Weather to Fly

Fall 2004 | National Weather Service | Volume 1, Number 4

WFO Medford, OR, Tries New Aviation Outreach Option

By James Reynolds, WCM, Medford, OR
James.Reynolds@noaa.gov

Like any other Warning Coordination Meteorologist, I am constantly on the lookout for new outreach opportunities, particularly for the aviation program.

Because of the relatively small population in our forecast area, efforts to make inroads on aviation outreach have been especially challenging.

For our offices, visits to local airports staff have been particularly helpful. Many of these airport staff have never been out to a WFO or have been not been out recently.

Earlier this year, NWS Medford Aviation Focal Point Robert Cramp and I conducted a familiarization and outreach trip to the Roseburg and Grants Pass, OR, airports.

Our visit to the Roseburg Airport was especially relevant because we may begin issuing a Terminal Aerodrome Forecast (TAF) for the site.

Because of this new product, we needed information regarding local topography and the airport layout. This information would later be used to prepare aviation forecasters to issue the product for these sites.

We began our outreach effort by visiting Adventure Aviation at the Roseburg Airport,

Figure 1. *Roseburg Airport: Looking west-southwest near railroad tracks that run parallel to the eastern perimeter of the field.*

In this Issue:

Weather to Fly

Managing Editor: Michael Graf
Editor/Layout: Melody Magnus
Melody.Magnus@noaa.gov

Figure 2. *Roseburg Airport: Looking north from a small bluff at the south end of the field.*

where we talked with a flight instructor and the office assistant.

We asked them what weather information they regularly obtained for flight planning purposes, how they obtained it and to what extent they used the Medford NWS website.

They, in turn, asked us if tours of the NWS office were available. We strongly encouraged them to call the office to schedule one.

They also revealed a potential error in our data of which we were unaware. Some pilots believed the nearby Sexton Summit Automated Surface Observing System (ASOS) was reporting pressure inaccurately.

Next, we visited Ameristar Air Center Incorporated. We spoke with the president of the company who was already a fan of the NWS Medford website. He told us he and other pilots used the page extensively.

We then completed a survey of the airfield. We circled the airport to develop a mental picture of the topography surrounding the airport and took some photos from different perspectives around the airfield (photos 1 and 2). We also made note of the volume of air traffic during our visit.

On our way back to Medford, we visited the Grants Pass Airport. We were able to visit Great Pacific Aviation Services a Fixed Base Operator on the field.

Here we spoke with a couple of flight instructors about what weather services their company uses for flight planning purposes. This company also used the Medford NWS website extensively.

After encouraging the flight instructors to come to our office for a tour, they echoed the same comments that we had heard in Roseburg about the Sexton Summit ASOS incorrectly reporting pressure.

In the end, this turned out to be an extremely educational trip. Not only did we make some valuable aviation contacts, but we were also able to develop mental pictures of the fields that we visited which should enhance future TAF forecasting.

Additionally, our customers made us aware of a perceived inaccuracy of an ASOS unit in the area. This problem was investigated and it was determined that no actual in accuracy existed. ✈

Writing Effective Terminal Aerodrome Forecasts (TAFs)

By Allan Fisher, Aviation Focal Point, WFO, Chicago, IL
Allan.Fisher@noaa.gov

The verification program developed by Andrew Rorke has been a huge help to me as an aviation focal point. It has allowed me to visually plot out how this station is writing its TAFs.

At Chicago, we write for the busiest aircraft operations area in the world, the combined traffic of O'Hare and Midway (MDW). Southwest Airlines, which uses MDW heavily, is a major user of our TAF.

With this in mind, I use the verification program for three reasons:

- To identify products key airport operations we can improve
- To compare models as a guide in preparing the TAF
- As background for the Chicago CWSU to know how the WFO forecaster is thinking and also for CWSU to use in its forecast preparation

One major area of concern for the airline industry is the NWS thunderstorms product (TS). TS reports are a major disputation to airport operations. Thunderstorms in the TAF also help determine "go/no go" decisions by general aviation pilots and require airlines to load extra fuel .

Most forecasters forecast an event, even if the chances of it occurring are low. Most forecasters are afraid of missing a storm so they forecast conditions a little lower just in case ceilings or visibilities drop more than expected.

The problem with this type of cautious forecasting in aviation is that every time a TS is included in the forecast, the airlines add extra fuel. A private pilot might decide not to fly.

The old rule of fuel usage is that for every extra unit of fuel added, 10 percent is burned up carrying it. So if we have TS in the forecast and nothing happens, 10 percent of that extra fuel is burned up regardless. At O'Hare, just 1 hour of false TS or false alternate weather, cost the airlines $100,000 per hour as of fall 2003. The cost now is much higher, somewhere around $140,000 per hour.

The airlines already carry extra fuel to handle unexpected situations and meet FAA mandates. Adding a TS to the TAF means even more fuel and more dollars.

A clear message from some of our customers at aviation conferences is that we reduce not only TS false hours, but also false hours of conditions requiring alternate airport filing and conditions below minimums.

Reducing false alarms has been a major goal at this office, and we are having success.

Figures 1-3 show how we are doing in the TS arena. The data show that the MAV was superior to us in forecasting TS, with much lower false hour counts. It also shows the MAV superior to the FWC guidance for TS. (See **Figure 1**, Thunderstorm False Alarm Hours.)

These numbers are only for the 0-12 hour time, for the 12-24 hour time the FWC number comparison is even worse.

This year, as you can see on the MAVFTS012 chart, **Figure 2**, our false numbers have come way down compared to past numbers.

Figure 3, MAVTS, shows how much we have reduced our false hours in the 0-12 and 12-24 hour time frame. Finally, for the months of May and July, we actually had a higher POD and lower false hour count compared to the MAV in the 0-12 and 12-24 hour window. We use this thought process also in our conference calls with the CWSU when preparing the TAF.

I do not believe we would have had this turnaround without both the CWSU calls and the verification program.

The verification program considers a TS forecast when the pop is 30 percent or higher in either the TAF or the guidance. We have made this turnaround by adopting two rules:

- When in doubt, leave them out.
- If MAV does not have them, especially in the 12-24 hour range, leave them out.

The other area I reviewed is model comparison, brought up in TS forecasting. I also use it for the all important IFR and LIFR conditions, since this impacts runway usage, airport capacity, and even airlines shutdown at MDW since they are only a CAT I field.

It's interesting to note the overall trend is for the old FWC

to be better than the newer MAV in both conditions, with a higher POD and about even FAR. (See **Figures 4 and 5**, IFR and LIFR.)

This is not an outcome I would have expected considering the MAV is from a more advanced model updated every 6 hours. But at least for O'Hare and Midway combined, it does not seem to matter.

Figure 6, POD IFR, shows we are beating the better FWC most months, with a higher POD and lower FAR.

Part of our success can also be attributed to the adoption of the "Practically Perfect TAF" ideology, where we mainly deal with the operationally significant changes, like MVFR to IFR, etc., and greatly limit the Prob and Tempo groups, especially in the last 12 hours of the TAF.

Our next step is to bring our improved TAFs to the attention of the aviation community, such as airport operations.

The data shows that in aviation the models still have room for improvement. This is where the NWS forecaster can really make a difference and find a niche. It is increasingly hard for us to beat the models in the public arena, but we have a major opening in the TAF arena.

The TAF is only for a tiny 5-mile circle, as opposed to the much larger area the public forecast covers. It takes more effort in aviation weather to pick out the tiny variation that makes a big difference in ceilings, visibility and TS, but the payoff to our customers is huge.

At O'Hare, IFR cuts the hourly acceptance rate from about 100 to 75. At Midway, a 200-1/2 puts the field below minimums and shuts them down. So we need to be right at these critical levels, without issuing false alarms. ✈

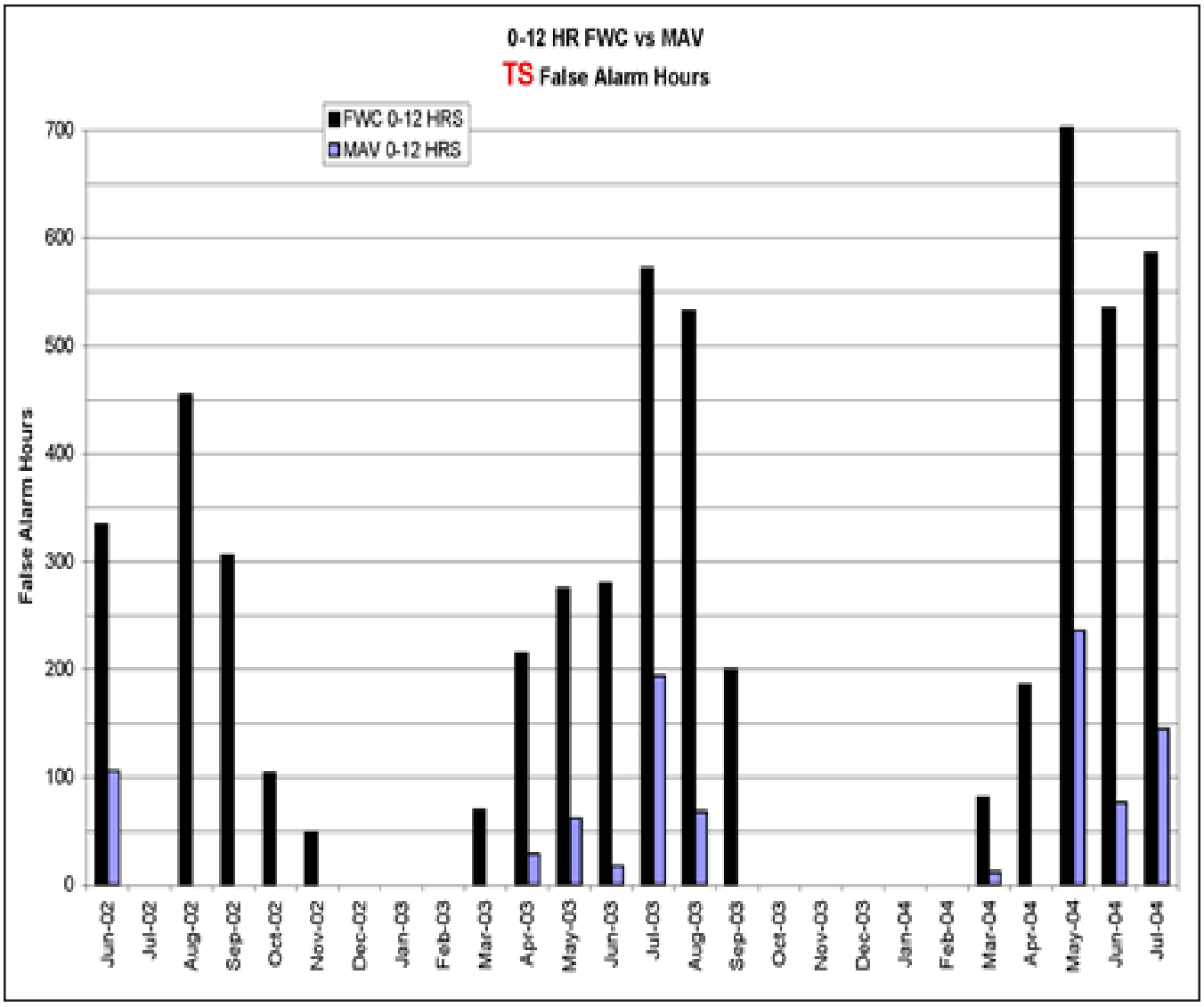


Figure 1. *Thunderstorm False Alarm Hours*

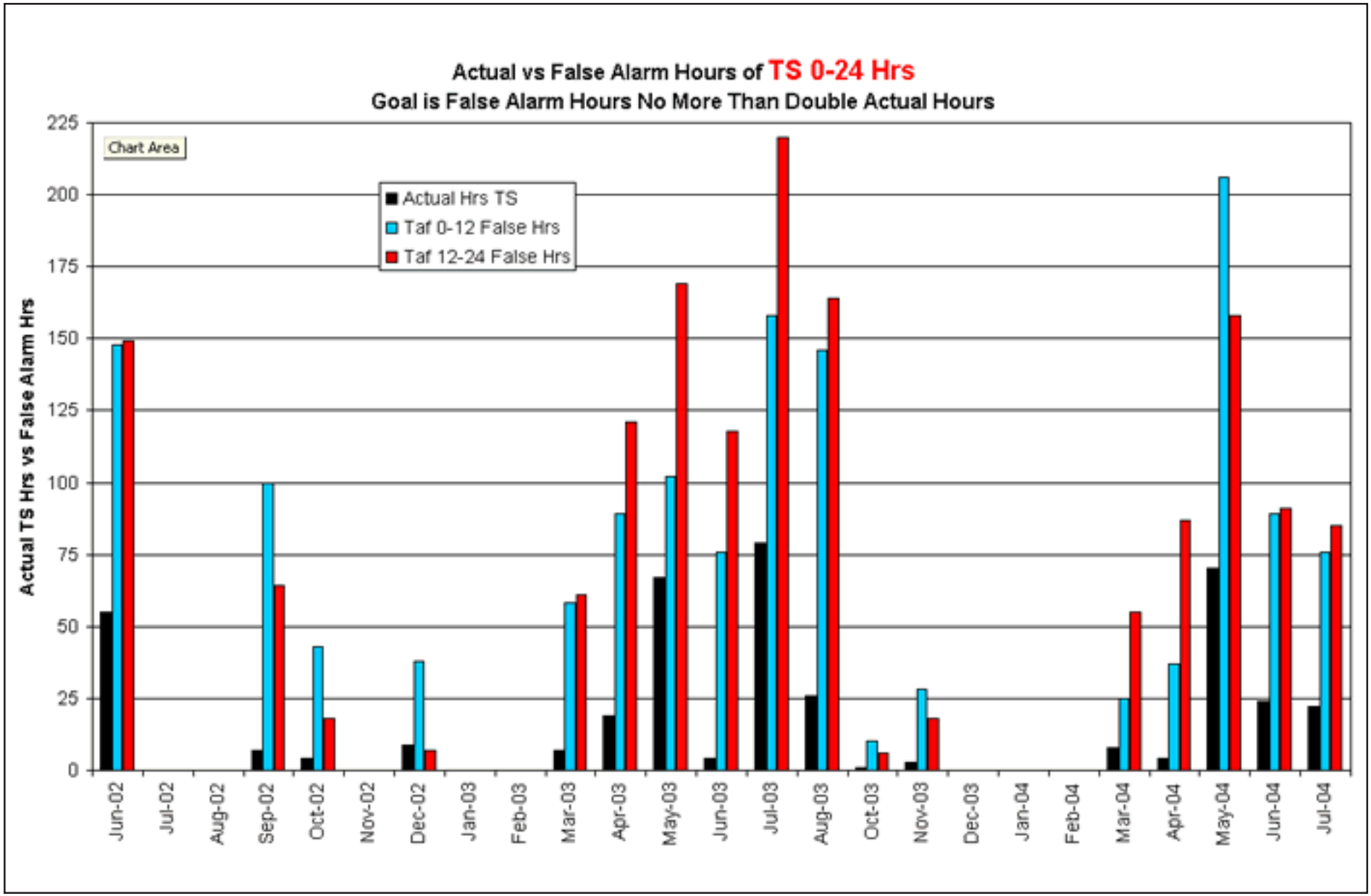


Figure 2. *MAV Thunderstorms: Actual vs. False Alarm Hours*

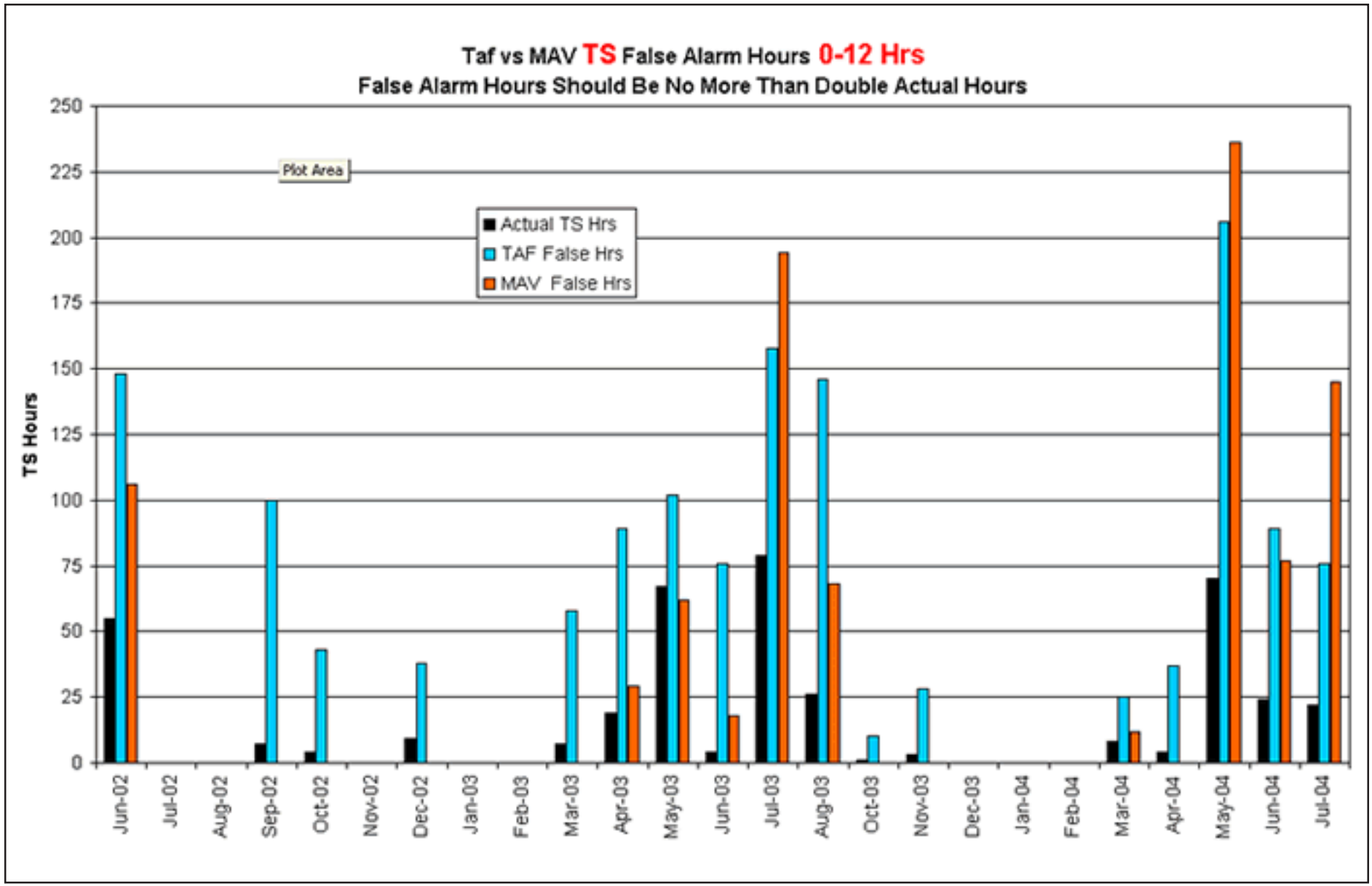


Figure 3. *MAV: Terminal Aerodrome Forecast vs. MAV Thunderstorm Storm Forecast*

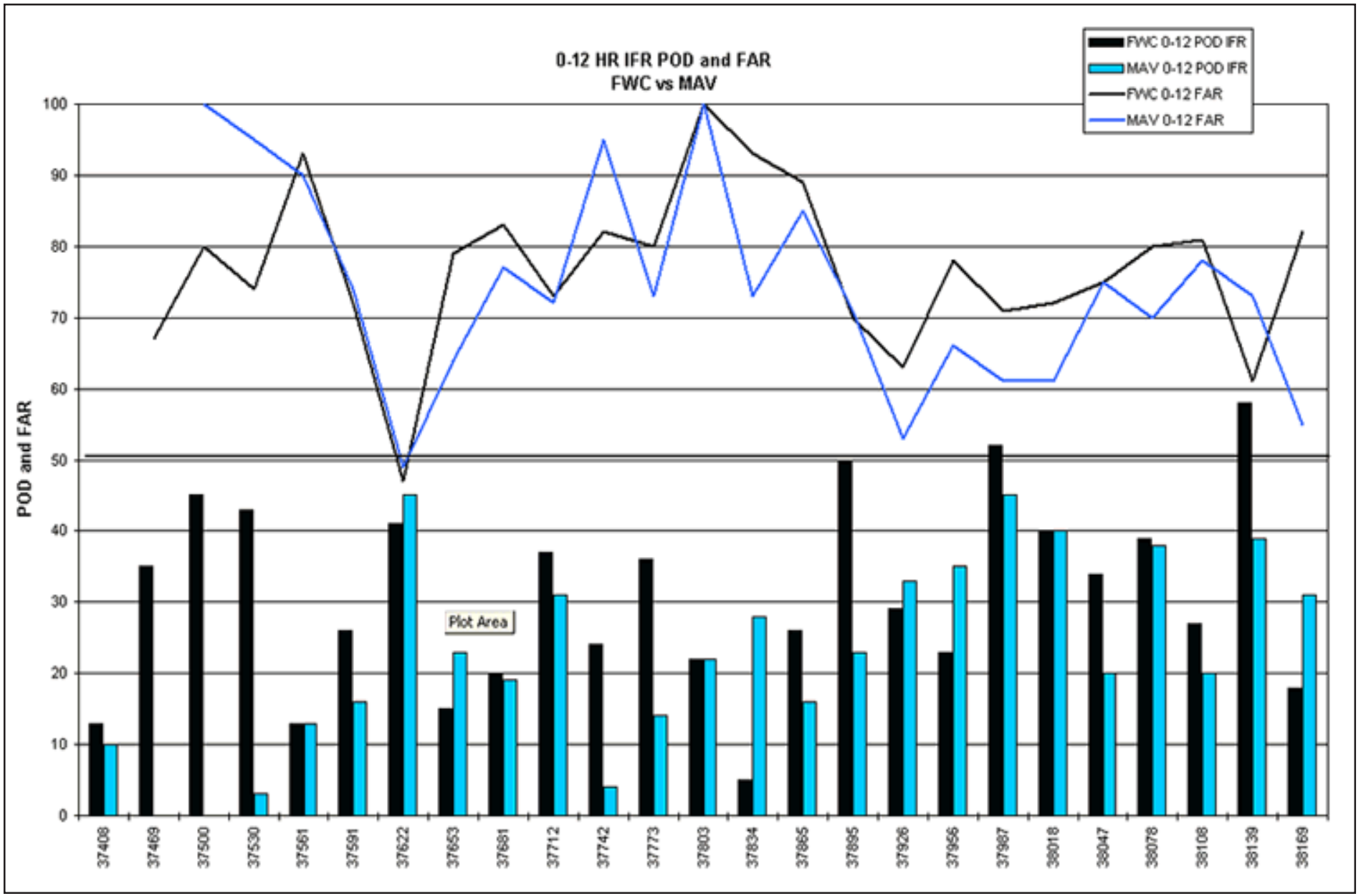


Figure 4. *Instrument Flight Rules: Probability of Detection vs. False Alarm Rates*

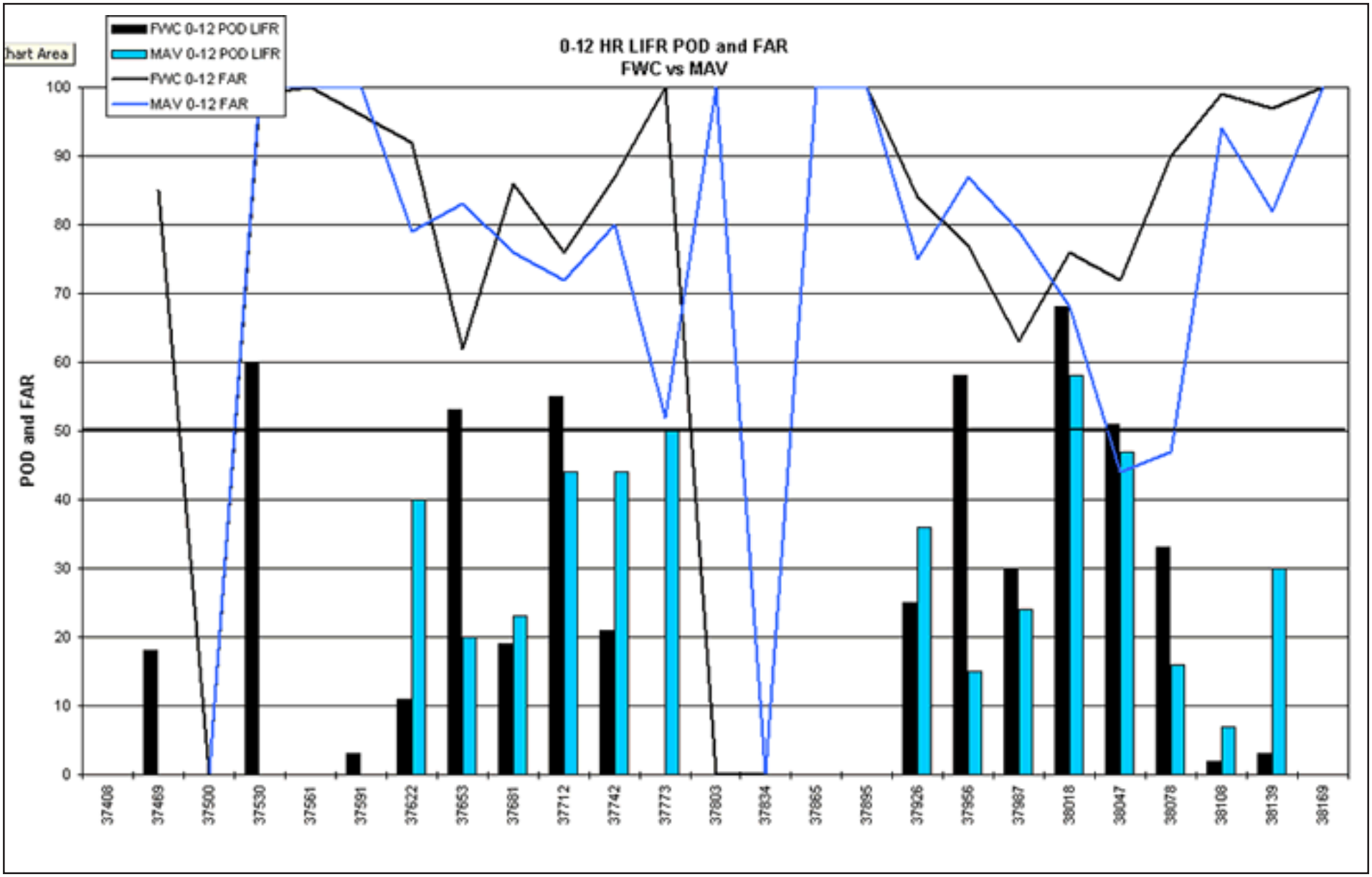


Figure 5. *Probability of Detection Low Instrument Flight Rules*

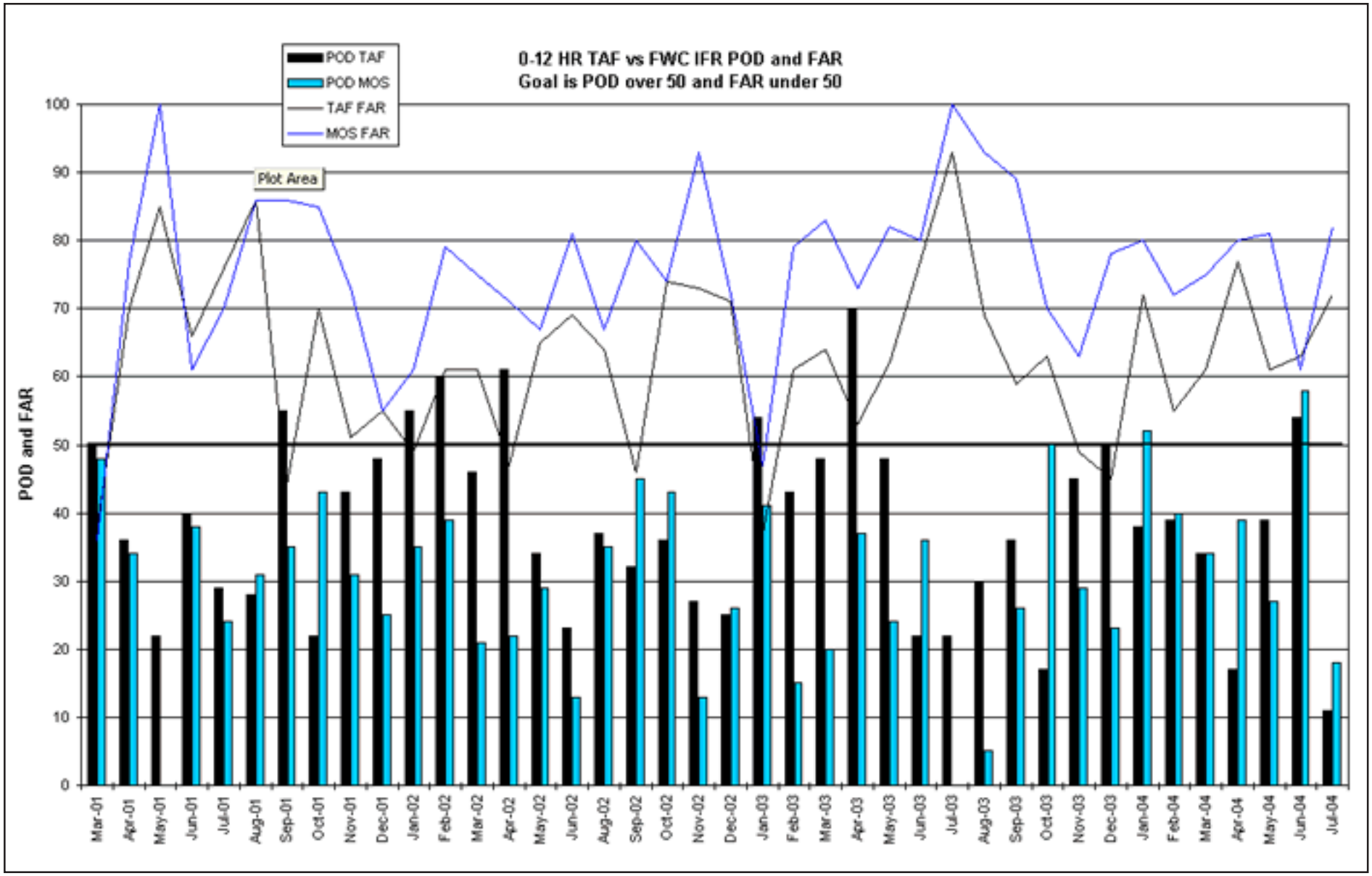


Figure 6. *Low Instrument Flight Rules: Probability of Detection vs. False Alarm Rates*

Improving Aviation Verification Statistics by Reducing False Alarms

By Craig Sanders, WFO Duluth, MN

"What gets measured gets done" is an Air Force axiom that focuses on performance and improvement of any task. The NWS is using this practice through verification statistics in nearly all aspects of its operation.

A newly released Instruction, NWSI 10-1601, contains an overview of all verification topics, including aviation forecasts.

The Aviation Services Branch (ASB) has been ramping up its efforts to provide forecasters with meaningful statistics and forecasting practices to improve the statistics. Call it bean counting but improving our products offers some very real economic benefits for pilots.

Verification Tools

The Weather Service's new national standard program, named Stats on Demand, will calculate aviation TAF performance statistics as well as stats for other NWS programs.

Stats on Demand creates a ton of data. Each WFO renders these data down to a succinct set of meaningful tables or graphs and charts to aid forecasters interpret the monthly results.

The verification site does not produce easy-to-use MS Excel spreadsheets from which charts can be created quickly. So some data still has to be extracted and entered into spreadsheets. That delays the work a bit, but it is vital that aviation focal points and SOOs enter the data.

Currently, the stats are generalized for all TAF sites served by each WFO, and all forecasters in that WFO.

In the future, Stats on Demand will offer stats for individual forecasters based on their forecaster numbers. You will be able to track your personal performance as well.

Not to worry, though! Access will be by password over a secure server. Neither managers nor aviation F/Ps will be able to view individual forecaster stats. Also, since it is a secure server, no one outside the NWS can get at the data. There will be more on that later when this feature becomes available.

ASB plans to track False Alarm Ratio (FAR) and Probability of Detection (POD) for TAFs for IFR conditions. No national or regional goals have been set yet, but you can bet that will come because the NWS has to meet Government Performance Review (GPR) goals. It is imperative that forecasters understand the impact of the service we now provide and learn how to improve our products.

A Little History

About 3 years ago, Andrew Rorke, forecaster at WFO Los Angeles created an outstanding MS Excel-based program called Aviation Verify.

The program produced an exhaustive array of aviation verification stats. There were so many, in fact, that no one was really sure what was most important or how to track it.

NWS brought in Northwest Airlines, the Aircraft Owners and Pilots' Association (AOPA), the FAA and other groups to advise the NWS about what is important to pilots and how better forecasts could help the aviation.

Forecasts of ceiling, visibility, and thunderstorms have a great impact on pilots because the FAA regulations under which pilots operate require that enough fuel is loaded so that the pilot can fly to the destination and hold for an additional 45 minutes.

Weight in an aircraft means fuel burn, and fuel burn means money either because of the fuel consumed or because of space used that could have carried passengers or cargo.

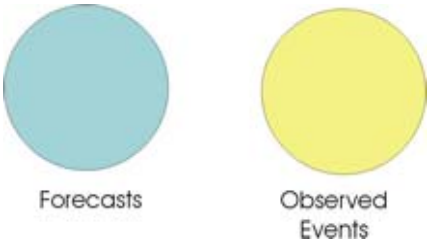
For years forecasters have been writing TAFs that not only try to describe future weather events but also make them feel good. We want to make sure that if any weather event happens, no matter how fleeting, it was mentioned in the TAF. That practice has led to long TEMPO-ladened or PROB30-ladened TAFs. Numerous sessions with aviation groups point to the need to avoid these practices.

This practice has negatively impacted pilots excessively because they have to carry fuel based on the worst weather case presented in the TAF whether it is in the main body of the TAF or in the TEMPO group. Numerous dialogue sessions with these aviation groups point to the need to get away from these practices whenever possible.

If we can break from past practices and learn to write TAFs differently, we will improve our verification and help the pilot. That's our mission in the aviation verification effort.

FAR and POD

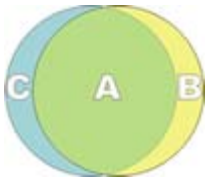
The FAR stats describe our shortcomings by telling us if we're crying wolf. POD is success oriented and tells us how well we are forecasting. These two



stats are derived differently for aviation forecasts than they are for severe weather and hydrologic events.

Forecasts and observed events can be thought of as two sets of data. Forecasts without supporting observations offer no confidence to the pilot. Observations without forecasts may only endanger pilots or at least cause them to take a problematic flight into adverse weather conditions.

When forecasts are verified with observations, these two data sets intersect (area A) by some amount.



Perfect forecasts would cause both of these circles to coincide exactly. Obviously, that never happens. So in the real world we have forecasts that don't verify (C) and observed events that were not forecasted (B).

False Alarm Ratio

False Alarm Ratio is defined as the ratio of forecasts that

$$FAR = \frac{C}{A+C}$$

don't verify, or misses, (area C), to all forecasts that were issued. All forecasts whether they verify or not are represented by areas A and C. Area A represents hits. Area C represents misses.

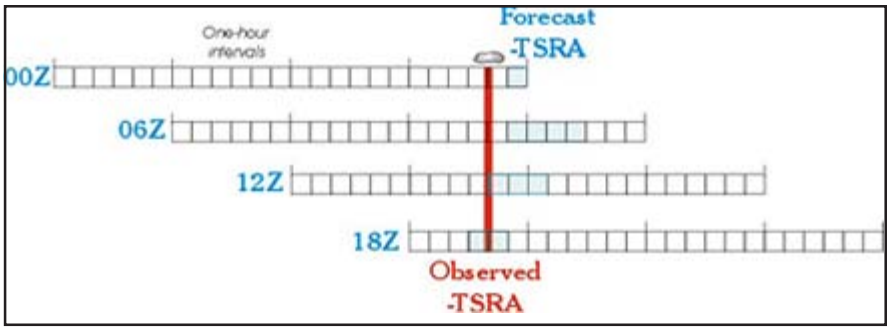


Figure 1

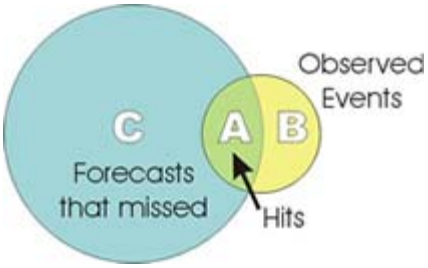
In Aviation Verify, the FAR is actually measured by False Alarm Hours of a weather element rather than as a ratio. The FAR represents the actual number of hours that an event occurred but was not forecast. So you won't see a nice little decimal number between zero and one when talking about false alarms in aviation forecasting as you would in severe weather.

But that's not the complete story. Events are over-counted by Aviation Verify and Stats on Demand (See **Figure 1**). TAFs can represent up to 16 hours of forecasted thunderstorms for a single, relatively quick event such as a frontal passage.

As an example, let's say that thunderstorms are deemed possible tomorrow afternoon. Each TAF issuance may then contain either a TEMPO or PROB30 group to describe the event as shown in the adjacent figure.

The OOZ TAF may mention thunderstorms in the fourth 6 hour period. The 12Z will mention thunderstorms in the second 6 hour period. From the verification perspective, one thunderstorm lasting 20 minutes may actually be forecasted for 16 hours if each TEMPO group is inserted with a maximum of 4 hours for that event.

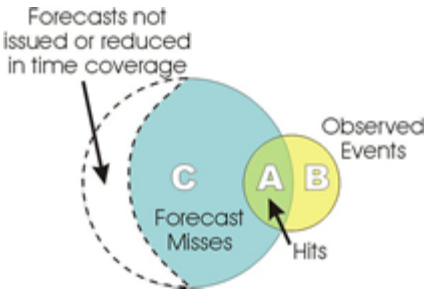
If NWS frequently forecasts events, or our forecasts cover a large period, and the weather event does not occur, the



forecasts are inaccurate. "C" becomes large relative to "A," the verified events, and the FAR hours are huge.

The larger the number of hours, the more "cry wolf" applies, costing airlines money. Also, if we frequently forecast events that are rare or occurring briefly, we create a high FAR.

We can reduce the FAR by cutting back on how frequently



we forecast events or on the number of hours we specify in TEMPO and PROB groups.

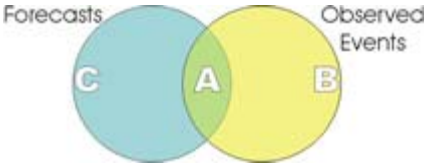
This allows the Hits (A) to assume a larger proportional value relative to the misses and thus reduces the FAR. We can improve TAF verification stats by cutting back on TEMPO or PROB group periods, or by not mentioning these groups at all when the probability of occurrence is low due to the spatial coverage or time duration of an event. In latter hours of a TAF, we could pick a good target value for the FM group and not even mention a TEMPO or PROB group.

FAR indicates poor performance because it is weighted with misses. Reducing the misses or the number of hours an event is mentioned in the TAF will lower the FAR (hours).

Probability of Detection

POD is the fraction of correctly forecasted events (hits) relative to all the weather events that occurred. This includes all those that verified and those that were not forecasted, but did occur.

$$POD = \frac{A}{A+B}$$



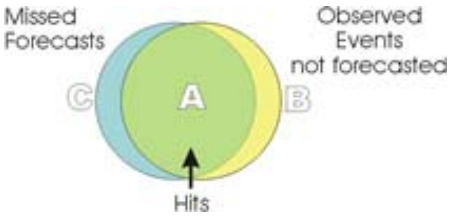
When an event is occurring over a long period, i.e., their are a large number of observations, you should be able to forecast it well. The probability of hitting the forecast is high because of the frequency or duration of the events. The larger the number of verified forecasts (hits), the larger the POD. As we improve our forecast skill, B becomes very small.

In other words, events occurring without a forecast become minimal. POD infers skill and success because it is weighted with hits.

The maximum value of POD is 1 which indicates success.

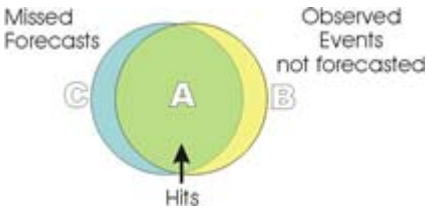
Four statistical scenarios

- This **high POD and low FAR** condition represents good forecasting. You have few



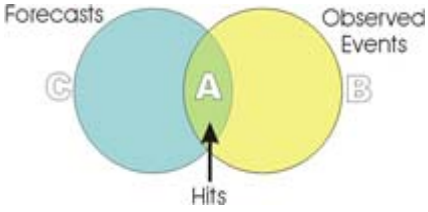
observations that occurred without warnings. And the warnings you issued were, in large part, verified.

- With **low POD and high FAR**, we have low quality or lazy forecasting. Forecasts are not hitting. Plus, the events that do occur are not being met

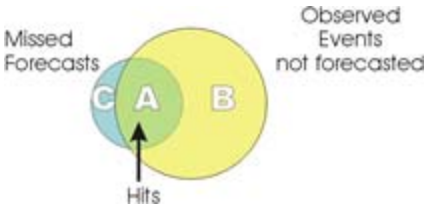


with a forecast that would likely verify.

- When **POD and FAR are both low**, the event may have occurred over a small time period for which you correctly

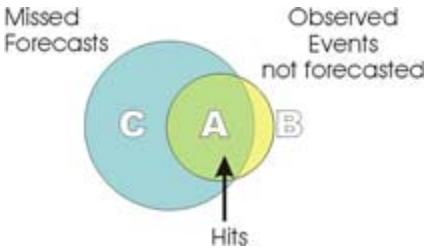


forecasted, but other similar events occurred which you failed to forecast or chose not to issue a forecast. Small



increases in time duration of the forecasted event will gradually improve these statistics. A higher FAR, but with benefit of a higher POD is permissible. Knowing climatology of specific events will help in this case.

- When **both POD and FAR are high,** you have a shotgun approach to forecasting. You succeeded in hitting the events that occurred, but you did cry wolf (high FAR). Despite the POD success, you still required an increased



expense or inconvenience to the pilot or airline. You may be pleased with the POD alone, but did you really improve meaningful service to the customer?

Results of Aviation Verification

ASB has decided to verify only the first 6 hours of a scheduled TAF. This is the period where most flight planning is conducted and most flights occur.

In other words, a pilot uses the 12Z TAF to plan and fly to his destination. Most flights last 4 hours or less.

Certainly pilots would want to know TAFs for their return trip, but when departing on the first leg, the first 6 hours is the key period. They will likely use the 18z TAF for the return leg, likely between noon and 6 p.m.

Verification data are divided into two 3-hour periods. The two charts below suggest that the FAR is lower in the first 3 hours, but is consistently higher in the 3-6 hour range. These numbers may be typical at many WFOs.

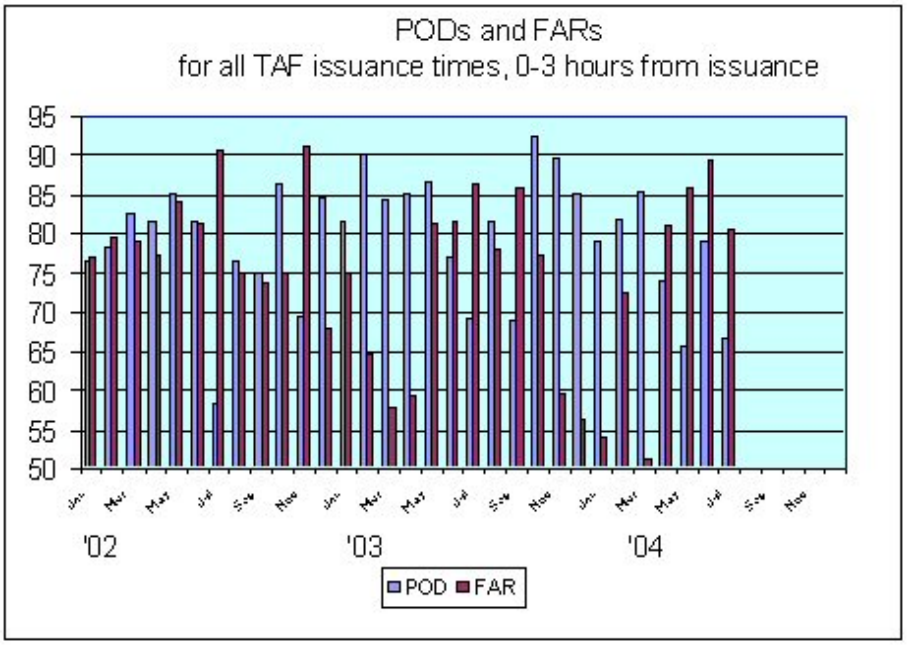
These graphs show that FARs are both large themselves and greatest during the wet months of the year. The year 2002 was wetter than 2003 or 2004 and shows a well distributed FAR, but still larger in the wet months, likely because of the cautionary forecasting with TEMPO and PROB groups.

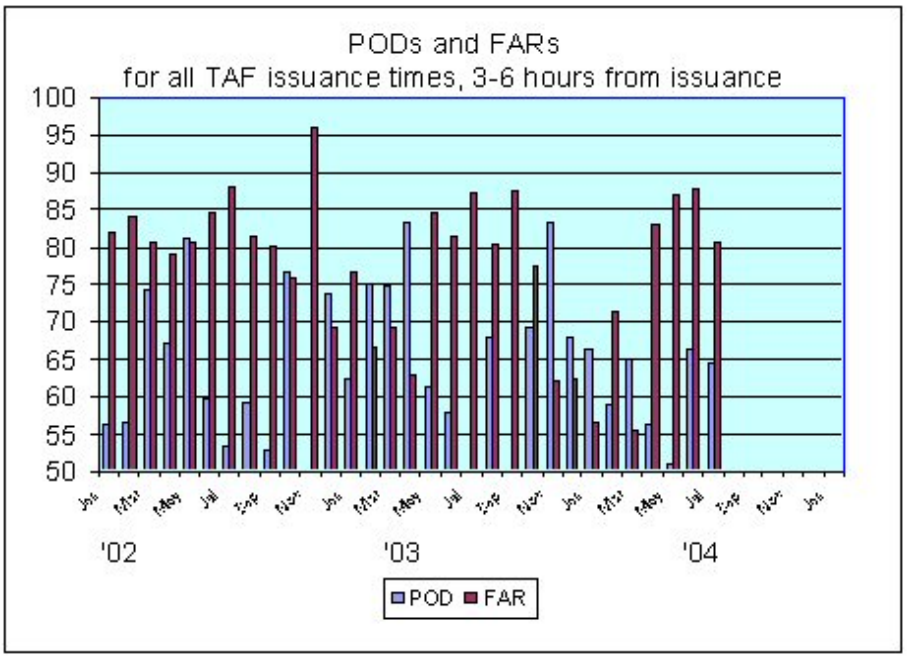
One assumption is that we over forecast precipitation. Fog may also be present more often during a wet year. GFS is the only model MOS evaluated at this writing, but the ETA MOS will be included in the future.

These graphs show that 06Z and 12Z TAFs normally cover the most *observed* IFR conditions as a general rule. The hours increase quickly for the period November through January. Knowledge of this climatological risk may prompt us to frequently over forecast IFR conditions.

These graphs clearly show thunderstorms are over forecast. Between 2002 and 2004 the office in question Forecasted thunderstorms frequently in the summer months; however, thunderstorms actually occurred less than 5 percent of the time.

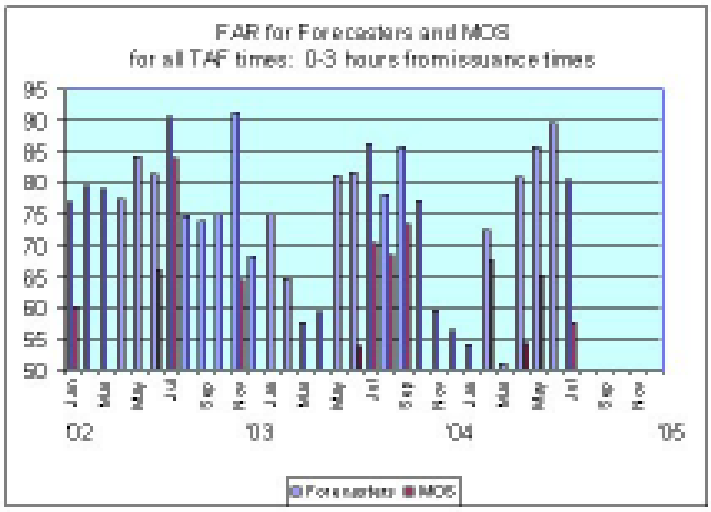
This would be a case of huge FAR with perhaps high POD. The forecaster no doubt mentioned thunder over a 4 hour period even a day in advance for an event that would likely last only a few minutes.

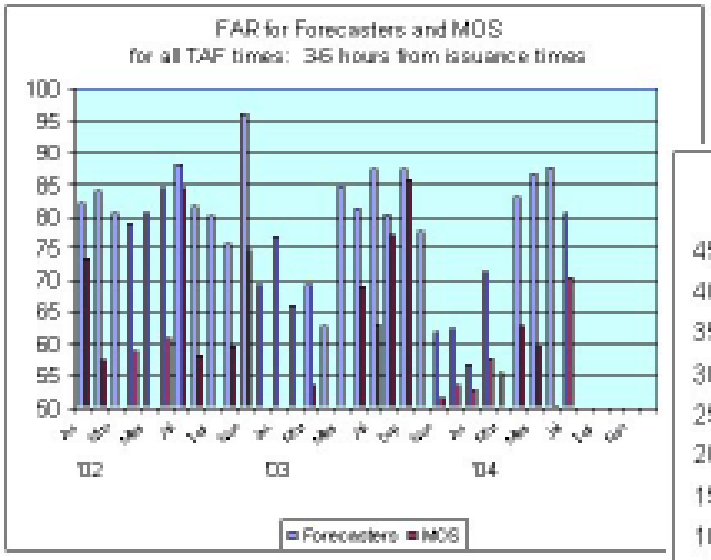




WFOs are aware of this situation have either eliminated TEMPO groups beyond the first 12 hours of the TAF or routinely mention VCTS in the body of the FM forecast group. VCTS is not counted as being a formal forecast of thunderstorms as –TSRA would be.

As forecasters, you may feel TEMPO groups add some better definition or awareness, timing or severity of an event. These charts show that for the office in question, TEMPO groups rarely actually help in meaningful ways. Since we know that we over forecast short term thunderstorm events as mentioned earlier, the TEMPOs actually hurt the TAFs in the summer because of the





often routine inclusion of a three or four hour TEMPO group.

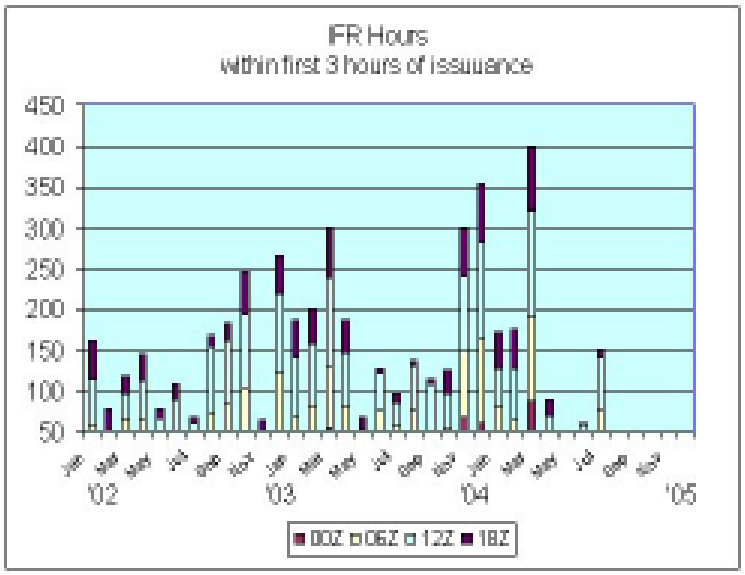
The extra caution in the forecast actually hurt the overall accuracy of the TAF. However, in the winter when longer term events are more likely to persist and to be forecasted, our TEMPO groups have a better chance of adding useful information.

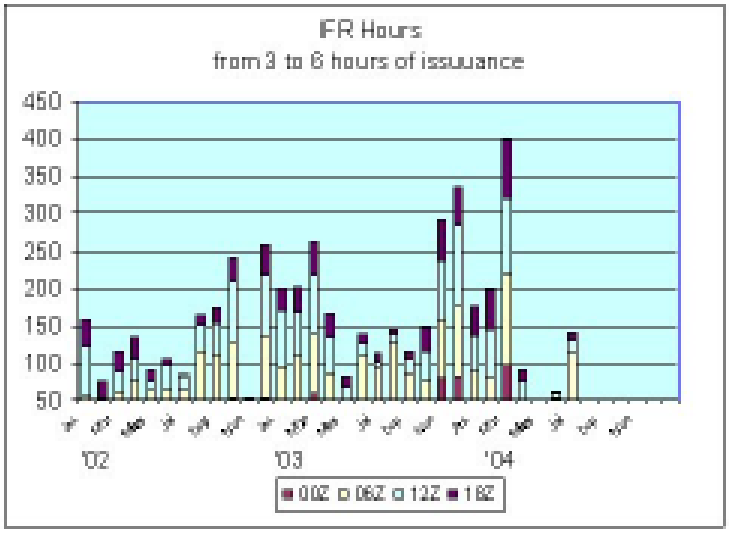
The "Hurt TAF" stats are lower in the winter. Visibility is also much more variable than ceiling. The units depicted in the chart are not as important as is the message of forecasters not helping the forecast with numerous TEMPO groups.

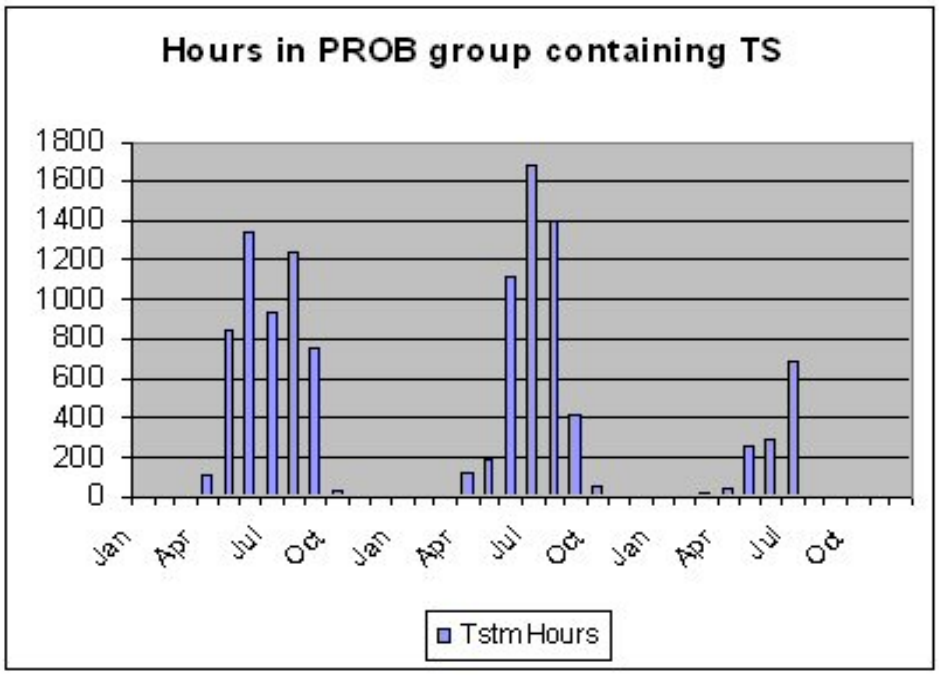
Conclusions

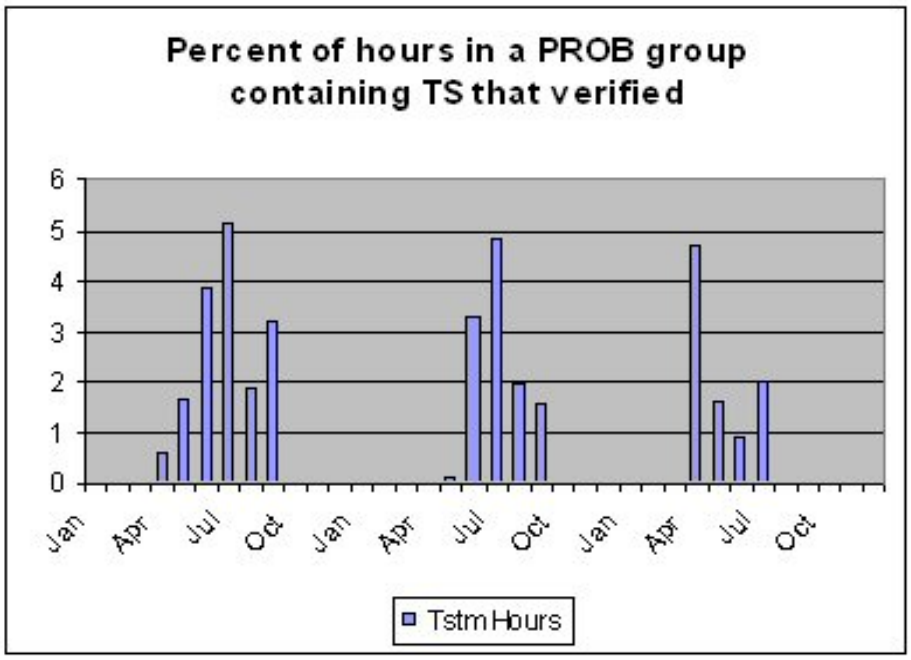
TAFs have a serious impact on the economics of flying. The national GPR goals require that we improve our statistics to serve pilots better. Over forecasting with TEMPO and PROB groups obviously degrades the perceived value of TAFs.

We can improve WFO stats by reducing the length of TEMPO or PROB groups in the first 6 hours of the TAF by knowing local climatology. We can also eliminate TEMPO and PROB groups from TAFs beyond 12 hours while using a best guess value in the FM group covering the time of the expected event.







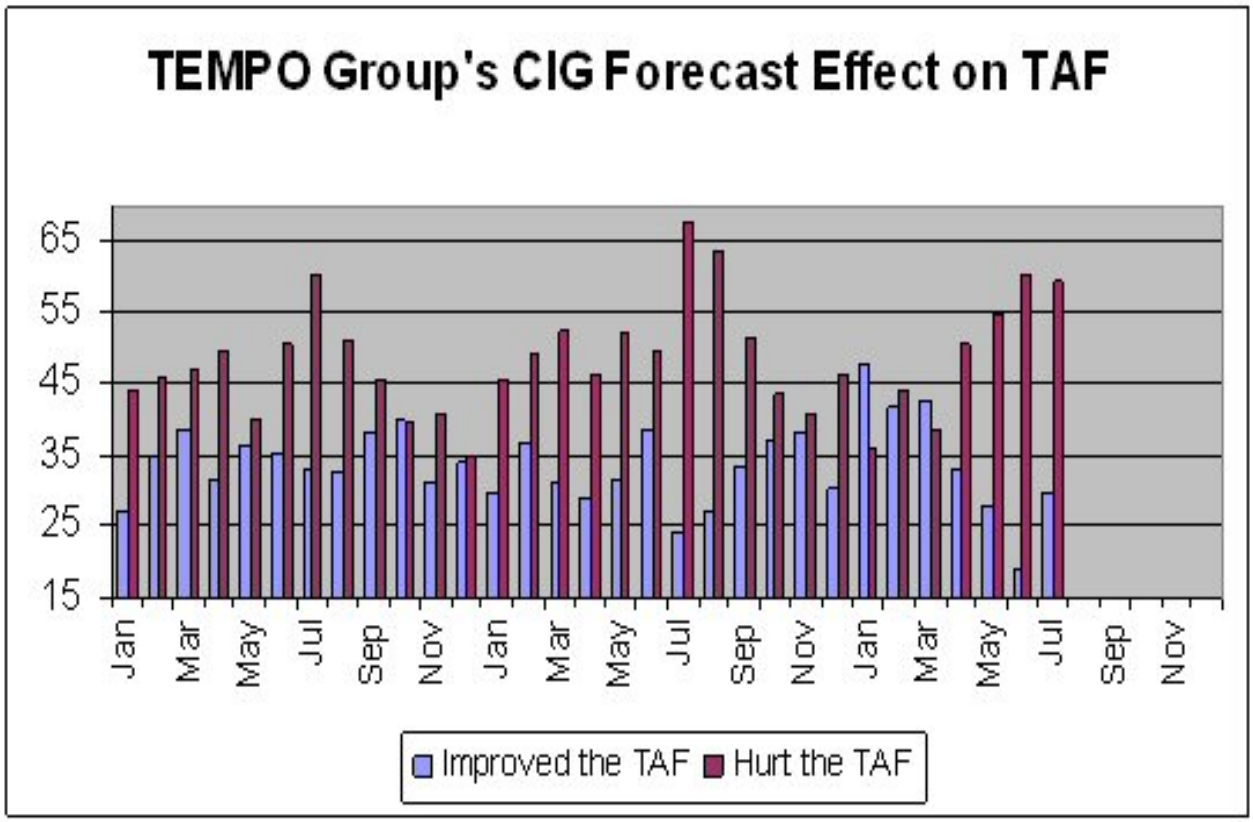


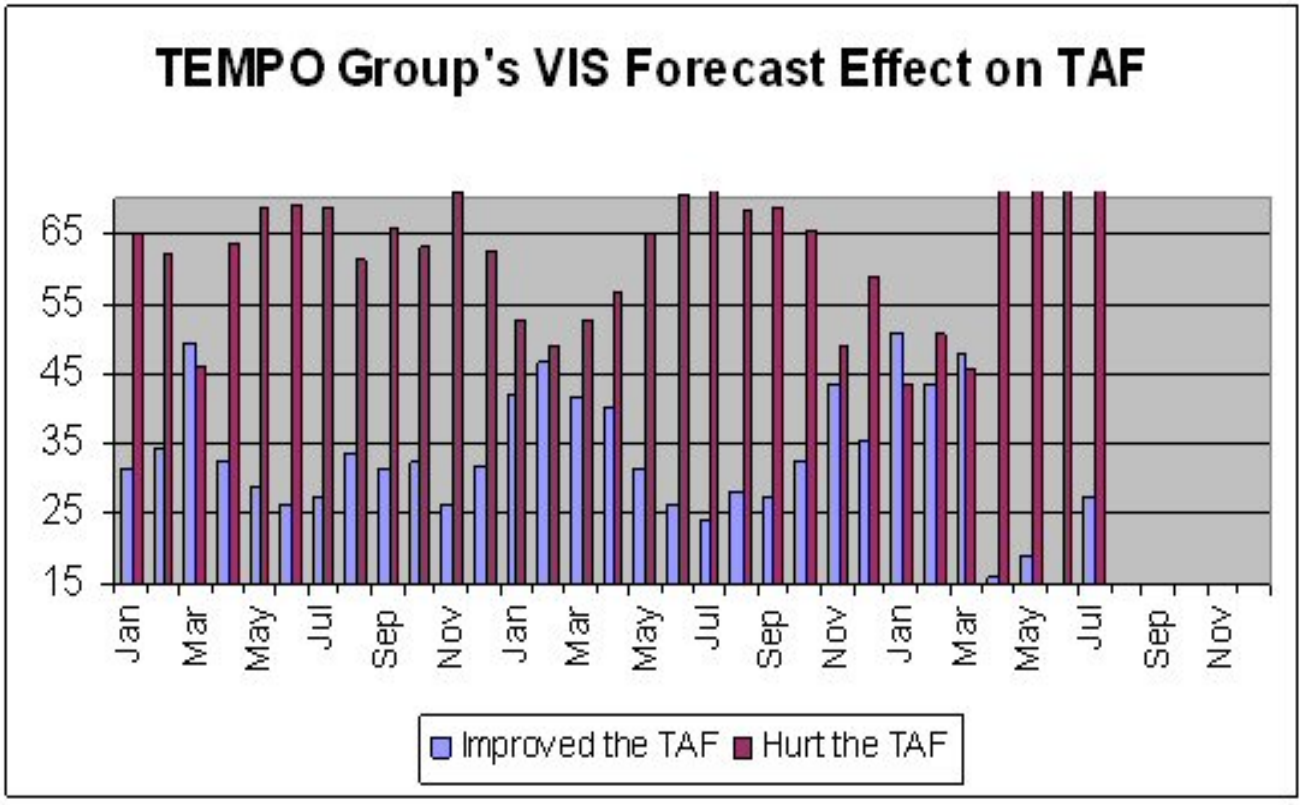
Forecasters could use later TAFs to focus on the event in a more confident, accurate and meaningful way.

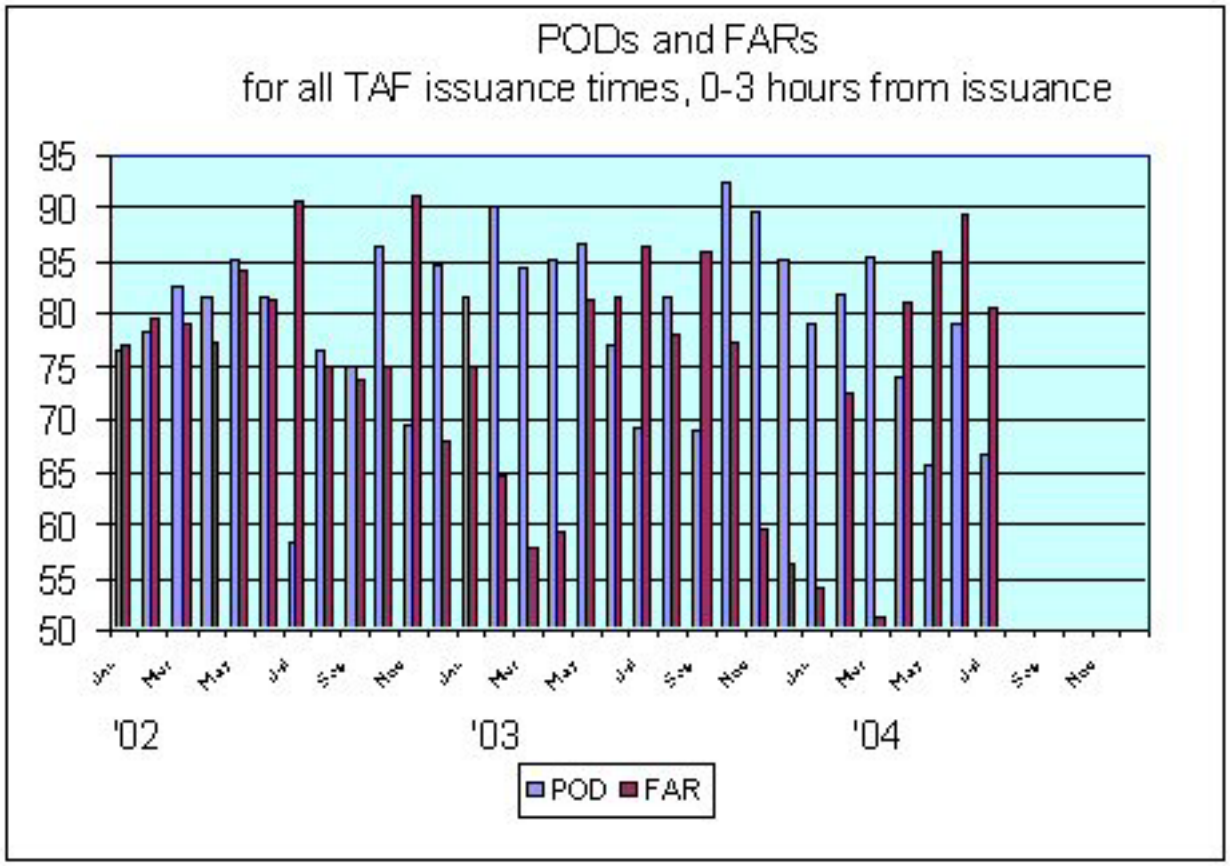
FAR is defined as the ratio of forecasts that don't verify, or misses, (area C), to all forecasts issued. All forecasts, whether they verify or not, are represented by areas "A" and "C." Area A represents hits. Area C represents misses.

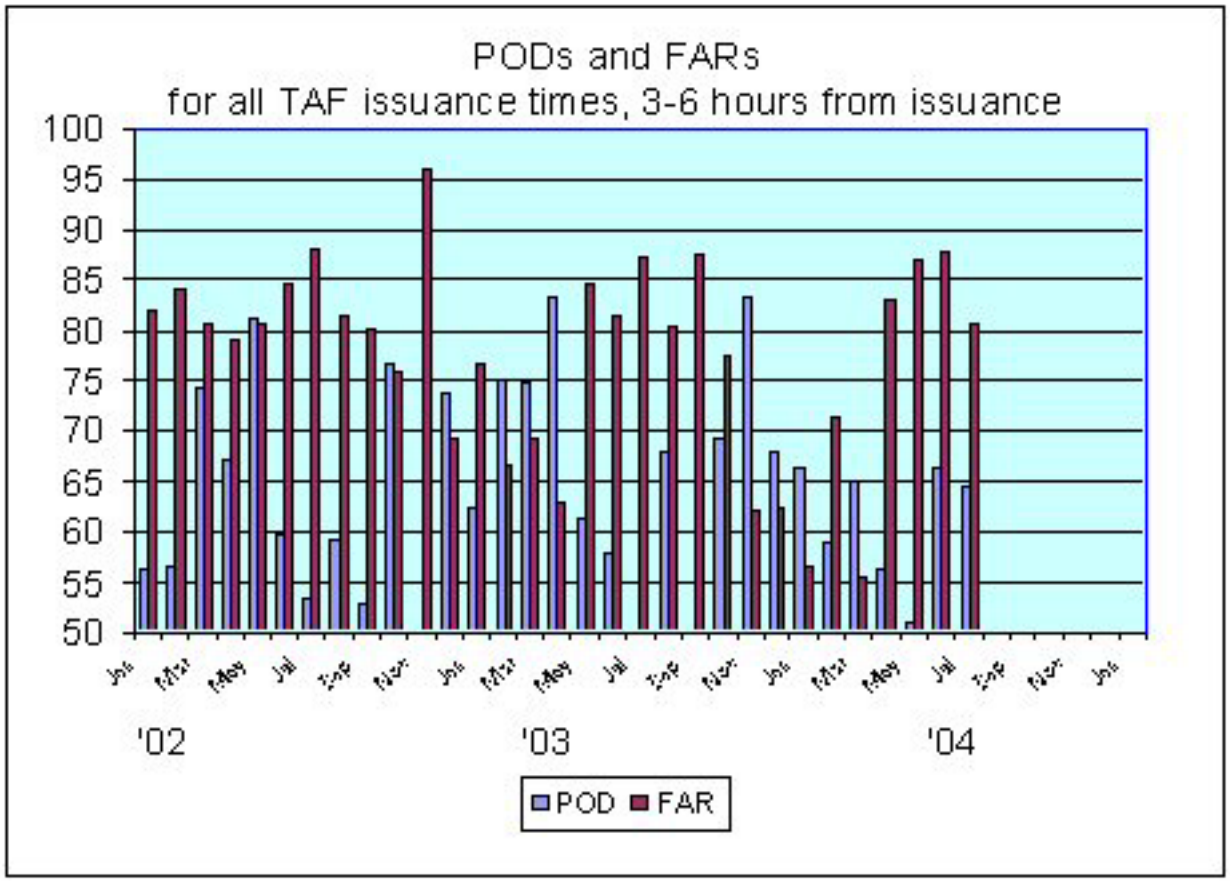
It documents whether we succeeded in our warnings for those events we chose to warn for. It doesn't take into account all the other events that might have occurred and which may have required a warning.

POD is the fraction of correctly forecasted events (hits) relative to all the weather events that occurred. This includes all those that verified and those that were not forecasted, but did occur. Q


TEMPO Group's CIG Forecast Effect on TAF
65
55
45
35
25
15
Jan
Mar
May
Jul
Sep
Nov
Jan
Mar
May
Jul
Sep
Nov
Jan
Mar
May
Jul
Sep
Nov
Improved the TAF
Hurt the TAF


TEMPO Group's VIS Forecast Effect on TAF
65
55
45
35
25
15
Jan
Mar
May
Jul
Sep
Nov
Jan
Mar
May
Jul
Sep
Nov
Jan
Mar
May
Jul
Sep
Nov
Improved the TAF
Hurt the TAF


PODs and FARs
for all TAF issuance times, 0-3 hours from issuance
95
90
85
80
75
70
65
60
55
50
'02
'03
'04
POD
FAR


PODs and FARs
for all TAF issuance times, 3-6 hours from issuance
100
95
90
85
80
75
70
65
60
55
50
'02
'03
'04
POD
FAR