Post Sim Run

This takes the simulation output and does some maths to get it to look exactly like the GPS output. This is intended to allow both sim and gps datasets to run trhough the same code so that 1:1 comparisons can be made.

Config and Functions

Read Sim Output

Read several CSV files that have been created using the "make benchmark" command. These files are created using a couple of different config files. The goal here is to compare the different parameters in the benchmark config files.

For this experiment one month of data was generated using different "movement methods" for each run. The different movement methods are:

- Random Movement: No behaviour states, no landscape interaction, just basic step and turn movement
- HMM Movement: Multiple behaviour states that are set the step and turn parameters. Transistions between states is a result of the landscape data.
- DLD movement: Complex movement influenced by behaviour states, home ranges, landscape info etc

The csv's are read into a dataframe in chunks and then saved to a parquet file. This should allow larger simulations to be processed on a smaller machine (like my laptop) without running into memory errors... up to a point.

Convert the parquet file into a geopandas dataframe and eventually a geopackage. This will allow it to be easily added to a QGIS project

Read GPS Data

Once all the preprocessing has been done, lets read the smaller/compressed parquet file into memory,do some analysis, and start plotting it.

Calculate HomeRange Timeseries

How much area is covered every week/month/year? Is it similar between the sim and GPS data?

Let's take both the GPS data and SIM data and group it by Agent ID, Month and, in the case of the simulation data, the Filename of the data. This is in order to keep the different benchmark results seperate.

So now that the homeranges are written to geopackage files, they can be visualised in QGIS. Here's a quick example:

Homerange sizes

Are the monthly home range areas similar to the GPS data? Let's plot the monthly home range area (convex hull of GPS/Simulated points per agent, per month). This homerange area is not specified in the model parameters and so "should" be similar if the model is behaving as expected.

The above figure shows that the simulated data is, obviously, more regular than the GPS data. There are a few GPS devices that have far fewer samples in a given month than expected. These few GPS points would alter the expected monthly home range size. Let's drop all rows that have a points/month that is too low.

The above figure shows that the GPS data home ranges are much smaller than any of the modelled home ranges... Not a good result...

The above figure shows that the male/female homeranges in the GPS data is significantly different. It would therefor make sense to model the male/female movement parameters differently.

Those same step and turn calcs again

Below is the method used to calculate step and turn values from a timeseries of points. It must be known that the values calculated are often for the previous points, so when looking at it with agents mixed up it'll just look plain wrong

Look at Movement Params

Similar to the initial data analysis, lets take a look at the step and turn distributions and then some other statistical measures like home range size.

Note:

There is a spike at 0 degrees turn angle, this happens when the agent slides up against the edge of the grid boundary, the x/y of the grid value get's set to 0 (or max) and resulting turning angle becomes different from model's chosen turning angle (and is also often 0)

HMM Step and Turns vs Behavioural State

Let's compar the step and turn vs behavioural state for a specific run

Timeseries Plots