Selecting Parameters in Performance-Based Ground Delay Program Planning

Alexander Estes, David Lovell and Michael Ball
University of Maryland – College Park

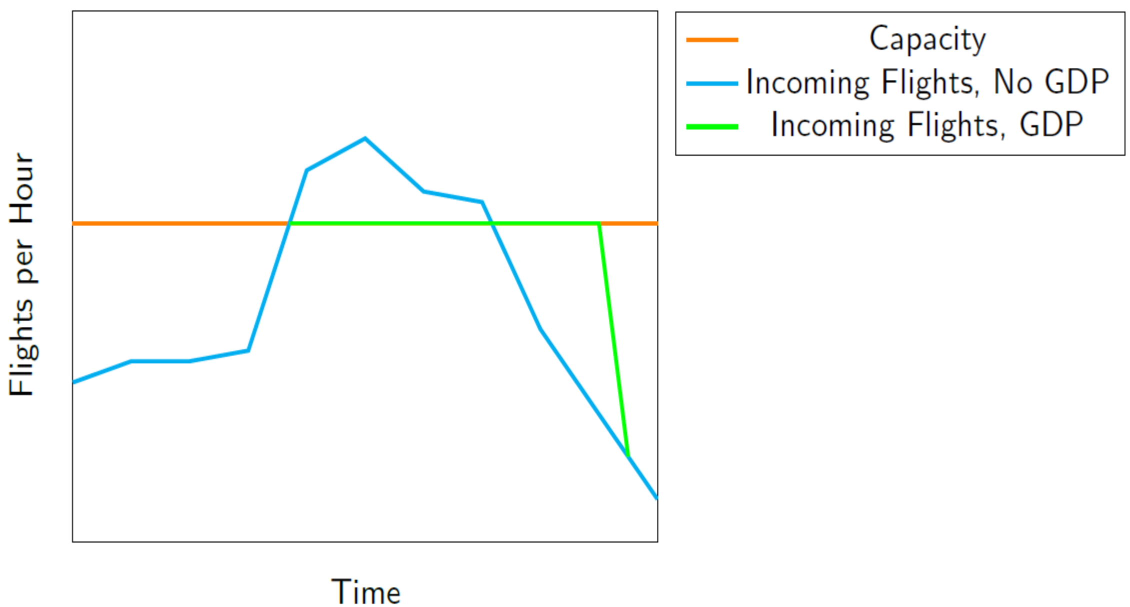
Research funded by NASA grant no. NNX14AJ79A

Outline

- **Background**
- Method
- Computational Results
- Conclusion

Ground Delay Programs (GDPs)

- Scheduled traffic can exceed capacity of airport to handle arrivals.
- To prevent unsafe situations, FAA will delay flights on the ground.
- FAA must select parameters such as time of affect and number of flights allowed to access resource.



Background I

- IP Models for GDP planning:
  - Many IP models proposed (e.g. Mukherjee and Hansen 2007, Ball et al. 2003, Richetta and Odoni 1993).
  - In most existing IP models the objective is to minimize weighted sum of air delays or ground delays.
  - Tend to offer little flexibility in the objectives used.

Background II

- GDP performance is not one dimensional:
  - For example, Liu and Hansen (2013) proposed 5 types of metrics designed to measure different aspects of GDP performance (capacity utilization, efficiency, predictability, equity and flexibility).
  - It is not generally possible to optimize all of these aspects simultaneously; a trade-off must be selected.

Background III

- GDP performance as a multi-objective problem:
  - Liu and Hansen (2015) proposed a method for planning GDPs whose objective combines predictability and delays.
  - Ball et al. (2017) proposed the COuNSEL mechanism, which would solicit preferences from flight operators and then select a 3-dimensional consensus vector. This vector indicated a desired tradeoff between three aspects of GDP performance (efficiency, throughput and predictability)

Goal

- Given:
  - A set of weather and traffic conditions.
  - A "target vector" specifying the desired performance.
  - Data on past GDPs/weather/traffic.
- Select:
  - A set of GDP parameters
- Goal:
  - Resulting performance should be close to desired performance.
- Proposed method is data-driven, can be used with any desired performance metrics.
- Could be used to implement a COuNSEL-like mechanism
- More generally, can support GDP decisions.

Outline

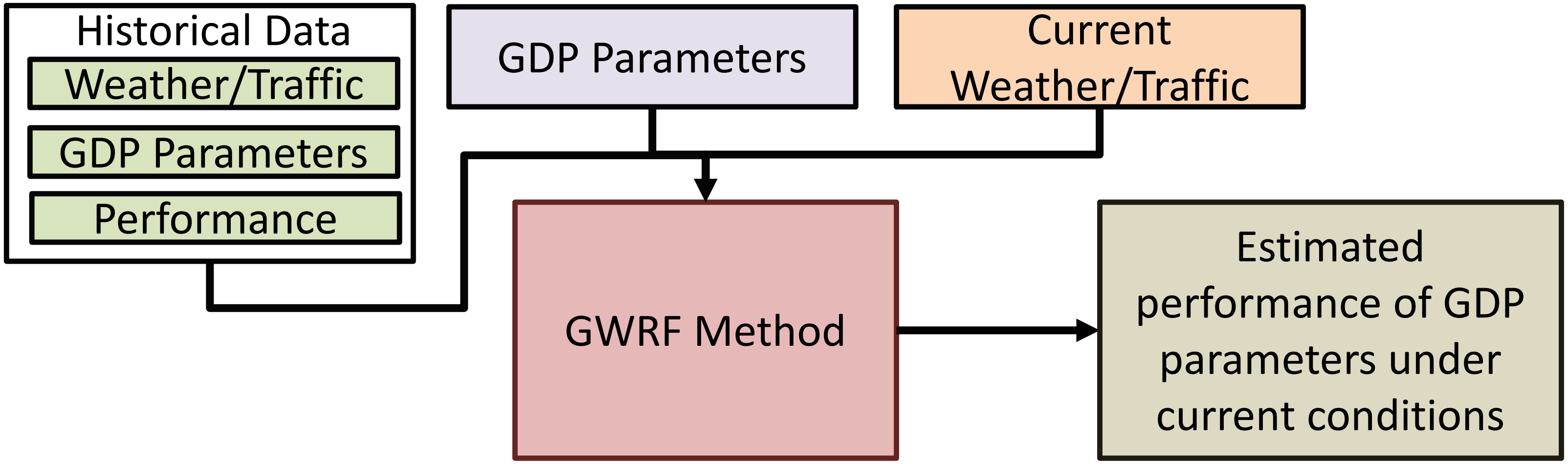
- Background
- **Method**
- Computational Results
- Conclusion

Previous Work

- Distance score between weather/traffic situations at a given airport (Gorripaty, Hansen & Podznukhov 2016)
  - Provides numerical measure of distance between pair of weather/traffic situations at an airport.
  - Features considered include convective weather near airport, weather at terminal, scheduled arrivals at airport.

Previous Work II

- GWRF Method (Estes, Lovell and Ball 2017)
  - Estimates performance of a GDP in a given set of weather and traffic conditions at a specific airport
  - Uses aforementioned distance score, converted into a similarity score.



Naïve Method I

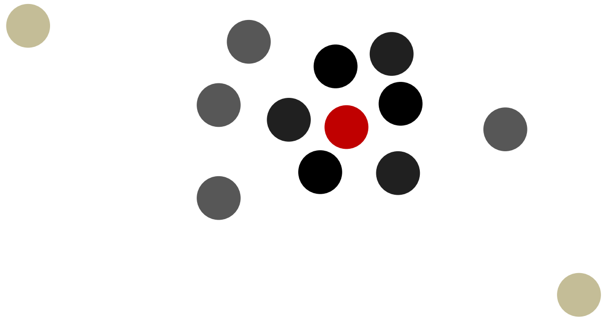
- Take each GDP that has been run in our entire dataset at the given airport.
- Use GWRF to estimate performance of that GDP under current conditions.
- Select the GDP whose estimated performance is closest to the target.

Naïve Method II

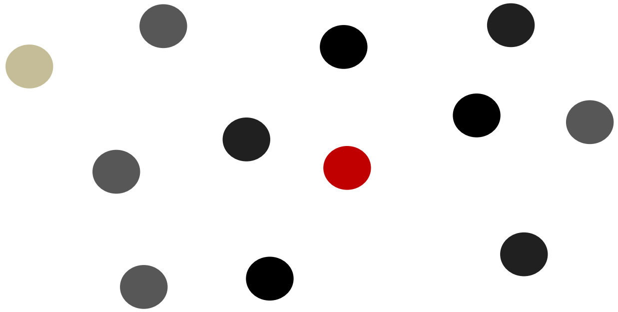
- Problem:
  - Accuracy of estimates varies with GDP parameters
  - If similar GDPs have been implemented many times in similar situations to current situation, then we expect the estimate to be accurate.
  - If not many similar GDPS have been implemented, or have only been implemented in very different situations, we expect estimate to be inaccurate.

Prediction-Weighted Similarity Score I

- Idea: develop a scoring system.
  - Higher score if estimate is based mainly on data from situations similar to current situation.
  - Lower score if estimate is based mainly on data from situations dissimilar to current situation.
  - Restrict GDPs to those with high score.



Higher Score



Lower Score

Prediction-Weighted Similarity Score II

- Definition of scoring system:
  - We have a measure of similarity between weather/traffic conditions (from distance measure)
  - Estimates are weighted sums of historical performance observations (weights depend on current situation and GDP parameters)
  - Apply same weights to similarity score between historical situation and current situation

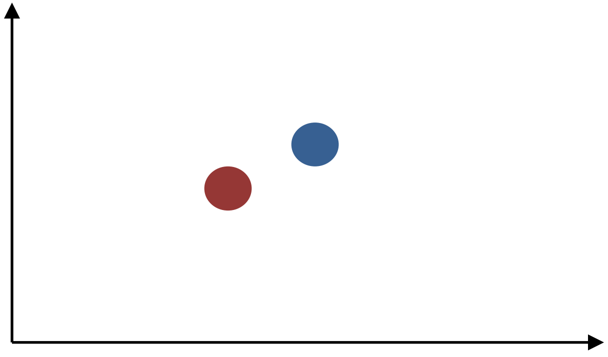
$$\hat{g}(x;z) = \sum w_i(x;z)\, y_i \qquad \hat{s}(x;z) = \sum w_i(x;z) s_i\,(z)$$

Prediction-Weighted Similarity Score III

- We can use this score to restrict our attention to GDPs whose performance estimates are well-supported by data.

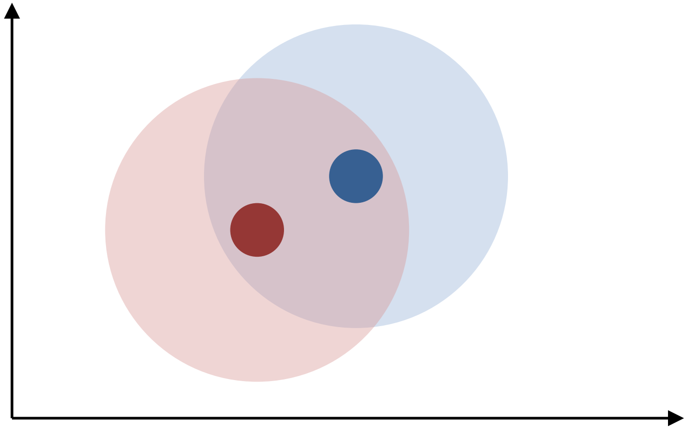
Naïve Method III

- Another downside:
  - The target vector might not be on the efficient frontier
  - It might be possible to find a set of GDP parameters that is better than the target vector in all performance measures.



Moving Closer to Efficient Frontier

- Difficulty:
  - Performance estimates are not certain.



Moving Closer to Efficient Frontier I

- Idea - adapt a ranking scheme used in evolutionary algorithms (Hughes 2000).
  - Low rank indicates that the point is probably closer to efficient frontier.
  - High rank indicates that the point is probably further from the efficient frontier.

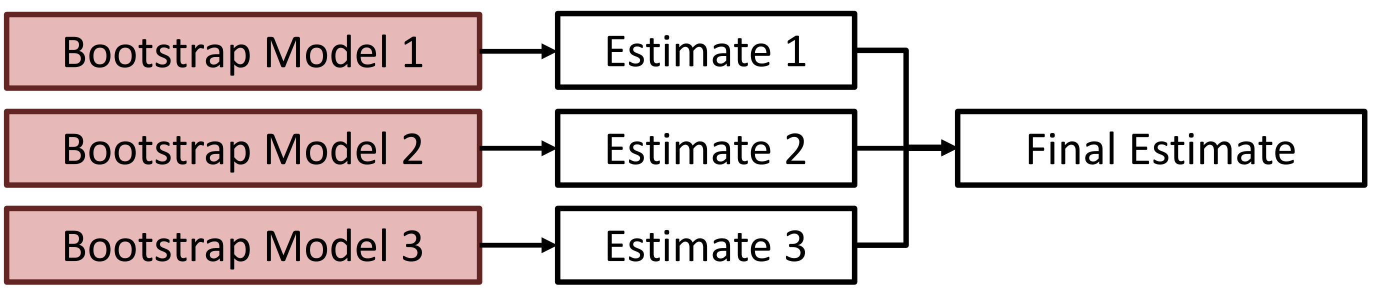
Dominance Rank I

- For each pair of GDP parameters *t* and *s*, we estimate the probability that *t* dominates *s*.
- Rank:

$$r(s;z) = \left(\sum_{t \in S} P(t \leq s; z) + \frac{1}{2} P(t \sim s; z) \right) - \frac{1}{2}$$

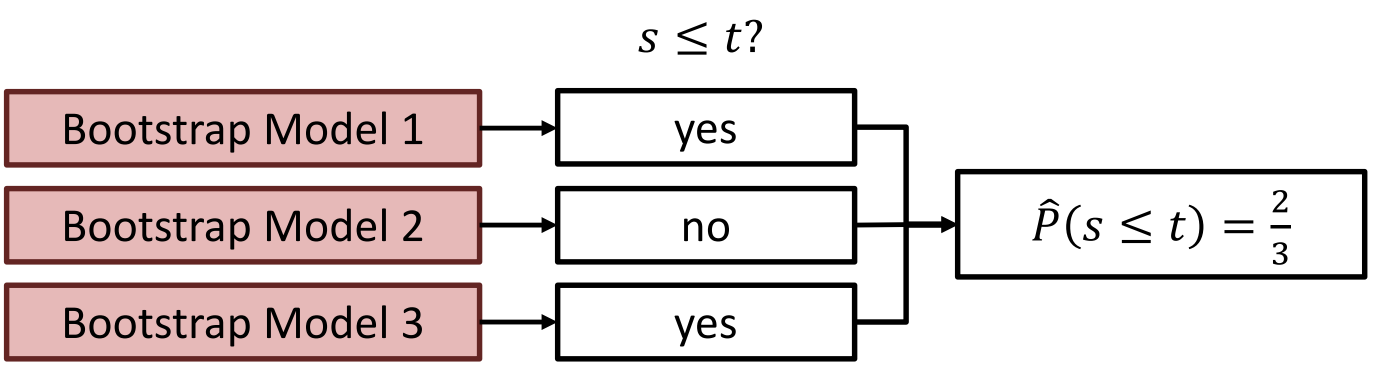
Dominance Rank II

- Probabilities are estimated as follows:
  - Performance estimation is done by bootstrapping.
  - Bootstrapping: fitting the same model multiple times with randomness introduced.
  - The estimates of these models are then averaged.



Dominance Score III

- Probabilities are estimated as follows:
  - For each models in this collection, estimate performance of parameters *s* and *t*.
  - $P(s \leq t)$ is the proportion of models in which performance of *s* dominates *t*.



Constrained Greedy Selection

- Set a threshold $s*$ for the prediction-weighted similarity score.
- Remove choices of GDP parameters with similarity score larger than $s*$.
- If all choices have similarity score larger than $s*$, then there is insufficient data to use this method.
- Form list of options that is increasing in predicted distance to target vector but have decreasing dominance rank:
  - Let x^1 be choice of GDP parameters with predicted performance closest to target vector.
  - If there is a choice of GDP parameters whose dominance rank is lower than x^k, we define x^{k+1} to be such an element whose predicted distance to the target vector is lowest.

Outline

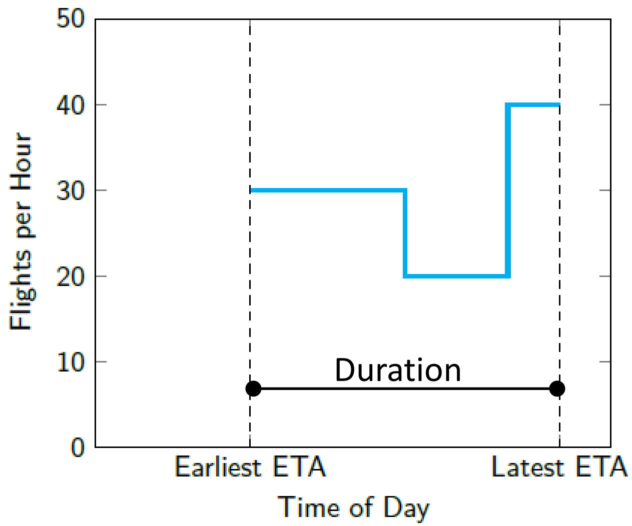
- Background
- Method
- **Computational Results**
- Conclusion

Experiment Setup

- 480 days at EWR where GDPs were run.
  - 80% training, 20% test
- Performance measures:
  - Average arrival delay
  - Holding events
  - Data source: ASPM
- We use distance scores provided by U.C. Berkeley (Gorripaty, Hansen & Podznukhov 2016)

GDP Features I

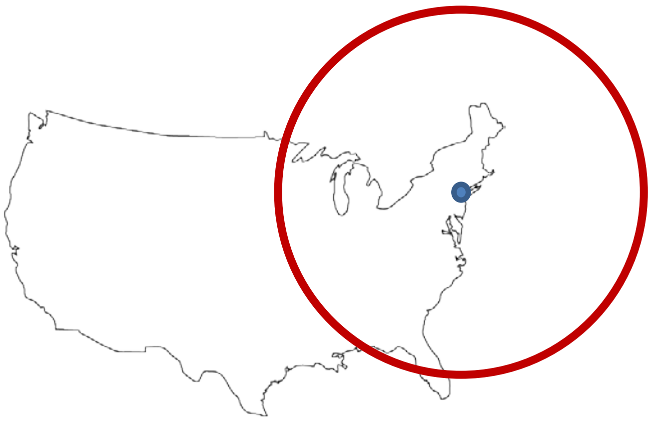
- All features reflect first planned GDP. Revisions are not included.
- File time - minutes after 4:00 a.m. local
- Start Time - minutes after 4:00 a.m. local
- Duration - minutes
- Average called rate
- Data source: NTML



GDP Features II

- Scope - number of CORE30 airports

Scope declared as radius



Scope declared by center

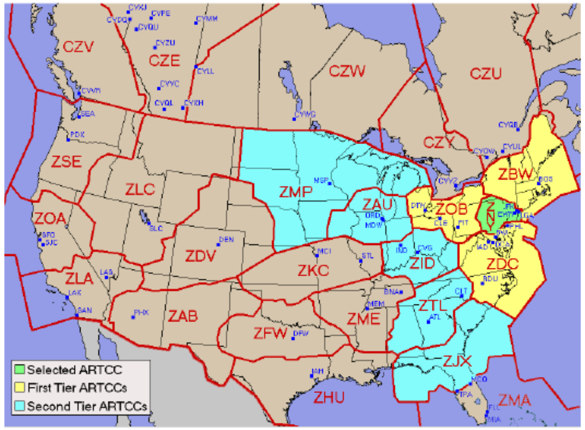
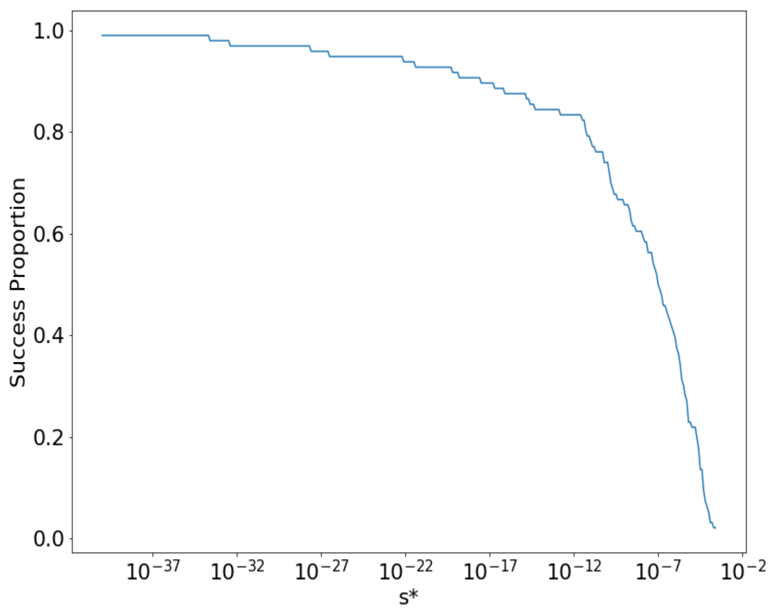


Image from FAA: http://www.fly.faa.gov/ois/tier/themap.htm

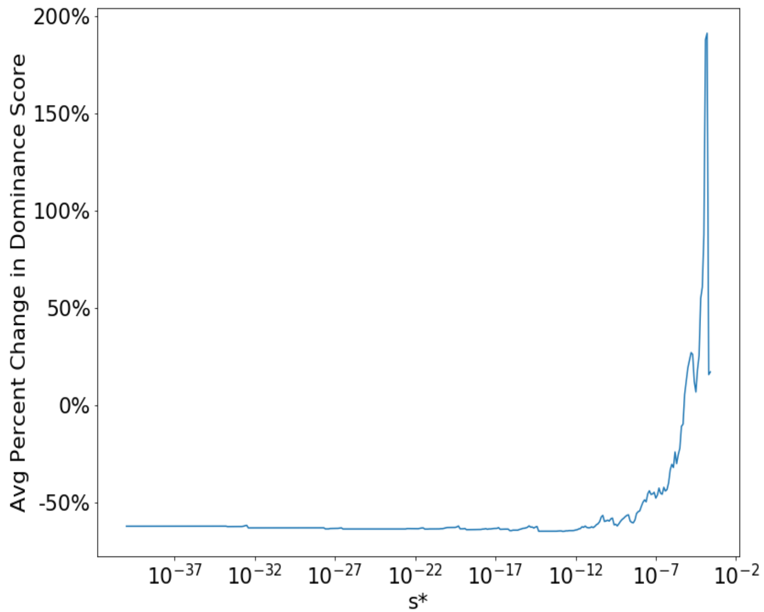
Experiment Setup II

- Performance estimation method is tuned on training set.
- For each day in test set:
  - Ran the naïve process to select a single choice of GDP parameters
  - Ran the constrained greedy selection to select a list of choices of GDP parameters

Results – Dominance score

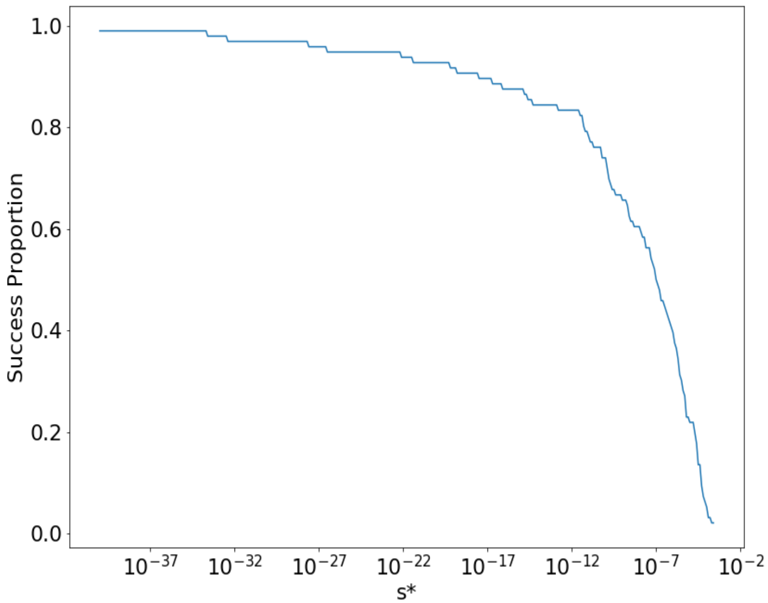


Proportion of test where proposed method returns at least one element vs. similarity threshold.

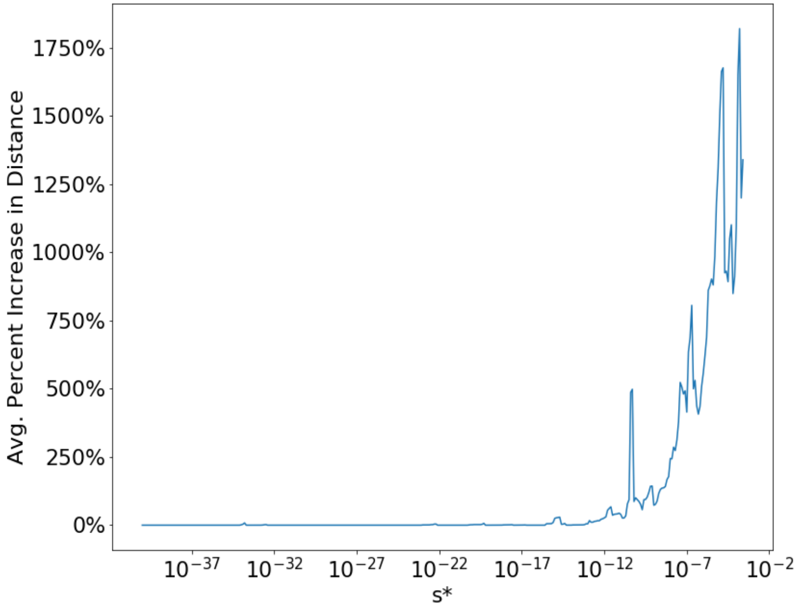


Improvement in dominance score vs. similarity threshold.

Results – Est. distance to target vector



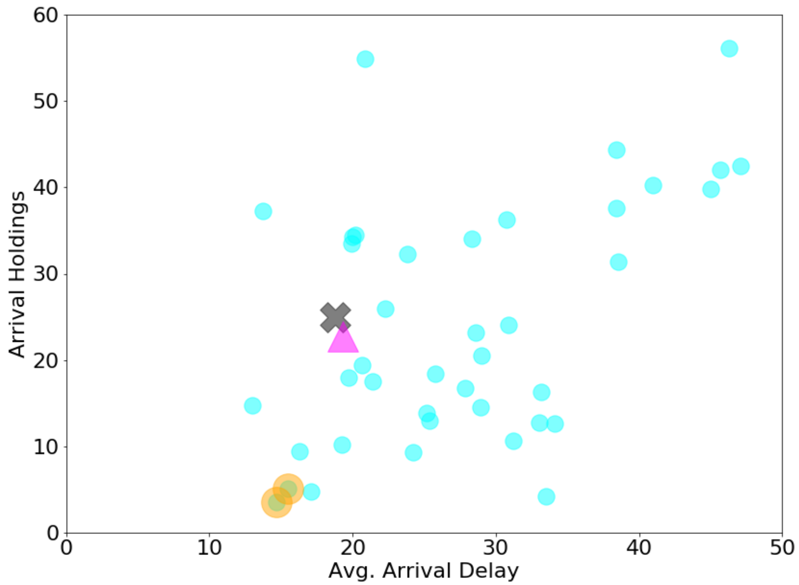
Proportion of test where proposed method returns at least one element vs. similarity threshold.



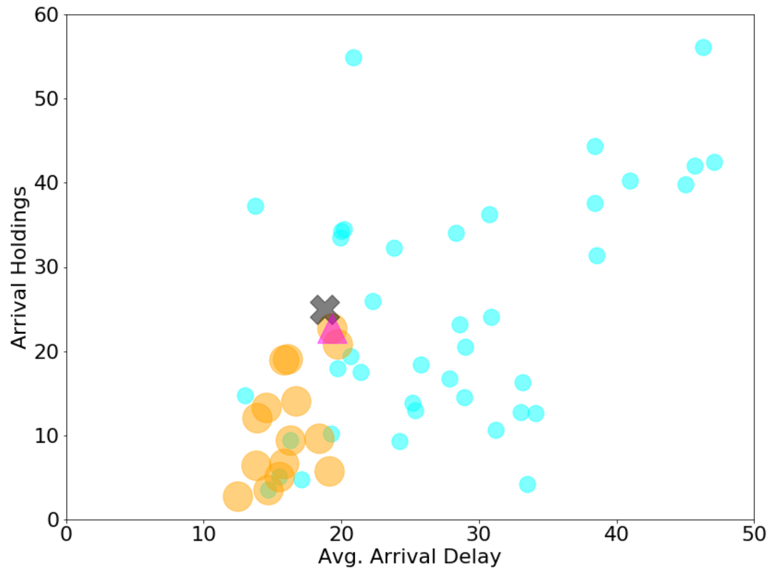
Increase in distance vs. similarity threshold.

Example

November 14, 2011



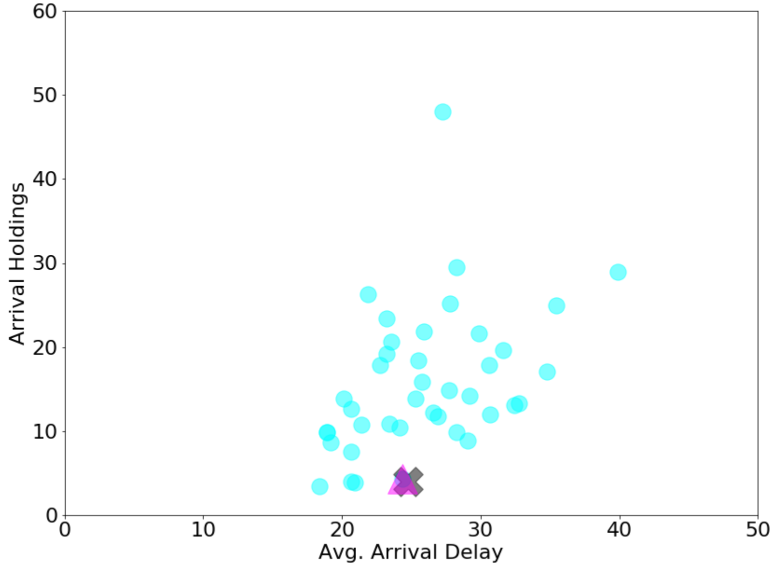
Stricter similarity threshold



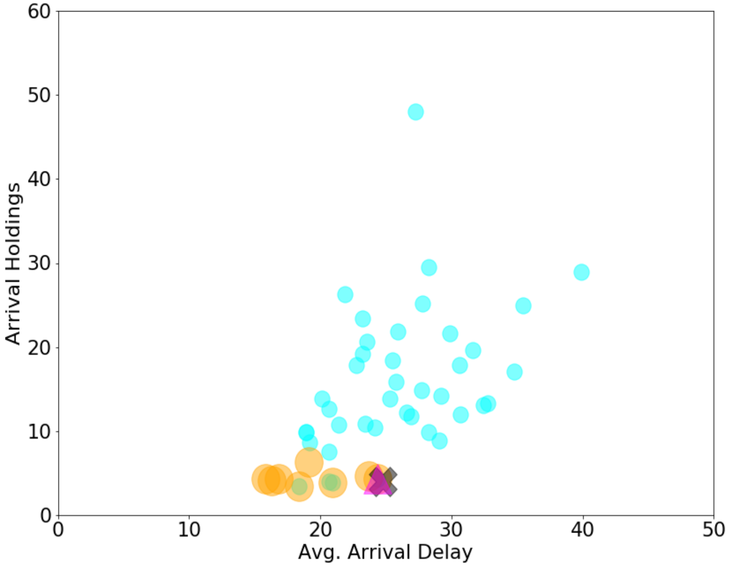
More lenient similarity threshold

Example

May 28, 2013



Stricter similarity threshold



More lenient similarity threshold

Outline

- Background
- Method
- Computational Results
- **Conclusion**

Conclusions

- New data-driven method in support of multi-objective GDP planning
- Can be used with any type of performance measure
- Can ensure that suggested GDP options are well-supported by data
- Can identify solutions whose estimated performance is closer to efficient frontier than greedy solutions.

Further Work

- How to choose similarity threshold?
- Alternate method for when method does not provide any suggestions.
- Combining with other mechanisms (e.g. COuNSEL).
- Presenting results to decision-makers.
- Other types of traffic management initiatives