NOAA Technical Report NWS 28

GEM: A Statistical Weather Forecasting Procedure

Silver Spring, Md.
November 1981

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Weather Service

NOAA TECHNICAL REPORTS

National Weather Service Series

The National Weather Service (NWS) observes and measures atmospheric phenomena; develops and distributes forecasts of weather conditions and warnings of adverse weather; collects and disseminates weather information to meet the needs of the public and specialized users. The NWS develops the national meteorological service system and improves procedures, techniques, and dissemination for weather and hydrologic measurements, and forecasts.

NWS series of NOAA Technical Reports is a continuation of the former series, ESSA Technical Report Weather Bureau (WB).

Reports listed below are available from the National Technical Information Service, U.S. Department of Commerce, Sills Bldg., 5285 Port Royal Road, Springfield, Va. 22161. Prices vary. Order by accession number (given in parentheses).

ESSA Technical Reports

WB 1 Monthly Mean 100-, 50-, 30-, and 10-Millibar Charts January 1964 through December 1965 of the IQSY Period. Staff, Upper Air Branch, National Meteorological Center, February 1967, 7 p, 96 charts. (AD 651 101)

WB 2 Weekly Synoptic Analyses, 5-, 2-, and 0.4-Mb Surfaces for 1964 (based on observations of the Meteorological Rocket Network during the IQSY). Staff, Upper Air Branch, National Meteorological Center, April 1967, 16 p, 160 charts. (AD 652 696)

WB 3 Weekly Synoptic Analyses, 5-, 2-, and 0.4-Mb Surfaces for 1965 (based on observations of the Meteorological Rocket Network during the IQSY). Staff, Upper Air Branch, National Meteorological Center, August 1967, 173 p. (AD 662 053)

WB 4 The March-May 1965 Floods in the Upper Mississippi, Missouri, and Red River of the North Basins. J. L. H. Paulhus and E. R. Nelson, Office of Hydrology, August 1967, 100 p.

WB 5 Climatological Probabilities of Precipitation for the Conterminous United States. Donald L. Jorgensen, Techniques Development Laboratory, December 1967, 60 p.

WB 6 Climatology of Atlantic Tropical Storms and Hurricanes. M. A. Alaka, Techniques Development Laboratory, May 1968, 18 p.

WB 7 Frequency and Areal Distributions of Tropical Storm Rainfall in the United States Coastal Region on the Gulf of Mexico. Hugo V. Goodyear, Office of Hydrology, July 1968, 33 p.

WB 8 Critical Fire Weather Patterns in the Conterminous United States. Mark J. Schroeder, Weather Bureau, January 1969, 31 p.

WB 9 Weekly Synoptic Analyses, 5-, 2-, and 0.4-Mb Surfaces for 1966 (based on meteorological rocketsonde and high-level rawinsonde observations). Staff, Upper Air Branch, National Meteorological Center, January 1969, 169 p.

WB 10 Hemispheric Teleconnections of Mean Circulation Anomalies at 700 Millibars. James F. O'Connor, National Meteorological Center, February 1969, 103 p.

WB 11 Monthly Mean 100-, 50-, 30-, and 10-Millibar Charts and Standard Deviation Maps, 1966-1967. Staff, Upper Air Branch, National Meteorological Center, April 1969, 124 p.

WB 12 Weekly Synoptic Analyses, 5-, 2-, and 0.4-Millibar Surfaces for 1967. Staff, Upper Air Branch, National Meteorological Center, January 1970, 169 p.

NOAA Technical Reports

NWS 13 The March-April 1969 Snowmelt Floods in the Red River of the North, Upper Mississippi, and Missouri Basins. Joseph L. H. Paulhus, Office of Hydrology, October 1970, 92 p. (COM-71-50269)

NWS 14 Weekly Synoptic Analyses, 5-, 2-, and 0.4-Millibar Surfaces for 1968. Staff, Upper Air Branch, National Meteorological Center, May 1971, 169 p. (COM-71-50383)

NWS 15 Some Climatological Characteristics of Hurricanes and Tropical Storms, Gulf and East Coasts of the United States. Francis P. Ho, Richard W. Schwerdt, and Hugo V. Goodyear, May 1975, 87 p. (COM-75-11088)

(Continued on last page)

NOAA Technical Report NWS 28

GEM: A Statistical Weather Forecasting Procedure

Robert G. Miller
Techniques Development Laboratory
Systems Development Office

Silver Spring, Md.
November 1981

U.S. DEPARTMENT OF COMMERCE
Malcolm Baldrige, Secretary

National Oceanic and Atmospheric Administration
John V. Byrne, Administrator

National Weather Service
Richard E. Hallgren, Director

CONTENTS

Preface vi

Abstract 1

1. Introduction and background 1

What is GEM 1
What does the acronym stand for 1
Why GEM 2
An example of a GEM forecast 2
Analysis of the example 2
Overview of the report 9

2. Creating GEM 11

Mathematical model 11
Data 12
Preparation 12
Transformations 17
Statistical analyses 24

3. Experimental results, old and new 39

Air Weather Service single-station experiment 39
Conditional climatology experiment 40
GEM experiments 41
Testing the value of two observations in the predictor set . . 55
Analyzing anomaly effects 56
Conclusions 56

4. Independent verification results 58

5. Operational GEM systems 63

6. Summary 67

Characteristics of GEM that deserve special emphasis 67
Possible areas of research for enhancing GEM 68
Data preparation 68
Data transformation 68
Computational 69
Statistical analysis 69
Output 70

7. New results 73

Improving the model 73
Including local-hourly climatology 77

Acknowledgments 81

Bibliography 81

Glossary of terms 83

Glossary of symbols 87

Appendix: "A beta classification model," R. G. Miller and D. L. Best . . A-1

Figures

1-1. Reproduction of the 1200 GMT, March 21, 1980, Daily Weather Map . . 3

1-2. Reproduction of part of the official March 21, 1980, Washington National Airport WBAN form 4

1-3. GEM's predicted hourly probability distributions for March 21, 1980, at Washington National Airport 5

1-4. GEM categorical predictions for the March 21, 1980, example 8

2-1. Locations selected to provide data for GEM 16

2-2. Criterion for specifying each dummy predictor and predictand 18

3-1. Plot of difference between anomaly additive constants for NO WX/WX . 57

5-1. Example of minicomputer output of GEM 64

5-2. Example of TSO output of GEM 65

5-3. Example of batch output of GEM 66

Tables

2-1. A display of quantities derived for GEM for the predictand Y=NO WX/WX 25

2-2. A display of quantities derived for GEM, for all predictands and for a 1-hr projection 33

3-1. Analysis of covariance Brier scores and significance-test results . . 44

4-1. Independent sample Brier scores from 26,328 cases at seven stations for GEM and persistence 59

4-2. Brier score and hit comparisons between GEM and persistence 61

4-3. Brier score comparison of GEM with station-derived climatology versus GEM with generalized climatology for Minneapolis-St. Paul airport 62

7-1. Normalized weights for exponential GEM model for t=2,...,12 and from 1 to 24 powers of A 74

7-2. Comparison of Brier scores and hits between exponential GEM and geometric GEM 78

7-3. Brier score comparison between GEM, with exponential decay and local-hourly climatology, and persistence 80

Microfiche

A $\underline{Z}'\underline{Z}$ matrix, labeled "GROUPS 123 ZZ"

B $\underline{Y}'\underline{Z}$ matrix, labeled "GROUPS 123 YZ"

C $\underline{A}$ matrix, labeled "GROUPS 123 EQUATIONS"

D $\underline{B}$ matrix, labeled "PLODITE A NO. 1, NO. 2"

E Beta coefficient of $\underline{B}$, labeled "BETA COEFFICIENTS"

F $\underline{Aa}$ anomaly $\underline{A}$ matrix, labeled "A ANOMALY"

G $\underline{Ba}$ anomaly $\underline{B}$ matrix, labeled "B ANOMALY"

H μ_0 μ_1 R^2, labeled "MU1, MU0, R SQUARE"

I P* thresholds, labeled "BETA THRSHLD"

J A_0 additive constants, labeled "A0 ADDITIVE CONSTANTS EACH STATION"

PREFACE

The philosophy underlying GEM has its roots in the writings and lectures of the late Professor Norbert Wiener of the Massachusetts Institute of Technology (1948, 1950, 1956). He cites the case for a probabilistic approach to prediction in meteorology and for a linear solution to the problem. Much of his argument is abstract, but his personal assurance that efforts such as GEM are on the right track is encouraging.

The first detailed description of a GEM model appeared in a 1964 proposal to the U.S. Air Force's Air Weather Service (AWS) in response to a need to incorporate specials and other randomly observed weather conditions such as those provided by pilot reports, radar, and satellites. (See Miller, 1968.) AWS did not fund the proposed effort at that time. However, in 1977, the work was undertaken by AWS in conjunction with St. Louis University. (See Miller et al., 1977.)

This Technical Report gives computational details and results of a direct followup to the AWS effort. The data bases have been enlarged and the scope increased to include the formulation and testing of a generalized operator--applicable anywhere, any time, for any element in a surface weather observation, and for any projection into the future.

A Glossary of Terms and a Glossary of Symbols are provided at the end of the report for clarification of some of the specialized nomenclature employed in the text.

GEM: A STATISTICAL WEATHER FORECASTING PROCEDURE

Robert G. Miller
Techniques Development Laboratory
Systems Development Office
National Weather Service, NOAA
Silver Spring, Md.

ABSTRACT. A procedure is developed for providing weather forecasting guidance over the short period between 0 and 12 hours. It uses only the local surface observation elements as predictors. The same equations are used for any location and project probabilistic predictions iteratively hour by hour. 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. Experimental results that probe the basic characteristics of the approach are presented, followed by independent verification of results. Features of the model's operational implementation are discussed under a variety of possible configurations. Certain future efforts are proposed for enhancing the technique.

1. INTRODUCTION AND BACKGROUND

What is GEM

GEM is a statistical technique for predicting the probability distribution of all local surface weather elements hour by hour. It uses only the current local surface weather conditions as predictors. From these probability distributions, categorical predictions are made for each surface weather element.

What Does the Acronym Stand For

"G" means that the technique is generalized. The same statistical equations can be applied at any location and for any time period. "E" stands for equivalent,* because of its equivalence (as a linear approximation) to a Markov chain. "M" is for its being a Markov process, 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.

*For reasons that are given in chapter 7, New Results, the "E" is more recently for exponential.

Why GEM

The Techniques Development Laboratory (TDL) of the National Weather Service has the responsibility for providing statistical weather guidance to field forecasters. Model output statistics (MOS) is the accepted procedure for providing this guidance. (See Glahn and Lowry, 1972.) However, since the input to MOS requires data from analyzed dynamical models, there is a gap of about 6 hours between the taking of observations and the availability of MOS. In general, persistence has represented the most skillful guidance available during the 0- to 6-hr period. Since GEM could incorporate all weather element information contained in the surface observation, including persistence, it seemed reasonable to expect that it would provide predictive information between 0 and 6, or possibly 12, hours with some skill. The results of the experiments reported here confirm this surmise.

An Example of a GEM Forecast

- Observation Time: 0700 LST, March 21, 1980
- Location: Washington National Airport (DCA)
- Forecast projection: 1 to 12 hours

Figure 1-1 shows the 1200 GMT, March 21, 1980, Daily Weather Map.

Figure 1-2 gives a reproduction of part of the official March 21, 1980, Washington National Airport WBAN form for verification purposes.

Figure 1-3 gives GEM's predicted hourly probability distributions (GEMTRIX) of all subsequent weather conditions from 1 to 24 hours for the March 21, 1980, example.

Figure 1-4 shows the GEM hourly categorical predictions (GEM) for the March 21, 1980, example.

Analysis of the example

Note: The daily synoptic weather map is provided only to show the reader the situation and, except for DCA's 0700 LST surface observation, was not used anywhere in GEM.

GEM's forecasts for the 12-hr period show good agreement with the actual record and special observations on the official WBAN form for temperature, dewpoint temperature, pressure, weather, wind, and clouds, with a definite indication of a frontal passage at about noon.

In particular, a complicated system was approaching the Washington, D.C., area. The GEM forecast anticipated DCA's entry into the warm sector before noon, with an increase in precipitation intensity, the onset of showers, and a fairly determined wind shift around the noon hour. An accompanying pressure rise and a continuing fall in temperature and dewpoint were predicted through the period along with a lessening of precipitation.

FRIDAY, MARCH 21, 1980

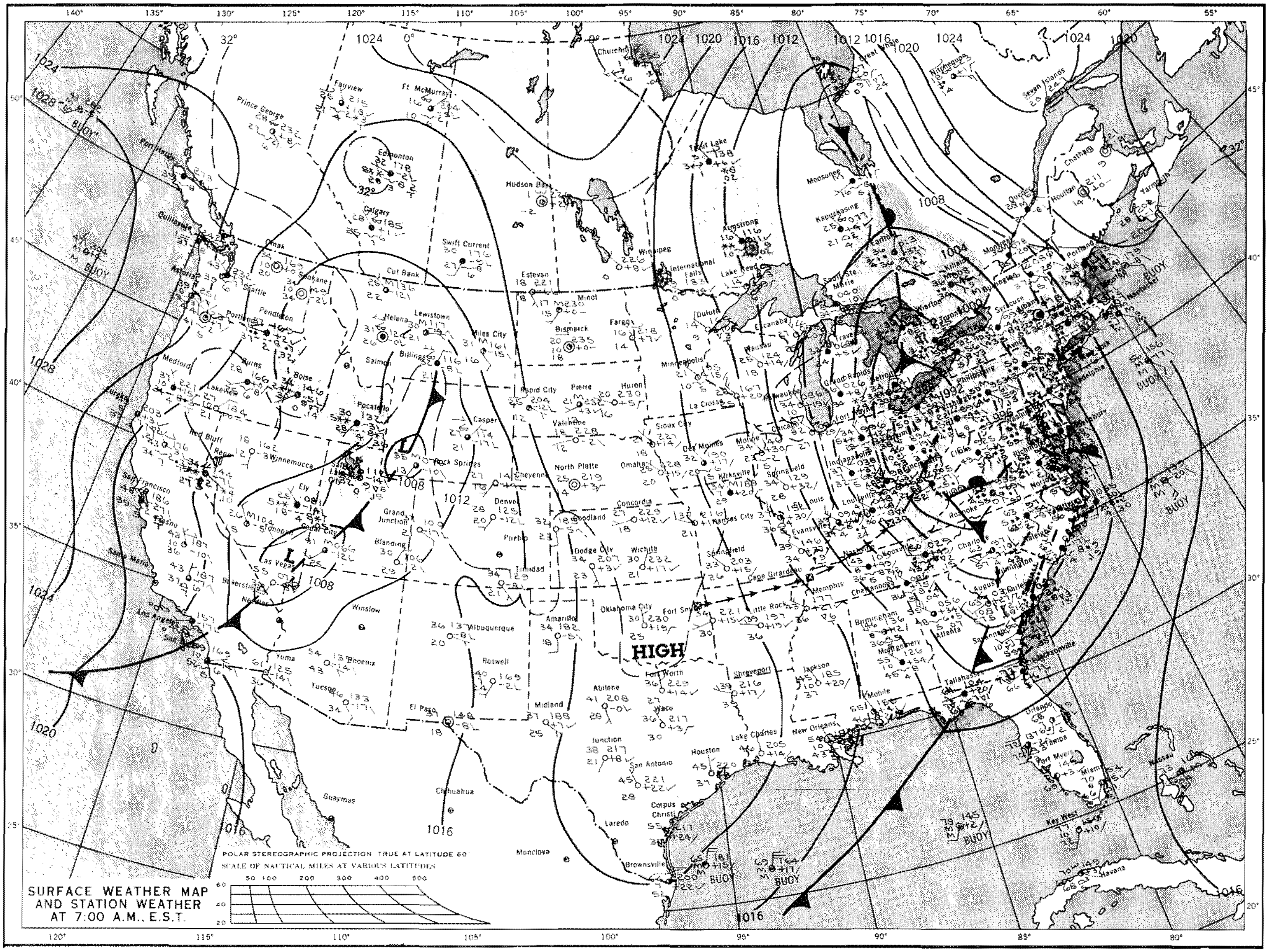


Figure 1-1.--Reproduction of the 1200 GMT, March 21, 1980, Daily Weather Map.

Type	Time (LST)	Sky and Ceiling	Visibility	Weather and Obstructions to Vision	Sea Level Press.	Temp.	Dew Pt.	Wind Dir.	Wind Speed	Wind Char./Gusts	Altimeter Setting	Remarks and Supplemental Coded Data	Initials
SA	0053	M12 OVC	7	R-	107	58	58	19	10		983	6300417// 67	CS
/		72405 81916 61616 10214	874//	14630	30	78		70420				20004 46745 0	CS
SA	0152	M13 OVC	7	R-	085	58	58	19	10		978		CS
SA	0252	M13 BKN 20 OVC	6	R-F	064	58	58	17	10		972	PRESFR / 98455	CS
SA	0353	M11 OVC	3	R-F	044	59	59	18	14		966	PRESFR / 75833 17//	CS
SP	0416	7 SCT E11 OVC	7	R-				17	12		965		CS
SA	0452	6 SCT E11 OVC	8	R-	027	60	59	16	12		961		CS
SA	0552	6 SCT E10 OVC	8	R-	010	60	59	17	12		956		CS
SP	0607	E7 BKN 10 OVC	5	R-F				16	11		954	PRESFR	CS
SA	0654	M8 BKN 11 OVC	6	R-F	990	60	59	15	13		950	PRESFR / 75461 172/	CS
/												57 20065	/
/		72405 81513 59616 99016	6732/	15754	69966			76180				20065 46757 0	CS
L	0723	M9 BKN 16 OVC	7	R-				15	14		946	PRESFR	NP
SA	0752	M9V BKN 12 OVC	7	R-	956	60	59	14	15	G22	940	CIG 8V10 PRESFR	NP
SP	0811	M10 BKN 12 OVC	15					16	14	G21	939		NP
SA	0852	9 SCT M10 BKN 16 OVC	15		936	61	60	15	15	G20	934	RE02 PRESFR	NP
L	0927	9 SCT M11 BKN 16 OVC	4	R-F				14	13		927	PRESFR	NP
SP	0930	8 SCT M10 OVC	2	RF				15	14		927	PRESFR	NP
SP	0937	M12 BKN 25 OVC	4	RF				13	13		925	PRESFR (FIBI)	NP
SA	0952	9 SCT M11 BKN 25 OVC	15		902	61	60	14	11		920	RB05E45 PRESFR	NP
~~SA~~	~~1052~~	~~10 SCT M14 BKN 25 OVC~~	~~15~~								~~919~~	/ 78810 172/	/
SA	1052	10 SCT M14 BKN 25 OVC	15		885	63	62	17	17		919		NP
SA	1152	14 SCT E25 BKN 90 BKN	10		861	65	63	18	12		912	MDT CU ALQDS PRESFR	NP
SP	1206	14 SCT E25 OVC	1	RW+				17	13		912	R36VR50V60+	NP
SP	1211	14 SCT E25 OVC	4	RW-				22	15	G25	913		NP
SP	1240	25 SCT E50 BKN 90 BKN	15					21	10		913	(FIBI)	NP
SA	1252	25 SCT E50 BKN 90 BKN	15		868	66	63	20	12		914	RB53E31/ 53420 [illegible]	NP
/		72405 62012 74258 86819	58570	17534	69844	92010		20085				46745 0 57 [illegible]	NP
SP	1334	E25 BKN 50 BKN 90 BKN	15					29	15	G20	913	MDT CU ALQDS	NP
SA	1352	13 SCT E25 BKN 90 BKN	15		865	69	60	29	14		913	MDT CU ALQDS RB36E45	NP
X	1441	15 SCT M25 OVC	5	RW-				29	28	G38	918	PRESRR [illegible]	NP
/												(FIBI)	/
SA	1453	15 SCT M25 BKN 50 OVC	7	RW-	885	58	51	30	22	G34	919	RB00 PRESRR [illegible]	NP
/												PK WND 2938/37	/
SP	1521	25 SCT E40 BKN	15					29	25	G38	921	~~PK WND 2938/37~~ PRESRR	NP
SA	1553	25 SCT 40 ~~[illegible]~~ SCT E100 BKN	15		905	58	43	31	20	G34	925	RE04 PK WND 3140/33	TH
/												/ ~~[illegible]~~ 33706 ~~[illegible]~~	
/												M70	
SP	1625	25 SCT E40 BKN 100 BKN	15					30	20	G48	932	PRESRR	TH
SA	1653	E40 BKN 100 BKN	15		936	52	36	30	24	G33	934	PK WND 3048/22	TH
SP	1738	E26 BKN 40 OVC	20					29	28	G42	937	~~PK WND 3042/29~~	NP
/												BINOVC (FIBI)	/
SA	1752	E26 BKN 40 BKN 100 OVC	20		949	50	35	30	24	G34	938	26 BKN V SCT PK WND	NP
/												2842/29	/
SP	1820	30 SCT E60 OVC	20	RW-				30	24	G38	944	PRESRR ~~[illegible]~~	NP
/												(FIBI)	/
SA	1853	30 SCT E60 OVC	15	RW-	966	47	36	30	25	G33	943	RB08 PK WND 2843/40	NP
/		72405 83025 74808 96608	4959/	02261	69942	90610						/ 26106 171/ 69	/
/												20091 46947 0	NP
SA	1954	E40 BKN 90 OVC	15	RW-	960	46	27	31	16	G26	947		CS

U.S. GPO 1978-665-010/1109 Region No. 6

Figure 1-2.--A reproduction of part of the official March 21, 1980, Washington National Airport WBAN form.

GEMTRIX

GENERALIZED EQUIVALENT MARKOV (GEM) MATRIX

TECHNIQUES DEVELOPEMENT LABORATORY

FOR STATION: DCA

VALID FOR 24 HOURS AFTER MAR 21,1980   7 LOCAL

CATEGORY	OBS 7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	0	1	2	3	4	5	6	7
		24 HOURLY FORECASTS AFTER OBSERVATION (LOCAL STANDARD TIME)																							
DB TEMP (F) -130 TO -31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DB TEMP (F) -30 TO -26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DB TEMP (F) -25 TO -21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DB TEMP (F) -20 TO -16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DB TEMP (F) -15 TO -11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DB TEMP (F) -10 TO -6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DB TEMP (F) -5 TO -1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DB TEMP (F) 0 TO 4	0	0	0	0	0	0	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0
DB TEMP (F) 5 TO 9	0	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	0	0	0	0	0	0
DB TEMP (F) 10 TO 14	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	0
DB TEMP (F) 15 TO 19	0	0	-1	-1	-2	-2	-2	-2	-2	-2	-2	-2	-2	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	0
DB TEMP (F) 20 TO 24	0	0	-1	-2	-2	-2	-3	-3	-3	-3	-3	-2	-2	-2	-2	-1	-1	-1	0	0	0	0	1	1	1
DB TEMP (F) 25 TO 29	0	-1	-2	-3	-3	-3	-4	-4	-3	-3	-3	-2	-2	-1	-1	0	1	1	2	2	2	3	3	4	4
DB TEMP (F) 30 TO 34	0	-1	-1	-2	-2	-2	-2	-1	-1	0	1	2	3	3	4	5	6	6	7	8	8	9	9	9	9
DB TEMP (F) 35 TO 39	0	1	2	2	2	2	2	3	3	4	4	5	6	7	8	9	9	10	10	11	11	12	12	12	12
DB TEMP (F) 40 TO 44	0	0	1	1	2	2	3	4	4	5	6	7	8	9	10	11	11	12	12	13	13	14	14	14	13
DB TEMP (F) 45 TO 49	0	1	1	2	3	4	5	6	7	8	9	10	10	11	12	12	13	13	14	14	14	14	14	14	14
DB TEMP (F) 50 TO 54	0	0	3	6	8	10	11	12	12	13	13	13	14	14	15	15	15	15	15	15	15	15	15	15	14
DB TEMP (F) 55 TO 59	0	16	22	23	23	22	21	20	20	19	19	18	18	18	18	18	18	17	17	16	16	16	15	15	14
DB TEMP (F) 60 TO 64	100	66	49	40	34	30	27	25	23	22	21	20	20	19	19	18	18	17	16	16	15	15	14	14	13
DB TEMP (F) 65 TO 69	0	16	21	23	23	22	21	20	19	18	18	17	16	16	16	15	15	14	14	13	13	12	12	12	11
DB TEMP (F) 70 TO 74	0	1	4	6	8	9	10	10	10	10	10	10	10	11	10	10	10	9	9	8	8	7	7	7	7
DB TEMP (F) 75 TO 79	0	2	3	3	4	4	4	5	5	5	5	5	6	6	5	4	3	3	2	1	0	0	-1	-1	1
DB TEMP (F) 80 TO 84	0	1	2	3	3	3	2	2	2	3	3	2	2	0	-1	-3	-3	-4	-4	-4	-5	-5	-5	-5	-4
DB TEMP (F) 85 TO 89	0	-1	0	1	3	3	3	3	2	2	1	0	-2	-3	-4	-5	-5	-5	-5	-5	-5	-5	-5	-5	-4
DB TEMP (F) 90 TO 94	0	0	-1	0	0	2	2	3	3	2	1	-1	-2	-3	-3	-3	-4	-4	-4	-4	-3	-3	-3	-3	-3
DB TEMP (F) 95 TO 99	0	0	0	0	0	0	1	1	1	1	0	0	-1	-1	-1	-1	-2	-2	-2	-2	-2	-1	-1	-1	-1
DB TEMP (F) 100 TO 104	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DB TEMP (F) 105 TO 109	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DB TEMP (F) 110 TO 140	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DPT DPR (F) 0	0	13	15	14	12	11	10	10	9	9	9	9	9	9	9	9	9	9	9	9	10	10	10	10	9
DPT DPR (F) 1	100	51	33	25	21	18	17	15	14	14	13	13	13	13	13	13	13	13	13	14	14	15	15	15	12
DPT DPR (F) 2 TO 4	0	33	43	42	40	39	37	36	35	34	34	34	35	37	38	40	41	43	44	45	45	46	46	45	41
DPT DPR (F) 5 TO 7	0	-5	0	4	7	9	10	10	11	12	13	15	18	20	22	23	24	24	25	24	24	23	23	23	24
DPT DPR (F) 8 TO 11	0	2	2	0	-1	-1	0	1	2	4	6	9	12	14	15	16	15	15	15	14	13	13	12	13	15
DPT DPR (F) 12 TO 15	0	2	4	4	2	0	-1	-1	0	1	2	4	6	7	7	6	6	5	5	4	4	4	3	4	5
DPT DPR (F) 16 TO 19	0	1	3	4	4	4	2	2	1	1	2	3	3	2	1	1	0	0	0	0	-1	-1	-1	0	1
DPT DPR (F) 20 TO 25	0	0	1	3	6	7	7	6	5	5	4	3	2	-1	-2	-2	-3	-3	-3	-3	-3	-3	-3	-3	-2
DPT DPR (F) 26 TO 35	0	-1	0	2	5	7	10	11	11	10	8	4	1	-2	-3	-3	-4	-4	-4	-4	-4	-4	-4	-4	-3
DPT DPR (F) 36 TO 50	0	0	0	2	3	5	6	8	8	8	6	4	2	1	0	-1	-1	-1	-1	-1	-2	-2	-2	-2	-1
DPT DPR (F) 51 TO 99	0	0	0	0	0	1	2	3	3	3	2	1	1	0	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1
VSBY (ST MI) .00 TO .49	0	-1	-2	-2	-2	-1	-1	0	0	1	1	1	1	2	2	2	2	2	2	2	3	3	3	3	3
VSBY (ST MI) .50 TO .74	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2
VSBY (ST MI) .75 TO .99	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
VSBY (ST MI) 1.0 TO 1.49	0	2	3	4	4	4	4	4	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
VSBY (ST MI) 1.5 TO 1.99	0	2	3	4	4	4	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	3	3
VSBY (ST MI) 2.0 TO 2.49	0	4	5	6	6	5	5	5	5	5	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
VSBY (ST MI) 2.5 TO 2.99	0	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	1	1	2	2	2
VSBY (ST MI) 3.0 TO 3.99	0	4	6	7	7	6	6	6	6	6	6	6	5	5	5	5	5	5	5	5	5	5	5	6	6
VSBY (ST MI) 4.0 TO 4.99	0	4	6	6	6	6	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6
VSBY (ST MI) 5.0 TO 5.99	0	10	9	7	6	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6
VSBY (ST MI) 6.0 TO 6.99	100	37	16	8	5	4	3	3	3	3	3	4	4	4	4	4	4	3	4	4	4	4	4	4	4
VSBY (ST MI) 7.0 TO 100.	0	34	49	57	61	63	64	64	65	65	64	65	65	66	66	66	66	66	66	66	65	65	63	60	59
NO F,IF	0	21	36	46	53	59	63	66	69	71	72	74	75	76	77	78	79	79	79	80	80	79	79	79	79
F,IF	100	79	64	54	47	41	37	34	31	29	28	26	25	24	23	22	21	21	21	20	20	21	21	21	21
NO GF	100	100	102	102	102	102	101	101	101	101	100	100	100	100	99	99	99	98	98	97	97	96	95	94	95
GF	0	0	-2	-2	-2	-2	-1	-1	-1	-1	0	0	0	0	1	1	1	2	2	3	3	4	5	6	5
NO K,H,D,KH,KD,HD,KHD	100	112	114	113	111	109	107	105	103	101	100	98	98	97	97	96	96	96	95	95	95	94	94	92	90
K,H,D,KH,KD,HD,KHD	0	-12	-14	-13	-11	-9	-7	-5	-3	-1	0	2	2	3	3	4	4	4	5	5	5	6	6	8	10
NO BS,BD,BN	100	100	100	99	99	99	99	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98
BS,BD,BN	0	0	0	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
NO L	100	96	94	93	93	93	94	94	94	94	95	95	95	95	95	96	96	96	96	96	96	96	96	96	96
L-	0	4	6	6	7	7	6	6	6	6	5	5	5	5	5	4	4	4	4	4	4	4	4	4	4
L,L+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NO R	0	22	38	49	58	64	69	73	76	78	79	81	82	83	84	84	85	85	86	86	86	87	87	87	88
R-	100	74	58	46	38	33	28	25	22	20	19	18	17	16	15	15	14	14	13	13	13	12	12	12	11
R	0	4	4	4	3	3	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
R+	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NO RW	100	96	93	92	92	91	91	91	90	90	91	91	91	91	92	92	92	92	93	93	93	93	93	93	94
RW-	0	3	6	7	7	8	8	9	9	9	9	8	8	8	8	7	7	7	7	6	6	6	6	6	6
RW	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
RW+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NO S,IC	100	99	98	97	97	96	96	96	95	95	95	95	95	94	94	94	94	94	94	94	94	94	94	94	94
S-,IC-	0	1	2	2	3	3	4	4	4	4	5	5	5	5	5	5	5	5	6	5	5	6	6	6	5
S,IC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S+,IC+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NO SW,IP	100	100	99	99	99	98	98	98	98	98	97	97	97	97	97	97	97	97	97	97	97	97	97	97	97
SW-,IP-	0	0	1	1	1	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	2
SW,IP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SW+,IP+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NO ZL	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ZL-,ZL,ZL+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NO ZR	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
ZR-,ZR,ZR+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NO TSTM,A	100	100	99	99	99	99	99	98	98	98	98	98	98	98	98	99	99	99	99	99	99	99	99	99	99
TSTM,A	0	0	1	1	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
NO TSTM+	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
TSTM+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 1-3.--GEM's predicted hourly probability distributions (GEMTRIX) of all subsequent weather conditions, 1-24 hr, for March 21, 1980, at Washington National Airport. Negative and greater than unity estimates of the probabilities are due to nonadditivity and should not be of great concern, since thresholding alleviates the problem.

WIND CALM / LT 2	0	-3	-4	-5	-5	-5	-5	-5	-5	-4	-4	-3	-1	0	1	2	3	4	4	5	6	6	6	6	4
WIND NNE TO NE / LE 10	0	0	1	1	1	1	1	1	1	2	2	3	3	3	3	3	3	3	3	3	3	4	4	4	4
WIND NNE TO NE / 11 TO 19	0	3	4	5	5	5	5	5	5	5	4	4	4	3	3	3	3	3	3	3	3	3	3	3	3
WIND ENE TO E / LE 10	0	-1	0	1	1	1	1	2	2	2	2	3	3	4	4	4	4	4	3	4	4	4	4	4	4
WIND ENE TO E / 11 TO 19	0	1	2	3	4	4	4	4	4	4	4	3	3	3	3	2	2	2	2	2	2	2	2	2	2
WIND ESE TO SE / LE 10	0	4	4	4	4	3	3	2	2	2	2	3	4	5	5	4	4	4	4	4	3	2	2	2	4
WIND ESE TO SE / 11 TO 19	0	9	10	10	9	8	7	7	6	6	6	5	4	4	4	3	3	3	2	2	2	2	2	2	2
WIND SSE TO S / LE 10	0	10	10	9	7	6	4	3	3	3	3	4	5	6	6	6	5	5	5	5	4	4	4	4	3
WIND SSE TO S / 11 TO 19	100	63	43	33	26	22	19	17	16	15	13	12	10	10	9	9	8	8	7	7	6	6	6	6	6
WIND SSW TO SW / LE 10	0	-4	-1	1	2	2	2	2	2	2	2	3	4	4	5	5	5	6	6	6	6	5	5	5	5
WIND SSW TO SW / 11 TO 19	0	7	12	15	16	16	16	15	14	13	12	11	9	9	9	9	9	9	8	8	8	8	8	8	8
WIND WSW TO W / LE 10	0	-2	-3	-2	-1	0	1	1	2	2	3	4	5	5	5	5	5	5	6	6	6	6	6	6	6
WIND WSW TO W / 11 TO 19	0	1	3	5	8	10	11	13	13	13	12	11	10	9	9	9	9	9	9	9	9	5	5	5	9
WIND WNW TO NW / LE 10	0	1	0	-1	-1	-1	-1	0	0	1	2	3	3	3	3	3	4	4	4	4	5	5	5	5	5
WIND WNW TO NW / 11 TO 19	0	2	2	3	4	5	7	8	9	10	10	10	9	9	9	9	9	9	9	9	9	10	10	10	10
WIND NNW TO N / LE 10	0	1	0	-1	-1	-1	-1	0	0	0	1	2	2	3	3	3	3	3	4	4	4	5	5	5	5
WIND NNW TO N / 11 TO 19	0	0	1	2	3	3	4	5	5	6	6	5	5	5	5	5	5	5	5	5	5	5	5	5	6
WIND NNE TO E / GE 20	0	1	1	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1
WIND ESE TO S / GE 20	0	4	5	5	5	5	4	4	4	4	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2
WIND SSW TO W / GE 20	0	3	6	8	9	10	10	10	10	10	9	8	8	8	8	7	7	7	7	7	7	6	6	6	6
WIND NNW TO N / GE 20	0	1	2	3	4	4	5	5	5	6	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
SLP (MB) 800.0 TO 985	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
SLP (MB) 985.1 TO 990	0	0	0	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
SLP (MB) 990.1 TO 995	0	4	7	9	10	11	11	11	11	11	11	10	10	10	9	9	9	8	8	8	8	7	7	7	7
SLP (MB) 995.1 TO 1000	100	85	72	63	55	49	44	40	37	34	31	29	27	26	24	23	22	20	20	19	18	17	16	16	15
SLP (MB) 1000.1 TO 1005	0	12	21	27	31	34	36	37	38	38	37	37	36	34	33	32	31	30	29	29	28	27	26	25	24
SLP (MB) 1005.1 TO 1010	0	0	2	4	7	11	14	18	20	23	24	25	26	26	26	26	27	27	28	28	28	28	28	27	26
SLP (MB) 1010.1 TO 1015	0	-1	-1	-1	0	2	4	7	8	10	11	11	12	12	12	13	14	15	16	17	19	19	19	19	18
SLP (MB) 1015.1 TO 1020	0	0	0	0	0	1	1	1	1	1	1	2	3	5	6	7	7	8	8	9	9	10	11	12	12
SLP (MB) 1020.1 TO 1025	0	0	-1	-1	-2	-3	-5	-6	-7	-7	-7	-6	-5	-4	-3	-2	-2	-2	-2	-2	-2	-1	0	1	3
SLP (MB) 1025.1 TO 1030	0	-1	-1	-2	-3	-4	-6	-7	-7	-8	-8	-8	-7	-7	-7	-6	-6	-6	-6	-6	-6	-6	-5	-5	-4
SLP (MB) 1030.1 TO 1035	0	0	0	0	-1	-1	-2	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
SLP (MB) 1035.1 TO 1040	0	0	0	0	0	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
SLP (MB) 1040.1 TO 1090	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CLD CVR #1 CLR	0	-1	-2	-4	-7	-9	-11	-12	-13	-13	-11	-8	-5	-1	2	6	7	9	10	11	11	10	6	3	1
CLD CVR #1 SCD	0	25	31	34	37	39	40	42	44	45	45	42	40	36	33	31	29	28	27	26	26	28	31	34	36
CLD CVR #1 BKN	100	33	39	34	33	33	33	32	31	30	29	28	27	26	25	24	23	23	22	22	22	22	23	25	26
CLD CVR #1 OVC	0	24	32	35	36	36	36	35	34	34	33	33	34	34	35	35	35	35	35	35	35	34	33	32	31
CLD CVR #1 TOT OBSC	0	0	0	0	1	1	2	3	3	4	4	4	5	5	5	5	5	5	5	6	6	6	6	6	6
CLD HGT #1 0 TO 1	0	-1	-1	-1	-1	0	0	0	1	1	1	1	2	2	2	2	2	2	2	2	2	3	3	3	2
CLD HGT #1 2 TO 4	0	10	15	16	16	16	14	13	13	12	11	11	10	10	10	9	9	9	9	9	9	9	9	9	9
CLD HGT #1 5 TO 6	0	20	22	20	18	16	14	13	12	11	10	10	9	9	8	8	8	8	7	7	7	7	7	7	7
CLD HGT #1 7 TO 9	100	57	39	30	25	21	19	17	16	15	14	13	13	12	11	11	11	11	10	10	10	10	10	10	10
CLD HGT #1 10 TO 14	0	18	24	24	23	21	20	19	18	17	16	15	15	14	14	13	13	13	13	13	12	12	12	12	13
CLD HGT #1 15 TO 19	0	2	7	10	11	12	12	12	12	11	11	11	10	10	10	9	9	9	9	9	9	9	9	9	9
CLD HGT #1 20 TO 24	0	1	3	5	7	7	8	8	8	8	8	8	8	7	7	7	7	7	7	7	7	7	7	7	7
CLD HGT #1 25 TO 29	0	0	1	3	5	6	7	7	7	7	7	7	6	6	6	6	6	6	6	6	6	6	6	6	6
CLD HGT #1 30 TO 39	0	-2	-1	2	6	9	10	11	11	11	11	10	9	9	8	8	8	8	8	8	8	8	8	8	8
CLD HGT #1 40 TO 49	0	-1	-1	0	2	5	7	8	9	9	9	8	6	5	5	5	4	4	4	4	4	4	4	5	5
CLD HGT #1 50 TO 59	0	-1	-1	-1	0	1	3	4	5	6	6	5	4	3	3	2	2	2	2	2	2	2	3	3	3
CLD HGT #1 60 TO 75	0	-1	-2	-2	-1	0	1	2	3	4	4	4	3	3	3	2	2	2	2	2	2	3	3	3	3
CLD HGT #1 76 TO 99	0	0	-1	-1	0	0	0	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3
CLD HGT #1 100 TO 150	0	1	0	-1	-2	-3	-3	-3	-2	-1	0	1	3	3	3	4	4	4	4	4	4	5	6	7	7
CLD HGT #1 151 TO UNL	0	1	0	-3	-7	-10	-13	-14	-14	-13	-11	-7	-3	2	6	9	10	11	12	12	12	11	8	6	5
CLD HGT #1 PART OBSC	0	-2	-2	-2	-1	0	0	1	1	2	2	2	2	3	3	3	3	3	3	3	4	4	4	4	4
CLD CVR #2 CLR	0	22	29	31	32	33	33	33	33	33	35	38	43	49	54	57	59	61	62	63	63	62	59	55	52
CLD CVR #2 SCD	0	0	2	4	5	6	7	8	8	9	9	9	7	5	4	3	3	3	2	2	2	3	4	5	6
CLD CVR #2 BKN	0	13	16	17	18	19	19	20	21	21	20	19	17	14	12	11	10	10	9	9	9	9	11	13	15
CLD CVR #2 OVC	100	65	53	48	45	43	41	39	38	37	36	34	33	31	30	29	28	27	27	26	26	26	26	27	28
CLD HGT #2 0 TO 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CLD HGT #2 2 TO 4	0	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CLD HGT #2 5 TO 6	0	3	3	3	3	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CLD HGT #2 7 TO 9	0	9	8	6	5	4	4	4	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2
CLD HGT #2 10 TO 14	100	37	20	14	11	10	8	7	7	6	6	5	5	5	4	4	4	4	4	4	4	4	4	4	4
CLD HGT #2 15 TO 19	0	12	11	9	7	6	6	5	5	4	4	4	4	4	3	3	3	3	3	3	3	3	3	3	3
CLD HGT #2 20 TO 24	0	6	7	7	6	6	5	5	4	4	4	4	4	3	3	3	3	3	3	3	3	3	3	3	3
CLD HGT #2 25 TO 29	0	4	5	5	5	5	4	4	4	4	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2
CLD HGT #2 30 TO 39	0	4	6	7	7	7	6	6	6	6	5	5	5	5	4	4	4	4	4	4	4	4	4	4	4
CLD HGT #2 40 TO 49	0	2	3	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	3	3	3	3
CLD HGT #2 50 TO 59	0	1	2	2	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2
CLD HGT #2 60 TO 75	0	1	2	3	4	4	4	5	5	5	5	5	4	4	4	4	4	4	4	4	4	4	4	4	4
CLD HGT #2 76 TO 99	0	1	2	3	3	4	4	4	5	5	5	5	4	4	3	3	3	3	3	3	3	3	3	3	4
CLD HGT #2 100 TO 150	0	-1	0	2	4	6	7	8	9	10	10	10	8	7	6	5	4	4	4	4	4	4	5	6	7
CLD HGT #2 151 TO UNL	0	21	29	33	36	38	40	41	42	43	45	48	51	55	58	60	62	63	64	65	65	65	64	62	61
TOTAL CLD CVR CLR	0	-1	-2	-4	-7	-9	-11	-12	-13	-13	-11	-8	-5	-1	2	6	7	9	10	11	11	10	6	3	1
TOTAL CLD CVR SCD	0	-3	-3	-1	1	3	5	7	8	9	9	8	8	8	7	7	7	6	6	6	6	7	9	10	10
TOTAL CLD CVR BKN	0	2	3	6	9	12	15	16	17	18	17	17	16	14	12	11	11	11	11	11	11	12	13	16	17
TOTAL CLD CVR OVC	100	102	102	99	97	94	92	89	88	86	85	83	81	80	78	76	75	74	73	73	72	72	71	72	71
CEILING 0 TO 1	0	-1	-2	-1	-1	0	0	0	1	1	1	1	1	2	2	2	2	2	2	2	2	3	3	3	2
CEILING 2 TO 4	0	6	10	12	13	12	12	11	10	10	9	9	9	8	8	8	8	7	7	7	7	7	7	7	7
CEILING 5 TO 6	0	16	18	16	14	13	12	11	10	9	8	8	8	7	7	7	6	6	6	6	6	6	6	6	6
CEILING 7 TO 9	100	50	32	24	20	17	16	14	13	12	11	11	10	10	9	9	9	9	9	8	8	8	8	8	8
CEILING 10 TO 14	0	20	23	22	20	18	17	16	15	14	13	13	12	12	11	11	11	11	11	10	10	10	10	10	10
CEILING 15 TO 19	0	5	8	10	10	10	10	10	9	9	9	8	8	8	8	8	7	7	7	7	7	7	7	7	7
CEILING 20 TO 24	0	3	5	6	7	7	7	7	7	6	6	6	6	6	6	6	6	6	6	6	5	5	5	5	5
CEILING 25 TO 29	0	2	3	4	5	6	6	6	6	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
CEILING 30 TO 39	0	2	3	5	6	8	8	8	8	8	8	8	8	8	7	7	7	7	7	7	7	7	7	7	7
CEILING 40 TO 49	0	1	2	3	3	4	5	6	6	6	6	6	6	6	6	6	5	5	5	5	5	5	5	5	5
CEILING 50 TO 59	0	0	1	1	2	2	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3
CEILING 60 TO 75	0	0	1	1	2	2	3	3	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5
CEILING 76 TO 99	0	1	1	1	1	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
CEILING 100 TO 150	0	-1	-1	0	0	1	1	2	3	4	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6
CEILING 151 TO UNL	0	-3	-4	-4	-3	-3	-2	0	2	4	7	9	11	12	14	16	17	18	18	19	19	20	20	20	20
MONTH (LOCAL) JANUARY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MONTH (LOCAL) FEBRUARY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MONTH (LOCAL) MARCH	100	100	100	100	99	99	99	99	99	99	98	98	98	96	98	98	98	97	97	97	97	97	97	97	96
MONTH (LOCAL) APRIL	0	0	0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3	3	3	3	3	3
MONTH (LOCAL) MAY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MONTH (LOCAL) JUNE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MONTH (LOCAL) JULY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MONTH (LOCAL) AUGUST	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MONTH (LOCAL) SEPTEMBER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MONTH (LOCAL) OCTOBER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MONTH (LOCAL) NOVEMBER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MONTH (LOCAL) DECEMBER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HOUR (LOCAL) 0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0
HOUR (LOCAL) 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0
HOUR (LOCAL) 2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0
HOUR (LOCAL) 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0
HOUR (LOCAL) 4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0
HOUR (LOCAL) 5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0
HOUR (LOCAL) 6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
HOUR (LOCAL) 7	100	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HOUR (LOCAL) 8	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HOUR (LOCAL) 9	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HOUR (LOCAL) 10	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HOUR (LOCAL) 11	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HOUR (LOCAL) 12	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HOUR (LOCAL) 13	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HOUR (LOCAL) 14	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HOUR (LOCAL) 15	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HOUR (LOCAL) 16	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HOUR (LOCAL) 17	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0
HOUR (LOCAL) 18	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0
HOUR (LOCAL) 19	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0
HOUR (LOCAL) 20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0
HOUR (LOCAL) 21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0
HOUR (LOCAL) 22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0
HOUR (LOCAL) 23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0

Figure 1-3.--(continued)

NO WX	0	24	36	42	46	48	49	50	50	50	51	51	51	52	53	53	54	54	54	54	53	53	51	49	49
WX	100	76	64	58	54	52	51	50	50	50	49	49	49	48	47	47	46	46	46	46	47	47	49	51	51
AUTWTR / DAY7-18 FALSE	0	2	4	6	8	10	11	13	14	16	17	18	70	70	71	72	73	73	74	74	74	75	75	75	25
AUTWTR / DAY7-18 TRUE	100	98	96	94	92	90	89	87	86	84	83	82	30	30	29	28	27	27	26	26	26	25	25	25	75
AUTWTR / HUMID FALSE	0	-2	8	17	25	31	36	40	44	46	48	49	50	50	50	51	51	51	51	51	52	52	52	52	54
AUTWTR / HUMID TRUE	100	102	92	83	75	69	64	60	56	54	52	51	50	50	50	49	49	49	49	49	48	48	48	48	46
AUTWTR / STHWIND FALSE	0	9	17	24	30	34	38	42	44	47	48	50	51	52	52	53	54	55	56	57	57	58	58	59	59
AUTWTR / STHWIND TRUE	100	91	83	76	70	66	62	58	56	53	52	50	49	48	48	47	46	45	44	43	43	42	42	41	41
AUTWTR / ESTWIND FALSE	0	27	40	48	54	58	62	64	66	67	67	67	68	68	69	70	70	71	72	72	72	72	72	72	72
AUTWTR / ESTWIND TRUE	100	73	60	52	46	42	38	36	34	33	33	33	32	32	31	30	30	29	28	28	28	28	28	28	28
AUTWTR / OVCSKY FALSE	0	-2	-1	2	5	8	11	14	17	20	22	25	27	29	31	32	34	35	36	37	38	39	40	40	41
AUTWTR / OVCSKY TRUE	100	102	101	98	95	92	89	86	83	80	78	75	73	71	69	68	66	65	64	63	62	61	60	60	59
AUTWTR / HISKY FALSE	100	103	104	102	100	98	96	93	91	88	86	84	81	80	78	76	75	74	73	72	71	70	70	69	69
AUTWTR / HISKY TRUE	0	-3	-4	-2	0	2	4	7	9	12	14	16	19	20	22	24	25	26	27	28	29	30	30	31	31
AUTWTR / FARVSBY FALSE	100	77	64	57	51	47	44	42	40	39	38	37	35	34	34	33	33	32	32	32	32	31	31	32	33
AUTWTR / FARVSBY TRUE	0	23	36	43	49	53	56	58	60	61	62	63	65	66	66	67	67	68	68	68	68	69	69	68	67
AUTWTR / NO WX FALSE	100	82	70	61	55	51	47	44	42	40	38	37	36	35	34	33	32	32	31	31	30	30	30	30	29
AUTWTR / NO WX TRUE	0	18	30	39	45	49	53	56	58	60	62	63	64	65	66	67	68	68	69	69	70	70	70	70	71
DAY7-18 / HUMID FALSE	0	3	18	32	42	50	55	60	62	64	65	65	80	81	82	83	84	85	86	86	87	88	88	89	48
DAY7-18 / HUMID TRUE	100	97	82	68	58	50	45	40	38	36	35	35	20	19	18	17	16	15	14	14	13	12	12	11	52
DAY7-18 / STHWIND FALSE	0	12	21	28	33	37	40	43	46	48	49	50	97	98	98	99	99	100	100	100	100	101	101	101	58
DAY7-18 / STHWIND TRUE	100	88	79	72	67	63	60	57	54	52	51	50	3	2	2	1	1	0	0	0	0	-1	-1	-1	42
DAY7-18 / ESTWIND FALSE	0	30	45	53	58	62	65	67	68	68	68	67	105	106	106	106	106	106	106	106	106	106	106	106	72
DAY7-18 / ESTWIND TRUE	100	70	55	47	42	38	35	33	32	32	32	33	-5	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	28
DAY7-18 / OVCSKY FALSE	0	3	8	14	20	24	28	31	34	36	37	39	73	75	76	77	78	79	79	80	81	81	82	83	45
DAY7-18 / OVCSKY TRUE	100	97	92	86	80	76	72	69	66	64	63	61	27	25	24	23	22	21	21	20	19	19	18	17	55
DAY7-18 / HISKY FALSE	100	98	94	90	87	84	81	78	75	72	70	67	130	129	128	127	126	125	124	123	123	122	121	120	62
DAY7-18 / HISKY TRUE	0	2	6	10	13	16	19	22	25	28	30	33	-30	-29	-28	-27	-26	-25	-24	-23	-23	-22	-21	-20	38
DAY7-18 / FARVSBY FALSE	100	66	50	40	34	31	28	27	26	25	25	25	112	112	112	111	111	110	110	110	109	109	109	108	34
DAY7-18 / FARVSBY TRUE	0	34	50	60	66	69	72	73	74	75	75	75	-12	-12	-12	-11	-11	-10	-10	-10	-9	-9	-9	-8	66
DAY7-18 / NO WX FALSE	100	76	60	50	43	39	35	32	30	29	27	26	115	114	114	113	112	112	111	111	111	110	110	110	20
DAY7-18 / NO WX TRUE	0	24	40	50	57	61	65	68	70	71	73	74	-15	-14	-14	-13	-12	-12	-11	-11	-11	-10	-10	-10	80
HUMID / STHWIND FALSE	0	19	36	48	56	61	65	68	70	71	72	73	73	72	72	71	70	70	70	70	69	69	69	70	73
HUMID / STHWIND TRUE	100	81	64	52	44	39	35	32	30	29	28	27	27	28	28	29	30	30	30	30	31	31	31	30	27
HUMID / ESTWIND FALSE	0	27	45	56	63	68	71	73	75	76	77	78	77	77	77	77	77	76	76	76	76	76	75	76	78
HUMID / ESTWIND TRUE	100	73	55	44	37	32	29	27	25	24	23	22	23	23	23	23	23	24	24	24	24	24	25	24	22
HUMID / OVCSKY FALSE	0	-4	3	12	20	27	32	37	40	43	45	46	47	48	49	50	51	51	52	52	52	52	52	53	54
HUMID / OVCSKY TRUE	100	104	97	88	80	73	68	63	60	57	55	54	53	52	51	50	49	49	48	48	48	48	48	47	46
HUMID / HISKY FALSE	100	104	109	110	109	107	106	105	104	103	102	101	99	97	94	92	89	87	85	83	82	80	80	81	87
HUMID / HISKY TRUE	0	-4	-9	-10	-9	-7	-6	-5	-4	-3	-2	-1	1	3	6	8	11	13	15	17	18	20	20	19	13
HUMID / FARVSBY FALSE	100	76	73	73	74	75	76	77	78	78	78	78	76	74	72	70	68	66	65	63	62	61	62	65	72
HUMID / FARVSBY TRUE	0	24	27	27	26	25	24	23	22	22	22	22	24	26	28	30	32	34	35	37	38	39	38	35	28
HUMID / NO WX FALSE	100	82	82	82	82	81	80	80	80	79	79	77	76	73	70	68	65	63	60	58	56	54	53	55	62
HUMID / NO WX TRUE	0	18	18	18	18	19	20	20	20	21	21	23	24	27	30	32	35	37	40	42	44	46	47	45	38
STHWIND / ESTWIND FALSE	0	39	56	65	71	75	79	81	83	84	84	84	84	84	84	85	86	88	89	89	90	90	90	90	90
STHWIND / ESTWIND TRUE	100	61	44	35	29	25	21	19	17	16	16	16	16	16	16	15	14	12	11	11	10	10	10	10	10
STHWIND / OVCSKY FALSE	0	12	23	31	38	44	48	52	54	56	58	60	61	63	64	65	66	66	67	67	68	69	69	69	69
STHWIND / OVCSKY TRUE	100	88	77	69	62	56	52	48	46	44	42	40	39	37	36	35	34	34	33	33	32	31	31	31	31
STHWIND / HISKY FALSE	100	94	91	88	87	87	88	88	89	89	90	89	89	89	88	88	88	89	90	91	92	93	93	94	94
STHWIND / HISKY TRUE	0	6	9	12	13	13	12	12	11	11	10	11	11	11	12	12	12	11	10	9	8	7	7	6	6
STHWIND / FARVSBY FALSE	100	64	49	43	42	43	45	49	52	54	57	58	60	60	61	62	62	64	65	67	68	70	72	73	74
STHWIND / FARVSBY TRUE	0	36	51	57	58	57	55	51	48	46	43	42	40	40	39	38	38	36	35	33	32	30	28	27	26
STHWIND / NO WX FALSE	100	70	58	53	52	53	54	56	58	60	62	63	63	63	63	64	64	65	67	68	69	70	71	71	70
STHWIND / NO WX TRUE	0	30	42	47	48	47	46	44	42	40	38	37	37	37	37	36	36	35	33	32	31	30	29	29	30
ESTWIND / OVCSKY FALSE	0	27	41	50	56	61	64	67	68	69	70	71	72	73	74	74	75	76	76	76	77	77	76	76	76
ESTWIND / OVCSKY TRUE	100	73	59	50	44	39	36	33	32	31	30	29	28	27	26	26	25	24	24	24	23	23	24	24	24
ESTWIND / HISKY FALSE	100	108	107	105	103	103	103	103	103	102	101	100	99	98	98	99	99	100	101	101	102	102	101	101	101
ESTWIND / HISKY TRUE	0	-8	-7	-5	-3	-3	-3	-3	-3	-2	-1	0	1	2	2	1	1	0	-1	-1	-2	-2	-1	-1	-1
ESTWIND / FARVSBY FALSE	100	95	88	84	83	83	83	84	85	85	85	85	84	84	84	85	86	87	88	89	90	90	90	90	91
ESTWIND / FARVSBY TRUE	0	5	12	16	17	17	17	16	15	15	15	15	16	16	16	15	14	13	12	11	10	10	10	10	9
ESTWIND / NO WX FALSE	100	99	94	90	88	88	87	87	87	87	86	85	84	84	84	85	86	87	88	88	88	88	88	87	86
ESTWIND / NO WX TRUE	0	1	6	10	12	12	13	13	13	13	14	15	16	16	16	15	14	13	12	12	12	12	12	13	14
OVCSKY / HISKY FALSE	100	99	99	99	99	98	98	98	97	97	96	97	97	97	98	98	98	98	98	98	98	98	98	98	97
OVCSKY / HISKY TRUE	0	1	1	1	1	2	2	2	3	3	4	3	3	3	2	2	2	2	2	2	2	2	2	2	3
OVCSKY / FARVSBY FALSE	100	74	63	57	55	54	53	53	53	53	53	54	54	55	55	56	56	57	57	57	58	59	59	60	61
OVCSKY / FARVSBY TRUE	0	26	37	43	45	46	47	47	47	47	47	46	46	45	45	44	44	43	43	43	42	41	41	40	39
OVCSKY / ON WX FALSE	100	84	73	67	63	60	59	58	57	57	56	56	57	58	58	59	59	59	59	59	59	59	59	59	58
OVCSKY / NO WX TRUE	0	16	27	33	37	40	41	42	43	43	44	44	43	42	42	41	41	41	41	41	41	41	41	41	42
HISKY / FARVSBY FALSE	100	93	90	90	91	92	93	93	92	91	90	89	88	87	86	85	85	84	84	84	84	85	86	88	88
HISKY / FARVSBY TRUE	0	7	10	10	9	8	7	7	8	9	10	11	12	13	14	15	15	16	16	16	16	15	14	12	12
HISKY / NO WX FALSE	100	103	104	104	103	103	102	100	98	96	93	91	89	88	86	85	83	82	82	81	81	81	81	81	80
HISKY / NO WX TRUE	0	-3	-4	-4	-3	-3	-2	0	2	4	7	9	11	12	14	15	17	18	18	19	19	19	19	19	20

Figure 1-3.--(concluded)

```
 XXX   XXXXX X   X
X      X     XX XX
X XXX  XXX   X X X
X   X  X     X   X
 XXX   XXXXX X   X
```

TECHNIQUES DEVELOPMENT LABORATORY

FOR STATION: DCA

VALID FOR 12 HOURS AFTER MAR 21,1980 7 LOCAL

	OB	12 HOURLY FORECASTS (LOCAL STANDARD TIME)											
	7	8	9	10	11	12	13	14	15	16	17	18	19
TEMPERATURE(F)	60	61	64	66	67	67	68	68	67	66	64	62	59
DEW POINT TP(F)	59	60	61	60	60	59	58	57	56	55	54	53	53
VSBY(100THS SM)	0600	0600	0600	0600	0600	0600	0600	0500	0500	0400	0400	0400	0400
FOG, ICE FOG	F	F	F	F	F	F	F	F	F	F	F	F	F
GROUND FOG													
SMOKE, HAZE													
BLOWING													
DRIZZLE													
RAIN	R-	R-	R-	R-	R-	R	R	R-	R-	R-	R-	R-	R-
RAIN SHOWER					RW-	RW-	RW-	RW-	RW-	RW-	RW-	RW-	RW-
SNOW, IC													
SNOW SHOWER, IP													
FREEZE DRIZZLE													
FREEZE RAIN													
THUNDERSTORM													
THUNDERSTORM+													
WIND(DDFF)	1513	1718	1719	1820	1921	2021	2121	2221	2321	2321	2420	2419	2418
SLP(10THS MB)	9990	9993	9999	10000	10000	9997	9995	9993	9994	9996	10000	10004	10010
CLOUD COVER #1	BKN	BKN	BKN	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC
CLOUD HEIGHT #1	7	7	7	7	7	7	7	7	7	7	7	7	7
CLOUD COVER #2	OVC	OVC	OVC	CLR	CLR	CLR	CLR	CLR	CLR	CLR	CLR	CLR	CLR
CLOUD HEIGHT #2	10	10	10	UNL	UNL	UNL	UNL	UNL	UNL	UNL	UNL	UNL	UNL
TOT CLOUD COVER	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC
CEILING 100S FT	7	7	7	7	7	7	7	7	7	7	7	7	7

Figure 1-4.--GEM categorical predictions for the March 21, 1980, example.

The actual sequence of events was very much in keeping with the forecast. A front or squall line passed around noon and showed an even sharper drop in temperature and dewpoint than predicted. The wind shifted and increased in speed as expected, but in a slightly more dramatic manner. Visibility improved much beyond that predicted by GEM.

In all, the GEM forecast contained useful guidance information. Particularly encouraging was the way the synoptic situation was inferred from only the 0700 LST observation. Incidentally, when GEM was projected out another 12 hours from the same 0700 LST observation, the temperature was predicted to fall another 13°F to 46°F, and this was closely in line with what actually occurred. Moreover, GEM's wind forecast showed a further veering of 30° in direction, which was in line with what was observed.

Overview of the Report

The work reported here is the culmination of three decades of research in the application of statistics to meteorological prediction. GEM is a multivariate linear regression system in which all variables, both predictors and predictands, are zero-one. The model underlying the system is Markovian. It uses only the most recent observation of the local surface weather elements to predict the probability distribution of those same weather elements. It does this in hourly increments. A categorical forecast is then made of each element, satisfying an arbitrary constraint of balancing the number of times an element category is predicted with the number of times it is observed to occur.

In the period leading up to the development of GEM, a number of findings--sometimes contrary to common belief--were uncovered. Principal among these is the notion of a generalized operator, by which one can use the same equation to forecast anywhere at any time. Early experimental results at the Massachusetts Institute of Technology began breaking down the notion of stratification of data. The procedure of stratifying data was thought to be advantageous in effecting a kind of nonlinearity in the prediction scheme. Deemed desirable was a synoptic climatology, in which past situations similar to the current situation are grouped and predictions are based on these data. However, it later became evident that dissimilar past events were as useful for prediction as similar past events. Furthermore, using <u>all</u> kinds of events (similar and dissimilar) yielded the best results of all--partly, perhaps, because of the larger sample afforded the scheme.

Experiments that followed, notably one by Harris (1962), boldly predicted temperatures at stations all around the contiguous United States by using only one equation. This equation had the predictors and predictands in standard units (accounting for local means and standard deviations), but the same coefficients were applicable to all locations (including the independent test locations). Even before this remarkable result, it was already becoming the common practice of many researchers not to stratify by the season of the year or the time of day. (See, for example, papers in Shorr 1958.)

In view of this earlier work, the results reported here (on the reliability of generalized operators) are not unexpected. However, this represents the first occasion on which a well founded statistical procedure, <u>the analysis of covariance</u>, has been employed in this context to give convincing evidence of its truth.

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.

Earlier in this chapter an example was given of the consequences of using the GEM procedure. Chapter 2 describes the mathematical model and explains how the data were prepared for constructing GEM. This is followed by a detailed explanation of how each weather element was transformed into zero-one events. Discussed also are some of the computational conveniences for the resulting binary data set.

The statistical analyses and data manipulations are given in the subsequent sections of chapter 2, ending with a selected set of material on the procedure's characteristics, for interpretation by the reader. Essentially all of the necessary matrices and other computed quantities are on microfiche and appear in a pocket inside the report's back cover.

Chapter 3 presents results of both old and new experiments in which GEM or its forerunners have been used. Some of these pertain only to independent verifications. Others give details of attempts to resolve the issue of single-station versus generalized operators in an elaborate analysis of covariance experiment. At the end of the chapter, conclusions are drawn from the results of the experiments.

In chapter 4 an independent verification of GEM is presented along with comparative statistics against persistence over the 1- to 12-hr period.

Chapter 5 deals with operational configurations of GEM under a variety of circumstances--involving a large-scale computer, time sharing option (TSO), and minicomputer.

Chapter 6 gives a projected view of GEM from the standpoint of enhancement and other possible applications. The report is summarized in this chapter.

Finally, chapter 7 covers new results-modifications to improve the model and their applications to the independent verification sample showing comparative statistics.

2. CREATING GEM

This chapter describes GEM in its entirety, from the mathematical model to the first step in data selection, and through the making of operational forecasts. It is suggested that Miller, 1968, be read as an introduction to GEM and, following that, Whiton, 1977, for an excellent and exhaustive presentation of the equivalent and Markov aspects of GEM. This should adequately cover all of how GEM was conceived and how it extends in mathematical form. Miller et al, 1977, and Miller, 1979b, might then be read to appraise the consequences of GEM's early comparative capabilities, for ceiling and visibility, under single-station rather than generalized circumstances.

Mathematical model

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

$$
\begin{aligned}
\hat{Y}_1 &= a_{1,0} + a_{1,1}Z_1 + a_{1,2}Z_2 + \ldots + a_{1,p}Z_p + \ldots + a_{1,P}Z_P \\
\hat{Y}_2 &= a_{2,0} + a_{2,1}Z_1 + a_{2,2}Z_2 + \ldots + a_{2,p}Z_p + \ldots + a_{2,P}Z_P \\
&\vdots \\
\hat{Y}_q &= a_{q,0} + a_{q,1}Z_1 + a_{q,2}Z_2 + \ldots + a_{q,p}Z_p + \ldots + a_{q,P}Z_P \\
&\vdots \\
\hat{Y}_Q &= a_{Q,0} + a_{Q,1}Z_1 + a_{Q,2}Z_2 + \ldots + a_{Q,p}Z_p + \ldots + a_{Q,P}Z_P
\end{aligned}
\tag{2-1}
$$

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

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

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 ε_q^2 ($q = 1, 2, \ldots, Q$). This is done by taking the partial derivatives of the ε^2's with respect to the unknown a's and 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):

$$
\underline{A} = (\underline{Z}'\underline{Z})^{-1}(\underline{Y}'\underline{Z})
\tag{2-3}
$$

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

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

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):

$$\hat{\underline{y}}_1 = \underline{z}_0{}'\underline{A} \tag{2-4}$$

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

$$\hat{\underline{y}}_T = \underline{z}_{T-1}\underline{A} \qquad \text{(multiplicative form)} \tag{2-5}$$

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:

$$\hat{\underline{y}}_T = \underline{z}_0\underline{A}^T \qquad \text{(additive form)} \tag{2-6}$$

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 $\underline{A}$, hour by hour. 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 T's desired to complete a forecast.

The GEM model has been demonstrated to converge to climatology when projected out to a large T. (See Whiton, 1977, for further discussion of this point.)

A word about the computing of $\underline{Z}'\underline{Z}$ and $\underline{Y}'\underline{Z}$: With all observed elements being only zeros and ones, the data can be packed into the bits of computer words, and all arithmetic operations performed by very speedy, logical, machine-language instructions. The data need only to be transposed initially from map form to vector form.

Data

Preparation

Steps 1-4 are data preparation activities. Step 5 is data transformation. Steps 6-12 include the statistical analyses.

Step 1 Select Weather Predictors

Notation	Predictor name
X_0	Unity (always one)
X_1	Month of year
X_2	Hour of day
X_3	Sea level pressure
X_4	Dry bulb temperature
X_5	Dew point depression
X_6	Lowest sky cover
X_7	Visiblity
X_8	No weather
X_9	Fog, ice fog
X_{10}	Ground fog

Notation	Predictor name
X_{11}	Smoke, haze, or dust
X_{12}	Blowing snow, dust or spray
X_{13}	Drizzle--light
X_{14}	Drizzle--moderate or heavy
X_{15}	Rain--light
X_{16}	Rain--moderate
X_{17}	Rain--heavy
X_{18}	Rain showers--light
X_{19}	Rain showers--moderate
X_{20}	Rain showers--heavy
X_{21}	Snow or ice--light
X_{22}	Snow or ice--moderate
X_{23}	Snow or ice--heavy
X_{24}	Snow or ice showers--light
X_{25}	Snow or ice showers--moderate
X_{26}	Snow or ice showers--heavy
X_{27}	Freezing drizzle
X_{28}	Freezing rain
X_{29}	Thunderstorm or light hail
X_{30}	Thunderstorm, heavy
X_{31}	Lowest cloud layer height
X_{32}	Middle sky cover
X_{33}	Middle cloud layer height
X_{34}	Total sky cover
X_{35}	Ceiling
X_{36}	Wind
X_{37}	Interactions (gross)

Step 2 Select Weather Predictands

Notation	Predictand name
U_1	Month of year
U_2	Hour of day
U_3	Sea level pressure
U_4	Dry bulb temperature
U_5	Dew point depression
U_6	Lowest sky cover
U_7	Visiblity
U_8	No weather
U_9	Fog, ice fog
U_{10}	Ground fog
U_{11}	Smoke, haze, or dust
U_{12}	Blowing snow, dust or spray
U_{13}	Drizzle--light
U_{14}	Drizzle--moderate or heavy
U_{15}	Rain--light
U_{16}	Rain--moderate
U_{17}	Rain--heavy
U_{18}	Rain showers--light
U_{19}	Rain showers--moderate

Notation	Predictand name
U_{20}	Rain showers--heavy
U_{21}	Snow or ice--light
U_{22}	Snow or ice--moderate
U_{23}	Snow or ice--heavy
U_{24}	Snow or ice showers--light
U_{25}	Snow or ice showers--moderate
U_{26}	Snow or ice showers--heavy
U_{27}	Freezing drizzle
U_{28}	Freezing rain
U_{29}	Thunderstorm or light hail
U_{30}	Thunderstorm, heavy
U_{31}	Lowest cloud layer height
U_{32}	Middle sky cover
U_{33}	Middle cloud layer height
U_{34}	Total sky cover
U_{35}	Ceiling
U_{36}	Wind
U_{37}	Interactions (gross)

Step 3 Select Weather Stations

Symbol		City	State
L_1	I	Albuquerque	New Mexico
L_2		Waco	Texas
L_3		Atlantic City (A)	New Jersey
L_4		Atlantic City (B)	New Jersey
L_5		Albany	New York
L_6		Atlanta	Georgia
L_7	I	Bismarck	North Dakota
L_8		Boise	Idaho
L_9	I	Boston	Massachusetts
L_{10}		Buffalo	New York
L_{11}		Baltimore	Maryland
L_{12}		Columbia	South Carolina
L_{13}		Cleveland	Ohio
L_{14}	I	Denver	Colorado
L_{15}		Duluth	Minnesota
L_{16}		Des Moines	Iowa
L_{17}		Sioux Falls	South Dakota
L_{18}		Great Falls	Montana
L_{19}		Wilmington	Delaware
L_{20}		Jackson	Mississippi
L_{21}	I	Jacksonville	Florida
L_{22}	I	Los Angeles	California
L_{23}		Lubbock	Texas
L_{24}	I	Memphis	Tennessee
L_{25}	I	Milwaukee	Wisconsin
L_{26}	I	Oklahoma City	Oklahoma
L_{27}		Norfolk	Virginia
L_{28}	I	Portland	Oregon

Symbol		City	State
L_{29}		Phoenix	Arizona
L_{30}	I	Pittsburgh	Pennsylvania
L_{31}	I	Raleigh-Durham	North Carolina
L_{32}	I	Reno	Nevada
L_{33}		Roanoke	Virginia
L_{34}	I	San Antonio	Texas
L_{35}		Savannah	Georgia
L_{36}		Louisville	Kentucky
L_{37}		Seattle-Tacoma	Washington
L_{38}	I	Saint Louis	Missouri
L_{39}		Tallahassee	Florida
L_{40}		Topeka	Kansas
L_{41}		Knoxville	Tennessee

Depicted spatially on the map in figure 2-1. The symbol I denotes station is part of analyses of variance and covariance sample.

Step 4 Select Sample of Observations

The following observation samples came from the years 1954-1965. Atlantic City appears in two forms because of a change in observation site during the period.

Weather station Symbol	Sample size Notation	Sample size Actual
L_1	N_1	105,002
L_2	N_2	101,521
L_3	N_3	47,662
L_4	N_4	56,879
L_5	N_5	103,673
L_6	N_6	105,000
L_7	N_7	105,011
L_8	N_8	101,105
L_9	N_9	104,989
L_{10}	N_{10}	103,371
L_{11}	N_{11}	87,562
L_{12}	N_{12}	104,341
L_{13}	N_{13}	104,951
L_{14}	N_{14}	104,401
L_{15}	N_{15}	104,999
L_{16}	N_{16}	105,025
L_{17}	N_{17}	105,047
L_{18}	N_{18}	98,902
L_{19}	N_{19}	43,275
L_{20}	N_{20}	87,147
L_{21}	N_{21}	104,890
L_{22}	N_{22}	105,052
L_{23}	N_{23}	103,321
L_{24}	N_{24}	105,063

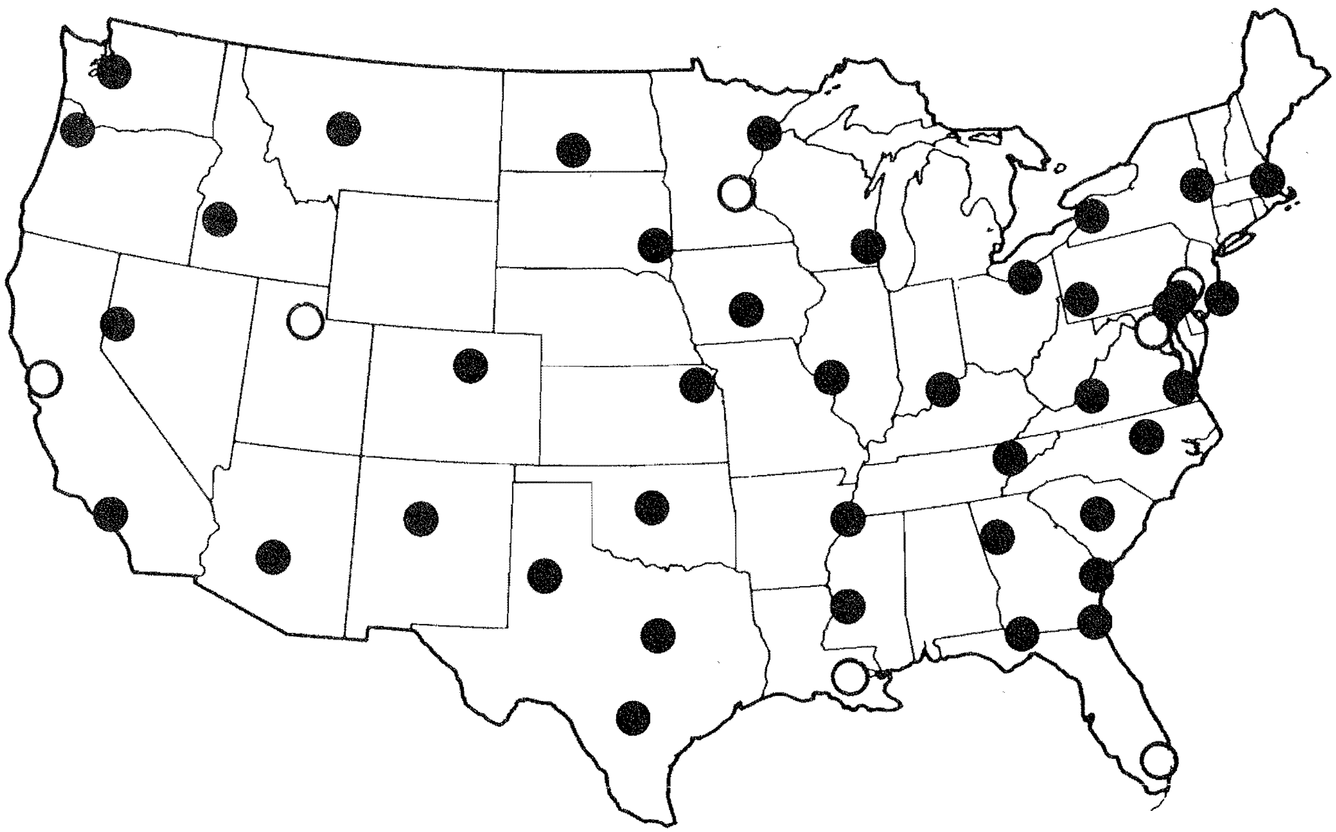


Figure 2-1.--Locations selected to provide data for GEM. An open circle denotes verification stations (7), and a filled-in circle denotes stations comprising the total (41) dependent sample stations.

Weather station Symbol	Sample size Notation	Sample size Actual
L_{25}	N_{25}	98,865
L_{26}	N_{26}	105,001
L_{27}	N_{27}	84,070
L_{28}	N_{28}	104,056
L_{29}	N_{29}	102,307
L_{30}	N_{30}	103,156
L_{31}	N_{31}	103,602
L_{32}	N_{32}	101,962
L_{33}	N_{33}	86,467
L_{34}	N_{34}	102,016
L_{35}	N_{35}	86,251
L_{36}	N_{36}	104,450
L_{37}	N_{37}	104,919
L_{38}	N_{38}	103,908
L_{39}	N_{39}	87,118
L_{40}	N_{40}	102,564
L_{41}	N_{41}	85,612
TOTAL	N	3,964,513

Transformations

Step 5 Transform the original predictors to zero-one variables (dummies). Leave out one from each original predictor because of redundancy.

Figure 2-2 shows a computer printout of the criterion used to dummy each predictor and predictand variable.

INDEX	POSITION IN EXPAND	POSITION IN COLLAP	WITHIN GROUP	RE-ARRANGE	DESCRIPTION		
2	2	2	1	159	MONTH (LOCAL)	JANUARY	
3	3	3	2	160	MONTH (LOCAL)	FEBRUARY	
4	4	4	3	161	MONTH (LOCAL)	MARCH	
5	5	5	4	162	MONTH (LOCAL)	APRIL	
6	6	6	5	163	MONTH (LOCAL)	MAY	
7	7	7	6	164	MONTH (LOCAL)	JUNE	
8	8	8	7	165	MONTH (LOCAL)	JULY	
9	9	9	8	166	MONTH (LOCAL)	AUGUST	
10	10	10	9	167	MONTH (LOCAL)	SEPTEMBER	
11	11	11	10	168	MONTH (LOCAL)	OCTOBER	
12	12	12	11	169	MONTH (LOCAL)	NOVEMBER	
13	13		12		MONTH (LOCAL)	DECEMBER	<-- LEFT-OUT
14	14	13	1	170	HOUR (LOCAL)	0	
15	15	14	2	171	HOUR (LOCAL)	1	
16	16	15	3	172	HOUR (LOCAL)	2	
17	17	16	4	173	HOUR (LOCAL)	3	
18	18	17	5	174	HOUR (LOCAL)	4	
19	19	18	6	175	HOUR (LOCAL)	5	
20	20	19	7	176	HOUR (LOCAL)	6	
21	21	20	8	177	HOUR (LOCAL)	7	
22	22	21	9	178	HOUR (LOCAL)	8	
23	23	22	10	179	HOUR (LOCAL)	9	
24	24	23	11	180	HOUR (LOCAL)	10	
25	25	24	12	181	HOUR (LOCAL)	11	
26	26	25	13	182	HOUR (LOCAL)	12	
27	27	26	14	183	HOUR (LOCAL)	13	
28	28	27	15	184	HOUR (LOCAL)	14	
29	29	28	16	185	HOUR (LOCAL)	15	
30	30	29	17	186	HOUR (LOCAL)	16	
31	31	30	18	187	HOUR (LOCAL)	17	
32	32	31	19	188	HOUR (LOCAL)	18	
33	33	32	20	189	HOUR (LOCAL)	19	
34	34	33	21	190	HOUR (LOCAL)	20	
35	35	34	22	191	HOUR (LOCAL)	21	
36	36	35	23	192	HOUR (LOCAL)	22	
37	37		24		HOUR (LOCAL)	23	<-- LEFT-OUT
38	38	36	1	94	SLP (MB)	800.0 TO 985	
39	39	37	2	95	SLP (MB)	985.1 TO 990	
40	40	38	3	96	SLP (MB)	990.1 TO 995	
41	41	39	4	97	SLP (MB)	995.1 TO 1000	
42	42	40	5	98	SLP (MB)	1000.1 TO 1005	
43	43	41	6	99	SLP (MB)	1005.1 TO 1010	
44	44	42	7	100	SLP (MB)	1010.1 TO 1015	
45	46	43	8	101	SLP (MB)	1020.1 TO 1025	
46	47	44	9	102	SLP (MB)	1025.1 TO 1030	
47	48	45	10	103	SLP (MB)	1030.1 TO 1035	
48	49	46	11	104	SLP (MB)	1035.1 TO 1040	
49	50	47	12	105	SLP (MB)	1040.1 TO 1090	
50	45		8		SLP (MB)	1015.1 TO 1020	<-- LEFT-OUT

Figure 2-2.--Criterion for specifying each dummy predictor and predictand. The first five columns represent indexes for referencing various matrix rows and columns on microfiche.

INDEX	POSITION IN EXPAND	COLLAP	WITHIN GROUP	RE-ARRANGE	DESCRIPTION	
51	51	48	1	2	DB TEMP (F) -140 TO -31	
52	52	49	2	3	DB TEMP (F) -30 TO -26	
53	53	50	3	4	DB TEMP (F) -25 TO -21	
54	54	51	4	5	DB TEMP (F) -20 TO -16	
55	55	52	5	6	DB TEMP (F) -15 TO -11	
56	56	53	6	7	DB TEMP (F) -10 TO -6	
57	57	54	7	8	DB TEMP (F) -5 TO -1	
58	58	55	8	9	DB TEMP (F) 0 TO 4	
59	59	56	9	10	DB TEMP (F) 5 TO 9	
60	60	57	10	11	DB TEMP (F) 10 TO 14	
61	61	58	11	12	DB TEMP (F) 15 TO 19	
62	62	59	12	13	DB TEMP (F) 20 TO 24	
63	63	60	13	14	DB TEMP (F) 25 TO 29	
64	64	61	14	15	DB TEMP (F) 30 TO 34	
65	65	62	15	16	DB TEMP (F) 35 TO 39	
66	66	63	16	17	DB TEMP (F) 40 TO 44	
67	67	64	17	18	DB TEMP (F) 45 TO 49	
68	68	65	18	19	DB TEMP (F) 50 TO 54	
69	69	66	19	20	DB TEMP (F) 55 TO 59	
70	71	67	20	21	DB TEMP (F) 65 TO 69	
71	72	68	21	22	DB TEMP (F) 70 TO 74	
72	73	69	22	23	DB TEMP (F) 75 TO 79	
73	74	70	23	24	DB TEMP (F) 80 TO 84	
74	75	71	24	25	DB TEMP (F) 85 TO 89	
75	76	72	25	26	DB TEMP (F) 90 TO 94	
76	77	73	26	27	DB TEMP (F) 95 TO 99	
77	78	74	27	28	DB TEMP (F) 100 TO 104	
78	79	75	28	29	DB TEMP (F) 105 TO 109	
79	80	76	29	30	DB TEMP (F) 110 TO 140	
80	70		20		DB TEMP (F) 60 TO 64	<-- LEFT-OUT
81	81	77	1	31	DPT DPR (F) 0	
82	82	78	2	32	DPT DPR (F) 1	
83	84	79	3	33	DPT DPR (F) 5 TO 7	
84	85	80	4	34	DPT DPR (F) 8 TO 11	
85	86	81	5	35	DPT DPR (F) 12 TO 15	
86	87	82	6	36	DPT DPR (F) 16 TO 19	
87	88	83	7	37	DPT DPR (F) 20 TO 25	
88	89	84	8	38	DPT DPR (F) 26 TO 35	
89	90	85	9	39	DPT DPR (F) 36 TO 50	
90	91	86	10	40	DPT DPR (F) 51 TO 99	
91	83		3		DPT DPR (F) 2 TO 4	<-- LEFT-OUT
92	92	87	1	106	CLD CVR #1 CLR	
93	94	88	2	107	CLD CVR #1 BKN	
94	95	89	3	108	CLD CVR #1 OVC	
95	96	90	4	109	CLD CVR #1 TOT OBSC	
96	93		2		CLD CVR #1 SCD	<-- LEFT-OUT
97	97	91	1	41	VSBY (ST MI) .00 TO .49	
98	98	92	2	42	VSBY (ST MI) .50 TO .74	
99	99	93	3	43	VSBY (ST MI) .75 TO .99	
100	100	94	4	44	VSBY (ST MI) 1.0 TO 1.49	
101	101	95	5	45	VSBY (ST MI) 1.5 TO 1.99	
102	102	96	6	46	VSBY (ST MI) 2.0 TO 2.49	
103	103	97	7	47	VSBY (ST MI) 2.5 TO 2.99	
104	104	98	8	48	VSBY (ST MI) 3.0 TO 3.99	
105	105	99	9	49	VSBY (ST MI) 4.0 TO 4.99	
106	106	100	10	50	VSBY (ST MI) 5.0 TO 5.99	
107	107	101	11	51	VSBY (ST MI) 6.0 TO 6.99	
108	108		12		VSBY (ST MI) 7.0 TO 100.	<-- LEFT-OUT
109	109	102	1	193	NO WX	
110	110		2		WX	<-- LEFT-OUT
111	112	103	1	52	F,IF	
112	111		1		NO F,F	<-- LEFT-OUT

Figure 2-2.--(continued)

INDEX	POSITION IN EXPAND	POSITION IN COLLAP	WITHIN GROUP	RE-ARRANGE	DESCRIPTION	
113	114	104	1	53	GF	
114	113		1		NO GF	<-- LEFT-OUT
115	116	105	1	54	K,H,D,KH,KD,HD,KHD	
116	115		1		NO K,H,D,KH,KD,HD,KHD	<-- LEFT-OUT
117	118	106	1	55	BS,BD,BN	
118	117		1		NO BS,BD,BN	<-- LEFT-OUT
119	120	107	1	56	L-	
120	121	108	2	57	L,L+	
121	119		1		NO L	<-- LEFT-OUT
122	123	109	1	58	R-	
123	124	110	2	59	R	
124	125	111	3	60	R+	
125	122		1		NO R	<-- LEFT-OUT
126	127	112	1	61	RW-	
127	128	113	2	62	RW	
128	129	114	3	63	RW+	
129	126		1		NO RW	<-- LEFT-OUT
130	131	115	1	64	S-,IC-	
131	132	116	2	65	S,IC	
132	133	117	3	66	S+,IC+	
133	130		1		NO S,IC	<-- LEFT-OUT
134	135	118	1	67	SW-,IP-	
135	136	119	2	68	SW,IP	
136	137	120	3	69	SW+,IP+	
137	134		1		NO SW,IP	<-- LEFT-OUT
138	139	121	1	70	ZL-,ZL,ZL+	
139	138		1		NO ZL	<-- LEFT-OUT
140	141	122	1	71	ZR-,ZR,ZR+	
141	140		1		NO ZR	<-- LEFT-OUT
142	143	123	1	72	TSTM,A	
143	142		1		NO TSTM,A	<-- LEFT-OUT
144	145	124	1	73	TSTM+	
145	144		1		NO TSTM+	<-- LEFT-OUT
146	146	125	1	110	CLD HGT #1 0 TO 1	
147	147	126	2	111	CLD HGT #1 2 TO 4	
148	148	127	3	112	CLD HGT #1 5 TO 6	
149	149	128	4	113	CLD HGT #1 7 TO 9	
150	150	129	5	114	CLD HGT #1 10 TO 14	
151	151	130	6	115	CLD HGT #1 15 TO 19	
152	152	131	7	116	CLD HGT #1 20 TO 24	
153	153	132	8	117	CLD HGT #1 25 TO 29	
154	154	133	9	118	CLD HGT #1 30 TO 39	
155	155	134	10	119	CLD HGT #1 40 TO 49	
156	156	135	11	120	CLD HGT #1 50 TO 59	
157	157	136	12	121	CLD HGT #1 60 TO 75	
158	158	137	13	122	CLD HGT #1 76 TO 99	
159	159	138	14	123	CLD HGT #1 100 TO 150	
160	161	139	15	124	CLD HGT #1 PART OBSC	
161	160		15		CLD HGT #1 151 TO UNL	<-- LEFT-OUT

Figure 2-2.--(continued)

INDEX	POSITION IN EXPAND	COLLAP	WITHIN GROUP	RE-ARRANGE	DESCRIPTION	
162	163	140	1	125	CLD CVR #2 SCT	
163	164	141	2	126	CLD CVR #2 BKN	
164	165	142	3	127	CLD CVR #2 OVC	
165	162		1		CLD CVR #2 CLR	<-- LEFT-OUT
166	166	143	1	128	CLD HGT #2 0 TO 1	
167	167	144	2	129	CLD HGT #2 2 TO 4	
168	168	145	3	130	CLD HGT #2 5 TO 6	
169	169	146	4	131	CLD HGT #2 7 TO 9	
170	170	147	5	132	CLD HGT #2 10 TO 14	
171	171	148	6	133	CLD HGT #2 15 TO 19	
172	172	149	7	134	CLD HGT #2 20 TO 24	
173	173	150	8	135	CLD HGT #2 25 TO 29	
174	174	151	9	136	CLD HGT #2 30 TO 39	
175	175	152	10	137	CLD HGT #2 40 TO 49	
176	176	153	11	138	CLD HGT #2 50 TO 59	
177	177	154	12	139	CLD HGT #2 60 TO 75	
178	178	155	13	140	CLD HGT #2 76 TO 99	
179	179	156	14	141	CLD HGT #2 100 TO 150	
180	180		15		CLD HGT #2 151 TO UNL	<-- LEFT-OUT
181	181	157	1	142	TOTAL CLD CVR CLR	
182	182	158	2	143	TOTAL CLD CVR SCT	
183	183	159	3	144	TOTAL CLD CVR BKN	
184	184		4		TOTAL CLD CVR OVC	<-- LEFT-OUT
185	185	160	1	145	CEILING 0 TO 1	
186	186	161	2	146	CEILING 2 TO 4	
187	187	162	3	147	CEILING 5 TO 6	
188	188	163	4	148	CEILING 7 TO 9	
189	189	164	5	149	CEILING 10 TO 14	
190	190	165	6	150	CEILING 15 TO 19	
191	191	166	7	151	CEILING 20 TO 24	
192	192	167	8	152	CEILING 25 TO 29	
193	193	168	9	153	CEILING 30 TO 39	
194	194	169	10	154	CEILING 40 TO 49	
195	195	170	11	155	CEILING 50 TO 59	
196	196	171	12	156	CEILING 60 TO 75	
197	197	172	13	157	CEILING 76 TO 99	
198	198	173	14	158	CEILING 100 TO 150	
199	199		15		CEILING 151 TO UNL	<-- LEFT-OUT
200	200	174	1	74	WIND CALM / LT 2	
201	201	175	2	75	WIND NNE TO NE / LE 9	
202	202	176	3	76	WIND NNE TO NE / 10 TO 19	
203	203	177	4	77	WIND ENE TO E / LE 9	
204	204	178	5	78	WIND ENE TO E / 10 TO 19	
205	205	179	6	79	WIND ESE TO SE / LE 9	
206	206	180	7	80	WIND ESE TO SE / 10 TO 19	
207	208	181	8	81	WIND SSE TO S / 10 TO 19	
208	209	182	9	82	WIND SSW TO SW / LE 9	
209	210	183	10	83	WIND SSW TO SW / 10 TO 19	
210	211	184	11	84	WIND WSW TO W / LE 9	
211	212	185	12	85	WIND WSW TO W / 10 TO 19	
212	213	186	13	86	WIND WNW TO NW / LE 9	
213	214	187	14	87	WIND WNW TO NW / 10 TO 19	
214	215	188	15	88	WIND NNW TO N / LE 9	
215	216	189	16	89	WIND NNW TO N / 10 TO 19	
216	217	190	17	90	WIND NNE TO E / GE 20	
217	218	191	18	91	WIND ESE TO S / GE 20	
218	219	192	19	92	WIND SSW TO W / GE 20	
219	220	193	20	93	WIND NNW TO N / GE 20	
220	207		8		WIND SSE TO S / LE 9	<-- LEFT-OUT
221	222	194	1	194	AUTWTR / DAY7-18 TRUE	
222	221		1		AUTWTR / DAY7-18 FALSE	<-- LEFT-OUT

Figure 2-2.--(continued)

INDEX	POSITION IN EXPAND	POSITION IN COLLAP	WITHIN GROUP	RE-ARRANGE	DESCRIPTION	
223	224	195	1	195	AUTWTR / HUMID TRUE	
224	223		1		AUTWTR / HUMID FALSE	<-- LEFT-OUT
225	226	196	1	196	AUTWTR / STHWIND TRUE	
226	225		1		AUTWTR / STHWIND FALSE	<-- LEFT-OUT
227	228	197	1	197	AUTWTR / ESTWIND TRUE	
228	227		1		AUTWTR / ESTWIND FALSE	<-- LEFT-OUT
229	230	198	1	198	AUTWTR / OVCSKY TRUE	
230	229		1		AUTWTR / OVCSKY FALSE	<-- LEFT-OUT
231	232	199	1	199	AUTWTR / HISKY TRUE	
232	231		1		AUTWTR / HISKY FALSE	<-- LEFT-OUT
233	234	200	1	200	AUTWTR / FARVSBY TRUE	
234	233		1		AUTWTR / FARVSBY FALSE	<-- LEFT-OUT
235	236	201	1	201	AUTWTR / NO PRECIP TRUE	
236	235		1		AUTWTR / NO PRECIP FALSE	<-- LEFT-OUT
237	238	202	1	202	DAY7-18 / HUMID TRUE	
238	237		1		DAY7-18 / HUMID FALSE	<-- LEFT-OUT
239	240	203	1	203	DAY7-18 / STHWIND TRUE	
240	239		1		DAY7-18 / STHWIND FALSE	<-- LEFT-OUT
241	242	204	1	204	DAY7-18 / ESTWIND TRUE	
242	241		1		DAY7-18 / ESTWIND FALSE	<-- LEFT-OUT
243	244	205	1	205	DAY7-18 / OVCSKY TRUE	
244	243		1		DAY7-18 / OVCSKY FALSE	<-- LEFT-OUT
245	246	206	1	206	DAY7-18 / HISKY TRUE	
246	245		1		DAY7-18 / HISKY FALSE	<-- LEFT-OUT
247	248	207	1	207	DAY7-18 / FARVSBY TRUE	
248	247		1		DAY7-18 / FARVSBY FALSE	<-- LEFT-OUT
249	250	208	1	208	DAY7-18 / NO PRECIP TRUE	
250	249		1		DAY7-18 / NO PRECIP FALSE	<-- LEFT-OUT
251	252	209	1	209	HUMID / STHWIND TRUE	
252	251		1		HUMID / STHWIND FALSE	<-- LEFT-OUT
253	254	210	1	210	HUMID / ESTWIND TRUE	
254	253		1		HUMID / ESTWIND FALSE	<-- LEFT-OUT
255	256	211	1	211	HUMID / OVCSKY TRUE	
256	255		1		HUMID / OVCSKY FALSE	<-- LEFT-OUT
257	258	212	1	212	HUMID / HISKY TRUE	
258	257		1		HUMID / HISKY FALSE	<-- LEFT-OUT

Figure 2-2.--(continued)

INDEX	POSITION IN EXPAND	COLLAP	WITHIN GROUP	RE-ARRANGE	DESCRIPTION	
259	260	213	1	213	HUMID / FARVSBY TRUE	
260	259		1		HUMID / FARVSBY FALSE	<-- LEFT-OUT
261	262	214	1	214	HUMID / NO PRECIP TRUE	
262	261		1		HUMID / NO PRECIP FALSE	<-- LEFT-OUT
263	264	215	1	215	STHWIND / ESTWIND TRUE	
264	263		1		STHWIND / ESTWIND FALSE	<-- LEFT-OUT
265	266	216	1	216	STHWIND / OVCSKY TRUE	
266	265		1		STHWIND / OVCSKY FALSE	<-- LEFT-OUT
267	268	217	1	217	STHWIND / HISKY TRUE	
268	267		1		STHWIND / HISKY FALSE	<-- LEFT-OUT
269	270	218	1	218	STHWIND / FARVSBY TRUE	
270	269		1		STHWIND / FARVSBY FALSE	<-- LEFT-OUT
271	272	219	1	219	STHWIND / NO PRECIP TRUE	
272	271		1		STHWIND / NO PRECIP FALSE	<-- LEFT-OUT
273	274	220	1	220	ESTWIND / OVCSKY TRUE	
274	273		1		ESTWIND / OVCSKY FALSE	<-- LEFT-OUT
275	276	221	1	221	ESTWIND / HISKY TRUE	
276	275		1		ESTWIND / HISKY FALSE	<-- LEFT-OUT
277	278	222	1	222	ESTWIND / FARVSBY TRUE	
278	277		1		ESTWIND / FARVSBY FALSE	<-- LEFT-OUT
279	280	223	1	223	ESTWIND / NO PRECIP TRUE	
280	279		1		ESTWIND / NO PRECIP FALSE	<-- LEFT-OUT
281	282	224	1	224	OVCSKY / HISKY TRUE	
282	281		1		OVCSKY / HISKY FALSE	<-- LEFT-OUT
283	284	225	1	225	OVCSKY / FARVSBY TRUE	
284	283		1		OVCSKY / FARVSBY FALSE	<-- LEFT-OUT
285	286	226	1	226	OVCSKY / NO PRECIP TRUE	
286	285		1		OVCSKY / NO PRECIP FALSE	<-- LEFT-OUT
287	288	227	1	227	HISKY / FARVSBY TRUE	
288	287		1		HISKY / FARVSBY FALSE	<-- LEFT-OUT
289	290	228	1	228	HISKY / NO PRECIP TRUE	
290	289		1		HISKY / NO PRECIP FALSE	<-- LEFT-OUT

Figure 2-2.--(concluded)

Statistical Analyses

Step 6 Compute the $\underline{Z}'\underline{Z}$ and $\underline{Y}'\underline{Z}$ matrices.

Step 7 Solve for $\underline{A}$ from $\underline{A} = (\underline{Z}'\underline{Z})^{-1} (\underline{Y}'\underline{Z})$.

Step 8 Construct PLODITE (Putting Left Out Dummy In The Equation) matrix $\underline{B}$ by adding in left-out coefficients and left-out equations.

Step 9 Solve for μ_0's and μ_1's. (For details, see appendix.)

Step 10 Solve for R^2's where $R^2 = \mu_1 - \mu_0$.

Step 11 Solve for threshold probabilities P*. (For details, see appendix.)

The method selected to describe the steps that were performed in the statistical analyses will be by way of deriving the quantities actually obtained for a particular predictand, NO WX/WX at a 1-hr projection. An entire display of these quantities for all 289 predictands for a 1-hr projection is contained on microfiche given in the pocket inside the back cover of this report.

Derivation of the two crossproduct matrices $\underline{Z}'\underline{Z}$ and $\underline{Y}'\underline{Z}$, in step 6, was accomplished, as was mentioned previously, by packing the zero-one observations in $\underline{Z}$ and $\underline{Y}$ and obtaining the products by logical "anding" two computer words together and counting the number of resulting bits. This gives a two-order-of-magnitude improvement in computing efficiency over ordinary floating-point multiplication, since it treats simultaneously as many observations as can fit into a computer word. These two matrices are on microfiches A and B, respectively.

For the labeled predictors in table 2-1, column 1 gives the sum row of the $\underline{Z}'\underline{Z}$ matrix and column 2 the NO WX/WX row of the $\underline{Y}'\underline{Z}$ matrix. This gives the products between the Y variable for NO WX/WX times each of the 290 predictors over the sample N.

Solving for the regression coefficient matrix $\underline{A}$ in step 7 was performed using the Crout method (Crout, 1941). This method does not require solving for the inverse matrix, $(\underline{Z}'\underline{Z})^{-1}$, but instead accomplishes deriving the regression coefficients by a forward and then a backward solution, avoiding many of the computational instabilities encountered by inverting large matrices. This matrix solution yields a 228 x 228 matrix--228 predictor coefficients for each of 228 predictands. In step 8 this matrix is expanded to include the otherwise redundant left-out dummy variables by simple arithmetic to a 290 x 290 PLODITE matrix called $\underline{B}$. Both $\underline{A}$ and $\underline{B}$ are on microfiches C and D, respectively.

The NO WX/WX equations for the $\underline{A}$ and $\underline{B}$ matrices appear as columns 3 and 4, respectively, in table 2-1. One further variation is presented in column 5 of this table, namely, the BETA coefficient form of the PLODITE equation in column 4. That is,

$$\beta_{iy} = B_{iy} \frac{\sigma_y}{\sigma_{Z_i}} \quad (i=1,2,...,290) \qquad (2\text{-}7)$$

where σ_Y and σ_{Z_i} are the standard deviations of Y and the predictor Z_i,

Table 2-1.--A display of quantities derived for GEM for the predictand Y = NO WX/WX at 1 hour. Included in the six columns are: 1) sum of Z's, 2) sum of cross-products Y and Z's, 3) generalized operator equation, A, 4) PLODITE generalized operator equation, B, 5) PLODITE beta coefficients, β , and 6) anomaly generalized operator equation, A_a. No entries indicate left out elements as described in the text.

Predictor Z			1	2	3	4	5	6
Number	Element	Category	ΣZ	ΣYZ	A	B	β	A_a
1	Always	Unity	3964513	3163668	.38544	.79800		.00000
2	Month	Jan	338217	244842	-.00137	-.00778	-.00541	-.00183
3		Feb	307968	225026	-.00057	-.00698	-.00465	-.00012
4		Mar	337739	260983	.00123	-.00518	-.00360	.00381
5		Apr	326031	268881	.01326	.00685	.00469	.01899
6		May	334902	281645	.01358	.00717	.00497	.02314
7		June	322102	270724	.01327	.00686	.00467	.02633
8		July	334584	281778	.00991	.00350	.00242	.02507
9		Aug	334753	277415	.00894	.00253	.00175	.02347
10		Sept	325820	270242	.01349	.00708	.00484	.02515
11		Oct	337465	274164	.00302	-.00338	-.00235	.01095
12		Nov	326774	256558	.00234	-.00407	-.00279	.00534
13		Dec	338158	251410		-.00641	-.00446	
14	Hour (LST)	00	166568	134684	-.00055	-.03368	-.01683	-.00058
15		01	166726	132855	-.00083	-.03396	-.01698	-.00120
16		02	166735	130876	.00020	-.03292	-.01646	-.00047
17		03	166689	128316	-.00213	-.03526	-.01762	-.00309
18		04	166317	123926	-.01213	-.04526	-.02260	-.01341
19		05	165737	118783	-.01728	-.05040	-.02513	-.01913
20		06	165186	116016	-.00718	-.04031	-.02006	-.00968
21		07	164787	117551	.06922	.03610	.01795	.06360
22		08	164506	122028	.08391	.05078	.02522	.07857
23		09	164340	127147	.08712	.05399	.02681	.08277
24		10	164174	131377	.08334	.05021	.02492	.08034
25		11	164109	134157	.07648	.04336	.02151	.07477
26		12	164148	136042	.07171	.03858	.01914	.07111
27		13	164137	137009	.06750	.03438	.01706	.06787
28		14	164144	137407	.06365	.03053	.01515	.06470
29		15	164144	137286	.05991	.02679	.01329	.06133
30		16	164149	137104	.05899	.02587	.01284	.06035
31		17	164109	136898	.06014	.02701	.01340	.06100
32		18	164250	136787	.06093	.02781	.01380	.06100
33		19	164867	137466	-.00219	-.03531	-.01756	.00043
34		20	165625	138116	-.00044	-.03357	-.01673	.00135
35		21	166239	138225	-.00009	-.03322	-.01658	.00112
36		22	166419	137428	-.00033	-.03346	-.01671	.00041
37		23	166408	136184		-.03313	-.01655	
38	SLP (MB)	800.0-985.0	1033	401	-.04081	-.03965	-.00159	-.03425
39		985.1-990.0	3330	1453	-.05600	-.05484	-.00396	-.05146
40		990.1-995.0	12091	6404	-.03520	-.03403	-.00467	-.03150
41		995.1-1000.0	40561	25369	-.02730	-.02613	-.00655	-.02566
42		1000.1-1005.0	131828	94263	-.01723	-.01607	-.00717	-.01735
43		1005.1-1010.0	417276	326015	-.00966	-.00850	-.00650	-.01093
44		1010.1-1015.0	977206	776339	-.00481	-.00365	-.00392	-.00561

Table 2-1.--(continued)

Number	Element	Category	1 ΣZ	2 ΣYZ	3 A	4 B	5 β	6 A_a
45	SLP (MB)	1015.1-1020.0	1215826	977442		.00116	.00134	
46		1020.1-1025.0	698126	565405	.00447	.00563	.00535	.00533
47		1025.1-1030.0	320069	265407	.00868	.00984	.00668	.00891
48		1030.1-1035.1	111202	94859	.01415	.01532	.00630	.01384
49		1035.1-1040.0	29005	25259	.02001	.02117	.00449	.02017
50		1040.1-1090.0	5960	5052	.01762	.01878	.00181	.01726
51	DBT (°F)	-140 - -31	58	49	-.05010	-.04796	-.00046	-.04502
52		- 30 - -26	200	150	-.12243	-.12029	-.00213	-.11731
53		- 25 - -21	605	486	-.06280	-.06067	-.00187	-.05735
54		- 20 - -16	1554	1239	-.06883	-.06670	-.00329	-.06419
55		- 15 - -11	3593	2965	-.04363	-.04150	-.00311	-.03878
56		- 10 - - 6	6389	5045	-.03865	-.03651	-.00365	-.03379
57		- 5 - - 1	10824	8495	-.03496	-.03282	-.00427	-.02927
58		0 - 4	16616	12689	-.03409	-.03195	-.00514	-.02743
59		5 - 9	24249	17893	-.03454	-.03240	-.00629	-.02563
60		10 - 14	38117	27921	-.02734	-.02520	-.00612	-.01709
61		15 - 19	58450	42606	-.02226	-.02012	-.00604	-.01206
62		20 - 24	95590	69632	-.01301	-.01087	-.00415	-.00347
63		25 - 29	150006	111102	-.00299	-.00085	-.00040	.00626
64		30 - 34	228311	163105	.00302	.00516	.00300	.01129
65		35 - 39	260201	197412	.00017	.00231	.00142	.00695
66		40 - 44	287560	220306	-.00183	.00031	.00020	.00309
67		45 - 59	299105	231217	-.00288	-.00074	-.00048	.00055
68		50 - 54	320497	248365	-.00262	-.00048	-.00033	-.00019
69		55 - 59	339288	264102	.00010	.00224	.00156	.00166
70		60 - 64	357114	279059		.00214	.00153	
71		65 - 69	364476	288963	-.00046	.00168	.00121	-.00276
72		70 - 74	369781	303117	.00442	.00656	.00475	-.00145
73		75 - 79	296915	263132	.00783	.00997	.00654	-.00189
74		80 - 84	204536	187816	-.00559	-.00346	-.00190	-.01801
75		85 - 89	132182	122965	-.01461	-.01247	-.00558	-.03041
76		90 - 94	68166	64526	-.01683	-.01469	-.00476	-.03610
77		95 - 99	22412	21726	-.00812	-.00598	-.00112	-.03087
78		100 - 104	5883	5774	-.00265	-.00051	-.00005	-.02736
79		105 - 109	1608	1586	-.00320	-.00107	-.00005	-.03116
80		110 - 140	227	225	.00127	.00341	.00006	-.02842
81	DPD (°F)	0	109186	17940	-.02448	-.01174	-.00478	-.02942
82		1	174045	58378	-.03100	-.01825	-.00931	-.03401
83		2 - 4	701496	417802		.01274	.01211	
84		5 - 7	607722	474044	-.03113	-.01838	-.01650	-.02437
85		8 - 11	634664	548258	-.01887	-.00612	-.00559	-.01182
86		12 - 15	479162	437078	-.01252	.00022	.00018	-.00566
87		16 - 19	363171	341999	-.00782	.00492	.00354	-.00148
88		20 - 25	373899	359623	-.00510	.00765	.00557	.00022
89		26 - 35	323068	315346	-.00217	.01058	.00721	.00124
90		36 - 50	156091	152289	-.00207	.01067	.00517	-.00070
91		51 - 99	42009	40911	.00494	.01768	.00451	.00742
92	CC #1	CLR	1120221	1047709	.00534	.00165	.00186	.00829
93		SCD	1433874	1182437		-.00369	-.00442	

Table 2-1.--(continued)

	Predictor Z		1	2	3	4	5	6
Number	Element	Category	ΣZ	ΣYZ	A	B	β	A_a
94	CC #1	BKN	723024	551984	.00969	.00600	.00577	.00953
95		OVC	615688	379718	.00009	-.00360	-.00325	.00111
96		TOT OBSC	71706	1820	.02204	.01835	.00609	.02025
97	VIS (M)	.00 - .49	38648	764	-.38011	-.33544	-.08209	-.35506
98		.50 - .74	16166	485	-.36691	-.32224	-.05115	-.34335
99		.75 - .99	15970	409	-.36683	-.32216	-.05082	-.34190
100		1.00 - 1.49	36608	1162	-.35939	-.31472	-.07498	-.33702
101		1.50 - 1.99	32702	1023	-.36081	-.31614	-.07122	-.33781
102		2.00 - 2.49	52298	2355	-.35117	-.30650	-.08710	-.33174
103		2.50 - 2.99	26827	1092	-.36235	-.31768	-.06487	-.34149
104		3.00 - 3.99	84881	6176	-.32985	-.28518	-.10281	-.31292
105		4.00 - 4.99	100832	10398	-.30306	-.25839	-.10132	-.28658
106		5.00 - 5.99	117557	20596	-.23455	-.18988	-.08022	-.22051
107		6.00 - 6.99	99064	29629	-.11556	-.07089	-.02756	-.10211
108		7.00 -100.00	3342960	3089579		.04467	.04045	
109	WEATHER	NO WX	3164088	3022632	.46116	.09311	.09309	.46196
110		WX	800425	141036		-.36805	-.36798	
111	FOG	NO FOG	3715675	3145063		-.00045	-.00027	
112		FOG	248838	18605	.00716	.00671	.00405	.00861
113	GROUND FOG	NO GF	3894272	3154942		.00013	.00004	
114		GF	70241	8726	-.00747	-.00733	-.00241	-.00777
115	HAZE, SMOKE	NO H, K	3707903	3126531		.00532	.00326	
116		H, K	256610	37137	-.08212	-.07681	-.04707	-.07152
117	BLOWING	NO B	3953950	3162175		-.00011	-.00001	
118		B	10563	1493	.04212	.04200	.00539	.03641
119	DRIZZLE	NO L	3921226	3158802		-.00073	-.00019	
120		L-	42654	4842	.06678	.06605	.01697	.06445
121		L, L+	633	24	.04754	.04681	.00147	.04231
122	RAIN	NO R	3816374	3140084		-.00025	-.00012	
123		R-	139674	23170	.00623	.00597	.00274	.00725
124		R	7365	361	.00977	.00952	.00102	.01110
125		R+	1100	53	.05508	.05483	.00227	.05805
126	RAIN SHOWERS	NO RW	3865835	3126202		-.00325	-.00126	
127		RW-	90735	35887	.13083	.12758	.04752	.13301
128		RW	5343	1062	.11266	.10941	.01000	.11061
129		RW+	2600	517	.16532	.16207	.01033	.16125
130	SNOW	NO S	3887264	3154652		.00007	.00002	
131		S-	73929	8915	-.00414	-.00407	-.00137	-.00588
132		S	2812	96	.00374	.00381	.00025	.00577
133		S+	508	5	.02768	.02775	.00078	.03343
134	SNOW SHOWERS	NO SW	3928234	3155246		-.00012	-.00003	
135		SW-	35777	8343	.01166	.01154	.00272	.01941
136	SNOW SHOWERS	SW	422	65	.09462	.09450	.00243	.09749
137		SW+	80	14	.11854	.11842	.00132	.12639
138	FREEZING DRIZZLE	NO ZL	3960295	3163455		-.00002	-.00000	
139		ZL-, ZL, ZL+	4218	213	.02176	.02173	.00176	.01551
140	FREEZING RAIN	NO ZR	3961427	3163426		.00002	.00000	
141		ZR-, ZR, ZR+	3086	242	-.01939	-.01938	-.00135	-.02168
142	THUNDERSTORM,A	NO TSM, A	3934524	3154044		.00032	.00007	
143		TSM, A	29989	9624	-.04187	-.04155	-.00897	-.04706

Table 2-1.--(continued)

Predictor Z Number	Element	Category	1 ΣZ	2 ΣYZ	3 A	4 B	5 β	6 A_a
144	THUNDERSTORM+	NO TSM+	3964343	3163621		-.00000	-.00000	
145		TSM+	170	47	.02605	.02605	.00042	.02226
146	CH #1 (00')	0 - 1	30238	727	-.00973	-.00755	-.00164	-.01678
147		2 - 4	99175	12704	-.00798	-.00580	-.00226	-.01337
148		5 - 6	82305	22041	-.00627	-.00409	-.00145	-.00968
149		7 - 9	117536	46459	-.00763	-.00545	-.00230	-.01029
150		10 - 14	167404	91377	-.00618	-.00400	-.00200	-.00834
151		15 - 19	137539	89861	-.00207	.00011	.00005	-.00332
152		20 - 24	126142	90903	-.00302	-.00084	-.00037	-.00321
153		25 - 29	124929	95429	-.00623	-.00404	-.00176	-.00564
154		30 - 39	255504	207213	-.00544	-.00326	-.00199	-.00461
155		40 - 49	239335	201578	-.00673	-.00455	-.00270	-.00586
156		50 - 59	179660	154800	-.00464	-.00246	-.00127	-.00443
157		60 - 75	196152	169871	-.00297	-.00079	-.00043	-.00405
158		76 - 99	146333	127535	-.00056	.00162	.00076	-.00034
159		100 - 150	393002	353357	.00467	.00685	.00510	.00399
160		151 - UNL	1606392	1495066		.00218	.00267	
161		PART OBSC	62867	4747	-.01608	-.01390	-.00432	-.02243
162	CC #2	CLR	2767330	2291124		.00010	.00012	
163		SCD	248836	214265	.00137	.00148	.00089	.00089
164		BKN	429316	345235	.00312	.00322	.00249	.00316
165		OVC	519031	313044	-.00403	-.00393	-.00330	-.00284
166	CC #2 (00')	0 - 1	463	16	.00982	.01192	.00032	.00951
167		2 - 4	10179	528	.01813	.02023	.00255	.02062
168		5 - 6	10982	1026	.01058	.01268	.00166	.01482
169		7 - 9	18773	2913	.00493	.00704	.00120	.00930
170	CC #2 (00')	10 - 14	39841	10612	-.00617	-.00407	-.00101	-.00162
171		15 - 19	32803	12422	-.00567	-.00357	-.00081	-.00165
172		20 - 24	33036	14724	-.00885	-.00675	-.00153	-.00549
173		25 - 29	31732	15708	-.01281	-.01071	-.00238	-.00928
174		30 - 39	56921	31344	-.01254	-.01044	-.00309	-.00961
175		40 - 49	51003	30933	-.01675	-.01464	-.00411	-.01455
176		50 - 59	42636	27895	-.01347	-.01137	-.00292	-.01237
177		60 - 75	74059	51567	-.01663	-.01453	-.00490	-.01621
178		76 - 99	82201	59923	-.01650	-.01440	-.00511	-.01546
179		100 - 150	263050	214315	-.01092	-.00882	-.00547	-.01152
180		151 - UNL	3216834	2689742		.00210	.00205	
181	TOTAL CLOUD COVER	CLR	1120039	1047568	.10674	.03642	.04084	.11059
182		SCD	781373	708165	.11068	.04036	.03999	.11982
183		BKN	722434	634205	.10070	.03038	.02921	.10919
184		OVC	1340667	773730		-.07032	-.08286	
185	CEILING (00')	0 - 1	29306	409	-.02309	-.02692	-.00574	-.01136
186		2 - 4	82348	6190	-.02128	-.02511	-.00892	-.00948
187		5 - 6	63621	12297	-.01730	-.02113	-.00661	-.00560
188		7 - 9	91444	30329	-.00459	-.00842	-.00315	.00687
189		10 - 14	124886	58656	.00723	.00340	.00148	.01874
190		15 - 19	97079	55230	.00856	.00473	.00182	.02066
191		20 - 24	84488	51830	.00633	.00250	.00900	.01906
192		25 - 29	81316	52922	.00819	.00435	.00154	.02104

Table 2-1.--(continued)

Number	Predictor Z Element	Category	1 ΣZ	2 ΣYZ	3 A	4 B	5 β	6 A_a
193	CEILING (00')	30 - 39	138928	93757	.00438	.00055	.00025	.01758
194		40 - 49	117208	82160	.00482	.00099	.00042	.01787
195		50 - 59	86289	62269	.00659	.00276	.00100	.01936
196		60 - 75	123976	91868	.01212	.00829	.00360	.02512
197		76 - 99	106856	80800	.01967	.01584	.00639	.03235
198		100 - 150	278263	230392	.03901	.03518	.02238	.05187
199		151 - UNL	2458505	2254559		-.00383	-.00463	
200	WIND	CALM	246054	181207	-.01431	-.01227	-.00737	-.01680
201		NNE-NE < 11	246345	187201	-.01373	-.01168	-.00703	-.01816
202		NNE-NE 11-19	124812	90701	-.01244	-.01040	-.00452	-.01755
203		ENE-NE < 11	236015	177244	-.01772	-.01568	-.00924	-.02210
204		ENE-NE 11-19	97973	68699	-.02232	-.02027	-.00784	-.02664
205		ESE-SE < 11	296348	230062	-.00417	-.00213	-.00139	-.00406
206		ESE-SE 11-19	125249	97250	-.00679	-.00475	-.00207	-.00856
207		SSE-S < 11	333410	266918		.00204	.00141	
208		SSE-S 11-19	235668	199396	.00423	.00627	.00369	.00220
209		SSW-SW < 11	308593	251306	.00207	.00411	.00275	.00286
210		SSW-SW 11-19	221594	187791	.00377	.00581	.00333	.00936
211		WSW-W < 11	274823	223645	.00459	.00663	.00420	.00767
212		WSW-W 11-19	183671	155349	.00807	.01011	.00529	.01737
213		WNW-NW < 11	264684	220670	.00399	.00604	.00375	.00410
214		WNW-NW 11-19	221901	193870	.01304	.01509	.00864	.01554
215		NNW-N < 11	242261	193639	-.00135	.00069	.00041	-.00285
216		NNW-N 11-19	162243	129082	.00427	.00631	.00311	.00319
217		NNE-E > 19	17012	9166	-.03147	-.02942	-.00479	-.03644
218		ESE-S > 19	22770	17875	-.01313	-.01109	-.00209	-.01578
219		SSW-W > 19	52815	43452	-.00237	-.00033	-.00009	.00402
220		WNW-N > 19	50272	39145	.00482	.00686	.00191	.00232
221	AUTWTR/DAY 7-18	F	2976307	2406499		.00141	.00152	
222		T	988206	757169	-.00568	-.00426	-.00459	-.00772
223	AUTWTR/HUMID	F	3423800	2924115		.00095	.00082	
224		T	540713	239553	-.00700	-.00604	-.00517	-.00546
225	AUTWTR/STHWIND	F	3117865	2509058		-.00024	-.00025	
226		T	846648	654610	.00113	.00089	.00091	.00172
227	AUTWTR/ESTWIND	F	3310339	2702304		.00073	.00068	
228		T	654174	461364	-.00445	-.00372	-.00344	-.00286
229	AUTWTR/OVCSKY	F	3173036	2734234		-.00058	-.00058	
230		T	791477	429434	.00290	.00232	.00231	.00343
231	AUTWTR/HISKY	F	2813734	2114116		.00038	.00043	
232		T	1150779	1049552	-.00132	-.00094	-.00106	-.00275
233	AUTWTR/FARVSBY	F	2353778	1689631		.00295	.00360	
234		T	1610735	1474037	-.00725	-.00430	-.00527	-.00497
235	AUTWTR/NO PRECIP	F	2238847	1694084		-.00998	-.01232	
236	AUTWTR/NO PRECIP	T	1725666	1469584	.02293	.01295	.01599	.02022
237	DAY 7-18/HUMID	F	3660849	3048541		-.00092	-.00061	
238		T	303664	115127	.01208	.01115	.00739	.01159
239	DAY 7-18/STHWIND	F	3062803	2429687		-.00083	-.00086	
240		T	901710	733981	.00364	.00281	.00294	.00403
241	DAY 7-18/ESTWIND	F	3279867	2640071		.00049	.00046	

Table 2-1.--(concluded)

	Predictor Z		1	2	3	4	5	6
Number	Element	Category	ΣZ	ΣYZ	A	B	β	A_a
242	DAT 7-18/ESTWIND	T	684646	523597	-.00283	-.00234	-.00220	-.00197
243	DAY 7-18/OVCSKY	F	3290996	2766486		.00218	.00204	
244		T	673517	397182	-.01285	-.01067	-.00998	-.01341
245	DAY 7-18/HISKY	F	2770508	2056488		-.00063	-.00072	
246		T	1194005	1107180	.00209	.00146	.00167	.00554
247	DAY 7-18/FARVSBY	F	2313266	1622299		.02676	.03286	
248		T	1651247	1541369	-.06425	-.03749	-.04603	-.06182
249	DAY 7-18/NO PRECIP	F	2194022	1617489		.00334	.00413	
250		T	1770491	1546179	-.00748	-.00414	-.00512	-.00755
251	HUMID/STHWIND	F	3536739	2932143		-.00174	-.00135	
252		T	427774	231525	.01614	.01440	.01113	.01748
253	HUMID/ESTWIND	F	3581001	2994083		.00068	.00050	
254		T	383512	169585	-.00703	-.00635	-.00468	-.00575
255	HUMID/OVCSKY	F	3389470	2971521		.00318	.00279	
256		T	575043	192147	-.02190	-.01872	-.01642	-.01973
257	HUMID/HISKY	F	3597156	2892308		.00056	.00041	
258		T	367357	271360	-.00606	-.00550	-.00397	-.00511
259	HUMID/FARVSBY	F	3400485	2708249		.00540	.00469	
260		T	564028	455419	-.03792	-.03253	-.02830	-.03478
261	HUMID/NO PRECIP	F	3254549	2711930		.00477	.00456	
262		T	709964	451738	-.02666	-.02189	-.02090	-.02774
263	STHWIND/ESTWIND	F	3289904	2631037		.00079	.00074	
264		T	674609	532631	-.00463	-.00384	-.00360	-.01123
265	STHWIND/OVCSKY	F	3350249	2788220		.00141	.00127	
266		T	614264	375448	-.00908	-.00767	-.00691	-.00689
267	STHWIND/HISKY	F	2825851	2115806		-.00319	-.00360	
268		T	1138662	1047862	.01112	.00792	.00893	.01099
269	STHWIND/FARVSBY	F	2500541	1721777		-.00174	-.00212	
270		T	1563972	1441891	.00441	.00267	.00325	.00428
271	STHWIND/NO PRECIP	F	2314388	1729057		.00886	.01087	
272		T	1650125	1434611	-.02128	-.01242	-.01525	-.02076
273	ESTWIND/OVCSKY	F	3407845	2862200		.00109	.00094	
274		T	556668	301468	-.00774	-.00665	-.00576	-.00725
275	ESTWIND/HISKY	F	3163955	2430289		-.00133	-.00133	
276		T	800558	733379	.00661	.00527	.00527	.00878
277	ESTWIND/FARVSBY	F	2825763	2123829		-.00075	-.00085	
278		T	1138750	1039839	.00262	.00187	.00211	.00381
279	ESTWIND/NO PRECIP	F	2745453	2128237		-.00208	-.00239	
280		T	1219060	1035431	.00676	.00468	.00538	.00612
281	OVCSKY/HISKY	F	3772933	2993018		-.00264	-.00141	
282		T	191580	170650	.05466	.05201	.02778	.05463
283	OVCSKY/FARVSBY	F	3043461	2429841		.00482	.00507	
284		T	921052	733827	-.02077	-.01594	-.01677	-.01705
285	OVCSKY/NO PRECIP	F	3008067	2464303		-.01959	-.02088	
286		T	956446	699365	.08120	.06161	.06566	.08379
287	HISKY/FARVSBY	F	1691683	939270		-.03571	-.04399	
288		T	2272830	2224398	.06229	.02658	.03275	.06369
289	HISKY/NO PRECIP	F	1513640	912557		.01602	.01938	
290		T	2450873	2251111	-.02591	-.00989	-.01197	-.01744

respectively. Constructing the beta coefficients is a common way for statisticians to give the coefficients relative status through standardizing; the higher the absolute value, the more important the predictor. The author gets more satisfaction in judging the importance of a predictor by realizing, in the B form, that the coefficient shows what a predictor (when it is "on") contributes to the estimated probability of Y=1, all other things equal. Such an appraisal is not ironclad either, owing to the effects of partial correlation, so let the reader beware of misinterpretation.

Another interesting equation, both for prediction and for interpretation, is the anomaly equation for NO WX/WX, where the station means have been removed. This equation appears in column 6 of table 2-1. A full version of the anomaly matrix $\underline{Aa}$ and its PLODITE form are given on microfiches F and G, respectively. When station-climatology adjustments are desired, the $\underline{Aa}$ matrix is employed with one additional ingredient: The additive constants, which are zero in $\underline{Aa}$, are replaced by the appropriate additive constants for the station desired. For 48 stations the additive constants have been determined from their respective climatologies and the $\underline{Aa}$ matrix and are on microfiche J.

Observations regarding Table 2-1 :

Note: Some of the calculations performed below are applicable only because the observed values of Z's and Y's are zero or one; e.g., $\Sigma Z = \Sigma Z^2$.

• Simple calculations that are possible--NO WX/WX both as predictor and predictand as an example:

Sample size is N = 3964513

- Predictor means: $\overline{Z} = \Sigma Z/N$, $\dfrac{3164088}{3964513} = .79810$

- Predictand mean: $\overline{Y} = \Sigma Y/N$, $\dfrac{3163668}{3964513} = .79800$

- Simple correlation coefficient squared:

$$R^2 = \frac{[\Sigma YZ - (\Sigma Y)\ (\Sigma Z)/N]^2}{(\Sigma Y - (\Sigma Y)^2/N)\ (\Sigma Z - (\Sigma Z)^2/N)}$$

$$= \frac{[3022632 - (3163668)\ (3164088)/3964513]^2}{(3163668 - (3163668)^2/3964513)\ (3164088 - (3164088)^2/3964513)} = .60675$$

- Since in Table 2-2 the multiple correlation coefficient squared is .65004, then (.65004-.60675) = .04329 or 4.33% is added to the reduction in variance over persistence by the other predictors.

• The beta coefficients reflect the influence of the predictor variances especially for visibility and weather when compared to PLODITE coefficients.

- Most elements have the same size coefficients for anomaly and regular regression.

- Some strong interactions are evident based on their coefficients. For example, OVCSKY/NO PRECIP = .06566, HISKY/FARVSBY = .03275, and DAY7-18/FARVSBY = -.04603. This last coefficient's sign is strange, but it is more acceptable realizing FARVSBY is =.04045, which tends to diminish the apparent strength of that interaction, giving a kind of nonadditivity correction.

- Month is stronger for anomaly equation of NO WX/WX predictand than regular regression.

- Higher temperatures show more of an effect on anomalies also.

The next important quantities, required for converting a probability forecast into a categorical forecast, are in step 9. These are μ_0 and μ_1. μ_0 is the mean of the predicted values $\hat{Y}$ over the sample N when the event was observed not to have occurred. Similarly, μ_1 is the mean of the predicted values $\hat{Y}$ over the sample N when the event was observed to have occurred. Their principal value is in the fact that the multiple correlation coefficient squared, R^2, for a particular predictand is

$$R^2 = \mu_1 - \mu_0 \tag{2-8}$$

(See the appendix.) This then satisfies step 10.

An important additional point to make here is as follows:

> R^2 for one hour is easily obtained from $\underline{A}$ and $\underline{Y}'\underline{Z}$. However, for subsequent hours such as 2, 3, ..., 24, the values for μ_0 and μ_1, and thereby R^2, cannot be obtained exactly from the quantities thus far derived. However, since $(\underline{Z}'\underline{Z})\underline{A} = (\underline{Y}'\underline{Z})_1$ with a 1 subscript on $(\underline{Y}'\underline{Z})$ to denote that Y is a one-hour prediction, a reasonable estimate of $(\underline{Y}'\underline{Z})_T$ for time T can be obtained from $(\underline{Z}'\underline{Z})\underline{A}^T \approx (\underline{Y}'\underline{Z})_T$.

This method of approximation was employed to get subsequent R^2's after the first hour.

The final derived quantity, in step 11, is the threshold probability P* for converting a probability forecast into a categorical forecast. That is, if the predicted probability of the first category exceeds the threshold of the first category, it becomes the category of the element that is predicted categorically. If it fails to exceed the threshold, the procedure is to accumulate probabilities, by adding the probability of the next category, and then to compare that accumulated probability against its threshold and so forth. A very detailed presentation on the thresholding method employed here is given in the appendix. The μ's and R^2s and P*'s for the hours 1-24 are given on microfiches H and I. Table 2-2 contains the values of μ_0, μ_1, R^2, and P* for hour 1 for demonstration purposes.

Table 2-2.--A display of quantities derived for GEM, for all predictands and for a 1-hr projection. Included in the four columns are: 1) μ_0 -- the mean of $\hat{Y}$ when Y did not occur, 2) μ_1 -- the mean of $\hat{Y}$ when Y did occur, 3) R^2 -- the multiple correlation coefficient squared (cumulative), and 4) P* -- the cumulative threshold probability for tripping categorical prediction, if exceeded by cumulative predicted probabilities. Month, hour of day, and interaction values are not shown for obvious reasons. SLP, DBT, DPD, and WIND P*s are not shown, because their categorical values are derived by a weighted-mean procedure, not by thresholding.

Number	Predictand		μ_0	μ_1	R^2	P*
	Element	Category				
38	SLP (MB)	800.0-985.0	.00007	.71912	.71904	
39		985.1-990.0	.00025	.77570	.77545	
40		990.1-995.0	.00076	.81783	.81707	
41		995.1-1000.0	.00229	.84308	.84079	
42		1000.1-1005.0	.00666	.86677	.86011	
43		1005.1-1010.0	.01848	.89760	.87913	
44		1010.1-1015.0	.04207	.93671	.89464	
45		1015.1-1020.0	.06805	.97170	.90365	
46		1020.1-1025.0	.08621	.98851	.90229	
47		1025.1-1030.0	.11342	.99566	.88224	
48		1030.1-1035.0	.13788	.99877	.86089	
49		1035.1-1040.0	.17613	.99973	.82360	
50		1040.1-1090.0	1.00000	1.00000		
51	DBT (°F)	-140 - -31	.00001	.52471	.52471	
52		-30 - -26	.00002	.63842	.63840	
53		-25 - -21	.00006	.71410	.71404	
54		-20 - -16	.00015	.76036	.76021	
55		-15 - -11	.00031	.79567	.79536	
56		-10 - -6	.00055	.82562	.82507	
57		-5 - -1	.00085	.85549	.85463	
58		0 - 4	.00128	.87390	.87262	
59		5 - 9	.00186	.88655	.88469	
60		10 - 14	.00288	.89142	.88855	
61		15 - 19	.00445	.89474	.89029	
62		20 - 24	.00718	.89614	.88896	
63		25 - 29	.01141	.90010	.88869	
64		30 - 34	.01734	.90903	.89169	
65		35 - 39	.02287	.92157	.89870	
66		40 - 44	.02883	.93218	.90335	
67		45 - 49	.03560	.94034	.90474	
68		50 - 54	.04434	.94680	.90246	
69		55 - 59	.05578	.95251	.89673	
70		60 - 64	.07176	.95790	.88614	
71		65 - 69	.09591	.96311	.86720	
72		70 - 74	.13716	.96896	.83181	
73		75 - 79	.18145	.97765	.79620	
74		80 - 84	.23464	.98552	.75088	
75		85 - 89	.30018	.99236	.69219	
76		90 - 94	.35057	.99731	.64675	
77		95 - 99	.35431	.99931	.64500	

Table 2-2.--(continued)

Number	Predictand Element	Predictand Category	μ_0	μ_1	R^2	P*
78	DBT (°F)	100 - 104	.39564	.99982	.60418	
79		105 - 109	.52118	.99997	.47879	
80		110 - 140	1.00000	1.00000		
81	DPD (°F)	0	.01451	.48796	.47345	
82		1	.03022	.60747	.57725	
83		2 - 4	.07031	.78750	.71720	
84		5 - 7	.10010	.85105	.75094	
85		8 - 11	.12829	.90001	.77173	
86		12 - 15	.15220	.92929	.77709	
87		16 - 19	.17851	.94798	.76947	
88		20 - 25	.22087	.96658	.74572	
89		26 - 35	.27878	.98534	.70656	
90		36 - 50	.37146	.99602	.62456	
91		51 - 99	1.00000	1.00000		
92	CC #1	CLR	.08246	.79062	.70816	.47000
93		SCD	.36028	.80124	.44095	.54400
94		BKN	.45660	.90427	.44768	.59800
95		OVC	.46891	.99136	.52246	.62000
96		TOT OBSC	1.00000	1.00000		
97	VIS (M)	.00 - .49	.00483	.50962	.50479	.37200
98		.50 - .74	.00660	.52953	.52293	.37900
99		.75 - .99	.00823	.54739	.53916	.38600
100		1.00 - 1.49	.01185	.57459	.56273	.39500
101		1.50 - 1.99	.01486	.59437	.57950	.40200
102		2.00 - 2.49	.01943	.61912	.59968	.41000
103		2.50 - 2.99	.02158	.63150	.60992	.41400
104		3.00 - 3.99	.02727	.67196	.64469	.42700
105		4.00 - 4.99	.03385	.70268	.66883	.43700
106		5.00 - 5.99	.04134	.72784	.68650	.44500
107		6.00 - 6.99	.04657	.74968	.70311	.45200
108		7.00 -100.00	1.00000	1.00000		
109	WEATHER	NO WX	.27926	.92930	.65004	.55000
110		WX	1.00000	1.00000		
111	FOG	NO FOG	.26936	.98195	.71259	.61028
112		FOG	1.00000	1.00000		
113	GROUND FOG	NO GF	.54583	.99013	.44430	.68307
114		GF	1.00000	1.00000		
115	HAZE, SMOKE	NO H, K	.31622	.97811	.66189	.62497
116		H, K	1.00000	1.00000		
117	BLOWING	NO B	.43409	.99884	.56474	.62489
118		B	1.00000	1.00000		
119	DRIZZLE	NO L	.59113	.99347	.40235	.72844
120		L	.92270	.99985	.07715	.87368
121		L, L+	1.00000	1.00000		
122	RAIN	NO R	.45163	.98246	.53083	.71714
123		R-	.82629	.99823	.17194	.81235
124		R	.93822	.99974	.06152	.89231
125		R+	1.00000	1.00000		

Table 2-2.--Continued

Number	Predictand Element	Predictand Category	μ_0	μ_1	R^2	P*
126	RAIN SHOWERS	NO RW	.73079	.98133	.25055	.81162
127		RW-	.94572	.99810	.05237	.91063
128		RW	.98506	.99935	.01429	.97673
129		RW+	1.00000	1.00000		
130	SNOW	NO S	.32060	.99363	.67303	.64591
131		S-	.74376	.99938	.25562	.75443
132		S	.84568	.99989	.15421	.80694
133		S+	1.00000	1.00000		
134	SNOW SHOWERS	NO SW	.57411	.99470	.42059	.71637
135		SW-	.94327	.99988	.05661	.89612
136		SW	.96917	.99998	.03081	.99363
137		SW+	1.00000	1.00000		
138	FREEZING DRIZZLE	NO ZL	.54380	.99942	.45562	.66785
139		ZL-, ZL, ZL+	1.00000	1.00000		
140	FREEZING RAIN	NO ZR	.60584	.99953	.39368	.69000
141		ZR-, ZR, ZR+	1.00000	1.00000		
142	THUNDERSTORM, A	NO TSM, A	.80106	.99390	.19283	.81687
143		TSM, A	1.00000	1.00000		
144	THUNDERSTORM +	NO TSM+	.99501	.99996	.00494	.99628
145		TSM+	1.00000	1.00000		
146	CH #1 (00')	0 - 1	.00403	.47510	.47107	.36000
147		2 - 4	.01460	.56754	.55295	.39300
148		5 - 6	.02167	.61597	.59430	.41000
149		7 - 9	.02931	.67639	.64708	.42900
150		10 - 14	.04063	.71636	.67574	.44200
151		15 - 19	.04890	.74326	.69436	.45100
152		20 - 24	.05731	.75853	.70123	.45700
153		25 - 29	.06470	.77501	.71031	.46300
154		30 - 39	.08200	.79708	.71508	.47200
155		40 - 49	.09816	.81626	.71810	.48000
156		50 - 59	.11076	.82931	.71854	.48600
157		60 - 75	.12499	.84286	.71787	.49300
158		76 - 99	.13687	.85168	.71480	.49700
159		100 - 150	.16962	.87668	.70706	.51100
160		151 - UNL	.60225	.99027	.38801	.66900
161		PART OBSC	1.00000	1.00000		
162	CC #2	CLR	.37015	.83975	.46960	.55600
163		SCD	.44941	.85867	.40926	.58500
164		BKN	.60896	.90828	.29931	.65600
165		OVC	1.00000	1.00000		
166	CH #2 (00')	0 - 1	.00011	.03649	.03638	.08091
167		2 - 4	.00221	.17946	.17725	.21800
168		5 - 6	.00429	.21987	.21558	.24400
169		7 - 9	.00762	.26086	.25324	.26700
170		10 - 14	.01418	.31402	.29985	.29500
171		15 - 19	.01925	.34480	.32555	.31000
172	CH #2 (00')	20 - 24	.02430	.36540	.34110	.32000
173		25 - 29	.02892	.38460	.35568	.33000

Table 2-2.--Continued

Number	Predictand Element	Predictand Category	μ_0	μ_1	R^2	P*
174	CH #2 (00')	30 - 39	.03698	.41296	.37598	.34300
175		40 - 49	.04387	.43572	.39186	.35300
176		50 - 59	.04957	.45160	.40202	.36000
177		60 - 75	.05937	.47501	.41564	.37100
178		76 - 99	.07010	.49117	.42707	.38100
179		100 - 150	.10420	.55217	.44797	.40700
180		151 - UNL	1.00000	1.00000		
181	TOTAL CLOUD COVER	CLR	.08244	.79063	.70819	.47000
182		SCD	.14692	.84060	.69368	.49700
183		BKN	.21239	.89148	.67909	.52500
184		OVC	1.00000	1.00000		
185	CEILING (00')	0 - 1	.00383	.48589	.48207	.36400
186		2 - 4	.01177	.59383	.58206	.40100
187		5 - 6	.01691	.63444	.61753	.41400
188		7 - 9	.02323	.67799	.65477	.42800
189		10 - 14	.03188	.70914	.67726	.43800
190		15 - 19	.03846	.72646	.68800	.44400
191		20 - 24	.04479	.73498	.69019	.44800
192		25 - 29	.05087	.74270	.69183	.45100
193		30 - 39	.06230	.75100	.68870	.45600
194		40 - 49	.07195	.75871	.68676	.46000
195		50 - 59	.07883	.76533	.68651	.46300
196		60 - 75	.08782	.77718	.68936	.46800
197		76 - 99	.09504	.78813	.69309	.47200
198		100 - 150	.11576	.81102	.69527	.48300
199		151 - UNL	1.00000	1.00000		
200	WIND	CALM	.04679	.29338	.24659	
201		NNE-NE < 11	.09370	.33947	.24577	
202		NNE-NE 11-19	.10888	.40968	.30081	
203		ENE-NE < 11	.13679	.50128	.36449	
204		ENE-NE 11-19	.14030	.55559	.41528	
205		ESE-SE < 11	.17143	.62665	.45522	
206		ESE-SE 11-19	.17696	.66589	.48893	
207		SSE-S < 11	.21899	.71017	.49118	
208		SSE-S 11-19	.23799	.75211	.51412	
209		SSW-SW < 11	.28091	.78605	.50514	
210		SSW-SW 11-19	.30517	.81577	.51059	
211		WSW-W < 11	.35446	.84289	.48843	
212		WSW-W 11-19	.39551	.86048	.46497	
213		WNW-W < 11	.46160	.88889	.42729	
214		WNW-NW 11-19	.57672	.90763	.33091	
215		NNW-N < 11	.57349	.95218	.37869	
216		NNW-N 11-19	.59861	.97762	.37901	
217		NNE-E > 19	.60779	.98007	.37229	
218		ESE-S > 19	.60357	.98389	.38032	
219		SSW-W > 19	.66175	.99150	.32975	
220		WNW-N > 19	1.00000	1.00000		

Remarks regarding table 2-2:

Computationally,

$$\mu_1 = \sum_{i=1}^{290} \left[\frac{\sum^N YZ}{\sum^N Z}\right] \cdot B_i \quad \text{or} \quad \mu_1 = \sum_{j=1}^{228} \left[\frac{\sum^N YZ}{\sum^N Z}\right] \cdot A_j$$

or

$$\mu_1 = \frac{3163668}{3964513}(.79800) + \frac{244842}{338217}(-.00778) + \frac{225026}{307968}(-.00698) + \ldots + \frac{2251111}{2450873}(-.00989)$$

$$= .92930$$

Also,

$$\mu_0 = \sum_{i=1}^{290} \left[1 - \frac{\sum^N YZ}{\sum^N Z}\right] \cdot B_i \text{ or } \mu_0 = \sum_{j=1}^{228} \left[1 - \frac{\sum^N YZ}{\sum^N Z}\right] \cdot A_j$$

or

$$\mu_0 = (1 - \frac{3163668}{3964513})(.79800) + (1 - \frac{244842}{338217})(-.00778) + (1 - \frac{225026}{307968})(-.00698) + \ldots$$

$$+ (1 - \frac{2251111}{2450873})(-.00989)$$

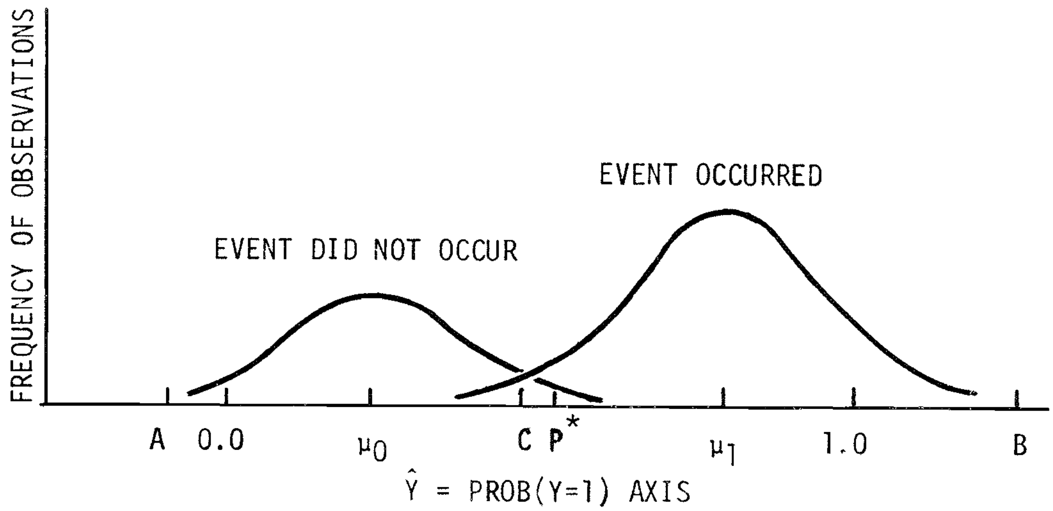
$$= .27926$$

Thus,

$$R^2 = \mu_1 - \mu_0$$

$= .65004$ or 65.004 percentage reduction in variance.

Furthermore, these parameters can be represented diagrammatically as:



SCHEMATIC OF THE SITUATION

Given: R^2, c

then: $\mu_0 = c\ (1-R^2)$

$\mu_1 = R^2 + \mu_0$

$\sigma_w^2 = R^2\ (1-R^2)\ c\ (1-c)$

P* Point at which area of total distribution to the left equals (1-c)

where σ_w^2 is the pooled within variance

c is the climatology

R^2 is the square of the multiple correlation coefficient

Depicted here are two distributions of the predicted value $\hat{Y}$, for when the event did not occur and the other for when the event did occur. μ_0 and μ_1 are the respective means of these distributions, while c is the grand mean of the total of the two distributions. The terminuses A and B are discussed in the appendix.

3. EXPERIMENTAL RESULTS, OLD AND NEW

Certain questions regarding GEM's capabilities have already been answered--if not completely, at least in part. The first question tested was: Can a comprehensive multiple-regression equation improve upon persistence in the very difficult problem of short-range forecasting of ceiling and visibility? The answer is that it can. At first, a screening of predictors succeeded in showing that this was true (Miller, 1964; Crisci and Lewis, 1973). For a single location, a similar answer was obtained in an equivalent Markov system on independent data using over 100 predictors at 3 hours and with an iterative scheme out to 6 hours. (See Miller et al., 1977.)

Another equivalent Markov approach, still not a generalized operator, yielded an affirmative answer on a large independent sample at 7 weather stations scattered over the continental United States. (See Miller, 1979b.) This Markov approach compared favorably with a regression-estimation-of-event-probabilities (REEP) method that made its projections directly.

These encouraging results prompted a series of GEM experiments designed to test 1) the value in a generalized operator of using all available predictors over a screened set, 2) the significance in a generalized operator of interactive predictors, 3) the importance in a generalized operator of including a location's climatology, and 4) the significance of a single-station set of equations over a generalized operator where climatology of the station has been included. The following sections will give detailed results of these experiments.

Air Weather Service Single-Station Experiment

The results in the Rickenbacker Air Force Base, Ohio, ceiling and visibility study yielded the following comparative Brier scores (Brier, 1950):

Weather element	GEM-like statistical technique	Conditional expectancy of persistence	Percent improvement
3-hr ceiling	.3755*	.4043	+7.1
3-hr visibility	.2564*	.2732	+6.1
6-hr ceiling	.4397*	.4763	+7.7
6-hr visibility	.2998*	.3175	+5.6

*Signifies superiority

where the statistical technique is a single-station (rather than generalized operator) iterative Markov approach, and where persistence utilizes probabilities conditioned on the hour of the day, month of the year, and the observed condition of the element at forecast time. The above figures were based on an independent sample of 29,154 forecasts. Other comparable figures were obtained for the other weather elements in the observation for the same independent sample tested.

Conditional Climatology Experiment

From a subsequent experiment, again applying single-station equations, a set of Brier scores, given below, compares the GEM-like procedure with the terminal-alert procedure (see Vercelli and Heffernan, 1978), which has already been shown to be more skillful than persistence. The terminal-alert procedure uses a REEP model.

Weather element	GEM-like statistical technique	Terminal alert procedure
DCA 1-hr ceiling	.193*	.198
DCA 1-hr visibility	.173*	.176
DCA 6-hr ceiling	.320	.319*
DCA 6-hr visibility	.306*	.310
SFO 1-hr ceiling	.192*	.200
SFO 1-hr visibility	.128*	.129
SFO 6-hr ceiling	.336*	.337
SFO 6-hr visibility	.215*	.216
SLC 1-hr ceiling	.133*	.135
SLC 1-hr visibility	.073	.072*
SLC 6-hr ceiling	.224	.223*
SLC 6-hr visibility	.121	.121
MSP 1-hr ceiling	.193*	.199
MSP 1-hr visibility	.109*	.110
MSP 6-hr ceiling	.354*	.357
MSP 6-hr visibility	.180	.180
MSY 1-hr ceiling	.196*	.201
MSY 1-hr visibility	.143*	.144
MSY 6-hr ceiling	.294*	.296
MSY 6-hr visibility	.222	.221*
PHL 1-hr ceiling	.237*	.245
PHL 1-hr visibility	.267*	.273
PHL 6-hr ceiling	.381	.380*
PHL 6-hr visibility	.453*	.461
MIA 1-hr ceiling	.212*	.216
MIA 1-hr visibility	.066*	.069
MIA 6-hr ceiling	.284	.282*
MIA 6-hr visibility	.091	.091

*Signifies superiority

These results are based on an independent sample of approximately 50,000 forecasts for each location. GEM-like forecasts, from data at the station being tested, were made for one hour on a direct basis, while the 6-hr forecasts were iterated hour by hour. The terminal-alert procedure forecasts were also single station, but the 6-hr forecasts were made directly. Paired comparison t tests were performed on each Brier score comparison. The conclusion was that the GEM-like technique was statistically significantly better than the terminal-alert procedure.

GEM Experiments

Analyses of variance and covariance experiments have been designed to test, in a hierarchical fashion, levels 1 through 5 (implicit here is a level 0 which uses climatological averages as a base):

Experiment 1.--Using all noninteractive predictors versus screened noninteractive predictors (level 2 versus level 1)

Experiment 2.--Adding interactive predictors versus no interactive predictors (level 3 versus level 2)

Experiment 3.--Station-adjusted climatology versus no station-adjusted climatology (level 4 versus level 3)

Experiment 4.--Single-station equations versus station-adjusted climatology (level 5 versus level 4)

The first two tests employ the analysis of variance in regression, while the last two tests use the analysis of covariance.

At the outset, the question is how many independent observations there are in the sample, considering the likelihood of high serial correlation in a set of consecutive hourly observations. This will have a decided bearing on the degrees of freedom specified in the statistical tests.

While serial correlation can be measured directly, there appears to be no available procedure for relating it to the issue of determining the number of independent observations in a sample. There is, however, a rational approach to the problem of determining the degree of "serial correlation," since all of the observations are zero-one. That is, calculate the number of runs in the sample for each predictor; then determine the sample size n that would, with no correlation, be expected to yield the number of runs r in that predictor having the fewest number of runs r_{min}. The determination of n is:

$$n = r_{min}/(2pq) \tag{3-1}$$

because the expected value is 2npq (see Mood, 1950) where p is the ratio of ones in the sample and q is the ratio of zeros in the sample. Finally, a factor f is determined to suggest the separation needed between observations to deem them independent:

$$f = N/n \tag{3-2}$$

In lieu of doing a random sampling of one out of f observations, a simpler but equivalent scheme is employed here: Divide each term in the $\underline{Z}'\underline{Z}$ and $\underline{Y}'\underline{Z}$ matrices by f. In this way the means, variances, and covariances would remain unbiased; however, the degrees of freedom in the test would be commensurate with the number of independent sample cases. Furthermore, it was considered unnecessary to use more than 1 1/2 million observations in performing these experiments. This degree of economy was accomplished by using data from only 15 representative stations of the original 41. The 15 chosen are identified in the station list in step 3 of chapter 2 by a I alongside the station name.

For this smaller sample (N=1,556,974) the factor f was found to be 18. Specifically, the predictor variable was the interactive term cold season (AUTWTR) and visibility $\geq$ 7 miles (FARVSBY) where

$$n = r_{min}/(2pq) = 40768/.48315 = 84380 \tag{3-3}$$

then

$$f = 1556974/84380 = 18.45 \tag{3-4}$$

Thus f = 18 was used as the divisor of $\underline{Z}'\underline{Z}$ and $\underline{Y}'\underline{Z}$.

It needs to be pointed out that the following tests apply only to the prediction scheme set up for 1-hr projections; retesting would be needed on other projections for which inferences are desired.

EXPERIMENT 1.--Using all non-interactive predictors versus screened noninteractive predictors (level 2 versus level 1)

The analysis-of-variance test is that of comparing the Brier score before and after adding all remaining non-interactive predictors to those screened non-interactive predictors. In particular, the F statistic is:

$$F \text{ (all predictors vs screening)} = \frac{[\text{BS (screening)} - \text{BS (all predictors)}] \cdot [n - P - 1]}{[\text{BS (all predictors)}] \cdot [(P - 1) - \text{ave. \# screened}]} \tag{3-5}$$

where

$$n = 86499$$

$$P = 193 \tag{3-6}$$

$$\text{Ave. \# screened} = 18$$

and where

$$F_{crit\ .01}\ (174{,}86305) = 1.28 \tag{3-7}$$

The results from this test are given in the fourth column of table 3-1 with the two Brier scores, BS (screening) and BS (all predictors), shown in the first and third columns, respectively. An asterisk in column 4 indicates a significant F value (1% level) was obtained and thereby suggests that adding all remaining predictors is important. Incidentally, for all predictands the use of screened predictors (level 1) was shown to be significant over climatological probability (level 0) and is reflected by all asterisks in column 2.

EXPERIMENT 2.--Adding interactive predictors versus no interactive predictors (level 3 versus level 2)

The appropriate procedure for testing the effects of adding interactive predictors to the set of all non-interactive predictors is again the analysis of variance; here the F statistic is:

$$F \text{ (with interactions vs no interactions)} = \frac{[BS \text{ (no interactions)} - BS \text{ (with interactions)}] \cdot [n - P - 1]}{[BS \text{ (with interactions)}] \cdot Q} \tag{3-8}$$

where

$$n = 86499$$

$$P = 228 \tag{3-9}$$

$$Q = \text{Number of interactive predictors} = 35$$

and where

$$F_{crit\ .01}\ (35,\ 86270) = 1.64 \tag{3-10}$$

The results from performing this test are given in the sixth column of table 3-1 with the Brier score, BS (with interactions), shown in the fifth column. An asterisk in the sixth column denotes the computed F statistic exceeded F_{crit}, thereby suggesting that adding these interactive predictors is important.

The interactive predictor set just tested and found to be significant for most predictands was initiated out of a discrete likelihood function study. (See Miller, 1979a.) Results from that study showed, in predicting NO WX/WX at Rickenbacker AFB, that there was a significant amount of interactive information--in the order of 4 percent of the remaining Brier score--over not using interactions. As a consequence, a set of very gross boolean interactive terms were constructed and used in the above test.

EXPERIMENT 3 and EXPERIMENT 4.-Station-adjusted climatology versus no station-adjusted climatology (level 4 versus level 3) and Single-station equations versus station-adjusted climatology (level 5 versus level 4)

One of the objectives in designing such a short-range forecasting procedure as GEM is to permit its use on a minicomputer. Efficiency in storage space would be achieved if individual station forecast equations would give way to a universal or generalized operator, applicable anywhere. For this to be possible, the usual stratification of data by location would have to be shown to be unnecessary.

The early concepts of restricting statistical prediction equations to particular seasons and hours of the day have already been shown to be questionable in this context. In fact, the enhancement in sample size afforded by the elimination of stratifying the data has more than compensated for the implied nonlinear effect in the system. However, rather than to accept this concept on faith, a statistical experiment was conducted to confirm or deny the desirability of station destratification.

Table 3-1.--Analyses of variance and covariance Brier scores and significance of test results. (Asterisk indicates significant result.)

Categories	(1) Screening/No interactions	(2) 0-1	(3) All predictors/ No interactions	(4) 1-2	(5) All predictors w/interactions	(6) 2-3	(7) All predictors/ Stn. adj. climatol.	(8) 3-4	(9) All predictors/ single station	(10) 4-5
DRY BULB TEMPERATURE (°F)										
-140 - -26	.00003	*	.00003	*	.00003	*	.00003		.00003	
-25 - -21	.00009	*	.00008		.00008	*	.00008		.00008	
-20 - -16	.00018	*	.00018		.00018		.00018		.00018	
-15 - -11	.00036	*	.00036		.00036		.00036		.00036	
-10 - -6	.00068	*	.00067		.00067		.00067		.00067	
-5 - -1	.00112	*	.00112		.00112		.00112		.00111	
0 - 4	.00171	*	.00171		.00171		.00171		.00169	
5 - 9	.00255	*	.00254		.00254		.00254		.00251	
10 - 14	.00396	*	.00396		.00396		.00396		.00391	
15 - 19	.00628	*	.00627		.00627		.00627		.00619	
20 - 24	.01033	*	.01032		.01032		.01032		.01014	
25 - 29	.01638	*	.01636		.01635		.01635		.01601	
30 - 34	.02372	*	.02365		.02364		.02364		.02304	
35 - 39	.02976	*	.02970		.02969		.02969		.02892	
40 - 44	.03452	*	.03448		.03448		.03447		.03364	
45 - 49	.03824	*	.03822		.03821		.03820		.03734	
50 - 54	.04254	*	.04251		.04250		.04250		.04150	
55 - 59	.04639	*	.04636		.04631	*	.04629	*	.04512	
65 - 69	.04913	*	.04910		.04906	*	.04905		.04777	
70 - 74	.04668	*	.04661		.04657	*	.04657		.04520	
75 - 79	.03970	*	.03961		.03957	*	.03954	*	.03841	
80 - 84	.02890	*	.02875	*	.02867	*	.02866		.02811	
85 - 89	.01795	*	.01781	*	.01775	*	.01775		.01742	
90 - 94	.00884	*	.00877	*	.00876	*	.00876		.00859	
95 - 99	.00264	*	.00263	*	.00263		.00263		.00257	
100 - 104	.00037	*	.00037		.00037		.00037		.00036	
105 - 140	.00003	*	.00003	*	.00003		.00003		.00002	*
Total asterisks:		27		6		9		2		1

DEWPOINT DEPRESSION
(°F)

0	.01131	*	.01115	*	.01110	*	.01108	*	.01069	
1	.02533	*	.02506	*	.02493	*	.02490	*	.02434	
5 - 7	.09086	*	.08999	*	.08795	*	.08786	*	.08615	
8 - 11	.09565	*	.09512	*	.09486	*	.09483	*	.09326	
12 - 15	.08090	*	.08059	*	.08049	*	.08046	*	.07918	
16 - 19	.06506	*	.06483	*	.06479		.06477		.06396	
20 - 25	.05948	*	.05915	*	.05910	*	.05908	*	.05808	
26 - 35	.04345	*	.04305	*	.04299	*	.04296	*	.04184	
36 - 50	.02114	*	.02094	*	.02092	*	.02088	*	.02023	
51 - 99	.00586	*	.00580	*	.00579	*	.00578	*	.00540	*
Total asterisks:		10		10		9		9		1

VISIBILITY
(St. mi.)

.00 - .49	.00443	*	.00436	*	.00433	*	.00433		.00423	
.50 - .74	.00331	*	.00329	*	.00329		.00329		.00324	
.75 - .99	.00336	*	.00335	*	.00335		.00335		.00331	
1.00 - 1.49	.00702	*	.00699	*	.00698		.00698		.00688	
1.50 - 1.99	.00743	*	.00741	*	.00741	*	.00740	*	.00729	
2.00 - 2.49	.01063	*	.01061		.01061	*	.01060		.01046	
2.50 - 2.99	.00738	*	.00737		.00737		.00736	*	.00724	
3.00 - 3.99	.01624	*	.01621		.01620	*	.01619		.01598	
4.00 - 4.99	.01980	*	.01976		.01974	*	.01973	*	.01973	
5.00 - 5.99	.02195	*	.02190		.02189	*	.02187		.02187	
6.00 - 6.99	.01870	*	.01866		.01861	*	.01859	*	.01833	
Total asterisks:		11		5		7		4		0

WEATHER

NO WX/WX	.05703	*	.05547	*	.05505	*	.05458	*	.05329	
F	.01632	*	.01568	*	.015570	*	.01554	*	.01504	
GF	.00755	*	.00737	*	.00728	*	.00727	*	.00706	
H,K	.02244	*	.02200	*	.02193	*	.02169	*	.02089	*
B	.00099	*	.00098	*	.00098		.00098		.00094	*
L-	.00642	*	.00630	*	.00628	*	.00628	*	.00614	
L, L+	.00009	*	.00009		.00009		.00009		.00008	*

Table 3-1.--(continued)

Categories	(1) Screening/No interactions	(2) 0-1	(3) All predictors/ No interactions	(4) 1-2	(5) All predictors w/interactions	(6) 2-3	(7) All predictors/ Stn. adj. climatol.	(8) 3-4	(9) All predictors/ single station	(10) 4-5
WEATHER (cont.)										
R-	.01620	*	.01594	*	.01587	*	.01585	*	.01555	
R	.00159	*	.00158		.00158	*	.00158		.00156	
R+	.00027	*	.00027		.00027		.00027		.00026	
RW-	.01768	*	.01736	*	.01724	*	.01722	*	.01692	
RW	.00135	*	.00134	*	.00133	*	.00133		.00132	
RW+	.00063	*	.00062	*	.00062		.00062		.00062	
S-	.00616	*	.00600	*	.00596	*	.00595	*	.00575	
S	.00054	*	.00053	*	.00053	*	.00053		.00051	*
S+	.00008	*	.00008	*	.00008		.00008		.00007	*
SW-	.00472	*	.00466	*	.00464	*	.00462	*	.00446	
SW, SW+	.00009	*	.00009		.00009		.00009	*	.00008	*
ZL-, ZL, ZL+	.00051	*	.00051	*	.00051		.00051	*	.00050	
ZR-, ZR, ZR+	.00043	*	.00042	*	.00042		.00042	*	.00042	
TSM-	.00530	*	.00523	*	.00522	*	.00522	*	.00510	
TSM+	.00005	*	.00005		.00005		.00005		.00004	*
Total asterisks:		22		17		13		13		7
WIND										
Calm	.04542	*	.04471	*	.04461	*	.04392	*	.04258	
NNE-NE LE 10	.04645	*	.04637		.04633	*	.04624	*	.04482	
NNE-NE 11-19	.02030	*	.02023	*	.02019	*	.02013	*	.01937	*
ENE-E LE 10	.04543	*	.04535		.04529	*	.04508	*	.04353	
ENE-E 11-19	.01736	*	.01731		.01728	*	.01725	*	.01673	
ESE-SE LE 10	.05794	*	.05783		.05760	*	.05743	*	.05578	
ESE-SE 11-19	.02418	*	.02405	*	.02390	*	.02384	*	.02294	*
SSE-S 11-19	.03659	*	.03644	*	.03636	*	.03598	*	.03466	
SSW-SW LE 10	.06190	*	.06176		.06126	*	.06102	*	.05905	
SSW-SW 11-19	.03351	*	.03339	*	.03308	*	.03293	*	.03191	
WSW-W LE 10	.05628	*	.05596	*	.05585	*	.05536	*	.05329	*

WSW-W 11-19	.03067	*	.03037	*	.03026	*	.02999	*	.02838	*
WNW-NW LE 10	.05522	*	.05512		.05496	*	.05447	*	.05260	
WNW-NW 11-19	.03358	*	.03336	*	.03320	*	.03306	*	.03190	
NNW-N LE 10	.05091	*	.05078		.05071	*	.05043	*	.04873	
NNW-N 11-19	.02709	*	.02701		.02698	*	.02690	*	.02613	
NNE-E GE 20	.00325	*	.00324	*	.00323	*	.00323	*	.00308	*
ESE-S GE 20	.00467	*	.00466		.00466	*	.00465	*	.00454	
SSW-W GE 20	.00628	*	.00623	*	.00622	*	.00621	*	.00606	
NNW-N GE 20	.00816	*	.00810	*	.00809	*	.00807	*	.00778	
Total asterisks:		20		11		20		20		5

SEA LEVEL PRESSURE
(mb)

800.0 - 985.0	.00008	*	.00008		.00008		.00008		.00008	*
985.1 - 990.0	.00032	*	.00032		.00032		.00032		.00032	
990.1 - 995.0	.00099	*	.00098		.00098		.00098		.00098	
995.1 - 1000.0	.00305	*	.00305		.00305		.00305		.00303	
1000.1 - 1005.0	.00873	*	.00871		.00871		.00871		.00866	
1005.1 - 1010.0	.02256	*	.02248	*	.02248		.02246	*	.02232	
1010.1 - 1015.0	.04262	*	.04246	*	.04246		.04239	*	.04209	
1020.1 - 1025.0	.02946	*	.02937	*	.02937		.02937		.02917	
1025.1 - 1030.0	.01403	*	.01399		.01399		.01399		.01390	
1030.1 - 1035.0	.00536	*	.00534		.00534		.00534		.00531	
1035.1 - 1040.0	.00145	*	.00145		.00145		.00145		.00144	
1040.1 - 1090.0	.00026	*	.00025	*	.00025		.00025		.00025	
Total asterisks:		12		4		0		2		1

CLOUD COVER #1

Clear	.06313	*	.06212	*	.06196	*	.06185	*	.06105	
Broken	.12003	*	.11930	*	.11896	*	.11875	*	.11741	
Overcast	.07603	*	.07437	*	.07390	*	.07351	*	.07214	
Total observation	.00759	*	.00745	*	.00741	*	.00741		.00725	
Total asterisks:		4		4		4		3		0

Table 3-1.--(continued)

Categories	(1) Screen-ing/No inter-actions	(2) 0-1	(3) All pre-dictors/ No inter-actions	(4) 1-2	(5) All pre-dictors w/inter-actions	(6) 2-3	(7) All pre-dictors/ Stn. adj. climatol.	(8) 3-4	(9) All pre-dictors/ single station	(10) 4-5
CLOUD HEIGHT #1 (100 ft)										
0-1	.00342	*	.00340	*	.00338	*	.00338		.00330	
2-4	.01298	*	.01273	*	.01270	*	.01270	*	.01248	
5-6	.01456	*	.01440	*	.10438	*	.01438		.01421	
7-9	.02119	*	.02102	*	.02097	*	.02096	*	.02061	
10-14	.02940	*	.02919	*	.02910	*	.02908	*	.02850	
15-19	.02633	*	.02617	*	.02614	*	.02613	*	.02575	
20-24	.02417	*	.02397	*	.02395	*	.02394		.02366	
25-29	.02326	*	.02306	*	.02305		.02304		.02276	
30-39	.03678	*	.03630	*	.03628	*	.03623	*	.03566	
40-49	.03379	*	.03341	*	.03338	*	.03337	*	.03286	
50-59	.02783	*	.02759	*	.02757		.02755	*	.02717	
60-75	.03028	*	.02997	*	.02995		.02993	*	.02935	
76-99	.02368	*	.02339	*	.02338		.02326	*	.02244	
100-150	.04696	*	.04646	*	.04640	*	.04633	*	.04577	
Partial obscuration	.01065	*	.01044	*	.01043	*	.01042	*	.01007	
Total asterisks:		15		15		11		11		0
CLOUD COVER #2										
Scattered	.05294	*	.05229	*	.05226		.05204	*	.05124	
Broken	.07650	*	.07564	*	.07534	*	.07517	*	.07423	
Overcast	.07813	*	.07712	*	.07701	*	.07688	*	.07591	
Total asterisks:		3		3		2		3		0
CLOUD HEIGHT #2 (100 ft)										
0-1	.00016	*	.00016	*	.00016		.00016		.00015	*
2-4	.00257	*	.00254		.00254		.00254	*	.00248	

5-6	.00248	*	.00247	*	.00247		.00247		.00243	
7-9	.00421	*	.00417	*	.00416		.00416	*	.00409	
10-14	.00851	*	.00836	*	.00835	*	.00835	*	.00820	
15-19	.00762	*	.00756	*	.00755	*	.00755	*	.00746	
20-24	.00765	*	.00760	*	.00759		.00759		.00730	
25-29	.00747	*	.00743	*	.00743		.00743		.00737	
30-39	.01305	*	.01295	*	.01294	*	.01292	*	.01276	
40-49	.01167	*	.01157	*	.01156	*	.01154	*	.01137	
50-59	.00986	*	.00979	*	.00979		.00977	*	.00966	
60-75	.01471	*	.01470	*	.01470		.01467	*	.01445	
76-99	.01392	*	.01387	*	.01386		.01386	*	.01374	
100-150	.04347	*	.04284	*	.04275	*	.04263	*	.04177	
Total asterisks:		14		13		5		10		1

TOTAL CLOUD COVER

Clear	.06311	*	.06211	*	.06195	*	.06184	*	.06105	
Scattered	.11021	*	.10924	*	.10909	*	.10894	*	.10775	
Broken	.10863	*	.10740	*	.10691	*	.10684	*	.10567	
Total asterisks:		3		3		3		3		0

CEILING
(100 ft)

0-1	.00327	*	.00324	*	.00322	*	.00322		.00315	
2-4	.01061	*	.01043	*	.01041	*	.01040	*	.00974	*
5-6	.01130	*	.01121	*	.01120	*	.01120	*	.01105	
7-9	.01686	*	.01676	*	.01673	*	.01672	*	.01643	
10-14	.02219	*	.02209	*	.02205	*	.02204		.02164	
15-19	.01870	*	.01863	*	.01861	*	.01861	*	.01839	
20-24	.01649	*	.01638	*	.01637	*	.01636		.01619	
25-29	.01599	*	.01589	*	.01588		.01588	*	.01575	
30-39	.02463	*	.02444	*	.02443		.02441		.02418	
40-49	.02224	*	.02211	*	.02209		.02206	*	.02185	
50-59	.01730	*	.01721	*	.01721		.01719	*	.01705	
60-75	.02283	*	.02275	*	.02273	*	.02272	*	.02251	
76-99	.01840	*	.01834	*	.01833		.01833	*	.01818	
100-150	.04112	*	.04092	*	.04087	*	.04086		.04047	
Total asterisks:		14		14		9		9		1

The appropriate model for testing the effects of grouping data is that of R. A. Fisher's analysis of covariance. For a lucid exposition of the analysis of covariance see Tatsuoka (1971).

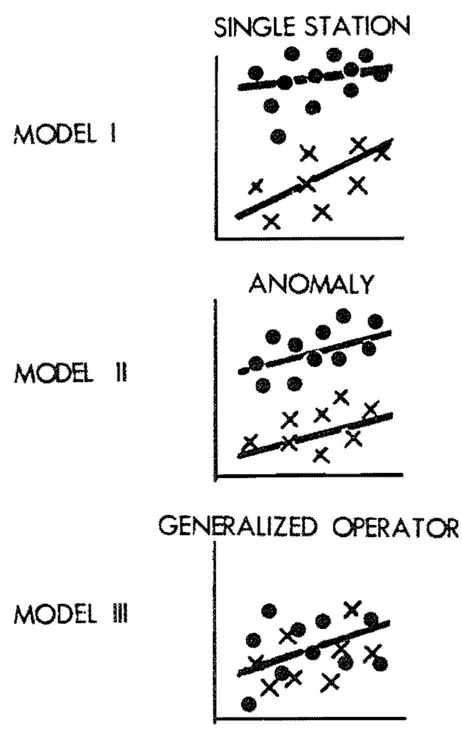
The effort here will be to determine which one of the following three models is most appropriate for representing the true situation:

Model I: The prediction of a weather element one hour hence should be represented by an individual-station (single-station) regression equation.

Model II: The prediction of a weather element one hour hence should be represented by the same regression equation everywhere except the station's climatology should be accounted for (anomaly).

Model III: The prediction of a weather element one hour hence should be represented by the same regression equation without restriction (generalized operator).

A schematic representation of these models for the analysis of covariance is depicted in the following:



Symbolized are data from two stations on a predictor-predictand graph. Dots are for one station and crosses are for the other. Model I denotes fitting is required for each station separately. Model II denotes that the same function between predictor and predictand suffices, but there is a difference in means. Model III denotes a single relationship applies for all of the data.

The analysis of covariance, in helping to decide which model to use, takes into account the important fact that the predictor observations differ from one location to another and therefore could account for the apparent predictand variations. Briefly, the procedure is to create cross-product matrices among all of the predictors and predictands, $\underline{Z}'\underline{Z}_k$ and $\underline{Y}'\underline{Z}_k$, for station k's data. Then each matrix is made into an anomaly matrix for each station by removing the mean values. Finally, composite anomaly matrices are made by summing these k (k=1, 2,..., K) station matrices.

Using Tatsuoka's nomenclature, the procedure is written for one of the Y's and one of the Z's as:

Y_{ki} = Predictand value of observation i at station k.

Z_{ki} = Predictor value of observation i at station k.

$Y_{k.} = \sum_{i=1}^{n_k} Y_{ki}$ = Sum of Y values for kth station, where n_k equals the number of observations from the k^{th} station.

$Z_{k.} = \sum_{i=1}^{n_k} Z_{ki}$ = Sum of Z values for k^{th} station.

$Y_{..} = \sum_{k=1}^{K} Y_{k.}$ = Grand total of Y values in entire sample of K stations combined.

$Z_{..} = \sum_{k=1}^{K} Z_{k.}$ = Grand total of Z values in entire sample of K stations combined.

(3-11)

In the present situation, the number of stations is K=15, and the individual station sample sizes n_k (k=1, 2, ... K) are given in step 4 of chapter 2, Creating GEM.

The analysis of covariance proceeds by computing the customary within-station and total sums-of-squares of Y as given by

$$SS_w = \sum_{k=1}^{K} \left[\sum_{i=1}^{n_k} Y_{ki}^2 - Y_{k.}^2/n_k \right] \qquad (3\text{-}12)$$

and

$$SS_t = \sum_{k=1}^{K} \sum_{i=1}^{n_k} Y_{ki}^2 - Y_{..}^2/N \qquad \text{where } N = \sum_{k=1}^{K} n_k \qquad (3\text{-}13)$$

respectively.

Again following Tatsuoka, similar quantities are needed for each of the Z's. In Tatsuoka's revised notation:

$$S_{k,yy} = \sum_{i=1}^{n_k} Y_{ki}^2 - Y_{k.}^2/n_k \tag{3-14}$$

and

$$S_{k.zz} = \sum_{i=1}^{n_k} Z_{ki}^2 - Z_{k.}^2/n_k \tag{3-15}$$

with

$$S_{k,zy} = S_{k,yz} = \sum_{i=1}^{n_k} Z_{ki}\, Y_{ki} - Z_{k.}\, Y_{k.}/n_k \tag{3-16}$$

Needed now is a pooling of these within-group quantities, letting W and T represent their values as:

$$W_{yy} = \sum_{k=1}^{K} S_{k,yy}$$

$$W_{zz} = \sum_{k=1}^{K} S_{k,zz}$$

$$W_{zy} = W_{yz} = \sum_{k=1}^{K} S_{k,zy} \tag{3-17}$$

and

$$T_{yy} = \sum_{k=1}^{K} \sum_{i=1}^{n_k} Y_{ki}^2 - Y_{..}^2/N$$

$$T_{zz} = \sum_{k=1}^{K} \sum_{i=1}^{n_k} Z_{ki}^2 - Z_{..}^2/N$$

$$T_{zy} = T_{yz} = \sum_{k=1}^{K} \sum_{i=1}^{n_k} Z_{ki}\, Y_{ki} - Z_{..}Y_{..}/N \tag{3-18}$$

Extensions of the notation for P predictors $Z_1, \ldots, Z_P$, and letting Z_0 denote Y (for the moment), which is still only a single predictand, gives

$Z_{\alpha ki}$ = The value of Z_α ($\alpha=0,1,\ldots,P$) of the i^{th} observation at the k^{th} location

$$Z_{\alpha k.} = \sum_{k=1}^{n_k} Z_{\alpha ki} \qquad (\alpha=0,1,\ldots,P)$$

$$Z_{\alpha ..} = \sum_{k=1}^{n_k} Z_{\alpha k.} \qquad (\alpha=0,1,\ldots,P) \tag{3-19}$$

Now the quantities are prepared for testing whether Model I, II, or III obtains. That is,

$$S_{k,\alpha\beta} = \sum_{i=1}^{n_k} Z_{\alpha ki}\ Z_{\beta ki} - Z_{\alpha k.}\ Z_{\beta k.}/n_k$$

$$(k = 1,\ldots,K;\ \alpha,\beta = 0,1,\ldots,P)$$

$$W_{\alpha\beta} = \sum_{k=1}^{K} S_{k,\alpha\beta} \qquad \text{(within locations)}$$

$$T_{\alpha\beta} = \sum_{k=1}^{K} \sum_{i=1}^{n_k} Z_{\alpha ki}\ Z_{\beta ki} - Z_{\alpha..}\ Z_{\beta..}/N \tag{3-20}$$

These terms are collected into several matrices--$\underline{S}_k$ (k=1, ..., K), $\underline{W}$, and $\underline{T}$. Ultimately, for testing, the following quantities are needed:

$$S_1 = W_{00} - \sum_{k=1}^{K} \underline{S}_{k,0p}\ \underline{S}_{k,pp}^{-1}\ \underline{S}_{k,p0}$$

$$S_2 = W_{00} - \underline{W}_{0p}\ \underline{W}_{pp}^{-1}\ \underline{W}_{p0}$$

$$S_3 = T_{00} - \underline{T}_{0p}\ \underline{T}_{pp}^{-1}\ \underline{T}_{p0} \tag{3-21}$$

with

$$S_4 = S_2 - S_1$$

$$S_5 = S_3 - S_2 \tag{3-22}$$

then

$$F_\eta = (S_4/\nu_4)\ /\ (S_1/\nu_1) \tag{3-23}$$

is the test statistic for judging whether the hypothesis in Model II is acceptable. Here the degrees of freedom, ν_1 and ν_4, are:

$$\nu_1 = n - (P+1)\ K$$

$$\nu_4 = P(K-1) \tag{3-24}$$

Also,

$$F_\mu = (S_5/\nu_5)\ /\ (S_2/\nu_2) \tag{3-25}$$

is the test statistic for judging whether the hypothesis in Model III is acceptable, provided the hypothesis in Model I was not accepted, where the appropriate degrees of freedom, ν_2 and ν_5, are:

$$\nu_2 = n - K - P$$

$$\nu_5 = K - 1 \tag{3-26}$$

In the particular analysis of covariance problem analyzed here,

$$n = 86499$$

$$P = 228$$

$$K = 15$$

Furthermore, tests were conducted for all predictand Y's, excluding one predictand in each weather element.

The results of the F_η and F_μ tests are presented in columns 10 and 8, respectively, in table 3-1. An asterisk is used to show significance at the 1-percent level. For example, if an asterisk appears in column 10, then accept Model I; if an asterisk is in column 8 (provided one does not appear in its corresponding column 10), then accept Model II. By default, Model III is accepted when neither column 10 nor 8 has an asterisk for that predictand variable.

An example of the calculations performed in this series of tests for NO WX/WX is given in the following:

Predictand--NO WX/WX 1 hour hence

k	Weather station	n_k	Single-station Brier score
1	MKE	98865	.06068
2	DEN	104401	.03561
3	LAX	105052	.06474
4	BIS	105011	.04787
5	BOS	104989	.06377
6	ABQ	105002	.02499
7	MEM	105063	.04853
8	STL	103908	.05728
9	JAX	104890	.06369
10	OKC	105001	.03715
11	PIT	103156	.08902
12	SAT	102016	.03787
13	RDU	103602	.05641
14	PDX	104056	.08782
15	RNO	101962	.02407

BS (single-station) = .05329

BS (anomaly) = .05458

BS (generalized operator) = .05505

Then

$$F_\eta = \frac{[\text{BS (anomaly)} - \text{BS (single station)}] \cdot [n - (P+1)K]}{[\text{BS (single station)}] \cdot [P(K-1)]} \tag{3-28}$$

Thus

$$F_\eta = \frac{(.05458 - .05329) \cdot (83064)}{(.05329) \cdot (3192)} = .63, \tag{3-29}$$

which is not significant, since $F_{crit.01}$ $(\infty,\infty) = 1.00$.

The hypothesis of Model II is not rejected, and therefore no asterisk appears in column 10 for NO WX/WX in table 3-1.

Proceeding now to test whether Model III should be rejected, F_μ is tested. That is,

$$F_\mu = \frac{[\text{BS (generalized operator)} - \text{BS (anomaly)}] \cdot [n - K - P]}{[\text{BS (anomaly)}] \cdot [K-1]} \tag{3-30}$$

Thus

$$F_\mu = \frac{(.05505 - .05458) \cdot (86256)}{(.05458) \cdot (14)} = 53.05 \tag{3-31}$$

This causes a rejection of Model III, because F exceeds the $F_{crit.01}$ $(14,\infty) = 2.08$. This leaves Model II as the appropriate one to accept. This rejection appears as an asterisk in column 8 of table 3-1 for NO WX/WX. All of the other predictand elements were tested in a similar manner, with their results in columns 8 and 10. It may be noted that the left-out predictand dummy was not tested along with the others. This was considered a redundant test and, if it is of special interest the test result may be inferred from the results of those that were tested for that weather element.

In summary, the proper way to interpret the results in table 3-1 is to:

- Accept Model I (single-station equation is best) if an asterisk is in column 10.
- Accept Model II (station-adjusted climatology, anomaly generalized operator) if an asterisk appears in column 8 but not in column 10.
- Accept Model III (straight generalized operator) if no asterisk appears in column 8 or 10.
- Prefer including interactive predictors to not including interactive predictors if an asterisk appears in column 6.
- Prefer including all predictors over screening if an asterisk appears in column 4.
- Prefer using a screened set of predictors over using climatological probabilities if an asterisk appears in column 2.

Testing the value of two observations in the predictor set

Another experiment included predictors from two consecutive observations. This scheme is more powerful than explicitly including one-hour tendencies as predictors, since the coefficients for each term can vary, while a tendency coefficient is fixed on both terms. Only single-station data from DCA were used in the two-observation experiment. It amounts to solving a 377-predictor regression problem with the usual 227 predictands for one hour hence.

The results were surprising but definitive. They showed that only 1 of the 227 predictands was aided significantly by these 89 additional predictors (not double the original 228, since month of year, hour of day, and the gross interactions were not entered again). The one significant situation that was encountered could have been expected by chance, since a 1-percent test was performed.

Analyzing anomaly effects

A number of worthwhile investigations can be made from the quantities prepared for GEM, in the matrices and in the equations. One of these will be demonstrated.

In the station-adjusted climatology (anomaly) set of equations, the additive constant is always zero for each predictand, because the predictand equations estimate the deviation of the predictand from its mean, just as the predictors are deviations from their means. However, by taking any station's climatology for each predictor and any particular predictand, a station-tailored additive constant can be determined. The overall climatology (including all 41 stations) also yields an additive constant for each predictand. When this is done for each station for, say, NO WX/WX, the additive constants can be compared in a meaningful way. In particular, a plot can be made of the differences between each station's and the overall additive constant. This has been done in figure 3-1. Positive differences mean that the station would have a higher probability of NO WX by that amount, and vice versa, all other things equal. Note the concentration of negative differences in the northeast, and in other industrialized areas.

Another point that is worth mentioning about these differences is that the squares of the differences are equal to the Brier score reductions that could be realized if station-adjusted climatology equations were invoked in place of straight, generalized-operator equations.

Conclusions

The Brier score results presented in table 3-1 provide evidence upon which the following observations are based:

- Screening predictors yields a significant improvement over climatology on all elements.

- Adding the remaining predictors to the screened set also provides a significant improvement in 105 of the 155 elements in the predictand set.

- Including interactive predictors to the total set of predictors was significant in 92 of the 155 predictands.

- Adjusting for station climatology was significant in 89 of the 155 predictands.

- Single-station equations were shown to be significantly better in only 17 of the 155 predictands over station-adjusted climatology.

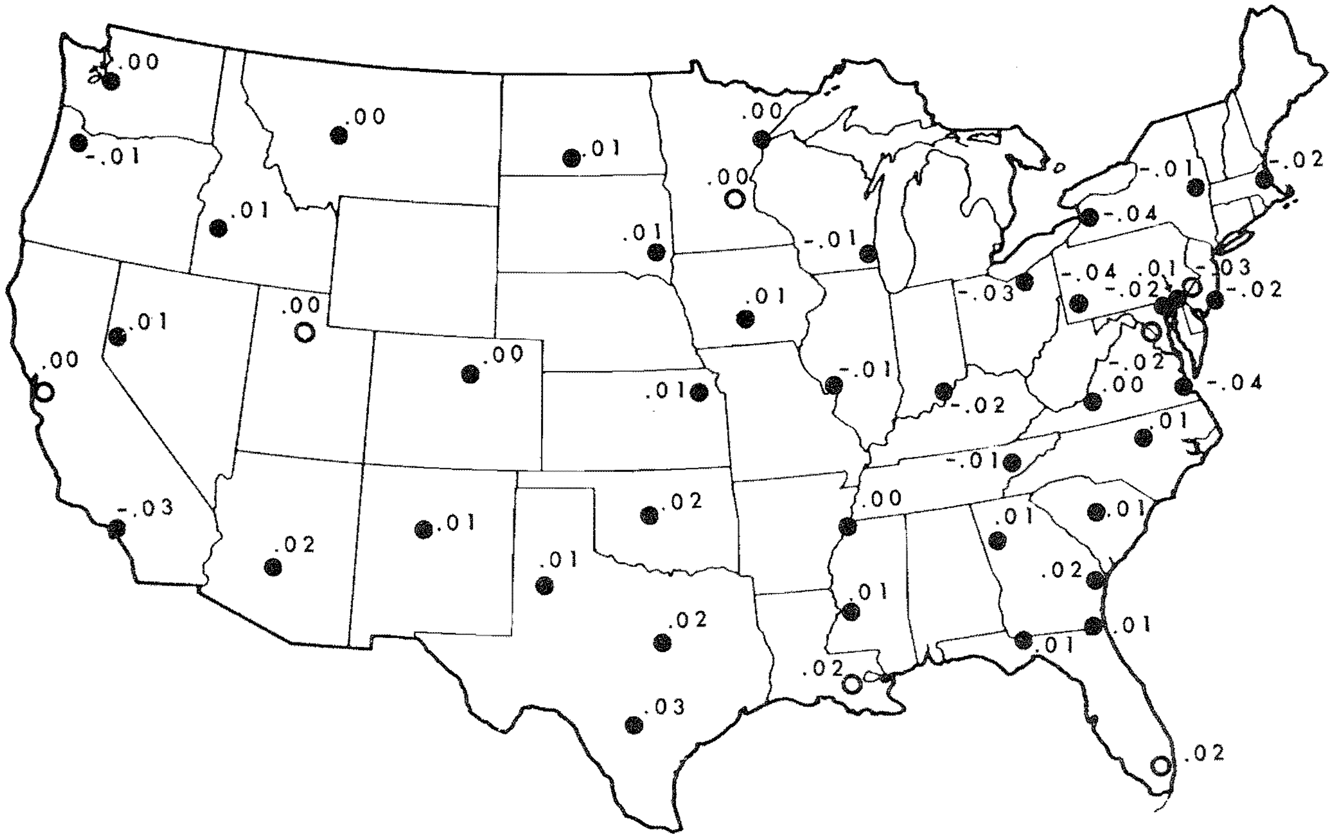


Figure 3.1.--Plot of difference between anomaly additive constants [A_0 (station) - A_0 (total)], for NO WX/WX.

It is thereby concluded that adding more predictors in the regression equations increases the skill of the predictions for most of the elements and should be preferable to screening. Adding interactive predictors, even though only grossly representing nonlinear input, has been shown to increase the accuracy of the forecasts and is therefore a recommended procedure to follow.

Station-adjusted climatology is important in improving the results from a statistically significant standpoint.

It is concluded that deriving equations to predict only at individual single stations will not enhance the skill of the forecast system over that of station-adjusted climatology generalized operators when the number of degrees of freedom consumed in the process is duly accounted for. It is concluded, therefore, that effects of local conditions--terrain, proximity to water, latitude, longitude, altitude, and the like--be accounted for by a station-adjusted climatology generalized operator.

Since inclusion of another observation failed to provide a significant improvement in skill, it is concluded also that a Markov model is appropriate in making a 1-hr prediction.

4. INDEPENDENT VERIFICATION OF RESULTS

Demonstrating the skill of a new statistical weather prediction system or any prediction system can be accomplished by subjecting it to a large, independent, historical sample or by evaluating its usefulness on a day-by-day exposure to the ultimate users of the guidance product--the practicing forecasters. A feedback of their observations could be most beneficial for tailoring its form and ultimate acceptance. Because of time considerations, however, the verification scheme selected here was the former.

A set of seven locations, not part of the 41 stations making up the dependent sample, was selected for a large-scale verification. The stations selected were the same seven tested and discussed in another context in chapter 3, Experiments. Since GEM predicts for any hour and any month, it was believed desirable to process all the approximately 700,000 independent forecasts. The processing time for making this many hourly forecasts out to 12 hours would have taken excessive computer time. To implement a practical subset verification, the following effort was carried out:

- Seven locations: DCA, PHL, SFO, SLC, MSP, MSY, and MIA.

- 26,328 independent forecasts covering all locations for the years 1954-1965.

- All hours of the day and all months of the year sampled, the scheme being to begin on the first day of the period sampled at 00, the second day at 01, the third day at 02, etc., separating the observations adequately to assure an even distribution without getting involved in randomizing.

- Projections for 1, 3, 6, 9, and 12 hours.

- All predictand elements in GEM except NO WX/WX were tested: T, DPD, V, P, F, GF, HK, B, L, R, RW, S, SW, ZL, ZR, TSM, TSM+, CC#1, CH#1, CC#2, CH#2, TCA, C, and W.

- The comparative method was persistence--measured primarily from the independent sample contingency table conditional probabilities.

- Statistics computed were: Brier score, percent correct (hits), Heidke skill score, and a contingency table of observed versus categorically forecasted conditions. Tables of summarizing statistics have been compiled for easy appraisal of the results.

Brier scores for each projection and for all elements are presented in table 4-1. For comparison, Brier scores were calculated for the conditional probability given persistence, derived from the same observational data used as input for the GEM forecast process for projections of 3, 6, 9, and 12 hours. Since these persistence Brier scores were computed from conditional persistence tables of the independent sample, they are biased favoring persistence. A persistence Brier score for a 1-hr projection, computed from the dependent sample used to develop GEM, is readily available and is also presented. The persistence Brier scores, for each projection and element, are also displayed in table 4-1.*

*The reader is directed to chapter 7, New Results, for the most recent verifications.

Table 4-1.--Independent sample Brier scores from 26,328 cases at seven stations for GEM and persistence. Projections are for 1, 3, 6, 9, and 12 hours. Persistence Brier scores are computed from a conditional persistence table of independent samples (except 1-hr), thus producing a bias favoring persistence.

		BRIER SCORE									
		GEM					PERSISTENCE				
Weather element		1 hr.	3	6	9	12	1 hr.	3	6	9	12
T	1	.22827	.35550	.40768	.42421	.42923	.22884	.35524	.40724	.42397	.42948
DPD	2	.27447	.36235	.39533	.40473	.40871	.27953	.37361	.41315	.42427	.42727
V	3	.08232	.10912	.12628	.13250	.13776	.08379	.11187	.12951	.13458	.13874
F	4	.01304	.02652	.03690	.04093	.04488	.01422	.02926	.03949	.04330	.04735
GF	5	.00901	.01389	.01479	.01566	.01675	.00932	.01467	.01554	.01619	.01723
K,H	6	.02597	.05242	.07157	.07828	.08425	.02735	.05427	.07174	.07619	.08044
B	7	.00052	.00072	.00083	.00077	.00105	.00054	.00072	.00084	.00077	.00105
L	8	.00602	.00814	.00913	.00837	.00935	.00615	.00834	.00926	.00846	.00944
R	9	.01891	.02593	.03045	.03368	.03392	.01961	.02646	.03099	.03419	.03434
RW	10	.01890	.02285	.02356	.02344	.02313	.01950	.02349	.02415	.02387	.02349
S	11	.00603	.00920	.01233	.01358	.01343	.00630	.00970	.01296	.01423	.01409
SW	12	.00292	.00351	.00420	.00323	.00369	.00295	.00350	.00423	.00319	.00369
ZL	13	.00032	.00040	.00061	.00086	.00072	.00033	.00040	.00062	.00086	.00072
ZR	14	.00019	.00049	.00059	.00045	.00053	.00019	.00050	.00059	.00046	.00053
TSM	15	.00725	.00763	.00705	.00802	.00684	.00742	.00777	.00715	.00813	.00690
TSM+	16	.00000	.00004	.00000	.00008	.00000	.00000	.00004	.00000	.00008	.00000
W	17	.35686	.42507	.44965	.45840	.46194	.35948	.41183	.43909	.45064	.45556
P	18	.07517	.17198	.24436	.27796	.30150	.07548	.17329	.24577	.27587	.29659
CC#1	19	.20712	.27120	.30048	.31643	.32415	.21565	.28215	.31423	.33127	.33793
CH#1	20	.23330	.32247	.35985	.37805	.38574	.23924	.32809	.36821	.38670	.39391
CC#2	21	.16575	.20936	.22581	.23572	.23971	.17733	.22276	.24016	.24952	.25269
CH#2	22	.12151	.15467	.16503	.16881	.17114	.12681	.16081	.17125	.17504	.17659
TCA	23	.18167	.25949	.30247	.32417	.33517	.18611	.26635	.31173	.33369	.34407
C	24	.16527	.21647	.23999	.25465	.25946	.17222	.22534	.24774	.26221	.26520

For ease in identifying GEM's relative performance against persistence, table 4-2 displays a comparison of the two for each projection and element. A "+" indicates a Brier score favoring GEM, a "-" indicates a Brier score favoring persistence, a "0" indicates the same Brier score for both, and a blank signifies no comparison is justified. A tabulation of pluses, minuses, and ties for each projection appears at the bottom of each column with an asterisk assigned to the technique that performs best overall for each projection.

To convert the probabilistic output of GEM into categorical forecasts for each element, two techniques were used. For the 1- and 3-hr projections, the category within each element with the highest probability was selected. For the 6-, 9-, and 12-hr projections, the category which first exceeds the cumulative P* threshold was selected. The P* thresholding procedure is based on a Beta distribution integration which yields categorical forecasts in the same frequency as those observed in nature while maximizing hits.

Within the constraints of the research effort carried on thus far, this combination of techniques for converting probabilities to categorical forecasts maximizes "hits." The results are displayed in table 4-2. For each projection, GEM scores more hits than persistence. For the 1-hr projection, GEM scores more hits in forecasting ten of the elements, persistence scores more hits for two of the elements, and the two processes tie in forecasting 12 elements.

GEM equations were derived by aggregating data together from nearly 4,000,000 observations from 41 locations in the United States to generate a general climatology. To test the hypothesis of whether forecast performance versus persistence would be improved by deriving the GEM equations using individual station-adjusted climatologies, the following experiment was performed. Station-adjusted climatologies were derived for Washington, D.C., (National) and Minneapolis-St. Paul airports. Brier scores produced by forecasts which resulted from the GEM process using the station-adjusted climatologies were compared with those using the general climatology. The results for Minneapolis-St. Paul are displayed for each projection in table 4-3.

The results for Washington, D.C., are similar. For this table, a "+" signifies a better (lower) Brier score using station-adjusted climatology, while a "-" signifies a better Brier score using the generalized climatology. Use of the localized climatology improves the Brier score, but at a cost of needing to generate a separate climatology for each station for which GEM forecasts are to be made. The reader will find a more refined use of climatology in chapter 7.

Although the total improvements in tables 4-2 and 4-3 appear comparable, the actual Brier score differences in the latter comparison are generally of smaller size. Incidentally, the equations are virtually the same for all locations, whether station-adjusted or generalized climatologies are used; only a climatology constant in each equation changes, depending on whether a generalized or station-adjusted climatology is used.

Conclusion

The conclusion is that GEM produces forecasts with better Brier scores and hits than does persistence for 24 weather elements in projections for 1, 3, 6, 9, and 12 hours. Station-adjusted climatology (anomaly) equations show improved skill as was suggested by the analysis of covariance tests.

Table 4-2.--Brier score and hit comparisons between GEM and persistence. (A "+" indicates superiority for GEM, and a "-" superiority for persistence, while a "0" shows equivalence between the two procedures.)

Weather element	Brier score Projections 1	3	6	9	12	Hits Projections 1	3	6	9	12
T	+	-	-	-	+	0	+	+	+	+
DPD	+	+	+	+	+	0	+	+	+	+
V	+	+	+	+	+	+	+	+	+	+
F	+	+	+	+	+	0	+	+	+	+
GF	+	+	+	+	+	0	+	+	+	+
K,H	+	+	+	-	-	0	+	+	+	+
B	+	+	+	-	-	0	-	-	-	-
L	+	+	+	+	+	+	+	+	+	+
R	+	+	+	+	+	+	+	+	+	+
RW	+	+	+	+	+	+	+	+	+	+
S	+	+	+	+	+	-	+	+	+	+
SW	+	-	+	-	+	0	+	+	+	+
ZL	-	-	+	+	+	0	+	-	0	+
ZR	+	+	-	+	+	0	+	-	-	+
TSM	+	+	+	+	+	+	+	+	+	+
TSM+	-	+		+		+	+	+	+	+
W	+	-	-	-	-	0	-	-	-	-
P	+	+	+	-	-	0	0	0	+	+
CC#1	+	+	+	+	+	+	+	-	+	+
CH#1	+	+	+	+	+	0	-	-	+	+
CC#2	+	+	+	+	+	+	+	+	+	+
CH#2	+	+	+	+	+	+	+	-	-	+
TCA	+	+	+	+	+	+	+	+	+	+
C	+	+	+	+	+	-	+	-	-	-
+	22*	20*	20*	18*	19*	10*	20*	15*	18*	21*
0	0	0	0	0	0	12	1	1	1	0
-	2	4	3	6	4	2	3	8	5	3

Table 4-3.--Brier score comparison of GEM with station-adjusted climatology versus GEM with generalized climatology for Minneapolis-St. Paul airport. ("+" favors the former, "-" favors the latter.

Element	Projections (h)				
	1	3	6	9	12
T	+	+	+	+	+
DPD	+	+	+	+	+
V	+	+	+	+	+
F	-	-	-	-	+
GF	+	+	+	+	+
H,K	+	+	+	+	+
B	+	+	+	+	+
L	+	-	-	-	+
R	+	+	-	+	+
RW	+	+	+	+	+
S	+	-	-	+	+
SW	+	+	+	+	+
ZL	+	-	+	+	+
ZR	-	+	+	+	-
TSM	-	-	-	-	-
TSM+	+	+	+	+	+
W	+	+	+	+	+
P	-	-	-	-	-
CC#1	+	+	+	+	+
CH#1	+	+	+	+	+
CC#2	+	+	+	+	+
CH#2	+	+	+	+	+
TCA	+	+	+	+	+
CIG	+	+	+	+	+
Total +'s	20*	18*	18*	20*	19*
Total -'s	4	6	6	4	5

5. OPERATIONAL GEM SYSTEM

The original format for GEM that appeared in the National Weather Digest (see Miller, 1979b) has been greatly improved. Instead of displaying categories within which the forecast is predicted to fall, the new scheme displays data that are far more readable and which require no legend for translation. Other changes include the following:

• Temperature forecasts are expressed as a value obtained by computing a weighted average--accumulating the product between the estimated probability of temperature falling inside an interval times the midvalue of the interval. At the first projection the observed temperature is applied as the midvalue.

• Dewpoint depressions are also expressed as weighted averages--similar to those for temperature--and the final output is an estimate of the actual dewpoint temperature, which is derived by subtraction from the estimated temperature.

• Pressure is also predicted with a weighted average similar to temperature.

• Wind direction and speed are expressed in degrees and knots, respectively. The direction is derived from trigonometric considerations through U and V components and from weighting the average of these with the predicted probability. The speed is also a weighted-average estimate, similar to temperature, computed from midvalues.

• Hydrometeors such as L, R, RW, S, SW, ZL, ZR, and the elements TSM, A, and TSM+ are treated in a manner suggesting a maximum-threat consideration. More specifically, in the appendix reference is made to predicted probabilities with a predicted lowest value A and highest value B. These have arbitrarily been set to two standard deviations (pooled within group) below and above the values of μ_0 and μ_1, respectively. A and B are not allowed to lie inside the interval 0-1 except at the end points.

• Obstructions to vision are handled in a manner similar to other hydrometeors, except that A and B are determined using one standard deviation.

• Visibility and clouds are also like hydrometeors but use zero standard deviations.

The above procedures have come about from subjecting the GEM output to daily exposure to "live forecasting." Feedback has been the main motivation for the present output form of GEM. In addition, some analyses of large-sample verification, none of which has been severe enough to vitiate the further use of the verification sample, have aided in developing the present form.

GEM is capable of accommodating a variety of operational computing configurations. It was designed primarily to function at short range, with the local observation entered manually or automatically into a minicomputer such as the Data General Eclipse in an Automation of Field Operations and Services (AFOS) (see National Weather Service, 1976) environment. It has been shown to possess this capability, and an example of this kind of output is given in figure 5-1. (For this example, threshold probabilities were used with A=0 and B=1 to arrive at categorical forecasts for all elements.)

```
 GGG  EEEEE M   M
G     E     MM MM
G GGG EEE   M M M
G   G E     M   M
 GGG  EEEEE M   M
```

TECHNIQUES DEVELOPMENT LABORATORY
FOR STATION: DCA
VALID FOR 12 HOURS AFTER MAR 21, 7 LOCAL

HOUR	TT	DPD	VV	WEATHER	DDFF	PPP	C1	H1	C2	H2	TS	CIG
7	62	1	0600	R- F	1715	9976	BKN	7	OVC	10	OVC	7
10	62	1	0600	R- F	1715	9976	OVC	5	OVC	10	OVC	7
13	62	1	0600	R- F	2615	9976	OVC	5	OVC	10	OVC	7
16	62	3	0400	R- RW- F	2425	9976	OVC	5	OVC	10	OVC	7
19	57	3	0400	R- RW- F	2425	9976	OVC	5	OVC	15	OVC	7

Figure 5-1.--Example of minicomputer output of GEM.

Conversely, using a large computer, the GEM system can employ a Time Sharing Option (TSO) terminal with an assumed observational data base with call letters used in a request-reply mode where the forecast is made in real time. An example of this output is given in figure 5-2.

Another large computer version uses a batch mode. Here the observation is entered with the program. Figure 5-3 shows an example of this output. Both of these large computer versions are tied to the NOAA IBM 360/195.

The small and large computer modes of calculation differ. The minicomputer uses an additive version of GEM, while the large computer versions use a multiplicative version.

Of great promise and potentially wide interest is the capability of the operational GEM to produce its forecasts on a microcomputer or even a hand computer. It is entirely practicable for a person having knowledge of the local weather conditions to make a NO WX/WX, ceiling, or visibility forecast for any projection, in a matter of seconds, on the hand-held computer. The mode visualized here is additive, not multiplicative, and limited to the elements and projections of most concern.

At the other end of the operational spectrum, there is no technological obstacle to the implementation of a telephone system with a real-time, voice response to a specific weather inquiry, whether current or predicted, for any place, any time, and for any weather element in the local observation.

```
 GGG    EEEEE M    M
G       E     MM  MM
G  GGG  EEE   M MM M
G    G  E     M    M
 GGG    EEEEE M    M
```

TECHNIQUES DEVELOPMENT LABORATORY

FOR STATION: DCA

VALID FOR 12 HOURS AFTER MAR 21,1980 7 LOCAL

	OB	12 HOURLY FORECASTS (LST)					
	7	8	9	10	11	12	13
TEMPERATURE(F)	60	61	64	66	67	67	68
DEW POINT TP(F)	59	60	61	60	60	59	58
VSBY(100THS SM)	0600	0600	0600	0600	0600	0600	0600
FOG, ICE FOG	F	F	F	F	F	F	F
GROUND FOG							
SMOKE, HAZE							
BLOWING							
DRIZZLE							
RAIN	R-	R-	R-	R-	R-	R	R
RAIN SHOWER					RW-	RW-	RW-
SNOW, IC							
SNOW SHOWER, IP							
FREEZE DRIZZLE							
FREEZE RAIN							
THUNDERSTORM							
THUNDERSTORM+							
WIND(DDFF)	1513	1719	1719	1820	1921	2021	2121
SLP(10THS MB)	9990	9993	9999	10000	10000	9997	9995
CLOUD COVER #1	BKN	BKN	BKN	OVC	OVC	OVC	OVC
CLOUD HEIGHT #1	7	7	7	7	7	7	7
CLOUD COVER #2	OVC	OVC	OVC	CLR	CLR	CLR	CLR
CLOUD HEIGHT #2	10	10	10	160	160	160	160
TOT CLOUD COVER	OVC	OVC	OVC	OVC	OVC	OVC	OVC
CEILING 100S FT	7	7	7	7	7	7	7

		14	15	16	17	18	19
TEMPERATURE(F)		68	67	66	64	62	59
DEW POINT TP(F)		57	56	55	54	53	53
VSBY(100THS SM)		0500	0500	0400	0400	0400	0400
FOG, ICE FOG		F	F	F	F	F	F
GROUND FOG							
SMOKE, HAZE							
BLOWING							
DRIZZLE							
RAIN		R-	R-	R-	R-	R-	R-
RAIN SHOWER		RW-	RW-	RW-	RW-	RW-	RW-
SNOW, IC							
SNOW SHOWER, IP							
FREEZE DRIZZLE							
FREEZE RAIN							
THUNDERSTORM							
THUNDERSTORM+							
WIND(DDFF)		2221	2321	2321	2420	2419	2418
SLP(10THS MB)		9993	9994	9996	10000	10004	10010
CLOUD COVER #1		OVC	OVC	OVC	OVC	OVC	OVC
CLOUD HEIGHT #1		7	7	7	7	7	7
CLOUD COVER #2		CLR	CLR	CLR	CLR	CLR	CLR
CLOUD HEIGHT #2		160	160	160	160	160	160
TOT CLOUD COVER		OVC	OVC	OVC	OVC	OVC	OVC
CEILING 100S FT		7	7	7	7	7	7

Figure 5-2.--Example of TSO output of GEM.

```
  XXX   XXXXX X   X
 X      X     XX XX
 X XXX  XXX   X X X
 X   X  X     X   X
  XXX   XXXXX X   X
```

TECHNIQUES DEVELOPMENT LABORATORY

FOR STATION: DCA

VALID FOR 12 HOURS AFTER MAR 21,1980 7 LOCAL

	OB	12 HOURLY FORECASTS (LOCAL STANDARD TIME)											
	7	8	9	10	11	12	13	14	15	16	17	18	19
TEMPERATURE(F)	60	61	64	66	67	67	68	68	67	66	64	62	59
DEW POINT TP(F)	59	60	61	60	60	59	58	57	56	55	54	53	53
VSBY(100THS SM)	0600	0600	0600	0600	0600	0600	0600	0500	0500	0400	0400	0400	0400
FOG, ICE FOG	F	F	F	F	F	F	F	F	F	F	F	F	F
GROUND FOG													
SMOKE, HAZE													
BLOWING													
DRIZZLE													
RAIN	R-	R-	R-	R-	R-	R	R	R-	R-	R-	R-	R-	R-
RAIN SHOWER					RW-	RW-	RW-	RW-	RW-	RW-	RW-	RW-	RW-
SNOW, IC													
SNOW SHOWER, IP													
FREEZE DRIZZLE													
FREEZE RAIN													
THUNDERSTORM													
THUNDERSTORM+													
WIND(DDFF)	1513	1718	1719	1820	1921	2021	2121	2221	2321	2321	2420	2419	2418
SLP(10THS MB)	9990	9993	9999	10000	10000	9997	9995	9993	9994	9996	10000	10004	10010
CLOUD COVER #1	BKN	BKN	BKN	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC
CLOUD HEIGHT #1	7	7	7	7	7	7	7	7	7	7	7	7	7
CLOUD COVER #2	OVC	OVC	OVC	CLR	CLR	CLR	CLR	CLR	CLR	CLR	CLR	CLR	CLR
CLOUD HEIGHT #2	10	10	10	UNL	UNL	UNL	UNL	UNL	UNL	UNL	UNL	UNL	UNL
TOT CLOUD COVER	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC	OVC
CEILING 100S FT	7	7	7	7	7	7	7	7	7	7	7	7	7

Figure 5-3.--Example of batch output of GEM.

6. SUMMARY

Characteristics of GEM that Deserve Special Emphasis

• GEM predicts for a point in space and at an instant in time--at a weather station location and at the time of observation--which suggests an inherent limitation in the skill obtainable.

• It uses a generalized operator and can therefore be applied to any location in the conterminous United States, on any day or hour, and for any projection (1-12 hours being preferred). It has instantaneous updating capabilities for any weather element any time a surface observation is taken.

• A prediction is made of the total conditional probability distribution at every hour into the future for each element. A categorical forecast is also made for each element. This tends to maximize the number of correct forecasts while maintaining a good fit between the number of times an event is predicted and the number of times it is observed to occur over time. The probability estimates made by the regression equations in GEM occasionally lie outside the 0-1 interval. This is only an aesthetic nuisance, which is duly accounted for in the method that is used to make categorical forecasts.

• The particular GEM configuration described here can very easily be reduced in size (in the number of predictors and predictands) by merely accumulating any subset of elements, except weather like fog and rain, since they can occur simultaneously in nature. This might be required to accommodate a smaller operational forecasting instrument such as a hand held computer or calculator.

• With such a large sample used to develop GEM (nearly 4,000,000 cases), the loss in Brier score when going from a dependent sample to an independent sample should be nil.

• Renormalizing or doing "enhancements" on the probabilities after each iteration has been deemed unnecessary and at times harmful. It is best to keep the probabilities in their original form. In fact, the equivalence between the multiplicative and additive forms would not be maintained under such circumstances.

• A complete set of results has been provided in the microfiche packet in the back cover of this report for any type of interpretation or possible modification that might be desired. For example, a spectral decomposition (Eigenfunctions) could be beneficial for interpreting the results, but this kind of solution has been hard to come by for such large matrices.

• The zero-one or dummy system of variables in GEM is completely nonparametric, implying that no assumptions regarding distributional forms, such as normality, have been made nor are they required. The tests of significance have an underlying assumed form, but they are classified as being robust.

• GEM is quite capable of predicting record events, since the data base covers a broader spectrum than the history of any station in question.

Possible Areas of Research for Enhancing GEM

Data Preparation

The present GEM system of predictor-predictand variables does not include cloud types, past precipitation occurrence, ground cover, gustiness of winds, nor any type of observational remarks. Perhaps some of these would provide predictive information unaccounted for by the current set of variables. Tests have denied the existence, however, of predictive information in tendencies, through inclusion of a previous observation, or in cloud types.

Interactive boolean predictors are shown to yield otherwise unaccounted-for information in this report. Perhaps a concerted effort using a screening lattice algorithm (SLAM; see Miller, 1969) or a more exhaustive use of discrete likelihood functions (DLF; see Miller, 1979a), which accounts for all two-variable interactions automatically, can bring new information to bear--even if only to account for the nonadditivity among the present predictors. A set of boolean predictors that should yield important information is hour of the day logically "anded" with other elements that have strong diurnal variations, such as temperature and dewpoint depression. The ultimate method for uncovering interactive sources of information lies in the total enumeration of observed combinations of dummy predictors--their number being certain to be constrained to something under the size of the sample. Obviously, this is a labor-intensive undertaking, and it is not being recommended here.

Upper-air predictors, while inviting as a source of important information, are unavailable except at the two times of the day that soundings are taken. This restriction would limit the present updating capabilities which, of course, are available at any time. When automatic sounding equipment, like that being used by the Prototype Regional Observing and Forecasting Service (PROFS) Project (see Beran, 1980) in Boulder, Colo., can be initiated at any time, this logistical problem will be overcome.

Network observations are also appealing as a potential source of information, possibly in the form of gridpoint data. Interpolations of zero-one observed data would be easy to perform, since they would be like probabilities of the event occurring at the gridpoint. However, more information might be lost by divorcing the system from straight observational data. Nonetheless, the concept has produced useful hurricane forecasting equations when a moving grid is employed. (See Veigas, Miller, and Howe, 1959.) This work also substantiates a generalized-operator formulation.

Data Transformation

An enhancement of GEM would be to employ a finer specification of event categories--more zero-one variables than are currently being included, especially in time of the year. With the present large sample size, or even one that is easily made larger through the additivity features of the cross product matrices, the resolution of each weather variable can be made as fine as desired. For those who believe that zero-one predictors fail to capture all that a corresponding continuous variable might offer, this feature should dispel that fear entirely. In fact, the ability of the regression coefficients to fit the individual zero-one pieces of the original variable gives it nonlinear capabilties that are not available in the continuous variable, unless the precise nonlinear form is specified a priori.

One type of seemingly important transformation to perform is a weighted regression. For example, 1) a variance stabilization with the ARCSIN, 2) a 0-1 constrainer with the logistic, 3) a standardization with beta coefficients, 4) a spectral decomposition with eigenfunctions, or 5) a normalization transformation. Cox (1970) has pointed out that when the predictors and predictands are all zero-one binary variables, like those employed here, the process of solving for such a weighted regression is simple to perform. Using Cox's approach, however, all efforts have uncovered nothing useful over that achieved by straight unweighted regression. The failure seems to be in overweighting the tails of the element's distribution.

Computational

Of the two mathematical versions of GEM--multiplicative and additive--the context of its use would dictate the proper mode to employ. If the computer is limited in the space available, then storing one matrix to perform an iterative solution is advantageous. Should speed be the primary consideration, then an additive version is recommended. For such a configuration, the coefficient matrix must be powered to as many iterations as may be desired. This solution requires that only the predictors in the observation that are unity need to have their respective coefficients added together. In an integer form this procedure can be made extremely fast. In contrast, the multiplicative (iterative) version cannot be so conveniently dealt with, since the form of computation would most likely need to be in floating point.

Statistical Analysis

Variations on the time steps in GEM should be tried. The 1-hr step used here could give way to 3, 12, 24, or even more hours, depending upon the application. Certainly a longer-range forecast system applying the GEM principle would be inefficient if performed hour by hour for situations where time and space averagings were desired.

For certain computing facilities it might be wiser to abandon the principle in GEM of using time-step iterations. Certainly a direct projection to particular hours would have to yield improved results, since the Brier score is minimized at those projections, not just in the first hour as in GEM.

The screening of predictors, for efficiency reasons, has been attempted in GEM. It suffers from the fact that time information is forsaken in the selection process. This causes the elimination from the GEM forecasts of many interesting and useful characteristics, such as manifestation of diurnal variations, deviations from persistence, onset and duration of weather, frontal passages, and discontinuities. Perhaps forcing time elements into the equation while screening would solve this problem.

Other multivariate statistical models may prove to be more powerful than regression. Canonical correlation, discriminant analysis, discrete likelihood functions, or a distance-neighborhood framework might enhance the technique. The simple elegance of the present model would require a substantial improving upon to be supplanted.

One area that has latitude for improvements is the application of mathematical programming methods--geometric, stochastic, integer, pseudo-boolean, and dynamic. In particular, a derivation of the appropriate utility function would permit a Bayesian solution of the probability-to-categorical forecasting problem under constraints of any type. The need for such a solution is evident from the consistent superiority of GEM's Brier score but with less success on hits. The predictive skill is evident but not fully captured.

Finally, an effort toward a quantitative-precipitation forecast should be attempted, using an expected amount over time based on the intensity of the type and its forecasted probability.

Output

The variety of output forms of GEM seems to be unlimited. The user's requirements would dictate the form. As guidance to the local forecaster, several versions are obvious. The array of hourly forecasted probability distributions for each element, called GEMTRIX, reflects the conditional climatology given the current observation. This gives the forecaster a quantitative measure of the risk he would be taking in his own "final" forecast should he or she deviate from GEM.

An interesting form of guidance output would be to plot and analyze (manually or automatically) the hourly categorical forecasts made by GEM in, say, a sectional map. The analysis could be based on either one element at one forecast time or on all elements taken jointly at all times in a kind of time lapse. The forecaster could superpose the immediate radar echoes to help resolve the important issue of timing the onset or offset of hydrometeors, frontal passages, squall lines, and the like. A future refinement could be the depicting of the previous or most recent error fields as a feedback source. Initially this might best be done subjectively.

Another application of graphical depiction would be to infer the climatology of stations not in the inventory for implementing station-adjusted climatology (anomaly) equations, since the anomaly equations have been shown to be more skillful than straight generalized operators.

An important use of GEM would be in monitoring and updating automatically in a minicomputer whenever a new record or special observation is received for a particular location. (See Vercelli and Heffernan, 1978.) Automated observing equipment could play an important rôle here. This is made possible by the real-time capabilities of the GEM model.

A future form of GEM would be its merging with other forecasts in an objective way. Ultimately it should be combined with all that is available--the human forecaster with his experience, MOS with its organization of dynamic model output, radar with its capacity to reflect immediate areal occurrence of precipitation, and satellite information with its timely and wide coverage of certain atmospheric events. A variety of models exist for such a blend, but statistical regression methods will probably be the most effective.

Variations in the form of input and output are also in need of testing. Perhaps fractional times (less than hourly time steps) would be of value in such critical situations as the landing or taking off of aircraft, or in military operations. A possible solution is the eigenfunction version of GEM. The types of short-period observing performed by PROFS and the Federal Aviation Administration (FAA) would make a good starting place. Another variation to test would be to input the observations as probabilities (Unger, 1980), depending upon an observed value's relationship to the interval in which it falls. This suggests a source of "free" information available for the taking.

GEM comes already equipped with a "what if" capability. This could increase our understanding of the forecasting problem if not further our understanding of the atmosphere.

It does not require much imagination to foresee the potential applications of GEM as a procedure for making on-demand telephone forecasts for any location in the observational data base. Furthermore, the many home computers now on the market or already in use are ideally suitable for this weather forecasting capability. Cable TV seems to be a natural form of output.

Finally, PROFS and the FAA are planning to use a GEM model, while the AWS (Kelly, 1978) and Air Force Geophysics Laboratory (AFGL) (Geisler, 1979) have already done work on a single-station GEM-like procedure. In the PROFS application, numerous other weather elements are being considered over those in the usual surface observation. In particular, soundings of the temperature, humidity, and wind conditions will be introduced from automated observing equipment at very short time intervals. The FAA also intends to use short-period automatic instrument readings at airfield locations. Data with such high frequencies can be accumulated very rapidly to expedite the implementation of GEM for the purposes desired. Systems such as the Automation of Field Operations and Services (AFOS), Automated Weather Distribution System (AWDS), Naval Environmental Display System (NEDS), Modular Automated Weather System (MAWS), Army field installations, ships at sea, and a standard telephone can quite easily make use of a GEM system for automatic forecasts or for monitoring official forecasts needing revision based on a recent observation. Developing countries might well find GEM inexpensive and easy to implement as a basic forecasting system.

7. NEW RESULTS

Improving the Model

Until now, the Markov process modeled by GEM has accommodated changes only at discrete times. Led partly by empirical evidence and by the appreciation that weather changes can occur at any time, GEM has now been altered to model a continuous-time Markov process. Feller (1950) discusses the change required in a model to switch from discrete time to continuous time--namely, from a geometric to an exponential representation. Howard (1960) gives all of the necessary details for accommodating changes over continuous time.

Specifically, the discrete-time representation of a Markov chain, predicting the probability vector $\underline{\Pi}$ at time t with $\underline{P}$ as the transition probability matrix, is:

$$\underline{\Pi}(t) = \underline{\Pi}(0)\underline{P}^t \qquad (t=0,1,\ldots) \tag{7-1}$$

which is from the recursion of $\underline{\Pi}(t+1) = \underline{\Pi}(t)\underline{P}$, t=0, 1, ... In the GEM context (7-1) can be represented equivalently as

$$\underline{\Pi}(t) = \underline{\Pi}(0)\underline{A}^t \qquad (t=0,1,\ldots) \tag{7-2}$$

where $\underline{A}$ is the transition-rate matrix of multiple regression equations.

In the continuous-time case, the difference equations underlying (7-1) and (7-2) give way to a set of differential equations underlying

$$\frac{d}{dt}\underline{\Pi}(t) = \underline{\Pi}(t)\underline{A} \tag{7-3}$$

Integrating (7-3) yields

$$\underline{\Pi}(t) = \underline{\Pi}(0)e^{\underline{A}t} \tag{7-4}$$

Equation (7-4) can be written in exponential-series form as

$$\underline{\Pi}(t) = \underline{\Pi}(0)\left[\underline{I} + t\underline{A} + \frac{t^2}{2!}\underline{A}^2 + \frac{t^3}{3!}\underline{A}^3 + \ldots\right] \tag{7-5}$$

where $\underline{I}$ is the identity matrix. For any given t the relationship in (7-5) imposes a set of weights onto the powers of $\underline{A}$. Observe that when t=1 there is an alteration made to the straight application of the regression equations in $\underline{A}$. Since these equations represent the best-linear-unbiased estimates that yield minimum residual variance, based on the least squares principle, a boundary condition will be set to maintain the use of an unweighted $\underline{A}$ at t=1. That is, the model to accomplish this is

$$\underline{\Pi}(1) = \underline{\Pi}(0)\underline{A} \tag{7-6}$$

$$\underline{\Pi}(t) = \underline{\Pi}(0)e^{\underline{A}t} \qquad t > 1$$

Empirical evidence has shown this model is to be preferred to (7-4) or to one that begins dampening after the first hour, such as $\Pi(t) = \Pi(0)\underline{A}e^{\underline{A}(t-1)}$, where $t \geq 1$.

A table of normalized weights, which sum to unity, is given in table 7-1 for t=2 ..., 12 and for powers of $\underline{A}$ from 1 to 24. Note that the crest of this set of weights appears around the power of $\underline{A}$ that corresponds to the projection time.

Table 7-1.--Normalized weights for exponential GEM model for t=2,...,12 and from 1 to 24 powers of $\underline{A}$.

VALUES FOR TIME 2

1 .13534D+00	2 .27067D+00	3 .27067D+00	4 .18045D+00	5 .90224D-01
6 .36089D-01	7 .12030D-01	8 .34371D-02	9 .85927D-03	10 .19095D-03
11 .38190D-04	12 .69436D-05	13 .11573D-05	14 .17804D-06	15 .25434D-07
16 .33913D-08	17 .42391D-09	18 .49872D-10	19 .55413D-11	20 .58329D-12
21 .58329D-13	22 .55552D-14	23 .50502D-15	24 .43914D-16	

VALUES FOR TIME 3

1 .49787D-01	2 .14936D+00	3 .22404D+00	4 .22404D+00	5 .16803D+00
6 .10082D+00	7 .50409D-01	8 .21604D-01	9 .81015D-02	10 .27005D-02
11 .81015D-03	12 .22095D-03	13 .55238D-04	14 .12747D-04	15 .27315D-05
16 .54631D-06	17 .10243D-06	18 .18076D-07	19 .30127D-08	20 .47569D-09
21 .71354D-10	22 .10193D-10	23 .13900D-11	24 .18131D-12	

VALUES FOR TIME 4

1 .18316D-01	2 .73263D-01	3 .14653D+00	4 .19537D+00	5 .19537D+00
6 .15629D+00	7 .10420D+00	8 .59540D-01	9 .29770D-01	10 .13231D-01
11 .52925D-02	12 .19245D-02	13 .64151D-03	14 .19739D-03	15 .56397D-04
16 .15039D-04	17 .37598D-05	18 .88465D-06	19 .19659D-06	20 .41387D-07
21 .82775D-08	22 .15767D-08	23 .28667D-09	24 .49855D-10	

VALUES FOR TIME 5

1 .67379D-02	2 .33690D-01	3 .84224D-01	4 .14037D+00	5 .17547D+00
6 .17547D+00	7 .14622D+00	8 .10444D+00	9 .65278D-01	10 .36266D-01
11 .18133D-01	12 .82422D-02	13 .34342D-02	14 .13209D-02	15 .47174D-03
16 .15725D-03	17 .49139D-04	18 .14453D-04	19 .40146D-05	20 .10565D-05
21 .26412D-06	22 .62886D-07	23 .14292D-07	24 .31070D-08	

VALUES FOR TIME 6

1 .24788D-02	2 .14873D-01	3 .44618D-01	4 .89235D-01	5 .13385D+00
6 .16062D+00	7 .16062D+00	8 .13768D+00	9 .10326D+00	10 .68838D-01
11 .41303D-01	12 .22529D-01	13 .11264D-01	14 .51990D-02	15 .22281D-02
16 .89126D-03	17 .33422D-03	18 .11796D-03	19 .39320D-04	20 .12417D-04
21 .37251D-05	22 .10643D-05	23 .29026D-06	24 .75721D-07	

Table 7-1.--(continued)

VALUES FOR TIME 7

1 .91188D-03	2 .63832D-02	3 .22341D-01	4 .52129D-01	5 .91226D-01
6 .12772D+00	7 .14900D+00	8 .14900D+00	9 .13038D+00	10 .10140D+00
11 .70983D-01	12 .45171D-01	13 .26350D-01	14 .14188D-01	15 .70942D-02
16 .33106D-02	17 .14484D-02	18 .59640D-03	19 .23193D-03	20 .85449D-04
21 .29907D-04	22 .99690D-05	23 .31720D-05	24 .96538D-06	

VALUES FOR TIME 8

1 .33546D-03	2 .26837D-02	3 .10735D-01	4 .28626D-01	5 .57252D-01
6 .91604D-01	7 .12214D+00	8 .13959D+00	9 .13959D+00	10 .12408D+00
11 .99262D-01	12 .72190D-01	13 .48127D-01	14 .29617D-01	15 .16924D-01
16 .90260D-02	17 .45130D-02	18 .21238D-02	19 .94389D-03	20 .39743D-03
21 .15897D-03	22 .60561D-04	23 .22022D-04	24 .76598D-05	

VALUES FOR TIME 9

1 .12341D-03	2 .11107D-02	3 .49981D-02	4 .14994D-01	5 .33737D-01
6 .60727D-01	7 .91091D-01	8 .11712D+00	9 .13176D+00	10 .13176D+00
11 .11858D+00	12 .97021D-01	13 .72766D-01	14 .50376D-01	15 .32385D-01
16 .19431D-01	17 .10930D-01	18 .57864D-02	19 .28932D-02	20 .13705D-02
21 .61671D-03	22 .26430D-03	23 .10812D-03	24 .42309D-04	

VALUES FOR TIME 10

1 .45402D-04	2 .45402D-03	3 .22701D-02	4 .75670D-02	5 .18918D-01
6 .37835D-01	7 .63058D-01	8 .90083D-01	9 .11260D+00	10 .12512D+00
11 .12512D+00	12 .11374D+00	13 .94785D-01	14 .72911D-01	15 .52080D-01
16 .34720D-01	17 .21700D-01	18 .12765D-01	19 .70914D-02	20 .37323D-02
21 .18662D-02	22 .88865D-03	23 .40393D-03	24 .17562D-03	

VALUES FOR TIME 11

1 .16705D-04	2 .18376D-03	3 .10107D-02	4 .37057D-02	5 .10191D-01
6 .22420D-01	7 .41103D-01	8 .64590D-01	9 .88811D-01	10 .10855D+00
11 .11940D+00	12 .11940D+00	13 .10945D+00	14 .92613D-01	15 .72767D-01
16 .53363D-01	17 .36687D-01	18 .23739D-01	19 .14507D-01	20 .83987D-02
21 .46193D-02	22 .24196D-02	23 .12098D-02	24 .57861D-03	

VALUES FOR TIME 12

1 .61484D-05	2 .73781D-04	3 .44269D-03	4 .17707D-02	5 .53122D-02
6 .12749D-01	7 .25499D-01	8 .43712D-01	9 .65568D-01	10 .87424D-01
11 .10491D+00	12 .11445D+00	13 .11445D+00	14 .10564D+00	15 .90551D-01
16 .72441D-01	17 .54331D-01	18 .38351D-01	19 .25567D-01	20 .16148D-01
21 .96887D-02	22 .55364D-02	23 .30198D-02	24 .15756D-02	

The consequence of employing (7-6) in contrast to (7-2) will now be demonstrated in an illustrative example.

Given:

- Predictands

Y_1	Total cloud cover clear	○
Y_2	Total cloud cover scattered	⦶
Y_3	Total cloud cover broken	⦷
Y_4	Total cloud cover overcast	⊕

- Predictors

X_1	Total cloud cover clear	○
X_2	Total cloud cover scattered	⦶
X_3	Total cloud cover broken	⦷
X_4	Total cloud cover overcast	⊕

- Location Washington, D.C. (DCA)

- Data (same sample as employed in GEM test)

		t_0 ○	⦶	⦷	⊕	Total
t_{+1}	○	19133	3166	267	63	22629
	⦶	2894	10983	3490	805	18172
	⦷	508	3343	7840	3316	15007
	⊕	94	679	3409	27556	31738
Total		22629	18171	15006	31740	87546

- Transition probability matrix $\underline{P}$

		t_0 ○	⦶	⦷	⊕
t_{+1}	○	.84551	.17423	.01779	.00198
	⦶	.12789	.60442	.23257	.02536
	⦷	.02245	.18397	.52246	.10447
	⊕	.00415	.03737	.22718	.86818

- Regression equations (omitting ⊕ as redundant)

$$\hat{Y}_1 = .00198 + .84352\ X_1 + .17225\ X_2 + .01581\ X_3$$
$$\hat{Y}_2 = .02536 + .10253\ X_1 + .57906\ X_2 + .20721\ X_3$$
$$\hat{Y}_3 = .10442 - .08199\ X_1 + .07953\ X_2 + .41802\ X_3$$

• Comparing the two models, under the separate initial conditions of being clear, scattered, broken, or overcast at a 3-hr projection, gives:

	○	⦶	⦷	⊕
Model $\underline{\Pi}(3) = \underline{\Pi}(0)\underline{P}^3$	.65787	.34595	.26265	.71224
Model $\underline{\Pi}(3) = \underline{\Pi}(0)e^{\underline{A}\cdot 3}$	.68532	.41254	.33764	.73472
Actual	.68651	.39494	.32891	.76654

Thus, in each instance the exponential model improved upon the geometric model for total cloud cover at DCA for a 3-hour projection. A similar study at DCA was conducted for 21 categories of wind at 3, 6, 9, and 12 hours. The same comparative results were obtained. In fact, a full-scale verification on the 26,328 sample described in chapter 4 yielded a convincing improvement by the exponential model over the geometric model, in Brier scores and hits, comparing all weather elements at all projections--excluding the 1-hr projection, where the forecasts are equivalent. These results are presented in table 7-2. It must be pointed out that a direct method of forecasting (noniterative) would yield the exact answer; however, it does require separate equations for the desired projections.

Furthermore, employing the continuous-time version of GEM permits predictions to be made for any time into the future beyond the first hour. For example, should a need arise for a 2 1/2-hr forecast, say for a takeoff or landing of an aircraft, such a requirement can be met very easily. No longer is it required to predict in whole-hourly units.

Because of these improved results, henceforth the model's acronym will stand for Generalized Exponential Markov.

Including Local-Hourly Climatology

Among the predictors used in GEM's regression equations is the hour of the day. Any diurnal variation in the aggregated sample of 41 stations is duly accounted for. However, individual station data possessing diurnal variation, different from the aggregate, might not be accounted for. Evidence from the analysis of covariance indicates that single-station analyses were not sufficiently statistically significant to warrant their use. This judgment, however, was made with regard to utilizing all predictors. Further evidence, primarily from the verification, suggests that individual station hourly climatological effects are significant. Meteorological reasoning also contributes to this surmise.

Fortunately, the inclusion of local-hourly climatology fits into the GEM model very conveniently when viewed in the following manner. Using (7-6) the model can be partitioned as

$$\underline{\Pi}(t) = \underline{\Pi}(0)\ [\underline{S} + \underline{T}(t)] \qquad (7\text{-}7)$$

Table 7-2.--Comparison of Brier scores and hits between exponential GEM and geometric GEM. A "+" favors the exponential, while a "-" favors the geometric. A "0" indicates a tie. Hour 1 is not compared, because the two models are equivalent for that projection.

Weather element	Brier score Projections 3	6	9	12	Hits Projections 3	6	9	12
T	-	-	-	-	+	+	+	+
DPD	-	-	-	-	-	-	-	-
V	+	+	+	+	-	-	+	+
F	+	+	+	-	+	+	+	+
GF	+	+	-	-	+	+	+	+
K,H	+	+	+	+	+	+	-	+
B	+	+	+	+	0	+	+	+
L	+	+	+	+	+	+	+	+
R	+	+	+	+	+	+	+	+
RW	+	+	+	+	+	+	-	+
S	+	+	+	+	+	+	+	+
SW	+	+	+	+	+	+	+	+
ZL	-	-	+	+	0	+	+	-
ZR	+	-	+	+	0	+	+	-
TSM	-	-	+	+	+	-	+	+
TSM+	+	-	+	+	+	-	0	0
W	+	+	+	+	+	+	+	+
P	+	+	+	+	+	+	+	-
CC#1	+	+	+	-	-	+	-	-
CH#1	+	-	+	-	+	+	+	-
CC#2	+	+	-	-	-	+	+	+
CH#2	+	+	+	-	+	+	+	+
TCA	+	+	+	+	-	-	-	+
C	+	+	+	+	+	+	+	+
+	20*	17*	20*	16*	16*	19*	18*	17*
0	0	0	0	0	3	0	1	1
-	4	7	4	8	5	5	5	6

where $\underline{S}$ is the steady state component and $\underline{T}(t)$ is the transient component of the Markov process. $\underline{S}$ is a stochastic matrix whose elements are non-negative and whose rows sum to unity and $\underline{T}(t)$ are differential matrices whose rows sum to zero. In this new context, local-hourly climatology is treated in $\underline{S}$, while $e^{\underline{A}t}$, $t>1$, and $\underline{A}$, $t=1$, are treated in $\underline{T}(t)$.

A comparative test of this new concept yields results that are superior to the original geometric form of GEM, for essentially all variables and all projections in the Brier score over the verification sample.

A final comparative test incorporating the exponential weighting and local-hourly climatology against persistence is shown in table 7-3. These results demonstrate GEM's superiority in 117 of the 120 comparisons and with an average improvement of 5 percent in the Brier score, despite the fact that persistence Brier scores from 3 to 12 hours are computed using the independent-sample conditional probabilities.

Table 7-3.--Brier score comparison between GEM, with exponential decay and local-hourly climatology, and persistence for the sample in table 4-1.

		BRIER SCORE									
		GEM					PERSISTENCE				
Weather element		1 hr.	3	6	9	12	1 hr.	3	6	9	12
T	1	.22684	.35097	.39826	.41197	.41519	.22884	.35524	.40724	.42397	.42948
DPD	2	.27253	.35554	.38323	.39089	.39418	.27953	.37361	.41315	.42427	.42727
V	3	.08184	.10709	.12199	.12712	.13189	.08379	.11187	.12951	.13458	.13874
F	4	.01297	.02599	.03586	.03963	.04337	.01422	.02926	.03949	.04330	.04735
GF	5	.00894	.01360	.01453	.01541	.01654	.00932	.01467	.01554	.01619	.01723
K,H	6	.02535	.04924	.06412	.06898	.07373	.02735	.05427	.07174	.07619	.08044
B	7	.00052	.00071	.00082	.00076	.00104	.00054	.00072	.00084	.00077	.00105
L	8	.00601	.00805	.00903	.00830	.00928	.00615	.00834	.00926	.00846	.00944
R	9	.01890	.02554	.03006	.03337	.03359	.01961	.02646	.03099	.03419	.03434
RW	10	.01888	.02269	.02346	.02334	.02306	.01950	.02349	.02415	.02387	.02349
S	11	.00603	.00920	.01227	.01351	.01335	.00630	.00970	.01296	.01423	.01409
SW	12	.00291	.00348	.00415	.00317	.00364	.00295	.00350	.00423	.00319	.00369
ZL	13	.00032	.00041	.00064	.00086	.00071	.00033	.00040	.00062	.00086	.00072
ZR	14	.00019	.00049	.00059	.00045	.00053	.00019	.00050	.00059	.00046	.00053
TSM	15	.00722	.00764	.00699	.00795	.00681	.00742	.00777	.00715	.00813	.00690
TSM+	16	.00000	.00004	.00000	.00008	.00000	.00000	.00004	.00000	.00008	.00000
W	17	.35324	.40537	.43342	.44254	.44592	.35948	.41183	.43909	.45064	.45556
P	18	.07501	.17094	.24254	.27499	.29792	.07548	.17329	.24577	.27587	.29659
CC#1	19	.20386	.26169	.28726	.30209	.30946	.21565	.28215	.31423	.33127	.33793
CH#1	20	.23088	.31251	.34725	.36421	.37230	.23924	.32809	.36821	.38670	.39391
CC#2	21	.16348	.20062	.21518	.22453	.22846	.17733	.22276	.24016	.24952	.25269
CH#2	22	.12070	.15056	.16105	.16518	.16738	.12681	.16081	.17125	.17504	.17659
TCA	23	.18004	.25285	.29285	.31312	.32395	.18611	.26635	.31173	.33369	.34407
C	24	.16453	.21382	.23577	.25028	.25517	.17222	.22534	.24774	.26221	.26520

ACKNOWLEDGMENTS

Many individuals assisted in this effort. I wish to thank especially the current team of Albert Forst, William Griner, William Herrmann, and Thomas Perrone, without whose efforts this work on GEM would not have been accomplished. Help was also received from Frank Lewis and David Vercelli. Thanks are given also to the AWS and SLU team of Majors Michael Kelly and Roger Whiton, who worked out many of the details and computer programs.

Fundamental ideas were provided by a former colleague and a pioneer in statistical weather prediction, Joseph Bryan.

My sincere appreciation to Larry Best and Jack Jalickee for some helpful ideas.

A very special thank-you to Evelyn Liddel, whose excellent editing and typing brought the report to life.

BIBLIOGRAPHY

Beran, D. W.. 1980: PROFS Proposed policy, mission, goals, and organizational structure. NOAA Technical Memorandum, ERL PROFS-2, Environmental Research Laboratories, NOAA, U.S. Department of Commerce, 49 pp.

Brier, G. W., 1950: Verification of forecasts in terms of probability. Mon. Wea. Rev., 78, 1-3.

Cox, D. R., 1970: Analysis of Binary Data. Chapman and Hall, London, 142 pp.

Crisci, R. L., and F. Lewis, 1973: Automated probability forecasts of ceiling and visibility based on single-station data. Nat. Wea. Service Report No. FAA-RD-73-13, Silver Spring, Md., 68 pp.

Crout, P. D., 1941: A short method for evaluating determinants and solving systems of linear equations with real or 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.

Geisler, E. B., 1979: Development and evaluation of a tower slant visual range system. Air Force Geophysics Laboratory, Instrumentation Paper No. 281, AFGL-TR-79-0209, 42 pp.

Glahn, H.R., and D. A. Lowry, 1972: The use of Model Output Statistics (MOS) in objective weather forecasting. J. Appl. Meteor., 11, 1203-1211.

Harris, R., 1962: Personal communication.

Howard, R., 1960: Dynamic Programming and Markov Processes. The Technology Press, MIT, John Wiley and Sons, New York, 136 pp.

Kelly, M. J., 1978: Development of single station and area statistical short-range forecast technique. Unpublished manuscript.

Miller, R. G., 1964: Regression estimation of event probabilities. Travelers Research Center Tech. Report, 7411-121, Contract Cwb 10704, Hartford, Conn., 153 pp.

______1968: A stochastic model for real-time on-demand weather prediction. Proceedings, First Statistical Meteorological Conference, Amer. Meteor. Soc., 48-51.

______1969: SLAM: A screening lattice algorithm for non-linear regression estimation of event probabilties, Travelers Research Center, Inc., Hartford, Conn.,23 pp.

______1979a: Estimating event probabilities by discrete likelihood functions. Proceedings, Sixth Conference on Probability and Statistics in Atmospheric Sciences, Amer. Meteor. Soc., 93-97.

______1979b: Conditional climatology given the local observation. Nat. Wea. Digest, 4, 2-15.

______, R. C. Whiton, and M. J. Kelly, Jr., 1977: Results of a single station forecasting experiment. Proceedings, Fifth Conference on Probability and Statistics in Atmospheric Sciences, Amer. Meteor. Soc., 37-40.

Mood, A. M., 1950: Introduction to the Theory of Statistics. McGraw-Hill, New York, 433 pp.

National Weather Service, 1976: Program development plan. Automation of field operations and services (AFOS). NOAA Report 76120603, 129 pp.

Shorr, B., 1958: Studies in statistical weather prediction. Travelers Ins. Co. Final Report, Contract No. AF19(604)-1590, Hartford, Conn., 238 pp.

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

Unger, D., 1980: Personal communication.

Veigas, K., R. G. Miller, and G. Howe, 1959: Probabilistic prediction of hurricane movements by synoptic climatology. Occasional Papers in Meteorology No. 2, Travelers Ins. Co., Hartford, Conn., 54 pp.

Vercelli, D. J., and M. M. Heffernan, 1978: AFOS aviation local monitoring and updating program. Preprints Conference on Weather Forecasting and Analysis and Aviation Meteorolgoy, Amer. Meteor. Soc., 369-373.

Whiton, R. C., 1977: Markov processes. Selected Topics on Statistical Meteorology, Ed., R. G. Miller, Air Weather Service Tech. Report AWS-TR-77-273, Scott AFB, Ill., 45 pp.

Wiener, N., 1948: Cybernetics. The Technology Press, MIT, John Wiley and Sons, New York, 194 pp.

______, 1950: Extrapolation, Interpolation, and Smoothing of Stationary Time Series. The Technology Press, MIT, John Wiley and Sons, New York, 163 pp.

______, 1956: Nonlinear prediction and dynamics. Proceedings of the Third Berkeley Symposium, Ed., J. Neyman, University of California Press, Berkeley, 247-252.

GLOSSARY OF TERMS

Additive: Requires only simple additions to obtain a solution.

AFB: Air Force Base

AFGL: Air Force Geophysics Laboratory

AFOS: Automation of Field Operations and Services

Analysis of covariance: R. A. Fisher's statistical method for testing treatment effects, taking into account concomitant variables through regression

Analysis of variance: R. A. Fisher's statistical method for testing treatment effects

Anding: Boolean operation where the resultant is a one only if both conditions are ones; otherwise it is zero.

Anomaly: A condition in which the mean--or climatology--has been removed from the original observations

AWDS: Automated Weather Distribution System

AWS: Air Weather Service

Bayes Solution: A decision-theoretic principle of minimizing risk or maximizing expected gain

Bias: Systematic distortion over a sample

Binary: Having only the value zero or one

Blending: Bringing together two or more predictions superior to any single prediction

Booleans: An interactive variable created by a logical operation of Boolean algebra

Brier score: A verification score for probability forecasts where

$$BS = \sum_{i=1}^{N} \sum_{g=1}^{G} (\hat{P}_{ig} - \Theta_{ig})^2/2N$$

$\hat{P}_{ig}$ is the predicted probability, Θ_{ig} is a one or zero, depending upon whether the event occurred or not, and where there are G categories and a sample of N. Actually 1/2 the original score defined by Brier.

Canonical correlation: A multivariate statistical method applied to two sets of variables

Categorical: An unambiguous choice of predicted weather-element category

Continuous variable: An ordered variable on a scale, in contrast to a discrete variable

CPU: Central processing unit

DCA: Washington, D.C.

Degrees of freedom: Parameters of the F distribution

Direct: A type of forecast that attempts to predict for a specific projection, in contrast to one that is obtained by iterating shorter-time projections

Discriminant analysis: A multivariate statistical method in which consideration is given to groups of data conditioned on the predictand

Distance neighborhood: A property of closeness in a Euclidean space

DLF: Discrete likelihood functions

Dummy variable: Having either the value zero or one in all observations

ECLIPSE: A minicomputer (made by Data General), which is an integral part of AFOS

Eigenfunction: The mathematical operation of decomposition into orthogonal components

FAA: Federal Aviation Administration

GEM: Generalized equivalent Markov--more recently, generalized exponential Markov

GEM-like: Other than a pure GEM procedure. Usually not generalized but based on the Markov assumption and capable of iteration

GEMTRIX: Matrix of hourly GEM-forecast probabilities of each weather-element category

Generalized operator: A fixed set of equations applicable anywhere

GMT: Greenwich mean time

Gross predictors: A simple Boolean interactive variable between two coarsely defined weather conditions

Hits: Number of correct forecasts

Interactive: A joint condition among two or more variables

Left out dummy: In categorizing a weather element into G categories, there is always one of the G that is redundant, since if all of the others are off, the left-out one must be on.

LST: Local standard time

Map form: Data arrayed where all observed elements for one particular time are together

Markov process: A stochastic process that uses only knowledge of the present state and nothing from any prior state

MAWS: Modular automated weather system

MIA: Miami, Florida

MIT: Massachusetts Institute of Technology

Models I, II, III: Models underlying the analysis of covariance

MOS: Model output statistics

MSP: Minneapolis-St. Paul, Minnesota

MSY: New Orleans, Louisiana

Multiplicative: Requiring multiplication operations to obtain a solution

Multivariate regression: Linear regression where the number of dependent variables regressed on a fixed set of independent variables exceeds one

NEDS: Naval Environmental Display System

NMC: National Meteorological Center

Nonadditivity: The principle that prevents the simple summing of two effects because of synergism

NWS: National Weather Service

OB: Observation

PERSIS: Persistence

PHL: Philadelphia, Pennsylvania

PIREP: Pilot report

PLODITE: Putting left out dummy in the equation

Predictand: A variable for which a forecast is made

Predictor: A variable used to make a forecast

PROFS: Prototype regional observing and forecasting service

REEP: Regression estimation of event probabilities

Renormalizing: Creating a situation where the sum of a set of numbers is made to be unity

Runs: The number of times in a binary string there is a switch from 0 to one or vice versa

Screening: A procedure which chooses a subset of predictors from a larger set

Serial correlation: The property that sequential observations are usually related to one another and are therefore not independent observations

SFO: San Francisco, California

Single station: A statistical operator based on only data from a certain location or station

SLAM: Screening lattice algorithm

SLC: Salt Lake City, Utah

SLU: St. Louis University

Spectral decomposition: A mathematical technique for arriving at orthogonal components

Station-adjusted climatology: The procedure of superimposing the local climatology on an otherwise generalized operator

Stratification: Grouping of data usually under some antecedent condition such as season

TDL: Techniques Development Laboratory

Threat: A verification scoring system that is defined as $H/(F+\Theta-H)$ where H is the number of hits, F is the number of forecasts, and Θ is the number of observed cases

Threshold: A probability value that, if exceeded by the forecast probability, would initiate a categorical forecast of the event

TSO: Time sharing option

TRC: Travelers Research Center

Vector form: Data arrayed where the same weather element appears over all observations

WBAN: Weather Bureau-Air Force-Navy observation form

GLOSSARY OF SYMBOLS

A	Extended limit below 0.0 in beta distribution; or hail
$\underline{A}$	Matrix of generalized operator regression coefficients one hour hence
$\underline{Aa}$	Matrix of anomaly regression coefficients for predicting one hour hence
B	Extended limit above 1.0 in beta distribution; or blowing weather condition
$\underline{Ba}$	Matrix of anomaly regression coefficients in $\underline{Aa}$ transformed to PLODITE form
$\underline{B}$	Matrix of regression coefficients in $\underline{A}$ transformed to PLODITE form
B_{iy}	Element i of $\underline{B}$ matrix for predictand Y
$\underline{\beta}$	Matrix of beta coefficients generated from $\underline{B}$ matrix
β	Beta coefficient in regression analysis; or beta distribution
BS	Brier score
C	Ceiling
CC#1	Lowest cloud cover
CC#2	Second cloud cover
CH #1	Lowest cloud height
CH #2	Second cloud height
DPD	Dew point depression
ε^2	Sum of squares of forecast errors
f	Factor for determining the number of independent observations
F	Computed F statistic; or fog
F_{crit}	Critical F value
F_η	Test statistic for Model II in the analysis of covariance
F_μ	Test statistic for Model III in the analysis of covariance
GF	Ground fog
Γ	Gamma function
H,K	Haze, smoke, dust, or any combination of these

I	Denotes station which was part of the analysis of variance and covariance tests
K	Number of stations in sample
L	Drizzle
L_k	Station k
μ_0	Mean of $\hat{Y}$ when event did not occur
μ_1	Mean of $\hat{Y}$ when event occurred
n	Estimated number of independent observations in a sample based on considering serial correlation
N	Total sample size
N_k	Sample size from station k
ν	Degrees of freedom
NO WX	No hydrometeors
Θ	Observation (0 is event not observed, 1 if event observed)
P*	Threshold probability
p	Predictor index
P	Total number of predictors; or pressure
$\underline{\Pi}(t)$	A probability vector at time t
q	Predictand index
Q	Total number of predictands
r	Number of runs
R	Rain
R^2	Correlation coefficient squared
RW	Rain showers
S	Snow
$\underline{S}$	Steady-state component in GEM
SSEX	Sum of squares explained
SSR	Sum of squares residual or within
SST	Sum of squares total

SSW	Sum of squares within or residual
SW	Snow showers
σ	Standard deviation
Σ	Summation
T	Temperature; or matrix power when superscript
TCA	Total cloud amount
TSM,A	Thunderstorm or hail
TSM+	Thunderstorm heavy
$\underline{T}(t)$	Transient-state component in GEM
U	Raw predictand
V	Visibility
W	Wind
WX	Hydrometeor
X	Raw predictor
Y	Dummy predictand
$\underline{Y}'\underline{Z}$	Predictand-predictor crossproduct matrix
Z	Dummy predictor
ZL	Freezing drizzle
ZR	Freezing rain
$\underline{Z}'\underline{Z}$	Predictor-predictor crossproduct matrix
^	Signifies a predicted or estimated value
'	Transpose of a matrix
$\underline{\ \ }$	Underscoring signifies a vector or matrix

APPENDIX

A BETA CLASSIFICATION MODEL

Robert G. Miller and Donald L. Best

1. INTRODUCTION

This paper introduces a new classification procedure using beta probability density functions (pdf) to compute threshold probability values. The classification problem is this: given a probability distribution for the occurrence of an event, how does one make a categorical decision? In decision theory, such classifications are made under the control of some underlying utility function. The decisionmaker may then choose categorical selections that either maximize some gain or minimize some loss. In weather forecasting, utility is usually some verification statistic which is to be optimized (e.g., percent correct, hits, threat score, or skill score). This paper departs from the decision-theoretic approach by using a much simpler, albeit approximate, procedure incorporating threshold probabilities and a successive pair-wise comparison test. Using threshold probability values is not new; however, what has yet to be achieved is a threshold model that would provide a wide range of desired categorical responses that in turn control the verification statistic. The Beta classification model presented here accomplishes this objective. This procedure can maximize threat score, and can produce a marginal distribution balance (i.e., the number of forecast events equals the number of events observed).

2. REGRESSION PROBABILITY MODEL

The first step in the classification problem is to establish a function which can provide event probabilities. Linear regression of a selected dependent variable onto the desired independent variables accomplishes this. Here we define the independent variables, or predictors, as X_1, X_2, X_3, ...X_K. We represent the dependent variable, the predictand, as Y; its estimate is $\hat{Y}$. The desired probability model is then:

$$\hat{Y} = d_0 + d_1X_1 + d_2X_2 + ... + d_KX_K \tag{1}$$

The solution of the coefficients (d_i's) is obtained through regular multiple regression techniques with or without screening. The definition of the predictand values is absolutely necessary. The event must be exhaustive and mutually exclusive of all other possible events. If the event over the developmental data sample is observed to fall within this preselected definition of occurrence, the Y-value is assigned a "1"; otherwise it is assigned a "0." The Y-data are, therefore, binary variables representing whether the event occurred or not. The predictor variables may be either scalar, binary, or some combination of either.

Introduction of a binary predictand Y into a least-squares linear regression program produces a model which then will estimate probabilities of future events. Since there are many possible combinations of the predictors, the probability model produces a range of probability values. These values can be grouped according to verification and examined through their frequency distributions as illustrated in figure 1. This figure also shows several features that are important to the understanding of the following discussion.

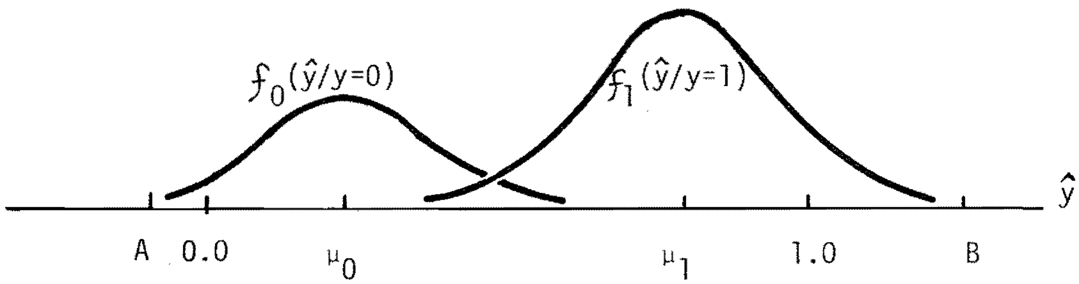


Figure 1.--Schematic depiction of the probability-value ($\hat{y}$) distributions when Y=1 and Y=0. The μ values represent distribution means.

3. CLASSIFICATION BY THRESHOLDING

There are two well defined clusters of probability values grouped into occurrence $f_1(\hat{Y}/Y=1)$ and non-occurrence $f_0(\hat{Y}/Y=0)$ of the event. The respective means of these distributions are μ_1 and μ_0. Some values fall outside the (0,1) range. The (A,B) interval represents the lower and upper bounds of possible probability values. The property that the "probability" estimate can fall outside the (0,1) range is more a nuisance to the classification problem than a mystical fact.

This property is actually of little concern, because the two distributions' overlapping values are of greater concern to us than the out-of-range values. Figure 2 portrays the overlapping problem with a given threshold value, p*.

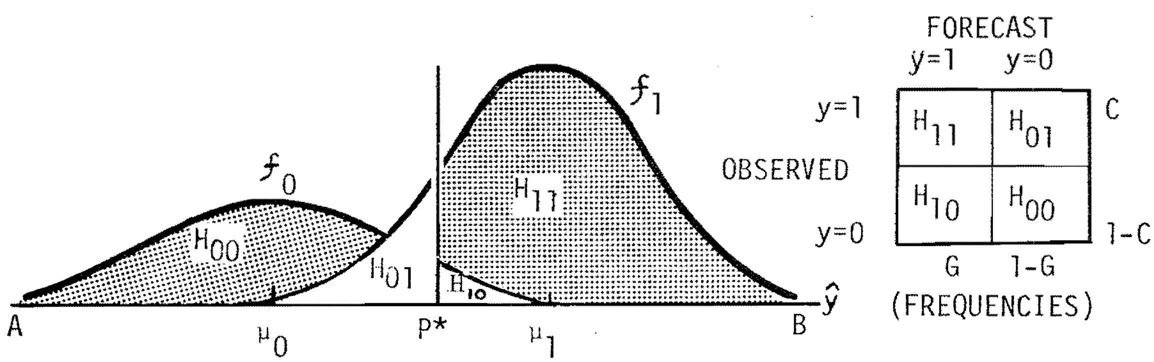


Figure 2.--Illustration of how a chosen p* (threshold probability) would control the frequency of positive classifications. A verification table is also shown. Subscripts on densities H_{ij} represent forecast category i and verified category j.

Since these two distributions describe the forecast model's response in an expected sense, we can construct an expected verification table upon which various statistical scores can be computed. The verification table's entries (H_{ij}) are estimated from the two distributions and the selected p* by these relationships:

$$
\begin{aligned}
H_{11} &= C \int_{p*}^{B} f_1 d\hat{Y} \\
H_{10} &= (1-C) \int_{p*}^{B} f_0 d\hat{Y} \\
H_{01} &= C \int_{A}^{p*} f_1 \, d\hat{Y} = C - H_{11} \\
H_{00} &= (1-C) \int_{A}^{p*} f_0 \, d\hat{Y} = (1-C) - H_{10}
\end{aligned}
\tag{2}
$$

To control the frequency of positive classifications (the G measure in figure 2), simply solve for the p* that gives the desired frequency result:

$$G = H_{11} + H_{10} \tag{3}$$

For example, classification control to balance the classification table's margins can be accomplished by finding the p* which yields G = C. Other scores can likewise be maximized by stepping p* through the (A,B) interval, deriving the expected verification table (the H_{ij} values will change), computing the desired statistical score, and stopping where the desired maximum or minimum score is found. For example, to maximize the threat score find the p* which yields $T_{max} = H_{11}/ (H_{11} + H_{10} + H_{01})$, or to maximize the Heidke skill score find *p such that

$$S_{MAX} = \frac{H_{11} + H_{00} - CG - (1-C)(1-G)}{1 - CG - (1-C)(1-G)} \tag{4}$$

A decision-theory application is also available. If a user has a known utility or value-assessment to apply against the expected verification table, one merely varies the p* until an expected maximum gain or minimum loss value results.

4. STATISTICS OF THE PROBABILITY VALUE DISTRIBUTIONS

Specifying the analytic form of the underlying distributions is a vital component of a threshold model because the H_{ij} values defined previously require some analytic function to integrate. The properties of the distributions in question are examined:

Definitions:

- C Relative frequency of the predictand event when Y=1.
- R The correlation between the Y and $\hat{Y}$ over the dependent sample (also known as the multiple correlation coefficient).
- f_i Shorthand notation for the distributions $f_i(\hat{Y}/Y=i)$, i=0,1.
- μ_i Mean value of the distribution f_i, i=0,1.
- σ_i^2 Variance of $\hat{Y}$ about μ_i when Y=i, i=0,1.
- σ^2 Total predictand variance.
- σ_w^2 Pooled predictand variance.

Computations and relationships:

$$C = \frac{1}{N} \sum_{j=1}^{N} Y_j \quad \text{(N=sample size)}$$

$$R^2 = (SST-SSR)/SST; \text{ SST=sum of squares of total, } \sum_{j=1}^{N} (Y_j - C)^2$$

$$\text{SSR=sum of squares of residuals, } \sum_{j=1}^{N} (\hat{Y}_j - Y_j)^2$$

SST-SSR=SSEX or sum of squares explained.

$\mu_0 = C \ (1-R^2)$ (see proof #1)

$\mu_1 = R^2 + C \ (1-R^2)$ (see proof #1) (Notice that: $\mu_1 - \mu_0 = R^2$)

$\sigma^2 = C \ (1-C)$ (see proof #2)

$\sigma_w^2 = C \ (1-C) \ R^2 \ (1-R^2)$ (see proof #3)

We have reason to suspect the distributions f_0 and f_1 to be beta pdf's, but to prove this is quite another matter. We postulate, therefore, that if we could parameterize the constants (also known as shape parameters) of the beta pdf using only the basic statistics described and defined above, we could compute likelihoods and use the Bayes theorem to test whether the input probability value ($\hat{Y}$) is unaltered after being transformed through a beta pdf. We surmise that, if an input value is transformed into a form which accomplishes desired results, then the transformation function is appropriate. In this case the input is the probability $\hat{Y}$, and the transformation function is the Bayes theorem using likelihoods (β_i) generated from the beta pdf's. That is, we want to show that

$$\hat{Y} = \frac{C \ \beta_1 \ (\hat{Y}|Y=1)}{C \ \beta_1(\hat{Y}|Y=1) + (1-C) \ \beta_0(\hat{Y}|Y=0)}, \tag{5}$$

with

$$\beta_i \ (\hat{Y}|Y=i) = \frac{\Gamma(\alpha_i+\nu_i)}{\Gamma(\alpha_i) \cdot \Gamma(\nu_i)} \ \hat{Y}^{\alpha_i-1} \ (1-\hat{Y})^{\nu_i-1}, \ (i=0,1) \tag{6}$$

Several empirical results substantiated that the beta pdf was the required distribution, but with the relationships given above we can also demonstrate it mathematically. (See proof #4.)

5. HANDLING THE OUT-OF-RANGE PROBLEM

The beta pdf is defined over the (0,1) interval, but figure 1 illustrates the true situation where some probability values can fall outside these bounds. One could argue, therefore, that any model which produces probabilities outside of the permissible range of the beta pdf must in fact not be replicating a beta pdf. Wadsworth and Bryan (1960) show, however, that a beta pdf can be "stretched" to other bounds such as (A,B). Stretching is performed by a transformation $U = (\hat{Y}-A)/(B-A)$ from the $\hat{Y}$-scale to a U-scale. The range of (0,1) thereby expands to (A,B). Wadsworth and Bryan also state that the solution of the stretched beta pdf uses the same shape parameters α_i and ν_i. The proper beta pdf for integration to solve the H_{ij} terms becomes:

$$\beta_i\,(\hat{Y}|Y=i) = \frac{\Gamma(\alpha_i+\nu_i)}{\Gamma(\alpha_i)\cdot\Gamma(\nu_i)}\quad U^{\alpha_i-1}\,(1-U)^{\nu_i-1}\ ,\ (i=0,1) \tag{7}$$

where proof #4 shows that:

$$\begin{aligned}\alpha_i &= \mu_i(\mu_i(1-\mu_i) - S_i^2)/S_i^2\ , \quad i=0,1\\ \nu_i &= \alpha_i(1-\mu_i)/\mu_i\ , \quad i=0,1\end{aligned} \tag{8}$$

if

$$S_i^2 = \frac{R^2}{(1+R^2)}\ \mu_i(1-\mu_i)\ , \quad i=0,1 \tag{9}$$

This information allows us to solve the H_{ij} verification values from the standard beta pdf.

An important corollary to the transformation of $\hat{Y}$ to a standard beta variate U is that any value of $\hat{Y}$ lying between A and B can be transformed to lie between 0 and 1 through the formula

$$U = \frac{\hat{Y} - A}{B - A} \tag{10}$$

Since A and B are not normally precisely known, a set of reasonable values has been found:

$$\begin{aligned}A &= \mu_0 - 2\sigma_w \quad \text{for } \mu_0<2\sigma_w\\ A &= 0 \text{ elsewhere}\\ B &= \mu_1 + 2\sigma_w \quad \text{for } (1-\mu_1)<2\sigma_w\\ B &= 1 \text{ elsewhere}\end{aligned} \tag{11}$$

also, set

$$\begin{aligned}U &= 0 \text{ when } \hat{Y} < A\\ U &= 1 \text{ when } \hat{Y} > B\end{aligned} \tag{12}$$

Proof #5 demonstrates some relationships which pertain to estimating the beta distribution parameters from known sample estimates.

6. SUMMARY

In problems such as weather forecasting it is often important to make a categorical decision about a future event. Given that we have a probability estimate of the future state of the atmosphere, we are left with the challenge of deciding whether the probability value is sufficiently large to warrant a categorical "yes it will occur" forecast. To do this we need something to compare the probability forecast against, hence the need for a critical value called the threshold probability.

When there are various users of weather-forecast information, the same probability of occurrence can evoke different categorical responses because each will most likely have different "thresholds of pain," so to speak. For example, if a 20-percent chance of a severe thunderstorm is forecast, one customer with a threshold probability of 30 percent will pick a "no it will not happen" category while another with a 15-percent threshold will definitely make plans for its occurrence. The simplicity of this classification procedure is to answer the question: does the probability forecast exceed the threshold probability? If it does, forecast an occurrence; otherwise do not. The beta pdf threshold model allows us to specify the threshold probability value needed by the user through the control of the expected frequency of positive classification (or "yes" forecasts).

APPENDIX

Proof #1: Prove that

$$\mu_0 = C(1-R^2) \tag{1}$$

and that

$$\mu_1 = R^2 + C(1-R^2). \tag{2}$$

Given that

$$R^2 = \frac{SSEX}{SST}, \tag{3}$$

where the sum of squares explained can be obtained from

$$SSEX = \sum_{k=1}^{K} d_k \sum_{j=1}^{N} X_{jk}Y_j - NC^2 \tag{4}$$

and (see proof #2)

$$SST = NC(1-C). \tag{5}$$

In addition, the mean of $\hat{Y}$ when the event occurs can be obtained from

$$\mu_1 = \sum_{k=1}^{K} d_k \sum_{j=1}^{N} X_{jk}Y_j/NC \tag{6}$$

Then, using (3), (4), and (5) we get

$$R^2 = (NC\mu_1 - NC^2)/NC(1-C). \tag{7}$$

Combining (7) with (6) will yield

$$\mu_1 = R^2 + C(1-R^2). \tag{8}$$

Now the mean of $\hat{Y}$ equals that of Y, because $\hat{Y}$ is an unbiased estimate of Y. Hence

$$C = C\mu_1 + (1-C)\mu_0 , \tag{9}$$

and (9) with (8) yields

$$\mu_0 = C(1-R^2). \qquad \text{QED} \tag{10}$$

Proof #2

$$\sigma^2 = C(1-C). \tag{1}$$

Given that Y is a binary variable (0 or 1)

$$\sigma^2 = \frac{1}{N} \cdot \text{SST}$$

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^{N} (Y_j - \overline{Y})^2 \qquad (2)$$

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^{N} (Y_j - 2Y_j\overline{Y} + \overline{Y}^2)$$

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^{N} Y_j - \frac{2Y}{N} \sum_{j=1}^{N} Y_j + \overline{Y}^2$$

Since $Y^2 = Y$ then $\sum_{j=1}^{N} Y_j^2 = \sum_{j=1}^{N} Y_j$ and $\overline{Y} = C$.

Thus,

$$\sigma^2 = C - 2C^2 + C^2 \qquad (3)$$

or

$$\sigma^2 = C(1-C). \qquad \text{QED} \qquad (4)$$

Proof #3: Prove that for $\hat{Y}$

$$\sigma_w^2 = C(1-C)\ R^2\ (1-R^2) \qquad (1)$$

given that

$$\sigma_w^2 = \frac{1}{N} \text{SSR}. \qquad (2)$$

Further, from the Analysis of Variance in regression,

$$\text{SSR} = \text{SST} - \text{SSEX} \qquad (3)$$

However, we know that

$$\text{SST} = NC(1-C)R^2 \qquad (4)$$

and

$$\text{SSEX} = n_0\ (\mu_0 - C)^2 + n_1(\mu_1 - C)^2 \qquad (5)$$

where

$$n_0 = N(1-C) \qquad (6)$$

$$n_1 = NC$$

Thus,

$$\text{SSR} = NC(1-C)R^2 - N(1-C)(\mu_0 - C)^2 - CN(\mu_1 - C)^2\ . \qquad (7)$$

But, from proof #1

$$\mu_0 = C(1-R^2) \tag{8}$$

$$\mu_1 = R^2 + C(1-R^2).$$

We then get

$$\sigma_w^2 = C(1-C)R^2-(1-C)(C-CR^2-C)^2-C(R^2+C-CR^2-C)^2$$

$$\sigma_w^2 = C(1-C)R^2-(1-C)C^2R^4-C(1-C)^2R^4$$

$$\sigma_w^2 = C(1-C)[R^2-CR^4-(1-C)R^4]$$

$$\sigma_w^2 = C(1-C)(R^2-CR^4-R^4+CR^4)$$

$$\sigma_w^2 = C(1-C)(1-R^2)R^2$$

<u>Proof #4</u>: Prove that

$$\hat{Y} = \frac{C \cdot \beta_1(\hat{Y}|Y=1)}{C \cdot \beta_1(\hat{Y}|Y=1)+(1-C) \cdot \beta_0(\hat{Y}|Y=0)} \tag{1}$$

where

$$\beta_i(\hat{Y}|Y=i) = \frac{\Gamma(\alpha_i+\nu_i)}{\Gamma(\alpha_i) \cdot \Gamma(\nu_i)} \hat{Y}^{\alpha_i-1}(1-\hat{Y})^{\nu_i-1}, \quad (i=0,1) \tag{2}$$

This is tantamount to showing that event probability forecasts, $\hat{Y}$, in the beta distribution produce likelihoods which, when applied to the Bayes theorem, yields itself.

Or, that

$$\hat{Y} = \frac{Cf_1}{Cf_1+(1-C)f_0} \tag{3}$$

Basic relationships and definitions:

$$f_1 = \frac{\Gamma(\alpha_1+\nu_1)}{\Gamma(\alpha_1)\ \Gamma(\nu_1)} \hat{Y}^{\alpha_1-1}(1-\hat{Y})^{\nu_1-1} \tag{4}$$

$$f_0 = \frac{\Gamma(\alpha_0+\nu_0)}{\Gamma(\alpha_0)\Gamma(\nu_0)} \hat{Y}^{\alpha_0-1}(1-\hat{Y})^{\nu_0-1} \tag{5}$$

$$\alpha_i = \mu_i \left(\mu_i(1-\mu_i)-S_i^2\right)/S_i^2 \qquad i=0,1 \qquad (6)$$

$$\nu_i = \left(\frac{1-\mu_i}{\mu_i}\right)\alpha_i \qquad i=0,1 \qquad (7)$$

where

μ_1 = mean of Y when $\hat{Y}$=1

μ_0 = mean of $\hat{Y}$ when Y=0

S_1^2 = variance of $\hat{Y}$ about μ_1 when Y=1

S_0^2 = variance of $\hat{Y}$ about μ_0 when Y=0

with

$$\mu_1 = R^2 + C(1-R^2) = R^2 + \mu_0 \qquad \text{(Proof \#1)} \qquad (8)$$

$$\mu_0 = C(1-R^2) \qquad \text{(Proof \#1)} \qquad (9)$$

$$S_i^2 = \frac{R^2}{1+R^2}\,\mu_i(1-\mu_i), \qquad (i=0,1) \qquad (10)$$

and

R^2 = Reduction of variance of the forecast equation, or the square of the correlation between the forecast probabilities and the dependent variable over the dependent sample.

Before we solve (3) simplify some of the above parameters:

Putting (10) into (6) reduces $\alpha_i = \dfrac{\mu_i}{R^2}$, $\qquad i=0,1 \qquad (11)$

Putting (8) or (9) into (7) reduces $\nu_i = \dfrac{1-\mu_i}{R^2}$, $\qquad i=0,1 \qquad (12)$

Now, $\alpha_i + \nu_i = \dfrac{1}{R^2}$ $\qquad i=0,1 \qquad (13)$

Rewriting (3) as

$$\frac{1}{1 + \dfrac{(1-C)}{C}\cdot\dfrac{f_0}{f_1}} = \frac{1}{1 + D}$$

and reducing the term D: Returning to (4) and (5), D becomes:

$$D = \frac{1-C}{C}\cdot\frac{\Gamma(\alpha_0+\nu_0)}{\Gamma(\alpha_1+\nu_1)}\cdot\frac{\Gamma(\alpha_1)}{\Gamma(\alpha_0)}\cdot\frac{\Gamma(\nu_1)}{\Gamma(\nu_0)}\cdot\hat{Y}^{\alpha_0-\alpha_1}\,(1-\hat{Y})^{\nu_0-\nu_1} \qquad (14)$$

From (11) $\alpha_0 - \alpha_1 = \dfrac{\mu_0-\mu_1}{R^2}$ (15)

and from (12) $\nu_0 - \nu_i = \dfrac{\mu_1-\mu_0}{R^2}$ (16)

But we also see from (8) that $\mu_1-\mu_0 = R^2$ (17)

Therefore, (15) and (16) become
$$\alpha_0 - \alpha_1 = -1$$
$$\nu_0 - \nu_1 = 1 \tag{18}$$

From (13) we see that $\Gamma(\alpha_0 + \nu_0) = \Gamma(\alpha_1 + \nu_1) = \Gamma(\frac{1}{R^2})$ (19)

Now (14) becomes, with (15), (16), and (17):

$$D = \frac{1-C}{C} \cdot \frac{\Gamma(\alpha_1}{\Gamma\ (\alpha_0)} \cdot \frac{\Gamma\ (\nu_1)}{\Gamma\ (\nu_0)} \cdot \frac{(1-\hat{Y}}{\hat{Y}} \tag{20}$$

Next we look at the ratio $\dfrac{\Gamma(\alpha_1)}{\Gamma(\alpha_0)}$: (21)

from (11) and (8)

$$\Gamma(\alpha_1) = \Gamma\ (\frac{\mu_1}{R^2}) = \Gamma(1+ \frac{\mu_0}{R^2}) \tag{21}$$

From (11) $$\Gamma(\alpha_0) = \Gamma(\frac{\mu_0}{R^2}) \tag{22}$$

Using the feature of the Gamma function that $\Gamma(1+Z) = Z\ \Gamma(Z)$, $Z>0$

we change (21) to

$$\Gamma(1+ \frac{\mu_0}{R^2}) = \frac{\mu_0}{R^2}\quad \Gamma(\frac{\mu_0}{R^2}) \tag{23}$$

Now from (22), (23), and (9)

$$\frac{\Gamma(\alpha_1)}{\Gamma(\alpha_0)} = \frac{\mu_0}{R^2} = \frac{C(1-R^2)}{R^2} \tag{24}$$

Next look at the ratio $\dfrac{\Gamma(\nu_1)}{\Gamma(\nu_0)}$:

From (12) and (8)

$$\Gamma\ (\nu_1) = \Gamma\ (\frac{1-\mu_1}{R^2}) = \Gamma\ (\frac{1-\mu_0-R^2}{R^2}) = \Gamma\ (-[1-\frac{1-\mu_0}{R^2}]). \tag{25}$$

From (12)

$$\Gamma(\nu_0) = \Gamma\ (\frac{1-\mu_0}{R^2}) \tag{26}$$

Using the feature of the Gamma function that

$$\Gamma\ (-Z) = -\frac{\Gamma(1-Z)}{Z}\ , \ Z > 0$$

Change 25 to

$$\Gamma\ (-[1-\frac{1-\mu_0}{R^2}]) = \frac{\Gamma(\frac{1-\mu_0}{R^2})}{\frac{1-\mu_0}{R^2}-1} \tag{27}$$

and using (26) and (27)

$$\frac{\Gamma(\nu_1)}{\Gamma(\nu_0)} = \frac{1}{\frac{1-\mu_0}{R^2}-1} \tag{28}$$

Before returning to solve D, (28) can be simplified further:

From (9)

$$\frac{\Gamma\ (\nu_1)}{\Gamma\ (\nu_0)} = \frac{1}{\frac{1-C(1-R^2)}{R^2}-1} = \frac{R^2}{1-C+CR^2-R^2} \tag{29}$$

$$= \frac{R^2}{(1-C)-(1-C)R^2}$$

$$= \frac{R^2}{(1-C)(1-R^2)}$$

Returning (24) and (29) to (20) yields:

$$D = \frac{1-\hat{Y}}{\hat{Y}} \tag{30}$$

Now reordered the form of (4) using (30), we finally prove

$$\hat{Y} = \frac{1}{\frac{1+1-\hat{Y}}{\hat{Y}}} = \frac{\hat{Y}}{\hat{Y}+1-\hat{Y}} = \hat{Y} \qquad \text{QED}$$

Proof #5: Show that

$$\hat{\alpha}_i = \hat{\mu}_i\left(\hat{\mu}_i(1-\hat{\mu}_i)-\sigma_1^2\right)/\sigma_1^2 \qquad i=0,1 \qquad (1)$$

$$\hat{\nu}_i = \hat{\alpha}_i(1-\hat{\mu}_i)/\hat{\mu}_i \qquad i=0,1 \qquad (2)$$

Given, from the Beta distribution (see Feller 1966, p. 49) that

$$\mu_i = \frac{\alpha_i}{\alpha_i+\nu_i} \qquad i=0,1 \qquad (3)$$

and

$$\sigma_i^2 = \frac{\alpha_i\nu_i}{(\alpha_i+\nu_i)^2(\alpha_i+\nu_i+1)} \qquad i=0,1 \qquad (4)$$

From (3) and the estimates $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ of μ_i and σ_i^2, respectively, we satisfy (2) by

$$\hat{\nu}_i = \frac{\hat{\alpha}_i(1-\hat{\mu}_i)}{\hat{\mu}_i} \qquad i=0,1 \qquad (5)$$

Now from (4) with μ_i and σ_i^2 replaced by their estimates $\hat{\mu}_i$ and $\hat{\sigma}_i^2$, respectively,

$$\hat{\sigma}_i^2 = \frac{\hat{\mu}_i^2 - \hat{\mu}_i^3}{\hat{\alpha}_i + \hat{\mu}_i} \qquad i=0,1 \qquad (6)$$

Therefore (1) is satisfied by using (4) and (6) or

$$\hat{\alpha}_i = \hat{\mu}_i\left(\hat{\mu}_i(1-\hat{\mu}_i)-\hat{\sigma}_i^2\right)/\hat{\sigma}_i^2 \qquad i=0,1 \qquad (7)$$

It is practical to employ σ_W^2 in place of σ_1^2 and σ_0^2, since the latter two require reference to the raw data and σ_w^2 does not. In fact,

$$\sigma_w^2 = R^2(1-R^2)\ C(1-C), \qquad (8)$$

from proof #3 QED

Experimental evidence has shown that using σ_w^2 for the individual group beta distributions or using σ^2 for the total beta distribution, with $\hat{Y}$ providing the likelihood ratios, performs equally well on the integration needed to determine P*.

References

Feller, William, An Introduction to Probability Theory and Its Applications, Volume II., John Wiley & Sons, New York, 1966, p. 49.

Wadsworth, G. P., and Joseph G. Bryan, Introduction to Probability and Random Variables, McGraw-Hill Book Co., 1960, pp. 101-104.

(Continued from inside front cover)

NWS 16 Storm Tide Frequencies on the South Carolina Coast. Vance A. Myers, June 1975, 79 p. (COM-75-11335)

NWS 17 Estimation of Hurricane Storm Surge in Apalachicola Bay, Florida. James E. Overland, June 1975. 66 p. (COM-75-11332)

NWS 18 Joint Probability Method of Tide Frequency Analysis Applied to Apalachicola Bay and St. George Sound, Florida. Francis P. Ho and Vance A. Myers, November 1975, 43 p. (PB-251123)

NWS 19 A Point Energy and Mass Balance Model of a Snow Cover. Eric A. Anderson, February 1976, 150 p. (PB-254653)

NWS 20 Precipitable Water Over the United States, Volume 1: Monthly Means. George A. Lott, November 1976, 173 p. (PB-264219)

NWS 20 Precipitable Water Over the United States, Volume II: Semimonthly Maxima. Francis P. Ho and John T. Riedel, July 1979, 359 p. (PB-300870)

NWS 21 Interduration Precipitation Relations for Storms - Southeast States. Ralph H. Frederick, March 1979, 66 p. (PB-297192)

NWS 22 The Nested Grid Model. Norman A. Phillips, April 1979, 89 p. (PB-299046)

NWS 23 Meteorological Criteria for Standard Project Hurricane and Probable Maximum Hurricane and Probable Maximum Hurricane Windfields, Gulf and East Coasts of the United States. Richard W. Schwerdt, Francis P. Ho, and Roger R. Watkins, September 1979, 348 p. (PB-80 117997)

NWS 24 A Methodology for Point-to-Area Rainfall Frequency Ratios. Vance A. Myers and Raymond M. Zehr, February 1980, 180 p. (PB80 180102)

NWS 25 Comparison of Generalized Estimates of Probable Maximum Precipitation With Greatest Observed Rainfalls. John T. Riedel and Louis C. Schreiner, March 1980, 75 p. (PB80 191463)

NWS 26 Frequency and Motion of Atlantic Tropical Cyclones. Charles J. Neumann and Michael J. Pryslak, March 1981, 64 p.

NWS 27 Interduration Precipitation Relations for Storms--Western United States. Ralph H. Frederick, John F. Miller, Francis P. Richards, and Richard W. Schwerdt, September 1981, 158 p.