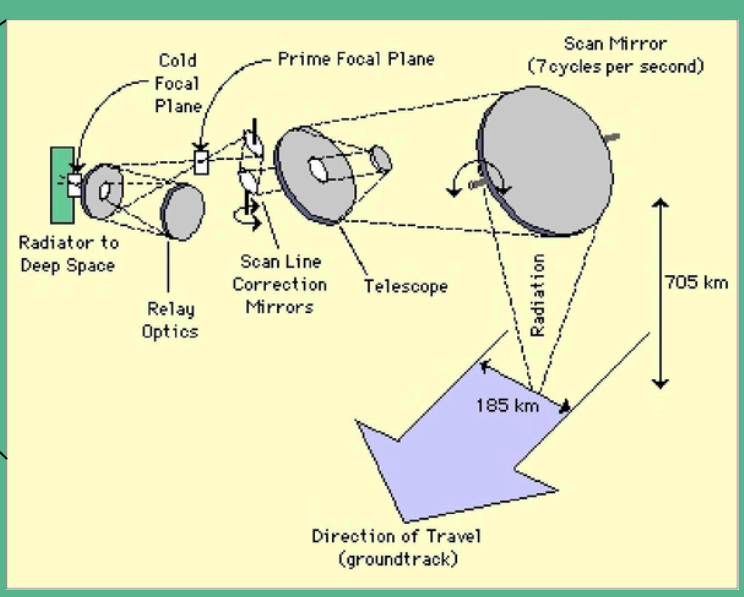
Submarine Groundwater Discharge using Landsat 7 and K-Means Clustering

Landsat 7 Data

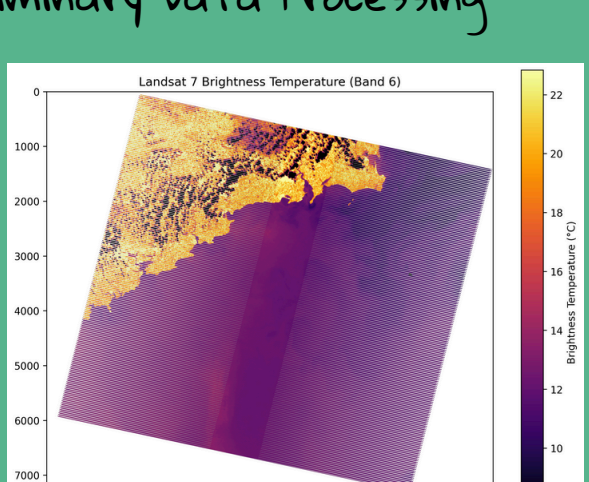
Landsat-7 Satelite



ETM+ Instrument

Preliminary Data Processing

Convert radiance to temperatures



Publicly Available Product

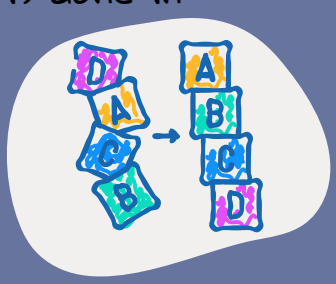
8 Bands

uses: surface temperature, soil moisture, vegetation identification, mineral and rock discrimination

ckmeans-1d-dp Implementation

1. Sort SST data

- for a 1-D implementation such as this, sorting is crucial. It is done in ascending order



2. Compute SSE for every interval

- Calculated for each interval [i, j]
- Cost of placing i and j in the same cluster

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

3. Build Dynamic Programming (dp) Table

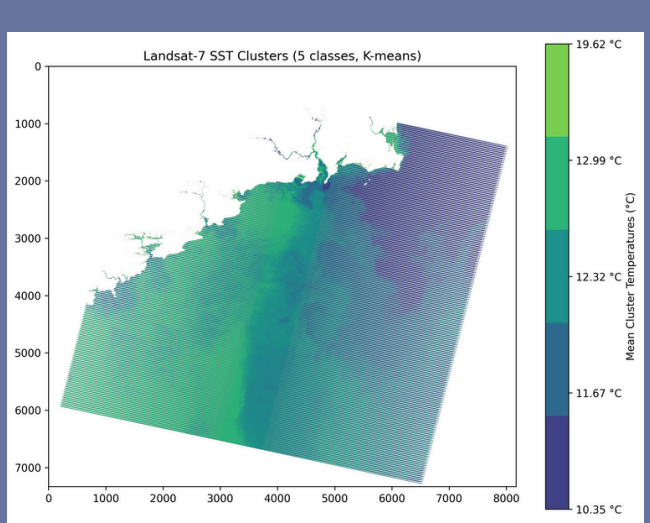
- Build a 2D table D[k][i] that stores the SSE of dividing the first i elements into k clusters
- Another table T[k][i] stores the index where the last cluster begins

4. Populate Tables Using Recurrence

- Use dynamic programming to fill in D and T by trying all possible partition points
- Allows for a globally optimal solution (unlike regular k-means which is heuristic and can get stuck in a local minimum).

5. Recover Clusters

- Use the T table to go back and determine where each cluster starts and ends, yielding the final cluster assignments.



Sources:
https://landsat.gsfc.nasa.gov/article/landsat-7/
https://pmc.ncbi.nlm.nih.gov/articles/PMC5148156/pdf/nihms833138.pdf
https://www.eoportal.org/satellite-missions/landsat-7#etm-enhanced-thematic-mapper-plus