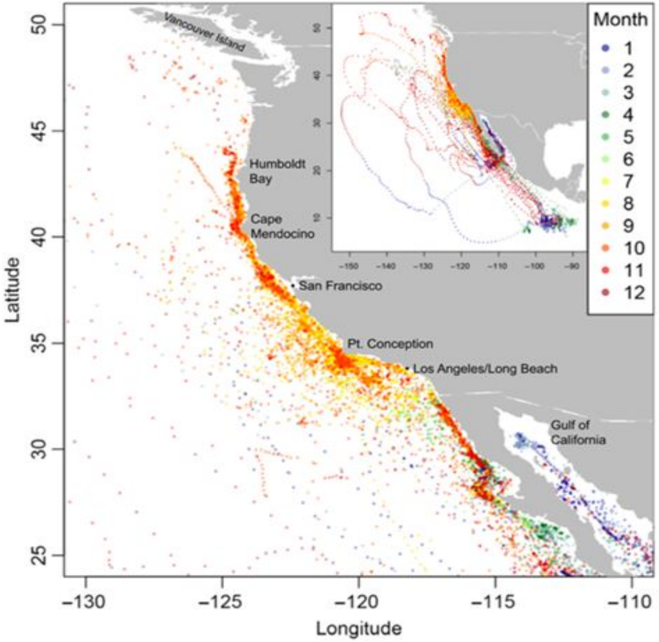
Recovering Individual Based Model Outcomes on Spatiotemporally Coarsened Data

Sameerah Helal

Blue Whales

Very social

Forage on krill



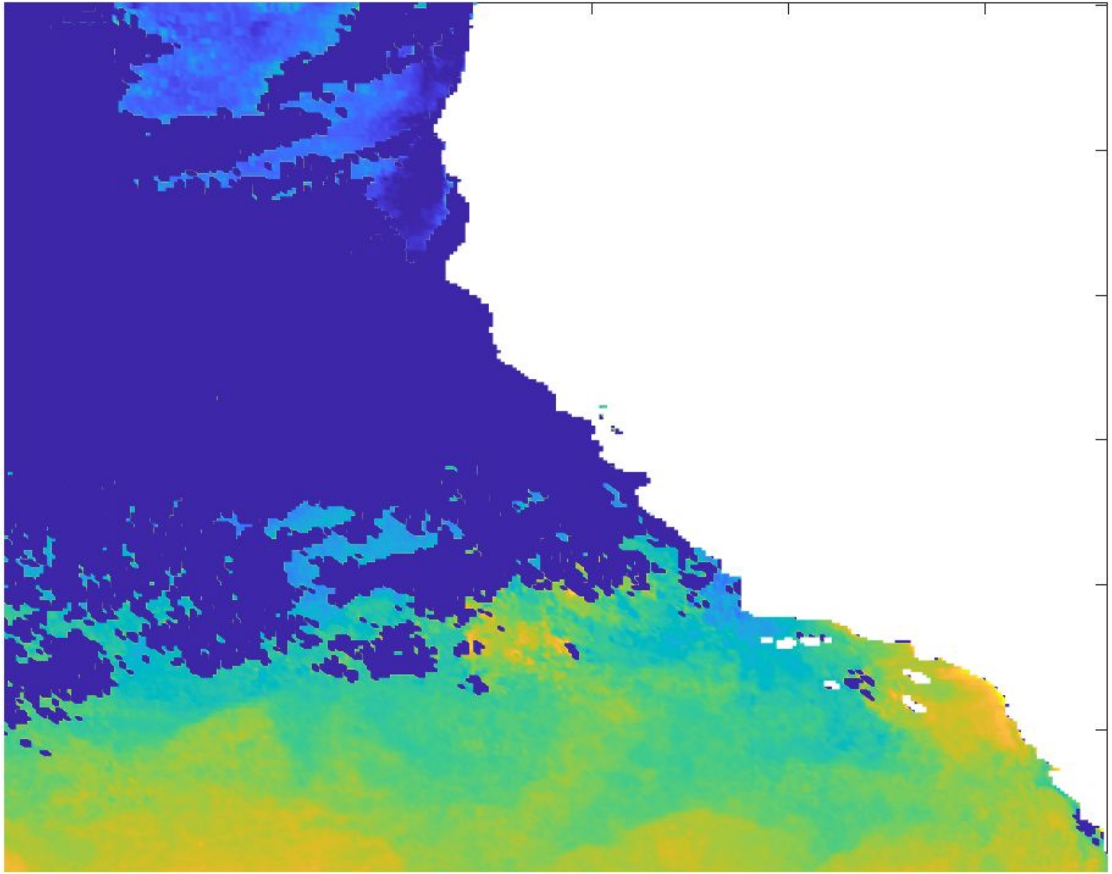
Seasonal migrators

Shipping lanes

Fishing gear

The Blue Whale IBM

- Individual based models are commonly used to study animal behaviours
- Process:
  - Input put environmental data into the IBM
  - The whales move through and interact with the domain based on the data
- When fed coarse environmental data, IBMs can produce inaccurate model outcomes
  - Can be caused by satellite’s movement, type of readings, and presence of cloud cover
  - Showing unrealistic behaviour or movement
- Goal of our research: develop modifications to coarse data and the IBM that will produce accurate model outcomes



Outline

1. How the Blue Whale IBM Works
2. The problem with coarse data
3. Algorithm and data modifications to mitigate the effects of coarse data
4. Results

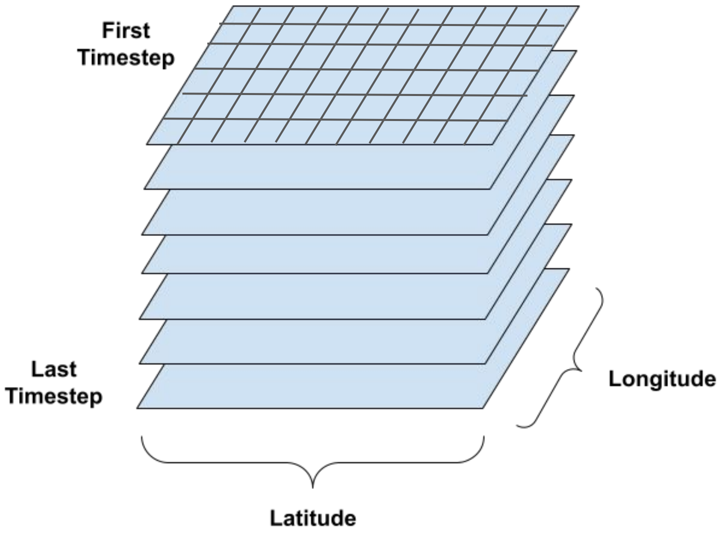
Inside the Blue Whale IBM

Data Dimensions

SST and krill are two 3 dimensional matrices of the same shape

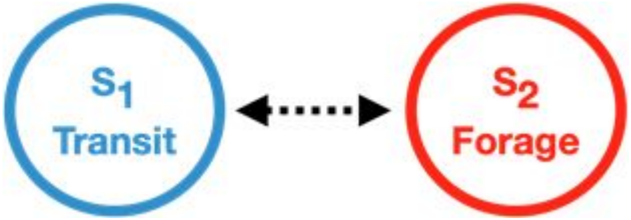
(domain width, domain height, number of timesteps)

Imagine the environmental data to be two dimensional maps stacked vertically, one for each time interval



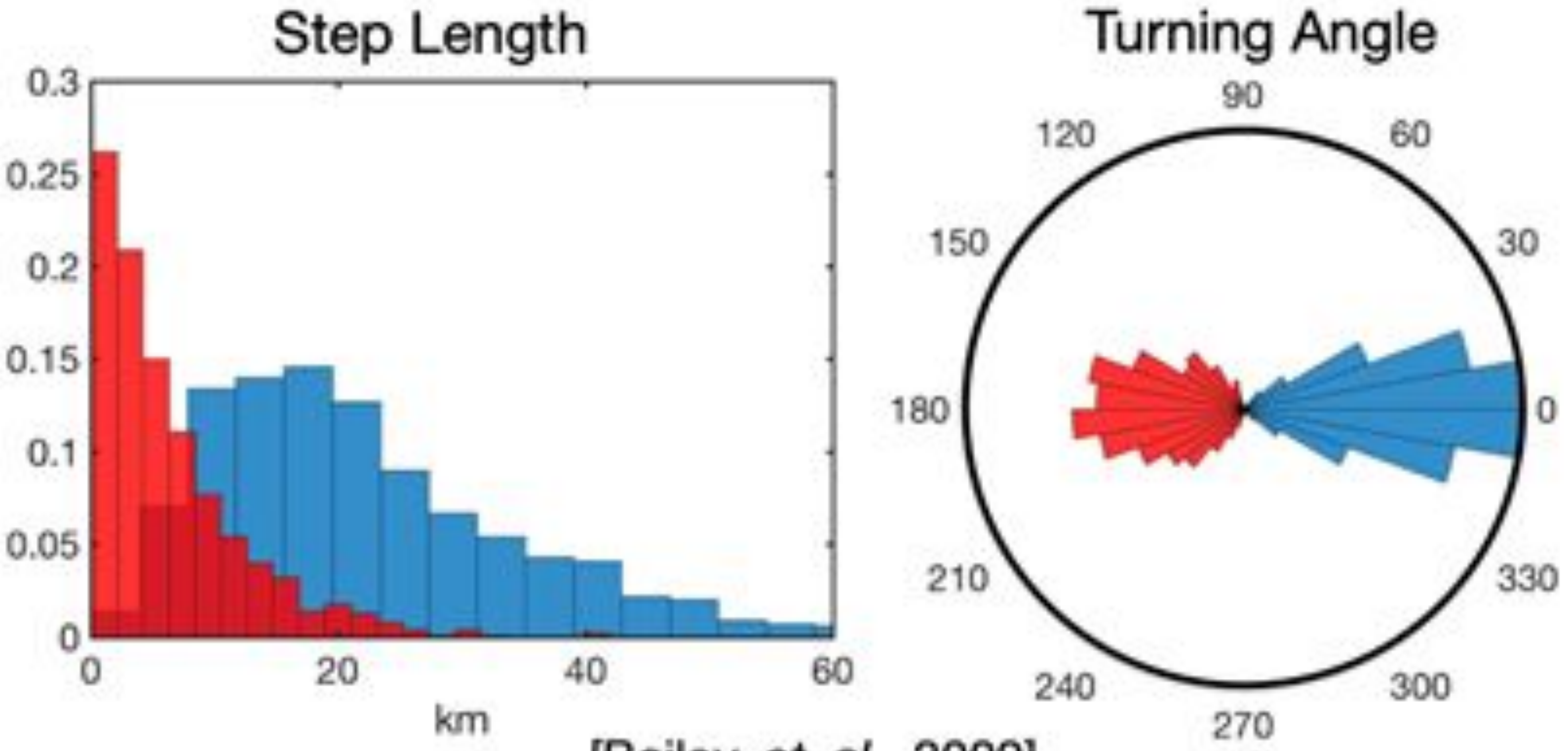
Whale Movement and State Switching

- Throughout the IBM, whales can occupy two two states: transit or foraging
  - Transit state (S1): travel through the domain
  - Foraging state (S2): where they “forage” for krill
- Whale state depends on local SST and krill
  - Lower SST and higher krill densities make a whale more likely to be in the foraging state
- Whale movement is constructed with step lengths and turning angles
  - Turning angle determines at what angle from its current trajectory a whale should turn
  - Step length determines how far in that direction to travel



Movement Distributions

- Step lengths and turning angles drawn from Gamma and Von Mises distributions
- Distinct distributions defined for whales in the foraging versus transit state
- At every location update, each individual reads its local SST and krill conditions, which influence the probability of foraging



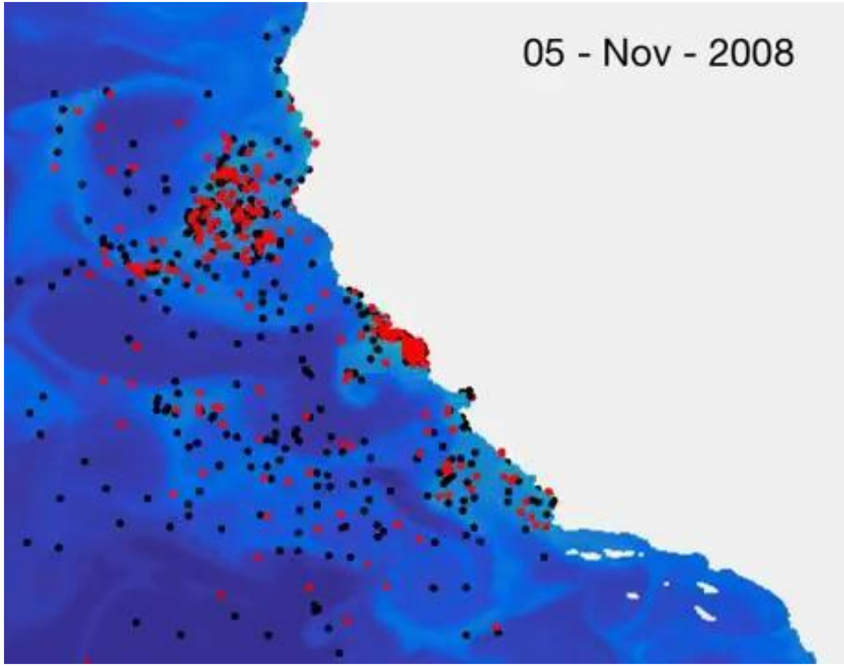
[Bailey, *et. al.*, 2009]

Summary

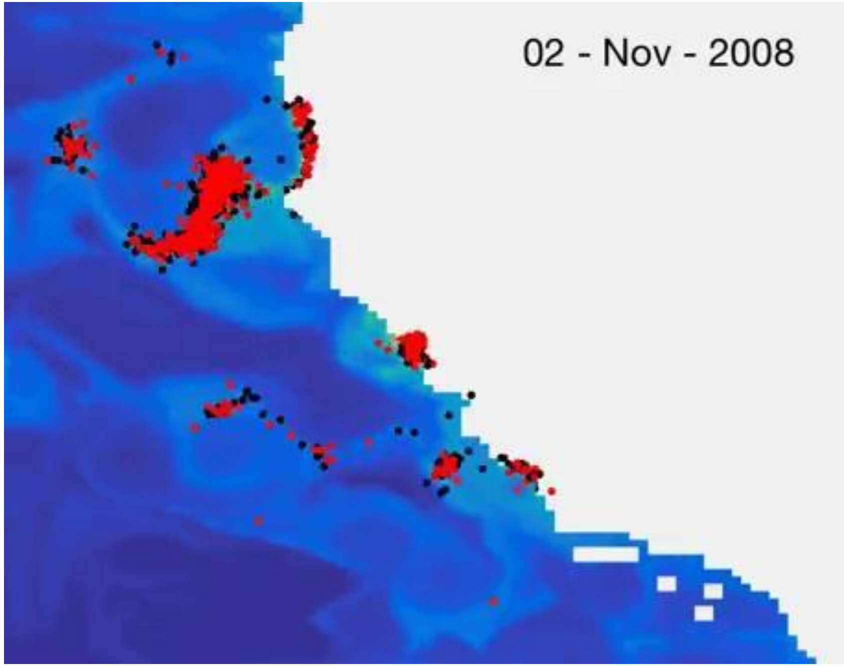
- IBM initiated on May 1st in the southern portion of the domain with 2000 whales
- At each update, the whales select:
    - A movement distance and direction based on the movement distributions and their current location
    - A foraging state based on their environmental conditions.
- Update cycle repeats until the end of the simulation

The Problem with Coarse Data

Spatially Coarse Data

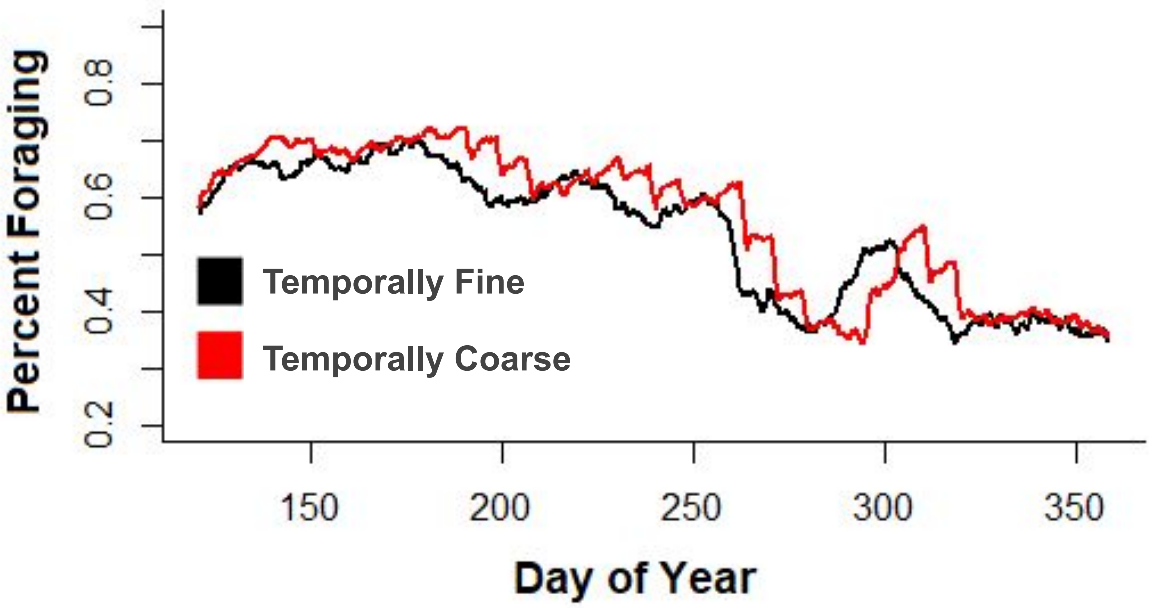


Spatially fine data: whales spread out



Spatially coarse data: whales cluster

Temporally Coarse Data

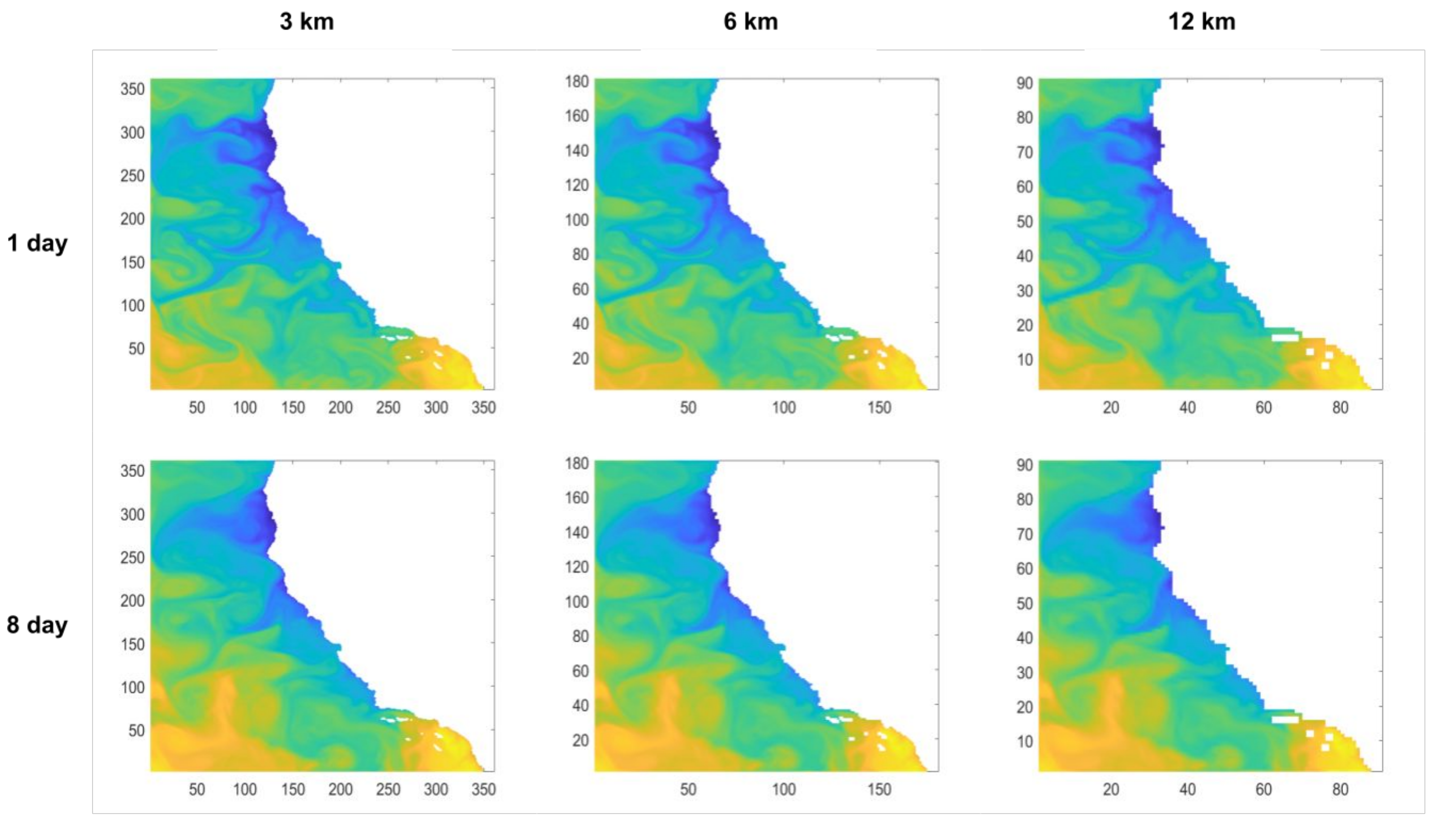


Temporally coarse data causes large, abrupt changes in the percentage of whales foraging at the end of each time-slice

Solution

Coarsening the Data

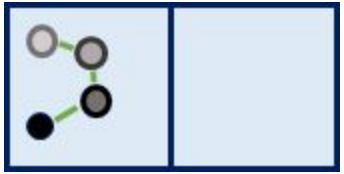
- Started with a real dataset: ROMS Data from 2008 (Gold Standard Data)
  - Gold standard: high resolution: 3 km spatial resolution and 1 day temporal resolution
- We manipulated the ROMS data to mimic the data collection shortcomings of satellites
- Temporal coarsening by averaging the data over time
  - Average the data in every n time-slices
  - Compress every n slices into a single slice
- Spatial coarsening with interpolation (interp3 in Matlab)



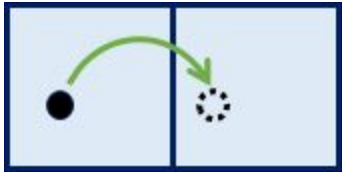
3 km
6 km
12 km
1 day
8 day

Model Adaptations: Spatial

- To mitigate the issues from spatially coarse data, we adjusted the rate at which the whales updated their locations and behavioural states
- In the gold standard model, whales take 4 steps per day
  - Step lengths selected from a Gamma distribution with parameters μ_0 and σ_0
- Solution: scale this distribution by a the number of hours in a timestep
  - Accommodates for the increased size of grid cell caused by coarse spatial data
- From the default IBM rate of 4, we changed our rates to 1 and 2



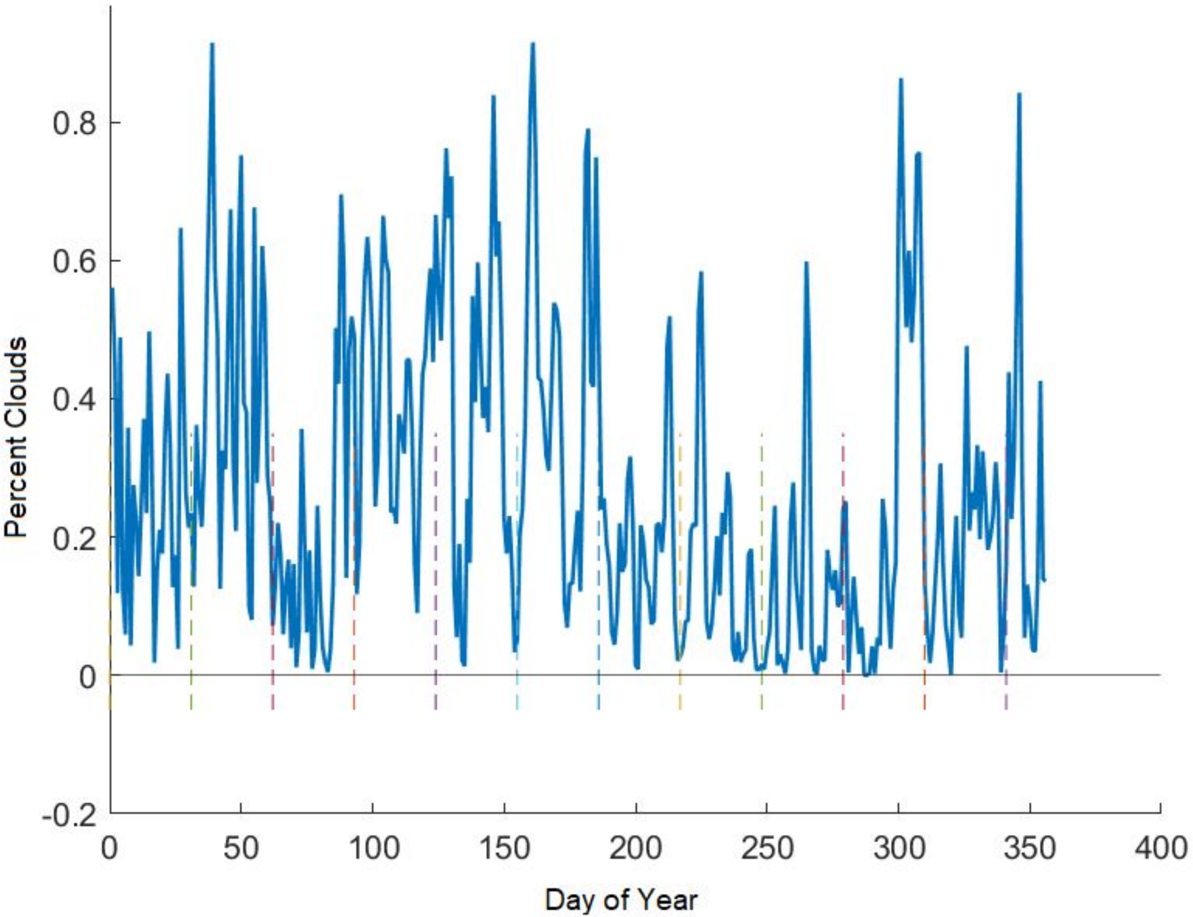
Short step length



Large step length

Model Adaptations: Temporal

- We augmented the temporally coarse data with available temporally finer resolutions.
- Randomly replaced 30% in the coarse data by fine data: given temporally coarse data, we backed it up, or combined it with data of equivalent spatial, but finer temporal resolution
- For example, we might modify the 6 km 8 day coarse data by replacing 30% of it with the finer 6 km 1 day data



Percent Clouds
0.8
0.6
0.4
0.2
0
-0.2
0
50
100
150
200
250
300
350
400
Day of Year

Results

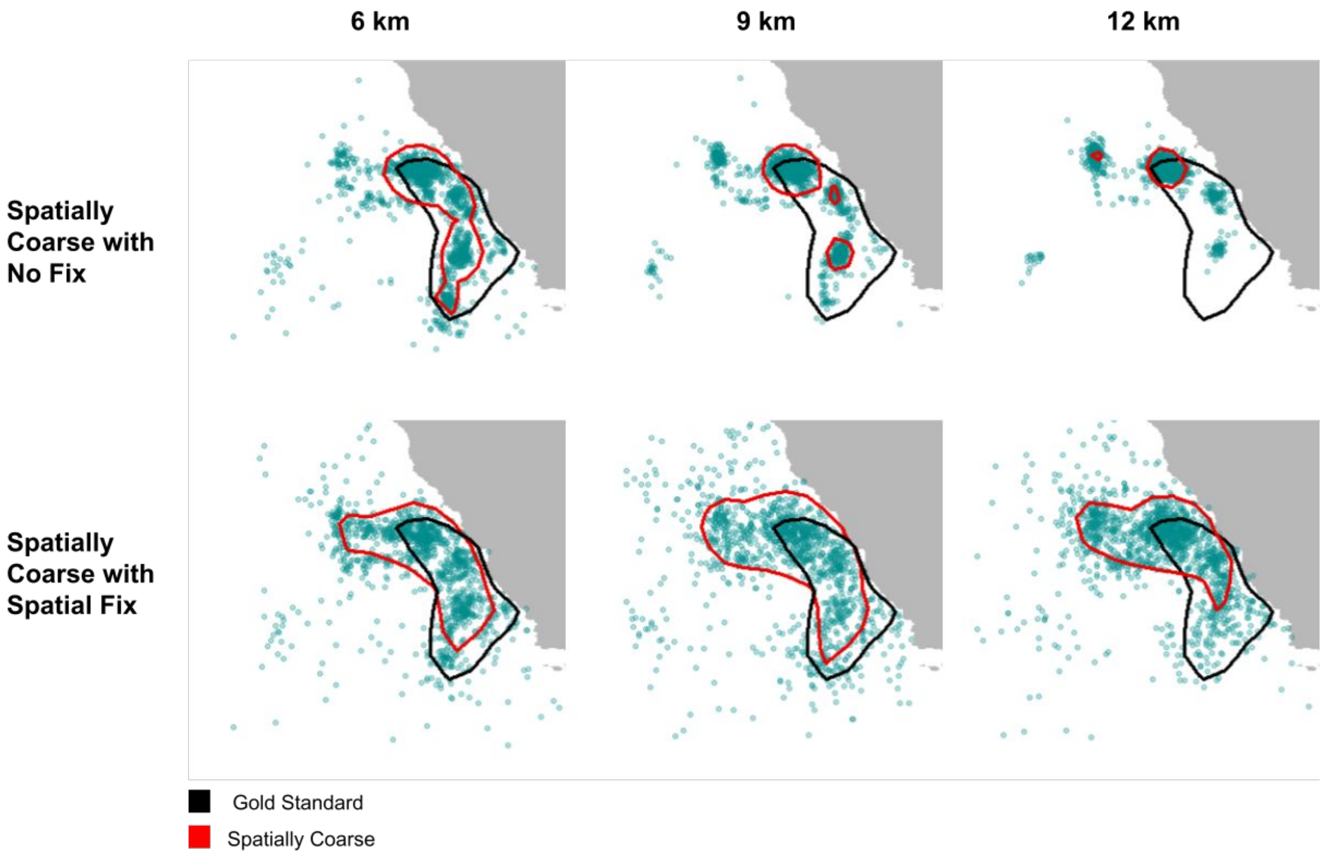
Metrics

- The L2 norm, or the Mean Square Error - minimize
- Create utilization distributions of whale positions from September 1 for output of each model - maximize
  - Used the volume intersection between the utilization distribution of the final positions of the whales from the given model versus the gold standard model
- The Δ metric is the combination of the L2 and VI norms - minimize
  - Our goal was to minimize the Δ of the model; i.e. minimize the L2 and maximize the VI values
  - Measured the effectiveness of a model adjustment with the fold decrease in Δ between the models before and after the adjustment was applied

Changing Rate Counteracts Spatially Coarsened Data

- With new rates in IBM, whales foraged in generally the same areas with coarse data as they did with the gold standard data
- The utilization distributions of the whales' positions on the date September 1, became more similar to that of the gold standard both visibly and numerically

	6 km	9 km	12 km
VI of Original	0.705	0.480	0.256
VI of Corrected	0.710	0.645	0.549
Fold Decrease Δ	1.617	1.920	1.800

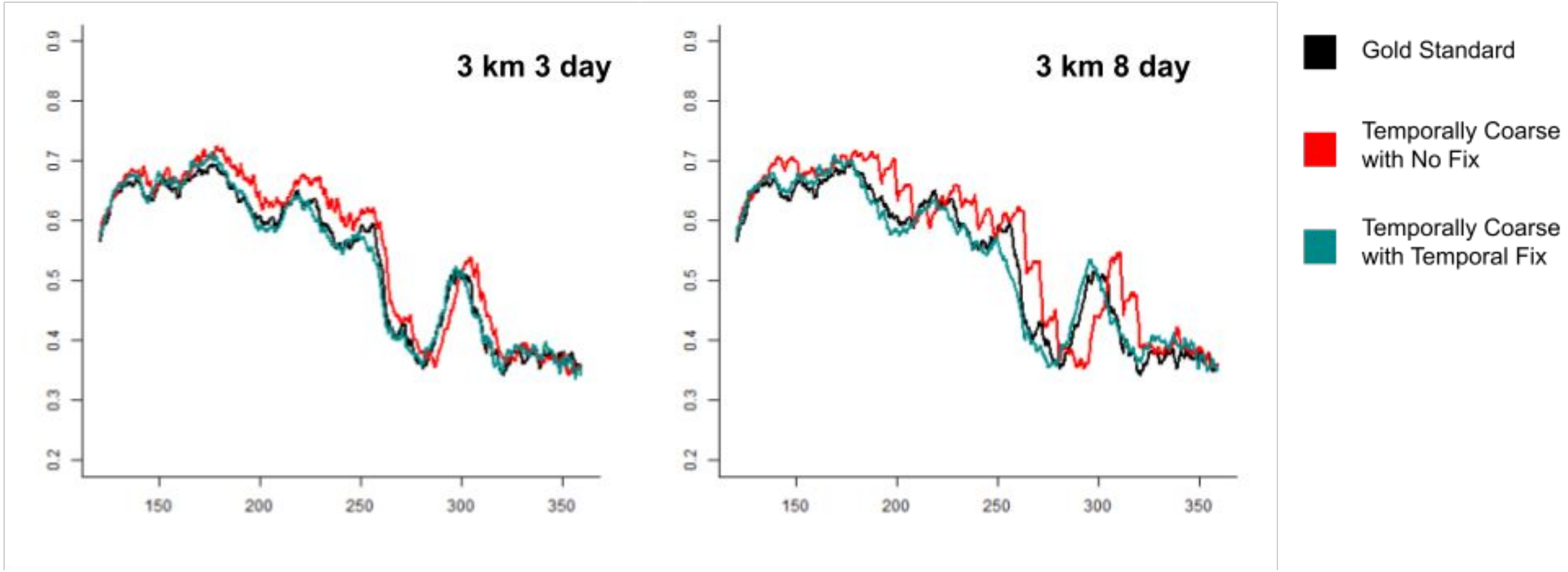


6 km
9 km
12 km
Spatially Coarse with No Fix
Spatially Coarse with Spatial Fix
Gold Standard
Spatially Coarse

Added Fine Data Counteracts Temporally Coarsened Data

- To mitigate abrupt changes in population level foraging behaviour caused by temporally coarse data, we replaced 30% of the coarse data with fine
- The foraging percentages were visually less step-like and numerically closer to the gold standard results

	3 day	8 day
L2 or Original	1.211	1.787
L2 of Corrected	0.772	0.935
Fold Decrease Δ	1.623	2.004


3 km 3 day
3 km 8 day
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
150
200
250
300
350
Gold Standard
Temporally Coarse with No Fix
Temporally Coarse with Temporal Fix

Combined Spatial and Temporal Fixes

- Negative influences on whale foraging behaviours were amplified when the environmental data had both poor spatial and temporal resolutions
  - Spatiotemporally coarsened data had a minimum Δ of 3 across all resolutions
- Combining the fixes by substituting available fine data, then changing model rate was successful in improving the behaviours of doubly coarse data

Δ Fold Decrease	3 km	6 km	9 km	12 km
1 day	-	1.617	1.920	1.800
3 day	1.623	2.963	2.700	3.220
8 day	2.004	3.305	3.257	3.770

Conclusion

- Solution for spatial coarseness is logical and easily applied
- Solution for temporal coarseness only required 30% of the coarse data to be randomly replaced with the fine to see great improvement
    - Analysis of GOES satellite data showed that the average missing data during summer and fall is usually less than 40%
- The methods are not unique to blue whales
    - The solutions are useful for models have inputs that are dynamic in space and time

Conclusion

- Sameerah Helal (me): shelal@ucdavis.edu
- Dr. Stephanie Dodson (my advisor): sadodson@ucdavis.edu

- Access the code: https://github.com/s-helal/Blue-Whale-Data-Research