Evaluation of Ensemble Forecast
- Performance metrics

Yuejian Zhu

EMC/NCEP/NOAA

5th NAEFS workshop, Mexican City
May 18 2010
Ensemble forecast metrics

• NCEP GEFS
  – NCEP GFS based ensemble forecast system
  – Half resolution of GFS forecast
  – ETR initial perturbation method

• NAEFS
  – Combine NCEP/GEFS, CMC/GEFS (current)
  – Adding FNMOC/GEFS (by Q4FY2010 or Q1FY2011)
  – About 60 members for 00UTC and 12UTC
  – Output every 6 hours out to 16 days
  – Exchange about 80 variables
  – Post products
    • Bias correction (49 variables)
    • Downscaling (8 variables)
NCEP/GEFS current evaluations (1)

- **Variables:**
  - 500hPa and 1000hPa heights
  - 850hPa and 2-meter temperatures
  - 10-meter U and V
  - 850hPa and 200hPa U and V

- **Domains:**
  - Globally, NH (20N-80N), SH (20S-80S) and TROP (20N-20S)
  - NA (Northern American), EU (Europe) and AS (Asia)

- **Resolution:**
  - Based on 2.5*2.5 degree resolution

- **Frequency:**
  - Every 12/24 hours, out to 16 days

- **Against:**
  - Analysis (or best analysis), and observations (possible for future)

- **Verifying:**
  - PAC, RMS errors, ABS errors for ensemble mean
  - Ensemble spread and histogram
  - BS (in terms of reliability and resolution)
  - RPSS and CRPS
  - ROC and EV (economic values)

- **Skills:**
  - Use NCEP/NCAR 40y reanalysis as references
Example of score cards for ensembles evaluation

Using 95% confidence interval (2.5%-97.5%), BLUE means NAEFSb+FNMOCb is significantly better than NAEFSb. RED means otherwise.

The reliability (Rel) and resolution (Res) are from Brier Score decomposition.
NCEP/GEFS current evaluations (2)

- Precipitation
- CONUS only
- Based on 1*1 degree resolution
- Every 24 hours, out to 15 days
  - From 1200UTC – 1200UTC
- Verify against: observations (Gauges)
  - About 8000-10000 reports for every 24 hours
- Verifying statistics for:
  - RMS errors, ABS errors for ensemble mean
  - Ensemble spread
  - CRPS (and reliability)
  - ETS, TSS and Bias for ensemble mean
    - different thresholds (0.01, 0.2, 2.0, 5.0, 10.0, 15.0, 25.0, 35.0 and 50.0 mm/days)
NCEP/GEFS current evaluations (3)

- **Tropical storm tracks** (or storm tracks)
- **Domains:**
  - Globally, Atlantic basin, East Pacific basin and West Pacific Basin
- Based on model output – surface pressure and other variables at 1.1 degree
- Out to 7 days (will be extended to 10 days)
- Against: observed position (best position)
- Verifying statistics for:
  - Track errors for ensemble mean
  - Ensemble spread
- Plotting:
  - Case by case for individual ensemble member
  - Statistics for case average
  - Statistics for seasonal average
## NCEP GEFS Probabilistic Evaluation Metrics (drafted plan)

<table>
<thead>
<tr>
<th>Fields</th>
<th>Levels</th>
<th>Area</th>
<th>Variable</th>
<th>Measures</th>
<th>Lead-time</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grid</td>
<td>500hPa</td>
<td>NH</td>
<td>Height</td>
<td>AC/RMS/sprd</td>
<td>3,5,8 days</td>
<td>3</td>
</tr>
<tr>
<td>Grid</td>
<td>500hPa</td>
<td>NH</td>
<td>Height</td>
<td>Skillful fcst</td>
<td>Days</td>
<td>1</td>
</tr>
<tr>
<td>Grid</td>
<td>500hPa</td>
<td>NH</td>
<td>Height</td>
<td>CRPS</td>
<td>3,5,8 days</td>
<td>3</td>
</tr>
<tr>
<td>Grid</td>
<td>1000hPa</td>
<td>NH</td>
<td>Height</td>
<td>CRPS</td>
<td>3,5,8 days</td>
<td>3</td>
</tr>
<tr>
<td>Grid</td>
<td>850hPa</td>
<td>NH</td>
<td>Temperature</td>
<td>CRPS</td>
<td>3,5,8 days</td>
<td>3</td>
</tr>
<tr>
<td>Grid/OBS</td>
<td>Surface</td>
<td>NH/Tropical</td>
<td>Pressure/winds</td>
<td>Track Error</td>
<td>1,3,5 days</td>
<td>3</td>
</tr>
<tr>
<td>Grid</td>
<td>850hPa</td>
<td>Tropical</td>
<td>Winds</td>
<td>CRPS</td>
<td>3,5 days</td>
<td>2</td>
</tr>
<tr>
<td>Grid</td>
<td>200hPa</td>
<td>Tropical</td>
<td>Winds</td>
<td>CRPS</td>
<td>3,5 days</td>
<td>2</td>
</tr>
<tr>
<td>Grid/OBS</td>
<td>Surface</td>
<td>CONUS</td>
<td>Precipitation</td>
<td>ETS/CRPS</td>
<td>1,3,5 days</td>
<td>3</td>
</tr>
<tr>
<td>Grid</td>
<td>Surface</td>
<td>NH</td>
<td>Temperature</td>
<td>CRPS</td>
<td>3 days</td>
<td>1</td>
</tr>
<tr>
<td>Grid</td>
<td>Surface</td>
<td>NH</td>
<td>Winds</td>
<td>CRPS</td>
<td>3 days</td>
<td>1</td>
</tr>
</tbody>
</table>
GFS and GEFS performance

- Skillful forecast
Day at which fcst loses useful skill (AC=0.6)
N. Hemisphere calendar year means

GFS has better performance of first 6 days in 2009 comparing to 2008
e.g. Day-6: .755(2008), .758(2009)
Monthly Average for NH 500hPa Height, 5-day forecasts

RPS scores

D5=0.357
2001

D6=0.334
2007

RPSS=0.396

EV (1:10 ratio)

D5=0.356

D6=0.361

EVAL=0.394
General evaluations for ensemble mean

- Comparison for implementation
NH Anomaly Correlation for 500hPa Height
Period: August 1st – September 30th 2007

- GEFSg is better than GFS at 48 hours
- GEFSg could extend skillful forecast (60%) for 9+ days
- 24 hours better than current GEFS
- 48 hours better than current GFS
Summary of the important cases of Bill Jimena, Rick and Ida

**TS track errors (2009)**

- **Cases**
  - 00UTC: 28
  - 12UTC: 26
  - 24UTC: 24
  - 36UTC: 22
  - 48UTC: 20
  - 72UTC: 15
  - 96UTC: 9
  - 120UTC: 6

For 00UTC only
NH 500hPa Height

Fall 2008 (AC)

FNMOC is about 12h behind CMC and NCEP

E20s – NCEP 20 members raw ensemble mean
E20m – CMC 20 members raw ensemble mean
E16f – FNMOC 16 members raw ensemble mean

Forecast days
Tropical Cyclone Track Error vs. Fhr – NCEP Ensemble

STATS for Atlantic basin 00UTC only
2 months (August and September 2008)
Evaluations for bias correction and downscaling

- Demonstrate the values added
RTMA Region 2m Temperature
Averaged From 2007090100 to 2007093000

NCEP/GEFS raw forecast

4+ days gain from NAEFS

NAEFS final products

From Bias correction (NCEP, CMC)
Dual-resolution (NCEP only)
Down-scaling (NCEP, CMC)
Combination of NCEP and CMC
RTMA Region 2m Temperature
Averaged From 2007090100 to 2007093000

NCEP/GEFS raw forecast

4+ days gain from new products

Final products: NCEPbc+CMCbc + dual-resolution + down-scaling

<table>
<thead>
<tr>
<th>Line Style</th>
<th>Legend</th>
</tr>
</thead>
<tbody>
<tr>
<td>+----------</td>
<td>naefs_bcds</td>
</tr>
<tr>
<td>o</td>
<td>gefs_raw</td>
</tr>
<tr>
<td>o</td>
<td>gefs_rawds</td>
</tr>
<tr>
<td>+</td>
<td>gefs_drbcds</td>
</tr>
<tr>
<td>+</td>
<td>gefs_bcds</td>
</tr>
<tr>
<td>x</td>
<td>cmc_bcds</td>
</tr>
</tbody>
</table>

BO CUI, GCWMB/EMC/NCEP/NOAA
From Bias correction (NCEP, CMC)
Dual-resolution (NCEP only)
Down-scaling (NCEP, CMC)
Combination of NCEP and CMC

NAEFS NDGD Probabilistic 2m Temperature Forecast Verification For 2007090100 – 2007093000

Continuous Ranked Probability Score (C)

From NCEP/GEFS raw forecast
NAEFS final products

8+ days gain
General evaluations for precipitation

- Ensemble mean and distribution
Continuous Rank Probability Score

\[
\text{CRPS} = \int_{-\infty}^{+\infty} [F(x) - H(x - x_0)]^2 dx
\]

**CRP Skill Score is**

\[
\text{CRPSS} = \frac{\text{CRPS}_c - \text{CRPS}_f}{\text{CRPS}_c}
\]

Order of 10 ensemble members (p01, p02, ..., p10)

Heaviside Function H

\[
H(x - x_0) = \begin{cases} 0 & (x \leq x_0) \\ 1 & (x > x_0) \end{cases}
\]
Evaluation for probabilistic forecast

- Still working on ……
2 Meter Temperature Forecast

Init: 2008042300 (solid line: 50% shaded: 10–90%)

Surface Pressure Forecast

500hPa Geopotential Height Forecast

Location: Washington DC (37N 77W)
2-meter temperature 10/90 probability forecast verification

- P10-expect
- P10-ncep raw
- P10-naefs

- P90-expect
- P90-ncep raw
- P90-naefs

Lead time (hours)

Probabilities (percent)

~40%
~80%
2-meter temperature 10/90 probability forecast verification
Northern Hemisphere, seasonal variation for NAEFS

NAEFS final 10% and 90% probability forecast for Dec. 2007 and Feb 2008
Values from adding additional ensembles

- For decision making
Value-added by including FNMOC ensemble into NAEFS T2m: Against analysis (NCEP’s evaluation, 1 of 4)

Northern Hemisphere 2 Meter Temp.
Continuous Ranked Probability Skill Scores
Average For 20081201 - 20090228

Raw NCEP ensemble has modest skill (3.4d)
0.5 CRPS skill
Value-added by including FNMOC ensemble into NAEFS
T2m: Against analysis (NCEP’s evaluation, 2 of 4)

Northern Hemisphere 2 Meter Temp.
Continuous Ranked Probability Skill Scores
Average For 20081201 - 20090228

- Raw NCEP ensemble has modest skill (3.4d)
- Statistically corrected NCEP ensemble has improved skill (4.8d)

Stat. corr.

0.5 CRPS skill

0 1 2 3 4 5 6 7 8 9 10
Forecast days

Skill Scores
Value-added by including FNMOC ensemble into NAEFS
T2m: Against analysis (NCEP’s evaluation, 3 of 4)

Northern Hemisphere 2 Meter Temp.
Continous Ranked Probability Skill Scores
Average For 20081201 – 20090228

Raw NCEP
Stat. corr.
NAEFS

Raw NCEP ensemble has modest skill (3.4d)
Statistically corrected NCEP ensemble has improved skill (4.8d)
Combined NCEP – CMC (NAEFS) show further increase in skill (6.2d)
Value-added by including FNMOC ensemble into NAEFS

T2m: Against analysis (NCEP’s evaluation, 4 of 4)

Northern Hemisphere 2 Meter Temp.
Continous Ranked Probability Skill Scores
Average For 20081201 – 20090228

- Raw NCEP ensemble has modest skill (3.4d)
- Statistically corrected NCEP ensemble has improved skill (4.8d)
- Combined NCEP – CMC (NAEFS) show further increase in skill (6.2d)
- Addition of FNMOC to NAEFS leads to modest improvement (6.7d)