

The Relationship Between 2011 METAR and TAF Data at Chicago-Midway and Seattle-Tacoma Airports

Michael A. Gallo and Matthew Kepto
Florida Institute of Technology

Abstract

This study examined the relationship between expected meteorological conditions as specified by TAF reports and actual ground conditions as specified by hourly METAR reports for Chicago-Midway (MDW) and Seattle-Tacoma (SEA) airports for the period September–December 2011. MDW and SEA were targeted because they had the highest and lowest percentage of delays, respectively, for 2011. The rationale was to determine if one of the contributing factors for the difference in percentage delays was because of the relationship between TAF and METAR reports. The primary hypothesis was that the relationship between the forecasts and actual ground conditions at MDW would be weaker than the corresponding relationship at SEA. TAF and METAR data were acquired from the respective TAF and METAR products pages at “Aviation Weather Charts Archive” (2012). Descriptive statistics revealed that MDW had less total departures than SEA (86,834 vs. 100,133) for all of 2011, but it also had nearly five times as many weather-related departure delays than SEA. Chi square analyses indicated that although the relationship between TAF and METAR at each airport was statistically significant, the corresponding Kappa agreement coefficients showed that this relationship was nearly twice as strong at MDW (.60) than at SEA (.35). Plausible explanations include that 70% of the weather conditions at MDW were VFR as opposed to only 56% at SEA, MDW had one-third the number of special METARS than SEA (374 vs. 917), and MDW had approximately one-fifth the number of LIFR conditions than SEA (70 vs. 337). The analysis also revealed that SEA had difficulty correctly forecasting IFR and LIFR conditions, especially under rapidly changing conditions. Based on the study’s findings, it appears that the relationship between TAF and METAR was not a contributing factor to departure delays at both MDW and SEA during the September–December 2011 period.

Introduction and Background

Airport delays are a common and often expected occurrence within the airline industry. The Bureau of Transportation Statistics (BTS), which collects aviation data, organizes the reason for airport delays into five main categories: carrier, weather, National Airspace System (NAS), security, and late aircraft arrival. Of these, “weather has been identified as the most important causal factor for NAS delays” (Sridhar & Kulkarni, 2008, p. 1) and has the greatest impact on airports.

“According to FAA statistics, weather is the cause of approximately 70% of the delays in the National Airspace System” (Kulesa, N.D., p. 1). Kulesa also reported “weather continues to play a significant role in a number of aviation accidents and incidents,” contributing to 23% “of all aviation accidents” (p. 1). Kulesa indicated that the total impact of weather “is an estimated national cost of \$3 billion for accident damage and injuries, delays, and unexpected operating costs” (p. 1). Some of the weather-related delays cited by Kulesa (N.D.) included thunderstorms and other convective weather, in-flight icing, turbulence, ceiling and visibility, ground de-icing, and volcanic ash.

Klein, Craun, and Lee (2010) reported “understanding airport delays, their causes and their relationship with inclement weather has been the subject of research for many years, especially since the late 90’s” (p. 1). This research has benefited from the combined efforts of federal organizations such as the National Weather Service (NWS), Federal Aviation Administration (FAA), Department of Defense (DOD) and NASA, private organizations such as MITRE, MIT Lincoln Lab, and academic institutions such as MIT, University of Maryland, and George Mason University (Klein et al., 2010).

Most of the research focus has been on developing models of delay. For example, using data from BTS and an open-source package called Weka (Hall et al., 2009), which is a collection of machine learning algorithms for data mining purposes, Stefanski (2009) developed models for predicting flight delays based on various attributes of a particular flight. Because of the voluminous amount of data, Stefanski limited his analysis to departing flights during the month of February 2008, and focused on seven attributes: day of week, airport origin, carrier, departure time, departure delay time, and distance the flight must travel after departure. As part of his findings, Stefanski reported that “airports and carriers may play a key role in determining whether a flight will be delayed or not” (p. 4), and “it is possible to make fairly good predictions on the basis of a few key attributes, such as carrier, departure time, date, and airport” (p. 7). A drawback to Stefanski’s study, though, is that the data were limited to a single month and he did not include weather as one of his attributes.

At the 26th International Congress of the Aeronautical Sciences, Sridhar and Kulkarni (2008) reported on their research, which involved developing models relating national delay, center level delays, and weather. They developed their models using the Weather Impacted Traffic Index (WITI), which is a metric of the number of aircraft affected by weather at a given instant of time. WITI uses National Convective Weather Diagnostic reports as well as METAR, which is “the primary observation code used in the U.S. to satisfy World Meteorological Organization (WMO) and International Civil Aviation Organization (ICAO) requirements for reporting surface meteorological data” (Aviation Weather Services, 2010, p. 3-1). Sridhar and Kulkarni restricted their analysis to traffic data for the 5-month period between April and August for the years 2004–2006, inclusive. The centers Sridhar and Kulkarni targeted were the 20 FAA Air Route Traffic Control Centers (ARTCC) within the continental U.S.

Sridhar and Kulkarni (2008) found mostly small correlations ($< .30$) between national WITI and center delays, which indicate that the national WITI is not a good predictor of weather delays at these centers. The centers with the lowest correlations ($-.02$ to $.03$) included Seattle (ZSE), Oakland (ZOA), Salt Lake City (ZLC), Albuquerque (ZAB), and Minneapolis (ZMP). The center with the largest correlation was New York (ZNY) at $.40$, which indicates that national WITI may be a good predictor of weather delays at the New York center. When Sridhar and Kulkarni examined the relationship between each center's respective WITI and center delays, they found that all centers except Seattle ($-.02$) had a positive correlation that ranged between $.17$ (Jacksonville, ZJX) and $.72$ (Houston, ZHU). This finding suggests that the Seattle center, which covers Washington, most of Oregon, and parts of California and Idaho, is unique because the weather delays at the center were related to neither the national WITI nor the center's own WITI.

Sridhar and Kulkarni (2008) also examined the impact of weather in each center on NAS delays. They found that the Oakland and Seattle centers had the lowest average daily WITI, 79 and 83, respectively, which indicates that these regions had the fewest number of aircraft affected by weather at a given instant of time. Among the 20 centers, though, the Seattle center was the only one with a zero average daily contribution to national delays. On the other hand, the center with the highest average daily contribution to national delays was the Chicago (ZAU) center, with an average WITI of 1,476. This center covers the northern half of Illinois, the southern Wisconsin, the eastern Iowa, and parts of Indiana and Michigan.

Instead of focusing on ARTCC as Sridhar and Kulkarni (2008) did, Klein et al. (2010) used WITI for predicting airport delays. Klein et al. initially used a 3-component WITI that included en route component (E-WITI), the terminal component (T-WITI), and the queuing delay component (Q-DELAY). E-WITI "reflects the impact of convective weather on routes connecting major airports," T-WITI "captures capacity degradation resulting from surface weather impact, proportional to the number of operations at an airport," and Q-DELAY "measures the cumulative effect of traffic demand in excess of capacity" (Klein et al., p. 2). They found that this 3-component WITI was insufficient to identify the weather's impact on individual airports.

As a result, Klein et al. (2010) modified this model and developed a 12-component airport WITI that included: E-WITI, which does not depend on the airport's terminal weather; volume WITI, which is based only on traffic; local convective weather; wind; snow; IMC data, which includes ceiling or visibility below airport specific minima, fog, and heavy rain; and other, which includes "minor impacts due to light/moderate rain or drizzle but ceilings/visibility above VFR minima (and) unfavorable RWY configuration usually due to light-to-moderate winds (15-20 Kt or even 10 Kt) that prevent optimum-capacity runway configurations from being used" (p. 3). These latter five components were then converted into T-WITI (linear) and Q-DELAY (nonlinear). Klein et al. (2010)

tested their model comparing predicted delays “for several major airports and for two different seasons (summer, winter)” (p. 7) to actual delays for specific dates in 2008 and 2009. They found the model to be robust and “sufficiently sensitive to weather forecast inaccuracies (and therefore) [...] can be used for convective and non-convective forecast product evaluation” (p. 12).

Pearson (2002) reported that among the fatal general aviation aircraft accidents that occurred between 1995 and 2000, two significant factors in 63% of the accidents were low ceilings and visibilities, which means the accidents occurred during Instrument Flight Rules (IFR) conditions. In an effort to shed light on the importance of IFR conditions in Terminal Aerodrome Forecasts (TAFs), Thompson and Baumgardt (2009) examined hourly METARs from 1961–2009 “to achieve climatological averages and percentiles for IFR conditions for two airports: La Crosse, WI (LSE) and Rochester, MN (RST). In contrast to a METAR, which contains the current meteorological conditions, a TAF contains the forecasted conditions. It is a “concise statement of the expected meteorological conditions significant to aviation for a specified time period within 5 statute miles of the center of the airport’s runway complex (terminal)” (Aviation Weather Services, 2010, p. 7-19).

Thompson and Baumgardt (2009) found that “IFR conditions have the highest frequency of occurrence in the Upper Mississippi Valley during the winter months, November through March, with fog being the major weather contributor” and that snow also contributes about 30% to IFR conditions in cool season (p. 4). Thompson and Baumgardt also reported that METAR data from 1961–1990, showed that “measurable snow events have a direct correlation to IFR conditions and approximately 90% of IFR visibilities occur rapidly, or within 2 hours of snow onset with little difference between the two airports investigated” (p. 4).

Thompson and Baumgardt’s (2009) findings suggest that with respect to weather involving snow, “TAF utilize IFR as prevailing conditions when confidence is high in light measurable snow events (and) that IFR conditions be forecast quickly after snow onset” (p. 4). To do this, Thompson and Baumgardt suggested that meteorologists use a variety of data, including historical METAR data as well as data from the Localized Aviation Model Output Statistics (MOS) Program (LAMP) in their forecast preparation process. Thompson and Baumgardt further suggested that because hourly historical METAR records are available at many locations that researchers examine historical METAR data to see how they related to TAF.

There is no argument that air traffic delays are a common phenomenon within the aviation field and that “inclement weather is the single biggest factor causing air traffic delays in the U.S.” (Klein, Kavoussi, & Lee, 2009, p. 1). The literature reviewed here demonstrates both the diversity of studies being conducted with respect to this issue as well as some of the limitations. For example, Stefanski (2009) focused only on delays in general without regard to weather and restricted his study to only a single month “to

reduce the sheer size of the dataset” (p. 1). Although Sridhar and Kulkarni (2008) focused on weather delays specifically, they restricted their studies to those occurring at the 20 ARTCCs within the continental U.S. They also focused on 5 non-winter months (April to August) for a 3-year period. Klein et al. (2010) developed a robust airport delay prediction model and applied this model to predict delays using past delay data, but they did not examine differences in delays among airports. Lastly, Thompson and Baumgardt (2009) concentrated on historical METAR data at two airports to see how they were related to IFR conditions in corresponding TAFs.

Purpose Statement and Operational Definitions

Following Thompson and Baumgardt’s (2009) recommendation to examine historical METAR data to see how they relate to TAF, the purpose of the current study was to examine the relationship between TAF and METAR at two airports: Seattle-Tacoma (SEA) and Chicago-Midway (MDW). The reason for selecting these airports was based on Sridhar and Kulkarni’s (2008) findings with respect to the Seattle and Chicago ARTCCs, and on data from the Bureau of Transportation and Statistics, which showed MDW and SEA had the highest and lowest percentage of delays, respectively, for 2011. The rationale was to determine if the relationship between TAF and METAR was a contributing factor to weather delays at these airports in 2011.

The current study was guided by the following research questions: (1) What is the relationship between METAR and TAF at SEA and MDW, respectively, in 2011? and (2) To what extent was the relationship between METAR and TAF a contributing factor to weather delays at SEA and MDW, respectively, in 2011? In the context of the study, weather conditions were defined with respect to visibility and ceiling height, which were classified as Low Instrument Flight Rules (LIFR), Instrument Flight Rules (IFR), Marginal Visual Flight Rules (MVFR), and Visual Flight Rules (VFR). Federal Aviation Regulations define LIFR conditions as a ceiling below 500 feet above ground level (AGL) and/or less than 1 statute mile visibility. IFR conditions are defined as 500 feet AGL to below 1000 feet AGL and/or 1 statute mile to below 3 statute miles visibility. MVFR conditions are defined as a ceiling 1000 to 3000 feet AGL and/or 3 to 5 statute miles visibility. VFR conditions are defined as a ceiling greater than 3000 feet AGL (or no ceiling) and greater than 5 statute miles visibility. Furthermore, “ceiling” was defined as overcast conditions or broken cloud cover, and “no ceiling” was defined as clear skies, few clouds, or scattered clouds. An overcast cloud layer covers 8/8 of the sky, a broken layer covers 5/8-7/8 of the sky, scattered clouds cover 3/8-4/8, few clouds cover 1/8-2/8, and clear skies are no clouds (Aviation Weather Services, 2010, pp. 3-13). The cloud cover was automatically reported as overcast, broken, scattered, few or clear in the TAF and METAR reports.

Methodology

The population for this study was all TAF and METAR reports for 2011 from MDW and SEA airports. Because of limited access to 2011 TAF and METAR reports, we used a convenience sampling strategy that delimited the sample to the 4-month period September through December 2011. Data collection consisted of accessing the on-time flight performance database (“Research and Innovative Technology Administration,” 2012). We then selected the “Get Lookup Table” link in the “OriginAirportID” section under the “Origin” headline to acquire the codes for Chicago-Midway (KMDW = 13232) and Seattle-Tacoma (KSEA = 14747). To retrieve airport data, we selected the following field names and descriptors for the targeted airports: dep_delay_new, arr_delay_new, cancelled, cancellation_code, diverted, flights, carrier_delay, weather_delay, NAS_delay, security_delay, late_aircraft_delay. Once all the appropriate descriptors were selected, we downloaded the corresponding data file and then prepared data tables using Excel. We collected TAF and METAR data from the respective TAF and METAR products pages at “Aviation Weather Charts Archive” (2012). The METARs were placed into Excel and labeled with the appropriate weather category and then matched with the corresponding TAF. The completed Excel tables were then loaded into the statistical software JMP Pro (2012) for data analysis.

Data Analysis

Descriptive Statistics

A summary of the number and minutes of flight delays by category departing MDW and SEA for the last 4 months of 2011 are provided in Table 1 and Table 2, respectively. As reported in these tables, MDW had 1,073 weather-related delays and SEA had 231 weather-related delays. Thus, MDW had 4.65 times more the number of weather-related delays than SEA. When examined with respect to total delays, the percentage of weather-related delays for MDW was $1073/49141$, or about 2.1%, and the percentage of weather-related delays for SEA was $231/31842$, or about 0.7%. When these data were examined with respect to number of minutes of delay, MDW had 47,958 minutes of weather-related delay out of 1,179,938 total delays, which is about 4.1%. Similarly, SEA had 7,814 minutes of weather-related delay out of 693,912 total delays, which is about 1.1%. Finally, when total departing flights are considered, MDW had 86,834 flights in all of 2011, of which 1073 were weather-related delays (1.2%) whereas SEA had 100,133 flights in all of 2011, of which 231 were weather-related delays (0.2%). In short, MDW had nearly five times as many weather-related delays as SEA despite less total departures.

A summary of the comparison between TAF and METAR for MDW and SEA for September–December 2011 is provided in Table 3. As reported in Table 3, each airport had 2,928 regularly scheduled METARs over the 4-month period. MDW had 3,302 total METAR reports of which 374 were special, unscheduled issuances. SEA had 3,845 total

METAR reports of which 917 were special, unscheduled reports giving SEA 2.45 times more special reports. Of the 3,302 METARs at MDW, the TAF correctly forecasted 2,137 (65%) VFR conditions, 396 (12%) MFR conditions, 22 (0.6%) LIFR conditions, and 142 (4%) IFR conditions. Of the 3,845 METARs at SEA, the TAF correctly forecasted 1,799 (47%) VFR conditions, 540 (14%) MFR conditions, 68 (1.8%) LIFR conditions, and 53 (1.4%) IFR conditions.

A summary of the TAF vs. METAR agreements for MDW for the last quarter of 2011 is provided in Table 4. The METAR reported VFR conditions 2,322 times (70.3%), MVFR occurred 638 times (19.3%), LIFR occurred 70 times (2.1%), and IFR conditions occurred 272 times (8%). The TAF was correct in forecasting 92% of the time for VFR conditions, 62.1% for MVFR, 31.4% for LIFR, and 52.2% of the time for IFR.

A summary of the TAF vs. METAR agreements for SEA for September–December of 2011 is provided in Table 5. The METAR reported VFR conditions 2,173 times (56.5%), MVFR occurred 1,042 times (27.1%), LIFR occurred 337 times (8.8%), and IFR conditions occurred 293 times (7.6%). The TAF was correct in forecasting 82.8% of the time for VFR conditions, 51.8% for MVFR, 20.2% for LIFR, and 10.9% for IFR.

Table 1

Number and Minutes of Delayed Flights by Category Departing Chicago-Midway (MDW) in 2011 by Month

Month	Type of Delay											
	Carrier		Weather		NAS		Security		Late Aircraft		Total Delays ^a	
	N	Min.	N	Min.	N	Min.	N	Min.	N	Min.	N	Min.
Jan.	1443	32498	103	2264	718	10829	0	0	1286	34565	4543	113843
Feb.	1075	24793	74	2140	867	15958	6	54	1002	27836	3501	83975
Mar.	1070	26391	56	2890	732	13164	8	108	957	32939	4297	98114
Apr.	1341	33908	95	4900	1006	20971	2	79	1395	57945	4511	136827
May	1344	31944	191	7275	955	21876	11	108	1416	66321	4872	151464
June	1213	31719	167	8298	776	19178	0	0	1230	56022	4781	140095
July	825	20681	73	3873	534	13093	1	9	835	32814	4376	101367
Aug.	917	21644	155	7532	554	15235	0	0	947	35852	4258	108973
Sep.	846	18968	69	2979	443	8437	1	125	833	29828	3957	92396
Oct.	690	16284	24	1826	412	8194	3	66	588	15541	3753	69914
Nov.	420	11743	24	1168	252	5007	1	7	377	11727	2950	20810
Dec.	444	13425	42	2813	248	5759	4	46	406	13389	3342	62160
Total	11628	283998	1073	47958	7497	157701	37	602	11272	414779	49141	1179938

Note. N = Total number of flights per category. Source: U.S. Dept. of Transportation's Bureau of Transportation Statistics (<http://www.transtats.bts.gov>).

^aTotal number of delayed flights and corresponding minutes represent *all* delays for 2011, including those not represented in the table. For example, NAS delays often include post-takeoff delays such as holds, which are not reported here.

Table 2

Number and Minutes of Delayed Flights by Category Departing Seattle-Tacoma (SEA) in 2011 by Month

Month	Type of Delay										Total Delayed ^a		Total Flights ^b
	Carrier		Weather		NAS		Security		Late Aircraft		N	Min.	N
	N	Min.	N	Min.	N	Min.	N	Min.	N	Min.			
Jan.	483	16145	37	1492	474	12960	2	27	471	16704	2617	60686	7585
Feb.	496	17967	60	1903	628	15744	2	17	446	16455	2286	58723	6812
Mar.	585	24723	17	571	651	17091	3	48	453	17750	2989	70508	7859
Apr.	443	16365	17	403	475	12583	6	93	407	16351	2449	58327	7719
May	572	17715	6	211	855	23402	8	109	376	14135	2579	54989	8537
June	540	19896	10	741	677	14922	5	45	496	20458	3018	66396	9337
July	563	20679	5	227	719	16923	3	48	515	22303	3022	68276	9828
Aug.	548	19396	4	258	554	15177	8	93	399	17299	3335	65791	9735
Sep.	376	14205	3	29	597	14444	1	24	214	8585	2251	39205	8558
Oct.	340	13928	7	356	497	12731	2	22	251	9892	2129	43485	8337
Nov.	425	18219	33	1130	615	16580	0	0	314	11572	2511	56141	7752
Dec.	457	15405	32	493	709	16592	1	37	368	13348	2656	51385	8074
Total	5828	214643	231	7814	7451	189149	41	563	4710	184852	31842	693912	100133

Note. N = Total number of flights per category. Source: U.S. Dept. of Transportation's Bureau of Transportation Statistics (<http://www.transtats.bts.gov>).

^aTotal number of delayed flights and corresponding minutes represent *all* delays for 2011, including those not represented in the table. For example, NAS delays often include post-takeoff delays such as holds, which are not reported here. ^bTotal flights represent *all* flights for each month and overall.

Table 3

Comparison between TAF and METAR for MDW and SEA (September–December 2011)

TAF	Airport							
	Chicago-Midway (MDW) ^a				Seattle-Tacoma (SEA) ^b			
	METAR				METAR			
	VFR	MVFR	LIFR	IFR	VFR	MVFR	LIFR	IFR
VFR	<u>2137</u>	162	12	32	<u>1799</u>	406	152	93
MVFR	165	<u>396</u>	20	76	332	<u>540</u>	63	115
LIFR	0	3	<u>22</u>	22	21	34	<u>68</u>	32
IFR	20	77	16	<u>142</u>	21	62	54	<u>53</u>
Total	2322	638	70	272	2173	1042	337	293

Note. ^aN = 3302 METARS of which 374 were special reports. ^bN = 3845 METARS of which 917 were special reports. The frequencies along the diagonals (underscored) represent the number of times TAF forecasts matched corresponding METARS. These are further elaborated in Table 4 and Table 5.

Table 4

TAF vs. METAR Agreements for MDW (September–December 2011)

Report	VFR		MVFR		LIFR		IFR	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
TAF ^a	2137	92.0%	396	62.1%	22	31.4%	142	52.2%
METAR	2322	70.3%	638	19.3%	70	2.1%	272	8%

Note. ^aTAF percentages represent the ratio of TAF to METAR. For example, of the 2322 VFR METARs, TAF was correct 2137 times, or 92%. ^bMETAR percentages represent the ratio of METAR frequencies (*N*) to the total number of METAR reports (3302). For example, VFR conditions were observed 2322/3302, or 70% of the time.

Table 5

TAF vs. METAR Agreements for SEA (September–December 2011)

Report	VFR		MVFR		LIFR		IFR	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
TAF ^a	1799	82.8%	540	51.8%	68	20.2%	32	10.9%
METAR	2173	56.5%	1042	27.1%	337	8.8%	293	7.6%

Note. ^aTAF percentages represent the ratio of TAF to METAR. For example, of the 2173 VFR METARs, TAF was correct 1799 times, or 82.8%.

^bMETAR percentages represent the ratio of METAR frequencies (*N*) to the total number of METAR reports (3845). For example, VFR conditions were observed 2173/3845, or 56.5% of the time.

Inferential Statistics

A summary of the results of the Chi-square analysis and agreement statistics between TAF and METAR for MDW and SEA by month is provided in Table 6. As reported in Table 6, the relationship between TAF and METAR was statistically significant for each of the targeted months for both MDW and SEA. In September for MDW, $\chi^2 = 403.87$, $df = 9$, $p < .0001$, and for SEA, $\chi^2 = 280.02$, $df = 9$, $p < .0001$. In October for MDW, $\chi^2 = 352.99$, $df = 4$, $p < .0001$, and for SEA, $\chi^2 = 204.03$, $df = 9$, $p < .0001$. In November for MDW, $\chi^2 = 606.85$, $df = 9$, $p < .0001$, and for SEA, $\chi^2 = 208.65$, $df = 9$, $p < .0001$. In December for MDW, $\chi^2 = 571.73$, $df = 9$, $p < .0001$, and for SEA, $\chi^2 = 334.68$, $df = 9$, $p < .0001$.

Table 6

Results of Chi-Square Analyses and Agreement Statistics between TAF and METAR for MDW and SEA by Month (September–December 2011)

Month	Airport							
	Chicago-Midway (MDW)				Seattle-Tacoma (SEA)			
	<i>N</i>	<i>df</i>	χ^2 ^a	Kappa ^b	<i>N</i>	<i>df</i>	χ^2	Kappa
September	821	9	403.87*	0.54*	830	9	280.02*	0.51*
October	790	4	352.99*	0.71*	1040	9	204.03*	0.25*
November	823	9	606.85*	0.63*	920	9	208.65*	0.35*
December	868	9	571.73*	0.53*	1055	9	334.68*	0.30*

Note.^aChi-square reflects likelihood ratio. ^bKappa coefficient is an agreement statistic that varies between 0 and 1 where 0 = no agreement between the factors and 1 = perfect agreement between the factors.

* $p < .0001$.

Table 7

Results of Chi-Square Analyses and Agreement Statistics between TAF and METAR for MDW and SEA Overall (September–December 2011)

Airport	<i>N</i>	<i>df</i>	χ^2 ^a	Kappa ^b
Chicago-Midway (MDW)	3302	9	2021.27*	0.60*
Seattle-Tacoma (SEA)	3845	9	1077.46*	0.35*

Note.^aChi-square reflects likelihood ratio. ^bKappa coefficient is an agreement statistic that varies between 0 and 1 where 0 = no agreement between the factors and 1 = perfect agreement between the factors.

* $p < .0001$.

In addition to the Chi-square analyses, the Kappa coefficient also was calculated for each airport. The Kappa coefficient is an agreement statistics that varies between 0 and 1 where 0 signifies no agreement between factors and 1 signifies perfect agreement between factors. As indicated in Table 6, the monthly Kappa coefficients were significant, which indicates there was a significant relationship between TAF and METARs at each airport. In September for MDW, Kappa = .54, $p < .0001$, and for SEA, Kappa = .51, $p < .0001$. In October for MDW, Kappa = .71, $p < .0001$, and for SEA, Kappa = .25, $p < .0001$. In November for MDW, Kappa = .63, $p < .0001$, and for SEA, Kappa = .35, $p < .0001$. In December for MDW, Kappa = .53, $p < .0001$, and for SEA, Kappa = .30, $p < .0001$.

A summary of the results of the Chi-square analysis and agreement statistics between TAF and METAR for MDW and SEA for the 4-month period, September–December 2011 is provided in Table 7. As reported in Table 7, the overall relationship between TAF and METAR was statistically significant. For MDW, $\chi^2 = 2021.27$, $df = 9$, $p < .0001$, and for SEA, $\chi^2 = 1077.46$, $df = 9$, $p < .0001$. The corresponding overall agreement statistics also were significant: For MDW, Kappa = .60, $p < .0001$; for SEA, Kappa = .35, $p < .0001$.

Discussion

When the study's results are applied to the first research question, the relationship between METAR and TAF data from September–December 2011 at Seattle-Tacoma and Chicago-Midway, respectively, was statistically significant, which indicates that forecasts at both airports were related to actual weather conditions. The TAF-METAR relationship at MDW, however, was nearly twice as strong as the TAF-METAR relationship at SEA as given by the Kappa coefficient. One plausible reason why MDW had a better TAF-METAR agreement was because 70% of the weather conditions at MDW during the targeted 4-month period were VFR as opposed to only 56% at SEA. SEA also had LIFR conditions 337 times (9%) whereas MDW only had LIFR conditions 70 times (2.1%). The greater prevalence of VFR conditions at MDW coupled with the greater prevalence of LIFR conditions at SEA suggests that MDW's forecasts would be more accurate than SEA's forecasts. This makes sense from a meteorological standpoint because it is typically easier to forecast good weather conditions (i.e., VFR) than bad weather conditions (i.e., IFR or LIFR).

With respect to the second research question, because the forecasts at both airports had strong statistical agreements with the actual ground conditions, it appears that the weather forecasts at MDW and SEA were not a contributing factor to weather delays from September–December 2011. What was surprising, though, was that MDW's Kappa agreement coefficient of .60 was nearly twice as high as SEA's Kappa coefficient of .35. Although SEA had more challenging weather than MDW as evidenced by more LIFR conditions and less VFR conditions, and SEA also had more total departures than MDW, SEA still had less weather related departure delays than MDW. Thus, we expected SEA to have a higher Kappa coefficient than MDW.

A plausible explanation for this finding is the number of METAR reports. Referencing Table 6, the number of METARs at MDW and SEA for the month of September was nearly the same at 821 and 830, respectively. The corresponding Kappa coefficients also were similar at .54 and .51, respectively. However, for October–December, SEA had considerably more METAR observations than MDW, which equated to lower Kappa coefficients. The Kappa coefficients for MDW during the last 3 months were much higher than those for SEA. These findings suggest an inverse relationship between the number of METARs and the Kappa coefficient: As the number of METARs increases, the agreement statistics between TAF and METAR decreases. This inverse

relationship is plausible when the special METAR (SPECI), which is non-routinely issued when weather is changing rapidly, is taken into consideration because it is easier to forecast weather that is relatively constant than weather that is changing rapidly in a short time period. More concretely: Unchanging, good weather (VFR), which was the general case at MDW, is much easier to forecast than rapidly changing bad weather (SPECI with LIFR), which was the general case at SEA.

Accenting the TAF-METAR relationship discussion, the data also suggest that during the targeted 4-month period SEA, when compared to MDW, was challenged in forecasting IFR and LIFR conditions, especially when conditions changed rapidly. For example, referencing Table 4 and Table 5, SEA correctly forecasted LIFR conditions 20% of the time vs. 31% of the time for MDW, and SEA correctly forecasted IFR conditions only 11% of the time compared to 52% of the time for MDW.

In conclusion, the data indicate that a plausible explanation for Chicago-Midway's weather-related departure delays compared to that of Seattle-Tacoma's is not poor forecasting because the results reflect the opposite: MDW had the higher frequency of weather-related delays but it also had a stronger agreement between TAF and METAR when compared to SEA. Although surprising, the data provided plausible explanations for the outcome. A better understanding of the reasons for the weather related departure delays at MDW and SEA may be possible with further research on what causes the weather delays, given that the forecasts were not a contributing factor. Replicating the current study using 2012 and 2013 data, and data from a time period other than September–December could reveal further insight and possibly different results. Because of the high number of unscheduled SPECI reports (particularly at SEA), it also may be beneficial to replicate the study with only regularly scheduled METAR reports. This modification may reveal different results because the rapidly changing weather between observations would be not taken into consideration.

Limitations and Delimitations

Limitations refer to circumstances or events that are beyond the control of the researcher. One limitation to the current study is that we used data provided by the respective airports. Other studies that use different weather data (e.g., “dominant” weather only might be reported) might get different results. A second limitation is TAF data were provided directly from the targeted airports and were not prepared by a different airport. Therefore, any subsequent study involving an airport that relies on TAF data from a region and not directly from the airport itself might get different results. A third limitation is that our findings are relevant to U.S. TAF data and not to the European model, which issues short TAFs (Jacobs, 1998). A final limitation is that the findings are reflective of the last quarter of 2011, not the entire year. Thus, similar studies that use an entire year's worth of data might get different results.

Delimitations refer to circumstances or events the researcher imposes on the study that further limits the generalizability of the results. One delimitation of this study is that although TAF reports include wind, visibility, weather, and cloud reports, we considered only visibility and clouds. Furthermore, only the more severe weather phenomenon was considered. For example, if thunderstorms and rain were reported, then only thunderstorms were included. Thus, other studies that use all of the data provided by TAF might not get the same results. A second delimitation is that we restricted METAR and TAF data to only September–December of 2011. Other studies involving the same targeted airports but use historical METAR and TAF data from a different time period might not get the same results. A third delimitation is that we focused on the two airports with the least and most delays in 2011. Studies that use the same selection criteria for a different time period will not necessarily involve the same airports. A final delimitation is that we restricted this study to weather-related delays on departure. Studies that focus on weather-related delays on arrival might not get the same results.

References

- Aviation weather charts archive. (2012). Retrieved from <http://aviationwxchartsarchive.com/page/home>
- Aviation Weather Services (2010). *Advisory circular 00-045G, change 1*. U.S. Department of Commerce, National Oceanic and Atmosphere Administration, National Weather Service, U.S. Department of Transportation's, and Federal Aviation Administration. Retrieved from http://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentID/215166
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations (11)*, 1.
- Jacobs, A. (2008). First guess TAF—FGTAF: Semi-automation in TAF production. Retrieved from <http://www.knmi.nl/bibliotheek/knmipubTR/TR212.pdf>
- JMP Pro (Version 10.0.0) [Software]. (2012). Available from <http://jmp.com/software/pro>
- Klein, A., Craun, C., & Lee, R. S. (2010). *Airport delay prediction using weather-impacted traffic index (WITI) model*. Paper presented at the 29th Digital Avionics Systems Conference (October 3–7). DOI: 10.1109/DASC.2010.5655493. Retrieved from http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5655493&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D5655493
- Klein, A., Kavoussi, S., & Lee, R. S. (2009). *Weather forecast accuracy: Study of impact on airport capacity and estimation of avoidable costs*. Paper presented at the 8th USA/Europe Air Traffic Management Research and Development Seminar (ATM2009) Retrieved from http://www.atmseminar.org/seminarContent/seminar8/papers/p_008_W.pdf
- Kulesa, G. (N.D.). *Weather and aviation: How does weather affect the safety and operations of airports and aviation, and how does FAA work to manage weather-related effects?* Retrieved from <http://climate.dot.gov/documents/workshop1002/kulesa.pdf>
- Pearson, D. C. (2002). *VFR flight not recommended: A study of weather-related fatal aviation accidents*. NOAA/NWS Technical Attachment, Southern Region Headquarters, Dallas, TX, TA 2002-18.

- Research and innovative technology administration: Bureau of transportation statistics. (2012). Retrieved from http://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Name=On-Time
- Sridhar, B., & Kulkarni, D. (2008). *Prediction of weather related delays in the continental United States*. Paper presented at the 26th International Congress of the Aeronautical Sciences (UCAS 2008, September 14–19). Retrieved from <http://www.icas-proceedings.net/ICAS2008/ABSTRACTS/435.HTM>
- Stefanski, T. (2009). *Predicting flight delays through data mining*. Retrieved from http://cs-people.bu.edu/dgs/courses/cs105/hall_of_fame/timoteo.html
- Thompson, S., & Baumgardt, D. (2009). Improving forecasts of instrument flight rule conditions over the upper Mississippi valley and beyond. Paper presented at the 23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction. Retrieved from <https://ams.confex.com/ams/pdfpapers/154154.pdf>