AD-A067 093   WEATHER WING (1ST)   HICKAM AFB HI   F/G 4/2
THE USE OF 1WW HISTORICAL TERMINAL FORECAST PERFORMANCE TABLES.(U)
MAR 79   P D WOOD, J A ELLINGTON

UNCLASSIFIED   1WW-TN-79-1   NL

| OF |
AD
A067093

END
DATE
FILMED
6--79
DDC

1.0

1.1

1.25 1.4 1.6

2.8 2.5

3.2 2.2

3.6

4.0 2.0

1.8

MICROCOPY RESOLUTION TEST CHART

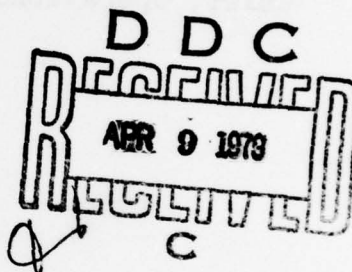NATIONAL BUREAU OF STANDARDS-1963-A

14

LEVEL

2

The Use of 1WW Historical Terminal Forecast Performance Tables.

10

Phillip D. /Wood Major, USAF and
John A. /Ellington, CMSgt, USAF

1st Weather Wing/DON
Hickam AFB, Hawaii 96853

D D C
RECEIVED
APR 9 1979
C

11   15 March 1979

9   Technical note.

**APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED**

12   16 p.

Headquarters 1st Weather Wing
United States Air Force
Hickam AFB, Hawaii 96853

AIR WEATHER SERVICE

AD A067093

DDC FILE COPY

FIRST WEATHER WING

372 450

79 04 05 068

This report approved for public release. There is no objection
to unlimited distribution of this report to the public at large,
or by DDC to the National Technical Information Service (NTIS).

This technical note has been reviewed and is approved for
publication.

FOR THE COMMANDER

ROBERT W. GOSSETT, JR, Colonel, USAF
Chief, Operations Division

PREFACE

During 1976 and 1977, technical consultants assigned to
HQ 1 WW experienced difficulty accomplishing evaluations of
subordinate unit terminal forecast performance. It was
assumed that subordinate units were experiencing the same
difficulty. There was no diagnostic aid (except the unit PIOP
standard for the TAF) available to use in reviewing the end-of-
month summaries. In particular, there were no historical data
in existence with which to compare the various elements on the
end-of-month TAFVER summary--prefigurance, post-agreement,
optimistic and pessimistic bias, etc. It was felt that
historical terminal forecast performance data could serve many
useful purposes. This technical report describes the elements
of the historical tables that were prepared and various ways
the tables can be used.

Phillip D Wood

PHILLIP D. WOOD, Major, USAF

ACKNOWLEDGEMENTS

CMSgt John D. Ellington is commended for determining the
need for and designing the 1 WW historical terminal forecast
performance tables. I also wish to thank Capt Rod Yama,
Capt Al Simoncic, and, especially, CMSgt John Ellington for
dedication in meticulously compiling all past 1 WW TAFVER
statistics available at HQ 1 WW into the original 1 WW
historical terminal forecast performance tables.

Phillip D Wood

PHILLIP D. WOOD, Major, USAF

i

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 1WW-TN-79-1	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* The Use of 1WW Historical Terminal Fore-cast Performance Tables		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Phillip D. Wood and John A. Ellington		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS 1st Weather Wing/DON Hickam AFB HI 96853		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS 1st Weather Wing/DO Hickam AFB HI 96853		12. REPORT DATE 15 March 1979
		13. NUMBER OF PAGES 10
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) 1st Weather Wing/DON Hickam AFB HI 96853		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE NA

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited

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

NA

18. SUPPLEMENTARY NOTES

Heidke Skill Score	Prefigurance
Optimistic Bias	Postagreement
Pessimistic Bias	

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

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

This report describes the various elements of the 1WW historical terminal forecast performance tables. Numerous ways the tables can be used are described and illustrated. The 1WW historical terminal forecast performance data are a one-of-a-kind diagnostic aid used to measure and demonstrate unit technical ability and performance.

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

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

iii

I. INTRODUCTION

In early 1978, 1 WW/DON compiled and published historical terminal forecast performance tables for Det 2, 1 WW (Andersen AFB, GU); Det 5, 1 WW (Clark AB, PI); Det 7, 1 WW (Wheeler AFB, HI); Det 8, 30 WS (Kadena AB, JA); Det 10, 30 WS (Kunsan AB, KR); Det 15, 30 WS (Osan AB, KR); and Det 17, 30 WS (Yokota AB, JA). The period of record of the tables was as follows: 30 WS units--10 years; 1 WW direct reporting units--5 years (Det 7 tables only 2 years). The historical terminal forecast performance tables were designed to be used as objective diagnostic aids. The remainder of this technical note will describe the elements of historical tables and ways to use the tables.

II. ELEMENTS OF THE HISTORICAL TERMINAL FORECAST PERFORMANCE TABLES.

Figure 1 is an example of a historical terminal forecast performance table. By examining the legend one can quickly determine the month and period of record of that particular table. The data in the tables are forecast verification statistics of all terminal forecasts issued during the period of record. Considering this, the tables should provide a good indication of typical terminal forecast performance at the unit, strong and weak areas, and an excellent, diagnostic tool with which to perform technical evaluations of unit forecast performance.

The contingency tables for the 3, 6, 12, 24 and all hour periods show forecast and observed conditions for the period of record. Looking at the 3 hour table in Figure 1, the "33" means that Category C was forecast 33 times when Category D was observed during the 10 year period of record. Category A, B, C, and D have the same ceiling and visibility values as the AWS TAFVER categories.

Category	Cloud Ceiling (Ft)	Visibility (Statute Miles)
A	< 200	$< 1/2$
B	200 to < 1000	1/2 to < 2
C	1000 to < 3000	2 to < 3
D	≥ 3000	≥ 3

"SS" is the Heidke Skill Score computed using the data in the contingency table. Note that there are separate Heidke Skill Scores for 3, 6, 12, 24, and all hours by month. The

1

HISTORICAL TERMINAL FORECAST PERFORMANCE

3 HOUR	A	B	C	D	TOTAL
A	1	2	1	7	11
B	0	14	21	19	54
C	0	5	56	49	110
D	1	2	33	967	1003
TOTAL	2	23	111	1042	1178

SS= .49885   OB= .08404   PB= .03480

6 HOUR	A	B	C	D	TOTAL
A	0	2	2	4	8
B	1	10	23	22	56
C	0	10	39	60	109
D	1	7	34	963	1005
TOTAL	2	29	98	1049	1178

SS= .39104   OB= .09593   PB= .04499

12 HOUR	A	B	C	D	TOTAL
A	0	1	3	3	7
B	0	6	10	38	54
C	0	5	29	71	105
D	0	14	40	958	1012
TOTAL	0	26	82	1070	1178

SS= .26237   OB= .20705   PB= .04928

24 HOUR	A	B	C	D	TOTAL
A	0	0	1	5	6
B	0	0	3	49	52
C	0	2	12	89	103
D	0	4	36	977	1017
TOTAL	0	6	52	1120	1178

SS= .08369   OB= .12479   PB= .03565

ALL HOUR	A	B	C	D	TOTAL
A	1	5-	7	19	32
B	1	30	57	128	216
C	0	22	136	269	427
D	2	27	143	3865	4037
TOTAL	4	84	343	4281	4712

SS= .32626   OB= .10293   PB= .04138

PREFIGURANCE

	3 HR	6 HR	12 HR	24 HR	ALL HR
A	.09091	0	0	0	.03125
B	.25926	.17857	.11111	0	.13889
C	.50910	.35779	.27619	.11650	.31850
D	.96411	.95821	.94664	.96067	.95739

POST-AGREEMENT

	3 HR	6 HR	12 HR	24 HR	ALL HR
A	.50	0	0	0	.250
B	.60870	.34482	.23077	0	.35714
C	.50450	.39796	.35366	.23077	.39650
D	.92802	.91802	.89533	.87232	.90283

MISSED CATEGORY

	3 HR	6 HR	12 HR	24 HR	ALL HR
+3	.00595	.00340	.00255	.00424	.00403
+2	.01698	.02037	.03480	.04244	.02865
+1	.06112	.07216	.06961	.07809	.07025
0	.88115	.85908	.84295	.83956	.85563
-1	.03226	.03820	.03820	.03226	.03523
-2	.00170	.00594	.01188	.00340	.00573
-3	.00085	.00085	0	0	.00042

LEGEND

UNIT XXX

MONTH   January

PERIOD OF RECORD: 19 68 THRU 19 77

FIGURE 1.   EXAMPLE OF 1 WW HISTORICAL TERMINAL FORECAST PERFORMANCE TABLE

2

Heidke Skill Score (SS) is calculated by the following formula:

$$SS = \frac{F - D}{T - D}$$

    F = Number of correct forecasts.
    T = Total number of forecasts.
    D = Number of correct forecasts which could be expected purely by chance.

SS, then, determines the number of forecasts which could have been hit by chance, eliminates them, and computes a score based upon the remainder, "those not attributable to chance." In the 3 hour contingency table, the Heidke Skill Score was calculated as follows:

$$SS = \cfrac{(1+14+56+967) - \cfrac{(2\times11)+(23\times54)+(111\times110)+(1042\times1003)}{1178}}{1178 - \cfrac{(2\times11)+(23\times54)+(111\times110)+(1042\times1003)}{1178}}$$

SS = .49885

"OB" and "PB" are optimistic and pessimistic bias (i.e., percent of total forecasts that were optimistic and pessimistic). Looking at the 3 hour contingency table again, the OB and PB were calculated in the following manner:

$$OB = \frac{2+1+7+21+19+49}{1178} = .08404 \text{ (8.4\% of all forecasts were optimistic)}$$

$$PB = \frac{0+0+5+1+2+33}{1178} = .03480 \text{ (3.5\% of all forecasts were pessimistic)}$$

Next let's examine the "Prefigurance" table in the upper right corner of the form. Prefigurance is the <u>capability</u> of correctly forecasting any weather event.

$$\text{Prefigurance} = \frac{\text{number of correct forecasts}}{\text{number of observed occurrences}}$$

3

For the 3 hour contingency table and Category A, prefigurance is calculated as follows:

Prefigurance = $\dfrac{1}{11}$ = .09091

Now let's move on to the post-agreement table. Post-agreement is the <u>reliability</u> of the forecasts that were issued.

Post-agreement = $\dfrac{\text{number of correct forecasts}}{\text{number of forecasts issued}}$

For the 3 hour contingency table and Category A, the post-agreement is calculated as follows:

Post-agreement = $\dfrac{1}{2}$ = .50

In the "Missed Category" table, the percent of total forecasts that were 3, 2, and 1 category optimistic and pessimistic misses are listed. Using the 3 hour contingency table, the percent of 1 category optimistic misses is calculated as follows:

Percent of
1 Cat Missed = $\dfrac{2+21+49}{1178}$ (Fcst/Obsd: = .06112
B/A, C/B, D/C)
(Total Fcsts)   (or 6.1% of all 3 hr
fcsts missed
optimistically by
one category)

The .88115 in the "3 HR" column means that 88.115 percent of 3 hour forecasts were correct.

After studying the above, one must conclude that the historical terminal forecast performance tables merely summarize a unit's past terminal forecast performance. The data in the tables take into account (1) unit location and associated weather, (2) seasonal variations of weather over the period of record, (3) technical capabilities and limitations of all assigned forecasters, and (4) the effects of new developments, new equipment, and/or new forecasting techniques, forecast studies, and other aids used during the period of record.

This section was designed to acquaint you with the various items in Figure 1. In Section III, we suggest ways to use the historical terminal forecast performance data.

4

III. WAYS TO USE THE HISTORICAL TERMINAL FORECAST PERFORMANCE TABLES.

The historical terminal forecast performance tables can be used in a number of ways. We will first provide a list of the ways that the data can be used, then we will discuss each proposed use. Examples will be provided to illustrate some of the ways. You may identify additional applications to use in your local technical enhancement program.

LIST OF WAYS TO USE HISTORICAL TABLES

1. In conjunction with TAFVER end-of-month summary, to accomplish a technical evaluation of the unit's terminal forecast performance.

2. As an aid to alert forecasters to forecast problems that traditionally have occurred during the next month or next quarter.

3. As a tool, the station chief can use to technically evaluate individual forecaster performance.

4. As an aid to alert forecasters of the number of times the low categories occurred at the verification times during the period of record.

5. As a tool, to help the station chief direct technical improvement efforts or additional training.

6. As an aid to guide TAF preparation.

7. In numerous other unit programs or activities such as Metcons, TAF bust review program, end-of-month performance evaluation to Detco, forecaster indoctrination training program, seminar program, evaluating OPSVER performance where forecast thresholds coincide with TAFVER categories.

8. As an aid to determining the unit's capability and limitations for providing support to operational thresholds which closely approximate the AWS TAFVER categories.

9. As a tool to assist wing or squadron technical consultants in determining the need for special technical consultant visits and how to prepare for those visits.

1. Used in conjunction with the TAFVER end-of-month summary to accomplish a technical evaluation of the unit's terminal forecast performance. The information presented in the

historical terminal forecast performance data will make the information in the end-of-month TAFVER summary have meaning. Figure 2 is a portion of a Jan 78 TAFVER summary. We computed the Heidke Skill Score and entered the score achieved by the unit on its 3 and 6 hour forecasts during Jan 78. (You will note on the end-of-month TAFVER summaries you receive that we enter the Heidke Skill Scores for 3, 6, 12, 24 and all hours.)

The SS of .2225 for the "3 Hour Forecast Summary" in Figure 2 is less than the SS of .49885 in Figure 1. This is a clue to look closer at the performance. At this point we wish to emphasize that the Heidke Skill Score is just one measure of unit technical performance; all diagnostic data should be considered when evaluating unit technical perform-ance regardless of SS.

In Figure 2 the percent of correct forecasts is 87.1. In the "Missed Category" table of Appendix 1 the corresponding number is .88115 or 88.1 percent correct. Therefore, the percent correct achieved in Jan 78 by this unit is only one percent below historical performance.

In Figure 2 the percentages of optimistic and pessimistic forecasts are 8.87 and 4.03, respectively. Appendix 1 corresponding numbers are 8.40 and 3.48. These data indicate the lower percent correct forecasts versus historical data were due to larger than historical optimism and pessimism. Also, forecasters continued to be much more optimistic than pessimistic.

In Figure 2 the prefigurances at 3 hours for Categories A, B, C, and D are 0, 14.2, 14.2, and 96.3, respectively. The Figure 1 corresponding numbers are 9.09, 25.93, 50.91, and 96.41. Unit capability with Categories B and C was not good.

In Figure 2 the post-agreements at 3 hours for Categories A, B, C, and D are 0, 33.3, 20.0, and 91.3. The Figure 1 corresponding numbers are 50.0, 60.9, 50.45, and 92.80. Forecast reliability for all categories fell below historical performance at the unit.

Next, let's compare the percent of forecast misses by category:

	+3	+2	+1	-1	-2	-3
Figure 1	.59	1.69	6.11	3.23	.17	.08
Figure 2	0	4.03	4.84	3.23	.81	0

6

3 HOUR FORECAST SUMMARY — JAN 78 — FORECAST SKILL

L DETACHMENT — CIG/VSBY COMBINED

STATION FORECAST	A	B	C	D	TOTAL		PERSISTENCE FORECAST	A	B	C	D	TO
A	0	0	0	0	0	A (OBSERVED)		0	0	0	0	
B	0	1	1	5	7	B	0	0	1	2	4	
C	0	1	1	5	7	C	0	0	1	2	4	
D	0	1	3	106	110	D	0	0	5	3	102	1
	0	3	5	116	124	TOTAL	0	0	7	7	110	1

POE: .0 14.2 14.2 96.3   PoE: .0 14.2 28.5 92.7
POST: .0 33.3 20.0 91.3   POST: .0 14.2 28.5 92.7

ENT MISSED BY CATEGORIES

+3	.00
+2	4.03
+1	4.84
0	87.10
-1	3.23
-2	.81
-3	.00

PERCENT MISSED BY CATEGORIES

+3	.00
+2	3.23
+1	4.84
0	84.68
-1	3.23
-2	4.03
-3	.00

ENT CORRECT FORECASTS 87.10 — PERCENT CORRECT FORECASTS 84.68
ENT OPTIMISTIC FORECASTS 8.87 — PERCENT OPTIMISTIC FORECASTS 8.86
ENT PESSIMISTIC FORECASTS 4.03 — PERCENT PESSIMISTIC FORECASTS 7.26

SS = .147

6 HOUR FORECAST SUMMARY — JAN 78 — FORECAST SKILL

DETACHMENT — CIG/VSBY COMBINED

STATION FORECAST	A	B	C	D	TOTAL		PERSISTENCE FORECAST	A	B	C	D	TO
A	0	0	0	0	0	A (OBSERVED)		0	0	0	0	
B	0	0	2	6	8	B	0	0	0	2	6	
C	0	0	1	6	7	C	0	0	1	1	5	
D	0	1	2	106	109	D	0	0	6	4	99	1
	0	1	5	118	124	TOTAL	0	0	7	7	110	1

POE: .0 .0 14.2 97.2   PoE: .0 .0 14.2 96.8
POST: .0 .0 20.0 89.8   POST: .0 .0 14.2 90.0

ENT MISSED BY CATEGORIES

+3	.00
+2	4.84
+1	6.45
0	86.29
-1	1.61
-2	.81
-3	.00

PERCENT MISSED BY CATEGORIES

+3	.00
+2	4.84
+1	5.65
0	80.65
-1	4.03
-2	4.84
-3	.00

ENT CORRECT FORECASTS 86.29 — PERCENT CORRECT FORECASTS 80.65
ENT OPTIMISTIC FORECASTS 11.29 — PERCENT OPTIMISTIC FORECASTS 10.48
ENT PESSIMISTIC FORECASTS 2.42 — PERCENT PESSIMISTIC FORECASTS 8.87

Figure 2. Extract from TAFVER
End-of-Month Summary

These data indicate +2 misses were over twice as great as
historical performance; -1 misses were equal to past
performance; -2 misses were more than twice past performance.
A commendable observation is the lack of 3-category busts
during this period. So this illustrative example suggests
the need to reduce the optimism of 3 hour forecasts and work
on Category B and C misses. Since the +2 misses were more
than twice as great as historical performance, the station
chief might ask forecasters to do postanalyses of two
category optimistic misses involving Category B and one
category optimistic misses involving Category C when the
trend of the TAF was far off.

Without the historical terminal forecast performance data,
the technical evaluation by the station chief would only
include comparing station performance versus persistence
performance with an eye on the unit PIOP performance standard
developed by 1 WW. The historical data promote full use of
the forecast verification data in the TAFVER end-of-month
summary. Results of these postanalyses should be shared with
other forecasters in the manner determined best by the station
chief.

2. Used as an aid to alert forecasters to forecast problems
that traditionally have occurred during the next month or
next quarter. Prior to the beginning of a new forecast month
or quarter, the station chief could review the applicable
historical tables and emphasize to assigned forecasters likely
performance and "difficult-to-forecast for" categories and
time periods. For example, after a review in Dec 78 of the
Jan 78 historical data (see Figure 1), the station chief
might highlight the following items to unit forecasters:

    a. Historical data suggests that our lowest percent
correct in January will occur at the 24 hour period.

    b. OB and PB figures for the 3, 6, 12, and 24 hour
periods reflect that "optimism" is likely to cause the
majority of our forecast misses in January.

    c. Due to the infrequent occurrence of Category A and
low forecast skill, we need to review the synoptic situations
(i.e., case studies and bust reviews) for Category A condi-
tions. Also, noting the low skill forecasting Category B,
the station chief might direct use of a new forecasting technique
to try and improve capability and reliability of forecasts.

We think 1WW units with personnel assigned on short tours
will benefit by implementing these procedures.

8

3. **Used as a tool the station chief can use to technically evaluate individual forecaster performance.** To do this, the station chief or individual forecaster would have to compile performance data to compare with some or all of the historical data. If one or a couple of forecasters are logging the majority of forecast misses, the station chief might ask them to compile tables as in Figure 1 on their forecasts. They could gain insight into their weaknesses if they compare their performance to the historical tables and point out their shortfalls to the station chief. They could discover those categories and time frames for which it might benefit them to do postanalyses or question other forecasters on clues, hints, etc., that led them to successful forecasts. The station chief could use the historical data versus performance comparison to learn those areas in which to assist individual forecasters during metcons or preparation of the TAF worksheet.

4. **Used as an aid to alert forecasters of the number of times the low categories occurred at the verification times during the period of record.** The tables will clearly show that repeated forecasts of low categories (i.e., Cats A and B) are imprudent at certain locales and time periods during specified months. Based on Figure 1, we conclude that repeated forecasts of Category A at 6, 12, or 24 hours are imprudent.

5. **Used as a tool to help the station chief direct technical improvement efforts or additional training.** When the station chief's comparisons of historical data versus end-of-month TAFVER statistics over a couple or several months revealed a pattern of contemporary forecast weaknesses, he should direct one or more of the following actions.

　　a. Preparation of postanalyses focused on specified forecast weaknesses.

　　b. Inclusion of new or revised forecast techniques in the local analysis and forecast program (LAFP) and/or into the TAF preparation worksheet.

　　c. Preparation of case studies and presentation to or review by unit forecasters.

　　d. Renovation of the forecast discussions prior to TAF completion with special attention given to forecast weaknesses.

　　e. Initiation of a request to higher headquarters for assistance (e.g., copies of applicable forecast studies at other units, technical literature on forecasting techniques, or some type of special assistance from USAFETAC).

9

f. Initiation of a unit follow-on training effort on forecasting ceilings and visibilities associated with specified synoptic patterns or weather phenomena.

6. Used as an aid to guide TAF preparation. This has been touched on previously under item 4 and elsewhere. Basically, the historical tables give a forecaster a feel for the advisability of forecasting low categories, whether the forecaster can benefit by being less optimistic or pessimistic and at what time periods to spend the most time trying to devise a logical forecast.

7. Used in numerous other unit programs and activities such as metcons, TAF bust review program, end-of-month perform- ance evaluation to Detco, forecaster indoctrination training program, evaluating OPSVER performance where forecast thresholds coincide with TAFVER categories. Those activities not dealt with previously in the above discussions really require no clarification.

8. As an aid to determining the unit's capability and limitations for providing support to operational thresholds which closely approximate the AWS TAFVER categories. If a customer needs forecast support for some criteria near the TAFVER criteria and asks what you anticipate your capability is, you could use the historical data to estimate your capability.

9. As a tool to assist wing or squadron technical consultants in determining the need for special technical consultant visits (TCVs) and how to prepare for those visits. At wing and squadron PIOP standards are used primarily to accomplish performance evaluations. When those evaluations indicate a unit is faltering, a closer examination can be made by using the historical terminal forecast performance data. Use of those tables helps consultants isolate problems. Once this is done, preparation can be made for special TCVs.

We have enumerated numerous ways the historical terminal forecast performance data can be used. We are certain you can think of others. By proposing all these ways to use the data, we are not suggesting that the historical tables are a panacea (i.e., a cure-all) or that every unit should use the tables in the ways we have described. Your particular management style should dictate how and how frequently you use the tables. If this technical note has thoroughly acquainted you with the elements of the historical tables and shown you some useful ways to exploit the data that you hadn't previously considered, our purpose in writing this publication has been achieved.