AI for Science Town Hall (presented by Forest Hoffman)

- OSCAR Town Hall: identify challenges with which AI could help
    - Final report after a series of workshops: doi:10.2172/1604765
    - Chapter 2 of the report focuses on Earth and Environmental sciences
        - Project environmental risk and develop resiliency in a change environment
            - Weather extremes threaten infrastructure and built environment
        - Characterize and modify subsurface conditions
            - Subsurface process understanding is key to transition to renewable energy
            - Existing subsurface data is sparse and heterogenous
            - Existing subsurface models are saddled with uncertainty and unreliable for long time periods
            - Use AI to address these problems
        - Predictive understanding of Earth system under a changing environment
            - Observations are diverse and heterogeneous
            - Incorporate knowledge from these sources into earth system models
        - Water security
    - Expected outcomes
        - Data-driven and physics-constrained hybrid models
            - Address scaling and heterogeneity issues
        - Use AI for model testbeds and surrogate models
        - Combine energy models with earth system models
            - HPC intensive
        - Use large data streams for energy production and predictive process understanding
            - Focus on data acquisition and thinking of a framework where AI/ ML can help at each phase
            - Measurement/ observations to better simulations
    - S.C. Pryor: How should we tackle the training data problem?
        - Examples of areas with missing training data?
            - Production of wind farms
            - Transmission across ISOs
        - Forest Hoffman: much of the data is considered proprietary (property of energy production companies and large facilities)
            - Possible solution: regulate and mandate that the data be available for training
        - Paul Ulrich: two solutions to lack of training data
            - 1. Increase the amount of data
            - 2. Impose more constraints on the problem to reduce the solution space of possible outputs
                - Eliminate “unphysical” possibilities

<u>Topics for the panel</u>

Prospects for AI/ML

- **What major scientific advances should we strive for by deploying AI/ML?**
    - Katie Dagon - can we use ML to automate calibration of climate models?
        - Results are mixed based
        - Crucial question: are we gaining efficiency with ML? Expert-guided tuning is time-intensive
        - Benefit: gain computational efficiency
    - Katie Dagon - D&A of extremes
        - There is lot to learn from CS literature and experts
        - Interdisciplinary collaboration is key
        - Interpretation and validation of D&A is paramount.
    - Katie Dagon - how can we use ML to improve predictability and prediction?
    - Ruby Leung - Use AI/ ML to identify precursors to wildfires or other extremes
        - Currently, the precursors are unknown to us. Maybe there are precursors in fields that are less well-explored (e.g. land-surface fields or soil moisture). There tends to be more emphasis on fields relating to atmospheric circulation.
        - Usually, hindsight case studies are done after the storm hits
        - Steer observations to help change the monitoring systems
        - Will the precursors change in the future making it harder/ easier to predict the event
    - Ruby Leung - use AI/ ML to identify the most powerful constraints to improve predictive understanding and skill?
        - Question to ask ML: use ML to assign a probability to the projection coming from different models
            - Why did you assign the probability you did?
            - Example test case: S2S models
    - Balu Nadiga - improve predictive skill and put climate prediction on the same footing as weather prediction
        - Currently, climate modelling is done with the goal of process understanding
        - Use ML for a paradigm shift to improve predictive skill
            - Currently, climate prediction is beset by uncertainties
            - One possibility: *first*, get predictive skill using whatever methods necessary. *Then,* get better scientific understanding.
    - Alex Hall - look more at biogeochemical variables
- **What does success look like?**
    - Katie Dagon - reduced model bias is the goal
    - Alex Hall (on model and observational diagnostics) - nonlinear modelling
        - Development of understanding of modes of variability was based on linear statistics
        - Physics of the system are nonlinear. Variability modes should fit this better.

        - Example: how would ENSO have been characterized if ML was available at the time? Probably ML could have represented the oscillation and its impacts better.
    - Alex Hall - understand the three unknowns below
        - Relative contributions of parametric and structural uncertainty,
        - Use ML to understand internal variability
        - Use ML to determine the degree of model overlap because of shared code (how orthogonal are they?)
    - Forest Hoffman - use ML for gap-filling more effectively (especially in nonlinear cases)
        - Account for heterogeneity and diverse datasets
    - Forest Hoffman - use ML as a probe to better understand parameterizations
        - If ML can improve a specific parameterization, what can that tell us about the existing parameterization of the process?

Q&A

- **Michael Wehner - what are the prospects for unsupervised learning in climate?**
    - Katie Dagon - semi-supervised and unsupervised learning could be useful for working in new climate states
        - Good example: Naomi Goldenson's presentation on using self-organizing maps and neural networks to identify drivers of ARs
    - Balu Nadiga - generative adversarial networks offer promise in this area
        - GANs can be used to augment existing datasets by creating new data with similar characteristics of the original dataset. This way, it will be possible to have more training data
        - Leverage reinforcement learning for no-analogues situations
- **Angeline Pendergrass - is ML useful for relationships that are fundamentally linear?**
    - Alex Hall - the beauty of ML is that it does not make any assumptions about linearity or nonlinearity of the data
        - Libby Barnes's paper: uses NNs to identify anthropogenic signals
    - Balu Nadiga - even in a linear system, the memory of the system can make the problem space very complex
        - Cookman operators - raise a nondynamical system to an infinite dimensional space.
        - there are linear operators in the infinite dimensional space that can represent the nonlinear changes
    - Paul Ulrich - ML methods can be "forced" to be linear by using linear activation functions
    - Matt Newman - a nonlinear system may be predictably linear on some time scales
    - Paul Ullrich - ML methods have the potential to quantify nonlinearity
        - In streamflow forecasting, he uses the Nash-Sutcliffe coefficient to quantify the improvement of ML over regression over the same set of inputs

- Matt Newman - a danger of using ML to represent a linear system is that it may overfit to noise
  - The ML model would overfit unpredictable noise as predictable nonlinearity
  - For this reason, dynamical models may be a better approach, since they relate to the physics of the system
- **Jitu Kumar - are explanaibility approaches scalable and transferable to other problems? Or are they custom for each problem at hand?**
  - Katie Dagon - Layerwise relevance propagation and feature importance tests are transferable across topics
    - Some classes of problems have different interpretability methods: e.g. classification vs. regression
  - Colin Zarzycki - Libby Barnes and DJ Gagne are good examples of NN interpretability
  - Paul Ullrich - new tools from the ML community on interpretability
    - “Transformers” (which are a specific class of neural network architecture) give rise to activation maps naturally
  - Paul Ullrich and Jitu Kumar - they pose an open question on explainability
    - If 2 neural networks have different hyperparameters but similar accuracy, will the interpretability methods yield the same output?
    - Paul Ullrich - using a model with too many parameters leads to overfitting. Using a model with too few parameters leads to underfitting. Hyperparameter searching should be used to find the ideal model architecture.
- **Paul Ullrich - There is an open problem related to memory**
  - There needs to be data reduction in the input space. A key challenge with using ML/AI for climate problems is that there are memory bottlenecks during training, because the input is large.