A Lead-Time Metric for Assessing Skill in Forecasting the Onset of IFR Conditions

Andrew Loughe
Sean Madine
Jennifer Mahoney

NOAA/ESRL/GSD

September 2008

1

I will be describing a forecast lead-time metric which focuses on that part of a Terminal Aerodrome Forecast that warns of the potential for Instrument Flight Rules (IFR) from low ceiling and visibility forecast elements.

This metric has been developed by NOAA's Earth System Research Laboratory with funds provided by the National Weather Service's Aviation Services Branch in support of its continued efforts to improve the lead time in forecasting the onset of IFR conditions for aviation.

Definitions

TAF is… a concise statement of forecast conditions at an airport, *usually* out to 24 or 30h. TAFs contain information on ceiling, visibility, winds, and weather that can affect flight safety. *Issued: 00,06,12,18Z*

VFR – Visual Flight Rules +

IFR – Instrument Flight Rules –

2

National Weather Service TAFs are issued at over 600 airports. These text products, which are similar in format to a METAR report, contain a concise statement of forecast conditions usually out to 24 or even 30 hours.

TAFs are routinely issued at 00, 06, 12, and 18 UTC, and there is provision for them to be amended, corrected and then broadcast as needed. TAFs clearly report expected weather conditions over short time segments within the full valid period of the overall forecast. There are hourly and special METARs co-located with these forecasts, and these serve as a basis for verifying the TAFs.

For the development of this particular lead-time metric, we concentrate on the ceiling and visibility aspects of the forecasts and observations, and highlight those conditions which fall into the range of Instrument Flight Rules for aviation (IFR). For the statistical web tool, the weather is viewed as being in an IFR state (ceiling below 1,000 ft, or visibility below 3 miles), or in a more favorable, non-IFR state which we loosely refer to here as Visual Flight Rules (VFR). This is a binary approach to the problem, with no gradations beyond IFR and more favorable than IFR, which we call VFR.

Challenges to Overcome with TAFs

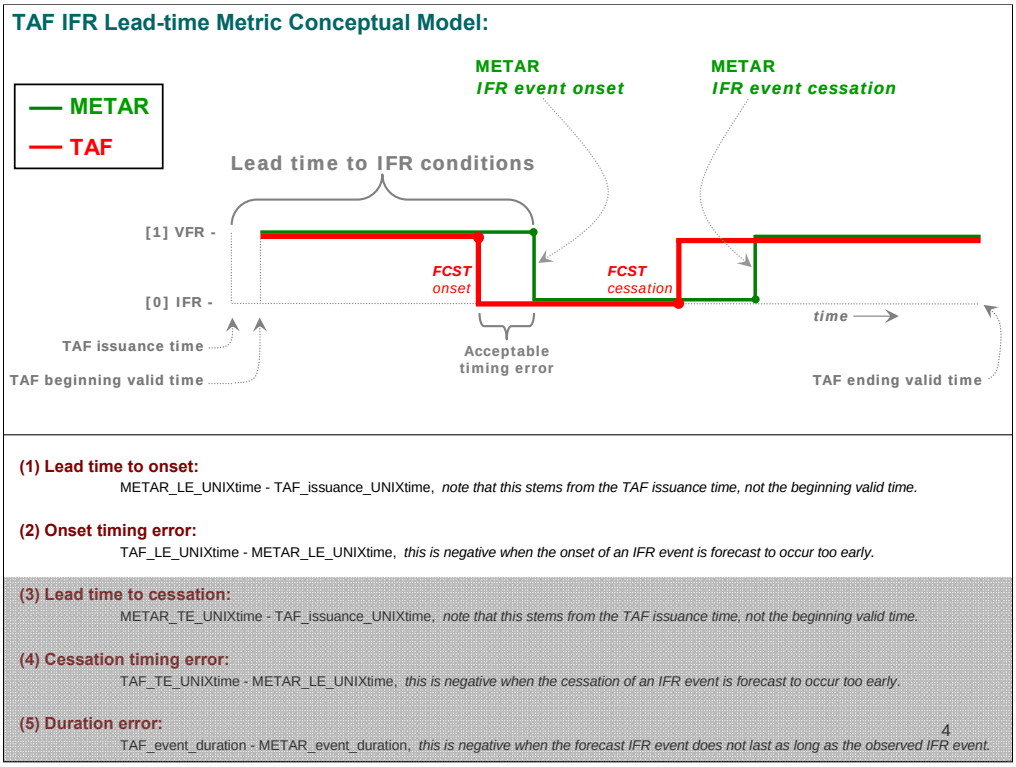
- Decoding TAFs and METARs:
  - Extracting flight rules information
  - Reducing TAFs toward the worst expected weather
  - Associating **F**/**O** IFR events with each another
- Handling:
  - Amended (AMD) and corrected (COR) TAFs
  - Greater fluctuation in the observed IFR event signal
  - Multiplicity of *hits* on observed IFR events
- Computing meaningful statistics by region, season, weather category, etc.

3

In-house, we decode National Weather Service (NWS) TAFs and extract data from accompanying METARs to obtain forecast and observed flight rules information. Using a relational database, we then store the information necessary to associate each observed IFR event with its forecast counterpart.

One important aspect of decoding TAFs is the reduction of expected weather to the worst expected condition for a given time segment. For example, if a forecaster believes that the weather will be poor TEMPORARILY, or that there is a PROBABILITY of poor weather that will cause conditions to fall within the IFR range, then we indicate this in our forecast database, and we strictly ignore the likelihood that the weather might be more favorable than this. For a given time segment, we reduce all expected weather to the worst possible weather condition stated in the forecast.

In this presentation we will also illustrate some of the major differences between TAF and METAR reporting that makes it challenging to associate the two– most importantly, the high rate of fluctuation between VFR and IFR conditions evident in many automated METAR reports, and the relatively low fluctuation rate found in nearly all Terminal Aerodrome Forecasts. Also, since TAFs are issued numerous times throughout the day, especially when amendments are required, we frequently experience a situation where many forecasts record a “hit” on a single, observed IFR event. Which of these forecasts is to be counted in the timing of the earliest forecast hit? We will explore at least two possible answers to this question.



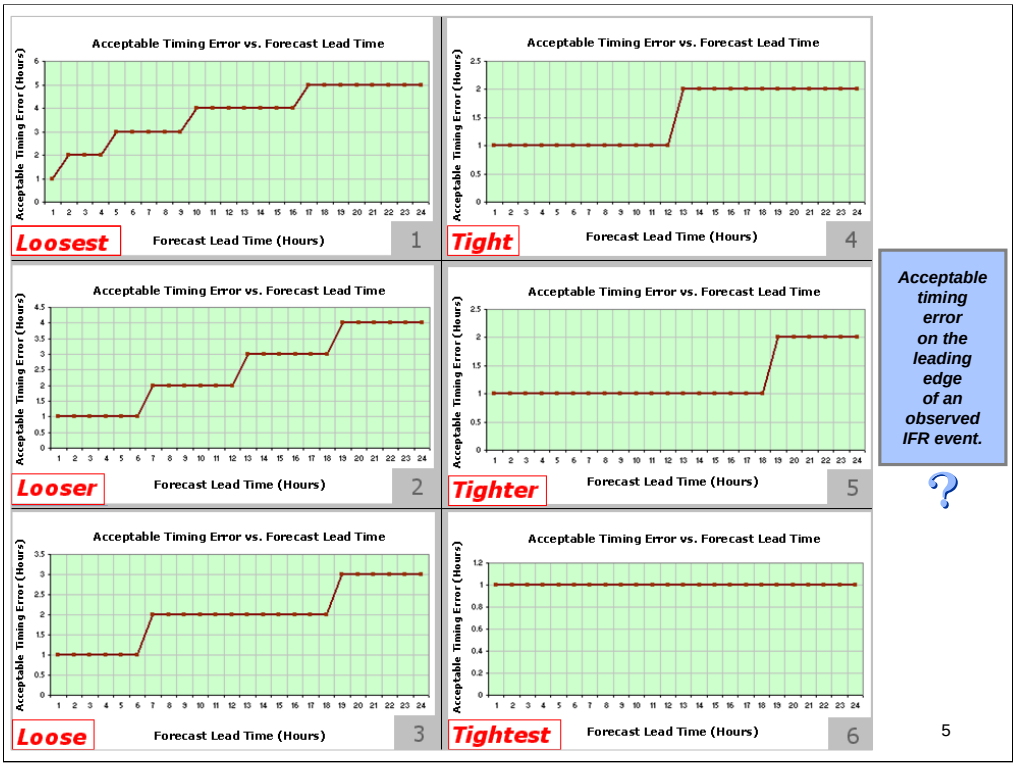
This timeline illustrates the mechanism used to compute a lead-time in forecasting IFR events.

An IFR “event” is bounded by vertical bars within this sample transition plot. You see an observed time series in green, and a forecast time series in red. Transitions from VFR to IFR are marked by a vertical bar that descends from the value of one to zero along the y-axis. The vertical bar which ascends from the value of zero to one marks the cessation of the observed or forecast IFR event.

Looking at the onset of the observed event, we see a timing error which represents the discrepancy between when an event was observed to occur, and when it was forecast to occur.

The lead time is measured from the onset of observed IFR conditions back in time to the issuance of the forecast that records a hit on the observed event, within the acceptable timing error. If no such forecast correctly predicts the event to occur, then we record a miss on the observed event.

The current web tool focuses its attention on the onset of IFR conditions, and an enhanced tool could also focus its attention on the cessation of such events. Additionally, one could study the duration aspects of observed and forecast IFR events. The statistical web tool that we have developed does allow one to compare duration attributes from the forecasts and observations.



The web tool that has been developed for the IFR lead time metric allows one to choose which timing error is acceptable when associating a forecast with an observed event to record a hit. In these timing error plots, acceptable timing error is represented along the y-axis, and forecast lead time is shown along the x-axis.

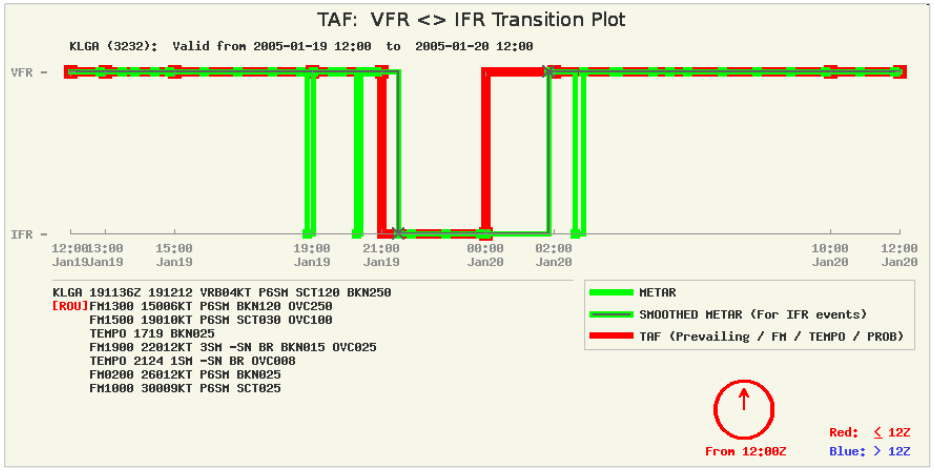
The “loose” case, number 3 in the lower left-hand corner, records a hit for any forecast onset that is within one hour of the observed event onset. For a forecast lead time beyond 6 hours, a 2 hour timing error would also be acceptable, and beyond an 18 hour lead time, this particular timing function (option #3) allows a forecast with a 3-hour timing error (or less) to register a hit on an observed event. Any timing error beyond 3 hours, regardless of forecast lead time, would not register a hit on the observed IFR event.

It will be of interest to users of the statistical tool, to vary parameters such as this one to see how it affects the hits and misses on observed events.

One could imagine that the hit rate on these events will rise with the loosest hit criteria for the timing (option #1 in the upper left-hand corner), and the hit rate will fall dramatically if one employs the tightest criteria (option #6 in the lower right-hand corner) which requires all forecast onsets to be within one hour of the observed onset, regardless of the lead time in forecasting IFR conditions.

METAR Multiplicity

The variability in METAR IFR events is much greater than for TAFs



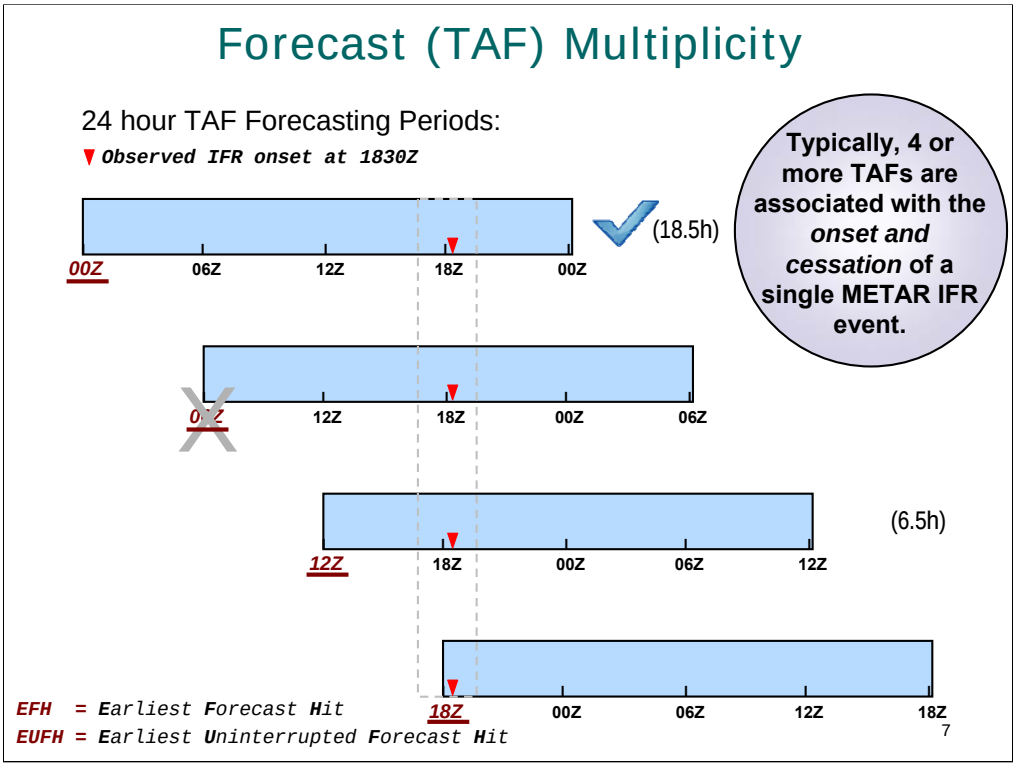
6

Many automated METAR sites exhibit a high degree of fluctuation into and out of IFR conditions throughout the course of the day. Unfortunately, we have no other observational source of IFR conditions, and so we strive, in developing the web tool, to make use of an observational quantity (the METAR shown here in green) whose characteristics differ greatly from their predictive counterpart (the TAF, shown here in red). In this figure, we see many observed IFR events associated with a single forecast event.

TAFs are not designed to predict fluctuations on such small, temporal scales as recorded by many automated METAR sites, and so we apply a simple smoother to the highly fluctuating observational signals to more closely align them with the more slowly varying (in time) TAF signals.

This smoother, at present, is rather simple. It ignores any observed IFR event that lasts less than 30 minutes in length, as can be seen in the diagram above. The actual, reported transitions from VFR to IFR and back are illustrated in this figure by the green line, while the smoothed METAR is illustrated by the green line with embedded black line.

The time of individual METAR reports is noted by thick green dots on the figure, and thick red dots mark end points in time of FROM, TEMPO, and PROB groups within a TAF that has been reduced to its worst weather condition.

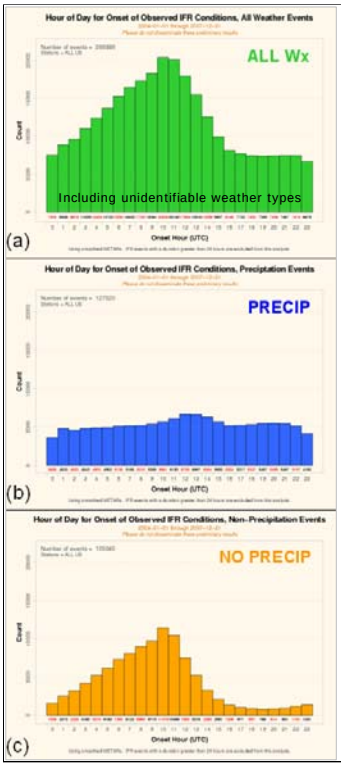


We have just described what we call “METAR multiplicity”, the fact that many short-lived observed IFR events can be associated with a single forecast event.

Now we will describe what we call forecast (TAF) multiplicity.

Assume that you have a single observed IFR event that begins at 1830 UTC. In this first illustration, forecasters issuing routine TAFs out to 24 hours might register a hit on the observed event from the 00Z forecaster, a hit from the 06Z forecaster, the 12Z forecaster, and even the 18Z forecaster– only 30 minutes prior to the beginning of the event. In our database we record either the Earliest Forecast Hit (EFH, 18.5 hours in this case), or the Earliest Uninterrupted Forecast Hit. The Earliest Uninterrupted Forecast Hit (EUFH) is the earliest forecast that was associated with the onset of an observed event, while all subsequent forecasts continued to properly record a hit on the IFR event. In an EUFH context, you will not get credit for a long lead time forecast on an event if you subsequently and incorrectly retreat from forecasting the event to occur.

In a second example shown here by animation, the Earliest Forecast Hit (EFH) is 18.5 hours, but the Earliest Uninterrupted Forecast Hit (EUFH) is only 6.5 hours, because (as shown in the animation) the 06Z forecaster has now incorrectly indicated that no IFR event will occur near the time of 1830UTC.



Observed (METAR) IFR Event Climatology 2004-2007

- Climatology reveals a weak diurnal signal for IFR events accompanied by **rain or snow**, and…
- A strong diurnal signal for **non-precip** (fog/stratus) events.

8

Prior to computing numerous hit statistics for forecasting IFR events, it is helpful to understand the climatology of these events.

The overall climatology shown here in green (figure a) contains information from all weather categories, including those weather types that are unidentifiable because present weather is not always reported in a METAR.

The distribution of onset hour throughout the day reveals a broad and flat distribution of events (the base, in the top figure), along with a pronounced peak near mid-day UTC.

From figure (b) it can be seen that the broad –based distribution from all weather events (in the top figure) is largely associated with those IFR events that occur during periods of precipitation. Figure (c) reveals that the peak in the top figure is largely associated with non-precipitation events– those IFR events that occur due to fog or stratus.

If you study figure (c), it becomes readily apparent that on average throughout the time period, a forecaster working the desk at, say 6 UTC is **most likely** to register a lead time on IFR events that is no greater than about 4-6 hours, whereas a forecaster working the desk at, say 12 or 18 UTC is **more** likely to register long lead times if he or she correctly forecasts an event to occur at the climatological peak, which is around 8 – 12 UTC, into the next day.

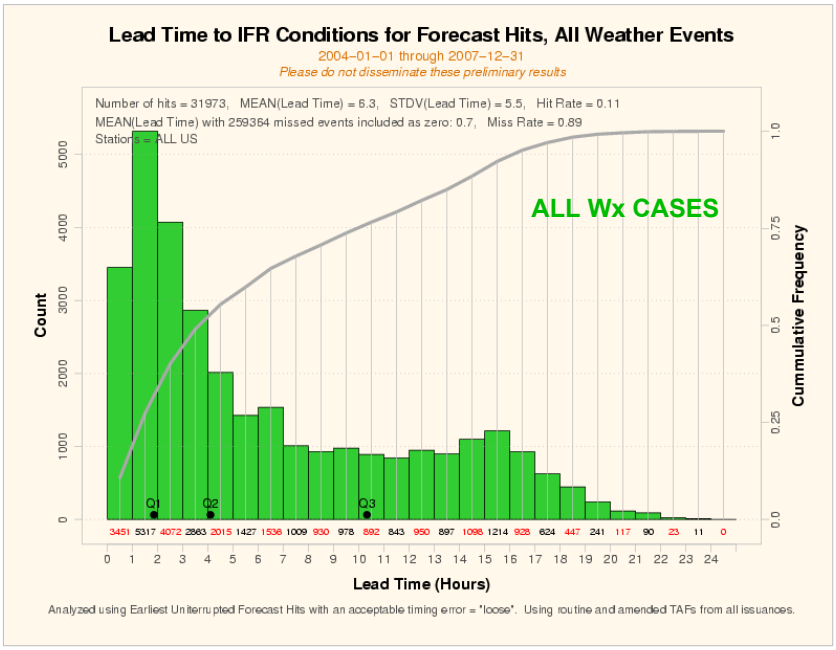
Lead-Time Distributions Stratified by Weather Type and Forecast Issuance Hour

9

Now we will discuss some actual results from the statistical web tool.

All of the results shown here, and even those climatological results we just discussed, come directly from the web tool.

Lead Time to IFR Onset for all Weather Categories



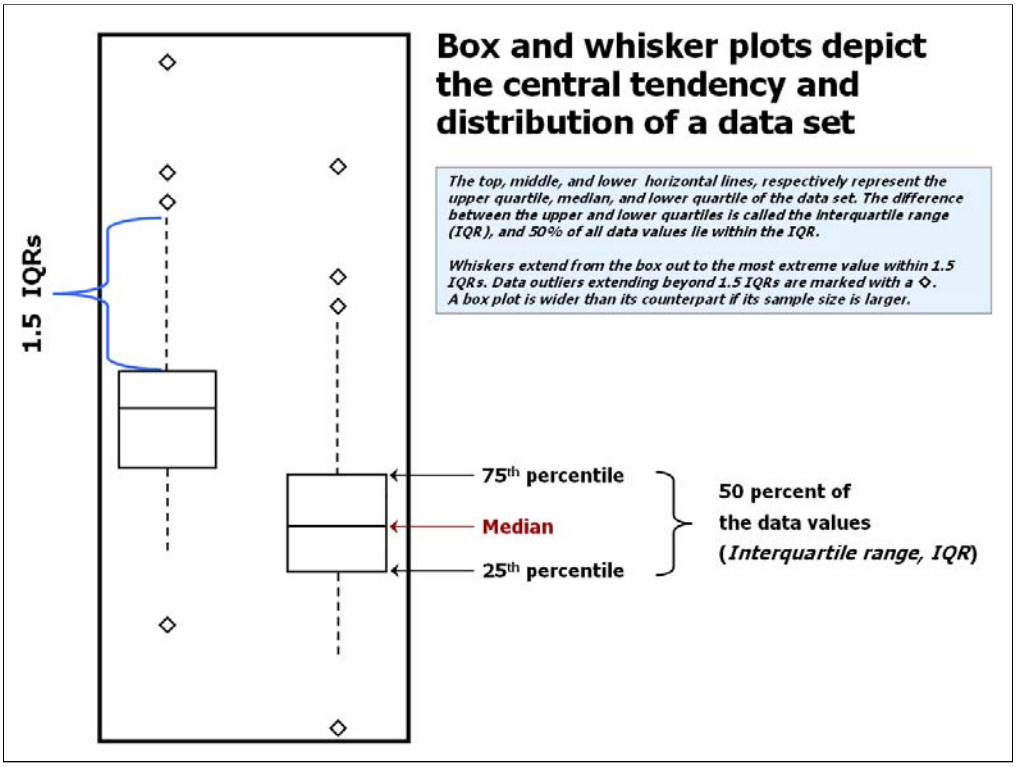
10

In this histogram, forecast hits are tallied in each bin along the x-axis. The x-axis represents lead time (forecast length) ranging from 0 to 24 hours. For this particular figure, hits were computed from the context of an Earliest Uninterrupted Forecast Hit, using “loose” matching criteria for the acceptable timing error. All TAF types were included (ROU and AMD), and the results have been tallied for TAFs that were issued every hour throughout the day, not just the routine issuance hours of 00, 06, 12, or 18 UTC, which is an option in the web tool.

This figure shows that the mean lead time for all forecasts that register EUFH hits on observed IFR events during the four year time period covering 2004-2007 is 6.3 hours and the standard deviation is 5.5 hours. The median lead time (Q2) is 4.1 hours,. The first and third quartiles are also shown on the histogram along the x-axis (Q1, Q3).

If you were to include all 259,364 missed IFR events as contributing a zero-hour lead time, and then re-compute the mean lead time with all of these zeros included, you would find that the mean lead time drops from 6.3 hours down to a sobering 0.7 hours, as shown in the plot title.

If you re-computed this entire diagram from the perspective of an Earliest Forecast Hit (EFH), the mean lead time would increase beyond 6.3 hours.



I will now briefly describe a box and whisker plot.

A box plot depicts the central tendency and distribution of a data set.

The top, middle, and lower horizontal lines of each box represents respectively, the upper quartile, median, and the lower quartile of a data set.

75% of all data values fall below the upper quartile (Q3 from the histogram).

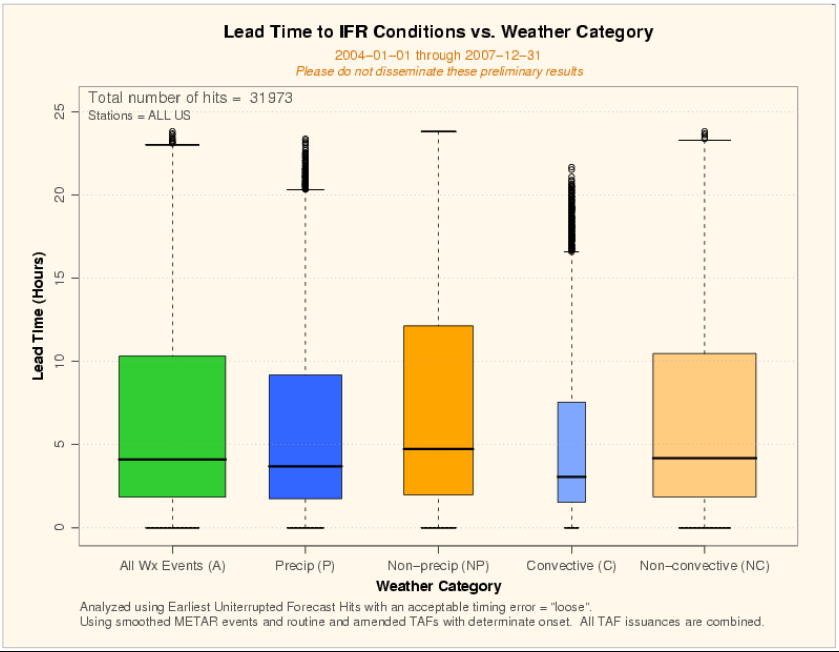
50% of all data values fall below the median (Q2 from the histogram).

25% of all data values fall below the lower quartile (Q1 from the histogram).

The range from 25% to 75% is called the inter-quartile range (IQR), where 50% of all data values are present.

Whiskers extend above and below the box to extreme values which extend out to 1.5 times the inter-quartile range. The most extreme values from the distribution are marked with a small circle or small diamond shape.

The Variation in Lead Time to IFR Onset by Weather Category



12

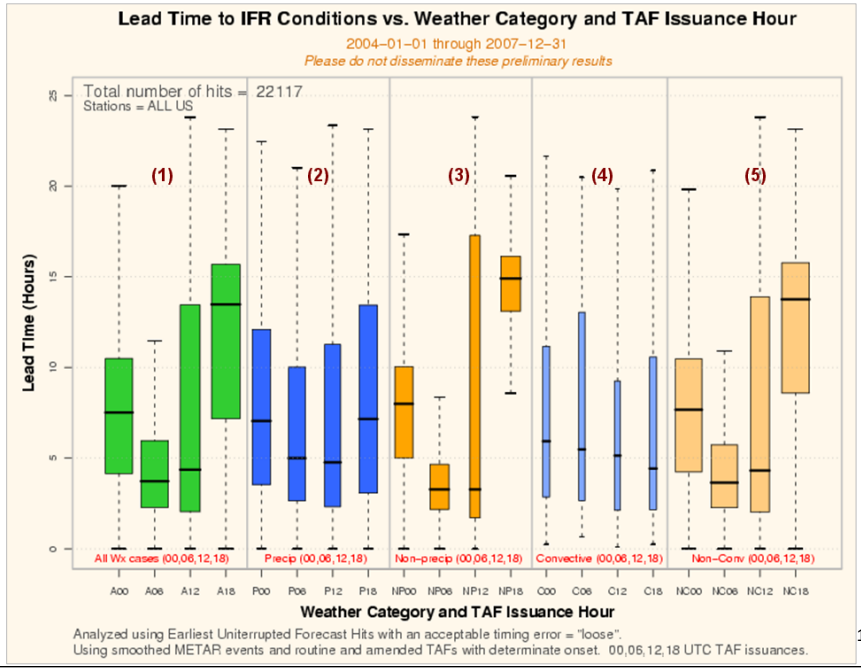
These box and whisker plots shows the distribution of forecast IFR event lead time in units of hours along the y-axis.

Color represents different weather types. The dark horizontal lines show the position of median values for the forecast lead time. The width of each box provides an indication of the relative number of events included in each analysis. Note that many METAR reports do not specify any present weather so, for example, the number of PRECIP + the number of NON-PRECIP events does not usually equal the total number of weather events because of this non-reporting problem. For this work, convection is recorded whenever thunderstorms or funnel clouds are reported in a METAR.

The plot titles indicate that this analysis is for all U.S. stations (over 600) covering the time period 2004-2007. Only TAFs with determinate IFR event onset time were included (some forecasts indicate IFR conditions from the very beginning of the forecast period, so it is not possible to determine the precise time the conditions were forecast to transition from VFR to IFR– the timing of the transition is indeterminate). Of the 31,973 TAFs that recorded “hits” over this time period (for all weather cases), the median lead time is approximately 4.1 hours. For these results, all TAF issuance hours are combined together.

Notice that when all issuance hours are combined together, as they are here, there is very little variation in forecast lead time with weather type. If however, you repeat this analysis, and stratify results by TAF issuance hour, a different interpretation is revealed…

Summary of the Variation in Lead Time by Weather Category and Forecast Issuance Hour



13

This series of box plots show the distribution of forecast IFR event lead time in units of hours along the y-axis. Color represents different weather types: (1) from all weather events in green; (2) all precipitation events in blue; (3) all non-precipitation (fog & stratus) events in dark orange; (4) convective events in light blue; and (5) non-convective events in light orange. Different TAF issuance hours for each group are indicated along the x-axis. The dark horizontal lines show the position of median values for the forecast lead time. The width of each box provides an indication of the relative number of events included in each analysis. The plot titles indicate that this analysis is for all U.S. stations (over 600) covering the time period 2004-2007. Only TAFs with determinate IFR event onset time were included. This plot shows how the distribution of forecast lead time varies by weather type <u>and</u> by TAF issuance hour.

If you look at the second group for precipitation events (in blue), you see that there is little variation in lead time for IFR events. Most of the variation in lead time by issuance hour is actually due to variations apparent in the non-precipitation (fog and stratus) events (group 3 in dark orange). There is little variation in lead time for precipitation cases (group 2) for the reasons noted when analyzing the climatological results: since there is little variation in the climatological occurrence of precipitation events with time-of-day, one expects little variation in forecast lead time with issuance hour for these cases. Similarly, we see much less variation of lead-time with issuance hour for convective cases (group 4 in light blue) versus that which is apparent for non-convective cases (group 5 in light orange).

Lead-Time Distributions by Station

14

Although the lead time metric statistical tool is best-suited for compiling aggregate statistics for a large number of stations, one can also use the web tool to study differences between climatological occurrences and hit rates at individual stations.

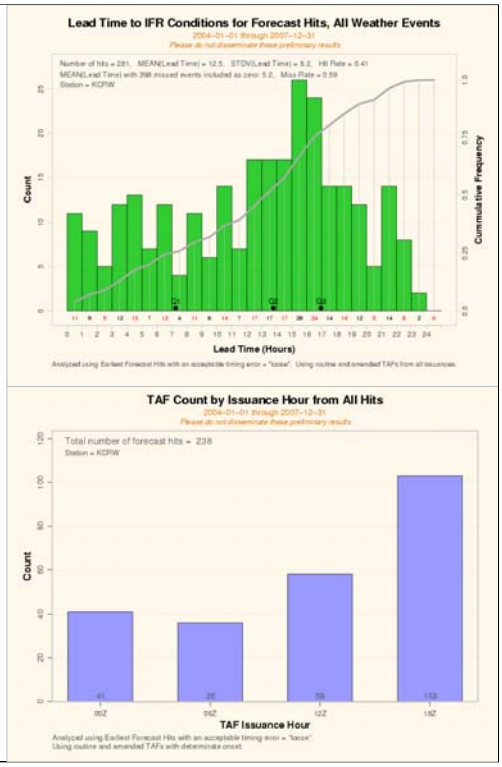
KCRW

(1) TAF count by issuance hour

(2) Lead time to IFR onset

EFH context

2004-2007



In this figure we see the distribution of lead times for hits at Charleston, WV from 2004-2007 in the top figure, and in the bottom figure, the TAF count by issuance hour for these hits. The TAF count includes only those hits from issuances at 00, 06, 12, 18 UTC.

The next slide presents this same information for San Francisco, CA. By comparing these results side-by-side, or by flipping back-and-forth between these two slides, one can see a dramatic difference between results at the two stations with regard to the lead time characteristics from the top figures, and the distribution of TAF count by issuance in the bottom figures.

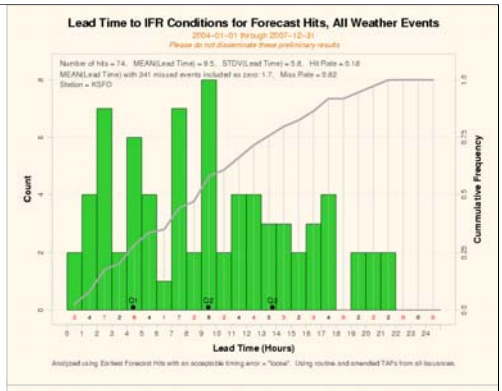
KSFO

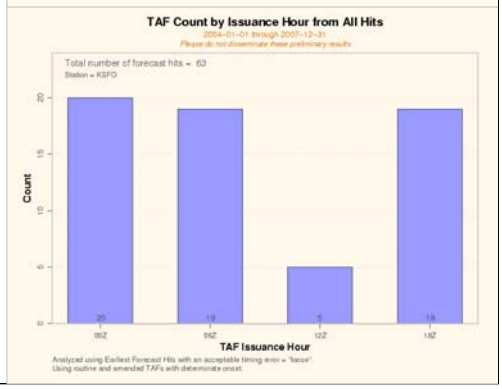
(1) TAF count by issuance hour

(2) Lead time to IFR onset

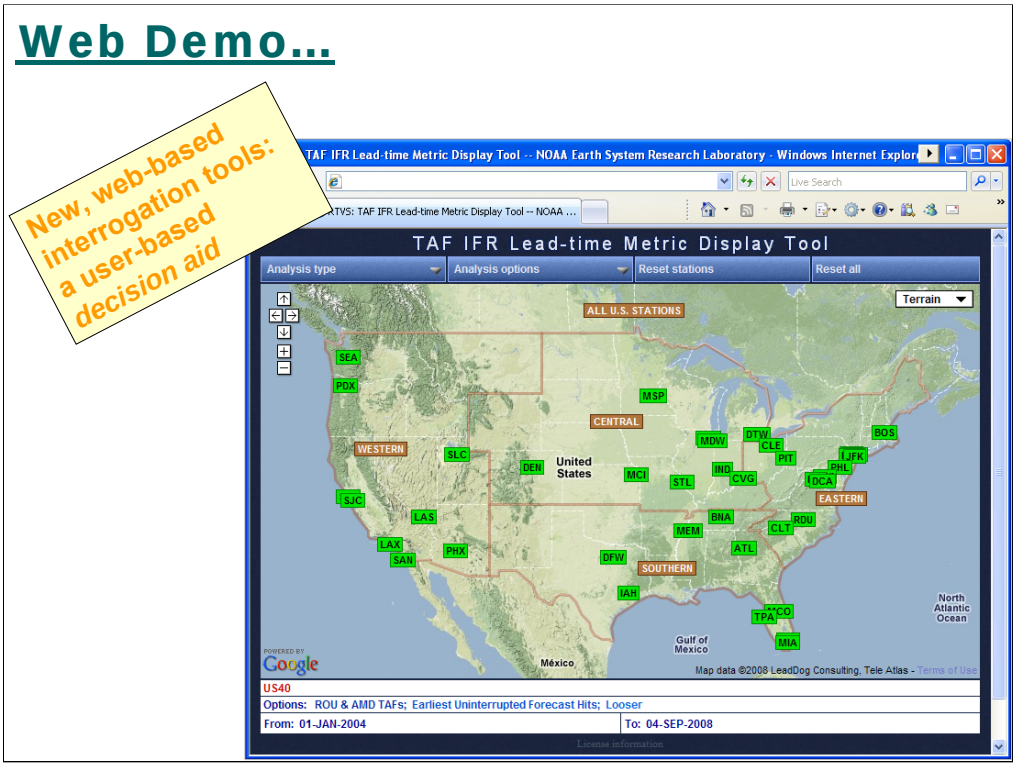
EFH context

2004-2007





The next slide presents this same information for San Francisco, CA. By comparing these results side-by-side, or by flipping back-and-forth between these two slides, one can see a dramatic difference between results at the two stations with regard to the lead time characteristics from the top figures, and the distribution of TAF count by issuance in the bottom figures.



This concludes our overview of the TAF lead-time metric. It will be insightful to continue along by viewing other demonstration videos which further elucidate the types of analyses that can be performed by accessing the web-based statistical tool that has been developed for your benefit.