**Guide to running scripts for the Scientific Reports LP Model**

**Running the simulations**

**Creating all the maps for the ICs (Python)**

**Note:** Larkin created the maps in Python using the set of 65 ICs. The code is included in the folders but may not be totally updated (<https://git.caslab.ece.vt.edu/hlarkin3/ags_grabber>)

1. Open up Pycharm and navigate to the ***importmap*** project folder
2. Open *feature\_set.py* and enter in all the desired ICs
3. Set the extent of the maps
4. Run *feature\_set.py* and the output will be *BW\_LFandInac\_Zelev\_[IC lat, IC lon].mat*

**Run all possible behavior probs for 65 ICs (MATLAB)**

1. Run *main\_hiker\_t100.m* (uses *parameters.m, randsmpl.m, run\_replicate.m, beh\_dist\_6.mat, mapdim20\_updatethresh.mat*)
   1. Open *parameters.m* and adjust number of probs and replicates
   2. Adjust the simulation time variable *simT* if necessary
   3. **Note:** if running on ARC, follow instructions for ARC below
2. Simulations save to a folder (\sims500t)

**Using ARC to run simulations**

1. Connect to VT VPN through Pulse Secure
2. Go to <https://ondemand.arc.vt.edu/pun/sys/dashboard/> and login
3. Navigate to *Interactive Apps* in the top tab and click MATLAB
4. Here you’ll submit a request for a MATLAB GUI on the Cascades Cluster
   1. My allocation details:
      1. **Account:** lost\_person\_march\_2021
      2. **Reservation:** blank
      3. **Partition:** normal\_q
      4. **MATLAB version:** R2020a
      5. **Number of Hours**: 120 (depends on number of ICs)
      6. **Number of cores:** 28
      7. **Number of GPUs per node:** 0
   2. Can open multiple sessions in different tabs to run simultaneously
5. Once the GUI opens, upload *main\_hiker\_t100.m, parameters.m, run\_replicate.m, beh\_dist\_6.mat, mapdim10\_updatethresh.mat, randsmpl.m*
6. Create a folder to save all the simulations
7. Run *main\_hiker\_t100.m*
   1. **Note**: you may have to clear the gcp at the end of the loop to be able to run multiple ICs, but sometimes you don’t have to

**Running the analysis on the simulations**

***PART 1: Energy Distance***

**Goal:** finds the energy statistic for the distance between the closest sim point and the find point from Bob's data for all 462 behaviors and reps

1. Navigate to folder \analysis energy dist
2. Run *bestbeh\_edist\_clpts.m* (uses *beh\_dist\_6.mat, mapdim20\_updatethresh.mat*)
   1. Calculates the best fits for all the simulations based on the energy dist
   2. Finds the best beh profile for all ICs using the weight, time of the closest reps
   3. Plots the trajectory points of the best fit behaviors
   4. Plots the final trajectory points of the best fit behaviors
   5. Plots these points sorted by weight
   6. Saves the best fit data in \analysis energy dist\best fit data
      1. Output: *allbesthiker\_edist500t.mat, allclosestpoints500t.mat, allweightedbesthiker\_edist500t.mat*
3. Run *plotbehaviors.m*
   1. Plots bar charts of behavior probability dists with weights, energy distances, and distances from the IPP to find
   2. Plots the sensitivity analysis for L
   3. Sorts the ICs by weights and behaviors
   4. Plots the weights versus distances
   5. Used to create some of the figures (those have been extricated to new scripts)

***PART 2: Leave-one-out cross validation***

**Goal:** Based on the distances from the actual find point, this script finds the best fitting behaviors based on the E-dist. It takes out one IC, calculates the average behavior using the weights for the rest of the ICs and saves this behavior. This behavior will be used to run the simulation for the IC that was left out.

1. Navigate to folder \analysis LOOCV
2. Run *bestbeh\_LOO1.m* (uses *beh\_dist\_6.mat, mapdim20\_updatethresh.mat*)
   1. Uses the simulation data to run a LOOCV and saves the best fits for each IC
   2. Saves the best fits (best\_EdL) in folder \analysis LOOCV\fits500\bestbeh\_icf#
3. Run *main\_LOO2.m* (uses *parameters.m, randsmpl.m, run\_replicate.m, beh\_dist\_6.mat, mapdim20\_updatethresh.mat*)
   1. Runs the best fit behavior from *bestbeh\_LOO1.m* on the IC that was left out of that analysis. The results from this simulation will used to find the percentiles
   2. Can change the number of simulation hours (100) and reps (500)
   3. Saves the output of the runs to folder \analysis LOOCV\simsfinal
4. Run *leave1outfinal3.m*
   1. Using the output of the simulations run with the probabilities found from leaving one out, this script finds the best behavior by energy distance for each IC
   2. By comparing the left-out energy distances to the original energy distances, we can find the percentiles
   3. Plots box plots, bar charts, and a histogram