Lattice Dynamical Wavelet Neural Networks Implemented Using Particle Swarm Optimization for Spatio-Temporal System Identification

Hua-Liang Wei, Stephen A. Billings, Yifan Zhao and Lingzhong Guo

***Abstract*—By combining an efficient wavelet representation with a coupled map lattice model, a new family of adaptive wavelet neural networks, called lattice dynamical wavelet neural networks (LDWNN), is introduced for spatio-temporal system identification. A new orthogonal projection pursuit (OPP) method, coupled with a particle swarm optimization (PSO) algorithm, is proposed for augmenting the proposed network. A novel two-stage hybrid training scheme is developed for constructing a parsimonious network model. In the first stage, by applying the orthogonal projection pursuit algorithm, significant wavelet-neurons are adaptively and successively recruited into the network, where adjustable parameters of the associated wavelet-neurons are optimized using a particle swarm optimizer. The resultant network model, obtained in the first stage, however, may be redundant. In the second stage, an orthogonal least squares algorithm is then applied to refine and improve the initially trained network by removing redundant wavelet-neurons from the network. An example for a real spatio-temporal system identification problem is presented to demonstrate the performance of the proposed new modeling framework.**

***Index Terms*—coupled map lattice, neural networks, particle swarm optimization, spatio-temporal systems, wavelets.**

I. Introduction

SPATIO-TEMPORAL systems are complex systems where the system states evolve spatially as well as temporally. The dynamics of such a system at a specific spatial location and at the present time instant are dependent on the dynamics of the system at other spatial locations and at previous times. Spatio-temporal systems can be viewed as an extension of classical nonlinear dynamical systems where the dynamics at the present time instant are determined by only the values of the associated input and state variables at previous times. The evolution trajectory of a traditional nonlinear dynamic system at each time instant consists of only one isolated point or just a few points, while the evolution trajectory of a spatio-temporal dynamical system at each specific time instant is a snapshot, pattern or image, formed by a collective of a great number of interacting points. Thus, the evolution trajectory of a spatio-temporal system can be viewed as a set of moving objectives (snapshots, patterns, images or pictures). Spatio-temporal phenomena are widely found in biology, chemistry, ecology, geography, medicine, physics, and sociology. In the literature, several efficient mathematical representations including the well known cellular automata (CA) [1], coupled map lattices (CML) [2], and cellular neural networks (CNN) [3] have been introduced to described spatio-temporal systems.

Whilst the forward problem of spatio-temporal systems has been extensively studied, with an assumption that the associated models are known to describe some specific dynamics, the inverse problem, which is concerned with finding models based on given observations for structure-unknown spatio-temporal systems, has received relatively little attention and relatively few results have been achieved. Identification plays an important role for solving the inverse problem for spatio-temporal systems, where the structure of analytical models is not available.

The central objective of this study is to introduce a new family of adaptive wavelet neural networks, where wavelet transforms will be incorporated into a specific type of CML model. This wavelet-based coupled map lattice model will be referred to as the lattice dynamical wavelet neural network (LDWNN). The construction procedure of the new network model is composed of two stages. At the first stage, linear combinations of a number of wavelet functions are chosen as the building blocks to form the initial candidate wavelet neurons. Inspired by the conventional projection pursuit regression method [4], a new orthogonal projection pursuit (OPP), implemented by a particle swarm optimization (PSO) algorithm, is introduced to augment the network by recruiting a number of optimized wavelet neurons in a stepwise manner. The OPP learning algorithm, similar to the conventional projection pursuit regression, may produce a redundant model. Thus, the objective of the second stage is to remove redundancy from the initially trained network, to produce a parsimonious representation. To achieve this aim, an orthogonal least squares learning algorithm [5]-[7] is applied to refine and improve the initially obtained network by removing potential redundant wavelet neurons from the network.

As will be noted from the proposed learning algorithm, the training procedure for the new network model does not need any pre-specified dictionary, as required by existing wavelet-based CML models [8]. Also, as will be seen later from the illustrative example, by combining the PSO based nonlinear

Manuscript received June 29, 2007. This work was supported by the Engineering and Physical Sciences Research Council (EPSRC), U.K. The authors gratefully acknowledge the help from Dr A. F. Routh, University of Sheffield, who supervised the B-Z experiments.

The authors are with the Department of Automatic Control and Systems Engineering, the Univerity of Sheffield, Sheffield, S1 3JD, UK. E-mail: {w.hualiang, s.billings, y.zhao, l.guo}@sheffield.ac.uk.

OPP training scheme with an effective forward orthogonal regression algorithm, the resultant wavelet network model can provide good generalization performance. Moreover, one feature of the new wavelet network, produced by the two-stage hybrid learning algorithm, is that now the resultant model is transparent to model users; involved wavelet neurons are ranked according to the capability of each neuron in representing the total variance in the system output signal. This is desirable for many application cases where physical insight on the individual variables and associated wavelet neurons are of interest. In summary, the objective of this study is to present an effective method that can produce sparse and transparent network models using wavelet basis functions, which can adaptively capture the variations of the associated dynamics. The proposed wavelet network is nearly self-implemented, that is, all within-network parameters can automatically be adjusted by the proposed algorithms. This is desirable for any structure-unknown or black-box modeling problems.

II. The Architecture of the New LDWNN

Coupled map lattice (CML) models are a class of dynamical models, with discrete time and discrete space, but with continuous state variables [3]. Take the 2-D CML model, involving the nearest-neighbour cell coupling on a squared lattice with *Moore neighbourhoods*, as an example, this can be expressed as

$$\begin{aligned} s_{i,j}(t) = & \sum_{-r\le p,q\le r} \alpha_{p,q}^{(1)} \Phi_{p,q}^{(1)}(s_{i+p,j+q}(t-1)) \\ & + \sum_{-r\le p,q\le r} \alpha_{p,q}^{(2)} \Phi_{p,q}^{(2)}(s_{i+p,j+q}(t-2)) + \cdots \\ & + \sum_{-r\le p,q\le r} \alpha_{p,q}^{(\tau)} \Phi_{p,q}^{(\tau)}(s_{i+p,j+q}(t-\tau)) \end{aligned} \quad (1)$$

where t=1,2, …, i=1,2, …, I, j=1,2, …, J, $s_{i,j} \in \mathbb{R}$ is the state representing the cell C(i,j), τ is the time lag, $\Phi_{p,q}^{(\tau)}$ are some linear or nonlinear functions, $\alpha_{p,q}^{(\tau)}$ are connecting coefficients, and $r \ge 0$ is referred to as neighborhood radius indicating how many neighborhood cells are involved in the evolution procedure for generating each centre cell $s_{i,j}(t)$ from the past state space. Clearly, if r=0, model (1) will become a pure temporal process. The evolution law for boundary cells often needs to be pre-specified. If both I and J are very large, boundary conditions may not affect the resultant patterns; if, however, one of the two numbers is small, boundary conditions may significantly distort the original patterns. For details about how to set boundary conditions, see [3].

Note that a total of $d = (2r+1)^2$ state variables are involved in the CML model (1). For convenience of description, introduce d single-indexed variables x_k, with k=1,2,…,d, to represent the d involved cells in the neighborhood. Also, let y represent the central cell $C(i,j)$. Then, the objective is to identify, from available data, a d-dimensional model

$$y(t) = f(\mathbf{x}(t)) = f(x_1(t), x_2(t), \cdots, x_d(t)) \quad (2)$$

or, in an explicit form, with respect to the state variables

$$\begin{aligned} s_{i,j}(t) = f(\mathbf{s}(t)) = f(& s_{i-r,j-r}(t-1), \cdots, s_{i-r,j}(t-1), \cdots, s_{i-r,j+r}(t-1), \cdots, \\ & s_{i,j-r}(t-1), \cdots, s_{i,j}(t-1), \cdots, s_{i,j+r}(t-1), \cdots, \\ & s_{i+r,j-r}(t-1), \cdots, s_{i+r,j}(t-1), \cdots, s_{i+r,j+r}(t-1)) \end{aligned} \quad (3)$$

where $\mathbf{x}(t)$ and $\mathbf{s}(t)$ are state vectors formed by the relative state variables.

One of the most commonly used approaches for constructing the high dimensional model (2) or (3) is to approximate the multivariate function f using a set of functions of fewer variables (often univariate)

$$f(\mathbf{x}(t)) = \sum_j w_j g_j(\mathbf{x}(t); \boldsymbol{\theta}_j) \quad (4)$$

where g_j are called the construction functions (hidden units), $\boldsymbol{\theta}_j$ are the associated parameter vectors, and w_j are the weight coefficients that can be assimilated into the parameter vector $\boldsymbol{\theta}_j$, as shown in the next section.

Wavelets [9], due to their inherent property and excellent capability, can be used as the elementary building blocks to represent these construction functions g_j in (4) as below

$$\begin{aligned} g_j(\mathbf{x}; \boldsymbol{\theta}_j) = & c_{1,j} \psi(x_1; a_{1,j}, b_{1,j}) + c_{2,j} \psi(x_2; a_{2,j}, b_{2,j}) + \cdots \\ & + c_{d,j} \psi(x_d; a_{d,j}, b_{d,j}) \end{aligned} \quad (5)$$

where $\psi(x_k; a_{k,j}, b_{k,j}) = \psi(a_{k,j} x_k - b_{k,j})$, with k=1,2, …, d, are wavelet basis functions, and $\boldsymbol{\theta}_j = [a_{1,j}, b_{1,j}, c_{1,j}, \cdots, a_{d,j}, b_{d,j}, c_{d,j}]^T$ are the parameter vectors that need to be optimized.

Assume that a total of m construction functions, $g_1, g_2, \cdots, g_m$, are involved in the network, then equation (4) can be expressed as

$$f(\mathbf{x}(t)) = \sum_{j=1}^{m} w_j g_j(\mathbf{x}(t); \boldsymbol{\theta}_j) = \sum_{j=1}^{m} \sum_{k=1}^{d} \tilde{c}_{k,j} \psi(x_k; a_{k,j}, b_{k,j}) \quad (6)$$

where $\tilde{c}_{k,j} = w_j c_{k,j}$. Now, the remaining key problem is how to construct the wavelet network model (6). Unlike in fixed grid wavelet network models, where a dictionary of candidate basis functions needs to be pre-specified, based on which some search and pruning algorithms are applied to find a set of significant basis functions [10]-[12], this study will consider a type of growing wavelet neural network, where a constructive learning algorithm that can be used to automatically and adaptively augment such a network will be provided.

III. Training The New LDWNN

Many constructive learning algorithms for constructing typical neural networks can be found in the literature [13]-[16].

The projection pursuit regression [4] and some variations [16] are among the class of the most commonly used approaches for augmenting single-hidden-layer neural networks. Inspired by the successful applications of these popular constructive learning algorithms, this study proposes a practical orthogonal projection pursuit (OPP) learning scheme, assisted by a particle swarm optimization (PSO) algorithm. Similar to other popular constructive algorithms, networks produced by the OPP algorithm may be redundant. To remove or reduce redundancy, an orthogonal least squares type learning algorithm [5]-[7] is applied to refine and improve the initially generated network by the OPP+PSO algorithm.

A. The OPP Algorithm aided by PSO for First Stage Network Training

Let $\mathbf{y}=[y(1),...,y(N)]^T \in \mathbb{R}^N$ be the observation vector of the output signal and $\mathbf{x}_k=[x_k(1),x_k(2),\cdots,x_k(N)]^T$ be the observation vector for the kth input variable, with k=1,2, …, d. For a given parameter vector $\boldsymbol{\theta}=[a_1,b_1,c_1,\cdots,a_d,b_d,c_d]^T$, let $\boldsymbol{\psi}_k=[\psi(x_k(1);a_k,b_k),\cdots,\psi(x_k(N);a_k,b_k)]^T$, $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_d]$ and $\mathbf{g}(\mathbf{X};\boldsymbol{\theta})=\sum_{k=1}^{d} c_k\boldsymbol{\psi}_k$.

The OPP algorithm is implemented in a stepwise fashion; at each step a construction vector that minimizes the projection error will be determined. Starting with $\mathbf{r}_0=\mathbf{y}$, find a construction function $\mathbf{g}_1=\mathbf{g}(\mathbf{X};\boldsymbol{\theta}_1)$ such that $\boldsymbol{\theta}_1=\arg\min_{\boldsymbol{\theta}}\{\|\mathbf{r}_0-\mathbf{g}(\mathbf{X};\boldsymbol{\theta})\|^2\}$. The associated residual vector may be defined as $\mathbf{r}_1=\mathbf{r}_0-\mathbf{g}_1$, which can be used as the "fake desired target signal" to produce the second construction vector $\mathbf{g}_2$. However, it should be noted that the coefficient $\boldsymbol{\theta}_1$ is not always identical to the true (theoretical) optimal value $\boldsymbol{\theta}_1^*$, no matter what optimization algorithms are applied. As a consequence, $\mathbf{r}_1=\mathbf{r}_0-\mathbf{g}_1$ may not be orthogonal with the construction vector $\mathbf{g}_1$. To make the associated residual orthogonal with the relevant construction vector, the residual is then defined as $\mathbf{r}_1=\mathbf{r}_0-\alpha_1\mathbf{g}_1$, where $\alpha_1=<\mathbf{r}_1,\mathbf{g}_1>/\|\mathbf{g}_1\|^2$. This procedure may need to be repeated many times. At the nth step, the residual sum of squares, also called the sum of squares error, $\|\mathbf{r}_n\|^2$, can be used to form a criterion to stop the growing procedure. For example, the criterion can be chosen as *error-to-signal ratio*: $\mathrm{ESR}=\|\mathbf{r}_n\|^2/\|\mathbf{y}\|^2$; when ESR becomes smaller than a pre-specified threshold value, the growing procedure can then be terminated.

The objective function in the OPP algorithm is defined as

$$\pi_{n-1}(\boldsymbol{\theta})=\|\mathbf{r}_{n-1}-\mathbf{g}(\mathbf{X};\boldsymbol{\theta})\|^2=\sum_{t=1}^{N}[r_{n-1}(t)-g(\mathbf{x}(t);\boldsymbol{\theta})]^2 \qquad (7)$$

where $\mathbf{x}(t)=[x_1(t),x_2(t),\cdots,x_d(t)]^T$, N is the number of training samples, $\mathbf{X}$ and $\boldsymbol{\theta}$ are defined as in the previous section, and $g(\mathbf{x}(t);\boldsymbol{\theta})=\sum_{k=1}^{d} c_k\psi(x_k(t);a_k,b_k)$

Equation (7) can be solved by using a particle swarm optimization method [17]. Let 'mPSO' be the maximum number of permitted iterations. The optimization procedure can then be terminated when either the iteration index exceeds 'mPSO', or when the parameter to be optimized becomes stable, that is, when $\|\boldsymbol{\theta}(t+1)-\boldsymbol{\theta}(t)\|^2\le\delta$, where δ is a pre-specified small number, say $\delta\le 10^{-5}$.

Note that for each step n in the above OPP algorithm, a PSO algorithm repeatedly runs 10 times, and the coefficients that produce the smallest value for the object function are chosen to be the parameters for the nth step search. By using this multi-time run strategy, the performance of the algorithm including the stability and convergence properties can then be guaranteed. It can be shown that the sequence $\|\mathbf{r}_n\|^2$ is strictly decreasing and positive; thus, by following the method given in [18],[19], it can easily be proved that the residual $\mathbf{r}_n$ is a Cauchy sequence, and as a consequence, $\mathbf{r}_n$ converges to zero. The algorithm is thus convergent.

B. Refine the Network Using the Forward Orthogonal Regression Algorithm

Assume that a total of m construction functions $g_j(\mathbf{x};\boldsymbol{\theta}_j)$, where $\mathbf{x}(t)=[x_1(t),x_2(t),\cdots,x_d(t)]^T$ and j=1,2, …, m, are involved in the network produced at the first stage. It is known that each g_j involves d individual wavelets, thus a total of $M=d\times m$ elementary wavelet neurons are involved in the network. Denote the set of these M wavelets by

$$\Omega=\{\psi_{k,j}:\psi_{k,j}(x_k)=\psi(x_k;a_{k,j},b_{k,j}),(k,j)\in\Gamma\} \qquad (7)$$

where $\Gamma=\{(k,j):k=1,2,\cdots,d;j=1,2,\cdots,m\}$. Note that all the parameters $a_{k,j}$ and $b_{k,j}$ have already been estimated at the first stage.

The objective of this refinement stage is to reselect the most significant wavelet functions from the set Ω, to form a more compact model for given nonlinear identification problems. Let $\mathbf{y}$ and $\mathbf{x}_k$ be defined as in the previous section, and let $\boldsymbol{\psi}_{k,j}=[\psi(x_k(1);a_{k,j},b_{k,j}),\cdots,\psi(x_k(N):a_{k,j},b_{k,j})]^T$, where $(k,j)\in\Gamma$. Also, let

$$\begin{aligned} D&=\{\boldsymbol{\varphi}_{(k-1)m+j}:\boldsymbol{\varphi}_{(k-1)m+j}=\boldsymbol{\psi}_{k,j},(k,j)\in\Gamma\}\\ &=\{\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_2,\cdots,\boldsymbol{\varphi}_M\} \end{aligned} \qquad (8)$$

The network refinement problem amounts to finding, from the dictionary D, a full dimensional subset $D_n=\{\mathbf{p}_1,\cdots,\mathbf{p}_n\}=\{\boldsymbol{\varphi}_{i_1},\cdots,\boldsymbol{\varphi}_{i_n}\}$, where $\boldsymbol{\alpha}_k=\boldsymbol{\varphi}_{i_k}$, $i_k\in\{1,2,\cdots,M\}$ and k=1,2, …,n (generally $n<<M$), so that $\mathbf{y}$ can be satisfactorily approximated using a linear combination of $\mathbf{p}_1,\mathbf{p}_2,\cdots,\mathbf{p}_n$. This can be solved by using the orthogonal least squares algorithm

given in [5]-[7]. Following the suggestion in [20], the penalized ESR criterion below is used for model size determination

$$\mathrm{PESR}_n = \frac{1}{(1-\lambda n/N)^2}(1-\sum_{i=1}^{n}\mathrm{ERR}_i) \tag{9}$$

It is suggested [20] that the adjustable parameter λ be chosen between 5 and 10.

IV. An Example

As an example of a real spatio-temporal system, the Belousov-Zhabotinsky [21] reaction was considered to illustrate the application procedure of the proposed new network modeling procedure. The BZ reaction, as an excitable medium, is an important class of chemical reactions exhibiting a spatio-temporal oscillatory behaviour. As a classical example of non-equilibrium thermodynamics, the BZ reaction provides an interesting chemical model of non-equilibrium biological phenomena, and the modeling and identification of these types of reactions is of extreme interest for theoretical analysis of relevant phenomena.

For a comparison purpose, the well-known Levenberg-Marquardt (LM) nonlinear least squares optimization algorithm [22], was also used to solve the nonlinear optimization problems. In order to fairly and objectively evaluate the proposed method, we used an existing LM algorithm available in Matlab, where finite difference gradients are considered.

A. The Data

By adopting the recipe given by Winfree [23], an experiment resulting in a thin layer BZ reaction was carried out, and a set of images were captured with equal time intervals during the experiment. The sampled images were pre-processed and saved as patterns with a resolution of 480 by 640 pixels. In this example, the LDWNN modeling framework was applied to these sampled images, and the objective is to apply the new network model for the identification of the BZ reaction.

A total of N=2500 data pairs, $\{\mathbf{x}(k), y(k)\}_{k=1,2,\ldots,N}$, were used for the network training, where $y(k)$ represents the values of the relevant central cell at the present time instant t, and $\mathbf{x}(k)=[x_1(k), x_2(k), \cdots, x_9(k)]^T$ represent the observations of the nine involved cells at a squared lattice with the *Moore neighbourhoods*, at the previous time instant t-1. These 2500 data pairs were formed as follows. Firstly, 5 adjacent pattern pairs were randomly chosen from the first 50 sampled patterns. Secondly, 500 data pairs were randomly chosen in each of these 5 pattern pairs.

B. Some Results

The Mexican hat wavelet, defined as $\psi(x)=(1-x^2)e^{-x^2/2}$, was used as the elementary building blocks for constructing the wavelet network model, where the time lag τ (defined in (1)) was set to be one. A total of 100 construction functions of the form (5) were optimized using the OPP+PSO algorithm, and the associated ESR index is shown in Fig. 1. The orthogonal least squares algorithm [5]-[7] was then applied to select significant individual wavelets from the pool that contains 900 individual candidate wavelets of the form $\psi_{k,j}=\psi(x_k; a_{k,j}, b_{k,j})$, with k=1,2, …, 9 and j=1,2, …, 100, and where both the dilation and translation parameters have already been optimized. The penalized error-to-signal ratio (PESR), produced by the orthogonal least squares algorithm suggested that a total of 15 wavelets should be included in the wavelet network model.

For a comparison, the Levenberg-Marquardt (LM) algorithm was also applied in the OPP procedure to solve the relevant nonlinear optimization problems. Similar to the PSO algorithm, we take ten time runs of the LM algorithm at each search step in the OPP procedure; at each time run, the initial values for the unknown parameters to be optimized were chosen in a way similar to that in the OPP algorithm; the corresponding ESR index is shown in Fig. 1. The associated PESR index suggested that the appropriate number of wavelets included in the network should be 17. From Fig. 1, it can be noted that PSO yields a better global convergence than the LM algorithm for the problem here, meaning that to achieve the same approximation accuracy, the number of wavelet basis functions in the PSO produced network is less than that in the LM produced network.

To evaluate the performance of the identified LDWNN model, the short-term predictive capability of model estimated using the OPP-PSO algorithm was inspected. Denote the observation of the image (pattern) measured at the present time instant t by $X(t)$. The k-step ahead predictions, denoted by $\hat{X}(t+k\,|\,t; f, X(t))$, where f represents the given identified model, are the iteratively produced results by the model, on the basis of $X(t)$ but without using information on observations for patterns at any other time instants. As an example, the one-step ahead predictions from the model produced by the OPP+PSO algorithm, on the basis of the measurements at a certain time instants t were calculated, and these are shown in Fig. 2. Clearly, the identified model provides good short-term predictions in the sense that these predictions capture the main features of the observed images.

To measure the performance of the identified wavelet network models, the 2-D spatially normalized mean square error (SNMSE) was considered

$$\mathrm{SNMSE}(t)=\frac{\sum_{i=1}^{I}\sum_{j=1}^{J}|\,s_{i,j}(t)-\hat{s}_{i,j}(t)\,|^2}{\sum_{i=1}^{I}\sum_{j=1}^{J}|\,s_{i,j}(t)-\bar{s}(t)\,|^2} \tag{10}$$

where $s_{i,j}(t)$ represent the observations at the time instant t, $\hat{s}_{i,j}(t)$ represent the corresponding predicted values from a given model, $\bar{s}(t)$ represents the spatial average value at the time instant t, and I and J define the size of the associated pattern. The value of SNMSE for the one-step ahead predictions, relative to the PSO+PSO produced model

consisting of 15 wavelet functions, was calculated to be 0.0787. For the LM algorithm produced model consisting of 17 wavelet functions, the value of SNMSE for the four cases was calculated to be 0.0808.

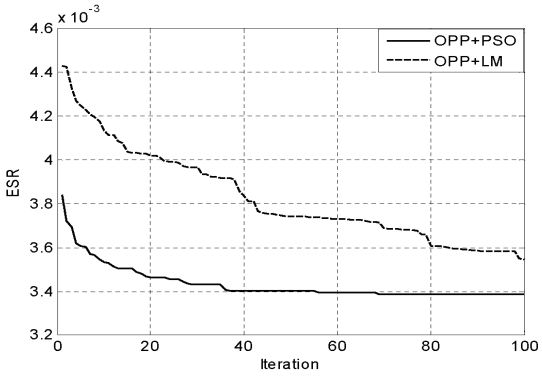

Fig. 1. The error-to-signal ratio (ESR) index, versus the number of the iterations, for the BZ reaction modeling problem. The solid line is for the OPP+PSO algorithm, and the dashed line is for the OPP+LM algorithm.

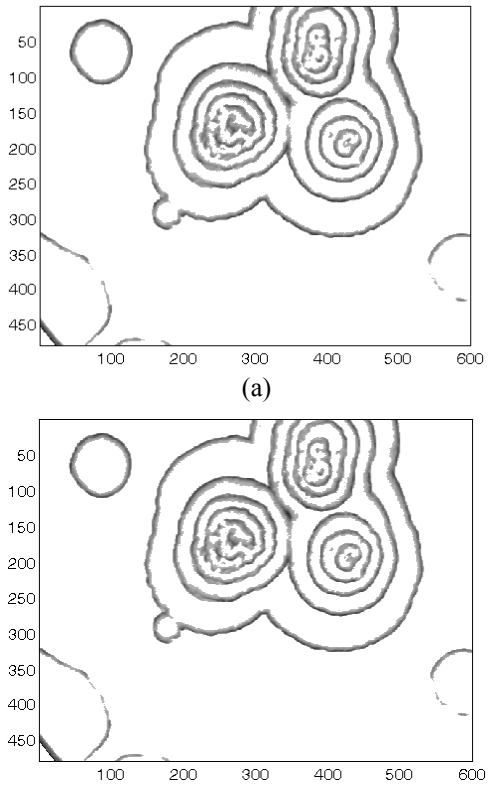

Fig. 2. One-step ahead predictions, on the basis of the observations at a certain time instant *t*, for the BZ reaction modeling problem. (a) True measurements at the time instant *t*+1; (b) Predictions.

V. Conclusion

A novel two-stage training scheme has been proposed for constructing a new class of lattice dynamical wavelet neural networks. It has been demonstrated using real data that the proposed network model is effective for spatio-temporal system identification. The proposed network possesses a few desirable features, for example, the network is almost self-implemented, meaning that by starting with some given conditions (initial, boundary and termination), all within-network parameters can be estimated and calculated by the proposed algorithms; the network provides a transparent model, where individual wavelet-neurons are explicitly available. The main drawback of the new network is perhaps the computation time, which is mainly spent on the nonlinear optimization procedure using the PSO algorithm. Theoretical analysis of the proposed algorithm and comprehensive comparisons between the proposed method and other approaches are worth giving in future studies.

References

[1] S. Wolfram, *Cellular Automata and Complexity*. New York: Addison-Wesley, 1994.

[2] K. Kaneko, *Theory and Application of Coupled Map Lattices*. New York: Wiley, 1993.

[3] L. O. Chua and T. Roska, *Cellular Neural Networks: Foundations and Applications*. Cambridge: Cambridge University Press, 2001.

[4] J. H. Friedman and W. Stuetzle, "Projection pursuit regression," *J. Amer. Statist. Assoc.*, vol. 76, no. 376, pp.817–823, Dec.1981.

[5] S. A. Billings, S. Chen, and M. J. Korenberg, "Identification of MIMO nonlinear systems using a forward regression orthogonal estimator," *Int. J. Control*, vol. 49, no.6, pp.2157-2189, Jun.1989.

[6] S. Chen, S. A. Billings and W. Luo, "Orthogonal least squares methods and their application to nonlinear system identification," *Int. J. Control*, vol. 50, no. 5, pp. 1873–1896, Nov. 1989.

[7] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least-squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Networks.*, vol. 2, no. 2, pp. 302–309, Mar. 1991.

[8] S. A. Billings, L. Z. Guo, and H. L. Wei, "Identification of coupled map lattice models for spatio- temporal patterns using wavelets," *Int. J. Control*, vol.14, No. 14, pp. 1021-1038, Nov 2005.

[9] I. Daubechies, *Ten Lectures on Wavelets*. Philaelphia, Pennsylvania: Society for Industrial and Applied Mathematics, 1992.

[10] Q. Zhang, "Using wavelet network in nonparametric estimation," *IEEE Trans. Neural Networks*, vol. 8, no.2, pp. 227-236, 1997.

[11] J. H. Xu and D. W. C. Ho, "A basis selection algorithm for wavelet neural networks," *Neurocomputing*, vol. 48, pp. 681-689, 2002.

[12] S. A. Billings and H. L. Wei, "A new class of wavelet networks for nonlinear system identification", *IEEE Trans. Neural Networks*, 16, no.4, pp. 862-874, July 2005.

[13] L. K. Jones, "A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training," *Ann. Statist.*, vol. 20, no. 1, pp. 608–613, 1992.

[14] T. Y. Kwok and D. Y. Yeung, "Constructive algorithms for structure learning in feedforward neural networks for regression problems," *IEEE Trans. Neural Netw.*, vol. 8, no. 3, pp. 630–645, May 1997.

[15] T. Y. Kwok and D. Y. Yeung, "Objective functions for training new hidden units in constructive neural networks," *IEEE Trans. Neural Netw.*, vol. 8, no. 5, pp. 1131–1148, Sep. 1997.

[16] J. N. Hwang, S. R. Lay, M.Maechler, R. D. Martin, and J. Schimert, "Regression modeling in back- propagation and projection pursuit learning," *IEEE Trans. Neural Networks*, vol. 5, no. 3, pp. 342–353, May 1994.

[17] J. Kennedy, R. C. Eberhart and Y. Shi, *Swarm Intelligence*, San Francisco: Morgan Kaufmann Publishers, 2001.

[18] Z. Zhang, *Matching Pursuit*. PhD dissertation, New York University, 1993.

[19] G. B. Huang, L. Chen, and C. K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Trans. Neural Netw.*, vol. 17, no. 4, pp. 879–892, Jul. 2006.

[20] S. A. Billings and H. L. Wei, "An adaptive orthogonal search algorithm for model subset selection and nonlinear system identification', *Int. J. Control*, vol.81, no.5, pp.714-724, May 2008.

[21] Y. Kuramoto, *Chemical Oscillations*, *Waves*, *and Turbulence*. Berlin: Springer, 1984.
[22] J. J. More, "The Levenberg-Marquardt algorithm: implementation and theory," in *Numerical Analysis*, G. A. Watson (e.d.), Lecture Notes in Mathematics 630, Springer-Verlag, 1977, pp. 105-116.
[23] A. T. Winfree, "Spiral waves of chemical activity," Science, vol. 175, no.4022, pp. 634-636, 1972.