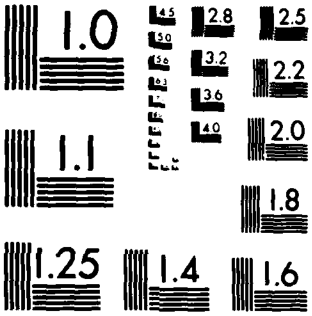
AD-A149 539 VERY SHORT RANGE STATISTICAL FORECASTING OF AUTOMATED 1/1
WEATHER OBSERVATIONS(U) NATIONAL WEATHER SERVICE SILVER
SPRING MD R G MILLER NOV 84 DOT/FAA/PM-84/31
UNCLASSIFIED DTFA01-83-Y-20625 F/G 4/2 NL

END
FILMED
DTIC



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

14

DOT/FAA/PM-84/31

Program Engineering
and Maintenance Service
Washington, D.C. 20591

Very Short Range Statistical Forecasting of Automated Weather Observations

Dr. Robert G. Miller

U.S. Department of Commerce
National Oceanic and
Atmospheric Administration
National Weather Service
Silver Spring, Maryland 20910

November 1984

Interim Report

This Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

Technical Report Documentation Page

1. Report No.
DOT/FAA/PM-84/31

2. Government Accession No.
AD-A149 539

3. Recipient's Catalog No.

4. Title and Subtitle
Very Short Range Statistical Forecasting of
Automated Weather Observations

5. Report Date
November 1984

6. Performing Organization Code

7. Author(s)
Robert G. Miller, Ph.D.

8. Performing Organization Report No.

9. Performing Organization Name and Address
U.S. Department of Commerce
National Oceanic and Atmospheric Administration
National Weather Service, Office of Systems Development,
Techniques Development Lab., 8060 13th St., S.S., MD 20910

10. Work Unit No. (TRAIS)
156-410-010

11. Contract or Grant No.
DTFA01-83-Y-20625

12. Sponsoring Agency Name and Address
U.S. Department of Transportation
Federal Aviation Administration
Program Engineering and Maintenance Service
Helicopter Program, APM-720; Washington, D.C. 20591

13. Type of Report and Period Covered
INTERIM REPORT
September 1983-Oct. 1984

14. Sponsoring Agency Code
FAA/APM-720

15. Supplementary Notes

Prepared under FAA/NWS Interagency Agreement No. DTFA01-83-Y-20625, managed by Weather Sensors Program, APM-440, Communications and Surveillance Division

16. Abstract

A procedure is developed for providing weather forecasting guidance over the short period between 1 to 60 minutes. It uses automated surface observation elements as predictors and predictands. The same equations project probabilistic predictions iteratively minute-by-minute. The model is founded on a Markov assumption and utilizes multivariate linear regression as the statistical operator. Details are given on how the model is constructed and how it compares with other objective methods such as climatology and persistence. Tests are performed on a new nonlinear approach.

17. Key Words
Short-range forecasts
Statistical forecasts
Markov
GEM
Multivariate linear regression

18. Distribution Statement
Document is available to the public through the National Technical Information Service, Springfield, Virginia 22151

19. Security Classif. (of this report)
Unclassified

20. Security Classif. (of this page)
Unclassified

21. No. of Pages
25

22. Price

Form DOT F 1700.7 (8-72)

Reproduction of completed page authorized

ACKNOWLEDGMENTS

Help on this project has been provided by Thomas J. Perrone in the initial stages of the work and in the maintenance of data collection at Sterling, Virginia. Others who have contributed include Wayne Huffman, Don Ouimette, John Edwards, Dick Reynolds, Jim Bradley, and Vicki Nadolski.

TABLE OF CONTENTS

LIST OF TABLES

1. INTRODUCTION

A recent analysis of rotorcraft operations (Adams, 1984, section 3.1, The Environment) indicates that the "typical flight mission consists of Point A to Point B flights that average 22 minutes in length, incorporating 5 interim stops for a total round-robin flight averaging 1 hour 48 minutes." The granularity (geographic and time) of weather information for rotorcraft operations is very small in comparison with what is required for fixed wing operations. Thus, the rotorcraft pilot has little interest in mid-range (4-6 hours) weather forecasts. What is needed is short range (10-120 minutes) forecasts for short range distances. The unique characteristics of rotorcraft allow them to land at places where airplanes can not. Many of these landing areas have low density traffic which does not justify a weather observer much less a forecaster. The rotorcraft community is extremely interested in an automatic weather sensor and an associated system for short term weather forecasting. It is for this reason that the FAA is sponsoring the NWS effort described in this report as an operational requirement.

The statistical technique for predicting the probability distribution of all surface weather elements minute-by-minute is called GEM for Generalized Equivalent Markov. It uses only the current local automated surface weather conditions as predictors. From these probability distributions, categorical predictions are made for each automated surface weather element. The technique is a Markov procedure which is briefly described in the following quotation from William Feller (1950):

> In stochastic processes the future is never uniquely determined, but we have at least probability relations enabling us to make predictions .... The term "Markov process" is applied to a very large and important class of stochastic processes .... Conceptually, a Markov process is the probabilistic analogue of the processes of classical mechanics, where the future development is completely determined by the present state and is independent of the way in which the present state has developed ... in contrast to processes ... where the whole past history of the system influences its future.

GEM is a multivariate linear regression system in which all variables, both predictors and predictands, are zero-one. It uses only the most recent observation of the automated surface weather elements to predict the probability distribution of those same automated weather elements. It does this in 1-minute increments. A categorical forecast is then made of each element, satisfying the constraint of balancing the number of times an element category is predicted with the number of times it is observed to occur.

If one were to approach the problem of predicting the probability distributions of future weather events by employing the classical Markov-chain model, it would soon become evident that enumerating the required states of nature, under a realistic number of characteristics, is infeasible. A new, or at least different, method must be tried. In GEM, a system of regression equations is set up to estimate the probability of all subsequent events at one time step. Then the transition probabilities in the usual Markov chain are essentially replaced by the regression-estimated probabilities. To accomplish this estimation of probabilities, all predictands are either a zero or a one in each observation. To facilitate the iterative characteristics of the chain, all predictors are similarly expressed as zero or one in each observation. The simplicity of such a system should be evident: Forecast all elements into the future by iterative steps, using only the present observed conditions of the events.

The mathematical model, data preparation, statistical analyses, and nonlinear prediction approach are given in Section 2. Section 3 presents results comparing GEM with climatology and persistence. Section 4 is a summary of work performed under the contract. Section 5 deals with future work to be performed.

2. TECHNIQUE DEVELOPMENT

This section describes the procedure from the mathematical model, through data preparation and statistical analyses, to a discussion of a nonlinear prediction method. The reader is referred to a NOAA Technical Report for further details (Miller, 1981).

2.1. Mathematical Model

Assumed given are measurements on a set of Z_1, Z_2, ..., Z_p predictor variables and a set of Y_1, Y_2, ..., Y_Q predictand variables for a group of N observations. The problem of multivariate regression is to construct a set of Q linear functions

$$\begin{aligned}
Y_1 &= a_{1,0} + a_{1,1}Z_1 + a_{1,2}Z_2 + \dots + a_{1,p}Z_p + \dots + a_{1,p}Z_p \\
Y_2 &= a_{2,0} + a_{2,1}Z_1 + a_{2,2}Z_2 + \dots + a_{2,p}Z_p + \dots + a_{2,p}Z_p \\
&\vdots \\
Y_q &= a_{q,0} + a_{q,1}Z_1 + a_{q,2}Z_2 + \dots + a_{q,p}Z_p + \dots + a_{q,p}Z_p \\
&\vdots \\
Y_Q &= a_{Q,0} + a_{Q,1}Z_1 + a_{Q,2}Z_2 + \dots + a_{Q,p}Z_p + \dots + a_{Q,p}Z_p
\end{aligned} \tag{1}$$

which have the property that the sum of the squares of the errors

$$\varepsilon_q^2 = \sum_{i=1}^{N} (Y_{i,q} - \hat{Y}_{i,q})^2 = \sum_{i=1}^{N} (Y_{i,q} - a_{q,0} - a_{q,1}Z_{i,1} - \tag{2}$$

$$\ldots - a_{q,p}Z_{i,p} - \ldots - a_{q,p}Z_{i,p})^2 \qquad (q=1,2,\ldots,Q)$$

are as small as possible. That is, the problem is to determine values of the $a_{q,p}$'s ($q = 1,2, \ldots, Q$; $p = 1,2, \ldots, P$) which minimize the quantities

$$\varepsilon_q^2 \; (q=1,2,\ldots,Q).$$

This is done by taking the partial derivatives of the Eq. (2) with respect to the unknown a's, setting each derivative equal to zero, and then solving for the a's. The process yields a set of normal equations which can be written in matrix notation as (underlining signifies a matrix or vector):

$$\underline{A} = (\underline{Z}'\underline{Z})^{-1}(\underline{Y}'\underline{Z}) \tag{3}$$

Expressed statistically this is the multivariate linear regression of the Y's on the Z's (Tatsuoka, 1971, pp. 26-38). In GEM, the Y values are advanced by one hour from the corresponding Z values. Thus

$$Y_{i+1,q} = Z_{i,q}$$

or

$$Y_{i+1,p} = Z_{i,p} (i=1,2,\ldots,N;\; q=1,2,\ldots,Q;\; p=1,2,\ldots,P).$$

Once $\underline{A}$ has been determined, it can then be used to estimate the value of $\underline{y}$ at one time step, given a set of $\underline{z}$ values at a zero time step (lower case values denote new observations of Y and Z):

$$\hat{\underline{y}}_1 = \underline{z}_0'\underline{A} \tag{4}$$

To employ an iterative scheme, such as in GEM, the estimate of $\underline{y}$ at time T can be expressed as

$$\hat{\underline{y}}_T = \underline{z}_{T-1}\underline{A} \qquad \text{(multiplicative form)} \tag{5}$$

with $\underline{z}$ at time T-1 taken to be the previous estimate $\hat{\underline{y}}_{T-1}$.

An equivalent alternative to estimating $\underline{y}$ at time T is to power $\underline{A}$ as follows:

$$\hat{\underline{y}}_T = \underline{z}_0\underline{A}^T \qquad \text{(additive form)} \tag{6}$$

The distinction between the two forms, multiplicative and additive, is that in the former, the operation required is to postmultiply the observation and then subsequent forecasts by $\underline{A}$, minute-by-minute. In the latter, since all observations in $\underline{z}_0$ are either zero or one, the operation only requires adding the coefficients whose observations are one, at any projection. To permit this, however, the powered versions of $\underline{A}$ must be determined initially, stored, and made available for the projections of interest.

2.2. Data Preparation

Data began to be collected at the National Weather Service's Techniques Development and Test Branch location at Sterling, Virginia, in April 1984. The following weather elements are observed once a minute by equipment similar to the FAA's Automated Weather Observing System (AWOS). The elements are:

- Lowest cloud hit
- Second cloud hit
- Third cloud hit
- Fourth cloud hit
- Visibility
- Station pressure
- Temperature
- Dew point temperature
- Wind speed
- Wind direction
- Precipitation amount in one minute
- Precipitation occurrence
- Frozen precipitation occurrence (when successfully measured)
- Date of the observation

The elements were transformed into categories, and dummy predictors and predictands were created. Table 1 shows the specific categories defined for each zero-one dummy predictor. Column 1 indicates the dummy variable number while column 4 gives the index of that variable. One dummy variable must be "left-out" because of mathematical redundancy.

2.3. Statistical Analyses

The statistical analyses which are performed on these data result from the processing of crossproduct matrices. The actual steps are as follows:

Step 1. Compute the $\underline{Z}'\underline{Z}$ and $\underline{Y}'\underline{Z}$ crossproduct matrices from the data matrices $\underline{Z}$ and $\underline{Y}$.

Step 2. Solve for $\underline{A}$ from $\underline{A} = (\underline{Z}'\underline{Z})^{-1}(\underline{Y}'\underline{Z})$ where $\underline{A}$ is the matrix of regression coefficients for making a 1-minute forecast.

Step. 3. Solve for the threshold probabilities p* for making categorical forecasts.

Derivation of the two crossproduct matrices $\underline{Z}'\underline{Z}$ and $\underline{Y}'\underline{Z}$, in step 1, was accomplished by using a pointer system which saved a considerable amount of computer time. This efficiency is made possible because of the zero-one nature of the observations.

For the labeled predictors in Table 2, Column 4 gives the sum row of the $\underline{Z}'\underline{Z}$ matrix and Column 5 the lowest ceiling row of the $\underline{Y}'\underline{Z}$ matrix. This gives the products between the Y variable for lowest ceiling hit times each of the 88 predictors over the sample N.

We solved for the regression coefficient matrix $\underline{A}$ in step 2 using the Crout method (Crout, 1941). This method does not require solving for the inverse matrix, $(\underline{Z}'\underline{Z})^{-1}$, but instead derives the regression coefficients by first a foreward and then a backward solution. Avoided are many of the computat instabilities encountered by inverting large matrices. The Crout metho yields an 88 x 87 matrix--88 predictor coefficients for each of 87 pred ands.

The lowest ceiling hit equation for the $\underline{A}$ matrix appears as Column 6 in Table 2.

2.4. Nonlinear Prediction Approach

Meteorologists have desired forecast guidance that is capable of predicting changes in the weather, such as frontal passages and their attendant variations, onset and discontinuation of severe weather (types and intensities), wind shifts and wind speed variations, as well as ceiling and visibility changes of a critical nature for aviation. Classical statistical approaches like regression have not succeeded in completely satisfying this desire, partly due to the *additive* nature of the statistical model currently employed. What seems to be needed is a model which will act in a *multiplicative* fashion--one capable of completely shutting down the prediction of an event when the antecedent conditions warrant. For example, when it rains, it is "never" preceded 1 minute before by a clear sky. However, a statistical-regression operator will fail to turn off the chance of rain fully if there are other antecedent conditions, say, easternly wind, high humidity, fog, and low visibility--conditions which are usually associated with future occurrences of rain. Regression would tend to increase the probability of rain because of each of these elements. In general with regression, the lack of any clouds would not be enough to negate completely the effect of these other elements.

Table 1. Predictor and predictand categories which specify the dummy variables used in GEM. Shown under the index column are the left-out categories not included because of redundancy.

Number	Weather Element	Category	Index
1	(Always unity)		1
2	Lowest cloud hit (00')	0 - 1	2
3		2 - 4	3
4		5 - 9	4
5		10 - 29	5
6		30 - 60	6
7		61 - UNL	Left out
8	Second cloud hit (00')	0 - 1	7
9		2 - 4	8
10		5 - 9	9
11		10 - 29	10
12		30 - 60	11
13		61 - UNL	Left out
14	Third cloud hit (00')	0 - 1	12
15		2 - 4	13
16		5 - 9	14
17		10 - 29	15
18		30 - 60	16
19		61 - UNL	Left out
20	Fourth cloud hit (00')	0 - 1	17
21		2 - 4	18
22		5 - 9	19
23		10 - 29	20
24		30 - 60	21
25		61 - UNL	Left out
26	Visibility (miles)	0 - 31/64	22
27		1/2 - 63/64	23
28		1 - 2 63/64	24
29		3 - 4 64/64	25
30		5 - 6 63/64	26
31		7 - 100	Left out
32	Station pressure (inches of Hg)	0 - 29.235	27
33		29.236 - 29.530	28
34		29.531 - 29.677	29
35		29.678 - 29.825	30
36		29.826 - 29.973	31
37		29.974 - 30.120	32
38		30.121 - 30.268	33
39		30.269 - 30.563	34
40		30.564 - 35.000	Left out
41	Temperature (°F)	-30 - 4	35
42		5 - 14	36
43		15 - 24	37
44		25 - 34	38
45		35 - 39	39
46		40 - 44	40

Table 1. Continued.

Number	Weather Element	Category	Index
47		45 - 49	41
48		50 - 54	42
49		55 - 59	43
50		60 - 64	44
51		65 - 74	45
52		75 - 84	46
53		85 - 94	47
54		95 - 110	Left out
55	Dew point depression (°F)	0 - 1	48
56		2 - 7	49
57		8 - 15	50
58		16 - 25	51
59		26 - 99	Left out
60	Wind speed (kt)	0 - 1	52
61		2 - 9	53
62		10 - 19	54
63		20 - 29	55
64		30 - 99	Left out
65	Wind direction (deg)	00 - 44	56
66		45 - 89	57
67		90 - 134	58
68		135 - 179	59
69		180 - 224	60
70		225 - 269	61
71		270 - 314	62
72		315 - 359	Left out
73	Precipitation amount (inches)	.002 - .100	63
74		.001 - .0019	64
75		.000 - .0009	Left out
76	Precipitation occurrence (Y or N)	Yes	65
77		No	Left out
78	Frozen precipitation (Y or N) (when successfully measured)	Yes	66
79		No	Left out
80	Month	January	67
81		February	68
82		March	69
83		April	70
84		May	71
85		June	72
86		July	73
87		August	74
88		September	75
89		October	76
90		November	77
91		December	Left out
92	Hour (LST)	00 - 01	78
93		02 - 03	79

Table 1. Continued.

Number	Weather Element	Category	Index
94		04 - 05	80
95		06 - 07	81
96		08 - 09	82
97		10 - 11	83
98		12 - 13	84
99		14 - 15	85
100		16 - 17	86
101		18 - 19	87
102		20 - 21	88
103		22 - 23	Left out

Table 2. Quantities derived for the designated dummy variables; the number of times each category occurred in the sample (ΣZ), the number of times each predictor occurred when it was followed by the lowest ceiling hit one minute later (ΣYZ), and the regression coefficient for each predictor when lowest ceiling hit was the predictand (A).

Index	Element	Category	ΣZ	ΣZY	A
1	(Always unity)		51882	1620	-.37821
2	Lowest cloud hit (00')	0 - 1	1620	684	.06854
3		2 - 4	2954	167	-.00320
4		5 - 9	2348	19	.00741
5		10 - 29	3342	21	.00465
6		30 - 60	5771	40	-.00437
7	Second cloud hit (00')	0 - 1	646	536	.33712
8		2 - 4	1442	227	.01249
9		5 - 9	1638	3	-.01058
10		10 - 29	2773	9	-.00494
11		30 - 60	4777	36	-.00751
12	Third cloud hit (00')	0 - 1	474	375	.08702
13		2 - 4	1332	339	.02272
14		5 - 9	1575	5	.01517
15		10 - 29	2655	4	.00173
16		30 - 60	4002	31	.00283
17	Fourth cloud hit (00')	0 - 1	251	188	-.02576
18		2 - 4	1245	433	.00099
19		5 - 9	1505	3	-.05200
20		10 - 29	2436	0	-.02365
21		30 - 60	3109	12	-.01352
22	Visibility (Miles)	0 - 31/64	508	413	.44201
23		1/2 - 63/64	544	306	.34377
24		1 - 2 63/64	2443	118	.02758
25		3 - 4 63/64	2132	78	.02444
26		5 - 6 63/64	2049	43	.00145
27	Station pressure (inches of Hg)	0 - 29.235	0	0	.00000
28		29.236 - 29.530	1461	26	-.00045
29		29.531 - 29.677	722	7	-.00125
30		29.678 - 29.825	8054	303	-.01045
31		29.826 - 29.973	15669	489	-.01004
32		29.974 - 30.120	19879	699	-.01522
33		30.121 - 30.268	5793	86	-.01149
34		30.269 - 30.563	304	10	.00000
35	Temperature (°F)	-30 - 4	0	0	.00000
36		5 - 14	0	0	.00000
37		15 - 24	0	0	.00000
38		25 - 34	216	9	.02964
39		35 - 39	549	9	.00781
40		40 - 44	1937	26	.00826
41		45 - 49	3454	52	.00659
42	Temperature (°F) cont.	50 - 54	5955	288	.00636
43		55 - 59	9335	517	.00017

Table 2. Continued.

Index	Element	Category	ΣZ	ΣZY	A
44		60 - 64	8601	236	.00433
45		65 - 74	12494	236	.00662
46		75 - 84	7648	151	.01197
47		85 - 94	1692	32	.01222
48	Dew point depression (°F)	0 - 1	2943	581	.00123
49		2 - 7	18062	550	-.00827
50		8 - 15	13835	195	-.00682
51		16 - 25	12817	214	-.00670
52	Wind speed (kt)	0 - 1	1357	84	.41415
53		2 - 9	40844	1386	.40452
54		10 - 19	9420	150	.40351
55		20 - 29	260	0	.31831
56	Wind direction (deg)	00 - 44	2932	93	-.00687
57		45 - 89	2435	121	-.00160
58		90 - 134	4893	234	.00886
59		135 - 179	5913	392	-.00877
60		180 - 224	11272	356	-.00500
61		225 - 269	4655	93	-.00009
62		270 - 314	11514	184	-.00896
63	Precipitation amount (inches)	.002 - .100	22	2	-.01724
64		.001 - .0019	97	4	.00564
65	Precipitation occurrence (Y,N)	Yes	2766	141	.01106
66	Frozen precipitation (Y,N) (when successfully measured)	Yes	0	0	.00000
67	Month	January	0	0	.00000
68		February	0	0	.00000
69		March	0	0	.00000
70		April	5655	92	.00684
71		May	37790	1342	.00688
72		June	8437	186	.00000
73		July	0	0	.00000
74		August	0	0	.00000
75		September	0	0	.00000
76		October	0	0	.00000
77		November	0	0	.00000
78	Hour (LST)	00 - 00	4334	209	.00684
79		02 - 03	4314	172	.00688
80		04 - 05	4103	215	.00000
81		06 - 07	4254	270	.00978
82		08 - 09	4223	156	-.00461
83		10 - 11	4370	50	-.01057
84		12 - 13	4425	70	-.00257
85		14 - 15	4389	65	-.00758
86		16 - 17	4376	78	-.00703
87		18 - 19	4380	62	-.00346
88		20 - 21	4373	99	-.00787

Fortunately, there is a statistical model or operator which possesses this necessary capability. The discrete likelihood function (DLF) approach is fairly new (see Miller, 1979), but the basis for its existence is founded on the work of the eminent statistician, Sir Ronald A. Fisher, whose own work and ideas on this subject were derived in the mid-eighteenth century from the inverse probability notions of Bayes. Basically, the concept is this: given that we observe a set of current conditions of the weather, the question to be asked is "What is the likelihood that these current conditions are those that would be the conditions preceeding rain and, conversely, what is the likelihood that these current conditions are those that would be the conditions preceeding no rain?" The two likelihoods are obtained by multiplying the conditional probabilities of each antecedent condition thus getting the joint probability of the entire observation. It should be emphasized that the presence of any antecedent condition which is incongruous with an event of interest (say, rain) will have a dramatic effect on that likelihood: it will force the likelihood to zero. Such a nonlinear system would seem to conform to meteorologists' desires. Should the usual conditional probabilities (posteriors) be of interest, they can be gotten directly from Bayes' theorem and the climatological frequencies of the possible events (priors). The likelihoods are obtained from a set of regression estimated probabilities (REEP) (see Miller, 1964). Empirical evidence has shown that rarely if ever is a REEP probability of an event < 0 when the event occurs and > 1.0 when it does not occur. Certainly the situations arise when REEP forecasts $P < 0$ and $P > 1$. However, truncating these REEP forecasts to 0 and 1.0, respectively, will not invalidate the reliability of the estimates.

Finally, a method which makes optimum use of these likelihoods for selecting categorical forecasts is an event selection based on a function of the likelihood ratio (see Von Mises, 1945).

3. RESULTS

To demonstrate the ability of the GEM equations to predict at a 1-minute projection, Brier scores have been computed for climatology, persistence, and GEM for each of the predictands of interest. These are given in Table 3 for the specified dummy variables. At the present time, only the dependent sample scores are presented. When one year's data has been compiled, Brier scores will be computed on a running sample of that next independent year. The Brier score for persistence as defined here uses only that dummy element corresponding to the specific predictand dummy. A greater reduction (lower values are better) in Brier score for persistence could have been achieved if all dummies of the predictand element were used as predictors. All dummies of a predictand element were not used as predictors in computing persistence's Brier score for two reasons: a) the procedure is so complex that it would severely strain the resources available to this project, and b) more importantly, persistence's

Table 3. Brier scores of each specified predictand for climatology, persistence, and GEM based on the developmented sample of 51882 cases. Dashes denote inapplicability.

Index	Element	Category	Climatology	Persistence	GEM
1	(Always unity)				
2	Lowest cloud hit (00')	0 - 1	.03025	.02532	.01969
3		2 - 4	.05370	.04582	.03732
4		5 - 9	.04328	.03256	.02578
5		10 - 29	.06037	.03229	.02481
6		30 - 60	.09876	.04538	.03768
7	Second cloud hit (00')	0 - 1	.01224	.00510	.00413
8		2 - 4	.02691	.01204	.00954
9		5 - 9	.03068	.01218	.01006
10		10 - 29	.05061	.01449	.01219
11		30 - 60	.08361	.03010	.02495
12	Third cloud hit (00')	0 - 1	.00898	.00449	.00373
13		2 - 4	.02496	.00899	.00734
14		5 - 9	.02935	.01008	.00838
15		10 - 29	.04859	.01236	.01034
16		30 - 60	.07127	.02483	.02082
17	Fourth cloud hit (00')	0 - 1	.00474	.00314	.00276
18		2 - 4	.02337	.00813	.00661
19		5 - 9	.02826	.01050	.00877
20		10 - 29	.04475	.01441	.01179
21		30 - 60	.05633	.02305	.01999
22	Visibility (Miles)	0 - 31/64	.00979	.00143	.00137
23		1/2 - 63/64	.01030	.00333	.00318
24		1 - 2 63/64	.04478	.00909	.00854
25		3 - 4 63/64	.03946	.01384	.01312
26		5 - 6 63/64	.03800	.01702	.01649
27	Station pressure (inches of Hg)	0 - 29.235	-	-	-
28		29.236 - 29.530	.02737	.00008	.00008
29		29.531 - 29.677	.01370	.00029	.00029
30		29.678 - 29.825	.13112	.00203	.00203
31		29.826 - 29.973	.21082	.00357	.00356
32		29.974 - 30.120	.23634	.00259	.00259
33		30.121 - 30.268	.09923	.00129	.00128
34		30.269 - 30.563	-	-	-
35	Temperature (°F)	-30 - 4	-	-	-
36		5 - 14	-	-	-
37		15 - 24	-	-	-
38		25 - 34	.00415	.00070	.00069
39		35 - 39	.01041	.00335	.00325
40		40 - 44	.03594	.00856	.00820
41		45 - 49	.06228	.01488	.01441
42	Temperature (°F) cont.	50 - 54	.10162	.02654	.02577
43		55 - 59	.14739	.04132	.03965
44		60 - 64	.13928	.04266	.04152
45		65 - 74	.18288	.02990	.02923

Table 3. Continued.

Index	Element	Category	Climatology	Persistence	GEM
46		75 - 84	.12567	.01705	.01655
47		85 - 94	.03160	.00679	.00666
48	Dew point depression (°F)	0 - 1	.05339	.03719	.03084
49		2 - 7	.22691	.06910	.05609
50		8 - 15	.19571	.04583	.04504
51		16 - 25	.18597	.03262	.03178
52	Wind speed (kt)	0 - 1	.02565	.01008	.00995
53		2 - 9	.16756	.04999	.04867
54		10 - 19	.14856	.04221	.04069
55		20 - 29	.00497	.00231	.00225
56	Wind direction (deg)	00 - 44	.05342	.01359	.01330
57		45 - 89	.04450	.01226	.01200
58		90 - 134	.08548	.01578	.01541
59		135 - 179	.10103	.02544	.02490
60		180 - 224	.17003	.02868	.02759
61		225 - 269	.08169	.02400	.02369
62		270 - 314	.17264	.03634	.03520
63	Precipitation amount (inches)	.002 - .100	.00044	.00036	.00034
64		.001 - .0019	.00189	.00186	.00174
65	Precipitation Occurrence (Y,N)	Yes	.05037	.00503	.00498
66	Frozen precipitation (Y,N) (when successfully measured)	Yes	-	-	-
67	Month	January	-	-	-
68		February	-	-	-
69		March	-	-	-
70		April	-	-	-
71		May	-	-	-
72		June	-	-	-
73		July	-	-	-
74		August	-	-	-
75		September	-	-	-
76		October	-	-	-
77		November	-	-	-
78	Hour (LST)	00 - 00	-	-	-
79		02 - 03	-	-	-
80		04 - 05	-	-	-
81		06 - 07	-	-	-
82		08 - 09	-	-	-
83		10 - 11	-	-	-
84		12 - 13	-	-	-
85		14 - 15	-	-	-
86		16 - 17	-	-	-
87		18 - 19	-	-	-
88		20 - 21	-	-	-

function is as a simple readily-available "no skill" statistical control. The more complex procedure is neither "readily available" nor simple, but a full-blown statistical forecasting procedure unto itself. The development of such a procedure is beyond the scope of this project.

4. BACKGROUND MATERIAL AND SUMMARY

Work on this contract began with a familiarization of the microcomputer programming language S Basic (structured compiler Basic) for the KAYPRO 10--a Z80 machine with a 10 megabyte Winchester hard disk, one floppy drive, and two RS232C ports plus a centronics port for a printer. Two such computers were acquired along with a letter quality printer about 3 months into the contract.

We engaged ARTAIS, Inc. through a subcontract to modify the experimental system at Sterling, Virginia. As a consequence, we now receive raw minute-by-minute sensor data plus observations derived from an algorithm developed for the Automated Surface Observation System (ASOS). One KAYPRO 10 computer was wired to the ARTAIS equipment at one of the KAYPRO's RS232C ports and was dedicated to the Sterling facility.

Capturing these data into files on the hard disk could not be done through the S Basic language. It was necessary to seek other ways of performing this task. Two such ways were found. One was through a C program written by Donald Ouimmette and the other through the purchase of MITE commercial telecommunication software. Both approaches succeeded; however, the former way was chosen for use because the program better suited our needs. We began collecting live data before the end of April and have collected data almost continuously since that time. Data collection has been interrupted very infrequently. The only serious type of interruption was caused by lightning striking elements of the observing system. When there is an interuption, we lose data until the outage has been brought to our attention or until we arrive at the Sterling facility to download the data onto floppies once a week. Most important: such outages will here bias the observations collected (e.g., deficiency in thunderstorm cases) to an, as yet, unknown degree.

Processing of the ASCII data, collected through the C program, is performed in the S Basic computer language. Gross error checking is performed on both the fixed and variable length data records. Eighty-eight predictors are set up to predict 87 predictands (described in Section 2.B). A pointer system was employed to get the crossproducts needed to solve the statistical equations for making a 1-minute forecast. Such a system is very efficient when dummy variables, such as are employed with GEM, are used. Nevertheless, it was necessary to acquire an additional KAYPRO 10 in June to permit the testing of the nonlinear DLF approach. Further details on DLF can be found in

Section 2.D. DLF can enhance the project in two ways: a) the DLF approach captures all the information contained in first-order interactions between each pair of predictors, avoiding the need to add such terms in the regular minute-by-minute GEM, and (b) the two methods, GEM and DLF, are compatible and will be used together should the contribution made by DLF be deemed worthwhile, based on further testing.

At the present time, we have exercised all the necessary development programs on as much data as have been collected. We will monitor the equations as they are produced on more and more data. Tests will be made to judge the value of DLF.

5. FUTURE WORK

Our plans for the remainder of the contract are:

- Complete the collection of a full year of AWOS and ASOS data at Sterling.
- Process these data for making a set of minute-by-minute (for 10-, 20-, 30-, 40-, 50-, and 60-minute projections) GEM equations for both AWOS and ASOS variables, both probabilities and categorical forecasts. These efforts will specifically predict ceiling, visibility, wind, precipitation, and temperature.
- Perform a verification on these equations on observations independent of the original sample.
- Test the effectiveness of Discrete Likelihood Functions (DLF).
- Prepare a plan for demonstration of the GEM system.
- Process any data acquired from other locations akin to the manner in which the Sterling data were processed.

One of the objectives during the period of this contract is to develop a prototype computer facility that will be self-standing as a:

- Real-time collector of automatic weather observations data minute-by-minute, both AWOS and ASOS.
- Decoder of each observation into dummy variables for processing into GEM.

- Accumulator of the statistical crossproduct into statistical covariance matrices within each predictand category.

- Creator of updated regression prediction equations.

- On-demand predictor of each element out to 60 minutes in 10-minute intervals.

Features of this facility will be that maintenance will be at a minimum. Only hardware breakdowns will disrupt the facility. Power breakdowns will not affect the operation, and it will not be required to periodically maintain the facility as was once thought necessary.

REFERENCES

Adams, R. J., and Pitts, J. T., 1984: Investigation of hazards of helicopter operations and root causes of helicopter accidents.

Crout, P. D., 1941: A short method for evaluating determinants and solving systems of linear equations with real and complex coefficients. Trans. AIEE, 60, 1235-1241.

Feller, W., 1950: An Introduction to Probability Theory and its Applications. John Wiley and Sons, New York, 419 pp.

Miller, Robert G., 1964: Regression estimation of event probabilities. Travelers Research Center Tech. Report, 7911-121, Contract Cwb10704, Hartford, Conn., 153 pp.

_____, 1979: Estimating event probabilities by discrete likelihood functions. Proceedings Sixth Conference on Probability and Statistics in Atmospheric Sciences, Banff, Amer. Meteor. Soc., 93-97.

_____, 1981: GEM: A statistical weather forecasting procedure. NOAA Technical Report NWS-28, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, 103 pp.

von Mises, R., 1945: On the classification of observation data into distinct groups. Ann. Math. Stat., 16, 68-73.

Tatsuoka, M. M., 1971: Multivariate Analysis: Techniques for Educational and Psychological Research. John Wiley and Sons, New York, 310 pp.

END

FILMED

2-85

DTIC