Pattern Recognition using Clustering Analysis to Support Freeway Management, Operations, and Modeling.

Rajib Saha[1], Mohammed Hadi[2], and Yan Xiao[3]
[1]Graduate Research Assistant, [2]Professor, [3]Research Assistant Professor
Department of Civil and Environmental Engineering, Florida International University

Background

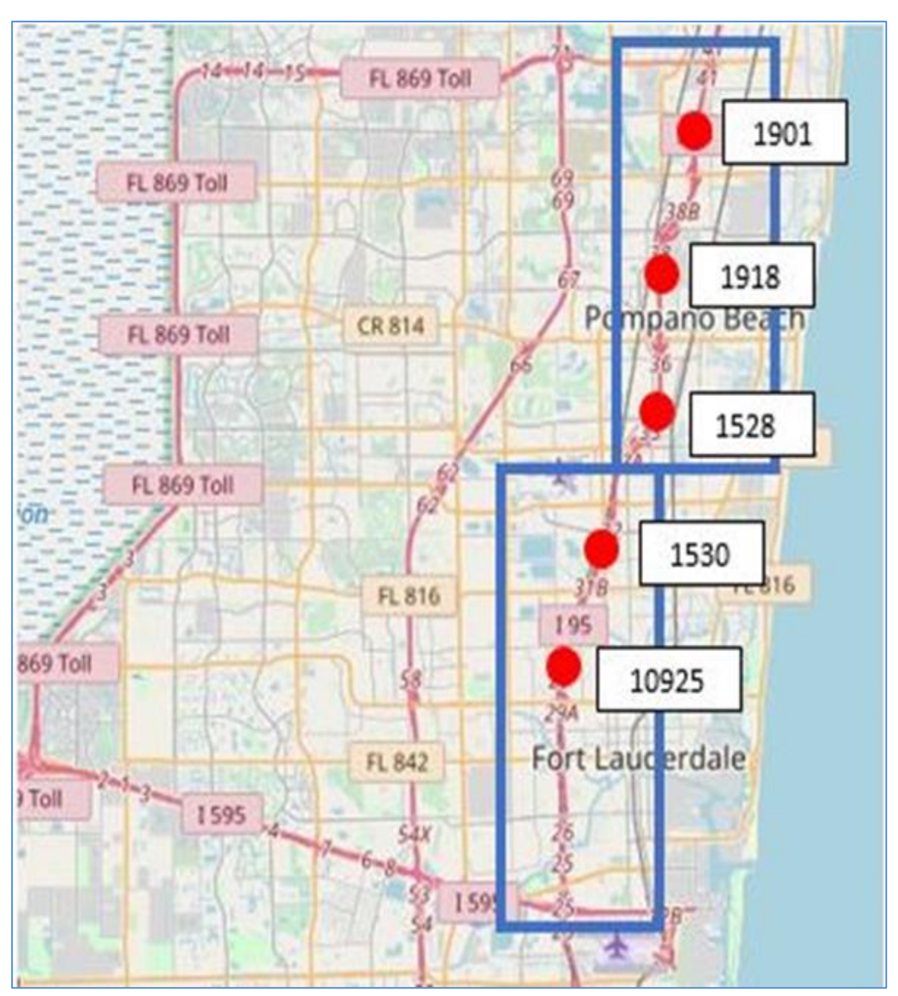
- The assessment of transportation systems including management and operation strategies are currently limited to one scenario reflecting what is supposed to be the **typical condition**.
- FHWA realizes the need for multi-scenario modeling by updating their guidance for utilizing Analysis, Modeling, and Simulation (AMS) to include **clustering analysis**.
- However, transportation agencies have limited information on

  i. Appropriate clustering technique(s),
  ii. Associated parameters (dissimilarity measure)
  iii. Optimal number of clusters, and
  iv. Selection of the representative observations

Objectives

- Investigate and demonstrate the use of a number of existing clustering methods for traffic pattern identifications.
- Support transportation agencies in identifying operational scenarios.

Study Area, Utilized Data, and Variables



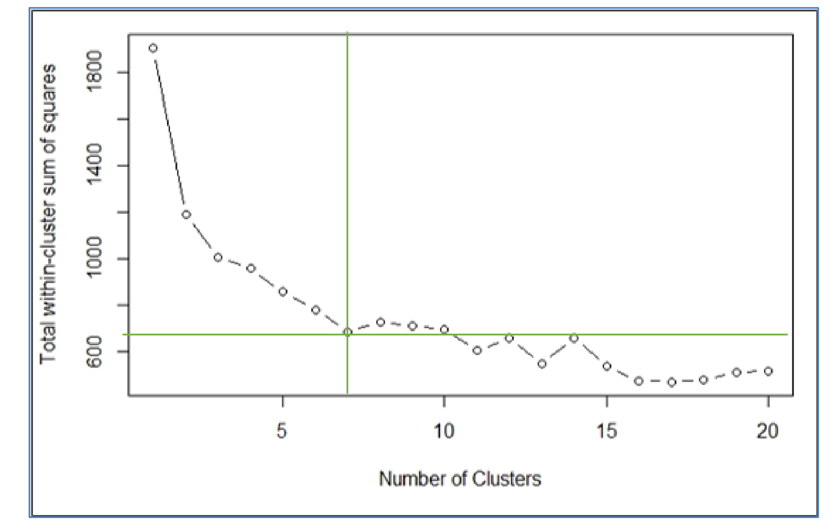
- **Freeway:** I-95, SB, AM Peak Period
- **Location:** Fort Lauderdale
- **Horizon:** Jan-Dec,2017 (excluding weekend and holidays)
- **Resolution:** 15 mins
- **Continuous Variables:**
  i. Volume Count
  ii. Speed
  iii. Occupancy
- **Categorical Variables:**
  i. Travel Lane Blockage
  ii. Incident Severity
  iii. Precipitation

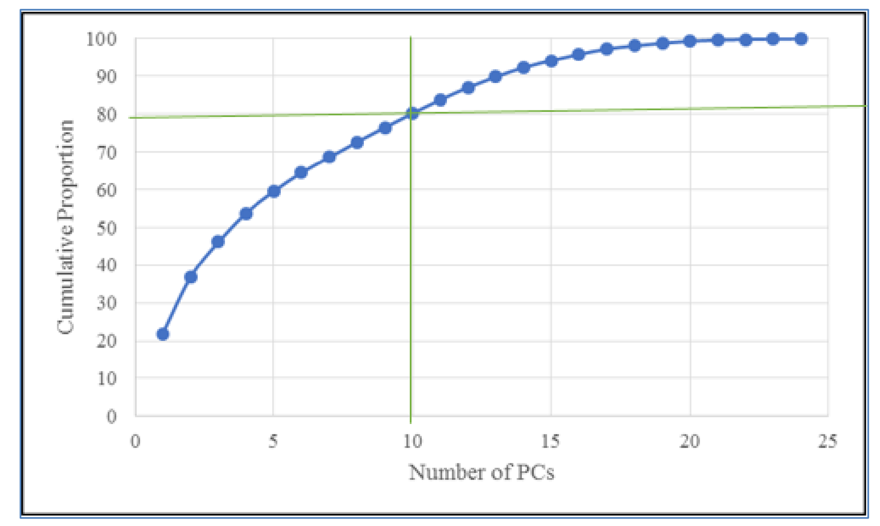
Clustering Methods

I. K-prototype
II. K-medoids
III. Hierarchical-Single
IV. Hierarchical-Complete
V. Hierarchical-Average
VI. Hierarchical-Centroid
VII. K-means with Principal Components (PCs)

Number of Clusters and PCs selection

- **Seven Clusters** were selected based on the within-cluster sum of squares producing by k-prototype.

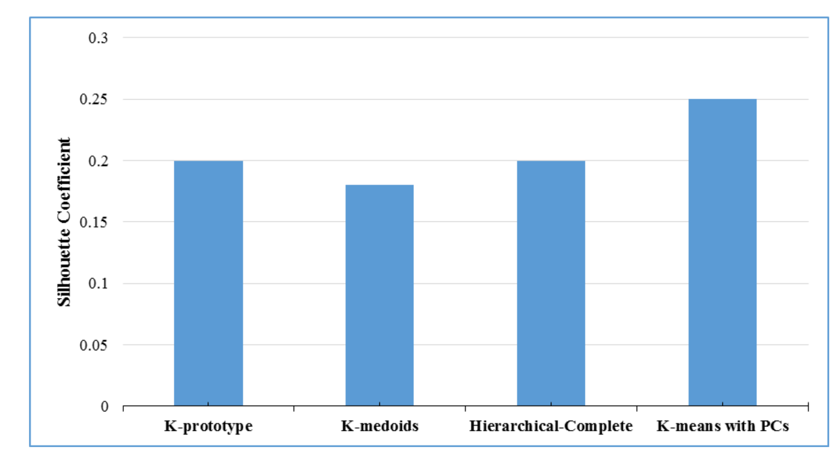


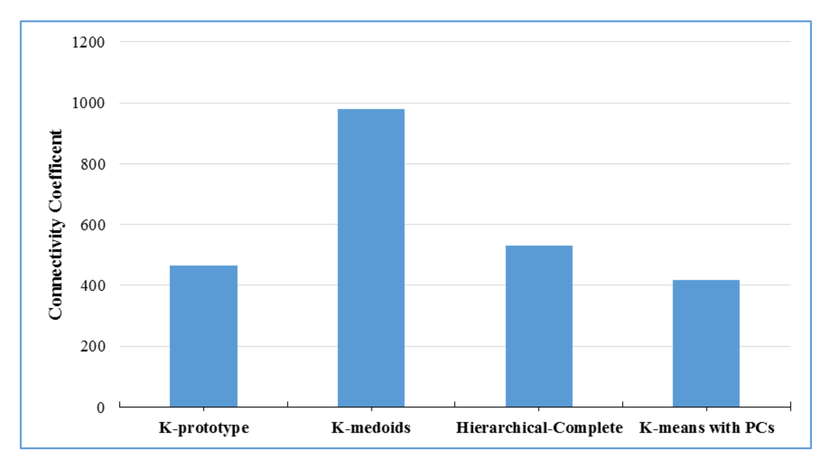


- **Ten PCs** were selected from PCAmix which retain 80% variation lies in data set.

Internal Performance Measures

- High **Silhouette coefficient** reflects compacted cluster based on intra-cluster homogeneity.

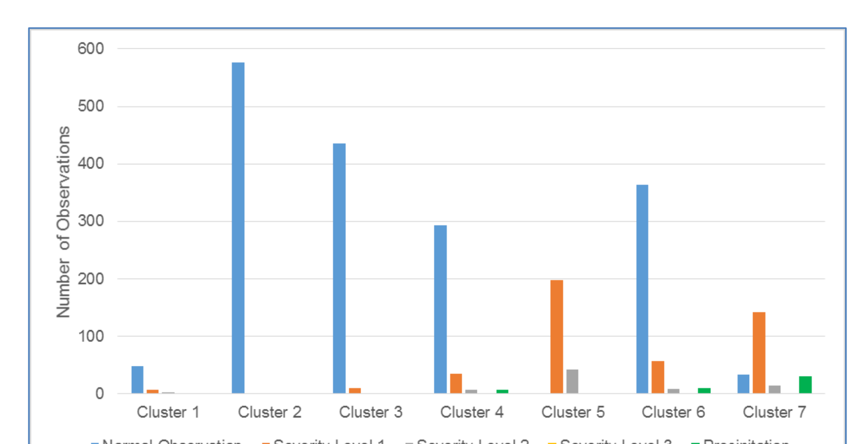




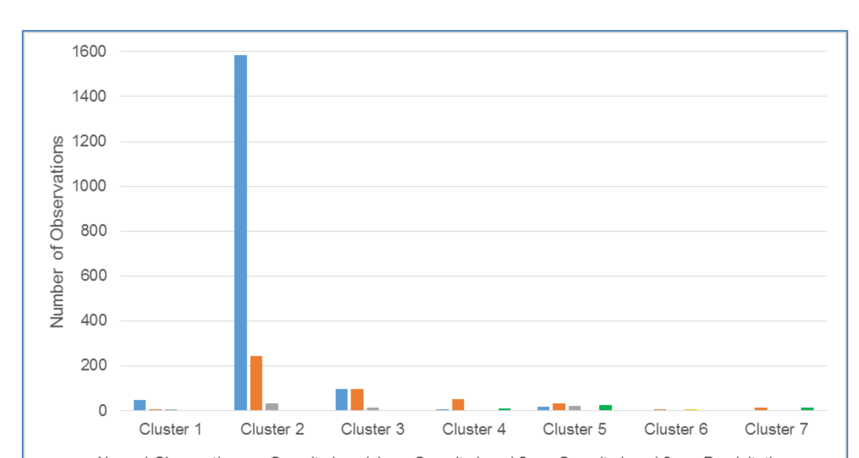
Low **Connectivity** determines nearest neighbors of the observations are placed in the same cluster.

External Performance Measures

- K-medoids failed to discern between incident and non-incident as well as precipitation and non-precipitation observations.



K-medoids

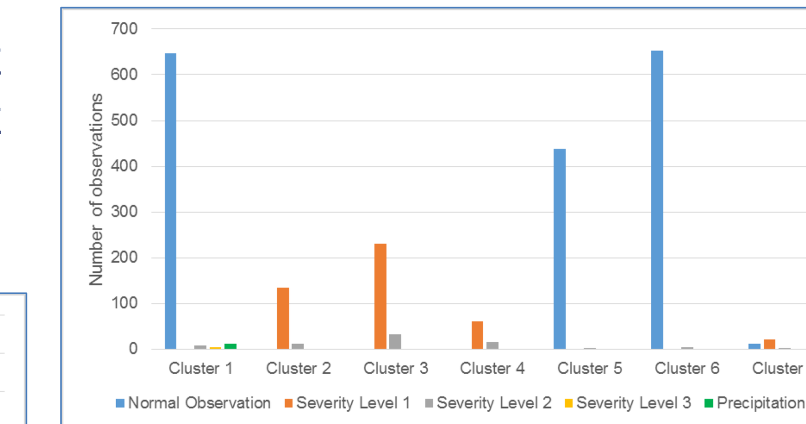


Hierarchical-Complete

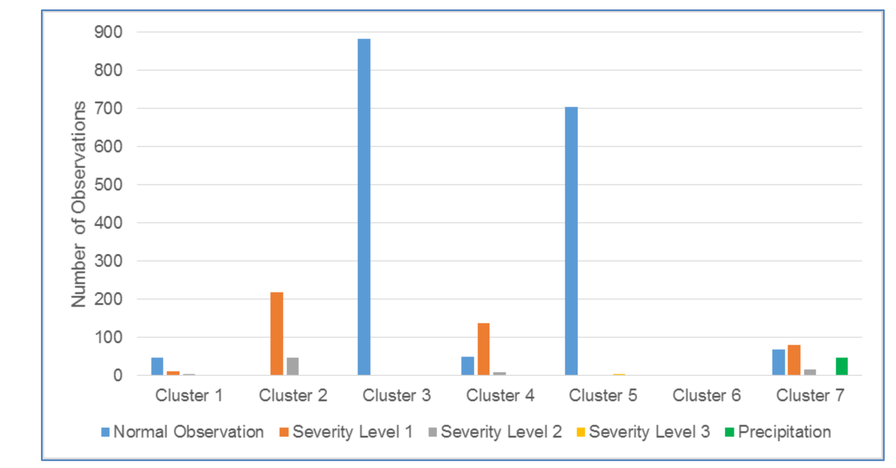
- Hierarchical method failed to differentiate among the traffic patterns and more than 80% observation placed in single cluster.

External Performance Measures (Contd.)

- K-prototype produced three distinct normal clusters, three distinct incident clusters, and one precipitation cluster.



K-prototype

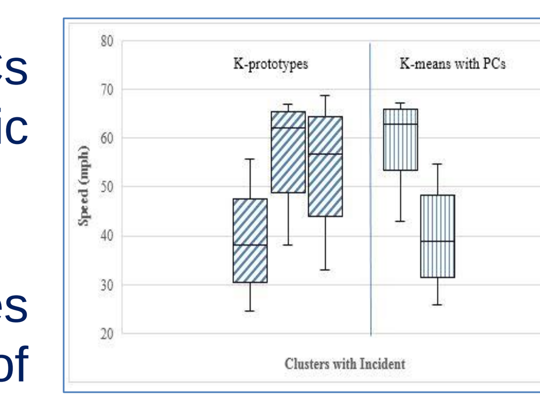


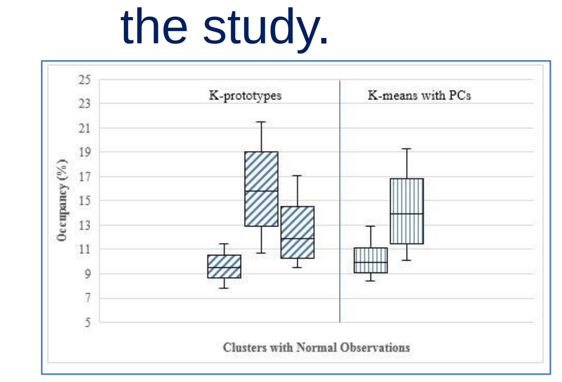
K-means with PCs

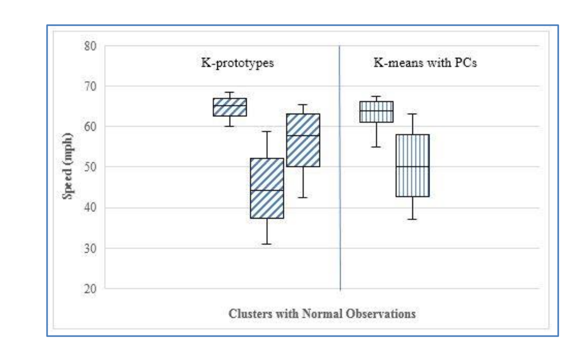
- K-means with PCs produced two distinct normal clusters, two distinct incident clusters, and one precipitation cluster.

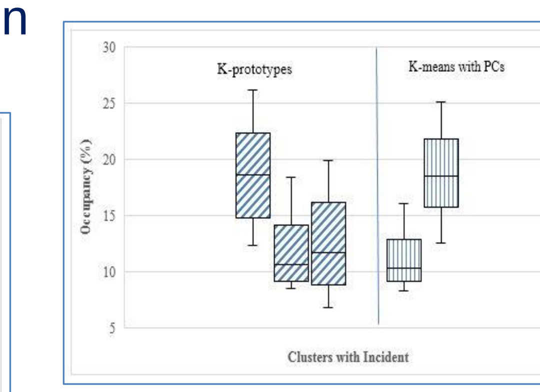
Patterns for AMS

- Both the K-prototypes and K-means with PCs produced good clustering of the traffic patterns.
- However, further investigation of the attributes of the clustered data ensures the superiority of the K-means with PCs over the K-prototypes in the study.









Operational Scenarios for AMS

1. Minor incident with high volume, moderate to high speed, and low to medium occupancy.
2. Major incident with low volume, low speed, and high occupancy.
3. Normal traffic pattern with high volume, high speed, and low occupancy.
4. Normal traffic pattern with high volume, low speed, and high occupancy.
5. Precipitation traffic pattern with volume, low speed, and high occupancy.

Acknowledgement

The presented work in this paper is part of a research project funded by FDOT and a cost share project funded by STRIDE.