Comparing and Clustering Ensemble Forecast Members to Support Strategic Planning in Air Traffic Flow Management

Shin-Lai (Alex) Tien[1], Christine Taylor[2] and Craig Wanke[3]
The MITRE Corporation, McLean, VA, 22102

This paper proposes a methodology for using ensemble weather forecasts to assist in air traffic flow contingency management. Specifically, the weather ensemble members are converted into scenarios of weather impact, and performance metrics are formulated to assess the similarity of these scenarios. Metrics for measuring weather impacts on both en route and terminal environments are considered in scenario clustering. Representative scenarios are selected using a proposed index, which quantifies the representativity of scenarios and addresses the requirements of representative selections. A numerical experiment is conducted using a simulation of historical traffic and weather forecast data to demonstrate the proposed methodology. It also indicates that when combining en route and terminal impact metrics, a proper weighting approach between two metric categories is needed to reflect operational preferences since their tradeoff may influence the clustering results as well as the representative selection.

I. Introduction

STRATEGIC traffic flow management (TFM) addresses predictions of significant capacity/demand imbalances four or more hours in the future. The current strategic planning process relies heavily on the mental translation of weather forecasts into traffic impact. However, at these look-ahead times, forecast uncertainty is a major challenge and requires a formal and integrated approach for predicting weather impact. Although probabilistic forecasts are available, specifying effective strategies for delay mitigation requires more explicit traffic impact information in both space and time dimensions.

In Tien et al. [4], the ensemble members from the Short Range Ensemble Forecast (SREF) product have been used to represent a wide range of deterministic weather scenarios for the en route airspace. While each of the scenarios is considered to have the same likelihood of occurrence, subsets of the members can demonstrate similar characteristics of impact which permits clustering and identification of a small number of representative weather scenarios. Such a limited but representative number of scenarios can significantly reduce the effort required by decision makers to develop mitigation strategies for each scenario.

In this study, we will extend the research scope by integrating the weather impact prediction in both en route and terminal airspace. Section II describes the translation from weather forecast variables to capacity reduction of the National Airspace System (NAS). The capacity reduction models developed respectively for en route ATC sectors and airports will be employed to facilitate the estimation of weather impact as well as the system-wide delays via a fast-time simulation tool. In Section III the performance metrics for classifying weather scenarios are explored. An ad hoc clustering algorithm as well as a representative member selection method is proposed. Section IV summarizes the numerical experiment and includes a sensitivity analysis that explores the tradeoff between en route and terminal impact. The difference between the clustering results and representative member selection will be discussed.

II. Evaluation of Weather Impact

The SREF[6] is composed of 21 87-hour deterministic forecasts and is the primary source of weather data for this study. Each ensemble member represents one trajectory of weather development through the National Airspace System (NAS) and each member is assumed equally likely to occur.

---

[1] Senior Operations Research Analyst, 7515 Colshire Drive, M/S N450.
[2] Lead Modeling/Simulation Engineer, 7515 Colshire Drive, M/S N450.
[3] Senior Principal Aviation Systems Engineer, 7515 Colshire Drive, M/S N450.

To estimate the impact on air traffic, simulation tools are often employed for quantifying NAS performance. Flow Contingency Management (FCM) developed by The MITRE Corporation is a decision support concept and prototype that integrates weather-impact forecasts with a NAS queuing network model to aid decision makers in the development of strategic plans for multiple potential outcomes of weather impact[1,2,3]. FCM estimates congestion by simulating the propagation of demand through the network, subject to capacity reduction of NAS resources due to weather or other constraints.

To leverage FCM's queuing network model for simulating weather impact, the 21 SREF forecasts need to be translated into capacity reduction of Air Traffic Control (ATC) sectors and airports. For quantifying the capacity loss on ATC sectors from SREF, Tien et al. [4] constructed a functional relationship between the reduction rate of sector capacity and the hourly precipitation, which is one of the SREF forecast variables. The SREF weather grids are matched with and aggregated by ATC sectors, so the gridded precipitation data can be converted into a sector-based format to predict a capacity reduction ratio. The loss of capacity of an ATC sector would be the reduction ratio multiplied by the nominal capacity estimated by FCM.

Predicting airport capacity at a strategic look-ahead time is a challenging task. Weather forecasts of wind, ceiling, visibility, and convective activities may jointly determine runway configuration and the associated arrival rate (AAR) and departure rate (ADR) for the airport. To estimate the impact from weather forecast, a translation from forecast variables to available airport capacity must be built. Currently we are developing a prototype that accounts for critical SREF variables as well as local characteristics for predicting airport-specific AAR and ADR. Such a prototype will be applied in our future work for modeling terminal weather impact.

In this study we employ a simplistic approach of deriving airport capacity for the purpose of concept validation: For each airport considered, the precipitation forecast (mm/hr) of the weather grid cell in which it resides is converted into radar reflectivity (dBZ) using Eq. (1)[10]. Then, we use Table 1 for categorizing the impact and determining the capacity reduction ratio. The loss of capacity of an airport can thus be derived from multiplying the predicted reduction ratio with the nominal arrival and departure capacity assumed in FCM.

With the estimated capacity of sectors and airports based on SREF forecasts, 21 weather-impact scenarios can be generated in FCM, each of which are provided with delay statistics for ATC sectors and airports by time. These scenarios will allow quantitative evaluation for comparison and clustering.

$$dBZ = 10\left(\log_{10} 200 + \frac{8}{5}\log_{10} Precip\right) \qquad \text{Eq. (1)}$$

Table 1 Terminal Impact Categorization Table

Reflectivity (dBZ)	**Description**	**Reduction Rate**
Below 30	Light	0 %
30 ~ 40	Moderate	20 %
40 ~ 50	Heavy	40 %
Above 50	Extreme	60 %

III. Clustering Methodology

This section describes a series of steps employed for evaluating weather impacts, clustering weather-impact scenarios, and selecting representative scenarios.

A. Impact Metrics for En Route ATC Sectors

After FCM generates 21 SREF weather-impact scenarios, proper aggregation of spatiotemporal delay data is needed to facilitate scenario comparison. For ATC sectors, the idea is to summarize the spatiotemporal delays with two aggregate values: spatially weighted delay and temporally weighted delay. Let $e(s,t)$ denote the delay value associated with sector s at time t, where $s \in \{1, \dots, S\}, t \in \{1, \dots, T\}$. Two sets of summary metrics are first defined to capture the impact in the spatial and temporal dimensions.

- Spatial summary metric: $SI(s) = \sum_{t=1}^{T} e(s,t)$, and
- Temporal summary metric: $TI(t) = \sum_{s=1}^{S} e(s,t)$.

The next step is to aggregate the summary metrics by time and sectors. The spatial summary metric SI renders a $1 \times S$ row vector, corresponding to S sectors in the weather forecast region. To incorporate the geographical relationship of sectors into metric aggregation, a summation technique is proposed:

$$Adjacency\ Weighted\ Summation = \sqrt{V \cdot\ (\boldsymbol{\delta} + \boldsymbol{I}) \cdot V^T}$$

where V is a $1 \times S$ row vector, whose transpose is V^T, $\boldsymbol{I}$ is a $S \times S$ identity matrix, and δ is the first-order $S \times S$ adjacency matrix of sector with zeros in the diagonal.

Similarly, time series data could also be weighted by “adjacency”, which needs to be defined. The temporal summary metric TI renders a $1 \times T$ row vector, corresponding to T time periods over the weather forecast horizon. The adjacency matrix δ for TI is a $T \times T$ matrix such that:

$$\boldsymbol{\delta} = \left\{ \delta_{i,j} \left| \begin{matrix} 1, \text{if j} = \text{i} + 1 \text{ or j} = \text{i} - 1 \text{ for } i, j = 1, \dots, T. \\ 0, otherwise. \end{matrix} \right. \right\}$$

Spatially and temporally weighted delay metrics can thus be computed for a spatiotemporal profile of en route delays and used for scenario comparison. More detailed discussion on metric characteristics can be found in Tien et al. [4].

B. Impact Metrics for Terminal

In FCM, airport delay statistics are collected for both departure and arrival operations at the boundaries of terminal airspace. Thus, for each airport, there is a temporal profile for departure delays as well as for arrival delays. The temporally weighted metric introduced in III.A can used to summarize these two delay profiles in order to facilitate scenario comparison.

C. Clustering Weather-Impact Scenarios

Enriquez and Kurcz [7] develop an iterative version of the Spectral Clustering algorithm that was used to cluster subjects that represent 2D aircraft trajectories. The subjects are first evaluated by pair-wise “similarity” defined by Euclidean norm and a Gaussian kernel and then iteratively partitioned into two groups until the stopping criteria are reached. Thus, there is no need to pre-specify a desired number of clusters.

Specifically, assuming that there are N subjects to be clustered, the main input of the algorithm is a “similarity” matrix $\boldsymbol{W} = \{w_{i,j} |\ i, j = 1, \dots, N\}$, which is typically constructed with a Gaussian kernel:

$$w_{i,j} = \exp\left(\frac{\|x_i - x_j\|}{-2\sigma^2}\right) \tag{5}$$

where x_i is the metric vector for an individual subject, for $i = 1, \dots, N$, and $\|\cdot\|$ is the Euclidean norm. This ensures that if two subjects are dissimilar in their metric values, their $w_{i,j}$ will be small. The scale parameter σ in Eq. (5) determines the width of the neighborhood and thus plays a critical role in computing similarity.

Tien et al. [4] tailored the algorithm for the requirements of FCM by determining the scale parameter σ based on the input data, as opposed to using a pre-defined parameter value. As a result, the algorithm is more adaptive to the magnitude of input data and avoids the need to tune the scale parameter whenever there is a change of metric or a change of weather day, which is desirable for real-time application as weather impact intensity varies day by day.

With the proposed impact metrics for both en route and terminal airspace, the clustering algorithm in Tien et al. [4] can be applied. Scenarios with similar characteristics will be identified so that the representative scenarios could be determined.

D. Selection of Representative Scenarios

After obtaining the clustering results, a representative scenario from each group may be identified with both subjective and objective judgments. For the TFM purpose, the representative scenarios should demonstrate some of the characteristics proposed in Xue et al. [8]:

1. Each representative scenario should correspond to an original SREF member.
2. Each representative scenario should be significantly different from other representative scenarios.
3. The representative scenarios together should adequately span all the scenarios.
4. The probability associated with each representative scenario indicates the fraction of representation in the ensemble by the associated cluster.

Molteni et al. [9] introduced the concept of representativity index (*RI*) for objectively selecting the representative scenarios from clusters of scenarios. For this study, the RI of scenario i is reformulated as follows:

$$RI_i = \frac{\sum_{j \notin C} w_{i,j} / (N - |\boldsymbol{C}|)}{\sum_{j \in C} w_{i,j} / |\boldsymbol{C}|}$$

where $w_{i,j}$ is the similarity metric between scenarios i and j defined in Eq.(5), $\boldsymbol{C}$ is the set of scenario indices in the same cluster as i, and N is the total number of scenarios.

The proposed *RI* of a scenario takes into account the similarity measurement inside and outside of a cluster. It is the ratio between its average similarity metrics from the scenarios of all other clusters and that from the scenarios in its own cluster. As a high $w_{i,j}$ means high similarity between two scenarios, a scenario with a low *RI* value indicates that its similarity to those in the same cluster is high, and its similarity to those out of the cluster is low.

A representative scenario of a cluster is thus defined as the one with the smallest *RI*, and its probability of occurrence is the percentage of the clustered scenarios to the total. This definition also meets the four characteristics listed above. Specifically, the smallest *RI* ensures that a representative scenario is significantly different from other representative scenarios, and together the representative scenarios span the entire range of scenarios.

IV. Numerical Results

A. Source of Data

To demonstrate the proposed methodology, the weather forecast and traffic demand on June 18, 2013 were analyzed: For each of the 21 SREF members, the hourly precipitation forecast, generated at 09:00Z, from 10:00Z to 08:00Z next day is used to generate capacity loss of sectors and airports. Traffic demands were based on the first filed flight plans and simulated via the FCM's queuing network model under forecast weather constraints. Figure 1 illustrates the historical weather on this day that primarily developed over the airspace of the ARTCCs of Cleveland (ZOB), Chicago (ZAU), Indianapolis (ZID), New York (ZNY), Boston (ZBW), Washington (ZDC), Atlanta (ZTL) and Jacksonville (ZJX). We will use FCM's simulation delays of sectors and airports in this airspace for computing the impact metrics.

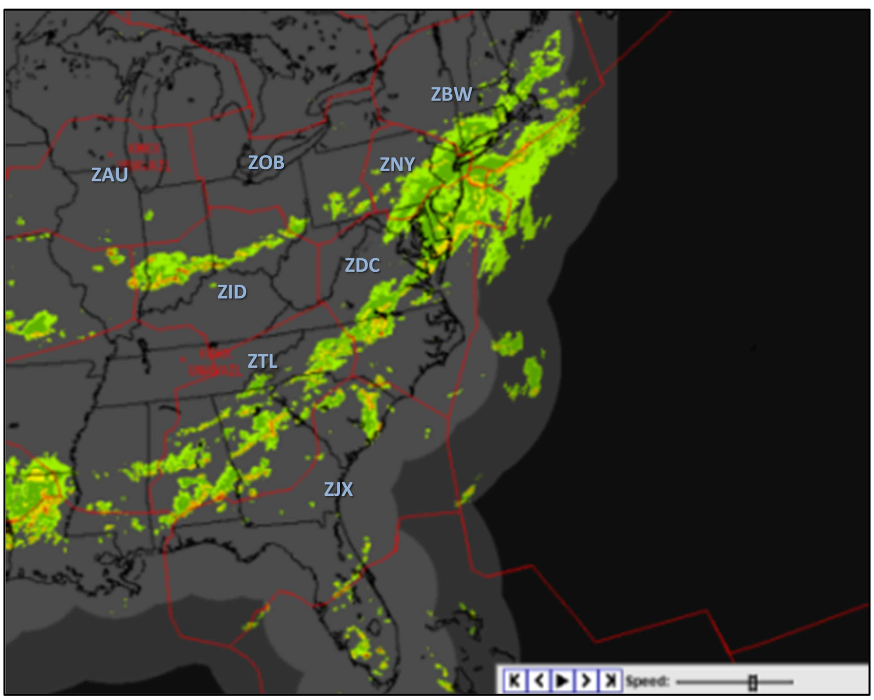


Figure 1. CIWS Snapshot of Historical Weather at June 18, 2013 22:00

The simulation delays produced by FCM are the backlog statistics from the underlying queuing model, which can be interpreted as potential congestion workload. Figure 2 shows the spatiotemporal distributions of sector delays from four selected scenarios. Visual inspection of all the scenarios indicates that the ZDC sectors are most heavily impacted and that the impact becomes more severe later in the day. However, the coverage and magnitude of impact differ among scenarios.

The impact metrics proposed in III.A can help numerically differentiate the impact by time or by space for systematically comparing multiple scenarios when visual inspection becomes unmanageable. To prepare the data for running the clustering algorithm, a temporal impact metric and a spatial impact metric are computed for sectors delays while only five impact metrics are selected for airport delays because of their significant variation among 21 scenarios. The selected airport delay metrics are LGA's arrival delay, EWR's departure delay, EWR's arrival delay, ATL's departure delay, and ORD's arrival delay.

The radar plot in Figure 3 illustrates the impact metrics for all 21 SREF scenarios, where each metric is normalized by its maximum value and mapped onto a scale of 0 to 100. Such a normalization step is necessary as the values of the various metrics can be drastically different. For example, in this case study, the temporal sector delay metric, which is derived from 378 sectors, could have a much higher value than the arrival delay metric for LGA. With the original scale, LGA arrival delays might not be as significant as sector delays, resulting in no influence on scenario clustering, while in fact LGA arrival delays would not only distinguish scenarios but also signal the need for a ground delay program for LGA.

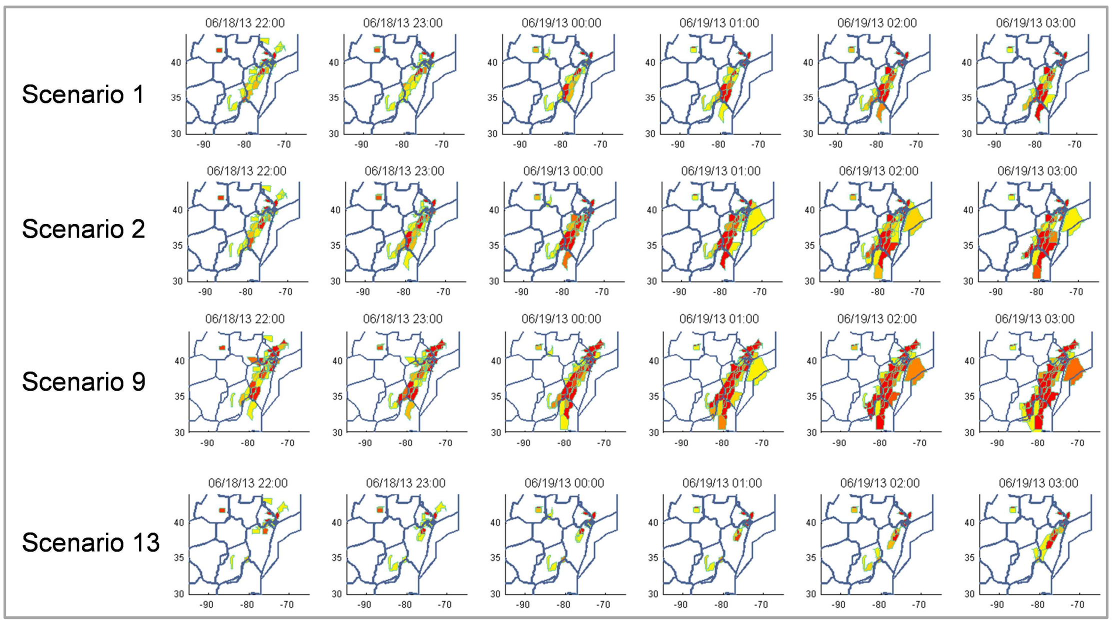


Figure 2. En Route Delays from FCM Simulation for Selected Weather Scenarios

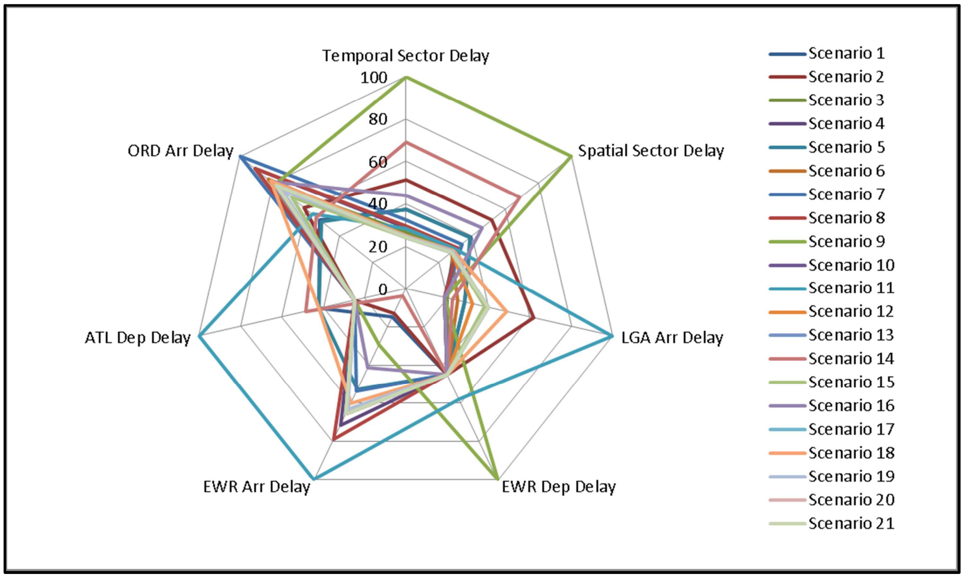


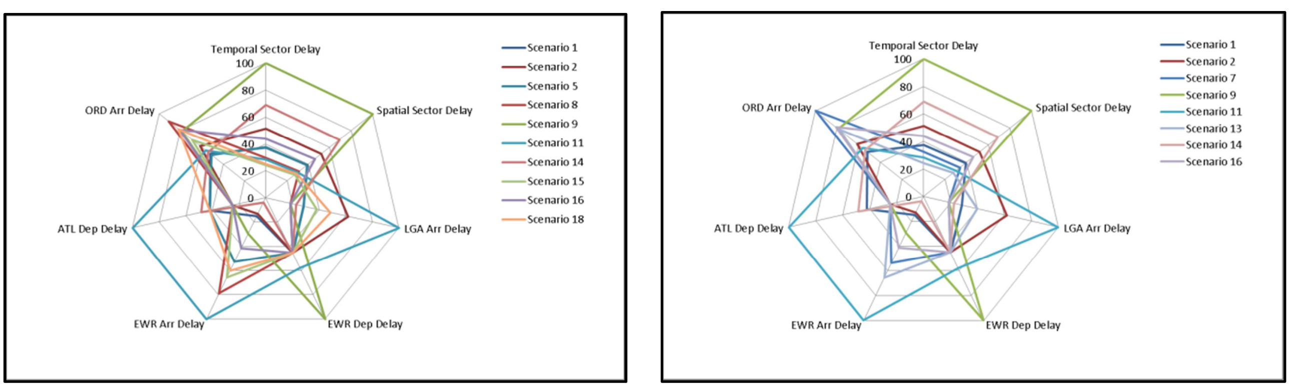
Figure 3. Visualization of Clustering Data Prepared for June 18, 2013

B. Clustering Results

Each weather-impact scenario is evaluated with 2 sector delay metrics and 5 airport delay metrics. To address the differences between sector delays and airport delays, we apply a weighting factor S/A on sector delays when computing similarity using Eq. (5). Operational preferences could be reflected in this tradeoff coefficient as well. The S/A value represents the degree to which sector delays are weighted more heavily than airport delays, where $S/A = 1$ means they are equally weighted, $S/A = 2$ means sector delays are weighted twice as much as airport delays.

With $S/A = 1$, there are 10 clusters and thus10 representative scenarios identified, as illustrated in Figure 4(a). Thus by clustering the 21 member ensemble in this manner, a significant reduction in weather futures is achieved. This reduction allows decision makers to focus only on distinctive but representative scenarios, as opposed to all 21 ensemble members.

It is expected that increasing the weighting on sector delays would result in fewer clusters as differences in airport delays become less significant. With $S/A = 8$, there are 8 representative scenarios identified, illustrated in Figure 4(b). It can be observed that all the distinctive contours regarding sector delays still exist while some representatives in Figure 4(a) are now either clustered with other scenarios or are no longer identified as the representative scenario.



(a) 10 representative scenarios when $S/A = 1$.

(b) 8 representative scenarios when $S/A = 8$.

Figure 4. Visualization of Representative Scenarios

C. Sensitivity of Sector-Airport Delay Tradeoff

To understand the sensitivity of the tradeoff coefficient to the cluttering results, we evaluated how changing the S/A value changes the specific clusters and representative members. Figure 5 depicts the evolution of clustering results as S/A increases. As expected, increasing the weight on sector delay metrics changes the number of the resulting clusters as well as the representative scenarios.

When S/A changes from 1 to 2, the cluster of {4, 7, 8} changes its representative from Scenario 8 to Scenario 7, as shown in Figure 6(a). In addition, Scenario 18 is included in another cluster and results in a change of representative, from Scenario 15 to Scenario 13, shown in Figure 6(b).

Figure 6(c) shows that when $S/A = 4$, the difference of EWR arrival delays between Scenarios 1 and 5 is not as influential as a lower S/A. Thus, two scenarios are combined in the same cluster.

After S/A is above 4, there is no change in clusters and representative selection.

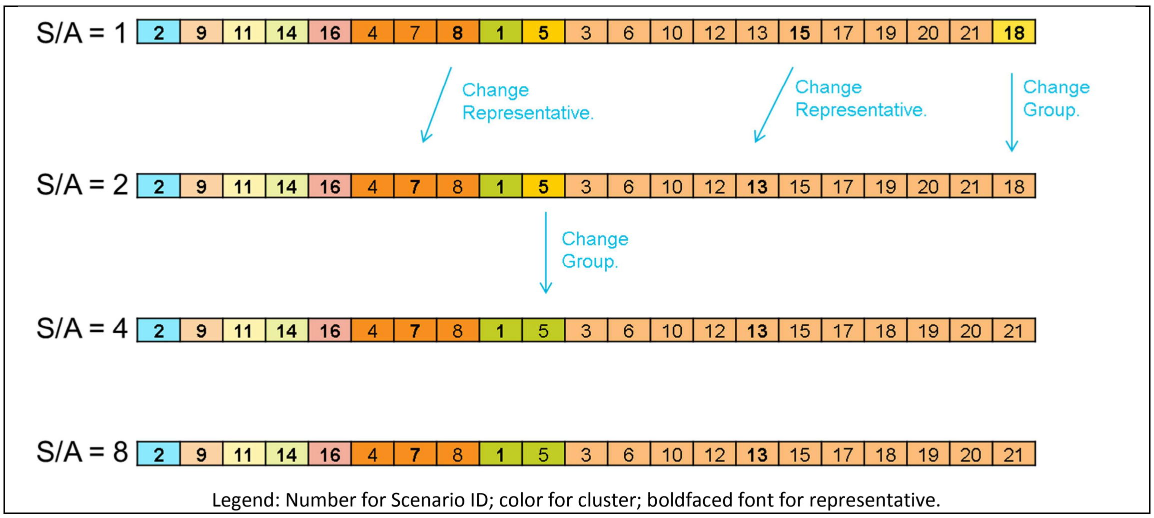
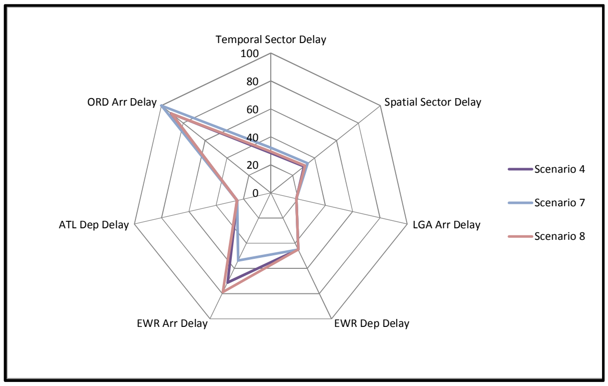
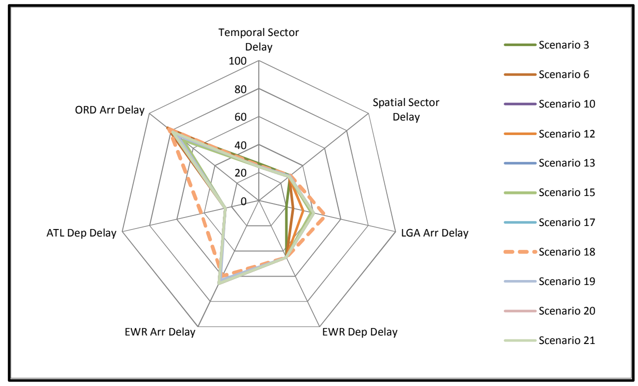


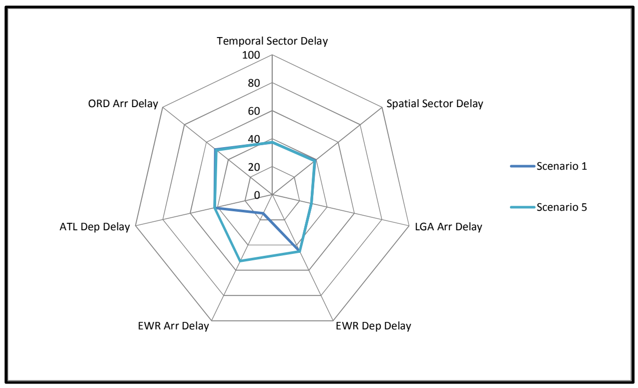
Figure 5. Sensitivity of S/A to the Clustering Results



(a) Representative changed from 8 to 7.



(b) Representative changed from 15 to 13 after Scenario 18 joined the cluster.



(c) Weighting more on sector delays results in less significance in airport delays.

Figure 6. Changes of Representativity within Individual Clusters

V. Summary

The ensemble members from the Short Range Ensemble Forecast product can be used to represent a wide range of deterministic weather scenarios for NAS strategic planning. The aggregation method for spatiotemporal data is employed to facilitate numerical comparison among scenarios. With the proposed representativity index (*RI*), representative scenarios that meet preferred characteristics can be selected after clustering. In the numerical experiment, sector and airport delays are incorporated into the proposed clustering approach, but the tradeoff between the two delay categories may influence the clustering results as well as the representative selection.

For future work, impact modeling for airport capacity needs to be improved for real-time decision support. Also, the performance of the clustering results could be examined with the traffic management initiatives designed specifically for the representative scenarios. It is expected that the initiatives that work best for the representative scenarios would also have similar response on the scenarios in the same cluster. Lastly, more weather days can be analyzed to fine-tune model parameters as well as understand the scale of historical weather severity.

Acknowledgments

The authors thank Dr. David Bright and Ben Schwedler at NOAA Aviation Weather Center for their suggestions and support on SREF data collection and processing, Dr. Sandip Roy of Washington State University and Dr. Yan Wan of North Texas University for their constructive comments, and Dr. Marco Enriquez of The MITRE Corporation for his insights on the clustering method. The authors would like to thank Tudor Masek, Rachel Ethier, John Huhn, Mary Hokit, Scott Mills, Matt Fronzak of The MITRE Corporation for their significant contributions in FCM concept and software development, Jonathan Hammer, Mike Klinker, Dr. Lixia Song, Joe Hollenberg of The MITRE Corporation for their valuable comments and support, and Kevin Workman, Dave Chaloux, James Woodside, Ali Rakei of QinetiQ North America for software development and integration.

Notice

References

[1]Federal Aviation Administration, *NextGen Mid-Term Concept of Operations for the National Airspace System*, Washington, DC, 2009.

[2]Taylor, C., Wanke, C., Wan, Y., and Roy, S., "An Operational Concept for Flow Contingency Management," *AIAA Aviation Technology, Integration, and Operations Conference*, Virginia Beach, VA, September, 2011.

[3]Taylor, C., Wanke, C., Wan, Y., and Roy, S., "A Decision Support System for Flow Contingency Management," *AIAA Guidance, Navigation, and Control Conference*, Minneapolis, MN, August, 2012.

[4]Tien, S., Taylor, C., and Wanke, C., "Identifying Representative Weather-Impact Scenarios for Flow Contingency Management," *2013 Aviation Technology, Integration, and Operations Conference*, August, 2013.

[5]Grabbe, S., Sridhar, B., and Mukherjee, A., "Similar Days in the NAS: an Airport Perspective," *2013 Aviation Technology, Integration, and Operations Conference*, August, 2013.

[6]Du, J., DiMego, G., Toth, Z., Jovic, D., Zhou, B., Zhu, J., Chuang, H., Wang, J., Juang, H., Rogers, E., and Lin, Y., "NCEP Short-Range Ensemble Forecast (SREF) System Upgrade in 2009," *The 19th Conf. on Numerical Weather Prediction and the 23rd Conf. on Weather Analysis and Forecasting*, Omaha, Nebraska, June 1-5, 2009.

[7]Enriquez, M. and Kurcz, C., "A Simple and Robust Flow Detection Algorithm Based on Spectral Clustering," *International Conference on Research in Air Transportation (ICRAT)*, Berkeley, CA, 2012.

[8]Xue, M., Roy, S., Zobell, S., Wan, Y., Taylor, C., and Wanke, C., "A Stochastic Spatiotemporal Weather-Impact Simulator: Representative Scenario Selection," *AIAA Aviation Technology, Integration, and Operations Conference*, Virginia Beach, VA, September, 2011.

[9]Molteni, F., Buizza, R., Marsigli, C., Montani, A., Nerozzi, F., and Paccagnella, T., "A Strategy for High-Resolution Ensemble Prediction. I: Definition of Representative Members and Global-Model Experiments," *Quarterly Journal of the Royal Meteorological Society*, 127, pp. 2029-2094, 2001.

[10]Raghavan, S., Radar Meteorology: Issue 27 of Atmospheric and Oceanographic Sciences Library, Springer, 2003, pp. 267.