Evaluation of Experimental Models for Tropical Cyclone Forecasting in Support of the NOAA Hurricane Forecast Improvement Project (HFIP)

Barbara Brown, Paul Kucera, Louisa Nance, and Christopher Williams

Joint Numerical Testbed
Tropical Cyclone Modeling Testbed
National Center for Atmospheric Research
(bgb@ucar.edu)
Tropical Cyclone Modeling Testbed (TCMT): HFIP Model evaluation activities

- **Main focus**: Independent diagnostic evaluation of HFIP models
- Planning and evaluation of HFIP Retrospective evaluations
  - Implementation of meaningful diagnostic verification approaches: *Focus on NHC requirements*
  - Evaluation of Stream 1.5 candidates; report to NHC and HFIP
- Real-time demonstration during Demo period (Aug-Nov)
- Evaluation of Demo models
  - Intensive evaluations of specific storms
  - Overall evaluation of forecasts for all storms
- Development of verification methods and tools
HFIP 2011 Retrospective Cases

**Goal:** Select Stream 1.5 models for HFIP Demo exercise

Modeling groups ran retrospective cases for more than 600 cases from 2008-2010

Collaboration with NHC
Retrospective evaluations

- Comparisons with baseline models
  - Significant and “practical” differences
  - Frequencies of large error differences
- Contributions to consensus forecasts
- Performance relative to “top flight” models
2011 Retrospective Evaluation

Example: 3 km HWRF

Error Distribution Evaluation

Stream 1.5 model: Fewer cases w/ large track errors
Similar or larger errors for intensity
## Practical Significance Evaluation

### 2011 Hurricane Retrospective Evaluation - Results

#### Example: 3 km HWRF

<table>
<thead>
<tr>
<th>Forecast Hour</th>
<th>0</th>
<th>12</th>
<th>24</th>
<th>36</th>
<th>48</th>
<th>60</th>
<th>72</th>
<th>84</th>
<th>96</th>
<th>108</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GFSI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Track</td>
<td>-2</td>
<td>-8</td>
<td>-12</td>
<td>-16</td>
<td>-20</td>
<td>-24</td>
<td>-28</td>
<td>-32</td>
<td>-36</td>
<td>-40</td>
<td>-44</td>
</tr>
<tr>
<td>(Land and Water)</td>
<td>0.816</td>
<td>0.987</td>
<td>0.986</td>
<td>0.961</td>
<td>0.936</td>
<td>0.819</td>
<td>0.837</td>
<td>0.876</td>
<td>0.929</td>
<td>0.999</td>
<td></td>
</tr>
<tr>
<td><strong>GMI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Track</td>
<td>4%</td>
<td>2%</td>
<td>2%</td>
<td>3%</td>
<td>5%</td>
<td>4%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>-3%</td>
<td>-3%</td>
</tr>
<tr>
<td>(Land and Water)</td>
<td>0.880</td>
<td>0.426</td>
<td>0.540</td>
<td>0.682</td>
<td>0.667</td>
<td>0.420</td>
<td>0.201</td>
<td>0.185</td>
<td>0.083</td>
<td>0.307</td>
<td></td>
</tr>
<tr>
<td><strong>LGEM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intensity</td>
<td>6%</td>
<td>8%</td>
<td>13%</td>
<td>11%</td>
<td>9%</td>
<td>4%</td>
<td>9%</td>
<td>0%</td>
<td>0%</td>
<td>-3%</td>
<td>-1%</td>
</tr>
<tr>
<td>(Land and Water)</td>
<td>0.987</td>
<td>0.986</td>
<td>0.999</td>
<td>0.984</td>
<td>0.532</td>
<td>0.390</td>
<td>0.000</td>
<td>0.279</td>
<td>0.292</td>
<td>0.094</td>
<td></td>
</tr>
</tbody>
</table>

### SS differences

- **Δ < -20**: 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36
- **Δ < -10**: 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36
- **Δ < 0**: 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36
- **Δ > 0**: 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36

### Not SS

- **Δ < -20**: 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36
- **Δ < -10**: 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36
- **Δ < 0**: 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36
- **Δ > 0**: 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36, 0.36

### Summary

- **Mean error difference**: 16.8
- **% Improve (+) / degrade (-)**: 15%
- **p-value**: 0.999
Top Flight Model Comparison

UW-NMS Example

Average Errors

Rank
Real-Time Demonstration and Evaluation

- Operational and experimental models provided in real-time to TCMT website in support of the HFIP Demonstration
- Multi-model ensemble mean computed and displayed using the HFIP experimental models
- Track plots and verification results available on the HFIP web site
- Near real time diagnostic evaluation of individual storms (Irene, Maria)
Real-Time Demonstration and Evaluation

- Operational and experimental models provided in real-time to TCMT website in support of the HFIP Demonstration
- Multi-model ensemble mean computed and displayed using the HFIP experimental models
- Track plots and verification results available on the HFIP web site
- Near real time diagnostic evaluation of individual storms (Irene, Maria)

Sample Size Sensitivity

Irene

Mean Intensity Error
Atlantic Basin (Land and Water)

Lead Time (h)

Intensity Error (kt)

2011 Error Distribution (All Cases)
Hurricane Maria Error Distribution
Hurricane Maria Mean
Error Target for 20% HFIP Goal
Real-Time Demonstration and Evaluation

• Operational and experimental models provided in real-time to TCMT website in support of the HFIP Demonstration
• Multi-model ensemble mean computed and displayed using the HFIP experimental models
• Track plots and verification results available on the HFIP web site
• Near real time diagnostic evaluation of individual storms (Irene, Maria)
New Methods for Evaluation of TC Forecasts

Developing new methods to evaluate ensemble and multi-model forecasts

Example: GFDL 18-member Ensemble

"Wind Rose" Plots for Track and Wind Speed Errors

Rank Histograms

Wind Speed/Track Error Diagrams
Summary

- TCMT supports HFIP efforts by providing independent, diagnostic, forecast verification of experimental model forecasts
- New approaches for evaluation
  - To meet needs of NHC for selecting promising models for Stream 1.5
  - To evaluate new kinds of forecasts (e.g., ensembles)
- Results available for 2009-2012 at http://www.rap.ucar.edu/jnt/tcmt/
- Currently gearing up for 2012 retrospective (Stream 1.5) evaluation