Article:
Mercieca, J. and Kadirkamanathan, V. orcid.org/0000-0002-4243-2501 (2016) Estimation and Identification of Spatio-Temporal Models with Applications in Engineering, Healthcare and Social Science. Annual Reviews in Control, 42. pp. 285-298. ISSN 1367-5788

https://doi.org/10.1016/j.arcontrol.2016.09.011

Estimation and Identification of Spatio-Temporal Models with Applications in Engineering, Healthcare and Social Science

Julian Mercieca[a], Visakan Kadirkamanathan[a,*]

[a]*Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, S1 3JD, UK*

Abstract

Several natural phenomena are known to exhibit a spatio-temporal evolution process. The study of such processes, which is pivotal to our understanding of how best to predict and control spatio-temporal systems, has motivated researchers to develop appropriate tools that infer models and their parameters from observed data. This paper reviews this active area of research by providing an insight into the fundamental ideas spanning the development of spatio-temporal models, dimensionality reduction methods and techniques for state and parameter estimation. Recent advances are discussed in the context of novel spatio-temporal approaches proposed for applications in three specific domains – engineering, healthcare and social science. They illustrate the wide applicability of estimation and identification of spatio-temporal processes as novel advances in sensor systems and data collection are used to observe them.

1. Introduction

Spatio-temporal systems are systems with variables whose evolution spans both space and time (Hoyle, 2006). Consequently, they are ubiquitous in several science and engineering disciplines including environmental science (Pirani et al., 2014; Moradkhani et al., 2005; Hooten and Wikle, 2008; Bocquet et al., 2010), bacterial and viral infection spread (Bhatt et al., 2013; Brooks-Pollock et al., 2014), biology (de Munck et al., 2002; Matani et al., 2003; Dewar and Kadirkamanathan, 2007), neuroscience (Aram et al., 2013), conflict dynamics (Zammit-Mangion et al., 2012a), mobile sensor networks (Gu and Hu, 2012) and video image processing (Kokaram and Godsill, 2002). This widespread interest has fuelled several research efforts over recent decades in an attempt to obtain mathematical models that best describe the behaviour of spatio-temporal

*Corresponding author

Email addresses: julian.mercieca@sheffield.ac.uk (Julian Mercieca), visakan@sheffield.ac.uk (Visakan Kadirkamanathan)

phenomena, where space and time data should not be treated as statistically independent variables (Cressie and Wikle, 2011). Such models form the fundamental building blocks for system simulation, design and analysis.

Whilst several spatio-temporal models are apparent in literature, these may be mainly grouped into two classes: geostatistical models (Gneiting et al., 2007) and dynamic models (Cressie and Wikle, 2011, Chapters 7-9). The former approach to modelling data makes use of a statistical description, typically in terms of mean and covariance functions (Guttorp and Sampson, 1994; Gneiting et al., 2007), such as an underlying Gaussian random function. On the other hand, dynamic models typically comprise difference or differential equations that would explicitly describe the temporal or spatio-temporal evolution. Dynamic models therefore usually allow for a mechanistic approach to systems, where parameters inferred typically constitute a physical meaning or a direct relation to the system behaviour (Duan et al., 2009) such as partial differential equations. The two modelling strategies may occasionally be interchangeable descriptions of the same process (Lindgren et al., 2011). Storvik et al. (2002) highlight numerous advantages of dynamic models over geostatistical models, including the more computationally efficient parameter estimation process when using signal processing tools with dynamic models and the employed covariance functions that could correspond to models that represent unnatural features. Furthermore, inference mechanisms associated with dynamic models are capable of readily handling missing or incomplete data. Due to their amenability to control and engineering applications (Zammit-Mangion et al., 2011), dynamic models shall be the main focus of this paper.

In several situations, most spatio-temporal processes can only be partially observed and an estimation problem naturally arises. The estimation of internal states is crucial for control, monitoring and fault diagnosis of several engineering processes. A cost-effective approach to monitor such variables employs model-based state estimation methods to estimate unmeasured and/or infrequently measured variables quickly and regularly. With ever-increasing computation speeds, state estimation is increasingly being performed for on-line monitoring and control in various application domains such as robotics, digital communications, computer vision and process control (Chen, 2003; Soroush, 1998). Historical developments in this state estimation are excellently introduced by Sorenson (1970). The seminal publications by Kalman (1960) and Luenberger (1966) spurred great research efforts in the area of dynamic model online state estimation. While such initial developments used only linear dynamic models, their nonlinear counterpart was the main topic of research in later years. It is noteworthy that despite significant advances employing dynamic models considering continuous observations (Stroud et al., 2001; Dewar et al., 2009; Freestone et al., 2011), very few efforts have considered the problem of having observations available as isolated events, i.e. point-process observations (Zammit-Mangion et al., 2012a).

State estimators may be developed using deterministic (Dochain, 2003; Misawa and Hedrick, 1989) or stochastic (Bayesian) approaches, with the latter approach being our principal focus throughout this article. A detailed exposure

to nonlinear Bayesian state estimation is provided in various books (Jazwinski, 1970; Gelb, 1974; Maybeck, 1979; Sorenson, 1985; Söderström, 2003), while more recent advances are described in (Ristic et al., 2004; Simon, 2006; Patwardhan et al., 2012; Särkkä, 2013). Recent works also include the emergence of deterministic and random sampling-based estimation approaches as part of the unconstrained sequential estimation algorithm development efforts. Overviews of considerable achievements in sigma point and particle filters are given in (Arulampalam et al., 2002; Chen, 2003; Ching et al., 2006; Daum, 2005; Rawlings and Bakshi, 2006).

All state estimation algorithms assume the accurate knowledge of the corresponding model parameters. However, if in addition to the states, a number of unknown parameters must be estimated, a joint state-parameter estimation algorithm is needed (Soroush, 1998). Inferring and constructing system models from experimental data is known as system identification and is paramount for the emulation of the system, prediction of the system response for particular inputs and investigation of different design situations (Billings, 2013). Consequently, the accuracy of such system representation would affect the validity of all the analysis, design and simulation and has therefore formed the basis of numerous efforts to develop and improve joint state-parameter estimation methods, namely the Markov Chain Monte Carlo (MCMC) methods (Robert and Casella, 2004), expectation-maximization (EM) algorithms (McLachlan and Krishnan, 1997) and their variants.

The objective of this article is to review methods that dominate the spatio-temporal estimation and identification literature with the scope of providing a window over current and future avenues of research that lie within important application areas. This paper is organised in seven sections. Section 2 presents a review of existing spatio-temporal models and accompanying theoretical properties from the literature that allow for its use in practical applications. Model reduction methods are discussed to show how the spatio-temporal field may be adequately represented by a finite-dimensional model which, in turn, allows for a state-space representation for which a number of estimation schemes may be readily applied. State estimation and joint state and parameter estimation methods are discussed in Section 3. Recent advances in the application of spatio-temporal estimation and identification methods for engineering, health and social science are presented in Sections 4, 5 and 6, respectively. Finally, Section 7 gives final remarks and outlines potential future research directions.

2. Spatio-temporal Models

Several dynamic spatio-temporal models have been proposed, however we shall only be reviewing the most common models appearing in the literature, namely, the space-time auto-regressive moving-average (STARMA) model, the coupled map lattice (CML), the integro-difference equation (IDE), the partial differential equation (PDE) and the spatio-temporal descriptor system formulation. Along the way, the strengths and weaknesses of each modelling schemes are highlighted and compared. These developments will be employed in Section

3 in conjunction with dimensionality reduction schemes typically used for the aforementioned spatio-temporal models, a brief review of which is provided in this section.

2.1. Space-Time Auto-Regressive Moving-Average (STARMA) Models

Following the successful introduction of the auto-regressive moving-average (ARMA) class of models for stochastic temporal processes (Box and Jenkins, 1970), extensions to the ARMA models were proposed in the 1970s to consider spatial dynamics in the time series evolution, leading to the development of space-time ARMA (STARMA) models (R. L. Martin, 1975; Pfeifer and Deutrch, 1980).

A STARMA model is described by a linear relationship lagged in both space and time. To obtain a STARMA formulation, a number of observations $y_{i,k}$ of the random variable $Y_{i,k}$ are required at each of the N sites (or fixed locations) located in the spatial field, over K time instants of the discrete time k. The spatio-temporal auto-regressive form expresses $y_{i,k}$ as a linear combination of past observations at site i (Pfeifer and Deutrch, 1980) and neighbouring sites. Spatial stationarity or homogeneity would exist if an identical relationship holds for every site.

In the classical STARMA formulation by Pfeifer and Deutrch (1980), a spatial lag operator of order l is first defined such that

$$L^{(0)} y_{i,k} = y_{i,k}, \tag{1}$$

$$L^{(l)} y_{i,k} = \sum_{j=1}^{N} w_{ij}^{(l)} y_{i,k}, \tag{2}$$

where $w_{ij}^{(l)}$ are a set of weights such that

$$\sum_{j=1}^{N} w_{ij}^{(l)} = 1 \tag{3}$$

for all i and $w_{ij}^{(l)} \neq 0$ if sites i and j are l^{th} order neighbours. STARMA model are generally represented in vector form, where the observations are given by the vector $\mathbf{y}_k = [y_{1,k}\ y_{2,k}\ \cdots\ y_{N,k}]^\top$. The superscript $^\top$ denotes the transpose operator. Representing the weights $w_{ij}^{(l)}$ in matrix form as $\mathbf{W}^{(l)} \in \mathbb{R}^{N \times N}$, the spatial lag operator for stacked observations may be written as

$$L^{(0)} \mathbf{y}_k = \mathbf{W}^{(0)} \mathbf{y}_k = \mathbf{I}_N \mathbf{y}_k, \tag{4}$$

$$L^{(l)} \mathbf{y}_k = \mathbf{W}^{(l)} \mathbf{y}_k, \text{ for } l > 0, \tag{5}$$

where $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. The STARMA model may now

be expressed in vector form as

$$\mathbf{y}_k = \sum_{\tau=1}^{p} \sum_{l=0}^{\lambda_\tau} \phi_{\tau l} \mathbf{W}^{(l)} \mathbf{y}_{k-\tau} - \sum_{\tau=1}^{q} \sum_{l=0}^{m_\tau} \phi_{\tau l} \mathbf{W}^{(l)} \epsilon_{k-\tau} + \epsilon_k, \tag{6}$$

where ϵ_k is a random normal error vector. This formulation is known as a STARMA (p,q) model.

Frequently used neighbourhood definitions were described by Besag (1974), typically assuming a local interaction hypothesis which results in fewer parameters to estimate. Although unrestricted models were proposed (Bennett, 1979), such spatio-temporal models required a significantly larger parameter space, forcing smaller spatial dimensions and fewer observation locations to be employed (Di Giacinto, 2006). Even though STARMA models were shown to outperform univariate ARMA models in forecasting applications (Pfeifer and Bodily, 1990), the researchers' interest in this class of models grew weaker over time, largely due to their inadequate treatment of spatial dependence and heterogeneity of observations (Anselin, 1988; Cressie, 1993).

Over the last decade, however, the potential of STARMA models was revisited as a consequence of increased computational power. Di Giacinto (2006) addressed the issue of instantaneous spatial correlation by starting the first summation in both terms of equation (6) from $\tau = 0$ so that innovations may represent a spatial spread within one sampling instant. Furthermore, heterogeneous model definitions became possible with model modifications that allowed the use of larger data sets, such as the toroidal space definition proposed by Glasbey and Allcroft (2008). The STARMA models' limitations, including that of having model dimension depending on the number of observation locations, however, still stand and alternative models were proposed in the 1980s where the coupled map lattices became particularly popular for their ability of representing spatio-temporal dynamics of systems that are too complex or not well understood.

2.2. Coupled Map Lattices

The complex and chaotic spatio-temporal behaviour exhibited by various natural phenomena (Kaneko, 1992, 1993) was the main motivation for the development of Coupled Map Lattices (CML) (Kaneko, 1985, 1986, 1989). The wide applicability of CMLs is evidenced by the plethora of uses reported in the literature and has been the modelling paradigm for studying chemical and physical processes, including modelling the physics of boiling (Yanagita, 1992), describing cloud dynamics (Yanagita and Kaneko, 1997), modelling reaction-diffusion dynamics (Levine and Reynolds, 1992), studying Bénard convection (Yanagita and Kaneko, 1993, 1995), modelling open fluid flow (Deissler, 1987; Deissler and Kaneko, 1987; Kaneko, 1985) and modelling crystal growth (Kessler et al., 1990; Levine and Reynolds, 1992). The use of CMLs for the analysis of complex spatio-temporal interactions occurring in ecology is reported in several works (Ikegami and Kaneko, 1992; Marcos-Nikolaus et al., 2002; Sole et al.,

1992). Other application areas include computer theory (Holden et al., 1992), image processing (Price et al., 1992) and electroencephalography (EEG) signal processing (Shen et al., 2006, 2008).

A subset of the more general class of lattice dynamic systems (Billings and Coca, 2002), CMLs are closely related to cellular automata (CA), with the relaxed constraint that the system states may not necessarily be discrete (Pan and Billings, 2008). CMLs are defined to be in discrete time and discrete space. Denoting a set of lattice points by $i = 1, \ldots, N$, where each element identifies a discrete location in space, and letting the field be $z_{i,k}$ at discrete-time instant k, the temporal evolution at site i is given by the nonlinear mapping $M_i : \mathbb{R}^N \to \mathbb{R}$, so that $z_{i,k+1} = M_i \mathbf{z}_k$, where $\mathbf{z}_k = [z_{1,k}\ z_{2,k}\ \ldots\ z_{N,k}]^\top$. Although a few works consider spatial heterogeneity (Parekh et al., 1998), a spatially homogeneous process is usually assumed so the standard nonlinear evolution equation becomes $z_{i,k+1} = M\mathbf{z}_k$, where the mapping dependence on i has been omitted.

The behaviour of the CML is clearly governed by the mapping M. The most commonly used mapping is the nearest neighbour coupling map (Kaneko, 1992; Bunimovich, 1995; Kaneko, 1989; Richter, 2008) that comprises a spatial coupling function f_c and a local interaction term f_l, as follows:

$$\begin{aligned} z_{i,k+1} &= f_c(z_{i-1,k}, z_{i+1,k}) + f_l(z_{i,k}) \\ &= \frac{\epsilon}{2}\left(f(z_{i-1,k}) + f(z_{i+1,k})\right) + (1-\epsilon) f(z_{i,k}), \end{aligned} \tag{7}$$

where $\epsilon \in [0, 1]$ and $f(\cdot)$ represents a pre-defined nonlinear function, for instance the logistic map $f : z_{i,k} \to 1 - a z_{i,k}^2$. This logistic map (May, 1976) is the most commonly used local map (Kaneko, 1989), however several other mappings can represent chaotic behaviour (Jost and Joy, 2001). Other mappings which consider larger neighbourhoods result in significantly different output patterns and are referred to as 'global coupling' (Jost and Joy, 2001).

A CML is generally derived through the natural laws obeyed by the system under study. However, if the mapping M is undetermined or cannot be derived, model structure detection and parameter estimation may be carried out (Coca and Billings, 2001; Pan and Billings, 2008; Billings et al., 2006; Coca and Billings, 2003; Guo et al., 2007). Although most CMLs described in the literature are deterministic, a stochastic CML was reported to use randomly perturbed lattice points (Coca and Billings, 2003).

Despite being dynamic, capable of representing systems exhibiting large uncertainties and highly representative of the system's underlying processes, CMLs are built bottom-up on a discrete grid. This means that observations must be taken on a regular lattice, which may be impossible in certain situations, such as control scenarios involving mobile agents. Although heterogeneous CMLs may provide a spatially varying mapping, proceeding with parametrising the heterogeneity and choosing the appropriate inference mechanism to cater for the heterogeneity in parameter estimation is unclear.

2.3. Integro-difference Equation Models

The discrete spatial lattice construction is the main weakness of the CML and the integro-difference equation (IDE) (Wikle, 2002; Kot et al., 1996) remedies the situation by employing a continuous-space representation. The deterministic IDE was first proposed by Kot and Schaffer (1986); Kot et al. (1996) to model the spread of invading organisms. The IDE was formulated by modelling a population in two separate stages. The first stage is referred to as the sedentary stage and is represented by a nonlinear map $f(\cdot)$ that determines local growth. The second stage, known as the dispersion stage, is described through an integral operator which represents physical diffusion or migration effects in a population. Based on such applications, the IDE was shown to model these systems better than the reaction-diffusion equation proposed by Fisher (1937). Since then, IDEs have been employed to model different phenomena such as cloud dynamics (Christopher K. Wikle and Berliner, 2001; Wikle, 2002) and precipitation nowcasting (Xu et al., 2005).

The IDE, which is continuous in space and discrete in time, represents the evolution of the spatio-temporal field z given by

$$z_k(\mathbf{s}) = \int_{\mathcal{S}} \kappa_k(\mathbf{s}, \mathbf{r}) f(z_{k-1}(\mathbf{r})) d\mathbf{r}, \tag{8}$$

where $k \in \mathbb{Z}^+$ is discrete time, $\mathbf{s} \in \mathcal{S} \subset \mathbb{R}^n$ represents the spatial location in an n-dimensional space and $\kappa_k(\mathbf{s}, \mathbf{r}) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a time-varying, heterogeneous spatial convolution kernel that controls the spatio-temporal interactions of the system.

Equation (8) represents a heterogeneous IDE model with nonlinear growth. In environmental literature, however, this is usually simplified and linear growth and homogeneity is assumed, yielding the following model:

$$z_k(\mathbf{s}) = f'(0) \int_{\mathcal{S}} \kappa_k(\mathbf{s} - \mathbf{r}) z_{k-1}(\mathbf{r}) d\mathbf{r}, \tag{9}$$

where f' represents the first derivative of f. Such simplification was desirable for the development of spatio-temporal methods such as spatio-temporal Kalman filtering (Cressie and Wikle, 2002; Wikle and Cressie, 1999) and new classes of non-separable covariance functions for geostatistical models (Brown et al., 2000, 2001; Storvik et al., 2002). In time, IDEs were recognized for the ability of representing complex spatio-temporal behaviour spanning several fields such as ecology (Kot et al., 1996), signal processing (Dewar et al., 2009; Scerri et al., 2009) and environmental applications (Cressie and Wikle, 2002; Xu et al., 2005; Wikle, 2002; Wikle and Cressie, 1999; Wikle and Hooten, 2005; Christopher K. Wikle and Berliner, 2001).

Most of the research, particularly in ecological literature, has focused on analysing the effect of the shape and growth term of the convolution kernel on the spatio-temporal process stability (Kot and Schaffer, 1986) and the shape and speed of the invading waves generated (Kot et al., 1996; Kot, 1992; Neubert et al., 1995; Billings et al., 2004; Veit and Lewis, 1996; Wang and Kot, 2001).

More recent efforts in modelling population dynamics using IDE models have been aimed at improving the basic representations given by equations (8) and (9) by employing the Allee effect (Kot et al., 1996; Veit and Lewis, 1996; Wang and Kot, 2001) and by analysing the effect of environmental variables and population structure on propagation (Neubert and Caswell, 2000; Billings et al., 2004). Other works report the estimation of the travelling wave shape (Medlock and Kot, 2003), the numerical estimation of the invading wave speed (Wang and Kot, 2001) and the prediction of the future invasion speed (Billings et al., 2004).

The IDE was put into a stochastic framework by Wikle (2002) who includes additive spatial noise using spatial Gaussian processes (GP) (Rasmussen and Williams, 2006). At each time instant of this stochastic IDE, the propagated field is superimposed by draws from a zero-mean spatial GP, $\epsilon_k(\mathbf{s}) \sim \mathcal{GP}(0, \Sigma(\mathbf{s}, \mathbf{r}))$. This stochasticity enables the modelling of uncertainties and caters for any random forcing functions or model mismatch. The set generated $\epsilon_k(\mathbf{s})$ is typically assumed to be independently and identically distributed (i.i.d.) over time, so that the behaviour of the model is largely dictated by the mixing kernel and the form of $f(\cdot)$. For instance, in EEG studies, $f(\cdot)$ is set to be a sigmoid function (Freestone et al., 2011), whilst in ecology, the Ricker growth models or the standard logistic models are frequently used (Kot and Schaffer, 1986). The function $f(\cdot)$, however, may also be taken to be of Gompertz, Beverton-Holt or Malthusian form (Hooten et al., 2007).

Dewar et al. (2009) derived a novel basis function decomposition for the IDE, where a state-space representation that decouples the number of states from the number of observation locations or parameters is presented. By using a state-space representation for the IDE, Scerri et al. (2009) employ ideas from multidimensional sampling theory to develop a method that provides the minimum model and parameter vector dimensions needed for an adequate system representation using the spatial bandwidth of the system and the frequency support of the redistribution kernel of the IDE.

When modelling systems using the IDE, the kernel provides an intuitive insight into the system dynamics. Several works have estimated basis functions which shape $\kappa_k(\mathbf{s}, \mathbf{r})$ (Scerri, 2010; Freestone et al., 2011; Zammit-Mangion et al., 2011; Dewar et al., 2009). However, as discussed in (Zammit-Mangion, 2011), a key limitation of the IDE is its inability to describe the evolution process at a physical level. The IDE may obscure the physical mechanism and consequently also presents significant challenges in describing heterogeneity. A more mechanistic approach to modelling is therefore required if a principled way of representing spatially varying systems is required. This leads to the consideration of another class of models, the partial differential equations (PDEs).

2.4. Partial Differential Equation Models

PDE models, which are continuous in both space and time, enjoy a widespread interest due to their extensive range of natural phenomena that they describe. These include fluid dynamics, mechanics, elasticity, quantum physics, thermodynamics and electromagnetic theory (Farlow, 1993; Helling, 1960; Mitchell and

Wait, 1973; Smith, 1967). Applications are as diverse as oceanography (Bennett, 2002), ecology (Holmes et al., 1994), wildfire control (Asensio and Ferragut, 2002) and flexible structures (Banks and Kunisch, 1989).

The formal definition of a PDE is any equation that involves an unknown function of two or more independent variables and one or more of its partial derivatives (Evans, 1998, Section 1.1). The independent variables for the case of spatio-temporal systems are restricted to be space and time. Letting space $s \in \mathcal{S} \subset \mathbb{R}$ and time $t \in \mathcal{T} \subset \mathbb{R}^+$ and considering a single-dimension spatio-temporal field $z(s,t) : \mathcal{S} \times \mathcal{T} \to \mathbb{R}$, we have the following general form of the PDE:

$$F\left(s, t, z, \frac{\partial z}{\partial s}, \frac{\partial z}{\partial t}, \frac{\partial^2 z}{\partial s^2}, \frac{\partial^2 z}{\partial t^2}, \frac{\partial^2 z}{\partial s \partial t}, \ldots\right) = 0. \tag{10}$$

The PDE is said to be linear if $F(\cdot)$ is a linear function, otherwise it is said to be nonlinear or quasilinear. Furthermore, the system is said to be space and time invariant if $F(\cdot)$ is independent of s and t. The study of linear PDEs has been quite extensive given the breadth of applicability to several areas of mathematical physics including vibrations, heat flow (Hill, 1987; Helling, 1960; Smith, 1967) and so on. However, several other phenomena modelled using nonlinear PDEs include fluid pressure effects solved using Navier-Stokes equations, superconductivity based on the Ginzburg-Landau equation and general relativity described by Einstein's field equations and the Dym equation (Debnath, 2005; Logan, 2008).

Despite the fact that PDEs represent spatio-temporal dynamics of physical phenomena, for which experiments prove the existence of a stable unique solution, their mathematical representation might not yield such a solution. While for the ODE case the general solution to an nth-order equation is described by a family of functions with n independent arbitrary constants, this is certainly not the case for PDEs. In fact, even the solution space for linear homogeneous PDEs is infinite dimensional. In systems literature, such systems resulting in an infinite dimensional solution space are referred to as distributed parameter systems (Omatu and Seinfeld, 1989).

PDEs are generally defined on some bounded domain. In such situations, the PDE formulation must include some prescribed conditions for z that must be satisfied on the domain boundary $\partial\mathcal{S}$. The conditions are either Dirichlet (first-type), whereby z takes on fixed values on $\partial\mathcal{S}$, or Neumann (second-type), where z is required to have fixed derivatives on $\partial\mathcal{S}$. If both boundary conditions and initial conditions are specified, the problem of determining field z which satisfies the PDE is referred to as the initial/boundary-value problem.

Despite several methods that exist for finding an analytical solution to PDEs (Mitchell and Wait, 1973; Farlow, 1993), most practical physical systems cannot be solved analytically and therefore numerical methods are adopted. Two main techniques are described in the literature, namely the finite element (Mitchell and Wait, 1977) and the finite difference methods (Smith, 1969).

The solution of PDEs is further complicated when the model parameters,

such as the thermal conductivity of a material in a heat flow equation, are unknown. The system identification community has focused significant research efforts towards obtaining models of spatio-temporal systems directly from data obtained by measurement, oftenly assuming little or no knowledge of the underlying physical processes (Guo and Billings, 2006; Guo et al., 2009; Coca and Billings, 2000; Niedzwecki and Liagre, 2003). This problem was first tackled by Travis and White (1985) by developing tests for the identifiability of PDE model parameters. Assuming identifiability, an estimation method based on alternating conditional algorithms was later proposed by Voss et al. (1998) and shown to estimate the Swift-Hohenberg equation successfully.

More recent works progressively allowed more assumptions to be relaxed. The assumption of a known structural form of the PDE taken by Coca and Billings (2000) is relaxed by Guo and Billings (2006); Guo et al. (2009), where PDE estimation is performed using the orthogonal least squares algorithm and Adams integration.

Whenever the initial or boundary conditions are stochastic (Carmona, 1998, Section 1.1) or where the forcing term is random in nature (Dalang and Frangos, 1998) or when the physical system is not fully known, stochastic PDEs (SPDEs) become the required form of representation. This intricate model can be used to describe all kinds of dynamics having a stochastic influence in nature or man-made complex systems (Prévôt and Röckner, 2007). This is clearly evidenced by several works reporting the use of SPDEs for modelling purposes in a vast array of application areas including hydrology (Unny, 1989), neurophysiology (Walsh, 1981), geophysics (Duan and Goldys, 2001) and signal denoising (Krim and Bao, 1999). Even though their applicability is extensive, choosing SPDEs for analysis presents significant challenges in the context of parameter estimation. Here, most of the literature treats deterministic PDEs observed in noise (Coca and Billings, 2002; Guo and Billings, 2006; Banks and Kunisch, 1989), whilst for the stochastic case, fewer works have been published (Solo, 2002). New estimation and identification tools for SPDEs have been recently explored by Zammit-Mangion et al. (2012b), where the use of the variational approximation and the consideration of both continuous and point-process observations where investigated for SPDEs.

2.5. Partial Differential-Algebraic Equation Models

An even more general class of models to the PDEs are partial differential-algebraic equation (PDAE) models, which is a descriptor formulation (also known as an implicit form, singular form or generalised state-space form). The models of a number of natural processes, such as fluid flow (Stull, 1988) and electrochemical reactions in a molten carbonate fuel cell (Chudej et al., 2003), result in a PDAE system of the form

$$\frac{\partial \mathbf{z}_d(\mathbf{s},t)}{\partial t} = \mathbf{F}\left(D^m\mathbf{z}(\mathbf{s},t), D^{m-1}\mathbf{z}(\mathbf{s},t), \ldots, D\mathbf{z}(\mathbf{s},t), \mathbf{z}(\mathbf{s},t)\right), \tag{11}$$

$$\mathbf{0} = \mathbf{G}\left(D^m\mathbf{z}(\mathbf{s},t), D^{m-1}\mathbf{z}(\mathbf{s},t), \ldots, D\mathbf{z}(\mathbf{s},t), \mathbf{z}(\mathbf{s},t)\right), \tag{12}$$

where $\mathbf{z}(\mathbf{s},t) = [\mathbf{z}_d^\top(\mathbf{s},t)\ \mathbf{z}_a^\top(\mathbf{s},t)]^\top$ and $\mathbf{z}_d$ and $\mathbf{z}_a$ denote the differential and algebraic states, respectively. For a non-negative integer m, $D^m\mathbf{z}(\mathbf{s},t)$ is the set of all partial spatial derivatives of order m.

It is noteworthy to highlight the difference between a PDAE and a constrained PDE system. For constrained PDE systems, the evolution of all process states $z(\mathbf{s},t) = \mathbf{z}_d(\mathbf{s},t)$ is described by PDEs, subject to algebraic constraints that confine their evolution. For a PDAE system, there exist some states $\mathbf{z}_a(\mathbf{s},t)$, known as algebraic, whose evolution is not governed by PDEs, but is completely dictated by the evolution of differential states $\mathbf{z}_d(\mathbf{s},t)$, such that all algebraic constraints are satisfied (Patwardhan et al., 2012). Although numerous researchers convert the descriptor problem to a standard state-space formulation (e.g. (Towers and Jones, 2016; Soleimanzadeh et al., 2014)), this process may introduce significant numerical errors where ODE solvers are used (Mandela et al., 2010). The conversion may also lead to the violation of algebraic constraints and makes measurements that are functions of algebraic states redundant for state estimation (Mandela et al., 2010). This has motivated a number of researchers to retain a descriptor formulation for estimation purposes (Mercieca et al., 2015; Mandela et al., 2010; Becerra et al., 2001), as will be described in Section 3.

2.6. Model Reduction and State-Space Modelling

Since most standard signal processing techniques are generally tailored for finite-dimensional systems, model reduction methods were developed to reduce infinite-dimensional spatio-temporal models to a finite-dimensional form. A common spatial and temporal discretisation scheme is the method of finite differences typically employed for PDE-based models (Grossmann et al., 2007). This method approximates spatial and temporal derivatives of the PDE using difference quotients.

The method of moments is another model reduction technique usually employed for spatial dimensionality reduction (Hausenblas, 2003), where a finite set of linearly independent basis functions $\{\phi_i(\mathbf{s})\}_{i=1}^{n_\phi}$ are used to decompose a spatio-temporal field $z(s,t)$ such that

$$z(\mathbf{s},t) \approx \sum_{i=1}^{n_\phi} \phi_i(\mathbf{s}) x_i(t) = \boldsymbol{\phi}^\top(\mathbf{s})\mathbf{x}(t), \tag{13}$$

where $\mathbf{x}(t) \in \mathbb{R}^{n_\phi}$ is a state vector that weights the field basis functions $\phi_i(\mathbf{s})$. The spatio-temporal field is then projected under an inner-product transformation with respect to a set of test functions $\{\chi_i(\mathbf{s})\}_{i=1}^{n_\phi}$ (Hausenblas, 2003). A popular choice for the test functions sets $\{\phi_i(\mathbf{s})\}_{i=1}^{n_\phi} = \{\chi_i(\mathbf{s})\}_{i=1}^{n_\phi}$, which is a special case of the method of moments known as the Galerkin method. The method of moments has a number of advantages over standard finite-difference schemes, particularly due to their easier use in complex geometry spaces and their ability to handle Dirichlet boundary conditions systematically by an appropriate choice of basis functions (Zammit-Mangion et al., 2012b). Since the

observation process is usually temporally discrete, an Euler step is used to obtain a discrete-time representation for the finite-dimensional system (Freestone et al., 2011).

By following finite-dimensional reduction, the popular stochastic state-space model framework is obtained for which several signal processing techniques are readily available and algorithm development is greatly facilitated:

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k) + \mathbf{q}_k, \tag{14}$$

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{r}_k, \tag{15}$$

where $\mathbf{x}_k := \mathbf{x}(k\Delta_t)$ and $\mathbf{y}_k := \mathbf{y}(k\Delta_t) \in \mathbb{R}^{n_y}$ are the system state and observation vectors, respectively, $\mathbf{q}_k$ and $\mathbf{r}_k$ are noise sequences, $\mathbf{f}_k(\cdot)$ is a dynamic model function, $\mathbf{h}_k(\cdot)$ is a measurement model function and Δ_t is the time step.

3. Estimation and Identification

One critical problem in spatio-temporal systems is reconstructing the field in some spatial domain at any given time instant from some observation process. If the data is sufficiently informative, as is the case with data obtained from an infrared camera for example (Demetriou et al., 2003), then the spatio-temporal field may be assumed to be entirely known with no need of any further signal processing. If, however, the field is measured at isolated points, such as in neural field (Freestone et al., 2011) or ocean (Leonard et al., 2007) sampling, state estimation for $\mathbf{X} = \mathbf{x}_{0:K} = \{\mathbf{x}_0, \ldots, \mathbf{x}_K\}$ (for K regularly spaced time intervals) is performed using observed data $\mathbf{Y} = \mathbf{y}_{1:K} = \{\mathbf{y}_1, \ldots, \mathbf{y}_K\}$ required for field reconstruction. The optimal estimation of the states from some data set is known as the smoothing problem. This problem is generally solved using either the forward-backward algorithm, where the forward pass represents filtering and the backward pass represents smoothing, or the two-filter smoother that combines forward messages (identical to those obtained using filtering) with backward messages computed in reverse time to get smoothed estimates. Here, we shall only be reviewing filtering strategies. If, in addition to the states, the model parameters need to be estimated, the problem is known as a joint state-parameter estimation problem.

3.1. State Estimation

The models that shall be considered here are discrete-time state space models of the form given by equations (14) and (15). In the context of conditional-density-approximation-based state estimators, when the model is linear, the analytical solution is the Kalman filter that describes the optimal recursive solution to the problem of sequential state estimation (Kalman, 1960). Combining the predicted state estimate $\hat{\mathbf{x}}_{k+1|k}$ with the measurement $\mathbf{y}_{k+1}$, we obtain the optimal state estimate which is constructed recursively as follows:

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{L}_{k+1}\mathbf{e}_{k+1}, \tag{16}$$

where $\mathbf{e}_{k+1} = \mathbf{y}_{k+1} - \mathbf{C}_{k+1}\hat{\mathbf{x}}_{k+1|k}$ is the innovation and $\mathbf{L}_{k+1}$ is the Kalman gain matrix given by

$$\mathbf{L}_{k+1} = \mathbf{P}_{k+1}^{(\epsilon,e)}[\mathbf{P}_{k+1}^{(e,e)}]^{-1}, \tag{17}$$

where

$$\mathbf{P}_{k+1}^{(\epsilon,e)} = \mathrm{E}[(\boldsymbol{\epsilon}_{k+1|k})(\mathbf{e}_{k+1})^T] \tag{18}$$

and

$$\mathbf{P}_{k+1}^{(e,e)} = \mathrm{E}[(\mathbf{e}_{k+1})(\mathbf{e}_{k+1})^T]. \tag{19}$$

The a priori estimation error is denoted by $\boldsymbol{\epsilon}_{k+1|k} = \mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k}$. A fundamental feature of the Kalman filter is that whenever $\mathbf{q}_k$ and $\mathbf{r}_k$ are additive Gaussian noise processes and $\mathbf{x}_0$ is Gaussian distributed, then the conditional densities $p[\mathbf{x}_{k+1}|\mathbf{Y}^k]$ and $p[\mathbf{x}_{k+1}|\mathbf{Y}^{k+1}]$, and the innovation sequence $\mathbf{e}_{k+1}$, are also Gaussian. This is a result of the preservation of the Gaussian distributions under linear transformations.

When dealing with nonlinear systems, the sequential Bayesian estimation problem requires the development of approximate and computationally tractable sub-optimal solutions. This time, when $\mathbf{q}_k$, $\mathbf{r}_k$ and $\mathbf{x}_0$ are Gaussian, the conditional densities $p[\mathbf{x}_{k+1}|\mathbf{Y}^k]$ and $p[\mathbf{x}_{k+1}|\mathbf{Y}^{k+1}]$ are non-Gaussian. In this class of nonlinear stochastic observers, the most popular approach is the extended Kalman filter (EKF), which uses the Taylor series approximation of the nonlinear function vectors $\mathbf{F}(\cdot)$ and $\mathbf{h}(\cdot)$ to construct the conditional densities. The EKF has been a successful solution for many industrial problems (Muske and Edgar, 1997), however many nonlinear systems remain problematic. A serious limitation of the first-order EKF formulation is that the prediction step requires approximating the expected value of a nonlinear function of a random variable by the propagation of the mean of the random variable through the nonlinear function (Daum, 2005), as follows:

$$\mathrm{E}[\mathbf{F}(\mathbf{x}_k, \mathbf{u}_k)] \approx \mathbf{F}(\mathrm{E}[\mathbf{x}_k|\mathbf{Y}^k], \mathbf{u}_k). \tag{20}$$

Such approximation is certainly invalid in view of Jensen's inequality (Casella and Berger, 1990), which states that $\phi[\mathrm{E}(\mathbf{x})] \leq \mathrm{E}[\phi(\mathbf{x})]$. Although the second-order EKF attempts to mitigate the significant error involved by correcting estimates of the mean $\hat{\mathbf{x}}_{k+1|k}$ and $\hat{\mathbf{x}}_{k+1|k+1}$ using the second-order terms in the Taylor series expansion of $\mathbf{F}(\mathbf{x}_k, \mathbf{u}_k)$, severe nonlinearities remain a problem. The algorithm is also computationally expensive particularly for high-dimensional systems. Moreover, the Taylor series approximation requires smooth and at least twice differentiable nonlinear function vectors.

This requirement for smooth or differentiable nonlinear functions is relaxed when using approximations based on statistical linearization (Gelb, 1974), which is a better alternative for approximating the nonlinear function of a random variable. Furthermore, the state and measurement uncertainties are not required to be linearly additive signals. This class of nonlinear filters yields better estimates of the moments of a distribution using samples rather than the Taylor series approximation of the nonlinear function. This, however, requires the knowledge

of the probability density function of the random variable. A very popular approach in this class of filters (that are often referred to as sigma point filters) is the unscented Kalman filter (UKF) (Julier et al., 2000; Julier and Uhlmann, 2004) for which several application studies have appeared (Romanenko et al., 2004; Romanenko and Castro, 2004; Vachhani et al., 2006; Wan and Van der Merwe, 2000). In the unscented transformation, a number of deterministic samples are chosen such that their weighted mean and covariance would equate the mean and covariance of the random variable undergoing a nonlinear transformation. The transformed sample points (or sigma points) are then used to calculate the a posteriori mean and covariance. The difficulty associated with Jacobian computations in EKF are therefore alleviated with derivative-free sigma point filters. In most of the formulations, however, the conditional densities are still approximated as Gaussian. Due to nonlinear transformations, the conditional densities are in fact non-Gaussian, thereby requiring other methods to overcome these simplifying assumptions.

Particle filtering is a new class of filtering techniques that can deal with state estimation problems arising from non-Gaussian and multi-modal distributions (Arulampalam et al., 2002; Rawlings and Bakshi, 2006). A particle filter (PF) uses Monte Carlo sampling to approximate the multi-dimensional integration involved in the prediction and update steps or the moments of the conditional distributions. Belonging to the class of particle filters is the ensemble Kalman Filter (EnKF) (Burgers et al., 1998; Evensen, 2003) which is based on the idea of obtaining estimates for $\mathbf{P}_{k+1}^{(\epsilon,e)}$ and $\mathbf{P}_{k+1}^{(e,e)}$ using random samples rather than deterministic ones. Detailed expositions on algorithmic and theoretical aspects of particle filtering are included in (Chen, 2003; Arulampalam et al., 2002; Rawlings and Bakshi, 2006; Cappe et al., 2007).

In addition to conditional-density-approximation-based estimators, various nonlinear state estimation methods that use an optimisation approach to solve nonlinear state estimation problems have been proposed. These techniques were developed with the specific goal of handling constraints on states and parameters in estimation (Patwardhan et al., 2012). One estimator that uses an explicit optimisation-based approach for state and parameter estimation of nonlinear dynamic processes (described by ODEs) is the moving horizon estimator (MHE) (Robertson et al., 1996). At every time step, the MHE solves an optimisation problem, thereby easily handling constraints and bounds on state variables. The standard MHE formulation considers an arrival cost term that accounts for the accumulated state estimate uncertainties till the current window of interest. For general nonlinear constrained state estimation, the arrival cost cannot be calculated analytically whilst for constrained linear ODE systems, the arrival cost may be approximated using the one for the unconstrained problem (Qu and Hahn, 2009). For nonlinear systems, the arrival cost is usually determined by approximating a constrained nonlinear ODE system as an unconstrained linear time-varying system (Tenny and Rawlings, 2002).

3.2. State and Parameter Estimation

All state estimation algorithms assume the accurate knowledge of the corresponding model parameters. If in addition to the states, a number of unknown parameters $\boldsymbol{\theta}$ must be estimated, a joint state-parameter estimation algorithm is needed (Soroush, 1998). These methods are the subject of the following subsections.

3.2.1. The Expectation-Maximization Algorithm

The expectation-maximization (EM) algorithm is a maximum likelihood (ML) estimation algorithm first introduced by Dempster, Laird and Rubin (Dempster et al., 1977) to solve incomplete-data or latent-data problems. Its first applications in identification were linear discrete stochastic state-space systems (Shumway and Stoffer, 1982), whilst more recent work in a more general setting is reported in (Gibson and Ninness, 2005). A comprehensive review treating the algorithm and its extensions is provided in (McLachlan and Krishnan, 1997) and summaries of the method are given in several sources (Dellaert, 2002; Bishop, 2006). In brief, EM has a dual role; that of estimating the unknown parameters $\boldsymbol{\theta}$ and that of estimating the latent states $\mathbf{X}$ using an iterative algorithm. An important point to recall is that in the context of spatio-temporal systems represented by state-space models, the hidden states $\mathbf{X}$ relate to the spatial field. The parameters $\boldsymbol{\theta}$ govern the statistics and dynamics of the evolution and observation processes. It is noteworthy that in the state-space formulation, the E-step requires solving the smoothing problem. Denoting the observed data as $\mathbf{Y}$, the EM algorithm can be summarized in the following four steps:

1. An initial guess is chosen for the parameter vector, $\boldsymbol{\theta}^{(0)}$. i is set to 0.

2. E-step: given the parameter vector and the measurements, the states are estimated using the joint log-likelihood function of the states and the measurements (Q-function), as follows:

$$\mathcal{Q}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}) = \int \log[p(\mathbf{X}, \mathbf{Y}|\boldsymbol{\theta})] p(\mathbf{X}|\mathbf{Y}, \boldsymbol{\theta}^{(i)}) d\mathbf{X}. \tag{21}$$

3. M-step: $\mathcal{Q}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta})$ is maximized with respect to $\boldsymbol{\theta}$. The maximizing value is $\boldsymbol{\theta}^{(i+1)}$.

4. The steps 2 and 3 are repeated until the change in parameter vector is within specified tolerance limits.

Note that the Q-function is the lower bound on the marginal likelihood $\log[p(\mathbf{Y}|\boldsymbol{\theta})]$. The EM algorithm may be slightly modified to become a semi-Bayesian approach to parameter estimation, known as the Maximum-a-posteriori (MAP)-EM algorithm. This involves including a parameter prior distribution $p(\boldsymbol{\theta})$ in the M-step. The resulting algorithm is only considered semi-Bayesian since at each step the posterior distribution probability mass is focused at its mode, thereby still being treated as a point estimate. In many cases, this is unrepresentative of

the true distribution. This problem is overcome using the variational Bayesian expectation-maximization (VBEM) algorithm, where the distributional properties of the parameter estimates are maintained throughout the E-step.

3.2.2. The Variational Bayesian Expectation-Maximization Algorithm

The elegant framework of the variational Bayesian expectation-maximization (VBEM) algorithm carries out analytic computations of approximate posterior distributions over parameters and latent variables (Attias, 1999, 2000). The posterior distributions are calculated iteratively (termed Iterative VB in (Šmídl and Quinn, 2005)), as in the EM algorithm, with guaranteed convergence. This technique inherits the advantages of a Bayesian approach whilst being deterministic with no sampling required. The approximate posterior distribution is unique for a given set of data, likelihood and prior distribution, making the VB method much faster than Markov chain Monte Carlo (MCMC) techniques such as that discussed in Section 3.2.3. It has therefore been widely applied to a vast range of problems that include vision tracking (Vermaak et al., 2003), neuroimaging (Penny et al., 2003), blind source separation (Cemgil et al., 2007) and the modelling of the cell's regulatory network (Beal et al., 2005; Sanguinetti et al., 2006).

The VBEM method yields a convenient functional form for approximating the joint posterior distribution $p(\mathbf{X}, \boldsymbol{\theta}|\mathbf{Y})$. This is usually obtained using the conditionally independent distributions $\tilde{p}(\mathbf{X})$ and $\tilde{p}(\boldsymbol{\theta})$ (also referred to as variational posterior distributions) so that $p(\mathbf{X}, \boldsymbol{\theta}|\mathbf{Y}) \approx \tilde{p}(\mathbf{X})\tilde{p}(\boldsymbol{\theta})$. The algorithm operates by taking a parameter distribution $\tilde{p}(\boldsymbol{\theta})^{(i)}$ into consideration and then determining $\tilde{p}(\mathbf{X})^{(i+1)}$ such that the lower bound is maximised. This VBE-step is given by

$$\tilde{p}(\mathbf{X})^{(i+1)} \propto \exp(E_{\tilde{p}(\boldsymbol{\theta})^{(i)}}[\ln p(\mathbf{X}, \boldsymbol{\theta}, \mathbf{Y})]) \tag{22}$$

Next, $\tilde{p}(\mathbf{X})^{(i+1)}$ is fixed and $\tilde{p}(\boldsymbol{\theta})^{(i+1)}$ is calculated such that the lower bound is maximised. This VBM-step is given by

$$\tilde{p}(\boldsymbol{\theta})^{(i+1)} \propto \exp(E_{\tilde{p}(\mathbf{X})^{(i+1)}}[\ln p(\mathbf{X}, \boldsymbol{\theta}, \mathbf{Y})]) \tag{23}$$

As in the EM algorithm, convergence may be monitored by observing the change in the mean of the parameter posterior distributions through the iterations. The VBEM algorithm shares most of its framework with its conventional EM counterpart. In fact, the EM algorithm may be viewed as a special case of VBEM, referred to as functionally constrained VB (Šmídl and Quinn, 2008; Beal, 2003). A significant difference worth noting is that $\tilde{p}(\mathbf{X})^{(i+1)}$ is obtained using expectations of $\boldsymbol{\theta}$ rather than only its ML point estimate. Whenever the posterior mode differs from the posterior mean, the two techniques will differ considerably. The advantage of the VBEM algorithm is therefore that through averaging, the method does not give excessive consideration to the mode of the parameter posterior distribution. In classical control terminology, the EM algorithm takes advantage of a certainty equivalence property, while the VBEM technique is more cautious and incorporates knowledge of second

and higher order moments in state estimation (Milito et al., 1982; Fabri and Kadirkamanathan, 2001).

3.2.3. Gibbs Sampling

With guaranteed convergence to the target posterior distributions (Robert and Casella, 2010), their simple implementation and wide applicability to most models, MCMC methods have been the most popular class of distributional approximation methods. The advent of parallel computing and several novel technologies has rendered MCMC techniques applicable to large-scale inference problems (Suchard et al., 2010; Lee et al., 2010). MCMC techniques obtain the desired posterior distribution from the stationary distribution of a generated Markov chain (a detailed overview is given in (Robert and Casella, 2004, Section 7.1)). Two of the most popular methods are the Gibbs sampler (Gelfand et al., 1990) and the Metropolis-Hastings algorithm (Hastings, 1970). The former strategy is ideal whenever the functional form of the joint posterior distribution $p(\mathbf{X}, \boldsymbol{\theta}|\mathbf{Y})$ is unknown, or difficult to sample from, but where the conditional densities $p(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Y})$ and $p(\mathbf{X}|\boldsymbol{\theta}, \mathbf{Y})$ are known, or easy to sample from. Since this is usually true for state-space models, the Gibbs sampling algorithm has been extensively used in this scenario (Carter and Kohn, 1994; Geweke and Tanizaki, 2001).

Considering a parameter sample $\boldsymbol{\theta}^{(i)}$, a basic two-state Gibbs sampler generates a state sample $\mathbf{X}^{(i+1)}$ from $\mathbf{X}_{i+1} \sim p(\mathbf{X}|\boldsymbol{\theta}^{(i)}, \mathbf{Y})$. Next, $\boldsymbol{\theta}^{(i+1)}$ is sampled from $\boldsymbol{\theta}^{(i+1)} \sim p(\boldsymbol{\theta}|\mathbf{X}^{(i+1)}, \mathbf{Y})$. The procedure is repeated until some termination criterion is met.

Despite the advantages of MCMC methods, they are all stochastic approximation techniques where the final distributional approximation is obtained from Markov chain paths, which are random in nature. Associated methods exhibit computational inefficiency and it is difficult to establish chain convergence with acceptable error. Such limitations become even more pronounced in high-dimensional systems (Mackay, 1998), such as spatio-temporal systems, thereby driving research efforts into approximate deterministic inference techniques such as EM and VBEM.

4. Applications in Engineering

Several engineering problems require the monitoring and control of spatio-temporal systems. Problems arising in spatio-temporal contamination (e.g. oil spills and contaminating fluid leaks), spatio-temporal monitoring and identification (such as monitoring the carbon footprint), search and rescue missions and wildfire control (Demetriou and Hussein, 2009) lead to an estimation problem that must be addressed.

The spatio-temporal monitoring problem is oftenly treated in the context of scheduling mechanisms that ensure adequate field monitoring by sensors (Gupta et al., 2006; Choi, 2009). Since practical situations involve sensors that cannot operate simultaneously, such as sonars occupying the same frequency band, or mobile sensors covering a limited geographical area, an optimal estimation

strategy is desired. Most of the associated monitoring problems are therefore formulated as an optimal control problem minimising the expected error covariance and is usually solved using dynamic programming techniques. A sensor planning strategy that minimises the field estimation error using Lyapunov techniques was proposed by Demetriou and Hussein (2009); Demetriou (2010). Sensor trajectory design and measurement strategies for the parameter estimation problem has been considered by Uciński (2005); Tricaud et al. (2008).

One spatio-temporal application in engineering is a smart structure system, where a number of sensors and actuators are installed into a structure such that this is able to interact with the external environment (Anderson et al., 2014). Apichayakul and Kadirkamanathan (2011) show how a spatio-temporal model may be constructed as an alternative way of modelling the structure. A robust EM algorithm is used for the estimate the parameters of the the spatio-temporal state-space model from experimental data.

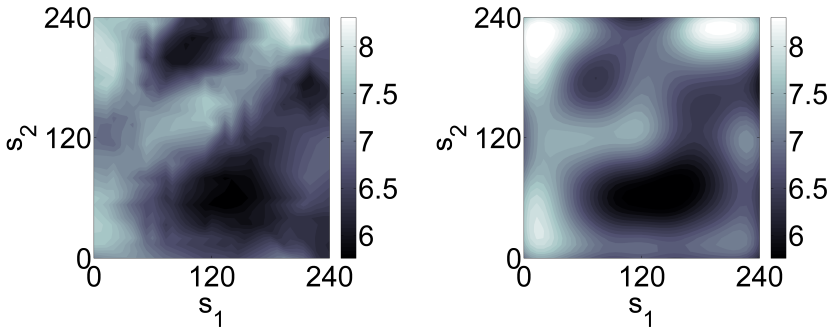
An important spatio-temporal system is fluid flow which has been a topic of increased recent interest due to its part in acting as the main disturbance in the control system of wind turbines. Maximising energy production and mitigating structural loads using the knowledge of oncoming wind is deeemed essential for this renewable energy source. Such information is critical for the preview control of wind turbines, as described by Schlipf and Pao (2014) in one of the current control research challenges documented in *The Impact of Control Technology* report published by the IEEE Control Systems Society. The accurate prediction of wind flow is further required for strategic placing of wind farms which should be located in regions of optimal potential load factor (Jefferson, 2008).

Although sampling an oncoming wind field has become possible with recent advances in flow measurement, such as light detection and ranging instrumentation (LiDAR), we are still left with the compelling questions of how best to use such sparse measurements to predict wind gusts and to incorporate such knowledge within a preview control scheme (Wang et al., 2012; Schlipf et al., 2013). These controllers will rely upon the accuracy of wind field prediction, which therefore calls for wind velocity estimation tools that predict wind turbine gusts using limited spatio-temporal wind velocity measurements (Angelou et al., 2010), thereby mitigating the possible blade damage due to severe wind gusts if the blade pitch is altered in a timely manner (Dunne et al., 2011; Wang et al., 2012; Kragh et al., 2013). This would of course link measurements to regions of flow which are not directly observed. Figure 1 shows a visual comparison of a typical generated and estimated atmospheric boundary layer wind flow field at a single time instant, based on wind velocity data obtained from 49 wind velocity sensors arranged over a regular grid.

The Navier-Stokes equations that govern fluid flow have a PDAE formulation, which for viscous incompressible flow is given by

$$\frac{\partial \mathbf{U}(\mathbf{s},t)}{\partial t} = -\nabla P(\mathbf{s},t) - \mathbf{U}(\mathbf{s},t)\cdot\nabla\mathbf{U}(\mathbf{s},t) + \frac{1}{Re}\nabla^2\mathbf{U}(\mathbf{s},t), \tag{24}$$

$$0 = \nabla\cdot\mathbf{U}(\mathbf{s},t), \tag{25}$$



(a) (b)

Figure 1: A typical result for wind velocity estimation is visualised here by showing a single time instant of (a) the generated wind data and (b) the estimated flow field, based on data obtained from 49 wind velocity sensors arranged on a regular grid. The contours represent the wind velocity magnitude in ms^{-1} within a 240m square domain at a height of 100m above sea level.

where $\mathbf{U}(\mathbf{s}, t)$ and $P(\mathbf{s}, t)$ denote the velocity and pressure fields, respectively, evolving over domain $\Omega \in \mathbb{R}^d$ for d-dimensional flow, with time $t \in \mathbb{R}_+$ and $\mathbf{s} \in \Omega$. The term Re denotes Reynolds number and the superscript $\top$ is the transpose operator. A notable feature of equations (24) and (25) is that no explicit equation exists for the pressure P. Also, since the pressure P is only determined up to an additive constant, the system is said to be undetermined and the concept of the differentiation index (Brenan et al., 1996) cannot be readily applied (Weickert, 1997).

Owing to the intractable nature of the Navier-Stokes equations in their original PDAE form, the state estimation of spatio-temporal systems governed by these equations remains a challenging task since the majority of established estimation techniques are designed for finite-dimensional systems in ordinary differential equation form. Retaining the full DAE formulation is a very attractive consideration due to the resulting pressure field description that would become essential for yet another wind flow application: flow control for reduced drag in transport vehicles. A pressure difference across a vehicle amounts to pressure drag (Bertin and Cummings, 2013), which constitutes 80% of ground transportation drag (Wood, 2004). It is estimated that 16% of the total energy consumed in the United States is used to overcome aerodynamic drag (Wood, 2004).

We note that a simplified wind model has been proposed by Soleimanzadeh et al. (2014), however this is derived as the spatial discretisation of the linearised incompressible Navier-Stokes equations. A recent work reported by Towers and Jones (2016) has derived a simplified deterministic state-space model of atmospheric boundary layer flow but is based on spatial discretisation and excludes pressure. If the spatio-temporal descriptor formulation of the Navier-Stokes equations is to be retained, an appropriate estimation framework is needed to

overcome the problems highlighted here.

5. Applications in Healthcare

The human brain is a highly sophisticated system which exhibits complex spatio-temporal dynamics. The electrical activity in the brain can be observed across space and through time via different electrographic modalities such as electroencephalography (EEG) using scalp or subdural array of electrodes. The underlying mechanisms of spatio-temporal pattern formation associated with normal and abnormal neural activities are normally hidden in electrophysiological recordings.

Theoretical brain models have been designed to explain brain's function and to predict complex neural rhythms. These are defined based on known biophysical principles which enables a physiologically relevant interpretation of model parameters.

The neural activity of the cortex at the mesoscopic scale can be represented by neural field models which describe spatiotemporal neurodynamics on a continuous cortical sheet. These mean-field models can be used in a data-driven framework where the patient-specific clinical measurements are incorporated into the model to estimate unmeasured system properties or parameters. Estimation of patient-specific parameters has the potential to transform understanding and treatment of neurological diseases. Here we briefly review data-driven approaches for neural field modeling using intracranial EEG (iEEG) recordings. Alternative mean-field approximations are neural mass models which have also been used in model-based frameworks using electrophysiological data (Freestone et al., 2013, 2014).

The stochastic IDE formulation of the Amari neural field model is given by Freestone et al. (2011)

$$z_k(\mathbf{s}) = \xi z_{k-1}(\mathbf{s}) + \Delta_t \int_{\mathcal{S}} \kappa(\mathbf{s}-\mathbf{r})\, f(z_{k-1}(\mathbf{r}))\, \mathrm{d}\mathbf{r} + e_{k-1}(\mathbf{s}), \tag{26}$$

where ξ is the time constant parameter of the membrane, and Δ_t is the sampling time step. The index $k \in \mathbb{Z}_0$ is discrete time and $\mathbf{s} \in \mathbb{R}^2$ are spatial locations in two-dimensional cortical surface, $\mathcal{S}$.

The field of postsynaptic potentials at time k and location $\mathbf{s}$ is denoted by $z_k(\mathbf{s})$ and is mapped through time via its convolution with the connectivity kernel, $\kappa(\mathbf{s}-\mathbf{r})$, and $f(z_k(\mathbf{s}))$ is a sigmoidal firing rate function. The connectivity kernel is commonly assumed to have a Mexican-hat shape with central excitation and surrounding inhibition (see Fig. 2). The disturbance $e_k(\mathbf{s})$ is a zero-mean normally distributed noise process.

The measurement equation is given by

$$y_k(\mathbf{s}_n) = \int_{\mathcal{S}} m(\mathbf{s}_n - \mathbf{s})\, z_k(\mathbf{s})\, \mathrm{d}\mathbf{s} + \varepsilon_k(\mathbf{s}_n), \tag{27}$$

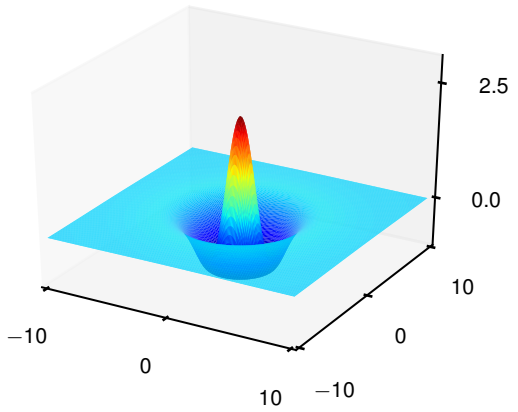


Figure 2: Mexican-hat connectivity kernel with central excitation and surrounding inhibition. The spatial mixing kernel governs the spatiotemporal dynamics of the neural field.

where $m(\cdot)$ models the intracranial sensor at spatial location $\mathbf{s}_n$ and $\varepsilon_k(\mathbf{s}_n)$ denotes a zero mean Gaussian white noise.

A state-space representation of (26) and (27) can be derived by decomposing the spatial field, $z_k(\mathbf{s})$, and the connectivity kernel, $\kappa(\mathbf{s})$, using a basis function decomposition. This can be done by writing the field and the kernel as sums of weighted basis functions. The weights on the field basis functions and the connectivity kernel form the states and the parameters of the state-space model respectively. The spacing and the width of basis functions can be determined using spatial spectral analysis (Aram et al., 2015b).

In (Freestone et al., 2011) Gaussian radial basis functions were used for the decomposition, allowing analytic computations of the model terms in a reduced state-space form. Alternatively B-spline scaling and wavelet functions can be adopted for simultaneous reconstruction of the neural field at different spatial scales (Aram et al., 2013). In addition to muitiresolution approximation (MRA) property, other advantages of B-splines scaling and wavelet functions include partition of unity and compact support (Goswami and Chan, 1999). However, the increase in computational requirements for the estimation of the state-space model will be significant.

The state-space formulation of the model allows the application of standard techniques for iterative estimation of the states (spatial fields) and the parameters (connectivity kernel). The resulting state equation will be non-linear due to the non-linear sigmoidal firing rate in (26), and therefore non-linear filtering or smoothing techniques are required for the state estimation step. Although the resulting state equation is non-linear the parameters of the system, connectivity weights, are linear with respect to the state. Therefore, a least squares method can be used for the parameter estimation step. By assuming a linear

form for the activation function, $f(z_k(\mathbf{s}))$, one can utilise the standard EM algorithm, applying the smoothing for the E-step and forming and maximizing the expected log-likelihood in the M-step (Aram et al., 2013). The results in these works showed that it is theoretically plausible to estimate the neural field, connectivity kernel and the membrane dynamics.

Using complicated iterative algorithms to estimate spatiotemporal characteristics of the neural field equation can potentially limit its application in practice. An efficient approach for computing a closed-form estimation of the connectivity kernel from average (over time) spatial correlations of iEEG recordings were developed by Aram et al. (2015a). The proposed method was then used to monitor the connectivity changes during normal activity and an epileptic seizure. The estimates of the connectivity kernel provided a plausible description for differences in seizure spread in different epilepsy patients. In particular it was shown that the loss of surround inhibition in the connectivity structure can contain epileptic events.

The estimated algorithm based on spatial correlation technique was further developed to solve general IDE model of the form described in equation (8) (Aram and Freestone, 2016). This work does not provide an estimate of the spatial field, however, if one is interested in the field reconstruction, the kernel estimate can be used as an initialisation for state-space estimation frameworks developed by Dewar et al. (2009); Scerri et al. (2009), improving the speed and the convergence of the estimation procedure.

The importance of data-driven neural models is increasing with technological advances in neural recording devices with enhanced spatial and temporal resolutions. These models provide patient-specific insight into normal and abnormal cortical functions by estimating the spatial field and physiologically meaningful parameters. Furthermore, such models can be potentially used to facilitate the application of therapeutic electrical stimulation where the feedback from the model will allow for closed-loop brain stimulation to control epileptic seizures.

6. Applications in Social Science

Social phenomena are being increasingly analysed for their spatio-temporal behaviour (King, 2011). The Arab springs, which is a revolutionary wave of demonstrations in the Arab world, represent one such social pattern which is affected by social media and global spatial interactions. Another social phenomenon is armed conflict (Johnson et al., 2011). These spatio-temporal interactions may not only be studied to describe but also predict future activity with confidence measures.

Social processes are typically investigated in terms of spatio-temporal point processes, which are stochastic processes where samples are described by a countable collection of space-time points. Such events having random locations and times may be assumed to be generated by a non-homogenous (with respect to a spatio-temporally varying intensity) Poisson process described by an intensity function $\lambda(\mathbf{s}, t)$, which is usually a function of a secondary continuous stochastic process $z(\mathbf{s}, t)$ (Smith and Brown, 2003). The intensity func-

tion is commonly represented by a log-Gaussian Cox process (LGCP), where the logarithm of the event intensity is taken to be a Gaussian process, i.e. $\ln \lambda(\mathbf{s}, t) \sim \mathcal{GP}(\cdot, \cdot)$ (Møller et al., 1998). Although this approach is advantageous due to the simplicity of its parameter inference scheme, known as the method of contrast (Diggle et al., 2005), this spatio-temporal intensity function does not allow for appropriate parametrisation that reveals an adequate interpretation of the underlying physical mechanism, rendering it less convenient for control purposes (Zammit-Mangion, 2011).

A dynamic systems approach solves this issue by having the intensity governed by a stochastic state-space model (Smith and Brown, 2003; Eden et al., 2004), which provides the added benefit of generally having unimodal state and parameter probability densities facilitating the use of approximative techniques (Yuan and Niranjan, 2010). The dynamic systems approach is adopted by Zammit-Mangion et al. (2012a) for predicting armed conflict and dynamic spatio-temporal modelling tools are proposed for the identification of complex underlying conflict processes including volatility, diffusion, relocation and heterogeneous escalation. Since conflict data is typically available in discrete-time form specifying the event date, rather than time, a discrete-time series of the continuous-space LGCP is considered. For each discrete-time index $k \in \mathcal{K}$, $\mathcal{K} = \{1, 2, \ldots, K\}$, the point-process intensity function employed is $\lambda_k(\mathbf{s}) = \exp(\mathbf{a}^\top \mathbf{b}(\mathbf{s}) + z_k(\mathbf{s}))$, where $z_k(\mathbf{s})$ is a spatial Gaussian process, $\mathbf{b}$ is a vector of spatially referenced covariates and $\mathbf{a}$ is the associated regression parameter vector. This allows the mean function of $z_k(\mathbf{s})$ to be related to descriptive variables such as population density, that enhance prediction accuracy.

To model the complex conflict dynamics, Zammit-Mangion et al. (2012a) make use of the stochastic IDE (SIDE), which is of the form given by equation (8) with the inclusion of an additive disturbance $e_k(\mathbf{s})$ assumed to be a Gaussian process. The study analyses the correlation between conflict events by determining probabilities of having a conflict occurrence at $\mathbf{r}$ given that another occurrence happened at $\mathbf{s}$ at time frame k or $k-1$. Such quantities are calculated using pair auto-correlation functions (PACFs). The bayesian inference of the SIDE-driven LGCPs requires a finite-dimensional reduction method which Zammit-Mangion et al. (2012a) carry out using a basis function decomposition method that follows the form given by equation (13). A general basis function selection method required for LGCPs is proposed by providing a relationship between the point-process frequency content and the PACF, following the ideas presented by Scerri et al. (2009); Freestone et al. (2011) which were discussed in Section 2.3.

The SIDE is then represented as a standard linear state-space model of the form

$$\mathbf{x}_{k+1} = A(\boldsymbol{\eta}_I)\mathbf{x}_k + \mathbf{q}_k(\boldsymbol{\vartheta}, \boldsymbol{\eta}_Q), \tag{28}$$

where $A(\boldsymbol{\eta}_I) \in \mathbb{R}^{n_\phi \times n_\phi}$, n_ϕ is the number of basis functions and $\mathbf{q}_k \in \mathbb{R}^{n_\phi}$ denotes a Gaussian coloured noise term having mean $\mathbb{E}[\mathbf{q}_k] = \boldsymbol{\vartheta}$ and covariance $\mathrm{cov}[\mathbf{q}_k] = \boldsymbol{\eta}_Q$. The joint state and parameter estimation problem that arises

is to estimate the states $\mathbf{X}_K = \{\mathbf{x}_k\}_{k=0}^K$ and the unknown parameters $\boldsymbol{\Theta} = \{\boldsymbol{\eta}_I, \boldsymbol{\vartheta}, \boldsymbol{\eta}_Q^{-1}\}$ given the data $\mathbf{Y}_K = \{\mathbf{y}_k\}_{k=1}^K$ which take the form of a set of coordinates for a logged event occurring at the kth time frame. A variational Bayes method is used to approximate the full posterior distribution

$$p(\mathbf{X}_K, \boldsymbol{\Theta}, \mathbf{a}|\mathbf{Y}_K) \approx \tilde{p}(\mathbf{X}_k)\tilde{p}(\boldsymbol{\vartheta})\tilde{p}(\boldsymbol{\eta}_I)\tilde{p}(\boldsymbol{\eta}_Q^{-1})\tilde{p}(\mathbf{a}), \tag{29}$$

with the variational marginals (Beal, 2003; Šmídl and Quinn, 2005) each revealing critical properties of conflict progression.

Zammit-Mangion et al. (2012a) show how their methods successfully model and predict conflict using data from the WikiLeaks Afghan War Diary (AWD), which contains around 77,000 logs of conflict events (such as gunfights or security checks) complete with their time and location. In particular, $\mathbf{X}_K$ is used for spatio-temporal field reconstruction and state inference therefore provides information about where and how the conflict intensity changes over space and time. The regression parameter $\mathbf{a}$ associated with population density and proximity to the closest major city revealed how most of the AWD logs were located in urban areas and regions of high population. Conflict escalation in Afghanistan was possible to identify using the parameter $\boldsymbol{\vartheta}$, which describes the spatially varying conflict escalation. This feature further allows the user to distinguish between isolated events or increasingly alarming situations. The parameter $\boldsymbol{\eta}_Q$ represents conflict volatility and is therefore used to determine the predictability of conflict. In other words, a large diagonal value in $\boldsymbol{\eta}_Q$ is a sign of significant volatility in the region, meaning that any future predictions made will be highly uncertain.

The developed dynamic point process modelling strategy further allows statistical predictions of the system's behaviour. By using the obtained generative spatio-temporal model, it is shown that the predictions made are statistically accurate, that is, a close match across Afghan provinces is obtained for the predicted and observed distribution of the armed opposition groups growth. A key result is the statistically accurate prediction of conflict dynamics for a whole year following the end of the available AWD data used to train the model.

7. Concluding remarks

This paper has reviewed estimation and identification methods for spatio-temporal systems. The inference techniques considered adopt a dynamic systems approach that make them amenable to control applications. Advances in sensor systems are enabling novel spatio-temporal processes to be partially observed and leading to new estimation and identification problems in a variety of applications. Three classes of applications, in engineering, healthcare and social science, were briefly reviewed. The examples serve to highlight the breadth of applicability of estimation and identification of spatio-temporal processes. Such recent advances have paved the way for future work that may potentially answer questions that we could not answer before. The ideas presented for the

unconventional application of predicting armed conflict, for instance, may provide predictive abilities that could influence decisions in peace-keeping efforts. This and several other applications may be better served by further advances in this area of research.

Acknowledgement

The authors would like to thank Dr. Parham Aram for his contribution in the review of spatio-temporal estimation and identification applications in healthcare and providing insightful comments that improved the overall quality of the paper. This work was partially supported by the University of Sheffield PhD Scholarship to Julian Mercieca and the EPSRC project (EP/K03877X/1) to Visakan Kadirkamanathan.

References

Anderson, S., Aram, P., Bhattacharya, B., Kadirkamanathan, V., 2014. Analysis of composite plate dynamics using spatial maps of frequency-domain features described by Gaussian processes. IFAC Proceedings Volume 47 (1), 949 – 954, 3rd International Conference on Advances in Control and Optimization of Dynamical Systems (2014).

Angelou, N., Mann, J., Courtney, M., Sjöholm, M., 2010. Doppler Lidar mounted on a wind turbine nacelle - upwind deliverable D6.7.1. Tech. rep., Risø-R-1757(EN), Risø DTU.

Anselin, L., 1988. Spatial econometrics: Models and applications. Kluwer Academic, Dordrecht, Netherlands.

Apichayakul, P., Kadirkamanathan, V., 2011. Spatio-temporal dynamic modelling of smart structures using a robust expectation-maximization algorithm. Smart Materials and Structures 20 (4), 045015.

Aram, P., Freestone, D., 2016. Estimation of the mixing kernel and the disturbance covariance in IDE-based spatiotemporal systems. Signal Processing 121, 46–53.

Aram, P., Freestone, D., Cook, M., Kadirkamanathan, V., Grayden, D., 2015a. Model-based estimation of intra-cortical connectivity using electrophysiological data. NeuroImage 118, 563–575.

Aram, P., Freestone, D., Dewar, M., Scerri, K., Jirsa, V., Grayden, D., Kadirkamanathan, V., 2013. Spatiotemporal multi-resolution approximation of the Amari type neural field model. NeuroImage 66, 88–102.

Aram, P., Kadirkamanathan, V., Anderson, S., 2015b. Spatiotemporal system identification with continuous spatial maps and sparse estimation. IEEE Transactions on Neural Networks and Learning Systems 26 (11), 2978–2983.

Arulampalam, M., Maskell, S., Gordon, N., Clapp, T., 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Transactions on Signal Processing 50 (2), 174–188.

Asensio, M. I., Ferragut, L., 2002. On a wildland fire model with radiation. International Journal for Numerical Methods in Engineering 54 (1), 137–157.

Attias, H., 1999. Inferring parameters and structure of latent variable models by variational bayes. In: Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence. pp. 21–30.

Attias, H., 2000. A variational Bayesian framework for graphical models. In: Advances in Neural Information Processing Systems. Vol. 12. pp. 209–215.

Banks, H. T., Kunisch, K., 1989. Estimation Techniques for Distributed Parameter Systems. Birkhäuser, Boston, USA.

Beal, M. J., 2003. Variational algorithms for approximate Bayesian inference. Ph.D. thesis, Gatsby Computational Neuroscience Unit, University College London, London, UK.

Beal, M. J., Falciani, F., Ghahramani, Z., Rangel, C., Wild, D. L., 2005. A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. Bioinformatics 21 (3), 349–356.

Becerra, V., Roberts, P., Griffiths, G., 2001. Applying the extended Kalman filter to systems described by nonlinear differential-algebraic equations. Control Engineering Practice 9 (3), 267–281.

Bennett, A., 2002. Inverse Modeling of the Ocean and Atmosphere. Cambridge University Press, Cambridge, UK.

Bennett, R. J., 1979. Spatial Time Series. Pion, London, UK.

Bertin, J. J., Cummings, R. M., 2013. Aerodynamics for Engineers, Sixth Edition. Prentice Hall, NJ, USA.

Besag, J., 1974. Spatial interaction and statistical analysis of lattice systems. Journal of the Royal Statistical Society B 36 (2), 192–236.

Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Drake, J. M., Brownstein, J. S., Hoen, A. G., Sankoh, O., Myers, M. F., George, D. B., Jaenisch, T., Wint, G. R. W., Simmons, C. P., Scott, T. W., Farrar, J. J., Hay, S. I., 2013. The global distribution and burden of dengue. Nature 496 (7446), 504–507.

Billings, S. A., 2013. Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains. John Wiley & Sons., West Sussex, UK.

Billings, S. A., Coca, D., 2002. Identification of coupled map lattice models of deterministic distributed parameter systems. International Journal of Systems Science 33 (8), 623–634.

Billings, S. A., Guo, L. Z., Wei, H. L., 2004. Projecting rates of spread for invasive species. Risk Analysis 24 (4), 817–831.

Billings, S. A., Guo, L. Z., Wei, H. L., 2006. Identification of coupled map lattice models for spatio-temporal patterns using wavelets. International Journal of Systems Science 37 (14), 1021–1038.

Bishop, C. M., 2006. Pattern Recognition and Machine Learning. Springer, New York.

Bocquet, M., Pires, C. A., Wu, L., 2010. Beyond Gaussian statistical modeling in geophysical data assimilation. Monthly Weather Review 138 (8), 2997–3023.

Box, G., Jenkins, G., 1970. Time Series Analysis: Forecasting and Control. Holden-Day, San Fransisco, CA, USA.

Brenan, K. E., Campbell, S. L., Petzold, L. R., 1996. Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations. SIAM, Philadelphia, PA, USA.

Brooks-Pollock, E., Roberts, G. O., Keeling, M. J., 2014. A dynamic model of bovine tuberculosis spread and control in great britain. Nature 511, 228231.

Brown, P. E., Diggle, P. J., Lord, M. E., Young, P. C., 2001. Space-time calibration of radar rainfall data. Journal of the Royal Statistical Society. Series C (Applied Statistics) 50 (2), 221–241.

Brown, P. E., Roberts, G. O., Kresen, K. F., Tonellato, S., 2000. Blur-generated non-separable spacetime models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62 (4), 847–860.

Bunimovich, L. A., 1995. Coupled map lattices: one step forward and two steps back. Physica D: Nonlinear Phenomena 86 (12), 248 – 255.

Burgers, G., Leeuwen, P., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. Monthly Weather Review 126, 1719–1724.

Cappe, O., Godsill, S., Moulines, E., 2007. An overview of existing methods and recent advances in sequential Monte Carlo. Proceedings of the IEEE 95 (5), 899–924.

Carmona, R. A., 1998. Stochastic Partial Differential Equations: Six Perspectives. American Mathematical Society.

Carter, C. K., Kohn, R., 1994. On Gibbs sampling for state space models. Biometrika 81 (3), 541–553.

Casella, G., Berger, R., 1990. Statistical inference. Brooks/Cole Publishing Company, Pacific Grove, CA, USA.

Cemgil, A. T., Févotte, C., Godsill, S. J., 2007. Variational and stochastic inference for Bayesian source separation. Digit. Signal Process. 17 (5), 891–913.

Chen, Z., 2003. Bayesian filtering: From Kalman filters to particle filters, and beyond. Tech. rep., Adaptive Syst.Lab., McMaster University, Ontario, Canada.

Ching, J., Beck, J. L., Porter, K. A., 2006. Bayesian state and parameter estimation of uncertain dynamical systems. Probabilistic Engineering Mechanics 21 (1), 81 – 96.

Choi, H., 2009. Adaptive sampling and forecasting with mobile sensor networks. Ph.D. thesis, MIT, Cambridge, MA, USA.

Christopher K. Wikle, Ralph F. Milliff, D. N., Berliner, L. M., 2001. Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. Journal of the American Statistical Association 96 (454), 382–397.

Chudej, K., Petzet, V., Scherdel, S., Pesch, H. J., Schittkowski, K., Heidebrecht, P., Sundmacher, K., 2003. Index analysis of a nonlinear PDAE system describing a molten carbonate fuel cell. Proceedings in Applied Mathematics and Mechanics 3 (1), 563–564.

Coca, D., Billings, S., 2001. Identification of coupled map lattice models of complex spatio-temporal patterns. Physics Letters A 287 (12), 65 – 73.

Coca, D., Billings, S. A., 2000. Direct parameter identification of distributed parameter systems. International Journal of Systems Science 31 (1), 11–17.

Coca, D., Billings, S. A., 2002. Identification of finite dimensional models of infinite dimensional dynamical systems. Automatica 38 (11), 1851 – 1865.

Coca, D., Billings, S. A., 2003. Analysis and reconstruction of stochastic coupled map lattice models. Physics Letters A 315 (12), 61 – 75.

Cressie, N., 1993. Statistics for spatial data. John Wiley & Sons., New York, NY, USA.

Cressie, N., Wikle, C. K., 2002. Space-time Kalman filter. Encyclopedia of Environmetrics 4, 2045–2049.

Cressie, N., Wikle, C. K., 2011. Statistics for Spatio-Temporal Data. Wiley, Hoboken, NJ.

Dalang, R. C., Frangos, N. E., 1998. The stochastic wave equation in two spatial dimensions. The Annals of Probability 26 (1), 187–212.

Daum, F., 2005. Nonlinear filters: beyond the Kalman filter. IEEE Aerospace and Electronic Systems Magazine 20 (8), 57–69.

de Munck, J., Huizenga, H., Waldorp, L., Heethaar, R., 2002. Estimating stationary dipoles from MEG/EEG data contaminated with spatially and temporally correlated background noise. IEEE Transactions on Signal Processing 50 (7), 1565–1572.

Debnath, L., 2005. Nonlinear Partial Differential Equations for Scientists and Engineers, 2nd Edition. Birkhauser, New York, USA.

Deissler, R. J., 1987. Spatially growing waves, intermittency, and convective chaos in an open-flow system. Physica D: Nonlinear Phenomena 25 (13), 233 – 260.

Deissler, R. J., Kaneko, K., 1987. Velocity-dependent lyapunov exponents as a measure of chaos for open-flow systems. Physics Letters A 119 (8), 397 – 402.

Dellaert, F., 2002. The expectation maximization algorithm, Technical Report number GIT-GVU-02-20, Georgia Institute of Technology. Tech. rep.

Demetriou, M., 2010. Guidance of mobile actuator-plus-sensor networks for improved control and estimation of distributed parameter systems. IEEE Transactions on Automatic Control 55 (7), 1570–1584.

Demetriou, M., Paskaleva, A., Vayena, O., Doumanidis, H., 2003. Scanning actuator guidance scheme in a 1-D thermal manufacturing process. IEEE Transactions on Control Systems Technology 11 (5), 757–764.

Demetriou, M. A., Hussein, I. I., 2009. Estimation of spatially distributed processes using mobile spatially distributed sensor network. SIAM Journal on Control and Optimization 48 (1), 266–291.

Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39 (1), 1–38.

Dewar, M., Kadirkamanathan, V., Oct 2007. A canonical space-time state space model: State and parameter estimation. IEEE Transactions on Signal Processing 55 (10), 4862–4870.

Dewar, M., Scerri, K., Kadirkamanathan, V., 2009. Data-driven spatio-temporal modeling using the integro-difference equation. IEEE Transactions on Signal Processing 57 (1), 83–91.

Di Giacinto, V., 2006. A generalized space-time ARMA model with an application to regional unemployment analysis in italy. International Regional Science Review 29 (2), 159–198.

Diggle, P., Rowlingson, B., Su, T.-l., 2005. Point process methodology for on-line spatio-temporal disease surveillance. Environmetrics 16 (5), 423–434.

Dochain, D., 2003. State and parameter estimation in chemical and biochemical processes: a tutorial. Journal of Process Control 13 (8), 801 – 818.

Duan, J., Goldys, B., 2001. Ergodicity of stochastically forced large scale geophysical flows. International Journal of Mathematics and Mathematical Sciences 28 (6), 313–320.

Duan, J. A., Gelfand, A. E., Sirmans, C. F., 2009. Modeling space-time data using stochastic differential equations. Bayesian Analysis 4 (4), 733–758.

Dunne, F., Simley, E., Pao, L., 2011. Lidar wind speed measurement analysis and feed-forward blade pitch control for load mitigation in wind turbines. Tech. rep., NREL Report NREL/SR-5000-52098.

Eden, U. T., Frank, L. M., Barbieri, R., Solo, V., Brown, E. N., 2004. Dynamic analysis of neural encoding by point process adaptive filtering. Neural Computation 16 (5), 971–998.

Evans, L., 1998. Partial Differential Equations (Graduate Studies in Mathematics, Vol. 19). AMS, Providence, RI, USA.

Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dynamics 53 (4), 343–367.

Fabri, S., Kadirkamanathan, V., 2001. Functional Adaptive Control: an Intelligent Systems Approach. Springer-Verlag, London, UK.

Farlow, S., 1993. Partial Differential Equations for Scientists and Engineers. Dover Publications.

Fisher, R. A., 1937. The wave of advance of advantageous genes. Annals of Eugenics 7 (4), 355–369.

Freestone, D., Aram, P., Dewar, M., Scerri, K., Grayden, D., Kadirkamanathan, V., 2011. A data-driven framework for neural field modeling. NeuroImage 56 (3), 1043–1058.

Freestone, D., Kuhlmann, L., Chong, M., Nesic, D., Grayden, D., Aram, P., Postoyan, R., Cook, M., 2013. Patient-specific neural mass modelling - stochastic and deterministic methods. World Scientific, Ch. 5, pp. 63–82.

Freestone, D. R., Karoly, P. J., Nešić, D., Aram, P., Cook, M. J., Grayden, D. B., 2014. Estimation of effective connectivity via data-driven neural modeling. Frontiers in Neuroscience 8.

Gelb, A., 1974. Applied optimal estimation. The MIT Press, Cambridge, MA, USA.

Gelfand, A. E., Hills, S. E., Racine-Poon, A., Smith, A. F. M., 1990. Illustration of Bayesian inference in normal data models using Gibbs sampling. Journal of the American Statistical Association 85 (412), 972–985.

Geweke, J., Tanizaki, H., 2001. Bayesian estimation of state-space models using the MetropolisHastings algorithm within Gibbs sampling. Computational Statistics Data Analysis 37 (2), 151 – 170.

Gibson, S., Ninness, B., 2005. Robust maximum-likelihood estimation of multivariable dynamic systems. Automatica 41 (10), 1667 – 1682.

Glasbey, C. A., Allcroft, D. J., 2008. A spatiotemporal auto-regressive moving average model for solar radiation. Journal of the Royal Statistical Society: Series C (Applied Statistics) 57 (3), 343–355.

Gneiting, T., Genton, M. G., Guttorp, P., 2007. Geostatistical space-time models, stationarity, separability, and full symmetry. In: Finkenstaedt, B., Held, L., Isham, V. (Eds.), Statistical Methods for Spatio-Temporal Systems. Chapman and Hall/CRC Press, Boca Raton, FL, USA, Ch. 4, pp. 151–175.

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

Grossmann, C., Roos, H. G., Stynes, M., 2007. Numerical Treatment of Partial Differential Equations. Springer-Verlag, Berlin, Germany.

Gu, D., Hu, H., 2012. Spatial Gaussian process regression with mobile sensor networks. IEEE Transactions on Neural Networks and Learning Systems 23 (8), 1279–1290.

Guo, L., Billings, S. A., 2006. Identification of partial differential equation models for continuous spatio-temporal dynamical systems. IEEE Transactions on Circuits and Systems II: Express Briefs 53 (8), 657–661.

Guo, L. Z., Billings, S. A., Coca, D., 2009. Consistent recursive parameter estimation of partial differential equation models. International Journal of Control 82 (10), 1946–1954.

Guo, L. Z., Mei, S. S., Billings, S. A., 2007. Neighbourhood detection and identification of spatio-temporal dynamical systems using a coarse-to-fine approach. International Journal of Systems Science 38 (1), 1–15.

Gupta, V., Chung, T. H., Hassibi, B., Murray, R. M., 2006. On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage. Automatica 42 (2), 251–260.

Guttorp, P., Sampson, P. D., 1994. 20 methods for estimating heterogeneous spatial covariance functions with environmental applications. In: Patil, G. P., Rao, C. R. (Eds.), Environmental Statistics. Vol. 12 of Handbook of Statistics. Elsevier, Amsterdam, Netherlands, pp. 661–689.

Hastings, W. K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57 (1), 97–109.

Hausenblas, E., 2003. Approximation for semilinear stochastic evolution equations. Potential Analysis 18 (2), 141–186.

Helling, G., 1960. Partial Differential Equations. An Introduction. Blaisdell Publishing.

Hill, J., 1987. Heat Conduction. Blackwell Scientific, Oxford, UK.

Holden, A., Tucker, J., Zhang, H., Pole, M., 1992. Coupled map lattice as computational systems. Chaos 2 (3), 367–376.

Holmes, E. E., Lewis, M. A., Banks, J. E., Veit, R. R., 1994. Partial differential equations in ecology: spatial interactions and population dynamics. Ecology 75 (1), 17–29.

Hooten, M. B., Wikle, C. K., 2008. A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. Environmental and Ecological Statistics 15 (1), 59–70.

Hooten, M. B., Wikle, C. K., Dorazio, R. M., Royle, J. A., 2007. Hierarchical spatiotemporal matrix models for characterizing invasions. Biometrics 63 (2), 558–567.

Hoyle, R., 2006. Pattern Formation. Cambridge University Press, Cambridge, UK.

Ikegami, T., Kaneko, K., 1992. Evolution of host-parasitoid network through homeochaotic dynamics. Chaos 2 (3), 397–407.

Jazwinski, A., 1970. Stochastic processes and filtering theory. Dover Publications, NY, USA.

Jefferson, M., 2008. A critical note on the efficacy of the uk's renewable obligation system as it applies to onshore wind energy developments in england. London Metropolitan University - Centre for International Business and Sustainability (CIBS), Working Papers (1).

Johnson, N., Carran, S., Botner, J., Fontaine, K., Laxague, N., Nuetzel, P., Turnley, J., Tivnan, B., 2011. Pattern in escalations in insurgent and terrorist activity. Science 333 (6038), 81–84.

Jost, J., Joy, M. P., 2001. Spectral properties and synchronization in coupled map lattices. Phys. Rev. E 65 (1), 016201.

Julier, S., Uhlmann, J., 2004. Unscented filtering and nonlinear estimation. Proceedings of the IEEE 92 (3), 401–422.

Julier, S., Uhlmann, J., Durrant-Whyte, H., 2000. A new method for the nonlinear transformation of means and covariances in filters and estimators. IEEE Transactions on Automatic Control 45 (3), 477–482.

Kalman, R. E., 1960. A new approach to linear filtering and prediction problems. Transactions of the ASME Journal of Basic Engineering 82, 35–45.

Kaneko, K., 1985. Spatial period-doubling in open flow. Physics Letters A 111 (7), 321 – 325.

Kaneko, K., 1986. Lyapunov analysis and information flow in coupled map lattices. Physica D: Nonlinear Phenomena 23 (13), 436 – 447.

Kaneko, K., 1989. Pattern dynamics in spatiotemporal chaos: Pattern selection, diffusion of defect and pattern competition intermittency. Physica D: Nonlinear Phenomena 34, 1–41.

Kaneko, K., 1992. Overview of coupled map lattices. Chaos 2 (3), 279–282.

Kaneko, K., 1993. Theory and Applications of Coupled Map Lattices. John Wiley & Sons., New York, USA.

Kessler, D. A., Levine, H., Reynolds, W. N., Nov 1990. Coupled-map lattice model for crystal growth. Phys. Rev. A 42, 6125–6128.

King, G., 2011. Ensuring the data-rich future of the social sciences. Science 331 (6018), 719–721.

Kokaram, A., Godsill, S., 2002. MCMC for joint noise reduction and missing data treatment in degraded video. IEEE Transactions on Signal Processing 50 (2), 189–205.

Kot, M., 1992. Discrete-time travelling waves: Ecological examples. Journal of Mathematical Biology 30 (4), 413–436.

Kot, M., Lewis, M. A., van den Driessche, P., 1996. Dispersal data and the spread of invading organisms. Ecology 77 (7), 2027–2042.

Kot, M., Schaffer, W. M., 1986. Discrete-time growth-dispersal models. Mathematical Biosciences 80 (1), 109 – 136.

Kragh, K. A., Hansen, M. H., Mikkelsen, T., 2013. Precision and shortcomings of yaw error estimation using spinner-based light detection and ranging. Wind Energy 16 (3), 353–366.

Krim, H., Bao, Y., 1999. A stochastic diffusion approach to signal denoising. In: Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on. Vol. 4. pp. 1773–1776.

Lee, A., Yau, C., Giles, M. B., Doucet, A., Holmes, C. C., 2010. On the utility of graphics cards to perform massively parallel simulation of advanced Monte Carlo methods. Journal of Computational and Graphical Statistics 19 (4), 769–789.

Leonard, N., Paley, D., Lekien, F., Sepulchre, R., Fratantoni, D., Davis, R., 2007. Collective motion, sensor networks, and ocean sampling. Proceedings of the IEEE 95 (1), 48–74.

Levine, H., Reynolds, W., 1992. Coupled map lattice techniques for simulating phenomena in reaction-diffusion systems. Chaos 2 (3), 337–342.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73 (4), 423–498.

Logan, J., 2008. An Introduction to Nonlinear Partial Differential Equations, 2nd Edition. John Wiley & Sons., Hoboken, NJ, USA.

Luenberger, D., 1966. Observers for multivariable systems. IEEE Transactions on Automatic Control 11 (2), 190–197.

Mackay, D., 1998. Introduction to Monte Carlo methods. In: Jordan, M. I. (Ed.), Learning in Graphical Models. Vol. 89 of NATO ASI Series. Springer Netherlands, pp. 175–204.

Mandela, R. K., Rengaswamy, R., Narasimhan, S., Sridhar, L. N., 2010. Recursive state estimation techniques for nonlinear differential algebraic systems. Chemical Engineering Science 65 (16), 4548 – 4556.

Marcos-Nikolaus, P., Martin-González, J. M., Solé, R. V., 2002. Spatial forecasting: Detecting determinism from single snapshots. International Journal of Bifurcation and Chaos 12 (2), 369–376.

Matani, A., Masuda, Y., Okubo, H., Chihara, K., 2003. Hierarchical clustering and filtering in half-inverse space for MEG and/or EEG hypothesis-free analysis. IEEE Transactions on Signal Processing 51 (2), 350–361.

May, R. M., 1976. Simple mathematical models with very complicated dynamics. Nature 261, 459–467.

Maybeck, P., 1979. Stochastic models, estimation and control. Academic Press, New York, USA.

McLachlan, G. J., Krishnan, T., 1997. The EM Algorithm and Extensions. John Wiley & Sons., New York, USA.

Medlock, J., Kot, M., 2003. Spreading disease: integro-differential equations old and new. Mathematical Biosciences 184 (2), 201 – 222.

Mercieca, J., Aram, P., Kadirkamanathan, V., 2015. Unscented Rauch-Tung-Striebel smoothing for nonlinear descriptor systems. In: European Control Conference (ECC). pp. 491–496.

Milito, R., Padilla, R., Padilla, R., Cadorin, D., 1982. An innovations approach to dual control. IEEE Transactions on Automatic Control 27 (1), 132–137.

Misawa, E., Hedrick, J., 1989. Nonlinear observersa state-of-the-art survey. Journal of Dynamic Systems, Measurement, and Control 111 (3), 344–352.

Mitchell, A., Wait, R., 1973. Partial Differential Equations of Mathematical Physics. Elsevier Science.

Mitchell, A., Wait, R., 1977. The Finite Element Methods in Partial Differential Equations. John Wiley & Sons.

Møller, J., Syversveen, A. R., Waagepetersen, R. P., 1998. Log Gaussian Cox processes. Scandinavian Journal of Statistics 25 (3), 451–482.

Moradkhani, H., Sorooshian, S., Gupta, H. V., Houser, P. R., 2005. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. Advances in Water Resources 28 (2), 135–147.

Muske, K. R., Edgar, T. F., 1997. Nonlinear process control. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, Ch. Nonlinear state estimation, pp. 311–370.

Neubert, M., Kot, M., Lewis, M., 1995. Dispersal and pattern formation in a discrete-time predator-prey model. Theoretical Population Biology 48 (1), 7 – 43.

Neubert, M. G., Caswell, H., 2000. Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. Ecology 81 (6), 1613–1628.

Niedzwecki, J., Liagre, P.-Y., 2003. System identification of distributed-parameter marine riser models. Ocean Engineering 30 (11), 1387 – 1415.

Omatu, S., Seinfeld, J., 1989. Distributed parameter systems: Theory and Applications. Clarendon Press, Oxford, UK.

Pan, Y., Billings, S. A., 2008. The identification of complex spatiotemporal patterns using coupled map lattice models. International Journal of Bifurcation and Chaos 18 (04), 997–1013.

Parekh, N., Parthasarathy, S., Sinha, S., 1998. Global and local control of spatiotemporal chaos in coupled map lattices. Phys. Rev. Lett. 81, 1401–1404.

Patwardhan, S. C., Narasimhan, S., Jagadeesan, P., Gopaluni, B., Shah, S. L., 2012. Nonlinear Bayesian state estimation: A review of recent developments. Control Engineering Practice 20 (10), 933–953.

Penny, W., Kiebel, S., Friston, K., 2003. Variational Bayesian inference for fMRI time series. NeuroImage 19 (3), 727–741.

Pfeifer, P. E., Bodily, S. E., 1990. A test of space-time ARMA modelling and forecasting of hotel data. Journal of Forecasting 9 (3), 255–272.

Pfeifer, P. E., Deutrch, S. J., 1980. A three-stage iterative procedure for space-time modeling. Technometrics 22 (1), 35–47.

Pirani, M., Gulliver, J., Fuller, G. W., Blangiardo, M., 2014. Bayesian spatiotemporal modelling for the assessment of short-term exposure to particle pollution in urban areas. Journal of Exposure Science and Environmental Epidemiology 24, 319327.

Prévôt, C., Röckner, M., 2007. A Concise Course on Stochastic Partial Differential Equations. Springer-Verlag, Berlin, Germany.

Price, C., Wambacq, P., Oosterlinck, A., 1992. The plastic coupled map lattice: A novel image-processing paradigm. Chaos 2 (3), 351–366.

Qu, C. C., Hahn, J., 2009. Computation of arrival cost for moving horizon estimation via unscented Kalman filtering. Journal of Process Control 19 (2), 358 – 363.

R. L. Martin, J. E. O., 1975. The identification of regional forecasting models using space: Time correlation functions. Transactions of the Institute of British Geographers (66), 95–118.

Rasmussen, C. E., Williams, C. K. I., 2006. Gaussian Processes for Machine Learning. MIT Press, Cambridge, Massachusetts, USA.

Rawlings, J. B., Bakshi, B. R., 2006. Particle filtering and moving horizon estimation. Computers and Chemical Engineering 30 (1012), 1529 – 1541.

Richter, H., 2008. Coupled map lattices as spatio-temporal fitness functions: Landscape measures and evolutionary optimization. Physica D: Nonlinear Phenomena 237 (2), 167 – 186.

Ristic, B., Arulampalam, S., Gordon, N., 2004. Beyond the Kalman filter - Particle filters for tracking applications. Artech House, Boston, MA, USA.

Robert, C. P., Casella, G., 2004. Monte Carlo Statistical Methods. Springer-Verlag, New York, USA.

Robert, C. P., Casella, G., 2010. Introducing Monte Carlo Methods with R. Springer-Verlag, New York, USA.

Robertson, D. G., Lee, J. H., Rawlings, J. B., 1996. A moving horizon-based approach for least-squares estimation. AIChE Journal 42 (8), 2209–2224.

Romanenko, A., Castro, J. A., 2004. The unscented filter as an alternative to the EKF for nonlinear state estimation: a simulation case study. Computers Chemical Engineering 28 (3), 347 – 355.

Romanenko, A., Santos, L. O., Afonso, P. A. F. N. A., 2004. Unscented Kalman filtering of a simulated ph system. Industrial Engineering Chemistry Research 43 (23), 7531–7538.

Sanguinetti, G., Lawrence, N. D., Rattray, M., 2006. Probabilistic inference of transcription factor concentrations and gene-specific regulatory activities. Bioinformatics 22 (22), 2775–2781.

Särkkä, S., 2013. Bayesian Filtering and Smoothing. Cambridge University Press, Cambridge, UK.

Scerri, K., 2010. A systems approach to spatio-temporal modelling. Ph.D. thesis, University of Sheffield, Sheffield, UK.

Scerri, K., Dewar, M., Kadirkamanathan, V., 2009. Estimation and model selection for an ide-based spatio-temporal model. IEEE Transactions on Signal Processing 57 (2), 482–492.

Schlipf, D., Pao, L., 2014. Preview control of wind turbines. In: Samad, T., Annaswamy, A. M. (Eds.), The Impact of Control Technology, 2nd edition. IEEE Control Systems Society.

Schlipf, D., Schlipf, D. J., Kühn, M., 2013. Nonlinear model predictive control of wind turbines using LIDAR. Wind Energy 16 (7), 1107–1129.

Shen, M., Chang, G., Wang, S., Beadle, P., 2006. Nonlinear dynamics of EEG signal based on coupled network lattice model. In: Wang, J., Yi, Z., Zurada, J., Lu, B.-L., Yin, H. (Eds.), Advances in Neural Networks - ISNN 2006. Vol. 3973 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 560–565.

Shen, M., Lin, L., Chang, G., 2008. Novel coupled map lattice model for prediction of EEG signal. In: Sun, F., Zhang, J., Tan, Y., Cao, J., Yu, W. (Eds.), Advances in Neural Networks - ISNN 2008. Vol. 5263 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 347–356.

Shumway, R. H., Stoffer, D. S., 1982. An approach to time series smoothing and forecasting using the EM algorithm. Journal of Time Series Analysis 3 (4), 253–264.

Simon, D., 2006. Optimal state estimation - Kalman, H_∞ and nonlinear approaches. John Wiley & Sons., Hoboken, NJ, USA.

Smith, A. C., Brown, E. N., 2003. Estimating a state-space model from point process observations. Neural Computation 15 (5), 965–991.

Smith, G., 1969. Numerical Solution of Partial Differential Equations. Oxford University Press, Oxford, UK.

Smith, M., 1967. Introduction to the Theory of Partial Differential Equations. Van Nostrand Company.

Söderström, T., 2003. Discrete-time stochastic system. Advanced textbooks in control and signal processing. Springer-Verlag.

Sole, R., Bascompte, J., Valls, J., 1992. Nonequilibrium dynamics in lattice ecosystems: Chaotic stability and dissipative structures. Chaos 2 (3), 387–395.

Soleimanzadeh, M., Wisniewski, R., Brand, A., 2014. State-space representation of the wind flow model in wind farms. Wind Energy 17 (4), 627–639.

Solo, V., 2002. Identification of a noisy stochastic heat equation with the EM algorithm. In: Proceedings of the 41st IEEE Conference on Decision and Control. Vol. 4. pp. 4505–4508.

Sorenson, H., 1970. Least-squares estimation: from Gauss to Kalman. IEEE Spectrum 7 (7), 63–68.

Sorenson, H., 1985. Kalman Filtering: Theory and applications. IEEE Press, New York, USA.

Soroush, M., 1998. State and parameter estimations and their applications in process control. Computers and Chemical Engineering 23 (2), 229 – 245.

Storvik, G., Frigessi, A., Hirst, D., 2002. Stationary space-time Gaussian fields and their time autoregressive representation. Statistical Modelling 2 (2), 139–161.

Stroud, J. R., Müller, P., Sansó, B., 2001. Dynamic models for spatiotemporal data. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (4), 673–689.

Stull, R. B., 1988. An Introduction to Boundary Layer Meteorology. Vol. 13. Kluwer Academic Publishers, Dordrecht, Netherlands.

Suchard, M. A., Wang, Q., Chan, C., Frelinger, J., Cron, A., West, M., 2010. Understanding gpu programming for statistical computation: studies in massively parallel massive mixtures. Journal of Computational and Graphical Statistics 19 (2), 419–438.

Tenny, M., Rawlings, J., 2002. Efficient moving horizon estimation and nonlinear model predictive control. In: Proceedings of the 2002 American Control Conference. Vol. 6. pp. 4475–4480.

Towers, P., Jones, B. L., 2016. Real-time wind field reconstruction from LiDAR measurements using a dynamic wind model and state estimation. Wind Energy 19 (1), 133–150.

Travis, C., White, L., 1985. Parameter identification of distributed parameter systems. Mathematical Biosciences 77 (12), 341 – 352.

Tricaud, C., Dariusz, M. P., Yang, U., Chen, Q., 2008. D-optimal trajectory design of heterogeneous mobile sensors for parameter estimation of distributed systems. In: 2008 American Control Conference. pp. 663–668.

Uciński, D., 2005. Optimal Measurement Methods for Distributed Parameter System Identification. CRC Press, Boca Raton, Florida, USA.

Unny, T., 1989. Stochastic partial differential equations in groundwater hydrology. Stochastic Hydrology and Hydraulics 3 (2), 135–153.

Vachhani, P., Narasimhan, S., Rengaswamy, R., 2006. Robust and reliable estimation via unscented recursive nonlinear dynamic data reconciliation. Journal of Process Control 16 (10), 1075 – 1086.

Veit, R. R., Lewis, M. A., 1996. Dispersal, population growth, and the allee effect: Dynamics of the house finch invasion of eastern north america. The American Naturalist 148 (2), 255–274.

Vermaak, J., Lawrence, N., Perez, P., 2003. Variational inference for visual tracking. In: Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. Vol. 1. pp. I–773–I–780 vol.1.

Voss, H., Bünner, M. J., Abel, M., Mar 1998. Identification of continuous, spatiotemporal systems. Phys. Rev. E 57, 2820–2823.

Šmídl, V., Quinn, A., 2005. The Variational Bayes Method in Signal Processing. Springer-Verlag, New York, USA.

Šmídl, V., Quinn, A., 2008. Variational Bayesian filtering. IEEE Transactions on Signal Processing 56 (10), 5020–5030.

Walsh, J. B., 1981. A stochastic model of neural response. Advances in Applied Probability 13 (2), pp. 231–281.

Wan, E., Van der Merwe, R., 2000. The unscented Kalman filter for nonlinear estimation. In: The IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium. pp. 153–158.

Wang, M.-H., Kot, M., 2001. Speeds of invasion in a model with strong or weak allee effects. Mathematical Biosciences 171 (1), 83 – 97.

Wang, N., Johnson, K. E., Wright, A. D., 2012. FX-RLS-based feedforward control for LIDAR-enabled wind turbine load mitigation. IEEE Transactions on Control Systems Technology 20 (5), 1212–1222.

Weickert, J., 1997. Applications of the theory of differential-algebraic equations to partial differential equations of fluid dynamics. Ph.D. thesis, Technische Universität Chemnitz-Zwickau, Chemnitz, Germany.

Wikle, C., Cressie, N., 1999. A dimension-reduced approach to space-time Kalman filtering. Biometrika 86 (4), 815–829.

Wikle, C., Hooten, M., 2005. Hierarchical Bayesian Spatio-Temporal Models for Population Spread. Oxford University Press, Oxford, UK.

Wikle, C. K., 2002. A kernel-based spectral model for non-Gaussian spatio-temporal processes. Statistical Modelling 2 (4), 299–314.

Wood, R., 2004. Impact of advanced aerodynamic technology on transportation energy consumption. In: SAE Technical Paper.

Xu, K., Wikle, C. K., Fox, N. I., 2005. A kernel-based spatio-temporal dynamical model for nowcasting weather radar reflectivities. Journal of the American Statistical Association 100 (472), 1133–1144.

Yanagita, T., 1992. Phenomenology of boiling: A coupled map lattice model. Chaos 2 (3), 343–350.

Yanagita, T., Kaneko, K., 1993. Coupled map lattice model for convection. Physics Letters A 175 (6), 415 – 420.

Yanagita, T., Kaneko, K., 1995. Rayleigh-Bénard convection patterns, chaos, spatiotemporal chaos and turbulence. Physica D: Nonlinear Phenomena 82 (3), 288 – 313.

Yanagita, T., Kaneko, K., Jun 1997. Modeling and characterization of cloud dynamics. Phys. Rev. Lett. 78, 4297–4300.

Yuan, K., Niranjan, M., 2010. Estimating a state-space model from point process observations: A note on convergence. Neural Computation 22 (8), 1993–2001.

Zammit-Mangion, A., 2011. Modelling from spatiotemporal data: A dynamic systems approach. Ph.D. thesis, University of Sheffield, Sheffield, UK.

Zammit-Mangion, A., Dewar, M., Kadirkamanathan, V., Sanguinetti, G., 2012a. Point process modelling of the Afghan war diary. Proceedings of the National Academy of Sciences 109 (31), 12414–12419.

Zammit-Mangion, A., Sanguinetti, G., Kadirkamanathan, V., 2011. A variational approach for the online dual estimation of spatiotemporal systems governed by the ide. In: Proceedings of the 18th IFAC World Congress. pp. 3204–3209.

Zammit-Mangion, A., Sanguinetti, G., Kadirkamanathan, V., July 2012b. Variational estimation in spatiotemporal systems from continuous and point-process observations. IEEE Transactions on Signal Processing 60 (7), 3449–3459.