This is an excerpt from the report of the CBS COORDINATION GROUP ON FORECAST VERIFICATION (CG-FV) meeting (Reading, UK, 15-17 May 2012).
The full report can be accessed at http://www.wmo.int/pages/prog/www/CBS-Reports/documents/CG-FV_Final-Report_May2012.pdf

Appendix A.II.2.3.1

I – STANDARDIZED VERIFICATION OF DETERMINISTIC NWP PRODUCTS

1. Introduction

This Appendix presents detailed procedures for the production and exchange of a standard set of verification scores for deterministic NWP forecasts produced by GDPFS centres. The goal is to provide consistent verification information on the NWP products of GDPFS participating centres for forecasters in the NMHSs and to help the GDPFS Centres compare and improve their forecasts. Scores will be exchanged between the participating producing centres via the Lead Centre for DNV. The Lead Centre functions, as described in II.2.3.1, include creating and maintaining a website for Deterministic NWP verification information, so that potential users will benefit from a consistent presentation of the results.

The term “deterministic NWP” refers to single integrations of NWP models providing products defining single future states of the atmosphere (as distinct from ensemble prediction systems where multiple integrations provide a range of future states).

The standardized verification should provide key relevant information appropriate to the state-of-the-art in NWP, while being as simple and as easy to implement as possible, and ensuring a consistent implementation across participating centres, in particular in the interpolation to verification grid, and use of a common climatology and set of observations.

2. Verification statistics

The following subsections define two sets of verification statistics. A mandatory set shall be provided by all participating centres. A set of additional recommended statistics is also defined which all centres should provide if possible. The current specifications are for the verification of upper-air fields. The specifications will be expanded as recommended procedures for surface parameters (including 2m Temp, wind, total precipitation) are developed and in response to changing user requirements. The detailed procedures are required to ensure it is possible to compare results from the different participating centres in a scientifically valid manner.

3. Parameters

Extra-tropics

Mandatory

- Mean sea-level pressure (verification against analysis only)
- Geopotential height at 850, 500 and 250 hPa
- Temperature at 850, 500 and 250 hPa
- Wind at 925, 850, 700, 500 and 250 hPa
- Relative humidity at 850 and 700 hPa

Additional recommended

- Geopotential height, temperature, wind at 100 hPa

Tropics

Mandatory

- Geopotential height at 850 and 250 hPa
- Temperature at 850 and 250 hPa
- Wind at 850 and 250 hPa
- Relative humidity at 850 and 700 hPa

4. Forecast times

Scores shall be computed daily for forecasts initialised at 00 UTC and 12 UTC separately. For those centres not running forecasts from either 00 UTC or 12 UTC, scores may be provided for forecasts initiated at other times and must be labelled as such.

5. Forecast steps

Every 12h to the end of the forecast range.

6. Areas

Northern hemisphere extra-tropics	90°N - 20°N, inclusive, all longitudes
Southern hemisphere extra-tropics	90°S - 20°S, inclusive, all longitudes
Tropics	20°N - 20°S, inclusive, all longitudes
North America	25°N–60°N 50°W–145°W
Europe/North Africa	25°N–70°N 10°W–28°E
Asia	25°N–65°N 60°E–145°E
Australia/New Zealand	10°S–55°S 90°E–180°E
Northern polar region	90°N - 60°N, inclusive, all longitudes
Southern polar region	90°S - 60°S, inclusive, all longitudes

Verification against analyses for grid points within each area, including points on the boundary.

7. Verification against analyses

7.1 Grid and interpolation

All parameters shall be verified against the centre's own analysis on a regular 1.5° x 1.5° grid.

In selecting the verification grid, consideration has been given to the variety of resolutions of current global NWP models, the resolved scales of models (several grid-lengths), the resolution of the available climatologies, the potential to monitor long-term trends in performance (including earlier, lower resolution forecasts) and computational efficiency.

Interpolation of higher resolution model fields to the verification grid shall be performed to retain features at the scale of the verification grid but not to introduce any additional smoothing. The following procedures shall be used:

- Spectral fields: truncate to equivalent spectral resolution (T120) for verification grid
- Grid point fields: use area-weighting to interpolate to verification grid

For scores requiring a climatology the climatology is made available via the LC-DNV website on the verification grid and needs no further interpolation.

8. Verification against observations

8.1 Observations

All parameters listed in section 3, except mean sea-level pressure, shall be verified against a common set of radiosondes. The list of radiosonde observations for each area is updated annually by the CBS Lead Centre for radiosonde monitoring. The chosen stations' data must be available to all the centres and be of sufficient quality on a regular basis. Consultation with all centres (usually by electronic mail) is desirable before establishing the final list. The current list is available via the website of the LC-DNV. The LC-DNV will contact all participating centres when the new list is available and inform them of the date from which the new list shall be used.

The observations used for verification shall be screened to exclude those with large errors. In order to do this, it is recommended that centres exclude values rejected by their objective analysis. Moreover, centres which apply a correction to the observations received on the GTS to remove biases (e.g. radiation correction), should use the corrected observations to compute verification statistics.

8.2 Interpolation

Verification shall be made using the nearest native model grid point to the observation location.

8.3 Areas

The networks used in verification against radiosondes consist of radiosonde stations located in the areas listed in Section 6.

The list of radiosonde stations to be used for each area is updated annually by the CBS Lead Centre for radiosonde monitoring (see subsection 8.1)

9. Scores *[move to Part II, 1.4]*

The following scores are to be calculated for all parameters against both analysis and (except mean sea-level pressure) observation.

Wind

Mandatory:

- rms vector wind error
- mean error of wind speed

Other parameters:

Mandatory

- Mean error
- Root mean square (rms) error
- Correlation coefficient between forecast and analysis anomalies (not required for obs)
- S1 score (only for MSLP and only against analysis)

Additional recommended

- mean absolute error
- rms forecast and analysis anomalies (not required for observations)
- standard deviation of forecast and analysis fields (not required for observations)

9.1 Score definitions

The following definitions should be used

Mean error $$M = \left(\sum_{i=1}^{n} w_i (x_f - x_v)_i \right) / \sum_{i=1}^{n} w_i$$

Root mean square (rms) error $$rms = \sqrt{\sum_{i=1}^{n} w_i (x_f - x_v)_i^2} \; / \sqrt{\sum_{i=1}^{n} w_i}$$

Correlation coefficient between forecast and analysis anomalies

$$r = \frac{\sum_{i=1}^{n} w_i (x_f - x_c - M_{f,c})_i (x_v - x_c - M_{v,c})_i}{\left(\sum_{i=1}^{n} w_i (x_f - x_c - M_{f,c})_i^2 \right)^{1/2} \left(\sum_{i=1}^{n} w_i (x_v - x_c - M_{v,c})_i^2 \right)^{1/2}}$$

rms vector wind error $$rms = \sqrt{\sum_{i=1}^{n} w_i (\vec{V}_f - \vec{V}_v)_i^2} \; / \sqrt{\sum_{i=1}^{n} w_i}$$

Mean absolute error $$MAE = \sum_{i=1}^{n} w_i \left| x_f - x_v \right|_i / \sum_{i=1}^{n} w_i$$

rms anomaly $$rmsa = \sqrt{\sum_{i=1}^{n} w_i (x - x_c)_i^2} \; / \sqrt{\sum_{i=1}^{n} w_i}$$

standard deviation of field $$sd = \sqrt{\sum_{i=1}^{n} w_i (x - M_x)_i^2} \; / \sqrt{\sum_{i=1}^{n} w_i} \quad \text{where} \quad M_x = \sum_{i=1}^{n} w_i x_i$$

S1 score $$S_1 = 100 \frac{\sum_{i=1}^{n} w_i (e_g)_i}{\sum_{i=1}^{n} w_i (G_L)_i}$$

Where:

x_f	=	the forecast value of the parameter in question
x_v	=	the corresponding verifying value
x_c	=	the climatological value of the parameter
n	=	the number of grid points or observations in the verification area

$M_{f,c}$ = the mean value over the verification area of the forecast anomalies from climate

$M_{v,c}$ = the mean value over the verification area of the analysed anomalies from climate

$\vec{V}_f$ = the forecast wind vector

$$e_g = \left\{ \left| \frac{\partial}{\partial x}\left(x_f - x_v\right) \right| + \left| \frac{\partial}{\partial y}\left(x_f - x_v\right) \right| \right\}$$

$$G_L = \max\left(\left| \frac{\partial x_f}{\partial x} \right|, \left| \frac{\partial x_v}{\partial x} \right| \right) + \max\left(\left| \frac{\partial x_f}{\partial y} \right|, \left| \frac{\partial x_v}{\partial y} \right| \right)$$

where the differentiation is approximated by differences computed on the verification grid.

The weights w_i applied at each grid point or observation location are defined as

Verification against analyses: $w_i = \cos\phi_i$, cosine of latitude at grid point i

Verification against observations: $w_i = 1/n$, all observations have equal weight

10. Exchange of scores

Each centre shall provide scores monthly to the LC-DNV. Details of the procedure and the required format for the data are provided on the website of the LC-DNV. All scores (12-hourly) for all forecasts verifying within a month shall be provided as soon as possible after the end of that month.

11. Climatology

To ensure consistency between results from different centres a common climatology shall be used for those scores requiring a climatology. All centres shall use the climatology provided via the LC-DNV website.

A daily climatology of upper-air parameters are available for both 00 UTC and 12 UTC. This provides an up-to-date estimate of climate characteristics for each day of the year, including climate mean, standard deviation and selected quantiles of the climate distribution. These latter statistics are required for the CBS standardized verification of EPS forecasts.

The data is made available in Grib format. Information on access to the data and further documentation are provided on the LC-DNV website.

12. Monthly and annual averaged scores

Where average scores are required over a defined period, the averaging shall be made using the following procedures:

Linear scores (mean error, mean absolute error) - mean
Non-linear score should be transformed to appropriate linear measure for averaging
mean of MSE;
Z-transform for correlation

For a defined period, the average shall be computed over all forecasts verifying during the period. Averages shall be computed separately for forecasts initiated at 00 UTC and 12 UTC and both sets of average values provided.

13. Confidence Intervals

Confidence intervals will be computed by the LC-DNV using the daily scores. The method used will be documented on the LC-DNV web site.

14. Documentation

Participating centres shall provide to the LC-DNV information on their implementation of the standardized verification system annually, shall confirm to the LC-DNV any changes to its implementation (including the annual change of station list, changes in additional statistics) and changes in their NWP model.