Wavelet Multiresolution Spatio-Temporal Modelling Using the Integro-Difference Equation

Parham Aram*, Michael Dewar and Visakan Kadirkamanathan, *Member, IEEE*

Abstract

The Integro Difference Equation (IDE) is an increasingly popular model of spatio-temporal processes. Here we develop a Multiresolution Approximation (MRA) framework for the IDE based on semi-orthogonal cardinal B-spline wavelets. State and parameter estimation is approached in a Maximum Likelihood (ML) framework using the Expectation Maximisation (EM) algorithm. Examples are given to demonstrate the features of the model.

Index Terms

Dynamic spatio-temporal modelling, Expectation Maximisation (EM) algorithm, Integro-Difference Equation (IDE), Multiresolution Approximation (MRA), wavelets.

P. Aram* and V. Kadirkamanathan are with the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, S1 3JD, U.K. (e-mail:p.aram@sheffield.ac.uk; visakan@sheffield.ac.uk).

M. Dewar is with the School of Informatics, University of Edinburgh, Edinburgh, EH8 9AB, U.K. (e-mail:mike.dewar@inf.ed.ac.uk)

Wavelet Multiresolution Spatio-Temporal Modelling Using the Integro-Difference Equation

I. Introduction

Complex spatio-temporal behaviour is found in many different systems such as aquatic ecosystems [1], meteorological forecasting [2], air pollution [3], soil temperature [4], thermal processes [5], disease spread [6] and real estate markets [7]. Spatio-temporal modelling aims to represent dynamic system behaviour across time and space.

Techniques for modelling spatio-temporal systems are generating growing interest, both in the applied and theoretical literature. Of particular interest are the class of dynamic spatio-temporal models, including Cellular Automata (CA) [8], Coupled Map Lattices (CMLs) [9], Lattice Dynamical Wavelet Neural Network (LDWNN) [10] and spatially correlated time series [11]–[13] which have all been used in a system identification context. The extent of the spatial interactions in these models is determined by a discrete neighbourhood structure.

Another such model is the Integro-Difference Equation (IDE). These models have received much attention in population ecology as models for the spatio-temporal spread of organisms [14], [15]. IDE models combine discrete temporal dynamics with a continuous spatial representation. The dynamics of this model are given by a spatial mixing kernel, which defines the mapping between the current spatial field and the previous spatial field. The IDE model can be applied to data sampled both regularly and irregularly in space due to the continuous representation of each spatial field. Estimating the underlying spatial field in time is of particular interest, and is not straightforward especially when the field displays structure at different spatial scales. Estimation of the underlying spatial field can be achieved by estimating the associated spatial mixing kernel from measured data, and then using this in a filtering framework.

Wikle [16] describes the IDE using a state-space formulation by decomposing the kernel and the field using a set of spectral basis functions. An alternative approach for the decomposition of the IDE was introduced in [17] where the resulting state and parameter space dimensions are independent of the number of observation locations. This is crucial not only for model parsimony but also has implications for studying systems observed at a high spatial resolution. In this method an Expectation-Maximisation

(EM) algorithm was used to estimate both the field and the kernel from the observed data. A similar approach is adopted by [18] which developed a model selection procedure by considering various spatial scales at which to represent the system.

Behaviour of the spatial field is often of interest at a variety of scales [19]–[22]. At each scale, an appropriate model should be adopted to adequately reflect the underlying phenomenon. The spatial field may have features at multiple scales either naturally or as a consequence of the measurement process [23] and hence modelling methods that represent the field at multiple scales are essential. Multiresolution Approximation (MRA) [24]–[26] provides such a modelling framework in which information at different scales can be represented simultaneously.

In practice, MRA can be implemented using different types of wavelet and scaling functions. Semi-orthogonal B-spline wavelets [27], [28] benefit from distinctive properties which are useful for system identification purposes. They are compactly supported and have a closed-form analytic expression which allows simple manipulations such as integration [29], [30]. This class of wavelets has been successfully applied in approximation schemes such as wavelet networks [31], [32] and wavelet based CML models [33].

In this paper a multiresolution approximation of the IDE based on the model proposed in [17] is presented. B-spline scaling and wavelet functions are used to decompose both the kernel and the spatial field. Maximum likelihood estimates of the derived parametric state-space model are obtained using the EM-based algorithm of [17]. In this model, the kernel and the spatio-temporal dynamics of the field are viewed at different spatial scales. Whereas in [18] different spatial scales were examined and one chosen, the work presented herein represents the system at a number of different spatial scales concurrently.

The rest of this paper is set out as follows: in Section 2, the IDE model is briefly reviewed and the MRA of the IDE is presented. Section 3 introduces the B-spline wavelet IDE model and provides necessary formulation to construct such a model in state-space. Section 4 briefly reviews the EM-based approach for parameter and state estimation. In Section 5, synthetic examples are given to demonstrate the ability of the developed MRA model in estimation of the hidden spatio-temporal field and the spatial mixing kernel. Finally conclusions are drawn in Section 6.

II. MRA OF THE IDE IN STATE-SPACE

The spatially homogeneous, linear IDE is given by

$$z_{t+1}(\mathbf{s}) = \int_{\mathcal{S}} k(\mathbf{s}-\mathbf{r})\, z_t(\mathbf{r})\, d\mathbf{r} + e_t(\mathbf{s}) \tag{1}$$

where $t \in \mathbb{Z}_0$ denotes discrete time, $\mathbf{s}, \mathbf{r} \in \mathcal{S} \subset \mathbb{R}^{n_S}$ are spatial locations in $n_{\mathcal{S}}$-dimensional physical space, where $n_{\mathcal{S}} \in \{1, 2, 3\}$. The continuous spatial field at time t and at location $\mathbf{s}$ is denoted $z_t(\mathbf{s})$. The model dynamics are defined by the homogeneous time invariant spatial mixing kernel, $k(\mathbf{s} - \mathbf{r}) : \mathbb{R}^{n_S} \rightarrow \mathbb{R}$ which maps the current spatial field to the next spatial field via the integral (1). The disturbance $e_t(\mathbf{s})$ is a zero-mean normally distributed noise process, spatially coloured but temporally independent, with covariance

$$cov\left(e_t(\mathbf{s}), e_{t+\tau}(\mathbf{r})\right) = \begin{cases} \eta(\mathbf{s} - \mathbf{r}), & \tau = 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

for any $\tau \in \mathbb{Z}$, where $\eta(\mathbf{s} - \mathbf{r})$ is a spatially homogeneous covariance function. The spatial field is observed by

$$\mathbf{y}_t = \mathbf{z}_t + \mathbf{v}_t \tag{3}$$

where the observation vector $\mathbf{y}_t = [y_t(\mathbf{s}_1) y_t(\mathbf{s}_2) \cdots y_t(\mathbf{s}_{n_y})]^\top$, compiled at n_y spatial locations and at time t, is corrupted by an i.i.d normally distributed zero-mean white noise $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_v)$ with $\mathbf{\Sigma}_v = \sigma_v^2 \mathbf{I}_{n_y}$. It is assumed here that $\mathbf{\Sigma}_v$ and $\eta(\mathbf{s} - \mathbf{r})$ are known. The superscript $\top$ denotes the transpose operator.

Assuming $z_t(.)$ and $k(.)$ are square-integrable functions, MRA of the one-dimensional IDE can be obtained by expanding both the field and the kernel in the integral (1) in terms of the translations and dilations of a scaling function $\phi(s)$ and a mother wavelet $\psi(s)$. For the one-dimensional kernel this is given by,

$$k(s - r) = \sum_{l \in \mathbb{Z}} \alpha_{j_0,l} \phi_{j_0,l}(s - r) + \sum_{j \geq j_0}^{\infty} \sum_{l \in \mathbb{Z}} \beta_{j,l} \psi_{j,l}(s - r) \tag{4}$$

In this expansion, $\alpha_{j_0,l}$ are the approximation coefficients at the lowest scale j_0 and $\beta_{j,l}$ represent the detail coefficients at different scales j, with $\phi_{j,l}(s) = 2^{\frac{j}{2}} \phi\left(2^j s - l\right)$ and $\psi_{j,l}(s) = 2^{\frac{j}{2}} \psi\left(2^j s - l\right)$. Integers j and l are called scale and translation parameters accordingly. The one-dimensional field can similarly be decomposed as

$$z_t(s) = \sum_{l \in \mathbb{Z}} x_{t,j_0,l} \phi_{j_0,l}(s) + \sum_{j \geq j_0}^{\infty} \sum_{l \in \mathbb{Z}} \check{x}_{t,j,l} \psi_{j,l}(s) \tag{5}$$

where $x_{t,j_0,l}$ and $\check{x}_{t,j,l}$ are the coefficients of the expansion at time t. A two-dimensional MRA can be implemented using 2-D scaling and wavelet functions built up using the tensor-product approach [34], [35]. We write the 2-D scaling function as

$$\phi_{j,\mathbf{l}}(s_1, s_2) = \phi_{j,l_1}(s_1) \phi_{j,l_2}(s_2) \tag{6}$$

and three 2-D wavelet functions $\psi^{(n)}(\mathbf{s})$ where $n = 1, 2, 3$ as

$$\psi^{(1)}_{j,\mathbf{l}}(s_1, s_2) = \phi_{j,l_1}(s_1) \psi_{j,l_2}(s_2) \tag{7}$$

$$\psi^{(2)}_{j,\mathbf{l}}(s_1,s_2)=\psi_{j,l_1}(s_1)\,\phi_{j,l_2}(s_2) \tag{8}$$

$$\psi^{(3)}_{j,\mathbf{l}}(s_1,s_2)=\psi_{j,l_1}(s_1)\,\psi_{j,l_2}(s_2) \tag{9}$$

for $j\in\mathbb{Z}$ and, $\mathbf{l}\in\mathbb{Z}^2$, where $\psi^{(1)}_{j,\mathbf{l}},\psi^{(2)}_{j,\mathbf{l}}$ and $\psi^{(3)}_{j,\mathbf{l}}$ extract fine features of the 2-D field at vertical, horizontal and diagonal orientations respectively. This can be generalised to d-dimensional MRA by introducing the scaling function

$$\phi_{j,\mathbf{l}}(s_1,s_2,\cdots s_d)=\phi_{j,l_1}(s_1)\,\phi_{j,l_2}(s_2)\cdots\phi_{j,l_d}(s_d) \tag{10}$$
$$j\in\mathbb{Z},\quad \mathbf{l}\in\mathbb{Z}^d$$

and 2^d-1 wavelet functions $\psi^{(n)}(\mathbf{s}),n=1,\cdots,2^d-1$ with a procedure similar to the 2-D case. Using these basis functions, multi-dimensional MRA of the kernel and the field can be obtained by

$$\begin{aligned}k(\mathbf{s}-\mathbf{r})=&\sum_{\mathbf{l}\in\mathbb{Z}^d}\alpha_{j_0,\mathbf{l}}\phi_{j_0,\mathbf{l}}(\mathbf{s}-\mathbf{r})\\&+\sum_{j\geq j_0}^{\infty}\sum_{\mathbf{l}\in\mathbb{Z}^d}\sum_{i=1}^{2^d-1}\beta^{(i)}_{j,\mathbf{l}}\psi^{(i)}_{j,\mathbf{l}}(\mathbf{s}-\mathbf{r})\end{aligned} \tag{11}$$

$$z_t(\mathbf{s})=\sum_{\mathbf{l}\in\mathbb{Z}^d}x_{t,j_0,\mathbf{l}}\phi_{j_0,\mathbf{l}}(\mathbf{s})+\sum_{j\geq j_0}^{\infty}\sum_{\mathbf{l}\in\mathbb{Z}^d}\sum_{i=1}^{2^d-1}\check{x}^{(i)}_{t,j,\mathbf{l}}\psi^{(i)}_{j,\mathbf{l}}(\mathbf{s}) \tag{12}$$

Equations (11) and (12) are infinite series expansions. For practical implementation, (11) and (12) must be truncated at some level j, therefore we have

$$k(\mathbf{s}-\mathbf{r})\approx\boldsymbol{\theta}^{\top}\boldsymbol{\lambda}(\mathbf{s}-\mathbf{r}) \tag{13}$$

$$z_t(\mathbf{s})\approx\boldsymbol{\mu}(\mathbf{s})^{\top}\mathbf{x}_t \tag{14}$$

where the unknown parameter vector $\boldsymbol{\theta}\in\mathbb{R}^{n_\theta}$ and the unknown state vector $\mathbf{x}_t\in\mathbb{R}^{n_x}$ are defined as

$$\boldsymbol{\theta}^{\top}=\begin{bmatrix}\boldsymbol{\alpha}^{\top}_{j_0} & \boldsymbol{\beta}^{\top}_{j_0} & \boldsymbol{\beta}^{\top}_{j_0+1}\cdots\boldsymbol{\beta}^{\top}_{j}\end{bmatrix} \tag{15}$$

$$\mathbf{x}^{\top}_t=\begin{bmatrix}\mathbf{x}^{\top}_{t,j_0} & \check{\mathbf{x}}^{\top}_{t,j_0} & \check{\mathbf{x}}^{\top}_{t,j_0+1}\cdots\check{\mathbf{x}}^{\top}_{t,j}\end{bmatrix} \tag{16}$$

and where the kernel approximation coefficient vector $\boldsymbol{\alpha}_{j_0}$ and the field approximation coefficient vector $\mathbf{x}_{t,j_0}$ comprise all the coefficients $\{\alpha_{j_0,\mathbf{l}}:\mathbf{l}\in\mathbb{Z}^d\}$ and $\{x_{t,j_0,\mathbf{l}}:\mathbf{l}\in\mathbb{Z}^d\}$ respectively. The kernel detail coefficient vectors $\boldsymbol{\beta}_j$ and the field detail coefficient vectors $\check{\mathbf{x}}_{t,j}$ are defined by

$$\boldsymbol{\beta}^{\top}_j=\begin{bmatrix}\boldsymbol{\beta}^{(1)\top}_j & \boldsymbol{\beta}^{(2)\top}_j\cdots\boldsymbol{\beta}^{(2^d-1)\top}_j\end{bmatrix} \tag{17}$$

$$\check{\mathbf{x}}^{\top}_{t,j}=\begin{bmatrix}\check{\mathbf{x}}^{(1)\top}_{t,j} & \check{\mathbf{x}}^{(2)\top}_{t,j}\cdots\check{\mathbf{x}}^{(2^d-1)\top}_{t,j}\end{bmatrix} \tag{18}$$

where the vector $\boldsymbol{\beta}_j^{(i)}$ contains all the coefficient $\left\{\beta_{j,\mathbf{l}}^{(i)} : \mathbf{l} \in \mathbb{Z}^d\right\}$ and the vector $\check{\mathbf{x}}_{t,j}^{(i)}$ contains all the coefficient $\left\{\check{x}_{t,j,\mathbf{l}}^{(i)} : \mathbf{l} \in \mathbb{Z}^d\right\}$. The vector of the kernel scaling and wavelet functions $\boldsymbol{\lambda} : \mathbb{R}^{n_{\mathcal{S}}} \to \mathbb{R}^{n_\theta}$ and the vector of the field scaling and wavelet functions $\boldsymbol{\mu} : \mathbb{R}^{n_{\mathcal{S}}} \to \mathbb{R}^{n_x}$ are defined by

$$\begin{aligned}\boldsymbol{\lambda}^\top(\mathbf{s}-\mathbf{r}) &= \Big[\boldsymbol{\phi}_{j_0}^\top(\mathbf{s}-\mathbf{r}) \quad \boldsymbol{\psi}_{j_0}^\top(\mathbf{s}-\mathbf{r}) \quad \boldsymbol{\psi}_{j_0+1}^\top(\mathbf{s}-\mathbf{r}) \\ &\qquad \cdots \quad \boldsymbol{\psi}_{j}^\top(\mathbf{s}-\mathbf{r})\Big] \end{aligned} \tag{19}$$

$$\boldsymbol{\mu}^\top(\mathbf{s}) = \Big[\boldsymbol{\phi}_{j_0}^\top(\mathbf{s}) \quad \boldsymbol{\psi}_{j_0}^\top(\mathbf{s}) \quad \boldsymbol{\psi}_{j_0+1}^\top(\mathbf{s}) \cdots \boldsymbol{\psi}_{j}^\top(\mathbf{s})\Big] \tag{20}$$

Each vector of basis functions on the right hand-side of (19) and (20) are constructed in a way similar to their corresponding weight vectors in (15) and (16). Note that the identified IDE model will approximate the underlying spatio-temporal dynamics at different spatial scales. In this way both macroscopic and microscopic behaviour of the system can be represented simultaneously.

The state-space representation of the spatio-temporal IDE proposed in [17] is derived for an arbitrary set of basis functions and hence this derivation can be applied using scaling and wavelet basis functions. Let

$$\boldsymbol{\Lambda}_x = \int_{\mathcal{S}} \boldsymbol{\mu}(\mathbf{s})\, \boldsymbol{\mu}(\mathbf{s})^\top d\mathbf{s} \tag{21}$$

$$\boldsymbol{\Lambda}_\theta = \iint_{\mathcal{S}} \boldsymbol{\mu}(\mathbf{s})\, \boldsymbol{\theta}^\top \boldsymbol{\lambda}(\mathbf{s}-\mathbf{r})\, \boldsymbol{\mu}(\mathbf{r})^\top d\mathbf{r} d\mathbf{s} \tag{22}$$

where $\boldsymbol{\Lambda}_x, \boldsymbol{\Lambda}_\theta \in \mathbb{R}^{n_x \times n_x}$. The spatial IDE model in (1) and (3) can be represented as linear state-space model of the form

$$\mathbf{x}_{t+1} = \mathbf{A}(\boldsymbol{\theta})\, \mathbf{x}_t + \mathbf{w}_t \tag{23}$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{v}_t \tag{24}$$

where

$$\mathbf{A}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}_x^{-1}\boldsymbol{\Lambda}_\theta \tag{25}$$

and

$$\mathbf{C} = \begin{pmatrix} \boldsymbol{\phi}_{j_0}(\mathbf{s}_1) & \boldsymbol{\phi}_{j_0}(\mathbf{s}_2) & \cdots & \boldsymbol{\phi}_{j_0}(\mathbf{s}_{n_y}) \\ \boldsymbol{\psi}_{j_0}(\mathbf{s}_1) & \boldsymbol{\psi}_{j_0}(\mathbf{s}_2) & \cdots & \boldsymbol{\psi}_{j_0}(\mathbf{s}_{n_y}) \\ \boldsymbol{\psi}_{j_0+1}(\mathbf{s}_1) & \boldsymbol{\psi}_{j_0+1}(\mathbf{s}_2) & \cdots & \boldsymbol{\psi}_{j_0+1}(\mathbf{s}_{n_y}) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\psi}_{j}(\mathbf{s}_1) & \boldsymbol{\psi}_{j}(\mathbf{s}_2) & \cdots & \boldsymbol{\psi}_{j}(\mathbf{s}_{n_y}) \end{pmatrix}^\top$$

and where

$$\mathbf{w}_t = \mathbf{\Lambda}_x^{-1} \int_{\mathcal{S}} \boldsymbol{\mu}(\mathbf{s})\, e_t(\mathbf{s})\, d\mathbf{s} \tag{26}$$

is a vector valued, zero-mean normally distributed white noise process with covariance given by (Lemma 1, Appendix)

$$\mathbf{\Sigma}_w = \mathbf{\Lambda}_x^{-1} \mathbf{\Pi} \mathbf{\Lambda}_x^{-\top} \tag{27}$$

and

$$\mathbf{\Pi} = \iint_{\mathcal{S}} \boldsymbol{\mu}(\mathbf{s})\, \eta(\mathbf{s}-\mathbf{r})\, \boldsymbol{\mu}(\mathbf{r})^{\top}\, d\mathbf{r} d\mathbf{s} \tag{28}$$

By substituting the parameter vector $\boldsymbol{\theta}$, the state vector $\mathbf{x_t}$, the vector of kernel basis functions $\boldsymbol{\lambda}$ and the vector of field basis functions $\boldsymbol{\mu}$ in (21-28), the state-space representation of the multiresolution approximation of the IDE can be achieved. It should be noted that the complexity of the model is independent of the number of observation locations, providing more flexibility in handling high resolution data.

III. B-SPLINE WAVELET IDE MODEL

Multiresolution approximation can be implemented using different classes of scaling and wavelet functions. In order to determine the matrices $\mathbf{\Lambda}_x$ and $\mathbf{\Lambda}_\theta$ convolution and inner product expressions of the basis functions need to be computable. B-spline wavelet and scaling functions are particularly convenient in the above decomposition as their convolution and inner product expressions can be found analytically.

The mth order cardinal B-spline scaling function is defined by the recurrence relation [27]

$$\begin{aligned} N_m(s) &= (N_{m-1} * N_1)(s) \\ &= \int_0^1 N_{m-1}(s-r)\, dr \quad m > 1 \end{aligned} \tag{29}$$

Here $*$ denotes the convolution operator and $N_1(s)$ is the characteristic function of the unit interval $[0, 1)$

$$N_1(s) = \begin{cases} 1 & \text{if } 0 \le s < 1, \\ 0 & \text{elsewhere} \end{cases} \tag{30}$$

In order to compute the integrals (21) and (22) we need to be able to calculate the convolution and inner product of B-spline scaling functions. Equation (29) can be rewritten as $(m-1)$ times convolution of indicator function with itself

$$N_m(s) = \underbrace{(N_1 * N_1 * \cdots * N_1)}_{m-1 \quad \text{convolutions}}(s) \tag{31}$$

then by using the associativity property of the convolution operator, it is clear that

$$\begin{aligned} N_m(s) * N_{m'}(s) &= \underbrace{\overbrace{(N_1 * \cdots * N_1)}^{m-1 \quad \text{convolutions}}(s) * \overbrace{(N_1 * \cdots * N_1)}^{m'-1 \quad \text{convolutions}}(s)}_{m+m'-1 \quad \text{convolutions}} \\ &= N_{m+m'}(s) \end{aligned} \tag{32}$$

This result can be used to show that (Lemma 2, Appendix)

$$\begin{aligned} \langle N_m(s-l_1), N_{m'}(s-l_2)\rangle &= N_{m+m'}(m+l_1-l_2) \\ &= N_{m+m'}(m'+l_2-l_1) \end{aligned} \tag{33}$$

where $\langle .,. \rangle$ denotes the inner product.

The mth order compactly supported B-spline wavelet function is defined as [27]

$$\varphi_m(s) = \sum_{n=0}^{3m-2} q_n N_m(2s-n) \tag{34}$$

with the coefficients given by:

$$q_n = \frac{(-1)^n}{2^{m-1}} \sum_{l=0}^{m} \binom{m}{l} N_{2m}(n-l+1) \qquad 0 \le n \le 3m-2 \tag{35}$$

By exploiting properties (32) and (33) the integrals (21) and (22) can be computed. In order to calculate elements of $\mathbf{\Lambda}_x$ and $\mathbf{\Lambda}_\theta$, scaling and wavelet basis functions should be expanded in terms of N_m at the appropriate scale. It is important to note that B-spline scaling and wavelet functions possess the following orthogonality properties [28]:

$$\langle \psi_{m;j_1,l_1}(s), \psi_{m;j_2,l_2}(s)\rangle = 0 \quad \text{for} \quad j_1 \neq j_2 \tag{36}$$

$$\langle \phi_{m;j_1,l_1}(s), \psi_{m;j_2,l_2}(s)\rangle = 0 \quad \text{for} \quad j_1 \leq j_2 \tag{37}$$

Where the subscript m is introduced to $\phi_{j,l}(s)$ and $\psi_{j,l}(s)$ to indicate the mth order B-spline scaling and wavelet functions. In the two dimensional case, the convolution integral can be factorised as follows,

e.g.

$$\phi_{m;j,\mathbf{l}}(s_1,s_2) * \psi^1_{m;j,\mathbf{l}'}(s_1,s_2) = \left[\phi_{m;j,l_1}(s_1) * \phi_{m;j,l_1'}(s_1)\right] \times \left[\phi_{m;j,l_2}(s_2) * \psi_{m;j,l_2'}(s_2)\right] \tag{38}$$

The preceding factorisation can be also applied to the 2D-inner product integral. Non-zero quantities can be then expressed by

$$\left\langle \phi_{m;j,\mathbf{l}}(s_1,s_2), \phi_{m;j,\mathbf{l}'}(s_1,s_2)\right\rangle = \left\langle \phi_{m;j,l_1}(s_1), \phi_{m;j,l_1'}(s_1)\right\rangle \times \left\langle \phi_{m;j,l_2}(s_2), \phi_{m;j,l_2'}(s_2)\right\rangle \tag{39}$$

likewise we can compute $\left\langle \psi^1_{m;j,\mathbf{l}}(s_1,s_2), \psi^1_{m;j,\mathbf{l}'}(s_1,s_2)\right\rangle$, $\left\langle \psi^2_{m;j,\mathbf{l}}(s_1,s_2), \psi^2_{m;j,\mathbf{l}'}(s_1,s_2)\right\rangle$ and $\left\langle \psi^3_{m;j,\mathbf{l}}(s_1,s_2), \psi^3_{m;j,\mathbf{l}'}(s_1,s_2)\right\rangle$.

Using these expansions it becomes straightforward to calculate the integrals (21) and (22). In this work, 4th order cardinal B-spline scaling and wavelet functions are used therefore, to compute the integrals (21) and (22) the 8th and 12th order B-spline values at integer points are required and can be found in [36].

IV. Estimation of the Multiresolution Spatio-temporal IDE Model

A state-space representation of the IDE model allows the use of the Expectation Maximisation (EM) algorithm to infer both the kernel and the spatial field from the noise corrupted spatio-temporal data. In this context, the EM algorithm is a two part iterative algorithm that yields the maximum likelihood estimate of the spatial mixing kernel parameters and the posterior distribution of the field weights. In the M-step a lower bound $\mathcal{Q}\left(\boldsymbol{\theta}, \boldsymbol{\theta}'\right)$ of the complete-data log-likelihood $p\left(\mathbf{X}, \mathbf{Y} \mid \boldsymbol{\theta}\right)$, is maximised under the assumption of the system parameters being $\boldsymbol{\theta}'$ [37]. This can be achieved for the state-space representation of the IDE model via Theorem 2 of [17] under very mild conditions.

In the E-step, given an estimate of $\boldsymbol{\theta}$, the Rauch Tung Striebel (RTS) smoother can be employed to determine the posterior state sequence. It can be shown that [37], [38] every EM iteration increases the lower bound on the likelihood function and hence improving the state and parameter estimates.

The stopping criterion is usually associated with either the change in the parameter estimates or the log-likelihood variation [39]. The stopping rule adopted herein is

$$\left(\parallel \mathbf{A} \parallel_F^{(i)} - \parallel \mathbf{A} \parallel_F^{(i-1)} \right) < \varepsilon \tag{40}$$

where ε is a threshold value and $\parallel \mathbf{A} \parallel_F^{(i)}$ and $\parallel \mathbf{A} \parallel_F^{(i-1)}$ are the Frobenius norm of the successive estimates of $\mathbf{A}$ matrices. It has been shown [17] that the state dimension is a critical factor in the complexity of the estimation procedure, however using the above decomposition, this is decoupled from the number of observation locations.

V. SIMULATION

In this section, we compare the performance of the different algorithms mentioned above on a synthetic example .....This section demonstrates the identification of the proposed IDE model. Primarily, the MRA of a synthetic model is demonstrated, which is able to represent the system at different spatial scales. In addition, the complexity of the field decomposition is shown to be independent to the number and spatial regularity of the observation locations. This allows the construction of parsimonious models using spatially dense observations, and demonstrates the utility of the approach to situations where the observations are not arranged on a regular lattice, i.e. sensor networks [40].

A. One dimensional example

Consider the following homogeneous multiscale spatial mixing kernel

$$\begin{aligned} k\left(s-r\right) = & -\phi_{4;0,-3}\left(s-r\right) + \phi_{4;0,-2}\left(s-r\right) \\ & - 0.5\psi_{4;1,-6}\left(s-r\right) + 0.5\psi_{4;1,-3}\left(s-r\right) \end{aligned} \tag{41}$$

The kernel and its decomposition are plotted in Fig.1. The dynamics associated with the IDE defined by (41) will be oscillatory and dissipative to initial conditions. The field is decomposed using scaling functions $\phi_{4;0,k}$ and the associated wavelet functions $\psi_{4;j,k}$ with $j = 0, 1$. Observations are then generated from (21-28) for spatial domain defined as $\mathcal{S} = [0, 10]$ in response to an initial condition where $\mathbf{\Sigma}_v = 0.2 \times \mathbf{I}_{n_y}$ and $\eta(s-r) = 2.5 \times \phi_{4,3,-2}(s-r)$. The response was sampled at 50 irregularly spaced random locations, at $t = 1, \cdots 400$. In order to estimate the spatial mixing kernel and the spatio-temporal field, the kernel and the field are decomposed using

$$k\left(s-r\right) = \sum_{l\in\mathbb{Z}} \alpha_{0,l}\phi_{4;0,l}\left(s-r\right) + \sum_{j=0}^{1}\sum_{l\in\mathbb{Z}} \beta_{j,l}\psi_{4;j,l}\left(s-r\right) \tag{42}$$

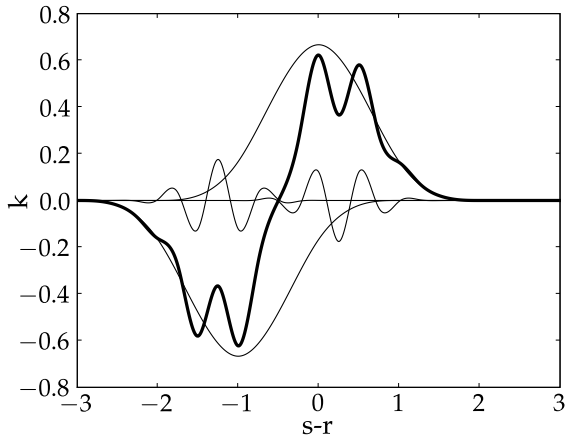


Fig. 1. Multiresolution decomposition of a complex spatial mixing kernel (dark line) and the corresponding scaling and wavelet basis functions (light lines).

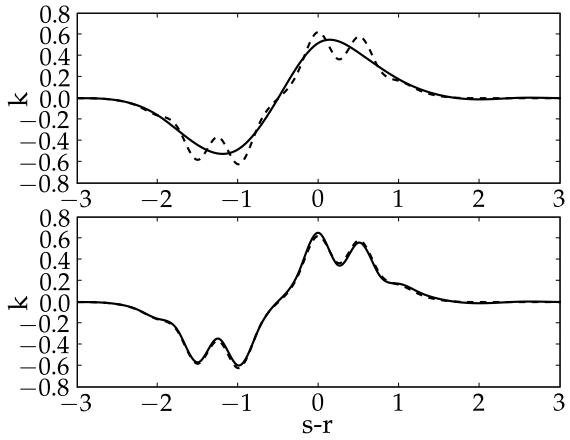


Fig. 2. Spatial mixing kernel; estimated and true kernels are shown by solid and dashed line respectively; top panel is estimation with maximmum scale $j = 0$ and bottom panel is estimation with maximum scale $j = 1$.

$$z_t(s) = \sum_{l\in\mathbb{Z}} x_{t,0,l}\phi_{4;0,l}(s) + \sum_{j=0}^{1}\sum_{l\in\mathbb{Z}} \check{x}_{t,j,l}\psi_{4;j,l}(s) \tag{43}$$

Note that the number of translation operations depends on the spatial range of the data. It follows that the total number of terms in (42) and (43) are 8 and 47 respectively. The estimated kernel in Fig.2 is obtained for the cubic B-spline, with initial scale 0 and maximum scale 1. The EM algorithm is allowed to run for 30 iterations to avoid any issues regarding early stopping, though typically the change in $\| \mathbf{A} \|_F$ falls below 10^{-4} after less than 15 iterations. A total number of 500 estimation experiments are

performed, where v_t and w_t are re-generated each trial.

The performance measure employed here is the Mean Integrated Squared Error (MISE) defined as

$$\mathbf{E}\left[\int_{\mathcal{S}} \left[e\left(s\right)\right]^2 ds\right] \tag{44}$$

where $e\left(s\right)$ is the error between the unknown function, $f\left(s\right)$ and its estimate, $\hat{f}\left(s\right)$. In this example the original kernel and the field are known and hence the integral in (44) can be computed analytically. For spatial mixing kernel we have

$$\begin{aligned}&\int_{\mathcal{S}} \left[e\left(s\right)\right]^2 ds \\ &\quad = \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)^\top \left[\int_{\mathcal{S}} \boldsymbol{\lambda}\left(s\right) \boldsymbol{\lambda}\left(s\right)^\top ds\right] \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)\end{aligned} \tag{45}$$

Here, we are particularly interested in assessing the performance improvement as the scale of the details in MRA is increased. Performance based on (45) at different spatial scales and over 500 trials have been evaluated. The MISE of the model at $j = 0$ is 0.024 and reduces to 0.005 when $\psi_{4;1,k}$ terms are

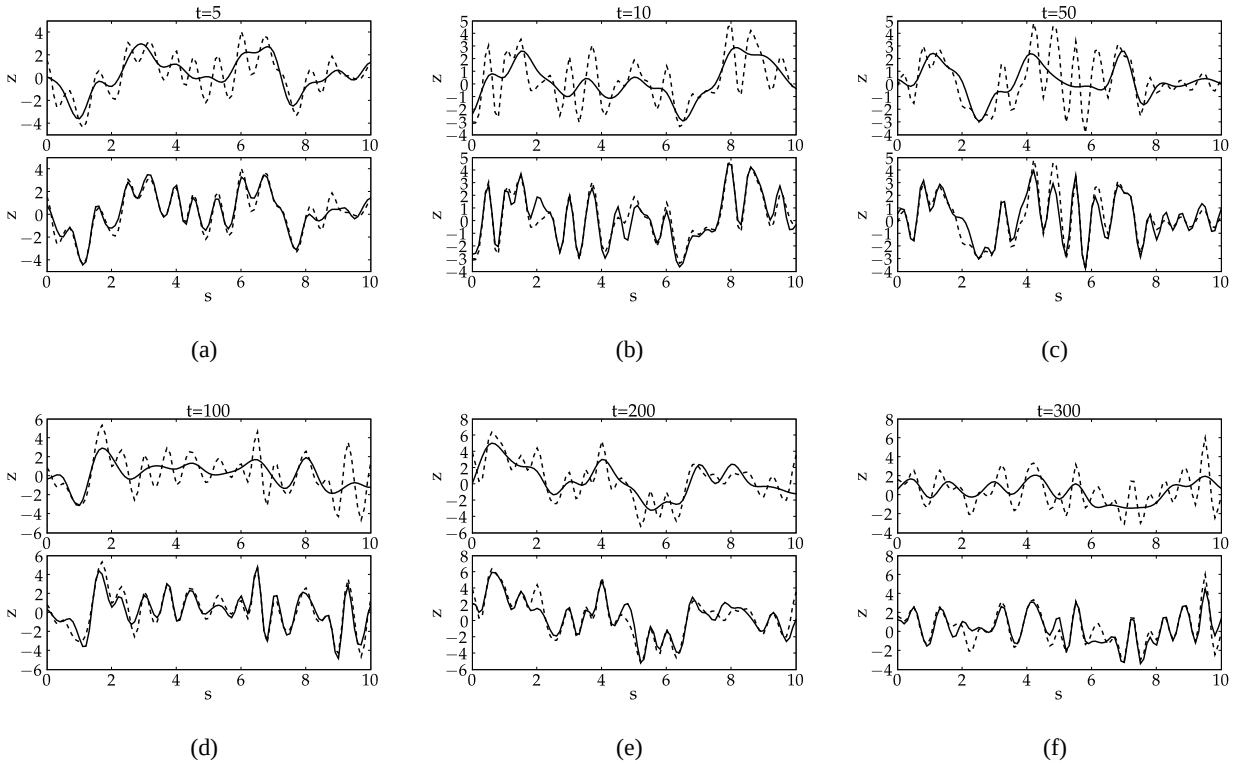


Fig. 3. Spatial field at a number of time instants; estimated and true fields are shown by solid and dashed line respectively; the top panel of each plot is the estimate with maximmum scale $j = 0$ and the bottom panel is the estimate with maximum scale $j = 1$.

also included in the model, in fact estimation performance is improved by a factor of 4.8. The coarsest approximation of the kernel along with its approximation at $j = 1$ are illustrated in Fig.2. It can be seen that the model is able to estimate both slowly and rapidly varying segments of the kernel with very high accuracy. Field estimates at a selection of time instants are shown in Fig.3. It is clearly observed that the coarse and fine features of the original field have been captured accurately over regions where observations are available. Observation locations are generated randomly, as shown in Fig. 4. Assuming the observed field can be fully described by the decomposed IDE model, the variance of the field can be computed to express the uncertainty of the estimation, at each point in space and at time t we have

$$\text{Var}(\hat{z}_t(s_i)) = \boldsymbol{\mu}^\top(s_i)\mathbf{P}_{t|T}\boldsymbol{\mu}(s_i) \tag{46}$$

where $\mathbf{P}_{t|T}$ is the covariance matrix associated with the state estimates obtained from the RTS smoother. Variance based on (46) at each spatial location and at $t = 50$ is shown in Fig.4. The uncertainty of the estimation is higher at almost every spatial location at $j = 1$ compared to $j = 0$. This is to be expected as there are more wavelet basis functions at higher resolution and hence more associated weights need to be estimated. Peaks in the variance of the field coincide with regions where no observations have been made. Increasing the number of observation locations, or using equally spaced observation locations would reduce the variance of the field estimation.

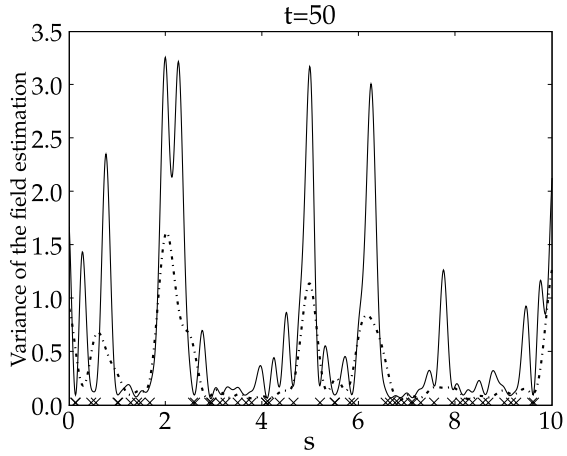


Fig. 4. Variance of the field at each spatial location for $j = 0$ (dotted line) and $j = 1$ (solid line). The observation locations are described with a cross.

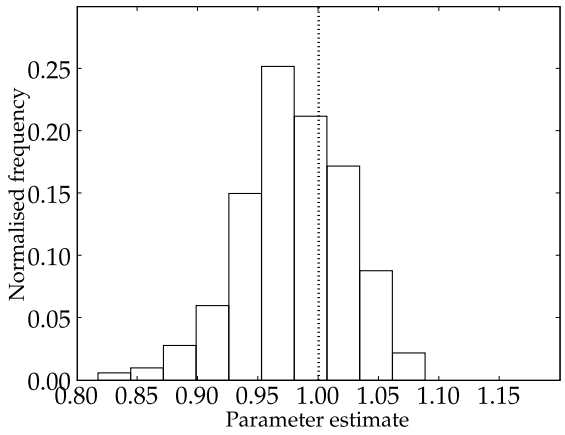


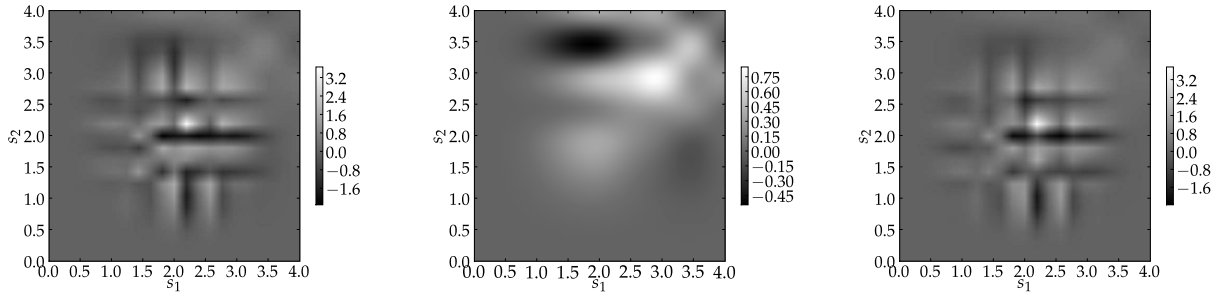
Fig. 5. Distribution of the parameter estimate for 2-D Example over 500 runs of the EM algorithm. The dotted line indicates the true parameter.

B. Two dimensional example

In order to demonstrate the ability of the developed model in higher dimensions, a two dimensional IDE model with a single B-spline kernel $\phi_{4;0,\mathbf{l}}(s_1 - r_1, s_2 - r_2)$ and $\mathbf{l} = [-2, -2]^\top$ is used with parameter $\theta = 1$. In this example spatial region $\mathcal{S} = [0, 4]^2$ is simulated over $t = \{1 \cdots 200\}$ using 7 scaling and wavelet functions at $j = 0$ together with 24 wavelet functions at $j = 1$. Univariate cubic B-spline scaling and wavelet functions are used to generate 2-D basis functions using (6-9). A set of 250 random observation locations is used; $\mathbf{\Sigma}_v$ is taken the same as that in example 1 and $\eta(s_1 - r_1, s_2 - r_2) = 2.5 \times \phi_{4,3,\mathbf{l}}(s_1 - r_1, s_2 - r_2)$ where $\mathbf{l} = [-2, -2]^\top$. The histogram of the parameter estimates in Fig.5 is generated using 500 runs of the EM algorithm. For comparison, snapshots for the original and the corresponding coarse and fine field estimates at time instants 5, 50 and 150 are illustrated in Fig.6-8. In addition, the MISE of the dynamic field at some time instants are presented in Table I. This shows that the multiresolution IDE model can estimate the spatio-temporal dynamics of the original system accurately. The accuracy of the estimation remains consistent over time. Performance is improved significantly at $j = 1$ compared to the case where $j = 0$.

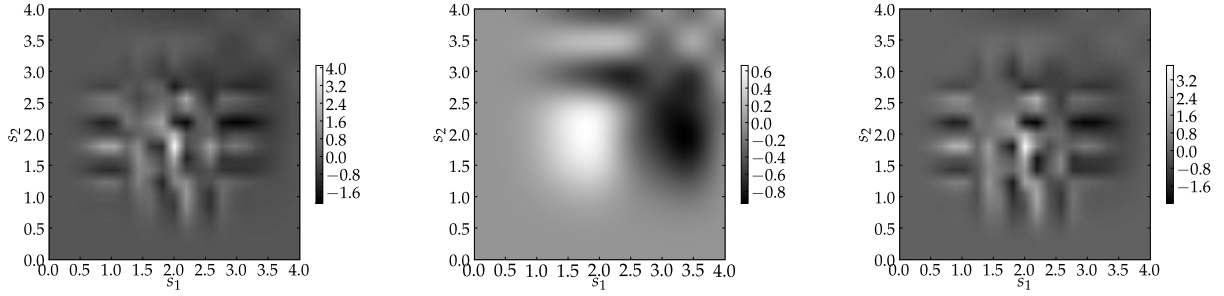
VI. CONCLUSION

A multi-resolution approach to modelling spatio-temporal systems has been presented. This model is able to represent continuous-space, discrete-time dynamics at a number of spatial scales simultaneously. This ability greatly extends the class of system which the Integro-Difference Equation can represent.



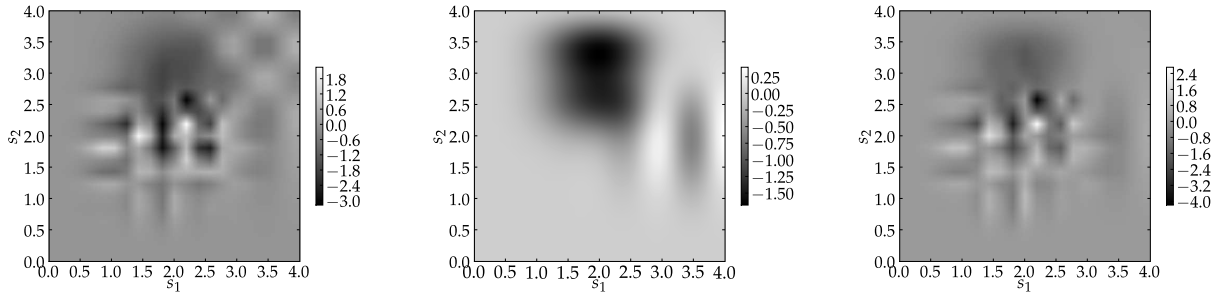
(a) (b) (c)

Fig. 6. Spatial field at t=5; a: True field; b: Estimated field with maximmum scale $j = 0$; c: Estimated field with maximmum scale $j = 1$.



(a) (b) (c)

Fig. 7. Spatial field at t=50; a: True field; b: Estimated field with maximmum scale $j = 0$; c: Estimated field with maximmum scale $j = 1$.



(a) (b) (c)

Fig. 8. Spatial field at t=150; a: True field; b: Estimated field with maximmum scale $j = 0$; c: Estimated field with maximmum scale $j = 1$.

TABLE I

MISE AT PARTICULAR SCALE BETWEEN MEASURED AND ESTIMATED 2D-FIELD AT A PARTICULAR TIME

Time	MISE	
	$j=0$	$j=1$
5	5.424	0.609
20	5.422	0.596
50	5.519	0.626
100	5.488	0.611
150	5.494	0.587

By decomposing both the spatial field and the spatial mixing kernel using a wavelet decomposition, it becomes possible to represent small scale details in the field at the same time as large scale details. This is an essential component of any practical spatio-temporal model, without which the model must be artificially separated into global and local modes a-priori.

The complexity of the model is not affected by the spatial resolution of the observation process, rather it reflects the underlying complexity of the system under study. Therefore, increasing the resolution at which the system is observed does not necessarily increase the complexity of the system identification problem. However the developed algorithm is sensitive to the state dimension, equivalent to the detail represented in the field, and hence intelligent approaches to reduce the number of basis functions is noted as future work.

By considering different levels of decomposition, the proposed approach could be used as a method of determining the appropriate scales of decomposition, within a model-selection framework. Combined with an approach to sparsely modelling spatial heterogeneity, such as boundary conditions, this would allow the application of this work to a real-world data set.

APPENDIX

Lemma 1: Consider the zero-mean process $e_t(s)$ with covariance as defined by (2) then

$$\mathbf{w}_t = \mathbf{\Lambda}_x^{-1} \int_{\mathcal{S}} \boldsymbol{\mu}(\mathbf{s})\, e_t(\mathbf{s})\, d\mathbf{s} \tag{47}$$

is a vector valued, zero mean normally distributed white noise process with covariance

$$\mathbf{\Sigma}_w = \mathbf{\Lambda}_x^{-1} \mathbf{\Pi} \mathbf{\Lambda}_x^{-\top} \tag{48}$$

and

$$\mathbf{\Pi} = \iint_{\mathcal{S}} \boldsymbol{\mu}(\mathbf{s})\, \eta(\mathbf{s}-\mathbf{r})\, \boldsymbol{\mu}(\mathbf{r})^{\top} d\mathbf{r} d\mathbf{s} \tag{49}$$

Proof: Equation (47) is a linear function of $e_t(\mathbf{s})$ and hence $\mathbf{w}_t$ is also normally distributed. The expected value of $\mathbf{w}_t$ is given by

$$\begin{aligned} \mathbf{E}[\mathbf{w}_t] &= \mathbf{\Lambda}_x^{-1} \int_{\mathcal{S}} \boldsymbol{\mu}(\mathbf{s})\, \mathbf{E}[e_t(\mathbf{s})]\, d\mathbf{s} \\ &= 0 \end{aligned} \tag{50}$$

and

$$\begin{aligned} \mathbf{\Sigma}_w &= \mathbf{E}[\mathbf{w}_t \mathbf{w}_t^{\top}] \\ &= \mathbf{\Lambda}_x^{-1} \mathbf{E}[\int_{\mathcal{S}} \boldsymbol{\mu}(\mathbf{s})\, e_t(\mathbf{s})\, d\mathbf{s} \int_{\mathcal{S}} \boldsymbol{\mu}(\mathbf{r})^{\top} e_t(\mathbf{r})\, d\mathbf{r}] \mathbf{\Lambda}_x^{-\top} \\ &= \mathbf{\Lambda}_x^{-1} \iint_{\mathcal{S}} \boldsymbol{\mu}(\mathbf{s})\, \mathbf{E}[e_t(\mathbf{s})\, e_t(\mathbf{r})] \boldsymbol{\mu}(\mathbf{r})^{\top} d\mathbf{r} d\mathbf{s} \mathbf{\Lambda}_x^{-\top} \\ &= \mathbf{\Lambda}_x^{-1} \iint_{\mathcal{S}} \boldsymbol{\mu}(\mathbf{s})\, \eta(\mathbf{s}-\mathbf{r})\, \boldsymbol{\mu}(\mathbf{r})^{\top} d\mathbf{r} d\mathbf{s} \mathbf{\Lambda}_x^{-\top} \\ &= \mathbf{\Lambda}_x^{-1} \mathbf{\Pi} \mathbf{\Lambda}_x^{-\top} \end{aligned} \tag{51}$$

■

Lemma 2: Let $N_m(s-l_1)$ and $N_{m'}(s-l_2)$ be shifted B-spline functions of order m and m' respectively. Then the inner product of $N_m(s-l_1)$ and $N_{m'}(s-l_2)$ can be calculated by

$$\begin{aligned} \langle N_m(s-l_1), N_{m'}(s-l_2) \rangle &= N_{m+m'}(m+l_1-l_2) \\ &= N_{m+m'}(m'+l_2-l_1) \end{aligned} \tag{52}$$

Proof: The support of $N_m(s)$ is $[0, m]$ and is symmetric with respect to $s = \frac{m}{2}$, i.e.

$$N_m\left(\frac{m}{2}+s\right) = N_m\left(\frac{m}{2}-s\right) \tag{53}$$

A direct consequence of (53) is

$$N_m(s) = N_m(m-s) \tag{54}$$

Therefore

$$\begin{aligned}
&\int_{-\infty}^{+\infty} N_m\left(s-l_1\right) N_{m'}\left(s-l_2\right) ds \\
&\quad = \int_{-\infty}^{+\infty} N_m\left(m-s+l_1\right) N_{m'}\left(s-l_2\right) ds \\
&\quad = \int_{-\infty}^{+\infty} N_m\left(m+l_1-l_2-u\right) N_{m'}\left(u\right) du \\
&\quad = \left(N_m * N'_m\right)\left(m+l_1-l_2\right) \\
&\quad = N_{m+m'}\left(m+l_1-l_2\right) \\
&\quad = N_{m+m'}\left(m'+l_2-l_1\right)
\end{aligned}$$

■

REFERENCES

[1] O. Schofield, T. Bergmann, P. Bissett, J. Grassle, D. Haidvogel, J. Kohut, M. Moline, and S. Glenn, "The long-term ecosystem observatory: an integrated coastal observatory," *Oceanic Engineering, IEEE Journal of*, vol. 27, no. 2, pp. 146–154, Apr. 2002.

[2] K. Xu, W. C. K, and F. Neil, "A kernel-based spatio-temporal dynamical model for nowcasting weather radar reflectivities," *Journal of the American Statistical Association*, vol. 100, no. 472, pp. 1133–1144, Dec. 2005.

[3] R. Romanowicz, P. Young, P. Brown, and P. Diggle, "A recursive estimation approach to the spatio-temporal analysis and modelling of air quality data," *Environmental Modelling & Software*, vol. 21, no. 6, pp. 759–769, Jun. 2006.

[4] B. Bond-Lamberty, C. Wang, and S. Gower, "Spatiotemporal measurement and modeling of stand-level boreal forest soil temperatures," *Agricultural and Forest Meteorology*, vol. 131, no. 1-2, pp. 27–40, Jul. 2005.

[5] H. Deng, H.-X. Li, and G. Chen, "Spectral approximation based intelligent modelling for distributed thermal process," *IEEE Trans. Control Syst. Technol.*, vol. 13, no. 5, pp. 686–700, May. 2005.

[6] H.-I. Kuo, C.-L. Lu, W.-C. Tseng, and H.-A. Li, "A spatiotemporal statistical model of the risk factors of human cases of H5N1 avian influenza in South-east Asian countries and China," *Public Health*, vol. 123, no. 2, pp. 188–193, Feb. 2009.

[7] H. Sun, Y. Tu, and S.-M. Yu, "A spatio-temporal Autoregressive model for multi-unit residential market analysis*," *The Journal of Real Estate Finance and Economics*, vol. 31, no. 2, pp. 155–187, Sep. 2005.

[8] S. Wolfram, "Universality and complexity in cellular automata," *Physica D*, vol. 10, no. 1-2, pp. 1 – 35, 1984.

[9] S. A. Billings and D. Coca, "Identification of coupled map lattice models of deterministic distributed parameter systems," *Int J Syst. Sci.*, vol. 33, no. 8, pp. 623–634, 2002.

[10] Hua-Liang Wei, S. A. Billings, Y. Zhao, and L. Guo, "Lattice dynamical wavelet neural networks implemented using particle swarm optimization for spatio-temporal system identification," *Neural Networks, IEEE Transactions on*, vol. 20, no. 1, pp. 181–185, Jan. 2009.

[11] P. E. Pfeifer and S. J. Deutsch, "Identification and interpretation of first order space-time ARMA models," *Technometrics*, vol. 22, no. 3, pp. 397–408, 1980.

[12] C. A. Glasbey and D. J. Allcroft, "A spatiotemporal auto-regressive moving average model for solar radiation," *Journal of the Royal Statistical Society*, vol. 57, no. 3, pp. 343–355, Jun. 2008.

[13] M. Dewar and V. Kadirkamanathan, "A canonical space-time state space model: State and parameter estimation," *Signal Processing, IEEE Transactions on*, vol. 55, no. 10, pp. 4862–4870, Oct. 2007.

[14] M. Kot, "Discrete-time travelling waves: Ecological examples," *Journal of Mathematical Biology*, vol. 30, no. 4, pp. 413–436, Feb. 1992.

[15] M. Kot, M. A. Lewis, and P. van den Driessche, "Dispersal data and the spread of invading organisms," *Ecology*, vol. 77, no. 7, pp. 2027–2042, Oct. 1996.

[16] C. K. Wikle, "A kernel-based spectral model for non-Gaussian spatio-temporal processes," *Statist. Model.*, vol. 2, pp. 299–314, 2004.

[17] M. Dewar, K. Scerri, and V. Kadirkamanathan, "Data-driven spatio-temporal modeling using the integro-difference equation," *Signal Processing, IEEE Transactions on*, vol. 57, no. 1, pp. 83–91, Jan. 2009.

[18] K. Scerri, M. Dewar, and V. Kadirkamanathan, "Estimation and Model Selection for an IDE-Based Spatio-Temporal Model," *Signal Processing, IEEE Transactions on*, vol. 57, no. 2, pp. 482–492, Feb. 2009.

[19] J. A. Wiens, "Spatial scaling in ecology," *Functional Ecology*, vol. 3, no. 4, pp. 385–397, 1989.

[20] D. B. Chelton, "Physical oceanography: A brief overview for statisticians," *Statistical Science*, vol. 9, no. 2, pp. 150–166, May 1994.

[21] M. Schmidt, O. Fabert, and C. Shum, "On the estimation of a multi-resolution representation of the gravity field based on spherical harmonics and wavelets," *Journal of Geodynamics*, vol. 39, no. 5, pp. 512–526, Jul. 2005.

[22] B. Ribba, T. Colin, and S. Schnell, "A multiscale mathematical model of cancer, and its use in analyzing irradiation therapies," *Theoretical Biology and Medical Modelling*, vol. 3, no. 1, p. 7, Feb. 2006.

[23] A. S. Willsky, "Multiresolution markov models for signal and image processing," *Proc. IEEE*, vol. 90, no. 8, pp. 1396?–1458, Aug. 2002.

[24] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 2, no. 7, pp. 674–693, Jul. 1989.

[25] ——, "Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbf{R})$," *Trans. Amer. Math. Soc.*, vol. 315, no. 1, pp. 69–87, 1989.

[26] D. F. Walnut, *An introduction to wavelet analysis*. Boston, MA: Birkhäuser, 2002.

[27] C. K. Chui and J. Wang, "On compactly supported spline wavelets and a duality principle," *Trans. Amer. Math. Soc.*, vol. 330, no. 2, pp. 903–915, 1992.

[28] M. Unser, A. Aldroubi, and M. Eden, "A family of polynomial spline wavelet transforms," *Signal Process.*, vol. 30, no. 2, pp. 141–162, Jan. 1993.

[29] ——, "On the asymptotic convergence of B-spline wavelets to Gabor functions," *Information Theory, IEEE Transactions on*, vol. 38, no. 2, pp. 864–872, Mar. 1992.

[30] C. K. Chui, *An Introduction to Wavelets*. Boston, MA: Academic Press, 1992.

[31] G. P. Liu, S. A. Billings, and V. Kadirkamanathan, "Nonlinear system identification using wavelet networks," *International Journal of Systems Science*, vol. 31, no. 12, pp. 1531–1541, Dec. 2000.

[32] W. Jun and P. Hong, "Constructing fuzzy wavelet network modeling," *International Journal of Information Technology*, vol. 11, no. 6, pp. 68–74, 2005.

[33] S. A. Billings, L. Z. Guo, and H. L. Wei, "Identification of coupled map lattice models for spatio-temporal patterns using wavelets," *Intern. J. Syst. Sci.*, vol. 37, no. 14, pp. 1021–1038, Nov. 2006.

[34] Y. Meyer, *Wavelets and Operators*. New York: Cambridge Univ. Press, 1992.

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

[36] J. C. Goswami and A. Chan, *Fundamentals of Wavelets: Theory, Algorithms, and Applications*. New York: Wiley, 1999.

[37] S. Gibson and B. Ninness, "Robust maximum-likelihood estimation of multivariable dynamic systems," *Automatica*, vol. 41, no. 10, pp. 1667–1682, Oct. 2005.

[38] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J, Roy. Statist., Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.

[39] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York: Wiley, 1997.

[40] S. K. Sahu and P. Challenor, "A space-time model for joint modeling of ocean temperature and salinity levels as measured by Argo floats," *Environmetrics*, vol. 19, no. 5, pp. 509–528, Aug. 2008.