Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations

16th WRF User workshop, Jun 15-19rd, 2015

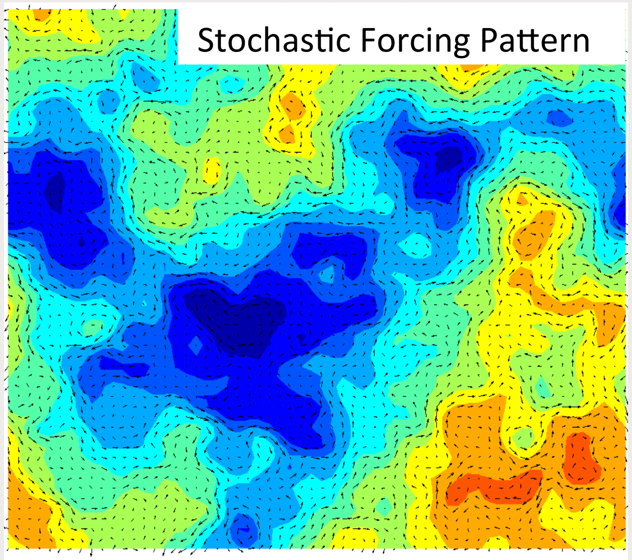
Judith Berner, Kate Fossell, Soyoung Ha, Josh Hacker, Chris Snyder

Outline

- Model-error representations increase the reliability of ensemble systems and improve forecast skill.
- Is this simple the result of increased reliability and decreased bias does the benefit of model-error schemes go beyond that?
  - Forecast are post-processed to remove bias and calibrated to have the same spread
  - Quantify the skill of post-processed models with and without model-error

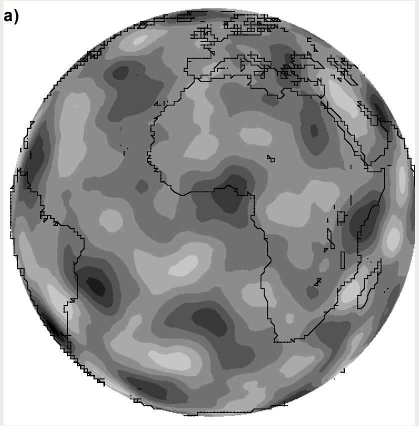
Summary of experiments

Experiment	Model-error representation	Color	Reference
CNTL	Control Physics	blue	Hacker et al. (2011b)
SKEBS	Stochastic kinetic-energy backscatter scheme	red	Berner et al. (2011)
PARAM	Multi-parameter	cyan	Hacker et al. (2011a)
SPPT	Stochastically perturbed physics tendencies	orange	Palmer et al. (2009)
PHYS4	Limited multi-physics (4 packages)	light green	Hacker et al. (2011b)
PHYS10	Multi-physics (10 packages)	dark green	Hacker et al. (2011b) Berner et al. (2011)
PHYS10_SKEBS	Multi-physics (10 packages) + + SKEBS	magenta	Berner et al. (2011)
PHYS4_SKEBS_PARAM	Limited multi-physics + (4 packages) + PARAM + SKEBS	black	Hacker et al. (2011b)



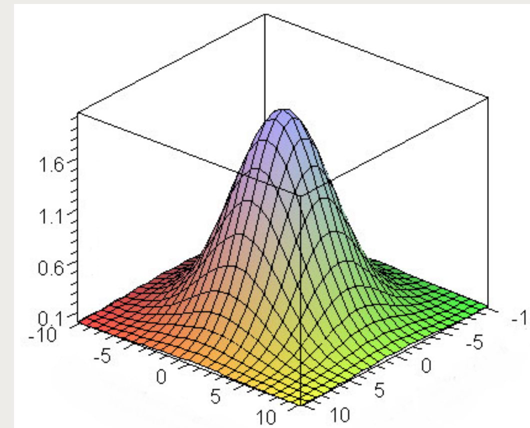
Stochastic-kinetic energy backscatter scheme (SKEBS)

- Rationale: A fraction of the subgrid-scale energy is scattered upscale and acts as random streamfunction and temperature forcing for the resolved-scale flow. Here: simply considered as additive noise with spatial and temporal correlations
- Similar to ECMWF global ensemble system (Shutts 2005, Berner et. al 08,09) but with constant dissipation rate and potential temperature perturbations (Berner et al. 2011).



Stochastically perturbed tendency scheme (SPPT)

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, Berner et al. 2014)

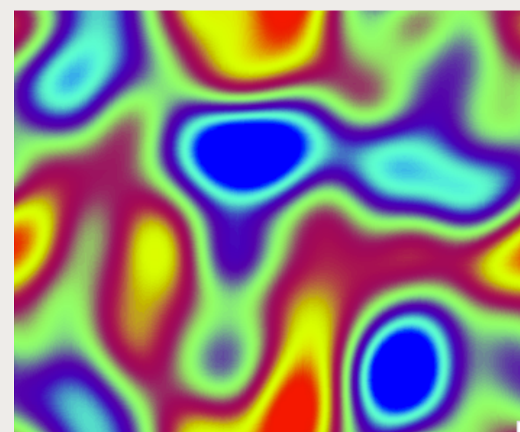


$$\frac{\partial X}{\partial t} = D_X + (r+1)P_X$$

Local tendency for variable X

Dynamical tendencies => Resolved scales

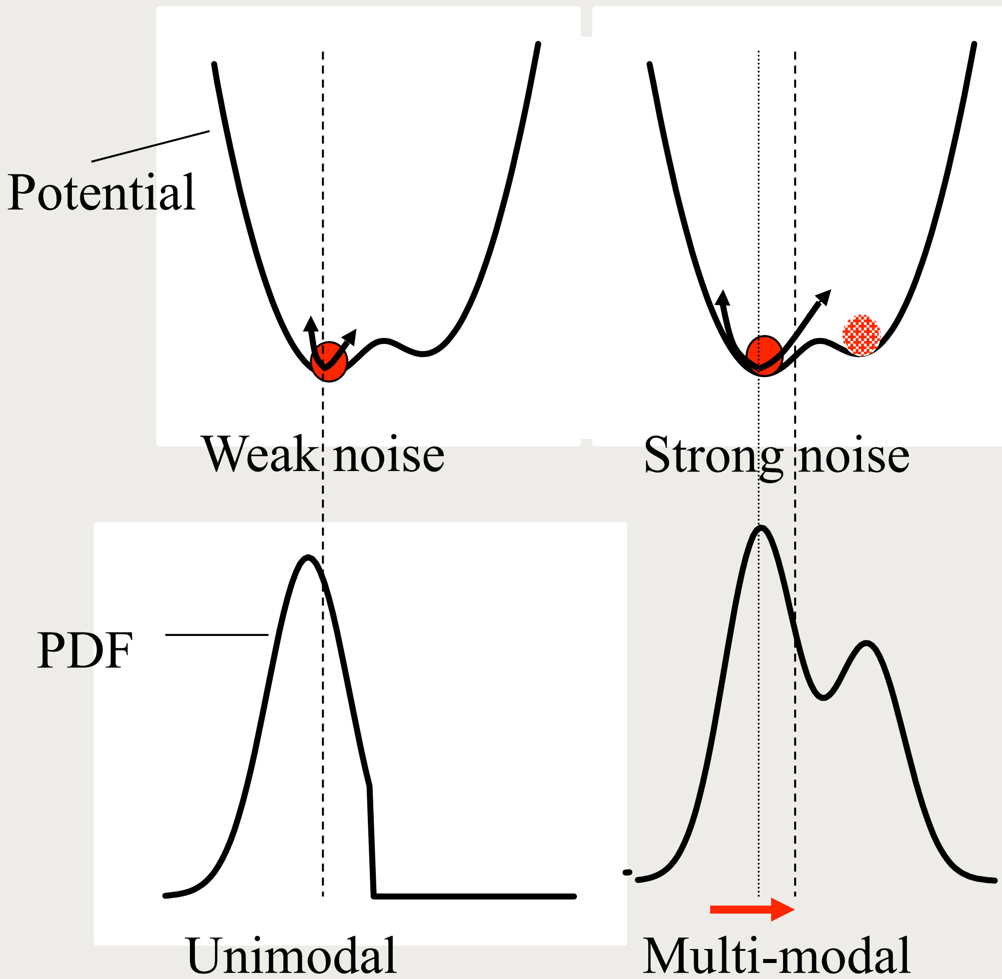
Physical tendencies => Unresolved scales



- ✧ Perturbs accumulated U,V,T,Q tendencies from physical parameterizations packages
- ✧ Same pattern for all tendencies to minimize introduction of imbalances

Potential of stochastic parameterizations to reduce model error

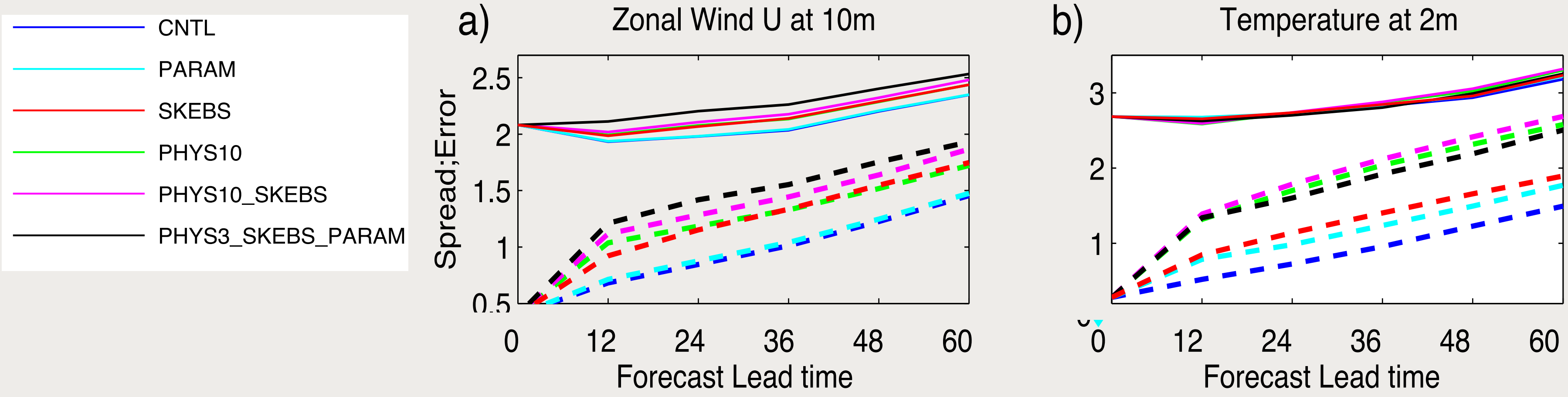
- Stochastic parameterizations can change the mean and variance of a PDF
- Impacts variability of model (e.g. internal variability of the atmosphere)
- Impacts systematic error (e.g. blocking precipitation error)
- Can trigger noise-induced regime transitions



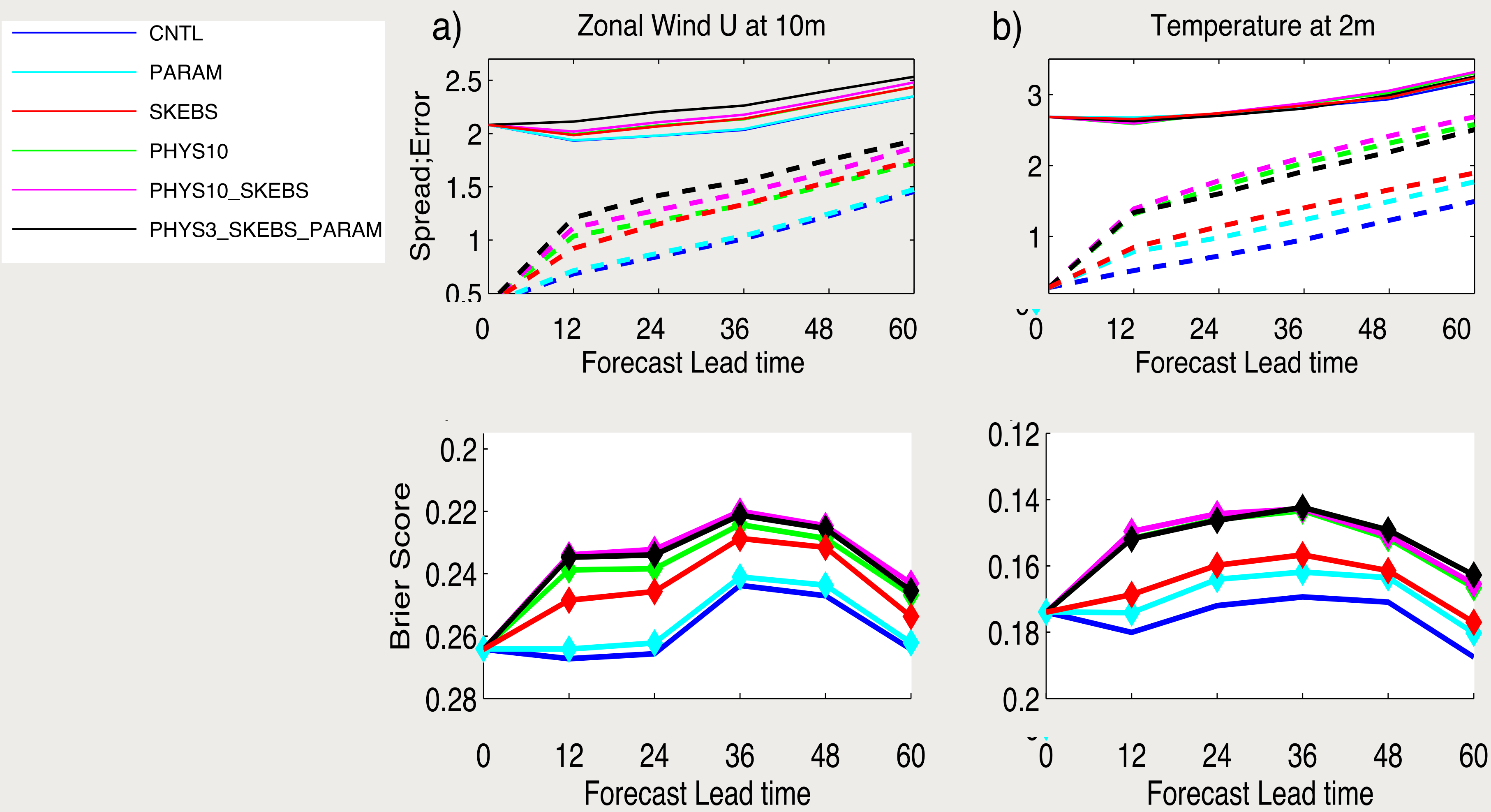
Experiment setup

- Weather Research and Forecast Model WRFV3.1.1 (or WRFV3.3.1)
- 45km horizontal resolution and 41 vertical levels
- 10-member ensemble, integrated for 60h (short-range forecast)
- 15 dates in Nov-Dec 2009, 00Z and 12Z, amounting to 30 cycles
- Limited area model: Contiguous United States (CONUS)
- Boundary and initial conditions are taken from GEFS
- Verification against observations (soundings and METAR)

Spread and error near the surface



Brierscore near the surface



Decomposition of the brier score

- Reliability is small (good) if number forecast probability in bin k equals the observed frequency

- Resolution is large (good) if forecast bins are different from the mean over the verification period.

$$\mathrm{BS} = \underbrace{\frac{1}{N}\sum_{k=1}^{K} n_k (p_k - o_k)^2}_{\text{Reliability}} - \underbrace{\frac{1}{N}\sum_{k=1}^{K} n_k (o_k - \bar{o})^2}_{\text{Resolution}} + \underbrace{\bar{o}(1-\bar{o})}_{\text{Uncertainty}}$$

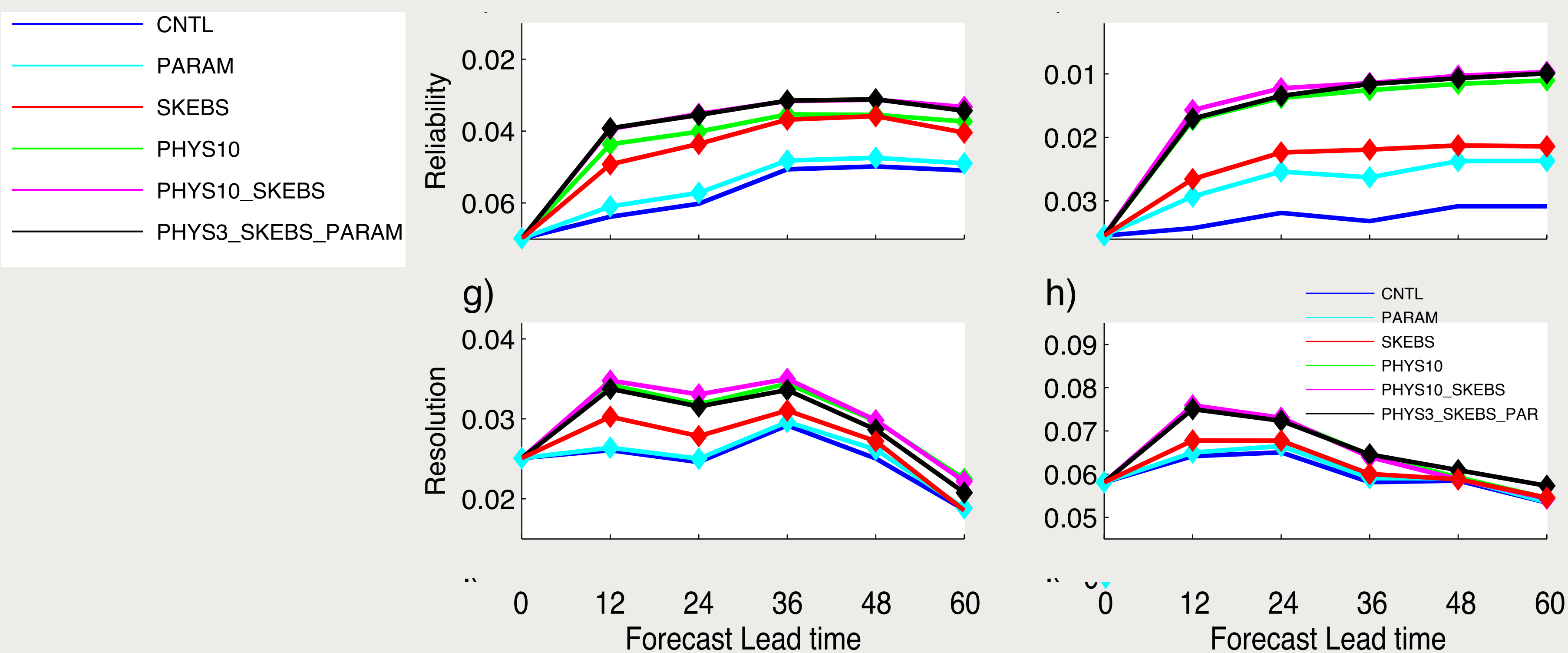
p_k : forecast probability value for bin k

o_k: observed frequency in bin k here observations)

n_k : number of forecasts that fall into bin k

N: total number of forecasts

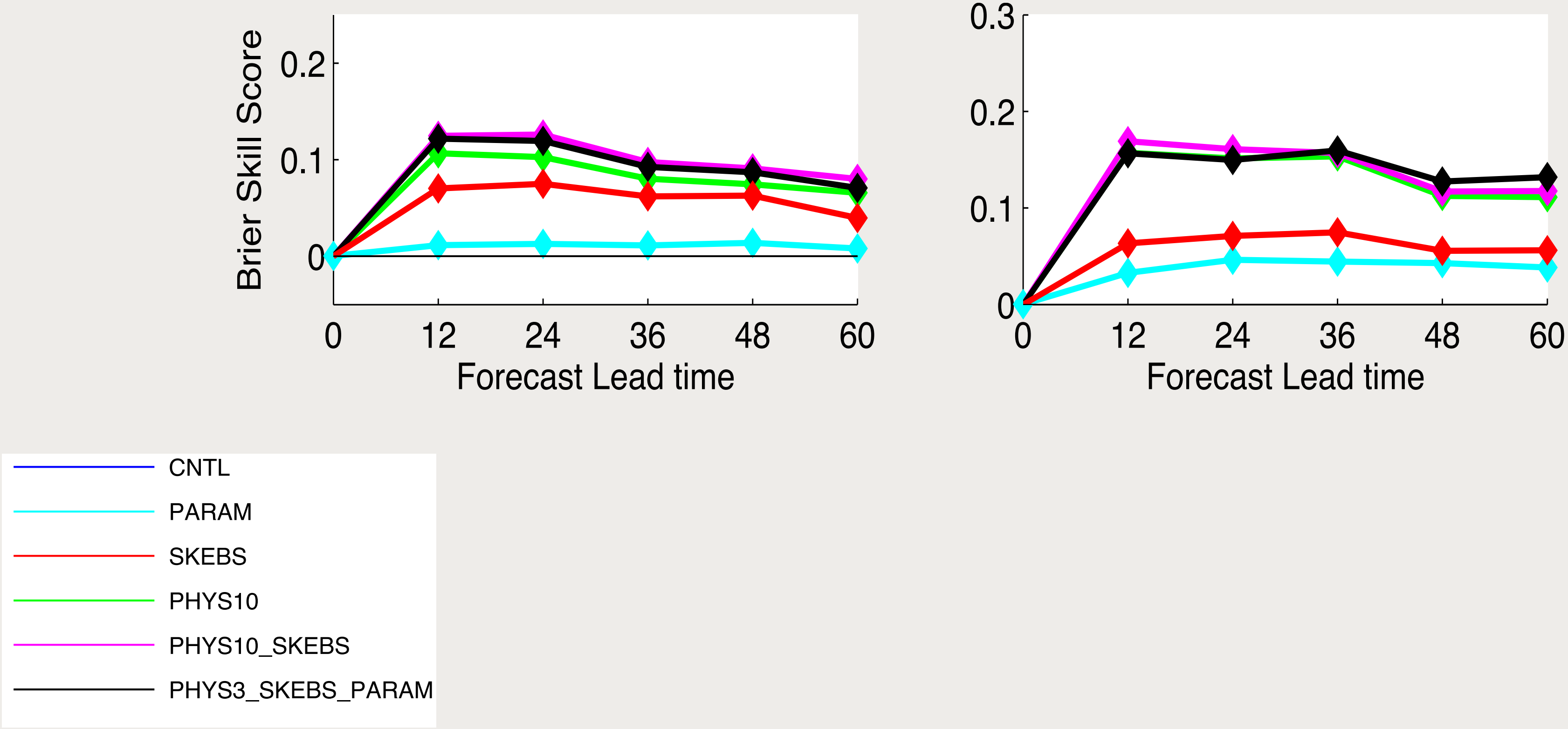
Reliability and resolution



Brier skill score

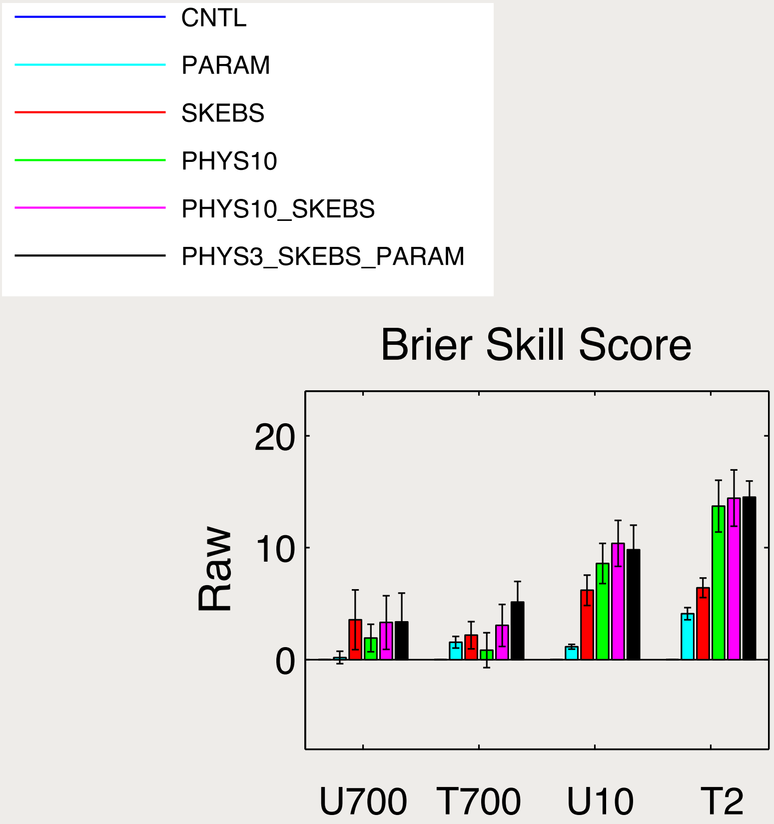
$$BSS_{exp} = \frac{BS_{ref} - BS_{exp}}{BS_{ref}}$$

where BS_{ref} is brier score of raw (unprocessed) CNTL

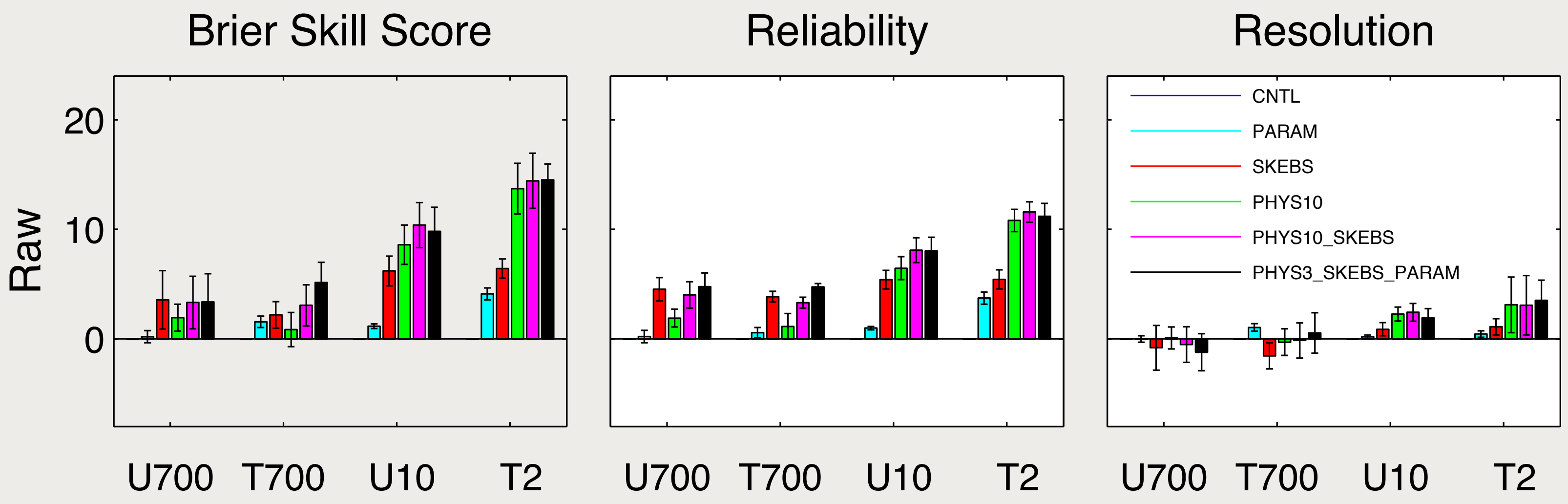


Relative skill improvement

- Average over all forecast lead times
- Variables are U700, T700, U10, T2



Relative skill improvement



Outline

- Model-error representations increase the reliability of ensemble systems and improve forecast skill.
- Is this simple the result of increased reliability and decreased bias does the benefit of model-error schemes go beyond that?
  - Ensemble forecasts are post-processed to remove bias and calibrated to have the same spread

Outline

- Model-error representations increase the reliability of ensemble systems and improve forecast skill.
- Is this simple the result of increased reliability and decreased bias does the benefit of model-error schemes go beyond that?
  - Ensemble Forecasts are calibrated to have the same spread
  - Since postprocessing methods are used as a an analytic tool, they are applied in-sample

Calibration

- Form of variance inflation but additionally insures that the potentially predictable signal after calibration is equal to the correlation of the ensemble mean with the observations (von Storch, 1999)

- Each calibrated ensemble member z_{ij} at each observation location is expressed as

$$z_{ij} = \alpha\mu_i + \beta x_{ij}$$

with

$$\alpha = \rho\frac{s_r}{s_{em}} \quad \text{and} \quad \beta = s_r\frac{\sqrt{1-\rho^2}}{s_e}.$$

- x_{ij}: : ensemble member j at time i before calibration
- μ_i :: ensemble mean
- α and β calibration parameters
- Index denoting observation location has been omitted
- ρ : correlation of ensemble mean with the reference
- s_e : spread
- s_{em} : standard deviation of ensemble mean
- s_r : standard deviation of reference

Calibration

Fullfills two conditions:

- the variance of each ensemble member is the same as that of a reference (here observations)

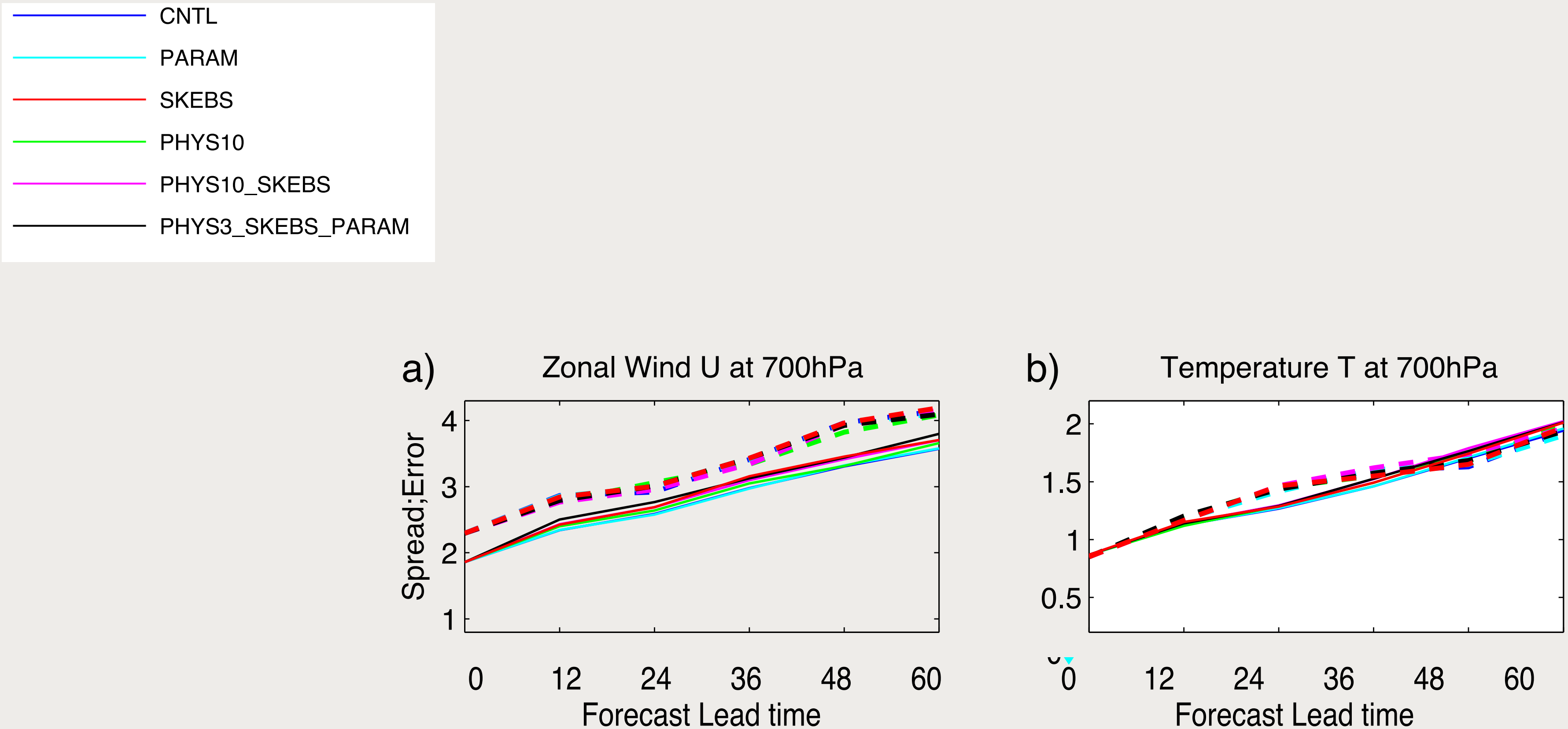
$$s_r^2 = \alpha^2 s_{\mathrm{em}}^2 + \beta^2 s_e^2 \,.$$

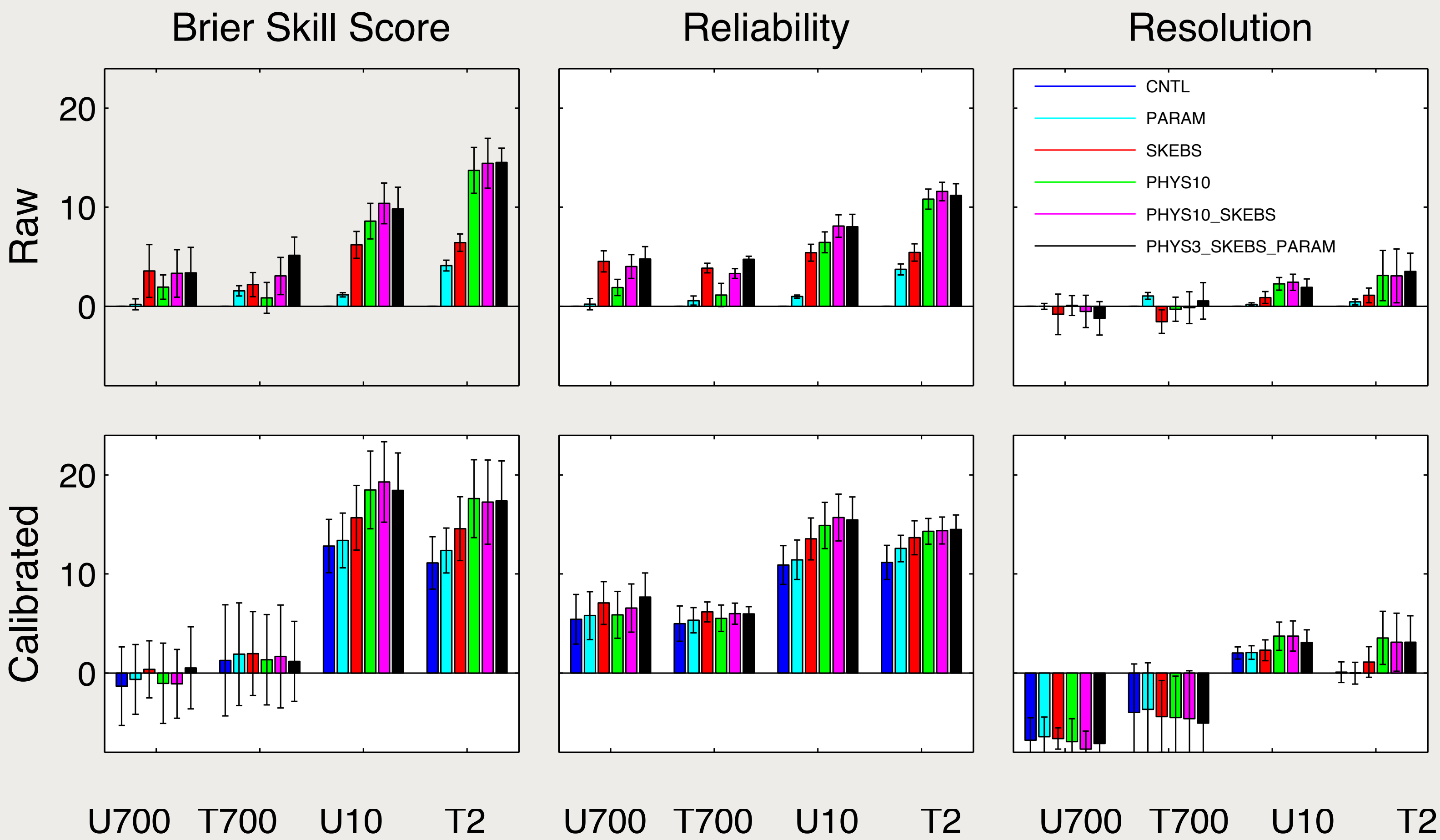
- the potentially predictable signal after calibration is equal to the correlation ρ of the ensemble mean with the observations (von Storch, 1999)

$$\rho = \frac{\mathrm{cov}(\mu, \mathrm{r})}{s_{\mathrm{em}}\, s_r} = \frac{\mathrm{cov}(\mu_{\mathrm{calib}}, \mathrm{r})}{s_{\mathrm{em,calib}}\, s_r}$$

  - s_e : spread
  - s_{em} : standard deviation of ensemble mean
  - s_r : standard deviation of reference

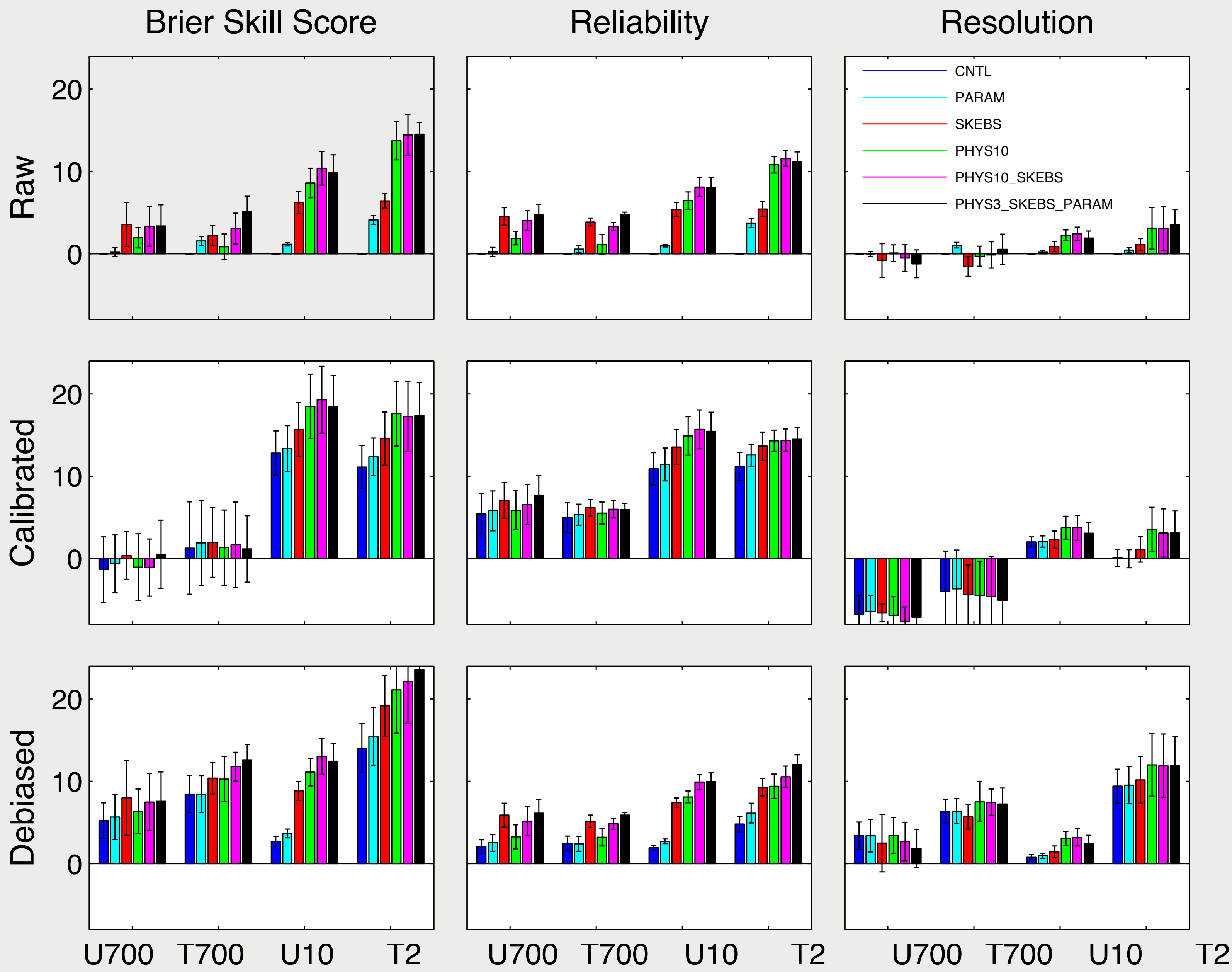
Impact of calibration in 700hPa

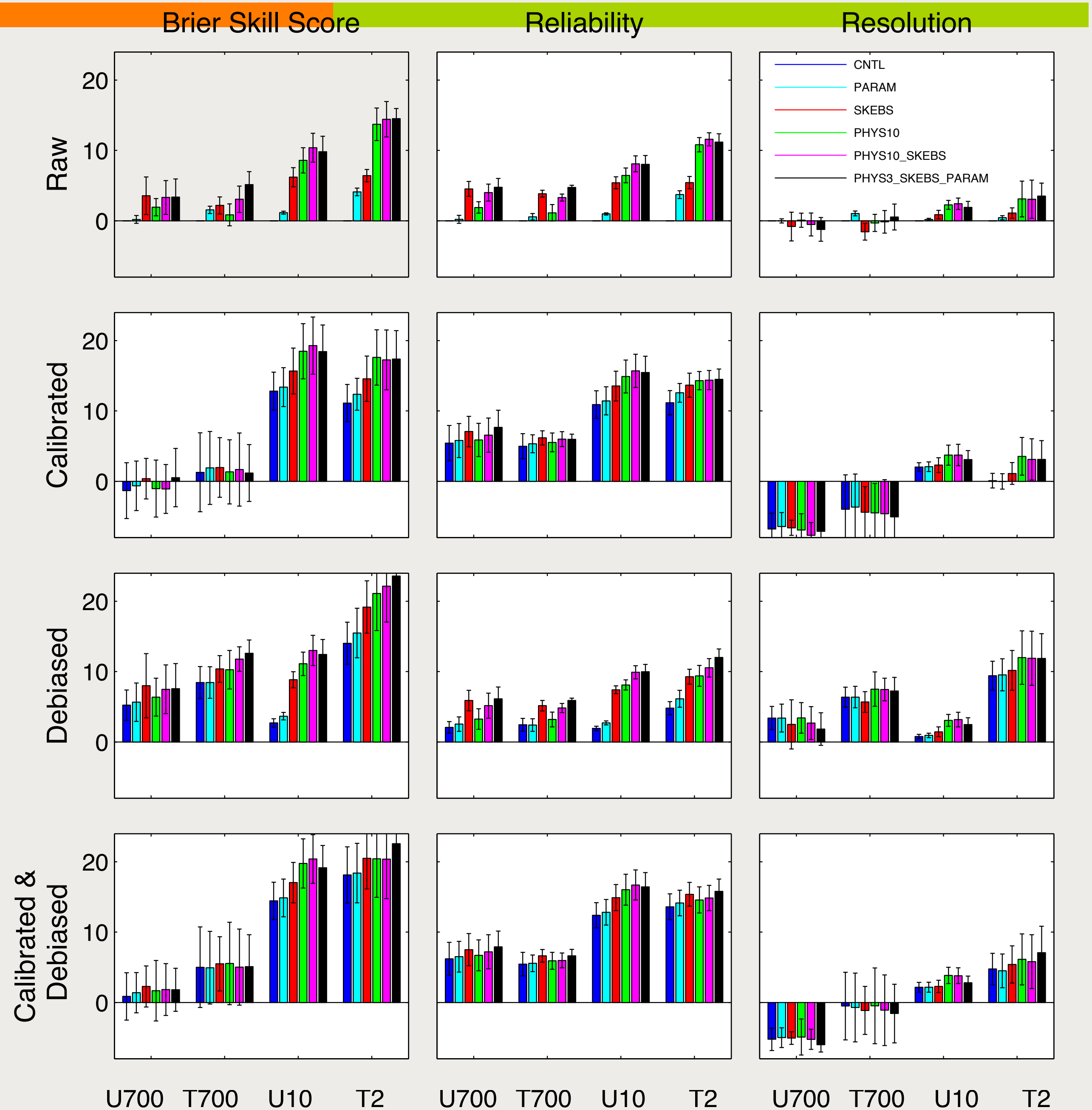



Brier Skill Score
Reliability
Resolution
Raw
Calibrated
20
10
0
CNTL
PARAM
SKEBS
PHYS10
PHYS10_SKEBS
PHYS3_SKEBS_PARAM
U700
T700
U10
T2

Outline

- Model-error representations increase the reliability of ensemble systems and improve forecast skill.
- Is this simple the result of increased reliability and decreased bias does the benefit of model-error schemes go beyond that?
  - Forecast are calibrated to have the same spread
  - Ensemble Forecasts are debiased with monthly mean bias


Brier Skill Score
Reliability
Resolution
CNTL
PARAM
SKEBS
PHYS10
PHYS10_SKEBS
PHYS3_SKEBS_PARAM
Raw
Calibrated
Debiased
20
10
0
U700
T700
U10
T2



Brier Skill Score
Reliability
Resolution
Raw
Calibrated
Debiased
Calibrated & Debiased
CNTL
PARAM
SKEBS
PHYS10
PHYS10_SKEBS
PHYS3_SKEBS_PARAM
20
10
0
U700
T700
U10
T2

(Model-specific) Conclusions

- Model-error representation increase forecast skill
- Including a model-error representation remains beneficial even if the ensemble systems are calibrated and/or debiased. This suggests that the merits of model-error representations go beyond increasing spread and removing the mean error and can account for certain aspects of structural model uncertainty.

Acknowledgements

Berner, J, K. Fossell, S.-Y. Ha, J. P. Hacker, C. Snyder 2015: "Increasing the skill of probailistics forecasts: Understanding performance improvements from model-error representations, *Mon. Wea. Rev.*, **143**, 1295–1320

Multi-Physics combinations

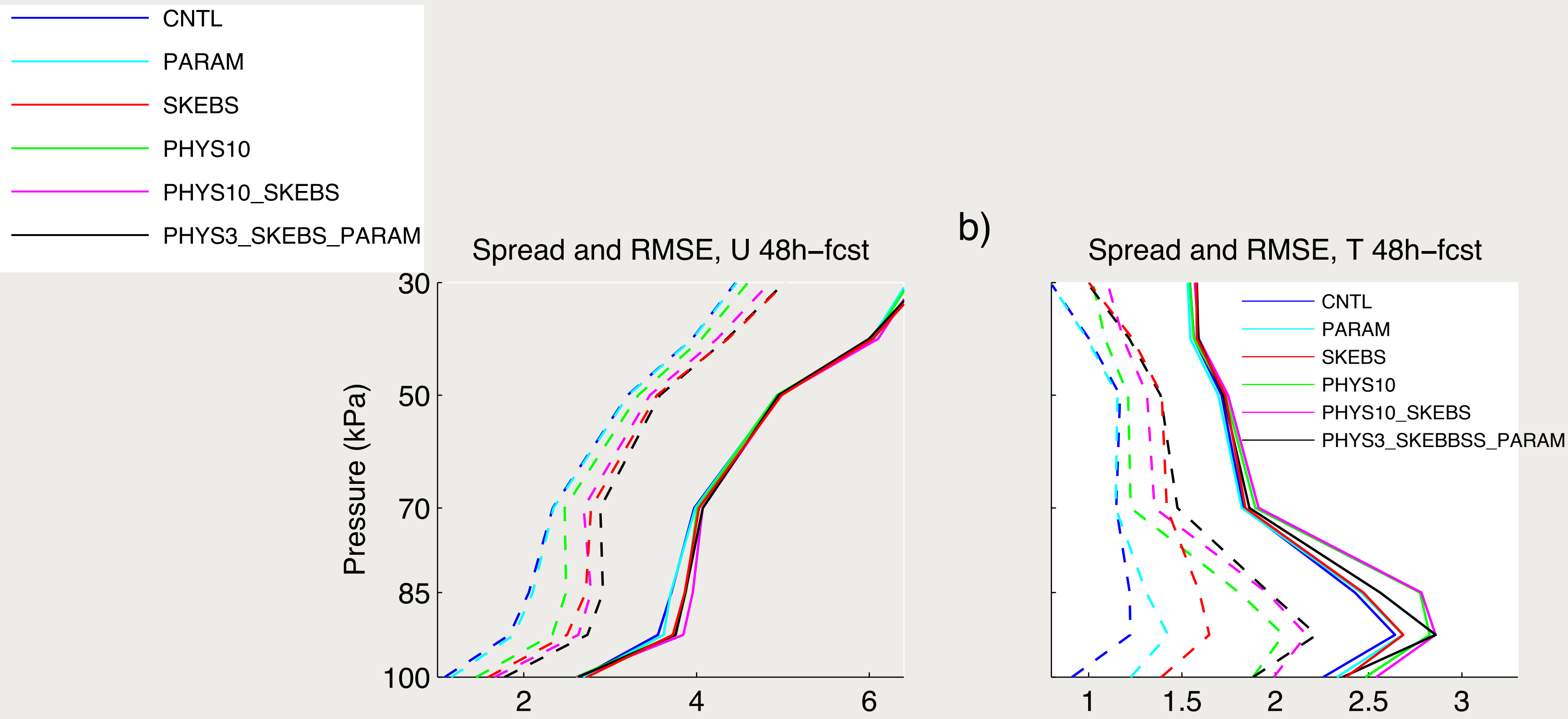
Member	Land Surface	Microphysics	PBL	Cumulus	Longwave	Shortwave
1	Thermal	Kessler	YSU	KF	RRTM	Dudhia
2	Thermal	WSM6	MYJ	KF	RRTM	CAM
3	Noah	Kessler	MYJ	BM	CAM	Dudhia
4	Noah	Lin	MYJ	Grell	CAM	CAM
5	Noah	WSM6	YSU	KF	RRTM	Dudhia
6	Noah	WSM6	MYJ	Grell	RRTM	Dudhia
7	RUC	Lin	YSU	BM	CAM	Dudhia
8	RUC	Eta	MYJ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

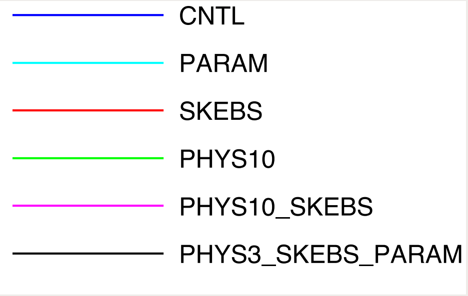
Decomposition of brier skill score

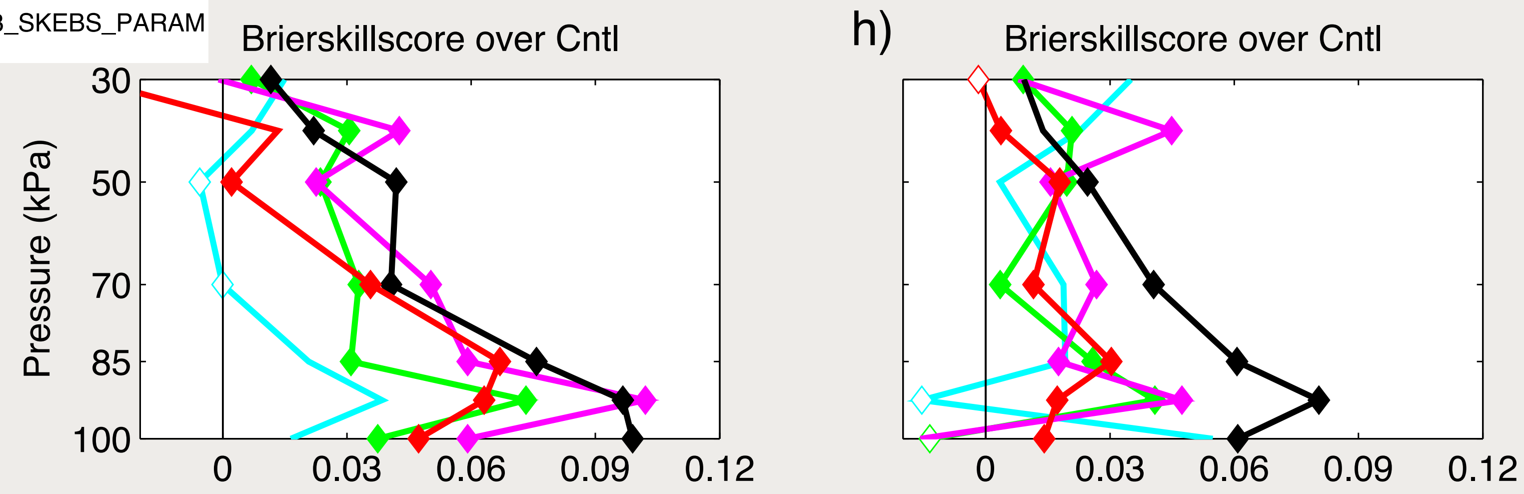
$$
\begin{aligned}
\mathrm{BSS}_{\mathrm{exp}} &= \frac{\mathrm{BS}_{\mathrm{CNTL,raw}} - \mathrm{BS}_{\mathrm{exp}}}{\mathrm{BS}_{\mathrm{CNTL,raw}}} \\
&= \frac{\mathrm{Rel}_{\mathrm{CNTL,raw}} - \mathrm{Rel}_{\mathrm{exp}}}{\mathrm{BS}_{\mathrm{CNTL,raw}}} + \frac{\mathrm{Res}_{\mathrm{CNTL,raw}}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

spread and error profiles @ 48h



Brier skill score profiles@ 48h





Reliability and resolution@ 48h