Description of the TAF Verification EXCEL files

Guenter Mahringer, Austro Control Aviation Met Service, guenter.mahringer@austrocontrol.at

September 2022

1. Introduction

In the TAF Verification EXCEL result files, data produced by the verification routines are imported, and verification results are produced as contingency tables, scores and graphics.

EXCEL result files are produced for ceiling – shortcut: CLD, visibility – VIS, present Weather – WX, wind direction – DD, wind speed (threshold based) – FF, and/or wind speed (deviations OBS-FCST) – FFD, wind gusts (threshold based) – FFX, and wind gusts (deviations OBS-FCST) – FFXD.

All result files contain information about ICAO location indicator, begin and end of investigated period, number of TAFs and TAF hours investigated and not investigated due to coding errors or lack of METAR observations.

They contain these **worksheets**:

- **Results:** contains summary contingency tables, scores and graphics. In lines 13 and 14, the user may select TAF issue time ranges and the time axis in graphics - lead-time or UTC (except for DD).
- **Results-10h (except for DD):** contains summary contingency tables, scores and graphics for a range of lead-time (default: 0 to 10h, the user may select lead-time range in line 15).
- **Comparison:** collection of scores re-ordered in two rows (all lead-times and range of lead-times) for easier production of comparisons. All scores are also contained in tables “results” and “results-10h”.
- **Intermediate:** scores calculated and ordered for the production of graphics.
- **Calculations:** summary contingency tables and scores, as well as detail contingency tables and scores for each forecast hour. Source for “results”, “results-10h” and “intermediate”.
- **Data:** input from verification routines

The main results are contained in the “results” and “results-10h” worksheet. The other worksheets are of use for comparisons and detailed special studies.

2. Wind direction DD

Wind direction is verified when the OBS FF and/or the lowest FCST FF is above a defined limit (FF ≥ METAR FF threshold). Forecasts are counted correct if the deviation between OBS and FCST wind direction is not greater than the allowed deviation (4), otherwise they are counted incorrect (2).

VRB is counted correct only when observed and forecast, otherwise it is counted incorrect (1). Both the allowed DD deviation and the FF limit are shown at the top of the results table.

When the OBS FF and the lowest FCST FF is < METAR FF threshold, the wind direction is not regarded significant and therefore counted correct (5).

The DD results table contains the fraction of correct DD forecasts when FF is ≥ METAR FF threshold, and for all cases. The values are detailed for each hour of TAF lead-time and UTC time. The figures show the dependence of correct forecasts on lead-time (top) and UTC (bottom).

3. Wind speed and wind gusts using OBS-FCST deviation – FFD, FFXD

For wind speed and wind gusts, deviations between observed and forecast hourly maxima / minima are investigated. Differences OBS – FCST are grouped in classes. Tables are set up dependent on FCST wind speed in order to see if under- or overforecasting depends on the forecast value. However, it has to be noted that METARs will not show real hourly maxima and minima of wind speed in most cases.

For wind gusts, the observation is the maximum gust of each hour, data have to be provided and inserted, as they are not contained in routine observations. Therefore, the comparison with the maximum forecast is primarily interesting.

- **Tables:**

  For maxima and minima, absolute and relative tables are shown. They can be used for forecaster training.

  The frequency of differences below a certain value (usually, <5kt and <10kt) may be used as performance scores.

- **Figures:**

  Figure (1) and (2): Frequencies of differences OBS - FCST in classes, depending on FCST, for maximum and minimum wind speed. For gusts, only the maximum is useful.

  Figure (3): Dependence of the fraction of correct forecasts (within deviation limits ±4kt (<5kt) and ±9kt (<10kt)) on lead-time or UTC, as selected in line 13.

4. Cloud ceiling – CLD, Visibility – VIS, Wind speed using thresholds – FF, Wind gusts using thresholds

Result file contain:

- **Contingency tables:**

  Contingency tables are shown for maximum and minimum ceiling, visibility and wind speed (hourly maximum / minimum FCST compared with hourly maximum / minimum OBS).

  Contingency tables can be used for forecaster training to show strengths and weaknesses of forecasts.

  For a detailed description of contingency tables and their calculation see Mahringer, 2008 and cited literature on forecast verification. Classes are delimited by pre-defined threshold values.

- **Total Scores:**

  KPI: Key performance indicator: Summary score. Rare events get high weight and frequent events get low weight. Calculation: Mean Event (PSS+HSS)/2 (from Table "Scores for Events") for all events with p(E) > value defined in cell F8. Properties: high correlation with hits and false alarms, especially in rare events. The score may be fluctuating dependent on statistics of events.

  The KPI along lead-time or UTC – as selected in line 13 – is displayed in the graphics.

GS: Gerrity Score. Score calculated from contingency table. Rare events get high weight and frequent events get low weight. Calculation: Mean Event PSS (from Table "Scores for Events"). Properties: high correlation with hits, especially in rare events. May be fluctuating dependent on statistics of events.

HSS: Heidke Skill Score calculated from contingency table. Description: see literature.

PSS: Peirce Skill Score calculated from contingency table. Description: see literature.

Diagonal: fraction of values on the diagonal of contingency table (green cells in table). This may be interpreted as fraction of correct forecasts in respect to defined categories.

Diagonal ±1 class: fraction of values within 1 class from the diagonal of contingency table (small deviation).

Within Forecast Range: Fraction of hours for which the maximum and minimum observations are not lying outside the maximum and minimum forecast values.

FC < OBS: fraction of hours with observations in classes above the forecasts (orange cells in maximum table and yellow cells in minimum table).

FC > OBS: fraction of hours with observations in classes below the forecasts (orange cells in minimum table and yellow cells maximum in table).

- **Scores for Events:**

Events are observations or forecasts below / above one selected threshold, e.g.: wind speed >15kt, visibility <800m, ceiling <1000ft. For wind, maximum values above a threshold will be the most interesting events. For visibility and ceiling, minima below certain limits are important events.

- Event: description of event.
- p(E): event frequency in investigated data.
- H or POD: hit rate, probability of detection. Fraction of observed events that were forecast.
- PC: Proportion of correct forecasts in respect to the selected event criterion.
- F: False Alarm Rate: Frequency of a false forecast in case of a non-event observed.
- FAR: False Alarm Ratio: Frequency of a false forecast in case of an event forecast.
- Bias: Fraction of the frequencies of event forecasts and event observations. Bias >1: more events forecast than observed. Bias <1: less events forecast than observed.
- PSS: Peirce skill score calculated from "collapsed" 2x2 contingency table. High correlation with hits for the event. Calculation: see literature.
- ORSS: Odds Ratio Skill Score calculated from "collapsed" 2x2 contingency table. High correlation with hits for the event. Calculation: see literature.
- HSS: Heidke skill score calculated from "collapsed" 2x2 contingency table. High correlation with false alarms. Calculation: see literature.
- p(E)|E=V: Probability of an event in case it is forecast.
- p(E)|E≠V: Probability of an event in case it is not forecast.
- p(E), p(E)|E=V and p(E)|E≠V are shown in the graphics.
- KPI: For events, the mean of PSS and HSS is calculated. Averaging these values gives the total KPI.

- **Figures:**

  Figure 1: “Probabilities of Events and Dependence on Forecast” is showing the climatological frequency of the events, the frequency of the events when forecast (success ratio) and the frequency of the events when not forecast (miss frequency). This graph is designed for specific customer information who have certain operational thresholds for flights.

  Figures 2, 3 and 4: “Scores along forecast range”

  Figure 2 is showing the KPI along the forecast range of the TAF (lead-time or UTC; as selected in line 13). Fluctuations occur when event frequencies are low for certain thresholds.

  To focus on sufficiently frequent events, Figure 3 (hourly) and Figure 4 (3-hourly) show the scores for selected thresholds. The user can adapt the selection using the switches besides Figure 3. Set “1” for “on” and “0” for “off”. “KPI selected” is the average KPI for the selected events.

  To show / hide curves, double-click the figure, open “Data selection” and select / unselect the requested data rows.

  Figure 5: Performance diagram: It shows the hit rate (POD) dependent on the success ratio SR (1 – FAR) for all (selected) events. Dashed lines in the diagram show the bias. The optimum position in the diagram is at the top right, where POD and SR are 1.

5. Present weather – WX

Result file contain:

- **Contingency tables:**

  Contingency tables are shown for the maximum (=worst) and minimum (=best) hourly forecast versus observed weather conditions.

  Contingency tables can be used for forecaster training to show strengths and weaknesses of forecasts.

  For a detailed description of contingency tables and their calculation see Mahringer, 2008 and cited literature on forecast verification. Classes are predefined in accordance with ICAO Annex 3 significant weather phenomena for TAFs.

  1 – Freezing fog
  2 – Showers in the vicinity: this class is used in order to count forecasts of moderate precipitation correct also if VCSH is observed.
  3 – Moderate liquid precipitation (including showers)
  4 – Blowing or drifting snow, dust or sand
  5 – Moderate solid precipitation (including showers)
  6 – Freezing precipitation
  7 – Thunderstorms, Squall lines
  0 – All other events (mainly “no significant weather”)

- **Total Scores:**

  KPI: Key performance indicator: Summary score. Rare events get high weight and frequent events get low weight. Calculation: Mean Event (PSS+HSS)/2 (from Table “Scores for Events) for all events with p(E) > value defined in cell F8. Properties: high correlation with

hits and false alarms, especially in rare events. The score may be fluctuating dependent on statistics of events. The KPI along forecast lead-time is displayed in the graphics.

Mean ORSS: Mean event Odds Ratio Skill Score.

Mean HSS: Mean event Heidke Skill Score. See Section 4.

Mean PSS: Mean event Peirce Skill Score. See Section 4.

Correct forecasts: Fraction of correct forecasts

- **Scores for Events, Figures:** See Section 4

6. Discussion on Scores

The TAF Verification System is producing a variety of different scores for weather elements and events. The general opinion of verification experts is that single scores alone are never able to account for all aspects in a forecasting system. It is the opinion of the author that scores of the type "percentage of correct forecasts" are not suitable in situations where rare events are forecast.

In TAFs, most interesting events are rare as they are usually observed in a minority of hours. For this reason, the TAF Verification System produces some scores, which are sensitive for hits and false alarms in respect to rare events, as for example the KPI score. On the other hand, this score may be fluctuating, as the behaviour of these events is not the same in each calculation period. Their short-term variability does not necessarily reflect long-term improvements in a forecasting environment, which are interesting from management point-of-views. To track these trends, more long-term verification calculations (e.g. annual scores) are more suitable.

7. References

International Civil Aviation Organization. Meteorological Services for International Air Navigation. Annex 3 to the Convention on International Civil Aviation.

Jolliffe IT, Stephenson DB (Ed.). 2003. Forecast verification: A Practitioner´s Guide in Atmospheric Science. John Wiley & Sons, Chichester, England, UK.

Mahringer, G. 2008. Terminal aerodrome forecast verification in Austro Control using time windows and ranges of forecast conditions. Meteorol. Appl. **15**: 113–123.