Under the centroid method, the flexibility of the energy extraction strategy is limited. We tried to create a more flexible strategy in three ways, all using machine learning: table-based Q-learning, neural fitted Q-learning, and dynamic programming with neural interpolation. Table-based Q-learning discretizes the state space, and estimates the value of each possible action in each state space chunk. We used this method successfully in a low dimensional setting, but we had difficulty scaling it up. Neural-fitted Q-learning uses a neural network to stores value estimates, but we had difficulty properly incorporating new information into the neural network. Dynamic programming with neural interpolation uses a system model to simulate interactions. To allow for a continuously varying policy, we store the resulting value and policy estimates using neural networks.

RESULTS. Using dynamic programming with neural interpolation, the learning agent learned a plausible and flexible policy, which adjusted in response to thermal position and shape. This was carried out in an idealized environment, with the following state variables: distance from center of thermal, height of UAV, and direction of UAV. Current work is focused on testing the developed algorithms in the more sophisticated CRRCSim simulator, with additional states and additional actions.