Transportation Research Part C 125 (2021) 103049

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Prediction of runway configurations and airport acceptance rates for multi-airport system using gridded weather forecast

Yuan Wang, Yu Zhang *

Department of Civil and Environmental Engineering, University of South Florida, 4202 E. Fowler Ave. ENB118, 33620, USA

ARTICLE INFO

Keywords:
Airport capacity
Air traffic flow management
Weather forecast
Deep-learning method
Multi-layer Convolutional Neural Networks

ABSTRACT

Accurate prediction of real-time airport capacity, a.k.a. airport acceptance rates (AARs), is key to enabling efficient air traffic flow management. AARs are dependent on selected runway configurations and both are affected by weather conditions. Although there have been studies tackling on the prediction of AARs or runway configurations or both, the prediction accuracy is relatively low and only single airport is considered. This study presents a data-driven deep-learning framework for predicting both runway configurations and AARs to support efficient air traffic management for complex multi-airport systems. The two major contributions from this work are 1) the proposed model uses assembled gridded weather forecast for the terminal airspace instead of an isolated station-based terminal weather forecast, and 2) the model captures the operational interdependency aspects inherent in the parameter learning process so that proposed modeling framework can predict both runway configuration and AARs simultaneously with higher accuracy. The proposed method is demonstrated with a numerical experiment taking three major airports in New York Metroplex as the case study. The prediction accuracy of the proposed method is compared with methods in current literature and the analysis results show that the proposed method outperforms all existing methods.

1. Introduction

The number of air passengers has been steadily increasing in past decades, with the average annual rate reaching 4.4% in the past five years (2014–2018) in the U.S. However, on average, only 77.95% of airline flights have been on-time in those five years. According to the Bureau of Transportation Statistics in 2018, 19.95% of flights encountered delays and 2.2% were canceled. Of the delayed flights, 55% were caused by inclement weather conditions and 36% by heavy traffic volume (Bureau of Transportation Statistics (BTS), 2018). Runways are bottlenecks of National Airspace System (NAS), and the situation becomes worse when adverse weather affects airfield capacity. To alleviate congestion, either the physical airfield capacity needs to be enhanced or air traffic management (ATM) initiatives are needed to improve the utilization of existing capacity. Increasing airport capacity through constructing new runways, taxiways, or terminal buildings requires massive capital investment and rigorous feasibility studies, which can take up to or more than 10 years and sometimes can be strongly opposed by local communities due to environmental concerns. Comparably, ATM initiatives to improve operations could be a timely and effective solution. The U.S. Federal Aviation Administration (FAA) and the National Aeronautics and Space Administration (NASA) have been working on airfield-related ATM initiatives such as an integrated arrival, departure, and surface tool as a traffic management system for complex terminal environments. These initiatives are expected to

* Corresponding author.
E-mail address: yuzhang@usf.edu (Y. Zhang).

https://doi.org/10.1016/j.trc.2021.103049
Received 13 April 2020; Received in revised form 31 December 2020; Accepted 15 February 2021
Available online 24 February 2021

improve the utilization of the airfield capacity. Thus, airfield capacity, especially real-time airfield capacity prediction, is key to these initiatives. As elaborated in a review of the literature, extensive efforts have been made for developing methods to predict real-time airfield capacities. The performance of prediction methods, however, still need improvement. Also, simultaneous prediction of real-time airfield capacity of airports in a multi-airport system, where operational interference exist due to the sharing of airspace, has not been well studied yet.

Airport capacity is usually defined as maximum sustainable throughput, i.e., the number of aircraft operations an airport can accommodate under continuous demand for a certain time period. Depending on the purpose, methods for defining airport capacity can be quite different. For example, for planning purposes, the airport capacity tool runwaySimulator sponsored by FAA and developed by MITRE, is widely used. For evaluating large capital expenditures, users may choose simulation software and perform more sophisticated what-if analysis to determine airport capacity by considering different runway layouts and operations, traffic demand, and also arrival/departure balance. Airport capacity obtained from these approaches is called theoretical airport capacity. For air traffic management initiatives to improve the use of airfield capacity, however, real-time suitable runway configuration and airfield capacity are needed, which are dynamically predicted given forecasted weather conditions, air traffic control staffing, and other operational environment features. Among these factors, forecasted weather is a major influential factor, and the combined prediction of runway configuration and airfield capacity is a challenging problem.

To complicate things even more, in areas where multiple airports are present, operational interdependency among airports makes accurate prediction much more complex and mathematically difficult to formulate. These areas, called metroplexes (with multi-airport systems), are composed of two or more major airports that usually serve the air traffic of a metropolitan area and have coordinated operations. For other non-metropolitan areas, they may also have multi-airport systems where some airports act as reliever airports in case of congestion or reduced capacity at other airport(s). Airports in a multi-airport system need to have synchronized operations to effectively use the limited terminal airspace, especially when capacity is constrained by inclement weather conditions or other factors. These airports usually share only one Terminal Radar Approach Control (TRACON) facility; for example, three major airports in New York are in TRACON N90. Nowadays, air traffic control personnel at each airport select a runway configuration, whereas air traffic managers/planners in the TRACON determine airport acceptance rates (AARs) for these airports for the next one or several hours based on various sources of weather forecast and airport conditions (e.g., runway closure). Accurate prediction of AARs is key to enabling efficient air traffic flow management. Underestimating AARs could cause unnecessary ground delays and underuse airport/airspace resources; overestimating AARs could increase airborne delays where aircraft must be held or even diverted until receiving confirmation of entering terminal airspace for safe landing.

Runway configuration is the combination of active runways in a time period that an airport uses for serving arrivals and departures. The sequence of configurations selected by controllers greatly influences an airport's capacity. Thus, it is important to predict both runway configuration and AAR.

In multi-airport systems, the task of selecting runway configurations and setting appropriate AARs is challenging because both heavily depend on weather forecasts in a large-scale terminal area. Currently, air traffic control personnel lack sufficient tools to assist them in translating weather forecast information from multiple sources into real-time runway configuration and AARs. In addition, under current operational protocol, runway configuration and AARs are determined by the coordination of different entities, which requires intensive verbal communications. Thus, a decision-making tool that can predict runway configuration and AARs in real-time for a metroplex is an urgent need.

This study presents a data-driven deep-learning framework for predicting both runway configurations and AARs to support efficient air traffic management for complex multi-airport systems. The two major contributions from this work are 1) the proposed model uses assembled gridded weather forecast for the terminal airspace instead of an isolated station-based terminal weather forecast, and 2) the model captures the operational interdependency aspects inherent in the parameter learning process so that proposed modeling framework can predict both runway configuration and AARs simultaneously with higher accuracy than shown in other studies in the existing literature.

The remainder of this paper is organized as follows. Section 2 reviews the existing literature of prediction of runway configurations and AAR. Section 3 provides the background of the New York Metroplex, the case study for demonstrating the proposed method. Section 4 describes the data collection and processing techniques used in this study. Section 5 describes the deep-learning methodology. Section 6 presents the experimental results from case study of New York Metroplex and compares the model performance with the state of the art. Finally, Section 7 summarizes this study and points out future research directions.

2. Literature reviews

Airport capacity is defined as the maximum sustainable throughput, i.e., the number of aircraft operations an airport can accommodate under continuous demand for a certain period (1 h or 15 min). Airport capacity can be categorized into two types—theoretical capacity and dynamic or real-time capacity (Fisher, 2012). Theoretical capacity is the optimal capacity an airport can achieve theoretically, given a specific runway configuration, meteorological condition (leading to different separation requirements of leading and trailing aircraft), arrival and departure fleet mix, etc. Different analytical methods and simulation models have been developed to obtain theoretical capacity. Such capacity is usually estimated for planning purposes—for example, to determine if new runway construction is needed or slot allocation should be changed. Real-time capacity, currently determined by ATCs (airport acceptance rate, AAR), is the capacity affected by real-time operational conditions and used for decision-making of air traffic flow management (ATFM). It is usually determined several hours ahead for tactical planning (look-ahead horizon 0–2 h) or strategic planning (look-ahead horizon 1–8 h).

Previous studies explored data-driven methods to predict real-time capacity (AAR) for strategic planning (i.e., forecast horizon 1–8 h). Data-driven methods can be categorized into two groups—scenario-based approach and machine learning approach.

The scenario-based approach aims to create arrival capacity distribution profiles over a period (usually one day) that can be used in one of the TMI Ground Delay Program (GDP) models. For example, Liu et al. (Liu et al., 2008) studied single airport GDP using an airport capacity scenario-based tree method. In their study, the capacity scenario was defined as a time series of AARs. Clustering analysis was used for analyzing historical AARs and identifying airport specific capacity scenarios, which were used as the input for optimizing GDP. Their research was expanded by Buxi et al. (Buxi and Hansen, 2011) by including weather forecast elements from Terminal Aerodrome Forecast (TAF) when estimating capacity profile. The authors used Principle Component Analysis (PCA) to generate primary and uncorrelated components of meteorological variables obtained from TAF on day-of-operation basis; then, the principle components were combined with historical AAR to identify capacity profiles through clustering analysis.

The machine learning approach aims to translate weather forecasts and historical AARs into probabilistic AARs for a given forecast horizon. Most studies used aviation weather products such as weather observations from the Meteorological Terminal Aviation Routine Weather Report (METAR) and weather forecast from TAF and the Localized Aviation Model Output Statistical Program (LAMP). Prediction of AARs is usually formulated as a time-dependent problem, where future AARs are dependent on previous AARs. For example, Provan et al. (Provan et al., 2011) proposed an AAR prediction model called Weather Translation Model for GDP Planning (WTMG) using an assembled decision-tree model that took weather forecast elements obtained from TAF and LAMP and previous-hour AAR as the inputs to predict the AAR for a look- ahead horizon of 1–12 h. Cox and Kochenderfer (Cox and Kochenderfer, 2016) adopted the same inputs of WTMG and proposed to use a Bayesian network-based AAR prediction model called AAR Distribution Prediction Model (ADPM).

In addition to aviation weather forecasts, other studies used weather information obtained from the comprehensive weather forecast product, Rapid Cycle (RUC, now called Rapid Refresh or RAP) from the National Oceanic and Atmospheric Administration (NOAA) to predict AARs. Wang et al. (Wang, 2011) (Wang, 2012) selected 32 surface weather forecast elements from RUC to predict hourly and 15-min AARs. The authors applied two methods, quadratic response surface linear (QRS) regression and bagging decision tree (BDT) regression, to study three airports with high air traffic demand—EWR, ORD, and ATL. Results from the study showed that using weather elements from RUC to predict AAR could achieve better accuracy than those using METAR.

Some studies also investigated factors that had impacts on the determination of AARs. Chung and Murphy (Chung and Murphy, 2010) applied linear regression to estimate AARs by combining airport operational information from Aviation System Performance Metrics (ASPM) and controller input information from the National Traffic Management Log (NTML) to obtain the best fit model by including different input variables. Airports studied in their paper included Newark Liberty International (EWR), John F. Kennedy International (JFK), LaGuardia (LGA), Dallas/Fort Worth International (DFW), Chicago O'Hare International (ORD), Los Angeles International (LAX), Seattle-Tacoma International (SEA), and San Francisco International (SFO). The authors found that a combination of factors providing the best predicted AAR were endogenous variables including runway configuration, controller input weather, and AAR adjustments. Wind is normally a significant factor in runways selection as some aircraft cannot accept any tailwind component. Maximum allowable tailwind is 10 knots for runway selection (FAA Order 8400.9). DeLaura et al. (DeLaura, et al., 2014) analyzed the impacts of surface wind and wind loft obtained from RUC to key factors (i.e., runway configuration, aircraft ground speed, spacing on final approach) associated with AARs. Their study found that the impacts of several wind metrics derived from wind loft were significant, e.g., surface headwinds and downstream capture box headwinds.

These aforementioned studies considered only an airport as an isolated entity without considering interactions between airports within the same metroplex if applicable. Only one study in the existing literature studied AARs in the context of a multi-airport system (Murça, 2018). This study identified trajectory patterns in a New York metroplex (Murça, 2018), then combined the patterns with convections along the arrival route and weather elements from TAF to predict AARs for three major airports—JFK, LGA, and EWR—using Gaussian process regression.

There are several caveats in existing studies on predicting AARs using data-driven methods: (1) although AARs are highly dependent on selected runway configurations, this characteristic was not well-captured in previous studies; only a few studies explicitly included the runway configuration selection process in the determination of AARs, but the algorithm used cannot be applied to airports with a high variation of AARs; (2) most studies were focused only on a single airport scenario, although in the context of a multi-airport system, both runway configurations selection and AARs determination of one airport are correlated to each other; and (3) the only existing literature predicting AARs in the context of a multi-airport system was dependent on real-time flight trajectories; however, flight trajectory patterns were processed off-line so their method cannot be used for real-time prediction. Also, flight trajectory patterns are heavily dependent on weather conditions, which makes these two factors have an endogenous effect. Including both in prediction causes an endogeneity problem in regression models.

Studies on runway configuration selection have focused on finding the best runway configuration sequencing based on optimizing of the airport as a queueing system or studying historical patterns to predict runway configurations using statistical methods.

Descriptive methods try to optimize the runway configuration selection process in a deterministic or stochastic queueing model by considering meteorological condition, wind information, controller workload (as a time buffer or penalty cost), etc. A significant challenge to the implementation of these operational descriptive methods is the modeling of the constraints and objectives of the human air traffic operators. Li et al. (Li and Clarke, 2010) proposed a stochastic dynamic programming to model runway configuration planning as a sequential decision-making process to maximize total arrival and departure rates. In each look-ahead horizon, a decision tree was constructed using stochastic wind information. Bertsimas et al. (Bertsimas et al., 2011) proposed a Mixed Integer Programming model to solve the single airport runway selection and arrival/departure balance problem for an entire day of operation by optimizing queuing delays, assuming the knowledge of respective capacity and feasible runways at a given time, and adding the changeover time of runway configuration as a constant value. They extended the work by considering marginally decreasing penalties associated with runway configuration switches in (Weld et al., 2010:). NASA's System-Oriented Runway Management concept (Lohr and Atkins, 2015) aimed at designing decision-support tools considering both strategic prediction of runway configuration and tactical decision support that helped air traffic controllers choose the optimal runway configuration and even the associated arrival-departure mix for every 15-min interval. A study by Jacquillat et al. (Jacquillat et al., 2016) proposed a dynamic programming model that combined runway configuration and airport service rate to minimize the congestion cost in a stochastic queuing system.

Other studies investigated historical data related to runway configurations and used statistical models to predict runway configurations by considering weather conditions. For example, Ramanujam et al. (Ramanujam and Balakrishnan, 2015) and Avery et al. (Avery and Balakrishnan, 2016) proposed a nested discrete choice model and used empirical observations to estimate the configuration selection process for look-ahead times of up to 6 h. Huston et al. (Houston and Murphy, 2012) applied logistic regression to predict runway configuration using airport operation information; however, their studies used weather observations instead of weather forecasts to predict runway configurations, which were not the actual situation.

A series of studies also attempted to model runway configuration selection and airport capacity jointly (Dhal, 2014) (Tien, 2015) (Tien, 2018). They developed a multi-stage framework to determine the runway configuration and AAR sequentially. First, a database with historical frequent runway configurations and associated wind and meteorological conditions was developed. Then, the eligible runway configuration was determined by the forecasted wind condition and frequency of previous usage; AAR was determined based on forecasted meteorological condition and average AARs of selected runway configurations. Tien et al. (Tien, 2015) validated the model by applying it to 35 major airports in the U.S., and the study was extended by including weather forecast elements obtained from Short Range Ensemble Forecast (SREF) (Tien, 2018).

Existing studies on runway configuration selection have several caveats: (1) similar to literature on AAR prediction, all studies in the existing literature on runway configuration selection focused only on a single-airport scenario; no interdependent operation aspects for multi-airports system were considered; and (2) both descriptive optimization methods and statistical methods in the literature did not use weather forecasts well; instead, they applied very limited wind-related variables from observed weather to model the runway configuration.

A multi-airport system (or metroplex) is a set of two or more airports operating in close proximity. Because they share the same terminal airspace and usually the same TRACON, these airports need to synchronize their operations, including coordinating runway configuration selection and capacity allocation to de-conflict air traffic flows (Ren, 2009).

The only existing study on the prediction of AARs for a multi-airport system is (Murça, 2018). The authors first identified six primary traffic flow patterns in a New York multi-airport system by clustering historical flight trajectory data. The authors fed the flow pattern together with visibility and wind condition data from TAF and convection status from Arrival Route Status and Impact (ARSI) forecast into a Gaussian process regression model to predict hourly probabilistic AARs for the three airports. The outcomes of the study were dependent on the accurate prediction of trajectory patterns; however, it is difficult to predict trajectory patterns, and they cannot be generated in real time. Also, flight trajectory patterns are heavily dependent on weather conditions. Furthermore, this study did not consider runway configuration impact on the determination of AARs.

Overall, existing data-driven methods on the prediction of runway configurations and AARs focus primarily on a single airport scenario, but the interdependency aspects among airports within a multi-airport system have not been well studied. To predict runway configuration and AARs of airports in a multi-airport system, the following must be considered: (1) other weather forecast sources than TAF and LAMP, because they are station-based, both using one data point to represent weather conditions in the vicinity of an airport; for a multi-airport system, however, larger-scale weather information need be obtained and used; and (2) because runway configurations are selected in a synchronized fashion and airport capacity is heavily related to the selected runway configurations, the prediction of runway configurations and AARs should be modeled jointly.

3. New York metroplex

The New York multi-airport system (NY-MAS) is composed of three major commercial airports—JFK, EWR, and LGA—as well as several secondary commercial and general aviation airports. According to the (Bureau of Transportation Statistics, 2018), approximately 5.4% of domestic flights and 9.0% of international flights in the U.S were served by these three major commercial airports. Increasing air traffic volume and capricious weather conditions make NY-MAS one of busiest and most complex systems for air traffic management and operations in the U.S. and worldwide.

TRACON-N90 is a consolidated operation center that provides air traffic control and management service to NY-MAS. For JFK, EWR, and LGA, each has an Air Traffic Control Tower (ATCT) that is responsible for supervising safe air traffic control, which includes but is not limited to directing aircraft landing, take-off, and taxi and runway scheduling. Given the limited airspace and high density of

Table 1
Primary runway configurations at NY-MAS in 2015.

Frequency	JFK Runway Configuration	LGA Runway Configuration	EWR Runway Configuration
9.1%	13L,22L|13R	22|13	22L|22R
5.8%	31L,31R|31L	31|4	4R|4L
4.2%	31L,31R|31L	22|31	22L|22R
3.3%	22L,22R|22R,31L	22|31	22L|22R

air traffic, airports in NY-MAS need to coordinate runway scheduling and capacity planning. Table 1 lists the most frequently used runway configuration sets in NY-MAS in 2015, and Fig. 1 shows the top two primary used runway configurations in NY-MAS. Depending on spatial weather conditions and airport operations, the metroplex will have different dominant arrival and departure streams (Murça, 2018). The runway configurations shown in red are those used 9.1% of the time in 2015, serving south arrival flow, i. e., aircraft mainly entering the terminal airspace from the south corner posts. The runway configurations used 5.8% of the time in 2015 are shown as blue arrows, which correspond to the north arrival flow (Murça, 2018).

4. Data collection and processing

Weather is a critical component for determining runway configuration and airport capacity. Current weather forecast products available at TMUs are fragmented. One important NextGen program is to integrate weather information and translate weather conditions into a decision support tool for traffic management.

Rapid Refresh (RAP) is a numerical weather model that covers the continental U.S. (CONUS) domain and is run by the National Centers for Environmental Prediction (NCEP); the previous version is called Rapid Update Cycle (RUC). RAP data are the continental-scale numerical weather forecast data disseminated and maintained by NOAA and have been used by numerous entities to obtain accurate short-term weather forecast information. They provide severe weather and hazard information together with other general weather condition elements aloft and on the surface. Currently, RAP has two versions regarding weather data resolutions; the first generated weather forecast on a 13-km (8-mi) resolution horizontal grid, and the second, called the High-Resolution Rapid Refresh (HRRR), generates data on a 3-km (2-mi) resolution grid (not archived by NOAA). RAP forecasts are generated and disseminated each hour with a forecast horizon from the current hour to up to 23 h. Radar data, surface observations, and satellite data are the sources used for the generation of RAP forecasts. In this study, RAP data with 13-km resolution were used, as archived by the NOAA National Center for Environmental Information and can be requested to download from the server.

RAP weather forecasts are processed and archived in gridded binary format (.grib2 file). In the RAP 13-km resolution version, each grid has location information represented by latitude and longitude and 315 weather forecast elements/variables distributed in multiple vertical layers—for example, surface precipitation and 255–0 mb above ground pressure. (A subset of the elements from RAP data is later shown in Table 2.)

Selection of elements that could be used in this study are discussed in Section 5.2. RUC data (the previous version of RAP) were used in the literature to predict AAR (Wang, 2011) (Wang, 2012); however, the authors extracted only surface weather forecast elements from the closest grid of the study airport. RAP data can be better used for two reasons: (1) they include spatial weather information, which can be very important for multi-airport system runway configuration and AAR prediction where weather information for a large terminal airspace is needed; and (2) they provide a comprehensive dataset with many aloft weather elements such as convection inhibition and storm relative helicity. Such information can be critical variables for runway configuration selection and airport capacity prediction.

The RAP weather forecast is ensembled gridded data and is different from a station-based aviation weather forecast, where a single point represents the regional weather condition. To use the spatial feature, the RAP data were decoded and reshaped from the Gridded binary format to three-dimensional tensor data, as illustrated in Fig. 2. Each 13 km $\times$ 13 km grid is represented by one corresponding latitude and longitude. The height dimension indicates different layers that contain different weather forecast elements—a total of 315 in the sample illustration. Note that the height does not represent altitude; rather, it represents layers defined in RAP, such as surface, 90 mbar above ground, etc.

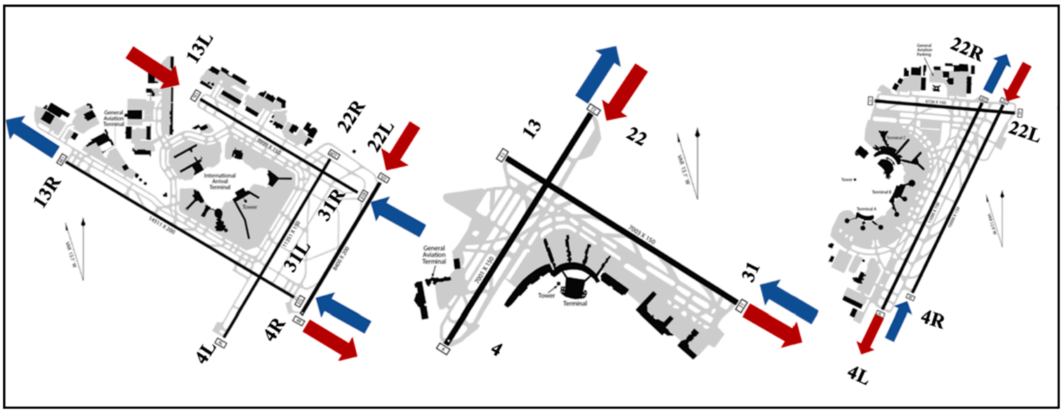


Fig. 1. Top 2 primary runway configurations of NY-MAS in 2015.

Table 2
Selected weather elements from RAP.

Layer	Acronym	Description	Unit
1000 m above ground	REFD	Derived radar reflectivity	dbZ
1000 m above ground	HLCY	Storm relative helicity	m2/s2
180–0 mb above ground	4LFTX	Best (4-layer) lifted index	K
180–0 mb above ground	CAPE	Convective Available Potential Energy	J/kg
180–0 mb above ground	CIN	Convective inhibition	J/kg
2 m above ground	DEPR	Dew point depression (or deficit)	K
2 m above ground	DPT	Dew point temperature	K
2 m above ground	POT	Potential temperature	K
2 m above ground	RH	Relative humidity	%
2 m above ground	SPFH	Specific humidity	kg/kg
2 m above ground	TMP	Temperature	K
255–0 mb above ground	CAPE	Convective Available Potential Energy	J/kg
255–0 mb above ground	CIN	Convective inhibition	J/kg
255–0 mb above ground	PLPL	Pressure of level from which parcel was lifted	Pa
3000–0 m above ground	HLCY	Storm relative helicity	m^2/s^2
4000 m above ground	REFD	Derived radar reflectivity	dbZ
500 mb	ABSV	Absolute vorticity	/s
500 mb	HGT	Geopotential height	gpm
500 mb	RH	Relative humidity	%
500 mb	TMP	Temperature	K
500 mb	UGRD	u-component of wind	m/s
500 mb	VGRD	v-component of wind	m/s
500 mb	VVEL	Vertical velocity (pressure)	Pa/s
500–1000 mb	LFTX	Surface lifted index	K
6000–0 m above ground	USTM	u-component of storm motion	m/s
6000–0 m above ground	VSTM	v-component of storm motion	m/s
6000–0 m above ground	VUCSH	Vertical u-component shear	/s
6000–0 m above ground	VVCSH	Vertical v-component shear	/s
90–0 mb above ground	CAPE	Convective Available Potential Energy	J/kg
90–0 mb above ground	CIN	Convective inhibition	J/kg
Convective cloud top level	HGT	Geopotential height	gpm
Entire atmosphere	PWAT	Precipitable water	kg/m^2
Entire atmosphere	REFC	Maximum/Composite radar reflectivity	
Entire atmosphere	RETOP	Radar Echo Top (18.3 DBZ)	m
Entire atmosphere	RHPW	Relative humidity with respect to precipitable water	%
Entire atmosphere	TCDC	Total cloud cover	%
Equilibrium level	HGT	Geopotential height	gpm
High cloud layer	HCDC	High cloud cover	%
Low cloud layer	LCDC	Low cloud cover	%
Middle cloud layer	MCDC	Medium cloud cover	%
Surface	ACPCP	Convective precipitation	kg/m^2
Surface	BGRUN	Subsurface runoff (baseflow)	kg/m^2
Surface	CAPE	Convective Available Potential Energy	J/kg
Surface	CFRZR	Categorical freezing rain (yes = 1; no = 0)	non-dim
Surface	CICEP	Categorical ice pellets (yes = 1; no = 0)	non-dim
Surface	CIN	Convective inhibition	J/kg
Surface	CRAIN	Categorical rain (yes = 1; no = 0)	non-dim
Surface	CSNOW	Categorical snow (yes = 1; no = 0)	non-dim
Surface	EPOT	Potential temperature	K
Surface	GUST	Surface wind gust	m/s
Surface	HGT	Geopotential height	gpm
Surface	HINDEX	Haines Index (dryness)	
Surface	HPBL	Planetary boundary layer height	m
Surface	LTNG	Lightning	non-dim
Surface	MSTAV	Moisture availability	%
Surface	NCPCP	Large scale precipitation (non-conv.)	kg/m^2
Surface	PRATE	Precipitation rate	$kg/m^2/s$
Surface	PRES	Pressure	Pa
Surface	PTEND	Pressure tendency	Pa/s
Surface	SNOD	Snow depth	m
Surface	SSRUN	Storm surface runoff	kg/m^2
Surface	TMP	Temperature	K
Surface	VIS	Visibility	m

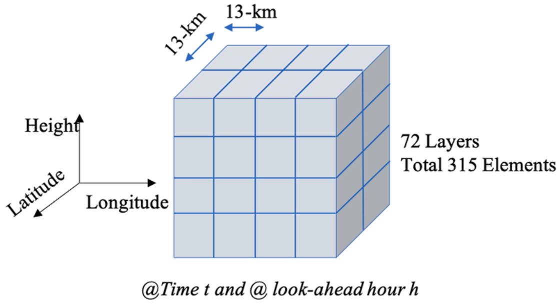


Fig. 2. Illustration of processed RAP data.

5. Methodology

5.1. Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a class of Deep Neural Network that is applied widely in image learning and video processing. Different from DNN, which takes vectors as inputs, the inputs of CNN are 3-D tensors, where the first two dimensions are horizontal and vertical positions and the third is the features or channels. For image classification and schematic analysis, the horizontal and vertical positions are the location of a certain pixel, and the features are usually RGB values that comprise the one pixel. In this problem, the inputs are also 3-D tensor—the positions are latitude and longitude of a certain grid, and features are weather forecast elements in each grid.

The core technique of a CNN layer is using a 3-D matrix called a kernel, taking the dot product of the kernel and the extracted windows of the same size from the input tensor and applying a non-linear activation function to obtain a feature map that contains various representations of the input tensor. Each kernel is set an initial weight, and the weight is updated after each learning iteration during the training process using a stochastic gradient decent algorithm. In this way, the model can handle complex non-linear spatial problems in which the network can learn to transform the input into the most representative features that can be used for classification and regression.

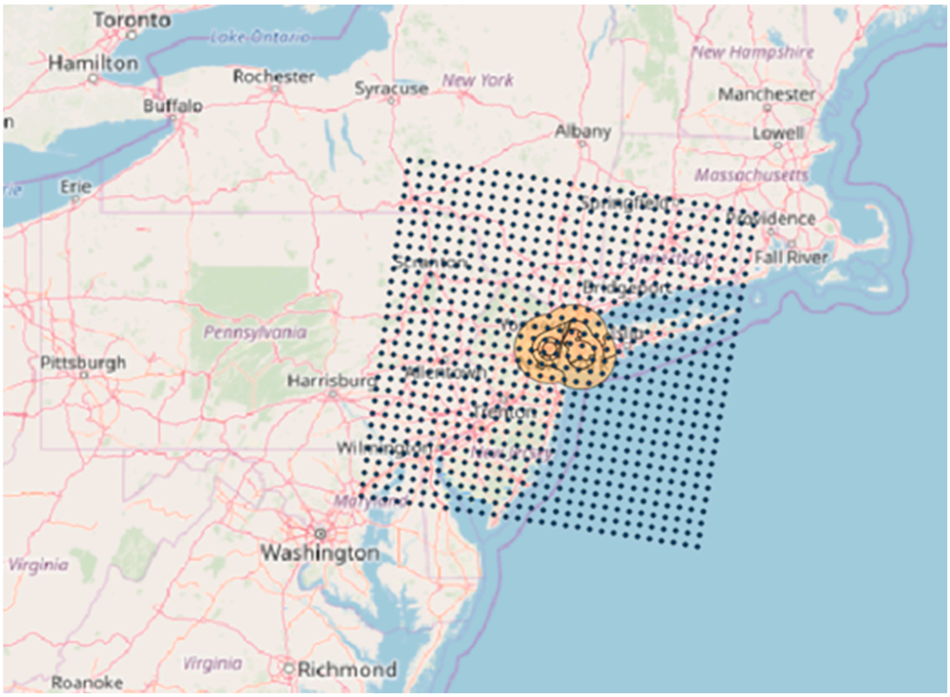


Fig. 3. Selected RAP weather forecast area in grids.

5.2. Model inputs

Data used in this study were the hourly RAP weather forecast in 2015 within 200 nautical miles in the vicinity of NY-MAS, as illustrated in Fig. 3. RAP weather forecast data of a 1–18-h look-ahead horizon in 2015 were requested from NOAA's National Centers for Environmental Prediction (NCEP) servers. The data were then decoded and pre-processed by filtering latitude and longitude to include an area of a 200 × 200 nautical miles rectangular terminal area in the vicinity of New York, as shown in Fig. 3. The 200 × 200 nautical miles area is represented by 29 × 29 grid points, each containing weather elements representing the weather condition of the 13 × 13 km grid; the associated latitude and longitude are the centroid of each grid.

Based on the literature review discussed in Section 2, 64 weather forecast elements from 18 vertical layers were selected. These selected weather-forecast elements included 24 surface variables and 20 aloft variables, as shown in Table 2. Among the 64 elements, 4 are categorical variables and 60 are numerical variables. The numerical variables selected are on a wide range of scales. Input and output variables that have large scale discrepancies can make the optimization process difficult because of the large gradient during CNN training. Thus, a maximin normalization method was applied to scale all input weather elements into (Bureau of Transportation Statistics (BTS), 2018), following the function $X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$, where X_{norm} is the normalized vector of input variable, X is the original vector of input elements, and X_{min} and X_{max} are the minimal and maximum vector of input features.

5.3. Model outputs

Airport hourly operation reports from the ASPM database were used to obtain runway configuration and AARs for JFK, LGA, and EWR. In total, 41 runway configurations were recorded at JFK throughout 2015; however, only 17 were used more than 1% of the time. At LGA and EWR, 14 and 10 primary runway configurations were used more than 1% of the time, respectively; all other runway configurations used less than 1% of the time at each airport were noted as “Other.” The primary runway configurations for the three airports and the corresponding AAR distributions are shown in Figs. 4, 5, and 6. JFK and EWR did not have a case in which AARs were equal to 0; however, LGA had 15 AARs equal to 0 caused by airport closure or other operational concerns. These data were not excluded from the model because they can capture the impacts of convective weather on airport capacity. From the distribution of AARs at these three airports, AARs are not standard distributed for most runway configurations. Also, the layout of runway configuration could have an impact on the variations of AAR distributions. For JFK and LGA, with more intersecting runways involved, such as 22L, 22R|22R, 31L and 22|13, the higher variations of AAR distribution are presented.

5.4. Runway configuration and AAR net – Multi-CNN

The overall modeling framework is shown in Fig. 7. The model includes four multi-layer CNNs that take in the same input tensor

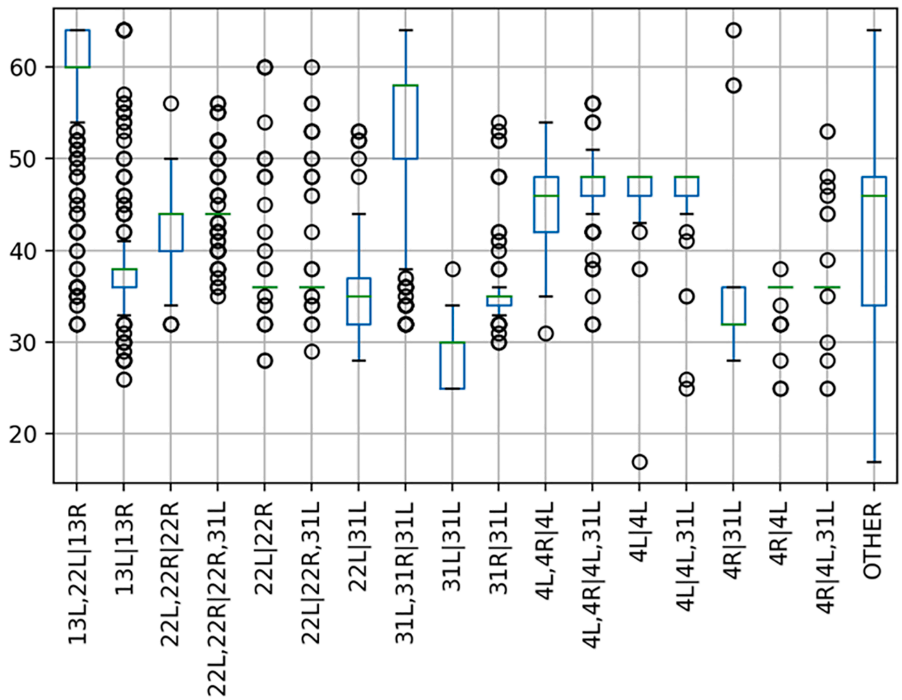


Fig. 4. JFK primary runway configurations (>1%) and corresponding AAR distribution.

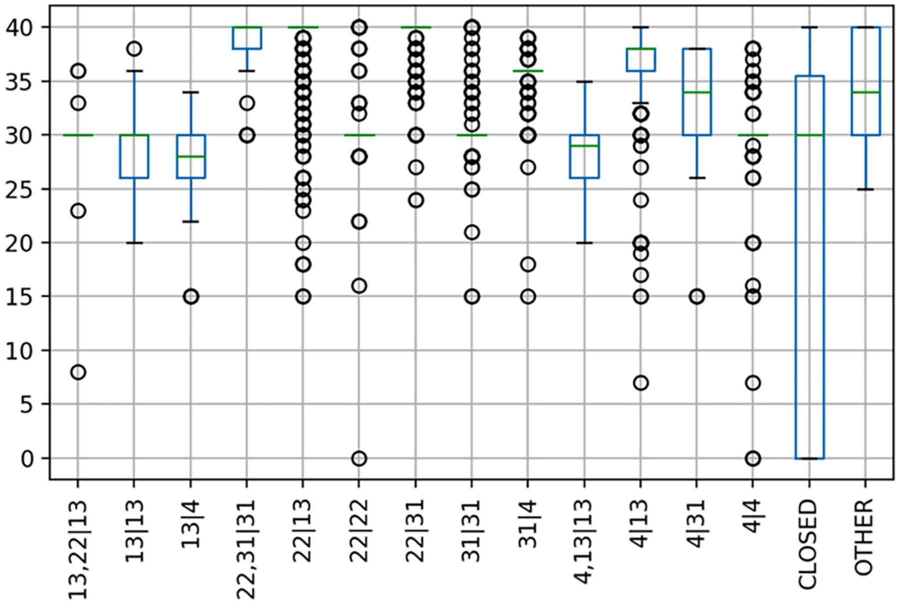


Fig. 5. LGA primary runway configurations (>1%) and corresponding AAR distribution.

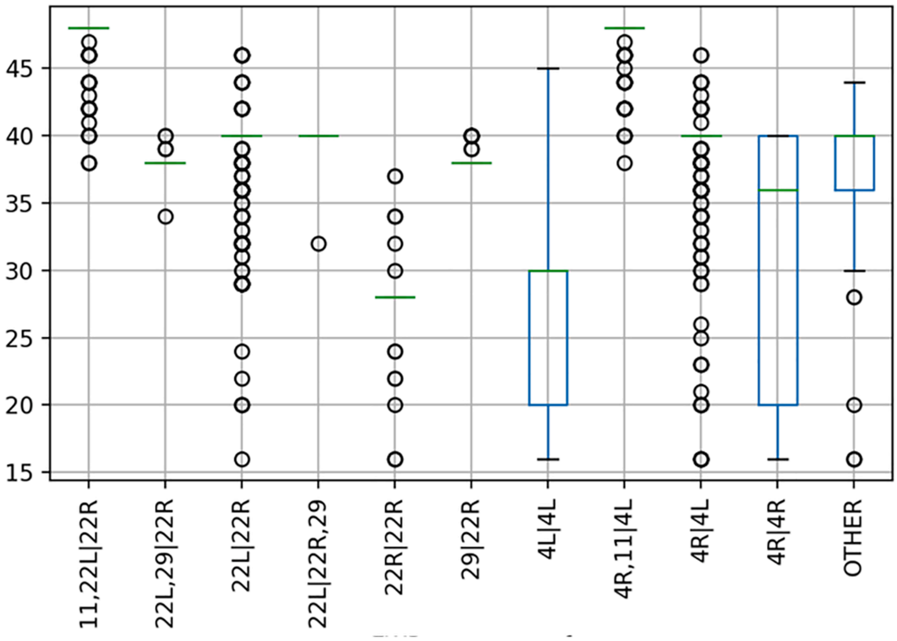


Fig. 6. EWR primary runway configurations (>1%) and corresponding AAR distribution.

and then feed into different multi-layer CNN models for different training purposes. The top three CNN branches shown in Fig. 7 have the same model architecture (see Fig. 8) and are used for predicting hourly runway configurations for each airport. The fourth CNN branch is used for predicting AARs of the three airports (see Fig. 8 for model architecture). The overall model input is a 3-D tensor with a size of $29 \times 29 \times 64$, where 64 is the normalized weather elements produced from processed RAP gridded data and the outputs are airport runway configurations and AARs within the study multi-airport system. The output of each runway configuration CNN branch is an $n \times 1$ vector, where n is the number of runway configuration categories (classes) for that airport. The outputs of AAR CNN branches are AAR values for each airport.

Prediction of runway configurations is a classification problem and prediction of AARs is a regression problem; thus, two types of function were used to calculate the loss for each model branch. In the runway configuration CNN branch shown in Fig. 8, we applied

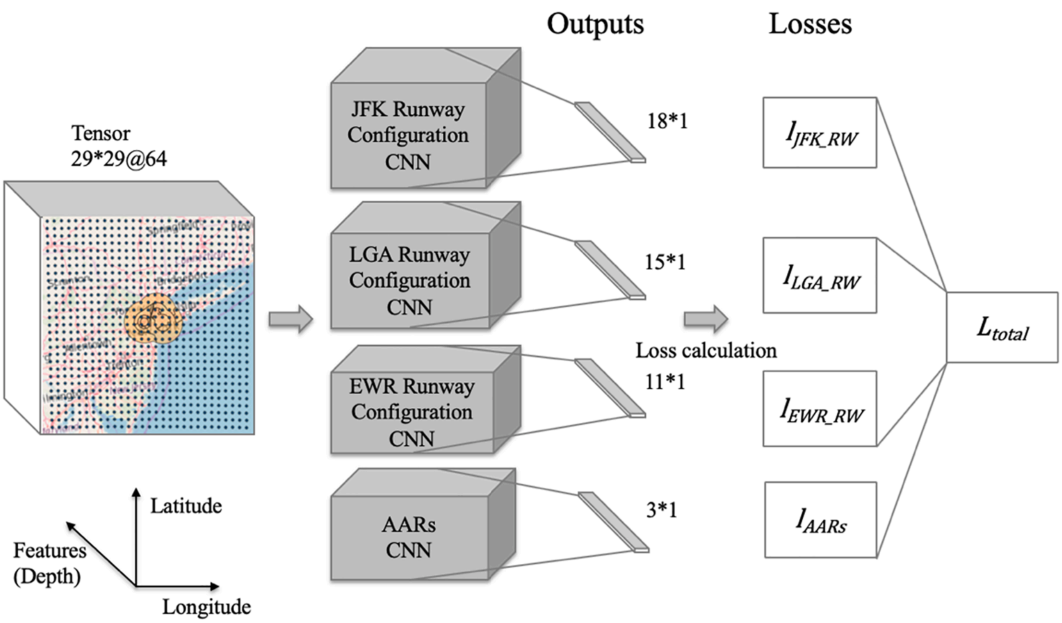


Fig. 7. Multi-airport system RWAAR modeling framework.

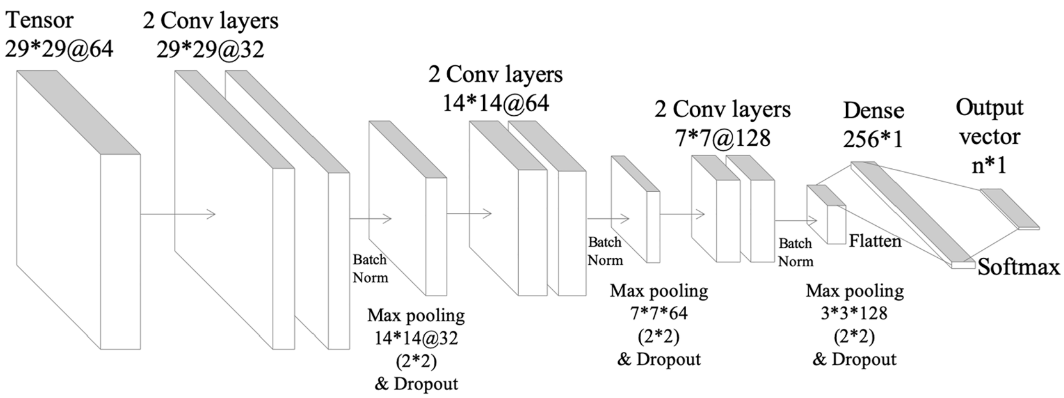


Fig. 8. Runway configuration CNN branch.

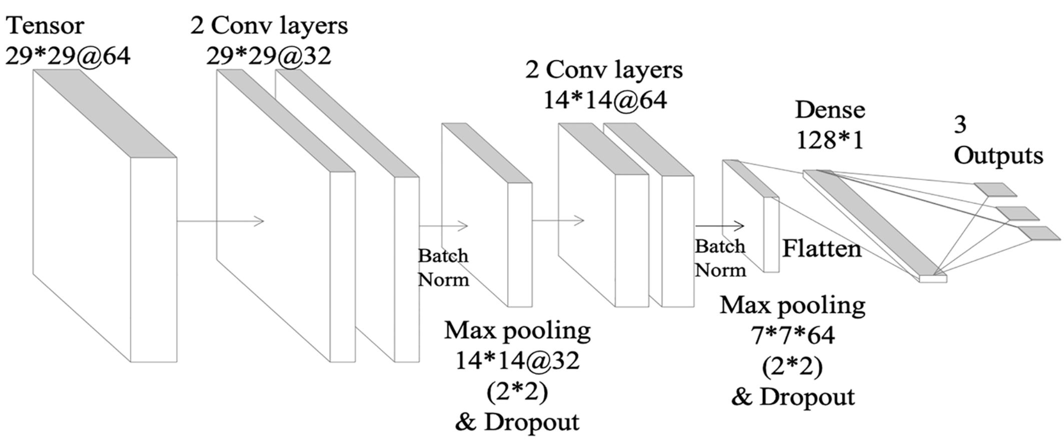


Fig. 9. AARs CNN branch.

softmax $\frac{exp(y_i)}{\sum_j y_j}$ to compute the probability that the training sample i belongs to class j. Then we used cross entropy $-\sum_i^K y_i log(\hat{y}_i)$ to calculate overall model loss, where y_i is i^{th} element of target vector, $\hat{y}_i$ is i^{th}element of probability obtained from softmax function, and K is the number of classes (also the length of output vector, in this case the class of runway configurations). In the AAR CNN branch illustrated in Fig. 9, the loss function used is the root mean square error $\sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$, where $y_i$is the i^{th}element of the true AAR value and $\hat{y}_i$ is the i^{th} element of the predicted value of AARs in a batch. The following weighted loss was used to calculate the total loss of the whole RWAAR net using function:

$$L_{Total} = W_{JFK\_RW} l_{JFK\_RW} + W_{LGA\_RW} l_{LGA\_RW} + W_{EWR\_RW} l_{EWR\_RW} + W_{AARs} l_{AARs}$$

where l_{JFK_RW}, l_{LGA_RW}, l_{EWR_RW}, and l_{AARs} are the losses calculated from each CNN branch, as shown in Fig. 7.

The runway configuration CNN branch takes in the 3-D input tensor and feeds it into a two-layer CNN with a kernel size of 3×3 and using ReLU activation function $f(z) = max(0, z)$. The convoluted outputs are then normalized and filtered through a max pooling layer with dropout rate 0.3. Finally, two more double convolutional layers are applied followed by max pooling and dropout layers. The output of this branch is a $n \times 1$ vector, where n is the number of runway configurations. Fig. 9 shows the AAR model branch with only two double convolutional layers. The hyperparameters used in this branch are the same as those in runway configuration branches.

6. Experiment results

6.1. Results analysis

In this study, each sample is a daily series of hourly data from 6am to 12am (18 observations each day). The total number of samples for each look-ahead hour is 6,570, among which 80% of the days were randomly selected as the training set and the rest 20% of the days as testing set. In such way, potential temporal orders in the dataset can be handled. After tuning different weights of the total loss function, the weight ratio leading to best performance was W_{JFK_RW}: W_{LGA_RW}: W_{EWR_RW} : W_{AARs} $= 10 : 10 : 10 : 0.75$. The predictive performance of the proposed RWAAR net was evaluated for look-ahead horizons for the next 1–8 h. Two statistical performance metrics—root mean square error (RMSE) and mean absolute error (MAE)—were used to evaluate the predictive performance.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

$$MAE = \frac{1}{n}\sum_{i=1}^n |\hat{y}_i - y_i|$$

where $\hat{y}_i$ is the predicted AAR for sample i, y_i is the true value for sample i, and n is the total number of samples in the testing dataset.

The accuracy of runway configuration prediction for NY-MAS is shown in Fig. 10. Throughout the forecast horizons, the average accuracies of predicting airport hourly runway configurations for JFK, LGA, and EWR are 79.21%, 85.86%, and 87.25%, respectively.

The performance of AAR prediction for NY-MAS is shown in Fig. 11. The model can predict AARs at JFK with an average root mean square error 5.85 or mean absolute error 4.2 throughout the look-ahead horizons. The proposed modeling framework works better for predicting AARs of LGA and EWR, with the root mean square error 2.91 and 2.29, respectively, and the mean absolute error 2.13 and 1.67, respectively. Model performance kept steady throughout the forecast horizons. It is reasonable that the learning model achieved better predictive performance for LGA and EWR because JFK has more runway configurations from which air traffic control personnel can select and JFK also has more disperse AAR distributions, as shown in Fig. 4. Overall performance measurements for these three airports are shown in Table 3.

As discussed in the literature review, the most recent study on AAR prediction for airports in a metroplex is (Murça, 2018), which applied a Gaussian process regression model to study the three major airports in NY-MAS. The authors used the trajectory tracks and weather forecast in the selected 69 days from 2013 to 2015 to predict AARs for JFK, LGA, and EWR for 1–8-h look-ahead horizons and applied a mean absolute percentage error (MAPE) $\frac{100\%}{n}\sum_{i=1}^n \left|\frac{\hat{y}_i - y_i}{y_i}\right|$ to measure the model performance. To compare the predictive performance of the proposed RWAAR with the model in (Murça, 2018), MAPE from JFK and EWR AAR predictions were calculated. For LGA, because AAR $= 0$ cases were presented, MAPE is not an appropriate measurement for this airport. Previous studies on runway configuration selections for a single airport took NY airports as their case study, and (Avery and Balakrishnan, 2016) applied discrete choice model to predict 15-min runway configurations for SFO, EWR, and LGA using weather observations from ASPM in 2011 and 2012. (Houston and Murphy, 2012) used logistic regression to predict runway configurations for JFK and LGA. The prediction accuracy of runway configuration can be compared from the proposed RWAAR net with the results from these two studies.

Comparison of performance of the proposed RWAAR net with those from the existing literature is summarized in Table 4. For AAR prediction, the proposed RWAAR net can achieve lower MAPE for JFK and EWR. Especially for EWR, the proposed model can achieve 5.2% lower error. Although MAPE for LGA could not be calculated, as shown in Table 4, prediction of AARs for LGA and EWR have a similar level of predictive performance; thus, presumably, AAR prediction for LGA would have the similar level of MAPE with EWR, which is better than that from the existing literature. For runway configuration prediction, compared with current-state-of-art, RWAAR

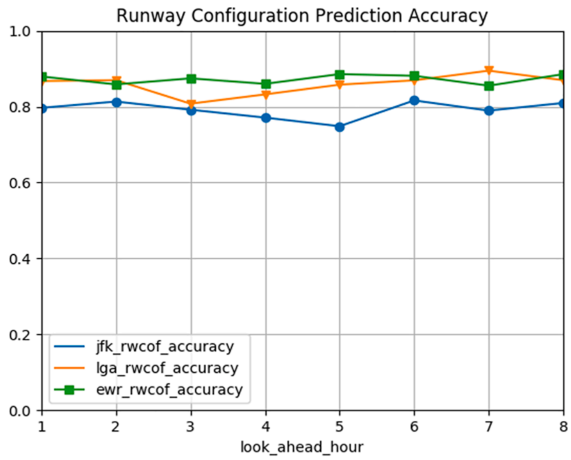


Fig. 10. Runway configuration prediction performance for New York multi-airport system.

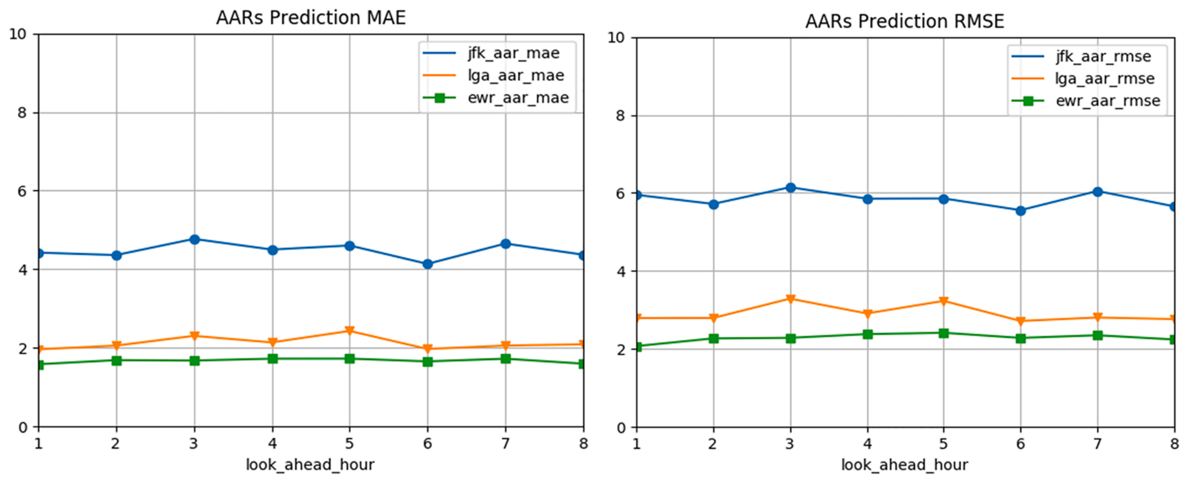


Fig. 11. AAR prediction performance for New York multi-airport system.

Table 3
RWAAR net overall performance.

Look-ahead hour	JFK runway configuration accuracy	LGA runway configuration accuracy	EWR runway configuration accuracy	JFK AAR		LGA AAR		EWR AAR	
				RMSE	MAE	RMSE	MAE	RMSE	MAE
1	79.7%	86.7%	87.9%	5.95	4.42	2.79	1.96	2.07	1.58
2	81.3%	87.0%	85.9%	5.72	4.36	2.79	2.06	2.27	1.69
3	79.2%	80.8%	87.5%	6.14	4.77	3.29	2.31	2.29	1.68
4	77.1%	83.2%	86.0%	5.85	4.50	2.91	2.14	2.38	1.73
5	74.8%	85.8%	88.6%	5.86	4.60	3.23	2.43	2.41	1.73
6	81.6%	86.9%	88.1%	5.56	4.14	2.72	1.97	2.28	1.66
7	79.0%	89.5%	85.5%	6.05	4.65	2.81	2.06	2.35	1.72
8	81.0%	87.0%	88.5%	5.65	4.37	2.76	2.09	2.24	1.60
Avg.	79.21%	85.86%	87.25%	5.85	4.48	2.91	2.13	2.29	1.67

can achieve 16.2% higher accuracy for JFK, 4.6% higher accuracy for LGA, and 9.45% higher accuracy for EWR. Note that in this study, weather forecast was used to predict runway configuration for the strategic planning horizon (1–8 h), whereas previous studies used weather observation in future hours, which were unknown at the time when the prediction needed to be performed.

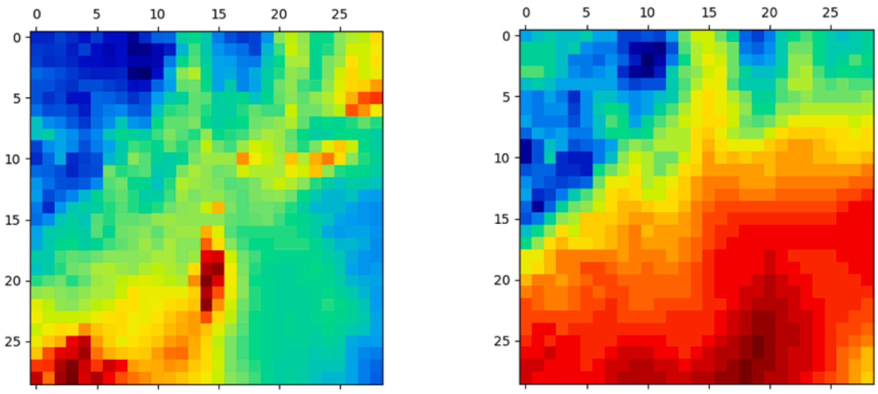
Overall, the proposed RWAAR net can predict runway configurations and AARs simultaneously for a metroplex using an ensembled gridded weather forecast. The results from the case study of NY-MAS show that the RWAAR net achieves better performance with regard to prediction of runway configurations and AARs compared with the outcomes of studies in the existing literature.

Table 4
Runway configurations and AAR prediction performance compared with current state-of-the art.

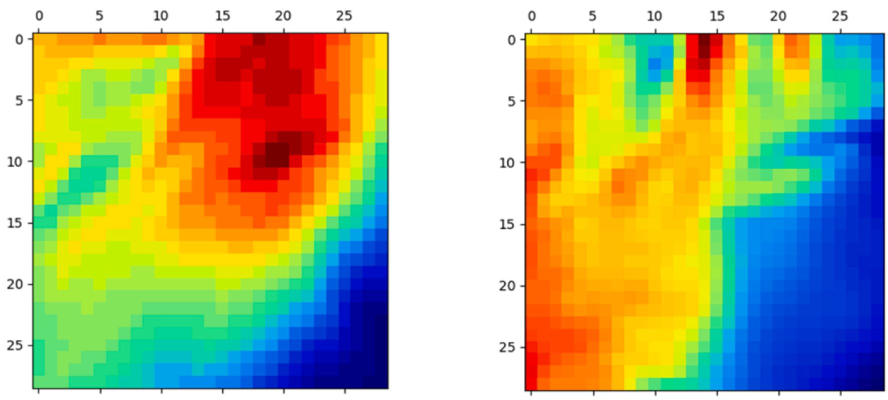
	Runway configuration prediction accuracy			AAR prediction MAPE		
Airport	JFK	LGA	EWR	JFK	LGA	EWR
RWAAR Net	79.2%	85.9%	87.25%	10.6%	N/A	4.3%
Murça and Hansman (2018)				13.6%	8.6%	9.5%
Jacob and Balakrishnan (2016)		81.3%	77.8%			
Houston, Stephanie et al (2012)	63%	75%				

6.2. Layer visulization

To obtain insights of how the proposed RWAAR works, a sample was selected from a three-hour look-ahead testing dataset, 10:00 am on 03/05/2015. First, four convection-related weather elements—180mb above convective inhabitation (J/kg), 1000 m above storm relative helicity (m^2/s^2), 6000 m u component of storm motions (m/s), and input_6000m v component of storm motions(m/s)—were extracted from the sample data. Fig. 12 shows the normalized weather forecasts in the 200 nautical miles surrounding New York metroplex. From the weather forecast, it can be seen that at 7:00 am on 03/05/2015, 10:00 am is forecasted to have convective storm motion from the south side of the New York metroplex. From airport operation data, LGA was closed on 03/05/2015 for four hours, from 11:00 am to 3:00 pm, which is explained by the storm activity forecasted for 10:00 am. Fig. 13 shows the feature maps obtained from selected layers—1st convolutional layer, 3rd convolutional layer, and 5th convolutional layer. Each small graph in one feature map shows the output after kernel convolution; 32 kernels were applied in the model. Although the feature map cannot be interpreted

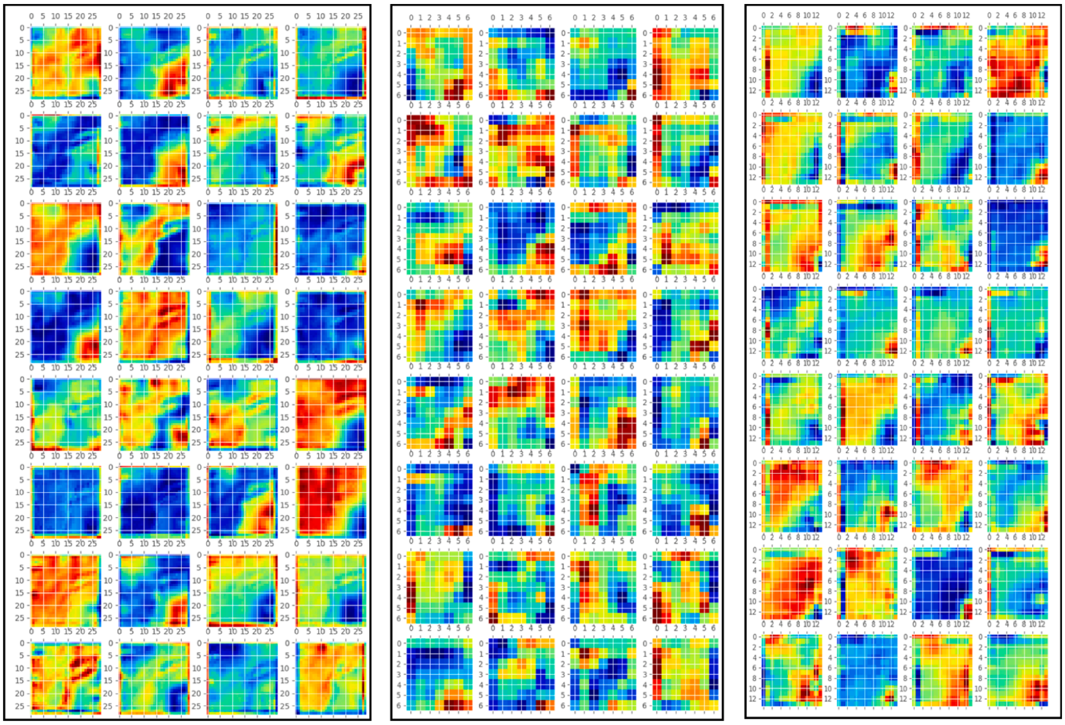


(1) Input_180mb above convective inhabitation (J/kg)
(2) Input_1000m above Storm relative helicity (m^2/s^2)



(1) Input_6000m u component of storm motions (m/s)
(2) Input_6000m v component of storm motions(m/s)

Fig. 12. Convection-related weather forecast elements visualization for selected sample (2015/3/5, 10:00 am, 3-h look-ahead).



(1) Conv2D_1 feature map 29*29 (2) Conv2D_3 feature map 14*14
(3) Conv2D_5 feature map, 7*7

Fig. 13. Feature maps for selected sample (2015/3/5, 10:00 am, 3-hour look-ahead).

intuitively in this problem, because the convoluted features have no mathematical meaning, it still can be observed that the CNN model is trying to generalize the weather patterns from input tensors into representative features.

7. Conclusion

Airport Acceptance Rate (AAR) is an important input for air traffic flow management. Inaccurate estimation of AAR can lead to either the waste of scarce airfield resources or the imposition of avoidable delays on airfield and airspace operation. ICAO provides a method of determining AARs, mainly adapted from the method applied by FAA. Air traffic control personnel usually follow the guidance to determine optimal AARs and then make adjustments based on their experience and operational condition, including weather forecast information at hand. There are no existing tools to assist them to translate weather forecast data into real-time airport capacity automatically. Furthermore, runway configurations and AARs of airports in a multi-airport system are determined by different air traffic control personnel. The lack of synchronization may lead to the loss of efficiency of the sharing airspace and airfield capacity. Thus, a decision-making tool that can better utilize weather forecast information and translate it into real-time runway configuration and airport acceptance rate is an urgent need.

This paper presented a multi-CNN modeling framework called RWAAR net to predict runway configurations and AARs simultaneously for airports in a multi-airport system (metroplex). The proposed modeling framework uniquely accounts for operational interdependency existing in multi-airport systems in the training process. The model takes in selected elements from high-precision ensembled gridded weather forecast data—Rap Refresh (RAP)—of the metroplex as the inputs. A total of 64 elements within 17 horizontal layers were carefully selected based on statistical analysis and feature importance analysis studies from previous chapters. An experimental study was applied to the New York metroplex with a defined terminal area of 200 nautical miles surrounding the metroplex. It was observed from the experiment results that the average root mean square error throughout the forecast horizon from 1 to 8 h for JFK was 5.85, 2.91 for LGA, and 2.29 for EWR. As for the predictive performance of runway configuration, the accuracy achieved was 79.21% for JFK, 85.86% for LGA, and 87.25% for EWR, respectively. The RWAAR net outperformed the model proposed in (Murça, 2018) for AAR prediction for the New York metroplex. The RWAAR net also outperformed the studies from (Avery and Balakrishnan, 2016) (Houston and Murphy, 2012) in regard to runway configurations for JFK, EWR, and LGA.

This study also has limitations, operational features, such as scheduled arrival and departure demand, have not been included in the model. In future work of this study, how to incorporate non-spatial operational data into the proposed CNN model architecture to achieve better model performance will be further explored.

CRediT authorship contribution statement

Yuan Wang: Methodology, Software, Validation, Formal analysis, Data curation, Visualization, Writing - original draft. **Yu Zhang:** Project administration, Funding acquisition, Conceptualization, Supervision, Methodology, Investigation, Resources, Writing - review & editing.

Acknowledgement

The authors gratefully acknowledge the support provided by FAA System Operations Service through National Center of Excellence for Aviation Operations Research (NEXTOR II). The views, opinions, findings, and conclusions reflected in this paper are the responsibility of the authors only and do not represent the official policy or position of FAA, or any other entity. The authors would also like to thank Judy Ghirardelli and Michael Allard from Office of Science and Technology Integration of National Weather Service for assisting weather data acquisition.

References

Bureau of Transportation Statistics (BTS), 2018. Airline on-time statistics and delay causes, https://www.transtats.bts.gov/OT_Delay/ot_delaycause1.asp?type=5&pn=1.
Fisher, L. (Ed.), 2012. Evaluating Airfield Capacity. Transportation Research Board, Washington, D.C.
Liu, P.-C., Hansen, M., Mukherjee, A., 2008. Scenario-based air traffic flow management: From theory to practice. Transport. Res. Part B: Methodolog. 42 (7), 685–702.
Buxi, G., Hansen, M., 2011. Generating probabilistic capacity profiles from weather forecast: A design-of-experiment approach. Proc. of USA/Europe Air Traffic Management Research & Development Seminar.
Provan, Christopher A., Lara Cook, Jon Cunningham,2011. A probabilistic airport capacity model for improved ground delay program planning. Digital Avionics Systems Conference (DASC), 2011 IEEE/AIAA, 30th IEEE, 2011.
Cox, J., Kochenderfer, M.J., 2016. Probabilistic airport acceptance rate prediction. AIAA Modeling and Simulation Technologies Conference.
Wang, Yao, 2011. Prediction of weather impacted airport capacity using ensemble learning. Digital Avionics Systems Conference (DASC), 2011 IEEE/AIAA, 30th IEEE.
Wang, Yao, 2012. Prediction of weather impacted airport capacity using RUC-2 forecast. Digital Avionics Systems Conference (DASC), 2012 IEEE/AIAA 31st IEEE, 2012.
Chung, S., Murphy, D., 2010. Developing a model to determine called rates at airports. 10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference.
DeLaura, Richard A., et al., 2014. Initial assessment of wind forecasts for Airport Acceptance Rate (AAR) and Ground Delay Program (GDP) planning. Massachusetts Inst. of Technology Lincoln Lab, Project Rept. ATC-414, Lexington, MA, 2014.
Murça, Mayara Condé Rocha, John Hansman, R., 2018. Predicting and planning airport acceptance rates in metroplex systems for improved traffic flow management decision support. Transport. Res. Part C: Emerg. Technolog., 97, 301–323.
Murça, Mayara Condé Rocha, John Hansman, Robert, 2018. Identification, characterization, and prediction of traffic flow patterns in multi-airport systems. IEEE Trans. Intell. Transport. Syst., 20.5, 1683–1696.
Li, Leihong, Clarke, John-Paul, 2010. A stochastic model of runway configuration planning. AIAA Guidance, Navigation and Control Conference.
Bertsimas, Dimitris, Frankovich, Michael, Odoni, Amedeo, 2011. Optimal selection of airport runway configurations. Oper. Res. 59 (6), 1407–1419.
Weld, C., Duarte, M., Kincaid, R., 2010:. A runway configuration management model with marginally decreasing transition capacities. Adv. Oper. Res., 436765
Lohr, Gary W., Stephen Atkins, 2015. System-oriented runway management concept of operations.
Jacquillat, Alexandre, Odoni, Amedeo R., Webster, Mort D., 2016. Dynamic control of runway configurations and of arrival and departure service rates at JFK airport under stochastic queue conditions. Transport. Sci. 51 (1), 155–176.
Ramanujam, Varun, Balakrishnan, Hamsa, 2015. Data-driven modeling of the airport configuration selection process. IEEE Trans. Hum.-Mach. Syst. 45 (4), 490–499.
Avery, Jacob, Balakrishnan, Hamsa, 2016. Data-driven modeling and prediction of the process for selecting runway configurations. Transp. Res. Rec. 2600 (1), 1–11.
Houston, Stephanie, Daniel Murphy, 2012. Predicting runway configurations at airports. No. 12-3682.
Dhal, Rahul, et al., 2014. An operations-structured model for strategic prediction of airport arrival rate and departure rate futures. 14th AIAA Aviation Technology, Integration, and Operations Conference.
Tien, S., et al., 2015. Evaluation of an airport capacity prediction model for strategizing air traffic management. American Meteorological Society 95th Annual Meeting.
Tien, Shin-Lai, et al., 2018. Using ensemble weather forecasts for predicting airport arrival capacity. J. Air Transport. 26 (3), 123–132.
Ren, Liling, et al., 2009. Contrast and comparison of metroplex operations: An air traffic management study of Atlanta, Los Angeles, New York, and Miami. 9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO) and Aircraft Noise and Emissions Reduction Symposium (ANERS).
Bureau of Transportation Statistics, 2018. Airport demand statistics. Available at https://www.transtats.bts.gov/Data_Elements.aspx?Data=1 (accessed August 2019).
FAA Order 8400.9 - National Safety and Operational Criteria for Runway Use Programs Document Information. (Accessed 12 August 2020).