Dynamics of Disruption and Recovery in Air Transportation Networks

Max Z. Li*, Karthik Gopalakrishnan
Hamsa Balakrishnan
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, MA, USA
{maxli, karthikg, hamsa}@mit.edu

Sang Shin, Darsh Jalan, Aritro Nandi
Lavanya Marla
Department of Industrial and Enterprise Systems Engineering
University of Illinois at Urbana-Champaign
Urbana, IL, USA
{sshin51, djalan2, aritron2, lavanyam}@illinois.edu

***Abstract*—Flight delays occur in the air transportation system when disruptive events such as weather, equipment outage, or congestion create an imbalance between system capacity and demand. These cycles of disruptions and subsequent recoveries can be viewed from a dynamical systems perspective: exogenous inputs (convective weather, airspace restrictions, etc.) *disrupt* the system, inducing delays and inefficiencies from which the system eventually *recovers*. We study these disruption and recovery cycles through a state-space representation that captures the severity and spatial impact of airport delays. In particular, using US airport delay data from 2008-2017, we first identify representative disruption and recovery cycles. These representative cycles provide insights into the common operational patterns of disruptions and recoveries in the system. We also relate these representative cycles to specific off-nominal events such as airport outages, and elucidate the differing disruption-recovery pathways for various off-nominal events. Finally, we explore temporal trends in terms of when and how the system tends to be disrupted, and the subsequent recovery.**

Keywords: Aviation disruptions; irregular operations; disruption recovery; spectral methods; flight delays

I. Introduction

The size and scale of the air transportation system render disruptions inevitable. In the US, more than 2 million flight operations were delayed in 2018, caused by everything from extreme weather to equipment outages [1]. Some examples of extremely disruptive events include the Chicago area control center fire on September 26, 2014; the March 2017 nor'easter; Hurricane Sandy and its effects on the New York-area airports; and computer outage issues at Delta Air Lines' Atlanta headquarters in August of 2016. Such disruptions, along with the associated flight delays and cancellations, result in significant monetary and environmental losses [2]. However, as with any resilient engineering system, a robust design enables swift recoveries with minimal secondary impacts. To this end, the Federal Aviation Administration (FAA) has a goal of "[achieving] 90% capacity at the top 30 airports with the most passenger activity within 24 hours [of a disruption], and 90% capacity at facilities that manage air traffic at high altitude and in the vicinity of airports within 96 hours" [3].

Disruptions and subsequent recoveries in the air transportation system vary in their geographical extent (number of airports affected), intensity (severity of resultant delays and cancellations), and duration (ranging from hours to days). The inherent variability of factors such as weather in the operating environment, along with the complex interconnectivity of the system, make it difficult to extract actionable insights from past events. Our work focuses on formalizing and analyzing disruptions and recoveries in the air transportation system by leveraging techniques that evaluate aviation disruptions from the perspective of signal processing in networked systems [4].

A. Motivation

We now discuss and motivate two questions that we will address in our work. The first relates to defining and formalizing periods of disruptions and recoveries comprehensively. A simple way to do so may be to consider the total delay as a measure of system disruptions, and then define any time interval in which delays exceed some threshold as a disrupted period. However, this approach is unable to capture spatial information regarding the geographical extent of the disruption. Suppose only one strongly-connected airport – and no other airport in the system – is experiencing high delays. The total delay metric may not classify this as a disruption, even though this scenario is unexpected, and may indicate an impending propagation of delays. Similar approaches that monitor temporal trends in delays at specific airports or origin-destination (OD) pairs are also unable to account for network connectivity-based information. Finally, in the context of extreme events such as hurricanes and nor'easters, the start of an event may not always coincide with the start of the system disruption. For example, airlines may proactively delay or cancel flights before the event, in which case the disruption precedes the event; on the other hand, airlines may opt to continue operations that progressively deteriorate, in which case the opposite order occurs. Hence, we address this question by providing a holistic method to identify both disruptions and the subsequent recovery phases, based on the magnitude and geographical distribution of delays.

The second motivation relates to understanding broader trends and patterns in historical disruption-recovery cycles in order to improve system predictability and resilience. As discussed earlier, a significant challenge to analyzing past disruptions and recoveries is their inherently large variability.

This work was supported in part by the National Science Foundation under CPS Award No. 1739505. Max Z. Li was also supported by the Wellington and Irene Loh Fund Fellowship, and an NSF Graduate Research Fellowship.

Thus, seemingly simple questions – what is the typical duration of disruptions due to specific off-nominal events; do two different events with similar delay impacts recover in different ways; is the recovery phase longer than the disruption phase; can we predict the onset and duration of the recovery phase; and so on – become very difficult to answer. Addressing these questions requires not only precise definitions for the onset and progression of disruptions and recoveries, but also the identification of "typical" or "representative" patterns that disruption-recovery cycles tend to follow.

Our main contributions lie in formalizing a framework for examining disruptions and recoveries in networks, drawing from graph signal processing and the state-space representation of dynamical systems. Specifically, we consider disruptions and recoveries not only in terms of signal magnitudes, but also their spatial distribution and temporal evolution. Our framework is of potential use to air traffic managers who might be interested in characterizing and improving the resilience of air transportation systems. Airlines and passengers may also benefit from an improved understanding of disruptions and subsequent recoveries. Finally, our work generates a reference data set of disruption-recovery cycles that can be used to benchmark system recovery (available upon request).

B. Prior work

Prior work has defined aviation system recovery in terms of flight delays, displaced or delayed crews, and disrupted or delayed passengers [5, 6]. However, these recovery definitions and strategies are airline-specific and of limited use in defining and measuring system-wide characteristics. We expand on this line of work by incorporating *system-wide spatio-temporal information* [4, 7], leading to our comprehensive definitions of disruptions and recoveries.

More broadly, analyzing disruptions and recoveries in a networked system is a growing field of study, and is closely related to understanding system resilience. Significant prior work has focused on developing models for such systems, and then analyzing them theoretically or through simulations [8, 9]. In [10], a system-level analysis was performed to model different types of synthetic disruptions and recoveries for simplified network models; they found that in simple aviation networks, there was asymmetry in the disruption and recovery timescales. Our work complements this literature by developing a framework that enables a data-driven analysis of historical disruption-recovery cycles.

Finally, while post-hoc analyses of singular off-nominal events [11, 12] and tactical disruption management strategies [13] have been explored, there have been no cohesive efforts to classify multiple types of disruption-recovery events. There also remains the lack of a formal framework for defining disruptions and recoveries in networked systems, a concern that we address in Section II.

C. Contributions

The key contributions of this paper are as follows:

1) We leverage techniques from graph signal processing to comprehensively define the start, progression, and end of flight delay disruption-recovery cycles in a network of airports. Our method not only considers the magnitude of delays, but also their spatial distribution, their relation to historical delay patterns, and temporal trends.
2) We identify *disruption-recovery trajectories* using operational data, and develop appropriate features in order to cluster them into representative groups. One of our key technical contributions is the choice of incorporating spectral graph-theoretic and temporal features to describe disruptions and recoveries in airport networks.
3) We uncover and interpret two interesting observations related to: (1) the behavior of disruption and recovery during off-nominal events (e.g., airport outages), and (2) the temporal trends in disruption-recovery trajectories.

II. Data and methodology

A. Data sources and processing

We use hourly airport delay data for the years 2008 to 2017 from the FAA's Aviation System Performance Metrics (ASPM) database for our analysis, focusing on the US Core 30 airports (see Figure 10 in the appendix for a geographical overview), which were responsible for 72% of all US enplanements in 2017 [14]. For each hour in this data set, we construct a graph with the airports as nodes, and the signal at each node being the total average arrival and departure delay experienced by all scheduled flights in that hour at the airport. The adjacency matrix for these graphs are the hourly 30×30 correlation matrices evaluated by considering the hour-by-hour subsets of the 10-year airport delay data set. Thus, there are 24 adjacency matrices, corresponding to each hour of the day. For each graph, the graph Laplacian is the difference between the degree and the adjacency matrix. With the hourly graph Laplacian, we can compute the total variation (TV) for each hour. The graph signal vector $\mathbf{x}(t) = [x_{i,t}] \in \mathbb{R}^{30\times 1}$ represents the delay $x_{i,t}$ at time t for airport i. The total delay (TD) is the 1-norm of $\mathbf{x}(t)$, and the total variation is $\mathrm{TV}\left(\mathbf{x}(t)\right) = \mathbf{x}(t)^{\intercal}\mathcal{L}\mathbf{x}(t)$, where $\mathcal{L}$ is the Laplacian for the hour-of-day of time t [4].

B. Disruption-recovery trajectories (DRTs)

We represent disruptions and the subsequent recoveries as *disruption-recovery trajectories* (DRTs) in the TV-TD state space. We define a DRT $\mathbb{T}_{t^*}$ to be a chronologically ordered set of TD and TV values, capturing the evolution of the magnitude (TD) and spatial distribution (TV) of airport delays. DRTs project the state of the system in a qualitatively interpretable manner. For example, we can assess how the system evolves in terms of airport delays from t_1 to t_2 by looking at the progression of the TV-TD state space trajectory $(\|\mathbf{x}(t_1)\|, \mathrm{TV}\left(\mathbf{x}(t_1)\right)) \rightarrow (\|\mathbf{x}(t_2)\|, \mathrm{TV}\left(\mathbf{x}(t_2)\right))$.

We further divide the TV-TD space into *regions* according to operationally-interesting regimes. We illustrate one potential (disjoint) partition that we utilize for our analysis in Figure 1(*a*). This partition distinguishes a nominal region (i.e., nominal TD and TV levels), a high-TD region (i.e., delay

magnitudes are high), and an unexpected TV region (i.e., the spatial distribution of delays is unexpected). Prior work has shown that unexpected spatial distributions correspond to regions with very high or low TV [4, 7]. In particular, given an observation $\mathbf{x}(t)$ belonging to a certain hour, we can compute the bounds on the TV for identifying delay distributions that are spatially perturbed at time t, denoted as $\left[\underline{\widehat{\Theta}}, \overline{\widehat{\Theta}}\right]_{\text{hour}(t)}$, using the methods proposed in [4, 7, 15]. We also compute a delay threshold $f_{\text{TD}}^{\text{hour}(t)} \in \mathbb{R}_{\geq 0}$ for each hour to identify delay distributions that have a very high magnitude of TD. The regions in Figure 1(a) are defined as follows:

- $\mathcal{N}$ Region (nominal): $\text{TV}\left(\mathbf{x}(t)\right) \in \left[\underline{\widehat{\Theta}}, \overline{\widehat{\Theta}}\right]_{\text{hour}(t)}$ (i.e., spatial distribution is nominal), and $\|\mathbf{x}(t)\|$ is less than $f_{\text{TD}}^{\text{hour}(t)}$ (i.e., the magnitude of delay is not abnormally high).
- $\mathcal{S}$ Region (scale): $\text{TV}\left(\mathbf{x}(t)\right) \in \left[\underline{\widehat{\Theta}}, \overline{\widehat{\Theta}}\right]_{\text{hour}(t)}$ (i.e., the spatial distribution is nominal), but $\|\mathbf{x}(t)\|$ is greater than $f_{\text{TD}}^{\text{hour}(t)}$ (i.e., the magnitude of delay is currently elevated).
- $\mathcal{D}$ Region (distribution): $\text{TV}\left(\mathbf{x}(t)\right) \notin \left[\underline{\widehat{\Theta}}, \overline{\widehat{\Theta}}\right]_{\text{hour}(t)}$ (i.e., the spatial distribution of delays is unexpected).

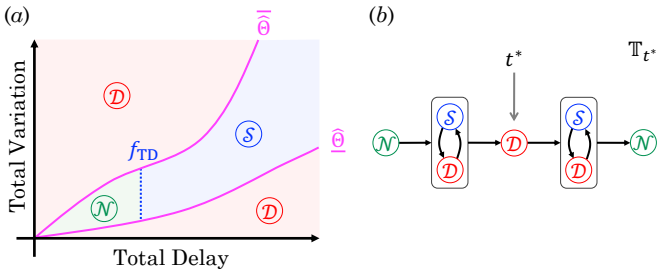

Fig. 1. (a) A disjoint partition of the TV-TD state space into three regions; (b) Schematic representation of a disruption-recovery trajectory $\mathbb{T}_{t^*}$ constructed via Algorithm 1 anchored in time at t^*.

Algorithm 1 describes a method for constructing an operationally-significant DRT $\mathbb{T}_{t^*}$, given the 3-region partition from Figure 1(a). Specifically, it *anchors* a DRT at a particular time index t^* such that $\left(\|\mathbf{x}(t^*)\|, \text{TV}\left(\mathbf{x}(t^*)\right)\right) \in \mathcal{D}$. Algorithm 1 is $O(T)$ where T is the total number of hours, and it constructs t^*-anchored DRTs forward in time. This algorithm identifies minimal-length trajectories that have at least one state in the region with unexpected spatial delay distributions, i.e., region $\mathcal{D}$, and the start and end state in region $\mathcal{N}$.

C. Nomenclature and definitions

The *state* at time t is the TD and TV of the system, i.e., $\left(\|\mathbf{x}(t)\|, \text{TV}\left(\mathbf{x}(t)\right)\right)$. We refer to a sequence of consecutive states as a *trajectory*, and a trajectory of length 2 as a *maneuver*. A DRT of length N consists of $N-1$ *maneuvers* between consecutive hours. The TV-TD state space is partitioned into *regions* $\mathcal{N}$, $\mathcal{S}$, and $\mathcal{D}$. Hence, every point on the TV-TD space (i.e., every state) belongs to one of three regions. A

Algorithm 1 Constructing DRTs given a 3-region disjoint decomposition of the TV-TD state space.

Input: Labeled states indexed by time $t \in [0 : \Delta t : T]$, where $\Delta t = 1$ hour; Region labels $\mathcal{R}(t) \in \{\mathcal{N}, \mathcal{S}, \mathcal{D}\}$
Output: Set of DRTs $\mathbb{T}_{t^*} \in \mathbb{T}$

```
t_s ← ∅; 𝕋 ← ∅; 𝟙_DRT ← FALSE
for t ∈ [0 : Δt : T] do
    if R(t) = N then
        t_s ← t
    end
    if R(t) = D ∧ t_s ≠ ∅ then
        𝟙_DRT ← TRUE
        t* ← t
    end
    if R(t) = N ∧ 𝟙_DRT = TRUE then
        𝕋_t* := {(‖x(τ)‖, TV(x(τ))) | τ ∈ [t_s, t]}
        t_s ← t
        𝟙_DRT ← FALSE
    end
end
```

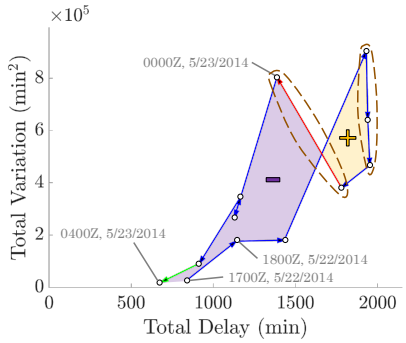

Fig. 2. A 12-hour long DRT transitions out of the nominal region at 1700Z. Arrows denote maneuvers, and their colors denote the succeeding region (see Figure 1). Select DRT features from Section II-E such as signed enclosed area and trade-off maneuvers (brown dashed indicators) are annotated. Note that each state is a one hour interval.

transition is a maneuver where the two consecutive states are in different regions. Note that a DRT is the minimal trajectory (shortest-length trajectory) that starts and ends in region $\mathcal{N}$ and contains at least one state in region $\mathcal{D}$. We further classify each maneuver into two categories (Figure 3). A *symbiotic* maneuver is one in which both TD and TV are increasing, or that both are decreasing. On the other hand, a *trade-off* maneuver is one in which the TV and TD change in opposite directions. Symbiotic maneuvers indicate pure disruptions or recoveries, whereas trade-off maneuvers are more nuanced, as one quantity is recovering at the detriment of the other. For example, a trade-off maneuver could indicate that although the system delay is decreasing, its spatial variability is increasing.

D. Disruptions and recoveries

In order to define disruptions and recoveries, we take into account the maneuver type (i.e., symbiotic versus trade-off) and whether or not a transition has occurred. For a DRT of length N, we define the *start of a disruption* to be a transition out of $\mathcal{N}$, and the *end of a recovery* to be a transition into $\mathcal{N}$. Among the remaining maneuvers, symbiotic maneuvers with

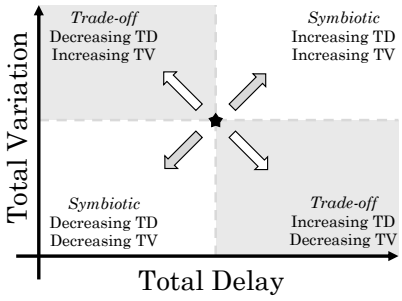

Fig. 3. Symbiotic and trade-off maneuvers in TV-TD space. The star at the center indicates the current state.

increasing TV and TD are defined as a *disruption segment*, and symbiotic maneuvers with decreasing TV and TD are defined as a *recovery segment*. Trade-off maneuvers inherit disruption or recovery classifications from the previous maneuver. The four possible trade-off maneuvers are defined as follows:

i. *Disruption-in-TV segment:* Increasing TV and decreasing TD following a disruption.
ii. *Disruption-in-TD segment:* Increasing TD and decreasing TV following a disruption.
iii. *Recovery-in-TD segment:* Increasing TV and decreasing TD following a recovery.
iv. *Recovery-in-TV segment:* Increasing TD and decreasing TV following a recovery.

Trade-off maneuvers are, by definition, a recovery along one axis and a disruption along the other. Our convention therefore assumes that a trade-off maneuver predominantly follows the trend of the preceding maneuver.

E. DRT clustering features

We cluster DRTs using twelve features that capture various operational characteristics:

DRT length: This feature, denoted by $|\mathbb{T}_{t^*}|$, captures the duration of a DRT. The minimum DRT length is 3 hours; the shortest DRT is given by $\textcircled{\mathcal{N}} \rightarrow \textcircled{\mathcal{D}} \rightarrow \textcircled{\mathcal{N}}$.

Duration in $\textcircled{\mathcal{S}}$ and $\textcircled{\mathcal{D}}$ regions: These features represent the number of hours in a DRT during which the system is either experiencing high delays (region $\textcircled{\mathcal{S}}$) or unexpected spatial delay distributions (region $\textcircled{\mathcal{D}}$). The minimum number of states in region $\textcircled{\mathcal{D}}$ is 1.

Average TD and TV intensity: For each DRT, we calculate the average TV and TD, and normalize them by their respective maximum values. The resultant features reflect the *intensity* in terms of the magnitude or spatial distribution of delay. A DRT where every hour attains a TD and/or TV value close to the maximum is considered to be more intense, with TD and/or TV intensity values close to 1.

Signed enclosed area: Figure 2 illustrates how in some DRTs, the disruption phase is characterized by higher spatial variability in the delay patterns, whereas in others, the spatial variability is higher during recovery. We use the *signed* enclosed area of a DRT as a feature:

$$\frac{1}{2}\sum \left(\|\mathbf{x}(t_{i+1})\| - \|\mathbf{x}(t_i)\|\right)\left(\mathrm{TV}\left(\mathbf{x}(t_{i+1})\right) + \mathrm{TV}\left(\mathbf{x}(t_i)\right)\right),$$

where the summation is over all maneuvers in the DRT. This computation is shown in Figure 4. If the area is negative, then more unexpected spatial delay distributions are associated with decreasing TD, whereas if the area is positive, these unexpected spatial delay distributions are associated with increasing TD. Note that this is an *aggregate* measure over an entire DRT.

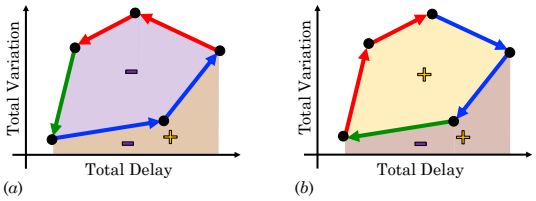

Fig. 4. A DRT where higher TV is associated with recovery (*left*), and a DRT where higher TV is associated with disruption (*right*).

Maximum TD and TV values: These features are the maximum observed TD and TV values for each DRT.

Number of symbiotic and trade-off maneuvers: These features are the counts of each type of maneuver in a DRT.

Length of symbiotic and trade-off maneuvers: The length of a maneuver is defined as the Euclidean norm of the maneuver in $\mathbb{R}^2$. We use the total length of the symbiotic and trade-off maneuvers as features that indicate the dominance of each maneuver type. For example, given a maneuver from (1, 1) to (10, 10) and a maneuver from (1, 1) to (100, 100), both are symbiotic maneuvers, but the latter is a more pronounced, dominant evolution within the TV-TD state space.

III. Representative DRTs

A. Average DRT characteristics

Algorithm 1 yields 2,322 DRTs composed of 12,350 hours (approximately 14% of all hours) within the 10-year span contained in our data set. The average length of a DRT, i.e., the average time between the system state leaving and returning to the nominal $\textcircled{\mathcal{N}}$ region, is 5.3 hours. Hence, the average duration during which the system is either in the disruption or recovery phase is 3.3 hours (subtracting the start and end nominal hours). Two of these hours are in the high-delay region $\textcircled{\mathcal{S}}$, and one hour is in the unexpected distribution region $\textcircled{\mathcal{D}}$. In other words, although most of the duration of a typical DRT involves only high magnitudes of delay (the conventional measure of a disruption), one-third of the duration is associated with the unusual spatial distribution of delay, and not necessarily its magnitude.

Recall that the TD and TV intensities measure how the values for each hour within a DRT compare to the maximum TD or TV values for that DRT. Operationally, a higher intensity indicates that both the disruption *as well as* the recovery of the system happened in a shorter time span, or in other words, most of the hours were spent close to the peak disruption state. Figure 5 shows the histograms of the average TD and TV intensities for each DRT. The distribution of $\|\mathbf{x}\|$ (i.e., the TD) is left-skewed with a mean of 0.83, whereas the distribution of

TV values is more symmetric with a mean of 0.57. The figure implies that when disruptions (and subsequently, recoveries) occur, the TD increases (and decreases) rapidly in time, but the effect on the spatial distribution of delay is more variable and evolves slowly. In other words, DRTs display faster changes along the horizontal (TD) axis than the vertical (TV) axis.

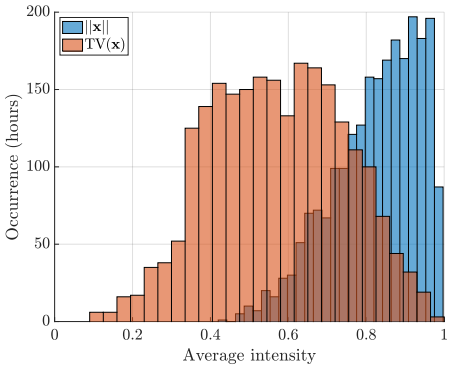

Fig. 5. Distribution of TD and TV intensity values.

Recall that the signed enclosed area of a DRT reflects whether the TV was higher during increasing delays (positive area) or decreasing delays (negative area), aggregated across the entire DRT. The average signed area is positive (6.72×10^6 min^3), indicating that the spatial distribution of delays tends to be more varied and unexpected during the disruption phase as compared to the recovery phase.

The last feature that we discuss in an average sense across all 2,322 DRTs is the number of symbiotic and trade-off maneuvers. Since the average length of a DRT is approximately 5 hours, the average number of maneuvers in the TV-TD state space is 4. Out of these, the average number of symbiotic maneuvers is 3, with 1 maneuver being a trade-off between TD and TV. Although the system prefers to evolve such that both TD and TV are increasing or decreasing, 25% of the times the system state exhibits a decrease in TD and an increase in TV, or vice versa. Since $\mathrm{TV}(\mathbf{x}) = \mathbf{x}^\intercal \mathcal{L} \mathbf{x}$, there is a positive quadratic relationship between TV and TD, indicating that the system typically will evolve symbiotically. The 25% of times where the system state exhibits trade-off maneuvers form an interesting set of airport delay behaviors, possibly reflecting the influence of external inputs (Traffic Management Initiatives or TMIs, airline recovery actions, etc.) in the disruption-recovery process.

B. Clustering DRTs

We used k-means clustering with the 12 features from Section II-E in order to determine *representative* DRTs from the set of 2,322 DRTs. While other clustering methods such as DBSCAN could be used, we chose k-means clustering for its interpretable parameter choice (i.e., number of clusters) and simplicity. Prior to clustering, we standardize all feature observations by the feature mean and standard deviation. We select $k = 7$ clusters, taking into account the within cluster sum-of-square (WCSS) error (Figure 6), and the cluster population and interpretability. Each cluster centroid provides an average representation of the DRTs that belong to that cluster. We list the centroids, along with pertinent DRT features, descriptive labels, and cluster population in Table I.

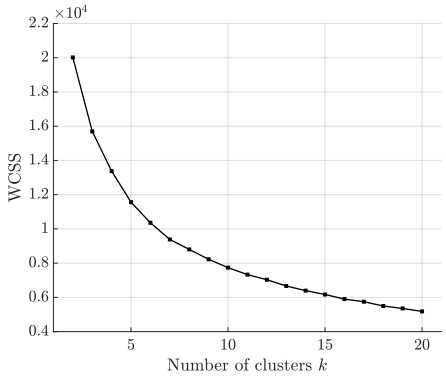

Fig. 6. Within cluster sum-of-square (WCSS) error versus the number of clusters.

C. Analyzing representative DRTs

For each representative DRT listed in Table I, we discuss the operational characteristics that describe the disruption and subsequent recovery. Furthermore, these representative DRTs help identify when a disruption begins, when the recovery begins, and when the event ends, using historical data.

We now list the representative DRTs, along with a shortened tag that we will use to refer to them:

Short DRTs with spatially-perturbed disruption segments (`Short_Dis`): This type of DRT is the most prevalent (50% of all DRTs), and is a short-duration (3 hours) event. `Short_Dis` DRTs indicate brief disruptions; for example, a transient pop-up thunderstorm around the vicinity of a major airport. In comparison to the other short representative DRT (`Short_Rec`), the average area for `Short_Dis` is positive, indicating that the airport delays were spatially distributed in a more unexpected manner during disruption than recovery. Furthermore, the maximum TV value observed for `Short_Dis` is significantly higher than `Short_Rec`, even though their maximum TD values are comparable.

Short DRTs with spatially-perturbed recovery segments (`Short_Rec`): This DRT type accounts for 33% of all DRTs. Similar to `Short_Dis`, these DRTs represent transient off-nominal conditions, with an average length of 4 hours. The average TD and TV intensity values for `Short_Rec` are smaller than those of `Short_Dis`, indicating that the system state does not typically attain the maximum TD and TV values. Furthermore, the area is negative but of the same magnitude as `Short_Dis`, meaning that the spatial distribution of airport delays was more unexpected during recovery segments than disruption segments.

Medium-length DRTs (`Med`): These DRTs have an average length of around 6 hours, indicating that these disruptions and subsequent recoveries account for significant portions of an operational day in the US airspace system. The relative rarity

TABLE I
THE SEVEN REPRESENTATIVE DRTS AND THEIR FEATURES.

DRT name	$\lvert\mathbb{T}_{t^*}\rvert$ (hours)	Region $\mathcal{S}$ (hours)	Region $\mathcal{D}$ (hours)	Avg. TD intensity	Avg. TV intensity	Area (min^3)	Max. TD (min)	Max. TV (min^2)	Symbiotic	Trade-off	Pop. (%)
Short_Dis	3	0	1	0.91	0.70	1.04×10^6	6.50×10^2	4.66×10^4	1	1	1163 (*50%*)
Short_Rec	4	1	1	0.76	0.47	-1.11×10^6	6.65×10^2	1.05×10^5	2	1	777 (*33%*)
Med	6	2	2	0.76	0.43	-5.74×10^6	1.19×10^3	5.50×10^5	4	1	196 (*8%*)
OpsDay_Dis	15	11	2	0.62	0.26	5.43×10^8	2.21×10^3	2.19×10^6	12	2	21 (*1%*)
OpsDay_Rec	18	15	1	0.68	0.35	-4.98×10^6	1.68×10^3	5.02×10^5	13	4	142 (*6%*)
MultiDay	55	49	4	0.59	0.22	2.20×10^8	2.28×10^3	1.07×10^6	40	14	22 (*1%*)
Dec08Event	229	221	6	0.57	0.23	8.46×10^8	3.09×10^3	1.21×10^6	173	55	1

of these longer-duration events are reflected in its cluster population: only 196 out of 2,322 DRTs (about 8%) are classified as Med. We also note that, similar to Short_Rec, the airport delays were spatially distributed in a more unexpected manner during recovery segments than disruption segments.

Operational day-long DRTs with spatially-perturbed disruption segments (OpsDay_Dis): With average DRT lengths of approximately 15 hours, these disruptions and subsequent recoveries account for a major portion of an operational day. For example, a DRT in OpsDay_Dis beginning in the morning would not recover back into the nominal $\mathcal{N}$ region until well into the evening. Similar to the difference between Short_Dis and Short_Rec DRTs, OpsDay_Dis and OpsDay_Rec DRTs differ by the signed enclosed area. The spatial distribution of airport delays was more unexpected during disruption segments for OpsDay_Dis DRTs. This DRT type accounts for less than 1% of all DRTs.

Operational day-long DRTs with spatially-perturbed recovery segments (OpsDay_Rec): The temporal persistence of these DRTs is similar to OpsDay_Dis, with an average length of 18 hours. As we have noted in OpsDay_Dis, the spatial delay distribution for OpsDay_Rec is more unexpected during recovery segments. Furthermore, the maximum TV value is significantly lower than OpsDay_Dis. Both OpsDay_Rec and MultiDay DRTs tend to occur in winter months, as we will discuss further in Section IV when we combine information regarding specific off-nominal events (nor'easters, hurricanes, etc.) and month-of-occurrence.

Multi-day DRTs (MultiDay): This cluster of DRTs represents a prolonged disruption and subsequent recovery event, with average lengths of over 2 days (55 hours). The maximum observed TD and TV values are also some of the highest among all clusters, indicating that these lengthy DRTs impact the system severely in terms of both magnitude and spatial distribution of delays. We also note that the spatial distribution of delays tend to be more unexpected during disruption segments for MultiDay DRTs, as signified by the positive average area. The unique characteristic of these MultiDay DRTs is that there was no recovery back to a nominal $\mathcal{N}$ region even during the overnight hours, when the system typically has low traffic and enough slack to reset the disruption.

We refer to the last cluster in Table I as Dec08Event; the fact that one unique DRT was placed in a cluster by itself indicates that it differs significantly from the other representative DRTs. Since it is a singular, extreme disruption-recovery event spanning almost 10 days (229 hours) in December 2008, we analyze it separately and present it as a case study.

In Section III-A, we saw that the average DRT was composed of 75% symbiotic maneuvers and 25% trade-off maneuvers. Even though this average does not have to hold within each representative DRT cluster, the ratio of symbiotic to trade-off maneuvers is robust to variations in the actual length of the DRT, assuming that it is long enough to observe such behavior. This indicates that even during prolonged disruption-recovery events, the preferred evolution of the system state is still in symbiotic directions, with maneuvers occurring in the trade-off direction at a frequency of only 20-25%.

IV. EVALUATING OFF-NOMINAL EVENTS AND TEMPORAL TRENDS

A. Mapping off-nominal events to DRTs

We identify the following types of *off-nominal* events using weather data and news sources: nor'easters, hurricanes, thunderstorms, and airline- or airport-specific outages. This yields a set of 178 days which are then cross-referenced with the set of DRTs, allowing us to examine what type of DRTs are common during each of these events. Figure 7 shows a normalized bar plot depicting the DRT type breakdown for each of the four off-nominal events.

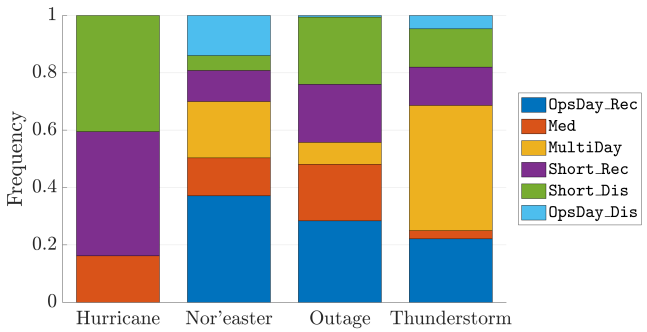

Fig. 7. Frequency of representative DRTs, given the occurrence of an off-nominal event. The extreme Dec08Event cluster is not shown.

Long-lasting DRTs (i.e., OpsDay_Rec, OpsDay_Dis, and MultiDay) are present in over 70% of nor'easter- and thunderstorm-type days, but only around 37% of days with an airline- or airport-specific outage. In particular, 37% of DRTs during nor'easters are of type OpsDay_Rec, and 44% of DRTs during thunderstorms are MultiDay. The spatial distribution of airport delays during nor'easter days, particularly during recovery segments, tends to be more unexpected than during disruption segments. This could be indicative of airline-specific recovery efforts that result in airport delays at unusual combinations of airports. Examining the average representative DRT lengths, the time it takes the system to be disrupted and recover from nor'easter-type days tends to be shorter than for thunderstorm-type days, which are dominated by MultiDay DRTs. This may be explained by the more volatile and disruptive nature of thunderstorm squall lines compared to large winter storms, resulting in more unpredictable DRTs.

63% of DRTs are Med-length or shorter on outage-type days, indicating that the disruption and recovery of the system during these events are short-lived. Similar to nor'easters, the spatial distribution of airport delays is higher during recovery segments, with 68% of DRTs having a negative area. Finally, we note that for many hurricane-type days, due to pre-emptive cancellations and airport closures, both TD and TV values are suppressed. Hence, most DRTs (84%) during hurricane-type days are short-term disruptions and recoveries.

B. Monthly distribution of DRTs

In order to observe temporal trends in DRT occurrences, we plot the frequency of occurrence of representative DRT types in Figure 8, splitting the data set into a 2008-2016 subset, and a 2017 subset. The reason for this split is that certain representative DRTs in the year 2017 behaved differently than in the preceding 9 years. Specifically, MultiDay DRTs primarily appeared only in the winter months prior to 2017. By contrast, 42%, 30% and 23% of all DRTs in April, July, and August 2017 were MultiDay DRTs. Furthermore, MultiDay DRTs in April and July 2017 are predominantly thunderstorm-type off-nominal days. This indicates an increased vulnerability of the system to thunderstorms in the summer of 2017. Further investigation would be needed to determine what specific initiatives and policies might have caused this shift in disruption-recovery dynamics in 2017.

C. December 2008 DRT: Case study

Dec08Event was an extremely long DRT (229 hours) with a sequence of disruptions and subsequent partial recoveries, occurring between December 15 and December 25, 2008. To better understand this DRT, we superimpose FAA-issued advisories related to Airspace Flow Programs (AFP), Ground Stops (GS), and Ground Delay Programs (GDP) for the duration of the Dec08Event DRT (Figure 9). The combined number of GS- and GDP-related advisories, a measure of airport capacity reductions, remained at, or above, 29 for most of the Dec08Event DRT. There was a brief drop in the number of GS- and GDP-related advisories on December 22,

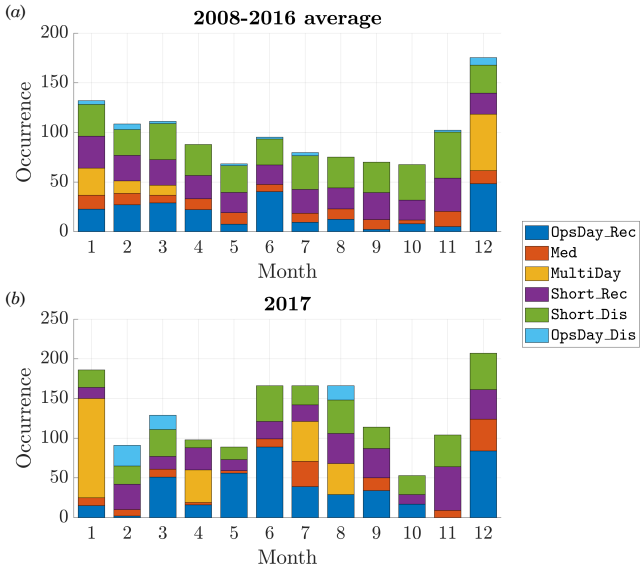

Fig. 8. Occurrence counts of DRT hours for each month, split by representative DRT clusters; counts (*a*) averaged across 2008-2016 and (*b*) for 2017.

but the continuity in the Dec08Event DRT indicates that the system was unable to return to a nominal TV-TD state before undergoing another disruption-recovery event between December 23 to December 25. The system returned to a nominal state for about 48 hours, before entering into a OpsDay_Rec-type DRT between December 27 and 28.

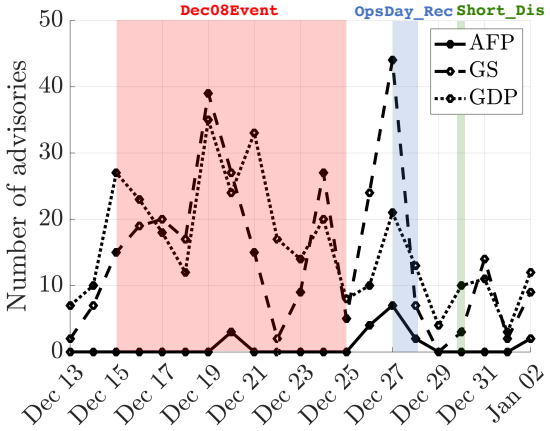

Fig. 9. Plot of AFP-, GS-, and GDP-related advisories issued by the FAA during the primary December 2008 DRT and subsequent shorter DRTs.

The Dec08Event DRT captures a series of disruptions caused by winter weather; it began with widespread ice storms throughout the US on December 11-12, and was followed by a separate, larger weather system that resulted in heavy rains in the West Coast, before transforming into a disruptive winter storm over the Midwest. The inability of the system state to return to the nominal $\mathcal{N}$ region, even during late evening and early morning periods when the system typically resets, was pronounced during this event.

V. Concluding remarks

This paper defined disruptions and subsequent recoveries using the delay magnitude (TD) and the spatial distribution (TV) of the delays, in conjunction with a low-dimensional state-space trajectory representation. We presented a partition of the TV-TD state space into various regions, representing nominal conditions, high-delay conditions, and conditions with unexpected spatial distributions of delay. We then focused on the problem of finding representative DRTs. The seven representative DRT clusters identified had interpretable characteristics in terms of lengths (i.e., the duration of disruptions and subsequent recovery), intensities, and delay behavior during the disruption or recovery segments.

The next steps will be to examine each of the individual DRTs at a more microscopic level, with a focus on time periods involving trade-off maneuvers. Doing so will help reveal whether these trade-off maneuvers correspond to the implementation of certain TMIs, driving the TD and TV values in a direction not normally traversed by them. Another direction of future research is to leverage the representative DRTs as features for predicting future system behavior, both at the system-wide and airline-specific levels.

References

[1] Bureau of Transportation Statistics, "Airline On-Time Statistics and Delay Causes," 2020.

[2] Z. A. Shavell, "Effects of schedule disruptions on the economics of airline operations," *Progress in Astronautics and Aeronautics*, vol. 193, pp. 115–126, 2001.

[3] Office of Inspector General, "Although FAA has taken steps to improve its operational contingency plane, significant work remains to mitigate the effects of major system disruptions," OIG, Tech. Rep., 2017.

[4] M. Z. Li, K. Gopalakrishnan, K. Pantoja, and H. Balakrishnan, "A spectral approach towards analyzing air traffic network disruptions," in *13th USA/Europe Air Traffic Management Research and Development Seminar*, 2019.

[5] J. Clausen, A. Larsen, J. Larsen, and N. J. Rezanova, "Disruption management in the airline industry—concepts, models and methods," *Computers & Operations Research*, vol. 37, no. 5, pp. 809–821, 2010.

[6] M. Ball, C. Barnhart, G. Nemhauser, and A. Odoni, "Air transportation: Irregular operations and control," *Hand-*

[8] M. Janić, "Modelling the resilience, friability and costs of an air transport network affected by a large-scale disruptive event," *Transportation Research Part A: Policy and Practice*, vol. 71, pp. 1 – 16, 2015.

books in operations research and management science, vol. 14, pp. 1–67, 2007.

[7] K. Gopalakrishnan, M. Z. Li, and H. Balakrishnan, "Identification of outliers in graph signals," in *58th Conference on Decision and Control (CDC)*. IEEE, 2019.

[9] B. Mukherjee, M. F. Habib, and F. Dikbiyik, "Network adaptability from disaster disruptions and cascading failures," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 230–238, 2014.

[10] K. L. Clark, U. Bhatia, E. A. Kodra, and A. R. Ganguly, "Resilience of the US national airspace system airport network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 12, pp. 3785–3794, 2018.

[11] E. J. Palin, A. A. Scaife, E. Wallace, E. C. Pope, A. Arribas, and A. Brookshaw, "Skillful seasonal forecasts of winter disruption to the UK transport system," *Journal of Applied Meteorology and Climatology*, vol. 55, no. 2, pp. 325–344, 2016.

[12] S. Kim, J. Lee, S. Oh, and Y. Yoon, "Assessment of the volcanic hazard of Mt. Paektu explosion to international air traffic using South Korean airspace," *Natural Hazards*, vol. 96, no. 2, pp. 647–667, Mar 2019.

[13] D. Suh and M. S. Ryerson, "A large neighborhood search heuristic to establish an optimal ad-hoc hubbing strategy in the wake of a large-scale airport outage," *Journal of Air Transport Management*, vol. 65, pp. 156 – 165, 2017.

[14] Federal Aviation Administration (FAA), "Aviation System Performance Metrics (ASPM) website," 2018.

[15] M. Z. Li, K. Gopalakrishnan, Y. Wang, and H. Balakrishnan, "Outlier Analysis of Airport Delay Distributions in US and China," in *IEEE Artificial Intelligence and Data Analytics in Air Transportation Conference*, February 2020.

Appendix

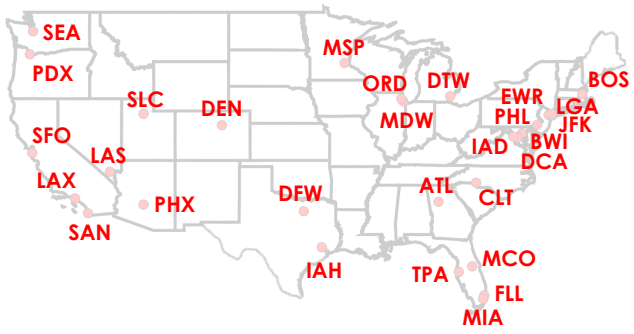

Fig. 10. Geographic locations of the US airports considered in our analysis (IATA code given). Note that HNL is not shown for simplicity.