Western Region Technical Attachment
NO. 04-06
July 14, 2004

The Prediction of Minimum Overnight Visibilities at the Fresno-Yosemite International Airport Utilizing Multiple Linear Regression

Mark Burger, WFO San Joaquin Valley, Hanford, CA
May 12, 2004

Introduction

a. Overview. Fresno-Yosemite International Airport (KFAT), located in the city of Fresno, California, is the principal commercial airport for the San Joaquin Valley, home to some 3.3 million people. Even with the ever-present threat of dense fog during the winter months resulting in numerous delayed or canceled flights, more than one million passengers passed through the airport during 2003, making this facility an especially important and notoriously challenging Terminal Aerodrome Forecast (TAF) point. Objective guidance has proven largely inadequate in the timing and severity of fog in the vicinity of this airport, thus providing the impetus for developing additional and improved measures, both qualitative and quantitative, in the forecasting of this fog.

b. Methodology. A previous technical attachment written by the author demonstrated a moderate relationship between dense fog events on the San Joaquin Valley floor and the 700 mb height level. As an example, for observations comprising 120 days from November 2002 through February 2003, at two separate valley locations, the Pearson correlation between widespread dense fog events and the 700 mb height level at Oakland, California (KOAK) stood at +0.46 (Burger 2003). Based on these encouraging results, this study seeks to expand the sampling time from the previous study, along with supplementing the upper air data with surface observations in order to increase the predictive capability of the previous study. Similar to the previous study, since no routine upper air sounding is available at KFAT, a site 153 miles upstream of the Fresno airport, KOAK, was utilized for the collection of 700 mb data. Lower height levels from this sounding are unrepresentative as mountains between Oakland and Fresno disrupt and modify the air flow and a marine environment at KOAK contrasts sharply with the near-surface atmosphere of the San Joaquin Valley. For additional information regarding the data collection process, results, and assumptions included herein, please refer to the Appendix at the end of this study.

c. Data. Automated Surface Observing System (ASOS) hourly data are compiled and archived locally and were utilized for this study. Specifically, observations from KFAT for December and the following January of each year from 1998 through 2003, inclusive, were incorporated as the first objective data set. From these observations, the 1600 local standard time (LST) surface dewpoint depression and surface horizontal visibility were noted. A late afternoon sampling time was desirable since it would reflect recent meteorological changes induced by daylight, yet maintain some predictive capability so that advance warning of fog events for the upcoming night could be assured. KFAT ASOS observations from the following night, 0000 LST through 0800 LST, were noted, and the lowest visibility attained was recorded, with both hourly and special observations considered. The second data set, the 12Z (0400 LST) radiosonde observations from KOAK, were also retained for the aforementioned period, with the 700 mb height level recorded. The morning upper height level and the lowest attained surface visibility at KFAT were then paired with the surface observations from the preceding afternoon to help gauge the predictive capability of all upper air and surface elements. Once all relevant data were gathered, a multiple linear (least-squares) regression analysis was performed, with the independent variables being the 700 mb KOAK height, 1600 LST surface dewpoint depression at KFAT, and 1600 LST horizontal surface visibility at KFAT; the dependent variable was the lowest visibility actually achieved during the 0000 LST to 0800 LST time frame. Suspicious or uncharacteristic data were excluded from the study; likewise, if any required surface elements were not present at the specified observation time, or radiosonde observations were missing for the 12Z sounding, the affected day was removed from the study. In all, 276 days constitute the data available for this study.

Results

a. Statistics. Table 1 includes basic descriptive statistics regarding observed upper heights, dewpoint depressions, and surface visibilities. Table 2 compares actual lowest visibility data with lowest predicted visibility data derived from the linear regression technique. The Pearson correlation between observed lowest visibilities and those suggested by linear regression, utilizing the same data, is +0.61.

b. Formulas. Although several different versions of this formula were derived from linear regression techniques, each utilizing different inputs and initial conditions, the following formula appears to provide the most plausible determination of expected lowest visibility:

$$\mathbf{V_L = 45.99133 - 0.01493H + 0.132621T_{dd} + 0.299727V}$$

where

V_L is the lowest predicted surface visibility (statute miles) at KFAT for the upcoming night
H is the 700 mb height level (meters) from the most recent 12Z KOAK radiosonde observation
T_{dd} is the surface dewpoint depression (degrees celsius) at 1600 LST from KFAT
V is the surface visibility (statute miles) at 1600 LST from KFAT

As the equation illustrates, there is an inverse relationship between upper height and lowest predicted visibility, with a positive relationship between the other independent variables and lowest predicted visibility. In both situations, the objective analysis supports subjective reasoning; that is, as upper heights climb, it would be expected that the lowest predicted surface visibility should decrease as upper level subsidence warms temperatures aloft and increases the strength of the low level inversion. As the surface dewpoint depression and afternoon surface visibility increase, it seems logical that the lowest predicted surface visibility overnight should also increase.

Further analyses were also undertaken; dates with large standardized residual values and those with statistically large influences were scrutinized and, through a series of iterations, progressively eliminated from the data set. The result was an equation which provided the highest Pearson correlation possible between lowest predicted visibility and lowest attained visibility, +0.69, given the constraints of the selected independent variables and scope of data. This equation is referred to as the modified equation, and is depicted below:

$$\mathbf{V_{LM} = 17.5 - 0.00575H + 0.155T_{dd} + 0.0898V}$$

where

V_{LM} is the modified lowest predicted visibility (statute miles) at KFAT for the upcoming night and all other variables are defined as previously depicted.

Important to note is that the latter equation dismisses 86 dates, leaving 190 cases available to compute the regression curve. Many of the eliminated cases were those where predicted visibilities were relatively high, as noted by the large drop in the three measures of central tendency (mean, median, and mode) demonstrated between the original equation and the modified equation; a summary of these data is available in Table 2 and Table 3. For both tables, statistics are calculated for all 276 cases (although the modified equation was developed using 190 cases, the resultant equation was applied to all dates to develop descriptive information).

Since the modified equation is biased towards the less anomalous, as well as lower, predicted visibility values, this equation is not especially recommended in the forecasting of lowest overnight visibility values, despite the improved correlation. Rather, since the correspondence between the lowest predicted visibility and the lowest visibility actually attained is higher, this equation would consequently represent an improvement for forecasting visibility trends from one night to the next. To help facilitate this end use, and to discourage comparisons to actual visibility measures, it is suggested that output from the modified equation be converted to standardized anomaly (Z) scores using the following transformation (Wilks 1995):

$$Z = (X – X_M) / S_x$$

where

Z is the standardized anomaly
X is the numerical input
X_M is the mean value of the numerical input data set
S_x is the standard deviation of the numerical input

Thus, substituting appropriate variables from the modified equation into the standardized anomaly formula:

$$\mathbf{Z = (V_{LM} - 1.35) / 0.82}$$

where

Z is the standard score representing a relative lowest predicted visibility from the modified equation
V_{LM} is the modified lowest predicted visibility (statute miles) at KFAT for the upcoming night

Of course, with this configuration, the standard deviation is 1.00; the minimum value attained from applying this conversion to the 276 cases is -2.34, while the maximum value is +2.95. Although the distribution is not normal, it does show a low degree of skewness; thus, percentiles can be calculated as a rough approximation of the severity of fog in relation to the normal distribution. Based upon applying the Z-score transformation to all days, scores of -2 or lower always produced actual minimum visibilities of a mile or less, with most days witnessing visibilities falling below one-quarter mile; for scores of +2 or higher, actual visibilities never fell below 3 miles.

c. Lowest Observed Visibility in Relation to the Independent Variables. Figure 1 shows a scatter plot of lowest visibilities experienced during the morning (vertical axis) along with the 700 mb height associated with the occurrence (horizontal axis). Several regression techniques were then performed, utilizing linear, quadratic, and cubic functions. Although the third-degree polynomial indicated the best correlation (function plotted), it performed only marginally better than its counterparts. The tendency is for morning visibilities to be highest when 700 mb heights were around 2950 meters, with the expected general downward trend in visibility as heights increase. Interesting to note, however, is the slight increase in visibility accompanying the highest heights. Although this may purely be the result of sparse data, it is conceivable that as the highest upper level heights are achieved, very strong subsidence aloft helps squelch the low level inversion, providing for greater visibility than may otherwise be expected.

Figure 2 depicts the relationship between lowest visibility observed during the morning (vertical axis) and the dewpoint depression from 1600 LST the previous afternoon (horizontal axis). Once again, the cubic function provided slightly higher correlation than either linear or quadratic methods, but only slightly so. However, as can be seen from the plot, this correlation is very weak and provides only a minimal influence on the equations put forth formerly.

Figure 3, a chart of the lowest visibility observed during the morning (vertical axis) in relation to the horizontal surface visibility observed at 1600 LST the previous day (horizontal axis), also was best fitted by a cubic function. But, as in all cases, linear regression techniques proved to be similarly correlated. Important to note is the congregation of cases at the lower left, where low afternoon visibility almost certainly implies low morning visibility. However, relatively high surface visibility during the afternoon does not necessarily translate into higher minimum visibilities overnight, as evidenced by the large spread of lowest attained visibilities at vis = 10 on the graph.

Summary

Although a linear assumption represents the least mathematically complicated technique to determine the relationship between a set of variables, such a method provides substantively comparable predictive capability for fog at the Fresno airport as higher-degree methods. The late afternoon surface dewpoint depression and horizontal visibility, along with the following morning's 700 mb height level, constitute the basis for developing the various equations set forth herein. When these parameters were compared to the observed lowest morning visibility at KFAT, and the output was statistically screened (the modified lowest predicted visibility equation), the Pearson correlation reached +0.69, indicating a good relationship between selected surface and upper air data and the tendency for fog. This correlation represents a substantial improvement over the previous study, and also affords a more quantitative measure of the likely fog severity for the upcoming night. By employing the standardized anomaly equation, an unbiased method for predicting minimum visibility trends from night-to-night emerges; the numerical output can be compared to the mean and standard deviation developed from a large sample, as well as to the output generated from previous nights, to determine appropriate forecaster response and lead to improved forecasting of dense fog for the Fresno TAF.

Acknowledgments

Thanks to Larry Greiss and Western Region's Scientific Services Division for reviewing this study.

References

Burger, M., 2003: The Application of Upper Level Heights in Diagnosing and Forecasting San Joaquin Valley Dense Fog Episodes. National Weather Service, Western Region Technical Attachment No. 03-09.

Wilks, D.S., 1995: Statistical Methods in the Atmospheric Sciences, Academic Press, Inc. pp. 42.

Appendix : Additional Considerations

a. Data Collection. As stated in the Introduction, only cases from December and January were examined for the purposes of this study. These months were chosen since they occur during the middle of the local "wet season," when boundary layer moisture is readily available and abundant. Given the relatively short daylight hours and low solar insolation during this time of year, the potential for pronounced and severe dense fog is at a maximum. Additionally, an attempt was made during the course of this study to isolate overnight conditions which supported "purely radiative" (i.e., clear) opportunities and differentiate those events from ones where clouds prevented substantial overnight radiative losses. However, maintaining objective parameters as to what constitutes good radiative conditions proved to be a formidable challenge, with the best attempts resulting in correlations much lower than when all observations were considered. Finally, although surface winds and pressure gradient were initially considered as being possible independent variables for this study, the hourly deviation for winds tended to be too great to be considered a reliable predictor; likewise, surface pressure data for selected hours and locations proved to be too variable and too often unavailable.

b. Data Analysis. Regression equations, figures, and descriptive statistical data were derived from Minitab Incorporated's Minitab 14 Demonstration application.

c. Equation Results. Since the regression technique involves fitting a line to the entire data set, some predicted visibilities will come out as negative numbers, which is an obvious fallacy in relation to ground truth. In this case, the interpretation should be that the independent variables input suggest an extremely high propensity for dense fog. Another characteristic to consider is that an observed 1600 LST surface visibility of 10 statute miles may imply a visibility of just that, or may, in fact, suggest much higher visibilities (such as 50 miles), which simply can not be reported by existing ASOS equipment. Thus, even though the formula treats both cases exactly the same, in reality, the differences in actual visibility between these two cases may prove substantially high as to seriously undermine the effectiveness of the formula on such days.

d. Other Assumptions. Since the input variables include parameters from the previous afternoon, the formulas are not likely to accurately predict the first night of a sustained fog episode. At the surface, the late afternoon visibility tends to be relatively high preceding the first night of fog, and the dewpoint depression may still be low owing to afternoon turbulent mixing and advection of drier air aloft in the wake of frontal passage. In an operational use of the equations, the 700 mb heights input are those from earlier in the day, and thus may not accurately reflect the rapid building of upper level heights which typically occurs as a multiple-day fog event begins. As an aside, through daily testing of the equations in January 2004, it appears the formulas have some skill in predicting the extensiveness and longevity of fog across other portions of the San Joaquin Valley, even though the equations were not specifically designed for that purpose. However, no quantitative data have as yet been developed to show as such.

Tables and Figures

Table 1. Descriptive statistics for the independent variables of the Fresno fog equation.

Element	Mean	Median	Mode	Standard deviation	Minimum	Maximum
12Z KOAK 700 mb height (m)	3091.00	3095.50	3076.00	69.23	2888.00	3215.00
1600 LST KFAT dewpoint depression (c)	6.95	7.00	8.00	3.89	0.00	19.00
1600 LST KFAT surface visibility (sm)	6.17	6.00	10.00	3.31	0.25	10.00

Table 2. Comparison of observed lowest visibility with that predicted by the Fresno fog equation.

Element	Mean	Median	Mode	Standard deviation	Minimum	Maximum
Observed lowest visibility (sm)	2.60	1.75	0.00	2.85	0.00	10.00
Predicted lowest visibility (sm)	2.62	2.63	2.76	1.74	-1.05	6.29

Table 3. Descriptive statistics for the modified Fresno fog equation.

Element	Mean	Median	Mode	Standard deviation	Minimum	Maximum
Modified predicted lowest visibility (sm)	1.35	1.38	0.90	0.82	-0.57	3.77

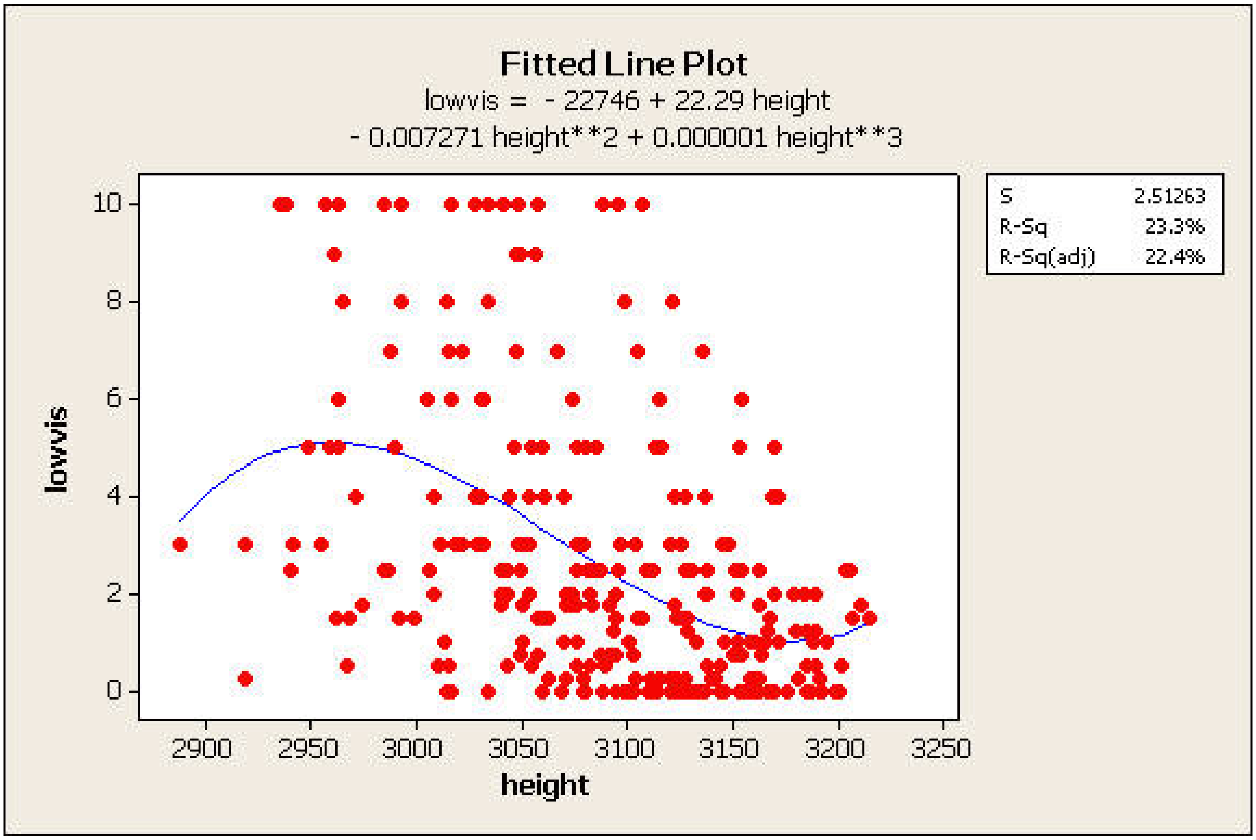
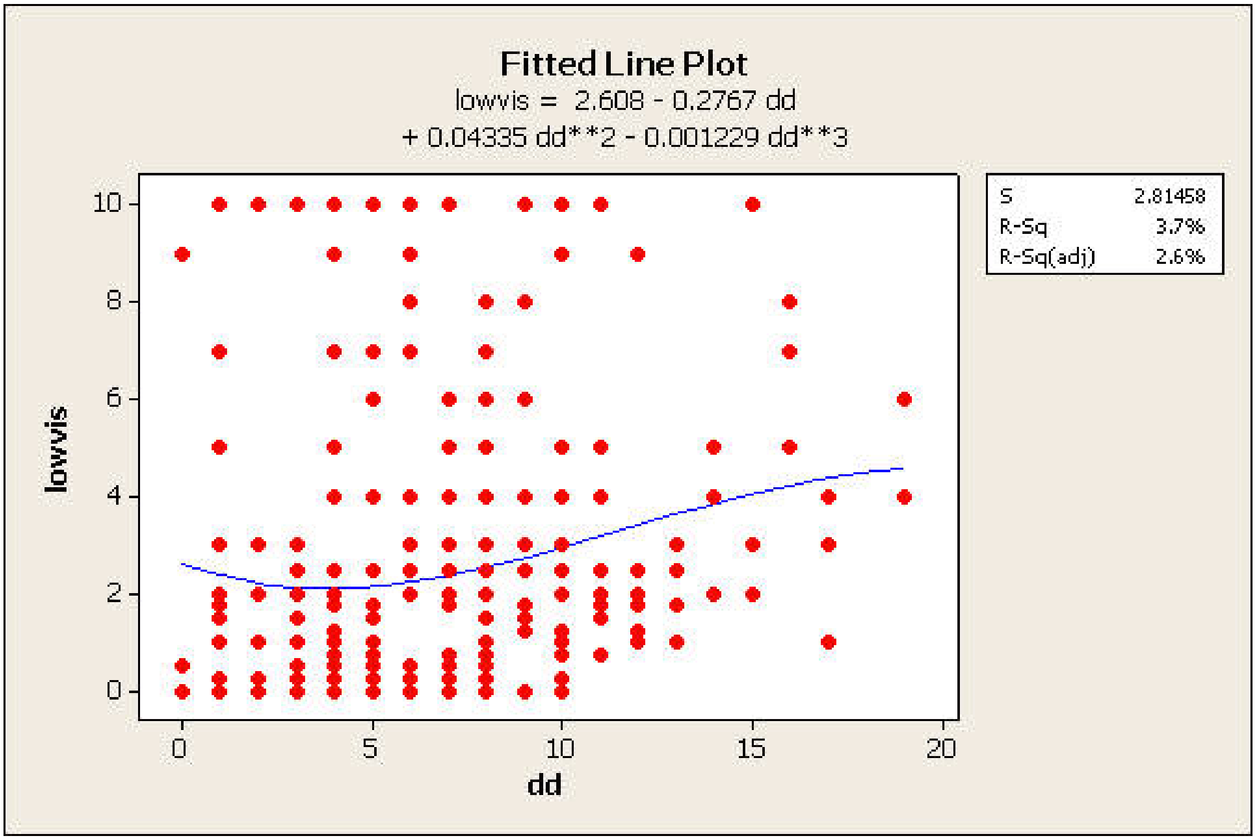

Fitted Line Plot
lowvis = - 22746 + 22.29 height
- 0.007271 height**2 + 0.000001 height**3
S 2.51263
R-Sq 23.3%
R-Sq(adj) 22.4%
lowvis
height
0
2
4
6
8
10
2900
2950
3000
3050
3100
3150
3200
3250

Figure 1



Fitted Line Plot
lowvis = 2.608 - 0.2767 dd
+ 0.04335 dd**2 - 0.001229 dd**3
S 2.81458
R-Sq 3.7%
R-Sq(adj) 2.6%
lowvis
dd
0
2
4
6
8
10
0
5
10
15
20

Figure 2

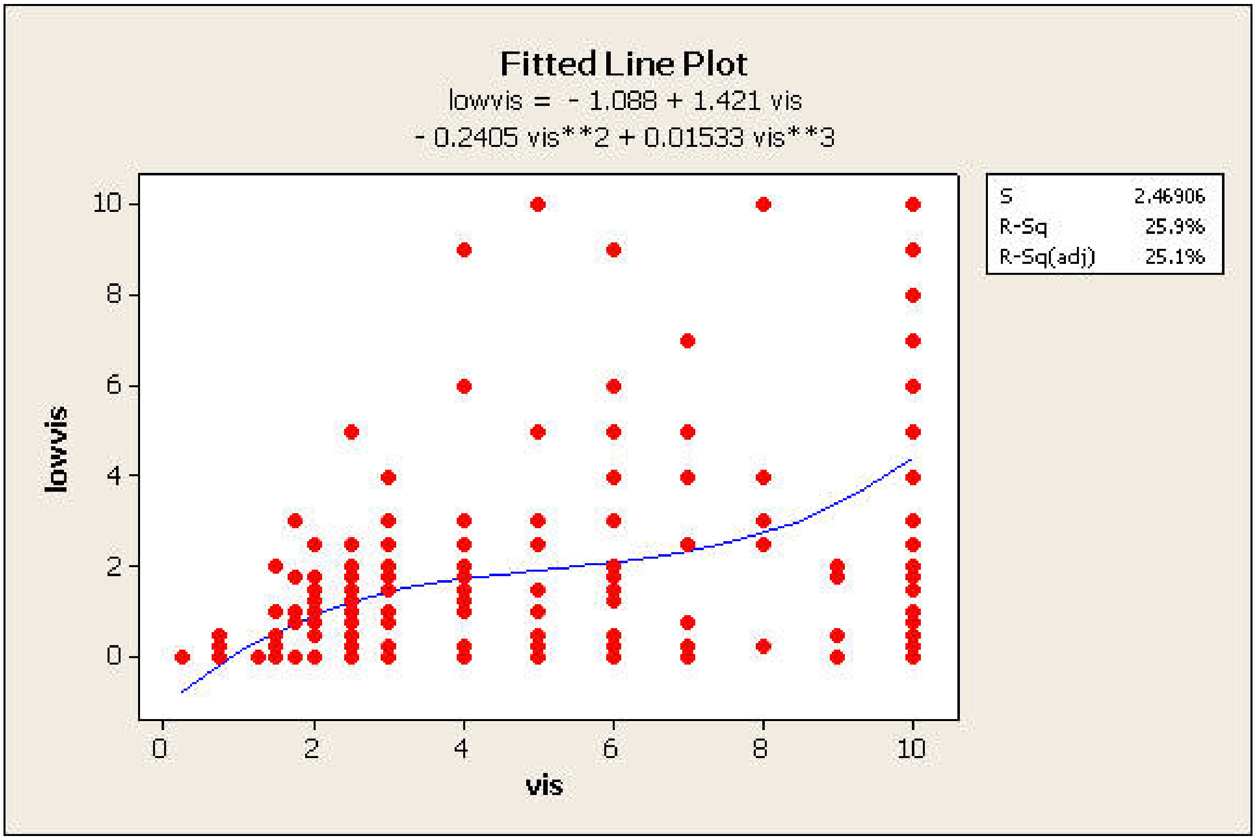

Fitted Line Plot
lowvis = - 1.088 + 1.421 vis
- 0.2405 vis**2 + 0.01533 vis**3
S 2.46906
R-Sq 25.9%
R-Sq(adj) 25.1%
lowvis
vis
0
2
4
6
8
10

Figure 3