# Optimizing Earth Science Observations: Developing Reinforcement Learning Techniques for Autonomously Determining Priority Observations in a Dynamic Environment

Collective Design Lab

Shashwat Raj, Computer Systems Engineering

Mentor: Paul Grogan, *Associate Professor*

*School of Computing and Augmented Intelligence, Fulton School of Engineering*

## Introduction

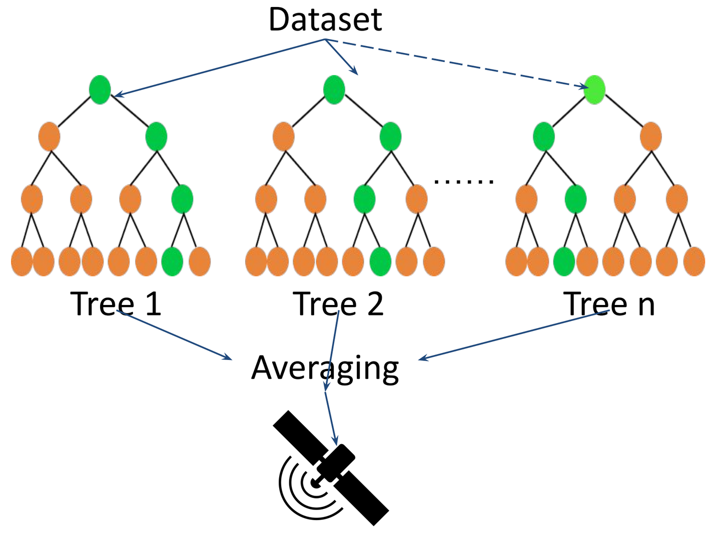
In an attempt to make the conventional system of scheduling for earth-science observations autonomous, dynamic targeting has been a highly discussed topic of interest. Enabling satellites to autonomously determine the priority observations in space reduces costs, labor and improves efficiency of Earth-Science Missions. A reinforcement learning technique is proposed to predict valid observations in a dynamic environment like space. The research question the project addresses is how to devise a near optimal way to prioritize areas with high atmospheric activity, precisely observing convective precipitation, resulting in enhanced environmental monitoring and providing deeper insights to Earth's atmospheric dynamics.

## Tools and Technologies

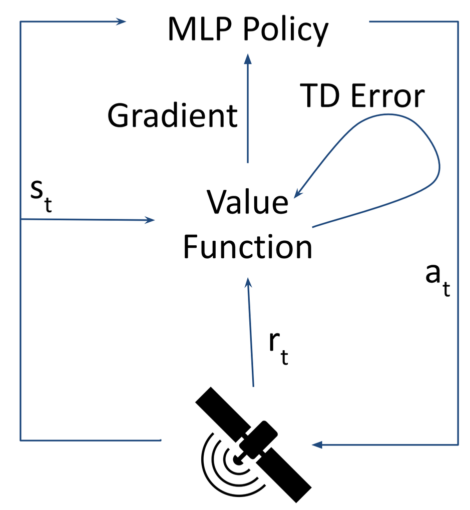
- Scikit-Learn, library to get performance measuring functions and data preprocessing functions..
- NASA GEOS5 G5NR Dataset for preparing training dataset with satellite orbit, instrument, time, geolocation, topology i.e CNPRCP (Convective Precipitation) over 4 months (May-Sept 2022)
- GeoPandas, open source library to use pandas objects with geospatial data types.
- Matplotlib, Cartopy and Seaborn, libraries for data visualization.
- Stable Baselines3, a pytorch based library for implementing RL algorithms.
- OpenAI Gymansium, a library used to creating the reinforcement learning (RL) environment.

## Methodology

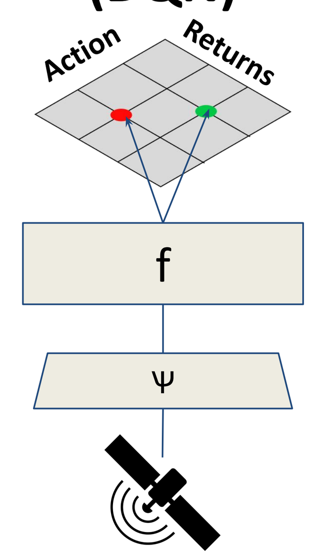
### Random Forest Classifier



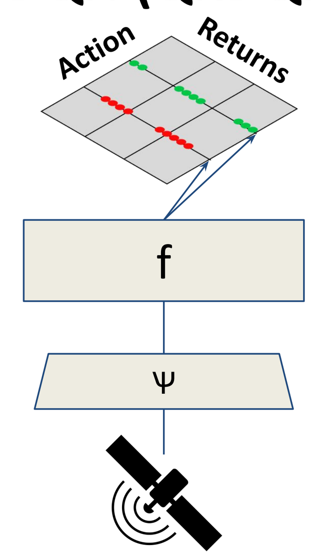

### Advantage Actor-Critic



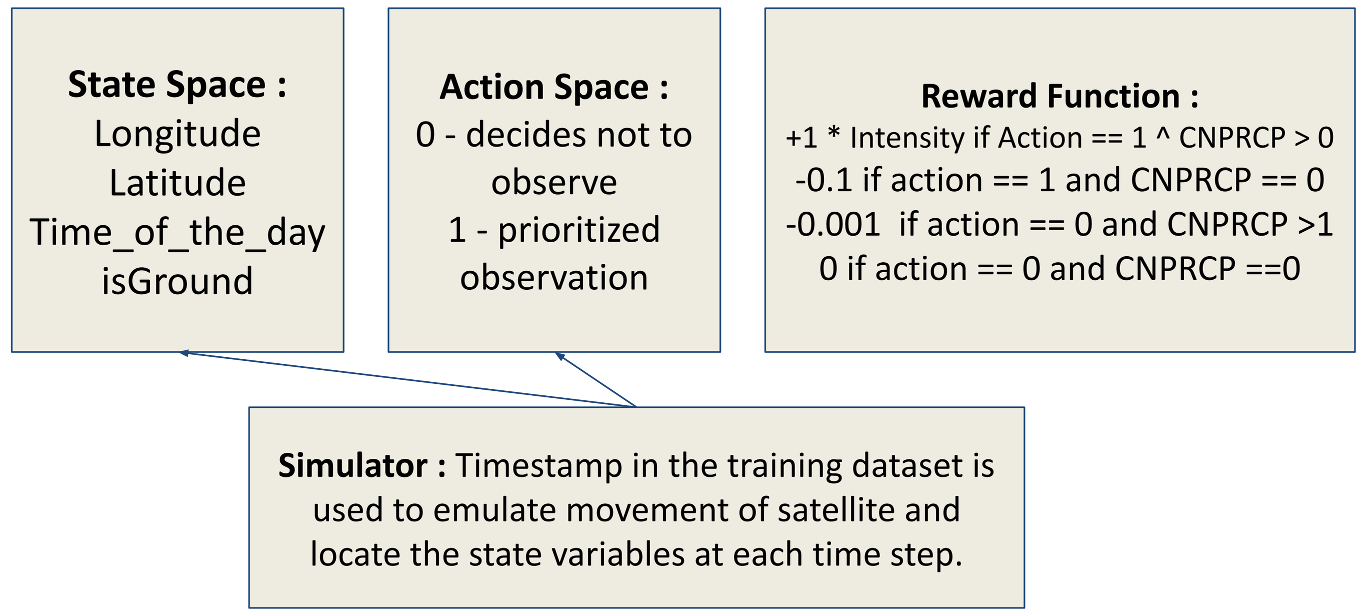

### Deep Q-Network (DQN)




### Quantile Regression DQN (QRDQN)

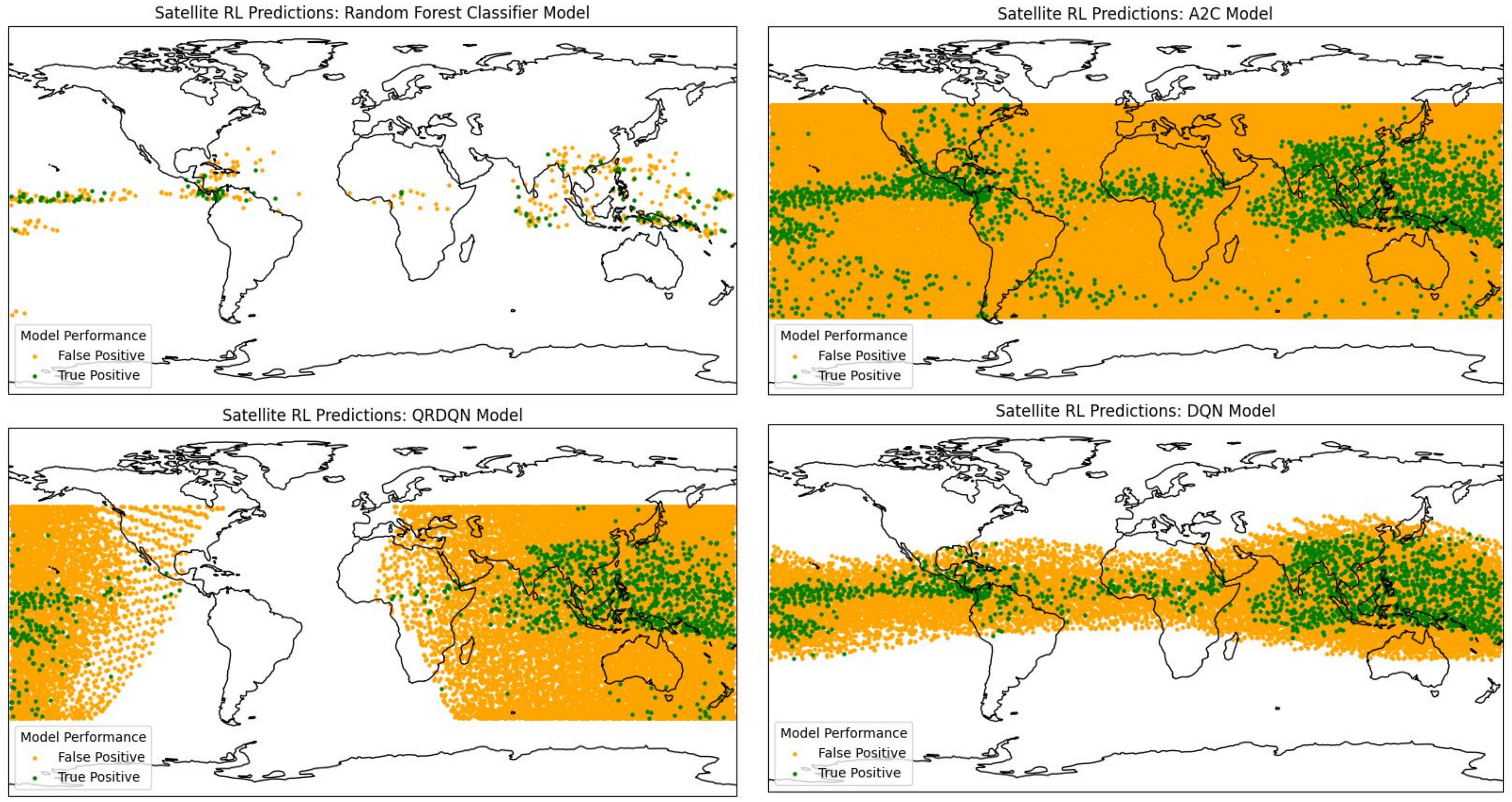


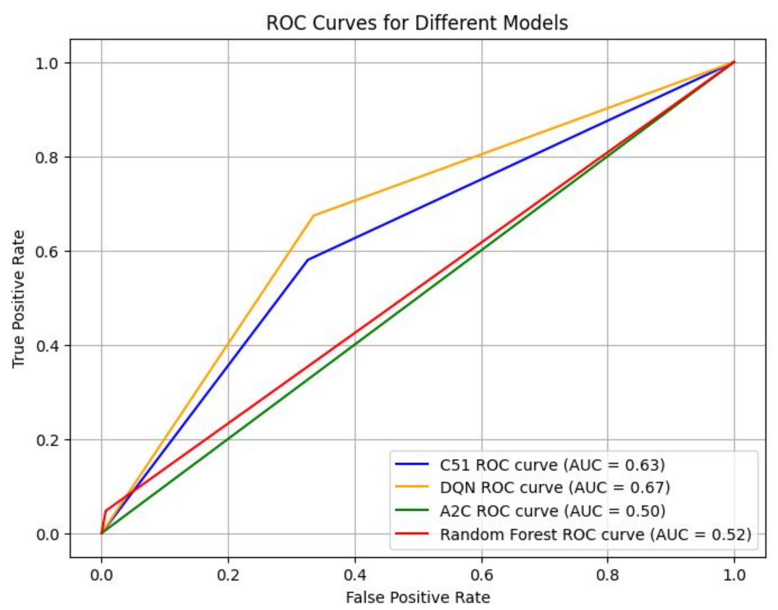

### DQN Agent (Setting up Environment)




### Monitoring model's Output :

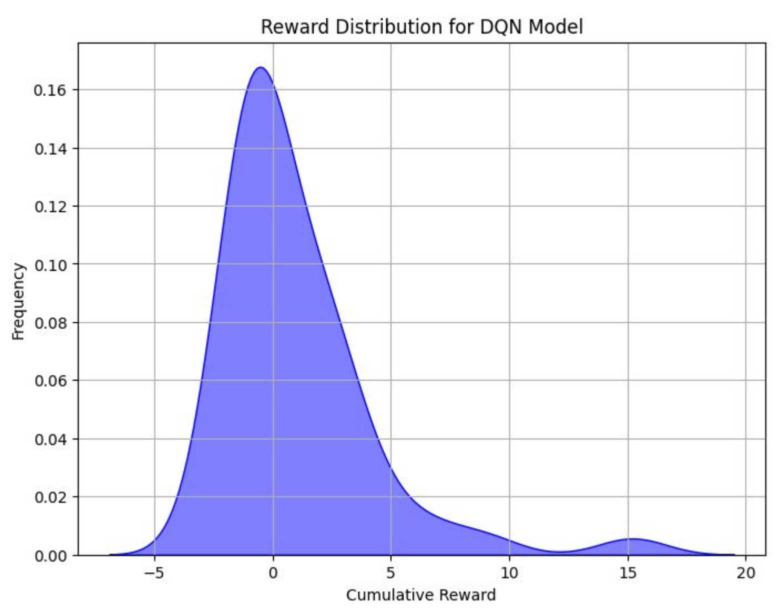
The Reinforcement learning models' performance was monitored on the outline of a world map over the entire dataset spanning over 4 months as well as during every orbit cycle. This was compared to the baseline Random Forest Classifier model.




### Receiver-Operating Characteristic Curve :

The curve maps out the true positive observations against the false positive observations. The closer the area is towards 1, the better the model performs.




### Cumulative Reward Distribution Curve:

Despite outperforming other models, the total reward sum is frequently below 0. This could be due to insufficient training dataset or suboptimal reward function.

## Conclusion

The QRDQN, which works on reward distributions rather than scalar rewards, achieved a recall of 84%, compared to 79% from the DQN model. However, it may not effectively represent real-world satellite performance, as false negatives are less critical in practical scenarios. Instead, precision serves as a more appropriate metric. The DQN model with a multi-layered perceptron (MLP) policy showed better precision (12.8%). Overall, policy-based RL models outperform the comparison baseline – Random Forest Classification Model.

## Future Work

Furthermore, future work revolves around addition of more rows, state space features like cloud cover and humidity and constraints like battery life cycle. This would require further prioritization based on intensity. Lastly, an optimal reward function could boost the model's predictions of priority observations. Different policies like CNN and Multi-Input as well as the use of Transformers and Ensemble Learning are also promising approaches.

FURI

**Acknowledgements**

Expressing immense gratitude to Dr. Paul Grogan's mentorship and the support provided by members in his CoDe lab. Special thanks to Mr. Suvan Kumar and Dr. Ismael Tapia Tamayo for their constant guidance

ASU Ira A. Fulton Schools of Engineering
Arizona State University