Similar Days and Representative TMIs

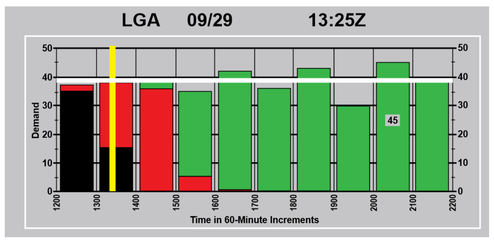
Alexander Estes, Michael Ball, and David Lovell

INSTITUTE FOR SYSTEMS RESEARCH
A. JAMES CLARK SCHOOL OF ENGINEERING

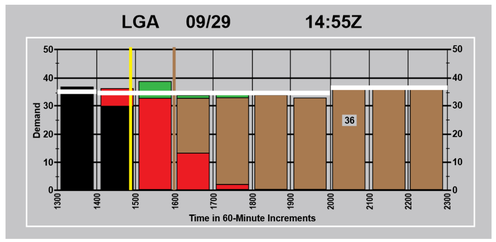
Background: Traffic Management Initiatives

In the national airspace system, imbalances between demand and capacity frequently occur. The Federal Aviation Administration addresses these by issuing traffic management initiatives (TMIs) that restrict demand.

Imbalance in capacity and demand



Imbalances smoothed by TMI

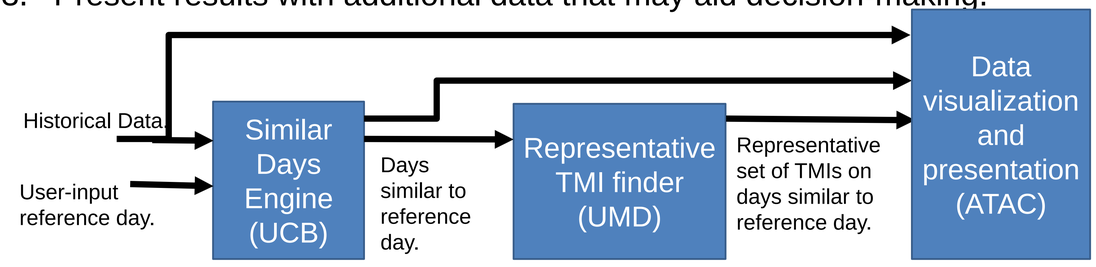


These images produced by CSC for FSM user guide

Similar Days Project Overview

Goal: data support for planning of Traffic Management Initiatives (TMIs).
Approach:

1. Identify days similar to reference day, and find the TMIs that occurred.
2. Choose a set of TMIs that are representative of the TMIs found in step 1.
3. Present results with additional data that may aid decision-making.



Identifying Representatives: *k*-center method.

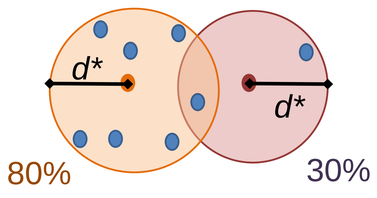
We find representative TMIs by solving a min-max facility location problem:

$$\min_{R \subseteq X, |R|=k} \left(\max_{r \in R} \left(\min_{x \in X} d(r,x) \right) \right)$$

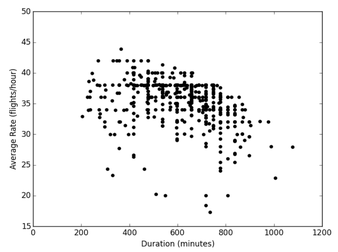
where X is the vector of data points, $d(x,y)$ is the distance between x and y and k is the desired number of representatives.

Some representatives represent common regions while others represent unusual regions. *Prevalence* measures the probability of receiving an observation similar to a representative. Prevalence is defined to be the proportion of observations within the optimal distance *d** of a representative.

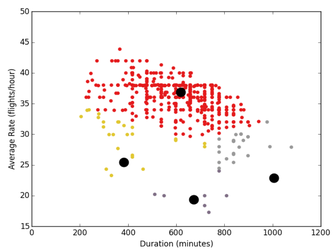
Prevalences:



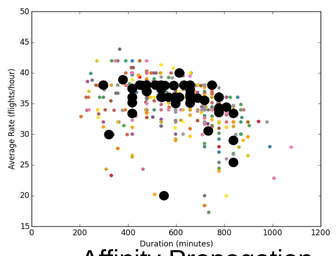
Visual comparison: *K*-center vs. others



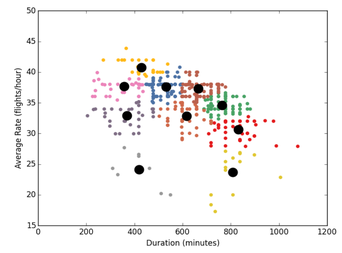
Scatter Plot: EWR GDPs



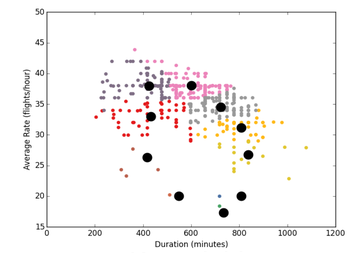
Mean shift



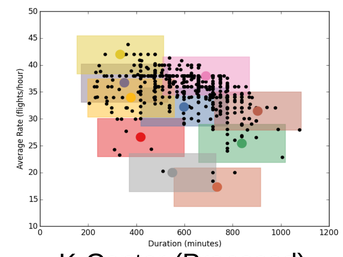
Affinity Propagation



K-means



K-medoids



K-Center (Proposed)

Predicting TMI performance

Once representative TMIs have been identified for a particular day, we produce estimates of the cancellations and the delay that would result if each representative TMI were run on that day. We developed two methods in which we combined a weighting scheme from Geographically Weighted Regression with a Random Forest and a Gradient-Boosted Forest respectively.

Results from Prediction of Avg. Delay		
Method	**Avg. Error**	**Improvement**
Unweighted Average	16.982	0.0%
Weighted Average	12.865	-24.2%
Average of k-NN	14.139	-16.7%
Unweighted Random Forest	11.759	-30.8%
Weighted Random Forest	11.612	-31.6%
Unweighted GB Forest	12.471	-26.6%
Weighted GB Forest	12.381	-27.1%

Results from Prediction of Cancellations		
Method	**Avg. Error**	**Improvement**
Unweighted Average	16.381	0.0%
Weighted Average	10.511	-35.8%
Average of k-NN	13.297	-18.8%
Unweighted Random Forest	10.924	-33.3%
Weighted Random Forest	9.310	-43.2%
Unweighted GB Forest	10.437	-36.3%
Weighted GB Forest	8.443	-48.5%

Funded by NASA NRA NNH14ZEA001N

In collaboration with University of California Berkeley and ATAC, Inc.