METEOROLOGY GROUP OF THE EANPG (METG)

EIGHTEENTH MEETING

(Paris, 22 – 26 September 2008)

Agenda Item 9: Institutional Issues

TAF VERIFICATION SYSTEM USED IN THE MET ALLIANCE

(Presented by Austria)

Summary

A common TAF Verification System was adopted by the MET Alliance Board in May 2008. In this method, two conditions are verified for each hour of the TAF: The highest (or most favourable) observed value is used to score the highest forecast value, and the lowest (or most adverse) observed value is used to score the lowest forecast value. This enables the use of the verification results for quality management, forecaster training and user-oriented issues.

1. Introduction

1.1 In ICAO Annex 3 para 2.2.2, the implementation of a quality system is recommended. In this context, the assessment of the quality of TAFs is of key importance. Verification results should be made available to management, forecasters and users.

1.2 The MET Alliance, which is formed by the national aeronautical meteorological service providers from Austria, Belgium, Ireland, the Netherlands, Switzerland and Germany, has agreed on a common TAF verification method in its Board Meeting in May, 2008. This method has originally been developed by Austro Control and is operational there. The system is currently implemented to fit the requirements of all MET Alliance members.

2. The TAF Verification Method

2.1 The TAF verification system is tailor-made in respect to the properties of TAFs. In a TAF, the forecaster gives a range of possible meteorological conditions by using different types of change groups. These conditions are valid for time intervals, the shortest being 1 hour. A TAF thus contains a range of forecast conditions for each hour.

2.2 Point verification has proved difficult for TAFs. To ease these difficulties, time and meteorological state constraints are relaxed. This is done by verifying two conditions for each hour of the TAF: The highest (or most favourable) observed value is used to score the highest forecast value, and the lowest (or most adverse) observed value is used to score the lowest forecast value. All available observations within the respective hour are used (METAR and SPECI).

2.3 This approach avoids the need of assumptions about probabilities for conditions forecast by TEMPO and PROB TEMPO, or ambiguous conditions during a BECMG period.

2.4 Visibility, ceiling height, present weather, wind speed and wind gusts are verified in categories delimited by the TAF amendment criteria of Annex 3. The following example shows the principles of the MET Alliance TAF Verification method.

Table 1: TAF Verification Example for Visibility

TAF VIS	4000 TEMPO 0709 0700 BCFG				OBS (m)	8000 2000	0400 1800	3000 6000	8000 9999
VIS \ TIME	06-07	07-08	08-09	09-10		06-07	07-08	08-09	09-10
5000 - 9999									
3000 - <5000									
1500 - <3000									
0800 - <1500									
0600 - <0800									
0350 - <0600									
0150 - <0350									
0000 - <0150									

The highest FCST / OBS category AND the lowest FCST / OBS category are verified for each hour.

2.5 Wind direction is verified only when the wind speed is ≥ 7 kt by checking if the observed wind direction lies within ± 20° from any forecast direction (ICAO Annex 3, Attachment B, Operationally desirable accuracy of forecasts).

2.6 Entries into contingency tables are made for each pair of OBS / FCST values. The size of the contingency tables is determined by the number of classes resulting from the TAF amendment criteria. There are pairs of contingency tables (one for highest and one for lowest values) for each weather element. Additionally, separate tables are set up for each lead time to investigate the dependence of TAF quality on lead time.

Table 2: Schematic of a 2*2 Contingency Table

Forecast	Observation		
	Yes	No	Total
Yes	a	b	a+b
No	c	d	c+d
Total	a+c	b+d	n=a+b+c+d

2.7 For management information, common verification measures for categorical events are calculated from the contingency tables.

Table 3: Scores Calculated from 2*2 Contingency Tables

Name of measure	Definition
Probability of Event (base rate)	p(E) = (a+c) / n
Bias	Bias = (a+b) / (a+c)
Hit rate (Probability of Detection)	H = POD = a / (a+c)
Proportion Correct	PC = (a+d) / n
False Alarm Ratio	FAR = b / (a+b)
False Alarm Rate	F = b / (b+d)
Conditional probability of an event, given:	
- the event was forecast	p(E) when fcst = a / a+b
- the event was not forecast	p(E) when not fcst = c / c+d
Heidke Skill Score (HSS)	HSS = (a+d – E) / (n – E)
with E = PC for random forecasts	E = ((a+b)*(a+c) + (b+d)*(c+d)) / n
Peirce´s Skill Score (PSS)	PSS = H - F = (a*d – b*c) / ((a+c)*(b+d))
Odds ratio skill score ORSS (Yule´s Q)	ORSS = (a*d – b*c) / (a*d + b*c)

For n*n contingency tables, the Gerrity Score (GS), the Heidke Skill Score (HSS) are determined. Further, frequencies of highest / lowest forecast classes being higher, equal and lower than the highest / lowest observed classes are computed. Additionally, measures are produced similar to the requirements of ICAO Annex 3, Attachment B (Operationally desirable accuracy of forecasts) are available. However, it must be mentioned that these measures have to be discussed as they show a strong dependence on airport climatology and they only reflect TAF amendment criteria to a certain extent.

2.8 For customer-oriented verification, the user´s flight planning procedures are considered. Often, only the lowest (most adverse) forecast conditions being below a certain threshold are used for flight planning. These can be evaluated directly by looking at “events” from the minimum value contingency tables, thus reducing the n*n- to a 2*2-contingency table, e.g. the event of visibility below 1500m (Table 4c).

2.9 Table 4 contains an example of TAF verification results. For shortness, only 4 thresholds (350m, 800m, 1500m and 3000m) are used.

Table 4: Example of TAF Visibility Verification Results for Hamburg EDDH, October 2007 – January 2008.

a) 5*5 Contingency Table for Maximum Visibility (m)

FCST \ OBS	<350	350 - <800	800 - <1500	1500 - <3000	≥3000	SUM
<350	3	0	0	0	0	3
350 - <800	1	7	4	1	7	20
800 - <1500	2	9	8	16	14	49
1500 - <3000	3	13	22	15	50	103
≥3000	18	18	17	42	7994	8089
SUM	27	47	51	74	8065	8264

b) 5*5 Contingency Table for Minimum Visibility (m)

FCST \ OBS	<350	350 - <800	800 - <1500	1500 - <3000	≥3000	SUM
<350	23	27	15	7	10	82
350 - <800	14	11	26	31	86	168
800 - <1500	10	4	3	20	95	132
1500 - <3000	7	9	9	47	516	588
≥3000	8	4	8	54	7220	7294
SUM	62	55	61	159	7927	8264

c) 2*2 Contingency Table for Minimum Visibility (m), Threshold 1500 m (Derived from Table 4b)

FCST \ OBS	< 1500	≥ 1500	SUM
< 1500	133	249	382
≥ 1500	45	7837	7882
SUM	178	8086	8264

d) Scores for the 5*5 Contingency Tables (Computed from Table 4a and 4b)

Measure	Maximum Vis (Table 4a)	Measure	Minimum Vis (Table 4b)
Gerrity Score GS	0.260		0.598
Heidke Skill Score HSS	0.357		0.234
Peirce Skill Score PSS	0.335		0.447
FC max < OBS max	0.011	FC min < OBS min	0.101
FC max = OBS max	0.971	FC min = OBS min	0.884
FC max > OBS max	0.018	FC min > OBS min	0.015

e) Scores for Minimum Visibility <1500 m (Computed from Table 4c)

p (E)	0.022
Bias	2.146
H = POD	0.747
PC	0.964
FAR	0.652
F	0.031
p(E) when fcst	0.348
p(E) when not fcst	0.006
HSS	0.459
PSS	0.716
ORSS	0.979

2.10 There are many conclusions that can be drawn from the material in Table 4a-e. Here, only some of them are discussed.

Table 4a shows that in most cases with poor visibility, the forecast maximum value is (partly considerably) higher than the observed maximum value. Still, that does not mean that poor visibility has not been forecast in these cases, as it may have been contained in a (PROB) TEMPO or BECMG group.

Table 4b shows a relatively large number of cases (516) with forecast minimum visibility between 1500 m and <3000 m and observed minimum visibility ≥3000 m. It may be assumed that many of these cases are associated with visibility reductions forecast to go along with convective precipitation, which fails to appear in the observations.

Table 4c shows that 249 cases appear where minimum visibilities below 1500 m were forecast which were not observed. These cases are mostly associated with (PROB) TEMPO forecasts of low visibilities like fog or fog patches which then fail to appear.

The Gerrity Score GS is a measure that puts higher weights on forecasts of rare events. Thus, it is higher for minimum visibility forecasts, which are much better in poor visibility cases. On the other hand, most deviations between observed and forecast visibility values are due to too low forecast minimum values (Table 4d). This is confirmed by the bias value of 2.146 (Table 4e). The combination of p(E), p(E) when forecast and p(E) when not forecasts is suitable to give good information about the relevance of the phenomenon, and about the benefit from using the forecasts compared to climatological information.

2.11 TAFs are also checked for syntax errors and for the occurrence of more than one change group of the same type with overlapping validity. Such incorrectly coded TAFs are counted and their percentage is one of the results, but they are excluded from the verification of TAF quality.

3. Action by the METG

3.1 The group is invited to note the information in this paper.

3.2 The MET Alliance TAF Verification System should be considered in the context of a community-wide standard TAF verification scheme.

– END –