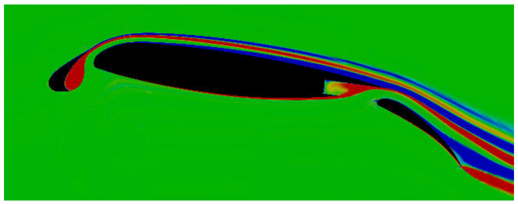
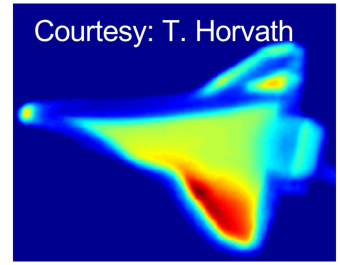
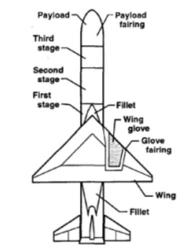
Exploring Data-driven Modeling of Boundary Layer Transition

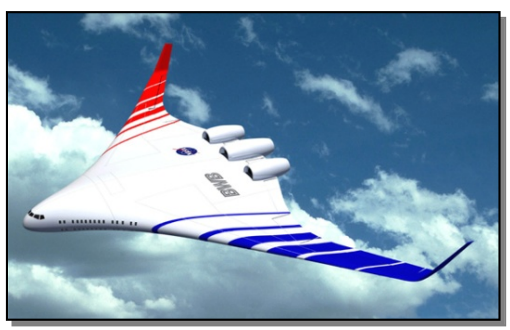






M. Choudhari, M. R. Malik, P. Paredes,
J. Berninger, K. Hausknecht, K. Perlin, R. Qiu
NASA Langley Research Center

M. I. Zafar, H. Xiao
Virginia Tech



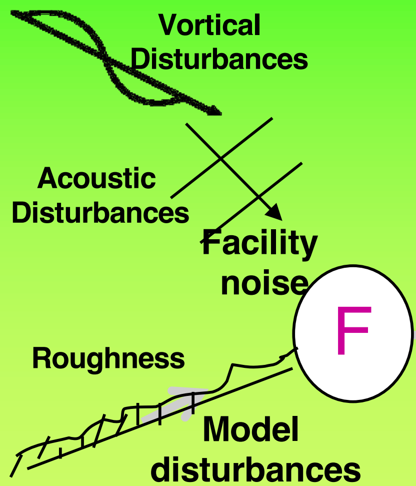
Motivation

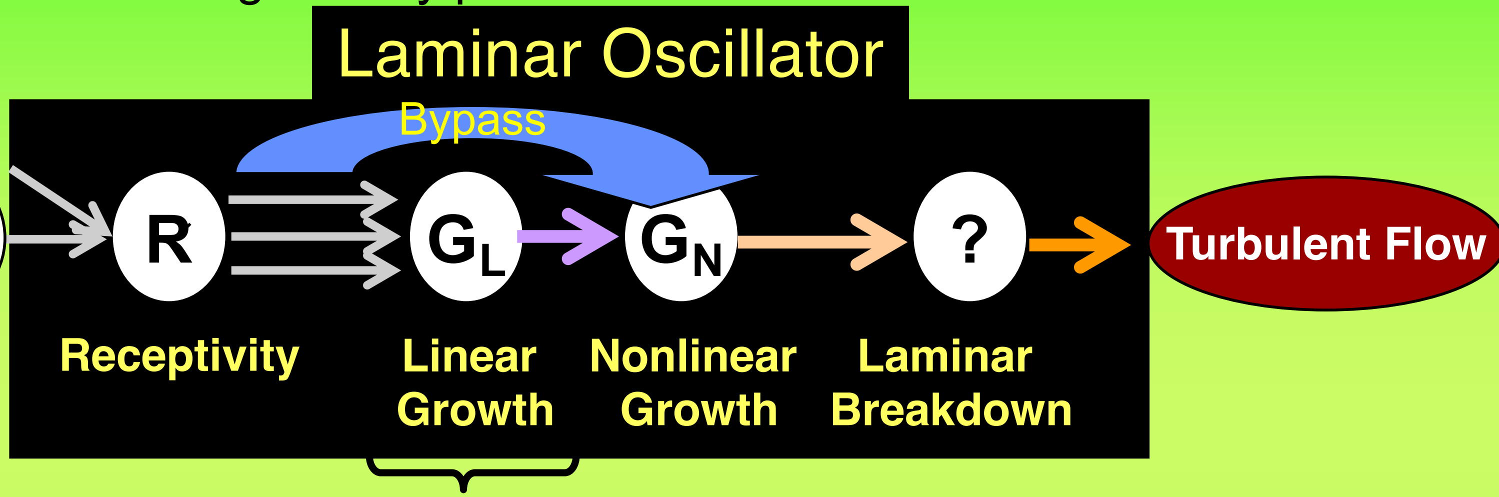
- **CFD Vision 2030 (Slotnick et al. 2014)**
  - "*The most critical area in CFD simulation capability that will remain a pacing item is the ability to adequately predict viscous turbulent flows with possible transition and flow separation*."
- **Greener aircraft concepts for fuel burn reduction**
  - In 2012, U.S. commercial carriers burned 10.6 billion gallons of jet fuel while the defense operations in 2011 accounted for approximately 3.0 billion gallons, with an estimated annual fuel cost of $42 billion within U.S. alone.
  - Laminar flow technology can provide breakthrough reductions in fuel burn
  - Uncertainty in transition modeling can offset benefits from laminar flow concepts
- **TPS design for hypersonic re-entry vehicles**
  - Transition accompanied by 3x-8x increase in surface heat transfer
  - Transition modeling accounts for highest uncertainty in surface heating
- **Challenge: Many Paths for Transition**
  Multiple Relevant Flow Configurations

Physics Based Transition Modeling

Transition = Forced response of laminar BL

- Challenge: Many paths for transition!

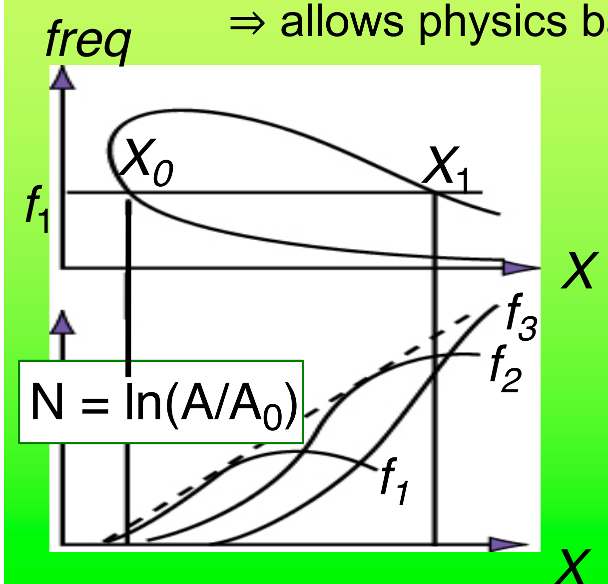




Slow phase accounts for majority of distance to transition onset

⇒ allows physics based correlation between x_{tr} and logarithmic amplitification factor N

$$\underbrace{N_{tr} = \underset{f}{\mathrm{Max}}\,[\ln (A/A_0)]}$$



Well-established approach in transition research,
but difficult to integrate with CFD codes!

Challenge: Linear stability calculations tend to be non-robust,
place high demands on basic state accuracy,
and often require significant user expertise

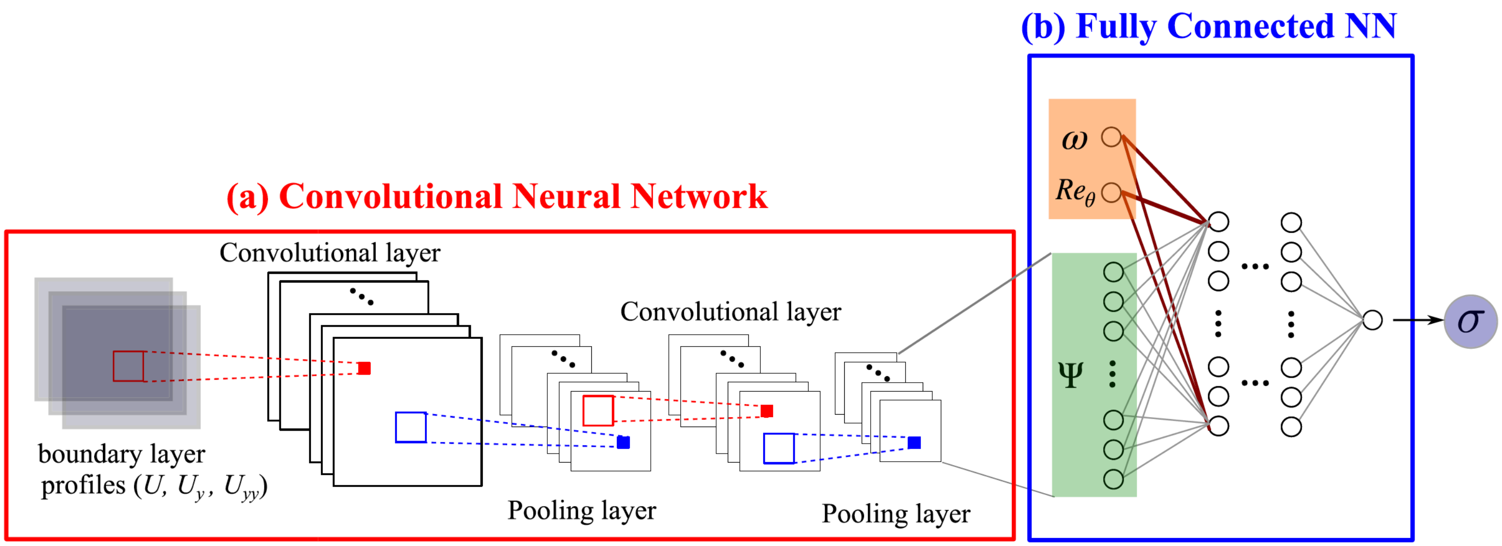
**Metamodels based on deep learning can alleviate this difficulty,
To help enable the CFD Vision 2030 goal of integrating physics based correlations with CFD solvers**

Outline

- **Motivation**
- **Physics Based Transition Modeling**
- **Data Driven Transition Modeling**
  - – Linear stability correlations
    - ➢ Examples from subsonic to hypersonic flows
      - ▪ TS Waves
      - ▪ Crossflow Vortices
      - ▪ Mack Modes
    - ➢ **Physical interpretation of latent features from CNN**
    - ➢ **Dealing with large datasets, selection of training data**
  - – **Modeling with UQ** (Gaussian Process Regression)
  - – Higher fidelity modeling: secondary instability correlations for crossflow transition
    - ➢ Multi-mode models
    - ➢ Mode shape prediction
- **Concluding Remarks**

Modeling the Amplification of T-S Instabilities in 2D BLs

- Existing data-driven models based on analytical curve fits or rapid interpolation techniques
  - Not well-suited for large number of “stability modifiers”
  - Do not allow easy modifications for custom/new data
- CNN-encoder architecture provides a computationally efficient alternative to conventional fully connected networks
  - Can also enable physical interpretation of learned features of BL



Encodes boundary layer profiles to latent physical parameters

Maps physical parameters to instability amplification rate

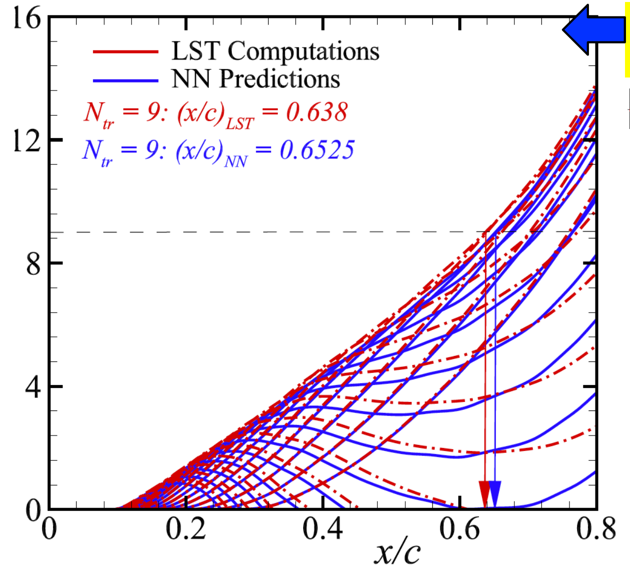
Zafar et al., Phys. Rev. Fluids, 2020

Selection of Training Dataset

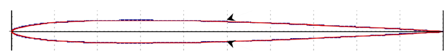
- Existing model for TS amplification by Drela and Giles (1988)
  - Based on analytical curve fits to stability calculations for selfsimilar Falkner-Skan (FS) boundary layers
  - BL profiles parameterized in terms of H (shape factor of velocity profile) and Re_θ (local Reynolds number based on momentum thickness)
  - Generally successful extrapolation to non-selfsimilar BLs over airfoils, but invalid for complex flows such as confluent boundary layers over high-lift configurations

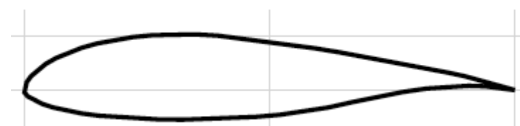
→ First application of CNN architecture used FS flows for training database

Validation of CNN based N-factor Predictions for Airfoils

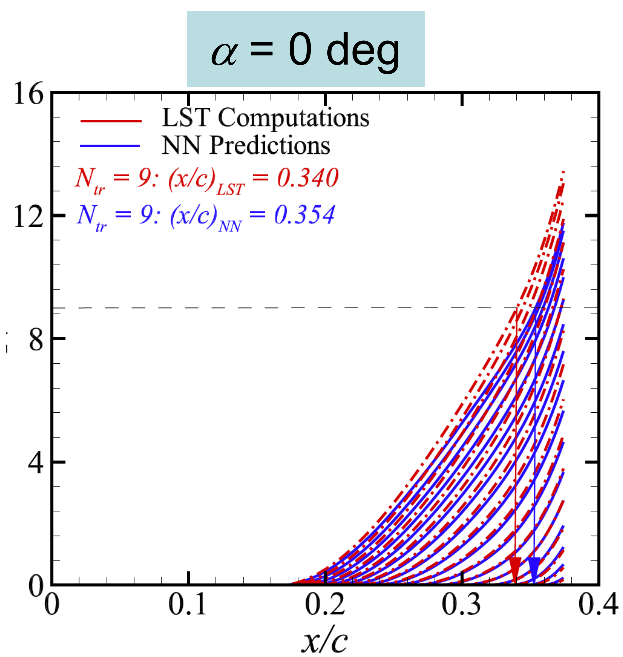


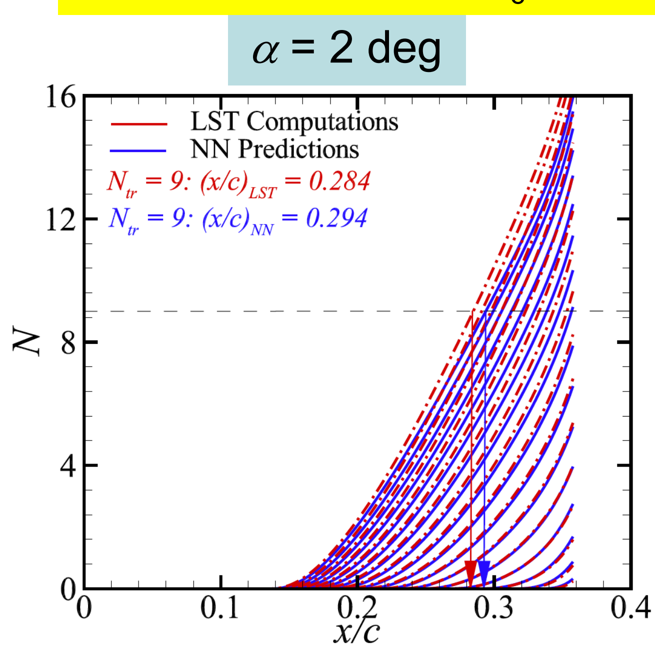
HSK Airfoil: α = 0 deg, $Re_c = 1.23 \times 10^6$

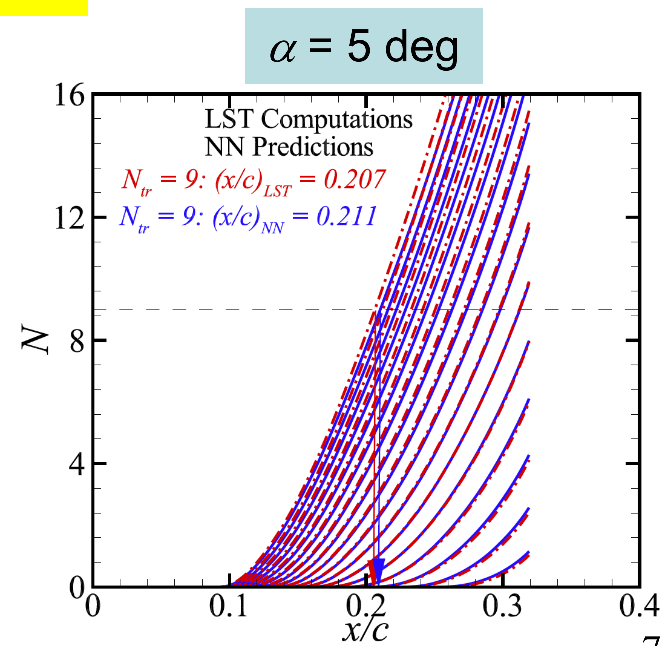




NLF-0416 Airfoil: $Re_c = 9 \times 10^6$

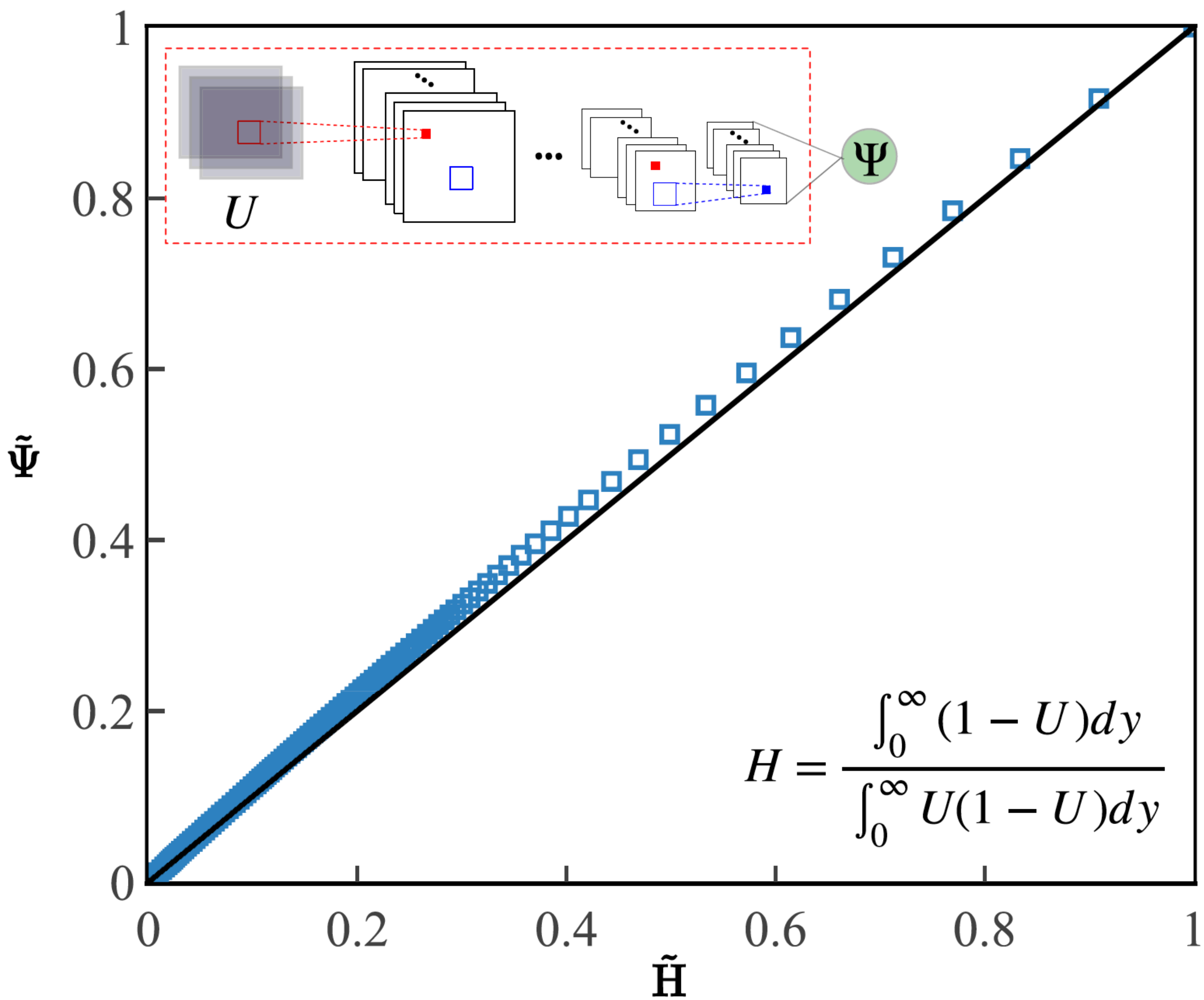






Physical Interpretation of CNN Extracted Latent Features

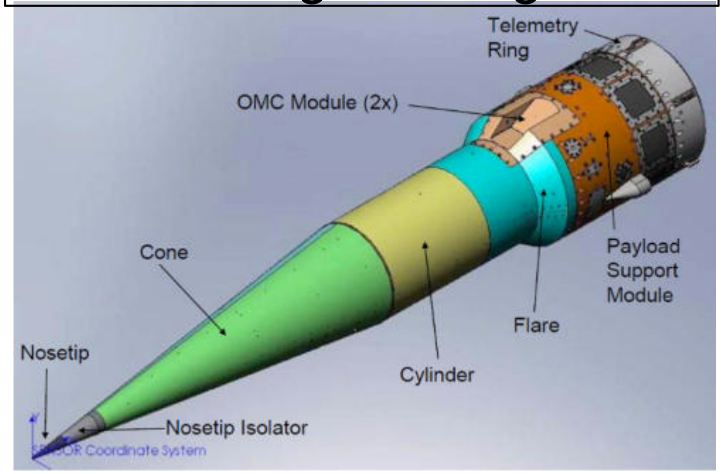
- Correlation between feature (Ψ) extracted by CNN from airfoil velocity profiles and profile shape factor (H) shows that **CNN is learning the velocity shape factor**

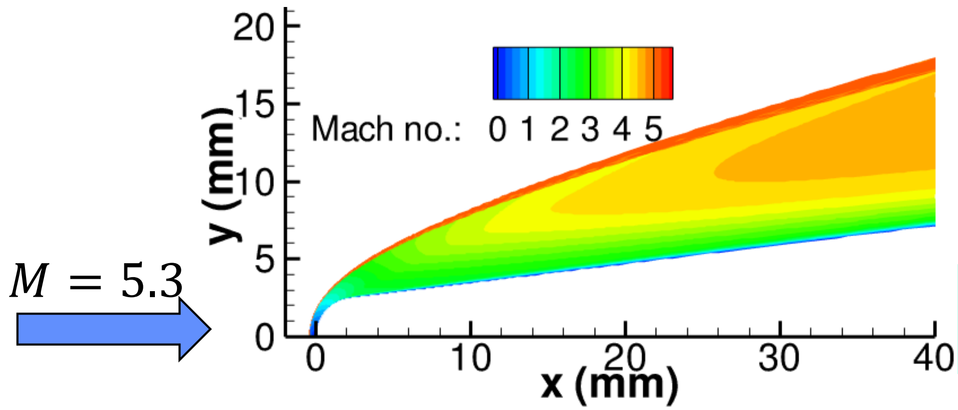


Normalized shape factor

Surrogate Models for Mack Mode Amplification

HIFiRE-1 flight configuration





1.0 m long cone

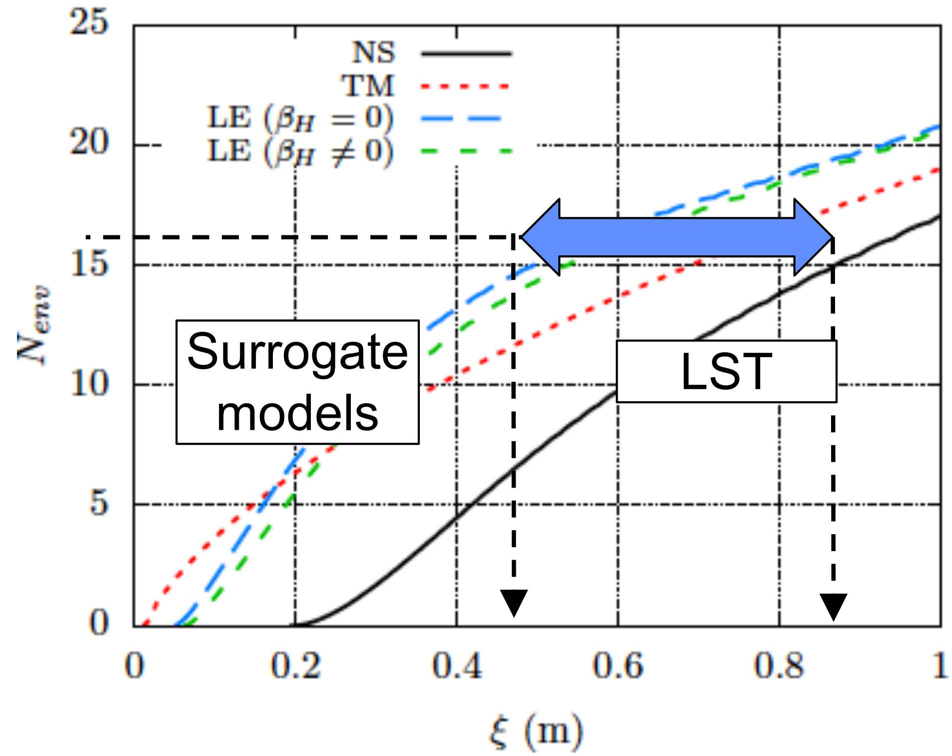
- Nonzero (albeit small) nose radius alters the mean flow characteristics relative to self-similar boundary layer over a sharp cone (e.g., HIFiRE-1 cone with R_n=2.5 mm)
  - Presence of entropy layer leads to challenges in reliable evaluations of integral profile characteristics
  - Local approximations as self-similar profiles that match the local boundary-layer properties can match the unstable frequency range of actual flow, but significantly overpredict the Mack mode growth rates, leading to large discrepancies in stability correlations
- Convolutional neural network (CNN) model uses entire input profiles, instead of scalar, integral parameters
- Automated generation of relatively large database of N-S computations

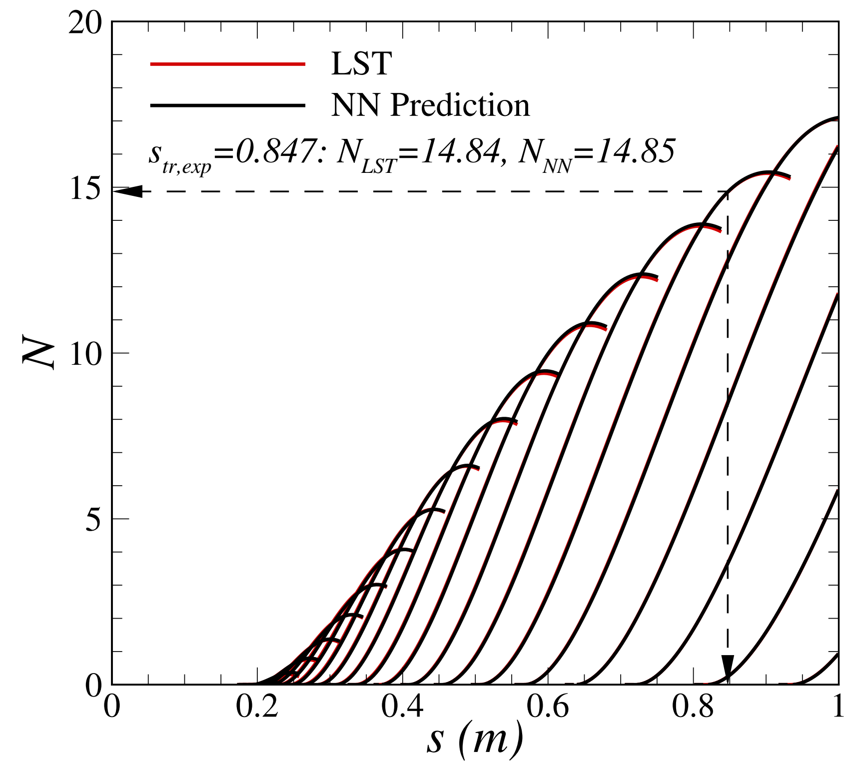
Conventional Database Lookup vs. CNN Model

Transition Correlation for HIFiRE1 in Ascent phase (t = 21.5 sec)

Correlations based on integral parameters
➔ O(40%) error in transition location

Deep learning ⇒ excellent agreement not just in x_{tr}, but in all N-factor curves





Local approximations as self-similar profiles that match the local boundary-layer properties can match the unstable frequency range of actual flow, but significantly overpredict the Mack mode growth rates, leading to large discrepancies (up to 40%) in predicted transition location

→ No such difficulty for CNN based predictions for both dominant and subdominant Mack modes

(Paredes et al., AIAA J., 2020)

N-Factor Prediction Based on CNN Model

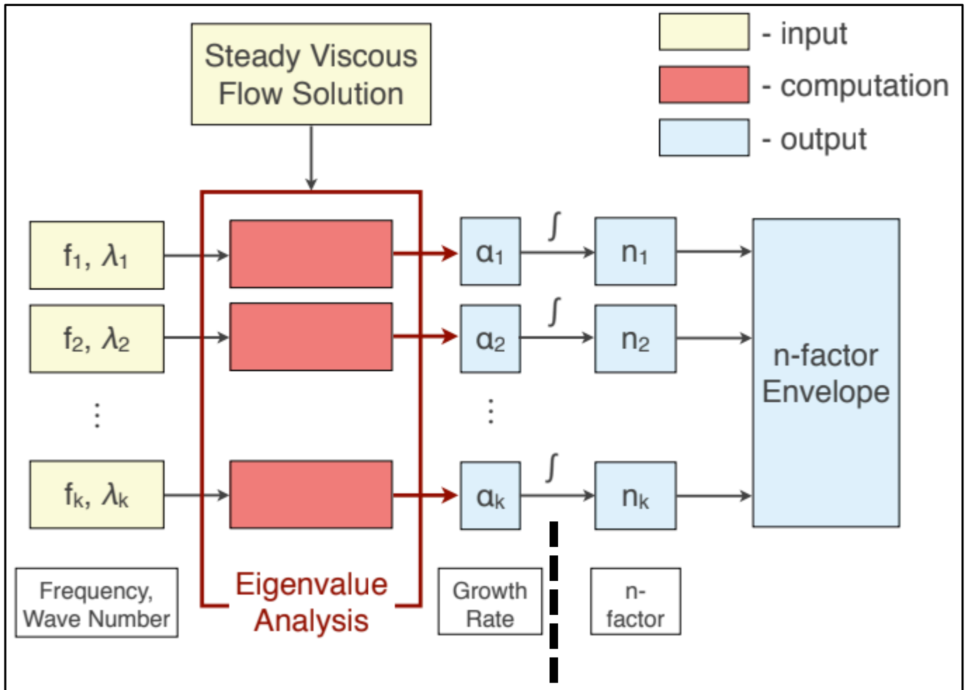


Figure 4. Overview of linear stability computations.

Rajnarayan and Sturdza (2011)

$$\text{N-factor} = \int_{s_{l.b.}}^{s} \sigma(s'; \omega^*, \beta^*) ds'$$

CNN model predicts local growth rate for a given instability mode

➔ Extra workflow requirements for N-factor calculations:

- Define range of frequencies and/or wavelengths to compute growth rates for
- Compute individual N-factor curves and then the envelope

⇒ Extra overhead in parallel codes

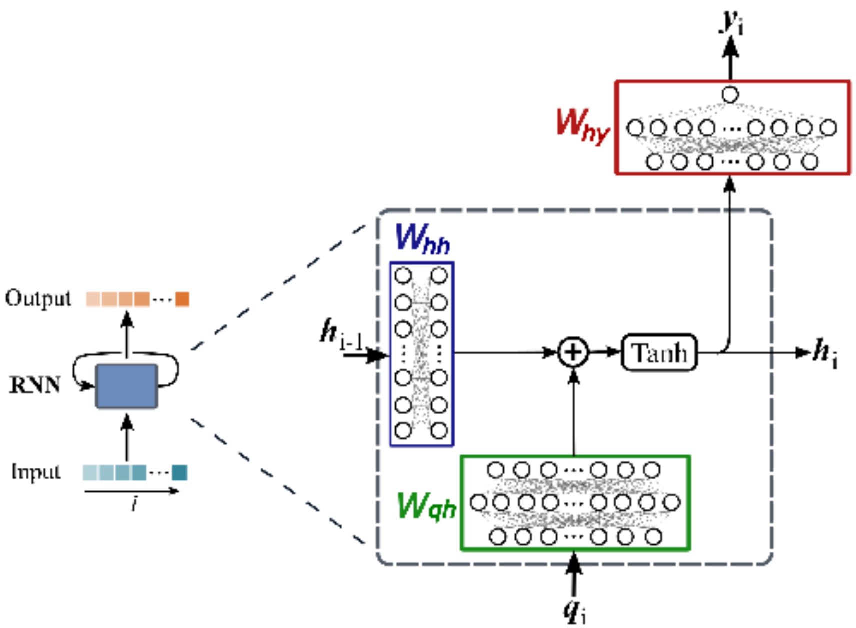
Recurrent Neural Network (RNN) for End-to-end Transition Modeling

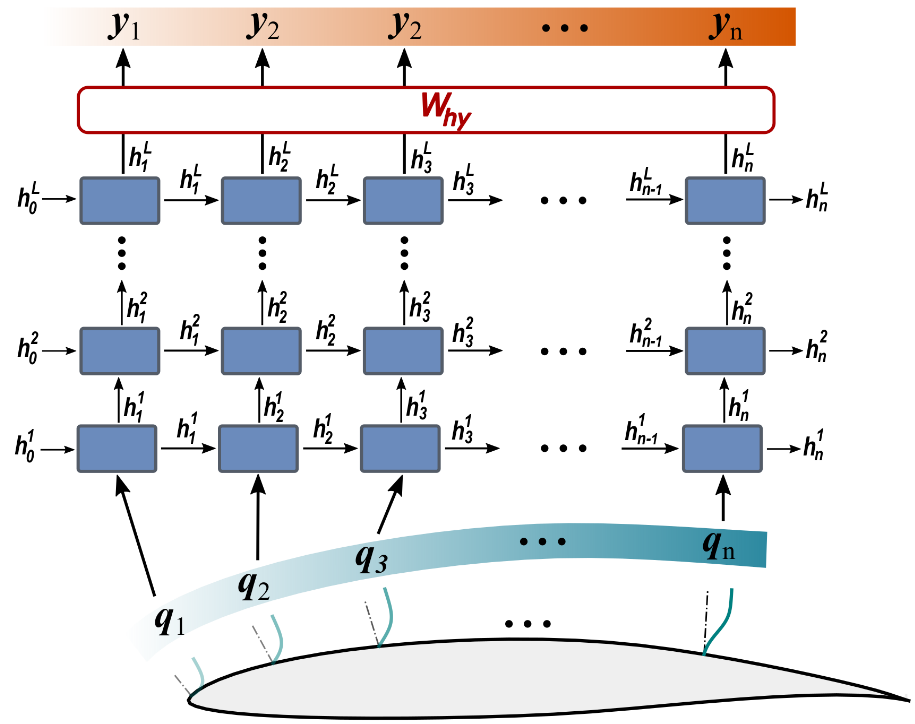
- Directly predicts N-envelope and transition location, providing end-to-end transition prediction based on e^N method

- Development of transition model using a large dataset of realistic airfoil BLs
  - Training data more representative of application space
  - But, also accompanied by other complexities involving the choice of training data
  - Provides insight into data sampling for practical applications, similarity/redundancy between different families of airfoils

Recurrent Neural Network (RNN)

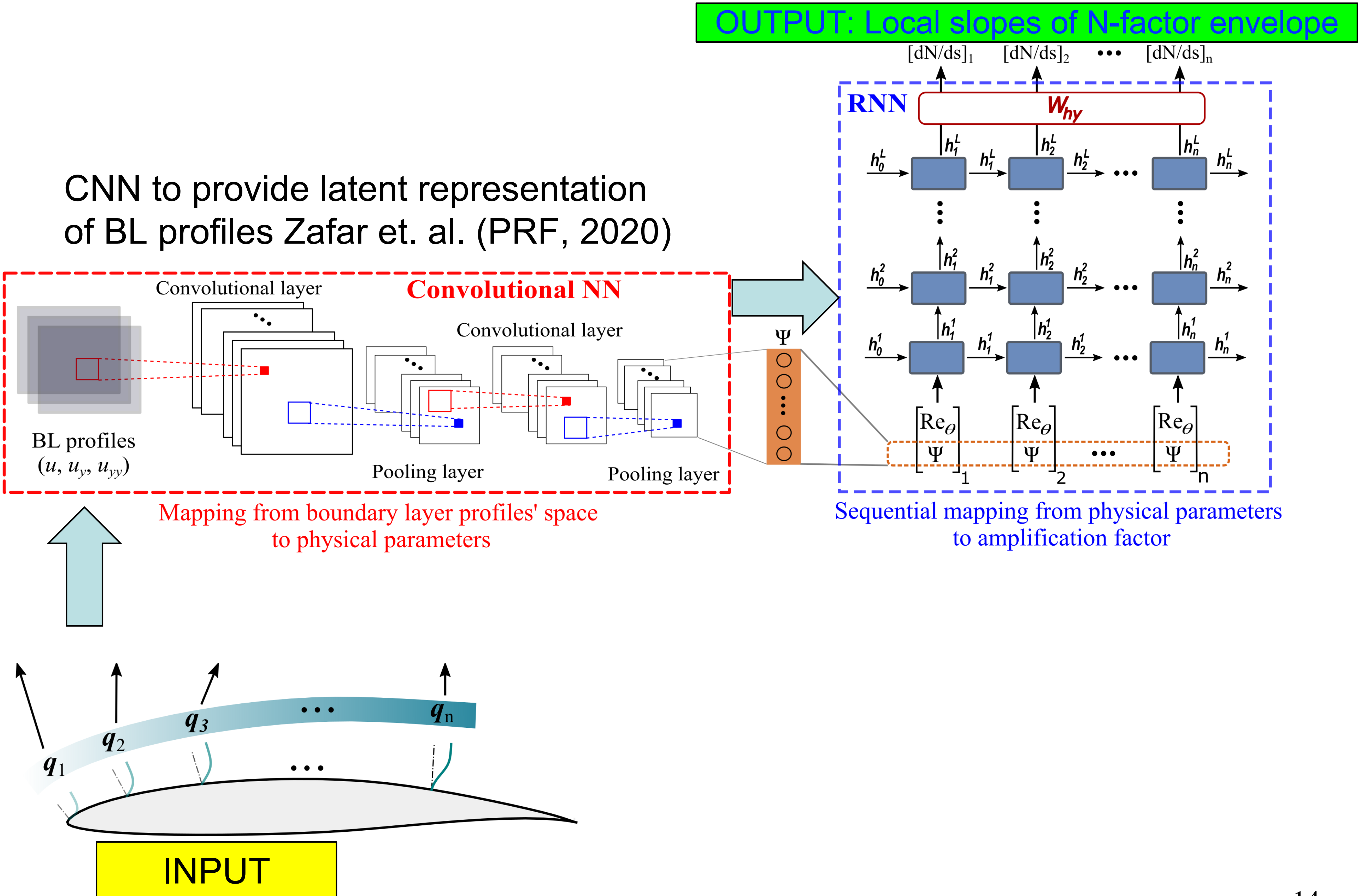
- A class of neural networks that have an internal memory (hidden state), and is well-suited for sequential data
- RNN directly maps the sequence of mean boundary layer-flow profiles to instability growth rate along the N-factor envelope





RNN Model

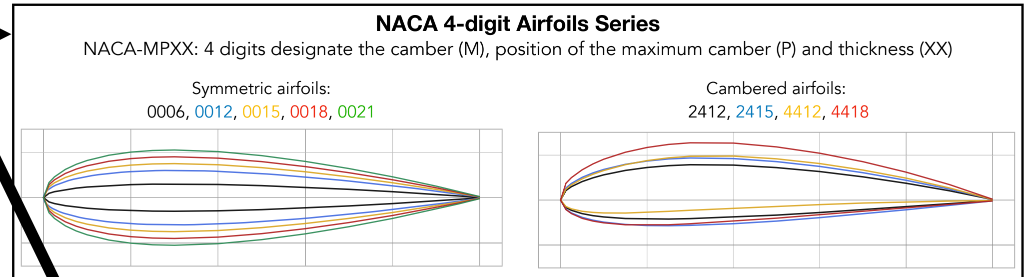
CNN to provide latent representation of BL profiles Zafar et. al. (PRF, 2020)

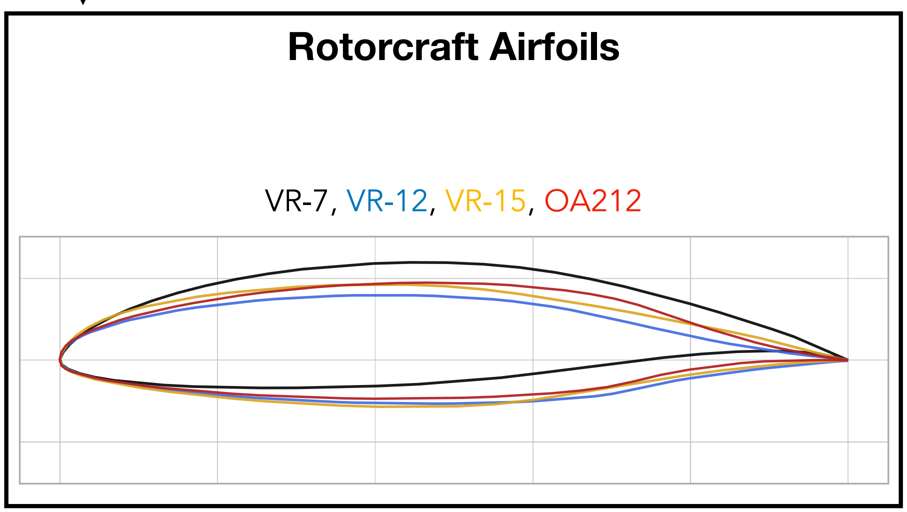


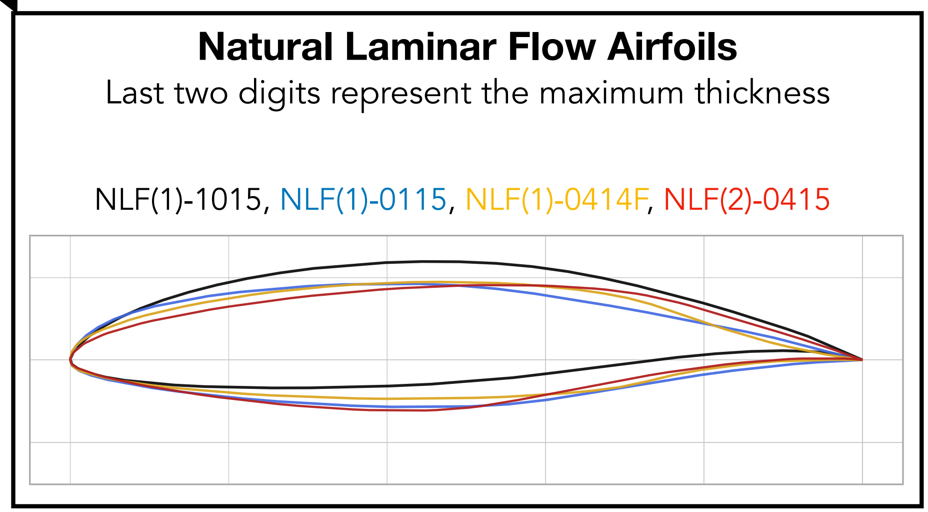
Database of TS Amplification over Airfoils

- Large database of amplification characteristics for Tollmien-Schlichting (TS) instabilities in 2D boundary layer flows over a large set of airfoils: **>40K flows with 120 stations each**

- 53 airfoils from different families:
  - NACA 4, 5 & 6-digit airfoil series
  - Natural laminar flow (NLF) airfoils
  - Rotorcraft airfoils
  - Low Reynolds number airfoils
  - General Aviation airfoils
  - Transonic airfoil contours, etc.



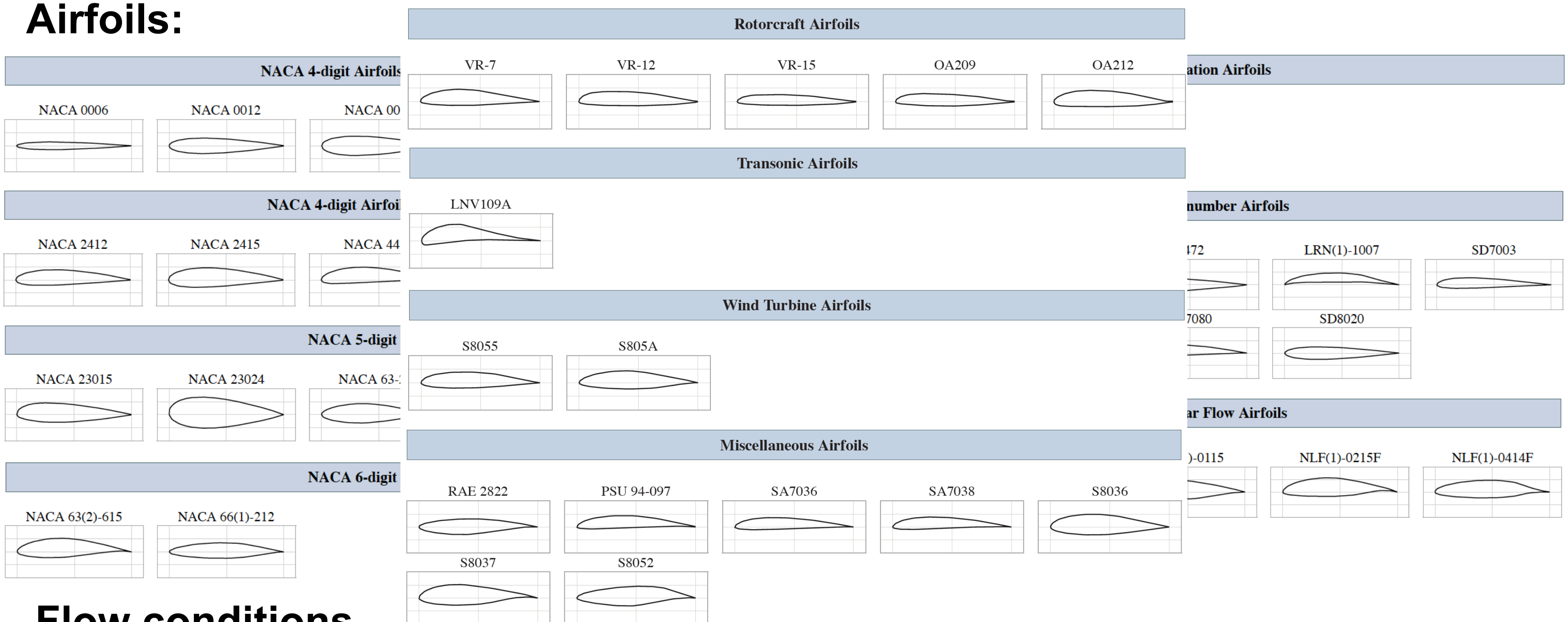




- >100K boundary layer flows at various flow conditions (airfoil contour, AoA, Re_c)

Dataset Available for Training

- **Airfoils:**



- **Flow conditions**

Angles of Attack (deg)	Reynolds Numbers
–6°, –5°, –4.5°, –4°, –3.5°, –3°, –2.5°, –2°, –1.5°, –1°, –0.5°, 0°, 0.5°, 1°, 1.5°, 2°, 2.5°, 3°, 3.5°, 4°, 5°, 6°, 7°, 8°	3.5×10^4, 5.0×10^4, 7.0×10^4, 1.0×10^5, 1.4×10^5, 2.8×10^5, 4.0×10^5, 5.6×10^5, 8.0×10^5, 1.1×10^6, 1.6×10^6, 2.3×10^6, 3.2×10^6, 4.5×10^6, 6.4×10^6, 9.0×10^6, 1.3×10^7, 1.8×10^7, 3.6×10^7, 5.1×10^7, 7.2×10^7, 1.0×10^8, 1.4×10^8

- **Error metrics:** Error in envelope N-factor

$$E_{N_{env}} = 100\times\frac{1}{m}\sum_{j=1}^{m}\left(\frac{\left\|N_{env}-\widehat{N}_{env}\right\|_{ds}}{\left\|N_{env}\right\|_{ds}}\right)_j$$

Error in transition location

$$E_{tr} = 100\times\frac{1}{M}\sum_{j=1}^{M}\left(\left|x/c_{tr}-\tilde{x/c}_{tr}\right|_j\right)$$

Big Data Issues: Selection of Optimal Training Data

- Quality of deep learning model strongly influenced by the selection of training data
- Relatively obvious criteria for selection of training data:
  - Small enough to maximize computational efficiency
  - Representative of application space
  - Balanced across relevant parameter space (to avoid bias)
- Difficult to apply these criteria in practice due to lack of formal procedures
  - **Casewise evaluation essential**

Training Cases

- E_{tr} refers to absolute error of transition location

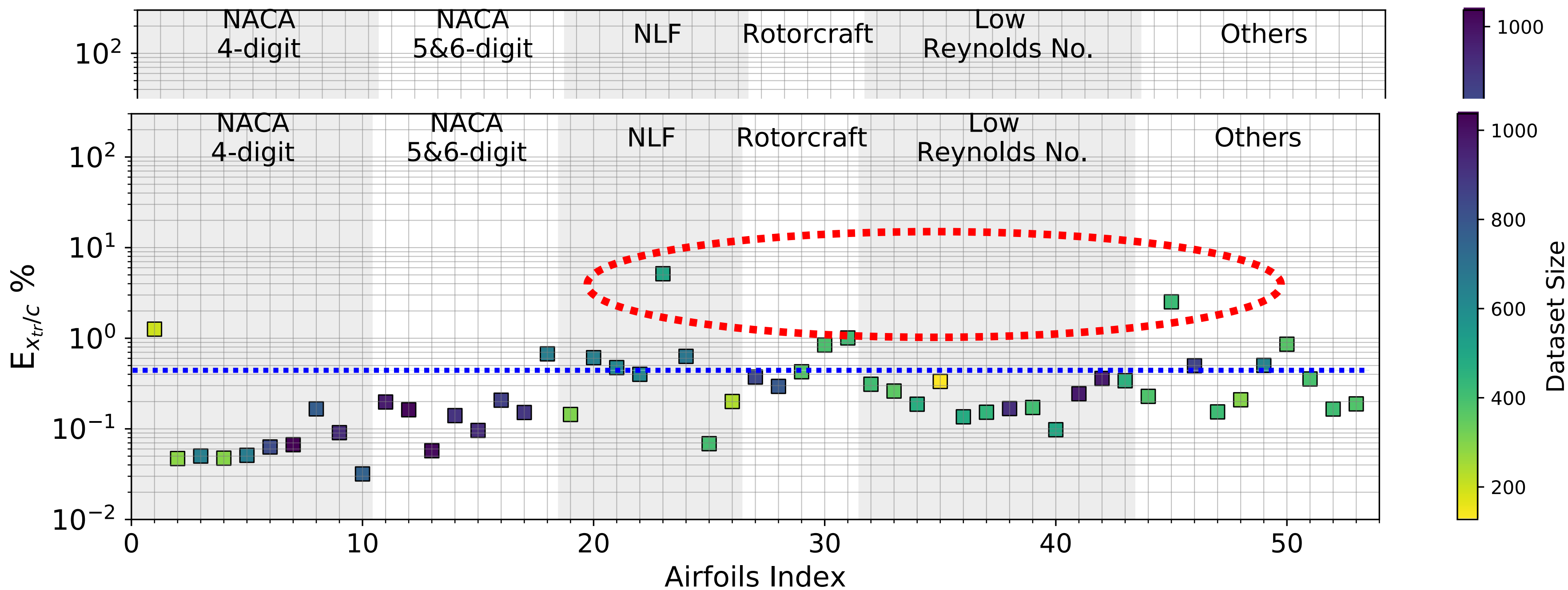
Case Index	Case Name	Training Dataset	Size - No. of BLs	Max Average Error for a single airfoil		Average Error of whole database	
				E_{env}	E_{tr}	E_{env}	E_{tr}
I	Base case	NACA 0012, NACA 2412, NACA 63-415, NLF(1)-0416, ONERA M6	2624	39.2%	6.52%	2.95%	0.42%
II	**Error based augmentation**	**Case I + specific BLs over other airfoils with $E_{env}\% > 3$ in Case I**	5024	5.38%	0.70%	1.53%	0.15%
III	Augmented set of airfoils based on error	Case I + all BLs over 5 other largest-error airfoils (LRN(1)-1007, NACA 6712, NLF(1)-1015, NLF(2)-0415, CLARK Y)	4455	14.5%	5.15%	2.71%	0.43%
IV	**Random augmentation**	**Case I + 100 random flow cases from each of the other airfoils**	**7026**	**8.95%**	**1.40%**	**1.92%**	**0.26%**
V-A	**Randomly selected BLs**	**Randomly selected 20% flow cases of each airfoil**	**6233**	**10.0%**	**1.51%**	**1.66%**	**0.22%**
V-B	**Randomly selected BLs**	**Randomly selected 100 flow cases of each airfoil**	**5300**	**5.49%**	**0.70%**	**1.60%**	**0.19%**

Absolute Error in Predicted Transition Location

Plots show mean absolute error in x_{tr}/c prediction for all flow cases of each airfoil

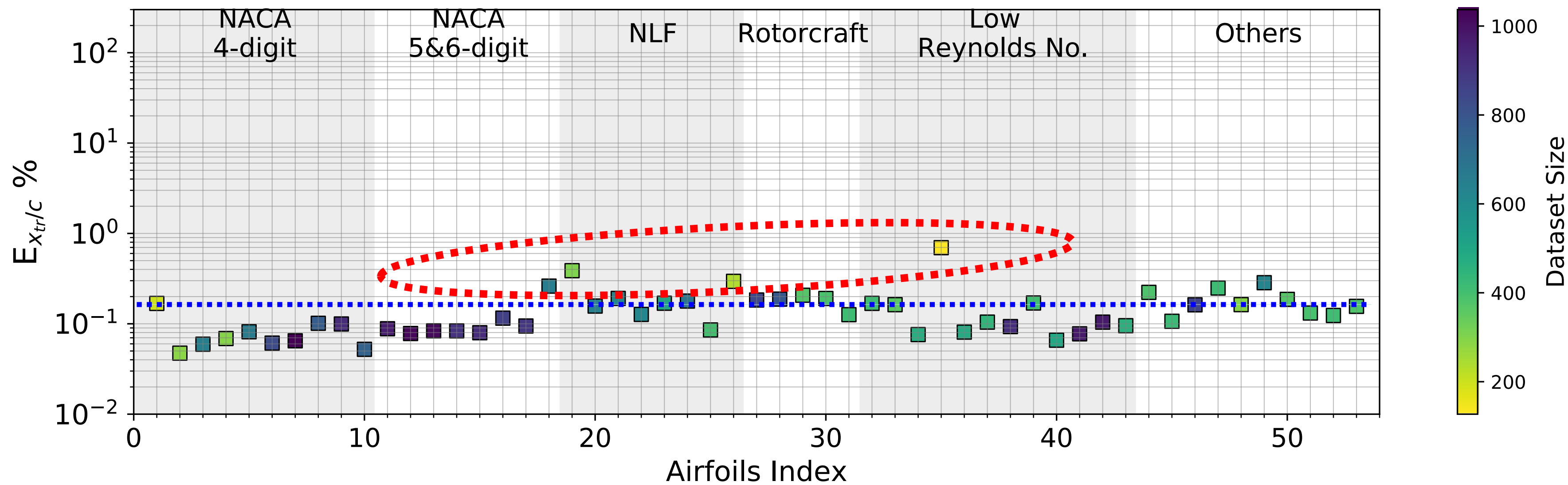
Case I

Case II

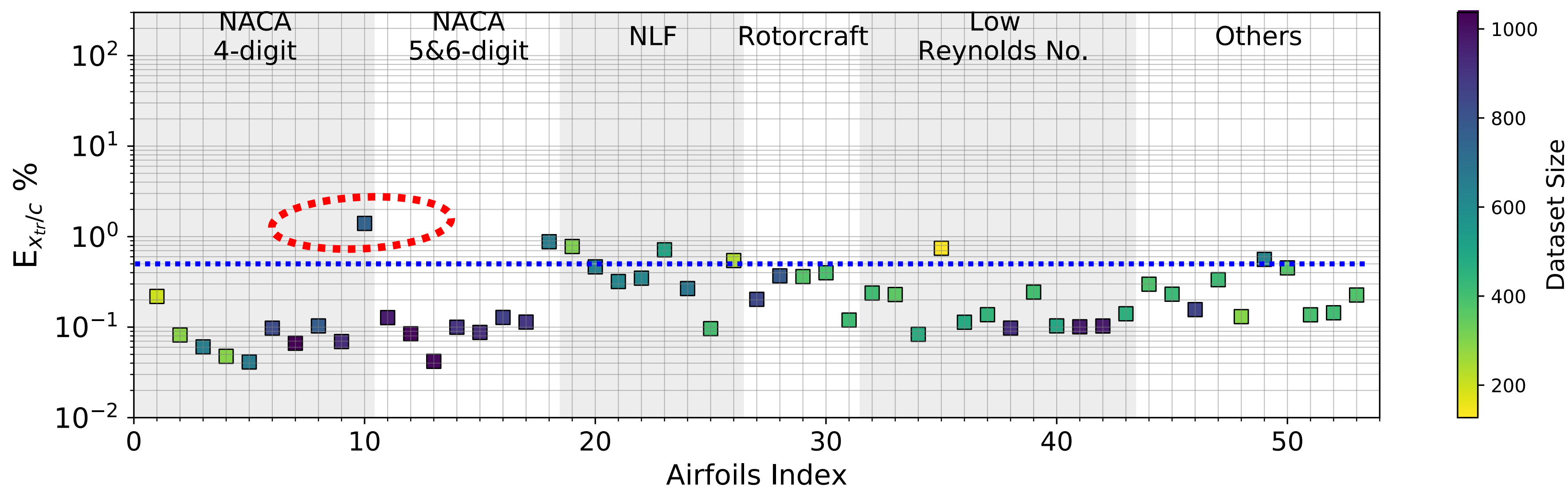


Absolute Error in Predicted Transition Location

Case III



Case IV



Uncertainty Quantification

- **Regular NN models make “confident” predictions even when highly incorrect**
- **Gaussian Process Regression (GPR) models provide built-in UQ via PDFs of predicted variables**
  - UQ based on proximity of test point to training data
  - O(N^3) computational complexity

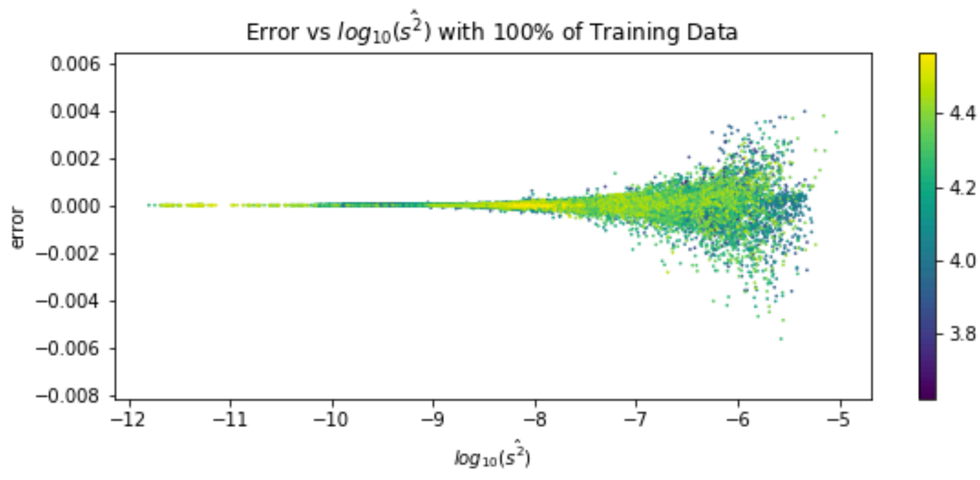
    ⇒ GPR does not scale well for large data and large number of variables
- **Case-by-case evaluation essential**

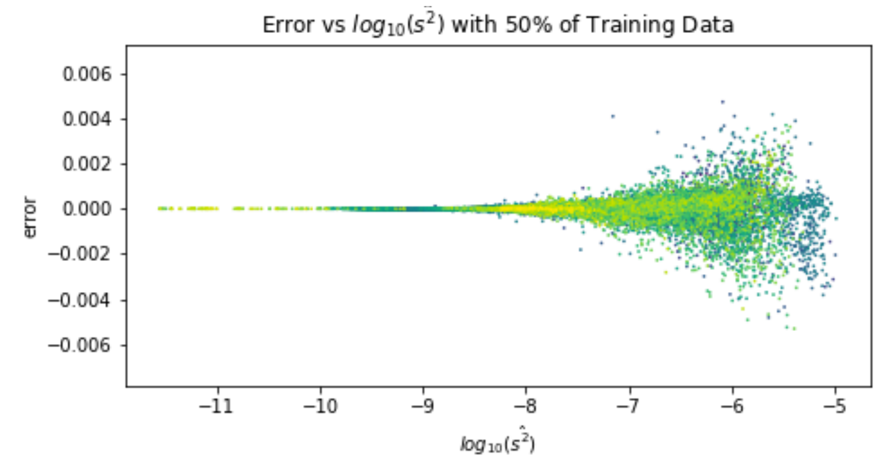
Local Approximate Gaussian Processes (laGP) Gramacy 2020

- laGP finds "local ad hoc kriging neighborhood" of size n, which is used to fit a GP
  - Neighborhood size n is the most important model parameter
  - Cubic complexity now applies in terms of n
- Uses a greedy algorithm to determine the local kriging neighborhoods
  - Neighborhood size has a trade-off between accuracy, efficiency, and susceptibility to overfitting
  - Inclusion of a new point within the neighborhood is determined not only by its proximity to test point but also how potentially different information it brings in comparison with what is already in the neighborhood
- **Output: Expected mean prediction and variance**
  - **Physical interpretation of variance is not fully clear**
  - **Performed out-of-distribution testing to gain insights**
    - ➢ Train laGP on the Falkner Skan database and predict the growth rates for a given set of airfoils
    - ➢ Monte Carlo sample from laGP's predictive distributions to generate an interval estimate for the N-factor envelope.

Findings from laGP application to Falkner Skan and High Mach Boundary Layers (cont.)

- High estimated variance (i.e., relative uncertainty) in laGP prediction correlates with high observed absolute error and vice versa

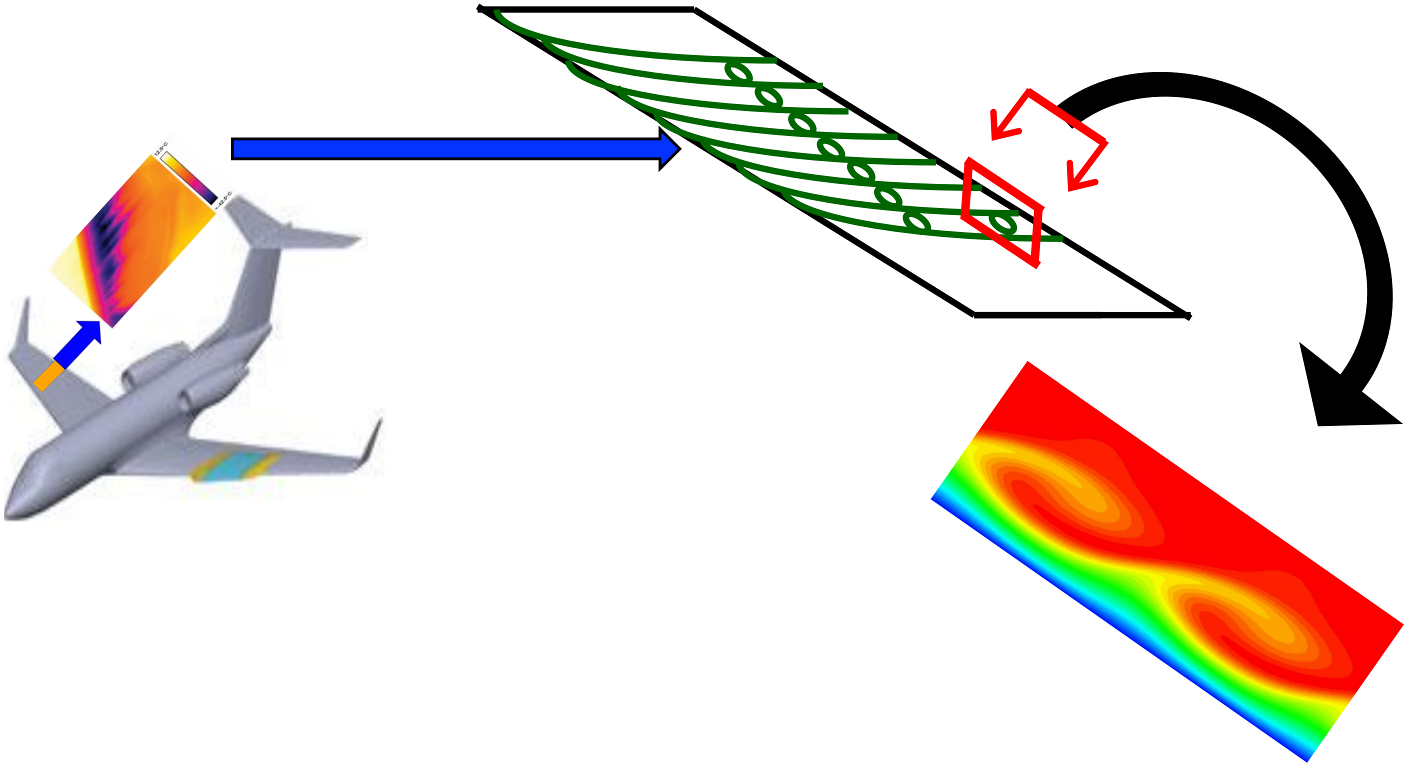




- Increasing neighborhood size from 50 to 100 more beneficial than additional increase from 100 to 150.
- laGP knows when it is accurate and when it may be inaccurate. This could help identify redundant data within a large database.

Reliable Prediction of Crossflow Transition Requires a Holistic Approach

Stationary crossflow vortices
in swept wing boundary layer



Crossflow transition =

Receptivity/Linear growth

+

<u>Prolonged region of nonlinear growth</u>

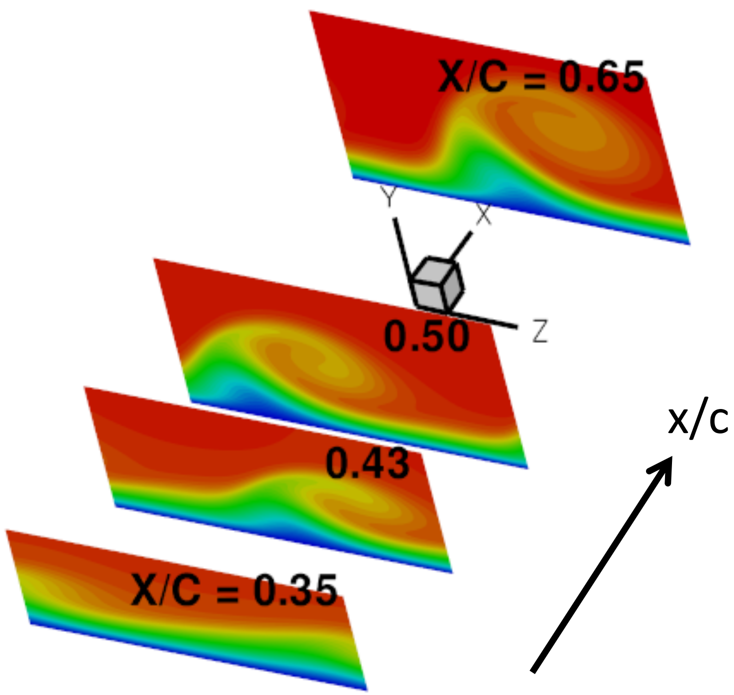
+

.
.
.

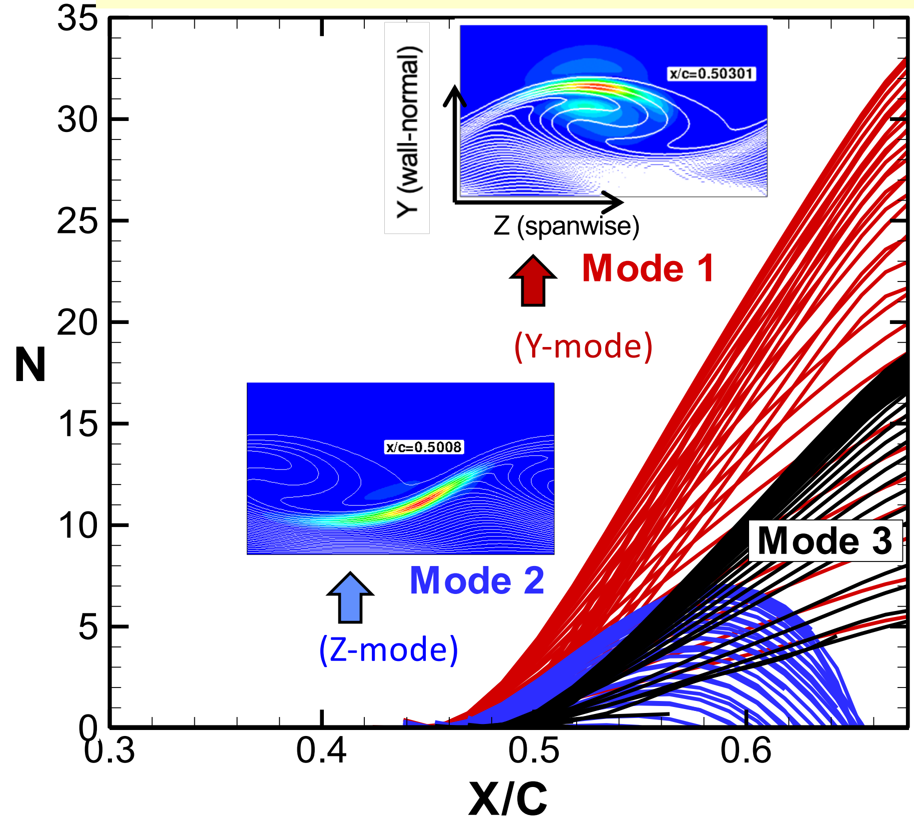
N-factor correlations based on secondary instabilities yield a More robust transition criterion

Secondary Instability of Stationary Crossflow Vortices

Primary Instability: Stationary Crossflow

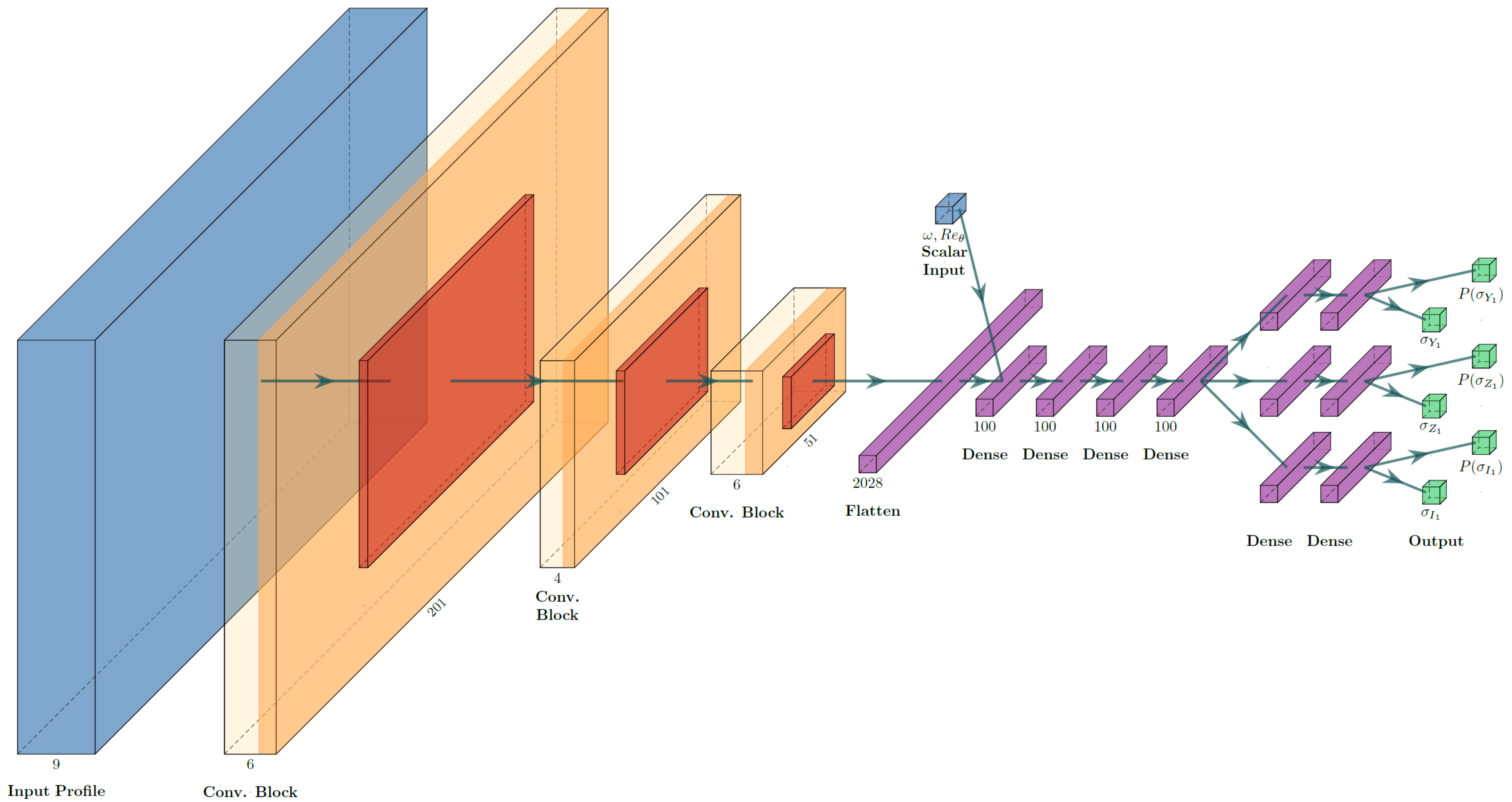


Secondary Instability (Multiple Unstable Modes)



- Input profiles are now planar distributions instead of line profiles
- A variable number of multiple unstable modes must be modeled
- Primary distinguishing feature of different modes ≡ mode shape
  - Desirable to predict both local growth rate and associated mode shape

Multi-mode Prediction with Mode Shape



Network learns shared convolutional and fully-connected latent features from flow profile and scalar inputs.

Network branches into separate fully connected networks that use the learned latent features as input to predict the probability for the existence of each mode and the associated growth rate.

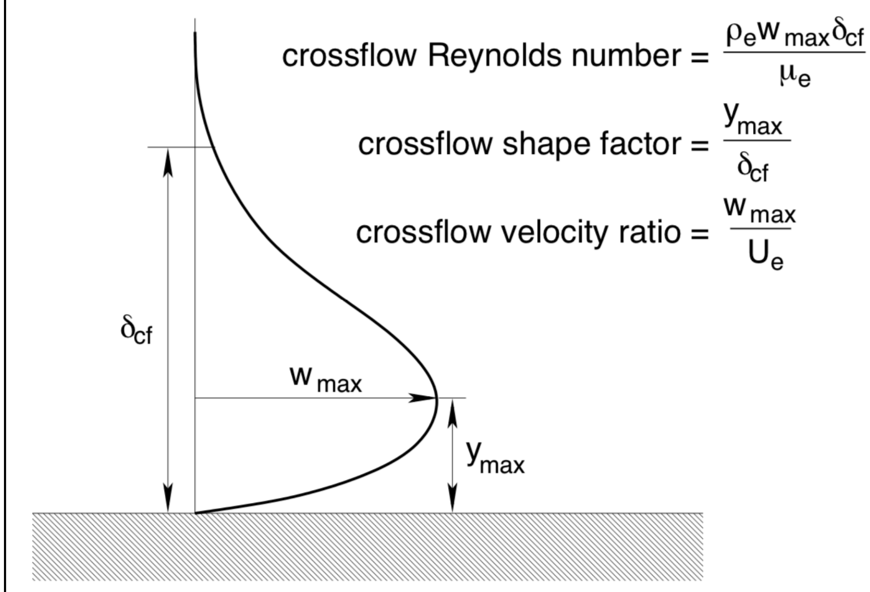
Summary

- Deep learning offers an attractive option to develop transition models based on instability amplification
  - CNN architecture allows the high-dimensional input space to be truncated to a small dimension
- Possible to develop models that work for multiple instability mechanisms and a broad class of boundary layers
  - Need large set of training data, but limited success with selection of
- For T-S instabilities, with prior knowledge, it was possible to interpret the primary latent feature identified by the CNN
  - However, still looking for ways to extend the model interpretation to other instability mechanisms
- Future work
  - Multi-mode prediction
  - Integration into CFD codes
  - Physics Informed Neural Networks (Karniadakis et al.)?

Extra Charts

Database for Stationary Crossflow Instability
Falkner-Skan-Cooke Selfsimilar Boundary Layers

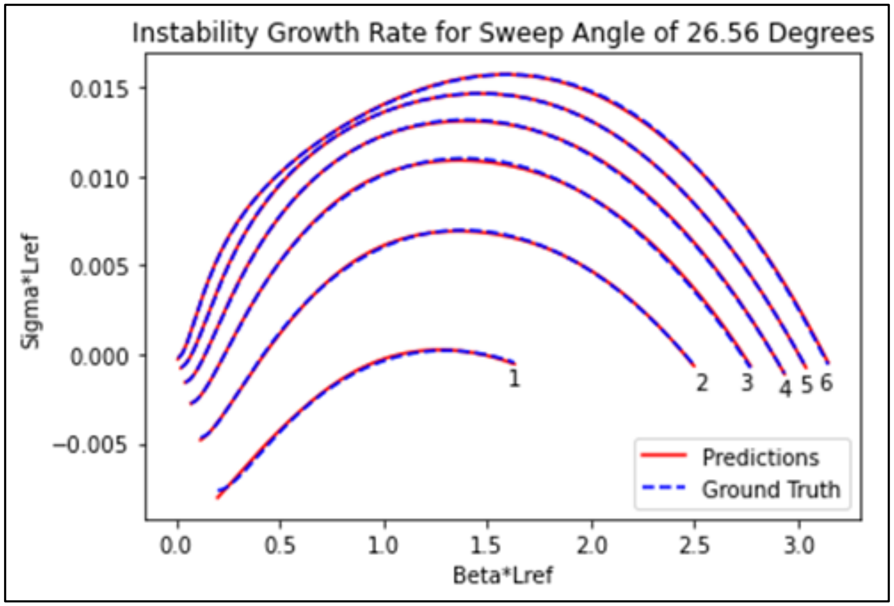
- 40 self-similar boundary layer profiles and 12,000 eigenvalue computations based on LST
- 37 boundary layer profiles were used for training and 3 were reserved for validation
- The crossflow shape factors, H_{cf} , encountered in more realistic flows exceed the range included in the FSC database (0.30 – 0.38)

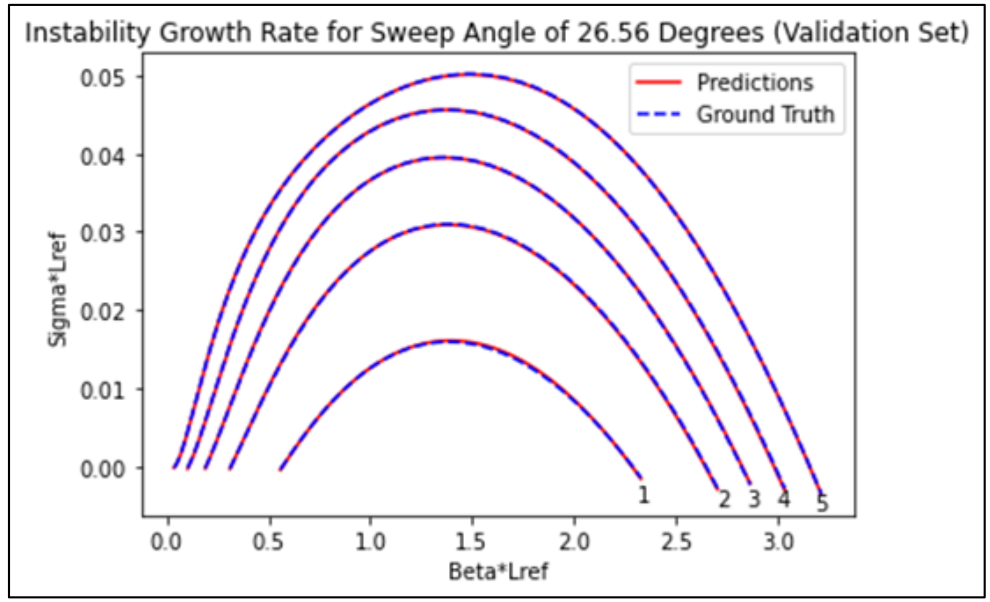


Crossflow velocity profile

Rajnarayan and Sturdza (2011)

Stationary Crossflow Instability
Falkner-Skan-Cooke Selfsimilar Boundary Layers





Summary of FSC Neural Network Models

Model	Input Profiles	Input Parameters	Sigma Validation Error (%)
Orr-Sommerfeld Eq.	U/q_e , W/q_e	β_θ, Re_θ, H_{CF} , Λ, $u_{cf\ max}/q_e$	0.637
Rajnarayan & Sturdza	W_s/q_e	β*Lref, Re_{Lref}, Λ, $R_{cf\ 0}$	0.312
Crouch et al.	U_s/q_e , W_s/q_e	β_θ, Re_θ, Λ, $u_{cf\ max}/q_e$	0.405

Stability Database for Swept Wing Boundary Layers "Big data"

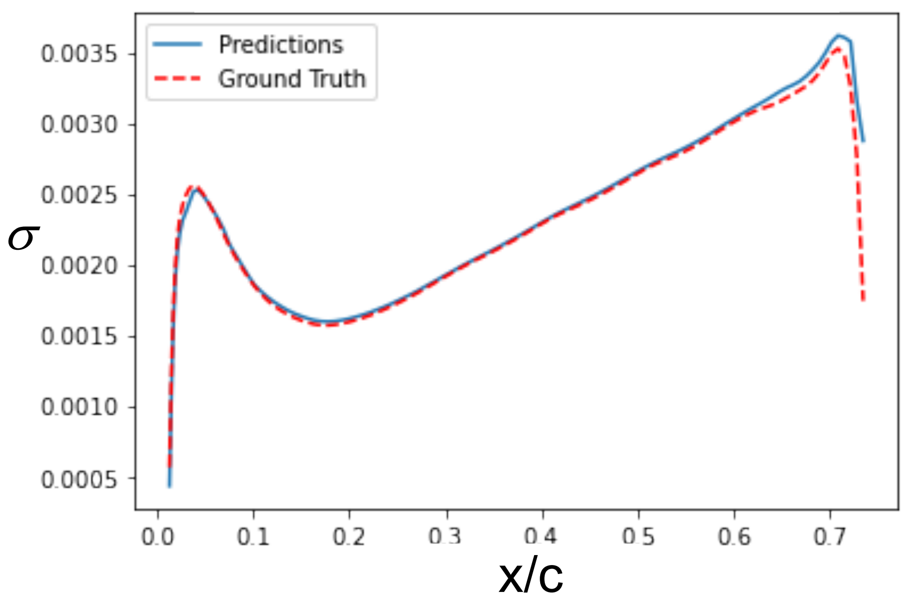
- **106,584 boundary layers over swept airfoils (stationary crossflow instability)**
- **Preliminary Data Sampling Technique for Training on NLF(2)-0415**

Re_c : [	2.5e4	3.5e4	5.0e4	7.0e4	…	1.024e8	1.44e8	2.048e8]
AoA : [	-6.0°	-5.0	-4.5°	-4.0°	…	8.0°	9.0°	10.0°]

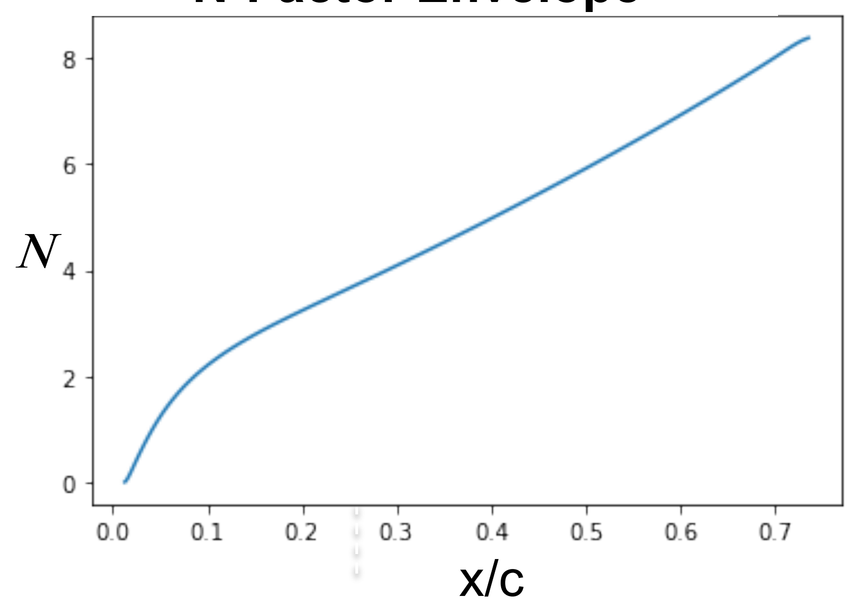
- Neural networks achieved 3.37% error with 1/3 of the crossflow cases, selected randomly from the sampled data

NLF(2)-0415 Airfoil, Re_c = 3.2e6, AoA = -3.5 deg, 40-deg sweep

Growth Rate Variation



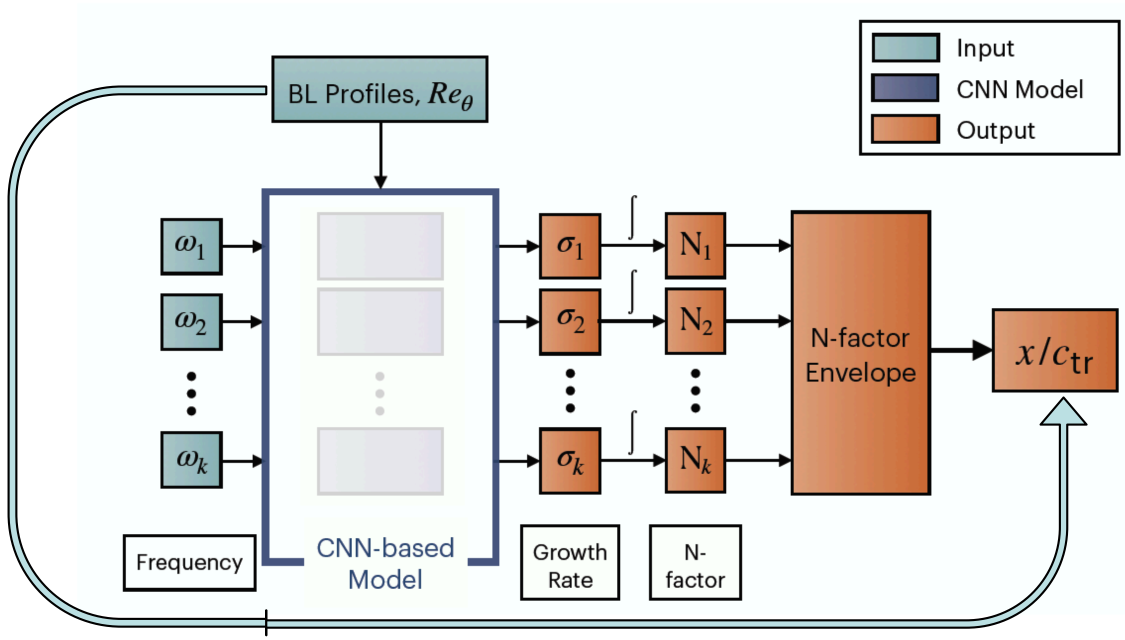
N-Factor Envelope



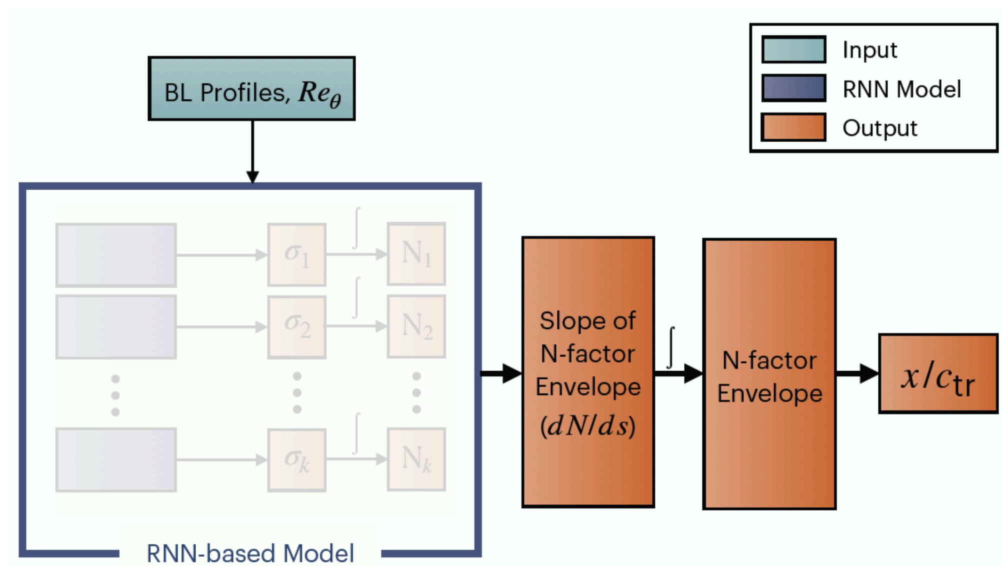
Hausknecht et al., 2021

Workflow Simplification via RNN

Previous NN (Zafar et al. 2020)



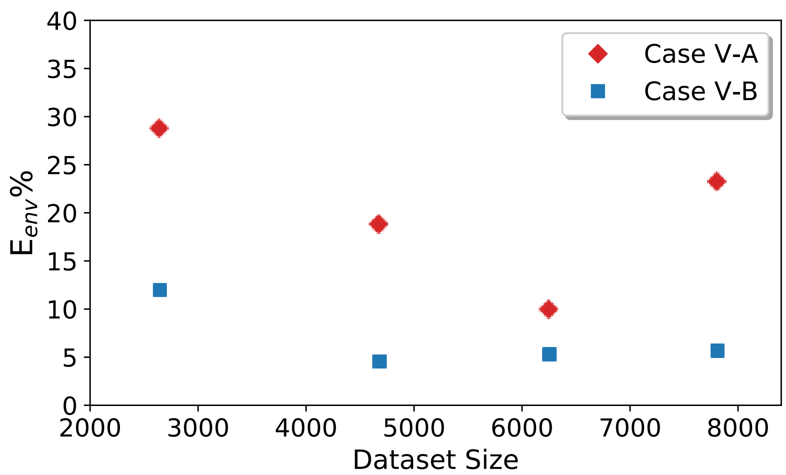
RNN (Zafar et al. 2021)

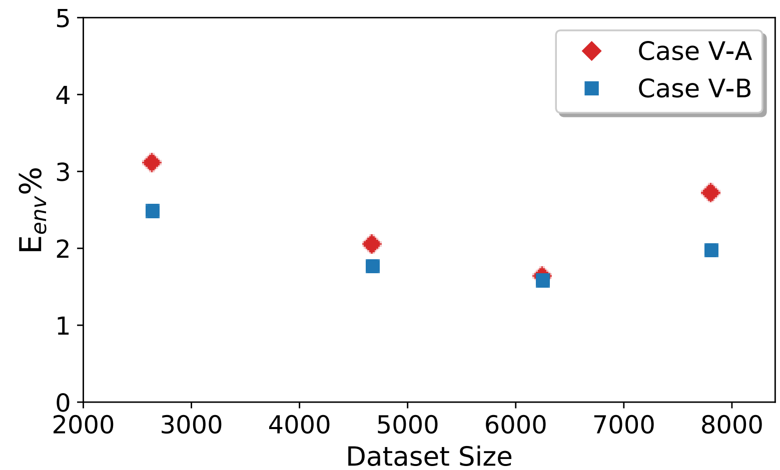


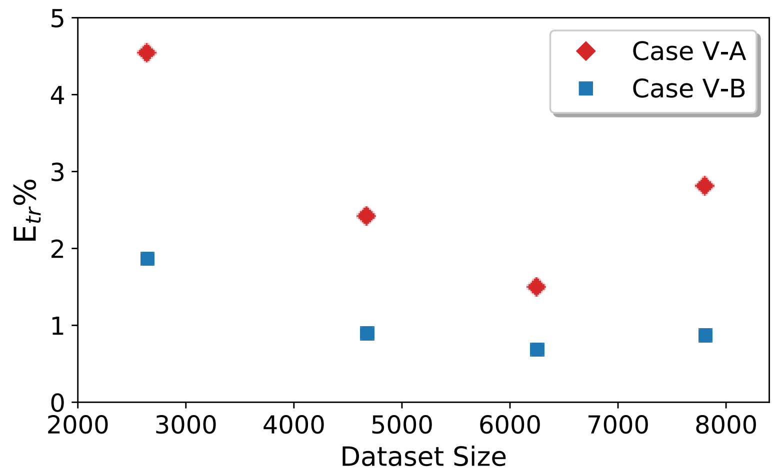
Case V

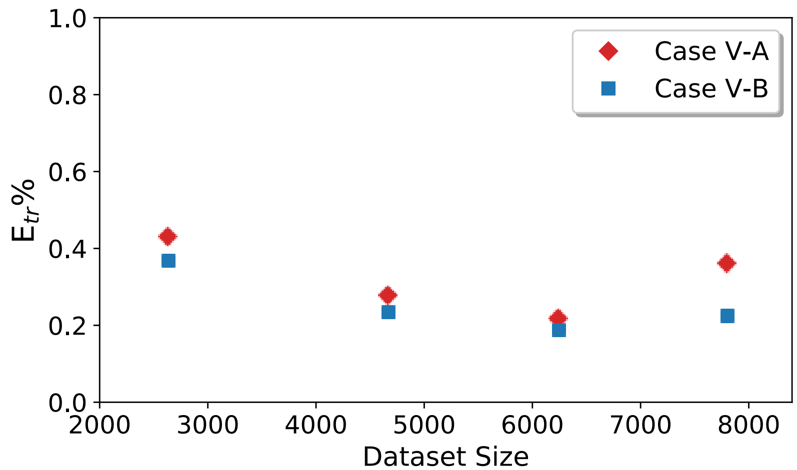
Max Average Error for a single airfoil

Average Error of whole database



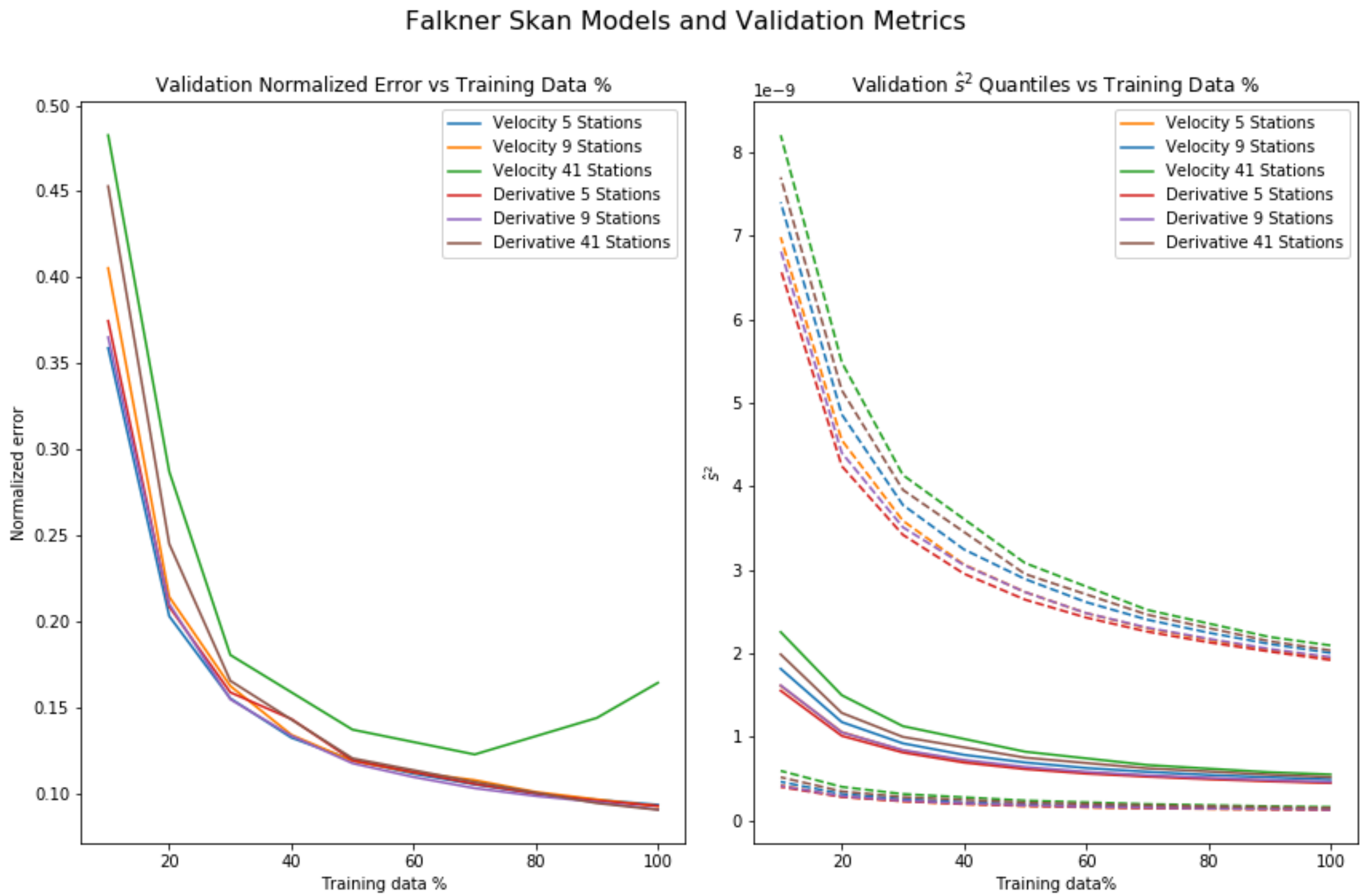




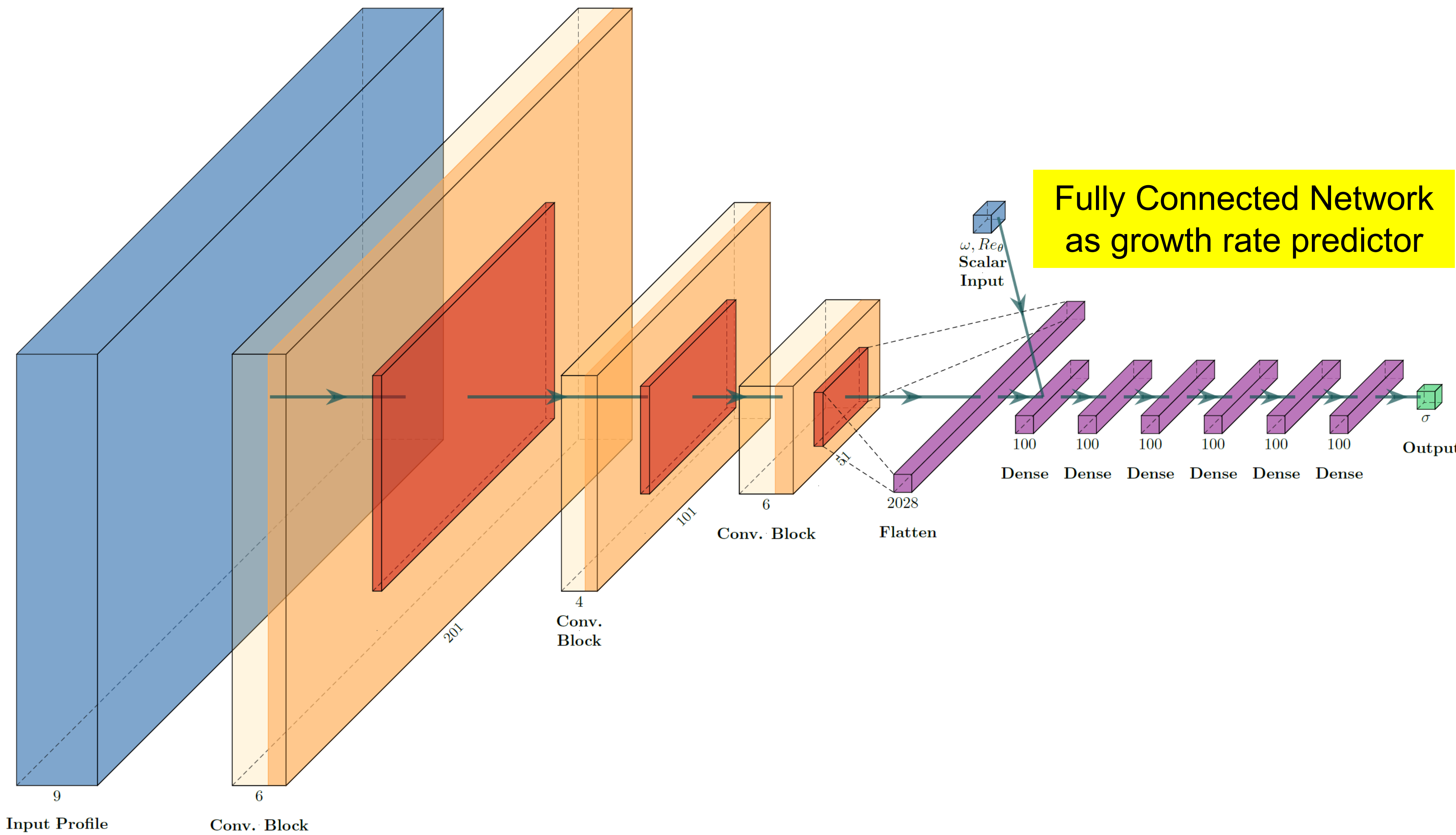


Findings from laGP application to Falkner Skan and High Mach Boundary Layers

- Increase in training data ⇒ progressive decrease in normalized error and estimated uncertainty



Single Mode Growth Rate Prediction



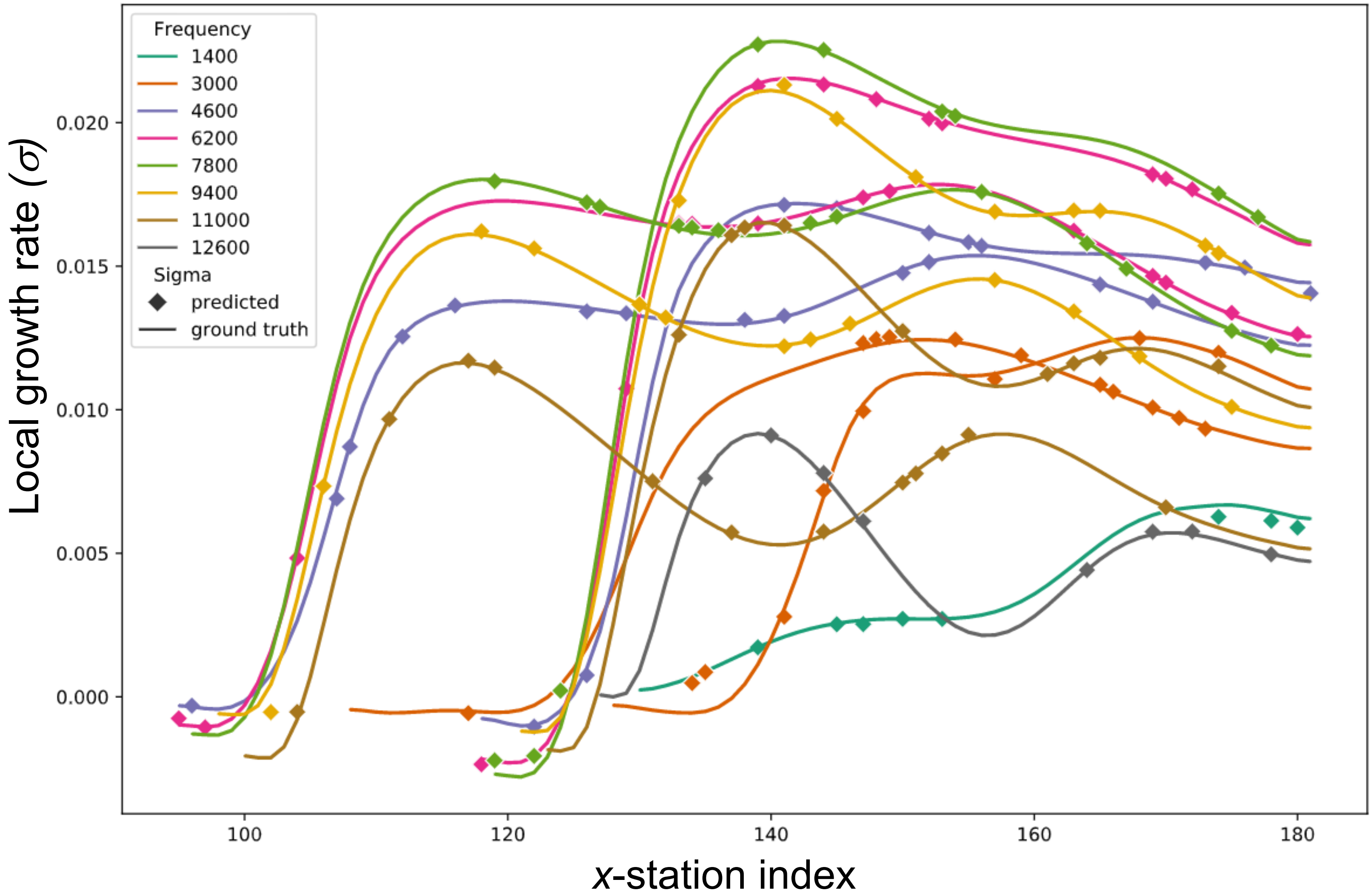
Fully Connected Network as growth rate predictor

CNN with planar (2D) basic state ”profile”

Validating CNN Model for Crossflow Transition

NLF(2)-0415 Airfoil ($Re_c = 2.4 \times 10^6$, $\lambda = 12$ mm)

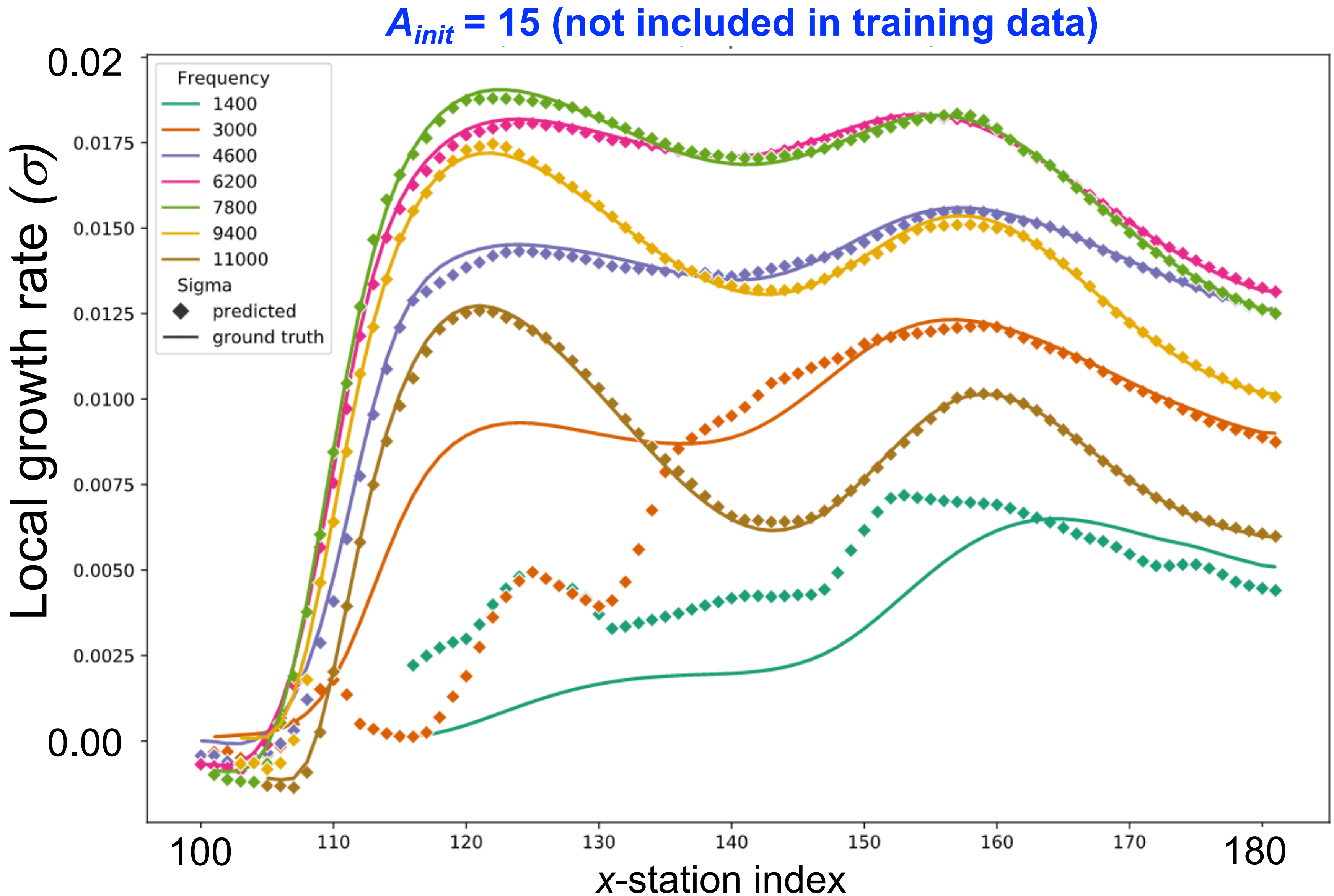
Within-distribution test error (A_{init} = 5, 10, 20)



- **Excellent performance on within-distribution test cases**

Validating CNN Model at Other Crossflow Amplitudes

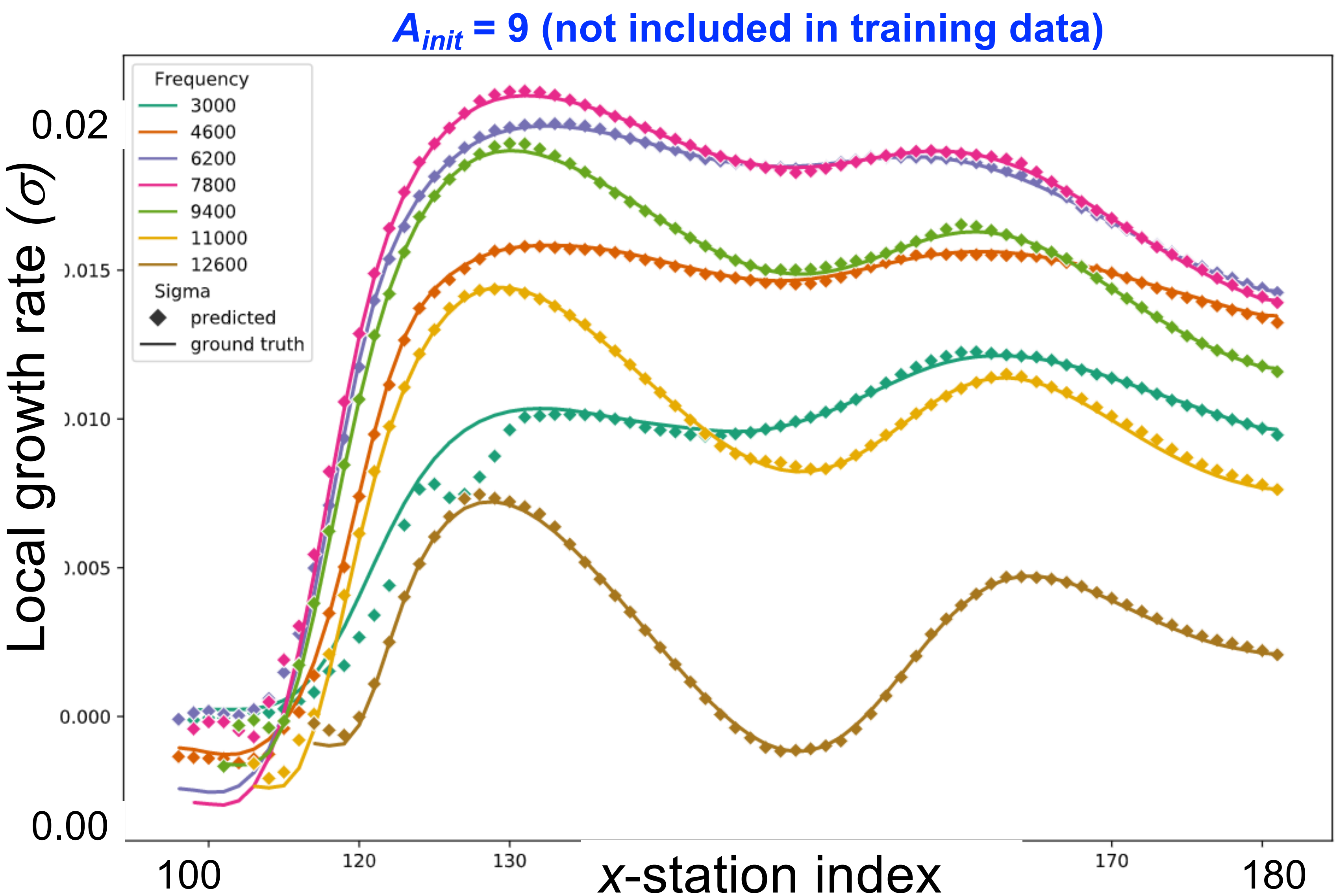
A_{init} = 15



- **Extrapolation to other crossflow vortex amplitudes shows visible discrepancies, but only for subdominant modes**

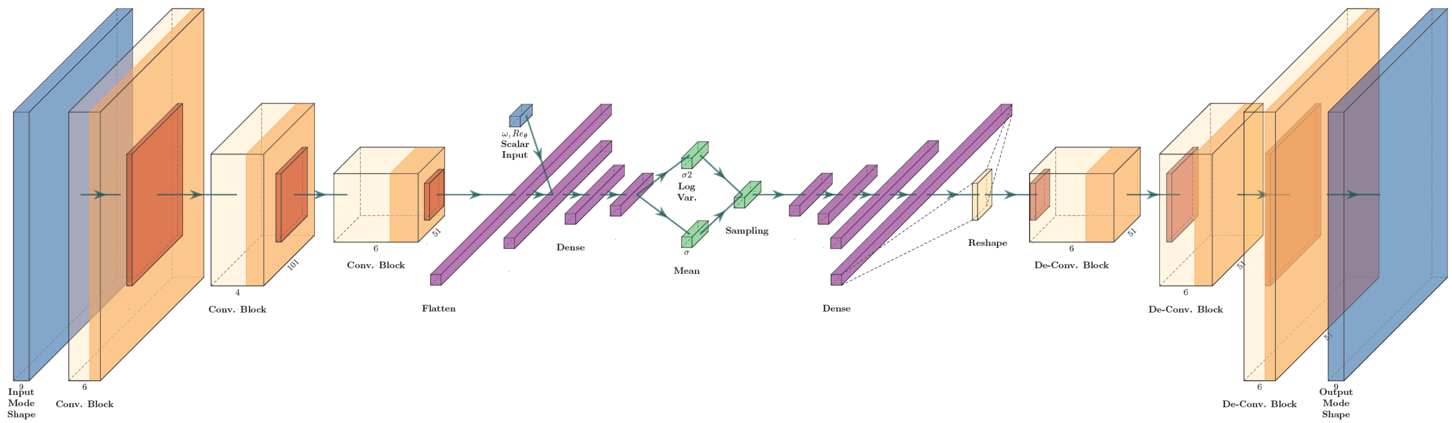
Validating CNN Model at Other Crossflow Amplitudes

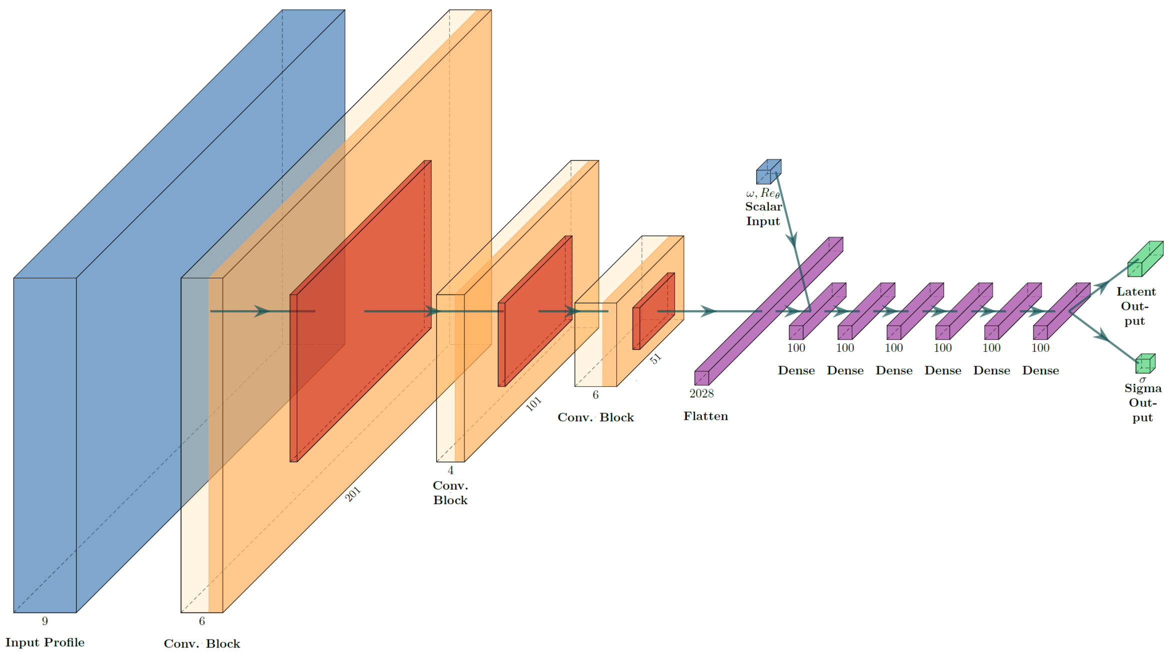
$A_{init} = 9$



Single Mode Prediction with Mode Shape

1. Unsupervised pretraining of variational autoencoder

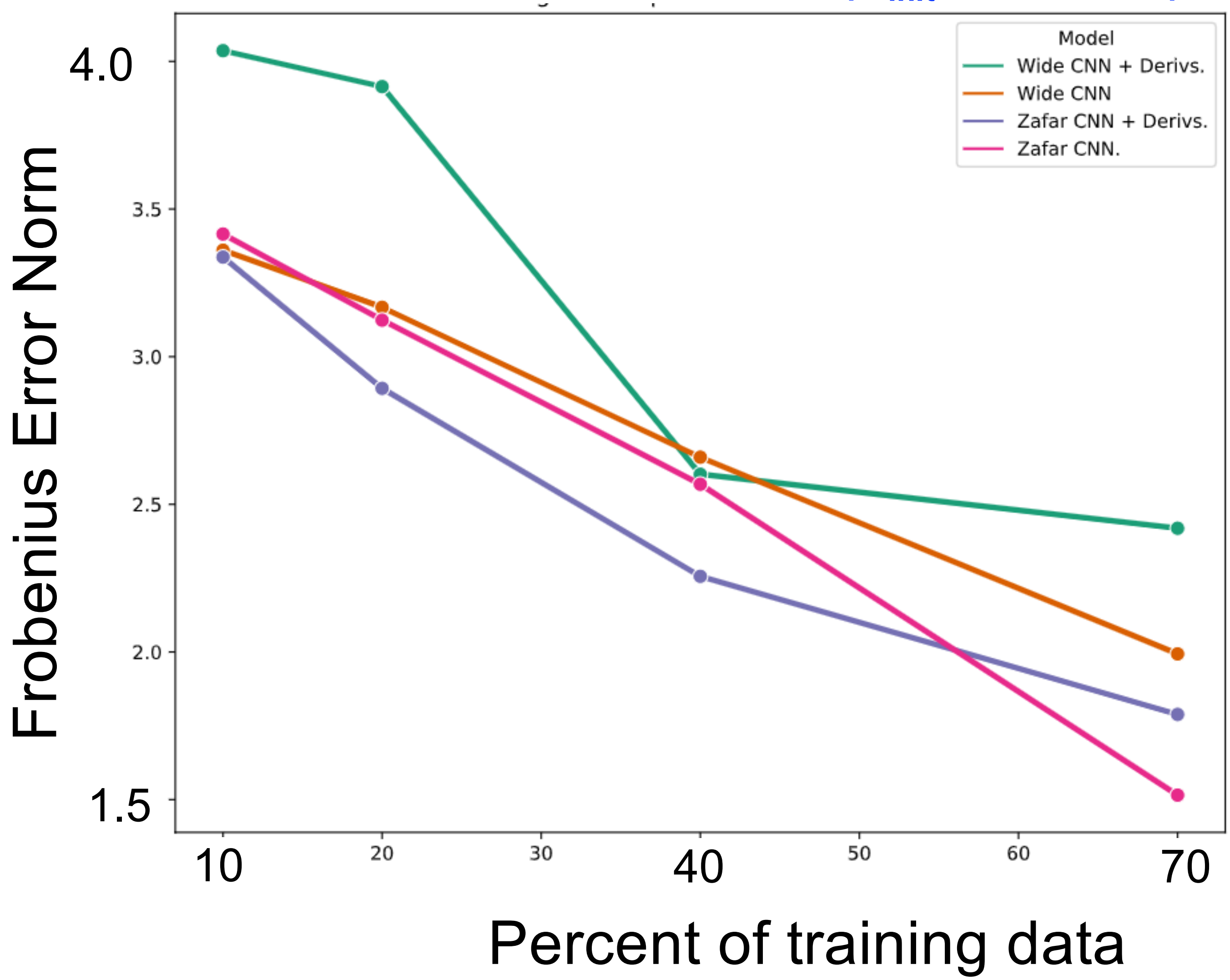




2. CNN prediction of growth rate + latent space representation of mode shape

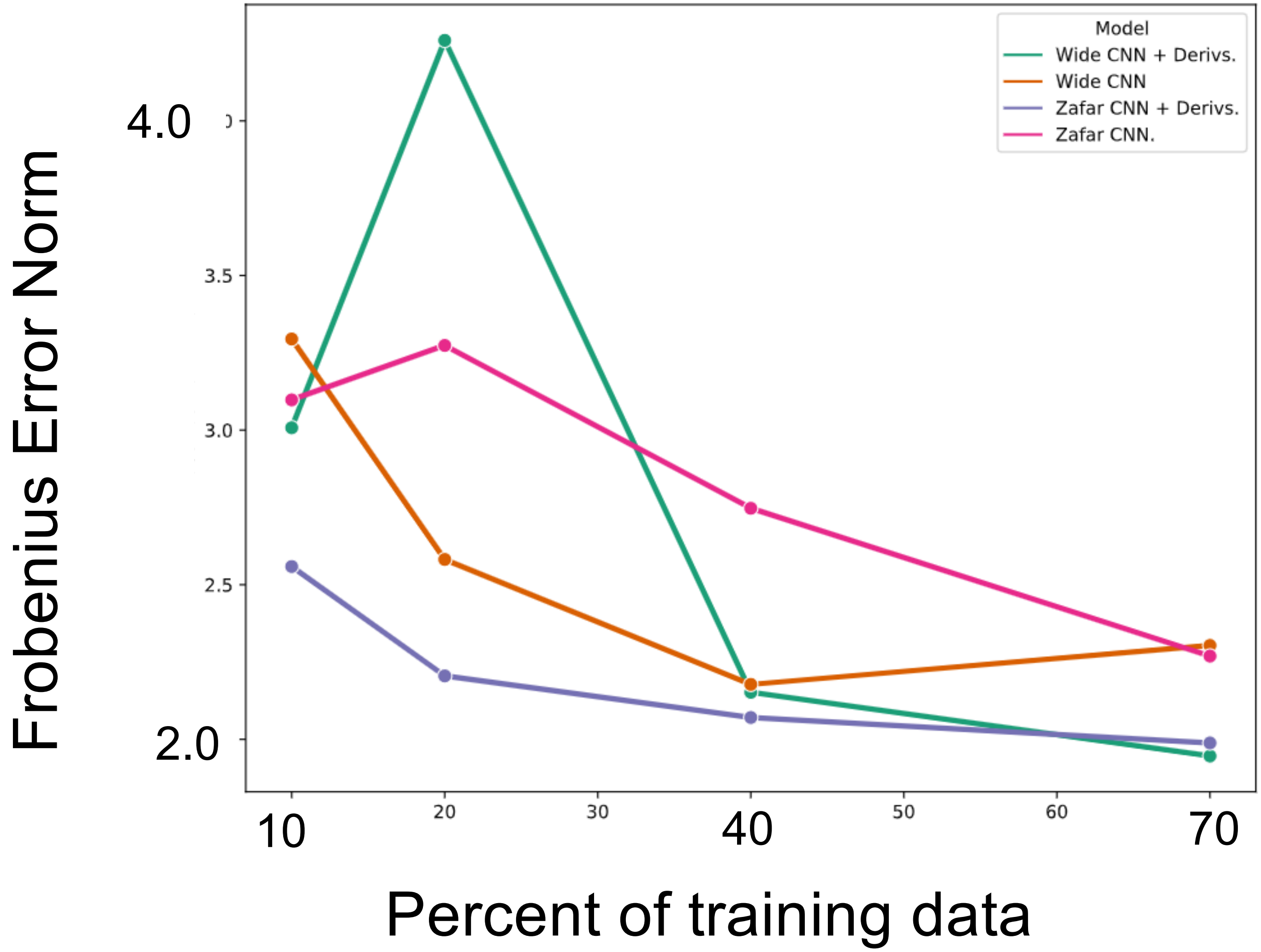
Performance of Candidate Architectures Test Accuracy vs. Volume of Training Data

Within-distribution test error (A_{init} = 5, 10, 20)



Performance of Candidate Architectures
Test Accuracy vs. Volume of Training Data

A_{init} = 15 (not included in training data)



Performance of Candidate Architectures Test Accuracy vs. Volume of Training Data