A Pilot Study for Short-Range Ensemble Rainfall Forecasting: Post-Processing, Verification and Prototype Implementation

D.-J. Seo\textsuperscript{1}, Ridwan Siddique\textsuperscript{1}, Geoffrey Shaughnessy\textsuperscript{2}, Scott Huebner\textsuperscript{2}, Sarah Noorjahan\textsuperscript{3} and John Raymond\textsuperscript{3}

\textsuperscript{1}Dept. Of Civil Eng., The Univ. of Texas at Arlington
\textsuperscript{2}Operation Control, SFWMD
\textsuperscript{3}Hydro data Management, SFWMD
### Synopsis:
Widespread moderate rains focused southwest through north of the Lake. A squall line of showers and embedded thunderstorms is moving through the District this morning. This forecast is reduced mainly to a timing exercise. Look for thunderstorms to decrease some this afternoon as the squall outruns its upper level support; but, some seabreeze enhancement will continue to aid development through late afternoon. Activity will exit eastward as the trailing front pushed through early evening. Chilly and breezy conditions will invade overnight through tomorrow with dry and cool conditions then persisting through the weekend.

<table>
<thead>
<tr>
<th>Quantitative Precipitation Forecast</th>
<th>24Hr Begin Thu</th>
<th>Local Max</th>
<th>24Hr Begin 7AM Fri</th>
<th>24Hr Begin 7AM Sat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper Kissimmee</td>
<td>.65</td>
<td>1.5”</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>Lower Kissimmee</td>
<td>.65</td>
<td>1.5”</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>Lake Okeechobee</td>
<td>.55</td>
<td>1.5”</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>Eastern Agricultural Areas</td>
<td>.45</td>
<td>1”</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>Western Agricultural Areas</td>
<td>.55</td>
<td>1”</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>Conservation Areas 1&amp;2</td>
<td>.40</td>
<td>1”</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>Conservation Area 3</td>
<td>.40</td>
<td>1”</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>Martin/St Lucie Counties</td>
<td>.45</td>
<td>1”</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>Eastern Palm Beach County</td>
<td>.40</td>
<td>1”</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>Eastern Broward County</td>
<td>.30</td>
<td>1”</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>Eastern Miami-Dade County</td>
<td>.25</td>
<td>&lt;1”</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>East Caloosahatchee</td>
<td>.70</td>
<td>1.5”</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>Big Cypress Preserve</td>
<td>.60</td>
<td>1”</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>Southwest Coast</td>
<td>.75</td>
<td>1.5”</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>District Overall</td>
<td>0.54</td>
<td>-</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
DAY-1 QPF, WET SEASON, UK

FORECAST RAINFALL (IN)

OBSERVED RAINFALL (IN)

1998-2011
2-DAY QPF, WET SEASON, EAAE

1998-2011
2-DAY QPF, WET SEASON, UK

1998-2011

Fig. 1. (a) Scatter plots of 6-h MAP and QPF for Day 1 for HUNP1 (1 August 2002–12 July 2008), TIFM7 (1 July 2003–20 June 2008), and NFDC1 (1 October 2000–30 September 2004). (b) Same as (a) but on base 10 log-log scales.
FY11

- Adapted/Developed the prototype statistical post processor, referred to as the Ensemble Post-Processor (EPP)
  - Given the District-produced Day-1 QPF, estimates the conditional probability distribution of the (unknown) observed rainfall and generates an ensemble forecast from it
    - Correct/Reduce bias
    - Quantify uncertainty
- Formulated prototype graphical products
  - Capture the probabilistic information for use by the District’s water managers and meteorologists
Experimental graphical products

DADE Day-1 QPF valid 07/02/2010

- Probability Less Than
- Rainfall (in)

DADE Day-1 QPF valid 07/02/2010

- Frequency
- Rainfall (in)

Legend:
- Single-Value QPF
- Probabilistic QPF
FY12

- Enhance stochastic bias correction
- Obtain and analyze historical archive of District-produced Day-2 QPF
  - 24-hr QPF
  - 48-hr QPF
- Develop the calibration component of EPP (EPP_PE)
- Develop the real-time component of EPP (EPP_EG)
- Install EPP at the District for pilot operation for selected rain areas
  - Jun 1 through Sep 30, 2012
  - DADE, EAAE, UK
Ensemble Post-Processor (EPP)

Hist. obs. rainfall (District)

Hist. Day-1 single-valued QPF (District)

Hist. Day-2 single-valued QPF (District)

EPP calibration processor

EPP parameters

Hist. post-processed 1- and 2-day ensemble QPF

Verification

Verification info.

EPP Calibration Component

Real-time Day-1 single-valued QPF (District)

Real-time Day-2 single-valued QPF (District)

EPP real-time processor

Real-time post-processed 1- and 2-day ensemble QPF

Product generation

Real-time 1- and 2-day probabilistic QPF products

EPP Real-Time Component
EPP

- Models the conditional probability distribution of observed rainfall given (single-valued) forecast rainfall
  - Normal quantile transform
  - Linear regression in bivariate normal space
  - Parameter optimization
  - Inverse-transform
- Samples random deviates from the conditional distribution
- Builds on Seo et al. (2006), Schaake et al. (2007), Wu et al. (2011), Seo (2012)
NWS Hydrologic Ensemble Forecast System (HEFS)

Forecasters add value

HEFS will enable seamless hydrologic ensemble prediction from weather to climate scales and translate weather and climate prediction into uncertainty-quantified water information.
Off line, models joint distribution between single-valued QPF for each lead time and verifying observation

Multi-year archive of single-value QPF necessary
EPP

In real-time, given single-value QPF, generates ensemble traces from the conditional distribution for each lead time.
EPP – Novel aspects

• Co-minimization of unconditional error variance and Type-II conditional bias (CB)
  – Improves estimation of large (and small) rainfall amounts, but only to the extent that the error is attributable to CB

• Heteroscedastic error modeling
  – Improves reliability (i.e. unbiased in probability) of ensemble forecast
What is conditional bias (CB)?

From Seo D-J. 2012. Conditional bias-penalized kriging. Stochastic Environmental Research and Risk Assessment. DOI 10.1007/s00477-012-0567-z (online first), [http://www.springerlink.com/content/f5q475570q08252t/](http://www.springerlink.com/content/f5q475570q08252t/).
CB-penalized optimal estimation

Minimize

\[ J = E_{X^*,X}[(X^* - X)^2] + \alpha \cdot E_X[(E_{X^*}[X^* \mid X] - X)^2] \]

where

\[ X = NQT^{-1}(bZ_{QPF} + \varepsilon) + a \]

\[ X^* = E[X \mid X_{QPF}] = E[X \mid Z_{QPF}] \]

\[ = \int \{NQT^{-1}(bZ_{QPF} + \varepsilon_o) + a\} f_\varepsilon(\varepsilon_o) d\varepsilon_o \]
Using Bayesian optimal estimation,

\[ E[X^* | X] = E[X^*] + \rho_{x^*x} \frac{\sigma_{x^*}}{\sigma_X} (X - E[X]) \]

we have:

\[ E[(E[X^* | X] - X)^2] = (E[X^*] - E[X])^2 + \sigma_X^2 (1 - \rho_{x^*x} \frac{\sigma_{x^*}}{\sigma_X})^2 \]

\[ J = \sigma_{\varepsilon}^2 + \alpha (E[X^*] - E[X])^2 + \alpha \cdot \sigma_X^2 (1 - \rho_{x^*x} \frac{\sigma_{x^*}}{\sigma_X})^2 \]
If you “scatter-plotted” ensemble forecasts with box-and-whisker plots:

Modified box-and-whisker plot
1) Subtract verifying obs from ensemble members, and
2) sort all “residual” ensemble forecasts in the ascending order of obs
(Just about) Perfect ensemble forecast

Reliability Diagram

False Alarm Rate
Relative Operating Characteristic (ROC) curve
Climatological ensemble forecast is unconditionally reliable, and has no resolution or discrimination.
Ensemble hindcasting results

- Leave-one-year-out cross validation
- For the entire period of record
  - DADE: 1998-2010
  - EAAE: 1998-2011
  - UK: 1998-2010
- Verified using the Ensemble Verification System (EVS)
Modified box plot of ensemble forecast errors against observed value.
DADE_ens.Precipitation at lead hour 24.0

1998-2010
Modified box plot of ensemble forecast errors against observed value.
DADE_sva.Precipitation at lead hour 24.0
Modified box plot of ensemble forecast errors against observed value.
EAAE_ens.Precipitation at lead hour 24.0
Modified box plot of ensemble forecast errors against observed value.
EAAE_sva.Precipitation at lead hour 24.0
Modified box plot of ensemble forecast errors against observed value.
UK_ens. Precipitation at lead hour 24.0
Modified box plot of ensemble forecast errors against observed value.

UK_sva.Precipitation at lead hour 24.0
Modified box plot of ensemble forecast errors against observed value.
DADE2_ens.Precipitation at lead hour 48.0
Modified box plot of ensemble forecast errors against observed value.
DADE2_sva.Precipitation at lead hour 48.0
Modified box plot of ensemble forecast errors against observed value.
EAAE2_ens.Precipitation at lead hour 48.0
Modified box plot of ensemble forecast errors against observed value.
EAAE2_sva.Precipitation at lead hour 48.0
Modified box plot of ensemble forecast errors against observed value.
UK2_ens.Precipitation at lead hour 48.0
Modified box plot of ensemble forecast errors against observed value.
UK2_sva.Precipitation at lead hour 48.0
ROC area is closely related to the economic value of a forecast system (Zhu et al. 2001)

Results based on simple cost/loss models have indicated that probabilistic forecasts have a higher value than single deterministic forecasts.

From Hagedorn (2005)
Gold=1 inch threshold

Navy blue=1 inch threshold

Reliability diagram for various event thresholds (upper) and sample counts (lower).
Conclusions and recommendations

• The Ensemble Post-Processor (EPP) has been developed, verified and installed for pilot testing from Jun 1 through Aug 31, 2012, for DADE, EAAE and UK.
  – The results will be reported at ICHE 2012 in early Nov, Orlando, FL.
• The EPP-generated ensemble QPFs are reliable (i.e. probabilistically unbiased) up to thresholds of 1 and 2 inches for 24- and 48-hr rainfall, respectively, and improves over the single-valued QPFs in information content.
• To translate the increased information content to increased economic value, however, the ensemble paradigm must be embraced and practiced in hydrologic modeling and water management.
Ensemble prediction for risk-based management of water hazards & resources

Adapted from Hagedorn (2005)
Some examples of hydrologic ensemble forecasting

- http://www.erh.noaa.gov/mmefs/
- http://www.erh.noaa.gov/mmefs/serfc/
NYCDEP ensemble project
Conclusions and recommendations (cont.)

- Additional sources of info for future rainfall (e.g. SREF) should be sought to help reduce missed QPFs for very large rainfall amounts
  - The skill in such information is usually small
  - The only way to make use of such info is through ensemble forecasting
- Codify the “operation control” rules and generate the derived distributions of flow and/or net loss based on the ensemble QPF
- Partner with NWS SERFC and jointly produce ensemble QPFs for South Florida
Thank you

For more information, contact:

djseo@uta.edu