Available online at http://docs.lib.purdue.edu/jate

JATE

Journal of Aviation Technology and Engineering 10:1 (2021) 21–29

A Comparison of the Localized Aviation MOS Program (LAMP) and Terminal Aerodrome Forecast (TAF) Accuracy for General Aviation

Douglas D. Boyd and Thomas A. Guinn

Embry Riddle Aeronautical University

Abstract

Background. For general aviation (GA) pilots, operations in instrument meteorological conditions carry an elevated risk of a fatal accident. As to whether a GA flight can be safely undertaken, aerodrome-specific forecasts (TAF, LAMP) provide guidance. Although LAMP forecasts are more common for GA-frequented aerodromes, nevertheless, the Federal Aviation Administration recommends that for such aerodromes (and for which a TAF is not issued) a pilot use the TAF generated for the geographically closest airport for pre-flight weather evaluation. Herein, for non-TAF-issuing airports, the LAMP (sLAMP) predictive accuracy for visual flight rules (VFR) and instrument flight rules (IFR) flight category was determined.

Method. sLAMP accuracy was evaluated over 12 months using the fractions of forecasts which were correct or false alarms. Statistical differences employed chi-square/Fisher exact tests.

Results. sLAMP forecasts ($n = 570$) across 43 states were accrued. The fraction of correct sLAMP forecasts for VFR (0.53) and IFR (0.68) exceeded ($p < 0.002$) that for the TAF (0.35 and 0.45 respectively) the latter generated at proximate aerodromes. Conversely the fraction of false alarms for VFR and IFR generated by the sLAMP and TAFs were comparable or lower.

Conclusion. Our findings indicate the forecast superiority of the sLAMP at non-TAF-issuing airports compared with the TAF generated at a proximate aerodrome.

Practical Application. For non-TAF-issuing sLAMP airports, these findings argue for greater integration of the latter tool in pre-flight weather briefings for GA operations.

Keywords: general aviation, LAMP, forecast, aviation weather

Introduction

Civil aviation can be broadly divided into revenue-generating enterprises (e.g., air carrier and freight) and general aviation, the latter largely made up of light (<12,500 lbs) reciprocating engine-powered airplanes (Federal Aviation Administration [FAA], 2015a). Unfortunately, almost all (97%) of civil aviation accidents involve general aviation aircraft (Boyd, 2017) with weather-related mishaps constituting the largest fraction of fatal accidents (AOPA Air Safety Institute, 2017).

Operation of aircraft in instrument meteorological conditions (IMC) represents one of the biggest challenges to general aviation safety (i.e., in the absence of external visual cues) and even more so where ceilings are low (Fultz & Ashley, 2016).

http://dx.doi.org/10.7771/2159-6670.1230

Under such conditions a pilot must be able to operate the aircraft by sole reference to instruments. While instrument flight rule (IFR)-rated general aviation pilots are trained to do so, most (72%) general aviation pilots do not hold this rating (FAA, 2015b). For the latter pilots (as well as IFR-rated pilots who have not maintained their currency/proficiency), an IMC encounter often leads to spatial disorientation (Benson, 1999; Partmet & Ercoline, 2008) and subsequently to a loss of control (Partmet & Ercoline, 2008). Such mishaps carry up to a nine-fold elevated risk of a fatal outcome compared with those in visual conditions (Boyd, 2015; Grabowski et al., 2002; Li & Baker, 1999). Put another way, although only 9% of general aviation accidents occur in IMC, nevertheless they constitute 25% of fatalities (Li & Baker, 2007). Accordingly, FAA regulations prohibit pilots who are not IFR-rated (or not legally current) from operating in such conditions. Rather, flights under visual flight rules (VFR—defined (FAA, 2018), partly, by a ceiling in excess of 3,000 ft above the ground (AGL)) are recommended for such pilots. This altitude is partly predicated on man-made towers some in excess of 2,000 ft AGL.

It is noteworthy that IFR-certificated general aviation pilots are not immune to the risks associated with IMC. Degraded proficiency in flying by instruments, partly due to the infrequency of operating under such conditions (Weislogel, 1983), represents the major cause of fatal mishaps in this unforgiving environment (Shao et al., 2014). For such IFR-rated pilots, operations in low IFR (LIFR; ceiling $<$ 500 ft AGL) (FAA, 2018) conditions should be avoided.

To ascertain whether a flight can be safely undertaken with respect to the aforementioned weather and IFR certification or lack thereof, FAA regulations (Electronic Code of Federal Regulation, 2015) mandate that pilots perform a weather briefing prior to any flight away from the aerodrome vicinity. Towards this end, a battery of aviation-specific forecasting weather tools are available, e.g., surface analyses/synopses, and two airport-specific forecast tools: terminal aerodrome forecasts (TAFs) (FAA, 2010) and the Localized Aviation Model Output Statistics Program (LAMP) (Ghirardelli, 2015; NOAA National Weather Service, 2017). The former (embracing an area extending to five statute miles from the airport) are produced by National Weather Service (NWS) Weather Forecast Office meteorologists every six hours (FAA, 2010). In contrast hourly-issued LAMPs are entirely automated (Glahn et al., 2017). The merits of these two forecasting tools in the context of typical general aviation operations (commonly leisure flights of $<$100 nautical miles distance) warrant discussion. For such relatively short flights these aerodrome-specific forecasts for the departure/destination aerodromes as well as stations along the route of the flight (Vasquez, 2018) are considered of particular utility in the pre-flight weather briefing.

Unfortunately, of about 5,100 civilian aerodromes (herein "stations" is used synonymously) in the USA (FAA, 2017a), TAFs are issued for a minority (approximately 750) (FAA, 2010) with a bias towards airports (which also are the focus of greater weather forecasting research; Verlinden & Bright, 2017) serving mainly air carriers but less so for general aviation (FAA, 2017a). In contrast, the LAMP (Ghirardelli et al., 2015) represents a potential alternative to the TAF due to a wider availability at airports more frequented by non-revenue light aircraft. As of 2018, 1,853 stations provide LAMP forecasts (National Weather Service, 2018). LAMP represents an automated forecast hybridized from the current observation, three advective models, and the Global Forecast System Modeled Output System (GFS MOS) (Glahn *et al.*, 2017). However, at the present time, for aerodrome-specific forecasts, the FAA recommends that pilots (FAA, 2017b) only use LAMP forecasts in conjunction with TAFs. For airports for which a TAF is not issued, the TAF from the nearest airport is recommended for a pre-flight weather briefing.

Considering the substantially larger number of general aviation-frequented aerodromes issuing LAMP forecasts (compared with those issuing TAFs) the study herein was undertaken to determine the predictive accuracy of this forecast tool for VFR ($>$3,000 ft) and IFR (500–1000 ft) ceiling-based flight categories. Specifically, the accuracy of the LAMP generated at non-TAF-issuing aerodromes (hereafter referred to as satellite-LAMP) was compared with the TAF produced by the geographically proximate airport.

Methods

Selection of LAMP Aerodromes

Over a 12-month period (21 March 2019–26 March 2020), geographical areas of the contiguous USA favoring low cloud ceilings were identified daily using two approaches: (i) the Skew-T-derived lifting condensation level (LCL) (NOAA National Weather Service, 2018b) and (ii) the surface prognosis chart (NOAA National Weather Service, 2018a) as described elsewhere (Boyd & Guinn, 2019). Contoured areas corresponding to the 500 m LCL height, low-pressure, frontal systems and troughs (surface prognosis charts) guided the selection of areas for LAMP evaluation. LAMP-issuing aerodromes (National Weather Service, 2018) located in the aforementioned regions were then selected (by progressively sequencing through the alphabetized list of stations) for that day (one per state) with the proviso that a TAF was not issued for that airport but relied on a geographically proximate station-issued TAF (FAA, 2017b). LAMPs from airports satisfying this criterion are hereafter referred to as satellite-LAMP. The most proximate aerodrome-specific TAF was identified using an aviation subscription service: Foreflight^{R}.

Comparison of Forecast Tools

For a comparison of the satellite-LAMP forecast with the TAF generated for the nearest aerodrome, a block of time ("flight category block period"; Figure 1) was chosen daily for each satellite-LAMP station as follows. This evaluation period ("flight category block period"; Figure 1) constituted a time frame within the corresponding TAF valid period (NOAA National Weather Service, 2018a) starting at any TAF time element (e.g., FM, TEMPO group but excluding BCMG) and ending (Figure 1) at the subsequent time element the latter also within the TAF valid period (FAA, 2010). No priority was accorded to any specific time element. The fact that LAMPs and TAFs are issued hourly and every six hours, respectively, had two experimental ramifications. First, two satellite-LAMP forecasts (NOAA, 2018) were used to assess its accuracy: one generated (i) at the hour concurrent with the TAF issue time (satellite-LAMPsync) and the other (ii) one hour prior to the flight category block period (satellite-LAMP-1h; Figure 1). Second, the initial TAF time element defining the start of the "flight category block period" was always prior to the issue of subsequent TAF (six hours later). Weather forecasts and observations were collected once daily to include warm (April–September) and cool (October–March) periods (Ghirardelli & Glahn, 2010). Station selection for cool months was restricted to those employed in the warm month analysis.

Flight Categories and Forecast Tool Accuracy

Ceiling (broken or overcast) height was used to operationally define flight categories herein. VFR: >3,000 ft; marginal VFR (MVFR): >1,000–3,000 ft; IFR: 500–1,000 ft; LIFR: <500 ft (FAA, 2018) with all altitudes AGL. For each forecast method, the flight category assigned daily was based on the lowest ceiling over the aforementioned "flight category block period" (Figure 1). Forecast satellite-LAMP and TAF flight categories were verified using the lowest reported ASOS ceiling data for the aforementioned block (Iowa Environmental Mesonet, 2018). It should be noted that data collection was restricted to those for which a change in ceiling-based flight category at the TAF time elements defining the flight category block period was forecast (Figure 1).

For the VFR flight category, a forecast was considered "correct" when VFR conditions were forecast and VFR conditions were observed, while a forecast was considered "incorrect" (also referred to as a "miss") when VFR conditions were forecast and other than VFR conditions (MVFR/IFR/LIFR) were observed. If other than VFR conditions were forecast, and VFR was observed, this was considered a "false alarm." For the case of IFR conditions, a forecast was considered "correct" if IFR conditions were forecast and IFR or better (VFR/MVFR) conditions were observed. An incorrect forecast was recorded if LIFR conditions were observed. If LIFR conditions were forecast, and IFR or better (VFR/MVFR) conditions were observed, this was considered a "false alarm." The rationale for this approach is that a LIFR situation puts the general aviation instrument-rated pilot at risk while VFR/MVFR does not.

The fraction of correct forecast (range 0.0–1.0) represents the count of accurate forecasts (for either VFR or IFR conditions as described above) divided by the sum of correct and missed forecasts. Similarly, the fraction of false alarms was the count of VFR (or IFR) conditions which were not forecast divided by the sum of this value and count of correct forecasts. Alternatively, fractions were stated as the corresponding percentages.

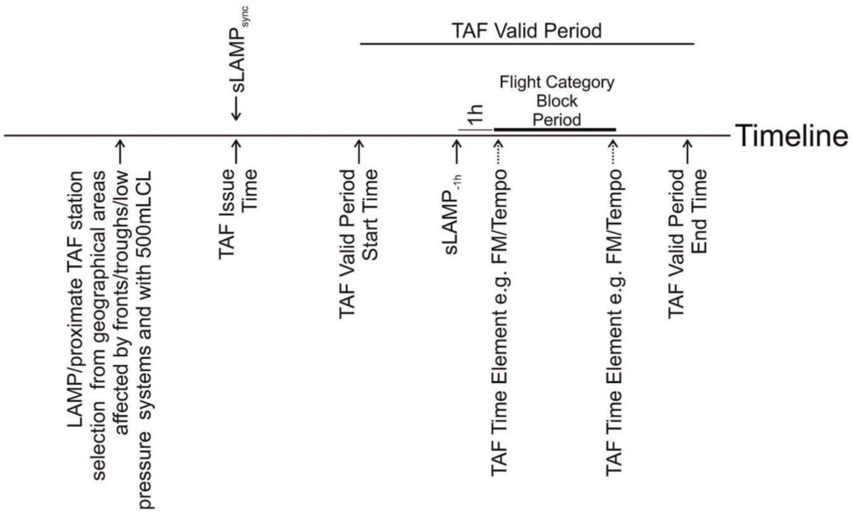

Figure 1. Experimental strategy: sLAMPsync and sLAMP-1h, satellite-LAMPs issued either concurrent with the TAF or one hour prior to the flight category block period, respectively. Time elements: FM, from; Tempo, temporary.

Statistical Analysis

Contingency tables (2 × 2 or 2 × 3) were used to test for differences in proportions using either two-sided Pearson chi-square or Fisher's exact tests (Field, 2009). Statistical analyses were undertaken with SPSS (v24) software.

Results

A total of 570 satellite-LAMP forecasts across 43 states were accrued over 218 days of the period spanning March 2019–March 2020. Of 570 satellite-LAMP forecasts, 260 and 310 were generated for the warm (April–September) and cool (October–March) periods, respectively.

VFR Forecast Accuracy by LAMP

The pre-flight weather briefing is crucial for the VFR-only pilot as to whether a flight can be undertaken safely, i.e., with ceilings >3,000 ft AGL (FAA, 2018), as >90% of accidents involving unintended IMC encounters have a fatal outcome (AOPA Air Safety Institute, 2017). Thus, we first compared the accuracy of the satellite-LAMP in forecasting the VFR flight category over the one-year study period with that of the TAF. Two different satellite-LAMP–TAF determinations for VFR forecasts were made: the satellite-LAMP issued (i) concurrent (rounding up to the next hour if within 15 minutes prior to the TAF issue time) with that of the TAF issue and (ii) one hour prior to the beginning of the evaluation period (flight category block period; Figure 1). Hereafter, these two forecasting tools are referred to as satellite-LAMPsync and satellite-LAMP-1h, respectively.

Regarding forecasting accuracy for VFR conditions, the satellite-LAMP issued concurrently with the TAF (satellite-LAMPsync) showed a correct fraction value of 0.53 (Figure 2). In proportion testing, this value was higher ($p = 0.002$) than the 0.35 for the corresponding TAF. Similarly, the satellite-LAMP issued one hour prior to the flight category block period (satellite-LAMP-1h) was also superior ($p < 0.001$) to the TAF in the percentage of correct forecasts for the ceiling-based VFR flight category (58 and 35%, respectively). Although the satellite-LAMP-1h correct fraction trended higher than the satellite-LAMPsync, this difference was not statistically significant ($p = 0.402$).

A corollary of the above findings is that the fraction of incorrect forecasts (miss fraction) in which VFR was forecast, but ceilings subsequently verified as <3,000 ft AGL, was lower for both satellite-LAMP relative to that for the TAF. This latter observation is important from a safety perspective for the non-IFR-rated pilot who may rapidly experience spatial disorientation upon an IMC encounter often with fatal consequence (AOPA Air Safety Institute, 2017).

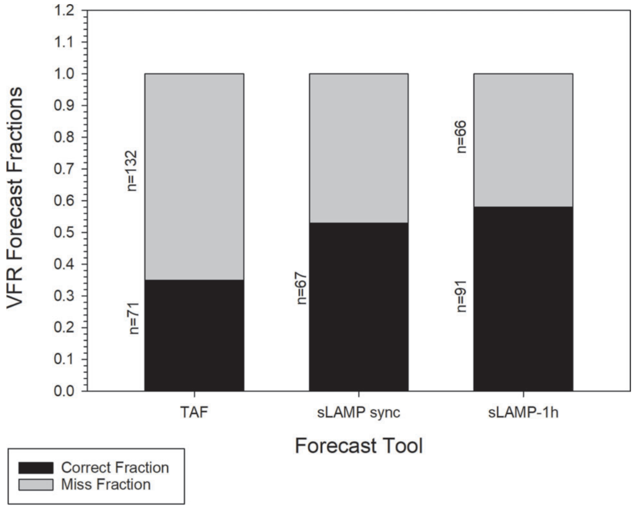

Figure 2. Accuracy of the VFR flight category forecast over the one-year study period. The accuracy of the various forecast instruments for the VFR flight category (ceiling >3,000 ft. AGL) is shown as fractions. The LAMP forecast was that generated either at the hour of the TAF issue (sLAMPsync) or 1 hour prior (sLAMP-1h) to the flight category block period per Figure 1. A "Miss" refers to a forecast in which verified ceilings were 3,000 ft. or lower. n, count of events. Statistical differences in proportions was undertaken using a Pearson Chi-Square (2-sided) testing of a 2 × 2 contingency table.

IFR Flight Category Accuracy of Satellite-LAMP Forecasts

Although IFR-rated general aviation pilots are trained to fly solely by reference to instruments, maintaining this skillset represents a continuing challenge for such pilots (Shao et al., 2014; Weislogel, 1983). Indeed, this lack of proficiency has been identified as a major cause of accidents involving IFR-rated general aviation pilots operating in degraded visibility (Shao et al., 2014). Consequently forecasts distinguishing between IFR (ceiling 500–1,000 ft AGL) and the more challenging LIFR (ceiling <500 ft AGL) flight categories are of importance for decision-making by these pilots as to whether, or not, an IMC flight should be undertaken.

Accordingly, the ability of the satellite-LAMP forecasts to segregate these two flight categories was determined over the 12-month study period. As described previously (Boyd & Guinn, 2019), since from a safety perspective an event in which a verified ceiling was higher than forecast (i.e., MVFR or VFR flight categories) presents no safety hazard, such forecast flight categories were binned with those which accurately predicted IFR and recorded as "correct." Using a similar strategy to that undertaken for VFR flight category forecasting, both the satellite-LAMPsync and the satellite-LAMP-1h were compared with the TAF.

Of 132 forecasts for IFR, the satellite-LAMPsync more often (Figure 3) accurately forecasted an IFR (or MVFR/VFR) flight category than the TAF (correct fractions 0.68 and 0.45, respectively), a difference which was significant ($p < 0.001$). Similarly, the satellite-LAMP-1h also proved

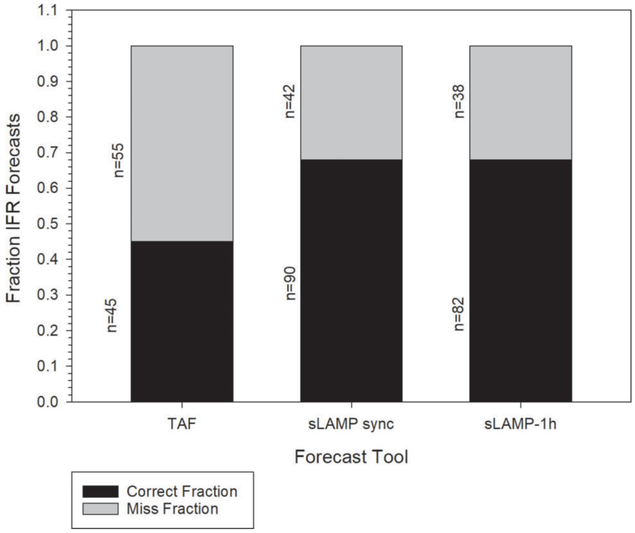

Figure 3. A comparison of the accuracy of the forecast tools for the IFR flight category over one year study period. The fraction of forecasts for IFR conditions which were correct and the corresponding miss fraction are shown. The LAMP forecast was that generated either at the hour of the TAF issue (sLAMPsync) or 1 hour prior (sLAMP-1h) to the flight category block period per Figure 1. A "Miss" refers to a forecast in which verified ceilings were lower than 500 ft. n, count of events. Statistical differences in proportions was undertaken using a Pearson Chi-Square (2-sided) testing of a 2 × 2 contingency table.

more efficacious ($p < 0.001$) in forecasting IFR (or MVFR/VFR) conditions (68% correct) using the TAF as referent. However, the satellite-LAMP-1h did not prove superior to the satellite-LAMPsync in IFR flight category forecasting as evidenced by their identical correct fraction scores (0.68). As a corollary, both satellite-LAMP measures showed disproportionately fewer (32%) "misses" (i.e., verified LIFR where IFR or better conditions were forecast compared with the TAF).

False Alarm Fraction for Satellite-LAMP and TAF forecasts

Whilst the aforementioned method used for evaluating LAMP accuracy is relevant to real-world operational decision-making by general aviation pilots, it suffers from one limitation. Specifically, it excludes weather events (VFR and IFR ceiling-based flight categories in the current study) which were not forecast but did occur (false alarms). In the mind of a pilot, any tool which excessively forecasts worse-than-actual conditions may undermine its credibility and ultimately lead pilots to disregard such forecasts (in common parlance: "crying wolf" too often). To address this shortcoming we used the fraction of false alarms to evaluate the satellite-LAMPs for forecasts of VFR and IFR ceiling-based flight categories again using the TAF as a comparator.

For VFR forecasts, the fraction of false alarms generated by the satellite-LAMPsync was not (Figure 4) statistically different ($p = 0.694$) from that of the corresponding TAF (0.43 and 0.40, respectively). In contrast, the percentage of false alarms for satellite-LAMP-1h was diminished ($p = 0.003$) relative to the TAF (22 and 40%, respectively). Regarding IFR forecasts (Figure 4), although the fraction of false alarms for the satellite-LAMPsync was elevated (0.18) compared with the TAF (0.08), this difference was not statistically significant ($p = 0.149$). In comparing satellite-LAMP-1h with the TAFs, the fractions of false alarms were comparable (0.11 and 0.08, respectively) and again no statistical difference noted in proportions ($p = 0.0771$). Thus, taken together, these data suggest that, compared with TAF forecasts, the satellite-LAMP does not predict worse-than-actual conditions in excess.

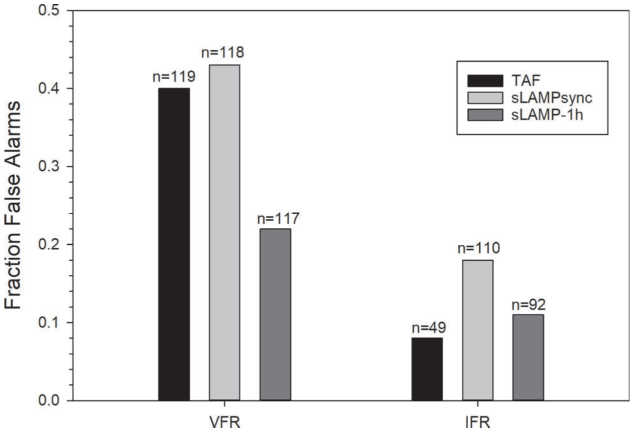

Figure 4. Fraction of false alarms for VFR and IFR flight category forecast for the one-year study period. The fraction of false alarms for the entire one year study duration are shown for VFR and IFR ceiling-based flight categories by the various forecast tools. Event n, total number of observations. Statistical differences in proportions were tested per Figure 2.

Critical Success Index Scores for Satellite-LAMP and TAF Forecasts

From a meteorological perspective, we also computed the critical success index (CSI) which embraces both forecast accuracy and false alarms (Schaefer, 1990). The satellite-LAMPsync and satellite-LAMP-1h both showed higher scores than the TAFs for VFR and IFR flight category forecasts. Thus, for VFR forecasts, the CSI scores for the satellite-LAMP-1h and satellite-LAMPsync were 0.50 and 0.83 (respectively), higher than the value (0.28) determined for the TAF. Similarly, for IFR forecasts, the CSI scores for the satellite-LAMP-1h and satellite-LAMPsync were superior to the TAF with scores of 0.63 and 0.59 (respectively), again improving on the value computed (0.43) for the corresponding TAF.

LAMP Forecasts for Warm and Cool Periods

Due to seasonal differences between warm and cool periods (e.g., ceilings related to convective versus non-convective

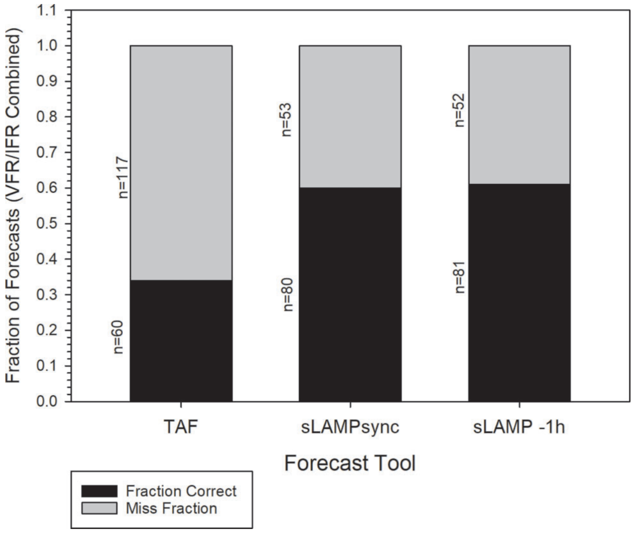

Figure 5. Forecast accuracy for the VFR/IFR flight categories over the cool months. The fraction of correct or missed forecasts for combined VFR-IFR flight categories (per Figures 2 and 3) collected over the Oct-Mar period were pooled and are shown. n, count. Statistical differences in proportions were undertaken using a Pearson Chi-Square (2-sided) testing of a 2 × 2 contingency table.

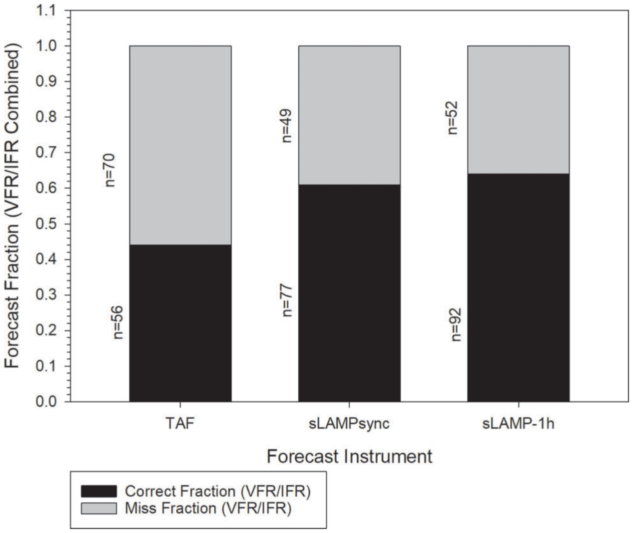

Figure 6. Comparison of forecast accuracy for the LAMP and TAF for the VFR/IFR flight categories over the warm months. The fractions corresponding to correct and missed forecasts for the VFR/IFR (combined) flight categories (per Figures 2 and 3) collected over the Apr-Sep period were pooled and are shown. n, count. Statistical differences in proportions were undertaken using a Pearson Chi-Square (2-sided) testing of a 2 × 2 contingency table.

clouds), we entertained the notion that a more robust satellite-LAMP forecasting accuracy in one period could offset an inefficacy for the other period. To address this possibility, satellite-LAMP forecasts for VFR and IFR flight categories were segregated into warm (April–September) and cool (October–March) periods (Ghirardelli & Glahn, 2010). Considering the limited count of IFR forecasts for the warm period, we combined VFR and IFR forecasts for each period.

For the cool months, the percentages of correct VFR/IFR-pooled forecasts (Figure 5) produced by both satellite-LAMPsync and satellite-LAMP-1h were higher (60 and 61%, respectively) than that for the TAF (34%). These differences were highly statistically significant (both $p < 0.001$). A similar analysis was performed for satellite-LAMP forecasts of VFR/IFR conditions for the warm months. Akin to observations made for the cool period, satellite-LAMP forecasts for the combined IFR/VFR flight categories proved more efficacious than the TAFs (Figure 6). Thus, the fractions of correct VFR/IFR forecasts determined for satellite-LAMPsync and satellite-LAMP-1h were 0.61 and 0.64, respectively, both statistically higher ($p = 0.011$ and $p = 0.001$, respectively) than that for the TAF (0.44).

Therefore, these data argue against the possibility that the efficacy of the satellite-LAMPs in forecasting VFR and IFR across the annual evaluation period is reflective of a bias in accuracy for either the cool or warm period.

Relative Satellite-LAMP–TAF Accuracy as a Function of Distance Between Stations

In our study, the distance between TAF and satellite-LAMP stations varied between 4 and 51 nautical miles (nmi) with a median of 24 nmi distance. We speculated that the VFR/IFR forecast accuracy of the satellite-LAMP, relative to that of the TAF, would improve as a function of distance between the two issuing stations. To determine if this was the case, the fraction of correct VFR and IFR flight category forecasts by the satellite-LAMP was determined for stations of either <23 nmi or >23 nmi separation distance from the TAF-issuing station. For increased statistical strength, satellite-LAMPsync and satellite-LAMP-1h data were aggregated.

The fraction of correct satellite-LAMPsync/-1h forecasts was superior ($p = 0.005$ or better) to the TAF in forecasting VFR and IFR flight categories (Table 1) irrespective of the distance between stations as determined in a 3 × 2 contingency table and controlling for distance. We then used the ratio of the correctly forecast conditions, generated by the satellite-LAMPsync/-1h:TAF ratio, as a measure of the efficacy of the satellite-LAMP for flight category forecasting based on distance between stations. Somewhat surprisingly, the efficacy of the combined satellite-LAMP for stations at greater distance (>23 nmi) was not superior to that for aerodromes more proximate to the TAF-issuing stations (satellite-LAMPsync/-1h:TAF ratios of 1.02 and 0.99 for VFR and IFR forecasts, respectively).

Discussion

The current study representing 12 months of data demonstrates the satellite-LAMP to be superior to the TAF issued at a proximate airport (4–51 miles distant in our study) in forecasting accuracy for VFR and IFR

Table 1
Flight category accuracy of forecast tools: dependence on distance between TAF and satellite-LAMP stations

Distance TAF–LAMP station (nmi)	Forecast instrument		Miss (n)	Correct (n)	Total	Aggregate satellite-LAMP sync/-1h fraction correct	Pearson chi-square *p* value	LAMPsync/-1h:TAF fraction correct (>23/≤23 nmi ratio)
				VFR				
<24	TAF	Count	65	40	105	0.38	0.005	102
		%	62	38	100			
	Combined LAMPsync/-1h	Count	62	81	143	0.57		
		%	43	57	100			
>23	TAF	Count	55	45	100	0.45	<0.001	
		%	55	45	100			
	Combined LAMPsync/-1h	Count	80	172	252	0.68		
		%	32	68	100			
				IFR				
<24	TAF	Count	30	22	52	0.42	0.011	0.99
		%	58	42	100			
	Combined LAMPsync/-1h	Count	43	77	120	0.64		
		%	36	64	100			
>23	TAF	Count	25	23	48	0.48	0.004	
		%	52	48	100			
	Combined LAMPsync/-1h	Count	37	95	132	0.72		
		%	28	72	100			

Note. Differences in proportions were tested in this 3 × 2 contingency table with distance as a constant. *n*, number of events; nmi, nautical miles.

ceiling-based flight categories. This finding is of particular importance since aerodromes frequented by general aviation pilots are more likely to be served by a LAMP than a TAF (FAA, 2010, 2017b). Currently, for such airports, the FAA recommends that pilots use the LAMP only in "combination with the TAF" issued for the geographically closest airport as the former "may not be as accurate as a forecast generated with human involvement" (FAA, 2017b).

Two relevant studies published prior merit discussion. While the first (Ghirardelli & Glahn, 2010) researched LAMP flight category forecast accuracy, it differed from ours in several respects. First and foremost, a single flight category (IFR and in that study the ceiling employed was any below 1,000 ft) was investigated, and not distinguishing between IFR and LIFR. Consequently, it did not address the operational needs of the (i) VFR-only pilot who should restrict flights to ceilings in excess of 3,000 ft (AGL) and (ii) the IFR-rated pilot, often lacking in instrument proficiency (FAA, 2018; Shao et al., 2014), and whom should eschew LIFR (<500 ft) operations (Weislogel, 1983). Additionally, the study (Ghirardelli & Glahn, 2010) was undertaken before redevelopment of the ceiling and sky cover algorithm in 2012 (Ghirardelli, 2015). Finally, by including data from U.S. stations (Ghirardelli & Glahn, 2010) in areas with low seasonal/ diurnal variability, verification data may have led to positive bias. While the second study (Boyd & Guinn, 2019) did not suffer these limitations, nevertheless, LAMP VFR/IFR forecast accuracy evaluations were restricted to stations issuing both TAF and LAMPs. Thus, the accuracy of satellite-LAMP at airports not issuing a TAF was not determined, an important practical limitation of that study. In addition, data were collected over a shorter period (six months) compared with the one year for the present study. Notwithstanding these differences, the prior study (Boyd & Guinn, 2019) demonstrated equal or superior efficacy of the LAMP in forecasting VFR and IFR flight categories for aerodromes which issued both forecasts.

The finding of minimal improvement in VFR/IFR flight category forecasts of the satellite-LAMP-1h compared with the satellite-LAMPsync was unexpected since weather observations by a station's ASOS contribute strongly to the LAMP forecast in earlier time frames (0–5 hours) (Ghirardelli & Glahn, 2010). We suspect this may be related, in part, to persistence in (ASOS) ceiling data over the intervening period between issue of the satellite-LAMPsync and satellite-LAMP-1h forecasts. Indeed, this supposition is in line with an earlier report documenting that persistence is a very competitive system in the very short term in forecasting ceiling height (Dallavale & Dagostaro, 1995).

We found little evidence of distance-dependent improved efficacy of satellite-LAMP (relative to the TAF) in forecasting flight categories. This observation was initially surprising as we had hypothesized that the former would hold an advantage with increasing distance. We suspect however that the lack of distance-dependent improved satellite-LAMP efficacy reflects limitations of the research design. A superior approach would have been to anchor various satellite-LAMP-issuing aerodromes to a single proximate TAF-issuing station and comparing the corresponding forecasts. However, such an approach was not undertaken for two reasons: (i) a greater distance from a TAF station may now put the satellite-LAMP-issuing aerodrome closer in distance to a separate TAF-issuing station and (ii) this was not the primary objective of the current study.

An area of weakness of the satellite-LAMP was the fraction of forecasted false alarms which in one instance (IFR forecasts by the satellite-LAMPsync) trended higher than that for the TAF. This is of some concern since false calls (the proverbial "crying wolf") have the potential to habituate pilots to such forecasts with a consequence of diminished safety.

Why would entirely automated satellite-LAMP observations be superior (based on the fraction of correct forecasts) in ceiling-based flight category forecasting compared with a trained meteorologist who may draw on several sources of weather data (including the LAMP) as well as experience? One possibility is that a TAF from a proximate airport may not address any local geographical features (e.g., body of water, terrain) affecting weather at the LAMP-issuing aerodrome. Also, and as described elsewhere (Boyd & Guinn, 2019), it is at the sole discretion of each NWS Weather Forecast Office meteorologist as to whether he/she employs LAMP data to generate the TAF. Since the LAMP forecast tool is relatively new (Ghirardelli et al., 2015) and validation studies few (Boyd & Guinn, 2019) it may be that NWS Weather Forecast Office meteorologists have been hesitant to make use of this tool. Also, since the geographical area covered by each NWS Weather Forecast Office is extensive (122 NWS offices cover the entire USA (FAA, 2010)), it is difficult to have expertise in the microscale environmental effects for all aerodrome locations.

The current study was not without limitations. First, in some analyses the number of events was low necessitating aggregation of some data. A second limitation was that since the study focused on geographical areas likely to experience marginal weather conditions on a given day based largely on synoptically driven features (LCL heights and/or frontal/trough regions), aerodromes affected by their own microclimates could have escaped evaluation. Nevertheless, we justify our strategy to avoid a positive bias associated with stations located in areas of low temporal (seasonal or diurnal) weather variability. Another potential shortcoming pertains to the omission of forecasts with the BCMG group. We elected to exclude these based on their gradual transitory nature (weather change over 1–2 hours) but realize that, as a consequence, we may have missed some observations.

The current study, together with prior published research (Boyd & Guinn, 2019), argues strongly for a greater integration of satellite-LAMP forecasts into the pre-flight weather briefing undertaken by general aviation pilots than currently advocated by the FAA (2017b). Presently, the FAA advises that LAMP data for a non-TAF-issuing aerodrome be used only to supplement the TAF issued at the nearest airport. Notwithstanding our recommendation, considering the imprecise nature of weather forecasting and that LAMP data (and TAFs) are for geographically discrete areas (i.e., an aerodrome), pilots should always avail themselves of all data applicable to a planned flight even for short-distance (<100 nmi) operations commonly undertaken by general aviation pilots.

References

AOPA Air Safety Institute. (2017). 27th Joseph T. Nall Report; General Aviation Accidents in 2015 (pp. 1−43).

Benson, A. J. (1999). Spatial disorientation—General aspects. In Ernsting, J. Nicholson, A. N., & Rainford D. J. (Eds.), *Aviation medicine* (pp. 419−454). Oxford, UK: Butterworth Heinemann.

Boyd, D. D. (2015). Causes and risk factors for fatal accidents in non-commercial twin engine piston general aviation aircraft. *Accident Analysis and Prevention*, *77*, 113−119.

Boyd, D. D. (2017). A review of general aviation safety (1984-2017). *Aerospace Medicine and Human Performance*, *88*, 657−664.

Boyd, D. D., & Guinn, T. (2019). Efficacy of the Localized Aviation MOS Program in ceiling flight category forecasts. *Atmosphere*, 10.

Dallavale, J. P., & Dagostaro, V. J. (1995). The accuracy of ceiling and visibility forecasts produced by the National Weather Service. *Sixth Conference on Aviation Weather Systems* (p. 213).

Electronic Code of Federal Regulation. (2015). *General operating and flight rules*. http://www.ecfr.gov/cgi-bin/text-idx?node=14:2.0.1.3.10

Federal Aviation Administration. (2010). Aviation Weather Services. AC 00-45G. 7-22-8-7.

Federal Aviation Administration. (2015a). General Aviation and Part 135 Activity Surveys. http://www.faa.gov/data_research/aviation_data_statistics/general_aviation.

Federal Aviation Administration. (2015b). U.S. Civil Airmen Statistics. https://www.faa.gov/data_research/aviation_data_statistics/civil_airmen_statistics/

Federal Aviation Administration. (2017a). Airport Data and Contact Information. https://www.faa.gov/airports/airport_safety/airportdata_5010/

Federal Aviation Administration. (2017b). Localized Aviation Model-Output Statistics (MOS) Program (LAMP), Weather Product. https://www.faa.gov/other_visit/aviation_industry/airline_operators/airline_safety/info/all_infos/media/2017/InFO17006 pdf.

Federal Aviation Administration. (2018). Safety of flight. Meteorology. In *Aeronautical information manual* (pp. 1-16–1-17). Washington, DC: Federal Aviation Administration.

Field, A. (2009). *Discovering statistics using IBM SPSS Statistics*. Thousand Oaks, CA: SAGE Publications.

Fultz, A. J., & Ashley, W. S. (2016). Fatal weather-related general aviation accidents in the United States. *Physical Geography*, *37*, 291−312.

Ghirardelli, J. E. (2015). Localized Aviation MOS Program (LAMP): A statistical post-processing system for the past, present, and future. https://www.weather.gov/mdl/lamp_presentations

Ghirardelli, J. E., Charba, J. P., Im, J.-S., Samplatsky, F. G., & Glahn, B. (2015). Improving Localized Aviation MOS Program (LAMP) guidance by utilizing emerging forecast and observation resources. *95th Annual AMS Meeting, Special Symposium on Mode-Post processing and Downscaling*.

Ghirardelli, J. E., & Glahn, B. (2010). The Meteorological Development Laboratory's aviation weather prediction system. *Forecasting*, *25*, 1027−1051.

Glahn, B., Schnapp, A. D., Ghirardelli, J. E., & Kim, J.-S. (2017). A LAMP-HRRR MELD for improved aviation guidance. *Weather and Forecasting*, *32*, 391−405.

Grabowski, J. G., Curriero, F. C., Baker, S. P., & Guohua, L. (2002). Exploratory spatial analysis of pilot fatality rates in general aviation crashes using geographic information systems. *American Journal of Epidemiology*, *155*, 398−405.

Iowa Environmental Mesonet. (2018). ASOS Network. https://mesonet.agron.iastate.edu/request/download. phtml?network=AZ_ASOS.

Li, G., & Baker, S. P. (1999). Correlates of pilot fatality in general aviation crashes. *Aviation, Space, and Environmental Medicine*, *70*, 305−309.

Li, G., & Baker, S. P. (2007). Crash risk in general aviation. *Journal of the American Medical Association*, *297*, 1596−1598.

National Weather Service. (2018). *Localized Aviation MOS Program*. https://www.weather.gov/mdl/lamp_stations_v2.1.0

NOAA. (2018). Localized Aviation MOS Product. http://www.nws.noaa.gov/mdl/gfslamp/meteoform php.

NOAA National Weather Service. (2017). Localized Aviation MOS Program. https://www.weather.gov/mdl/lamp_home

NOAA National Weather Service. (2018a). Aviation Weather Center. https://www.aviationweather gov/.

NOAA National Weather Service. (2018b). SPC Mesoscale Analysis. https://www.spc.noaa.gov/exper/mesoanalysis/new/viewsector.php?sector=19#.

Partmet, A. J., & Ercoline, W. R. (2008). Spatial orientation in flight. In Davis, J. R. Johnson, R. Stepanek, J. Fogarty J. A. (Eds.), *Fundamentals of aerospace medicine* (pp. 143−205). Philadelphia, PA: Wolters Kluwer.

Schaefer, J. T. (1990). The critical success index as an indicator of warning skill. *Weather and Forecasting*, *5*, 570–575.

Shao, B. S., Guindani, M., & Boyd, D. D. (2014). Fatal accident rates for instrument-rated private pilots. *Aviation, Space, and Environmental Medicine*, *85*, 631−637.

Vasquez, T. (2018). Preflight briefings. *IFR*, 19−21.

Verlinden, K. L., & Bright, D. R. (2017). Using the second-generation GEFS reforecasts to predict ceiling, visibility and aviation flight categories. *Weather and Forecasting*, *32*, 1765−1780.

Weislogel, G. S. (1983). Study to determine the IFR operational profile and problems of the general aviation single pilot. National Aeronautics and Space Administration NASA-CR-3576, NAS 1.26:3576.