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

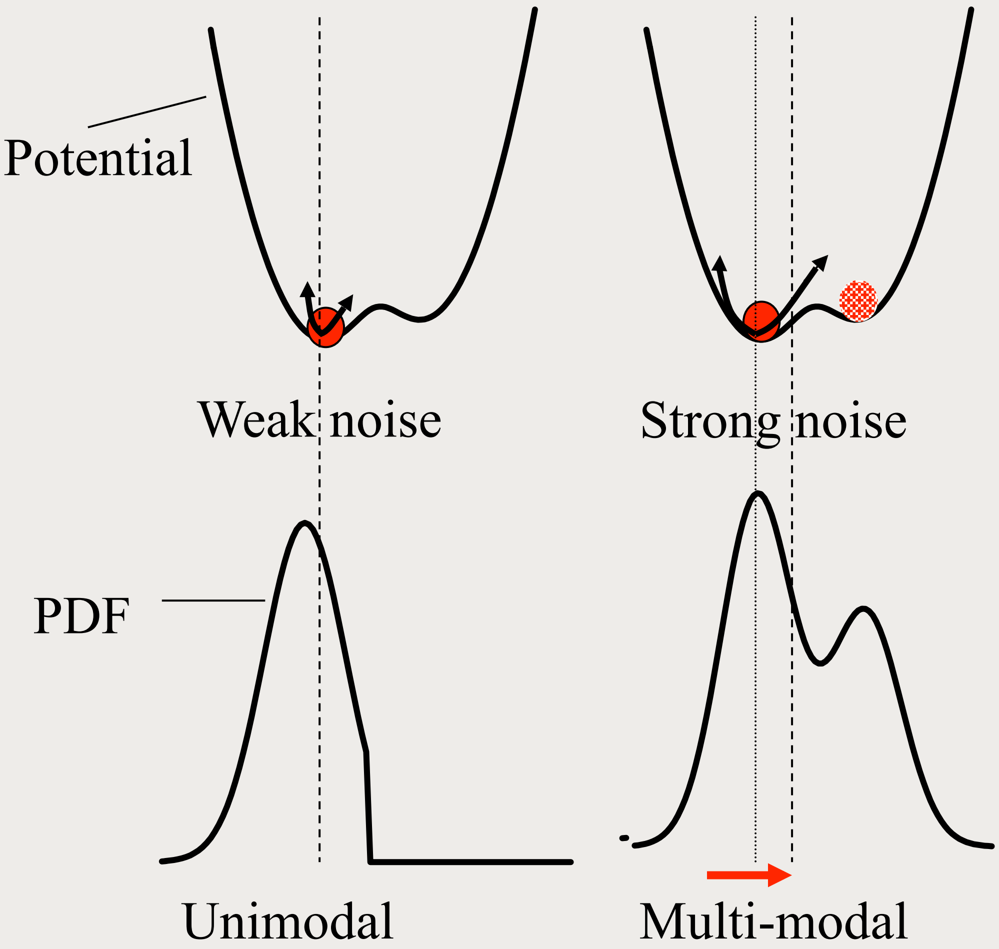
Experiment	Model-error representation	Color	Reference
CNTL	Control Physics	blue	Hacker et al. (2011b)
PARAM	Multi-parameter scheme	cyan	Hacker et al. (2011a)
SKEBS	Stochastic kinetic-energy backscatter scheme	red	Berner et al. (2011)
SPPT	Stochastically perturbed physics tendencies	orange	Palmer et al. (2009)
PHYS10	Multi-physics (10 packages)	green	Hacker et al. (2011b) Berner et al. (2011)
PHYS10_SKEBS	Multi-physics (10 packages) + + SKEBS	magenta	Berner et al. (2011)
PHYS3_SKEBS_PARAM	Limited multi-physics + (3 packages) + PARAM + SKEBS	black	Hacker et al. (2011b)

Stochastic parameterization schemes

Stochastic kinetic-energy backscatter scheme (SKEBS)	Stochastically perturbed parameterization scheme (SPPT)
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(Shutts, 2005, Mason and Thomson 1992)	Rationale: Especially as resolution increases, the equilibrium assumption is no longer valid and fluctuations of the subgrid-scale state should be sampled (Buizza et al. 1999, Palmer et al. 2009)

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

Member	Land Surface	Microphysics	PBL	Cumulus	Longwave	Shortwave
1	Thermal	Kessler	YSU	KF	RRTM	Dudhia
2	Thermal	WSM6	MYJ	KF	RRTM	CAM
3	Noah	Kessler	MYJ	BM	CAM	Dudhia
4	Noah	Lin	MYJ	Grell	CAM	CAM
5	Noah	WSM6	YSU	KF	RRTM	Dudhia
6	Noah	WSM6	MYJ	Grell	RRTM	Dudhia
7	RUC	Lin	YSU	BM	CAM	Dudhia
8	RUC	Eta	MYJ	KF	RRTM	Dudhia
9	RUC	Eta	YSU	BM	RRTM	CAM
10	RUC	Thompson	MYJ	Grell	CAM	CAM

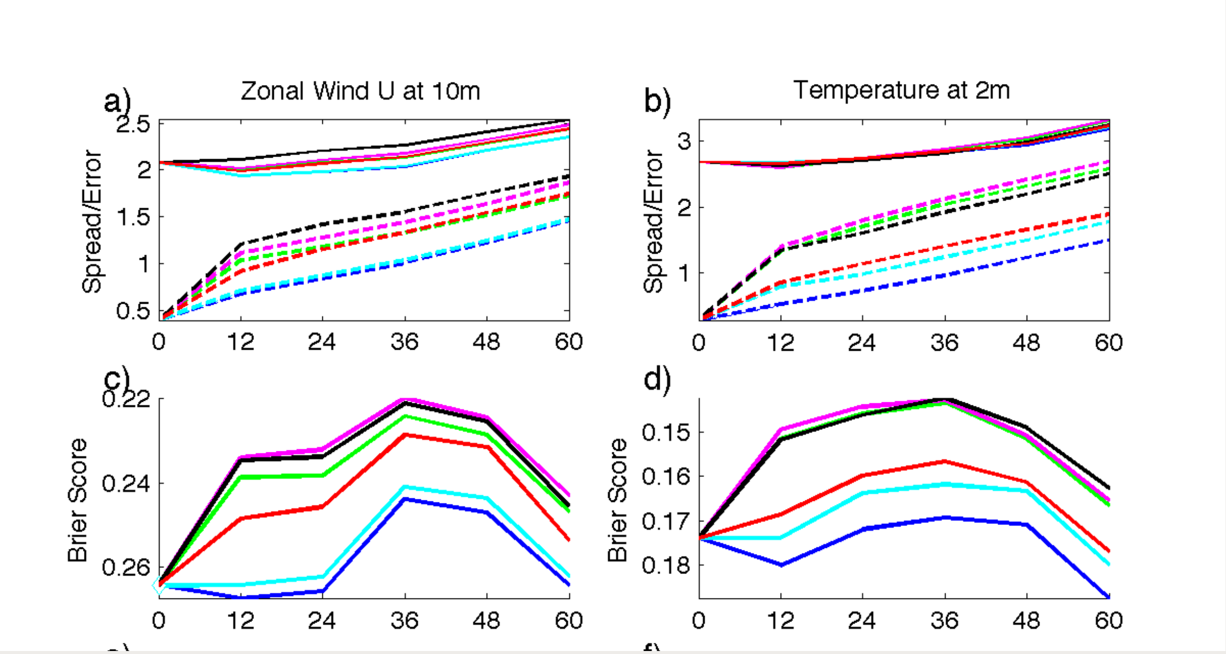
Table 2. Configuration of the multi-physics ensemble. Abbreviations are: BM – Betts-Miller; CAM – Community Atmosphere Model; KF – Kain-Fritsch; MYJ – Mellor-Yamada-Janjic; RRTM – Rapid Radiative Transfer Model; RUC – Rapid Update Cycle; WSM6 – WRF Single-Moment Six-class; YSU – Yonsei University. For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

Spread, Error and Brierscore

(Threshold o<x< +σ)

CNTL	blue
PARAM	cyan
SKEBS	red
PHYS10	green
PHYS10_SKEBS	magenta
PHYS3_SKEBS_PARAM	black

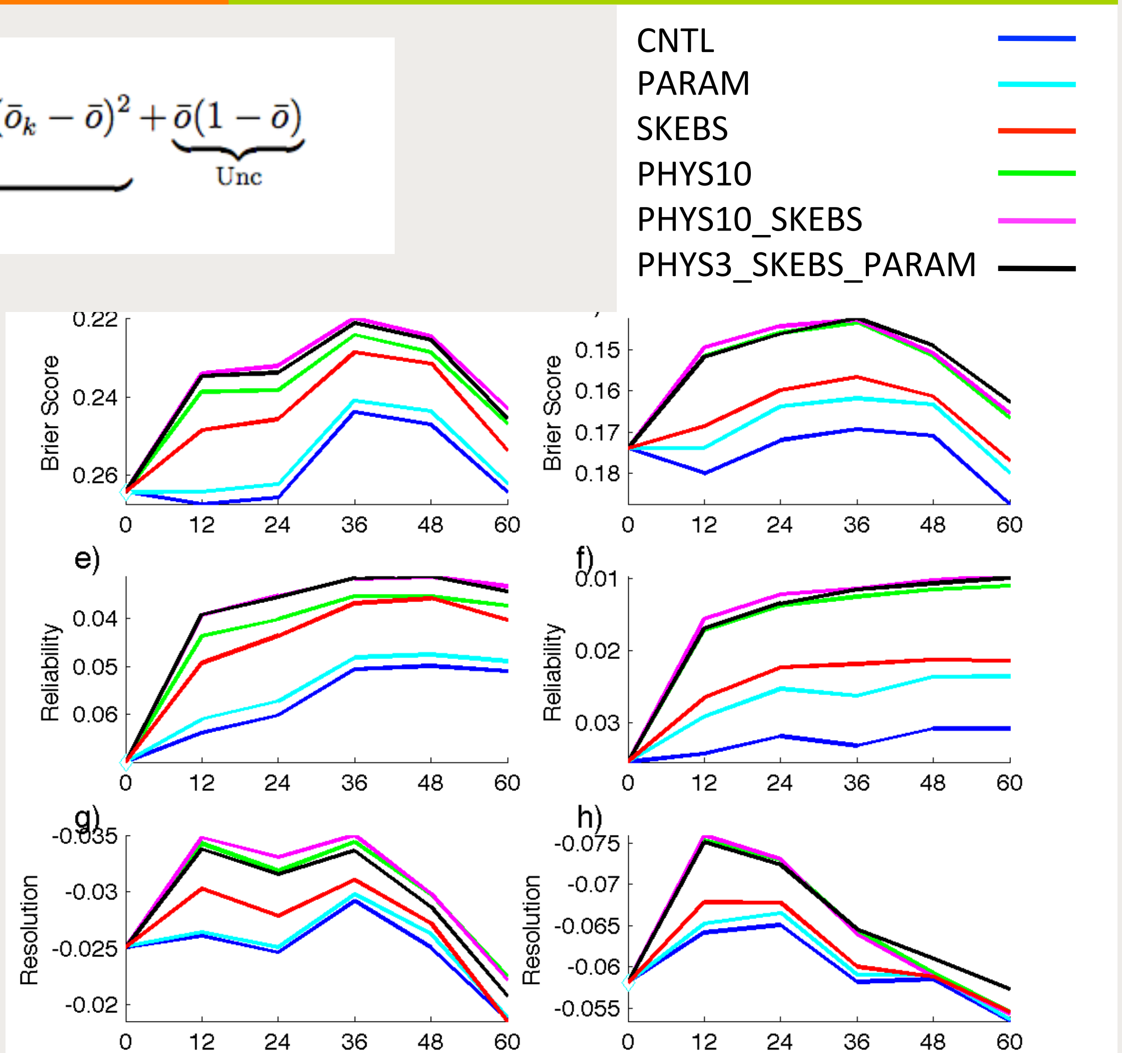
Spread – dashed
RMS Error – solid



"CNTL < PARAM < SKEBS < PHYS10 < PHYS3_SKEBS_PARAM < PHYS10_SKEBS"

Decomposition of Brierscore

$$BS = \underbrace{\frac{1}{N}\sum_{k=1}^{K} n_k(p_k - \bar{o}_k)^2}_{\text{Res}} + -\underbrace{\frac{1}{N}\sum_{k=1}^{K} n_k(\bar{o}_k - \bar{o})^2}_{\text{Rel}} + \underbrace{\bar{o}(1-\bar{o})}_{\text{Unc}}$$



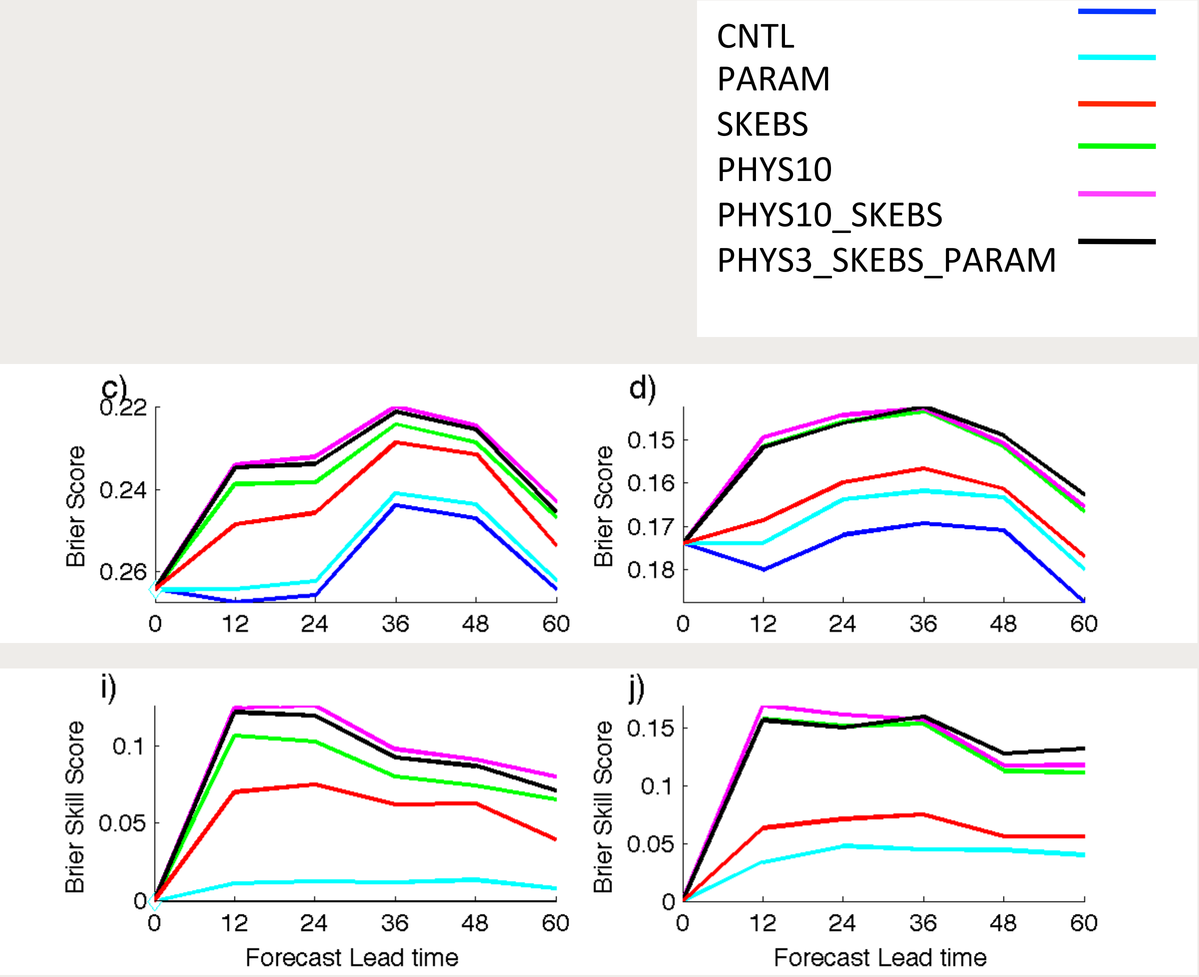
Are model-error schemes merely influencing reliability (linked to spread) or also resolution?

- Resolution and reliability are both increased
- Order stays more or less the same

Brier skillscore

$$\mathrm{BSS_{exp}} = \frac{\mathrm{BS_{ref}} - \mathrm{BS_{exp}}}{\mathrm{BS_{ref}}}.$$

- 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

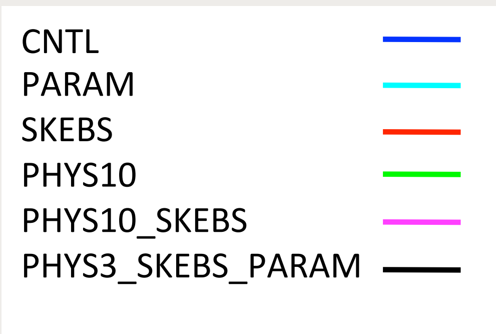
$$z_{ij} = \alpha\mu_i + \beta x_{ij}; \quad \text{with} \quad \alpha = \rho\frac{s_r}{s_{em}} \quad \text{and} \quad \beta_i = s_r\frac{\sqrt{1-\rho^2}}{s_{e,i}},$$

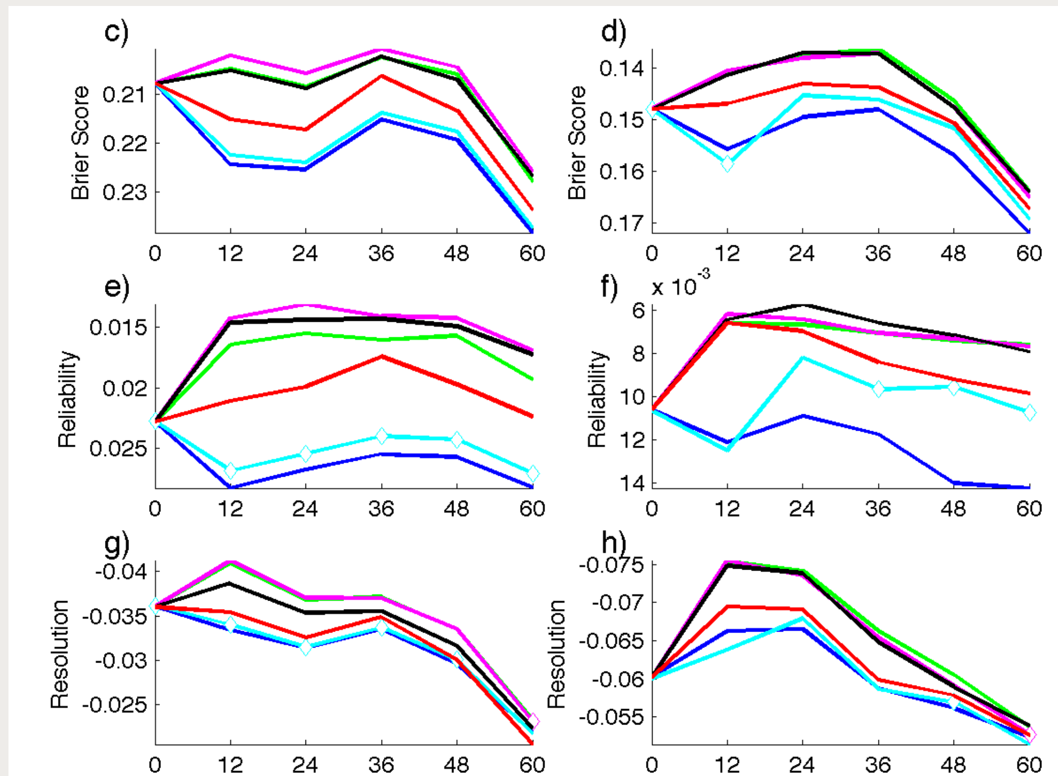
Fullfills two conditions:

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

Brier score of calibrated ensemble systems



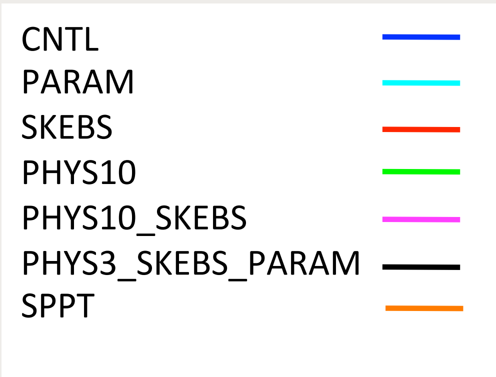


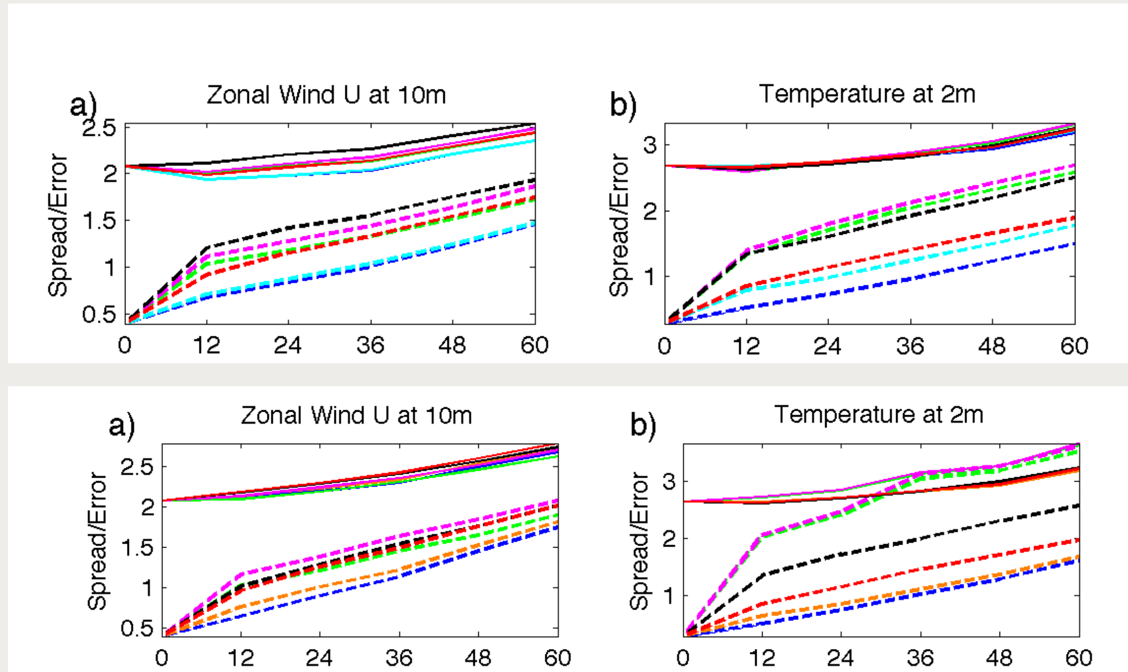
- Model-error schemes have different resolution and reliability
- Ability to represent structural uncertainty better?

"CNTL < PARAM < SKEBS < PHYS10 < {PHYS3_SKEBS_PARAM < PHYS10_SKEBS}"

Impact of other postprocessing methods

- Debiasing
- A combination of debiasing

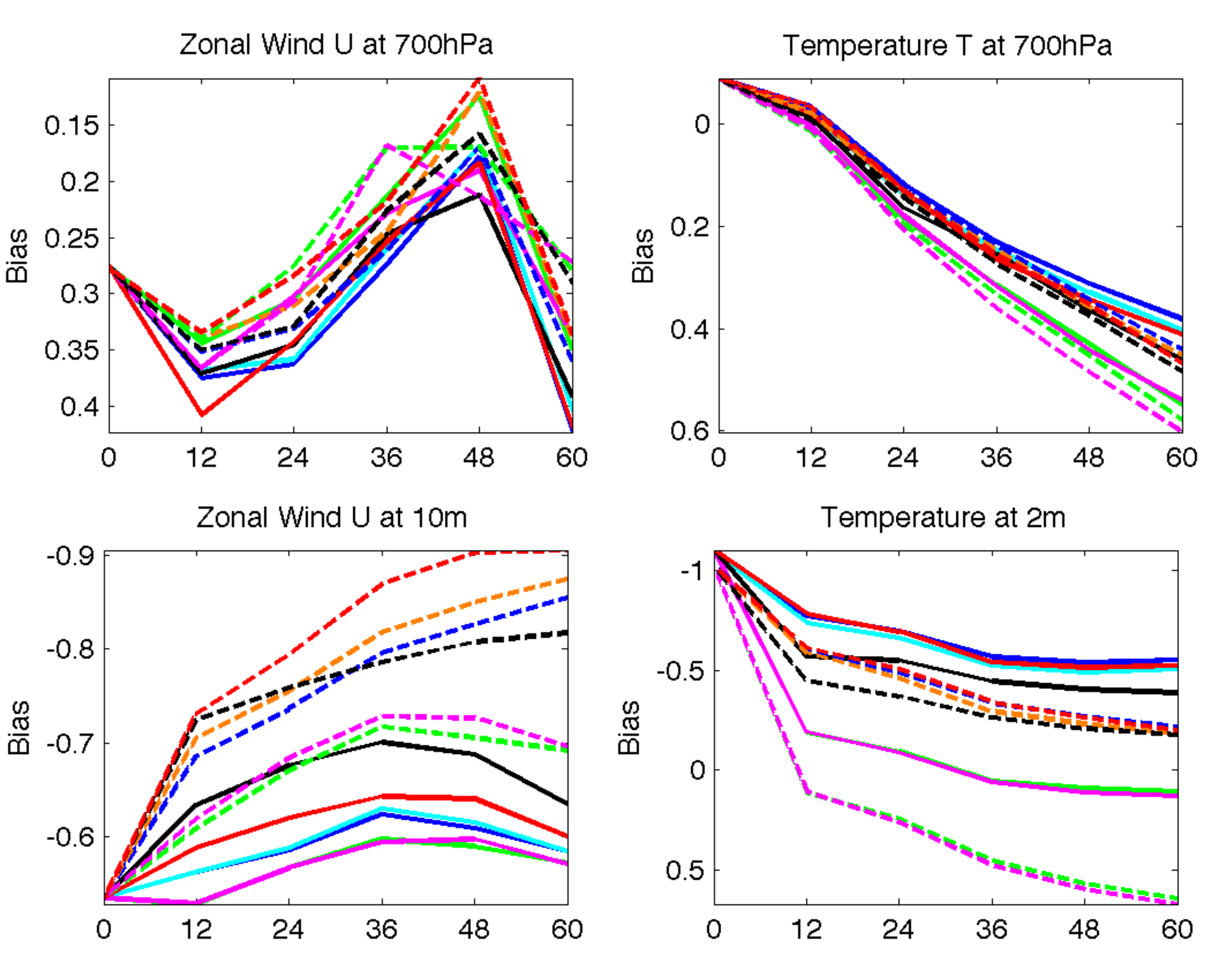




Impact of changing model-versions

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

Bias



Zonal Wind U at 700hPa
Bias
0.15
0.2
0.25
0.3
0.35
0.4
0
12
24
36
48
60
Temperature T at 700hPa
Bias
0
0.2
0.4
0.6
0
12
24
36
48
60
Zonal Wind U at 10m
Bias
-0.9
-0.8
-0.7
-0.6
0
12
24
36
48
60
Temperature at 2m
Bias
-1
-0.5
0
0.5
0
12
24
36
48
60

Impact Summary

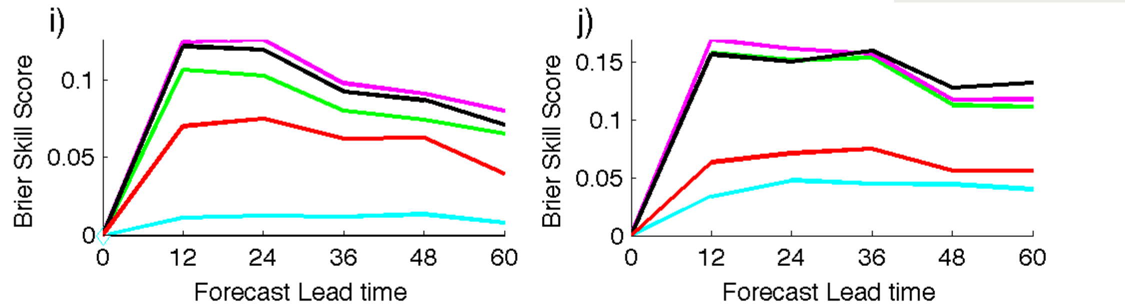
$$\begin{aligned}\mathrm{BSS}_{\mathrm{exp,post}} &= \frac{\mathrm{BS}_{\mathrm{exp,raw}} - \mathrm{BS}_{\mathrm{exp,post}}}{\mathrm{BS}_{\mathrm{exp,raw}}} \\ &= \frac{\mathrm{Rel}_{\mathrm{exp,raw}} - \mathrm{Rel}_{\mathrm{exp,post}}}{\mathrm{BS}_{\mathrm{exp,raw}}} + \frac{\mathrm{Res}_{\mathrm{exp,raw}} - \mathrm{Res}_{\mathrm{exp,post}}}{\mathrm{BS}_{\mathrm{exp,raw}}} + \frac{\mathrm{Unc}_{\mathrm{exp,raw}} - \mathrm{Unc}_{\mathrm{exp,post}}}{\mathrm{BS}_{\mathrm{exp,raw}}}\end{aligned}$$

- 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)

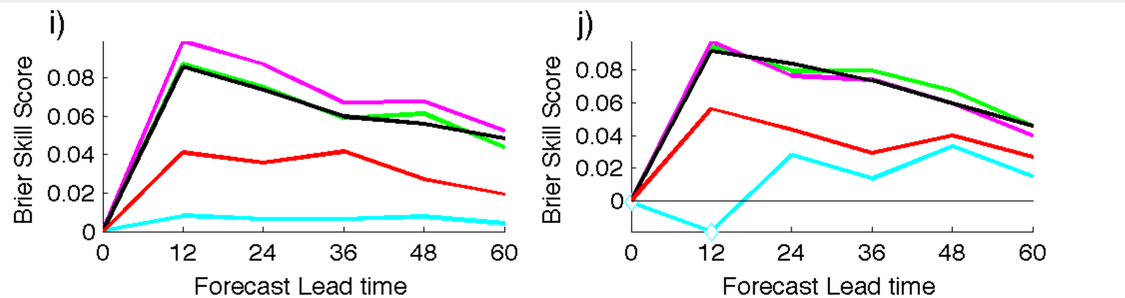
Impact Summary

$$\begin{aligned}\mathrm{BSS}_{\mathrm{exp,post}} &= \frac{\mathrm{BS}_{\mathrm{exp,raw}} - \mathrm{BS}_{\mathrm{exp,post}}}{\mathrm{BS}_{\mathrm{exp,raw}}} \\ &= \frac{\mathrm{Rel}_{\mathrm{exp,raw}} - \mathrm{Rel}_{\mathrm{exp,post}}}{\mathrm{BS}_{\mathrm{exp,raw}}} + \frac{\mathrm{Res}_{\mathrm{exp,raw}} - \mathrm{Res}_{\mathrm{exp,post}}}{\mathrm{BS}_{\mathrm{exp,raw}}} + \frac{\mathrm{Unc}_{\mathrm{exp,raw}} - \mathrm{Unc}_{\mathrm{exp,post}}}{\mathrm{BS}_{\mathrm{exp,raw}}}\end{aligned}$$

Raw



Calibrated



Brierskillscore | Reliability Component | Resolution component

CNTL
PARAM
SKEBS
PHYS10
PHYS10_SKEBS
PHYS3_SKEBS_PARAM

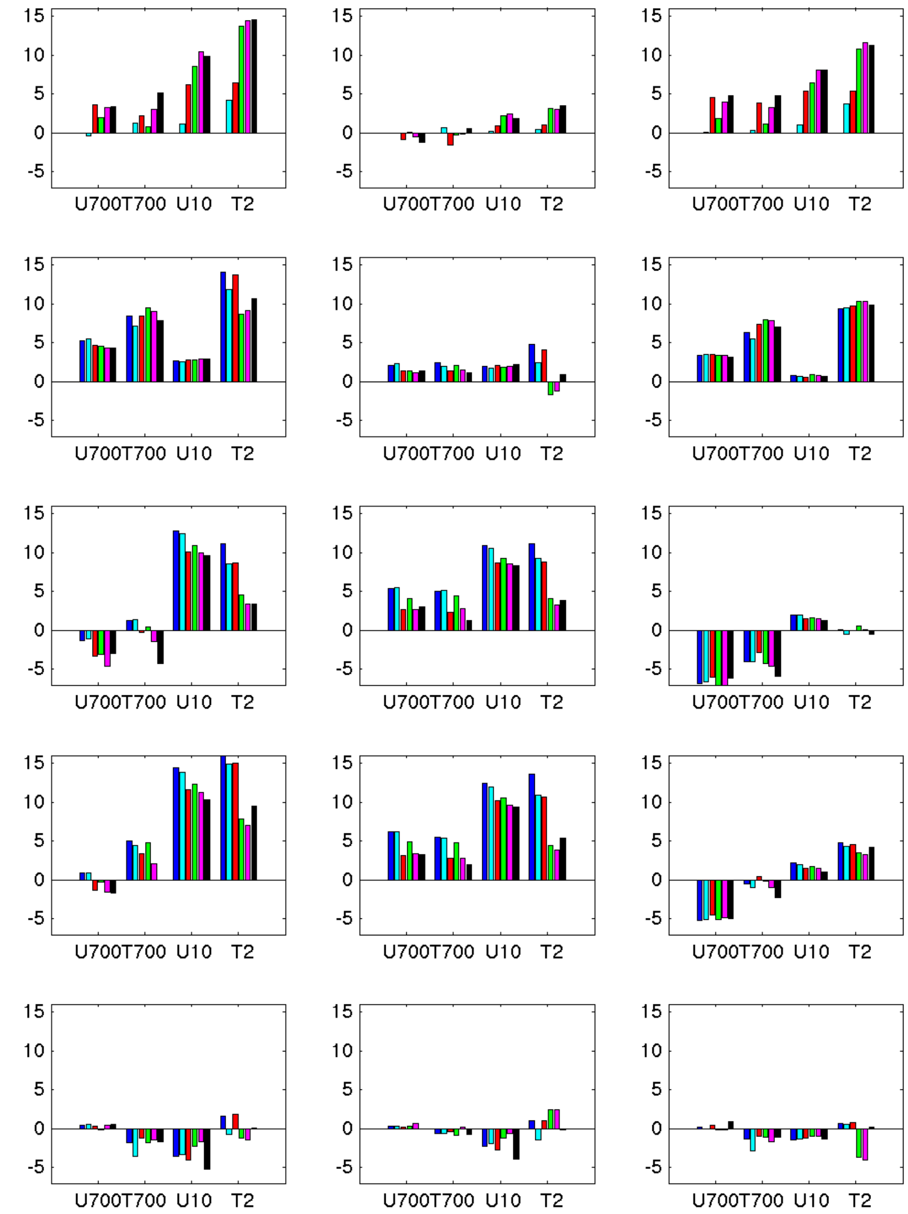
Model-error scheme

Debiasing

Calibration

Calibration & Debiasing

Changing model-versions



Comments

- 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 outperfrom PHYS10

Conclusions

- 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 SKEBS was not directly developed for

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