APPLICATION OF ARTIFICIAL INTELLIGENCE IN THE OPTIMIZATION OF MOBILITY IN DYNAMIC AIRSPACE CONFIGURATIONS DURING EMERGENCY SITUATIONS

FINAL REPORT

MAY 2023

AUTHORS

KE FENG, YONGXIN LIU, MARC MEIER-DORNBERG,

WAIRIMU MWANGI, DAHAI LIU

EMBRY-RIDDLE AERONAUTICAL UNIVERSITY

HOUBING SONG

UNIVERSITY OF MARYLAND, BALTIMORE COUNTY

US DEPARTMENT OF TRANSPORTATION GRANT 69A3551747125

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Application of Artificial Intelligence in the Optimization of Mobility in Dynamic Airspace Configurations During Emergency Situations		5. Report Date May 2023 6. Source Organization Code ERAU Cost Center 61603
7. Author(s) Ke Feng, Yongxin Liu, Marc Meier-Dornberg, Wairimu Mwangi, Dahai Liu, Houbing Song		8. Source Organization Report No. CATM-2023-R3-ERAU
9. Performing Organization Name and Address Center for Advanced Transportation Mobility Transportation Institute 1601 E. Market Street Greensboro, NC 27411		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 69A3551747125
12. Sponsoring Agency Name and Address University Transportation Centers Program (RDT-30) Office of the Secretary of Transportation–Research U.S. Department of Transportation 1200 New Jersey Avenue, SE Washington, DC 20590-0001		13. Type of Report and Period Covered Final Report: October 2020 – May 2023 14. Sponsoring Agency Code USDOT/OST-R/CATM
15. Supplementary Notes:		

16. Abstract

In emergency situations, the timely evacuation of people is crucial, and air traffic plays a significant role. Dynamic airspace configuration (DAC) is a promising approach to maximizing air traffic throughput while accommodating dynamic traffic changes, making it suitable for emergency evacuation. This study aims to use artificial intelligence to develop practical DAC for emergency evacuation. The proposed modeling method constructs the airspace as a graph and applies a spectral clustering algorithm to group airports and balance the workload among sectors. Sharing air traffic control resources within the cluster can lead to more efficient air traffic control. Two sets of experiments conducted under different air traffic conditions show that our approach reduces the workload unbalance level among sectors by 50%. This study has the potential to be further developed into a recommendation system to assist with airspace configuration during emergency evacuations.

17. Key Words Emergency; Evacuation; Dynamic Airspace Configuration	18. Distribution Statement Unrestricted; Document is available to the public through the National Technical Information Service; Springfield, VT.

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 80	22. Price …

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY

As the aviation industry continually grows and the demand for air travel increases, one important factor that determines the sustainability of this growth is Air Traffic Control (ATC). ATC is the heart of air operations and is a ground-based global system that has the primary purpose of providing air traffic services that include preventing the collision of aircraft; expediting the flow of air traffic in an orderly manner; providing en route flight advisory to pilots; ground taxiing, landing, and take off procedures; among many other functions. In addition, by following Air Traffic Management (ATM) policies, ATC is responsible for the operation of aircraft within a given airspace structure. Dynamic Airspace Configuration (DAC) offers a new airspace structure that is dynamic. This means that DAC has the ability to be flexible based on user demand while meeting the constraints of elements such as weather, fleet diversity, congestion, and sector complexity (Kopardekar, Bilimor, & Sridhar, 2007). Since DAC is able to maximize air traffic throughput while accommodating dynamic traffic changes, it is suitable for emergency evacuation when the traffic demand changes rapidly and under a lot of constraints. Given that DAC is a new paradigm within the aviation industry, this study is looking at ways to effectively transition the static airspace configuration to this newly proposed configuration and its potential to apply it for emergency evacuation.

To provide an optimal solution to realize DAC, the literature review section discusses and provides insight into how artificial intelligence (AI) can be used to optimize mobility in DAC during emergency situations by looking at current ground transportation practices that could potentially be transferred to DAC. Then the implementation is conducted through two phases. First, a Spatial-temporal graph neural network is proposed to make predictions on each airport's workload. The purpose of the second stage is to realize practical DAC by balancing the workload among sectors to improve air traffic control efficiency, thus maximizing the air traffic throughput. Our proposed approach is evaluated under two different experiment settings for different traffic conditions. The result shows it is able to decrease the unbalanced level of ATC workload by over 50%. It is believed that our approach has the potential to be further developed into a recommendation system to assist with airspace configuration during emergency evacuations.

2 BACKGROUND

This literature review aims to present the application of machine learning algorithms in DAC emergency procedures by identifying algorithms and practices in transportation systems that could potentially be carried over to DAC. At the time of writing this review and to the best of our knowledge, DAC is still a new field that is currently under investigation and there are not many resources and tangible real-time applications of emergency response procedures in DAC. Therefore, we use the ground transportation approach in addition to existing airspace-focused research to establish a theoretical foundation for DAC in emergency situations. The goal of creating a DAC aligns with the U.S. Department of Transportation's goal of creating Intelligent Transport Systems (ITS) that have improved efficiency, safety, and operational capabilities.

2.1 EMERGENCY EVACUATION OVERVIEW

2.1.1 Definition of Emergencies

The U.S. Federal Emergency Management Agency (FEMA) defines the term emergency as "any incident, whether natural, technological or human caused, that necessitates responsive action to protect life or property " (Federal Emergency Management Agency & United States Department of Homeland Security, 2022). Certain types of emergencies necessitate evacuation, "the organized, phased and supervised withdrawal, dispersal or removal of civilians from dangerous or potentially dangerous areas, and their reception and care in safe areas" (Federal Emergency Management Agency & United States Department of Homeland Security, 2022). Closely related to the term "emergency" is the term "disaster." Since there is no universally accepted definition, the terms "disaster" and "emergency" are sometimes used interchangeably. However, disasters typically cause injury or loss of life on a larger scale. A disaster is usually dealt with under a state of emergency and can require evacuation (Shaluf, Ahmadun, & Mat Said, 2003).

Ground and air transportation play vital roles in emergency management since the transportation systems can be the cause of the emergency and/or be used to alleviate the impacts of an emergency, for example through disaster relief and evacuation efforts. Ground transportation can be considered as any form of transport that occurs through the mode of water

and land transportation. Air transportation is any form of transport that is performed through the air, for example by commercial aircraft or rescue helicopters. While the focus of this work lies on the response to emergencies through air transportation, emergency ground transportation is recognized as a well-researched field and is therefore used as an additional reference.

2.1.2 Causes of Emergencies

Emergencies can be caused by numerous and a diverse set of factors. Shaluf et al. (2003) state that disasters are often caused by "accumulated unnoticed events." The authors differentiate five disaster-type events:

1. *Conflict type situation 'Political Crisis'* – These events can be caused by factors such as war, embargoes, internal conflicts, hostile takeovers, and terrorist attacks.
2. *Non-conflict type situation* – Examples of this event are financial crisis, sabotage, poor or faulty training, and loss of proprietary information.
3. *Socio technical [human caused] disasters* – This category is named by the FEMA as a common cause for emergencies. The FEMA names nuclear explosions or cyberattacks as examples (Federal Emergency Management Agency & United States Department of Homeland Security, 2022).
4. *Natural disasters* – Along with human caused disasters, natural disasters are named by the FEMA as a common cause for emergencies. The FEMA lists numerous hazards including natural events such as hurricanes, thunderstorms, earthquakes, tsunamis, and volcanic eruptions (Federal Emergency Management Agency, 2020). Scientific advancements allow for prediction of potential natural disasters with varying lead time and accuracy. Depending on the type of event and lead time, pre-event warnings allow individuals to seek shelter or evacuate (National Research Council, 1991).
5. *Transportation accidents* – These events impact the transportation system but do not have an immediate emergency-type impact on the community.

All of these disaster-type events can result in situations that need to be dealt with under a state of emergency. There is ample precedent that shows the different roles that civilian or military airspace users, as well as the airspace system itself, can play in these situations:

a. Airspace users and/or the airspace system cause the situation. Boeing (2022) counts 307 commercial jet airplane accidents worldwide between 2012 and 2021. Many of these accidents occur while operating on ground and do not result and any fatalities. However, these accidents may result in temporary airport and/or airspace capacity reduction through closed runways, taxiways, or gates.
b. Airspace users and/or the airspace system are impacted by the situation. The NAS and its users frequently deal with severe weather and other events that can impact air traffic routes. The FAA issues the National Playbook, a collection of Sever Weather Avoidance Plan (SWAP) routes for common scenarios, including non-weather events such as military operations (Department of Transportation, 2021). In addition to this, events ranging from volcanic eruptions (Eyjafjallajokull's global fallout.2010) to armed conflicts (Kuipers, Verolme, & Muller, 2020) can impact routes, airports, and other parties involved.
c. Airspace users and/or the airspace system are needed for evacuation or relief, for example during the evacuation efforts in Afghanistan in 2021 (Sune, Donati, & Youssef, 2021) or after natural disasters (Royal Aeronautical Society, 2006). With sufficient lead time prior to natural disasters, airlines may decide to move aircraft to other locations (Seet, 2022).

2.1.3 Emergency Response Plans

Emergency response plans are part of the Emergency Response Management (ERM) systems commonly used in transportation emergency practices. An emergency plan can be described as an official document that establishes the policies and procedures to be followed during unexpected incidents with the objective of preventing fatalities, injuries, as well as property and environmental damage. The International Civil Aviation Organization (ICAO) recommends [air navigation] service providers to "establish and maintain an emergency response plan for accidents and incidents." Service providers shall also ensure that the

emergency plan is properly coordinated with the organizations involved (ICAO, 2016). Furthermore, ICAO recommends operators of aerodromes (ICAO, 2004) and heliports (ICAO, 2013) establish an emergency plan for the coordination of actions taken in an emergency. These emergency plans include frameworks on how to communicate with other agencies.

In addition to emergency response plans, ICAO asks its members to establish contingency plans for the event of disruptions of air traffic services and related supporting services to preserve availability of major world air routes. Contingency plans make provisions for temporary disruptions of primary [air navigation] services and facilities (ICAO, 2001). Considering the complex structure of the NAS and DAC, an in-built emergency response plan as well as a contingency plan should be considered in its model. This would provide a streamlined process for controllers and the organization to follow that is comprehensive and outlines the "specific roles, set of actions, and timeframes to respond to unexpected situations, disruptions, or potential disruptions" (International Civil Aviation Organization, 2019). Given that DAC offers a dynamically changing airspace by "adapting to user demand while meeting changing constraints of weather, traffic congestion and complexity, as well as a highly diverse aircraft fleet,", it is important that controllers and other parties are provided with the necessary resources such as emergency plans to sustain their operations safely within the DAC model (Kopardekar et al., 2007). Emergency response plans are essential when preparing traffic controllers to effectively handle unexpected incidents in a manner that controls the stress of the situation while keeping the controllers focused and efficient while diffusing the emergency.

2.1.4 Emergency Operations

In addition, the following procedures as per the U.S Department of Transportation's Federal Highway Administration's report Best Practices in Emergency Transportation Operations Preparedness and Response could also be considered in DAC emergency situations (Houston & Hamilton, 2006). These practices include:

- *Virtual Emergency Operations Center* – According to the U.S FEMA, an Emergency Operations Center (EOC) is described as "a [physical, virtual or hybrid] location from which leaders of a jurisdiction or organization coordinate information and resources to

support incident management activities" (FEMA, 2022). EOCs aim to gather highly trained experts who have access to state-of-the art technology that enables them to mitigate an emergency and reduce potential adverse effects of its outcome. With respect to DAC, EOCs can potentially be used to coordinate and establish the immediate actions followed during an emergency within DAC sectors. According to Kopardekar et al. (2007), the current static structure of airspace has three main limitations one of which is controller workload limitations which reduces the efficiency of air traffic management and airspace usage. Due to the stressful nature of emergencies, air traffic controllers can easily become overwhelmed, thus such incidents are examples of events that contribute to controller workload limitations. EOCs would provide an organized and proficient manner of dealing with emergencies without overwhelming air traffic controllers managing the affected sectors. They would also contribute to the dynamic and adaptable structure that DACs provide in improving airspace practices and management. An example of how EOCs can do this is by having protocols that immediately redirect non-emergency aircraft to other sectors, hence allowing controllers to focus specifically on a distressed aircraft. This eradicates the potential danger of distressed aircraft affecting other aircraft in the sector while ensuring that controllers are not overwhelmed and reducing human error which could potentially worsen the situation. Finally, an added benefit of virtual EOCs is that multiples of them that are not collocated can be set up, which allows for and increases the coordinated efforts of emergency services.

- *EOC Monitoring of Radio Transmissions* – Once an EOC is activated, it monitors radio transmissions between entities involved in attending to an emergency. It is important that the reception of the transmissions is strong and clear to avoid any miscommunication or actions that could further increase the severity of the emergency. This ensures that the EOC is constantly updated on the progress of events during an emergent event. Given that there are multiple attributes that change without notice when carrying out operations, monitoring radio transmissions during emergencies maintains a cohesive emergency response effort among those involved in mitigating the emergency and restoring normal operational conditions.

- *Transfer of Dispatching Functions* – When an EOC is activated, emergency services dispatch is transferred to the EOC for close coordination. In DAC, airport operators and other relevant entities can be included in a sector's allocated EOC to keep them involved while they liaise with emergency response teams. This ensures that the necessary entities required to mitigate the outcome of the emergency are involved and if ground emergency services are required, the necessary personnel can be dispatched.

- *Incident Response Protocols* – A set of procedures can be determined beforehand and activated in the event of an emergency. This ensures organization, accountability, and smooth running of the emergency efforts applied to diffuse the emergency and its outcomes.

Many of the practices listed above are already part of day-to-day NAS operations. Others are considered in plans, such as the Airport Emergency Plans or Operational Contingency Plans. For example, airports are asked to set up emergency plans in which an EOC typically plays a central role in coordinating the emergency with ATC and other affected parties (Department of Transportation, 2010). Emergencies handled through airport EOCs can occur at airports (e.g., emergency landing) or can have their origin outside of the airport but affect airport operations (e.g., natural disasters). In addition to Airport Emergency Plans, there are pre-coordinated Operational Contingency Plans to ensure continuation of services when an ATC facility is temporarily unable to perform all or some of its tasks. Contingency measures include temporary transfer of responsibilities to other facilities and communication to airspace users through Notice to Air Missions (NOTAMs) (Department of Transportation, 2020). It is worth mentioning that an airspace that uses DAC needs to incorporate failure of DAC system components into contingency plans.

A commonly used tool for air traffic related emergency management in the NAS are Temporary Flight Restrictions (TFR). In case of a disaster, hazard, or special event, a TFR can be issued by the FAA to protect persons or property on the ground or in the air. TFR can also be used to create a safe environment for disaster relief operations. Examples of scenarios that

may warrant the use of TFR include volcanic eruptions, nuclear accidents, hijackings, and the need for relief activities after earthquakes, floods, or other disasters. TFR are implemented through NOTAMs (Department of Transportation, 2021).

While there are well established emergency procedures for traffic management and at airports and other facilities in the NAS, emergency situations are often accompanied by a sudden change of air traffic patterns. DAC can make emergency management more efficient and effective by dynamically adjusting the airspace and airspace management, thus equally balancing workload across available resources, so that bottlenecks are reduced.

2.1.5 Application of Ground Transportation Emergency and Safety Procedures in DAC

To the best of our knowledge, there is little research or publicly available knowledge on handling emergency situations involving air traffic systems. This concerns all three possible emergency situations described above (i.e., airspace users and/or the airspace system cause the situation, airspace users and/or the airspace system are impacted by the situation, and airspace users and/or the airspace system are needed for evacuation or relief). Examples include rapidly changing traffic patterns, sudden change of airspace user demand on certain routes, and best practices for rerouting airspace users to avoid disaster zones. However, comparable emergency situations have been thoroughly researched for ground transportation systems. Common characteristics between air transportation emergencies and ground transportation emergencies are the temporary loss of infrastructure, such as roads, traffic lights, ATC services, and runways, due to accidents or disasters. Often, traffic needs to be guided around the disaster/accident zone, which is done by blocking roads or by issuing TFR. Furthermore, both air and ground transportation are critical in evacuation, rescue, and relief efforts in the event of a major disaster, such as hurricanes. Due to the lack of information about handling air transportation emergencies, research on ground transportation emergencies is evaluated in this chapter to assess if knowledge on handling ground transportation emergencies can be transferred to air transportation systems.

Road transportation is known to have the highest number of incidents in comparison to other ground transportation systems, as well as water and air transportation. Most of these incidents tend to be fatal; hence, there is a global effort to try and find a way to manage and improve road infrastructure and practices. Some of the shared characteristics of ground transportation emergencies include:

- Fatalities
- Fire explosions
- Severe wreckage of the vehicle or vessel
- Collisions with other vehicles, vessels, wildlife, or other environmental factors, such as icebergs
- Destruction of the surrounding environment due to impact or elements, such as oil spills

Examples of road transportation emergencies include engine failures, brakes malfunctioning, tires blowing out, collision with other vehicles or wildlife, driving off the road or out of a traffic lane, hydroplaning due to severe weather, and failure of brakes among other instances that lead to accidents that are severe and are likely fatal. Examples of maritime emergencies include damage to the vessel's structure due to severe weather, icebergs, reefs, docks, etc.; on board fires; siege by pirates; and active shooters among many other situations.

According to the Occupational Safety and Health Administration (OSHA), a workplace emergency is described as "a situation that threatens workers, customers, or the public; disrupts or shuts down operations; or causes physical or environmental damage" (Occupational Safety and Health Administration, 2018). As discussed in the previous sections, emergencies can be caused by a myriad of things. The priority in the event of an emergency or accident is to minimize the impact of the incident while protecting those involved. It is therefore important to ensure that there are contingencies in place to deal with emergencies prior to their occurrence. This section examines some of the AI applications followed in ground transportation, particularly road transportation, when responding to emergencies that could potentially be applied to DAC emergency protocols.

The application of AI in transportation revolutionizes the sector by providing a superior way to improve current transport systems in terms of safety, efficiency, and sustainability. Recent studies have been carried out to test the application and viability of AI in transportation. According to the World Health Organization (WHO), "road traffic injuries are the leading causes of death for children and young adults aged 5-29 years" and lead to approximately 1.3 million deaths annually (World Health Organization, 2022). As a result, there is a worldwide push for solutions that provide safer roads and systems that reduce deaths or injuries caused by road accidents (Bibi et al., 2021). In a bid to find efficient and long-lasting solutions to this issue, AI is at the forefront of some of the methods being proposed to reduce road traffic incidents as well as improve the road emergency response practices. Currently, the major AI algorithms being investigated, such as Artificial Neural Networks (ANN), Decision Trees (DT), and Support Vector Machines (SVM), have been applied to different fields, such as computer vision. These techniques are being used to determine Crash Prediction Models (CPM) (Caliendo, Guida, & Parisi, 2007), which provide information on factors that promote Road Traffic Accidents (RTA) and the appropriate emergency response actions required to deal with them as well as possibly prevent them. This segment will discuss the application of these algorithms in road emergency response and their use in improving road safety through traffic management and how these applications can similarly be used in DAC.

ANN which are commonly referred to as Neural Networks (NN) (Bishop, 1994) are algorithms that have been inspired by and mimic the complex biological operations of the human brain. Just like the human brain, their architecture consists of a network of nodes and layers whose goal is to work together to recognize patterns, which are used to solve a myriad of problems related to AI. What makes ANN unique when trying to build crash prediction models that highlight factors leading to RTAs is that they do not "require any pre-defined underlying relationship between dependent and independent variables," thus making it a powerful tool in prediction and classification problems which make up most of these models (Chang, 2005). According to Chang, ANN offer a great and better means of analyzing the frequency of accidents on freeways in comparison to traditional models, such as negative binomial regression that has commonly been used in the past (Chang, 2005). Chang uses the two models

to investigate non-behavioral risk factors, such as environmental conditions that play an important role in traffic safety and contribute to freeway accidents. The nonlinearity between variables in ANN increases the algorithm's robustness when creating prediction and interpretable models that are not limited to sensitivity and result fluctuance when variables change. This is particularly beneficial in CPM because "an accident is rarely due to a single risk factor but is rather the outcome of a series of factors" and maintaining the correlation between these variables reduces accuracy in the analysis (Chang, 2005). ANN provide an alternative and more accurate means to analyze risk factors related to freeway accidents and the CPM obtained are used to mitigate the effects of these factors and reduce the occurrence of accidents. Although ANN are powerful algorithms, a major disadvantage is that they are computationally expensive. Despite this, investigations by Silva et al. show that ANN are at the forefront of road CPM that provide information on crash frequency and severity (Silva, Andrade, & Ferreira, 2020). These models use several variables, such as road-environmental conditions, human factors, vehicle type and age, weather, traffic volume, etc., to determine how these factors contribute to the increasing frequency of road accidents and the severity of these accidents depending on which factors are present at the time. The information obtained from these models is then used to establish road safety conditions and practices that could either be enhanced or created as way of reducing road RTA. This analysis compared the application of the backpropagation and Levenberg-Macquardt (LM) algorithms (Chauhan, Dahiya, & Sharma, 2019) in ANN. Backpropagation is a learning algorithm that trains and refines the weights of a NN by computing the gradient descent of the errors from the outputs obtained. The LM algorithm is also a neural network training algorithm that works to minimize a model's sum of squared error. The comparison of the two algorithms by Silva et al. revealed that the application of the LM algorithm along with the sigmoid function improved the model's convergence, thus confirming the validity of using ANN in CPM. Additionally, with an increase in traffic congestion resulting from growth in population and ecommerce, ANN can be used to determine driver behavior and detect accidents on freeways, which can be used to establish road safety policies and transportation decision tools that relieve congestion and traffic incidents (Liu & Shetty, 2021).

ANN can be applied in DAC through Trajectory Based Operations (TBO), which is an "air traffic management concept that enhances strategic planning of aircraft flows to reduce capacity-to-demand imbalances in the National Airspace System (NAS)" (Federal Aviation Administration, 2020). ANN can be used to create prediction models that analyze the capacity-to-demand features of air traffic sectors based on aircraft trajectories and air traffic in real time. It is important to note that "actual aircraft trajectories are modified by air traffic control depending on potential conflicts with other traffic" (Gallego et al., 2019). Given that aircraft trajectories are not fixed and continuously change due to unrelated external factors, such as weather, traffic conflict, unexpected emergencies, etc., creating models that account for these factors as a proactive measure helps to decrease negative effects, such as traffic congestion and potential accidents. This type of model could increase the effectiveness of a dynamic airspace being able to support changes that affect traffic flow and air traffic management, which are major components to the DAC problem. Finally, given that DAC is comprised of time varying features, ANN could be used to apply spatio-temporal graphs that use data from the Automatic Dependent Surveillance – Broadcast (ADS-B) and flight tracking data to determine aircraft trajectories (Han, Wang, Shi, & Yue, 2021), effects of changes to the trajectories (Kumar, Corrado, Puranik, & Mavris, 2021), and outcomes from changes in the trajectories. This would provide insight on how to improve DAC sector operations during regular operations and emergencies, which is what our model investigates.

In addition, DT are a form of supervised machine learning algorithm that makes use of ensemble learning to perform classification and regression tasks (Myles, Feudale, Liu, Woody, & Brown, 2004). The goal of DT is to discover and extract patterns from a given database that can be used to classify and make predictions based on an expected or known output. When looking at their application in ground transportation emergency response and procedures, DT have been widely used to analyze and improve road safety infrastructure as well as practices during road emergencies. DT in the form of random forests can be used to predict road accident severity with a 94% prediction accuracy (Shanmugam, Raheem, & Batcha, 2021). Road accident severity is essential in determining how external factors, such as environment, contribute to the severity of an accident. This is important information when deciding how to

improve infrastructure that promotes road safety and reduces casualties or fatalities resulting from accidents. DT are also used to establish road departure crash mitigation solutions, such as Lane Support Systems (LSS). A road departure crash is one that occurs when a vehicle is operated outside of the edges of a traffic lane, and it is one of the most common causes of road traffic accidents. According to Pappalardo et al., LSS is an automated in-built vehicle system that uses cameras to detect lane marking and alerts drivers when the vehicle is too close to the markings or when it is outside of those markings (Pappalardo, Cafiso, Di Graziano, & Severino, 2021). LSS are a form of road technology that helps reduce accidents while improving motorist's driving behavior and safety. Pappalardo et al., created a predictive model using DT to determine the probability of the absence or presence of faults in LSS based on selected independent variables, such as horizontal curvature of the road, average speed of the vehicle, and the marking coefficient (Pappalardo et al., 2021). The results from this model were used to refine LSS, which is part of the effort behind ITS. DT can be similarly applied in DAC to predict sector saturation based on operational conditions and the expected emergency outcome centered on the present conditions. For example, if a sector is in a mountainous region that tends to have poor visibility, DT can be used to classify the expected emergency outcomes correlated to these conditions. This data can be used to train a DAC model to provide predictions of which sectors would be affected based on current weather reports. The information obtained could then be used to reroute traffic scheduled to be in the affected sectors. This improves the efficiency of the controllers overseeing those sectors, reduces delays, and decreases the potential of aircraft accidents in the affected sectors. Furthermore, to accommodate the fact that research in DAC is still ongoing and there are limited resources in its functionality, we propose combining an ensemble of DT with a 3D graph-based optimization that caters to spatio-temporal consistency and automatically creates segments of sectors that do not meet the base operational requirements of a sector that could be established by a governing body, such as the FAA. These segments would then be passed through a model that identifies the factors that disqualify them from meeting the base operational requirements and provide mitigating solutions to alleviate the situation and restore smooth operation.

Moreover, another algorithm used in road transportation traffic management and emergency response is SVM. SVM is a linear ML model that is supervised and used to perform classification and regression problems (Hearst, Dumais, Osuna, Platt, & Scholkopf, 1998). SVM is a discriminative classification algorithm that can learn either the linear or non-linear decision boundaries in the attribute space which it uses to separate the classes. It is discriminative because it is impacted only by the training instances near the boundaries of the classes. It also does not require any assumptions of the data distribution. SVM has been applied to predict crash injury severity of road accidents. A study carried out by Assi et al. to predict the gravity of traffic injuries in correlation to the severity of an accident combined SVM with the Fuzzy-C Means (FCM) technique uses cluster analysis to subdivide a dataset into two or more clusters and each cluster is considered a fuzzy set and is assigned a membership grade by each training vector that is measured by a membership function (Assi, Rahman, Mansoor, & Ratrout, 2020). The parameters considered in this study included vehicle attributes (e.g., model of the car, age, speed range, etc.) and road attributes (e.g., sharp bends, blind spots, potholes, pedestrian crossings, etc.). Carrying out this analysis provided insight on traffic management strategies, such as speed limits, road architecture, speed bumps, and speeding cameras, that could be improved to reduce the severity of accidents and the resulting injuries. Streamlining these elements of traffic management contributes to improving road safety procedures which when optimize decrease RTA. Additionally, SVM can be used to predict fatalities resulting from a road incident. Gu et al. apply SVM and use non-linear data to predict traffic fatalities (Gu et al., 2018). They also refine the model's prediction accuracy based on parameter optimization through the introduction of particle swarm optimization, which finds the optimal parameters for the model (Gu et al., 2018). This approach and use of SVM can be used to optimize the prediction parameters used in DAC to determine the severity of sector traffic saturation under different conditions.

Finally, another way that AI is being applied in transportation is to reduce traffic-related fatalities on roads in urban areas through an automated scheme that applies computer vision methods (Bustos et al., 2021). The motivation behind the Bustos et al. research and the application of computer vision in reducing RTA and fatalities is the fact that majority of the

urban plans in most cities do not account for the safety of pedestrians, yet they are the most exposed and vulnerable to road risks than any of the road users at any given time (Bustos et al., 2021). The use of computer vision in managing RTA introduces a modern solution to decreasing the number of car accidents and fatalities that happen annually through road transportation. This is achieved through the "adaptation and training of a Residual Convolutional Neural Network to determine a hazard index for each given urban scene, as well as interpretability analysis based on image segmentations and class activation mapping on those same images" (Bustos et al., 2021). This method of AI application provides an avenue to digitally map hazardous regions within urban areas that promote road accidents. This information can be used to improve urban plans that create road networks that are safer and easy to improve digitally as urban areas evolve. This technique can be applied to DAC, which is dynamic and constantly changing, by applying it to map sectors and provide information on airspace regions that are prone to emergencies due to factors such as terrain, weather, traffic, etc.

2.2 Dynamic Airspace Configuration (DAC) Overview

2.2.1 Airspace Capacity Management and Airspace Structures

Airspace capacity is determined by Air Traffic controller workload (Lee et al., 2008). In both the U.S. NAS and in the European airspace, controlled airspace sectors can be combined with others or split into smaller controlled airspace sectors to balance the workload equally across available ATC resources and sectors (Zelinski & Chok, 2011).

A controlled airspace sector is operated by a small team of controllers (Gianazza, 2010; Lee et al., 2008). A controlled airspace sector is comprised of one or more airspace modules. These airspace modules are usually referred to as sectors. To avoid confusion with the term 'controlled airspace sector,' we adapt the naming convention from Gianazza (2010) and refer to an individual sector as 'airspace module.' We refer to a set of one or more airspace modules that are controlled by one team of controllers as 'controlled airspace sector.'

It is important to note that Air Traffic controllers need training to work on a specific set of airspace modules. Controllers are trained in certain regions or areas of specialization which are usually comprised of groups of airspace modules (Zelinski & Chok, 2011).

Airspace configuration describes the partitioning of controlled airspace sectors (Gianazza, 2010). Configuration schedules are generated days in advance (Zelinski & Chok, 2011) to outline the planned airspace configuration and triggers to changes of the configuration prior to operations (Lee et al., 2008). During operations, ATC managers may decide to tactically change the pre-planned airspace configuration to better balance the workload across controlled airspace sectors (Gianazza, 2010).

Since tactical changes to the airspace configuration schedules are taken by ATC managers, these changes depend on the manager's individual experience (Kopardekar, Bilimor, & Sridhar, 2007). In addition to this, these configurations are based on estimated and not actual demand (Gianazza, 2010). Researchers and practitioners call for "more fluid and dynamic airspace structures that are able to accommodate changes in both traffic and user demands" (Lee et al., 2008).

2.2.2 Introduction to Dynamic Airspace Configuration

DAC is an operational paradigm in which the boundaries of controlled airspace sectors are dynamically reconfigured to meet the airspace user demand and increase available airspace capacity limits (Zelinski & Chok, 2011). The partitioning of controlled airspace sectors and the ATC resource allocation to these controlled airspace sectors is changed with the goal of equally balancing the workload across available Air Traffic Controllers (Gianazza, 2010). Doing so enables ANSPs to better accommodate user-preferred trajectories, i.e., trajectories that take prevailing weather patterns and other security and environmental constraints into account (Kopardekar et al., 2007).

There are three major components to DAC: the overall organization of the airspace, dynamically adapting airspace to meet airspace user demand, and a generic airspace design (Lee et al., 2008).

First, restructuring today's airspace may be necessary to build a solid foundation to dynamically adapting airspace. Many of today's airspace structures are based on historical use profiles and may not be ideal for today's demand pattern (Lee et al., 2008). Furthermore, the widespread availability of technologies, such as GPS and ADS-B, creates possibilities that may require changes to the airspace structure (Kopardekar et al., 2007). Other approaches to DAC require changes in the way that ATC works with airspace module boundaries (Klein, Rodgers, & Kaing, 2008).

The second part of DAC is the dynamic adaptation of airspace on a tactical level to reduce demand and capacity imbalances. Researchers differentiate two approaches (Zelinski & Chok, 2011). The first of these approaches is dynamically creating entirely new controlled airspace sectors for every period. Controlled airspace sector boundaries are newly created every time without relying on pre-existing structures. This approach is sometimes also referred to as dynamic sectorization. This is not desirable from an ATC perspective since ATC personnel cannot familiarize themselves with the airspace (Sergeeva, Delahaye, Mancel, & Vidosavljevic, 2017). The second and more desirable approach from an operational perspective is to use existing building blocks (e.g., airspace modules) which can be dynamically combined to form a controlled airspace sector (Sergeeva et al., 2017).

Third, a generic design of the airspace would promote interchangeability between facilities so that more controllers can manage different airspaces (Lee et al., 2008). However, it is important to point out that the airspace also consists of highly complex and congested airspaces, such as the New York Center for which interchangeability is considered unrealistic and impractical (Kopardekar et al., 2007).

Taking all three parts together, the concept of DAC offers promising approaches to maximize throughput by flexibly adjusting airspace configurations to weather and other constantly evolving constraints. The remainder of this review focuses on the second part of DAC, i.e., approaches and algorithms to dynamically adjust airspace configuration.

2.2.3 A Review of Promising Experimental Studies and Simulations

To better meet the dynamic traffic demand in airspace management, one approach is to redraw airspace module boundaries at every time step, known as dynamic sectorization. However, this approach has operational drawbacks due to safety requirements and the limited time for air traffic controllers to familiarize themselves with new airspace configurations. Therefore, this section will focus on approaches that utilize fixed airspace modules as building blocks. In particular, we will review two previous research frameworks that represent different approaches to solving the DAC problem. These approaches have shown promising results and are worth exploring further. By using fixed airspace modules, these frameworks provide a stable foundation for air traffic controllers to work with, thus avoiding the safety risks associated with constantly changing airspace configurations.

2.2.3.1 Forecasting workload and airspace configuration with neural networks and tree search methods (Gianazza, 2010)

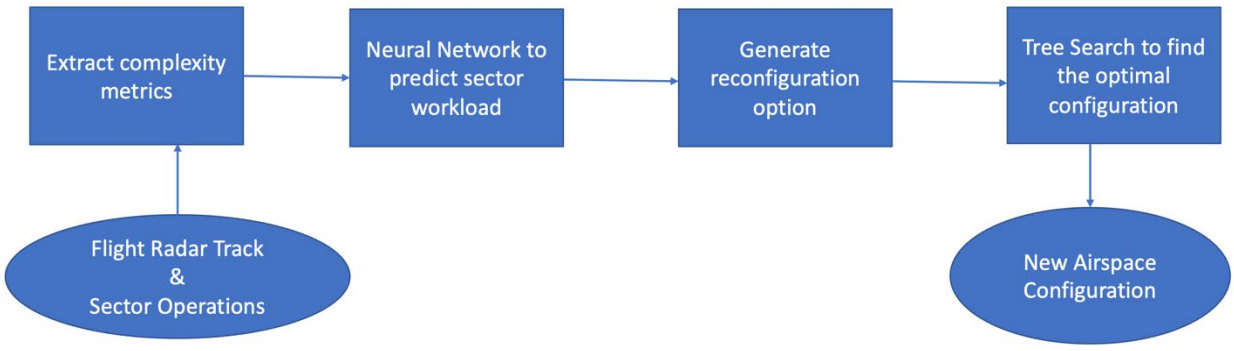


Figure 1: Flowchart of Gianazza's model (Gianazza, 2010)

Series research have been done by David Gianazza with a focus of improving airspace configuration. In their work, a promising framework is proposed to realize DAC. Here is the brief idea introduced in their research, the flowchart is shown in Figure 1: First the flight radar track and sector operation history are used as input data. Then some metrics are extracted from the raw data and feed into a neural network shown in the feature selection section. The neural network is trained to provides a workload indication for the air traffic control sector, whether the workload in this sector is high, normal, or low. With this prediction, the algorithm generates different new configurations by splitting sectors into several smaller airspace modules when the workload is high or merged with other sectors when the workload is low. Next, the tree search methods explore all possible partitions, while restricting them to being operationally valid. They make sure the algorithm will build an optimal airspace partition where the workload is balanced across the sectors and use the restrictions to lower the reconfiguration cost.

2.2.3.1.1 Data Source

Data used to conduct experiments are sampled from five French air traffic control centers on June 2, 2003. The radar tracks of the traffic and sector opening archives are recorded as raw data. The metrics used for training the model are then computed based on the recorded flight trajectories and sector geometry very minute of that day. The labels are in one hot encoding form, where:

- $d = (1,0,0)^T$ stands for low workload where sector needs to be merged,
- $d = (0,1,0)^T$ stands for normal workload where no change needs to be made,
- $d = (0,0,1)^T$ stands for high workload where sector needs to be split.

Later when the model is used to predict future workload, instead of real-time data, flight plans are used to simulate flight trajectories, which are then used as raw data to calculate the features. The bias of simulated data and real data are discussed in the Results and Discussion section.

2.2.3.1.2 Feature Selection

In the previous publication (Gianazza & Guittet, 2006), they explained the details about how they selected the most relevance features to predict air traffic controllers' workload. A dependent variable is selected to represent their workload, and the model needs to be trained in order to maximize the correlation of the selected metrics with that dependent variable. An assumption made here is that splitting and merging history of an ATC sector can represent the controllers' workload, because sectors will be splitting into smaller sectors when the controllers' workload is too heavy or merged when it is too light. Another advantage of sector opening history is that it is stored in air traffic control centers, which does not require extra experiments to collect data. Thus, the sector status is the dependent variable, which belongs to either merge, split, or normal.

To select metrics from 28 candidates, principal component analysis is conducted to reduce the feature dimension and AIC is used to select the most relevant components. Next, Bayesian Information Criterion (BIC) (Neath & Cavanaugh, 2012) is used to find the best metrics in each relevant component. As a result, the following 6 metrics are chosen:

- Sector volume V
- Number of aircraft within the sector Nb
- Average vertical speed avg_vs
- Incoming flows with time horizons of 15 minutes and 60 minutes (F15, F60)
- Number of potential trajectory crossings with an angle greater than 20 degrees (inter_hori18).

The metrics are smoothed by moving average to avoid too frequent reconfiguration caused by high variations. The author mentioned that although these 6 features are relatively basic and intuitive, the result shown that they are good enough for this task. A possible explanation for this is that our labels are not fine-grained, which only have three class. More sophisticated metrics are needed to capture small variations in the workload.

2.2.3.1.3 Model Definition

As shown in Figure 1, the main component of the Gianazza model is the neural network, which is used to predict the air traffic controllers' workload and used as an indicator to change the sector status, and the tree search model to find the best reconfiguration.

The neural network has a basic architecture of 3 layers: an input layer with 6 units to input 6 features, one single hidden layer with 15 units and an output layer with 3 units representing 3 classes. The output layer has the Softmax activation function (Gao, Liu, & Lombardi, 2020), the output can be considered as the posterior probabilities of each class, where $y = (P_{low}, P_{normal}, P_{high})^T$ and the sum of $P_{low,}P_{normal}, P_{high}$ is 1.

Those workload prediction are then input into the tree search algorithm, which would explore all possible configurations and find the best one according to a cost function. The cost function is used to evaluate each configuration and consider the number of sectors and the workload probabilities as the main criteria. The tree is built from the root node. Each tree node is a list of couples. The first element of the couple is the sectors combined, and the second is the list of valid sectors of that combination but without sectors in other couples. When expending to the next level, adding a new sector into different couples leads to a new branch. In order to improve the search efficiency, they proposed an algorithm called "Branch & Bound." The improvement of it, compared to exhaustive tree search, is that the lowest cost of all leaves developed from this node is compared to the best cost so far. When it is higher than the best cost then this whole branch will be deleted.

When combining the two models together, the algorithm is iterated as below:

```
1. Initial configuration (t=0):
   1 ATC sector ⟵ all elementary airspace modules
2. At each time step (1 minute):
     • Decide if airspace must be reconfigured
       ⟶ check workload in each ATC sector
     • If so, reconfigure airspace:
         – select modules to recombine,
         – explore all combinations,
         – select configuration with minimum cost.
```

Figure 2: Sudo code of the Gianazza's algorithm (Gianazza, 2010)

2.2.3.1.4 Results

The neural network has an overall accuracy of 85%, where 90% for class merge, 68% for class normal and 93% for class split. Overfitting is not likely since the model is not complex and the test results are consistent with the training result.

Experiments were conducted to test its performance. To assess the performance, the predicted configurations are compared to the real ones. There are two measures used to quantify the difference, which are the dissimilarity of the number of sectors between them and the number of configuration changes. First, the model is trained on the sampled data on June 2nd and tested on June 6th and 7th. The result shows that the dissimilarity of number of sectors ae 0.191 on 6th and 0.115 on 7th, which is pretty close to the reality. The number of configuration change is above reality, which has 46 against 33 on 6th and 37 against 33 on 7th. Next the model is tested on the trajectories data simulated from flight plans. Compared to the predicted configuration made by flow management operators, the model outputs are more realistic. The dissimilarity between the model outputs and reality is 0.137, and 0.568 for the flow management operators' prediction on June 7th.

To improve this work, future work can focus on the following points:

- Use real-time sector volume as input data instead of the estimated traffic made from flight plan. Because the real sector volume can be affected by a lot of factors, such as weather and military activities, which are unseen during estimation but in fact have a

considerable influence. Attempts of mixing radar tracks with simulated trajectories may help reduce the bias to improve model accuracy.

- Consider the temporal correlation of the data by input the data as a time series and use more complexed model to make the prediction, promising candidates are RNN (Zaremba, Sutskever, & Vinyals, 2014) and LSTM (Sherstinsky, 2020).

2.2.3.2 Dynamic airspace configuration by genetic algorithm (Sergeeva et al., 2017)

In collaboration with the EU and Eurocontrol, researchers (Sergeeva et al., 2017) proposed to model dynamic airspace configuration as a graph partitioning problem that can be optimized with a genetic algorithm. The authors define two different types of airspace modules. Those airspace modules that "are permanently busy areas with a high traffic load" are designated to be "Sector Building Blocks" (SBBs). Less busy and more generic airspace modules are called "Sharable Airspace Modules" (SAMs). A controlled airspace sector should consist of at least one SBB and multiple SAMs. Instead of incorporating re-configuration into the cost function, this approach ensures stability of the configuration by making the busiest airspace modules (SBBs) a fixed central component of each controlled airspace sector. Only the generic SAMs change from one configuration to the next. The approach works best when the airspace is divided into relatively small SAMs and SBBs. However, results better than the current system can also be obtained without introducing SAMs but by using existing airspace module boundaries and treating those relatively large airspace modules as SBBs.

2.2.3.2.1 Data Source

The approach has been tested successfully on simulated free route trajectories for the Maastricht Area Control Center, which consists of 8 airspace modules. The simulated free route trajectories are based on data from July 11, 2014. Two baseline scenarios are compared to two solution scenarios built by the algorithm. The first baseline scenario is the actual configurations from July 11, 2014. The second baseline scenario is the configurations suggested by the tool used to optimize airspace configurations, the Improved Configuration

Optimizer (ICO) system tool (Guégain & Quinton, 2022). Scenario 1 builds configurations based on the existing 8 airspace modules, each treated as SBB. Scenario 2 is based on newly created airspace module boundaries where the Maastricht Area Control Center is divided into 32 SBBs and SAMs.

2.2.3.2.2 Feature Selection

The genetic algorithm uses a cost function to minimize the following criteria:

- Workload imbalance between all controlled sectors: Workload is computed as the number of controlled flights per defined timeframe. A maximum occupancy count per hour of 360 flights was used in the experiments as target workload.
- Coordination workload: Count of flights that cross a sector boundary within a specified timeframe.
- Number of re-entry flights: Count of flights that enter a sector more than once. Such cases should be avoided because they unnecessarily increase the coordination workload.
- Number short-transit flights: Count of flights that enter a sector for only a short period of time.
- Number of controlled sectors: There is a maximum number of controlled sectors that can be opened in each configuration. This number typically depends on the number of available controllers.

The cost function consists of these five criteria. All criteria are normalized to values between 0 and 1. They are then multiplied by proportion coefficients to control each criterion's individual importance to the overall cost. Furthermore, the following constraints must be fulfilled, whereas the last two constraints are soft ones:

- All airspace modules in a controlled sector must be connected.
- There should be continuity between resulting configurations.
- Shapes of sectors (in a lateral view) such as "stairs" or "balconies" should be restricted.

2.2.3.2.3 Model Overview

The three-dimensional airspace is modeled as a weighted graph in which each node is an airspace module, i.e., either a SAM or an SBB. The weight of each node represents the workload in the respective airspace module. The link to other nodes represents the coordination workload between two adjacent airspace modules.

Since this is an NP hard combinatorial problem, a genetic algorithm is used. While such approaches do not guarantee the correct solution, they have proven to find approximate solutions within a reasonable time. The used genetic algorithm starts with a randomly generated initial population, i.e., several first guesses for a possible solution. A random set of individuals from the first population is picked from which a few best solutions are selected. Then, three possible operators are applied to the individuals of this population: nothing to carry over the same solution to the next generation, crossover, and mutation. Crossover combines the solution of two individuals. Mutation is applied to avoid local minima. The results from these three operations form the next generation. These steps are typically repeated for a pre-defined number of iterations or until another termination condition has been reached.

A possible solution in the genetic algorithm (Whitley, 1994) used for airspace configuration consists of two layers. In the first layer, the number of controlled sectors and their root nodes is defined. After every iteration, this first layer is altered through crossover and mutation operations. Based on this first layer, the second layer uses graph partitioning techniques to allocate all other nodes (SBBs that are not root nodes and SAMs) to one of the root nodes. A special mutation operator is used for the second layer in which the algorithm starts with the worst performing sector (either overload or underload). The algorithm then swaps airspace modules of adjacent airspace sectors to balance their workload.

2.2.3.2.4 Results

The simulated free route trajectories based on Maastricht Area Control Center traffic data from July 11, 2014, are used to compare two scenarios to two baseline scenarios. The first baseline

scenario is the actual configuration used on July 11, 2014. The second baseline scenario is the configuration suggested by the existing tool used for airspace configuration. Scenario 1 uses existing airspace module boundaries. Since the airspace modules are rather large, all of them treated as SBB. Scenario 2 produces a configuration based on a further division of the airspace into 32 SAMs and SBBs located on two altitude layers.

Both solution scenarios show significant improvements compared to the baseline scenarios. Since the quality of workload balancing depends on the number of input modules, scenario 2 produces solutions that are much better balanced in terms of the workload than scenario 1. However, scenario 2 results in solutions that have undesirable geometric shapes like "balconies". Despite this, the authors claim that operational experts have found these solutions to be acceptable. Improving the algorithm to find solutions that are characterized by a more ideal geometric shape could be the focus of future work. In addition to this, the authors identify a more advanced workload metric as a possibility for further improvement.

2.2.4 Summary of literature review

Although various research has been conducted to implement DAC, many of these methods redraw the boundaries of airspace modules frequently, which is not operationally desirable due to the high collaboration cost. An approach that uses fixed airspace modules as building blocks and generates new configurations by combining or splitting these blocks into groups is preferable. However, there are limited research following this idea and we have reviewed the most promising research of them in detail and identified improvements that can leverage the benefits of both methods. In summary, we are inspired to combine the two works mentioned in section 3.3 together by utilizing both temporal and spatial information. On the one hand, Gianazza's approach in 3.3.1 has shown the ability of neural network to predict the air traffic controllers' workload which is very important when evaluating the quality of a configuration and its performance can be further improved if input the features as a time series. On the other hand, the research from Sergeeva et al. in 3.3.2 provide us a new idea to include spatial information of adjacent sectors into the model by constructing the problem into a graph

partition problem. Therefore, we chose the research direction of utilizing both temporal and spatial information.

3 PHASE 1: DELAY PREDICTION WITH SPATIAL-TEMPORAL GRAPH NEURAL NETWORK

We used the genetic algorithm to generate all possible partitions and used a spatial-temporal graph neural network to evaluate the air traffic controllers' workload and used that as the indicator to choose the best configuration. For the first step, we tried to realize the spatial-temporal neural network. We assigned the flight delay prediction task to this model as the evaluation criteria for the following reasons: first, in the previous research on dynamic airspace configuration, choosing the appropriate metrics to evaluate the Air Traffic Controllers workload was critical for the model performance. The predicted total flight delay in an ATC sector could be then used to evaluate the sector workload and used as an indicator when choosing the optimal configuration. Second, once the model was completed, its ability to learn the underlying pattern of the spatial-temporal information could be confirmed, and we would continue to the next stage.

3.1 Rationale

The accuracy of the flight delay prediction task is still restricted mainly due to three factors: the lack of available flight data, the high diversity of the delay causes, and severe imbalance of the delayed flight versus normal flight (Ball et al., 2010). Some of the work feed the data as a time series without taking the spatial information into consideration, which prevents the model from obtaining higher performance (Wang, Bi, Xie, & Zhao, 2022). Because the air traffic system is a complex system with both temporal and spatial information, i.e., the level of flight delay will not only be influenced by the past throughput of the same airport but also be affected by the upcoming flights from other airports. Previous research (Jiang et al., 2022) re-construct the flight data into a graph and proposed a spatial-temporal neural network to predict the average delay of an airport. However, since most of the flights take off and arrive on time, the flight delay data is highly unbalanced. Previous research has shown that with different sequential models, such as RETAIN (Choi et al., 2016) and recurrent neural network

(Zaremba et al., 2014), the contrastive loss (Khosla et al., 2020) could improve the model performance than cross entropy loss for data sets with the severe class imbalance and the results are quite robust in different tasks (Chaitanya, Erdil, Karani, & Konukoglu, 2020) (Taleb, Kirchler, Monti, & Lippert, 2022) (You et al., 2020).

To fulfill the research gap mentioned, we used a spatial-temporal graph neural network to classify the flight delay of the air transportation network in the United States and used the contrastive loss to tackle the data imbalance problem. We used the real flight data, retrieved from the Airline On-Time Performance (AOTP) Dataset, which provides a more realistic evaluation of the model's performance. There are several key steps included in developing our model. First, in order to apply contrastive loss, we reformatted the data to a classification dataset by setting a threshold where only the top 20% of flights with the most severe delays are labeled as delay. Second, the spatial-temporal graph neural network is built with two main modules. The graph convolution layer was used to capture the pattern of the flight propagation. The temporal convolution was leveraged to capture the time dimension node features aggregated by the graph convolution layer. Next, the conservative loss was implemented. The objective of the model is to minimize the contrastive loss by pulling the similar pairs together and pushing dissimilar pairs apart. Finally, the model gave binary output of 1 and 0, which represents the prediction result of delay and not delay separately. Data source, model structure and experiments are discussed in the following section in this chapter.

3.2 Data Source and Preprocessing

The data set used is from the Airline On-Time Performance (AOTP) Dataset provided by Bureau of Transportation Statistics. It contains the detailed non-stoop flight information of year 2016 in the United States. However, not all the information in the dataset is useful. First, we cleaned the dataset by removing the irrelevant information and only keeping the below features for each data record, including flight date, flight distance, departure and arrival airports code, actual time and delay time when departed and arrived, time of taxiing in and taxiing out. Flight records that miss those features were removed.

Next, we processed the data into spatial and temporal features to feed into the model. The spatial feature is a node graph which is represented by an adjacent matrix. To generate the spatial feature, each node is an airport and individual flights were aggregated as pairs. Then the adjacent relationship of each pair of nodes was calculated as the total number of flights between two nodes. The temporal information was extracted by inputting data as a time-series. For each node, the departure and arrival data were separated. At each time point we assigned a sequence number to each data record. Then we regrouped the data records based on the node and the sequence number and calculated three features for every thirty minutes. The features are the average number of flights, delay time and taxing time to evaluate the airport capacity and mobility. For each node, the feature values are the sum of both departure and arrival data. As a result, the data is in a form of a 3-D array $X \in R^{N*L*F} = R^{285*17275*3}$, where N is the number of nodes, L is the length of the time sequence and F is the number of features.

Finally, the data set was transformed into binary classes of delay and not delay. By setting a threshold of delay time acceptance, flights that had delay time longer than or equal to the threshold were labeled as delay and the rest were labeled as not delayed. In the experiments, the threshold is set to the top 50% with longest delays.

3.3 Model Overview

3.3.1 Problem Statement

The air traffic network is defined as a graph $G = (N, E, A)$ to gather spatial information. N is a set of nodes representing each airport. E is a set of edges which represent the connection of two nodes. $A \in R^{n*n}$ is the adjacent matrix where each element is the number of flights between a pair of nodes, n is the number of nodes. Each node has F features, which are the average number of flights, delay time and taxing time, to indicate its capacity and mobility. The input data is defined as $X = (X_1, X_2, \dots, X_L)^T \in R^{N*L*F}$, where N is the number of nodes, L is the length of the time sequence and F is the number of features. The predicted delay status in the future at time step τ is defined as $Y_\tau = (y_\tau^1, y_\tau^2, \dots, y_\tau^n)^T \in R^{n*1}$. As a classification problem, y_τ^i is 0 if it will not be a delay or 1 if there will be a delay.

With above definition, the flight delay classification problem is formulated as: At time step t, given the historical data records in the past k time steps until t, the objective is to classify whether there will be a delay or not at the next m time steps:

$$\hat{Y}_{t+m} = P(Y_{t+m} | X_{t-k}, \dots, X_t)$$

3.3.2 Graph Convolution Network

The air transportation network is modeled as a graph, in order to learn the spatial dependency of different airports, the graph convolution layer (Wu et al., 2019)was used to capture the pattern of the flight propagation among different airports. According to Henaff, Bruna, and LeCun (2015), the graph convolution is defined as:

$$g_\theta * x = g_\theta(L)x = g_\theta(U\Lambda U^T)x$$

Where g_θ is the kernel, L is the Laplacian matrix and its eigenvalue decomposition is $L = U\Lambda U^T$, Λ is a diagonal matrix and U is Fourier basis. Considering the large amount of nodes of airports which result in huge amount of parameters, to improve the computation efficiency, Chebyshev polynomials were leveraged to approximate the computation (Simonovsky & Komodakis, July 2017), which is:

$$g_\theta * x \approx \sum_{k=0}^{K-1} \theta_k T_k(\hat{L})x$$

Where θ_k is the polynomial coefficients that will be learned, $\hat{L} = \frac{2}{\lambda_{max}}L - I$, λ_{max} is the maximum eigenvalue of Laplacian matrix. T_k is the recursive Chebyshev polynomials and defined as:

$$T_0(x) = 1$$
$$T_1(x) = x$$
$$\dots$$
$$T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$$

As a result, the final definition of the graph convolution is:

$$GConv = \sigma(g_\theta * x) = \sigma(\sum_{k=0}^{K-1} \theta_k T_k(\hat{L})x)$$

Where σ is the activation function.

3.3.3 Temporal Convolution Network

The temporal convolution was leveraged to capture the time dimension node features aggregated by the graph convolution layer. This temporal convolution is a standard 1-D convolution and defined as:

$$TConv = \sigma(\Phi * GConv)$$

Where the convolution operation, Φ is the convolution kernel to be learned and σ is the activation function.

3.3.4 Temporal Attention Mechanism

Attention mechanism (Niu, Zhong, & Yu, 2021) was adopted to assist the spatial-temporal convolution network, which is used to dynamically capture the temporal correlations of the air traffic for each airport (Guo, Lin, Feng, Song, & Wan, 2019). The attention score matrix is defined as below:

$$Att = V \cdot Sigmoid({h_1}^T W h_2 + b)$$

Where V, W, and b are learnable parameters, ${h_1}^T$ and h_2 are the product of the current input and different learnable kernels. Each element in this attention matrix can be considered as the level of dependencies between two timesteps. Then the temporal attention was obtained after applying a Softmax function, which is used for normalization. It was directly applied to the input then fed into the model, which could capture relevant temporal information adaptively.

3.3.5 Contrastive Loss

Previous research (Wanyan et al., 2021) has shown that with different sequential models, such as RETAIN (Choi et al., 2016) and recurrent neural network (Zaremba et al., 2014), the contrastive loss could improve the model performance than cross entropy loss (Zhang & Sabuncu, 2018) for data sets with the severe class imbalance and the results are quite robust in different tasks (Chaitanya et al., 2020; Taleb et al., 2022; You et al., 2020). The main concept

is to minimize the contrastive loss by pulling the similar pairs together and pushing dissimilar pairs apart. The contrastive loss is defined as (Hadsell, Chopra, & Lecun, 2006):

$$L(W, Y, X_1, X_2) = \frac{1}{2}(1 - Y)(D_w)^2 + \frac{1}{2}(Y)max(0, m - D_w)^2$$

Where $X1$, $X2$ is a pair of input vectors, Y is a binary label assigned to this pair of input and Y is 0 when the pair belong to the same class, Y is 1 when the pair belong to different classes. D_w is the function that calculates the distance between the pair, here the euclidean distance is used. m is a positive margin and only dissimilar pairs within the margin contribute to the loss.

3.3.6 Final Framework Structure

The framework of the spatial-temporal graph neural network model is shown in Figure 3. The spatial-temporal convolution block is composed of a graph convolution layer and a temporal convolution layer, which realize the main function of learning both from both the spatial and temporal information simultaneously. By stacking multiple blocks together, the model is capable of capturing more complex dynamics in the data. However, there is a tradeoff between the computation efficiency and the model performance. At the end, a fully connected layer is appended to compile the result into classification prediction.

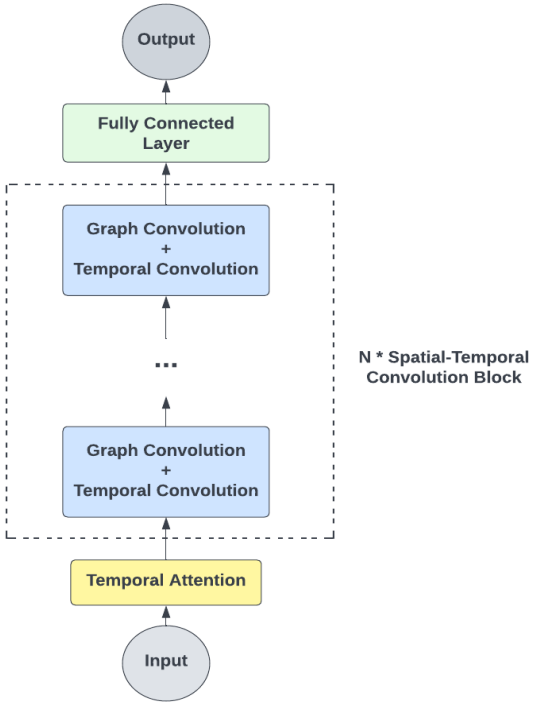


Figure 3: Framework of Spatial -temporal graph neural network model

3.4 Experiment and Result Analysis

For the experiment, the task was to use the air traffic data of the past 9 hours to predict if there will be a delay after 2 hours. Thus, the input is a 4D array, where the added dimension is the chopped time sequence with 18 time steps. The label is binary, where 1 represents there will be a delay after 2 hours for this airport and 0 means not. All the input data is processed with a zero-mean normalization. The whole dataset is divided into training set, validation set, and testing set with a ratio of 6:2:2.

The model includes 3 spatial-temporal blocks. The max order of Chebyshev polynomials K is set to 3. The keener size for temporal convolution is also set to 3. The activation function for the spatial and temporal convolution layer are Rectified Linear Unit (ReLU) (Agarap, 2018). The batch size is set to 64 for training. The optimizer is Adamax (Llugsi, El Yacoubi, Fontaine,

& Lupera, 2021) starting at 0.0003. Figure 4 presents the confusion matrix of the classification test result:

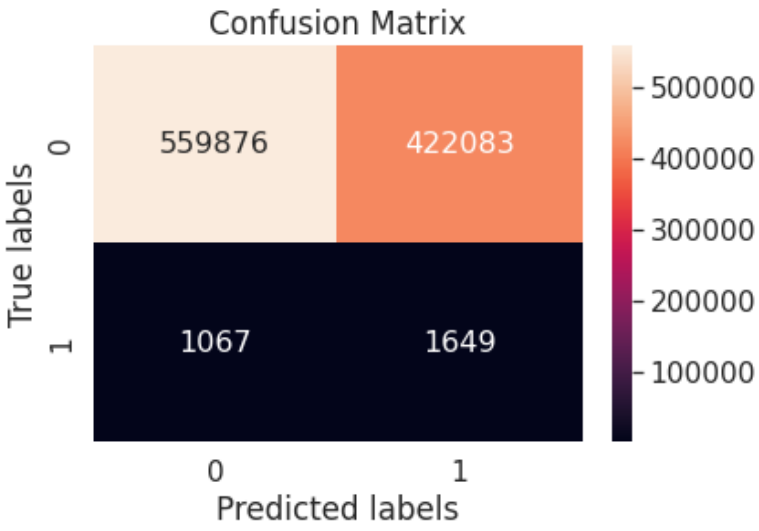


Figure 4: Confusion Matrix of the classification test

The following equations define the accuracy and recall score of this classification test:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} = \frac{559876 + 1649}{559876 + 422083 + 1067 + 1649} = 0.57$$

$$Recall = \frac{TP}{TP + FN} = \frac{1649}{1649 + 1067} = 0.61$$

$$Precision = \frac{TP}{TP+FP} = \frac{1649}{1649+422083} = 0.003$$

The result is not as good as expected, probably due to below reasons and some exploring are undergoing to improve the performance: first, considering the complexity of the model, the data sample size is limited, including more data can provide the model with more information to learn the underlying pattern, considering there are only about 9,000 delay flight right now. Larger high-quality dataset with similar features for the other years and nations would be better.

Second, according to Federal Aviation Administration (FAA), from 2008 to 2013, there are 69% of the flight delay caused by weather. The result in previous study (Sun Choi, Young Jin Kim, Briceno, & Mavris, Sep 2016) also showed that combining weather data with a machine learning model could improve the performance of flight delay prediction. We have not included the weather information in this experiment initially, because we are looking forward to seeing to what extent does the combination of spatial and temporal information help address this classification problem since it shows great potential on the regression task. However, seen from the results of our experiment, it needs some other information to improve the performance. Finally, while constructing the adjacent matrix which contains the spatial information, the weight between the nodes is now only using the number of flights without considering the distance. But, when the flight distance is longer, the flight is more likely to be affected, because the impact of a traffic event could be shared within a region (Jiang et al., 2022). We are looking into aspects to further improve the model's performance.

4 PHASE 2: SPECTRAL CLUSTERING FOR DYNAMIC AIRSPACE CONFIGURATION

4.1 Overview

As discussed in section 3.2.1, a dynamic approach to airspace configuration is needed that avoids redrawing airspace boundaries for each reconfiguration. To achieve this, we applied a novel approach that utilizes fixed airspace modules as building blocks that can be combined or split based on workload. Our approach is based on a spectral clustering algorithm that constructs a graph representation of the airspace using geographic information.

In this graph, each module was represented as a node, and the edges between nodes were weighted based on the connection and workload of the modules in the current time window. The spectral clustering algorithm was then applied to predict the new configuration for the next time window based on the current workload. To promote efficient traffic flow, the adjacency matrix was constructed in a way that traffic centers with high traffic load are pushed away from each other, making the light-workload modules more attractive to the traffic center. Our

algorithm automatically identifies this pattern, where busy modules are in the center and surrounded by less busy ones. It then separates the busy sectors and combines adjacent idle modules to create new sectors, maximizing air mobility and reducing congestion during emergency evacuations. The results of our experiment outperform static airspace configuration.

Our proposed approach presents a promising solution for implementing dynamic airspace configuration in real-world scenarios, particularly in emergency evacuation situations. This is achieved by first clustering airspace into modules that prioritize heavy workload centers surrounded by modules with spare resources. Then reallocating ATC resources within each cluster, we can balance the workload among modules to maintain maximum traffic capacity. Our algorithm employs fixed airspace modules as building blocks, eliminating the need for frequent airspace boundary redraws and facilitating efficient ATC resource allocation in response to changing traffic loads. Overall, this approach demonstrates significant potential for realizing dynamic airspace configuration in practical applications.

4.2 Model Overview

4.2.1 Data Preprocessing and feature selection

Flight Delays and Cancellations were published by the U.S. Bureau of Transportation in 2015 (Department of Transportation Bureau of Transportation Statistics, 2017). This dataset includes the statistics tracking the on-time performance of domestic flights operated by large air carriers. Data preprocessing was conducted before feeding into the model. The original data has total of 30 attributes, as shown in Figure 5:

Column	Dtype
YEAR	int64
MONTH	int64
DAY	int64
DAY_OF_WEEK	int64
AIRLINE	object
FLIGHT_NUMBER	int64
TAIL_NUMBER	object
ORIGIN_AIRPORT	object
DESTINATION_AIRPORT	object
SCHEDULED_DEPARTURE	int64
DEPARTURE_TIME	float64
DEPARTURE_DELAY	float64
TAXI_OUT	float64
WHEELS_OFF	float64
SCHEDULED_TIME	float64
ELAPSED_TIME	float64
AIR_TIME	float64
DISTANCE	int64
WHEELS_ON	float64
TAXI_IN	float64
SCHEDULED_ARRIVAL	int64
ARRIVAL_TIME	float64
ARRIVAL_DELAY	float64
DIVERTED	int64
CANCELLED	int64
CANCELLATION_REASON	object
AIR_SYSTEM_DELAY	float64
SECURITY_DELAY	float64
AIRLINE_DELAY	float64
LATE_AIRCRAFT_DELAY	float64
WEATHER_DELAY	float64

Figure 5: Flight data attributes

However, not every attribute was recorded for each flight, thus columns with more than 25% missing values are removed, which including the weather delay. Also, there is strong correlation relationship between departure delay and arrival delay. Since there are a lot of factors that are not related to airport workload could influence the arrival time of a flight, such as weather, departure delay is used as the workload indicator. For the rest of data, only the related attributes are kept, including the scheduled date, airline, origin and destination airport, departure time and delay time. Canceled or diverted flights are removed. To reduce computation load, departure delay is transformed into binary label, 0 is non-delay flight and 1 is delay flight.

We focused on 21 airports in Florida to test and evaluate the model performance. The location of 21 airports in Figure 6 are used to construct the graph as shown in Figure 7:

name	type	local_code	latitude_deg	longitude_deg
Pensacola International Airport	medium_airport	PNS	30.473400	-87.186600
Destin-Fort Walton Beach Airport	medium_airport	VPS	30.483200	-86.525398
Northwest Florida Beaches International Airport	medium_airport	ECP	30.357106	-85.795414
Tallahassee Regional Airport	medium_airport	TLH	30.396500	-84.350304
Jacksonville International Airport	large_airport	JAX	30.494101	-81.687897
Gainesville Regional Airport	medium_airport	GNV	29.690100	-82.271797
Daytona Beach International Airport	medium_airport	DAB	29.179899	-81.058098
Orlando Sanford International Airport	large_airport	SFB	28.777599	-81.237503
Orlando International Airport	large_airport	MCO	28.429399	-81.308998
Tampa International Airport	large_airport	TPA	27.975500	-82.533203
St. Petersburg Clearwater International Airport	medium_airport	PIE	27.910200	-82.687401
Melbourne Orlando International Airport	medium_airport	MLB	28.102800	-80.645302
Vero Beach Regional Airport	medium_airport	VRB	27.655600	-80.417901
Sarasota Bradenton International Airport	medium_airport	SRQ	27.395399	-82.554398
Charlotte County Airport	medium_airport	PGD	26.920200	-81.990501
Southwest Florida International Airport	large_airport	RSW	26.536200	-81.755203
Palm Beach International Airport	large_airport	PBI	26.683201	-80.095596
Fort Lauderdale Executive Airport	medium_airport	FXE	26.197300	-80.170700
Fort Lauderdale Hollywood International Airport	large_airport	FLL	26.072599	-80.152702
Miami International Airport	large_airport	MIA	25.793200	-80.290604
Key West International Airport	medium_airport	EYW	24.556101	-81.759598

Figure 6: 21 Florida Airport location data

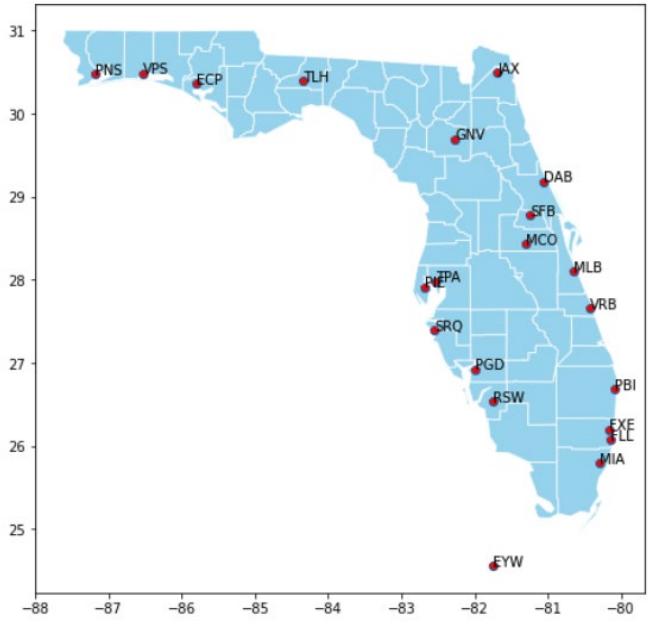


Figure 7: Location of 21 Florida Airports

4.2.2 Airspace Graph Construction

To utilize the geographic information and constrain the clustering algorithm to only reconfigure a sector using modules around it, the airspace was constructed as a graph, represented by adjacency matrix. In the graph, each smallest undividable module is the node and the edge weights between nodes are determined by the connection and the workload of the two modules. For two connected nodes, the edge weight measures the connectivity of them, in another word the distance in the graph space. This connectivity is inversely related to the workload of the nodes: as workload increases, distance between nodes also increases. Thus, when both nodes are busy, they are further apart on the graph. However, each of the become more attractive to neighboring nodes with lighter workloads that are geographically close to them.

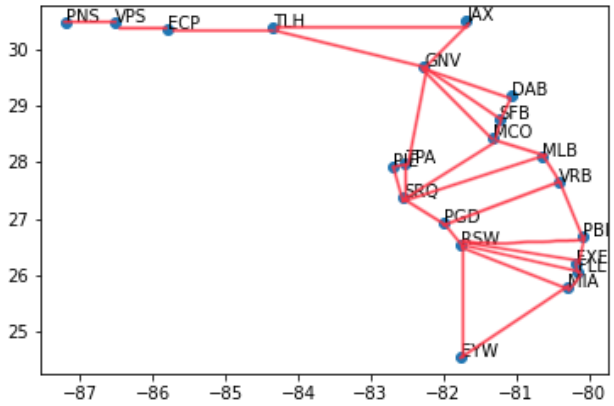


Figure 8: Connection of 21 Florida airports

There are two steps to constructing an adjacency matrix.

Step 1: Generating Initial Airport Adjacency Graph (IAG). We identify if two nodes are connectible based on the geographic location. Based on the location of 21 airports in Florida, the assumption of the 21 Florida airports connection is shown in Figure 8. In a real situation, a pre-tactical evacuation route should be prepared and used as a connection assumption. Mathematically, we define that if two airports, v_1 and v_2 are connected if v_2's is the closest neighbor geographical neighbor of v_1 at the same azimuth.

Step 2: Creating Hybrid Airport Adjacency Graph (HAG): We assigned edge weight for each connection of airports. To balance the workload of each sector, on the graph, airports that are less busy should have stronger connections to each other, and the busy airports should be far away from each other to avoid being clustered together. In other words, the edge weight is negatively related to the total workload quantified by the predicted delay flights and delay ratio in the future 2 hours between the nodes. A modified radial-based kernel is chosen to encode the workload into edge weight, defined as below:

$$load_{ij} = \frac{d_i + d_j}{f_i + f_j}$$

$$w_{i,j} = B^{((1-\lambda)\cdot(100\cdot load_{ij})+(\lambda)\cdot d_{ij}-shift)}$$

Where *i* and *j* are two connected nodes, *d* is the number of delayed flights in this time window, *f* is the total number of flights in this time window. Thus, $load_{ij}$ is the percentage of delayed flights of node i and j and normalized between 0 and 1. We only use the percentage of delayed flights as an indicator of the ATC workload, more comprehensive workload can be explored in the future and substituted here. The variable d_{ij} is the geographical distance between the two airports. λ is the geographical weight factor in balancing between considering geographical distance and gross delay ratio, this factor is incrementally tuned in our program. $B \in (0,1)$ is the base. Both *B* and shift are hyperparameters that can be fine-tuned, in our experiment, we use 300 as it is numerically more stable. where *B* is to satisfy the negative relation between workload and the edge weight and shift is to reduce computation load. Mathematically, the larger the edge weight between two nodes (airports), the more likely the two nodes should be connected and stay in the same cluster.

4.2.3 Spectral Clustering on Hybrid Airport Adjacent Graph

After constructing the Hybrid Adjacency Graph based on the prediction of futuristic delayed and incoming flights, spectral clustering is performed to cluster the nodes into different clusters, based on the strength of connections in the graph. Our assumptions here are stated as follows:

***Assumption I**: In emergency evacuation or other busy scenarios, all flight plans are known at least two hours in advance.*

In other words, we try to find an optimal plan to allocate non-busy airport's air traffic control resources to assist busy airports which are under emergency with possibly more delays. We believe this method can be generalized not only to emergency situations but can also be used in daily routines when passenger flows are distributed in a temporally and spatially unbalanced manner.

Mathematically, spectral clustering first performs Eigen decomposition of the adjacency matrix of the Hybrid Airport Adjacent Graph to project data from a higher dimension to a lower dimension, then, with only the limited information of the data, the clustering is easier and more accurate because the unimportant information has been eliminated. In general, we use the following steps to perform spectral clustering algorithm (von Luxburg, 2007):

Step 1: Construct Hybrid Airport Adjacency Graph with $\lambda = 0$, in this way, the initial clustering will not consider the geographical location of airports.

Step 2: we calculate the degree matrix of the graph; The degree matrix is a diagonal matrix where the value at entry (i, i) is the degree of node i.

Step 3: we calculate the eigenvalues and eigenvectors of the degree matrix; then we sort them based on the eigenvalues; finally, we perform k-means clustering algorithm (Hartigan & Wong, 1979) with an initial $k = 5$ on first n-mean vectors with nonzero eigenvalues to get the initial clustering result.

Step 4: scan each cluster in the initial clustering result, if any cluster contains more than three airports or with a diameter greater than 100 nautical miles (the typical transmission range of ADS-B transponder). We increase the number of clusters by 1 and simultaneously, increase the geographical weight λ by 0.1. However, the maximum value can only be 0.5.

Step 5: We repeat the clustering procedure as described in Step (1) until all clusters satisfy the criteria defined in Step (4).

Step 4 and 5 makes the clustering process adaptive and this algorithm can gradually evolve to use geographical constraints to create clusters with reasonable spatial size. Therefore, we do not require select dedicate k and λ values for different scenarios.

After spectral clustering on HAG, the airports that are geographically close and with relatively low workloads are combined as a new sector. Simultaneously, the busy airports with more delayed flights will be picked up and isolated. We can then move forward to select and merge the airspace and resources of non-busy airports with busy airports. Our goal of balancing the workload of different airports during emergency evacuations or other busy scenarios.

4.2.4 Fine-Tuning for Dynamic Workload Balancing

In this stage, we aim to merge different airports' governing regions to rebalance the workload of the area when there is an abrupt increase in the travel demand or flight delays. For this purpose, we have the following assumptions:

Assumption II: *the abrupt increase of travel demand under emergency situations could cause significant delay flights. In our experiment, if an airport's delayed flights within the predicted time window are greater than 2 delayed flights per hour (with a regional airport) or with a delay percentage within this time window being 30% (medium or large airports), we then mark this airport as a busy airport and needs external assistance.*

Assumption III*: the nearby airport that used to assist a busy airport should: (a) has fewer delayed flights if the category of the airports is identical, or a lower percentage of delay if the category of the airport is different.*

Based on assumptions II and III, we develop the fine-tuning algorithm for each busy airport as follows:

Step 1: We created a ranked list of busy airports based on (a) a user-defined priority level with a default value of zero, (b) the number of delayed flights, (c) delay ratio and (d) number of scheduled flights within the predicted time window. Here, the user-defined priority level can be filled when there's an emergent situation.

Step 2: We scanned all airports within 100 nautical miles of the busy airport and determine if a specific airport can be merged to assist a busy airport based on these criteria: (a) distance, (b) less number of predicted delayed flights, (c) lower delay ratio. Specifically, we created a ranked list of non-busy candidate airports and picked the closest one.

Step 3: If any two airports are selected as a collaborative pair, we created a new cluster containing only the two airports, to prevent airspace conflict, we also ensure that

there's no other busy airport within the combined airspace before establishing the collaboration relationship.

In general, the algorithm could automatically separate currently busy sectors and combine the adjacent modules that are not busy as a new sector, allowing reallocating more resources from idle sectors to busy sectors to realize Dynamic Airspace Configuration with maximum air mobility.

4.2.5 Evaluation Metric

To assess how well the model balances workload across different sectors in the selected area, we use a metric called the "Regional Unbalanced level." This metric can be used towards the workload of handling regular or delayed flights. This metric can be calculated in two steps:

Step 1: We calculated the average number of non-delayed and delayed flights handled by airports within each cluster

$$F_i = \frac{\sum_{k=1}^{m} f_k}{m} \text{ and } D_i = \frac{\sum_{k=1}^{m} d_k}{m}$$

Step 2: We calculated the variance of the F_i and D_i over all clusters, noted as S. This value helps to quantify the workload unbalance within the whole airspace in scope. n is the number of clusters in the configuration.

$$S_D = \frac{\sum_{i=i}^{n}(D_i - \underline{D})^2}{n-1} \text{ and } S_F = \frac{\sum_{i=i}^{n}(F_i - \underline{F})^2}{n-1}$$

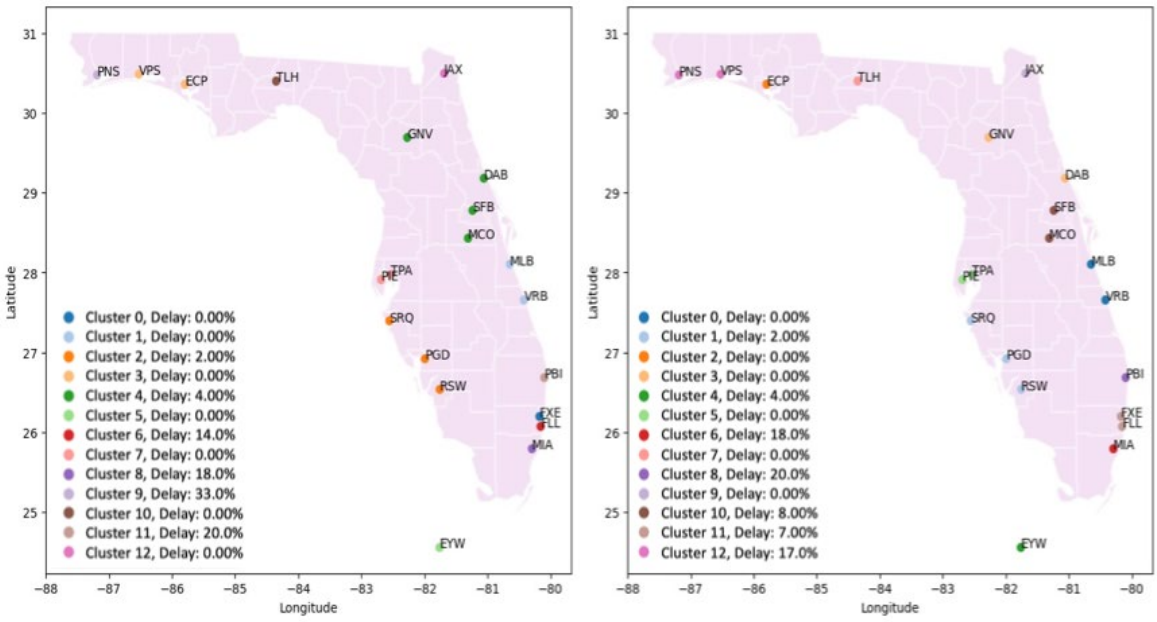
Specifically, high S_D or S_F indicate the workload of handling delayed or regular flights are highly different in across different clusters. We assume that airports within the same cluster are collaborating with each other to handle emerging workloads. Therefore, smaller S_D or S_F are preferred, because we aim to combine busy airports and surrounding idle airports through clustering to achieve a reasonable allocation of ATC resources, that is, to allocate resources from less busy airports in a cluster to busy airports. When the performance of the algorithm is good enough, each cluster can have enough resources, that is, the workload will be reduced, and the number of delayed flights of each cluster will be similar, so the variance will also be reduced.

4.3 Experiment and Analysis

Traffic volume is affected by various factors such as the season, day of the week, and time of day, as well as weather conditions, holidays, and rush hour periods. To assess the model's performance under different traffic loads, we conducted two sets of experiments. The first set evaluates the model under various time windows on the same day, each representing a different level of traffic load. The second set evaluates the model's ability to generalize to different high-traffic days by testing it on the same time windows across multiple days. By conducting these experiments, we can evaluate the robustness and generalization capabilities of the model under various traffic conditions.

4.3.1 Different times on the same day

To achieve a more accurate simulation of traffic conditions during an emergency evacuation where abrupt travel demands are causing huge workloads, the model is tested multiple times on the same day. Specifically, the testing is conducted on December 24th, a day with heavy flight traffic due to the holiday season. Three distinct 2-hour time windows are selected to represent different traffic conditions: 7 AM - 9 AM for low traffic (Figure 9 and Table 1), 12 PM - 2 PM for high traffic (Figure 10 and Table 2), and 7 PM - 9 PM (Figure 11 and Table 3) for medium traffic.



(a) (b)

Figure 9: Clustering and airspace reconfiguration using flight plan and delay prediction for 7:00 to 9:00AM Dec 24, 2015. (a) Airspace allocation based on flight plan and delay prediction. (b) Fine-tuning of airspace configuration for better workload balancing

As in Figure 9a and 9b, the airspace configuration algorithm combines MCO with SFB, FXE with FLL, and PNS with VPS to balance the workload; In the meantime, MIA is isolated independently because there's no non-busy airport within a reasonable range and without airspace overlap with busy airports. As shown in Table 1, the unbalanced level in terms of handling regular and delayed flights has been reduced.

Cluster	Airport	Avg. flights handled by each airport within the cluster	Avg. delayed flights handled by each airport within the cluster
0	VRB,MLB	0.5	0
1	RSW,PGD,SRQ	5.33	0.33
2	ECP	2	0
3	GNV,DAB	1	0
4	EYW	0	0
5	PIE,TPA	13	0
6	MIA	40	7
7	TLH	2	0
8	PBI	15	3
9	JAX	13	0
10	MCO,SFB	22	3.5
11	FXE,FLL	14.5	2
12	VPS,PNS	4	1
Overall variance after reconfiguration		130	4.46
Overall variance before reconfiguration		192	5

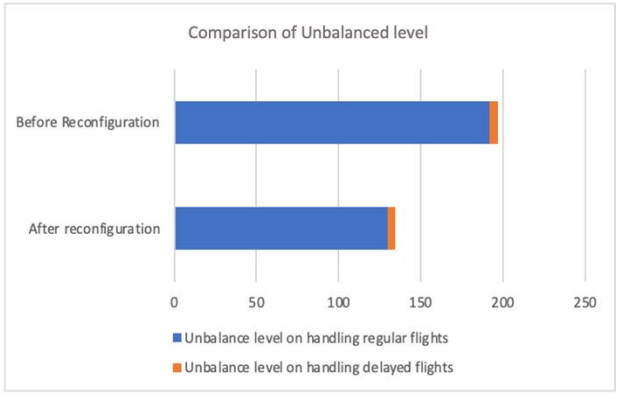
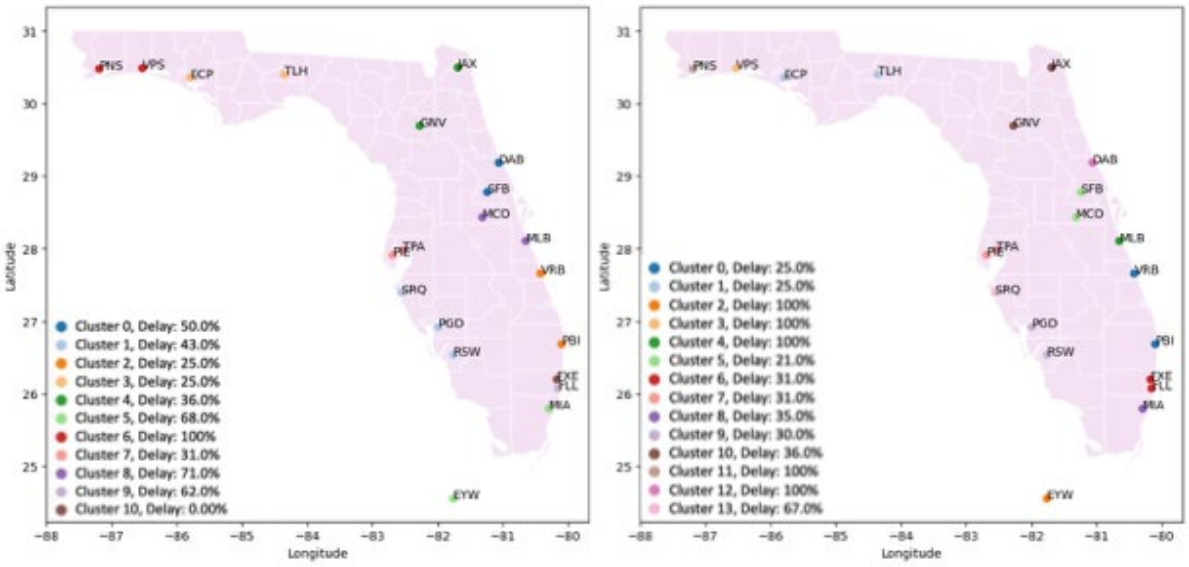


Table 1: Predicted workload distribution within each cluster (as in Figure 9b) and comparison of workload unbalance level for 7:00 AM to 9:00 AM Dec 24, 2015

As in Figures 10a and 10b, the airspace configuration algorithm combines PNS and VPS, RSW with PGD, FXE with FLL to balance the workload. In the meantime, MIA is isolated independently because there's no non-busy airports within reasonable range and without airspace overlap with busy airports. As seen in Table 2, the unbalanced level in terms of handling regular and delayed flights has been reduced. Compared with Figure 9, we have more clusters since the algorithm isolates some extremely busy airports and creates dedicated clusters for collaborating airports.



(a) (b)

Figure 10: Clustering and airspace reconfiguration using flight plan and delay prediction for 12:00 PM to 2:00 PM Dec 24, 2015. (a) Airspace allocation based on flight plans and delay prediction. (b) Fine-tuning of airspace configuration for better workload balancing

Cluster	Airport	Avg. flights handled by each airport within the cluster	Avg. delayed flights handled by each airport within the cluster
0	ECP	1	1
1	PBI,VRB	13.5	6.5
2	TLH,GNV	1	0.5
3	EYW	1	1
4	DAB	0	0
5	MLB	0	0
6	SFB,MCO	41	16
7	FLL,FXE	23	10
8	PIE,TPA	19	8
9	MIA	42	16
10	PGD,RSW	14	5
11	JAX	12	6
12	VPS,PNS	1.5	1.5
13	SRQ	2	2
Overall variance after reconfiguration		213	30
Overall variance before reconfiguration		494	77

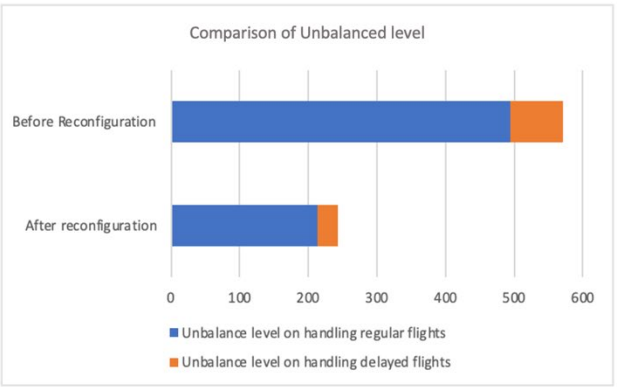
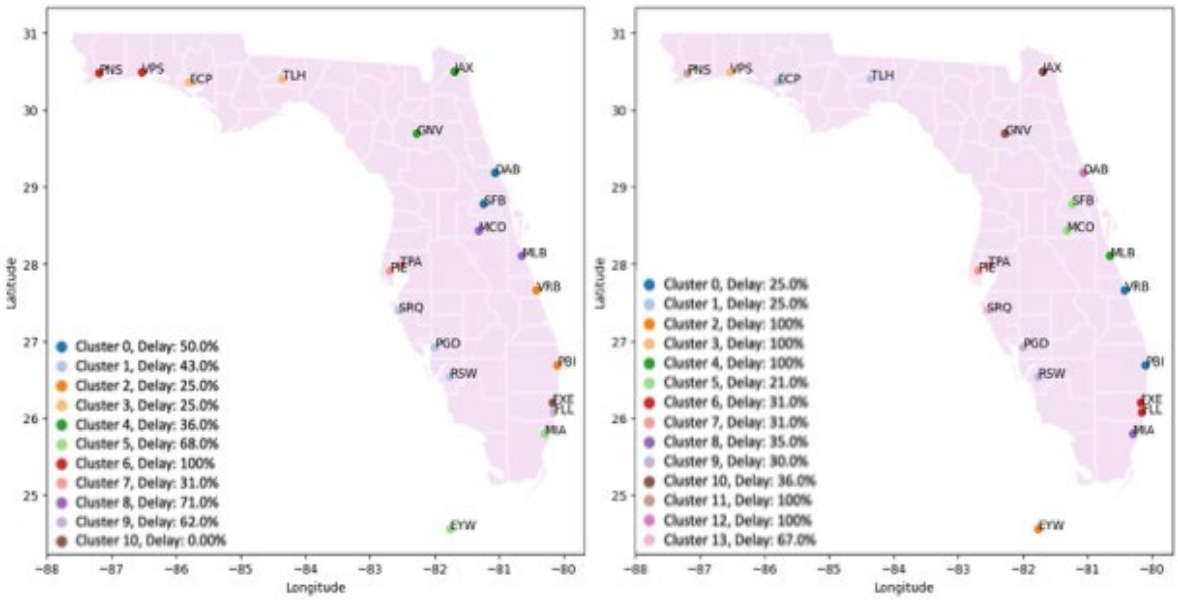


Table 2: Predicted workload distribution within each cluster (as in Figure 10b) and comparison of workload unbalance level for 12:00 PM to 2:00 PM Dec 24, 2015

As shown in Figure 11a and 11b, the airspace configuration algorithm combines ECP with TLH, MCO with SFB, RSW with PGD and FLL with FXE to balance the workload; As seen in Table 2, the unbalance level in terms of handling regular and delayed flights has been reduced. Compared with Figure 9 and Figure 10, even though the travel demand at this period is less than in the busy hours but the unbalanced level in terms of the average delayed flight handled by each airport is even more severe.



(a) (b)

Figure 11: Clustering and airspace reconfiguration using flight plan and delay prediction for 19:00 PM to 21:00 PM Dec 24, 2015. (a) Airspace allocation based on flight plan and delay prediction. (b) Fine-tuning of airspace configuration for better workload balancing

Cluster	Airport	Avg. flights handled by each airport within the cluster	Avg. delayed flights handled by each airport within the cluster
0	PBI,VRB	8	4
1	ECP,TLH	1	0.5
2	EYW	1	1
3	VPS	1	1
4	MLB	1	1
5	SFB,MCO	37.5	16
6	FLL,FXE	22.5	14
7	PIE,TPA	17	10.5
8	MIA	54	19
9	PGD,RSW	9	5.5
10	GNV,JAX	5.5	4
11	PNS	2	2
12	DAB	2	2
13	SRQ	3	2
Overall variance after reconfiguration		260	39
Overall variance before reconfiguration		455	99

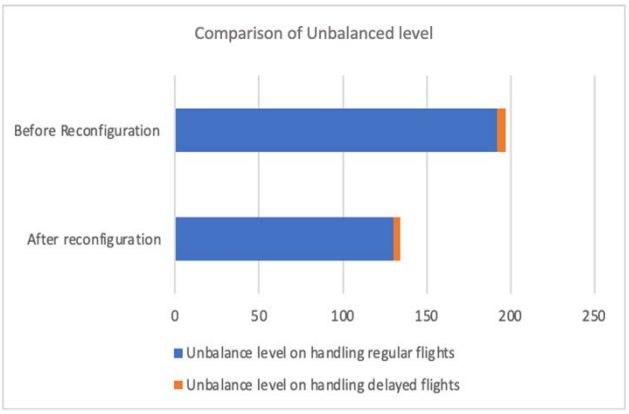


Table 3: Predicted workload distribution within each cluster (Figure 11b) and comparison of workload unbalance level for 19:00 PM to 21:00 PM Dec 24, 2015

Table 4 shows the improvement of ATC's workload unbalance level.

Time on Dec 24th, 2015	Traffic	Reduction on unbalanced level of handling regular flights	Reduction on unbalanced level of handling delayed flights
7:00 AM to 9:00 AM	Low	42.85 %	10.8%
12:00 PM to 2:00 PM	High	56.9%	61.04%
19:00 PM to 21:00 PM	Medium	42.86%	60.1%

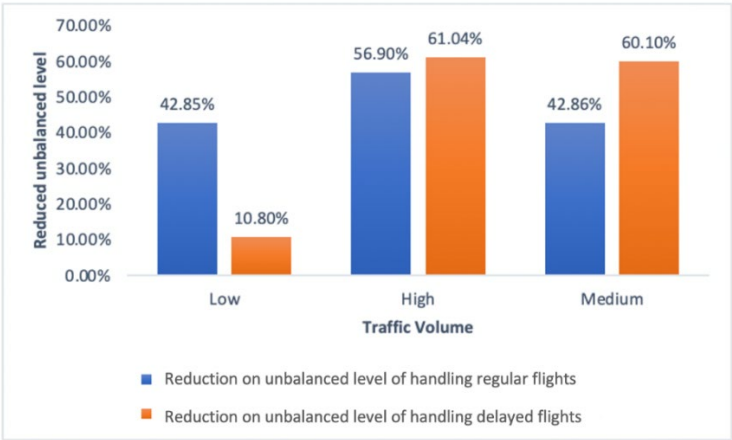
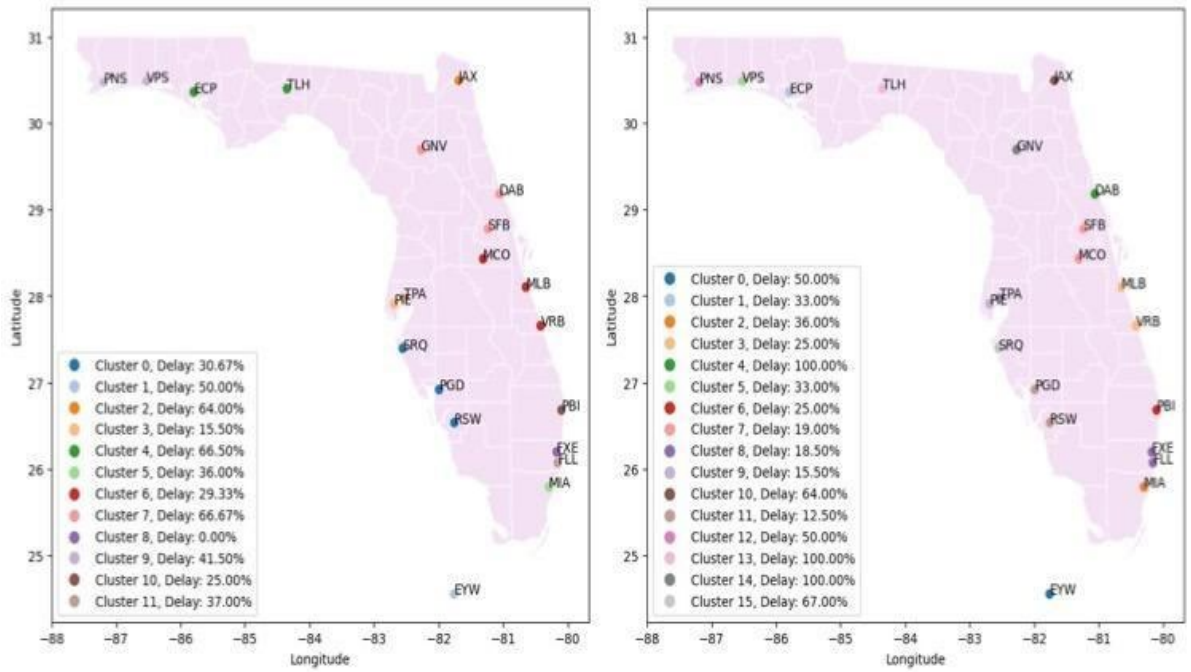


Table 4: Comparison of the DAC algorithm in handling regular and delayed flights

4.3.2 Same time for different high traffic days

To evaluate the performance of our airspace reconfiguration algorithm on different dates, three high traffic volume dates are chosen, which are 7/3, 11/25 (one day before Thanksgiving), 12/31 (one day before new year). As shown in table 2, our algorithm still outperformed static airspace configuration on all three days.



(a) (b)

Figure 12: Clustering and airspace reconfiguration using flight plan and delay prediction for 12:00 PM to 14:00 PM July 3, 2015. (a) Airspace allocation based on flight plan and delay prediction. (b) Fine-tuning of airspace configuration for better workload balancing

As shown in Figure 12, the airspace configuration algorithm combines MCO with SFB, FLL with FXE, TPA with PIE, and RSW with PGD to balance the workload. In the meantime, our algorithm also identifies that the following airports' airspace cannot be merged and should be considered independently: MIA, JAX, PNS, PBI, TLH, GNV, SRQ. As presented in Table 5, the unbalance level in terms of handling regular and delayed flights has been reduced.

Cluster	Airport	Avg. flights handled by each airport within the cluster	Avg. delayed flights handled by each airport within the cluster
0	EYW	2	1
1	ECP	3	1
2	MIA	45	16
3	VRB,MLB	1	0.5
4	DAB	1	1
5	VPS	3	1
6	PBI	12	3
7	SFB,MCO	38.5	14.5
8	FLL,FXE	20.5	7.5
9	PIE,TPA	21	6.5
10	JAX	14	9
11	PGD,RSW	8	2
12	PNS	6	3
13	TLH	2	2
14	GNV	2	2
15	SRQ	3	2
Overall variance after reconfiguration		185	55
Overall variance before reconfiguration		428	24
Improvement		56.77%	56.37%

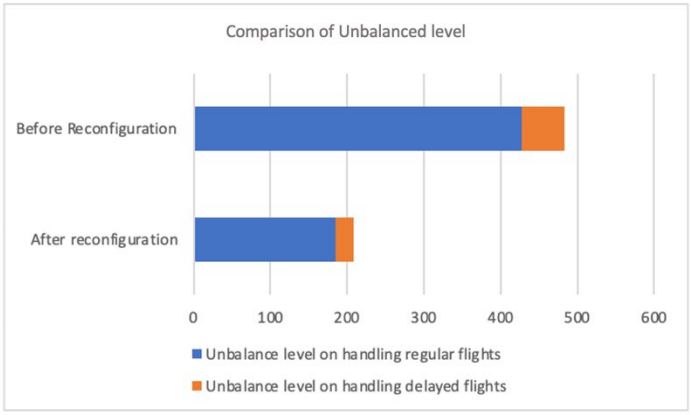
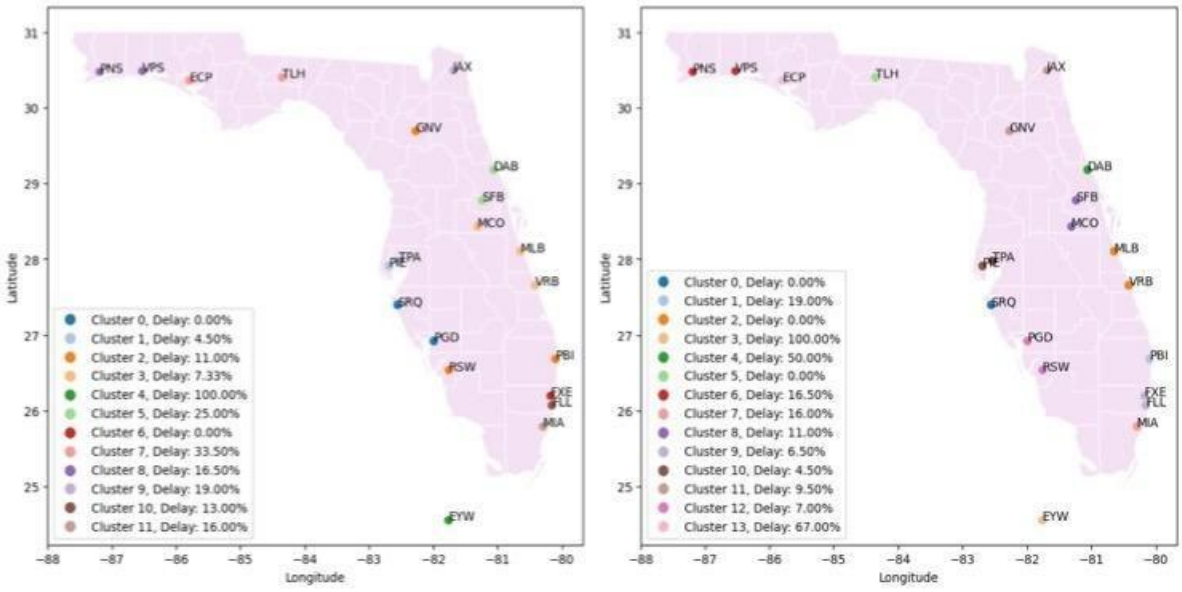


Table 5: Predicted workload distribution within each cluster (Figure 12b) and comparison of workload unbalanced level for 12:00 PM to 14:00 PM July 3, 2015



(a) (b)

Figure 13: Clustering and airspace reconfiguration using flight plan and delay prediction for 12:00 PM to 14:00 PM Nov 25, 2015. (a) Airspace allocation based on flight plan and delay prediction. (b) Fine-tuning of airspace configuration for better workload balancing

As in Figure 13, the airspace configuration algorithm combines MCO with SFB, FLL with FXE, JAX with GNV, TPA with PIE, and RSW with PGD. In the meantime, our algorithm also identifies that the following busy airports' airspace are not merged and should be considered independently: MIA, PBI, ECP. As in Table 6, the unbalance level in terms of handling regular and delayed flights has been reduced.

Cluster	Airport	Avg. flights handled by each airport within the cluster	Avg. delayed flights handled by each airport within the cluster
0	SRQ	3	0
1	PBI	21	4
2	VRB,MLB	1	0
3	EYW	1	1
4	DAB	2	1
5	TLH	1	0
6	PNS,VPS	2.5	0.5
7	MIA	38	6
8	MCO,SFB	41	9
9	FXE,FLL	22.5	3
10	TPA,PIE	23	2
11	GNV,JAX	8.5	1.5
12	PGD,RSW	11	1.5
13	ECP	3	2
Overall variance after reconfiguration		196	6
Overall variance before reconfiguration		479	16
Improvement		59.1%	62.5%

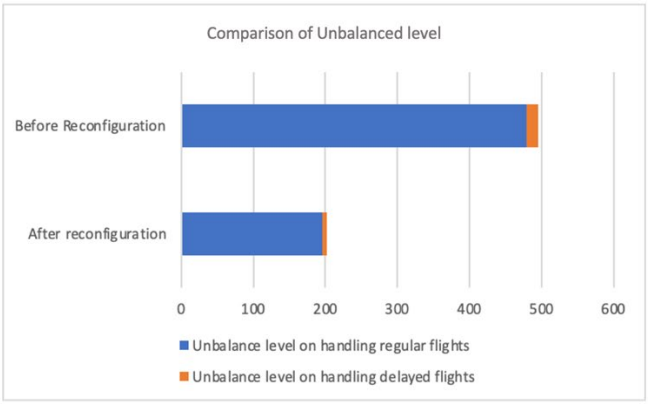
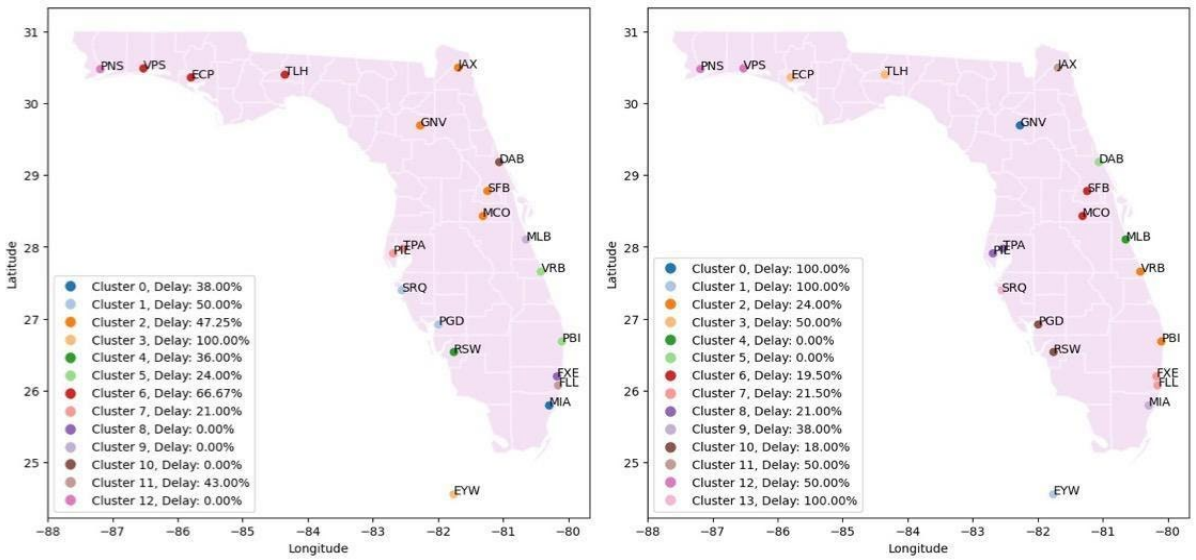


Table 6: Predicted workload distribution within each cluster (Figure 13b) and comparison of workload unbalance level for 12:00 PM to 14:00 PM Nov 25, 2015

As in Figure 14, the airspace configuration algorithm combines MCO with SFB, FLL with FXE, TPA with PIE, VPS with PNS, and RSW with PGD to balance the workload. In the meantime, our algorithm also identifies that the following busy airports' airspace are not merged and should be considered independently: MIA, JAX, SRQ. As in Table 7, the unbalance level in terms of handling regular and delayed flights has been reduced.



(a) (b)

Figure 14: Clustering and airspace reconfiguration using flight plan and delay prediction for 12:00 PM to 14:00 PM Dec 24, 2015. (a) Airspace allocation based on flight plan and delay prediction. (b) Fine-tuning of airspace configuration for better workload balancing

Cluster	Airport	Avg. flights handled by each airport within the cluster	Avg. delayed flights handled by each airport within the cluster
0	GNV	1	1
1	EYW	1	1
2	PBI,VRB	13.5	6.5
3	ECP,TLH	1	0.5
4	MLB	0	0
5	DAB	0	0
6	SFB,MCO	41	16
7	FLL,FXE	23	10
8	PIE,TPA	19	8
9	MIA	42	16
10	PGD,RSW	14	5
11	JAX	12	6
12	VPS,PNS	1.5	1.5
13	SRQ	2	2
Overall variance after reconfiguration		213	30.9
Overall variance before reconfiguration		494	77
Improvement		56.9%	59.9%

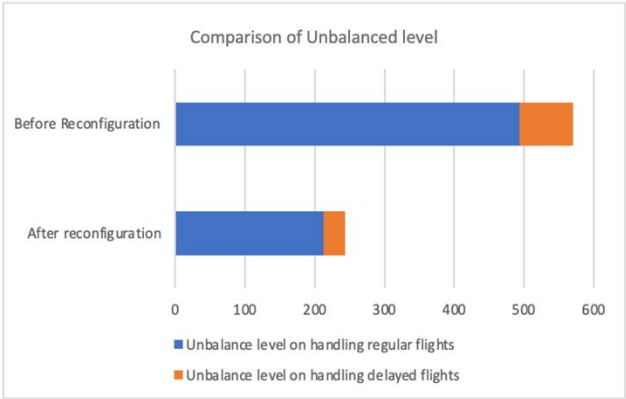
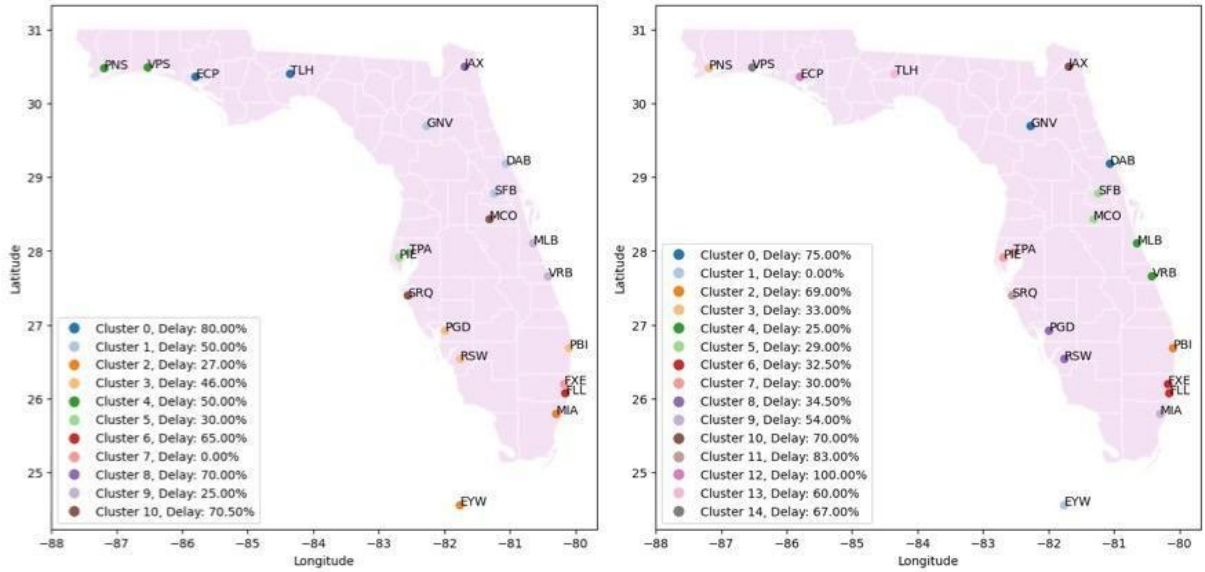


Table 7: Predicted workload distribution within each cluster (Figure 14b) and comparison of workload unbalanced level for 12:00 PM to 14:00 PM Dec 24, 2015

Comparing the results, we can find some common merging strategies are employed by our algorithm multiple times, such as MCO with SFB, FXE with FLL and etc. The airspace between MIA and SRQ may need to be further divided by introducing more ATC centers. In this experiments, our algorithm show its effectiveness by significantly decrease unbalance level the ATC's workload in busy scenarios.

4.3.3 Same time for different low traffic days

As a comparison to section 5.3.2, we also compare the airspace configuration results on the busy hours (12:00 PM to 14:00 PM) on these low-traffic days: Feb 17, June 9, and Sep 8, 2015.



(a) (b)

Figure 15: Clustering and airspace reconfiguration using flight plan and delay prediction for 12:00 PM to 14:00 PM Feb 17, 2015. (a) Airspace allocation based on flight plan and delay prediction. (b) Fine-tuning of airspace configuration for better workload balancing

As in Figure 15, the airspace configuration algorithm combines MCO with SFB, FLL with FXE, TPA with PIE, VPS with PNS, and RSW with PGD to balance the workload. In the meantime, our algorithm also identifies that the following busy airports' airspace are not merged and should be considered independently: MIA, JAX, SRQ, ECP, TLH, VPS. As in

Table 8, the unbalance level in terms of handling regular and delayed flights has been reduced. Even though the number of flights on this specific date is low, as indicated by the result of our clustering algorithm, the reduced traffic is due to the massive delay or other extreme situations.

Cluster	Airport	Avg. flights handled by each airport within the cluster	Avg. delayed flights handled by each airport within the cluster
0	DAB,GNV	1.5	1
1	EYW	1	0
2	PBI	16	11
3	PNS	3	1
4	MLB,VRB	1	0.5
5	SFB,MCO	39.5	23
6	FXE,FLL	23	15
7	TPA,PIE	25	15
8	PGD,RSW	14.5	10
9	MIA	37	20
10	JAX	10	7
11	SRQ	6	5
12	ECP	3	3
13	TLH	5	3
14	VPS	3	2
Overall variance after reconfiguration		169.5	56
Overall variance before reconfiguration		471	168
Improvement		64.1%	66.7%

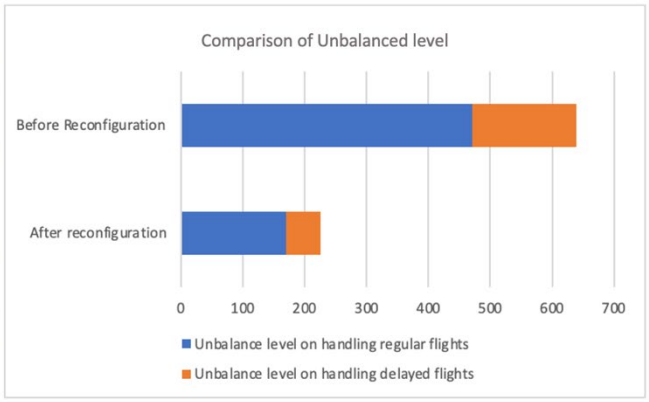
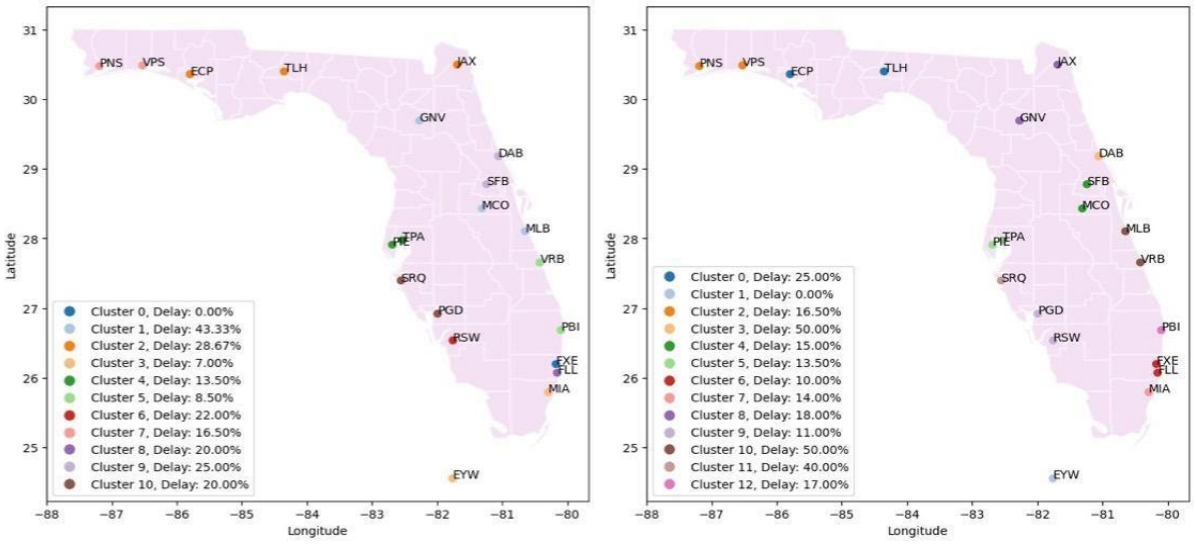


Table 8: Predicted workload distribution within each cluster (Figure 15b) and comparison of workload unbalanced level for 12:00 PM to 14:00 PM Feb 17, 2015

As in Figure 16, the airspace configuration algorithm combines MCO with SFB, FLL with FXE, TPA with PIE, JAX with GNV, MLB with GRB, and RSW with PGD to balance the workload. In the meantime, our algorithm also identifies that the following busy airports' airspace are not merged and should be considered independently: MIA, SRQ, PBI. As in Table 9, the unbalance level in terms of handling regular and delayed flights has been reduced. Compared with Table 8, the air traffic volume is normal because there's no sign of massive delay.



(a) (b)

Figure 16: Clustering and airspace reconfiguration using flight plan and delay prediction for 12:00 PM to 14:00 PM June 9, 2015. (a) Airspace allocation based on flight plan and delay prediction. (b) Fine-tuning of airspace configuration for better workload balancing

Cluster	Airport	Avg. flights handled by each airport within the cluster	Avg. delayed flights handled by each airport within the cluster
0	TLH,ECP	1	0.5
1	EYW	1	0
2	PNS,VPS	3.5	0.5
3	DAB	2	1
4	SFB,MCO	38	11.5
5	PIE,TPA	22.5	6
6	FLL,FXE	25	5
7	MIA	35	5
8	JAX,GNV	6	2
9	PGD,RSW	9	2
10	MLB,VRB	1	1
11	SRQ	5	2
12	PBI	12	2
Overall variance after reconfiguration		175	10.2
Overall variance before reconfiguration		439	31
Improvement		60%	67.1%

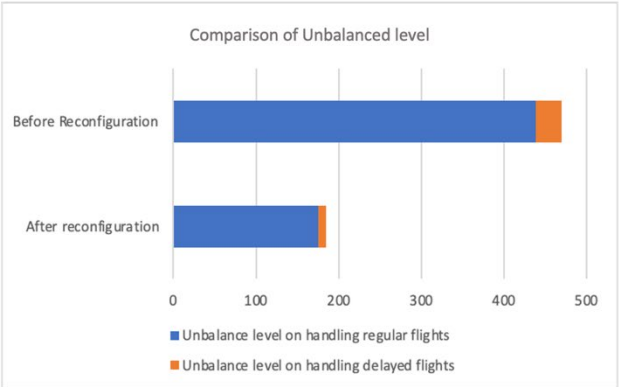
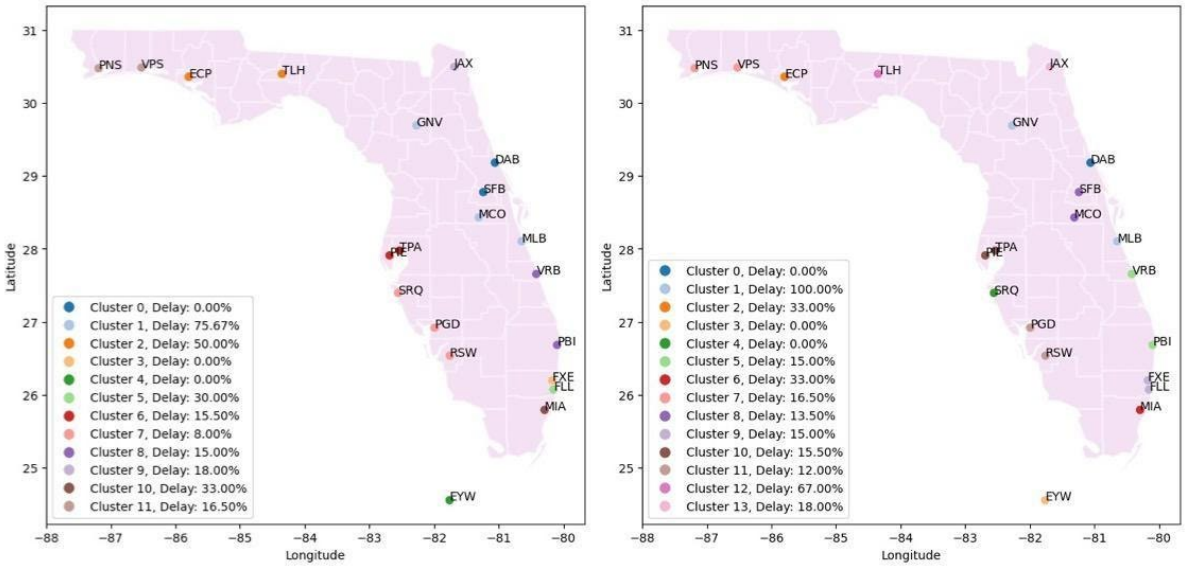


Table 9: Predicted workload distribution within each cluster (Figure 16b) and comparison of workload unbalanced level for 12:00 PM to 14:00 PM June 9, 2015



(a) (b)

Figure 17: Clustering and airspace reconfiguration using flight plan and delay prediction for 12:00 PM to 14:00 PM Sep 8, 2015. (a) Airspace allocation based on flight plan and delay prediction. (b) Fine-tuning of airspace configuration for better workload balancing

Cluster	Airport	Avg. flights handled by each airport within the cluster	Avg. delayed flights handled by each airport within the cluster
0	DAB	2	0
1	GNV,MLB	1	1
2	ECP	3	1
3	EYW	1	0
4	SRQ	3	0
5	PBI,VRB	5	1.5
6	MIA	42	14
7	PNS,VPS	3.5	0.5
8	MCO,SFB	35	9.5
9	FXE,FLL	23	7
10	TPA,PIE	18	5.5
11	PGD,RSW	8.5	2
12	TLH	3	2
13	JAX	11	2
Overall variance after reconfiguration		176	17.76
Overall variance before reconfiguration		379	32
Improvement		53%	44.5%

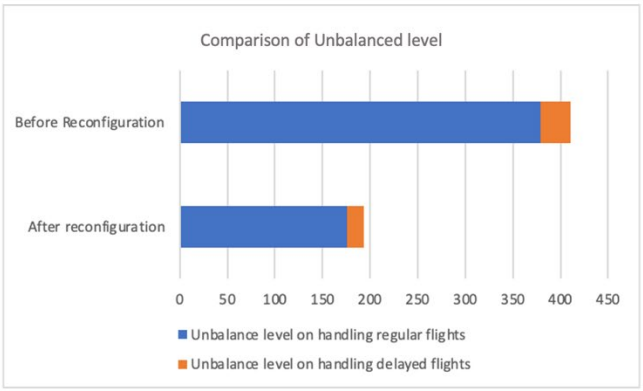
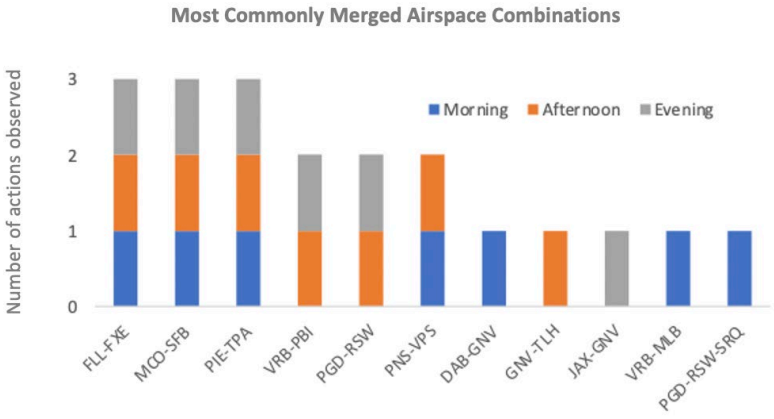


Table 10: Predicted workload distribution within each cluster (Figure 15b) and comparison of workload unbalanced level for 12:00 PM to 14:00 PM June 9, 2015

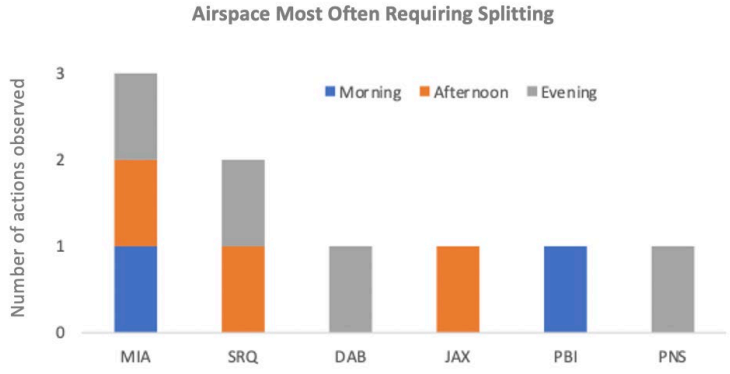
As in Figure 17, the airspace configuration algorithm combines MCO with SFB, FLL with FXE, TPA with PIE, JAX with GNV, MLB with GRB, and RSW with PGD to balance the workload. In the meantime, our algorithm also identifies that the following busy airports' airspace are not merged and should be considered independently: MIA, TLH, JAX. As in Table 10, since this is not a busy day, our algorithm's efficiency in rebalancing the workload of ATC seems some decrease, but its performance is still promising.

4.4 Result discussion

4.4.1 Analysis of airspace merging within the same day



(a)

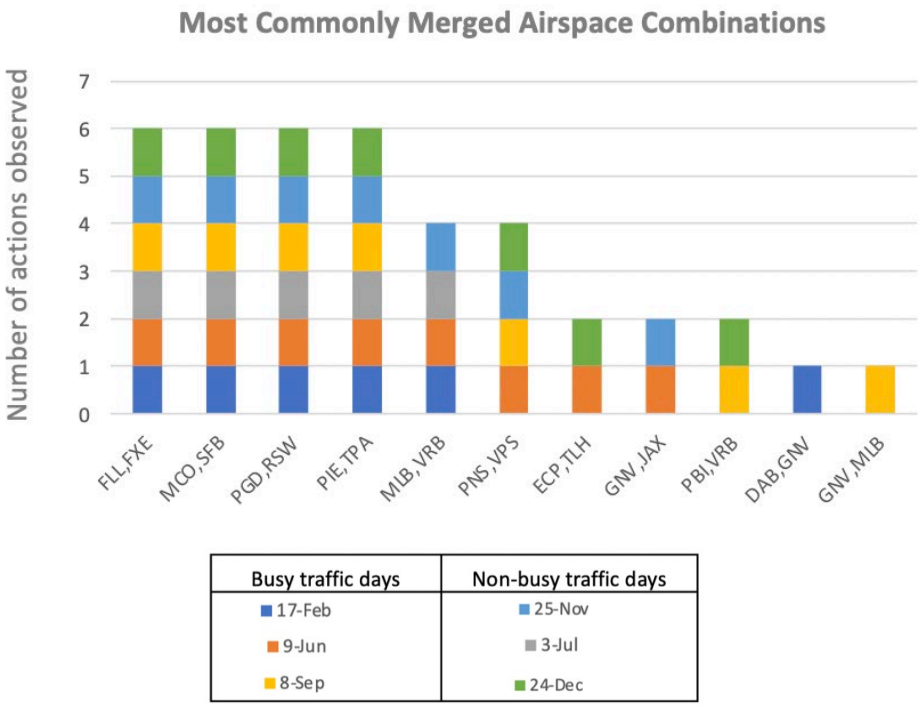


(b)

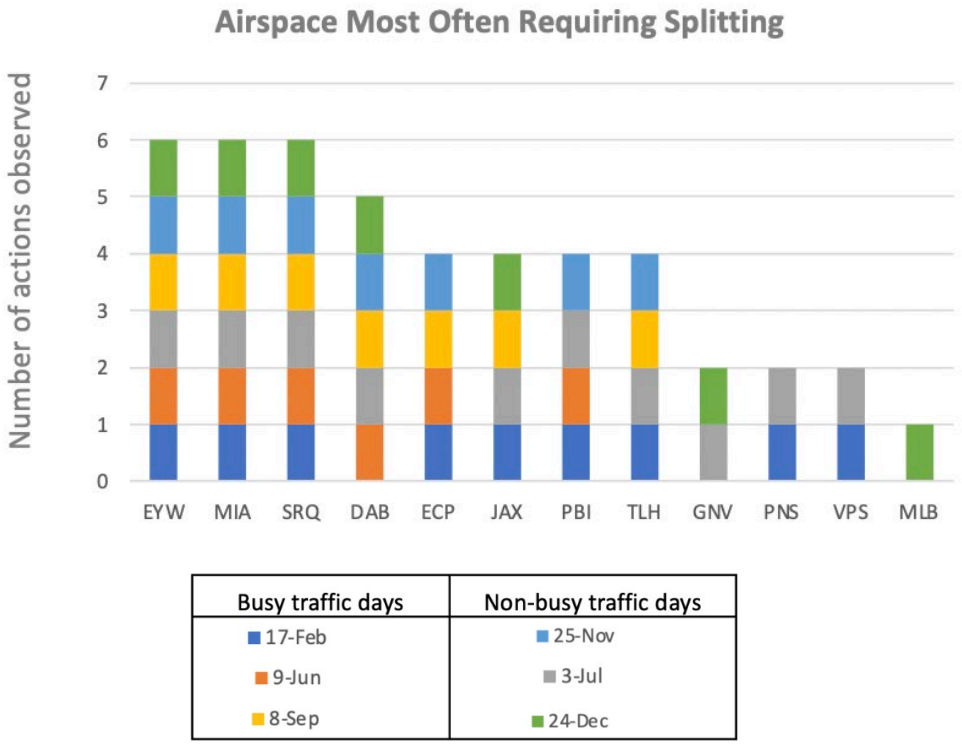
Figure 18: Summary of airspace reconfiguration actions within Dec 24, 2015 (a) Merged airspace. (b) Airspaces that need further separation.

We summarize the airspace merging actions taken by the algorithm as in Figure 11. The three most successful merges are: FLL-FXE, MCO-SFB, and TPA-PIE, these airports that are selected to merge into collaborative pairs because: a) They are closely located within the communication range of ADS-B transponders and b) executive airports are not usually as busy as large international airports. This frequent merging also indicates that the three airports, FLL, TPA, MCO, are busy and need assistance all day round. Comparably, PBI and PNS are busy in the from afternoon to the evening, RSW requires merging in the morning and afternoon. Interestingly, GNV is dynamically assigned to collaborate with three different airports (DAB, TLH, JAX) in three different time periods. Figure b shows that MIA and its nearby airports are extremely busy all day round and makes it impossible to find collaborative airport that is less busy than it. Comparably, SRQ's ATC team needs assistance in the afternoon and evening. Additionally, DAB and PNS needs additional assistance in the evening while JAX and PBI needs assistance in morning and afternoon, respectively.

4.4.2 Analysis of airspace merging in one typical busy hour of different days



(a)



(b)

Figure 19: Summary of airspace reconfiguration actions observed on 12:00 PM to 14:00 PM on different days of 2015: (a) airspaces that are merged. (b) airspaces that need further separation

The actions taken most frequently are the merging of FLL-FXE, MCO-SFB, PGD-RSW, and PIE-TPA, the reason for these mergings is that the international airports are always busy all years round and need assistances other regional or executive airports. GNV is surrounded by medium and busy airports, therefore, our algorithm dynamically assigned GNV to collaborate with airports for different scenarios. We also realized that EYW, MIA, SRQ are always assigned to not merge with other airports because MIA and SRQ locates in busy airspace of southern Florida, and they cannot find a less busy airport to collaborate with them. EYW is located too far away from other existing commercial airports, making it difficult to merge its airspace with other airports.

5 FINDINGS, CONCLUSIONS, RECOMMENDATIONS

In emergency situations, the timely and efficient evacuation of people is of utmost importance, and aviation transportation plays a vital role in achieving this goal. However, assuring the safety of airplanes without creating huge and unbalanced load to the national airspace system during emergency evacuations can be challenging, particularly with the current airspace configuration method which aims mainly to satisfy the demands of normal operations. In this context, automated and dynamic airspace configuration (DAC) presents a promising and innovative. DAC has the ability to increase air traffic throughput while still balancing the workload of air traffic controllers, in this way, we ensure that the ATC operators in emergent airspaces will not have to face the surging workloads independently. Specifically, we integrate deep neural networks and unsupervised machine learning algorithms to develop a practical and effective DAC framework that can dynamically select collaborative airports to assist busy airports that is under emergent situations. Our experiments show that the workload imbalance degree of handling regular and delayed flights can by reduced by up to 50%.

In the first phase, we reviewed the application of AI in ground transportation emergencies and emergency management response operations that could be used in DAC. From literature review, it is evident that further research still needs to be carried out especially when considering AI applications. After the surveys, we considered combining the two works mentioned in section 3.3 together by utilizing both temporal and spatial information.

In the second phase, we proposed a Spatial-Temporal graph neural network model. By constructing a graph of the air transportation network, we want to let the DAC framework adjust the airspace based on the predicted situations in the future. The neural network learns the spatial information about the air traffic propagation from other airports. Also, the historical data of each airport was also included in the input, in order to let the model learn the temporal information about the past traffic condition. Experiments were conducted to evaluate the model's performance. The task is to the use the air traffic data of past 9 hours to predict if there will be a delay after 2 hours for a total of 285 airports in the U.S. The results show that the model has a classification accuracy score of 57% and recall score of 61%. The performance is

not as good as expected, some further research is undergoing to improve the model performance: first, finding a larger dataset which contains much more delay flight information; second, include weather information to assist model prediction as the majority of the delay flights are affected by this factor; finally, improve the construction method of the adjacent matrix by combining the distance and flight number for the weights between the nodes.

In the third phase, we proposed a novel approach that utilizes the future prediction of flight delays to reconfigure airspace to balance the workload of air traffic controllers. The assumption here is that airports needing assistance will face a lot of predictable delays in the near future. Our approach was based on a spectral clustering algorithm that constructs a graph representation of the airspace using geographic information. The spectral clustering algorithm was applied to predict the new configuration for the next time window based on the current workload. To promote efficient traffic flow, our algorithm first clustering airspace into groups that prioritize heavy workload modules as centers surrounded by modules with spare resources. Then reallocating ATC resources within each cluster to balance the workload among modules for new configuration of each time window. Thus, maximizing air mobility and reducing congestion during emergency evacuations. The results of our experiment outperform static airspace configuration. To evaluate the model ability of generation new configuration that can maintain balanced workload among modules, it is tested under three set of settings: different time windows on the same day, same time window for different high traffic days and same time window for different low traffic days. The result shows that, it reduces the unbalance level and number of delayed flights over 55% under high traffic volume.

Our simulations on real data also revealed that the air space in Miami and Sarasota need to be further sliced to increase its capability, since our DAC algorithm has a low success rate in finding collaborative airports that are less busy within 60 miles. Other than that, the airspace in Key West airport is located too far away from other airports making it almost impossible to collaborate with other busy airports to share emergent workloads.

There are several limitations for our study at the current stage: (a) Since we were not able to fetch detailed information of ATC controllers in many airports, it makes it impossible to create a precise model to quantify the capability of each tower and balance their workloads. (b) We leveraged the delay ratio as a metric to roughly estimate the busy level of airports, better metrics can be used to make the model more practical. (c) The validity of data from public sources is quite limited, as airports' operational data are seldom archived and shared properly, which makes the simulation less realistic and applicable.

There is scope for future work to improve this approach in several ways. Firstly, sector workload evaluation can be enhanced by exploring more effective features beyond the number of delayed flights and total flights used in this study. Secondly, the performance of flight delay prediction can be improved by addressing the limitations of training the current Spatial-temporal Graph Neural Network, such as the highly unbalanced dataset and the lack of weather data, which is one of the most important causes of flight delay. Additionally, the effectiveness of the contrastive loss could be improved by training with larger batch sizes, though hardware limitations currently prevent this. Lastly, the linear operations currently used in the spectral clustering algorithm to reduce the graph dimension may not be able to capture complex air traffic patterns at scale. Therefore, it would be worth exploring nonlinear methods such as autoencoders to reduce the graph dimension.

6 REFERENCES

Agarap, A. F. (2018). Deep learning using rectified linear units (relu). *arXiv Preprint arXiv:1803.08375.*

Assi, K., Rahman, S. M., Mansoor, U., & Ratrout, N. (2020). Predicting crash injury severity with machine learning algorithm synergized with clustering technique: A promising protocol. *International Journal of Environmental Research and Public Health, 17*(15), 5497.

Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A. R., . . . Zou, B. (2010). Total delay impact study: A comprehensive assessment of the costs and impacts of flight delay in the United States.

Bibi, R., Saeed, Y., Zeb, A., Ghazal, T. M., Rahman, T., Said, R. A., . . . Khan, M. A. (2021). Edge AI-based automated detection and classification of road anomalies in VANET using deep learning. *Computational Intelligence and Neuroscience, 2021*, 1-16. doi:10.1155/2021/6262194.

Bishop, C. M. (1994). Neural networks and their applications. *Review of Scientific Instruments, 65*(6), 1803-1832.

Boeing (2022). Statistical summary of commercial jet aircraft accidents. worldwide operations (online). *Statistical Summary of Commercial Jet Aircraft Accidents. Worldwide Operations (Online).*

Bustos, C., Rhoads, D., Solé-Ribalta, A., Masip, D., Arenas, A., Lapedriza, A., & Borge-Holthoefer, J. (2021). Explainable, automated urban interventions to improve pedestrian and vehicle safety. *Transportation Research Part C: Emerging Technologies, 125*, 103018.

Caliendo, C., Guida, M., & Parisi, A. (2007). A crash-prediction model for multilane roads. *Accident Analysis & Prevention, 39*(4), 657-670. doi: 10.1016/j.aap.2006.10.012.

Chaitanya, K., Erdil, E., Karani, N., & Konukoglu, E. (2020). Contrastive learning of global and local features for medical image segmentation with limited annotations. *Advances in Neural Information Processing Systems, 33*, 12546-12558.

Chang, L. (2005). Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network. *Safety Science, 43*(8), 541-557.

Chauhan, V. K., Dahiya, K., & Sharma, A. (2019). Problem formulations and solvers in linear SVM: A review. *Artificial Intelligence Review, 52*(2), 803-855.

Choi, E., Bahadori, M. T., Sun, J., Kulas, J., Schuetz, A., & Stewart, W. (2016). Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. *Advances in Neural Information Processing Systems, 29.*

Department of Transportation (2010). *FAA advisory circular - airport emergency plan*

Department of Transportation (2020). *FAA order JO 1900.47F - air traffic control operational contingency plans.*

Department of Transportation (2021). *FAA order JO 7210.3CC - facility operation and administration.*

Department of Transportation Bureau of Transportation Statistics (2017). Flight delays and cancellations. Retrieved from https://www.kaggle.com/datasets/usdot/flight-delays.

Eyjafjallajokull's global fallout (2010). 69.

Federal Aviation Administration (2020). United States department of transportation. trajectory based operations (TBO). Retrieved from https://www.faa.gov/air_traffic/technology/tbo/.

Federal Emergency Management Agency (2020). United States department of homeland security. hazard information sheets. Retrieved from https://www.ready.gov/sites/default/files/2021-01/ready_full-suite_hazard-info-sheets.pdf.

Federal Emergency Management Agency & United States Department of Homeland Security (2022). Emergency operations center how-to quick reference guide.

Gallego, C. E. V., Comendador, V. F. G., Carmona, M. A. A., Valdés, R. M. A., Nieto, F. J. S., & Martínez, M. G. (2019). A machine learning approach to air traffic interdependency modelling and its application to trajectory prediction. *Transportation Research Part C: Emerging Technologies, 107*, 356-386.

Gao, Y., Liu, W., & Lombardi, F. (2020). Design and implementation of an approximate softmax layer for deep neural networks. Paper presented at the *2020 IEEE International Symposium on Circuits and Systems (ISCAS),* 1-5.

Gianazza, D. (2010). Forecasting workload and airspace configuration with neural networks and tree search methods. *Artificial Intelligence, 174*(7), 530-549. doi: 10.1016/j.artint.2010.03.001.

Gianazza, D., & Guittet, K. (2006). *Selection and evaluation of air traffic complexity metrics* IEEE. doi:10.1109/dasc.2006.313710.

Gu, X., Li, T., Wang, Y., Zhang, L., Wang, Y., & Yao, J. (2018). Traffic fatalities prediction using support vector machine with hybrid particle swarm optimization. *Journal of Algorithms & Computational Technology, 12*(1), 20-29.

Guégain, E., & Quinton, C. (2022). The ICO tool suite: Optimizing highly configurable systems.

Hadsell, R., Chopra, S., & Lecun, Y. (2006). *Dimensionality reduction by learning an invariant mapping* IEEE. doi:10.1109/cvpr.2006.100.

Han, P., Wang, W., Shi, Q., & Yue, J. (2021). A combined online-learning model with K-means clustering and GRU neural networks for trajectory prediction. *Ad Hoc Networks, 117*, 102476.

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their Applications, 13*(4), 18-28.

Houston, N., & Hamilton, B. A. (2006). Best practices in emergency transportation operations preparedness and response: Results of the FHWA workshop series. *Best Practices in Emergency Transportation Operations Preparedness and Response: Results of the FHWA Workshop Series, Annotated.*

ICAO (2001). *Annex 11to the convention on international civil aviation,* International Civil Aviation Organization.

ICAO (2004). *Annex 14 to the convention on international civil aviation - volume I aerodrome design and operations* International Civil Aviation Organization.

ICAO (2013). *Annex 14 to the convention on international civil aviation - volume II heliports*. 999 University Street, Montréal, Quebec, Canada H3C 5H7: International Civil Aviation Organization.

ICAO (2016). *Annex 19 to the convention on international civil aviation*. 999 Robert-Bourassa Boulevard, Montréal, Quebec, Canada H3C 5H7: International Civil Aviation Organization.

International Civil Aviation Organization (2019). Emergency response planning. Retrieved from https://www.icao.int/sustainability/ERP/Pages/default.aspx.

Jiang, Y., Niu, S., Zhang, K., Chen, B., Xu, C., Liu, D., & Song, H. (2022). Spatial-temporal graph data mining for IoT-enabled air mobility prediction. *IEEE Internet of Things Journal, 9*(12), 9232-9240. doi:10.1109/JIOT.2021.3090265.

Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., . . . Krishnan, D. (2020). Supervised contrastive learning. *Advances in Neural Information Processing Systems, 33*, 18661-18673.

Klein, A., Rodgers, M. D., & Kaing, H. (2008). Dynamic FPAs: A new method for dynamic airspace configuration. Paper presented at the 1-11. doi:10.1109/ICNSURV.2008.4559176.

Kopardekar, P., Bilimor, K., & Sridhar, B. (2007). Initial concepts for dynamic airspace configuration. *AIAA Paper,* doi: https://doi.org/10.2514/6.2007-7763.

Kuipers, S., Verolme, E., & Muller, E. (2020). Lessons from the MH17 transboundary disaster investigation. *Journal of Contingencies and Crisis Management, 28*(2), 131-140. doi:10.1111/1468-5973.12288.

Kumar, S. G., Corrado, S. J., Puranik, T. G., & Mavris, D. N. (2021). Classification and analysis of go-arounds in commercial aviation using ADS-B data. *Aerospace, 8*(10), 291.

Lee, P., Mercer, J., Gore, B., Smith, N., Lee, K., & Hoffman, R. (2008). Examining airspace structural components and configuration practices for dynamic airspace configuration. doi:10.2514/6.2008-7228.

Liu, H., & Shetty, R. R. (2021). Analytical models for traffic congestion and accident analysis.

Llugsi, R., El Yacoubi, S., Fontaine, A., & Lupera, P. (2021). Comparison between adam, AdaMax and adam W optimizers to implement a weather forecast based on neural networks for the Andean city of Quito. Paper presented at the *2021 IEEE Fifth Ecuador Technical Chapters Meeting (ETCM),* 1-6.

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society, 18*(6), 275-285.

National Research Council (1991). *A safer future: Reducing the impacts of natural disasters* National Academies Press.

Neath, A. A., & Cavanaugh, J. E. (2012). The bayesian information criterion: Background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics, 4*(2), 199-203.

Niu, Z., Zhong, G., & Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing, 452*, 48-62.

Occupational Safety and Health Administration (2018). Emergency preparedness and response: Getting started. evacuation & shelter-in-place. *United States Department of Labor, 9.*

Pappalardo, G., Cafiso, S., Di Graziano, A., & Severino, A. (2021). Decision tree method to analyze the performance of lane support systems. *Sustainability, 13*(2), 846.

Royal Aeronautical Society (2006). *Aviation in support of humanitarian and disaster relief operations*. London, U.K: Royal Aeronautical Society.

Seet, C. (2022, Sep 26). Hurricane Ian approaches the United States: Will it impact air travel? Retrieved from https://simpleflying.com/hurricane-ian-air-travel-impact/.

Sergeeva, M., Delahaye, D., Mancel, C., & Vidosavljevic, A. (2017). Dynamic airspace configuration by genetic algorithm. *Journal of Traffic and Transactions Engineering（English Edition）, 4*(3), 300-314. doi: 10.1016/j.jtte.2017.05.002.

Shaluf, I. M., Ahmadun, F., & Mat Said, A. (2003). A review of disaster and crisis. *Disaster Prevention and Management: An International Journal, 12*(1), 24-32.

Shanmugam, B., Raheem, M., & Batcha, N. (2021). Road accident and emergency management: A data analytics approach.*8*, 122-128.

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena, 404*, 132306.

Silva, P. B., Andrade, M., & Ferreira, S. (2020). Machine learning applied to road safety modeling: A systematic literature review. *Journal of Traffic and Transportation Engineering (English Edition), 7*(6), 775-790.

Sun Choi, Young Jin Kim, Briceno, & Mavris. (Sep 2016). Prediction of weather-induced airline delays based on machine learning algorithms. Paper presented at the 1-6. doi:10.1109/DASC.2016.7777956 Retrieved from https://ieeexplore.ieee.org/document/7777956.

Sune, E. R., Donati, J., & Youssef, N. (2021). U.S. carries out drone strike in Kabul as Afghanistan evacuation flights end: WSJ. *Barron's (Online).*

Taleb, A., Kirchler, M., Monti, R., & Lippert, C. (2022). Contig: Self-supervised multimodal contrastive learning for medical imaging with genetics. Paper presented at the *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 20908-20921.

von Luxburg. (2007). A tutorial on spectral clustering. *Statistics and Computing, 17*(4), 395-416. doi:10.1007/s11222-007-9033-z.

Wang, F., Bi, J., Xie, D., & Zhao, X. (2022). Flight delay forecasting and analysis of direct and indirect factors. *IET Intelligent Transport Systems, 16*(7), 890-907.

Wanyan, T., Honarvar, H., Jaladanki, S. K., Zang, C., Naik, N., Somani, S., . . . Glicksberg, B. S. (2021). Contrastive learning improves critical event prediction in COVID-19 patients. *Patterns (New York, N.Y.), 2*(12), 100389. doi:10.1016/j.patter.2021.100389.

Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and Computing, 4*, 65-85.

World Health Organization, (. (2022). Road traffic injuries. Retrieved from https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

Wu, F., Souza, A., Zhang, T., Fifty, C., Yu, T., & Weinberger, K. (2019). Simplifying graph convolutional networks. Paper presented at the *International Conference on Machine Learning,* 6861-6871.

You, Y., Chen, T., Sui, Y., Chen, T., Wang, Z., & Shen, Y. (2020). Graph contrastive learning with augmentations. *Advances in Neural Information Processing Systems, 33*, 5812-5823.

Zaremba, W., Sutskever, I., & Vinyals, O. (2014). Recurrent neural network regularization. *arXiv Preprint arXiv:1409.2329.*

Zelinski, S., & Chok, F. L. (2011). Comparing methods for dynamic airspace configuration. Paper presented at the 3A1-13; 1. doi:10.1109/DASC.2011.6096043.

Zhang, Z., & Sabuncu, M. (2018). Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in Neural Information Processing Systems, 31.*

7 APPENDIX

Publications, presentations, posters resulting from this project:

Ke Feng, Dahai Liu, Yongxin Liu, Hong Liu, Houbing Song. GraphDAC: A Graph-Analytic Approach to Dynamic Airspace Configuration. **Accepted for Publication** by IEEE IRI'23 on June 13, 2023.