Improving Nuclear Contamination Monitoring Through Microbial Composition

Han Zhang[1], David Bernstein[2], Adam Arkin[1][3]

[1]University of California, Berkeley, [2]University of Vermont,[3]Lawrence Berkeley National Lab

Motivation

- Nuclear processes emit long-lasting nuclear chemicals
- Microbes are sensitive responders to the environment
- Therefore, microbial communities will change compositions in response to these signals
- It's feasible to use machine learning approaches to predict contamination signals from microbial composition
- To improve data quality, we combined field data with high-resolution lab isolate data to improve field predictions with informed data augmentation.

Mission Relevance

- This work will further the development of biology-based methods for the classification and prediction of environmental contaminants, specifically, products of nuclear materials processing.
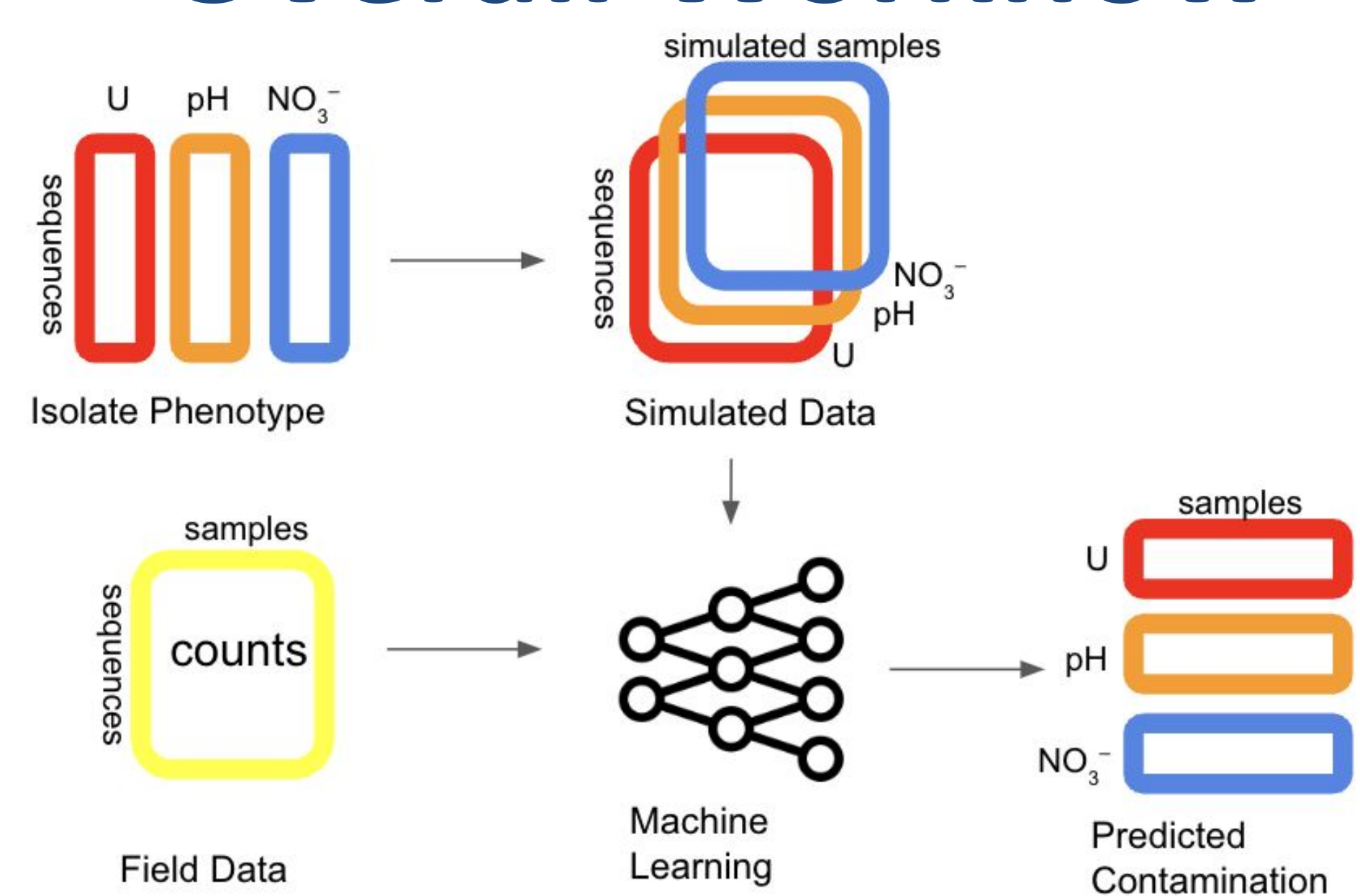
Overall Workflow

Figure 1: Schematic representation of data integration process. Isolate phenotype data measures the growth of each strain across contaminants. This data is used to generate simulated contamination specific operational taxonomic units (OTUs) by samples count matrices. Simulated samples are then combined with field samples to improve machine-learning based prediction of environmental contamination.
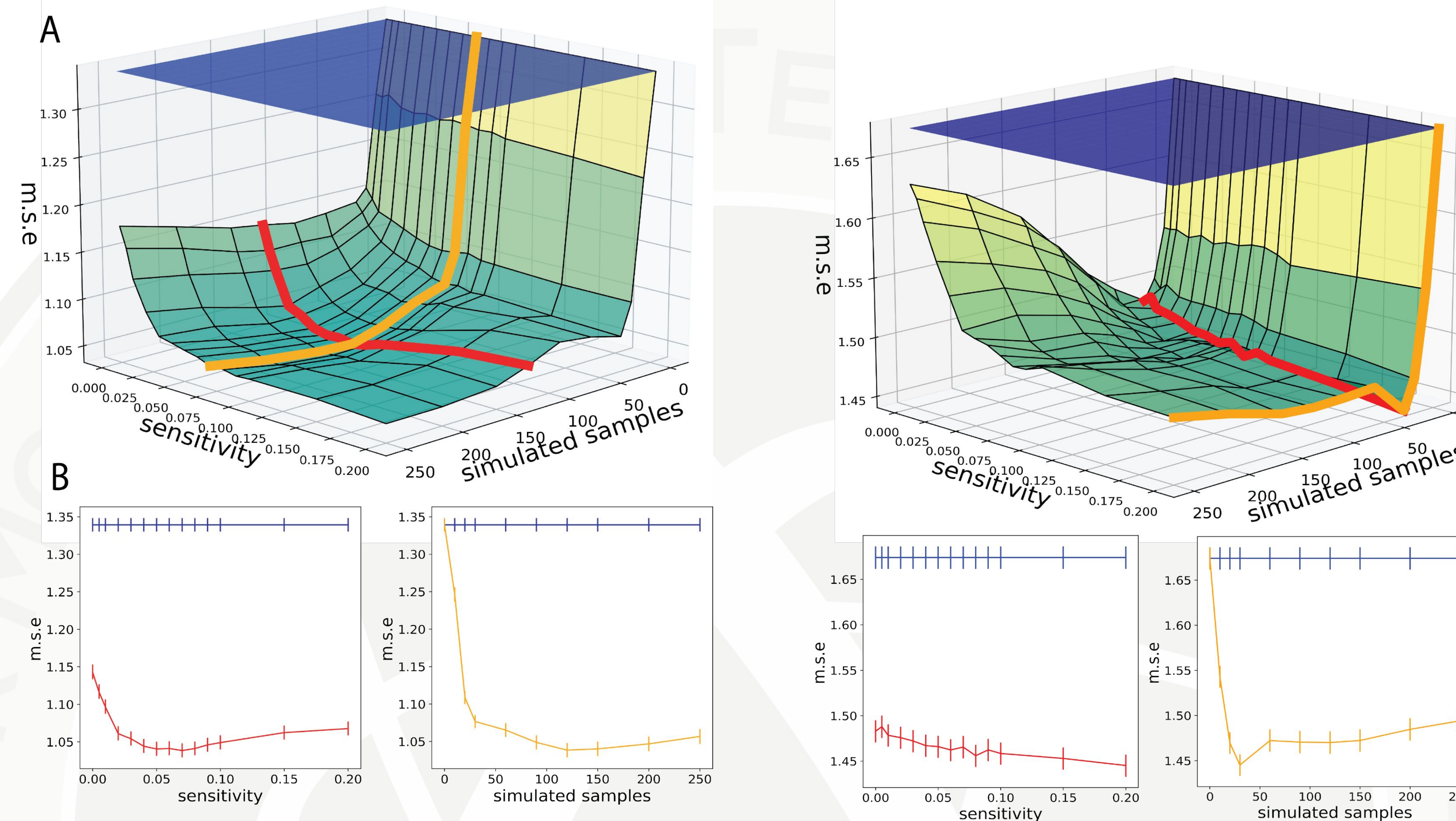
Results

Figure 2: Improving prediction of uranium and pH contamination
(A) Machine learning prediction accuracy (mean squared error MSE) of uranium(left) and pH(right) is plotted as a function of the sensitivity parameter and number of simulated samples. The blue surface shows the baseline accuracy of the machine learning prediction with only the field data. The green surface shows the accuracy with simulated samples that incorporate prior information.
(B) Cross-sections of Fig 2(A) when using 100 simulated samples(left) or 0.1 sensitivity(right).

Technical Approach

Counts are drawn from binomial distribution:
$$W_{i,j} \sim binomial(M_i, Z_{i,j})$$

Z is drawn from a beta distribution (for over-dispersion):
$$Z_{i,j} \sim beta(a_{i,j}, b_{i,j})$$

Beta distribution is parameterized by mean and over-dispersion parameters:
$$\mu_{i,j} = \frac{a_{i,j}}{a_{i,j} + b_{i,j}}$$
$$\phi_{i,j} = \frac{1}{a_{i,j} + b_{i,j} + 1}$$

Beta parameters generated from linear regression with logit link function:
$$g(\mu_{i,j}) = \beta_{0,j} + X_i\beta_j$$
$$h(\phi_{i,j}) = \beta_{0,j}^* + X_i\beta_j^*$$
$$g(c) = h(c) = log\left(\frac{c}{1-c}\right)$$

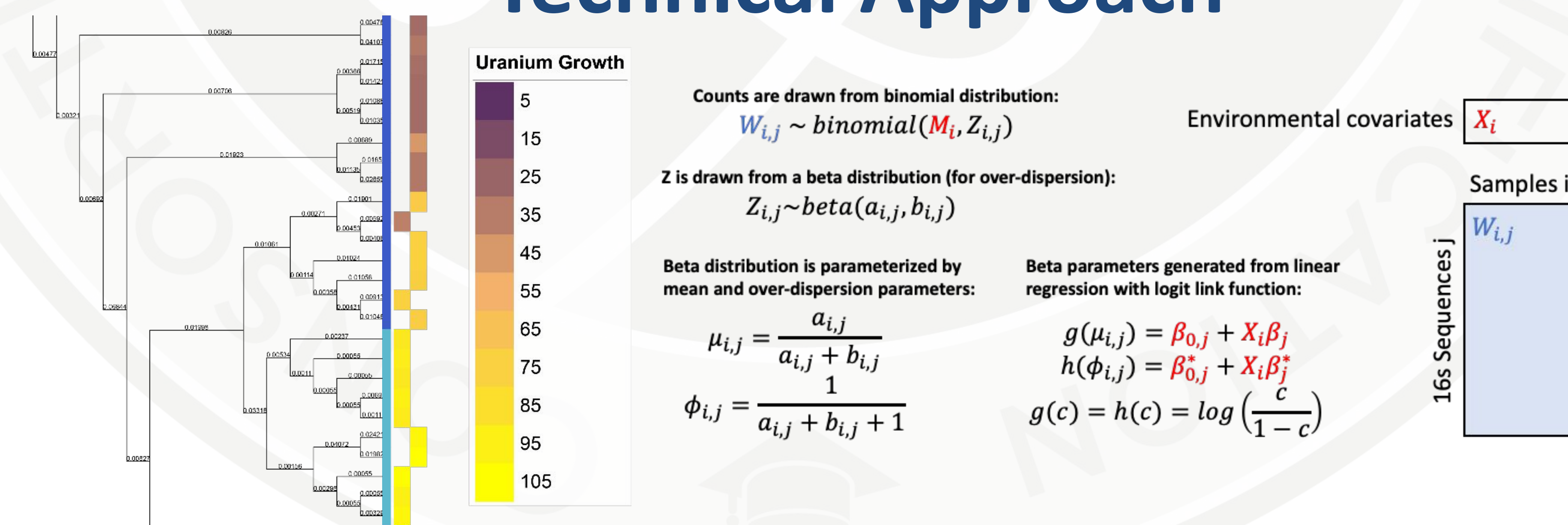Environmental covariates X_i

Samples i
$W_{i,j}$
16s Sequences j

Figure 3: Phylogenetic diffusion of uranium growth data. The inner ring represents isolate growth values, and the outer ring represents the diffused values of the field sequences.

Figure 4: Beta-binomial Regression model to simulate microbial community data. The impact of nuclear contamination on microbial composition is incorporated into the beta sampling process.

MTV Impact

- Hired UC Berkeley undergraduate student through the Undergraduate Research Apprentice Program.
- Collaboration with Hazen (U Tennessee Knoxville, Oak Ridge National Lab), Alm (MIT), and Duff (Savanah River National Lab) labs for microbial community sampling and genomic data analysis.

Expected Impact

- The integration of mechanistic modeling will help us understand complex ecological processes in the context of nuclear contamination and guide monitoring and remediation strategies.

Conclusion

- The predictive power of microbiomes on environmental contamination can be improved by incorporating mechanistic data as prior knowledge.

Next Steps

- Explore the sensitivity of our results to different steps in our pipeline
- Expand cross-validation with additional data

References

- 1.Smith, M. B., Rocha, A. M., Smillie, C. S., Olesen, S. W., Paradis, C., Wu, L., ... & Hazen, T. C. (2015). Natural bacterial communities serve as quantitative geochemical biosensors. MBio, 6(3).
- 2.He, Z., Zhang, P., Wu, L., Rocha, A. M., Tu, Q., Shi, Z., ... & Zhou, J. (2018). Microbial functional gene diversity predicts groundwater contamination and ecosystem functioning. MBio, 9(1).
- 3.Bahram, M., Netherway, T., Frioux, C., Ferretti, P., Coelho, L. P., Geisen, S., ... & Hildebrand, F. (2021). Metagenomic assessment of the global diversity and distribution of bacteria and fungi. Environmental Microbiology, 23(1), 316-326.

National Nuclear Security Administration