Basics of Uncertainty Quantification

Going Beyond Model Accuracy

Ashley S. Dale, Ph.D.

6 Jun 2025

Schmidt Sciences

UNIVERSITY OF TORONTO

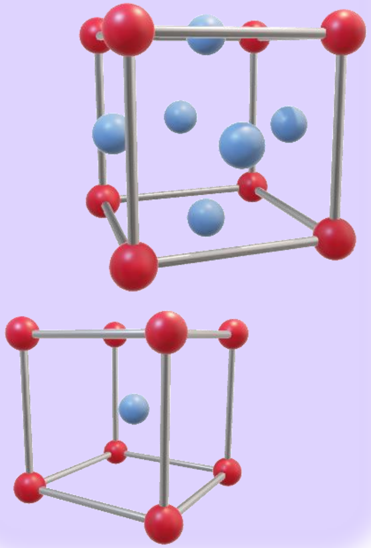
Overview

- Uncertainty 101
- Uncertainty Characterization Workflow
- Case Study
  - Code Demonstration Using Random Forest

Uncertainty 101

Example: Machine Learning for Materials Property Prediction

Chemistry Search Space



Feature Identification

phase

bandgap

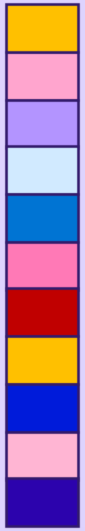
spacegroup

valence electrons

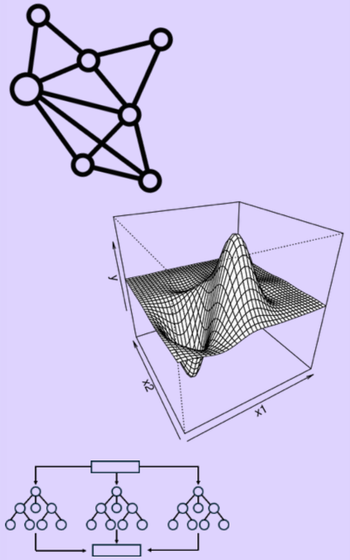
atomic weight

covalent radius

Feature Embedding



Modeling



Each of these tasks may have an uncertainty associated with it

Image Credit: Surrogates: Gaussian Process Modeling, Design and Optimization for the Applied Sciences. R. B. Gramarcy

Do your model parameters have uncertainty? Do your data points?

- Think about these questions for a second.

If your model parameters do not have uncertainty…

- Congratulations! You might be using **frequentist** statistics!
- As a frequentist, you will draw an infinite number of samples from your data distribution
- Those samples will bring your uncertainty to zero because any observed uncertainty is just due to missing data
- **Methods:** *confidence intervals, significance tests, hypothesis testing*

UNCERTAINTY IS A PROPERTY OF THE DATA

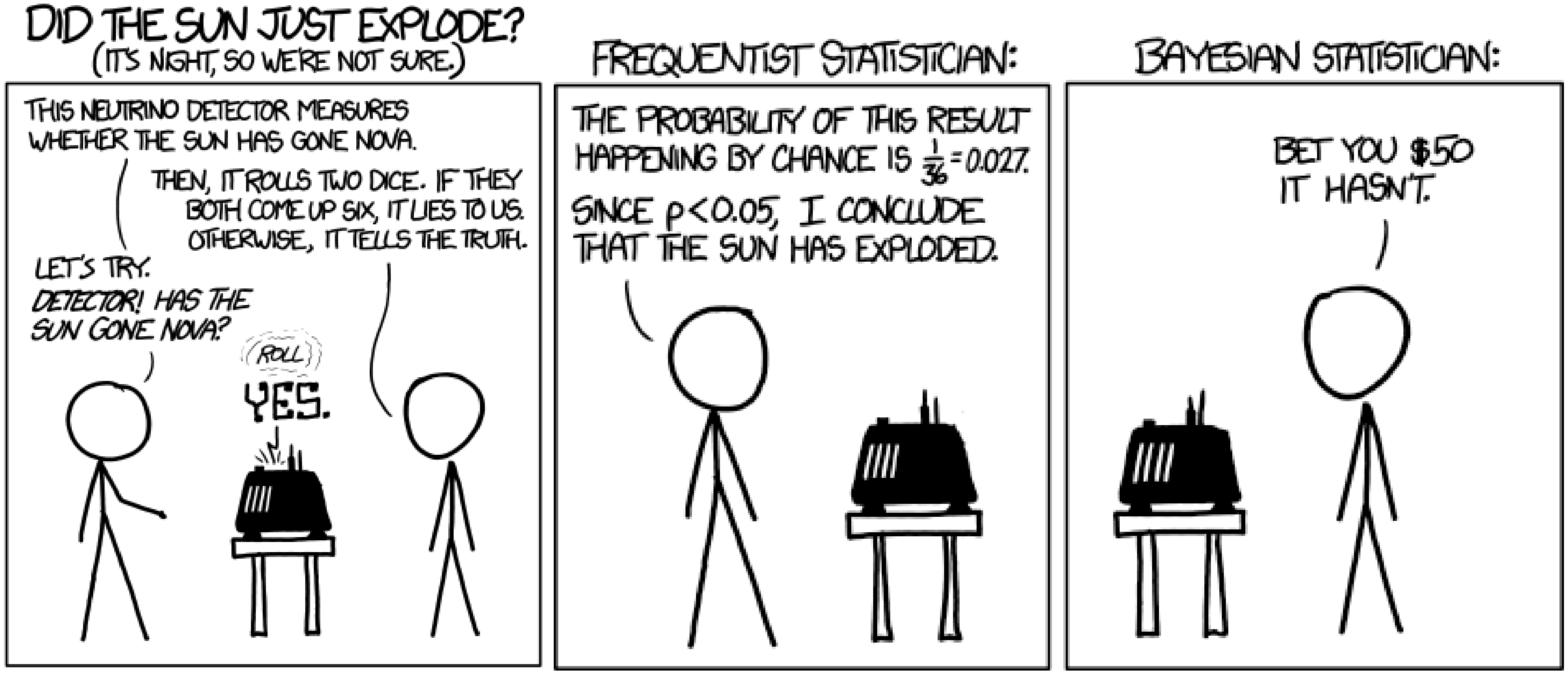
If your model paramters **do** have uncertainty..

- Congratulations! You might be using **Bayesian** statistics!
- As a Bayesian, you will keep increasing your number of samples conditioned on previous knowledge
- Your uncertainty might never be zero, but you are okay with that as long as you know what it is
- **Methods**: *Really good at estimating prior knowledge*

UNCERTAINTY IS A PROPERTY OF THE MODEL

Frequentist & Bayesian Statistics

https://xkcd.com/1132



*If the sun has gone nova, everyone would be dead. Therefore, continued life is the prior P(H), not the dice roll

Likelihood

$L(H;D) = P(D|H)$

Probability

$P_{Pos}(H|D) = P(D|H)P_{Prior}(H)/P(D)$

Controversy @ https://statmodeling.stat.columbia.edu/2012/11/10/16808/

If neither of these feel like a good fit....

CONGRATULATIONS!!!

- You might be doing **machine learning**!
- Machine learning uses ideas from both Bayesian and Frequentist statistics
- Some statisticians might consider this “unprincipled”. Machine learning people just make it work.
- **Methods**: *Everything. Some methods work better than others sometimes.*

<u>**UNCERTAINTY IS EVERYWHERE**</u>

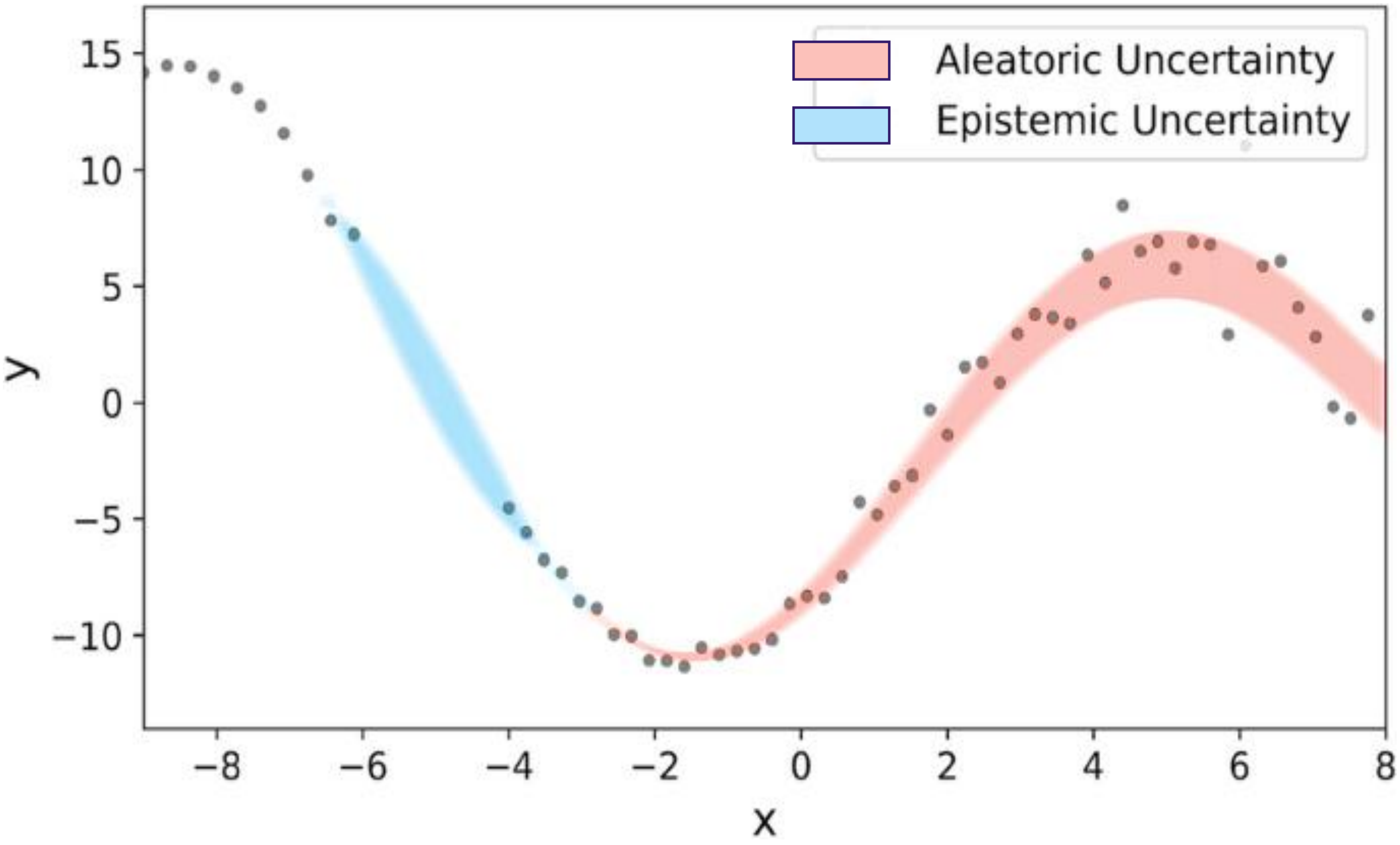
How to think about uncertainty while doing machine learning

- Uncertainty from the data
- Uncertainty from the model parameters
- Uncertainty from the optimization process
- Uncertainty from model choice, hyperparameters

- Uncertainty we can fix
- Uncertainty we can’t fix

- Acceptable uncertainty
- Unacceptable uncertainty

Different Kinds of Uncertainty...



<u>Aleatoric uncertainty</u>: noisy parameters; measures stability

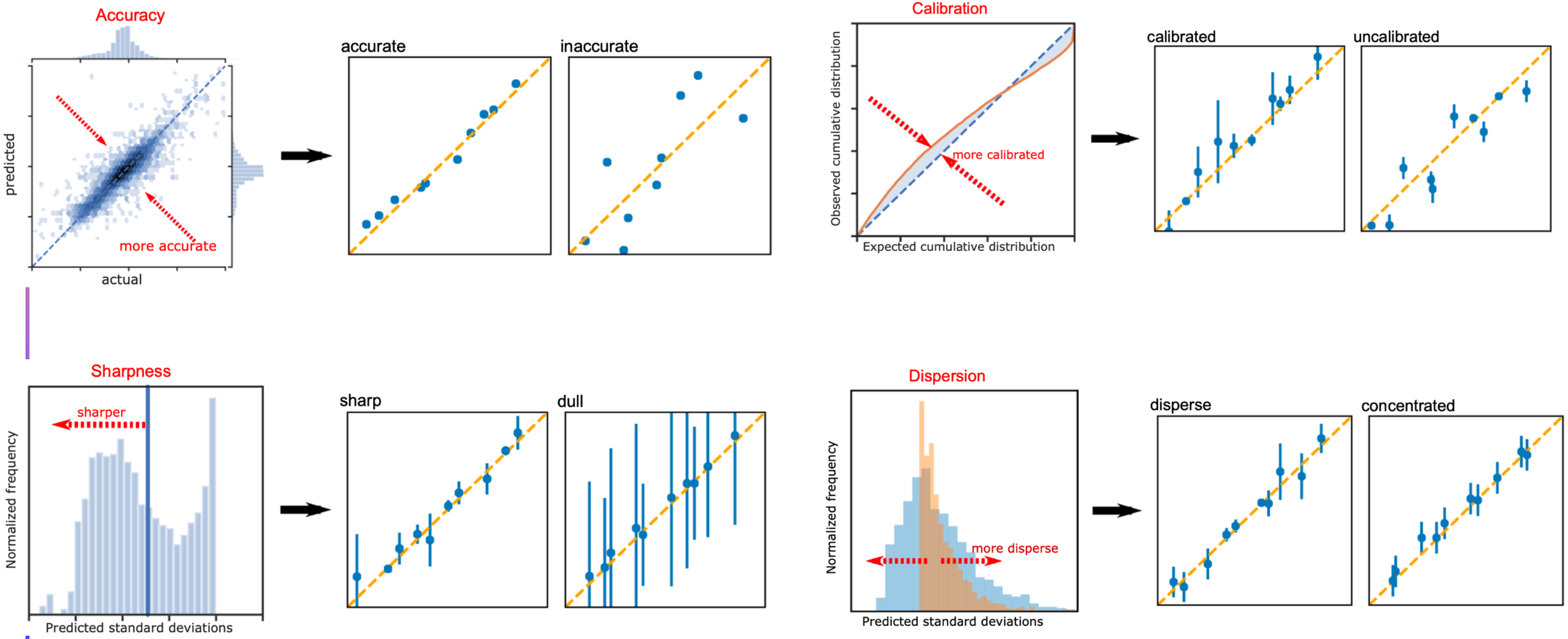
<u>Epistemic uncertainty</u>: missing information; measures robustness

Total Uncertainty = Aleatoric Uncertainty + Epistemic Uncertainty

We would like our model to be robust to both; improving model performance

Yang, Chu-I., and Yi-Pei Li. "Explainable uncertainty quantifications for deep learning-based molecular property prediction." *Journal of Cheminformatics* 15.1 (2023): 13.

...Lead to Different Measurements



Tran, Kevin, et al. "Methods for comparing uncertainty quantifications for material property predictions." *Machine Learning: Science and Technology* 1.2 (2020): 025006.

Uncertainty Characterization Workflow

Uncertainty Characterization

Identify potential sources of uncertainty → Consider where the uncertainty *can* be measured → Quantify Uncertainty → Calibrate Uncertainty → Analyze Uncertainty

Quantifying uncertainty requires the prediction to have an underlying distribution

Not all regression models return this distribution for the prediction

State of the Art for Quantifying Uncertainty

1. Retraining with different hyperparameters (e.g., random seeds, learning rate, etc.) to create an ensemble of models
   - **Pros**: good understanding of the role the hyperparameters play
   - **Cons**: computationally expensive; large search space

2. Using models with intrinsic uncertainty estimates,e.g., Bayesian Neural Nets
   - **Pros:** gold standard
   - **Cons:** very small class of models; also can be computationally expensive

3. Modifying the model after it is trained to provide an uncertainty estimate
   - **Pros:** You can do this to almost any model
   - **Cons**: The model may no longer perform the way you want it to

More Info:

- Tavazza, Francesca, et al. "Uncertainty prediction for machine learning models of material properties." *ACS omega* 6.48 (2021): 32431-32440.
- Tran, Kevin, et al. "Methods for comparing uncertainty quantifications for material property predictions." *Machine Learning: Science and Technology* 1.2 (2020): 025006

Total Variance for a Single Sample

- n : number of predictions
- y : target value
- y_i: i^{th} predicted value

$$\sigma^2 = \frac{1}{n}\sum_{i}^{n}(y - y_i)^2$$

Uncertainty Decomposition Through Variance Conservation

Assumption: variance and uncertainty are proportional

$$U_{Total} = U_{Epistemic} + U_{Aleatoric}$$

Epistemic Uncertainty: Can be reduced if more knowledge is available → explained variability

Aleatoric Uncertainty: Cannot be reduced; it is a property of the model-data system → unexplained variability

We want to obtain estimates for $U_{Epistemic}$ and $U_{Aleatoric}$

https://murraylax.org/rtutorials/regression_anovatable.html

Epistemic Uncertainty

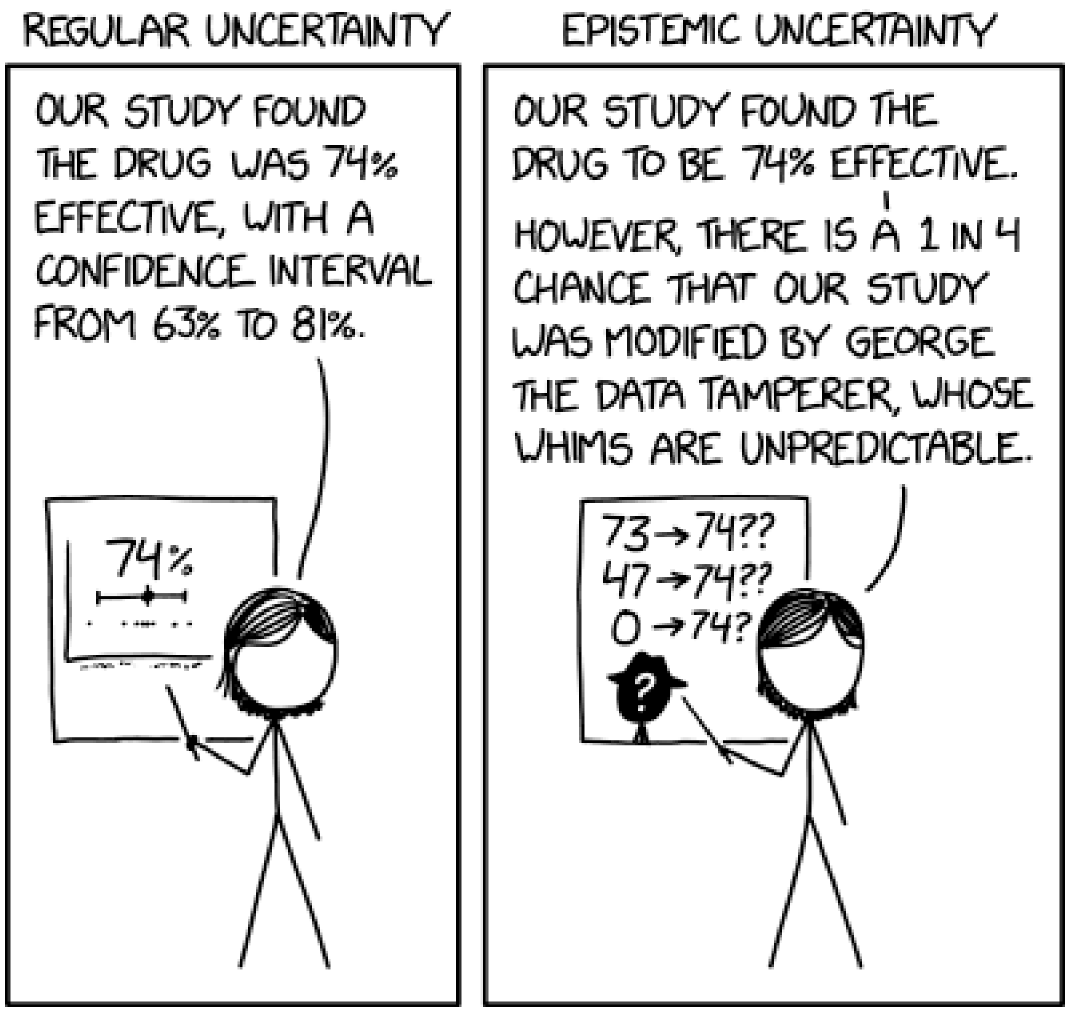
- Epistemic Uncertainty is "good" because it is
  - Explainable
  - Fixable

$$\bar{y} = \frac{1}{n}\sum_{i}^{n}(y_i)^2$$

Mean of predicted values

$$\sigma_{epi}^2 = \frac{1}{n}\sum_{i}^{n}(\bar{y} - y_i)^2$$

Variance of predicted values w.r.t. the mean prediction



https://xkcd.com/2440/

Aleatoric Uncertainty is the Bad Kind

- Aleatoric Uncertainty cannot be minimized
  - Noisy data
  - Stochastic optimization
  - Random seed hyperparameters
  - Has a probability distribution associated with it

- Aleatoric Uncertainty should be constant or increasing given more information

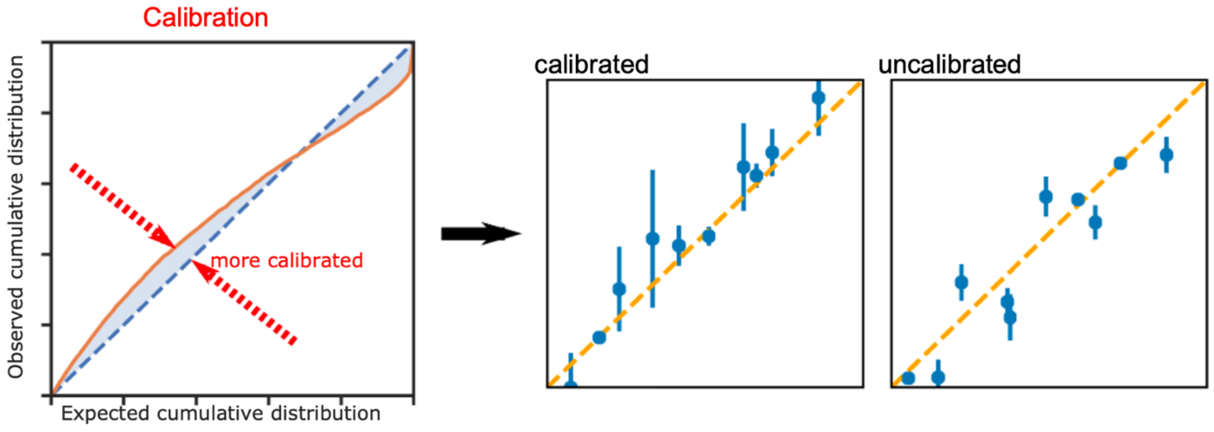
$$\sigma_{al}^2 = \sigma^2 - \sigma_{epi}^2$$

Uncertainty Calibration

Calibration: *The model has learned the correct distribution for predicting values*

- Errobars should overlap with the parity line.
- If error bars do not overlap, then model is not calibrated
- BUT Just because error bars do overlap doesn't mean the model is calibrated

Good error bars are an outcome of good calibration; they don't indicate it



Tran, Kevin, et al. "Methods for comparing uncertainty quantifications for material property predictions." *Machine Learning: Science and Technology* 1.2 (2020): 025006.

Uncertainty Calibration for Regression Models

This is difficult to do correctly; no clear consensus on best practices in the community

- Resource list:
  - https://arize.com/blog-course/what-is-calibration-reliability-curve/
  - https://medium.com/@anuj_shah/model-calibration-for-regression-c9633eab061d
  - **Evaluating and Calibrating Uncertainty Prediction in Regression Tasks**

    by Dan Levi [1,*], Liran Gispan [1], Niv Giladi [1,2] and Ethan Fetaya [3]
- Will present one method in the code, but not in detail and it is not definitive

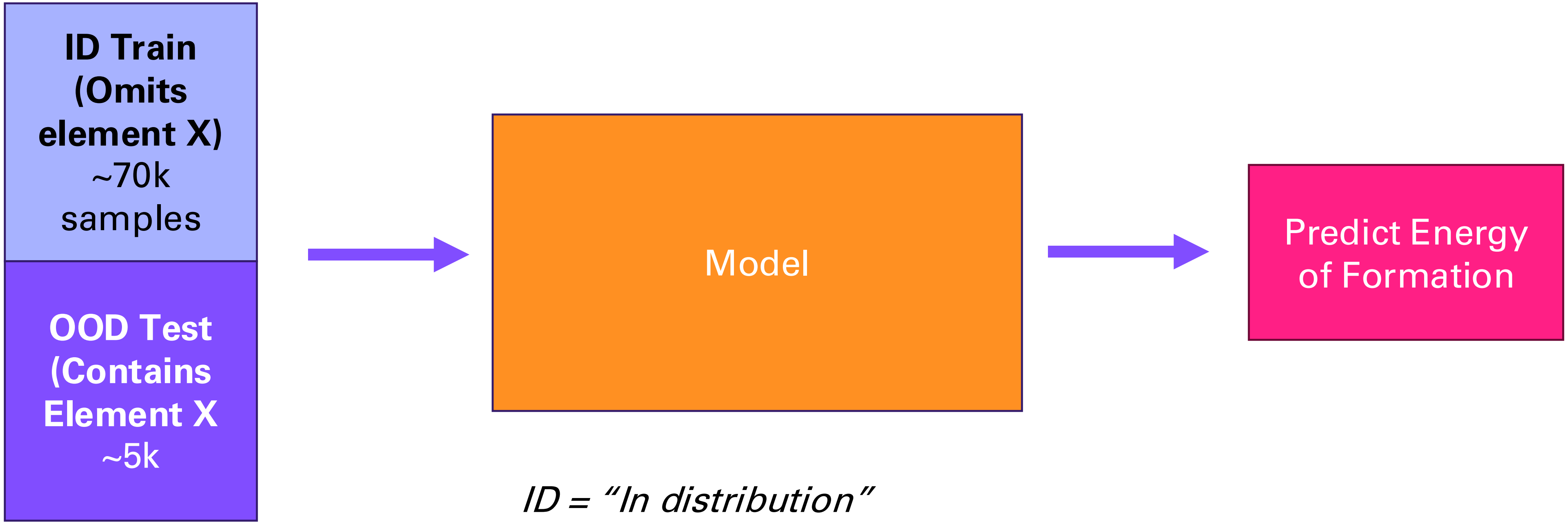
Case Study: Leave-one-element-out for Epistemic Uncertainty

Predict the bandgap for a chemistry containing an element not present in the training data

Li, K., Rubungo, A.N., Lei, X. *et al.* Probing out-of-distribution generalization in machine learning for materials. *Commun Mater* **6**, 9 (2025). https://doi.org/10.1038/s43246-024-00731-w

Setting Up the Experiment

Jarvis 3D DFT Dataset



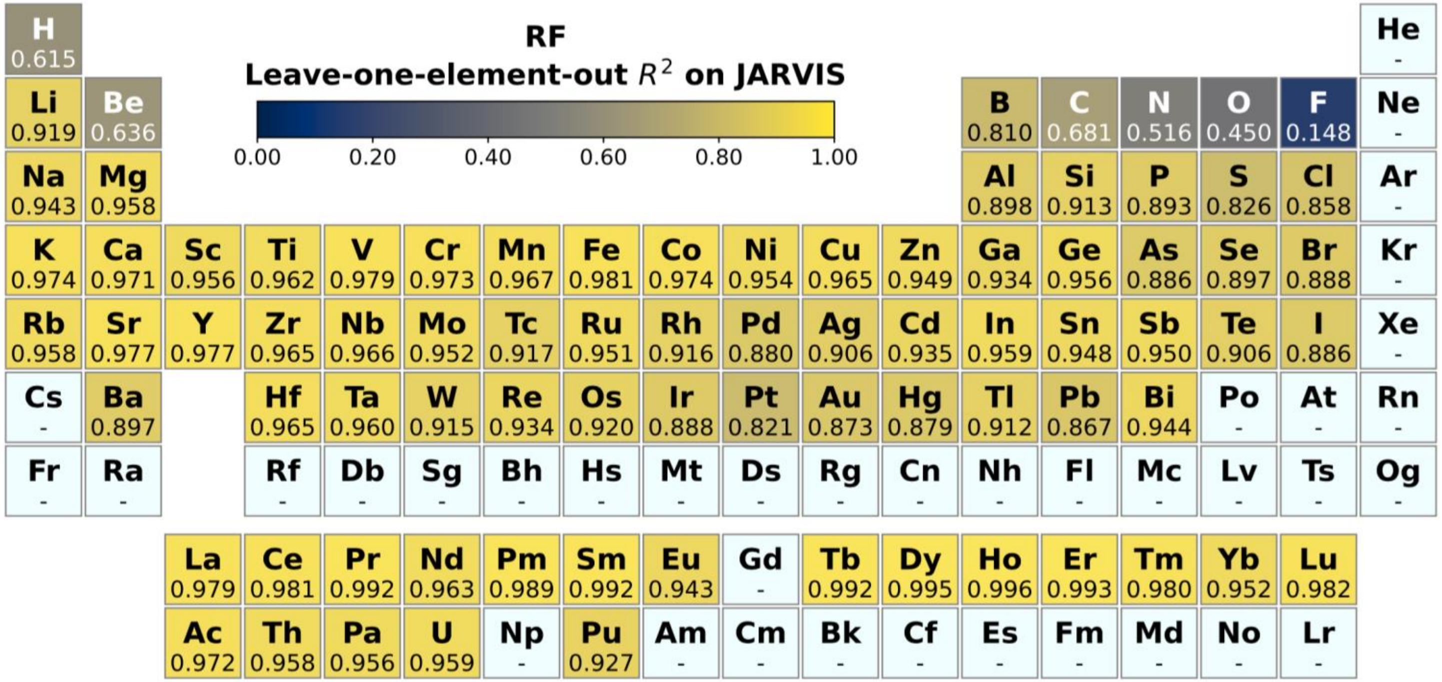
ID = “In distribution”
OOD = “Out of distribution”

Kangming’s Experiment:
Train on data that omits element X from chemistries. Does it generalize to element X?

Model Generalizes Differently Depending on Omitted Element

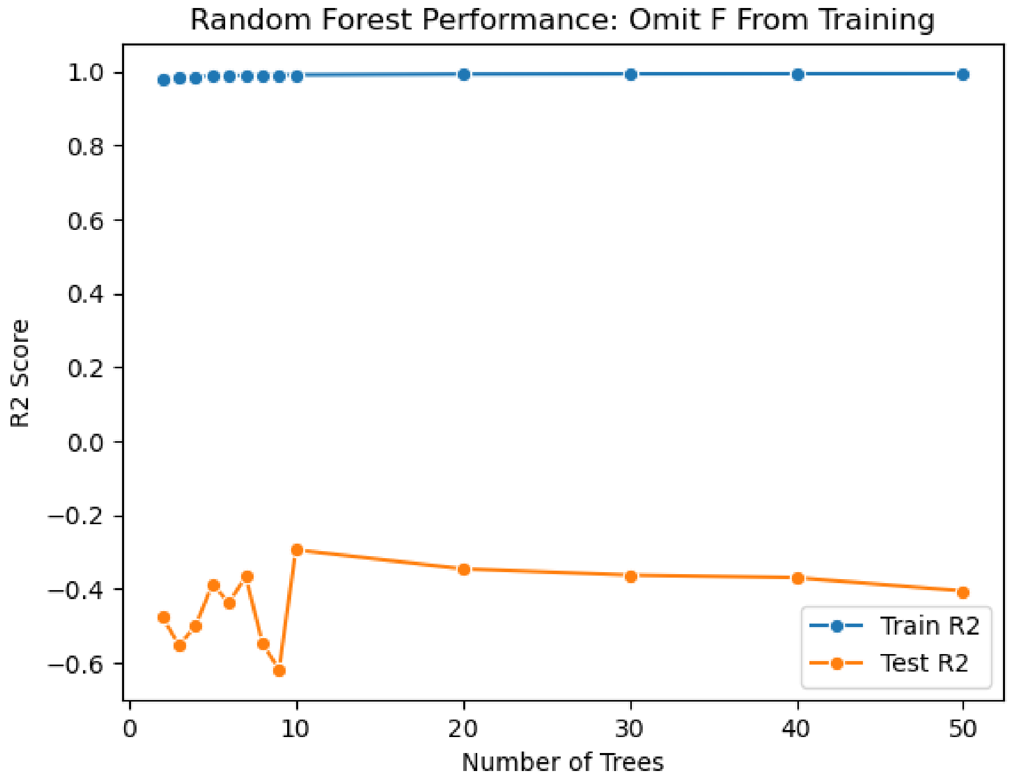
Regression Task: Predict the DFT-calculated **enthalpy** for a chemistry containing an element not present in the training data

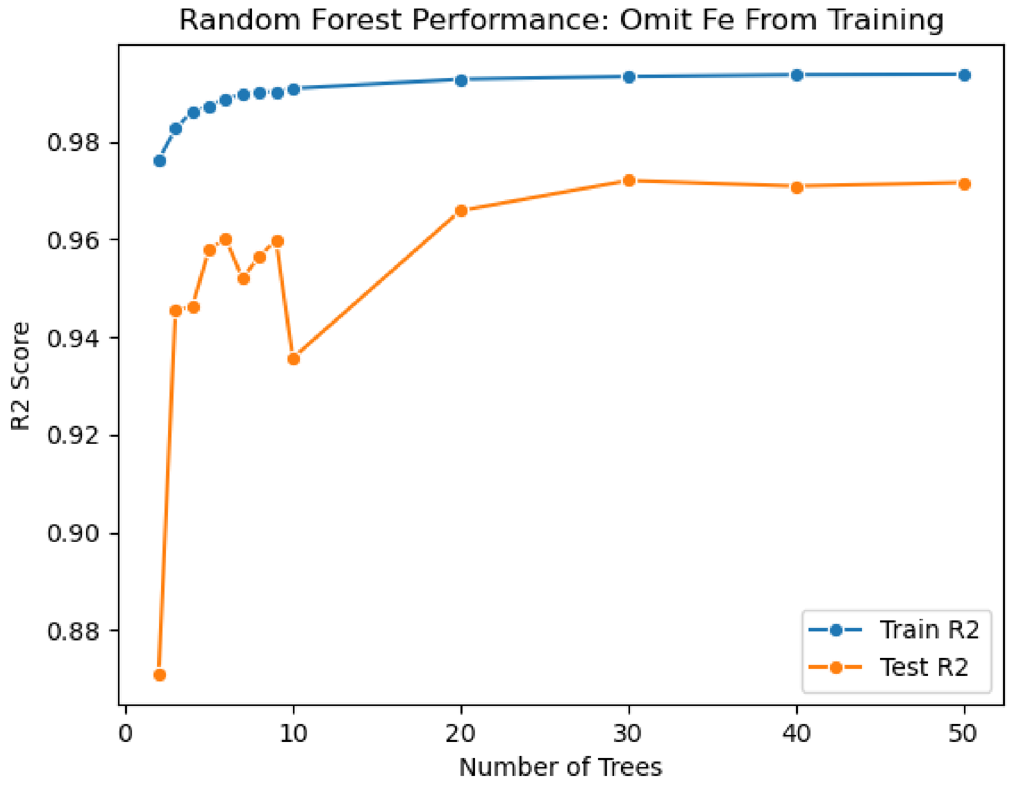
Result: Depends on which element was omitted from the training data



Li, K., Rubungo, A.N., Lei, X. *et al.* Probing out-of-distribution generalization in machine learning for materials. *Commun Mater* **6**, 9 (2025). https://doi.org/10.1038/s43246-024-00731-w

Random Forest Prediction Results





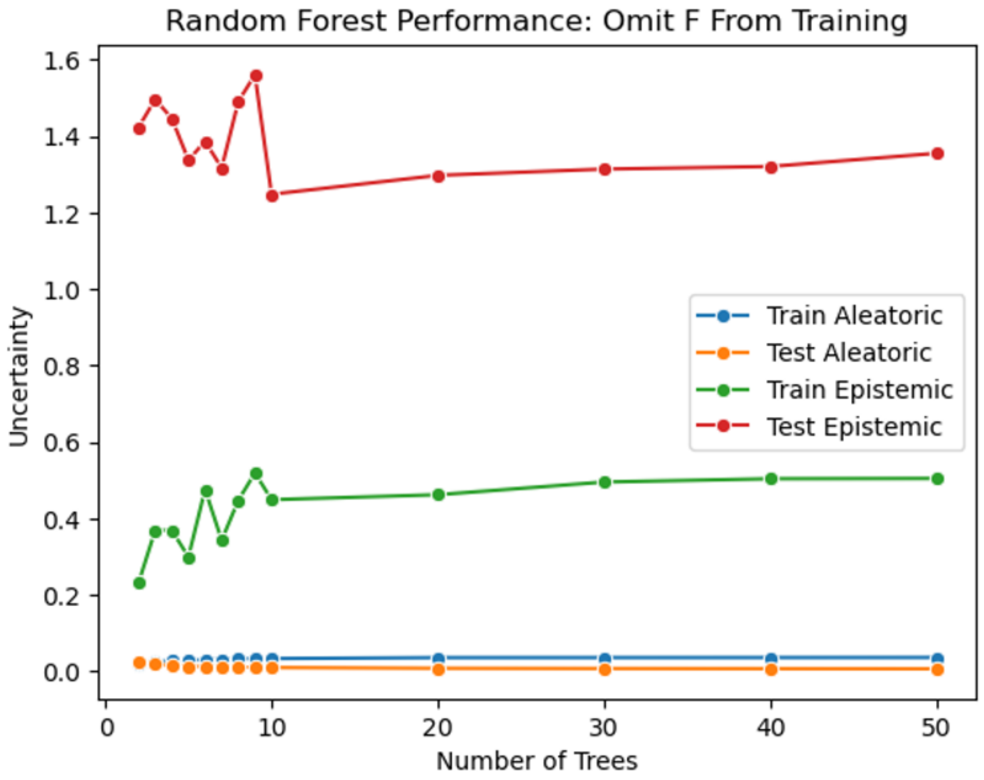
Model does not generalize to F when F is omitted from training data

Likely to have High **Epistemic Uncertainty**

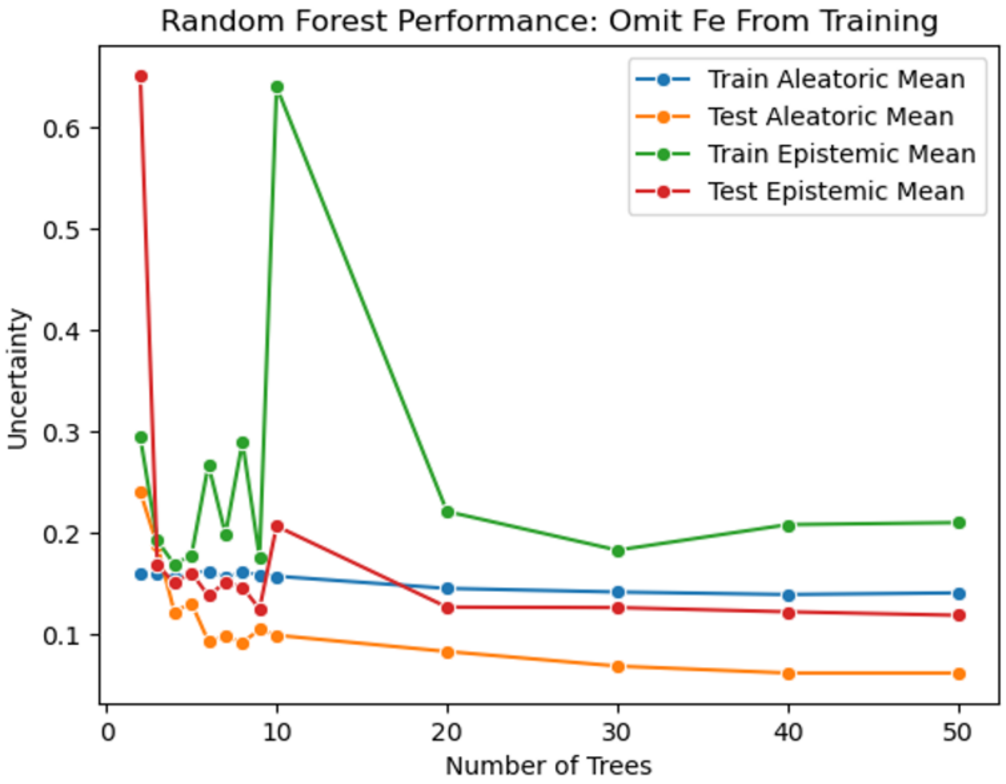
Model generalizes to Fe when Fe is omitted from training data

Likely to have Low **Epistemic Uncertainty**

Random Forest Uncertainty Results

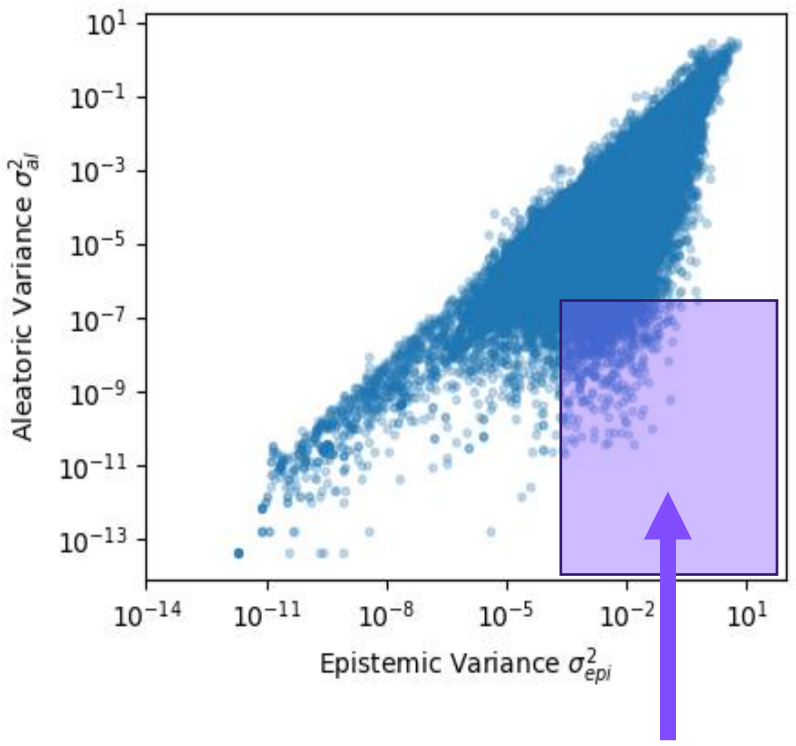


Epistemic uncertainty is high
Aleatoric uncertainty is low and constant



This much messier. Decomposition is unclear.

Deeper Dive: Accelerated Materials Discovery

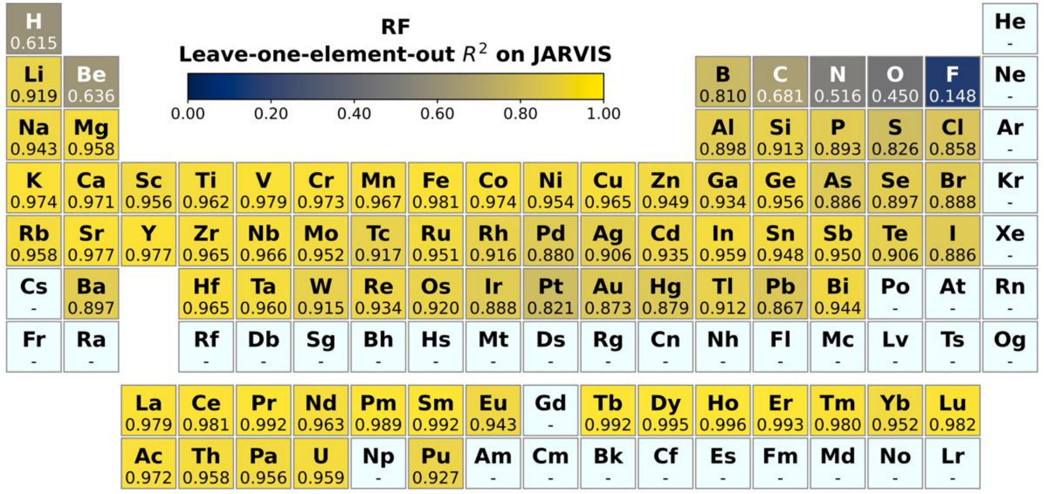


High epistemic uncertainty
Low aleatoric uncertainty

Consider active learning paradigm with Bayesian Optimization:

- Sparse data problem; where to sample next?
- Need to balance exploitation and exploration
- Selecting predictions with high epistemic uncertainty and low aleatoric uncertainty

Coding Exercise



Choose an element X.

Track the change in average uncertainties as the Random Forest Model gets larger.

Do you observe good separation between epistemic and aleatoric uncertainties?

Thank you so much!

Thank you to my collaborators:

- Kangming Li @ KAUST, Brian DeCost @ NIST,

- Jason Hattrick-Simpers and the AUTODIAL group @ U Toronto

Shout out to the following works:

- *Meinshausen, N., & Ridgeway, G. (2006). Quantile regression forests*
- *Kuleshov, et al. (2018) Accurate uncertainties for deep learning using calibrated regression*

Email: ashley.dale@utoronto.ca

Website: https://daleas0120.github.io/