Northern American Ensemble Forecast System (NAEFS)-Bias Correction

Yuejian Zhu,
Bo Cui and Zoltan Toth

Environmental Modeling Center
NOAA/NWS/NCEP

Acknowledgements:
DingChen Hou  EMC
NAEFS Background Information

- First of a kind project
  - Operational multi-center ensemble system
  - Bias correction, climate percentiles never computed on such a scale operationally

- Timetable
  - Mar 2003  Project started
  - Oct 2003  Draft Research, Development and Implementation Plan
  - Sep 2004  Initial Operational Capability – Operational data exchange
  - May 2006  First Operational Implementation
  - Mar 2007  NAEFS upgrade

- Challenges
  - Developed joint plan with MSC personnel
  - Arranged operational data exchange
  - Coordinated GEFS development with international NAEFS developments
  - Coordinated software development & operational implementation with MSC
  - Worked with less THORPEX resources than planned originally

- Future expansion
  - Develop sustainable plans
    - Coordinate with partners
    - Rename NAEFS and position it as prototype GIFS system
  - Resource concerns
    - Computational (telecommunication, disc, etc)
First Implementation of NAEFS – Summary

1. Bias corrected members of joint MSC-NCEP ensemble
   • Decaying accumulated bias (~past 50 days) for each var. for each grid point
   • For selected 35 of 50 NAEFS variables
   • 32(00Z), 15(06Z), 32(12Z) and 15(18Z) joint ensemble members
   • Bias correction against each center’s own operational analysis

2. Weights for each member for creating joint ensemble
   (equal weights now – unequal weights to be added later)
   • Weights don’t depend on the variables
   • Weights depend on geographical location (low precision packing)
   • Weights depend on the lead time

3. Climate anomaly percentiles for each member
   • Based on NCEP/NCAR 40-year reanalysis
     • Used first 4 Fourier modes for daily mean,
     • Estimated climate pdf distribution (standard deviation) from daily mean
   • For selected 19 of 50 NAEFS variables
   • 32(00Z), 15(06Z), 32(12Z) and 15(18Z) joint ensemble members
   • Adjustment made to account for difference between oper. & re-analysis
   • Provides basis for downscaling if local climatology available
     – Non-dimensional unit
Bias Correction Method & Application

- Bias Assessment: adaptive (Kalman Filter type) algorithm

\[
\text{decaying averaging mean error} = (1-w) \times \text{prior t.m.e} + w \times (f - a)
\]

For separated cycles, each lead time and individual grid point, t.m.e = time mean error

- Test different decaying weights. 0.25%, 0.5%, 1%, 2%, 5% and 10%, respectively

- Decide to use 2% (~ 50 days) decaying accumulation bias estimation

- Bias Correction: application to NCEP operational ensemble 15 members
## List of Variables for Bias Correction, Weights and Forecast Anomalies for CMC & NCEP Ensemble

<table>
<thead>
<tr>
<th>Ensemble</th>
<th>CMC &amp; NCEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRID</td>
<td>1x1 deg (360x180 lat-lon)</td>
</tr>
<tr>
<td>DOMAIN</td>
<td>Global</td>
</tr>
<tr>
<td>FORMAT</td>
<td>WMO Grib Format</td>
</tr>
<tr>
<td>HOURS</td>
<td>6 hourly out of 384 hours (current 240 hours for CMC Ensemble)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>CMC &amp; NCEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>GZ</td>
<td>200, 250, 500, 700, 850, 925, 1000</td>
</tr>
<tr>
<td>TT</td>
<td>200, 250, 500, 700, 850, 925, 1000</td>
</tr>
<tr>
<td>U,V</td>
<td>200, 250, 500, 700, 850, 925, 1000</td>
</tr>
<tr>
<td>TT</td>
<td>2m</td>
</tr>
<tr>
<td>U, V</td>
<td>10m</td>
</tr>
<tr>
<td>MSLP</td>
<td>Sea Level Pressure</td>
</tr>
<tr>
<td>Sfc Pres</td>
<td>Surface Pressure</td>
</tr>
<tr>
<td>Tmax</td>
<td>2m</td>
</tr>
<tr>
<td>Tmin</td>
<td>2m</td>
</tr>
</tbody>
</table>

Note: 35 Variables in total, red variables are for climate anomalies only.
Summary of NAEFS First Implementation

• Period:
  – 04/10/2006 – Current (NCO real time parallel)

• Maps comparison for bias (before and after)
  – 500hPa height, 2m temperature

• Statistics for
  – Bias reduction in percentage
    • Height, temperature, winds
  – RMS errors
  – Probabilistic verifications (ROC)
    • NH, SH and tropic

• Conclusions
  – Bias reduced (approximately 50% at early lead time)
  – RMS errors improved by 9% for d0-d3
  – Probabilistic forecast
    • Improved for all area, all lead time
    • Typically for NH, 20-24 hours improvement from d7
500hPa height: 120 hours forecast (ini: 2006043000)

Shaded: left – raw bias  
right – bias after correction
2 meter temperature: 120 hours forecast (ini: 2006043000)

Shaded: left – raw bias right – bias after correction
Bias Improvement (absolute value) after Bias correction

500hPa height

Overall bias reduction:
(globally)
D0-3: 50%
D3-8: 40%
D8-15: 30%

850hPa temperature

Sea level pressure

There is daily variation after bias correction, more bias reduced for valid 12Z cycle

2m Temperature
Bias Improvement (absolute value) after Bias correction

- **10m U-component**
  - Overall bias reduction:
    - (Tropic)
    - D0-3: 50%
    - D3-8: 45%
    - D8-15: 40%

- **10m V-component**

- **Sea level pressure**

- **2m temperature**
Evaluation after bias correction (16 cases)

Northern Hemisphere 500 mb Height (ROC area)
Average For 20060425 – 20060510

Southern Hemisphere 500 mb Height (ROC area)
Average For 20060425 – 20060510

Black-operational ensemble (10m)
Red-real time parallel ensemble (14m)
Green-real time parallel ensemble after bias correction (14m)

Probabilistic skill
Extended 20-h for d-7

RMS errors for ensemble mean
reduced for 48-h forecast (~9%)
NAEFS verification

- Reference: NCEP/NCAR 40y reanalysis (next slide)
- Variables:
  - 1000hPa, 500hPa heights, 850hPa, 2m temperature, 10m u and v
- Verified for ensemble mean:
  - RMS errors, spread, mean error (bias) and absolute error
- Verified for ensemble distribution:
  - Histogram (Talagrand)
- Verified for ensemble probabilistic forecast
  - ROC, RPSS, CRPS, BSS (Resolution and Reliability), EV
- Regions:
  - NH, SH, Tropical, Asia, Europe and Northern American
- Statistics from seasonal average
Climatological Data

- NCEP/NCAR 40 years (1958-1997) reanalysis
- Monthly Sampling
  - For example: $40 \times 30 = 1200$
- Generating 10 equally-a-likely, based on monthly sampling
- Projected to verify date
- All forecast skills will base on 10 equally-a-likely climatological bins.
Global Ensemble Model Evaluation: (NCEP against NCEPb)

### 500 hPa Height Scores NCEP .vs NCEPb

<table>
<thead>
<tr>
<th>Region</th>
<th>ROC</th>
<th>EV</th>
<th>RPSS</th>
<th>BSS</th>
<th>CRP</th>
<th>CRPS</th>
<th>RMS/SPRD</th>
<th>ERR/ABSE</th>
<th>HISTOGRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TROP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 1000 hPa Height Scores (NCEP .vs NCEPb)

<table>
<thead>
<tr>
<th>Region</th>
<th>ROC</th>
<th>EV</th>
<th>RPSS</th>
<th>BSS</th>
<th>CRP</th>
<th>CRPS</th>
<th>RMS/SPRD</th>
<th>ERR/ABSE</th>
<th>HISTOGRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TROP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 850 hPa Temperature Scores (NCEP .vs NCEPb)

<table>
<thead>
<tr>
<th>Region</th>
<th>ROC</th>
<th>EV</th>
<th>RPSS</th>
<th>BSS</th>
<th>CRP</th>
<th>CRPS</th>
<th>RMS/SPRD</th>
<th>ERR/ABSE</th>
<th>HISTOGRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TROP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 2 Meters Temperature Scores (NCEP .vs NCEPb)

<table>
<thead>
<tr>
<th>Region</th>
<th>ROC</th>
<th>EV</th>
<th>RPSS</th>
<th>BSS</th>
<th>CRP</th>
<th>CRPS</th>
<th>RMS/SPRD</th>
<th>ERR/ABSE</th>
<th>HISTOGRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TROP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ISSUES ADDRESSED

• Effect of bias-correction
  – Different variables

• Comparing of NCEP and CMC’s forecasts
  – Before & after bias correction

• Impact of combined ensemble (NAEFS)
  – Before & after bias correction
  – Gains from bias correction + combination =
    • NAEFS advantage
Northern Hemisphere 500hPa Height Histogram Distribution
Average For 20060601 – 20060831

HISTOGRAM

1-day

3-day

5-day

8-day

12-day

16-day

Good spread, but more biased
RMSE and Spread

Mean and absolute errors

10 meter wind (u-component)
Less biased,
There is less room to improve the skill by bias-correction only
ISSUES ADDRESSED

• Effect of bias-correction
  – Different variables

• Comparing of NCEP and CMC’s forecasts
  – Before & after bias correction

• Impact of combined ensemble (NAEFS)
  – Before & after bias correction
  – Gains from bias correction + combination =
    • NAEFS advantage
**Continuous Rank Probability Score**

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

**CRP Skill Score is**

\[ CRPSS = \frac{CRPS_c - CRPS_f}{CRPS_c} \]

**Heaviside Function H**

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

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

Ranked (ordered) Probability Score (RPS) is to verify multi-category probability forecasts, to measure both reliability and resolution which based on climatologically equally likely bins.

\[
RPS = 1 - \frac{1}{k-1} \left[ \sum_{i=1}^{k} \left( \sum_{n=1}^{i} P_n - \sum_{n=1}^{i} O_n \right)^2 \right]
\]

and

\[
RPSS = \frac{RPS_f - RPS_c}{1 - RPS_c}
\]

Example of 10 climatologically equally likely bins, 10 ensembles

<table>
<thead>
<tr>
<th>OBS</th>
<th>On</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCST PROB</td>
<td>(P_n)</td>
<td>0%</td>
<td>10%</td>
<td>20%</td>
<td>30%</td>
<td>10%</td>
<td>20%</td>
<td>0%</td>
<td>10%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

\[
\sum_{n=1}^{i} P_n - \sum_{n=1}^{i} O_n \]

\(i=10 = k: number of categories\)
**Northern Hemisphere 500hPa Height**
Continuous Ranked Probability Skill Scores
Average For 20061201 - 20070228

**500hPa height**

**Northern Hemisphere 1000hPa Height**
Continuous Ranked Probability Skill Scores
Average For 20061201 - 20070228

**1000hPa height**

**Northern Hemisphere 850hPa Temp.**
Continuous Ranked Probability Skill Scores
Average For 20061201 - 20070228

**850hPa temperature**

**Northern Hemisphere 2 Meter Temp.**
Continuous Ranked Probability Skill Scores
Average For 20061201 - 20070228

**2 meter temperature**

Black = NCEP bias-corrected
Red = CMC bias-corrected
Green = NAEFS combined
ISSUES ADDRESSED

• Effect of bias-correction
  – Different variables

• Comparing of NCEP and CMC’s forecasts
  – Before & after bias correction

• Impact of combined ensemble (NAEFS)
  – Before & after bias correction
  – Gains from bias correction + combination =
    • NAEFS advantage
Solid: RMS error
Dash: Spread

Solid: Mean error (bias)
Dash: Mean absolute error

36h improvement by NAEFS
Northern Hemisphere 2 Meter Temp.
Ranked Probability Skill Scores (RPSS)
Average For 20061201 – 20070228

24h improvement by NAEFS

RPSS vs CRPSS

Winter 2006-2007
NH 2m temperature
For
NCEP raw forecast (black)
NCEP bias corrected forecast (red)
NAEFS forecast (pink)
Background !!!!
Relative Operating Characteristics area (ROC area)

- **f(noise)**
- **f(signal)**

- Near perfect forecast
- No skill forecast
- Real forecast

- False alarm rate
- Hit rate

- Decision threshold

- False alarm rate

- 0 1

- Real forecast
## Appendix 6

### KEY PERFORMANCE MEASURES

#### Improvement in Ensemble Forecasts

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Threshold</th>
<th>Actual 25Apr-10May06</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias Reduction (%)</td>
<td>50%</td>
<td>30-70%</td>
<td>Met or exceeded in Tropics &amp; up to D3 elsewhere; slightly below otherwise</td>
</tr>
<tr>
<td>RMS Error Reduction (%)</td>
<td>10%</td>
<td>Up to 10%</td>
<td>Met up to D3, below expected D4 and beyond</td>
</tr>
</tbody>
</table>

#### Improvement in Ensemble-based Probabilistic Forecasts

<table>
<thead>
<tr>
<th></th>
<th>3 Day</th>
<th>6 Hours</th>
<th>12 hrs</th>
<th>Exceeded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7 Day</td>
<td>12 Hours</td>
<td>16 hrs</td>
<td>Exceeded</td>
</tr>
<tr>
<td></td>
<td>10 – 14 Days</td>
<td>24 Hours</td>
<td>48 hrs</td>
<td>Exceeded</td>
</tr>
</tbody>
</table>
### Appendix 8
**MINIMAL (PREFERRED) CONFIGURATION FOR THE GLOBAL ENSEMBLE FORECAST SYSTEMS OPERATIONAL AT CMC AND NCEP**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast lead time (days)</td>
<td>16</td>
<td>16 (35)</td>
<td>16</td>
</tr>
<tr>
<td>Number of cycles per day</td>
<td>2 (4)</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Number of ensemble members</td>
<td>10 (20)</td>
<td>20 (50)</td>
<td>14 / 20</td>
</tr>
<tr>
<td>Model resolution (km)</td>
<td>120 (90)</td>
<td>80 (60)</td>
<td>120 / ?</td>
</tr>
<tr>
<td>Number of vertical levels</td>
<td>28 (42)</td>
<td>42 (64)</td>
<td>28 / ?</td>
</tr>
</tbody>
</table>

**NAEFS Configuration Review (NCEP)**