Review

Modeling of distributed parameter systems for applications—A synthesized review from time–space separation

Han-Xiong Li\textsuperscript{a,}\textsuperscript{*}, Chenkun Qi\textsuperscript{b}

\textsuperscript{a} Department of Manufacturing Engineering & Engineering Management, City University of Hong Kong, Hong Kong, China
\textsuperscript{b} School of Mechanical Engineering, Shanghai Jiao Tong University, State Key Laboratory of Mechanical System and Vibration, Shanghai 200240, China

\textbf{Abstract}

Many industrial processes belong to distributed parameter systems (DPS) that have strong spatial–temporal dynamics. Modeling of DPS is difficult but essential to simulation, control and optimization. The first-principle modeling for known DPS often leads to the partial differential equation (PDE). Because it is an infinite-dimensional system, the model reduction (MR) is very necessary for real implementation. The model reduction often works with selection of basis functions (BF). Combination of different BF and MR results in different approaches. For unknown DPS, system identification is usually used to figure out unknown structure and parameters. Using various methods, different approaches are developed. Finally, a novel kernel-based approach is proposed for the complex DPS. This paper provides a brief review of different DPS modeling methods and categorizes them from the view of time–space separation.

© 2010 Elsevier Ltd. All rights reserved.

\textbf{Keywords:}

Distributed parameter system
System modeling
Time–space separation
System identification
Model reduction
Parameter estimation

\section{Introduction}

\section{Modeling of known DPS}

\subsection{Philosophy of time–space separation}

\section{Model reduction}

\subsection{Galerkin method}

\subsection{Collocation method}

\subsection{Approximated inertial manifold method}

\section{Spatial basis functions selection}

\subsection{Finite element method}

\subsection{Finite difference method}

\subsection{Spectral method}

\subsection{Karhunen–Loève method}

\section{Modeling of unknown DPS}

\subsection{Equation residual minimization}

\subsection{Prediction error minimization}

\subsection{Kernel-based modeling}

\section{Conclusions and challenges}

\section{Acknowledgements}

\section{References}

\textsuperscript{*} Corresponding author. Tel.: +86 852 34428435; fax: +86 852 34420172.
\textit{E-mail address: mehxli@cityu.edu.hk} (H.-X. Li).

0959-1524/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jprocont.2010.06.016
1. Introduction

Advanced technological needs such as semiconductor manufacturing, nanotechnology, biotechnology, material engineering and chemical engineering, have motivated control of material microstructure, fluid flows, spatial profiles (e.g., temperature field) and product size distributions [1]. These physical, chemical or biological processes all lead to so-called distributed parameter systems (DPS), where the input, output and even parameters can vary both temporally and spatially. Some typical examples include thermal process [2,3], fluid process [4–6], convection–diffusion–reaction process [7] and flexible beam [8–10]. As the significant progress in the sensor, actuator and computing technology, the studies of distributed parameter processes becomes more and more active in science and engineering. Recently, several special issues for control of DPS have been organized by Dochain et al. [11], Christofides [12], Christofides [13], Christofides and Armaou [14], and Christofides and Wang [15].

A proper mathematical model of the process is essential to many applications such as system analysis, numerical simulation, control design and optimization. The DPS has the time–space coupled nature that is often described in partial differential equations (PDE). This spatially distributed feature requires an infinite-dimensional modeling, which is more difficult and complicated than modeling of lumped parameter systems (LPS). However, because of a finite number of actuators and sensors for practical sensing and control and limited computing power for implementation, such infinite-dimensional systems need to be approximated by finite-dimensional systems. It should be noted that the study of infinite-dimensional systems is very fundamental in the theoretical study of the DPS [16–18], e.g., the analysis of controllability and observability, the proof of existence and uniqueness of solutions, the validity of the approximation by finite-dimensional models. For engineering applications, a finite-dimensional modeling is usually required, which makes the model reduction essential to DPS modeling. In control design, even if the DPS could be unknown, the nominal model of the process is often known (or available) and thus used as a design platform. Many of modeling work is based on the known nominal DPS, upon which control can be further developed for the original unknown DPS. In general, modeling of DPS can be classified into two groups, modeling of known DPS and modeling of unknown DPS.

- For the known DPS, its PDE description derived from the first-principle knowledge can provide a rigorous description of the system. Using proper spatial basis functions, the time–space nature can be decoupled. Then the appropriate model reduction method can approximate the infinite-dimensional system into a finite-order of ordinary differential equations (ODE).
- For the unknown DPS, system identification is further developed from the modeling work of the known DPS. Extra work is needed for two following situations. One is that the PDE structure is available with only some parameters unknown, which requires parameter estimation of the DPS. Another situation is that the PDE structure is unknown, which requires structure design and parameter estimation of the DPS.

Based on the above descriptions, model reduction and system identification can be summarized as some of fundamental issues for the DPS modeling. In the last several decades, many researchers in the field of mathematics and engineering have done much effort on these problems. For each problem, though many different methods have been developed, most of these methods can be synthesized into several categories according to their underlined philosophy. Moreover, different problems and their corresponding methods perhaps share some common properties. Until now, there is little systematic review work presented in this area.

Since it is a very large and complex field, it is almost impossible to review every work in the area, and difficult to do so because different people may have different view. The aim of this paper is to review the applicable methods and classify them under the unified framework of the time–space separation, so that their underlined philosophy can be clearly shown and fairly compared, and easily understood by a wide range of readers. Due to our limited capacity, some of the complicated mathematical theory such as existence and uniqueness of the PDE solution, modeling error analysis, parameter identifiability, convergence of the estimation algorithm, etc., may not be included in this paper.

2. Modeling of known DPS

When the PDE description of the DPS is known, it can be easily transformed into an infinite-dimensional system of ordinary differential equation or difference equation (DE). Then the model reduction will be critical to derive a low-order model for practical application. All these work can be synthesized into a time–space separation framework. For an easy understanding, the following PDE

\[
\frac{\partial y(x,t)}{\partial t} = \alpha \frac{\partial^2 y}{\partial x^2} + \beta \frac{\partial y}{\partial x} + f(y) + w(x)u(t),
\]

(1)

with the boundary conditions \(y(0,t) = 0\), \(y(\pi, t) = 0\), and the initial condition \(y(x, 0) = y_0(x)\), will be selected as an example. Here for simplicity we take the PDE with the first-order in time as the example. However, the following modeling could be also applicable for the PDE with the second-order in time, e.g., flexible structures [19–21].

2.1. Philosophy of time–space separation

It is well known that a continuous function can be approximated using Fourier series [22]. Based on this principle, the spatio-temporal variable \(y(x, t)\) of the DPS can be expanded by a set of spatial basis functions (BFs) \(\{\phi_i(x)\}_{i=1}^{\infty}\) as follows

\[
y(x,t) = \sum_{i=1}^{\infty} \phi_i(x)y_i(t).
\]

(2)

Similar to Fourier series, the spatial BFs are often ordered from slow to fast in the spatial frequency domain. Because the fast modes contribute little to the whole system, only the first \(n\) slow modes in the expansion will be retained in practice [23]

\[
y_n(x,t) = \sum_{i=1}^{n} \phi_i(x)y_i(t).
\]

(3)

Thus, the spatio-temporal variable is separated into a set of spatial BFs and the corresponding temporal model, as depicted in Fig. 1. The key is to select proper spatial BFs, and construct the finite-order (low-order) temporal model as explained in Fig. 2, which turns a PDE into a finite-dimensional system. Finally, through the time–space synthesis, the spatio-temporal system will be recovered.

Under the time–space separation framework, different approaches and methods will arise according to the combination of proper model reduction approaches and spatial BFs selection.

2.2. Model reduction

The weighted residual method (WRM) [23,24] is the most often used method for model reduction. The equation residual
of the model (1) generated from the truncated expansion can be expressed as

$$ R(x, t) = \dot{y}_n - \left( \alpha \frac{\partial^2 y_n}{\partial x^2} + \beta \frac{\partial y_n}{\partial x} + f(y_n) + wbu \right), $$

which can be made small in the sense that

$$ (R, \phi_i) = 0, \quad i = 1, \ldots, n, $$

where \( \{\phi_i(x)\}_{i=1}^n \) are a set of weighting functions to be chosen. As shown in Fig. 3, the minimization of the residual \( R \) actually turns to minimizing its projections to weighting functions. This is an easy way to obtain a \( n \)-order ODE model for temporal coefficients \( \{y_i(t)\}_{i=1}^n \). For the PDE with the high-order (e.g., second-order) in time, the obtained ODE will be also high-order in time [19,20].

The accuracy and efficiency of WRM depend very much on the basis and weighting functions chosen [23]. Many methods have been proposed based on the selection of weighting functions. The most popular approach appears to be Galerkin method and collocation method.

### 2.2.1. Galerkin method

If the weighting functions \( \{\phi_i(x)\}_{i=1}^n \) are chosen to be the BFs \( \{\phi_i(x)\}_{i=1}^n \), then the method is called the Galerkin method [23,24]. It has the advantage that the residual is made orthogonal to each BF and is, therefore, the best solution possible in the space made up of the \( n \) functions \( \phi_i(x) \). Because it does not need to find other weighting functions, so this method is relatively simple and most often used.

### 2.2.2. Collocation method

The weighting functions of the collocation method [24] are chosen to be Dirac delta functions \( \delta(x - x_i), i = 1, \ldots, n \). The residuals vanish at collocation nodes \( \{x_i\}_{i=1}^n \), i.e., \( R(x_i) = 0 \), so the collocation nodes are very critical for the modeling performance. Fortunately, some mathematical theories show that they can be specified automatically in an optimal way, for example, the roots of the orthogonal polynomials (e.g., [25,26]).

Both Galerkin and collocation methods are linear reduction methods that work fine for the linear DPS. Since the fast modes \( \{y_i(t)\}_{i=n+1}^{\infty} \) are completely ignored, some information of the slow modes may get lost for the nonlinear DPS because of the coupling between the slow and fast modes. To improve the accuracy without using a high-order model, nonlinear reduction methods are often used to compensate some of the fast modes ignored \( \{y_i(t)\}_{i=n+1}^{\infty} \). One of them is based on inertial manifold (IM) [27], where the fast modes are described by the slow modes, and then the DPS can be transformed into a finite-dimensional system. However, for many nonlinear DPS, IM may not exist or may be difficult to find.

### 2.2.3. Approximated inertial manifold method

To overcome the above problems, the approximated inertial manifold (AIM) is used to approximately compensate the fast modes with the slow modes [28–30]. Even for the systems with unknown existence of IM, the AIM method can often achieve a better performance than linear Galerkin or collocation methods. There are several approaches to obtain the AIM.

- To assume that the fast modes are at pseudo-steady state, then the so-called steady manifold [31] can approximate IM easily, by ignoring the dynamic information of the fast modes.
- To consider the dynamic information, another AIM is used by integrating the fast modes for a short time using an implicit Euler method [28].
- To further improve the approximation accuracy, a novel procedure based on singular perturbations method can be used to construct the AIM with an arbitrary accuracy under certain conditions [7,30].
The above AIM can be implemented in either the Galerkin or the collocation approach, which leads to Galerkin AIM and collocation AIM methods.

2.3. Spatial basis functions selection

Selection of spatial BFs is critical to the model reduction, and has a great impact to the modeling performance. As shown in Table 1, the spatial BFs can be classified into local and global types, and further into analytical and data-based functions based on applications. In general, there are four major approaches, the finite difference method (FDM), the finite element method (FEM), the spectral method and the Karhunen–Loève (KL) method. The FDM can be considered as a special case of FEM if the BFs are selected as Dirac delta functions; the data-based KL BFs are often global; and no local data-based BFs have been reported so far. For different applications, different modeling methodologies can be formulated through an integration of proper spatial BFs and model reduction approaches.

2.3.1. Finite element method

The spatial BFs of the FEM are local. The spatial domain is first discretized into sub-domains. Then the low-order piecewise polynomials [32], splines [33], and wavelets [34,35] are often used as local BFs in sub-domains. Further combined with model reduction methods, different applications are reported.

- Examples include, the spline-Galerkin method for the order reduction of the controller [36] and flexible structure [10], the wavelet-Galerkin method and the wavelet-Galerkin-AIM method [35] for the reduced-order modeling.
- Other examples include, the piecewise polynomials-Galerkin method for an AIM implementation [37], the wavelet-collocation method for numerical simulation [38], and the wavelet-collocation-AIM method for the reduced-order DPS modeling [39,40].

The most attractive feature of the FEM is its flexible ability to handle complex geometries and boundaries because of the local BFs used. Due to its flexibility, many FEM software products are developed for the DPS simulation. FEM can provide an upper bound on the error associated with the finite-dimensional approximation. Because of the local BFs, the FEM often requires a high-order model for a good approximation.

2.3.2. Finite difference method

FDM is a popular method to provide the numerical solution of the PDE [41]. The spatio-temporal variables are discretized within the time–space domain. Derivatives at each discretization node are approximated by the difference over a small interval, which can be forward, backward and central difference derived often from Taylor expansion. Then the PDE system is transformed into a set of difference equations, whose order is proportional to the number of spatial discretization nodes. The most attractive feature of FDM is that it can work for all kinds of DPS with various boundary conditions and regular domain. However, it usually requires a high-order model for an accurate solution, and has the disadvantages of heavy computation burden.

The method of lines (MOL) [42] is a special case of the FDM, with only partial derivatives in the spatial direction replaced by finite difference approximations. It results in a system of ODE, which has the similar strength and weakness as FDM. Both FDM and MOL can fall into the framework of time–space separation, if their spatial BFs and weighting functions are both chosen as Dirac delta functions. Thus, the FDM and MOL can be viewed as a subset of the FEM approach.

2.3.3. Spectral method

The spatial BFs of the spectral method are global and orthogonal in the whole spatial domain [43,44]. Due to the global nature of BFs, the spectral method can achieve the lower-order model than the FEM. However, for the same reason, an efficient spectral method often requires that the system has a regular space domain and smooth output. In particular, most of parabolic systems have the spectral gap between the slow and fast modes, thus it is possible to derive an accurate low-order model using the spectral method. To obtain a satisfactory model, the BFs should be carefully designed according to some practical situations such as boundary condition and space domain. Some typical BFs are discussed as follows.

2.3.3.1. Fourier series. Fourier series has some good properties such as approximation capability, infinite differential and periodic functions. So it is often used for the processes with periodic boundary conditions and a finite domain [43,44].

2.3.3.2. Eigenfunctions. The eigenfunctions (EF) of the linear or linearized spatial operator are suitable for most of parabolic PDE systems [45,46] because their eigenspectrum display a separation of the eigenvalues, i.e., slow and fast modes as shown in Fig. 4. Many control applications are based on the finite-order ODE models derived from the EF-Galerkin method, especially for quasi-linear parabolic systems. For example, control of a class of quasi-linear parabolic diffusion-reaction processes [47], and control of more complex distributed systems, which include the cases of parameter uncertainties [46,48] and input constraints [49,50]. To further improve the modeling performance, the EF-Galerkin-AIM method is reported to control quasi-linear parabolic processes [30], the process in time-dependent spatial domains [51], and the process with parameter uncertainties [45].
Except the above successful applications, the use of EF also has its limitations.

- When the spatial operator is self-adjoint [52], the model may be low-order. However, for the non-self-adjoint system (e.g., convection-diffusion processes), the EF will satisfy a weighted orthonormality relation, thus the resulting model could be of higher order or even unstable because of the slow convergence of the EF solution [35,52,53].
- For many typical operators and regular boundary conditions, the eigenvalue and EF can be easily derived [54]. However, for the system with nonlinear spatial operators, spatially varying parameters, complex boundary conditions and irregular domain, it will be very difficult and even impossible to get analytical EF.
- EF may be hardly applied to hyperbolic systems, because they do not show the clear separated eigenspectrum. Thus, the method of characteristics [24] might be a good choice, where the system is transformed into a set of ODEs, which describe the original DPS along with their characteristic lines.

2.3.3.3. Orthogonal polynomials. Orthogonal polynomials are also popular in model reduction of the DPS. For example, the polynomial-collocation method is used to derive a low-order DPS model for the purpose of simulation and control [25], and for the adaptive control [26]. In general, Chebyshev polynomials [43,44] and Legendre polynomials [43] are suitable for non-periodic problem defined on a finite domain. Laguerre polynomials work well on the semi-infinite domain. Hermite polynomials [44] are useful for the problems with an infinite domain. These functions are very flexible and can be applicable to a broad class of systems. Some polynomials are optimal in some cases, e.g., Chebyshev or Legendre polynomials are optimal in a non-periodic problem with a finite domain. However, they may not be always optimal since these general polynomials do not utilize any specific knowledge about the system.

2.3.4. Karhunen–Loève method

The KL method, also known as proper orthogonal decomposition or principal component analysis (PCA), is a statistical analysis technique of obtaining the so-called empirical eigenfunctions (EEFs) from the numerical or experimental data [4,55,56]. The basic idea of the KL expansion is to find those modes which represent the dominant character of the system. The problem is to compute the most characteristic spatial structure \( \phi_i(x) \) among the representative process data \( y(x,t) \) that can be performed by minimizing the following objective function

\[
\min_{\phi_i(x)} \left \{ \left( y(x,t) - \sum_{i=1}^{n} \phi_i(x) \phi_i(x), y(x,t) \right)^2 \right \}
\]

subject to \( \phi_i(x), \phi_i(x) = 1, i = 1, \ldots, n \)

where the inner product is defined as \( \langle f(x), g(x) \rangle = \int f(x)g(x)dx \) and the ensemble average is defined as \( \langle f(t) \rangle = \frac{1}{L} \sum_{t=1}^{L} f(t) \). The constraint \( \phi_i(x), \phi_i(x) = 1 \) is imposed to ensure that the function \( \phi_i(x) \) is unique. In fact, the solution of this problem is closely related to singular value decomposition. The spatio-temporal dynamics can be separated into orthonormal spatial and temporal modes using the singular value decomposition

\[
y(x,t) = \sum_{i=1}^{m} \sigma_i \phi_i(x) \psi_i(t), \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_m.
\]

where singular values \( \sigma_i \) denote the importance of the modes, left singular functions \( \phi_i(x) \) represent spatial modes, and right singular functions \( \psi_i(t) \) are temporal modes. Among the mth-order decomposition, the first n spatial modes will be the solution. The embedded separations can be geometrically depicted in Fig. 5.

The KL method is actually implemented in different ways, e.g., spatial correlation method and method of snapshots [55]. The EEFs are found in an ordered manner based on the values of \( \sigma_i \), with the first EEF as the most dominant behavior, the second as the next dominant and so on. Usually, only the first few modes can capture the most important dynamics of the system, thus a small number of EEFs can be selected to yield a low-order model.

The KL-Galerkin method could be one of the most commonly used DPS modeling methods. It has been applied to system analysis, model reduction, numerical simulation of many complex distributed systems, e.g., fluid flow [4,57,58], thermal process [2,59–61], and diffusion–convection–reaction process [62]. Many control applications can be carried out, e.g., control of the growth of thin films in a chemical vapor deposition reactor [63], control of a diffusion–reaction process [64], control of a thin shell system [65], and optimization of diffusion–convection–reaction processes [66].

The applications of KL-Galerkin–AIM method [67,68] and the KL-collocation method [69–71] are also studied for simulation, control and optimization of thermal processes.

Remark. Compared with the spectral method, the KL method is applicable to a wider range of complex distributed systems, including those with irregular domain, nonlinear spatial operator and nonlinear boundary conditions. Because BFs of the KL method can provide an optimal linear representation of spatio-temporal data themselves [4,72], it may generate a lower-order and more accurate model than the FEM and the spectral method. However, its major drawback is that the KL method depends on cases and lacks a systematic solution for one class of systems. Thus, experiment settings such as the input signal, the time interval, the number of snapshots, the values of system parameters [59], and the initial conditions [73] have to be carefully chosen for an efficient application.

3. Modeling of unknown DPS

All the previous model reduction methods require that the PDE description of the system is known. However, in many cases it is difficult to obtain an exact PDE description of the process only from the physical and chemical laws, so the system identification is often employed to estimate the unknown DPS from data.

Under the philosophy of the time–space separation, as shown in Fig. 6 the unknown DPS modeling could be summarized in two main approaches: equation residual minimization and prediction error minimization.

- The equation residual minimization is to first reconstruct the left and right sides of the PDE as in (1) from data, then to estimate the model by minimizing the equation residual as given in (4). This approach does not need to solve the PDE system.
In the prediction error minimization approach, the model is figured out by minimizing the error between the measured output and predicted output as shown in Fig. 7. The numerical solution of the PDE is required to obtain the predicted output, so its optimization computation may be complex.

Ljung’s book [74] is very good and useful in the area of system identification. The concepts, theory and algorithms in this book are also suitable for DPS though the identification of DPS is more difficult and complex than lumped parameter systems. There are lots of studies on the system identification of LPS, which can give much help on the estimation and identification of DPS. Similar to LPS, in general, problems in the DPS can be classified in two different cases: parameter estimation for known structure, structure design and parameter estimation for unknown structure.

When the PDE structure of the system is known, the unknown parameters can be estimated from the experimental data [75]. Once the parameters are determined, the PDE model reduction can be used for real applications. Some earlier survey papers are given by Kubrusly [76] and Polis and Goodson [77].

For completely unknown processes which widely exist in real world, the black-box identification has to be used. More sensors are often required in spatial locations to collect enough information. Modeling becomes extremely difficult because both the structure and parameters need to be figured out.

Due to the time/space nature, there are some special problems on the DPS identification which will be emphasized here. In either case, the model reduction and the BF selection introduced in Section 2 will play an essential role as depicted in Fig. 6.

3.1. Equation residual minimization

After each term in the PDE is computed from the data, the unknown uncertainties can be compensated by minimizing the equation error. The equation residual minimization is often transformed into a linear or nonlinear regression problem. For the term reconstruction, the finite difference is popular and simple method [78], and other schemes include spline or kernel-based differentiation method [79]. Some authors used orthogonal functions for each term reconstruction in the PDE, e.g., Fourier series [80], Chebyshev series [81], Walsh functions [82], Laguerre polynomials [83,84], Taylor series [85], general orthogonal polynomials [86] and block-pulse functions [87].

• When the model structure is known, the least-squares method or other traditional optimization algorithms can be popularly used to estimate parameters [78, 80, 83, 84, 86–88]. It actually becomes a traditional parameter estimation problem.

• When the model structure is unknown, Voss et al. [89] presented alternating conditional expectation algorithm for parameter estimation and model selection. Bär et al. [90] used linear regression method for parameter estimation and backward elimination or exhaustive search for model selection. Guo and Billings [91] and Guo et al. [79] estimated the parameters and structure using orthogonal least-squares algorithm. There is no standard solution.

Because it does not need to solve the PDE, the equation residual minimization approach is relatively simple. However, some de-noise and smoothing techniques may be useful to reduce noise sensitivity, e.g., filter [78], spline function based differentiation [79], and implicit Adams integration [91]. The resolution of the sensors also has inevitable impact on the accuracy of the estimated derivatives. More sensors are usually used to obtain good parameter estimation if the parameters are temporally or spatially varying. The convergence and consistency analysis is still a difficult problem because of the inaccurate derivatives estimation.

3.2. Prediction error minimization

When the structure is known, the predicted output can be computed from the dynamical PDE model with FDM, FEM, spectral and KL methods, and thus the system parameters can be found by minimizing the error between the measurement and the prediction [75]. If a low-order prediction model is possible, the spectral or KL based estimation may be effective and computationally efficient compared with FDM and FEM.

• For constant parameters, the traditional optimization techniques can be directly used, e.g., Levenberg–Marquardt method [92,93], Gauss–Newton method [94], Newton-like method [95], quasi-linearization method [96], and multiple shooting optimization algorithm [97]. The convergence analysis and estimation error problem are also studied by Rannacher and Vexler [94] and Banks et al. [93].

• For spatially and/or temporally varying parameters, a finite-dimensional parameterization such as the expansion onto a set of basis functions or spatial points are often needed. Then the problem becomes the constant parameter estimation problem as described above and the traditional optimization techniques can be used, e.g., least-squares method [98], quasi-linearization method [99], and Levenberg–Marquardt method [100,101]. Some fundamental issues are also studied, such as, the identifiability analysis [102] and convergence problem [99–101].

When the structure is unknown, after choosing the proper BFs for the time/space separation, the corresponding unknown temporal model can be estimated using traditional system identification.
techniques. The methodology is similar to that shown in Fig. 2, where the temporal model is identified from the data instead of derived from the PDE description. Various lumped system identification techniques can be applied by combining different selections of BFs.

- Most of methods are proposed with the parametric modeling with model structure pre-selected. The neural networks are popularly integrated with finite difference method [103], or KL method [104–106] to compensate the structure uncertainty. To better model DPS with more complex nonlinearity, a nonlinear PCA based neural method is proposed [107]. For the partly known DPS, a neural compensation method is integrated with the spectral method to estimate unknown nonlinearities and parameters [108,109]. This hybrid neural spectral method can significantly reduce the number of sensors used for identification. However, there is still no systematic method to pre-select the model structure.

- Some works are related with nonparametric modeling method with the model structure determined from data [110–112]. Some considers work on the structure design together with parameter estimation, such as, the modeling with a fuzzy partition method for network design [113]. The nonlinear autoregressive with exogenous input (NARX) model is estimated by Coca and Billings [114], Billings and Coca [115] and Guo and Billings [116] using the orthogonal forward regression (OFR) algorithm for model selection and parameter estimation. Theoretical results concerning the existence, stability and convergence of a finite-dimensional model are also studied [117].

In general, for the unknown structure, modeling is still not systematic and lots of efforts need to spend.

3.3. Kernel-based modeling

The kernel model is another type of prediction error minimization approach for the DPS modeling with unknown structure.

- The linear DPS can be represented using the Green’s function (i.e., impulse response function or kernel)

$$y(x, t) = \int_0^t \int_0^t g(x, \zeta, \tau)u(\zeta, t - \tau)d\tau d\zeta. \quad (8)$$

The Green’s function can be derived from the first-principle knowledge [54]. If the analytical Green’s function is not available, it can be estimated from the input–output data using the time–space separation. For example, based on the singular function expansion, a time-invariant Green’s function kernel can be estimated using SVD method [52]. To avoid the time-invariant assumption, a time-varying Green’s function kernel can be obtained from the singular function and KL basis function expansion using SVD–KL method [118–120]. The Green’s function can also be estimated using the least-squares method [121]. Because the Green’s function model uses one single kernel, it may only be able to approximate the linear DPS or the nonlinear DPS at the given working condition.

- For modeling the nonlinear DPS, a spatio-temporal Volterra model is proposed recently with a series of spatio-temporal (3D) kernels [122,123]

$$y(x, t) = \sum_{r=1}^R \int_0^t \int_0^t \int_0^t \int_0^t g_r(x, \zeta_1, ..., \zeta_r, t_1, ..., t_r)$$

$$\times \prod_{p=1}^r u(\zeta_p, t - t_p)d\tau pd\zeta p. \quad (9)$$

Now the rth-order 3D kernel $g_r$ denotes the influence of the input $u$ at location $\zeta_1, ..., \zeta_r$ and time $t - t_1, ..., t - t_r$ on the output $y$ at location $x$ and time $t$. For easy understanding, a simple example of 3D kernel model for $R = 2$ is given as below

$$y(x, t) = \int_0^t \sum_{r=0}^t g_1(x, \zeta, t_1)u(\zeta, t - t_1)d\zeta_1$$

$$+ \int_0^t \int_0^t \sum_{r_1=0}^t \sum_{r_2=0}^t g_2(x, \zeta_1, \zeta_2, t_1, t_2)u(\zeta_1, t - t_1)$$

$$\times u(\zeta_2, t - t_2)d\zeta_1d\zeta_2. \quad (10)$$

The dynamics can be first decomposed into a set of kernels, upon which the previous time–space separation of the kernels

$$g_r(\zeta) = \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^l \sum_{l=1}^m \sum_{j'=1}^l \sum_{k'=1}^l d_{i,j,k,l,j',k'}$$

$$\times \prod_{s=1}^r \phi_i(x)\psi_{j,s}(\zeta_s)\psi_{k,s}(\zeta_s), r = 1, ..., R, \quad (11)$$

will be carried out as shown in Fig. 8. After estimating unknown parameters $d_{i,j,k,l,j',k'}$, the kernels can be reconstructed using the time–space synthesis. The modeling capability will improve as the kernel numbers increase. Since multiple kernels are used, it should be more capable to approximate a wide range of nonlinear complex DPS.

- To model the block-oriented nonlinear DPS, the kernel-based spatio-temporal Hammerstein and Wiener models are proposed [124–126]. The dynamics are first decomposed into a static nonlinear block $N(\cdot)$ and a dynamic linear spatio-temporal block $g(\cdot)$, where the spatio-temporal Hammerstein model (see Fig. 9) consists of a static nonlinear block connected by a dynamic linear spatio-temporal kernel (i.e., Green’s function)

$$y(x, t) = \int_0^t \int_0^t g(x, \zeta, \tau)N(u(\zeta, t - \tau))d\tau d\zeta, \quad (12)$$

$$u(x, t) = \frac{N(\cdot)}{\|N(\cdot)\|}, v(x, t) = \frac{g(x, \zeta, \tau)}{\|g(x, \zeta, \tau)\|}, \quad (13)$$

Fig. 9. Kernel-based spatio-temporal Hammerstein model.
and the spatio-temporal Wiener model (see Fig. 10) will arise if the connection order is reversed

\[ y(x, t) = N \left( \int_0^\pi \int_0^t g(x, \xi, t - \tau) d\tau d\xi \right). \] (13)

They can be considered as a nonlinear extension of the Green’s function model and also a spatial extension of traditional temporal Hammerstein and Wiener models. The modeling can be simplified to traditional Hammerstein and Wiener modeling after the time–space separation as shown in Fig. 11. This kind of models is suitable for approximating a wide range of block-oriented nonlinear DPS, and also convenient for the control design due to their linear/nonlinear separated structure.

Remark. In addition to the equation residual minimization and prediction error minimization approaches, other approaches include maximum likelihood method [127], adaptive parameter and state estimation method [128–131], etc. However, they mainly focus on the unknown parameter problem.

4. Conclusions and challenges

This paper presents a brief survey on advances in DPS modeling from the view of time–space separation. The DPS modeling has two key issues: model reduction, and system identification. Methods have been classified and their strength and weakness have been explained under the principle of time–space separation. For engineering applications, the infinite-dimensional nature must be transformed to the finite-dimensional problem, where the model reduction plays an essential role. The system identification will be crucial to the unknown DPS.

Modeling of DPS with unknown structure is still not mature due to its complexity and large varieties. Some of difficult issues are summarized below that may worth a future research.

- All the existing DPS identification methods can only work in the vicinity of the operating conditions due to the nonlinearity in both temporal and spatial domain. It is still very challenging to obtain a model which can work well in a wide range of operating conditions.

- Control application is one of the major purposes for DPS identification. A proper integration of both system identification and control for an optimal performance is very important and needs much more effort. Though there are many studies reported for LPS (e.g., [132–134]), only a few studies have been presented for DPS [135–137].

- Optimal experiment design (e.g., selection of input-output variables, input signal design, and selection of working conditions) to obtain informative data is very important for the identification of DPS. The placement (number and locations) of sensors and actuators may be also included in this topic. It is a difficult problem, particularly in nonlinear, multivariable, or hybrid case (with both continuous and discrete variables). Some studies of sensor placement are reported for parameter estimation of DPS in known structure [138–144]. See survey papers [145,146] for sensor/actuator placement and one book [147] for a summary of sensor placement for parameter estimation. However, the extension of these works to DPS in unknown structure need further research. Moreover, whether the results from experiment design extensively and systematically studied in LPS identification (e.g., [148–150]) can be extended to DPS is still an open question.

Acknowledgements

Authors would like to thank the editor and the anonymous referees for their valuable comments and suggestions. The work presented in this paper is partially supported by a grant from RGC of Hong Kong (CityU: 117208) and a grant from NSF China (50775224).

References


