Proceedings of the 10th International Conference on Computer Science & Education, July 22-24, 2015.
Fitzwilliam College, Cambridge University, UK. DOI:10.1109/ICCSE.2015.7250228

A Wavelet Neural Network Model for Spatio-Temporal Image Processing and Modeling

Hua-Liang Wei, Yifan Zhao, and Richard Jiang

***Abstract*—Spatio-temporal images are a class of complex dynamical systems that evolve over both space and time. With increased interest in the investigation of nonlinear complex phenomena, especially spatio-temporal data analysis governed by evolutionary laws that are dependent on both spatial and temporal dimensions, there has been an increased need to investigate data-driven modeling methods for this class of complex systems. Compared with pure temporal processes, the identification of spatio-temporal models from observed images is much more difficult and quite challenging. Starting with an assumption that there is no a priori information about the true model but only observed data are available, this work introduces a new type of wavelet network that utilizes the easy tractability and exploits the good properties of multiscale wavelet decompositions to represent the rules of the associated spatio-temporal evolutionary system. An application to a chemical reaction exhibiting a spatio-temporal evolutionary behaviour, is investigated to demonstrate the application of the proposed modeling and learning approaches.**

I. Introduction

IN the real world, spatio-temporal (ST) phenomena exist widely in biology, chemistry, ecology, geography, medicine, physics, and sociology and so on. To simulate and analyze spatio-temporal phenomena, several efficient representations, for example the well known cellular automata (CA) [1], cellular neural networks (CNNs) [2] and coupled map lattice (CML) models [3], have been proposed. In these representations, it is often assumed that the associated mathematical model structure, along with the model parameters, is known, so that the model can be used to describe or analyze some specific phenomena. However, the evolution laws associated with real-world ST phenomena may not always be completely known, and evolution rules need to be acquired from observed data of images or patterns. Hence, in recent years, the identification of ST models from observed data has received much attention and several efficient identification methods and algorithms have been proposed, see for example [4],[5].

Manuscript received December xx, 2014. This work was supported in part by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/I011056/1 and Platform Grant EP/H00453X/1.

H. L. Wei is with the Department of Automatic Control & Systems Engineering, and with the INSIGNEO Institute for *in silico* Medicine, University of Sheffield, Sheffield, S1 3JD, United Kingdom (e-mail: w.hualiang@sheffield.ac.uk).

Y. Zhao is with the Engineering Sciences Division, School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, Bedfordshire, MK43 0AL,UK (email: yifan.zhao@cranfield.ac.uk).

R. Jiang is with the School of Computer Science and Digital Technologies, Northumbria University, Newcastle, UK (email: richard.jiang@northumbria.ac.uk)

This study considers the ST model identification problem, where it is assumed that there is no a priori information about the true model structure and only imaged data are available. Motivated by the successful applications of the multiscale and multiresolution analysis approaches in classical signal and image processing [6], as well as in dynamical process modeling [7]-[12], and also inspired by the easy tractability of conventional coupled map lattice models [3], this study aims to introduce a new type of wavelet network model that utilizes the easy tractability and exploits the good properties of multiscale wavelet decompositions for ST system identification. Unlike in a typical wavelet based multiscale or multiresolution dynamical modeling approaches, where the elementary building blocks are strictly chosen to be some dyadic wavelets, in the new wavelet model, the choice of the prototype functions are permitted to be very flexible, any functions including wavelets, B-splines and Gaussian type functions can be chosen as the elementary building blocks as long as there is strong evidence that the functions possess desirable properties and can lead to a good model for a given modeling problem. In most existing wavelet models for dynamical systems, the scale and shift parameters are restricted to a dyadic lattice. Dyadic wavelet models are proved to be perfect for general static signal representation, in that dyadic wavelets, along with associated scale functions, can often form orthogonal (orthonormal) bases that play an important role in wavelet multiresolution analysis (MRA)[6]. An important property of an orthonormal decomposition is that the well known Pareval's theorem holds, that is, the energy of a signal is conserved, without any loss, in the wavelet coefficients. For the ST dynamical system modeling problem, where observations are often sparse in the problem space, a dyadic lattice may not usually be an optimal choice. In addition, data used in dyadic wavelet models for nonlinear dynamical systems often need to be compressed or normalized to some specific finite interval, to simplify the associated modeling procedures. Although data normalization is frequently used in many modeling approaches and can often simplify the associated modeling procedures, normalization may, at the same time, change the physical meanings of the signals to be modeled. This may be undesirable for some applications where variables are required to preserve their physical dimension. In this study, data normalization is unnecessary.

The proposed model is composed of a number of basis functions; the feature of each individual function is determined by three factors: the scale (dilation) parameter, the shift (translation) parameter, and the coefficient weighted on the associated function. For a chosen

elementary building block (the prototype function), the task of the model identification involves as least three aspects: the determination of the scale and shift parameters; the determination of the model structure and complexity, that is, the determination of which and how many basis functions should be included in the model; and the estimation of the weight coefficients. A new simple unsupervised histogram-based clustering algorithm is introduced, which can be used to determine the scale and shift parameters of individual functions that will be used to form an initial full network model. The initial full model is in general highly redundant. A forward orthogonal regression learning algorithm [13]-[15], implemented using a mutual information method, is then applied to refine and improve the initial full model by removing redundant basis functions.

II. Wavelet-Based Additive Neural Networks

The 2D case, which has obvious physical meanings and is widely applied in practice, is taken as an example to illustrate how to construct the network model. For simplicity, only the autonomous case is considered. In an autonomous ST system, no external input image is imposed, and the output image at any time t is due exclusively to the initial conditions and the evolution of the pattern. Model representations for these situations can easily be extended in a straightforward way to other more complex cases.

A. Lattice Cell Model

Assume that the 2-D image or pattern produced by an ST system, at the time instant t, consists of an $I\times J$ rectangular array of cells, $C^t(i,j)$, with Cartesian coordinates (i,j), i=1,2, …, I, j=1,2, …, J. Following [1], let $S_r^t(i,j)$ be the sphere of influence of the radius r of cell $C^t(i,j)$, at the time instant t, defined as

$$S_r^t(i,j)=\{C^t(i,j):\max_{1\le p\le I,1\le q\le J}\{|i-p|,|j-q|\}\le r\} \quad (1)$$

where t=1,2…, i=1,2, …, I, j=1,2, …, J, and r is a non-negative integer number indicating how many neighborhood cells are involved in the evolution procedure. The sphere $S_r^t(i,j)$ is sometimes referred to as the (2r+1) ×(2r+1) neighbourhood. Let $s_{i,j}(t)\in\mathbb{R}$ be the state variable representing the cell $C^t(i,j)\in S_r^t(i,j)$. From the definition of $S_r^t(i,j)$, a total of (2r+1)^2 state variables are involved in (1), see Table I, where the symbol $C(i,j)$ will be used to indicate cells at an arbitrary evolution time instant.

Let $s_{i,j}(t)$ be the (i,j)th cell to be updated at time t. A wide range of ST systems can be described by the discrete-time, discrete-space and continuous-state spatio-temporal difference equation of the form below

$$\left.\begin{aligned} s_{i,j}(t)&=f(\mathbf{s}(t-1),\mathbf{s}(t-2),\cdots\quad{}_{ag}))\\ &=f(s_{i-r,j-r}(t-1),...,s_{i,j}(t-1),...,s_{i+1,j+r}(t-1),...,\\ &\quad s_{i-r,j-r}(t-2),...,s_{i,j}(t-2),...,s_{i+1,j+r}(t-2),...,\\ &\quad s_{i-r,j-r}(t-n_{lag}),...,s_{i,j}(t-n_{lag}),...,s_{i+1,j+r}(t-n_{lag}))\end{aligned}\right\} \quad (2)$$

TABLE I
The $(2r+1)\times(2r+1)$ Neighborhood Cells, Represented by Variables x_k, with $k=1,2,...,(2r+1)^2$.

$C(i\text{-}r,j\text{-}r)$	…	$C(i\text{-}r,j)$	…	$C(i\text{-}r,j+r)$
x_1	…	x_r	…	x_{2r+1}
…	…	…	…	…
$C(i,j\text{-}r)$		$C(i,j)$		$C(i,j+r)$
$x_{r(2r+1)+1}$	…	$x_{r(2r+1)+(r+1)}$	…	$x_{(r+1)(2r+1)}$
…	…	…	…	…
$C(i+r,j\text{-}r)$	…	$C(i+r,j)$	…	$C(i+r,j+r)$
$x_{2r(2r+1)+1}$		$x_{2r(2r+1)+(r+1)}$		$x_{(2r+1)(2r+1)}$

where f is some nonlinear function, n_{lag} is the time lag, defined as a positive integer, indicating how many past images or patterns are involved in the evolution procedure, and $\mathbf{s}(t-k)$ is the state vector formed by the (2r+1)^2 state variables relative to the patterns at the time instant (t-k) with k=1,2, …, n_{lag}, that is,

$$\mathbf{s}(t-k)=[s_{i-r,j-r}(t-k),...,s_{i,j}(t-k),s_{i+r,j+r}(t-k)] \quad (3)$$

Note that the general lattice cell representation (2) includes, as special cases, most typical coupled map lattice models. For convenience of description, introduce d single-indexed variables $x_k(t)$ as below

$$\mathbf{x}(t)=[x_1(t),x_2(t),...,x_d(t)]=[\mathbf{s}(t-1),\mathbf{s}(t-2),...,\mathbf{s}(t-n_{lag})] \quad (4)$$

where $\mathbf{s}(t-k)=[x_{1+(k-1)(2r+1)^2}(t),...,x_{k(2r+1)^2}(t)]$ for k=1,2, …, n_{lag}. Also, let $y(t)$ represent the state variable $s_{i,j}(t)$ corresponding to the central cell $C^t(i,j)$. Then, Eq. (2) becomes

$$y(t)=f(\mathbf{x}(t))=f(x_1(t),x_2(t),...,x_d(t)) \quad (5)$$

In conventional coupled map lattice models, the nonlinear function f in model (2) is often assumed to be known as some deterministic function. However, for real-word complex ST systems, a pre-determined function f may not sufficiently characterize the underlying dynamics. It may be better to learn, from available real observations, an appropriate model for a given ST system. The task of ST system identification is to construct, based on available data, a model that can represent, as close as possible, the observed evolution behavior. Unlike constructing static models for typical data fitting, the objective of dynamical modeling is not merely to seek a model that fits the given data well, it also requires, at the same time, that the model should be capable of capturing the underlying system dynamics carried by the observed data, so that the resultant model can be used in simulation, analysis, and control studies.

B. New Multiscale Wavelet Based Lattice Cell Model

Inspired by the idea behind the traditional coupled map lattice models [3], the present study employs an additive model structure to approximate the nonlinear function (5)

$$\begin{aligned} y(t) &= \hat{f}(\mathbf{x}(t)) + \mathbf{e}(t) \\ &= f_1(x_1(t)) + f_2(x_2(t)) + \ldots + f_d(x_d(t)) + e(t) \end{aligned} \tag{6}$$

where $f_i(\cdot)$ are some univariate nonlinear functions that need to be identified, and $e(t)$ is some modeling error that can be treated as an independent identical distributed noise sequence. A common approach used for effectively reconstructing the nonlinear functions $f_i(\cdot)$ is to construct a nonlinear approximator $\hat{f}_i$ using some specific types of basis functions.

The present study, however, employs a family of shifted multiscale basis functions to approximate the nonlinear maps $f_1, f_2, ..., f_d$ in (6). Let ψ be a chosen mother wavelet and consider the family

$$\psi(x; b, a) = \psi\left(\frac{x-b}{a}\right) \tag{7}$$

where $a \in \mathbb{R}^+$, $b \in \mathbb{R}$, and the wavelet ψ is admissible. The admissibility condition is depicted using the Fourier transform $\hat{\psi}(\xi)$ of ψ as $C_\psi = \int_{-\infty}^{\infty} \xi^{-1} |\hat{\psi}(\xi)|^2 d\xi < \infty$ [16]. Each of these d functional components $f_i(x_i)$ can now be represented as

$$f_i(x_i(t)) = \sum_k \sum_j w_{k,j}^{(i)} \psi(x_i(t); b_{k,j}^{(i)}, a_{k,j}^{(i)}) \tag{8}$$

where $a_{k,j}^{(i)}$ and $b_{k,j}^{(i)}$ are pre-determined scale and shift parameters, and $w_{k,j}^{(i)}$ are the associated weight coefficients. Equation (6) can now be written as

$$\begin{aligned} y(t) &= \hat{f}(\mathrm{x}(t)) + e(t) = \sum_{i=1}^{d} f_i(x_i(t)) + e(t) \\ &= \sum_{i=1}^{d} \sum_k \sum_j w_{k,j}^{(i)} \psi(x_i(t); b_{k,j}^{(i)}, a_{k,j}^{(i)}) + e(t) \end{aligned} \tag{9}$$

This is the initial full model for the associated ST system (2). In this model the scale and shift parameters $a_{k,j}^{(i)}$ and $b_{k,j}^{(i)}$ are restricted to a pre-specified non-dyadic lattice, which is directly determined by using the information given by the original observation dataset. The weight coefficients $w_{k,j}^{(i)}$, however, need to be estimated by solving the associated regression equation problem.

III. Determination of Shift and Scale Parameters

A. Data Grouping

Consider a given time series $\{x(t) : t = 1, \cdots$. Let $x_{\min} = \min\{x(t)\}_{t=1}^{N_0}$, $x_{\max} = \max\{x(t)\}_{t=1}^{N_0}$ and $R = [x_{\min}, x_{\max}]$. Now the objective is to partition all the data points in the time series $\{x(t)\}$ into K groups. The grouping criterion and the associated partitioning procedure are as follows:

- Divide the interval R into k equally-spaced subintervals (bins); the kth bin is defined as $R_k = [r_k, r_{k+1})\ (1 \le k \le K\text{-}1)$ and $R_k = [r_k, r_{k+1}]$ for k=K, where $r_k = x_{\min} + (k-1)h$ and $h = (x_{\max} - x_{\min}) / K$ (h is referred to as the bin width).
- If $x(t) \in R_k$, then the data point is included in the kth bin.
- Denote the number of data points in the kth bin R_k by g_k.

These are the basic ideas of a histogram method used for data grouping. However, the determination of the bin width and bin number K is still an open issue in histogram analysis. The famous Sturges' rule [17], which has been used in default in some popular software, suggests that the total bin number be chosen as $K = 1 + \log_2(N_0)$. It has been pointed out [18] that the Sturges' rule is more of a number-of-bins rule rather than a bin-width-oriented rule itself, and it has been shown that the bin width produced by the Sturges' rule leads to an over-smoothed histogram, especially when the number of samples is large. In [18], a cross-validation (CV) criterion, called the biased CV function, was suggested for bin width choice

$$\mathrm{BCV}(h(K)) = \frac{5}{6N_0 h(K)} + \frac{1}{12N_0^2 h(K)} \sum_{k=1}^{K-1} (g_{k+1} - g_k)^2 \tag{10}$$

where $h(K) = (x_{\max} - x_{\min}) / K$ is the bin width of K partitions. The bin width for the associated dataset is defined to be the one that minimises the BCV criterion. This cross validation criteria can be used to determine the scale and shift parameters of the associated basis functions.

B. Choice of the Shift and Scale Parameters

For a given time series $\{x(t) : t = 1, 2, \cdots, N_0\}$, assume that a total of K groups have been determined using the above histogram-based grouping algorithm, and let c_k be the centre (midpoint) of the kth interval R_k. In the present study, wavelet shift parameters are chosen to be these K centers.

It is known that wavelet basis functions are compactly supported or nearly compactly supported. For example the B-splines and associated wavelets [19] are compactly supported, while the Gaussian and the Mexican hat wavelets [16] are nearly compactly supported. Hence, to the wavelet transform (7), for any scale and shift parameters a and b, there must exist a positive number δ such that

$$|\psi(x; b, a)| = \left|\psi\left(\frac{x-b}{a}\right)\right| \approx 0 \text{ , for } \left|\frac{x-b}{a}\right| \ge \delta \tag{11}$$

For example, for the Mexican hat wavelet $\psi(x)=(1-x^2)\exp(-x^2/2)$, $|\psi(x)|\le 0.0001$ for $|x|\ge\delta=5$. Thus, for a fixed shift parameter *b*, the scale parameter *a* should not be chosen too small, because a very scale will 'disable' many useful data points. On the other hand, the scale parameter *a* should not be chosen too large, because a very large scale could make the associated functions become too smooth to capture detailed dynamics of the signal.

From the above discussion, the shift and scale parameters *b* and *a* are chosen as follows:

- The number of wavelet shift parameters is chosen to be *K*, which is defined using the histogram-based grouping algorithm.
- The *K* shift parameters $b_1,b_2,\ldots,b_k$ are chosen to be the centers of the *K* bins (groups).
- For each shift parameter b_k, allocate a total of *J* scale parameters to the associated basis functions, denote these scale parameters by $a_1,a_2,\ldots,a_J$, where $a_j=2^j a_0/4$ for j=1,2, …, J and $a_0=h(K)/\sqrt{2K}$.

Here, the idea of choosing $a_0=h(K)/\sqrt{2K}$ comes from [20], where it is suggested that the scale (bandwidth) parameter be chosen as $\sigma=\max_{1\le i<j\le K}\{|b_i-b_j|\}/\sqrt{2K}$. Applying the above procedure to equation (6), yields,

$$\begin{aligned} y(t) &= \sum_{i=1}^{d}\sum_{k}^{K_i}\sum_{j}^{J} w_{k,j}^{(i)}\psi(x_i(t);b_{k,j}^{(i)},a_{k,j}^{(i)})+e(t) \\ &= \sum_{i=1}^{d}\sum_{k=1}^{K_i}\sum_{j=1}^{J} w_{k,j}^{(i)}\psi\left(\frac{x_i(t)-b_k^{(i)}}{a_j^{(i)}}\right)+e(t) \end{aligned} \tag{12}$$

where $a_j^{(i)}$ and $b_k^{(i)}$ are the scale and shift parameters, relative to the *i*th variable $x_i(t)$. In model (12), it is assumed that the time series $\{x_i(t)\}$ produced by the *i*th variable $x_i(t)$ has been partitioned into K_i groups, and thus a total of K_i shift parameters are involved for the variable $x_i(t)$; for each of these shift parameters, there are a total of *J* scale parameters associated with the wavelets. Therefore, the initial full wavelet model (12) contains a total of $M=J(K_1+K_2+\ldots+K_d)$ basis functions. Model (12) can easily be converted to a linear-in-the-parameters form

$$y(t)=\sum_{m=1}^{M}\theta_m\phi_m(t)+e(t)=\boldsymbol{\varphi}^T(t)\boldsymbol{\theta}+e(t) \tag{13}$$

where $\varphi_m\in\{\psi(x_i;b_k^{(i)},a_{k,j}^{(i)}):i=1,\ldots,d;k=1,\ldots,K_i;j=1,\ldots,J\}$, θ_m are model parameters, and $\boldsymbol{\varphi}(t)=[\varphi_1(t),\ldots,\varphi_M(t)]^T$ and $\boldsymbol{\theta}$ are the associated regressor and parameter vectors, respectively. Notice that in most cases the initial full regression equation (13) might be highly redundant, some of the regressors or model terms can thus be removed from the initial regression equation without any effect on the predictive capability of the model, and this elimination of the redundant regressors usually improves the model performance. Generally, only a relatively small number of model terms need to be included in the regression model for most nonlinear dynamical system identification problems. An efficient model term selection algorithm is thus highly desirable to detect and select the most significant regressors.

IV. Model Refinement

Let $\boldsymbol{\varphi}_m=[\varphi_m(1),\ldots,\ \phi_m(N)]^T$ be a vector formed by the *m*th candidate model term in the initial full model (13), where *m*=1,2, …, *M*. Let $D=\{\boldsymbol{\varphi}_1,\cdots,\boldsymbol{\varphi}_M\}$ be a dictionary composed of the *M* candidate bases. Note that the set *D* is often highly redundant. The model refinement problem amounts to finding, from the vector dictionary *D*, a full dimensional subset $D_m=\{\mathbf{p}_1,\ldots,\mathbf{p}_m\}=\{\boldsymbol{\varphi}_{i_1},\ldots,\boldsymbol{\varphi}_{i_m}\}$, where $\mathbf{p}_k=\boldsymbol{\varphi}_{i_k}$, $i_k\in\{1,2,\ldots,M\}$ and *k*=1,2,…,*m* (generally $m<<M$), so that $\mathbf{y}$ is satisfactorily approximated using a linear combination of $\mathbf{p}_1,\mathbf{p}_2,\ldots,\mathbf{p}_m$ as

$$\mathbf{y}=\beta_1\mathbf{p}_1+\beta_2\mathbf{p}_2+\ldots+\beta_m\mathbf{p}_m+\mathbf{e}_m \tag{14}$$

where $\mathbf{e}_m$ is the associated model residual vector.

The orthogonal least squares (OLS) algorithm [21],[22] can be used to determine model basis functions (model terms). In this study, however, a variation of the OLS algorithm, called the forward orthogonal regression (FOR) algorithm, implemented using a mutual information method [13],[14], is employed for model refinement.

V. An Application in Chemical Reactions

The proposed multiscale additive wavelet regression model can be applied to identify some SP phenomena, where the true models are unknown and the initial full model involves a great number of 'input' or 'independent' variables. To illustrate the application of the new modeling method, the Belousov-Zhabotinsky [23] reaction was considered here as an example. As a classical example of nonequilibrium thermodynamics, the BZ reaction provides an interesting chemical model of nonequilibrium phenomena, and the modeling and identification of these types of reactions is of extreme interest for theoretical analysis of such phenomena.

By adopting the recipe given by Winfree [24], an experiment resulting in a thin layer BZ reaction was carried out in the laboratory, and a set of images were sampled with equal time intervals during the experiment, using a digital video camera that is connected to a PC via a USB socket. The sampled images were processed and saved as patterns with a resolution of 300 by 500 pixels. Some of these patterns are shown in Fig. 1. These sampled images were used to estimate a parsimonious wavelet network model.

A. Experimental Data

Consider the model of form (2), where the total number of model variables is determined by two factors: the radius of the neighbourhood, r, and the time lag, n_{lag}. In the present study, the two coefficients were chosen to be r=1 and n_{lag}=2. Thus, the model involves a total of 18 model variables.

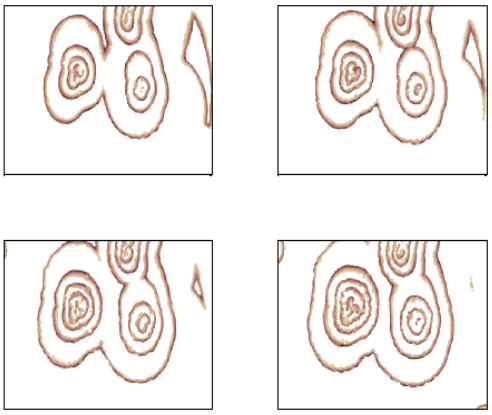


Fig. 1. Sampled snapshots for the BZ reaction at different time instants. The size of each template is 300×500 (300 pixels in the vertical direction and 500 pixels in the horizontal direction). (a) t=10; (b) t=20; (c) t=30; (d) t=40.

The state variable $s_{i,j}(t)$, at the present time instant t, was initially assumed to be associated with state variables in the past two adjacent neighbourhoods at the previous time instants t-1 and t-2. Any two patterns, at the abutting time instants t and t-1 are called an adjacent pattern group. For an arbitrary time instant, the data pair, $\{\mathbf{x}(t), y(t)\}$, where $\mathbf{x}(t)$ and $y(t)$ are defined by (4) and (5), is called a data pair. Notice that $\mathbf{x}(t)$ and $y(t)$ are also implicitly associated with the spatial location indices i and j. As a consequence, for any given time instant t, there would be a large number of data pairs available.

A training dataset, consisting of a total of N=6000 data pairs, $\{\mathbf{x}(t),y(t)\}$(t=1,2,…,N), was generated for model identification, where $y(k)$ represents the value of the relevant central cell at the present time instant, and $\mathbf{x}(t)=[x_1(t),x_2(t),\cdots$ represent the values of the 18 involved cells on a squared lattice, at the previous time instants. Data pairs $\{\mathbf{x}(t),y(t)\}$ in the training dataset were randomly chosen from ten adjacent pattern groups that were also randomly selected from the first 40 sampled patterns.

B. Determining Shift and Scale Parameters

Note that all the involved 18 variables come from the same system; in a statistical sense, these 18 variables, as well as the 'output' variable $y(t)$, should obey the same distribution. Thus the 18 variables could be allocated the same shift parameters. It was noticed that the values of a great number of observations are exactly equivalent to the maximum value y_{max}=255. Thus, the one shift parameter was in default chosen to be 255, and other shift parameters were determined by performing the histogram-based grouping algorithm on the associated time series $\{y(t)\}$ with t=1,2,…, N, where data points whose values are exactly equivalent to y_{max}, were excluded. The biased CV criterion defined by (10), shown in Fig. 2, suggests that the optimal number of groups for the associated dataset may be chosen to be 8.

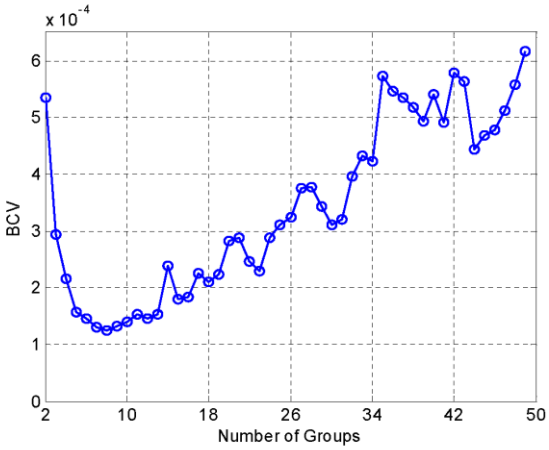


Fig. 2. The BCV criterion versus the number of groups, produced using the histogram-based clustering algorithm.

The Mexican hat wavelet function, $\psi(x)=(1-x^2)e^{-x^2/2}$, was used as the elementary building block for constructing the wavelet network model, and the primary bandwidth for each group was chosen to be a_0=6.25. For each shift parameter b_k, a total of 6 scales were to used to perform associated wavelet transforms; denote these scale parameters by $a_1,a_2,\ldots,a_6$, where $a_j=2^{j-2}a_0$. The initial full model was thus of the form

$$y(t)=\sum_{i=1}^{18}\sum_{k=1}^{9}\sum_{j=1}^{6}w_{k,j}^{(i)}\psi\left(\frac{x_i(t)-b_k}{a_j}\right) \tag{15}$$

C. Model Refinement and Performance Evaluation

The initial full wavelet network model (18) contains a total of 972 model terms; most of the model candidate model terms may be redundant. The initial full model thus needs to be refined. The FOR-MI algorithm was performed, over the given training dataset, to select significant individual basis functions from the initial model (15). An adjustable generalized cross-validation (readers are referred to [15] for the definition of the AGCV) suggests that a total of 26 basis functions should be included in the final model.

To evaluate the performance of the identified additive wavelet models, the short-term predictive capability of the models was inspected. Denote the observation of the image (pattern) measured at the time instant t by $X(t)$. The k-step-ahead prediction, denoted by $\hat{X}(t+k\,|\,X(t),X(t-1);f)$, where f represents the identified nonlinear function, is the iteratively produced result by the identified model, on the basis of $X(t)$ and $X(t$-1), but without using information on

observations for patterns at any other time instants. As an example, the measurements at the time instants $t = 41$ and 42 were used to calculate the 1-step-ahead predictions for the values at the time instant t=43, and this is shown in Fig. 3, from which it is clear the identified model provide a perfect representation for the BZ reaction data.

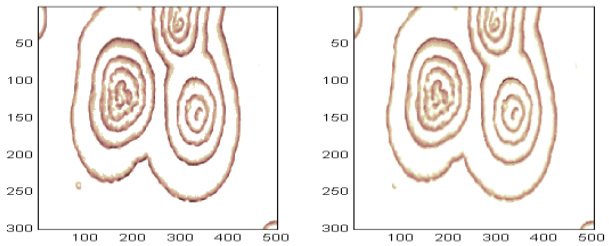


Fig. 3. Model prediction (one step ahead) for the BZ reaction at time instant t=43. (a) Real measurement at t=43; (b) Predicted image from the model.

VI. Conclusion

The proposed wavelet neural network model, by incorporating a multiscale wavelet decomposition approach into the traditional lattice cell model, provides a simple but powerful representation for spatio-temporal images. An initial full wavelet network model for a given model identification problem may involve a great number of basis functions. Experience has shown that in general only a relatively small number of basis functions are significant and need to be included in the model. This was achieved using the well known orthogonal least squares types of learning algorithms.

The wavelet network model identification procedure is performed on some scaled and translated basis functions, where two types of parameters need to be determined: the shift and the scale parameters. Although the present study provides some tips for choosing these parameters, efficient optimization methods and algorithm for these parameters still need to be considered in a future study, to produce more efficient models for complex spatio-temporal systems. A challenging topic in spatio-temporal system identification and modeling is the determination of the neighborhood cells - e.g. which and how many neighborhood cells should be considered and included in the system evolution models, this is believed to be a most difficult problem to solve where only observed data or images are available but the inherent dynamics of the system is not clear. This challenging issue would be investigated in future research.

References

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

[2] L. O. Chua and T. Roska, *Cellular Neural Networks and Visual Computing*. Cambridge: Cambridge University Press, 2002.

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

[4] S.A. Billings and Y. Y. Yang, ‘‘Identification of the neighbourhood and CA rules from spatio-temporal CA patterns,’’ *IEEE Trans. Syst. Man Cybern. B*, vol. 33, pp. 332–339, 2003.

[5] S. Wang, H. L. Wei, D. Coca, and S. A. Billings, “Model term selection for spatio-temporal system identification using mutual information,” *Int. J. Syst. Sci*, vol.44, no.2, pp.223-231, 2013.

[6] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego: Academic Press, 1998.

[7] H. L. Wei and S. A. Billings, “A unified wavelet-based modelling framework for nonlinear system identification: the WANARX model structure,” *Int. J. Control*, vol. 77, no. 4, pp.351-366, Mar. 2004.

[8] H. L. Wei and S. A. Billings, “Identification and reconstruction of chaotic systems using multiresolution wavelet decompositions,” *Int. J. Syst. Sci*, vol. 35, no. 9, pp. 511-526, July 2004.

[9] S. A. Billings and H. L. Wei, “The wavelet-NARMAX representation: A hybrid model structure combining polynomial models with multiresolution wavelet decompositions,” *Int. J. Syst. Sci.*, vol. 36, no. 3, pp. 137-152, Feb. 2005.

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

[11] S. A. Billings, L. Z. Guo, and H. L. Wei, “Identification of coupled map lattice models for spatio-temporal patterns using wavelets,” *Int. J. Sys. Sci*., vol. 37, no. 14, pp. 1021-1038, Nov 2006.

[12] H. L. Wei, S. A. Billings, Y. Zhao, and L. Z. Guo, L. Z., “Lattice dynamical wavelet neural networks implemented using particle swarm optimization for spatio-temporal system identification,” *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 181-185, 2009.

[13] S. A. Billings and H. L. Wei, “Sparse model identification using a forward orthogonal regression algorithm aided by mutual information,” *IEEE Trans. Neural Networks*, vol. 18, no. 1, pp. 306-310, 2007.

[14] Wei, H. L. and Billings, S. A., “Model structure selection using an integrated forward orthogonal search algorithm assisted by squared correlation and mutual information,” *Int. J. Modelling, Identification and Control*, vol. 3, no. 4, pp. 341-356, 2008.

[15] H. L. Wei, and S. A. Billings, “Generalized cellular neural networks (GCNNs) constructed using particle swarm optimization for spatio-temporal evolutionary pattern identification,” *Int. J. Bifurcation and Chaos*, vol. 18, no. 12, pp. 3611-3624, 2008.

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

[17] H. Sturegs, “The choice of a class-interval,” *J. Amer. Statisst. Assoc*., Vol. 21, pp. 65-66, 1926.

[18] D. W. Scott, *Multivariate Density Estimation*. New York: John Wiley, 1992.

[19] C. K. Chui, *An Introduction to Wavelets*. New York: Academic, 1992.

[20] S. Haykin, Neural Networks—A Comprehensive Foundation (2nd ed.). New Jersey: Prentice Hall, 1999.

[21] S. A. Billings, S. Chen, and M. J. Korenberg, “Identification of MIMO non-linear systems using a forward-regression orthogonal estimator,” *Int. J. Control*, vol. 49, no. 6, pp. 2157-2189, 1989.

[22] 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, 1989.

[23] B. P. Belousov, “A periodic reaction and its mechanism,” in Collection of Short Papers on Radiation Medicine (in Russian), Medgiz, Moscow, pp.145-152, 1959.

[24] A. T. Winfree, “Spiral waves of chemical activity,” *Science*, vol.175, no.4022, pp. 634-636, 1972.



https://dspace.org/

A wavelet neural network model for spatio-temporal image processing and modeling

Wei, Hua-Liang; Zhao, Yifan; Jiang, R.

2015
IEEE
http://dx.doi.org/10.1109/ICCSE.2015.7250228
Hua-Liang Wei, Yifan Zhao and R. Jiang. A wavelet neural network model for spatio-temporal image processing and modeling. Proceedings of the 10th International Conference on Computer Science & Education, July 22-24, 2015. Fitzwilliam College, Cambridge University, UK.

Downloaded from a DSpace repository. https://dspace.org/