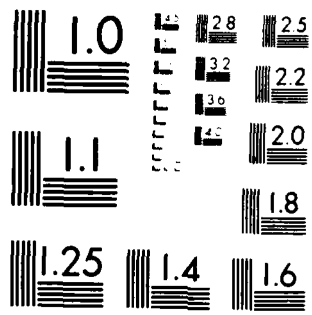
AD-A123 998 EVALUATION OF AN OBSERVATION-BASED CLIMATOLOGY MODEL FOR PREDICTING VISIB..(U) UNIVERSITY OF CENTRAL FLORIDA ORLANDO DEPT OF MATHEMATICS AND.. S J BEAN ET AL. 1/1

UNCLASSIFIED 06 JUL 82 SCIENTIFIC-1 AFGL-TR-82-0187 F/G 4/2 NL

END
DATE
FILMED
2-83
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A 123008

AFGL-TR-82-0187

EVALUATION OF AN OBSERVATION-BASED CLIMATOLOGY MODEL FOR PREDICTING VISIBILITY FOR DATA-VOID LOCATIONS IN GERMANY

S. J. Bean
P. N. Somerville

University of Central Florida
Department of Mathematics and Statistics
Orlando, Florida 32816

DTIC
ELECTE
JAN 31 1983
D
H

Scientific Report No. 1

6 July 1982

Approved for public release; distribution unlimited

DTIC FILE COPY

AIR FORCE GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANSCOM AFB, MASSACHUSETTS 01731

83 01

Qualified requestors may obtain additional copies from the Defense Technical Information Center. All others should apply to the National Technical Information Service.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFGL-TR-82-0187	2. GOVT ACCESSION NO. AD-A123998	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) EVALUATION OF AN OBSERVATION-BASED CLIMATOLOGY MODEL FOR PREDICTING VISIBILITY FOR DATA-VOID LOCATIONS IN GERMANY		5. TYPE OF REPORT & PERIOD COVERED Scientific Report No. 1
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) S. J. Bean, P. N. Somerville		8. CONTRACT OR GRANT NUMBER(s) F19628-82-K-0001
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Central Florida, Department of Mathematics and Statistics, Orlando, Florida 32816		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62101F, 667009AK
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Geophysics Laboratory, Hanscom AFB, Massachusetts 01731, Monitor/Donald Grantham/LYT		12. REPORT DATE 6 July 1982
		13. NUMBER OF PAGES 17
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Modeling	Data-void locations
Visibility	Climatological
Weibull Distribution	Evaluation
Germany	Constants Model
Sample Re-use	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

An observation-based climatological model for predicting visibility in data-void locations in Germany previously developed is evaluated. Two methods are used. First the model based on data from thirty stations (calibration set) is evaluated using a second set of thirty stations (evaluation set). Second, the model is evaluated using "sample re-use". The model was shown to give good results for estimating visibility results for nearly all of the stations.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ACKNOWLEDGEMENT

We should like to acknowledge the contribution of Jon Dittmer in the accomplishment of the results in this paper. Jon was responsible for writing several computer programs and made all of the numerous runs required for this report.

Accession For
NTIS GRA&I
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A

Evaluation of an Observation-based Climatology Model for Predicting Visibility for Data-void Locations in Germany

by

S. J. Bean and P. N. Somerville

University of Central Florida

1.0 INTRODUCTION

1.1 Review of Problem

A goal of the Air Weather Service has been to achieve a capability to determine the climatic probability of above-threshold conditions of the weather relative to the success of an Air Force flight mission, anywhere, at any time, expeditiously. Such a capability would materially heighten the effectiveness of a weapon system, since it is well-known that the environment can both degrade and enhance system effectiveness.

One method of summarizing or compacting the huge volume of historical records is by means of empirical cumulative distribution functions (cdf's). An empirical cumulative distribution function is simply the tabulated cumulative relative frequencies, or probabilities that a given variable will fall below specified values. Somerville and Bean (1979) have demonstrated that a number of climatological variables may be modeled with closed form distribution functions. For a given location and time (e.g., month and hour) the historical observations can be used to estimate specific model parameters.

Somerville and Bean (1981) used the Weibull distribution to model visibility in Germany for 30 stations. The cumulative distribution function for the Weibull is given by

$$F(x) = 1 - e^{-\alpha x^{\beta}}$$

where α and β are constants. Values of α and β are derived for each station, for each month and each of the eight 3-hour periods of the day. The probability of visibility less than x miles is then obtained by substitution in F(x).

1.2 Methods of Extending Visibility Probabilities to Data-Void Regions

A more difficult problem is to develop models which can be used to estimate probabilities for locations where records presently do not exist. Somerville and Bean (1981) developed two models for Germany. Thirty stations for which visibility records were available were used in the development of the models. For each of the stations, for a specified month and hour period, the Weibull distribution was used to model visibility. That is, a value for each of α and β was obtained. Having obtained the values for α and β, these values were regressed on a set of variables which were thought to have a possible influence on visibility. These included elevation, elevation relative to the average elevation on a circle whose radius is 20 kilometers from the station, east-west and north-south elevation difference, population density, relative humidity, proximity to a major body of water, mean wind speed, mean precipitation, latitude, longitude, and functions of and interactions between the above. A stepwise regression program was used to select which of the variables could be used as predictors for a specified month and hour period. Finally, a specially designed least squares non-linear regression program was used to simultaneously determine the regression coefficient in the formulas for α and β. This model was named the "variables model." The coefficients and the regression models are given in the above referenced paper.

A second model, named the "constants model" was also developed. Here only month and time of day were considered. That is, no information regarding elevation, humidity, wind speed, etc., was used. For each month, and time of day, non-linear regression was used to determine the "best" values for α and β. These were also tabulated in Somerville and Bean (1981).

In either model, extension of the climatology was accomplished by obtaining values for α and β at the data-void station, and then using the Weibull distribution to determine the desired visibility probabilities.

In this manuscript we will restrict ourselves to the evaluation of the constants model. Evaluation of the variables model will be given in a future report.

2.0 EVALUATION OF THE CONSTANTS MODEL

Two methods were used to evaluate the model. First, 30 stations were used as a "calibration" set, and the resulting constants model was used on a second independent "evaluation" data set of 30 West German stations. Second, sample re-use was used to evaluate the constants model. From the 60 evaluations one "overall" RMS (Root Mean Square) of the method was obtained. Sample re-use (sometimes called cross-validation) is a relatively new technique which makes it possible to use the same set of data for "calibration" and "evaluation". Briefly, if there is a total of n stations, n separate solutions are obtained. For each solution, one station is used as the evaluation set with the remaining n-1 used as the calibration set. Using the method one can obtain the Root Mean Square (RMS) error of the modeling procedure for each station. The individual RMS errors may then be combined to obtain an overall RMS error for the procedure. For future use the recommended model is the one using all the stations for the calibration set. For a good account of the sample re-use technique, papers by Stone (1974) and Geisser (1975) are recommended.

The first evaluation method gave some encouraging results, and it also indicated some ways we may improve the model. Exhibit 2.1 gives the RMS of the probability estimates for each station taken over all months and hour periods with the exception of hour periods 1 and 2. The data for these early pre-dawn hours are frequently missing, and they were, therefore, eliminated from the study. Stations 4, 13, and 26 stand out as very poor fits which inflate the overall RMS considerably. These three stations are much higher in elevation than the other stations in the study. Also, they are much higher than the surrounding area. These factors seem to give rise to much different visibility conditions than the other areas in the study.

Exhibit 2.2 gives the resulting RMS values averaged over all stations in the evaluation set for each month and hour periods 3 through 8. The overall RMS of .108 compares favorably with the calibration data set overall RMS of .063. The model obviously does not fit as well on the evaluation set, and we certainly could not expect that it would.

Station		RMS
1	AACHEN, DL	.074
2	BREMEN, GER	.055
3	LINGEN, GER	.050
4	KAHLER ASTEN, GER	.241
5	AIGEN ENNSTAL, GER	.109
6	PLEZEN / DOBRA, CZ	.110
7	BREMGARTEN, GER	.052
8	OBERSTDORF, GER	.107
9	KONSTANZ, GER	.054
10	INNSBRUCK, OS	.092
11	SALZBURG, OS	.071
12	PASSAU, GER	.055
13	FELDBERG, GER	.298
14	NEUHAUSEN, DL	.075
15	NURBURG, GER	.076
16	KOBLENZ, DL	.083
17	GIESSEN, GER	.085
18	HERSFIELD, DL	.068
19	KISSINLEN, DL	.063
20	COBURG, GER	.053
21	HOF, GER	.055
22	BERUS, GER	.061
23	KARLSRUHE, GER	.063
24	OHRINGEN, GER	.057
25	STAUBING, GER	.100
26	GROSSER FALK, GER	.249
27	LAHR, GER	.057
28	LAUDHEIM, GER	.054
29	KAUFBEUREN, DL	.066
30	MUHLDORF, GER	.051

Exhibit 2.1

Overall RMS For Each of The 30 Stations

in The Evaluation Set

	06-08	09-11	12-14	15-17	18-20	21-23	all
Jan	.137	.130	.130	.134	.134	.149	.136
Feb	.136	.126	.123	.125	.124	.138	.129
Mar	.124	.119	.107	.105	.113	.127	.116
Apr	.124	.112	.096	.092	.099	.104	.105
May	.114	.097	.079	.067	.076	.092	.089
Jun	.112	.092	.071	.058	.070	.084	.083
Jul	.106	.083	.063	.051	.056	.075	.075
Aug	.099	.082	.064	.058	.068	.077	.076
Sep	.115	.098	.074	.067	.085	.102	.092
Oct	.113	.111	.097	.095	.101	.121	.107
Nov	.136	.132	.130	.131	.136	.146	.135
Dec	.131	.124	.124	.127	.128	.145	.130
all	.121	.110	.100	.097	.103	.116	.108

Exhibit 2.2

RMS Over All Stations in the Evaluation Set

By Month and Hour Period (LST)

The results for the sample re-use evaluation were comparable to those for the first study. The RMS values for each of the 60 stations averaged over all months and hour periods 3 through 8 are given in Exhibit 2.3. The RMS values over all 60 stations for all months and hour periods 3 through 8 are given in Exhibit 2.4. The RMS values corresponding to sample re-use (Exhibit 2.4) are generally smaller than the results using the second 30 stations for evaluation (Exhibit 2.2). This is mainly due to the larger sample size in the sample re-use evaluation. The sample re-use evaluation makes use of 59 stations to build a model, whereas the first procedure makes use of only 30 stations in the model estimation

Exhibit 2.5 gives the values of α and β for the constants me[illegible]. for each month and hour period where all 60 stations are used for the cal[illegible]on set.

Exhibit 2.6 shows the location in Germany which were used in developing the models evaluated in this report.

WMO	Station	Lat	Long	RMS
10501	AACHEN, DL	50.78	-6.12	.081
10224	BREMEN, GER	53.05	-8.80	.046
10305	LINGEN, GER	52.52	-7.33	.127
10427	KAHLER ASTEN, GER	51.18	-8.50	.274
11157	AIGEN ENNSTAL, OS	47.53	-14.15	.1[illegible]
11446	PLEZEN/DOBRA, CZ	49.67	-13.30	.145
10906	BREMGARTEN, GER	47.90	-7.63	.05[illegible]
10948	OBERSTDORF, GER	47.40	-10.30	.1[illegible]
10929	KONSTANZ, GER	47.68	-9.20	.05[illegible]
11120	INNSBRUCK, OS	47.[illegible]	-11.3[illegible]	.081
11150	SALZBURG, OS	47.80	-13.02	.05[illegible]
10893	PASSAU, GER	48.58	-13.50	.066
10908	FELDBERG, GER	47.87	-8.02	.32[illegible]
10921	NEUHAUSEN, DL	47.99	-8.92	.108
10510	NURBURG, GER	50.33	-6.97	.134
10515	KOBLENZ, DL	50.35	-7.60	.102
10532	GIESSEN, GER	50.57	-8.72	.076
10542	HERSFELD, DL	50.67	-9.72	.135
10658	KISSINGEN, DL	50.20	-10.10	.126
10671	COBURG, GER	50.27	-10.97	.123
10685	HOF, GER	50.32	-11.90	.067
10704	BERUS, GER	49.27	-6.70	.125
10727	KARLSRUHE, GER	49.02	-8.40	.[illegible]
10742	OHRINGEN, GER	49.20	-9.53	.092
10786	STRAUBING, GER	48.8[illegible]	-12.60	.112
10791	GROSSER FALK, GER	49.08	-13.30	.275
10805	LAHR, GER	48.37	-7.85	.061
10837	LAUPHEIM, GER	48.22	-9.93	.094
10953	KAUFBEUREN, DL	47.87	-10.63	.065
10875	MUHLDORF, GER	48.25	-12.55	.091
10616	HAHN AB	49.95	-7.27	.040
10610	BITBURG AB	49.95	-6.57	.056
10614	RAMSTEIN AB	49.43	-7.58	.064
10607	SPANGDAHLEM AB	49.97	-6.70	.067
10384	TEMPELHOF ARPT	52.47	-13.40	.074
10755	ANSBACH AAF	49.32	-10.63	.071
10544	FULDA AAF	50.53	-9.63	.062
10869	ERDING AS	48.32	-11.93	.060
10765	FEUCHT AAF	49.38	-11.18	.055
10618	BAUMHOLDER AAF	49.65	-7.30	.126
10626	BAD KREUZNACH AAF	49.87	-7. 9	.111
10971	BAD TOLZ AAF	47.77	-11.6[illegible]	.123
10714	ZWEIBRUCKEN AB	49.22	-7.40	.24[illegible]
10633	WIESBADEN AB	50.05	-8.33	.050
10633	FINTHEN AAF	49.97	-8.15	.05[illegible]
10763	FURTH AAF	49.50	-10.95	.090
10642	HANAU AAF	50.17	-8.95	.04[illegible]
10852	GABLINGEN AAF	48.45	-10.67	.109
10653	GIEBELSTADT AUX AF	49.67	-9.88	.082
10687	GRAFENWOHR AAF	49.70	-11.95	.073
10734	HEIDELBERG AAF	49.40	-8.65	.041
10752	ILLESHEIM AAF	49.47	-10.38	.149
10659	KITZINGEN AAF	49.75	-10.20	.06[illegible]
10763	NURNBERG	49.50	-11.08	.0[illegible]6
10729	COLEMAN AAF	49.57	-8.47	.114
10657	WERTHEIM AAF	49.77	-9.48	.1[illegible]
10745	SCHWAEBISCH HALL AA	49.12	-9.78	.077
10712	SEMBACH AB	49.52	-7.87	.037
10862	SIEGENBERG GUNNERY	48.[illegible]	-11.80	.172
10738	ECHTERDINGEN ARPT	48.68	-9.22	.057

Exhibit 2.3

Overall RMS For Each of the Stations

(Sample Re-use)

	06-08	09-11	12-14	15-17	18-20	21-23	all
Jan	.126	.119	.114	.113	.124	.166	.128
Feb	.117	.105	.099	.099	.111	.152	.115
Mar	.103	.097	.087	.083	.094	.116	.097
Apr	.101	.089	.074	.072	.080	.089	.085
May	.093	.076	.060	.051	.060	.077	.071
Jun	.094	.073	.054	.045	.056	.070	.067
Jul	.088	.066	.049	.039	.047	.065	.061
Aug	.093	.070	.051	.046	.055	.080	.068
Sep	.106	.085	.064	.053	.074	.097	.082
Oct	.100	.097	.082	.078	.100	.134	.100
Nov	.112	.108	.104	.106	.121	.151	.118
Dec	.113	.107	.106	.108	.115	.167	.121
all	.104	.092	.082	.079	.090	.119	.096

Exhibit 2.4

RMS Over All Stations by Month and Hour Period

(Sample Re-use)

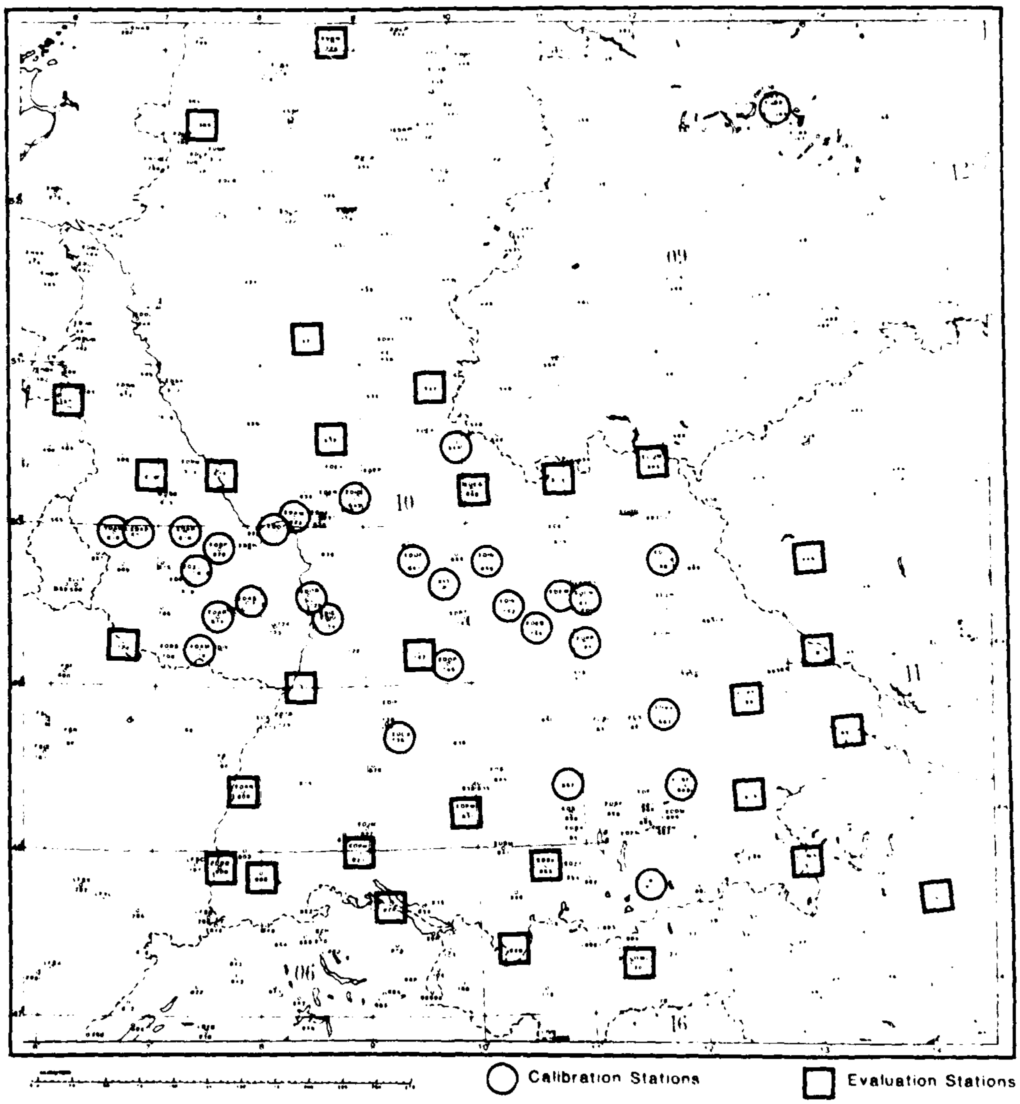
		06-08	09-11	12-14	15-17	18-20	21-23
Jan	alpha	.248	.253	.182	.172	.190	.217
	beta	.794	.799	.895	.891	.909	.866
Feb	alpha	.258	.239	.140	.123	.141	.162
	beta	.784	.834	.957	.743	.946	.887
Mar	alpha	.185	.131	.066	.056	.068	.072
	beta	.855	.954	1.078	1.036	1.043	1.089
Apr	alpha	.128	.067	.041	.039	.044	.045
	beta	.967	1.082	1.000	.901	.945	1.084
May	alpha	.100	.039	.029	.022	.029	.036
	beta	.953	1.134	.934	.939	.914	1.049
Jun	alpha	.093	.032	.021	.017	.022	.026
	beta	1.038	1.251	1.122	1.127	1.104	1.290
Jul	alpha	.079	.027	.016	.013	.016	.019
	beta	1.147	1.338	1.201	1.208	1.239	1.414
Aug	alpha	.137	.039	.016	.014	.018	.027
	beta	.952	1.341	1.408	1.342	1.352	1.370
Sep	alpha	.261	.091	.024	.018	.027	.047
	beta	.698	1.052	1.416	1.394	1.377	1.304
Oct	alpha	.351	.208	.078	.064	.095	.154
	beta	.561	.740	1.045	1.077	1.028	.849
Nov	alpha	.254	.223	.133	.129	.145	.178
	beta	.661	.727	.859	.846	.861	.761
Dec	alpha	.221	.226	.163	.169	.183	.210
	beta	.784	.798	.894	.857	.865	.808

Exhibit 2.5

Values of α and β for the Constants Model

by Month and Hour

Exhibit 2.6 **West German Stations**



5.0 CONCLUSIONS AND RECOMMENDATIONS

The constants model has been shown to give good results for estimating visibility probabilities with the exception of some higher elevation areas. Because the variables model makes use of a number of characteristics of the data-void location, including that of elevation, it is expected to improve the fit over the constants model.

One problem with the variables model, however, is that many of the input variables may be as hard to obtain as information about visibility. Two other models should be investigated to see if it is possible to improve on the constants model and at the same time not require too much information: one is a model that uses only topographical variables such as elevation and average elevation of the surrounding area; another is a model based on cluster analysis. That is, the known (α,β) parameters might be used to determine regions of homogeneous visibility characteristics. Constants models could then be used on the individual regions and, as before, sample re-use could be used to evaluate the results.

REFERENCES

Geisser, S. (1975) The Predictive Sample Re-use Method With Applications, Journal of the American Statistical Association 70, 320-328.

Somerville, P.N. and S. J. Bean, "Modeling Visibility For Locations In Germany Where No Records Exist" AFGL-TR-81-0313 December 1981. AD A111890

Somerville, P.N. and S. J. Bean, "Some Models For Visibility For German Stations", AFGL-TR-81-0144, 15 April 1981. AD A104167

Somerville, Paul N. and S. J. Bean, "Stochastic Modeling Of Climatic Probabilities", AFGL-TR-79-0222, November 1979. AD A080559

Stone, M. (1974) Cross-validatory Choice and Assessment of Statistical Predictions, Journal of the Royal Statistical Society B, 36 111-133.

END

DATE
FILMED

2-83

DTIC