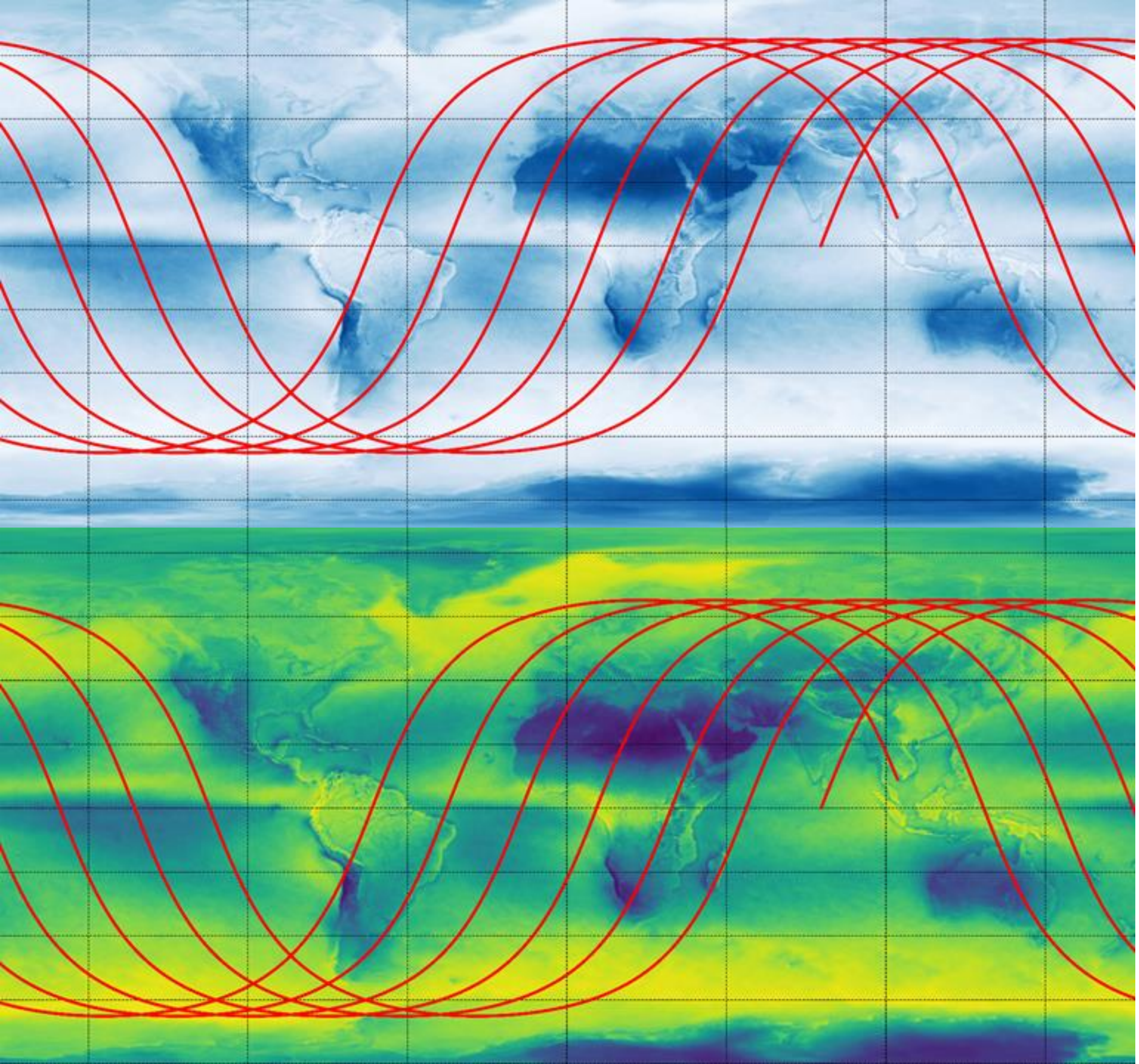


Learning-Based Planning for Improving Science Return of Earth Observation Satellites

Abigail Breitfeld[1], Alberto Candela[2], Juan Delfa[2], Akseli Kangaslahti[3], Itai Zilberstein[2], Steve Chien[2], David Wettergreen[1]

[1]Robotics Institute, Carnegie Mellon University
[2]Jet Propulsion Laboratory, California Institute of Technology
[3]University of Michigan, Ann Arbor

International Workshop on Planning & Scheduling for Space, 2025 (Presentation Only)

Carnegie Mellon University ROBOTICS INSTITUTE

Jet Propulsion Laboratory California Institute of Technology

UNIVERSITY OF MICHIGAN

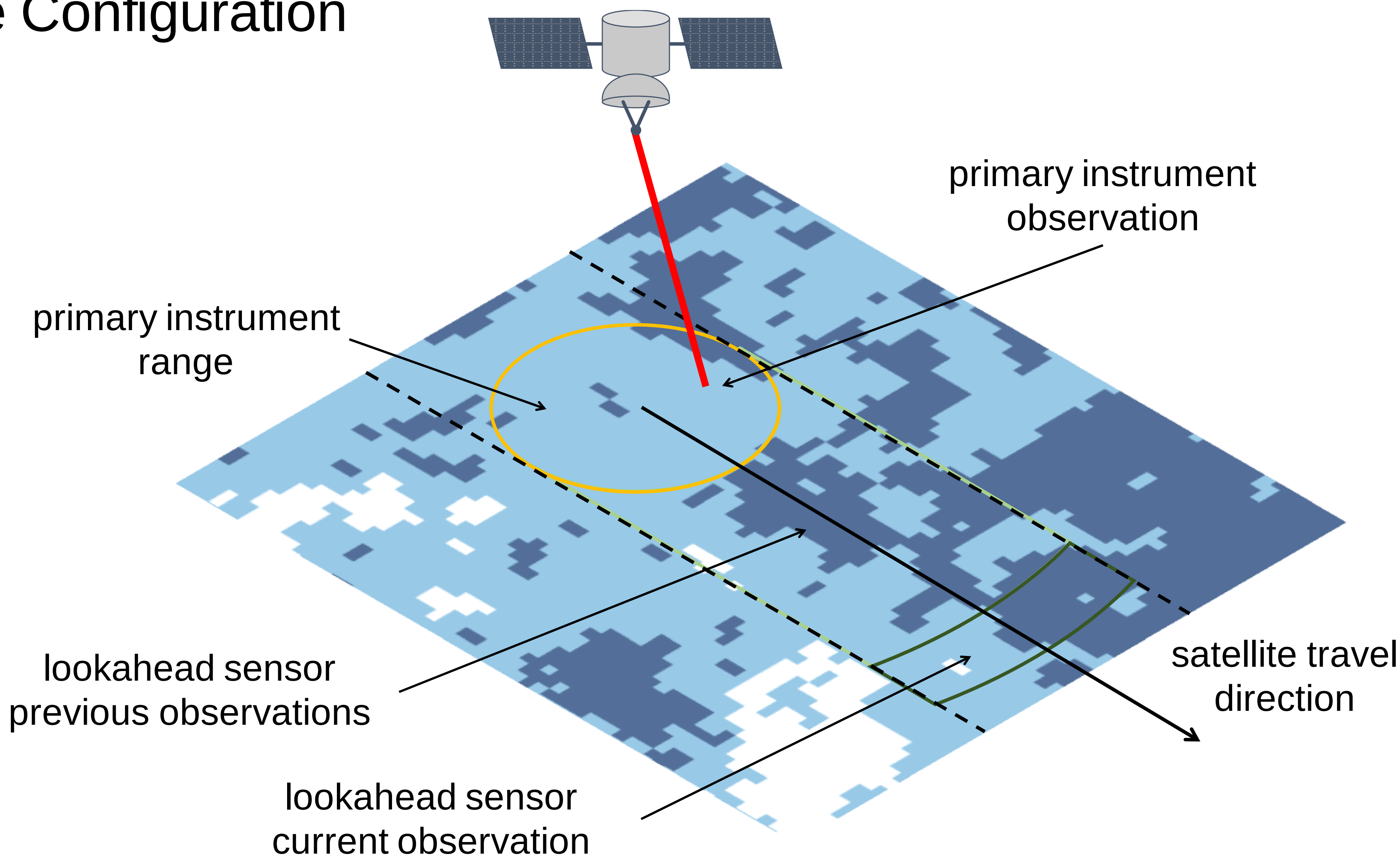
Dynamic Targeting

Problem Overview

Dynamic Targeting:
Pointing an instrument on an Earth-observing satellite to sample locations with the **most useful scientific information**

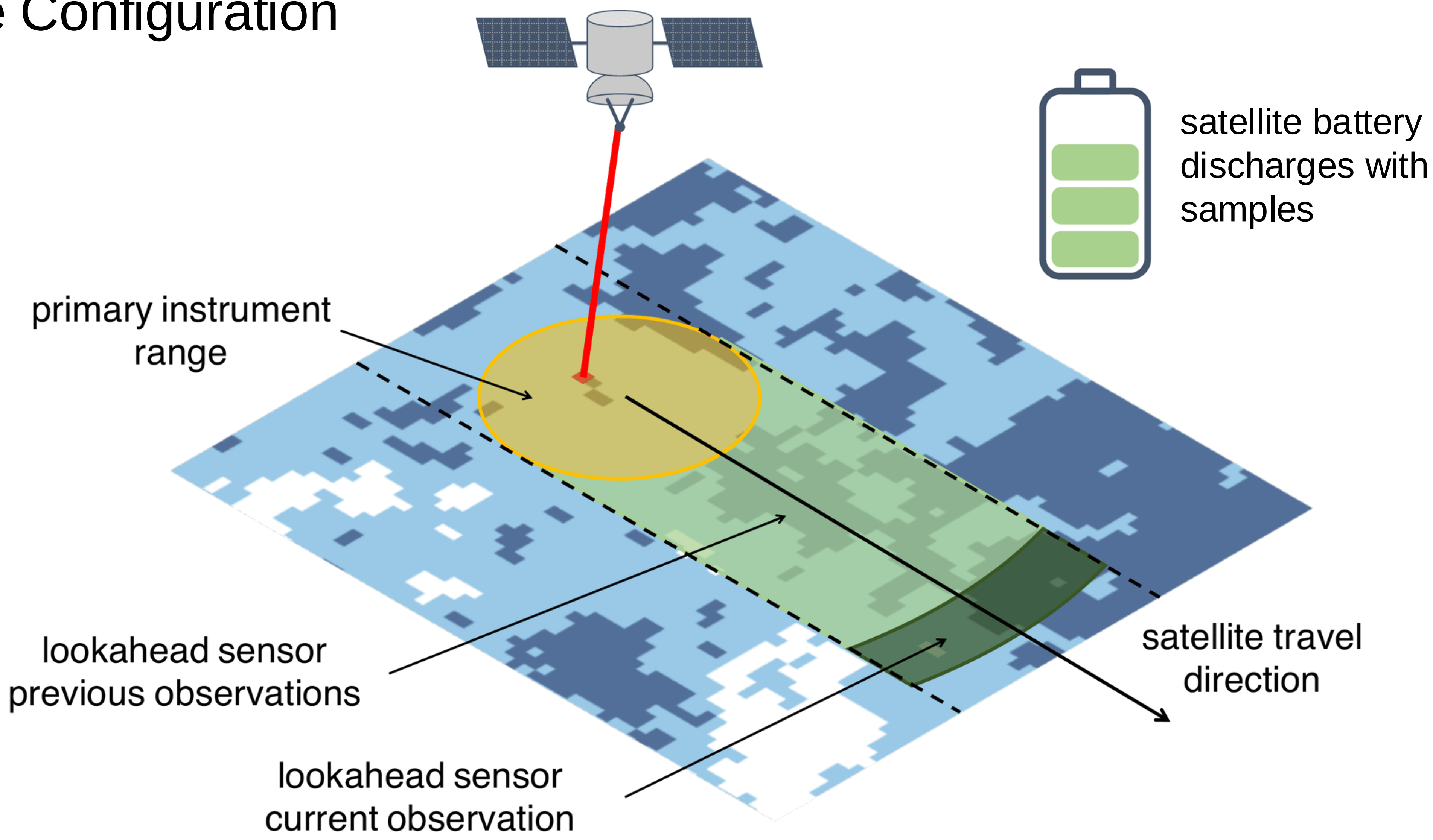
Dynamic Targeting

Satellite Configuration



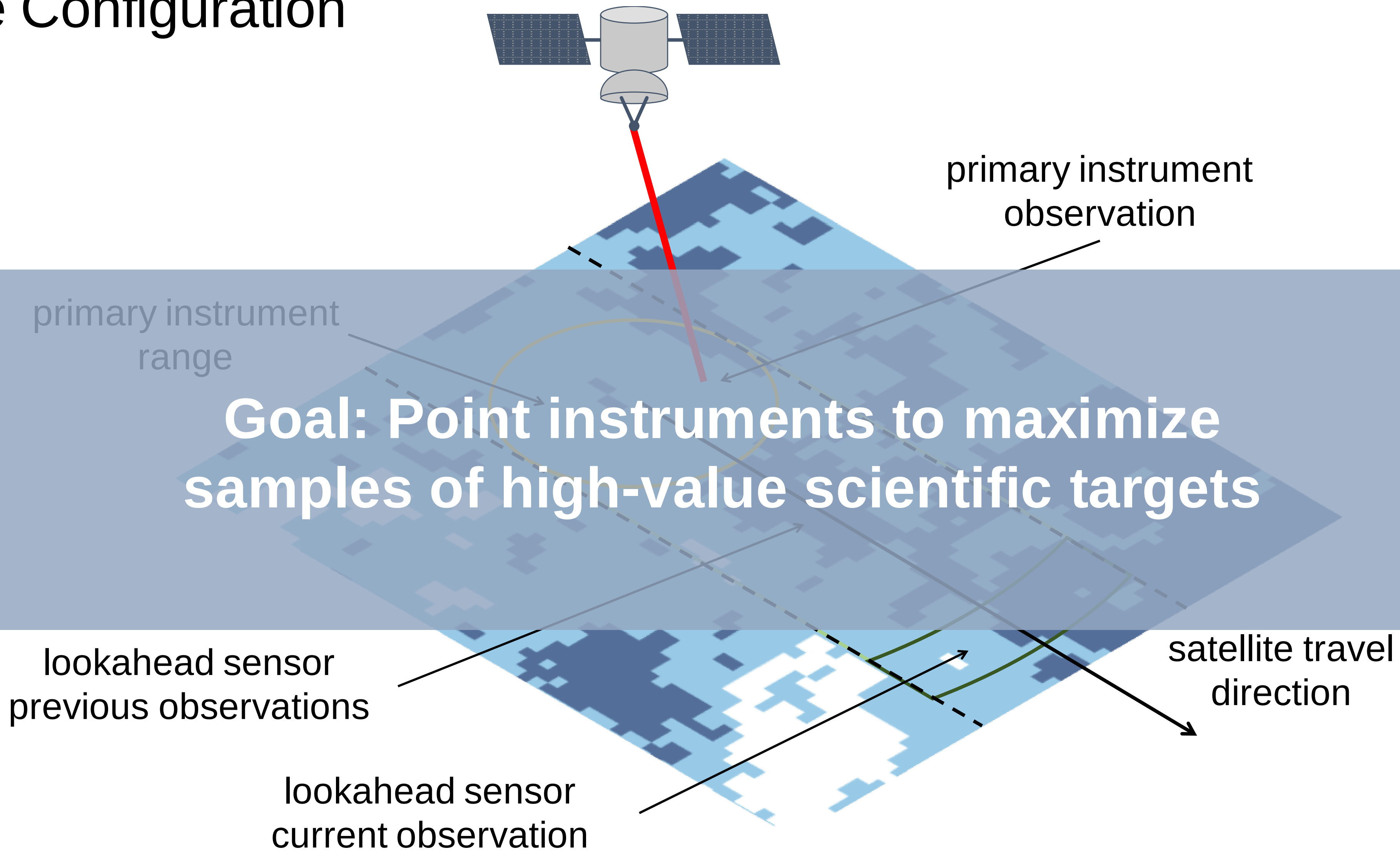
Dynamic Targeting

Satellite Configuration



Dynamic Targeting

Satellite Configuration

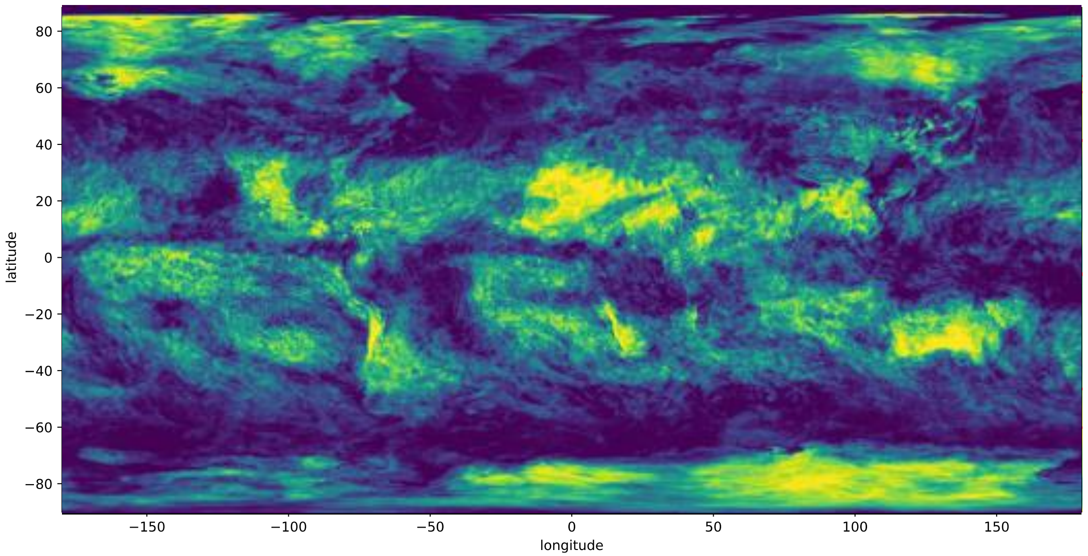


Goal: Point instruments to maximize samples of high-value scientific targets

Dynamic Targeting

Environment Overview

Cloud Avoidance:
Moderate Resolution Imaging Spectroradiometer (MODIS)

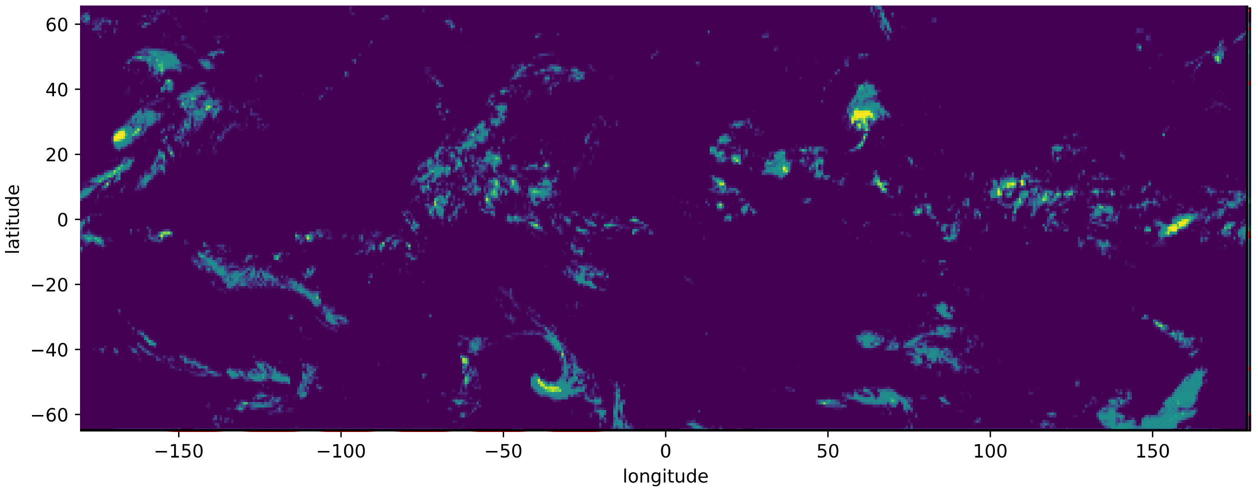


Cloud
Reward = 1

Mid-Cloud
Reward = 10

Clear
Reward = 100

Storm Hunting:
Global Precipitation Measurement (GPM)



No Storm
Reward = 1

Rainy Anvil
Reward = 10

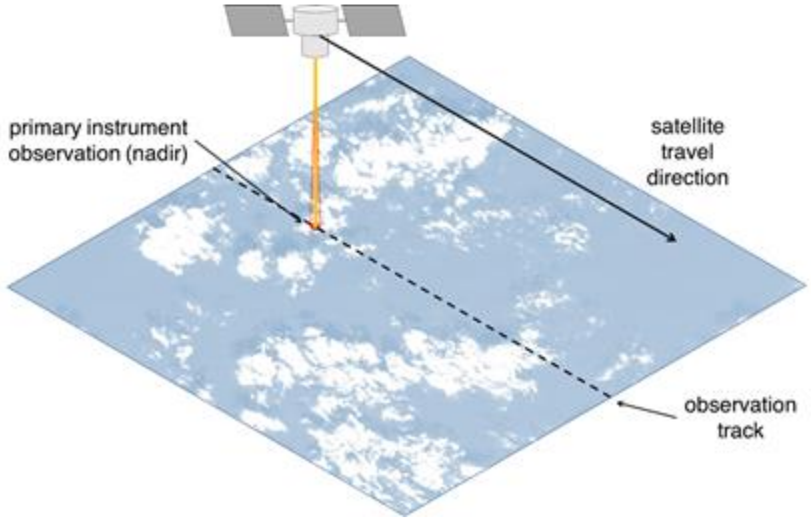
Convective Core
Reward = 100

Figure: Candela A., Swope J., Chien S. 2023

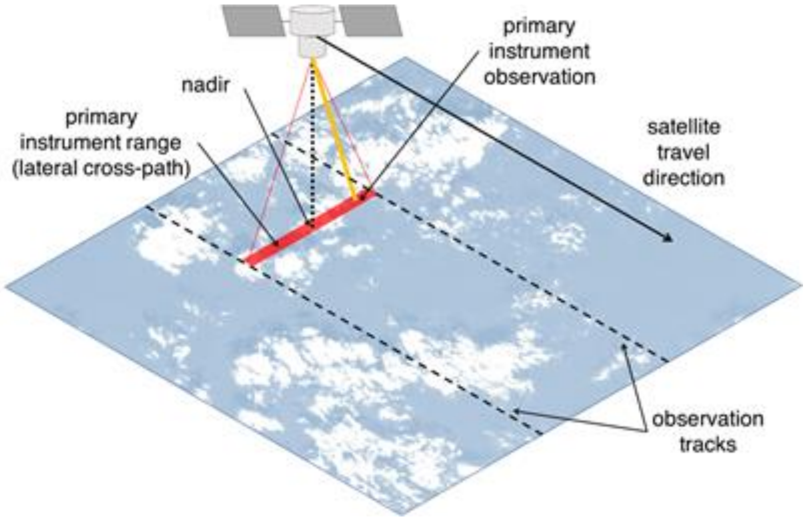
Dynamic Targeting

Prior Algorithms

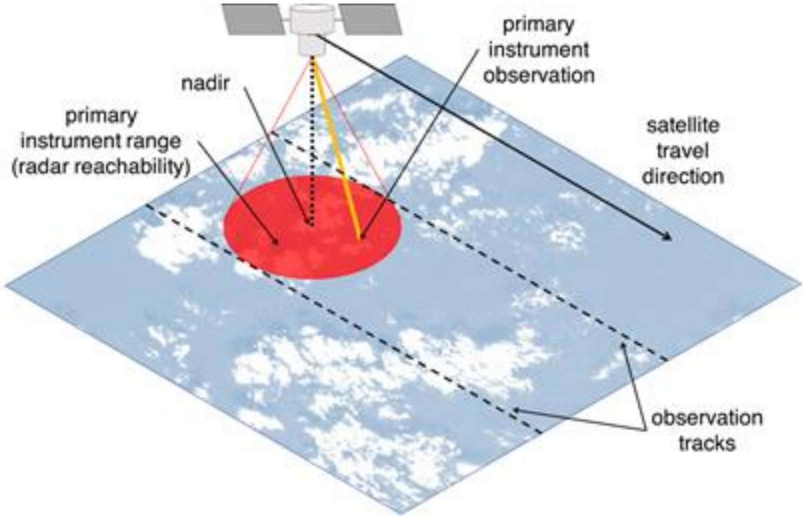
Greedy Nadir



Greedy Lateral



Greedy Radar



Greedy Window

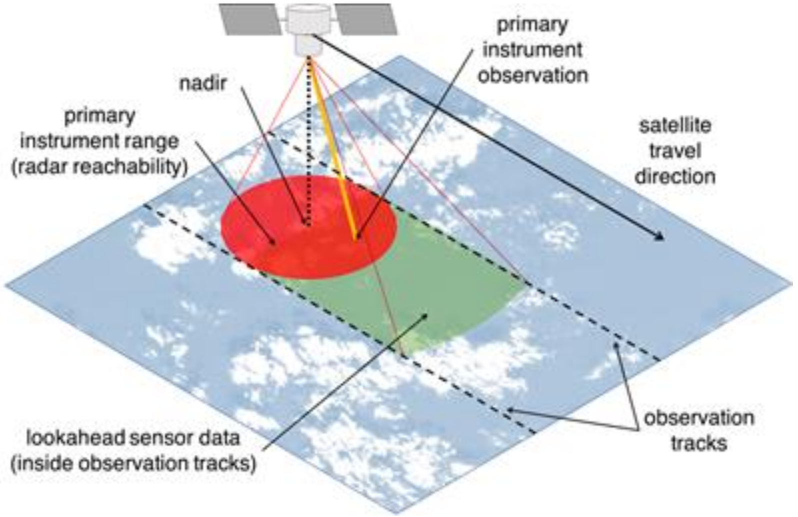


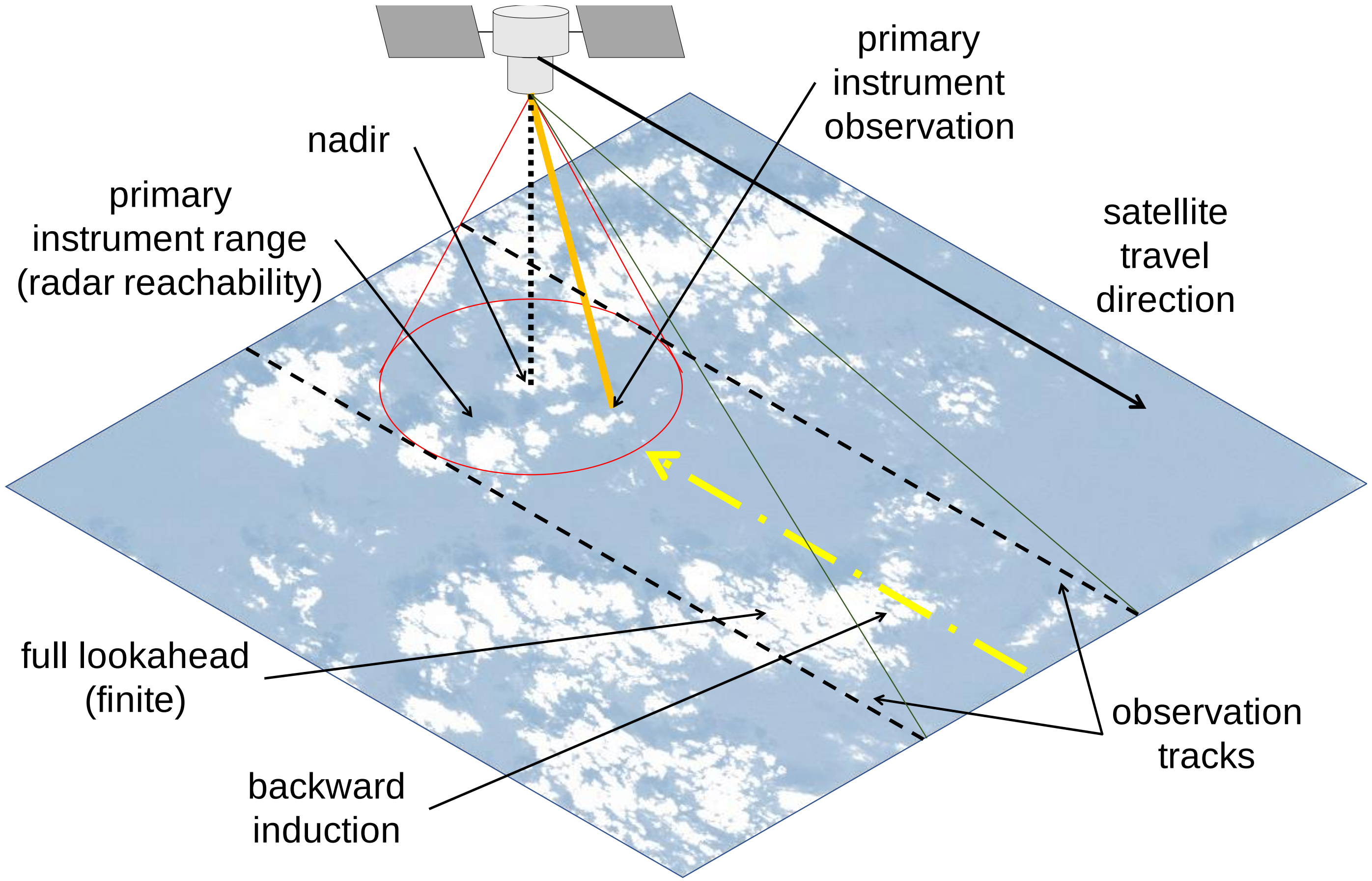
Figure: Candela A., Swope J., Chien S. 2023

Dynamic Targeting

Prior Algorithms

Dynamic Programming

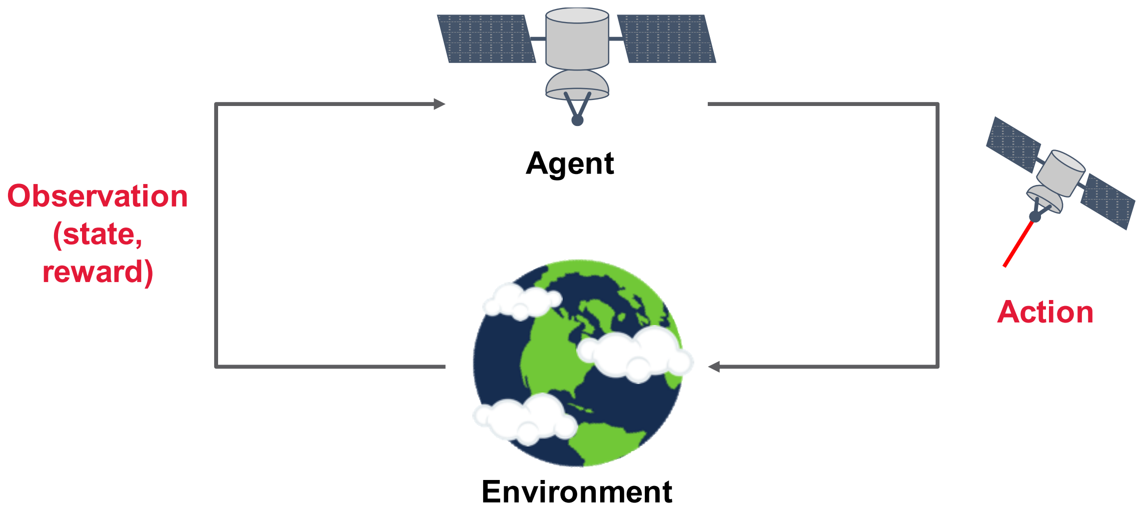
- Considers entire path of the satellite
- Not feasible for real-world use
- Can be used as an "oracle" baseline



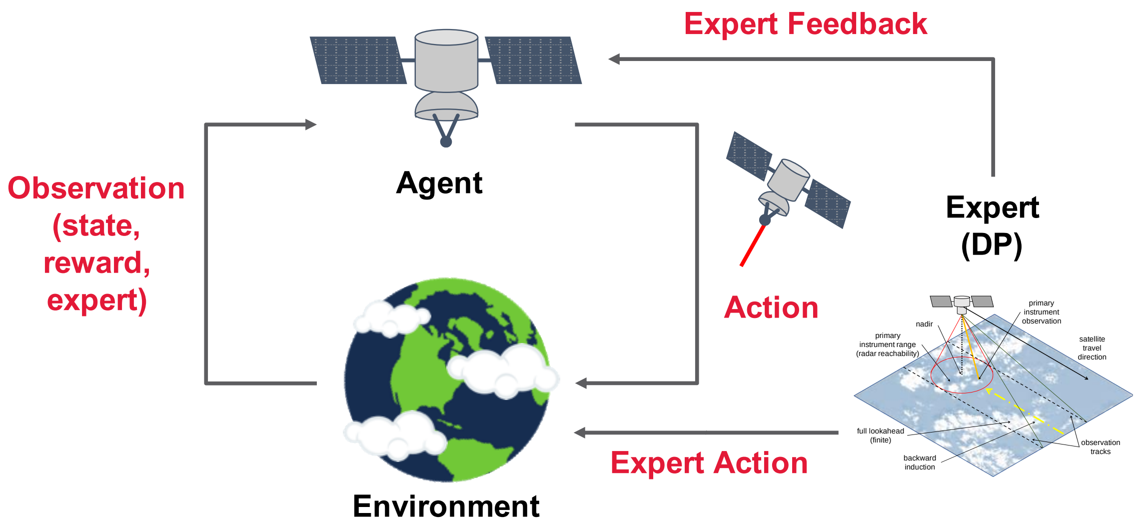
Learning for Dynamic Targeting

Goal: Use machine learning techniques to improve the science return of satellites as compared to existing heuristic methods

Reinforcement Learning

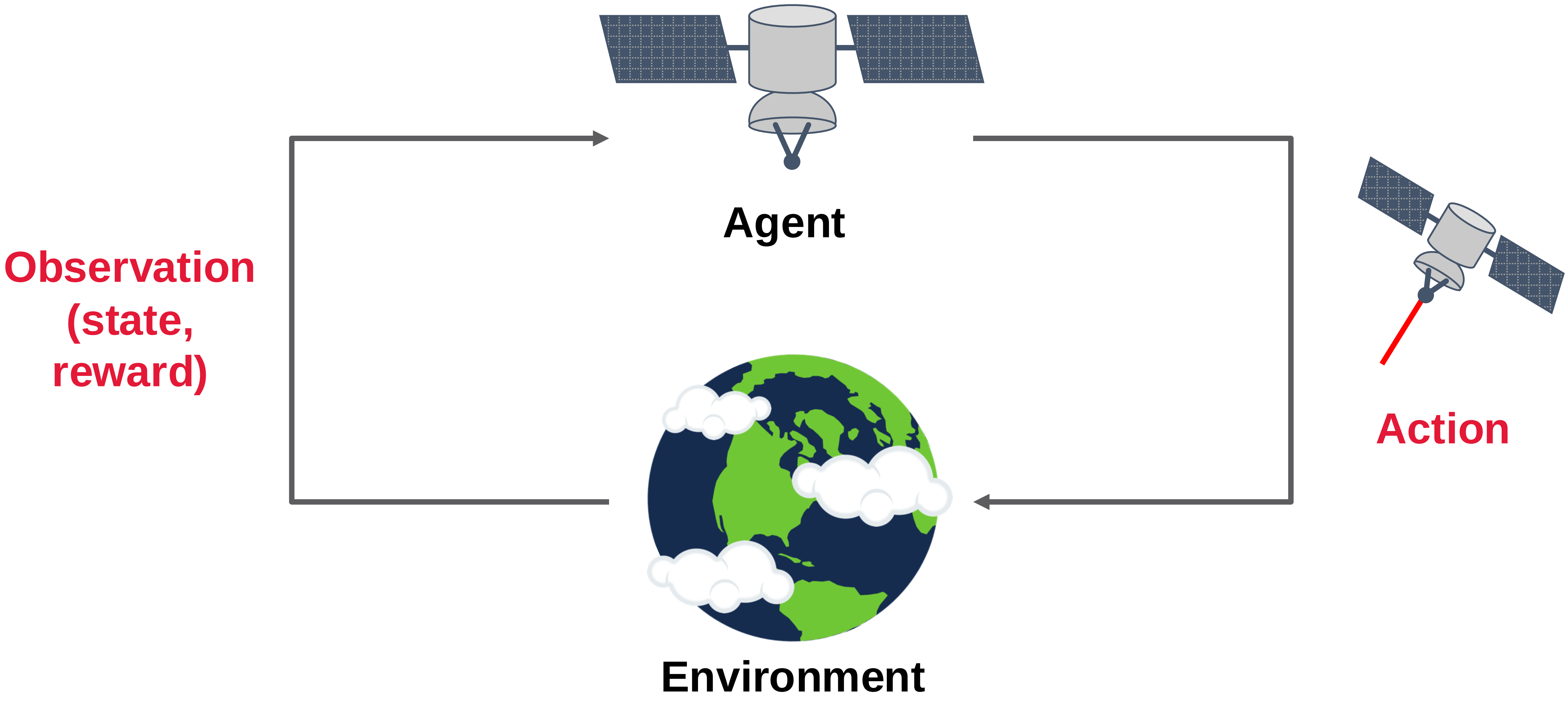


Imitation Learning

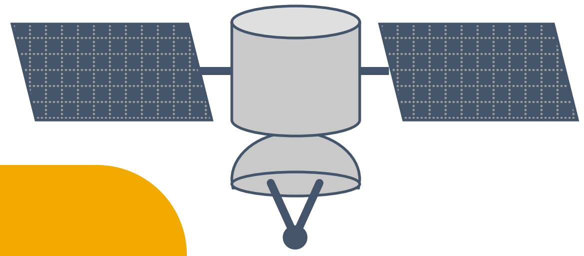


Learning for Dynamic Targeting

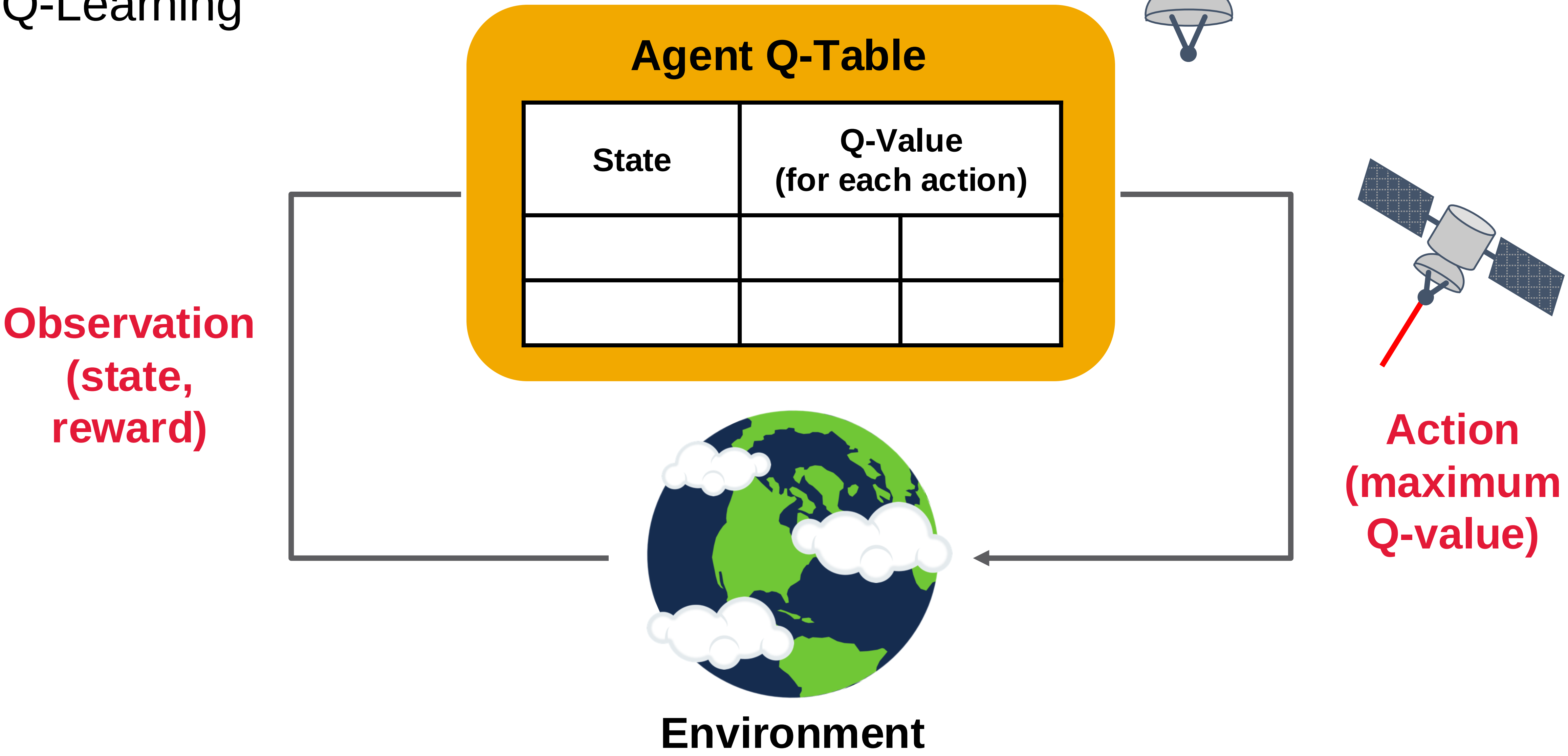
Reinforcement Learning



Learning for Dynamic Targeting



Q-Learning



Learning for Dynamic Targeting

Q-Learning Update Function

$$\underbrace{\boxed{Q(s_t,a_t)}}_{\text{new value}} \leftarrow \overbrace{\boxed{Q(s_t,a_t)}}^{\text{current value}} + \underbrace{\boxed{\alpha}}_{\text{learning rate}} \Big(\overbrace{\boxed{r_{t+1}}}^{\text{reward}} + \underbrace{\boxed{\gamma}}_{\text{discount rate}} \overbrace{\boxed{\max_a Q(s_{t+1},a)}}^{\text{maximum expected future value}} - \underbrace{\boxed{Q(s_t,a_t)}}_{\text{current value}} \Big)$$

Learning for Dynamic Targeting

Q-Learning State and Action

Cloud Avoidance

State = [State of Charge; Cloud, Mid-Cloud, Clear } Present in Primary; Cloud, Mid-Cloud, Clear } Present in Lookahead]

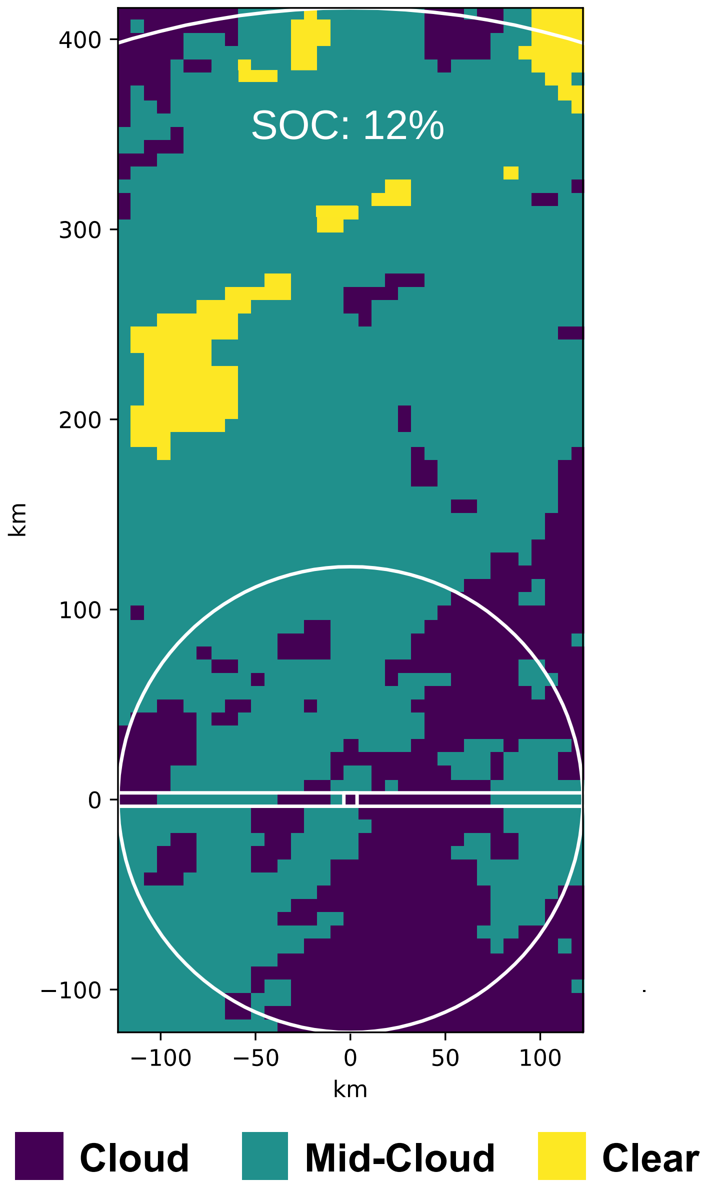
Storm Hunting

State = [State of Charge; No Storm, Rainy Anvil, Convective Core } Present in Primary; No Storm, Rainy Anvil, Convective Core } Present in Lookahead]

Possible Actions:

1. Sample highest reward cloud type closest to nadir
2. Do not sample

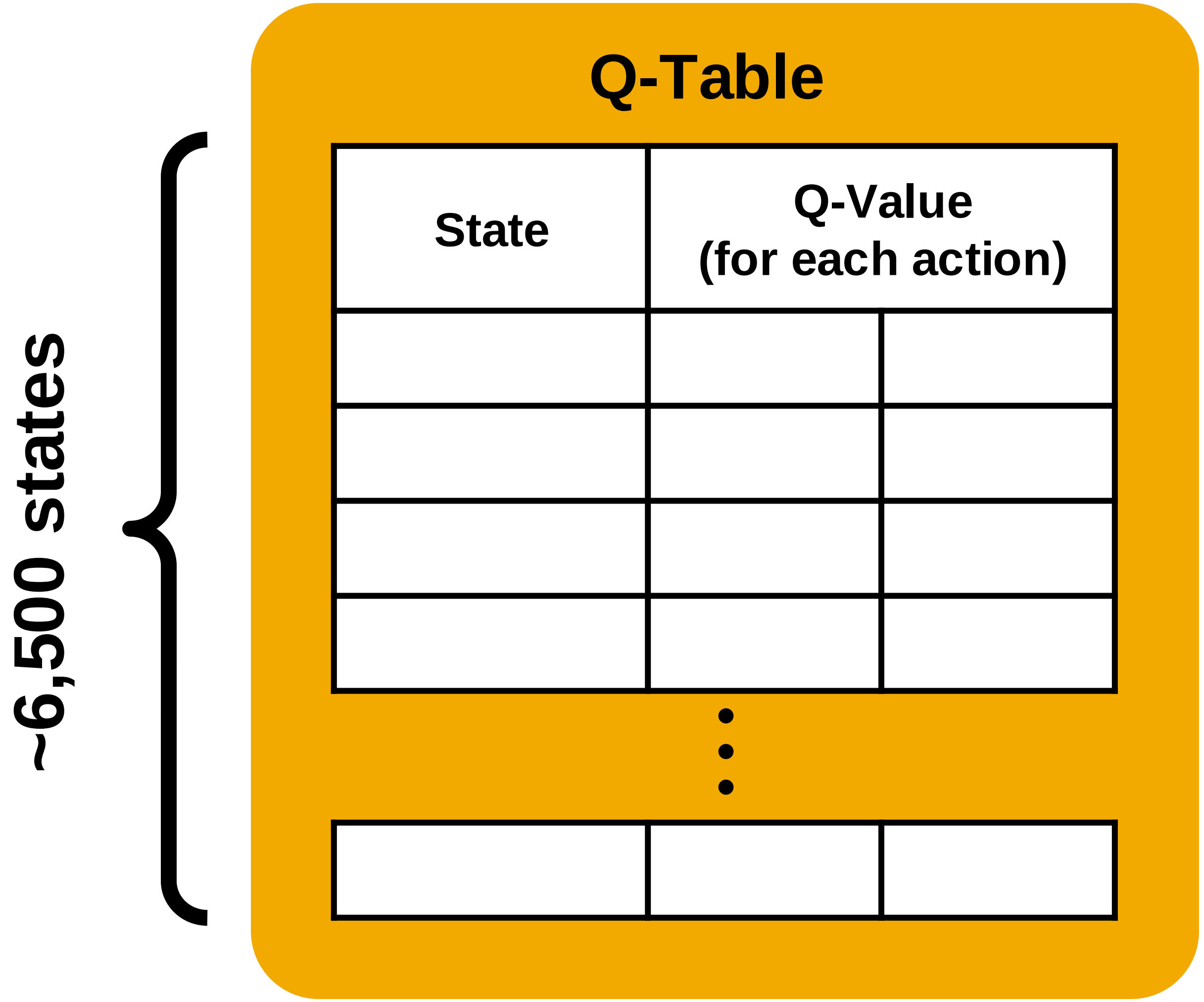
State: [12, 1, 1, 0, 1, 1, 1]



Learning for Dynamic Targeting

Q-Learning Improvements

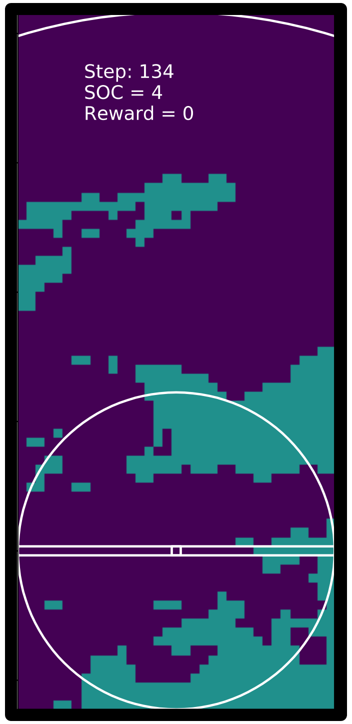
- It is unlikely that all 6,500 states will show up naturally in training data
- Each cloud image is only associated with one state of charge, which limits the number of states encountered



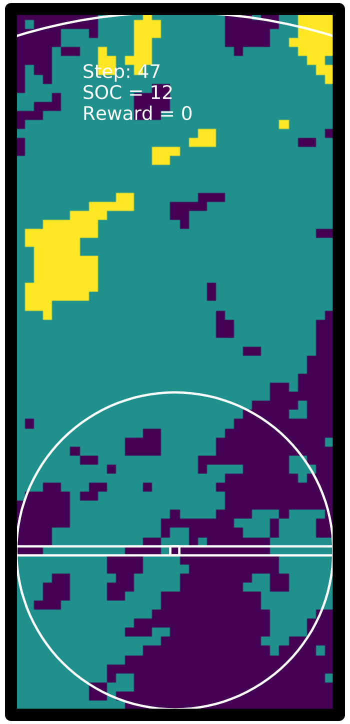
Learning for Dynamic Targeting

Q-Learning Improvements: Consider Every Possible State of Charge

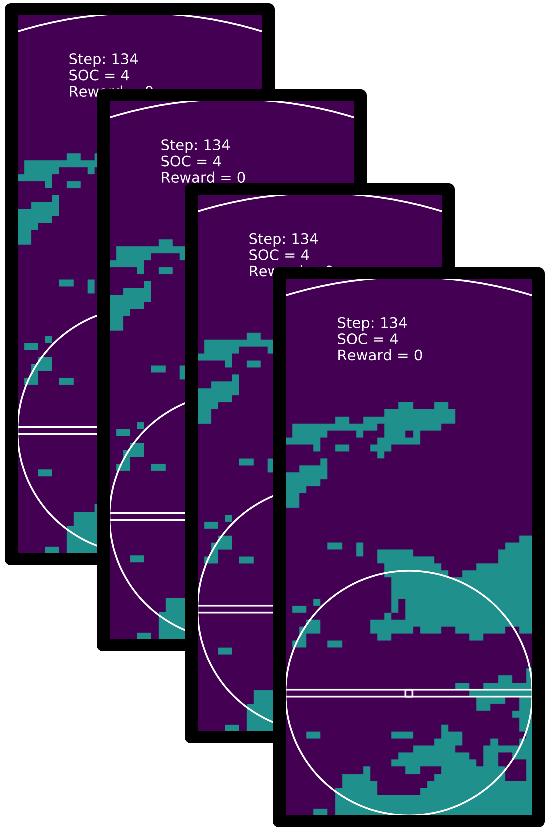
State:
[4, 1, 1, 0, 1, 1, 0]



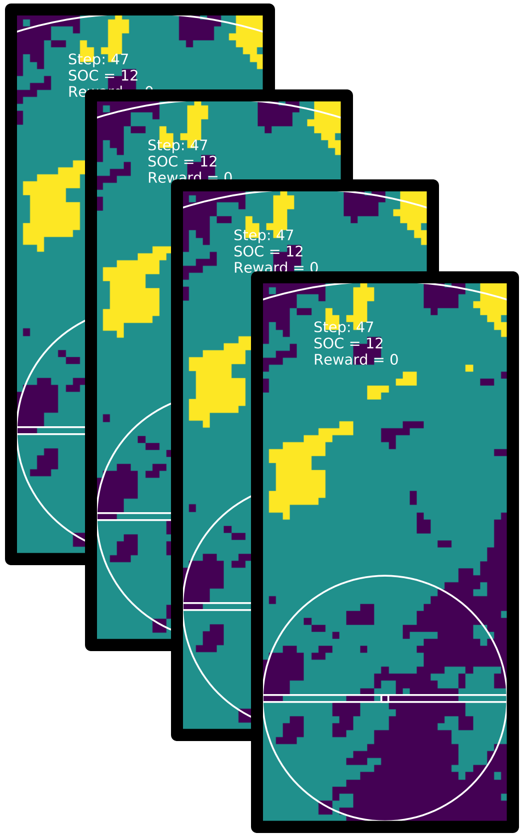
State:
[12, 1, 1, 0, 1, 1, 1]



States:
[[0-100], 1, 1, 0, 1, 1, 0]

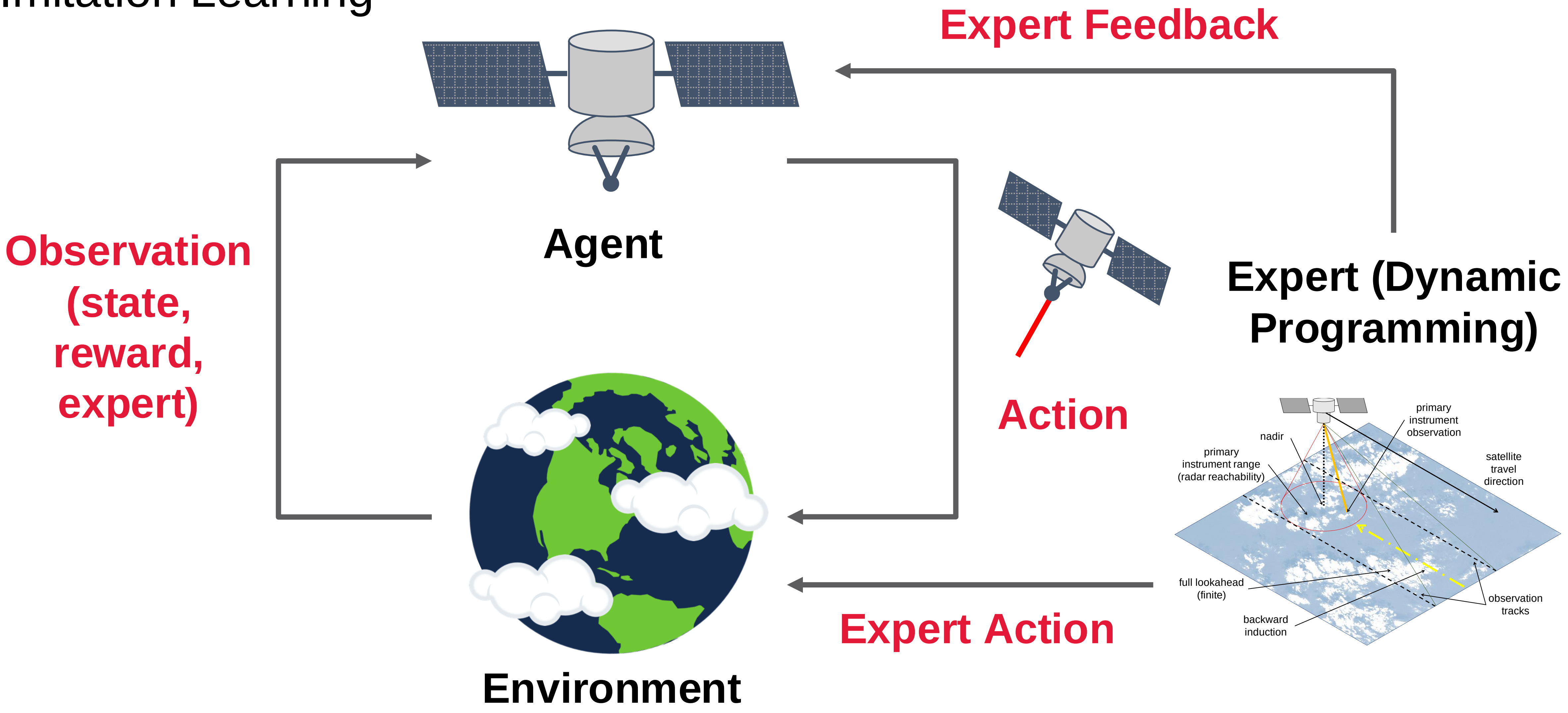


States:
[[0-100], 1, 1, 0, 1, 1, 1]



Learning for Dynamic Targeting

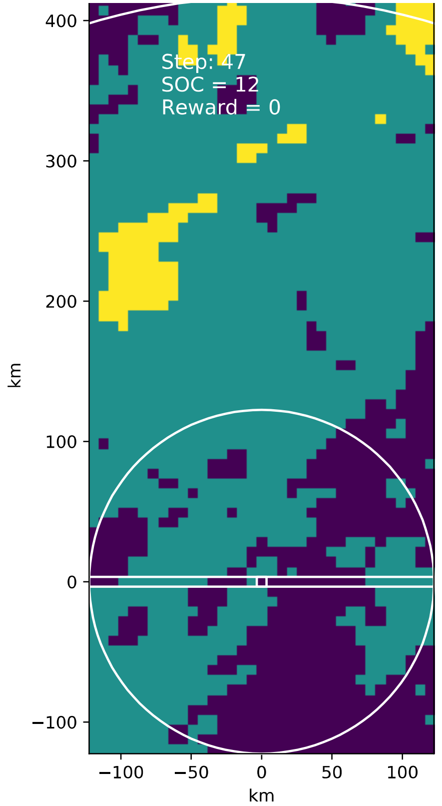
Imitation Learning



Learning for Dynamic Targeting

Behavioral Cloning

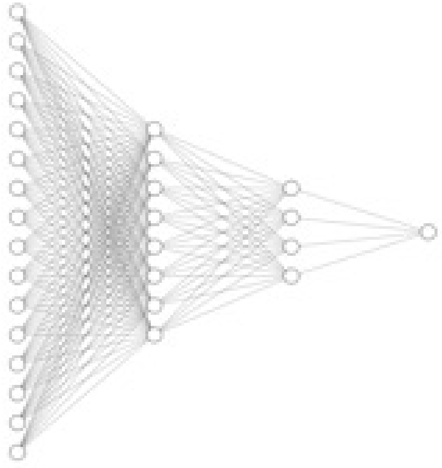
instrument data



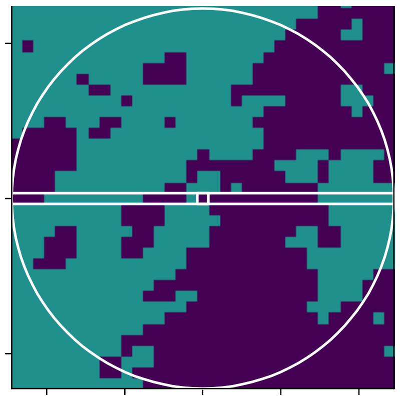
state
feature vector

$$s_t = \begin{bmatrix} 0.12 \\ 1.0 \\ 1.0 \\ 1.0 \\ 1.0 \\ 1.0 \\ 1.0 \\ 0.5 \\ 0.5 \\ 0.0 \\ 0.2 \\ 16 \\ 0.2 \end{bmatrix}$$

neural network
learns a policy π_θ that approximates the expert policy $\pi^\star$



decision making
draw action from policy
$a_t \sim \pi_\theta(s_t)$



do not sample

Learning for Dynamic Targeting

Behavioral Cloning State and Action

Cloud Avoidance

State = [
State of Charge
Cloud, *Mid-Cloud*, *Clear* } *Present in Primary*
Cloud, *Mid-Cloud*, *Clear* } *Present in Lookahead*
Cloud, Mid-Cloud, Clear } Fraction in Primary
Cloud, Mid-Cloud, Clear } Fraction in Lookahead
]

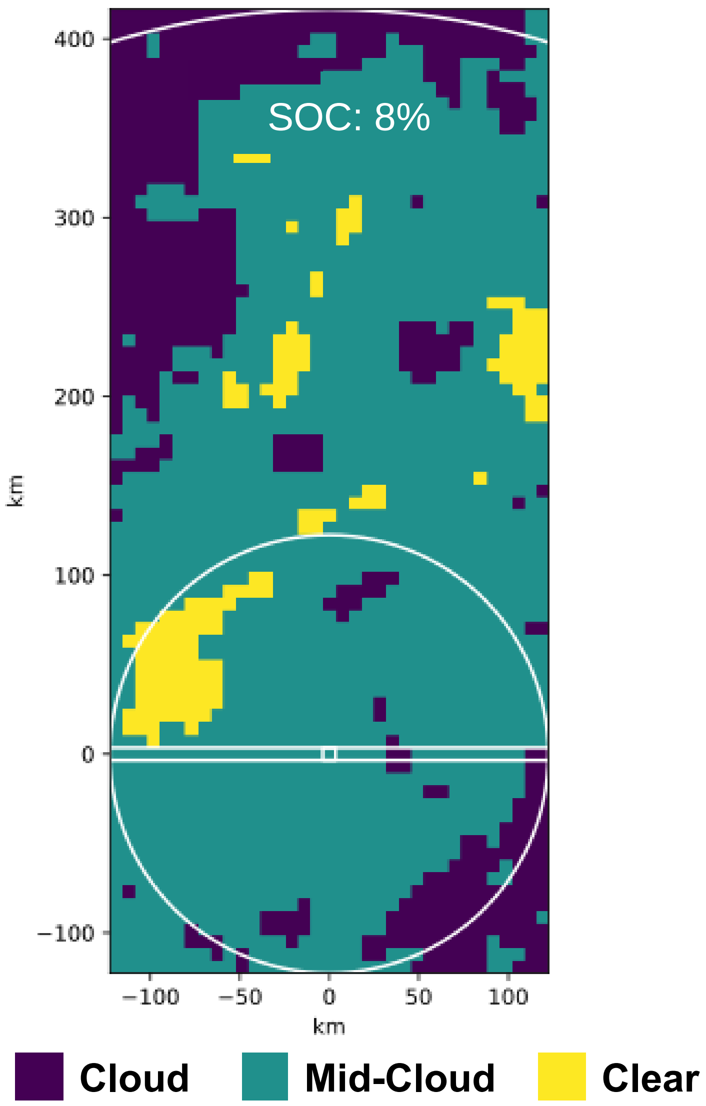
Storm Hunting

State = [
State of Charge
No Storm, *Rainy Anvil*, *Convective Core* } *Present in Primary*
No Storm, *Rainy Anvil*, *Convective Core* } *Present in Lookahead*
No Storm, Rainy Anvil, Convective Core } Fraction in Primary
No Storm, Rainy Anvil, Convective Core } Fraction in Lookahead
]

Possible Actions:

1. Sample highest reward cloud type closest to nadir
2. Do not sample

State: [8, 1, 1, 1, 1, 1, 1, 0.15, 0.71, 0.14, 0.09, 0.72, 0.19]

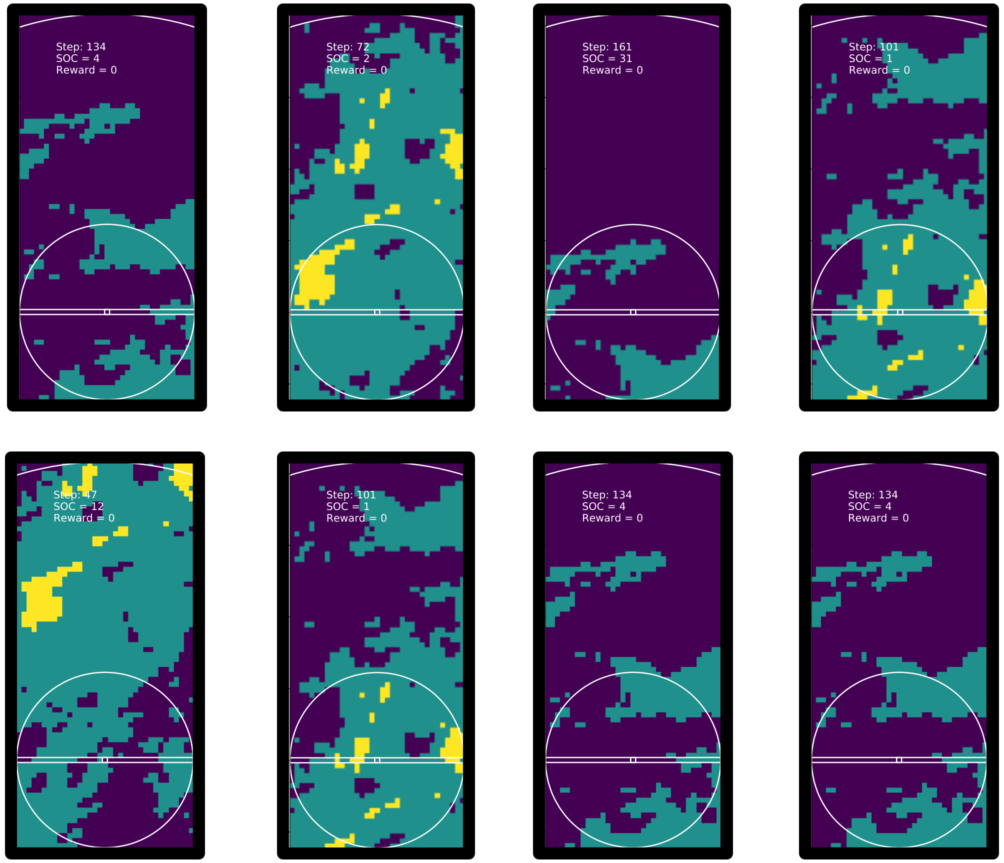


Learning for Dynamic Targeting

Behavioral Cloning Improvements: Balance Datasets

Unbalanced Data

Balanced Data



Results

Experimental Setup

- We use **real satellite data** from the MODIS and GPM missions
- Each dataset was collected during a different week of the year for experimental variety
- Each dataset contains 86,400 images, representing one day

Cloud
Reward = 1

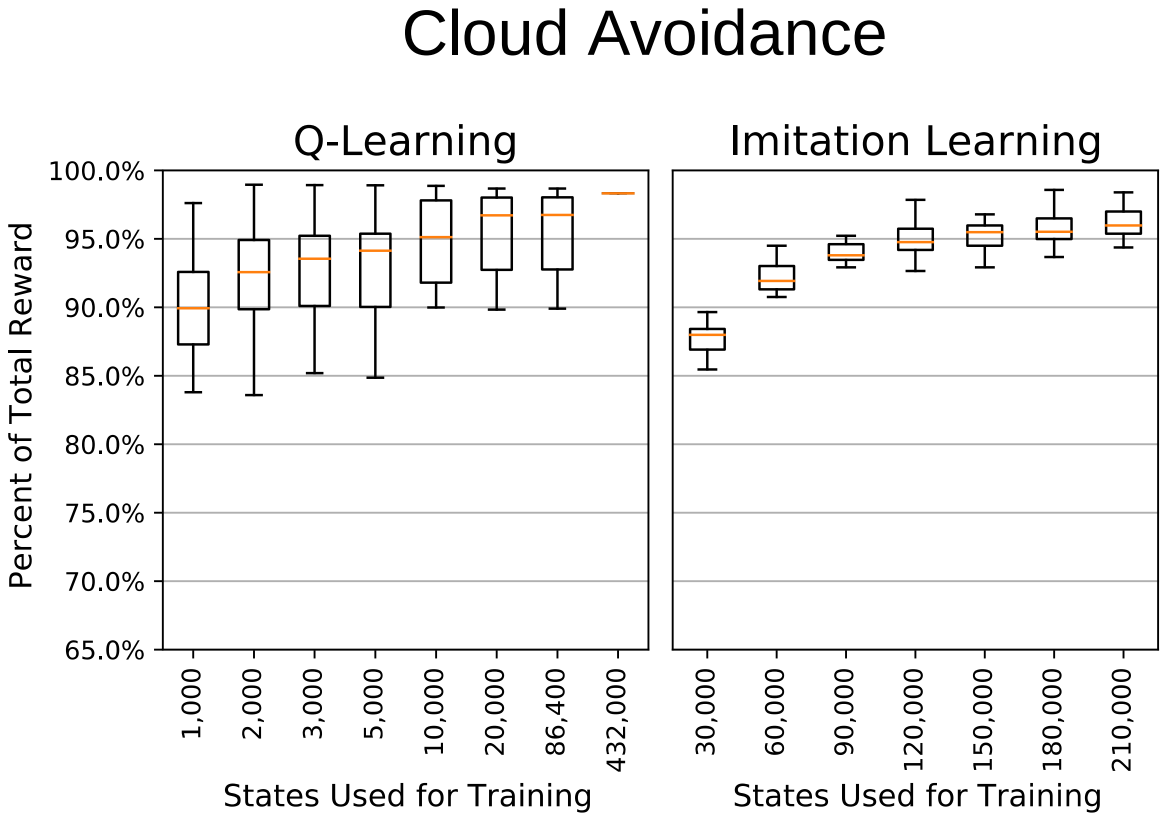
Mid-Cloud
Reward = 10

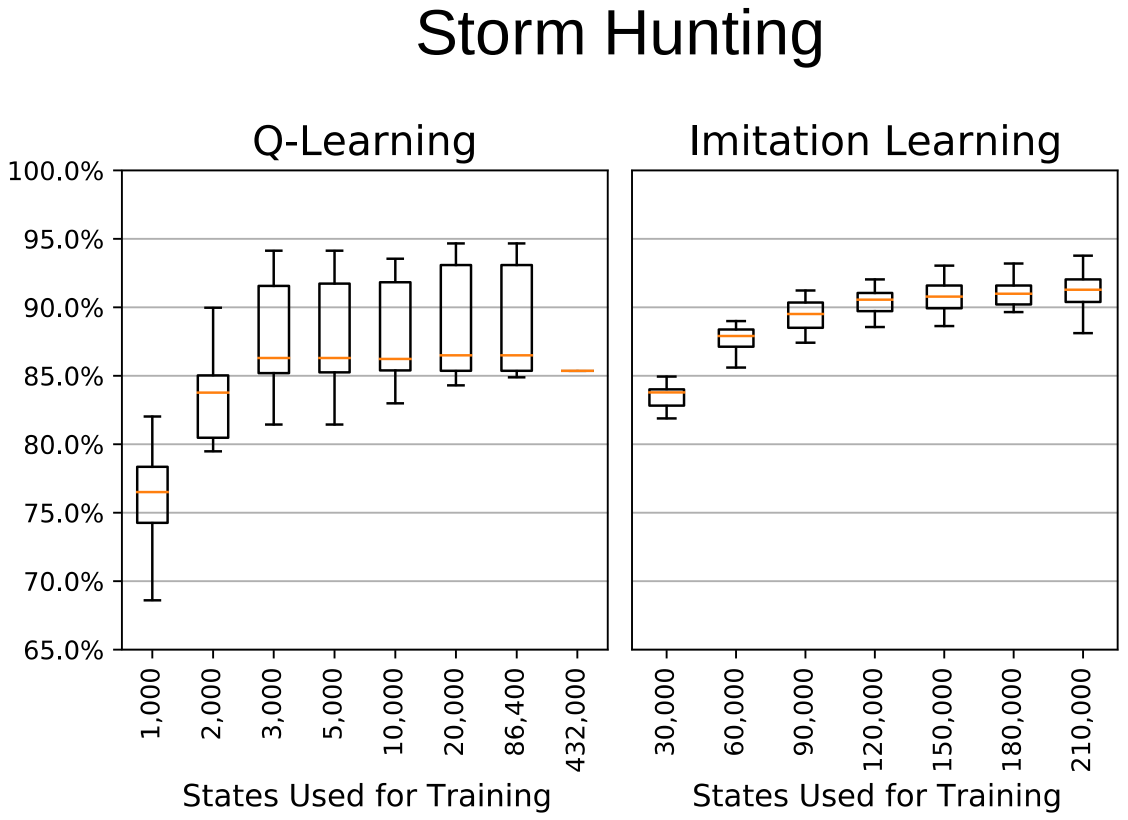
Clear
Reward = 100

Results

States Used for Training vs. Performance

Both methods can be effectively trained with relatively little data

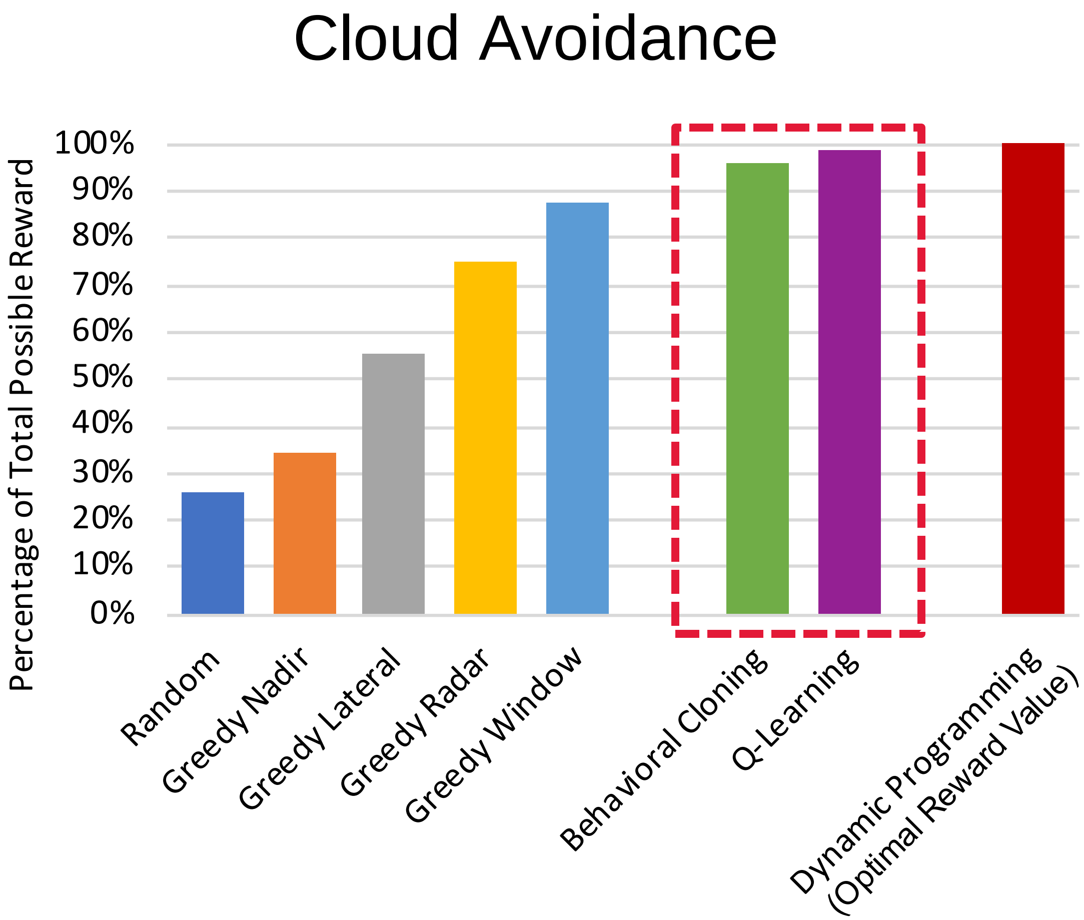


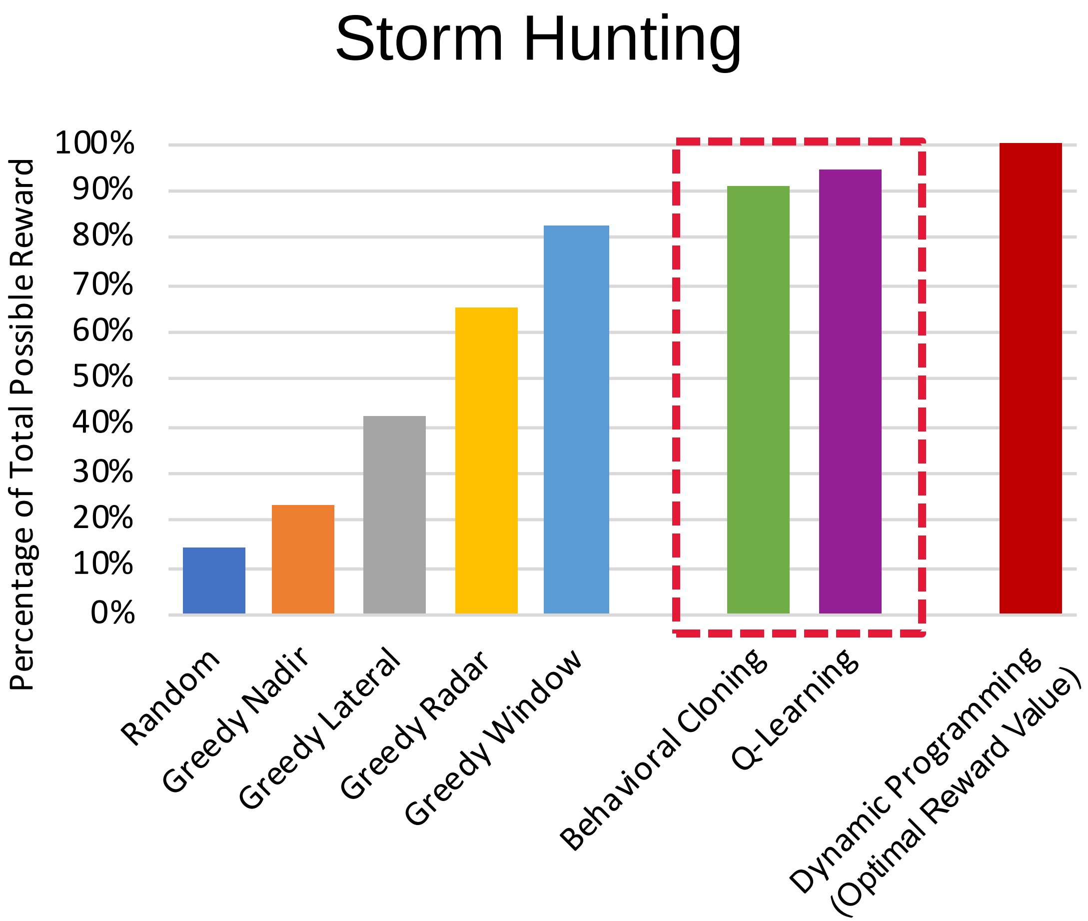


Results

Improvement in Scientific Reward

Both methods significantly improve science return over prior work





Conclusions and Next Steps

Key Takeaway

- Using learning methods improves the science return for satellite pointing tasks and requires relatively little data to train

Improvements and Next Steps

- Consider additional satellite constraints in the reward model
- Use an algorithm like Proximal Policy Optimization (PPO) for continuous state representations
- Test on real satellite

Acknowledgements

- This work was supported by the NASA Earth Science Technology Office

References

1. Candela A., Swope J., Chien S. 2023. Dynamic Targeting to Improve Earth Science Missions. AAIA Journal of Aerospace Information Systems.
2. Ogut,M.,Bosch-Lluis,X.,Kangaslahti,P.,Ramos-Perez,I.,Munoz-Martin,J.F.,Cooperrider,J.,Yue,Q.,Swope,J.,Tavallali, P., Chien, S., Pradhan, O., Deal, W., and Cooke, C., “Autonomous Capabilities and Command and Data Handling Design for the Smart Remote Sensing of Cloud Ice,” *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 7119–7122. https://doi.org/10.1109/IGARSS46834.2022.9883491.
3. Bosch-Lluis, X.,Kangaslahti, P.,Ramos, I.,Ogut,M.,Tanner,A.,Cooperrider,J.,Munoz-Martini,J.F.,Yue,Q.,Deal,W.,and Cooke, C., “Smart Ice Cloud Sensing (SMICES): An Overview of its Submillimeter Wave Radiometer,” *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 4296–4299. https://doi.org/10.1109/IGARSS46834.2022.9883671.
4. Justice, C., Vermote, E., Townshend, J., Defries, R., Roy, D., Hall, D., Salomonson, V., Privette, J., Riggs, G., Strahler, A., Lucht, W., Myneni, R., Knyazikhin, Y., Running, S., Nemani, R., Wan, Z., Huete, A., van Leeuwen, W., Wolfe, R., Giglio, L., Muller, J., Lewis, P., and Barnsley, M., “The Moderate Resolution Imaging Spectroradiometer (MODIS): land remote sensing for global change research,” *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 36, No. 4, 1998, pp. 1228–1249. https://doi.org/10.1109/36.701075.

References

6. Schulman, John, et al. “Proximal Policy Optimization.” *OpenAI*, 20 July 2017, https://openai.com/research/openai-baselines-ppo.
7. Wouter van Heeswijk. “Proximal Policy Optimization (PPO) Explained.” *Medium*, Towards Data Science, 31 Jan. 2023, https://medium.com/towards-data-science/proximal-policy-optimization-ppo-explained-ab ed1952457b.
8. Sutton, Richard S., and Andrew Barto. *Reinforcement Learning: An Introduction*. The MIT Press, 2020.
9. Huang, Shengyi, et al. “The 37 Implementation Details of Proximal Policy Optimization.” *The ICLR Blog Track*, 2022, iclr-blog-track.github.io/2022/03/25/ppo-implementation-details/.

Jet Propulsion Laboratory
California Institute of Technology

jpl.nasa.gov

Portions of this research were carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004). This work was supported by the Earth Science and Technology Office (ESTO), NASA.