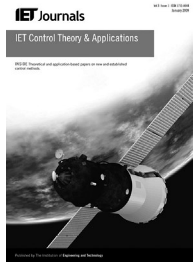
www.ietdl.org

Published in IET Control Theory and Applications
Received on 20th February 2013
Revised on 1st August 2013
Accepted on 11th August 2013
doi: 10.1049/iet-cta.2013.0150



ISSN 1751-8644

Identification of non-linear stochastic spatiotemporal dynamical systems

Hanwen Ning[1,2], *Xingjian Jing*[1], *Li Cheng*[1]
[1]*Department of Mechanical Engineering, Hong Kong Polytechnic University, HungHom, Kowloon, HongKong*
[2]*School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan, People's Republic of China*
E-mail: xingjian.jing@googlemail.com

Abstract: A systematic identification method for non-linear stochastic spatiotemproal (SST) systems described by non-linear stochastic partial differential equations (SPDEs) is investigated in this study based on pointwise observation data. A theoretical framework for a semi-finite element model approximating to an infinite-dimensional system is established, and several fundamental issues are discussed including the approximation error between the underlying infinite-dimensional dynamics and the model to be identified, and its rationality etc. Based on the proposed theoretical framework, a general identification method with irregular observation data is provided. These results not only provide an effective method for the identification of non-linear SST systems using measurement data (both offline and online), but also demonstrate a potential solution for the analysis, design and control of non-linear SST systems from a numerical point of view.

1 Introduction

The stochastic partial differential equation (SPDE) theory was initiated by the study of concrete model equations in physics and engineering and evolved gradually into a new branch of modern mathematics [1–3]. Different SPDEs have been proposed and successfully applied to describe the linear and non-linear spatiotemporal physical phenomena in different areas, especially in finance analysis, fluid dynamics, field quantum theory and statistical mechanics etc [4–6]. The solution of an SPDE is usually established in the framework of infinite dimensional abstract spaces, and the average qualitative properties in Banach spaces of the solutions such as existence, stability and asymptotic behaviours etc can be studied using stochastic analysis theory and generalised functional theory etc [5–7]. Some efficient numerical methods can also be employed to find an approximation of the solution of a SPDE system, including stochastic difference methods and stochastic finite element methods [8–10]. It should be noted that the approximation errors with the difference methods strongly rely on the differential form of the SPDE and can only be investigated case by case with respect to a specific SPDE. However, the analysis and approximation errors with the finite element methods can be more convenient to implement in the framework of infinite dimensional spaces, since the numerical stochastic finite element method itself is established with the idea of using finite dimensional basis spaces to approximate infinite dimensional spaces [11–14].

Nonlinear system identification is to build a mathematical model with measurement data for non-linear dynamical systems, and many methods have been proposed such as set membership, neural network, orthogonal least square etc [15–18]. Among those, the kernel learning method is a very promising one, which could realise the best estimation with least number of samples [19, 20]. For identification of deterministic spatiotemporal dynamical systems given by PDEs [21–26], continuous spatiotemporal dynamical systems can be transformed into coupled lattice dynamical systems, in which the variables with respect to each node represent the same set of physical quantities and orthogonal forward regression algorithms are utilised to obtain the numerical relationship of these variables. For non-linear spatiotemporal systems given by deterministic PDEs in variational formulation, orthogonal least squares algorithms and finite dimensional approximation methods are also exploited with experimental measurement data to identify a non-linear model which can reproduce the underlying dynamic phenomena directly [27]. On the one hand, although these results provide an important insight into the modelling, analysis and control of spatiotemporal dynamical systems, few results are noticed in the literature to address the identification problems of non-linear stochastic spatiotemproal (SST) systems. As discussed before, a non-linear SPDE or SST model would be more preferable in practice since it can effectively and reasonably take noise process into consideration. However, the existing identification methods for deterministic PDE systems may not be straightforward to be generalised to the stochastic cases. Some fundamental theoretical issues related to the numerical approximation of an SPDE should be addressed before potential identification methods can be reasonably proposed and applied. On the other hand, in conventional identification methods for spatiotemporal systems [21–24], it is usually required that the observation locations are uniformly distributed on the spatiotemporal plant. Obviously, this is an undesired restriction, which may lead the conventional methods to be infeasible or inapplicable in some applications, since sensors are not always able to be placed uniformly. Moreover, to the best of the authors' knowledge, there is still a few results available for the identification of spatiotemporal systems with irregular observation locations at present.

The contribution of this paper is in two aspects. The first one is that theoretical results are developed to address some fundamental problems related to the approximation errors between the estimated semi-finite element model and the original infinite-dimensional SPDE dynamic system. Some techniques from inverse semi stochastic finite element methods and stochastic approximation methods are employed for this purpose. The Galerkin approximation method is utilised to construct a finite-dimensional stochastic system, which is used to approximate the non-linear infinite system under study. Based on these theoretical results and techniques, a systematic identification framework for non-linear SST systems using available measurement data from both theoretical and practical viewpoints is thus developed. It is shown that the numerical relationship of the basis coefficients of the finite element model can be transformed into a regression equation of multi-input-multi-output (MIMO) partially linear model (PLM) instead of commonly used non-linear autoregressive network with exogenous inputs (NARX) one, which greatly facilitates the implementation of the proposed identification method. The second contribution of this paper is to provide a general identification method for non-linear spatiotemporal systems with irregular observation locations. Necessary discussions and numerical examples are given to illustrate the effectiveness of the proposed method and to point out potential applications of our results.

The rest of the paper is organised as follows. The stochastic evolution system for SPDEs is introduced in Section 2. The convergence and estimation theorems of finite-dimensional approximation models in semi-discretised and fully discretised cases are formulated in Section 3.1. The identification approach and the underlying idea that governs the identification process based on inverse semi-finite element method and stochastic analysis are discussed in Sections 3.2 and 3.3 introduces kernel based partially linear least-square regularised regression (PL-LSRR) algorithm for identification of SST systems. The identification method with irregular observation data is given in Sections 3.4 and 3.5 provides some discussion about the advantages and limitations of the proposed method. Examples are provided in Section 4 and some conclusions are drawn thereafter in Section 5.

2 Stochastic evolution equations

Let H be a separable Hilbert space. Denote by $|\cdot|$ the norm in H, by $(\cdot,\cdot)$ the scalar product in H. $\|\cdot\|$ is also used to denoted the norm in ordinary Banach space. Let $(\Omega, \Im, \Im_t, P)$ be a complete probability space with a filtration $\{\Im_t\}$ satisfying the usual condition (i.e. the filtration contains all P-null sets and is right continuous). $\mathbf{E}$ is used to denote the mathematical expectation with respect to the probability space. $W(t)$ is H valued, Q-Wiener process introduced in [3, 8, 12], with eigenvalues $\gamma_l > 0$ and corresponding eigenfunctions ξ_l. $\mathcal{B}_l, l = 1, 2, \ldots$, is a sequence of real valued independently and identically distributed Brownian motions. Then $W(t) = \sum_{l=1}^{\infty} \gamma_l^{1/2} \xi_l \mathcal{B}_l(t)$.

Consider the following non-linear stochastic evolution equation

$$\begin{cases} \mathrm{d}u(t) + Au(t)\,\mathrm{d}t = f(u(t))\,\mathrm{d}t + \mathrm{d}W(t), & t \in [0, T] \\ u(0) = u_0 \end{cases} \tag{1}$$

in Hilbert space H, where $u(t)$ is an H valued stochastic process, A is a linear, self-adjoin, compact, positive-definite operator densely defined on the corresponding definition domain $D(A) \subset H$. We also have the following basic assumptions

(I) $A : D(A) \subset H \to H$ is the infinitesimal generator of a strongly continuous analytic semi-group of bounded linear operators $E(t), t \geq 0$ on H. The initial data u_0 is an $\Im_0$ measurable H valued stochastic variable.

(II) f is a non-linear mapping from H to H satisfying the global Lipschitz condition and linear growth condition. There exists a constant L_0 such that

$$|f(x) - f(y)| \leq L_0|x - y|, \quad |f(x)| \leq L_0|x|, \quad \forall x,\ y \in H$$

Definition 1: An H valued process $u(t)$ is said to be a weak solution of system (1). Let A^* is the adjoint operator of A and defined on the corresponding definition domain $D(A^*)$, if $u(t)$ has Bochner integrable trajectories P-almost surely ($P - a.s.$) and for $\forall v \in D(A^*)$ and $\forall t \geq 0$

$$\begin{aligned}(u(t), v) = (u_0, v) + \int_0^t (u(s), A^* v)\,\mathrm{d}s + \int_0^t (f(u(s)), v)\,\mathrm{d}s \\ + \int_0^t (\mathrm{d}W(s), v)\,\mathrm{d}s, \quad P - a.s.\end{aligned} \tag{2}$$

Definition 2: An H valued $\Im_t$ measurable process $u(t)$ is said to be a mild solution of system (1), if $\mathbf{E}|u(t)|^2 < \infty$ and $u(t)$ satisfies the following equation

$$\begin{aligned}u(t) = E(t)u_0 + \int_0^t E(t - s) f(u(s))\,\mathrm{d}s \\ + \int_0^t E(t - s)\,\mathrm{d}W(s), \quad P - a.s.\end{aligned} \tag{3}$$

The phase space H studied here is an infinite-dimensional Hilbert space, which usually represent some function space on spatial domain $\mathcal{D} \subset R^d$ with boundary $\partial\mathcal{D}$ where the physical phenomena are generated. The stochastic evolution equation in (1) actually represents a large class of non-linear stochastic systems described by stochastic heat equations, stochastic wave equations, stochastic reaction–diffusion equations with appropriate boundary conditions (e.g. Dirichlet or Neumann boundary conditions) [14, 28, 29]. The existence and uniqueness of the solution for system (1) can be derived by the standard techniques in [3] with the assumptions presented above. In the following parts of this section, it is assumed that $\mathcal{D} = [0, 1]$ and $H = L^2(\mathcal{D})$ and A is a second-order differential operator on $D(A) = H^2(\mathcal{D}) \cap H_1^0(\mathcal{D})$ [14, 28]. The method to be developed could be easily extended to the case of multi-dimensional spatial domain. The corresponding deterministic system of (1) can be written as

$$\begin{cases} \mathrm{d}u(t)/\mathrm{d}t + Au(t) = f(u(t)), & t \in [0, T] \\ u(0) = u_0 \end{cases} \tag{4}$$

For system (1), we firstly give a result about the regularity of the solution. For convenience, C is used to represent any constants stemming from our computations.

doi: 10.1049/iet-cta.2013.0150

Lemma 1: Let u be the mild solution of system (1), then for $\forall 0 \le t_1 \le t_2 \le T$, there exists a constant $C = C(u_0, T)$, such that

$$\mathbf{E}|u(t_2) - u(t_1)|^2 \le C(t_2 - t_1) \tag{5}$$

Proof: The proof is straightforward by following Definition 2, and using Lemma 2 and Holder inequality etc., which is thus omitted for paper limitation. □

3 Identification method

The proposed identification method will be discussed in four steps. First, we establish a general identification framework for SPDE systems and propose a general finite-dimensional approximation regression model (Sections 3.1 and 3.2). Second, the deterministic part of the regression model for the deterministic part of the equation is estimated by using a kernel-based learning approach PLM–LSRR (Section 3.3). Third, a general identification method is proposed for identification problems with irregular observation data (Section 3.4). Finally, some necessary discussions are given for the established stochastic parametric MIMO–PLM-based model (Section 3.5).

3.1 Approximation errors

In this section, a finite-dimensional approximation model is proposed to approximate the infinite-dimensional stochastic dynamical system (1), and some fundamental results regarding the approximation error and convergence of the finite-dimensional approximation model are provided.

Let $\{\mathcal{T}_h\}_{0<h<1}$ be a regular family of triangulations on spatial domain $\mathcal{D}$ with the maximal edge length h. Let S_h be a family of finite-element spaces consisting of piecewise continuous polynomials. The semi-discrete model of system (1) is presented as follows

$$\begin{cases} \mathrm{d}u_h(t) + A_h u_h(t)\,\mathrm{d}t = P_h f(u_h(t))\,\mathrm{d}t + P_h\,\mathrm{d}W(t), \quad t \in [0, T] \\ u_h(0) = P_h u_0 \end{cases} \tag{6}$$

where $A_h : S_h \to S_h$ is the discrete analogue of A defined by $(A_h\eta_1, \eta_2) = A(\eta_1, \eta_2), \forall \eta_1, \eta_2 \in S_h$. Here $A(\cdot, \cdot)$ is bilinear form obtained from the differential operator A. P_h denotes the orthogonal projection from H onto S_h defined by $(P_h\xi_1, \xi_2) = (\xi_1, \xi_2), \forall \xi_1 \in H, \forall \xi_2 \in S_h$. Let $E_h(t), t \ge 0$ be the analytic semi-group on S_h. The mild solution of the approximation system above is given by

$$u_h(t) = E_h(t)P_h\xi + \int_0^t E_h(t-s)P_h f(u_h(s))\,\mathrm{d}s + \int_0^t E_h(t-s)P_h\,\mathrm{d}W(s), \quad P - a.s. \tag{7}$$

For any $s \in R$, let $\dot{H}^s = \dot{H}^s(\mathcal{D}) = D(A^{s/2})$ (see [14]) with norm $\|v\|_s = \|A^{s/2}v\|$. The following two lemmas present fundamental and useful properties of the analytic semi-group mentioned above.

Lemma 2 [14]: For $\|\zeta\|_{\dot{H}^2} < \infty$ and $\forall t \ge 0$, there exists a constant C such that

$$|E_h(t)P_h\zeta - E(t)\zeta| \le Ch^2\|\zeta\|_{\dot{H}^2}, \quad \text{and}$$
$$\int_0^t \|E_h(t-s)P_h - E(t-s)\|^2\,\mathrm{d}s \le Ch^2 \tag{8a}$$

The objective is to establish a semi-finite element approximation model to the stochastic spatiotemporal dynamical system in (1) using observation data. The dynamical behaviours of the semi-finite element approximation model to be identified could be different from the original SPDE, and the approximation error would also be closely affected by the number of dimensions of finite-element spaces and the length of time step. For these reasons, two theorems describing the approximation errors in mean-square manner are provided. The first one addresses the approximation error of the semi-discrete model in (6).

Theorem 1: If $E\|u_0\|^2_{\dot{H}^2} < \infty$, then for $\forall t \in [0, T]$, there exists a constant $C = C(u_0, T)$ such that

$$\mathbf{E}|u_h(t) - u(t)|^2 \le Ch^2 \tag{8b}$$

Proof: See Appendix 1. □

The following results provide an estimation of the approximation error in the fully discrete case. That is, not only the finite-element approximation is used with respect to the spatial variable, but also the difference method is used with respect to the time variable. Let Δt be the time step and $t_n = n\Delta t$ with $n \ge 1$. The backward Euler method for (6) is used here to obtain the approximation variable in semi finite-element model

$$\frac{u_h(t_n) - u_h(t_{n-1})}{\Delta t} + A_h u_h(t_n) = \frac{1}{\Delta t}\int_{t_{n-1}}^{t_n} P_h f(u_h(t_{n-1}))\,\mathrm{d}s + \frac{1}{\Delta t}\int_{t_{n-1}}^{t_n} P_h\,\mathrm{d}W(s), \quad n \ge 1$$
$$u_h(0) = P_h u_0 \tag{9}$$

Let $r(\lambda) = 1/(1+\lambda)$, (9) can be rewritten as a equivalent form as follow

$$u_h(t_n) = r(\Delta t A_h)u_h(t_{n-1}) + \int_{t_{n-1}}^{t_n} r(\Delta t A_h)P_h f(u_h(t_{n-1}))\,\mathrm{d}s + \int_{t_{n-1}}^{t_n} r(\Delta t A_h)P_h\,\mathrm{d}(s), \quad n \ge 1$$
$$u_h(0) = P_h u_0 \tag{10}$$

Here (10) is the discrete approximation system of the finite-element system (6), and then u_h is the approximate solution of the finite-element system (6). Therefore the solution obtained by iteration from (10) is also reasonably used to approximate $u(t_n)$. The main results regarding the approximation error are presented as follows.

Theorem 2: Let u and U^n be the solutions of system (1) and (10), respectively. If $E\|u_0\|^2_{\dot{H}^2} < \infty$, then there exits a constant $C = C(T, u_0)$, such that

$$\mathbf{E}|u_h(t_n) - u(t_n)|^2 \le C(\Delta t + h^2) \tag{11}$$

Proof: See Appendix 2. □

It can be concluded from Theorems 1 and 2 that the semi-finite element model converges to the original SPDE, and the approximation error are $O(h^2)$ and $O(\Delta t)$ with respect to spatial variable and time variable, respectively.

3.2 Identification framework

For the stochastic evolution system (1), define a partition of $[0, T] \times [0, 1]$ by rectangles $[t_k, t_{k+1}] \times [x_j, x_{j+1}]$ for $k = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, N$, $t_k = (k-1)\Delta t$, $x_j = (j-1)\Delta x$, $\Delta t = T/M$, and $h = \Delta x = 1/N$. Here, M is the number of samples. From the definition of the weak solution in (2), an equivalent finite-element variational formulation of system (1) is given as follows. For $\forall\phi \in S_h$

$$(\phi, u_h(t)) = (\phi, P_h\xi) + \int_0^t (A_h^*\phi, u_h(s))\,\mathrm{d}s + \int_0^t (\phi, P_hf(u_h(s))\,\mathrm{d}s) + \int_0^t (\phi, P_h\,\mathrm{d}W(s)) \tag{12a}$$

which implies

$$(\phi, u_h(t_k)) = (\phi, P_h\xi) + \int_0^{t_k} (A_h^*\phi, u_h(s))\,\mathrm{d}s + \int_0^{t_k} (\phi, P_hf(u_h(s))\,\mathrm{d}s) + \int_0^{t_k} (\phi, P_h\,\mathrm{d}(s)) \tag{12b}$$

$$(\phi, u_h(t_{k+1})) = (\phi, P_h\xi) + \int_0^{t_{k+1}} (A_h^*\phi, u_h(s))\,\mathrm{d}s + \int_0^{t_{k+1}} (\phi, P_hf(u_h(s))\,\mathrm{d}s) + \int_0^{t_{k+1}} (\phi, P_h\,\mathrm{d}(s)) \tag{12c}$$

and then, for the given time partition, we have

$$(\phi, u_h(t_{k+1}) - u_h(t_k)) = \int_{t_k}^{t_{k+1}} (A_h^*\phi, u_h(s))\,\mathrm{d}s + \int_{t_k}^{t_{k+1}} (\phi, P_hf(u_h(s))\,\mathrm{d}s) + \int_{t_k}^{t_{k+1}} (\phi, P_h\,\mathrm{d}(s)) \tag{13}$$

Then, a difference method is used with respect to the time variable

$$(\phi, u_h(t_{k+1})) = (\phi, u_h(t_k)) + \Delta t(A_h^*\phi, u_h(t_k)) + \Delta t(\phi, P_hf(u_h(t_k))) + \int_{t_k}^{t_{k+1}} (\phi, P_h\,\mathrm{d}(s)) \tag{14}$$

To find the approximation solution of system (1) in the finite element space. Let

$$u_h(t, x) = \sum_{l=1}^{N_h} u_l(t)\phi_l \tag{15}$$

where N_h is the dimensional number of finite element basis, $\phi_l(x)$ is the given finite-element bases used for approximation such as B-spline function basis [30] and $u_l(t)'s$ are the coefficients of the finite-element basis. Denoting P_hf by $f_h : S_h \to S_h$, and substituting (15) into (14), we have

$$\sum_{l=1}^{N_h} u_l(t_{k+1})(\phi, \phi_l) = \sum_{l=1}^{N_h} u_l(t_k)((\phi, \phi_l) + \Delta t(A_h^*\phi, \phi_l)) + \Delta t\left(\phi, f_h\left(\sum_{l=1}^{N_h} u_l(t_i)\phi_l\right)\right) + (\phi, P_h(W(t_{k+1}) - W(t_k))) \tag{16a}$$

and then

$$\sum_{l=1}^{N_h} (u_l(t_{k+1}) - u_l(t_k))(\phi, \phi_l) = \sum_{l=1}^{N_h} u_l(t_k)\Delta t(A_h^*\phi, \phi_l) + \Delta t\left(\phi, f_h\left(\sum_{l=1}^{N_h} u_l(t_k)\phi_l\right)\right) + (\phi, P_h(W(t_{k+1}) - W(t_k))) \tag{16b}$$

Let $\phi = \phi_l$, $l = 1, 2, \ldots, N_h$. Note that $F_0(U(t_k)) = (f_1(U(t_k)), f_2(U(t_k)), \ldots, f_{N_h}(U(t_k)))^{\mathrm{T}}$, $A_0 = [a_{ij}]$, $B = [b_{ij}]$ and $E_0(t_k) = e_1^\star(t_k), e_2^\star(t_k), \ldots, e_{N_h}^\star(t_k))$, where $a_{ij} = \Delta t(A_h^*\phi_j, \phi_i)$, $b_{ij} = (\phi_j, \phi_i)$, $U(t_k) = (u_1(t_k), u_2(t_k), \ldots, u_{N_h}(t_k))^{\mathrm{T}}$, $f_i(U(t_k)) = \Delta t(\phi_i, P_hf(\sum_{l=1}^{N_h} u_l(t_k)\phi_l))$ and $e_i^\star(t_k) = (\phi_i, P_h(W(t_{k+1}) - W(t_k)))$ for $1 \le i, j \le N_h$. Moreover, it is obvious that A_0 is a symmetrical, N_h dimensional, definite matrix. Then, based on (16b), we have

$$B(U(t_{k+1}) - U(t_k)) = A_0U(t_k) + F_0(U(t_k)) + E_0(t_k) \tag{17a}$$

which implies

$$U(t_{k+1}) = U(t_k) + B^{-1}A_0U(t_k) + B^{-1}F_0(U(t_k)) + B^{-1}E_0(t_k) \tag{17b}$$

The random term can be approximated as follows [11]

$$\int_{t_k}^{t_{k+1}} (\phi, P_h\,\mathrm{d}(s)) = (\phi, P_h(W(t_{k+1}) - W(t_k))) = \left(\sum_{l=1}^{N_h} \alpha_j^{1/2}\phi_l(\mathcal{B}_j(t_{k+1}) - \mathcal{B}_j(t_k)), \phi\right)$$

It can be seen that the random term above is normally distributed for the characteristics of $\mathcal{B}_j$. α_j can be computed by solving the following equation $\sum_{l=1}^{N_h} \alpha_l(\phi_l, \psi)^2 = \sum_{l=1}^{N_h} \gamma_l(\xi_l, \psi)^2$, $\forall\psi \in S_h$.

In this paper, first, it is assumed that the data in the spatial domain are sampled at the locations that are uniformly distributed on $\mathcal{D} = [0, 1]$, that is the data are spatially sampled at the points $(1/N), (2/N), \ldots, (N - 1/N)$ and thereafter $N_h = N$. Similarly, in the time domain, the data are uniformly sampled over the interval $[0, T]$ of observation with a sampling period Δt.

Based on (16), let $\phi = \phi_l$, $l = 1, 2, \ldots, N_h$. Then, a multi-dimensional non-linear discrete difference equation is obtained, and the finite-element basis coefficients $u_l(t_i)$ could be computed in an iterative manner. It can be seen that if the finite element basis ϕ_l, the spatial step Δx and the time step Δt are given, the corresponding finite-element basis coefficients $u_l(t_i)$ are determined by a resulting regression equation as (17b). In the finite-element approximation

IET Control Theory Appl., pp. 1–15
doi: 10.1049/iet-cta.2013.0150

theory for SPDEs, $u_l(t_i)'s$ are usually considered as the approximation value of $u(t_i, x_l)'s$, which provides a possibility to investigate the stochastic dynamical system from a system identification perspective. That is, we could estimate the approximation model for the stochastic system directly from measurement data. Note that the matrix B is known for given finite-element basis, to use this prior knowledge for identification, the regression model (17b) can be rewritten as follows

$$Y(t_k) = A_0 U(t_k) + F_0(U(t_k)) + E_0(t_k) \tag{18}$$

where $Y(t_k) = B(U(t_{k+1}) - U(t_k))$. From (17) and (18), it is clear that even if the specific form of the SPDE is completely unknown, the numerical relationship of the finite-element coefficients can still be described by the following MIMO–PLM

$$Y_l(t_k) = \beta_l^{\mathrm{T}} U(t_k) + F_l(U(t_k)) + e_l(t_k), \quad \text{for } l = 1, \ldots, N_h \tag{19}$$

where Y_l, is the corresponding lth term of Y, β_l^{T}, $F_l's$ and $e_l's$ are corresponding coefficient vector of the linear part, unknown non-linear mappings and noise terms respectively for $N_h = N$ dimensional system that is to be determined directly from the discrete observation data of $u(t_i, x_l)$. For more details of the PLM, it can be referred to the references [18, 19]. Then, it can be seen that (17b) could be used as a predictor for the finite-element coefficients as the model (18) and (19) are identified by using the observations.

B-spline finite-element basis is used here, and let $\{\phi_l\}_{l=0}^{N}$ be the standard l_0th order B-spline basis. If the model (19) is estimated from the observation data, the approximation semi-finite element model for system (1) can be constructed by two parts. The first part is to estimate the regression model (19) for the finite-element basis coefficients. The second part is the interpolation, in which the cubic (three order) B-spline interpolation and linear B-spline interpolation are used in the spatial domain and time domain, respectively.

$$\hat{u}(t, x) = \sum_{i=1}^{M} \sum_{l=1}^{N} \hat{u}_h(t_i, x_l) \Phi_{i,l}^{1,3}(t, x) \tag{20}$$

where $\Phi^{1,3} = \phi_{M,1}(t) \otimes \phi_{N,3}(x)$ is two-dimensional tensor splines and $\hat{u}_h(t_i, x_l)$ is obtained by (21). $\hat{u}(t, x)$ is the estimate for $u(t, x)$ using the identified semi-finite element model. It should be noted that $\hat{u}_h(t_i, x_l)$ is in fact a random variable, and the approximation error of the semi-finite element model to be identified in mean-square manner is described by Theorems 1 and 2.

3.3 Kernel learning method and the modelling of the system

To investigate the identification problem of non-linear spatiotemporal system approximated by (19) and not to lose the generality, a general input vector that consists of all possible variables could be constructed as $z(k) = [U^{\mathrm{T}}(k), \ldots, U^{\mathrm{T}}(k - n_U + 1)]^{\mathrm{T}}$, where $U(k) = [u_1(t_k), \ldots, u_N(t_k)]^{\mathrm{T}}$. Then, for each channel

$$Y_l(t) = \beta_l^{\star T} z(t) + F_l^{\star}(z(t)) + e_l(t), \quad \text{for } l = 1, \ldots, N_h \tag{21a}$$

where $\beta_l^{\star T}$'s and $F_l^{\star}$'s are coefficients of the linear part and unknown non-linear functions to be estimated. This leads to a MIMO–PLM model, which includes all the possible combinations of regression terms. It is shown in [31] that it is possible to use a PLM with particular model structure to successfully identify a non-linear system containing a linear part and a non-linear component with potentially better performance by using the corresponding kernel-based learning algorithm (PL-LSRR). For identifying the SST system of this study, the PL-LSRR [31] is adopted here.

To introduce PL-LSRR, we give the definition of the reproducing kernel Hilbert space (RKHS). Let $K : Z \times Z \to R$ be continuous, symmetric and positive semi-definite, where Z is a vector space. Such a function is called a Mercer Kernel. The RKHS H_K associated with the kernel K is defined to be the closure of the linear span of the set of functions $\{K_z := K(z, \cdot), z \in Z\}$ with inner product $\langle \cdot, \cdot \rangle_{H_K} = \langle \cdot, \cdot \rangle_K$ satisfying $\langle K_z, K_{z'} \rangle_K = K(z, z')$. If the kernel function is chosen as a linear kernel: $K_{\mathrm{lin}}(z, z') = z^{\mathrm{T}} z'$, The RKHS is the space of linear functions. If the kernel function is chosen as a Gaussian radial basis function (RBF) kernel: $K_{\mathrm{RBF}}(z, z') = \exp(-\|z - z'\|_2^2 / \sigma^2)$, the RKHS is a space of non-linear functions and has good approximation performance for smooth non-linear functions. For model (21) and given sample set $\{Y_l(t), z(t)\}_{t=1}^{M}$, where M is the number of samples. For given $l \in \{1, 2, \ldots, N_h\}$, we note

$$Y_l(t) = g_l(z(t)) = \beta_l^{\star T} z(t) + F_l^{\star}(z(t)) + e_l(t) \tag{21b}$$

$\hat{g}_l = g_{l,1} + g_{l,2}$ as the model for the estimation of the regression function g, where $g_{l,1}$ and $g_{l,2}$ are used to approximate the linear and non-linear part of the PLM, respectively, and they are obtained by solving the following PL-LSRR optimisation problem in two RKHSs

$$\min_{g_{l,1} \in H_{K_{\mathrm{lin}}}, g_{l,2} \in H_{K_{\mathrm{RBF}}}} \frac{1}{M} \sum_{t=1}^{M} \Big\{ (Y_l(t) - g_{l,1}(z(t)) - g_{l,2}(z(t)))^2 + \nu_1 \|g_{l,1}\|_{K_{\mathrm{lin}}}^2 + \nu_2 \|g_{l,2}\|_{K_{\mathrm{RBF}}}^2 \Big\} \tag{22}$$

where $\|g_{l,1}\|_{K_{\mathrm{lin}}}^2 = \langle g_{l,1}, g_{l,1} \rangle_{K_{\mathrm{lin}}}, \|g_{l,2}\|_{K_{\mathrm{RBF}}}^2 = \langle g_{l,2}, g_{l,2} \rangle_{K_{\mathrm{RBF}}}$, and ν_1, ν_2 are regularisation constants. The PL-LSRR uses the sum of two functions (one is in $H_{K_{\mathrm{lin}}}$ and the other is in $H_{K_{\mathrm{RBF}}}$) to approximate the regression function. The regularisation constants ν_1, ν_2 are used to make a trade-off between the empirical prediction errors and the complexity of the model. The ratio $\nu = \nu_1/\nu_2$ is important for how much linear function component should appear in the predicting function. When ν is small, the non-linear function component is restrained more, and more linear function component appears in the predicting function. The solution of (22) can be obtained as follows

$$g_{l,1}(z) = \sum_{t=1}^{M} a_t K_{\mathrm{lin}}(z, z(t)), \quad g_{l,2}(z) = \sum_{t=1}^{M} b_t K_{\mathrm{RBF}}(z, z(t)) \tag{23}$$

where $a = (a_1, \ldots, a_M)^{\mathrm{T}}$ and $b = (b_1, \ldots, b_M)^{\mathrm{T}}$ is the unique solution of the well-posed linear system

$$\begin{bmatrix} M\nu_1 I + K_{\mathrm{lin}}[z] & K_{\mathrm{RBF}}[z] \\ K_{\mathrm{lin}}[z] & M\nu_2 I + K_{\mathrm{RBF}}[z] \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} Y_l \\ Y_l \end{bmatrix} \tag{24}$$

where $K_{\mathrm{lin}}[z]$ is the $M \times M$ matrix whose (i,j)th entry is $K_{\mathrm{lin}}(z(i), z(j)), K_{\mathrm{RBF}}[z]$ is the $M \times M$ matrix whose (i,j)th

doi: 10.1049/iet-cta.2013.0150

entry is $K_{\text{RBF}}(z(i), z(j))$ and $Y_l = (Y_l(t_1), Y_l(t_2), \ldots, Y_l(t_M))^{\text{T}}$. Consequently, the estimated model is given by

$$\hat{Y}_l(t) = \hat{g}_l(z(t)) = \sum_{t=1}^{M} a_t K_{\text{lin}}(z, z(t)) + \sum_{t=1}^{M} b_t K_{\text{RBF}}(z, z(t)) \tag{25a}$$

Let $\hat{g} = (\hat{g}_1, \hat{g}_2, \ldots, \hat{g}_{N_h})^{\text{T}}$. Then, for the prediction of the coefficients of finite-element basis, we have

$$\hat{U}(t+1) = U(t) + B^{-1}\hat{g}(z(t)) \tag{25b}$$

For more details of PL-LSRR, it can be referred to [31, 32].

Remark 1: Theorems 1 and 2 demonstrate the convergence and approximation error for the semi-finite element model (16a) to approximate the original SPDE system, and the approximation error here is not the error for the identification model (but this approximation error directly influences the identification error). It can be seen that, with the observation data, the kernel learning method (PL-LSRR) is applied to obtain a specific parametric form (25a) and (25b) to approximate (16a), and then to approximate the underlying SPDE system. It is obvious that the identification error (or the approximation error of the established model (25) to the original SPED (1)) not only depend on h and Δt but also the selection of the number of observations, the type of kernel approximation function, the selection of the kernel and regularisation parameters, even the effectiveness of the PL-LSRR method itself.

3.4 Identification with irregular observation data

Generally, there are two key steps in modelling dynamical behaviours of spatiotemporal systems. The first step is to collect appropriate spatiotemporal data from the plant, and the second step is to develop efficient identification algorithm based on the sampling data. In existing identification methods, all the spatiotemporal nodes are often numerically related with each other using a MIMO–NARX model, and in order to apply the searching based algorithm (such as OLS and OFR [21–24, 27]), the data are usually required to be sampled from the nodes that are regularly distributed on the spatiotemporal dynamical plant, which means that the sensors must be uniformly distributed. However, many practical spatiotemporal systems may not allow us to place sensors regularly. Therefore it is necessary to develop identification methods with irregular observation data. Based on the identification framework proposed above, an identification method is developed for this purpose in this section.

Note that, in the approximation of SPDEs by the finite element method, (20) is used to compute the spatiotemporal states of corresponding SPDEs. When the finite element coefficients $u_h(t_i, x_l)$'s are known, $u(t, x)$ could be computed based on linear scheme (20), where x could be any node in the spacial domain. Inversely, when $u_h(t_i, x_l)$'s are unknown, since the value of spline function $\Phi^{1,3}(t, x) = \phi_{M,1}(t) \otimes \phi_{N,3}(x)$ is fixed for given (t, x), (20) could also be considered as a linear regression equation, where $u(t, x)$'s are the output on the left-hand side of the equation, $\Phi^{1,3}(t, x)$'s are the input on the right-hand side of the equation and $u_h(t_i, x_l)$'s are unknown regression coefficients to be estimated. That is, if we could have enough proper observation data $u(t, x)$ (these data need not to be sampled from regular observation locations), the corresponding unknown finite-element coefficients $u_h(t_i, x_l)$'s could be obtained by a linear regression process.

Assume that we have observations $u(t_i, x_1^\star), u(t_i, x_2^\star), \ldots, u(t_i, x_{N_0}^\star), i = 1, 2, \ldots, M$, where $x_l^\star, l = 1, 2, \ldots, N_0$ could be any node in the spacial domain (of course could be irregularly distributed). Considering the regression characteristics of (20), if the number of observation locations N_0 is greater than the dimensional number of the approximation model N_h, then the data are sufficient for us to compute the finite element coefficients $u_h(t_i, x_l)$'s. Therefore this could be an effective method to obtain the finite-element coefficients $u_h(t_i, x_l)$'s by using irregular observation data. That is to add a linear regression step to estimate the finite-element coefficients $u_h(t_i, x_l)$'s by using data from irregular observation such as $u(t_i, x_l^\star)$s based on the scheme (20), and then, establish the approximation model by using the estimated $\hat{u}_h(t_i, x_l)$'s. To show our method, an illustrative example is given as follows

$$\begin{cases} \dfrac{\partial}{\partial t}u(t,x) - \dfrac{\partial^2}{\partial x^2}u(t,x) + 0.5u(t,x) = \dfrac{\partial^2 W}{\partial t \partial x}, & 0 < t \le T \\ u(t,0) = u(t,1) = 0, & t \ge 0 \\ u(s,x) = u_0(x) & 0 \le x \le 1 \end{cases} \tag{26}$$

where $(\partial^2 W/\partial t \partial x)$ denotes the second-order derivative of the Brownian sheet. Here, for the corresponding evolution operator A and the phase space H, we have $A = A^* = -\Delta$, where $\Delta = (\partial^2/\partial x^2)$ is the Laplace operator and $H = L^2(0, 1)$.

To summarise the discussion above, we provide a general procedure to identify a practical SST system with regular or irregular observation data.

Step 1: Sample data from observation locations, which are not necessary regularly distributed on the spatiotemporal plant. For example, to establish a five-dimensional numerical model for system (26), data from seven locations such as $u(t_i, 0.1), u(t_i, 0.2), u(t_i, 0.33), u(t_i, 0.52), u(t_i, 0.6), u(t_i, 0.85), u(t_i, 0.95), i = 1, 2, \ldots, M$, could be used.

Step 2: Estimate the coefficients of finite-element basis $u_h(t_i, x_l)$'s by following linear regression with the help of the sampled data.

$$u(t, x) = \sum_{i=1}^{M} \sum_{l=1}^{N} \hat{u}_h(t_i, x_l)\Phi_{i,l}^{1,3}(t, x) + e(t, x) \tag{27}$$

For system (26), to establish a five-dimensional model, x here could be 0.1, 0.2, 0.33, 0.52, 0.6, 0.85, 0.95 and $i = 1, 2, \ldots, M$ as presented in the last step. Note that for given x and t, $\Phi_{i,l}^{1,3}(t, x)$'s are fixed, $u_h(t_i, x_l)$'s can be rapidly solved computationally.

Step 3: By using the estimated data $\hat{u}_h(t_i, x_l)$'s ($\hat{u}_l(t_i)$'s or $\hat{u}(t_i, x_l)$'s) in step 2, identification algorithms such as PL-LSRR can be applied to achieve a specific parametric model of (25b) for the original spatiotemporal dynamical system.

3.5 Some further discussions

Usually, for an unknown non-linear spatiotemporal system, it can be assumed that all the spatiotemporal nodes are

doi: 10.1049/iet-cta.2013.0150

numerically related with each other using a MIMO–NARX model, and the identification of spatiotemporal systems given by PDEs or SPDEs has been investigated using the MIMO–NARX model [21–23]. There are two fundamental problems yet to be addressed. That is, (a) whether the approximation model to be identified could be theoretically closely related to the original SPDE, and (b) if they are closely related, what the upper bound of approximation error could be if they are accurately estimated. The theoretical results presented in this study clearly provide the answers to these two questions. The theoretical basis for the MIMO PLM based approximation to a non-linear SPDE system is established, which is shown to be a more straightforward model generated from the governing SPDE model of the underlying spatio-temporal system. Theorems 1 and 2 give a general upper bound of the approximation error considering the length of sampling intervals if the MIMO PLM are accurately estimated. Moreover, from Theorems 1 and 2, it can also be seen that the approximation theory may offer a useful guidance for choosing the proper number of sensors and sampling time required for a practical identification task for spatiotemporal systems subject to a desired estimation accuracy.

In some existing methods (e.g. [21–23]), the data sampled from regular observation locations is directly used to construct the MIMO approximation model, and the model dimension is often required to equal the number of observation locations. If the number of sensors placed is greater than the dimensional number of the model that is to be established, then some data sampled from certain locations can not be used for identification. Thus, these unused data which also contain information of system dynamics has to be discarded. For example, if the data sampled from 12 observation locations is used to construct ten-dimensional approximation model, data sampled from two additional locations can not be used. In the proposed method presented in Section 3.5, all the data regardless of observation locations could be used in the linear regression step. Thus, the data considered to be redundant in some existing methods could be utilised to achieve a more accurate approximation model. Moreover, it is noted that most existing methods are offline ones, whereas the proposed method can be further developed to be an online algorithm to trace the changing dynamics of underlying systems (see [33]).

In addition, although stochastic evolution equations are focused in this study, the proposed methodology could actually be applied to identification of many other different kinds of SPDEs, since the semi-finite element approximation method can be widely applied to approximate many different kinds of PDE systems and the proposed method can be considered as a reasonable inverse semi-finite element method. In practical implementation, although the underlying SPDE is not known but only some experimental data of the system available, we need only to adjust kernel learning parameters and model parameters of the corresponding MIMO PLM model accordingly to achieve a better approximation model, using some available model selection techniques [33, 34]. For example, consider the following damped advection equation

$$\frac{\partial}{\partial t}u(t,x) + u(t,x)\frac{\partial}{\partial x}u(t,x) = -\kappa_d u(t,x) + \frac{\partial^2 W}{\partial t \partial x}$$

where damping coefficient $\kappa_d > 0$, the dynamics of the system develop on the spatial domain $[0, 1]$ and it is obvious that this equation could not be considered as a stochastic evolution equation governed by (1). For $\forall \phi \in S_h$, by using the finite-element method, we have

$$(\phi, u_h(t_{k+1}) - u_h(t_k)) + \int_{t_k}^{t_{k+1}} \left(\phi, u_h \frac{\partial}{\partial x} u_h\right) \mathrm{d}s$$
$$= -\kappa_d \int_{t_k}^{t_{k+1}} (\phi, u_h) + \int_{t_k}^{t_{k+1}} \left(\phi, \frac{\partial^2 W}{\partial s \partial x}\right) \mathrm{d}s$$

Then, the difference method can be used to obtain the following equation

$$(\phi, u_h(t_{k+1}) - u_h(t_k)) = \frac{1}{2}\Delta t(u_h^2(t_k), \phi) - \kappa_d \Delta t(u_h(t_k), \phi)$$
$$+ \int_{t_k}^{t_{k+1}} \left(\phi, \frac{\partial^2 W}{\partial s \partial x}\right) \mathrm{d}s$$

Let $u_h(t,x) = \sum_{l=1}^{N_h} u_l(t)\phi_l$, with the similar technique shown in pages 8 and 9, it can be concluded that the finite-element representation of the damped advection equation above can also be considered as a partially linear regression model with the coefficients as the finite-element basis. Therefore the proposed method can be applied for system identification. For more details of the application of finite-element method to different kinds of PDE, the readers are referred to [35, 36].

It is known that, in practice, the finite-element method is the most commonly used one for the simulation of spatiotemporal dynamical system. The reason could be that the finite element basis can be adjusted to make the finite element method efficient for dynamical systems with different irregular or complicated boundaries. Since in the proposed method, the dynamical systems to be identified are not necessary with regular boundaries, and the identified model could be considered as inverse semi-finite element regression model for finite-element basis coefficients, it can be seen that the proposed method is actually also applicable for systems with irregular boundary conditions.

Usually, most spatiotemporal dynamical systems could be identified by using the regularly spaced data sampled by CCD camera [37, 38]; however, there are also some exceptions. It is well known that sensor networks have been used in many engineering practices such as weather forecast [39], sea surface temperature monitoring [40] and structure health monitoring etc. [41]. In all these applications, SPDEs have been proved to be effective models to describe the dynamics, and irregular spaced data would be encountered because of many realistic reasons such as sensor malfunction, lost data, corrupted data (have to be discarded) or not possible to place sensors at some positions etc. Therefore the method proposed in this paper is more relevant to these problems.

It is also noted that in some realistic situations, correlated noise or non-Gaussian noise instead of white noise may perturb the spatiotemporal dynamics, which are usually modelled by SPDEs driven by correlated noise [42], coloured noise [43] or Levy noise [44]. The qualitative theory, numerical algorithm, computation scheme of these non-white noise driven SPDEs are still open problems and hot research topics [44–46]. Recently, some numerical results show that these SPDEs could be still translated into PLM or NARX model with non-Gaussian random error term. Based on these proposed models [47, 48], it is interesting and important to develop corresponding new kernel regression algorithms to address the identification problems for non-Gaussian or correlated noise perturbed spatiotemporal dynamics. These will definitely be studied in a future study.

The advantages of the proposed method of this study could be summarised as follows. (1) The identification framework in the proposed method is reasonably supported by the finite-element theory, and the approximation error of the underlying semi-finite element model (to be identified by a parametric PLM eventually) to the original spatiotemporal system is clearly given. (2) Our method is applicable for the identification problems with irregular observations and potentially allows irregular or complex boundary conditions. (3) It is very possible to develop online identification algorithm based on our method (see [33]).

In this work, for convenience in discussion, we restrict the examples to the case of one-dimensional SPDE systems. Since we identify the model channel by channel, for the 2D or 3D or even more complex SPDE systems, when using the proposed method, it needs only to increase the N_h (the dimensional number of the model) to achieve a specific parametric model.

Compared with some existing methods such as the coupled map lattice (CML) based identification method [23–25], the proposed method may also have its own disadvantages. On the one hand, for identification problems with irregular observations, the data available for (21a) are obtained by the least-square regression equation (27), which implies that the data used to construct parametric inverse semi-finite element model (25a) are not the direct observations of the dynamics. Moreover, the identified parametric model (25a) is a finite-element approximation of the underlying infinite-dimensional stochastic dynamical system. Therefore the identified model of the proposed method can be understood as an indirect approximation model to the underlying SPDE system, compared with the CML method [23, 24]. On the other hand, when complex spatiotemporal dynamical systems are encountered, computational problems may arise with the proposed method since all possible observation locations should be considered to collect sufficient data for model approximation. However, when only the main features of the underlying dynamics are focused, the computation load could be reduced and the proposed method could still be applicable for online tasks [33] with a particularly designed sparse and irregular sensor network. All these issues will be further investigated in a future study.

4 Numeral simulations

4.1 Example A

Consider the following stochastic heat equation

$$\frac{\partial u}{\partial t} = 0.5\frac{\partial^2 u}{\partial x^2} + 6\ln(1 + |u(x,t)|) + d_0\frac{\partial^2 W}{\partial t \partial x} \quad (28)$$

with the initial and boundary conditions $u(x,0) = 2\sin(\pi x)$, $u(0,t) = 0, u(1,t) = 0$. The corresponding deterministic system is

$$\frac{\partial u}{\partial t} = 0.5\frac{\partial^2 u}{\partial x^2} + 6\ln(1 + |u(x,t)|) \quad (29)$$

with the same initial and boundary conditions of (28). The parameter d_0 is used to denote the noise intensity. When $d_0 = 0$, system (28) becomes the corresponding deterministic system (29). When $d_0 = 0.1$, the output data of system (28) in simulation are plotted in Fig. 1*a*. The data from $t = 0$ to $t = 1$ will be used in model estimation. The simulation output data corresponding to the deterministic system in (29) is plotted in Fig. 1*b*. To apply the proposed identification method, the space domain is sampled at 24 points evenly expanded over $[0,1](0.04, 0.08, \ldots, 0.96)$. Therefore a 24D MIMO PLM model is established as shown in Section 3.4, and the maximum lag adopted here is $n_U = 1$. The time domain is sampled at 350 evenly expanded points over [0,0.35]. Thus, 350 data points used in the regression model for each dimension are generated.

Polynomial kernel ($K(x_i,x_j) = (x_i^{\mathrm{T}}x_j + r^\star)^{d^\star}$, $d^\star$ is polynomial degree and $r^\star$ is tuning parameter), and Gaussian (RBF) kernel ($K(x_i,x_j) = \exp(-\|x_i - x_j\|_2^2/\sigma_0^2)$, σ_0 is the bandwidth) are used in the regression. The selection of kernel parameters may have some influence on identification performance [24, 27, 37]. It is efficient to identify a more complex non-linear process by using a larger degree $d^\star$ and a smaller bandwidth σ_0^2. But it may result in over-fitting problems, when the non-linearity of the process is relatively weak. When a small $d^\star$ and a larger bandwidth σ_0^2 are used, overfitting problems may be avoided, but it is difficult to trace a complex non-linear process. The regularisation constants ν_1, ν_2 are pre-selected to make a trade-off between the empirical prediction errors and the complexity of the model. The ratio $\nu = \nu_1/\nu_2$ is important for how much linear function component should appear in the predicting function. When ν is small, the non-linear function component is restrained more, and more linear function component appears in the predicting function. For the selection of the parameters of PL-LSRR, we refer the readers to the references [20, 31, 32]. In this example, the polynomial kernel function is chosen with parameters $d^\star = 2$ and $r^\star = 0.001$. To obtain a relatively good identification performance, the regulation parameters ν_1 and ν_2 are selected to be 0.05 and 1, which implies the ratio ν_1/ν_2 is 0.05.

The one-step-ahead (OSA) predicted outputs ($\hat{U}(k+1) = U(k) + B^{-1}\hat{g}(z(k))$) and predicted error ($U(k+1) - \hat{U}(k+1)$) over the time domain $[0,1]$ are plotted in Fig. 2, respectively. It can be seen that the dynamical behaviours of the system are traced very well with the established model.

Not to be confused with the one-step-ahead prediction, it should be noted that for all the following simulations in this example, by default, the model prediction is calculated by using the multi-step-ahead (MSA) model prediction output shown as follows, tat is

$$\begin{aligned}\hat{U}(k+1) &= U\hat{(k)} + B^{-1}\hat{g}(\hat{U}(k),\\ &\hat{U}(k-1), \ldots, \hat{U}(k-n_U+1))\end{aligned} \quad (30)$$

with only the initial data $U(0) = u_0$ known, which is given by the initial and boundary conditions of the system under study. The predicted outputs of the identified model for the deterministic part of the system and model estimation errors over the time domain $[0,1]$ are shown in Fig. 3, respectively. It can be seen that the identified model (30) can predict system deterministic output behaviour very well, which implies the accuracy of the proposed method.

To further demonstrate the effectiveness of our method in identifying the deterministic part of the equation, the identification results of the corresponding deterministic system ($d_0 = 0$) are also given. The prediction error over the time domain $[0,1]$ is shown in Fig. 4*a*. For the identification with irregular observation locations (i.e. 0.01, 0.03, 0.05, 0.07, 0.09, 0.11, 0.13, 0.15, 0.17, 0.19, 0.21, 0.23, 0.77, 0.79, 0.81, 0.83, 0.85, 0.87, 0.89, 0.91, 0.93, 0.95, 0.97, 0.99), the prediction error is shown in Fig. 4*b*. The kernel parameters and the number of data selected are the same as the ones used

IET Control Theory Appl., pp. 1–15
doi: 10.1049/iet-cta.2013.0150

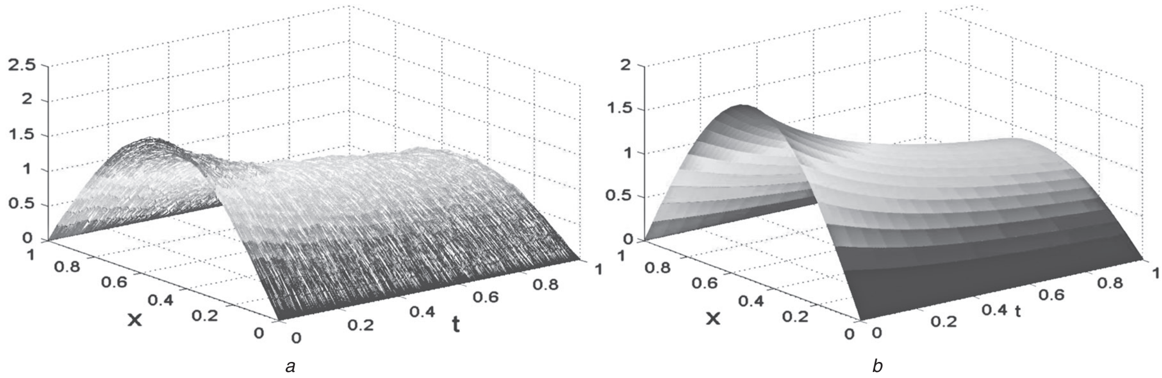


Fig. 1 *Simulation output of system (28) and (29)*

a Simulation output of system (28)
b Simulation output of system (29)

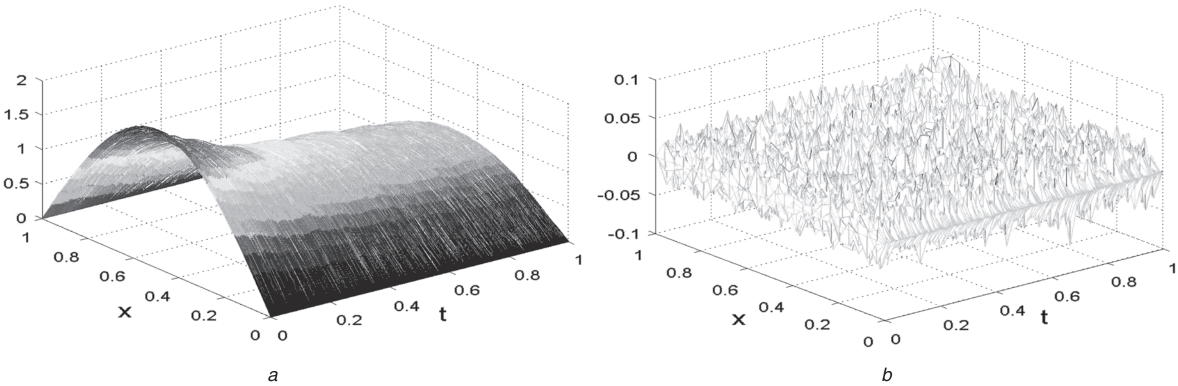


Fig. 2 *Model prediction output and error (OSA)*

a Model prediction output (OSA)
b Model prediction error (OSA)

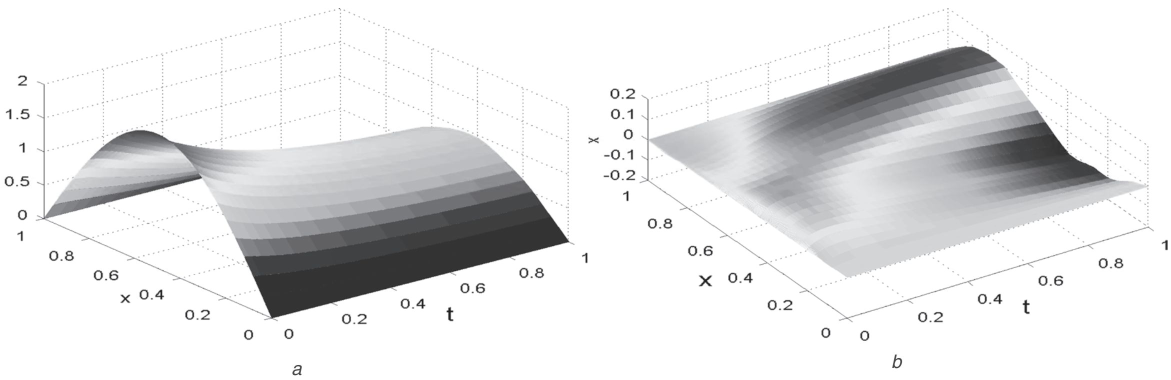


Fig. 3 *Model prediction output and error*

a Model prediction output (MSA)
b Model prediction error (MSA)

before. It can be seen that the identification performance with irregular observation locations is almost the same with the performance with regular ones, and the identification results using the proposed method are very accurate.

4.2 Example B

Consider the following stochastic wave equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \sqrt{1 + u^2(x,t)} + d_0 \frac{\partial^2 W}{\partial t \partial x} \qquad (31)$$

with boundary conditions and initial conditions $u(0,t) = u(1,t) = 0, u(x,0) = 8x(1-x), \partial u/\partial t(x,0) = 0$. The corresponding deterministic system is

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \sqrt{1 + u^2(x,t)} \qquad (32)$$

Equation (32) is with the same boundary conditions and initial conditions as (31). Although the right-hand side of (31) is the second-order derivative with respect to time, system (31) can still be considered as an abstract stochastic evolution equation of the form (1) (for more details of

doi: 10.1049/iet-cta.2013.0150

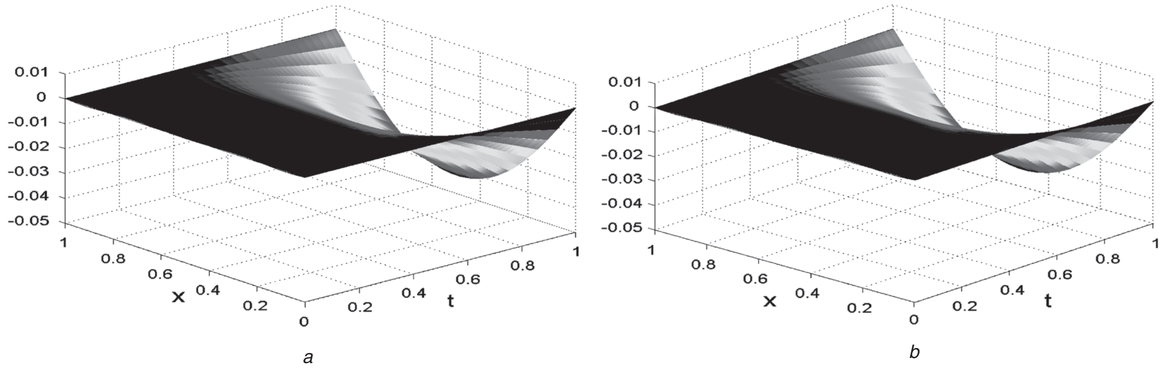


Fig. 4 *Model prediction error (MSA) for regular and irregular observation data*

a Model prediction error (MSA, regular)
b Model prediction error (MSA, irregular)

stochastic wave equations, the readers are referred to [12]). In this example, the Gaussian kernel function is chosen with parameters $\sigma_0^2 = 0.5$. The regulation parameters ν_1 and ν_2 are selected to be 0.4 and 20, which implies the ratio ν_1/ν_2 is 0.02.

When $d_0 = 0.2$, the output data of system (31) is shown in Fig. 5*a*, the first one third of which are used for model identification. The output data from the deterministic part of the system are given in Fig. 5*b*. To apply the proposed identification method, the space domain is sampled with 25 points not uniformly distributed over $[0, 1]$ (i.e. 0.2, 0.22, 0.24, 0.26, 0.28, 0.3, 0.32, 0.34, 0.36, 0.38, 0.5, 0.52, 0.54, 0.56, 0.58, 0.60, 0.62, 0,64, 0.66, 0.68, 0.8, 0.82, 0.84, 0.86, 0.88, 0.9). Then, these data are used to establish a 19D MIMO-PLM model to approximate the original wave equation, and the maximum lag adopted here is $n_U = 2$. The time domain is sampled evenly with 300 points over [0,3]. Thus, 300 data points for each dimension are generated.

The one-step-ahead predicted outputs and errors over the time domain $[0, 10]$ are plotted in Fig. 6. When $d_0 = 0.2$, The multi-step-ahead predicted errors are shown in Fig. 7*a*. For the same reason as Example A, the identification results of the deterministic part ($d_0 = 0$) are also shown. The prediction errors over the time domain $[0, 10]$ are given in Fig. 7*b*. It can be seen that these results still demonstrate

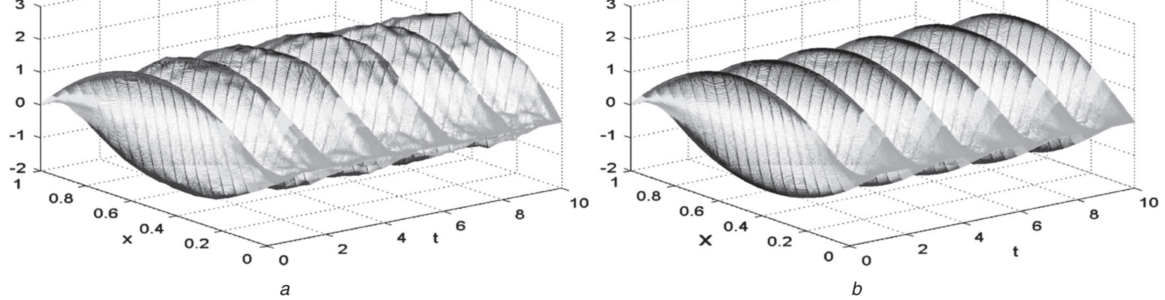


Fig. 5 *Simulation outputs of system (31) and (32)*

a Simulation outputs of system (31)
b Simulation outputs of system (32)

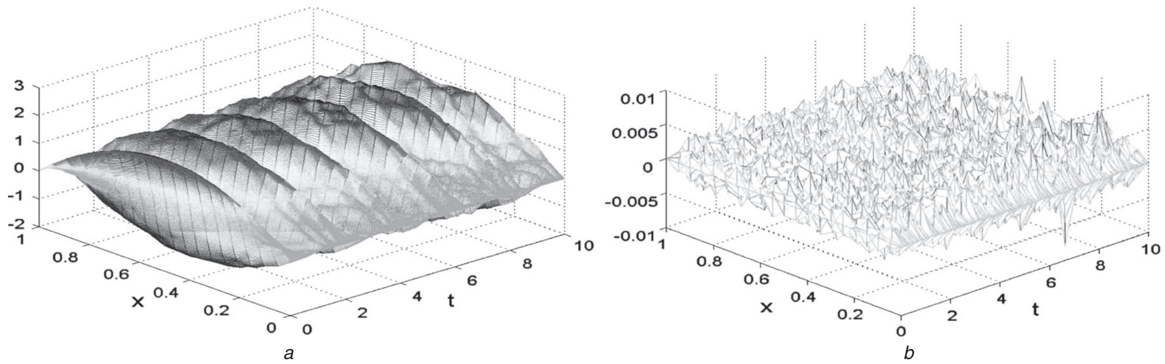


Fig. 6 *Model prediction output and error (OSA)*

a Model prediction output (OSA)
b Model prediction error (OSA)

doi: 10.1049/iet-cta.2013.0150

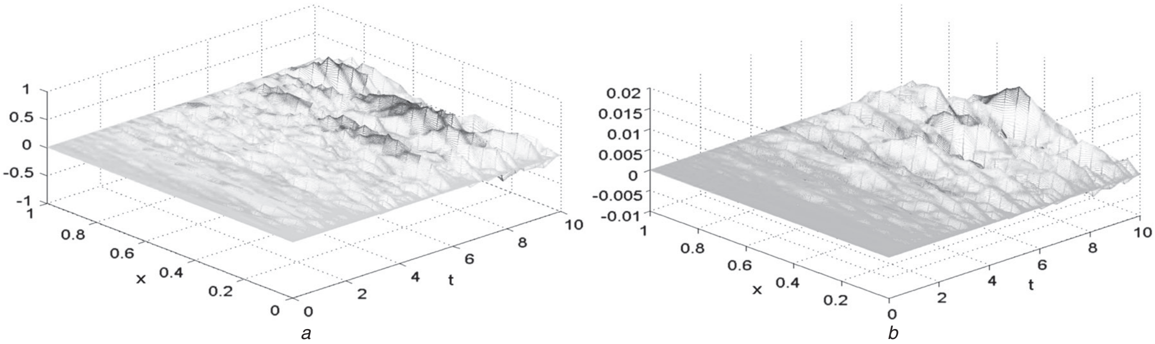


Fig. 7 *Model prediction output and error (MSA)*

a Model prediction error (MSA, noise)
b Model prediction error (MSA, noise free)

the effectiveness of the established stochastic approximation model for the SPDE system in (31).

4.3 Some further discussions on the simulation results

To further demonstrate the identification performance of the proposed method, the relative mean square error (RMSE) is introduced here as follows

$$\mathrm{RMSE}_j = \sqrt{\frac{1}{M}\sum_{k=1}^{M}(\hat{u}(x_j, t_k) - u(x_j, t_k))^2 / u(x_j, t_k)^2}$$

with respect to each spatial node of the dynamical system $j = 1, 2, \ldots, N, M$ and N are numbers of spatial and temporal nodes, respectively. The OSA RMSE for the Example A and Example B are given in Fig. 8, indicating a very small prediction error.

For the MSA prediction, it can be seen for example from Fig. 7 that the MSA prediction error seems becoming larger and larger after some time steps. One major reason could be seen from the MSA prediction method that we adopt in (30). Since only the initial boundary condition is known in our MSA prediction, that is, only $U(0)$ is known, the prediction error (even very very small at beginning) and noise effect must be accumulated in the predicted state of $U(k)$, and therefore, the error would spontaneously become larger and larger (even though the exactly accurate model is used here). Especially, for the MSA prediction with noisy data, the noise eventually drives the system dynamics to be more and more like a stochastic system. However, no matter from Fig. 4 (no noise) or from Fig. 7 (with noise) it can be seen that the overall MSA prediction error is quite small in the given time slot (the overall RMSE is less than 2% in Fig. 4 and less than 15% in Fig. 7), which indicate clearly that the model is well identified.

Moreover, it can also be seen from Figs. 3*b* and 7*a* that the identification results for different spatial nodes are different. Figs. 1*a* and 5*a* show that the dynamics developed around the middle of spatial domain (point 0.5) is different from the dynamics developed around the ends of the spatial domain. Theoretically, the dynamics and characteristics of non-linearity for each spatial point is different. Then, different kernel and regularisation parameters should be selected for each channel of the model in order to achieve better identification performance. Since in the proposed method, the spatiotemporal dynamical systems are identified channel by channel and the identification performance of PL-LSRR would rely on the choice of parameters [31]. However, if complicated identification problems for example for 2D spatial domain are encountered, the dimensional number of the model could be very large. For instance, for an image modelling problem of 1024*768 pixels with 1000 snapshots in time, the model to be estimated could be 786432-dimensional, which implies that it is extremely difficult to manually and empirically select the best parameters for each channel in practice. Therefore in the examples presented above, for each channel of the model, the same parameters are selected to have a relatively good performance as a whole. It is therefore noted that the self-selection of the kernel and regularisation parameters, together with an

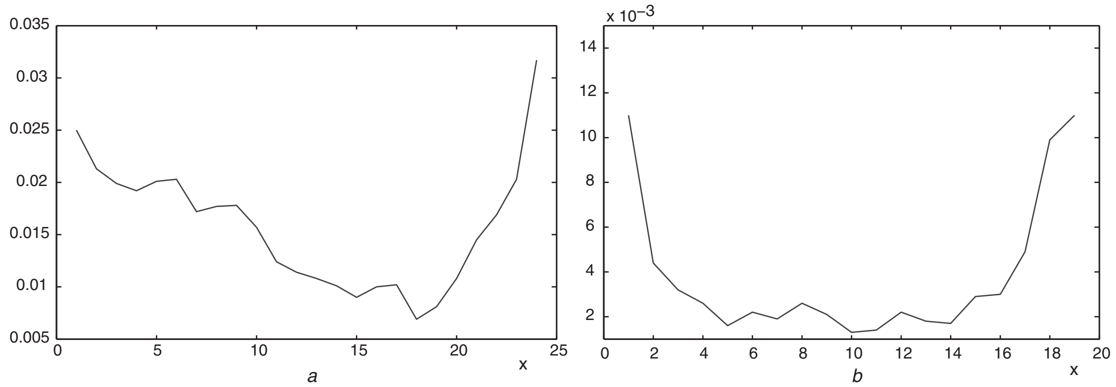


Fig. 8 *RMSE for Examples A and B (OSA)*

a RMSE for Example A (OSA)
b RMSE for Example B

doi: 10.1049/iet-cta.2013.0150

optimisation of sensor locations, is another important topic that should be focused to improve MSA prediction accuracy in future studies.

5 Conclusions

In this paper, a systematic and theoretical approach is proposed for the identification of a semi-finite element model for an underlying infinite-dimensional stochastic spatial-temporal system given by SPDEs based on observation data. It provides a general theoretical framework for the identification of non-linear SST systems using MIMO–PLMs. Fundamental theoretic results are established to deal with the approximation issues of the identified model to the underlying SPDE system. The identification problem with irregular observation locations is also discussed to make the proposed method more effective and realistic in practice. The proposed identification method should have a wide applicability in many areas of scientific and engineering practices where the analysis and design of non-linear SST systems are involved. Future studies will focus more on improvement of MSA prediction with smart parameter selection procedure, reduction of computation load with strategic irregular distribution of observation locations, and identification methods for non-Gaussian noise or correlated noise perturbed spatiotemporal dynamics. A continuous-time identification method with some existing techniques [49–51] will also be studied in the context of the proposed identification scheme.

Moreover, it is noted that some valuable theoretical results of numerical computations for SPDEs are reported recently in the framework of stochastic evolution equations using the semigroup approach ([11–13, 15, 16] and references therein). Many interesting analysis and numerical simulations are also obtained, which provide a significant insight into the dynamics of SPDEs. However, the topics such as data based optimisation, control and modelling, signal processing and filter design for SPDEs, which have great potential in application, have not yet received much attention [52–55]. On the other hand, in the past decades, a number of powerful PLM-based algorithms such as PL-LSRR, PLM-based support vector machine and maximal likelihood estimation et al are developed. These methods have been noted increasing reports for system identification, pattern recognition and optimal control of dynamical systems [19, 21, 22] and references therein). Therefore, the results developed and demonstrated in this paper provide a useful insight into these interesting research topics mentioned above, and alternatively potential methods to solve the mentioned problems could be found by following a similar technical line. These will be focused in future studies.

6 Acknowledgments

The authors would like to thank the editor and reviewers for your useful comments and constructive suggestions on this paper. This work was supported by the GRF project of Hong Kong RGC (Ref. no. 517810), Internal (competitive) research grants of Hong Kong Polytechnic University.

7 References

1 Debnath, L.: ‘Nonlinear partial differential equations for scientists and engineers’ (Birkhauser, Boston, 2005)

2 Farlow, S.J.: ‘Partial differential equations for scientists and engineers’ (John Wiley and Sons Inc., New York, 1982)

3 Da Prato, G., Zabczyk, J.: ‘Stochastic equations in infinite dimensions’ (Cambridge University Press, Cambridge, UK, 1992)

4 Chow, P.: ‘Stochastic partial differential equations’ (Chapman and Hall, Boca Raton, FL, 2007)

5 Peszat, S., Zabczyk, J.: ‘Stochastic partial differential equations with levy noise: an evolution equation approach’ (Cambridge University Press, Cambridge, UK, 2007)

6 Friedman, A.: ‘Stochastic differential equations and applications’ (Academic Press, New York, 1975)

7 Mahmudov, N.I.: ‘Controllability of linear stochastic systems’, *IEEE Trans. Autom. Control*, 2001, **41**, pp. 724–731

8 Allen, E.J., Novosel, S.J., Zhang, Z.: ‘Finite element and difference approximation of some linear stochastic partial differential equations’, *Stoch. Stoch. Rep.*, 1998, **64**, pp. 117–142

9 Yoo, H.: ‘Semi-discretization of stochastic partial differential equations on R^1 by a finite-difference method’, *Math. Comput.*, 1999, **69**, pp. 653–666

10 Higham, D.J.: ‘An algorithmic introduction to numerical simulation of stochastic differential equations’, *SIAM Rev.*, 2001, **3**, pp. 525–546

11 Yan, Y.: ‘Galerkin finite element methods for stochastic parabolic partial differential equations’, *SIAM J. Numer. Anal.*, 2005, **4**, pp. 1363–1384

12 Kovacs, M., Larsson, S., Saedpanah, F.: ‘Finite element approximation of the linear stochastic wave equation with additive noise’, *SIAM J. Numer. Anal.*, 2010, **2**, pp. 408–427

13 Cao, Y., Yang, H., Yin, L.: ‘Finite element methods for semilinear elliptic stochastic partial differential equations’, *Numer. Math.*, 2007, **1**, pp. 181–198

14 Thomee, V.: ‘Gelerkin finite element methods for parabolic problems’ (Springer-Verlag, Berlin, 2006, 2nd edn.)

15 Sjoberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P.-Y. et al.: ‘Nonlinear black-box modeling in system identification: a unified overview’, *Automatica*, 1995, **12**, pp. 1691–1724

16 Milanese, M., Novara, C.: ‘Set membership identification of non-linear systems’, *Automatica*, 2004, **40**, pp. 957–975

17 Chen, S., Billings, S.A., Luo, W.: ‘Orthogonal least squares methods and their application to non-linear system identification, *Int. Control J.*, 1989, **50**, pp. 1873–1896

18 Lind, I., Ljung, L.: ‘Regressor selection with the analysis of variance method’, *Automatica*, 2005, **41**, pp. 693–700

19 Vapnik, V.: ‘The Nature of statistical learning theory’ (Springer-Verlag, New York, 1995)

20 Suykens, J.A.K., Gestel, V.T., Brabanter, J.D., Moor, B.D., Vandewalle, J.: ‘Least squares support vector machines’ (World Scientific, Singapore, 2002)

21 Guo, L., Billings, S.A.: ‘State-space reconstruction and spatiotemporal prediction of lattice dynamical systems’, *IEEE Trans. Autom. Control*, 2007, **52**, pp. 622–632

22 Guo, L., Billings, S.A., Coca, D.: ‘Identification of partial differential equation models for a class of multiscale spaio-temporal dynamical systems’, *Int. Control J.*, 2010, **83**, pp. 40–48

23 Wei, H., Billings, S.A., Zhao, Y., Guo, L.: ‘Lattice dynamical wavelet neural networks implemented using particle swarm optimization for spatio-temporal system identification’, *IEEE Trans. Neural Netw.*, 2009, **20**, pp. 181–185

24 Billings, S.A., Guo, L., Wei, H.: ‘Identification of coupled map lattice models for spatiotemporal patterns using wavelets’, *Int. J. Syst. Sci.*, 2006, **37**, pp. 1021–1038

25 Guo, L., Billings, S.A.: ‘Identification of partial differential equation models for continuous spatiotemporal dynamical systems’, *IEEE Trans. Circuits Syst. II, Express Briefs*, 2008, **53**, pp. 657–661

26 Li, H., Qi, C.: ‘Spatiotemporal modeling of nonlinear distributed parameter systems’ (Springer, 2011)

27 Coca, D., Billings, S.A.: ‘Identification of finite dimensional models of infinite dimensional dynamical systems’, *Automatica*, 2002, **38**, pp. 1851–1865

28 Barth, A.: ‘A finite element method for martingale-driven stochastic partial differential equations’, *Commun. Stoch. Anal.*, 2010, **3**, pp. 355–375

29 Jentzen, A., Kloeden, P.E.: ‘Overcoming the order barrier in the numerical approximation of stochastic partial differential equations with additive space-time noise’, *Proc. R. Soc. A*, 2009, **465**, pp. 649–667

30 de Boor, C., Hollig, K., Riemenschneider, S.: ‘Box splines’ (Springer-Verlag, Berlin, 1993)

31 Xu, Y., Chen, D.: ‘Partially linear least squares regularized regression for system identification’, *IEEE Trans. Autom. Control*, 2009, **54**, pp. 2637–2641

doi: 10.1049/iet-cta.2013.0150

32 Espinoza, M., Suykens, J.A.K., Moor B.D.: 'Kernel based partially linear models and non-linear identification', *IEEE Trans. Autom. Control*, 2005, **50**, pp. 1602–1606

33 Ning, H., Jing, X., Chen, L.: 'Online identification of nonlinear spatiotemporal systems using kernel learning approach', *IEEE Trans. Neural Netw.*, 2011, **22**, pp. 1381–1394

34 Hong, X., Mitchell, R.J., Chen, S., Harris, C.J., Li, K., Irwin, G.W.: 'Model selection approaches for non-linear system identification: a review', *Int. J. Syst. Sci.*, 2008, **39**, pp. 925–946

35 Girault, V., Raviart, P.A.: 'Finite element method for Navier–Stokes equation: theory and algorithms' (Springer-Verlag, Berlin, 1986)

36 Vegt, J.J.W.V.D., Bokhove, O.: 'Finite element method for partial differential equations'. Working paper, 2009

37 Goljan, M.: 'Digital camera identification from imagesCestimating false acceptance probability', *Lect. Notes Comput. Sci.*, 2009, **5450**, pp. 454–468

38 Lucas, J., Fridrich, J., Goljan, M.: 'Digital camera identification from sensor pattern noise', *IEEE Trans. Inf. Forensics Sec.*, 2006, **52**, pp. 205–214

39 Atak, A., Linton, O., Xiao, Z.: 'A semiparametric panel model for unbalanced data with application to climate change in the United Kingdom', *J. Econ.*, 2011, **164**, pp. 92–115

40 Ricardo, A., Liliana, B., Josselin, G.: 'Wave propagation in waveguides with random boundaries', arXiv:1111.1118v1

41 Charles, R., Farrar, R., Keith, W.: 'An introduction to structural health monitoring', *Phil. Trans. R. Soc. A*, 2007, **365**, pp. 303–315

42 Carreras, D.M., Mellou, M., Sarra, M.: 'On stochastic partial differential equations with spatially correlated noise: smoothness of the law', *Stoch. Process Appl.*, 2001, **93**, pp. 269–284

43 Ferrante, M., Martaa, S.: 'SPDEs with coloured noise: analytic and stochastic approaches', *ESAIM: Prob. Stat.*, 2006, **10**, pp. 380–405

44 Peszat, S., Zabczyk, J.: 'Stochastic partial differential equations with levy noise' (Cambridge, 2007)

45 Barth, A.: 'Stochastic partial differential equations: approximations and applications'. submitted for PhD dissertation, Center of Mathematics for Applications (CMA) University of Oslo, 2009

46 Abdulle, A., Pavliotis, G.A.: 'Numerical methods for stochastic partial differential equations with multiple scales', *J. Comput. Phys.*, 2012, **231**, pp. 2482–2497

47 Kamrani, M., Blomker, D.: 'Numerical solution of stochastic partial differential equations with correlated noise' (Institution fur Mathematik der University Augsburg, Preprints-Herausgeber, 2013)

48 Doering, C.R., Sargsyan, K.V., Smereka, P.: 'A numerical method for some stochastic differential equations with multiplicative noise', *Phys. Lett. A*, **344**, 2005, pp. 149–155

49 Garnier, H., Wang, L.: 'Identification of continuous-time models from sampled data' (Springer, 2008)

50 Walsh, J.B.: 'An introduction to stochastic partial differential equations, lecture notes in mathematics 1180' (Springer-Verlag, Berlin, 1986)

51 Chen, S., Billings, S.A.: 'Representation of non-linear systems: the NARMAX model', *Int. Control J.*, 1989, **49**, pp. 1013–1032

52 Scholkopf, B., Smola, A.J.: 'Learning with kernels: support vector machines, regularization, optimization and beyond' (MIT Press, Cambridge, MA, 2002)

53 Suykens, J.A.K., Vandewalle, J., Moor, B.D.: 'Optimal control by least squares support vector machines', *Neural Netw.*, 2001, **14**, pp. 23–35

54 Qi Chenkun, C., Li, H., Zhang, X., Zhao, X., Li, S., Gao, F.: 'Time/space-separation-based SVM modeling for nonlinear distributed parameter processes', *Ind. Eng. Chem. Res.*, 2011, **50**, pp. 332–342

55 Xia, Y., Leung, H.: 'Nonlinear spatial-temporal prediction based on optimal fusion', *IEEE Trans. Neural Netw.*, **17**, 2006, pp. 975–988

56 Pazy, A.: 'Semigroups of linear operators and applications to partial differential equations' (Springer-Verlag, Berlin, 1983)

8 Appendix

8.1 Appendix 1: Proof of Theorem 1

By the definition of the solutions (3) and (7), we have

$$\begin{aligned}
&\mathbf{E}|u_h(t)-u(t)|^2\\
&\quad=\mathbf{E}|(E_h(t)P_h-E(t))u_0\\
&\qquad+\int_0^t(E_h(t-s)P_h-E(t-s))\,\mathrm{d}w(s)\\
&\qquad+\int_0^t E_h(t-s)P_h(f(u_h(s))-f(u(s)))\,\mathrm{d}s\\
&\qquad+\int_0^t(E_h(t-s)P_h-E(t-s))f(u(s))\,\mathrm{d}s|^2\\
&\quad\le 4\mathbf{E}|(E_h(t)P_h-E(t))\xi|^2\\
&\qquad+4\mathbf{E}|\int_0^t(E_h(t-s)P_h-E(t-s))\,\mathrm{d}w(s)|^2\\
&\qquad+4\mathbf{E}|\int_0^t E_h(t-s)P_h(f(u_h(s))-f(u(s)))\,\mathrm{d}s|^2\\
&\qquad+4\mathbf{E}|\int_0^t(E_h(t-s)P_h-E(t-s))f(u(s))\,\mathrm{d}s|^2\\
&\quad=I_1+I_2+I_3+I_4 \qquad (33)
\end{aligned}$$

It follows from Lemma 3.2

$$I_1=4\mathbf{E}|(E_h(t)P_h-E(t))u_0|^2\le Ch^2\mathbf{E}||u_0||^2_{\dot H^2}$$

By Lemma 3.2 and Burkholder–Davis–Gundy inequality [3], we have

$$\begin{aligned}
I_2&=4\mathbf{E}|\int_0^t(E_h(t-s)P_h-E(t-s))\,\mathrm{d}w(s)|^2\\
&\le 4\mathbf{E}\int_0^t\|E_h(t)P_h-E(t)\|^2\,\mathrm{d}s\le Ch^2
\end{aligned}$$

For I_3, with the help of the Hölder inequality, the Lipschitze condition and growth condition imposed on f, we have

$$\begin{aligned}
I_3&=4\mathbf{E}|\int_0^t(E_h(t-s)P_h-E(t-s))f(u(s))\,\mathrm{d}s|^2\\
&\le 4\int_0^t\|E_h(t-s)P_h-E(t-s)\|^2\,\mathrm{d}s\mathbf{E}\int_0^t|f(u(s))|^2\,\mathrm{d}s\\
&\le Ch^2\mathbf{E}\int_0^t|u(s)|^2\,\mathrm{d}s\le Ch^2\sup_{0\le s\le T}\mathbf{E}|u(s)|^2
\end{aligned}$$

For the last term I_4

$$\begin{aligned}
I_4&=4\mathbf{E}|\int_0^t E_h(t-s)P_h(f(u_h(s))-f(u(s)))\,\mathrm{d}s|^2\\
&\le 4\mathbf{E}\left(\int_0^t\|E_h(t-s)P_h\||f(u_h(s))-f(u(s))|\,\mathrm{d}s\right)^2\\
&\le C\int_0^t\|E_h(t-s)P_h\|^2ds\int_0^t\mathbf{E}|u_h(s)-u(s)|^2\,\mathrm{d}s\\
&\le CT\int_0^t\mathbf{E}|u_h(s)-u(s)|^2\,\mathrm{d}s
\end{aligned}$$

By the estimations obtained above, we have

$$\begin{aligned}
\mathbf{E}|u_h(t)-u(t)|^2&\le Ch^2(1+E\|u_0\|^2_{\dot H^2}+\sup_{0\le s\le T}\mathbf{E}|u(s)|^2)\\
&\quad+CT\int_0^t\mathbf{E}|u_h(s)-u(s)|^2\,\mathrm{d}s \qquad (34)
\end{aligned}$$

Therefore, using the Gronwall's inequality, it can be obtained that there exists a constant $C=C(u_0,T)$ such that

$$\mathbf{E}|u_h(t)-u(t)|^2\le Ch^2 \qquad (35)$$

It can be seen that for any given time instance $t\in[0,T]$, the finite-element space approximation model with respect

doi: 10.1049/iet-cta.2013.0150

to spatial variables is convergence to the original SPDE in the order of $O(h^2)$. The proof is completed. □

8.2 Appendix 2: Proof of Theorem 3.3

For simplicity, let $E_h^n = r(\Delta t A_h)^n$ and $U^n = u_h(t_n)$, and by (10), we have

$$U^n = E_h^n P_h u_0 + \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} E_h^{n-j+1} P_h f(U^{j-1})\,\mathrm{d}s + \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} E_h^{n-j+1} P_h\,\mathrm{d}(s) \tag{36}$$

With the help of the definition of mild solution for system (1), we have

$$u(t_n) = E(t_n)u_0 + \int_0^{t_n} E(t_n - s)f(u(s))\,\mathrm{d}s + \int_0^{t_n} E(t_n - s)\,\mathrm{d}(s) \tag{37}$$

which implies that the approximation error in mean-square manner can be represented as

$$\begin{aligned}
\mathbf{E}|U^n - u(t_n)|^2 \le C\Bigg(&\mathbf{E}|(E_h^n P_h - E(t_n))u_0|^2 \\
&+ \mathbf{E}\left|\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j} E_h^{n-j+1}P_h(f(U^{j-1}) - f(u(t_{j-1})))\,\mathrm{d}s\right|^2 \\
&+ \mathbf{E}\left|\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j} E_h^{n-j+1}P_h(f(u(t_{j-1})) - f(u(s)))\,\mathrm{d}s\right|^2 \\
&+ \mathbf{E}\left|\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j} (E_h^{n-j+1}P_h - E(t_n - t_{j-1}))f(u(s))\,\mathrm{d}s\right|^2 \\
&+ \mathbf{E}\left|\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j} (E(t_n - t_{j-1}) - E(t_n - s))f(u(s))\,\mathrm{d}s\right|^2 \\
&+ \mathbf{E}\left|\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j} (E_h^{n-j+1}P_h - E(t_n - t_{j-1}))\,\mathrm{d}(s)\right|^2 \\
&+ \mathbf{E}\left|\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j} (E(t_n - t_{j-1}) - E(t_n - s))\,\mathrm{d}(s)\right|^2\Bigg)
\end{aligned} \tag{38}$$

Directly from the results in [14] (Theorem 7.8), for the first term of above equation, we have

$$\mathbf{E}|(E_h^n P_h - E(t_n))u_0|^2 \le C(\Delta t + \Delta h^2)$$

By the Lipschitze conditions imposed on f, the stability of $r(\lambda)$ and the Hölder inequality, for the second term, we have

$$\begin{aligned}
&\mathbf{E}\left|\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j} E_h^{n-j+1}P_h(f(U^{j-1}) - f(u(t_{j-1})))\,\mathrm{d}s\right|^2 \\
&\le n\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j} 1^2\,\mathrm{d}s\mathbf{E}\int_{t_{j-1}}^{t_j} ||E_h^{n-j+1}P_h||^2|f(U^{j-1}) \\
&\quad - f(u(t_{j-1}))|^2\,\mathrm{d}s \\
&\le \frac{T}{\Delta t}\Delta t\sum_{j=1}^{n}\Delta t C\mathbf{E}|f(U^{j-1}) - f(u(t_{j-1}))|^2 \\
&\le C\Delta t\sum_{j=1}^{n}\mathbf{E}|U^{j-1} - u(t_{j-1})|^2
\end{aligned}$$

For the third term, by Lemma 1, it can be obtained that

$$\begin{aligned}
&\mathbf{E}\left|\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j} E_h^{n-j+1}P_h(f(u(t_{j-1})) - f(u(s)))\,\mathrm{d}s\right|^2 \\
&\le n\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j} 1^2\,\mathrm{d}s\mathbf{E}\int_{t_{j-1}}^{t_j} |E_h^{n-j+1}P_h|^2|f(u(s)) \\
&\quad - f(u(t_{j-1}))|^2\,\mathrm{d}s \\
&\le C\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\mathbf{E}|u(s) - u(t_{j-1})|^2\,\mathrm{d}s \\
&\le C\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}(s - t_{j-1})^2\,\mathrm{d}s \le C\Delta t^2
\end{aligned}$$

To obtain the estimation of the fourth term, an estimation result on E_h^n is borrowed from [13]. We have

$$\begin{aligned}
&\Delta t\sum_{j=1}^{n}|(E_h^{n-j+1}P_h - E(t_n - t_{j-1}))u_0|^2 \\
&= \Delta t\sum_{j=1}^{n}|(E_h^jP_h - E(t_j)u_0|^2 \le C(\Delta t + \Delta x^2)|u_0|^2
\end{aligned}$$

Then, by the growth condition imposed on f, it follows that

$$\begin{aligned}
&\mathbf{E}|\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}(E_h^{n-j+1}P_h - E(t_n - t_{j-1}))f(u(s))\,\mathrm{d}s|^2 \\
&\le n\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j} 1^2\,\mathrm{d}s\mathbf{E} \\
&\quad \times\int_{t_{j-1}}^{t_j} C\|E_h^{n-j+1}P_h - E(t_n - t_{j-1})\|^2|u(s)|^2\,\mathrm{d}s \\
&\le C \le \left(\sup_{0\le s\le T}\mathbf{E}|u(s)|^2\right)(\Delta t + \Delta x^2)
\end{aligned}$$

For the fifth term, Theorem 6.13 introduced in [56] can be utilised for the estimation, and we have

$$\begin{aligned}
&\mathbf{E}\left|\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}(E(t_n - t_{j-1}) - E(t_n - s))f(u(s))\,\mathrm{d}s\right|^2 \\
&\le n\sum_{j=1}^{n}\mathbf{E}\left(\int_{t_{j-1}}^{t_j}|E(t_n - s)(E(s - t_{j-1}) - I)f(u(s))|\,\mathrm{d}s\right)^2 \\
&\le C\left(\sup_{0\le s\le T}\mathbf{E}\|u(s)\|_{\dot{H}^2}^2\right)\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}(s - t_{j-1})^2\,\mathrm{d}s \\
&\le C\left(\sup_{0\le s\le T}\mathbf{E}\|u(s)\|_{\dot{H}^2}^2\right)\Delta t^2
\end{aligned}$$

Since the cross terms of the square sum are zero, which can be guaranteed by the results in [3], for the last two terms,

doi: 10.1049/iet-cta.2013.0150

with the similar techniques demonstrated above, it follows that

$$\mathbf{E}\left|\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}(E_h^{n-j+1}P_h - E(t_n - t_{j-1}))\,\mathrm{d}(s)\right|^2 \le \sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\|E_h^{n-j+1}P_h - E(t_n - t_{j-1})\|^2\,\mathrm{d}s \le C(\Delta t + \Delta x^2)$$

For the last term

$$\mathbf{E}|\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}(E(t_n - t_{j-1}) - E(t_n - s))\,\mathrm{d}(s)|^2 \le \sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}||E(t_n - t_{j-1})(I - E(s - t_{j-1}))||^2\,\mathrm{d}s \le C\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}(s - t_{j-1})^2\,\mathrm{d}s \le C\Delta t^2$$

By adding all estimations obtained above, it is shown that there exists a constant $C = C(T, u_0)$, such that

$$\mathbf{E}|U^n - u(t_n)|^2 \le C(\Delta t + \Delta t^2 + \Delta x^2) + C\Delta t\sum_{j=1}^{n}\mathbf{E}|U^j - u(t_j)|^2 \qquad (39)$$

Then, by using the discrete form of Grönwall inequality, we can obtain that

$$\mathbf{E}|U^n - u(t_n)|^2 \le C(\Delta t + \Delta x^2) \qquad (40)$$

Thus

$$\mathbf{E}|u_h(t_n) - u(t_n)|^2 \le C(\Delta t + \Delta x^2) \qquad (41)$$

The proof is completed. □

doi: 10.1049/iet-cta.2013.0150