Hydrologic Ensemble Forecasting Service (HEFS)

Seminar B
What’s New for HEFSv1

Hank Herr

HEFS Training Workshop (4), Silver Spring, MD
September 19, 2013
HEFS Components

- **MEFP: pre-processor**
  - Raw weather and climate forecasts (GEFS, CFSv2,..)

- **MEFP PE: parameters**

- **EVS: verification**
  - Unbiased forcing (basin scale)

- **GraphGen: products**
  - "Corrected flow"

- **EnsPost: post-processor**
  - "Raw flow"

- **EnsPost PE: parameters**

- **Verification results**
  - Ensemble products

**Legend**:
- = Forecast tool (real-time/hindcast)
- = Supporting tool
- = Future capability

**Data assimilator**

**Hydrologic Ensemble Processor**

**Hydrologic data**
Overview of HEFSv1 Changes

- **Diagnostics**
  - Tools have been added to allow for better QC of historical and RFC archived forecast/observed time series used as input to MEFPPE
    - Use MEFPPE to identify bad data
    - Use external tools to fix bad data
  - MEFPPE parameter diagnostic tools have been added
    - Based on earlier XEFS EPP3 work with John Schaake and Rob Hartman
    - Working on guidance for how to use diagnostic tools in decision making with MEFPPE

- **Other changes include configuration changes, bug fixes, and enhancements related to the science and software of HEFS**
  - Behavior of and output from MEFPPE, EnsPostPE, MEFP, and EnsPost all change as a result of this release
    - Example: By default, modulation events are no longer included when MEFP generates ensembles
Required Actions for HEFSv1

- Perform configuration changes
  - May have already been done
  - Read the release notes
  - Summarizing list provided in Appendix of these slides

- Re-estimate all parameter files
  - Use your existing parameter estimation stand-alone (with updated jars)
  - Perform QC for historical and RFC archived forecast/observed data
    - Described later
  - MEFP temperature parameters
    - GEFS reforecast data read incorrectly in previous release (FogBugz 1166)
  - MEFP precipitation parameters
    - CFSv2 used fixed ‘Data Window’ in previous release (FogBugz 1213)
  - EnsPost parameters
    - Bug fix related to how CDFs are computed in the EnsPostPE
Required Actions for HEFSv1

- Configuration changes when using raw climatology
  - CHPS transformations should be used to append raw climatology to the end of MEFP generated ensembles
    - Instructions in *MEFP Configuration Guide: Forecast Components Section 6.1.3*
  - Resampled climatology will still be generated using MEFP
  - Applies to MARFC and NERFC for NYCDEP

- MEFP adapter run file property change if ensemble members are constructed based on calendar year instead of water year
  - Set the memberIndexingYear property to be “calendarYear”
  - Hydrologic water year (Oct 1 – Sep 30) is the default setting
    - Matches standard ESP climatology forecasting
  - Applies to MARFC and NERFC for NYCDEP
    - To be consistent with the HEFS hindcasts generated at OHD
Things to Watch Out For

- The default settings for a new adapter run file property behaviorIfEventMissing changes the behavior of MEFP
  - If a canonical event cannot be computed for a source using the input time series provided due to missing data, an error is thrown
    - More sensitive to missing data
    - Old data more likely to yield errors in adapter run due to missing values at end of the time series
  - The option ‘fillClimatology’ will make it run as before, but then it is easier for bad/old data to be processed by MEFP and go unnoticed

- Preprocessing workflows (import and spatial interpolation) may not error out if data to import is not provided
  - Ex: GEFS workflow will use old data in spatial interpolation if import module fails to import any data
    - Error will not occur until MEFP uses the data as input (see above)
Time Series Diagnostics

Why?

MTRN6LWR Datacard Snippet

<table>
<thead>
<tr>
<th>Date</th>
<th>TMIN</th>
<th>TMAX</th>
<th>TMRN</th>
<th>TMRN</th>
<th>TMRN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1828146</td>
<td>27.135</td>
<td>13.485</td>
<td>3.495</td>
<td>33.280</td>
<td>33.280</td>
</tr>
<tr>
<td>1828147</td>
<td>39.500</td>
<td>33.280</td>
<td>33.280</td>
<td>33.280</td>
<td>33.280</td>
</tr>
<tr>
<td>2828148</td>
<td>33.585</td>
<td>33.280</td>
<td>33.280</td>
<td>33.280</td>
<td>33.280</td>
</tr>
<tr>
<td>2828149</td>
<td>33.280</td>
<td>33.280</td>
<td>36.708</td>
<td>33.280</td>
<td>33.280</td>
</tr>
<tr>
<td>2828150</td>
<td>33.280</td>
<td>33.280</td>
<td>33.280</td>
<td>33.280</td>
<td>33.280</td>
</tr>
<tr>
<td>2828151</td>
<td>33.280</td>
<td>33.280</td>
<td>33.280</td>
<td>33.280</td>
<td>4.869</td>
</tr>
<tr>
<td>2828152</td>
<td>33.280</td>
<td>33.280</td>
<td>-2.398</td>
<td>33.280</td>
<td>33.280</td>
</tr>
<tr>
<td>2828153</td>
<td>7.537</td>
<td>33.280</td>
<td>33.280</td>
<td>33.280</td>
<td>33.280</td>
</tr>
<tr>
<td>2828154</td>
<td>33.280</td>
<td>33.280</td>
<td>33.280</td>
<td>33.280</td>
<td>33.280</td>
</tr>
</tbody>
</table>

Many repeated 33.280 values
Does not fit diurnal pattern

Unreasonable Spikes

Historical TMIN/TMAX

TMIN Spike

MEFP Output

Unreasonable Spikes
Time Series Diagnostics
Why?

MTRN6LWR Datacard Snippet

<table>
<thead>
<tr>
<th>Time</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
<th>Value 4</th>
<th>Value 5</th>
<th>Value 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>182847</td>
<td>5.417</td>
<td>16.394</td>
<td>23.322</td>
<td>15.514</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2828152</td>
<td>1.337</td>
<td>10.990</td>
<td>16.736</td>
<td>5.610</td>
<td>0.384</td>
<td>8.048</td>
</tr>
<tr>
<td>2828153</td>
<td>12.795</td>
<td>5.378</td>
<td>1.866</td>
<td>11.441</td>
<td>17.117</td>
<td>8.858</td>
</tr>
<tr>
<td>2828154</td>
<td>4.936</td>
<td>18.668</td>
<td>26.480</td>
<td>12.085</td>
<td>5.312</td>
<td>22.893</td>
</tr>
</tbody>
</table>

Copied from another year to fit the diurnal pattern

Only one spike left

MEFP Output
Time Series Diagnostics (Demo)

- Quality control historic/archive data
  - Applicable to **Historical Data Panel** and **RFC Forecasts Panel** within the MEFPPE
  - Checks are made when the steps are performed (or data imported):
    - Missing data
    - Gross range (0 – 100 mm for precip, -100 – 100 degC for temperature)
    - Minimum temperature exceeds maximum temperature
    - Check results are stored in a file within the MEFPPE run area
  - **Location Summary Panels** are enhanced
    - Icon has been added to display for which locations historical or RFC archived forecast data is questionable
  - **Diagnostics Panel** is enhanced
    - Questionable data is highlighted by light red marks/zones within graphic
  - GraphGen chart viewing panel enhanced to allow for more easily navigating and examining questionable data values
Time Series Diagnostics (Demo)

Quality control steps

- Upon installing and configuring HEFSv1 in the parameter estimation stand-alone, perform the steps associated with the Historical Data Panel and RFC Forecasts Panel again so that MEFPPE can detect questionable data.

- Examine the Location Summary Panels to identify locations that have questionable data.

- Use the Diagnostics Panel to examine historical data and archived RFC forecast data to determine if questionable data is bad data.

- Replace bad data with reasonable data values at the source.
  - Historical time series
    - Modify datacard files and re-import the data into CHPS—or—use the FEWS Data Editor:
      [Link](https://publicwiki.deltares.nl/display/FEWSDOC/04+Data+Display+and+Data+Editor#04DataDisplayandDataEditor-DataEditor)
  
  - RFC archived forecast time series with observations
    - Modify data in archive database
    - Copy RFC forecast files, modify by hand, and import the files

- Perform step again to incorporate fixed data (or import RFC data).
Estimated Parameters Diagnostics (Demo)

- Diagnostics added to allow for a cursory examination of estimated parameters
  - Based on diagnostics employed by John Schaake and Rob Hartman
  - Block plot displays parameters in three dimensions (day of year, event, and parameter value)
    - Parameter value displayed based on color
    - Table view also available with same color coding for background of cells
  - In addition to viewing raw parameter values, the following are allowed:
    - Displaying difference between observed and forecast event means
    - Displaying difference between correlations across forecast sources
  - Clicking on blocks or table cells opens up a scatter plot of the raw canonical event values used in parameter estimation
  - Tools provided for navigating the chart and table and for selecting events for which to display values
Questions?
Appendix
List of Changes for HEFSv1

(see the release notes for a complete list with full details)
Changes for HEFSv1
MEFP/MEFPPE Configuration Changes

- **Grids.xml**
  - `<y>` changed to 50 from 56 (FogBugz 1086)

- **MEFP_MAP_to_GMT.xml**
  - Set `<ignoreMissing>` flag to false on sample transformation

- **MEFP_MAT_to_TAMN_TAMX.xml**
  - Set `<ignoreMissing>` flag to false on sample transformation

- **MEFPPE.xml** (PI-service configuration)
  - Includes GMT translated MAT time series query

- **MEFP_GEFS_TFMN_6to24.xml, MEFP_CFSv2_TFMN_6to24.xml**
  - Fixed so that minimum is being computed (FogBugz 1195)

- **FGroup_MEFP_CFSv2*_LaggedEnsemble.xml**
  - Removed some configuration causing warnings (FogBugz 1202)
Changes for HEFSv1
MEFP/MEFPPE Configuration Changes

- MEFP_Forecast.xml, MEFP_FMATS_Forecast.xml (workflows)
  - Changes made to allow for running diurnal pattern module in ensemble mode using multiple processors to improve performance

- IMPORTANT: Raw climatology option in MEFP will be removed in the near future
  - For HEFSv1, resampled climatology will be the default
  - Reason:
    - MEFP does not output true raw climatology for temperature (MAT)
    - Capability is already available through CHPS transformations
  - It is recommended that CHPS transformation be used to append raw climatology to MEFP output if raw must be used
    - See Section 5.1.3 of the MEFP Configuration Guide: Forecast Components
Changes for HEFSv1
Bug Fixes

- **MEFP/MEFPPE**
  - Fixed problem involving not removing tmp files (FogBugz 1091)
  - Fix to lessen chance of run-time information file being partially written when CHPS/MEFPPE crashes for any reason (FogBugz 1093)
  - Included missing jar in release (FogBugz 1119)
  - Logic problem due to the first day of the year with successfully computed parameters being the last day (FogBugz 1124)
  - RFC forecast source fixed to include correct unit in files constructed from archive database data (FogBugz 1133)
  - TMAX GEFS data being assigned to TMIN and vice versa (FogBugz 1166)

- **EnsPost/EnsPostPE**
  - Bug fix related to how CDFs are computed in the EnsPostPE
  - Bug fix related to how the CDFs and parameters are applied in EnsPost
Changes for HEFSv1
Small Enhancements

- MEFP/MEFPPE
  - Proper default start/end years assigned to the RFC, GFS, GEFS, and CFSv2 forecast source estimation options (FogBugz 1094)
  - Improved messaging related to failures in estimating parameters (for example, insufficient data) (FogBugz 1106)
  - Added Default Button for each estimation option; clicking the button recovers the delivered default setting (FogBugz 1186)
  - Run file property memberIndexingYear added (FogBugz 1200)
    - Other run file properties also added (FogBugz 1213)
  - Various other changes tracked in FogBugz 1213
    - Modulation events removed from default canonical event list and are not included by default when MEFP generates ensembles
    - behaviorIfEventMissing flag added to adapter properties (defaults to erroring out if an event cannot be computed)
    - CFSv2 pairing window uses standard estimation option instead of fixed window
Changes for HEFSv1
Small Enhancements

- EnsPost/EnsPostPE
  - Added an option to allow for outputting the model time-scale 24-hour results from EnsPost, instead of the 6-hourly disaggregated values (FogBugz 1159)
  - Various changes tracked in FogBugz 1191