Guidance for validating the HEFS at a small number of forecast locations

1. Purpose

This document suggests a minimum or "baseline" approach to validating the HEFS at two forecast locations in each RFC, as required by the NWS Annual Operating Plan (AOP) for Fiscal Year 2015 (FY15). For more general guidance on conducting hindcasting and verification, see the HEFS documentation (including the Hindcasting Guide) at: http://www.nws.noaa.gov/oh/hrl/general/indexdoc.htm

2. Basin selection

There is a milestone for the RFCs to validate the HEFS at **two headwater basins** in FY15, but other basins should be included as time allows. More complicated scenarios will be considered later; these include downstream basins and basins for which upstream regulations are important. Preferably, for simplicity, the two headwater basins should not contain sub-basins, as this will require additional CHPS configuration, increase hindcasting time, and will complicate the validation somewhat.

The selected basins should have sufficiently long and uninterrupted records of observations (MAP, MAT, QIN or QME), hydrologic simulations (SQIN), and precipitation and temperature forecasts for use in the MEFP. For example, when producing HEFS hindcasts with forcing inputs from the GEFS, the forcing and streamflow observations should be available for most of the period between 1985 and 2012 (ideally, much longer for MAT and MAP, because the MEFP produces one ensemble member for each historical year of record). Of course, some data may be missing for particular historical dates.

3. Suggested scenarios

The validation should include the MEFP temperature and precipitation forecasts and the HEFS streamflow forecasts. **The first priority is to conduct validation over a suitably long historical period**, with a forecast issued at 12Z each day, as this will support an evaluation of the HEFS at reasonably high precipitation and streamflow thresholds. It is also desirable to validate a HEFS configuration that reflects, as closely as possible, the proposed operational configuration in the selected basins. In practice, however, this may not be possible, given the short historical period for which the RFC and WPC forcing is available and the limited frequency of the CFSv2 reforecasts (every 5 days).

In keeping with the first priority, **it is recommended that the baseline configuration includes the GEFS forecasts only**, for which a suitably long historical record is available (01/01/1985-07/30/2012). By implication, the validation will consider a forecast horizon of 1-15 days. Time permitting, other validation scenarios may be considered. For example, if the operational configuration includes the RFC/WPC forcing, a separate hindcast run may be conducted for a shorter period (for which the RFC/WPC forcing is available). If the EnsPost is used operationally, it should be included in the hindcasting and validation. In order to measure forecast skill, a reference forecast will be required, and "MEFP climatology" is recommended. The streamflow forecasts generated with forcing inputs from "MEFP climatology" are analogous to those from Ensemble Streamflow Prediction (ESP). The EnsPost should not be applied to the reference forecasts. Thus, the recommended scenarios are:

1. MEFP precipitation, MEFP temperature, and HEFS streamflow forecasts (with EnsPost, if applicable) with forcing inputs from the **GEFS only**, a forecast horizon of 1-15 days (i.e. `gefsNumberOfForecastDays=15`), and a historical period of 1985-2012.
2. MEFP precipitation, MEFP temperature and HEFS streamflow forecasts (**without EnsPost**) with forcing inputs from **MEFP climatology only**, a forecast horizon of 1-15 days (i.e. `climatologyNumberOfForecastDays=15`), and a historical period of 1985-2012.

4. Data exports for verification

The following hindcast data exports are recommended:

1. MEFP precipitation and temperature forecasts
2. HEFS streamflow forecasts without EnsPost (the same SQIN export used to calibrate EnsPost)
3. If applicable, the HEFS streamflow forecasts with EnsPost

To begin with, it is recommended that the forecast files are exported in PI-XML format. While these files are much larger and more time consuming to read and write than Fast Infoset/binary, the PI-XML format is human readable and, therefore, simpler to quality control. When specifying a directory to read in the EVS, **all** files and subdirectories will be processed unless a filter (e.g. .xml) is defined (see `2b. Identify input data sources > More > Other options > Forecast file filter`).

5. Verification with the EVS

In order to verify the forcing and streamflow forecasts, a suggested EVS configuration is provided alongside this guide. The configuration includes two default basins (XXXXX and YYYYY), with references to their default data sources, and some suggested verification options. **These configurations will need to be adjusted to reflect the actual basins and datasets available**. They may be further adjusted to support any early applications of the HEFS at each RFC (e.g. additional validation thresholds or forcing sources). However, the verification should consider a broad range of conditions; it should not focus on a narrow range of thresholds or a specific time period or season. The suggested configurations include three separate EVS project files:

1. Default_precipitation.evs
2. Default_temperature.evs
3. Default_streamflow.evs

In principle, these could be combined into a single EVS project file, but a shorter configuration is simpler to diagnose when problems occur. For this reason, it is generally recommended (but certainly not compulsory) to produce a separate EVS file for each basin and variable that will be validated. Once validation has been implemented successfully at several locations, a more efficient configuration may be considered (e.g. a single EVS file for one variable across an entire forecast group).

While the HEFS outputs are 6-hourly or less, the streamflow observations are typically only available at a 1-day aggregation. Also, a 1-day aggregation is arguably more meaningful for streamflow verification, as it filters out intra-day variability/noise. **It is, therefore, recommended that the temperature, precipitation and streamflow forecasts are all validated at a temporal aggregation of 1 day**; that is, total daily precipitation, daily average temperature and daily average streamflow.

The following naming convention is used for the Verification Units (VUs) (i.e. `2a. Set unit identifiers`), but many others are possible:

- "Location identifier": the basin identifier (e.g. FTSC1) with placeholders of XXXXX/YYYYY
- "Environmental variable identifier": one of Precipitation, Temperature or Streamflow
- "Additional identifier": the forcing and EnsPost streamflow scenarios (e.g. GEFS_EnsPost)

The suggested configuration includes one VU for the main operational forecast (GEFS) and one VU for the baseline forecast with climatological forcing (CLIM). In addition, the streamflow configuration includes one VU for the raw streamflow forecasts (GEFS), one VU for the forecasts with EnsPost (GEFS_EnsPost), and one VU to measure the "value added" by the EnsPost (GEFS_EnsPost_Residual), where the EnsPost forecasts are benchmarked against the raw streamflow forecasts. **If the EnsPost is**

not being used operationally, the VUs for the EnsPost should be deleted. The configurations are likely to require several updates. As a minimum, the following will need to be updated in each file:

- The basin identifiers (`2a. Set unit identifiers`)
- The observed and forecast data sources (`2b. Identify input data sources`)
- The output path (`2d. Set location for output data`)

In addition, depending on RFC, the following may need to be adjusted:

- The file formats for the forecasts or observations (e.g. Datacard for the observations)
- The "Forecast file variable ID" and/or the "Observed file variable ID". These options are accessed via `2b. Identify input data sources` > `More` > `Other options`. For example, if the precipitation forecasts are contained in PI-XML files with a "parameterId" of MAPX, the "Forecast file variable ID" must be MAPX. Also, if a forecast file contains multiple "ensembleId" entries, the desired identifier should appear in the "Forecast file ensemble ID"
- The forecast or observed scale information, which is access through `2b. Identify input data sources` > `More`. For example, if a unit change is required to pair the forecasts and observations (e.g. if the forecasts are in MM and the observations are in INCH), the "Target attribute units" of **either** the forecasts **or** the observations will need to be set, depending on the desired units for the outputs. The "Attribute units" (i.e. the existing units) should be set in **all** cases, even when a unit change is not required
- The "Forecast time zone" under `2b. Identify input data sources`
- The "Observed time zone" under `2b. Identify input data sources` (see below)
- The "First lead time for aggregation of forecasts [hours]" under `2b. Identify input data sources` > `More` > `Pairing options` (see below)

6. Pairing the forecasts and observations for verification

Pairing of the MEFP temperature and precipitation forecasts with corresponding observations is generally straightforward, as the forecasts and observations are both available at a 6-hourly timestep. If the MAP/MAT observations are *not* available at the synoptic times of {0Z,6Z,12Z,18Z}, a false "Observed time zone" should be used to minimize any differences (CST in most cases). This will introduce a small timing error, but the effects should be negligible at a temporal aggregation of 1 day.

When verifying streamflow, it may not be straightforward to pair the forecasts and observations. In most cases, the streamflow observations will be 1-day averages (QME) from midnight to midnight in local time, whereas the HEFS streamflow forecasts may be 1-, 3- or 6-hourly instantaneous flows, with a forecast issue time of 12Z. In order to pair the forecasts and observations, a false time zone may be required for the "Observed time zone" and some of the earlier forecasts may need to be skipped. While undesirable, these assumptions are unavoidable without a more complicated pairing strategy (e.g. disaggregation and re-aggregation of the QME onto a 12Z-12Z clock), which is beyond the scope of this guide. If the forecasts are 1- or 3-hourly, there is greater flexibility to reduce timing errors in the pairs. When the forecasts are 6-hourly and the observations are 24-hourly, the default scenario is:

- Use <u>Central Standard Time</u> for the "Observed time zone". If the observed flows are contained in a PI-XML file, the `<timeZone>` must be CST (i.e. <u>-6</u>). For files formats that comprise a time zone, the EVS entry ("Forecast time zone") is cross-checked against the file entry
- The "First lead time for aggregation of forecasts [hours]" should be <u>24</u>

In the above scenario, the EVS will interpret the observed streamflow as an average over the period 6Z-6Z. By specifying a "First lead time for aggregation of forecasts [hours]" of 24, the EVS will conduct a four-point average of the forecasts with lead times of {24,30,36,42} hours, ending at 6Z each day (in keeping with the observations). These averages will continue throughout the forecast horizon from 6Z-6Z each day. Other approaches are possible (using CHPS or external tools), but they are beyond the scope of this guide. As indicated above, there is greater flexibility to minimize pairing errors when the observations are 6-hourly or less and when the forecast timestep is 1-hourly or 3-hourly (except in CST, where the forecasts and observations are naturally aligned). Known departures from the default pairing for streamflow are as follows:

- **CBRFC (MST):** the forecasts are 1-hourly and QME is available from 7Z-7Z. The first forecast that coincides with 7Z-8Z has a lead time of 19-20 hours. Thus, the "Observed time zone" should be MST and the "First lead time for aggregation of forecasts [hours]" should be 20
- **CNRFC (PST):** the forecasts are 1-hourly and QME is available from 8Z-8Z. The first forecast that coincides with 8Z-9Z has a lead time of 20-21 hours. Thus, the "Observed time zone" should be PST and the "First lead time for aggregation of forecasts [hours]" should be 21

7. Interpretation of the verification results

Interpretation of the verification results should focus on the HEFS forecasts being "reasonable" (i.e. without gross errors) at a range of precipitation, temperature, and streamflow thresholds, rather than a detailed evaluation of the different attributes of forecast quality.

In practice, the forecast quality will vary in space and time, depending not only on the correct implementation of the HEFS, but on river basin conditions, the quality of the hydrologic model calibration, and the quality of the raw forcing (e.g. from the GEFS), among other factors. The aim here is simply to "screen" the HEFS forecasts and to identify any major problems that would compromise early applications of these forecasts, whether by NWS or others (and for any given purpose, rather than a specific application). The forecasts should be screened by evaluating the following EVS outputs:

1. MEFP precipitation and temperature forecasts
2. HEFS streamflow forecasts without the EnsPost
3. If applicable, the HEFS streamflow forecasts with the EnsPost (including the "value-added")

While all the metrics should be explored, any major problems should be particularly apparent when examining:

1. Box-plots of forecast errors by observed value. These plots are contained in files whose names include ".GEFS_Modified_box_plots_per_lead_time_by_observed_value". There is one plot for each forecast lead time. Any major problems should be readily apparent in these plots. For example, the HEFS forecasts should consistently capture the observed value (i.e. the forecast spread should capture the zero error line), except, perhaps, at the largest observed flows
2. Forecast skill relative to climatology. In particular, the mean Continuous Ranked Probability Skill Score (CRPSS). The CRPSS is contained in a file whose name includes "GEFS.Mean_continuous_ranked_probability_**skill**_score" (NOTE: there are two metrics with similar names; the CRPSS contains "skill" in the name). The HEFS forecasts should significantly improve upon climatology, particularly during the first few days. This will be evidenced by positive values of the CRPSS (>0 and <=1). Consistently negative values of the CRPSS would indicate a problem, although negative values are more likely for very low or high thresholds at long forecast lead times. The CRPSS is shown for multiple thresholds, as well as "All data"
3. If applicable, the value added by the EnsPost when measured against the raw streamflow forecasts. In particular, the skill measured by the CRPSS, which is contained in a file whose name includes "GEFS_EnsPost_Residual.Mean_continuous_ranked_probability_**skill**_score"