

Unsupervised clustering of telemetry data with deep neural network

Natalya Rapstine and Jeff Tracey
U.S. Geological Survey

RMACC May 22, 2019

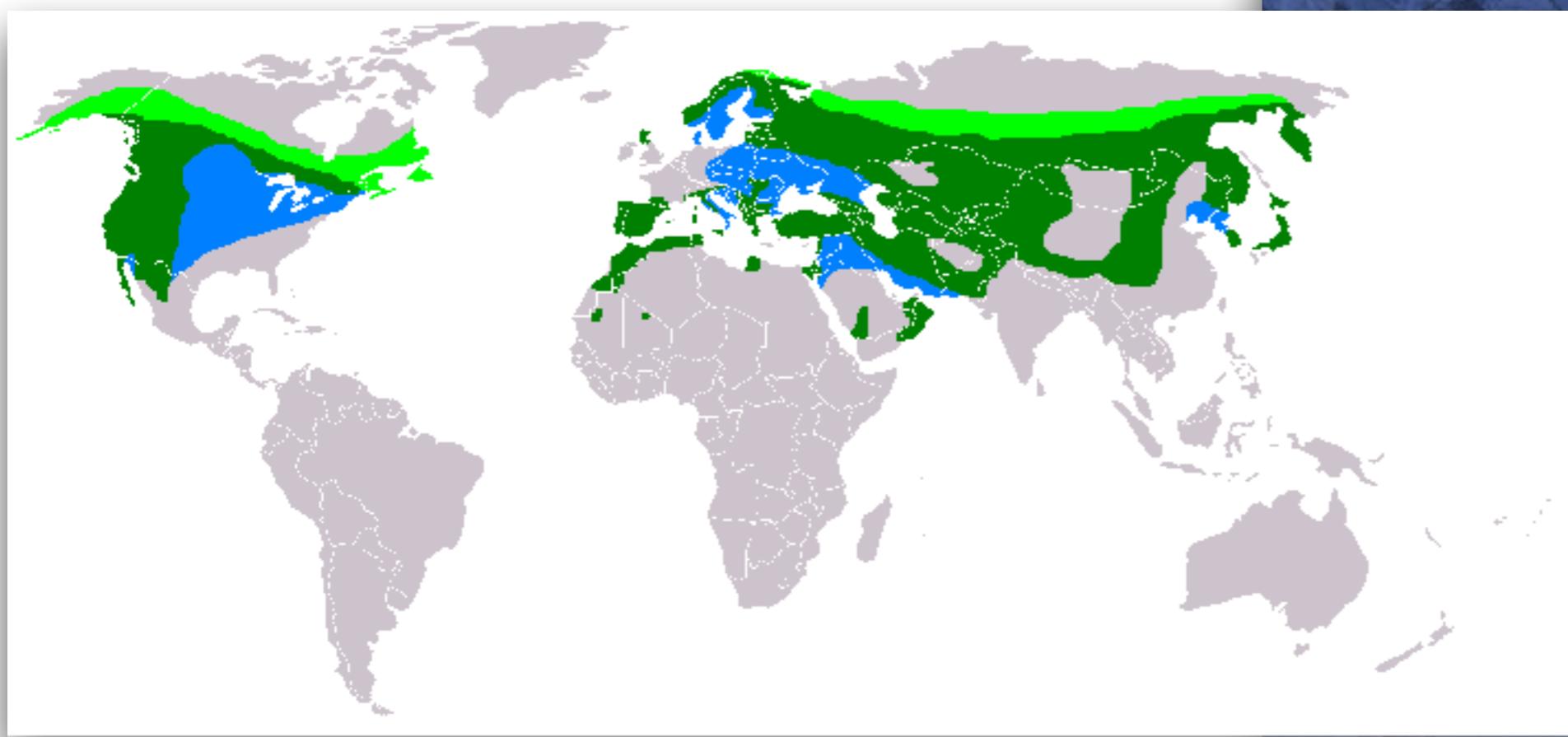


Golden Eagle

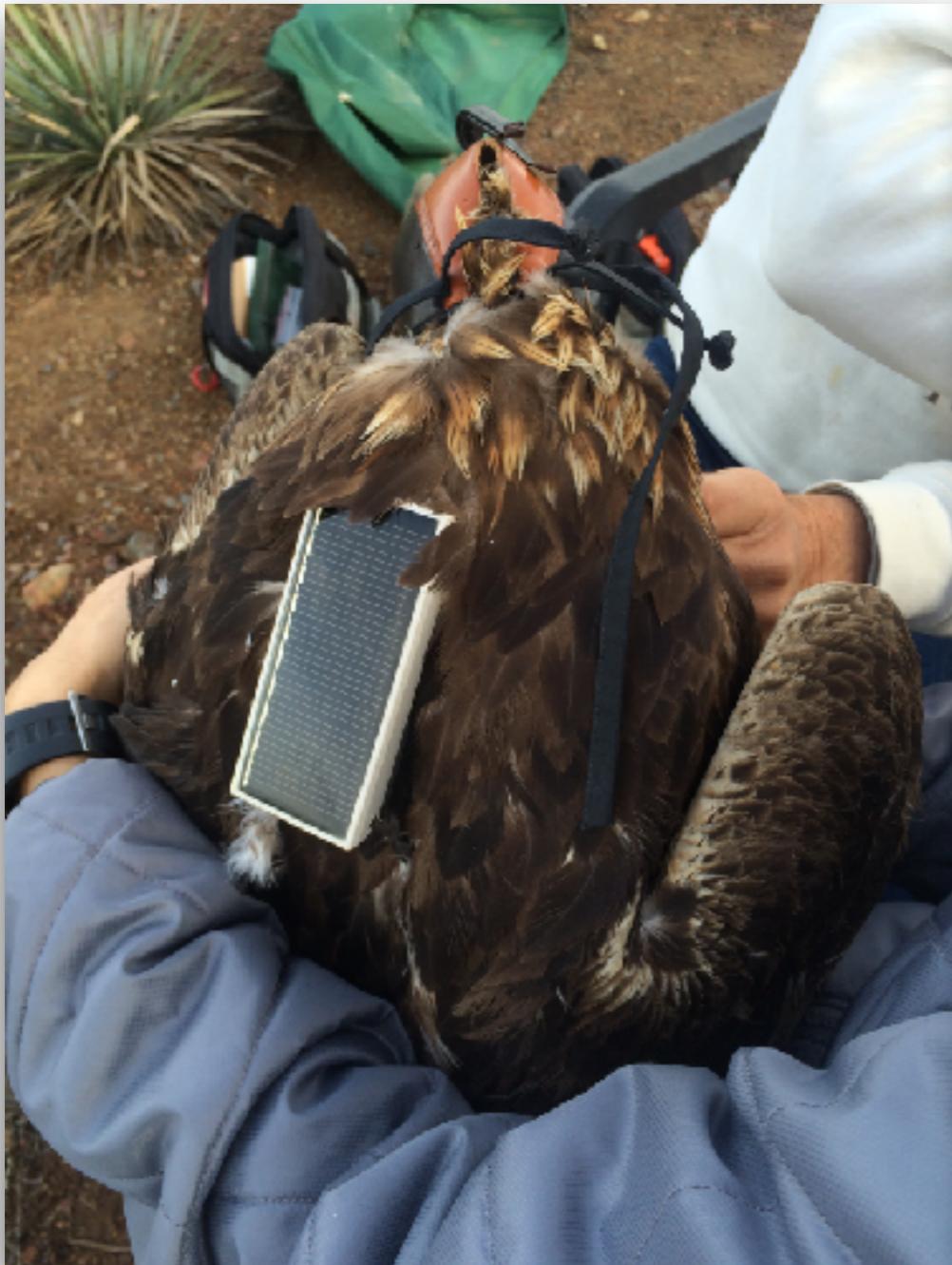
Aquila chrysaetos



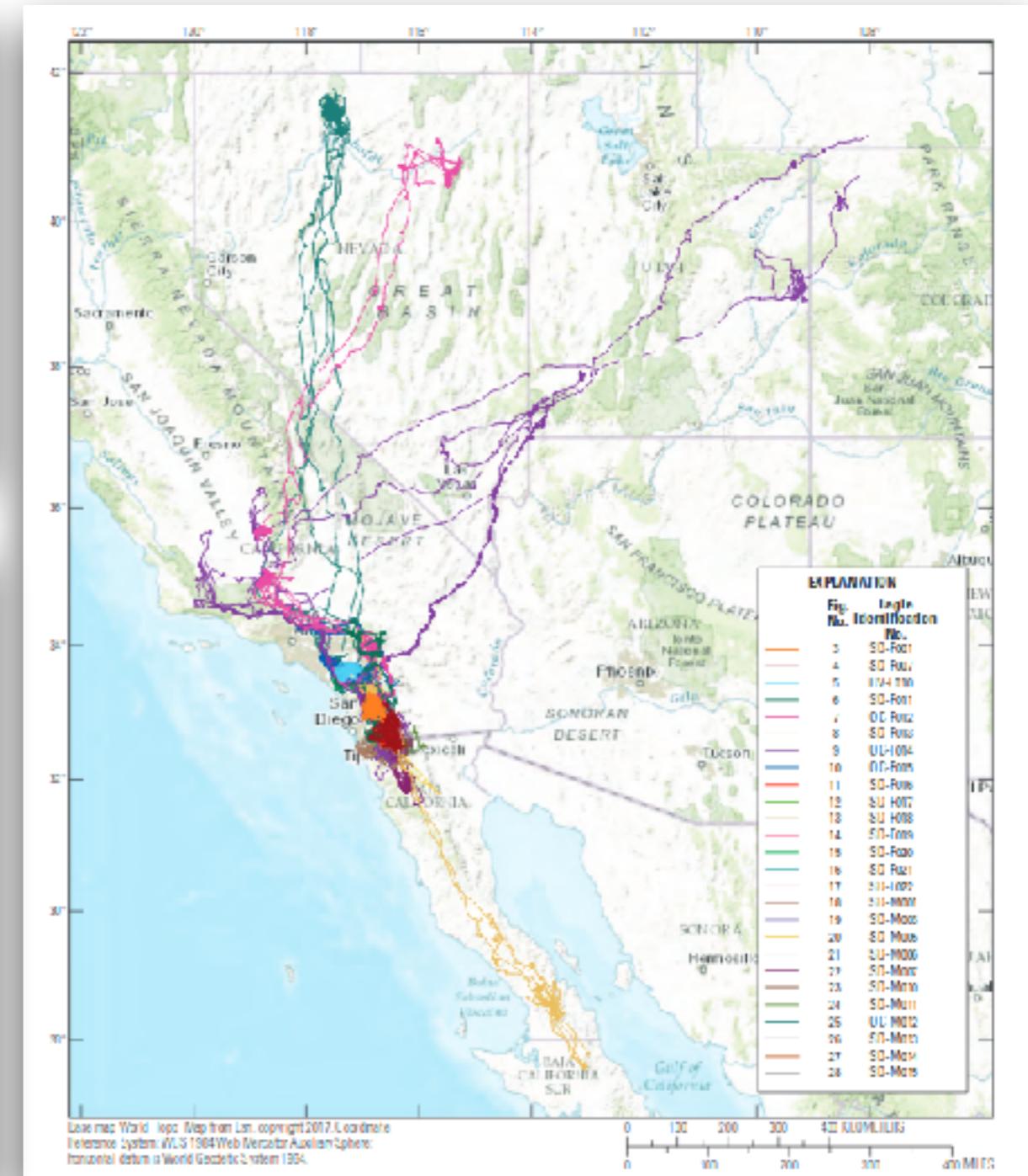
Distribution



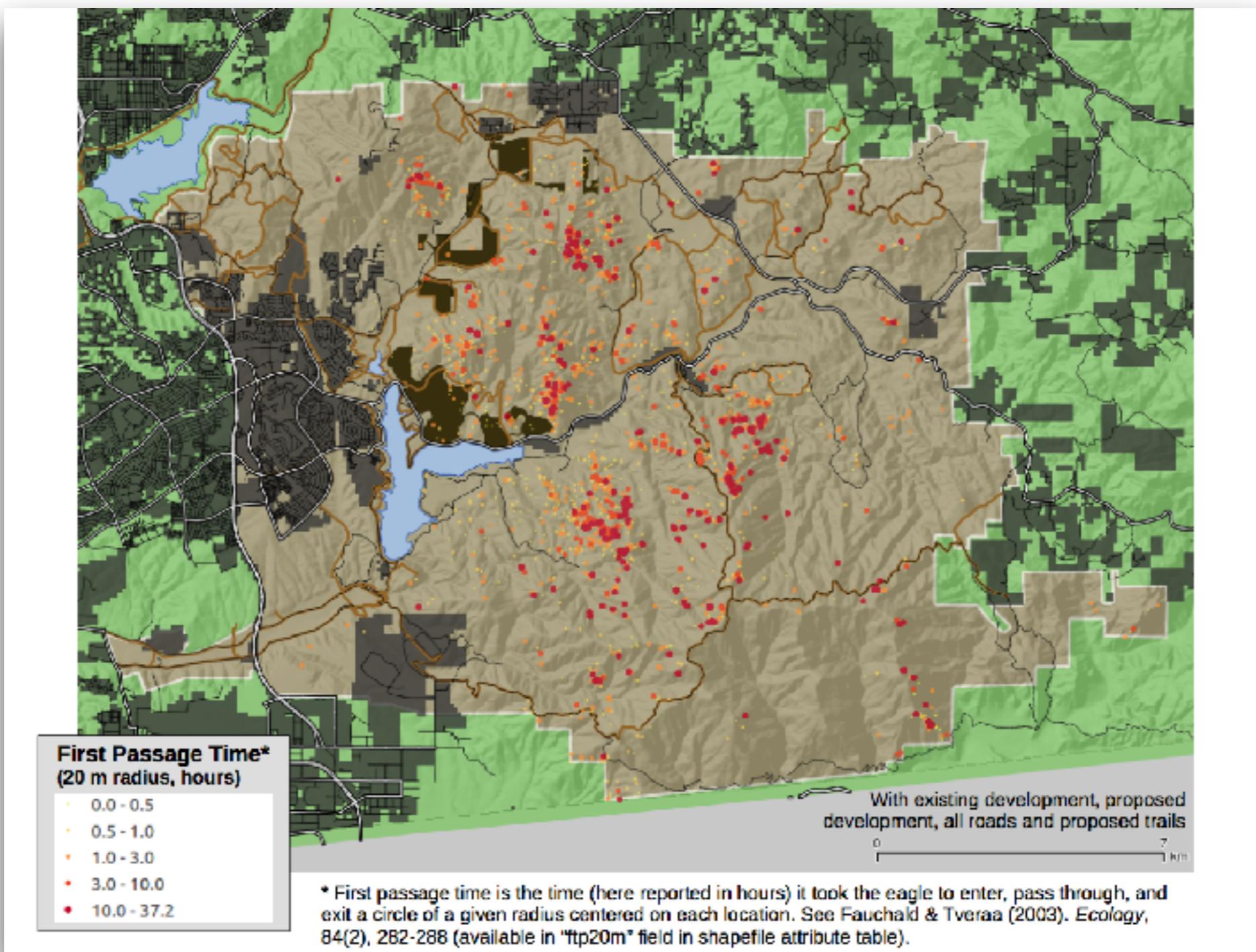
Telemetry



Telemetry data

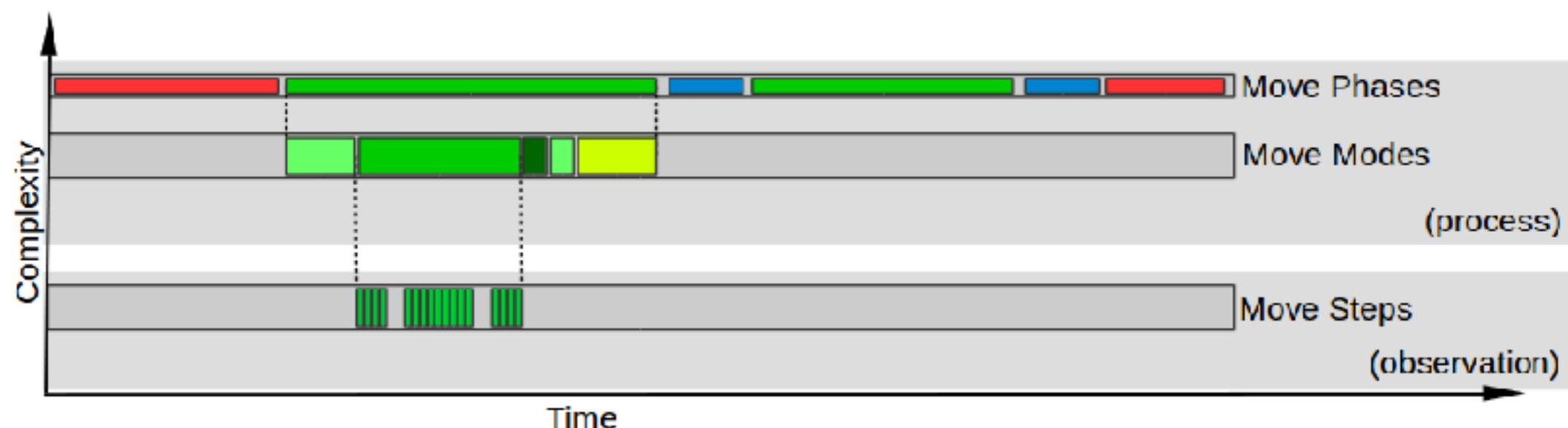


Habitat Loss

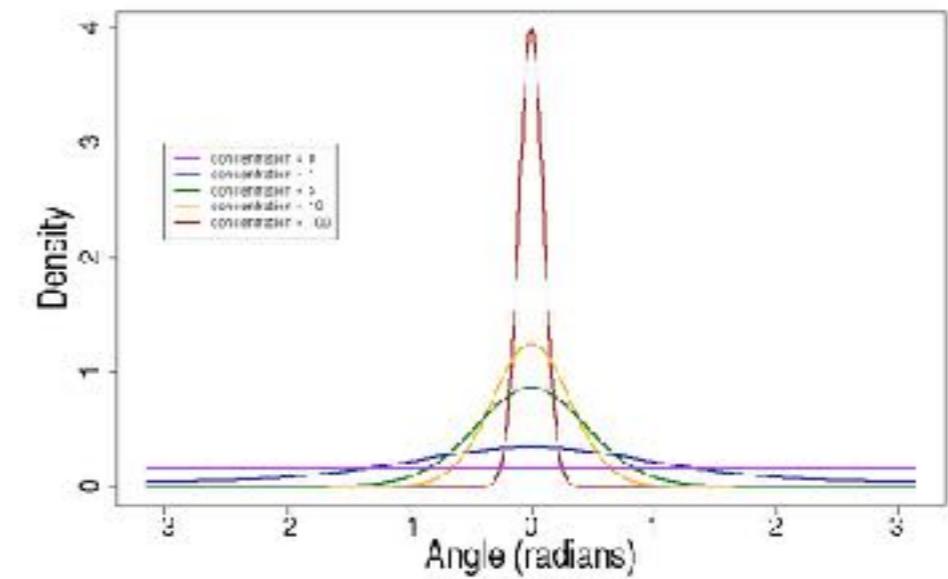
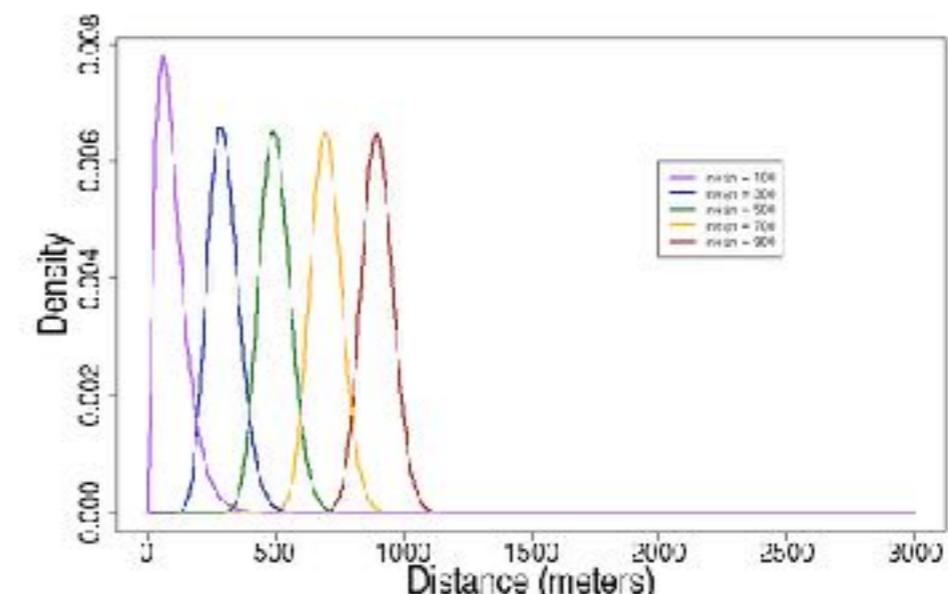
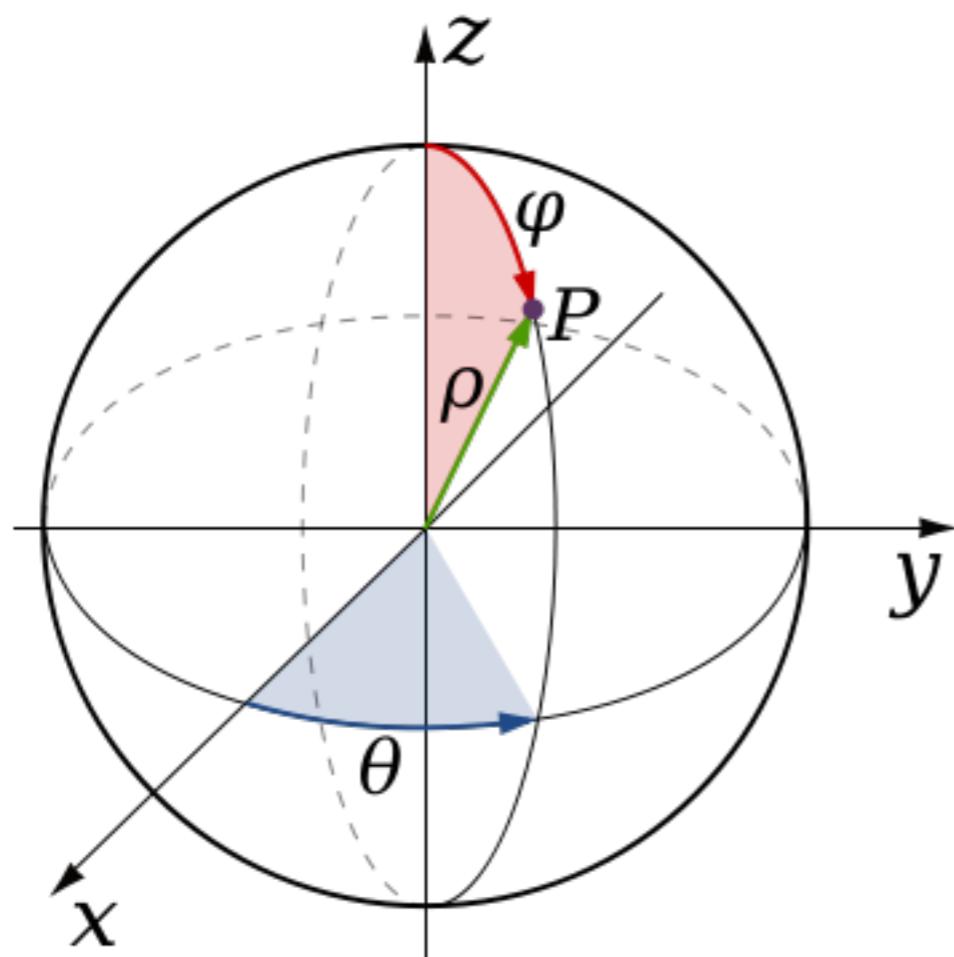


Research Challenge

Use deep learning neural nets + a classification algorithm to segment move data by eagle behavior



Simulated Data

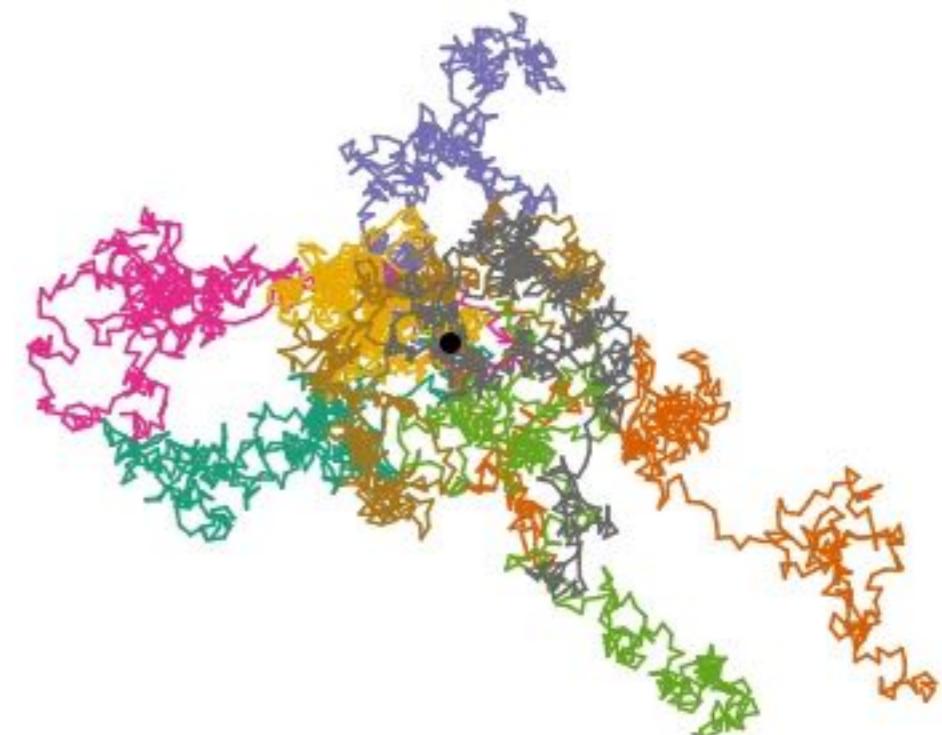
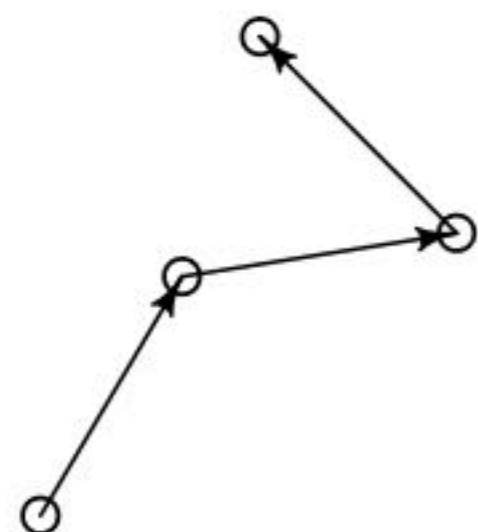


Sim Move Modes

	Move Distance	Mean Angle*	Stop
No Movement (NM)	0	NA	~ Bernoulli
Simple Random Walk (SRW)	~ gamma	~uniform	exp time
Correlated Random Walk (CRW)	~ gamma	previous angle	exp time
Point Bias (PB)	~ gamma	angle to point	within distance
Directional Bias (DB)	~gamma	constant angle	exp time

After exiting a mode, Markov matrix
for transitions between modes

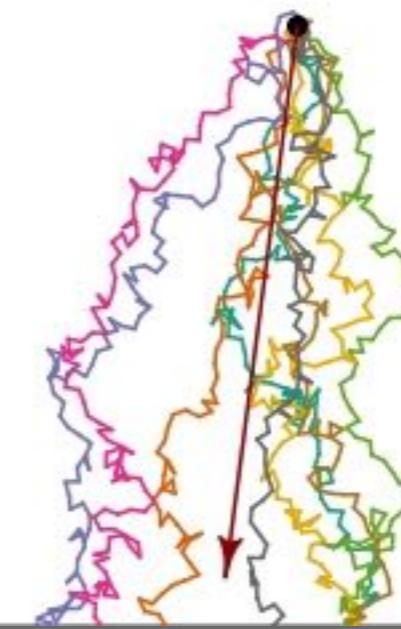
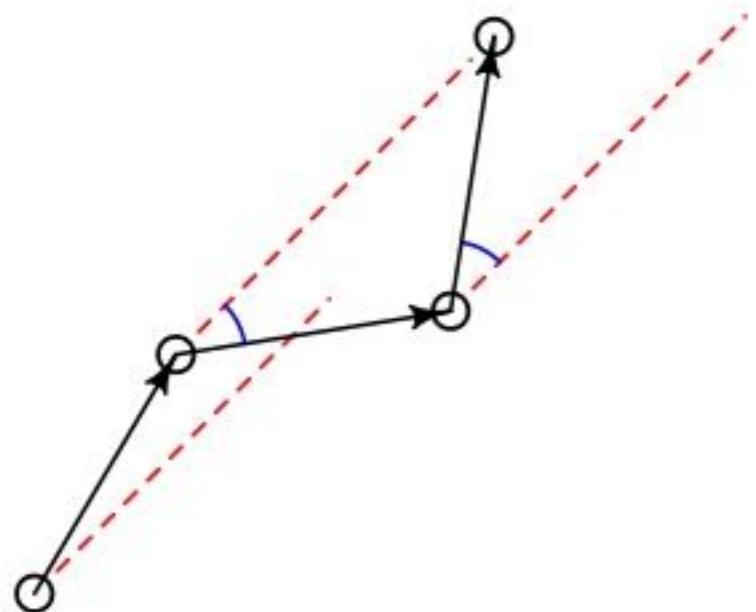
Simple Random Walk (SRW)



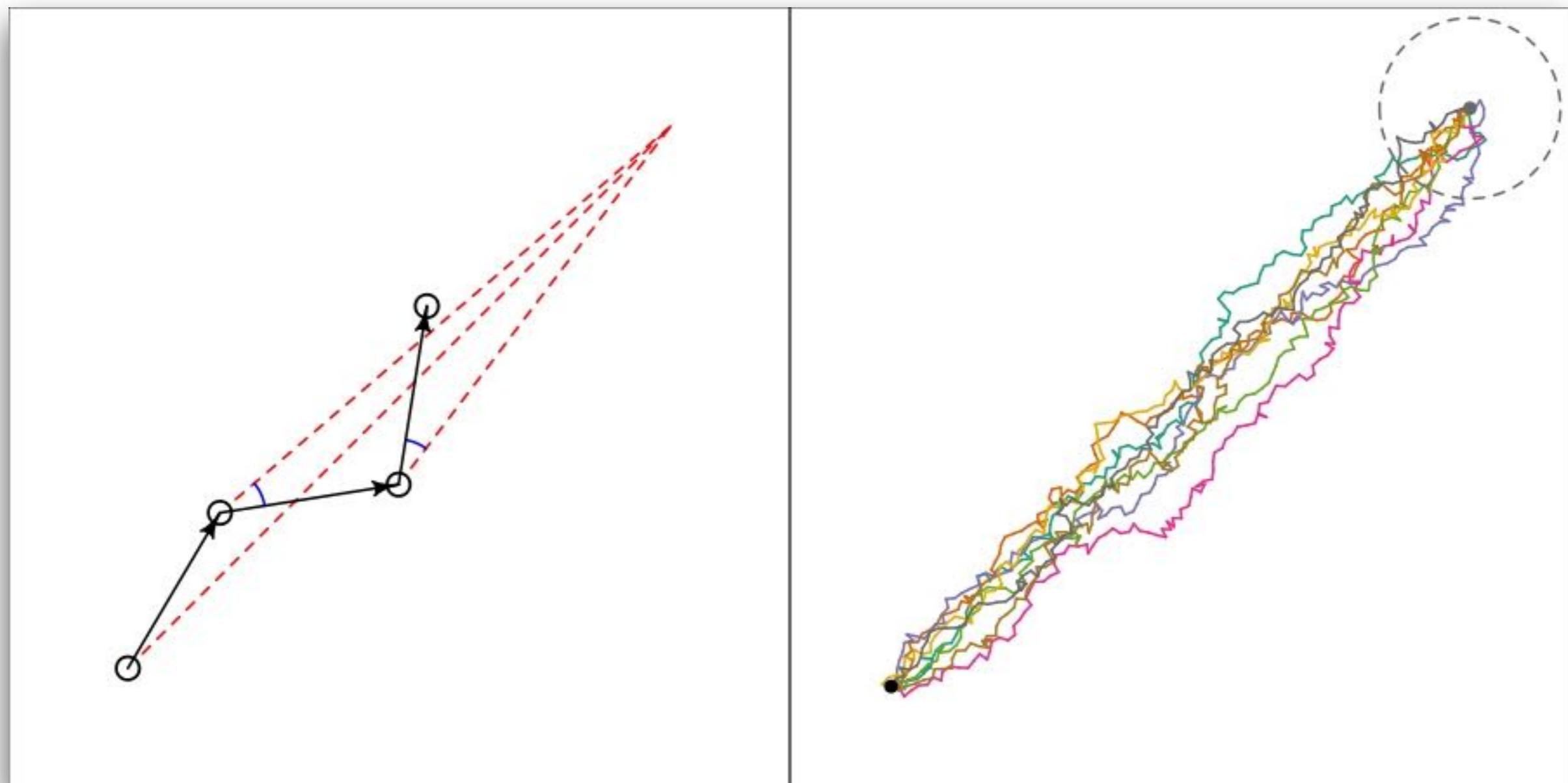
Correlated Random Walk (CRW)



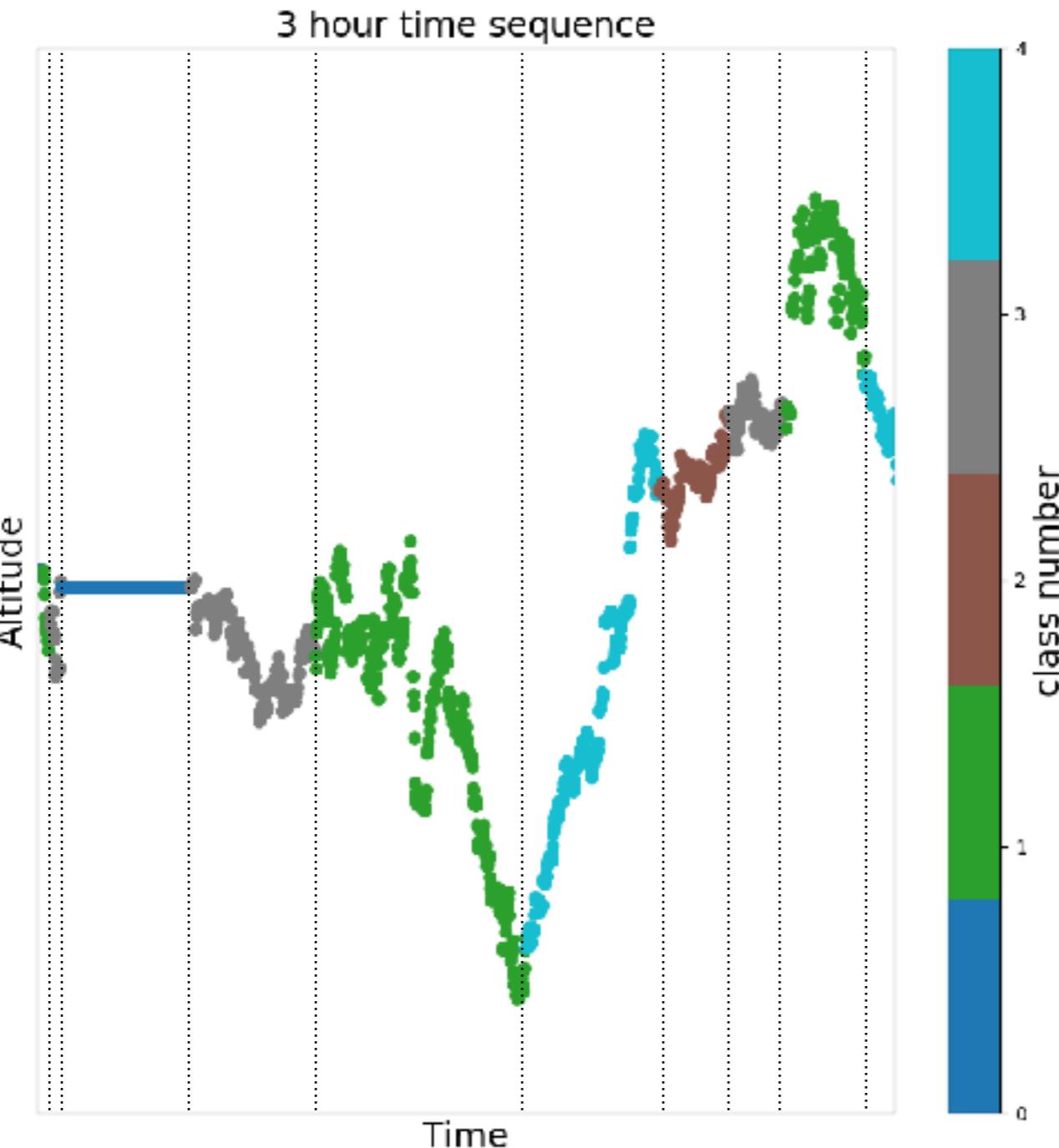
Directional Bias (DB)



Point Bias (PB)



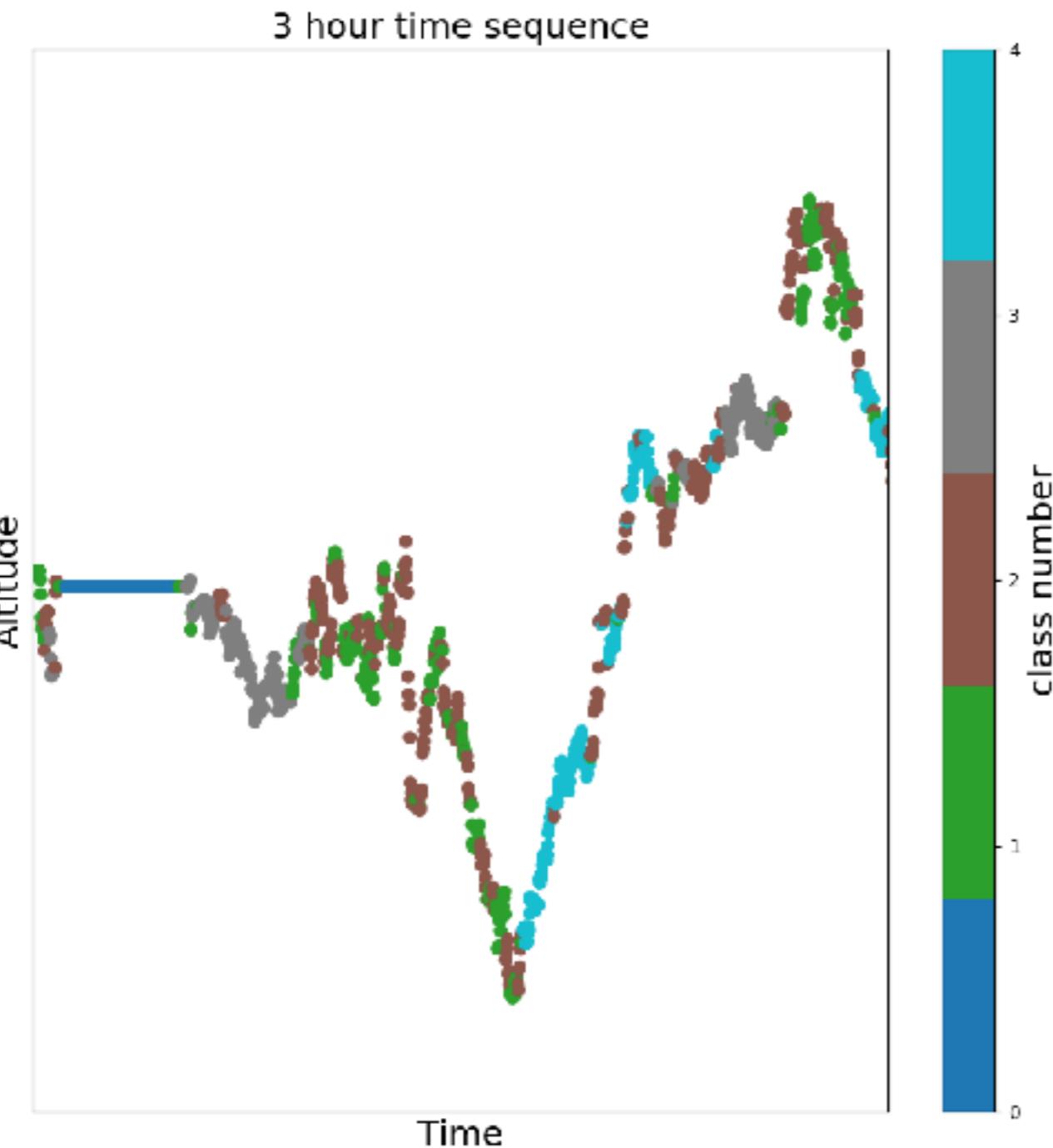
Research Challenge



Extract movement patterns
from sequences of GPS
data

Recognize behaviors

Our Approach



Autoencoder with LSTM layers

LSTM takes time dependence into account

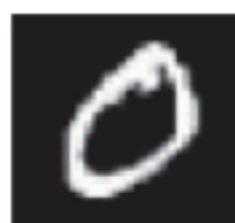
Cluster encoded data representation

Supervised

label = 5



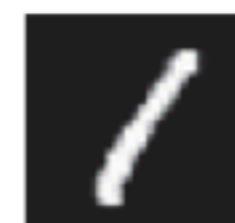
label = 0



label = 4



label = 1



label = 9



label = 2



label = 1



label = 3



label = 1



label = 4



label = 3



label = 5



label = 3



label = 6



label = 1



label = 7



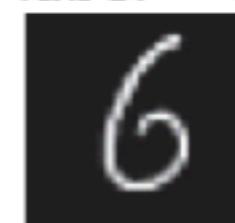
label = 2



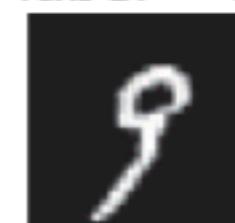
label = 8



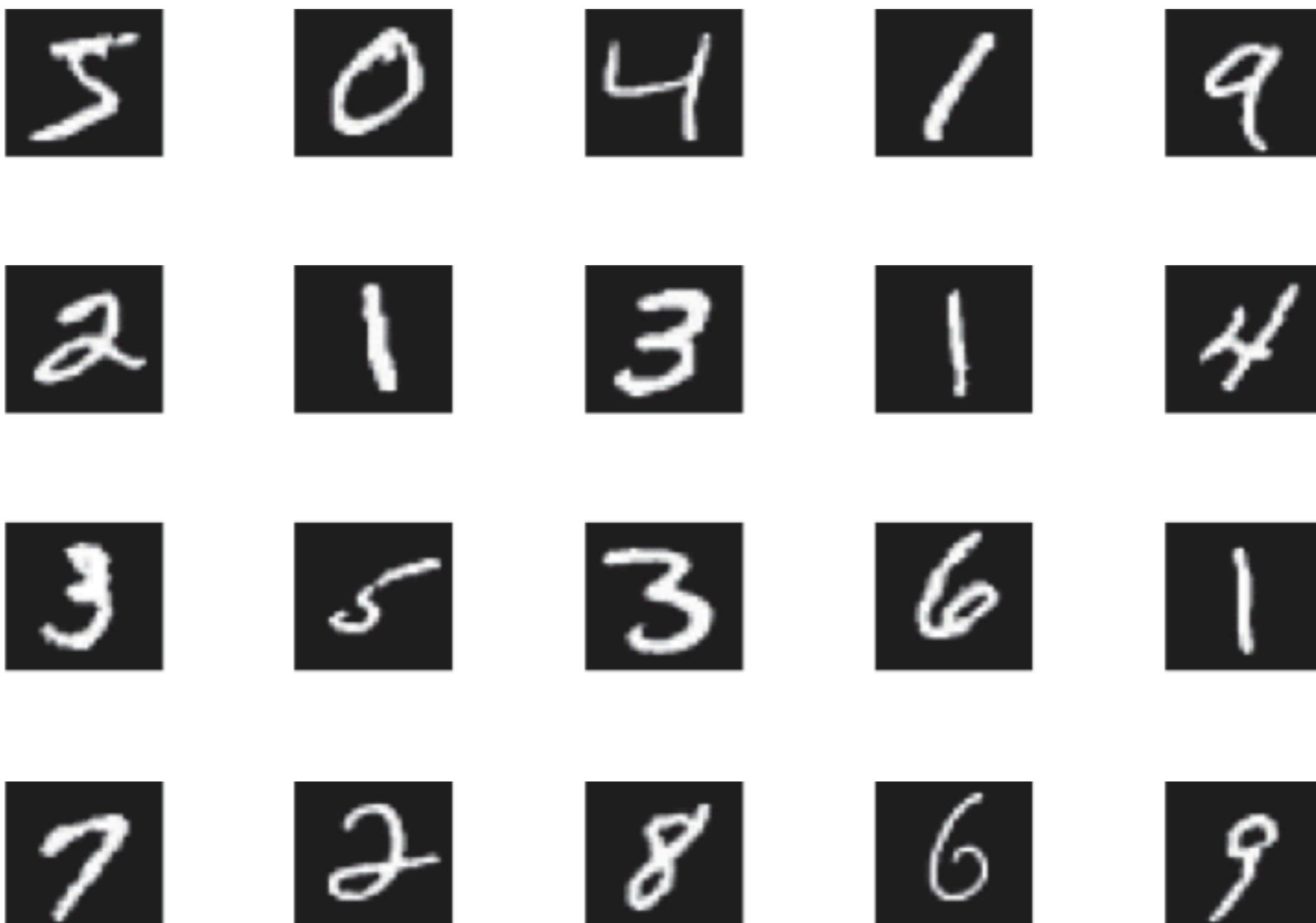
label = 6



label = 9

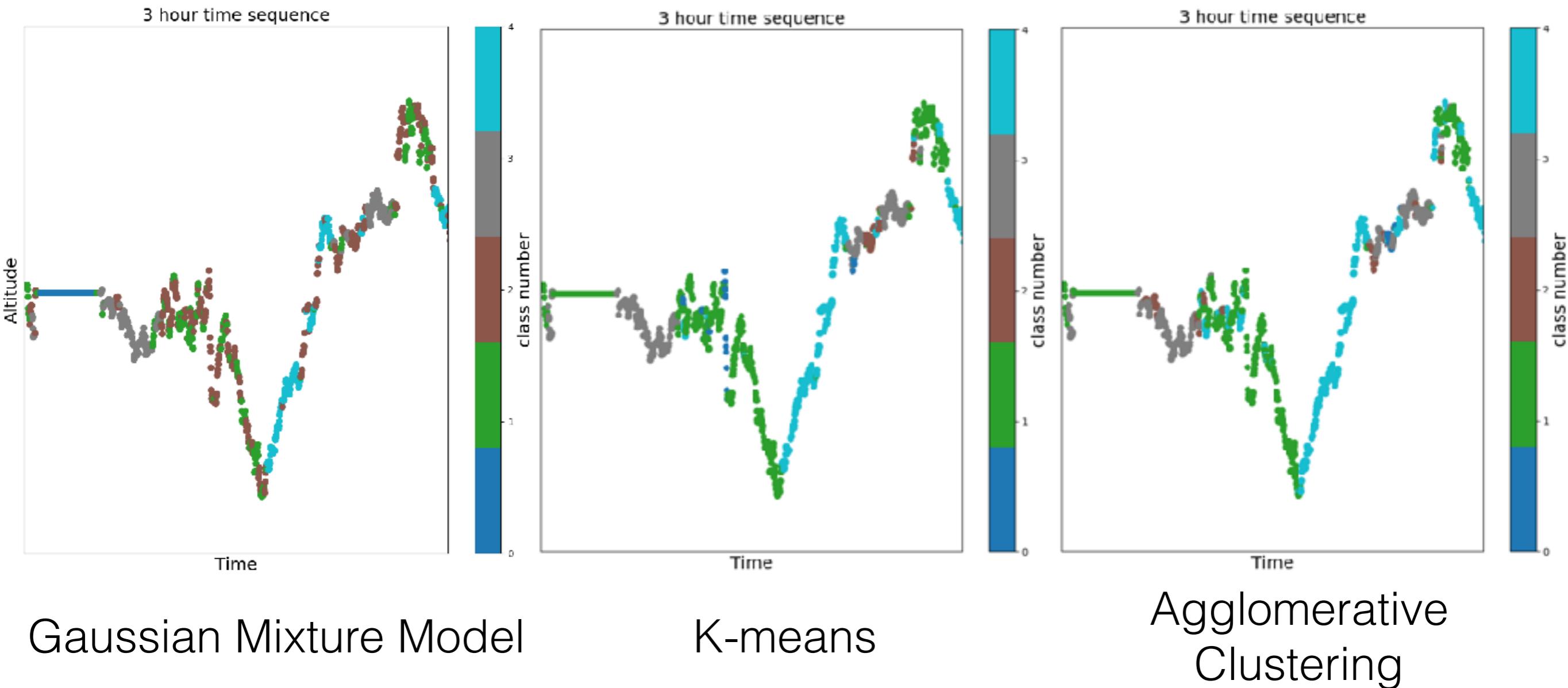


Unsupervised



No labels

Clustering

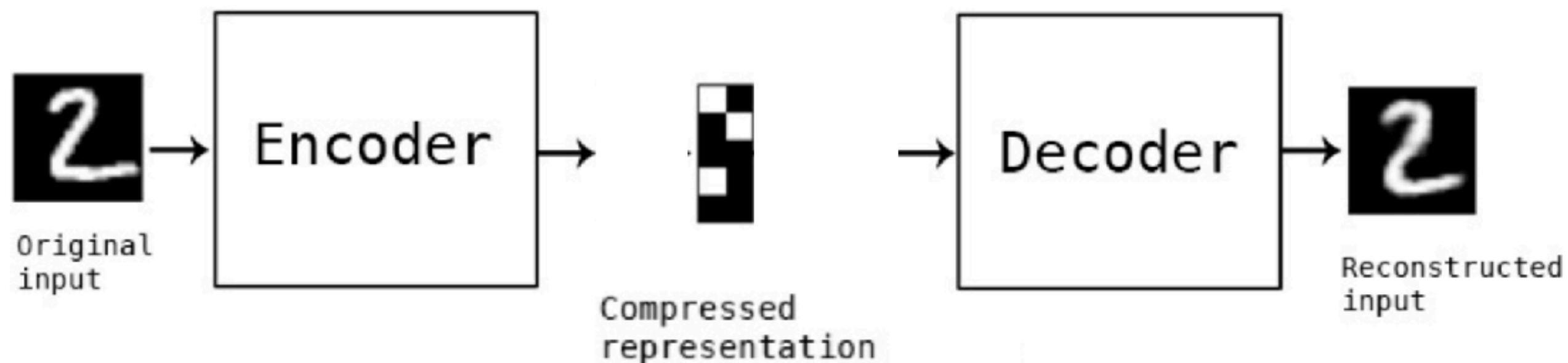


Gaussian Mixture Model

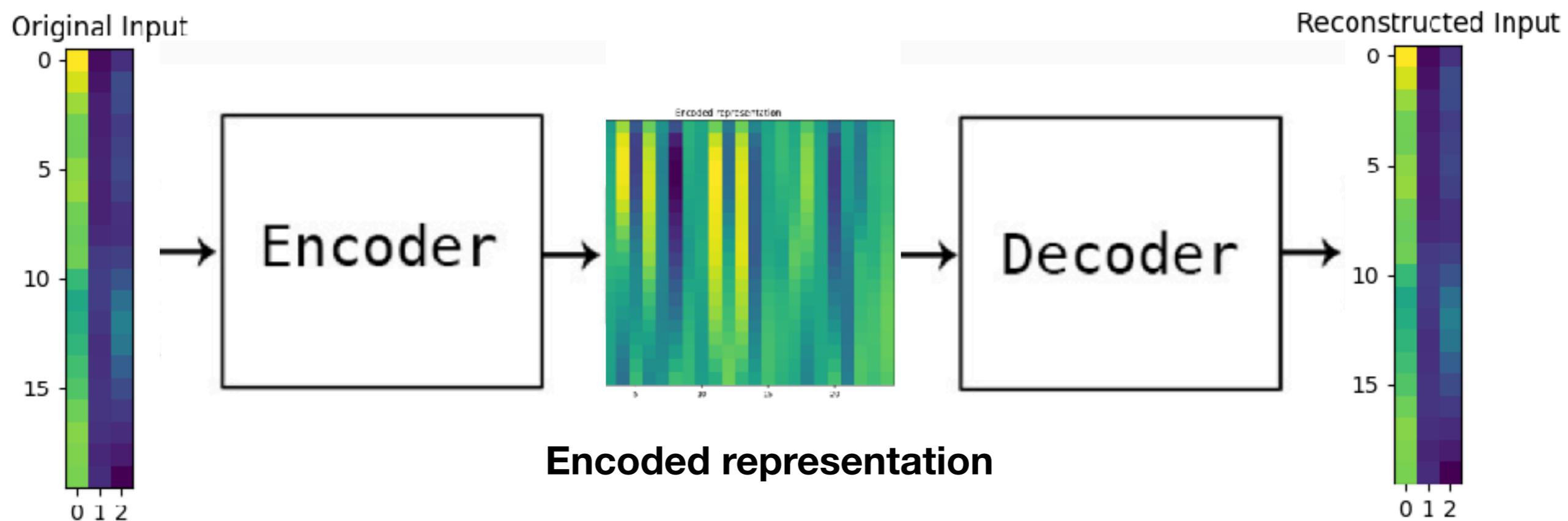
K-means

Agglomerative
Clustering

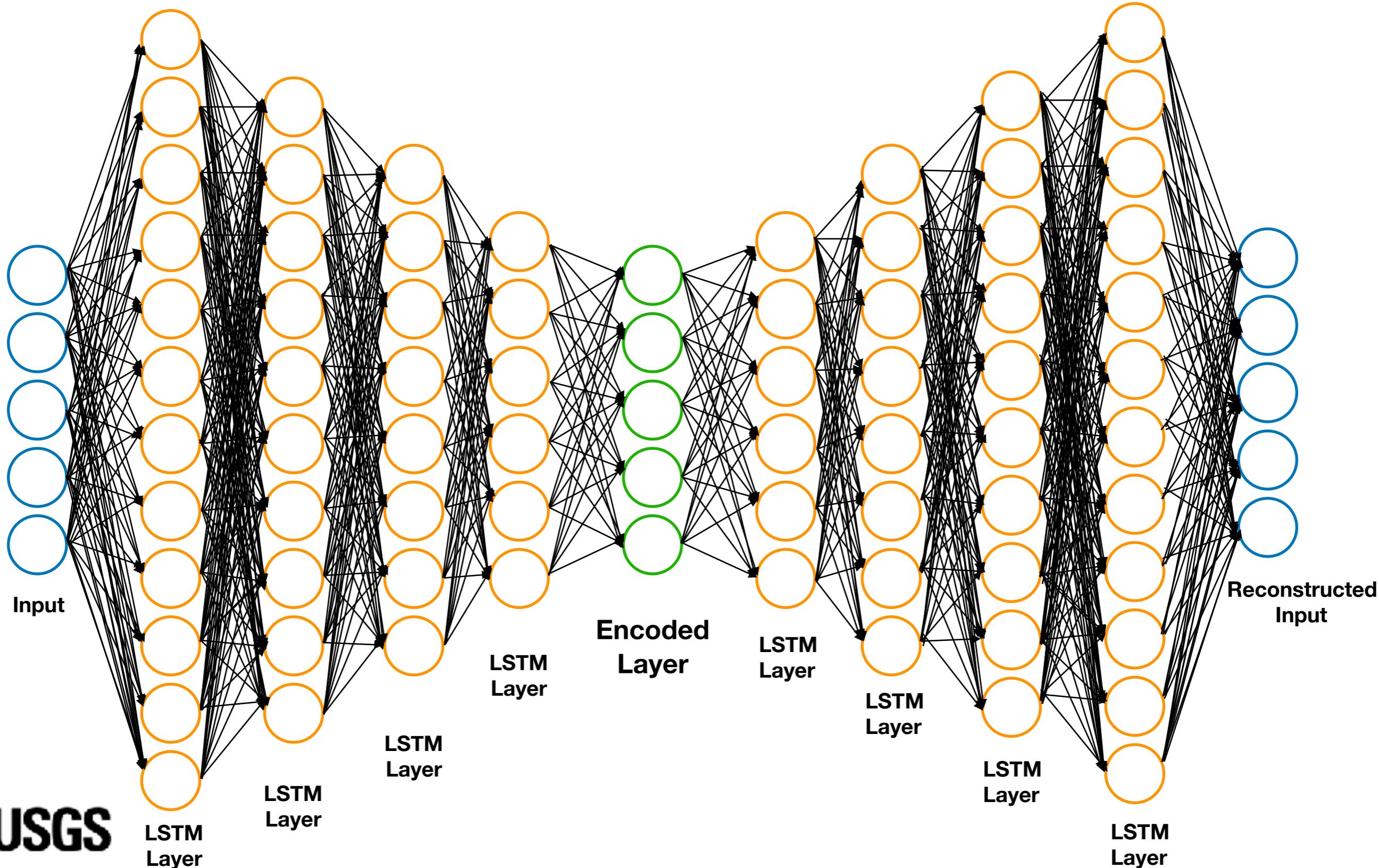
What is an autoencoder?



What is an autoencoder?

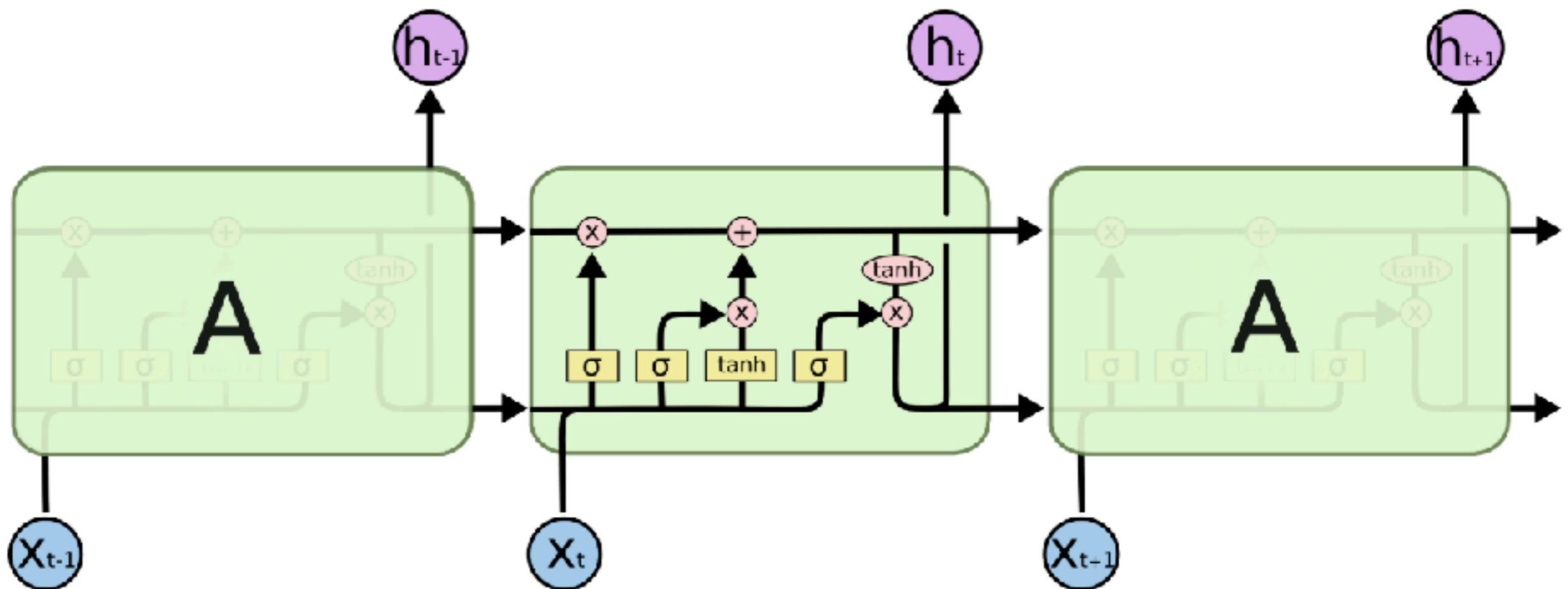


Autoencoder Architecture



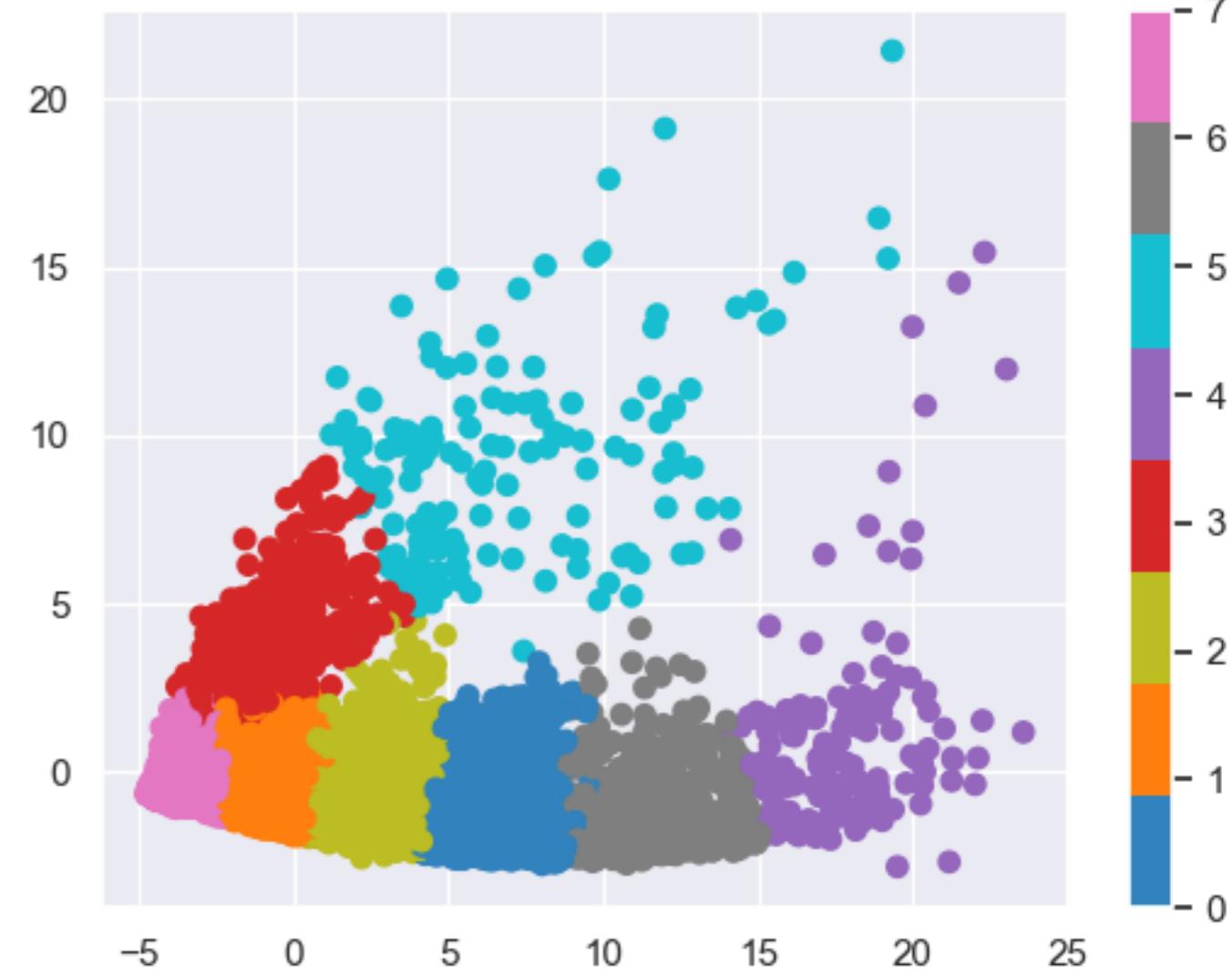
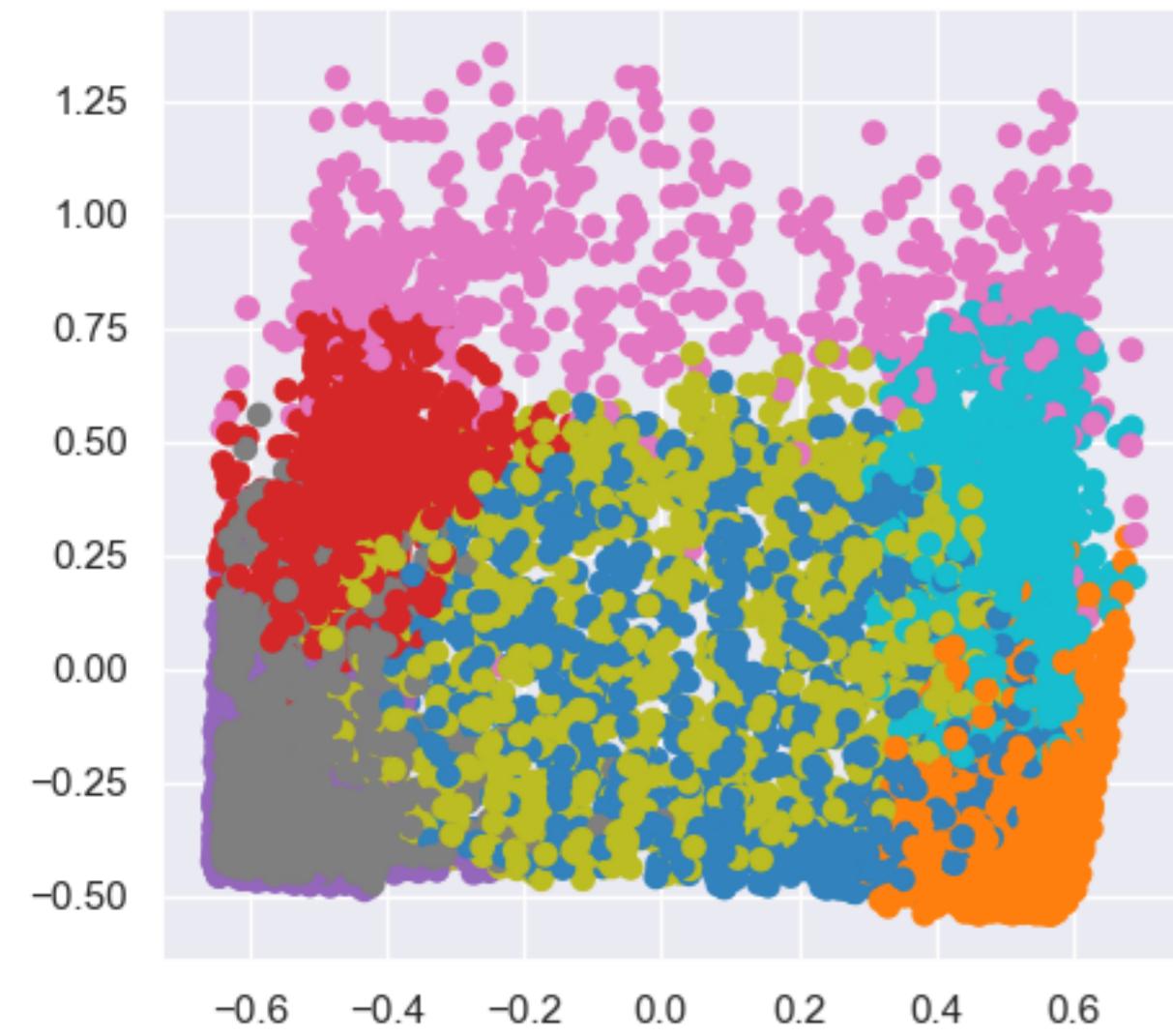
LSTM

Long short term memory cell: 12 parameter



Clustering after Autoencoding

We cluster the raw data and the “encoded” data

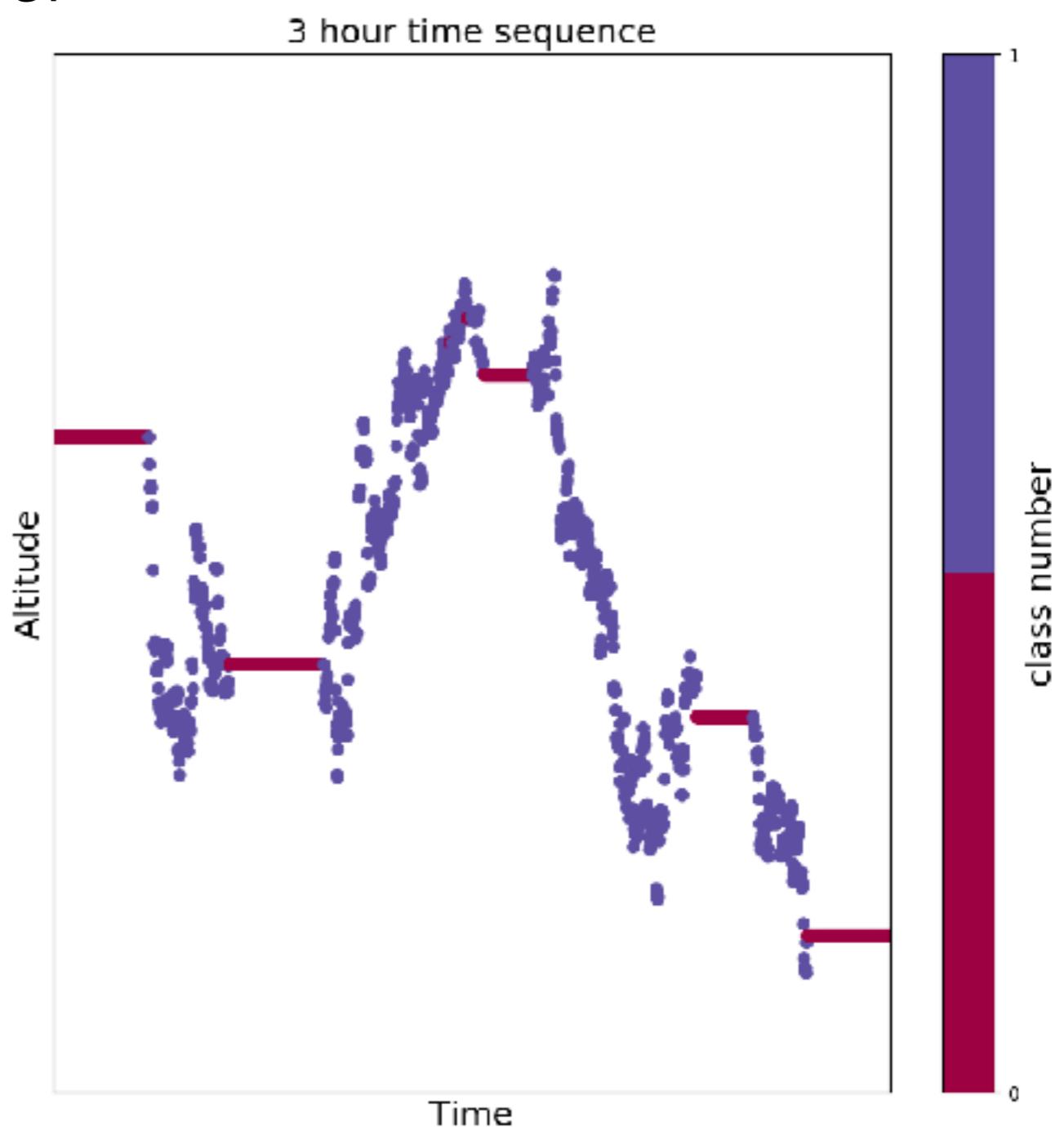


Simulated data experiments

2 Classes

Distinguish between two classes:

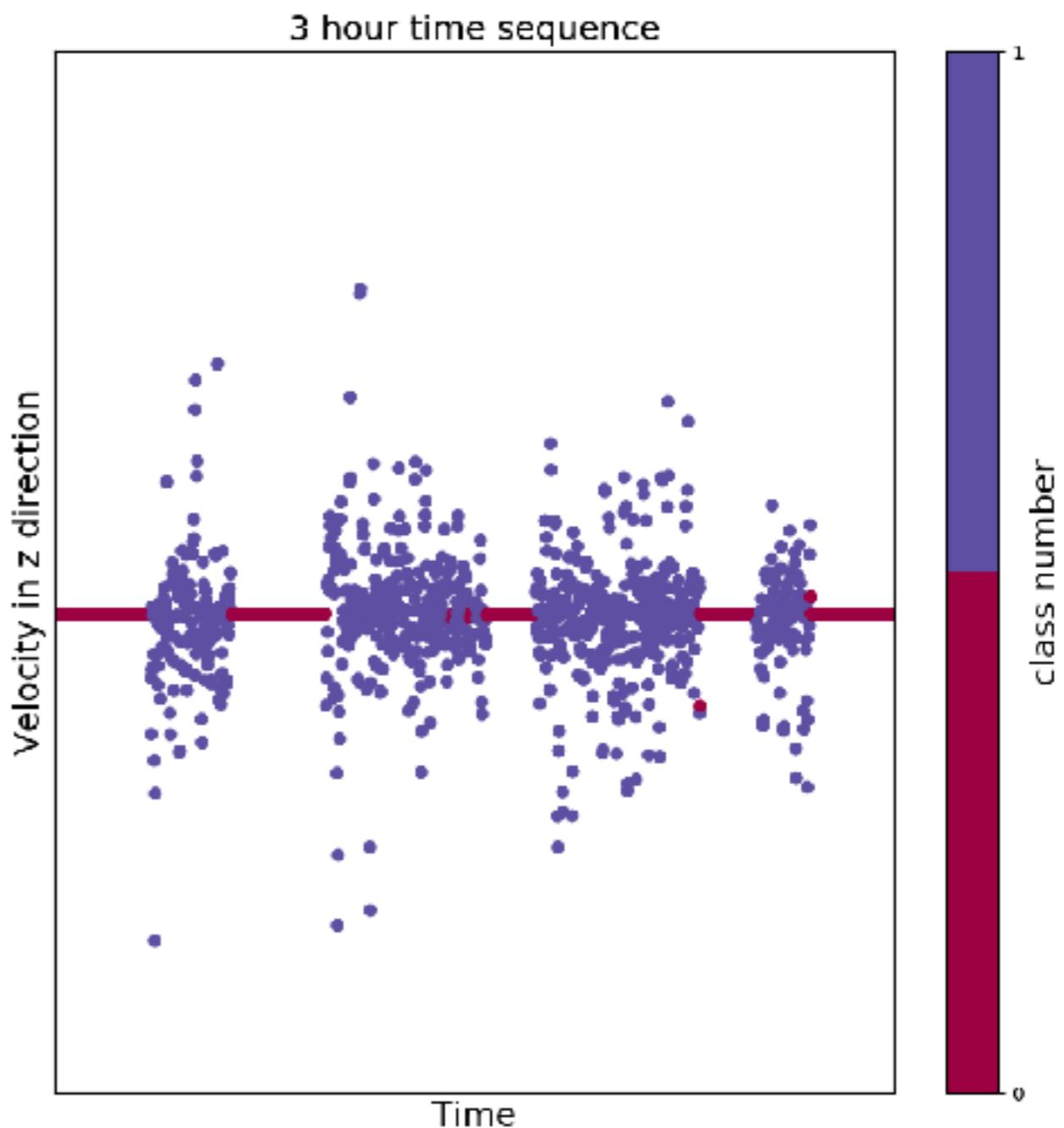
- No movement
- Movement



Clustering Raw Data

K-means:

- 0.64 accuracy
- 0.59 F1 score

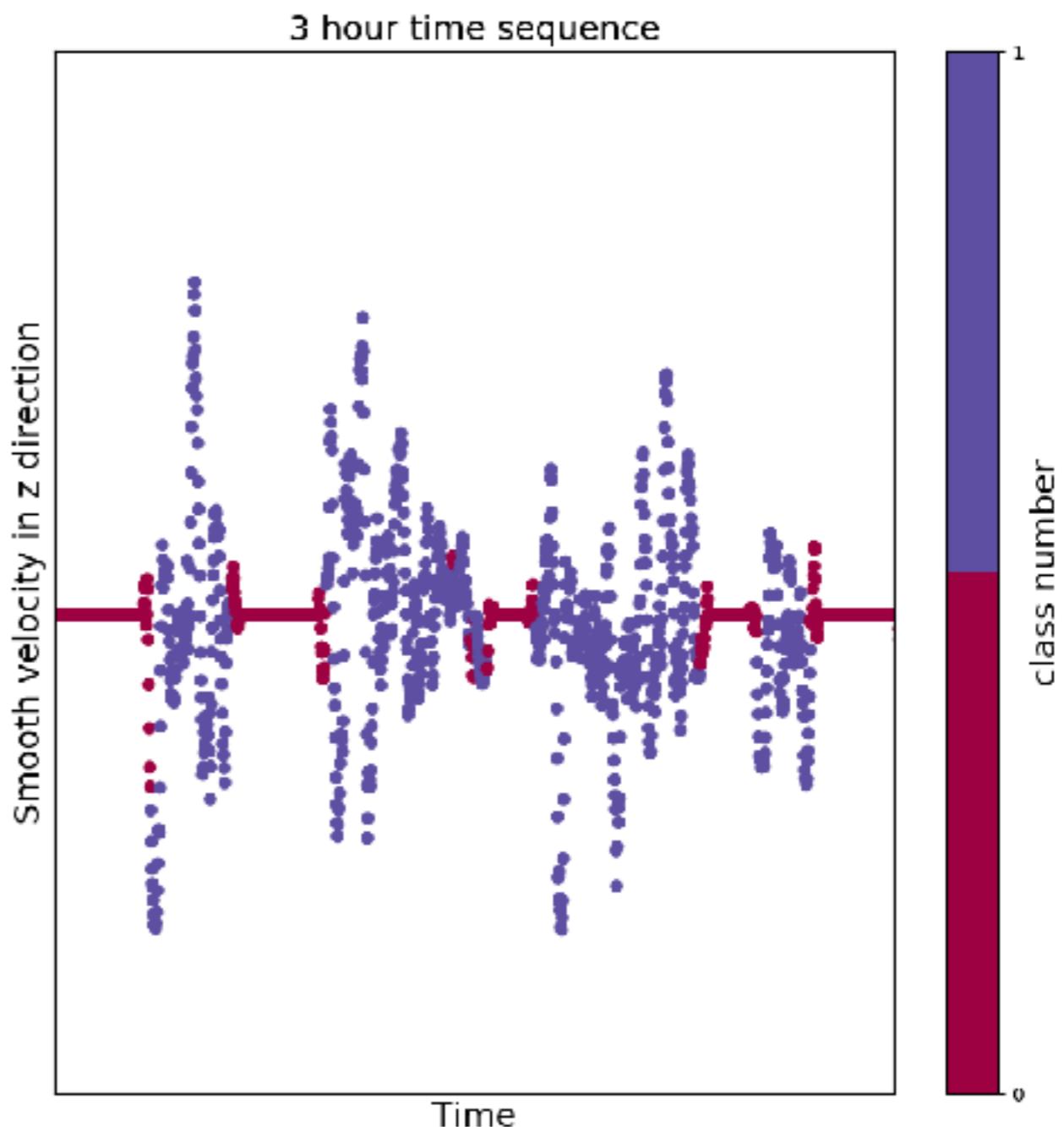


Clustering Raw Data

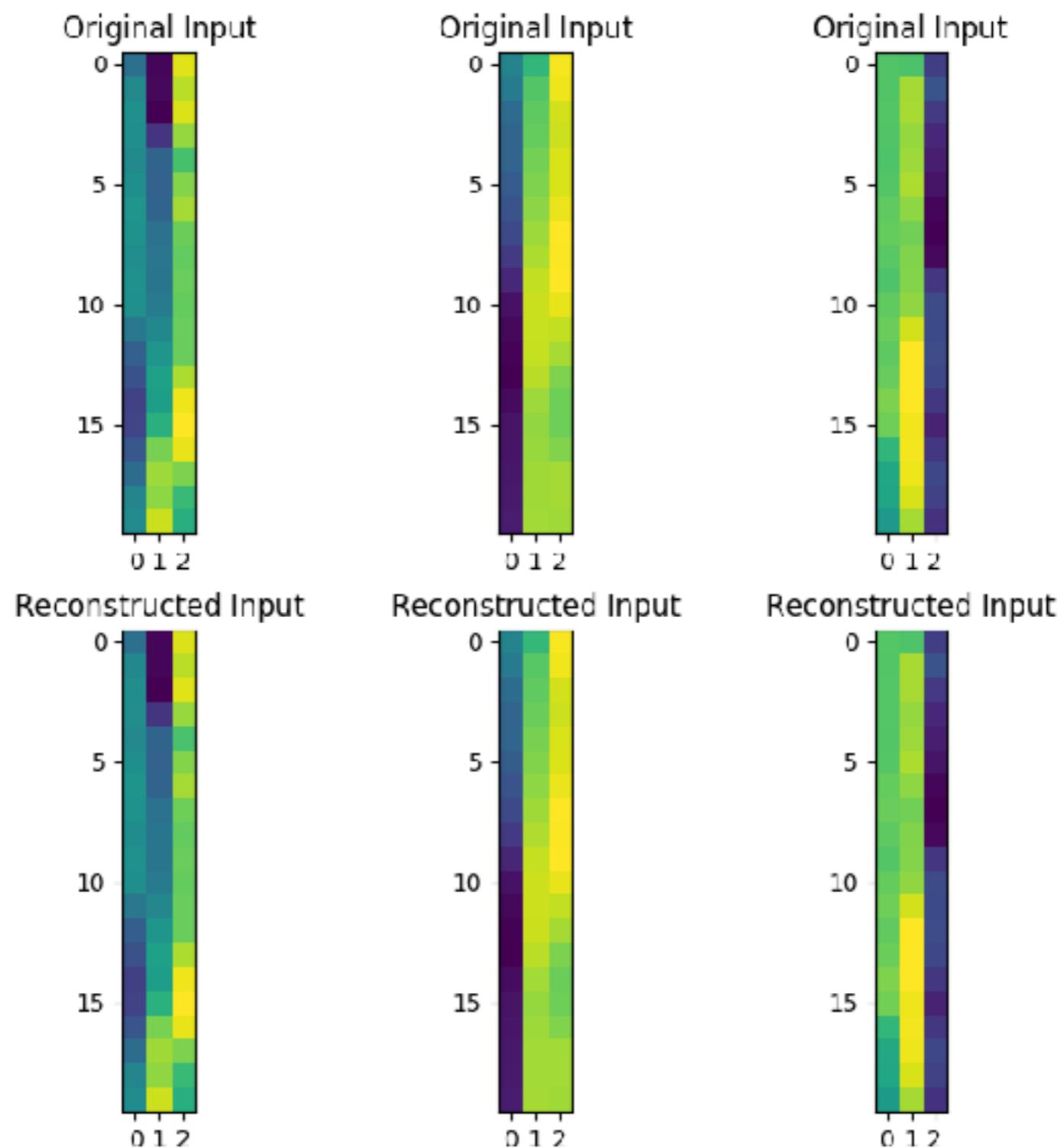
Applying a smoothing filter (Savitzky-Golay)

K-means:

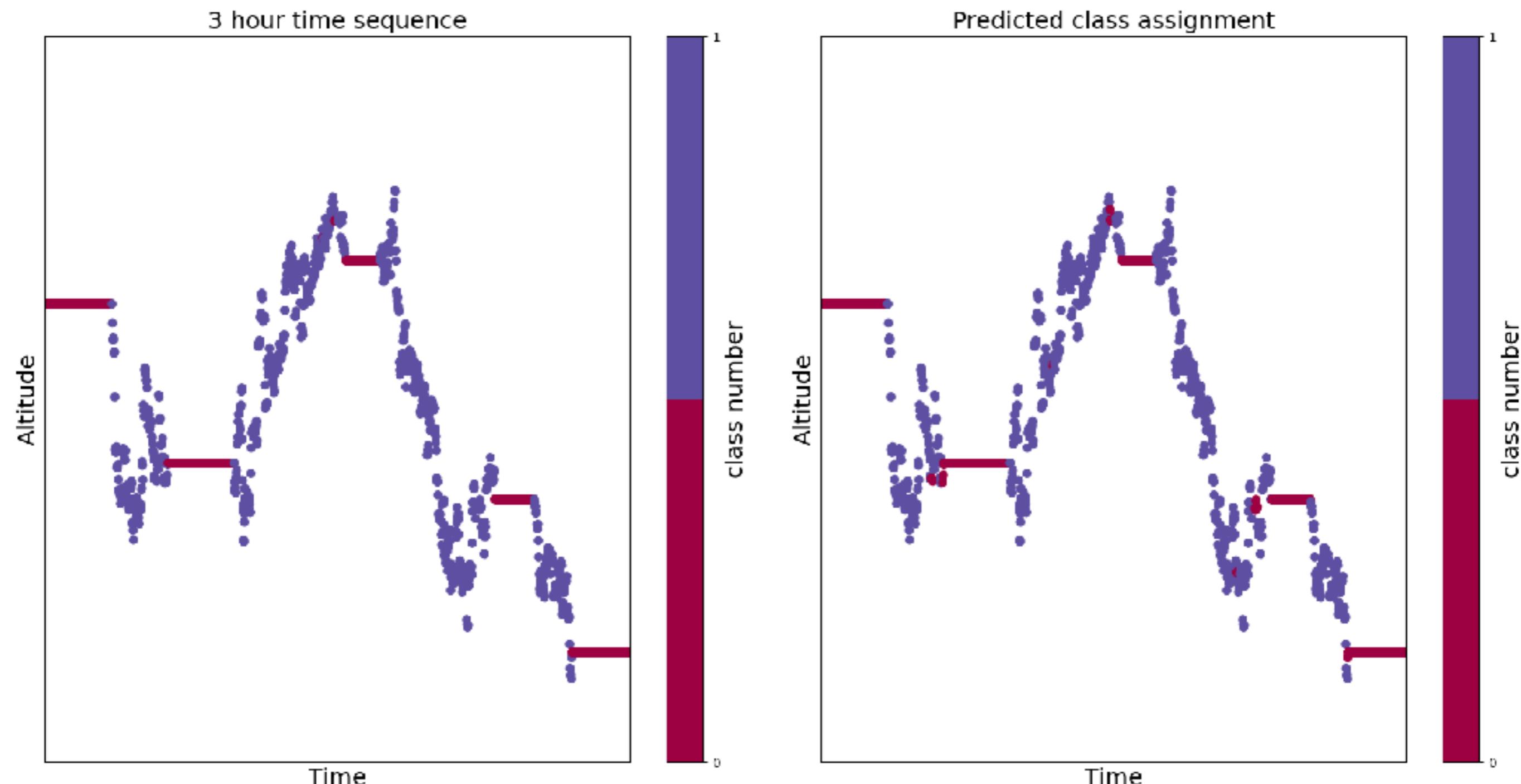
- 0.96 accuracy
- 0.97 F1 score



Input Reconstruction



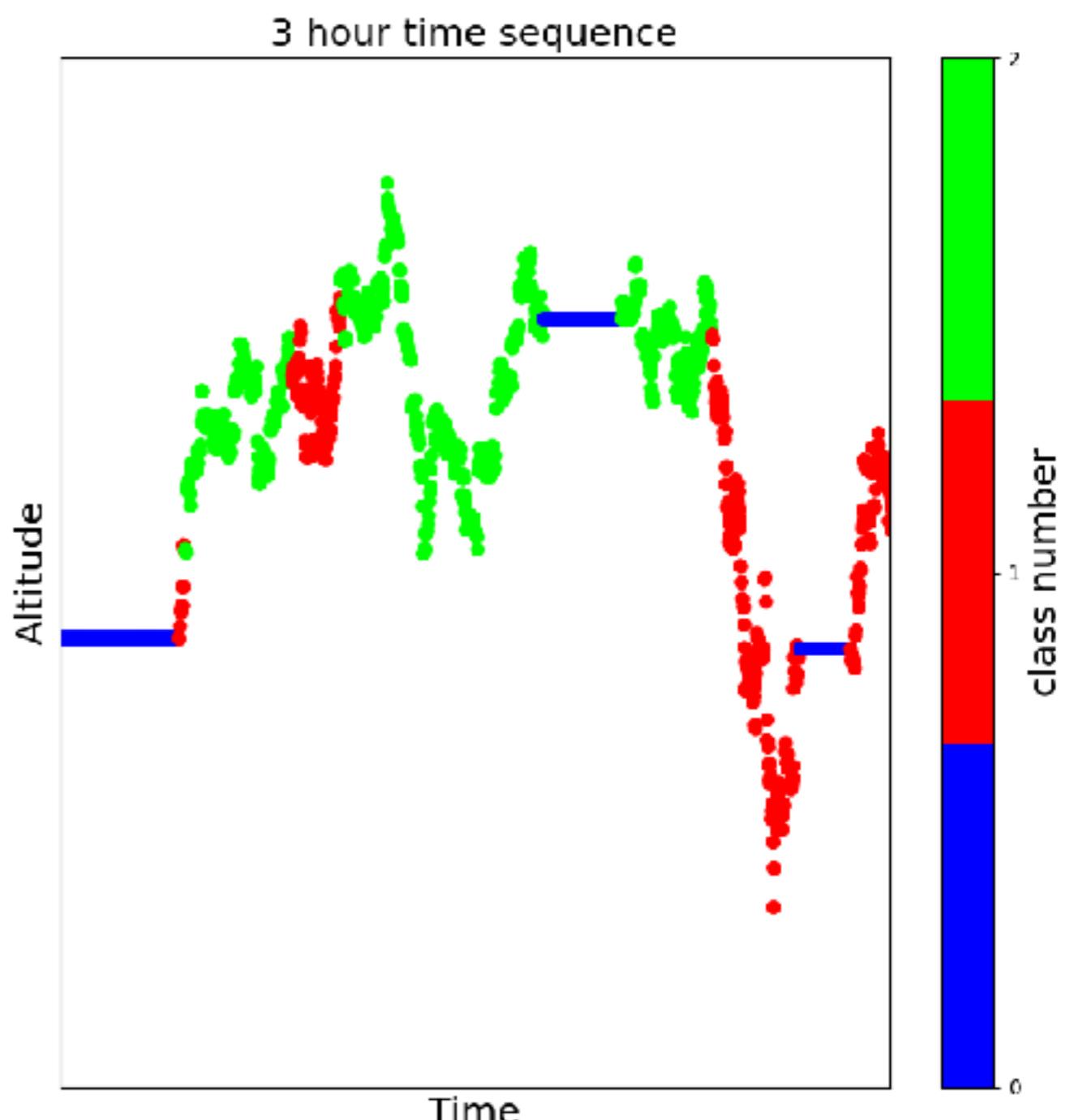
True and Predicted Classes



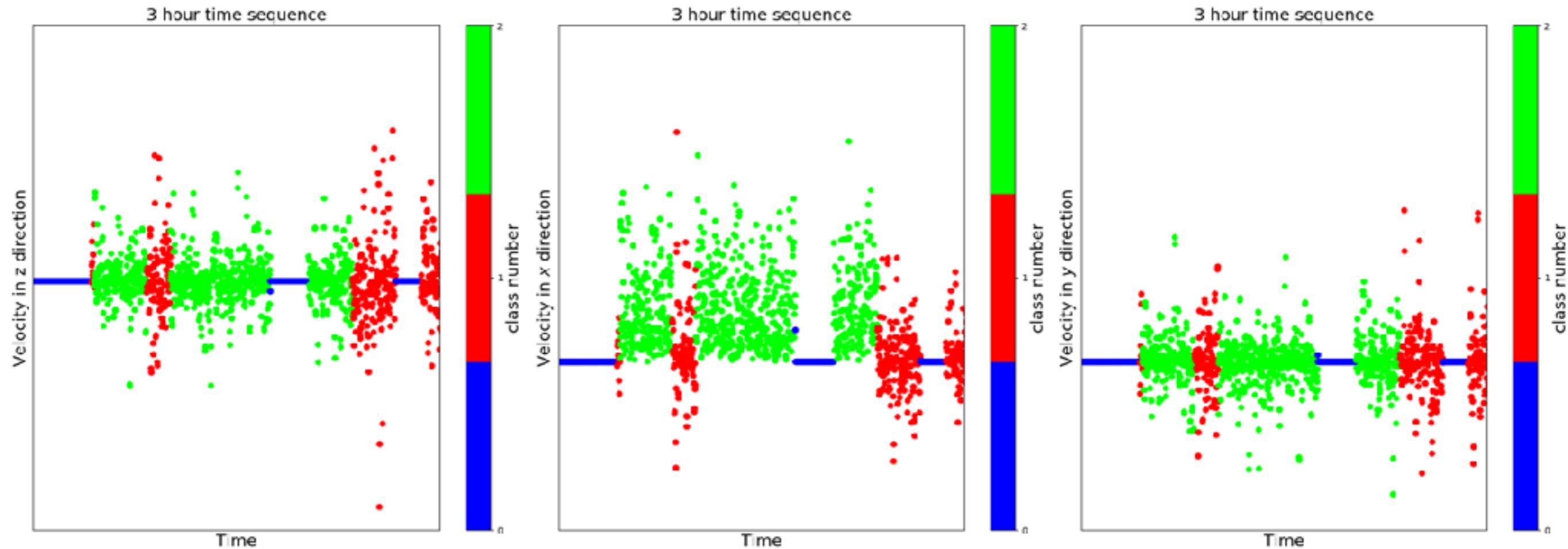
3 Classes

Distinguish between three classes:

- No movement
- Random Walk
- Directional Bias



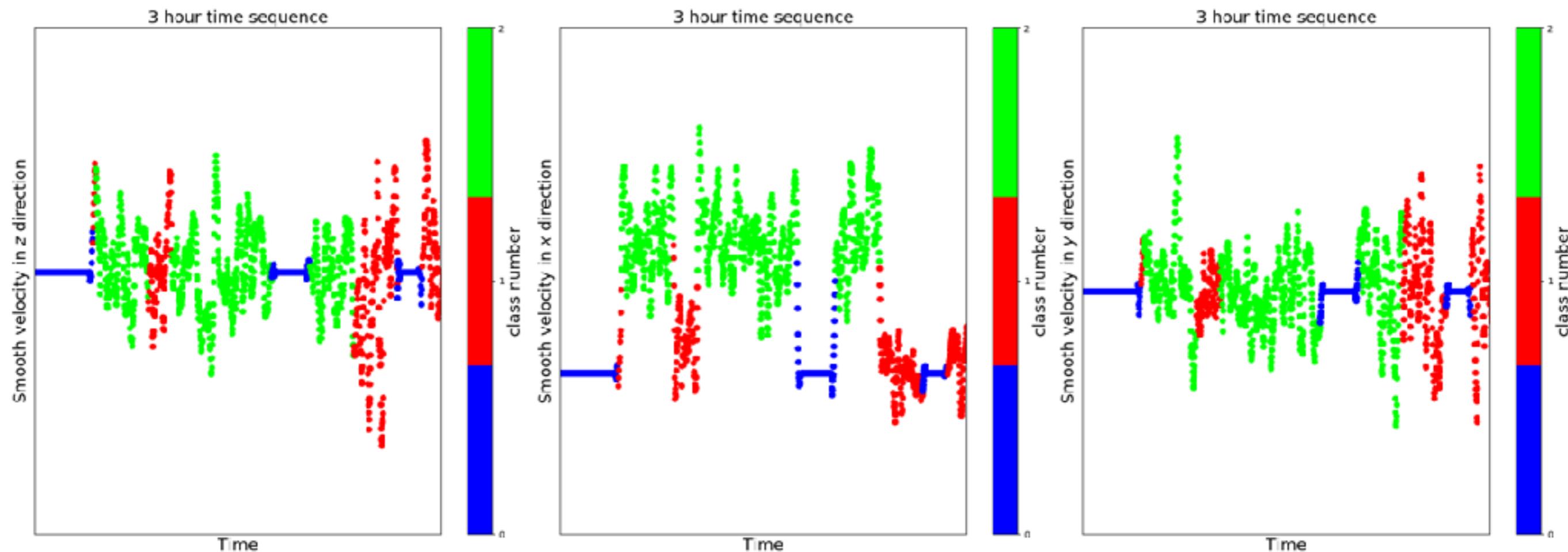
Clustering Raw Data



K-means: 0.45 accuracy and 0.38 F1 score

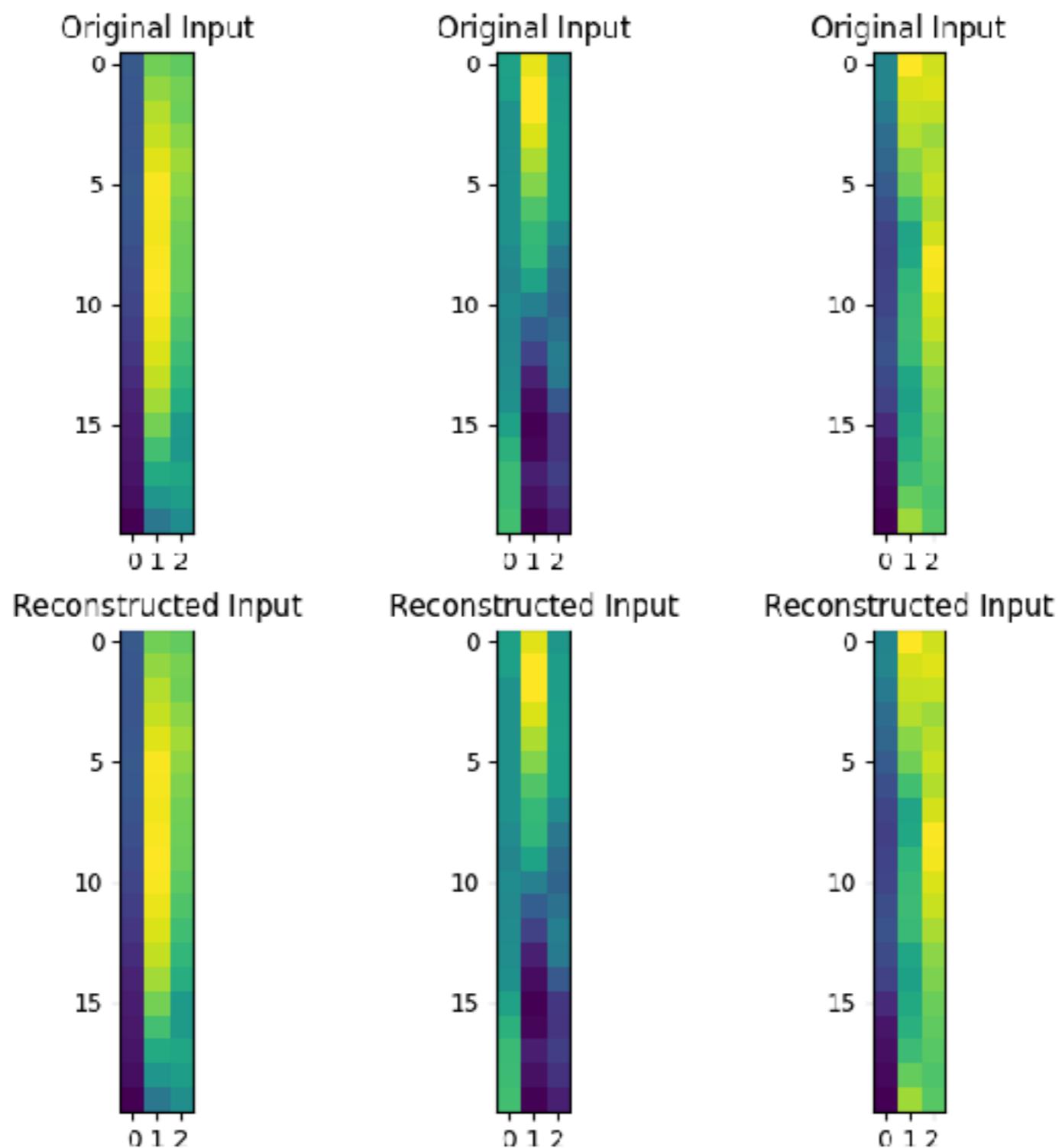
Clustering Raw Data

Applying a smoothing filter (Savitzky-Golay)

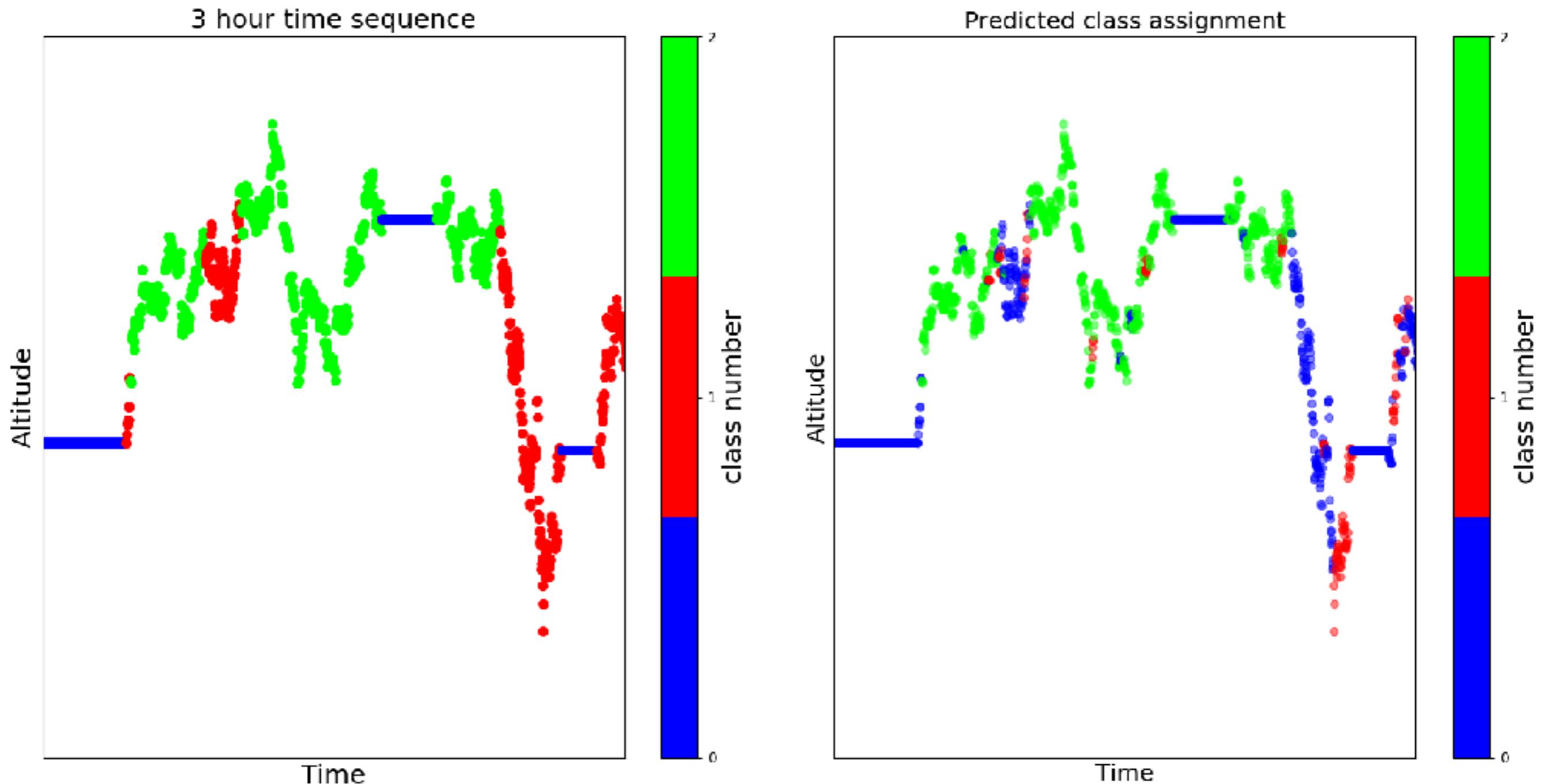


K-means: 0.53 accuracy and 0.47 F1 score

Input Reconstruction



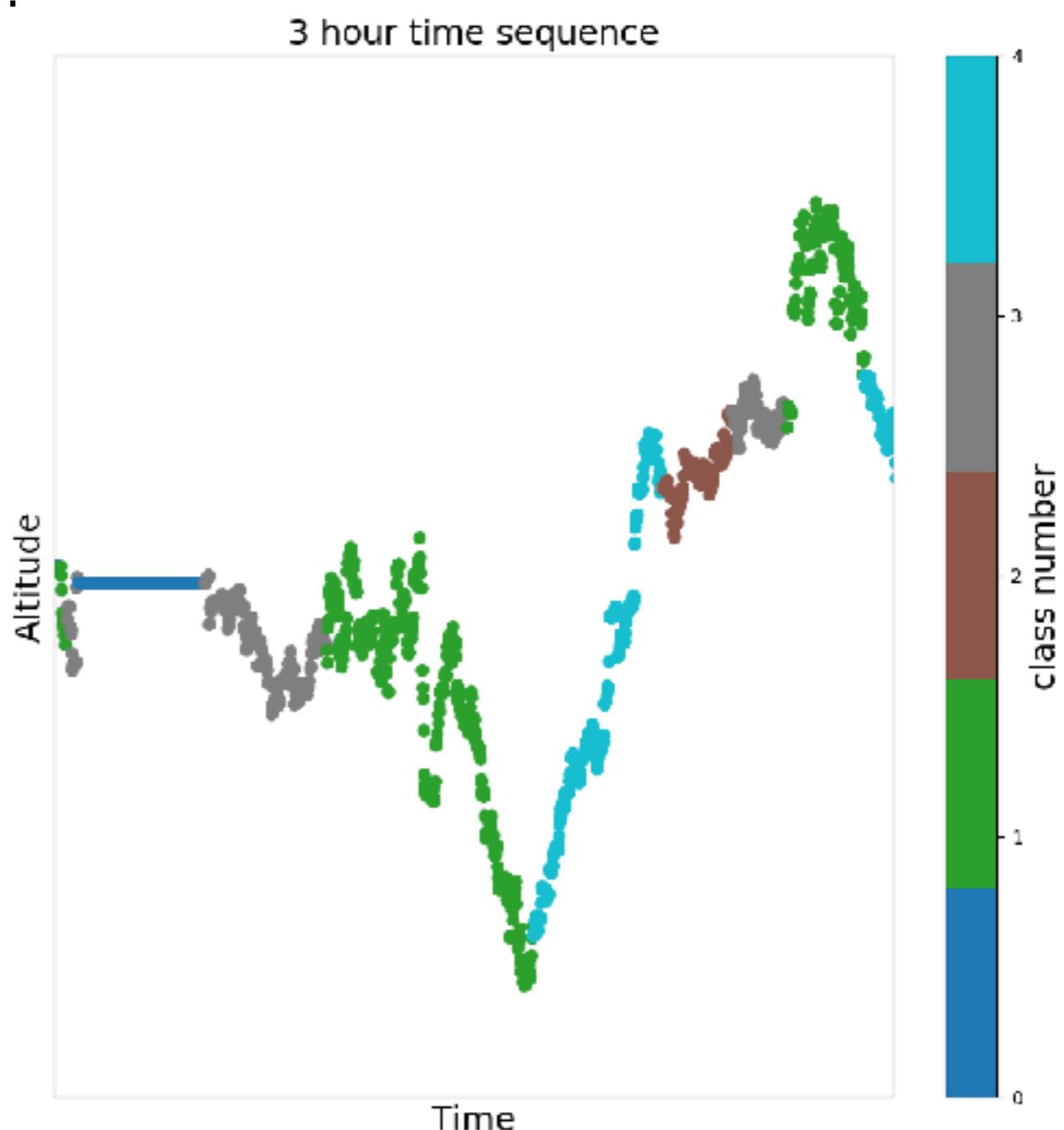
True and Predicted Classes



5 Classes

Distinguish between five classes:

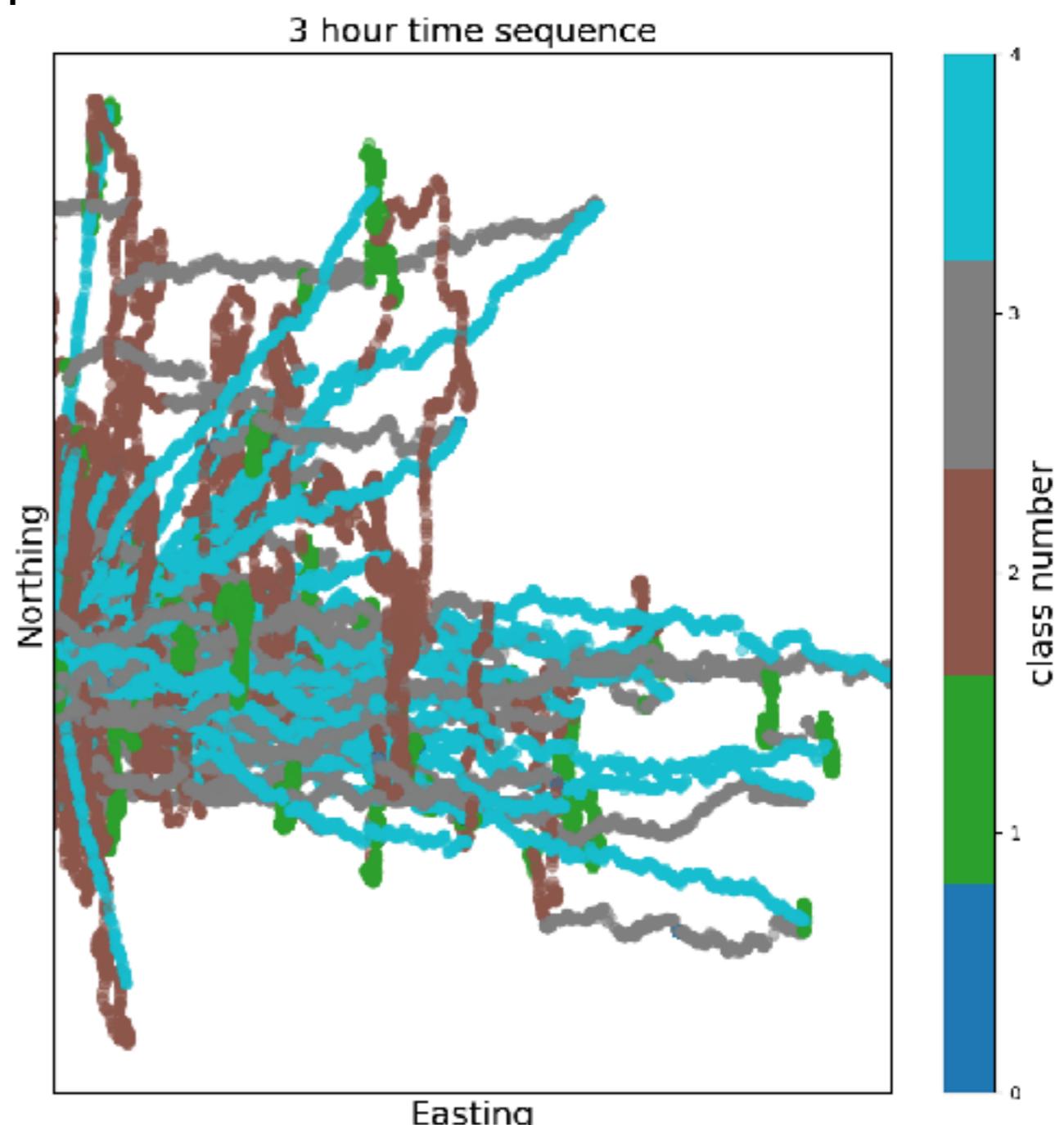
- No movement
- Simple Random Walk
- Correlated Random Walk
- Directional Bias
- Point Bias



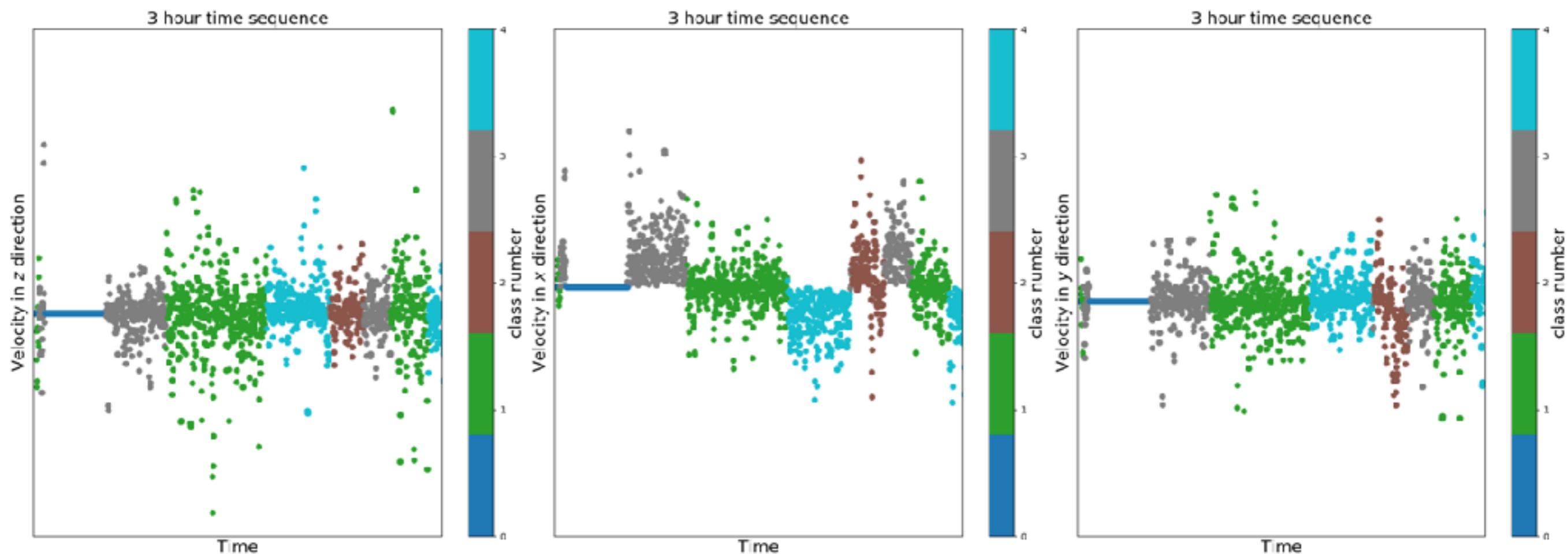
5 Classes

Distinguish between five classes:

- No movement
- Simple Random Walk
- Correlated Random Walk
- Directional Bias
- Point Bias



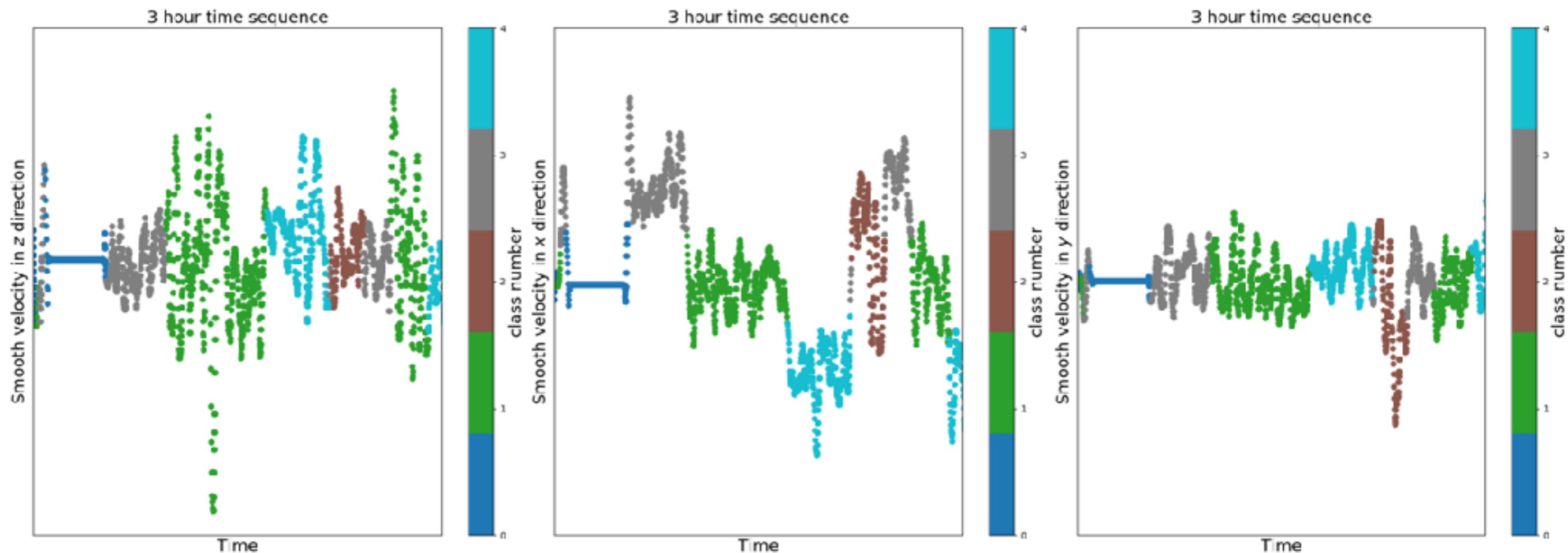
Clustering Raw Data



K-means: 0.40 accuracy and 0.39 F1 score

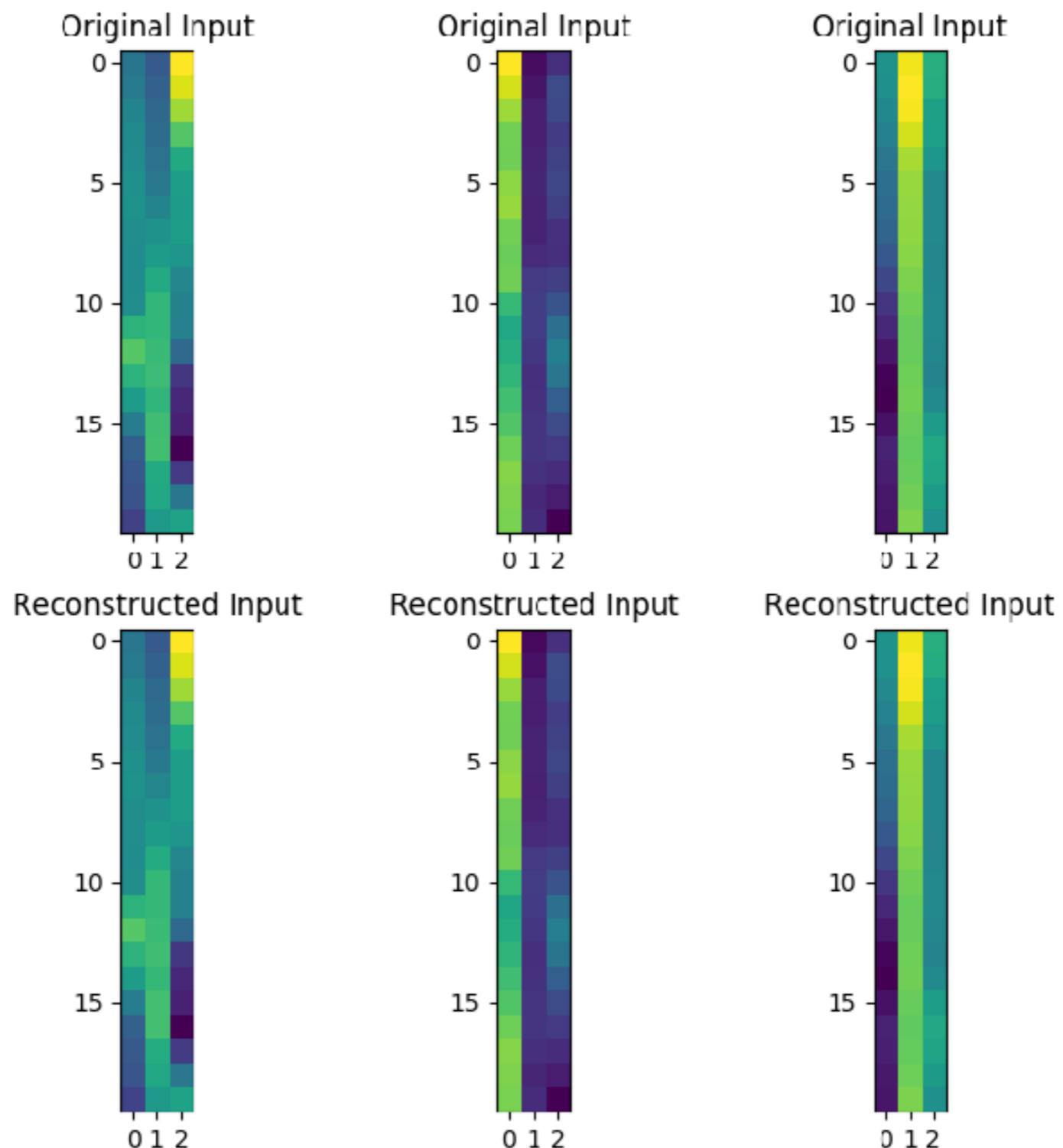
Clustering Raw Data

Applying a smoothing filter (Savitzky-Golay)

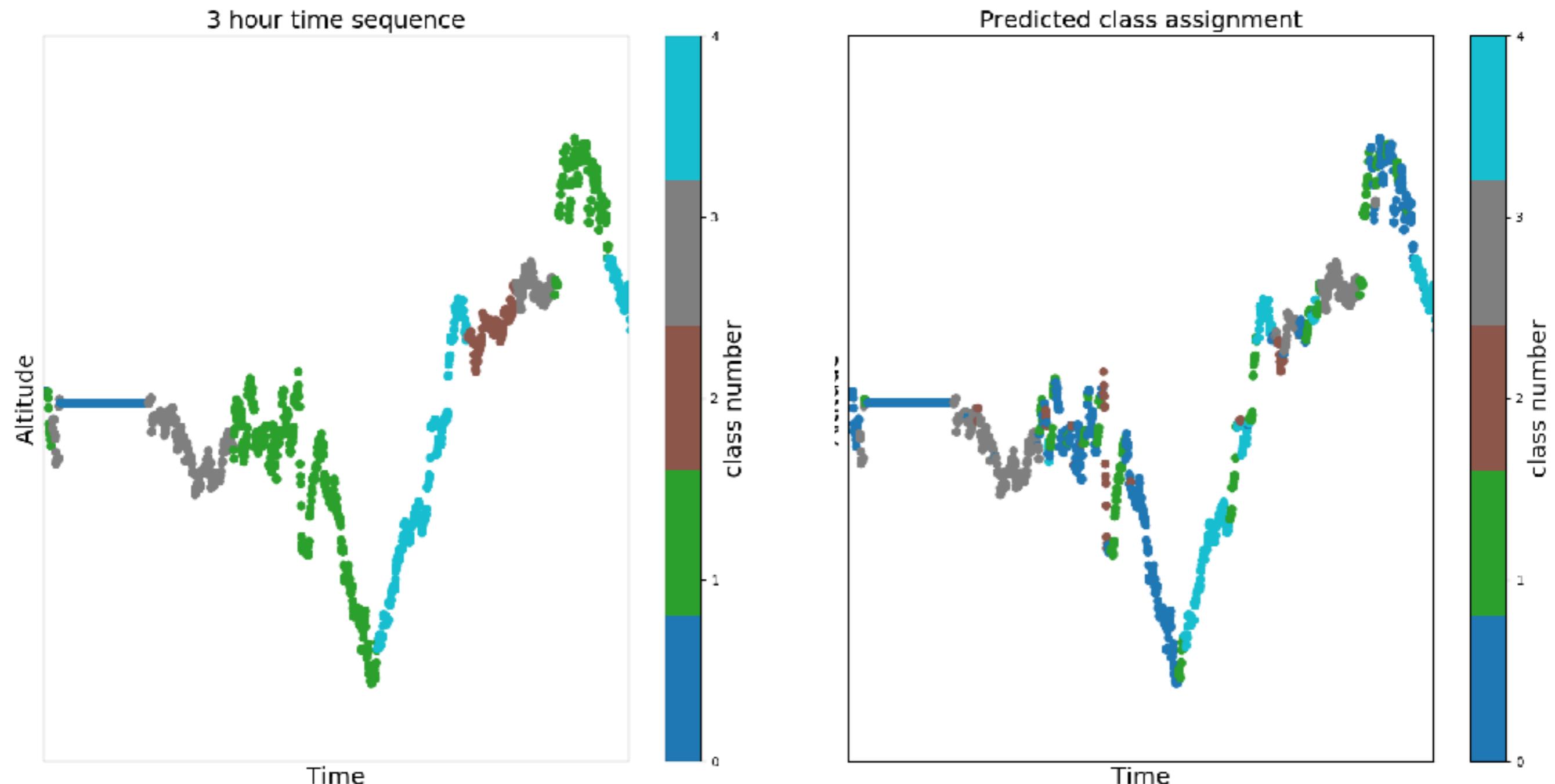


K-means: 0.52 accuracy and 0.48 F1 score

Input Reconstruction

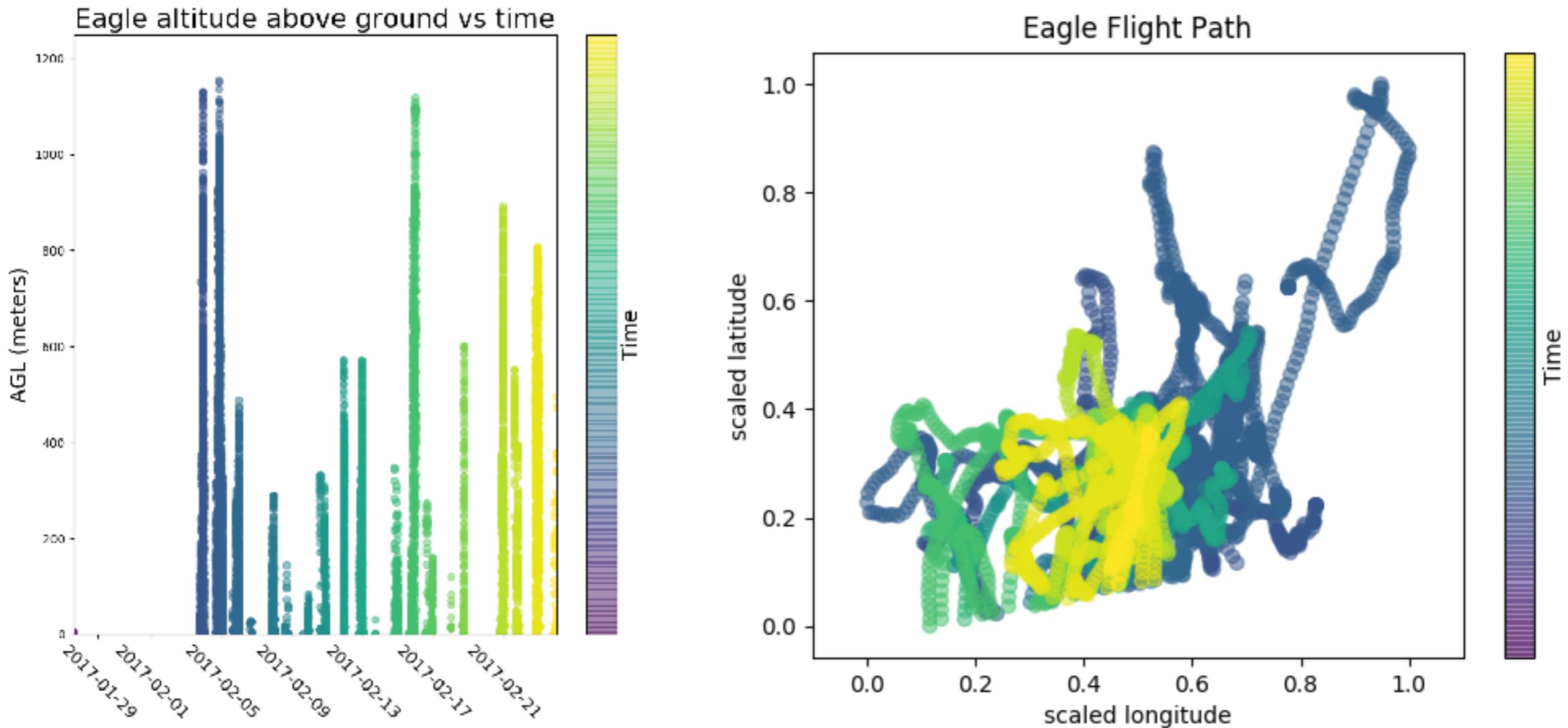


True and Predicted Classes

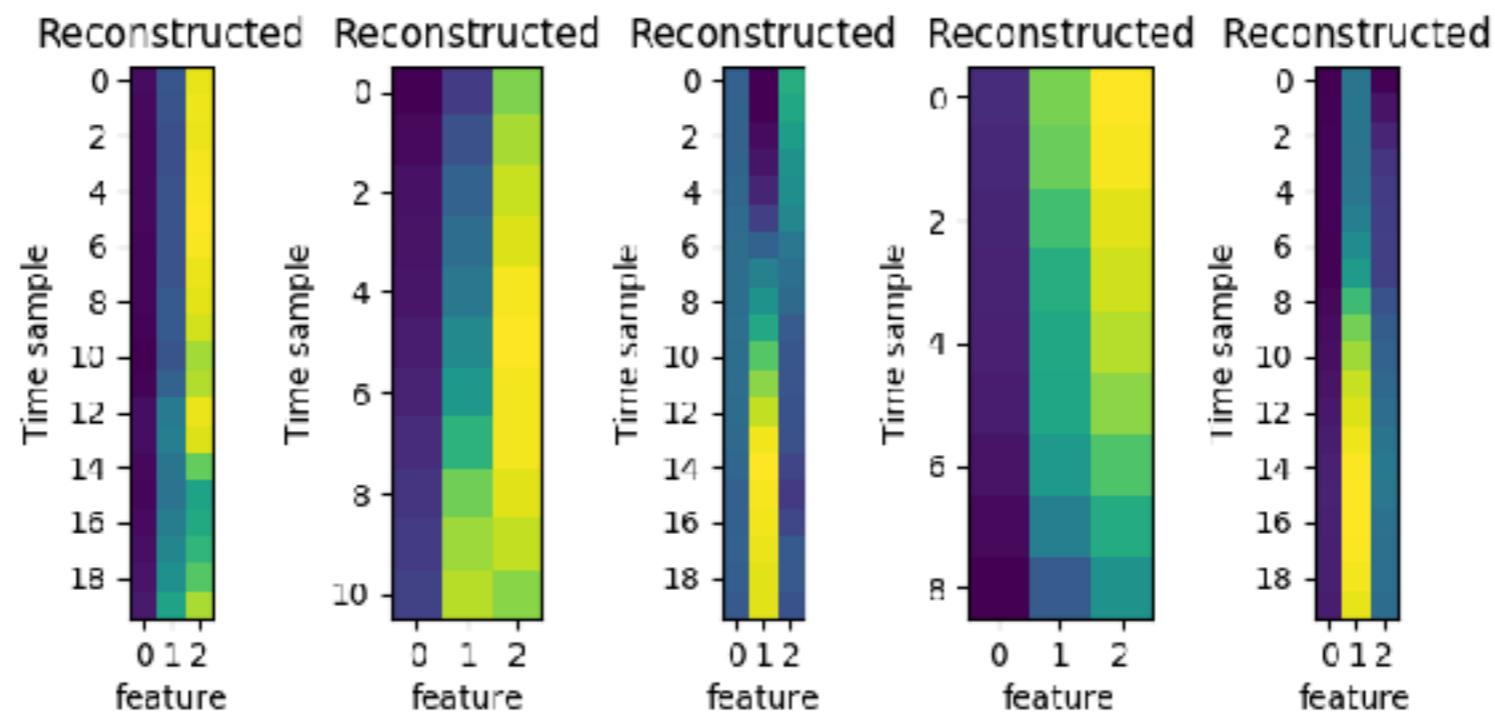
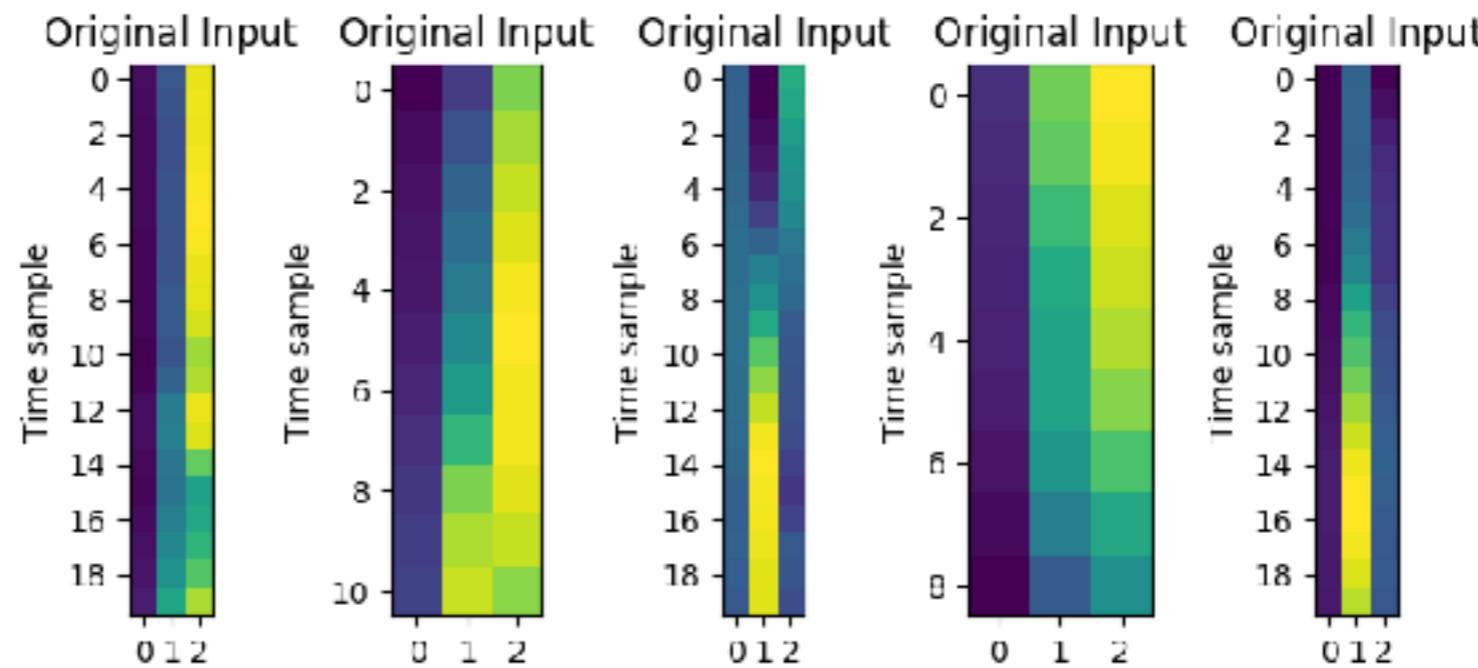


Real Data Application

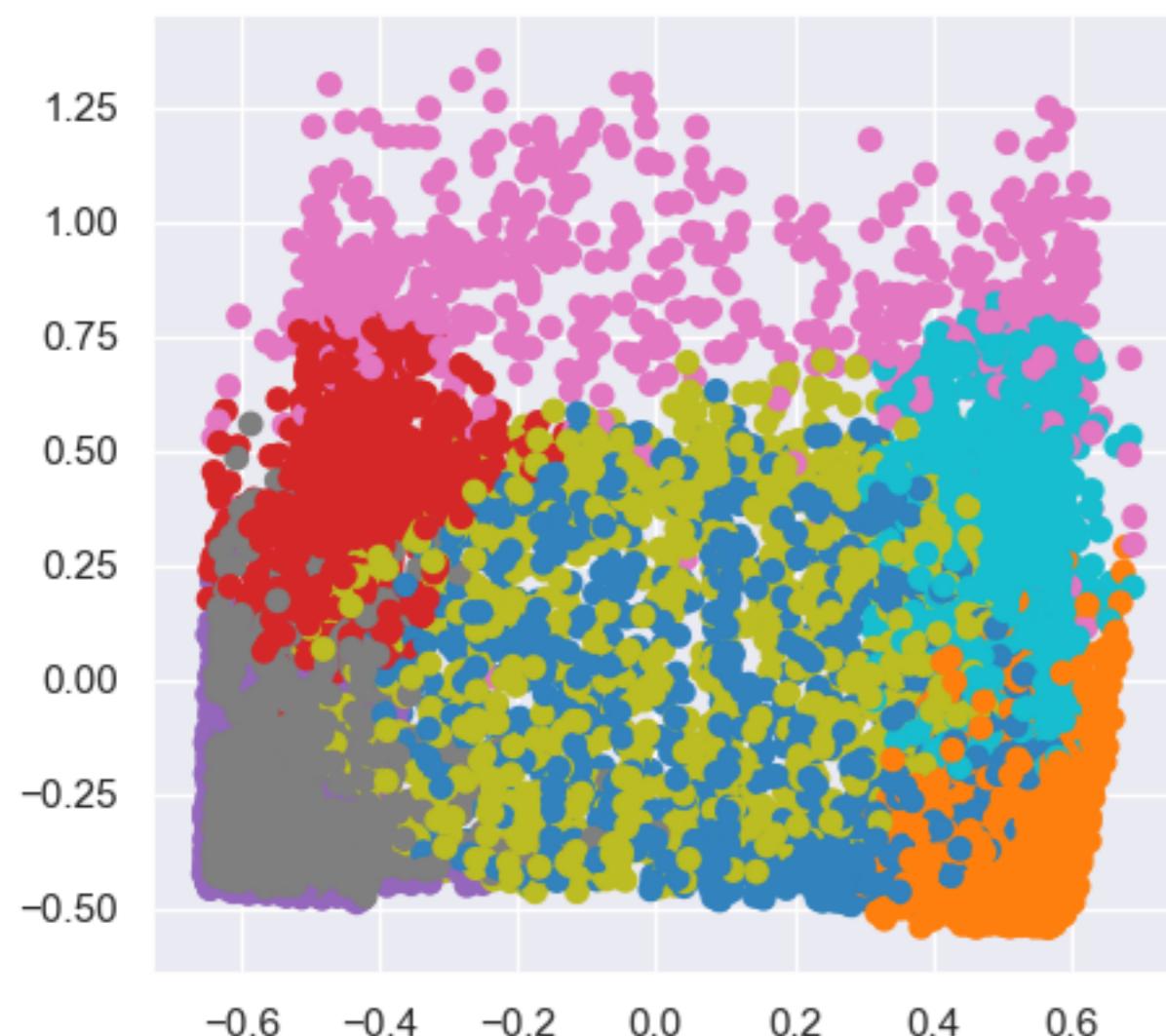
Telemetry Data for One Golden Eagle



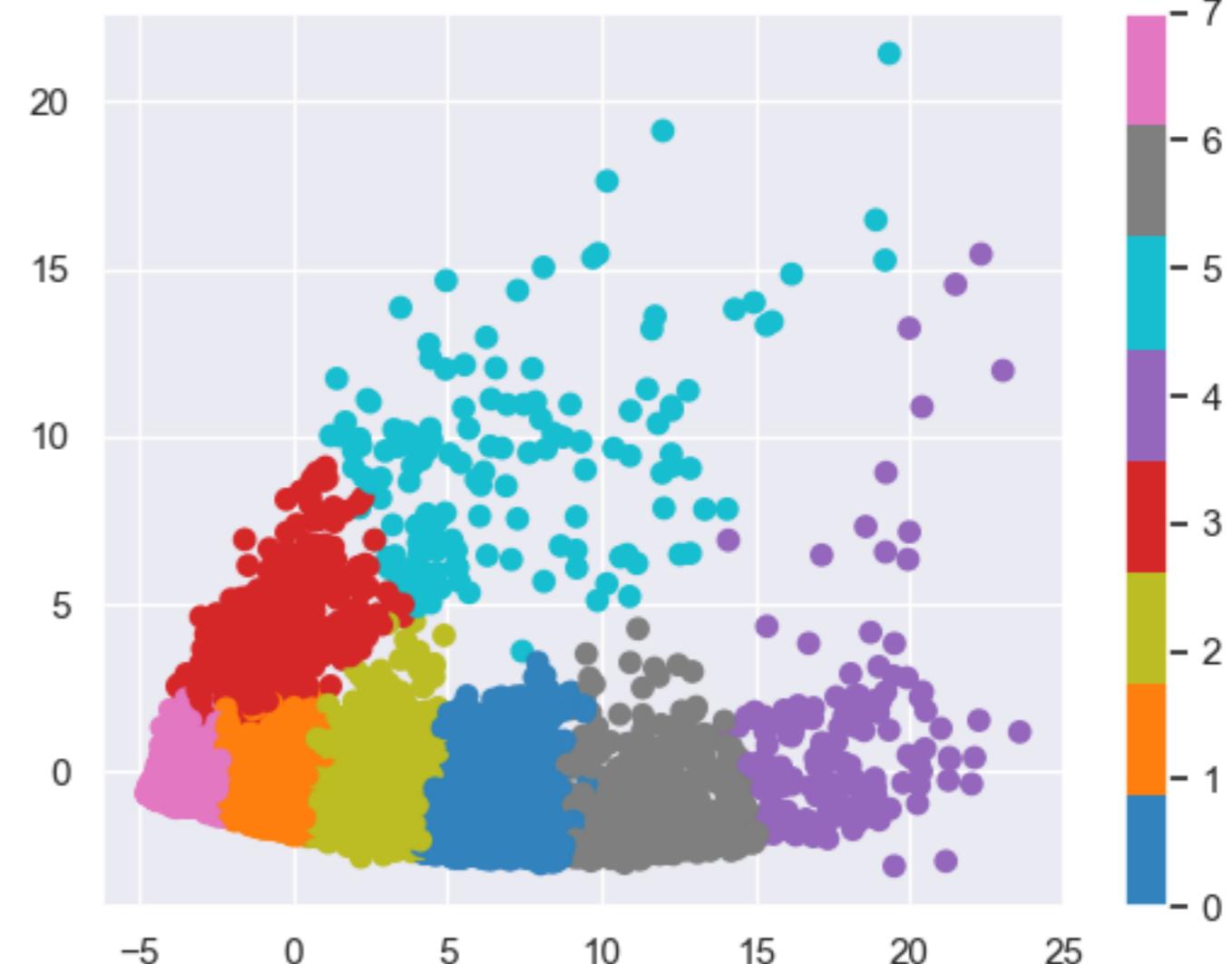
Input Reconstruction



Clustering

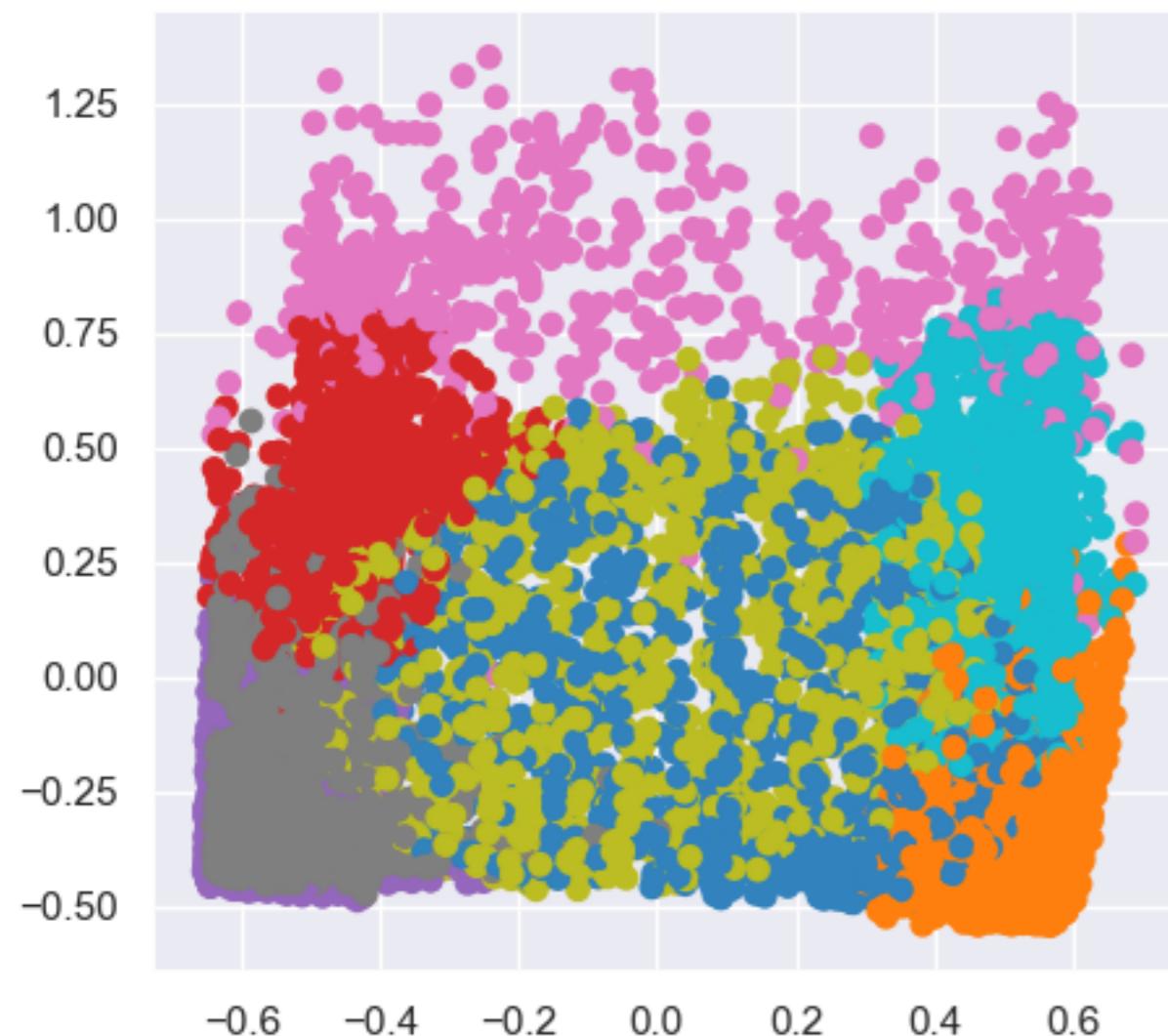


Original Data Space

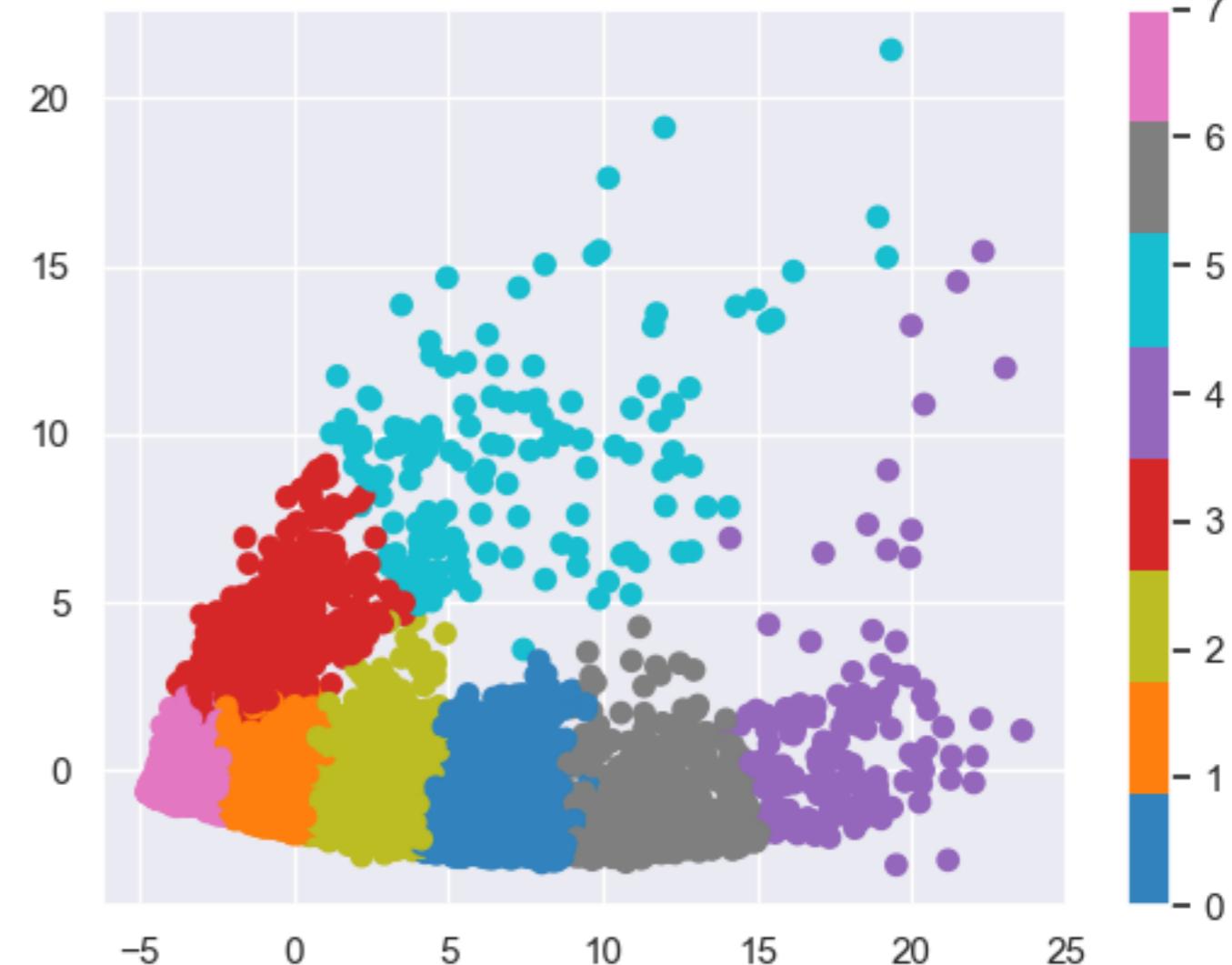


Encoded Space

Clustering



Original Data Space

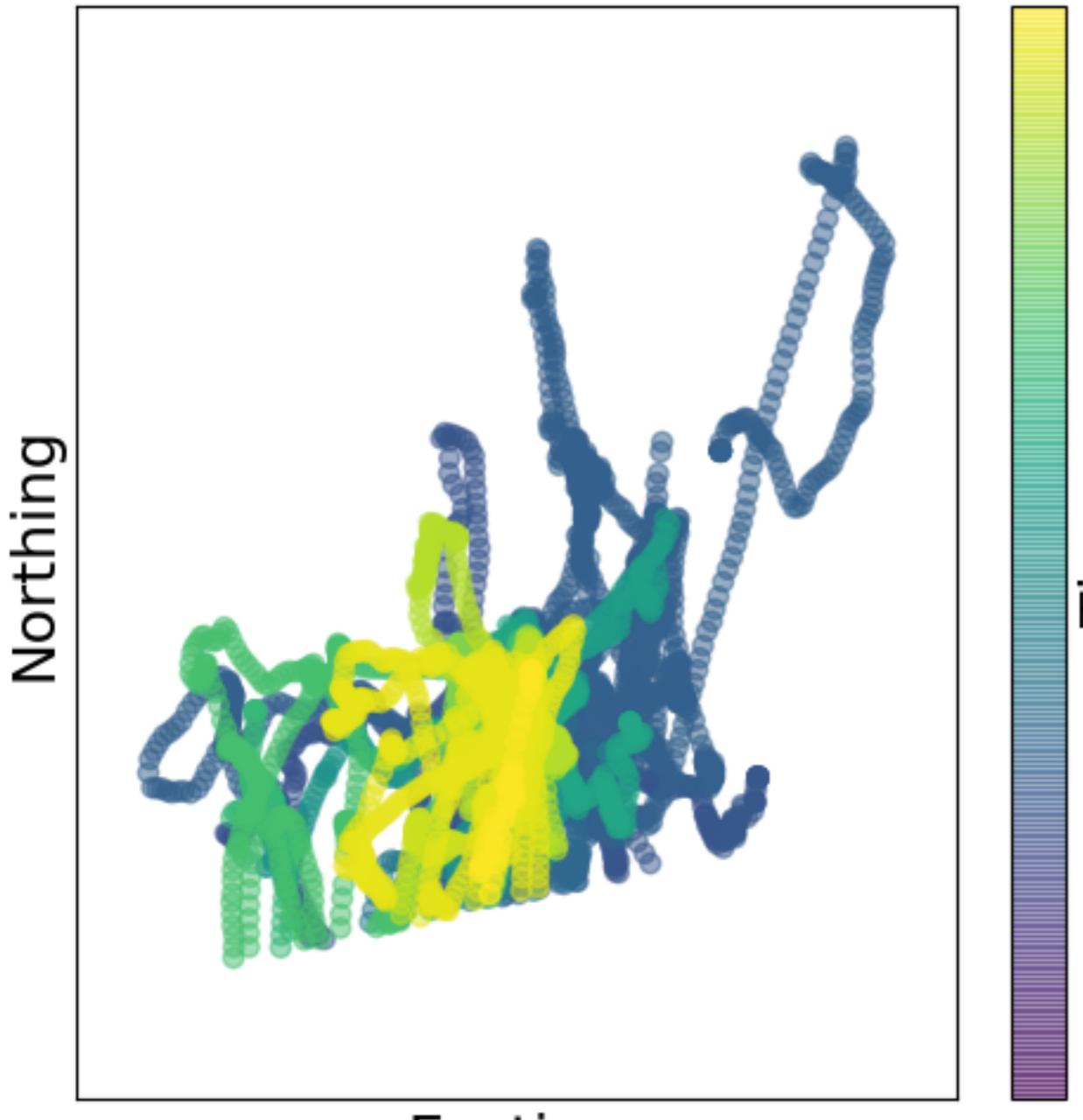


Encoded Space

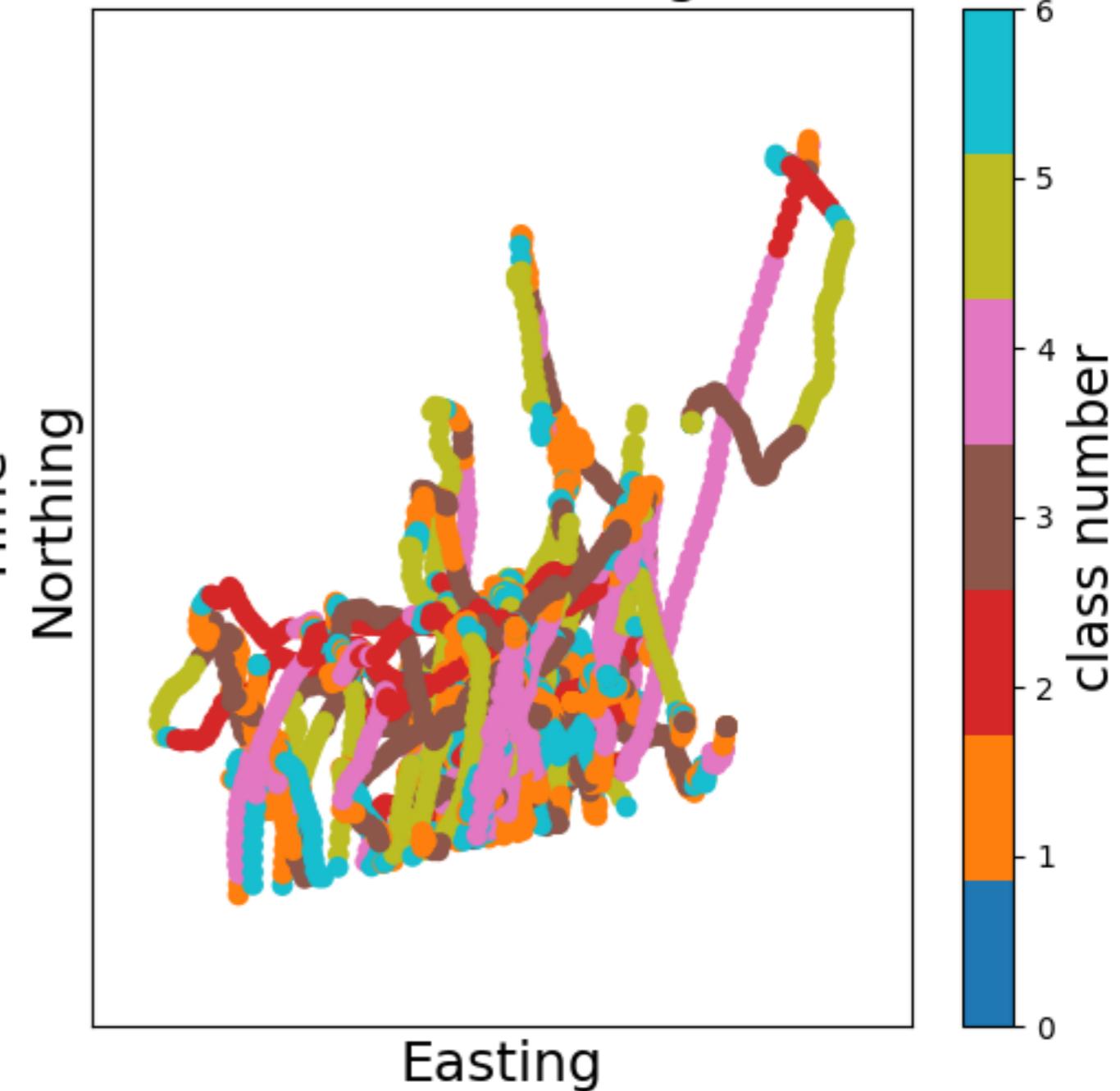
Better cluster separation

Flight Path with Assigned Classes

Eagle Flight Path



Predicted class assignment



Conclusions

Deep LSTM autoencoders can take account of time dependency in the data

Clustering in the encoded space leads to more separated clusters

Choice of clustering algorithm matters

Future Research

Eagle behavior from movement patterns

Variational autoencoders

Add noise to the simulations

Questions