# Analysis of Relationship Between Aircraft Departure Delays and Regional Weather Patterns

Sydney I. Dolan
*Purdue University, West Lafayette, Indiana, 47906, United States*

**The purpose of this research is to analyze the relationship between flight delays and weather patterns in order to develop a greater understanding of the factors behind flight delays. The factors analyzed include type of weather pattern, airport, carrier, and day. Due to the large quantity of data available, the data set chosen for this study was all flights from the top 35 U.S. airports in the month of January 2017. Data for this study was taken from the Department of Transportation, and the National Oceanic and Atmospheric Administration. The study developed a probability model to calculate the likelihood of probability of delay for the airport based on a given day. Additionally, the study analyzed the rates of delay between airports, to test the hypothesis that increased delays from one airport will result in more delays at other airports. The results of this analysis can be used to help the FAA understand some of the factors that affect flight delays, which can help them accommodate more flights without sacrificing passenger safety.**

## I. Introduction

Flight delays present a huge logistic problem to both airlines, passengers, and airports. According to a 2016 estimation, flight delays result in aircraft companies losing billions of dollars each year. In the U.S., the Federal Aviation Administration (FAA) estimates that the cost of delays in one year is 31 billion[1]. Weather patterns often result in more delays from the airport, as weather patterns can ground airplanes until the storm has passed. One airline consulting firm estimated that severe weather during one week in January 2014 cost the airline industry 1.4 billion[2]. These delays inconvenience both the passengers and create additional work complexity for the airlines and the FAA. Understanding more about how weather patterns can effect flight delays can lead to saved time for the passengers, saved money for the airlines, and a safer system for the FAA.

Rising passenger volume means that air transportation systems need to be more robust to accommodate more flights. Over the next 20 years, carrier passenger growth is expected to increase by 1.9 percent per year [2]. This equates to airline passenger travel nearly doubling in the next 20 years, to over 1.2 billion in 2032[3]. The FAA plans to increase its number of commercial operations and contract towers by 45% to accommodate this growth[4]. While these accommodations will help with air transportation infrastructure, they will not help resolve systemic issues with weather patterns. The current system used by FAA's Air Traffic Operations relies on the judgement of traffic managers, on a case-by-case basis[5]. More information on weather patterns and their effects on flight patterns could help traffic managers make informed decisions about the best way to redirect flights.

With rising passenger volume, understanding the causes of flight delays is important to allow for an infrastructure that can properly handle increased load. At its current state, the FAA needs to learn more about how to properly redirect aircraft along different paths, as well as increase the volume of aircraft within the air transportation system. The predicted growth of the industry requires the FAA to increase size of its current operations, and the number of flights it oversees. To do this safely, the FAA needs more research and modeling into what factors result in flight delays, and how to mitigate delays when these factors occur.

The objective of this work was to provide a high-level analysis on how flight departure delays and weather patterns are related, and what factors can result in higher delays for an airport. Ideally, the work will identify specific weather conditions and variables that result in a higher probability of flight delay. However, flight analysis is extremely complex and involves many changing variables, thus proving difficult to develop conclusive evidence.

Formidable components of the research include the large quantity of data available, as well as conclusively determining which factors create a given result.

For this research, the data set focused on flight data from the top 35 domestic airports in January 2017. This research involves analyzing delays at each airport, and factors like flight carrier, weather patterns, and day of week that could result in higher delays. A probability model is developed to calculate the likelihood of delay based off the known information of delays for the airport. The meaning of these results is analyzed to see if any meaningful information can be drawn from the research presented.

## II. Background

Weather radar is used to help flight traffic operators direct flights, but there is little in the system to incorporate historical data and trends into the decision making. Currently, NEXRAD weather radar is used to redirect flights. This system combines data from multiple sites to create an image of the weather over a map. This method allows for a more comprehensive approach to weather, which allows Air Traffic Controllers (ATCs) to make strategic decisions. There are limitations to this approach as the weather radar system has a 10 minute delay, and the radar on the plane is limited in range[6]. However, one of the areas that can be improved is providing ATCs with additional information so that they can make more informed decisions about how to react to different weather patterns.

Previous work in this field of study has focused on optimizing flight redirection to accommodate more flights within the system. Additionally, other studies have analyzed aircraft delay and on-time performance at specific airports. These studies have either analyzed the relationship between weather and flight traffic between a given airport and its destination, or all outgoing flights at one airport. Therefore, most previous methodology involves reducing the scope of data to provide more accurate results to an individual component. Then, the smaller scale analysis can be combined to provide context to the larger picture. This work is different than previous work as it is providing useful information about the full system rather than one aspect. By analyzing the top-level trends of the system, a greater understanding of individual airport delays can be developed.

The data in this analysis was taken from the National Oceanic and Atmospheric Administration (NOAA), and the Federal Aviation Administration (FAA). The FAA collects data regarding the performance of flights, numbers about the arrival time, delay time, cause of delay, and minutes delayed. These numbers are available publicly and can be downloaded from the Aviation, Data, and Statistics section of their website. The information about weather was compiled from the Storm Events Database created by the NOAA's National Weather Service. The database itself was created in 1950, but it was not until 1993 that the data was updated into modern file formats with more detailed supplemental information on the storm events. The data in the database comes from a variety of sources, county, state and federal emergency management officials, SKYWARN spotters, National Weather Service (NWS) damage surveys, insurance industry and the general public among others. The NWS receives data up to 75 days at the end of each month. Each piece of data collected is validated for accuracy by the NWS. Since the database relies on the reporting of sources outside of NOAA, the NWS does not guarantee the accuracy of the information. However, the storm database does mention where the data was acquired, and the NWS recommended reaching out to that source if need be to verify the data[7]. Another limitation of the database is that there is no metric for determining severe weather, it is up to the individuals or organizations reporting to determine if the weather is severe enough. The Storm Data database is formally defined as documenting the occurrence of storms and other significant phenomena having sufficient intensity to cause loss of life or significant property damage. Additionally, the database documents rare weather phenomena, and other significant meteorological events such as record temperatures or precipitation[7]. Due to the range of information that is acceptable for the database, the data used for this reseachmust be taken with these concerns in mind.

Despite these limitations, there are several benefits to using the NWS database. It is the most comprehensive of its kind. Data is split into over 48 different weather patterns and separated to each county of origin within each state. Additionally, it provides further location information within each county, as well as the time of day that the event occurred. For certain weather patterns, there is also supplemental information available, like windspeed magnitude, injuries, and property damage. With the supplemental information about the weather patterns available, this data becomes a more viable option to analyze the best way to look at high-level trends in severe weather patterns and flight data.

These results will be useful because they will provide insight into how flight delays and weather are related. Additionally, since this study analyzes a network of airports, as well as their carrier, the study will provide insight into how delays can be connected between airports, and the interconnectedness of delays from airline carriers. The results can be applied to smaller scale studies, with further points of interest both in system modeling and in data analytics.

## III. Developing Data Analysis Methodology

The first objective was to look at the dataset to see how representative it was for a typical month. Additionally, the factors that could potentially affect delay had to be reasonably divided up into sub groups to see if there was any direct relationship. The process to do this involved looking at which months of the year had the most delays, and if the weather in January 2017 could be classified as more severe than normal.

By eliminating external factors that are deemed not important enough to have a large effect on the system, the data set could be simplified further. According to the Bureau of Transportation Statistics, the top 10 airline carriers control almost 90% of the airline domestic market in 2017[8]. After the 2008 economic recession, many airline carriers merged in order to stay order to stay in business. The string of mergers resulted in the industry shrinking form about 8 to 9 carriers in 2000, to only four in 2012. It is estimated that Delta, United, Southwest, and American control about 85% of all domestic air travel[9]. One of the airlines analyzed in this study, Virgin American, began its merger in 2017 as well, so the data for the airline is affected by the merging process[10]. It was decided to eliminate other airline carriers past the top 10 because they do not have enough flights to serve as meaningful data points. Another factor that could be eliminated were smaller airports. The FAA states that its 35 Operational Evolution Partnership airports are commercial U.S. airports with significant activity. More than 70 percent of passengers move through these airports[11]. The large quantity of data points available from these airports makes them viable options for this analysis. Furthermore, some of the airports are hubs for certain airline carriers, so only analyzing the top ten airports was not a viable option because it did not provide enough data for how these airline carriers operate.

Another important data point was looking to see if there were any extenuating events in that month that could have caused a rise in delays. This would mean any current events, that could either potentially pose a security risk such that a no-fly zone was created, or that there was a large sporting event that would cause an unusual influx of passengers. After researching, the analysis determined that there were no abnormal current events in January 2017 that could have resulted in an abnormal influx of data.

### A. Calculating Delay Data and Weather Data

The first step in the data was to find the number of delays for each day for each airport. A script was compiled to scan through the FAA database for each airport ID and export it to an excel sheet.

The next step in the process was to search the FAA database for the delays for each carrier for a specific airport for a given month. This was because air carrier information was analyzed to see how particular aircraft carriers were affected by weather patterns. This information was used to test the hypothesis that certain air carriers are more prone to delay.

Once the delay data had been collected, I analyzed flight patterns for each state during the month of January. By using the NOAA database, the weather patterns were split into 9 categories: winter storm, high wind, wind chill freeze, thunderstorm, dense fog, heavy rain, winter weather, ice storm, and hail. More categories are listed as an option on the NOAA site, however, some of the weather patterns listed as options did not occur in January, or they are not severe enough to effect flights, such as droughts.

The weather data was divided into two categories: overall weather patterns for the state in that day, and weather for the county that the airport is located in. The specialized weather data can provide context to weather patterns directly over the airport. If there is severe weather on that day, then it will be impossible for flights to leave at their designated time. Since the database only reports severe weather patterns as defined by those in the area, the resultant data will be important markers of how airports can better respond during severe weather patterns. In the second category, that marks significant weather patterns in the entire state. Since the planes will be approaching the airport from all cardinal directions, it is useful to look at other patterns affecting the state. There are too many data points to individually analyze the weather patterns in each state that the aircraft fly through.

Similar to the delay data, the weather data was coded so that days with significant weather problems were highlighted. This was so that it would be easier to visually see which days had the most overlap of factors.

After the initial data had been collected from both databases, there was enough information to create a probability model. The probability model was developed using the formula below.

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)} \tag{1}$$

This formula is a probabilistic model that states the likelihood of one event in a given category happening based off the information provided in the data set. In the formula, A represents the event that has already happened, and B represents the other event that could potentially happen. Data falls into both categories, or delayed flights, is in the numerator, and data that only falls into the first category, of flights that day, is in the denominator. This information was used to develop graphs of how the probability of delay changes over the course of a month for a given airport. There is a common conception that certain days of the week are better to travel on, so the data was split into one graph that was over the course of the month, and one graph that had 4 lines representing each week of the month. This way, the hypothesis that certain days routinely have a higher delay can be tested.

### B. Isolating the Data Set to Find Severe Days

While calculating the number of delays for each factor and counting weather patterns helped to shrink the massive data set size, the data is still large enough that it needed to be divided further to provide context on how the relationship between airport delay and weather patterns. In order to determine which days had a particularly high number of delays, the data was split into quartiles. Fig. 1 shows a color-coded table for the month of January for each airport. The table was organized in order of most popular airport to least popular airport. From observing the figure, the general trend is that there were more delays, and thus a higher chance of delay, at the start of the month. Furthermore, the color-coded table supports the hypothesis that certain days of the week will have higher chance of delays. On every weekend of the month, there was a higher chance of delay compared to the other days in that week. This conclusion was supported when the probability data was split into weeks for each airport and graphed over time. Additionally, the 21, 22 and 23 had a higher rate of delay because a cold front swept through the country on those days. There was winter storms and high wind occurring in 13 different states on the 21$^{st}$ and 23$^{rd}$, and 16 different states on the 22$^{nd}$.

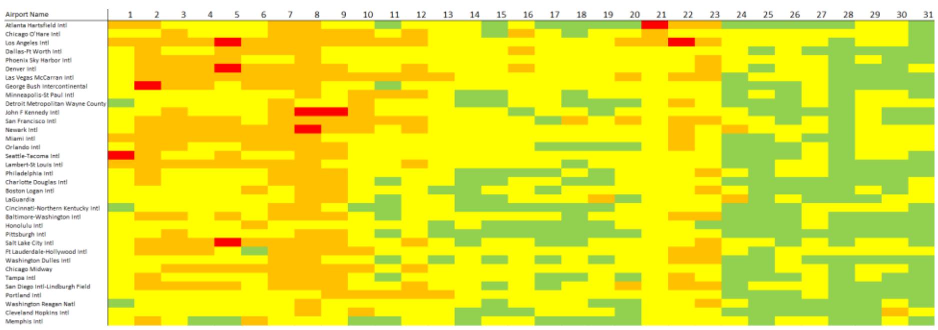

**Figure 1: Color coded Probability Table of Flight Delays**.

## IV. Results

### A. Analyzing Upper Quartile Probability Days and Airports

Days that were in the upper 75% quartile were analyzed first to test the hypothesis that weather can result in a higher chance of delay. There were only 7 days over the entire month that resulted in a probability of delay that was over 75% at one of the airports. These days and the predicted probability at each airport are shown in table 1.

**Table 1. Airports with Upper Quartile Probability of Delay**

| Day | Airport Name | Probability of Delay |
|---|---|---|
| 1 | Seattle-Tacoma International | 79.5875% |
| 2 | George Bush Intercontinental | 78.6967% |
| 5 | Los Angeles International | 76.9759% |
| 5 | Denver International | 77.8929% |
| 5 | Salt Lake City International | 84.8185% |
| 8 | John F. Kennedy International | 87.5969% |
| 8 | Newark International | 77.3852% |
| 9 | John F. Kennedy International | 77.5192% |
| 21 | Atlanta Hartsfield International | 75.3122% |
| 22 | Los Angeles International | 77.0073% |

There are a few interesting conclusions that can be drawn from table 1. The table demonstrates that it is not necessarily the most popular airport that will have the highest chance of delay, as Chicago O'Hare has the second most domestic flights and it did not have any days in the upper quartile. Another notable point about the table is that there are a few airports with several upper quartile days. This could be due to location and the climate that they are in, or logistic issues with the airport regarding how quickly airline carriers are able to access gates and prepare to fly. It could also be a compounding issue, with delays at other airports resulting in a later arrival time, and further delays for flights that need that plane.

In order to test to see if any of these issues were correlated with the upper Quartile Probability, the work analyzed extenuating factors during the particular day. Table 2 shows the rate of carrier delays, as well as any weather patterns during that day at the airport. The weather patterns are split up into weather throughout the state during that day, as well as weather affecting the individual county that the airport is in.

**Table 2. Additional Factors Contributing to Upper Quartile Delay**

| Day | Airport Name | Carriers with an Upper Quartile Delay | Regional Weather Patterns | State Weather Patterns |
|---|---|---|---|---|
| 1 | Seattle-Tacoma | American, Southwest, Delta, United, Jet Blue, Alaska, Spirit, SkyWest, Hawaiian | | Winter Storm |
| 2 | George Bush | American, Delta, United, Alaska, Spirit, SkyWest, Frontier, ExpressJet | Thunderstorm | High Wind |
| 5 | Los Angeles | American, Southwest, Delta, United, Jet Blue, Alaska, Spirit, SkyWest, Virgin America | | Heavy Rain |
| 5 | Denver | American, Delta, United, Jet Blue, Spirit, Frontier, ExpressJet | | Winter Storm |
| 5 | Salt Lake City | American, Southwest, Delta, United Jet Blue, Alaska, SkyWest, Frontier | | |
| 8 | John F. Kennedy | American, Delta, JetBlue, Alaska, Hawaiian, Virgin America | | |
| 8 | Newark | American, Southwest, Delta, United, Jet Blue, Spirit | | |
| 9 | John F. Kennedy | American, Delta, Jet Blue, Alaska, Virgin America | | |
| 21 | Atlanta Hartsfield | American, Southwest, Delta, Spirit | | High Wind |
| 22 | Los Angeles | American, Delta, United, Jet Blue Alaska, Spirit, SkyWest, Virgin America | Winter Storm | Winter Storm |

Table 2 correlates with the idea that with higher delays at a given airport, each carrier will likely to have more delays than normal. However, 40 percent of airports that had a high probability of delay did not have weather at the same time. There are several possible explanations for this phenomenon. One explanation is that the NOAA database only records severe weather patterns, and the airports may have been affected by weather that was otherwise not recorded by the database. Another possible explanation is that the upper quartile airports are affected by delays at other airports with severe weather patterns. While testing the first explanation is out of the scope of this research, it is possible to test the second.

### C. Analyzing Severe Weather Patterns on High Probability Delay Days

To test the hypothesis that severe weather throughout the country can result in higher probability of delay at airports, I first analyzed the weather patterns throughout the country on the days in question. Table 3 demonstrates the number of states that were experiencing severe weather on those days, as well as which regions of the country were experiencing severe weather, if any.

**Table 3. Weather Patterns Throughout the Country in January 2017**

| Day | States with Severe Weather | Regions With Severe Weather |
|---|---|---|
| 1 | CA,CO, FL,ID, MN, MT, NV, NM, OR, SD, TX, UT, WA | West, Southwest, Midwest |
| 2 | AL, CA, CO, FL, GA, IL, IN, KS, KY, LA, MN, MS, MO, NV, NH ND, OR, PA, SD, TX, UT, WY | Southeast, Midwest, Mid-Atlantic, Southwest |
| 5 | AR, CA, CO, DE, HI, IL, IN, KS, KY, MD, MI, MO, NE, NH, NM, NY, OH, OR, PA, TX, VA, WV, WI, WY | Southwest, Midwest, Mid-Atlantic |
| 8 | CA, CO, ID, MS, MT, NM, OR, UT, WA, WY | West, Southwest, Northwest |
| 9 | CA, CO, ID, MI, NV, NM, NC, OR, TX, UT, WI, WY | West, Midwest |

| | | |
|---|---|---|
| 21 | AL, AK, AZ, AR, CA, CO, FL, GA, HI, LA, MI, MS, NM, OR, SC TX, | Southwest, Southeast, Midwest |
| 22 | AL, AK, AZ, CA, CO, FL, GA, HI, ID, LA, MS, NV, NM, OR, TX, UT, WY | Southwest, West, Northwest Southeast |

From table 3, it can be concluded that although there was not severe weather over the high probability airports, there was significant weather patterns throughout the country on those days. This hypothesis is supported by previous figures, as well as the idea that days with a red block on them were more likely to have a higher probability of delay throughout.

The next factor to analyze was the effects of airline carriers on flight delays. If the flights being routed through one airport are delayed, then it can have negative effects on the rest of the system. Several of the aircraft carriers that were analyzed in this research use the spoke-hub model for transport optimization. One the drawbacks to this model is that it is relatively inflexible, and changes at the hub airport can have unexpected consequences throughout the model. While there are some aircraft carriers that do not use the hub model at all, most of them have at least one hub airport, and a few of them with several hub airports. Table 4 shows each airline carrier, and their top hub airport. If the carrier did not use the spoke-hub model, they were not included in this analysis. Additionally, the table has a section on if the aircraft carrier was dealing with more delays on that day, and if there was weather within that state on that day.

**Table 4. Airport Hub Delays and Weather Patterns in January 2017**

| Carrier | Airport Hub | Upper Quartile Days | Regional Weather | State |
|---|---|---|---|---|
| American | Dallas Ft. Worth | 2,5,8,9 | 2 | 1,2, 5, 9, 21,22 |
| Delta | Hartsfield Jackson Atlanta | 2, 8, 9, 21 ,22 | | 2, 21, 22 |
| United | Chicago O’Hare | 5, 8, 9 | | 2, 5 |
| Jet Blue | Kennedy | 1,2, 8 ,9 | | 5 |
| Alaska | Seattle-Tacoma | 1, 2, 5, 8, 9 | 8 | 1, 8 |
| Frontier | Denver | 1, 2, 5, 8, 9 | | 1, 2, 5, 8, 9, 21, 22 |
| Hawaiian | Honolulu | 2, 5, 21 | | 5, 21, 22 |
| Virgin America | San Francisco | 1, 2, 5, 8, 9, 22 | 22 | 1, 2, 5, 8,9 ,21, 22 |

The table is useful to providing context to how complex it is to determine the relationship between aircraft delays and weather patterns. Since most aircraft carriers use the spoke-hub model for transportation, any delays at one of their hubs can create a ripple effect throughout the rest of the system. Furthermore, when aircrafts delay, it is not always because of weather immediately in the area of the airport. The table also demonstrates that weather and flight delay is not necessarily a proportional relationship. There are several data points on the graph where the state had weather in the surrounding area and it did not result in higher delays at the airport. Conversely, there are days that had a higher probability of delay, and clear weather in the surrounding area.

## V. Conclusion and future work

In this paper, a comprehensive analysis of the busiest domestic airports and airline carriers was developed to see if there was any conclusive relationship between flight delays and weather patterns. After identifying days that had a statistically higher probability at a given airport, I analyzed factors that could have possibly affected this result. Severe weather patterns, both regionally and state-wise were analyzed. Additionally, I looked at the weather patterns for the hub airports for the airline carriers, to see if the weather at the hub location resulted in higher chance of delays at other locations. The work found that weather patterns are not linearly correlated with higher delays for airports, as airports could have severe weather patterns without more delays. Furthermore, severe weather patterns at hub airports do not guarantee a higher rate of delay either.

Future work will involve incorporating more sources for weather data to improve the robustness of the weather analysis. Prediction accuracy might improve if there were more sources to verify the weather patterns throughout the state, rather than one database. Additionally, more research on common flight patterns between hub airports and their spokes will need to be completed in order to gain an understanding of how airports are affected by weather patterns at the hub.

## Acknowledgments

I would like to thank my research professor, Dr. Dengfeng Sun, who helped me throughout this project with many valuable suggestions and statistical advice. Working with him is a wonderful experience.

## References

[1]"U.S. Passenger Carrier Delay Costs," http://airlines.org/dataset/per-minute-cost-of-delays-to-u-s-airlines/

[2]Belvedere, M., and LeBeau, P., "Weather flight disruptions cost $1.4 billion: Data" Available: https://www.cnbc.com/2014/01/08/weather-flight-disruptions-cost-14-billion-data.html.

[3]"Airline Passenger Travel to Nearly Double in Two Decades ," *Federal Aviation Administration News*, Mar. 2012.

[4]"FAA Aerospace Forecast Fiscal Years 2017-2037" Available: https://www.faa.gov/data_research/aviation/aerospace_forecasts/media/FY2017-37_FAA_Aerospace_Forecast.pdf

[5]Rumore, K., and Geib, P., "How the air traffic control system works (graphic)," *Chicago Tribune*, Sep. 2014.

[6]Vector, V., "How Air Traffic Controllers Help Planes Avoid Bad Weather," *The Points Guy* Available: https://thepointsguy.com/2015/09/insider-series-how-do-air-traffic-controllers-help-planes-avoid-bad-weather/.

[7]"Storm Data FAQ Page," *National Centers for Environmental Information* Available: https://www.ncdc.noaa.gov/stormevents/faq.jsp.

[8]"Carrier Snapshots," *Bureau of Transportation Statistics*Available: https://www.transtats.bts.gov/homepage.asp.

[9]Newman, R., "How Airline Mergers Saved an Industry-- and May Even Benefit Fliers," *U.S. News*, Feb. 2013.

[10]Goldman, D., and Ostrower, J., "Alaska Airlines is killing off the Virgin America Brand," *CNN Money*, Mar. 2017.

[11]"OEP 35" Available: http://aspmhelp.faa.gov/index.php/OEP_35.