SAVE POD AND FAR

Normalization for Event Frequency in Performance Metrics (IFR Example)
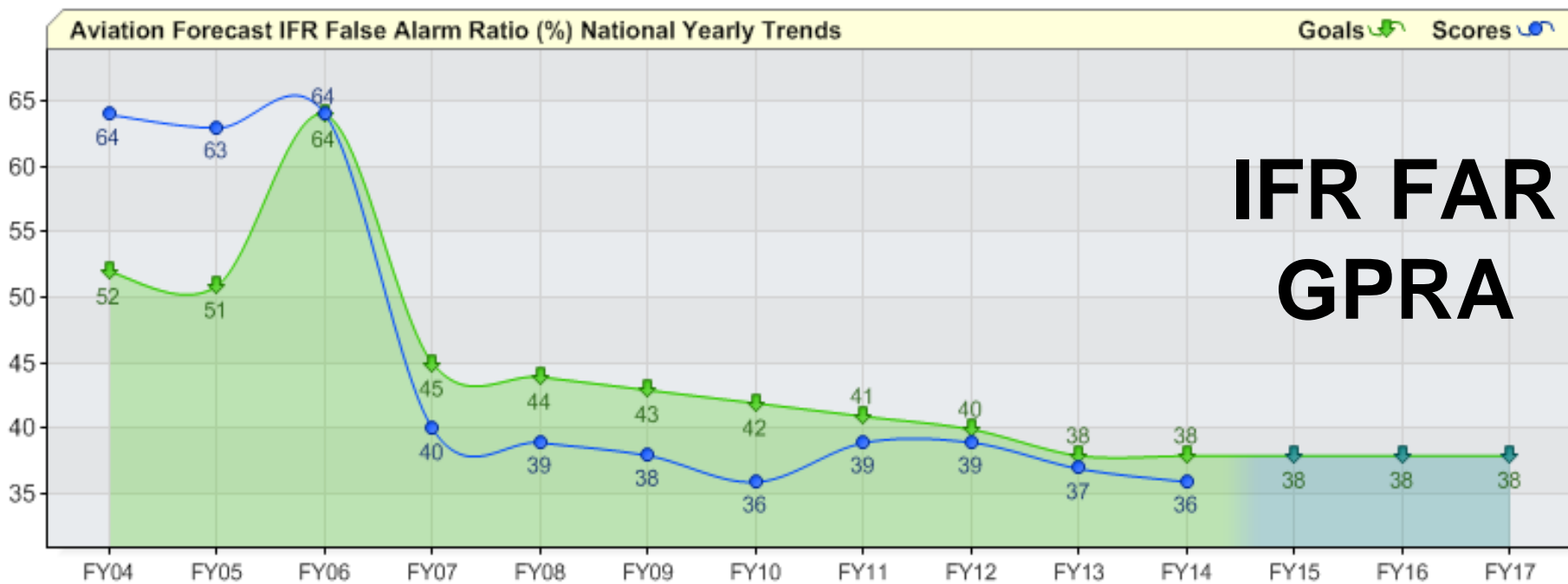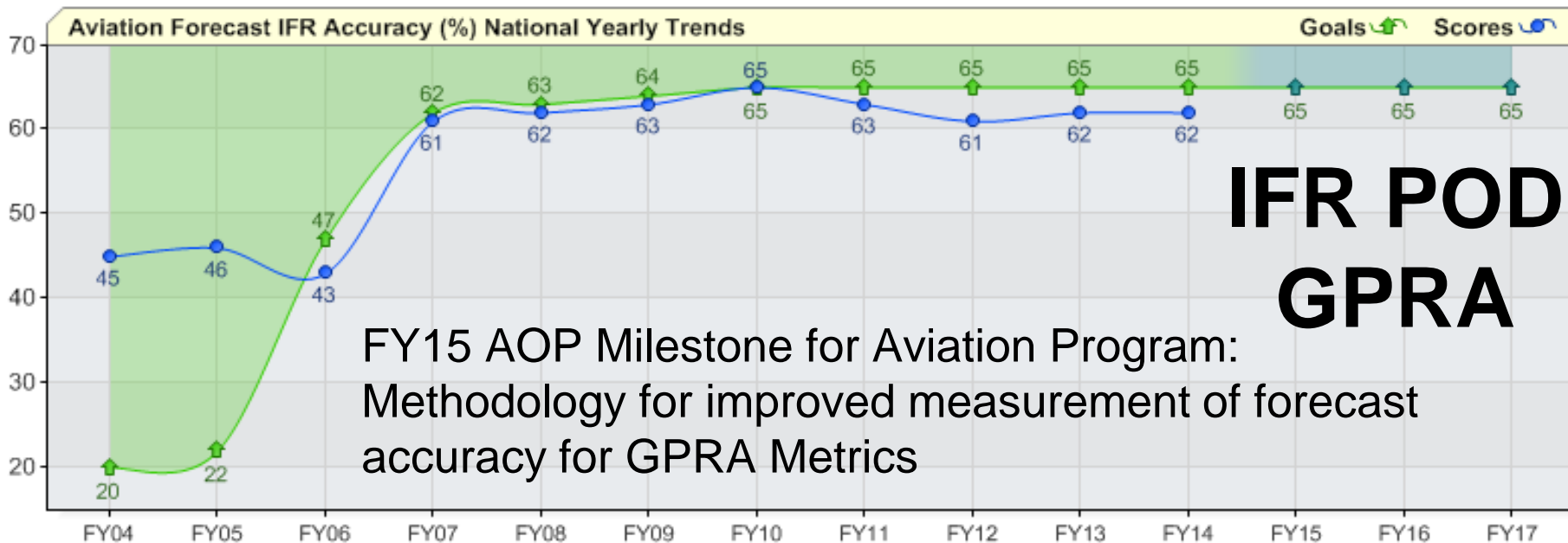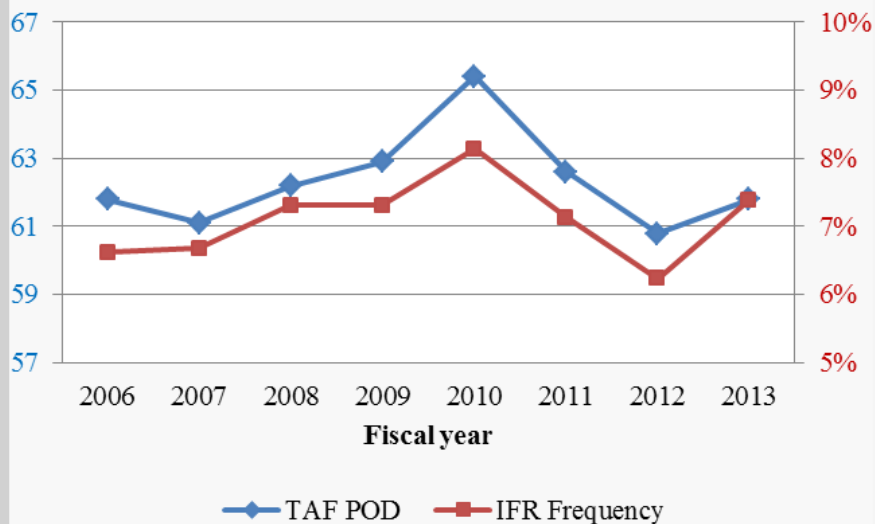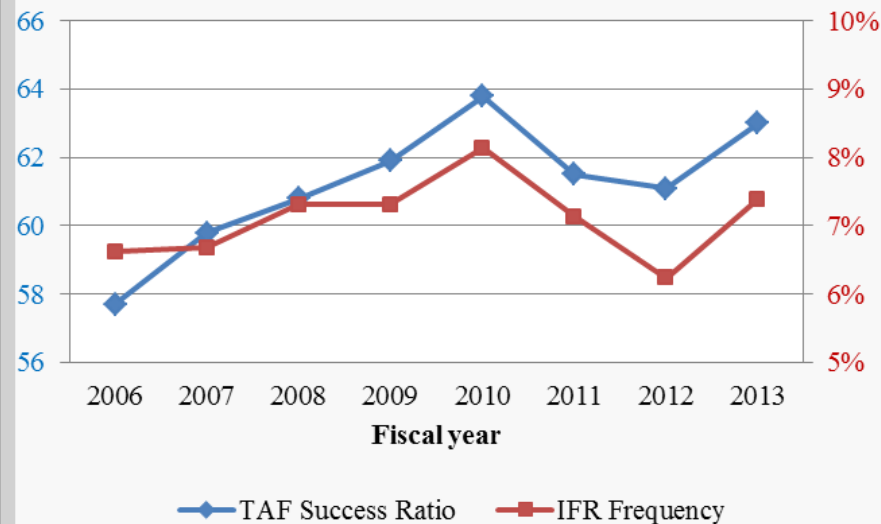
Matthew Lorentson

August 2015

High Points

- POD and FAR must not be used individually to *summarize* performance

- Performance metrics must be normalized to account for the gross influence of event frequency

- Use a moving average to evaluate progress and trends

Aviation Forecast IFR Accuracy (%) National Yearly Trends — Goals / Scores

IFR POD GPRA

FY15 AOP Milestone for Aviation Program: Methodology for improved measurement of forecast accuracy for GPRA Metrics

Scores (blue): FY04 45, FY05 46, FY06 43, FY07 61, FY08 62, FY09 63, FY10 65, FY11 63, FY12 61, FY13 62, FY14 62

Goals (green): FY04 20, FY05 22, FY06 47, FY07 62, FY08 63, FY09 64, FY10 65, FY11 65, FY12 65, FY13 65, FY14 65, FY15 65, FY16 65, FY17 65

Aviation Forecast IFR False Alarm Ratio (%) National Yearly Trends — Goals / Scores

IFR FAR GPRA

Scores (blue): FY04 64, FY05 63, FY06 64, FY07 40, FY08 39, FY09 38, FY10 36, FY11 39, FY12 39, FY13 37, FY14 36

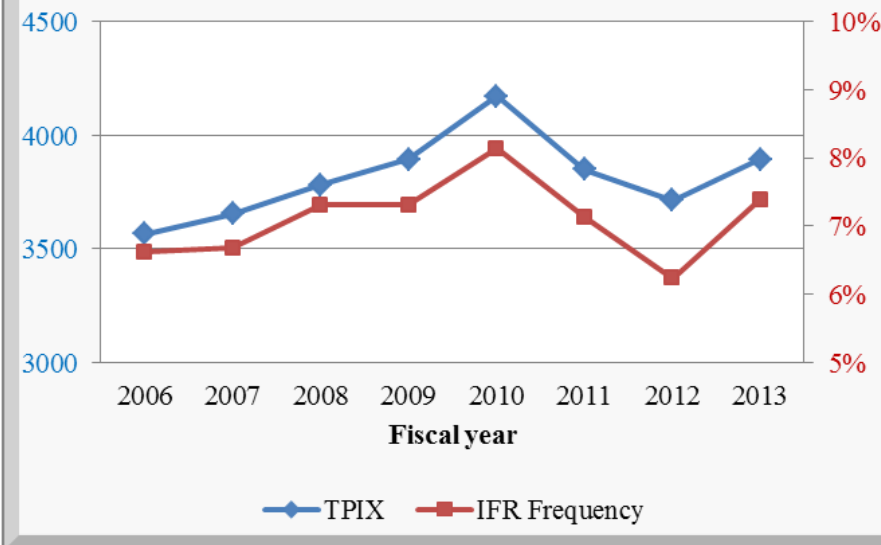Goals (green): FY04 52, FY05 51, FY06 64, FY07 45, FY08 44, FY09 43, FY10 42, FY11 41, FY12 40, FY13 38, FY14 38, FY15 38, FY16 38, FY17 38

3

TPIX = POD x SR

– Index use solves POD or FAR overemphasis

– TPIX easy to calculate (advantage over CSI)

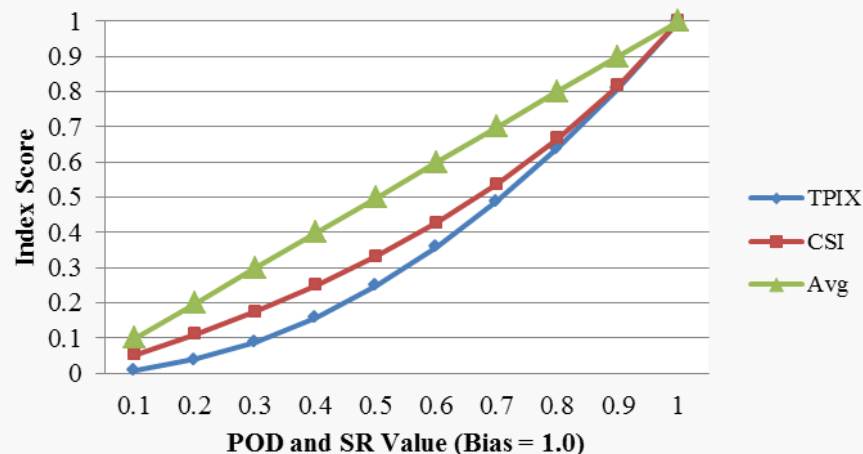– TPIX exhibits strong linear correlation with IFR Frequency

4

–IFR Frequency range (right axis, in red) is magnified compared to IFR Total Performance Index range (right axis, in blue) to emphasize comparable distribution shapes
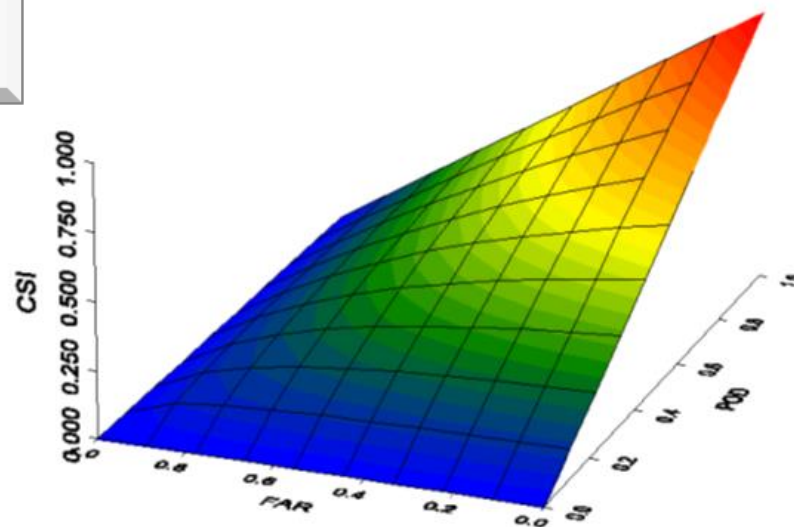
Performance Indices

CSI vs. TPIX

–"It is essential to recall that there is no universal approach to verification, but rather that the procedure selected needs to match the specific objectives of the study" - Roebber, 2009

–Avg = (POD + SR)/2
–CSI = 1/[1/POD+(1/SR)-1]
–TPIX = POD x SR

5

N80

	Month	POD	FAR	SR	TPIX	CSI	Avg.	IFR Freq.	IFR Freq^2	IFR Freq^3	K
65	Jan-11	66.40	33.50	66.50	4415.60	0.4976	66.45	12.38%	0.01533	0.0018974	
66	Feb-11	67.20	35.00	65.00	4368.00	0.4934	66.10	10.12%	0.01024	0.0010364	
67	Mar-11	63.70	37.70	62.30	3968.51	0.4598	63.00	8.25%	0.00681	0.0005615	
68	Apr-11	62.30	38.70	61.30	3818.99	0.4471	61.80	6.94%	0.00482	0.0003343	
69	May-11	63.30	37.70	62.30	3943.59	0.4577	62.80	6.40%	0.00410	0.0002621	
70	Jun-11	54.30	44.10	55.90	3035.37	0.3802	55.10	3.97%	0.00158	0.0000626	
71	Jul-11	55.10	48.20	51.80	2854.18	0.3642	53.45	3.81%	0.00145	0.0000553	
72	Aug-11	58.30	42.90	57.10	3328.93	0.4054	57.70	4.35%	0.00189	0.0000823	
73	Sep-11	61.30	39.50	60.50	3708.65	0.4378	60.90	6.86%	0.00471	0.0003228	
74	Oct-11	58.80	39.90	60.10	3533.88	0.4229	59.45	5.39%	0.00291	0.0001566	
75	Nov-11	59.60	41.60	58.40	3480.64	0.4184	59.00	6.99%	0.00489	0.0003415	
76	Dec-11	67.10	34.50	65.50	4395.05	0.4958	66.30	10.80%	0.01166	0.0012587	

Month	POD	FAR	SR	TPIX	CSI	Avg.	IFR Freq.
Jan-12	66.80	34.00	66.00	4408.80	0.4970	66.40	9.07%

Monthly TAF IFR Performance, 2006-2014

Scatter plot of TPIX = (1-FAR) x POD versus IFR Frequency. Jan. 2012 point highlighted in red.

$$y = -144436x^2 + 43772x + 1382.7$$
$$R^2 = 0.8816$$
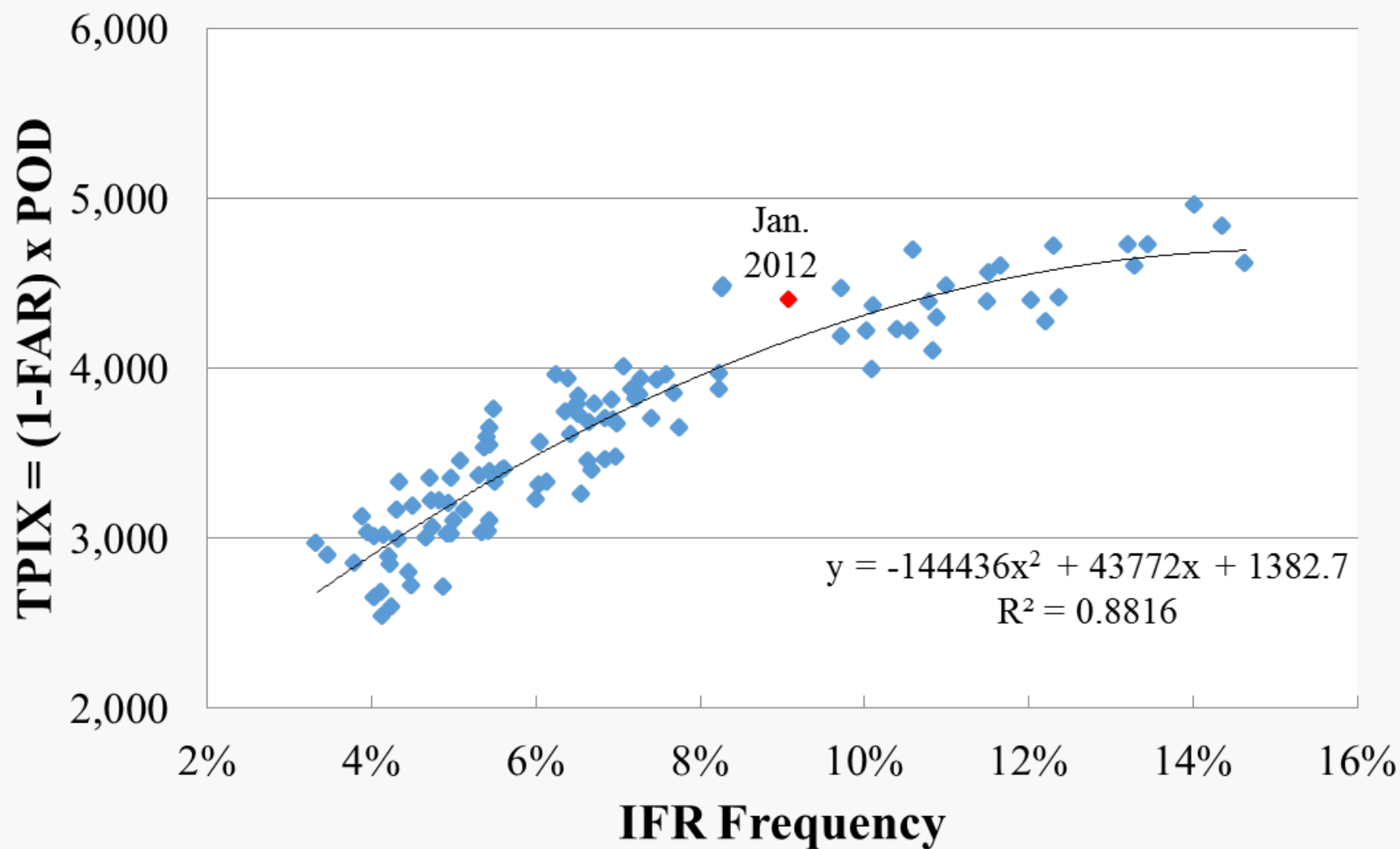
Month	POD	FAR	SR	TPIX	CSI	Avg.	IFR Freq.
Jan-12	66.80	34.00	66.00	4408.80	0.4970	66.40	9.07%

Moving Average

- TPIX residuals represent performance relative to all months →

- Month to month variability is high, standard deviation = ~200 TPIX Points

- Like stocks and commodities, single-day performance, or even monthly volatility, should not be used to make long-term investment decisions… performance is a long-term prospect, thus we should use moving averages (and 12-month lag) →

Improvement Over Predicted Performance
Nation 2005-2014

$y = 2.2228x - 122.26$

Nation TAF IFR Performance
12-Month Moving Average

$y = 2.3421x - 108.31$
$R^2 = 0.6885$

8
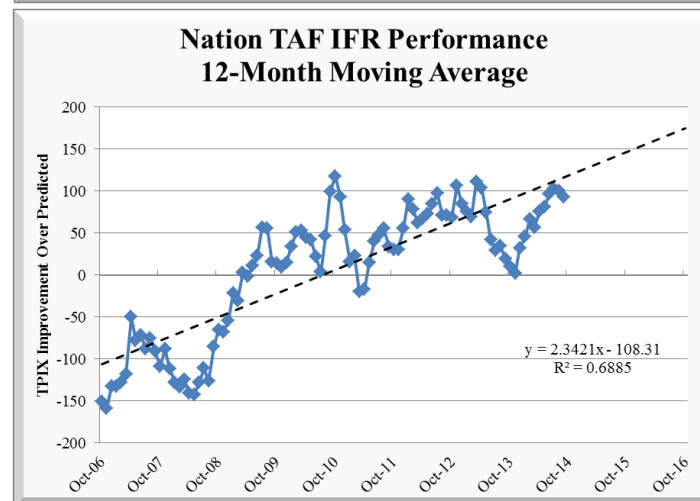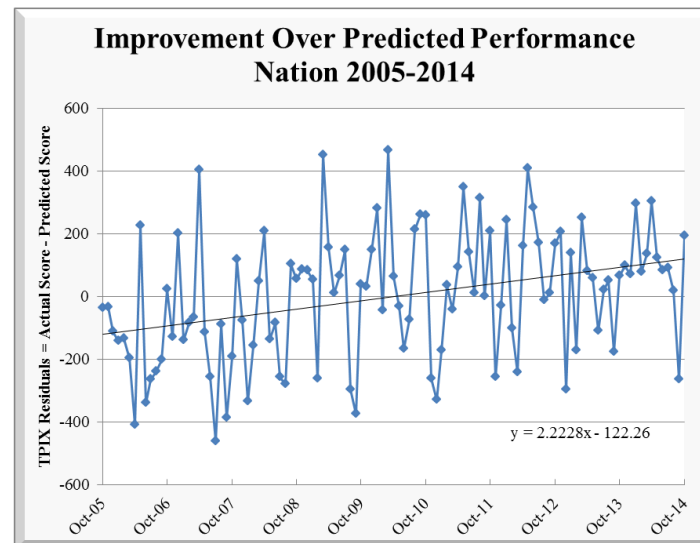
Single Location Example: ORD

- ORD exhibits a wide range of IFR Frequency, higher IFR Frequency typically in winter months

- Recent performance, according to the eight-year monthly sample, is very good in relative terms

ORD Monthly TAF IFR Performance 2006-2013

$$y = 699181x^3 - 394354x^2 + 74694x + 1073.2$$
$$R^2 = 0.6252$$

(Y-axis: TPIX; X-axis: IFR Frequency)

ORD TPIX Monthly Performance

(Y-axis: Regression Residual; X-axis: Oct-05 to Nov-12)

ORD TPIX Residual 12-Month Moving Average

(Y-axis: Average Regression Residual; X-axis: Oct-05 to Nov-12)

9
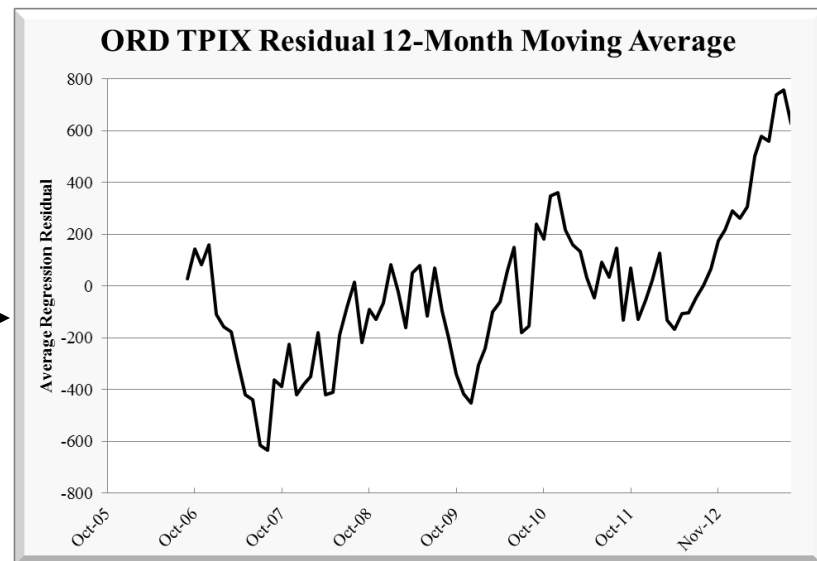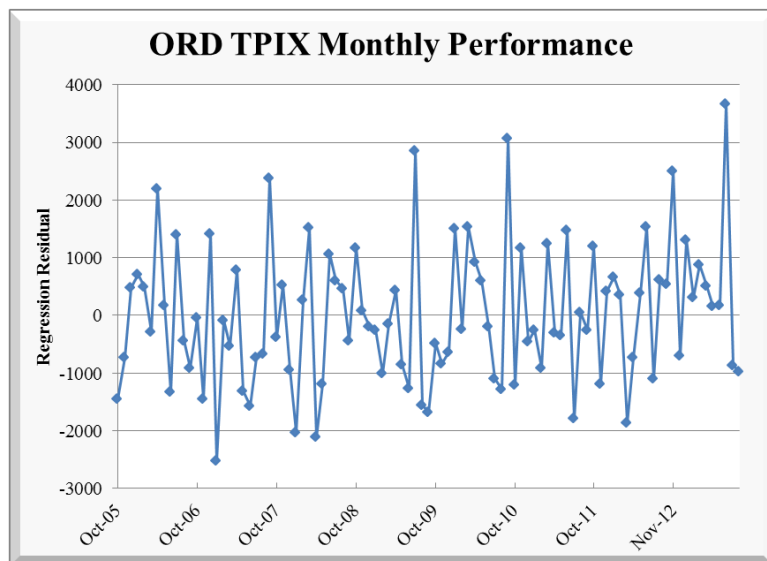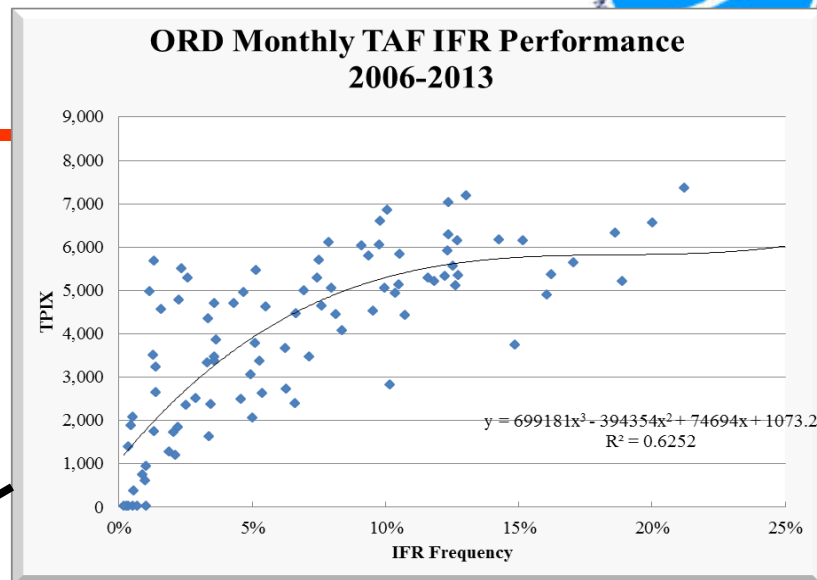
Summary

- POD and FAR must not be used individually to *summarize* performance

- Performance metrics must be normalized to account for the gross influence of event frequency

- Use a moving average to evaluate progress and trends

Details of Methodology

- Peer-reviewed article published Dec 2013

- Coordinated with Performance Branch

- Better represents actual forecast performance

Journal of Operational Meteorology
Short Contribution

Scale Normalization for IFR-Frequency Effects in Aviation Forecast Performance Statistics

MATTHEW LORENTSON
National Weather Service, Silver Spring, Maryland

ABSTRACT

The National Weather Service uses *probability of detection* and *false alarm ratio* to assess forecast performance. Statistical evidence indicates that a quantitative relationship exists between these forecast performance metrics and the frequency with which a forecasted condition occurs. Current national aviation performance goals do not account for this relationship, which reduces their utility. There is meaningful evidence that indicates the influence of low ceiling/visibility frequency on national aviation forecast performance metric averages can be neutralized through scale normalization.

1. Introduction

The United States' Government Performance and Results Act (GPRA) of 1993 requires government agencies to measure outcome-related goals for major functions and operations (Office of Management and

(henceforward termed "IFR Frequency") was related to the reduction in IFR forecast performance. The NWS Aviation Services Branch responded to this concern and investigated the relationship between IFR

http://www.nwas.org/jom/articles/2013/2013-JOM22/2013-JOM22.pdf

Questions

Thanks to Kevin Stone at Aviation & Space Weather Branch for input on this presentation

Backup Slides

Total Performance Index

- Total Performance Index (TPIX)

  POD = Probability of Detection

  FAR = False Alarm Ratio
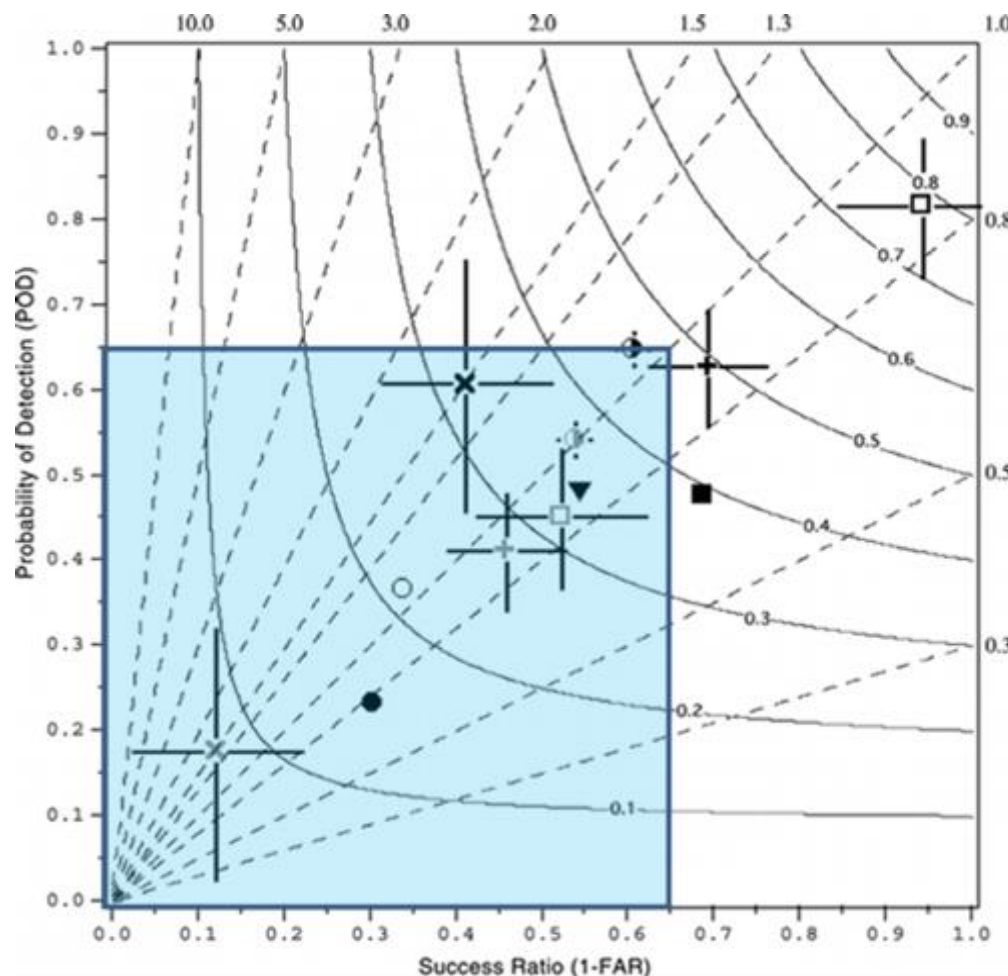
  1-FAR = Success Ratio (SR)

$$TPIX = POD * (1-FAR)$$

- Comparable to Critical Success Index (CSI) but easier to understand and calculate from POD and FAR

$$CSI = \frac{1}{[(100/POD)+(100/1-FAR)-1]}$$

Geometric Relationship*: POD, SR, CSI, TPIX, and Bias

TPIX can be visualized on this graph as a quadrilateral area calculated by multiplying Probability of Detection by Success Ratio.

- Dashed lines = Bias (POD/SR)
- Solid contour = CSI.
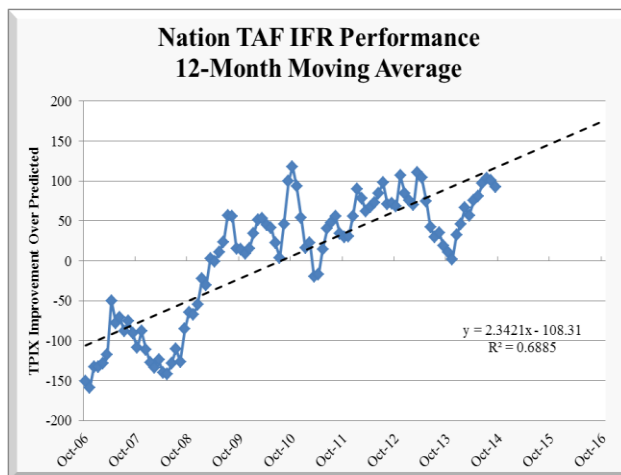- Blue square = TPIX example

Using whole numbers, the blue square area with POD and SR scores of 65 (Bias = 1.0) produces a TPIX of 4225 and CSI of ~4815. TPIX and CSI are maximized in the form of a square when bias = 1.0

Cross and shape figures represent various forecast averages discussed by Roebber; half circles represent TAF and MOS TAF averages (MOS in gray).
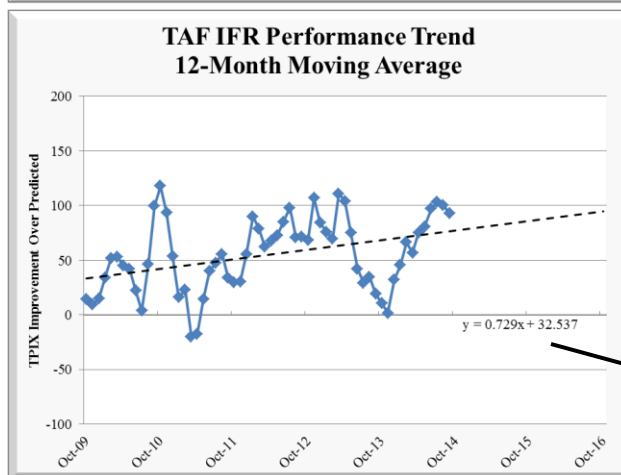
–Roebber, P., 2009: Visualizing Multiple Measures of Forecast Quality. Wea. Forecasting, 24, 601—608. [Available online at: http://journals.ametsoc.org/doi/full/10.1175/2008WAF2222159.1]

New GPRA Metric:
Improvement Over Predicted Score

2006-2014

Nation TAF IFR Performance
12-Month Moving Average

$y = 2.3421x - 108.31$
$R^2 = 0.6885$

2009-2014

TAF IFR Performance Trend
12-Month Moving Average

$y = 0.729x + 32.537$

Fiscal Year Improvement Over Predicted				
12 Month Moving Avg.				
IFR GPRA Goals (based on 2009-2014 trend)				
FY (Oct-Sep)	Month #	Predicted	Actual	Performance
2010	12	41.29	99.36	58.07
2011	24	50.03	33.96	-16.07
2012	36	58.78	71.55	12.77
2013	48	67.53	19.15	-48.38
2014	60	76.28	92.85	16.57
FY (Oct-Sep)	Month #	Goal	Actual	Performance
2015	72	85.03		
2016	84	93.78		
2017	96	102.52		
2018	108	111.27		
2019	120	120.02		
2020	132	128.77		

0.7290379(month #) + 32.537062

16

New GPRA Metric:
Improvement Over Predicted Score

TAF IFR Performance Trend
12-Month Moving Average

y = 0.729x + 32.537

2009-2014

Fiscal Year Improvement Over Predicted				
12 Month Moving Avg.				
IFR GPRA Goals (based on 2009-2014 trend)				
FY (Oct-Sep)	Month #	Predicted	Actual	Performance
2010	12	41.29	99.36	58.07
2011	24	50.03	33.96	-16.07
2012	36	58.78	71.55	12.77
2013	48	67.53	19.15	-48.38
2014	60	76.28	92.85	16.57
FY (Oct-Sep)	Month #	Goal	Actual	Performance
2015	72	85.03		
2016	84	93.78		
2017	96	102.52		
2018	108	111.27		
2019	120	120.02		
2020	132	128.77		

0.7290379(month #) + 32.537062 = Goal
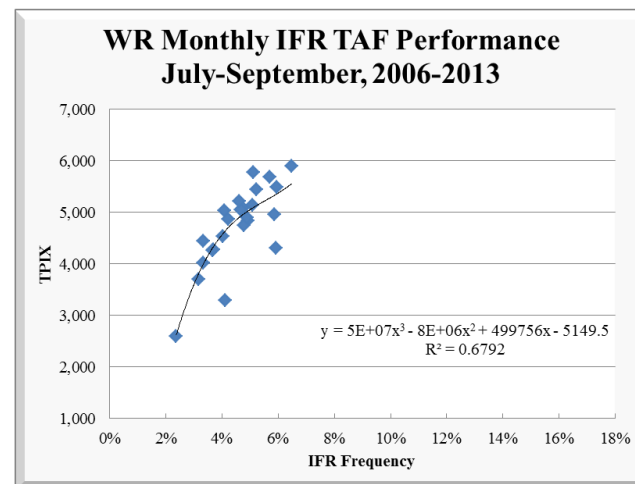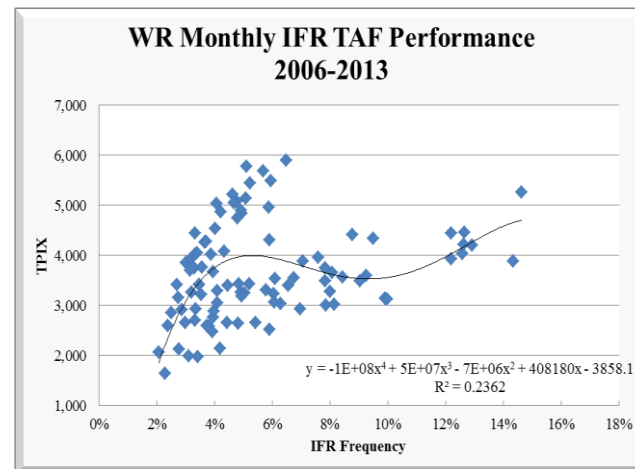Performance trend over time

Western Region Climate

- Artifacts of performance regimes are evident in large-samples

  ⟶

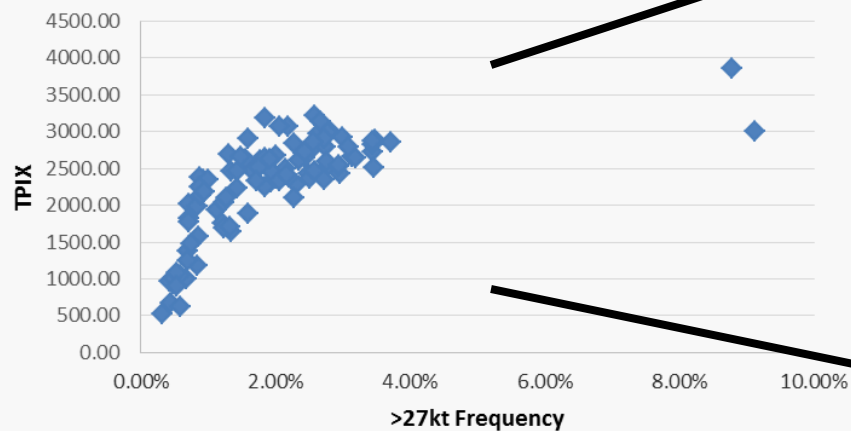- Western Region: significant diversity in its climate profile—coastal vs. mountain

- During stratus season, IFR conditions appear easier to forecast…thus seasons, and different climate regimes, should be measured separately…
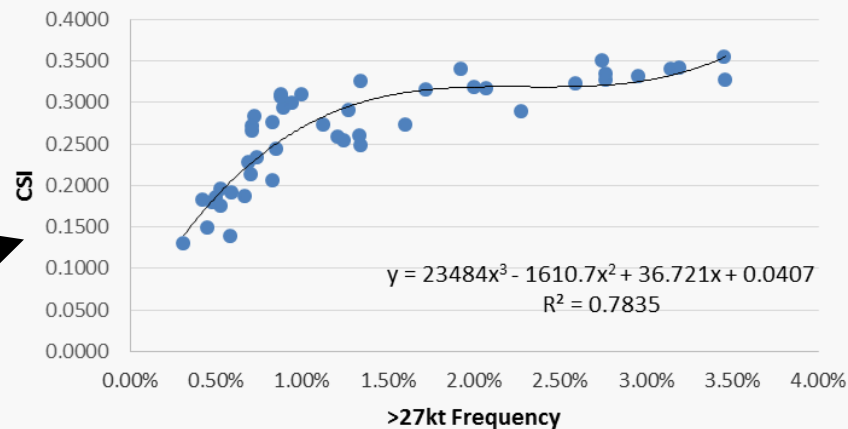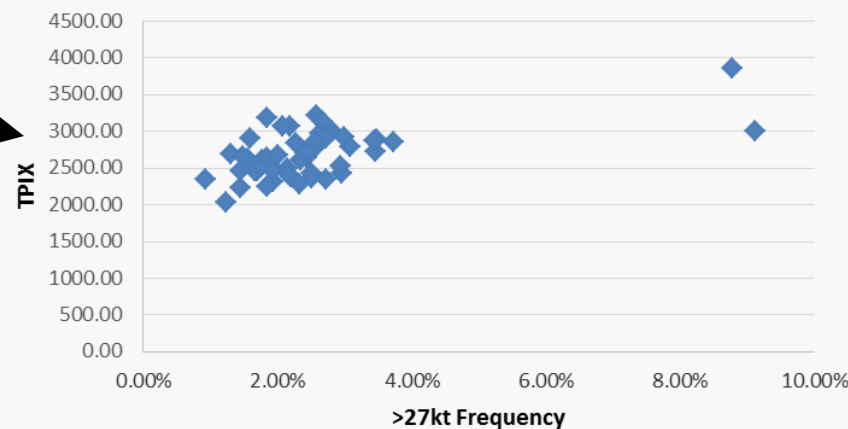
  ⟶

18

Wind Gust >27kt

Gust >27kt

Gust >27kt April - September

$y = 23484x^3 - 1610.7x^2 + 36.721x + 0.0407$
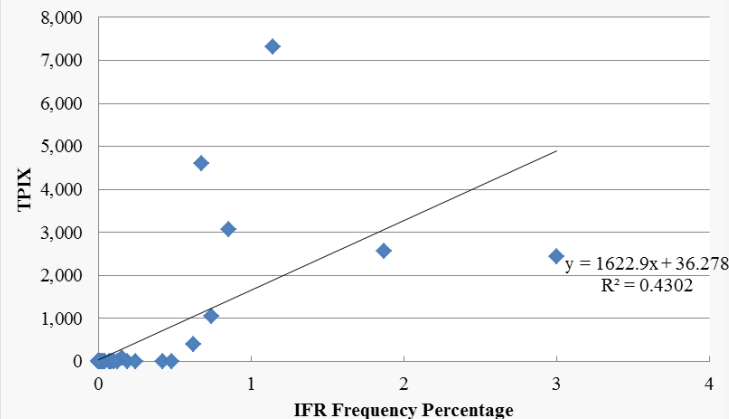$R^2 = 0.7835$

Gust >27kt October - March

Single Location Example: LAS

- LAS exhibits low IFR Frequency, thus an unusable sample

- Recent performance, according to the eight-year monthly sample, is very good in relative terms

LAS Monthly IFR TAF Performance 2006-2013

$y = 1622.9x + 36.278$
$R^2 = 0.4302$

(y-axis: TPIX; x-axis: IFR Frequency Percentage)

LAS Monthly <5k Freq. Performance 2006-2014

$y = 4E+06x^3 - 1E+06x^2 + 130964x + 63.203$
$R^2 = 0.6896$

20