GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Freie Universität Berlin

Calibrating probabilistic decadal predictions for categories

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

* Freie Universität Berlin, Institute for Meteorology
** MeteoSwiss

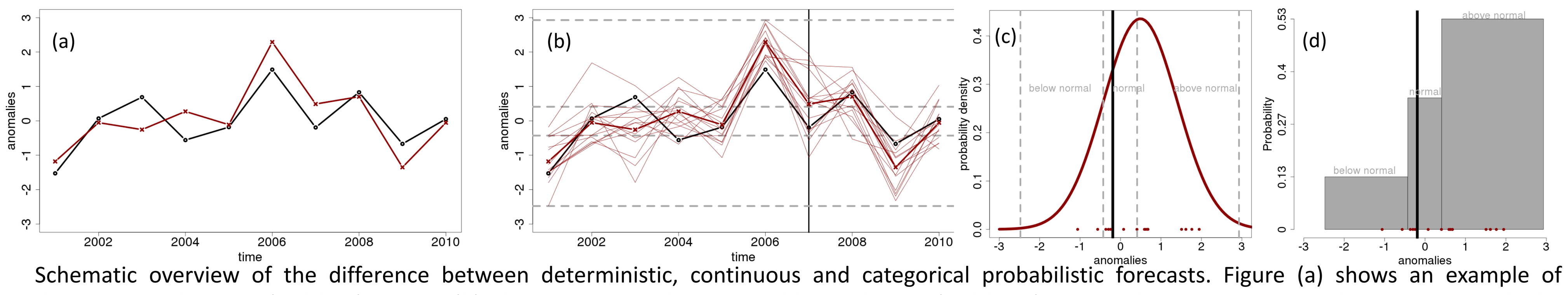
Abstract

Decadal climate predictions are of great socio-economic interest due to the corresponding planning horizons of several political and economic decisions. Due to uncertainties of weather and climate, forecasts (e.g. due to initial condition uncertainty), they should be and are increasingly issued in a probabilistic way. Similar to seasonal forecasts, also the decadal predictions are often issued in a categorical way, e.g. tercile (above normal, normal, below normal). Analogously to their continuous counterpart, also categorical probabilistic predictions need to be calibrated to ensure their usefulness.
While re-calibration methods for seasonal time scales are available and frequently applied, these methods still have to be adapted for decadal time scales and its characteristic problems like climate trend and lead time dependent biases. We propose a method to re-calibrate probabilistic categorical predictions that takes the above mentioned characteristics into account and apply this method to decadal predictions from the MiKlip (Germany's initiative for decadal prediction) system.

1. Introduction

Probabilistic forecast?

- Given the uncertainties due to, e.g. initial conditions, weather and climate forecasts should be and are increasingly issued in a probabilistic way.



Schematic overview of the difference between deterministic, continuous and categorical probabilistic forecasts. Figure (a) shows an example of deterministic forecast (red line). Figure (b) shows an exemplary ensemble prediction (red lines). For each year the ensemble members could be represented as PDF (e.g. 2007, figure (c)) or divided into categories (e.g. terciles, figure (d)).

What is a good Probabilistic forecast?

„... an important goal is to maximize sharpness without sacrificing calibration." [2,4]

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

Calibration or reliability:
Probabilistic forecasts „mean what they say", e.g. for days with a forecast of 30% chance of rain, we expect a relative frequency of 30% rainy days.

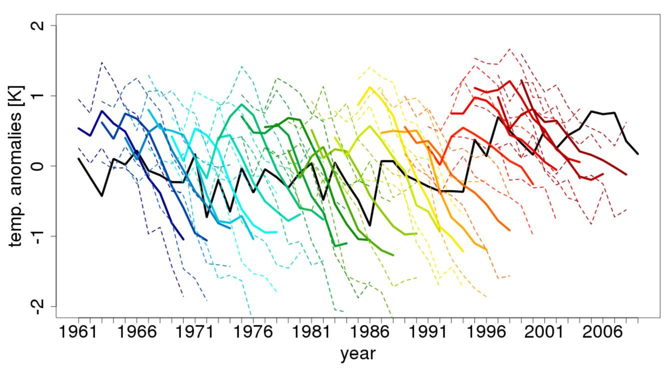
Problems of our probabilistic decadal forecasts:

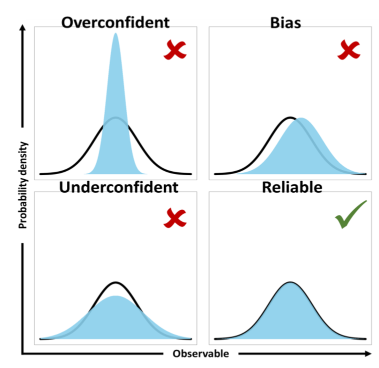
„... ensemble distributions typically underestimate the true forecast uncertainty and tend to be overconfident ..." [3]

Adjust ensemble spread

Characteristic problems of decadal forecasts:

- limited number of hindcasts
- different climate trends
- dependence on lead years (drift)





2. Method

Fit multinomial distribution with VGLM[5]:

- In contrast to GLMS, vector generalized linear models (VGLMs) allow for response variables outside the classical exponential family and for more than one parameter.
- VGLMs are estimated using iteratively reweighted least squares (IRLS).
- VGLM is implemented in the VGAM[5] package of R.
- ➢ Re-calibrate decadal tercile predictions by:

$$\eta_j = log\left(\frac{P(Y=s)}{P(Y=1)}\right) = \beta_{(j)1} + \beta_{(j)2}t + \left(\beta_{(j)3} + \beta_{(j)4}\tau\right)x_{(s)1}$$
$$+\left(\beta_{(j)5} + \beta_{(j)6}\tau^2\right)x_{(s)2} + \left(\beta_{(j)7} + \beta_{(j)8}\tau^3\right)x_{(s)3}$$

η_j: Linear predictor j=1,2 for categories normal (s=2) and above normal (s=3)
$\beta_{(j)p}$: Regression coefficients
$x_{(s)}$: Number of ensemble members in category s
τ: Lead year
t: Initialization year

Verification with Ranked Probability (Skill) Score:

- The RPS measures accuracy (RPS = REL – RES + UNC) of probability forecasts for more than two categories (e.g. terciles, below normal, normal, above normal).

$$RPS_\tau = \frac{1}{I}\sum_{i=1}^{I}\sum_{k=1}^{K}\left(F_{\tau,i,k}^2 - O_{t,k}^2\right)$$

$F_{\tau,i,k}$: Cumulative forecast probability of leadyear τ, initialization i and class k.

$O_{\tau,i,k}$: Cumulative probability of class k from observations for the time t(i,τ).

$$RPSS_\tau = 1 - \frac{RPS_{fc,\tau}}{RPS_{ref,\tau}}$$

$RPSS > 0$: Forecast is more skillful than reference.
$RPSS < 0$: Reference is superior to forecast.

- RPSS was calculated with FairRpss[1] function implemented in the SpecsVerification package.

3. Application

Data:

- Surface temperature
- Model: MPI-ESM-LR,
- Prototype (GECCO2)
- 15 ensemble members
- Initialization years: 1961-2011
- Lead years: 10
- Annual mean
- Reference: NCEP 20CR (1961-2009)

Cross-validation:

- 39 years were used for training and a 10 years window was left out for validation. The validation window was shifted year-wise
- $RPSS_{raw}$ (reference raw model) and $RPSS_{clim}$ (reference climatology) were calculated for the validation windows.

Results:

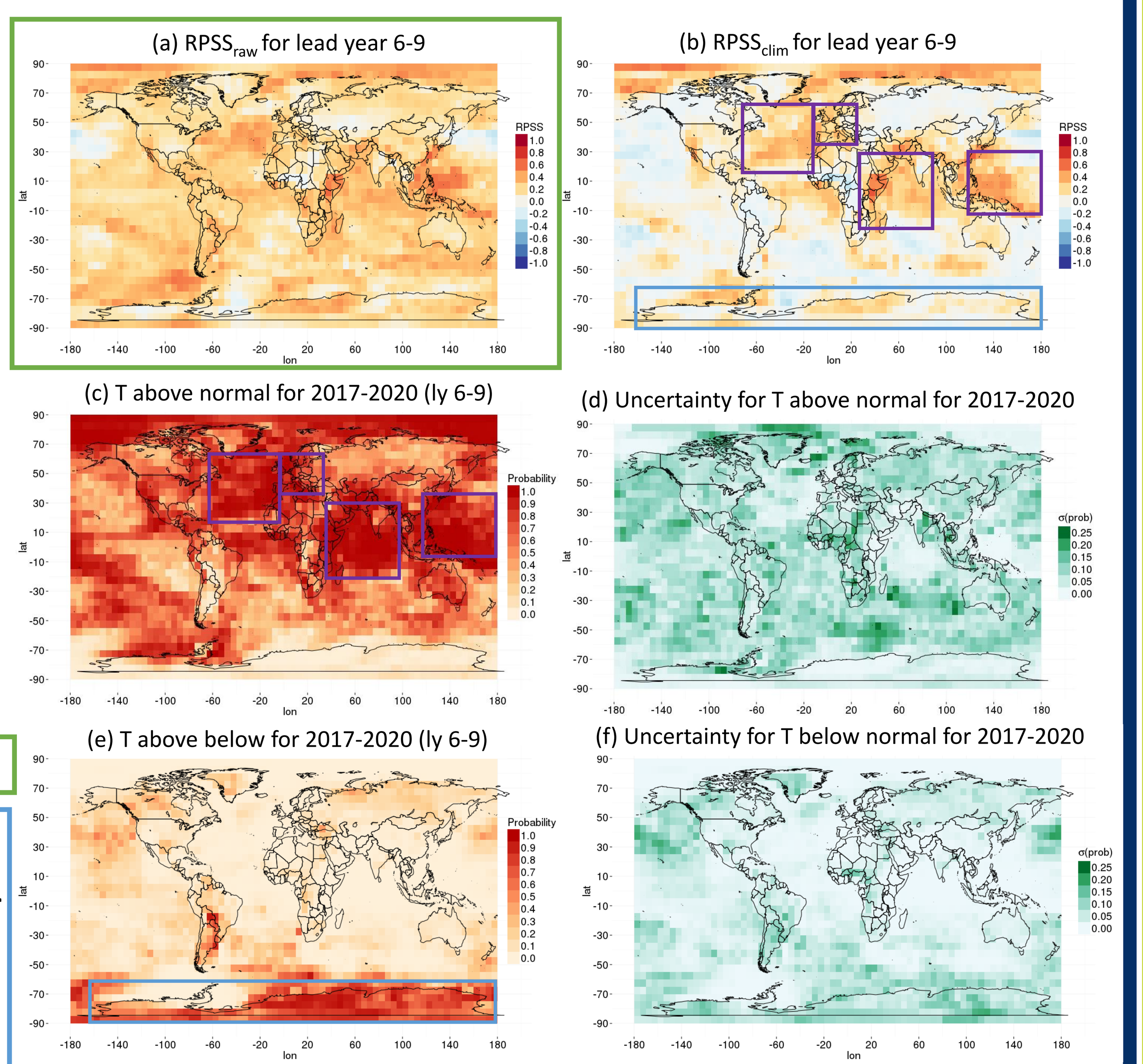
- Re-calibrated model is superior to raw model for most regions

High probability of warmer (than normal) temperature for years 2017-2020 over:

- North Atlantic ($RPSS_{clim} \approx 0.6$)
- Western Europe ($RPSS_{clim} \approx 0.4$)
- Indian Ocean ($RPSS_{clim} \approx 0.6$)
- West Pacific ($RPSS_{clim} \approx 0.8$)

High probability of colder (than normal) temperature for years 2017-2020 over:

- Antarctic ($RPSS_{clim} \approx 0.6$)



RPSS of the re-calibrated model w.r.t. raw model (a) and climatology (b) for lead years 6-9. (c) and (e) show the predicted probability of temperature above and below normal for years 2017-2020. The prediction was re-calibrated with regression coefficients estimated during training phases of cross-validation (1961-2009 -> 39 estimates for $\beta_{(j)p}$ and 39 re-calibrations). While (c) and (e) is the mean of this 39 re-calibrations, the corresponding uncertainties (d) and (f) were calculated by their standard deviation.

4. Summary

- Re-calibration shows an im-provement w.r.t. raw model for almost every region.
- Re-calibration also outperforms the climatology in many regions.
- Regions with both, high prob-abilities of warmer/colder temp-eratures and positive skill could be detected
- This regions also exhibit a small uncertainty due to parameter estimation.

Outlook:

- What happens if continuous predictions will be re-calibrated first and terciles calculated afterwards
- How great is the impact of the influencing factors lead year and initialization year?
- Use VGAM instead of VGLM

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., A. E. Raftery, A. H. Westveld and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Mon.Weath. Rev., 133, 1098–1118.
3. Weigel, A.P., M. A. Liniger, and C. Appenzeller, 2009: Seasonal ensemble forecasts: Are recalibrated single models better than multimodels? Mon. Weather Rev., 137(4):1460–1479.
4. Wilks, D. S., 2011: Statistical Methods in the Atmospheric Sciences. Academic Press.
5. Yee, T. W., 2015: Vector Generalized Linear and Additive Models: With an Implementation in R. New York, USA: Springer. ISBN 978-1-4939-2817-0.

FONA Mittelfristige Klimaprognosen BMBF
MiKlip

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

http://www.fona-miklip.de/de/257.php