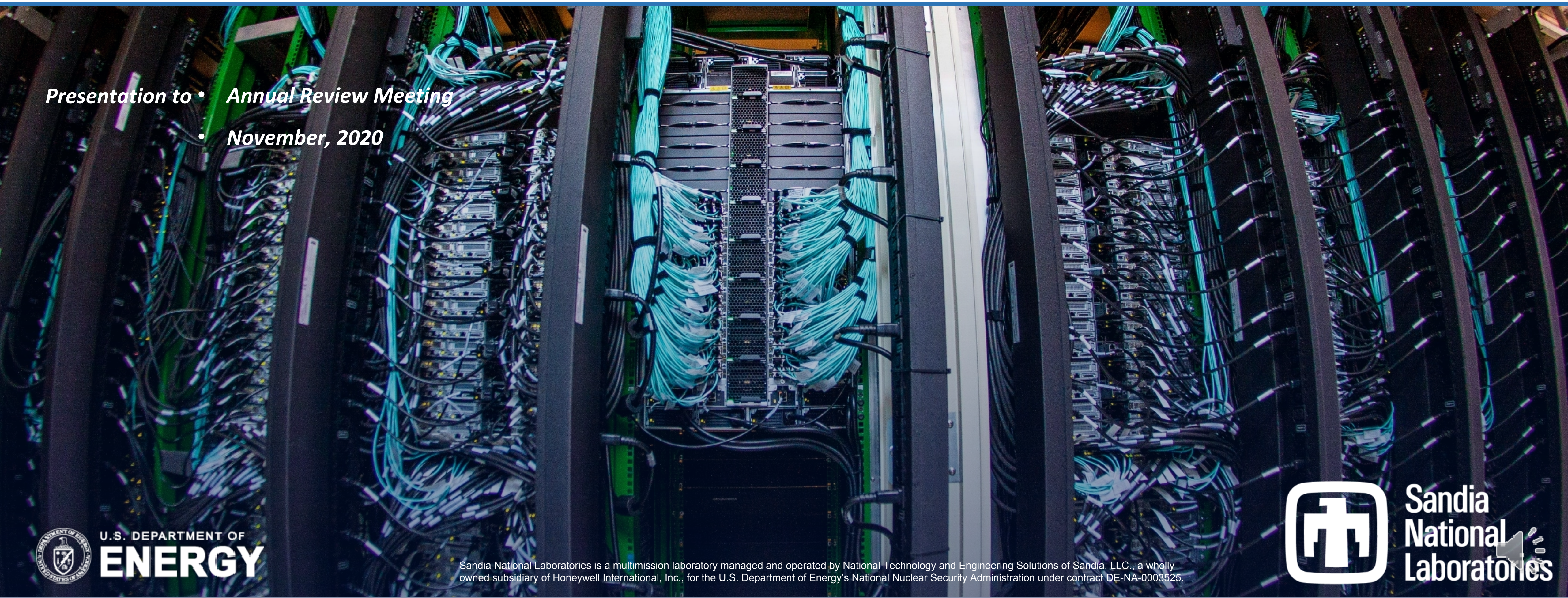
SAND2020-11801PE

Sensitivity and hyperparameter optimization for CNN-LSTM based architectures for CO_2 flow prediction

Science-informed Machine Learning to Accelerate Real Time (SMART) Decisions in Subsurface Applications

- *Joe Hogge and Hongkyu Yoon*

Sandia National Laboratories



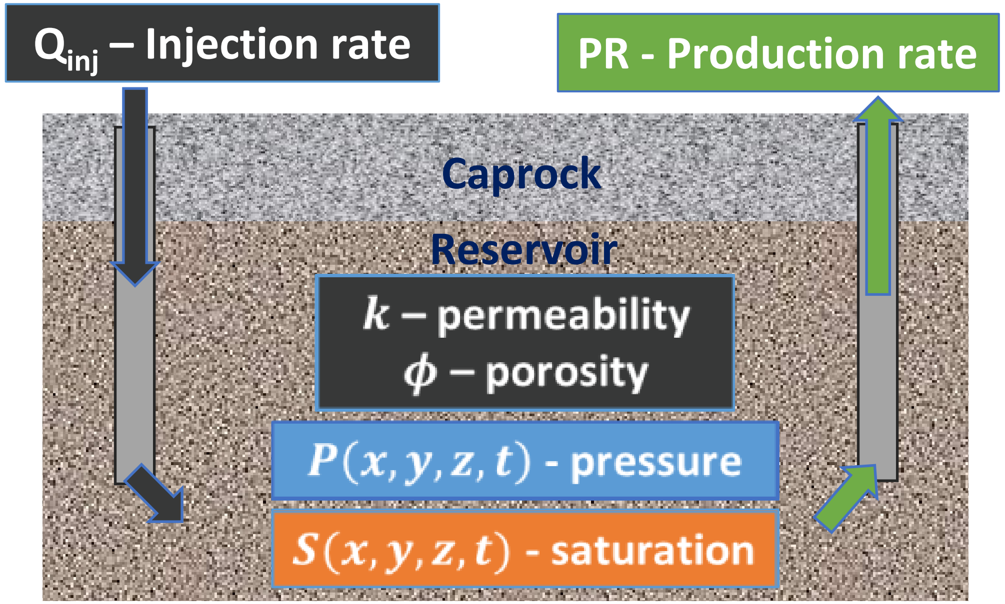
Presentation to
- ***Annual Review Meeting***
- ***November, 2020***

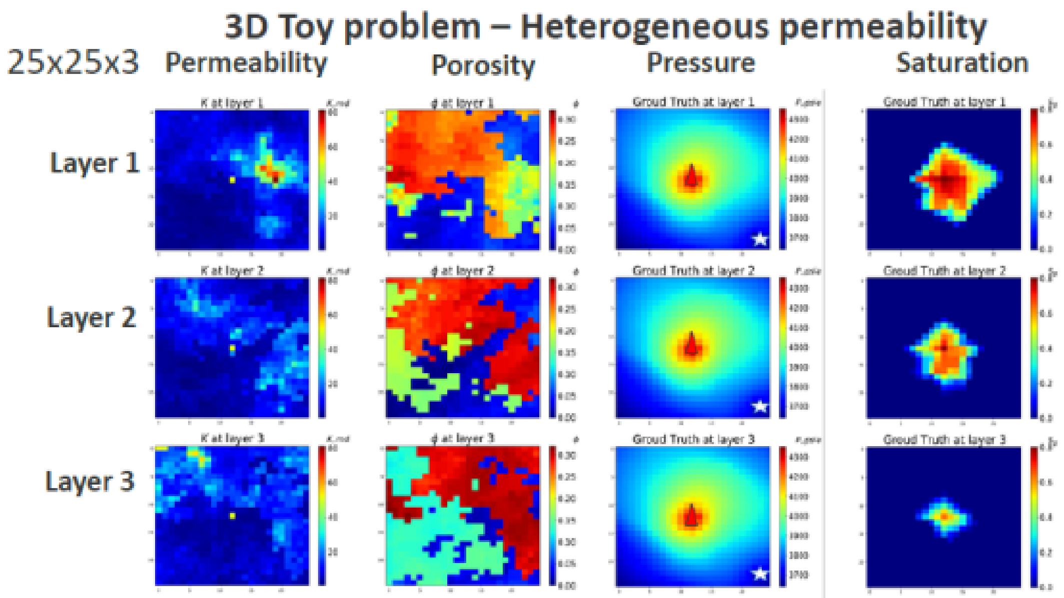
U.S. DEPARTMENT OF ENERGY

Sandia National Laboratories is 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-NA-0003525.

Model Setup: 3D Toy Problem [Task 5]

Sandia National Laboratories

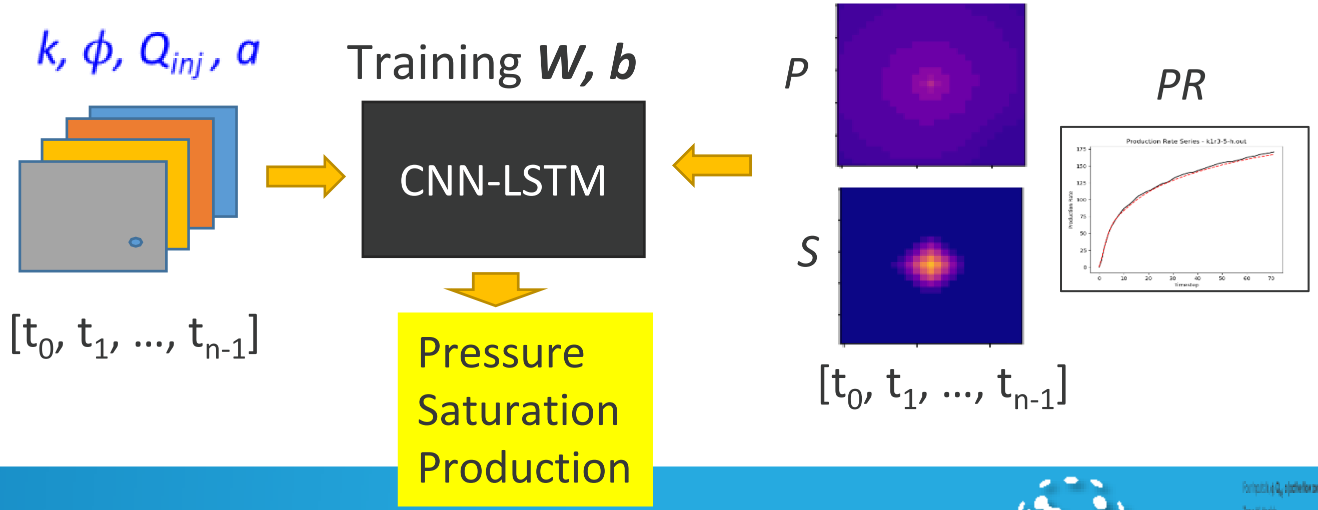




Four Inputs: **k**, **ϕ**, **Q_{inj}**, **a (active flow zone)**

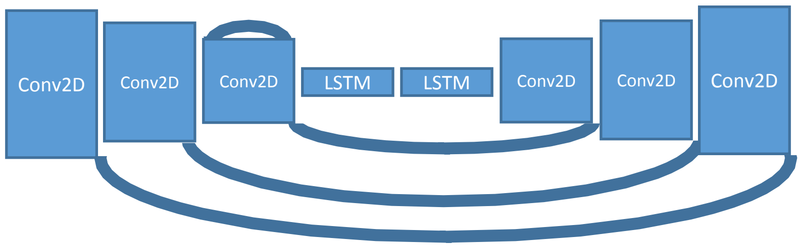
Three ML Models:

Pressure, **Saturation**, **Production rate (PR)**



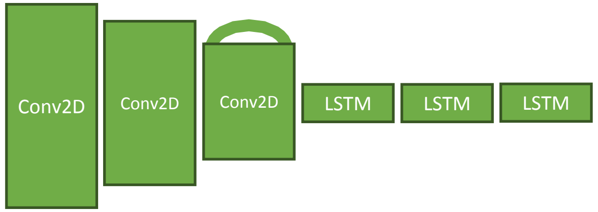
Model Architectures

- **Model for spatio-temporal data (Pressure & Saturation)**
  - Base architectures: Convolutional Neural Network (CNN)-LSTM (long short term memory)
  - Encorder-decorder with CNN-LSTM-CNN
  - "TimeDistributed" layer in Keras to reduce the number of trainable parameters by sharing weights for time-series data
- **Model for temporal data (Production rate)**
  - Base architectures: Convolutional Neural Network (CNN)-LSTM (long short term memory)
  - CNN-LSTM
  - "TimeDistributed" layer in Keras
- **Key hyperparameters**
  - The number of CNN layers (3 or 4)
  - The number of filters, filter size, stride, activation function, etc
  - The number of stacked LSTM units (0-3)
  - The size of hidden units in the LSTM layer (1-32)
  - Dense layer before the LSTM unit
  - Skip connections (easy to mitigate the vanishing gradient issue)
  - Loss function of the binary map of Saturation (binary crossentropy) to capture sharp interface of Saturation profile



(example) Total parameters: **103,302** ; Trainable params: 103,302

Loss = Mean Squared Error
+ λ * Binary Crossentropy
(λ: weight)

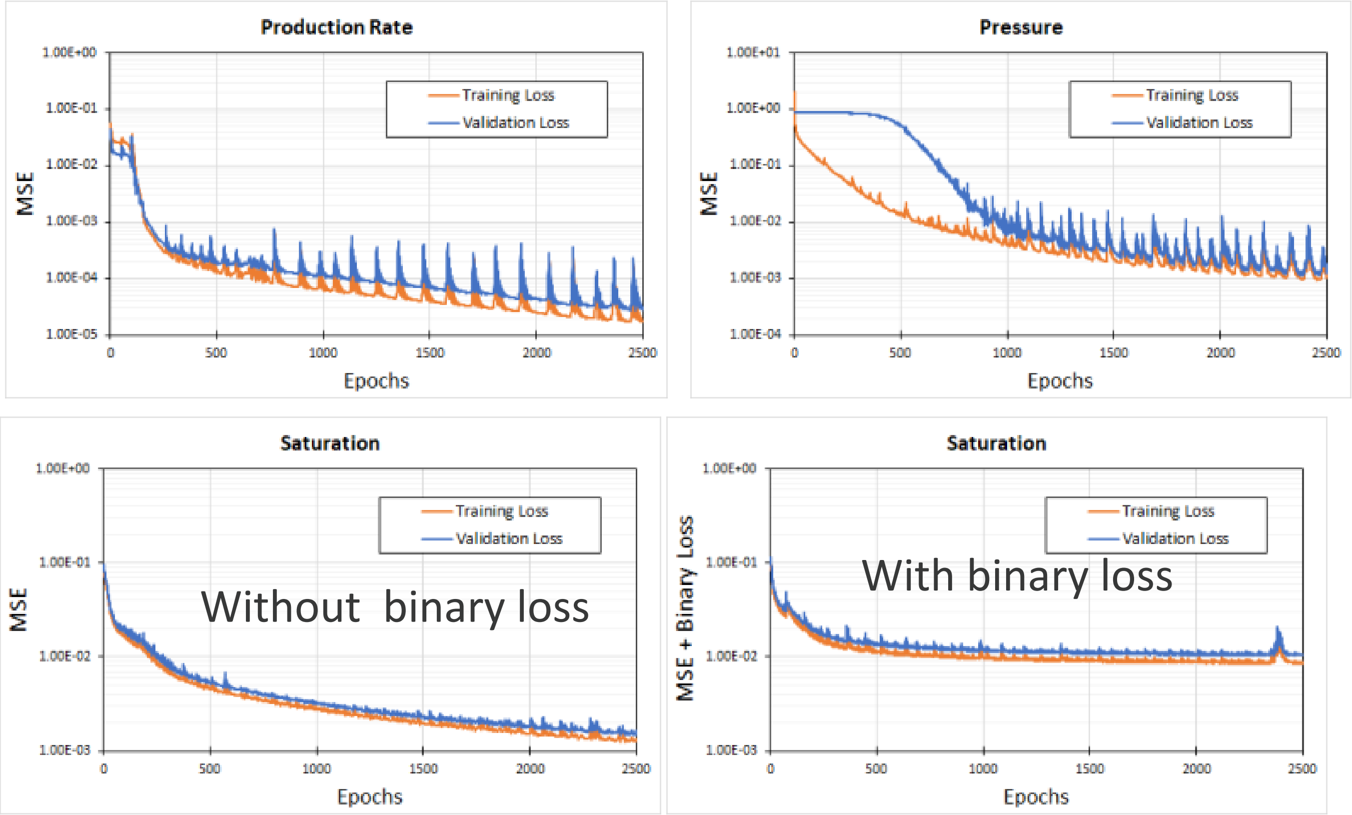


(example) Total params: **47,445** ; Trainable params: 47,445

Loss = Mean Squared Error

Results: Training, Validation, Testing

- Training (19 simulation data)
  - Tesla V100-SXM3-32GB GPU (Linux)
  - Quadro P4000 Windows Desktop
  - 15-20min training time for each target quantity (P, S, PR)
  - Actual training time: 15-25ms/epoch, so mostly data loading and saving time
- Validation & Testing (5 & 3 simulation data)
  - Trained model selected based on validation loss (validation data was not used during training)



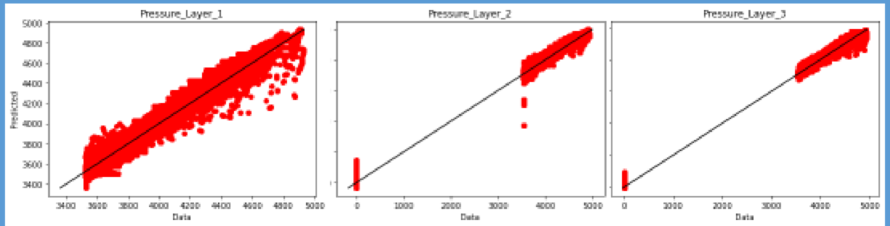
Note: The total loss of saturation with the binary loss case is higher than one without binary loss due to the additional loss.

- Binary loss increased learning efficiency (nearly flat after ~500 epochs)
- Actual performance of saturation prediction is better with binary loss

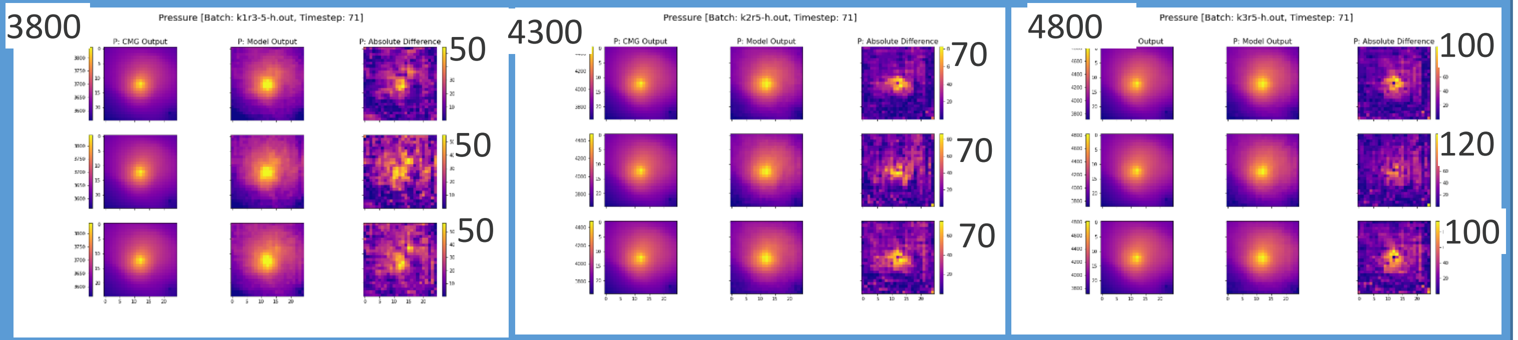
Results: Testing (3 cases)

Pressure

Layer 1 | Layer 2 | Layer 3

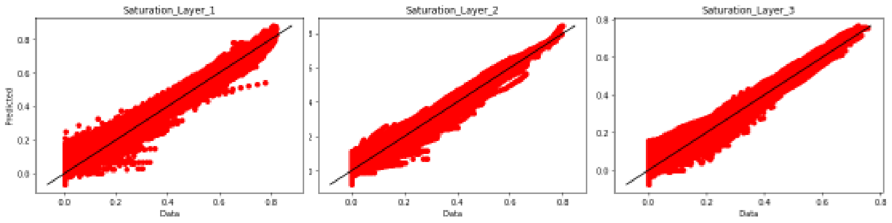


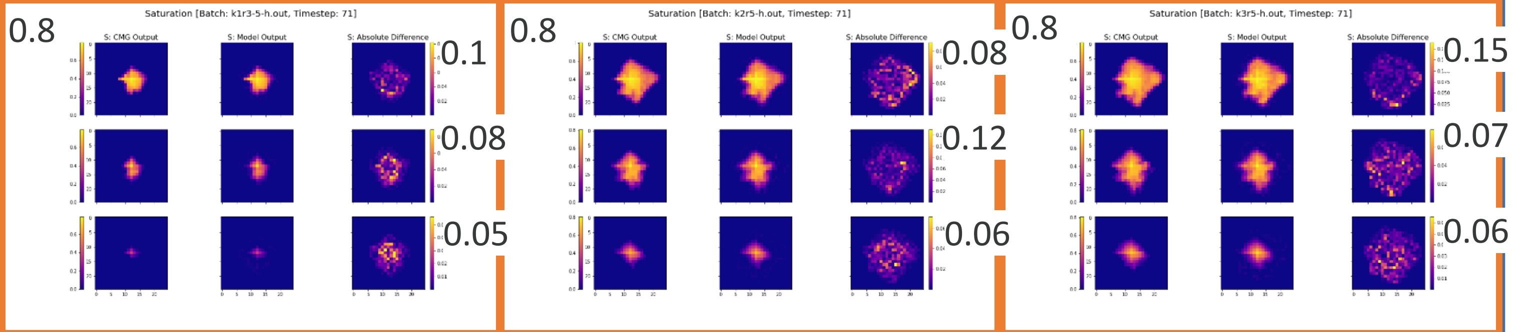
CMG Ground Truth, ML prediction, Difference



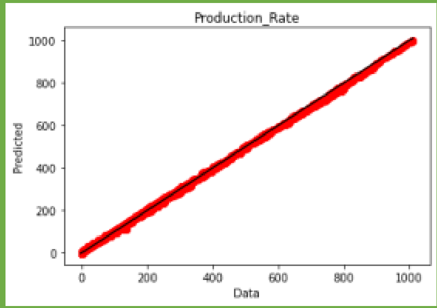
Saturation

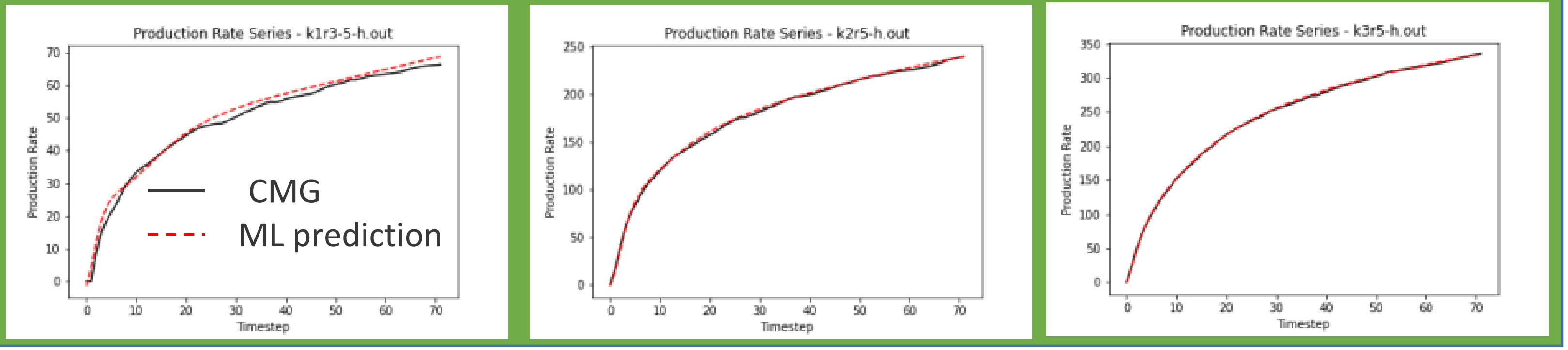
Layer 1 | Layer 2 | Layer 3





Production Rate





Future Work

Sandia National Laboratories

- **Computationally efficient CNN-LSTM architecture has been implemented**
  - Automatic hyperparameter optimization will be feasible (e.g., Sherpa or Grid-Search)
  - Scalability for large problems
- **Physics-based Loss Function**
  - Binary loss function boosted learning efficiency and improved predictions using the Saturation model
  - Flux based loss function can improve the model performance

Thank you!