



RESEARCH ARTICLE

Transfer learning for predicting source terms of principal component transport in chemically reactive flow

Ki Sung Jung^{1,2} , Tarek Echekki³ , Jacqueline H. Chen¹ and Mohammad Khalil¹

¹Combustion Research Facility, Sandia National Laboratories, Livermore, CA, USA

²Department of Mechanical Engineering, Pukyong National University, Busan, Republic of Korea

³Department of Mechanical and Aerospace Engineering, North Carolina State University, Raleigh, NC, USA

Corresponding author: Ki Sung Jung; Email: kjung@pknu.ac.kr

Received: 10 April 2024; **Revised:** 05 September 2024; **Accepted:** 07 October 2024

Keywords: artificial neural network; chemical kinetics; principal component analysis; reduced order modeling; transfer learning

Abstract

Transfer learning has been highlighted as a promising framework to increase the accuracy of the data-driven model in the case of data sparsity, specifically by leveraging pretrained knowledge to the training of the target model. The objective of this study is to evaluate whether the number of requisite training samples can be reduced with the use of various transfer learning models for predicting, for example, the chemical source terms of the data-driven reduced-order modeling (ROM) that represents the homogeneous ignition of a hydrogen/air mixture. Principal component analysis is applied to reduce the dimensionality of the hydrogen/air mixture in composition space. Artificial neural networks (ANNs) are used to regress the reaction rates of principal components, and subsequently, a system of ordinary differential equations is solved. As the number of training samples decreases in the target task, the ROM fails to predict the ignition evolution of a hydrogen/air mixture. Three transfer learning strategies are then applied to the training of the ANN model with a sparse dataset. The performance of the ROM with a sparse dataset is remarkably enhanced if the training of the ANN model is restricted by a regularization term that controls the degree of knowledge transfer from source to target tasks. To this end, a novel transfer learning method is introduced, Parameter control via Partial Initialization and Regularization (PaPIR), whereby the amount of knowledge transferred is systemically adjusted in terms of the initialization and regularization schemes of the ANN model in the target task.

Impact Statement

The training dataset of reduced-order modeling (ROM) for chemically reactive flows comprises multidimensional numerical simulations. Still the number of high-quality training samples could be sparse under practical scenarios. To resolve this issue, we adopt four different transfer learning methods to a data-based ROM for predicting the ignition of a homogeneous hydrogen/air mixture with a sparse dataset. The present work illustrates that the requisite amount of training samples for neural network model training can be remarkably reduced with the use of regularization-based transfer learning methods. Furthermore, a novel transfer learning model, PaPIR, is introduced in the present study. The PaPIR provides a unified transfer learning framework in terms of initialization and regularization of the neural network model.

1. Introduction

With the continuous advancement of chemical kinetic mechanisms, detailed chemical mechanisms of large-hydrocarbon fuels can be comprised of thousands of species and tens of thousands of elementary



chemical reactions (Lu and Law, 2009). The sheer size of the mechanisms and their wide range of temporal scales are major challenges for high-fidelity numerical simulations of turbulent reacting flows with large-hydrocarbon fuels. Various approaches have been developed to reduce the number of variables in composition space. A conventional way is to develop skeletal/reduced chemical kinetic mechanisms, in which key species and elementary chemical reactions are extracted from a detailed mechanism using techniques such as directed relation graph (DRG) (Lu and Law, 2005), DRG with error propagation (Pepiot-Desjardins and Pitsch, 2008), computational singular perturbation (Lam and Goussis, 1994), and path flux analysis (Sun et al., 2010), thereby reducing the overall size and computational cost of simulations.

More recently, a data-based dimensionality reduction method has also been applied to chemically reactive flows, where a low-dimensional manifold of the original thermochemical state variables is defined based on data-based dimensionality reduction techniques, including linear and non-linear principal component analysis (PCA) (Parente et al., 2009; Sutherland and Parente, 2009; Mirgolbabaee and Echehki, 2013, 2014). The distinct features of the data-driven technique compared with physics-based low-dimensional manifolds, such as the steady laminar flamelet model (Peters, 1983, 1984), the unsteady flamelet/progress variable approach (Pierce and Moin, 2004; Ihme et al., 2005), and flamelet generated manifolds (van Oijen and de Goey, 2000), are that the correlations of the thermochemical state vector are identified by a “training dataset” that is prepared a priori, and the rank of the low-dimensional manifold can be easily adjusted by the user depending on the trade-off between the compression ratio and accuracy. Either linear mapping or a non-linear regression method (e.g., artificial neural network [ANN] and Gaussian process regression [GPR]) is employed for the closure of the governing equations (Sutherland and Parente, 2009; Isaac et al., 2015). A PCA-based reduced-order modeling (ROM) has shown its applicability to replicate characteristics of turbulent flames through a priori (Sutherland and Parente, 2009; Mirgolbabaee and Echehki, 2013; Parente and Sutherland, 2013; Dalakoti et al., 2021) and a posteriori evaluations (Biglari and Sutherland, 2015; Echehki and Mirgolbabaee, 2015; Owoyele and Echehki, 2017; Malik et al., 2021, 2022; Abdelwahid et al., 2023; Kumar et al., 2023; Jung et al., 2024).

Despite the advantages of the data-based ROM for reactive flow simulations, one of the drawbacks of the model is associated with its strong dependency on the quality of training data. For instance, Owoyele and Echehki performed two-dimensional (2D) and three-dimensional (3D) surrogate direct numerical simulations (DNS) of a premixed methane/air flame in a vortical flow with the transport of principal components (PCs) (Owoyele and Echehki, 2017), revealing that a low-dimensional manifold defined from a one-dimensional (1D) training dataset fails to reproduce 2D flame characteristics because of missing curvature effects in the 1D training dataset. Our previous study of 2D surrogate DNS on the compression ignition of large hydrocarbon fuels with the data-based ROM (Jung et al., 2024) showed that the performance of the ROM significantly degrades if the initial temperatures between the training and target dataset are different. Dalakoti et al. (2021) also pointed out that a PCA-based ROM constructed using either a zero-dimensional (0D) homogeneous reactor or a 1D non-premixed igniting flamelet dataset is unable to fully represent the heat release characteristics of a 3D spatially developing turbulent *n*-dodecane jet flame at high-pressure conditions. These findings indicate that a data-driven ROM for chemically reactive flows requires high-quality training data, usually obtained by carrying out multidimensional simulations with a detailed chemical kinetic mechanism, to reproduce the characteristics of the full-order modeling (FOM) accurately. However, given that one of the main purposes of adopting ROMs for reactive flow simulations is to alleviate the computational cost, an argument can be made that it would be impractical to always obtain a sufficient number of high-quality training samples whenever operating conditions of a combustion system change (e.g., temperature, Reynolds number, and turbulent intensity). In other words, the amount of high-quality training data, necessary to optimize a ROM for chemically reactive flows with limited computational resources, can be sparse under practical conditions.

In the machine learning community, transfer learning has been highlighted as a promising framework to improve performance in the case of data sparsity, together with providing a robust initialization scheme and speeding up the learning process (Pan and Yang, 2009). The central idea of transfer learning in the context of machine learning is that a pretrained machine learning model, optimized with a sufficient

number of training samples, is utilized to train a target machine learning model that has sparse training data. Numerous studies demonstrate that the performance of the machine learning model with a sparse dataset is remarkably enhanced by applying transfer learning for clustering (Yang et al., 2009; Mieth et al., 2019; Wang et al., 2021), classification (Hosny et al., 2018; Quattoni et al., 2008; Yao and Doretto, 2010; Zhu et al., 2011), and regression cases (Salaken et al., 2019; Subel et al., 2021; Liu et al., 2022; Yang et al., 2022).

There are different ways of “transferring” knowledge from the previous model (or source model) to the target model, such as instance-based algorithms, feature-based algorithms, model-based algorithms, and relation-based algorithms (Yang et al., 2020). In the present study, a model-based transfer learning algorithm, also known as parameter-based transfer learning (Yang et al., 2020), is adopted to utilize the parameters obtained from the previous machine learning model for the optimization of the target model with sparse training samples. The straightforward way is to freeze all (or some of) the parameters of the target machine-learning model with those obtained from the pretrained model (Pan and Yang, 2009). The parameters of the previous machine learning model can also be used as an initial guess of the parameter values in the target machine learning model. A regularization-based transfer learning method has recently been introduced (Li et al., 2018, 2019; De and Doostan, 2022) in which the knowledge of the previous machine learning model can be “partially” transferred to the target model by adjusting the magnitude of the regularization parameter.

The main objective of the present study is to investigate the possibility of alleviating the requisite number of training samples for optimizing data-driven ROMs for chemically reactive flows by utilizing different transfer learning methods. It has been shown that an accurate prediction of the source term is one of the most challenging parts of the framework of data-driven ROMs (Dalakoti et al., 2021). Therefore, the main focus of the present study is to utilize transfer learning methods to mitigate the requisite number of training samples in the prediction of the 0D homogeneous ignition of a hydrogen/air mixture in a constant volume reactor. While the PC-transport ROM has non-stiff transport terms (Sutherland and Parente, 2009), these terms can be easily predicted by using a shallow neural network model, compared with the more complex neural network model required for predicting the chemical source term (Owoyele and Echeikki, 2017; Kumar et al., 2023).

The dimensionality of the hydrogen/air mixture in composition space is reduced by applying PCA, and the non-linear relationship between the PCs and their reaction rates is mapped by optimizing ANN models. The effect of the number of training samples on the performance of the data-driven ROM is first investigated, and subsequently, different transfer learning approaches are adopted to predict the reaction rates of the PCs with a sparse dataset. To this end, we introduce a novel transfer learning method called “Parameter control via Partial Initialization and Regularization (PaPIR),” where the amount of knowledge transferred from source to target ANN model can be systemically adjusted for the initialization and regularization of the target ANN model.

The outline of the article is as follows. **Section 2** presents the details of the data-driven ROM, ANN models, and various transfer learning methods. **Section 3** illustrates the results of the PCA-based data-driven ROM for the 0D ignition process of a hydrogen/air mixture with various initial conditions depending on the number of training samples. Transfer learning is not applied thus far to highlight the importance of the training data on the performance of the model. In **Section 4**, four different transfer learning methods are utilized for the training of the ANN model with a sparse dataset, from which the performance of the transfer learning methods for various target tasks is evaluated.

2. Methodology

Homogeneous ignition of a hydrogen/air mixture in a constant volume reactor is predicted by applying PCA-based data-driven ROM (PC-transport ROM). Since the integration of numerically stiff chemistry is a bottleneck for many reactive flow simulations, it is reasonable to consider that the present homogeneous reactor configuration is an important benchmark case for evaluating the efficiency of PC-transport ROM to accurately reproduce reactive flow simulations.

The temporal evolution of the original thermochemical-state vector with different initial conditions is first collected by performing a series of 0D simulations of the homogeneous hydrogen/air mixture, and subsequently, a low-dimensional manifold is defined by applying PCA to the collected data. Here, the new variables defined by PCA are denoted as the PCs. The reaction rates of the PCs are predicted as a function of PCs using an ANN. After training the ANN model on one task, the knowledge of the trained ANN model is transferred to another task where the size of the training data is forced to be sparse. The performance of the different transfer learning methods is then systemically investigated by varying (1) the task similarity between the source and target tasks, and (2) the degree of data sparsity in the target task. The methodology of these investigations is described in this section.

2.1. Zero-dimensional ignition dataset for a homogeneous hydrogen/air mixture

In a spatially homogeneous constant volume reactor, the temporal evolution of species and temperature starting from the initial time, $t = 0$, is computed by solving the system of ordinary differential equations (ODEs) defined by

$$\frac{d\theta}{dt} = \dot{\omega}_{\theta}, t \in [0, t_f] \quad (2.1)$$

where θ represents the thermochemical state vector (i.e., species mass fraction and temperature), $\dot{\omega}_{\theta}$ the reaction rate vector divided by the mixture density, and t_f the end time. For the initial conditions, the initial pressure of the system, p_0 , is fixed to be atmospheric, and different values for the initial temperature, T_0 , are used including 1000, 1050, 1100, 1300, and 1400 K. The initial mass fractions of the hydrogen/air mixture are determined by an equivalence ratio, ϕ , which ranges from 0.1 to 3.0. A detailed chemical kinetic mechanism for hydrogen/air mixtures, developed by Li et al. (2004), is used where the dimension of the original thermochemical state vector is 10. A six-stage, fourth-order Runge–Kutta method (Kennedy and Carpenter, 1994) with a uniform time step, dt , of 0.2 ns is adopted for time integration. The CHEMKIN library (Kee et al., 1996) is used to compute the chemical kinetics and thermodynamic properties of the mixture.

Figure 1 shows the ignition delay time, τ_{ig} , of the hydrogen/air mixture for various ϕ and T_0 . As expected, the variation in τ_{ig} exhibits a “U”-shaped profile as a function of ϕ . Here, τ_{ig} is defined as the time at which the temperature gradient is maximum. In accordance with Arrhenius Law, τ_{ig} notably changes with changes in T_0 .

In the present study, the objective of the PC-transport ROM is to replicate the ignition characteristics of the hydrogen/air mixture over a wide range of ϕ at a specific T_0 , meaning that training samples, the low-dimensional manifold, and the corresponding ANN model are separated by T_0 . To reasonably provide a data-sparse scenario, an underlying assumption of the present study is that there exists a sufficient number of training samples spanning over ϕ at T_0 of 1000 K, while the training data size for the cases where $T_0 > 1000$ K is assumed to be sparse (see the symbols in Fig. 1 as an example).

Specifically, at $T_0 = 1000$ K, the training dataset is collected by carrying out 30 different 0D simulations varying ϕ (i.e., $\Delta\phi = 0.1$; ϕ ranging from 0.1 to 3.0), and then the low-dimensional manifold is defined by applying PCA to the training dataset. Training of the ANN model by using a sufficient number of training samples at $T_0 = 1000$ K is considered as the “source task” for the present study. Here, N_{ϕ} is defined as the number of 0D simulations at a given T_0 such that N_{ϕ} of the source task is 30. For the “target tasks” where T_0 is higher than 1000 K (i.e., 1050, 1100, 1300, and 1400 K), N_{ϕ} is set to be ≤ 4 , such that the number of training data for the target task is forced to be sparse. In this study, a “sparse dataset” refers to a dataset with insufficient training samples such that the corresponding ANN model is unable to replicate the ignition characteristics of a fuel/air mixture with a wide range of ϕ (i.e., $\phi = 0.1 - 3.0$) at a given T_0 . The description of the dataset with different N_{ϕ} is summarized in Table 1. Note that the range of T_0 (i.e., $1000 \leq T_0 \leq 1400$ K) is carefully selected to ensure that the overall distributions of the input and

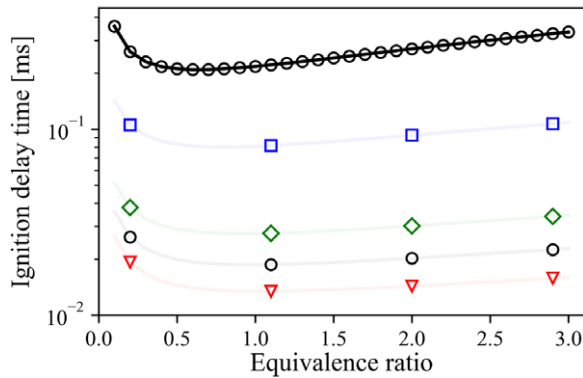


Figure 1. Variations in 0D ignition delay time, τ_{ig} , of the hydrogen/air mixture for different initial temperatures, T_0 , as a function of equivalence ratio, ϕ . In the present study, it is assumed that the number of training samples at the source task ($T_0 = 1000$ K) is sufficient, while the number of training samples at the target tasks ($T_0 > 1000$ K) is sparse.

Table 1. Description of the dataset with different N_ϕ

N_ϕ	ϕ	Number of samples (M)
2	0.5, 1.5	40,000
3	0.5, 1.5, 2.5	60,000
4	0.2, 1.0, 2.0, 3.0	80,000
30	0.1–3.0 ($\Delta\phi = 0.1$)	600,000

output datasets between the source and target tasks are similar, yet noticeably different. This difference becomes more pronounced as T_0 increases in the target task.

While the present study provides a data-sparse scenario based on a series of 0D simulations, such an imbalance in the number of training samples can also be observed from multidimensional simulations and experiments (Humbird et al., 2020; Subel et al., 2021). It is also noted that transfer learning methods used in the present study are not limited to specific source and target tasks. Rather, these methods can be applied to various scenarios (e.g., different pressure or equivalence ratio conditions between source and target tasks), provided that there is a task similarity between the two tasks.

For each 0D simulation, the thermochemical state vector and their reaction rate, θ and $\dot{\omega}_\theta$, respectively, are uniformly sampled from t of 0 to $2\tau_{\text{ig}}$. The number of samples for each 0D simulation is set to be 20,000 such that the first 10,000 samples are assigned to the preignition period and the remaining 10,000 samples are related to the postignition period. The “test dataset” at a given T_0 is also prepared to evaluate the accuracy of the PC-transport ROM. It consists of 29 different 0D simulation results at a given T_0 ($\Delta\phi = 0.1$; ϕ ranging from 0.15 to 2.95) and is separated from the training dataset.

2.2. Principal component analysis

Consistent with previous studies applying PC-transport ROMs (Parente et al., 2009; Sutherland and Parente, 2009; Mirgolbabaee and Echehki, 2013, 2014), the dimension of the original thermochemical vector is reduced by applying PCA. Assuming that M number of samples of the N -dimensional thermochemical state vectors are collected by performing multiple 0D simulations at a given T_0 , the dataset of the thermochemical vector, $\Theta = [\theta_1, \theta_2, \dots, \theta_M]$, is composed of a $N \times M$ dimensional matrix. Note that Θ is normalized based on the min–max normalization before PCA. The $N \times N$ dimensional

matrix of orthonormal eigenvectors, \mathbf{Q}^T , of the covariance matrix of Θ is constructed, and subsequently, the dataset of the PC vector, Ψ , can be defined as

$$\Psi = \mathbf{Q}^T \Theta \quad (2.2)$$

where $\Psi \in \mathbb{R}^{N \times M}$ represents the M numbers of collections of the PC vector, $\psi = [\psi_1, \psi_2, \dots, \psi_N]^T$.

Note that the first PC, ψ_1 , is a linear combination of the original thermochemical state vector that captures the maximum variance of the dataset. The second PC, ψ_2 , is then orthogonal to the first PC, and all the subsequent PCs follow the same concept. In the present study, the leading first five PCs (i.e., $N_{\text{PC}} = 5$) are retained from ψ such that the dimensionality of the system is reduced from 9 (except for N_2) to 5, which captures over 99% of the original total variance. In other words, a $N \times N_{\text{PC}}$ matrix of \mathbf{A} is constructed that contains the leading N_{PC} eigenvectors from \mathbf{Q} . The low-dimensional manifold then becomes

$$\Psi^{\text{red}} = \mathbf{A}^T \Theta \quad (2.3)$$

where $\Psi^{\text{red}} \in \mathbb{R}^{N_{\text{PC}} \times M}$ represents the dataset of the truncated PC vector, $(\psi^{\text{red}}) = [\psi_1, \psi_2, \dots, \psi_{N_{\text{PC}}}]^T$. Hereinafter, Ψ^{red} and ψ^{red} are referred to as Ψ and ψ , respectively, for the sake of brevity.

The system of ODEs for the low-dimensional manifold can be defined by projecting Eq. 2.1 on the matrix \mathbf{A}^T :

$$\frac{d\psi}{dt} = \dot{\omega}_\psi, t \in [0, t_f] \quad (2.4)$$

where $\dot{\omega}_\psi$ is the reaction rate term for ψ , defined by $\dot{\omega}_\psi = \mathbf{A}^T \dot{\omega}_\theta$ (Sutherland and Parente, 2009). In the framework of the PC-transport ROM, the time integration of Eq. 2.4 is solved instead of solving Eq. 2.1, and then a conversion from ψ into θ is carried out as a postprocessing step. An ANN model is used for the regression of $\dot{\omega}_\psi$ as a function of ψ .

Figure 2 shows the PC modes with respect to the original thermochemical vector defined by using different training datasets in terms of T_0 . For all the cases, N_ϕ is set to be 30. It is readily observed from the figure that the PC modes show a similar trend irrespective of the dataset, implying that transfer learning can be effectively applied to the dataset. This result is consistent with the previous study (Biglari and Sutherland, 2012) which shows that the PCA basis is insensitive to the spatial filter width in the context of large eddy simulations. Specifically, the first PC mode is negatively correlated with the fuel and oxidizer, while it is positively correlated with the product, H_2O , and temperature. Accordingly, the first PC represents the oxidation progress of the hydrogen/air mixture. The second PC mode is primarily correlated with the fuel, and the third PC mode is correlated with the formation of HO_2 . The results indicate that the PCs obtained through the data-driven approach are linked to a physical interpretation of the combustion system, consistent with previous findings (Owoyele and Echehki, 2017; Malik et al., 2021).

Nonetheless, it is important to note that the PC modes are slightly altered with a change of T_0 , which can have a significant impact on the application of transfer learning to the ROM. In other words, a unified definition of the low-dimensional manifold throughout tasks would be preferred to transfer the knowledge efficiently. In the present study, \mathbf{A}^T defined from the source task (i.e., $N_\phi = 30$ and $T_0 = 1000$ K) is applied to all target tasks to ensure consistency in the definitions of ψ and $\dot{\omega}_\psi$. Such an approach is based on the observation that despite the presence of slight differences in the PC modes, the first PC, which also accounts for most of the data variance, also exhibits the least difference when T_0 is varied.

Note, however, that using a unified \mathbf{A}^T has a potential risk of introducing noticeable errors during the conversion from ψ to θ in the target task, especially if the reconstruction is carried out by using a matrix conversion step, $\theta \approx \mathbf{A}\psi$. To address this issue, another non-linear ANN model is employed to convert from ψ to θ for all cases, instead of using the matrix inversion. This ensures that the performance of the reconstruction is mainly affected by the number of samples, M , in the target task rather than the choice of \mathbf{A}^T . As will be discussed, such an ANN model is found to require far fewer parameters compared with the other ANN models that predict $\dot{\omega}_\psi$. Hence, this ANN model shows a reasonable accuracy even when trained with a sparse dataset. It is also noted that the non-linear ANN was found to reduce the reconstruction error during the conversion from ψ to θ as compared with the matrix inversion, $\theta \approx \mathbf{A}\psi$ (Mirgolbabaei and Echehki, 2015).

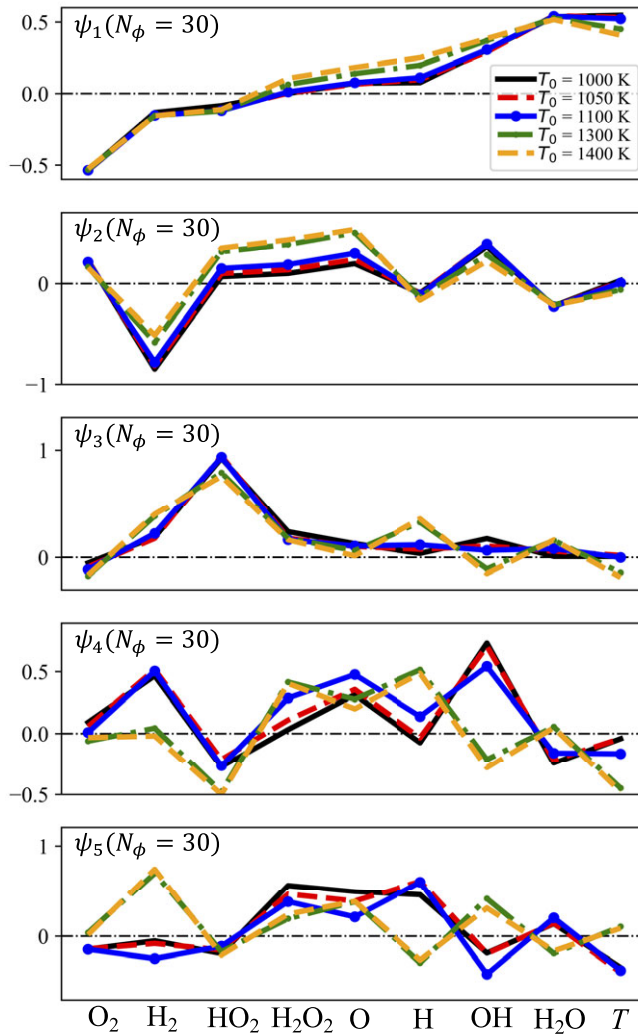


Figure 2. Modes of the first five PCs depending on the training dataset varying T_0 with N_ϕ of 30.

The sensitivity of the ANN model performance depending on the definition of the eigenvector matrix, \mathbf{A}^T , is shown in Fig. S1 in Supplementary Material (SM). It is shown that the effect of a PCA basis (i.e., \mathbf{A}^T) plays a secondary role in determining the optimal ANN model in the target task, while the main source of the error arises from the source term regression errors.

Figure 3 shows the temporal evolution of the first three PCs for three different ϕ of 0.85, 1.35, and 2.95 at T_0 of 1000 K, obtained by projecting the FOM result onto \mathbf{A}^T . As discussed earlier, the first PC represents the progress variable of the mixture, such that the first PC switches from negative to positive values near the ignition delay time. For the third PC, its mode is mainly correlated with the intermediate species, namely HO_2 , such that it is maximum just before ignition of the mixture.

2.3. Artificial neural network

A fully connected, multi-input, and multi-output ANN model is used to predict the reaction rates of the PCs. The PC vector is used as the input of the ANN, and the reaction rates of these PCs are the output. The architecture of the ANN model is determined by performing a grid search method from which the number

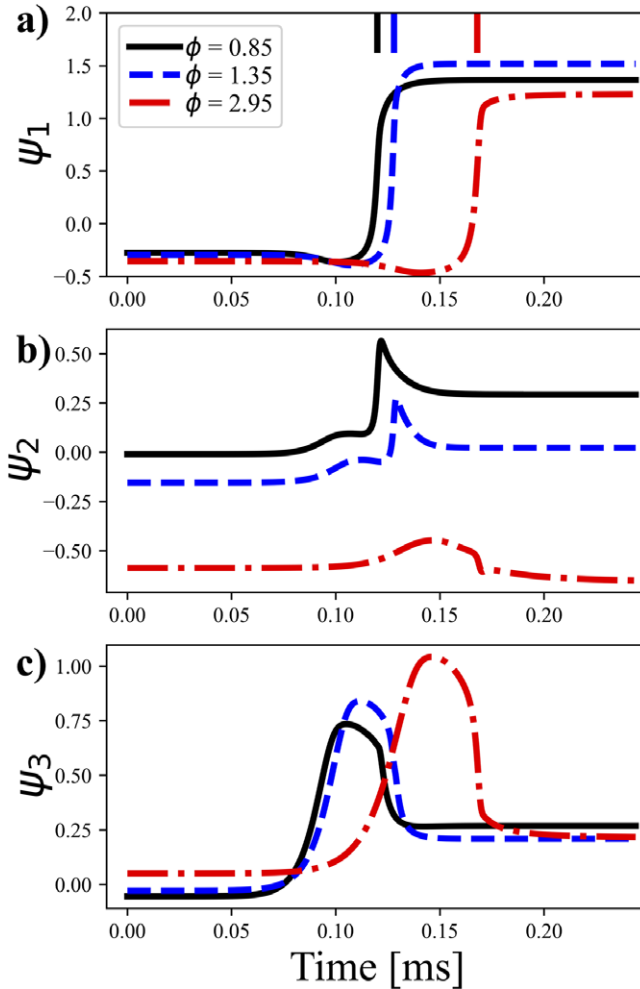


Figure 3. Temporal evolution of the first three PCs for three different equivalence ratios, ϕ , of 0.85, 1.35, and 2.95, obtained by projecting \mathbf{A}^T onto the FOM result. The vertical lines in (a) represent the ignition delay time for different ϕ . Here, the ignition delay time is defined by the time at which the temperature gradient reaches its maximum value.

of hidden layers and nodes are set to 3 and 30, respectively. The hyperbolic tangent activation function is adopted for all hidden layers. For cases without the application of transfer learning methods, the Xavier normal initialization method (Glorot and Bengio, 2010), which is a commonly used initialization scheme and is compatible with the hyperbolic tangent activation function, is employed. Note that the goal of training the ANN model is to optimize weights (\mathbf{w}) and biases (\mathbf{b}) vectors, which together make up the parameter vector, $\mathbf{h} = [\mathbf{w}, \mathbf{b}]$. We apply the early stopping callback function based on the validation loss, and a plateau learning rate scheduler is used to fine-tune the parameters.

As discussed in Section 2.2, another ANN model is also employed and trained for the reconstruction from $\boldsymbol{\psi}$ to $\boldsymbol{\theta}$. One hidden layer with 10 nodes is found to be sufficient for this ANN model to reconstruct the original thermochemical scalars with reasonable accuracy. Note that such an ANN model requires considerably fewer training samples, and, therefore, transfer learning is not applied to this model.

For both source and target tasks, 80% of the dataset is used as a training dataset, and the remaining 20% is allocated as the validation set to assess the model's performance and prevent overfitting. The mean absolute error (MAE) loss function is applied for the ANN training, consistent with previous studies

showing that the choice of the MAE results in a better performance than RMSE and MSE in capturing an ignition process of a fuel/air mixture (Han et al., 2020, 2022). The Adam optimizer (Kingma and Ba, 2014) is used for stochastic optimization. Once the ANN model is optimized using the training dataset, it is both a priori and a posteriori evaluated against the test dataset, which is not involved in the training process.

To efficiently capture the ignition process of the hydrogen/air mixture, the dataset is divided into three clusters (Cl#1–Cl#3): (Cl#1) earlier ignition period ($\psi_3 - \psi_{3,0} < 0.005$ and $\psi_1 < 0.0$), (Cl#2) later ignition period ($\psi_3 - \psi_{3,0} \geq 0.005$ and $\psi_1 < 0.0$), and (Cl#3) post ignition period ($\psi_1 \geq 0.0$), where the $\psi_{3,0}$ denotes the magnitude of ψ_3 at the initial condition. The clustering criteria are based on the observation that ψ_1 and ψ_3 effectively represent the progress variable and evolution of intermediate species, respectively, as depicted in Figure 3. For more complicated numerical configurations, unsupervised clustering algorithms (e.g., k -means algorithm or Vector Quantization Principal Component Analysis [VQPCA]) could be adopted to automatically determine the clustering criteria, as described in (D'Alessio et al., 2023; Savarese et al., 2024). Note that the data clustering method has been proven as an effective way to capture the ignition process of various fuel/air mixtures (Han et al., 2020; Han et al., 2022; Jung, Kumar, et al., 2024). The performance of the ROM will later be evaluated in each cluster.

2.4. Transfer learning methods

In this study, four different transfer learning methods are applied to the target tasks. Let \mathbf{h}^s denote the parameter vector (weights and biases vectors) extracted from the pretrained source task. The first transfer learning method (TL1) is that the knowledge of the pretrained ANN model obtained from the source task is fully shared with the target task. In other words, the parameter vector in the target task, denoted by \mathbf{h} , is identical to \mathbf{h}^s , and no fine-tuning step is performed in TL1. Therefore, it can be conjectured that TL1 is likely to show good performance only when the task similarity between the source and target is very high. The second transfer learning method (TL2) is to set the initial parameter vector in the target task, \mathbf{h}_0 , with \mathbf{h}^s , and then fine-tune the model using the sparse dataset in the target task. In other words, TL2 serves to initialize the ANN model in the target task by using \mathbf{h}^s , and as such, \mathbf{h} will be different from \mathbf{h}^s after fine-tuning.

As discussed in Li et al. (2029), a drawback of TL2 is that previous knowledge obtained from the source task may be lost during the fine-tuning step. To resolve this issue, the third transfer learning method (TL3), which is associated with parameter restriction, is applied. The total loss function, \mathcal{L} , in TL3 includes a regularization term (X. Li et al., 2018, 2029), which is slightly different from the conventional l_2 regularizer, as follows:

$$\mathcal{L} = \text{MAE} + \lambda_1 \|\mathbf{h} - \mathbf{h}^s\|_2^2. \quad (5)$$

In this expression, MAE represents the mean absolute error loss function, and λ_1 is the regularization parameter. Here, the magnitude of λ_1 mainly controls the degree of knowledge transferred from source to target task during the fine-tuning step in the target task. Consistent with TL2, \mathbf{h}_0 in TL3 is set to be \mathbf{h}^s , and subsequently, the ANN model is fine-tuned by using the dataset in the target task. It is evident that \mathbf{h} will be identical to \mathbf{h}^s as the magnitude of λ_1 is very high, whereas there is no penalty for \mathbf{h} to change during the fine-tuning step at $\lambda_1 = 0$. Therefore, it can be considered that TL3 becomes equivalent to TL1 and TL2 as the magnitude of λ_1 becomes very large or approaches zero, respectively.

Lastly, we introduce a novel transfer learning method called “Parameter control via Partial Initialization and Regularization (PaPIR).” The central idea of PaPIR is to provide a unified transfer learning framework in terms of initialization and regularization. In addition to applying λ_1 to adjust the effect of the regularization during training, another variable, λ_2 , is introduced in PaPIR so that the amount of previous knowledge transferred to the target task in terms of the initialization can also be controlled by changing the magnitude of λ_2 . The initialization method of the PaPIR is a combination of two initialization schemes, namely Xavier normal initialization (Glorot and Bengio, 2010) and initialization with \mathbf{h}^s . Xavier normal initialization is a family of the Gaussian-based initialization technique with zero mean and a determined variance. Thus, it is a

sort of random initialization strategy and is unrelated to the pretrained knowledge. On the other hand, initialization with \mathbf{h}^s is categorized as a data-driven initialization strategy (Narkhede et al., 2022).

The initialization process in PaPIR follows a normal distribution function, N , which is expressed as follows:

$$\mathbf{w}_0 = N\left(\lambda_2 \mathbf{w}^s, \left[\sqrt{\frac{2}{f_i + f_o}}(1 - \lambda_2)\right]^2\right) \quad (2.6)$$

$$\mathbf{b}_{0,b} = N(\lambda_2 \mathbf{b}^s, 0^2) \quad (2.7)$$

and \mathbf{b}_0 represent the initial weights and biases vector of the ANN model in the target task (i.e., $\mathbf{h}_0 = [\mathbf{w}_0, \mathbf{b}_0]$), and \mathbf{w}^s and \mathbf{b}^s represent the weight and biases vector extracted from the source task (i.e., $\mathbf{h}^s = [\mathbf{w}^s, \mathbf{b}^s]$). Here, f_i and f_o represent the number of incoming and outgoing nodes at each layer, respectively, which are identical to those used in the Xavier normal initialization scheme (Glorot and Bengio, 2010).

As λ_2 in Eqs. 2.6–2.7 approaches zero, the initialization scheme becomes equivalent to the Xavier normal initialization method. As λ_2 approaches unity, on the other hand, \mathbf{h}_0 simply becomes identical to \mathbf{h}^s . Thus, the degree of knowledge transfer for the initialization of the target task can be adjusted by varying the value of λ_2 between zero and unity, which equivalently represents a bound between the Xavier normal initialization method and \mathbf{h}^s , respectively. As demonstrated in Section 3, PC-transport ROM with a sparse dataset generally fails to capture the overall ignition process of a hydrogen/air mixture if the ANN model is trained from scratch. This shortcoming is mainly attributed to the propensity of the ANN models to get stuck in local minima, especially with a sparse dataset. Given that an appropriate initialization scheme can help avoid local minima (Narkhede et al., 2022), PaPIR has the potential advantage of enhancing the performance of transfer learning by introducing some degree of randomness during the initialization process. Table 2 summarizes and compares the four transfer learning methods used in the present study.

3. Results without transfer learning

The results of the PC-transport ROM for predicting the 0D ignition process of hydrogen/air mixture over a wide range of ϕ at a given T_0 are presented. Transfer learning is not applied in this section. For the source task ($T_0 = 1000$ K), a sufficient amount of training data is provided to train the ANN (i.e., $N_\phi = 30$), and thus, the PC-transport ROM is expected to accurately capture the overall ignition characteristics of the hydrogen/air mixture. Subsequently, the effect of the number of training samples on the performance of the PC-transport ROM is investigated by gradually decreasing the number of training samples.

3.1. Source task: $T_0 = 1000$ K

In the source task, the ANN models for predicting the reaction rate of PCs are trained by using the training dataset with N_ϕ of 30 ($\phi = 0.1 - 3.0$; $\Delta\phi = 0.1$) at T_0 of 1000 K. A system of ODEs, Eq. 2.4, is solved for 29 different 0D simulations listed in the test dataset, and then the performance of the PC-transport ROM is evaluated against the FOM by comparing τ_{ig} between two different simulations. As mentioned earlier, τ_{ig} is defined by the time at which the temperature gradient reaches its maximum value, and τ_{ig} in the PC-transport ROM can be predicted after reconstructing the temperature profile from the evolved PC solutions.

Figure 4 shows the variations in τ_{ig} for the hydrogen/air mixture at $T_0 = 1000$ K with various ϕ predicted by the PC-transport ROM and FOM. As shown in the figure, τ_{ig} predicted by the PC-transport ROM shows excellent agreement with the FOM. The relative percentage error is below 2% for the entire range of ϕ , demonstrating that PC-transport ROM with a sufficient number of training samples can accurately reproduce the ignition process of a hydrogen/air mixture over a wide range of ϕ at a given T_0 . Note that the relative percent error for the fuel-lean mixture is slightly higher than that for the fuel-rich

Table 2. Summary of the transfer learning methods used in this study. “ α ” in the PaPIR model represents $\alpha = \sqrt{(2/(f_i + f_o))(1 - \lambda_2)}$

Model	Description	Further training	Objective	Initialization	Similar to
TL1	Parameter sharing	No	–	–	TL3 with $\lambda_1 = \infty$ PaPIR with $\lambda_1 = \infty$
TL2	Fine-tuning	Yes	Target data misfit	$w_0 = w^s$ $b_0 = b^s$	TL3 with $\lambda_1 = 0$ PaPIR with $\lambda_1 = 0, \lambda_2 = 1$
TL3	Parameter restriction	Yes	Target data misfit with $\lambda_1 \ h - h^s\ _2^2$	$w_0 = w^s$ $b_0 = b^s$	PaPIR $\lambda_2 = 1$ with
PaPIR	Parameter control via partial initialization and regularization	Yes	Target data misfit with $\lambda_1 \ h - h^s\ _2^2$	$w_0 =$ $N(\lambda_2 w^s, \alpha^2)$ $b_0 = N(\lambda_2 b^s, 0^2)$	

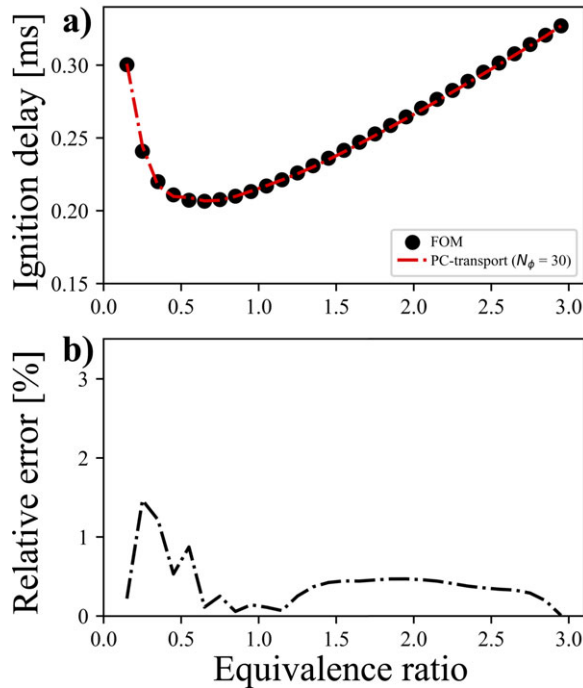


Figure 4. Variations in (a) 0D ignition delay time, τ_{ig} , predicted by FOM (solid symbol) and PC-transport ROM (dashed-dot line), and (b) the relative error of the PC-transport ROM compared with FOM for the homogeneous hydrogen/air mixture with various ϕ (i.e., $\phi = 0.15–2.95$; $\Delta\phi = 0.1$) at $T_0 = 1000$ K.

mixture, which is attributed to the fact that the number of training samples assigned to the fuel-lean mixture is fewer than that assigned to the fuel-rich mixture. Also, τ_{ig} shows a steeper variation with ϕ as ϕ becomes less than 0.5, which also affects the result.

Figure 5 presents the temporal evolution of the original thermochemical state vector of the hydrogen/air mixture at T_0 of 1000 K and ϕ of 1.35, as predicted by the PC-transport ROM and FOM. It is readily observed that the profiles of the thermochemical state variables reconstructed from the PC-transport ROM are in good agreement with the results from FOM for both major and minor species. This finding indicates

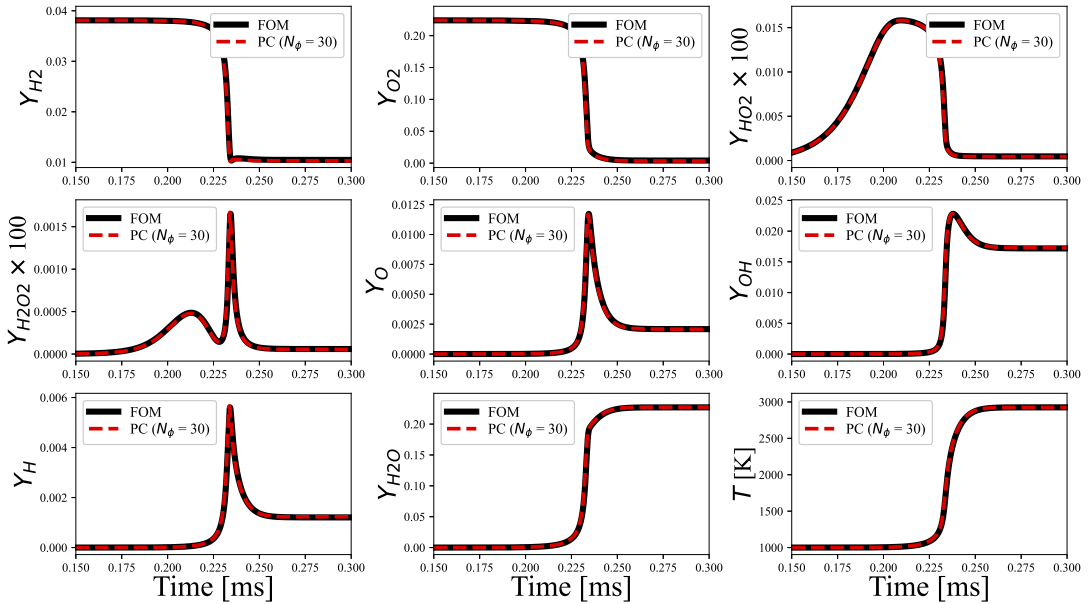


Figure 5. Temporal evolution of the thermochemical state scalars of a homogeneous hydrogen/air mixture at $T_0 = 1000$ K and $\phi = 1.35$. Solid line: FOM result, Dashed line: reconstructed from the PC-transport ROM result with $N_\phi = 30$.

that the number of PCs retained in this study ($N_{PC} = 5$) is sufficient to recover the original thermochemical state scalars, together with the successful validation of using the ANN model to convert from ψ to θ .

3.2. Target task with data sparsity

The results of the PC-transport ROM for a target task where T_0 is 1050 K are investigated depending on the number of training samples. In this task, ANN models are trained by using each of the different training datasets, each with different numbers of training samples (i.e., $N_\phi = 2-30$), and both a priori and a posteriori evaluations are carried out to assess the performance of the ROM depending on N_ϕ . At a given N_ϕ , the ANN model training is repeated 10 times to take into account the sensitivity of the model arising from the randomness of the initial parameters and the stochastic nature of the optimization process. The best achievable (or minimum) error is then evaluated over 10 repetitions at a given N_ϕ . The normalized root mean squared error (NRMSE) is adopted to a priori quantify the error of the ANN for predicting $\dot{\omega}_\psi$ and is defined by

$$\text{NRMSE}[\%] = \sqrt{\frac{\|\widehat{\omega}_{\psi,p} - \widehat{\omega}_\psi\|_2^2}{\|\widehat{\omega}_\psi\|_2^2}} \times 100 \quad (2.8)$$

where $\widehat{\omega}_{\psi,p}$ and $\widehat{\omega}_\psi$ represent the normalized reaction rates of the PC vector predicted by the ANN model and obtained from the FOM, respectively. The minmax normalization method is used to normalize the reaction rates of the PC.

Figure 6 shows the variations in NRMSE of the test set in the target task with T_0 of 1050 K and various N_ϕ . The NRMSE of the test set generally shows a decreasing trend with an increase of N_ϕ , such that the optimal value of NRMSE for the case with $N_\phi = 30$ approaches $O(10^{-1})$ [%] for all the clusters. This outcome clearly indicates that the number of training samples plays a crucial role in determining the performance of the PC-transport ROM. In addition, the variations in the NRMSE as a result of repeating the ANN model training 10 times exhibit a noticeable fluctuation at a given N_ϕ . Consequently, for Cl#1

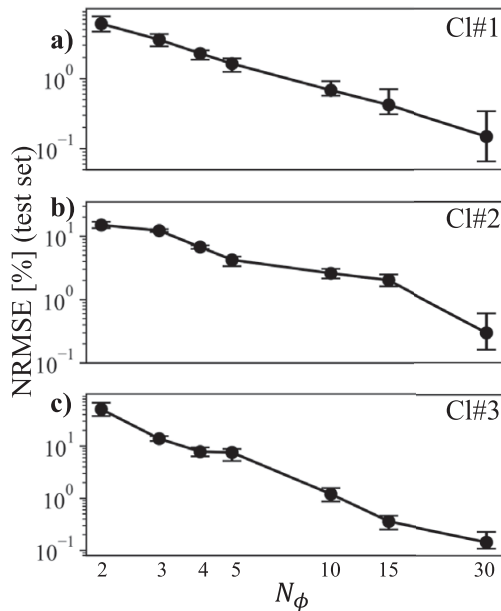


Figure 6. Variations in NRMSE of the test set in the target task with T_0 of 1050 K as a function of N_ϕ for (a) Cluster 1, (b) Cluster 2, and (c) Cluster 3. The closed circle symbol represents the averaged NRMSE obtained from 10 repetitions of the ANN model training.

and CI#3, the worst cases with $N_\phi = 30$ have a similar magnitude of NRMSE compared with the best cases with $N_\phi = 15$. This result suggests that a multistart-based optimization algorithm would be necessary to obtain the nearest optimal neural network model at a given N_ϕ .

Next, a series of 0D simulations is carried out by using the ANN models trained with different numbers of training samples. For the case with $N_\phi = 30$, τ_{ig} is predicted by using the best and worst ANN models out of 10 repetitions of the ANN model training. For the other cases, τ_{ig} is predicted by using the best ANN model only. Figure 7 summarizes the variations in τ_{ig} for the hydrogen/air mixture with T_0 of 1050 K and various ϕ listed in the test set, predicted by using the ANN models with different N_ϕ .

Figure 7 shows that the PC-transport ROM fails to capture the overall ignition characteristics of a hydrogen/air mixture with $N_\phi \leq 15$. In this regard, the datasets with $N_\phi \leq 15$ are regarded as “sparse datasets.” Note that the PC-transport ROM performs relatively well when the target equivalence ratio is adjacent to one of the equivalence ratios listed in the training dataset. For instance, the training dataset with $N_\phi = 3$ consists of the 0D simulation results with ϕ of 0.5, 1.5, and 2.5, where the relative error of the PC-transport ROM is relatively small near ϕ of 0.5, 1.5, and 2.5, while the performance of the PC-transport ROM declines as the target equivalence ratio moves farther from the training dataset. It is also important to note that even when a large amount of training samples are used (i.e., $N_\phi = 30$), the simulation results occasionally do not agree well with the results from the FOM, consistent with the a priori evaluation in Figure 6. This result not only highlights that the number of training samples is a crucial part of optimizing the ANN model but also indicates that the uncertainty of the ANN model training is noticeable and is primarily because of the stochastic nature of the training process and/or the randomness of the initial parameter vectors (i.e., weights and biases).

4. Results with transfer learning

From the previous discussion, the main issues associated with the PC-transport ROM for capturing the reaction rates of PCs with a sparse dataset are twofold: (1) the PC-transport ROM inaccurately predicts the reaction rate of the PCs over a wide range of ϕ because of scarcity of training samples, and (2) a multistart-

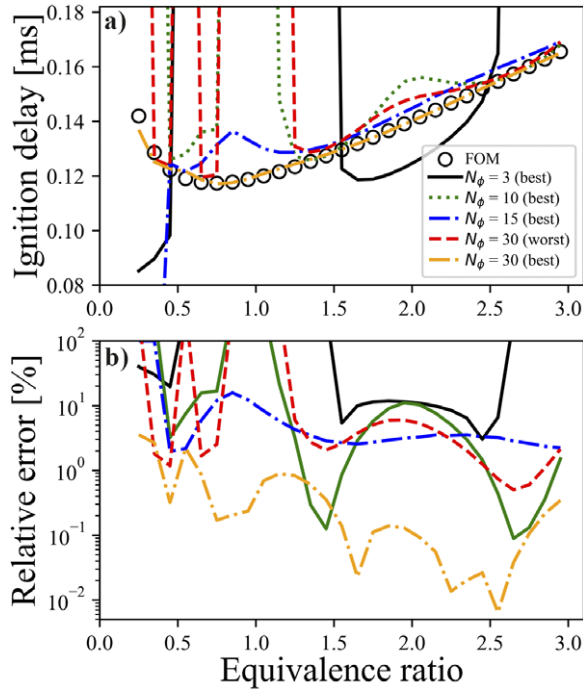


Figure 7. Variations in (a) 0D ignition delay time, τ_{ig} , predicted by the FOM (symbol) and PC-transport ROMs trained using a different number of training samples, and (b) the relative-error of the PC-transport ROMs compared with FOM for a homogeneous hydrogen/air mixture with various ϕ (i.e., $\phi = 0.15 - 2.95$; $\Delta\phi = 0.1$) at $T_0 = 1050$ K.

based optimization strategy is required to figure out the nearest optimal ANN model for a given training dataset. To address these issues, the previous knowledge gained from the source task ($T_0 = 1000$ K and $N_\phi = 30$; see Section 3.1) is transferred to the target tasks. Four different transfer learning methods, TL1, TL2, TL3, and PaPIR, are applied to different target tasks in terms of the task similarity (i.e., T_0 difference between source and target task) and the degree of data sparsity in the target tasks (i.e., N_ϕ in the target task).

4.1. TL1–TL3: general characteristics

As a baseline case, the result of applying three transfer learning methods (TL1, TL2, and TL3) to the target task where $T_0 = 1050$ K and $N_\phi = 4$ is discussed first. In this case, the difference of T_0 between source and target tasks is relatively small (i.e., $\Delta T = 50$ K), such that task similarity between the two tasks is considered to be high. The parameters obtained from the source task are used to train the ANN model in the target task in various ways.

As shown in Table 2, TL3 (parameter restriction) becomes equivalent to TL1 (parameter sharing) and TL2 (fine-tuning) as the value of the regularization parameter, λ_1 , in Eq. 2.5 approaches near infinity and zero, respectively. Thus, the performance of the ANN model using three different transfer learning methods can be evaluated by adjusting the magnitude of λ_1 . Similar to the previous cases where transfer learning is not employed, ANN training is repeated 10 times to evaluate the uncertainty of the ANN model training.

Figure 8 shows the NRMSE values against the training and test sets, along with the percentage differences in the optimized parameters between the source and target tasks, represented by $\|\mathbf{h} - \mathbf{h}^s\|_2^2 / \|\mathbf{h}^s\|_2^2 \times 100$, as a function of λ_1 for the case where $T_0 = 1050$ K and $N_\phi = 4$. Several points are noted in the figure.

First, given that λ_1 serves as a penalty term during the fine-tuning, the NRMSE of the training set generally decreases with a decrease of λ_1 for all the clusters. On the other hand, $\|h - h^s\|_2^2 / \|h^s\|_2^2$ continues to increase as λ_1 decreases, indicating that the parameters in the target task become more dissimilar to those in the source task with a decrease of λ_1 . This trend demonstrates that the magnitude of λ_1 mainly controls the degree of knowledge transfer from the source to the target task.

Second, as the magnitude of λ_1 becomes sufficiently large (i.e., $>10^4$), h becomes nearly identical to h^s , illustrating that the transfer learning method for this case is equivalent to TL1, where the knowledge gained from the source task is fully transferred to the target task. In that case, the NRMSE of the test set can be notably higher than the case without applying transfer learning (see Figure 8b and d as an example), indicating that TL1 would play a negative role in predicting the ignition delay of the hydrogen/air mixture for the target task even if the difference in T_0 is relatively small. This would be primarily attributed to the non-linear nature of chemical kinetics, where reaction rates are highly sensitive to temperature change.

Third, as the magnitude of λ_1 approaches near zero, the corresponding transfer learning method represents TL2, where the parameters obtained from the source task are used to initialize the parameters in the target task. In this scenario, $\|h - h^s\|_2^2 / \|h^s\|_2^2$ is relatively high, indicating that the knowledge acquired from the source task is prone to be lost during the fine-tuning process. Nonetheless, it is worth mentioning that the NRMSE of the test set using TL2 nearly equals that obtained from the best ANN model without applying transfer learning. Furthermore, the results of TL2 show less fluctuation from the 10 repetitions of training compared with the results without applying transfer learning. This suggests that initializing the

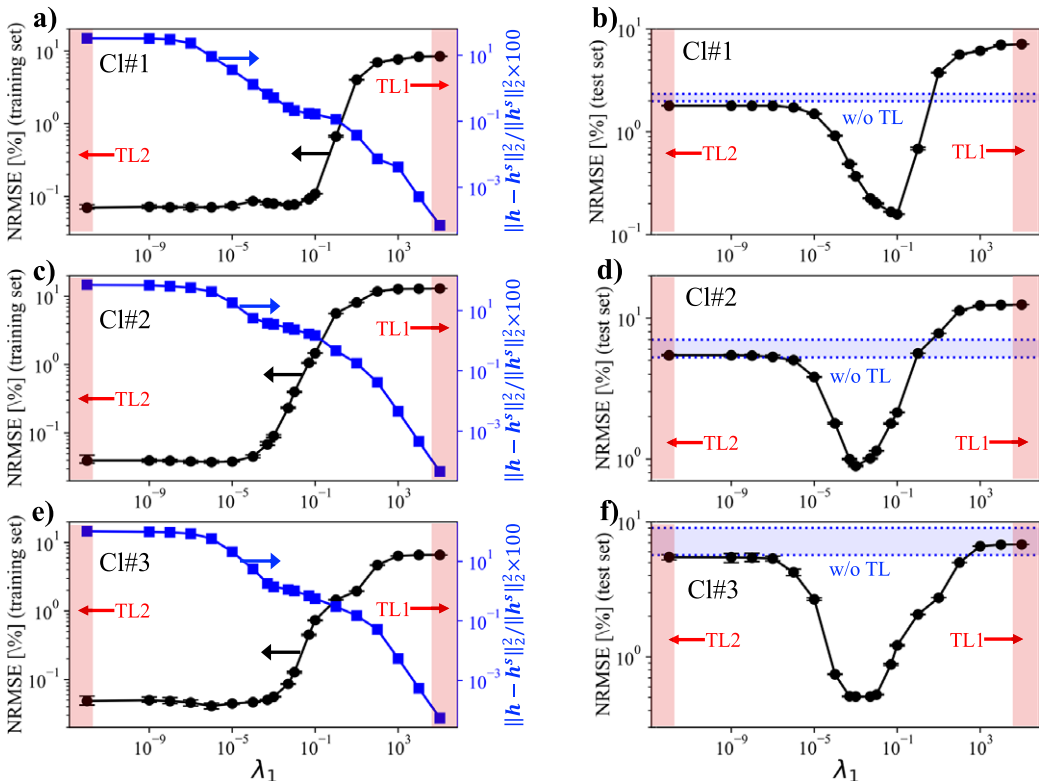


Figure 8. A priori evaluation of (left) the NRMSE for the training set (i.e., $T_0 = 1050$ K, $N_\phi = 4$) and $\|h - h^s\|_2^2 / \|h^s\|_2^2 \times 100$, and (right) the NRMSE for the test set for the target task with $T_0 = 1050$ K and $N_\phi = 29$ as a function of λ_1 . The highlighted regions on the right represent the range of NRMSE of the test set predicted by the PC transport model without applying transfer learning.

ANN model of the target task with the parameters gained from the source task can be considered an appropriate initialization scheme, provided that the similarity between the source and target tasks is high.

Lastly, the NRMSE of the test set reaches its minimum as the magnitude of λ_1 has a finite value (i.e., $\lambda_1 = O(10^{-4} - 10^{-1})$) for all the clusters. In this case, the resultant NRMSE is approximately an order of magnitude lower than that obtained from training the ANN model from scratch, clearly demonstrating that TL3 with an optimal value of λ_1 can remarkably improve the performance of the neural network model for the target task. Note that to achieve a comparable level of NRMSE as that obtained from TL3, the ANN model trained from scratch requires a larger number of N_ϕ , ranging from 15 to 30, indicating that TL3 can reduce the requisite number of training samples up to eight times. Furthermore, the ANN model training with the use of the optimal value of λ_1 is nearly insensitive to the number of training repetitions. This is attributed to the well-known effect of regularization on the stochastic optimization process. Hence, the overall number of training repetitions required to find the optimal ANN model in the target task can be significantly reduced by applying TL3 with an optimal value of λ_1 , provided that the source and target tasks are similar in parameter space.

Based on the observations in Figure 8, it can be inferred that the loss function in the target task is likely to contain multiple local minima such that the training result of the ANN model may not reach the global minima of the loss function, especially with the sparse dataset. One method to address this issue and improve the performance of the ANN model is to utilize the pretrained ANN model. If the source and target tasks are similar to each other in the parameter space, then initializing the parameter in the target task (\mathbf{h}_0) with that from the source task (\mathbf{h}^s) can assist in searching for the optimal parameters. Hence, the performance of the ANN model with TL2 has the potential to show better performance compared with the ANN model trained from scratch. Furthermore, TL3 with the optimal value of λ_1 can help prevent \mathbf{h} from significantly deviating from \mathbf{h}^s during the fine-tuning step, which can enhance the accuracy in predicting τ_{ig} in the target task.

Regarding the computational cost for training ANN models, the present study assumes that the number of training samples in the target task is small, resulting in a significantly lower training cost for the target task compared with the source task. For instance, the wall clock time required to train the source task at $T_0 = 1050$ K with $N_\phi = 30$ is 3150 s for Cl#2, while the wall clock time for training target task for $T_0 = 1050$ K with $N_\phi = 4$ is 328 s when using TL3 with $\lambda_1 = 10^{-4}$ (see Figure S2 in Supplementary Material). This is simply because the number of training samples in the target task is noticeably smaller than that in the source task. Moreover, the convergence rate for training the target task can be further improved if an ANN model trained with a specific λ_1 is used as an initial guess for the training of the next ANN model with a different λ_1 .

It is also worth mentioning that the computational cost of collecting training samples relative to the cost of performing the sensitivity evaluations by varying λ_1 is case-specific and mainly determined by the target configuration. For example, in our previous study, the computational cost for generating training samples accounts for 12.5% of the target 2D DNS run (Jung et al., 2024), and is expensive compared with the ANN training cost which accounts for $\sim 0.1\%$ of the 2D DNS run. Therefore, the proposed TL approach is particularly useful when the computational cost for obtaining high-fidelity training samples is expensive.

Next, a posteriori evaluation of the PC-transport ROM with different transfer learning strategies is carried out by performing a series of 0D simulations of a hydrogen/air mixture using the PC-transport ROM for the target task, where $T_0 = 1050$ K and $N_\phi = 4$. Figure 9 shows the variations in τ_{ig} as a function of ϕ depending on the different transfer learning methods. The PC-transport ROM fails to predict the overall ignition trend when the ANN is trained from scratch or trained with TL1. Note that the PC-transport ROM with TL1 inaccurately captures the early stage of 0D ignition, leading to error accumulation over time. Consequently, the PC-transport ROM with TL1 fails to undergo a thermal runaway crossing for the full range of ϕ and, hence, τ_{ig} approaches infinity.

The overall performance of the PC-transport ROM with TL2 is better than the PC-transport ROM without applying the transfer learning method or with TL1, demonstrating the importance of the initialization scheme and the fine-tuning step on the result, respectively. Here, τ_{ig} predicted by the PC-transport ROM with TL2 is in relatively good agreement with that from the FOM when ϕ is near

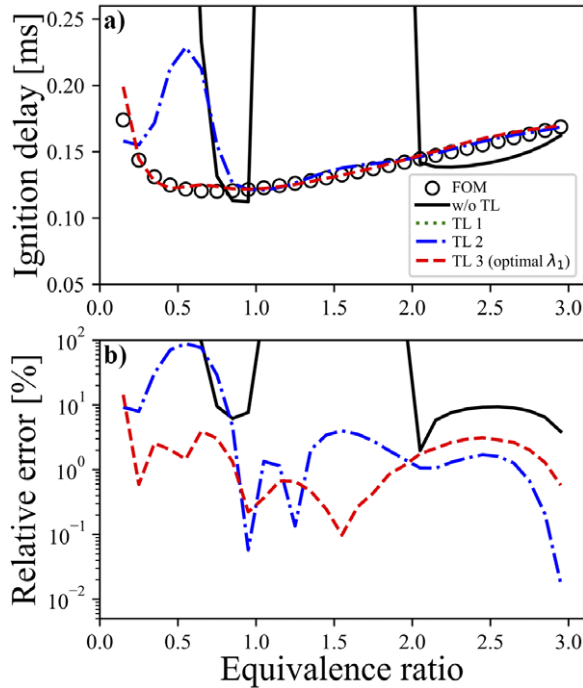


Figure 9. Variations in (a) 0D ignition delay time, τ_{ig} , predicted by FOM (solid symbol) and PC-transport ROMs trained by applying different transfer learning methods, and (b) the relative-error of the PC-transport ROMs compared with FOM for the homogeneous hydrogen/air mixture with various ϕ at $T_0 = 1050$ K. N_ϕ of the training set is set to 4. The values of the optimal λ_1 for TL3 are 5×10^{-2} , 1×10^{-3} , and 1×10^{-3} for Cl#1, Cl#2, and Cl#3, respectively. The ROM with TL1 fails to predict ignition, and therefore, the result with TL1 is not shown in the figure.

those in the target task training dataset (i.e., ϕ of 0.2, 1.0, 2.0, and 3.0), while the performance of the ANN model decreases as ϕ becomes farther from those in the training dataset of the target task.

For the PC-transport ROM with the optimal value of λ_1 (i.e., TL3), it is readily observed that the PC-transport ROM shows a good performance of predicting τ_{ig} over a wide range of ϕ . This result substantiates that the regularization-based transfer learning framework can increase the accuracy of the ANN model with the sparse training dataset. Figure 10 presents the temporal evolution of the PCs with four different values of ϕ of 0.15, 0.65, 1.55, and 2.55, predicted by the FOM and the PC-transport ROM with an optimal value of λ_1 . Although a slight time lag exists between the FOM and the PC-transport ROM results as ϕ deviates farther from values in the training dataset, the PC-transport ROM can reasonably capture the onset of ignition and the subsequent equilibrium period of the PCs.

4.2. TL1–TL3: parametric study in terms of task similarity and data sparsity

Additional parametric studies are carried out by varying N_ϕ or increasing T_0 to 1300 and 1400 K. Figure 11 shows τ_{ig} as a function of ϕ for the target task with T_0 of 1050 K and decreasing N_ϕ to 2 and 3 by applying different transfer learning methods. Note that as N_ϕ decreases to 2 and 3, the training dataset is intended not to cover the entire range of the test dataset (i.e., $\phi = 0.5$ and 1.5 for $N_\phi = 2$, and $\phi = 0.5, 1.5$, and 2.5 for $N_\phi = 3$) such that there exist several test cases where ϕ is outside of the range of the training dataset (see the highlighted regions in Figure 11). The overall variations in τ_{ig} predicted by applying different transfer learning methods show a similar trend regardless of the change of N_ϕ . The PC-transport ROM with TL1 fails to capture the onset of ignition of the hydrogen/air mixture for the entire range of ϕ , while

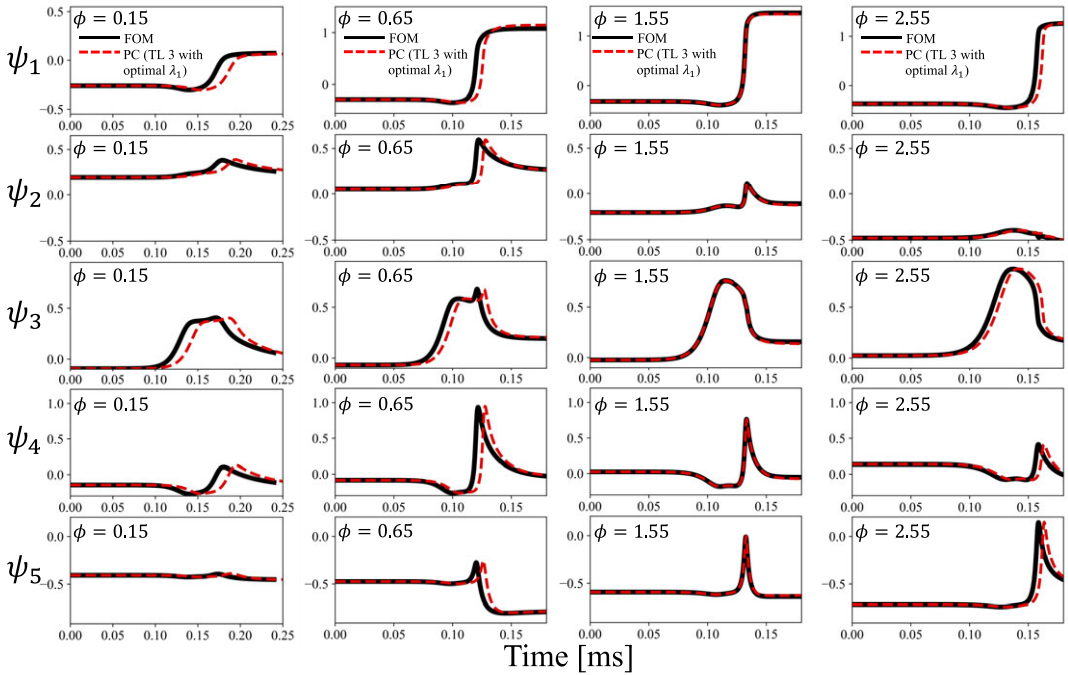


Figure 10. Temporal evolution of the PCs that represent the homogeneous hydrogen/air mixture at $T_0 = 1000$ K and ϕ of 0.15, 0.65, 1.55, and 2.55, respectively (left to right). Solid line: PCs projected from the FOM result, Dashed line: PC-transport ROM using the optimal λ_1 in TL3. The values of the optimal λ_1 for TL3 are 5×10^{-2} , 1×10^{-3} , and 1×10^{-3} for Cl#1, Cl#2, and Cl#3, respectively.

the results with TL2 show a better performance than those without applying transfer learning. The PC-transport ROM with the optimal value of λ_1 in TL3 outperforms all the other models.

As expected, the accuracy of the PC-transport ROM notably decreases as the target ϕ of the 0D simulation is outside of the range of the training dataset, which is a well-known drawback of machine learning models for extrapolation. Nonetheless, the result of the PC-transport ROM with TL3 shows relatively good performance even for the cases where ϕ is outside of the range of the training set. This result implies that the previous knowledge obtained from the source task helps increase the accuracy of the extrapolation of the ANN model, consistent with previous findings (Humbird et al., 2020).

Next, target tasks are considered where T_0 is increased further (i.e., $T_0 = 1300$, and 1400 K) such that the similarity between source and target tasks decreases. Figure 12 shows the variations in τ_{ig} for the hydrogen/air mixture with T_0 of 1300 and 1400 K and N_ϕ of 4 by using various ANN models with or without applying transfer learning methods. For the cases with $T_0 = 1300$ K, it is found that the PC-transport ROM with the optimal value of λ_1 in TL3 shows a reasonable performance over the entire range of ϕ , while the results without applying transfer learning or with applying TL2 exhibit a noticeable error in predicting ignition of a lean mixture. As T_0 in the target task further increases to 1400 K, on the other hand, the result with TL3 shows only a marginal improvement compared with the other cases. This result demonstrates that the performance of the regularization-based transfer learning method decreases with a decrease in task similarity between source and target tasks.

4.3. PaPIR: unified transfer learning method

Finally, the performance of the unified transfer learning method, PaPIR, in the different target tasks is investigated. As discussed in Section 2.4, the central idea of PaPIR is to control the degree of knowledge transfer from the source to the target task by adjusting the magnitudes of λ_1 and λ_2 , which are associated

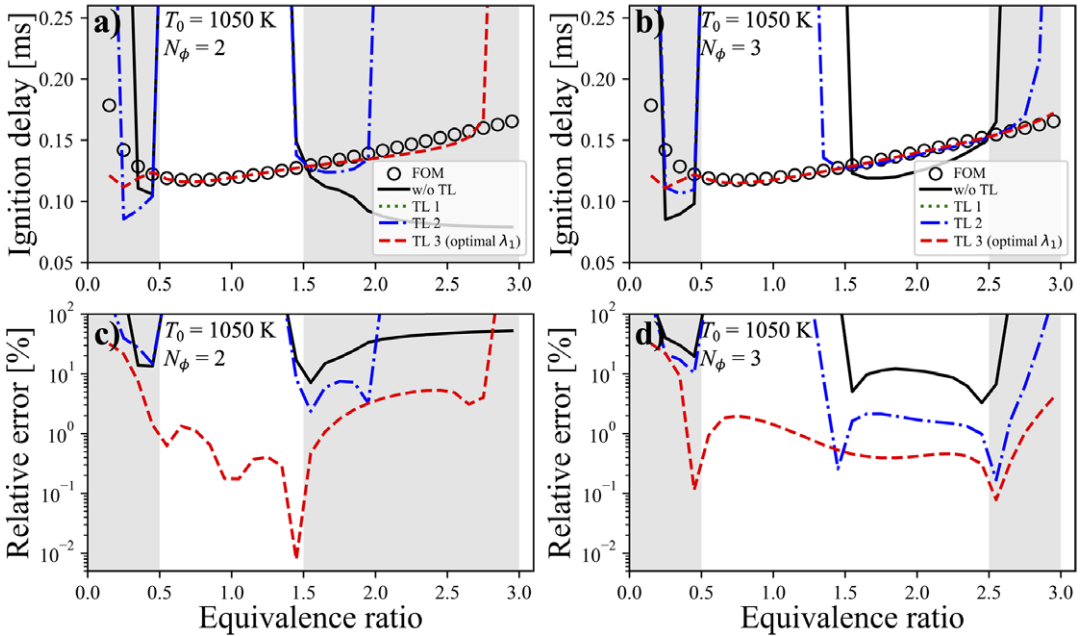


Figure 11. Variations in (top) τ_{ig} predicted by FOM (solid symbol) and PC-transport ROMs trained by applying different transfer learning methods, and (bottom) the relative error of the PC-transport ROMs compared with FOM for the homogeneous hydrogen/air mixture with various ϕ at $T_0 = 1050$ K. Here, N_ϕ of the training set is (left) 2, and (right) 3, respectively. The highlighted region represents the cases where ϕ is out of the range of the training dataset in the target task. The ROM with TL1 fails to predict ignition, and hence its results are not shown in the figure.

with the regularization and initialization of the ANN model in the target task, respectively. Unlike TL3, \mathbf{h}_0 in PaPIR can be distributed by either a normal distribution function following the Xavier normal initialization method ($\lambda_2 = 0$), or \mathbf{h}^s ($\lambda_2 = 1.0$), or in between the two ($0 < \lambda_2 < 1$). In this regard, the effect of the initialization on the performance of transfer learning can be investigated by varying λ_2 in PaPIR.

Figure 13 presents the a priori result of the best achievable NRMSE of $\dot{\omega}_{\psi}$ of the test set for three different target tasks (i.e., $T_0 = 1050, 1300,$ and 1400 K with N_ϕ of 4) as a function of λ_1 and λ_2 , conditional on each of the three different clusters. Consistent with the previous results, the best achievable (or minimum) value of NRSME is evaluated by repeating the ANN model training 10 times at a given λ_1 and λ_2 . As shown in Figure 13a, PaPIR covers all the transfer learning methods discussed in the present study, namely, TL1, TL2, and TL3. In general, the results with TL1 exhibit a large error and increase with an increase of T_0 in the target task. In TL3, there exists an optimal value of λ_1 that results in a lower NRMSE than for the results either without applying transfer learning or with TL2.

For the case where task similarity is relatively high (i.e., $T_0 = 1050$ K in the target task; see Figure 13a–c), the value of NRMSE is mainly governed by the regularization parameter λ_1 , whereas it is largely unaffected by a change of the initialization parameter, λ_2 . Since the source and target tasks are similar to each other in this case, the optimal value of λ_1 is relatively large (e.g., $\lambda_1 = 10^{-1}$ in Cl#1). Given that a regularization term serves to convexify the objective function, a relatively large magnitude of λ_1 leads the ANN model to be insensitive to a change in the initialization scheme. Consequently, PaPIR does not outperform TL3 when the similarity between source and target tasks is high. The best achievable values of the NRMSE depending on the different transfer learning methods are summarized in Table 3.

As T_0 in the target task increases to 1300 K, results with Cl#1 (Figure 13d) show that the NRMSE of the test dataset attains its minimum at a relatively low magnitude of λ_1 ($= 10^{-3}$). Although the overall variations of the NRMSE are still mainly governed by λ_1 , the NRMSE is no longer invariant to a change of

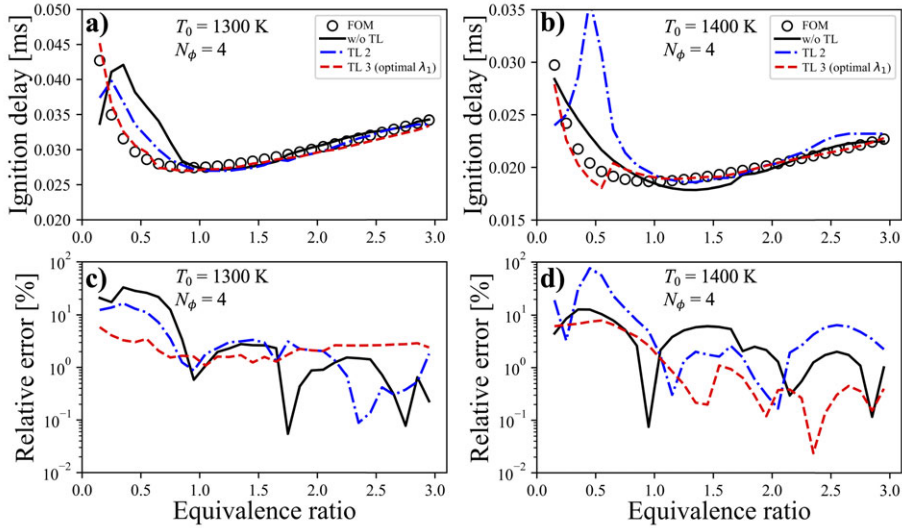


Figure 12. Variations in (top) τ_{ig} predicted by (solid symbol) FOM and PC-transport ROMs trained by applying different transfer learning methods, and (bottom) the relative error of the PC-transport ROMs compared with FOM for the homogeneous hydrogen/air mixture with various ϕ at (left) $T_0 = 1300$ K and (right) $T_0 = 1400$ K.

λ_2 at the optimal value of λ_1 , indicating that λ_2 starts to play a role in the optimization of the ANN model. Since the magnitude of the optimal value of λ_1 decreases compared with the case with $T_0 = 1050$ K, the complexity of the loss function at the optimal value of λ_1 increases, and consequently, the training result can be varied with the different initialization schemes. This finding indicates that an initialization scheme becomes important in the framework of transfer learning as the similarity between the source and target tasks becomes relatively low. Note that in Figure 13d, the ANN model exhibits a slightly lower magnitude of NRMSE at λ_2 of 0.7 compared with that of 1.0, illustrating the potential advantage of PaPIR over TL3. Readers are referred to Table 3 to quantify the difference of NRMSE between PaPIR and TL3. For the results of Cl#2 and Cl#3, on the other hand, λ_1 still plays a major role in determining the best achievable value of NRMSE (see Figure 13e and f). This would be because the decrease of task similarity for these clusters is not as pronounced as for Cl#1.

As T_0 in the target task further increases to 1400 K, it is found that λ_1 still shows a dominant effect on the NRSME compared with λ_2 , demonstrating that the primary factor of determining the performance of transfer learning is a regularization parameter in general. Nonetheless, there are several cases where the ANN model trained with $\lambda_2 < 1$ exhibits a lower magnitude of NRMSE compared with the best candidate obtained from TL3 (see Figure 13g). Note that at $T_0 = 1400$ K, the ratio of the best achievable NRMSE obtained from PaPIR to TL3 is 0.827, 0.984, and 0.866, for Cl#1, Cl#2, and Cl#3, respectively. This result shows that adjusting the initial values of the parameters in the target task can further enhance the performance of transfer learning in the target task with a sparse dataset, especially when the similarity between source and target tasks is low such that the optimal value of the regularization parameter, λ_1 , becomes relatively low.

To further investigate the advantage of PaPIR over other transfer learning methods, especially when the task similarity is relatively low, Figure 14 presents the variations in τ_{ig} for a hydrogen/air mixture with T_0 of 1400 K and N_ϕ of 4, predicted by the FOM and the PC-transport ROMs with different transfer learning methods. This figure clearly shows that τ_{ig} predicted by PaPIR shows excellent agreement with that from the FOM over the entire range of ϕ , which is clearly distinct from the other models. Although the relative error obtained from PaPIR is slightly higher than that from TL3 or the PC-transport ROM without

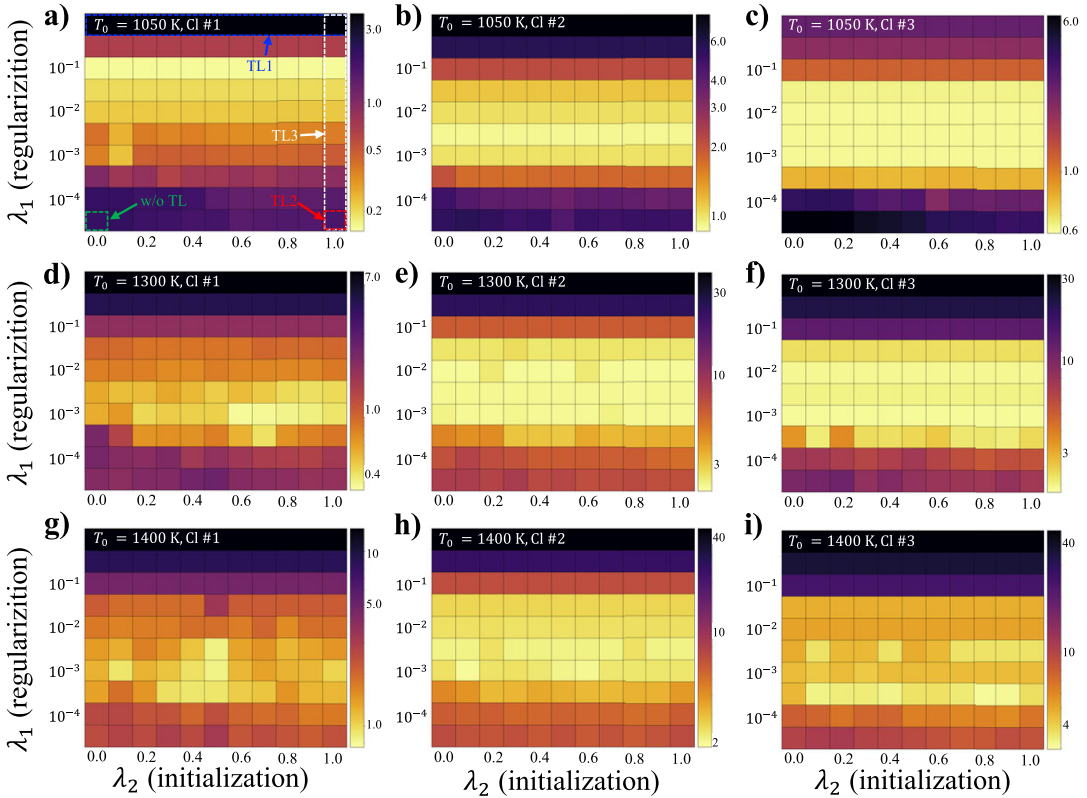


Figure 13. Distributions of the best achievable value of NRMSE [%] using PaPIR as a function of λ_1 and λ_2 for the different test datasets out of 10 repetitions of the ANN model training. The target task is varied ranging from $T_0 = 1050, 1300,$ and 1400 K (top to bottom) for Cluster 1, 2, and 3 (left to right) with N_ϕ of 4.

Table 3. Best achievable value of NRMSE [%] by using different transfer learning methods for the test dataset with various T_0 and $N_\phi = 4$, out of 10 repetitions of ANN model training

Model	$T_0 = 1050$ K			$T_0 = 1300$ K			$T_0 = 1400$ K		
	Cl#1	Cl#2	Cl#3	Cl#1	Cl#2	Cl#3	Cl#1	Cl#2	Cl#3
w/o TL	1.9818	5.4682	6.0114	1.7982	8.6369	10.961	2.2756	6.8892	11.000
TL1	3.7541	7.7887	2.7417	7.3993	38.760	31.030	13.134	44.246	46.182
TL2	1.7861	5.3703	5.3862	1.9408	6.9736	8.6645	2.1617	6.9680	11.038
TL3	0.1570	0.8949	0.5108	0.3893	2.2277	1.8973	1.0157	2.1242	3.6054
PaPIR	0.1570	0.8967	0.5105	0.3316	2.2227	1.8963	0.8401	2.0896	3.1229

applying transfer learning at $\phi > 1$, the PC-transport ROM with PaPIR shows a more robust performance for predicting the oxidation process of hydrogen/air mixture over a wide range of ϕ .

One may argue that under the data-sparse scenario, the number of test datasets is also likely to be insufficient, rendering it infeasible to find the optimal values of λ_1 and λ_2 by relying on the test dataset. In future work, a systematic way of estimating the optimal values of those two parameters without relying on a test dataset will be investigated. One practical example would be adopting the L-curve criterion, a well-known heuristic method, to find the optimal regularization parameter without relying on the test dataset

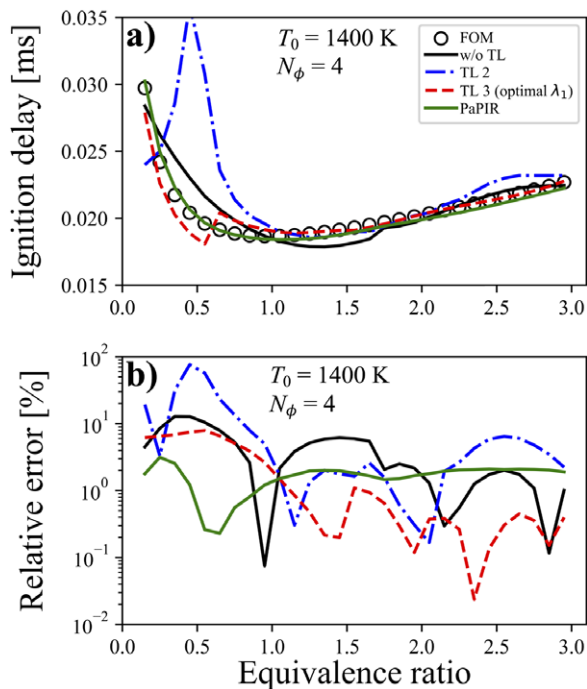


Figure 14. Variations in (top) τ_{ig} predicted by the (open symbol) FOM and PC-transport ROMs trained by applying different transfer learning methods, and (b) the relative error of the PC-transport ROMs compared with the FOM for a homogeneous hydrogen/air mixture with various ϕ at $T_0 = 1400$ K and $N_\phi = 4$.

(Hansen, 1992). Also as future work, we will compare the performance of different methods that “partially” transfer the knowledge of the pretrained neural network model in the target task. This includes applying Bayesian transfer learning methods (Soriano et al., 2024) or Hessian-based transfer learning methods (Jung et al., 2024), through which the efficiency of the TL methods will be elucidated in various scenarios.

Note also that the present study effectively assesses the efficiency of TL methods in chemical kinetics, as temperature is a key factor that significantly influences these processes. Regularization-based transfer learning has been successfully demonstrated as a robust algorithm not only in the machine learning community in general but also in combustion systems (e.g., across different equivalence ratios (Soriano et al., 2024) and numerical configurations (Jung et al., 2024)). Therefore, changes in operating conditions, such as equivalence ratio and pressure, are likely to provide similar insights as those found in the present study. The overall efficiency of TL will be further investigated with large hydrocarbon fuels to elucidate the effect of dimensionality and the complexity of combustion systems on transfer learning.

5. Conclusions

In this study, various transfer learning methods were applied to the prediction of the reaction rate of the PCA-based low-dimensional manifold that represents the ignition process of a homogeneous hydrogen/air mixture in a constant volume reactor. A sufficient number of training samples spanning a wide range of ϕ was provided in the source task where $T_0 = 1000$ K, whereas the number of training datasets was assumed to be sparse in the target task where $T_0 > 1000$ K. The effect of the number of training samples on the performance of the PC-transport ROM was first investigated, followed by the application of three different transfer learning approaches (i.e., TL1, TL2, and TL3) to the different target tasks. To this end, a

unified transfer learning framework was proposed in this study to elucidate the role of initialization and regularization on the performance of transfer learning. The following results are highlighted from the present study:

- In general, the number of training datasets played a primary role in determining the performance of the model. Without applying transfer learning, the PC-transport ROM failed to reproduce the ignition process of a hydrogen/air mixture with a sparse dataset (i.e., $N_\phi \leq 15$). It was also found that the PC-transport ROM without transfer learning shows relatively good accuracy for the test cases when the initial condition of the ROM is adjacent to that included in the training dataset.
- Three different transfer learning methods, parameter sharing (TL1), fine-tuning (TL2), and parameter restriction (TL3), were then applied to the target task where $T_0 = 1050$ K and $N_\phi = 4$. The PC-transport ROM using TL1 led to a significant error in predicting the PCs' reaction rates, while the PC-transport ROM with TL2 showed a slightly better performance than that without applying transfer learning approaches. An optimal value of the regularization parameter λ_1 in TL3 led to a remarkable decrease in the NRMSE of the test dataset. Moreover, the profiles of the 0D ignition delay predicted by the PC-transport ROM with TL3 exhibit good agreement with those obtained from the FOM, demonstrating the importance of the regularization-based transfer learning method.
- Parametric studies were performed by varying T_0 and N_ϕ in the target tasks to investigate the effect of task similarity and data sparsity in the target task on the performance of the different transfer learning methods, respectively. It was found that the knowledge from the source task helped predict the ignition process of a hydrogen/air mixture outside of the ϕ range in the training dataset, demonstrating the advantage of applying transfer learning for extrapolation. As T_0 in the target task was increased to 1400 K, the performance of TL3 is degraded because of the decrease of the similarity between the source and target tasks.
- A novel transfer learning approach, PaPIR, was applied to the various target tasks. When the task similarity between the source and target tasks is high, the effect of the initialization parameter, λ_2 , has a negligible effect on the NRMSE of the test set of the target task, while the minimum of the NRMSE is primarily determined by λ_1 . The optimal value of λ_1 decreased with a decrease in task similarity, such that the effect of different initialization schemes on the result became noticeable. Although λ_1 still had a dominant effect on the result, an additional performance improvement could be achieved by changing the magnitude of λ_2 , illustrating the potential advantage of PaPIR.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/dce.2024.50>.

Data availability statement. The data that support the findings of this study are available from the authors upon request.

Acknowledgements. The authors would like to acknowledge the contributions of Anuj Kumar at North Carolina State University for sharing the code for the PC-transport ROM.

Author contributions. Conceptualization: K.S.J., T.E., J.H.C., M.K.; Data Curation: K.S.J.; Investigation: K.S.J., Methodology: T.E., J.H.C., M.K.; Supervision: M.K.; Writing—Original Draft: K.S.J.; Writing—Review and Editing: T.E., J.H.C., M.K.; Funding acquisition: M.K.

Funding statement. This work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories (Project 222361), a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This report describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the report do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Competing interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Ethical standards. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

References

- Abdelwahid S, Malik MR, Al Kader Hammoud HA, Hernández-Pérez FE, Ghanem B and Im, HG (2023) Large eddy simulations of ammonia-hydrogen jet flames at elevated pressure using principal component analysis and deep neural networks. *Combustion and Flame*, 253, 112781.
- Biglari A and Sutherland JC (2012) A filter-independent model identification technique for turbulent combustion modeling. *Combustion and Flame* 159(5), 1960–1970.
- Biglari A and Sutherland JC (2015) An *a posteriori* evaluation of principal component analysis-based models for turbulent combustion simulations. *Combustion and Flame* 162, 4025–4035.
- Dalakoti DK, wehrfritz A, Savard B, Day MS, Bell JB and Hawkes ER (2021) An *a priori* evaluation of a principal component and artificial neural network based combustion model in diesel engine conditions. *Proceedings of the Combustion Institute* 38, 2701–2709.
- D'Alessio G, Sundaresan S and Mueller ME (2023) Automated and efficient local adaptive regression for principal component-based reduced-order modeling of turbulent reacting flows. *Proceedings of the Combustion Institute* 39(4), 5249–5258.
- De S and Doostan A (2022) Neural network training using l1-regularization and bi-fidelity data. *Journal of Computational Physics* 458, 111010.
- Echekki T and Mirgolbabaei H (2015) Principal component transport in turbulent combustion: *A posteriori* analysis. *Combustion and Flame* 162, 1919–1933.
- Glorot X and Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. In *Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 249–256.
- Han X, Jia M, Chang Y and Li Y (2020) A deep learning-based ODE solver for chemical kinetics [arXiv:2012.12654].
- Han X, Jia M, Chang Y and Li Y (2022) An improved approach towards more robust deep learning models for chemical kinetics. *Combustion and Flame* 238, 111934.
- Hansen PC (1992) Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review* 34, 561–580.
- Hosny KM, Kassem MA and Fouad MM (2018) Skin cancer classification using deep learning and transfer learning. In *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)*, pp. 90–93.
- Humbird KD, Peterson JL, Spears BK and McClarren RG (2020) Transfer learning to model inertial confinement fusion experiments. *IEEE Transactions on Plasma Science* 48(1), 61–70.
- Ilme M, Cha CM and Pitsch H (2005) Prediction of local extinction and re-ignition effects in nonpremixed turbulent combustion using a flamelet/progress variable approach. *Proceedings of the Combustion Institute* 30, 793–800.
- Isaac BJ, Thornock JN, Sutherland J, Smith PJ and Parente A (2015) Advanced regression methods for combustion modelling using principal components. *Combustion and Flame* 162, 2592–2601.
- Jung KS, Soriano BS, Chen JH and Khalil M (2024) A hessian-based transfer learning approach for artificial neural networks based chemical kinetics with a sparse dataset. *Proceedings of the Combustion Institute* 40, 105390.
- Jung KS, Kumar A, Echekki T and Chen JH (2024) On the application of principal component transport for compression ignition of lean fuel/air mixtures under engine relevant conditions. *Combustion and Flame* 260, 113204. <https://doi.org/10.1016/j.combustflame.2023.113204>.
- Kee RJ, Rupley FM, Meeks E and Miller JA (1996) CHEMKIN-III: A Fortran Chemical Kinetics Package for the Analysis of Gas-Phase Chemical and Plasma Kinetics. Tech. rep. No. SAND96-8216. Sandia National Labs., Livermore, CA, USA.
- Kennedy C and Carpenter M (1994) Several new numerical methods for compressible shear-layer simulations. *Applied Numerical Mathematics* 14, 397–433.
- Kingma DP and Ba J (2014) Adam: a method for stochastic optimization [arXiv:1412.6980].
- Kumar A, Rieth M, Owoyele O, Chen J and Echekki T (2023) Acceleration of turbulent combustion DNS via principal component transport. *Combustion and Flame* 255, 112903.
- Lam SH and Goussis DA (1994) The CSP method for simplifying chemical kinetics. *International Journal of Chemical Kinetics* 26, 461–486.
- Li J, Zhao Z, Kazakov A and Dryer FL (2004) An updated comprehensive kinetic model of hydrogen combustion. *International Journal of Chemical Kinetics* 36, 566–575.
- Li X, Grandvalet V and Davoine F (2018) Explicit inductive bias for transfer learning with convolutional networks. In *Proceedings of the 35th International Conference on Machine Learning*, vol. 80, pp. 2825–2834.
- Li X, Grandvalet Y and Davoine F (2029) A baseline regularization scheme for transfer learning with convolutional neural networks. *Pattern Recognition* 98, 107049.
- Liu X, Li Y and Chen G (2022) Transfer learning for regression via latent variable represented conditional distribution alignment. *Knowledge Based Systems* 240, 108110.
- Lu T and Law CK (2005) A directed relation graph method for mechanism reduction. *Proceedings of the Combustion Institute* 30, 1333–1341.
- Lu T and Law CK (2009) Toward accommodating realistic fuel chemistry in large-scale computations. *Progress in Energy and Combustion Science* 35, 192–215.
- Malik MR, Coussement A, Echekki T and Parente A (2022) Principal component analysis based combustion model in the context of a lifted methane/air flame: Sensitivity to the manifold parameters and subgrid closure. *Combustion and Flame* 244, 112134.

- Malik MR, Vega PO, Coussement A and Parente A** (2021) Combustion modeling using principal component analysis: *A posteriori* validation on Sandia flames D, E and F. *Proceedings of the Combustion Institute* 38, 2635–2643.
- Mieth B, Hockley JRF, Görnitz N, Vidovic MM-C, Müller K-R, Gutteridge A and Ziemek D** (2019) Using transfer learning from prior reference knowledge to improve the clustering of single-cell RNA-Seq data. *Scientific Reports* 9, 20353.
- Mirgolbabaei H and Echekki T** (2013) A novel principal component analysis-based acceleration scheme for LES–ODT: An *a priori* study. *Combustion and Flame* 160, 898–908.
- Mirgolbabaei H and Echekki T** (2014) Nonlinear reduction of combustion composition space with kernel principal component analysis. *Combustion and Flame* 161, 118–126.
- Mirgolbabaei H and Echekki T** (2015) The reconstruction of thermo-chemical scalars in combustion from a reduced set of their principal components. *Combustion and Flame* 162, 1650–1652.
- Narkhede MV, Bartakke PP and Sutaone MS** (2022) A review on weight initialization strategies for neural networks. *Artificial Intelligence Review* 55, 291–322.
- Owoyele O and Echekki T** (2017) Toward computationally efficient combustion DNS with complex fuels via principal component transport. *Combustion Theory and Modelling* 21, 770–798.
- Pan SJ and Yang Q** (2009) A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 20, 1345–1359.
- Parente A, Sutherland J, Tognotti L and Smith P** (2009) Identification of low-dimensional manifolds in turbulent flames. *Proceedings of the Combustion Institute* 32, 1579–1586.
- Parente A and Sutherland JC** (2013) Principal component analysis of turbulent combustion data: data pre-processing and manifold sensitivity. *Combustion and Flame* 160, 340–350.
- Pepiot-Desjardins P and Pitsch H** (2008) An efficient error-propagation-based reduction method for large chemical kinetic mechanisms. *Combustion and Flame* 154, 67–81.
- Peters N** (1983) Local quenching due to flame stretch and non-premixed turbulent combustion. *Combustion Science and Technology* 30, 1–17.
- Peters N** (1984) Laminar diffusion flamelet models in non-premixed turbulent combustion. *Progress in Energy and Combustion Science* 10, 319–339.
- Pierce CD and Moin P** (2004) Progress-variable approach for large-eddy simulation of non-premixed turbulent combustion. *Journal of Fluid Mechanics* 504, 73–97.
- Quattoni A, Collins M and Darrell T** (2008) Transfer learning for image classification with sparse prototype representations. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8.
- Salaken SM, Khosravi A, Nguyen T and Nahavandi S** (2019) Seeded transfer learning for regression problems with deep learning. *Expert Systems with Applications* 115, 565–577.
- Savarese M, Jung KS, Dave H, Chen JH and Parente A** (2024) A new index for the comparative evaluation of combustion local low-dimensional manifolds. *Combustion and Flame* 265, 113434.
- Soriano BS, Jung KS, Echekki T, Chen JH and Khalil M** (2024) Probabilistic transfer learning methodology to expedite high fidelity simulation of reactive flows. <https://arxiv.org/abs/2405.10944>
- Subel A, Chattopadhyay A, Guan Y and Hassanzadeh P** (2021) Data-driven subgrid-scale modeling of forced burgers turbulence using deep learning with generalization to higher Reynolds numbers via transfer learning. *Physics of Fluids* 33, 031702.
- Sun W, Chen Z, Gou X and Ju Y** (2010) A path flux analysis method for the reduction of detailed chemical kinetic mechanisms. *Combustion and Flame* 157, 1298–1307.
- Sutherland JC and Parente A** (2009) Combustion modeling using principal component analysis. *Proceedings of the Combustion Institute* 32, 1563–1570.
- van Oijen JA and de Goey LPH** (2000) Modelling of premixed laminar flames using flamelet-generated manifolds. *Combustion Science and Technology* 161, 111–138.
- Wang R, Zhou J, Jiang H, Han S, Wang L, Wang D and Chen Y** (2021) A general transfer learning-based gaussian mixture model for clustering. *International Journal of Fuzzy Systems* 23, 776–793.
- Yang K, Lu J, Wan W, Zhang G and Hou L** (2022) Transfer learning based on sparse gaussian process for regression. *Information Sciences* 605, 286–300.
- Yang Q, Chen Y, Xue G-R, Dai W and Yu Y** (2009) Heterogeneous transfer learning for image clustering via the social web. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pp. 1–9.
- Yang Q, Zhang Y, Dai W and Pan SJ** (2020) *Transfer Learning*. Cambridge: Cambridge University Press https://scholar.google.com/citations?view_op=view_citation&hl=en&user=P6WcnfkAAAAJ&citation_for_view=P6WcnfkAAAAJ-KUbv n5osdkgC.
- Yao, Y. and Doretto, G.** (2010) Boosting for transfer learning with multiple sources. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1855–1862.
- Zhu Y, Chen Y, Lu Z, Pan S, Xue G-R, Yu Y and Yang, Q** (2011) Heterogeneous transfer learning for image classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 25, 1304–1309.

Cite this article: Jung KS, Echekki T, Chen JH and Khalil M (2024). Transfer learning for predicting source terms of principal component transport in chemically reactive flow. *Data-Centric Engineering*, 5, e42. doi:10.1017/dce.2024.50