# PGE 383
# Machine Learning

**Lecture outline . . .**

- **Examples of Machine Learning**
- **Conclusions**

**Michael Pyrcz, The University of Texas at Austin**

# Motivation

Inspire you with some great opportunities to apply machine learning.

We are doing this in our new consortium – DIRECT.

**DI**gital **RE**servoir **C**haracterization **T**echnology

Contact Michael for more information about the consortium and Leilani for information about joining.


Michael Pyrcz

mpyrcz@Austin.utexas.edu

Leilani Swafford

Leilani.Swafford@austin.utexas.edu

# PGE 383
# Machine Learning

**Lecture outline . . .**

- **Examples of Machine Learning**

**Michael Pyrcz, The University of Texas at Austin**

# Examples of Machine Learning

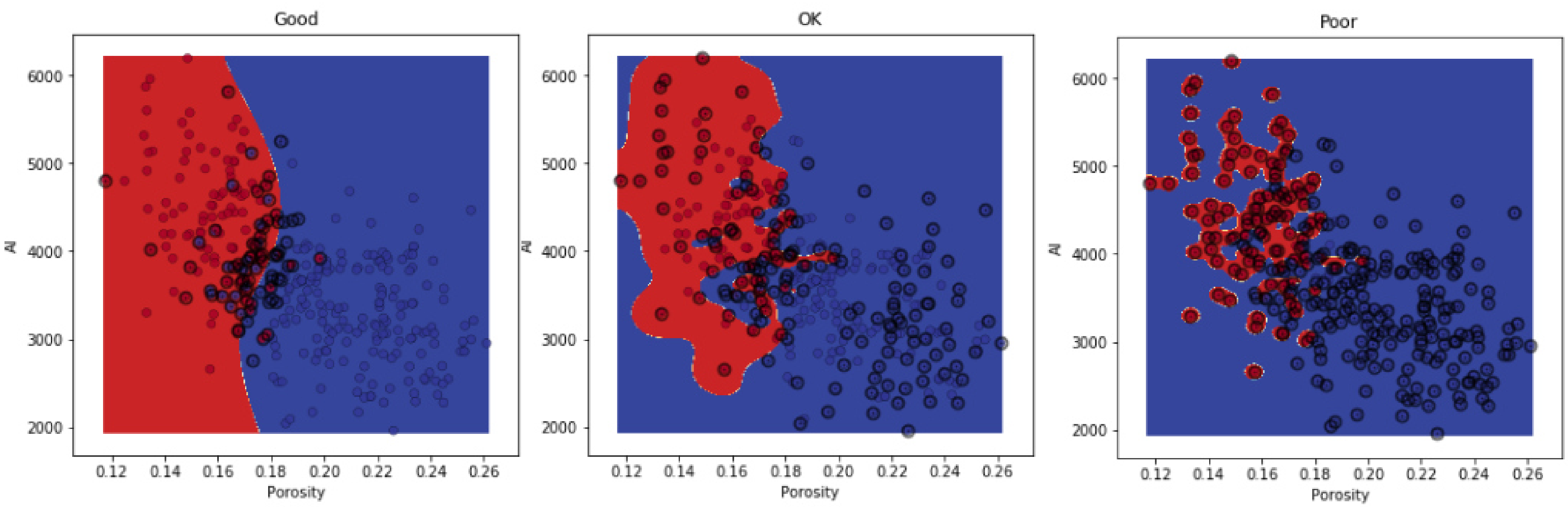
Provides a set of examples with machine learning to address subsurface challenges.

In the next couple of lectures we will cover:

- Dimensionality Reduction
- Tree-based Methods
- Support Vector Machines

# Support Vector Machines

Support vector machines for interpolating, extrapolating facies from data.

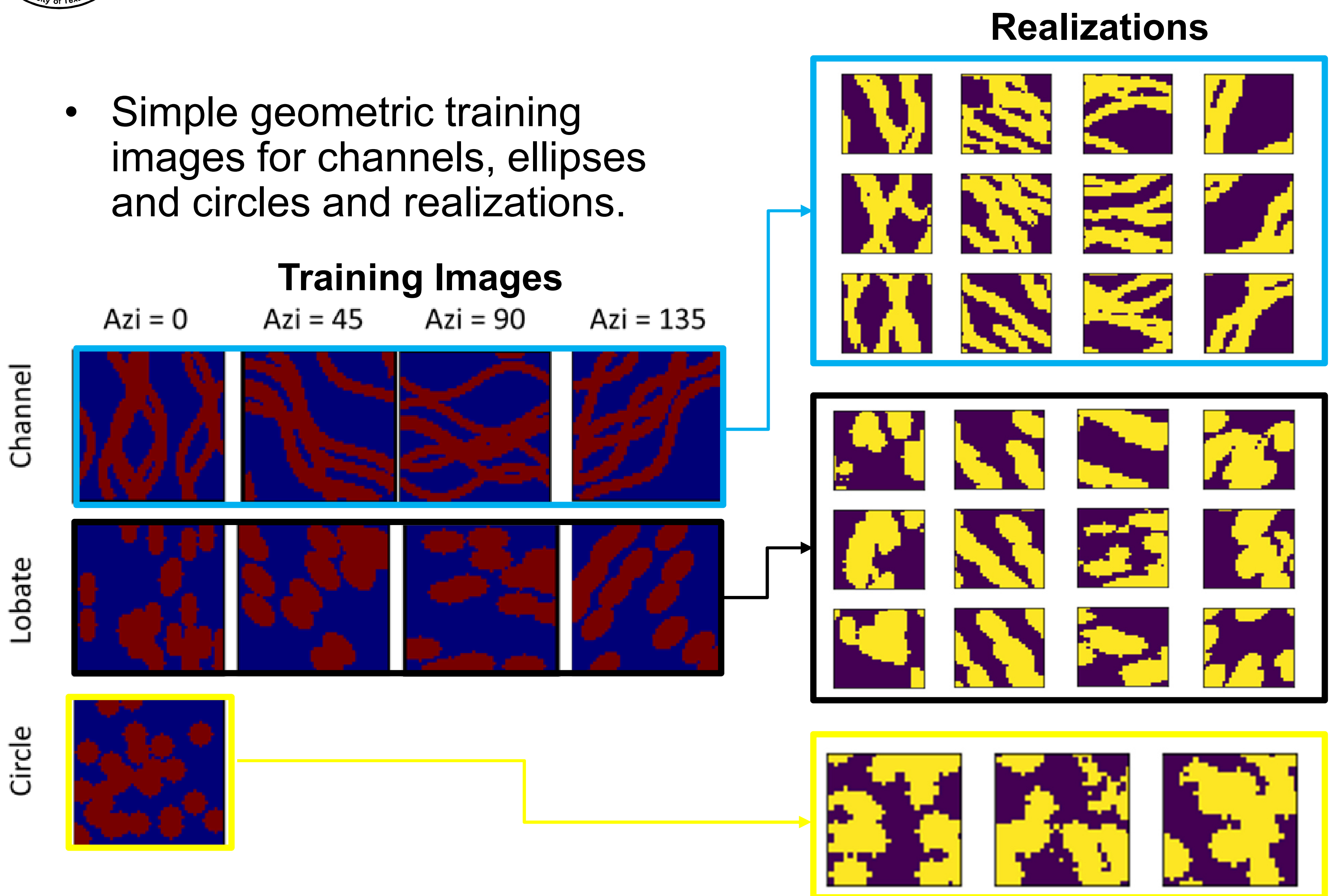



- A range of spatial models with radial basis function Kernels with a variety of penalties and Kernel parameters

Workflow developed by Wendi Liu, PhD student at The University of Texas at Austin.

# Neural Networks

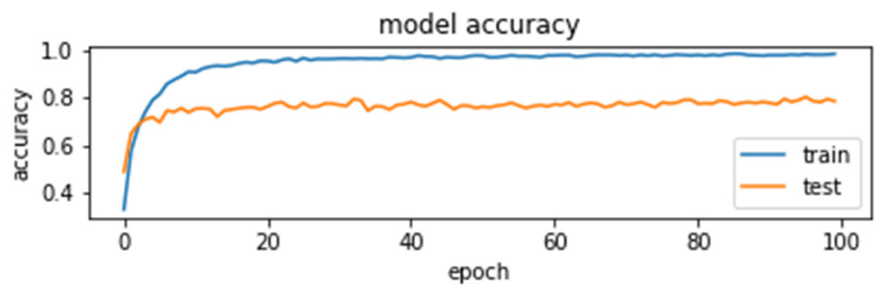
- Simple geometric training images for channels, ellipses and circles and realizations.



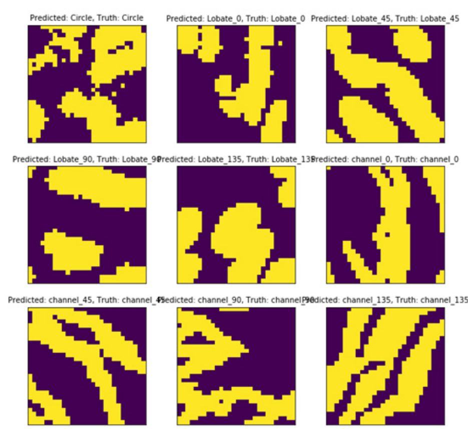

Workflow developed by Honggeun Jo and Javier Santos, PhD student at The University of Texas at Austin.

# Neural Networks

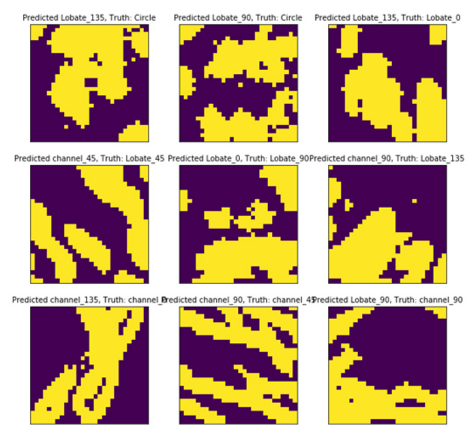
- The training and testing accuracy vs. number of training cycles, **Epoch**
- Levels off at about 78% accuracy




**Correct Identification**




**Incorrect Identification**

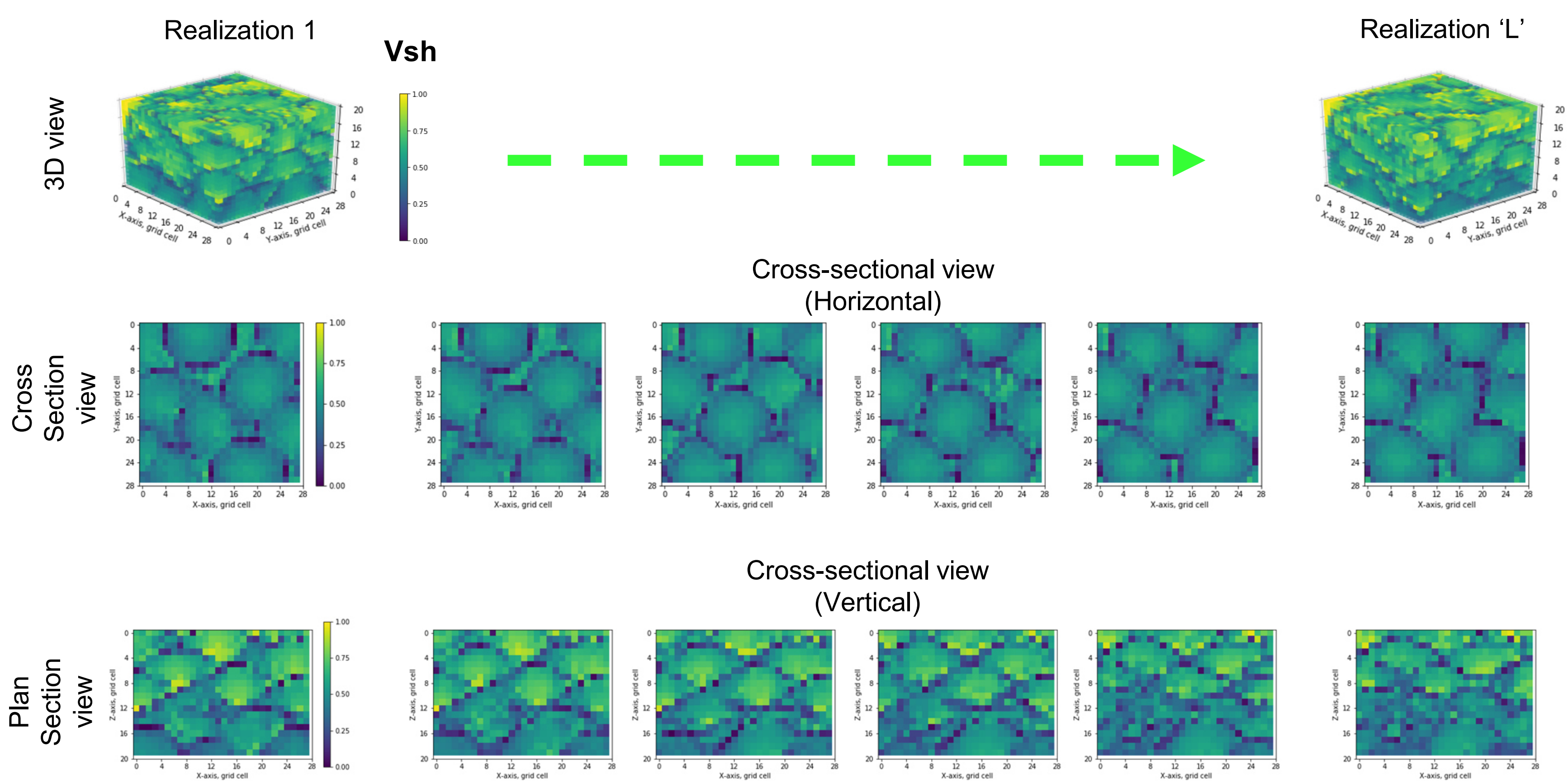



Workflow developed by Honggeun Jo and Javier Santos, PhD student at The University of Texas at Austin.

# Convolutional Neural Nets

Can explore the space of uncertainty along a continuous manifold.

- A latent reservoir manifold based on a single paramter




Workflow developed by Honggeun Jo and Javier Santos, PhD student at The University of Texas at Austin.

# Convolutional Neural Nets

Filling In Missing Spatial Information

- Semantic inpainting algorithm (Yeh et al., 2015).
- Using conceptual and perceptual information

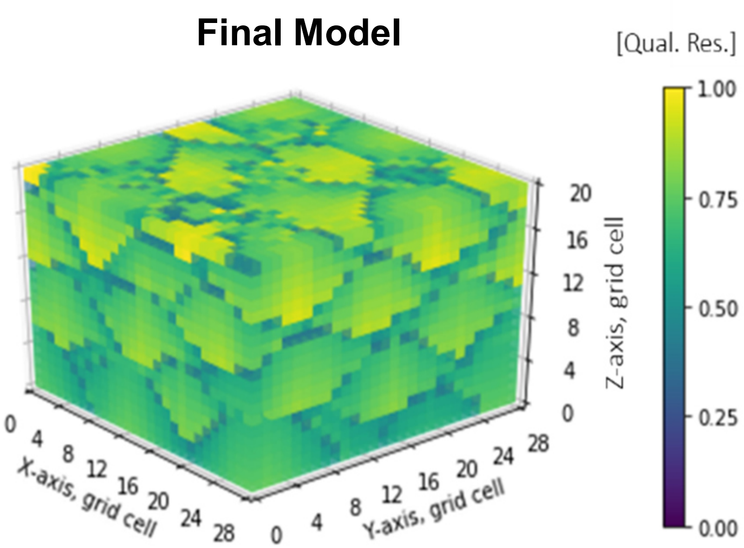

Examples of semantic image inpainting with DCGAN (Yeh et al., 2016)

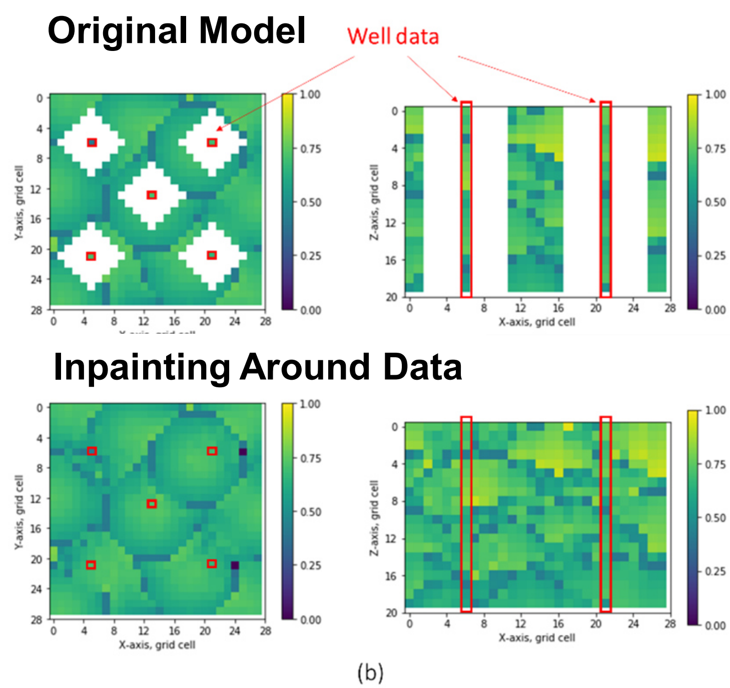
Workflow developed by Honggeun Jo and Javier Santos, PhD student at The University of Texas at Austin.

# Convolutional Neural Nets

Conditioning to Well Data?

- Remove model around data
- Use conceptual (model around mask) and perceptual (model elsewhere to fill in missing model consistent with data



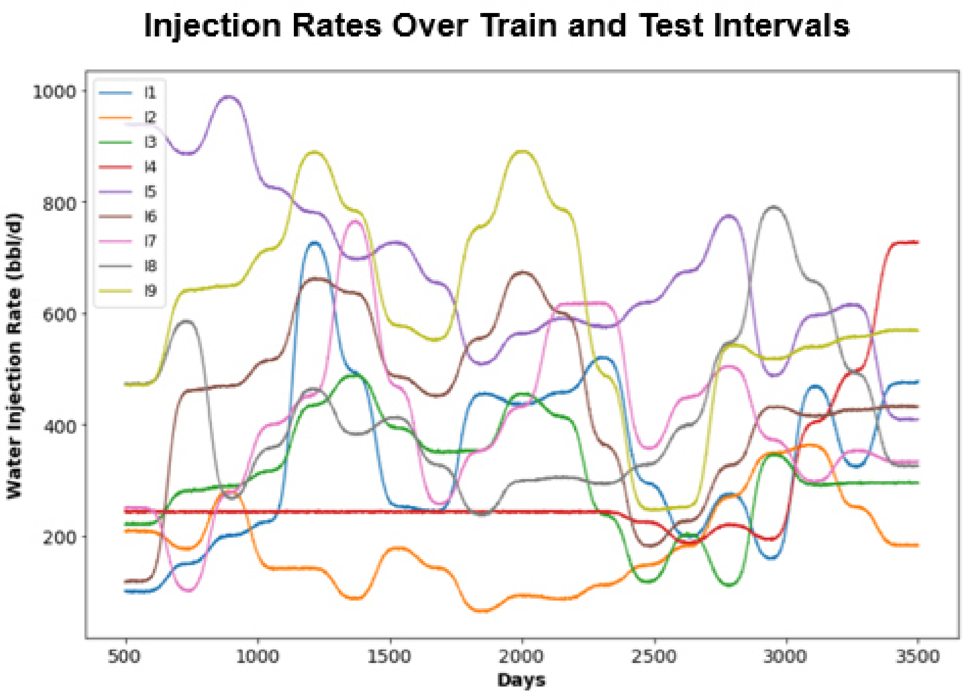



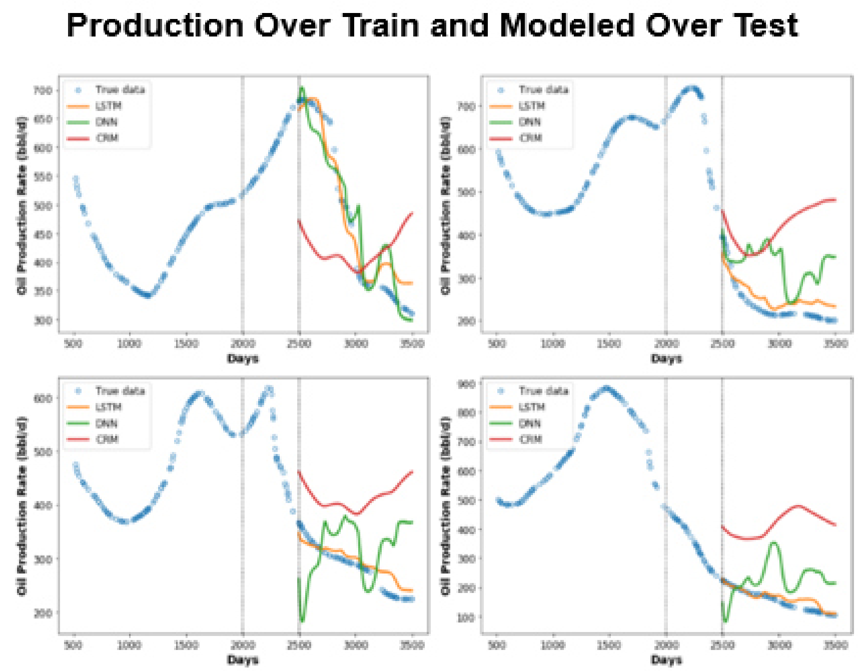

Workflow developed by Honggeun Jo, PhD student at The University of Texas at Austin.

# Long Short-Term Memory Networks

Prediction of producer flow rates based on complicated interactions of injectors.

- Train with 2500 days and predict future 100 days.




Workflow developed by Azor Nwachukwu, PhD student at The University of Texas at Austin.

# PGE 383
# Machine Learning

**Lecture outline . . .**

- **Conclusions**

**Michael Pyrcz, The University of Texas at Austin**

# Machine Learning Summary

**Subsurface is Unique**: due to: (1) sparse data, (2) heterogeneous spatial system, (3) high degree of uncertainty, (4) essential interpretation and physics, and (5) extremely high value development decisions.

**Data Analytics**: application of data cleaning and statistical analysis to support decision making. Robust use statistics and domain knowledge (geoscience and engineering) remain critical. **We must go beyond the data!**

**Data Cleaning**: is still 80 – 90% of the effort and data curation with massive variety and volume of metadata remains challenging. **Garbage In, Garbage Out**, and **Correlation is Not Causation** remain in effect.

# Machine Learning Summary

**Machine Learning Model Accuracy**: is based on these testing error components: (1) model variance, sensitivity of the model due to limited data, (2) model bias, error due to an inability to fit the complexity of the system, and (3) irreducible error due to missing variables, ranges of the variables in the training dataset.

**Model Complexity and Accuracy**: due to the trade-off between model variance and model bias, it is common for lower complexity models to outperform the accuracy of complicated models. **Don't jump to complicated!**

**Modeling Complexity and Interpreability**: more complicated models are generally more difficult to interrogate and communicate. **The model may work, but we may fail to learn from it and trust it!**

# Machine Learning Summary

**Nonparametric Models**: are typically 'parameter-rich', requiring a large number of implicit parameters; therefore, requiring a larger number of training data and resulting in greater risk of overfit.

**Overfit Models**: when a model explains almost all variance in the training, expressing high confidence, but performs poorly in testing with new data. **The model fits training data idiosyncracies!**

**Model Parameters** are set to maximize the fit with training data, and **Model Hyperparameters** determine the model complexity and are set by tuning with the withheld testing data to avoid overfit. **Overfit is insidious!**

# Machine Learning Summary

**Plus**

1. TBD

**Delta**

1. TBD

# Machine Learning Summary

**Plus**

1. TBD

**Delta**

1. TBD

# PGE 383
# Machine Learning

**Lecture outline . . .**

- **Examples of Machine Learning**
- **Conclusion**

Michael Pyrcz, The University of Texas at Austin