ELSEVIER

Journal of Process Control 12 (2002) 807–821

JOURNAL OF PROCESS CONTROL

www.elsevier.com/locate/jprocont

Robust model-order reduction of complex biological processes

T.T. Lee[a,*], F.Y. Wang[b], R.B. Newell[b]

[a]*Institute of High Performance Computing, 1 Science Park Road #01-01 The Capricorn, Singapore Science Park II, Singapore 117528, Singapore*
[b]*Computer-Aided Process Engineering Centre, Department of Chemical Engineering, The University of Queensland, St. Lucia 4072 Brisbane, Qld, Australia*

Abstract

This paper addresses robust model-order reduction of a high dimensional nonlinear partial differential equation (PDE) model of a complex biological process. Based on a nonlinear, distributed parameter model of the same process which was validated against experimental data of an existing, pilot-scale BNR activated sludge plant, we developed a state-space model with 154 state variables in this work. A general algorithm for robustly reducing the nonlinear PDE model is presented and based on an investigation of five state-of-the-art model-order reduction techniques, we are able to reduce the original model to a model with only 30 states without incurring pronounced modelling errors. The *Singular perturbation approximation balanced truncating* technique is found to give the lowest modelling errors in low frequency ranges and hence is deemed most suitable for controller design and other real-time applications.

Keywords: Model-order reduction; BNR activated sludge process; Singular perturbation approximation; Hankel singular values

1. Introduction

There has been an increased interest in applying models for complex biological processes such as nutrients removal activated sludge plants due to their usefulness in not only assisting plant trouble-shooting, operator training and optimising design of new plants, but also for real-time process monitoring and control. A number of models for the biological nutrient removal (BNR) activated sludge processes have been published in the literature [1–3]. Biological models that incorporate kinetics of the International Association of Water Quality (IAWQ) Model No. 1 [4] consisting of eight processes with 13 state variables and 19 processes with 19 state variables respectively are indeed large and complex. If we allow prediction of transient dynamics and coupling of the biological model with models of associated equipment such as secondary settling tank and return activated sludge tank, this will result in an increased complexity of the entire BNR activated sludge model. For real-time process monitoring or control, a reduction in the order of the high dimensional model is inevitable.

It is well known that model-order reduction has often been based on physical intuition. For instance, concentration of a species within a stirred-tank reactor and its outlet is assumed to be the same, and design of packed bed separation towers is approximated by calculating 'equivalent' number of trays. Based on a number of simplified, qualitative assumptions, a first attempt to reduce model of a biological system incorporating the IAWQ Model No. 1 kinetic has been reported by Jeppsson and Olsson [5]. Recently, Jannssen et al. [6] developed similar assumption-based reduced-order model for controller tuning with resulting decreased simulation time by a factor of three. Ayesa et al. [7] applied the reduced-order model of Jeppsson and Olsson to evaluate its *observability* for two separate wastewater treatment plant configurations, with the final objectives of development and application of advanced control strategies to the plants. Recently, Steffens et al. [8] applied the *singular perturbation approach* of Kokotovic et al. [9] to a carbon removal and nitrifying activated sludge process model in which 23 ordinary differential equations were reduced to 23 algebraic equations with resulting savings in process simulation time without incurring significant modelling errors. As indicated in their paper, the biological model has 32 states thus giving a total of 32 ordinary differential

* Corresponding author. Tel.: +65-770-9287; fax: +65-770-9280.
E-mail address: leett@ihpc.nus.edu.sg (T.T. Lee).

0959-1524/02/$ - see front matter
PII: S0959-1524(02)00006-9

equations (ODE). Total number of equations for the entire activated sludge system would be more since algebraic equations defining variables at various streams were not counted. Unfortunately the authors did not explicitly state total number of equations for the system.

Although design of controllers for these complicated biological processes is one of the main objectives of model-order reduction, none of them had addressed the important closed-loop "robustness" issues. From a control engineering point of view, this implies that:

- Simplification of the best available process model is such that design of a control system utilising the reduced-order model is able to meet certain specifications;
- Fictitious and unobservable or uncontrollable states generated by the model reduction algorithm can be systematically 'stripped' away; and
- If modern control method such as linear quadratic gaussian (LQG) or $\mathbf{H}_\infty$ (pronounced as "H infinity") control laws are to be employed for controlling the process, usage of the reduced-order model can significantly reduce the control law complexity without pronounced change in control system performance.

Besides Kokotovic's technique, there is a host of other established techniques that can be used to automate reduction of complex biological process models. Several time-domain and frequency-domain based model-order reduction techniques have emerged largely motivated by the needs for simplifying numerical simulation and control system design.

One of the earliest works on these techniques is due to Davison [10] who proposed simplification to linear, state-space models by retaining dominant system eigen values and their corresponding *eigen vectors*. Chidambara [11] provided some modifications to this early time-domain, model-order reduction technique. Similar development in model-order reduction automation using the frequency domain approach has also appeared in the early literature [12] but the resulting reduced-order models based on these early ideas have often been not optimised.

Following Wilson [13], optimal reduced-order model for multiple input-multiple output (MIMO) processes are obtained by minimising some cost functions and satisfying two algebraic matrix *Lyapunov* equations. Extensions and generalisation to this technique were later developed by other workers [14]. Development in optimising reduced-order model via the frequency domain has similarly been reported by Luus [15]. Despite being able to meet the primary objective of finding reduced order models that closely approximate the original high dimensional systems, these classical optimisation techniques nevertheless have several drawbacks. Nonlinearity in the parameters of the reduced order models is probably the most severe problem because these techniques not only require good starting guesses to avoid convergence problem but one may face with multiple local minima when employing the time consuming iterative minimisation algorithms.

Significant advancement in model-order reduction automation is achieved with the introduction of the so-called *balanced truncating* method of model reduction by Moore [16] and the Optimal Hankel norm approximations [17]. What makes these newer model-order reduction techniques attractive is that not only one could anticipate the extent of an error even before actually doing the model reduction but more importantly they are much simpler to apply and require significantly less computation time [18].

In many cases, it is desirable to achieve low modelling errors in a particular frequency region compared with other regions. For instance, controllers in the process industries typically operated at low to unity range of frequencies. In such case the original state-space model should be pretreated with appropriate frequency weighting functions to emphasise the lower frequency regions while de-emphasising the higher frequency regions prior to model reduction [19,20]. Alternatively, the singular perturbation approach of Kokotovic and Moore's *balanced realisation* theory may be employed to partition *slow* subsystem from *fast* subsystem to achieve similar results.

Some recent examples on the application of balanced realisation and Hankel norm approximation theories are modelling and control of Kraft digester [21] and design of robust controller [22]. A number of new computational algorithms have also been recently proposed, generally to improve the practicality of the developed theories. For instance, Halevi et al. [23] proposed an algorithm for updating optimum reduced-order models either on-line or off-line based on series expansion. Beck et al. [24] on the other hand employed a *linear fraction transformation* technique together with the balanced truncating technique of Moore to develop an algorithm for model-order reduction of MIMO system with uncertainties.

In this work, we shall develop reduced-order, linear state-space models for a nonlinear partial differential equation (PDE) model of a complex biological process (i.e. the BNR activated sludge system) based on an investigation of five model-order reduction techniques that contributes to achieving the three-point aims stated earlier. Based on a frequency-domain modelling error criterion, errors incurred due to model-order reduction will be explicitly quantified followed by selection of the 'best' reduced-order model. Verifiability of the selected reduced-order, state-space model of the BNR activated sludge process will be done by comparing its dynamics against a non-linear, distributed parameter-based model of the same process developed and experimentally-validated in our previous work [3]. We shall in a separate

study [25], address robust multivariable control of complex biological processes, using the results of reduced-order, state-space model of the BNR activated sludge process developed in this work.

2. Process description

Fig. 1 shows a BNR activated sludge plant operated in the Johannesburg configuration that consisted of anaerobic, anoxic and aerobic zones and a secondary settler in a back-to-back scheme with multi recycle streams. To ensure plug flow conditions prevailed in the bioreactors, the basins are normally partitioned, such that backmixing is minimised. In the anaerobic zone, fermentable organic from the influent wastewater are mixed with returned activated sludge (RAS) and converted to volatile fatty acids (VFAs) by heterotrophic organisms. The latter is consumed by phosphorus accumulating organisms (PAOs) and stored as *poly-hydroxy-alkanoate* (PHA) in the *internal organic cells*. Concurrently, poly-phosphate and hence energy for VFAs accumulation are released from *internal inorganic cells*. There is a net reduction of alkalinity and hence pH in this zone due to acids production. If the amount of VFAs is insufficient, additional acids may be added to maintain a maximum phosphate to VFAs ratio and a minimum concentration of biological phosphate organisms in the system. It is also not uncommon to install an 'activated' primary sedimentation tank to allow production of VFAs by fermentation of readily biodegradable substrate in the incoming sewage.

In the anoxic zone, nitrate that is recycled from the aerobic zone is converted to dinitrogen by facultative heterotrophic organisms. Denitrification results in released of alkalinity and hence increased in pH. There is also a pronounced removal of phosphorus in this zone.

On reaching the aerobic zone, virtually all the readily biodegradable organics [hereafter will be referred to as biodegradable *chemical oxygen demand* (COD)] in the partially-treated wastewater has been consumed by the heterotrophic organisms in the previous zones. In this zone, two major processes occurred. In the presence of dissolved oxygen, the released phosphate is taken up by the PAOs and stored as poly-phosphate in the *internal inorganic cells*. This results in a net growth of the PAOs. The second process that occurs in this zone is nitrification of ammonia in the wastewater by the autotrophic organisms. In order to minimise the amount of dissolved oxygen from going into the anoxic zone, the last compartment is typically not aerated. Part of the sludge that contains phosphorus to be removed is wasted while the remaining is returned to the anaerobic zone after thickening in the settler and additional denitrification in the RAS tank.

3. Theory

3.1. Original process model

A model of the Johannesburg BNR activated sludge process based on a distributed parameter approach was recently developed by the authors [3]. We present here only the salient results, which are reused in this work. Readers interested in the detailed development work should refer to the original paper [3]. Hydraulics of the main bioreactors was formulated based on the axial dispersion theory while the kinetics was based on a reduced-order model of the activated sludge model

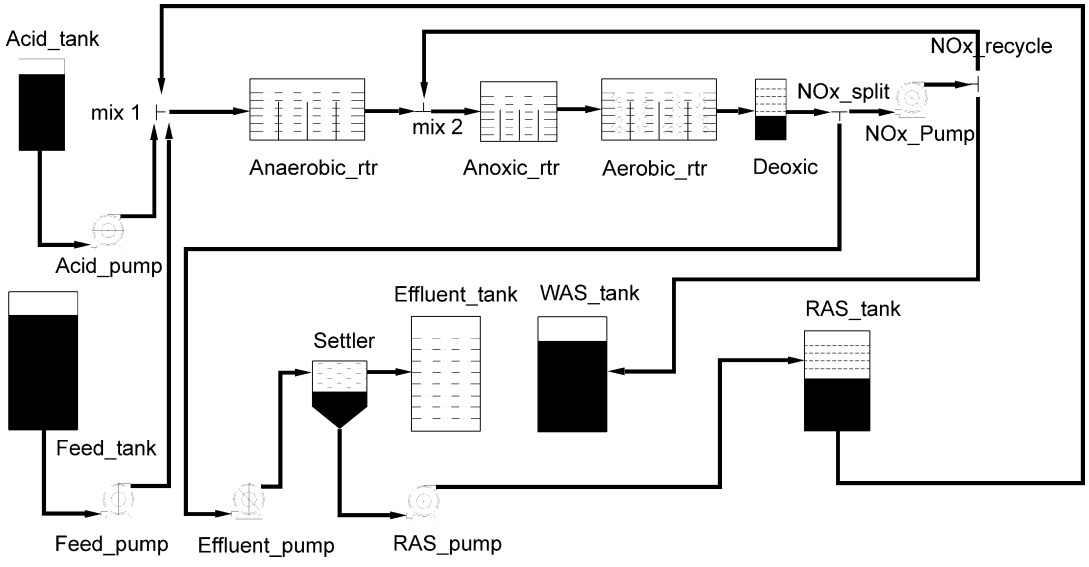

Fig. 1. A BNR activated sludge process.

(ASM) No. 2 of the IAWQ [1]. The reduced-order kinetics of ASM No. 2 employed in our work consisted of eight processes and 11 components for the anaerobic zone, 11 processes and 12 components for the anoxic zone, and 12 processes and 13 components for the aerobic zone. The general mass-balance equation for a component C_k within the compartmentalised anaerobic, anoxic, or aerobic zone is given as follow:

$$\frac{\partial C_k}{\partial t} = \frac{1}{HRT}\cdot\left(\frac{1}{Pe}\frac{\partial^2 C_k}{\partial Z^2} - \frac{\partial C_k}{\partial Z}\right) + \sum_{p=1}^{NP} v_{p,k}\rho_p + M_k \qquad (1)$$

where $v_{p,k}$ [−] is the stoichiometric coefficient for component k and process p; ρ_p (mg l^{-1} h^{-1}) is the rate equation for process p; NP is the number of processes; M is the external mass transfer [mg l^{-1} h^{-1}]; Pe [−] is the Peclet number for a bioreaction zone and HRT [h] is the hydraulic retention time of the zone. The boundary conditions for Eq. (1) are given by Eqs. (2) and (3) as follow:

$$\frac{1}{Pe}\frac{\partial C_k}{\partial Z} = C_k - C_{k,a} \quad Z = 0 \ \ \forall t \qquad (2)$$

$$\frac{\partial C_k}{\partial Z} = 0 \qquad Z = 1 \quad \forall t \qquad (3)$$

The secondary settler was modelled by dividing its length into two distinct zones, namely thickening and clarification. A general expression for solids dynamics in the thickening zone of the Double-Peclet settler model is given by:

$$\frac{\partial X_k}{\partial t} = -\frac{Q_U}{A}\frac{\partial X_k}{\partial Z} - \frac{\partial (V_S X_k)}{\partial Z} + \frac{Q_U L_{\text{bot}}}{A}\frac{1}{Pe_{\text{bot}}}\frac{\partial^2 X_k}{\partial Z^2} \qquad (4)$$

where Q_U (m^3 h^{-1}) is the downward volumetric flowrate, V_S (m s^{-1}) is the gravity settling velocity of Takacs et al. [26], L_{bot} [m] and Pe_{bot} [−] are length of the thickening zone and Peclet number of solids at the thickening zone respectively. Similarly, a general expression for solids dynamics in the clarification zone is given by:

$$\frac{\partial X_k}{\partial t} = -\frac{Q_O}{A}\frac{\partial X_k}{\partial Z} + \frac{\partial (V_S X_k)}{\partial Z} + \frac{Q_O L_{\text{up}}}{A}\frac{1}{Pe_{\text{up}}}\frac{\partial^2 X_k}{\partial Z^2} \qquad (5)$$

where Q_O (m^3 h^{-1}) is the overflow or upward volumetric flowrate, L_{up} (m) and Pe_{bot} (−) are length of the settler and the Peclet number of solids at the clarification zone respectively. We assumed the RAS tank and the deoxic or nonaerated compartment of the aerobic zone to be well mixed and modelled using simple ordinary differential equations (ODEs) for each of the components.

3.2. Discretised model

The distributed parameter systems described by partial differential equations such as Eqs. (1)−(3) can be conceptually viewed as the systems with an infinite dimension. Consequently, to convert a partial differential equation (PDE) into a set of ordinary differential equations (ODE) by using proper numerical schemes, such as method of lines and orthogonal collocation techniques, can also be classified as a model order reduction process. While it is rather straightforward to convert a PDE model to its ODE counterpart, such model reduction process has drawbacks in that it is difficult to ascertain the relationship between the original PDE model and its discretised version and the dynamical properties of both models may be different [27]. It is possible to conceive a reduction directly on the PDE for some systems such as the quasi-linear hyperbolic PDE system described by Christofides and Daoutidis [28,29] or the basic dynamical model described by Winkin [30]. However, knowledge on some classes of nonlinear PDE such as the ones under our study is still incomplete thus rendering this second approach infeasible at this stage.

In our previous work [31], a comparative study on various numerical schemes, including method of lines, orthogonal collocation, and orthogonal collocation on finite elements, has been carried out for solving models of activated sludge processes. It has been concluded that the orthogonal collocation on finite element method with three internal collocation points and four finite elements gives excellent results in all the cases under our study.

It has been noticed that Lefevre et al. [27] have carried out a systematic study on the optimal selection of orthogonal polynomials for solving chemical reactor equations by collocation methods with very useful guidelines. The first part in their work on the determination of the order of polynomials has been considered in our previous work [31]. The second part in their work on the optimal selection of parameters α and β in the Jacobi polynomials $p_N^{\alpha,\beta}$ has demonstrated the significance for the resolution of chemical reactor equations with hotspot. Since the activated sludge processes under our study are more or less operated in the isothermal conditions without sharp temperature changes, to our experience, the numerical results are not very sensitive to the parameters α and β. In this paper, we use the previous results on the order of polynomials with pre-specified parameters α and β, and concentrate on the model order reduction from higher order to reduced order ODE systems.

The standard orthogonal collocation solution [32] to a transient, non-symmetrical problem such as Eq. (1) can be expressed in terms of an interpolation polynomial given by:

$$C_k(z) = \sum_{i=1}^{N+2} C_{k,i}\ell_i(z) \tag{6}$$

where $C_{k,i}$ are concentration of component k at some discrete spatial positions (called collocation points) namely at points z_i and $l_i(z)$ are Lagrange interpolation polynomial defined as:

$$l_i(z) = \prod_{\substack{j=1 \\ j\neq i}}^{N+2} \frac{(z-z_j)}{(z_i-z_j)} = \frac{P_{N+2}(z)}{(z-z_i)\left(\dfrac{\mathrm{d}P_{N+2}(z_i)}{\mathrm{d}z}\right)} \tag{7}$$

where P_{N+2} is called node polynomial of degree $N+2$ and $P_{N+2}(z) = (z-z_1)(z-z_2)\ldots(z-z_N)(z-z_{N+1})$. The points z_i ($i=1, 2, \ldots, N+2$) are taken as zeros of the orthogonal Jacobi polynomials which can be computed using the Newton Raphson method [32].

Using Eq. (6), the first and second order partial derivatives of Eq. (1) can be approximated as follows:

$$\frac{\partial C_k}{\partial Z} = \begin{bmatrix} \dfrac{\mathrm{d}C_{k,2}}{\mathrm{d}Z} \\ \dfrac{\mathrm{d}C_{k,3}}{\mathrm{d}Z} \\ \vdots \\ \dfrac{\mathrm{d}C_{k,N+1}}{\mathrm{d}Z} \end{bmatrix}, \frac{\partial^2 C_k}{\partial Z^2} = \begin{bmatrix} \dfrac{\mathrm{d}^2C_{k,2}}{\mathrm{d}Z^2} \\ \dfrac{\mathrm{d}^2C_{k,3}}{\mathrm{d}Z^2} \\ \vdots \\ \dfrac{\mathrm{d}^2C_{k,N+1}}{\mathrm{d}Z^2} \end{bmatrix} \tag{8}$$

$$\frac{\mathrm{d}C_{k,j}}{\mathrm{d}Z} = \sum_{i=1}^{N+2} A_{ji}C_{k,i} \quad \text{and} \quad \frac{\mathrm{d}^2C_{k,j}}{\mathrm{d}Z^2} = \sum_{i=1}^{N+2} B_{ji}C_{k,i}$$

where $A_{ji} = \dfrac{\mathrm{d}l_i(z_j)}{\mathrm{d}Z}$, $B_{ji} = \dfrac{\mathrm{d}^2l_i(z_j)}{\mathrm{d}Z^2}$ and $j=2\ldots, N+1$, $i=1, \ldots, N+2$.

The ODEs for the compartmentalised anaerobic, anoxic and aerobic zones are obtained by inserting Eq. (8) into Eq. (1). Concentration of component k at the inlet and outlet of each bioreactors, $C_{k,1}$ and $C_{k,N+2}$ are calculated by applying Eq. (8) to Eqs. (2) and (3). Orthogonal collocation solutions to the general distributed parameter model represented by Eqs. (1)–(3) developed in our previous work [31,33] is given by:

$$\frac{\mathrm{d}C_{k,j}}{\mathrm{d}t} = \frac{1}{HRT}.\left(\frac{1}{Pe}\sum_{i=1}^{N+2} B_{j,i}.C_{k,i} - \sum_{i=1}^{N+2} A_{j,i}.C_{k,i}\right) + \sum_{p=1}^{NP} v_{p,k}\rho_{p,j} + M_{k,j} \tag{9}$$

$$C_{k,1} = \frac{A_{N+2,N+2}\left(\sum_{i=2}^{N+1} A_{1,i}.C_i + Pe.C_a\right) - A_{1,N+2}\sum_{i=2}^{N+1} A_{N+2,i}.C_i}{A_{1,N+2}A_{N+2,1} - A_{N+2,N+2}.(A_{1,1}-Pe)} \tag{10}$$

$$C_{k,N+2} = \frac{\left(\dfrac{1}{Pe}A_{1,1}-1\right)\sum_{i=2}^{N+1} A_{N+2,i}.C_i - A_{N+2,1}\left(\dfrac{1}{Pe}\sum_{i=2}^{N+1} A_{1,i}.C_i + C_a\right)}{\dfrac{1}{Pe}A_{N+2,1}A_{1,N+2} - \left(\dfrac{1}{Pe}.A_{1,1}-1\right)A_{N+2,N+2}} \tag{11}$$

Five collocation points or $N=3$ were found to be adequate to model each of the bioreaction zones [3].

For a finite constant length of ΔZ, a discretised, N-layer Double-Peclet settler model for Eqs. (4) and (5) is given by:

$$\frac{\mathrm{d}X_1}{\mathrm{d}t} = \frac{Q_O(X_2-X_1) - V_{S,1}X_1A + \dfrac{Q_OL_{\mathrm{up}}}{Pe_{\mathrm{up}}}\left(\dfrac{X_2-X_1}{\Delta Z}\right)}{A.\Delta Z} \tag{12}$$

for top layer;

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = \frac{Q_O(X_{i+1}-X_i) + V_{S,i-1}X_{i-1}A - V_{S,i}X_iA + \dfrac{Q_OL_{\mathrm{up}}}{Pe_{\mathrm{up}}}\left(\dfrac{X_{i+1}-2X_i+X_{i-1}}{\Delta Z}\right)}{A.\Delta Z} \tag{13}$$

for layers 2 to $(M-1)$ where $i=2, \ldots, (M-1)$ and $M=$feed layer;

$$\frac{\mathrm{d}X_M}{\mathrm{d}t} = \frac{Q_f(X_f-X_M) + V_{S,M-1}X_{M-1}A - V_{S,M}X_MA + \dfrac{Q_OL_{\mathrm{up}}}{Pe_{\mathrm{up}}}\left(\dfrac{X_{M+1}-X_M}{\Delta Z}\right) - \dfrac{Q_UL_{\mathrm{bot}}}{Pe_{\mathrm{bot}}}\left(\dfrac{X_M-X_{M-1}}{\Delta Z}\right)}{A.\Delta Z} \tag{14}$$

for feed layer (layer M);

$$\frac{\mathrm{d}X_j}{\mathrm{d}t} = \frac{Q_U(X_{j-1}-X_j) + V_{S,j-1}X_{j-1}A - V_{S,j}X_jA + \dfrac{Q_UL_{\mathrm{bot}}}{Pe_{\mathrm{bot}}}\left(\dfrac{X_{j+1}-2X_j+X_{j-1}}{\Delta Z}\right)}{A.\Delta Z} \tag{15}$$

for layers $(M+1)$ to $N(N-1)$ where $j=(M+1), \ldots, (N-1)$ and $N=$bottom layer;

$$\frac{\mathrm{d}X_N}{\mathrm{d}t} = \frac{Q_U(X_{N-1}-X_N) + V_{S,N-1}X_{N-1}A - \dfrac{Q_UL_{\mathrm{bot}}}{Pe_{\mathrm{bot}}}\left(\dfrac{X_N-X_{N-1}}{\Delta Z}\right)}{A.\Delta Z} \tag{16}$$

for bottom layer (layer N). A 20-layer Double-Peclet settler model requiring 20 ODEs gave the best results in terms of sum of the square of relative errors when compared with two other state-of-the-art layered settler models [3].

The entire discretised process model represented by Eqs. (9)–(16) and ODEs for the RAS tank and the deoxic consisted of 154 ODEs was simulated using NIMBUS process simulator [34]. This software package has recently been revised and its name changed to "DaeSim Studio" by the developer [35]. The number of ODEs in the model depends on the size of discretisation of the main bioreactors of the BNR activated sludge process (the anaerobic, anoxic and aerobic bioreactors) and the settler. In this paper, we use the previous results on choice of number of ODEs for the discretised model and readers keen in this numerical aspect should refer to our previous work [3,33].

Simulation results based on this model were validated against nine comprehensive experimental data sets from a pilot plant located at the Liverpool Sewerage Treatment plant in Sydney, Australia. The pilot-plant processes a constant flow of about 70 l/h of screened, degritted sewerage with about 1000 l capacity in the main bioreactors. The plant set-up is essentially similar to Fig. 1. Sewerage used in the study was municipal wastewater, sourced from a full-scale wastewater treatment plant located at the same site. Results from the study showed that model prediction was generally good with the exception of acetate in the anaerobic zone and soluble phosphate in the aerobic zone in two experiments.

3.3. Linear state-space model

Linearisation of the BNR activated sludge system is carried out around certain operational points corresponding to a pseudo steady state conditions of an experimental design. Consequently, the linear state space model representations are not unique. For large changes in operational conditions, multiple linearised models could be necesssary. In this paper, the deviations from the experimentally validated pseudo steady state [3] are within small ranges in which a single, linear state space model is valid.

A linear process model represented by the well-known state-space formula is given by:

$$\begin{aligned} \frac{dX}{dt} &= AX + BU \\ Y &= CX + DU \end{aligned} \tag{17}$$

where X is a vector of state variables, U is a vector of input variables and Y is a vector of output variables. A, B, C and D are constant matrices. The variables X, U and Y are usually expressed in terms of perturbation or deviation from their steady-state values. Similarly, we suppose that the reduced-order model for Eq. (1) using the state-space representation is given by:

$$\begin{aligned} \frac{d\hat{X}}{dt} &= A_{red}\hat{X} + B_{red}U \\ \hat{Y} &= C_{red}\hat{X} + D_{red}U \end{aligned} \tag{18}$$

Our basic objective in model-order reduction of state-space models is to find a set of reduced-order matrices A_{red}, B_{red}, C_{red} and D_{red} such that the approximate output $\hat{Y}$ is as close as possible to Y for all inputs U.

Fig. 2 shows a general algorithm for robustly reducing the order of a linear state-space model of a complex biological process based on the balanced or direct truncating model-order reduction of Moore. The first step of the algorithm involves biological model development followed by computation of the constant matrices A, B, C and D by evaluating the *Jacobians* (Step 1, Fig. 2) of the discretised process model. Once the constant matrices are computed, 'redundant' or unnecessary state variables if existed can be eliminated such that the state-

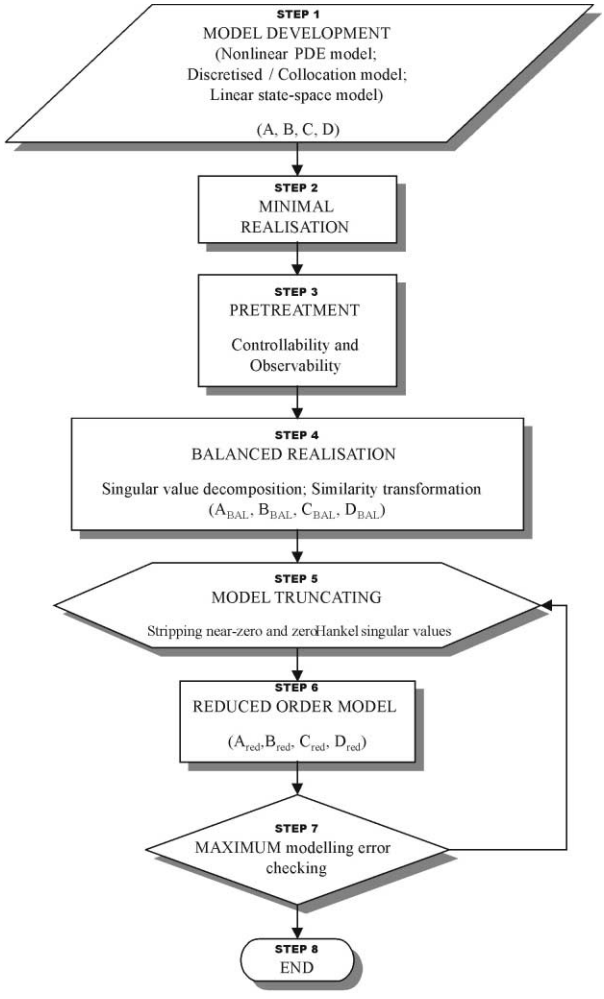

Fig. 2. A general algorithm for robust model-order reduction.

space system has *minimal realisation* (Step 2, Fig. 2). The number of state variables removed should correlate to the physical knowledge of the system. The minimum realisable state-space model must then be 'pretreated' to see how well the system is controllable and observable (Step 3, Fig. 2). These two properties are typically quantified by their *gramians*. The quantitative definitions for controllability and observability gramians of continuous systems are given in the Appendix.

Often it is desirable to acquire a *balanced* representation of state space models (see Appendix for definition) because a balanced system is guaranteed to be asymptotically stable, controllable and observable [16]. Pernebo and Silverman [36] showed that balanced truncated models similarly possess these desirable properties. Having obtained the observability and controllability properties of the system, we could then compute balanced realisation of the system under investigation such that the states are balanced between the input-to-state coupling (controllability) and the state-to-output coupling (observability), namely the system is *internally-balanced*. In quantitative sense, realisation A_{BAL} B_{BAL} and C_{BAL} is internally balanced if their controllability and observability gramians and the *singular values* are identical and diagonal [Eq. (A5), Appendix]. In this special case, they are essentially the *Hankel* singular values (see Appendix for definition). Formulations for the determination of a balanced realisation involving singular value decomposition and similarity transformation as stated in Step 4 of Fig. 2 are detailed in the Appendix.

Weak or least controllable and observable states are quantified by zero and close to zero Hankel singular values are thus conveniently eliminated from the internally balanced realisation (Step 5, Fig. 2). This is mathematically represented as follows:

$$\begin{aligned} G(s) &:= \left(\frac{A}{C}\,\middle|\,\frac{B}{D}\right) = \begin{pmatrix} A_{\mathrm{BAL}} & B_{\mathrm{BAL}} \\ C_{\mathrm{BAL}} & D_{\mathrm{BAL}} \end{pmatrix} \\ &= \begin{pmatrix} A_{\mathrm{BAL},11} & A_{\mathrm{BAL},12} & B_{\mathrm{BAL},1} \\ A_{\mathrm{BAL},21} & A_{\mathrm{BAL},22} & B_{\mathrm{BAL},2} \\ C_{\mathrm{BAL},1} & C_{\mathrm{BAL},2} & D_{\mathrm{BAL}} \end{pmatrix} \\ &\approx G_{\mathrm{red}}(s) = \begin{pmatrix} A_{\mathrm{BAL},11} & B_{\mathrm{BAL},1} \\ C_{\mathrm{BAL},1} & D_{\mathrm{BAL}} \end{pmatrix} \end{aligned} \tag{19}$$

The constants $A_{\mathrm{BAL},11}$, $B_{\mathrm{BAL},11}$, $C_{\mathrm{BAL},1}$ and D_{BAL} in Eq. (19) are the required reduced order matrices (Step 6, Fig. 2). In the final stage (Step 7, Fig. 2) of the robust model-order reduction algorithm, one must define an error bound at which maximum errors incurred in model-order reduction is explicitly quantified. For a k-th order reduced order model, $G_{\mathrm{red}}(s)$ of an n-th order system, $G(s)$, the maximum modelling error is defined as twice the sum of the tail of the Hankel singular values, namely:

$$\left\| G(s) - G_{\mathrm{red}}(s) \right\|_{\infty} \leqslant \text{modelling error} \tag{20}$$

where modelling error $= 2\sum_{i=k+1}^{n}\sigma_i$ and $\sigma =$ Hankel singular value.

A proof of Eq. (20) is detailed by Enns [37]. In addition to the original presentation at a conference, the method has also been explained in many publications in the literature, such as in Chapter 10, Section 10.2 of the book authored by Anderson and Moore [38]. Computation of Steps 5–7 is iterated until an 'acceptable' modelling error is reached. This completes the description for the robust model-order reduction algorithm.

Four other state-of-the-art model-order reduction techniques employed in this work for reducing the BNR activated sludge process model are briefly described as follows.

3.4. Frequency-weighted balanced realisation and model reduction (FW-DT)

The robust model-order reduction technique [hereafter referred to as direct truncating (DT) model reduction technique] can be extended to incorporate weighting functions to emphasise or de-emphasise frequency range of interest such that maximum modelling error is minimised. The method based on the frequency-weighted balanced truncation developed by Enns [37] involving frequency shaping of model reduction error has been widely accepted in the systems science for model and controller order reductions [38]. In quantitative sense, one may wish to compute a reduced-order model, $G_{\mathrm{red}}(s)$ such that:

$$\left\| W_1(G(s) - G_{\mathrm{red}}(s))W_2 \right\|_{\infty}$$

is small. The parameters W_1 and W_2 are given frequency-dependent weighting functions. The basic idea of this technique is to modify the controllability and observability gramians (Step 3, Fig. 2) prior to actual model-order reduction [37,38]. Except for Step 3 in Fig. 2, this technique uses essentially the same procedure for DT model-order reduction technique described earlier. The mathematical details for calculation of frequency-weighted controllability and observability gramians are given in the Appendix. Despite the reportedly good weighted reduced-order model of these two-sided frequency-weighted gramians, its asymptotic stability is not guaranteed. Only one-sided, input frequency-weighting function was proven to result in generically asymptotic stable, reduced model [38]. It is for this reason that we will employ the input frequency-weighted model-reduction in this work.

3.5. Schur decomposition for balanced-truncating model-order reduction (SCHUR BT)

Full-scale, channel-type activated sludge biorectors are known to exhibit significant process time delays [33].

For state-space models of such systems with non-minimum phase or time delay (ie. system which is characterised by right-half plane zero(s) or having Hankel singular value(s) of zero), Safanov and Chiang [18] showed that unitary transformation as used in Moore's balanced truncating technique causes numerical difficulties. *Schur* method addresses this problem by decomposition of the product of controllability and observability gramians, $L_C L_O$ (see Appendix).

3.6. Singular perturbation approximation (SPA) to model reduction

The idea of combining the matured singular perturbation approach of Kokotovic and the good absolute-error truncating properties of balanced realisation theory was first mooted by Fernando and Nicholson [39]. The singular perturbation approximation (SPA) to model reduction assumes that balanced states can be partitioned conformally into two subsystems namely: *strong* subsystem representing slow dynamic species (e.g. heterotrophs, PAO and autotrophs) and *weak* subsystem representing fast dynamic species (e.g. dissolved oxygen and soluble COD). We approximate low frequency behaviour of the system by setting the dynamics of strong subsystem to zero, i.e. a procedure that is termed *fast-mode* model-order reduction. Complete algorithm for the SPA model-order reduction technique is obtained by replacing the direct model truncating step (Step 5, Fig. 1) with computed realisation for strong subsystem.

In the algorithm of Fernando and Nicholson [39], they require that the weak and strong subsystems be checked by calculating the eigenvalues of $A_{\mathrm{BAL},11}$ and $A_{\mathrm{BAL},22}$ matrices such that the latter is an order magnitude higher than the former. Alternatively, one needs to satisfy the following time-scale separation property: $\min_i |\lambda_i(A_{\mathrm{BAL},22})| > \max_j |\lambda_j(A_{11} - A_{12}A_{22}^{-1}A_{21})|$. In this work, the time-scale separation property need not necessarily be satisfied because the Hankel singular value itself acts as a perturbation parameter. One only needs to strip away weakly controllable and weakly observable states ie states with Hankel singular values near-zero or zero to guarantee a good kth order singular perturbation approximant that gives a low infinity error norm bound. Besides being able to emphasise the low frequency ranges, Anderson and Liu [40] showed that the SPA enjoys the same error bound as the DT model-order reduction technique. A slightly less general derivation of the algorithm for SPA model-order reduction, specific to our requirement is detailed in the Appendix.

3.7. The optimal Hankel norm approximation model reduction technique

The optimal Hankel norm approximation (OHKAPP) is yet another important but more complicated model reduction technique considered in this study. As with all the techniques discussed earlier, the OHKAPP exploits the important property of Hankel singular values, namely given that $\sigma_k \gg \sigma_{k+1}$ then state X_{k+1} will affect the input–output system behaviour much less than state X_k. The OHKAPP involves complex manipulation of a balanced realisation G of degree n to find a reduced-order model, G_r of degree k such that Hankel norm (largest Hankel singular values) of the error system $(G-G_r)$ is minimised and specifically, $\|G - G_r\|_H = \sigma_{k+1}$. The kth optimal Hankel norm approximant of order k for G is recovered by separating G_r into stable and anti-stable parts by a stability projection treatment. A specific algorithm to compute the optimal Hankel norm approximants for linear multivariable continuous time system can be found in [17] and an improved algorithm of the OHKAPP that eliminates computation of ill-conditioned balanced transformation can be found in [41].

As with balanced truncating technique, a frequency weighting function can be introduced to the OHKAPP technique to shape the approximate error at frequency ranges of interest [19,40]. However unlike the non-weighted case, applications of the frequency-weighted OHKAPP (FW-OHKAPP) is limited because of difficulty in evaluating its error bound and hence, not necessarily results in optimal approximants. Moreover, the FW-OHKAPP generally requires a square system and in fact, Anderson and Liu [40] showed that the frequency-weighted direct truncating technique almost always yield better approximations than the FW-OHKAPP. Due to these shortcomings, we would not consider the FW-OHKAPP in this study.

4. Model-order reduction of the BNR activated sludge process

Using MATLAB's robust control toolbox [42], we will now investigate the feasibility of applying the direct truncating (DT), FW-DT, SCHUR BT, SPA and OHKAPP techniques to reduce the original, 154 states BNR activated sludge model developed in this work to lower order models. For the FW-DT technique, we use a first order, asymptotically stable input weight with a zero at -50 and a pole at 0.01. Robustness of the five model-order reduction technques to modelling errors will also be investigated.

We identified the inputs to the system as: influent flow; influent soluble acetate; influent soluble phosphate (SPO_4); influent ammonia (SNH_4); returned activated sludge (RAS) flow; external soluble acetate (S_A) flow and nitrate (SNO_3) and nitrite (SNO_2) recycle flow. The desired system outputs are identified as effluent SNH_4; effluent SPO_4 and effluent SNO_3. Based on these identified system inputs and outputs, a linear state-space

model in the form of Eq. (17) for the entire BNR activated sludge process is developed by evaluating the jacobians of the discretised process model. The jacobians of the original BNR activated sludge model and the constant matrices A, B, C and D of the state-space model were computed by using NIMBUS software package and NIMBUS toolbox for MATLAB [34].

5. Results and discussions

5.1. Frequency domain considerations

Fig. 3 shows the computed Hankel singular values (HSVs) of the BNR activated sludge model. Clearly, a large number of the HSVs are zero or close to zero and we should be able to strip or truncate them from the system without introducing pronounced modelling errors. Simulation experience shows that ill conditioning due to numerical problem occurred when we employed the OHKAPP technique to try to reduce the model order down to 40 states.

On the contrary, such problem did not arise when other techniques discussed earlier were employed and in fact, the resulting modelling error (or infinity error bound) are very low, namely less than 10^{-4}. This anomaly may be largely due to the effect of pretreatment of the original state-space model by balancing process. We may recall that truncated, balanced realisation is guaranteed to be not only asymptotically stable but also observable and controllable.

In order to evaluate the five model-order reduction techniques on a same basis, we reduced the model order by truncating another 10 states and the resulting modelling error for each methods is given in Table 1. Results in Table 1 show that modelling errors incurred are still considered remarkably low, despite reduction of the number of states by some 80%. Fig. 4 shows the singular value plots of SPO_4 over a wide range of frequencies for the original (or full) state-space model

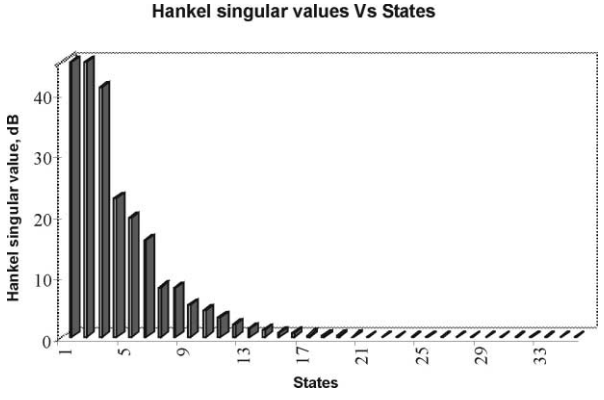

Fig. 3. Hankel singular values of the BNR activated sludge process model. (Total number of states is 154. States not shown have zero or near-zero HSV.)

compared with the reduced, 30-states models developed from the five different model reduction techniques. The resulting reduced-order models exhibited good matches with the full models up to a frequency of about 10 rad/h. Evidently from the singular value plot for SPO4, the DT and SCHUR BT model reduction approaches come closest to the full model at high frequencies, thus suggesting that they possess superior high frequency characteristics. On the other hand, the singular value curves based on the SPA model reduction technique appears to flattened out and fell apart from the full model at high frequencies. One could therefore expect largest truncating errors at these frequencies when the SPA technique is employed. These findings are also true for SNO_3 and SNH_4 (singular value plots not shown).

Results of frequency errors for SPO_4 is given in Fig. 5. Despite incurring the largest error at high frequencies, the SPA technique demonstrated the best results at unity range of frequencies, which is the region of interest in our application to the BNR activated sludge process. This finding is also true for SNO_3 and SNH_4 (frequency error plots not shown).

The FW-DT approach appears to give relatively low errors at low frequency and high frequency ranges, but its errors are undesirably high between 1 and 100 rad/h. This finding is also true for SNO_3 and SNH_4 (frequency error plots not shown). Moreover as shown in Table 1, its infinity error bound or loosely speaking overall modelling error is the highest amongst the methods investigated.

Fig. 5 also shows that employment of the FW-DT model-order reduction technique results in peak frequency response errors of -40 dB for SPO_4 at 6 rad/h. The peak frequency response errors for SNO_3 and SNH_4 (results not shown) are of the same order of magnitude. It is interesting to point out that over and above this frequency, the errors started to drop almost at constant rates and at frequency above 150 rad/h, the FW-DT model-order reduction technique produces smaller errors compared with the OHKAPP and SPA model-order reduction techniques.

The fact that the DT, SCHUR BT and OHKAPP model-order reduction techniques incurred the greatest

Table 1
Results of modelling error [a]

Method	Modelling error $\|G(s) - G_{red}(s)\|_\infty$
DT	0.0267
FW-DT	0.8697
SPA	0.0266
OHKAPP	0.0267
SCHUR BT	0.0267

[a] No. of original states = 154; No. of reduced states = 30. FW-DT uses first order, asymptotically stable input weight with zero = -50 and pole = 0.01.

approximation errors which are rather flat in the low frequency region (Fig. 5) makes them relatively less desirable for application to the BNR activated sludge process. Based on these results and in view of our requirement of obtaining reduced-order model that has good properties at low frequencies, the 30-state model of the BNR activated sludge process developed from the SPA model reduction technique is deemed the best and hence selected for further considerations.

It is interesting to note that results from a comparative study of seven algorithms for model reduction were presented at an international conference [43]. These include four singular value decomposition (SVD) based methods and three moments matching based methods. Three of the four SVD based methods namely: balanced model reduction (DT), singular perturbation approximation (SPA) method, and optimal Hankel norm approximation (OHKAPP) are investigated in this paper except the approximate balanced reduction method. The main advantage of the approximate balanced reduction method is the relatively less computational load of the same order of magnitude with certain degree of model quality deterioration.

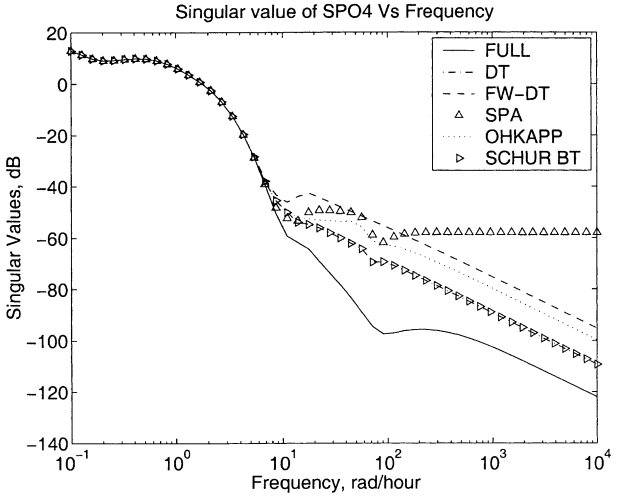

Fig. 4. Singular value plot for SPO_4 using full model and reduced-order models.

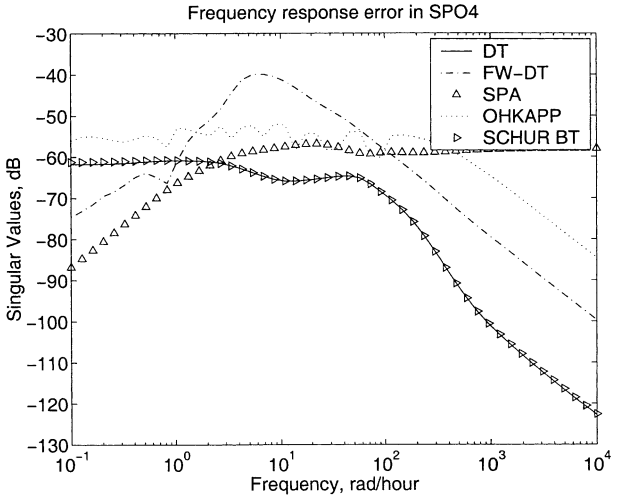

Fig. 5. Frequency model reduction error in SPO_4.

Since accuracy of model reduction far outweigh savings in computing time of the same order of magnitude, the approximate balanced reduction is not considered in our work. In this work, more SVD based methods, such as frequency weighted direct truncating (FW-DT), Schur decomposition for balanced truncating model order reduction, and the combination of singular perturbation approach with balanced truncating, have been addressed, leading to a more complete comparative study on model order reduction.

The three moment matching based methods lead to even lower qualities of reduced models but with relatively higher computational efficiencies than the SVD based methods as concluded by Guercin et al. [43]. They further pointed out that the three moment matching methods employed in their study has other critical drawbacks namely: (i) the reduced model is not guaranteed to be stable; (ii) no global error bounds exist and (iii) model reduction process is non automated requiring try and error procedure. Again, since accuracy of model reduction far outweigh savings in computing time of the same order of magnitude, the moment matching based methods are omitted from our work and the SVD based model-order reduction methods should be recommended.

In general, the singular perturbation approximation approach (SPA) always produces very good reduction errors at low frequencies, up to about 1 rad/h for BNR activated sludge processes but produces larger errors at high frequencies than the other SVD based model-order reduction methods investigated. The direct truncating (DT), SCHUR balanced truncating and the optimal Hankel norm approximation (OHKAPP) methods on the other hand tend to have smaller errors at high frequencies and larger errors at low frequencies. Performance of the frequency weighted direct truncating (FW-DT) method depends entirely on the user-supplied frequency weights and whether to emphasise or de-emphasise frequency range of interest and the frequency error is not guaranteed to be of the same order of magnitude as the other techniques. Results from this study nevertheless showed that a first order, asymptotically stable input weight is suitable for emphasising the lower frequency regions while de-emphasising the higher frequency regions prior to model reduction for the BNR activated sludge process model. Further tuning of the frequency weights by trial and error is necessary if one wishes to further minimise the maximum modelling errors.

5.2. Time domain considerations

In order to see how well the linear, reduced-order model behaves under dynamic conditions, step-tests were performed on the disturbance variables, identified as inputs to the system, namely influent S_A; influent flow; influent SPO_4 and influent SNH_4.

Fig. 6 shows that responses of the reduced-order model to a 30% reduction in the influent S_A are reasonably well verified, when compared to the nonlinear (collocation) model and the full linear (state space) model. Increased in the effluent SPO_4 is attributable to the reduced growth of the X_{PHA} in the anaerobic zone and hence reduction in the net uptake of SPO_4 by the X_{PP} in the aerobic zone. Increased in the effluent SNO_3 and SNH_4 may be largely due to not only reduced in the anoxic growth of X_H but also the denitrifying PAOs. SNH_4 in the aerobic zone is also affected by the reduction in X_{AUT} growth owing to the increased SPO_4 concentration in the aerobic zone.

Fig. 7 shows that responses of the reduced-order model to a 10% increase in the influent flow are also reasonably well verified, when compared with the nonlinear as well as the original (or full) linear models. Steady-state errors appear to be rather small since the absolute concentration of all species is almost unchanged. Increased in influent flow means that the hydraulic retention time (*HRT*) within the system is reduced. In the anaerobic zone, S_A uptake by X_{PP} and SPO_4 released by X_{PAO} drops. This results in a net reduction in SPO_4 uptake by PAOs in the aerobic zone. Similarly, growth of X_{AUT} in the aerobic zone drops due to reduction in *HRT* and results in less SNH_4 converted to SNO_3, namely concentration of both components will undoubtedly rises.

Fig. 8 shows that step responses to a 10% increase in the influent SPO_4 are similarly well fitted by the models. Clearly without additional S_A in the system, X_{PHA} cannot release energy sufficiently to metabolise X_{PP} for uptake of excess SPO_4 in the aerobic zone. Dynamics of SNO_3 and SNH_4 are relatively unchanged.

Fig. 9 shows that increased in the response of SPO_4 to a 30% increase in the influent SNH_4 produced a significant difference in the steady-state gains (about 25%). This difference should be treated as model uncertainty due to linearisation. We consider a step change of 30% in the influent SNH_4 concentration as the worst case scenario in the open-loop dynamic step response of the effluent SPO_4. This lost of process information is largely

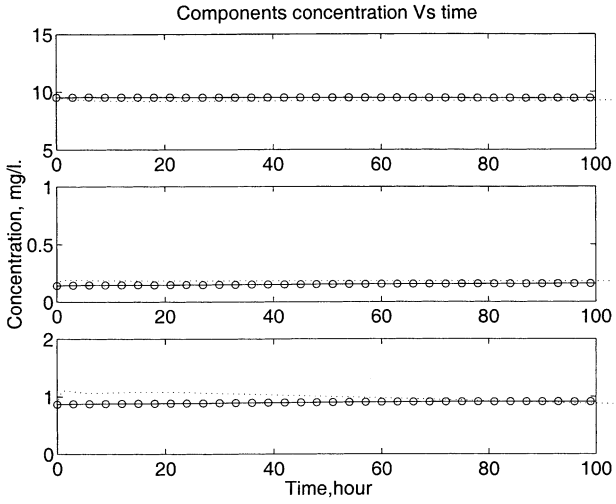

Fig. 7. Effects of influent flow disturbance (10% step up) to the dynamics of effluent SNO_3 (Top), SNH_4 (Mid) and SPO_4 (Bot). '–' = full linear (state-space) model; '○' = reduced linear (state-space) model; '...' = nonlinear (collocation) model.

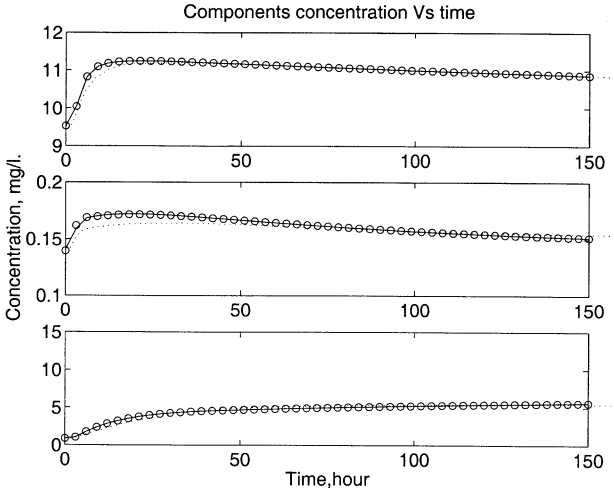

Fig. 6. Effects of influent S_A disturbance (30% step down) to the dynamics of effluent SNO_3 (Top), SNH_4 (Mid) and SPO_4 (Bot). '–' = full linear (state-space) model; '○' = reduced linear (state-space) model; '...' = nonlinear (collocation) model.

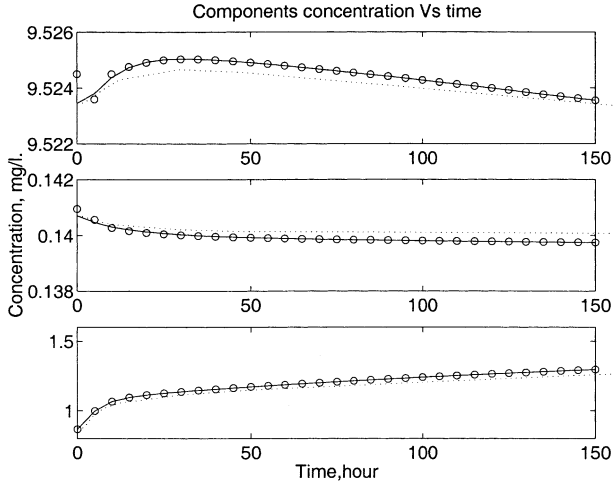

Fig. 8. Effects of influent SPO_4 disturbance (10% step up) to the dynamics of effluent SNO_3 (Top), SNH_4 (Mid) and SPO_4 (Bot). '–' = full linear (state-space) model; '○' = reduced linear (state-space) model; '...' = nonlinear (collocation) model.

related to the behaviour of phosphorus accumulating organisms (PAO), responsible for the release of SPO_4 at the aerobic zone or outlet of the BNR activated sludge process. It is envisaged that smaller step changes to the influent SNH_4 would reduce the mismatch between the nonlinear (collocation) and the linear state-space models. If the reduced-order, steady-state model is to be employed for control of the BNR activated sludge process, the controller design should be robust to such uncertainty.

6. Conclusions

Five state-of-the-art model-order reduction techniques namely: DT, FW-DT, SCHUR-BT, SPA and OHKAPP have been investigated in this study with the view of selecting an appropriate technique to reduce the order of a high-dimensional, nonlinear PDE model of a complex biological process to a 'manageably' smaller order model.

The original 154 dimensional, full linear (state-space) model of the BNR activated sludge process is reduced to 30 dimensional, reduced-order model with very low modelling errors. Based on the results of frequency–error plots, the reduced-order model generated from the Singular Perturbation Approximation technique gives the lowest errors in low frequency ranges and hence is deemed most suitable for controller design and other real-time applications.

Dynamics of species of interest, namely SNO_3, SNH_4 and SPO_4 using the selected reduced-order state-space model fits the original, nonlinear distributed parameter model reasonably well in most of the cases under study. However, there exist significant steady-state errors of as much as 25% in the response of SPO_4 to disturbance in the influent SNH_4. Such errors should nevertheless be treated as modelling uncertainty and design of controllers for this process should then be robust to this uncertainty.

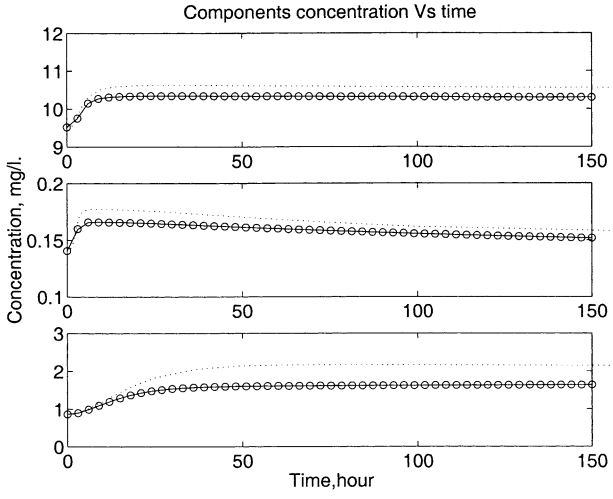

Fig. 9. Effects of influent SNH_4 disturbance (10% step up) to the dynamics of effluent SNO_3 (Top), SNH_4 (Mid) and SPO_4 (Bot). '–' = full linear (state-space) model; '○' = reduced linear (state-space) model; '...' = nonlinear (collocation) model.

A_{SP}, B_{SP}, C_{SP} and D_{SP} are the required reduced-order constant matrices.

Appendix. Controllability and observability gramians

Controllability is a property that represents the coupling between input and the states while observability is a property that represents the coupling between the outputs and the states. These properties are mathematically quantified by their gramians as:

$$L_c = \int_0^\infty e^{At} B B^T \mathrm{e}^{A^T t} \mathrm{d}t \tag{A1}$$

and

$$L_o = \int_0^\infty e^{At} C^T C \mathrm{e}^{At} \mathrm{d}t \tag{A2}$$

respectively. Eqs. (A1) and (A2) exist if and only if the system is asymptotically stable or in quantitative terms, the eigenvalues of A [denoted as $\lambda_i(A)$] are strictly in the left-half plane or having negative values including zero. Unique solutions to L_c and L_o are typically computed by solving two matrix Lyapunov equations given by:

$$AL_C + L_C A^T + BB^T = 0 \tag{A3}$$

and

$$A^T L_o + L_o A + C^T C = 0 \tag{A4}$$

respectively.

A1. Balancing and truncating

A linear continuous system is said to be asymptotically stable if its eigenvalues or poles lie within the *unit circle*, i.e. their absolute values are less than unity. Mathematical representation of a balanced system is given by:

$$\hat{L}_C = \hat{L}_O = T^T L_o T = T^{-1} L_C T^{-T} = \sum \tag{A5}$$

where $\sum = \mathrm{diag}\{\sigma_1, \sigma_2, .. , \sigma_n\}$ and $\sigma_1 > \sigma_2 > .. > \sigma_n > 0$. The variable σ represents singular values of the system. The variable Σ is obtained by performing a singular value decomposition on the product of square roots of the gramians, namely:

$$U\sum V = R_o R_c \tag{A6}$$

where U and V are unitary matrices ie. $U^TU=I$ and $V^TV=I$. The variable R_o is the product square root matrix of the observability gramian ie. $R_o=\sqrt{L_o}$ and the variable R_c is the product square root matrix of the controllability gramian ie. $R_c=\sqrt{L_c}$.

The variable T is the nonsingular similarity transformation matrix that is defined by:

$$T= R_c V \Sigma^{-1/2} \tag{A7}$$

The balanced realisation, A_{BAL}, B_{BAL}, C_{BAL} and D_{BAL} is obtained in step 4 of Fig. 2 as follow:

$$A_{\mathrm{BAL}} = T^{-1}AT;\ B_{\mathrm{BAL}} = T^{-1}B;\ C_{\mathrm{BAL}} = CT \text{ and } D_{\mathrm{BAL}} = D \tag{A8}$$

The similarity transformation in Eq. (A8) is an important matrix operation in that although the gramians are modified, the original transfer function is preserved, namely:

$$G_{\mathrm{BAL}}(s) := \left(\frac{A_{\mathrm{BAL}}}{C_{\mathrm{BAL}}}\middle|\frac{B_{\mathrm{BAL}}}{D_{\mathrm{BAL}}}\right) = \left(\frac{T^{-1}AT}{CT}\middle|\frac{T^{-1}B}{D}\right) = G(s) = C(sI-A)^{-1}B+D \tag{A9}$$

A2. Hankel singular value

Hankel singular values of a system are singular values that are invariant under state-space transformation and are indicative of the controllability and observability properties of the system states in a *balanced* state-space realisation. This property which is fundamental in the model reduction techniques used throughout this work is important in that it provides a mean for determining those states that can be retained or removed. In quantitative sense, the Hankel singular values of a stable process, $G(s)$ are defined as:

$$\sigma_i(G(s)) = \{\lambda_i(L_C L_O)\}^{1/2} \tag{A10}$$

where $i=1, .. , n$ and $\sigma_1 \geqslant \sigma_2 .. \geqslant \sigma_n \geqslant 0$.

A3. Frequency-weighted controllability and observability gramians

Consider the asymptotically stable input and output frequency weighting functions are respectively given by:

$$W_i(s) := \left(\frac{A_i}{C_i}\middle|\frac{B_i}{D_i}\right) \text{ and } W_o(s) := \left(\frac{A_o}{C_o}\middle|\frac{B_o}{D_o}\right) \tag{A11}$$

If we cascade the input weighting function $W_i(s)$ of Eq. (A11) with system of Eq. (17), we have augmented matrices given by:

$$\hat{A}_c(s) := \left(\frac{A}{0}\middle|\frac{BC_i}{A_i}\right) \text{ and } \hat{B}_c(s) := \begin{pmatrix} BD_i \\ B_i \end{pmatrix} \tag{A12}$$

Similarly, if we cascade system of Eq. (17) to the output weighting function, $W_o(s)$ of Eq. (A11) we have:

$$\hat{A}_c(s) := \left(\frac{A}{B_oC}\middle|\frac{0}{A_o}\right) \text{ and } \hat{C}_o(s) := [\ D_oC\ C_o\ ] \tag{A13}$$

The augmented controllability and observability gramians can be computed by solving two augmented system Lyapunov equations given by:

$$\hat{A}_c\hat{L}_c + \hat{L}_c\hat{A}_c + \hat{B}_c\hat{B}_c = 0 \tag{A14}$$

and

$$\hat{A}_o\hat{L}_o + \hat{L}_o\hat{A}_o + \hat{C}_o\hat{C}_o = 0 \tag{A15}$$

where

$$\hat{L}_c(s) := \begin{pmatrix} L_c & * \\ * & * \end{pmatrix} \text{ and } \hat{L}_o(s) := \begin{pmatrix} L_o & * \\ * & * \end{pmatrix} \tag{A16}$$

The frequency-weighted controllability and observability gramians, L_c and L_o can then easily be extracted from Eq. (A16) noting that the new matrices L_c and L_o should have identical size as with the system matrix A.

A4. Schur decomposition approach

The product of L_CL_O in Schur form is given by:

$$V^TL_CL_OV = \begin{pmatrix} \lambda_1 & * & * & \cdots & & * \\ 0 & \lambda_2 & * & \cdots & & * \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & * \\ 0 & \cdot & \cdot & & 0 & \lambda_n \end{pmatrix} \tag{A17}$$

where λ_i are eigenvalues of L_CL_O and V^T and V are left and right eigen spaces . The latter two variables are used for computation of similarity transformation matrix, T and hence the balance realisation.

A5. Singular perturbation approximation approach

Suppose we have a balanced and linear state-space model given by:

$$\begin{aligned}\frac{dX}{dt} &= A_{BAL}X + B_{BAL}U\\ Y &= C_{BAL}X + D_{BAL}U\end{aligned} \tag{A18}$$

Suppose again the states can be partitioned conformally into two subsystems namely: *strong* subsystem representing slow dynamic species (e.g. heterotrophs, PAO and autotrophs) and *weak* subsystem representing fast dynamic species (e.g. dissolved oxygen and soluble COD), namely:

$$\begin{pmatrix}\frac{dX_{strong}}{dt}\\ \frac{dX_{weak}}{dt}\end{pmatrix} = \begin{pmatrix}A_{BAL,11} & A_{BAL,12}\\ A_{BAL,21} & A_{BAL,22}\end{pmatrix}\begin{pmatrix}X_{strong}\\ X_{weak}\end{pmatrix} + \begin{pmatrix}B_{BAL,1}\\ B_{BAL,2}\end{pmatrix}U \tag{A19}$$

$$\begin{pmatrix}Y_{strong}\\ Y_{weak}\end{pmatrix} = \begin{pmatrix}C_{BAL,1} & C_{BAL,2}\end{pmatrix}\begin{pmatrix}X_{strong}\\ X_{weak}\end{pmatrix} + DU;\quad X_i(0) = 0 \tag{A20}$$

The derivatives of X_{weak} in Eq. (A19) represent vector of species with fast dynamics or eigen values with large negative real parts. Based on the singular perturbation approach, the low frequency behaviour of the system can be approximated by setting the former to their steady-state values. This is a procedure that is termed *fast-mode* model-order reduction. In control engineering, this means mode that is outside the control system bandwidth. Applying fast mode reduction to Eq. (A19), we have:

$$0 = A_{BAL,21}X_{strong} + A_{BAL,22}X_{weak} + B_{BAL,2}U$$

or

$$X_{weak} = -A_{BAL,22}(A_{BAL,21}X_{strong} + B_{BAL,2}U) \tag{A21}$$

Clearly Eq. (A21) exists if and only if $A_{BAL,22}$ is invertible or nonsingular. Green and Limebeer [44] showed that since $A_{BAL,11}$ is asymptotically stable and $A_{BAL,11}$, $B_{BAL,1}$ and $C_{BAL,1}$ are minimal, $A_{BAL,22}$ is also stable and is invertible or nonsingular. Inserting it into the strong subsytem in Eqs. (A19) and (A20), we have:

$$\frac{dX_{strong}}{dt} = (A_{BAL,11} - A_{BAL,12}A_{BAL,22}A_{BAL,21})X_{strong} + (B_{BAL,1} - A_{BAL,12}A_{BAL,22}B_{BAL,2})U \tag{A22}$$

and

$$Y_{strong} = \left(C_{BAL,1} - C_{BAL,2}A_{BAL,22}A_{BAL,21}\right)X_{strong} + \left(D - C_{BAL,2}A_{BAL,22}B_{BAL,2}\right)U \tag{A23}$$

The singularly perturbed reduced-order model is thus given by:

$$G_{red}(s) := \left(\begin{array}{c|c}A_{SP} & B_{SP}\\ \hline C_{SP} & D_{SP}\end{array}\right) \tag{A24}$$

where:

$$\begin{aligned}A_{SP} &= A_{BAL,11} - A_{BAL,12}A_{BAL,22}^{-1}A_{BAL,21}\\ B_{SP} &= B_{BAL,1} - A_{BAL,12}A_{BAL,22}^{-1}B_{BAL,2}\\ C_{SP} &= C_{BAL,1} - C_{BAL,2}A_{BAL,22}^{-1}A_{BAL,21}\\ D_{SP} &= D - C_{BAL,2}A_{BAL,22}^{-1}B_{BAL,2}\end{aligned}$$

References

[1] M. Henze, W. Gujer, T. Mino, T. Matsuo, M.C. Wentzel, G.R. van Marais, Activated Sludge Model No. 2—IAWQ Scientific and Technical Report No. 2, IAWQ, London, 1994.

[2] P.S. Barker, P.L. Dold, General model for biological nutrient removal activated sludge systems: model presentation, Water Environment Research 69 (5) (1997) 969–984.

[3] T.T. Lee, F.Y. Wang, R.B. Newell, Distributed parameter approach to the dynamics of complex biological processes, AIChE Journal 45 (10) (1999) 2245–2268.

[4] M. Henze, C.P.L. Grady, W. Gujer, G.R. van Marais, T. Matsuo, Activated Sludge Model No. 1—IAWPRC Scientific and Technical Report No. 1, IAWPRC, London, 1987.

[5] U. Jeppsson, G. Olsson, Reduced order models for on-line parameter identifications of the activated sludge process, Water Science Technology 28 (11/12) (1993) 173–183.

[6] M.J.M. Jannssen, L.N. Hopkins, B. Petersen, P.A. Vanrolleghem, Reduction of an activated sludge process model to facilitate controller tuning, in: R. Van Landeghem (Ed.), Proceedings of the 14th European Simulation Multiconference, Society for Computer Simulation International (SCS), 2000, pp. 697–701.

[7] E. Ayesa, G. Oyarbide, L. Larrea, J.L. Garcia-Heras, Observability of reduced order models—application to a model for control of Alpha process, Water Science Technology 31 (2) (1995) 161–170.

[8] M.A. Steffens, P.A. Lant, R.B. Newell, A systematic approach for reducing complex biological wastewater treatment models, Water Research 31 (3) (1997) 590–606.

[9] P.V. Kokotovic, R.C. O'Malley, P.I. Sannuti, Singular perturbation and order reduction in control theory—an overview, Automatica 12 (1976) 123–132.

[10] E.J. Davison, A method for simplifying linear dynamic systems, IEEE Transactions Automatic Control AC-11 (1966) 93–101.

[11] M.P. Chidambara, Further comments by M.R. Chidambara, IEEE Transactions Automatic Control AC-12 (1969) 799–800.

[12] G. Langholz, D. Feinmesser, Model reduction by Routh Approximations, International Journal Systems Science 9 (5) (1978) 493–496.

[13] D.A. Wilson, Model reduction for multivariable systems, International Journal Control 20 (1974) 57–64.
[14] A. Lepschy, G.A. Mian, U. Viaro, A method for optimal linear model reduction, System Control Letters 8 (1987) 405–410.
[15] R. Luus, Optimization in model reduction, International Journal Control 32 (5) (1980) 741–747.
[16] B.C. Moore, Principal components analysis in linear systems: Controllability, observability and model reduction, IEEE Transactions Automatic Control AC-26 (1981) 17–32.
[17] K. Glover, All optimal Hankel-norm approximations of linear multivariable systems and their L_∞ error bounds, International Journal Control 39 (6) (1984) 1115–1193.
[18] M.G. Safanov, R.Y. Chiang, A schur method for balanced-truncation model reduction, IEEE Transactions Automatic Control AC-34 (1989) 729–733.
[19] A. Latham, B.D.O. Anderson, Frequency-weighted optimal Hankel-norm approximation of stable transfer functions, System Control Letters 5 (1985) 229–236.
[20] U.M. Al-Saggaf, G.F. Franklin, Model reduction via balanced realisations: an extension and frequency weighting techniques, IEEE Transactions Automatic Control AC-33 (7) (1988) 687–692.
[21] P.A. Wisnewski, F.J. Doyle III, A reduced model approach to estimation and control of a Kamyr digester, Computers Chemical Engineering, 20 (Suppl.), 1996, S1053-S1058.
[22] T. Watanabe, K. Yasuda, R. Yokoyama, Balanced truncating preserving poles in a specified disk and it's applicaiton to the reduction of H_ω controller, IECON Proceedings, 3 (1996) 1359–1364.
[23] Y. Halevi, A. Zlochevsky, T. Gilat, Parameter-dependent model order reduction, International Journal Control 66 (3) (1997) 463–485.
[24] C. Beck, J. Doyle, K. Glover, Model reduction of multidimensional and uncertain systems, IEEE Transactions Automatic Control, 41(10) (1996) 1466–1477.
[25] T.T. Lee, F.Y. Wang, R.B. Newell, Robust multivariable control of complex biological process (in preparation).
[26] L. Takacs, G.G. Patry, D. Nolasco, A dynamic model of the clarification-thickening process, Water Research 25 (1991) 1263–1271.
[27] L. Lefevre, D. Dochain, S. Feyo de Azevedo, A. Magnus, Optimal selection of orthogonal polynomials applied to the integration of chemical reactor equations by collocation methods, Computers and Chemical Engineering 24 (2000) 2571–2588.
[28] P.D. Christofides, P. Daoutidis, Feedback control of hyperbolic PDE systems, AIChE Journal 42 (11) (1996) 3063–3086.
[29] P.D. Christofides, P. Daoutidis, Robust control of hyperbolic PDE systems, Chemical Engineering Science 53 (1) (1998) 85–105.
[30] J.J. Winkin, D. Dochain, P. Ligarius, Dynamical analysis of distributed parameter tubular reactors, Automatica 36 (2000) 349–361.
[31] T.T. Lee, F.Y. Wang, R.B. Newell, Dynamic simulation of bioreactor systems using orthogonal collocation on finite elements, Computers and Chemical Engineering 23 (1999) 1247–1262.
[32] J. Villadsen, M.L. Michelsen, Solution of Differential Equation Models by Polynomial Approximation, Prentice Hall, Englewood Cliffs, NJ, 1978.
[33] T.T. Lee, F.Y. Wang, R.B. Newell, Dynamic modelling and simulation of activated sludge process using orthogonal collocation approach, Water Research 32 (1) (1999) 73–86.
[34] R.B. Newell, I.T. Cameron, NIMBUS Users manual, Computer-Aided Process Engineering Centre, Department of Chemical Engineering, The University of Queensland, Brisbane, Australia, 1991.
[35] R.B. Newell and I.T. Cameron, DaeSim Studio Modeling and Simulation Environment, DaeSim Technologies Pty Ltd, Brisbane, Australia, 2000 Available from <http://www.daesim.com>.
[36] L. Pernebo, L.M. Silverman, Model reduction via balanced state space representations, IEEE Transactions Automatic Control AC-27 (1982) 2.
[37] D.F. Enns, Model reduction with balanced realizations: an error bound and a frequency weighted generalization, in: Proceedings 23rd Conference Decision Control, Las Vegas, NV, 1984, pp. 127–1321.
[38] B.D.O. Anderson, J.B. Moore, Linear Quadratic Methods, Prentice Hall, 1989.
[39] K.V. Fernando, H. Nicholson, Singular perturbation model reduction of balance systems, IEEE Transactions Automatic Control AC-27 (2) (1982) 466–468.
[40] B.D.O. Anderson, Y. Liu, Controller reduction: concepts and approaches, IEEE Transactions Automatic Control 34 (8) (1989) 802–811.
[41] M.G. Safanov, R.Y. Chiang, D.J.N. Limebeer, Optimal Hankel model reduction for nonminimal systems, IEEE Transactions Automatic Control 35 (4) (1990) 494–502.
[42] M.G. Safanov, R.Y. Chiang, Robust Control Toolbox Users' Guide, The Mathworks Inc, Natick, 1991.
[43] S. Gugercin, A.C. Antoulas, A comparative study of 7 algorithms for model reduction, in: Proc. 39th IEEE Conf, Decision and Control, 2000, pp. 2367–2372.
[44] M. Green, D.J.N. Limbeer, Linear Robust Control, Prentice Hall, Inc., Englewood Cliffs, NJ, 1995.