Examples of Mission-driven Data Science from Jefferson Lab and ACES

Diana McSpadden, Data Scientist – Data Steward

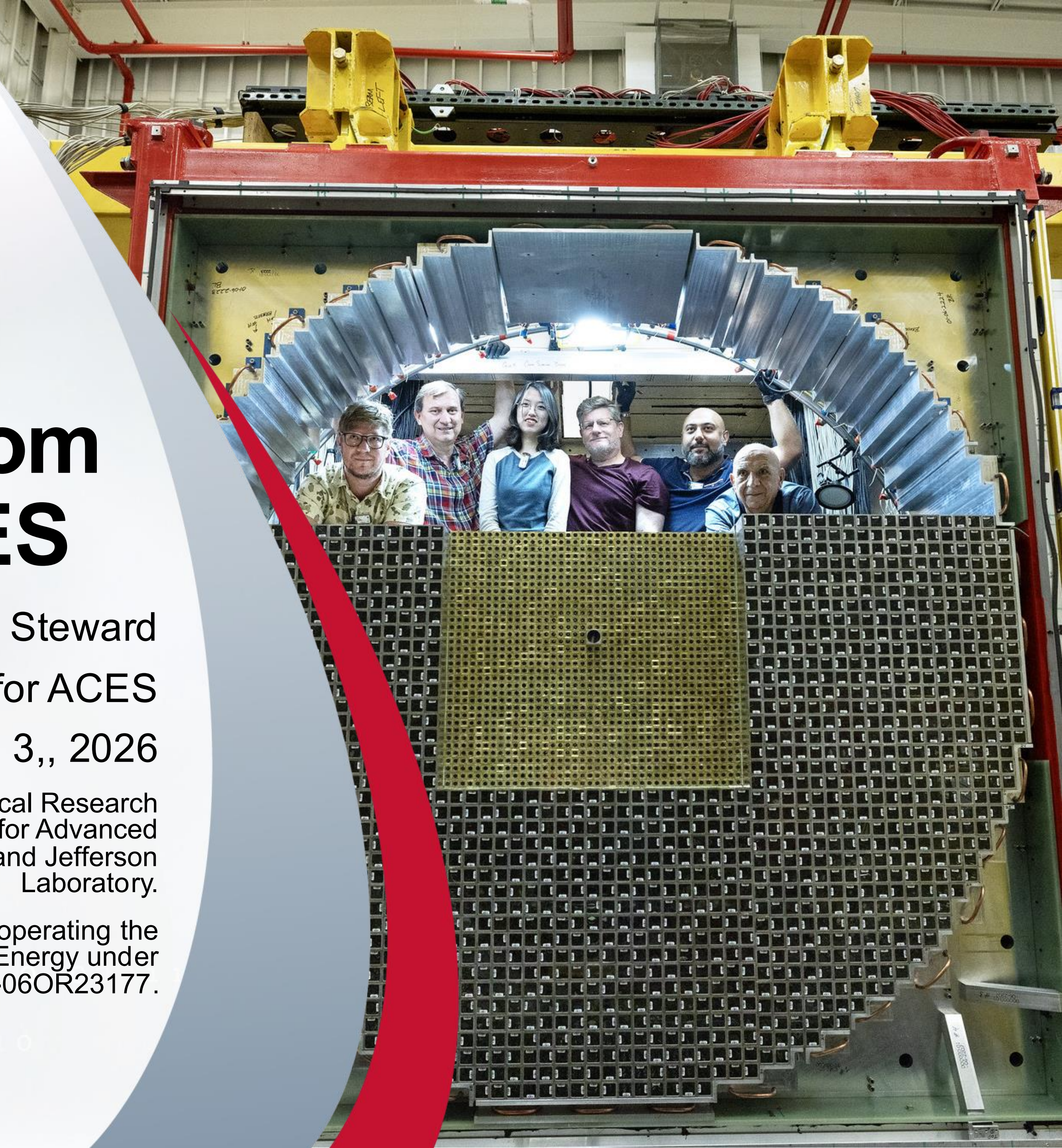
Jefferson Lab and Joint Institute for ACES

March 3,, 2026

Funding for this effort was provided by The Hampton Roads Biomedical Research Consortium as part of the efforts associated with the Joint Institute for Advanced Computing on Environmental Studies between Old Dominion University and Jefferson Laboratory.

This presentation has been authored by Jefferson Science Associates (JSA), operating the Thomas Jefferson National Accelerator Facility for the U.S. Department of Energy under Contract No. DE-AC05-06OR23177.

Jefferson Lab



Diana McSpadden
Data Scientist – Data Steward

dianam@jlab.org

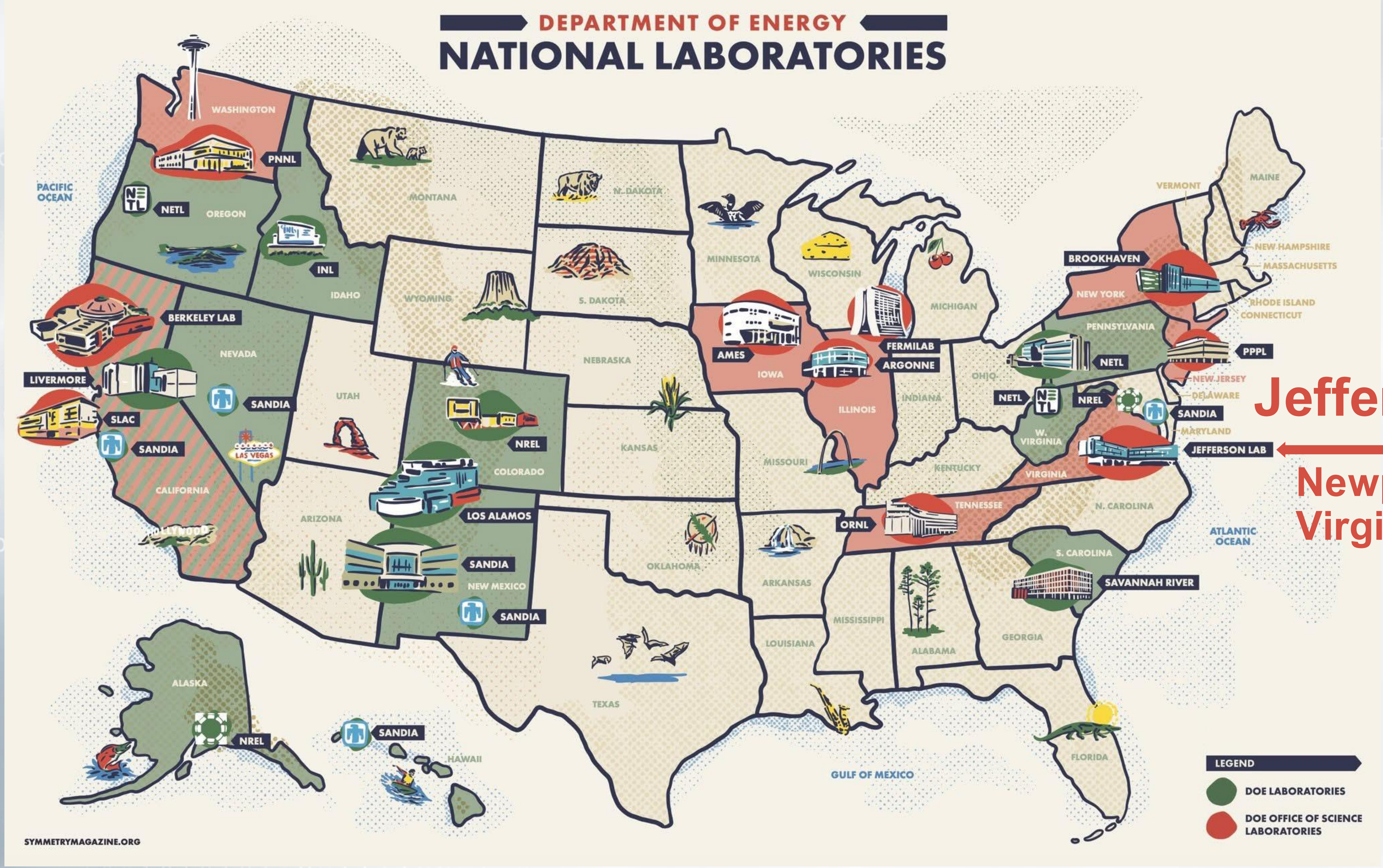
data.jlab.org

My interests:

- *Scientific data stewardship*
- *FAIR and AI-Ready data*
- *Uncertainty quantification (UQ) for risk-aware decision support*
- *Machine learning (ML) for regional flooding*

Today:

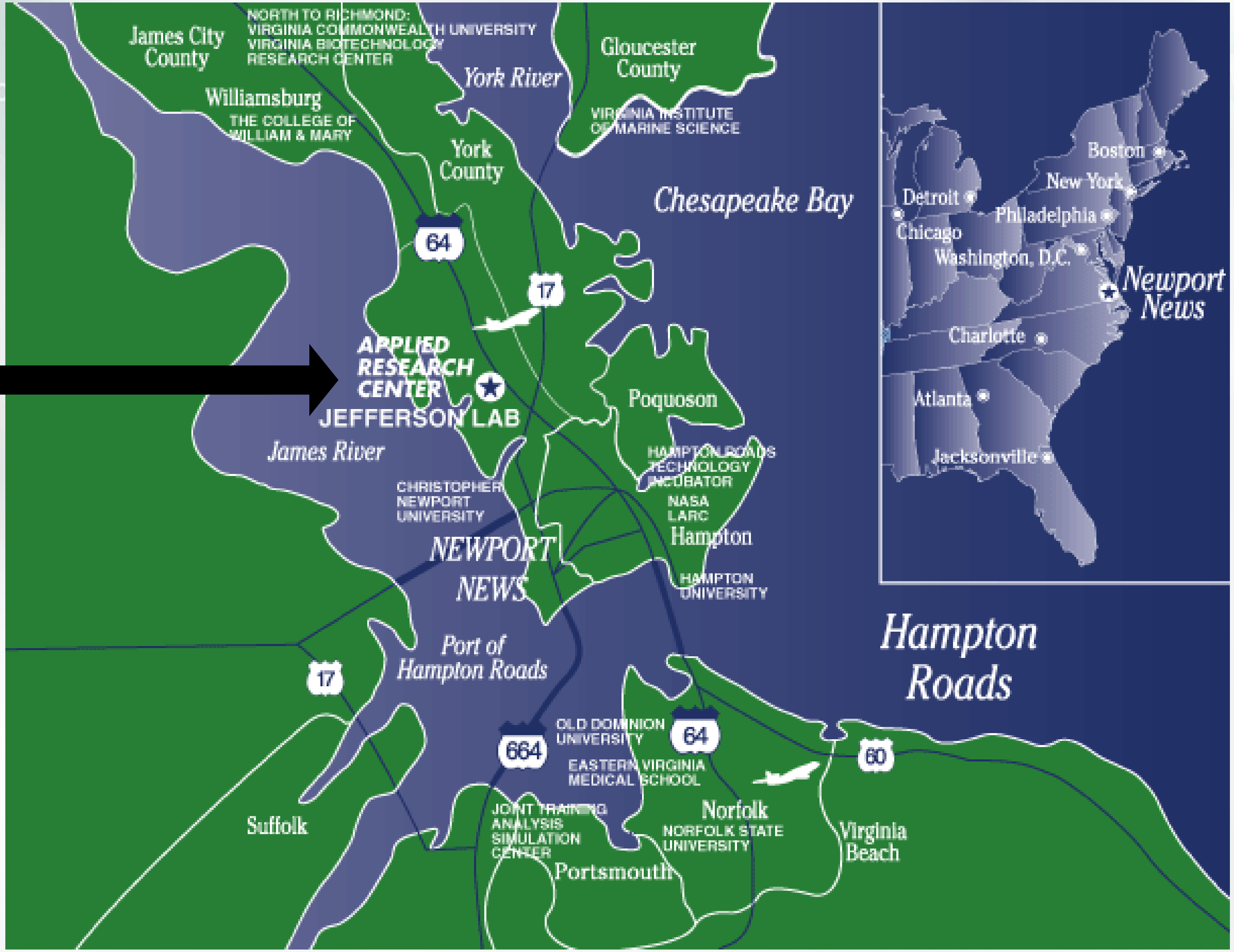
- *Introduction to Jefferson Lab*
- *ML and UQ for risk-aware coastal flood management*



Jefferson Lab

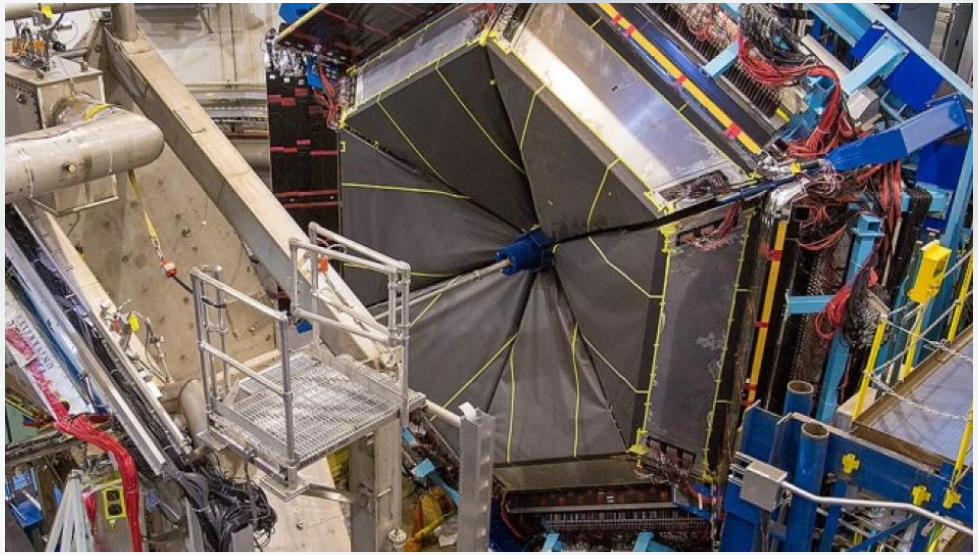
Newport News, Virginia

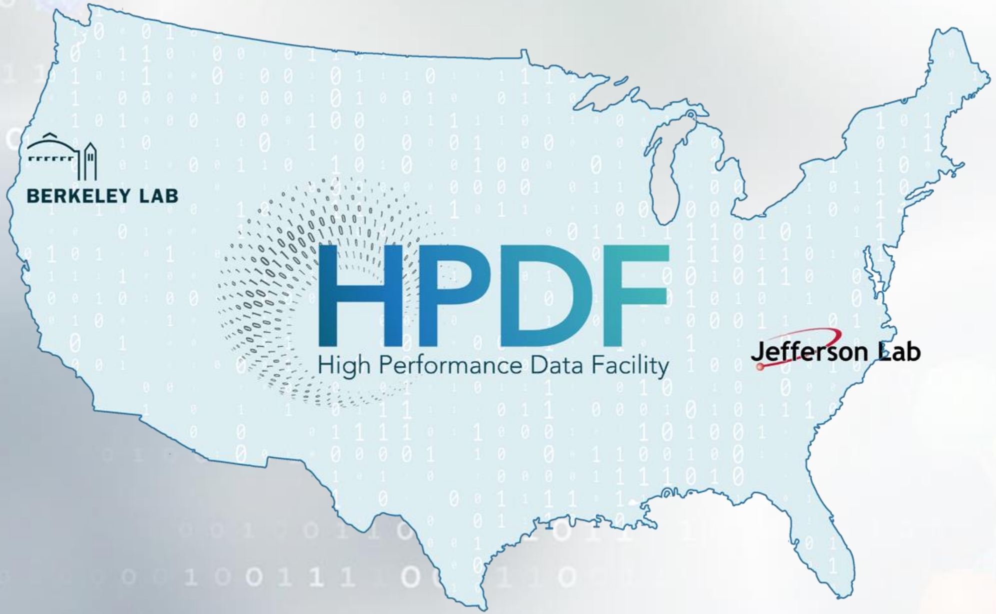
Jefferson Lab



Jefferson Lab

Thomas Jefferson National Accelerator Facility (Jefferson Lab) is a U.S. Department of Energy Office of Science national laboratory. Scientists worldwide utilize the lab's unique particle accelerator, known as the **Continuous Electron Beam Accelerator Facility (CEBAF)**, to probe the most basic building blocks of matter.

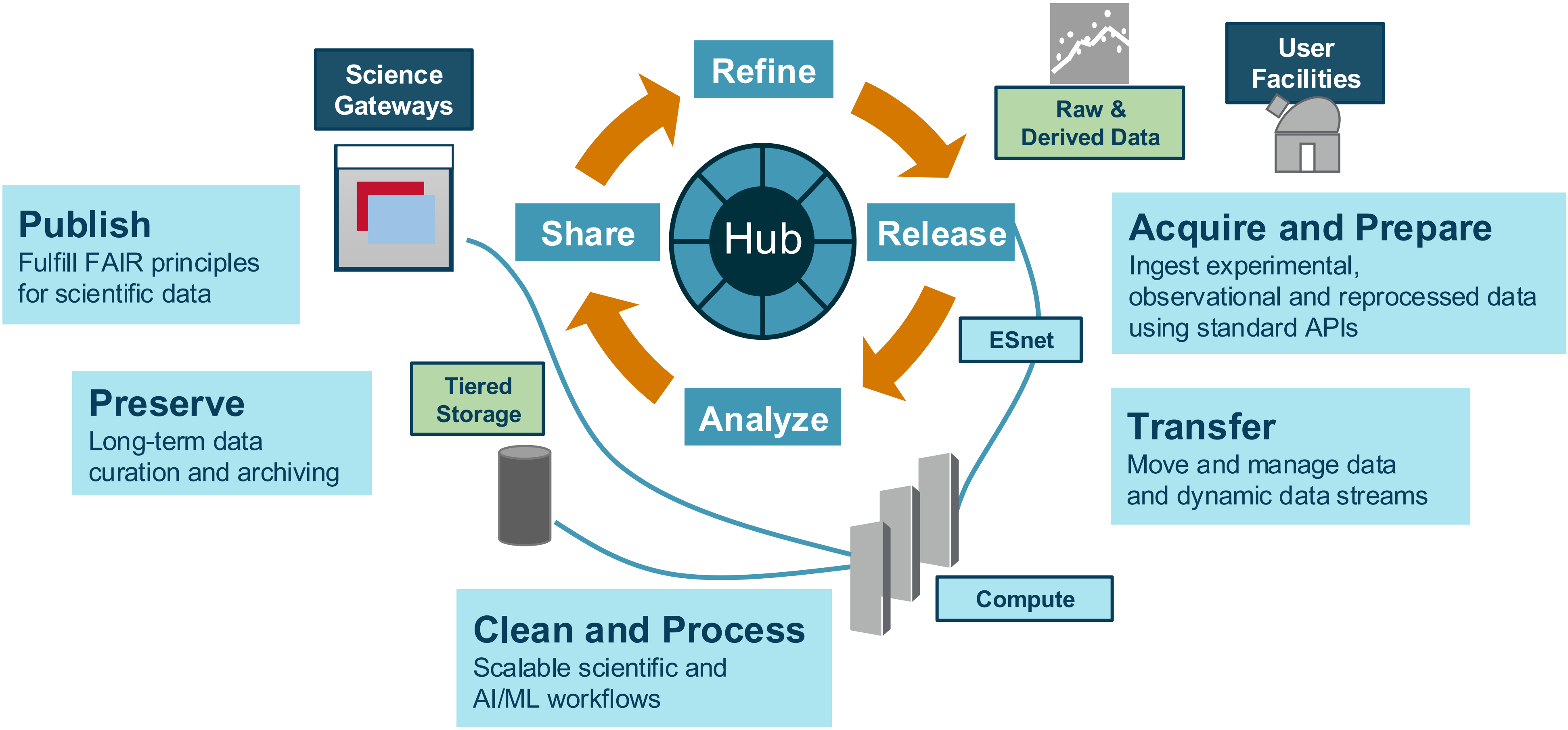




DOE Selected Thomas Jefferson National Accelerator Facility as the lead for its new **High Performance Data Facility (HPDF)** Hub. Jefferson Lab will partner with DOE's Lawrence Berkeley National Laboratory to form a joint project team led by Jefferson Lab. The HPDF will be a computing and data infrastructure resource that will provide transformational capabilities for data analysis, networking and storage for the nation's research enterprise.

https://www.jlab.org/
https://science.osti.gov/np/Facilities/User-Facilities/CEBAF
https://www.hpdf.science/

HPDF will support data lifecycle management



The HPDF Hub is a cutting-edge high-performance data platform that provides advanced storage, processing, and analytical services, empowering data-intensive research and AI-driven discoveries.

JLab and the Genesis Mission

Energy Department Advances Investments in AI for Science

The U.S. Department of Energy (DOE) today announced over $320 million in investments to rapidly advance the Genesis Mission's artificial intelligence (AI) capabilities.

Energy.gov

December 10, 2025

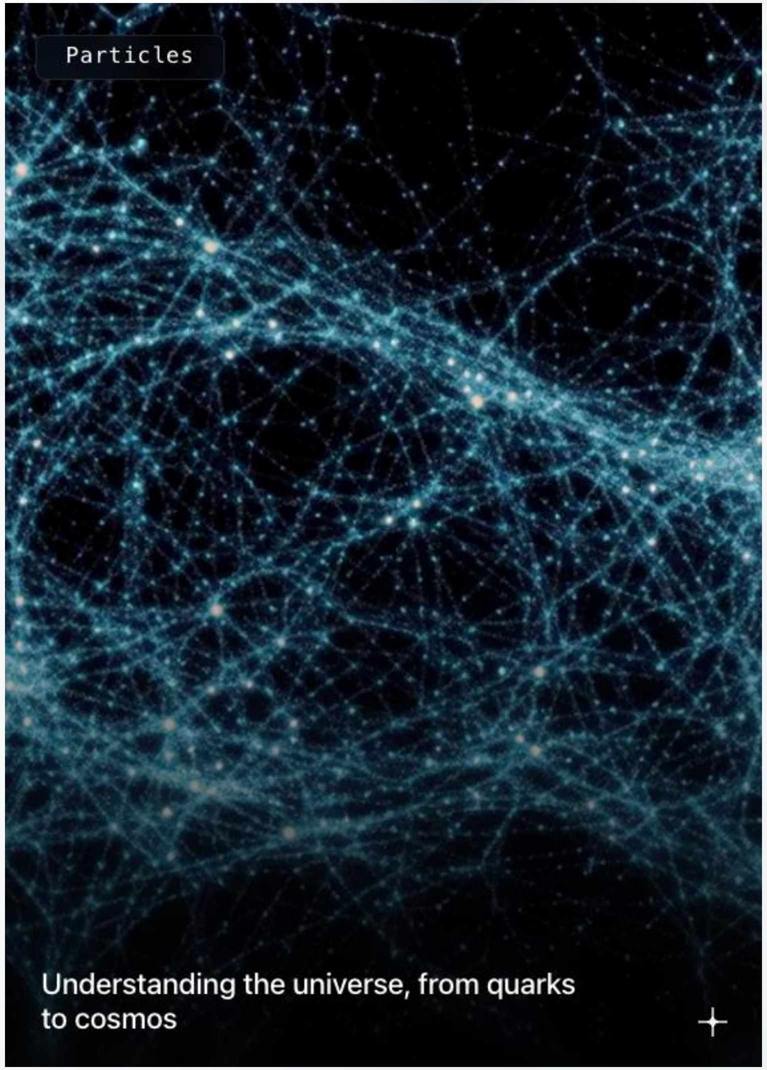
JLab participates at every level:

✓ HPDF

✓ Infrastructure

✓ Models

✓ Data



https://genesis.energy.gov/
https://american-science-cloud.github.io/amsc-site/

Data-driven flood forecasting and some potential implications

Jefferson Lab

aces ADVANCED COMPUTING for ENVIRONMENTAL STUDIES

Generated by ChatGPT

Joint Institute on Advanced Computing for Environmental Studies (ACES)

2 Research Arcs

- Health Informatics
- Environment
  - Air modeling
  - Flood modeling

Interdisciplinary Research Teams

- Public Health
- Climate Modeling
- Computer Science & Data Science
- Physics

Collaboration is essential!

Data providers are vital!

National labs are key!

ACES flood modeling


Daniel Lersch
Staff Data Scientist


Diana McSpadden
Staff Data Scientist – Data Steward


Steven Goldenberg,
Staff Data Scientist


Ahmed Mohammed
Staff Data Scientist


Khan M. Iftekharuddin
Co-Director JI-ACES


Chetan Kumar,
Research Scientist

Coastal flooding in Hampton Roads, Virginia

- Norfolk is one of the most vulnerable communities to coastal flooding
- Can expect over 200 flood events annually by 2049[1]

"A father and daughter kayak down Magnolia Avenue in Norfolk, Virginia, in 2009 after a severe nor'easter."

Hyunsoo Leo Kim -- The Virginian-Pilot

[1] AG Burgos, BD Hamlington, Philip R Thompson, and Richard D Ray. Future nuisance flooding in Norfolk, VA, from astronomical tides and annual to decadal internal climate variability. Geophysical Research Letters, 45(22):12–432, 2018.

Physics-based models and ML-based surrogate models

Why use machine learning?

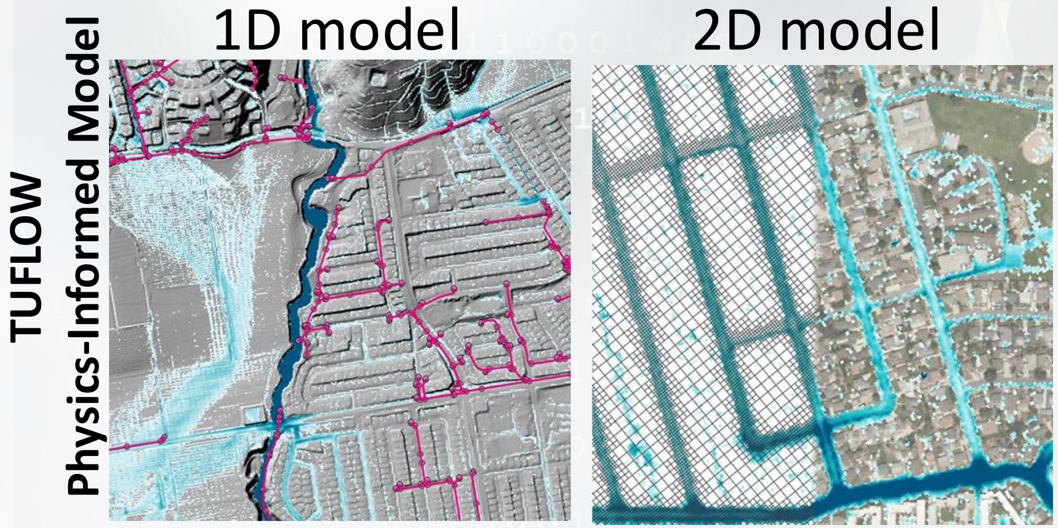
- Physics-based models:
  - **Highly accurate**
  - **Computationally expensive**: 4.5 – 6 hours for a 24-hour rainfall event
  - Impractical for real-time decision support

ML methods:

- Computationally efficient
- Techniques for estimating uncertainty

Implications of Data-Driven Methods:

- Data collection and processing
- Estimating model uncertainty

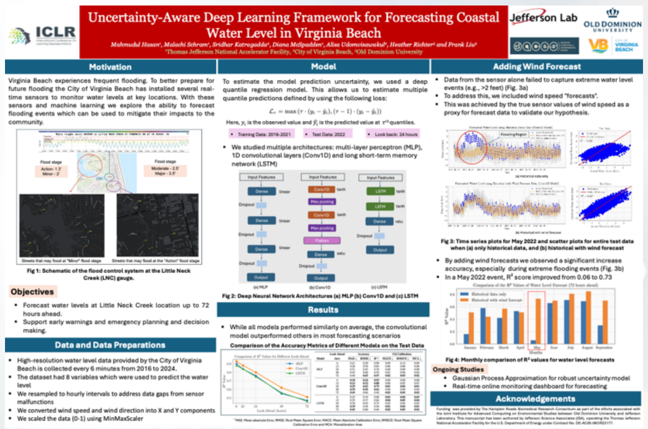


TUFLOW is the physics-informed hydrodynamic model used to model "true" water depth for JI-ACES-UVa research in Norfolk, Virginia.

Use physics principles. (e.g., conservation of energy, friction, turbulence). UVa colleagues ran our TUFLOW simulations.

https://tuflow.com

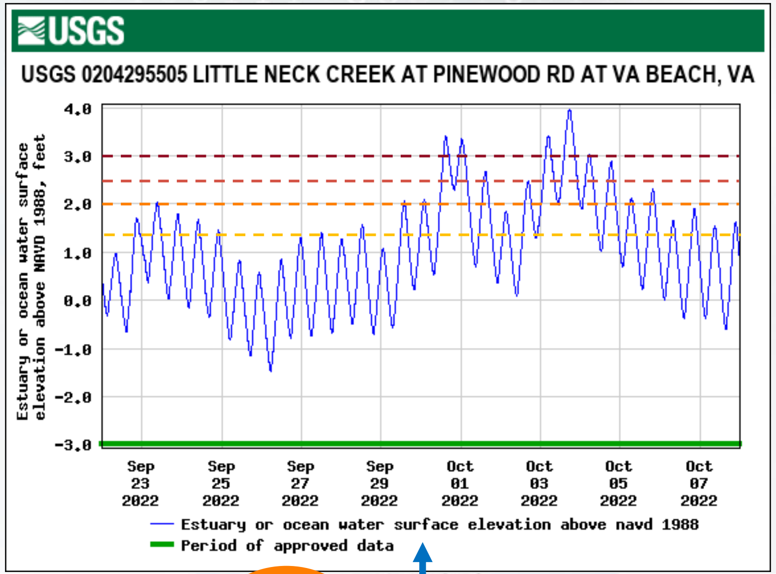
Uncertainty-Aware Deep Learning Framework for Forecasting Coastal Water Level in Virginia Beach



City of Virginia Beach requires at least 72-hour water-level forecasts at key stream gauge locations.

Little Neck Creek stream gauge:

- Water level → flooding conditions in the surrounding streets



Water Level & Impact stage

- **Used by City Public Works:**
  - Close streets
  - 3 days to drain stormwater storage

Flood Stage:

Action: 1.3 ft
Minor: 2.0 ft

Flood Stage:

Moderate: 2.5 ft
Major: 3.5 ft

Streets expected to flood at "Minor" stage

Streets expected to flood at "Action" stage

https://waterdata.usgs.gov/monitoring-location/USGS-0204295505/

Forecasting Coastal Water Level in Virginia Beach: Gauges and Sensors

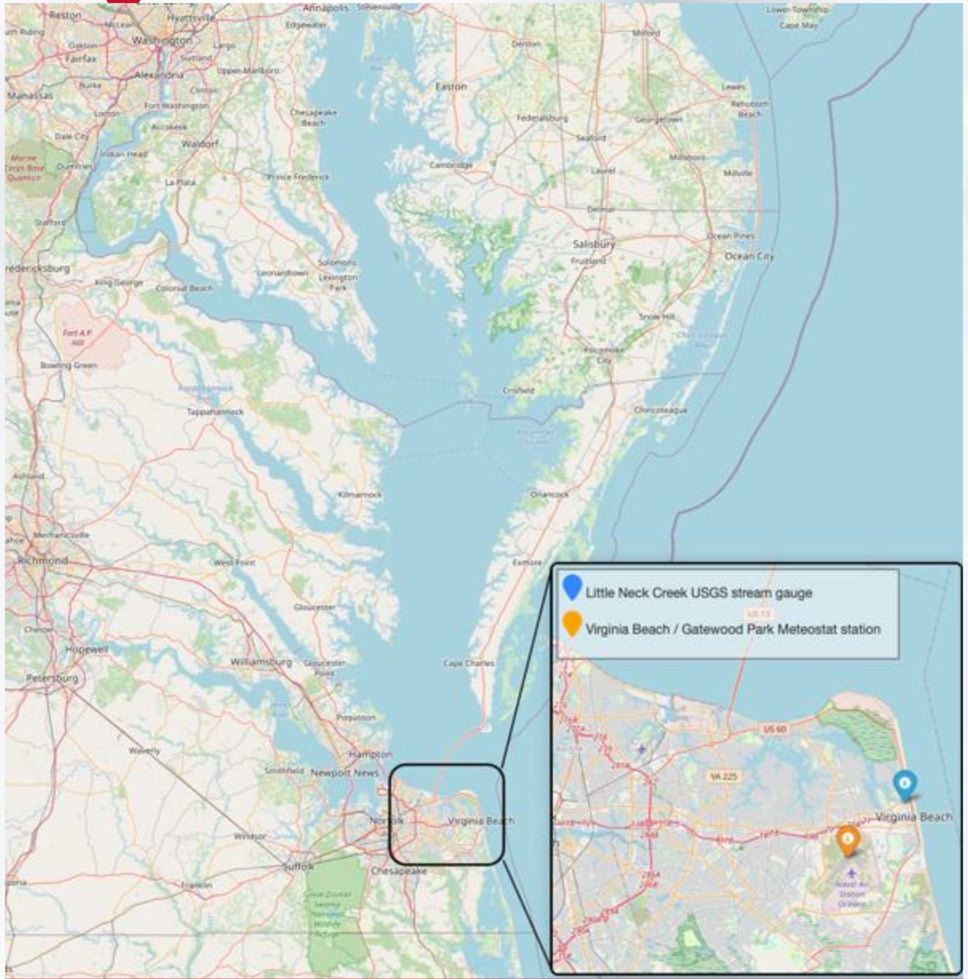
Forecasting 72-hour water level

Historical Inputs (Previous 24 hours)

- Water Level (feet)
- Wind Speed/Direction
- Precipitation (inches)

Historical Weather (Previous 24 hours)

- Atmospheric Pressure (hPa)
- Temperature (°C)
- Dew Point (°C)



USGS/Virginia Beach LNC gauge*: water level, wind speed/direction, and precipitation data at* ***six-minute*** *intervals.*
Virginia Beach/Gatewood Park OpenMeteo station*: air pressure, temperature, and dew point, were retrieved* ***hourly****.*

Data-Driven Implications: Data Preparation & AI-Readiness

Encourage the documentation of preparation of shared data, and AI-readiness pipelines.

1. **Down sampling & alignment**
2. Transformations
3. Including Wind Forecast
4. Train / Val / Test Split
5. Sequencing
6. Scaling data

Data-Driven Implications: Data Preparation

The City of Virginia Beach supplied data sources previously used for comparison.

Data includes:

1. City of Virginia Beach / USGS stream gauge:
   - Recorded every six minutes
   - Inconsistent stream gauge availability
1. OpenMeteo weather station
   - Recorded every sixty minutes

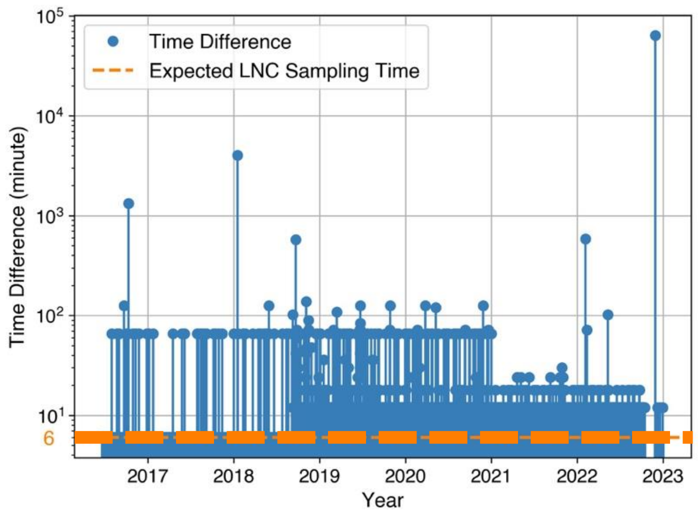
Data Alignment:

- Adjusted LNC data to match hourly intervals of OpenMeteo historical weather data.

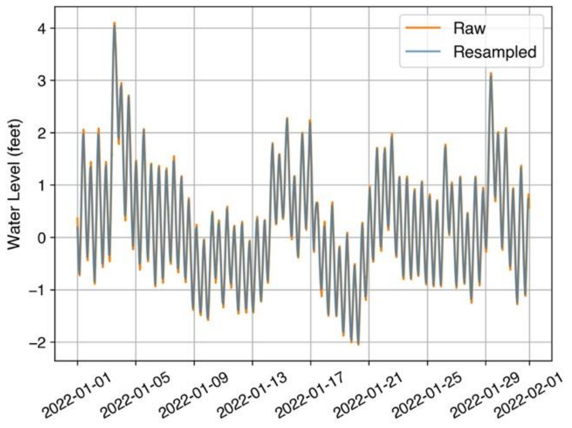
Aggregation Method:

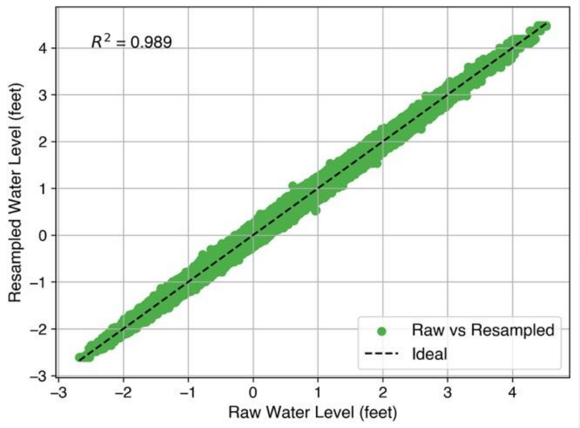
- Hourly averages for most variables.
- Maximum value for precipitation to capture infrequent rainfall events.
- No interpolation applied for missing samples.

Visualization demonstrates minimal info loss



Ideally, the USGS tide gauge samples every 6 minutes (orange line). The plot shows a non-consistent sampling interval.





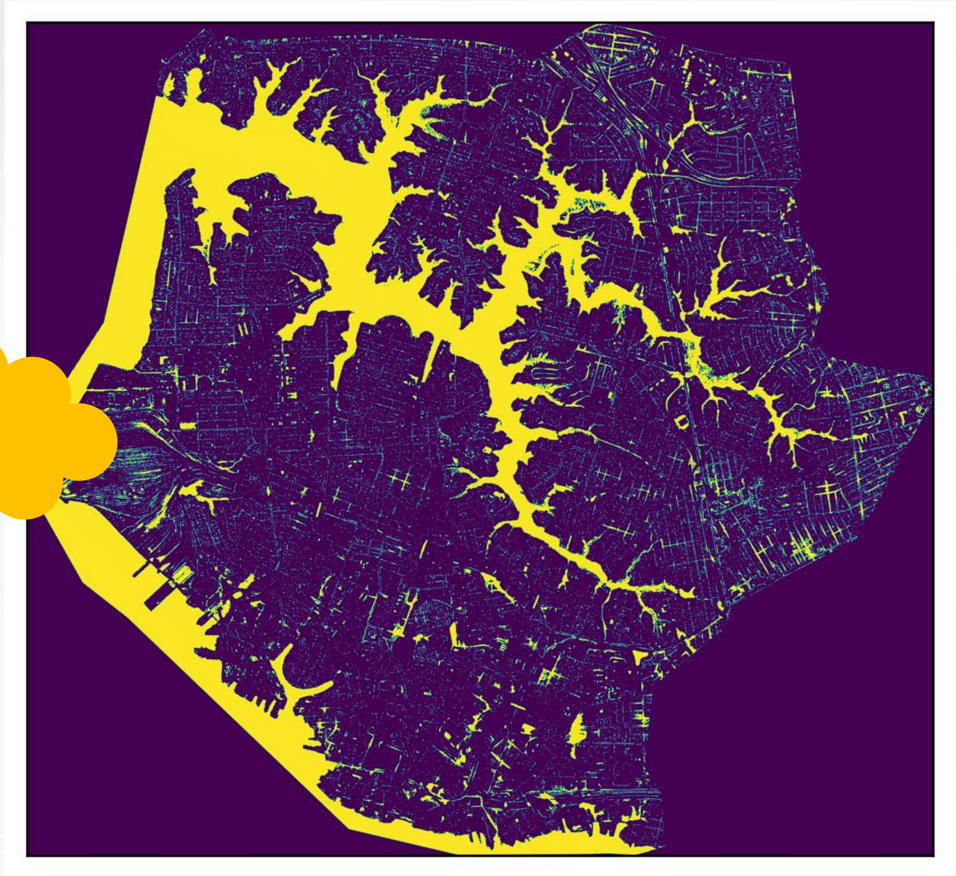
Comparison between raw water level (6-minute frequency) and resampled water level (1-hour frequency) is shown in terms of time series plot (left) and scatter plot (right).

AI Uncertainty Estimation – Why should I care?

Training Data

Operating Conditions

**Similar?
Not?
Out-of-Distribution?**

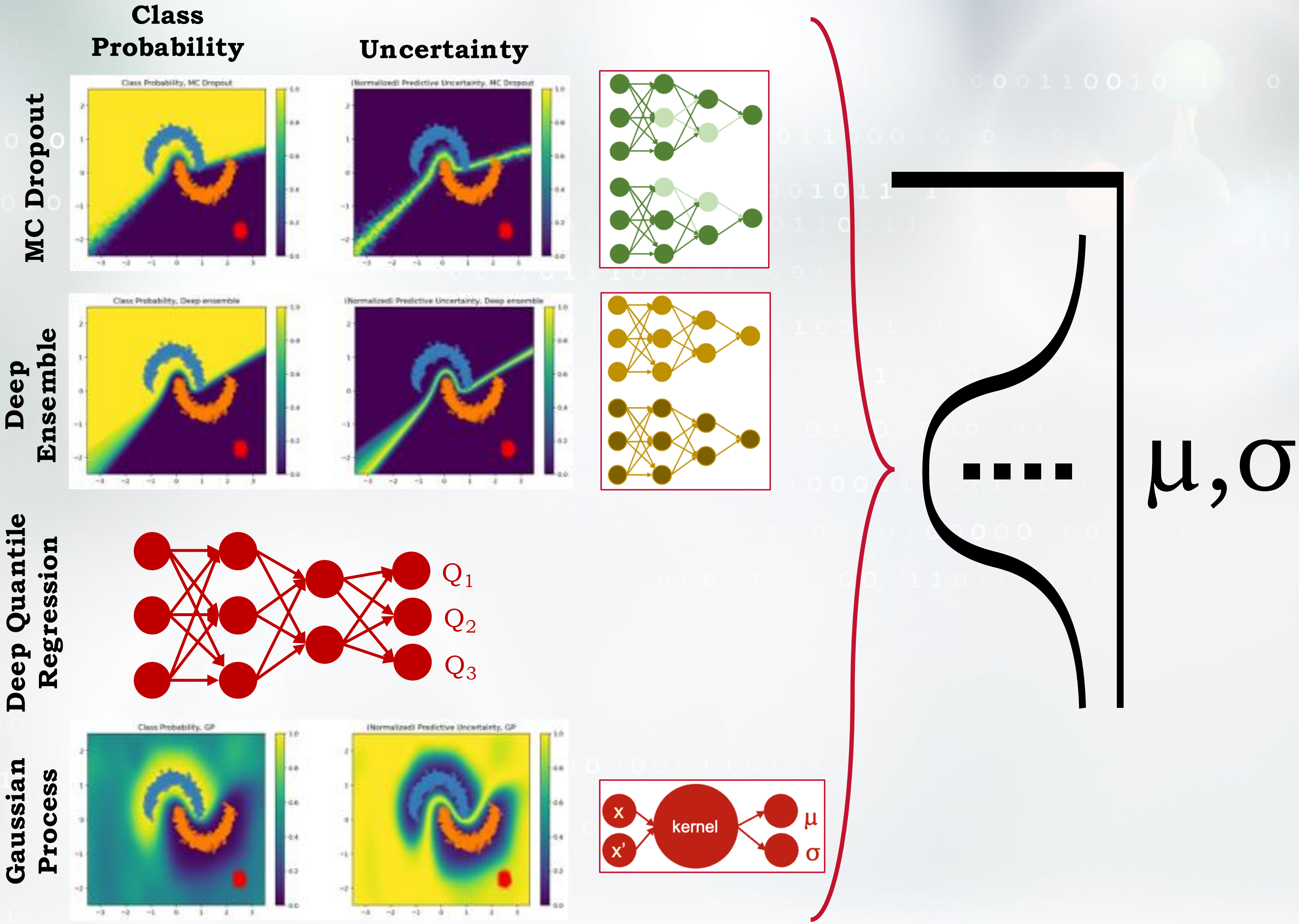


National Oceanic and Atmospheric Administration (2024, May 14). *April 2024 was Earth's warmest on record*. NOAA.gov. Retrieved November 3, 2024, from https://www.noaa.gov/news/april-2024-was-earths-warmest-on-record

Examples of ML Uncertainty Quantification Techniques

Half Moon Toy Problem

1. Categorize (x,y) coordinates into **BLUE** and **ORANGE** classes
2. Uncertainty (standard deviation) estimated differently by different models
3. Especially for samples not seen in training data (**RED**).



Evaluating Uncertainty Aware Models

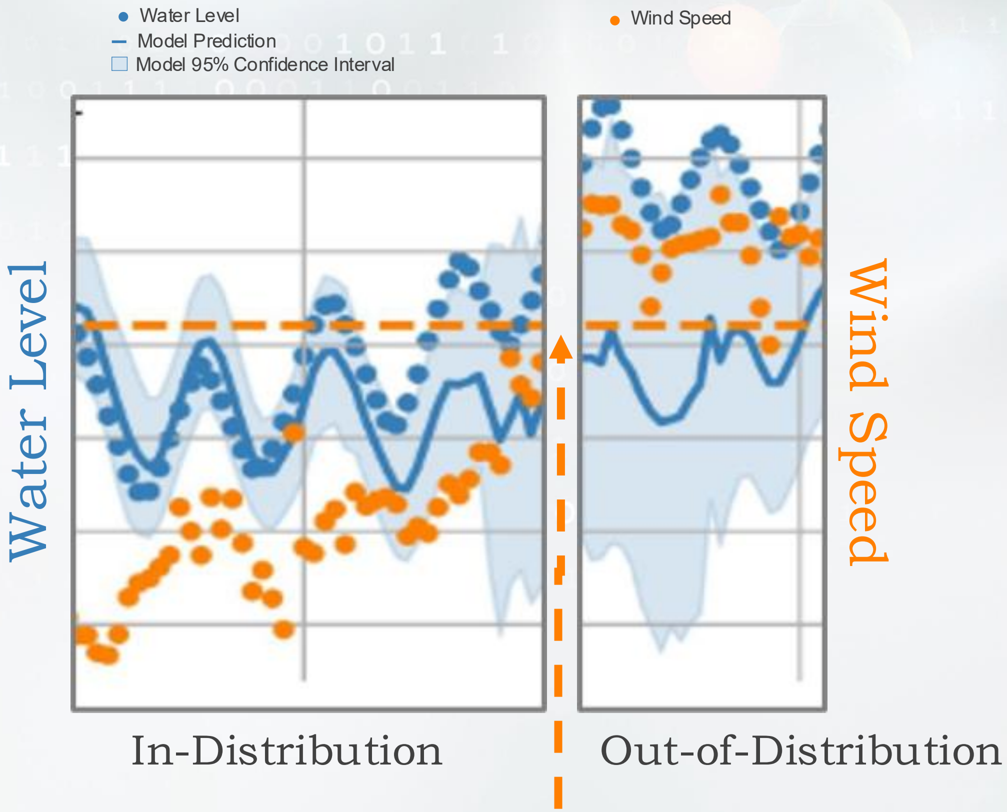
Evaluations:

1. Accuracy
2. Uncertainty Calibration
   - 10% confidence yields correct predictions 10% of the time.
   - 90% confidence yields correct predictions 90% of the time.
3. Uncertainty Distance Awareness
   - Uncertainty increases with greater distance from training data, i.e. increasing unfamiliarity.
   - GP methods are distance-aware, comparing data points with mathematical similarity decreasing as data distance increases.
   - DQR, MC Dropout, and ensemble methods show similar distance awareness, require empirical evaluation for reliability.

Encourage documentation of a dataset's limitations, and appropriate operational use.

Uncertainty quantification for scientific AI

- JLab focuses on **distance-aware UQ** for scientific AI
- Necessary for real-time inference & diagnostics, **scientific system control**, and real-time, **risk-aware** decision support
- Critically important when **safety constraints should rarely or never be violated**
- Intersection of **data**, **AI**, **risk**, and the **built environment**
- Implications for **critical infrastructure assets near water,** and data-driven methods are used for real-time decisions of consequence



Out-of-Distribution Boundary Encountered.
High uncertainty requires risk assessment.

JLab-ACES selected publications

JLab-ACES Dataset Publications

Roy, B., S. Goldenberg, D. McSpadden (2025). Input Data for LSTM Model for Street-Scale Nuisance Flood Forecasting in Norfolk, Virginia using Transfer Learning, HydroShare, http://www.hydroshare.org/resource/cdaaadf3e934466a85327a9c3ee1f3e0

Roy, B., S. Goldenberg, D. McSpadden (2025). Input data for LSTM and seq2seq LSTM surrogate models for multi-step-ahead street-scale flood forecasting in Norfolk, VA, HydroShare, http://www.hydroshare.org/resource/e5d6d32a320f4bcca679e0bf388c2bcc

McSpadden, D., S. Katragadda, C. Kumar, A. H. Mohammed, M. Hasan, M. Schram, A. Udomvisawakul, H. Richter (2025). USGS Stream Gauge 0204295505 at Little Neck Creek, Virginia Beach, VA, and Associated Weather Data (2016–2022) in CSV Format, HydroShare, https://doi.org/10.4211/hs.94a65edc37c943e490a25347a3744a16

Selection of JLab-ACES publications

McSpadden, D., Goldenberg, S., Roy, B., Schram, M., Goodall, J. L., & Richter, H. (2024). A comparison of machine learning surrogate models of street-scale flooding in Norfolk, Virginia. Machine Learning with Applications, 15, 100518. https://doi.org/10.1016/j.mlwa.2023.100518

Goldenberg, Steven, McSpadden, Diana, Roy, Binata, Schram, Malachi, Goodall, Jonathan L., & Richter, Heather (2023). Uncertainty Quantified Machine Learning for Street Level Flooding Predictions in Norfolk, Virginia. https://www.climatechange.ai/papers/neurips2023/33/paper.pdf

Roy, B., Goodall, J. L., McSpadden, D., Goldenberg, S., & Schram, M. (2023, December). Application of LSTM and seq2seq LSTM surrogate models for forecasting multi-step-ahead nuisance flooding of flood-vulnerable streets in Norfolk, Virginia. In AGU Fall Meeting Abstracts (Vol. 2023, No. 2, pp. H41V-02). 2023AGUFM.H41V..02R

Kumarm C., McSpadden, D., Goldenberg, S., et al. Location Invariant Flood Forecasting using Fourier Neural Operator. Authorea. January 09, 2025. DOI: 10.22541/au.173645444.46575894/v1

Hasan, M. M., Schram, M., Katragadda, S., McSpadden, D., Udomvisawakul, A. N., Richter, H., & Liu, F. (2025). Uncertainty-Aware Deep Learning Framework for Forecasting Coastal Water Level in Virginia Beach. ICLR 2025 Workshop on Tackling Climate Change with Machine Learning. Retrieved from https://www.climatechange.ai/papers/iclr2025/49

Questions

AI-Ready Data Preparation: Sequencing

Periodic Water Level Pattern: Short-term dependencies observed at the experimental site.

Dataset Question:

How Many Timesteps of Look-Back?

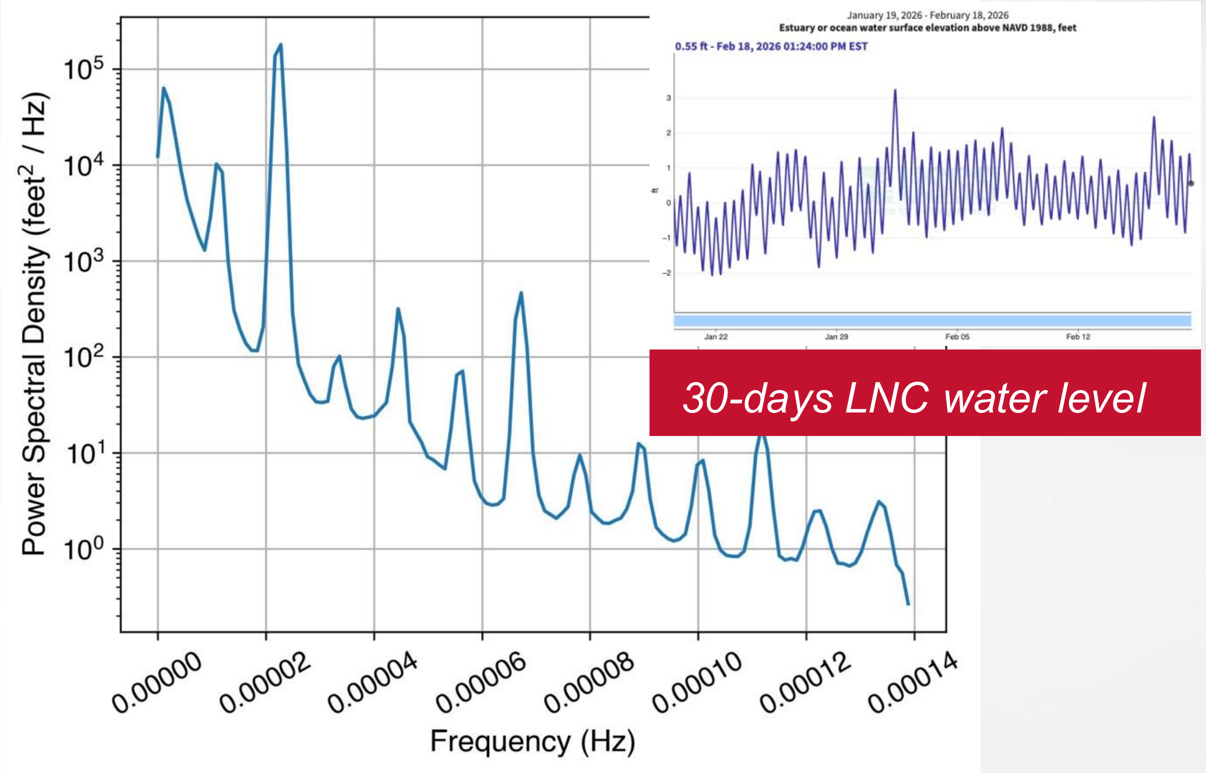
Visualization:

- PSD shows signal power distribution across frequencies.

Key Insight:

- Highest spectral density ≈ 0.00002 Hz → ~12-hour cycle.
- Confirms need for 24-hour historical data window.

Encourage documentation of preparation of shared data, and AI-readiness pipelines.



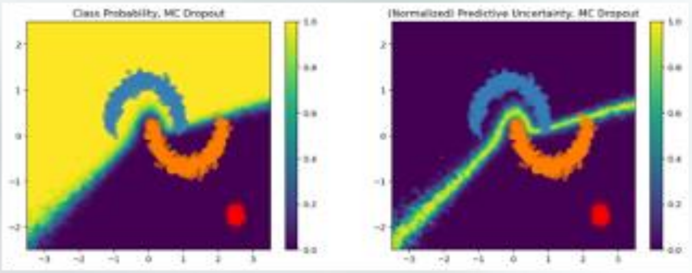
30-days LNC water level

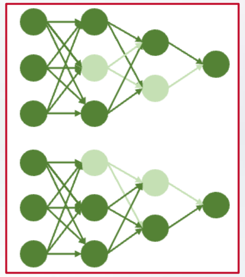
*Frequency analysis of the water level using power spectral density (PSD). **Welch's Method Basics:** Splits signal into overlapping segments; Computes FFT for each segment and averages squared magnitudes.*

Examples of ML Uncertainty Quantification Techniques

Class Probability | **Uncertainty**

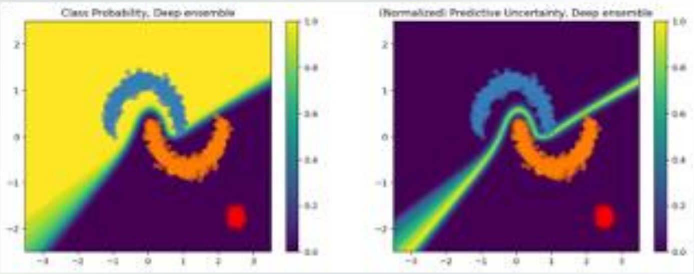
MC Dropout

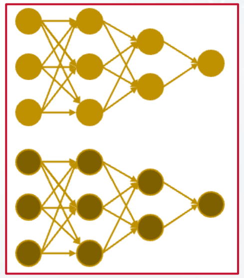




Randomly drops neurons; distribution of predictions

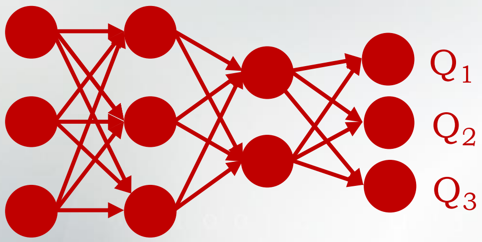
Deep Ensemble





Multiple model initializations; distribution of preds; post UQ calibration

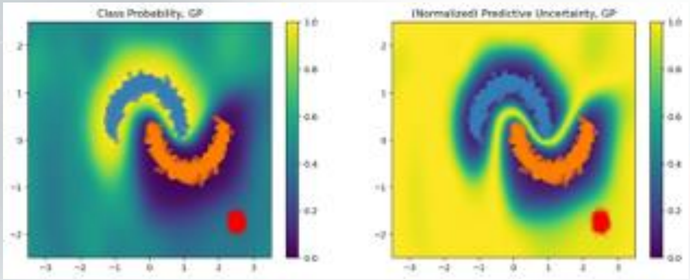
Deep Quantile Regression

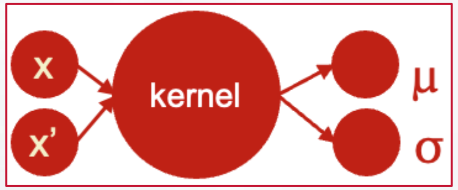


$L_\tau = \max(\tau\cdot(y-\hat{Q}_\tau(y|x)),(\tau-1)\cdot(y-\hat{Q}_\tau(y|x)))$

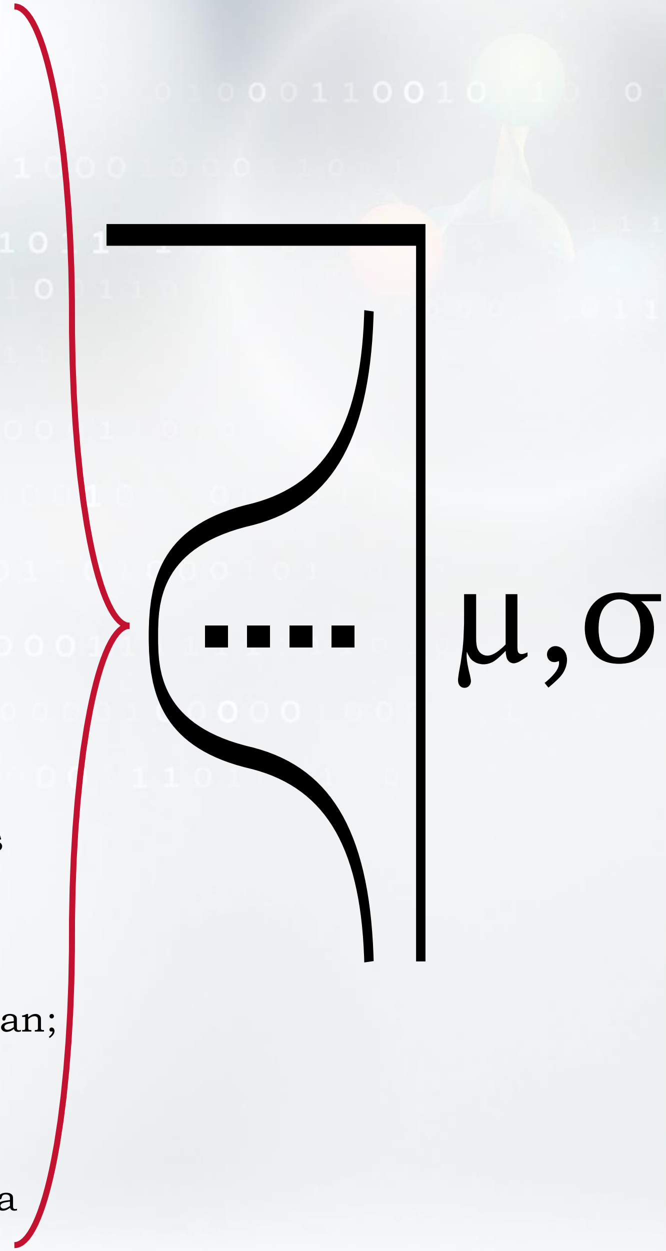
Loss penalizes over- and under-predictions asymmetrically; estimates selected quantiles rather than only the mean.

Gaussian Process



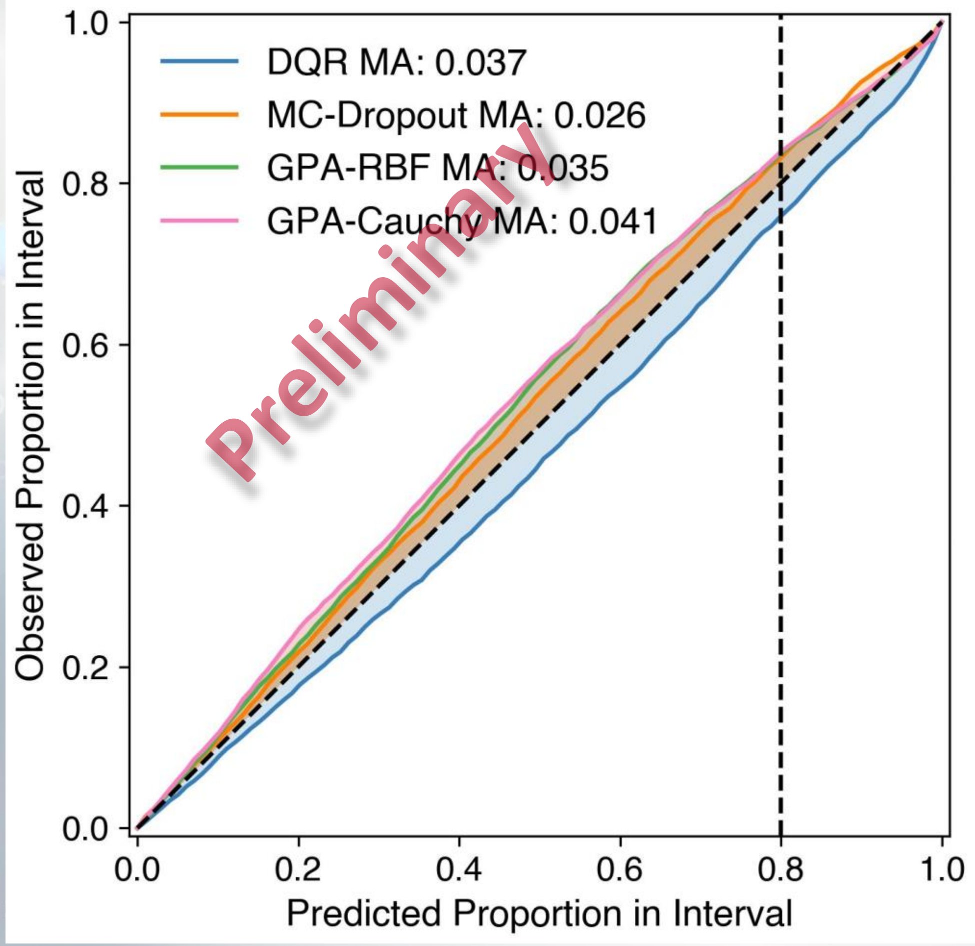


$x \sim N(\mu,\Sigma)$ estimates mean; Covariance function (uncertainty) given the data



Uncertainty & Miscalibration

Example Preliminary In-Distribution Results

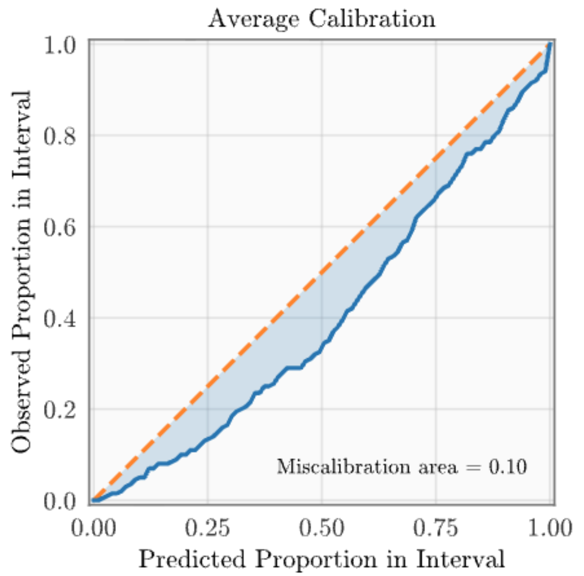


Uncertainty Calibration

- 10% confidence yields correct predictions 10% of the time.
- 90% confidence yields correct predictions 90% of the time.
- Uncertainty Toolbox used for visualization and UQ calibration metrics: https://github.com/uncertainty-toolbox/uncertainty-toolbox

Example toy problem 10% miscalibrated, and overly confident:

When predicting 50% of samples in the interval, only about 32% were observed.



Miscalibration Area

$$p \in P = \{0, 0.01, 0.02, \ldots, 1.0\},$$

In practice, for each nominal coverage probability, p ∈ P = {0, 0.01, 0.02, . . . , 1.0}, we compute the corresponding z_p/2 value that denotes the critical value (quantile) of the standard normal distribution corresponding to a two-sided coverage probability.

To compute Miscalibration Area, we plot p on x-axis against hat_p_p on y-axis to yield what is known as the average calibration plot.

A perfectly calibrated model yields points lying on the diagonal, indicating exact agreement between predicted and observed proportions.

Deviations below the diagonal indicate overconfident predictions, whereas deviations above it indicate underconfident behavior.

The average MA metric measures the area between the ideal diagonal line and the curve with 0 indicating a perfectly calibrated model for all values of p.