FORECASTING OPEN OCEAN FOG AND VISIBILITY [1]

Lawrence D. Burroughs
Ocean Products Center, NMC

1. INTRODUCTION

A large percentage of the accidents at sea occur with visibilities under one kilometer (Tremant, 1987). Although other obstructions may lower visibility below 1 km, the most prevalent obstruction is fog. Fog is only reported if the visibility is 1 km or less. Until now there has been no objective guidance available to National Weather Service forecasters for fog or lowered visibilities. The Open Ocean Fog and Visibility Forecast Guidance System was developed to fill this need. This system is designed to provide fog and visibility guidance over the North Pacific and North Atlantic during the prime season for fog and lowered visibilities (April - September). The guidance is not applicable to coastal areas.

2. DEVELOPMENT

The perfect prog technique (Klein *et al*, 1959) was used to develop the fog and visibility forecast equations. With this approach all data used in the development of relationships are analyzed or observed data. Usually, the predictor and predictand are concurrent in time. When the equations are used to predict, forecast values of the predictors must be obtained and substituted into the equations to give a forecast of the predictand. The name perfect prog comes from the fact that forecast data are entered into the equation as if it were equivalent to the analyzed data that were used to develop the equations (Wilson and Macdonald, 1985).

2.1 Predictand Data

The predictand data were taken from ship data for the years 1980 - 1983 at 12 hour intervals. Fog data included observed fog of any kind and observed drizzle when the past weather indicated fog and the visibility was less than 1 km. The fog was categorized as no fog or fog. Visibility data were corrected to be consistent with the observed weather and the WMO code and include poor visibility due to fog. Visibility is categorized to delineate areas with a visibility of less than or greater than 3 n mi.

2.2 Predictor Data

The predictor data came primarily from the analyzed fields of the Global Data Assimilation System (GDAS) (Kistler and Parrish, 1982 and Dey and Morone, 1985) or computations made by using the GDAS data. In addition a boundary layer diagnostic model was used to create air temperatures and equivalent potential temperatures at 19 m (the nominal height of most ship instrumentation). In all, 20 basic GDAS fields and 45 computed fields were used plus four climatological parameters and two location parameters. The data were interpolated to the location of the ship data to derive the forecast equations.

2.3 Regions

Burroughs (1987) describes the development of open ocean fog forecasting regions based on a National Climatic data Center fog climatology (Guttman, 1971), the frequency of fronts over the open ocean and the frequency of high and low centers over the open ocean. Thirteen warm and 13 cool season regions were delineated. These regions were then modified by the inclusion of drizzle and observed fog data for 1980 - 1984. The number of regions was reduced to those shown in Figs. 1 and 2 after test equations were derived and evaluated, and the results showed the equations for the regions excluded to be too unreliable to use.

2.4 Discriminant Analysis

Equations were developed for fog and visibility with discriminant techniques. A short discussion on these techniques, following Tatsuoka (1971), is given below. For additional information see Anderson (1958), Miller (1962), or Cooley and Lohnes (1971).

In meteorology, categorical events are often related to continuous predictors by the use of discriminant analysis. The events are grouped by category. Then equations are developed by using the continuous predictors which best separate the groups.

Having developed the equations, the next question is how to classify a particular observation (vector of predictor values) into a particular group. The objective is to minimize the number of misclassifications. The chi-squared statistic is used for this purpose. The smaller the value of chi-squared for a particular group the nearer the observation is to the average value for that group. Since the equations for each group differ only by the coefficients and constants used, the observation is categorized into the group whose equation has the smallest chi- squared value. This assumes that the groups have

[1] OPC contribution no. 34

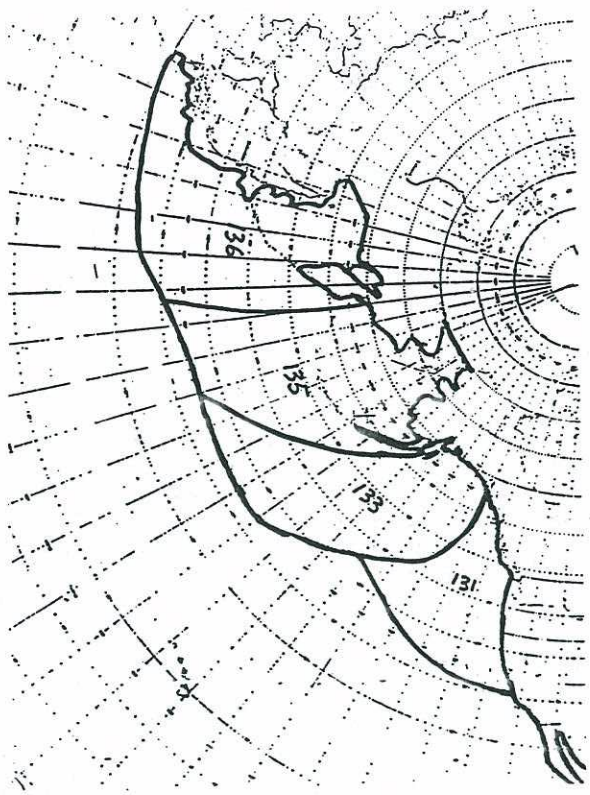

Figure 1: Warm season fog and visibility regions for the North Pacific.

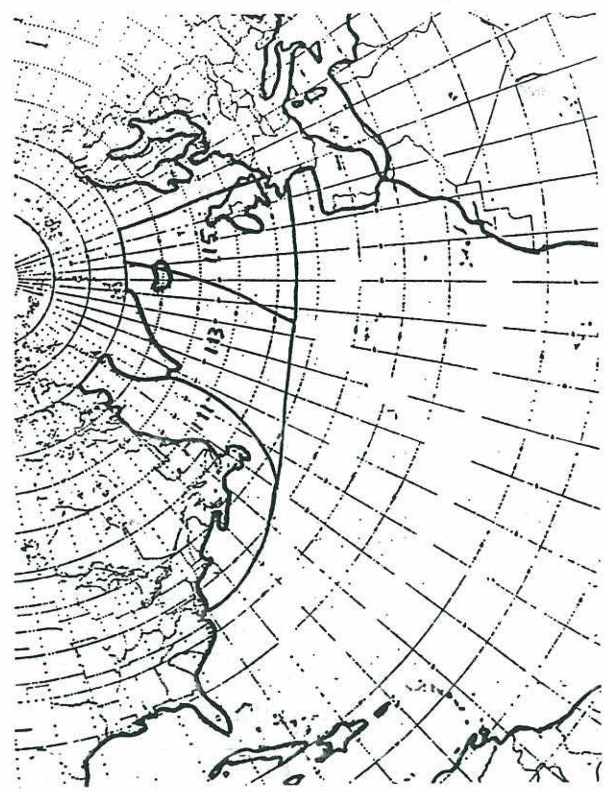

Figure 2: Warm season (April - September) fog and visibility regions for the North Atlantic.

multivariate normal distributions with equal dispersion matrices, and, therefore, that a pooled covariance matrix can be used to compute chi-squared for each group.

When the covariance matrix cannot be pooled, the group equations are corrected for the differences in dispersion by using the individual group covariance matrices.

Thus far the assumption has been that the relative frequencies of all the groups are equal. For fog and visibility, this is not the case. Chi-squared can be further adjusted to take into account the prior probabilities (relative frequencies of group membership) which have been determined from the dependent predictand samples.

We have already assumed that the group dispersions are multivariate normal; if we further assume that an observation fits into one of the groups and does not fall outside the ensemble of groups, then chi-squared can be related to the posterior probability that an observation belongs to a particular group. The event forecast is assigned to the group whose posterior probability is greatest.

2.5 Inflation and Thresholding

If the discrimination was perfect, no further adjustments would be necessary. This, generally, is not the case, and two other procedures can be used to help minimize the number of misclassifications: inflation (Klein *et al*, 1959 and Miller, 1988) and thresholding.

Inflation adjusts the forecasts so that the variance of the forecasts and the variance in the climatology are approximately equal. This procedure is applied to the fog and visibility posterior probabilities prior to classification.

Thresholding means that a given category may not be chosen unless a certain predictand value is reached (in this case a given inflated posterior probability forecast). These values are normally determined empirically. The thresholds for both the fog and visibility were determined by an iterative procedure which optimized the equations for each projection, so that for each catagory the bias was as close to 1.0 as possible. (A value of 1.0 means the forecasts in a given catagory are neither over nor under forecast.) By optimizing the equations for each projection, the thresholds become model dependent. Thresholding partially accounts for differences between the developmental data set and the model output being used as input to the equations.

When a model changes significantly, or the equations are used with output from a different model, the thresholds must be recomputed. In a sense the equations, through the use of this kind of thresholding, become MOS (model output statistics)-like. Because of the periodic necessity to recompute the thresholds, there is a cost in time. But this cost is less than that required to redevelop the entire equation set periodically as would be necessary with MOS, and a stationary numerical model (one that is not greatly changing over long periods of time) is not required. The benefits of this type of thresholding in terms of better forecasts outweigh the additional cost in time to compute the thresholds.

Thresholding is applied during the classification of the group forecasts after the posterior probabilities have been inflated.

2.6 Equation Development

To determine which variables to use as predictors, a stepwise discriminant analysis procedure was employed. The predictand was a classification variable which defined the groups with which the predictors were to be associated. Potential predictors were chosen to enter or stay in the discriminant model according to specified criteria for the squared partial correlation of the predictor with the classification variable, accounting for the effects of the predictors already selected.

The procedure began by selecting the variable that contributed most to the discriminatory power (group separation) of the equations and met the criterion to enter (a partial squared correlation of 0.0005). At each succeeding step, if a variable already selected failed to meet the criterion to stay (a partial squared correlation of 0.0025), it was removed. Otherwise, the variable, not already selected, that contributed most to the discriminatory power of the equations and met the criterion to enter was added. When all variables selected met the criterion to stay, and none of the remaining variables met the criterion to enter, the stepwise selection stopped (SAS, 1982). This procedure was used to determine which predictors to use in the final equations and provided the canonical correlation used in the inflation procedure. This procedure assumes that the variables are multivariate normal with equal dispersions.

Once the predictors for the fog and visibility equations were chosen, a second procedure was used. This procedure derived the final group equations which took into account the differences in dispersion matrices between the groups and the group prior probabilities. These equations produce the posterior probabilities for each group. The posterior probabilities are then inflated, and a threshold is applied.

To summarize the development sequence:

1. A stepwise discriminant analysis was performed to determine which predictors to use in the group posterior probability equations.
2. A second discriminant procedure was used, taking into account the group dispersions and prior probabilities, to develop the equations to forecast the group posterior probabilities, and
3. Thresholds were developed for each equation at each projection hour.

There are two equation sets for fog and visibility in each region shown in Figs. 1 and 2: one for the 0000 UTC cycle and one for the 1200 UTC cycle. The reason for this is that the amount of predictand data and its quality varies between cycles. There is always less data available on the night cycle than the day cycle. The reason is that ships report fewer night observations than day observations. Also the day observations tend to be better than the night observations. In the Pacific the day observations occur on the 1200 UTC cycle, while the night observations occur on the 0000 UTC cycle. In the Atlantic the reverse is true although the number of observations per cycle are more balanced than in the Pacific.

2.7 Equation Application

Equations are applied to the NMC Aviation Model (Sela, 1988). The forecast procedure includes:

1. Computing the group posterior probabilities from an observation vector which contains the values of each predictor at a given point on the grid for a given model projection.
2. Inflating the posterior probabilities, and
3. Classifying the posterior probabilities into a particular group with the use of the thresholds.

Forecasts are made at 24-h intervals for projections from 24 through 72 hours. At present only the 24- and 48-h forecasts on the 0000 UTC cycle are used on the Marine Significant Weather Chart (NWS, 1988). Eventually the 72-h forecast will be added and the 1200 UTC cycle forecasts.

3. EVALUATION

Operational equations were derived for the warm season only. Several statistics were used to evaluate the forecasts made with the equations. They include: the Heidke skill score against climatology, bias, and threat score.

The Heidke Skill Score against climatology is given by the number of correct forecasts minus the number of correct forecasts that would have been predicted by climatology divided by the total number of forecasts minus the number of correct forecasts that would have been predicted by climatology. The bias was defined earlier. The threat score is the number of correct forecasts in a catagory divided by the total number of forecasts less the correct forecasts of the other catagories. Skill is exhibited if the skill score is greater than zero and less than or equal to one. The closer to one the better. Threat scores vary between zero and one; a score of one is perfect.

Evaluations were made with two months of data from the 1987 warm season (June and July). Tables 1 and 2 show the probabilities of occurrence for the Atlantic and Pacific Oceans for fog and visibility respectively. Individual evaluations of fog and visibility were done for each region and each cycle, but only the combined results for the Atlantic and Pacific for the 0000 UTC cycle are shown here. Tables 3 - 6 show the results of the evaluations of fog and visibility respectively.

Region	fog	no fog
Atlantic	.1518	.8482
Pacific	.2312	.7688

Table 1: Fog climatology for the months of June and July. Probabilities were determined from data for the years 1980 through 1984.

Table 2: Visibility climatology for the months of June and July. Probabilities were determined from data for the years 1980 through 1984.

Region	< 5.5 km	≥ 5.5 km
Atlantic	.3198	.6802
Pacific	.4058	.5942

Table 3: Evaluation of fog forecasts with independent data from June and July 1987 for the North Atlantic. Tau is the projection hour; N is the sample size; S. S. is the skill score; T. S. is the threat score; F stands for fog, and N F stands for no fog. Results are for the 0000 UTC cycle.

Tau	N	S. S.	Bias		T. S.	
			F	N F	F	N F
00	1883	.251	0.99	1.00	.22	.80
24	1883	.271	1.00	1.00	.17	.82
48	1889	.255	0.98	1.00	.17	.82
72	1889	.260	0.98	1.00	.18	.81

Table 4: Same as Table 3 except for the North North Pacific.

Tau	N	S. S.	Bias		T. S.	
			F	N F	F	N F
00	4940	.297	0.93	1.02	.28	.73
24	4876	.315	0.98	1.01	.31	.73
48	4845	.307	0.89	1.03	.29	.74
72	4772	.258	1.04	0.99	.28	.70

Table 5: Evaluation of visibility forecasts with independent data from June and July 1987 for the North Atlantic. Category labels for bias and threat score are in kilometers. Results are for the 0000 UTC cycle.

Tau	N	S. S.	Bias		T. S.	
			< 5.5	≥ 5.5	< 5.5	≥ 5.5
00	1883	.324	1.18	0.94	.33	.66
24	1883	.332	1.17	0.95	.30	.68
48	1889	.314	1.12	0.96	.30	.67
72	1889	.270	1.23	0.92	.27	.64

Table 6: Same as Table 5 except for the North Pacific.

Tau	N	S. S.	Bias		T. S.	
			< 5.5	≥ 5.5	< 5.5	≥ 5.5
00	4940	.315	1.08	0.96	.39	.58
24	4876	.333	1.07	0.96	.41	.59
48	4845	.302	0.96	1.02	.38	.59
72	4772	.269	1.07	0.95	.38	.55

In the Atlantic and Pacific the fog forecasts are nearly unbiased. The threat scores and skill scores are better in the Pacific than in the Atlantic. The skill scores peak at 24 hours. Part of this may be due to the thresholding technique, and part may be due to different biases in the initialization than in the model itself. The 1989 data may show a different story because of improvements in both the initialization and the model since 1987. All the forecasts show reasonable skill against climatology.

The visibility equations tend to overforecast low visibilities at all projections but the 48 hour projection in the Pacific. There the biases are nearly balanced. Again the threat scores are better in the Pacific, while the skill scores are better in the Atlantic. Skill scores again peak at 24 hours probably for the same reasons as the fog above. All forecasts show reasonable skill against climatology. It is clear from the biases that more optimization has to be done: never the less the skill scores and threat scores are very good.

4. AVAILABILITY

Output from the Fog and Visibility Forecast Guidance System is depicted on the Marine Significant Weather Chart. Figure 3 shows a sample chart. The scalloped lines are where visibility decreases to 3 n mi or less and the hatched areas are where fog is predicted to occur.

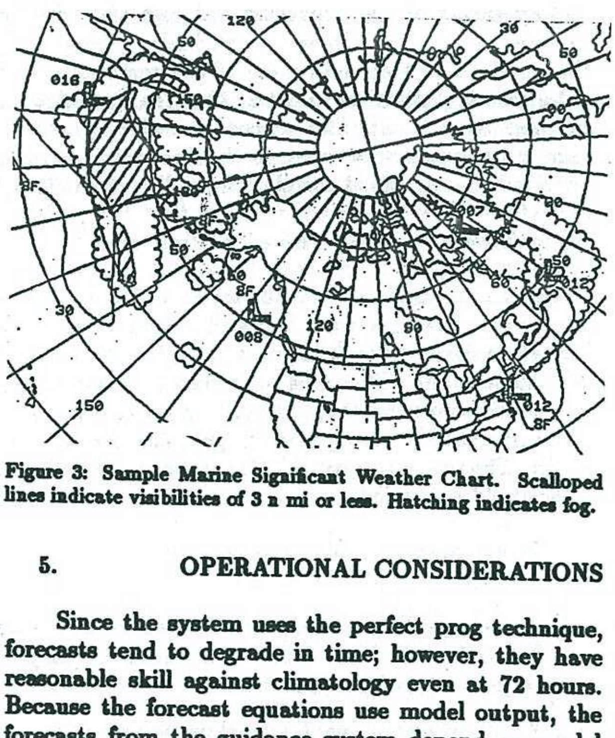

Figure 3: Sample Marine Significant Weather Chart. Scalloped lines indicate visibilities of 3 n mi or less. Hatching indicates fog.

5. OPERATIONAL CONSIDERATIONS

Since the system uses the perfect prog technique, forecasts tend to degrade in time; however, they have reasonable skill against climatology even at 72 hours. Because the forecast equations use model output, the forecasts from the guidance system depend on model performance; therefore, the forecaster needs to be aware of how the model is doing in his or her area of interest in order to make any necessary adjustments to the statistically derived forecast.

The forecaster should also keep in mind that system is designed for open ocean fog and visibility forecast guidance only. Its guidance should not be extrapolated into coastal areas. The system is also not designed to delineate lowered visibilities due to isolated mesoscale phenomena.

6. REFERENCES

Anderson, T. W., 1958: *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, Inc., New York, 354 pp.

Burroughs, L. D., 1987: Development of open ocean fog forecasting regions, *Ocean Products Center Technical Note/NMC Office Note* No. 323, National Weather Service, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 36 pp.

Cooley, W. W., and P. R. Lohnes, 1971: *Multivariate Data Analysis*. John Wiley and Sons, Inc., New York, pp 243-250.

Dey, C. H., and L. L. Morone, 1985: Evolution of the National Meteorological Center Global Data Assimilation System: January 1982-December 1983, *Mon. Wea. Rev.*, 113, pp 304- 318.

Guttman, N. B., 1971: Study of worldwide occurrence of fog, thunderstorms, supercooled low clouds and freezing temperatures. NAVAIR 50-1c-60. National Climatic Center, Environmental Data Service, NOAA for the Commander, Naval Weather Service Command, 64 pp.

Kistler, R. E., and D. F. Parrish, 1982: Evolution of the NMC Data Assimilation System: September 1978-January 1982, *Mon. Wea. Rev.*,110, pp 1335-1346.

Klein, W. H., B. M. Lewis, and I. Enger, 1959: Objective prediction of five-day mean temperatures during winter, *J. Meteor.*, 16, 672-682.

Miller, R. G., 1962: Statistical prediction by discriminant Analysis, *Meteorological Monographs*, 4, 54 pp.

——, 1988: Very short range statistical forecasting of automated weather observations. *Final Report* No. DOT/FAA/PS-88/3, Techniques Development Laboratory, National Weather Service, NOAA, U.S. Department of Commerce, 240 pp.

National Weather Service, 1988: Marine Significant Weather Chart, *NWS Technical Procedures Bulletin* No. 376, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 3 pp.

SAS Institute, 1982: *SAS User's Guide*: Statistics. SAS Institute Inc., 584 pp.

Sela, J. G., 1988: The new T80 NMC operational spectral model, *Preprint* - 8th Conference on Numerical Weather Prediction, Am. Meteor. Soc., pp 312-315.

Tatsuoka, M. M., 1971: *Multivariate Analysis*, Techniques for educational and psychological research. John Wiley and Sons, Inc., New York, pp 217-232.

Tremant, M., 1987: La prevision du brouillard en mer, *Meteorologie Maritime et Activites Oceanographiques Connexes* Rapport No. 20, World Meteorological Organisation, 127 pp.

Wilson, L. J., and K. Macdonald, 1985: Assessment of perfect prog marine wind forecasts, *Proceedings International Workshop on Offshore Winds and Icing* Halifax, Nova Scotia, October 7-11, 1985, Environment Canada, Atmospheric Environment Service, Canada, pp 353-363.