Calibration project at ECMWF

Stephan Hemri

CST group, Heidelberg Institute for Theoretical Studies

12 February 2015

Heidelberg Institute for Theoretical Studies

Overview

- motivation for statistical post-processing
- post-processing design of the calibration project at ECMWF
- presentation of EMOS post-processing for T2M forecasts
- verification results for T2M and PPT24
- is there a temporal trend in the skill gap between raw ensemble and post-processed forecasts?

Why post-processing of ECMWF forecasts?

Global ensemble forecasting systems:

- are prone to probabilistic biases
- tend to underdispersion for surface variables

Post-processing methods:

- aim to remove probabilistic biases
- should maximize sharpness subject to calibration
- ensemble model output statistics (EMOS, Gneiting et al. (2005))
- Bayesian model averaging (BMA, Raftery et al. (2005))

Post-processing design

- ECMWF 12 UTC forecasts (1 HRES, 50 ENS, 1 CTRL)
- forecast lead times $1, \ldots, 10$ days
- variables: **T2M**, **PPT24**, V10, and TCC
- station-wise post-processing on global domain
- observation and forecast data from January 2002 to March 2014
- sliding window training periods
- used mainly EMOS because of its low computational cost

EMOS for T2M

Let $\boldsymbol{f} = (f_1, f_2, \ldots, f_K)^T$ be the vector of the K member raw ensemble forecasts (here: HRES, the mean of ENS, and CTRL):

- then the EMOS predictive density is

  $$y|\boldsymbol{f} \sim g(m, \sigma),$$

  where $g(\cdot)$ is a parametric density function with location and scale parameters m and σ.
- m and σ are functions of ensemble statistics.

EMOS for T2M

- For a training period $T = \{t_1, \ldots, t_n\}$ we fit

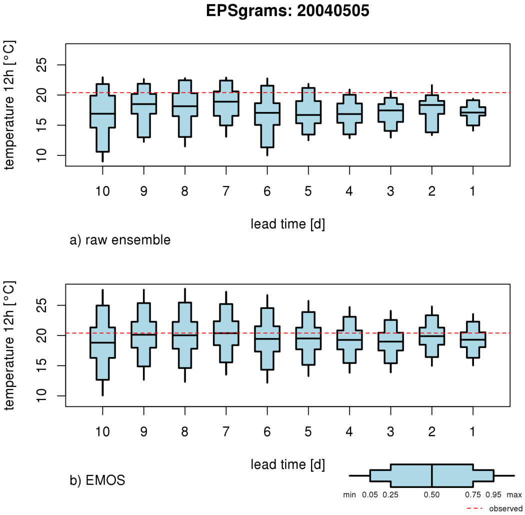
$$y_{t_j} = c_0 + c_1 \sin\left(\frac{2\pi j}{365}\right) + c_2 \cos\left(\frac{2\pi j}{365}\right) + \varepsilon_{t_j}, \quad j = 1, \ldots, n$$

- with fitted value $\tilde{y}_t$ as the climatological mean on day t
- $\tilde{f}_{\overline{\mathrm{ENS}},t}$, $\tilde{f}_{\mathrm{CTRL},t}$, and $\tilde{f}_{\mathrm{HRES},t}$ "climatological" forecast means
- With this:

$$m = \tilde{y} + a_1(f_{\mathrm{HRES}} - \tilde{f}_{\mathrm{HRES}}) + a_2(f_{\mathrm{CTRL}} - \tilde{f}_{\mathrm{CTRL}}) + a_3(f_{\overline{\mathrm{ENS}}} - \tilde{f}_{\overline{\mathrm{ENS}}})$$

$\sigma^2 = b_0 + b_1 s^2,$ where s^2 is the ensemble variance

Example: T2M forecasts for Vienna



Verification

Verification against observations:

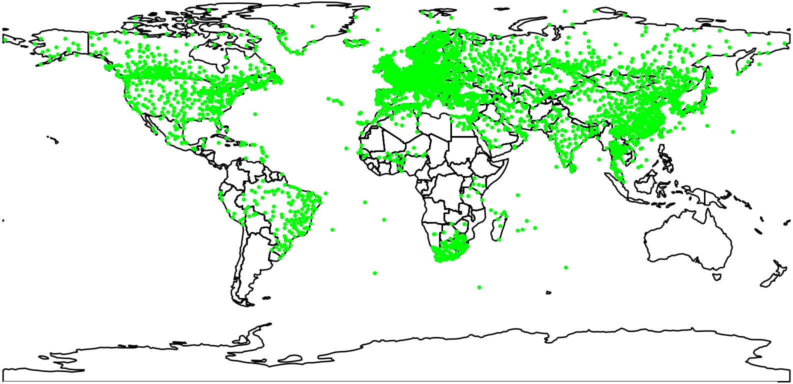
- continuous ranked probability score as main measure of skill:

$$\text{crps}(P, y) = \int_{-\infty}^{\infty} \left[P(x) - \mathbb{1}_{[x \geq y]} \right]^2 dx$$

- station-wise block bootstrapping to check for differences in the mean CRPS between raw ensemble and EMOS forecasts

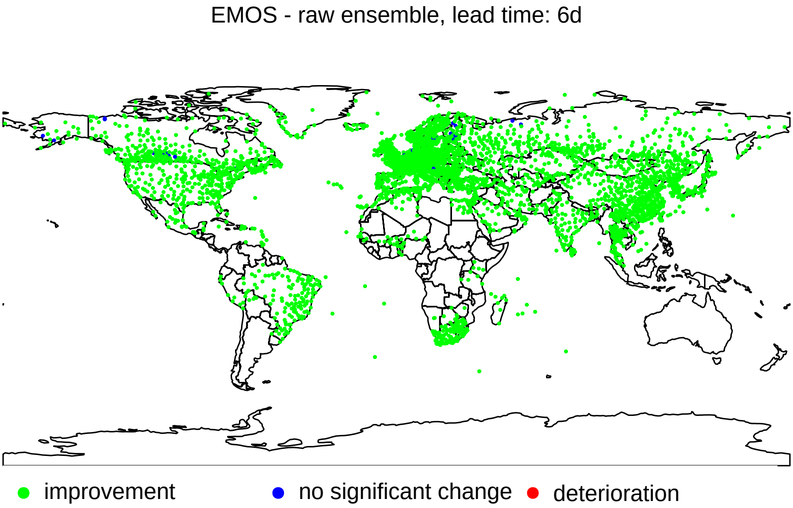
Significant changes in CRPS for T2M

EMOS - raw ensemble, lead time: 3d



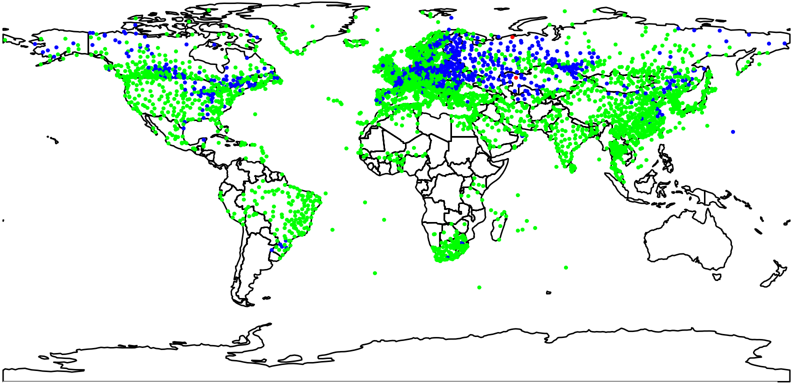
● improvement ● no significant change ● deterioration

Significant changes in CRPS for T2M



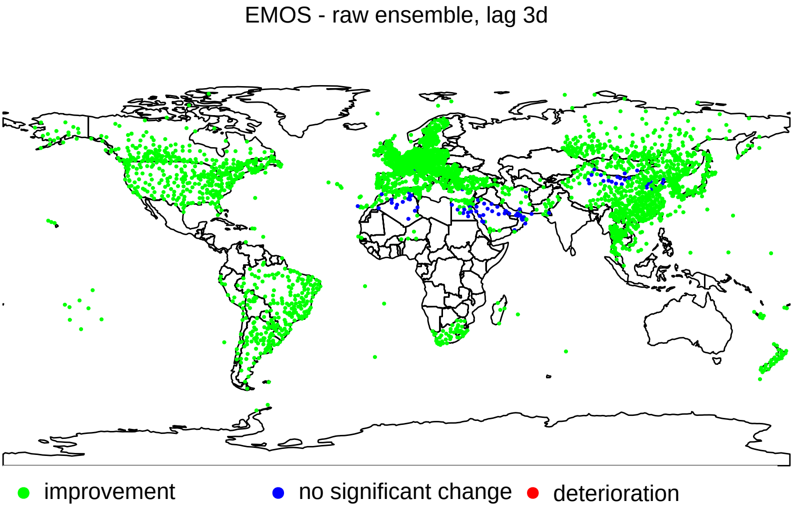
Significant changes in CRPS for T2M

EMOS - raw ensemble, lead time: 10d

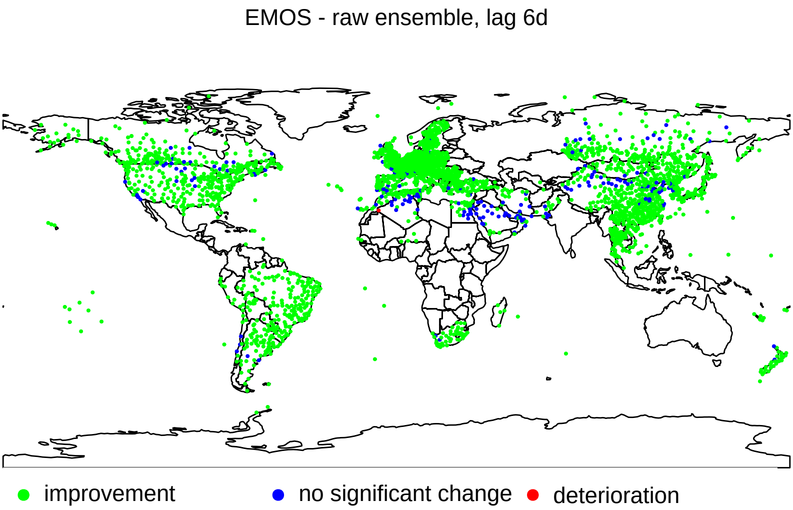


● improvement ● no significant change ● deterioration

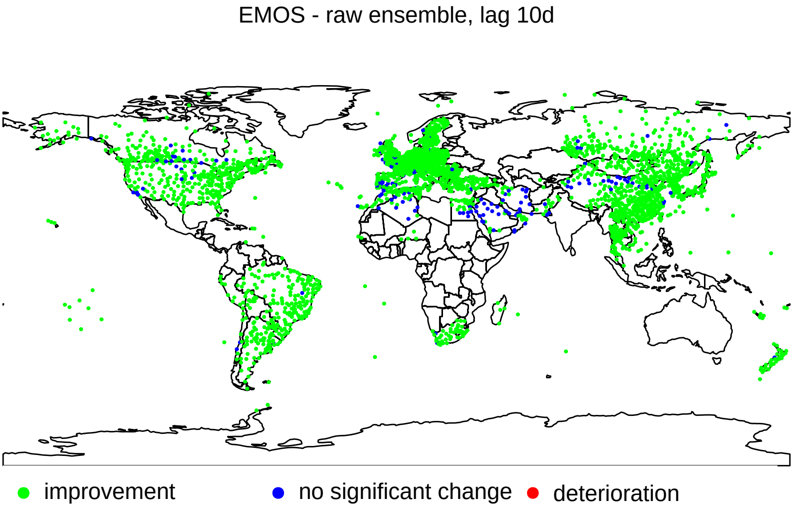
Significant changes in CRPS for PPT24



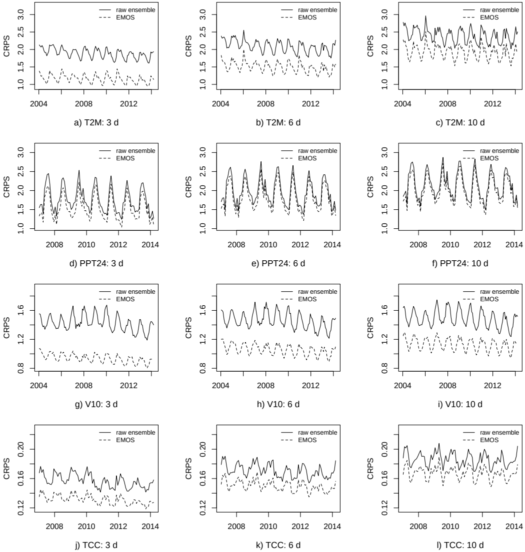
Significant changes in CRPS for PPT24



Significant changes in CRPS for PPT24



Global average CRPS



But: skill of raw ensemble improves over time...

- ECMWF ensemble is under continuous development
- Hypothesis:
  - these improvements also reduce systematic errors, which can also be reduced by post-processing
  - hence, the gap in skill between raw ensemble and post-processed forecasts narrows over time
  - and post-processing might lose its justification

Evaluation of the evolution of ΔCRPS

- Monthly time series of CRPS differences:

$$\Delta\text{CRPS}_t = \text{CRPS}_{\text{raw},t} - \text{CRPS}_{\text{EMOS},t}$$

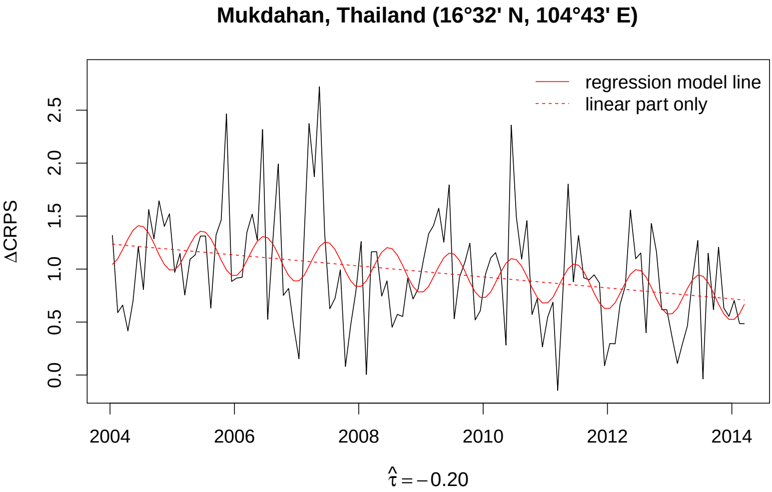
- Model 1 - parametric:

$$\Delta\text{CRPS}_t = \beta_0 + \beta_1 t + \beta_2 \sin\left(\frac{2\pi t}{12}\right) + \beta_3 \cos\left(\frac{2\pi t}{12}\right) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

- Model 2 - Kendall's τ rank correlation coefficient for pairs (t, r_t), where r_t are the residues of:

$$\Delta\text{CRPS}_t = \gamma_0 + \gamma_1 \sin\left(\frac{2\pi t}{12}\right) + \gamma_2 \cos\left(\frac{2\pi t}{12}\right) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

Example station



Percentages of stations with significant trend in ΔCRPS[a]

			parametric model		Kendall's τ statistics	
			T2M	PPT24	T2M	PPT24
forecast lead time	3 d	no significant trend	42 %	76 %	44 %	77 %
		negative trend	34 %	19 %	32 %	18 %
		positive trend	24 %	5 %	24 %	5 %
	6 d	no significant trend	46 %	82 %	48 %	82 %
		negative trend	31 %	14 %	29 %	13 %
		positive trend	23 %	4 %	23 %	5 %
	10 d	no significant trend	54 %	83 %	54 %	82 %
		negative trend	27 %	11 %	26 %	11 %
		positive trend	19 %	6 %	20 %	7 %

[a]Percentages of stations (totals are 4160 (T2M) and 2917 (PPT24)) showing no, negative, or positive trend in monthly ΔCRPS values against time at a significance level of 0.05

Conclusions

- univariate post-processing using EMOS improves skill at almost any station and lead time for the variables T2M, PPT24, V10, and TCC
- probabilistic skill of both raw ensemble and EMOS forecasts improves over time
- gap in skill (ΔCRPS) remains almost constant over time
    - improvements to the atmospheric model increase potential skill
    - statistical post-processing will keep adding skill in the foreseeable future

References

[1] Gneiting, T., and A. E. Raftery, 2007: Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc. 102*, 359–378.

[2] Gneiting, T., A. E. Raftery, A. H. Westveld, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Weather Rev. 133* (5), 1098-1118.

[3] Hemri, S., M. Scheuerer, F. Pappenberger, K. Bogner, and T. Haiden, 2014: Trends in the predictive performance of raw ensemble weather forecasts. *Geophys. Res. Lett. 41*, 9197-9205.

[4] Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev. 133* (2), 1155-1174.

[5] M. Scheuerer, 2014: Probabilistic quantitative precipitation forecasting using ensemble model output statistics. *Q. J. R. Meteorol. Soc. 140*(680), 1086-1096.

[6] Scheuerer M., and L. Büermann, 2014: Probabilistic quantitative precipitation forecasting using ensemble model output statistics. *J. R. Stat. Soc. Ser. C 63*(3), 405-422.

[7] Thorarinsdottir, T. L., and T. Gneiting, 2010: Probabilistic forecasts of wind speed: Ensemble model output statistics using heteroscedastic censored regression. *J. R. Stat. Soc. Ser. A 173*, 371-388.