WORLD METEOROLOGICAL ORGANIZATION

COMMISSION FOR BASIC SYSTEMS
OPAG on DPFS

MEETING OF THE CBS (DPFS) TASK TEAM ON SURFACE VERIFICATION

GENEVA, SWITZERLAND
20-21 OCTOBER 2014

DPFS/TT-SV/Doc. 4.2

(X.IX.2014)

Agenda item : 4.2

ENGLISH ONLY

Joint Working Group for Forecast Verification Research perspective on global surface verification plans

(Submitted by Laurie Wilson and Marion Mittermaier,

co-chairs of JWGFVR)

Summary and purpose of document

This document provides a commentary from the WWRP working group on the draft proposal circulated to the group.

Action Proposed

The meeting is invited to discuss the comments provided with the view of potentially amending the proposed framework, where appropriate.

Reference: -

G. Candille, C. Côté, P. L. Houtekamer and G. Pellerin, 2007: Verification of an Ensemble Prediction System against Observations, *Monthly Weather Review*, Vol. 135, pp. 2688–2699.

Input is provided as per the headings of the proposal headings.

The group asks whether the verification will be done in one place, e.g. by the lead centre, or by individual global modelling centres with an exchange of scores.

1. Introduction

The group suggests that the document states clearly that this proposal covers surface fields *from global models.*

3. Parameters

- Would it be better to use 2-m specific humidity instead of relative humidity? Or perhaps dewpoint?
- Should smaller accumulation periods (6h, 12h) be considered for precipitation?

The group feels that the above, RH and sub-daily precipitation, should be encouraged, but not mandatory since, especially for precipitation, there are far fewer sub-daily measurements than there are daily measurements.

6. Grid and interpolation

Nearest grid-point is probably a good choice for precipitation. Bilinear interpolation may be more suitable for 2-m temperature and humidity, though with improving grid resolution the need for this is less clear. To keep it as simple as possible, and avoid differences in the implementation of interpolation schemes between centres (which plagued the upper-air scores for years), perhaps it is advisable to use the nearest grid point for all quantities.

7. Observations

Achieving consistency will be challenging. The GTS has to be the approved source of observations which is accessible to all. Whilst work at the Met Office (and elsewhere) has explored various methods it is clear that though the GTS is the common source of observations for all centres, considerable differences exist between centres, in terms of the observations actually received, and the positional information available. Quality control and ensuring uniformity is non-trivial, e.g. Met Office uses the data assimilation system for real-time quality control, as surface observations are assimilated. QC procedures will vary from centre to centre, and we suspect it will not be easy to standardise these.

8. Areas

Areal averages are preferred for exchanging scores, as exchanging scores for individual sites would be too problematic. This could be possible if one centre did all the calculations. It will be important to aggregate over (relatively) homogeneous regions.

9. Scores

Other than SEEPS the other scores require thresholds to be set, which could be challenging, to achieve climatological consistency, especially for aggregation.

10. Exchange

Monthly would be consistent with what is currently done. But is this too frequent for surface parameters? How about 3-monthly or 6-monthly? See also 12 below.

11. Climatology

Whilst it is recognised that SEEPS is very useful for comparing and aggregating scores which cover many climatic regions, are scores which require climatology adding too much complexity? Even if the climatology is provided by one centre, to be used by everyone, the Met Office has shown that it is far from straightforward to use such a 3rd party climatology.

12. Temporal and spatial aggregation

A period of 1 month may not be sufficient to obtain meaningful scores, e.g. in the case of precipitation. Should 3-month or 6-month periods be used? e.g. results for DJF, MAM, JJA and SON.

If contingency table based scores are included, exchanging the contingency table counts (hits, misses, etc.), would be better as they would easily support longer aggregation.

13. Confidence intervals

The recommended block length is taken from upper air recommendations. Is there a good reason to propose a different block length? Based on experience with precipitation, confidence intervals were not particularly sensitive to choice of block length, though this may be different for temperature, and needs checking.

14. Observation errors

It would be nice to take this into consideration, but increases the complexity of the recommendations, which is probably unrealistic. A key related issue is that the location information of sites *must* be improved, so that the true nearest model grid points can be extracted. Lack of precision in this information may make it hard for code to extract the nearest grid point to an observing site, and could introduce an error far greater than any other observation error.

It is nearly impossible to back out "true" model error, but reporting observation errors, when known, could be an important step towards recognising the existence of observation errors.