**Monograph:**
Guo, L.Z., Billings, S.A. and Coca, D. (2007) Multiscale identification of spatio-temporal dynamical systems using a wavelet multiresolution analysis. Research Report. ACSE Research Report no. 947 . Automatic Control and Systems Engineering, University of Sheffield

# Multiscale identification of spatio-temporal dynamical systems using a wavelet multiresolution analysis

Guo, L. Z., Billings, S. A., and Coca, D.

Department of Automatic Control and Systems Engineering
University of Sheffield
Sheffield, S1 3JD
UK

Research Report No. 947
March 2007

# Multiscale identification of spatio-temporal dynamical systems using a wavelet multiresolution analysis

Lingzhong Guo, Stephen A Billings, and Daniel Coca

***Abstract*— In this paper, a new algorithm for the multiscale identification of spatio-temporal dynamical systems is derived. It is shown that the input and output observations can be represented in a multiscale manner based on a wavelet multiresolution analysis. The system dynamics at some specific scale of interest can then be identified using an orthogonal forward least-squares algorithm. This model can then be converted between different scales to produce predictions of the system outputs at different scales. The method can be applied to both multiscale and conventional spatio-temporal dynamical systems. For multiscale systems, the method can generate a parsimonious and effective model at a coarser scale while considering the effects from finer scales. Additionally, the proposed method can be used to improve the performance of the identification when measurements are noisy. Numerical examples are provided to demonstrate the application of the proposed new approach.**

***Index Terms*— Multiscale identification, spatio-temporal system, orthogonal least squares algorithm, multiresolution analysis**

## I. Introduction

THE identification of spatio-temporal systems has recieved increasing attention in recent years with applications in a variety of scientific and engineering areas (Voss, Bunner, Abel 1998, Coca and Billings 2001, 2002, McSharrya, Ellepolab, von Hardenberga, Smitha, Kenning 2002, Yan, Hu, Zhou, and Liu 2004, Billings, Guo, and Wei 2006, Guo and Billings 2006, Guo, Billings, and Wei 2006). Both discrete and continuous time models have been developed including coupled map lattice (CML) models, lattice dynamical system (LDS) models, and partial differential equations (PDE) to describe the underlying spatio-temporal dynamical systems based on observations of the system response. In addition, many useful identification algorithms have also been derived for the detection of the system structure, determination of the unknown system parameters, and the removal or modelling of noise. However, identifying a model of a spatio-temporal system from observations is still far from straightforward. Among the major difficulties that remain, two will be given special attention in this paper. The first arises from multiscale problems, and the second is associated with noise on the observations.

Multiscale phenomena have been widely observed in many diverse fields including physics, chemistry, biology, ecology, and network traffic systems (Mandelbrot 1967, Car and Parrinello 1985, Littlewood and Maniatty 2005, Lorenz 1996, Chaudhari, Yan, and Lee 2003, Bindal, Khinast, and Ierapetritou 2003, Muller-Buschbaum, Bauer, Pfister, Roth, Burghammer, Riekel, David, and Thiele 2006, Louie and Kolaczyk 2006, Huerta, Rabinovich, Abarbanel, Bazhenov 1997, Feldmann, Gilbert, Willinger, and Kurtz 1998, Eck 2004). There are also a variety of multiscale systems including chaotic systems, molecular dynamics, and solar systems. Due to the presence of different scales, modelling, analysis and prediction of multiscale problems often involves the problem of how to describe the interactions between the different laws of physics at different scales, how to analyse the properties of such systems both qualitatively and quantitatively, how to cope with irregularly sampled data, and how to obtain exact/approximate solutions either analytically or numerically. All these problems require new methods and tools to find a solution. Existing methods of modelling, analysis and computation of multiscale systems include Fourier analysis, multigrid methods, domain decomposition methods, fast multipole methods, adaptive mesh refinement methods, wavelet-based methods, homogenisation methods, quasi-continuum methods, and the Heterogeneous Multiscale Method (HMM) (see reviews given by E and Engqiust (2003), E, Engquist, Li, Ren, and Vanden-Eijnden (2006) and references therein). Considering the difficulties of modelling such systems, it would be advantageous if a model could be identified from the observed multiscale data. The model could then be used for the analysis of system behaviour or in control. To the best of our knowledge, there is currently very little research on the identification problem for multiscale systems from observed data and this important problem will therefore be studied in this paper.

Noise is another important issue in system identification. It is well known that the presence of noise on the measured data will affect the accuracy of the identified models for both multiscale systems and conventional systems. In some cases filtering techniques can be used to improve the performance of the identification. However, modelling with directly filtered data does not usually provide satisfactory identification results because applying data filtering without taking into account the dynamical properties of the original systems may result in the removal of important features from the data, this is particularly severe for nonlinear and multiscale data because the signals generated by these systems generally contain features and noise that have varying contributions over both time, space and frequency.

The authors are with the Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield, S1 3JD, UK.

In this paper, a new solution to the identification problem of spatio-temporal dynamical systems directly from observations is proposed to tackle the above mentioned two difficulties. The idea behind the proposed method is that an infinite dimensional spatio-temporal system is projected onto a finite dimensional subspace using a wavelet multiresolution analysis so that the system can be viewed at different scales. It has been shown in Coca and Billings (2002) that the wavelet coefficients at a scale form a finite dimensional system of ordinary differential equations which can be used to represent an finite dimensional approximation at some scale of the original system . The system dynamics at a specific scale of interest can then be identified using an orthogonal forward regression (OFR) least-squares algorithm (Billings, Chen, and Kronenberg 1989). It is shown that this model can then be converted between different scales to produce predictions of the system outputs at different scales using wavelet decomposition and reconstruction techniques. Because of the filtering properties of wavelets it is also shown that the new method naturally combines the identification procedure with a filtering process.

Section 2 presents a multiscale representation of spatio-temporal dynamical systems using a wavelet multiresolution analysis. In section 3, the identification method and the implementation strategy are presented including a discussion about the properties of the OFR algorithm. Section 4 illustrates the proposed approach using some examples. Finally conclusions are drawn in section 5.

## II. A MULTISCALE REPRESENTATION OF SPATIO-TEMPORAL DYNAMICAL SYSTEMS

Consider the following evolution equation of a spatio-temporal dynamical system

$$\frac{du}{dt} + Lu = f, u(0) = u_0 \tag{1}$$

where $L : V \rightarrow V^*$ is a nonlinear operator with $V \subset L^2(\Omega)$ a Sobolev space, and where $\Omega \subset R^n$ is a nice spatial domain. The evolution of equation (1) can represent a partial differential equation model for both conventional and multiscale spatio-temporal dynamical systems.

To generate a multiscale representation of the system (1), let $V_j \subset V, j \in Z$ be a multiresolution approximation of the space $V$. That is, $V_j, j \in Z$ is an increasing sequence of closed subspaces of $L^2(\Omega)$ with the following properties (Chui, 1992)

1) $V_j \subset V_{j+1}$,
2) $f(x) \in V_j \Longleftrightarrow f(2x) \in V_{j+1}, j \in Z$,
3) $\bigcup_{j\in Z} V_j$ is dense in $L^2(R^n)$ and $\bigcap_{j\in Z} V_j = \emptyset$,
4) A scaling function $\phi(x) \in V_0$ exists such that the set $\{\phi(x-k)|k \in Z^n\}$ forms a Riesz basis of $V_0$.

Following the definition of the multiresolution analysis, the set of functions $\{\phi_{j,k} = 2^{j/2}\phi(2^j x - k)\}$ is a Riesz basis of $V_j$. Let $W_j$ be a complementary space of $V_j$ in $V_{j+1}$, such that $V_{j+1} = V_j \bigoplus W_j$. Consequently

$$\bigoplus_{j\in Z} W_j = L^2(R^n). \tag{2}$$

$W_j$ is called a wavelet subspace. A function $\psi(x)$ is a wavelet if the set of functions $\{\psi(x-k)|k \in Z^n\}$ is a Riesz basis of $W_0$. It follows that the set of wavelet functions $\{\psi_{j,k} = 2^{j/2}\psi(2^j x - k)\}$ is a Riesz basis of $L^2(R^n)$.

At resolution $j$ the projection $P_j$ (resp. $Q_j$) of a function $f$ onto $V_j$ (resp. $W_j$) that corresponds to the above split of $L^2(R^n)$ can be written with the use of a dual scaling function $\tilde{\phi}$ (resp. dual wavelet function $\tilde{\psi}$) as follows

$$\begin{aligned} P_j f(x) &= \sum_k < f, \tilde{\phi}_{j,k} > \phi_{j,k}(x) \\ Q_j f(x) &= \sum_k < f, \tilde{\psi}_{j,k} > \psi_{j,k}(x) \end{aligned} \tag{3}$$

where $< \cdot >$ denotes the inner product. Such wavelets are called biorthogonal wavelets. Generally, $\tilde{\phi} \neq \phi$ and $\tilde{\psi} \neq \psi$ except when orthogonality holds. The definition of a multiresolution analysis implies that for any $f(x) \in L^2(\Omega)$

$$\begin{aligned} &\lim_{j\rightarrow\infty} P_j f(x) = f(x) \\ &f(x) = \sum_j Q_j f(x). \end{aligned} \tag{4}$$

Since $W_j$ is the complementary subspace of $V_j$ in $V_{j+1}$, that is $V_{j+1} = V_j \bigoplus W_j$, it follows that $P_{j+1}f(x) = P_j f(x) + Q_j f(x)$. This gives an alternative representation of the projection of a function $f \in L^2(\Omega)$ at resolution $j+1$ using both the scaling and wavelet functions as

$$P_{j+1}f(x) = \sum_k < f, \tilde{\phi}_{j,k} > \phi_{j,k}(x) + \sum_k < f, \tilde{\psi}_{j,k} > \psi_{j,k}(x). \tag{5}$$

In this way, it is understood that the projection $P_j$ provides an approximation of the function $f$ at some resolution $j$ and the details left by this approximation are contained in $Q_j$. By iteration, a wavelet decomposition can be obtained as follows

$$\begin{aligned} P_{j+1}f(x) &= \sum_k < f, \tilde{\phi}_{j-l,k} > \phi_{j-l,k}(x) \\ &\quad + \sum_{i=j-l}^{j} \sum_k < f, \tilde{\psi}_{i,k} > \psi_{i,k}(x). \end{aligned} \tag{6}$$

It should be noted that a discretisation of (1) can be considered as replacing (1) with an approximation in $V_{j+1}$. The operator $L$ is replaced by a operator $L_{j+1} : V_{j+1} \rightarrow V_{j+1}$ defined by

$$< Lu_{j+1}, w_{j+1} > = < L_{j+1}u_{j+1}, w_{j+1} >, w_{j+1} \in V_{j+1}. \tag{7}$$

Then the discretised dynamics of the system (1) at a scale/resolution $j+1$ can be described as the following initial value problem in $V_j$

$$\frac{du_{j+1}}{dt} + L_{j+1}u_{j+1} = f_{j+1}, u_{j+1}(0) = u_0 \tag{8}$$

where $u_{j+1}(t,x) = P_{j+1}u(t,x) = \sum_k < u(t,x), \tilde{\phi}_{j+1,k}(x) > \phi_{j+1,k}(x)$ and $f_{j+1}(t,x) = P_{j+1}f(t,x) = \sum_k < f(t,x), \tilde{\phi}_{j+1,k}(x) > \phi_{j+1,k}(x)$.

Consider the decomposition $V_{j+1} = V_j \oplus W_j$, the projection from $V_{j+1}$ onto $V_j$ and $W_j$ yields

$$\begin{aligned} u_j(t,x) &= P_j u_{j+1}(t,x) = \sum_k u_{j,k}(t)\phi_{j,k}(x) \quad (9) \\ f_j(t,x) &= P_j f_{j+1}(t,x) = \sum_k f_{j,k}(t)\phi_{j,k}(x) \end{aligned}$$

on $V_j$ with $u_{j,k}(t) =< u_{j+1}(t,x), \tilde{\phi}_{j,k}(x) >$ and $f_{j,k}(t) =< f_{j+1}(t,x), \tilde{\phi}_{j,k}(x) >$, and

$$\begin{aligned} u'_j(t,x) &= Q_j u_{j+1}(t,x) = \sum_k u'_{j,k}(t)\psi_{j,k}(x) \quad (10) \\ f'_j(t,x) &= Q_j f_{j+1}(t,x) = \sum_k f'_{j,k}(t)\psi_{j,k}(x) \end{aligned}$$

on $W_j$ with $u'_{j,k}(t) =< u_{j+1}(t,x), \tilde{\psi}_{j,k}(x) >$ and $f'_{j,k}(t) =< f_{j+1}(t,x), \tilde{\psi}_{j,k}(x) >$, which satisfies

$$\begin{aligned} u_{j+1}(t,x) &= u_j(t,x) + u'_j(t,x) \quad (11) \\ f_{j+1}(t,x) &= f_j(t,x) + f'_j(t,x). \end{aligned}$$

Applying the projections $P_j$ and $Q_j$ to eqn. (8) generates the following two equations

$$\begin{aligned} \frac{du_j}{dt} + P_j L_{j+1}(u_j + u'_j) &= f_j \quad (12) \\ \frac{du'_j}{dt} + Q_j L_{j+1}(u_j + u'_j) &= f'_j. \end{aligned}$$

Following the Corollary 3.1 (Coca and Billings 2002), there exists a finite dimensional system of ordinary differential equations with the wavelet coefficients $\{u_{j,k}, u'_{j,k}\}_j$ and $\{f_{j,k}, f'_{j,k}\}_j$ as the output and input of the system such that $u_j(t,x) = \sum_k u_{j,k}(t)\phi_{j,k}(x)$ forms a finite dimensional approximation at the scale $j$ of the solution of the original system. Furthermore, this finite dimensional system of ordinary differential equations can generally be decoupled into two effective equations: one describes the coarse dynamics at scale $j$

$$\frac{du_{j,k}}{dt} = F_j(u_{j,k}(t), f_{j,k}(t), f'_{j,k}(t)) \quad (13)$$

and the second describes the corresponding detailed dynamics for this scale

$$\frac{du'_{j,k}}{dt} = F'_j(u'_{j,k}(t), f_{j,k}(t), f'_{j,k}(t)). \quad (14)$$

By repeating the above decomposition process, a multiscale representation as eqns. (13) and (14) of the original spatio-temporal dynamical system is obtained.

It should be noted that eqn. (13) represents an approximation to the original dynamical systems at scale $j$. The presence of the details $f'_{k,j}$ in eqn. (13) indicates that the influence from a finer scale is accommodated within this model.

## III. Multiscale identification of spatio-temporal dynamical systems

For spatio-temporal dynamical systems, experimental measurements are often available in the form of a series of snapshots $u(x, n\Delta t)$, $n = 0, 1, 2, \cdots$, $x \in \Omega$, where $\Delta t$ is the time sampling interval. Assume the system to be considered is spatially sampled at a sampling interval $\Delta x$, then the observations are discrete measurements $u(x_i, t_j)$ both in time and space. This is equivalent to replacing the original system in the infinite dimensional space $V$ with an approximation in some finite dimensional subspace $V_j$. The objective of the identification is to obtain one model for the system from these observations.

### A. Comments on multiscale identification problems

As mentioned earlier, the objective of multiscale identification for spatio-temporal dynamical systems is to obtain an effective model or a set of models from observed spatio-temporal patterns. Eqns. (13) and (14) at any scale can be used to describe the underlying spatio-temporal dynamics at that specific scale once the representation is estimated or extracted by using some identification algorithm. Ideally, the identification technique should be able to produce a concise model structure with a low spatial dimension. This ensures that the obtained model is parsimonious and can readily be interpreted either for simulation or analysis. Once eqns. (13) and (14) are approximated by some function space such as polynomials, they will be in the form of a linear-in-the-parameters model, therefore, theoretically any least-squares-type algorithm can be employed to produce a model. However, there are several problems related to multiscale identification which also need to be addressed:

1) Choosing the proper approximation subspace $V_j$, that is the mesh size over a spatial domain is very important. In identification, the mesh size represents the sampling period in the spatial domain and reflects the number of measurement locations and determines the scale of interest. In the case that the mesh size can be made sufficiently small, a system model can theoretically be identified using the data from the finest scale, which can resolve the dynamical behaviours at all scales. However, the finer the mesh size is, the higher the dimension of the approximation subspace becomes, that means the dimension of the resulting finite dimensional model will be extremely large and computationally expensive if not formidable. In this case, it is worth considering projecting the data into a coarser scale to obtain a lower dimensional and effective model. In some other cases, it will be of interest to construct system equations on a coarse scale that account for the contributions from these finer scales, such as in molecular dynamics. These requirements translate into the need to identify an effective and economical model for the coarse scale with a lower dimension.
2) Often the measured data from the system of interest can not be obtained at the finest scale. In this case, it is impossible to identify a system model at the finest

scales directly from the measurements. However, the observations from the best obtainable scale can be used to identify a model at the specific scale, which represents a coarse behaviour of the original system and it would be beneficial if an equivalent finer model could be obtained whose solutions have the same coarser behaviour as the original unknown complicated systems.

3) The presence of noise on the measured data can affect the accuracy of the identified models.

To overcome these problems, in this paper a new approach is proposed by applying the Orthogonal Forward Regression (OFR) least-squares algorithm to data at some available scale. The proposed method has the following characteristics.

1) The proposed method does not assume any a priori knowledge regarding the structure or parameters of the underlying spatio-temporal dynamical system.
2) The measured data are not limited to the finest scale. In the case that the system can be observed at the finest scale, the proposed method can be used to obtain a simpler model at some coarse scale and this identified model can be used to predict the outputs of the system at different scales including the finest scale. In the case that the system can not be measured at its finest scale, a coarser model can be identified at some observable scale, by which an approximation of the finest scale behaviour can be made.
3) The multiscale representation using wavelets provides a natural filter for systems with measurement noise. This is because the coarser the data are, the less the noise is.
4) The OFR least-squares algorithm can effectively determine the model structure and provide parameter estimates in a forward term selection manner.

In what follows, the multiscale identification method and some simulation examples will be presented.

The multiscale identification algorithm proposed in this paper can be summarised as follows.

1) Perform a multiresolution analysis for the measured data to obtain a multiscale representation of the system.
2) Choose a suitable scale and apply the OFR algorithm to generate a model structure and parameter estimates.
3) Using wavelet decomposition and reconstruction methods predict the outputs of the system at different scales.

### *B. The OFR algorithm*

Given a set of (candidate terms) basis functions from a regressor class, the objective of the identification algorithm is to select the significant terms from this set while estimating the corresponding monomial coefficients. In this paper, the OFR least-squares algorithm is applied to a set of polynomial basis functions. The OFR algorithm involves a stepwise orthogonalisation of the regressors and a forward selection of the relevant terms based on the Error Reduction Ratio criterion. The algorithm provides the optimal least-squares estimate of the polynomial coefficients.

Formally, the classical OFR least-squares algorithm can be stated as follows (Billings, Korenberg, and Chen 1988).

Consider the following linear relationship

$$y(t) = \sum_{i=1}^{n} \theta_i p_i(t) + \xi(t) \tag{15}$$

where $p_i(t)$ are regressors and $y$ is the dependent variable.

Let $p_0(t), p_1(t), \cdots, p_n(t)$ and $y(t)$, $t = 1, 2, \cdots, N$ be the series of observations. Denote $Y = (y(1), y(2), \cdots, y(N))^T$ and $P_i = (p_i(1), p_i(2), \cdots, p_i(N))^T$, $i = 0, 1, \cdots, n$, then the following linear regression model can be formed

$$Y = P\theta + \Xi \tag{16}$$

where $P = (P_0, P_1, \cdots, P_N)$ is the regression matrix, $\theta = (\theta_1, \theta_2, \cdots, \theta_n)^T$ represents the unknown parameters to be estimated, and $\Xi = (\xi(1), \xi(2), \cdots, \xi(N))^T$ is some modelling error vector.

Assume that the matrix $P^T P$ is symmetric and positive definite, the matrix decomposition theorem states that the matrix $P^T P$ can be repressed as

$$P^T P = A^T D A \tag{17}$$

where $A$ is a unit upper-triangle matrix with $A^{-1}A = I$ and $D$ is diagonal with all positive elements. Then (16) can be written as follows

$$Y = P\theta + \Xi = P(A^{-1}A)\theta + \Xi = XA\theta + \Xi = Xg + \Xi \tag{18}$$

where $X = PA^{-1}$ is an $N \times (n+1)$ matrix with orthogonal columns $X_i$ such that $X^T X = D$, and $g = A\theta$. The orthogonal least squares solution $\hat{g}$ to (18) is then given by

$$\hat{g} = D^{-1} X^T Y. \tag{19}$$

The parameters $\hat{g}$ and $\hat{\theta}$ satisfy the triangular equation

$$A\hat{\theta} = \hat{g}. \tag{20}$$

Because of the orthogonality properties, the term $X_i$ and the quantity $g_i$ can be calculated in an independent manner. This is achieved by applying the algorithm in a forward way with the error reduction ratio (ERR) as selection criterion at each step. The ERR caused by term $i$, $i = 0, 1, \cdots, n$ is defined as

$$ERR_i = \frac{g_i^2 X_i^T X_i}{Y^T Y}. \tag{21}$$

### *C. Properties of the OFR least-squares algorithm*

In order to discuss the properties of the OFR algorithm, some assumptions are needed.

**Assumption 1:** There is no undermodelling, that is, the input-output data was generated by some true dynamic system that can be represented by (15) with the parameters $\theta_i, i = 1, \cdots, n$, and the system is input-output bounded uniformly with probability one.

**Assumption 2:** $\xi(t)$ is a zero mean white sequence with a finite covariance $\sigma^2$ and is uncorrelated with $p_i(t), i = 1, \cdots, n$.

**Assumption 3:** All processes involved are (jointly) ergodic of at least second order and the inputs are persistently exciting of sufficiently high order with probability one.

**Assumption 4:** The matrix $P^TP$ is symmetric and positive definite.

From (19)

$$\hat{g} = D^{-1}X^TY = D^{-1}X^T(Xg + \Xi) = g + D^{-1}X^T\Xi. \quad (22)$$

Let $\tilde{g} = \hat{g} - g$, then it is easy to show that under Assumptions 1 and 2 the estimates will be unbiased $E\{\hat{g}\} = g$ (Korenberg, Billings, Liu, and Mcilroy 1988).

In what follows, the convergence issue of the OFR algorithm will be addressed. Here, $\hat{g}$ will be written as $\hat{g}^{[N]}$ to indicate that there are $N$ observations used for identification. Note that the covariance of the parameter error $\tilde{g}^{[N]}$ is given by

$$cov(\tilde{g}^{[N]}) = E\{(\hat{g}^{[N]} - g)(\hat{g}^{[N]} - g)^T\} = \sigma^2(X^TX)^{-1}. \quad (23)$$

From (23) and $P^TP = A^TX^TXA$ with $A$ a unit upper-triangle matrix, the following result can be given

**Theorem 1.** Under Assumptions 1 and 2,

$$\lim_{N\to\infty} \|\tilde{g}^{[N]}\| = 0 \quad (24)$$

if and only if

$$\lim_{N\to\infty} \underline{\lambda}(P^TP) = \infty \quad (25)$$

or equivalently

$$\lim_{N\to\infty} \underline{\lambda}(X^TX) = \infty \quad (26)$$

where $\|\cdot\|$ denotes the Euclidean norm and $\underline{\lambda}(M)$ is the minimum eigenvalue of the matrix $M$.

**Theorem 2.** Under Assumption 1 to 4 $lim_{N\to\infty}\hat{g}^{[N]} = g$ with probability one.

Outline of the proof. Assumptions 1 and 3 mean the system is input-output bounded uniformly and there is no any particular eigenvector of $P^TP$ along with the energy of the system is concentrated with probability one and it follows that the ratio of the largest to the smallest eigenvalues of $P^TP$ is bounded uniformly in $N$ with probability one. Then according to Aderson and Taylor (1979), it only needs to prove that $\lim_{N\to\infty}\underline{\lambda}(P^TP) = \infty$ with probability one. This is equivalent to showing $\lim_{N\to\infty}\underline{\lambda}(X^TX) = \infty$ with probability one. Note that $X^TX$ is a diagonal matrix as

$$X^TX = \begin{bmatrix} \sum_{t=1}^N X_0^2(t) & & & \\ & \sum_{t=1}^N X_1^2(t) & & \\ & & \ddots & \\ & & & \sum_{t=1}^N X_n^2(t) \end{bmatrix} \quad (27)$$

Note that $\sum_{t=1}^N X_i^2(t), i = 0, 1, \cdots, n$ are the eigenvalues of $P^TP$. Without loss of generality, let $\sum_{t=1}^N X_0^2(t)$ be the minimum eigenvalue of $X^TX$. Suppose $\lim_{N\to\infty}\sum_{t=1}^N X_0^2(t) < \infty$ with probability one, then there must exist an integer $T > 0$ such that $X_0(t) = 0$ for all $t \geq T$ with probability one, which is in contradiction with Assumption 3.

### *D. Predictions between different scales*

Once a system model at certain scale is identified, it can be used to calculate the wavelet coefficients at different scales in terms of the wavelet decomposition and reconstruction method.

Decompostion $V_j = V_{j-1} \oplus W_{j-1}$:

$$u_{l,j-1}(t) = \sum_k h_{k-2l}u_{k,j}(t) \quad (28)$$
$$u'_{l,j-1}(t) = \sum_k g_{k-2l}u_{k,j}(t)$$

where $h$ and $g$ are the wavelet filter coefficients.

Reconstruction $V_{j+1} = V_j \oplus W_j$:

$$u_{l,j+1}(t) = \sum_k h_{l-2k}u_{k,j}(t) + \sum_k g_{l-2k}u'_{k,j}(t) \quad (29)$$

where $h$ and $g$ are the wavelet filter coefficients.

Note that if the details at all scales are available, then the reconstruction process (29) can be carried out up to the finest scale. If the details are not available, the reconstruction process can still be performed by just ignoring all the details at finer scales. In this case, the obtained prediction is an approximation of the original behaviours at these finer scales.

### *E. Noise reduction analysis*

The presence of noise in the observations generally prevents the OFR algorithm from selecting the correct terms in the model and consequently can produce erroneous parameter estimates. The relationship between the values of $ERR_i$ and the signal-to-noise ratio has been analysed in detail in Guo and Billings (2007). Here a brief discussion is given.

Note that from the definition of ERR (21), it can be observed that the OFR is equivalent to maximising the product moment correlation coefficient. In fact, the product moment correlation coefficient $\rho_i$ of term $i$ satisfies

$$\rho_i^2 = \frac{cov(Y, W_i)^2}{var(Y)var(W_i)} = ERR_i \quad (30)$$

Let $\bar{\rho}_i$ be the correlation coefficient of the term $i$ obtained from noisy data, then

$$\begin{aligned} \bar{\rho}_i &= \frac{\eta_Y}{\sqrt{\eta_Y^2+1}}\sqrt{\frac{< W_i, W_i >}{< \bar{W}_i, \bar{W}_i >}}\rho_i \quad (31) \\ &+ \sum_{j=0}^{i-1}\frac{< P_i, W_j >}{< W_j, W_j >}\frac{\eta_Y}{\sqrt{\eta_Y^2+1}}\sqrt{\frac{< W_j, W_j >}{< \bar{W}_i, \bar{W}_i >}}\rho_j \\ &- \sum_{j=0}^{i-1}\frac{< \bar{P}_i, \bar{W}_j >}{\bar{W}_j, \bar{W}_j}\sqrt{\frac{< \bar{W}_j, \bar{W}_j >}{< \bar{W}_i, \bar{W}_i >}}\bar{\rho}_j \end{aligned}$$

where $\bar{P}$ and $\bar{W}$ are the noisy version of $P$ and $W$. $\eta_Y = var(Y)/\sigma_\varepsilon^2$ with $\varepsilon$ the output noise. Note that $\eta_Y = var(Y)/\sigma_\varepsilon^2$ can be considered as the signal-to-noise ratio. From (31) it can be observed that when the signal-to-noise ratio is sufficiently high, then $\bar{\rho}$ will be sufficiently close to the true value $\rho$. Therefore it is highly desirable to have a signal with a high signal-to-noise ratio. Wavelet decomposition provides such a way to reduce the noise from observations and therefore increase the signal-to-noise ratio. By applying a wavelet multiresolution analysis to the data space, a set of coarser signals with high signal-to-noise ratios can be obtained. It follows that an accurate model can be obtained from data at a coarse scale. It can be expected that better predictions of the system behaviour at different scales can be given by using such an accurate model.

## IV. Simulation studies

### A. Example 1 - Non-homogeneous wave equation

Consider the following non-homogeneous wave equation

$$\frac{\partial^2 u(t,x)}{\partial t^2} - C\frac{\partial^2 v(t,x)}{\partial x^2} = f(t,x), x \in [0,1] \qquad (32)$$

with initial conditions

$$\begin{aligned} u(0,x) &= 0 \qquad (33)\\ \frac{du(0,x)}{dt} &= 4\text{exp}(-x) + \text{exp}(-0.5x) \end{aligned}$$

where

$$f(t,x) = -13\text{exp}(-x)\text{cos}(1.5t) - 9.32\text{exp}(-0.5x)\text{cos}(2.1t) \qquad (34)$$

For $C = 1.0$ the exact solution $u(t,x)$ of the initial value problem (32), (34) is

$$\begin{aligned} u(t,x) &= 4\text{exp}(-x)\text{cos}(1.5t) + 2\text{exp}(-0.5x)\text{cos}(2.1t) \qquad (35)\\ &\quad -4\text{exp}(-x)\text{exp}(-t) - 2\text{exp}(-0.5x)\text{exp}(-0.5t) \end{aligned}$$

The measurement function was taken as

$$y(t,x) = u(t,x) \qquad (36)$$

The reference solution was sampled at $2^4 = 16$ equally spaced points over the spatial domain $\Omega = [0,1], x = \{x_0, \cdots, x_{15}\}$. This sampling process can be viewed as replacing (35) with an approximation in a finite dimensional subspace $V_n, n = 4$. From each location, $1000$ input/output data points sampled at $\Delta t = \pi/105$ were generated. To test the performance of the proposed approach, a sequence of noise with zero-mean and variance $0.01$ is added to the output $y(t,x)$, which results in a signal-to-noise ratio of $26.212$dB. After applying a single level wavelet decomposition with Haar wavelets (other wavelets can be also used), a coarse approximation of the original data in subspace $V_{n-1}$ was obtained with a signal-to-noise ratio of $29.256$dB. Note that the signal-to-noise ratio has been increased for the signal at the coarse scale and this should significantly increase the accuracy of the model. The data are plotted in Fig.(1) and Fig.(2).

| Terms | Estimates | ERR | STD |
|---|---|---|---|
| $y_i(t-1))$ | 1.8166e-001 | 9.9598e-001 | 2.1673e-001 |
| $y_i(t-2))$ | 1.3340e-001 | 5.8750e-004 | 2.0024e-001 |
| $y_{i+1}(t-1))$ | 6.8620e-002 | 6.1609e-004 | 1.8140e-001 |
| $y_{i-1}(t-1))$ | 1.3783e-001 | 4.4141e-004 | 1.6658e-001 |
| $y_{i-1}(t-2))$ | 2.3862e-001 | 1.2535e-004 | 1.6212e-001 |
| $u_i'(t-2)$ | 8.4634e+000 | 1.3364e-005 | 1.6160e-001 |
| $u_i'(t-1))$ | -8.4172e+000 | 1.0446e-003 | 1.1793e-001 |
| $u_i(t-1)$ | -2.0234e-001 | 7.3185e-005 | 1.1429e-001 |
| $u_i(t-2)$ | 1.9207e-001 | 9.5179e-005 | 1.0932e-001 |
| $y_{i+1}(t-2)$ | 1.8647e-001 | 3.4899e-005 | 1.0744e-001 |
| $constant$ | 1.2969e-004 | 1.2338e-009 | 1.0744e-001 |

TABLE I

Example 1: The terms and parameters of the final model at the coarse space $V_{n-1}$

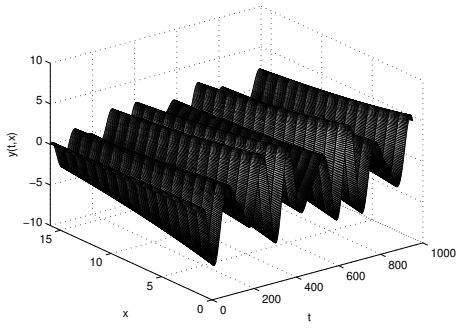


Fig. 1. Example 1: Spatio-temporal signal at the original scale $n = 4$

In this simulation, a model at the coarse scale is to be identified. The inital neighbourhood was selected to be $i-1$ and $i+1$ in the spatial domain and $t-1, t-2$ in the time domain. A set of 200 observations randomly selected among the data set was used for identification. In addition, 200 input and output data $u_i(t), y_{i-1}(t)$ and $y_{i+1}(t)$ from neighbouring locations $i-1, i+1$ acted as inputs during the identification. The identified model using the OFR least squares algorithm with a linear model structure, are listed in Table (I), where ERR denotes the Error Reduction Ratio and STD denotes the standard deviations.

The model predicted output of the identified model is plotted in Fig.(3). Fig.(4) shows the model predicted error between the exact solution and the identified CML model predicted output. Although there are errors between the real output and the model predicted output it is clearly observed that the identified CML model is able to reproduce the system behaviour in $V_{n-1}$ with very high fidelity. Furthermore, the obtained model in Table (I) was used to produce a prediction for the system behaviour at the original scale $n = 4$. The prediction has been done for two cases: one is with the original high frequency details and the other is where the high frequency details have been ignored. The predicted result and error for case one are shown in Figs. (5) and (6). Fig. (7) shows the model prediction without any high frequency information from the identified model.

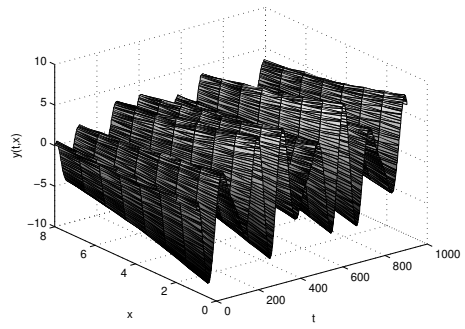


Fig. 2. Example 1: Spatio-temporal signal at single-level coarser scale $n-1$

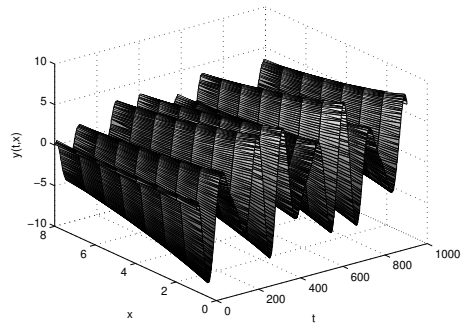


Fig. 3. Example 1: Model predicted output at the coarse scale

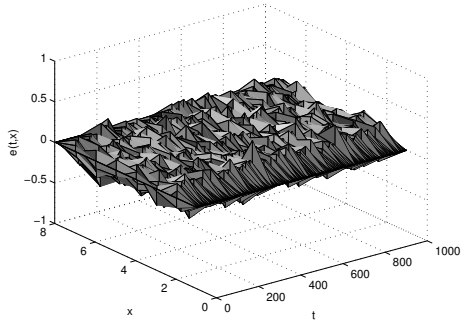


Fig. 4. Example 1: Model predicted error at the coarse scale

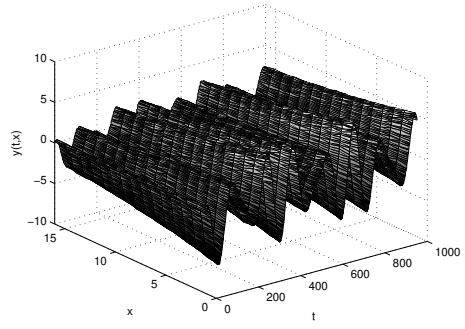


Fig. 5. Example 1: Model predicted output with high frequency detail at the original scale using the identified coarse model

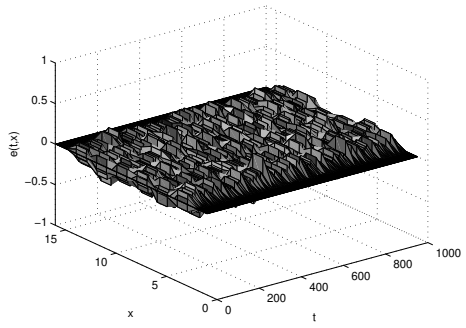


Fig. 6. Example 1: Model predicted error with high frequency detail at the original scale using the identified coarse model

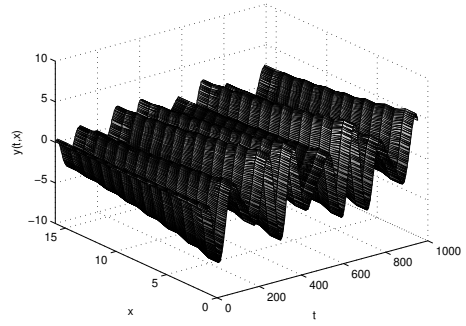


Fig. 7. Example 1: Model predicted output without high frequency detail at the original scale using the identified coarse model

### *B. Example 2 - Two-dimensional deterministic CML*

Consider the following two-dimensional deterministic CML with symmetrical nearest neighbour coupling

$$\begin{aligned} x_{i,j}(t) &= (1-\varepsilon)f(x_{i,j}(t-1)) + \frac{\varepsilon}{4}(f(x_{i,j-1}(t-1)) \\ &\quad + f(x_{i,j+1}(t-1)) + f(x_{i-1,j}(t-1)) + f(x_{i+1,j}(t-1))) \end{aligned} \tag{37}$$

where $x_{i,j}(t) i,j = 1, \cdots, N$ is the state of the CML located at site $(i,j)$ at discrete time $t$, $\varepsilon$ is the coupling strength, and $N$ is the size of lattice. Periodic boundary conditions, that is $x_{1,j}(t) = x_{N,j}(t), x_{i,1}(t) = x_{i,N}(t)$ for all $i,j$ and $t$, are used throughout this study. The evolution of the CML on the lattice sites is governed by the local map $f$, which is generally a nonlinear function chosen to be the logistic map in this simulation

$$f(x) = 1 - ax^2. \tag{38}$$

This model has been extensively studied. It has been observed that for small $\varepsilon$ ($< 0.3$) the system evolves from a frozen random state to pattern selection and to fully developed spatio-temporal chaos via spatio-temporal intermittency. For stronger coupling $\varepsilon > 0.3$ neither a frozen random pattern nor a pattern selection regime is formed which implies there are no pattern changes in this case (Kaneko 1989).

The model (37) with (38) was simulated for a lattice of the size $50 \times 50$ with random initial conditions, periodic boundary conditions, and parameters $\varepsilon = 0.4, a = 1.55$.

| Terms | Estimates | ERR | STD |
|---|---|---|---|
| $y_{i,j}(t-2)$ | -3.1531e-001 | 8.7128e-001 | 3.9711e-001 |
| $y^2_{i,j}(t-2)$ | 3.7868e-001 | 3.4649e-002 | 3.4637e-001 |
| $y_{i,j}(t-1)y_{i,j}(t-2)$ | 2.0228e-001 | 3.0380e-002 | 2.7692e-001 |
| $y_{i,j}(t-1)y_{i-1,j}(t-2)$ | -3.3424e-002 | 2.1891e-002 | 2.3042e-001 |
| $y_{i,j+1}(t-1)y_{i+1,j}(t-1)$ | -3.8419e-002 | 5.2035e-003 | 2.1488e-001 |
| $constant$ | 1.2349e+000 | 4.0395e-003 | 2.0377e-001 |
| $y^2_{i,j}(t-1)$ | -4.2798e-001 | 2.0705e-002 | 1.2293e-001 |
| $y^2_{i,j-1}(t-1)$ | -4.1401e-002 | 1.4045e-003 | 1.1542e-001 |
| $y^2_{i-1,j}(t-2)$ | 5.0840e-002 | 1.2054e-003 | 1.0855e-001 |
| $y^2_{i+1,j}(t-1)$ | -3.2564e-002 | 8.0166e-004 | 1.0373e-001 |

TABLE II

EXAMPLE 2: THE TERMS AND PARAMETERS OF THE FINAL MODEL AT THE COARSE SCALE

The observation variable was set to be $y_{i,j} = x_{i,j}$. Some snapshot patterns are shown in Fig.(8). With these parameters, the system is actually in a chaotic regime with a maximal Lyapunov exponent $\lambda_1 = 0.016046$, which was calculated from the spatial average values of the snapshots by using a numerical algorithm proposed by Rosenstein, Collins, and De Luca (1993). A double-level wavelet decomposition was applied to the data with Haar wavelets as a basis, the obtained coarse snapshots are shown in Fig. (9) at the first level and Fig. (10) at the second level. For the coarser data, the numerically calculated maximal Lyapunov exponent is $0.016046$ at the first scale which is the same as the one calculated from the finer data and $0.018615$. This shows that chaotic systems are essentially self-similar and multiscale.

In the identification, the same set of 200 observation pairs randomly selected among the coarse data set were used. The neighbourhood was set to be the nearest four sites, that is, $(i, j-1), (i, j+1), (i-1, j)$, and $(i+1, j)$ and the time lag was set to be 2. The identified model is listed in Table (II)

The model predicted snapshots of the identified model are plotted in Fig.(11) and the maximal Lyapunov exponent for the model predicted data is $0.014055$ which is quite close to the value $0.016946$. The obtained model in Table (II) was also used to produce predictions for the system behaviour at a finer scale and a coarser scale. The results are shown in Figs. (12) to (14). The corresponding maximal Lyapunov exponents are $0.014055$ and $0.014055$ for the finer scale predictions with and without high frequency details, and $0.015152$ for the coarser scale prediction. From the simulation results it can be observed that the identified model is able to reproduce the chaotic behaviour at different scales of the underlying spatio-temporal system with high performance.

## V. CONCLUSIONS

A novel approach to the multiscale identification of spatio-temporal dynamics has been introduced. It has been demonstrated that the wavelet multiresolution analysis provides a powerful approximation tool for the multiscale representation of the spatio-temporal dynamics. It is also shown that it is possible to extract a system model at some coarse scale, which can then be used to produce predictions for the system behaviour at different scales with or without high frequency

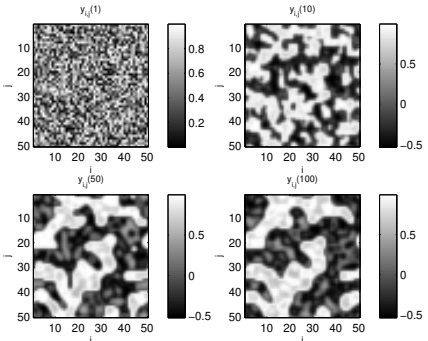

Fig. 8. Example 2: Some snapshots of data (at $t = 1, 10, 50$, and 100)

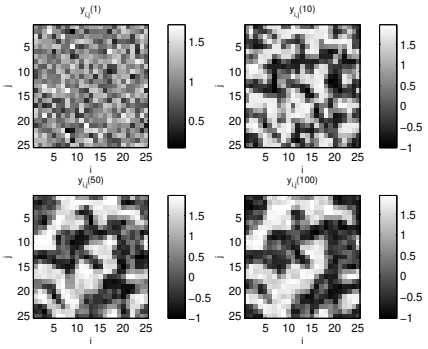

Fig. 9. Example 2: Some snapshots of data (at $t = 1, 10, 50$, and 100) at the first coarse scale

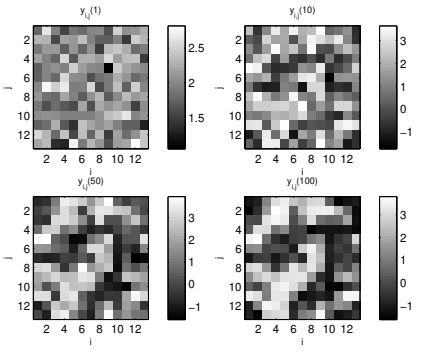

Fig. 10. Example 2: Some snapshots of data (at $t = 1, 10, 50$, and 100) from the second-level wavelet decomposition

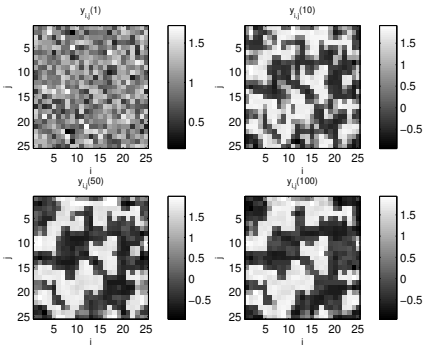

Fig. 11. Example 2: Some snapshots of model predicted outputs (at $t = 1, 10, 50$, and 100) at the first coarse scale

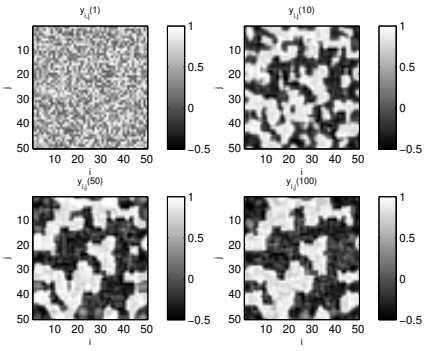

Fig. 12. Example 2: Some snapshots of reconstructed outputs (at $t = 1, 10, 50$, and 100) with details at the original scale using the identified coarse model

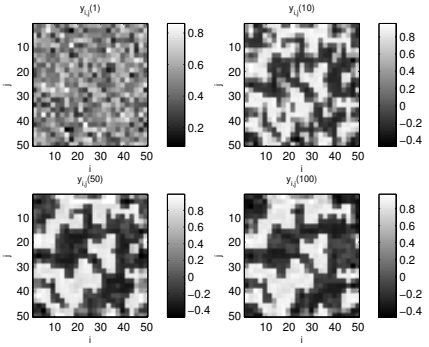

Fig. 13. Example 2: Some snapshots of reconstructed outputs (at $t = 1, 10, 50$, and 100) without details at the original scale using the identified coarse model

information of the original dynamics. The proposed approach can not only generate a simple, effective model for the system but can significantly reduce the noise contained in the signals. Simulation results were included to demonstrate that the new wavelet-based identification procedure can produce excellent models with a very good model predictive performance.

## ACKNOWLEDGMENT

The authors gratefully acknowledge financial support from EPSRC (UK).

## REFERENCES

[1] Aderson, T. W. and Taylor, J. B., (1979) Strong consistency of the least squares estimates in dynamic models, *The annals of Statistics*, Vol. 7, No.3, pp. 484-489.

[2] Billings,S. A., Chen, S., and Kronenberg, M. J. (1988) Identification of MIMO nonlinear systems using a forward-regression orthogonal estimator, *Int. J. Contr.*, Vol. 49, No. 6, pp. 2157-2189, 1989.

[3] Billings, S. A., Guo, L. Z., and Wei, H. L., (2006) Identification of coupled map lattice models of spatio-temporal patterns using wavelets, *Int. J. Syst. Sci.*, Vol. 37, No. 14-15, pp.1021-1038.

[4] Bindal, A., Khinast, J. G., and Ierapetritou, M. G., (2003) Adaptive multiscale solution of dynamical systems in chemical processes using wavelets, *Computers and Chemical Engineering*, Vol. 27, pp. 131-142.

[5] Car, R. and Parrinello, M., (1985) Unified approach for molecular dynamics and density-functional theory, *Physical Review Letters*, Vol. 55, No. 22, pp. 2471-2474.

[6] Chaudhari, A., Yan, C., and Lee, S. L., (2003) Multifractal scaling analysis of autopoisoning reaction over a rough surface, *Journal of Physics A: Mathematical and General*, Vol. 36, pp. 3757-3772.

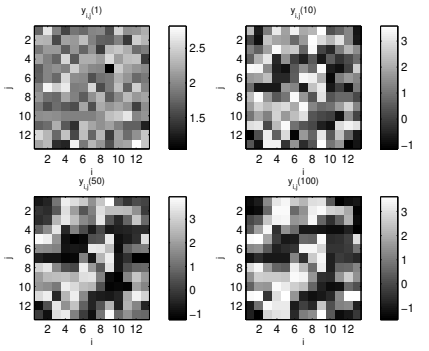

Fig. 14. Example 2: Some snapshots of reconstructed outputs (at $t = 1, 10, 50$, and 100) at second coarse scale using the identified coarse model

[7] Chui C. K., (1992) *An Introduction to Wavelets*, San Diego: Academic Press, Inc.

[8] Coca, D. and Billings, S. A., (2001) Identification of coupled map lattice models of complex spatio-temporal patterns, *Phys. Lett.*, A287, pp. 65-73.

[9] Coca, D. and Billings, S. A., (2002) Identification of finite dimensional models of infinite dimensional dynamical systems, *Automatica*, Vol. 38, pp. 1851-1856.

[10] E., W. N., Engquist, B., Li, X., Ren, W., and Vanden-Eijnden, E., (2006) The heterogeneous multiscale method: A review, preprint, http://www.math.princeton.edu/multiscale/review.pdf

[11] E., W. N., and Engquist, B. (2003) Multiscale modelling and Computation, *Notices of the AMS*, Vol. 50, No.9, pp.1062-1070.

[12] Eck, C., (2004) Analysis of a two-scale phase field model for liquid-solid phase transitions with equiaxed dendritic microstructures, *Multiscale Modeling and Simulation*, Vol. 3, No. 1, pp. 28-49.

[13] Feldmann, A., Gilbert, A. C., Willinger, W., and Kurtz, T. G., (1998) The changing nature of network traffic: scaling phenomena, *Computer Communications Review*, Vol. 28, No. 2, pp. 5-29.

[14] Kaneko, K., (1989) Spatiotemporal chaos in one- and two-dimensional coupled map lattices, *Physica*, D37, pp. 60-82.

[15] Korenberg, M., Billings, S. A., Liu, Y. P., and Mcilroy, P. J., (1988) Orthogonal parameter estimation algorithm for non-linear stochastic systems, Int. J. Contr., Vol. 48, No.1, pp. 193-210.

[16] Guo, L. Z. and Billings, S. A., (2006) Identification of partial differential equation models for continuous spatio-temporal dynamical systems, *IEEE Trans. Circuits and Systems – II: Express Briefs*, Vol. 53, No. 8, pp. 657-661.

[17] Guo, L. Z., Billings, S. A., and Wei, H. L., (2006) Estimation of spatial derivatives and identification of continuous spatio-temporal dynamical systems, *Int. J. Contr.*, Vol.79, No. 9, pp. 1118-1135.

[18] Guo, L. Z. and Billings, S. A., (2007) A modified orthogonal forward regression least-squares algorithm for system modelling from noisy regressors, *Int. J. Contr.*, Vol.80, No. 3, pp. 340-348.

[19] Huerta, R., Rabinovich, M. I., Abarbanel, H. D. I., and Bazhenov,M., (1997) Spike-train bifurcation scaling in two coupled chaotic neurons, *Physical Review E*, Vol. 55, No. 3, pp. R2108-2110.

[20] Littlewood, D. J. and Maniatty, A. M., (2005) Multiscale modelling of crystal plasticity in AL 7075-T651, in *Proceedings of VIII International Conference on Computational Plasticity*, Barcelona, Spain, pp. 618-621.

[21] Lorenz, E. N., (1996) Predictability: a problem partly solved, in *Proc. Seminar on Predictability*, Vol. 1, ECMWF, Reading, Berkshire, UK, pp. 1-18.

[22] Louie, M. M. and Kolaczyk, E. D., (2006) A multiscale method for disease mapping in spatial epidemiology, *Statistics in Medicine*, Vol. 25, No. 8, pp. 1287-1308.

[23] Mandelbrot, B. B., (1967) How long is the coast of Britain? Statistical self-similarity and fractional dimension, *Science*, Vol. 156, pp. 636-638.

[24] McSharrya, P. E., Ellepolab, J. H., von Hardenberga, J., Smitha, L. A., and Kenning, D., (2002) Spatio-temporal analysis of nucleate pool boiling: identification of nucleation sites using non-orthogonal empirical functions, *Int. J. Heat & Mass Transfer*, Vol. 45, No. 2, pp. 237-253.

[25] Muller-Buschbaum, P., Bauer, E., Pfister, S., Roth, S. V., Burghammer, M., Riekel, C., David, C., and Thiele, U., (2006) Creation of multi-scale strip-like patterns in thin polymer blend films, *Europhysics Letters*, Vol. 73, No. 1, pp. 35-41.

[26] Rosenstein, M. T., Collins, J. J., and De Luca, C. J., (1993) A practical method for calculating largest Lyapunov exponents from small data sets, *Physica*, D65, pp. 117-134.

[27] Schwartz, I. B., Morgan, D. S., Billings, L., Lai, Y. C., (2004) Multi-scale continuum mechanics: From global bifurcations to induced high-dimensional chaos, *Chaos*, Vol. 14, No. 2, pp. 373-386.

[28] Voss, H., Bunner, M. J. Bunner, and Abel, M., (1998) Identification of continuous, spatiotemporal systems, *Physical Review E*, Vol. 57, No.3, pp. 2820-2823.

[29] Yan, L. R., Hu, D. W., Zhou Z., and Liu Y., (2004) Spatio-temporal Identification of Hemodynamics in fMRI: A Data-Driven Approach, Chapter in *Medical Imaging and Augmented Reality*, Lecture Notes in Computer Science, Yang G. Z. and Jiang T. Z.(eds), Springer Berlin / Heidelberg, Vol. 3150/2004, pp.213-220.