Adaptive Exploration for Geological Classification

Team:
Aishanou Osha Rait
Shivam Gautam
Tushar Agrawal
Bishwamoy Sinha Roy

Overview

- Problem Introduction
- Simulation Environment
- Problem 1 – Sampling Point Selection
  - Approach 1 - Variance Method
  - Approach 2 - Clustering Method
- Problem 2 – Planning Path To Sampling Point
  - Approach 1 - Dynamic Programming (DP)
  - Approach 2 - Multi-Heuristic A*
- Conclusion

Problem Introduction

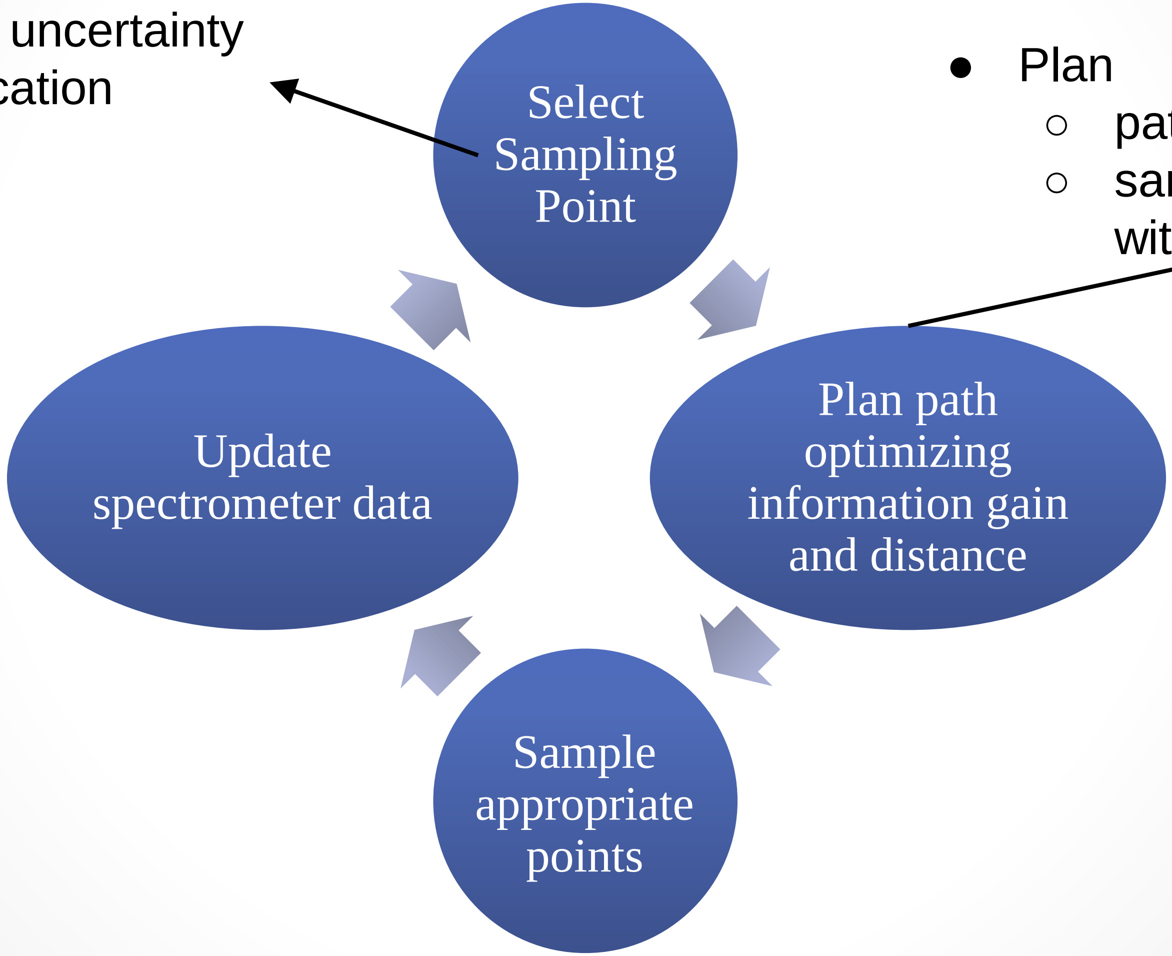
- Rover exploration of extraterrestrial bodies is slow
- One of the motives of exploration – **Rock classification**
- Spectrometer readings to classify rocks

- Satellite Spectrometers have **low wavelength and spatial resolution** and high sensor noise
- Ground based spectrometer sampling is **costly**
- Optimization Problem
- **Where to sample such that the rock classification has lowest uncertainty.**

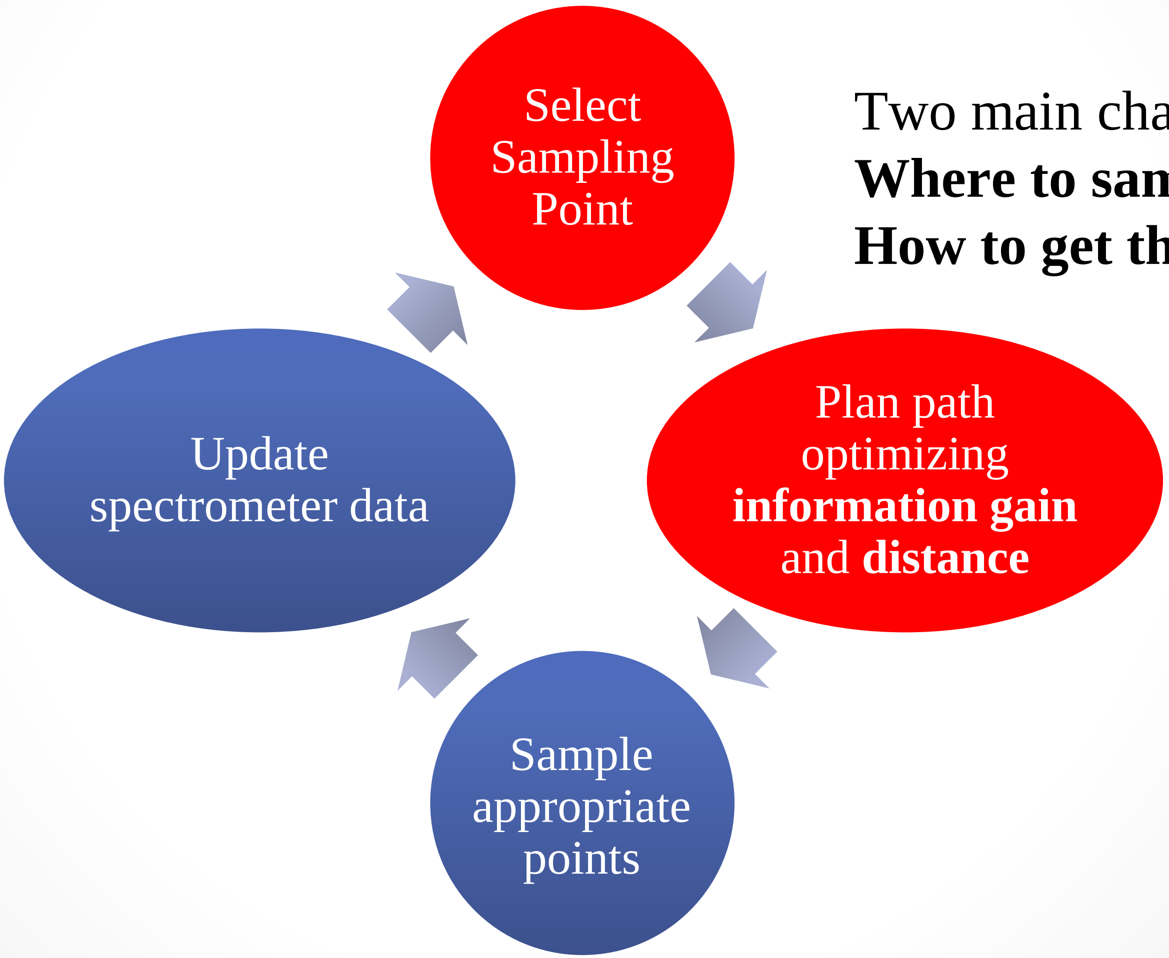
Problem Introduction

Decrease uncertainty in classification

- Plan
  - path
  - sampling points within path



Problem Introduction



Two main challenges:

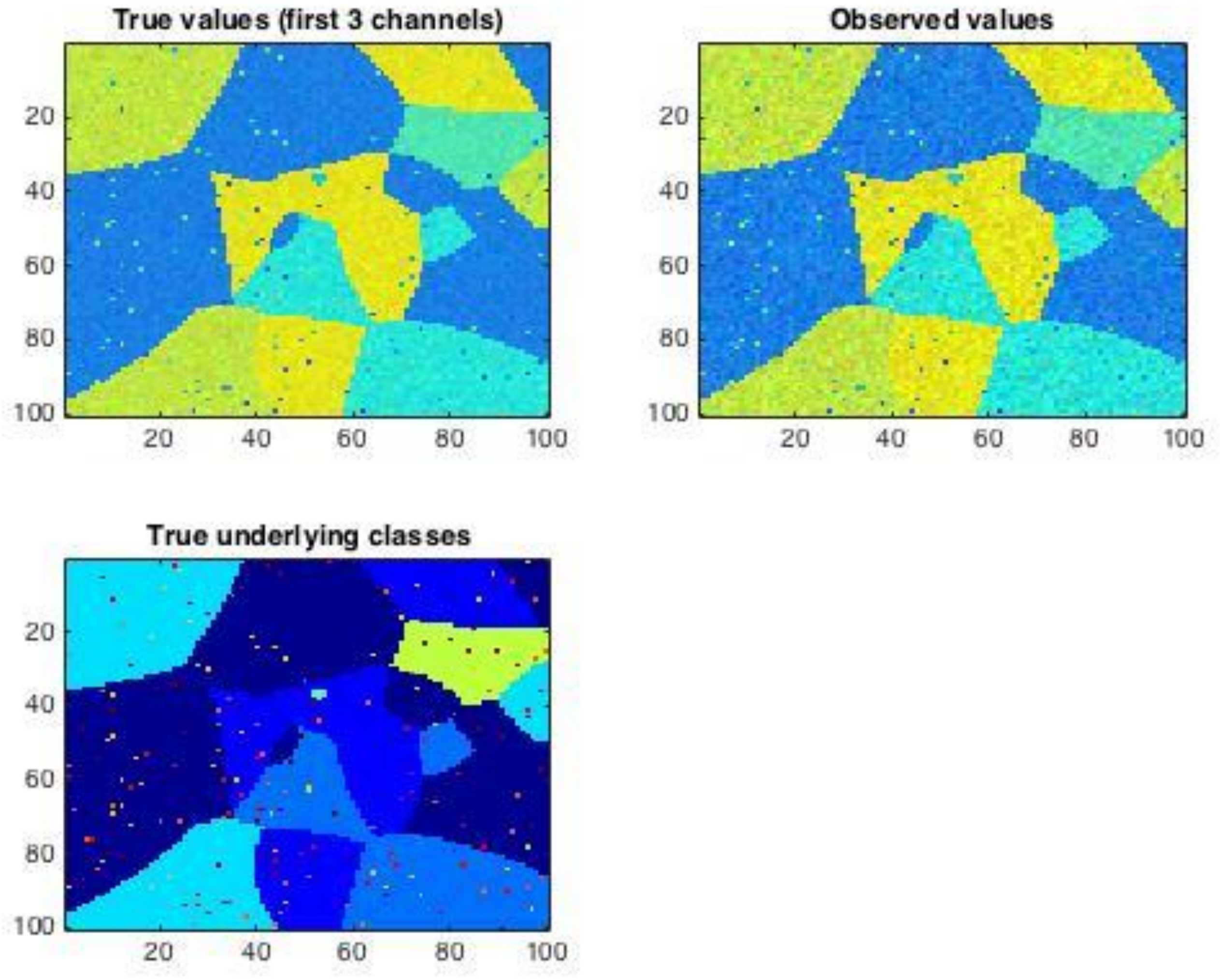
Where to sample

How to get there (just path)

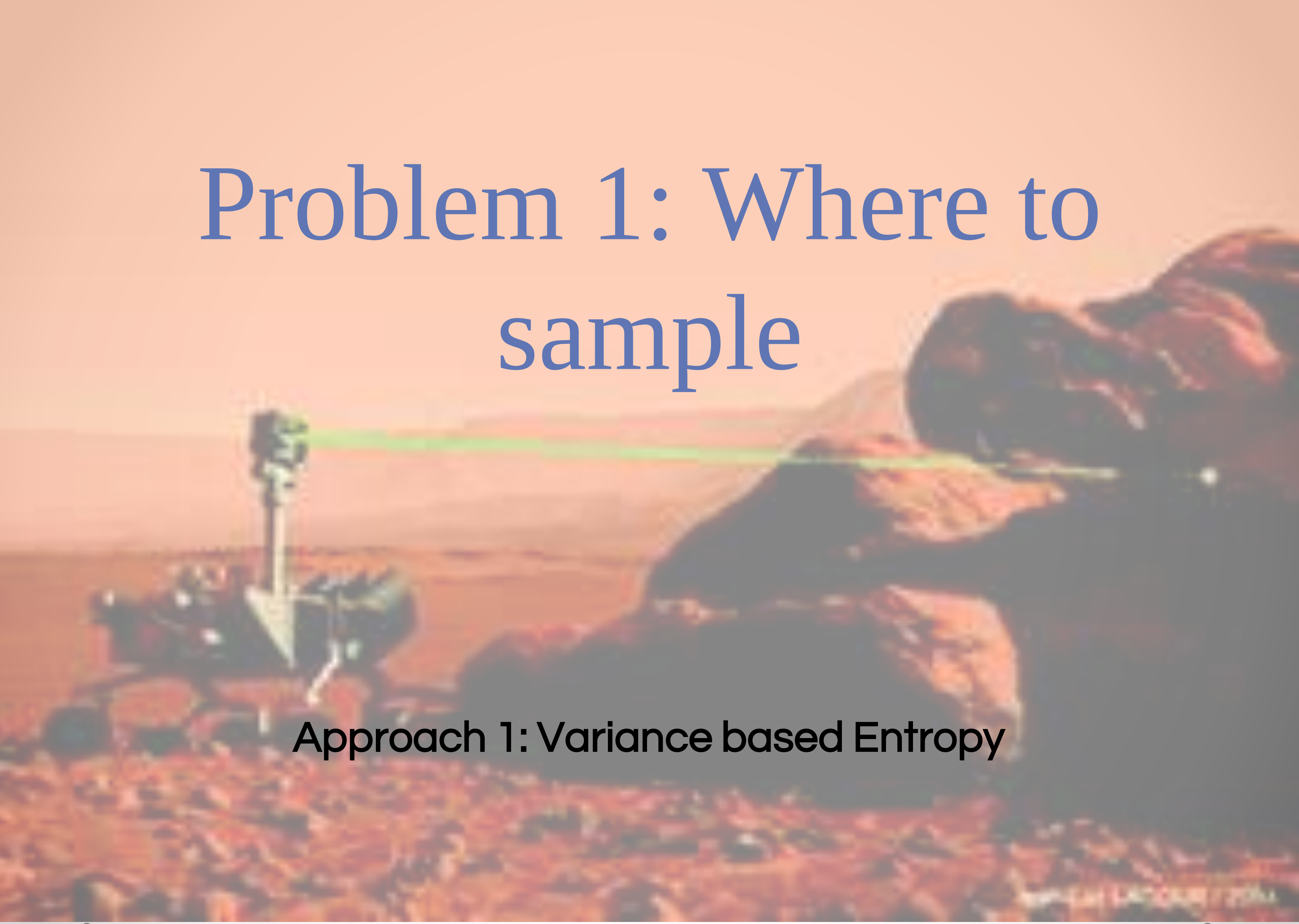
Simulation Environment

- Provides three pieces of information
- Information 1 – Satellite Reading for a region
  - Containing less number of channel
- Information 2 – Rover reading at each point in the region
  - Result of sampling a point – more number of channels (higher resolution)
- Information 3 – True classification of each point
- We control
  - Noise in sensor
  - Number of classes
  - Number of dominant and rare classes

Simulation Environment



Problem 1: Where to sample



Approach 1: Variance based Entropy

Differential Entropy-Greedy Approach

- Starting sample set: S={S1}
- Calculate Entropy of each point in map with the sample set using

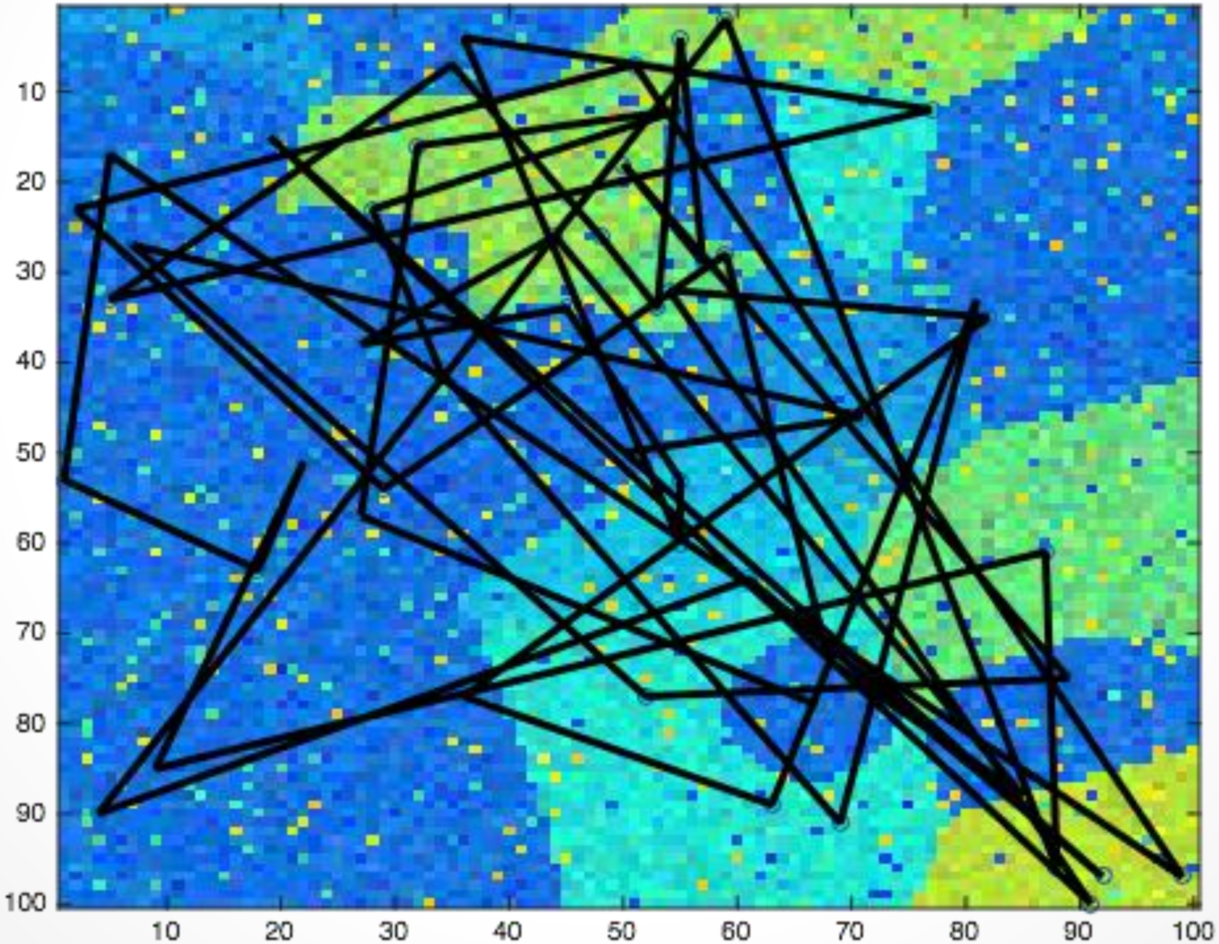
$$H(S) = \frac{1}{2} \sum_{b}^{B} \log(2\pi\sigma_b)$$

where, σ_b is the variance in the bth band of S

- Maximum Entropy point is the next sample point
- Travel to point S2, sample, and add it to the sample set: S={S1,S2}
- Repeat entropy calculation
- Stop after 50 samples

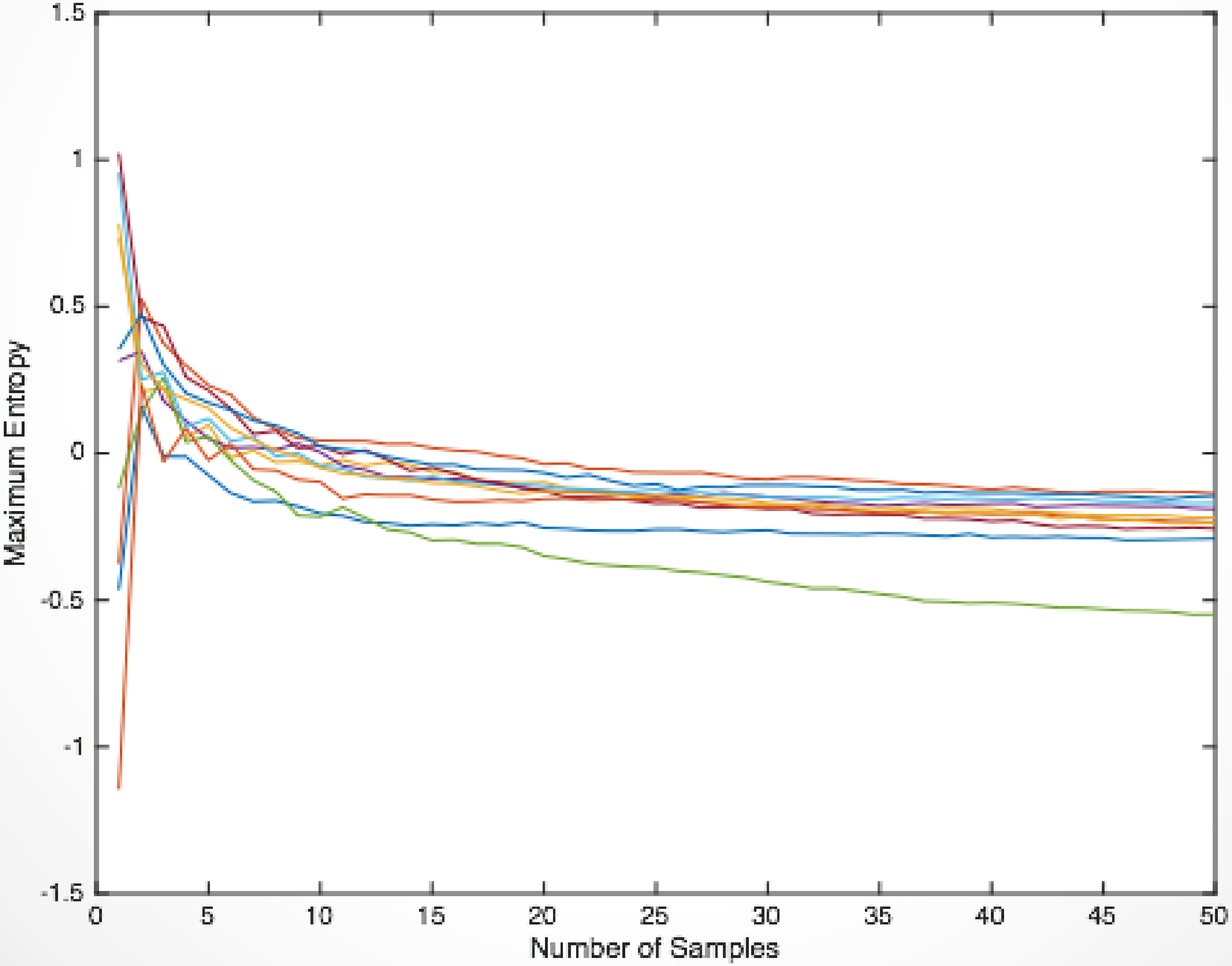
Greedy Approach - Full map

Starting point - center of map



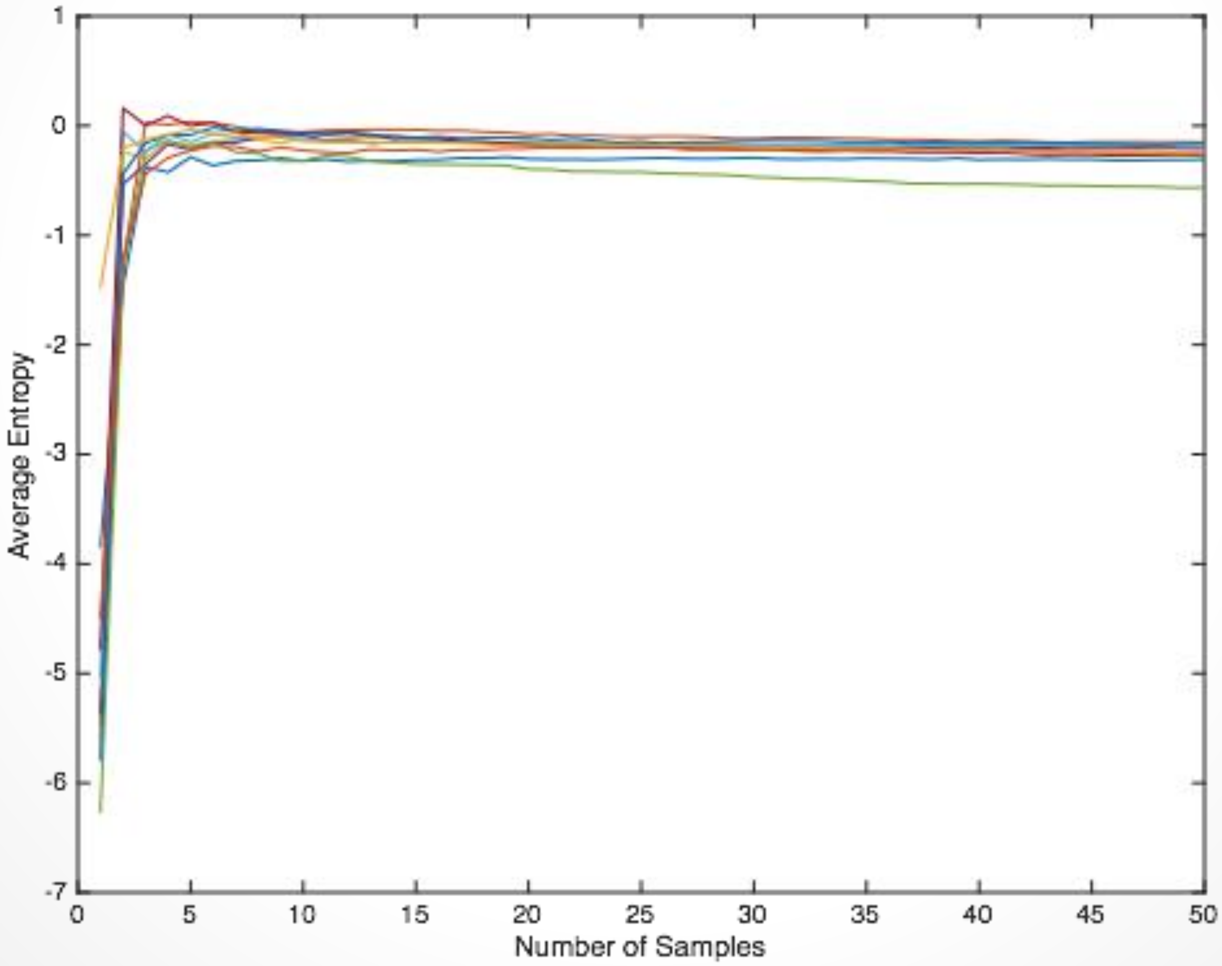
Greedy Approach - Full map

No. of Test Data = 10



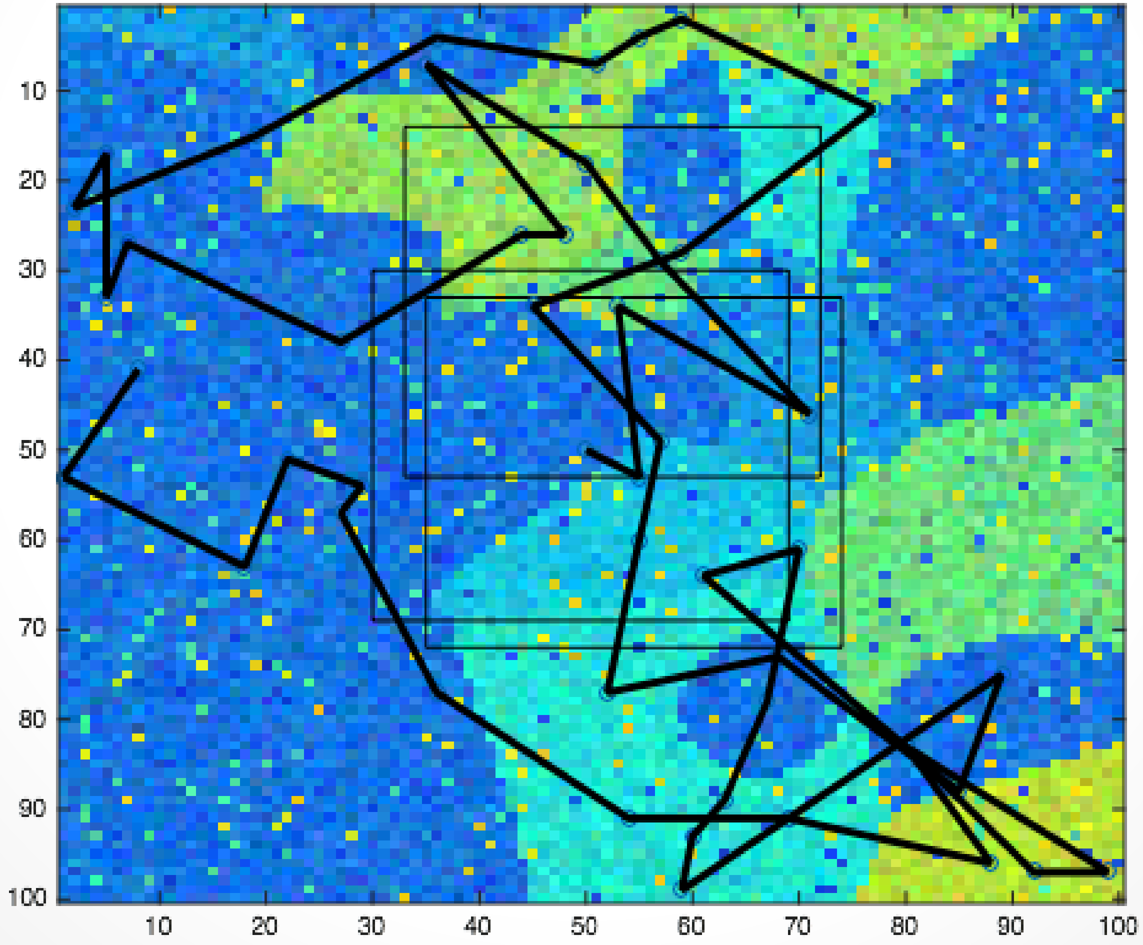
Greedy Approach - Full map

No. of Test Data = 10



Greedy Approach - window

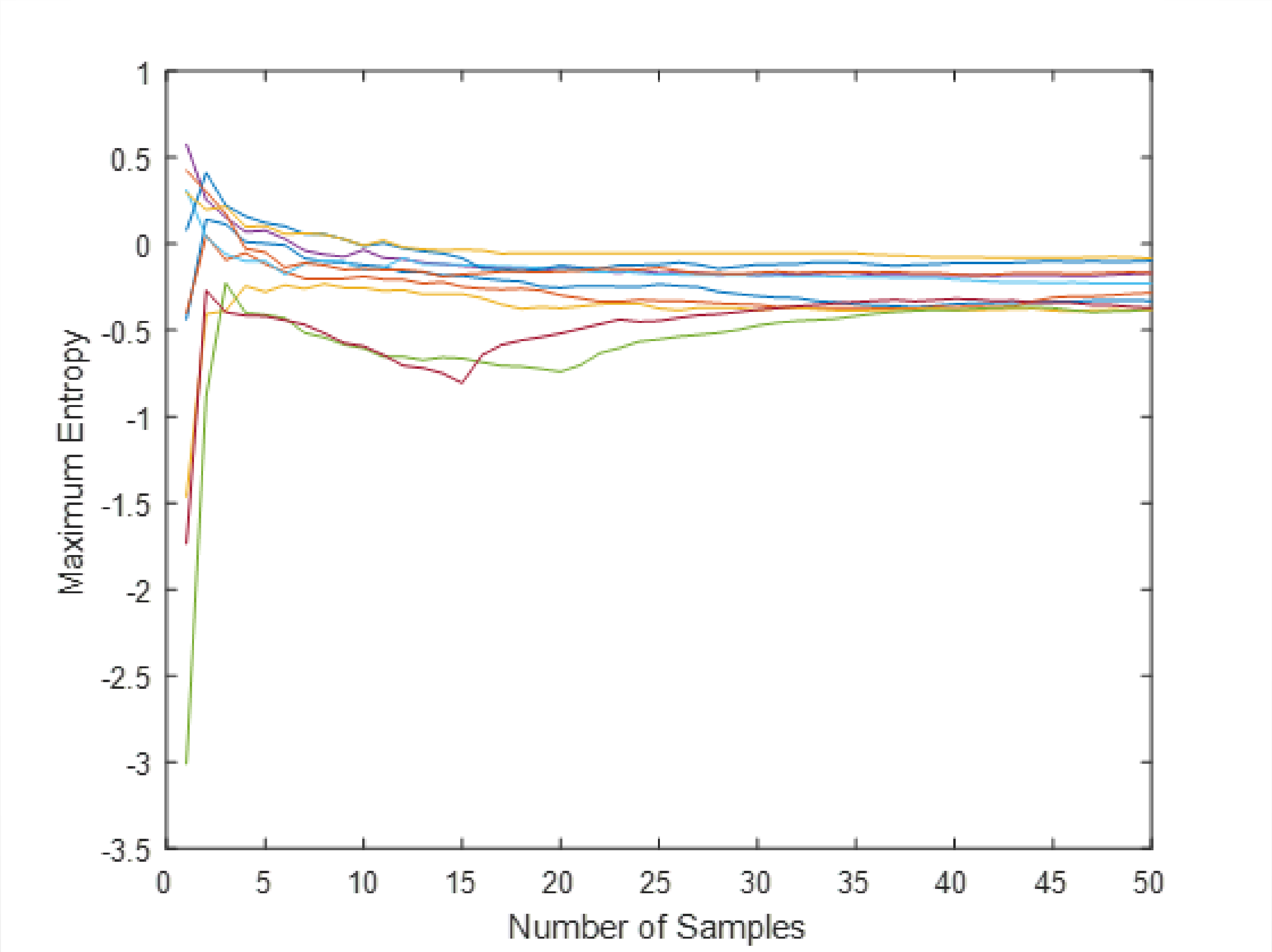
Window size +-20



Greedy Approach - window

Window size +-20

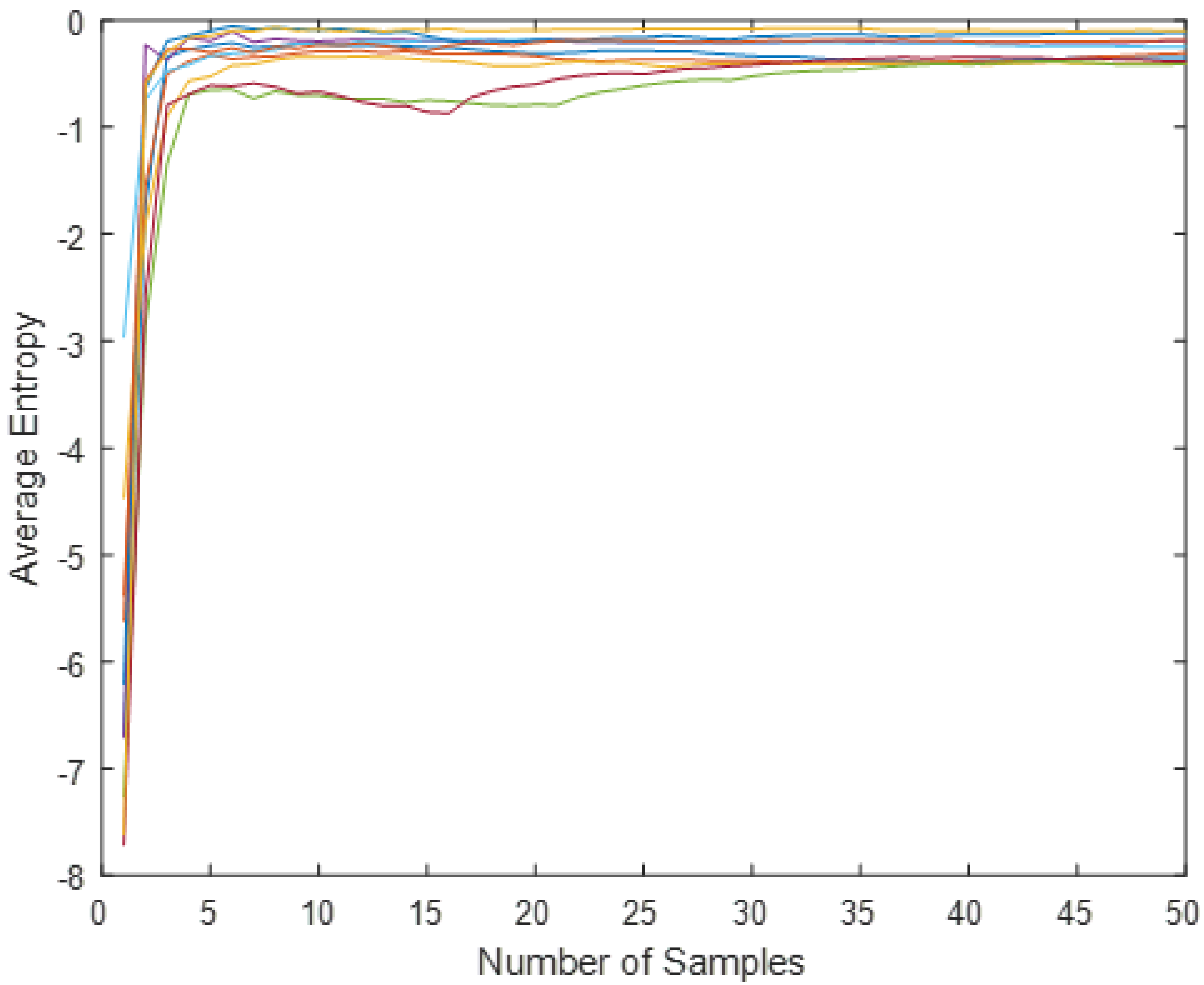
No. of Test Data = 10



Greedy Approach - window

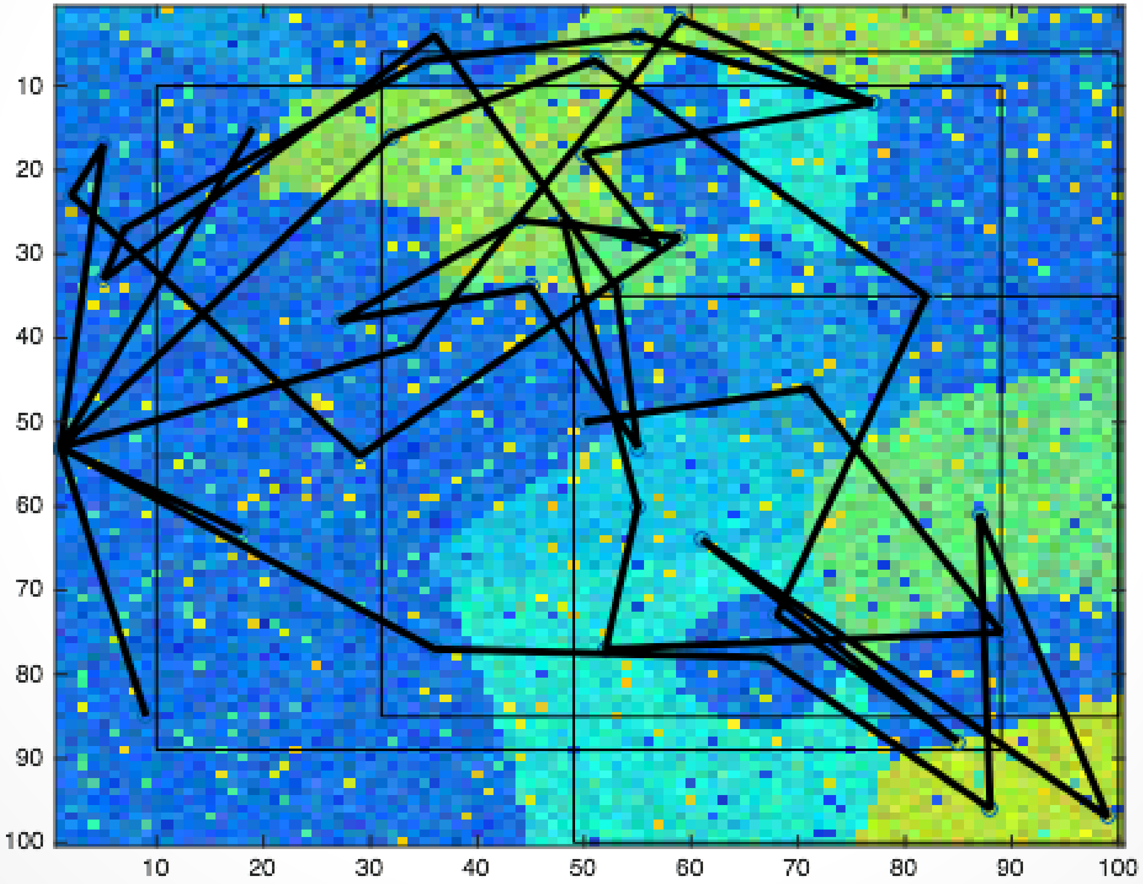
Window size +-20

No. of Test Data = 10



Greedy Approach - window

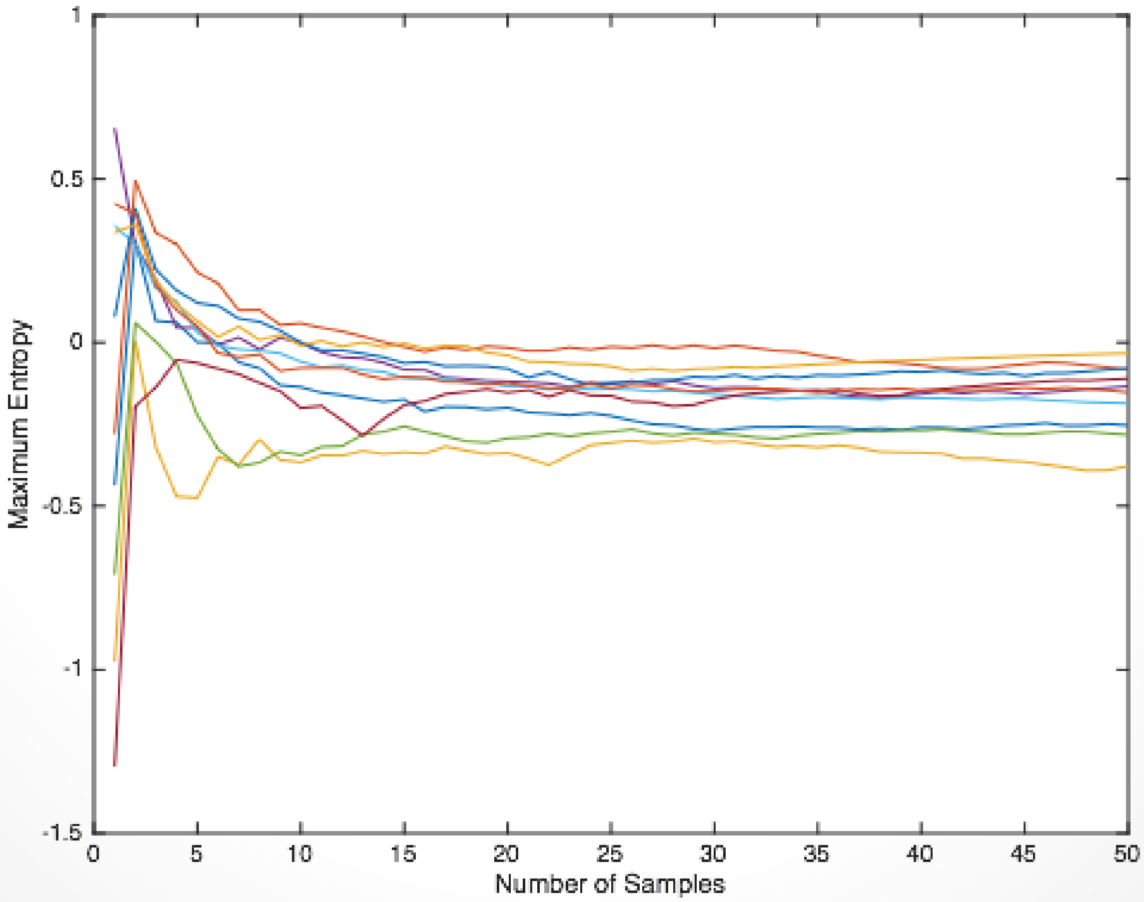
Window size +-40



Greedy Approach - window

Window size +-40

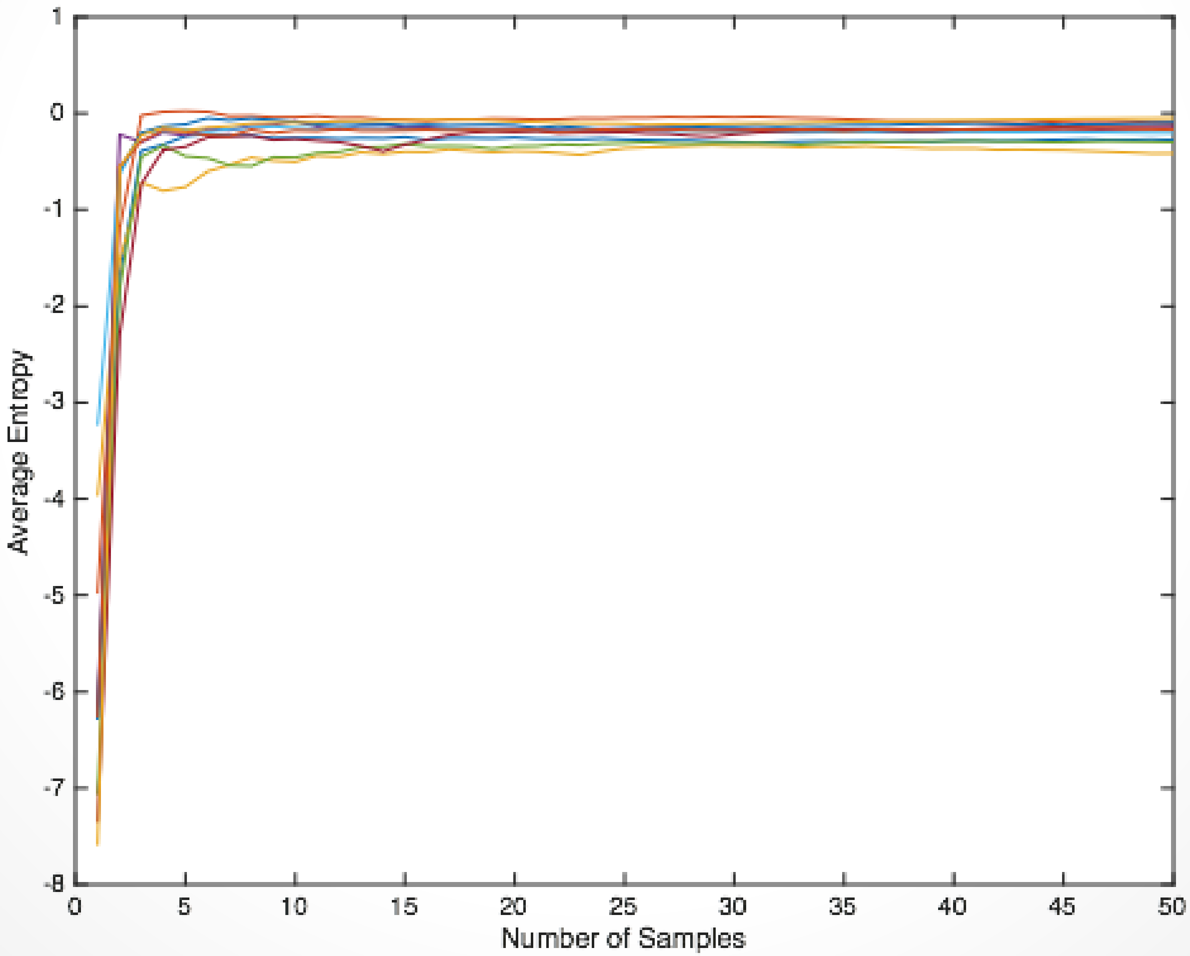
No. of Test Data = 10



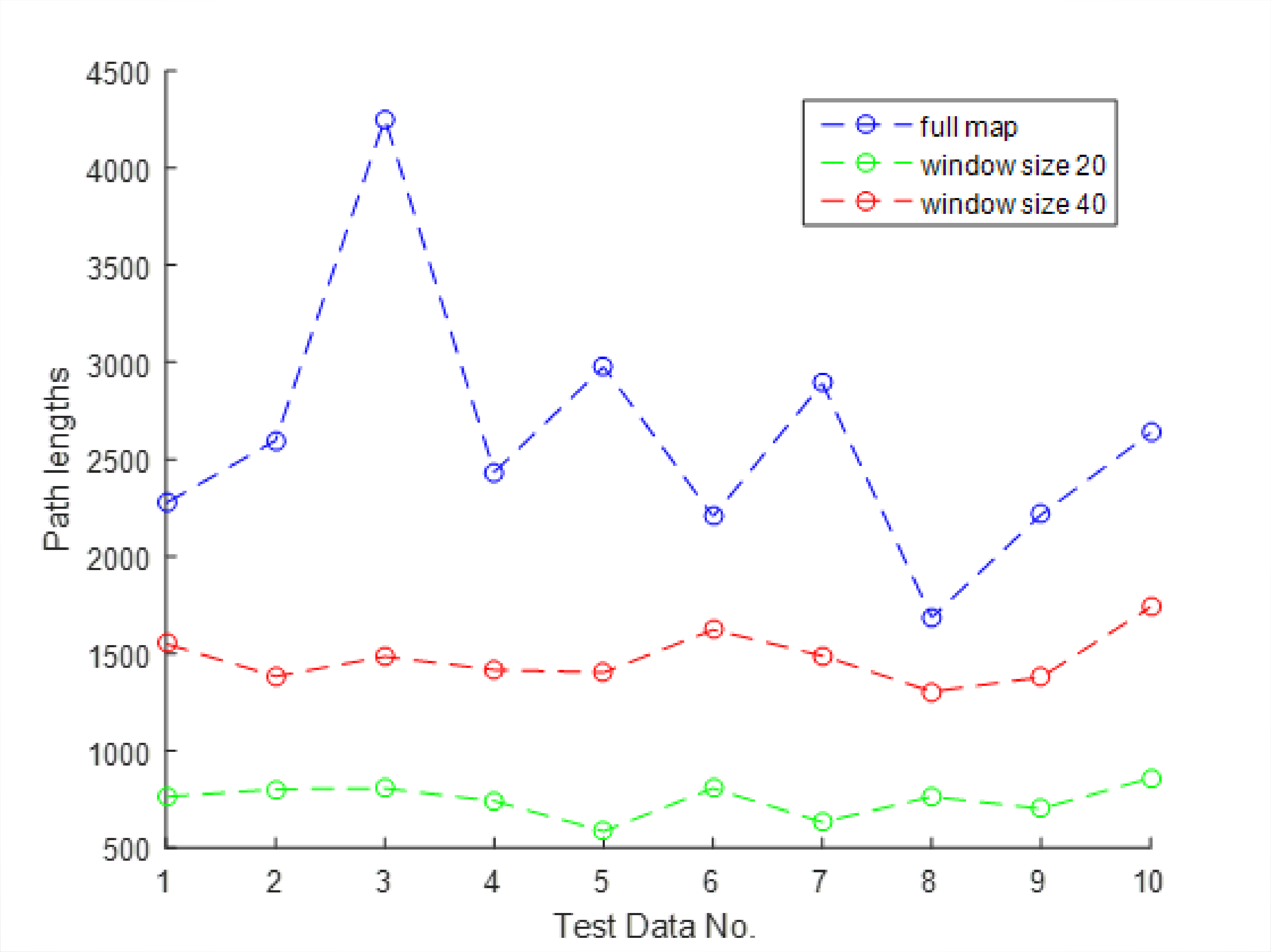
Greedy Approach - window

Window size +-40

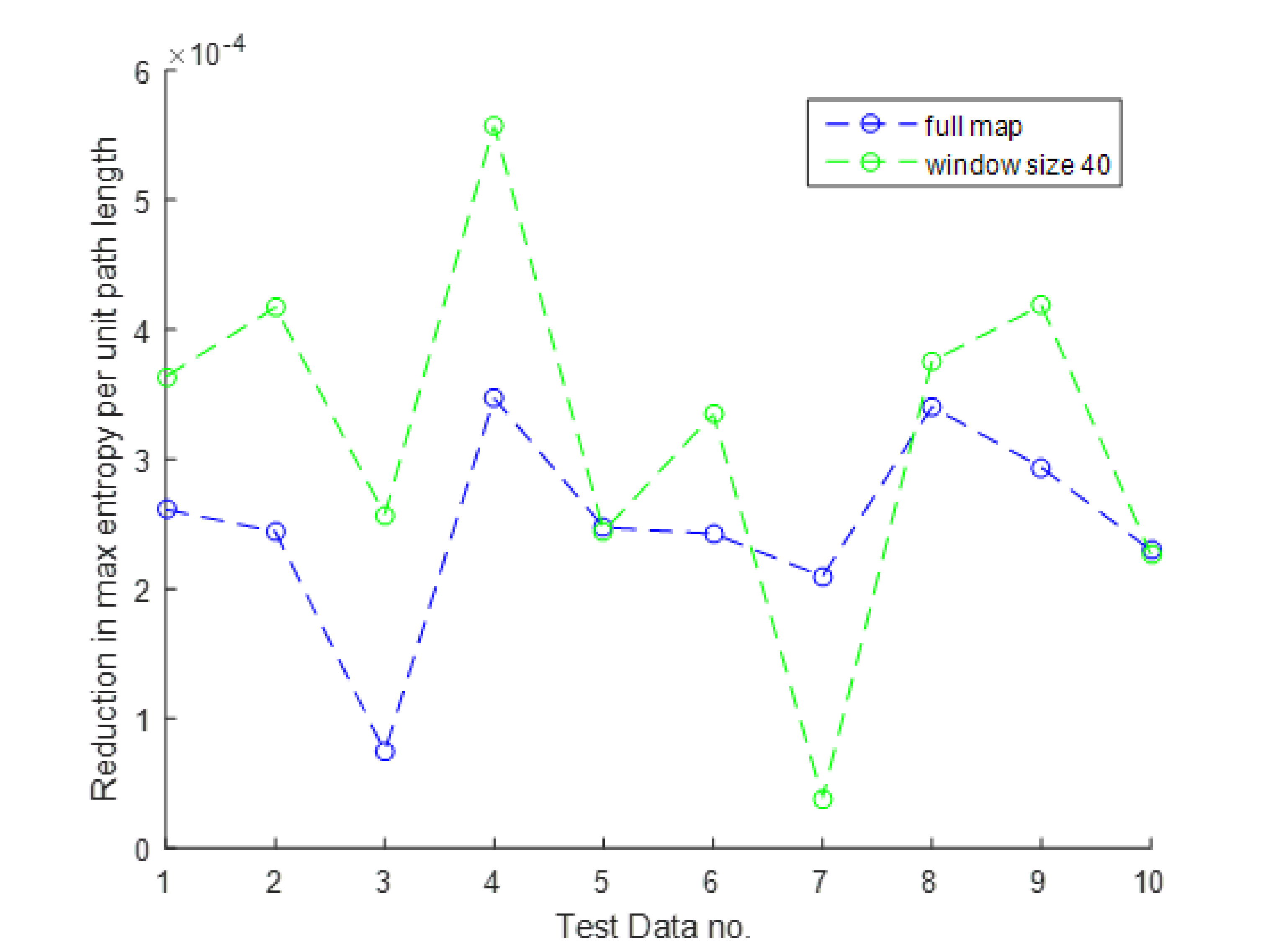
No. of Test Data = 10



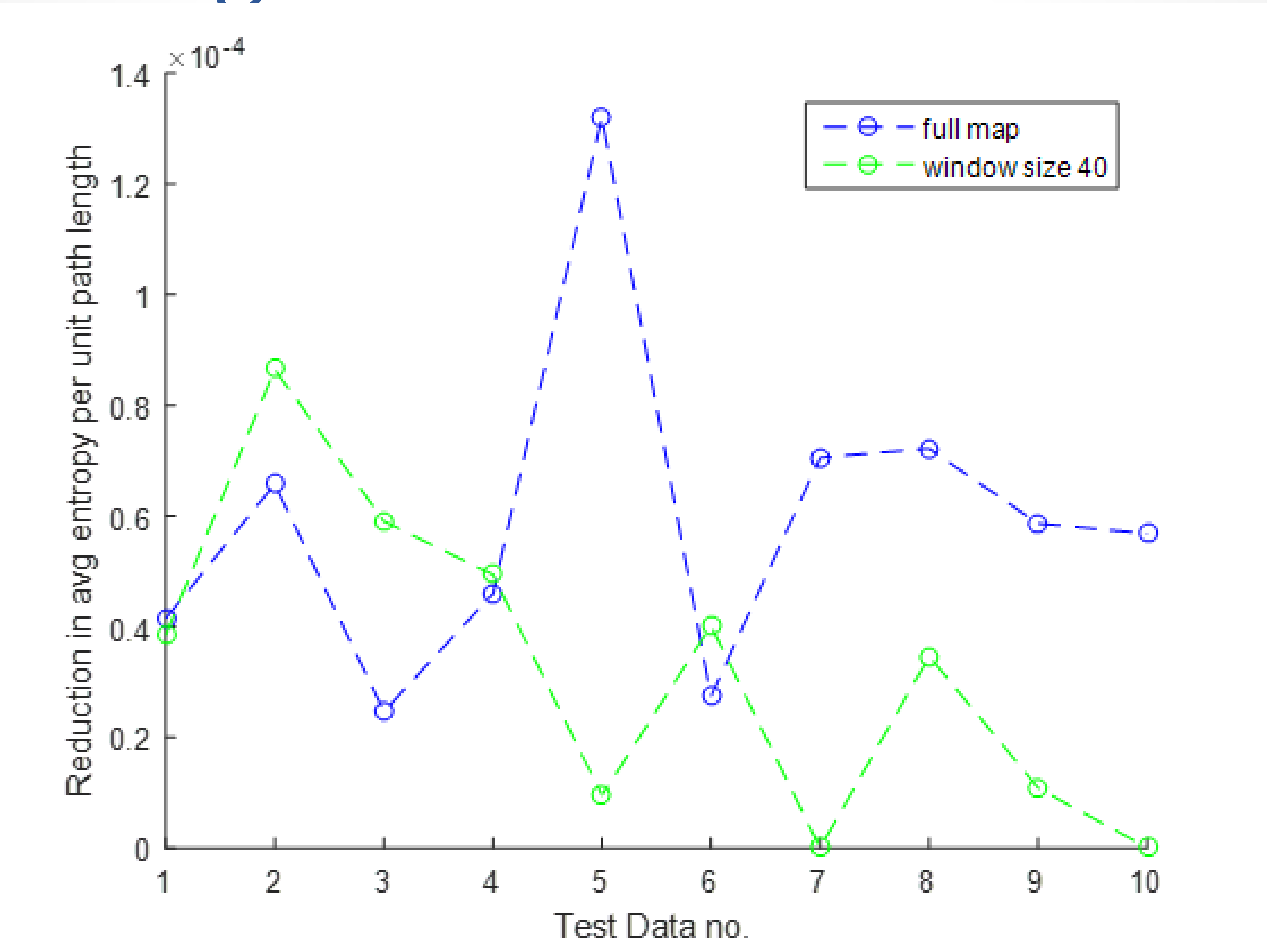
Paths lengths - Full vs window



Maximum Entropy reduction/path length - Full vs window



Average Entropy reduction/path length - Full vs window



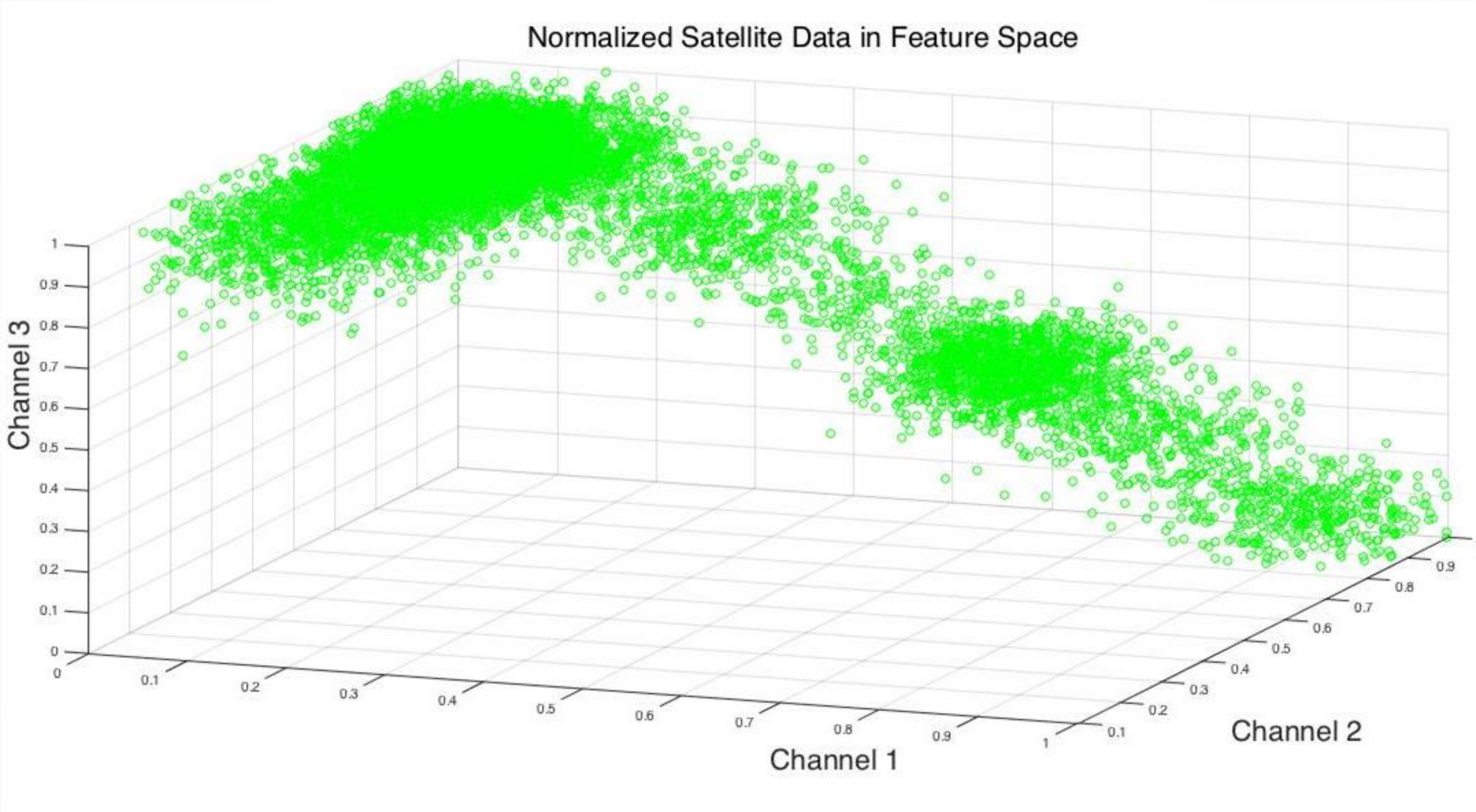
Problem 1: Where to sample

Approach 2: Feature Space based Entropy

General Flow

1. Satellite Data in feature space
2. Cluster
3. Probabilities and entropy calculation
4. Maximum entropy point sample
5. Replace sampled point with data and update

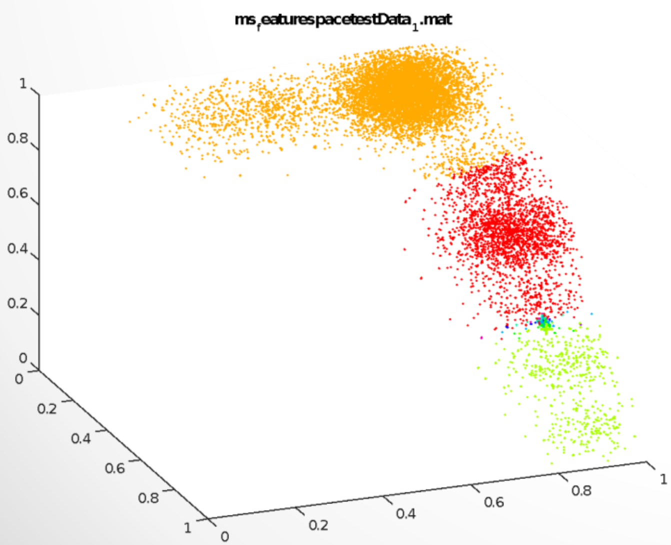
Data in Feature Space



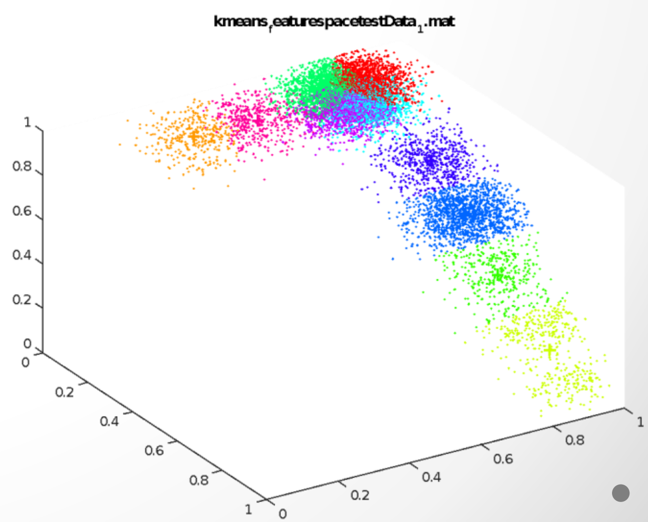
Clustering

- K-means
  - Parametric approach – assumes number of classes (not available)
- Mean-shift
  - Non-parametric approach – bandwidth keeps changing
    - Optimization parameter – people have done this...
  - might not capture the rare classes as they are sparse

Meanshift (0.2)



Kmeans



Probability and Entropy

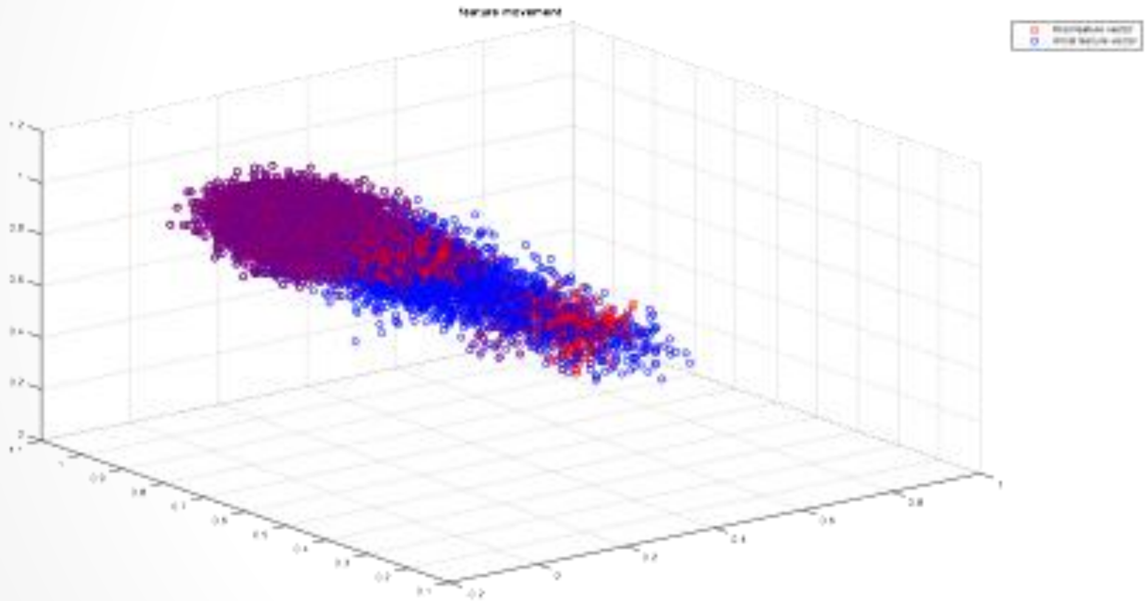
$$P_{ij} = \frac{1/d_{ij}}{\sum_{j=1}^{J} 1/d_{ij}}$$

i = feature point
j = cluster
J = total number of clusters

$$H(i) = \sum_{j=1}^{J} P_{ij} log_2(\frac{1}{P_{ij}})$$

Shannon's Entropy for each feature point
Units – bits (log base 2)

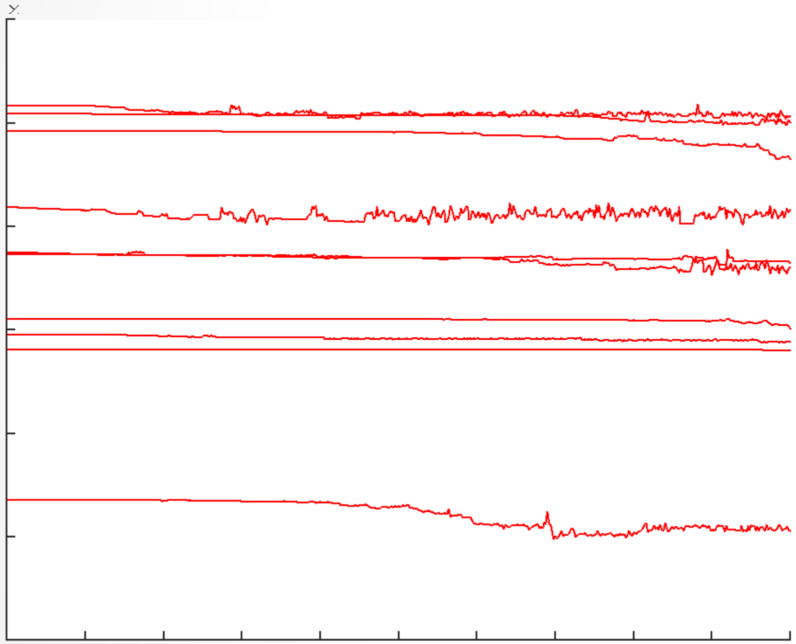
Re-sampling



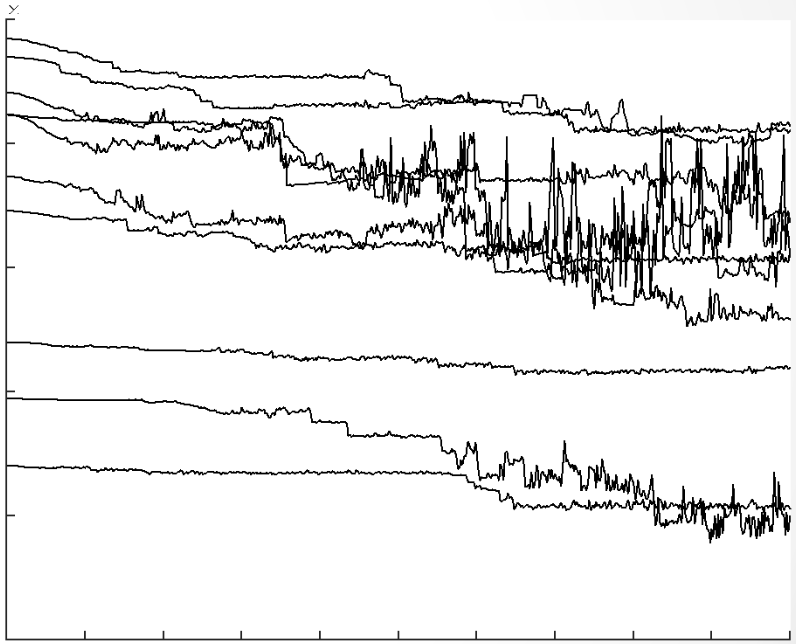
- Gaussian Kernel to move points around the sampled point

Results

2.96% decrease



1.25% decrease



Overall average entropy decrease in both methods..
Multiple parameters left to optimize!

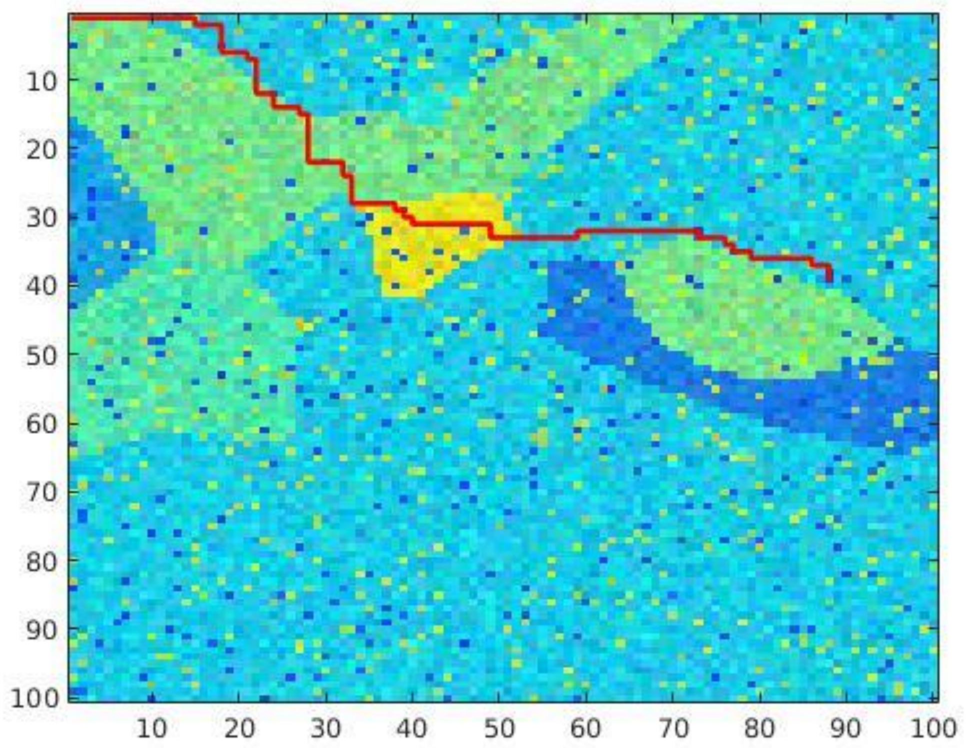
Problem 2: Planning Path to Sampling Point

Approach 1: DP path length constraint planning

DP approach

- Currently:
  - Budget based planning
  - Scientist guided
  - Heuristics to optimize
- Extreme case:
  - DP approach to find path with Maximum information gain, given a budget
  - Sets a reference for best solution for the budget
  - Cannot be used Dynamically -: Computation cost is very high

Paths Generated



Problem 2: Planning Path to Sampling Point

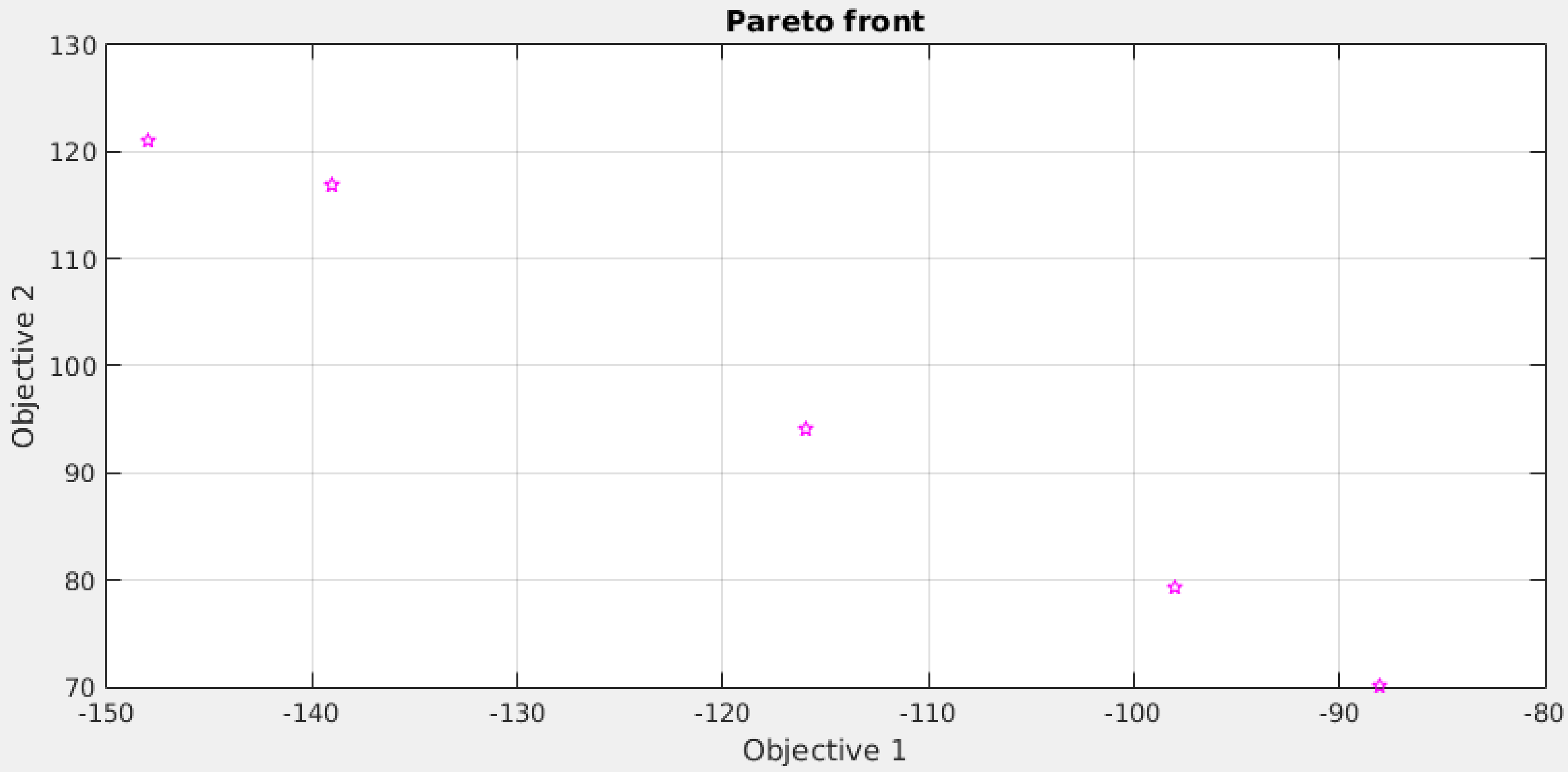
Approach 2: Multi-Heuristic A*

Multi-Heuristic A*

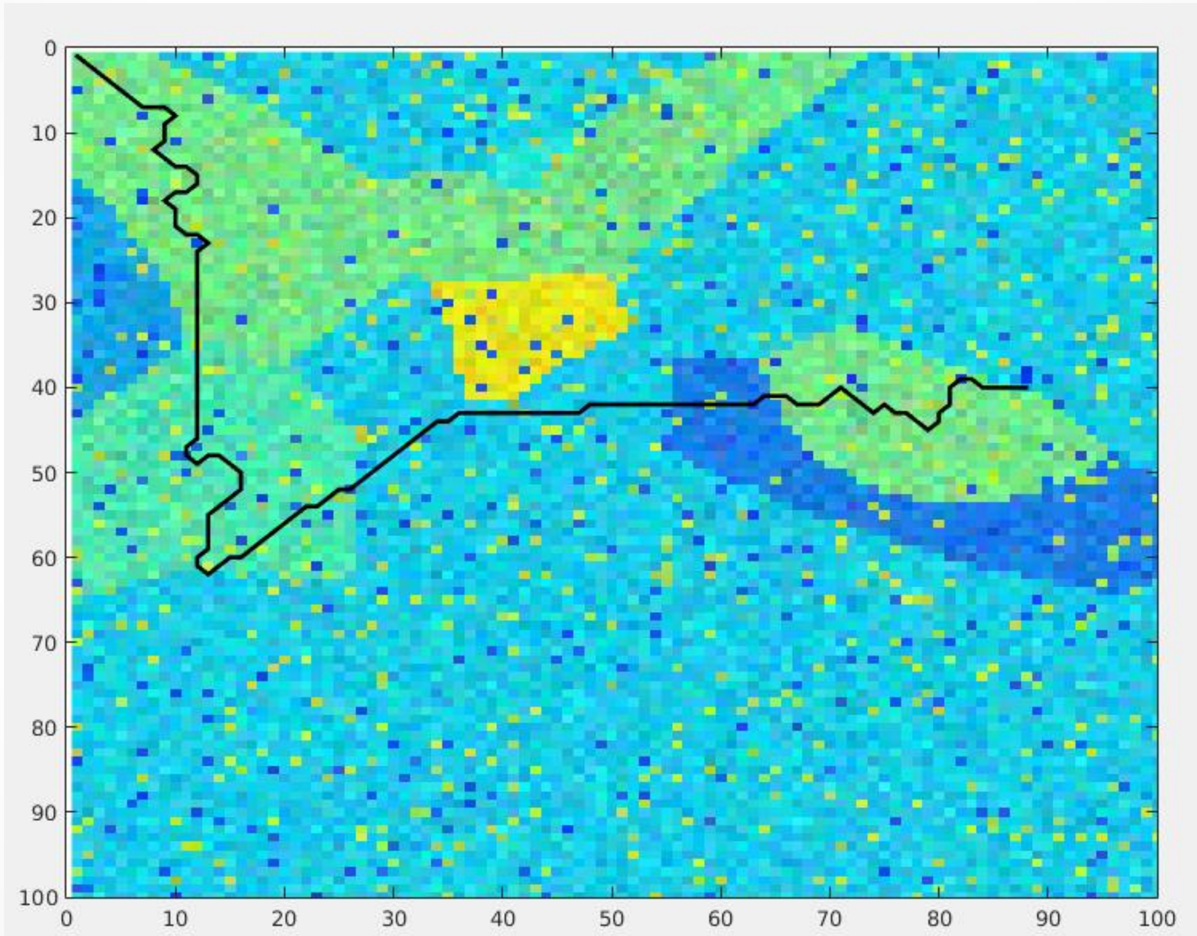
- Multiple Heuristics in Planning
  - Given a 'Global' Objective, Maximise Information Collected Along the Way
  - Distance as 'Anchor' heuristic
  - Information as an additional Heuristic

- Multi-Objective Optimisation
  - Maximise Information Gain
  - Minimize Path Length

Pareto Optimisation Curve

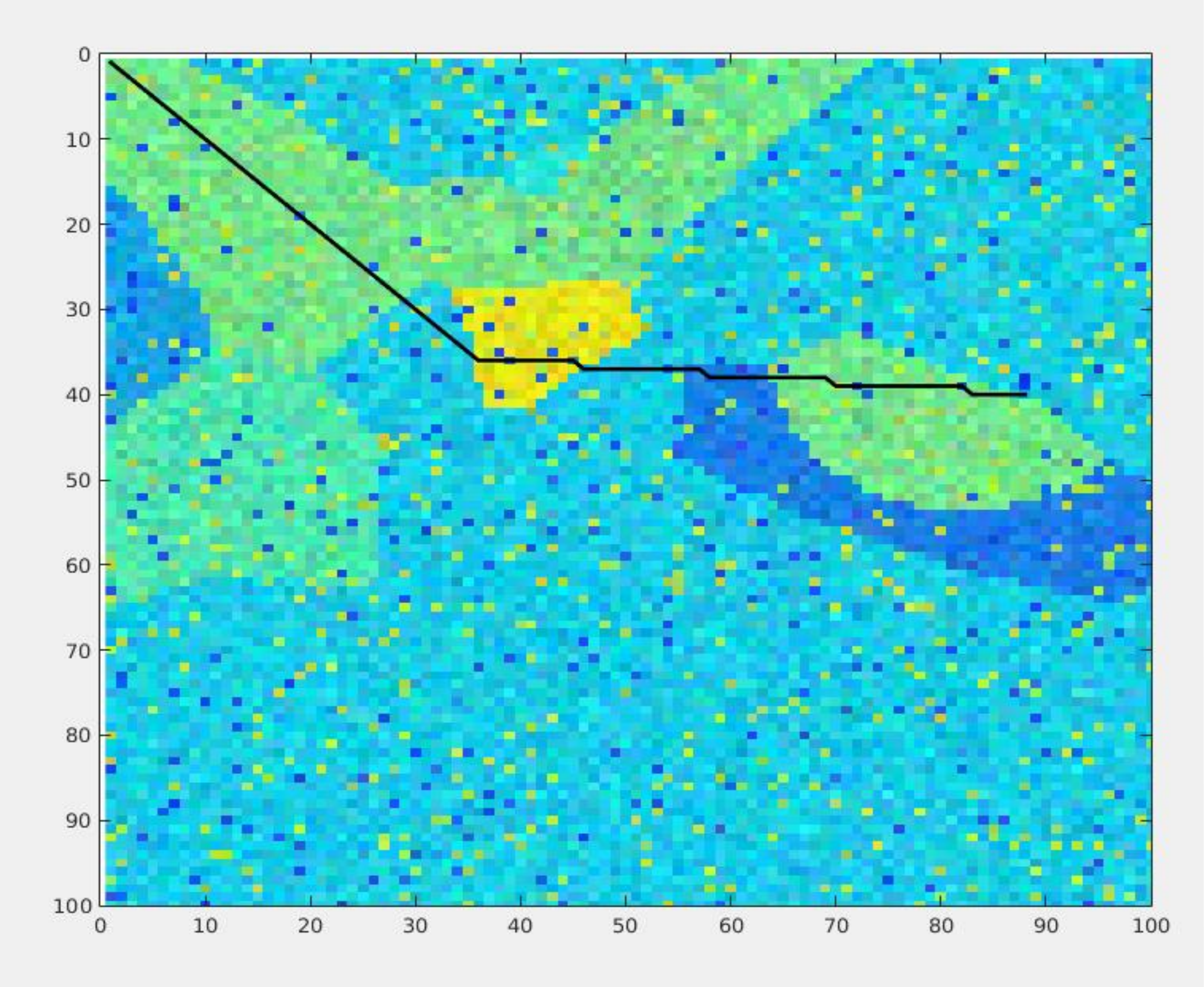


Paths Generated



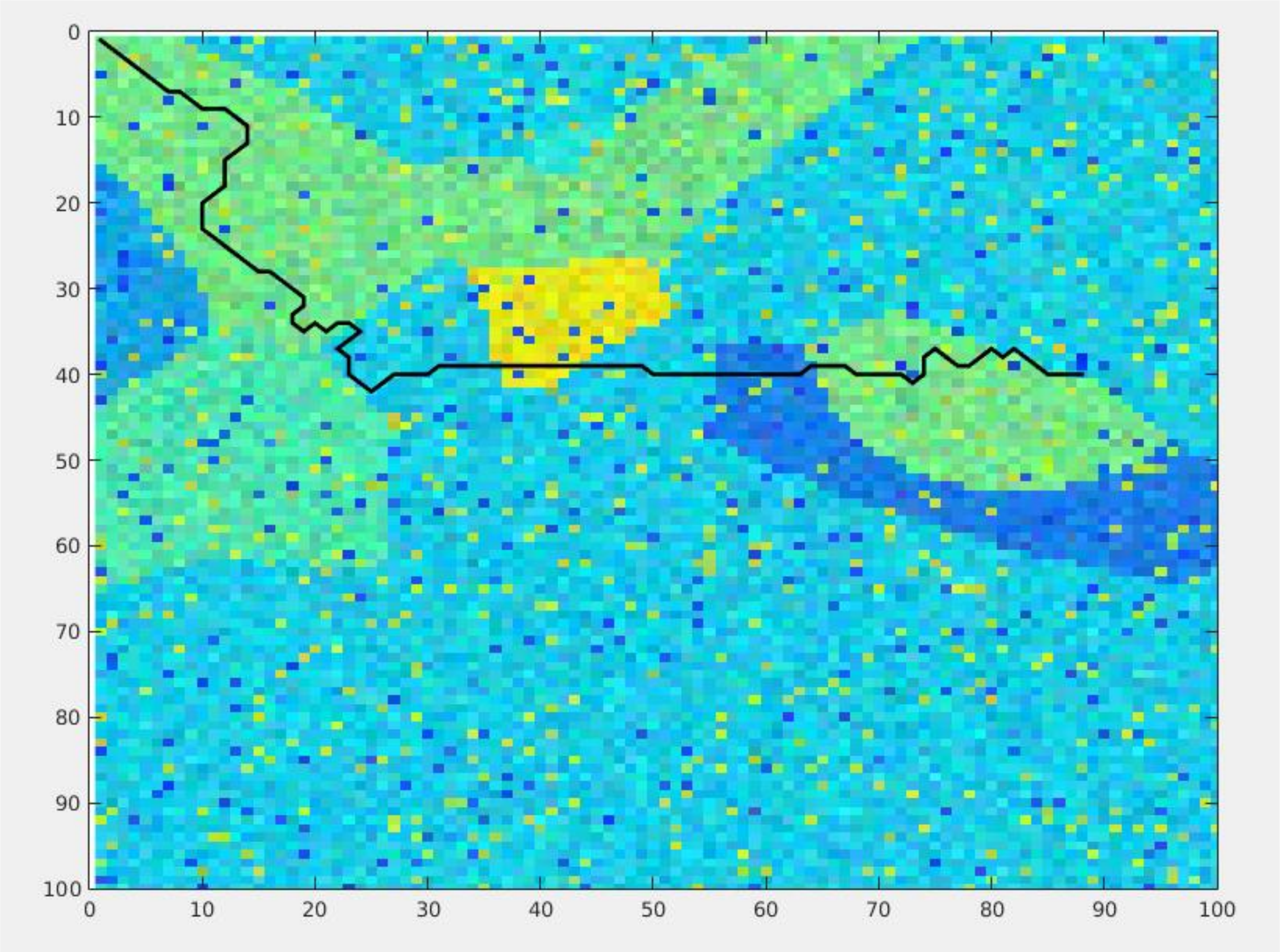
Weight of Information >> Weight of Distance

Paths Generated



Weight of Information << Weight of Distance

Paths Generated



Optimal Weights for Distance and Information

Statistical Comparison of Planning

Sno.	Technique	Percent Entropy Reduction	Entropy Reduction Per Unit Distance Travelled	Percent Entropy Reduction per Unit Distance Travelled	Runtime
1	Dynamic Programming	69.75%	211.66	0.5449%	5 hours
2	MHA*	46.875%	142.81	0.3982%	0.1s - 4s

Conclusion

- State of the art entropy calculation methods implemented
- Novel implementation of clustering based entropy done
- Planning in Euclidean space and Information space
- Multi Heuristic A*
- Dynamic Programming

Thank You!