Modeling Workgroup Meeting Quarterly Review

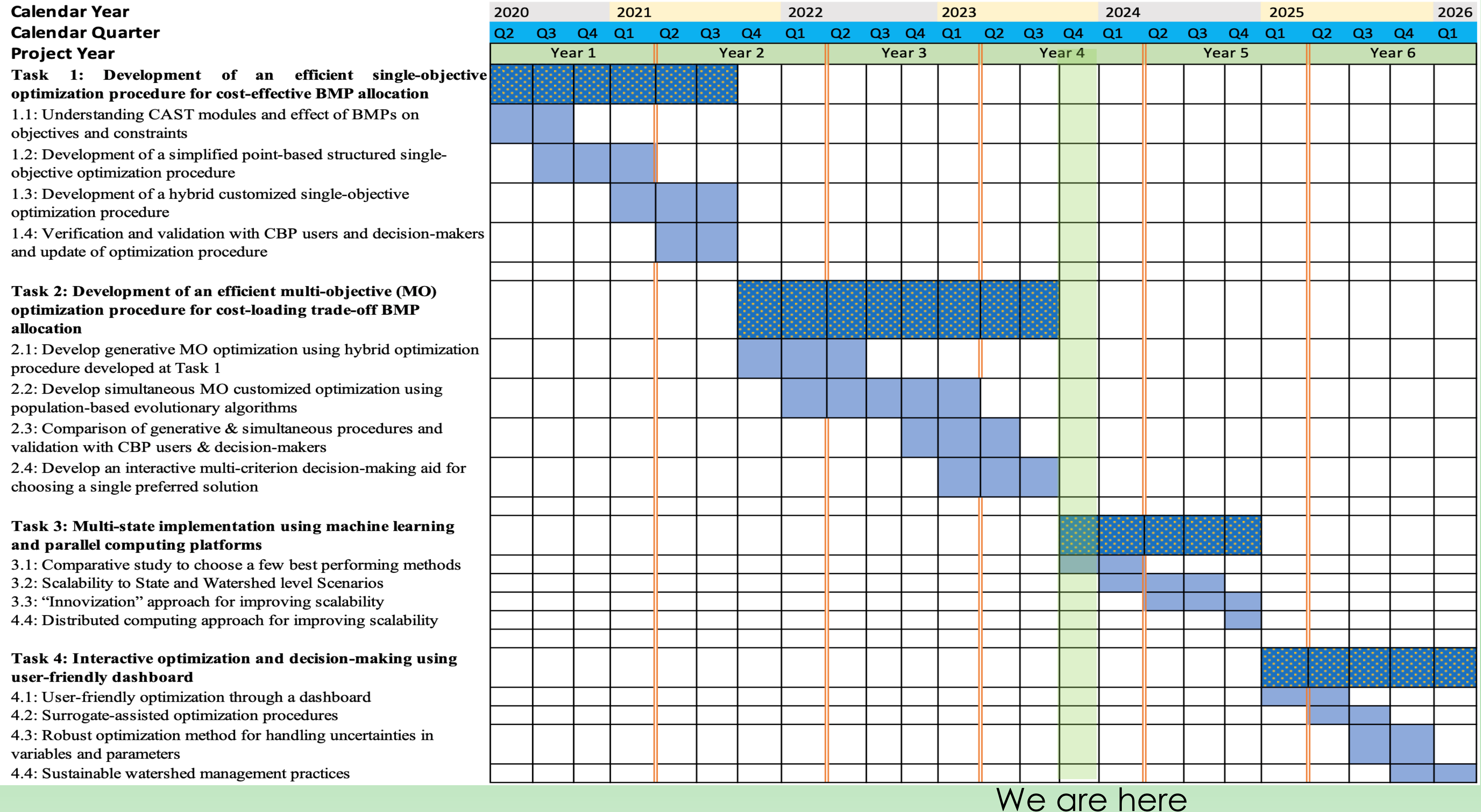
January 2024

Optimization Update

Pouyan Nejadhashemi, Kalyanmoy Deb, Gregorio Toscano, and Hoda Razavi

MICHIGAN STATE UNIVERSITY

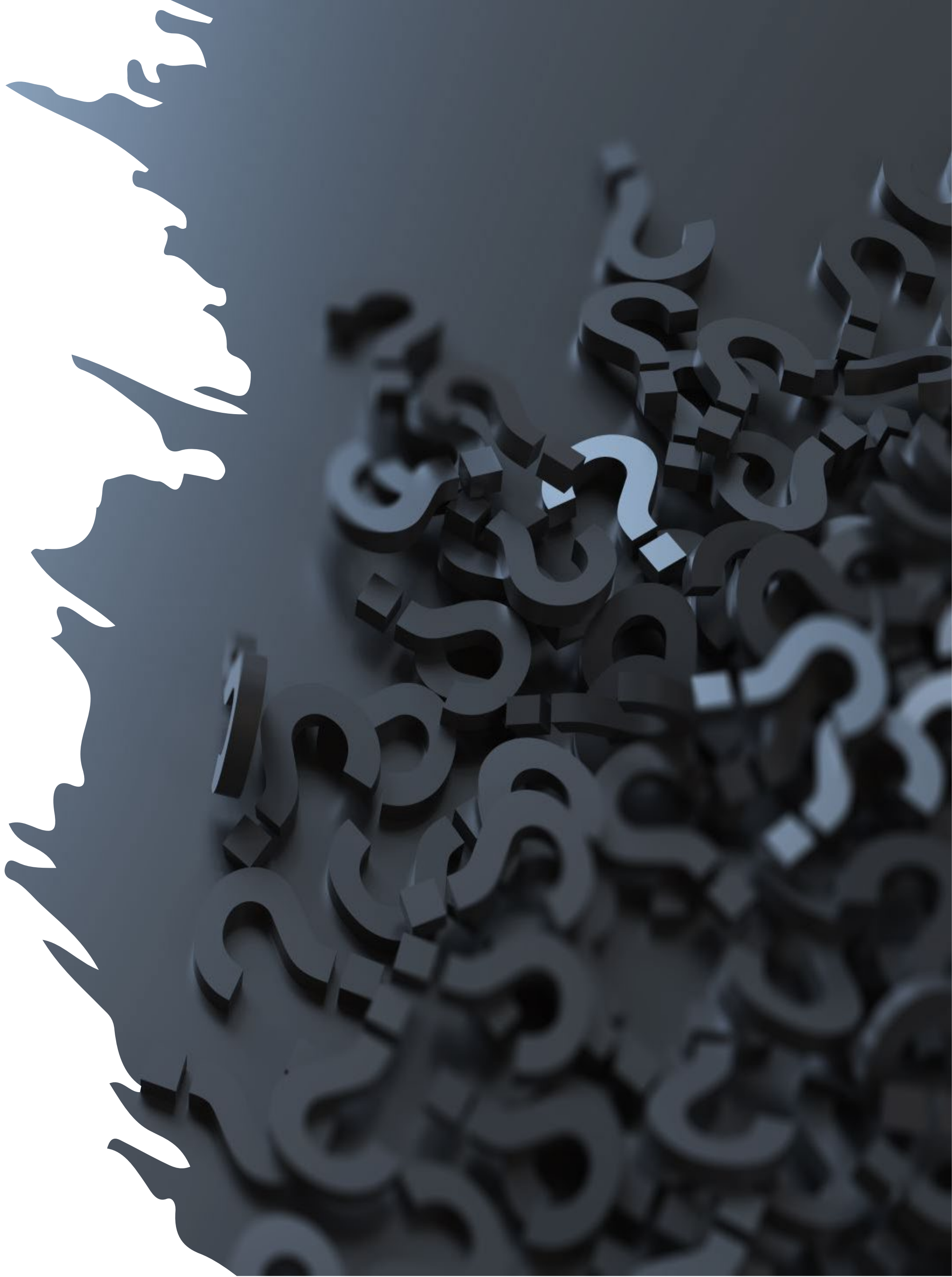
Timeline of the Project



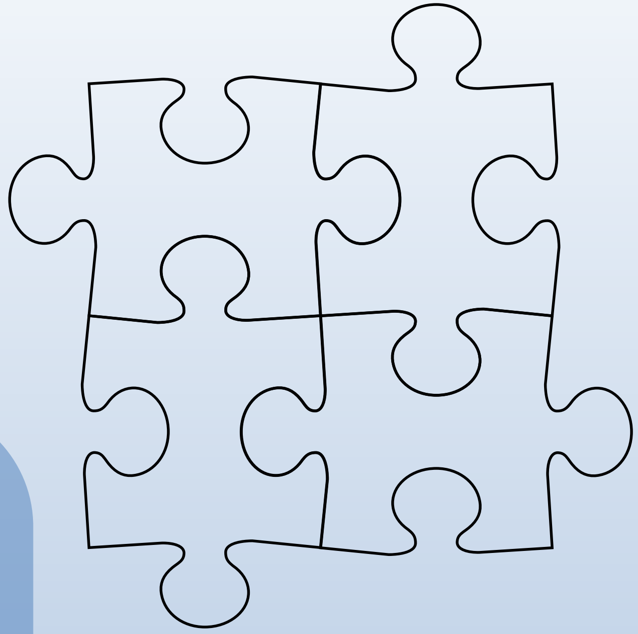

OVERViEW

Problem Statement

One of the most challenging issues for a real-world optimization problem is the ***Search Space Complexity*** (variables and constraints), especially for the CBW management problem.



Beating the "Curse of Dimensionality"



- Due to the **large dimension of the problem**, there is **no off-the-shelf optimization algorithm** capable of handling this problem.

- Therefore, we developed a **customized optimization** approach **to speed up computational time and reduce the size of optimization variables** to make the problem solvable in a reasonable time frame.

Innovation through Optimization

The major problem we are facing is **the large number of optimization variables.** These variables originate from three major components:

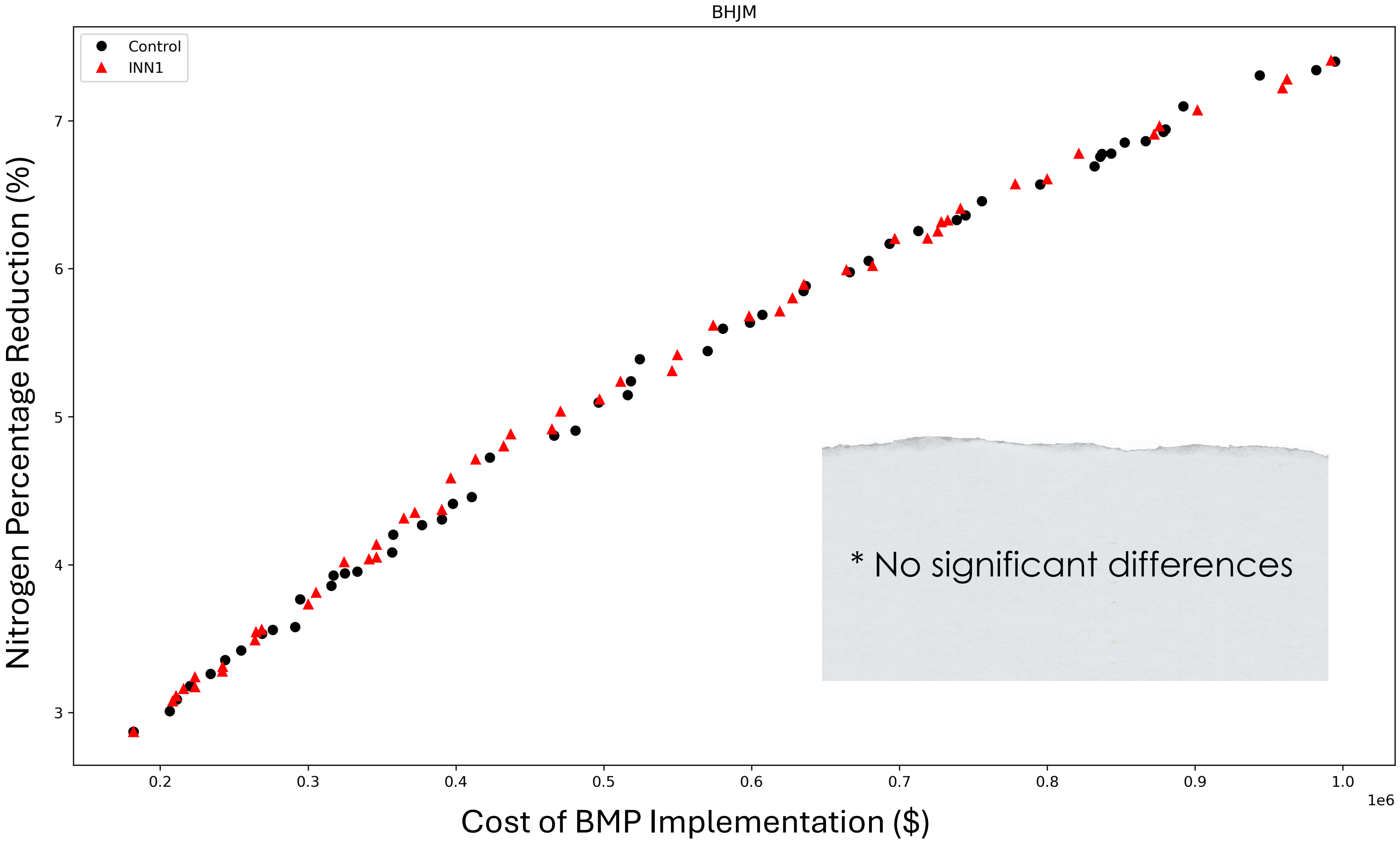
- Type of BMPs
- BMP implementation location
- Size of BMPs

Innovization

By understanding **the common characteristics of the group of BMPs or locations of implementation**, we can reduce the number of BMPs and ultimately **reduce the total number of variables**.

Compare Original vs. Innovized Optimization

Berkeley, Hardy, Jefferson, and Mineral Counties



* No significant differences

Innovization- for efficiency BMPs-

By reducing the number of BMP from about **200 types of efficient BMPs to only 10 BMPs,** the size of the problem was significantly reduced.

The innovizaed approach can reduce the number of variables by **about 96%,** which is a promising result.

Goal

- Our goal is to develop a more comprehensive approach that encompasses the entirety of the Chesapeake Bay watershed.

Step 2: Variable Reduction

VARIABLE REDUCTION TECHNIQUE CLUSTERING

How do you simplify your data analysis?

Various Techniques for Variable Reduction

1. Principal Component Analysis (PCA): PCA is a statistical technique that transforms the original variables into a new set of uncorrelated variables ordered so that the first few retain most of the variation present in the original variables.

2. Bayesian variable reduction:

Bayesian variable reduction is a method that leverages Bayesian statistics to pinpoint the most relevant variables, incorporating prior knowledge and data evidence. It is useful for managing large variable sets and model uncertainty.

3. LASSO (Least Absolute Shrinkage and Selection Operator)

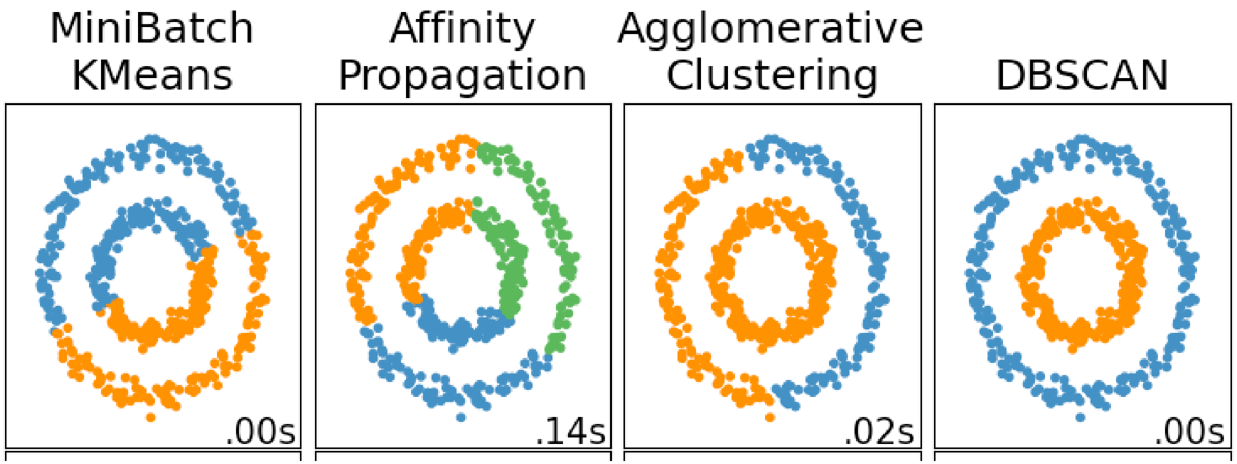
LASSO is a method that combines variable selection and regularization to improve model accuracy and simplicity by penalizing the absolute size of coefficients, reducing some to zero, ideal for datasets with more variables than observations.

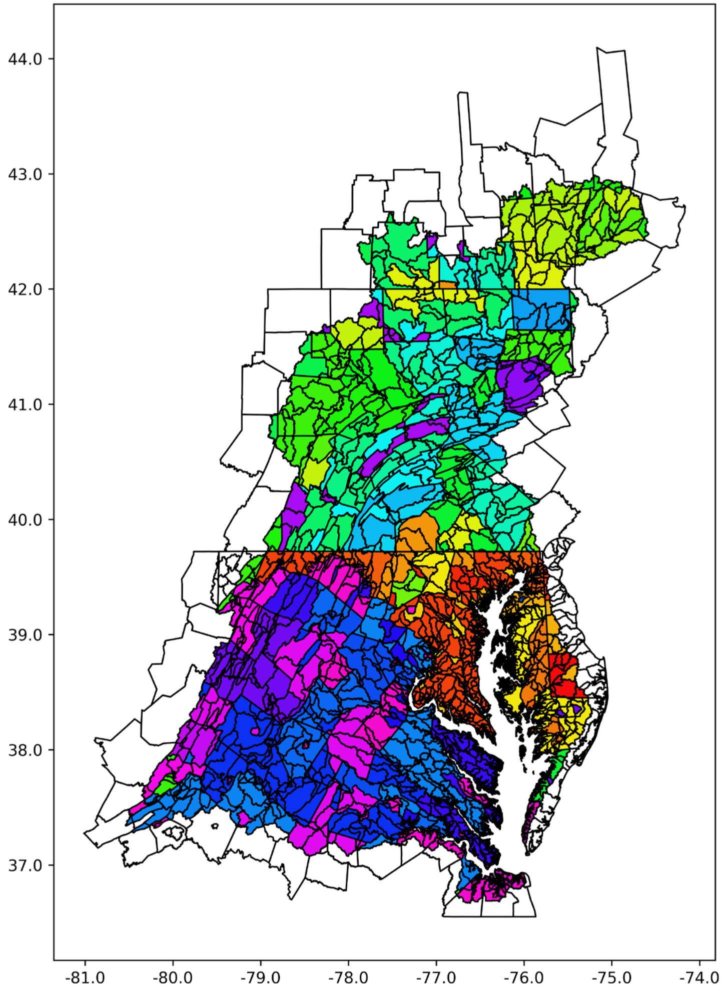
4. Ridge Regression

Ridge regression is a technique for analyzing multiple regression data that suffer from multicollinearity. Multicollinearity occurs when predictor variables in a model are highly correlated and can lead to unstable estimates of the regression coefficients. Ridge regression stabilizes these coefficients.

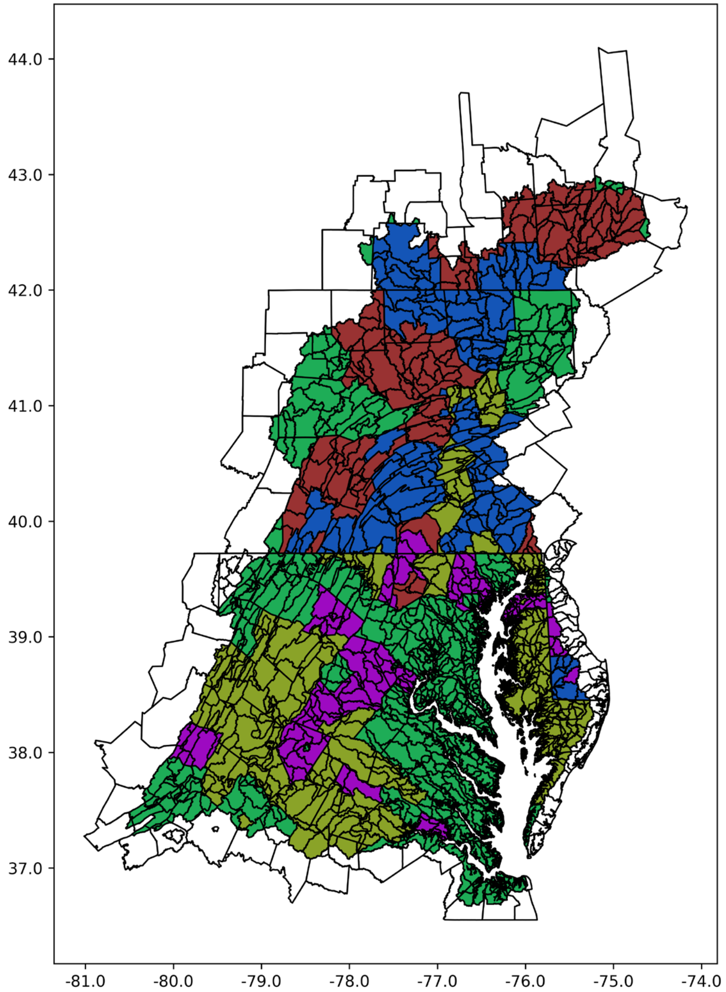
Step 3: Clustering

Clustering is an unsupervised/supervised learning method in machine learning and data analysis for grouping similar objects into clusters, widely used for statistical analysis across various fields.

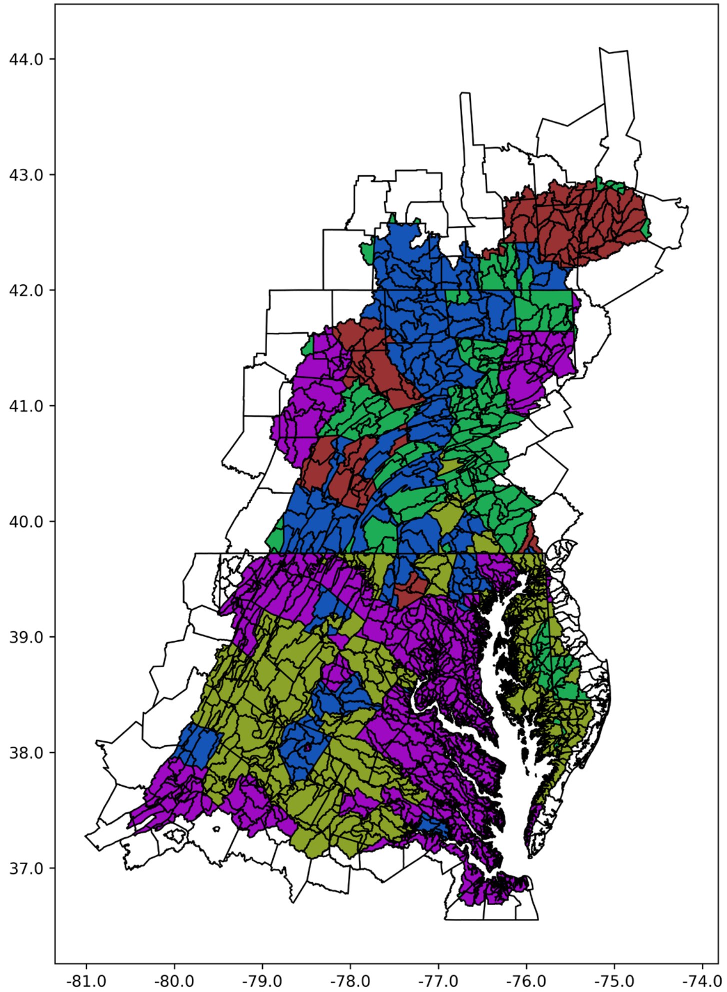




Affinity Propagation



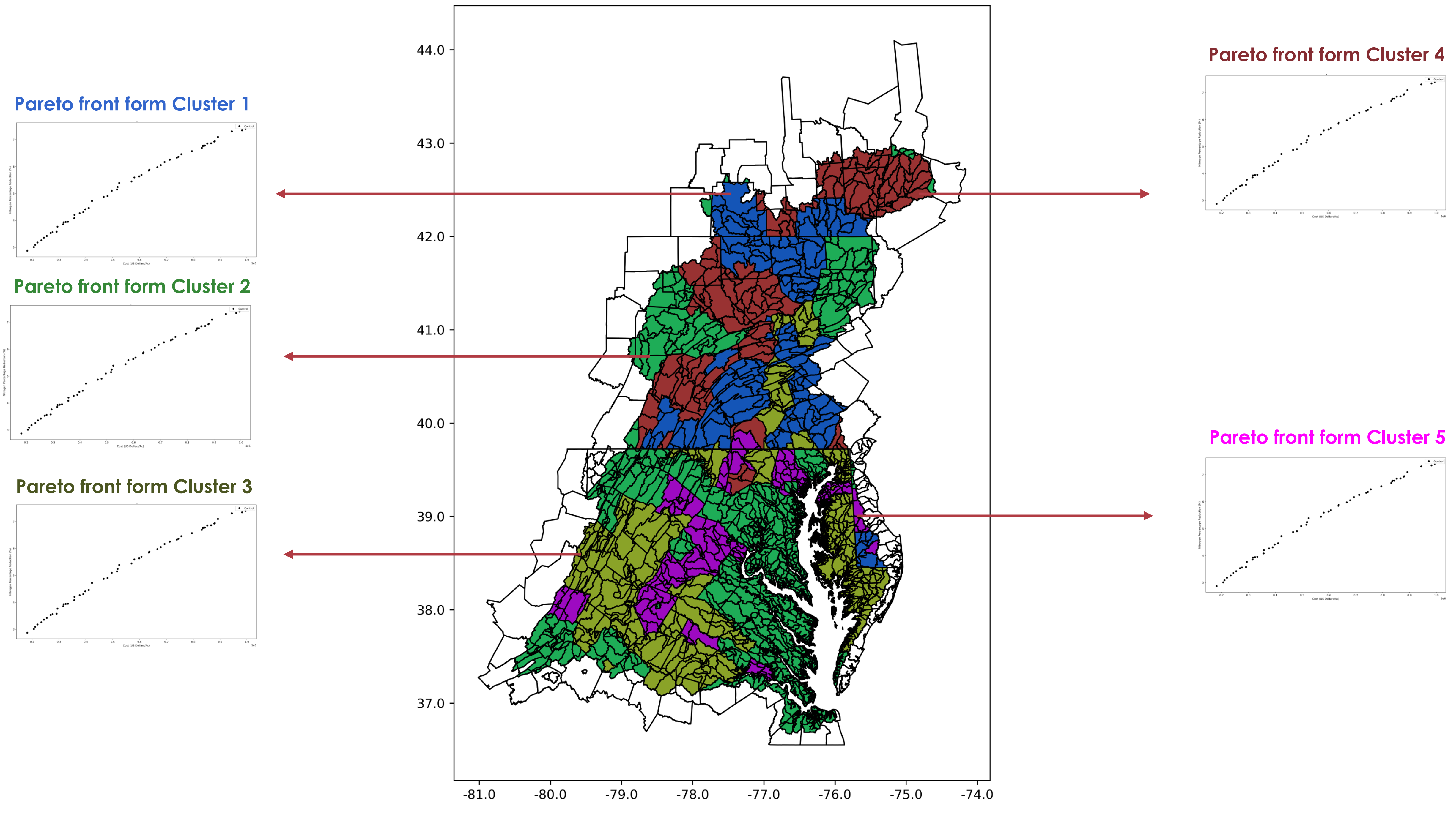
Agglomerative



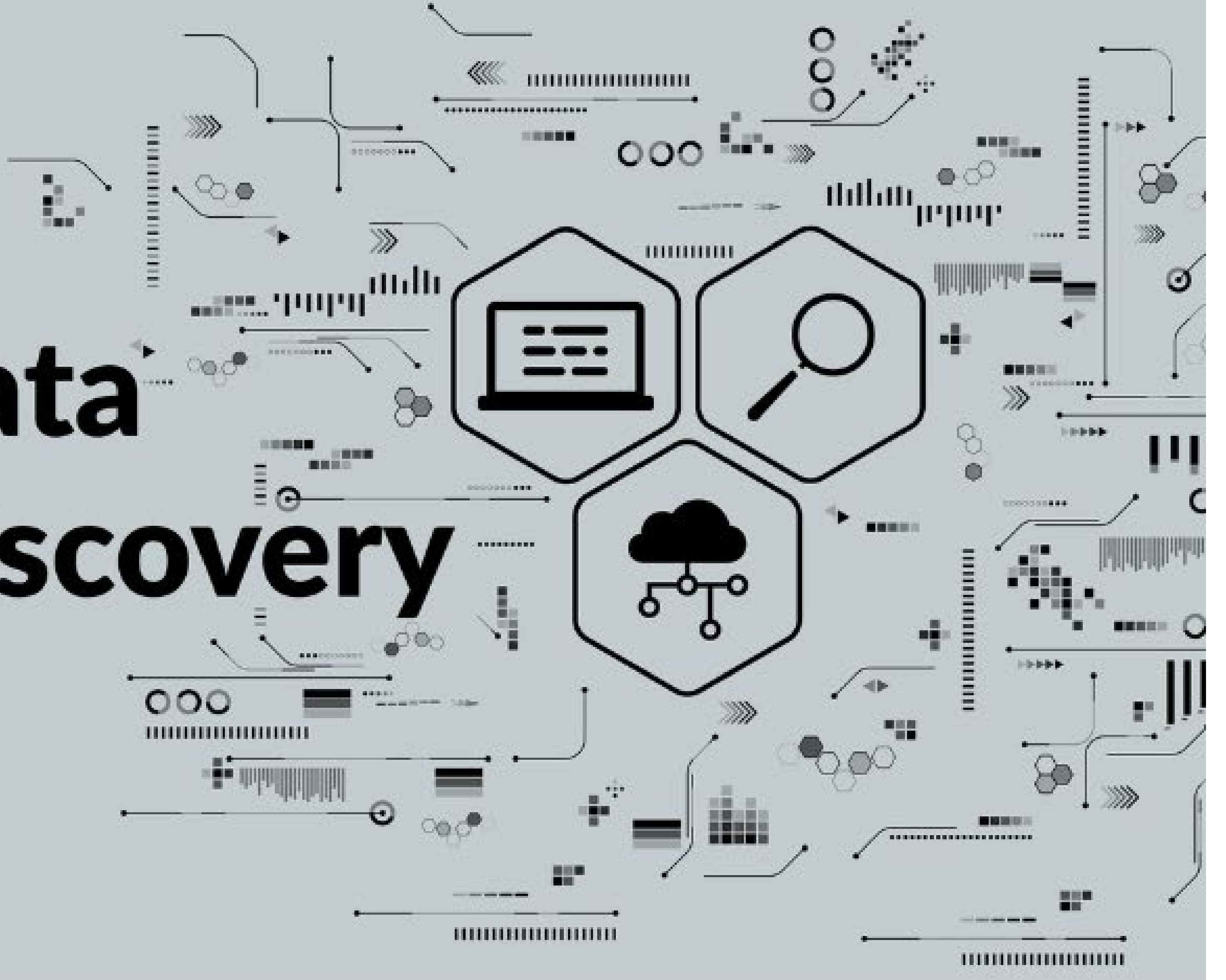
K-means

Clustering and Variable Reduction

Step 4: Optimization

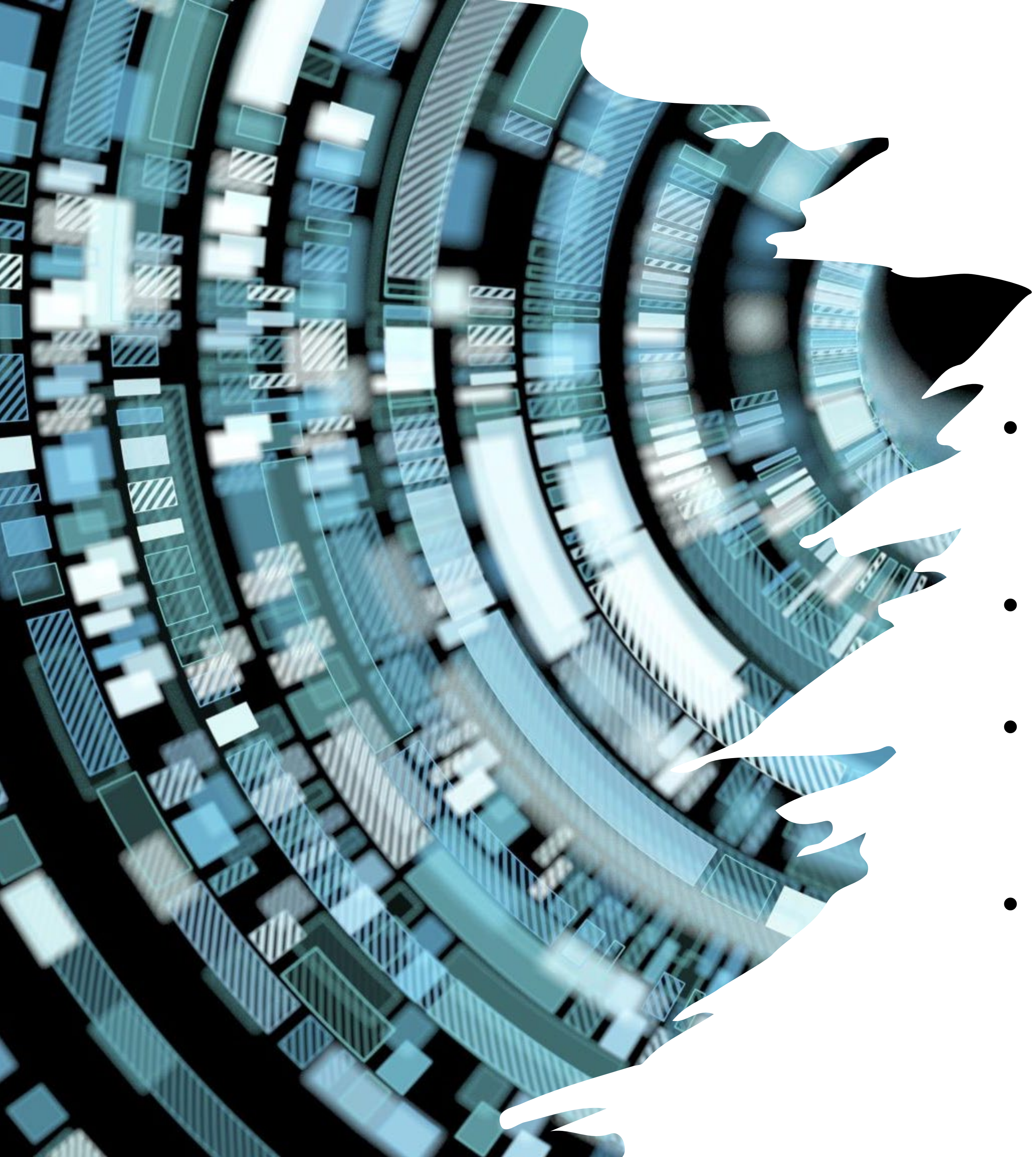

Pareto front form Cluster 1
Pareto front form Cluster 2
Pareto front form Cluster 3
Pareto front form Cluster 4
Pareto front form Cluster 5
44.0
43.0
42.0
41.0
40.0
39.0
38.0
37.0
-81.0
-80.0
-79.0
-78.0
-77.0
-76.0
-75.0
-74.0

Step 5: Data Discovery



Data Discovery

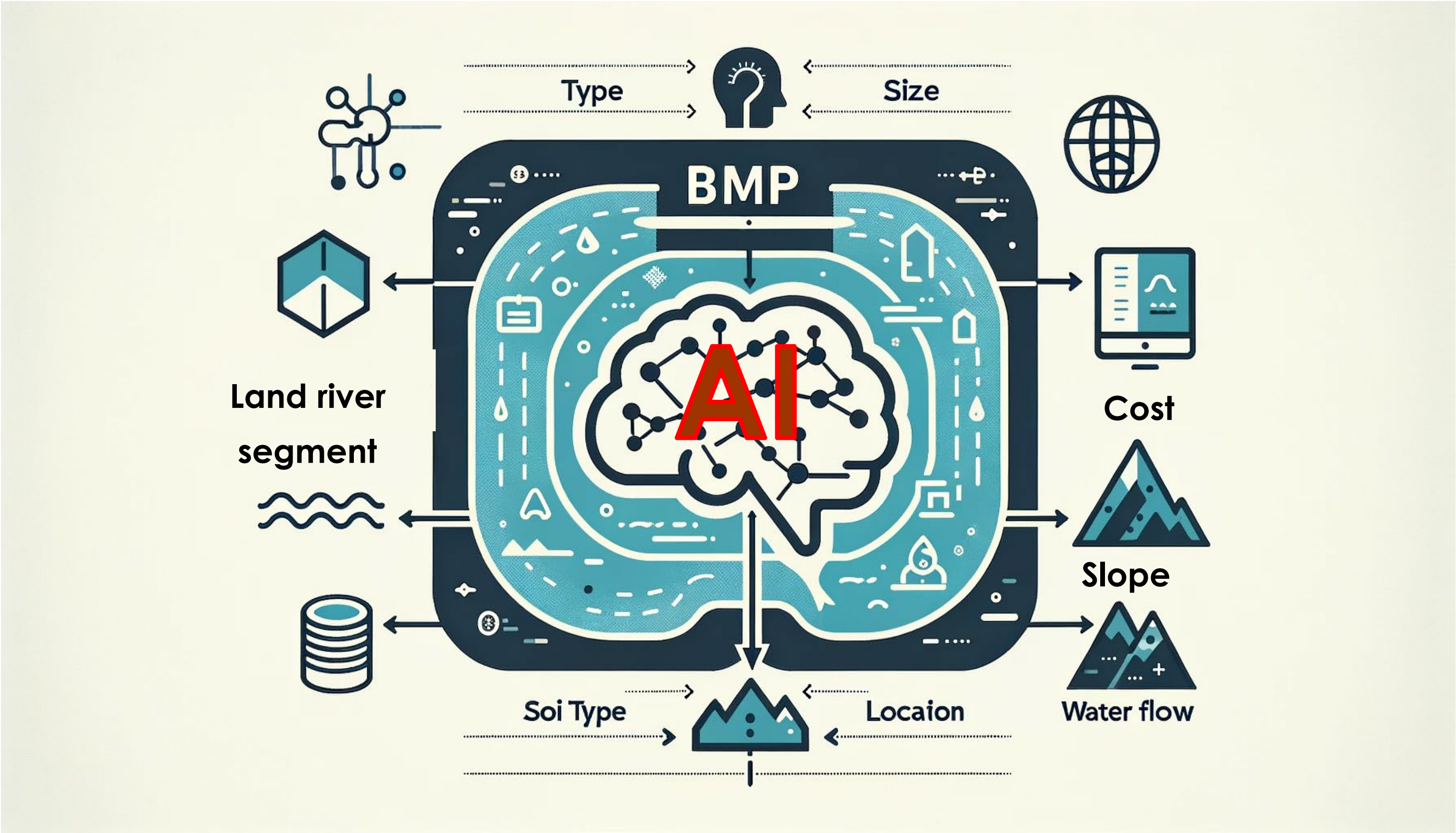
Employ **artificial intelligence** methods to recognize the **relationships between clustering, variable combinations, land river segments, and BMP selection.**



Artificial Intelligence

- **Machine Learning Approaches** (e.g., Random Forests, support vector machines, deep learning).
- **Optimization and Search Algorithms** (e.g., Genetic Algorithms).
- **Bayesian Networks and Probabilistic Models** (e.g., Markov chain Monte Carlo and Gaussian Processes).
- **Decision Support Systems** (e.g., Automated Decision Making and Cognitive Computing).

Connecting BMP and Land River Segment Characteristics

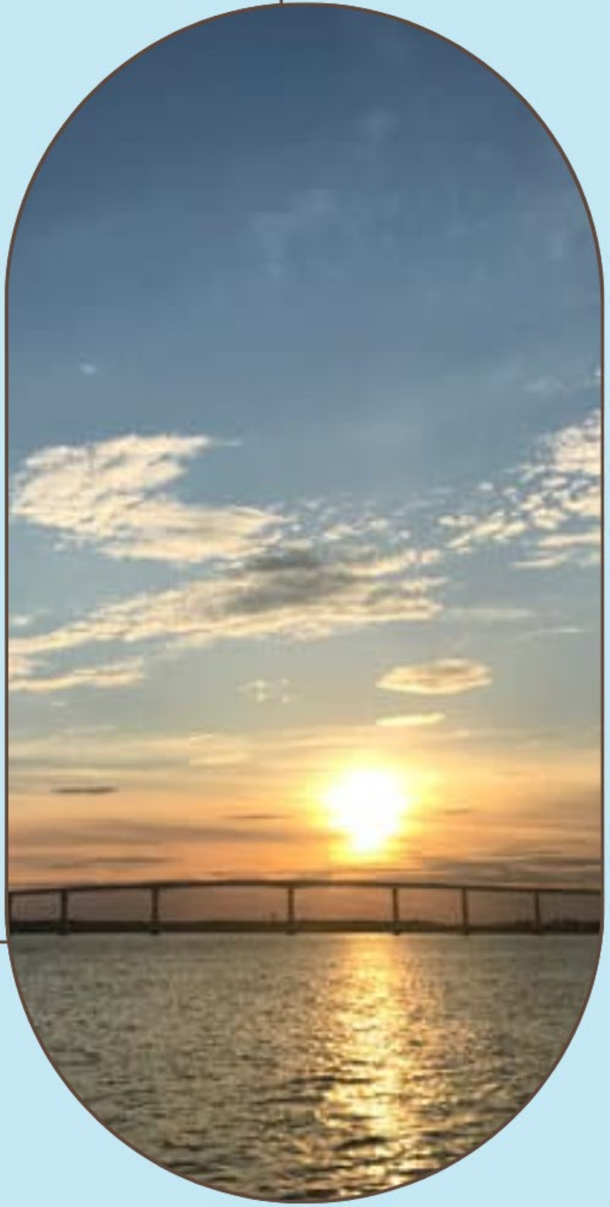


NEXT STEP

- Using artificial intelligence to enhance optimization
- Extensions to more counties and states
- Validation of our results on some critical county/state cases
- Parallel computing platform for faster execution
- Uncertainty and other practicality handling
- Workshops with CBP users for feedback and improvement of our approaches

Chesapeake Bay Optimization Webinars

Join experts in the discussion of bay preservation.



2024 Chesapeake Bay Optimization Webinars

Objective:

* Raise awareness about our BMP optimization framework.

Features:

* Interactive show of the framework capabilities

Benefits:

* Enhanced decision-making
* Foster community collaborations