Available online at www.sciencedirect.com
ScienceDirect
Journal of Process Control xxx (2008) xxx–xxx
JOURNAL OF PROCESS CONTROL
www.elsevier.com/locate/jprocont

A multi-channel spatio-temporal Hammerstein modeling approach for nonlinear distributed parameter processes

Chenkun Qi [a], Hai-Tao Zhang [b,c], Han-Xiong Li [a,*]

[a] Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong, China
[b] Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, PR China
[c] Department of Engineering, University of Cambridge, Cambridge CB1 1PZ, UK

Received 9 August 2007; received in revised form 15 January 2008; accepted 21 January 2008

Abstract

Modeling of distributed parameter processes is a challenging problem because of their complex spatio-temporal nature, nonlinearities and uncertainties. In this study, a spatio-temporal Hammerstein modeling approach is proposed for nonlinear distributed parameter processes. Firstly, the static nonlinear and the distributed dynamical linear parts of the Hammerstein model are expanded onto a set of spatial and temporal basis functions. In order to reduce the parametric complexity, the Karhunen–Loève decomposition is used to find the dominant spatial bases with Laguerre polynomials selected as the temporal bases. Then, using the Galerkin method, the spatio-temporal modeling will be reduced to a traditional temporal modeling problem. Finally, the unknown parameters can be easily estimated using the least squares estimation and the singular value decomposition. In the presence of unmodeled dynamics, a multi-channel modeling framework is proposed to further improve the modeling performance. The convergence of the modeling can be guaranteed under certain conditions. The simulations are presented to show the effectiveness of this modeling method and its potential to a wide range of distributed processes.

Keywords: Distributed parameter system; Spatio-temporal modeling; Hammerstein model; Galerkin method; Karhunen–Loève decomposition; Laguerre polynomial

1. Introduction

Many of industrial processes such as fluid flow, convection and diffusion reaction process, and thermal process belong to distributed parameter systems (DPS). They are a class of infinite-dimensional dynamical systems because the inputs, outputs, states and parameters vary both temporally and spatially. Traditional lumped models cannot properly describe this class of spatially distributed systems. Although partial differential equations (PDEs) can accurately predict the nonlinear and spatially distributed dynamical behavior, it is difficult to apply due to limited computation capacity for numerical implementation and finite actuators/sensors for practical control. Therefore a finite-dimensional model is often desirable. Practically, obtaining a proper finite-dimensional model is a difficult and challenging problem due to the spatio-temporal coupling and nonlinear uncertainties existing in the processes.

If the process is known, traditional methods such as finite difference and finite element can be easily applied to discretization of the process. But they will lead to a high-order ordinary differential equation (ODE) which is not suitable for real-time control. Under some conditions, a low-order ODE model may be possible by using weighted residual methods [24] including the Galerkin method, the collocation method and the approximate inertial manifold method [6], etc. Due to complex boundary conditions and external disturbances, it is very difficult to obtain a good model only from physical insights. Therefore, it is necessary to develop a data-driven identification method.

* Corresponding author. Tel.: +852 27888435; fax: +852 27888423.
E-mail address: mehxli@cityu.edu.hk (H.-X. Li).

0959-1524/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jprocont.2008.01.006

For the unknown linear DPS, a linear model can be derived from the input–output data using the singular value decomposition (SVD) method [14] for a linear time-invariant distributed parameter system or the Karhunen–Loève (KL) expansion combination with the SVD [36] for a linear time-varying distributed parameter system. However, a linear model is only able to approximate the nonlinear system around a given working point.

For the unknown nonlinear DPS, some basis function expansion based models such as nonlinear autoregressive with exogenous input model [7] and artificial neural networks [25,28,10] can be used to establish the dynamical relationship between the coefficients of basis functions and the inputs. However such models may be very complex which are not suitable for practical control.

In the traditional nonlinear system identification, the block-oriented nonlinear models have been widely used because of their simple structures, abilities to approximate a large class of nonlinear processes and efficient control schemes. They consist of the interconnection of linear time invariant (LTI) systems and static nonlinearities. Within this class, two common model structures are: the Hammerstein model, which consists of the cascade connection of a static nonlinearity followed by a LTI system, and the Wiener model, in which the order of the linear and the nonlinear blocks is reversed. These models have been successfully used to represent many practical nonlinear processes [12,13]. To the authors' knowledge, so far the block-oriented nonlinear models are only studied for lumped parameter systems. This study will extend the traditional Hammerstein modeling into nonlinear distributed parameter systems.

For the traditional Hammerstein modeling, several methods have been proposed in the literature [21,29,1,2,4,37,11,17,31,16]. It is notable that an algorithm based on the least squares estimation and the singular value decomposition (LSE–SVD) is proposed for Hammerstein–Wiener systems [1] and extendedly studied for Hammerstein systems [16]. The algorithm is derived from the use of basis functions for the representation of the linear and the nonlinear parts. In the case of model matching, the consistency of the estimates can be guaranteed under certain conditions. However, in the presence of unmodeled dynamics, further studies are required.

In this study, in order to model the nonlinear distributed parameter system, a spatio-temporal Hammerstein model is proposed for the first time by adding the space variables into the traditional Hammerstein model, which consists of the cascade connection of a static nonlinearity followed by a distributed dynamical linear time-invariant system.

A basic identification algorithm based on LSE–SVD is designed as follows. Firstly, the nonlinear and the distributed linear parts are expanded onto spatial and temporal basis functions with unknown coefficients. In order to reduce the parametric complexity, the KL decomposition is used to find the dominant spatial basis functions and Laguerre polynomials are selected as the temporal basis functions. Then, using the Galerkin method, the spatio-temporal modeling will turn into a traditional modeling problem in time domain. Subsequently, least squares techniques can be used to identify a parameter matrix charactering the product of parameters of the linear and the nonlinear parts. Finally, by using SVD, optimal estimates of the parameters of each part can be obtained. This basic identification algorithm can provide consistent estimates under some assumptions in the case of model matching.

In the presence of unmodeled dynamics, a multi-channel identification algorithm is proposed to compensate the residuals of the single-channel model and further reduce the modeling error. This algorithm is noniterative and numerically robust since it is based only on the least squares estimation and the singular value decomposition. The convergent estimates can be guaranteed under proper assumptions. The spatio-temporal Hammerstein model can be easily used for many applications such as model predictive control due to its simple nonlinear structure. The simulations demonstrate the effectiveness of the proposed modeling method.

The main contributions can be summarized as follows:

(1) The traditional Hammerstein modeling is extended into nonlinear distributed parameter systems.
(2) A basic identification algorithm based on LSE–SVD is developed by using the Galerkin method with the KL decomposition and the Laguerre polynomials expansion.
(3) In the presence of unmodeled dynamics, a multi-channel identification approach is proposed to further reduce the modeling error.
(4) The estimates are proved to be convergent under some conditions.

The rest of the paper is organized as follows. In Section 2, the spatio-temporal Hammerstein system is proposed. The single-channel identification algorithm is derived in Section 3. The multi-channel modeling approach and analysis are provided in Section 4. Simulation examples illustrating the performance of the proposed modeling approach are presented in Section 5, and finally, some conclusions are provided in Section 6.

2. Spatio-temporal Hammerstein system

A spatio-temporal Hammerstein system is shown in Fig. 1. The system consists of a static nonlinear element $N(\cdot) : \mathbb{R} \to \mathbb{R}$ followed by a distributed linear time-invariant system

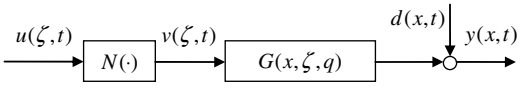

Fig. 1. Spatio-temporal Hammerstein system.

Please cite this article in press as: C. Qi et al., A multi-channel spatio-temporal Hammerstein modeling approach ..., J. Process Contr. (2008), doi:10.1016/j.jprocont.2008.01.006

$$y(x,t)=\sum_{\tau=0}^{t}\int_{\Omega}g(x,\zeta,\tau)v(\zeta,t-\tau)\mathrm{d}\zeta, \tag{1}$$

with transfer function $G(x, \zeta, q)$ (1×1), where x and ζ are spatial variables defined on the domain Ω, and q stands for the forward shift operator. The input–output relationship of the system is then given by

$$y(x,t)=\sum_{\tau=0}^{t}\int_{\Omega}g(x,\zeta,\tau)N(u(\zeta,t-\tau))\mathrm{d}\zeta+d(x,t), \tag{2}$$

where $u(x,t)\in\mathbb{R}$ and $y(x,t)\in\mathbb{R}$ are the input and output at time t, and $d(x,t)\in\mathbb{R}$ includes the unmodelled dynamics and the stochastic disturbance. For easy understanding, the integral operator is used for spatial operation and sum operator for temporal operation. In this study, only the single-input–single-output (SISO) system is considered. The extension of the results to the multi-input–multi-output (MIMO) system is straightforward.

The problem is to estimate N and G from the input–output data $\{u(\zeta_i, t), y(x_j, t)\}$, $(i=1,\ldots,n_u, j=1,\ldots,n_y, t=1,\ldots,n_t)$, where $\zeta_i, x_j\in\Omega$, n_u and n_y are the number of sampled spatial points of the input and output, and n_t is the time length. For simplicity, assume that the spatial points ζ_i and x_j are uniformly distributed over the spatial domain.

3. Basic identification approach

3.1. Basis function expansion

In general, the input $u(x, t)$ has finite degrees of freedom since only a finite number of actuators are available in practice. Thus assume that the input $u(x, t)$ can be formulated in terms of a finite number of spatial input basis functions $\{\psi_i(x)\}_{i=1}^{m}$ as follows

$$u(x,t)=\sum_{i=1}^{m}\psi_i(x)a_i(t), \tag{3}$$

where $a_i(t)=\int_{\Omega}u(x,t)\psi_i(x)\mathrm{d}x$ is the time coefficient (implemental input signal), $\psi_i(x)$ describes how the control action $a_i(t)$ is distributed in the spatial domain Ω, and m is the number of actuators, which can be determined by physical knowledge.

Ideally, the output $y(x,t)$ and the error $d(x,t)$ can be expressed by an infinite set of orthonormal spatial output basis functions $\{\varphi_i(x)\}_{i=1}^{\infty}$ as follows

$$y(x,t)=\sum_{i=1}^{\infty}\varphi_i(x)b_i(t), \tag{4}$$

$$d(x,t)=\sum_{i=1}^{\infty}\varphi_i(x)d_i(t), \tag{5}$$

where $b_i(t)=\int_{\Omega}y(x,t)\varphi_i(x)\mathrm{d}x$ and $d_i(t)=\int_{\Omega}d(x,t)\varphi_i(x)\mathrm{d}x$ are the time coefficients, respectively. This is because of inherently infinite dimensional characteristic of distributed parameter system. Practically, both $y(x, t)$ and $d(x, t)$ can be truncated into n dimensions as below

$$y_n(x,t)=\sum_{i=1}^{n}\varphi_i(x)b_i(t), \tag{6}$$

$$d_n(x,t)=\sum_{i=1}^{n}\varphi_i(x)d_i(t). \tag{7}$$

$\varphi_i(x)$ are usually selected as standard orthonormal functions such as Fourier series, Legendre polynomials, Jacobi polynomials and Chebyshev polynomials [9]. In this study, the KL decomposition [23] is used to identify the empirical dominant basis functions from the process data. Among all linear expansions, the KL expansion is the most efficient in the sense that for a given approximation error, the number of KL bases required is minimal. Owing to this, the KL decomposition can help to reduce the number of estimated parameters. Details of the KL method are presented in Appendix 1.

Assume that the intermediate output $v(x,t)\in\mathbb{R}$ can be described as

$$v(x,t)=N(u(x,t))=\sum_{i=1}^{m}\sum_{j=1}^{v}\psi_i(x)\beta_j h_j(a_i(t)), \tag{8}$$

where $h_j(\cdot):\mathbb{R}\to\mathbb{R}(j=1,\ldots,v)$ are nonlinear basis functions and $\beta_j\in\mathbb{R}(j=1,\ldots,v)$ are coefficients. Typically, the nonlinear functions $h_j(\cdot)$ can be chosen as polynomials, radial basis functions, wavelets [27] and so on.

Assuming that the kernel $g(x, \zeta, \tau)$ in (2) is absolutely integrable on time domain $[0,\infty)$ at any spatial point x and ζ, which means that the corresponding model is stable, then it can be represented by means of orthonormal temporal basis functions. Theoretically, the kernel is supposed to be expanded onto spatial output bases $\{\varphi_i(x)\}_{i=1}^{\infty}$, spatial input bases $\{\psi_i(x)\}_{i=1}^{m}$ and temporal bases $\{\phi_i(t)\}_{i=1}^{\infty}$ as follows

$$g(x,\zeta,\tau)=\sum_{i=1}^{\infty}\sum_{j=1}^{m}\sum_{k=1}^{\infty}\alpha_{i,j,k}\varphi_i(x)\psi_j(\zeta)\phi_k(\tau), \tag{9}$$

where $\alpha_{i,j,k}\in\mathbb{R}(i=1,\ldots,\infty,j=1,\ldots,m,k=1,\ldots,\infty)$ are constant coefficients of basis functions $\varphi_i(x)\psi_j(\zeta)\phi_k(\tau)$. Practically, a finite-dimensional truncation

$$g_{n,l}(x,\zeta,\tau)=\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{l}\alpha_{i,j,k}\varphi_i(x)\psi_j(\zeta)\phi_k(\tau), \tag{10}$$

is often good enough for a realistic approximation, where n and l are the dimension of output bases and temporal bases, respectively. $\phi_i(t)$ can be selected as Laguerre functions [32], Kautz functions [33] and generalized orthonormal basis functions [18]. Here, Laguerre functions are chosen for the development, due to their simplicity and robustness to the choice of sampling period and model order [32]. Details of Laguerre functions are shown in Appendix 2.

Substitution of (4), (5), (8) and (9) into (2) with a n-dimensional truncation of output bases will have

$$y_n(x,t)=\sum_{\tau=0}^{t}\int_{\Omega}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{\infty}\alpha_{i,j,k}\varphi_i(x)\psi_j(\zeta)\phi_k(\tau)\sum_{r=1}^{m}\times\sum_{s=1}^{v}\psi_r(\zeta)\beta_s h_s(a_r(t-\tau))\mathrm{d}\zeta+d_n(x,t),\tag{11}$$

To make the kernel $g_{n,l}(x, \zeta, \tau)$ explicit, (11) can be rewritten as

$$y_n(x,t)=\sum_{\tau=0}^{t}\int_{\Omega}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{l}\alpha_{i,j,k}\varphi_i(x)\psi_j(\zeta)\phi_k(\tau)\sum_{r=1}^{m}\times\sum_{s=1}^{v}\psi_r(\zeta)\beta_s h_s(a_r(t-\tau))\mathrm{d}\zeta+\tilde{d}_n(x,t),\tag{12}$$

where

$$\tilde{d}_n(x,t)=\sum_{i=1}^{n}\varphi_i(x)\tilde{d}_i(t),$$

$$\tilde{d}_i(t)=\sum_{\tau=0}^{t}\int_{\Omega}\sum_{j=1}^{m}\sum_{k=l+1}^{\infty}\alpha_{i,j,k}\psi_j(\zeta)\phi_k(\tau)\times\sum_{r=1}^{m}\sum_{s=1}^{v}\psi_r(\zeta)\beta_s h_s(a_r(t-\tau))\mathrm{d}\zeta+d_i(t).$$

3.2. Temporal model

Eq. (12) can be further simplified into

$$\sum_{i=1}^{n}\varphi_i(x)b_i(t)=\sum_{i=1}^{n}\varphi_i(x)\sum_{j=1}^{m}\sum_{k=1}^{l}\alpha_{i,j,k}\sum_{r=1}^{m}\times\sum_{s=1}^{v}\beta_s\psi_{j,r}\ell_{k,s,r}(t)+\sum_{i=1}^{n}\varphi_i(x)\tilde{d}_i(t),\tag{13}$$

where

$$\psi_{j,r}=\int_{\Omega}\psi_j(\zeta)\psi_r(\zeta)\mathrm{d}\zeta,\tag{14}$$

$$\ell_{k,s,r}(t)=\sum_{\tau=0}^{t}\phi_k(\tau)h_s(a_r(t-\tau)),\tag{15}$$

Using the Galerkin method [6], the projection of (13) onto the output basis functions $\varphi_h(x)$ $(h=1, \ldots, n)$ will lead to the following n equations

$$\sum_{i=1}^{n}\int_{\Omega}\varphi_h(x)\varphi_i(x)\mathrm{d}x b_i(t)=\sum_{i=1}^{n}\int_{\Omega}\varphi_h(x)\varphi_i(x)\mathrm{d}x\sum_{j=1}^{m}\sum_{k=1}^{l}\alpha_{i,j,k}\times\sum_{r=1}^{m}\sum_{s=1}^{v}\beta_s\psi_{j,r}\ell_{k,s,r}(t)+\sum_{i=1}^{n}\times\int_{\Omega}\varphi_h(x)\varphi_i(x)\mathrm{d}x\tilde{d}_i(t).$$

Since $\{\varphi_i(x)\}_{i=1}^{n}$ are orthonormal, then we have

$$b(t)=\sum_{j=1}^{m}\sum_{k=1}^{l}\sum_{s=1}^{v}\alpha_{j,k}\beta_s L_{j,k,s}(t)+\tilde{d}(t),\tag{16}$$

where

$$b(t)=[b_1(t),\cdots,b_n(t)]^T\in\mathbb{R}^n,\tag{17}$$

$$\tilde{d}(t)=[\tilde{d}_1(t),\cdots,\tilde{d}_n(t)]^T\in\mathbb{R}^n.\tag{18}$$

$$\alpha_{j,k}=[\alpha_{1,j,k},\cdots,\alpha_{n,j,k}]^T\in\mathbb{R}^n,\tag{19}$$

$$L_{j,k,s}(t)=\sum_{r=1}^{m}\psi_{j,r}\ell_{k,s,r}(t)\in\mathbb{R},\tag{20}$$

3.3. Least square estimation

Eq. (16) can be expressed in a linear regression form

$$b(t)=\Theta^T\Phi(t)+\tilde{d}(t),\tag{21}$$

where

$$\Theta=[\alpha_{1,1}\beta_1,\cdots,\alpha_{1,1}\beta_v,\cdots,\alpha_{m,l}\beta_1,\cdots,\alpha_{m,l}\beta_v]^T\in\mathbb{R}^{n\times mlv},\tag{22}$$

$$\Phi(t)=[L_{1,1,1}(t),\cdots,L_{1,1,v}(t),\cdots,L_{m,l,1}(t),\cdots,L_{m,l,v}(t)]^T\in\mathbb{R}^{mlv}.\tag{23}$$

In practice, u and y are uniformly sampled over the spatial domain. In this case, $b(t)$ can be computed from the pointwise data $y(x, t)$ using spline interpolation in the spatial domain. The accurate $a(t)=[a_1(t), \cdots, a_m(t)]^T$ can be obtained from $u(x, t)$ using the inversion operation of a matrix formed by the basis functions provided that $n_u \geqslant m$. Then, $\Phi(t)$ can be constructed from $a(t)$.

Considering n_t set of temporal data $\{\Phi(t), b(t)\}_{t=1}^{n_t}$, it is well known from [20] that by minimizing a quadratic criterion on the prediction errors

$$\widehat{\Theta}=\arg\min_{\Theta}\left\{\frac{1}{n_t}\sum_{t=1}^{n_t}\|b(t)-\Theta^T\Phi(t)\|^2\right\},\tag{24}$$

Θcan be estimated using the least squares method as follows

$$\widehat{\Theta}=\left(\frac{1}{n_t}\sum_{t=1}^{n_t}\Phi(t)\Phi^T(t)\right)^{-1}\left(\frac{1}{n_t}\sum_{t=1}^{n_t}\Phi(t)b^T(t)\right),\tag{25}$$

provided that the indicated inverse exists.

The next problem is how to estimate the parameters $\alpha_{j,k}$ $(j=1, \ldots, m,\ k=1, \ldots, l)$ and β_s $(s=1, \ldots, v)$ from the estimate $\widehat{\Theta}$ in (25).

3.4. Singular value decomposition

For convenience, we define $\alpha=[\alpha_{1,1}^T,\cdots,\alpha_{m,l}^T]^T\in\mathbb{R}^{nml}$ and $\beta=[\beta_1,\cdots,\beta_v]^T\in\mathbb{R}^v$. It is clear that the parametrization (8) and (9) is not unique, since any parameter vectors $\alpha\sigma$and $\beta\sigma^{-1}$, for some nonzero constant σ, provide the same input/output equation (12). A technique that can be used to obtain uniqueness is to normalize the parameter vectors α (or β), for instance assuming that $\|\beta\|_2=1$. Under this assumption the parametrization (8) and (9) is unique.

From the definition of the parameter matrix Θ in (22), it is easy to see that

$$\Theta = \text{blockvec}(\Theta_{\alpha\beta}),$$

where blockvec($\Theta_{\alpha\beta}$) is the block column matrix obtained by stacking the block columns of $\Theta_{\alpha\beta}$ on the top of each other, and $\Theta_{\alpha\beta} \in \mathbb{R}^{v\times nml}$ has been defined as

$$\Theta_{\alpha\beta} \triangleq \begin{bmatrix} \beta_1\alpha_{1,1}^T & \cdots & \beta_1\alpha_{m,l}^T \\ \beta_2\alpha_{1,1}^T & \cdots & \beta_2\alpha_{m,l}^T \\ \vdots & \ddots & \vdots \\ \beta_v\alpha_{1,1}^T & \cdots & \beta_v\alpha_{m,l}^T \end{bmatrix} = \beta\alpha^T. \tag{26}$$

Thus an estimate $\widehat{\Theta}_{\alpha\beta}$ of the matrix $\Theta_{\alpha\beta}$ can be obtained from the estimate $\widehat{\Theta}$ in (25). The problem now is how to estimate the parameter matrices α and β from $\widehat{\Theta}_{\alpha\beta}$.

In order to solve this problem, an important fact should be uncovered first. It is clear that the closest, in the Frobenius norm sense, approximation of $\widehat{\Theta}_{\alpha\beta}$ is not just a single pair of $\hat{\alpha}$ and $\hat{\beta}$ but a series of pairs $(\hat{\beta}^c, \hat{\alpha}^c), (c = 1, \ldots, p)$ that solve the following optimization problem

$$(\hat{\beta}^c, \hat{\alpha}^c) = \arg\min_{\alpha^c, \beta^c} \left\{ \left\| \widehat{\Theta}_{\alpha\beta} - \sum_{c=1}^{p} \beta^c (\alpha^c)^T \right\|_F^2 \right\}, \tag{27}$$

where the Frobenius norm of a matrix $A \in \mathbb{R}^{m\times n}$ is defined as $\|A\|_F = (\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2)^{1/2}$.

To illustrate this fact, a lemma [15] should be introduced.

Lemma 1. *Let $\widehat{\Theta}_{\alpha\beta} \in \mathbb{R}^{v\times nml}$ have rank $\gamma \geqslant 1$, and let the economy-size SVD of $\widehat{\Theta}_{\alpha\beta}$ be given by*

$$\widehat{\Theta}_{\alpha\beta} = U_\gamma \Sigma_\gamma V_\gamma^T = \sum_{i=1}^{\gamma} \sigma_i \mu_i \upsilon_i^T, \tag{28}$$

where the singular matrix $\Sigma_\gamma = diag\{\sigma_i\}$ such that

$$\sigma_1 \geqslant \cdots \geqslant \sigma_\gamma > 0,$$

and where the matrices $U_\gamma = [\mu_1, \ldots, \mu_\gamma] \in \mathbb{R}^{v\times\gamma}$ and $V_\gamma = [\upsilon_1, \ldots, \upsilon_\gamma] \in \mathbb{R}^{nml\times\gamma}$ contain only the first γ columns of the unitary matrices $U \in \mathbb{R}^{v\times v}$ and $V \in \mathbb{R}^{nml\times nml}$ provided by the full SVD of $\widehat{\Theta}_{\alpha\beta}$,

$$\widehat{\Theta}_{\alpha\beta} = U\Sigma V^T,$$

respectively. Then $\forall p \leqslant \gamma$, the following equation holds

$$\begin{aligned}(\hat{\beta}^c, \hat{\alpha}^c) &= \arg\min_{\alpha^c, \beta^c} \left\{ \left\| \widehat{\Theta}_{\alpha\beta} - \sum_{c=1}^{p} \beta^c (\alpha^c)^T \right\|_F^2 \right\} \\ &= (\mu_c, \upsilon_c\sigma_c), (c = 1, \ldots, p),\end{aligned} \tag{29}$$

where $(\hat{\beta}^c, \hat{\alpha}^c)$ is defined as the cth channel, and p is the number of channels. The parameter approximation error is given by

$$\varepsilon^p = \left\| \widehat{\Theta}_{\alpha\beta} - \sum_{c=1}^{p} \beta^c (\alpha^c)^T \right\|_F^2 = \sum_{c=p+1}^{\gamma} \sigma_c^2. \tag{30}$$

Based on Lemma 1, the estimated parameters $\hat{\alpha}$ and $\hat{\beta}$ can be obtained by

$$(\hat{\beta}, \hat{\alpha}) = \arg\min_{\alpha,\beta} \{ \| \widehat{\Theta}_{\alpha\beta} - \beta(\alpha)^T \|_F^2 \} = (\mu_1, \upsilon_1\sigma_1). \tag{31}$$

The consistency in the previous work [1,16] can be extended to this basic identification approach under certain conditions (e.g., model matching and zero-mean disturbance). However, if such conditions are not satisfied, a multi-channel identification approach proposed in the following section may provide a better solution.

4. Multi-channel identification approach

Definition 1. The system (2) is named as a single-channel Hammerstein system. The multi-channel Hammerstein system is formed by the parallel connection of p single-channel Hammerstein systems.

For a single-channel Hammerstein model, we can see from (26) that $rank(\Theta_{\alpha\beta}) = 1$, since $\Theta_{\alpha\beta}$ is the product of a column vector β and a row vector α^T. However, generally speaking, its estimate $\widehat{\Theta}_{\alpha\beta}$ from process data cannot be exactly expressed as the product of a column vector and a row vector due to unmodeled dynamics and disturbance. That is $rank(\widehat{\Theta}_{\alpha\beta}) > 1$. In addition, σ_2/σ_1 (see (28)) cannot always be small enough to make the parameter approximation error (30) acceptable. Thus, it is very necessary to add more channels to compensate the modeling residuals.

The proposed multi-channel identification methodology is shown in Fig. 2.

A sequential identification algorithm is designed as follows. Firstly, the 1st channel model is estimated using the basic identification algorithm from the input–output data $\{u(x,t), y(x,t)\}_{t=1}^{n_t}$. Secondly, the 1st channel model error $e^1(x,t) = y(x,t) - \hat{y}_n^1(x,t)$ is regarded as the new output, and then the 2nd channel model is identified. Similarly, $e^2(x,t), \cdots, e^{p-1}(x,t)$ can determine the 3rd, $\cdots$, pth channel models, and so on.

However the sequential identification algorithm for each channel may lead to a computational burden problem. According to Lemma 1, the multi-channel identification

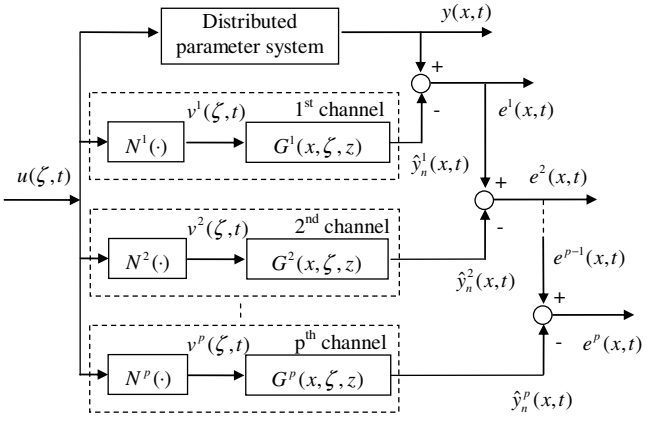

Fig. 2. Multi-channel identification of spatio-temporal Hammerstein model.

algorithm can be easily implemented simultaneously as below

Algorithm 1

Step 1 : Determine the input basis functions $\{\psi_i(x)\}_{i=1}^{m}$, find the output basis functions $\{\varphi_i(x)\}_{i=1}^{n}$ using the KL method, choose the Laguerre polynomials $\{\phi_i(t)\}_{i=1}^{l}$, then obtain the corresponding temporal coefficients $\{a(t)\}_{t=1}^{n_t}$ and $\{b(t)\}_{t=1}^{n_t}$ of the input $\{u(x_i,t)\}_{i=1,t=1}^{n_u,n_t}$ and output $\{y(x_j,t)\}_{j=1,t=1}^{n_y,n_t}$ respectively.

Step 2 : Compute the linear regressors $\Phi(t)$ according to (15) and (23) using $\{a(t)\}_{t=1}^{n_t}$, then compute the least squares estimate $\widehat{\Theta}$ as in (25), and construct the matrix $\widehat{\Theta}_{\alpha\beta}$ such that $\widehat{\Theta} = \text{blockvec}(\widehat{\Theta}_{\alpha\beta})$.

Step 3 : Compute the economy-size SVD of $\widehat{\Theta}_{\alpha\beta}$ as in Lemma 1, and the partition of this decomposition as in (28).

Step 4 : Compute the estimates of the parameter vectors $\hat{\alpha}^c$ and $\hat{\beta}^c$ as $\hat{\beta}^c = \mu_c$ and $\hat{\alpha}^c = \upsilon_c\sigma_c (c = 1,\ldots,p)$, respectively.

Remark 1. It is important to note that the algorithm intrinsically delivers estimates that satisfy the uniqueness condition $\|\hat{\beta}^c\|_2 = 1$, since the matrix μ_c in the SVD of $\widehat{\Theta}_{\alpha\beta}$ is a unitary matrix.

4.1. Multi-channel Hammerstein model

Based on Algorithm 1, a multi-channel spatio-temporal Hammerstein model consisting of p channels

$$\hat{y}_n(x,t) = \sum_{c=1}^{p}\sum_{\tau=0}^{t}\int_{\Omega}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{l}\hat{\alpha}^c_{i,j,k}\varphi_i(x)\psi_j(\zeta)\phi_k(\tau) \times \sum_{r=1}^{m}\sum_{s=1}^{v}\psi_r(\zeta)\hat{\beta}^c_s h_s(a_r(t-\tau))\mathrm{d}\zeta, \quad (32)$$

is constructed to approximate the nonlinear DPS as shown in Fig. 3. Each channel consists of the cascade connection of a static nonlinear block represented by basis functions, followed by a dynamic linear block represented by spatio-temporal Laguerre model as shown in Fig. 4. The transfer functions in Fig. 4 can be derived from (48) in Appendix 2 as follows

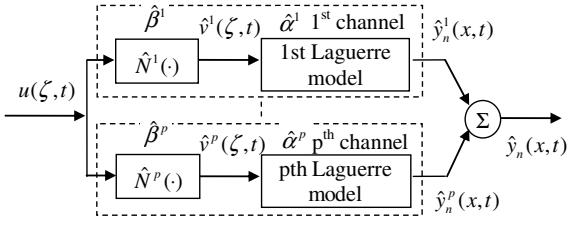

Fig. 3. Multi-channel spatio-temporal Hammerstein model.

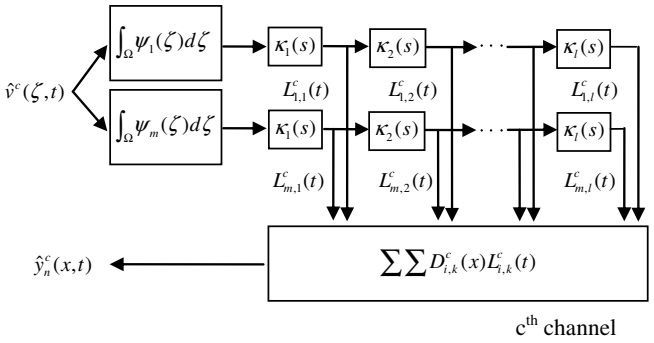

Fig. 4. Spatio-temporal Laguerre model of the cth channel.

$$\kappa_1(s) = \frac{\sqrt{2\xi}}{s+\xi}, \kappa_2(s) = \cdots = \kappa_l(s) = \frac{s-\xi}{s+\xi},$$

where ξ is the time-scaling factor.

Note that in Figs. 3 and 4, $\hat{y}_n(x,t) = \sum_{c=1}^{p}\hat{y}^c_n(x,t)$, $\hat{v}^c(\zeta,t) = \sum_{r=1}^{m}\sum_{s=1}^{v}\psi_r(\zeta)\hat{\beta}^c_s h_s(a_r(t))$, $\hat{D}^c_{j,k}(x) = \sum_{i=1}^{n}\hat{\alpha}^c_{i,j,k}\varphi_i(x)$ and $L^c_{j,k}(t) = \sum_{r=1}^{m}\sum_{s=1}^{v}\psi_{j,r}\hat{\beta}^c_s\ell_{k,s,r}(t)$.

Remark 2. If the time scale ξ is chosen suitably, then the Laguerre series can efficiently model any stable linear plant [34]. Usually, the parameter ξ which gives a good performance is obtained from trials. Many studies have dealt with the time scale selection problem using such as offline optimization [3] and online adaptation [30] methods. For another parameter l, there are no theoretical methods but only some empirical ones so far.

Based on Lemma 1, we can give the following theorem to show the advantage of multi-channel mechanism.

Theorem 1. *For a spatio-temporal Hammerstein system (2), if the estimated matrix $\widehat{\Theta}_{\alpha\beta}$ has rank $\gamma \geqslant 1$ and the parameters of the cth channel $(\hat{\beta}^c, \hat{\alpha}^c)(c = 1,\ldots,p, p \leqslant \gamma)$ are obtained by (29), then the parameter approximation error (30) will satisfy*

$$\varepsilon^1 > \varepsilon^2 > \cdots \varepsilon^p > \cdots > \varepsilon^\gamma = 0.$$

Proof. This can be easily drawn from Lemma 1. □

Theorem 1 means that the parameter approximation error will be reduced by increasing the channel number p. Moreover, the model complexity can also be controlled by the number of channels. There is a tradeoff between the complexity and accuracy. Due to the property of the SVD, the parameter of the cth channel is the cth principal component of the whole parameter space. Therefore, only the first few dominant channels can construct a good model.

An important issue is the convergence of the estimated parameters as the number of data points n_t tends to infinity. Now, we will give a convergence theorem to support the proposed algorithm.

For simplicity, let $y_n(x,t) = H(x,t,\Theta,\{u(\zeta,\tau)\})$ denote a multi-channel Hammerstein model with $n, l, v < \infty$

$$y_n(x,t)=\sum_{c=1}^{p}\sum_{\tau=0}^{t}\int_{\Omega}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{l}\alpha_{i,j,k}^{c}\varphi_i(x)\psi_j(\zeta)\phi_k(\tau)\times\sum_{r=1}^{m}\sum_{s=1}^{v}\psi_r(\zeta)\beta_s^c h_s(a_r(t-\tau))\mathrm{d}\zeta, \tag{33}$$

where $\Theta=\sum_{c=1}^{p}[\alpha_{1,1}^{c}\beta_1^c,\ldots,\ \alpha_{1,1}^{c}\beta_v^c,\ldots,\alpha_{m,l}^{c}\beta_1^c,\ldots,\alpha_{m,l}^{c}\beta_v^c]^T\in\mathbb{R}^{n\times ml v}$, $\alpha_{j,k}^{c}(j=1,\ldots,m,k=1,\ldots,l)$ are defined by (19), and $\{u(\zeta,\tau)\}=\{u(\zeta,\tau)|\zeta\in\Omega,\tau=1,\ldots,t\}$.

We always assume that there is an optimal model

$$y_n^*(x,t)=H(x,t,\Theta^*,\{u(\zeta,\tau)\}), \tag{34}$$

with an optimal parameter matrix Θ^* such that

$$\Theta^*=\arg\min_{\Theta\in D_\Theta}\{\overline{E}(y(x,t)-H(x,t,\Theta,\{u(\zeta,\tau)\}))^2\}, \tag{35}$$

where $\overline{E}f(x,t)^2=\lim_{n_t\to\infty}\frac{1}{n_t}\sum_{t=1}^{n_t}\frac{1}{A}\int_{\Omega}Ef(x,t)^2\mathrm{d}x$, $A=\int_{\Omega}\mathrm{d}x$ and E is expectation operator. Let D_Θ be compact. Define $\Theta_{\alpha\beta}^*\in\mathbb{R}^{v\times nml}$ such that $\Theta^*=\mathrm{blockvec}(\Theta_{\alpha\beta}^*)$.

Under the uniform spatial discretization, $\frac{1}{A}\sum_{i=1}^{\infty}b_i(t)^2=\frac{1}{A}\int_{\Omega}y(x,t)^2\mathrm{d}x$ can be replaced by $\frac{1}{n_y}\sum_{j=1}^{n_y}y(x_j,t)^2$. However, the accurate $a(t)$ can be obtained provided that $n_u\geqslant m$. Therefore, according to the details of the developed identification algorithm, the minimization problem (24) is indeed equivalent to the following problem

$$\widehat{\Theta}=\arg\min_{\Theta\in D_\Theta}\left\{\frac{1}{n_t}\frac{1}{n_y}\sum_{t=1}^{n_t}\sum_{j=1}^{n_y}(y(x_j,t)-H(x_j,t,\Theta,\{u(\zeta,\tau)\}))^2\right\}. \tag{36}$$

It should be mentioned that, (24) can be considered as a practical implementation of (36) in order to reduce the involved spatial complexity. However, the theoretical analysis should be performed in the spatio-temporal domain.

Assumption 1. Let $W(x,t)$ be the σ-algebra generated by $(d(x,t),\cdots,d(x,0))$. For each t, τ ($t\geqslant\tau$) and any $x,\zeta\in\Omega$, there exist random variables $y_\tau^0(x,t)(y_t^0(x,t)=0)$, $u_\tau^0(\zeta,t)(u_t^0(\zeta,t)=0)$, that belong to $W(x,t)$, but are independent of $W(x,\tau)$, such that

$$E|y(x,t)-y_\tau^0(x,t)|^4<M\lambda^{t-\tau},$$
$$E|u(\zeta,t)-u_\tau^0(\zeta,t)|^4<M\lambda^{t-\tau},$$

for some $M<\infty$, $\lambda<1$.

Assumption 2. Assume that the model $y_n(x,t)=H(x,t,\Theta,\{u(\zeta,\tau)\})$ is differentiable with respect to Θ for all $\Theta\in D_\Theta$. Assume that

$$|H(x,t,\Theta,\{u_1(\zeta,\tau)\})-H(x,t,\Theta,\{u_2(\zeta,\tau)\})|\leqslant M\sum_{\tau=0}^{t}\lambda^{t-\tau}\sup_{\zeta\in\Omega}|u_1(\zeta,\tau)-u_2(\zeta,\tau)|,$$

and $|H(x,t,\Theta,\{0(\zeta,\tau)\})|\leqslant M$, where Θ belong to an open neighborhood of D_Θ, $M<\infty$ and $\lambda<1$.

Assumption 3. Define $\varepsilon(x,t,\Theta)=y(x,t)-H(x,t,\Theta,\{u(\zeta,\tau)\})$ and there exists

$$\left|\frac{\partial\varepsilon(x,t,\Theta)^2}{\partial\Theta}\right|\leqslant M\varepsilon(x,t,\Theta)^2,\Theta\in D_\Theta,\quad\forall x\in\Omega,\ \forall t.$$

Remark 3. Assumption 1 means that the system to be identified is exponentially stable, i.e., the remote past of the process is "forgotten" at an exponential rate. Assumption 2 has three meanings. Firstly, the model is differential with respect to the parameters. Secondly, the model may not increase faster than the linear one. Thirdly, the model is also exponentially stable. Regarding Assumption 3, the derivative of the modeling error with respect to the parameters is bounded by the modeling error. Such conditions are required to make the parameter optimization procedure feasible, and guarantee the following convergence.

Theorem 2. *For a spatio-temporal Hammerstein system (2), the multi-channel model (32) is estimated using Algorithm 1. If Assumptions 1–3 are satisfied, then* $\sum_{c=1}^{p}\hat{\beta}^c(\hat{\alpha}^c)^T\to\Theta_{\alpha\beta}^*$ *and* $\hat{y}_n(x,t)\to y_n^*(x,t)$ *w. p. 1 as* $n_t\to\infty$, $n_y\to\infty$, *and* $p\to\gamma$, *where* $\gamma=rank(\Theta_{\alpha\beta}^*)$.

Proof. The proof is given in Appendix 3. □

5. Case study

In order to evaluate the proposed modeling method, two typical distributed processes are studied.

For an easy comparison, some performance indexes are set up as follows:

- Spatio-temporal error $e(x,t)=y(x,t)-\hat{y}_n(x,t)$,
- Spatial normalized absolute error, $SNAE(t)=\int|e(x,t)|\mathrm{d}x/\int\mathrm{d}x$,
- Temporal normalized absolute error, $TNAE(x)=\sum|e(x,t)|/\sum\Delta t$,
- Root of mean squared error, $RMSE=\left(\int\sum e(x,t)^2\mathrm{d}x/\int\mathrm{d}x\sum\Delta t\right)^{1/2}$.

5.1. Catalytic rod

A long thin rod in a reactor shown in Fig. 5 is a typical transport-reaction process in chemical industry [6]. The reactor is fed with pure species A and a zero-th order exothermic catalytic reaction of the form $A\to B$ takes place in the rod. Since the reaction is exothermic, a cooling medium that is in contact with the rod is used for cooling.

Under the assumptions of constant density and heat capacity of the rod, constant conductivity of the rod, and constant temperature at both sides of the rod, and excess of species A in the furnace, the mathematical model which describes the spatio-temporal evolution of the rod temperature consists of the following parabolic PDE [6]:

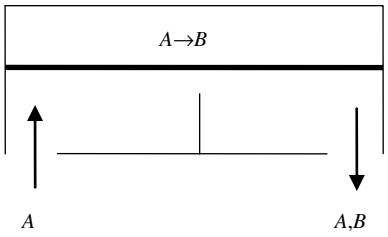

Fig. 5. Catalytic rod.

$$\frac{\partial y(x,t)}{\partial t} = \frac{\partial^2 y(x,t)}{\partial x^2} + \beta_T\left(e^{-\frac{\gamma}{1+y}} - e^{-\gamma}\right) + \beta_u(\psi(x)^T a(t) - y(x,t)), \quad (37)$$

subject to the Dirichlet boundary and initial conditions:

$$y(0,t) = 0, y(\pi,t) = 0, \quad (38)$$

$$y(x,0) = 0, \quad (39)$$

where $y(x, t)$, $\psi(x)$, β_T, β_u, γ and a denote the temperature in the reactor, the actuator distribution, the heat of reaction, the heat transfer coefficient, the activation energy, and the manipulated input (temperature of the cooling medium).

The process parameters are often set as

$$\beta_T = 50, \beta_u = 2, \gamma = 4.$$

There are available four control actuators $a(t) = [a_1(t), \cdots, a_4(t)]^T$ with the spatial distribution function $\psi(x) = [\psi_1(x), \cdots, \psi_4(x)]^T$, where $a_i(t) = 1.1 + 5\sin(t/10 + i/10)$ $(i = 1, \ldots, 4)$ and $\psi_i(x) = H(x - (i-1)\pi/4) - H(x - i\pi/4)$ $(i = 1, \ldots, 4)$ ($H(\cdot)$ being the standard Heaviside function). Due to the infinite-dimensional feature, sufficient sensors should be used to measure the representative spatial features of the distributed parameter system, which depends on the required modeling accuracy. In this case, nineteen sensors are used. The random process noise is bounded by 0.001 with zero-mean. The sampling period Δt is 0.01 and the simulation time is 5.

The process output $y(x, t)$ is shown in Fig. 6, while the obtained KL basis functions are shown in Fig. 7 with $n = 5$. The temporal bases $\phi_i(t)$, $(i = 1, \ldots, 10)$ are chosen as Laguerre series with time-scaling factor $\xi = 2.05$. The nonlinear bases are polynomials as $h_i(a) = a^i$ $(i = 1, \ldots, 4)$.

The prediction output $\hat{y}_n(x,t)$ of the *3-channel Hammerstein model* is shown in Fig. 8, with the prediction error $e(x, t)$ presented in Fig. 9. It is obvious that the *3-channel Hammerstein model* can satisfactorily model the original process. As shown in Figs. 10 and 11, the prediction error $SNAE(t)$ and $TNAE(x)$ will be smaller when using more channels. It is verified that the *multi-channel Hammerstein modeling* approach can improve the performance of the *single-channel model*.

Finally, the influence of the sensor number on the modeling error *RMSE* is studied. As seen from Fig. 12, the *3-channel Hammerstein modeling* performance will be better as the number of sensors increases since the more spatial information is available. However, little improvement is seen if more than 19 sensors are used.

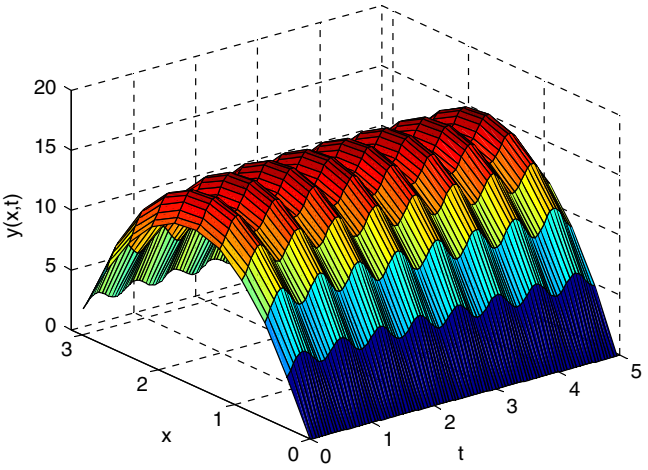

Fig. 6. Process output.

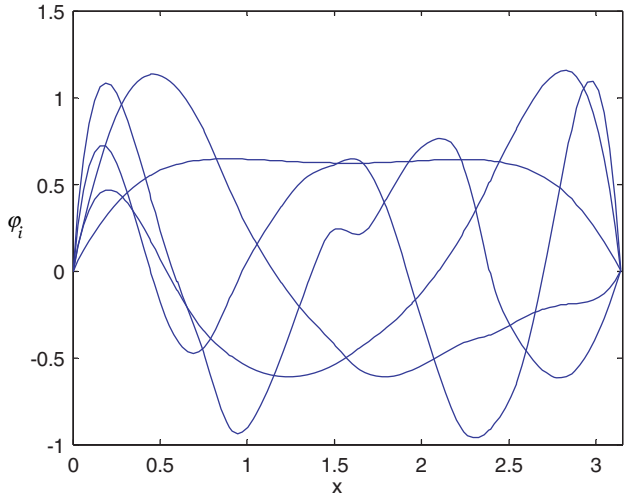

Fig. 7. Karhunen–Loève basis functions.

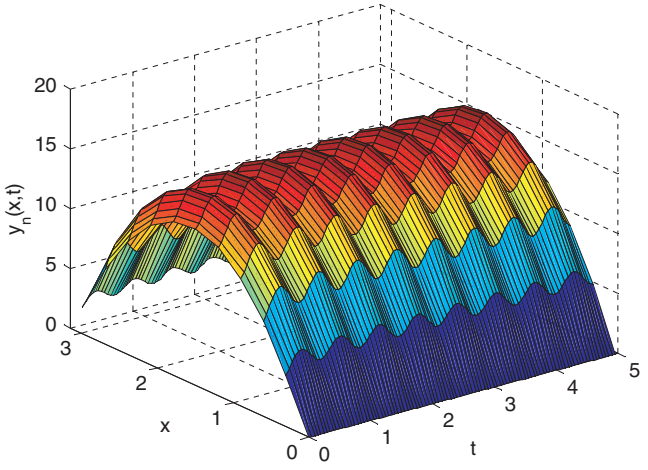

Fig. 8. Prediction output of 3-channel Hammerstein model.

Please cite this article in press as: C. Qi et al., A multi-channel spatio-temporal Hammerstein modeling approach ..., J. Process Contr. (2008), doi:10.1016/j.jprocont.2008.01.006

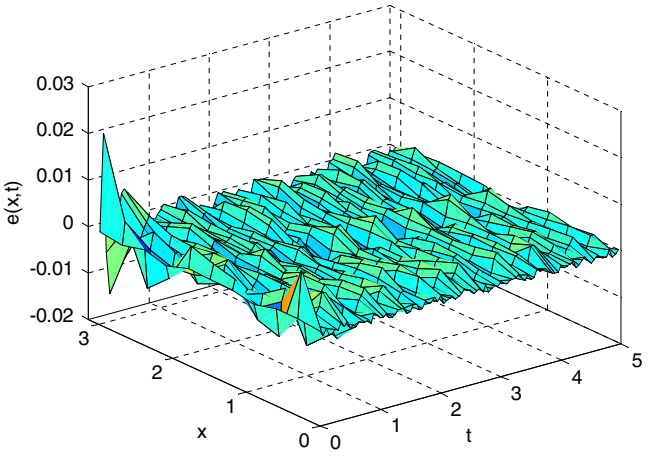

Fig. 9. Prediction error of 3-channel Hammerstein model.

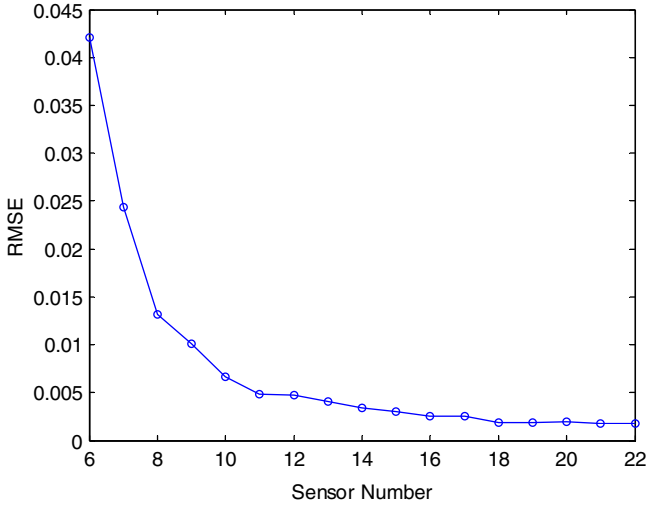

Fig. 12. RMSE of 3-channel Hammerstein model.

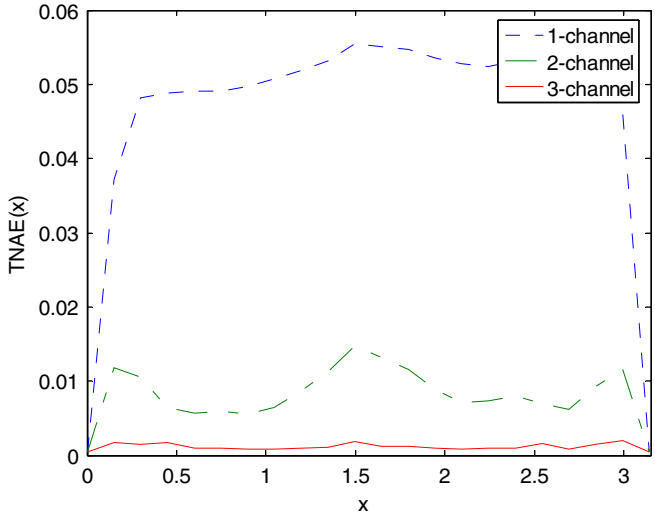

Fig. 10. TNAE(x) of Hammertein models.

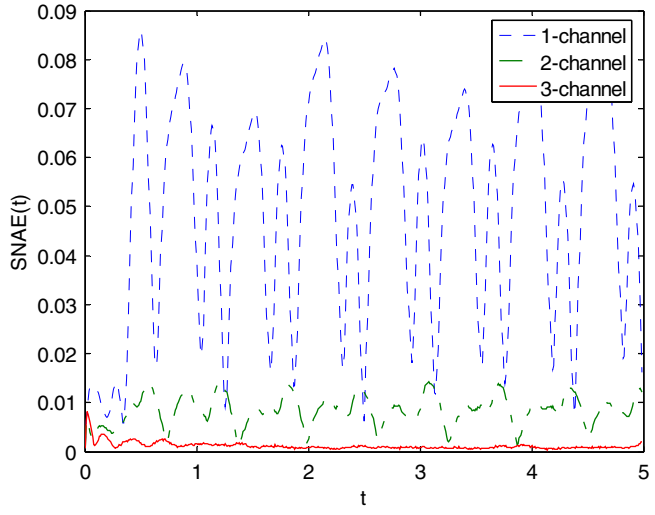

Fig. 11. SNAE(t) of Hammertein models.

5.2. Catalytic packed-bed reactor

Consider the temperature distribution in a long, thin non-isothermal catalytic packed-bed reactor as shown in Fig. 13 [5]. A reaction of the form $C \rightarrow D$ takes place on the catalyst. The reaction is endothermic and a jacket is used to heat the reactor.

A dimensionless model that describes this nonlinear tubular chemical reactor is provided as follows

$$\begin{aligned} \varepsilon_p \frac{\partial y_g}{\partial t} &= -\frac{\partial y_g}{\partial x} + \alpha_c(y - y_g) - \alpha_g(y_g - \psi(x)a(t)), \\ \frac{\partial y}{\partial t} &= \frac{\partial^2 y}{\partial x^2} + \beta_0 e^{\frac{\gamma y}{1+y}} - \beta_c(y - y_g) - \beta_p(y - \psi(x)a(t)), \end{aligned} \tag{40}$$

subject to the boundary conditions

$$\begin{aligned} &x = 0, y_g = 0, \frac{\partial y}{\partial x} = 0, \\ &x = 1, \frac{\partial y}{\partial x} = 0, \end{aligned} \tag{41}$$

where y_g, y and a denote the dimensionless temperature of the gas, the catalyst and jacket, respectively. It is assumed that only catalyst temperature measurements are available.

The values of the process parameters are given below

$$\varepsilon_p = 0.01, \gamma = 21.14, \beta_c = 1.0, \beta_p = 15.62, \beta_0 = -0.003, \alpha_c = 0.5 \quad \text{and} \quad \alpha_g = 0.5.$$

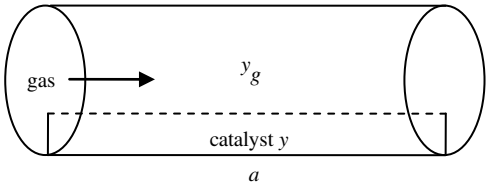

Fig. 13. Catalytic packed-bed reactor.

A heater is used with the spatial distribution $\psi(x) = \sin(\pi x)$, $0 \leqslant x \leqslant 1$. In the numerical calculation, we set the input $a(t) = 1.1 + 1.5\sin(t/20)$. In the simulations, sixteen sensors are used in order to capture the sufficient spatial information. The random process noise is bounded by 0.0005 with zero-mean. The sampling period Δt is 0.0001 and the simulation time is 0.1.

The process output $y(x, t)$ is shown in Fig. 14, while the obtained KL basis functions are shown in Fig. 15 with $n = 3$. The temporal bases $\phi_i(t)$, $(i = 1, \ldots, 10)$ are chosen as Laguerre series with time-scaling factor $\xi = 20.5$. The nonlinear bases are polynomials as $h_i(a) = a^i$ $(i = 1, \ldots, 4)$.

The prediction output $\hat{y}_n(x, t)$ of the *3-channel Hammerstein model* is shown in Fig. 16, with the prediction error $e(x, t)$ presented in Fig. 17. It is obvious that the*3-channel Hammerstein model* can approximate the original spatio-temporal dynamics very well. As shown in Figs. 18 and 19, as the channel number increases, the prediction error $SNAE(t)$ and $TNAE(x)$ will decrease, which is consistent with the theoretical analysis. As illustrated in Fig. 20, the

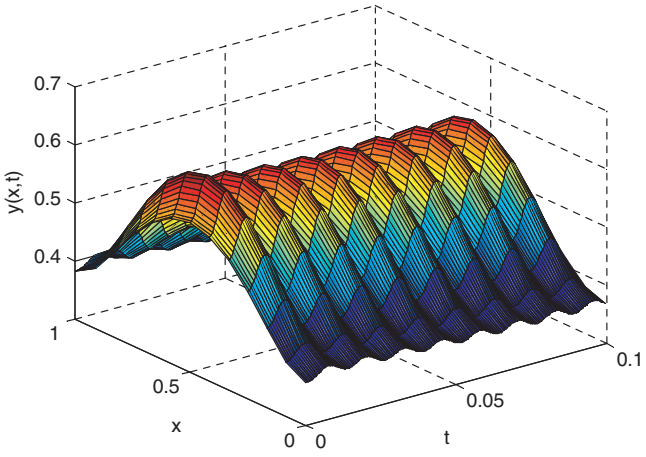

Fig. 14. Process output.

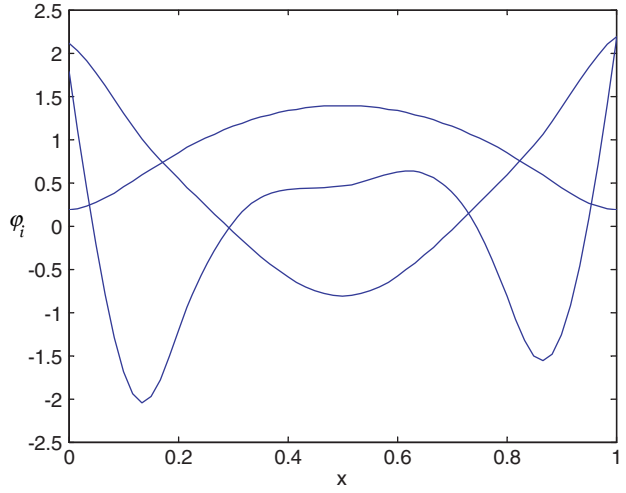

Fig. 15. Karhunen–Loève basis functions.

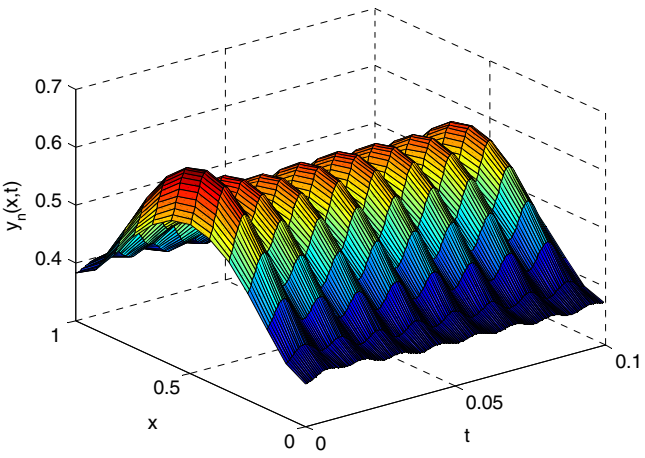

Fig. 16. Prediction output of 3-channel Hammerstein model.

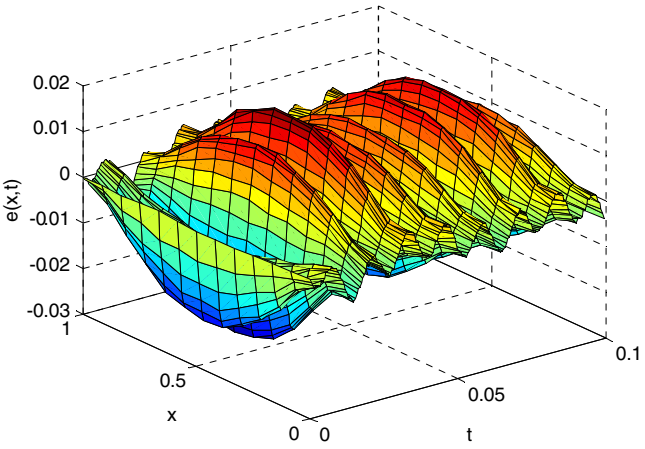

Fig. 17. Prediction error of 3-channel Hammerstein model.

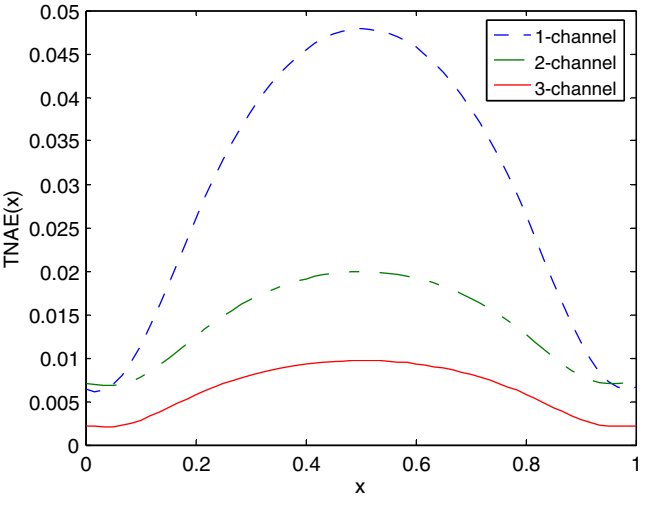

Fig. 18. TNAE(x) of Hammertein models.

3-channel Hammerstein modeling error *RMSE* will become smaller when using more sensors, and but more than 16 sensors will have less effect on the modeling performance.

Please cite this article in press as: C. Qi et al., A multi-channel spatio-temporal Hammerstein modeling approach ..., J. Process Contr. (2008), doi:10.1016/j.jprocont.2008.01.006

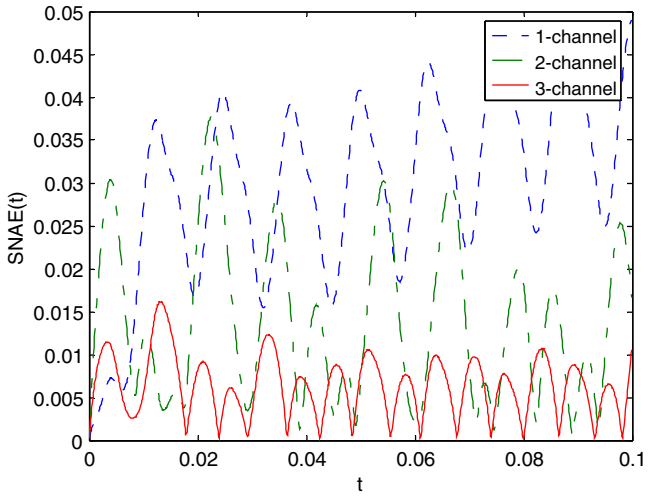

Fig. 19. SNAE(t) of Hammertein models.

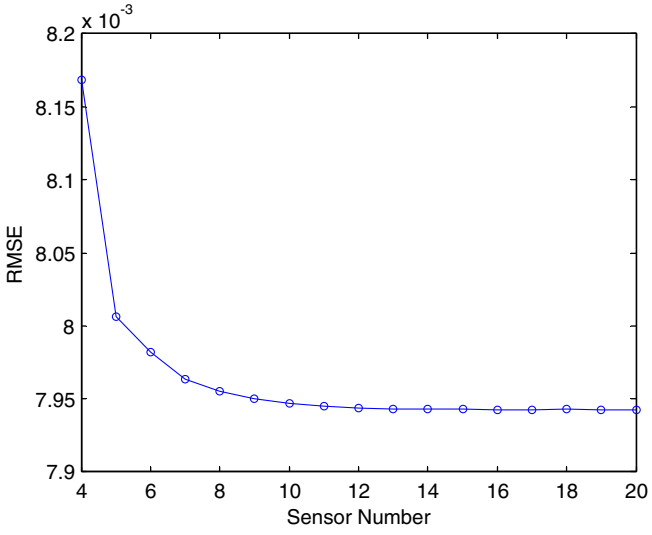

Fig. 20. RMSE of 3-channel Hammerstein model.

6. Conclusion

A spatio-temporal Hammerstein modeling approach is proposed for nonlinear distributed parameter systems. The Hammerstein model consists of the static nonlinear and the distributed dynamical linear parts. Using the Galerkin method with the expansion onto KL spatial bases and Laguerre temporal bases, the spatio-temporal modeling is reduced to a traditional temporal modeling problem. The unknown parameters can be easily estimated using the least squares estimation and the singular value decomposition. In the presence of unmodeled dynamics, a multi-channel modeling approach is proposed to reduce the single-channel modeling error. This modeling method provides convergent estimates under some conditions. The simulations of two examples are presented to show the effectiveness of this modeling method and its potential to industrial applications.

Acknowledgements

Authors would like to thank the associate editor and the anonymous referees for their valuable comments and suggestions. The work is partially supported by a grant from RGC of Hong Kong (CityU: 116905), a SRG grant from City University of Hong Kong (7002219), and National Natural Science Foundation of China (NNSFC) with granting No. 60704041.

Appendix 1. Karhunen–Loève method

The basic idea of the KL expansion is to find those modes which represent the dominant character of the system from process data.

Suppose we have n_t observations $y(x, t)$ at time $t = 1, \ldots, n_t$ (called snapshots). The problem is compute the most characteristic structure $\varphi(x)$ among these snapshots $y(x, t)$ that can be formulated as the one of maximizing the following objective function [6]

$$\begin{aligned} &\text{maximize} \frac{\langle (\varphi(x), y(x,t))^2 \rangle}{(\varphi(x), \varphi(x))} \\ &\text{subject to } (\varphi(x), \varphi(x)) = 1, \varphi(x) \in L^2(\Omega). \end{aligned} \tag{42}$$

where the ensemble average and the inner product are defined as $\langle f(x,t) \rangle = \frac{1}{n_t} \sum_{t=1}^{n_t} f(x,t)$, $(\varphi_1(x), \varphi_2(x)) = \int_\Omega \varphi_1(x)\varphi_2(x)\mathrm{d}x$, respectively. The constraint $(\varphi, \varphi) = 1$ is imposed to ensure that the function $\varphi(x)$ is unique. The Lagrangian functional corresponding to this constrained optimization problem is

$$J = \langle (\varphi(x), y(x,t))^2 \rangle - \lambda((\varphi, \varphi) - 1),$$

and necessary condition for extremes is that the functional derivative vanishes for all variation $\varphi + \delta\psi \in L^2(\Omega)$, where δ is a real number:

$$\frac{\mathrm{d}J(\varphi + \delta\psi)}{\mathrm{d}\delta}(\delta = 0) = 0, (\varphi, \varphi) = 1.$$

Using the definitions of inner product and ensemble average, computing $\mathrm{d}J(\varphi + \delta\psi)/\mathrm{d}\delta$ $(\delta = 0)$, and using that $\psi(x)$ is an arbitrary function, the following necessary conditions for optimality can be obtained:

$$\int_\Omega \langle y(x,t), y(\zeta,t) \rangle \varphi_i(\zeta)\mathrm{d}\zeta = \lambda_i \varphi_i(x), (\varphi, \varphi) = 1. \tag{43}$$

where $\varphi_i(x)$ is the ith eigenfunction, λ_i is the corresponding eigenvalue.

A computationally efficient way to obtain the solution of the above integral equation is provided by the method of snapshots [26,22] where the requisite eigenfuction $\varphi_i(x)$ is expressed as a linear combination of the snapshots as follows

$$\varphi_i(x) = \sum_{t=1}^{n_t} \gamma_{ti} y(x,t), \tag{44}$$

Substituting (44) into (43) gives the following eigenvalue problem:

$$\int_{\Omega} \frac{1}{n_t} \sum_{t=1}^{n_t} y(x,t) y(\zeta,t) \sum_{\tau=1}^{n_t} \gamma_{\tau i} y(\zeta,\tau) \mathrm{d}\zeta = \lambda_i \sum_{t=1}^{n_t} \gamma_{ti} y(x,t), \tag{45}$$

Defining

$$Y_{t\tau} = \frac{1}{n_t} \int_{\Omega} y(\zeta,t) y(\zeta,\tau) \mathrm{d}\zeta,$$

The eigenvalue problem of (45) can be equivalently written as

$$Y\gamma_i = \lambda_i \gamma_i, \tag{46}$$

where γ_i is the ith eigenvector with elements γ_{ti} which can be used in (44) to construct the eigenfunctions $\varphi_i(x)$. Because the matrix Y is symmetric and positive semifinite, thus its eigenvalues λ_i are real and non-negative. Furthermore, the computed eigenfunctions are orthogonal. The empirical eigenfunctions are found in an ordered manner. That is, the first EEF captures the dominant behavior, the second EEF the next dominant and so on. Since these EEFs are usually small in number, the model derived is of low-dimension.

Appendix 2. Laguerre function

Laguerre function is defined as a functional series [35]

$$\phi_i(t) \triangleq \sqrt{2\xi} \frac{e^{\xi t}}{(i-1)!} \cdot \frac{\mathrm{d}^{i-1}}{\mathrm{d}t^{i-1}} [t^{i-1} \cdot e^{-2\xi t}], \quad i = 1,2,\ldots,\infty, \ \xi > 0, \tag{47}$$

where ξ is the time-scaling factor and $t \in [0, \infty)$ is time variable. The Laplace transform of the ith Laguerre function is given by (48)

$$\phi_i(s) = \sqrt{2\xi} \frac{(s-\xi)^{i-1}}{(s+\xi)^i}, \quad i = 1,2,\ldots,\infty, \ \xi > 0. \tag{48}$$

Laguerre functions (47) and (48) form a complete orthonormal basis in the function space $L_2(R_+)$ and $H_2(C_+)$, respectively.

Appendix 3. The proof of Theorem 2

In order to obtain the convergence with probability 1, the following lemma, which is the direct extension of the previous work [8,19], is needed in the proof of Theorem 2.

Lemma 2. *Let $\xi(x, t)$ be a random variable with zero-mean value and with*

$$|E(\xi(x,t)\xi(x,\tau))| \leqslant M \frac{t^{\alpha} + \tau^{\beta}}{1 + |t-\tau|^{\beta}}, \quad x \in \Omega, M < \infty, \ 0 \leqslant 2\alpha < \beta < 1. \tag{49}$$

Then

$$\frac{1}{n_t} \sum_{t=1}^{n_t} \xi(x,t) \to 0, \quad \text{w. p.1 as } n_t \to \infty. \tag{50}$$

where w. p. 1 means with probability 1.

We now turn to the proof of Theorem 2. The convergence of the estimation $\widehat{\Theta}$

$$\widehat{\Theta} \to \Theta^*, \quad \text{w. p.1 as } n_t \to \infty, \quad n_y \to \infty, \tag{51}$$

implies that

$$\widehat{\Theta}_{\alpha\beta} \to \Theta^*_{\alpha\beta}, \quad \text{w. p.1 as } n_t \to \infty, \quad n_y \to \infty. \tag{52}$$

By Lemma 1, we have

$$\sum_{c=1}^{p} \hat{\beta}^c (\hat{\alpha}^c)^T \to \widehat{\Theta}_{\alpha\beta}, \quad \text{as} p \to \gamma,$$

where $\gamma = rank(\widehat{\Theta}_{\alpha\beta})$. Therefore

$$\sum_{c=1}^{p} \hat{\beta}^c (\hat{\alpha}^c)^T \to \Theta^*_{\alpha\beta}, \quad \text{w. p.1 as } n_t \to \infty, \quad n_y \to \infty, p \to \gamma. \tag{53}$$

Since $H(x, t, \Theta, \{u(\zeta, \tau)\})$ is continuous with respect to Θ, the convergence of parameters as in (53) naturally leads to the convergence of the model to its optimum

$$\hat{y}_n(x,t) \to y^*_n(x,t), \quad \text{w. p.1 as } n_t \to \infty, \quad n_y \to \infty, \quad p \to \gamma. \tag{54}$$

Define

$$Q_{n_y n_t}(\Theta) = \frac{1}{n_y} \sum_{j=1}^{n_y} \left\{ \frac{1}{n_t} \sum_{t=1}^{n_t} \varepsilon(x_j,t,\Theta)^2 \right\}.$$

As define in (35), Θ^* minimizes

$$\overline{E}(y(x,t) - H(x,t,\Theta,\{u(\zeta,\tau)\}))^2 = \lim_{n_y \to \infty} \{ \lim_{n_t \to \infty} E Q_{n_y n_t}(\Theta) \},$$

and the estimate $\widehat{\Theta}$ minimizes $Q_{n_y n_t}$ as defined in (36).

In order to prove (51), we should prove the convergence as follows

$$\sup_{\Theta \in D_\Theta} |Q_{n_y n_t}(\Theta) - E Q_{n_y n_t}(\Theta)| \to 0, \quad \text{w. p.1 as } n_t \to \infty, n_y \to \infty. \tag{55}$$

One feasible solution is to achieve the following convergence at any fixed spatial variable x before working at the spatio-temporal space.

$$\sup_{\Theta \in D_\Theta} \frac{1}{n_t} \sum_{t=1}^{n_t} |\varepsilon(x,t,\Theta)^2 - E\varepsilon(x,t,\Theta)^2| \to 0, \quad \text{w. p.1 as } n_t \to \infty. \tag{56}$$

To achieve the convergence of (56), we have to obtain the convergence first at the pre-defined small open sphere, and then extend it to the global domain D_Θ using Heine-Borel's theorem.

Please cite this article in press as: C. Qi et al., A multi-channel spatio-temporal Hammerstein modeling approach ..., J. Process Contr. (2008), doi:10.1016/j.jprocont.2008.01.006

3.1. Convergence of modeling error ε to its optimum over B

Define the supremum between the model error and its optimum as a random variable

$$\eta(x,t) = \eta(x,t,\Theta^0,\rho) = \sup_{\Theta\in B}[\varepsilon(x,t,\Theta)^2 - E\varepsilon(x,t,\Theta)^2].$$

Let D be the open neighborhood of D_Θ and choose $\Theta^0 \in D_\Theta$. We can define a small open sphere centered at Θ^0 as

$$B(\Theta^0,\rho) = \{\Theta \| \Theta - \Theta^0| < \rho\}.$$

Let $B = B(\Theta^0, \rho) \cap D$, then

$$\sup_{\Theta\in B}\frac{1}{n_t}\sum_{t=1}^{n_t}[\varepsilon(x,t,\Theta)^2 - E\varepsilon(x,t,\Theta)^2] \leqslant \frac{1}{n_t}\sum_{t=1}^{n_t}\eta(x,t). \tag{57}$$

Define $\xi(x, t) = \eta(x, t) - E\eta(x, t)$. *If we can prove*

- $\xi(x, t)$ satisfies Lemma 1 and
- the mean of $\eta(x, t)$ is infinitesimal,

then $\eta(x, t)$ is also infinitesimal.

Firstly, we consider

$$|E(\xi(x,t)\xi(x,\tau))| = Cov[\eta(x,t),\eta(x,\tau)].$$

Define $\eta_\tau^0(x,t) = \sup_{\Theta\in B}[\varepsilon_\tau^0(x,t,\Theta)^2 - E\varepsilon(x,t,\Theta)^2]$, with

$$\varepsilon_\tau^0(x,t,\Theta) = y_\tau^0(x,t) - H(x,t,\Theta,\{u_\tau^0(\zeta,j)\}), t > \tau,$$

where $\{u_\tau^0(\zeta,j)\}$ denotes the input set $(u_\tau^0(\zeta,t),\ldots, u_\tau^0(\zeta,\tau+1),0,\ldots,0)$ for all $\zeta \in \Omega$, $y_\tau^0(x,t)$ and $u_\tau^0(\zeta,j)$ are the variables introduced in Assumption 1. For convenience, let $u_\tau^0(\zeta,j) = 0$ and $y_\tau^0(x,j) = 0$ for $j < \tau$. Obviously $\eta_\tau^0(x,t)$ is independent of $\eta(x,\tau)$ from Assumption 1.

Hence

$$Cov[\eta(x,t),\eta(x,\tau)] = Cov[\eta(x,t) - \eta_\tau^0(x,t),\eta(x,\tau)].$$

Then using Schwarz's inequality, we have

$$|E(\xi(x,t)\xi(x,\tau))| \leqslant [E\eta(x,\tau)^2 E(\eta(x,t) - \eta_\tau^0(x,t))^2]^{1/2}. \tag{58}$$

Since

$$\begin{aligned}|\eta(x,t) - \eta_\tau^0(x,t)| &\leqslant \sup_{\Theta\in B}|\varepsilon(x,t,\Theta)^2 - \varepsilon_\tau^0(x,t,\Theta)^2| \\ &\leqslant \sup_{\Theta\in B}\{|\varepsilon(x,t,\Theta)| + |\varepsilon_\tau^0(x,t,\Theta)|\} \\ &\quad\times \sup_{\Theta\in B}|\varepsilon(x,t,\Theta) - \varepsilon_\tau^0(x,t,\Theta)|,\end{aligned}$$

using Assumption 2, we can further have

$$\begin{aligned}|\eta(x,t) - \eta_\tau^0(x,t)| \leqslant M\Bigg[&\sum_{j=0}^{t}\lambda^{t-j}\{|y(x,j)| + |y_\tau^0(x,j)| \\ &+\sup_{\zeta\in\Omega}|u(\zeta,j)| + \sup_{\zeta\in\Omega}|u_\tau^0(\zeta,j)|\}\Bigg] \\ &\times\Bigg[\sum_{j=0}^{t}\lambda^{t-j}\{|y(x,j) - y_\tau^0(x,j)| \\ &+\sup_{\zeta\in\Omega}|u(\zeta,j) - u_\tau^0(\zeta,j)|\}\Bigg].\end{aligned}$$

Using Assumption 1 and Schwarz's inequality, we can finally derive

$$E|\eta(x,t) - \eta_\tau^0(x,t)|^2 \leqslant M\lambda^{t-\tau}. \tag{59}$$

Following the similar derivation above and using Assumptions 2 and 1, we can also derive

$$E\eta(x,\tau)^2 \leqslant M. \tag{60}$$

Placing (59) and (60) into (58), we can easily derive that $\xi(x, t)$ satisfies Lemma 2, that is

$$\begin{aligned}\frac{1}{n_t}\sum_{t=1}^{n_t}\xi(x,t) &= \frac{1}{n_t}\sum_{t=1}^{n_t}(\eta(x,t) - E\eta(x,t)) \\ &\to 0, \quad \text{w. p.1 as } n_t \to \infty.\end{aligned} \tag{61}$$

Secondly, we derive the mean value of η

$$E\eta(x,t) = E\sup_{\Theta\in B}[\varepsilon(x,t,\Theta)^2 - E\varepsilon(x,t,\Theta)^2].$$

Since the right-hand side is continuous with respect to Θ, $E\eta(x, t)$ should be small if B is small. Furthermore, by Assumption 3,

$$\begin{aligned}\left|\frac{\partial\varepsilon(x,t,\Theta)^2}{\partial\Theta}\right| &\leqslant M|\varepsilon(x,t,\Theta)|^2 \\ &\leqslant M\left[\sum_{j=0}^{t}\lambda^{t-j}\{|y(x,j)| + \sup_{\zeta\in\Omega}|u(\zeta,j)|\}\right]^2,\end{aligned}$$

where we again have used the uniform bounds in Assumption 2. Consequently, by Assumption 1,

$$E\sup_{\Theta\in B}\left|\frac{\partial\varepsilon(x,t,\Theta)^2}{\partial\Theta}\right|^2 \leqslant M.$$

Now

$$\begin{aligned}E\eta(x,t) &= E\sup_{\Theta\in B}[\varepsilon(x,t,\Theta)^2 - E\varepsilon(x,t,\Theta)^2] \\ &\leqslant E\sup_{\Theta\in B}[\varepsilon(x,t,\Theta)^2 - \varepsilon(x,t,\Theta^0)^2] \\ &\quad + \sup_{\Theta\in B}E[\varepsilon(x,t,\Theta^0)^2 - \varepsilon(x,t,\Theta)^2] \\ &\leqslant \left[E\sup_{\Theta\in B}\left|\frac{\partial\varepsilon(x,t,\Theta)^2}{\partial\Theta}\right| + \sup_{\Theta\in B}E\left|\frac{\partial\varepsilon(x,t,\Theta)^2}{\partial\Theta}\right|\right] \\ &\quad\times\sup_{\Theta\in B}|\Theta - \Theta^0| \leqslant M^0\rho.\end{aligned} \tag{62}$$

Finally, from (62), (57) becomes

$$\sup_{\Theta\in B}\frac{1}{n_t}\sum_{t=1}^{n_t}[\varepsilon(x,t,\Theta)^2-E\varepsilon(x,t,\Theta)^2]$$
$$\leqslant\frac{1}{n_t}\sum_{t=1}^{n_t}(\eta(x,t)-E\eta(x,t))+M^0\rho. \quad (63)$$

It is clear to see from (61) that the first term of the right-hand side is arbitrarily small for sufficiently large n_t. Since ρ can also be arbitrarily small, therefore

$$\sup_{\Theta\in B}\frac{1}{n_t}\sum_{t=1}^{n_t}|\varepsilon(x,t,\Theta)^2-E\varepsilon(x,t,\Theta)^2|$$
$$\to 0,\quad \text{w.p.1 as } n_t\to\infty. \quad (64)$$

3.2. Convergence extension to global D_Θ

Since D_Θ is compact, by applying Heine–Borel's theorem, from (64) the following result is easily concluded

$$\sup_{\Theta\in D_\Theta}\frac{1}{n_t}\sum_{t=1}^{n_t}|\varepsilon(x,t,\Theta)^2-E\varepsilon(x,t,\Theta)^2|$$
$$\to 0,\quad \text{w. p.1 as } n_t\to\infty. \quad (65)$$

3.3. Extension to spatio-temporal domain

Obviously

$$\sup_{\Theta\in D_\Theta}|Q_{n_yn_t}(\Theta)$$
$$-EQ_{n_yn_t}(\Theta)\Big|\leqslant\frac{1}{n_y}\sum_{j=1}^{n_y}\sup_{\Theta\in D_\Theta}\frac{1}{n_t}\sum_{t=1}^{n_t}\Big|\varepsilon(x_j,t,\Theta)^2$$
$$-E\varepsilon(x_j,t,\Theta)^2|,$$

therefore

$$\sup_{\Theta\in D_\Theta}|Q_{n_yn_t}(\Theta)-EQ_{n_yn_t}(\Theta)|\to 0,\quad \text{w. p.1 as } n_t$$
$$\to\infty, n_y\to\infty. \quad (66)$$

References

[1] E.W. Bai, An optimal two-stage identification algorithm for Hammerstein–Wiener nonlinear systems, Automatica 34 (3) (1998) 333–338.
[2] E.W. Bai, D. Li, Convergence of the iterative Hammerstein system identification algorithm, IEEE Transactions on Automatic Control 49 (11) (2004) 1929–1940.
[3] R.J.G.B. Campello, G. Favier, W.C. Amaral, Optimal expansions of discrete-time Volterra models using Laguerre functions, Automatica 40 (5) (2004) 815–822.
[4] H.F. Chen, Pathwise convergence of recursive identification algorithms for Hammerstein systems, IEEE Transactions on Automatic Control 49 (10) (2004) 1641–1649.
[5] P.D. Christofides, Robust control of parabolic PDE systems, Chemical Engineering Science 53 (16) (1998) 2949–2965.
[6] P.D. Christofides, Nonlinear and Robust Control of PDE Systems: Methods and Applications to Transport-Reaction Processes, Birkhauser, Boston, 2001.
[7] D. Coca, S.A. Billings, Identification of finite dimensional models of infinite dimensional dynamical systems, Automatica 38 (11) (2002) 1851–1865.
[8] H. Cramér, M.R. Leadbetter, Stationary and Related Stochastic Processes, Wiely, New York, 1967.
[9] K.B. Datta, B.M. Mohan, Orthogonal Functions in Systems and Control, World Scientific, Singapore, 1995.
[10] H. Deng, H.-X. Li, G. Chen, Spectral approximation based intelligent modeling for distributed thermal processes, IEEE Transactions on Control Systems Technology 13 (5) (2005) 686–700.
[11] F. Ding, T.W. Chen, Identification of Hammerstein nonlinear ARMAX systems, Automatica 41 (9) (2005) 1479–1489.
[12] E. Eskinat, S. Johnson, W. Luyben, Use of Hammerstein models in identification of nonlinear systems, AIChE Journal 37 (2) (1991) 255–268.
[13] K. Fruzzetti, A. Palazoglu, K. McDonald, Nonlinear model predictive control using Hammerstein models, Journal of Process Control 7 (1) (1997) 31–41.
[14] D.H. Gay, W.H. Ray, Identification and control of distributed parameter systems by means of the singular value decomposition, Chemical Engineering Science 50 (10) (1995) 1519–1539.
[15] G. Golub, C. Van Loan, Matrix Computations, 3rd ed., The Johns Hopkins University Press, Baltimore and London, 1996.
[16] J.C. Gómez, E. Baeyens, Identification of block-oriented nonlinear systems using orthonormal bases, Journal of Process Control 14 (6) (2004) 685–697.
[17] W. Greblicki, Continuous-time Hammerstein system identification from sampled data, IEEE Transactions on Automatic Control 51 (7) (2006) 1195–1200.
[18] P.S.C. Heuberger, P.M.J. Van den Hof, O.H. Bosgra, A generalized orthonormal basis for linear dynamical systems, IEEE Transactions on Automatic Control 40 (3) (1995) 451–465.
[19] L. Ljung, Convergence analysis of parametric identification methods, IEEE Transaction on Automatic Control 23 (5) (1978) 770–783.
[20] L. Ljung, System Identification: Theory for the User, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1999.
[21] K. Narendra, P. Gallman, An iterative method for the identification of nonlinear systems using a Hammerstein model, IEEE Transactions on Automatic Control 11 (3) (1966) 546–550.
[22] A.J. Newman, Model reduction via the Karhunen–Loève expansion Part II: Some elementary examples. Technical Report T.R.96-33. College Park, Maryland: University of Maryland, 1996.
[23] H.M. Park, D.H. Cho, The use of the Karhunen–Loève decomposition for the modeling of distributed parameter systems, Chemical Engineering Science 51 (1) (1996) 81–98.
[24] W.H. Ray, Advanced Process Control, McGraw-Hill, New York, 1981.
[25] R.A. Sahan, N. Koc-Sahan, D.C. Albin, A. Liakopoulos, Artificial neural network-based modeling and intelligent control of transitional flows, in: Proceeding of the 1997 IEEE International Conference on Control Applications, Hartford, CT, 1997.
[26] L. Sirovich, New Perspectives in Turbulence, first ed., Springer, New York, 1991.
[27] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Glorennec, H. Hjalmarsson, A. Juditsky, Nonlinear black-box modeling in system identification: a unified approach, Automatica 31 (12) (1995) 1691–1724.
[28] N. Smaoui, S. Al-Enezi, Modelling the dynamics of nonlinear partial differential equations using neural networks, Journal of Computational and Applied Mathematics 170 (1) (2004) 27–58.
[29] P. Stoica, On the convergence of an iterative algorithm used for Hammerstein system identification, IEEE Transactions on Automatic Control AC-26 (4) (1981) 967–969.
[30] N. Tanguy, R. Morvan, P. Vilbé, L.C. Calvez, Online optimization of the time scale in adaptive Laguerre-based filters, IEEE Transactions on Signal Processing 48 (4) (2000) 1184–1187.
[31] J. Vörös, Recursive identification of Hammerstein systems with discontinuous nonlinearities containing dead-zones, IEEE Transactions on Automatic Control 48 (12) (2003) 2203–2206.
[32] B. Wahlberg, System identification using Laguerre models, IEEE Transactions on Automatic Control 36 (5) (1991) 551–562.

Please cite this article in press as: C. Qi et al., A multi-channel spatio-temporal Hammerstein modeling approach ..., J. Process Contr. (2008), doi:10.1016/j.jprocont.2008.01.006

[33] B. Wahlberg, System identification using Kautz models, IEEE Transactions on Automatic Control 39 (6) (1994) 1276–1282.
[34] L. Wang, W.R. Cluett, Optimal choice of time-scaling factor for linear system approximation using Laguerre models, IEEE Transactions on Automatic Control 39 (7) (1995) 1463–1467.
[35] C.C. Zervos, G.A. Dumont, Deterministic adaptive control based on Laguerre Series representation, International Journal of Control 48 (6) (1988) 2333–2359.
[36] D. Zheng, K.A. Hoo, M.J. Piovoso, Low-order model identification of distributed parameter systems by a combination of singular value decomposition and the Karhunen–Loève expansion, Industrial and Engineering Chemistry Research 41 (6) (2002) 1545–1556.
[37] Y.C. Zhu, Identification of Hammerstein models for control using ASYM, International Journal of Control 73 (18) (2000) 1692–1702.

Please cite this article in press as: C. Qi et al., A multi-channel spatio-temporal Hammerstein modeling approach ..., J. Process Contr. (2008), doi:10.1016/j.jprocont.2008.01.006