AIAA AVIATION Forum
June 25-29, 2018, Atlanta, Georgia
2018 Aviation Technology, Integration, and Operations Conference

10.2514/6.2018-3673

Modeling Key Predictors of Airport Runway Configurations Using Learning Algorithms

Alphan Altinok,* Ravi Kiran,* and Brian Bue*
NASA Jet Propulsion Laboratory – California Institute of Technology, Pasadena, CA 91109

and

Karl D. Bilimoria†
NASA Ames Research Center, Moffett Field, CA 94035

Advanced traffic flow management automation will need accurate predictions of airport runway configurations. Terminal area weather and traffic demand are generally considered to be the most significant factors in predicting runway configuration. Weather information is forecasted across multiple features, including wind direction, wind speed, gusts, cloud ceilings, visibility, temperature, and precipitation, among many others. We use machine learning techniques on historical weather and runway data to determine weather features that correlate well with runway configurations. We analyze the predictive capability of weather features using different learning models trained on data from four major U.S. airports: Atlanta (ATL), Washington – Dulles (IAD), New York – Kennedy (JFK), and San Francisco (SFO). Wind direction alone is strongly correlated with runway configurations above all other examined factors, as expected. This correlation is the most significant component of the ~80% prediction accuracy in selecting between the two most frequently used runway configurations. However, individual airports show variations on how well the runway configuration decisions correlate with wind direction. While wind direction was identified as the most significant indicator of configuration decisions in ATL, IAD, and JFK, it did not emerge as such at SFO. Traffic demand was not found to be a strong factor in predicting runway configurations at any of the airports analyzed. In rare instances, when high demand cannot be accommodated within the current configuration, temporary changes are likely to be attributable to demand. However, these occurrences are so limited in number that their overall effect is not sufficient to consider traffic demand as a major indicator of runway configuration at the airports analyzed.

I. Introduction

Major airports have multiple runways that can be used in many possible configurations. Airport traffic managers configure the airport by selecting an appropriate combination of runways for landing and takeoff operations. The operational capacity (number of landings/takeoffs per hour) of an individual runway can vary from its nominal value, depending on the usage scenario such as configuration of other runways in the airport, aircraft weight class distribution, flight rules (e.g., visual or instrument) in effect, and available ground instrumentation such as landing aids[1]. The total arrival/departure capacity of an airport is a function of the airport configuration and the operational capacities of individual runways. The nature of arrival/departure operations is typically unique to specific airports because of local terrain features, meteorological conditions, infrastructure (number and orientation of runways, landing aids), noise abatement procedures, and local arrival/departure flows. Thus, factors influencing runway configuration decisions at one airport may or may not be relevant at other airports, or may influence different outcomes according to local characteristics.

* Machine Learning and Instrument Autonomy, MS:158-242, 4800 Oak Grove Drive. *altinok@jpl.nasa.gov*
† Aerospace Research Engineer, Flight Trajectory Dynamics and Controls Branch, Mail Stop 210-10. Fellow, AIAA. *karl.bilimoria@nasa.gov*

Changing an airport's configuration disrupts its arrival, departure, and taxiing operations, and adversely affects its capacity in the short term while the change is being implemented; hence configuration changes are made sparingly. In the longer term, the new configuration may affect airport capacity in different ways, depending on its characteristics. For example, a switch from north flow to south flow due to a reversal in wind direction may leave airport capacity essentially unchanged after the configuration switch has been fully implemented. As another example, consider an airport using two closely-spaced runways for all its operations. If low ceiling/visibility limits the airport to single-runway operations for several hours, its capacity would essentially be cut in half until independent two-runway operations can be resumed when ceiling/visibility improves sufficiently.

Traffic management initiatives (TMIs) are used to regulate the flow of traffic in order to resolve situations where demand for a limited resource (e.g., runways) exceeds the available supply. An example of a TMI is a ground delay program (GDP), which assigns departure delays to flights bound for a destination airport where arrival demand is predicted to exceed capacity. Designing a GDP requires an estimate of the airport arrival capacity profile several hours into the future. In current operations, airport traffic managers use their experience and judgment to manually generate these estimates which are then used by traffic flow managers at the Air Traffic Control System Command Center to manually design GDPs. In the far-term future, TMIs may be designed by autonomous software agents that seek to provide an integrated solution that addresses multiple interacting demand/capacity imbalances (in airspace and/or runway resources) across the National Airspace System (NAS). These TMI solutions may also attempt to provide a degree of robustness to uncertainties in NAS data by considering various distributions of demand/supply profiles. One of the many inputs to such a future system is an accurate prediction of airport arrival/departure capacity profiles.

In this work, we focus on estimating runway configurations, from which airport arrival/departure capacity profiles can be calculated. In particular, we seek to identify *specific weather features* that correlate well with historically observed use of different runway configurations, using machine learning algorithms. Examples of weather features include wind speed, wind direction, cloud conditions, ceilings, and visibility. We also investigate the roles of *traffic demand* and *inertia* (tendency to stay in the same configuration) as potential indicators of runway configurations.

This paper is organized as follows. Section II provides an overview of previous work, and Section III covers some operational aspects of runway configurations. The data sources used in this study are summarized in Section IV, and major results of the study are described in Sections V and VI. The paper ends with conclusions presented in Section VII.

II. Previous Work

Early work on estimating airport operating capacity used operations research techniques to model runway capacities[2,3,4]. With known or estimated runway capacities, optimal runway configuration planning was studied using stochastic models[5,6].

Integrated Airport Capacity Model (IACM)[7] was introduced as a theoretical model to generate distributions of airport acceptance rates (AAR) based on terminal aerodrome forecast (TAF). IACM integrates results of several component models into an hourly forecast probability distribution for airport rates of both arrivals and departures. IACM considers runway capacity as well as the capacity of surrounding airspace. The model factors in several weather features including ceiling and visibility, wind, precipitation, and echo tops. The results of IACM predictions are compared to the FAA benchmark capacity values, as published in the 2004 Airport Capacity Benchmark Report[8]. For the airports ATL and Dallas / Fort Worth (DFW), the model provides capacity estimates with 95% confidence to be within FAA benchmark values, in the presence of continuous traffic demand. However, results for Chicago – O'Hare (ORD) showed that IACM significantly overestimated capacities because it does not adequately model the complexity of operations at this airport. This work provides evidence of the significantly different operational priorities of specific airports.

An extension[9] of IACM uses ensemble weather forecasts to characterize uncertainty in weather forecasts. Terminal weather types considered in this study were ceilings and visibility, and surface winds. Results were aligned with the FAA benchmark values[8], with estimated values about 5% higher than the benchmark, but exceeded the actual throughputs by 20-30%. In a follow up study[10], all weather forecast models were compared for their importance in predicting airport capacity. Authors state that maximum capacity predictions agreed reasonably well with actual traffic counts at ATL, especially when demand was near airport capacity. For Instrument Meteorological Conditions (IMC) and Marginal Visual Meteorological Conditions (MVMC), capacity predictions were found to be better than the case of Visual Meteorological Conditions (VMC). Predictions for VMC were typically over-estimates of airport capacity. None of the considered weather forecasts were found to generate better prediction results over the others.

An Empirical Airport Configuration Prediction Model[11] introduced separate deterministic and probabilistic prediction models for airport runway configurations. Seventy-seven airports were studied with runway configuration data from the Aviation System Performance Metrics (ASPM) database between 2009 and 2010. A detailed mapping section describes how runway usage was associated with runway configurations. Weather features include wind speed, wind direction, ceilings, and visibility, where wind direction was found to have the greatest influence on the airport configuration. Per-airport prediction accuracy ranges from 54% to 100% on a categorical match. Most frequent predictions were in the 75-80% range. A separate probabilistic model uses the Localized Aviation Model output statistics Program (LAMP) and TAF forecasts for a 2h look-ahead time. The probabilistic model computes conditional posteriors of different airport configurations given the forecasted weather distributions in 2h time. The model accuracy was assessed as comparing rank orders of actuals vs. prediction probabilities. The ranking was based on a runway use metric, which matches different airport configurations to predictions. Rank 1 defines 1 runway difference, rank 2 defines 2 runway difference, and so on. Within rank 3, a rather loose bound on airport configuration, prediction accuracies exceed 81%.

AAR Distribution Prediction Model (ADPM)[12] aims to predict AAR given weather from TAF and METAR, or Meteorological Terminal Aviation Routine Weather Report. The authors model forecast uncertainty using historical METAR observations and TAF forecasts. However, explicit modeling of weather uncertainty was not found to have a significant effect on AAR predictions. Current airport capacity and ceilings were reported to be the most determining features at San Francisco (SFO) and Newark (EWR).

Maximum-likelihood discrete-choice model[13] describes a statistical model to predict runway configuration. Model predictions were compared against baseline statistics as frequency of occurrence of different runway configurations. Considered factors influencing runway configuration decisions were *inertia*, as tendency to stay in the current configuration, headwind speeds, arrival/departure demand, noise abatement procedures, configuration switch proximity, and inter-airport coordination. New York – LaGuardia (LGA) and Newark (EWR) had 10 and 20 distinct configurations, respectively. Inertia and headwind speeds were found to be prominent features for both airports. Considered operational scenarios were restricted to VFR, time periods with no noise abatement, and simultaneous matching JFK configurations. Under these restrictions, prediction accuracy for both airports were evaluated statistically against the baseline. Configuration switches, a.k.a. airport flips, were categorized into six different types. Correct predictions were evaluated within and outside of temporal vicinity of switches. Outside of temporal vicinity of switches, the prediction accuracy ranges between 16% and 95%, and within the switch period the reported accuracy ranges between 21% and 76%. For the *most frequently observed* configuration at LGA, the authors report a 95% accuracy.

In a follow up study[14], wind direction and speed were explicitly categorized. Results for 15m and 3h lead times were given as averages of correct predictions of multiple runway utilizations. For SFO, this average figure is 85.1%, with a set of predictions ranging between 58.0% and 93.6% for a total of 4 runway configurations. For LGA, the result averages to 82.2% with a range of values between 67.0% and 89.0% for a total of 7 runway configurations. Both results assume perfect knowledge of weather and traffic demand.

Weather Translation Model for GDP planning (WTMG)[15] translates weather forecasts into probabilistic arrival capacity predictions over up to 12h time horizon. The model uses historical forecast and capacity data to build a regression tree. Static WTMG works with independent samples of each hourly AAR, while the dynamic WTMG conditions the current AAR on the value of the AAR from the previous hour. Overall, the model predictions were declared to reduce the Root Mean Square error by approximately 50% over the naive model.

Ensemble methods were compared using support vector machines[16] (SVM) to predict AAR during impacted weather. Bagging decision trees (BDT) were found to outperform SVM classifiers. Based on the FAA benchmark[8], a threshold was introduced to group AAR values into *normal* and *reduced* categories. AAR prediction targets for 2, 4 and 6 hour lead times were considered, each as a binary target with values of *normal* or *reduced*. Most frequently used runways at EWR were evaluated using SVM (67-80%) and BDT (76-85%) predictions. Consistently, one runway vs. all other runways is less predictable than selecting between two specific runway configurations. Prediction results on EWR and SFO were found to be less accurate than ORD and ATL. A follow up study[17] uses Rapid Update Cycle (RUC)-2 weather forecasts to compare linear regression models and bagging decision trees in AAR prediction. In order to evaluate the impact of adverse weather on airport capacity models for EWR, weather data was sampled from only bad weather and all weather conditions separately. For EWR, 95% of the relative errors on 15m AAR predictions were within 10%.

III. Operational aspects of airport runway configurations

Runway configurations of an airport are represented as a list of individual arrival and departure runways separated by a marker. For example, in Figure 1, configuration [26R|27R|28, 26L|27L] designates arrival runways as [26R|27R|28] and departure runways as [26L|27L]. Individual runways are named by numbers and optionally followed by a letter. The number corresponds to the angle between the runway direction and the magnetic north (taken as 0˚), rounded to ten degrees. Parallel runways are identified by the optional position indicator from the point of view of approaching pilots. L, R, and C indicators refer to left, right, and center runways, respectively. For example, a designator 27L identifies the left runway oriented at 270˚ with respect to magnetic north. Note that the same physical runway has two distinct names depending on the approach; for example, 27L and 9R refer to the same physical runway with opposite operating directions.

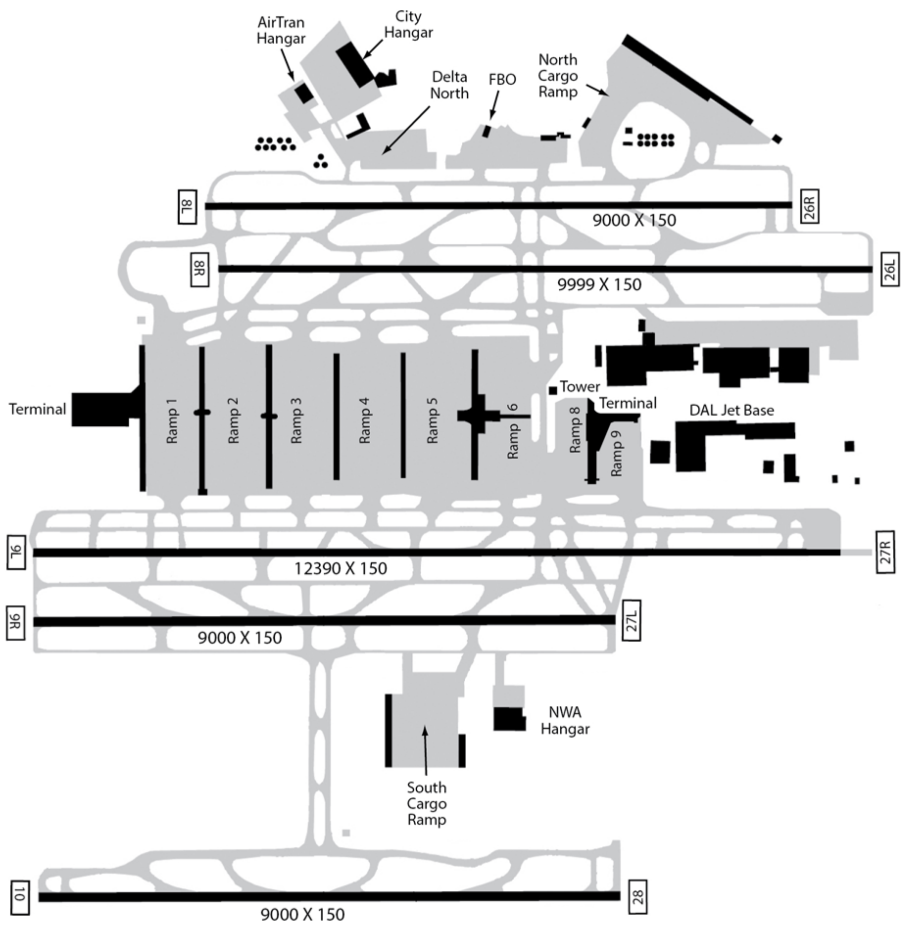


Figure 1. Runway chart of Atlanta Hartsfield–Jackson International Airport

A. Relationship between runway configuration and airport operating capacity

Maximum capacities of individual runways are calculated based on the physical capacity of the runway and subsequently adjusted according to prevailing meteorological conditions and spacing intervals required between successive aircraft[18]. Depending on the runway geometry and usage, e.g., shared arrivals and departures, taxiway layouts, local airspace configuration, and other limitations, the maximum capacity may be further reduced[1]. Capacity estimates are tabulated and published by the FAA for guidance for each major airport, e.g., ATL[19]. The complex nature of this estimation process injects a considerable amount of airport-specific, time-sensitive, and subjective information into the actual runway capacities.

Airport capacity is defined as the hourly throughput a runway system is able to sustain during periods of high demand. Airport capacity is typically expressed as a Pareto frontier also termed as the *capacity envelope*, Figure 2, with tradeoffs between arrival and departure capacities. A separate chart is calculated per meteorological condition, per runway configuration for each airport, e.g., ATL[20]. Complexities involved in airport capacity estimation led to the development of different model-based approaches[21] to aid in planning airport runway operations.

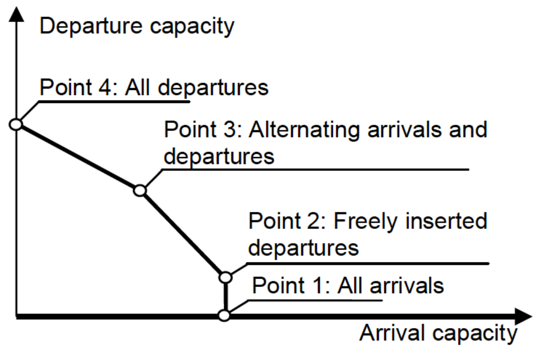


Figure 2. General form of airport capacity chart for runway configurations

Statistically, predicting runway configurations is a harder problem than that of predicting the airport capacity. Consider an airport operating with only two runway configurations at the same capacity. A predictor would be 100% accurate by always selecting this capacity figure, while the same strategy would not work for predicting the runway configurations. Different runway configurations can lead to the same operating capacity range, as is often the case due to the bi-directional naming of the same physical runway. Table 1 shows arrival capacity guidelines for different runway configurations at ATL. Different configurations of runways 26, 27, and 28 for arrival and departure use, shown in the first three rows of Table 1, are rated at the same capacity figures for each of the meteorological conditions. Similarly, in the opposite directions, runways 08, 09, and 10 are all rated at the same capacity figures, shown in rows 4 through 6.

Conversely, similar configurations can have different capacity estimates due to specific use of a runway. The runway configuration in row 7 of Table 1 is similar to row 3, with the only difference of runway 28 being used for arrivals vs. departures; however, the capacity estimates for these configurations differ by about 25%.

The same physical runway can have different capacity estimates depending on geographical, local, or other considerations. For example, ATL has two primary directional traffic flows[20]. With departure fanning conducted on the same runway to reduce the minimum time required between successive departures, one direction enables higher departure throughput than the opposite, leading to considerably different capacity estimates, as is the case for arrival configurations shown in the last two rows of Table 1. The complexities and uncertainties embedded in mapping runway configurations to operational capacity envelopes are the reason to examine the runway configurations in place of operational capacities of airports.

Major airports have select configurations analyzed and tabulated for airport capacity estimation purposes. Runway capacity models, Airfield Capacity Model (ACM) and Runway Simulator (rS), as validated[21], translate dominant runway configurations to capacity estimate envelopes in an effort to optimize airport operations and planning. Validation is reported to be limited to select airports and dominant runway configurations for those airports, with many of the actual configurations not verified with empirical data. In practice, several configurations are used for traffic management activities. Figure 3 shows individual runway configuration use as a percentage, between 2012 and 2015, for ATL, IAD, JFK, and SFO. Actual number of assigned configurations are shown in the top right corner for each airport. Notably, ATL uses runway configurations [26R,27L,28 | 26L,27R] and [8L,9R,10 | 8R,9L] in 91% of total hours of operation between 2012 and 2015, inclusive, while SFO operates with [28L,28R | 1L,1R] in 68% of total hours. This highly skewed use of specific configurations is not the case with IAD and JFK. For ATL, the top two dominant configurations, together accounting for 91% of usage, have identical capacity figures of 116 arrivals per hour under VMC. Thus, a trivial *capacity* predictor that always outputs 116 could achieve 91% accuracy. However, a trivial *configuration* predictor that always outputs the single most dominant runway configuration could only achieve about 55% accuracy. A similar configuration predictor could achieve 68% accuracy for SFO, which ranks as one of the most delayed airports among the 30 major airports in the U.S.[22]

	Arrival	Departure	VMC	LOW VMC	IMC	LOW IMC
RC1:	26R|27R|28	26L|27L	122	116	104	96
RC2:	26L|27R|28	26R|27L	122	116	104	96
RC3:	26L|27L|28	26R|27R	122	116	104	96
RC4:	08L|09L|10	08R|09R	122	116	104	96
RC5:	08R|09R|10	08L|09L	122	116	104	96
RC6:	08R|09L|10	08L|09R	122	116	104	96
RC7:	26L|27L	26R|27R|28	96	86	76	72
RC8:	27L		25	22	20	18
RC9:	9R		50	45	40	36

Table 1. Selected runway configurations and associated capacities for ATL

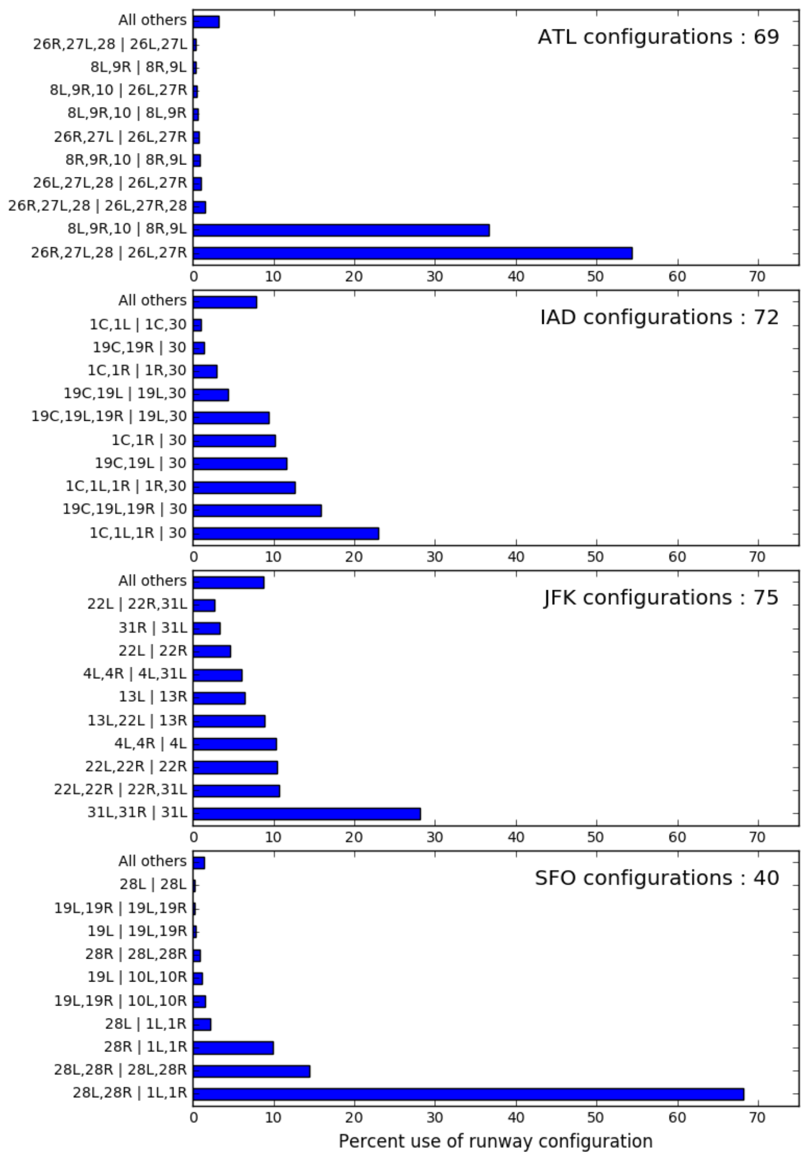


Figure 3. Runway configuration frequencies

B. Factors influencing runway configuration selections

A multitude of factors, relating to the physics of aviation, geographical features, airport layout, local economics, technological infrastructure, and incidental circumstances can influence the choice of runway configurations for efficient airport operations. Many of these factors vary with the airport, and relative priorities can change considerably based on the specific airport. In general, major airports achieve more efficiency by reserving separate runways for takeoffs and for landings as opposed to shared use.

Wind direction and speed are considered as the major factors in runway selections. Because aircraft can takeoff and land more efficiently into the wind, using runways oriented opposite to the prevailing wind direction is desired. To maintain safe operations, aircraft must observe limits on headwinds, tailwinds, and crosswinds. Hence runway configuration is influenced by both wind speed and direction, although sensitivity to wind direction is small for low wind speeds.

Ceiling and visibility are considered to be other major weather features in takeoff and landing operations. Most major airports have precision instrumentation for landing under low ceilings/visibility; however not all runways at a given airport have these landing aids, and unequipped runways would need to be removed from the airport configuration. At San Diego airport (SAN), terrain properties preclude instrumentation for precision landing to the west. Therefore, in bad visibility conditions, landings at SAN are aligned east rather than west. This is an example where terrain specifics could dictate airport operations unlike other major airports.

Demand patterns can vary significantly based on the geographic location of airports and which commercial routes are being served. In large metropolitan areas, such as the Washington D.C. – New York corridor or Southern California, multiple neighboring airports result in complex interacting arrival/departure flows that affect the choice of runway configurations. A runway configuration change at one airport could necessitate a runway configuration change at a neighboring airport.

Simultaneous takeoff and landing operations for busy airports with non-intersecting runways, e.g., Denver airport (DEN) or ATL, are easier to accommodate on separate runways, while smaller but busy airports, e.g., LGA and SFO, may have to conduct takeoffs and landings on intersecting runways. Thus, configuration changes in DEN would likely follow different causal patterns than those of SFO or LGA.

Periodic changes to runway configurations aim to accommodate external factors, such as noise issues with the surrounding communities. A typical example is the Los Angeles International airport (LAX) where aircraft normally land to the west because of the prevailing winds. But at night time, wind conditions permitting, aircraft will approach from the west over the Pacific and land to the east to avoid excessive noise impact on local residents.

Many other factors can influence configuration changes for an airport. Airport structure (the number, length, and orientation of runways), traffic patterns with a mixture of short and long distance flights, as well as seasonally changing aircraft weight mix will directly and indirectly influence the frequency of runway configuration changes. As airport runways are very expensive to design, build, and maintain[23], to even out wear and tear on the runways, controllers will attempt to distribute the usage across different runways by periodic configuration changes. Continuous impact of multiple large aircraft at speeds exceeding 120 mph will wear out the touchdown zones of major runways. Thus, there is an intrinsic incentive for balanced use of runways. Finally, occasional one-off type configuration changes can happen for reasons that are planned (e.g., nearby airspace use) or unplanned (e.g., equipment or emergency issues, even unruly passengers forcing aborted takeoffs).

Configuration changes due to some of these factors are expected *a priori* for specific airports while other factors such as weather and traffic demand are forecasted. Naturally, incidental configuration changes are much less frequent than changes due to other reasons. Runway configuration data contains all of these reasons intermixed, and there is no clear way of assigning rationale to individual configuration selections for model-building purposes. Therefore, a machine learning approach should focus on which of the forecasted features are relatively correlated with runway configurations based on data, before attempting to optimize accuracy levels.

IV. Data sources

Our focus is on how a subset of forecasted factors influences the runway configuration decisions. In particular, we investigate weather features, as well as traffic demand patterns and inertia (the tendency to keep the same configuration), for their predictive capability of runway configurations. To evaluate different features, we train learning methods on permutations of a subset of the features and test the resulting models to predict runway configurations at different hours of operation. We collected weather, traffic demand, and runway configuration data between the beginning of 2012 and the end of 2015, corresponding to about 35,000 hours, as follows.

Weather. The National Oceanic and Atmospheric Administration (NOAA) publishes weather information in METAR (Meteorological Terminal Aviation Routine Weather Report) format for aviation[24]. The weather data is updated in regular 15-minute intervals and as needed in between. For pilot use the METAR notices are encoded in text blocks. An example METAR report and its decoding is presented in Table 2.

KATL 162052Z 29008KT 10SM SCT040TCU 32/23 A3003 RMK AO2	
KATL	Atlanta/Hartsfield-Jackson International Airport, GA, USA
162052Z	Date and time group. Z indicates UTC in 24-hour format.
29008KT	Wind direction and speed: from 290 degrees (WNW) at 08 knots
10SM	Visibility: 10 statute miles
SCT040TCU	Sky condition: scattered clouds (SCT) at 4,000 ft; towering cumulus clouds (TCU)
32/23	Temperature 32°C (89°F), dewpoint 23°C (73°F)
A3003	Altimeter setting: 30.03 in Hg (1017.0 mb) atmospheric pressure
RMK	Following blocks are remarks
AO2	Automated station with precipitation discriminator

Table 2. Example METAR report with decoding of each text block, excluding remarks

The Data Warehouse project[25] at NASA Ames Research Center provides a decoded version of historical METAR data organized in individual columns. Structured METAR records have over 300 potential fields, a significant fraction of which are sparsely populated remark identifiers. Based on previous work in related areas in air traffic management, we selected the following weather features to be potentially significant in predicting airport runway configurations. These features consist of wind direction, wind speed, variability of wind direction, visibility, ceiling height, temperature, dew point, altimeter setting, significant weather descriptors, significant precipitation, and presence of clouds. Occasional missing data points in these indicators were imputed by carrying forward the last known observation in that indicator.

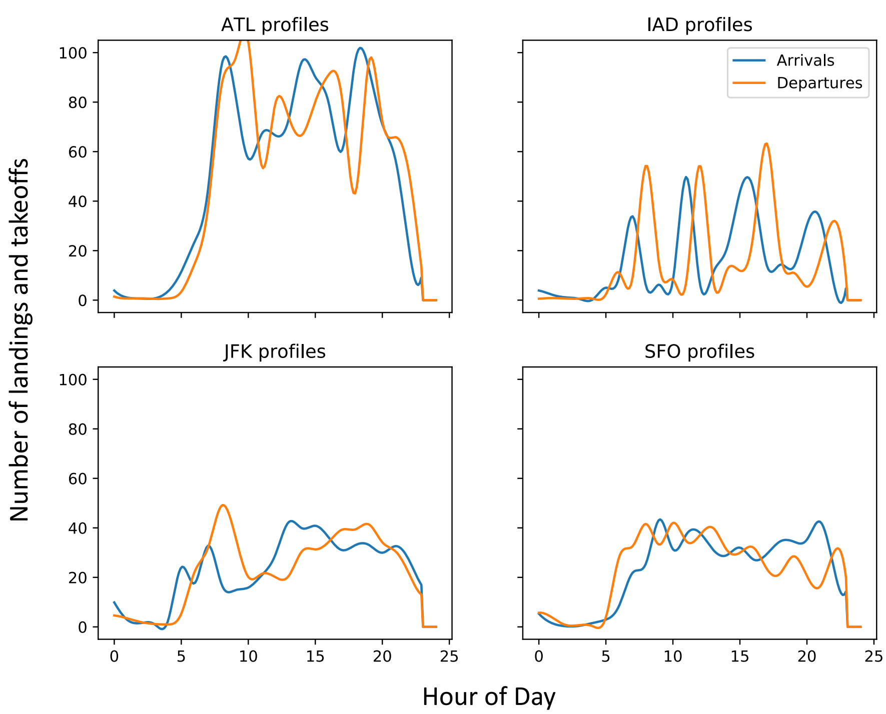


Figure 4. Demand profiles of ATL, IAD, JFK, and SFO as estimated from actual arrivals and departures over 24 hours

Traffic Demand. Actual landing and takeoff counts are available from the Aviation System Performance Metrics[26] (ASPM) database in the FAA Operations and Performance Data. ASPM provides hourly arrival and departure counts per airport. Actual operation counts are indicative of scheduled demand with the exception of flight cancellations. We retrieved traffic data for ATL, IAD, JFK, and SFO between the years 2012 and 2015 from the ASPM database. The traffic demand profile of each airport, shown in Figure 4, is unique due to location, physical capacity, and traffic patterns, among other factors. We experimented with raw arrival and departure statistics, as well as a categorized version of these as *above* or *below* the hourly expected demand for any given hour of the day.

Runway configurations. We retrieved runway configuration data from the Data Warehouse at NASA Ames Research Center for the years 2012 through 2015. Runway configurations are recorded hourly for major airports.

To construct training data sets, METAR and traffic demand information were aligned with the known runway configuration at the beginning of each hour where the last known METAR report was associated with decision making.

Examination of runway configuration data revealed that, even with multiple runways, some airports tend to utilize two or three dominant configurations out of tens of configuration combinations that are physically possible. Between 2012 and 2015 each of the analyzed airports was configured in ~70 distinct ways, with the exception of SFO which used only 40 distinct configurations. The dominant configuration use was more pronounced in ATL and SFO than in IAD or JFK, see Figure 3.

V. Visual assessment of features using dimensionality reduction

As weather information is high-dimensional, exploring this high-dimensional space in lower dimensions can be useful to develop an insight into the relative importance of different weather features as they correlate with runway configuration data. To do so, we use t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm[27] to perform the dimensionality reduction. The t-SNE algorithm maps data points from the high-dimensional space to a lower dimensional space such that the mapping aims to preserve local neighborhoods in the original high-dimensional space. The mapping depends on a single free parameter *perplexity*, which balances the influence of local versus global distances.

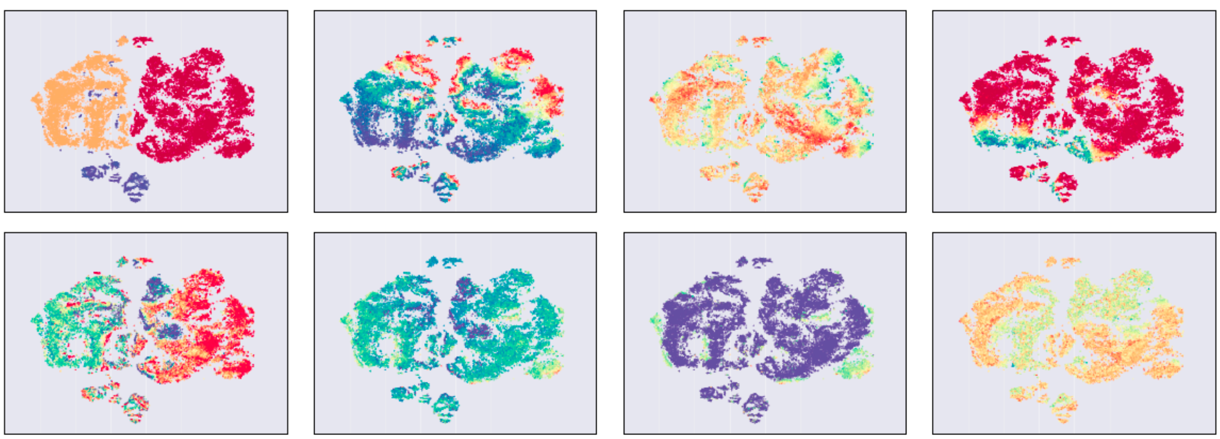


Figure 5. Projections of weather features in 2D for ATL
Top row left to right: runway configuration (as reference), ceiling, temperature, and visibility
Bottom row left to right: wind direction, wind speed, wind gust, and traffic demand

Mapping the distribution of individual features from the original space onto a 2D projection provides a visual assessment of how well the distributions of these features align with each other. Note that t-SNE does not provide the mapping as a function, unlike other dimensionality reduction methods such as the Principal Component Analysis (PCA), and it is non-deterministic. On the other hand, t-SNE projections often result in better visualizations than those of PCA. To capture temporal properties of each weather feature, we compute the deviation/change in the past 3 to 5 hours along with the cumulative deviation/change, followed by normalizing to [0, 1]. Aligning different weather features, traffic demand, and runway configurations along the same hour, we construct high-dimensional vectors, each of which corresponds to an observation point in this high-dimensional space. Using t-SNE, we map this space to a 2D representation, where closer points in the high-dimensional space appear closer in the 2D map. Note that the spatial position of each point represents all inputs in the combined vector, including weather, traffic demand, and runway configuration information. To highlight the distribution of an individual feature on the 2D maps, we use color shading according to the values of that feature. Thus, closer values of individual features correspond to similar colors on a predefined color palette, with the number of total colors used corresponding to the number of distinct values for that feature. The spatial distributions of points do not change with different color maps. This method provides a preliminary visual assessment of how well a feature is correlated with a target feature in the original space by comparing color distributions.

In Figure 5, the top left panel shows the distribution of runway configurations for ATL as the target projection in 2D, where the distribution of the two major configurations are clearly distinguishable as large clusters of two different colors. The third cluster at the bottom of the panel represents infrequent configurations. Features that contribute to this separation are expected to exhibit a similar color distribution (but not necessarily in the same individual colors). Comparing the three weather features in the top row, i.e., ceiling, temperature, and visibility, it is observed that they do not show a similar separation of colors for these features. Variations in their color distributions do not indicate significant correlations with the target distribution, i.e., the major runway configurations. The bottom left panel, showing wind direction, is the only feature with significant separation of color values similar to the separation of different runway configurations. The remaining weather features, wind speed and wind gusts, show as almost uniformly distributed, indicating no significant correlation with the distribution of the runway configurations. Finally, traffic demand is shown in the bottom right-most panel, where a non-uniform distribution is observed but a correlation with the runway configuration is not readily evident.

VI. Weather, traffic demand, and configuration inertia as predictors of runway configuration

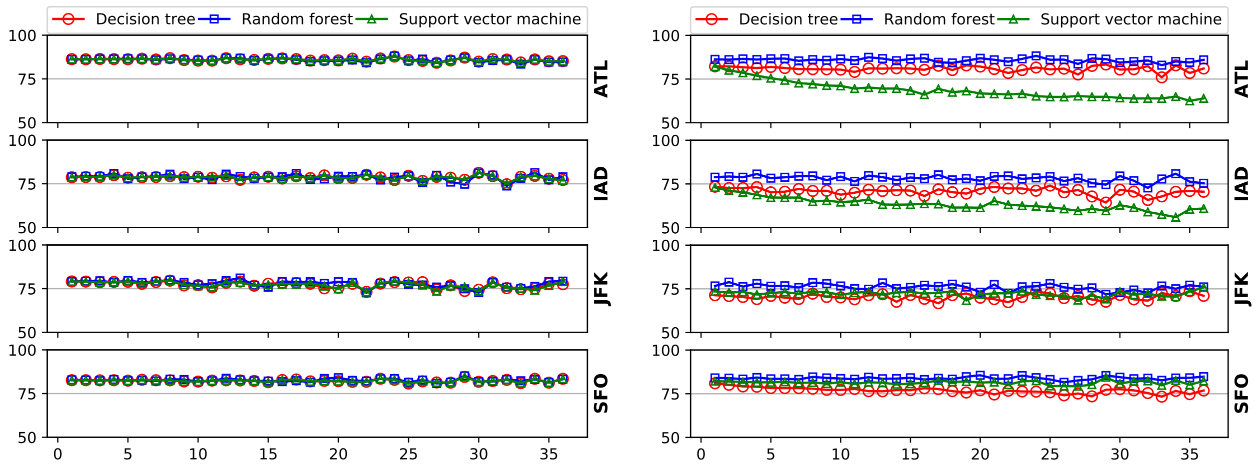
We first aim to answer the question if runway configuration of an airport can be predicted based on information about weather, traffic demand, and configuration inertia. We consider input features as *predictors* if a trained model can output correct predictions of major runway configurations beyond random chance. It is consistent with common practice to eliminate infrequent runway configurations from data for this analysis, on the basis that their cumulative frequency corresponds to less than 6% use over all hours of operation. However, all configuration changes represent some runway balancing activity with an associated cost that could increase the relative importance of such infrequent instances. Our objective is not necessarily to optimize the accuracy of a configuration predictor, but to identify how specific features contribute to configuration selections as gleaned from historical data in a machine learning setting.

A. Weather

To determine the relative performance of weather and other indicators within the context of machine learning algorithms, we experimented with random forests[28] (RF), which are attractive because: i) random forests can efficiently work with multiple and potentially correlated features, ii) an estimate of variable importance can be computed[28,29], and, iii) they are robust to noise. We also inspected decision tree (DT) and support vector machine (SVM) predictions to validate random forest predictions. For each airport, the retrieved data between 2012 and 2015 corresponds to more than 35,000 hours of operation. Routine cleaning procedures, such as eliminating unrecognized or unrecorded runway configurations, did not change data size significantly. For all weather features, missing values were forward-filled with the last known record in time, within the last hour. The total amount of forward-fills were less than 2.4% of total data volume.

To study the effects of temporal correlation in consecutive weather reports we subsample hourly observations at increasing window sizes between 1 and 36 hours. For example, for a window of 20 hours, we sample observations from a subset of all data that are 20 hours apart from each other, without replacement. As the number of sampling hours grows, the number of observations available for training and testing reduces accordingly.

For training datasets, we balance the representation of target configurations by including additional observations such that each runway configuration is represented equally. We cross-validate models over five random sets of 50% training and test data pulled from these subsets. Figure 6 (a) shows predictions using wind direction as the sole predictor of runway configuration, while Figure 6 (b) shows predictions using multiple features consisting of wind direction, wind speed, wind gusts, visibility, cloud ceilings, and temperature. Each data point corresponds to a prediction accuracy, as percentage of correct predictions, reported by the respective models over increasing windows of operation hours. Note that we examined predictions only between the top-2 most commonly used configurations in each airport to study time-correlation and feature relevance. With wind direction as the only feature, predictions by all models are consistent across the four airports. However, considering additional features causes DT and SVM accuracy levels to drop, and in an uneven manner across the airports. Indeed, considering multiple weather features as opposed to the single weather feature (wind direction), we observe a measureable decline in prediction accuracy as the number of sampling hours increases. In other words, the different features of weather input could be temporally correlated. In addition, variations in decline are not uniform across the airports, suggesting that operating characteristics of individual airports use weather features in different ways.



(a) Wind direction as the only feature

(b) Multiple weather features

Figure 6. Predictions by decision tree, random forest, and support vector machines over increasing subsampling windows as the number of hours between samples

	Single feature			Multiple features		
	DT	RF	SVM	DT	RF	SVM
ATL	86.87	86.98	86.87	82.21	87.32	82.07
IAD	79.87	80.94	79.94	73.52	80.73	73.01
JFK	79.60	81.08	79.65	71.37	79.01	73.53
SFO	82.88	83.54	82.89	80.97	84.69	82.18
ATL	80.24	80.53	80.24	72.89	80.53	75.49
IAD	61.37	61.75	61.37	56.86	62.42	60.99
JFK	68.55	68.64	68.55	57.38	68.35	63.18
SFO	68.60	69.50	68.61	65.95	71.67	67.39
ATL	84.29	84.88	84.28	78.54	84.78	79.63
IAD	42.36	43.05	42.50	42.28	54.24	43.32
JFK	54.47	54.70	54.42	49.55	59.26	47.65
SFO	71.77	72.71	71.76	67.15	73.62	70.26

Table 3. Prediction accuracies. Top block shows predictions between the top-2 runway configurations; middle block shows predictions between top-2 and all other runway configurations as a combined third category; bottom block shows predictions between the top-5 runway configurations.

Since the number of dominant configurations and the relative usage of these configurations (see Figure 3) change across airports, binary classification between the top-2 configurations is insufficient to assess expected prediction accuracy levels. Table 3 shows a comparison of predictions for multiple runway configurations as raw accuracy scores; the single-feature case includes only wind direction, while the multi-feature case includes wind direction, wind speed, wind gusts, ceilings, visibility, and temperature. The top block shows results for predictions between the top-2 configurations (aggregating the data in Fig. 6), while the bottom block presents results for predictions between the top-5 configurations. The middle block presents results for predictions between three categories: the top-2 configurations, and all other configurations combined into a single category to examine robustness of predictions for infrequently used runway configurations. All predictions remain well above chances of random picks, 1/2, 1/3, and 1/5, for the top, middle, and bottom blocks, respectively. Out of the evaluated models, random forests achieve slightly better results in almost all cases. It is worth noting that inclusion of additional runway configurations affect IAD and JFK predictions significantly more than IAD and SFO. Relative dominance of additional configurations in IAD and JFK are consistent with these results.

To evaluate how well different weather features predicted runway configurations, we examined all combinations including two or more features. Figure 7 presents a comparison of relative feature importance in influencing runway configuration decisions, selecting between the top-2 most commonly used configurations. In each column, contributions of examined features are normalized to [0, 1]. For example, the left-most column in ATL shows that *wind direction* and *wind speed* were used as features, and *wind direction* was estimated to be responsible for more than 90% of correct runway configuration predictions. Although each column adds up to 100%, there is a wide variation in the total number of correct predictions associated with each column.

The right-most column shows the relative importance of all six weather features used. Again, wind direction emerges as the single most important predictor among all other weather features. In the absence of *wind direction* as a feature, there is no definitive weather feature emerging as important across different airports. In SFO, relative importance of *wind direction* over other weather conditions is not as strong, but it remains as one of the major factors. Across all airports, if present in the combination of evaluated features, *wind direction* is the basis of most decisions in selecting runway configurations. Notably, including other weather conditions in prediction reduces accuracy levels, as seen by comparing corresponding entries in Table 3 under *single feature* and *multiple features* panels.

Different weather features would be expected to improve the prediction accuracy if they contribute mutually exclusive information. Potentially correlated information between features will reduce the improvement. In the case of studied weather features, inclusion of additional weather features not only does not improve prediction accuracy, but in all four airports their combined effects sum up to less accurate predictions. One possible explanation could be that runway configuration decisions are based primarily on wind direction without consideration of other weather features in a significant way. Traffic demand, noise abatement, traffic patterns, and incidental issues are possible factors that could account for incorrect predictions.

B. Demand and inertia

Out of many other possible decision factors, traffic demand and tendency to remain in the same runway configuration, termed *inertia*[14], are thought to be important factors in determining runway configuration. Inspecting available data revealed numerous short-term configuration changes such as the ones shown in Table 4 for ATL. As increased departure and decreased arrival operations were expected at ATL, runway 28 was opened for departures temporarily. Changes to accommodate increases in arrival or departure demand only happen at peak demand hours, but the converse is not true, i.e., during off-peak hours, the airport is not reconfigured for a lower capacity configuration. We model traffic demand in two separate ways. First, regardless of the hour of the day, we look at when expected traffic exceeds the airport capacity for that airport at any given hour of the day. Unfortunately, capacity figures are only published for select runway configurations out of many configurations that were actually used. Therefore, we calculate nominal demand profiles for 24-hour periods retrieved from the Aviation System Performance Metrics[26] (ASPM) database, in the FAA Operations and Performance Data. We use it as a proxy for airport capacity when actual capacity figures are unknown. Figure 4 shows nominal demand profiles as generated by kernel density estimation for departures and arrivals at the four airports. When scheduled demand exceeds the nominal demand, we flag the observation hour incurring excessive demand. A second method relies on identifying only peak hours when demand exceeds known capacities of a subset of runway configurations.

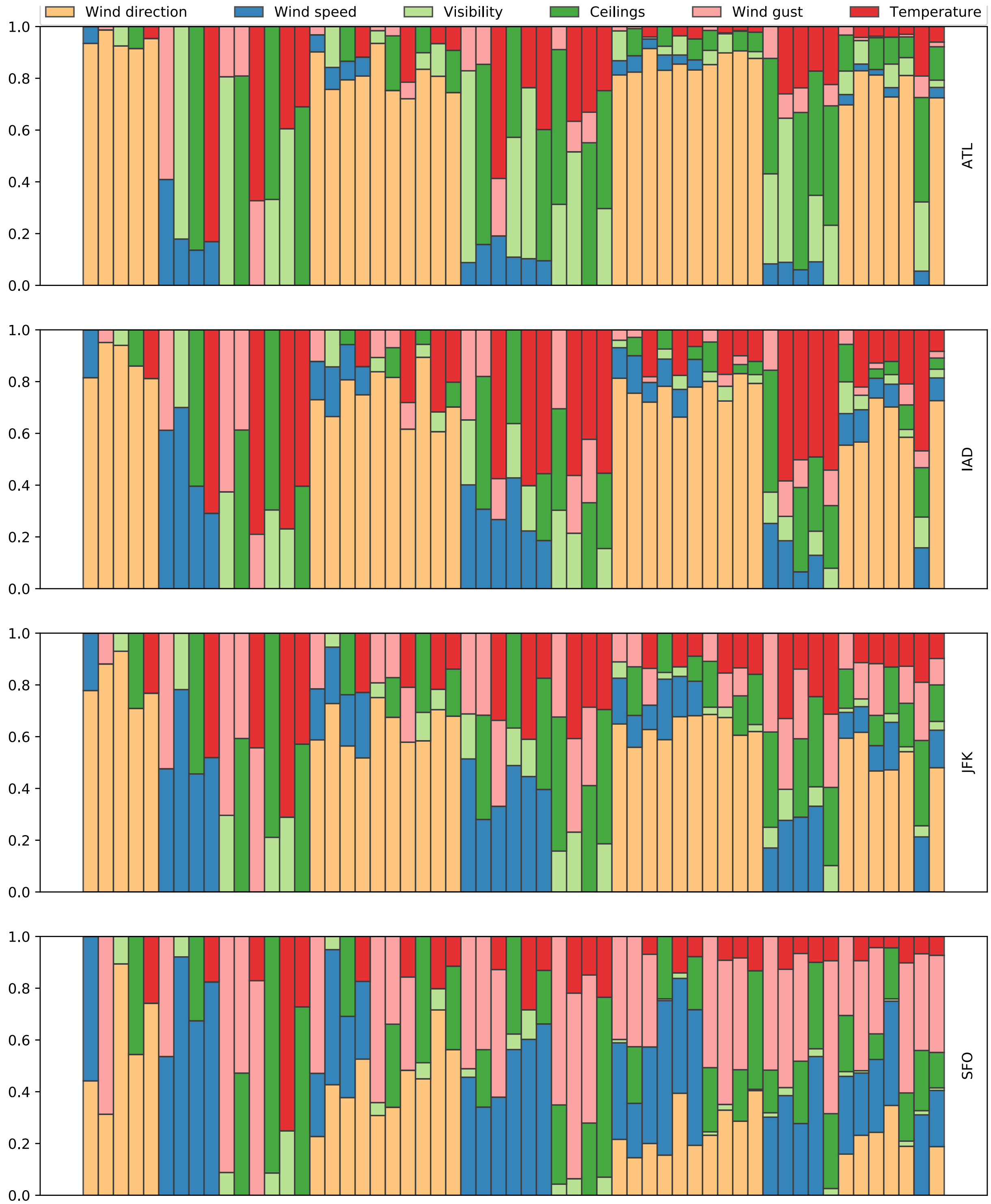


Figure 7. Combinations of top 6 weather features are evaluated to predict runway configurations. Each column represents a distinct combination of features, where the heights of different colored bars correspond to relative importance of weather features considered in that column.

Atlanta/Hartsfield-Jackson International Airport (01/02/2012)		
Runway configuration	Departures	Arrivals
26R,27L,28 | 26L,27R	103	82
26R,27L,28 | 26L,27R,28	113	55
26R,27L,28 | 26L,27R	53	67

Table 4. Runway configuration changes over three consecutive hours and corresponding departure and arrival counts

Tendency to stay in the same configuration should decrease over long periods of runway use. We experimented with linear, logarithmic, and exponential decay models to determine if *inertia* was a factor that can be used in predictions. Note that *inertia* captures time-correlated information by definition. Table 5 shows results with select *demand* and *inertia* features. In any of the experiments, neither demand nor inertia significantly improved upon wind direction as a predictor feature when wind direction was present. While spot checks of data tend to confirm changes due to peak-hour increases in demand, the total number of such changes are insignificant in the total hours of operation. Thus, input to learning algorithms should account for this imbalance by modifying the sampling regime, using priors, or a weighting mechanism to calculate posteriors. However, a difficulty arises in sampling observations where the configuration change is mainly attributed to demand fluctuations. Without confirmed examples of why specific configuration changes were made, there are no principled ways of applying one of the above-mentioned techniques to confirm *traffic demand* or *inertia* as good predictors of runway configurations.

	Wind direction (WD)			WD + Demand			WD + Inertia		
	DT	RF	SVM	DT	RF	SVM	DT	RF	SVM
ATL	86.87	87.02	86.87	83.99	86.98	85.04	84.67	86.54	86.82
IAD	79.87	80.94	79.95	75.61	80.58	76.93	77.50	81.09	80.54
JFK	79.60	80.24	79.65	79.53	83.61	81.34	78.36	79.96	79.65
SFO	82.89	83.76	82.89	81.58	83.76	82.26	82.52	85.32	82.87

Table 5. Wind direction, demand, and inertia as predictors

VII. Conclusions

Previous research documents different statistical methods for predicting airport runway configurations based on weather and traffic data. Reported results vary with specific airports, data sources and selection methods, domain-specific information, as well as with the use of specific metrics. Current prediction accuracy levels for most airports are in the 80–85% range for picking major runway configurations. As our prediction results (for ATL, IAD, JFK, SFO) support the accuracy levels previously reported, it is possible to predict *major* runway configurations well beyond random choice, independent of the hour of operation.

Our focus on the most important predictor also indicates that for at least three out of the four major airports analyzed, wind direction emerges as the sole significant predictor regardless of the underlying prediction model. Prediction accuracies, while consistently above the random choice threshold, vary measurably with the characteristics and the number of major runway configurations of the specific airport under study. The most significant contribution of our study is that while prediction accuracy could be improved through careful data selection or statistical techniques, the use of additional weather features, traffic demand, or configuration inertia will likely not improve predictions within similar problem settings. Based on current prediction levels, wind direction correlates with *major* runway configurations for 82.6% of the operational hours – see Table 3 top-block, averaged over airports and models. This is an expected outcome as the major runways are by design aligned with the prevailing winds in the geographic region of the airport. Use of less frequent configurations appear to be not as strongly correlated with wind direction.

Wind direction does not emerge as the sole predictor in the case of SFO, with the likely cause being a single major configuration leading over all other configurations. Thus, it is harder to resolve individual weather features reliably. Consequently, developing a general airport-agnostic prediction method requires further investigation that can account for relevant characteristics of individual airports. To learn statistical models from data, labeling the runway configuration selections with operational reasons at the time of selection may prove to be invaluable. It is possible that non-weather features not considered in this work may account for the remaining uncertainty in prediction outcomes. Unfortunately, there are no data sources available at this time to investigate this conjecture. Automated data collection procedures at towers could enable future research.

Thinking of runway configuration prediction as a proxy problem to estimating the airport capacity brings about interesting possibilities as the latter is mathematically simpler. However, capacity prediction would not be as informative to address congestive situations because multiple operational constraints can lead to different capacity estimates from the same runway configuration, and multiple configurations can lead to the same capacity figures. An alternative proxy to the capacity prediction problem is to predict the points in time where runway configuration was changed, even if the associated *airport capacity* figure may have stayed the same. Note that not all configuration changes bear the same cost in terms of workload and delays. For example, provisional opening of an auxiliary runway to accommodate a temporary issue is recorded as a runway configuration change, yet it is quite different in nature, purpose, and duration than switching the airport configuration (termed as *airport flip*) to accommodate the prevalent dominant traffic flow. Any time a decision is made to change the current runway configuration, the state of the airport must account for flights approaching the airport, other flights taxiing to runway thresholds for takeoff, and other flights still at their gates. Accordingly, switching the landing and takeoff directions of an airport could require temporary holding of takeoff and landing operations, taxiing the waiting aircraft to the new runway, landing airborne aircraft currently on final approach, and diverting other airborne flights towards new approaches until takeoff traffic clears away. These events can result in gate-holds, ground-delays, or even ground-stops. In certain cases, routing airborne traffic to neighboring airports could become necessary and will require coordination between the tower, approach, and center. This type of configuration change can cause significant delays and has a much higher cost than other types of configuration changes.

Our analysis indicates that most runway configuration decisions are based on prevailing wind directions, but different airports evaluate wind direction in conjunction with other influencing factors, potentially particular to the local characteristics. Further analysis could reveal which hours of operation are more likely to rely on specific features as the most significant determining factor over other features. This will guide the development of more descriptive models that can account for various factors in conjunction, at different hours of operation.

Acknowledgments

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. This work was funded by the Safe Autonomous System Operations Project under NASA's Airspace Operations and Safety Program.

References

[1]Office of Airport Planning and Programming, The Federal Aviation Administration. "Airport Capacity and Delay," *Advisory Circular 150/5060-5*, 1983.

[2]Hockaday, S.L.M. and Kanafani, A.K., "Developments in Airport Capacity Analysis," *Transportation Research*, Vol. 8, No. 3, 1974, pp. 171–180.

[3]Venkatakrishnan, C.S., Barnett A., and Odoni A.R., "Landings at Logan Airport: Describing and Increasing Airport Capacity," *Transportation Science*, Vol. 27, No. 3, 1993, pp. 211–227.

[4]Newell, G. F., "Airport Capacity and Delays," *Transportation Science*, Vol. 13, No. 3, 1979, pp. 201–241.

[5]Bertsimas, D., Frankovich, M., and Odoni, A., "Optimal Selection of Airport Runway Configurations," *Operations Research*, Vol. 59, No. 6, 2011, pp. 1407–1419.

[6]Li, L., and Clarke, J.P., "A Stochastic Model of Runway Configuration Planning," *AIAA Guidance, Navigation, and Control Conference*, 2010.

[7]Kicinger, R., Cross, C., Myers, T., Krozel, J., Mauro, C., and Kierstead, D., "Probabilistic Airport Capacity Prediction Incorporating the Impact of Terminal Weather," *AIAA Guidance, Navigation and Control Conference*, 2011, pp. 8–11.

[8]The Federal Aviation Administration and The MITRE Corporation, "Airport Capacity Benchmark Report," *Technical Report*, U.S. Dept. of Transportation, Federal Aviation Administration, Washington D.C., 2004.

[9]Kicinger, R., Krozel, J., Steiner, M., and Pinto, J., "Airport Capacity Prediction Integrating Ensemble Weather Forecasts," *AIAA Infotech Aerospace Conference*, Garden Grove, CA, USA, 2012.

[10]Kicinger, R., Chen, J., Steiner, M., and Pinto, J., "Probabilistic Airport Capacity Prediction Incorporating Weather Forecast Uncertainty," *AIAA Guidance, Navigation, and Control Conference*, 2014, pp. 1465.

[11]Ramamoorthy, K. and Hunter, G., "An Empirical Airport Configuration Prediction Model," *AIAA Modeling and Simulation Technologies Conference*, Minneapolis, Minnesota, 2012.

[12]Cox, J., and Kochenderfer, M. J., "Probabilistic Airport Acceptance Rate Prediction," *AIAA Modeling and Simulation Technologies Conference*, 2016, pp. 165.

[13]Ramanujam, V., and Balakrishnan, H., "Estimation of Maximum-Likelihood Discrete-choice Models of the Runway Configuration Selection Process," *IEEE American Control Conference,* 2011, pp. 2160–2167.

[14]Avery, J., and Balakrishnan, H., "Predicting Airport Runway Configuration – a Discrete-choice Modeling Approach," *Air Traffic Management Research and Development Seminar*, Lisbon, Portugal, 2015.

[15]Provan, C. A., Cook, L., and Cunningham, J., "A Probabilistic Airport Capacity Model for Improved Ground Delay Program Planning," *IEEE/AIAA Digital Avionics Systems Conference*, 2011, pp. 2B6–1.

[16]Wang, Y., "Prediction of Weather Impacted Airport Capacity Using Ensemble Learning," *IEEE/AIAA Digital Avionics Systems Conference*, 2011, pp. 2D6–1.

[17]Wang, Y., "Prediction of Weather Impacted Airport Capacity using RUC-2 Forecast," *IEEE/AIAA Digital Avionics Systems Conference*, 2012, pp. 3C3–1.

[18]The Federal Aviation Administration, "Order JO 7210.3W", *Facility Operation and Administration*, Feb. 2010.

[19]The Federal Aviation Administration, "Airport Acceptance Rates for Atlanta International Airport," *http://www.fly.faa.gov/Information/east/ztl/atl/atl_aar.htm*, 2016.

[20]The Federal Aviation Administration, "Hartsfield-Jackson Atlanta International Airport Capacity Profile," *https://www.faa.gov/airports/planning_capacity/profiles/*, 2014.

[21]Kim, A., and Hansen, M., "Validation of Runway Capacity Models," *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2177, 2010, pp. 69–77.

[22]Bureau of Transportation Statistics, "Airport Ranking and Summaries by On-time Performance." *https://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/subject_areas/airline_information/airline_ontime_tables/index.html*, 2017.

[23]The Federal Aviation Administration, "Standards for Specifying Construction of Airports," *Advisory Circular 150/5370-10g*, 2014.

[24]The Federal Aviation Administration, "Official Guide to Basic Flight Information and ATC Procedures, Aeronautical Information Manual," *https://www.faa.gov/air_traffic/publications/media/aim.pdf*, 2015.

[25]Eshow M., Lui M., and Ranjan S., "Architecture and Capabilities of a Data Warehouse for ATM Research," *IEEE/AIAA Digital Avionics Systems Conference*, 2014.

[26]The Federal Aviation Administration, "Aviation System Performance Metrics (ASPM)", *https://aspm.faa.gov/apm/sys/main.asp*, 2017.

[27]Maaten, L., and Hinton, G., "Visualizing Data using t-SNE," *Journal of Machine Learning Research*, Vol. 9, 2008, pp. 2579-2605.

[28]Breiman, L., "Random Forests," *Machine Learning*, Vol. 45, No. 1, 2001, pp. 5-32.

[29]Strobl, C., Boulesteix A., Kneib, T., Augustin, T., and Zeileis, A., "Conditional Variable Importance for Random Forests," *BMC Bioinformatics*, Vol. 9, No. 1, 2008, pp. 307.