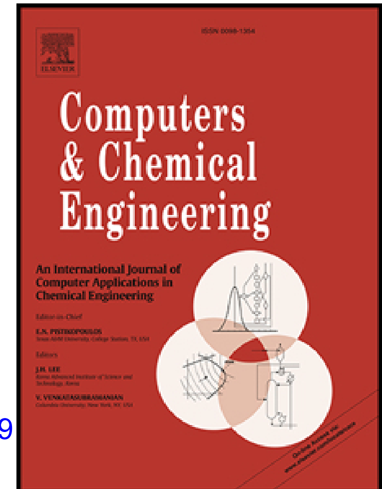
The final publication is available at Elsevier via https://dx.doi.org/10.1016/j.compchemeng.2018.08.029 © 2018. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/

Accepted Manuscript



Optimization and Control of a Thin Film Growth Process: A Hybrid First Principles/Artificial Neural Network Based Multiscale Modelling Approach

Donovan Chaffart , Luis A. Ricardez-Sandoval

PII: S0098-1354(18)30637-9
DOI: https://doi.org/10.1016/j.compchemeng.2018.08.029
Reference: CACE 6202

To appear in: *Computers and Chemical Engineering*

Received date: 21 June 2018
Revised date: 20 July 2018
Accepted date: 20 August 2018

Please cite this article as: Donovan Chaffart , Luis A. Ricardez-Sandoval , Optimization and Control of a Thin Film Growth Process: A Hybrid First Principles/Artificial Neural Network Based Multiscale Modelling Approach, *Computers and Chemical Engineering* (2018), doi: https://doi.org/10.1016/j.compchemeng.2018.08.029

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Highlights

- A novel hybrid multiscale model is developed to simulate thin film deposition
- Continuum models and stochastic PDE used to capture deposition multiscale behaviour
- Artificial neural networks used to predict SPDE's model parameters
- Hybrid multiscale model validated using kinetic Monte Carlo based thin film model
- Hybrid multiscale model used to perform optimization & control on thin film growth

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Optimization and Control of a Thin Film Growth Process: A Hybrid First Principles/Artificial Neural Network Based Multiscale Modelling Approach

Donovan Chaffart, Luis A. Ricardez-Sandoval*

Department of Chemical Engineering, University of Waterloo, Waterloo, Canada, N2L 3G1

ABSTRACT

This work details the construction and evaluation of a low computational cost hybrid multiscale thin film deposition model that couples artificial neural networks (ANNs) with a mechanistic (first-principles) multiscale model. The multiscale model combines continuum differential equations, which describe the transport of the precursor gas phase, with a stochastic partial differential equation (SPDE) that predicts the evolution of the thin film surface. In order to allow the SPDE to accurately predict the thin film growth over a range of system parameters, an ANN is developed and trained to predict the values of the SPDE coefficients. The fully-assembled hybrid multiscale model is validated through comparison against a kinetic Monte Carlo-based thin film multiscale model. The model is subsequently applied to a series of optimization and control studies to test its performance under different scenarios. These studies illustrate the computational efficiency of the proposed hybrid multiscale model for optimization and control applications.

Keywords: Artificial Neural Networks; Multiscale Modelling; Thin Film Deposition; Hybrid Modelling; Stochastic Partial Differential Equations.

*Corresponding author. Tel.: + 1 519 888 4567x38667; fax.: + 1 519 888 4347. *E-mail Address:* laricard@uwaterloo.ca (L.A. Ricardez-Sandoval)

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

1. Introduction

The thin-film deposition process is crucial to the fabrication and manufacture of semiconductors and microelectronic components (Baumann et al., 2001; Seshan and Schepis, 2012). This has driven the development of model-based optimization and control schemes to improve the quality of the thin films and minimize the deposition cost (Theodoropoulou et al., 1998; Lou and Christofides, 2003; Middlebrooks and Rawlings, 2007). The behaviour of a thin-film growth process is governed by the deposition, desorption, and migration of the film precursor species on a molecular level, which are themselves dependent upon macroscale processes such as mass, energy, and momentum transport. Accordingly, optimal thin film process improvement studies require control of the molecular-level events through manipulation of the system's macroscopic inputs (Ulissi et al., 2013). Consequently, process improvement studies for thin film growth are most readily performed using multiscale modelling techniques that capture the relevant molecular and bulk phenomena on the scales in which they occur (Vlachos, 2005). Multiscale models for thin film growth typically combine continuum partial differential equations (PDEs), which describe the macroscopic behaviour of the gas-phase precursor species, with kinetic Monte Carlo (kMC) simulations that capture the microscopic kinetic events responsible for the thin film growth (Lam and Vlachos, 2001). This multiscale approach has been previously applied to perform a number of different optimization and control studies for thin film deposition (Rasoulian and Ricardez-Sandoval, 2015a, 2015b, 2016; Christofides and Armaou, 2006; Christofides et al., 2009). Similar multiscale methodologies have been used to model a variety of different processes, such as plasma-enhanced chemical vapour deposition (Crose et al., 2015, 2017), atomic layer deposition (Dwivedi and Adomaitis, 2009; Adomaitis, 2010), polymer branch growth (Rawlston et al., 2011, 2010), and catalytic reactor systems (Chaffart et al., 2016; Chaffart and Ricardez-Sandoval, 2018). However, this PDE/kMC-based multiscale approach is limited in its application for process improvement studies, as the kMC model is not of closed form and results in prohibitive computational costs (Ricardez-Sandoval, 2011). Consequently, there is a need to develop techniques and tools to decrease the computational cost of process improvement studies for thin film deposition and other multiscale systems. One method is to develop efficient optimization and control policies in order to decrease the computational burden of process improvement in thin film deposition systems. For example, approximate dynamic programming techniques can be considered as a promising option to perform process control at reduced computational costs and in the presence of uncertainty (Lee and Lee, 2005, 2009; Lee and Wong, 2010). Another approach is to develop reduced-order and data-driven models that predict the values of key system observables as a function of designated control parameters. This approach has been previously used to perform efficient process improvement studies for systems with computationally-prohibitive dynamic models such as hydraulic fracturing (Siddhamshetty et al., 2018a, 2018b), copper electrodeposition (Rusli et al., 2006), and thin film deposition (Rasoulian and Ricardez-Sandoval, 2015a, 2015b). These reduced-order modelling techniques are useful as they can simulate the underlying multiscale models at reduced computational costs with minimal loss in the model accuracy, and they can be readily incorporated into existing optimization and control schemes. However, these low-order models are limited to predicting key properties and may be unable to capture detailed fine-scale information about

ACCEPTED MANUSCRIPT

the system that are readily available from kMC simulations, such as the temporal and spatial evolution of the microscopic structure of a thin film surface in thin film deposition models.

Stochastic partial differential equations (SPDEs) provide an alternative approach that can be used to efficiently simulate the surface growth in a thin film system. SPDEs use partial differential equations to capture the spatial and transient evolution of the system, and they include random terms to simulate stochastic process changes (Walsh, 1986). SPDEs are closed-form models, and thus they are computationally efficient compared to kMC-based models (Ni and Christofides, 2005a, 2005b). In addition, SPDEs are able to capture detailed information about the multiscale system, such as the microscopic thin film surface evolution, that may be extremely difficult to predict using reduced-order and data-driven models. Consequently, a number of SPDEs have been developed to describe the growth and evolution of thin film surfaces. Examples of this include the Edwards-Wilkinson equation developed to model surface height fluctuations (Edwards et al., 1982; Vvedensky, 2003), the Kardar-Parisi-Zhang equation developed to simulate randomly-growing interfaces (Kardar et al., 1986; Bertini and Giacomin, 1997), and a fourth-order SPDE model developed to predict the morphological evolution of epitaxial film growth (Vvedensky et al., 1993). SPDEs have additionally been incorporated into several optimization and control policies for thin film deposition processes. For example, Lou and Christofides (2004) adapted the Edwards-Wilkinson equation to perform feedback control on the surface roughness during thin film growth. Similarly, Ni and Christofides (2005a, 2005b) developed a methodology to construct linear SPDEs that were used to perform feedback control and multivariable predictive control on the film thickness and surface roughness of a thin film deposition process.

Despite the computational advantages of SPDEs, there still remain a number of challenges that impede the implementation of SPDE-based process improvement studies in thin film deposition systems. SPDE models are typically developed to capture the kinetic growth of the thin film surface, and thus they are incapable of predicting the macroscopic precursor gas behaviour. Hence, SPDEs need to be coupled with continuum transport models in order to capture the full multiscale nature of thin film growth. To the authors' knowledge, however, SPDEs have not been previously incorporated into multiscale thin film models. Another key issue is that it can be difficult to determine SPDE models that can adequately capture the microscopic surface growth subject to changes in the system parameters (Ni and Christofides, 2005b). This is most readily overcome through the determination of models that can predict the values of key SPDE model coefficients as a function of relevant process parameters. These models can either be constructed directly from microscopic process rules (Vvedensky et al., 1993), or by using data from kMC-driven thin film simulations (Ni and Christofides, 2005a). Another viable approach is to develop lookup tables of known SPDE coefficient values that can be combined with interpolation techniques to predict the SPDE coefficients at specific operating conditions. However, the relationship between the SPDE coefficients and the relevant system parameters can be nonlinear and subject to substantial noise, and it can therefore be difficult to predict the coefficient values through interpolation or low-order modelling approaches.

Artificial neural networks (ANNs) are an alternative approach that can be used to correlate the SPDE coefficients to changes in the multiscale system, as they are known to be able to determine relationships and predict data in highly nonlinear and noisy systems (Lapedes and Farber, 1987; Venkatasubramanian and Vaidyanathan, 1992). ANNs

ACCEPTED MANUSCRIPT

utilize robust mathematical models referred to as neurons in order to simulate the pattern-learning capabilities of the biological neural networks in a human brain (Dayhoff and DeLeo, 2001; Yegnanarayana, 2009). The neurons are assembled into an interconnected network of models that approximate the relationship between the network inputs (i.e. the process parameters) and its outputs (i.e. the SPDE coefficients). The network is trained using provided input-output data (i.e. specific process parameters and their corresponding SPDE coefficients), which is used to update the weights and biases of each neuron model to improve the overall network performance (Fine, 2006; Svozil et al., 1997). These ANNs, when coupled with an SPDE, form a hybrid model in order to capture the thin film surface growth subject to changes in the process parameters. This notion of hybrid modelling is attractive when simulating complex systems that cannot be readily described using mechanistic knowledge-based modelling approaches. The key idea in these models is to couple physically-relevant system knowledge with data-driven machine learning approaches such as ANNs to capture nonlinear, noisy, and otherwise convoluted process relationships (Venkatasubramanian, 2009; Oliveira, 2004). Accordingly, hybrid models have been developed for a number of complex applications, such as pharmaceutical milling (Akkisetty et al., 2010), bioreactor systems (Oliveira, 2004), and fuel additive design (Sundaram et al., 2001; Achenie et al., 2002). In principle, hybrid models provide a means to accurately and efficiently simulate the surface growth of thin film deposition processes. However, to the author's knowledge, such an approach has yet to be implemented for that process.

Motivated by this, the objective of this work is to present a novel hybrid multiscale model that combines SPDEs, ANNs, and continuum PDEs to simulate the evolution of a thin film deposition process. This model utilizes continuum transport equations to describe the behaviour of the precursor gas phase, coupled with an SPDE to capture the microscale thin film surface growth. The model additionally contains a feedforward ANN to predict key SPDE coefficients as a function of the substrate temperature and the surface mole precursor fraction; this hybrid model structure was implemented due to the noisy and nonlinear relationship between the SPDE coefficients and the relevant system parameters, which cannot be readily captured using mechanistic models. The proposed hybrid multiscale model is demonstrated to be able to simulate the thin film deposition process with similar accuracy to traditional kMC/PDE multiscale modelling techniques, and at a fraction of the computational cost. Furthermore, the hybrid multiscale model features are illustrated by implementing the model into a series of process improvement studies that seek to improve the quality of a deposited thin film.

This paper is organized as follows. Section 2 describes the continuum PDEs, the SPDE, and the ANN implemented in this work and the coupling scheme used to form a hybrid multiscale model that captures a thin film deposition process. In Section 3, the fully-assembled hybrid multiscale model is compared against a kMC-based thin film multiscale model in order to validate the model performance and demonstrate its computational superiority. Section 4 subsequently details two separate case studies in which the hybrid multiscale model is subjected to different optimization and control applications. As described in Section 4.1, the model is implemented into a bi-objective optimization formulation that seeks to determine the optimal temperature profile which will maximize the film's growth rate while simultaneously minimizing the final film roughness. In the case study described in Section 4.2, the hybrid multiscale model is coupled with a PI controller to control the surface roughness of the thin film deposition process. Concluding remarks are provided at the end.

ACCEPTED MANUSCRIPT

2. Hybrid multiscale model development

This work considers the epitaxial thin film growth of a gaseous precursor species onto a 1-dimensional substrate within a deposition chamber. The precursor species flow into the deposition chamber and form a boundary layer along the surface of the substrate. It is assumed for simplicity that the gaseous precursor species are evenly distributed along the axis parallel to the substrate surface, such that the precursor species concentration remains uniform across the entire substrate. As shown in Fig. 1, molecules within the boundary layer can adsorb onto the substrate to form the thin film. Adsorbed molecules can subsequently desorb back into the gas phase or migrate to another site on the film surface. This thin film deposition process is modelled using a multiscale approach that subdivides these phenomena into the macroscopic gas phase domain, which describes the macroscopic transport of the gas-phase precursor species to the thin film surface, and the microscopic thin film surface domain, which describes the kinetic events that control the evolution and growth of the deposited thin film. The subsequent sections will describe each of these domains and the modelling techniques used to simulate them.

2.1 Gas phase model

The gas phase domain of the thin film deposition process can be described using continuum transport equations as follows (Lam and Vlachos, 2001):

Momentum Transport: $$\frac{\partial}{\partial \tau}\left(\frac{\partial f}{\partial \eta}\right) = \frac{\partial^3 f}{\partial \eta^3} + f\frac{\partial^2 f}{\partial \eta^2} + \frac{1}{2}\left[\frac{\rho_b}{\rho} - \left(\frac{\partial f}{\partial \eta}\right)^2\right], \quad (1)$$

Heat Transport: $$\frac{\partial T}{\partial \tau} = \frac{1}{Pr}\frac{\partial^2 T}{\partial \eta^2} + f\frac{\partial T}{\partial \eta}, \quad (2)$$

Mass Transport: $$\frac{\partial y}{\partial \tau} = \frac{1}{Sc}\frac{\partial^2 y}{\partial \eta^2} + f\frac{\partial y}{\partial \eta}. \quad (3)$$

The boundary conditions for these equations are:

$$\frac{\partial f}{\partial \eta} = 1, T = T_{bulk}, y = y_{bulk}, \qquad \text{at } \eta \to \infty, \quad (4)$$

$$f = 0, \frac{\partial f}{\partial \eta} = 0, T = T_s, \qquad \text{at } \eta \to 0, \quad (5)$$

$$\frac{\partial y}{\partial \eta} = \frac{Sc(R_a - R_d)}{\sqrt{2a\mu_{bulk}\rho_{bulk}}}, \qquad \text{at } \eta \to 0, \quad (6)$$

where η and $\tau = 2at$ denote the dimensionless distance to the substrate surface and the dimensionless time respectively; f denotes the dimensionless stream function; ρ is the gas density; T is the local temperature; Pr is the Prandtl number; y is the mole fraction of the precursor gas; Sc is the precursor species Schmidt number; and a denotes the hydrodynamic strain rate. Similarly, ρ_{bulk}, T_{bulk}, y_{bulk}, and μ_{bulk} denote the density, temperature, mole

precursor fraction, and viscosity in the bulk phase; whereas T_s denotes the temperature of the substrate and y_s denotes the mole precursor fraction at the film surface, as depicted in Fig. 1. Furthermore, the terms R_a and R_d represent the rates of adsorption and desorption on the thin film surface, respectively, and therefore the surface boundary condition in Eq. (6) defines the net flux in y as the difference between the rates of adsorption and desorption (Lam and Vlachos, 2001). Note that R_a and R_d are dependent on the behaviour of the thin film surface; as a result, the mass transport equation in Eq. (3) cannot be solved without additional knowledge from the film surface model.

2.2 Thin film surface model

The thin film surface model details the morphological evolution of the deposited epitaxial film subject to the kinetic events (adsorption, desorption, and migration) that control the thin film growth. Simulating the kinetic events across an entire substrate is computationally prohibitive due to the large number of surface sites. However, due to the uniform boundary layer assumption described at the beginning of Section 2, the thin film growth across an entire substrate can be approximated using a small patch with a limited number of surface sites. This substrate patch can be located anywhere along the surface of the wafer, and its properties can be used to approximate the thin film properties over the entire substrate. Note that if a uniform precursor species boundary layer cannot be assumed across the substrate surface, then spatial coarse-graining techniques such as the gap-tooth model can be alternatively used to reduce the computational cost of simulating the spatially-varying film growth across the full substrate (Chaffart et al., 2016; Gear et al., 2003).

The microscopic evolution of the film growth on a 1-dimensional substrate patch can be described using the following SPDE (Vvedensky et al., 1993):

$$\frac{\partial h}{\partial t} = \nu \frac{\partial^2 h}{\partial x^2} + \lambda \left(\frac{\partial h}{\partial x}\right)^2 + K \frac{\partial^4 h}{\partial x^4} + \sigma \frac{\partial^2}{\partial x^2}\left(\frac{\partial h}{\partial x}\right)^2 + \eta. \tag{7}$$

In this equation, $h(x,t)$ denotes the height of the epitaxial film at time t and a position x along the surface of the substrate patch, where x is defined such that every integer of x corresponds to a site on the substrate, and every discretized point in x is a whole number. In addition, the terms ν, λ, K, and σ in Eq. (7) are SPDE coefficients that are problem-specific; whereas $\eta(x,t)$ is a random Gaussian variable with mean F and a shot noise covariance matrix:

$$\langle \eta(x,t)\eta(x',t')\rangle = V\delta(x-x')\delta(t-t'), \tag{8}$$

where $\delta(\cdot)$ is the dirac delta function. The SPDE in Eq. (7) has the following initial condition and periodic boundary conditions:

$$h(0,t) = h(L,t), \tag{9}$$

$$\frac{\partial h(0,t)}{\partial x} = \frac{\partial h(L,t)}{\partial x}, \tag{10}$$

$h(x, 0) = 0,$ (11)

where L denotes the length of the substrate patch in terms of the number of surface sites. Note that the SPDE was limited to one spatial dimension for simplicity; however, it can be extended to capture thin film growth on higher dimensional substrates (Vvedensky et al., 1993).

In Eq. (7), the stochastic variable η depicts the net deposition flux of the thin film, as dictated by the rates of adsorption and desorption. In addition, the PDE terms controlled by ν and λ are generally associated with regulating the desorption of deposited thin film molecules, whereas the terms controlled by K and σ are generally associated with regulating the surface migration events. Note that the surface desorption and migration behaviour is predominantly captured by the PDE terms associated with ν and K, respectively. Although the PDE terms proportional to λ and σ also describe the desorption and migration surface events, their contributions to the overall film surface evolution are minor compared to the terms proportional to ν and K, as illustrated in the derivation of the SPDE from the kinetic surface events (Vvedensky et al., 1993). The coefficients ν, λ, K, σ, F, and V in Eqs. (7)-(8) are model parameters that enable the SPDE to capture the evolution of the film surface. Consequently, the values of these SPDE coefficients are affected by process parameters that influence the surface growth, such as the substrate temperature, T_s, and the precursor mole fraction at the film surface, y_s. Note that for process optimization and control, T_s is required to change in time in order to improve the process performance. In addition, y_s is a multiscale parameter that varies in time depending on the mass transport behaviour within the system. Consequently, it is necessary to adjust the SPDE coefficients subject to the transient changes in these process conditions. The SPDE coefficients can be determined for given sets of T_s and y_s by fitting the SPDE model to data obtained from kMC simulations. This data can be used to develop low-order models to predict the coefficient values over a range of key process parameters (Ni and Christofides, 2005a). However, low-order models can be difficult to develop, as the relationship between the SPDE coefficients and the system parameters can be noisy and highly nonlinear. In this work, the process coefficient data can be used to train a supervised feedforward ANN to predict the SPDE coefficients for a set of process parameter values. Thus, the microscopic evolution of the thin film surface can be captured by a hybrid model in which an ANN is coupled with the aforementioned SPDE. The following two sections describe in detail the ANN approach implemented in this work. The next section provides a brief overview of the feedforward ANN architecture used in this work, whereas the following section details the ANN training and implementation.

2.2.1 Feedforward artificial neural network overview

The ANN algorithm implemented in this work utilizes separate, independent sub-networks to predict each SPDE coefficient individually as a function of the relevant system parameters. In each of these sub-networks, the neurons are subdivided into a two-layer structure consisting of a single hidden layer and an output layer, as illustrated in Fig. 2. Note that in this figure, ξ_i denotes the linear weighted function that categorizes the relationship between the ith neuron in each layer and its jth inputs, whereas f denotes the transfer function. The hidden layer transfer function in

ACCEPTED MANUSCRIPT

each of the sub-networks consists of a tan-sigmoid function that maps ξ_i onto a nonlinear curve, whereas a linear transfer function was selected for the output layer to map ξ_i onto the range of the output parameters. This ANN methodology has been successfully implemented for parameter approximation applications in previous studies (Bashir and El-Hawary, 2009; Ozkaya et al., 2007); further details concerning the general ANN structure can be found elsewhere (Svozil et al., 1997).

The values of the weight and bias terms $\omega_{i,j}$ and b_i in ξ_i (see Fig. 2) for each sub-network are initially randomly generated from a uniform distribution. These values are subsequently adjusted through backpropagation using sets of input data with known output parameter values. The sub-networks implemented in this work use the Levenberg-Marquardt algorithm to adjust $\omega_{i,j}$ and b_i and thus improve the network performance; further details can be found in (Hagan and Menhaj, 1994). In order to avoid overfitting and achieve good generalization in each sub-network, the implemented ANN training algorithm utilizes an early stopping approach that subdivides the input-output data into three sets: the training set, validation set, and testing set. The training set is used to improve the network fitting through backpropagation. After each epoch of training, the sub-network is presented with the validation set input data and tasked with predicting the corresponding outputs. The network-predicted outputs are compared with the known validation outputs in order to calculate the validation error. The sub-network is subsequently re-trained for another epoch using the training set. This cycle is repeated until it has been determined that further training will result in a loss of generalization in the sub-network, i.e. once the validation error fails to decrease over six consecutive epochs (Bashir and El-Hawary, 2009). After the training is completed, the ANN framework is assessed by passing the testing data through the network. If the errors in the testing or validation steps are insufficiently large, then the input-output data is redistributed between the three sets and the network is re-trained until the validation and testing errors are all below a user-defined criterion.

2.2.2 Artificial neural network implementation

The feedforward ANN algorithm described above was used to develop a neural network to predict the SPDE parameters ν, K, F, and V as a function of two primary system parameters that significantly affect the thin film growth: the surface mole precursor fraction, y_s, which varies due to the multiscale nature of the system; and the surface temperature, T_s, which is used to control the thin film growth process. Note that the ANN was not developed to predict the remaining two SPDE parameters λ and σ as these were observed from open-loop simulation data to not be significantly affected by T_s and y_s. These parameters were therefore held constant and their values (listed in Table 1) were determined by least squares fitting.

In order to improve the performance of the ANN, physical insights about the thin film deposition system and behaviour of the SPDE coefficients were considered when designing and implementing the network into the hybrid surface model. For example, the behaviour of ν, K, F, and V with respect to T_s and y_s was observed to differ significantly between each SPDE coefficient. In order to simplify the network training and improve upon the ANN performance, an independent sub-network was developed in order to predict each SPDE coefficient individually, as

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

illustrated in Fig. 3. An additional observation was that the behaviour of K with respect to T_s and y_s is highly nonlinear and difficult to predict, as opposed to the other SPDE parameters that are comparatively well-behaved. Therefore, the sub-networks used to predict K were designed to contain 15 neurons within their hidden layers in order to adequately capture this parameter's behaviour, as shown in Fig. 3. In contrast, the sub-networks for ν, F, and V were implemented with 10 neurons within their hidden layers.

Another key insight was that for a system at time $t = t_m$, temperature $T_s(t_m)$, and surface mole fraction $y_s(t_m)$, it was observed that the SPDE coefficients ν and K can vary significantly depending on whether the surface temperature was increased from a lower temperature $\left(T_s(t_m) > T_s(t_{m-1})\right)$ or decreased from a higher temperature $\left(T_s(t_m) < T_s(t_{m-1})\right)$. Specifically, it was observed that the values of these coefficients for a set T_s and y_s differ from the expected results if the temperature is increased and the change in temperature is sufficiently large, or if the time interval between temperature changes $(t_m - t_{m-1})$ is sufficiently small such that the transient profiles of the kinetic surface events have not converged to a constant value before the temperature is raised. This phenomenon is due to the effects of a change in temperature on the ratio of the kinetic events taking place. When the temperature is raised to $T_s(t_m)$ from a lower temperature, it results in a higher number of surface migration and desorption events taking place than if the temperature was lowered to the same temperature $T_s(t_m)$ from a higher temperature. This difference is more pronounced under large temperature increases and if the transient profiles of the kinetic events have not converged on a constant value before the temperature was raised. Consequently, the values of the coefficients ν and K must be increased under these conditions in order to account for the higher percentages of desorption and migration events taking place. For the present study, it was observed that sufficiently good coefficient predictions could be achieved by developing and training two unique pairs of sub-networks to predict ν and K as a function of only T_s and y_s subject to the different temperature change conditions, as illustrated in Fig. 3. The resulting ANN design was tested in numerous validation studies as detailed in Section 3, and it was observed to perform well under the range of operating conditions considered in this work.

The selection of appropriate process data is another critical factor in training the ANN to accurately predict the SPDE coefficients (Li et al., 2018). Accordingly, a cumulative total of 165 values for each SPDE coefficient were determined over a range of process parameters in order to train the sub-networks for the proposed hybrid multiscale model. These coefficient samples were determined over mole equidistant precursor fraction values between 1×10^{-6} and 3×10^{-6} and over increasing and decreasing temperature values evenly-spaced between 900 K and 1,300 K. These ranges cover the expected operating region for this process; in addition, the coefficient sample points were selected to prevent any unexplored regions within the operating search space. These SPDE coefficient values were determined by fitting the SPDE model to kMC data derived at each combination of T_s and y_s using least squares fitting. The kMC data was generated by simulating the thin film surface growth for a fixed T_s and y_s using a kMC model that has been previously reported (Rasoulian and Ricardez-Sandoval, 2014). The SPDE coefficients were subsequently determined by finding the coefficient values that minimized the sum of squared errors between key observables in the kMC data (i.e. the transient profiles of the thin film roughness and growth rate) and those generated by the SPDE. After the SPDE coefficients had been determined for each combination of T_s and y_s, the

coefficient data was randomly subdivided between the training, validation, and testing sets such that the training set contained 70% of the data, while the validation and testing sets each contained 15% of the data. These subsets were subsequently used to train and validate the ANN using the early stopping method described in Section 2.2.1. Note that the ANN would ideally be trained using larger datasets in order to ensure good network generalization; however, the number of sample points was practically limited by the computational cost required to determine the SPDE coefficients for a given set of process conditions. In order to ensure that the trained ANN and its sub-networks were not overfitted and that they had achieved good generalization, the fully-assembled hybrid multiscale model was subjected to numerous validation studies, as detailed in Section 3.

2.3 Thin Film Characteristics

The thin film deposition process is characterized by a number of parameters that are critical to achieving high quality films. Characteristics such as the film thickness and the surface roughness are often considered good performance indicators of the film quality, while the film growth rate provides a metric of the deposition productivity (Rasoulian and Ricardez-Sandoval, 2014). Note that due to the uniform boundary layer assumption, it can be assumed that the thin film characteristics are uniform across the entire substrate, and thus these metrics can be interpreted as wafer-averaged properties. The film thickness at time $t = t_m$ is defined as the mean film height over the substrate, and it can be calculated based off of the SPDE results as follows:

$$H(t_m) = \frac{\sum_{n=1}^{N} h(x_n, t_m)}{L}, \tag{12}$$

where $x = x_n$ denotes the nth discretized spatial point along the surface of the 1D substrate patch and N denotes the total number of spatial discretization points along the substrate patch. The film roughness can be characterized by the mean number of broken bonds at time $t = t_m$, as follows:

$$r(t_m) = 1 + \frac{\sum_{n=1}^{N} |h(x_{n+1}, t_m) - h(x_n, t_m)| + |h(x_{n-1}, t_m) - h(x_n, t_m)|}{2L}, \tag{13}$$

where the number of broken bonds is calculated by summing the difference in height between each surface site and their neighbouring sites, and dividing by 2 to prevent double-counting. Furthermore, the film growth rate at time $t = t_m$ can be calculated by calculating the slope of the film thickness over a time interval Δt, as follows:

$$Gr(t_m) = \frac{H(t_m) - H(t_{m-1})}{\Delta t}, \tag{14}$$

where $t_m - t_{m-1} = \Delta t$. Under constant operating conditions, the film thickness changes linearly in time, and therefore Eq. (14) is able to accurately capture the film growth rate. Note that Eqs. (12) and (13) above have been defined to capture the film thickness and the surface roughness of a thin film grown on a 1D substrate; however, these equations can be readily expanded to capture important film characteristics for higher dimensional substrates.

2.4 Multiscale Model Coupling and Assembly

In the hybrid multiscale model, neither the macroscopic gas phase model nor the microscopic thin film surface model can be solved independently, as the processes that occur in each domain are influenced by the events that occur in the other domain. In the macroscopic gas phase domain, the surface boundary condition in Eq. (6) depends on the difference in the rates of adsorption and desorption, $R_a - R_d$, which can be calculated from the thin film growth rate over a time interval Δt_{MS} as follows:

$$R_a(t) - R_d(t) = \frac{Gr(t)}{2a} = \frac{H(t) - H(t - \Delta t_{MS})}{2a\Delta t_{MS}}, \tag{15}$$

where Δt_{MS} denotes the multiscale coupling time between the gas phase and film surface models. Note that the growth rate is a function of the film thickness, as depicted in Eq. (14), and consequently it depends on the behaviour of the thin film surface model. On the microscale surface domain, the thin film deposition process is significantly affected by the surface mole precursor fraction, y_s, which is determined by solving the mass transport equation as shown in Eq. (3). Therefore, in order to accurately capture the film growth using the hybrid multiscale model, the ANN was trained as described in Section 2.2.2 to predict the SPDE coefficient values as a function of y_s, in addition to the controlled operational parameter T_s.

The fully-assembled hybrid multiscale model combines the macroscopic continuum PDEs, the microscopic SPDE model, and the coefficient-predicting ANN in order to adequately simulate the transient thin film deposition process, as illustrated in Fig. 4. Each of these models are coupled such that the relevant parameter information can be smoothly passed from one model to another. In the macroscopic gas phase, the continuum PDEs are solved in order to determine the transient behaviour of the precursor gas species within the deposition chamber over the multiscale coupling time interval Δt_{MS}. Subsequently, the surface mole precursor fraction obtained from the solution of the PDE model, y_s, is passed to the ANN, along with the substrate temperature T_s, in order to determine the model coefficient values that enable the SPDE to adequately describe the thin film growth subject to those surface conditions. The coefficients are then passed to the SPDE model, which simulates the evolution of the thin film surface for Δt_{MS}. These results are then used to calculate the difference in the rates of adsorption and desorption, $R_a - R_d$, based on the film growth rate according to Eq. (15). As shown in Fig. 4, the value of $R_a - R_d$ is passed back to the continuum PDEs, where it is used to update the surface boundary condition of the mass transport equation. The gas phase PDEs are subsequently re-evaluated for another interval Δt_{MS}. This cycle is repeated until the final thin film deposition time t_f has been reached. Note that the multiscale coupling time interval, Δt_{MS}, can be increased in order to reduce the computational cost of the hybrid multiscale model; however, sufficiently large values of Δt_{MS} can result in instabilities in the multiscale model coupling in addition to loss of information in the thin film properties. Consequently, the value of Δt_{MS} used in this work was determined *a priori* from trial-and-error simulations in order to best address this trade-off. The parameter values for the deposition process implemented in

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

this work are listed in Table 1. The kinetic models and their parameters used to determine the SPDE coefficients are not shown here for brevity but they can be found elsewhere (Rasoulian and Ricardez-Sandoval, 2014).

3. Validation of the Hybrid Multiscale Model

The objective of this section is to test and validate the performance of the hybrid multiscale model described in Section 2. The model performance was evaluated with respect to a continuum modelling/kMC-based multiscale thin film model that has been previously reported (Rasoulian and Ricardez-Sandoval, 2014). For this comparison, both the hybrid multiscale model and the kMC-based multiscale model were used to simulate the behaviour of an epitaxial deposition process subject to transient changes in the relevant process parameters. Note that the accuracy of the hybrid multiscale model is dependent on the predictive capabilities of the trained ANN; therefore, the hybrid multiscale model performance was validated subject to changes in the network input parameters, i.e. the substrate temperature T_s and the surface mole precursor fraction y_s. These parameters were allowed to vary every 200 s in this validation study, for a total thin film deposition time of 2,400 s. The substrate temperature is a controlled operational parameter that can be readily adjusted online. On the other hand, y_s is a system parameter determined by the gas phase mass transport equation and thus it cannot be controlled directly. However, y_s is affected by the state of the thin film surface due to the multiscale nature of the system, and hence it is influenced by other process parameters that affect the film deposition, such as T_s, the bulk mole precursor concentration y_{bulk}, and the precursor gas density in the bulk ρ_{bulk}. Therefore, in the context of this validation study, y_s was allowed to vary according to the changes made by the multiscale system. Another key consideration is that in order to ensure good generalization in the hybrid multiscale model results, it is necessary to validate the model using values of T_s and y_s that have not been previously considered during the network training. To comply with this requirement, the temperatures used to validate the model performance were randomly generated from a uniform distribution bounded by 900 K $\leq T_s \leq$ 1,300 K, as shown in Fig. 5a. These temperatures were subsequently inspected to ensure they differed sufficiently from those used to train the ANN and its sub-networks in Section 2.2.2. In addition, the surface mole precursor fraction was allowed to evolve based on the changes introduced in the multiscale model that impact the mass transport of the gas precursor species; the values of y_s were inspected *a posteriori* to verify that they were sufficiently different from the values used during training, as illustrated in Fig. 5b. Note that the temperature profile illustrated in Fig. 5a was generated solely with the intent of validating the hybrid multiscale model and to show that it is able to capture the actual phenomena at a wide range of operating conditions; i.e. it is not intended to convey an actual temperature profile that could be implemented in practice.

The performance of both the hybrid multiscale model and the kMC-based multiscale model were assessed through the evolution of the thin film roughness and film growth rate over the course of the deposition process, as illustrated in Figs. 5c and 5d respectively. In addition, Table 2 displays the sum of squared errors between the results of the hybrid multiscale model and the kMC-based multiscale model for each interval of different T_s and y_s values. These results show that the hybrid multiscale model is able to adequately capture the transient thin film behaviour, and that

the ANN in the hybrid model has good generalization and is not subject to overfitting. As shown in Fig. 5, the roughness and growth rate profiles determined using the hybrid multiscale model match those determined using the kMC-based multiscale model. Furthermore, Table 2 shows that the errors in the hybrid multiscale model performance remain sufficiently low (i.e. < 0.1 for the roughness profiles, and < 1 for most of the film growth rate profiles). Note that the sum of squared errors in the growth rate profiles are notably larger than the errors of the roughness profiles; this is due to the short-term fluctuations in the growth rate that occur following changes in T_s and y_s. In both multiscale models, these fluctuations arise due to the coupling between the film growth rate and the surface mole precursor fraction in the microscopic and macroscopic domains, as both parameters will continuously change and affect each other following a disturbance until they reach a new steady state, as shown in Figs. 5b and 5d. The magnitude of the fluctuations can vary significantly between the two multiscale modelling approaches, resulting in large sum of squared errors in the hybrid multiscale model growth rate. However, these growth rate errors decrease significantly as the fluctuations decay and the growth rate approaches steady state, as illustrated visually in Fig. 5d. With regards to the computational costs, the kMC-based multiscale model and the hybrid multiscale model required 3,540 s and 78.4 s of CPU time, respectively, in order to simulate the thin film growth process (3.4GHz Intel i7-4770 processor). This shows that the hybrid multiscale model is approximately 45 times faster than the kMC-based multiscale model, requiring only 2.2% of the computational time that would be required by the kMC-based multiscale model to run a single simulation. This demonstrates the computational efficiency of the hybrid multiscale model over the traditionally-used kMC-based multiscale model. Overall, this comparison study illustrates that the hybrid multiscale model is capable of simulating the thin film deposition process with sufficient accuracy, and at significantly lower computational costs. Note that in order to further explore the performance of the hybrid multiscale model, additional validation tests were performed under different operating conditions with similar performance results to the validation study showcased above; these studies are not shown here for brevity.

The results in Figs. 5c and 5d additionally illustrate the response of the surface roughness and the film growth rate to changes in the substrate temperature. According to the figures, the roughness and the growth rate both decrease when T_s is raised and increase when T_s is lowered over the temperature ranges considered in this work. This behaviour is due to the effects of the temperature on the ratio of the kinetic events taking place during thin film deposition. Increases in the surface temperature result in a higher percentage of migration and desorption events taking place because of their Arrhenius relationship with T_s (Lam and Vlachos, 2001). Consequently, the film surface roughness diminishes due to the increase in the number of migration events taking place, whereas the higher percentage of desorption events results in a decrease in the overall film growth rate. Note that this behaviour has been previously described in the literature (Aarik et al., 1999; Ott et al., 1997; Lam and Vlachos, 2001; Rasoulian and Ricardez-Sandoval, 2015b).

4. Optimization and Control Studies

ACCEPTED MANUSCRIPT

The aim of this section is to present a series of process improvement studies that demonstrate the computational advantages of the hybrid multiscale thin film growth model in an optimization or control setting. In the first case study, the hybrid multiscale model is implemented into a bi-objective optimization scheme that seeks the optimal compromise between two conflicting objectives in the thin film deposition process. In the second study, the hybrid multiscale model is coupled with a PI controller to control the thin film roughness and achieve targeted set-points.

4.1 Dynamic Optimization

The manufacturing of microelectronic and semiconductor devices requires the deposition of thin films that meet or exceed a targeted thickness H_D with minimal surface roughness, i.e. $\min R(t_f)$, where t_f is the thin film deposition time. In addition, it is desirable to optimize process productivity by maximizing the overall growth rate, i.e. $\max \int_0^{t_f} Gr(t)dt$. However, there exists a trade-off between these optimization objectives, as high temperatures are required to minimize roughness whereas low temperatures are required to maximize the growth rate (Rasoulian and Ricardez-Sandoval, 2014). This trade-off can be addressed with a 1-norm bi-objective optimization scheme, which seeks to find the point within the feasible search space that is closest to the infeasible point that satisfies both conflicting objectives, i.e. the utopia point (Grossmann et al., 1982). Motivated by this, the objective of this section is to use the hybrid multiscale model to perform a bi-objective optimization in order to identify the optimal substrate temperature profile that simultaneously minimizes the surface roughness and maximizes the deposition rate of the epitaxial thin film. Note that the substrate temperature is the only controlled parameter considered within this optimization study; all other parameters are either held constant at their nominal values as listed in Table 1 or determined from the solution of the mass transport PDE, e.g. y_s. The bi-objective algorithm implemented in this work initially performed a pair of optimization studies, one with respect to each optimization objective, as shown below.

Minimization of Roughness (*minR*):

$$\min_{T_s \in \mathbf{T_s}} R(t_f), \quad (16)$$

subject to:

Hybrid Multiscale Model (PDE-ANN-SPDE Model), Eqs. (1)-(15) (17)

$$900\text{ K} \leq \mathbf{T_s} \leq 1300\text{ K} \quad (18)$$

$$\mathbf{T_s} = [T_s(\Delta t_c), \dots, T_s(N_c \Delta t_c)] \quad (19)$$

$$T_s(n_c \Delta t_c) - T_s\big((n_c - 1)\Delta t_c\big) \leq 50\text{ K} \quad (20)$$

$$N_c \Delta t_c = t_f \quad (21)$$

$$H(t_f) \geq H_D \quad (22)$$

ACCEPTED MANUSCRIPT

Maximization of Growth Rate (*maxGr*):

$$\max_{T_s \in \mathbf{T_s}} Gr_\Sigma\,(t_f), \tag{23}$$

subject to:

Constraints, Eq. (17)-(22).

In the above expressions, Δt_c denotes the temperature change interval, N_c denotes the number of temperature changes considered in the optimization problem, and $Gr_\Sigma(t_f) = \int_0^{t_f} Gr(t)dt$ denotes the overall growth rate over the time interval $t = [0, t_f]$. For each optimization study, the system was initialized at a constant temperature $T_s(0) =$ 1,000 K, and the substrate temperature was not allowed to vary by more than 50 K over a single temperature change to avoid any infeasibly-large temperature changes in the system. Furthermore, the total deposition time was set to $t_f =$ 1,560 s, the targeted minimum film thickness was set to $H_D =$ 2,000 layers, and the temperature was allowed to change every 120 s, such that $N_c = 13$.

The results from problems *minR* and *maxGr* are shown in Table 3 and Fig. 6, whereas the optimal temperature profiles are illustrated in Fig. 7. These results show that the surface roughness can be minimized by step-increasing the temperature to 1,300 K (i.e. the maximum allowed temperature) whereas the growth rate can be maximized by step-decreasing the temperature to its lower bound of 900 K, as would be expected. These results motivate the development of the multi-objective optimization problem (*Multi-obj*) that aims to find the optimal compromise between these two conflicting objectives, i.e. minimization of surface roughness and maximization of film growth rate. This can be accomplished by minimizing the 1-norm distance between the compromise point and the utopia point as follows:

$$\min_{T_s \in \mathbf{T_s}} \frac{R(t_f) - R^l}{R^u - R^l} + \left(1 - \frac{Gr_\Sigma(\mathrm{t_f}) - Gr_\Sigma^l}{Gr_\Sigma^u - Gr_\Sigma^l}\right), \tag{24}$$

subject to:

Constraints, Eq. (17)-(22).

The results from *minR* and *maxGr* presented in Table 3 and Fig. 6 provide upper and lower bounds on the roughness (R^u and R^l) and total growth rate (Gr_Σ^u and Gr_Σ^l), which can be used to define the utopia point (R^l, Gr_Σ^u) within the objective search space. Once the objective bounds have been defined, they can be used to determine the optimal compromise point within the feasible objective search space. The results of this optimization study (*Multi-obj*) are listed in Table 3, and the optimal temperature profile is illustrated in Fig. 7. In addition, Fig. 8 depicts the transient variation in the surface roughness and film growth rate generated using the optimal temperature profile. Fig. 8 also compares the roughness and growth rate profiles obtained using the hybrid multiscale model with those obtained by operating the kMC-based multiscale model under the same temperature profile, which further validates the hybrid multiscale model performance. Figs. 7 and 8 illustrate that the optimal trade-off was achieved by keeping the substrate temperature predominantly low throughout the film deposition in order to maximize the overall growth

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

rate, while increasing the system temperature toward the end of the deposition process to minimize the final surface roughness. At the beginning of the process, the system temperature was lowered to 900 K to maximize the initial growth rate. After 720 s, the temperature was slightly increased in order to marginally decrease the overall film roughness. This was observed to minimize the time required to smoothen the surface during the final deposition stages, which contributed toward a lower final roughness and a higher overall growth rate than would have been obtained otherwise. At 1080 s, the system temperature was gradually raised to 1,094 K in order to minimize the final roughness of the thin film surface.

Fig. 9 illustrates the Pareto front formed by the trade-off between the final surface roughness and the overall growth rate, which serves as the boundary of the feasible search space accessible by the thin film deposition process. This figure additionally depicts the locations of the optimal trade-off point and the utopia point within the objective space. The Pareto front points were obtained by solving the following optimization problem (Gembicki and Haimes, 1975):

$$\min_{T_s \in \mathbf{T_s}, \gamma} \gamma, \quad (25)$$

subject to:

$$R(t_f) - w\gamma \leq R^l \quad (26)$$

$$Gr_\Sigma(t_f) + (1-w)\gamma \geq Gr_\Sigma^u \quad (27)$$

Constraints, Eq. (17)-(22).

In the expression above, γ is a slack variable that is used to minimize $R(t_f)$ and maximize $Gr_\Sigma(t_f)$ simultaneously subject to a weight term w, whose value is adjusted from 0 to 1 in order to obtain each of the points in the Pareto front shown in Fig. 9. The optimization problem in Eq. (25) required on average a CPU time of 300 min for each point considered in the Pareto front, whereas the optimal trade-off point determined from Eq. (24) required 40 min of computational time. This highlights the computational advantage of the 1-norm bi-objective approach for solving optimization problems with conflicting objectives. Note that the higher computational demands of the Pareto front optimization problem in Eq. (25) is because this optimization problem requires significantly more iterations to converge compared to the bi-objective optimization problem (i.e. Eq. (24)). This is due to the parameter γ in Eq. (25), which serves as both the objective function and an additional optimization variable for the Pareto front optimization problem. Fig. 9 demonstrates that the temperature profile determined by solving Eq. (24) provides an optimal trade-off point that is sufficiently close to the utopia point. In particular, Table 3 indicates that the optimal trade-off final surface roughness is at 84.79% of the utopia final roughness, whereas the overall trade-off growth rate is observed to be at 92.54% of the overall utopia growth rate. These results demonstrate that the optimization results from Eq. (24) satisfy the trade-off between the final surface roughness and the overall growth rate for the considered thin film deposition process in an optimal fashion.

ACCEPTED MANUSCRIPT

4.2 Control of Thin Film Surface Roughness through Set-point Tracking

The deposition of high-quality thin films is dependent on the online manipulation of key process parameters to achieve and maintain specific film properties during the deposition process, such as the surface roughness (Lou and Christofides, 2003). These parameters affect the microkinetic events responsible for thin film deposition, and consequently they can be adjusted on-line in order to control the thin film growth on a molecular level and achieve targeted process constraints. In this section, the hybrid multiscale model is used to perform feedback control on the thin film surface roughness by manipulating the substrate temperature. In this approach, the hybrid model is adopted as an estimator of the system states whereas a previously-developed kMC-based multiscale model is used to simulate the plant (Rasoulian and Ricardez-Sandoval, 2014). The hybrid multiscale model is coupled with a proportional integral (PI) controller to adjust T_s and achieve the targeted film surface roughness, as illustrated in the control scheme diagram in Fig. 10. Note that in this figure, r_m denotes the measured surface roughness at any time *t*, which is estimated using the hybrid multiscale model.

This case study considers three different scenarios in order to demonstrate the performance of the proposed control scheme using the hybrid multiscale model. These scenarios used the control scheme shown in Fig. 10 to perform set-point tracking changes on the thin film deposition process subject to nominal conditions, in the presence of disturbance effects in the bulk mole precursor fraction (y_{bulk}), and subject to uncertainty in key model kinetic parameters, respectively. In all three scenarios, the PI controller used the roughness values provided by the hybrid multiscale model to adjust the surface temperature. The proportional gain of the controller was set to 1.7 K/mL whereas the integral gain was set to 7 K/(mL·s); these values were determined *a priori* based on results obtained from preliminary simulations.

The first scenario considers controlling a thin film deposition process subject to nominal conditions, i.e. there are no uncertainties or disturbances in the system. In this scenario, the hybrid multiscale model-based control scheme was implemented thrice to achieve three separate surface roughness values. The roughness set-points for these simulations were set at $r_{sp,1} = 1.5$ mL, $r_{sp,2} = 1.25$ mL, and $r_{sp,3} = 1.175$ mL, respectively. The surface temperature control actions are depicted in Fig. 11a, whereas Fig. 11b displays the controlled thin film roughness trajectories as determined by the estimator (i.e. the hybrid multiscale model) and the plant (i.e. the kMC-based multiscale model). This figure illustrates that the hybrid multiscale model coupled with the PI controller was able to regulate the film roughness and achieve the targeted set-points. In addition, the hybrid multiscale model-based control scheme required 12.3 s of computational time to predict the entire deposition process, whereas a control scheme using the kMC-based multiscale model would have required 1,313.6 s. These results further highlight the computational efficiency of the proposed hybrid multiscale model in a control setting.

The second scenario examines the behaviour of the proposed control scheme in the presence of disturbances in a key system parameter (e.g. the bulk mole precursor fraction, y_{bulk}). Note that the bulk mole precursor fraction y_{bulk} affects the surface mole precursor fraction y_s through the mass transport equation (Eq. (3)) and therefore it has a significant effect on the thin film growth. It was assumed in this scenario that the disturbances in y_{bulk} are

ACCEPTED MANUSCRIPT

measured, and thus the parameter's fluctuations can be tracked and taken into account by the hybrid multiscale model-based control scheme. The bulk mole precursor fraction was subjected to a disturbance every 60 s according to a normal distribution around its nominal value ($\bar{y}_{bulk} = 2 \times 10^{-6}$) with a standard deviation $\sigma(y_{bulk}) = 2.5 \times 10^{-7}$. Note that the same disturbance profile was implemented for both the plant model (kMC-based multiscale model) and the control scheme model (hybrid multiscale model). This control study was implemented three times using the three roughness set-points specified for the first scenario. The resulting temperature control actions and their corresponding roughness profiles are illustrated in Fig. 12. Fig. 12a additionally compares the temperature control actions subject to disturbances with the temperature control actions determined under nominal conditions. These results demonstrate that the proposed control scheme is able to regulate the thin film roughness subject to measured disturbances in y_{bulk}. Following each disturbance, the PI controller adjusted the surface temperature according to the hybrid multiscale model predictions in order to meet the targeted specifications, as illustrated by the temperature profiles in Fig. 12a. These results confirm that the hybrid multiscale model is able to adequately ensure that the desired film roughness has been achieved despite the presence of measured disturbances in the system.

In the third scenario, the hybrid multiscale model-based control scheme was used to perform set-point tracking changes on the thin film deposition process subject to parametric uncertainty. For this study, it was assumed that the activation energies of the kinetic thin film growth events are not known due to measurement and experimental errors (Evans and Ricardez-Sandoval, 2014). These include the activation energy of a single surface bond E, the activation energy of desorption E_d, and the activation energy of migration E_m; further details concerning the kinetic events and the activation energies can be found in (Rasoulian and Ricardez-Sandoval, 2014). The uncertainty in these parameters was assumed to be normally-distributed with mean values of $[\bar{E}, \bar{E}_d, \bar{E}_m] = [17, 17, 10.2]$ kcal/mol, i.e. their nominal values according to (Rasoulian and Ricardez-Sandoval, 2014), and with a standard deviation of 0.25 kcal/mol for each parameter. In order to demonstrate the effects of uncertainty on the thin film deposition system, Monte Carlo sampling was used to simulate the surface roughness control scheme subject to 12 unique uncertainty realizations. Fig. 13a displays the temperature control actions generated by the control scheme for the three roughness set-points considered in the previous scenarios. Similarly, Fig. 13b illustrates the resulting roughness profiles using each of the different uncertainty realizations. In this figure, each of the controller actions and roughness profiles generated by the hybrid multiscale model correspond to a particular realization in the uncertain parameters. Note that the uncertainty is unmeasured, and thus it cannot be taken into account by the hybrid multiscale model-based control scheme. Consequently, the controller actions and corresponding roughness profiles generated by the hybrid multiscale model correspond to a particular realization. The results in Fig. 13 demonstrate that, although the proposed control scheme is unable to precisely achieve the targeted roughness set-points in the plant model, the control scheme is able to maintain a level of control over the surface roughness in the presence of parametric uncertainty. This demonstrates the need to expand the control scheme and the hybrid multiscale model itself to be able to account for uncertainties and unknown disturbances in the process parameters; such approaches, however, were not explored within the context of this work.

ACCEPTED MANUSCRIPT

5. CONCLUSIONS

The primary contribution of this work was to construct a hybrid multiscale model that combines continuum (PDE) models and SPDEs with ANNs to efficiently simulate a thin film deposition process. This model was developed to operate at a fraction of the computational cost required by previously-implemented kMC-based thin film multiscale models while still capturing the film parameters with sufficient accuracy. The thin film surface was modelled using a fourth-order non-linear SPDE with terms that simulate each of the relevant kinetic events. In addition, an ANN was developed and trained to predict the values of the SPDE coefficients as a function of the surface temperature and the surface precursor gas fraction, such that the SPDE could capture the changes in the film growth subject to changes these system parameters. The SPDE-based multiscale model was observed to perform well against the previously-studied kMC-based multiscale model, as it was able to simulate transient variations in the thin film properties with sufficient accuracy yet at a fraction of the computational cost. The SPDE-based multiscale model was subsequently implemented into a 1-norm bi-objective optimization problem to determine the temperature profile that maximizes the film growth rate while minimizing the final film roughness. The results of this optimization showed that these conflicting objectives could be met by keeping the overall temperature profile low and by increasing the temperature at the end of the deposition process. Furthermore, the SPDE-based multiscale model was coupled with PI controllers in order to control the thin film surface roughness. The hybrid multiscale model-based control scheme was able to adequately regulate the surface roughness and achieve targeted set-points for thin film deposition processes subject to nominal conditions and measured disturbances. In addition, the hybrid multiscale model-based scheme was able to maintain a level of control over the surface roughness of thin film processes subject to parametric uncertainty. However, the control scheme was unable to ensure that the thin film surface roughness fell below the targeted roughness values, as would be required in the fabrication of high-quality thin films. This motivates the need to expand the hybrid multiscale model to be able to account for parametric uncertainty in future works. Overall, both case studies showcased how the hybrid multiscale model can be implemented to perform accurate and efficient optimization and control studies for thin film deposition systems.

ACKNOWLEDGMENTS

The authors of this paper would like to graciously acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC) for their financial support granted for this research.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

REFERENCES

Aarik, J., Aidla, A., Kiisler, A.-A., Uustare, T., Sammelselg, V., 1999. Influence of substrate temperature on atomic layer growth and properties of HfO_2 thin films. Thin Solid Films 340, 110–116.

Achenie, L., Venkatasubramanian, V., Gani, R., 2002. Computer aided molecular design: theory and practice. Elsevier, Amsterdam.

Adomaitis, R.A., 2010. Development of a multiscale model for an atomic layer deposition process. J. Cryst. Growth 312, 1449–1452.

Akkisetty, P.K., Lee, U., Reklaitis, G.V., Venkatasubramanian, V., 2010. Population balance model-based hybrid neural network for a pharmaceutical milling process. J. Pharm. Innov. 5, 161–168.

Bashir, Z.A., El-Hawary, M.E., 2009. Applying wavelets to short-term load forecasting using PSO-based neural networks. IEEE Trans. Power Syst. 24, 20–27.

Baumann, F.H., Chopp, D.L., Rubia, T.D. de la, Gilmer, G.H., Greene, J.E., Huang, H., Kodambaka, S., O'Sullivan, P., Petrov, I., 2001. Multiscale modeling of thin-film deposition: applications to Si device processing. MRS Bull. 26, 182–189.

Bertini, L., Giacomin, G., 1997. Stochastic burgers and KPZ equations from particle systems. Commun. Math. Phys. 183, 571–607.

Chaffart, D., Rasoulian, S., Ricardez-Sandoval, L.A., 2016. Distributional uncertainty analysis and robust optimization in spatially heterogeneous multiscale process systems. AIChE J. 62, 2374–2390.

Chaffart, D., Ricardez-Sandoval, L.A., 2018. Robust dynamic optimization in heterogeneous multiscale catalytic flow reactors using polynomial chaos expansion. J. Process Control 96, 113–131.

Christofides, P.D., Armaou, A., 2006. Control and optimization of multiscale process systems. Comput. Chem. Eng. 30, 1670–1686.

Christofides, P.D., Armaou, A., Lou, Y., Varshney, A., 2009. Control and optimization of multiscale process systems. Birkhäuser, Boston, Massachusetts.

Crose, M., Kwon, J.S.-I., Tran, A., Christofides, P.D., 2017. Multiscale modeling and run-to-run control of PECVD of thin film solar cells. Renew. Energy, Special Issue: Control and Optimization of Renewable Energy Systems 100, 129–140.

Crose, M., Sang-Il Kwon, J., Nayhouse, M., Ni, D., Christofides, P.D., 2015. Multiscale modeling and operation of PECVD of thin film solar cells. Chem. Eng. Sci., Control and Optimization of Smart Plant Operations 136, 50–61.

Dayhoff, J.E., DeLeo, J.M., 2001. Artificial neural networks. Cancer 91, 1615–1635.

Dwivedi, V., Adomaitis, R.A., 2009. Multiscale simulation and optimization of an atomic layer deposition process in a nanoporous material. ECS Trans. 25, 115–122.

Edwards, S.F., S, F.R., Wilkinson, D.R., 1982. The surface statistics of a granular aggregate. Proc R Soc Lond A 381, 17–31.

Evans, R.D., Ricardez-Sandoval, L.A., 2014. Multi-scenario modelling of uncertainty in stochastic chemical systems. J. Comput. Phys. 273, 374–392.

Fine, T.L., 2006. Feedforward neural network methodology. Springer-Verlag New York, Inc., New York.

Gear, C.W., Li, J., Kevrekidis, I.G., 2003. The gap-tooth method in particle simulations. Phys. Lett. A 316, 190–195.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Gembicki, F., Haimes, Y., 1975. Approach to performance and sensitivity multiobjective optimization: The goal attainment method. IEEE Trans. Autom. Control 20, 769–771.
Grossmann, I.E., Drabbant, R., Jain, R.K., 1982. Incorporating Toxicology in the Synthesis of Industrial Chemical Complexes. Chem. Eng. Commun. 17, 151–170.
Hagan, M.T., Menhaj, M.B., 1994. Training feedforward networks with the Marquardt algorithm. IEEE Trans. Neural Netw. 5, 989–993.
Kardar, M., Parisi, G., Zhang, Y.-C., 1986. Dynamic scaling of growing interfaces. Phys. Rev. Lett. 56, 889–892.
Lam, R., Vlachos, D.G., 2001. Multiscale model for epitaxial growth of films: Growth mode transition. Phys. Rev. B 64.
Lapedes, A., Farber, R., 1987. Nonlinear signal processing using neural networks: Prediction and system modelling. Presented at the IEEE international conference on neural networks, San Diego, California.
Lee, J.H., Wong, W., 2010. Approximate dynamic programming approach for process control. J. Process Control, ADCHEM 2009 Special Issue 20, 1038–1048.
Lee, J.M., Lee, J.H., 2009. An approximate dynamic programming based approach to dual adaptive control. J. Process Control 19, 859–864.
Lee, J.M., Lee, J.H., 2005. Approximate dynamic programming-based approaches for input–output data-driven control of nonlinear processes. Automatica 41, 1281–1288.
Li, G., Liu, Z., Li, J., Fang, Y., Liu, T., Mei, Y., Wang, Z., 2018. Application of general regression neural network to model a novel integrated fluidized bed gasifier. Int. J. Hydrog. Energy 43, 5512–5521.
Lou, Y., Christofides, P.D., 2004. Feedback control of surface roughness using stochastic PDEs. AIChE J. 51, 345–352.
Lou, Y., Christofides, P.D., 2003. Estimation and control of surface roughness in thin film growth using kinetic Monte-Carlo models. Chem. Eng. Sci. 58, 3115–3129.
Middlebrooks, S.A., Rawlings, J.B., 2007. Model predictive control of $Si_{1-x}Ge_x$ thin film chemical-vapor deposition. IEEE Trans. Semicond. Manuf. 20, 114–125.
Ni, D., Christofides, P.D., 2005a. Construction of stochastic PDEs for feedback control of surface roughness in thin film deposition, in: Proceedings of the 2005, American Control Conference, 2005. Presented at the Proceedings of the 2005, American Control Conference, 2005., pp. 2540–2547 vol. 4.
Ni, D., Christofides, P.D., 2005b. Multivariable predictive control of thin film deposition using a stochastic PDE model. Ind. Eng. Chem. Res. 44, 2416–2427.
Oliveira, R., 2004. Combining first principles modelling and artificial neural networks: a general framework. Comput. Chem. Eng., ESCAPE 13 28, 755–766.
Ott, A.W., Klaus, J.W., Johnson, J.M., George, S.M., 1997. Al_3O_3 thin film growth on Si(100) using binary reaction sequence chemistry. Thin Solid Films 292, 135–144.
Ozkaya, B., Demir, A., Bilgili, M.S., 2007. Neural network prediction model for the methane fraction in biogas from field-scale landfill bioreactors. Environ. Model. Softw. 22, 815–822.
Rasoulian, S., Ricardez-Sandoval, L.A., 2016. Stochastic nonlinear model predictive control applied to a thin film deposition process under uncertainty. Chem. Eng. Sci. 140, 90–103.
Rasoulian, S., Ricardez-Sandoval, L.A., 2015a. A robust nonlinear model predictive controller for a multiscale thin film deposition process. Chem. Eng. Sci. 136, 38–49.

Rasoulian, S., Ricardez-Sandoval, L.A., 2015b. Robust multivariable estimation and control in an epitaxial thin film growth process under uncertainty. J. Process Control 34, 70–81.
Rasoulian, S., Ricardez-Sandoval, L.A., 2014. Uncertainty analysis and robust optimization of multiscale process systems with application to epitaxial thin film growth. Chem. Eng. Sci. 116, 590–600.
Rawlston, J.A., Schork, F.J., Grover, M.A., 2011. Multiscale modeling of branch length in butyl acrylate solution polymerization: molecular versus continuum kinetics. Macromol. Theory Simul. 20, 645–659.
Rawlston, J.A., Schork, F.J., Grover, M.A., 2010. Multiscale modeling of branch length in butyl acrylate solution polymerization. Macromol. Theory Simul. 19, 309–322.
Ricardez-Sandoval, L.A., 2011. Current challenges in the design and control of multiscale systems. Can. J. Chem. Eng. 89, 1324–1341.
Rusli, E., Drews, T.O., Ma, D.L., Alkire, R.C., Braatz, R.D., 2006. Robust nonlinear feedback–feedforward control of a coupled kinetic Monte Carlo–finite difference simulation. J. Process Control 16, 409–417.
Seshan, K., Schepis, D., 2012. Handbook of thin film deposition, 4th ed. William Andrew, Oxford.
Siddhamshetty, P., Wu, K., Kwon, J.S.-I., 2018a. Optimization of simultaneously propagating multiple fractures in hydraulic fracturing to achieve uniform growth using data-based model reduction. Chem. Eng. Res. Des. 136, 675–686.
Siddhamshetty, P., Yang, S., Kwon, J.S.-I., 2018b. Modeling of hydraulic fracturing and designing of online pumping schedules to achieve uniform proppant concentration in conventional oil reservoirs. Comput. Chem. Eng. 114, 306–317.
Sundaram, A., Ghosh, P., Caruthers, J.M., Venkatasubramanian, V., 2001. Design of fuel additives using neural networks and evolutionary algorithms. AIChE J. 47, 1387–1406.
Svozil, D., Kvasnicka, V., Pospíchal, J., 1997. Introduction to multi-layer feed-forward neural networks. Chemom. Intell. Lab. Syst. 39, 43–62.
Theodoropoulou, A., Adomaitis, R.A., Zafiriou, E., 1998. Model reduction for optimization of rapid thermal chemical vapor deposition systems. IEEE Trans. Semicond. Manuf. 11, 85–98.
Ulissi, Z.W., Strano, M.S., Braatz, R.D., 2013. Control of nano and microchemical systems. Comput. Chem. Eng. 51, 149–156.
Venkatasubramanian, V., 2009. DROWNING IN DATA: Informatics and modeling challenges in a data-rich networked world. AIChE J. 55, 2–8.
Venkatasubramanian, V., Vaidyanathan, R., 1992. Diagnosing noisy process data using neural networks, in: Fault Detection, Supervision and Safety for Technical Processes 1991, IFAC Symposia Series. Pergamon, Oxford, pp. 547–552.
Vlachos, D.G., 2005. A review of multiscale analysis: examples from systems biology, materials engineering, and other fluid–surface interacting systems, in: Marin, G.B., Advances in Chemical Engineering, Advances in Chemical Engineering: Multiscale Analysis. Academic Press, San Diego, California, pp. 1–61.
Vvedensky, D.D., 2003. Edwards-Wilkinson equation from lattice transition rules. Phys. Rev. E 67, 025102.
Vvedensky, D.D., Zangwill, A., Luse, C.N., Wilby, M.R., 1993. Stochastic equations of motion for epitaxial growth. Phys. Rev. E 48, 852–862.

ACCEPTED MANUSCRIPT

Walsh, J.B., 1986. An introduction to stochastic partial differential equations, in: École d'Été de Probabilités de Saint Flour XIV - 1984, Lecture Notes in Mathematics. Springer, Berlin, pp. 265–439.

Yegnanarayana, B., 2009. Artificial neural networks. PHI Learning Pvt. Ltd., New Delhi.

ACCEPTED MANUSCRIPT

Figure Captions

Figure 1 Schematic of the deposition of a thin film onto a substrate

Figure 2 The general structure of a two-layer ANN

Figure 3 The structure of the ANN implemented to predict the SPDE coefficients, including the sub-ANNs used to predict ν and K subject to changes in the system. Black: sub-ANNs used to determine coefficients independent of the direction of temperature change; blue: sub-ANNs used to determine coefficients subject to increasing temperature changes ≥25 K, over temperature change time intervals ≤200 s and temperatures ≥1000 K; red: sub-ANNs used to determine coefficients subject to all other conditions.

Figure 4 The overall structure of the hybrid multiscale model, including the coupling between the continuum PDEs, the ANN, and the SPDE

Figure 5 Hybrid multiscale model validation under random changes in the surface temperature: a) Temperature profile; b) Observed surface mole precursor fraction profile; c) Observed surface roughness profiles; d) Observed film growth rates

Figure 6 Variation in the surface roughness and growth rates obtained using the *minR* and *maxGr* temperature profiles

Figure 7 Optimal temperature profiles as determined from the *minR*, *maxGr*, and *Multi-obj* optimization studies

Figure 8 Variation in the surface roughness and growth rates obtained using the *Multi-obj* optimal trade-off temperature profile

Figure 9 Location of the optimal trade-off point (red) and the utopia point (blue) along the roughness/growth rate Pareto front (black)

Figure 10 Block diagram of the feedback control scheme applied to the thin film deposition process

Figure 11 Surface roughness feedback control subject to nominal conditions: a) Surface temperature control actions; b) film roughness trajectories from the estimator (hybrid multiscale model, green/yellow) and the plant (kMC-based multiscale model, blue) for each targeted set-point (red dashed lines)

Figure 12 Surface roughness feedback control subject to disturbances in y_s: a) Surface temperature control actions; b) film roughness trajectories from the estimator (hybrid multiscale model, green/yellow) and the plant (kMC-based multiscale model, blue) for each targeted set-point (red dashed lines)

Figure 13 Surface roughness feedback control subject to parametric uncertainty in the activation energies: a) Surface temperature control actions; b) film roughness trajectories from the estimator (hybrid multiscale model, green/yellow) and the plant (kMC-based multiscale model, blue) from 12 uncertainty realizations for each targeted set-point (red dashed lines)

ACCEPTED MANUSCRIPT

Table 1. Model parameter values

Symbol	Value
a	5 s^{-1}
L	10,001 sites
Sc	0.75
y_{bulk}	2×10^{-6}
Δt_{MS}	2 s
λ	6.5×10^{-3} m/s
$\mu_{bulk}\rho_{bulk}$	9×10^{11}
ρ_{bulk}/ρ	1
σ	7×10^{-4} m^3/s

ACCEPTED MANUSCRIPT

Table 2. Sum of squared errors between the hybrid multiscale model and the kMC-based multiscale model results subject to random changes in the substrate temperature.

Interval	**Substrate Temperature, T_s (K)**	**Mean Surface Mole Precursor Fraction, y_s**	**Sum of Squared Errors, Surface Roughness**	**Sum of Squared Errors, Film Growth Rate**
1	1,115.7	1.231×10^{-6}	1.128×10^{-2}	1.806
2	1,281.0	1.453×10^{-6}	4.264×10^{-2}	0.345
3	1,040.3	1.165×10^{-6}	5.590×10^{-2}	0.307
4	981.4	1.126×10^{-6}	8.965×10^{-2}	0.012
5	1,010.8	1.138×10^{-6}	3.337×10^{-2}	0.120
6	1,266.5	1.426×10^{-6}	1.667×10^{-2}	0.216
7	1,187.8	1.312×10^{-6}	1.761×10^{-2}	0.074
8	1,069.5	1.187×10^{-6}	1.630×10^{-2}	0.098
9	1,203.9	1.324×10^{-6}	4.491×10^{-2}	0.085
10	1,156.7	1.275×10^{-6}	2.794×10^{-2}	0.080
11	973.4	1.123×10^{-6}	6.134×10^{-2}	0.200
12	1,262.3	1.418×10^{-6}	0.308×10^{-2}	0.210

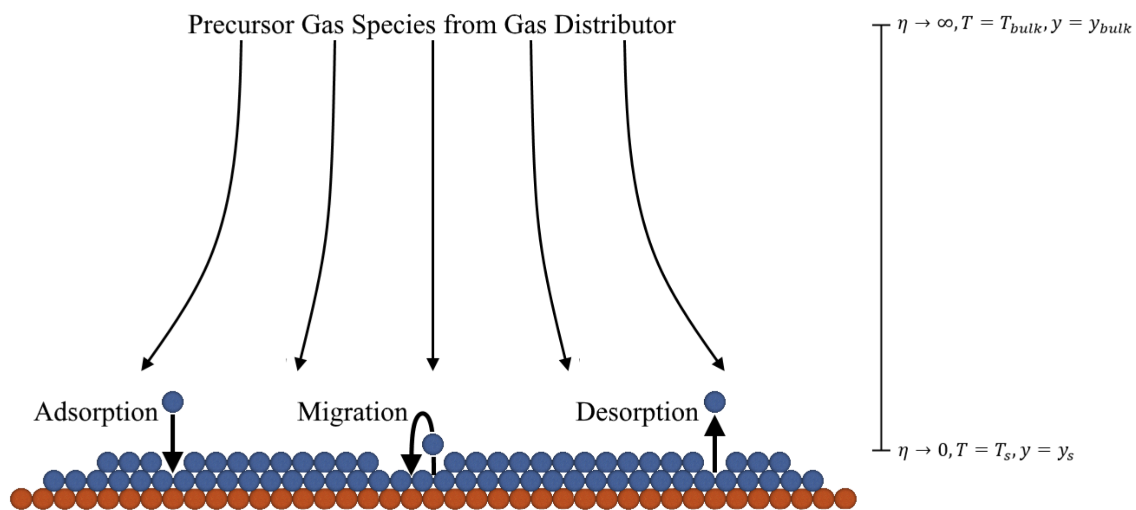
ACCEPTED MANUSCRIPT

Table 3. Optimization study results.

Optimization	$\boldsymbol{R(t_f)}$ **(mL)**	$\boldsymbol{Gr_\Sigma(t_f)}$ **(mL/s)**
minR	1.17271 (R^l)	939.331 (Gr_Σ^l)
maxGr	2.69693 (R^u)	1,392.40 (Gr_Σ^u)
Multi-obj	1.40450	1,358.64

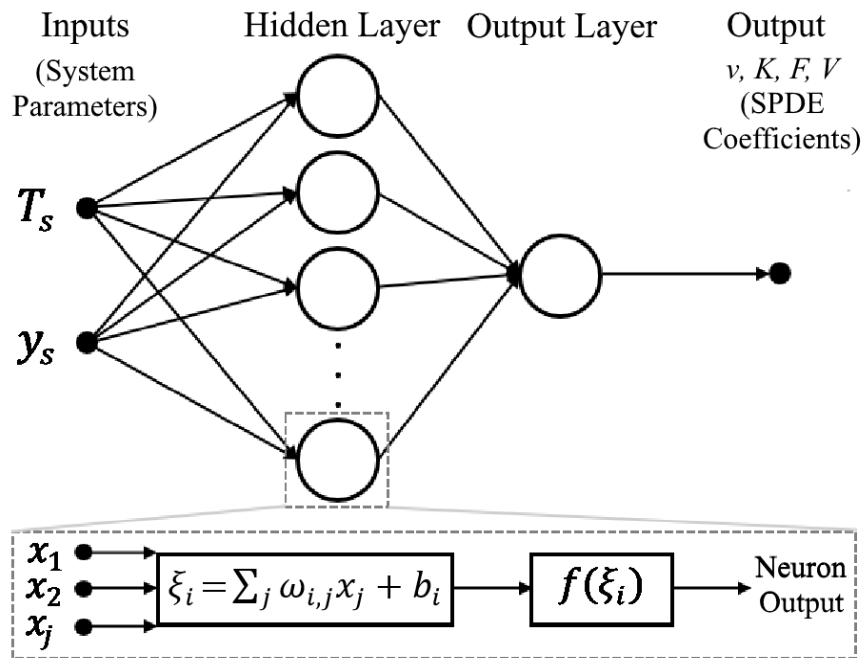
ACCEPTED MANUSCRIPT

Figure 1



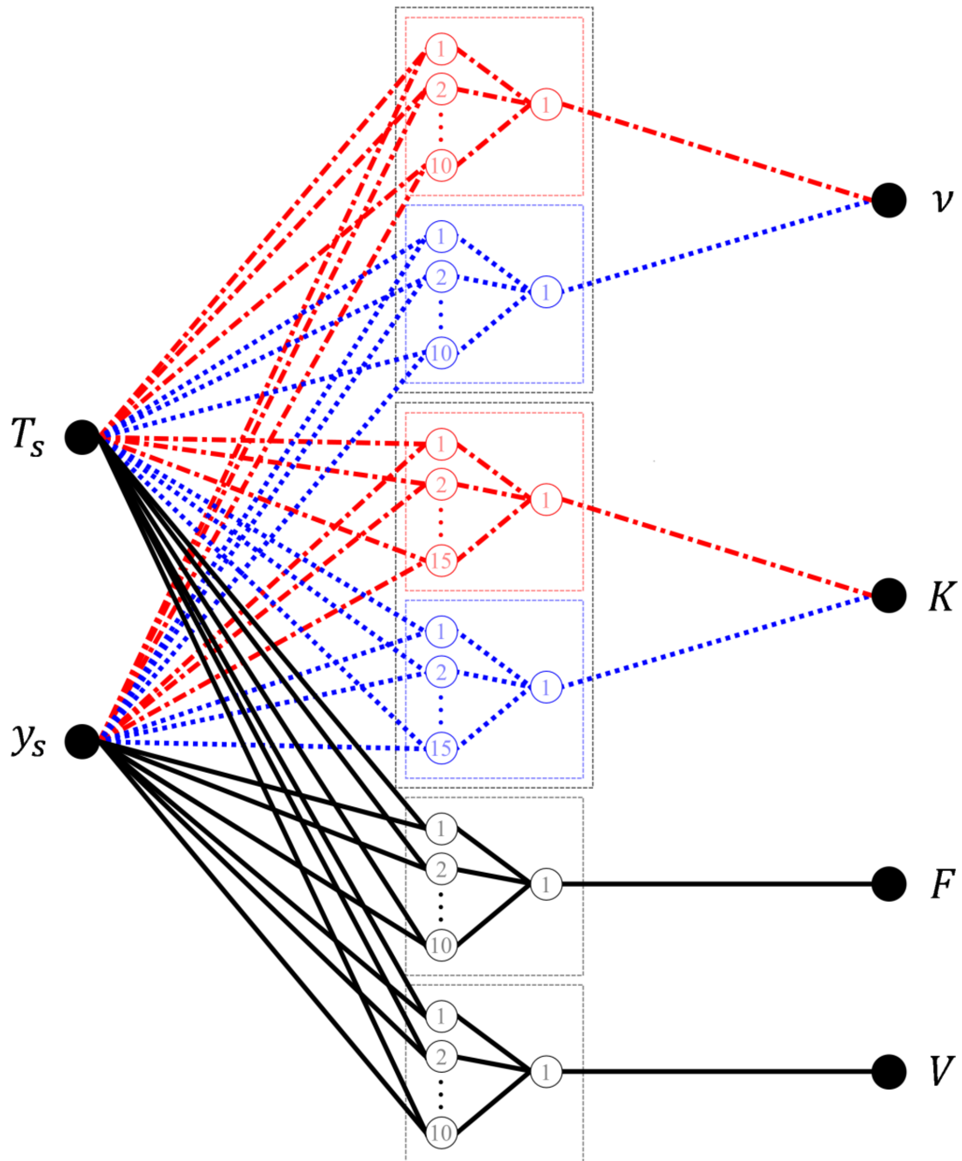
ACCEPTED MANUSCRIPT

Figure 2



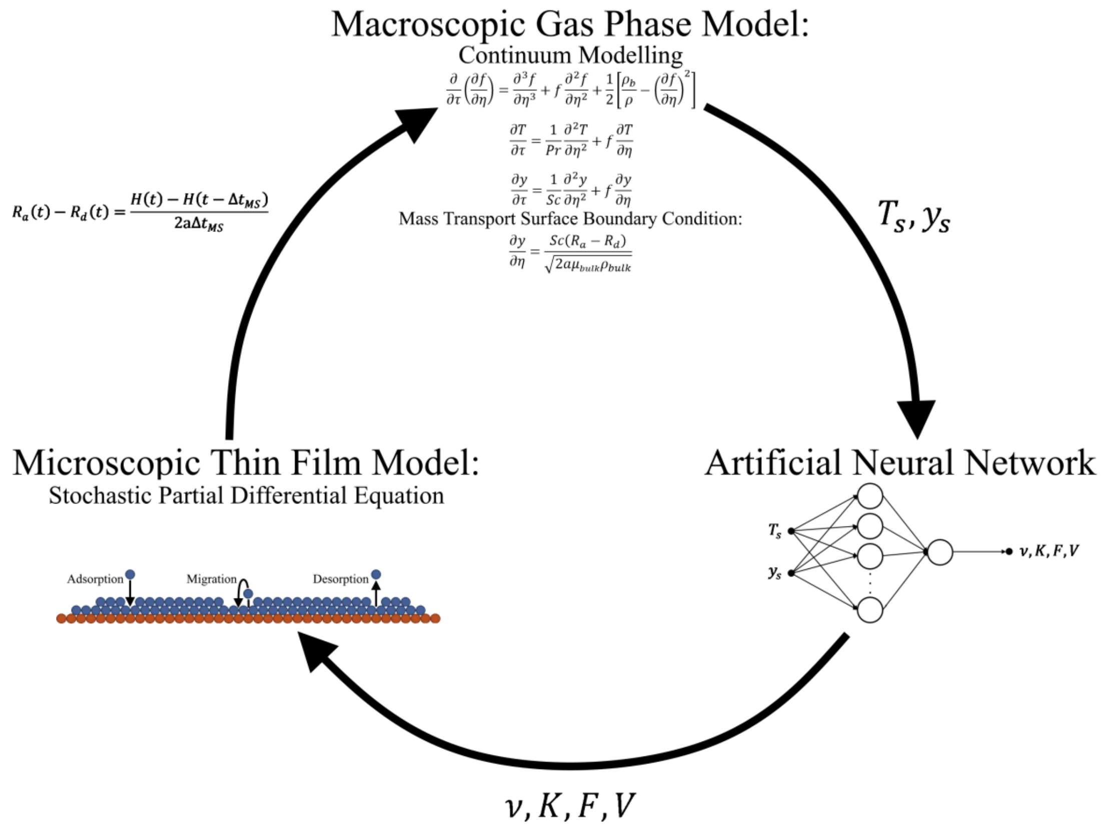
ACCEPTED MANUSCRIPT

Figure 3



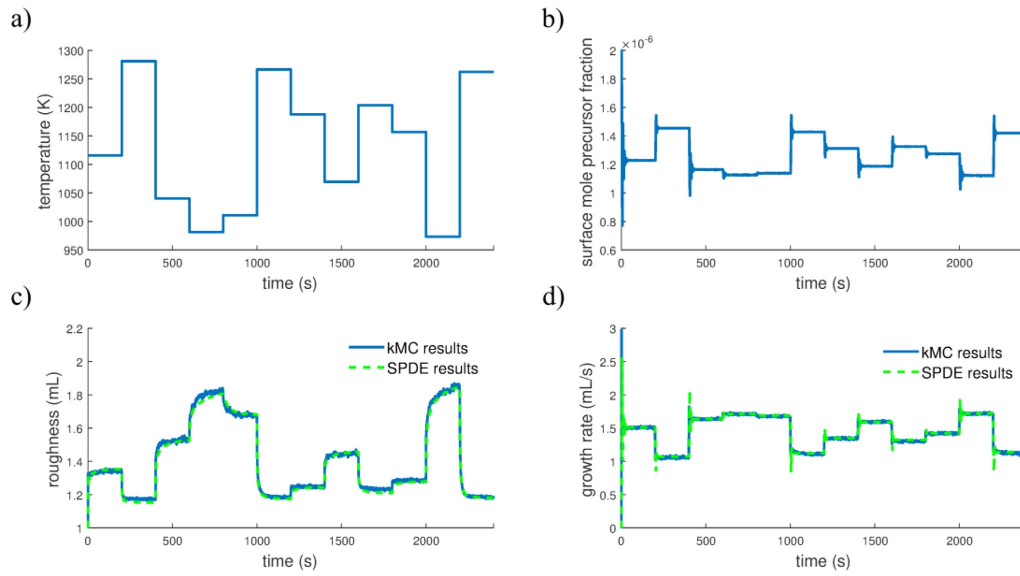
ACC

Figure 4



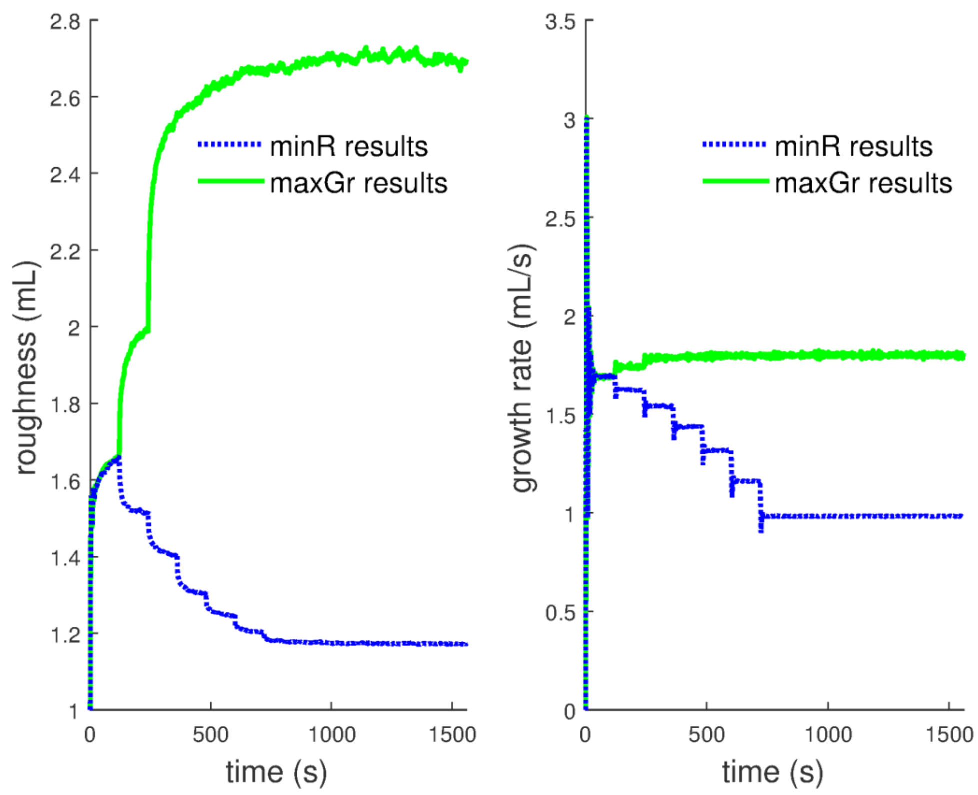
ACCEPTED

Figure 5



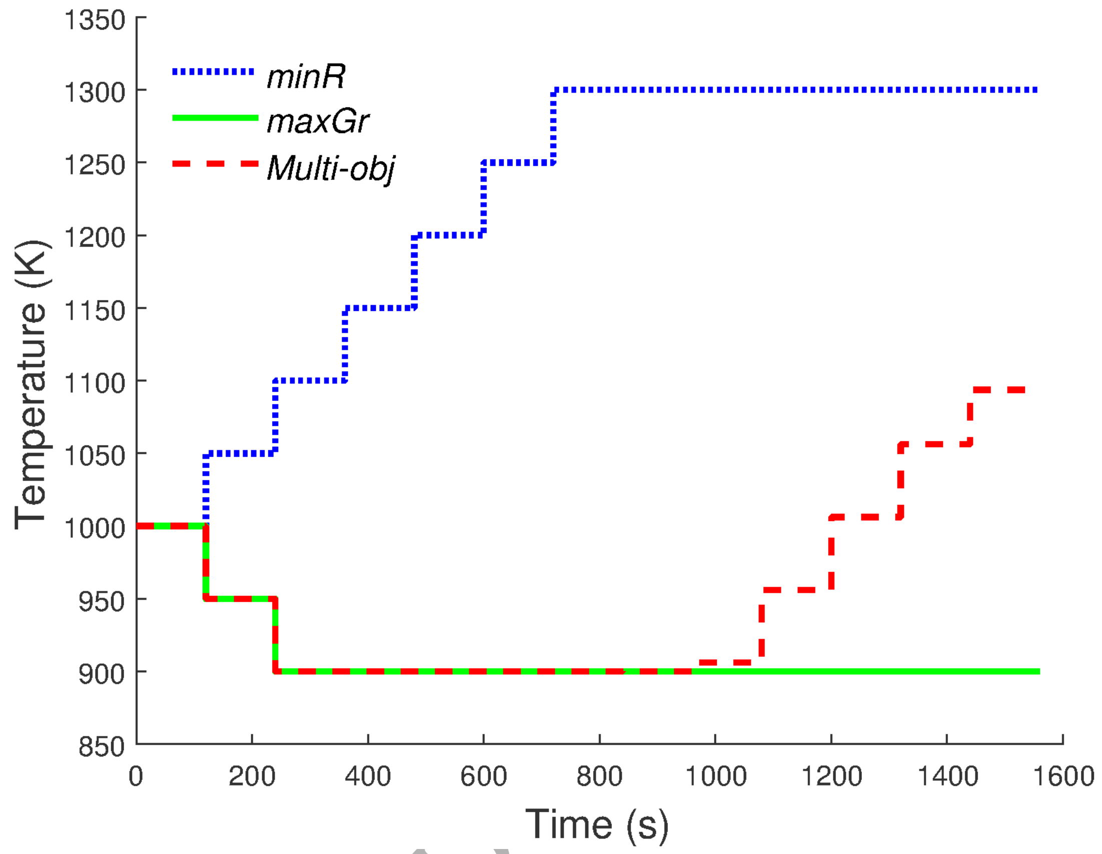
ACCEPTED MANUSCRIPT

Figure 6



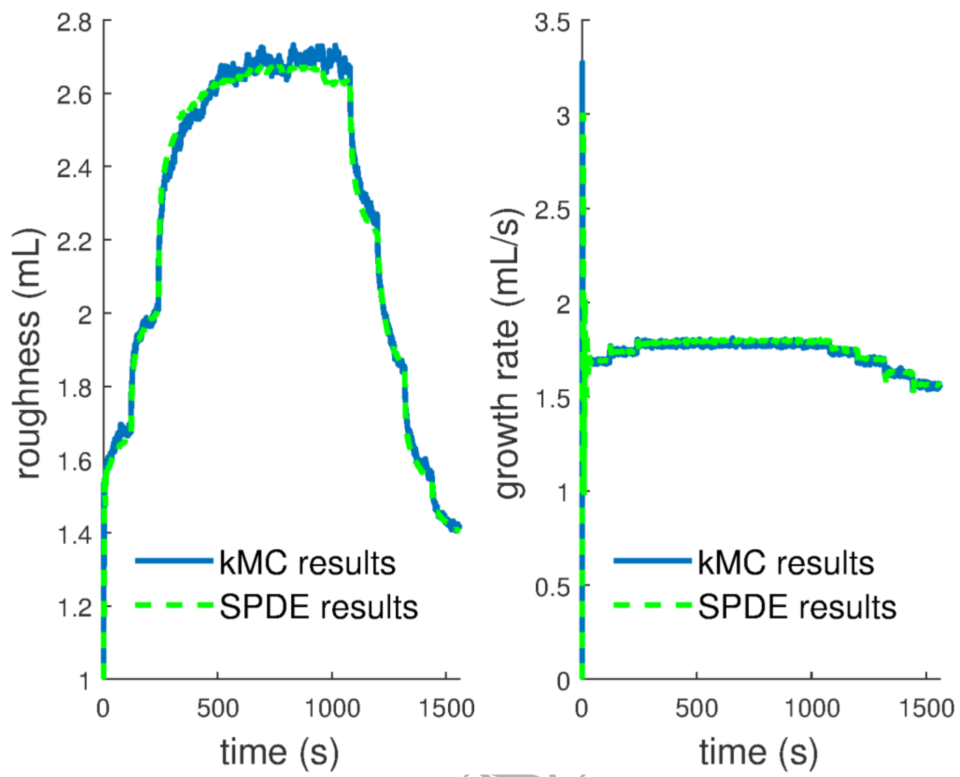
ACCEPTED

Figure 7



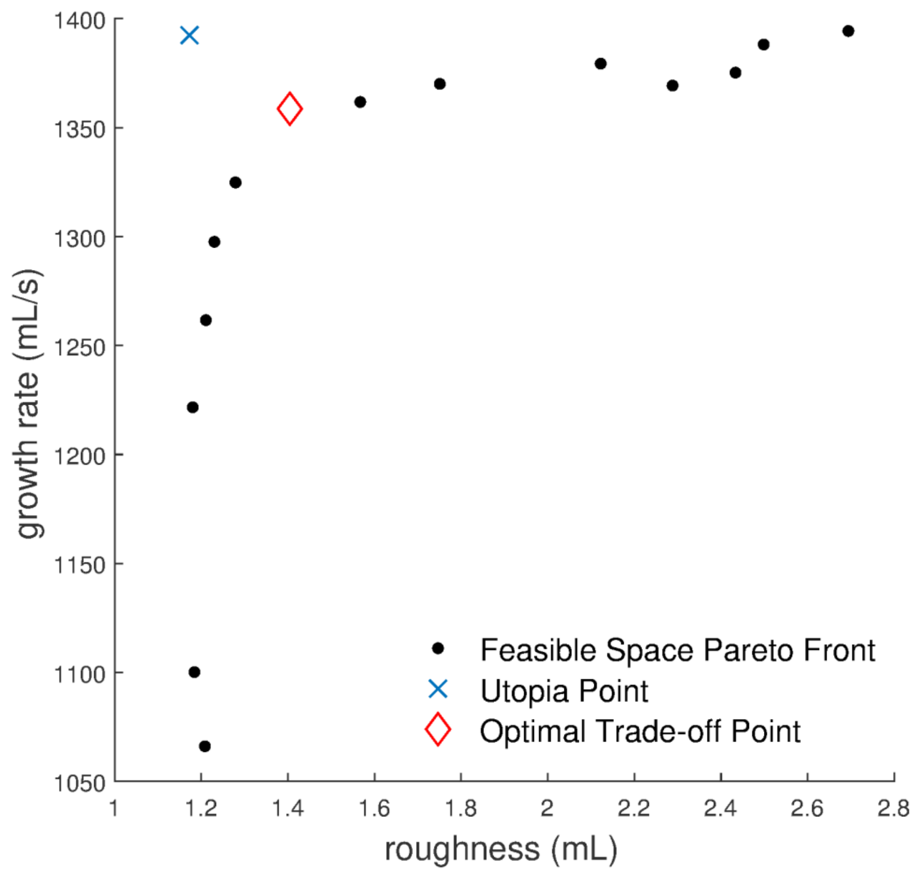
ACCEPTED

Figure 8



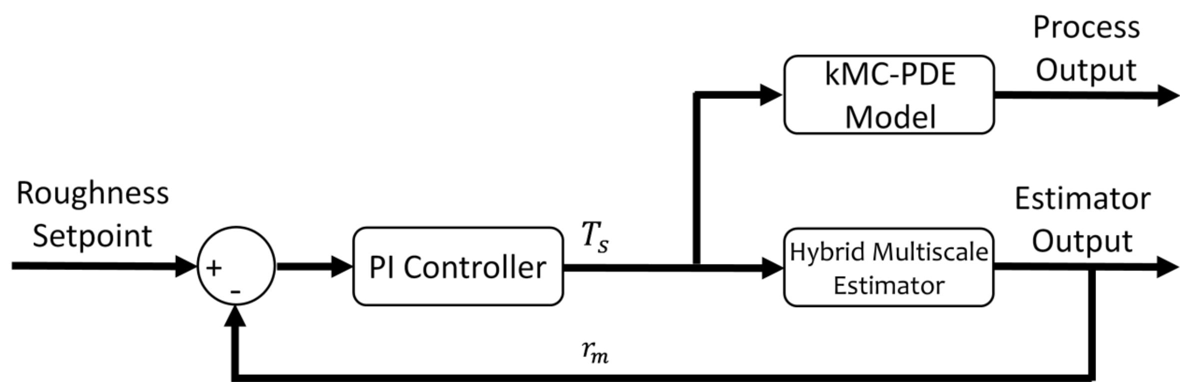
ACCEPTED MANUSCRIPT

Figure 9



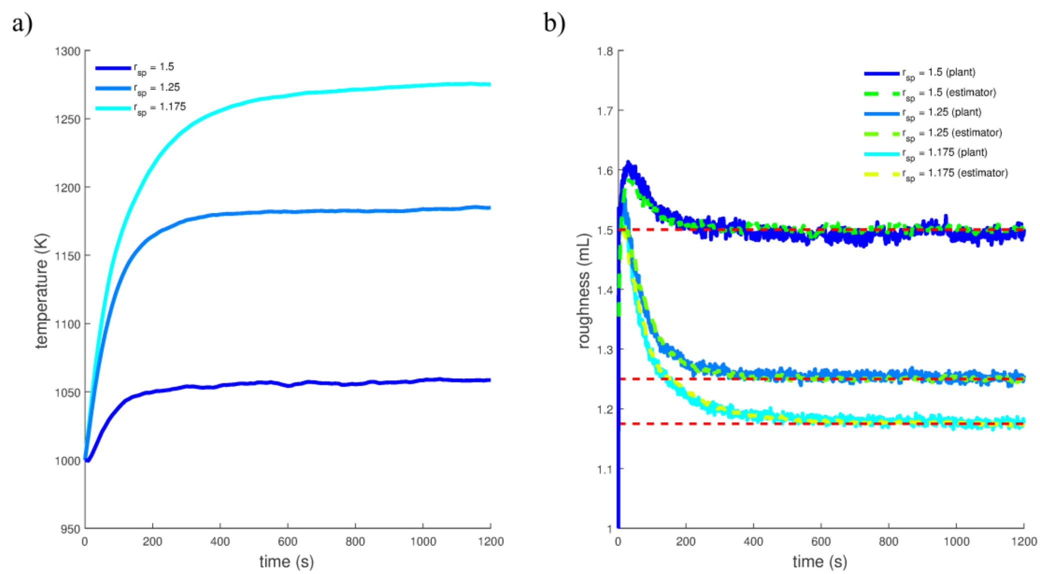
ACCEPTED MANUSCRIPT

Figure 10



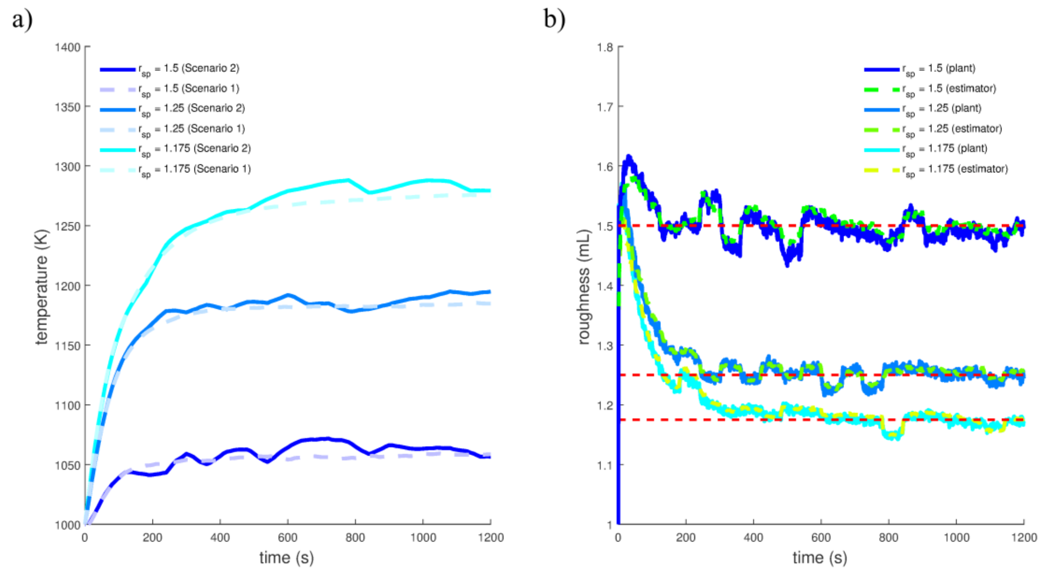
ACCEPTED MANUSCRIPT

Figure 11



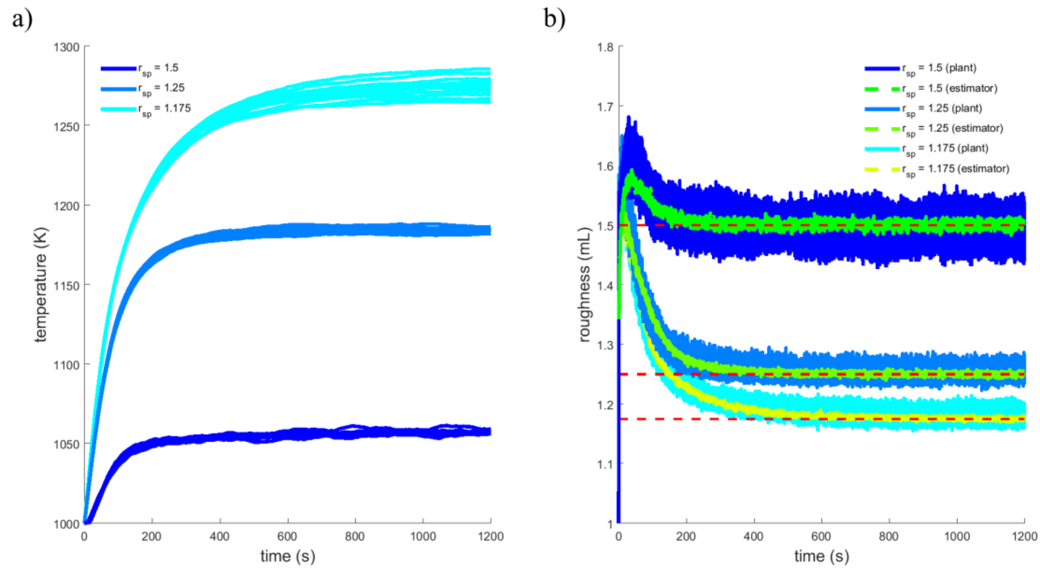
ACCEPTED MANUSCRIPT

Figure 12



ACCEPTED MANUSCRIPT

Figure 13



ACCEPTED MANUSCRIPT