GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Model selection for DeFoReSt: a strategy for recalibrating decadal predictions

Max-Planck-Institut für Meteorologie
Freie Universität Berlin

A. Pasternack*, H. W. Rust*, J. Grieger*, M. A. Liniger**, J. Bhend**, W. Müller***, U. Ulbrich*

* Freie Universität Berlin, Institute for Meteorology
** MeteoSwiss
*** Max-Planck-Institute for Meteorology, Hamburg

Abstract

One issue frequently observed for decadal probabilistic forecasts is that they tend to be not reliable and thus need to be re-calibrated. Methods for seasonal time scales have to be adapted for decadal time scales, e.g. climate trend and lead time dependent bias.
With DeFoReSt[3], we proposed a Decadal Climate Forecast Recalibration Strategy to tackle these problems. The original approach of DeFoReSt assumes 3rd and 2nd order polynomials to capture lead year dependent errors and 1st order for start time dependency. Here, we propose not to restrict orders a priori but use a systematic model selection strategy based on non-homogeneous boosting to identify the relevant predictors for recalibrating the MiKlip decadal prediction system.

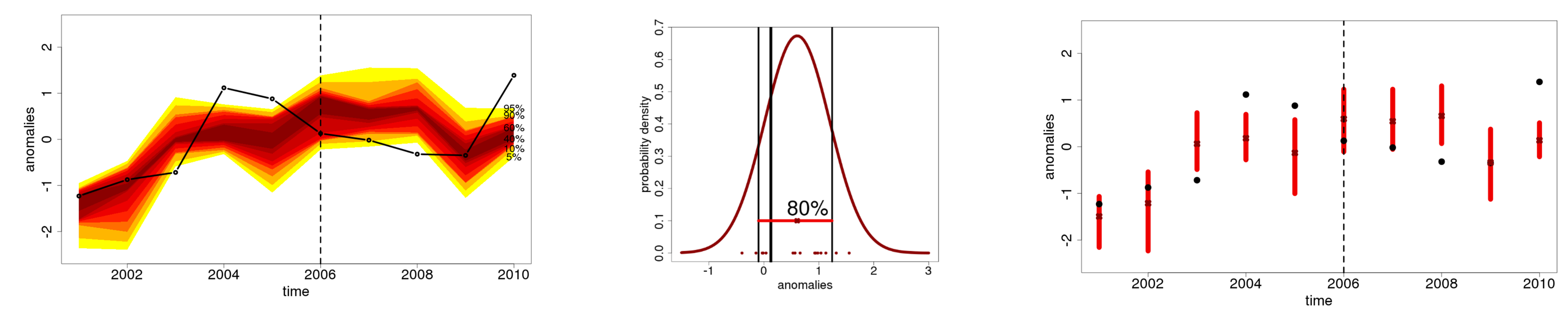
1. Introduction

What is a good Probabilistic forecast?

„... an important goal is to maximize sharpness without sacrificing calibration." [2,6]

Calibration or reliability:
„Ensemble members are reliable if the MSE between the ensemble mean and observations is identical to the time mean intra-ensemble variance."

Sharpness:
Forecasts take a risk, i.e. are frequently different from the climatological value?

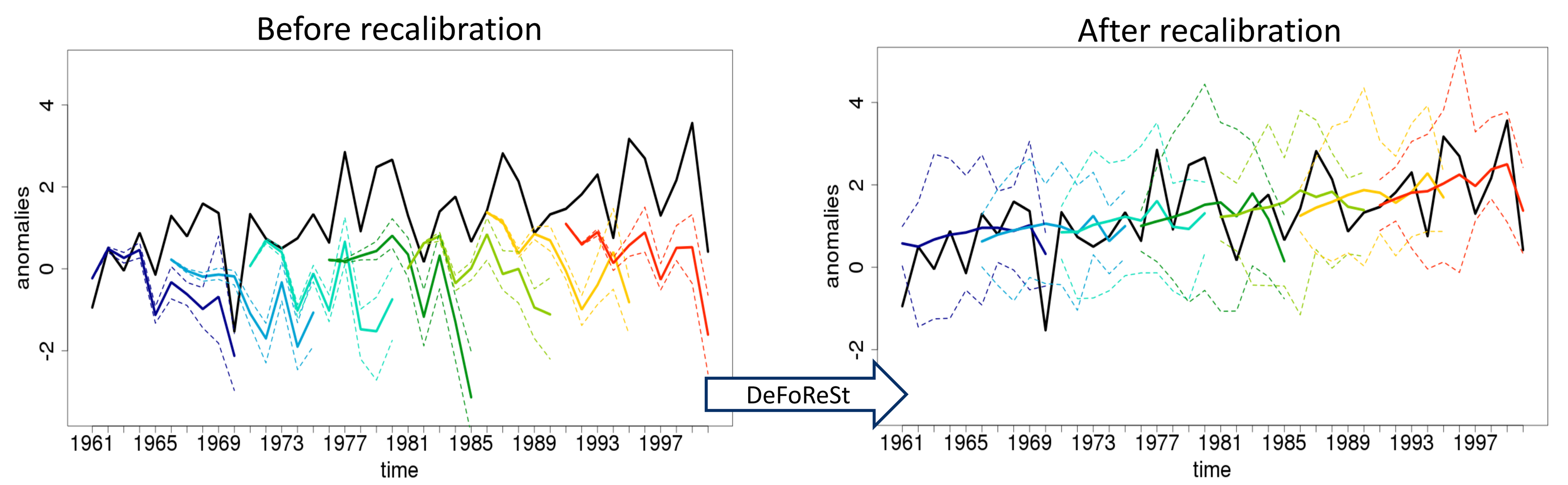


Problems of probabilistic decadal forecasts:

„... ensemble distributions typically underestimate the true forecast uncertainty and tend to be overconfident ..." [5] → Adjust ensemble spread

Characteristic problems of decadal forecasts:

- limited number of hindcasts
- dependence on lead years (drift)
- different climate trends
- Ens. spread is dependent from lead and start year



Exemplary drift of a decadal toy model (colored lines) with associated pseudo-observations (black line) before (left) and after recalibration with DeFoReSt (right). The dotted lines represent the ensemble minimum/maximum.

2. A model selection method for DeFoReSt

Decadal forecast recalibration strategy (DeFoReSt) [4]**:**

- This approach accounts for a lead and start year dependent unconditional bias (drift), conditional bias and conditional ens. dispersion.
- Assumption: the forecast distribution is Gaussian, thus

$$f^{Cal} \sim \mathcal{N}\left(\alpha(t,\tau) + \beta(t,\tau)\mu(t,\tau), \left(\delta(t,\tau)\sigma(t,\tau)\right)^2\right)$$

$$\alpha(t,\tau) = \sum_{k=0}^{3}(a_{2k} + a_{2k+1}t)\tau^k \qquad \delta(t,\tau) = \sum_{k=0}^{2}(d_{2k} + d_{2k+1}t)\tau^k$$

$$\beta(t,\tau) = \sum_{k=0}^{3}(b_{2k} + b_{2k+1}t)\tau^k$$

f^{Cal}: Calibrated Forecast t: Start year $\sigma^2(t,\tau)$: Ens. variance
$\mu(t,\tau)$: Ensemble mean τ: Lead year

➢ Find an a, b, d that minimize the negative log-Likelihood L:

$$L = -\frac{1}{N}\sum_{j=1}^{N} log\left(\frac{1}{\delta\sigma_j}\varphi\left(\frac{O_j - \alpha + \beta\mu_j}{\delta\sigma_j}\right)\right)$$

(For a Gaussian assumption & mean over N time steps)

O_j: Observation at time step j φ: PDF of std. norm. distribution

- Orders of α, β and δ are fixed a priori!
- Is there a better approach?

Model selection with Non-homogenous boosting[3]**:**

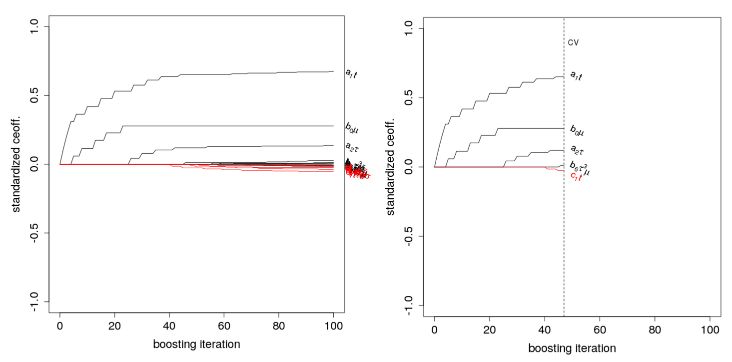
- Number of predictors is increased in order to catch all features:

$$f^{MS} \sim \mathcal{N}\left(\alpha(t,\tau) + \beta(t,\tau)\mu(t,\tau), \left(\gamma(t,\tau) + \delta(t,\tau)\sigma(t,\tau)\right)^2\right)$$

$$\alpha(t,\tau) = \sum_{k=0}^{6}(a_{2k} + a_{2k+1}t)\tau^k \qquad \gamma(t,\tau) = \sum_{k=0}^{6}(c_{2k} + c_{2k+1}t)\tau^k$$

$$\beta(t,\tau) = \sum_{k=0}^{6}(b_{2k} + b_{2k+1}t)\tau^k \qquad \delta(t,\tau) = \sum_{k=0}^{6}(d_{2k} + d_{2k+1}t)\tau^k$$

- Boosting iteratively increases model coefficients.
- Most relevant parameters are increased first.
- Best set of coeff. can be found by cross-validation (CV).
- Thus, not relevant parameters are zero.

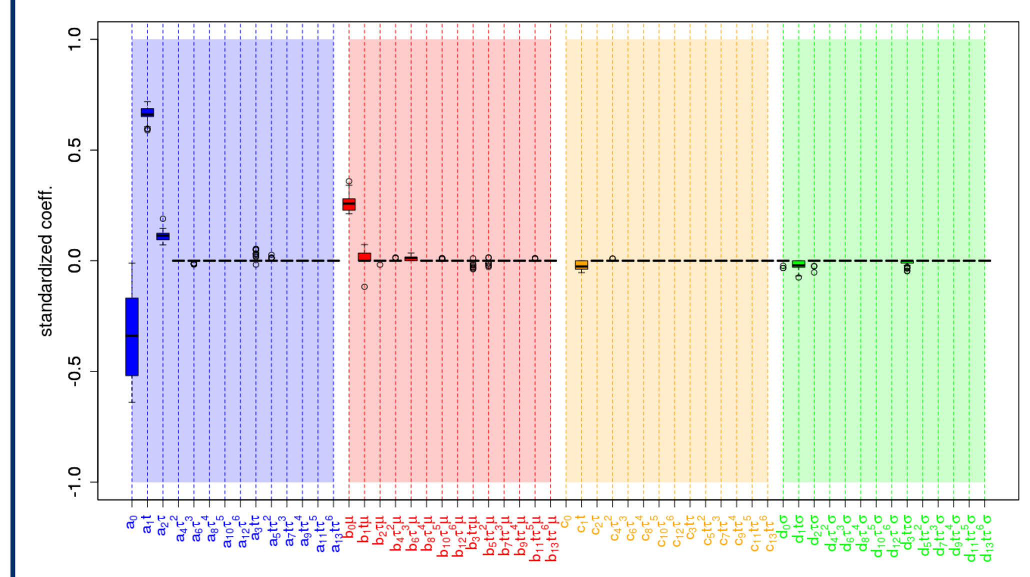


3. Apply model selection to decadal forecasts

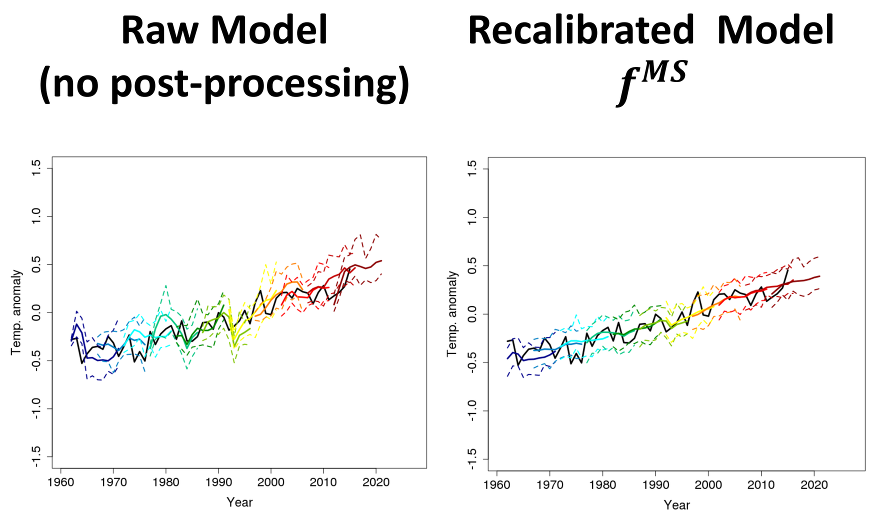
Data:

- Surface temperature
- MPI-ESM-LR, preop
- 10 ensemble member
- Start years: 1961-2015
- Annual mean
- Observation: HadCRUT4
- Reference: Historical runs

Identified coefficients after boosting. (Global mean)



Identified coefficients for uncond. bias (blue bars), cond. bias (red bars), uncond. dispersion (orange bars) and cond. dispersion (green bars) for preop. The coefficients are standardized, i.e. higher values implying a higher relevance.



Top: Global mean preop-Lr before (left) and after (right) recalibration with model selection (colored lines). The black line represents HadCRUT4.

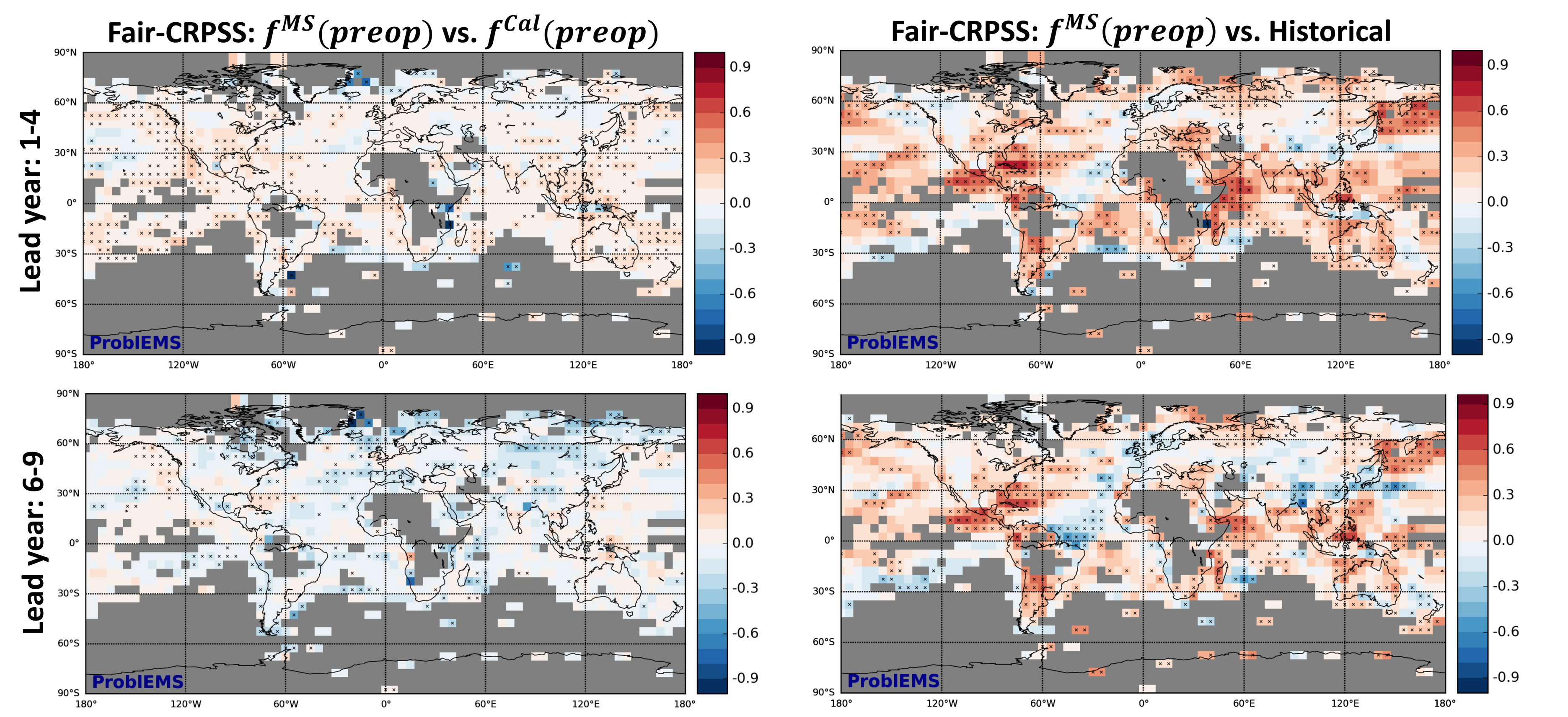
Validation of recalibration with model selection:

- Continuous ranked probability skill score (CRPSS) is used for validation:

$$CRPSS = 1 - \frac{CRPS_{Fc.}}{CRPS_{Ref.}}$$

$CRPSS > 0$: Forecast is more skillful than ref.
$CRPSS < 0$: Ref. is superior to forecast.

- Fair-CRPSS is independent from different ensemble sizes[1]
- All Scores were calculated in a cross-validation setting.



Fair-CRPSS of Preop with recalibration with model selection (f^{MS}) w.r.t. to recalibration without model selction (f^{Cal}) (left panels) and Historical runs (right panels) for lead years 1-4 (top) and 6-9 (bottom). The dotted grid points indicate significant changes w.r.t. to Historical runs. The significance was calculated with a 500-wise bootstrapping approach.

4. Summary & conclusions

- Model selection with boosting shows a major influence of linear start year dependency.
- Recalibration with model selection (f^{MS}) shows only a minor skill improvement w.r.t. DeFoReSt (f^{Cal}) but is more robust.
- Model selected recalibration is superior to Historical runs for lead years 1-4 and 6-9.

➢ **Outlook:** Also use a polynomial approach for start years.

FONA Mittelfristige Klimaprognosen BMBF
MiKlip

References / Acknowledgment

1. Ferro, C.A.T., 2013: Fair scores for ensemble forecasts. Q.J.R. Meteorol. Soc. DOI: 10.1002/qj.2270.
2. Gneiting, T. et al., 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Mon.Weath. Rev., 133, 1098–1118.
3. Messner et al., 2017: Nonhomogeneous boosting for predictor selection in ensemble postprocessing. Monthly Weather Review, 145(1):137–147.
4. Pasternack, A. et al. 2018. Parametric decadal climate forecast recalibration (DeFoReSt 1.0). , Geosci. Model Dev., 11, 351-368.
5. Weigel, A.P. et al., 2009: Seasonal ensemble forecasts: Are recalibrated single models better than multimodels? Mon. Weather Rev., 137(4):1460–1479.
6. Wilks, D. S., 2011: Statistical Methods in the Atmospheric Sciences. Academic Press.

A. Pasternack, alexander.pasternack@met.fu-berlin.de