Evaluation of Different Model-Error schemes in a WRF mesoscale ensemble

14th WRF Users’ Workshop, June 24-28, 2013, Boulder, CO

Judith Berner, Kate R. Smith, So-young Ha, Josh Hacker, Chris Snyder (NCAR)
Outline

- Evaluation of different model-error schemes in the WRF mesoscale ensemble: stochastic, multi-physics and combinations thereof
- Where is the additional skill coming from? Is it just the increased spread?
  - Decompose Brierscore into components
  - Ability of different model-error scheme to capture structural uncertainty
- How does the increased skill compare against that of postprocessing the forecasts
  - Impact of calibration
  - Impact of debiasing
  - Impact of model-version
Experimental Setup

- Weather Research and Forecast Model WRFV3.1.1.
- 15 dates between Nov 2008 and Dec 2009, 00Z and 12Z, 30 cycles or cases
- 45km horizontal resolution and 41 vertical levels
- Limited area model: Continuous United States (CONUS)
- Initial and boundary conditions from GFS (downscaled from NCEPs Global Forecast System)
- Verification against 3003 surface observations from the aviation routine weather report measurements (METAR) (and 106 soundings)
- Observation error not taken into account
### Model-error Experiments

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Model-error representation</th>
<th>Color</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNTL</td>
<td>Control Physics</td>
<td>blue</td>
<td>Hacker et al. (2011b)</td>
</tr>
<tr>
<td>PARAM</td>
<td>Multi-parameter scheme</td>
<td>cyan</td>
<td>Hacker et al. (2011a)</td>
</tr>
<tr>
<td>SKEBS</td>
<td>Stochastic kinetic-energy backscatter scheme</td>
<td>red</td>
<td>Berner et al. (2011)</td>
</tr>
<tr>
<td>SPPT</td>
<td>Stochastically perturbed physics tendencies</td>
<td>orange</td>
<td>Palmer et al. (2009)</td>
</tr>
<tr>
<td>PHYS10</td>
<td>Multi-physics (10 packages)</td>
<td>green</td>
<td>Hacker et al. (2011b)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Berner et al. (2011)</td>
</tr>
<tr>
<td>PHYS10.SKEBS</td>
<td>Multi-physics (10 packages) + + SKEBS</td>
<td>magenta</td>
<td>Berner et al. (2011)</td>
</tr>
<tr>
<td>PHYS3.SKEBS.PARAM</td>
<td>Limited multi-physics + (3 packages) + PARAM + SKEBS</td>
<td>black</td>
<td>Hacker et al. (2011b)</td>
</tr>
</tbody>
</table>
### Stochastic parameterization schemes

<table>
<thead>
<tr>
<th>Stochastic kinetic-energy backscatter scheme (SKEBS)</th>
<th>Stochastically perturbed parameterization scheme (SPPT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rationale: A fraction of the dissipated kinetic-energy is scattered upscale and available as forcing for the resolved flow (Shutts, 2005, Mason and Thomson 1992)</td>
<td>Rationale: Especially as resolution increases, the equilibrium assumption is no longer valid and fluctuations of the subgrid-scale state should be sampled (Buizza et al. 1999, Palmer et al. 2009)</td>
</tr>
</tbody>
</table>
Potential to reduce model error

- Stochastic parameterizations can change the mean and variance of a PDF
- Impacts variability of model (e.g. internal variability of the atmosphere)
- Impacts systematic error (e.g. blocking precipitation error)
## Multi-Physics combinations

<table>
<thead>
<tr>
<th>Member</th>
<th>Land Surface</th>
<th>Microphysics</th>
<th>PBL</th>
<th>Cumulus</th>
<th>Longwave</th>
<th>Shortwave</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Thermal</td>
<td>Kessler</td>
<td>YSU</td>
<td>KF</td>
<td>RRTM</td>
<td>Dudhia</td>
</tr>
<tr>
<td>2</td>
<td>Thermal</td>
<td>WSM6</td>
<td>MYJ</td>
<td>KF</td>
<td>RRTM</td>
<td>CAM</td>
</tr>
<tr>
<td>3</td>
<td>Noah</td>
<td>Kessler</td>
<td>MYJ</td>
<td>BM</td>
<td>CAM</td>
<td>Dudhia</td>
</tr>
<tr>
<td>4</td>
<td>Noah</td>
<td>Lin</td>
<td>MYJ</td>
<td>Grell</td>
<td>CAM</td>
<td>CAM</td>
</tr>
<tr>
<td>5</td>
<td>Noah</td>
<td>WSM6</td>
<td>YSU</td>
<td>KF</td>
<td>RRTM</td>
<td>Dudhia</td>
</tr>
<tr>
<td>6</td>
<td>Noah</td>
<td>WSM6</td>
<td>MYJ</td>
<td>Grell</td>
<td>RRTM</td>
<td>Dudhia</td>
</tr>
<tr>
<td>7</td>
<td>RUC</td>
<td>Lin</td>
<td>YSU</td>
<td>BM</td>
<td>CAM</td>
<td>Dudhia</td>
</tr>
<tr>
<td>8</td>
<td>RUC</td>
<td>Eta</td>
<td>MYJ</td>
<td>KF</td>
<td>RRTM</td>
<td>Dudhia</td>
</tr>
<tr>
<td>9</td>
<td>RUC</td>
<td>Eta</td>
<td>YSU</td>
<td>BM</td>
<td>RRTM</td>
<td>CAM</td>
</tr>
<tr>
<td>10</td>
<td>RUC</td>
<td>Thompson</td>
<td>MYJ</td>
<td>Grell</td>
<td>CAM</td>
<td>CAM</td>
</tr>
</tbody>
</table>

Table 2. Configuration of the multi-physics ensemble. Abbreviations are: BM – Betts-Miller; CAM – Community Atmosphere Model; KF – Kain-Fritsch; MYJ – Mellor-Yamada-Janjic; RRTM – Rapid Radiative Transfer Model; RUC – Rapid Update Cycle; WSM6 – WRF Single-Moment Six-class; YSU – Yonsei University. For details on the physical parameterization packages and references see Skamarock et al. (2008).
Spread, Error and Brierscore
(Threshold $0 < x < +\sigma$)

Spread — dashed
RMS Error — solid

"CNTL < PARAM < SKEBS < PHYS10 < PHYS3_SKEBS_PARAM < PHYS10_SKEBS"
Decomposition of Brierscore

\[ BS = \frac{1}{N} \sum_{k=1}^{K} n_k (p_k - \bar{p}_k)^2 + \frac{1}{N} \sum_{k=1}^{K} n_k (\bar{p}_k - \bar{p})^2 + \bar{p}(1 - \bar{p}) \]

- Are model-error schemes merely influencing reliability (linked to spread) or also resolution?
  - Resolution and reliability are both increased
  - Order stays more or less the same
Brier skill score

\[ BSS_{exp} = \frac{BS_{ref} - BS_{exp}}{BS_{ref}}. \]

- Normally climatology is reference, but here CNTL
- Measures skill improvements of model-error schemes over CNTL
What if all ensemble systems had the same spread?

- **Calibrate** all ensemble systems so that each member has the same variability as the observations (Doblas-Reyes et al., 2005; Kharin and Zwiers, 2003, Hamill and Colucci, 1998)

- Calibrated ensemble systems will have similar spread

- Just another way of assessing the role of spread
Calibration of ensemble systems

\[ z_{ij} = \alpha \mu_i + \beta x_{ij}; \quad \text{with} \quad \alpha = \rho \frac{s_r}{s_{em}} \quad \text{and} \quad \beta_i = s_r \frac{\sqrt{1 - \rho^2}}{s_{e,i}}, \]

Fullfills two conditions:

1. The variance of the inflated prediction is the same as that of the reference (here observations) (Hamill and Colucci, 1998)

2. The potentially predictable signal after inflation is made equal to the correlation of the ensemble mean with the observations (Kharin and Zwiers, 2003a)
   
   Correlation between ensemble mean and observations is not changed
Model-error schemes have different resolution and reliability

Ability to represent structural uncertainty better?

“CNTL < PARAM < SKEBS < PHYS10 < {PHYS3_SKEBS_PARAM < PHYS10_SKEBS}”
Impact of other postprocessing methods

- Debiasing
- A combination of debiasing
Impact of changing model-versions

- Now SPPT, but not PARAM
- Multi-physics has changed and is now clearly overdispersive => big problem for maintenance
Bias

![Zonal Wind U at 700hPa](image)

![Temperature T at 700hPa](image)

![Zonal Wind U at 10m](image)

![Temperature at 2m](image)
What is the relative impact of postprocessing (debiasing, calibration, changing model versions)

Compute skill of postprocesses ensemble forecast over raw ensemble forecasts

Done for each experiment separately, e.g. SKEBS postprocesses is compared to SKEBS raw

Take the average of the skill score over all forecast times (exclude initial time)
Impact Summary

\[
BSS_{\text{exp,post}} = \frac{BSS_{\text{exp,raw}} - BSS_{\text{exp,post}}}{BSS_{\text{exp,raw}}} = \frac{Rel_{\text{exp,raw}} - Rel_{\text{exp,post}}}{BSS_{\text{exp,raw}}} + \frac{Res_{\text{exp,raw}} - Res_{\text{exp,post}}}{BSS_{\text{exp,raw}}} + \frac{Unc_{\text{exp,raw}} - Unc_{\text{exp,post}}}{BSS_{\text{exp,raw}}}
\]

Raw

Calibrated
Results hold qualitatively for other verification thresholds

Results significant at 95% except PARAM

If obs error is included results still hold qualitatively, but significance is reduced

In free atmosphere SKEBS tends to outperform PHYS10
Model-error schemes improve forecast skill by improving both, reliability and resolution.

The impact is of comparable magnitude to that of common postprocessing methods.

Combining multiple model-error schemes yields consistently best results.
SKEBS tutorial and informal discussion group

- **Stochastic Kinetic-Energy Backscatter Scheme**
- Friday, 10:30 – 12:00
- Tutorial
- Informal discussion group
  - Feedback to developers
  - Network with other users, especially on applications
Skill of calibrated forecasts

- At 12h difference between RMSE and spread was ca. 1.3m/s and 2.0K, now 0.4m/s and 0.2K
- Spread between ensemble systems is much closer
- Brier score increases
- Combination of multiple model-error schemes performs still best, but has still most spread.
- Multi-physics performs very well