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

	timestamp	ID	x_proj	y_proj	x	y	step	t
0	2000-01-01T01:00:00	17a57649-35a3-4933-bba8-c509fae60855	1.619439e+06	1.967883e+06	114.813624	171.391292	12.698136	0.498
1	2000-01-01T01:00:00	b6d3dc59-cacf-46be-b198-1a715b0e8d97	1.616058e+06	1.967522e+06	2.090985	183.421964	8.271562	0.899
2	2000-01-01T01:00:00	ddf707ce-1585-4411-8aff-9fd8a239d417	1.617639e+06	1.969132e+06	54.792766	129.758149	37.628007	2.971
3	2000-01-01T01:00:00	df5333e6-f4c5-4d39-85dc-3e80c1c734ba	1.616515e+06	1.969465e+06	17.324505	118.650867	18.390228	0.136
4	2000-01-01T01:00:00	ceb632dc-dfff-4849-97a1-53fb03b0dcf8	1.618858e+06	1.967483e+06	95.439107	184.727471	35.841659	0.357

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:

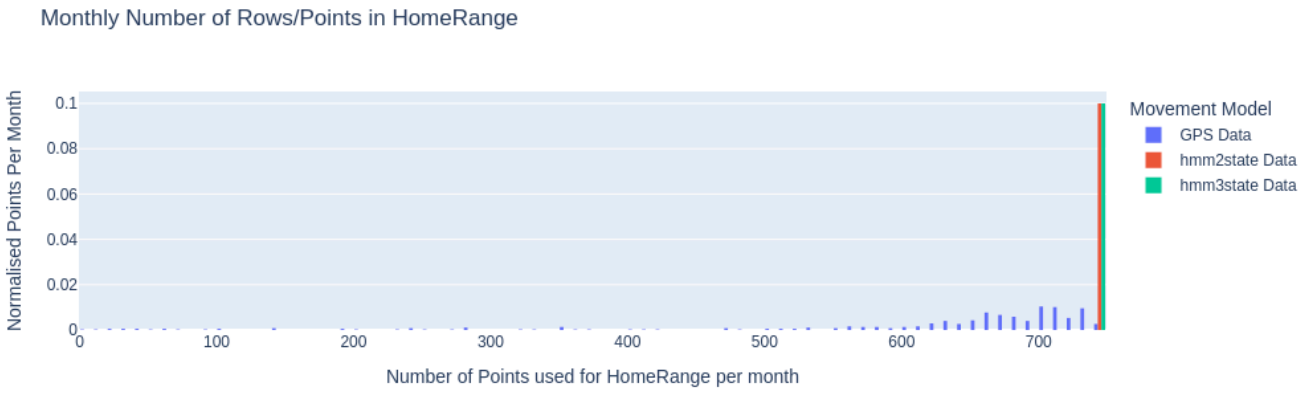
Homerange sizes

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

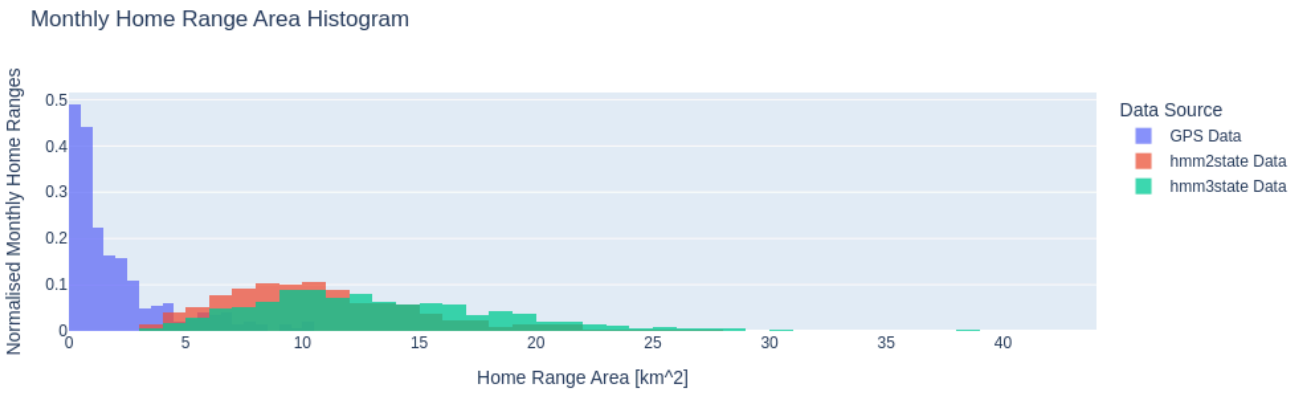
model parameters and so "should" be similar if the model is behaving as expected.

	ID	time_group	filename	0	samples_in_group	area
0	002ea5cf-21f5-4dfb-b1f6-702a81a61b3c	2000-01	hmm2state	POLYGON ((1628047.108 1966287.066, 1628034.074...	743	5.241009
1	005da896-da1c-4a7a-a608-81ff02778762	2000-01	hmm2state	POLYGON ((1621802.727 1968012.381, 1621604.35 ...	743	17.470278
2	0062db91-94bb-4cdd-92b4-7114960503c6	2000-01	hmm2state	POLYGON ((1618724.473 1964326.021, 1618217.709...	743	13.868732
3	0077c72d-6aa5-44c1-b00e-50b985dfb977	2000-01	hmm3state	POLYGON ((1618924.673 1961985, 1617941.26 1962...	743	7.477996
4	00c1e214-6b5a-4d72-8f91-5dc6550a98c3	2000-01	hmm3state	POLYGON ((1619237.008 1967943.974, 1618278.335...	743	15.881053
...	...	...	...	...	...	...
1995	ff32202a-a69b-476e-80aa-b6745cd640ca	2000-01	hmm2state	POLYGON ((1619090.739 1970248.322, 1618418.447...	743	7.522238
1996	ff41a61a-fe2a-47dd-9097-914dd0db8e7e	2000-01	hmm3state	POLYGON ((1624577.195 1962537.452, 1623971.075...	743	8.987130
1997	ff5c3c5f-dea7-4ac3-8196-967408a9392a	2000-01	hmm2state	POLYGON ((1617232.505 1970462.963, 1616391.988...	743	3.867634
1998	fff4a387-eddf-4d9e-afab-ecdb3abc031c	2000-01	hmm3state	POLYGON ((1617927.401 1967765.361, 1617774.868...	743	7.539432
1999	fffb3fdb-133d-4b76-94a3-503089e9a872	2000-01	hmm3state	POLYGON ((1619529.85 1966769.778, 1618305.6 19...	743	16.768998

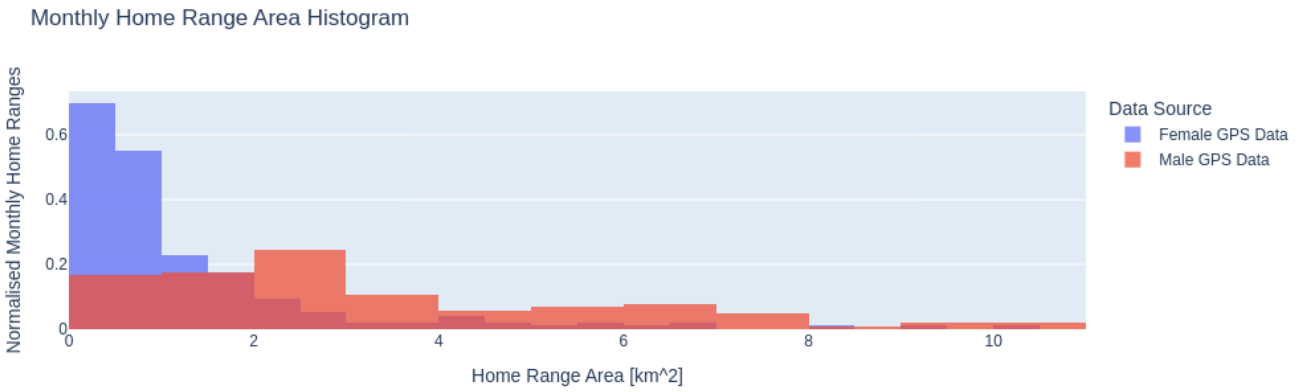
2000 rows × 6 columns



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...



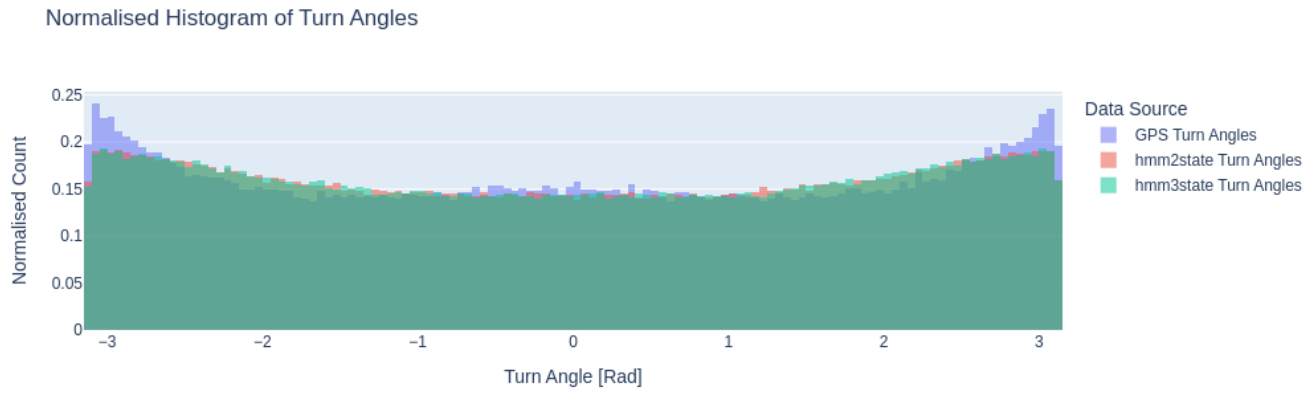
The above figure shows that the male/female homeranges in the GPS data is significantly different. It would therefor make sense to model the male/female movement parameters differently.

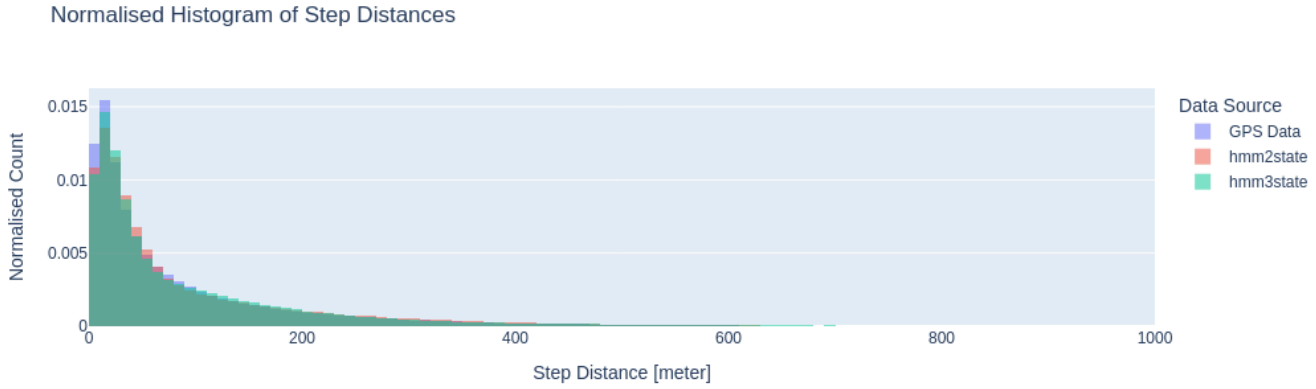
Those same step and turn calcs again

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.



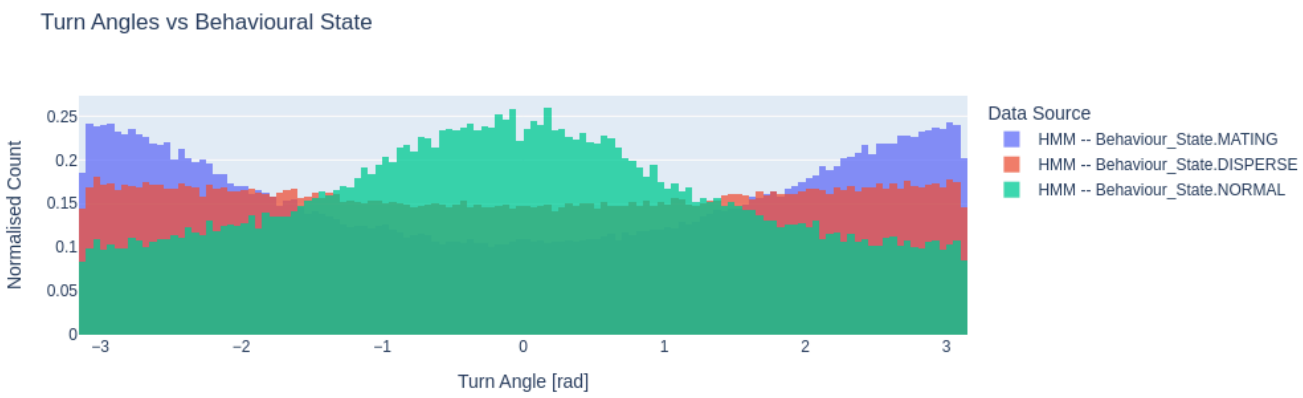


HMM Step and Turns vs Behavioural State

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

	filename	turn	turn_angle	step	step_distance
1000	hmm3state	-3.010763	1.714190	27.681707	7414.549560
1001	hmm3state	-3.139172	1.601724	31.457653	1290.042172
1002	hmm3state	1.023445	-1.642376	88.784482	2662.589871
1003	hmm3state	3.043519	1.506348	118.639849	5541.025398
1004	hmm3state	-2.912987	-3.030573	14.550588	1476.831121
...	...	...	...	...	...
741995	hmm3state	2.287063	1.908374	59.927193	5954.180329
741996	hmm3state	-1.131049	1.671017	488.279928	649.372515
741997	hmm3state	-0.124254	2.820899	133.917890	1730.981165
741998	hmm3state	0.336376	-1.789185	486.673241	67.973516
741999	hmm3state	-1.144727	-2.770909	17.664766	70.619251

741000 rows × 5 columns



	timestamp	ID	x_proj	y_proj	x	y	step
1000	2000-01-01 02:00:00	17a57649-35a3-4933-bba8-c509fae60855	1.619413e+06	1.967891e+06	113.935949	171.141013	27.379877
1001	2000-01-01 02:00:00	b6d3dc59-cacf-46be-b198-1a715b0e8d97	1.616078e+06	1.967553e+06	2.782393	182.394985	37.141063
1002	2000-01-01 02:00:00	ddf707ce-1585-4411-8aff-9fd8a239d417	1.617557e+06	1.969237e+06	52.059904	126.260046	133.171825
1003	2000-01-01 02:00:00	df5333e6-f4c5-4d39-85dc-3e80c1c734ba	1.616515e+06	1.969503e+06	17.341492	117.410357	37.218776
1004	2000-01-01 02:00:00	ef0608dd-5dac-494a-ad68-01bdc8e80b58	1.617842e+06	1.972655e+06	61.565713	12.325300	132.028720
...	...	...	...	...	...	...	...
741995	2000-01-31 22:00:00	b7d70e33-e232-4fc2-9633-6e9c4d618d2a	1.626012e+06	1.967498e+06	333.887603	184.232528	59.927193
741996	2000-01-31 22:00:00	4696fdb1-376d-487a-bdfd-1e2c3a65237a	1.624345e+06	1.965321e+06	278.335622	256.785682	488.279928
741997	2000-01-31 22:00:00	9265e40b-da73-4e39-905f-e4a9d3814bbd	1.625067e+06	1.965600e+06	302.394004	247.503196	133.917890
741998	2000-01-31 22:00:00	5042b581-da36-4760-9ea2-d7b5e54cd2a8	1.629015e+06	1.966345e+06	434.000000	222.679892	486.673241
741999	2000-01-31 22:00:00	3c3431dd-fd64-44a8-b83f-6d07465889e1	1.624964e+06	1.964797e+06	298.963790	274.265486	17.664766

1482000 rows × 20 columns

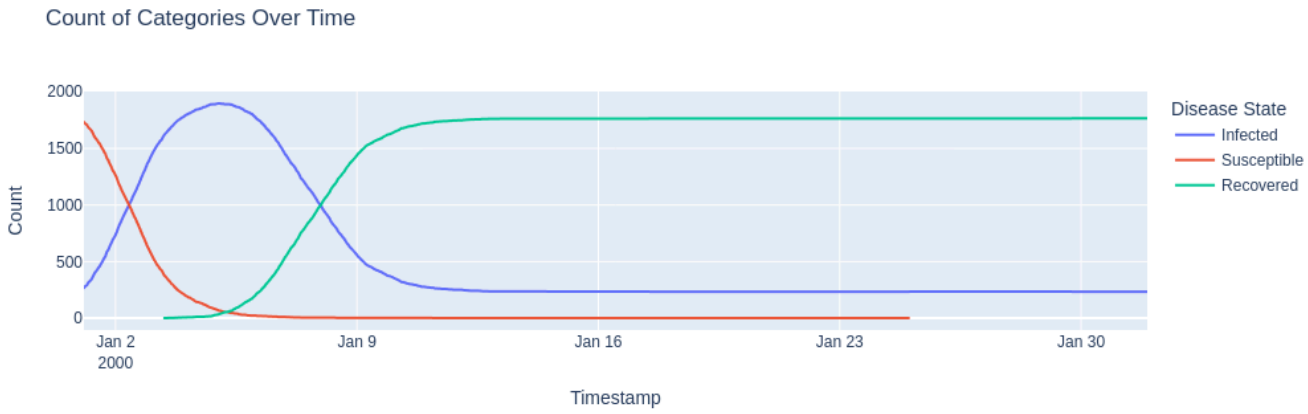
	turn	turn_angle
1714	1.519331	-2.588156
2713	-2.588156	2.346602
3713	2.346602	-3.098436
4710	-3.098436	-0.695807
5711	-0.695807	-2.492587
...	...	...
737836	2.774876	0.207339
738834	0.207339	-0.209262
739999	-0.209262	0.956633
740998	0.956633	0.266754
741998	0.336376	-1.789185

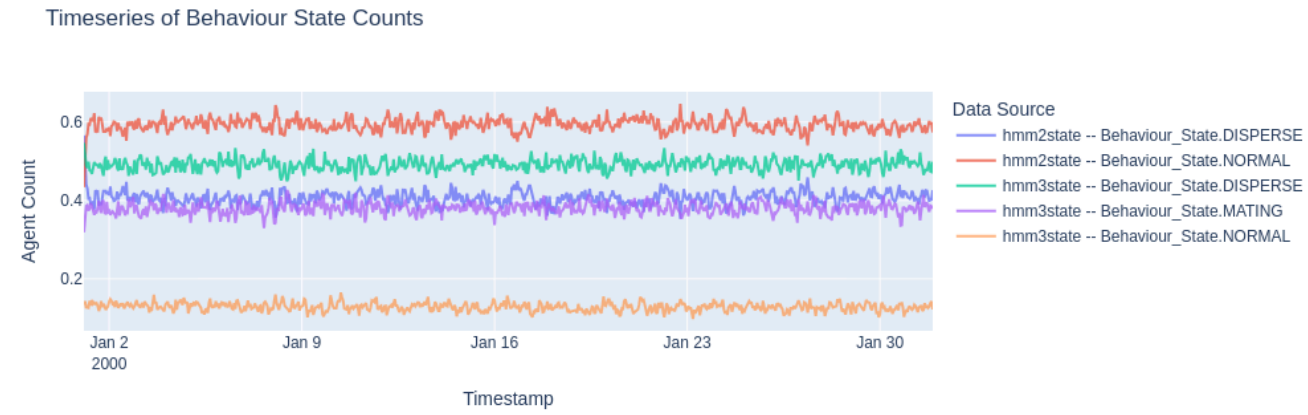
741 rows × 2 columns

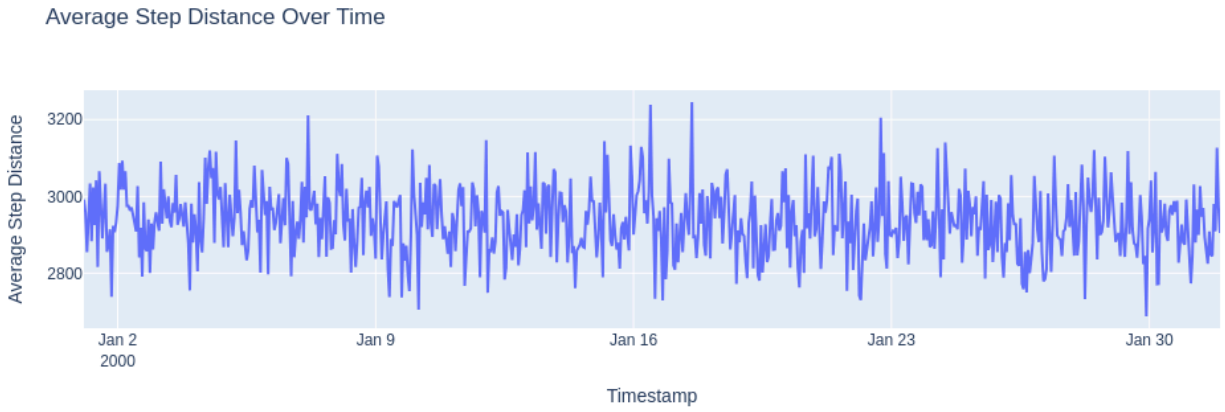
Timeseries Plots

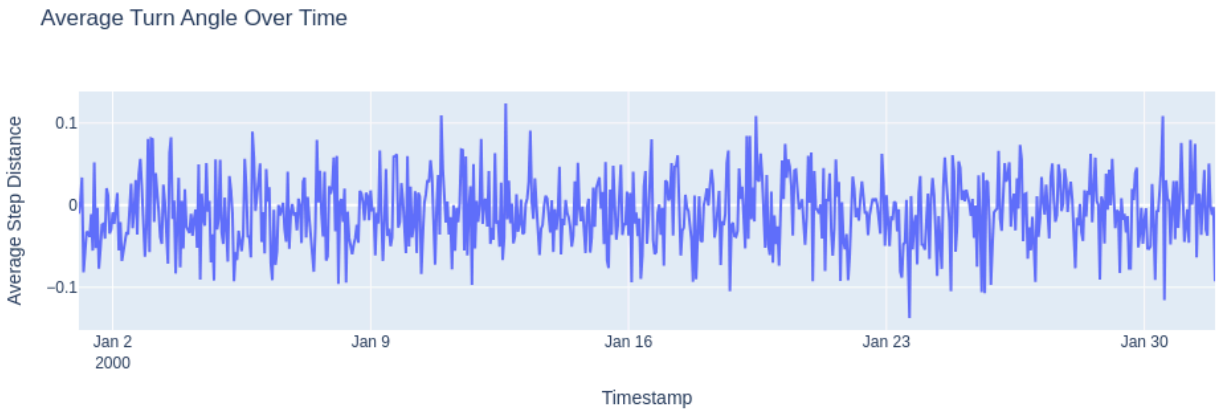
	step	step_distance
1000	27.379877	2085.102632
1001	37.141063	5577.152529
1002	133.171825	6669.809621
1003	37.218776	366.932380
1004	132.028720	5672.580143
...	...	...
741995	59.927193	5954.180329
741996	488.279928	649.372515
741997	133.917890	1730.981165
741998	486.673241	67.973516
741999	17.664766	70.619251

1482000 rows × 2 columns


Count of Categories Over Time
Count
2000
1500
1000
500
0
Jan 2
2000
Jan 9
Jan 16
Jan 23
Jan 30
Timestamp
Disease State
Infected
Susceptible
Recovered


Timeseries of Behaviour State Counts
Agent Count
0.6
0.4
0.2
Jan 2
2000
Jan 9
Jan 16
Jan 23
Jan 30
Timestamp
Data Source
hmm2state -- Behaviour_State.DISPERSE
hmm2state -- Behaviour_State.NORMAL
hmm3state -- Behaviour_State.DISPERSE
hmm3state -- Behaviour_State.MATING
hmm3state -- Behaviour_State.NORMAL


Average Step Distance Over Time
Average Step Distance
3200
3000
2800
Jan 2
2000
Jan 9
Jan 16
Jan 23
Jan 30
Timestamp




```
[NbConvertApp] Converting notebook 04-Post ABM Run.ipynb to webpdf
[NbConvertApp] WARNING | Alternative text is missing on 10 image(s).
[NbConvertApp] Building PDF
[NbConvertApp] PDF successfully created
[NbConvertApp] Writing 395291 bytes to 04-Post ABM Run.pdf
```