PAGEOPH, Vol. 120 (1982)

0033–4553/82/020203–08$01.50 + 0.20/0

Infrared Visibility Prediction by Statistical Methods

By Arnold Court, Gong-Yuh Lin and Stephen Zetsche[1]

Abstract – This study attempts to find statistical methods of predicting infrared visibility (IRV), as calculated from hourly meteorological observations from a North Atlantic weather ship. Simple and multiple regressions expressing IRV as a function of its component weather variables, and exponential data transformations, for time lags of 1 to 24 hours, gave R^2 values from 0.68 (1-hour lag) to 0.09 (24-hour lag). These have limited predictive power for lags up to 6 hours, almost none for longer lags. Two-category discriminant analysis, using class breaks at 2 km or 10 km is of little use, due to uneven data distribution.

Possibly more promising would be an application of Machine Output Statistics ('MOS'), used routinely for temperature forecasts, to this problem.

Key words: Statistical methods for visibility prediction; Atmospheric visibility.

1. Background

Predicting the distance at which objects can be detected by a forward-looking infrared device has become an important problem, especially at sea. 'Infrared visibility', or 'IRV', is not routinely measured and hence must be estimated from other meteorological observations (Table 1). An algorithm developed by the U.S. Navy (Katz *et al.*, 1979) expresses IRV as a function of wind speed, moisture, precipitation, air temperature, pressure, and visual range. To estimate IRV from hourly weather observations in which the basic moisture measurement is by psychrometer, the equations have been revised to use wet bulb temperature, rather than relative humidity or mixing ratio, as the basic quantity (Table 2).

Accepting without evaluation the computational estimate of IRV, this study seeks to describe the statistical properties of IRV in hopes of developing ways to forecast it 1 to 24 hours in advance. For this purpose, hourly IRV was computed for Ocean Weather Station 'I' (59 N, 19 W) 600 km south of Iceland for approximately 10 years of record (1961–1971). Mean IRV was 31.3 km, with a very slight diurnal variation from a mean minimum of 30.4 km at 3 a.m. to a mean maximum of 32.2 km at 3 p.m. The moderate seasonal variation shows IRV to be highest in spring (Apr: 35.1 km) and lowest in late summer and autumn (Oct: 28.3 km) (Fig. 1).

Four half years, two 'winters' (Nov–Apr) and two 'summers' (May–Oct) are

[1]) California State University, Northridge, CA 91330, USA.

Table 1

Symbols for weather variables used in computing infrared visibility (I), and related quantities

B	variable portion of visibility function
E	actual vapor pressure, mb
H	relative humidity, fractional (*not* %)
I	infrared visibility, km (also 'IRV')
M	mixing ratio, g/kg
P	atmospheric pressure, mb
R	correlation coefficient
T	dry bulb temperature, C
Θ	dry bulb temperature, K ($= T + 273.15$)
Ψ	web bulb temperature, C
U	0.1 (wind speed −3.5)
V	visual range, km (when $V < 1$ km)
W	precipitation intensity index (WW code)
Y	wind speed m/s ($= 10U + 3.5$)

Table 2

Formulas for estimating infrared visibility (I) from concurrent weather variables

$E = 6.108 \exp [17.27\Psi/(\Psi + 237.3)] - 0.066P(T - \Psi)(1 - 0.00115\,\Psi)$

$H = E/6.108 \exp [17.27T/(T + 237.3)]$

$M = 622E/(P - E)$

$I = \{0.000172 + [0.22B + 0.00225 + 0.00207M^{1/2} + 10^{-9}\,\Theta \exp (1800/\Theta)\,(9.834 - 0.4246M + 1.3638M^2 - 0.005713M^3)]^2\}^{-1/2}$

$V < 1$ km: $B = 3.91/V$

$V \geq 1$ km, wind ≤ 10.5 m/s: $B = W + (0.572992 + 0.403750U + 3.3178U^2)\,(-0.49008 + 4.627H - 16.82449H^2 + 29.8622H^3 - 25.9181H^4 + 8.82561H^5)$

$V \geq 1$ km, wind > 10.5 m/s: $B = W + (-22.55837 + 99.9136U - 171.1637U^2 + 146.2984U^3 - 62.03601U^4 + 10.7387U^5)\,(-1.9854 + 18.7238H - 68.0895H^2 + 120.9162H^3 - 105.0134H^4 + 35.7888H^5)$

'present weather, ww'	precipitation intensity index, W
00–49	0.000
50–59, 76–79	0.035
60, 61, 66, 68, 70, 71, 80, 83, 85, 87, 89, 91, 93	0.200
62, 63, 72, 73, 81, 84, 95, 96	0.600
64, 65, 67, 69, 74, 75, 82, 86, 88, 90, 92, 94, 97–99	1.600

examined in detail. Letters designate these half years; the periods they cover and the number of hours of data available out of the total possible are:

	Time period	Season	Available/Possible	
'A'	Nov 1961–Apr 1962	'winter'	3935/4344	90.6%
'B'	May 1962–Oct 1962	'summer'	4235/4416	95.9%
'C'	Nov 1962–Apr 1963	'winter'	4232/4344	97.4%
'D'	May 1963–Oct 1963	'summer'	4246/4416	96.2%

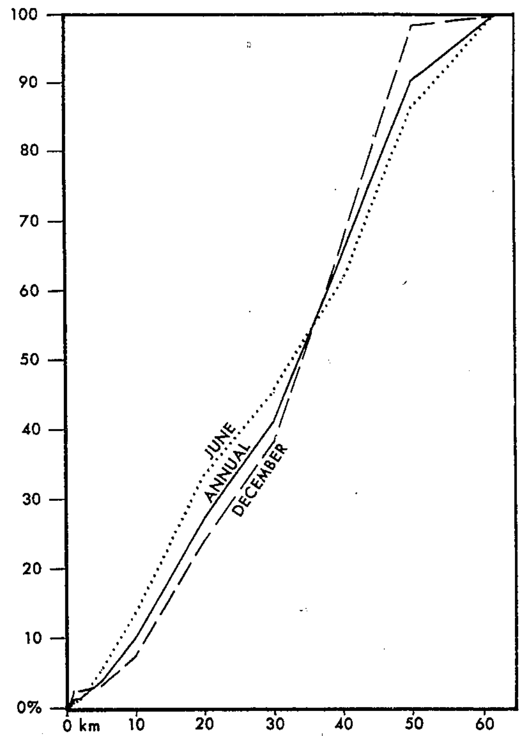

Figure 1
Cumulative distribution of IRV (1961–1971).

Over these two years, the computed IRV ranged from 0.06 to 61.4 km, with the mean for the 16 648 available observations at 31.2 km. Persistence of IRV, shown by simple serial correlations, separately for each of the four periods, decreases fairly smoothly from 0.8 for a 1-hour lag to 0.1 or 0.2 after 24 hours; the decrease is slower than in a simple Markov process (Fig. 2).

2. *Regressions*

Simple and multiple regressions were used to find the best linear prediction equations for selected time lags of 1 to 24 hours, and to evaluate the prediction accuracies of these equations. For a simple regression based on IRV only, for period A the 1-hour lag correlation of 0.79 indicates that 63% of the variance of IRV can be removed by predicting it from the previous hour's value:

$$IRV_k = 6.61 + 0.792 IRV_{k-1}$$

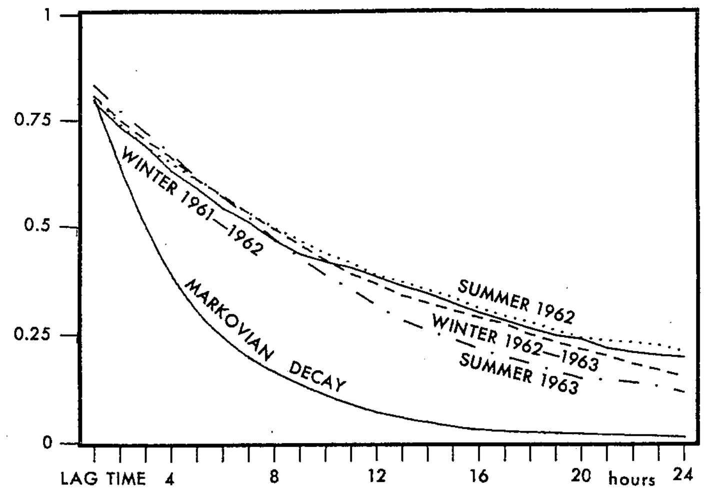

Figure 2
Autocorrelation of IRV.

Using three preceding hourly values improves the estimate only slightly, with a multiple correlation of 0.81:

$$\text{IRV}_k = 4.066 + 0.527\text{IRV}_{k-1} + 0.205\text{IRV}_{k-2} + 0.140\text{IRV}_{k-3}$$

For the four periods the simple regression R^2 decreases to 62 to 65% at 1-hour lag, to 1 to 5% at 24-hour lag (Table 3). Thus the predictive power of the simple regression is negligible for lags greater than 6 hours.

IRV may also be estimated by multiple regression directly on the individual weather elements used in its calculation. This procedure is much simpler than first computing IRV for each hour and then using it in a regression. The basic formula (Table 2) expresses IRV as the inverse square root of the sums and products of various polynomials, each involving one or more weather variables to assorted powers. Nevertheless, for simplicity the six variables were first entered, by stepwise regression, in linear combinations to express IRV. For zero lag, such a polynomial for period A, using symbols defined in Table 1, is:

$$\text{IRV}_0 = 7.15 - 1.64Y + 8.55T - 9.36\Psi - 21.1W + 0.203V + 0.034P$$

The multiple correlation of 0.92 yields a coefficient of determination of 0.85, to which the several terms contribute, respectively, 0.482 (Y), 0.209 (T), 0.108 (Ψ), 0.045 (W), and less than 0.01 for the last two (P, V).

A similar model for prediction 1 hour ahead in period A is:

$$\text{IRV}_1 = -31.0 - 0.160Y + 7.28T - 8.20\Psi - 3.92W + 0.179V + 0.073P$$

The multiple correlation of 0.81 is slightly more than the simple correlation of 0.79

Table 3

Coefficients of determination (R^2) and standard errors of prediction of infrared visibility (km) by three regressions

Lag	Period	Simple	Multiple	Mult. expon.
1	A	0.62/10	0.65/10	0.66/10
	B	0.65/8	0.64/5	0.67/8
	C	0.65/9	0.65/9	0.66/9
	D	0.68/8	0.63/8	0.66/8
3	A	0.47/12	0.49/12	0.50/12
	B	0.48/10	0.52/9	0.54/9
	C	0.50/11	0.52/11	0.52/11
	D	0.51/9	0.50/10	0.51/9
6	A	0.29/14	0.32/14	0.33/14
	B	0.34/11	0.39/10	0.40/10
	C	0.33/13	0.36/13	0.37/13
	D	0.32/11	0.31/11	0.33/11
12	A	0.14/15	0.16/15	0.16/15
	B	0.16/12	0.22/12	0.22/12
	C	0.14/15	0.16/15	0.18/14
	D	0.10/13	0.12/13	0.13/13
24	A	0.04/16	0.04/16	0.06/16
	B	0.05/13	0.08/13	0.09/13
	C	0.02/16	0.04/15	0.05/15
	D	0.01/14	0.02/14	0.02/14

obtained previously for predicting IRV from its value 1 hour earlier. For lags of 1 to 24 hours for the four periods, in most cases the amount of variation explained by this regression is equal to or very slightly higher than the figures for the simple regression, although the difference in R^2 ranges from a 5% improvement to a 5% detriment, and none of the differences is significant at the 5% level (Table 3).

Conforming more fully to the basic formula is a multiple regression involving various positive powers ($\frac{1}{2}$, 1, $1\frac{1}{2}$, 2, 3, 4, 5) of the weather variables. Once again, stepwise regression procedures were used, but for simplicity the resulting equations were limited to only the most important variables, i.e. those which increase R^2 by 0.01 or more. The results of this set of regressions are consistently better, by up to 3%, than those from the unmodified weather variables (Table 3).

These are the best regressions, but the differences among the three types discussed here are very slight. Other transformations, logarithmic and exponential, did not yield better results. None of the regressions is very useful as a predictive tool. The 95% confidence interval for the predicted IRV ranged from ± 31 km (24-hour lag) to ± 19 km (1-hour lag) for the two winters, and from ± 27 km to ± 16 km for the two summers. This differed only slightly according to the regression method used.

3. Discriminants

Two-group discriminant functions were sought for classifying or predicting IRV categories, breaking at 2 km and 10 km. Discriminant analysis is a useful technique to determine functions as linear combinations of a set of predictor variables, in order to achieve not only the maximum group differentiations but also the minimum probability of error in assigning cases to groups. A discriminant function is a vector which maximizes the ratio of between-groups to within-groups sums of squares of discriminant scores (JOHNSTON, 1978; TATSUOKA, 1971).

Classification can be achieved by using unstandardized discriminant functions in conjunction with group centroids that are group means on discriminant functions. For instance, the discriminant function for two IRV groups breaking at 2 km for period A and zero lag, using six weather variables is:

$$\mathrm{IRV}_0 = 3.599 + 0.315T - 0.271\Psi - 0.004P - 0.054Y - 4.328W + 0.037V$$

The centroid is -2.671 for IRV < 2 km and 0.053 for IRV ≥ 2 km. The midpoint of group centroids is -1.309. An hour for which this equation gives IRV > -1.309 is classified as having IRV ≥ 2 km. The overall accuracy of this equation is 93.7%, but only 76 hours have IRV < 2 km in period A, accounting for 1.9% of the 3942 hours. Therefore, the accuracy would be 98.1% even if every hour was classified as having IRV > 2 km.

Despite the low frequency, the occurrence of IRV < 2 km is of concern; an objective technique of predicting it is needed. Using the given discriminant function, approximately 62% (47 of 76 cases) of hours with IRV < 2 km are correctly classified. This implies that the discriminant function has very little power for differentiating IRV categories breaking at 2 km.

The low discriminant power is also evidenced by small canonical correlations and high Wilks' Lambda values for the various lag lengths and periods (Table 4). Canonical correlation is the maximum correlation between the linear combination function of independent variables and that of dependent variables that are coded 1 for IRV ≥ 2 km and 0 for IRV < 2 km. Wilks' Lambda, the ratio of the within-groups to the total sum of squares, is an inverse measure of discriminant power.

For separation at 10 km, the discriminant functions for two IRV categories yield moderate discriminant powers, as evidenced by modestly high values of canonical correlations and Wilks' Lambda for zero lag. As with regression models, the discriminant powers decrease as lag lengths increase, with very little predictive power for lag lengths > 6 hours for IRV categories breaking at either 10 km or 2 km. This varies little among the four periods.

Because of considerable differences in the sample sizes of IRV categories, Box's M statistic, employed to test the equality of group covariance matrices, showed significant differences in group covariances at the 100% level, which implies that the discriminant technique may not be an appropriate approach for group differentiations of IRV

Table 4

Canonical correlations, Wilks' Lambda, and accuracies of two-group discriminant functions with 2 km and 10 km as breaking criteria for four half-year periods

		2 km			10 km		
Lag	Period	Canon.	Wilks	Accur.	Canon.	Wilks	Accur.
0	A	0.35	0.88	93.7	0.65	0.58	91.7
	B	0.23	0.95	84.1	0.59	0.65	91.4
	C	0.39	0.85	96.5	0.67	0.54	95.6
	D	0.25	0.94	78.3	0.64	0.59	90.2
1	A	0.23	0.95	81.1	0.53	0.72	83.0
	B	0.20	0.96	81.5	0.48	0.77	86.9
	C	0.32	0.90	93.5	0.53	0.71	86.0
	D	0.24	0.94	77.0	0.55	0.70	84.3
3	A	0.19	0.96	78.2	0.46	0.78	78.7
	B	0.17	0.97	77.8	0.40	0.84	82.5
	C	0.27	0.93	91.0	0.45	0.80	80.5
	D	0.23	0.95	76.5	0.48	0.77	80.7
6	A	0.15	0.98	74.7	0.39	0.84	74.5
	B	0.15	0.98	73.2	0.34	0.86	78.8
	C	0.23	0.95	88.0	0.36	0.87	73.8
	D	0.20	0.96	71.6	0.36	0.87	73.8
12	A	0.10	0.99	66.8	0.30	0.90	69.7
	B	0.12	0.99	71.5	0.21	0.95	70.9
	C	0.17	0.97	84.4	0.22	0.95	63.8
	D	0.13	0.98	64.6	0.18	0.97	63.6
24	A	0.10	0.99	65.0	0.19	0.96	64.3
	B	0.09	0.99	68.3	0.18	0.97	70.0
	C	0.10	0.99	73.9	0.14	0.98	63.1
	D	0.10	0.99	59.7	0.10	0.99	58.2

categories breaking at 2 km or 10 km. Nevertheless, discriminant analysis provides descriptive statistical associations between weather variables and IRV group membership. Furthermore, discriminant analysis may still be a useful technique for differentiating IRV categories of almost equal sample size.

4. *Conclusions*

Standard statistical procedures cannot produce useful predictions of IRV; predictive power is limited for time lags up to 6 hours and almost nonexistent for longer lags. Surprisingly, IRV can be predicted slightly more precisely, at any lag, by a linear combination of antecedent weather variables than by their combination into IRV and the use of a simple lag correlation.

While persistence and linear regression appear to have inadequate predictive power for IRV, more refined statistical procedures may be more effective. These would be similar to the Machine Output Statistics ('MOS') used by the National Weather Service to predict local temperatures, rainfall, and other variables from forecast heights of the 700 and 500 millibar surfaces at various gridpoints. In actual operation, each temperature is predicted by a linear regression on predicted heights at ten nearby gridpoints. Perhaps IRV can be predicted similarly from large-scale synoptic forecasts.

Acknowledgments

This work was undertaken under contract N 00228-80-C-RW 53 with the Naval Environmental Prediction Research Facility, Monterey. Dr Andreas Goroch of NEPRF supplied much needed guidance and advice.

References

JOHNSTON, R. J., *Multivariate Statistical Analysis in Geography* (Longman, New York 1978), p. 243.

KATZ, B. S., HEPFER, K., and MAC MEEKIN, N., *Electro-Optical Meterological Sensitivity Study* (Naval Surface Weapons Center, White Oak, MD 1979), PR 79-67.

TATSUOKA, M. M., *Multivariate Analysis* (Wiley, New York 1971).

(Received 24th May 1982, revised 7th July 1982)