

Spatiotemporal Scenario Data-Driven Decision-Making Framework for Strategic Air Traffic Flow Management

Wen Zhang¹, Junfei Xie¹ and Yan Wan²

Abstract— With the growth of air traffic demand, the recent few years have witnessed increasingly frequent traffic delays. The strategic air traffic flow management (ATFM) aims to resolve this issue by planning flows at a long look-ahead time frame for better resource allocation. Existing studies on strategic ATFM have been focused on modeling and analyzing complicated flow dynamics and designing effective TMIs robust to uncertainties. Little effort has been made to overcome the challenge of how to quickly design optimal TMIs for air traffic systems of large state and decision spaces. In this paper, we introduce a novel spatiotemporal scenario data-driven decision-making framework, which conquers this challenge by leveraging historical TMIs for spatiotemporal weather-impact scenarios similar to the current scenario under evaluation. By moving most computations to offline and limiting online computations to fine-tuning of the control parameters in the historical TMIs, this framework significantly expedites the design speed and makes real-time decision-making for large-scale dynamical systems possible. The simulation results demonstrate the effectiveness and efficiency of the proposed framework.

I. INTRODUCTION

With the increase of air traffic demand, the capacity of the National Airspace System (NAS) is becoming saturated. According to the Federal Aviation Administration (FAA), 69% of system impacting delays are caused by weather and 19% of them are caused by volume between 2008 and 2013 [1]. To reduce traffic delays, the *strategic* air traffic flow management (ATFM) was proposed [2], which plans traffic flows 2-15 hours in advance and is considered a key component of the Next-Generation (NextGen) Air Transportation System. Planning at such a long look-ahead time frame, compared to the 0-2 hour *tactical* time frame, is advantageous in that it allows management of traffic flows over a larger spatial scale for better resource allocation [3]. However, the *large state and decision spaces*, *complicated flow dynamics* and existence of *a wide range of uncertainties* (e.g., convective weather and uncertain traffic demand) also make strategic ATFM very challenging.

Despite the abundant works in tactical air traffic management [4], studies on strategic ATFM are limited. As an initial step towards management design, stochastic queuing network models [3], [5], [6] were recently developed to capture the dynamics of air traffic flows under uncertain weather and management actions. These models are suitable for strategic

decision-making as they regard air traffic as aggregated flows rather than individual aircraft, whose scheduling details are less meaningful in the strategic time frame.

Besides system modeling, there have also been some efforts on evaluating the impact of uncertain weather on air traffic flows and designing strategic traffic management initiatives (TMIs) robust to weather uncertainties [7]–[14]. For instance, articles [3], [14] approximate the nonlinear air traffic flow model as a tractable jump-linear model and derive simple mathematical equations to measure system performance under uncertain weather. Based on this study, a sensitivity-based approach [10] was then developed to design optimal TMIs. This *analytical* approach allows analyzing and designing TMIs at a low computational cost, but is only applicable when tractable system models are available. For large-scale systems of complicated dynamics that can be hardly captured by tractable models, *simulation-based* approaches have been developed [9], [11], [12], [15], which select a set of samples of the uncertainty space and design TMIs optimal to these samples. For instance, the M-PCM-OFFD based approach introduced in [11], [15] allows the design of optimal TMIs to only consider a small set of samples, while retaining the optimality to the whole uncertainty space.

We note that existing studies on strategic ATFM, such as those mentioned above, have been focused on addressing the challenges of how to model complicated flow dynamics and how to analyze and design TMIs robust to uncertainties. How to conquer the challenge of large state and decision spaces has been rarely considered. In this paper, we aim to address this challenge to enable *real-time* decision-making for strategic ATFM.

In order to significantly expedite the design speed, our idea is to move most computations to offline and limit online computations to fine-tuning of the control parameters only. This can be achieved by leveraging the big-data techniques. In particular, under the assumption that similar spatiotemporal weather-impact scenarios lead to similar TMIs, we can make plans based on historical TMIs for scenarios similar to the current scenario, instead of designing a new TMI from scratch. This spatiotemporal scenario data-driven decision-making framework, first mentioned in our previous paper [11], is not easy to realize. It not only requires an effective system model to capture flow dynamics, an approach to design optimal TMIs and a database to store historical weather-impact scenarios and associated TMIs, but also an efficient query system to retrieve similar weather-impact scenarios. In the past few years, we have developed each

*This work was supported by the National Science Foundation under Grant 1839707 and 1839804.

¹Wen Zhang and Junfei Xie are with the Department of Computing Sciences, Texas A&M University-Corpus Christi, Corpus Christi, TX, 78412 junfei.xie@tamucc.edu

²Yan Wan is with the Department of Electrical Engineering, University of Texas at Arlington, Arlington, TX, 76010 yan.wan@uta.edu

component for this framework [3], [5], [11], [16], [17]. In this study, we integrate these components and realize the proposed framework for the first time. Various comparative simulation studies are also conducted, which verify the proposed framework and demonstrate its effectiveness and efficiency. Notably, this spatiotemporal scenario data-driven decision-making framework is not only suitable for strategic ATFM, but also applicable for many other large-scale dynamical systems, such as ground transportation systems, power networks, and complex information systems.

In the rest of the paper, we first briefly introduce an air traffic flow model and a weather-impact model in Section II, which serve as the evaluation and design foundation for this study. We then formulate the problem in the same section. In Section III, we describe in detail the spatiotemporal scenario data-driven decision-making framework for strategic ATFM. Extensive simulation studies are then conducted in Section IV to verify the proposed framework. Section V finally concludes the paper with discussions on future works.

II. SYSTEM MODELING AND PROBLEM FORMULATION

A. Air Traffic Flow Model

The stochastic queuing network model developed in our previous study [5] captures air traffic flow dynamics in the strategic time frame. It regards air traffic as aggregated flows and describes these flows as stochastic processes. To capture the routing of flows, it models the air traffic system as multiple overlapped sub-networks, distinguished by the origin and destination airports. In each sub-network, flows initiating from the origin choose a route and traverse through a set of *merging/splitting* points to the destination. As the NAS is partitioned into regions, the directed routes intersect with region boundaries at the *boundary intersection points*.

Let us now describe the mathematical formulation of this flow-based queuing network model. Consider flows from origin o to destination d , at each merging/splitting point j , their dynamics can be captured by following equations

$$\begin{aligned} g_{odjl}[k] &= \rho_{odjl} \sum_i f_{odij}[k] \\ \sum_l \rho_{odjl} &= 1, \quad \rho_{odjl} \in [0, 1] \end{aligned} \quad (1)$$

where $f_{odij}[k]$ and $g_{odjl}[k]$ represent the number of aircraft entering node j from node i and the number of aircraft leaving from node j to node l in sub-network od at time k , respectively. ρ_{odjl} denotes the fraction of flows leaving from node j to node l in sub-network od . In the special case, flows at the origin airport o are described by $g_{odjl}[k] = \rho_{odjl} f_{od}[k]$, where $f_{od}[k]$ is the total demand in sub-network od at time k .

When flows arrive at the boundary intersection point, they are modeled as entering a virtual buffer. Due to management actions and weather impact, only limited amount of flows are allowed to enter the region, with the rest waiting in the buffer. Mathematically, let ijm denote the boundary intersection point determined by region m and link (i, j) , and $N_{ijm}[k]$ be the number of aircraft allowed to enter region m from

ijm at time k . Also define $u_{ijm}[k]$, $e_{ijm}[k]$, $b_{ijm}[k]$ as the number of aircraft arriving at ijm , leaving ijm , and waiting in the buffer at time k , respectively. The flow dynamics at boundary intersection point ijm can then be captured by following equations

$$\begin{aligned} e_{ijm}[k] &= \min(N_{ijm}[k], b_{ijm}[k-1] + u_{ijm}[k]) \\ b_{ijm}[k] &= \max(0, b_{ijm}[k-1] + u_{ijm}[k] - N_{ijm}[k]) \\ u_{ijm}[k] &= \sum_{\text{all O-D pairs}} g_{odij}[k - k_{i,ijm}] \\ \sum_{i,j} N_{ijm}[k] &\leq \Gamma_m \end{aligned} \quad (2)$$

where Γ_{ijm} and Γ_m are the capacity of link (i, j) in region m and the total capacity of region m in clear weather, respectively. $k_{i,ijm}$ is the number of time steps for the aircraft to travel from node i to node ijm , where the speed of the aircraft is assumed to be constant.

As convective weather directly reduces the capacity of a region, a scaling factor $\omega_m[k] \in [0, 1]$ is introduced to model weather impact. In particular, the capacity of region m in convective weather is captured by

$$\Gamma_m^0[k] = \omega_m[k] \Gamma_m \quad (3)$$

This model allows the design of different TMIs. For instance, the minute-in-trail (MIT) strategy, a typical TMI, is captured by N_{ijm} .

B. Weather-Impact Model

In this section, we briefly review a stochastic weather-impact model [18], which captures the spatiotemporal evolution of uncertain weather events and their impact on traffic flows at the strategic time frame. This model will be used in simulation studies (Section IV) to generate weather-impact scenarios, i.e., $\omega_m[k]$ in (3), which are required for evaluating and designing TMIs.

This weather-impact model [18] describes the spatiotemporal evolution of weather impact as a networked Markov process [19]. In particular, each region m is considered to have a discrete-valued weather-impact state varying over time, indicating the changes of region capacity under weather impact. At each time k , the weather-impact state of region m is determined by its neighbors' previous states at time $k-1$. The influence of a neighbor n on region m is described by two parameters: 1) a scalar $c_{mn} \in [0, 1]$ that reflects how frequently region n influences region m , and 2) a *local transition matrix* $A_{nm} \in \mathbb{R}^{M \times M}$ that captures the influence of region n 's previous state on region m 's current state. In particular, let $\mathbf{s}_m[k] \in \mathbb{R}^{1 \times M}$ be the state vector of region m at time k , whose i -th element equals to 1 if region m is at state i and 0 otherwise. M is the total number of states. Then the state of region m at time k , $\mathbf{s}_m[k]$, is determined by following two steps.

- First, region m randomly selects a neighbor n as its *determining region* with probability c_{mn} . Note that $c_{mn} = 0$ if regions m and n are not neighbors and $\sum_n c_{mn} = 1$.

- Second, region m randomly picks its current state $\mathbf{s}_m[l]$ according to a probability mass function specified by $\mathbf{s}_m[k-1]A_{nm}$.

By modulating the values of C_{mn} and A_{nm} , we can generate weather-impact scenarios of different statistics.

C. Problem Formulation

In this study, we consider the optimal MIT design problem formulated as follows.

Consider an air traffic system with airspace partitioned into N regions. Our objective is to find the optimal MIT strategy for a given weather-impact scenario, such that total cost is minimized. To evaluate the performance of an MIT strategy, we use the total backlog, i.e., $\sum_{k=1}^{k_p} B[k] = \sum_{k=1}^{k_p} \sum_{i,j,m} b_{ijm}[k]$, as the performance metric, which reflects the severity of traffic congestion. $k_p > 0$ is the end time of consideration. As the unit cost incurred by management induced backlogs are typically smaller than that incurred by weather induced backlogs, we adopt the following cost function that assigns different unit costs to the management and weather induced backlogs [14]

$$J = \sum_{k=0}^{k_p} f_1(B_M[k]) + f_2(B_{NM}[k]), \quad (4)$$

where $B_M[k]$ and $B_{NM}[k]$ represent the management and non-management induced total backlogs at time k , respectively. $f_1(\cdot)$ and $f_2(\cdot)$ are functions that quantify the costs of management and non-management induced backlogs. To penalize high transient backlogs that are not desired in reality, we let $f_i(\cdot) = C_i x + H_i \mathbf{1}(x - \gamma_i)$, $i \in \{1, 2\}$, where C_i is the unit cost associated with each type of backlogs. γ_i is a threshold for high transient backlogs, and H_i is the extra unit cost associated with high transient backlogs. $\mathbf{1}(x - \gamma_i) = x - \gamma_i$, if $x \geq \gamma_i$, and 0 otherwise.

III. SPATIOTEMPORAL SCENARIO DATA-DRIVEN DECISION-MAKING FRAMEWORK

In this section, we first provide an overview of the spatiotemporal scenario data-driven decision-making framework for strategic ATFM. We then describe in detail its key components.

A. Overview

To conquer the challenge of efficiently managing air traffic systems at a large spatiotemporal scale, the spatiotemporal scenario data-driven decision-making framework is proposed that makes advantage of historical TMIs to speed up the management design, under the assumption that similar weather-impact scenarios lead to similar TMIs. An overview of the proposed decision-making framework is shown in Fig. 1, which consists of three components: *historical database*, *query subsystem* and *control subsystem*.

The *historical database* stores historical weather-impact scenarios and associated TMIs. These historical data are used to facilitate online decision-making. When a new TMI is designed for a new weather-impact scenario, both the TMI

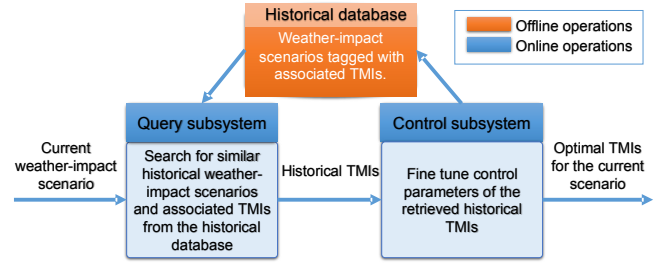


Fig. 1: Overview of the spatiotemporal scenario data-driven decision-making framework for strategic ATFM.

and corresponding scenario are inserted into the database to support future decisions.

The *query subsystem*, core component of our framework, fills the gap between online and offline operations. Given a new weather-impact scenario to be evaluated, it quickly retrieves similar scenarios and associated TMIs from the historical database. The retrieved historical TMIs are then fed into the control subsystem for further action.

The *control subsystem*, brain of our framework, takes the retrieved historical TMIs from the query subsystem and fine-tunes their control parameters to derive the optimal solution for the current weather-impact scenario under investigation.

In the following subsections, we describe the query and control subsystems in more detail.

B. Query Subsystem

The query subsystem is realized by a *multiresolution distance-based similarity search algorithm* developed in our previous studies [17], [20], which achieves fast query of similar weather-impact scenarios. Note that the weather-impact scenarios of spatiotemporal evolving dynamics represent a new data type, called *spatiotemporal scenario data*, which has been rarely studied [16], [21].

The effectiveness of the multiresolution distance-based similarity search algorithm is ensured by a multiresolution distance measure [12], [16] that captures the similarity between two scenarios by scanning and comparing the two scenarios at different resolutions using a spatiotemporal window of varying size. In particular, consider two weather-impact scenarios, S_i and S_q , each of which is described by a few snapshots of the weather impacts over an airspace G captured at a few continuous time points T . A distance $d_{i,q,w,h}$ is calculated after each round of scenario scan using a spatiotemporal window of size w and h . Mathematically,

$$d_{i,q,w,h} = \frac{\sum_{\varphi_{z,w} \in \Phi_w} \sum_{\varphi_{l,h} \in \Phi_h} \frac{1}{|\varphi_{z,w}| |\varphi_{l,h}| |\Phi_h|} \Delta l_{iq}}{\quad} \quad (5)$$

where $\Delta l_{iq} = \sum_{g_n \in \varphi_{z,w}} \sum_{t_m \in \varphi_{l,h}} \frac{P}{\lambda_{n,w} \tau_{m,h}} \cdot \frac{I_{i,n,m}}{\lambda_{n,w} \tau_{m,h}}$. $I_{i,z,l}$ is the intensity of weather impact on region $g_z \in G$ at time point $t_l \in T$ in scenario S_i . $\varphi_{z,w} \in \Phi_w$ is a spatial window of size w centered at region g_z and $\varphi_{l,h} \in \Phi_h$ is a temporal window of size h starting from time point t_l , where Φ_w and Φ_h are the

full sets of spatial windows of size w and temporal windows of size h , respectively. $\lambda_{z,w}$ and $\tau_{l,h}$ are contribution factors that equalize the contribution of each region g_z and each time point t_l to the distance calculation, respectively. $|A|$ represents the cardinality of set A . The overall distance $D_{i,q} = \sum_{h=1}^{h_{\max}} \sum_{w=1}^{w_{\max}} d_{i,q,w,h} \frac{\delta_w \alpha_h}{\sum_{h=1}^{h_{\max}} \sum_{w=1}^{w_{\max}} \delta_w \alpha_h}$ is the weighted sum of $d_{i,q,w,h}$ obtained at all resolutions, where $w_{\max}$ and $h_{\max}$ denote the maximum spatial and temporal window sizes to evaluate, respectively. δ_w and α_h are weighting factors that control the contributions of different spatial and temporal resolutions to the distance calculation.

The efficiency of the multiresolution distance-based similarity search algorithm is achieved by a progressive pruning procedure that prunes the search space after each round of scenario scan using gradually tightened lower and upper bounds of the distance measure. In particular, given a database S and a query scenario S_q , consider the problem of finding the top- K scenarios in the database S that are most similar to the query scenario S_q . Let $\underline{D}_{i,q}^{(j)}$ and $\overline{D}_{i,q}^{(j)}$ be the lower and upper bounds of the overall distance $D_{i,q}$ between scenarios $S_i \in S$ and S_q calculated after the j -th round of scenario scan, respectively, where $j \in \{1, 2, \dots, w_{\max} h_{\max}\}$. All scenarios in the database that satisfy $\underline{D}_{i,q}^{(j)} > M_K$ are safely discarded after each round of scenario scan, where $M_K = \max_{S_j \in A_K} \overline{D}_{i,q}^{(j)}$ is the K -th smallest upper bound value and A_K is the set of top K scenarios in S with the smallest upper bound values. The lower and upper bounds are updated using following equations after each round of scenario scan

$$\begin{aligned} \underline{D}_{i,q}^{(j)} &= \begin{cases} d_{i,q,w^*,h^*} + \frac{\delta_w \alpha_h}{\sum} (d_{i,q,1,1} - d_{i,q,w^*,h^*}), & \text{if } j = 1 \\ \underline{D}_{i,q}^{(j-1)} + \frac{\delta_w \alpha_h}{\sum} (d_{i,q,w^{(j)},h^{(j)}} - d_{i,q,w^*,h^*}), & \text{else} \end{cases} \\ \overline{D}_{i,q}^{(j)} &= \begin{cases} d_{i,q,1,1}, & \text{if } j = 1 \\ \overline{D}_{i,q}^{(j-1)} + \frac{\delta_w \alpha_h}{\sum} (d_{i,q,w^{(j)},h^{(j)}} - d_{i,q,1,1}), & \text{else} \end{cases} \end{aligned} \quad (6)$$

where $\sum = \sum_{h=1}^{h_{\max}} \sum_{w=1}^{w_{\max}} \delta_w \alpha_h$. $w^{(j)}$ and $h^{(j)}$ are the spatial and temporal window sizes at the j -th round of scenario scan, respectively. w^* and h^* represent the largest spatial and temporal window sizes that lead to full coverage of the spatial map G and time sequence T , respectively.

This similarity search algorithm also implements two data access schemes to further expedite the search speed in large databases. One is a modified Filter-Restart scheme that performs an initial pruning of the search space to find an initial candidate set S_c . Another scheme prioritizes window sizes w and h with large weights $\delta_w \alpha_h$ to increase the speed that the bounds are tightened. Let $W = \{(w^{(j)}, h^{(j)})\}_{j=1}^{w_{\max} h_{\max}}$ be the sorted list of window sizes, where $\delta_{w^{(j_1)}} \alpha_{h^{(j_1)}} \geq \delta_{w^{(j_2)}} \alpha_{h^{(j_2)}}$, $\forall j_1 < j_2$, $j_1, j_2 \in \{1, 2, \dots, w_{\max} h_{\max}\}$. Algorithm 1 summarizes the procedures of this similarity search algorithm.

C. Control Subsystem

In this pioneering study, we adopt the genetic algorithm [22], a typical evolutionary algorithm, to realize the control subsystem. The genetic algorithm searches for the (near) optimal solution to a given problem by imitating human

Algorithm 1: Multiresolution Distance-based Similarity Search Algorithm

Input: Query S_q , Database S , and K
Output: The top- K most similar scenarios to query S_q

- 1 Apply the Filter-Restart scheme to find an initial candidate set S_c ;
- 2 **for** $j = 2$ **to** $w_{\max} h_{\max}$ **do**
- 3 **foreach** $S_i \in S_c$ **do**
- 4 Calculate $d_{i,q,w^{(j)},h^{(j)}}$ using (5), where $(w^{(j)}, h^{(j)}) \in W$;
- 5 Calculate $\underline{D}_{i,q}^{(j)}$ and $\overline{D}_{i,q}^{(j)}$ using (6);
- 6 **end**
- 7 **if** $|S_c| > K$ **then**
- 8 Determine the value of M_K ;
- 9 Remove all scenarios S_i that satisfy $\underline{D}_{i,q}^{(j)} > M_K$ from the candidate set S_c ;
- 10 **else**
- 11 Exit from the **for** loop;
- 12 **end**
- 13 **end**
- 14 **if** $|S_c| > K$ **then**
- 15 $S_c \leftarrow K$ scenarios selected from S_c that have the smallest upper bound values $\overline{D}_{i,q}^j$;
- 16 **end**

evolution. It operates on a *population* of *individuals*, where each individual is described by a *chromosome* and each chromosome is composed of a set of *genes*. In a design problem, each gene corresponds to a control parameter, the chromosome corresponds to the full set of control parameters, and each individual represents a potential solution. After the population is initialized, the algorithm evolves the population from generation to generation through iteratively applying three operations: *selection*, *crossover*, and *mutation*. In particular, given the current population, the selection is first applied to pick several pairs of individuals as *parents* according to their *fitness*, where fitness corresponds to the cost of a solution. Individuals with better fitness (i.e., solutions of lower cost) are more likely to be selected. Then, the crossover exchanges parts of the parents' chromosomes to generate *children* and the mutation further adds randomness to some genes of the children to maintain diversity. The evolution continues until the (near) optimal solution is found or a specified number of generations have been evaluated.

In our problem, the MIT rate, its start time and duration are the control parameters to be optimized, and the MIT restriction can be applied at any boundary intersection point ijm . Let $\pi = \{ijm\}$ include the set of boundary intersection points to be controlled using MIT. A chromosome then contains $3|\pi|$ genes. The MIT rate, denoted as R_{ijm} , is allowed to take any integer from 0 up to Γ_m , under the condition that $\sum_{i,j} R_{ijm} \leq \Gamma_m$, and it takes effect when $\sum_{i,j} R_{ijm} \leq \Gamma_m^0$. The start time and duration are selected so that the span of control does not exceed the span of weather-

impact scenarios. To speed up the design, we initialize the population using historical MIT strategies for weather-impact scenarios similar to the current scenario under evaluation. After the algorithm terminates, we save the top K best solutions to the database to facilitate future decisions. Note that if the size of the population is P , then P historical MIT strategies for the top $K = \frac{P}{\kappa}$ most similar weather-impact scenarios will be retrieved by the query subsystem from the historical database.

IV. SIMULATION STUDIES

In this section, we conduct simulation studies to investigate the performance of the spatiotemporal scenario data-driven decision-making framework implemented in MATLAB. Two examples of different scales are studied on the Alienware Aurora with an Intel Core i7-7700K 4.20GHz processor and 32GB memory.

A. Small-Scale Study

We consider an air traffic system of four O-D sub-networks as shown in Fig. 2. Different sub-networks are differentiated by different colors, and the airspace is partitioned into $N = 8$ regions $\{g_1, g_2, \dots, g_8\}$. The merging/splitting points coincide with the boundary intersection points (white nodes). The red numbers on the links represent the transit time measured in minutes. Flow fractions are also marked on the links. To simulate air traffic flows, we randomly generate traffic demand $f_{od}[k]$ from a Poisson distribution with a mean of 5 at 15-minutes intervals and inject these flows into each O-D sub-network. The capacity of each region is set to $\Gamma_1 = 18, \Gamma_2 = 14, \Gamma_3 = \Gamma_5 = 5, \Gamma_4 = \Gamma_6 = \Gamma_8 = 15$, and $\Gamma_7 = 4$.

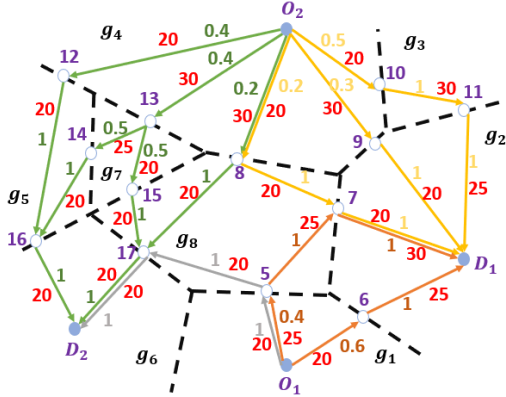


Fig. 2: An air traffic system of four O-D sub-networks.

To build the historical database, we generate 1000 weather-impact scenarios using the weather-impact model described in Section II-B. The local transition matrix in the model is set to $A_{nm} = \begin{bmatrix} 0.9 & 0.05 & 0.03 & 0.02 \\ 0.8 & 0.1 & 0.06 & 0.04 \\ 0.9 & 0.03 & 0.03 & 0.04 \\ 0.9 & 0.07 & 0.02 & 0.01 \end{bmatrix}$, $\forall n, m \in \{1, 2, \dots, 8\}$, and the scalars $c_{mn} \in [0, 1]$ are randomly generated. Each region has four possible states, corresponding to Full Capacity ($\omega_m[k] = 1$), High Capacity

($\omega_m[k] = 0.8$), Medium Capacity ($\omega_m[k] = 0.6$), and Low Capacity ($\omega_m[k] = 0.4$). Each scenario lasts for 6 hours from 10:00 to 15:00 with 1-hour intervals. For each scenario, we run the genetic algorithm to generate corresponding MIT strategies. In this small-scale study, we apply MIT restriction on nodes $\pi = \{5, 8, 15\}$. The optimization cost function in (4) is configured by setting $C_1 = 2, C_2 = 5, H_1 = 2, H_2 = 3$ and $V_1 = V_2 = 10$. For the genetic algorithm, we set the population size to $P = 60$, size of the parent pool to 30 and mutation rate to 0.1. The algorithm terminates when the best fitness score is unchanged for 12 successive generations or 100 generations have been evaluated. The top $\kappa = 10$ best solutions for each scenario are saved into the database.

Now consider a new weather-impact scenario visualized in Fig. 3(a), we apply the proposed decision-making framework to derive the (near) optimal MIT strategy. In particular, the query subsystem first retrieves historical MIT strategies for the top $K = 6$ scenarios that are most similar to the current scenario, with $w_{max} = 8, h_{max} = 6, \delta_w = e^{-0.8(w-1)}, \alpha_h = e^{-0.8(h-1)}$. The scenario that is most similar to the current scenario is visualized in Fig. 3(b). The control subsystem then takes the retrieved 60 MIT strategies to initialize the population and runs the genetic algorithm to find the (near) optimal MIT strategy for the current scenario. The result is provided in Table I, which also shows the best MIT strategy for the retrieved most similar scenario. Note that there are multiple optimal strategies with the lowest cost, and we randomly pick one. As we expected, similar weather-impact scenarios lead to similar MIT strategies.

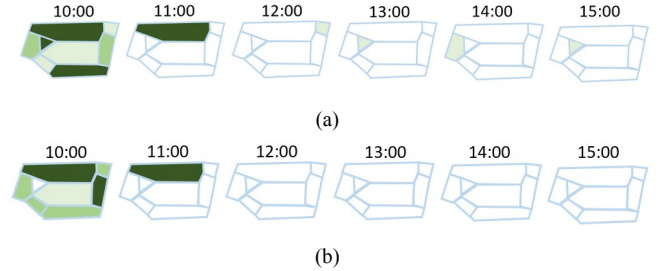


Fig. 3: Visualization of (a) the current weather-impact scenario to be evaluated and (b) the retrieved most similar scenario. Darker colors indicate higher weather intensities.

For comparison, we also run the traditional genetic algorithm with randomly initialized population. The results show that our method is significantly more efficient than the traditional genetic algorithm that designs from scratch. In particular, our method only takes around 15 generations (averaged over 20 experimental runs) and 35.07s (including the time for scenario retrieval) to find the (near) optimal solution, while the traditional genetic algorithm that designs from scratch takes around 31.5 generations and 61.12s. In addition, the MIT strategies found by our method are often better than the ones found by the traditional genetic algorithm, which have an average cost of 2229.46. Of note, the superiority of our method diminishes when the problem scale decreases, due to the overhead for scenario retrieval.

TABLE I: The (near) optimal MIT strategies for the current and the retrieved most similar weather-impact scenario in the small-scale study.

ijm	Current Scenario			Retrieved Most Similar Scenario		
	Start Time	Duration (hrs)	R_{ijm}	Start Time	Duration (hrs)	R_{ijm}
5	13:15	1	9	13:15	1	9
8	15:30	1	3	15:30	0.25	3
15	15:00	0.5	1	15:00	0.5	1
Cost	$J = 2228.05$			$J = 2019.65$		

B. Large-Scale Study

In this study, we consider the same air traffic system shown in Fig. 2, but apply MIT restrictions on seven nodes $\pi = \{5, 8, 15, 6, 7, 9, 11\}$ and extend the control horizon to 12 hours. The demand $f_{od}[k]$ is randomly generated from a Poisson distribution with a mean of 10, and the capacity of each region is set to $\Gamma_1 = 32$, $\Gamma_2 = 29$, $\Gamma_3 = 9$, $\Gamma_4 = 31$, $\Gamma_5 = 10$, $\Gamma_6 = 30$, $\Gamma_7 = 8$, and $\Gamma_8 = 28$. The historical database consists of 10000 weather-impact scenarios with each lasting for 12 hours. The population size, size of the parent pool and mutation rate set in the genetic algorithm are set to 80, 30, and 0.1, respectively. The other settings are same as the small-scale study.

Given a new weather-impact scenario, we run our method and the traditional genetic algorithm with randomly initialized population. As expected, our method requires much fewer generations (around 31.9 generations averaged over 20 experimental runs) and less time (around 3.28min) to find the (near) optimal solution than the traditional genetic algorithm that requires around 63.4 generations and 5.47min. Our method also finds better MIT strategies of lower cost (around 7506.65) than the traditional genetic algorithm that finds MIT strategies with an average cost of 7584.72

V. CONCLUSION

In this paper, we introduced a novel spatiotemporal scenario data-driven decision-making framework for strategic ATFM. By moving most computations to offline, this framework makes it possible to manage large-scale air traffic systems in real-time. Instead of designing from scratch, this framework utilizes historical TMIs for similar weather-impact scenarios to significantly increase the design speed, where the historical TMIs are retrieved from the database and fine tuned to derive the optimal solution for the current scenario. Simulation studies at different scales demonstrate the effectiveness and efficiency of the proposed framework. In the future, we will explore the difference between current and retrieved scenarios to reduce the number of control parameters to be fine tuned.

REFERENCES

- [1] "FAQ: Weather Delay," <https://www.faa.gov/nextgen/programs/weather/faq/>, 2017, [Online; accessed Jan. 17, 2019].
- [2] C. Taylor, C. Wanke, Y. Wan, and S. Roy, "A framework for flow contingency management," in *Proceedings of the 11th AIAA Aviation Technology, Integration, and Operations Conference*, 2011, p. 6904.
- [3] Y. Zhou, Y. Wan, S. Roy, C. Taylor, and C. Wanke, "A stochastic modeling and analysis approach to strategic traffic flow management under weather uncertainty," in *Proceedings of AIAA Guidance, Navigation, and Control Conference*, 2011, p. 6514.

- [4] J.-P. B. Clarke, S. Solak, Y. Chang, L. Ren, and A. Vela, "Air traffic flow management in the presence of uncertainty," in *Proceedings of the 8th USA/Europe Air Traffic Seminar*, 2009.
- [5] Y. Wan, C. Taylor, S. Roy, C. Wanke, and Y. Zhou, "Dynamic queuing network model for flow contingency management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1380–1392, 2013.
- [6] —, "Dynamic queuing network model for flow contingency management," in *Proceedings of AIAA Guidance, Navigation, and Control Conference*, p. 6361.
- [7] Y. Wan and S. Roy, "A scalable methodology for evaluating and designing coordinated air-traffic flow management strategies under uncertainty," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 4, pp. 644–656, 2008.
- [8] —, "Sensitivity of national airspace system performance to disturbances: modeling, identification from data, and use in planning," in *Proceedings of AIAA Guidance, Navigation and Control Conference and Exhibit*, 2008, p. 6323.
- [9] Y. Zhou, Y. Wan, S. Roy, C. Taylor, C. Wanke, D. Ramamurthy, and J. Xie, "Multivariate probabilistic collocation method for effective uncertainty evaluation with application to air traffic flow management," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 10, pp. 1347–1363, 2014.
- [10] J. Xie, Y. Zhou, Y. Wan, A. Mitchell, and S. Roy, "A jump-linear model based sensitivity study for optimal air traffic flow management under weather uncertainty," in *Proceedings of AIAA Infotech@Aerospace*, 2015.
- [11] J. Xie, Y. Wan, and F. L. Lewis, "Strategic air traffic flow management under uncertainties using scalable sampling-based dynamic programming and q-learning approaches," in *Proceedings of the 11th Asian Control Conference*, IEEE, 2017, pp. 1116–1121.
- [12] J. Xie and Y. Wan, "Scalable multidimensional uncertainty evaluation approach to strategic air traffic flow management," in *Proceedings of AIAA Modeling and Simulation Technologies Conference*, 2015, p. 2492.
- [13] —, "A network condition-centric flow selection and rerouting strategy to mitigate air traffic congestion under uncertainties," in *Proceedings of the 17th AIAA Aviation Technology, Integration, and Operations Conference*, 2017, p. 3427.
- [14] Y. Zhou, Y. Wan, C. Taylor, S. Roy, and C. Wanke, "Performance evaluation and optimal decision-making for strategic air traffic management under weather uncertainty," in *Proceedings of Infotech@Aerospace*, 2012, p. 2516.
- [15] J. Xie, Y. Wan, K. Mills, J. J. Filliben, and F. L. Lewis, "A scalable sampling method to high-dimensional uncertainties for optimal and reinforcement learning-based controls," *IEEE control systems letters*, vol. 1, no. 1, pp. 98–103, 2017.
- [16] J. Xie, Y. Wan, Y. Zhou, S.-L. Tien, E. P. Vargo, C. Taylor, and C. Wanke, "Distance measure to cluster spatiotemporal scenarios for strategic air traffic management," *Journal of Aerospace Information Systems*, vol. 12, no. 8, pp. 545–563, 2015.
- [17] J. Xie, H. Nguyen, and Y. Wan, "Similarity search of spatiotemporal scenario data for strategic air traffic management," in *Proceedings of Aviation Technology, Integration, and Operations Conference*, 2018, p. 3978.
- [18] S. Roy, Y. Wan, C. Taylor, and C. Wanke, "A stochastic network model for uncertain spatiotemporal weather impact at the strategic time horizon," in *Proceedings of the 10th AIAA Aviation Technology, Integration, and Operations Conference*, 2010, p. 9348.
- [19] C. Asavathiratham, S. Roy, B. Lesieutre, and G. Verghese, "The influence model," *IEEE Control Systems*, vol. 21, no. 6, pp. 52–64, 2001.
- [20] J. Xie, A. Reddy Kothapally, Y. Wan, C. He, C. Taylor, C. Wanke, and M. Steiner, "Similarity search of spatiotemporal scenario data for strategic air traffic management," *Journal of Aerospace Information Systems*, pp. 1–16, 2019.
- [21] J. Xie, Y. Wan, Y. Zhou, S.-L. A. Tien, C. Taylor, and C. Wanke, "A multi-resolution spatiotemporal weather impact scenario clustering algorithm for flow contingency management," in *Proceedings of the 14th AIAA Aviation Technology, Integration, and Operations Conference*, 2014, p. 2029.
- [22] A. P. Alves da Silva and D. M. Falcão, "Fundamentals of genetic algorithms," *Modern heuristic optimization techniques: theory and applications to power systems*, pp. 25–42, 2008.