This is an author produced version of *A canonical space-time state space model: state and parameter estimation* .

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/3466/

Article:
Dewar, M. and Kadirkarnanathan, V. (2007) A canonical space-time state space model: state and parameter estimation. IEEE Transactions on Signal Processing, 55 (10). pp. 4862-4870. ISSN 1053-587X

http://dx.doi.org/10.1109/TSP.2007.896245

promoting access to
White Rose research papers

eprints@whiterose.ac.uk
http://eprints.whiterose.ac.uk/

A Canonical Space-Time State Space Model: State and Parameter Estimation

Michael Dewar and Visakan Kadirkamanathan, *Member, IEEE*

***Abstract*—The maximum likelihood estimation of a dynamic spatiotemporal model is introduced, centred around the inclusion of a prior arbitrary spatiotemporal neighborhood description. The neighborhood description defines a specific parameterization of the state transition matrix, chosen on the basis of prior knowledge about the system. The model used is inspired by the spatiotemporal ARMA (STARMA) model, but the representation used is based on the standard state-space model. The inclusion of the neighborhood into an expectation-maximization based joint state and parameter estimation algorithm allows for accurate characterization of the spatiotemporal model. The process of including the neighborhood, and the effect it has on the maximum likelihood parameter estimate is described and demonstrated in this paper.**

***Index Terms*—Dynamic spatiotemporal modeling, expectation-maximization (EM) algorithm, maximum likelihood parameter estimation, state-space.**

I. Introduction

DYNAMIC spatiotemporal systems can be represented using models which describe the correlation information found in observations of the system. A typical assumption is that this correlation is local in nature, with the implication that global behavior in the observations is an emergent property of local interactions.

A popular technique to describe local correlations is to define a neighborhood which limits the modeled correlation to a local spatio–temporal region. Lattice based spatio–temporal models such as cellular automata [1] and coupled map lattices [2] use this technique, where the neighborhood is defined in terms of a spatio–temporal translation operator [3]. However, these models are restricted to data sets with a regular lattice structure [4].

The spatio–temporal autoregressive moving average (STARMA) model [5]–[7] avoids the restriction of lattice based models whilst maintaining the neighborhood structure. This motivates their use for real-world system identification [8], where spatially correlated signals are often gathered which, whilst being regularly sampled in time, are irregular in space [9].

STARMA models represent the spatio–temporal process using a set of correlated time series, an approach which lends itself to systems which are observed at a reasonably small number of observation locations but which are heavily sampled in time. For example, in oesophageal station manometry [10], a small number of pressure sensors are placed in the oesophagus in order to record the pressure across time at several locations during peristalsis. Similarly, although on a different spatial scale, multielectrode probes make ensemble recordings of neural activity at a number of different spatial locations over time [11]. A third example is given by [12] wherein measurements of truck flows were made monthly at eight locations on the Mexico-Texas border over a period of three years. In each of these cases, STARMA models are applicable as models of the underlying processes.

This paper's aim is to introduce the idea of a neighborhood description to the dynamic state space framework to model spatially correlated time series. The estimation of the state space model consists of using the observed field to estimate the hidden field and model parameters, both of which are constrained by the neighborhood description. From the point of view of estimation these unknown quantities are conditionally dependent, so an iterative technique is used to solve the joint estimation problem.

In a maximum likelihood framework, the natural solution to such a problem is to use the well-known expectation-maximization (EM) algorithm. The application of the EM algorithm to linear dynamic systems [13] has potential advantages over the more popular subspace methods [14], [15]. Importantly, the maximum-likelihood construction allows direct inclusion of a neighborhood-based parameterization of the state-space model which can subsequently be used to estimate the hidden field. In this context, the algorithm utilizes the Kalman Smoother [16] to perform expectation with respect to the hidden field, before analytically maximizing the resulting likelihood function.

This paper introduces a principled method of including this neighborhood information using a neighborhood-based parameterization mapping, and describes the resulting estimation algorithm within the EM framework. Section II describes the spatio–temporal model, the neighborhood definition and an algorithm to generate the necessary parameterization mapping. Section III describes the EM-based algorithm for estimation of the spatio–temporal model. Section IV illustrates the developed techniques using a selection of synthetic models. Finally, Section V concludes.

II. Spatio–Temporal Model

The spatio–temporal process to be modeled exists in the space formed by $\mathcal{S} \times \mathcal{T}$ where $\mathcal{S}$ is the spatial domain of interest and $\mathcal{T}$ is the temporal domain. The temporal domain is always assumed to be one dimensional and the process is assumed to be causal. The spatial domain can be up to three dimensions with

Manuscript received June 30, 2006; revised December 14, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zidong Wang. This work was supported by the EPSRC.

The authors are with the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, S1 3JD, U.K. (e-mail: m.dewar@sheffield.ac.uk; visakan@sheffield.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2007.896245

1053-587X/$25.00 © 2007 IEEE

no inherent causality. These fundamental properties give structure to the spatio–temporal problem, leading to much discussion about symmetry and separability of representation [17].

The spatio–temporal system is observed as a set of spatially arranged and correlated time series. This motivates the use of a spatio–temoprally indexed hidden variable, such that the current hidden field is comprised of current and past filtered values of the time series, where the number of included past values depends on the maximum autoregressive order of the spatio–temporal process. Let n_t denote the maximum temporal autoregressive order of the process and let $x(s,t)$ be a hidden variable at a specific spatio–temporal location $(s \in \mathcal{S}, t \in \mathcal{T})$.

Assumption 1: The dynamics of the hidden field are represented by

$$x_{t+1} = Ax_t + Ww_t \tag{1}$$

where $x_t \in \mathbb{R}^{n_x}$ denotes the state vector

$$\begin{aligned} x_t = [&x(s_1,t)x(s_2,t)\ldots x(s_{n_y},t) \\ &\times x(s_1,t-1)x(s_2,t-1)\ldots x(s_{n_y},t-1) \\ &\times x(s_1,t-2)\ldots\ldots x(s_{n_y},t-n_t+1)]^\top \end{aligned}$$

where n_y is the number of observation locations and where the superscript T symbol denotes the transpose operator. The state matrix is arranged in the following canonical form

$$A = \begin{bmatrix} \bar{A} & \\ I & \mathbf{0} \end{bmatrix} \tag{2}$$

where $\bar{A} \in \mathbb{R}^{n_y \times n_x}$ contains parameters and I and $\mathbf{0}$ denote the identity and zero matrices respectively, such that $A \in \mathbb{R}^{n_x \times n_x}$. The n_x by n_y matrix $W = [I\ \mathbf{0}]^\top$ maps the state disturbance $w_t \in \mathbb{R}^{n_y}$ onto the next state. The disturbance on the state is modeled using Gaussian white noise $w_t \sim \mathcal{N}(0, \Sigma_w)$ where $\Sigma_w \in \mathbb{R}^{n_y \times n_y}$ and $\mathcal{N}(\mu, \Sigma)$ denotes a Gaussian distribution with mean μ and covariance Σ. The collection of states up to time T is defined as $X = \{x_1 \ldots x_T\}$.

Assumption 2: The mapping between elements of the hidden field and the observed field is given by

$$y_t = Cx_t + v_t \tag{3}$$

where $t \in \mathcal{T} \subset \mathbb{Z}$ denotes discrete-time. The n_y by n_x observation matrix $C = [I\ \mathbf{0}]$ is constructed so that the current output is a noise corrupted version of the hidden variables $\bar{x}_t$. The observation vector $y_t \in \mathbb{R}^{n_y}$ is formed from the current value of the time series associated with each observation location

$$y_t = [y(s_1,t)\ldots y(s_{n_y},t)]^\top$$

where $s_i \in \mathcal{S} \subset \mathbb{R}^{n_\mathcal{S}}$ is a spatial location and $n_\mathcal{S}$ is the number of spatial dimension. The observation disturbance is denoted $v_t \in \mathbb{R}^{n_y}$ and is modeled by Gaussian white noise with distribution $v_t \sim \mathcal{N}(0, \Sigma_v)$ where $\Sigma_v \in \mathbb{R}^{n_y \times n_y}$. The collection of observations up to time T is defined as $Y = \{y_1 \ldots y_T\}$ and the collection of both the states and the observations is denoted $Z = \{X, Y\}$.

Assumption 3: The system is assumed to be stationary in time.

The above assumptions define the model of the spatio–temporal system and have a number of implications. The construction of W in Assumption 1 implies that state disturbance w_t is only associated with the current value of the filtered time series $x(s_1,t)\ldots x(s_{n_y},t)$ and not with the past time series values which together construct the hidden field at time t. This is due to the canonical structure of the model. Assumption 3 implies that the parameters of the model are time invariant. Note that this does not imply that the field to be modeled is completely homogeneous. Rather, each hidden variable dynamical process has spatial-location specific parameters, but these parameters remain invariant over time.

By partitioning the state vector into current and past hidden variables

$$x_t = \left[\bar{x}_t^\top \tilde{x}_t^\top\right]^\top$$

where the partition

$$\bar{x}_t = [x(s_1,t)x(s_2,t)\ldots x(s_{n_y},t)]^\top$$

and the remainder of the state vector is denoted $\tilde{x}_t$ allows the model be written

$$\bar{x}_t = \bar{A}x_{t-1} + w_t \tag{4}$$

$$\tilde{x}_t = [I\ 0]x_{t-1} \tag{5}$$

$$y_t = Cx_t + v_t. \tag{6}$$

This structured form of the model is observable and unique [18].

A. Neighborhood

Definition 1: The *neighborhood* associated with a hidden variable at spatio–temporal location $(\varsigma \in \mathcal{S}, \tau \in \mathcal{T})$ is a known subregion of $\mathcal{S} \times \mathcal{T}$. Hidden variables which fall within the neighborhood of $x(\varsigma,\tau)$ are known as *neighbors* of the variable at (ς,τ). The set of neighbors associated with the hidden variable $x(\varsigma,\tau)$ is denoted $N(x(\varsigma,\tau))$.

As an example neighborhood, consider the oesophageal peristalsis example mentioned in Section I. Station manometry allows the collection of time series at locations along the length of the oesophagus, as shown in Fig. 1. As a patient swallows, sensors measure the pressure as the peristaltic wave travels down the oesophagus. Due to this downward direction of the peristaltic wave, a reasonable assumption is to choose a neighborhood that describes the hidden variable at a particular location in relation to those above it. If the spatial location is measured as the distance from the top of the oesophagus, such a neighborhood could be described using

$$N(x(\varsigma,\tau)) = \bigcup_{i=1}^{n_t} \{x(s,\tau-i) : d_i < s < \varsigma\} \tag{7}$$

where, at each lag $\tau - i$, a spatial area above ς with upper bound d_i defines the neighborhood of $x(\varsigma,\tau)$.

Assumption 4: A hidden variable is conditionally independent of all variables outside its neighborhood, such that

$$\begin{aligned} &\mathrm{p}(x(\varsigma,\tau)|x(s,t);(s,t) \in \mathcal{S} \times \mathcal{T}) \\ &\qquad = \mathrm{p}(x(\varsigma,\tau)|x(s,t);(s,t) \in N(\varsigma,\tau)) \end{aligned} \tag{8}$$

where $\varsigma \in \mathcal{S}$ and $\tau \in \mathcal{T}$.

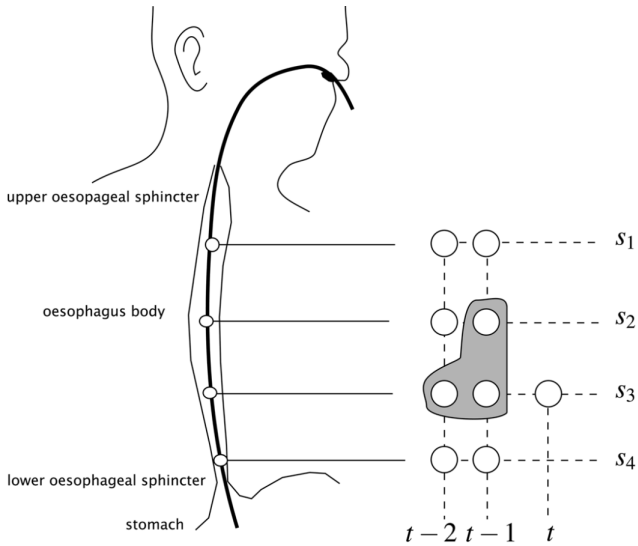

Fig. 1. Neighborhood construction to model the oesophagus from station manometry observations. Each circle represents a hidden variable at a specific space and time. The shaded area is the neighborhood of the hidden variable at (s_3, t).

TABLE I
CONSTRUCTION OF THE NEIGHBORHOOD MAPPING MATRIX

for $i = 1$ to n_y and $j = 1$ to n_x
if $[x_t]_j \in N([x_{t+1}]_i)$
$k = (j-1)n_y + i$
$\Delta = [\Delta\ e_k]$
Order Δ by k

B. Structure in $\bar{A}$

The neighborhood definition introduces extra structure into the parameter matrix $\bar{A}$ where, following Assumption 4, parameters representing relationships between non-neighboring states are known to be zero-valued. A mapping from the (known and unknown) parameter space defined by $\bar{A}$ to the unknown parameter space is developed.

Let n_ϕ denote the number of unknown parameters in $\bar{A}$ and let the unknown parameter vector be denoted $\phi \in \mathbb{R}^{n_\phi}$ such that $n_\phi \leq n_x n_y$ with equality when the neighborhood does not introduce structure to the $\bar{A}$ matrix.

Definition 2: Given an arbitrary neighborhood N the corresponding mapping $\Delta \in \{0,1\}^{n_y n_x \times n_\phi}$ between the unknown parameter vector ϕ and the matrix $\bar{A}$ is defined by

$$\mathrm{vec}(\bar{A}) = \Delta\phi \tag{9}$$

Let e_k be drawn from the usual $n_y n_x$-dimensional Euclidean basis, such that the kth element of e_k is equal to 1 and zero otherwise. Then Δ can be constructed using the algorithm given in Table I.

To further extend the oesophageal example, consider the scenario depicted in Fig. 1. Pressure measurements are taken at four locations within the oesophageal body. The system is considered to be homogenous, therefore, the neighborhood is the same shape for each hidden variable; shown on the diagram is the neighborhood of $x(s_3, t)$. Any variable outside the neighborhood implies a zero-valued element on the third row of $\bar{A}$. The neighborhood shown in Fig. 1 would generate

$$\bar{A} = \begin{bmatrix} a_{11} & 0 & 0 & 0 & a_{15} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 & a_{26} & 0 & 0 \\ 0 & a_{32} & a_{33} & 0 & 0 & 0 & a_{37} & 0 \\ 0 & 0 & a_{43} & a_{44} & 0 & 0 & 0 & a_{48} \end{bmatrix}$$

were Δ would be constructed via the algorithm given in Table I as

$$\Delta = [e_1\ e_2\ e_6\ e_7\ e_{11}\ e_{12}\ e_{16}\ e_{17}\ e_{22}\ e_{27}\ e_{32}].$$

III. ESTIMATION

The EM algorithm provides a well-known framework for approaching the joint state and parameter estimation problem for the general, linear state-space model. Introduced by Shumway and Stoffer [13] and recently revisited by Gibson and Ninness [19], it presents an alternative to subspace-based, dual filtering, and gradient descent techniques. In the context of the spatio–temporal model outlined earlier, the construction of the likelihood for the EM algorithm's M-step presents an opportunity to include the neighborhood information into the estimation procedure, without losing the beneficial properties of the estimator as described by Gibson and Ninness. This section describes the inclusion of the canonical form and spatio–temporal neighborhood based parameterization into the estimator and presents an algorithm to estimate the states and parameters of the spatio–temporal model described earlier.

A. The Likelihood Function

Maximum likelihood estimation seeks to find parameters

$$\phi_{\mathrm{ML}} = \arg\max_{\phi} \mathrm{p}(Y;\phi)$$

The EM algorithm approximates $\mathrm{p}(X,Y;\phi)$ with respect to a prior parameter estimate and, once approximated, a closed form solution to $\phi^* = \arg\max_\phi \mathrm{p}(X,Y;\phi)$ can be found. By exploiting the relationship between $\mathrm{p}(X,Y;\phi)$ and $\mathrm{p}(Y;\phi)$ it is possible to generate a sequence of parameter estimates that converges on the maximum likelihood parameter estimate [19]. The complete-data log-likelihood is defined as

$$\ell(\phi) = \ln \mathrm{p}(X,Y;\phi)$$

which can be written in terms of the model's component densities by repeated application of Bayes' rule

$$\ell(\phi) = \ln \mathrm{p}(x_0) + \sum_{t=1}^{T} \ln \mathrm{p}(y_t \mid x_t) + \sum_{t=0}^{T} \ln \mathrm{p}(\bar{x}_{t+1} \mid x_t) + \sum_{t=0}^{T} \ln \mathrm{p}(\tilde{x}_{t+1} | x_t) \tag{10}$$

noting that $\bar{x}_t$ and $\tilde{x}_t$ are conditionally independent and where the component densities are written

$$\begin{aligned} \mathrm{p}(x_0) &= \mathcal{N}(\mu, \Pi) \\ \mathrm{p}(\bar{x}_t|x_{t-1}) &= \mathcal{N}(\bar{A}x_{t-1}, \Sigma_w) \\ \mathrm{p}(\tilde{x}_t|x_{t-1}) &= \delta([I0]x_{t-1}) \\ \mathrm{p}(y_t|x_t) &= \mathcal{N}(Cx_t, \Sigma_v) \end{aligned}$$

where $\delta(a)$ denotes the Dirac delta function $\delta(\tilde{x}_t - a)$. Note that only $\mathrm{p}(\bar{x}_t|x_{t-1})$ is a function of ϕ.

B. The M-Step

The problem of concurrently estimating both the parameter set and the state sequence is solved by the EM algorithm through taking expectations of $\ell(\phi)$ with respect to an estimate of X, conditional on the current parameter set $\phi^{(i-1)}$.

Definition 3: The so-called Q-function is given by

$$Q(\phi, \phi^{(i-1)}) = \mathrm{E}[\ln \mathrm{p}(X, Y; \phi)] \tag{11}$$

where the expectation is taken with respect to the distribution $p(X|Y; \phi^{(i-1)})$.

After evaluating the expectation, the Q-function becomes a deterministic function of ϕ, which can be maximized. To introduce the neighborhood structure into the parameter estimation problem, the Q-function is expressed in terms of the parameter vector ϕ and the known neighborhood mapping Δ.

Lemma 1: The Q-function in Definition 3 for the spatio–temporal system (4–6) can be written in the following form

$$\begin{aligned} Q(\phi, \phi^{(i-1)}) = \alpha &+ 2\mathrm{vec}(I)^\top (\Xi_x^\top \otimes \Sigma_w^{-1})\mathrm{vec}(\bar{A}) \\ &- \mathrm{vec}(\bar{A})^\top (\Xi_{xx}^\top \otimes \Sigma_w^{-1})\mathrm{vec}(\bar{A}) \end{aligned} \tag{12}$$

where $\otimes$ denotes the Kronecker product, Ξ_x, Ξ_{xx} denote $\sum_{t=1}^T \mathrm{E}[x_{t-1}\bar{x}_t^\top]$, and $\sum_{t=1}^T \mathrm{E}[x_{t-1}x_{t-1}^\top]$ respectively and α denotes a constant conditionally independent of ϕ.

Proof: The component densities are substituted into (11) to give

$$Q(\phi, \phi^{(i-1)}) = \mathrm{E}\left[\mathrm{const.} - \sum_{t=1}^T (\bar{x}_t - \bar{A}x_{t-1})^\top \Sigma_w^{-1}(\bar{x}_t - \bar{A}x_{t-1})\right]$$

where the constant term collects together all the quantities not dependent on the unknown parameters. Using properties of the trace operator, the above equation can be expanded and rearranged to produce

$$\begin{aligned} Q(\phi, \phi^{(i-1)}) = \mathrm{const.} &+ 2\mathrm{tr}\left(\sum_{t=1}^T \mathrm{E}\left[x_{t-1}\bar{x}_t^\top\right] \Sigma_w^{-1}\bar{A}\right) \\ &- \mathrm{tr}\left(\sum_{t=1}^T \mathrm{E}\left[x_{t-1}x_{t-1}^\top\right] \bar{A}^\top \Sigma_w^{-1}\bar{A}\right). \end{aligned} \tag{13}$$

Note that the constant term is extended to include the term $\mathrm{E}[\bar{x}_t^\top \Sigma_w^{-1}\bar{x}_t]$ which is conditionally independent of ϕ. Further properties of the Kronecker product, vectorize and trace operators [20] are used to manipulate (13) to produce the given result. With $\Xi_x = \sum_{t=1}^T \mathrm{E}[x_{t-1}\bar{x}_t^\top]$ and $\Xi_{xx} = \sum_{t=1}^T \mathrm{E}[x_{t-1}x_{t-1}^\top]$ the first component of (13) can be written

$$\begin{aligned} \mathrm{tr}\left(\Xi_x \Sigma_w^{-1}\bar{A}\right) &= \mathrm{vec}\left(\Sigma_w^{-1}\Xi_x^\top\right)^\top \mathrm{vec}(\bar{A}) \\ &= \mathrm{vec}(I)^\top \left(\Xi_x^\top \otimes \Sigma_w^{-1}\right)\mathrm{vec}(\bar{A}). \end{aligned} \tag{14}$$

Similarly, the second component of (13) can be written as

$$\begin{aligned} \mathrm{tr}\left(\Xi_{xx}\bar{A}^\top \Sigma_w^{-1}\bar{A}\right) &= \mathrm{vec}(\bar{A})^\top \mathrm{vec}\left(\Sigma_w^{-1}\bar{A}\Xi_{xx}\right) \\ &= \mathrm{vec}(\bar{A})^\top \left(\Xi_{xx} \otimes \Sigma_w^{-1}\right)\mathrm{vec}(\bar{A}). \end{aligned} \tag{15}$$

The result of Lemma 1 follows by substituting (14) and (15) into (13). ■

Given the Q-function in terms of $\mathrm{vec}(\bar{A})$, the inclusion of the spatio–temporal neighborhood can be made using the result from Definition 2, allowing the estimation of ϕ. The following Lemma shows that the necessary inversion can be performed, followed by the main result.

Lemma 2: The matrix $\Delta^\top(\Xi_{xx} \otimes \Sigma_w^{-1})\Delta$ is invertible.

Proof: Let $F = \Delta^\top(\Xi_{xx} \otimes \Sigma_w^{-1})\Delta$ and let $G = \Xi_{xx} \otimes \Sigma_w^{-1}$. The sum of dyads Ξ_{xx} is, by definition, positive semidefinite and becomes positive definite under the persistent excitation condition, which is guaranteed by the disturbance w_t. By [20] the Kronecker product of two positive definite matrices is also positive definite therefore G is a positive definite matrix.

By definition, F is positive definite if and only if $\langle Fz, z\rangle > 0$ for any nonzero vector $z \in \mathbb{R}^{n_\phi}$, where $\langle \cdot, \cdot \rangle$ denotes the inner product. Substituting for F in the inner product gives

$$\begin{aligned} \langle Fz, z\rangle &= \langle \Delta^\top G\Delta z, z\rangle \\ &= z^\top \Delta^\top G\Delta z = \langle Gu, u\rangle > 0 \end{aligned}$$

since G is positive definite, where $u = \Delta z$. The inequality holds only for nonzero u, which is guaranteed as long as $\mathrm{rank}(\Delta) \geq \dim(z)$. By Definition 2, $\mathrm{rank}(\Delta) = \dim(z)$, hence F is positive definite, and thus invertible, completing the proof. ■

Theorem 1: The estimate of the unknown parameters $\hat{\phi}$ that locally maximizes the Q-function of the spatio–temporal model described by (4–6), (9) is given by

$$\hat{\phi} = \left(\Delta^\top \left(\Xi_{xx} \otimes \Sigma_w^{-1}\right)\Delta\right)^{-1} \Delta^\top \left(\Xi_x \otimes \Sigma_w^{-1}\right)\mathrm{vec}(I). \tag{16}$$

Proof: The Q-function expresses the expectation of the log-likelihood of the state and observation sequences given a candidate parameter set, conditional on the observations and current parameter set

$$Q(\phi, \phi^{(i-1)}) = \mathrm{E}[\ln \mathrm{p}(X, Y; \phi)].$$

Following Lemma 1 the Q-function can be written in terms of the vectorized $\bar{A}$ matrix

$$\begin{aligned} Q(\phi, \phi^{(i-1)}) = \alpha &+ 2\mathrm{vec}(I)^\top \left(\Xi_x^\top \otimes \Sigma_w^{-1}\right)\mathrm{vec}(\bar{A}) \\ &- \mathrm{vec}(\bar{A})^\top \left(\Xi_{xx} \otimes \Sigma_w^{-1}\right)\mathrm{vec}(\bar{A}). \end{aligned}$$

The vectorized $\bar{A}$ matrix can be written in terms of the unknown parameter vector and the known mapping Δ, by substituting (9) into (12), thereby enforcing the neighborhood definition

$$Q(\phi, \phi^{(i-1)}) = \alpha + 2\mathrm{vec}(I)^\top \left(\Xi_x^\top \otimes \Sigma_w^{-1}\right) \Delta\phi - \phi^\top \Delta^\top \left(\Xi_{xx} \otimes \Sigma_w^{-1}\right) \Delta\phi$$

Differentiating the Q-function with respect to ϕ gives

$$\frac{dQ}{d\phi} = 2\mathrm{vec}(I)^\top \left(\Xi_x^\top \otimes \Sigma_w^{-1}\right) \Delta - 2\phi^\top \Delta^\top (\Xi_{xx} \otimes \Sigma_w^{-1})\Delta.$$

Equating the above to zero and rearranging gives

$$\mathrm{vec}(I)^\top \left(\Xi_x^\top \otimes \Sigma_w^{-1}\right) \Delta = \phi^\top \Delta^\top \left(\Xi_{xx} \otimes \Sigma_w^{-1}\right) \Delta.$$

The result follows by premultiplying both sides by $(\Delta^\top(\Xi_{xx} \otimes \Sigma_w^{-1})\Delta)^{-1}$. The second derivative of the Q-function is given by

$$\frac{d^2Q}{d\phi^2} = -2\Delta^\top \left(\Xi_{xx} \otimes \Sigma_w^{-1}\right) \Delta$$

which, by Lemma 2, is negative definite and, therefore, the derived estimate is located at a local maximum of the Q-function. ■

There are three special cases which remove the dependence of the estimate of ϕ on Σ_w, given in the following corollaries.

Corollary 1: If each hidden variable is subject to uncorrelated disturbance from the same distribution, then the parameter estimate that maximizes the Q-function is given by

$$\hat{\phi} = \left(\Delta^\top (\Xi_{xx} \otimes I)\, \Delta\right)^{-1} \left(\Delta^\top (\Xi_x \otimes I)\right) \mathrm{vec}(I). \quad (17)$$

Corollary 2: If no spatial neighborhood structure is defined the maximum likelihood estimate of the unknown parameters is given by

$$\hat{\phi} = \mathrm{vec}\left(\Xi_{xx}^{-1}\Xi_x\right).$$

Corollary 3: If the neighborhood transformation matrix is restricted to $\Delta = \bar{\Delta} \otimes I$ where $\bar{\Delta} \in \mathbb{R}^{n_x \times (n_\phi/n_y)}$ is constructed from the n_x-dimensional Euclidean basis, then the parameter estimate reduces to

$$\phi = \mathrm{vec}\left(\left(\Xi_x^\top \bar{\Delta}\right)\left(\bar{\Delta}^\top \Xi_{xx} \bar{\Delta}\right)^{-1}\right).$$

The proofs of the above corollaries follow from direct substitution and algebraic manipulation. The use of $\bar{\Delta}$ in Corollary 3 implies a restricted neighborhood which introduces vertical bands of parameters in $\bar{A}$. This implies that the *same* subset of variables in the hidden field at time $t-1$ affects each hidden variable in $\bar{x}_t$. As an example, consider a process which is monitored over time at a large number of observation locations, and suppose that the majority of the dynamic process behavior can be explained by the observations at a subset of those locations. Then a neighborhood constructed as in Corollary 3 can be used to create a model of the process that only depends on the subset of informative observation locations, while still allowing the state of the process at all the observation locations to be estimated.

C. The E-Step

The expectation step consists of evaluating the Q-function, given the current parameter set $\phi^{(i-1)}$ and the observed field. Practically, this involves calculating the expectations in Ξ_x and Ξ_{xx}, given in the following Lemma.

Lemma 3: Conditional on the current parameter set $\phi^{(i-1)}$ and the hidden field Y, the values of Ξ_x and Ξ_{xx} are given by

$$\Xi_x = \sum_{t=1}^{T} \left(\bar{M}_t + \hat{\bar{x}}_t \hat{x}_{t-1}^\top\right)$$
$$\Xi_{xx} = \sum_{t=1}^{T} \left(P_{t-1} + \hat{x}_{t-1}\hat{x}_{t-1}^\top\right)$$

where $\hat{x}$ denotes the expected value of x, P_t the covariance of x_t and $\bar{M}_t$ denotes the first n_y rows of the covariance matrix of x_t and x_{t-1}.

Proof: The proof follows that of [19]. Recall that

$$\Xi_x = \sum_{t=1}^{T} \mathrm{E}\left[x_{t-1}\bar{x}_t^\top\right]$$
$$\Xi_{xx} = \sum_{t=1}^{T} \mathrm{E}\left[x_{t-1}x_{t-1}^\top\right].$$

Using the definition of covariance and linearity of the expectation operator it is straightforward to show that

$$P_t = E\left[x_t x_t^\top\right] - \hat{x}_t\hat{x}_t, \quad M_t = E\left[x_t x_{t-1}^\top\right] - \hat{x}_t\hat{x}_{t-1}.$$

The result follows by partitioning M_t such that

$$E\left[\bar{x}_t x_{t-1}^\top\right] = \bar{M}_t + \hat{\bar{x}}_t\hat{x}_{t-1}$$

■

Given the model and observed field, the expected value of the state at a given time can be calculated using the standard Kalman Smoother, with an extra recursion to calculate the covariance M_t [21]. This algorithm is given in Table II, where the notation $\hat{x}_{t|t-1}$ denotes the expected value of x_t given information up to time $t-1$.

D. The Estimation Algorithm

The E- and M-steps of Sections III-C and B are iterated until convergence. The algorithm requires a method to initialize either an initial parameter set or an initial state sequence. Typically, a mean-squared error parameter estimation technique is employed to generate an initial parameter set, however here the structure of the C-matrix can be exploited to populate a state sequence using the observed values of Y. This is then used in an M-step to generate an initial parameter set. Following [22], the change in a function of parameter values is used to generate stopping criteria. The algorithm will halt when

$$(\lambda^{(i)} - \lambda^{(i-1)}) > \epsilon \quad (18)$$

TABLE II
E-STEP: THE KALMAN SMOOTHER

Initialize $\hat{x}_0,\ P_0$

Forward iteration:

$P_{t|t-1} = AP_{t-1|t-1}A^\top + \Sigma_w$

$\hat{x}_{t|t} = A\hat{x}_{t-1|t-1} + P_{t|t-1}C^\top(CP_{t|t-1}C^\top + \Sigma_v)^{-1}(y_t - C\hat{x}_{t|t-1})$

$P_{t|t} = P_{t|t-1} - P_{t|t-1}C^\top(CP_{t|t-1}C^\top + \Sigma_v)^{-1}CP_{t|t-1}$

Backwards iteration:

$S_t = P_{t|t}A^\top P_{t+1|t}^{-1}$

$\hat{x}_{t|T} = \hat{x}_{t|t} + S_t(\hat{x}_{t+1|T} - \hat{x}_{t+1|t})$

$P_{t|T} = P_{t|T} + S_t(P_{t+1|T} - P_{t+1|t})S_t{}^\top$

Initialize:

$M_{T|T} = (I - P_{T|T-1}C^\top(CP_{T|T-1}C^\top + \Sigma_v)^{-1}C)AP_{T-1|T-1}$

Backwards iteration:

$M_{t|T} = P_{t|t}S_{t-1}^\top + S_t(M_{t+1|T} - AP_{t|t})S_{t-1}^\top$

TABLE III
MAXIMUM LIKELIHOOD STATE AND PARAMETER ESTIMATION FOR SPATIO–TEMPORAL STATE-SPACE MODELS

Initialize:

Generate Δ using the algorithm given in Table I

Populate the state sequence using observations.

Choose threshold ϵ and set $(\lambda - \lambda^{(i-1)}) > \epsilon$

While $(\lambda - \lambda^{(i-1)}) > \epsilon$

E-step:

Execute the algorithm of Table II to give

X, $\{P_{1|T} \ldots P_{T|T}\}$, and $\{M_{1|T} \ldots M_{T|T}\}$.

M-step:

Calculate the expectation matrices:

$\Xi_x = \sum_{t=1}^{T}\left(\bar{M}_t + \hat{\bar{x}}_t\hat{x}_{t-1}{}^\top\right)$

$\Xi_{xx} = \sum_{t=1}^{T}\left(P_{t-1} + \hat{x}_{t-1}\hat{x}_{t-1}{}^\top\right)$

Calculate the parameter estimates :

$\phi = -\left(\Delta^\top\left(\Xi_{xx}{}^\top \otimes \Sigma_w^{-1}\right)\Delta\right)^{-1}\Delta^\top\left(\Xi_x{}^\top \otimes \Sigma_w^{-1}\right)\mathrm{vec}(I)$

Form $\mathrm{vec}(\bar{A}) = \Delta\phi$

Calculate stopping criteria using (18)

where ϵ is a threshold and $\lambda^{(i-1)}$ and $\lambda^{(i)}$ are the maximum eigenvalue of the previous and the updated $A^\top A$ matrices, respectively.

Theorem 2: The algorithm given in Table III generates a sequence of parameter estimates such that

$$L(\phi^{(i)}) \geq L(\phi^{(i-1)})$$

with equality if and only if $\phi^{(i+1)} = \phi^{(i)}$ and which converges to a local maximum $L(\phi^*)$.

Proof: The expected incomplete-data log-likelihood function is given by

$$L(\phi) = Q(\phi, \phi^{(i-1)}) - \mathrm{E}[\ln p(Z\,|\,Y;\phi)].$$

The change in $L(\phi^{(i)})$ over each iteration of the algorithm in Table III is given by

$$\begin{aligned} L(\phi^{(i)}) - L(\phi^{(i-1)}) = \{&Q(\phi^{(i)}, \phi^{(i-1)}) \\ &- Q(\phi^{(i-1)}, \phi^{(i-1)})\} \\ &+ \{\mathrm{E}[\ln p(Z|Y;\phi^{(i)})] \\ &- \mathrm{E}[\ln p(Z|Y;\phi^{(i-1)})]\} \end{aligned}$$

which is always nonnegative as $\mathrm{E}[\ln p(Z|Y;\phi^{(i)})] > \mathrm{E}[\ln p(Z|Y;\phi^{(i-1)})]$ following the standard application of Jensen's inequality [23] and $Q(\phi^{(i)}, \phi^{(i-1)}) \geq Q(\phi^{(i-1)}, \phi^{(i-1)})$ by Theorem 1, hence, $L(\phi^{(i)})$ is an increasing function of i.

By Theorem 1 and Definition 3 $Q(\phi, \phi^{(i-1)})$ is continuous in both arguments, satisfying the condition of [24, Theorem 2], application of which demonstrates convergence of the sequence $\{L(\phi^{(i)}), L(\phi^{(i+1)})\ldots\}$ to a local maximum $L(\phi^*)$.

The equality $L(\phi^{(i)}) = L(\phi^{(i-1)})$ is clearly true if $\phi^{(i+1)} = \phi^{(i)}$; the "only if" condition is demonstrated for a general dynamic state space model in [19, Corollary 5.1], and, hence, for the parameterized model (1), (3) completing the proof. ■

IV. SIMULATION EXAMPLES

This section illustrates the above approach to modeling spatio–temporal systems via a set of simulated examples. First, a simple, homogeneous 2-D, four-state system is shown, followed by a homogeneous 2-D, 12-state system and a heterogeneous 3-D, eight-state system. All the examples use 500 simulated time points and a convergence threshold of $\epsilon = 1 \times 10^{-10}$. As a measure of parameter bias, the same norm used in the convergence criteria is used and is referred to as the λ-value, that is $\lambda = \max(\mathrm{eig}(A^\top A))$. The reported bias is the percentage error in the λ-value, namely $1 - (\hat{\lambda}/\lambda^*)$ where $\hat{\lambda}$ and λ^* are the estimated and true λ-values respectively.

A. Estimation of A 2-D, Homogeneous, Four-State System

Initially a simple spatio–temporal model with two observation locations and four states is used. Fig. 2 represents this graphically where the shaded area describes the neighborhood of $x_1(t+1)$. The neighborhood of $x_2(t+1)$ only contains $x_2(t)$ and $x_4(t)$-a result of translating the neighborhood down by the spatial distance between y_1 and y_2. Following the algorithm given in Table I, this neighborhood generates a transformation matrix

$$\Delta = [e_1\ e_4\ e_5\ e_7\ e_8]$$

and parameter values are chosen as

$$\bar{A} = \begin{bmatrix} 1.3 & 0 & -0.8 & 0.9 \\ 0 & 1.2 & 0 & -0.5 \end{bmatrix}.$$

The disturbance and noise covariances are chosen to be $\Sigma_w = 0.8 \times I^{n_x}$ and $\Sigma_v = 0.2 \times I$.

To demonstrate consistent behavior, the algorithm was run until convergence 10 times with different noise realizations each time. Fig. 3 shows the parameter bias at each iteration for each run. The algorithm converges in an average of 26 iterations.

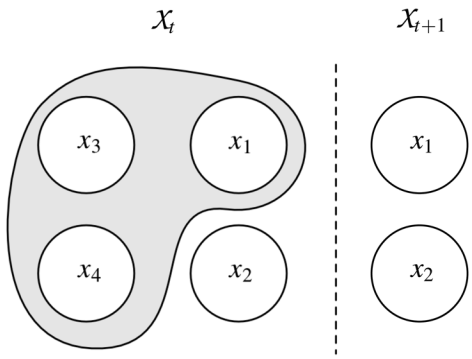

Fig. 2. Example neighborhood for Section IV-A. Part of the hidden field is shown for time $t+1$, and the whole hidden field is shown for time t. The neighborhood of $x_1(t+1)$ is shown by the shaded region.

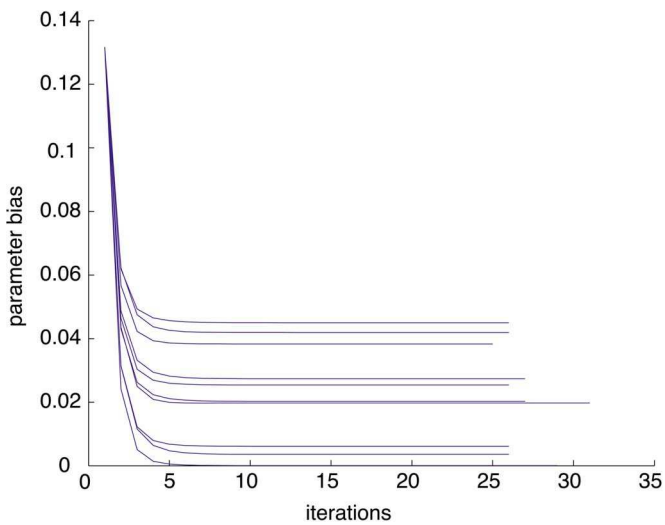

Fig. 3. Ten runs of the algorithm using different noise realizations. The figure shows the parameter bias as a percentage of the true λ-value.

TABLE IV
TRUE AND ESTIMATED PARAMETERS CORRESPONDING TO THREE CASES SHOWN IN FIG. 3

	Parameter Estimates		
True Parameters	Best	Median	Worst
1.3	1.29	1.31	1.33
-0.8	-0.78	-0.81	-0.83
0.9	0.93	0.91	0.91
1.2	1.30	1.27	1.31
-0.5	-0.54	-0.59	-0.57

Table IV presents the best, median and the worst parameter estimates of the 10 runs of the algorithm shown in Fig. 3.

Fig. 4 displays the sensitivity of the parameter estimate to changes in Σ_w used in the estimator. Here, 100 random Σ_w matrices were generated by choosing a random matrix R via Matlab's `rand()` function then setting $\Sigma_w = 0.1R^\top R$ to ensure a positive definite covariance matrix. Using the same state sequence, each randomly chosen Σ_w was presented to the maximization routine (16). This figure demonstrates the insensitivity

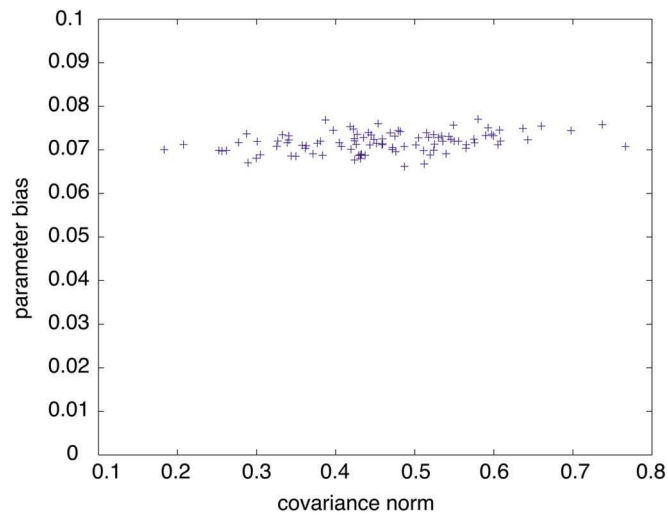

Fig. 4. Behavior of the maximize routine when presented with randomly chosen covariance matrices but the same data set.

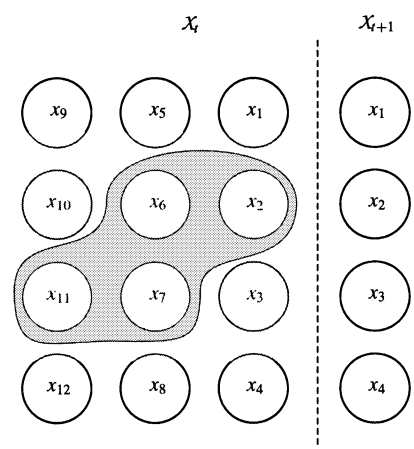

Fig. 5. Example neighborhood for Section IV-B. The neighborhood of $x_2(t+1)$ is shown by the shaded region.

of the algorithm to errors in the state disturbance covariance matrix. This is an important numerical property of the estimator, as Σ_w may not be well characterized in practice.

B. Estimation of a 2-D, Homogeneous, 12-State System

In order to emphasize the benefits of the neighborhood definition, Table V presents a comparison of the estimation of a standard vector-AR (VAR) model using least-squares with the result using the algorithm of Table III. The system has $n_x = 12$ states, $n_y = 4$ observations locations and a neighborhood definition as shown in Fig. 5. The neighborhood shape is the same for each hidden variable in $\bar{x}_{t+1}$. The disturbances are the same as for the previous example where $\Sigma_w = 0.8 \times I^{n_x}$ and $\Sigma_v = 0.2 \times I$.

Shown in the table are only the nonzero parameters and their estimates. Here, an estimate for the parameter vector of the VAR model is found using least squares on the original data set without the neighborhood definition and is displayed next to the result using the algorithm given in Table III.

The combination of the filtering inherent in the Kalman smoother and the dimension reduction in the parameter estimation problem due to the neighborhood definitions produces the overall increase in accuracy. Table V also indicates that

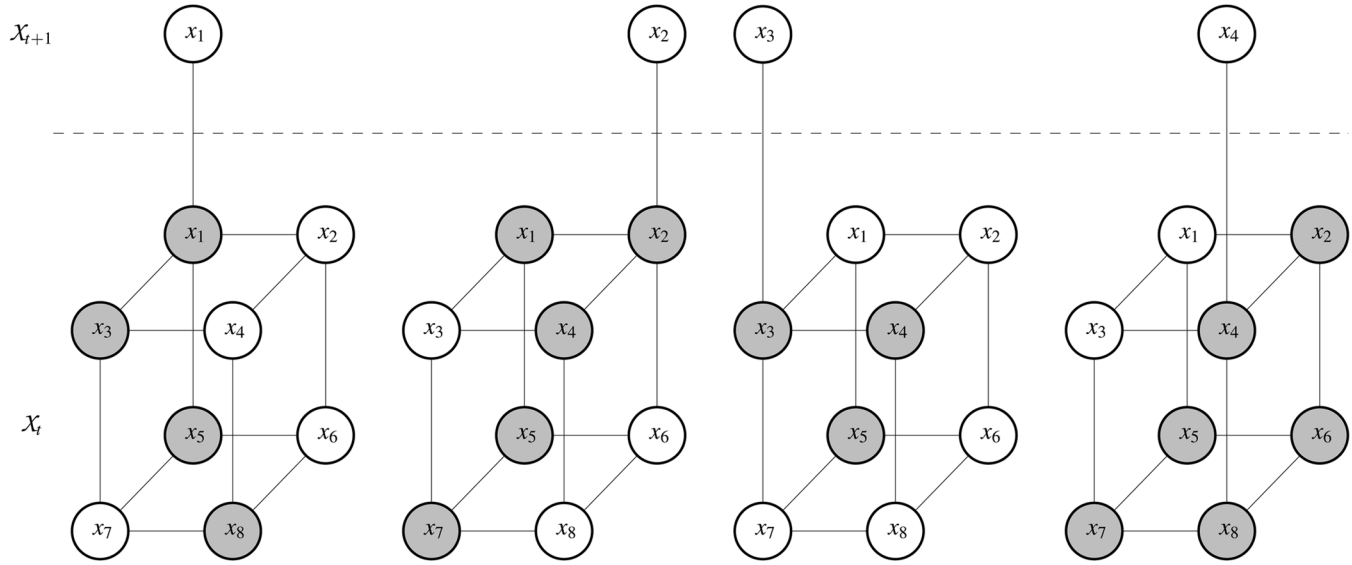

Fig. 6. A heterogenous 3-D example. The neighborhood of x_1 to x_4 is shown left-to-right by the shaded states.

TABLE V
PARAMETER ESTIMATES FOR A VAR MODEL COMPARED WITH THE SPATIO–TEMPORAL MODEL (Δ)

True	VAR	Δ	True	VAR	Δ
0.07	-0.01	-0.01	0.2	0.05	0.24
0.2	0.1	0.18	0.32	0.38	0.3
0.4	-0.04	0.39	0.38	0.27	0.36
0.45	0.07	0.48	0.1	-0.01	0.09
0.4	0.01	0.46	0.6	0	0.69
0.1	-0.03	0.11	0.4	0.02	0.5
0.1	-0.02	0.02	0.4	0.32	0.39

automatic neighborhood detection is infeasible by naïvely inferring Δ from zero-or small-valued parameter estimates from a nonparameterized system identification procedure.

To demonstrate the effect of the neighborhood mapping on the algorithm, note that the least-squares step applied above is of order $O(n_x^3) + O(Tn_x^2)$ and one M-step of Theorem 1 is of order $O(n_\phi^3) + O(n_y^2 n_x^2) + O(Tn_x^2)$, noting that the product associated with the neighborhood mappings can be considered a sorting operation. The order of the least-squares computation is typically lower, depending on the neighborhood parameterization. The E-step is typically less complex than the M-step (unless number of model parameters is significantly reduced by Δ and the observation sequence is either very long or consists of only a small number of observation locations). The complexity of the algorithm increases linearly with the number of iterations. For the example given above, a G4 PowerPC takes $\sim$3 s each for the E-step and the M-step, and $\sim$2.7 s for the least squares computation.

C. Estimation of a 3-D, Heterogeneous, Eight-State System

A 3-D example is shown in Fig. 6. Here $n_x = 8, n_y = 4$ and the parameters of $\bar{A}$ are given in Table VI. Again, the noise covariances $\Sigma_w = 0.8 \times I$ and $\Sigma_v = 0.2 \times I$. The neighborhood for each hidden variable in $\bar{x}_t$ is defined separately, rather than being translated versions of one another, creating a heterogeneous system whose behavior is dependent on the absolute spatial location. Using the algorithm given in Table I, this neighborhood generates a transformation matrix

$$\Delta = [e_1\ e_2\ e_6\ e_8\ e_9\ e_{11}\ e_{14}\ e_{15}\ e_{16}\ e_{17}\ e_{18}\ e_{19}\ e_{20}\ e_{24}\ e_{26}\ e_{28}\ e_{29}\ e_{32}]$$

leaving 18 unknown parameters to be estimated. The algorithm converges in 27 iterations.

TABLE VI
TRUE AND ESTIMATED VAR AND SPATIO–TEMPORAL MODEL (Δ) PARAMETERS FOR THE HETEROGENEOUS 3-D EXAMPLE

True	VAR	Δ	True	VAR	Δ
0.5	0.46	0.44	0.1	0.02	0.02
0.65	0.05	0.71	-0.25	-0.01	-0.22
-0.3	-0.02	-0.35	0.5	0.19	0.46
-0.4	0.30	-0.44	0.2	-0.58	0.27
0.3	0.73	0.35	-0.4	-0.02	-0.44
0.2	0.07	0.20	-0.35	-0.44	-0.38
-0.3	-0.01	-0.32	-0.5	0.47	-0.53
-0.6	-0.42	-0.60	0.25	0.29	0.29
0.4	0.00	0.46	0.2	0.12	0.12

The parameter estimates generated are shown in Table VI, compared with the estimated parameters of the VAR model. In this example, the VAR model parameters are estimated using a state sequence that has been smoothed using the true parameter values. The VAR parameter estimates still suffer due to the much larger space from which to select the parameter vector. This demonstrates the clear benefit of the model structure via incorporation of the neighborhood definition.

V. CONCLUSION

This paper has presented a framework to estimate the states and parameters of a spatio–temporal state space system. The emphasis has been on the neighborhood definition, how this

neighborhood effects the maximum likelihood parameter estimation problem and how it improves the modeling accuracy.

The class of model put forward can be considered nonstationary-in-space and stationary-in-time. Both the neighborhood and the associated parameters can be different at different points in the field at a given time, i.e., the homogeneity assumption can be broken without affecting the linearity or Gaussian assumptions and, therefore, without affecting the accuracy of the estimator. This class of model is also suitable for modeling processes observed at locations which are distributed irregularly across space.

The price paid for this flexibility is the potential loss of parsimony and an increase in computational complexity over standard techniques. In applications with a high number of observation locations a different mapping between the observed and hidden fields is required. However, for applications which use a low number of observation locations but which are detailed in time, the neighborhood based spatial time-series model provides a conceptually clear and easily implementable framework. The increase in complexity introduced by the neighborhood mapping is mitigated by the accuracy of the parameter estimates.

A number of extensions to the presented framework can be considered. A fully heterogeneous system could be modeled by considering nonstationary parameters, which would incorporate the standard Kalman smoother. Attention could also be paid to the parsimony issue; techniques such as making simplifying assumptions on the space to be modeled could allow for a greater number of observation locations to be dealt with before having to make the compromise for a more complex model. Finally, as in the coupled map lattice literature [25], the need for a systematic neighborhood detection scheme is clear, meaning that the neighborhood need not be treated as prior information.

ACKNOWLEDGMENT

The authors would like to acknowledge the anonymous reviewers for their valuable suggestions.

REFERENCES

[1] J. von Neumann, *Theory of Self-Reproducing Automata*. Chicago: Univ. Illinois Press, 1966.
[2] K. Kaneko, Ed., *Theory and Applications of Coupled Map Lattices*. New York: Wiley, 1993.
[3] D. Coca and S. A. Billings, "Analysis and reconstruction of stochastic coupled map lattice models," *Phys. Lett. A*, vol. 315, pp. 61–75, 2003.
[4] P. C. Kyriakidis and A. G. Journel, "Geostatistical space-time models: A review," *Math. Geol.*, vol. 31, no. 6, pp. 651–684, 1999.
[5] 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.
[6] P. E. Pfeifer and S. J. Deutsch, "A three-stage iterative procedure for space-time modeling," *Technometrics*, vol. 22, no. 1, pp. 35–47, 1980.
[7] D. S. Stoffer, "Estimation and identification of space-time ARMAX models in the presence of missing data," *J. Amer. Statist. Assoc.*, vol. 81, no. 395, pp. 762–772, 1986.
[8] B. K. Epperson, "Spatial and space-time correlations in ecological models," *Ecolog. Model.*, vol. 132, pp. 63–76, 2000.
[9] N. Cressie, *Statistics for Spatial Data*. New York: Wiley, 1993.
[10] A. Jenkinson, S. Kadirkamanathan, S. Scott, E. Yazaki, and D. Evans, "Postoperative factors affecting outcome following laparoscopic fundoplication for control of acid reflux," *Br. J. Surg.*, vol. 90, no. 3, pp. 372–373, 2003.
[11] M. A. L. Nicolelis and S. Ribeiroa, "Multielectrode recordings: The next steps," *Curr. Opin. Neurobiol.*, vol. 12, no. 5, pp. 602–606, Oct. 2002.
[12] R. A. Garrido, "Spatial interaction between the truck flows through the Mexico–Texas border," *Transport. Res. Part A*, vol. 34, pp. 23–33, 2000.
[13] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *J. Time Series Anal.*, vol. 3, no. 4, pp. 253–264, 1982.
[14] G. A. Smith and A. J. Robinson, A comparison between the EM and subspace identification algorithms for time-invariant linear dynamical systems Cambridge Univ., Eng. Dept., 2000, Tech. Rep..
[15] B. Ninness and S. Gibson, On the relationship between state-space-subspace-based and maximum-likelihood system identification methods Univ. Newcastle, Australia, 2000, Tech. Rep..
[16] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA J.*, vol. 3, no. 8, pp. 1445–1450, 1965.
[17] T. Gneiting, M. Genton, and P. Guttorp, Geostatistical space-time models, stationarity, separability and full symmetry Univ. Wash., Tech. Rep. 475, 2005.
[18] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1999.
[19] S. Gibson and B. Ninness, "Robust maximum-likelihood estimation of multivariable dynamic systems," *Automatica*, vol. 41, pp. 1667–1682, 2005.
[20] W. Steeb, *Matrix Calculus and Kronecker Product With Application and C++ Programs*. Singapore: World Scientific, 1997.
[21] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications*. New York: Springer, 2000.
[22] K. Xu and C. Wikle, "Estimation of parameterized spatio-temporal dynamic models," *J. Stat. Planning Inference*, vol. 137, no. 2, pp. 567–588, 2007.
[23] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York: Wiley, 1997.
[24] C. F. Wu, "On the convergence properties of the em algorithm," *The Ann. Stat.* , vol. 11, no. 1, pp. 95–103, 1983.
[25] S. A. Billings and Y. Yang, "Identification of the neighborhood and CA rules from spatio-temporal CA patterns," *IEEE Trans. Syst. Man Cybern. B*, vol. 3, no. 2, 2003.

Michael Dewar received the M.Eng. degree in control systems engineering and the Ph.D. degree in systems engineering both from The University of Sheffield, U.K., in 2002 and 2007, respectively.

He is currently working as a Research Associate in the Department of Automatic Control and Systems Engineering, The University of Sheffield. In 2004, he was with the Department of Electrical Power and Control Engineering, University of Malta. His interests surround the modeling of dynamic spatiotemporal systems, including parameter and state estimation for constrained state space models, modeling from irregularly sampled data sets, and the modeling of heterogenous spatiotemporal fields.

Visakan Kadirkamanathan (M'90) received the B.A. and Ph.D. degrees in electrical and information engineering from the University of Cambridge, U.K.

He held research associate positions with the University of Surrey, U.K. and the University of Cambridge before joining the Department of Automatic Control and Systems Engineering, The University of Sheffield, U.K., as a lecturer in 1993. He is currently Professor and is with the Signal Processing and Complex Systems Research Group. His research interests include nonlinear signal processing, particle filters, intelligent control, optimization and decision support, swarm intelligence and applications in systems biology, aerospace fault diagnosis, and wireless communication signal detection and *ad hoc* networking. He has coauthored a book on intelligent control and has published more than 110 papers in refereed journals and proceedings of international conferences.

Dr Kadirkamanathan is the coeditor of the *International Journal of Systems Science* and has served as an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS and the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, PART B.