

Final Project Write-Up

Grace Chang

Inferring Animal Densities from Tracking Data Using Markov Chains

The paper can be accessed [here](#).

1 Paper Description

1.1 Biological Context

Currently, animal tracking is done by capturing an animal, tagging it, and then releasing it into the wild to monitor movement patterns. From this information, we can glean some information about habitat use, foraging patterns, behavior, social structure, life history, and response to environmental changes. This information is value to both the study of and the preservation of animals. Movement data can also help businesses pick locations that do not interfere with the natural habitat.

1.2 Problem

The current problem with animal tracking is that with every tagging, there is a release bias, as the animals are consistently released from the same location. This means that all of the movement data is always biased as all of the animals start from the same location. There is unfortunately no current bias removal technique and as a result of this bias, it is hard to analyze the data and differentiate between what is and what isn't a result of the release bias.

1.3 Argument

The paper argues that Markov Chains are a lucrative tool in improving animal tracking methods. This is because the movement patterns of animals can be thought of as having a certain probability of moving from one place to another. The authors explain that Markov Chains are superior in tracking uses because it can better replicate the movement patterns while also negating starting position bias that are inherent in location-based animal studies. This method can be done over a geographic grid, water depth, or habitat type.

1.4 Model

1.4.1 Assumptions

1. Animals that are tracked are a random subset of population in terms of movement behavior.

2. The movements form a time-homogenous Markov chain in the that probability of an animal moving from i to j at time p is independent of time.

1.4.2 Theory

$$\begin{aligned}\pi_i &= \sum_j \pi_j p_{ji} \\ \pi_i &= P(\text{Animal randomly chosen from population is in cell } j \text{ at any time}) \\ p_{ji} &= P(\text{Animal moves from } j \text{ to } i)\end{aligned}$$

Assuming that $\pi = \{\pi_i\}$ and where $P = \{p_{ji}\}$,

$$\pi = \pi P.$$

We estimate by

$$\begin{aligned}\hat{p}_{ji} &= \frac{n_{ji}}{\sum_j n_{ji}} \\ n_{ji} &= \text{Number of steps taken from cell } j \text{ to cell } i \\ \sum_j n_{ji} &= \text{Total number of steps for all positions.}\end{aligned}$$

Where π is the left eigenvector P that corresponds to the eigenvalue of 1.

1.5 Experiment

It compares a standard track density method with a Markov density method by using both a homogenous density setup and a quadrant based setup with different weights on each quadrant. The paper consists of two main components: simulated path data and real sperm whale tracking data. It then uses both of these datasets to show that Markov Chains are more effective. I will be focusing on the simulated path data.

1.5.1 Data Creation

The data was done with 1,000 steps and 300 agents. It the observation space was one unit square divided into 6 different cells. One cell was used to represent all observations outside of the observation area.

They created four types of data:

1. Random starting position with move length 0.05
2. Random starting position with move length depending on quadrant
3. Random starting position within 0.1 units of the center with move length 0.05

0.05	$0.05/\sqrt{1.5}$
$0.05/\sqrt{0.5}$	0.05

Table 1: Weights of the Quadrants

4. Random starting position within 0.1 units of the center with move length depending on quadrant.

Analysis Of these four types of data there was Track Density, the relative density from summing the number of track positions in each cell, and Markov Density, the density using eigenvalue model.

2 My Methods

2.1 Data Creation

My data creation was done as close as possible from what was stated in the paper. A breakdown of the steps is located in the appendix. To sort the columns and rows, I created the equation

$$\frac{1}{6}(\text{column number}) - \frac{1}{6} \leq \text{x-coordinate} < \frac{1}{6}(\text{column number})$$

$$\frac{1}{6}(\text{row number}) - \frac{1}{6} \leq \text{x-coordinate} < \frac{1}{6}(\text{row number})$$

Where I used this numbering for the cells:

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36

To address the outside cells, I created a separate vector that held a count for all of steps in there.

3 My Results

I ended up creating 8 heat maps total: four for both Tracks and Markov. However, I ended up finding an issue between my results and the paper's results, so

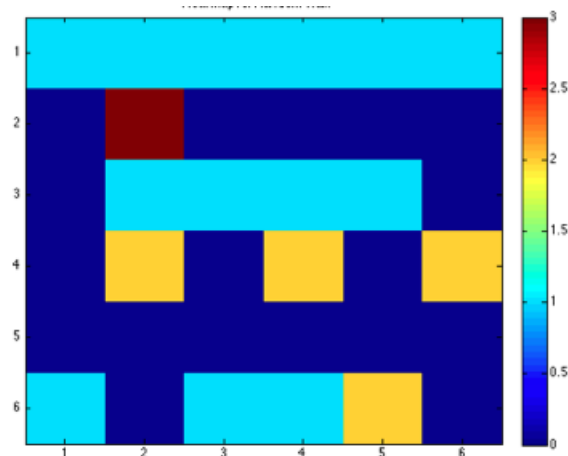


Figure 1: Heatmap for Random Starting Position With Move Length 0.05

I will discuss the first heat map in detail and then show the rest of my results in the appendix.

The following is the heat map for the random starting position with move length 0.05.

If this is compared to the paper’s results, which are located in the appendix. It is apparent that it looks nothing like the track density in the paper. While their track destiny displays a nice gradient of blues and greens, my heat map displays a disjointed, multicolor, heat map of chaos. Upon further analysis, it becomes evident that there are several issues within the paper.

3.1 Paper Rhetoric

Below is an image of the description of the methods illustrated in the paper. As indicated in my methods, I have simulated as close as possible to the methods discussed in the paper. However, the paper glosses over the details of the particular steps used to create the simulation. Most importantly, there is talk of how the “...estimates of the relative densities were normalized to have the same mean as the ideal densities, excluding the exterior cell.” However, there is no information included regarding the ideal densities, meaning that it is actually impossible to replicate the paper because the results include steps and information not given by the paper.

3.2 Heat map Weight

After observing that my heatmaps were starting to become very different from the rest of the paper’s heatmap, I looked critically at the different heatmap provided by the paper. The heatmaps of the paper only show red a few amount

1	1	1	1	1	1	
0	3	0	0	0	0	
0	1	1	1	1	0	
0	2	0	2	0	2	with exterior = 276
0	0	0	0	0	0	
1	0	1	1	2	0	

Figure 2: Data for Heatmap

of times, and the rest of the heatmaps show no red. Thus, we know that the paper is creating the heatmap where the heaviest weight is red, and that most of the heatmaps are not displaying the 'heaviest' value. In the paper, there is also talk about an exterior cell, where all of the steps that venture out of the observation area are contended. We also see that, from Figure 1, the ideal densities were not used to normalize the exterior cell. Figure 2 is the data that I created for the above heat map:

As we can see here, the exterior cell, which seems to also be included in the heatmap, has a very heavy weight. To replicate what seems to be done by the authors, I attached the exterior cell into the corner of the paper and then created a heatmap, shown in Figure 2. It seems that once the exterior cell is taken into account, it dominates the whole heatmap. The other questionable part of this paper is the lack of a heatmap bar. Heatmap bars are very crucial in the reading of a heatmap so this omission makes it even more difficult to figure out the authors' general process in creating this heatmap. The discussion of the exterior cell implies that they weighted it, but despite that, the results still did not match the data that was created by following every step dictated. For the rest of the heatmaps and weighted heatmaps, please refer to the Appendix.

3.3 Lack of Information

Not only does the paper not provide the idea densities, but it also does not provide any of its data. In addition, it does not provide the heat map weight. The MATLAB code that they provided was incomplete and is not particularly described. The only information that I was able to glean from the code was that they did the Markov density the same way that I did. If we refer to the second experiment in the paper, we see that the explanation is just as, if not even more, convoluted as the first paper. If the second part was detailed, this might imply a obvious glaring error on my part. However, the lack of detailed explanation throughout the paper means that there is most likely an overall deficiency of information in this paper.

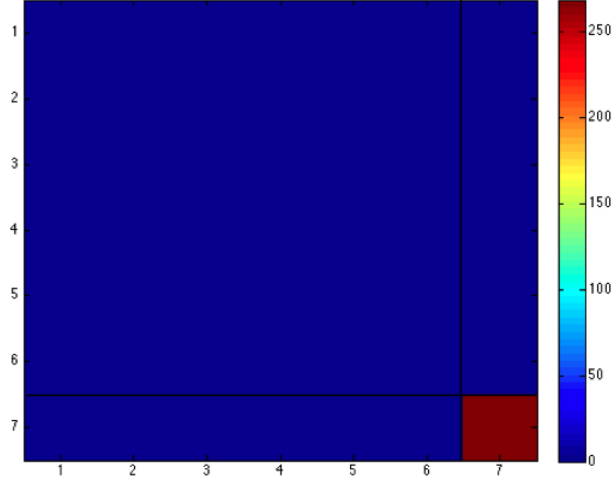


Figure 3: Scaled Heatmap for Random Starting Position With Move Length 0.05

3.4 Experiment Creation Issues

The paper says that the study area is a 1×1 grid. With each step being around 0.05, this means that an animal would only need to traverse 20 steps to leave the observation area. This means that over 300 agents and 1,000 steps, it is highly likely for the animals to leave. Probability speaking, the outside space that spans infinitely versus the 1×1 grid will eventually mean that all of the animals will leave in the end. This results in a much smaller dataset to work with. In this particular case, it is extremely detrimental because probability models, such as a Markov chain, are not as effective with smaller numbers where the difference is just one or two.

4 Conclusion

4.1 Data

Because of the difficulty in comparing results, it is hard to have a conclusion regarding the experiment. However, the idea of using a Markov chain to predict the movement patterns by using probabilities is a very interesting idea that could be full of potential. The model could prove to be extremely compelling if there is further analysis on better data.

4.2 Experiment Creation

There was a glaring error in the authors' creation of this experiment, as most of the sample size leaves after a while. Thus, I learned to be extremely careful and to think critically about the experiments I could make.

4.3 Paper Creation

When creating a paper, be extremely detailed and provide all data used. This minimizes and sources of conflict and will help make the paper better understood. The paper should have enough data and detailed enough steps for the paper to be replaceable, or else people could respect the paper less, despite possible innovation illustrated in the paper.

5 Appendix

5.1 Methods for Creating Track Density

1. Create matrix: 2 columns, # agents row
2. Generate random starting point
3. Input final position in matrix
 - (a) Roll to move x-direction or y-direction
 - (b) Roll to move forward or backward
4. Reassign coordinates to grid number (1-36)
 - (a) Created empty 6x6 matrix
 - (b) Loop across agents, rows, and columns
 - (c) If fulfilled both those conditions, then would +1 to 6x6 matrix (sum them)

5.2 Methods for Creating Markov Density

1. Create matrix: 2 columns, # agents row
2. Generate random starting point
3. Input final position in matrix
 - (a) Roll to move x-direction or y-direction
 - (b) Roll to move forward or backward
 - (c) Created count matrix each time move into cell
4. Calculate p_{ij}

- (a) Summed count matrix for denominator
- (b) Divided count matrix by totals
- 5. Calculate Eigenvalue for Final Density
 - (a) Calculate eigenvalue for p_{ji}
 - (b) Make script to find the eigenvector corresponding to eigenvalue 1
 - (c) Normalized eigenvector
- 6. Split into grid

5.3 Individual Heatmaps

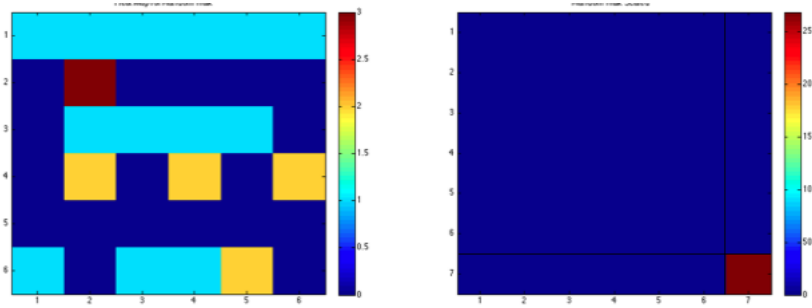


Figure 4: Tracks: Random 0.05

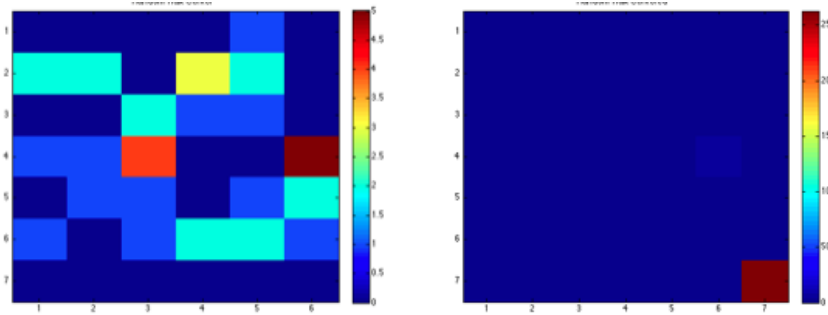


Figure 5: Tracks: Centered 0.05

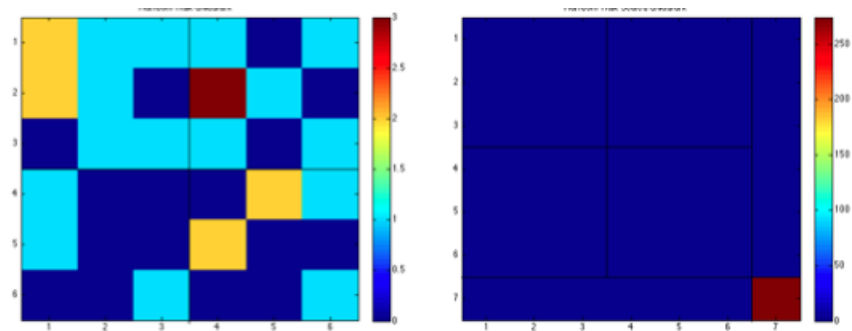


Figure 6: Tracks: Random Quadrant

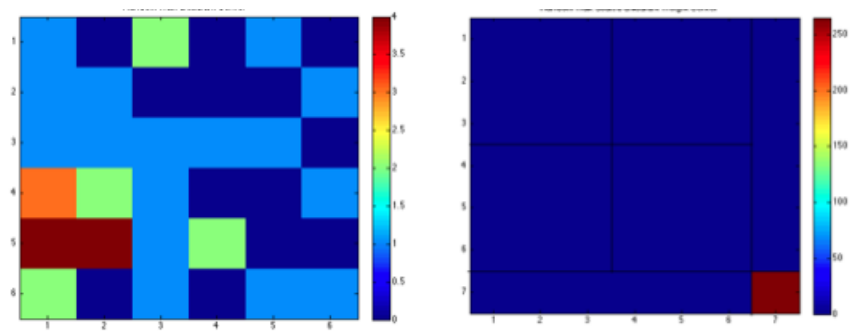


Figure 7: Tracks: Centered Quadrant

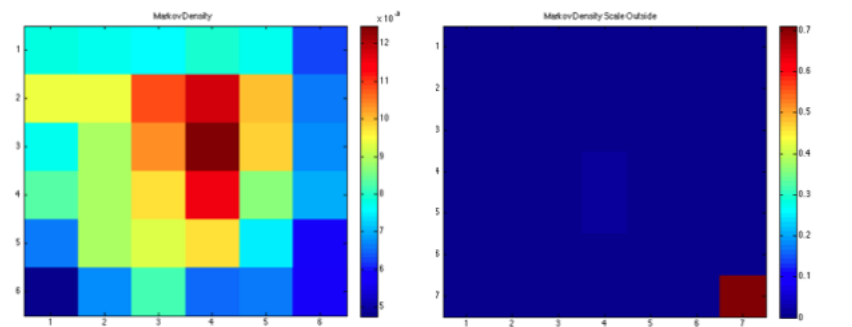


Figure 8: Markov: Random 0.05

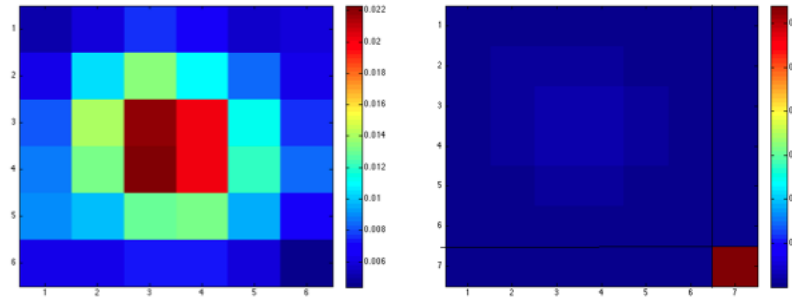


Figure 9: Markov: Centered 0.05

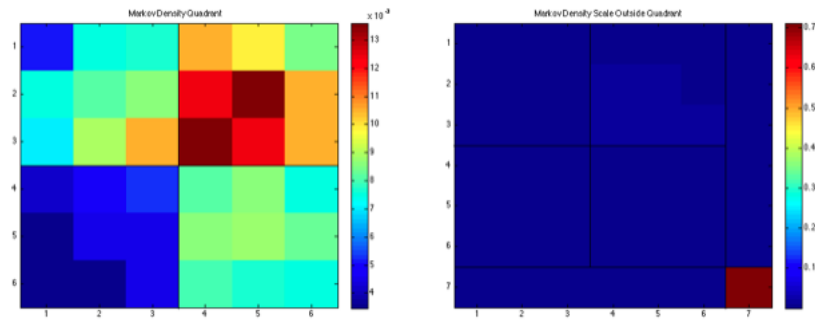


Figure 10: Markov: Random Quadrant

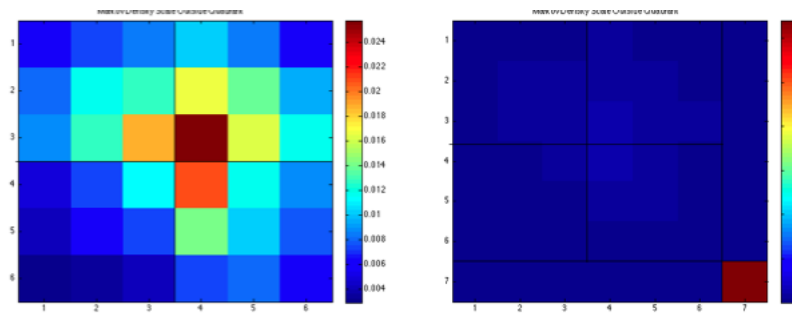


Figure 11: Markov: Centered Quadrant

5.4 Combined Heatmaps

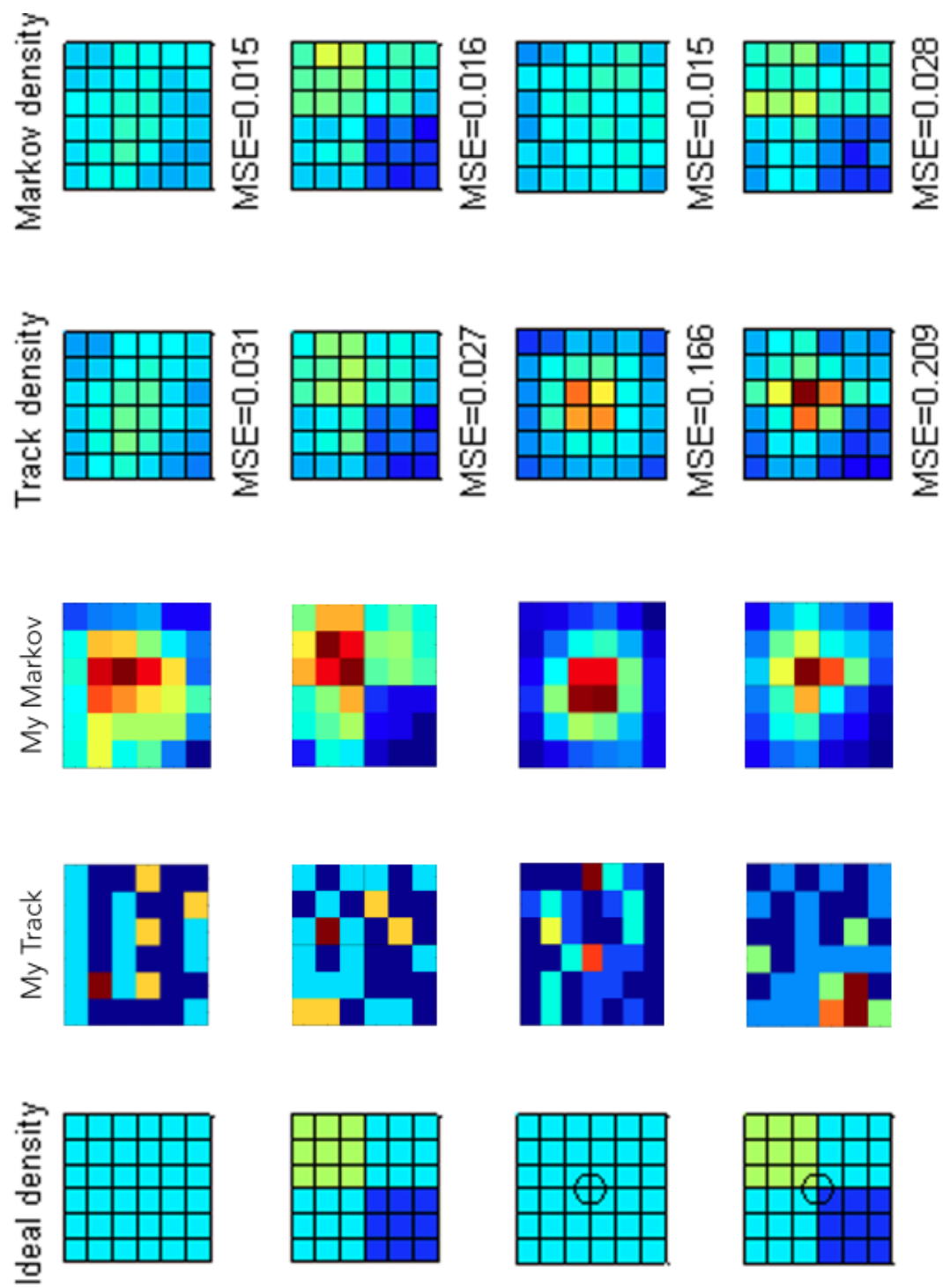


Figure 12: My Results and Paper's Results

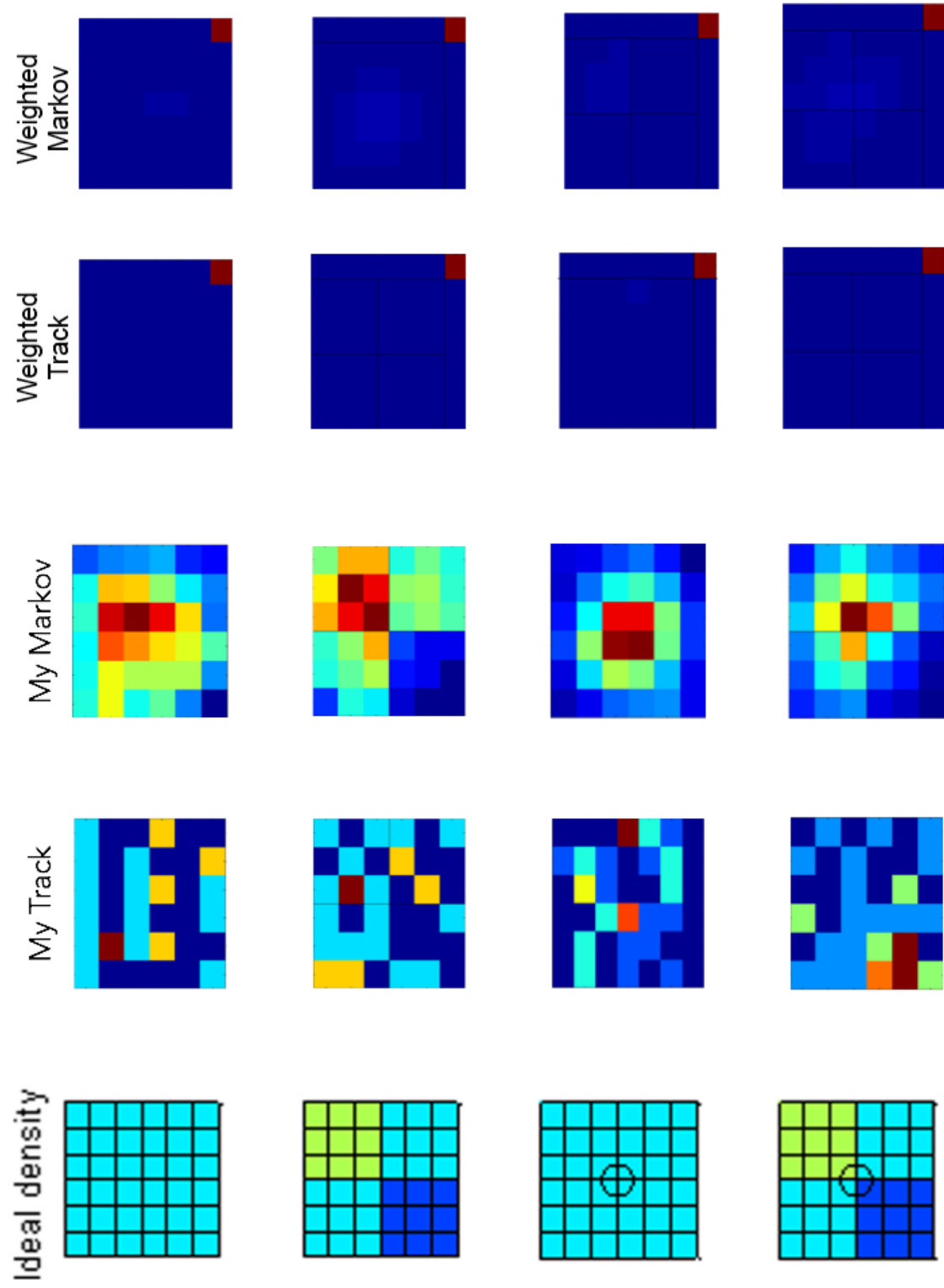


Figure 13: Weighted Heatmaps