Similar Historical Days and Air Traffic Flow Management Response Strategies

PROFESSOR DAVID LOVELL

NEXTOR II SYMPOSIUM

OCTOBER 2018

UNIVERSITY OF MARYLAND 1856

A. JAMES CLARK SCHOOL OF ENGINEERING

Background of similar days type activities

- 2013 paper by Shon Grabbe, Banavar Sridhar, and Avijit Mukherjee
  - National and airport-level clusters
  - Based on traffic, weather, and GDP existence
- Berkeley work on GDP decision-making (Mark Hansen, Yi Liu)
  - Develop performance criteria for GDPs
  - Relate performance goals to GDP parameter decisions: planned clearance time, scope, early cancellation policy
- Service level expectations (Berkeley (Mark Hansen, Yi Liu, Lei Kang, Maryland (Michael Ball, Prem Swaroop), MIT (Cynthia Barnhart, Vikrant Vaze, Chiwei Yan))
  - Build consensus amongst operators as to daily goals, given projected traffic, weather,etc.
  - Ultimately, the outputs were designed to serve as inputs to a GDP (or AFP) parameter design effort
- Stochastic optimization for GDP planning (Charles Glover, Michael Ball)
  - Better balance of equity and efficiency than ration-by-distance when there is uncertainty about weather duration

A. JAMES CLARK
SCHOOL OF ENGINEERING

Background of similar days type activities

- NASA Project
  - Berkeley, UMD, ATAC team
  - RAND project led by Kenneth Kuhn (also had earlier work on Pareto-optimal GDP decision-making)
- MITRE CAASD
- Avmet
- Others?

A. JAMES CLARK
SCHOOL OF ENGINEERING

Our focus on human decision support

- Trying to produce a “menu” of solution strategies
  - Each is not necessarily intended to be the deployed strategy, but represents a set of deployed strategies that have been used in the past
  - Nevertheless, each is a valid strategy
  - The menu should consist of things that are different enough from each other so as to be useful alternatives to the human, but should also in some sense “cover” the set of historical actions
  - Performance attributes can then be tagged to the representatives, although this is not part of what makes them similar to each other

Our team

- Berkeley
  - Mark Hansen, Alexey Pozdnukhov, Yi Liu, Michael Seelhorst, Sreeta Gorripaty
  - Traffic, weather similarity engine
- Maryland
  - Michael Ball, David Lovell, Alex Estes
  - GDP parameter similarity, performance prediction
- ATAC
  - John Schade, Kennis Chan, Corey Warner
  - Data, GUI design

A. JAMES CLARK
SCHOOL OF ENGINEERING

Hourly weather forecast similarity

- Build a reference set of similar hours:
  - Same runway configuration
  - Same meteorological conditions (VMC/IMC)
  - Small absolute difference in Airport Acceptance Rate
- WF has the following components:
  - TS: indicator for thunderstorms
  - Sn: indicator for snowstorms
  - $Vis1$: $\log(\text{visibility})$
  - $Vis2$: $\max\{0, \log(\text{visibility}/4)\}$
  - $Ceil1$: $\log(\text{ceiling})$
  - $Ceil2$: $\max\{0, \log(\text{ceiling}/3000)\}$
  - Ws: absolute wind speed
  - WN: north-south component of wind speed
  - WE: east-west component of wind speed

A. JAMES CLARK
SCHOOL OF ENGINEERING

Hourly weather forecast similarity

- Learning distance metrics

$$d_A(WF_i, WF_j) = \|WF_i - WF_j\|_A = \sqrt{(WF_i - WF_j)^{\mathrm{T}} A (WF_i - WF_j)}$$

  - Quadratic form, generalization of (co)variance that allows for learned weighting coefficients in the A matrix
- Determine A matrix by solving an optimization problem:

$$\min_A \sum_{(WF_i, WF_j) \in S} \|WF_i - WF_j\|_A^{\ 2}$$

$$s.t. \sum_{(WF_i, WF_j) \in D} \|WF_i - WF_j\|_A \geq 1$$

$$A \succcurlyeq 0$$

  - This ensures that A is nontrivial, produces a mathematically legitimate metric, and makes the greatest distinction between pairs of forecasts that are similar and those that are different
- This is a semi-supervised process because human input determines the sets S and D, but the learning algorithm learns the matrix A.

UNIVERSITY OF MARYLAND 18 56 A. JAMES CLARK SCHOOL OF ENGINEERING

Daily weather forecast similarity

- Estimation results for matrix A:

TABLE I. ESTIMATION RESULTS ON A MATRIX

Variables	*TS*	*Sn*	*Vis2*	*Vis1*	*Ceil2*	*Ceil1*	*Ws*	*WN*	*WE*
TS	61.15[a]	4.78	4.14	15.52	-11.74	12.30	2.22	-0.47	-0.56
Sn	4.78	7.05	20.28	-7.67	-3.80	3.84	-0.10	-0.07	-0.10
Vis2	4.14	20.28	66.55	-17.37	-29.90	30.09	2.91	-0.75	-0.81
Vis1	15.52	-7.67	-17.37	39.05	-40.69	41.07	8.46	-1.35	-1.31
Ceil2	-11.74	-3.80	-29.90	-40.69	87.12	-87.76	-15.24	2.64	2.61
Ceil1	12.30	3.84	30.09	41.07	-87.76	88.41	15.35	-2.66	-2.63
Ws	2.22	-0.10	2.91	8.46	-15.24	15.35	2.76	-0.47	-0.46
WN	-0.47	-0.07	-0.75	-1.35	2.64	-2.66	-0.47	0.08	0.08
WE	-0.56	-0.10	-0.81	-1.31	2.61	-2.63	-0.46	0.08	0.08

a. All the weights are scaled by a factor of 1000.

- $D_{J,K} = \sum_{i=T_s}^{T_e} \left\| WF_{J,i} - WF_{K,i} \right\|_A^2$

A. JAMES CLARK
SCHOOL OF ENGINEERING

Relative importance of weather

- Build five hypothetical special cases of A for weather scenarios
- Estimate hypothetical distances between hourly TAFs using these matrices

TABLE II. STATISTICS OF HYPOTHETICAL HOURLY DISTANCES

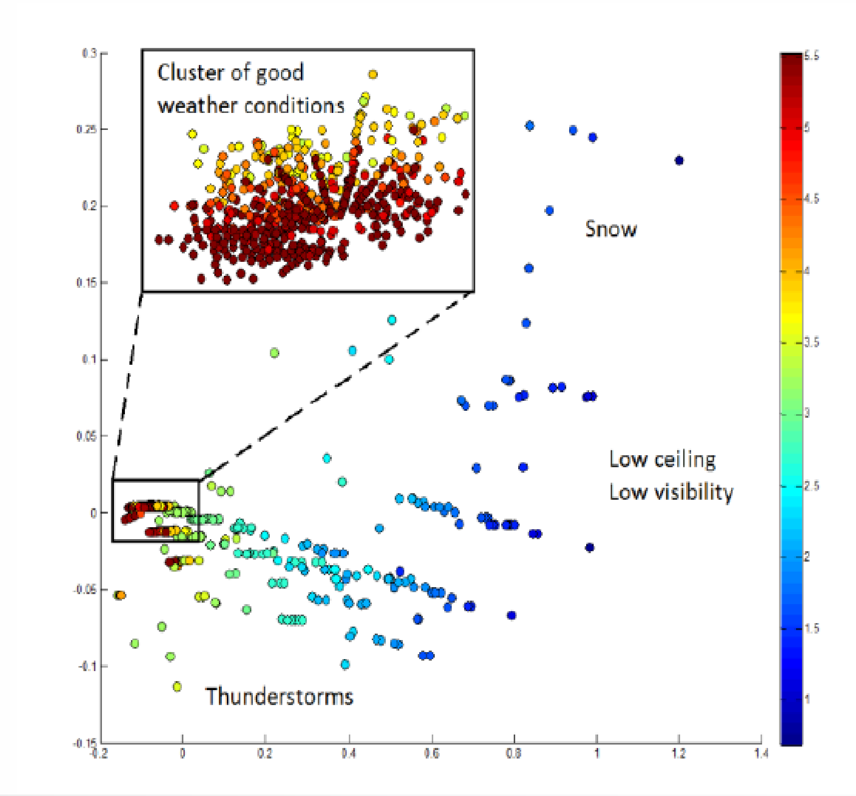
Statistics	Thunderstorm	Snow	Visibility	Ceiling	Wind
Mean	3.5[a]	2	47	94.6	24.5
Median	0	0	0	3.4	20.1
Max	247.3	84	491.5	809.9	320.8
% non-zero obs.	1.43	2.41	31.3	71.6	99.6
Mean of non-zero obs.	247.3	84	150.3	132.1	24.6
Median of non-zero obs.	247.3	84	149	5.5	20.2
Std. of non-zero obs.	0	0	103.4	190.9	21.8

a. All the weights are scaled by a factor of 1000.

A. JAMES CLARK
SCHOOL OF ENGINEERING

Visualizing distances under different weather conditions

- Distance-preserving projection from 9-dimensional space to 2-D space
- One picture for each hypothetical coefficient matrix
- On each picture, color corresponds to specific choice of weather variable



A. JAMES CLARK
SCHOOL OF ENGINEERING

Adding additional weather variables, and demand and capacity information

- Use METAR data, which gives us temperature, wind speed and direction, precipitation, and lightning (in addition to previous variables)
- Capacity is a censored variable, since it can only truly be observed when demand is high enough to trigger it
    - Model the cumulative distribution function of capacity using survival analysis
    - Use a Random Survival Forest to handle non-linear effects and the mix of continuous and discrete weather variables
- Model the influence of demand by combining an hourly demand profile (ASPM) with the capacity model in a queuing model to measure the difference in expected delay under IMC and VMC conditions

Validation of similarity engine

- Build an operational outcomes similarity matrix using data on
  - Cancellations
  - Diversions
  - Holding
  - Average delay
- Construct an aggregate weather-capacity-demand similarity matrix by taking the maximum of:
  - weather-capacity-induced distance (Euclidean distance between two CDFs)
  - demand-induced distance (difference in deterministic queuing delays)
- Validate by computing the correlation between the operational outcomes similarity matrix and the weather-capacity-demand similarity matrix

A. JAMES CLARK
SCHOOL OF ENGINEERING

Further refinements

- Use Principal Component Analysis to reduce the dimensionality of the capacity data
  - A case study at EWR suggests that 6 dimensions (instead of 61) can still capture 90% of the variance in the data
  - The principal components happen to represent meaningful chronological artifacts:
    - Mean capacity over the hour
    - Contrast between morning and evening
    - Contrast between middle of the day and other times
- PCA on demand data not found to be useful
- Clustering attempts inconclusive
  - This means that similarity is more meaningful as a continuous metric comparing two days, as opposed to a mechanism to group similar days into clusters

Is “similar historical days” even a good idea?

- Case studies suggest that for a given reference day, similar historical days can be found that exhibit a mix of levels of TMI intervention

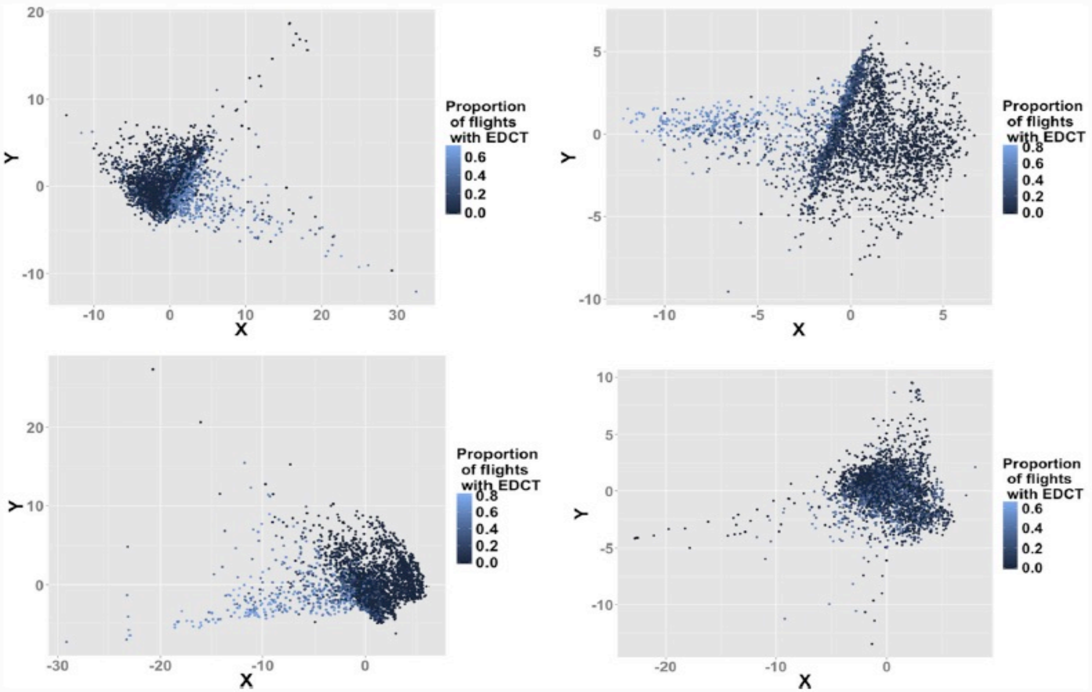


Fig. 9. MDS plot for maximum distance for EWR, SFO, ORD and JFK respectively.

A. JAMES CLARK
SCHOOL OF ENGINEERING

GDP similiarity

- Identify similar GDPs by:
  - Constituent parameters (rate, duration, etc.)
  - Actual performance (only for historical TMIs)
- Predict GDP performance if the same parameters were reproduced on the subject day
- Include both historical performance and predicted performance as fields in the "menu" of TMI choices presented to the decision-maker

Representative TMIs

- We presume that the similarity engine can generate a number of days similar to the reference day that is too large for consideration by the human decision-maker
- We want to reduce this to a smaller number of representative TMIs that can form the choice set
- We cannot do this by simple averaging or clustering because
  - Each of the results still needs to be a legitimately implementable TMI
  - This won't cover the variety of possible solutions we would like to present
  - The data don't tend to exhibit cluster structure

A. JAMES CLARK
SCHOOL OF ENGINEERING

Finding representatives

- Method 1:
  - Build a similarity graph for the set of similar days based on binary similarity
  - Trim that graph to a minimum dominating set
  - This guarantees that each data point will be similar to a representative
- Method 2:
  - Exploit a continuous similarity distance measure
  - For any threshold ϵ, we can find the ϵ-neighborhood similarity graph
  - Either fix the threshold ϵ ahead of time, or choose the number k of representatives we want (this is a human factors decision)
    - This then becomes the k-center problem
  - This guarantees that each data point will be within the distance ϵ of a representative, and that no other choice of k representatives provides a smaller distance

A. JAMES CLARK
SCHOOL OF ENGINEERING

Measuring the distance between TMIs

$$p_f(u,v) = \frac{\left|\left\{\{x,y\}: x,y \in X, \left|x_f - y_f\right| < \left|u_f - v_f\right|\right\}\right|}{\left|\{x,y\}: x,y \in X\right|}$$

This is a representation of the empirical CDF of the component difference. Once all those are known, then choose

$$d(u,v) \equiv \max_f p_f(u,v)$$

We use a small sample estimate of the empirical CDF in cases where the data set is prohibitively large

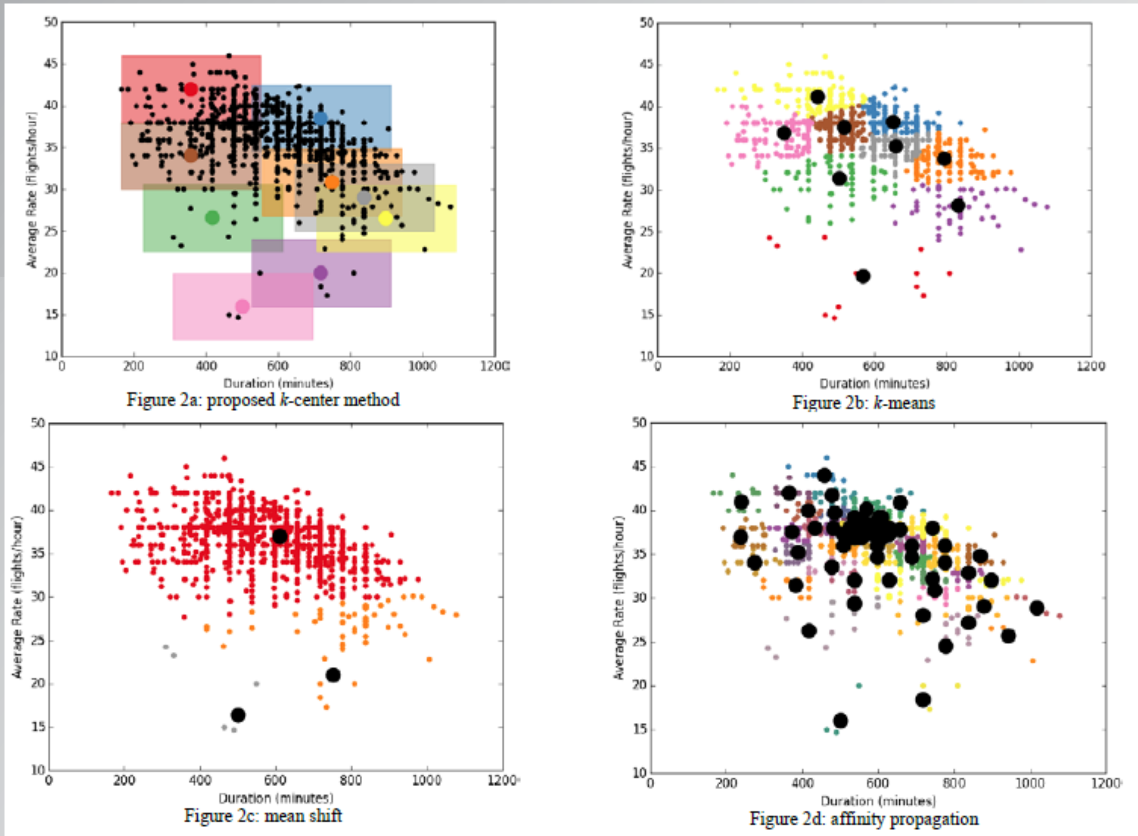
A. JAMES CLARK
SCHOOL OF ENGINEERING

Prevalence

- Having produced a set of representatives, we note that they may not all be of equal importance
- For any representative r, we define the set $S(r)$ to be the set of vertices adjacent to r in the similarity graph. Then the prevalence of r is given by $|S(r)|$.

A. JAMES CLARK
SCHOOL OF ENGINEERING

Results



- Covers the data set in a way that is not overly sensitive to local density
- Tuned to produce an appropriate number of representatives
- Representatives are original members of the data set
- Better represents the diversity of the original data set

Predicting GDP performance

- Envisioned both as a complementary feature of similar days work and as a follow-on of service level expectations work
- Statistical regression: {GDP timing parameters, scope, rate, etc.} → {performance metrics}
- Adapt geographically weighted regression (Gaussian kernel) to the distance norms from weather, traffic, and GDP parameters discussed earlier
- Regression methods: Random Forest and Gradient-Boosted Forest
- Estimate both mean and 90^{th} percentile values of average delay and number of cancellations

Results

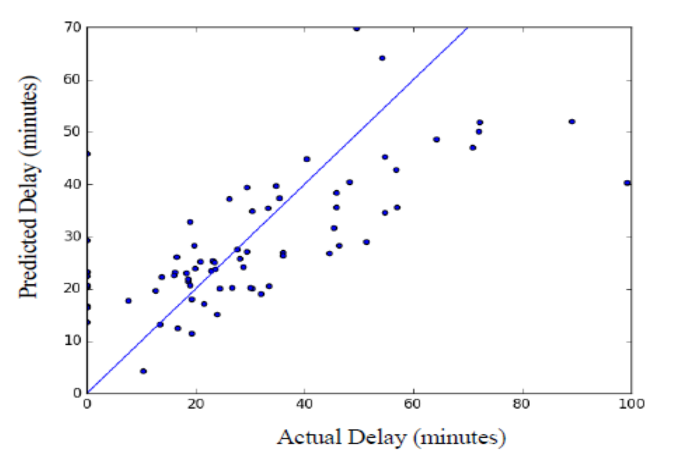


TABLE 1. RESULTS FROM ESTIMATION OF EXPECTED VALUE OF AVERAGE DELAY.

Method	Avg. Error	Improvement Over Unweighted Avg.
Unweighted Average	16.982	0.0%
Weighted Average	12.865	-24.2%
Average of *k*-Nearest Neighbors	14.139	-16.7%
Global Random Forest	11.759	-30.8%
Spatial Random Forest	11.612	-31.6%
Global Gradient-Boosted Forest	12.471	-26.6%
Spatial Gradient-Boosted Forest	12.381	-27.1%

TABLE 2. RESULTS FROM ESTIMATION OF 90% QUANTILE OF AVERAGE DELAY

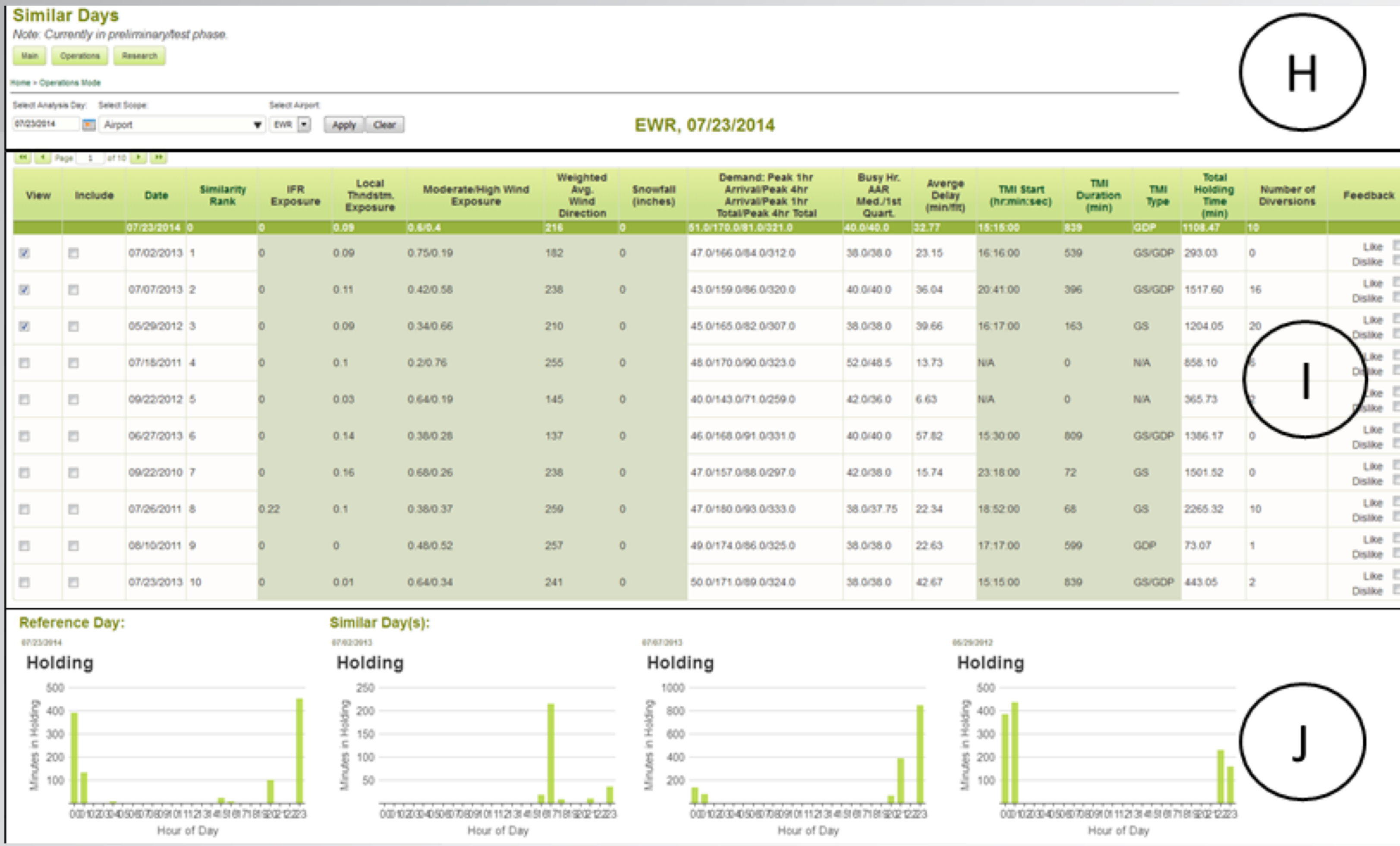
Method	Average Loss	Improvement Over Unweighted Average
Unweighted Quantile	8.589	0.0%
Weighted Quantile	7.044	-18.0%
Maximum of *k*-Nearest Neighbors	6.255	-27.2%
Global Gradient-Boosted Forest	3.356	-60.9%
Spatial Gradient-Boosted Forest	3.391	-60.5%

- Our methods do better than baseline methods chosen for comparison, but the real strength is in the prediction of quantiles
- Similar improvements were observed for predictions of the number of cancelled flights

Broader impacts: development of generalizable statistical machine learning techniques

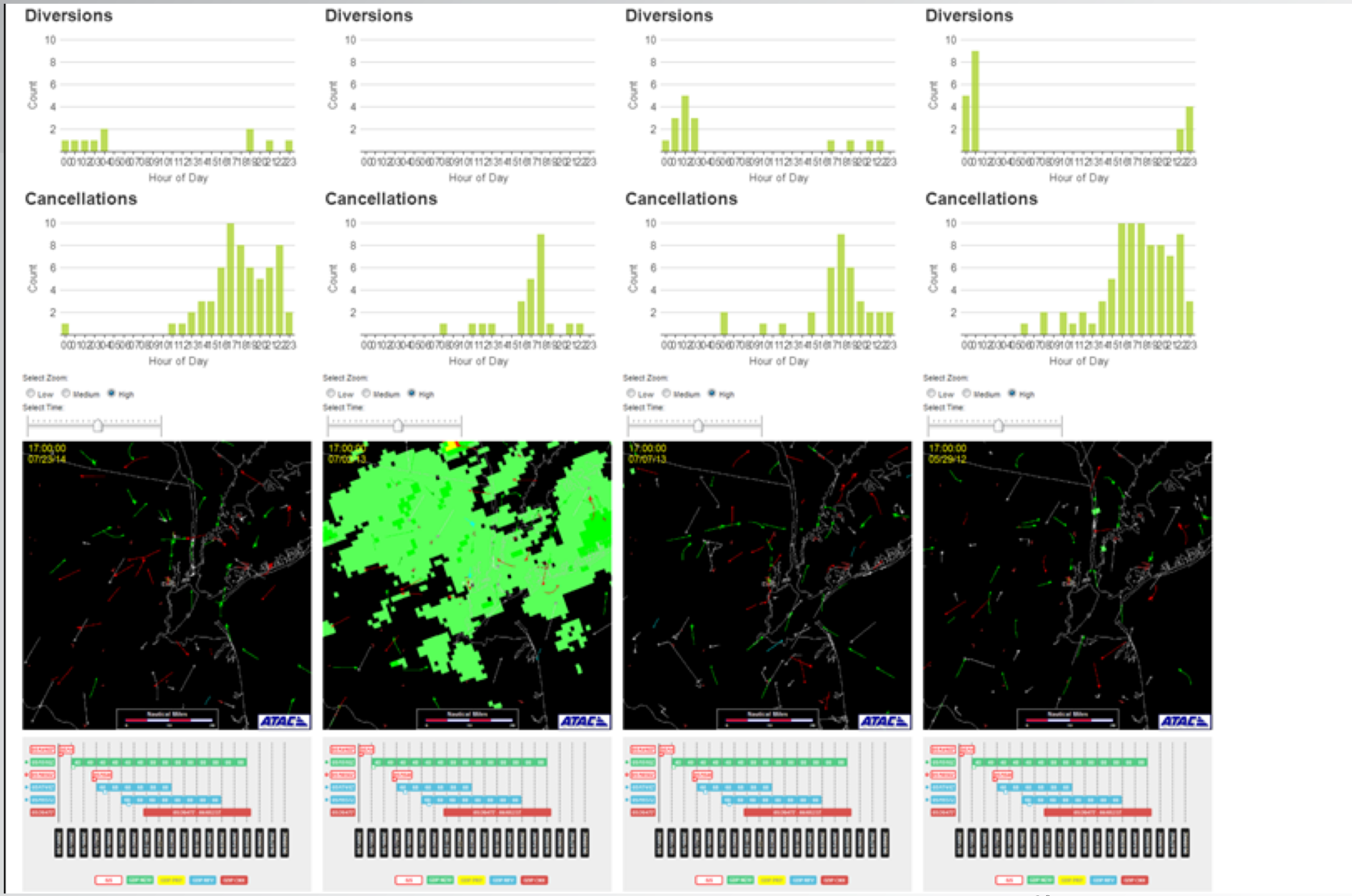
- Unsupervised prototype reduction
  - Cast the representative points problem as one of finding a small set of prototypes within a large dataset, as an unsupervised learning problem
  - Useful for a coarse description of the distribution of a multidimensional dataset, particularly in the context of human interpretation
- Representative regions
  - Generalization of points problem to a version involving regions instead
  - Also useful for data exploration by a human
  - Extensive theoretical results on axiomatic construction, convergence, coverage
  - Useful as a density estimation method

Graphical user interface



View	Include	Date	Similarity Rank	IFR Exposure	Local Thndstm. Exposure	Moderate/High Wind Exposure	Weighted Avg. Wind Direction	Snowfall (inches)	Demand: Peak 1hr Arrival/Peak 4hr Arrival/Peak 1hr Total/Peak 4hr Total	Busy Hr. AAR Med./1st Quart.	Averge Delay (min/flt)	TMI Start (hr:min:sec)	TMI Duration (min)	TMI Type	Total Holding Time (min)	Number of Diversions	Feedback
		07/23/2014	0	0	0.09	0.6/0.4	216	0	51.0/170.0/81.0/321.0	40.0/40.0	32.77	15:15:00	839	GDP	1108.47	10	
☑	☐	07/02/2013	1	0	0.09	0.75/0.19	182	0	47.0/166.0/84.0/312.0	38.0/38.0	23.15	16:16:00	539	GS/GDP	293.03	0	Like ☐ Dislike ☐
☑	☐	07/07/2013	2	0	0.11	0.42/0.58	238	0	43.0/159.0/86.0/320.0	40.0/40.0	36.04	20:41:00	396	GS/GDP	1517.60	16	Like ☐ Dislike ☐
☑	☐	05/29/2012	3	0	0.09	0.34/0.66	210	0	45.0/165.0/82.0/307.0	38.0/38.0	39.66	16:17:00	163	GS	1204.05	20	Like ☐ Dislike ☐
☐	☐	07/18/2011	4	0	0.1	0.2/0.76	255	0	48.0/170.0/90.0/323.0	52.0/48.5	13.73	N/A	0	N/A	858.10	5	Like ☐ Dislike ☐
☐	☐	09/22/2012	5	0	0.03	0.64/0.19	145	0	40.0/143.0/71.0/259.0	42.0/36.0	6.63	N/A	0	N/A	365.73	2	Like ☐ Dislike ☐
☐	☐	06/27/2013	6	0	0.14	0.38/0.28	137	0	46.0/168.0/91.0/331.0	40.0/40.0	57.82	15:30:00	809	GS/GDP	1386.17	0	Like ☐ Dislike ☐
☐	☐	09/22/2010	7	0	0.16	0.68/0.26	238	0	47.0/157.0/88.0/297.0	42.0/38.0	15.74	23:18:00	72	GS	1501.52	0	Like ☐ Dislike ☐
☐	☐	07/26/2011	8	0.22	0.1	0.38/0.37	259	0	47.0/180.0/93.0/333.0	38.0/37.75	22.34	18:52:00	68	GS	2265.32	10	Like ☐ Dislike ☐
☐	☐	08/10/2011	9	0	0	0.48/0.52	257	0	49.0/174.0/86.0/325.0	38.0/38.0	22.63	17:17:00	599	GDP	73.07	1	Like ☐ Dislike ☐
☐	☐	07/23/2013	10	0	0.01	0.64/0.34	241	0	50.0/171.0/89.0/324.0	38.0/38.0	42.67	15:15:00	839	GS/GDP	443.05	2	Like ☐ Dislike ☐

Graphical user interface



Where might this go from here?

- HITL trials
- Algorithm to automatically generate GDP/AFP parameters (Alex's recent paper)

Other future work?

- Apply a similar set of models (similarity, representative finding) to extract useful palettes of historical data to other decision-making contexts