Hydrological Modeling Where No Meteorological Stations Exist


**Introduction**

Hydrologic modeling at daily or sub daily resolution requires accurate precipitation and temperature data, which is particularly important when planning land use changes and live management operations such as irrigation allocation. Weather gage data pose a variety of challenges, including the fact that they are non-existent in many watersheds of interest. We show that by integrating global forecast products and reanalysis datasets from the atmospheric modeling community into the hydrologic models commonly used, we can circumvent some of the issues with point-of-measure weather gauges. Traditionally in hydrologic forecasting, a precipitation forecast is used from a nearby precipitation station on which the hydrologic model has been calibrated. Here we present the results of two studies. First, we compare calibrating a watershed model (SWAT) using weather forecasts from ensembles of raw gridded atmospheric models interpolated to the center of subbasins against using the closest weather gage measurements. Second, we perform a similar study using global reanalysis datasets. In addition, we look at what scales and radii using direct gridded model output may introduce equal or less error to watershed modeling projects than using the closest station.

**Issues in Hydrologic Modeling**

- **Spatial Issues**
  - Assuming nearest reporting stations for PPT.
    - 30km apart, PPT = 23
  - Finding a set of stations to run with real-time at 30km is unlikely.
  - This type of agreement would be a dream situation for us!
  - US has:
    - ~60km spacing between real-time stations
    - Africa
      - 600km-1000km regularly spaced stations in region

**Study Overview**

- **Goals**
  - Need consistent forcing variables (ppt, temp, etc).
  - And no missing time steps
  - We would like to produce short term forecasts.
  - We want to develop “live” management tools.
  - Live irrigation allocation planning
  - Applicable world wide.

- **Hypothesis**
  - Gridded models have some skill, thus substituting short range forecasts centered in the watershed for weather station data should better represent conditions in the watershed than stations at some X distance away.
  - Question: What is this distance?

**Study Overview, cont.**

**Data Overview**

- Climate Forecast System Reanalysis (CFSR)
  - Hourly Forecast System (NCEP)
  - ECMWF Archive.
  - NCEP
  - Closest Weather Stations.
  - Variables.
    - PPT, as well as Max and Min Temperature.

**Gridded Forecast Models from TIGGE**

- Date range: 4/1/2007 – present
  - 6Z runs, model hours 06 through 30
  - ECMWF Ensemble.
  - 0.45deg Res. Global.
  - 51 perturbations.
  - Median from 50 member pt forecast.
  - NCEP Global Ensemble Forecast System (GFE).
  - 1.0deg Res. Global.
  - 30 perturbations.
  - Median from 20 member pt. forecast.

**Study Watersheds and Model**

- Gumara Watershed
  - Tana Basin
  - 1200km²
- Town Brook watershed
  - New York State.
  - 37km² catchment of Cannonsville basin
  - Cross River catchment
  - 44km² catchment of NYC water supply
  - SWAT 2005, initialized with ArcSWAT.

**Results, cont.**

**Best Meteorological Station vs. CFSR**

<table>
<thead>
<tr>
<th>Location</th>
<th>Watershed Area (km²)</th>
<th>Best Met Distance (km)</th>
<th>Best Met NSE</th>
<th>CFSR NSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Town Brook</td>
<td>44</td>
<td>13</td>
<td>0.54</td>
<td>0.68</td>
</tr>
<tr>
<td>Dutchess</td>
<td>37</td>
<td>0/3</td>
<td>0.52</td>
<td>0.63</td>
</tr>
<tr>
<td>Westchester</td>
<td>4.2</td>
<td>0</td>
<td>0.29</td>
<td>0.17</td>
</tr>
<tr>
<td>Monticello</td>
<td>30</td>
<td>0</td>
<td>0.17</td>
<td>0.04</td>
</tr>
<tr>
<td>Cross River</td>
<td>30</td>
<td>0</td>
<td>0.17</td>
<td>0.06</td>
</tr>
</tbody>
</table>

**Ethiopian Watersheds**

<table>
<thead>
<tr>
<th>Location</th>
<th>Watershed Area (km²)</th>
<th>CFSR NSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lake Tana</td>
<td>45</td>
<td>0.56</td>
</tr>
<tr>
<td>Mefraj</td>
<td>40</td>
<td>0.59</td>
</tr>
<tr>
<td>Jijiga</td>
<td>30</td>
<td>0.61</td>
</tr>
</tbody>
</table>

**CFSR Results Town Brook**

Left, three scatter plots showing the optimal calibrations of SWAT on the Town Brook watershed. At the top a) the optimal calibration results using reanalysis data from CFSR, giving the best results when compared to an “Ideal meteorological station” calibration shown in the center b) and the previously published ideal calibration from Easton et al. 2008 shown in c). All calibrations were performed using the same period.

**CFSR Results Cross River**

Below, predicted vs. observed stream flow graph with the furthest distance station in the background in darkest/largest symbols and closer stations represented by lighter/smaller symbols. Legend shows the location along with locations respective NSE.

**Discussion**

- There is validity to using reanalysis datasets in place of weather station data to force hydrologic modeling systems when nearby weather stations are not available or have inconsistent records
- There is validity to using short term forecast archives in place of weather station data to force hydrologic modeling systems.
- The distance to the where the model skill produces the same accuracy as the weather stations is closer than we thought.
- Variables from the atmospheric models are physically linked, and we should likely keep them that way.

**Conclusion**

This study found that forecast archives and CFSR reanalysis data provide a potentially valuable alternative to “nearest weather station” data for hydrologists. Indeed, this study found that the CFSR data may actually be a better representation of the weather than point observations, at the scale of small watersheds (<50 km²). Because the data are available globally, they provide new opportunities for modeling hydrology in remote or data-poor regions of the world. The next steps this research will involve including a wide range of hydrologic models, watershed characteristics (e.g., size), and phenomena (e.g., soil moisture). In addition the CFSR data could make downscaling possible in large ungauged basins, as well as allow modeling shorter time steps.

**Acknowledgements and Interesting Side Points**

- Authors would like to acknowledge ECMWF, who supplied the data for this study via the TIGGE online user interface. [http://tigge.ecmwf.int/](http://tigge.ecmwf.int/)
- All computing for this study was performed in the publicly accessible Amazon EC2 grid.
- Authors are looking for collaborators who might wish to run similar studies without the pain of extracting the datasets themselves.