Understanding the link between ensemble mean error variance, spread-error ratio, mean error and the CRPS

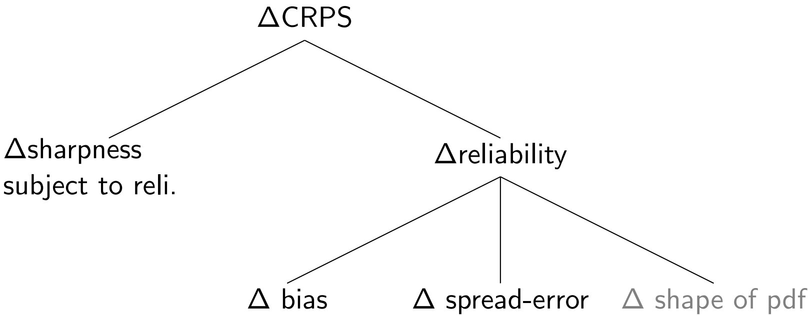
Martin Leutbecher and Thomas Haiden

2020-IVMW-O

Acknowledgements: Zied Ben Bouallègue, Jochen Bröcker, Tilmann Gneiting, David Richardson and Maxime Taillardat

Motivation

- Proper scores are essential for NWP development involving ensemble forecasts [Gneiting and Raftery, 2007]
- The Continuous Ranked Probability Score (CRPS) is a widely used proper score
- Users as well as developers may wish to better understand what causes a change in the CRPS, which can have several reasons



A homogeneous Gaussian forecast-observation distribution

Let us consider a stochastic model of the forecast-observation distribution

$$f_k = \overline{f} + \sigma \zeta_k \qquad \text{with} \tag{1}$$

$$\overline{f} = \alpha y + \beta \eta + b \tag{2}$$

- with ensemble members f_k, ensemble mean $\overline{f}$, ensemble variance σ^2 and observation y
- the model is homogeneous and Gaussian: α, β, σ, b are constants and $\zeta_k, \eta \sim N(0,1); \quad y \sim N(0, \omega^2)$ are independent Gaussian random variables
- based on extensions of toy models described by Weigel and Bowler [2009] and Leutbecher and Ben Bouallègue [2020], see Leutbecher and Haiden [2020] for details

ECMWF

The expected CRPS for the stochastic model

It can be shown that the expected CRPS $\mathcal{C}$ satisfies

$$\mathcal{C} = \frac{\epsilon}{\sqrt{\pi}} \left[\sqrt{2 + 2\sigma_*^2} \exp\left(-\frac{b_*^2}{2 + 2\sigma_*^2} \right) + \sqrt{\pi}\, b_* \operatorname{erf}\left(\frac{b_*}{\sqrt{2 + 2\sigma_*^2}} \right) - \sigma_* \right]$$

where

ϵ^2	is the variance of the error of the ensemble mean
b_*	is the mean error of the ensemble mean (i.e. the bias) normalised by ϵ
σ_*	is the spread-error ratio, i.e. ensemble stdev normalised by ϵ
erf	is the error function, $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-t^2)\, dt$

The expected CRPS for the stochastic model

It can be shown that the expected CRPS $\mathcal{C}$ satisfies

$$\mathcal{C} = \frac{\epsilon}{\sqrt{\pi}} \left[\sqrt{2 + 2\sigma_*^2} \exp\left(-\frac{b_*^2}{2 + 2\sigma_*^2} \right) + \sqrt{\pi}\, b_* \operatorname{erf}\left(\frac{b_*}{\sqrt{2 + 2\sigma_*^2}} \right) - \sigma_* \right]$$

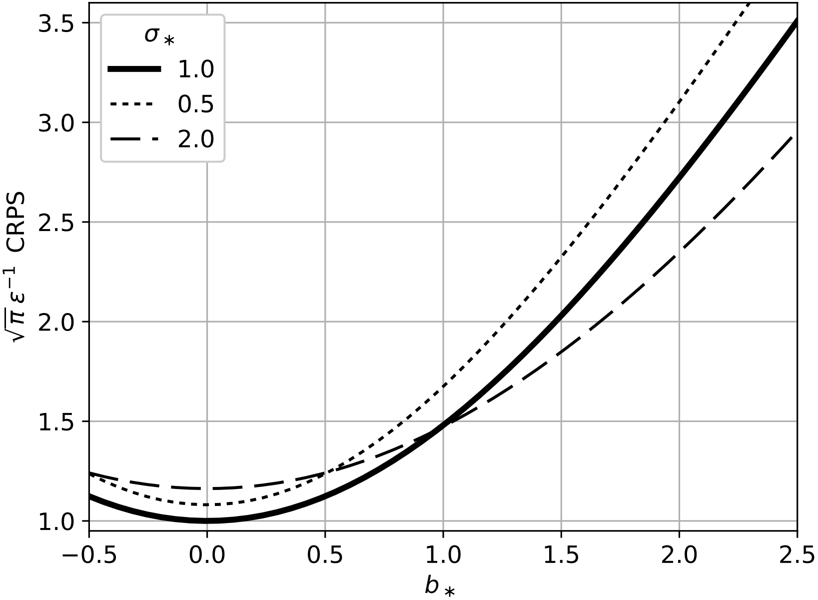
where

ϵ^2	is the variance of the error of the ensemble mean
b_*	is the mean error of the ensemble mean (i.e. the bias) normalised by ϵ
σ_*	is the spread-error ratio, i.e. ensemble stdev normalised by ϵ
erf	is the error function, $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-t^2)\, dt$

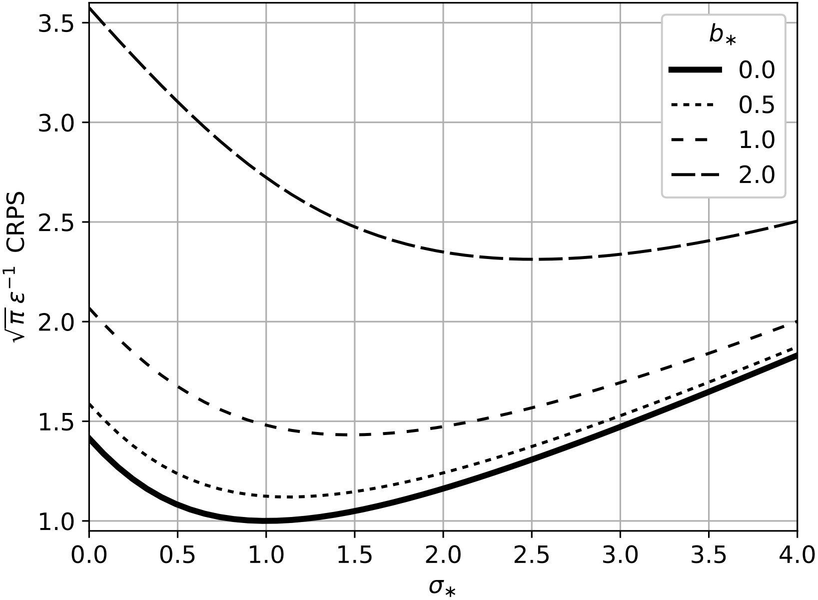
For a perfectly reliable ensemble, we have $b_* = 0$ and $\sigma_* = 1$. This implies $[\ldots] = 1$ and

$$\mathcal{C} = \frac{\epsilon}{\sqrt{\pi}}$$

Linking CRPS and normalized bias



Linking CRPS and spread-error ratio



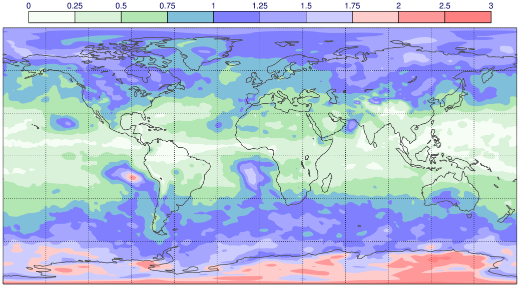
The homogeneous Gaussian model as an approximation for real NWP data

- Ensemble verification data can be analysed using the relationship derived for the homogeneous Gaussian model
- Doing so turns this relationship into an approximation of the exact CRPS.
- Is it useful even if real data are inhomogeneous and deviate from normality?

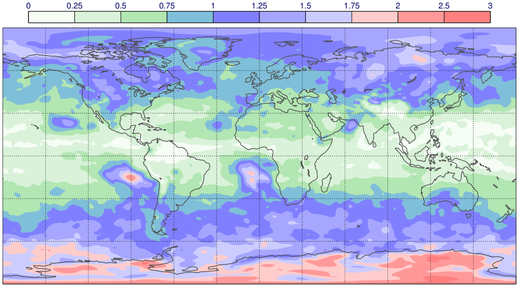
How good is the approximation?

5-day ENS forecast of 850 hPa temperature, JJA2019

CRPS (K)



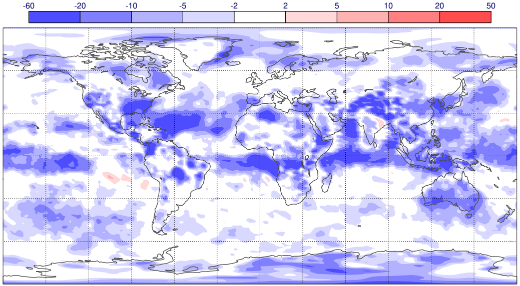
full Gaussian approximation $\mathcal{C}$ (K)



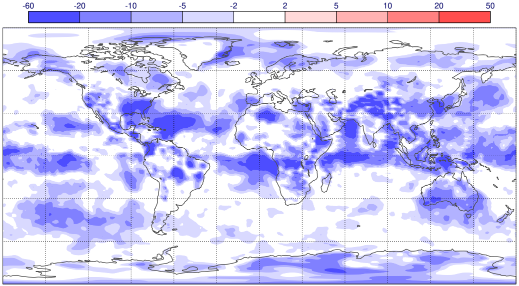
Impact of bias correction

5-day ensemble fc of T850, JJA2019

ΔCRPS (percent)



$\Delta\mathcal{C}$ (percent)

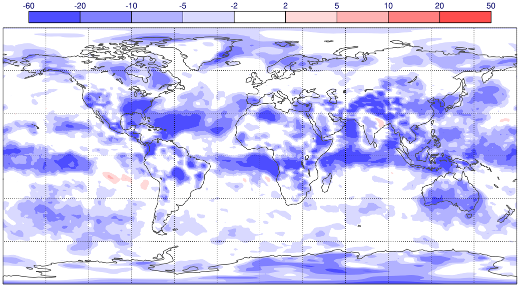


ECMWF

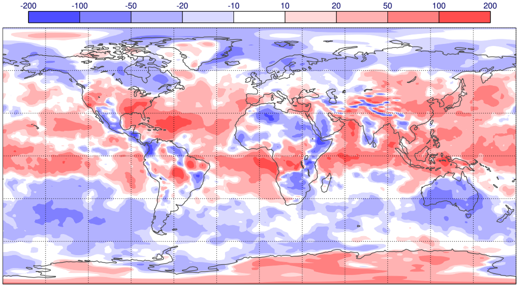
Change in CRPS versus normalised bias

5-day ensemble fc of T850, JJA2019

ΔCRPS (percent)



$100b_*$



Decompositions

Score changes can be decomposed exactly into three terms

$$
\begin{aligned}
\Delta\mathcal{C} &= \Delta_\epsilon + \Delta_{\sigma_*} + \Delta_{b_*} \quad \text{with} && \\
\Delta_\epsilon &= \mathcal{C}(\tilde{\epsilon}, \sigma_*, b_*) - \mathcal{C}(\epsilon, \sigma_*, b_*) && \text{due to ens. mean err. variance} \\
\Delta_{\sigma_*} &= \mathcal{C}(\tilde{\epsilon}, \tilde{\sigma}_*, b_*) - \mathcal{C}(\tilde{\epsilon}, \sigma_*, b_*) && \text{due to spread-error ratio} \\
\Delta_{b_*} &= \mathcal{C}(\tilde{\epsilon}, \tilde{\sigma}_*, \tilde{b}_*) - \mathcal{C}(\tilde{\epsilon}, \tilde{\sigma}_*, b_*) && \text{due to bias}
\end{aligned}
$$

Score changes can be decomposed exactly into three terms

$$
\begin{aligned}
\Delta\mathcal{C} &= \Delta_\epsilon + \Delta_{\sigma_*} + \Delta_{b_*} \quad \text{with} \\
\Delta_\epsilon &= \mathcal{C}(\tilde{\epsilon}, \sigma_*, b_*) - \mathcal{C}(\epsilon, \sigma_*, b_*) && \text{due to ens. mean err. variance} \\
\Delta_{\sigma_*} &= \mathcal{C}(\tilde{\epsilon}, \tilde{\sigma}_*, b_*) - \mathcal{C}(\tilde{\epsilon}, \sigma_*, b_*) && \text{due to spread-error ratio} \\
\Delta_{b_*} &= \mathcal{C}(\tilde{\epsilon}, \tilde{\sigma}_*, \tilde{b}_*) - \mathcal{C}(\tilde{\epsilon}, \tilde{\sigma}_*, b_*) && \text{due to bias}
\end{aligned}
$$

An elegant decomposition of the CRPS into reliability, resolution and uncertainty arises following Siegert [2017]

$$
\begin{aligned}
\mathrm{CRPS} &= \mathrm{REL} - \mathrm{RES} + \mathrm{UNC} \\
\mathrm{REL} &= \mathcal{C} - \epsilon/\sqrt{\pi} \\
\mathrm{RES} &= (\omega - \epsilon)/\sqrt{\pi} \\
\mathrm{UNC} &= \omega/\sqrt{\pi}.
\end{aligned}
$$

Summary

- Analytical expression for expected CRPS given a homogeneous Gaussian forecast-observation distribution
- Works well as approximation of the sample mean CRPS for medium-range weather forecasts of upper air variables
- Score changes can be understood in terms of changes in the variance of the ensemble mean, the spread-error ratio and the bias
- This diagnostic can be added fairly easily to existing verification without the need for postprocessing
- The sensitivity of the CRPS to day-to-day variations in ensemble spread (heteroscedasticity) appears rather weak.
- Optimising ensemble configurations based on the CRPS of direct model output leads to over-dispersion in the presence of bias

References

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction and estimation. *Journal of the American Statistical Association*, 102:359–378, 2007. doi: 10.1198/016214506000001437.

Martin Leutbecher and Zied Ben Bouallègue. On the probabilistic skill of dual-resolution ensemble forecasts. *Q. J. R. Meteor. Soc.*, 146:707–723, 2020. doi: 10.1002/qj.3704.

Martin Leutbecher and Thomas Haiden. Understanding changes of the continuous ranked probability score using a homogeneous Gaussian approximation. *Q. J. R. Meteor. Soc.*, 2020. doi: 10.1002/qj.3926. URL `http://doi.org/10.1002/qj.3926`.

Stefan Siegert. Simplifying and generalising Murphy's Brier score decomposition. *Q. J. R. Meteor. Soc.*, 143:1178–1183, 2017. doi: 10.1002/qj.2985.

A.P. Weigel and N.E. Bowler. Comment on 'Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts?'. *Q. J. R. Meteor. Soc.*, 135(639):535–539, 2009. doi: 10.1002/qj.381.