# Understanding Exoplanet Habitability: An AI framework for Predicting Atmospheric Absorption Spectra

***VASUDA TREHAN*** (Department of Information Science, University at Albany SUNY)
Dr Kevin Knuth (Department of Physics, University at Albany SUNY)
Michael Way (NASA Goddard Space Flight Center, Institute for Space Studies)

## *Abstract*

*Evolution of space technology, fueled by advancements in Artificial Intelligence and Machine Learning, has transformed our capacity to explore space. The James Webb Space Telescope (JWST) has made this information more accessible, resulting in valuable data. We are working to create an atmospheric absorption spectrum prediction model for exoplanets based on spectral data generated by General Circulation Models developed by NASA's Goddard Institute for Space Studies. Describing an atmospheric absorption spectrum as a set of bins, a model of the spectral bin heights can be generated as a function of the estimated 30 planetary parameters. Spline curves describe the bin heights as a function of the values of the planetary parameters. The main difficulty is the dimensionality of the parameter space, which will include up to 30 planetary parameters. Once the spectrum prediction model is created, Bayesian Adaptive Exploration will identify areas of the planetary parameter space for which more data are needed to constrain the model better. Specifically, parameter values for which the uncertainties in the predicted spectra are largest (highest entropy) are candidates for generating synthetic spectra to improve the system's precision. As more synthetic spectra are generated to train the system, the model's predictive capabilities will increase. This system will be used as a forward model to assist in forming likelihood functions so that the planetary parameters can be inferred given a planet's atmospheric absorption spectrum. This endeavor is expected to contribute to a better understanding of exoplanetary properties and general habitability.*

## *Basic Idea*

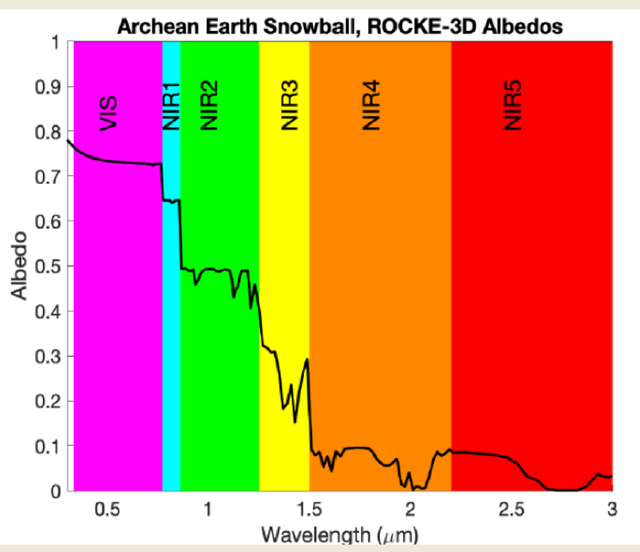
Basic idea is to utilize Measured and Synthetic Atmospheric Spectra sourced from NASA's Climate Models (GCMs) as training data for an AI system. This AI system will the work as a forward model to predict spectra based on a set of approximately 30 planetary parameters.

Further helping us understand the climatic conditions of any planet

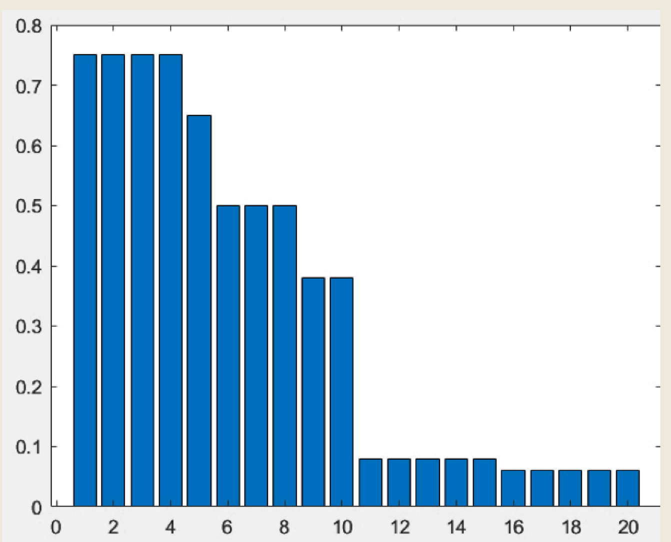
## *Objective*

- Develop an integrated framework utilizing machine learning algorithms.
- Utilize a trained AI system as a forward model to predict spectra from planetary parameters.
- Establish a secondary machine learning system to infer planetary parameters from spectral data.

## *Data*




The spectra in image is a sample of collected data. The spectrum on the left comes from the ROCKE3D simulation. Leftmost being the visible range (VIS) followed by Near InfraRed bands.



As part of the analysis these bands will then be represented as bar graph with bar for each spectral type.

*The spectra in image is a sample of collected data. The spectrum on the left comes from the ROCKE3D simulation.*

## *Approach*

### *Explored Models & Approaches*

- Linear interpolation
- Multivariate regression
- Multivariate interpolation
- Linear Shepherd method

### *Proposed Models & Approaches*

- Spline Curve
  - Cubic Spline Models
- Logarithmic interpolation
- Nested Sampling

## *Methodology*

### Spline Curve

Applied spline interpolation to the 1D synthetic data.

- Utilized a predefined value
- Generate spectra data
- Use spline curve function
- Store the output as y-values
- Visualized the interpolated curve

Mathematical Formula :

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$$

### Logarithmic interpolation

On previous version,

- Implemented logarithmic interpolation formula
- Calculated logarithmic values for splined value
- Visualize the curves

Mathematical Formula :

$$y = y_1 \cdot \left(\frac{x}{x_1}\right)^{\frac{log(\frac{y_2}{y_1})}{\log(\frac{x_2}{x_1})}}$$

### Nested Sampling

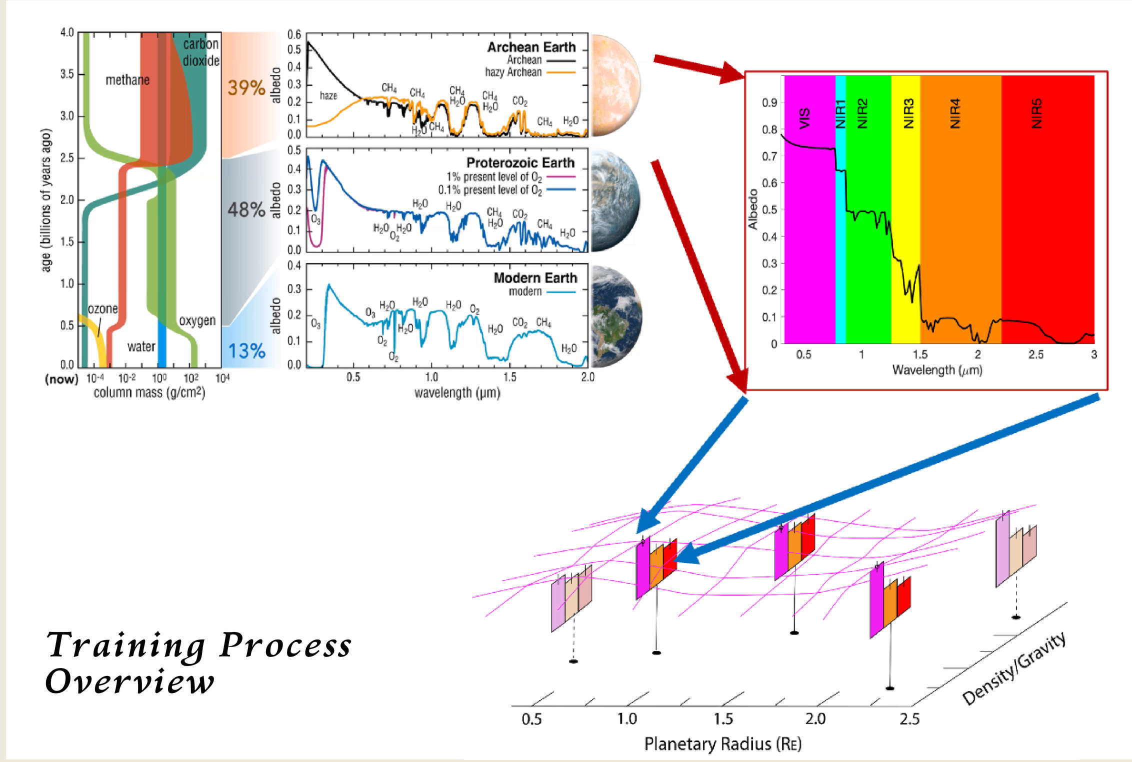
On addition to previous version,

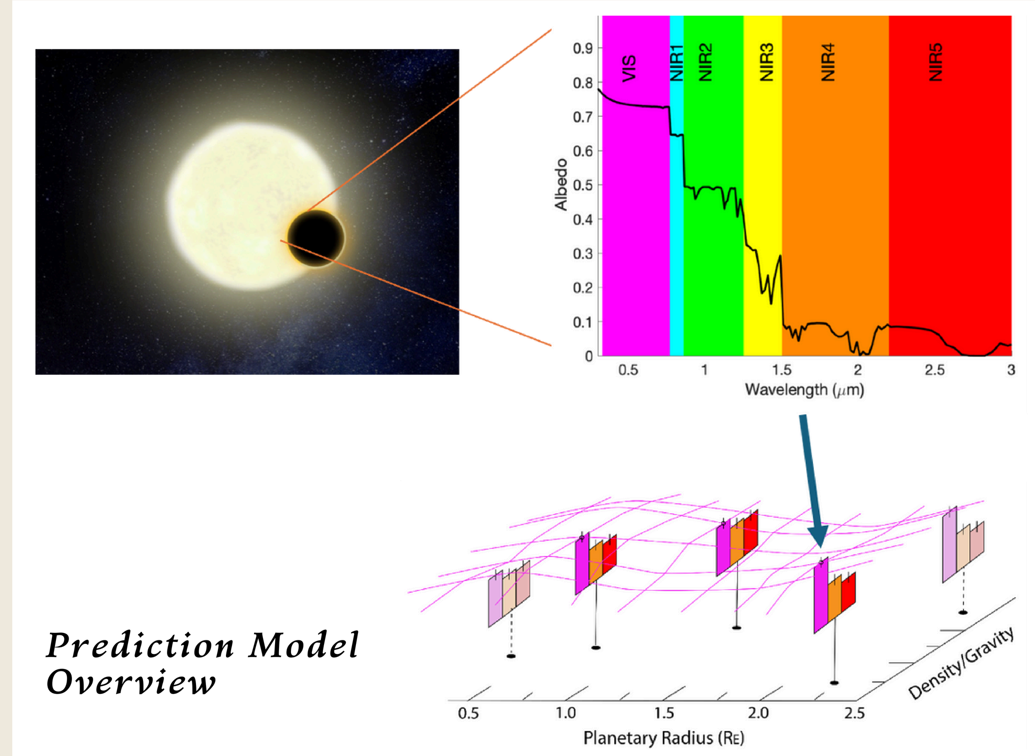
- Nested sampling is used as a Bayesian inference technique.
- Iteratively evaluate
- Implement likelihood function
- Narrowed down parameter space
- Used Bayesian approach to refine parameter

Mathematical Formula :

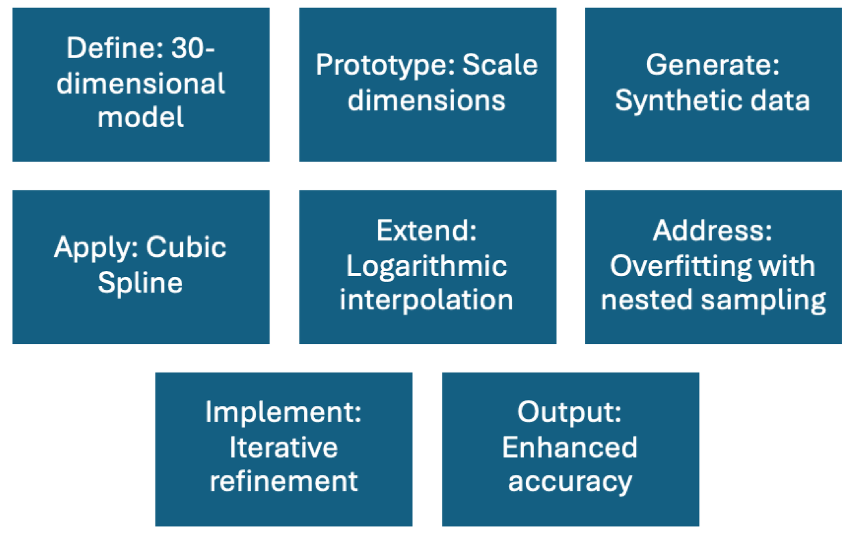
$$L(\theta \mid D, M) = P(D \mid \theta, M)$$

## *Ideology*




## *Designing Model*




## *Data generation*

**Generated synthetic data**

- Initiated five data points – Considered 6 evenly distributed points between [0,1]
- Generated 5 sets of planetary parameters histogram of the 6 evenly distributed values.

**Data generation function** – A combination of polynomial and trigonometric equations to generate synthetic data.

- Provides sample replica of data provided by GISS.

## *Conclusion*

Progress in A.I. and ML Models:

- Significant progress in quantifying planet habitability using A.I. and ML models.
- Challenges lie in accurately factoring and correlating 30+ parameters crucial for habitability assessment.

Methodological Limitations:

- Current approaches effective in one/two-dimensional spaces but face exponential complexity as dimensions increase.

## *Future Work*

- Develop standardized models to analyze and correlate 30+ parameters effectively.
- Decrease the complexity of the machine learning model for higher dimensions of data
- Explore Bayesian Convolutional Neural Networks (BCNN), Gaussian Processes, and deep learning models to overcome complexities and overfitting.

For further details contact : vtrehan@albany.edu

## *Discussion*

**Current Prototype Limitations**

- Limited to one/two-dimensional analysis with five spline knots per dimension (25 splines in two dimensions).

**High Dimensionality Challenge**:

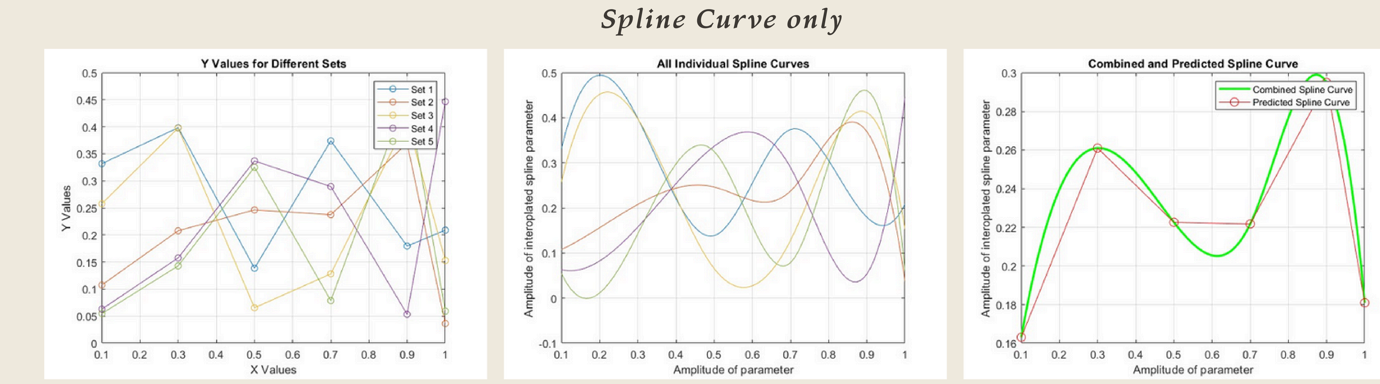
- Scaling to 30 dimensions would require impractically large numbers of splines (10^21), due to unaccounted correlations between spline knots.

**Optimization Goals:**

- Reduce model parameters from billions to thousands by factorization.

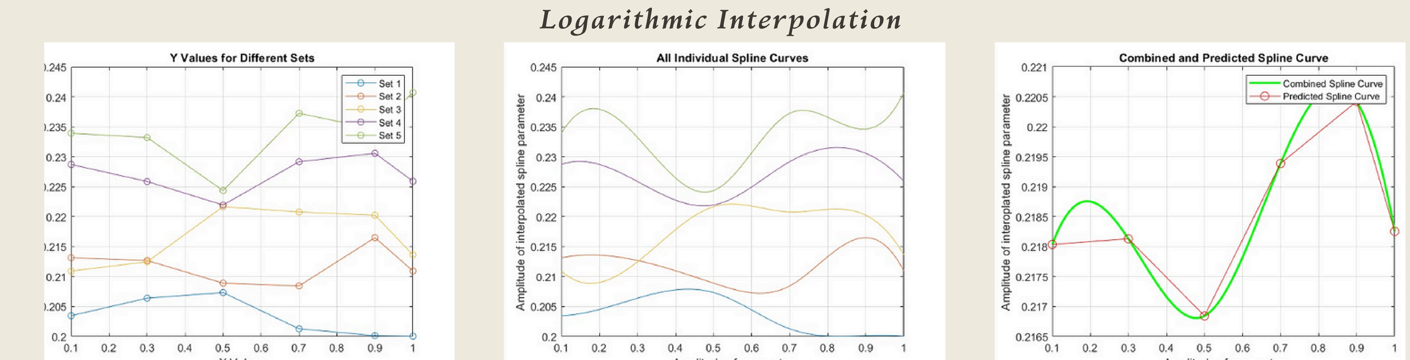
## *Results*

*Spline Curve only*



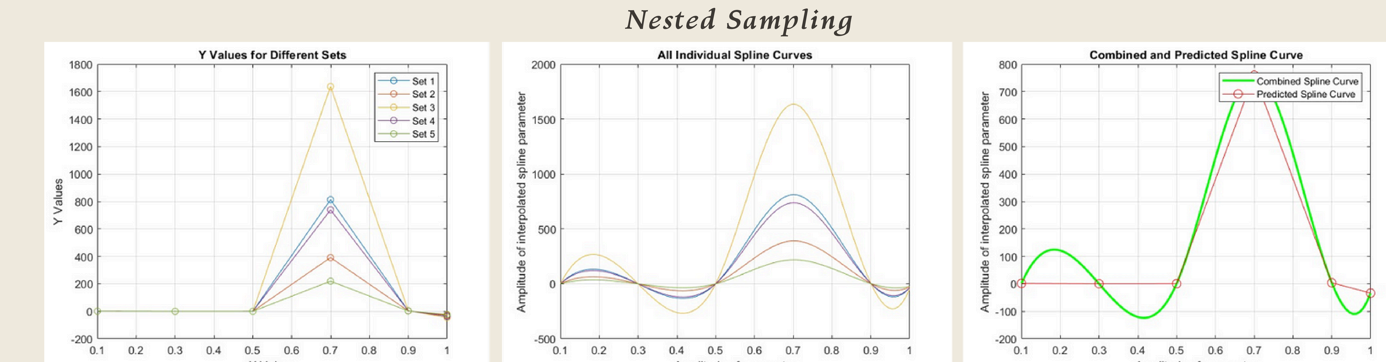

Input | Spline Curve | Evaluation spline curve

*Logarithmic Interpolation*




Input | Spline Curve with interpolation | Evaluation spline curve

*Nested Sampling*




Input | Spline Curve with sampling | Evaluation spline curve

The images represent the 5 sample parameters in consideration that were first passed through nested sampling and interpolated then represented as a spline curve model.

## *References*

Way, M., Ackerman, A., Aleinov, I., Barnes, R., Chandler, M., Colose, C., Fauchez, T., Guzewich, S., Harman, "S".C., Kiang, N., Knuth, K.H., Leboissetier, A. Scharf, C., Sohl, L., Tsigaridis, K., Villanueva, G., Wolf, E. (2023). Habitability Space: Exploring a New Frontier via Climate Models and Planetary Statistics, [ A grant proposal to the NASA Interdisciplinary Consortia for Astrobiology Research].