Contents lists available at ScienceDirect

Chemical Engineering Research and Design

journal homepage: www.elsevier.com/locate/cherd

Bayesian model averaging for estimating the spatial temperature distribution in a steam methane reforming furnace

Anh Tran[a], *Madeleine Pont*[a], *Andres Aguirre*[a], *Helen Durand*[a], *Marquis Crose*[a], *Panagiotis D. Christofides*[a,b,*]

[a] *Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90095-1592, USA*
[b] *Department of Electrical Engineering, University of California, Los Angeles, CA 90095-1592, USA*

ARTICLE INFO

Article history:
Received 21 July 2017
Received in revised form 19 September 2017
Accepted 22 September 2017
Available online 4 October 2017

Keywords:
Steam methane reforming
CFD modeling
Bayesian modeling
Sparse modeling techniques
Furnace operation
Process safety

ABSTRACT

In this work, we introduce a statistical-based model identification scheme that generates a high-fidelity model for the outer reforming tube wall temperature (OTWT) distribution as a function of the furnace-side feed (FSF) distribution, total FSF flow rate and interactions among neighboring reforming tubes from reformer data. The proposed scheme is structured to have two major components, namely, a prediction step and a correction step, which are designed to be parallelized so that the prediction and correction models of all reforming tubes are derived simultaneously from the reformer data and independently from one another. Initially, a computational fluid dynamics (CFD) model of an industrial-scale reformer created in our previous work is utilized to facilitate the generation of the training and testing data. Then, we propose the development of an algorithm for the prediction step based on Bayesian variable selection, Bayesian model averaging, sparse nonlinear regression, reformer geometry and theories of thermal radiation so that for each reforming tube, the prediction step can systematically identify predictors for the OTWT and simultaneously create a corresponding library of sub-prediction models. A collection of prediction models for all reforming tubes is defined as a prediction model for the OTWT distribution, which is expected to capture the dependence of the OTWT distribution on the FSF distribution and total FSF flow rate. Next, we propose an algorithm for the correction step designed based on ordinary Kriging so that for each reforming tube, the correction step creates a spatial model allowing the OTWT to be estimated from the predicted OTWT of the neighboring reforming tubes. A collection of correction models for all reforming tubes is defined as a correction model for the OTWT distribution, which is expected to adjust the predicted OTWT distribution to account for interactions among neighboring reforming tubes. Subsequently, the combined data-driven model for the OTWT distribution is created using the prediction and correction models for the OTWT distribution, which allows the combined model to account for the effect of interactions among neighboring reforming tubes while estimating the OTWT distribution based on the FSF distribution and total FSF flow rate. The proposed integrated model identification scheme is executed on the Hoffman2 cluster at UCLA to construct the data-driven model for the OTWT distribution from the training data, and the results from the goodness-of-fit and out-of-sample prediction tests of the data-driven model are used to demonstrate the effectiveness of the scheme proposed in this work.

* *Corresponding author.*
E-mail address: pdc@seas.ucla.edu (P.D. Christofides).
https://doi.org/10.1016/j.cherd.2017.09.027

1. Introduction

Hydrogen is proposed by many to be the fuel of the future as it is the key ingredient in a transition from a fossil fuel-based economy toward a zero carbon emission and sustainable energy economy. Hydrogen can serve as an efficient energy carrier for hydrogen-based technologies (e.g., fuel cells) and lead to substantial reduction of greenhouse gas emissions and great environmental benefits. Hydrogen can be produced by a variety of technologies (e.g., steam methane reforming (SMR), coal gasification, biomass gasification, electrolysis, partial oxidation, solar thermal cracking) from fossil (e.g., natural gas), non-fossil (e.g., biogas) and non-carbon (e.g., water) sources (Barreto et al., 2003), which highlights the great potential and flexibility of a hydrogen-based economy. Additionally, hydrogen is a key feedstock for the petroleum refining and fine chemical manufacturing industries (Tran et al., 2017a).

Among the technologies employed to produce high purity hydrogen on a commercial scale, SMR and coal gasification are the current major technologies. Barreto et al. (2003) conducted a survey of literature and publicly available databases associated with hydrogen manufacturing technologies and concluded that SMR required the least investment for hydrogen plants with a capacity above 1,000,000 N m^3 per day in comparison to the amount needed by the competing technologies. Our literature survey recognizes that SMR accounts for the largest share of world hydrogen production (Hydrogen Production Expert Panel, 2013; Barreto et al., 2003; Ewan and Allen, 2005); in 2005, for example, approximately 48% of worldwide hydrogen production was derived from SMR. Barreto et al. (2003) speculated that SMR would remain the leading technology in world hydrogen production in 2050.

SMR is a catalytic endothermic process, in which natural gas (i.e., methane) and high pressure superheated steam react to produce hydrogen and carbon oxides. As methane and steam are naturally stable compounds, SMR is typically carried out at high temperature (e.g., 1023.15–1223.15 K, Pantoleontos et al., 2012) to ensure that the process is economically viable, and a desired methane conversion is achieved; hence, SMR is carried out inside furnaces. A steam methane reforming furnace (which is referred to as a reformer in this manuscript) has two closed and thermally coupled domains (i.e., the tube side and furnace side); the tube side is defined by hundreds of nickel-based catalyst filled reforming reactors (which are referred to as reforming tubes in this manuscript), and the furnace side is a combustion chamber. Specifically, heat is generated in the furnace side by burning fresh natural gas and, typically, recycled effluent from the purification section (e.g., a pressure swing adsorption unit) (Kroschwitz and Howe-Grant, 1999) in excess air to create a high temperature environment facilitating the production of hydrogen in the tube side.

Reformers can be categorized with respect to the location of burners: top-fired, side-fired, bottom-fired and terrace wall-fired reformers, which in turn dictates interactions between the tube side and furnace side, the temperature distribution characteristics of the furnace-side flow inside the reformer and the heat flux profile along the reforming tubes, e.g., bottom-fired reformers are characterized by a constant heat flux along the reforming tubes (Ferreira-Aparicio et al., 2005). In SMR, the reformer configurations that facilitate a high heat transfer rate to the tube side near the reforming tube inlet are expected to require the shortest reforming tube length to achieve a desired set-point of the methane conversion and are preferable. Therefore, the top-fired reformer is frequently employed at hydrogen plants using SMR technology and is the subject of the present work.

The chemical manufacturing plants designed to produce high purity hydrogen based on SMR technology typically consist of two major sections: the synthesis section and purification section (Gupta, 2008). In the synthesis section, fresh natural gas feedstock is treated in a desulfurization unit to remove thiol compounds preventing catalysts used in downstream processes from being poisoned. Then, the effluent of the desulfurization unit typically undergoes a catalytic prereforming process to convert higher hydrocarbons in the treated feedstock into methane and byproducts (i.e., carbon oxides and hydrogen), preventing these high hydrocarbons from decomposing inside the reforming tubes, which causes coke formation and catalyst deactivation. Next, the treated feedstock (i.e., the effluent of the prereforming process) and high pressure superheated steam are fed into the reformer to undergo SMR producing hydrogen. The synthesis gas (i.e., the effluent of the reformer) is treated in a water-shift reactor to remove carbon monoxide and to produce a small additional amount of hydrogen, and the effluent of the water-shift reactor enters the purification section in which the process stream is stripped of unreacted reactants and byproduct (i.e., carbon dioxide) to produce high purity product.

Tran et al. (2017a) indicates that an increase in the total furnace-side feed (FSF) flow rate is expected to result in increases in both the averaged outer reforming tube wall temperatures (OTWTs) and the methane conversion of SMR in the reformer, and hence, it can be inferred that the highest reformer efficiency is achieved when the maximum total FSF flow rate is delivered. In the present work, a collection of all radially weighted averaged OTWTs of all reforming tubes at a fixed height is defined as an OTWT distribution, and an averaged temperature profile among all reforming tubes along the reforming tube length is defined as an averaged OTWT profile. In practice, it is difficult to find the FSF distribution to deliver the maximum total FSF flow rate without damaging the reforming tubes and shortening the reformer service life (which can occur at temperatures above the design temperature of the reforming tube wall) because the temperature distribution of the furnace-side flow in the reformer is spatially nonuniform and is tightly coupled with all OTWT distributions along the reforming tube length. Specifically, the temperature distribution of the furnace-side flow governs all OTWT distributions along the reforming tube length in the reformer, and therefore, a nonuniform temperature distribution of the furnace-side flow causes the OTWT distributions to also be nonuniform (i.e., to be platykurtic or multimodal with a wide temperature range). In addition, current technologies (e.g., an infrared camera system positioned around the reformer) allow the OTWT distributions to be continuously monitored, so OTWT distributions are frequently used to assess the degree of nonuniformity in the temperature distribution of the furnace-side flow, which is measured as the greatest temperature difference in the OTWT distribution at a fixed height among all reforming tubes. This definition of the degree of nonuniformity in the temperature distribution of the furnace-side flow has also been used in the SMR literature (Kumar et al., 2015; Tran et al., 2017a; Zheng et al., 2010); for instance, Kumar et al. (2015) reports that the variation in OTWT of all reforming tubes at a fixed height fluctuates between 30 K and 110 K.

In a reformer with a high degree of nonuniformity in the temperature distribution of the furnace-side flow (i.e., an imbalanced reformer) the total FSF flow rate is reduced so that the maximum temperature value of the averaged OTWT profile is often kept significantly lower than the design temperature of the reforming tube wall to avoid damaging the reforming tubes, which leads to a lower reformer efficiency. It is important to note that in the above scenario, although the maximum temperature value of the averaged OTWT profile is significantly lower than the design temperature of the reforming tube wall, the high degree of nonuniformity in OTWT distributions suggests that OTWTs of some reforming tubes may be near the design temperature. The typical operating strategy of reformers calibrates the FSF distribution and total FSF flow rate in an attempt to compensate for the nonuniform temperature distribution of the furnace-side flow inside the reformer to prevent the OTWT of all reforming tubes from exceeding the design temperature of the reforming tube wall and to extend the service life of the reformer at the expense of the reformer efficiency. The hydrogen manufacturing industry reasons that this trade-off is necessary because when the OTWT of a reforming tube exceeds the design temperature of the reforming tube wall, the reformer might have a significantly shorter service life, e.g., Pantoleontos et al. (2012) reports that an increase of 20 K above the design temperature of the reforming tube wall causes the reformer service life to be reduced by half, and the reforming tubes are at risk of rupturing; if rupture were to occur, it would lead to substantial production and capital losses. For decades, this ad hoc operating strategy of the reformer, which is designed to retain the reformer service life by allowing the reformer to be operated at suboptimal conditions, has been a solution that prevents significant unexpected capital costs as the cost of re-tubing is estimated to be approximately 10% of the total capital investment (Latham et al., 2011). Therefore, designing operating strategies to actively minimize the degree of nonuniformity in the temperature distribution of the furnace-side flow and maximize the total FSF flow rate is getting much attention from academics and industries. A number of references in the recent SMR literature have investigated furnace balancing methods, which are systematic operating strategies that search for an optimal FSF distribution that reduces the degree of nonuniformity in the OTWT distribution at a fixed height in the reformer in an attempt to improve the reformer efficiency. Therefore, it is apparent that an accurate relationship between the OTWT distribution at a fixed height, the FSF distribution and the total FSF flow rate of the reformer is required to design a robust furnace balancing method. The dependence of the OTWT distribution on the FSF distribution and the total FSF flow rate of the reformer can be constructed based on first principles modeling (Latham et al., 2011) and computational fluid dynamics (CFD) modeling (Tran et al., 2017b) of the reformer. However, models derived from these modeling techniques are typically unsuitable for designing a robust real-time furnace-balancing scheme (for which the need is evident as the reformer is constantly subjected to various disturbances, e.g., ambient temperature) (Kumar et al., 2015). In particular, first principles modeling often uses an overwhelming number of simplifying assumptions in the development of reformer models, which causes prediction data generated by first principles models often to fall short in terms of accuracy. While CFD modeling does not have the same issue (e.g., the simulation data generated by the reformer CFD model developed in Tran et al. (2017b) have been shown to be a reasonably accurate representation of the experimental data recorded from an on-line unit), the significant computational time needed to create a single CFD data set makes it unsuitable for designing a real-time furnace balancing scheme because the CFD model is required to be repeatedly solved with different FSF distributions to search for the optimized FSF distribution. Therefore, data-driven modeling is an appealing alternative as data-driven models are computationally inexpensive and can have reasonable accuracy.

In this work, data-driven modeling is used to discover the dependence of the OTWT distribution on the FSF distribution and the total FSF flow rate of the reformer. A direct approach is to model the OTWT at a fixed height of each reforming tube as a function of the FSF distribution and total FSF flow rate (i.e., the FSF flow rates of all burners), and estimates of the parameters associated with the regressors in the data-driven model can be determined by the ordinary least squares (OLS) regression method in which the sum of squared residuals between the OTWT data and fitted data is minimized. This naive approach is expected to create an uninterpretable (in the sense that it is not representing physical relationships between burners and reforming tubes within the reformer) data-driven model for the OTWT distribution with high prediction errors due to over fitting. Specifically, Tran et al. (2017b) notes that in the high temperature reformer, thermal radiation is expected to be the dominant mode of heat transfer, and Olivieri and Vegliò (2008) shows that 95% of the total heat transfer rate to the tube side in the reformer of their investigation is due to thermal radiation. Additionally, the inverse square law for thermal radiation suggests that the OTWT of each reforming tube is governed by the FSF flow rates of the nearby burners. Therefore, a smaller set of important regressors for each reforming tube should be identified and used to construct a more computationally efficient data-driven model.

Standard statistical practice employs shrinkage and subset selection techniques (e.g., LASSO, nonnegative garotte and ridge regression) to search for the set of important regressors and to calculate the estimates of their corresponding parameters based on some criteria (e.g., minimizing the sum of squared residuals) to derive a single best data-driven model. Subsequently, this standard statistical practice assumes this data-driven model to be the true model for the relationship between the OTWT of a reforming tube, FSF distribution and total FSF flow rates, and then utilizes the chosen model exclusively to make predictions. The greatest flaw of the standard statistical practice is that the approach overlooks the importance of competing data-driven models. Specifically, Hoeting et al. (1999) illustrates that two competing models with similar goodness of fit for given training data can yield substantially different predictions and suggests that predictions made based on a single data-driven model are unreliable. Bayesian statistics provides systematic straightforward methods to identify the set of important regressors for each reforming tube (i.e., Bayesian variable selection) and to account for model uncertainty in making predictions based on the observed database (i.e., Bayesian model averaging). Bayesian statistics is employed in many disciplines (e.g., chemistry, genetics, medicine and finance) and has led to over 587 publications between 1996 and 2014 (Fragoso and Neto, 2015).

In this work, we combine the Bayesian methods and sparse nonlinear regression technique (i.e., least absolute shrinkage and selection operator or LASSO) to derive a collection of data-driven models, each of which requires the minimum number of terms for a given basis set of regressors while

revealing the dependence of the OTWT of each reforming tube on the FSF distribution and total FSF flow rate with a reasonable accuracy. Specifically, it is assumed that a data-driven model for the OTWT of each reforming tube has the form of a linear combination of nonlinear/linear transformations of the original regressors (i.e., sparse nonlinear regression) so that the nonlinear characteristics in the observed data can be adequately described (Brunton et al., 2016; Wilson and Sahinidis, 2017). For instance, Brunton et al. (2016) shows that sparse nonlinear regression can be used to create data-driven models that describe the underlying dynamics of nonlinear systems (e.g., nonlinear oscillators and the chaotic Lorenz system). A typical library of nonlinear transformations contains monomial, exponential, logarithmic and trigonometric functions.

Lastly, it is recognized that the OTWT of each reforming tube is spatially correlated to those of the nearby neighbors because thermal radiation inside the reformer is expected to occur between the furnace-side flow, reforming tube walls and refractory wall (Tran et al., 2017b). It is also recognized from the reformer CFD simulation data reported in Tran et al. (2017a) that reforming tubes which are in close proximity to one another tend to have similar OTWTs. Therefore, it is logical to allow the data-driven model for the OTWT of each reforming tube to also utilize the information (i.e., the locations and OTWTs) of the neighboring reforming tubes, in addition to the FSF distribution and total FSF flow rate, to make predictions of the OTWT. In the geostatistics literature, there exist a variety of distance-weighted average interpolation algorithms (e.g., Kriging, triangulation and the inverse distance method), each of which can be used to estimate an OTWT of each reforming tube as a weighted average of the neighboring OTWTs based on a unique weighting function. In this work, Kriging is selected as it provides a straightforward approach to adjust the predicted OTWT distribution estimated based on the FSF distribution and total FSF flow rate to account for the effect of interactions among neighboring reforming tubes on the OTWT distribution. Kriging can also be used for clustering data, which occurs due to the arrangement of the reforming tubes inside the reformer, i.e., the reforming tubes are arranged in an irregular grid pattern (e.g., seven rows of 48 units), in which reforming tubes are situated close to one another along a row but are separated by a relatively large distance between rows. Additionally, the Kriging algorithm is designed to minimize the mean squared prediction error, which makes this interpolation algorithm more suitable for predicting the OTWT at unexplored operating conditions of the reformer.

Motivated by this, the present work focuses on developing a two-step prediction and correction model identification procedure that utilizes Bayesian methods with an efficient search algorithm (Occam's window), sparse nonlinear regression, Kriging, information on the reformer geometry and theories of thermal radiative heat transfer to derive a high-fidelity model from reformer data such that the model can account for interactions among neighboring reforming tubes in making predictions about the OTWT distribution based on the FSF distribution and total FSF flow rate. This manuscript is structured as followed: in Section 2, physical descriptions, process modeling and process simulation of a computational fluid dynamics (CFD) model of a reformer are briefly discussed to be used as supporting evidence that the reformer CFD data are adequate representations of the data from an on-line reformer, which allows us to use the reformer CFD model to facilitate the creation of the training and testing data. In Section 3, an overview of the integrated model identification scheme is presented, and the major components, namely, the prediction step and correction step, are introduced. In Section 4, a rundown of the prediction step that details the procedure for deriving the prediction model for the OTWT distribution from the training data is presented, and in Sections 4.1, 4.2 and 4.3 the integration of Bayesian variable selection, Bayesian model averaging, sparse nonlinear regression and theories of thermal radiation into an algorithm for the prediction step is described. In Section 5, a rundown of the correction step that details the procedure for creating the correction model for the OTWT distribution from the training data using ordinary Kriging is presented. In Section 6, the procedure to generate the combined data-driven model for the OTWT distribution from the prediction and correction models is detailed. Finally in Section 7, the goodness of fit and out-of-sample predictive performance of the data-driven model for the OTWT distribution generated from the integrated model identification scheme proposed in this work are evaluated using the training and testing data, respectively, and are discussed to highlight the potential of this work for being used in developing more optimal operating conditions of a reformer in a computationally efficient manner, e.g., it may be considered for use as the data-driven model for an on-line robust furnace balancing optimizer.

2. Reformer CFD database

In the present work, the high-fidelity reformer CFD model developed from our previous work Tran et al. (2017a,b) is used to represent an on-line reformer designed by Selas Fluid Processing Corporation at a hydrogen plant (Latham et al., 2011). This is because the geometry of the reformer model is created to have approximate dimensions of the on-line unit, which is 16 m wide, 16 m long and 13 m tall. The geometry of the reformer model also contains important features of its physical counterpart, which include the major components (i.e., 336 reforming tubes, 72 inner-lane burners, 24 outer-lane burners and 8 flue-gas tunnels) and the layout inside the reformer as shown in Fig. 1. Additional details of the reformer geometry can be found in Tran et al. (2017b). Furthermore, the mesh of the reformer CFD model has been verified to have acceptable mesh quality based on the criteria (i.e., min orthogonal factor and max ortho skew) suggested by ANSYS ICEM and to allow the reformer CFD model to produce mesh-independent solutions. The generation of the reformer mesh is detailed in Tran et al. (2017a,b). In addition, the modeling strategies for the known transport phenomena and chemical reactions associated with SMR and air combustion processes are used to create the reformer CFD model. For instance, the reformer CFD model is implemented with the empirical correlation for the furnace-side total emissivity (Maximov, 2012), Lambert Beer's law, Kirchoff's law and the discrete ordinate method to simulate radiative heat transfer between the furnace-side flow, outer reforming tube walls and refractory walls. Readers who are interested in the details of the modeling strategies for the reformer CFD model, efficient step-by-step converging strategy and data collection procedure are recommended to refer to our previous work (Tran et al., 2017a,b). Moreover, the CFD simulation data generated by the reformer CFD model have been shown to be in good agreement with simulation data generated by a reforming Gibbs reactor model, with typical plant data in the SMR literature and with plant data provided

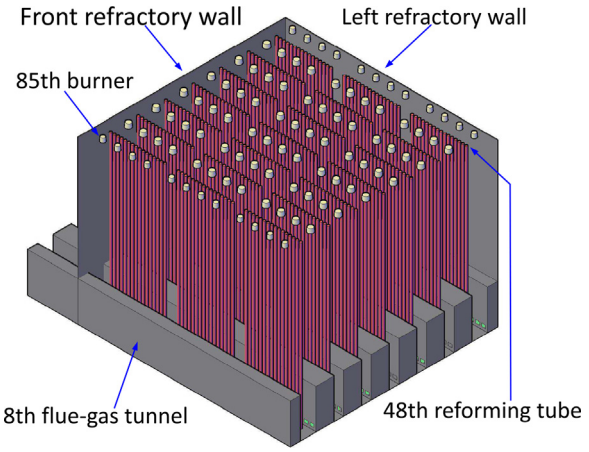

Fig. 1 – The isometric view of an industrial-scale, top-fired, co-current reformer. The right and back refractory walls with respect to the flue gas tunnel exits in this figure are removed to expose the interior of the reformer, which consists of 336 reforming tubes, which are represented by 336 cylinders, 96 burners, which are represented by 96 frustum cones, and 8 flue-gas tunnels, which are represented by 8 rectangular intrusions.

by a third party collaborator, and therefore, can be considered as an adequate representation of plant data collected from an on-line unit.

In the present work, the reformer CFD model is utilized to facilitate the creation of the reformer database. Specifically, the reformer CFD model is initially implemented with varying FSF distributions at a fixed typical total FSF flow rate and, subsequently, implemented with varying total FSF flow rates at the optimized FSF distribution reported in Tran et al. (2017a), which creates 21 CFD data sets. In this work, 18 CFD data sets (i.e., the training data), which is approximately 90% of the reformer database, are randomly chosen and used as inputs for the integrated model identification scheme to produce a data-driven model that adequately describes the dependence of the OTWT distribution on the FSF distribution and total FSF flow rate. The remaining 3 CFD data sets (i.e., the testing data), which is approximately 10% of the reformer database, are used to validate the performance on out-of-sample predictions of the data-driven model.

3. Overview of the prediction and correction model identification scheme

The statistical-based model identification scheme is a two step (i.e., prediction and correction) procedure that is designed to derive a model for the dependence of the OTWT distribution at a fixed distance of 6.5 m away from the reformer ceiling on the FSF distribution and total FSF flow rate from the training data of the reformer CFD database. In the prediction step, the model-building process for the ith OTWT as a function of the FSF distribution and total FSF flow rate from the training data is executed independently of other reforming tubes based on Bayesian methods, i.e., Bayesian variable selection and Bayesian model averaging (BMA), with an efficient search algorithm and sparse nonlinear regression (which will be further defined in Sections 4.1, 4.2 and 4.3). The data-driven model for the OTWT of the ith reforming tube generated by the prediction step is referred to as the ith prediction model and is formulated as follows,

$$\hat{T}_i^{P,n} = \sum_{k=1}^{K_i} w_{i,k}^P \tilde{T}_{i,k}^{P,n} \quad \forall n \in \{1,\ldots,N\} \tag{1a}$$

subject to

$$\sum_{k=1}^{K_i} w_{i,k}^P = 1 \tag{1b}$$

$$\vec{F}^n = \left[F_1^n, F_2^n, \ldots, F_{96}^n\right]^T \tag{1c}$$

$$||\vec{F}^n||_1 = F_{tot}^n \tag{1d}$$

$$\tilde{T}_{i,k}^{P,n} = \sum_{g=1}^{G} \left(\vec{\alpha}_i^{kg}\right)^T \cdot f_g(\vec{F}^n) + \alpha_i^k \tag{1e}$$

where n is the index of the nth data set in the training data, N is the number of data sets in the training data, $\hat{T}_i^{P,n}$ is the BMA predicted estimate of the ith OTWT based on the nth FSF distribution ($\vec{F}^n$) and nth total FSF flow rate (F_{tot}^n), $\tilde{T}_{i,k}^{P,n}$ is the kth predicted estimate of the ith OTWT based on $\vec{F}^n$, F_{tot}^n and the kth sub-prediction model for the ith reforming tube (where this model denoted by $M_{i,k}$ is defined and developed in Section 4), K_i is the total number of sub-prediction models in the ith library, $w_{i,k}^P$ is the BMA weighting factor of $M_{i,k}$ (developed in Section 4.2), $f_g(\cdot)$ is the gth basis function in the library of transformation functions (developed in Section 4.3), G is the number of functions in the library of transformation functions (developed in Section 4.3), $\vec{\alpha}_i^{kg}$ is the empirical parameter vector of $M_{i,k}$ corresponding to $f_g(\cdot)$ (defined in Section 4.3) and α_i^k represents the estimated ambient temperature of $M_{i,k}$ (defined in Section 4.3). The prediction step in the model identification algorithm is parallelized to simultaneously create 336 prediction models, each of which corresponds to a reforming tube of the reformer, and subsequently, these prediction models are combined to create the data-driven model (i.e., the prediction model) for the OTWT distribution to describe the dependence of the OTWT distribution on the FSF distribution and total FSF flow rate. However, the prediction model for the OTWT distribution does not account for the effects of interactions between neighboring reforming tubes on the OTWT distribution, and hence, the correction step of the model identification procedure is created. In the correction step, the predicted estimate of the ith OTWT is adjusted based on information of the neighboring reforming tubes extracted from the reformer geometry and predicted OTWT distribution. The correction model for the ith reforming tube developed based on ordinary Kriging (to be further developed in Section 5) is formulated as follows,

$$\hat{T}_i^{C,n} = \sum_{\substack{j=1 \\ j \neq i}}^{336} w_{i,j}^C \hat{T}_j^{P,n} \tag{2}$$

where $\hat{T}_i^{C,n}$ is the corrected estimate of the ith OTWT and $w_{i,j}^C$ is the Kriging weighting factor of the jth reforming tube. Finally, the estimated OTWT of the ith reforming tube given $\vec{F}^n$ and F_{tot}^n

is computed as the weighted sum of the BMA predicted and corrected estimates as follows,

$$\hat{T}_i^n = w^P \hat{T}_i^{P,n} + \left(1 - w^P\right) \hat{T}_i^{C,n} \quad (3)$$

where w^P is the weighting factor of the BMA estimates, of which the optimal value is determined by leave-out-one cross validation (developed in Section 6). In the remainder of this manuscript, the development of the prediction and correction model identification algorithm is elucidated in detail, and the accompanying assumptions are explicitly stated and demonstrated to be valid using the training data.

Remark 1. The decision to investigate the OTWT distribution at a fixed distance of 6.5 m away from the reformer ceiling was originally proposed in Tran et al. (2017a). In particular, this OTWT distribution has been found to have a high average OTWT, i.e., ~99% of the maximum value of the average OTWT profile (which is a collection of averages of OTWT distributions along the reforming tube length and is considered to be invariant at the fixed total FSF flow rate), so that in a scenario that the hydrogen plant is at risk of suffering significant capital and production losses due to a suboptimal FSF distribution, this OTWT distribution is expected to have a high degree of nonuniformity (indicated by a wide temperature range and a high standard deviation), and the maximum value of the OTWT distribution at a distance of 6.5 m away from the reformer ceiling may exceed the design temperature of the reforming tube wall. This rationale suggests that the reformer can be kept in a safe operating regime by reducing the degree of nonuniformity in the OTWT distribution and preventing the maximum value of the OTWT distribution from exceeding the design temperature, and therefore, a data-driven model for the OTWT distribution was created in the development of the furnace-balancing scheme in Tran et al. (2017a) that seeks to reduce the nonuniformity. In anticipation that an application of the modeling method in this work may be to derive the data-driven model for a furnace-balancing algorithm as in Tran et al. (2017a), we consider a distance of 6.5 m away for the ceiling in this work for consistency with Tran et al. (2017a).

4. Prediction models

In this work, Bayesian methods are integrated in the algorithm for the prediction step because of the twofold advantage that these methods offer (Section 1): the Bayesian variable selection method provides a straightforward approach to identify the important regressors for each reforming tube, and BMA allows model-selection uncertainty to be accounted for in making predictions about the OTWT distribution given $\vec{F}^n$ and F_{tot}^n (Hoeting et al., 1999). In the prediction step, the prediction model for the ith OTWT is derived from the training data independently of the model-building process for the remaining OTWTs. This approach is expected to effectively reduce the total computational time devoted for constructing the prediction model for the OTWT distribution because the model-building process can be simultaneously executed for 336 reforming tubes. In this section, the important terminologies used in the description of the algorithm for the prediction step are explicitly defined, and then, the algorithm for the prediction step is discussed in the context of constructing the ith prediction model from the training data.

In the jth iteration of the prediction step,

1. Checked predictors are the burners that are in the neighborhood of the ith reforming tube, and a collection of checked predictors in the jth iteration of the prediction step is denoted as $\mathbf{S}_{iC}^j$. However, checked predictors may or may not have a significant impact on the ith OTWT.
2. Potential predictors are elements of a subset of $\mathbf{S}_{iC}^j$, which is denoted as $\mathbf{S}_{iP}^j$. Potential predictors are candidate regressors of the ith reforming tube model and will be characterized by the regressor classification layer. Potential predictors that are classified as important regressors by the regressor classification layer are added to the basis set of regressors (denoted as $\mathbf{S}_{iR}$).
3. Predictors are important regressors of the ith prediction model and are elements of $\mathbf{S}_{iR}$.
4. A sub-prediction model library of the ith reforming tube is a collection of models that allows the ith OTWT to be estimated based on $\vec{F}^n$ and F_{tot}^n.

In addition, the concept of a multistage affecting zone is introduced to facilitate the search for and classification of checked predictors. Specifically, we define the multistage affecting zone of the ith reforming tube in the jth iteration of the prediction step as a cylindrical region with a radius that is denoted R_{iZ}^j and is evaluated as follows,

$$R_{iZ}^j = R_{iZ}^0 + j\Delta R_Z \quad \forall j = 1, 2, 3, \ldots \quad (4)$$

where R_{iZ}^0=0 m is the initial radius of the multistage affecting zone of the ith reforming tube and ΔR_Z=1.5 m is the corresponding step increment. The first stage of an affecting zone of the ith reforming tube is defined as a cylindrical volume of radius R_{iZ}^j centered at the ith reforming tube inside which burners are considered to be elements of $\mathbf{S}_{iC}^j$. The second stage of an affecting zone of the ith reforming tube is defined as a hollow cylindrical volume bounded by two concentric cylinders of radii R_{iZ}^j and R_{iZ}^{j-1} inside which burners are considered to be elements of $\mathbf{S}_{iP}^j$, which is defined as follows,

$$\mathbf{S}_{iP}^j = \mathbf{S}_{iC}^j \backslash \mathbf{S}_{iC}^{j-1} \quad (5)$$

where $\mathbf{S}_{iC}^j \backslash \mathbf{S}_{iC}^{j-1}$ is defined as all elements that are in $\mathbf{S}_{iC}^j$ but are not in $\mathbf{S}_{iC}^{j-1}$.

The algorithm for the prediction step consists of two layers, which are referred to as the regressor collection (Fig. 2) and regressor classification (Fig. 3) layers, respectively. The regressor collection layer is composed of five processes represented by five rectangular boxes, i.e., the location identifier, checked predictor identifier, potential predictor identifier, sparse nonlinear regression and termination checker. The layer is structured to have two sequential modes, which aim to obtain the default $\mathbf{S}_{iR}$ and to search for potential predictors to be used as inputs for the regressor classification layer, respectively. We note that the second mode is proposed to avoid having to re-look at all the potential predictors that have already been evaluated. The search algorithm for the regressor collection layer is developed based on the inverse square law for thermal radiation. It makes use of two fundamental guidelines, namely, the burners separated from the ith reforming tube by a significant distance should not be considered as potential predictors, and the nearby burners have greater influence on the ith OTWT than other burners that are situated farther away from the ith reforming tube, to reduce the

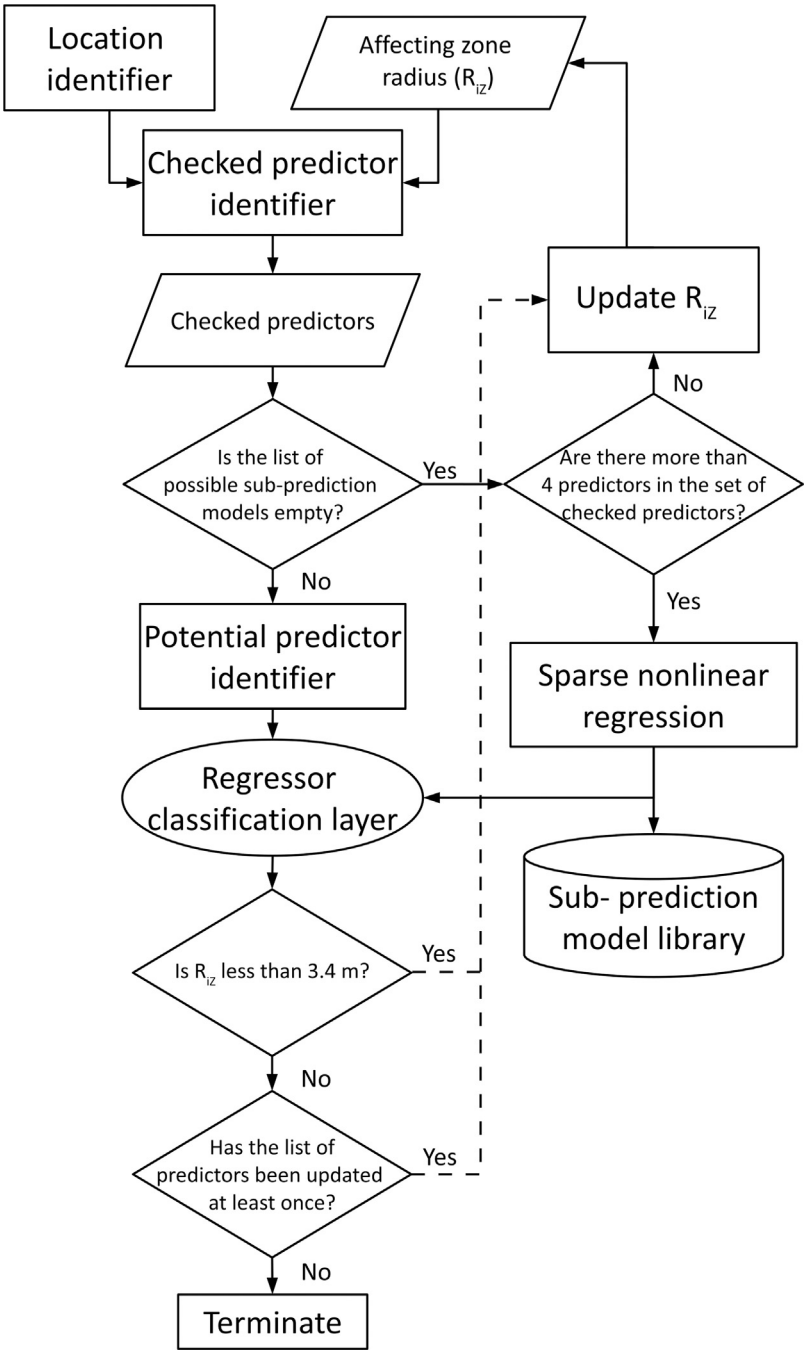

Fig. 2 – Flowchart of the regressor collection layer in the prediction step.

regressor space and define a unique searching path, which will be elaborated later in this section. The search algorithm is the most critical development of the present work and allows the model identification based on the Bayesian framework to be computationally efficient. This is because the search algorithm selectively creates a small number of sub-prediction models for the ith OTWT to be used by the BMA in making predictions as opposed to 2^{96} ($\approx 7.9 \times 10^{28}$) sub-prediction models, which would be generated by the alternative, i.e., the exhaustive search algorithm, which if it was implemented would make the model identification scheme computationally infeasible. In this work, the distance between the ith reforming tube and jth burner is denoted as d_{ij} and is calculated as the Euclidean distance between the projections of their centroids on a horizontal plane, and we assume that the default $\mathbf{S}_{iR}$ of the ith reforming tube consists of the four nearest burners.

In the first iteration of the prediction step, the regressor collection layer is operated under the first mode to obtain the default $\mathbf{S}_{iR}$. Initially, it begins with the location identifier, which calculates the relative distance from the ith reforming tube to burners inside the reformer, and simultaneously computes the radius of the affecting zone (denoted as R^1_{iZ}) of the ith reforming tube according to Eq. (4). Next, the checked predictor identifier uses the information about the relative location of the ith reforming tube to create a virtual reformer geometry, which consists of the ith reforming tube and 96 burners as shown in Fig 4 , and uses R^1_{iZ} to create the first stage of the affecting zone as shown in Fig. 5. The checked predictor identifier, then, uses the virtual reformer geometry and first-stage affecting zone to generate $\mathbf{S}^1_{iC}$, which contains information (i.e., distance to the ith reforming tube, burner IDs and FSF flow rates) of the checked predictors. In the first iteration, the library of ith sub-prediction models is expected to

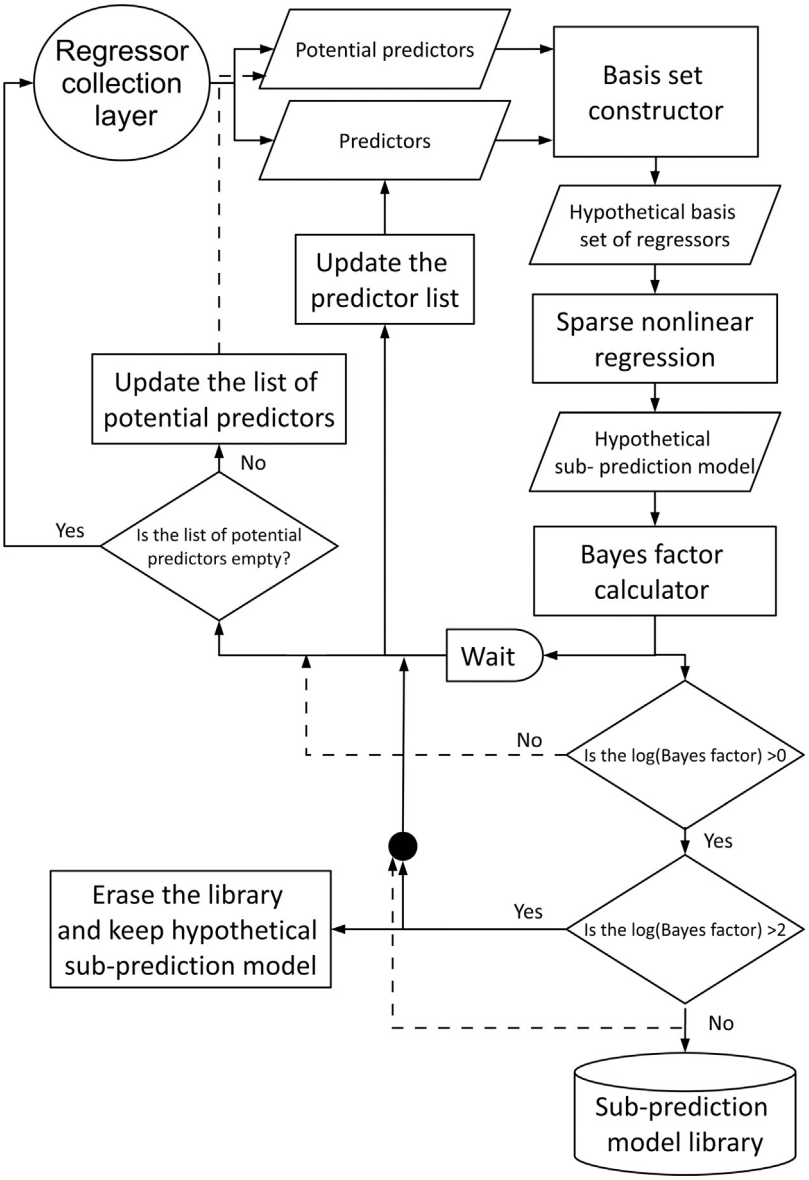

Fig. 3 – Flowchart of the regressor classification layer in the prediction step.

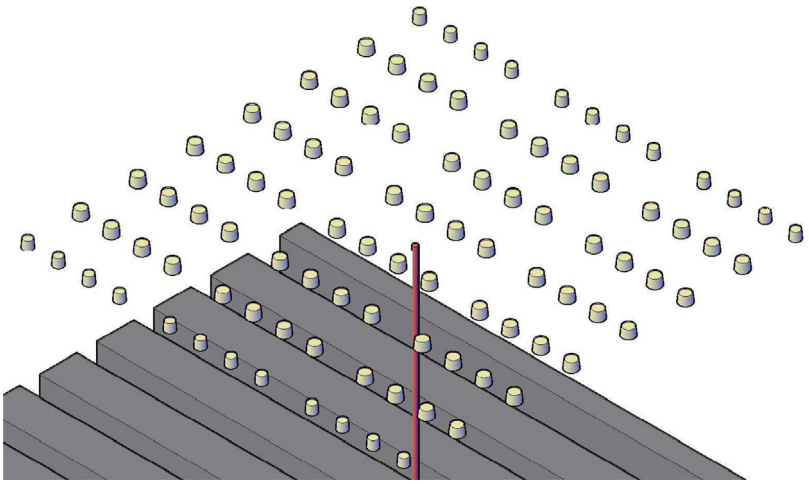

Fig. 4 – A virtual reformer geometry that is created based on the information generated by the location identifier in the regressor collection layer and consists of the ith reforming tube represented by a cylindrical tube and 96 burners represented by the frustum cones.

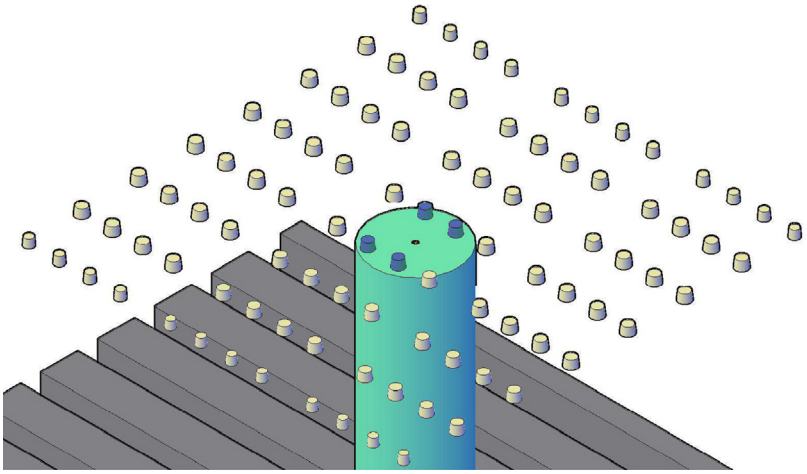

Fig. 5 – The first stage of the multistage affecting zone that is created based on R^1_{iZ} and the virtual reformer geometry in the first iteration of the prediction step. In this figure, the checked predictors in $\mathbf{S}^1_{iC}$ are represented by the four shaded frustum cones and can be visually identified.

be empty, and therefore, the cardinality of $\mathbf{S}^1_{iC}$ is evaluated. If the cardinality of $\mathbf{S}^1_{iC}$ is less than or equal to four, the prediction step is instructed to terminate the first iteration and to execute the next iteration. The second iteration of the prediction step begins to recompute the radius of the affecting zone (denoted as R^2_{iZ}) of the ith reforming tube according to Eq. (4). As in the first iteration of the prediction step, the checked predictor identifier in the second iteration uses R^2_{iZ} to create the updated first stage of the affecting zone and, then, uses it and the virtual reformer geometry created in the first iteration to generate $\mathbf{S}^2_{iC}$. Thereafter, the cardinality of $\mathbf{S}^2_{iC}$ is also evaluated since the library of ith sub-prediction models is expected to remain empty, and if its result is still less than or equal to four, the prediction step is again instructed to terminate the second iteration and to execute the next iteration. This procedure is repeated until the number of elements in the checked predictor set of the ith reforming tube is strictly greater than four.

This discussion is continued with the assumption that in the jth iteration of the prediction step, the cardinality of $\mathbf{S}^j_{iC}$ is greater than four as shown in Fig. 6. Initially, the four nearest burners to the ith reforming tube in $\mathbf{S}^j_{iC}$ are used to create the default $\mathbf{S}_{iR}$, which is utilized by the sparse nonlinear regression algorithm to create the default sub-prediction model for the ith OTWT. Then, the model is stored in the ith library of sub-prediction models and is assigned an index of 1. The remaining elements in $\mathbf{S}^j_{iC}$, i.e., $\mathbf{S}^j_{iC} \backslash \mathbf{S}_{iR}$, are considered to be potential predictors, which are elements of $\mathbf{S}^j_{iP}$, and are used as inputs for the regressor classification layer, which is the secondary layer of the prediction step algorithm.

The regressor classification layer is structured based on the Bayesian variable selection framework to systematically update $\mathbf{S}_{iR}$ using a given potential predictor set and to selectively create additional sub-prediction models for the ith OTWT. A brute-force Bayesian variable selection would require all possible hypothetical basis sets of regressors to be created from $\mathbf{S}^j_{iP}$ and $\mathbf{S}_{iR}$, followed by using the sparse nonlinear regression to generate all possible hypothetical sub-prediction models, from which the important sub-prediction models would be selected. This approach is expected to be computationally intensive and will not be implemented. Instead, the layer is designed to exploit Occam's window (Madigan and Raftery, 1994) and the two fundamental guidelines that are used in the regressor collection layer to avoid generating all possible hypothetical basis sets of regressors from $\mathbf{S}^j_{iP}$ and $\mathbf{S}_{iR}$ in the process of identifying the important sub-prediction models by assessing the impact of each element in the potential predictor set on the goodness of fit of the sub-prediction model in a step-wise fashion. Specifically, in the jth iteration of the regressor collection layer, the regressor classification layer starts out with the basis set constructor, which is programmed to strategically select an element in $\mathbf{S}^j_{iP}$ followed by adding it to the existing $\mathbf{S}_{iR}$ to create a hypothetical basis set of regressors (denoted as $\mathbf{S}^h_{iR}$), which is subsequently used by the sparse nonlinear regression to generate the corresponding hypothetical sub-prediction model. If the hypothetical sub-prediction model can explain the training data significantly better than the sub-prediction model created with $\mathbf{S}_{iR}$ (in a sense related to Bayes factors to be made clear in Section 4.1), it is either stored in or used to replace the ith library of sub-prediction models based on the supporting evidence for the model, and $\mathbf{S}^h_{iR}$ is used to replace $\mathbf{S}_{iR}$; otherwise, the hypothetical sub-prediction model and $\mathbf{S}^h_{iR}$ are discarded. In this work, the procedure of selecting an element in $\mathbf{S}^j_{iP}$ encoded in the basis set constructor begins by ranking elements in $\mathbf{S}^j_{iP}$ in the order of increasing distance to the ith reforming tube such that the potential predictor separated from the ith reforming tube by the shortest distance is considered to have the highest rank (i.e., the 1st rank), and the potential predictors separated from the ith reforming tube by an equal distance are considered to have the same rank. Then, the rank of each element in $\mathbf{S}^j_{iP}$ is utilized by the basis set constructor as an indicator for the order that the element is selected to construct $\mathbf{S}^h_{iR}$. Specifically, the basis set constructor starts with the high-ranking elements in $\mathbf{S}^j_{iP}$ because the burners that are situated closer to the ith reforming tube are expected to have greater influence on the ith OTWT than other burners that are situated farther away and are more likely to be considered as predictors; therefore, $\mathbf{S}^h_{iR}$ is more likely to be accepted. When multiple elements of $\mathbf{S}^j_{iP}$ have the same rank, the order that these elements are selected to construct $\mathbf{S}^h_{iR}$ is trivial because our proposed model (Eq. (1a)) does not contain any cross term, which suggests that burners are assumed to independently interact with the ith reforming tube. In this case, they are iteratively selected one by one to construct $\mathbf{S}^h_{iR}$ before the potential predictors in the next lower rank are selected. This procedure is repeated until

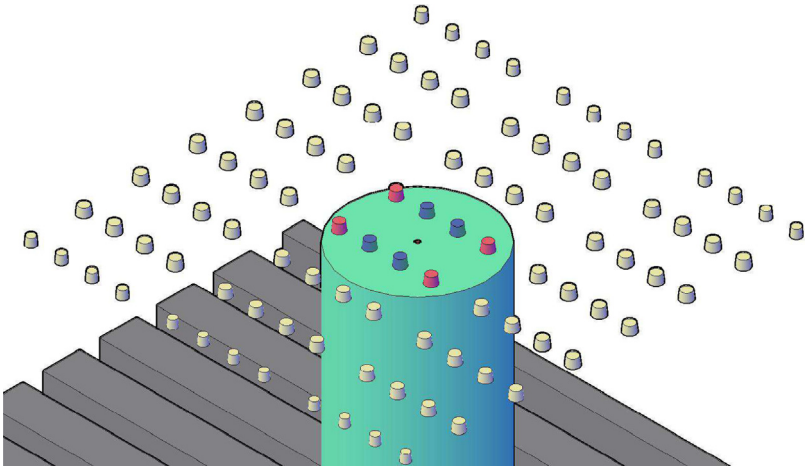

Fig. 6 – The first stage of the multistage affecting zone that is created based on R_{iZ}^{j} and the virtual reformer geometry in the *j*th iteration of the prediction step.

all elements in $\mathbf{S}_{iP}^{j}$ are considered, then the updated $\mathbf{S}_{iR}$ is sent back to the regressor collection layer.

Subsequently, in the regressor collection layer, the termination checker reviews R_{iZ}^{j} and the updated $\mathbf{S}_{iR}$ created in the regressor classification layer to decide if the prediction step should be terminated. Specifically, when R_{iZ}^{j} is less than the critical radius of the multistage affecting zone (denoted as R_{Z}^{c}), the termination checker always instructs the prediction step to execute the next iteration. In this work, R_{Z}^{c} is chosen to be 3.4 m to prevent the prediction step from exploring unnecessary regions in the regressor space because Tran et al. (2017a) shows that the faraway burners, which are defined as those separated from the ith reforming tube by a distance greater than 3.4 m, are expected to have small impact on the ith OTWT by demonstrating that data-driven models which include the faraway burners in the basis set of regressors have similar goodness of fit to that of the data-driven model that excludes the faraway burners from the basis set of regressors. On the contrary, when R_{iZ}^{j} becomes greater than or equal R_{Z}^{c}, the termination checker only allows the prediction step to execute the next iteration if at least one potential predictor in $\mathbf{S}_{iP}^{j}$ is added to $\mathbf{S}_{iR}$. This design of the termination checker is proposed to account for the potential influence of the furnace-side flow pattern on the OTWT distribution that allows the faraway burners to have long range effects on the ith OTWT and to simultaneously prevent the prediction step from exploring unnecessary regions in the regressor space.

After the *j*th iteration of the prediction step, the regressor collection layer is operated under the second mode to exclusively search for potential predictors to be used as inputs for the regressor classification layer. Initially, it begins to compute R_{iZ}^{j+1} of the ith reforming tube according to Eq. (4), which is used by the checked predictor identifier in the $(j+1)$th iteration to create the first stage of the affecting zone as done in the previous iterations of the prediction step. Next, the checked predictor identifier uses the first stage affecting zone and virtual reformer geometry created in the first iteration to generate $\mathbf{S}_{iC}^{j+1}$. In the $(j+1)$th iteration of the prediction step, because the library of ith sub-prediction models is no longer empty, the potential predictor identifier is executed for the first time; this step is the key difference between the first and second modes of the regressor collection layer. Specifically, the potential predictor identifier uses R_{iZ}^{j+1} and R_{iZ}^{j} to create the second stage of the affecting zone as shown Fig. 7 and, then, uses it and the virtual reformer geometry to generate $\mathbf{S}_{iP}^{j+1}$, which contains information of the potential predictors. After that, $\mathbf{S}_{iP}^{j+1}$ is used as an input for the regressor classification layer, which updates $\mathbf{S}_{iR}$. Finally, the termination checker utilizes R_{iZ}^{j+1} and the updated $\mathbf{S}_{iR}$ in the decision-making process of concluding the prediction step. This procedure is repeatedly executed until the radius of the multistage affecting zone is greater than or equal R_{Z}^{c}, and all elements in the potential predictor set of the ith reforming tube are rejected. Finally, BMA is utilized to determine the weighting factor for each member in the ith library of sub-prediction models (see Section 4.2).

4.1. *Bayesian variable selection*

In the present work, a Bayesian variable selection method with the search algorithm developed based on Occam's window and theories of thermal radiation is used to identify the predictors for the ith OTWT based on the training data and to simultaneously create a collection of sub-prediction models that can be used to explain the dependence of the ith OTWT on the FSF distribution and total FSF flow rate. At each step, an element in the set of potential predictors is added to $\mathbf{S}_{iR}$ to create a $\mathbf{S}_{iR}^{h}$, which is subsequently used to create the corresponding hypothetical sub-prediction model as detailed earlier in this section (the sparse nonlinear regression algorithm used to develop the hypothetical sub-prediction models will be described in Section 4.3). Then, the layer quantitatively assesses the goodness of fit of two competing data-driven models of the ith reforming tube, i.e., the sub-prediction model (denoted as $M_{i,k}$ where k is the model index in the ith library of sub-prediction models) created with $\mathbf{S}_{iR}$ and the hypothetical sub-prediction model denoted as $M_{i,h}$ created with $\mathbf{S}_{iR}^{h}$, by using the ratio of posterior model probabilities as the metric to determine if a potential predictor of interest can be classified as an important regressor. Specifically, the posterior model probability of $M_{i,k}$, which is denoted as $pr(M_{i,k}|\mathbf{T}_i)$ to represent the probability that $M_{i,k}$ is the true model for the dependence of the ith OTWT on the FSF distribution and total FSF flow rate after observing the training data, is computed as follows,

$$pr\left(M_{i,k}|\mathbf{T}_i\right) = \frac{pr\left(\mathbf{T}_i|M_{i,k}\right) pr\left(M_{i,k}\right)}{\sum_{l=1}^{K_i} pr\left(\mathbf{T}_i|M_{i,l}\right) pr\left(M_{i,l}\right)} \tag{6}$$

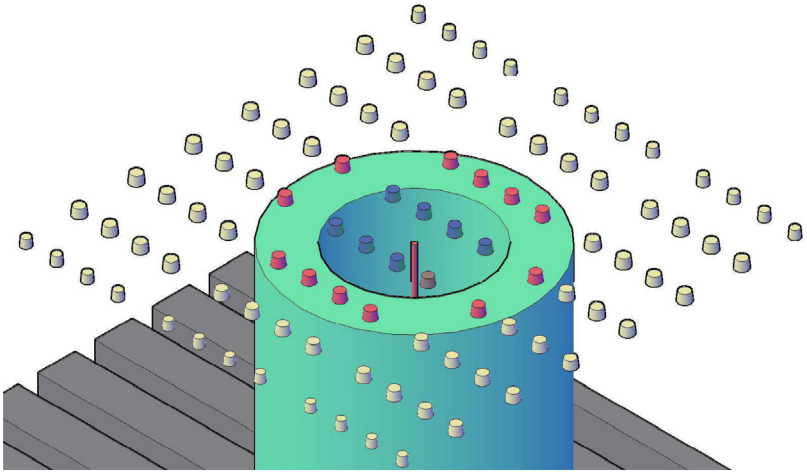

Fig. 7 – The first stage of the multistage affecting zone that is created based on R_{iZ}^{j+1}, R_{iZ}^{j} and the virtual reformer geometry in the (j + 1)th iteration of the prediction step.

where $\mathbf{T}_i = \left\{T_i^1, T_i^2, \ldots, T_i^N\right\}$ is the collection of all $N = 18$ training data sets, T_i^j is the ith OTWT extracted from the jth training data set, K_i is the total number of sub-prediction models in the ith library, $pr(M_{i,k})$ is the prior model probability of $M_{i,k}$ and $pr(\mathbf{T}_i|M_{i,k})$ is the model evidence in favor of $M_{i,k}$. It is worth noting that $pr(M_{i,k})$ reflects our beliefs that $M_{i,k}$ is the true model for the dependence of the ith OTWT on the FSF distribution and total FSF flow rate before observing the training data. If information about the true model for the ith reforming tube is available (e.g., the number of predictors in the true model is known), the prior model probability distribution could be designed to incorporate such information by assigning the sub-prediction models that have the same number of predictors as the true model with a higher prior model probability. However, this is typically not the case in practice. Therefore, the prior model probability distribution is chosen to be noninformative, which assigns equal prior probability to all sub-prediction models in the ith library and allows conclusions about the true model to be drawn directly from data. The noninformative prior model probability distribution for the ith reforming tube is designed as follows,

$$pr(M_{i,j}) = \frac{1}{K_i} \quad \forall j = 1, \ldots, K_i. \tag{7}$$

Hence, the ratio of posterior model probabilities between $M_{i,h}$ and $M_{i,k}$ can be evaluated as follows,

$$\frac{pr\left(M_{i,h}|\mathbf{T}_i\right)}{pr\left(M_{i,k}|\mathbf{T}_i\right)} = \frac{pr\left(\mathbf{T}_i|M_{i,h}\right)pr\left(M_{i,h}\right)}{pr\left(\mathbf{T}_i|M_{i,k}\right)pr\left(M_{i,k}\right)} = \frac{pr\left(\mathbf{T}_i|M_{i,h}\right)}{pr\left(\mathbf{T}_i|M_{i,k}\right)} = B_{i,hk} \tag{8}$$

where $B_{i,hk}$ is defined as a Bayes factor for $M_{i,h}$ against $M_{i,k}$. Eq. (8) suggests that $B_{i,hk}$ and the ratio of posterior model probabilities between $M_{i,h}$ and $M_{i,k}$ under the assumption of the noninformative prior model probability distribution (Eq. (7)) are equivalent, and therefore, the value of $B_{i,hk}$ can be used as the quantitative evidence in favor of incorporating the potential predictor into $\mathbf{S}_{iR}$. However, computing $B_{i,hk}$ is a nontrivial task because there is no analytical closed-formed expression for computing $pr(\mathbf{T}_i|M_{i,k})$. Specifically, $pr(\mathbf{T}_i|M_{i,k})$ is computed by integrating over all possible values of the parameters of $M_{i,k}$ as follows,

$$pr\left(\mathbf{T}_i|M_{i,k}\right) = \int pr\left(\mathbf{T}_i|M_{i,k}, \vec{\alpha}_i^k\right) pr\left(\vec{\alpha}_i^k|M_{i,k}\right) d\vec{\alpha}_i^k \tag{9}$$

where $\vec{\alpha}_i^k = [\vec{\alpha}_i^{k1}, \ldots, \vec{\alpha}_i^{kG}, \alpha_i^k]^T$ is the parameter vector of $M_{i,k}$ (see Section 4.3), $pr(\vec{\alpha}_i^k|M_{i,k})$ is the prior probability density of $\vec{\alpha}_i^k$ and $pr(\mathbf{T}_i|M_{i,k}, \vec{\alpha}_i^k)$ is the likelihood function of $\mathbf{T}_i$. It is important to note that the likelihood function of $\mathbf{T}_i$ is defined as the joint probability density function of $T_i^1, T_i^2, \ldots, T_i^N$ and is dependent on $\vec{\alpha}_i^k$ of $M_{i,k}$ as follows,

$$pr\left(\mathbf{T}_i|M_{i,k}, \vec{\alpha}_i^k\right) = \prod_{j=1}^{N} pr\left(T_i^j|M_{i,k}, \vec{\alpha}_i^k\right) \tag{10}$$

where $pr(T_i^j|M_{i,k}, \vec{\alpha}_i^k)$ is the probability density function of T_i^j. In addition, the probability density function of T_i^j is assumed to be computed as follows,

$$pr(T_i^j|M_{i,k}, \vec{\alpha}_i^k) = \frac{1}{\sqrt{2\pi(\sigma_i^j)^2}} \exp\left(-\frac{(T_i^j - \hat{T}_i^{P,j})^2}{2(\sigma_i^j)^2}\right) \tag{11}$$

where σ_i^j is the standard deviation of the noise in the ith OTWT in the jth training data set. Due to challenges in computing $B_{i,hk}$, many published works in the Bayesian statistics literature center on proposing methods to compute $pr(\mathbf{T}_i|M_{i,k})$ numerically, e.g., Markov Chain Monte Carlo, or to approximate it with an acceptable accuracy, e.g., Laplace approximation, Bayesian information criterion (BIC) approximation and maximum likelihood estimator (MLE) approximation (Fragoso and Neto, 2015). In this work, the BIC approximation is favored for two reasons: the BIC approximation is expected to provide a good approximation of $\log B_{i,hk}$ for linear models (Fragoso and Neto, 2015) and allows us to avoid making assumptions about the prior probability density of $\vec{\alpha}_i^k$, which allows the decision to incorporate the potential predictor into $\mathbf{S}_{iR}$ to be made based entirely on the training data (Kass and Raftery, 1995). Under the BIC approximation, $\log B_{i,hk}$ is computed as follows,

$$\log B_{i,hk} = \left[\log\left(pr(\mathbf{T}_i|M_{i,h}, \hat{\vec{\alpha}}_i^h)\right) - \log\left(pr(\mathbf{T}_i|M_{i,k}, \hat{\vec{\alpha}}_i^k)\right)\right] - \frac{1}{2}(d_h - d_k)\log(N) \tag{12}$$

where $\hat{\vec{\alpha}}_i^k$ is the MLE of $\vec{\alpha}_i^k$ and d_k is the number of elements in the basis set of regressors of $M_{i,k}$ (see Section 4.3). It is important to note that the first term in Eq. (12) can be interpreted as the assessment of the goodness of fit between two sub-prediction models, and the second term can be interpreted as the penalty for using a model with higher complexity. The approximated value of $\log B_{i,hk}$ is the evidence in favor of $M_{i,h}$ extracted from the training data, and higher values of $\log B_{i,hk}$ imply that the training data provides more evidence in support of $M_{i,h}$ and against $M_{i,k}$. Therefore, the value of $\log B_{i,hk}$ can be used in the decision-making process for determining whether to incorporate a potential predictor into $\mathbf{S}_{iR}$. In the present work, the task of interpreting the value of $\log B_{i,hk}$ is especially challenging because the improvement in the model goodness of fit by incorporating an additional term into $\mathbf{S}_{iR}$ is expected to be underestimated when a small number of training data sets is used. Hence, guidelines for interpreting Bayes factors proposed in the Bayesian statistics literature, namely, twice the natural logarithm of the Bayes factor suggested by Kass and Raftery (1995), half unit on the logarithm of the Bayes factor suggested by Jeffreys (1935), Occam's window with the lower bound of 1/20 and upper bound of 20 suggested by Raftery et al. (1996) and Occam's window with the lower bound of 1/20 and upper bound of 1 suggested by Madigan et al. (1994), were reviewed. In this work, Jeffreys' interpretation of the Bayes factor (Jeffreys, 1935) is selected and is tailored to account for the impact of using the small number of training data sets on the value of $\log B_{i,hk}$ as follows: we reason that when an insignificant regressor is incorporated into $\mathbf{S}_{iR}$ to create $\mathbf{S}_{iR}^h$, the two competing sub-prediction models are expected to have a similar goodness of fit, which causes the first term in Eq. (12) to be approximately zero, and the dimension of $\mathbf{S}_{iR}^h$ is greater than that of $\mathbf{S}_{iR}$, which causes the second term in Eq. (12) to be negative due to the increased model complexity. Therefore, a negative $\log B_{i,hk}$ can be viewed as a strong evidence against $M_{i,h}$, which implies that the ith OTWT is independent of the potential predictor. In this case, the regressor classification layer is instructed to discard $M_{i,h}$ and to dismiss the attempt to incorporate the potential predictor into $\mathbf{S}_{iR}$. On the contrary, when $\log B_{i,hk}$ is nonnegative and greater than 2.0, it can be viewed as a strong evidence against $M_{i,k}$, which indicates that the training data provides significant evidence in support of $M_{i,h}$ and against $M_{i,k}$ because $M_{i,h}$ has a significantly higher goodness of fit than $M_{i,k}$. In this case, the layer is instructed to discard $M_{i,k}$ along with its nested sub-prediction models (i.e., the sub-prediction models created and stored throughout the prediction step in the process of generating $M_{i,k}$) in the ith sub-prediction model library followed by accepting the attempt to incorporate the potential predictor into $\mathbf{S}_{iR}$ and storing $M_{i,h}$ in the ith library as the first sub-prediction model. It is worth noting that this action is inspired by the two principles of Occam's window, which suggests that when a sub-prediction model has a significantly lower goodness of fit than its competitor, it should be discredited (Fragoso and Neto, 2015). Furthermore, in the remaining case, i.e., when $\log B_{i,hk}$ is nonnegative and less than or equal to 2.0, it can be viewed as a weak evidence against $M_{i,k}$, which indicates that the training data suggests there is weak evidence in support of $M_{i,h}$ and against $M_{i,k}$ because $M_{i,h}$ has a slightly higher goodness of fit than $M_{i,k}$; however, this evidence is insufficient to discredit $M_{i,k}$. In this case, the layer is instructed to accept the attempt to incorporate the potential predictor into $\mathbf{S}_{iR}$ and store $M_{i,h}$ in the ith library as the $(k+1)th$ sub-prediction model.

Remark 2. Although the reformer data (i.e., the training data and testing data) is generated by simulating the high-fidelity reformer CFD model, it is expected to exhibit some stochastic behavior because in all simulations of the reformer CFD model, the final global normalized residuals have small nonzero values which suggests that the reformer data fluctuates around the true steady-state. Furthermore, even though the training data and testing data are expected to exhibit some stochastic behavior, each reformer data set only has one realization, which prevents the magnitude of the noise from being estimated from the reformer data. In this work, the noise in OTWTs of all reforming tubes in the jth training data set is assumed to be normal, independent and identically distributed with a mean of zero and a standard deviation of σ^j, which is assumed to be 5–10% of the corresponding OTWT range and is approximated as follows,

$$\sigma_i^j = \sigma^j = \eta_\sigma \left(\max(\vec{T}^j) - \min(\vec{T}^j)\right) \quad \forall i = 1, \ldots, 336 \tag{13}$$

where $\max(\vec{T}^j)$ and $\min(\vec{T}^j)$ represent the maximum and minimum OTWT in the jth training data set, respectively, and η_σ is a hyperparameter, which has a random value between 0.05 and 0.1 generated by our in-house uniform random number generator.

4.2. *Bayesian model averaging*

In the present work, Bayesian model averaging is used to account for model uncertainty in making predictions about the ith OTWT based on the FSF distributions and total FSF flow rates. In this section, it is assumed that K_i sub-prediction models for the ith reforming tube, that capture the dependence of the ith OTWT on the FSF distribution and total FSF flow rate reasonably well, are derived from the training data based on the Bayesian variable selection approach, and T_i^* is to be estimated given $\vec{F}^*$ and F_{tot}^*, which are the unexplored FSF distribution and total FSF flow rate, respectively, of the reformer. Therefore, the posterior mean of T_i^*, which represents the most likely value of T_i^* that is expected to be observed based on the training data, can be computed as follows,

$$E\left(T_i^* | \mathbf{T}_i, \vec{F}^*, F_{tot}^*\right) = \sum_{k=1}^{K_i} pr\left(M_{i,k} | \mathbf{T}_i\right) E\left(T_i^* | \mathbf{T}_i, \vec{F}^*, F_{tot}^*, M_{i,k}\right) \tag{14}$$

where $E(T_i^* | \mathbf{T}_i, \vec{F}^*, F_{tot}^*, M_{i,k})$ represents the posterior mean of T_i^* when $M_{i,k}$ is assumed to be the true model for the ith reforming tube. It is recognized that Eqs. (1a) and (14) resemble one another, and therefore, it can be inferred that $pr(M_{i,k}|\mathbf{T}_i)$ is the BMA weighting factor of $M_{i,k}$,

$$pr\left(M_{i,k} | \mathbf{T}_i\right) = w_{i,k}^P. \tag{15}$$

In addition, $pr(M_{i,k}|\mathbf{T}_i)$ can be expressed in terms of Bayes factors by dividing the numerator and denominator of Eq. (6) by the posterior probability of an arbitrary model (e.g., $M_{i,1}$) chosen from the ith sub-prediction model library,

$$pr\left(M_{i,k} | \mathbf{T}_i\right) = \frac{B_{i,k1}}{\sum_{l=1}^{K_i} B_{i,l1}}. \tag{16}$$

Therefore, the BMA weighting factor of $M_{i,k}$ can be approximated by the BIC approximation (Eq. (12)) using the training data, which allows $\hat{T}_i^{P,*}$ to be computed as a weighted average

of $\tilde{T}_{i,k}^{P,*}$ generated by sub-prediction models of the ith reforming tube (where the superscript star in place of n signifies that the predictions are associated with the unexplored operating condition).

4.3. *Sparse nonlinear regression*

In the present work, sparse nonlinear regression with maximum likelihood estimation (MLE) is used to construct data-driven models of the ith reforming tube that can be used to explain nonlinearities in the dependence of the ith OTWT on the FSF distribution and total FSF flow rate. The idea is inspired by the work in Brunton et al. (2016) which uses sparse nonlinear regression to extract governing equations of nonlinear systems from observed data. Additionally, Brunton et al. (2016) notes that governing equations of many systems typically consist of a few nonzero terms, which makes sparse nonlinear regression an especially appealing approach. To derive sub-prediction models that can be used to describe the dependence of the ith OTWT on the FSF distribution and total FSF flow rate from the reformer data, a library of linear and nonlinear transformations is designed based on expected interactions between the ith OTWT and its predictors. For instance, if the jth burner is a predictor of the ith OTWT, any nonzero value of the jth burner FSF flow rate is expected to cause the ith OTWT to rise above the ambient temperature, and when the FSF flow rate of the burner is increased, the ith OTWT is also expected to increase. Therefore, the transformations are restricted to nonnegative, monotonically increasing and continuously differentiable classes of functions. In this work, the library of transformation functions is proposed to consist of monomial, root and exponential functions as follows,

$$f_1\left(\vec{F}^n\right) = \left[F_1^n, F_2^n, \ldots, F_{96}^n\right]^T \tag{17a}$$

$$f_2\left(\vec{F}^n\right) = \left[(F_1^n)^2, (F_2^n)^2, \ldots, (F_{96}^n)^2\right]^T \tag{17b}$$

$$f_3\left(\vec{F}^n\right) = \left[(F_1^n)^3, (F_2^n)^3, \ldots, (F_{96}^n)^3\right]^T \tag{17c}$$

$$f_4\left(\vec{F}^n\right) = \left[\sqrt[2]{F_1^n}, \sqrt[2]{F_2^n}, \ldots, \sqrt[2]{F_{96}^n}\right]^T \tag{17d}$$

$$f_5\left(\vec{F}^n\right) = \left[\sqrt[3]{F_1^n}, \sqrt[3]{F_2^n}, \ldots, \sqrt[3]{F_{96}^n}\right]^T \tag{17e}$$

$$f_6\left(\vec{F}^n\right) = \left[\sqrt[4]{F_1^n}, \sqrt[4]{F_2^n}, \ldots, \sqrt[4]{F_{96}^n}\right]^T \tag{17f}$$

$$f_7\left(\vec{F}^n\right) = \left[\sqrt[5]{F_1^n}, \sqrt[5]{F_2^n}, \ldots, \sqrt[5]{F_{96}^n}\right]^T \tag{17g}$$

$$f_8\left(\vec{F}^n\right) = \left[\exp\left(F_1^n\right), \exp\left(F_2^n\right), \ldots, \exp\left(F_{96}^n\right)\right]^T \tag{17h}$$

where F_j^n is the FSF flow rate of the jth burner from the nth data set. Next, the library of transformations is used to formulate the generalized sub-prediction model for the ith reforming tube, which is assumed to be dependent on all 96 burners and in that case is denoted by $\tilde{T}_i^{P,n}$, as follows,

$$\tilde{T}_i^{P,n} = \sum_{g=1}^{8} \left(\vec{\alpha}_i^g\right)^T \cdot f_g\left(\vec{F}^n\right) + \alpha_i \tag{18}$$

where $\vec{\alpha}_i^g \in \mathrm{IR}^{96\times 1}$ is the generalized parameter vector associated with the gth transformation function, where $g = 1, \ldots, 8$ (Eq. (17)), and $\alpha_i \in \mathrm{IR}$ represents ambient air temperature. Then, a basis set of regressors (i.e., $\mathbf{S}_{iR}$) of the ith reforming tube created by the Bayesian variable selection method developed in the regressor classification layer is utilized in the model-building process which integrates information about the reformer layout and the knowledge that thermal radiation is expected to be the dominant mode of heat transfer in the reformer into the sub-prediction model (i.e., $M_{i,k}$) to set parameters associated with the burners that are not contained in S_{iR} to zero, effectively reducing the number of terms in Eq. (18) by a factor of $96/N_{iR}$, where N_{iR} is the cardinality of $\mathbf{S}_{iR}$, leading to significant reduction in the computational cost. Therefore, $M_{i,k}$ can be written as follows,

$$\tilde{T}_{i,k}^{P,n} = \sum_{g=1}^{8} \left(\vec{\alpha}_i^{kg}\right)^T \cdot f_g\left(\vec{F}^n|_{S_{iR}}\right) + \alpha_i^k \tag{19}$$

where $\vec{F}^n|_{S_{iR}} \in \mathrm{IR}^{N_{iR}\times 1}$ is a vector in the design matrix of $M_{i,k}$. The nonlinear sparse regression with MLE is formulated as a constrained optimization problem that is structured based on the L1 regularization technique (i.e., LASSO) because LASSO is known to shrink parameters associated with the irrelevant transformations to zero which further reduces the number of terms in the sub-prediction model (Eq. (18)) of the ith reforming tube. In addition, theories of thermal radiation discussed in Section 1 are integrated into the sub-prediction model by means of equality and inequality constraints (Eqs. (21a)–(21c)) in the optimization problem. The formulation for the sparse nonlinear regression with MLE is proposed as follows,

$$\min_{\substack{\alpha_i^k \in [298.15,\, 348.15] \\ \alpha_{ij}^{kg} \in [0,\infty)}} \sum_{n=1}^{N} \frac{\left(T_i^n - \tilde{T}_{i,k}^{P,n}\right)^2}{2\left(\sigma_i^n\right)^2} + \lambda_i \sum_{g=1}^{8} ||\vec{\alpha}_i^{kg}||_1 \tag{20}$$

subject to

$$\sum_{g=1}^{8} \alpha_{il}^{kg} f_g\left(\bar{F}^0\right) = \sum_{g=1}^{8} \alpha_{ij}^{kg} f_g\left(\bar{F}^0\right) \quad \text{if} \quad d_{il} = d_{ij} \tag{21a}$$

$$\sum_{g=1}^{8} \alpha_{il}^{kg} f_g\left(\bar{F}^0\right) \geq \left(\frac{d_{ij}}{d_{il}}\right)^{\beta_l} \sum_{g=1}^{8} \alpha_{ij}^{kg} f_g\left(\bar{F}^0\right) \quad \text{if} \quad d_{il} < d_{ij} \tag{21b}$$

$$\sum_{g=1}^{8} \alpha_{il}^{kg} f_g\left(\bar{F}^0\right) \leq \left(\frac{d_{ij}}{d_{il}}\right)^{\beta_u} \sum_{g=1}^{8} \alpha_{ij}^{kg} f_g\left(\bar{F}^0\right) \tag{21c}$$

$$\bar{F}^0 = \frac{F_{tot}^{typ}}{96} \tag{21d}$$

where l and j are indices of burners that are elements in $\mathbf{S}_{iR}$, α_{ij}^{kg} is the parameter in $M_{i,k}$ associated with the gth transformed FSF flow rate of the jth burner, $\alpha_i^k \in [298.15, 348.15]$ represents an ambient temperature parameter in $M_{i,k}$ and its typical range, λ_i is the tuning parameter in LASSO of the ith reforming tube, d_{ij} and d_{il} are distances from the ith reforming tube to the jth and lth burners, respectively, $\beta_u = 6.0$ and $\beta_l = 1.0$ are hyperparameters of the constraints and are chosen by trial and error, and F_{tot}^{typ} is the total FSF flow rate typically reported

in the SMR literature (Tran et al., 2017b). The constraints of Eqs. (21a)–(21c) are formulated in an effort to integrate the inverse square law of thermal radiation into $M_{i,k}$. Specifically, burners, which are separated from the ith reforming tube by an equal distance and are supplied with the same FSF flow rate, are expected to have the same impact on the ith OTWT as shown in the equality constraint (Eq. (21a)), whereas burners which are separated from the ith reforming tube by different distances but are supplied with the same FSF flow rate are expected to have different impacts on the ith OTWT as shown in the inequality constraint (Eq. (21b)). In this particular case, a burner that is situated closer to the ith reforming tube is expected to have a higher impact on the ith OTWT than those that are further away. It is noted that an additional inequality constraint (Eq. (21c)) is added to the optimization problem to prevent sparse nonlinear regression in an attempt to reduce the number of terms in Eq. (19) from falsely presuming that the impact due to a closer burner is indefinitely higher than that due to a further burner given the premise that the two are separated from the ith reforming tube by different distances and are supplied with the same FSF flow rate. It is noted that in the constrained optimization problem (Eqs. (20) and (21)), λ_i directly controls the degree of shrinkage for the parameter vector in $M_{i,k}$. Specifically, large values of λ_i result in a high degree of shrinkage and favor underfitting data-driven models with low levels of complexity. On the contrary, small values of λ_i result in a low degree of shrinkage and favor overfitting data-driven models with high goodness of fit. Therefore, it is desired to use the optimal value of λ_i to balance between the degree of complexity and goodness of fit in data-driven models. In this work, leave-out-one (LOO) cross validation is used to search for the optimal value of λ_i among the proposed values, $\mathbf{S}_\lambda = \{0.1, 0.2, \ldots, 1.0, 1.2, \ldots, 2.0, 5.0, 10\}$, because the fitting error (i.e., the mean-square error) might not be an adequate representation for the out-of-sample prediction error. In LOO cross validation, the training data ($\mathbf{T}_i$) is split into sub-training and sub-testing data in such a way that a reformer data set in the training data is assigned to the sub-testing data, and the remaining data sets are assigned to the sub-training data. This procedure generates N different pairs of sub-training and sub-testing data from the training data (where N is the cardinality of the training data), and then, each pair of the sub-training and sub-testing data is used to derive a sub-prediction model library for the ith reforming tube and to evaluate a corresponding out-of-sample prediction error for each value of λ_i in $\mathbf{S}_\lambda$. It is recognized that LOO cross validation is computationally intensive, e.g., the total CPU time is expected to be $\sim N$ times more than that required by an approach that uses the complete training data and the fitting error as the metric to select λ_i, but LOO cross validation provides evidence (i.e., the unbiased estimate of the prediction error) based on which the optimal λ_i can be identified. Specifically, the value of λ_i in $\mathbf{S}_\lambda$ that yields the least prediction error for out-of-sample predictions is considered to be the optimal λ_i.

5. Correction models

In the present work, an algorithm for the correction step is designed based on ordinary Kriging, which is superior to other common interpolation techniques and can yield estimates with minimum variance (Holdaway, 1996), to improve the predicted OTWT distribution, which is generated from the prediction model for the OTWT distribution using $\vec{F}^n$ and F^n_{tot}, by accounting for the effect of interactions between neighboring reforming tubes on the OTWT distribution. The correction step is also designed to be a distributed algorithm, which derives the correction model for the ith reforming tube from the training data independently of other reforming tubes allowing the correction models for the 336 reforming tubes to be simultaneously created, which expedites the model-building process of the correction model for the OTWT distribution. In the remainder of this section, the underlying process that leads to the temperature variation in the OTWT distribution is discussed to elucidate the use of ordinary Kriging in an attempt to account for the impact of interactions among neighboring reforming tubes on the OTWT distribution, and then, assumptions and equations associated with ordinary Kriging will be explicitly presented.

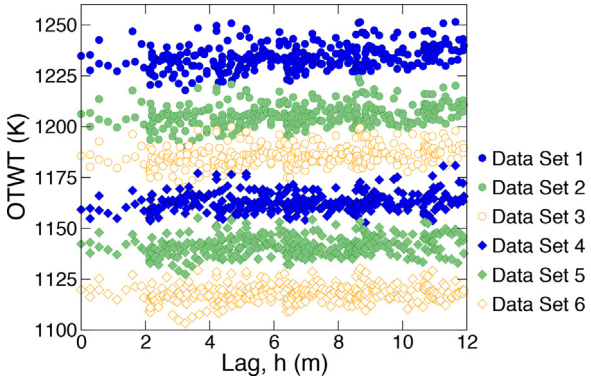

Fig. 8 – The scatter plot of OTWTs of all reforming tubes in six training data sets versus the corresponding distance to the 77th reforming tube, which is referred to as the 77th lag. It is noted that not all reformer CFD data sets in the training data are shown in this figure to avoid cluttering.

As noted in Section 1, the degree of nonuniformity in the OTWT distributions along the reforming tube length is controlled by the temperature distribution of the furnace-side flow, which is a product of many complex interacting transport phenomena and chemical reactions taking place simultaneously inside the reformer. The analytical deterministic solution for the OTWT distribution modeled as a function of the reformer inputs (e.g., the FSF distribution and FSF flow rate) cannot be easily obtained, and therefore, variations in the OTWT distribution observed in the training data appear to be as though they are the result of a random process. Fig. 8 displays the OTWT of each reforming tube as a function of the ith lag (i=77 for the 77th reforming tube is chosen in this figure for demonstration purposes), where the lag is defined as the Euclidean distance between the jth and ith reforming tubes where j is different from i. The figure suggests that the mean of the OTWT distribution in each reformer CFD data set in the training data is constant. This realization justifies the assumption that in a sufficiently small neighborhood the underlying process that governs the variations in the OTWT distribution is intrinsically stationary. Therefore, the spatial correlation of the OTWT among reforming tubes in a sufficiently small neighborhood can be summarized in a spatial variance function (i.e., semivariograms). Under the intrinsic stationarity assumption, the variance function only depends on lag (which is defined as a Euclidean distance between a pair of reforming tubes and is denoted as h) and is calculated using

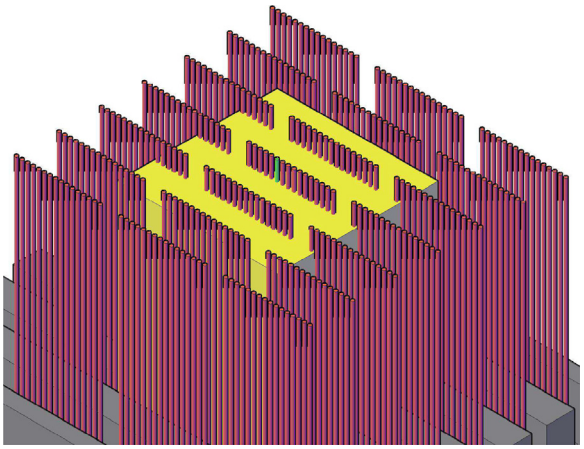

Fig. 9 – The isometric view of the ordinary Kriging neighborhood of the ith reforming tube that is denoted as S_i^K and is defined as an 8 m by 8 m region centered at the ith tube, inside which the underlying process that gives rise to the variations in the OTWT distribution is assumed to be intrinsically stationary.

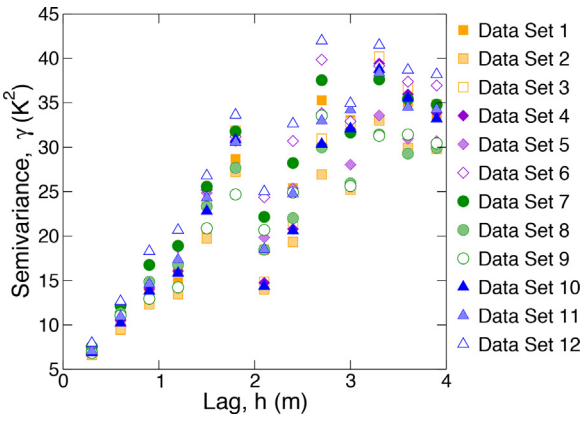

Fig. 10 – Omnidirectional sample semivariograms for S_{77}^K computed by the classical estimator as shown in Eq. (22) using the training data. It is noted that not all results generated using reformer CFD data sets in the training data are displayed in this figure to avoid cluttering.

the classical estimator (i.e., the method of moments) from the training data as follows,

$$\hat{\gamma}_i^n(h) = \frac{1}{2q_i(h)} \sum_{q_i(h)} (T_j^n - T_k^n)^2 \tag{22}$$

where $\hat{\gamma}_i^n(h)$ is the sample semivariogram capturing the spatial correlation of the OTWTs among reforming tubes that are separated by $h \pm 0.15$ m and in the neighborhood of the ith reforming tube (denoted as S_i^K) created from using the nth reformer CFD data set in the training data, T_j^n and T_k^n are the OTWTs of the reforming tubes that are separated by $h \pm 0.15$ m and in S_i^K, and $q_i(h)$ is the number of pairs of reforming tubes that are separated by $h \pm 0.15$ m and in S_i^K. In this work, S_i^K is defined to be an 8 m by 8 m region centered at the ith reforming tube as shown in Fig. 9 so that there are a sufficient number of reforming tubes in S_i^K, which allows relatively stable regional sample semivariograms to be created for each training data set. The sample semivariograms for S_{77}^K (the neighborhood of the 77th reforming tube) calculated based on Eq. (22) using the training data are shown in Fig. 10. It is recognized that

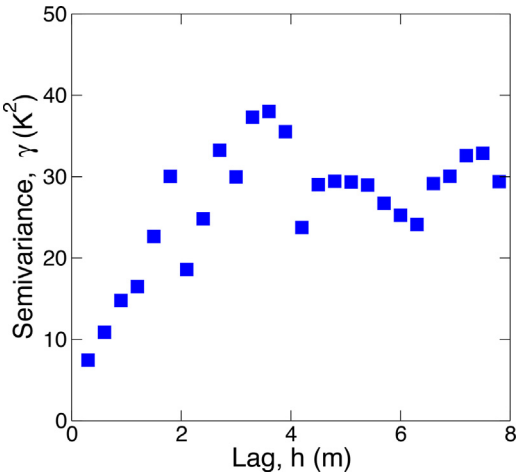

Fig. 11 – Omnidirectional average sample semivariograms for S_{77}^K computed by Eqs. (22) and (23) using information of all reforming tube pairs in S_{77}^K from the training data.

the sample semivariograms in S_{77}^K at a given lag are similar across all training data, which suggests that the spatial correlation among refoming tubes separated by a distance of h m can be modeled to be independent of the FSF distribution and the total FSF flow rate. Therefore, all individual sample semivariograms can be pooled together to create average sample semivariograms as follows,

$$\hat{\gamma}_i^*(h) = \frac{\sum_{n=1}^N \hat{\gamma}_i^n(h) q_i(h)}{\sum_{n=1}^N q_i(h)}. \tag{23}$$

The average sample semivariograms shown in Fig. 11 are expected to be a reasonable representation of the spatial correlation among reforming tubes in S_i^K (Bilonick, 1983).

It is noted from Fig. 11 that the average sample semivariograms typically increase with increasing lag, which indicates that the OTWTs of reforming tubes that are closely situated are more highly correlated than the OTWTs of those that are farther apart, with the exception at the lag of approximately 2 m. A detailed analysis of the reformer geometry shown in Fig. 1 reveals that the spatial correlation between reforming tubes might also be dependent on directionality in which the reforming tubes are separated. Two new terminologies are introduced to facilitate the discussion of anisotropic sample semivariograms: pairs of reforming tubes within a row are considered to be in the North-South direction and are referred to as North-South pairs, while pairs of reforming tubes in two adjacent rows are considered to be in the East-West direction and are referred to as East-West pairs. It is noted that the adjacent rows of reforming tubes are separated by a distance of approximately 2 m which is the smallest lag between any East-West pair. In addition, the East-West pairs in which reforming tubes are separated by approximately 2 m are expected to be under the influence of a number of common burners, specifically, under the assumption that the four nearest burners to the ith reforming tube are the ith default predictors, the regressor collection layer determines that their sets of regressors always have two common predictors. On the contrary, the North-South pairs in which reforming tubes are separated by approximately 2 m may or may not be under the influence of any common burner. This analysis suggests that at the lag of approximately 2 m, the OTWTs of the East-West pairs are expected to be more sim-

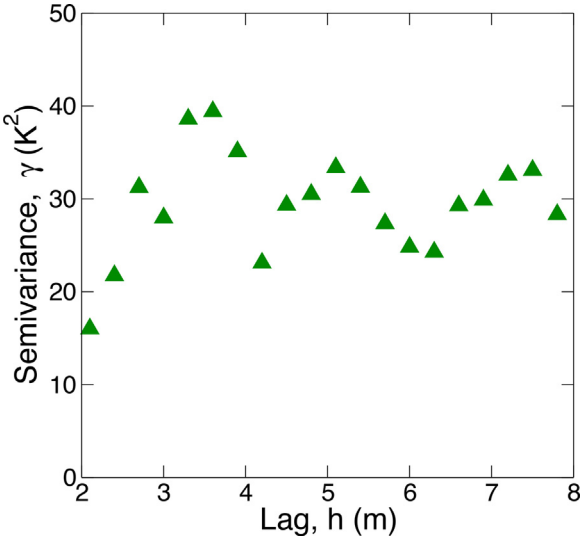

Fig. 12 – East-West anisotropic average sample semivariograms for S_{77}^{K} computed by Eqs. (22) and (23) using information of the East-West pairs in S_{77}^{K} from the training data.

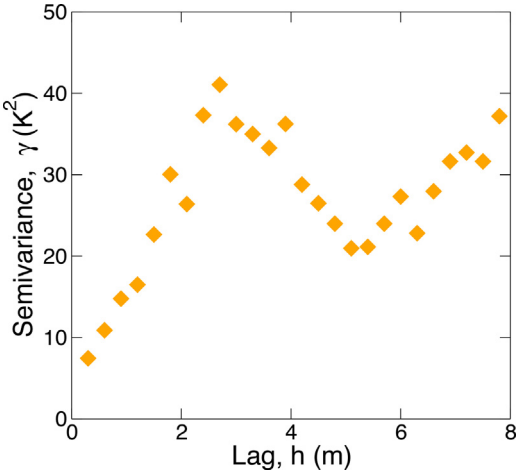

Fig. 13 – North-South anisotropic average sample semivariograms for S_{77}^{K} computed by Eqs. (22) and (23) using information of the North-South pairs in S_{77}^{K} from the training data.

ilar than those of the North-South pairs. Furthermore, the number of East-West pairs at the lag of approximately 2 m is significantly larger than the number of the North-South pairs separated by the same distance due to the reformer layout as shown in Fig. 1. Therefore, the average omnidirectional sample semivariograms at the lag of approximately 2 m as shown in Fig. 11 might not be used to represent the spatial correlation between the North-South pairs. In this work, North-South and East-West average anisotropic sample semivariograms are evaluated according to Eq. (23) using information of the North-South pairs and East-West pairs in $\mathbf{S}_i^K$, respectively, and are shown in Figs. 12–13 . It is recognized that with the exception at the lag of approximately 2 m at which the East-West pairs exhibit a strong spatial correlation, the scatter plot of East-West average anisotropic sample semivariograms shown in Fig. 12 indicates that when the lag is greater than or equal to 2.4 m, East-West pairs appear to be spatially uncorrelated. Therefore, information of the East-West average anisotropic sample semivariograms will not be utilized in the model-building process of the spatial model (i.e., the correction model) for the ith OTWT. Additionally, it is noted from Fig. 13 that when the lag is less than 4 m, North-South pairs are spatially correlated, and North-South average anisotropic sample semivariograms exhibit the expected trend in which semivariance increases with increasing lag; however, when the lag becomes greater than or equal to 4 m, North-South pairs suddenly appear to be spatially uncorrelated. The analysis of the average anisotropic sample semivariograms suggests that the OTWTs of reforming tubes separated by a distance greater than or equal to 4 m are spatially uncorrelated so that this information should not be utilized in the model-building process of the correction model for the ith reforming tube. Therefore, the present work only uses information of reforming tubes in $\mathbf{S}_i^K$ and separated by a distance that is strictly less than 4 m to construct the omnidirectional and isotropic average sample semivariograms.

Next, the omnidirectional and anisotropic average sample semivariograms are fitted with linear and exponential functions to generate 4 different theoretical semivariograms, namely, linear omnidirectional, exponential omnidirectional, linear anisotropic and exponential anisotropic theoretical semivariogram models. Then, LOO cross validation is used to identify the most suitable theoretical semivariogram model, which most accurately describes the spatial correlation among neighboring reforming tubes in $\mathbf{S}_i^K$, to be used in the model-building process of the ith reforming tube. The linear (Eq. (24a)) and exponential (Eq. (24b)) functions (Cressie, 1985) are given as follows,

for $h > 0$

$$\tilde{\gamma}_l(h) = \tau_{1,l} + \tau_{2,l} h \qquad 0 \le \tau_{1,l}, \tau_{2,l} < +\infty \tag{24a}$$

$$\tilde{\gamma}_e(h) = \tau_{1,e} + \tau_{2,e}\left[1 - \exp\left(\frac{h}{\tau_{3,e}}\right)\right] \qquad 0 \le \tau_{1,e}, \tau_{2,e} < +\infty \;\&\; 0 \le \tau_{3,e} < 4 \tag{24b}$$

for $h = 0$

$$\tilde{\gamma}_l(h) = \tilde{\gamma}_e(h) = 0 \tag{24c}$$

where $\tau_{1,l}$ and $\tau_{2,l}$ are parameters of the linear theoretical semivariogram model, and $\tau_{1,e}$, $\tau_{2,e}$ and $\tau_{3,e}$ are parameters of the exponential theoretical semivariogram model. It is noted that $\tau_{3,e}$ in Eq. (24b) is related to the range in which variations of the OTWTs are spatially correlated and, therefore, it is reasonable to assume that $\tau_{3,e}$ is less than 4 m. Parameters of theoretical semivariograms are estimated using the method of weighted least squares developed in Cressie (1985).

Finally, the correction model for the ith reforming tube is formulated as a weighted average OTWT of the neighbors in $\mathbf{W}_i^K$ as follows,

$$\hat{T}_i^{C,n} = \sum_{j}^{w_i^K} w_{i,j}^C T_j^n \tag{25}$$

where j is an index of a reforming tube in $\mathbf{W}_i^K$, which is a subset of $\mathbf{S}_i^K$, and $w_{i,j}^C$ is the correction weighting factor of the jth reforming tube. In this work, $\mathbf{W}_i^K$ is defined to consist of the three nearest northward, three nearest southward, one nearest eastward and one nearest westward neighbors of the ith reforming tube. In the event that the ith reforming

tube is missing any of the aforementioned neighbors, $\mathbf{W}_i^K$ is shrunk down accordingly, and the ith OTWT is estimated only using the existing neighbors. For instance, as the 1st reforming tube has neither three nearest northward neighbors nor one nearest eastward neighbor as shown in Fig. 1, the cardinality of $\mathbf{W}_1^K$ is reduced to four, and the 1st OTWT is directly computed as the weighted average OTWT of the three nearest southward and one nearest westward neighbors.

Two different approaches for obtaining $w_{i,j}^C$ associated with elements in $\mathbf{W}_i^K$ are proposed to account for the choice of incorporating directionality into the correction model for the ith reforming tube. Specifically, when an omnidirectional theoretical semivariogram model is used to represent the spatial correlation between reforming tubes in $\mathbf{W}_i^K$, it is assumed that the ith reforming tube is spatially correlated to all neighbors in $\mathbf{W}_i^K$, and all $w_{i,j}^C$ in the correction model (Eq. (25)) are the Kriging weights and can be calculated as follows,

$$\Gamma_i \vec{w}_i^C = \vec{\gamma}_i \tag{26a}$$

subject to

$$\Gamma_i \in \mathrm{IR}^{N_{i,K}+1 \times N_{i,K}+1} \text{ such that,}$$

$$\begin{aligned} &\Gamma_{i,kl} = 0; k = l \quad \text{where} \quad k,l \in \left[1, N_{i,K}+1\right] \\ &\Gamma_{i,kl} = \Gamma_{i,lk} = \tilde{\gamma}_i^*\left(\hat{h}_{kl}\right) \text{where} \quad k,l \in \left[1, N_{i,K}\right] \\ &\Gamma_{i,(N_{i,K}+1)l} = \Gamma_{i,l(N_{i,K}+1)} \text{where} \quad l \in \left[1, N_{i,K}\right] \end{aligned} \tag{26b}$$

$$\vec{w}_i^C = \left[w_{i,j}^C, \ldots, \lambda_i^C\right]^T \forall j \in \mathbf{W}_i^K \tag{26c}$$

$$\vec{\gamma}_i = \left[\tilde{\gamma}_i^*\left(h_{ij}\right), \ldots, 1\right]^T \quad \forall j \in \mathbf{W}_i^K \tag{26d}$$

where Γ_i represents the matrix of semivariances between the neighbors in $\mathbf{W}_i^K$, $\vec{\gamma}_i$ represents the vector of semivariances between the ith reforming tube and its neighbors in $\mathbf{W}_i^K$, $\tilde{\gamma}_i^*$ is the best omnidirectional theoretical semivariogram model, which is identified by LOO cross validation, to describe the spatial correlation of reforming tubes in $\mathbf{W}_i^K$, $N_{i,K}$ is the number of Kriging weights in the correction model for the ith reforming tube, λ_i^C is the Lagrangian multiplier, h_{ij} is the lag between the ith and jth reforming tubes and $\hat{h}_{kl}$ is the lag between the kth and lth neighbors in $\mathbf{W}_i^K$. It is noted that the subscripts in $\hat{h}_{kl}$ represent the order in which neighbors are arranged in $\mathbf{W}_i^K$ instead of indices of reforming tubes, and in this study, the neighbors in $\mathbf{W}_i^K$ are arranged in an increasing order of their reforming tube indices. When an anisotropic theoretical semivariogram model is used to represent the spatial correlation between reforming tubes in $\mathbf{W}_i^K$, it is assumed that the ith reforming tube is spatially correlated to the North-South neighbors but is spatially uncorrelated to the East-West neighbors in $\mathbf{W}_i^K$. In this scenario, the ith corrected OTWT is defined as an average of two distinct estimates calculated using information of the North-South neighbors and of the East-West neighbors, respectively. It is noted that, as the spatial correlation between the North-South neighbors and the ith reforming tube is captured in the anisotropic theoretical semivariogram model, the estimate calculated using information of the North-South neighbors can be obtained after all $w_{i,j}^C$ associated with the North-South neighbors are calculated as shown in Eq. (26). On the other hand, because the ith reforming tube is spatially uncorrelated to the East-West neighbors, the estimate calculated using information of the East-West neighbors is simplified to an average OTWT among the East-West neighbors.

6. Combined model uniting prediction and correction models

Upon the simultaneous creation of prediction models as discussed in Section 4 and correction models as discussed in Section 5 for 336 reforming tubes from the training data, the data-driven model for the OTWT distribution, which is a function of the FSF distribution and total FSF flow rate and is also able to account for interactions among neighboring reforming tubes, can be assembled. Initially, the prediction models for 336 reforming tubes are pooled to create the prediction model for the OTWT distribution, by which the predicted OTWT distribution can be estimated based on a given FSF distribution and total FSF flow rate, as follows,

$$\begin{pmatrix} \hat{T}_1^{P,n} \\ \hat{T}_2^{P,n} \\ \vdots \\ \hat{T}_{336}^{P,n} \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{K_1} w_{1,k}^P \tilde{T}_{1,k}^{P,n} \\ \sum_{k=1}^{K_2} w_{2,k}^P \tilde{T}_{2,k}^{P,n} \\ \vdots \\ \sum_{k=1}^{K_{336}} w_{336,k}^P \tilde{T}_{336,k}^{P,n} \end{pmatrix}. \tag{27}$$

Then, the correction models for 336 reforming tubes are also pooled to create the correction model for the OTWT distribution, by which the corrected OTWT distribution can be estimated based on the predicted OTWT distribution, as follows,

$$\begin{pmatrix} \hat{T}_1^{C,n} \\ \hat{T}_2^{C,n} \\ \vdots \\ \hat{T}_{336}^{C,n} \end{pmatrix} = \begin{pmatrix} \sum_{j}^{w_1^K} w_{1,j}^C \hat{T}_j^{P,n} \\ \sum_{j}^{w_2^K} w_{2,j}^C \hat{T}_j^{P,n} \\ \vdots \\ \sum_{j}^{w_{336}^K} w_{336,j}^C \hat{T}_j^{P,n} \end{pmatrix}. \tag{28}$$

Next, the data-driven model for the OTWT distribution is formulated as a weighted average of the prediction and correction models for the OTWT distribution as follows,

$$\begin{pmatrix} \hat{T}_1^{n} \\ \hat{T}_2^{n} \\ \vdots \\ \hat{T}_{336}^{n} \end{pmatrix} = w^P \begin{pmatrix} \hat{T}_1^{P,n} \\ \hat{T}_2^{P,n} \\ \vdots \\ \hat{T}_{336}^{P,n} \end{pmatrix} + \left(1 - w^P\right) \begin{pmatrix} \hat{T}_1^{C,n} \\ \hat{T}_2^{C,n} \\ \vdots \\ \hat{T}_{336}^{C,n} \end{pmatrix} \tag{29}$$

where the optimal value of w^P denoted by $\hat{w}^P$ is selected among the proposed values, $\mathbf{S}_w = \{0.0, 0.1, \ldots, 1.0\}$ by LOO cross validation, which can be carried out in the same manner as

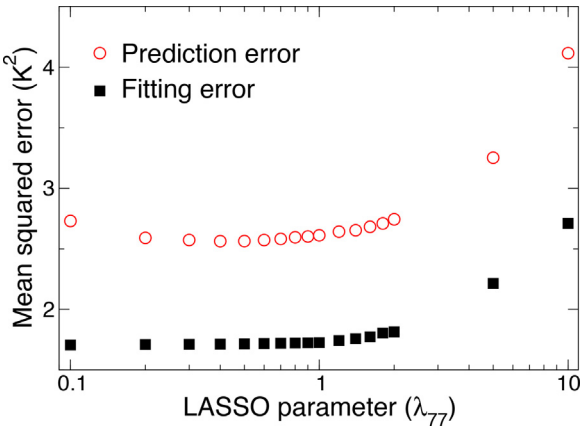

Fig. 14 – Summary of LOO cross validation for selecting the optimal value of λ_{77} from $\mathbf{S}_\lambda$ in which the mean squared prediction errors are denoted by the empty red circles and the mean squared fitting errors are denoted by the filled black squares. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

discussed in Section 4.3. Specifically, the value of w^P in $\mathbf{S}_w$ that yields the least prediction error for out-of-sample predictions is considered to be the optimal w^P.

7. Results

The statistical-based model identification approach developed in this work is structured to be entirely parallelized; specifically, the prediction and correction models of the ith reforming tube can be derived simultaneously from the training data and independently of the model-building process of other reforming tubes. This feature allows the optimal LASSO parameter and theoretical semivariogram model for the ith reforming tube and the optimal weighting factor of the BMA predicted estimates to be determined using leave-out-one cross validation, which is expected to improve the prediction accuracy of the OTWT distribution model for out-of-sample predictions. We note that the model-building process for the OTWT of each reforming tube is identically, independently and simultaneously executed on the shared computing cluster at UCLA, and therefore, the effectiveness of the proposed model identification scheme can be demonstrated using results generated from the model-building process of any reforming tube. In the remainder of this section, the 77th reforming tube is chosen as a representative example because the number of sub-prediction models with high goodness of fit (i.e., 4) and the number of predictors (i.e., 9) for the 77th OTWT make it possible to illustrate the effectiveness of the proposed approach to identify the important burners and to allow the prediction model for the 77th OTWT to account for model uncertainty while forecasting in a concise manner.

The results from LOO cross validation to select the optimal value of λ_{77} (denoted as $\hat{\lambda}_{77}$) from $\mathbf{S}_\lambda$ are summarized in Fig. 14. Specifically, the value of λ_{77} controls the model complexity and goodness of fit as discussed in Section 4.3 when the sparse nonlinear regression is formulated as an L1 penalized optimization problem, which is illustrated as shown in Fig. 14. It is recognized from Fig. 14 that the mean square fitting error increases with increasing values of λ_{77}, and specifically, the fitting error is the largest at the highest values of λ_{77} considered in this work as low-complexity models are favored, while the fitting error is the smallest at the lowest value of λ_{77} as models with high goodness of fit are favored. In addition, Fig. 14 reveals that the fitting error is over-optimistic because it is lower than the mean squared prediction error for all values of λ_{77}, and the fitting error should not be used as the metric for selecting $\hat{\lambda}_{77}$ from $\mathbf{S}_\lambda$ because the prediction error is not necessarily minimized at the value of λ_{77} which minimizes the fitting error. Specifically, Fig. 14 indicates that at λ_{77}=0.4, the prediction error is minimized, which suggests that the optimal value of λ_{77} is 0.4, i.e.,

$$\hat{\lambda}_{77} = 0.4. \tag{30}$$

Thereafter, $\hat{\lambda}_{77}$ is used as the LASSO parameter of the sparse nonlinear regression (Eq. (20)) in the prediction step algorithm, by which the prediction model for the dependence of the 77th OTWT on the FSF distribution and total FSF flow rate is derived from the complete training data. It is found that this procedure generates four nested sub-prediction models of the 77th reforming tube, which together represent the prediction model for the 77th reforming tube. The four sub-prediction models of the 77th reforming tube, each of which can be used to estimate the 77th OTWT based on a given FSF distribution and total FSF flow rate, are as follows,

$$\begin{aligned}\tilde{T}^{P,n}_{77,1} = \; & 134.70\sqrt[5]{F^n_{18}} + 225.49\sqrt[5]{F^n_{19}} + 141.73\sqrt[5]{F^n_{30}} + 234.28\sqrt[5]{F^n_{31}} \\ & +2.54\exp\left(F^n_{18}\right) + 3.17\exp\left(F^n_{19}\right) + 14.45\exp\left(F^n_{32}\right) + 348.15\end{aligned} \tag{31a}$$

$$\begin{aligned}\tilde{T}^{P,n}_{77,2} = \; & 51.80\sqrt[5]{F^n_{17}} + 131.61\sqrt[5]{F^n_{18}} + 177.63\sqrt[5]{F^n_{19}} + 134.87\sqrt[5]{F^n_{30}} \\ & +179.40\sqrt[5]{F^n_{31}} + 66.49\sqrt[5]{F^n_{32}} + 10.94\exp\left(F^n_{17}\right) + 1.17\exp\left(F^n_{18}\right) \\ & +0.64\exp\left(F^n_{19}\right) + 6.22\exp\left(F^n_{32}\right) + 348.15\end{aligned} \tag{31b}$$

$$\begin{aligned}\tilde{T}^{P,n}_{77,3} = \; & 61.26\sqrt[5]{F^n_{17}} + 125.22\sqrt[5]{F^n_{18}} + 166.56\sqrt[5]{F^n_{19}} + 125.22\sqrt[5]{F^n_{30}} \\ & +166.56\sqrt[5]{F^n_{31}} + 60.01\sqrt[5]{F^n_{32}} + 37.83\sqrt[5]{F^n_{43}} + 5.40\exp\left(F^n_{17}\right) \\ & +6.39\exp\left(F^n_{32}\right) + 6.63\exp\left(F^n_{43}\right) + 348.15\end{aligned} \tag{31c}$$

$$\begin{aligned}\tilde{T}^{P,n}_{77,4} = \; & 61.48\sqrt[5]{F^n_{17}} + 120.02\sqrt[5]{F^n_{18}} + 175.72\sqrt[5]{F^n_{19}} + 120.02\sqrt[5]{F^n_{30}} \\ & +175.72\sqrt[5]{F^n_{31}} + 69.04\sqrt[5]{F^n_{32}} + 21.96\sqrt[5]{F^n_{43}} + 10.00\exp\left(F^n_{15}\right) \\ & +8.43\exp\left(F^n_{43}\right) + 348.15.\end{aligned} \tag{31d}$$

At a glance, the four sub-prediction models successfully account for the reformer geometry (i.e., the reforming tube and burner arrangement) as only the parameters associated with burners that are situated nearby the 77th reforming tube are nonzero and also obey the inverse square law for thermal radiation as the parameters associated with burners that are closer to the 77th reforming tube are larger. The distance between each predictor of the 77th OTWT and the 77th reforming tube is detailed in Table 1. On closer inspection, it

Table 1 – Distance from burners in S_{77R} to the 77th reforming tube.

Burner ID	Distance (m)
19th	1.08
31st	1.08
18th	1.44
30th	1.44
32nd	2.32
17th	2.36
43rd	3.21
15th	5.24

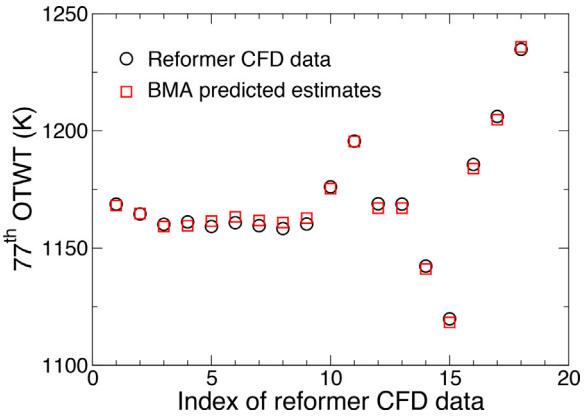

Fig. 15 – Comparison of the 77th OTWT between reformer CFD data from the training data and BMA predicted estimates generated from the prediction model for the 77th reforming tube.

is recognized that out of the eight transformation functions (Eq. (17)) proposed in the development of the sparse nonlinear regression, only the parameters associated with the quint root and exponential functions are nonzero, and the parameters associated with the quint root function are noticeably larger than those associated with the exponential function (Eq. (31)). We speculate that the use of the exponential function in the sub-prediction models by the sparse nonlinear regression is because the training data might have suggested that the predictors located further away from the 77th reforming tube have higher impacts on the 77th OTWT than we previously expect (Eq. (21b)). This analysis reveals the underlying function that governs the relationship between the 77th OTWT and FSF flow rates of its predictors. Next, the prediction model is created as the weighted average of the four sub-prediction models of the 77th reforming tube using Bayesian model averaging as follows,

$$\hat{T}_{77}^{P,n} = 0.01\tilde{T}_{77,1}^{P,n} + 0.23\tilde{T}_{77,2}^{P,n} + 0.29\tilde{T}_{77,3}^{P,n} + 0.47\tilde{T}_{77,4}^{P,n} \quad (32)$$

where the weighting factor associated with each sub-prediction model indicates the level of supporting evidence given by the training data. Then, the training data are used to evaluate the resubstitution accuracy of the prediction model for the 77th reforming tube. The comparison between reformer data from the training data and BMA predicted estimates generated from the prediction model (Eq. (32)) is shown in Figs. 15 and Fig. 17, and the corresponding residual plot is shown in Fig. 16. Specifically, Fig. 16 shows that the maximum and average residuals are 2.54 K and −0.01 K, respectively, and Fig. 17 shows that all plotted points are close to the diagonal line with the slope of 1 and y-intercept of zero.

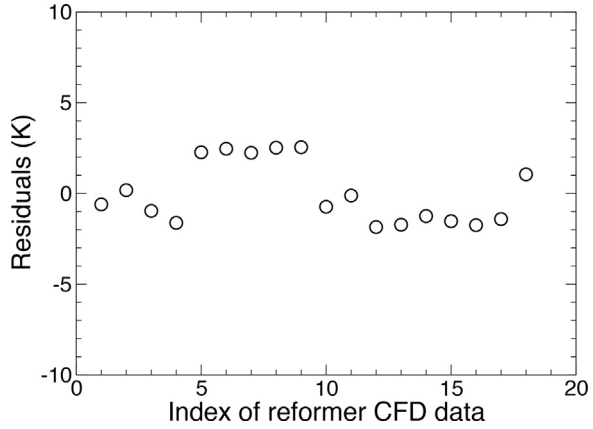

Fig. 16 – Residuals between reformer CFD data from the training data and BMA predicted estimates generated from the prediction model for the 77th reforming tube.

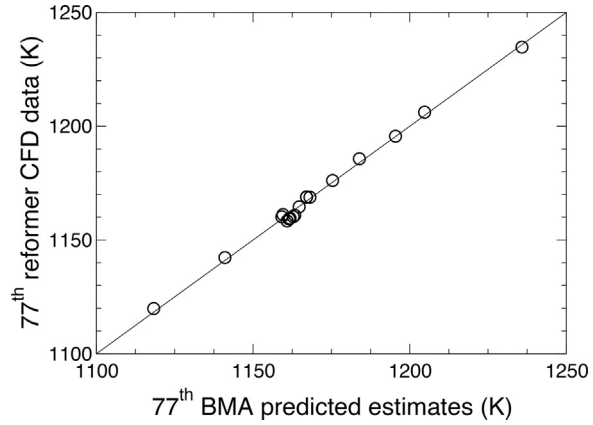

Fig. 17 – Reformer CFD data from the training data versus BMA predicted estimates generated from the prediction model for the 77th reforming tube scatter plots.

Therefore, Figs. 15, 16 and Fig. 17 show that the prediction model for the 77th reforming tube has a high goodness of fit and provides an excellent description for the dependence of the 77th OTWT on the FSF distribution and total FSF flow rate.

Next, the results from LOO cross validation, which allow the best theoretical semivariogram model (denoted as $\hat{\gamma}_{77}^{*}$) to be selected among the linear omnidirectional, linear anisotropic, exponential omnidirectional, and exponential anisotropic theoretical semivariogram models so that the spatial correlation among neighboring reforming tubes in $\mathbf{S}_{77}^{K}$ can be described with adequate accuracy, are summarized in Fig. 18. It is noted from Fig. 18 that the correction models for the 77th reforming tube created in LOO cross validation using the linear omnidirectional theoretical semivariogram model yield the smallest mean squared prediction error for out-of-sample predictions, and therefore, $\hat{\gamma}_{77}^{*}$ is assumed to be the linear omnidirectional theoretical semivariogram model. Thereafter, $\hat{\gamma}_{77}^{*}$ is used as the predetermined theoretical semivariogram model in the correction step algorithm, by which the correction model for the 77th reforming tube is derived from the complete training data as shown in Table 2. Table 2 shows the expected trend in spatial modeling, i.e., as the distance between a neighbor and the 77th reforming tube increases, their OTWT values becomes less correlated. Then, the training data are used to evaluate the resubstitution accuracy of the correction model for the 77th reforming tube. The comparison

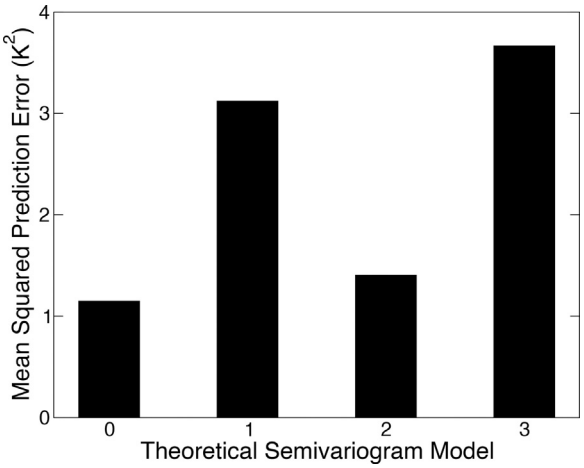

Fig. 18 – A plot of the mean squared prediction error associated with each of the four theoretical semivariogram models considered in the LOO cross validation to select the best theoretical semivariogram to model the spatial correlation among reforming tubes in the S^{K}_{77}. The values on the horizontal axis, i.e., 0, 1, 2, and 3, correspond to the linear omnidirectional, linear anisotropic, exponential omnidirectional, and exponential anisotropic theoretical semivariogram models, respectively. This figure indicates that the linear omnidirectional model has the least mean squared prediction error and is considered the best model.

Table 2 – Lags, Kriging weights, and directionality of reforming tubes used in the weighted average correction step in calculating the 77th OTWT.

Reforming tube ID	Lag (m)	Kriging weight	Directionality
29	2.13	0.03	Eastward
74	0.84	0.06	Northward
75	0.56	0.13	Northward
76	0.28	0.29	Northward
78	0.28	0.30	Southward
79	0.56	0.15	Southward
80	1.60	0.03	Southward
125	2.13	0.03	Westward

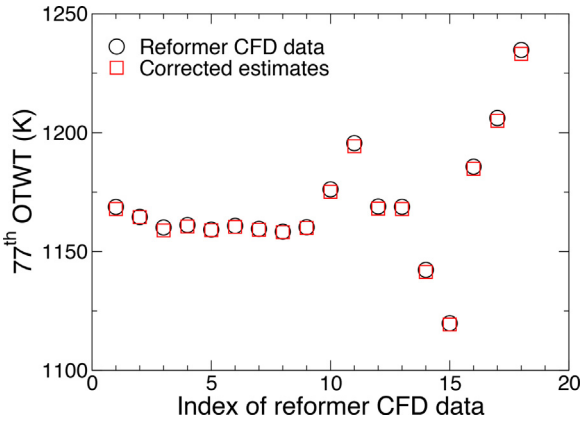

Fig. 19 – Comparison of the 77th OTWT between reformer CFD data from the training data and corrected estimates generated from the correction model for the 77th reforming tube.

between reformer data from the training data and corrected estimates generated from the corrected model (Table 2) is shown in Figs. 19 and 21 ; the corresponding residual plot is shown in Fig. 20. Specifically, Fig. 20 shows that the maximum

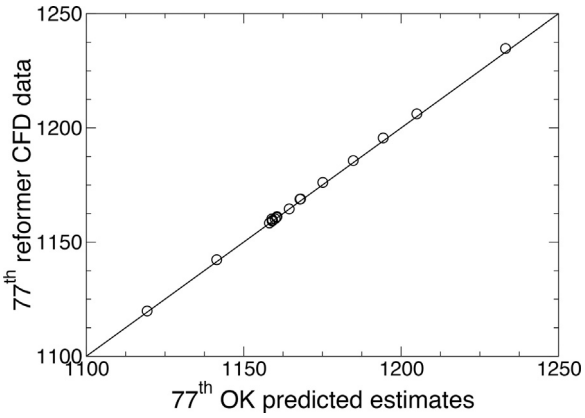

Fig. 21 – Reformer CFD data from the training data versus corrected estimates generated from the correction model for the 77th reforming tube scatter plots.

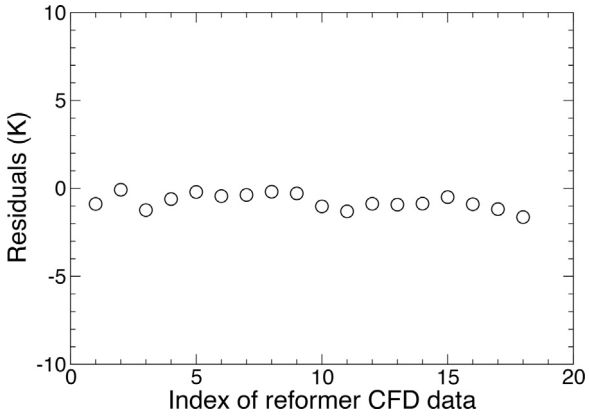

Fig. 20 – Residuals between reformer CFD data from the training data and corrected estimates generated from the correction model of the 77th reforming tube.

and average residuals are 1.63 K and −0.74 K, respectively, and Fig. 21 shows that all plotted points are close to the diagonal line with the slope of 1 and y-intercept of zero. Therefore, Figs. 19, 20 and 21 show that the correction model for the 77th reforming tube has a high goodness of fit and provides an excellent description of the 77th OTWT by using information of the neighbors.

The results from LOO cross validation to select the optimal weighting factor for the BMA predicted estimates from $\mathbf{S}_w$ are summarized in Fig. 22. It is noted from Fig. 22 that the fitting error decreases with increasing values of w^P, and specifically, the fitting error is at its smallest when the value of w^P becomes 1, which implies that estimates of the OTWT distribution are based entirely on the prediction model for the OTWT distribution, while the fitting error is the largest when the value of w^P becomes 0, which implies that estimates of the OTWT distribution are based entirely on the correction model for the OTWT distribution. This observation is expected because the correction model for the OTWT distribution is derived from the training data instead of from the predicted OTWT distributions, which allows it to have realistic knowledge about spatial correlation among neighboring reforming tubes so that the overall predictability performance of the data-driven model for the OTWT distribution for out-of-sample predictions can be improved. Fig. 22 demonstrates the necessity of using the correction model for the OTWT distribution, and specifically, the prediction error is minimized at w^P=0.9, which suggests that the correction model for the OTWT distribution improves

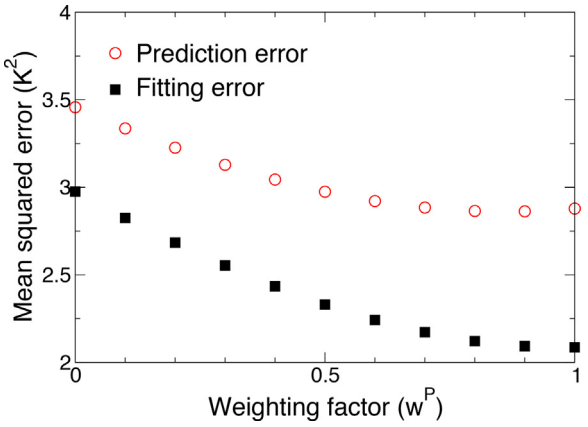

Fig. 22 – Summary of LOO cross validation for selecting the optimal value of w^P from S_w in which the mean squared prediction errors are denoted by the empty red circles and the mean squared fitting errors are denoted by the filled black squares. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

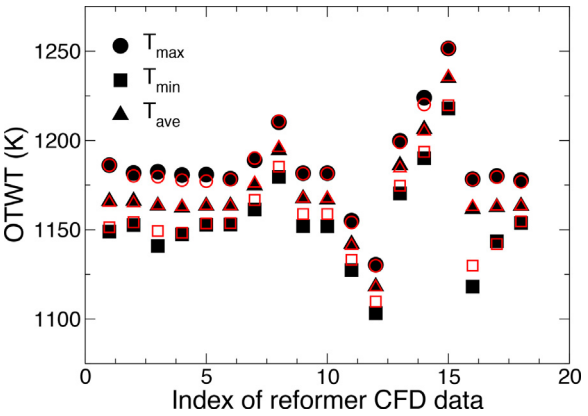

Fig. 23 – Comparison of the maximum (T_{max}), average (T_{ave}) and minimum (T_{min}) OTWTs in each pair of data sets between reformer CFD data in the training data, which are represented by the filled black symbols, and estimates generated from the data-driven model for the OTWT distribution, which are represented by the empty red symbols. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

the out-of-sample predictive performance of the data-driven model for the OTWT distribution. Subsequently, the prediction models and correction models for 336 reforming tubes derived from the complete training data and the optimal value of w^P are used to formulate the data-driven model for the OTWT distribution. We note that the model for the OTWT distribution depicts a multiple-input multiple-output (MIMO) system and is composed of 336 algebraic equations, which can be solved instantaneously on a standard compute node on a shared computing cluster at UCLA (i.e., the Hoffman2 cluster). Then, the training data are used to evaluate the resubstitution accuracy of the data-driven model for the OTWT distribution. The comparison between reformer data from the training data and estimates generated from the data-driven model for the OTWT distribution is shown in Figs. 23–24 and Table 3. Fig. 23 provides a descriptive comparison of the maximum, average and minimum OTWT in each pair of data sets between reformer

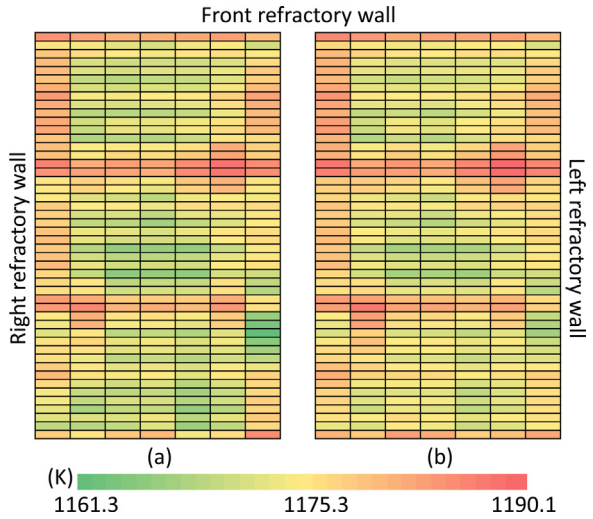

Fig. 24 – Comparison of the OTWT distributions between the 7th data set in the training data (a) and its corresponding estimate generated from the data-driven model for the OTWT distribution (b). In this figure, the layout of 336 reforming tubes in the reformer is represented by a table consisting of 336 rectangular cells, in which each cell corresponds to a unique reforming tube in the reformer, i.e., a cell at the bottom right corner of the table represents the 48*th* reforming tube in the reformer as shown in Fig. 1.

Table 3 – Analysis of the deviations between reformer CFD data in the training data and estimates generated from the data-driven model for the OTWT distribution.

Index of data set	$(\Delta T)_{max}$ K	$(\Delta T)_{ave}$ K	$\sigma_{\Delta T}$ K
1	8.9	2.1	2.5
2	8.1	1.8	1.9
3	18.5	2.2	7.2
4	7.3	1.5	1.9
5	6.7	1.1	1.1
6	6.9	1.1	1.1
7	5.6	1.4	1.1
8	8.1	1.6	2.2
9	7.3	0.9	0.9
10	7.3	1.1	1.0
11	7.5	2.0	2.7
12	7.2	1.3	2.0
13	8.5	1.9	1.6
14	7.7	1.4	1.1
15	2.6	0.7	0.3
16	12.7	2.2	6.1
17	7.2	1.3	1.2
18	8.2	1.0	1.6

CFD data in the training data and estimates generated from the data-driven model for the OTWT distribution, Fig. 24 provides a visual comparison of the OTWT distribution contour map between the 7th data set in the training data and its corresponding estimate generated from the data-driven model for the OTWT distribution, and Table 3 provides an analysis of the deviations between reformer CFD data in the training data and estimates generated from the data-driven model for the OTWT distribution. It is noted that the contour maps shown in Fig. 24 created from the 7th data set in the training data and its corresponding estimate generated from the data-driven model for the OTWT distribution are nearly identical. Specifically, the maximum, minimum and average OTWT of the estimated OTWT distribution are adequately close to those in the reformer CFD data as shown in Fig. 23, and the

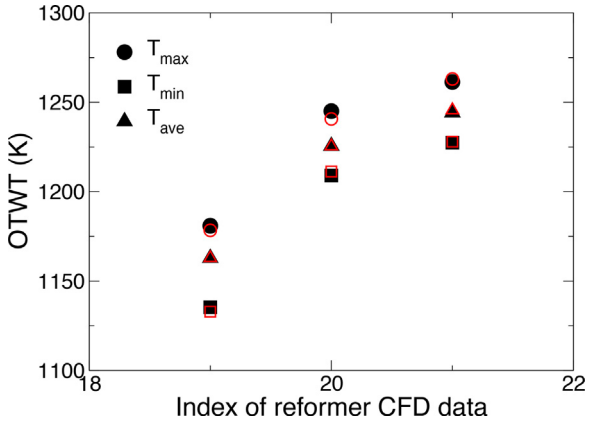

Fig. 25 – Comparison of the maximum (T_{max}), average (T_{ave}) and minimum (T_{min}) OTWTs in each pair of data sets between reformer CFD data in the testing data, which are represented by the filled black symbols, and estimates generated from the data-driven model for the OTWT distribution, which are represented by the empty red symbols. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 4 – Analysis of the deviations between reformer CFD data in the testing data and estimates generated from the data-driven model for the OTWT distribution.

Index of data set	$(\Delta T)_{max}$ K	$(\Delta T)_{ave}$ K	$\sigma_{\Delta T}$ K
19	8.2	1.6	2.3
20	5.7	0.7	0.6
21	4.3	1.4	1.0

maximum and average deviations between the pair as shown in Table 3 are 5.6 K and 1.4 K, respectively, with the standard deviation of 1.1 K which is approximately 0.1% of the average OTWT in this reformer data. This evidence indicates that the 7th OTWT distribution generated by the data-driven model is an accurate representation of the corresponding reformer CFD data. In addition, similar conclusions can be drawn from Fig. 23 and Table 3 for other data sets in the training data. Therefore, it can be concluded that the data-driven model for the OTWT distribution has a high goodness of fit and provides an excellent description for the dependence of the OTWT distribution on the FSF distribution, total FSF flow rate and interactions among neighboring reforming tubes.

Finally, the unbiased testing data are used to evaluate the out-of-sample predictive performance of the data-driven model for the OTWT distribution generated from the proposed integrated model identification scheme. The comparison between reformer data from the testing data and estimates generated from the data-driven model for the OTWT distribution is shown in Fig. 25 and Table 4. Fig. 25 provides a descriptive comparison of the maximum, average and minimum OTWTs in each pair of data sets between reformer CFD data in the testing data and estimates generated from the data-driven model for the OTWT distribution, and Table 4 provides an analysis of the deviations between reformer CFD data in the testing data and estimates generated from the data-driven model for the OTWT distribution. Specifically, the maximum, minimum and average OTWT of the estimated OTWT distributions are consistent with those in the reformer CFD data as shown in Fig. 25 and the maximum, average and standard deviation of the differences between the pairs as shown in Table 3 are nearly negligible compared to the corresponding average OTWT in each data set. It is noted that the maximum, average and standard deviation of the differences in the OTWT distributions between the reformer CFD data and estimates generated from the data-driven model for the OTWT distribution in Table 3 and Table 4 are similar. Therefore, it can be concluded that the data-driven model for the OTWT distribution also has a high out-of-sample prediction performance and provides an excellent description for the dependence of the OTWT distribution on the FSF distribution, total FSF flow rate and interactions among neighboring reforming tubes.

8. Conclusion

The present work developed an integrated model identification scheme that was able to derive a high-fidelity model for the dependence of the OTWT distribution on the FSF distribution, total FSF flow rate and interactions among neighboring reforming tubes from the reformer data. To this end, a high-fidelity reformer CFD model that had been developed and rigorously validated by typical plant data from the SMR literature and actual plant data from our third-party collaborator in our previous work was utilized to facilitate the generation of the reformer database, which was split into the training and testing data. Then, we used Bayesian variable selection, Bayesian model averaging, the BIC approximation, sparse nonlinear regression and theories of thermal radiation to develop the prediction step algorithm in the integrated model identification, from which the prediction model for the OTWT distribution that estimated a predicted OTWT distribution based on a FSF distribution and total FSF flow rate was generated using the training data. Next, we used ordinary Kriging to develop the correction step algorithm in the integrated model identification, from which the correction model for the OTWT distribution that estimated a corrected OTWT distribution based on a predicted OTWT distribution was derived using the training data. Thereafter, we created the data-driven model for the OTWT distribution as a weighted average of the prediction and correction models previously derived from data.

One of our primary interests regarding this work is integrating it in the development of an on-line robust furnace balancing optimizer, which searches for the optimized valve distribution to deliver an optimized FSF distribution and total FSF flow rate to improve the reformer thermal efficiency and compensate for impacts of disturbances on the reformer efficiency. Therefore, it was important that the integrated model identification scheme is computationally efficient so that the optimized FSF distribution and total FSF flow rate can be promptly generated to prevent disturbances from damaging reforming tubes and reducing the reformer service life. In this effort, the prediction and correction steps were structured to be entirely parallelized; specifically, the prediction and correction models of 336 reforming tubes could be derived simultaneously from the training data and independently of one another. This feature allowed the optimal LASSO parameter and the most suited theoretical semivariogram model for each reforming tube, and the optimal weighting factor of the prediction estimates to be determined using leave-out-one cross validation, which was demonstrated to improve the prediction performance of the data-driven model for the OTWT distribution. Finally, the results from the goodness-of-fit and out-of-sample prediction tests of the data-driven model for

the OTWT distribution generated from the integrated model identification scheme demonstrated the high effectiveness of the method proposed in this work. In future work, the integrated model identification scheme can be used to develop an advanced furnace balancing scheme that simultaneously optimizes the FSF distribution and maximizes the total FSF flow rate to decrease the degree of temperature nonuniformity inside the reformer and to increase the reformer efficiency without damaging the reforming tubes and reducing the reformer service life. The ability to adjust the total mass flow rate for the advanced furnace balancing scheme is of special interest for the hydrogen manufacturing industry as it can potentially lead to substantial savings in the re-tubing cost of a reformer.

Acknowledgements

Financial support from the Department of Energy is gratefully acknowledged.

References

Barreto, L., Makihira, A., Riahi, K., 2003. The hydrogen economy in the 21st century: a sustainable development scenario. Int. J. Hydrogen Energy 28, 267–284.

Bilonick, R.A., 1983. Risk qualified maps of hydrogen ion concentration for the New York state area for 1966–1978. Atmos. Environ. 17, 2513–2524.

Brunton, S.L., Proctor, J.L., Kutz, J.N., 2016. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proc. Natl. Acad. Sci. U.S.A. 113, 3932–3937.

Cressie, N., 1985. Fitting variogram models by weighted least squares. J. Int. Assoc. Math. Geol. 17, 563–586.

Ewan, B., Allen, R., 2005. A figure of merit assessment of the routes to hydrogen. Int. J. Hydrogen Energy 30, 809–819.

Ferreira-Aparicio, P., Benito, M., Sanz, J., 2005. New trends in reforming technologies: from hydrogen industrial plants to multifuel microreformers. Catal. Rev. 47, 491–588.

Fragoso, T.M., Neto, F.L., 2015. Bayesian model averaging: a systematic review and conceptual classification, Available form: arXiv:1509.08864.

Gupta, R.B., 2008. Hydrogen fuel: production, transport, and storage. CRC Press.

Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. Stat. Sci. 14, 382–401.

Holdaway, M.R., 1996. Spatial modeling and interpolation of monthly temperature using kriging. Clim. Res., 215–225.

Hydrogen Production Expert Panel, 2013. Report of the Hydrogen Production Expert Panel: A Subcommittee of the Hydrogen & Fuel Cell Technical Advisory Committee. Technical Report. United States Department of Energy.

Jeffreys, H., 1935. Some tests of significance, treated by the theory of probability. Math. Proc. Cambridge Philos. Soc. 31, 203–222.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Kroschwitz, J.I., Howe-Grant, M. (Eds.), 1999. Kirk-Othmer Encyclopedia of Chemical Technology. John Wiley and Sons Inc., New York, NY.

Kumar, A., Baldea, M., Edgar, T.F., Ezekoye, O.A., 2015. Smart manufacturing approach for efficient operation of industrial steam-methane reformers. Ind. Eng. Chem. Res. 54, 4360–4370.

Latham, D.A., McAuley, K.B., Peppley, B.A., Raybold, T.M., 2011. Mathematical modeling of an industrial steam-methane reformer for on-line deployment. Fuel Process. Technol. 92, 1574–1586.

Madigan, D., Raftery, A.E., 1994. Model selection and accounting for model uncertainty in graphical models using Occam's window. J. Am. Stat. Assoc. 89, 1535–1546.

Madigan, D., Raftery, A.E., York, J.C., Bradshaw, J.M., Almond, R.G., 1994. Strategies for graphical model selection. In: Cheeseman, P., Oldford, R.W. (Eds.), Selecting Models From Data: AI and Statistics IV. New York Springer-Verlag, 91.

Maximov, A., 2012. Thesis for the Degree of Doctor of Science: Theoretical Analysis and Numerical Simulation of Spectral Radiative Properties of Combustion Gases in Oxy/air-fired Combustion Systems. Lappeenranta University of Technology.

Olivieri, A., Vegliò, F., 2008. Process simulation of natural gas steam reforming: fuel distribution optimisation in the furnace. Fuel Process. Technol. 89, 622–632.

Pantoleontos, G., Kikkinides, E.S., Georgiadis, M.C., 2012. A heterogeneous dynamic model for the simulation and optimisation of the steam methane reforming reactor. Int. J. Hydrogen Energy 37, 16346–16358.

Raftery, A.E., Madigan, D., Volinsky, C.T., 1996. Accounting for model uncertainty in survival analysis improves predictive performance. Bayesian Stat. 5, 323–349.

Tran, A., Aguirre, A., Crose, M., Durand, H., Christofides, P.D., 2017a. Temperature balancing in steam methane reforming furnace via an integrated CFD/data-based optimization approach. Comput. Chem. Eng. 104, 185–200.

Tran, A., Aguirre, A., Durand, H., Crose, M., Christofides, P.D., 2017b. CFD modeling of a industrial-scale steam methane reforming furnace. Chem. Eng. Sci. 171, 576–598.

Wilson, Z.T., Sahinidis, N.V., 2017. The ALAMO approach to machine learning. Comput. Chem. Eng. 106, 785–795.

Zheng, D., Wu, B., Fleitz, J., Trajkovski, R., Zhou, C.Q., 2010. CFD simulation of a hydrogen reformer furnace. In: Proceedings of the 14th International Heat Transfer Conference. American Society of Mechanical Engineers, pp. 233–244.