Exoplanetary surface composition prediction using ML

- Train a model to predict percentage surface composition of exoplanets (mainly terrestrial) from the reflection photometric flux.

- Spectra generated using planetary models (ATMOS & PICASO) and spectral library (USGS, PSG and MODIS).

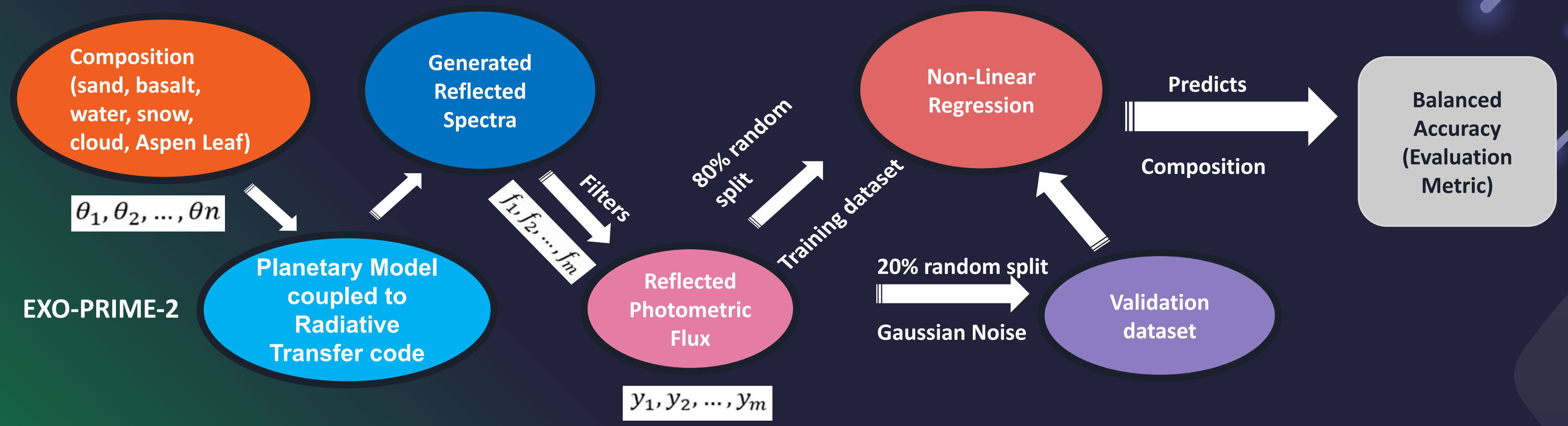
    - Generate the flux value for each filter by convoluting the spectra with the filters' profile.

    - This can help characterize future telescopes for predicting composition using photometric flux and follow up in time-intensive spectroscopic data.

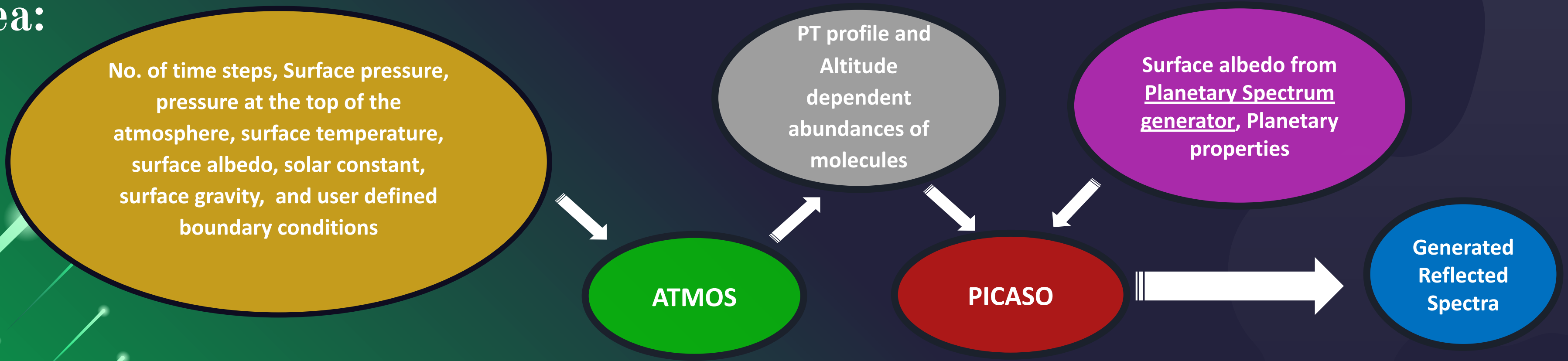
CONTENT

Paper	Model and Algorithms	What the paper does
Pham & Kaltenegger, 2022: Follow the water: finding water, snow, and clouds on terrestrial exoplanets with photometry and machine learning	Atmospheric Model ● ExoPrime-2 ML Algorithms ● XG Boost Albedos Libraries ● **USGS Spectral Library**	● Creates and analyse a grid of 53,130 reflection spectra of Earth-like planets for varying surface compositions and cloud coverage. ● Performs binary classification and predicts on the presence of snow, clouds, and water on the surface ● Found Five optimal filters (Feature importance ranking) ● Performs mock Bayesian analysis & MCMC with the identified five filters to retreive exact surface compositions
Pham & Kaltenegger, 2021: Color classification of Earth-like planets with machine learning	Atmospheric Model ● ExoPrime-2 ML Algorithms ● LDA, KNN, CART, LR, NB, SVM, RF, Majority voting classifier Albedo Libraries ● **ECOSTRESS – ASTER** ● **USGS Spectral Library** ● **The Colour catalogue of life**	● Analyzes a grid of 318,780 reflection spectra of hypothetical planets with different surface compositions and cloud coverage ● Uses six diverse biota samples, including vegetation, biofilm, and UV radiation resistant biota. ● Focuses on the Johnson B, V, R, I filters ● Uses 8 machine-learning algorithms to classify the existence (binary classification) and fraction of biota (multi-classification) in exoplanet spectra with varying S/N ratios.

Literature Review and Project Idea:



Idea:



Atmos: A coupled climate-photochemical model

Photochemical Model

- Generates an initial atmospheric state
- **User-specified boundary conditions:**

  gas mixing ratios or fluxes and deposition velocities, the stellar spectrum, the total atmospheric pressure, the initial temperature-pressure profile

- 233 chemical reactions and includes 50 chemical species, 9 of which are short-lived
- Output:

  altitude-dependent abundances of H_2O photochemically produced in, or transported to, the stratosphere, CO_2, O_3, CH_4, O_2, N_2, and C_2H_6.

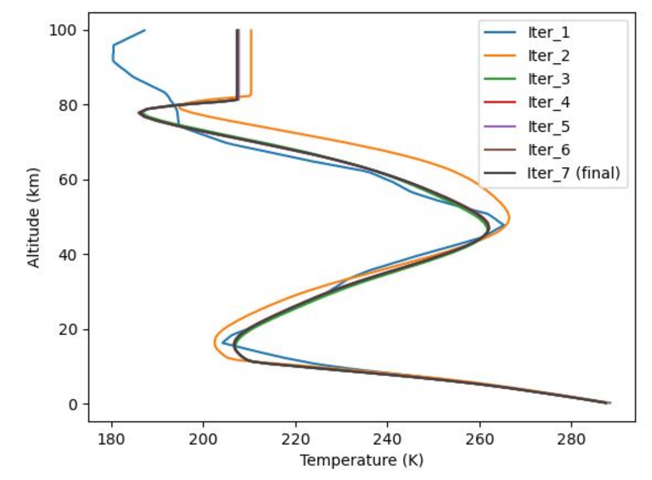
Climate Model

- The tropospheric temp. calculated by following a wet adiabatic lapse rate to the altitude at which the stratospheric temperature is reached
- **Input:**

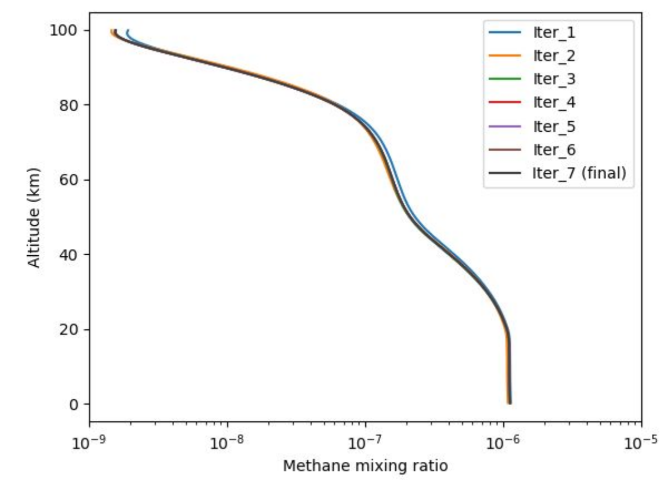
  the number of steps to run the model, pressure at the surface, pressure at the top of the atmosphere, surface temperature, surface albedo, solar constant and surface gravity.

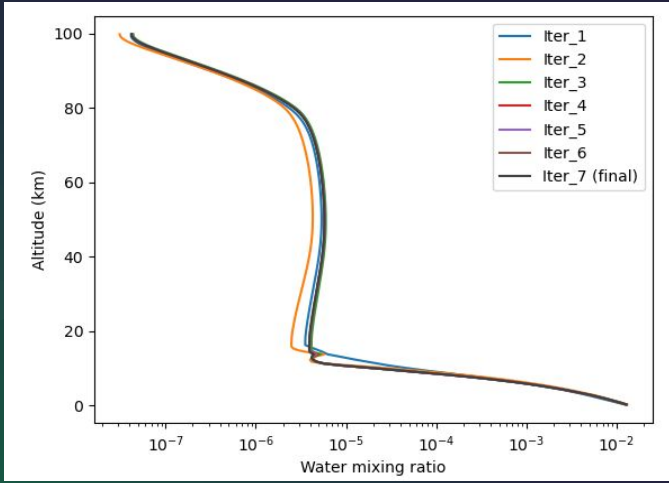
- **Output:**

  altitude, temperature, water mixing ratio



Plots showing Atmos finding Convergence for Modern Earth conditions





PICASO : A Radiative Transfer Model

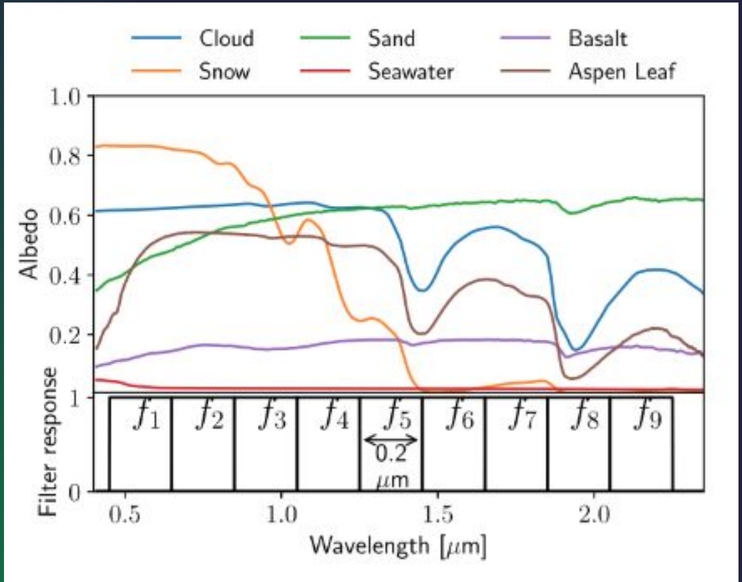
- PICASO is an atmospheric radiative transfer model to produce the reflection spectra. The original documentation was used as reference for using PICASO in our codes.
- PICASO can be used for obtaining transmission, emission and reflection spectra.
- Using PICASO for obtaining the reflection spectra of exoplanets with a certain wavelength dependent albedo function for its surface components.
- The main equation used in PICASO is the radiative transfer equation given below,

PICASO takes the following inputs:

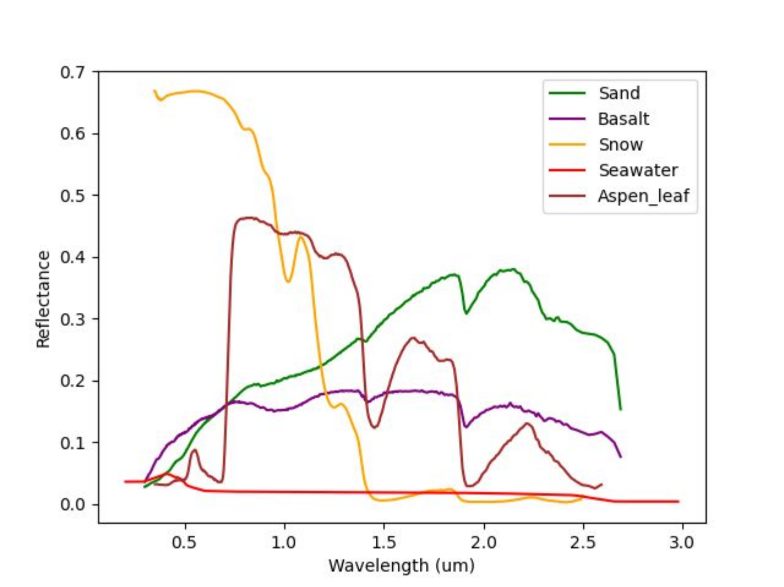
- Basic planetary properties (planet mass, radius, stellar spectra)
- PT profile and abundances (Obtained from ATMOS)
- Cloud profile (angle-scattering albedo, asymmetry and total extinction) (not used in our case).
  Instead albedo of the cloud is obtained from a model and is used along with other albedo functions for finding effective albedo.
- Surface albedo: can be average surface albedo or wavelength dependent surface albedo (in our case).

$$\boxed{I\left(\tau_i, \mu\right) = I\left(\tau_{i+1}, \mu\right) e^{\delta\tau_i/\mu} - \int_0^{\delta\tau_i} S\left(\tau', \mu\right) e^{-\tau'/\mu} d\tau'/\mu}$$

Comparing the albedos reference paper used and the same obtained from PSG



Albedo used by Pham & Kaltenegger, 2022 for different components.
Source: USGS and ASTER library



Albedos obtained from PSG

Data- Set Description

- The data set provided in the paper which we will be trying to reproduce with PICASO and ATMOS consists of reflection spectra for various surface combinations.
- Filters are applied on this spectra to obtain the photometric flux. In our case we took nine photometric flux values which was found to provide us with a good accuracy rate.

- Six surfaces were considered which are Snow, sand, basalt, cloud, vegetation and sea water
- Permutations with 5 percent steps for these six surfaces lead to a total of 53,130 different surfaces.
- The data is then divided into training and validation data with an 80-20 ratio.

Surface combinations

	cloud	snow	sand	seawater	basalt	veg
0	0.00	0.00	0.00	0.00	0.00	1.00
1	0.00	0.00	0.00	0.00	0.05	0.95
2	0.00	0.00	0.00	0.00	0.10	0.90
3	0.00	0.00	0.00	0.00	0.15	0.85
4	0.00	0.00	0.00	0.00	0.20	0.80
...	...	...	...	...	...	...
53125	0.95	0.00	0.00	0.00	0.05	0.00
53126	0.95	0.00	0.00	0.05	0.00	0.00
53127	0.95	0.00	0.05	0.00	0.00	0.00
53128	0.95	0.05	0.00	0.00	0.00	0.00
53129	1.00	0.00	0.00	0.00	0.00	0.00

Flux values for the surface combinations

	f1	f2	f3	f4	f5	f6	f7	f8	f9
0	66.450295	70.559679	60.551194	34.801215	13.230524	8.823439	6.337946	0.350842	1.599316
1	67.564830	69.209650	58.643643	33.787692	12.898762	8.762820	6.250520	0.381013	1.635478
2	68.679366	67.859620	56.736091	32.774169	12.567000	8.702201	6.163094	0.411184	1.671641
3	69.793902	66.509591	54.828539	31.760645	12.235237	8.641582	6.075669	0.441354	1.707803
4	70.908438	65.159561	52.920988	30.747122	11.903475	8.580963	5.988243	0.471525	1.743966
...	...	...	...	...	...	...	...	...	...
53125	234.117912	142.502061	90.005522	58.435347	33.283890	21.958330	16.991558	4.328158	5.985801
53126	232.629441	141.027892	89.131071	57.823231	33.000671	21.622128	16.787464	4.287007	5.882511
53127	238.863237	146.362171	92.777791	60.235069	34.058652	22.909336	17.573305	4.480611	6.353865
53128	245.008517	149.283036	93.169635	59.558788	33.361182	21.608670	16.806243	4.282940	5.877620
53129	241.769328	147.709586	93.563699	60.746116	34.688553	22.713449	17.644302	4.505732	6.178602

Implementing ML Algorithms

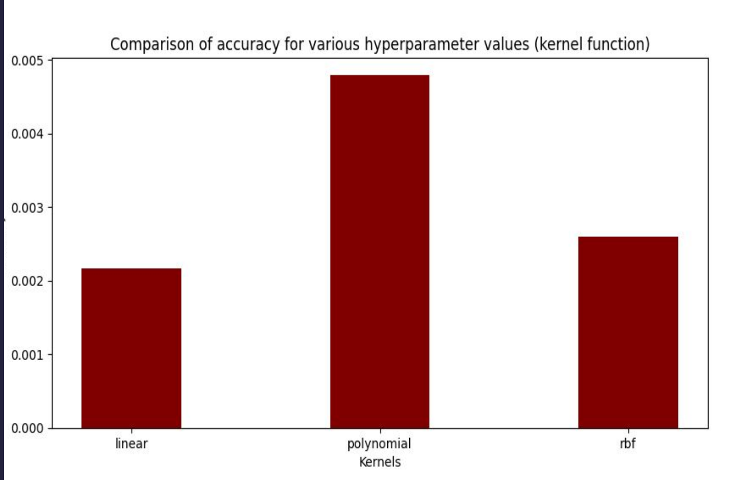
We implemented both SVR and Random forest with the following hyperparameters

- For SVR we used linear kernel function which provided the best accuracy
- For random forest the number of estimators was the hyperparameter which was set to 10 in our case

We used an 80-20 split for training and testing data.
The mean squared error in both the algorithms are,

- SVR: 0.002024
- Random forest: 0.00235

Pham & Kaltenegger, 2022 obtained a balanced accuracy of around unity for snow and cloud and 0.7 for water at S/N=100.



Statistics

- We compared how SVR performed for various kernels (SVR, Polynomial, RBF)
- We found that Linear kernel function provides the best accuracy for this problem

WORK DONE SO FAR

- Literature Review
- Managed to converge ATMOS, where we faced difficulty earlier
- Prepared the code to generate our new data-set after coupling the models (PSG, ATMOS and PICASO)
- Implemented preliminary non-linear regression algorithms (SVR and RF) on the available data-set of the paper.
- Compared the accuracies of both the algorithms with paper.
- Learned about Neural Network and will apply MLP when complete data is generated

FUTURE PLANS

Obtain albedo for cloud and data-set will be complete

Comparing spectra from paper with the one we generate

Augment noise, Training and Validation using SVR, RF and MLP

Surface Composition Prediction