Regime-dependent statistical post-processing of ensemble forecasts

Sam Allen (sa495@exeter.ac.uk), Chris Ferro, Frank Kwasniok

Weather regimes

- Statistical post-processing uses the behaviour of past forecasts to predict the errors that will be present in the current forecast.
- Previous studies have found that model biases are affected by the prevailing atmospheric circulation.
- Forecasters therefore often manually adjust their forecast depending on the prevailing flow.
- The synoptic-scale flow can be explained by a small number of persistent and recurrent weather patterns, called regimes.
- Incorporating the circulation directly into post-processing can account for regime-dependent model errors.

Motivation

- Atmospheric predictability and model biases depend on the prevailing weather regime.
- If the seasonal cycle of model errors can be attributed to the occurrence of certain regimes then regime-analogues may be a more sensible choice of training data than only recent forecasts.
- Regimes can explain relationships between different variables and spatial locations so are sensible for use in multivariate approaches.
- Certain high impact weather events can be attributed to the occurrence of specific weather patterns.
    - Using regime-dependent forecasts could improve predictions of extreme weather.

Regime-dependent statistical post-processing

- Let $\boldsymbol{f} = (f_1, \dots, f_M)$ denote an ensemble forecast for a weather variable y, with M exchangeable members.
- Assume R regimes have been identified. Let M_r be the number of ensemble members predicting regime r , for $r = 1, \dots, R$.
- Let $w(r)$ represent a probability that the atmosphere resides in regime r at the forecast validation time.
- One possible choice of $w(r)$ is the proportion of forecasts predicting regime r at the validation time, M_r/M.

Bayesian Model Averaging (BMA)

- BMA uses a mixture of forecast distributions specified for each ensemble member:

$$y|\boldsymbol{f} \sim \frac{1}{M} \sum_{m=1}^{M} g(y|f_m, \boldsymbol{\theta})$$

- Ensemble members could be dressed using different parametric families, g, or model parameters, $\boldsymbol{\theta}$, depending on the regime that they predict:

$$y|\boldsymbol{f} \sim \sum_{r=1}^{R} w(r) \sum_{m=1}^{M_r} g_r(y|f_m, \boldsymbol{\theta}_r)$$

Ensemble Model Output Statistics (EMOS)

- EMOS uses a heteroscedastic predictive distribution that depends on the ensemble $\boldsymbol{f}$, typically through its mean and variance:

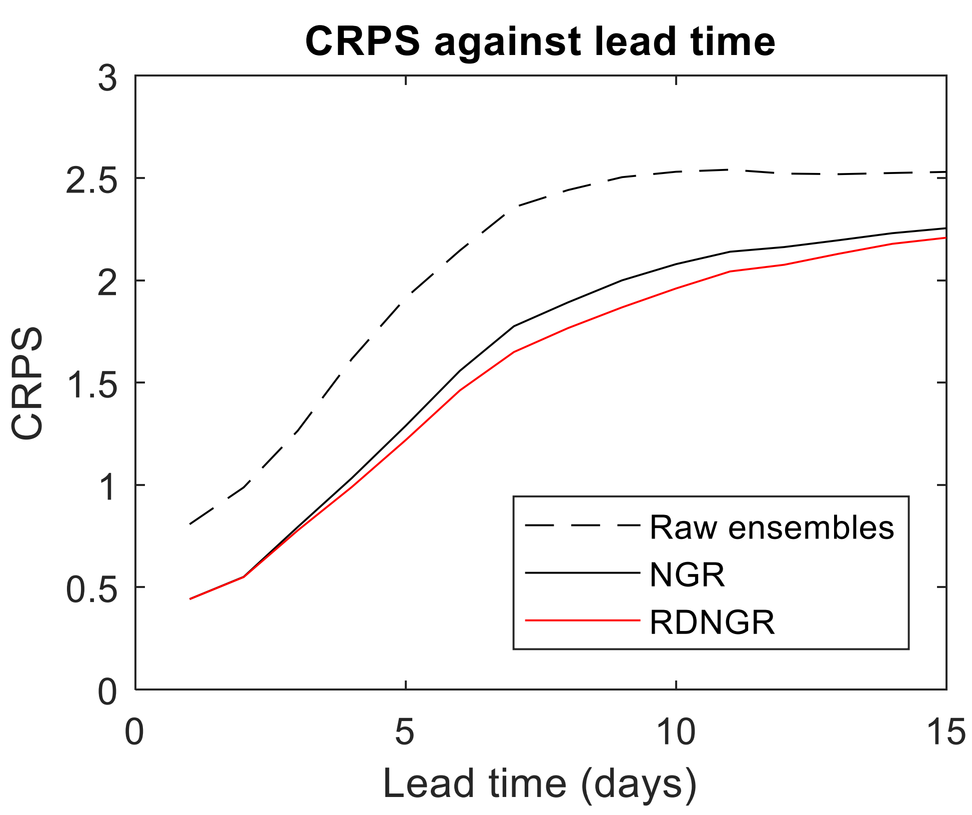
$$y|\boldsymbol{f} \sim g(y|\boldsymbol{f}, \boldsymbol{\theta})$$

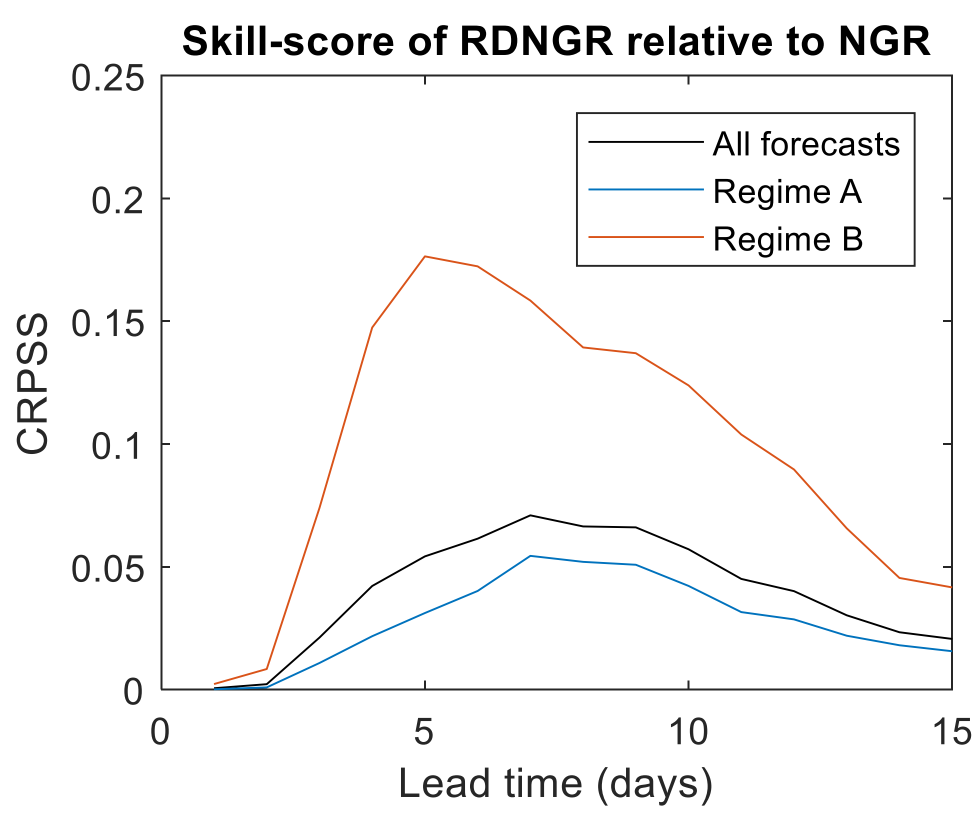
- Regime-specific forecast distributions, g_r, and parameters, $\boldsymbol{\theta}_r$, could be estimated for each regime:

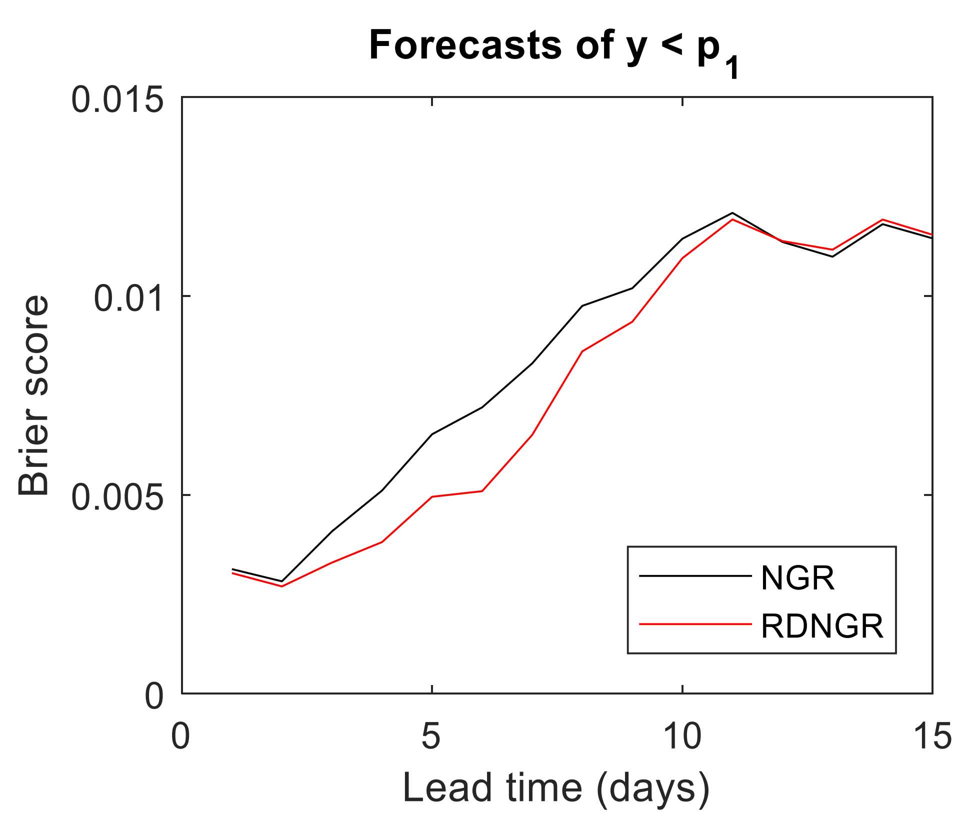
$$y|\boldsymbol{f} \sim \sum_{r=1}^{R} w(r)\, g_r(y|\boldsymbol{f}, \boldsymbol{\theta}_r)$$

Simulation study – Lorenz '96 system

- The Lorenz (1996) system exhibits regime-like behaviour, alternating between two regimes (regimes A and B) (Christensen et al., 2015).
- We implement the above BMA and NGR approaches in this system, with $R = 2$ and $w(r) = M_r/M$.
- We predict the energy (E) in the system, using Normal distributions for all g and g_r, but with regime-specific model parameters.
- Forecasts distributions are estimated and assessed using the continuous ranked-probability score (CRPS) and its skill-score (CRPSS).
- Results are shown for NGR and Regime-dependent NGR (RDNGR) but the same results are seen for BMA and RDBMA







- RDNGR can improve upon conventional NGR by more than 5% for lead times between 5 and 10 days, and up to 17% in one of the regimes.
- Forecasts of extremely low values of E are significantly more accurate too (p_1 is the first percentile of the unconditional distribution of E) .

- Allen S, Ferro CAT, Kwasniok F. 2019: Regime-dependent statistical post-processing of ensemble forecasts.
- Christensen HM, Moroz IM, Pamer TN. 2015: Simulating weather regimes: impact of stochastic and perturbed parameter schemes in a simple atmospheric model
- Lorenz EN. 1996: Predictability, a problem partly solved.