\frac{\pi}{T}]$, so $rank(\Phi_{m\times n})=n$ and $n\le m$. Then it can be verified that iteration of (12) and (13) converges to a fixed point and $\sum_k e_k^2$ reaches the minimum value at the fixed point.

Theorem 3 *Let* $e(k+1)=f-\Phi C(k+1)$ *and* $C(k+1)=C(k)+Ae(k)$ *. If* $rank(\Phi_{m\times n})=n$ *, then there exists the matrix* A *to make the iteration converge to a fixed point.*

Proof (12) is substituted into(13); then $C(k+1)=(I-A\Phi)C(k)+Af$. Since $rank(\Phi_{m\times n})=n$, then $\Phi_{m\times n}$ has the form of QR-Decomposition. Let $\Phi_{m\times n}=QR$, where $Q^TQ=I$ and R is a no-singular and upper triangular matrix. Let $A=\lambda R_{n\times n}^{-1}Q_{n\times m}^T$, then $I-A\Phi=I-\lambda R^{-1}Q^TQR=(1-\lambda)I$. Based on contraction mapping theorem, if $|1-\lambda|<1$, then there exists a unique fixed point $C=R^{-1}Q^Tf$ of iteration convergence.

For $e=f-\Phi C=f-QRC$, let $L=QRC$ and $y=\sum_k e_k^2=\sum(f_k-l_k)^2$. If $f_k=l_k$, then $\frac{\partial y}{\partial l_k}=-2(f_k-l_k)=0$ and $\frac{\partial^2 y}{\partial l_k^2}=2>0$. This indicates that $f=L$, i.e. $C=R^{-1}Q^Tf$, is the minimum point of y, which implies $e=(I-QQ^T)f$ at this point. □

By theorem3, there exists a neighborhood $D_0=|C-C_0|<L$ such that $C\in D_0$ and $(\sum_k|c_k|)^2<M$ during training if the initial values of iteration are limited, where $C_0=R^{-1}Q^Tf$. This satisfies the requirement of the energy of WNN localized in $[-2^j b_w, 2^j b_w]$.

Theorem1 and theorem3 mean that the iteration of (12), (13)converges to a fixed point at which the cost function is small enough only if $\mathrm{supp}(\tilde{f}(w))\subseteq[-\frac{\pi}{T},\frac{\pi}{T}]$. By Shannon theorem, for the approximator with small enough sample errors and the support of Fourier transform in $[-\frac{\pi}{T},\frac{\pi}{T}]$, errors between samples are decided on whether the energy of approximated function is concentrated in the frequency neighborhood $[-\frac{\pi}{T},\frac{\pi}{T}]$. This implies that overfit is avoided only if $\mathrm{supp}(\tilde{f}(w))\subseteq$

$[-\frac{\pi}{T},\frac{\pi}{T}]$ during training. By the analysis of section 2.2, $\mathrm{supp}(\tilde{f}(w))$ is only decided by the input weights, so the new algorithm is that only output weights are trained by errors back propagation, and the input weights are only decided by sample period, but not sample errors. we can choose the correct input weights to make $[-2^j b_w, 2^j b_w] \approx [-\frac{\pi}{T},\frac{\pi}{T}]$.

3 MULTIDIMENSIONAL WAVELET NEURAL NETWORK

Multidimensional father wavelet can be built with one-dimensional father wavelet by tensor product

$$\phi_{j,k_1,\cdots,k_n}(x_1,\cdots,x_n)=\phi_{j,k_1}\times\cdots\times\phi_{j,k_2}. \quad (15)$$

The multidimensional approximator is written as:

$$\tilde{f}(x_1,\cdots,x_n)=\sum_{k_1}\cdots\sum_{k_n}c_{k_1,\cdots,k_n}\phi_{j_1k_1}(x_1)\cdots\phi_{j_nk_n}(x_n). \quad (16)$$

For the special property of tensor product, the approximator $\tilde{f}(x_1,\cdots,x_n)$ can be represented in another form

$$\tilde{f}(x_1,\cdots,x_n)=\left(\Phi_1(x_1)\otimes\cdots\otimes\Phi_n(x_n)\right)C, \quad (17)$$

where $\Phi_i=\left(\phi_{j_ik_i}(x_i)\right)_{1\times m_i}$.

(17) indicates that the good prediction of WNN depends on good properties of every input direction and multidimensional wavelet neural network overfits only if one direction x_i overfits, so the condition of multidimensional wavelet neural network without overfit is $[-2^j b_{wi}, 2^j b_{wi}] \subseteq [-\frac{\pi}{T_i},\frac{\pi}{T_i}]$ in every direction x_i, where T_i is the sample period in the direction x_i.

4 SIMULATION DATA

Hourly NO_x and NO_2 are obtained from the monitoring site of the Bureau of the Environment of Heilong Jiang in Jiamu Si City in 1995 and 1998. The hourly metrological data are obtained for the same period from the Bureau of the Weather of Heilongjiang. About 39000 data are taken to train or test AWNN, for some data are missed or obviously wrong. The meteorological variables used in this work are similar to that used by Gardner(1999). They are Low cloud amount (LOW): oktas; Base of lowest cloud (BASE): synoptic code shown in Table 1; Visibility (VIS): synoptic code shown in Table 2; Dry bulb temperature(DRY): ^0C ; Vapour pressure(VP): mbar; Wind speed(WS): ms^{-1}. Instead of the emission factors, the network is given two additional time of day inputs consisting of the sine and cosine of the time of day normalized between 0 and 24 h(Gardner, 1998).

Table1: Synoptic Code for Reporting Height of Lowest Cloud

Code	Height/(m)
0	0-50
1	50-100
2	100-200
3	200-300
4	300-600
5	600-1000
6	1000-1500
7	1500-2000
8	2000-2500
9	Above 2500m or no clouds

Table 2: Synoptic Code for Reporting Visibility

Code	Visibility
00	Visibility<0.1km
01-50	Visibility over the range 0.1-5.0 km. at 0.1 km intervals e.g. 01=0.1 km. 02=0.2 km and 50=5.0 km
60-80	Visibility over the range 10-30 km. at 1 km intervals e.g. 60=10 km. 70=20 km and 80=30 km
81-88	Visibility over the range 35-37 km. at 5 km intervals e.g. 81=35 km. 82=40 km and 88=70 km
89	>70 km

In order to be used with the adaptive neural networks, all data were normalized into the range -1.0+1.0 except of height of lowest cloud and visibility. This is carried out by determining the maximum and minimum values of each variable over the whole data period and calculating normalized variables using the following formula: $x_{norm}=2\times(\frac{(x-x_{\min})}{(x_{\max}-x_{\min})})-1.0$.

Observations of the height of the low cloud base are not made during times of very poor visibility. In this work missing low cloud base information is substituted by a value of +1.0 which is the maximum value the normalized value can attain and represents a cloud base above 2500m or no cloud.

5 SIMULATION

5.1 Choice of Father Wavelet

The 3th order box spline father wavelet is used as the active function of AWNN for its Fourier transform with linear phase and symmetry, which are suitable for training and decision of b_w in this new algorithm. The Fourier transform of 3th order box spline is $\phi(w) = (\frac{\sin(w/2)}{w/2})^3$. b_w is taken as 4 so that $\int_{4}^{+\infty} |\phi(w)|^2 \, dw = 0.0029$ is small enough. The Synoptic codes of the height of lowest cloud and the visibility are directly used as inputs of neural network. Other input meteorological variables which are normalized into the range -1.0+1.0 are round off to one decimal place. Then, let

$$j_i = \log_2(\frac{\pi}{b_w \times T_i}), \tag{18}$$

where j_i and T_i are the dilation coefficient and sample period corresponding to input variable x_i, respectively. It should be pointed out that j_i is not required to be a integer.

5.2 Result of Simulation

The samples are divided into two sets—training set and checking set. The data in the training set are sampled in 1995; the data sampled in 1998 are in the checking set, which are used to check the adaptive performance of AWNN because two new roads were built near the monitor site in 1997, which led to variance of emission sources. The MAE-mean absolute error is used as performance statistics calculated over the whole year.

Figure 1 and Figure 2 show the static performance of AWNN during January 1995 comparing with adaptive performance of AWNN shown in Figure 3 and Figure 4. We let the model work well in 1995, and then the model is used to predict pollutant concentrations of 1998. MAE of NO_2 and NO_x of AWNN in 1995 is 9.9 and 38.0, respectively, compared to 9.8 and 38.5 presented in the work of Gardner(1999), which means that the static performance of AWNN is similar to that of MLP model, but Gardner (1999)

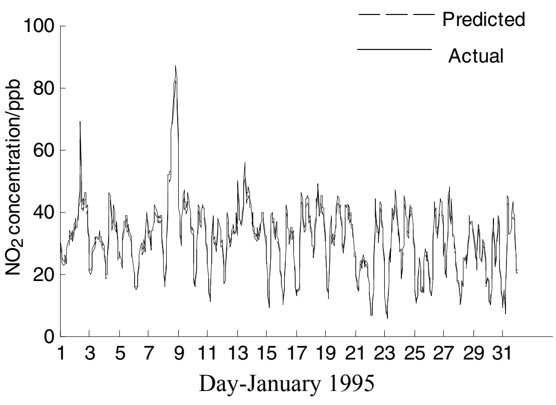

Figure 1: The Predicted NO_2 Concentration Compared to Actual Concentration

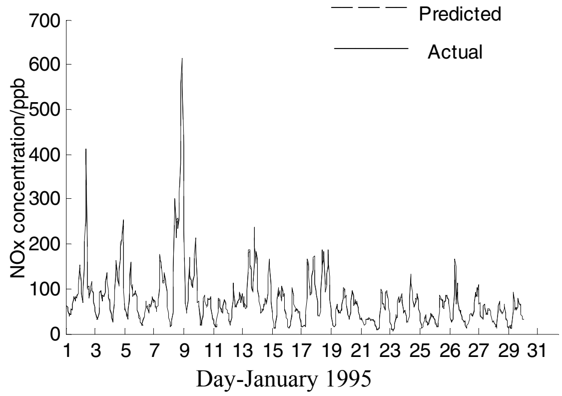

Figure 2: The Predicted NO_x Concentration Compared to Actual Concentration

have used the MLP trained by samples from 1990 to predict the air pollution of 1991and point out that the static model can not deal with the new conditions in 1991, which do not appear in 1990.

Due to these two roads built in1997, the surroundings of minor site varies much with the variance of emission sources. AWNN has to adjust itself by on-line learning in order to catch up with actual line. Figure 3 and Figure 4 show that the accuracy of AWNN is low in the beginning of learning, but the errors gradually decrease after about seven days learning. Figure 3 and Figure 4 also show that the predicted line derives from the actual line between about 10-15 January because the windy weather is very different from that of early month, which forces AWNN to learn again to suit the new weather condition; in the same time, this also indicates AWNN can predict well again after several days learning when it meets a new condition.

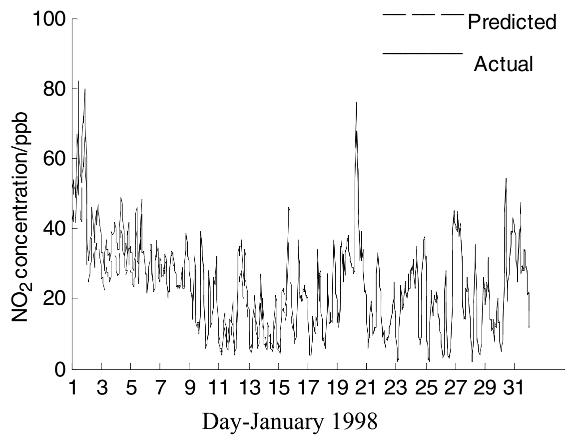

Figure 3: The Predicted NO_2 Concentration for AWNN in the New Condition

Figure 5 and Figure 6 show the performance of AWNN of which the input weights and output weights are trained by error back-propagation. First, the data from 1995 are divided into three partitions— a training, valida-

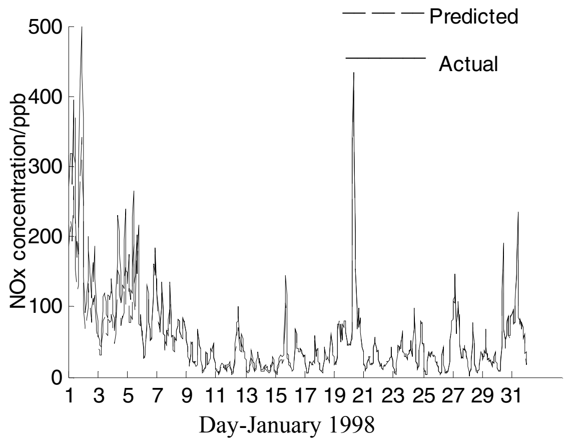

Figure 4: The Predicted NO_x Concentration for AWNN in the New Condition

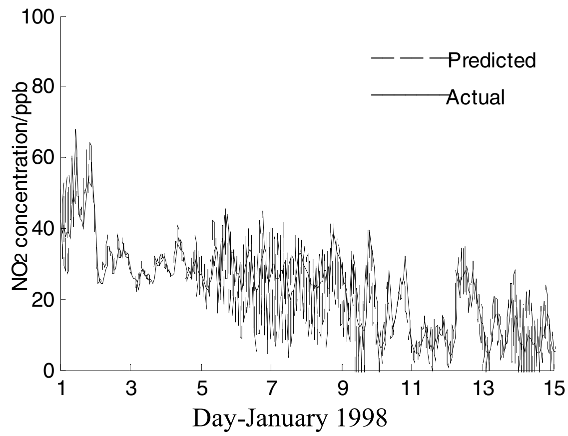

Figure 6: Overfit of AWNN in the New Condition

tion and test sets. The training set forms the bulk of the data. The validation set is used during training in order to check the generalization performance. Training can be stopped when the performance on the validation data reaches a maximum. The test data are the data upon which the final model is tested. Then the model trained by the data from 1995 is used to predict the air pollution of 1998 and, in order to judge the generalization, the middle points between samples from 1998 calculated by model compare with the line crossing the two samples. The simulations show that AWNN trained by error back-propagation is difficult to avoid overfit especially when the initial values of input weight are big during on-line learning, though it works well as a static model in 1995. The error back-propagation can make the sample errors very small (sometimes zero) during the on-line learning, but the errors between predicted line and actual line often become large with decrease of sample errors. The new algorithm ensures generalization of AWNN, so AWNN based on the new algorithm does not overfit during training. In the same time, the velocity of convergence is improved,

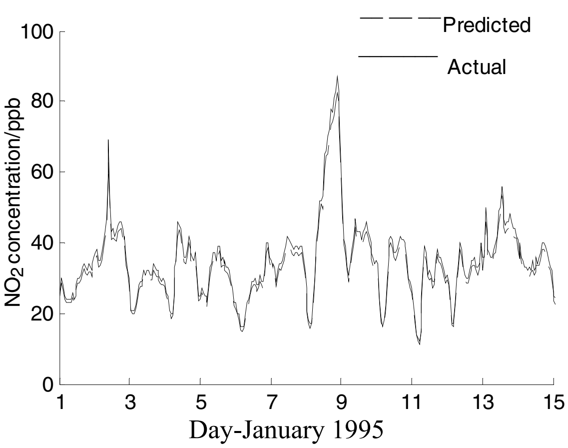

Figure 5: AWNN Trained by Error Back-Propagation

which help WNN suit the new conditions faster for the new algorithm only need adjustment of output weights instead of both input and output weights.

6 CONCLUSIONS

Adaptive neural network is a powerful tool for prediction of air pollution abatement scenarios. But, the same methods used by static NN often can not ensure generalization of adaptive neural network. To solve this problem, a new algorithm is proposed based on Fourier transform and wavelet theory. Through analysis of theory, this work proposes that choice of correct input weights can ensure generalization of AWNN without any other measures, so the new algorithm only trains the output weights, and the input weights are decided by sample period. This ensures simplicity of the structure and the algorithm of AWNN, which improves the efficiency of on-line learning. On the other hand, the new algorithm only trains output weights, so the velocity of convergence is faster than that of general NN.

This new algorithm is applied to AWNN for prediction of NO_2 and NO_x concentration. Results show that WNN can work as well as other static NN when the condition of prediction is stable. Results also show the new algorithm ensures good generalization of AWNN during online learning when the condition of prediction is unstable, but the classic algorithm can not ensure the generalization during the same course.

REFERENCES

Abdul-Wahab, S. A. and S. M. Al-Alawil. 2002. Assessment and prediction of tropospheric ozone concentration levels using artificial neural network. *Environmental Modelling and Software* 17 (3): 219-228.

Alessandro Fassò, and Ilia Negri. 2002a. Multi-step forecasting for nonlinear models of high frequency ground ozone data: a Monte Carlo approach. *Environmetrics* 13(4): 365-378.

Alessandro Fassò and Ilia Negri. 2002b.Non-linear statistical modeling of high frequency ground ozone data. *Environmentrics* 13 (3):225-241.

Andreas Ikonomopoulos, and Akira Endou. 1998. Wavelet decomposition and radial basis function networks for system monitoring. *IEEE Transactions on Nuclear Science* 45 (5): 2293 – 2301.

Asha B.Chelani, and C.V. Chalapati Rao. 2002. Prediction of sulphur dioxide concentration using artificial neural networks. *Environmental Modelling and Software* 17 (2):161-168.

Christophe P. Bernard and Jean-Jacques E. Slotine. 1997 . Adaptive control with multiresolution bases. In *Proceedings of the 1997 36th IEEE Conference on Decision and Control,* 3884 – 3889. San Diego. California USA

Claudio Silva, Patricio Pérez, and Alex Trier. 2001. Statistical modleling and prediction of atmospheric pollution by particulate material: two nonparametric approaches. *Environmentrics* 12 (2):147-159.

Farrell, J. A. 1996a. Approximators characteristics and their effect on training misbehavior in passive learning. In *Intelligent Control, 1996. Proceedings of the 1996 IEEE International Symposium,* 181 – 187. Dearborn, MI USA.

Farrell, J. A. 1996b. Motivations for local approximators in passive learning control. *Journal* of *Intelligent System Control* 1(2): 195-210.

Gardner, M. W., and S. R. Dorling, 1998. Artificial neural networks(the multilayer perception)-a review of applications in the atmospheric sciences. *Atmospheric Environment*. 32 (14/15): 2627-2636.

Gardner, M. W., and S. R. Dorling. 1999. Neural network modeling and prediction of hourly NO_x and NO_2 concentrations in urban air in London. *Atmospheric Environment* 33 (5):709-719.

Jun Zhang, G.G. Walter, and Y.Miao. 1995. Wavelet neural networks for function learning. *IEEE Transaction Signal Processing* 43 (6): 1485-1496.

Jian-xin Xu, and Ying Tan. 2001. Nonlinear adaptive wavelet control using constructive wavelet networks. In *American Control Conference, 2001. Proceedings of the 2001* , 25-27. Arlington, VA USA

Jinhua Xu, and Daniel W.C. Ho. 2002. A basis selection algorithm for wavelet neural networks. *Neurocomputing* . 48 (1/4):681-689.

Jaakko Kukkonen, and Leena Portanen. 2003. Extensive evaluation of neural network models for the prediction of NO_2 and PM_{10} concentrations, compared with a deterministic modelling system and measurements in contral Helsinki. *Atmospheric Environment* 37 (32): 4539-4550.

Kolehmainen, M., H. Martikainen, and J. Ruuskanen. 2001. Neural networks and periodic components used in air quality forecasting. *Atmospheric Environment* 35 (5): 815-825.

Kolehmainen, M., H. Martikainen, T. Hiltunen, and J. Ruuskanen. 2002. Forecasting air quality using hybrid neural network modeling. *Environmental Monitoring and Assessment* 65 (1/2): 277-286.

Licheng Jiao, Jin Pan and Yangwang Fang. 2001. Multiwavelet neural network and its approximation properties. *IEEE Transaction Neural Networks* 12 (5): 1060 - 1066.

Marija Božnar. 1997a. Pattern selection strategies for a neural network-based short term air pollution prediction model. In *Intelligent Information Systems, 1997. IIS '97. Proceedings,* 340 – 344. Grand Bahama Island Bahamas.

Marija Božnar. 1997b. Perceptron neural network-based model predicts air pollution**.** In *Intelligent Information Systems, 1997. IIS '97. Proceedings,* 345-349. Grand Bahama Island Bahamas

Mallat. S. 2003. *A Wavelet Tour of Signal Processing.* 2nd ed. Beijing: China Machine Press.

Patricia Perez and Jorge Reyes. 2001. Prediction of particulate air pollution using neural techniques. *Neural Computing and Applications* 10 (2):165-171.

Poggio, T. and F. Girosi. 1990a. Networks for approximation and learning. *Proceedings of the IEEE*. 78 (9): 1481 – 1497.

Poggio, T. and F. Girosi. 1990b. Regularization algorithms for learning that are equivalent to multilayer networks. *Science*. 247: 978-982.

Qianhua Zhang and A.Benvenist. 1992. Wavelet network. *IEEE Transaction Signal Processing* 13 (6): 889-898.

Qinghua Zhang. 1997. Using wavelet network in nonparameters estimation. *IEEE Transaction Neural Networks* 8 (2): 227~236.

Robert M.Sanner and Jean-Jacques E.Slotine. 1995. Structurally dynamic wavelet networks for the adaptive control of uncertain robotic systems. In *Proceedings of the IEEE Conference on Decision and Control*, 2460-2467. New Orleans, LA.

Sureshbabu, N. and J. A. Farrell. 1999. Wavelet-Based System Identification for Nonlinear Control. *IEEE Transactions on Automatic Control* 44 (2): 412-417.

Wenjian Wang, Weizhen Lu, Xiekang Wang and Andrew Y.T. Leung, Prediction of maximum daily ozeone level using combined neural network and statistical characterisitics. *Environment International* 29 (5):555-562.

Weizhen Lu, Huiyuan Fan and Wenjian Wang, et. 2002. Prediction of respirable suspended particulate level in Hong Kong downtown area using principal component analysis and artificial neural networks. In *Proceedings of the 4th World Congress on Intelligent Control and Automation,* 49-53. Shanghai, P.R.China.

AUTHOR BIOGRAPHIES

ZHIGUO ZHANG received his B.S. in Automation and Control Engineering from JiNan University(2000); and

received his M.S. in Control Science and Engineer from Harbin Institute of Technology(2002). He is studying in Harbin Institute of Technology for Ph. D. since 2002. His research interests include developing models using neural network and wavelet. His e-mail address is <zhangzg2002@hit.edu.cn>.

SAN YE is a Professor in Simulation Center at Harbin Institute of Technology. He graduated from Harbin Institute of Technology in 1986. He was a Lecturer from 1986 to 1991 and was an Associate Professor from 1991 to 1996. He has designed four simulation machines of power stations. His research interests include the design, optimization and simulation of complex system using neural network and fuzzy models. His e-mail address is <sanye@hit.edu.cn>.