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

In no particular order: P. Aram, M. Dewar, D.R. Freestone, K.Scerri, D.B. Grayden and V. Kadirkamanathan

***Abstract*—**

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

I. Introduction

To Do

II. Integro-Difference Equation Neural Field Model

The stochastic integro-difference equation (IDE) form of the model based on the Wilson-Cowan [1] or Amari neural field [2] formulation is given by (cite NeuroImage paper)

$$v_{t+T_s}(\mathbf{r}) = \xi v_t(\mathbf{r}) + T_s \int_\Omega w(\mathbf{r},\mathbf{r}')\, f(v_t(\mathbf{r}'))\, d\mathbf{r}' + e_t(\mathbf{r}), \tag{1}$$

where the local post-synaptic membrane voltage at time t of a population of neurons at position $\mathbf{r}$ is denoted $v_t(\mathbf{r})$, parameter $\xi = \left(1 - \frac{T_s}{\tau}\right)$ where τ is the synaptic time constant and T_s is the sampling time. The spatial connectivity between neural populations at a distance $\mid \mathbf{r} - \mathbf{r}' \mid$ is described by the connectivity kernel $w(\mathbf{r}-\mathbf{r}')$, for an isotropic and homogeneous neural field. The connectivity kernel is typically a "Mexican hat" function, which describes local excitation and lateral inhibition [2]. The firing rate of the presynaptic neurons is related to the post-synaptic membrane potential by the sigmoidal activation function $f(.)$. The term $e_t(\mathbf{r})$ is an $i.i.d.$ disturbance such that $e_t(\mathbf{s}) \sim \mathcal{GP}(\mathbf{0}, \eta(\mathbf{r}-\mathbf{r}'))$, where $\mathcal{GP}(\mathbf{0}, \eta(\mathbf{r}-\mathbf{r}'))$ denotes a zero mean spatial Gaussian process with covariance function $\eta(\mathbf{r}-\mathbf{r}')$ [3]. For simplicity, we denote hereafter the current time step by t and the future time step by $t+1$. Here we assume that the sigmoid activation function, $f(\cdot)$, can be approximated by the piecewise function

$$\hat{f}(v_t(\mathbf{r}')) = \begin{cases} 0, & v_t(\mathbf{r}') \le v_1 \\ \varsigma v_t(\mathbf{r}'), & v_1 < v_t(\mathbf{r}') < v_2 \\ 1, & v_t(\mathbf{r}') \ge v_2 \end{cases} \tag{2}$$

where the parameter ς governs the slope of the sigmoid. Furthermore, we assume the field is in the linear region of the piecewise function giving

$$v_{t+1}(\mathbf{r}) = \xi v_t(\mathbf{r}) + T_s \int_\Omega w(\mathbf{r}-\mathbf{r}')\, v_t(\mathbf{r}')\, d\mathbf{r}' + e_t(\mathbf{r}), \tag{3}$$

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)

Note that the parameter ς is absorbed into the kernel for simplicity. Oservation equation is given by

$$\mathbf{y}_t = \int_\Omega m\left(\mathbf{r}_{n_y} - \mathbf{r}'\right) v_t(\mathbf{r}')\, d\mathbf{r}' + \boldsymbol{\epsilon}_t, \tag{4}$$

where electrophysiological recordings $\mathbf{y}_t = [y_t(\mathbf{r}_1) y_t(\mathbf{r}_2) \cdots y_t(\mathbf{r}_{n_y})]^\top$, compiled at n_y spatial locations and at time t via the observation kernel $m\left(\mathbf{r}_{n_y} - \mathbf{r}'\right)$, is corrupted by an i.i.d normally distributed zero-mean white noise $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$ with $\boldsymbol{\Sigma}_\epsilon = \sigma_\epsilon^2 \mathbf{I}_{n_y}$. The superscript $\top$ denotes the transpose operator.

III. MRA of the IDE in State-Space

Assuming $v_t(.)$ and $w(.)$ are square-integrable functions, MRA of the one-dimensional IDE can be obtained by expanding both the field and the kernel in the equation (3) in terms of the translations and dilations of a scaling function $\phi(r)$ and a mother wavelet $\psi(r)$. For the one-dimensional kernel this is given by,

$$w(r-r') = \sum_{l\in\mathbb{Z}} \alpha_{j_0,l}\phi_{j_0,l}(r-r') + \sum_{j\ge j_0}^{\infty}\sum_{l\in\mathbb{Z}} \beta_{j,l}\psi_{j,l}(r-r') \tag{5}$$

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 $\phi_{j,l}(r) = 2^{\frac{j}{2}}\phi\left(2^j r - l\right)$ and $\psi_{j,l}(r) = 2^{\frac{j}{2}}\psi\left(2^j r - l\right)$. Integers j and l are called scale and translation parameters accordingly. The one-dimensional field can similarly be decomposed as

$$v_t(r) = \sum_{l\in\mathbb{Z}} x_{t,j_0,l}\phi_{j_0,l}(r) + \sum_{j\ge j_0}^{\infty}\sum_{l\in\mathbb{Z}} \breve{x}_{t,j,l}\psi_{j,l}(r) \tag{6}$$

where $x_{t,j_0,l}$ and $\breve{x}_{t,j,l}$ 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 [4], [5]. We write the 2-D scaling function as

$$\phi_{j,\mathbf{l}}(r_1, r_2) = \phi_{j,l_1}(r_1)\,\phi_{j,l_2}(r_2) \tag{7}$$

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

$$\psi^{(1)}_{j,\mathbf{l}}(r_1, r_2) = \phi_{j,l_1}(r_1)\,\psi_{j,l_2}(r_2) \tag{8}$$

$$\psi^{(2)}_{j,\mathbf{l}}(r_1, r_2) = \psi_{j,l_1}(r_1)\,\phi_{j,l_2}(r_2) \tag{9}$$

$$\psi^{(3)}_{j,\mathbf{l}}(r_1, r_2) = \psi_{j,l_1}(r_1)\,\psi_{j,l_2}(r_2) \tag{10}$$

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

$$\phi_{j,\mathbf{l}}(r_1, r_2, \cdots r_d) = \phi_{j,l_1}(r_1)\,\phi_{j,l_2}(r_2)\cdots\phi_{j,l_d}(r_d) \quad (11)$$
$$j \in \mathbb{Z}, \quad \mathbf{l} \in \mathbb{Z}^d$$

and $2^d - 1$ wavelet functions $\psi^{(n)}(\mathbf{r}), n = 1, \cdots, 2^d - 1$ 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

$$k(\mathbf{r}-\mathbf{r}') = \sum_{\mathbf{l}\in\mathbb{Z}^d} \alpha_{j_0,\mathbf{l}}\phi_{j_0,\mathbf{l}}(\mathbf{r}-\mathbf{r}') + \sum_{j\geq j_0}^{\infty}\sum_{\mathbf{l}\in\mathbb{Z}^d}\sum_{i=1}^{2^d-1}\beta_{j,\mathbf{l}}^{(i)}\psi_{j,\mathbf{l}}^{(i)}(\mathbf{r}-\mathbf{r}') \quad (12)$$

$$z_t(\mathbf{r}) = \sum_{\mathbf{l}\in\mathbb{Z}^d} x_{t,j_0,\mathbf{l}}\phi_{j_0,\mathbf{l}}(\mathbf{r}) + \sum_{j\geq j_0}^{\infty}\sum_{\mathbf{l}\in\mathbb{Z}^d}\sum_{i=1}^{2^d-1}\check{x}_{t,j,\mathbf{l}}^{(i)}\psi_{j,\mathbf{l}}^{(i)}(\mathbf{r}) \quad (13)$$

Equations (12) and (13) are infinite series expansions. For practical implementation, (12) and (13) must be truncated at some level j, therefore we have

$$w(\mathbf{r}-\mathbf{r}') \approx \boldsymbol{\theta}^\top \boldsymbol{\lambda}(\mathbf{r}-\mathbf{r}') \quad (14)$$

$$v_t(\mathbf{r}) \approx \boldsymbol{\mu}(\mathbf{r})^\top \mathbf{x}_t \quad (15)$$

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

$$\boldsymbol{\theta}^\top = \begin{bmatrix}\boldsymbol{\alpha}_{j_0}^\top & \boldsymbol{\beta}_{j_0}^\top & \boldsymbol{\beta}_{j_0+1}^\top \cdots \boldsymbol{\beta}_j^\top\end{bmatrix} \quad (16)$$

$$\mathbf{x}_t^\top = \begin{bmatrix}\mathbf{x}_{t,j_0}^\top & \check{\mathbf{x}}_{t,j_0}^\top & \check{\mathbf{x}}_{t,j_0+1}^\top \cdots \check{\mathbf{x}}_{t,j}^\top\end{bmatrix} \quad (17)$$

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

$$\boldsymbol{\beta}_j^\top = \begin{bmatrix}\boldsymbol{\beta}_j^{(1)\top} & \boldsymbol{\beta}_j^{(2)\top} \cdots \boldsymbol{\beta}_j^{(2^d-1)\top}\end{bmatrix} \quad (18)$$

$$\check{\mathbf{x}}_{t,j}^\top = \begin{bmatrix}\check{\mathbf{x}}_{t,j}^{(1)\top} & \check{\mathbf{x}}_{t,j}^{(2)\top} \cdots \check{\mathbf{x}}_{t,j}^{(2^d-1)\top}\end{bmatrix} \quad (19)$$

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 $\boldsymbol{\lambda} : \mathbb{R}^{n_S} \to \mathbb{R}^{n_\theta}$ and the vector of the field scaling and wavelet functions $\boldsymbol{\mu} : \mathbb{R}^{n_S} \to \mathbb{R}^{n_x}$ are defined by

$$\boldsymbol{\lambda}^\top(\mathbf{r}-\mathbf{r}') = \begin{bmatrix}\boldsymbol{\phi}_{j_0}^\top(\mathbf{r}-\mathbf{r}') & \boldsymbol{\psi}_{j_0}^\top(\mathbf{r}-\mathbf{r}') & \boldsymbol{\psi}_{j_0+1}^\top(\mathbf{r}-\mathbf{r}') \cdots & \boldsymbol{\psi}_j^\top(\mathbf{s}-\mathbf{r})\end{bmatrix} \quad (20)$$

$$\boldsymbol{\mu}^\top(\mathbf{r}) = \begin{bmatrix}\boldsymbol{\phi}_{j_0}^\top(\mathbf{r}) & \boldsymbol{\psi}_{j_0}^\top(\mathbf{r}) & \boldsymbol{\psi}_{j_0+1}^\top(\mathbf{r}) \cdots \boldsymbol{\psi}_j^\top(\mathbf{r})\end{bmatrix} \quad (21)$$

Each vector of basis functions on the right hand-side of (20) and (21) are constructed in a way similar to their corresponding weight vectors in (16) and (17). 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.

To derive the state-space representation of the model we substitute decompositions (14) and (15) into equation (3) and then cross-multiply the decomposed model by $\boldsymbol{\mu}(\mathbf{r})$ and integrate over the spatial domain, Ω, giving

$$\int_\Omega \boldsymbol{\mu}(\mathbf{r})\boldsymbol{\mu}(\mathbf{r})^\top d\mathbf{r}\mathbf{x}_{t+1} = \xi\int_\Omega \boldsymbol{\mu}(\mathbf{r})\boldsymbol{\mu}(\mathbf{r})^\top d\mathbf{r}\mathbf{x}_t + T_s\int_\Omega \boldsymbol{\mu}(\mathbf{r})\int_\Omega \boldsymbol{\theta}^\top\boldsymbol{\lambda}(\mathbf{r}-\mathbf{r}')\boldsymbol{\mu}^\top(\mathbf{r}')\mathbf{x}_t\,d\mathbf{r}'d\mathbf{r} + \int_\Omega \boldsymbol{\mu}(\mathbf{r})e_t(\mathbf{r})\,d\mathbf{r}, \quad (22)$$

Now defning

$$\boldsymbol{\Lambda}_x \triangleq \int_\Omega \boldsymbol{\mu}(\mathbf{r})\boldsymbol{\mu}^\top(\mathbf{r})\,d\mathbf{r} \quad (23)$$

$$\boldsymbol{\Lambda}_\theta \triangleq \int_\Omega \boldsymbol{\mu}(\mathbf{r})\int_\Omega \boldsymbol{\theta}^\top\boldsymbol{\lambda}(\mathbf{r}-\mathbf{r}')\boldsymbol{\mu}^\top(\mathbf{r}')\mathbf{x}_t\,d\mathbf{r}'d\mathbf{r} \quad (24)$$

and substituting these into (22) and cross-multiplying by $\boldsymbol{\Lambda}_x^{-1}$ yields

$$\mathbf{x}_{t+1} = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}_t + \mathbf{w}_t \quad (25)$$

where

$$\mathbf{A}(\boldsymbol{\theta}) = (T_s\boldsymbol{\Lambda}_x^{-1}\boldsymbol{\Lambda}_\theta + \xi\mathbf{I}) \quad (26)$$

and $\mathbf{I}$ is the identity matrix, and

$$\mathbf{w}_t = \boldsymbol{\Lambda}_x^{-1}\int_\Omega \boldsymbol{\mu}(\mathbf{r})e_t(\mathbf{r})\,d\mathbf{r} \quad (27)$$

which is a vector valued, zero-mean normally distributed white noise process with covariance [6]

$$\boldsymbol{\Sigma}_w = \boldsymbol{\Lambda}_x^{-1}\iint_{\boldsymbol{\Omega}} \boldsymbol{\mu}(\mathbf{r})\eta(\mathbf{r}-\mathbf{r}')\boldsymbol{\mu}^\top(\mathbf{r}')\,d\mathbf{r}'d\mathbf{r}\boldsymbol{\Lambda}_x^{-\top}. \quad (28)$$

Similarly, the observation equation of the state-space model is found by substituting decomposition (15) into equation (4) giving

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \boldsymbol{\varepsilon}_t, \quad (29)$$

where each element of the observation matrix is

$$\mathbf{C}_{ij} \triangleq \int_\Omega m(\mathbf{r}_i - \mathbf{r}')\boldsymbol{\mu}_j(\mathbf{r}')\,d\mathbf{r}'. \quad (30)$$

This completes the state-space representation of the multiresolution approximation of the neural field model. ***Mike you can add state-space derivation using Galerkin projection here.***

IV. B-SPLINE WAVELET IDE MODEL

Multiresolution approximation can be implemented using different classes of scaling and wavelet functions. In order to determine the matrices $\boldsymbol{\Lambda}_x$ and $\boldsymbol{\Lambda}_\theta$ 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 [7]

$$\begin{aligned} N_m(s) &= (N_{m-1} * N_1)(s) \\ &= \int_0^1 N_{m-1}(s-r)\,dr \quad m>1 \end{aligned} \tag{31}$$

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

$$N_1(s) = \begin{cases} 1 & \text{if } 0 \le s < 1, \\ 0 & \text{elsewhere} \end{cases} \tag{32}$$

In order to compute the integrals (**??**) and (**??**) we need to be able to calculate the convolution and inner product of B-spline scaling functions. Equation (31) can be rewritten as $(m-1)$ times convolution of indicator function with itself

$$N_m(s) = \underbrace{(N_1 * N_1 * \cdots * N_1)}_{m-1 \quad \text{convolutions}}(s) \tag{33}$$

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

$$\begin{aligned} N_m(s) * N_{m'}(s) &= \underbrace{\overbrace{(N_1 * \cdots * N_1)}^{m-1 \quad \text{convolutions}}(s) * \overbrace{(N_1 * \cdots * N_1)}^{m'-1 \quad \text{convolutions}}(s)}_{m+m'-1 \quad \text{convolutions}} \\ &= N_{m+m'}(s) \end{aligned} \tag{34}$$

This result can be used to show that (Lemma 1, Appendix)

$$\begin{aligned} \langle N_m(s-l_1), N_{m'}(s-l_2)\rangle &= N_{m+m'}(m+l_1-l_2) \\ &= N_{m+m'}(m'+l_2-l_1) \end{aligned} \tag{35}$$

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

The mth order compactly supported B-spline wavelet function is defined as [7]

$$\varphi_m(s) = \sum_{n=0}^{3m-2} q_n N_m(2s-n) \tag{36}$$

with the coefficients given by:

$$q_n = \frac{(-1)^n}{2^{m-1}} \sum_{l=0}^{m} \binom{m}{l} N_{2m}(n-l+1) \qquad 0 \le n \le 3m-2 \tag{37}$$

By exploiting properties (34) and (35) the integrals (**??**) and (**??**) 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 [8]:

$$\langle \psi_{m;j_1,l_1}(s), \psi_{m;j_2,l_2}(s)\rangle = 0 \quad \text{for} \quad j_1 \neq j_2 \tag{38}$$

$$\langle \phi_{m;j_1,l_1}(s), \psi_{m;j_2,l_2}(s)\rangle = 0 \quad \text{for} \quad j_1 \le j_2 \tag{39}$$

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.

$$\begin{aligned} \phi_{m;j,\mathbf{l}}(s_1,s_2) * \psi^1_{m;j,\mathbf{l}'}(s_1,s_2) = &\left[\phi_{m;j,l_1}(s_1) * \phi_{m;j,l'_1}(s_1)\right] \\ &\times \left[\phi_{m;j,l_2}(s_2) * \psi_{m;j,l'_2}(s_2)\right] \end{aligned} \tag{40}$$

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

$$\begin{aligned} \langle \phi_{m;j,\mathbf{l}}(s_1,s_2), \phi_{m;j,\mathbf{l}'}(s_1,s_2)\rangle = &\left\langle \phi_{m;j,l_1}(s_1), \phi_{m;j,l'_1}(s_1)\right\rangle \\ &\times \left\langle \phi_{m;j,l_2}(s_2), \phi_{m;j,l'_2}(s_2)\right\rangle \end{aligned} \tag{41}$$

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 (**??**) and (**??**). In this work, 4th order cardinal B-spline scaling and wavelet functions are used therefore, to compute the integrals (**??**) and (**??**) the 8th and 12th order B-spline values at integer points are required and can be found in [9].

V. 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 $p(\mathbf{X}, \mathbf{Y} \mid \boldsymbol{\theta})$, is maximised under the assumption of the system parameters being $\boldsymbol{\theta}'$ [10]. In the E-step, given an estimate of $\boldsymbol{\theta}$, filtering methods can be used to determine the posterior state sequence. Here we compare three techniques namely based on the Rauch Tung Striebel (RTS) smoother [11], Ensemble Kalman smoother (EnKs) [12] and a heuristic separable RTS smoother for state estimation problem. For the first method we modify the framework proposed in [13] by computing and storing the required elements prior to the iterative estimation algorithm, the modified algorithm is computationally less demanding (see Appendix **??**). In Ensemble Kalman filter (EKf) an ensemble of model states, $\mathbf{X}_t \in \mathbb{R}^{n_x \times N} \sim \mathcal{N}(\mathbf{x}_t, \mathbf{P}_t)$ is propogated through the state equation to calculate the forecast estimates $\hat{\mathbf{x}}_t^f$ and the error covariance of the predicted model state, $\mathbf{P}_t^f$ where N is the number of ensemble members. Then the model state is updated to compute analysed estimates, $\hat{\mathbf{x}}_t^a$. This is important for model with high dimensional state space where standard Kalman filter is computationally infeasible to implement as the state covariance matrix needs to be stored and propogated in time. The EnKs updates the filtered ensembles at the current and all prior times every time a

new set of measurements becomes available. The algorithm is given in Algorithm **??** for completeness. It is essential that the observations be also treated as an ensemble of random variables, $\mathbf{Y}_t$, having a distribution with mean equal to the observed value, $\mathbf{y}_t$ and covariance equal to $\mathbf{\Sigma}_v$ [14].

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 [11], [15] 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 [10]. The stopping rule adopted herein is

$$\left(\| \mathbf{A} \|_F^{(i)} - \| \mathbf{A} \|_F^{(i-1)} \right) < \varepsilon \tag{42}$$

where ε is a threshold value and $\| \mathbf{A} \|_F^{(i)}$ and $\| \mathbf{A} \|_F^{(i-1)}$ are the Frobenius norm of the successive estimates of $\mathbf{A}$ matrices. It has been shown [13] 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.

VI. 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 [16].

A. One dimensional example

Consider the following homogeneous multiscale spatial mixing kernel

$$\begin{aligned} k(s-r) = & -\phi_{4;0,-3}(s-r) + \phi_{4;0,-2}(s-r) \\ & - 0.5\psi_{4;1,-6}(s-r) + 0.5\psi_{4;1,-3}(s-r) \end{aligned} \tag{43}$$

The kernel and its decomposition are plotted in Fig.**??**. The dynamics associated with the IDE defined by (43) 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 (**??**-**??**) 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

$$k(s-r) = \sum_{l \in \mathbb{Z}} \alpha_{0,l} \phi_{4;0,l}(s-r) + \sum_{j=0}^{1} \sum_{l \in \mathbb{Z}} \beta_{j,l} \psi_{4;j,l}(s-r) \tag{44}$$

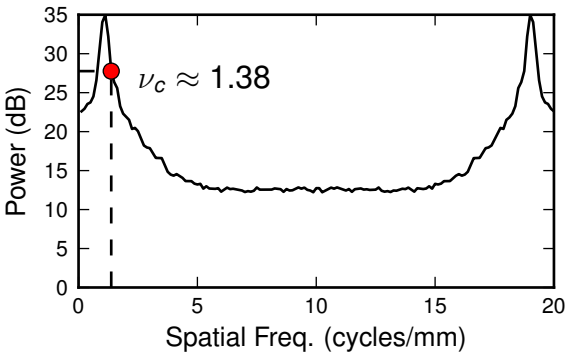

Fig. 1. The average (over time) power in dB of the spatial frequency of the observations. The dashed line shows the cutoff frequency (-3 dB point).

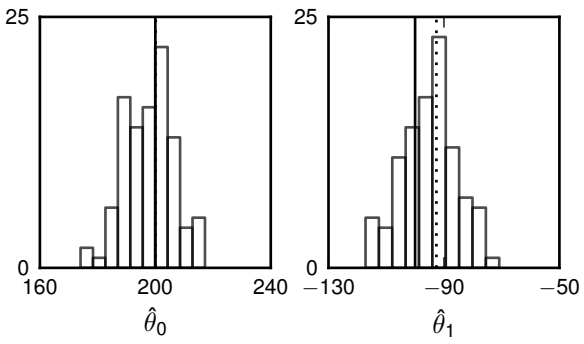

Fig. 2. Histograms of the parameter estimates over 100 realizations. In each of the subplots the solid lines show the actual parameter values and the dotted lines show the means of the estimated parameters. Left panel: The histogram of the central excitatory connectivity kernel basis function parameter estimates, θ_0. Right panel: The histogram of the surround inhibition connectivity kernel basis function parameter estimates, θ_1.

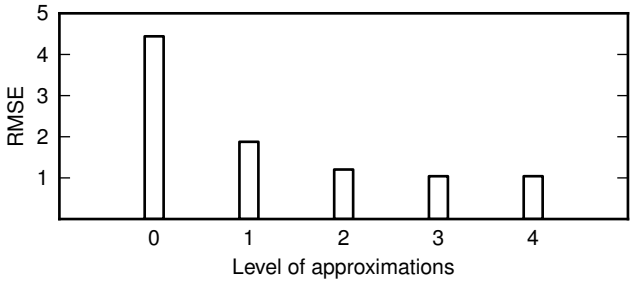

Fig. 3. RMSE of the estimated field at different spatial resolution.

$$z_t(s) = \sum_{l \in \mathbb{Z}} x_{t,0,l} \phi_{4;0,l}(s) + \sum_{j=0}^{1} \sum_{l \in \mathbb{Z}} \breve{x}_{t,j,l} \psi_{4;j,l}(s) \tag{45}$$

Note that the number of translation operations depends on the spatial range of the data. It follows that the total number of terms in (44) and (45) are 8 and 47 respectively. The estimated kernel in Fig.**??** 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

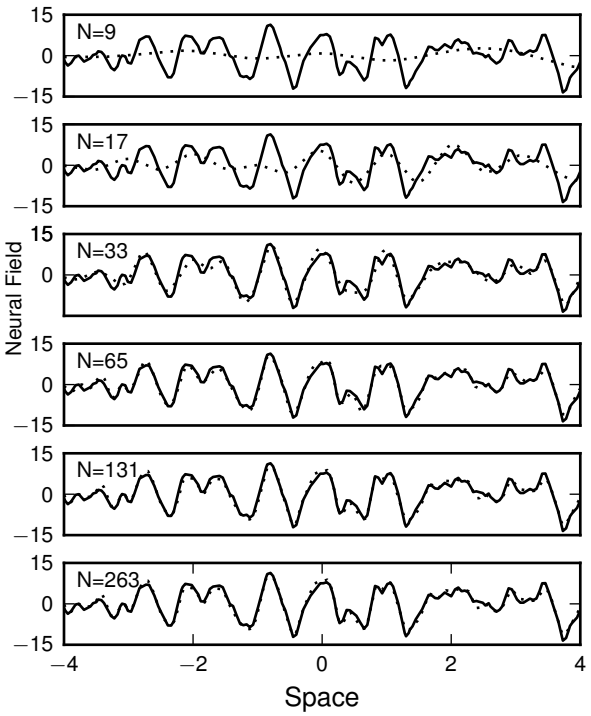

Fig. 4. Spatial field at one time instant for different spatial resolutions; estimated and true fields are shown by dashed and solid lines respectively.

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

$$\mathbf{E}\left[\int_{\mathcal{S}} \left[e\left(s\right)\right]^2 ds\right] \tag{46}$$

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 (46) can be computed analytically. For spatial mixing kernel we have

$$\begin{aligned}&\int_{\mathcal{S}} \left[e\left(s\right)\right]^2 ds \\ &\quad = \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)^\top \left[\textstyle\int_{\mathcal{S}} \boldsymbol{\lambda}\left(s\right) \boldsymbol{\lambda}\left(s\right)^\top ds\right] \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)\end{aligned} \tag{47}$$

Here, we are particularly interested in assessing the performance improvement as the scale of the details in MRA is increased. Performance based on (47) 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 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.**??**. 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.**??**. 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. **??**. 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

$$\text{Var}(\hat{z}_t(s_i)) = \boldsymbol{\mu}^\top(s_i)\mathbf{P}_{t|T}\boldsymbol{\mu}(s_i) \tag{48}$$

where $\mathbf{P}_{t|T}$ is the covariance matrix associated with the state estimates obtained from the RTS smoother. Variance based on (48) at each spatial location and at $t = 50$ is shown in Fig.**??**. 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.

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

VII. 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.

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: 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

$$\begin{aligned}\langle N_m(s-l_1), N_{m'}(s-l_2)\rangle &= N_{m+m'}(m+l_1-l_2)\\ &= N_{m+m'}(m'+l_2-l_1)\end{aligned} \quad (49)$$

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

$$N_m\left(\frac{m}{2}+s\right) = N_m\left(\frac{m}{2}-s\right) \quad (50)$$

A direct consequence of (50) is

$$N_m(s) = N_m(m-s) \quad (51)$$

Therefore

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

■

REFERENCES

[1] H. Wilson and J. Cowan, "A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue," *Biological Cybernetics*, vol. 13, no. 2, pp. 55–80, 1973.

[2] S. Amari, "Dynamics of pattern formation in lateral-inhibition type neural fields," *Biological Cybernetics*, vol. 27, no. 2, pp. 77–87, 1977.

[3] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, Cambridge MA, 2005.

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

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

[6] 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.

[7] 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.

[8] 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.

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

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

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

[12] G. Evensen, "The ensemble kalman filter: theoretical formulation and practical implementation," *Ocean Dynamics*, vol. 53, pp. 343–367, 2003.

[13] 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.

[14] P. Houtekamer and H. Mitchell, "Ensemble kalman filtering," *Quarterly Journal of the Royal Meteorological Society*, vol. 131, no. 613, pp. 3269–3289, 2005.

[15] 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.

[16] 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.