Adaptive Depth Neural Networks for Scale-Bridging Modeling of Multiphase Reacting Flows

Michael E. Mueller[1], Sankaran Sundaresan[2]

[1]Department of Mechanical and Aerospace Engineering

[2]Department of Chemical and Biological Engineering

Princeton University

Contributors:

G. D'Alessio (PD), C.E. Lacey (GS), M. Sulaiman (PD)

Shift Ctrfl

FY22 FECM Spring R&D Project Review Meeting

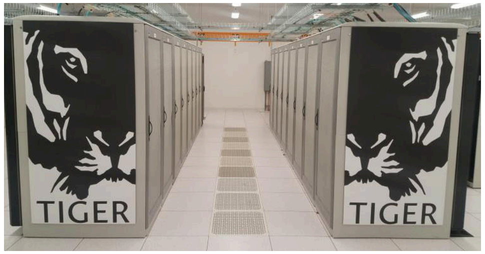
Crosscutting Research: Simulation-Based Engineering

May 6, 2022

Motivation

- Big Computing Means Big Data…

$O(\mathrm{GB})$



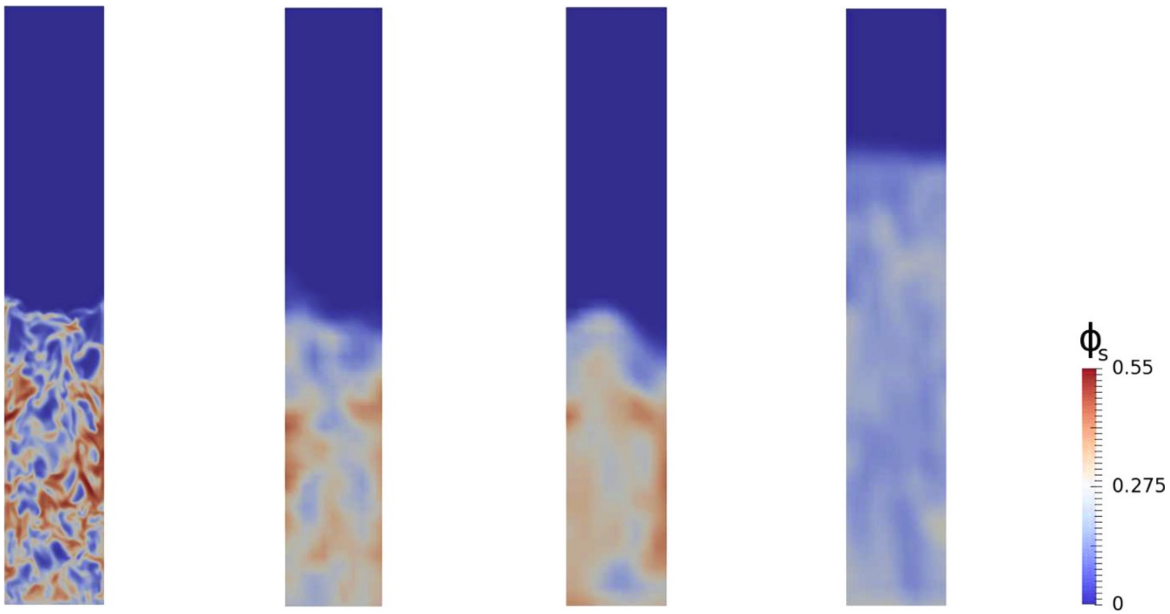
$O(\mathrm{TB})$

$> O(\mathrm{PB})$

- <u>Question</u>: Can we and how do we leverage this big data in modeling for coarse-grained multi-physics flow simulations?

Motivation

- Big Computing Means Big Data…

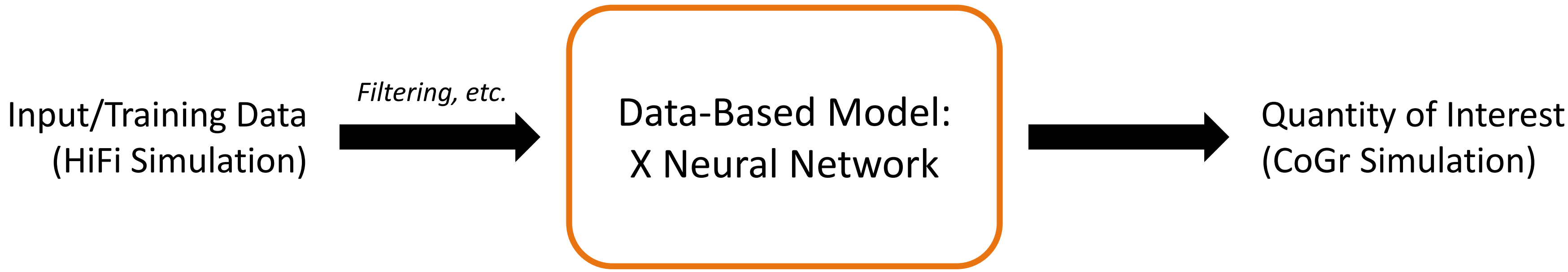


- Question: How do we use data from targeted high-resolution, full-fidelity simulations to develop efficient data-based models for coarse-grained simulations?

Image: Y. Jiang, J. Kolehmainen, Y. Gu, Y.g. Kevrekidis, A. Ozel, S. Sundaresan, Powd. Tech. 346 (2019) 403-413

Motivation

- Data-Based Modeling Framework: Part I (A Priori)



- The training of data-based models typically requires tremendous expertise to tune architecture/hyperparameters and massage the input data.

- <u>Program Objective I</u>: Develop a data-based modeling framework that is fully automated and efficient.

Motivation

- Data-Based Modeling Framework: Part II (A Posteriori)

Input
(CoGr Simulation) → Data-Based Model:
X Neural Network → Quantity of Interest
(CoGr Simulation)

- The use of data-based models in coarse-grained simulations requires that they be as efficient as possible lest any increase in accuracy be offset by an increase in computational cost.

- Program Objective II: Develop data-based models that are "simple" and quick to evaluate in coarse-grained simulations.

Data-Based Modeling

- Data-Based Modeling Workflow

Raw Input Data →

Data Pre-Processing: Outlier Removal, Data Sampling

→ Dimensionality Reduction: Principal Component Analysis, Dimensional Consistency

→ Modeling: Regression (NN), Hyperparameters/Architecture

→ **Output QoI**

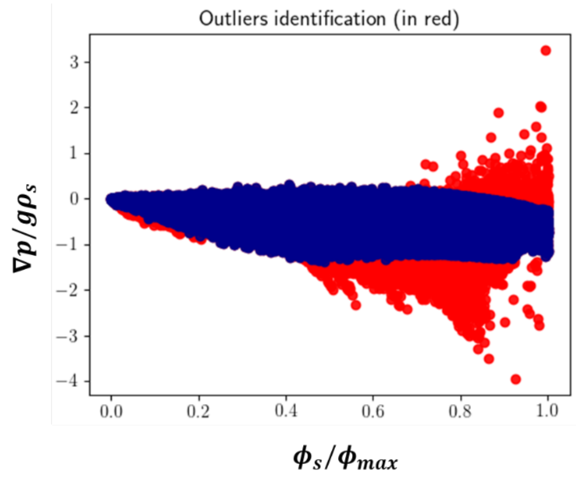
- Philosophy: Automated, Efficient, Simple

Data-Based Modeling

Data Pre-Processing: Outlier Removal

- Based on Principal Component Analysis (PCA)* and low-dimensional projection with large projection errors[1]:
  - Orthogonal outliers: Far from the low-dimensional projection but close to the bulk of the data when projected
  - Leverage outliers: Far from the low-dimensional project and far from the bulk of the data

- Application: Fluidized Bed Reactor[2]
  - Training data from fine-grid TFM.
  - Model target is the filtered particle drag.
  - Inputs: Reynolds number, filter size, solid volume fraction, pressure gradient, and slip velocity.



*Principal Component Analysis (PCA) finds the linear combination of input variables that contributes most to input variance.

[1]M. Hubert, P. Rousseeuw, T. Verdonck, Comput. Stat. Data Analy. 53 (2009) 2264-2274

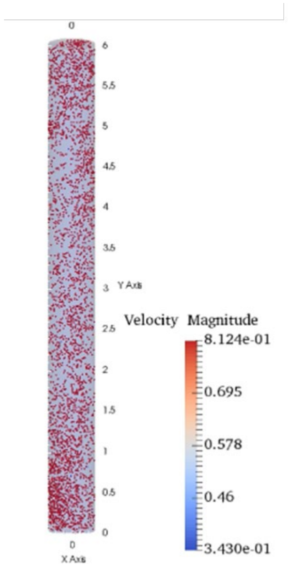
[2]Y. Jiang, J. Kolehmainen, Y. Gu, Y.g. Kevrekidis, A. Ozel, S. Sundaresan, Powd. Tech. 346 (2019) 403-413

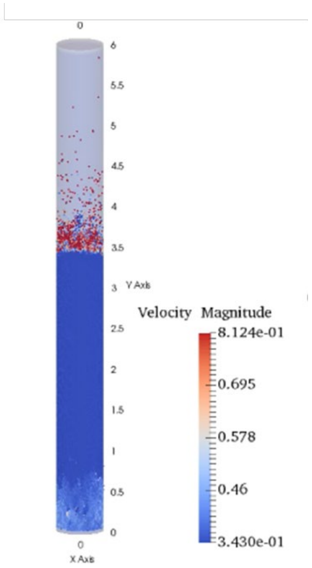
Shift
Ctrfl

Data-Based Modeling

- ## Data Pre-Processing: Outlier Removal
  - Application: Fluidized Bed Reactor

Without Outlier Removal: Unstable Simulation





With Outlier Removal: Stable Simulation

- Lesson Learned: Outlier removal can help prevent overfitting and poor model extrapolation when applied to forward simulation.

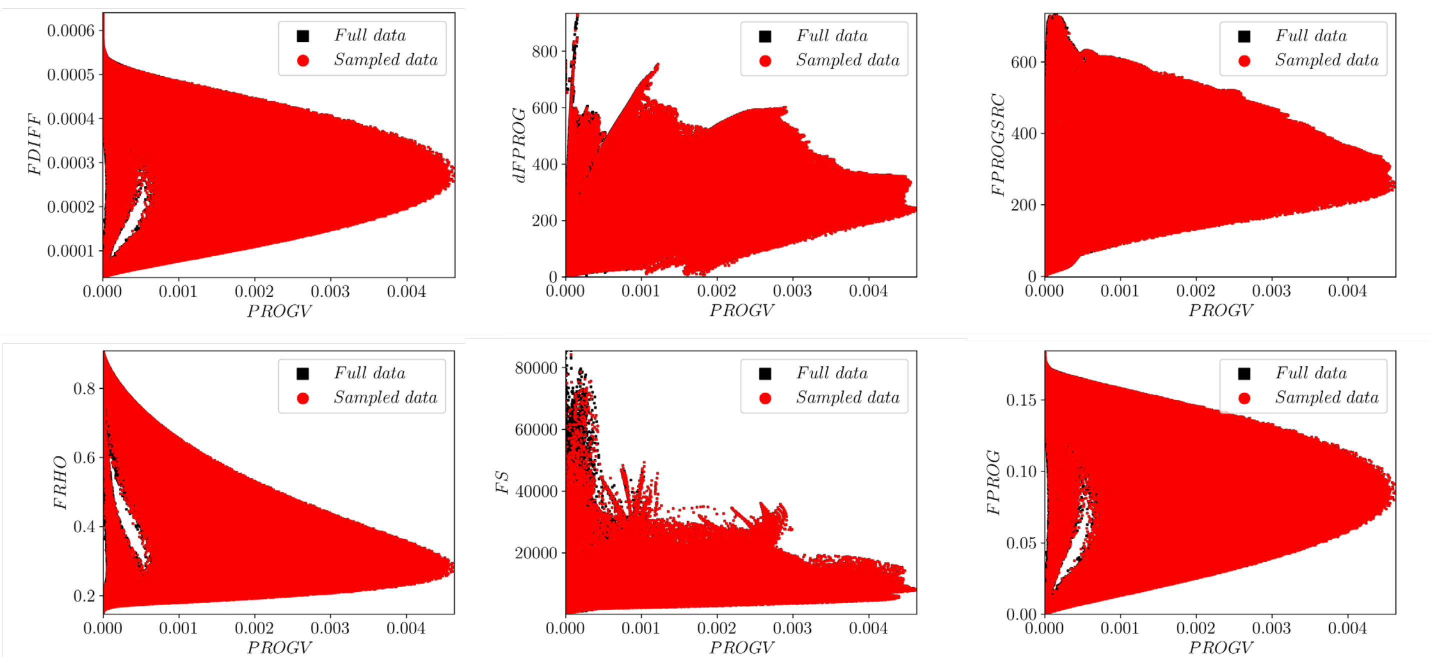
Data-Based Modeling

- Data-PreProcessing: Data Sampling
  - <u>Conditional Sampling</u>:
    - Cluster data with respect to one important variable
      - User input with physical intuition.
      - First Principal Component for full automation.
    - Subsample randomly from each cluster.
    - One cluster corresponds to purely random downsampling.
    - Too many clusters biases sampling away from highly sampled clusters so could influence prediction of most probably samples.

Shift
Ctrfl

Data-Based Modeling

- Data-PreProcessing: Data Sampling
  - Application: Turbulent Premixed Flames[1,2]
    - 4M observations downsampled to 750k observations in 7D space.
    - No apparent loss in “coverage” of the downsampled data.

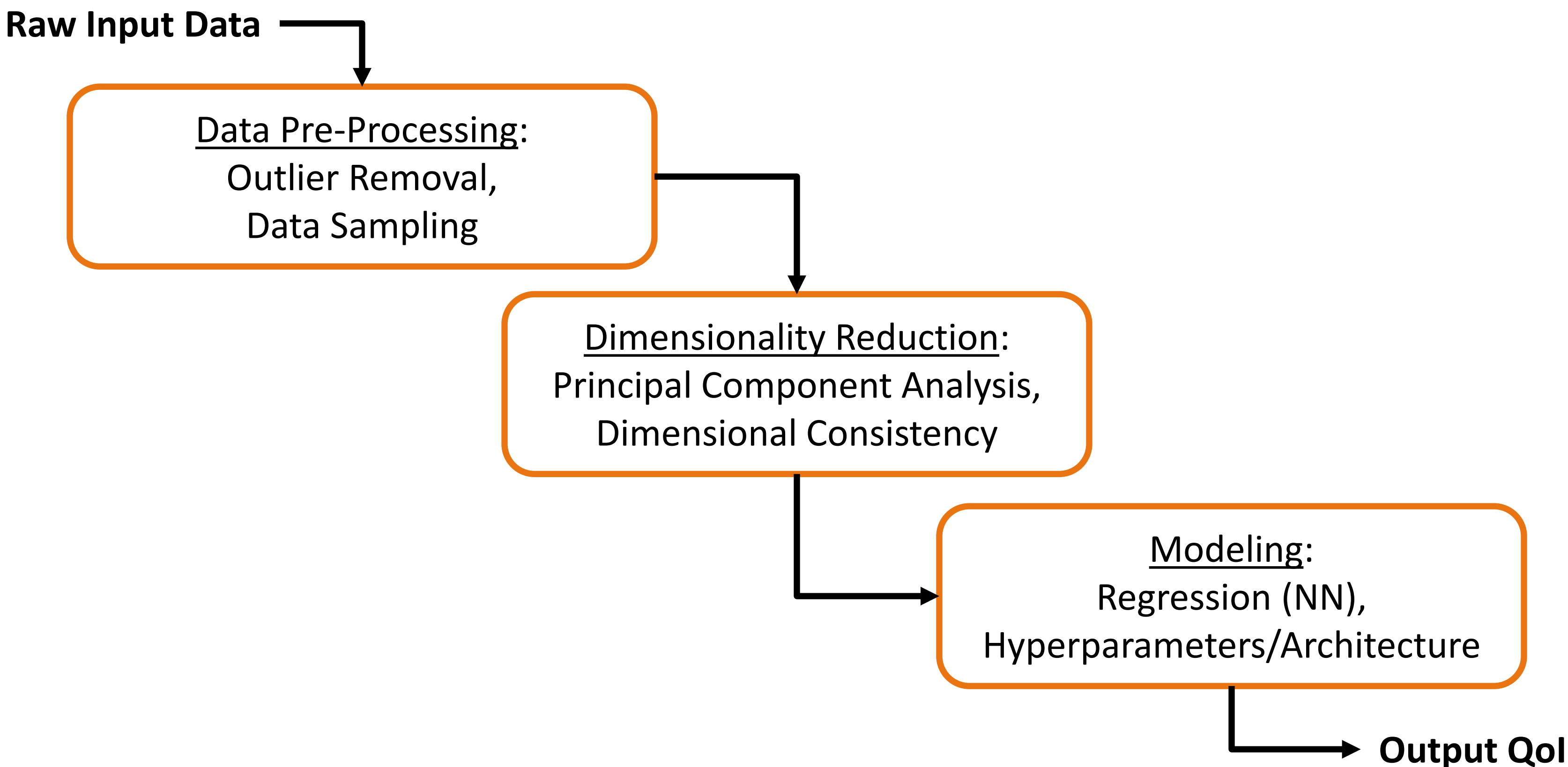


[1]J.F. MacArt, T. Grenga, M.E. Mueller, Combust. Flame 191 (2018) 468-485

[2]J. Lee, J.F. MacArt, M.e. Mueller, Combust. Flame 216 (2020) 1-8

Data-Based Modeling

- Data-Based Modeling Workflow



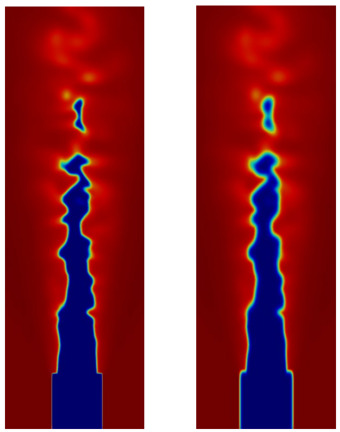
- Philosophy: Automated, Efficient, Simple

Shift
Ctrfl

Data-Based Modeling

- Dimensionality Reduction: Dimensional Consistency
  - Filtered Progress Variable Dissipation Rate Modeling[1]
    - Inputs: $\widetilde{\Lambda}, \Lambda_v, \left|\nabla\widetilde{\Lambda}\right|, |\tilde{S}|, \Delta_L, \widetilde{D}, \overline{\dot{m}}_\Lambda, \bar{\rho}$

DNS: Planar Jet Flames[2,3]

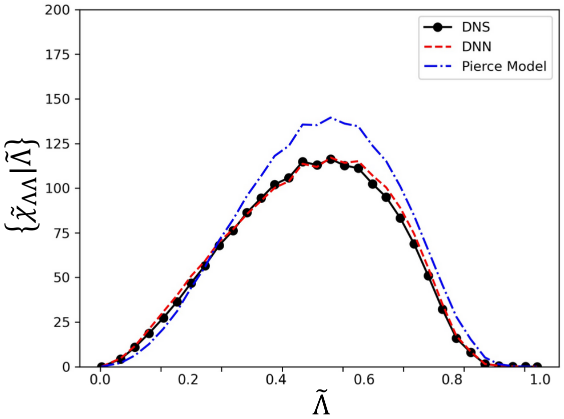


Train Dissipation Rate Model with DNN

[1]C.E. Lacey, G. D'Alessio, S. Sundaresan, M.E. Mueller (2022) in preparation
[2]J.F. MacArt, T. Grenga, M.E. Mueller, Combust. Flame 191 (2018) 468-485
[3]J. Lee, J.F. MacArt, M.e. Mueller, Combust. Flame 216 (2020) 1-8

Data-Based Modeling

- Dimensionality Reduction: Dimensional Consistency
  - Filtered Progress Variable Dissipation Rate Modeling[1]



8 Inputs
Network Architecture: 33 Layers, 36 Neurons
Mean Absolute Error: 0.208
Training Time: 3 hours

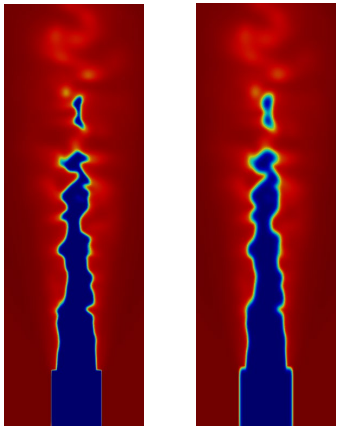
- This is a very large network that will be extremely expensive to evaluate in a forward coarse-grained simulation…

[1]C.E. Lacey, G. D'Alessio, S. Sundaresan, M.E. Mueller (2022) in preparation

Data-Based Modeling

- Dimensionality Reduction: Dimensional Consistency
  - Filtered Progress Variable Dissipation Rate Modeling[1]
    - Dimensional Inputs: $\widetilde{\Lambda}, \Lambda_v, \left|\nabla\widetilde{\Lambda}\right|, |\tilde{S}|, \Delta_L, \widetilde{D}, \overline{\dot{m}}_\Lambda, \bar{\rho}$
      - Different physical dimensions!
    - Dimensionally Consistent Inputs: $\frac{\widetilde{\Lambda}^2\widetilde{D}}{{\Delta_L}^2}, \frac{\Lambda_v\widetilde{D}}{{\Delta_L}^2}, \frac{\widetilde{\Lambda}|\nabla\widetilde{\Lambda}|\widetilde{D}}{\Delta_L}, \widetilde{\Lambda}^2\left|\tilde{S}\right|, \frac{\tilde{C}\overline{\dot{m}}_C}{\bar{\rho}}$
      - Smaller input space with the same information, and the results are not sensitive to the chosen combination of the inputs.

DNS: Planar Jet Flames[2,3]



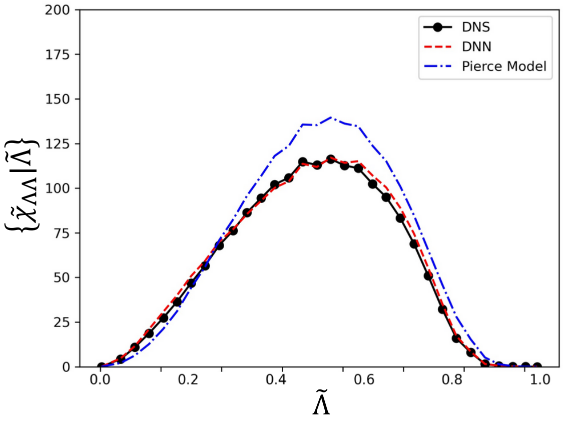
$\Delta = 4$ $\Delta = 16$

Train Dissipation Rate Model with DNN

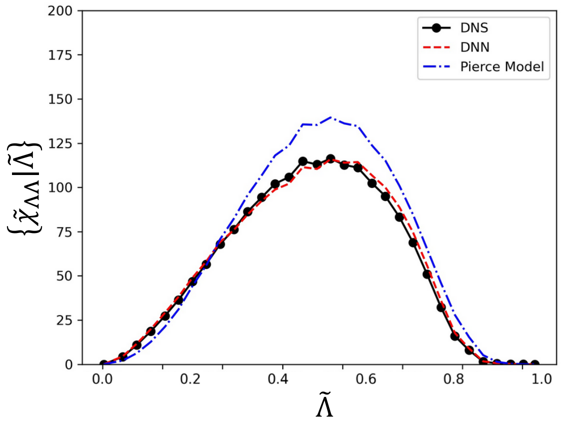
[1]C.E. Lacey, G. D'Alessio, S. Sundaresan, M.E. Mueller (2022) in preparation
[2]J.F. MacArt, T. Grenga, M.E. Mueller, Combust. Flame 191 (2018) 468-485
[3]J. Lee, J.F. MacArt, M.E. Mueller, Combust. Flame 216 (2020) 1-8

Data-Based Modeling

- Dimensionality Reduction: Dimensional Consistency
  - Filtered Progress Variable Dissipation Rate Modeling[1]



8 Dimensional Inputs
Network Architecture: 33 Layers, 36 Neurons
Mean Absolute Error: 0.208
Training Time: 3 hours



5 Dimensionally Consistent Inputs
Network Architecture: 13 Layers, 34 Neurons
Mean Absolute Error: 0.078
Training Time: 1.5 hours

- Input Choice: Use fewer inputs through considering physical dimensions without any loss of information/accuracy but with much smaller network (faster to evaluate) and faster training time.

Shift
Ctrfl

[1]C.E. Lacey, G. D'Alessio, S. Sundaresan, M.E. Mueller (2022) in preparation

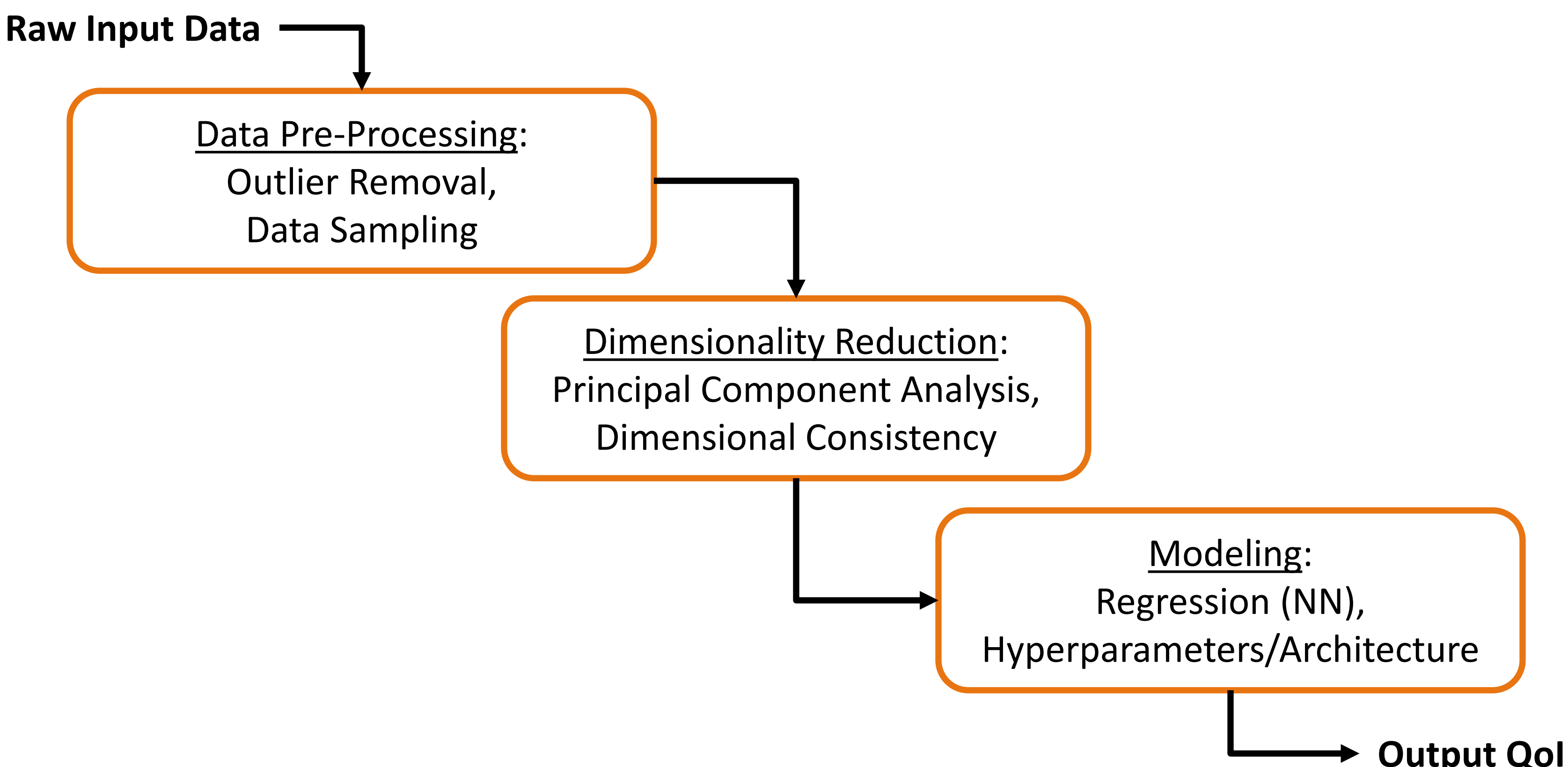
Data-Based Modeling

- Dimensionality Reduction: Dimensional Consistency
  - Vector Quantities of Interest
    - No need to limit approach to scalar quantities of interest!
    - We have also applied the same ideas to the local subfilter variation of the progress variable dissipation rate, that is, a vector quantity of interest[1].
  - Open Question
    - How does dimensional reduction via dimensional consistency compare to purely data-based Principal Component Analysis?

[1]C.E. Lacey, S. Sundaresan, M.E. Mueller, Combust. Flame (2022) in preparation

Data-Based Modeling

- Data-Based Modeling Workflow



- Philosophy: Automated, Efficient, Simple

Data-Based Modeling

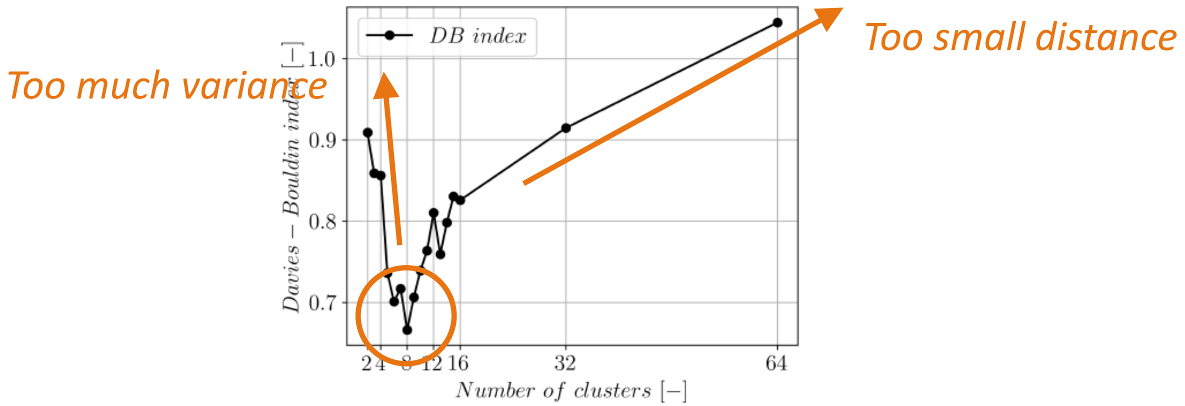
- Modeling: Efficient Regression[1]
  - What regression approach should be utilized?
    - Linear regression is very fast to train and evaluate but less accurate.
    - Neural networks are accurate but slower to train and evaluate.
  - Key Concept: Use both! Cluster the data and use linear regression in each cluster when sufficiently accurate (everything is linear locally enough…) or a neural network when needed.
    - Will this not just add yet more hyperparameters and require yet more hand-tuning?
    - Prevailing Strategy: Automate it!
    - Application[1]: Regression of reduced-order thermochemical state from turbulent nonpremixed flames[2] using Principal Component Analysis (PCA).

Shift
Ctrfl

[1] G. D'Alessio, S. Sundaresan, M.E. Mueller, Proc. Combust. Inst. (2022) submitted
[2] A. Attili, F. Bisetti, M.E. Mueller, H. Pitsch, Combust. Flame 161 (2014) 1849-1865

Data-Based Modeling

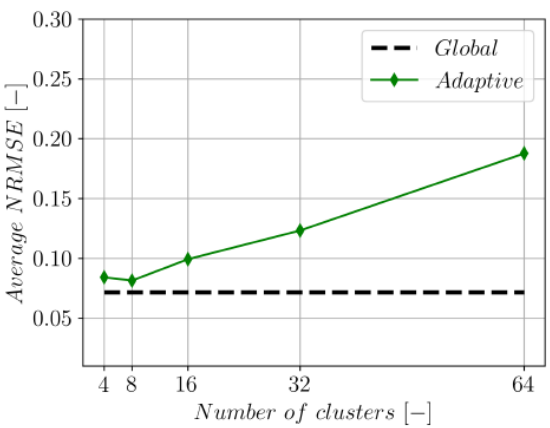
- Modeling: Efficient Regression[1]
  - Clustering – How many clusters should be chosen?
    - Too few clusters: Large variance in data within cluster
    - Too many clusters: Less distinction between clusters
    - Davies-Bouldin (DB) Index
      - Balances data variance in cluster with distance between clusters



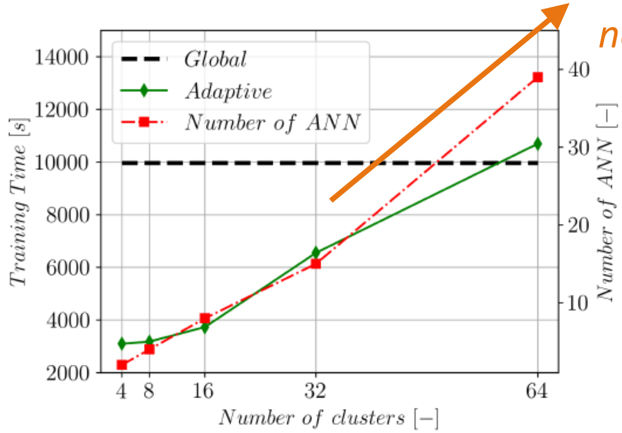
      - Choose minimum DB for optimal number of clusters

Data-Based Modeling

- ## Modeling Efficient Regression[1]
  - Regression – How accurate and how fast is the regression?



Too many networks

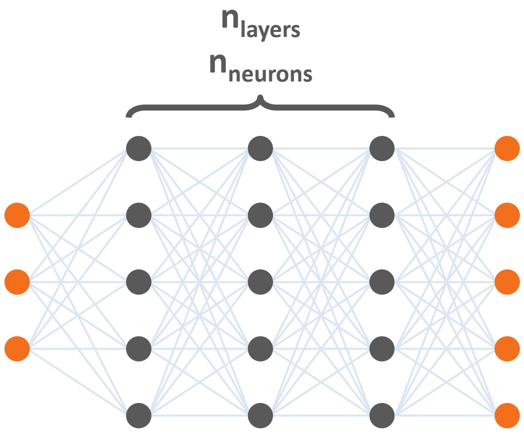


  - Remarkably, the local adaptive regression is most accurate when the clustering is best and is as accurate as one global neural network.
  - With local adaptive regression, the training time is ***<u>substantially reduced</u>*** even though more than one neural network (albeit simpler) must be trained due to the leveraging of local linear regression.

[1]G. D'Alessio, S. Sundaresan, M.E. Mueller, Proc. Combust. Inst. (2022) submitted

Data-Based Modeling

- Modeling: Hyperparameters/Architecture

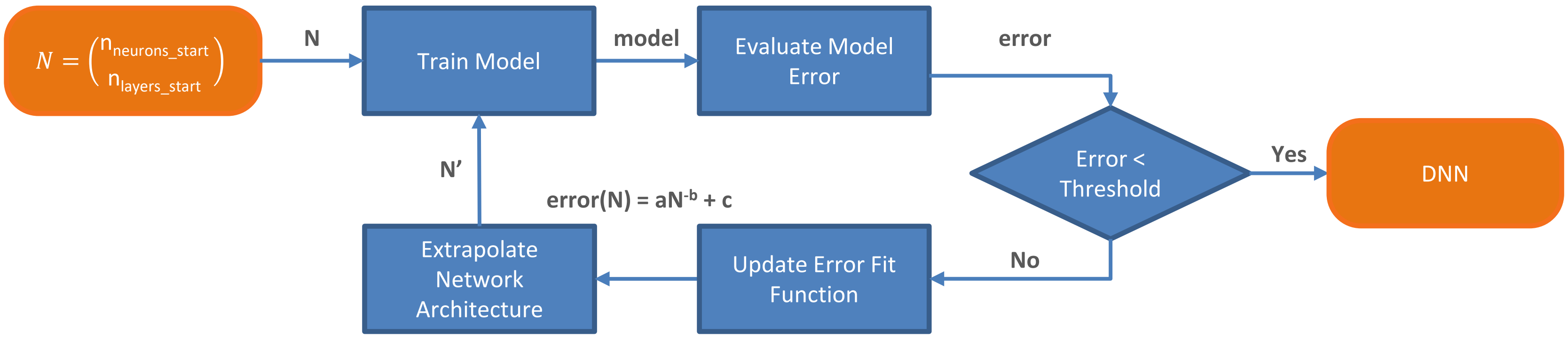


  - The "common" approach to neural network development is hand-tuning the architecture until an acceptable error is achieved.
    - Hand-tuning is not automatic!
  - This "common" approach typically leads to suboptimal, bloated, overfit networks that are expensive to train and expensive to evaluate.
    - Expensive is not efficient and simple!

Shift
Ctrfl

Data-Based Modeling

- Modeling: Hyperparameters/Architecture
  - First Approach: “Convergence” Rate



- Fixed activation function, optimizer, and learning rate.
- This approach was automatic and worked but tends to lead to excessively large networks because the power law fit to the error is just not very good.

Shift
Ctrfl

Data-Based Modeling

- Modeling: Hyperparameters/Architecture
  - Second Approach: Bayesian Optimization[1]
    - Basic idea is to construct a probabilistic model in which some acquisition function is minimized to arrive at an optimal model with highest accuracy.
      - The acquisition function leads the approach to the next sampling location within the model parameter space.
    - Can include not only the architecture but also the activation function and the learning rate as part of the optimization process.

    - Key Question: What acquisition function should be chosen?
      - Probability of Improvement – Targets regions with highest uncertainty
      - Expected Improvement – Targets regions with highest uncertainty without straying too far from known "best" model
      - Lower Confidence Bound – "Greedy" algorithm that targets regions with potentially optimal model (mean minus standard deviations)

Shift Ctrfl

1J. Snoek, H. Larochelle, R.P. Adams, Adv. Neural Inform. Proc. Sys. 25 (2012)

Data-Based Modeling

- Modeling: Hyperparameters/Architecture
  - Application: Regression of reduced-order thermochemical state from turbulent nonpremixed flames[1] using Principal Component Analysis (PCA), specifically the Principal Component source terms.

Network accuracy

	ST 1	ST 2	ST 3	ST 4	ST5
a_{PI}	0.0424	0.1440	0.0407	0.0577	0.0697
a_{EI}	0.0410	0.1442	0.0416	0.0564	0.0736
a_{LCB}	0.0400	0.1312	0.0391	0.0559	0.0676

Network design

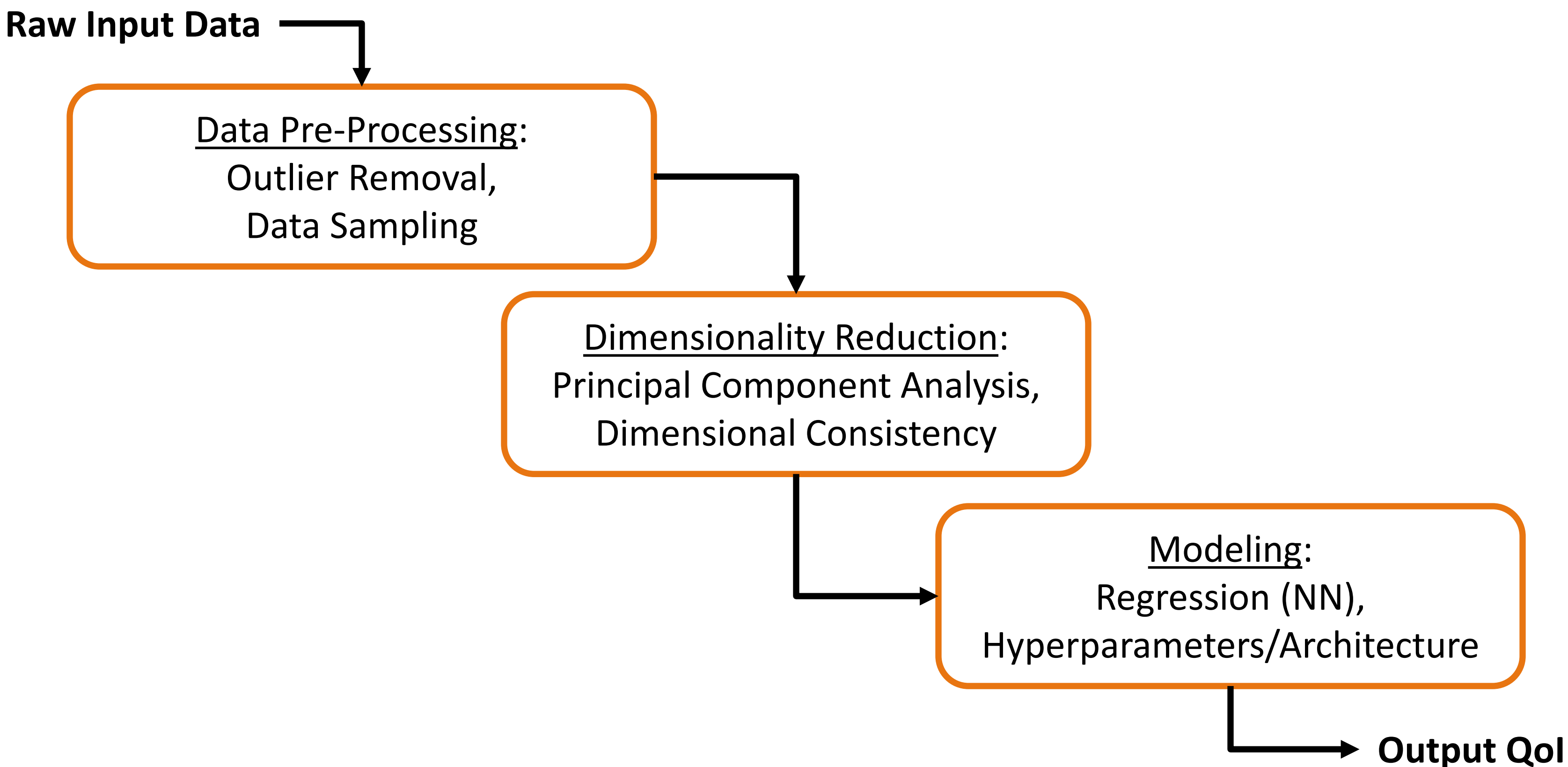
	Neurons [-]	Layers [-]	Activation [-]	Learning rate [-]	Training time [s]
a_{PI}	175	13	ELU	0.00131	14443.5
a_{EI}	414	11	SELU	$1.199\ 10^{-5}$	21202.4
a_{LCB}	**189**	**2**	SELU	0.00538	**10583.6**

- Lower Confidence ound acquisition function leads to most accurate and smallest network with the fastest training time!

[1]A. Attili, F. Bisetti, M.E. Mueller, H. Pitsch, Combust. Flame 161 (2014) 1849-1865

Data-Based Modeling

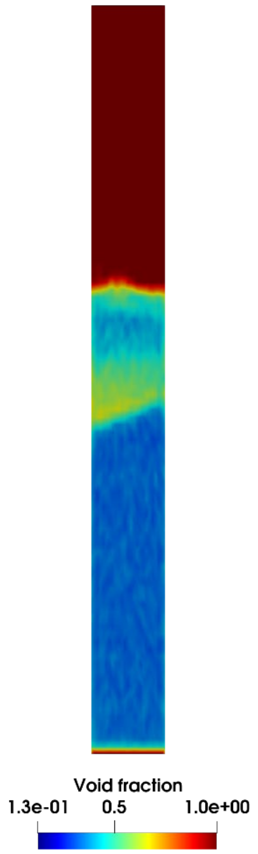
- Data-Based Modeling Workflow



- Philosophy: Automated, Efficient, Simple

Application

- Fluidized Bed Reactor (Work-in-Progress)
  - Training data from fine-grid TFM simulations[1].
  - No data pre-processing.
  - Inputs (non-dimensional):
    - Reynolds Number: Re
    - Filter Size: Δ/L
    - Solid Volume Fraction: $\phi_s/\phi_{\max}$
    - Pressure Gradient: $\nabla p/\rho_s g$
    - Slip Velocity: $U_{\mathrm{slip}}/U_{\mathrm{terminal}}$
  - Quantity of interest is filtered particle drag.
  - Monolithic neural network with Bayesian optimization.

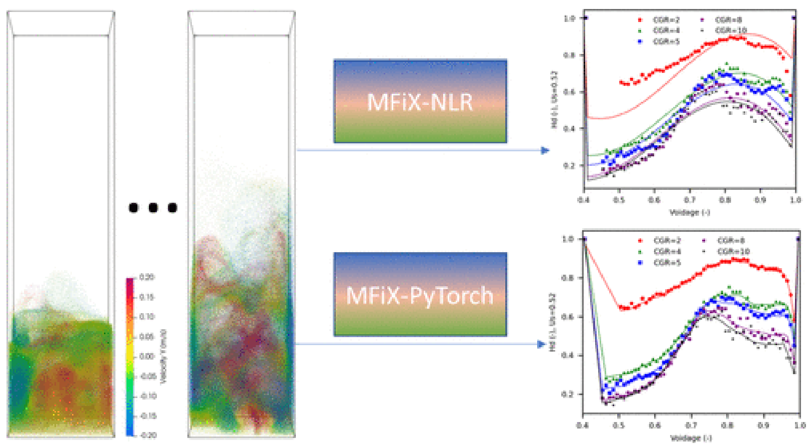


  - Application to coarse-grid Euler-Lagrange simulation:

[1]Y. Jiang, J. Kolehmainen, Y. Gu, Y.g. Kevrekidis, A. Ozel, S. Sundaresan, Powd. Tech. 346 (2019) 403-413

Collaboration

- Collaboration with NETL
  - Prior Work at NETL[1]:
    - Coupling MFiX with data-based filtered drag model derived from fine-grid CFD-DEM simulations.
    - Limited success with neural networks due to challenges in finding the best hyperparameters/architecture.



    - Goal: Use our automated approach with their training data and implement in MFiX.

[1]L. Lu, X. Gao, J.-F. Dietiker, M. Shahnam, W.A. Rogers, Ind. Eng. Chem. Res. 61 (2022) 882-893

Data-Based Modeling

- Data-Based Modeling Workflow

Raw Input Data

<u>Data Pre-Processing</u>:
Outlier Removal,
Conditional Sampling

<u>Dimensionality Reduction</u>:
Principal Component Analysis,
Dimensional Consistency

<u>Modeling</u>:
Adaptive Regression,
Bayesian Optimization

Output QoI

- <u>Philosophy</u>: Automated, Efficient, Simple

Shift
Ctrfl

Acknowledgements

- Princeton Postdocs and Graduate Students
  - Dr. Giuseppe D’Alessio
  - Dr. Mostafa Sulaiman
  - Mr. Cristian E. Lacey

- Funding
  - UCFER (S001342-USDOE)
  - BES/SCGSR
  - Princeton University (Research Computing)