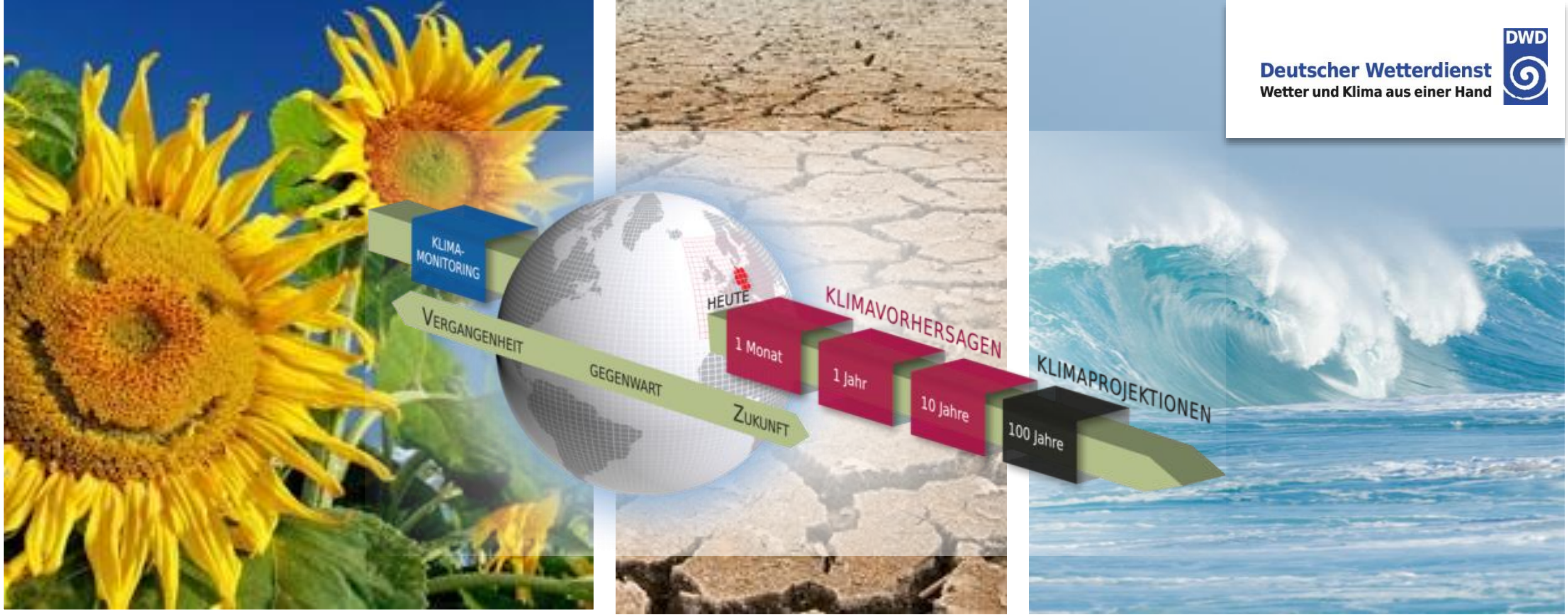


Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Postprocessing at DWD: subsampling the ensemble, recalibration, and statistical downscaling

C. Dalelane, A. Pasternack, P. Lorenz

Overview

1 Subsampling

2 Recalibration

3 Statistical Downscaling

Statistically Selected Seasonal Climate Prediction

ERA5 SST, SND, SIC, TA100

ERA5 SLP

ERA5 TAS

- In theory the model propagates the initial conditions in analogy to nature – not so in the Euro-Atlantic sector → low skill
- Simultaneous relationships between circulation and sensible weather more reliable

- **Dobrynin et al. (2018)** and **Dalelane et al. (2020)**: Improvement of the dynamial forecasts using statistically predicted circulation indices

GCFS assimilates SST, SND, SIC, TA100

GCFS propagates SLP

GCFS TAS forecast

Statistially Selected Seasonal Climate Prediction

ERA5 SST, SND, SIC, TA100

GCFS assimilates SST, SND, SIC, TA100

- **Dobrynin et al. (2018)** and **Dalelane et al. (2020)**: Improvement of the dynamial forecasts using statistically predicted circulation indices

ERA5 SLP

Prediction of circulation indices for leadtime

GCFS SLP select most accurate members

ERA5 TAS

- **Statistical prediction** of circulation indices based on past states of predictor fields
- Members of dynamical forecast ensemble realize indices during leadtime → **subselection** of members with „correct" indices
- Seasonal forecast based on subensemble

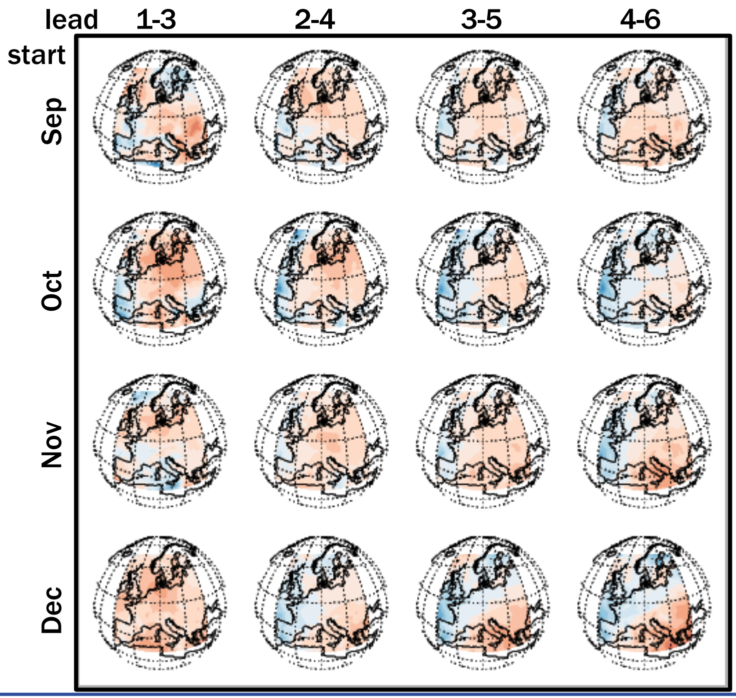
GCFS subsampled TAS forecast

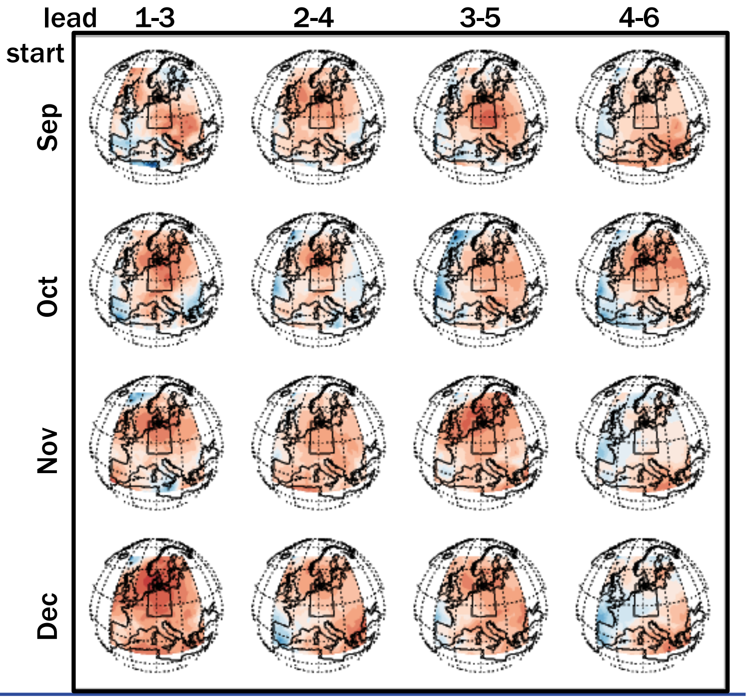
Statistially Selected Seasonal Climate Prediction

Skill vs. climatology

left: ensemble mean

right: current version of subselection on the DWD website



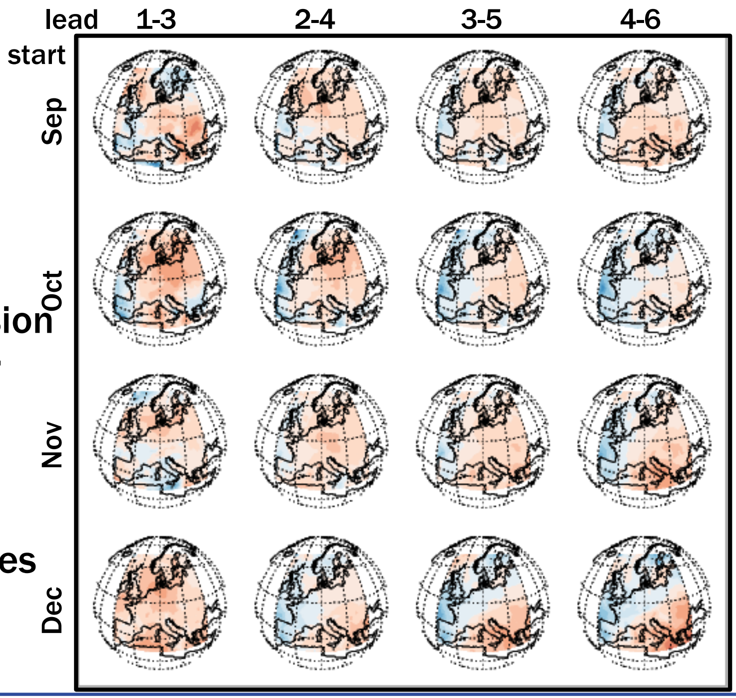


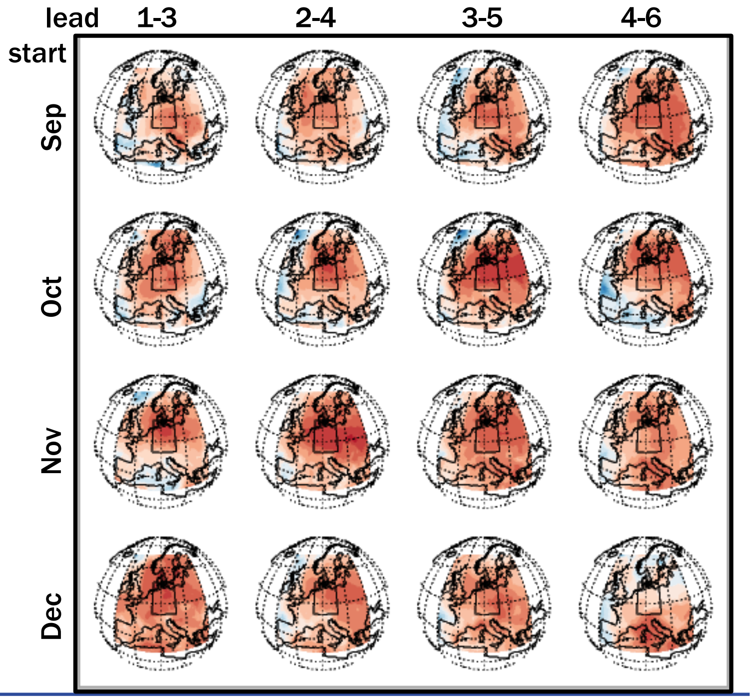
Statistially Selected Seasonal Climate Prediction

Skill vs. climatology

left: ensemble mean

right: latest version using 3D Tensor Partial Least Squares and modified circulation indices predictive for temperature





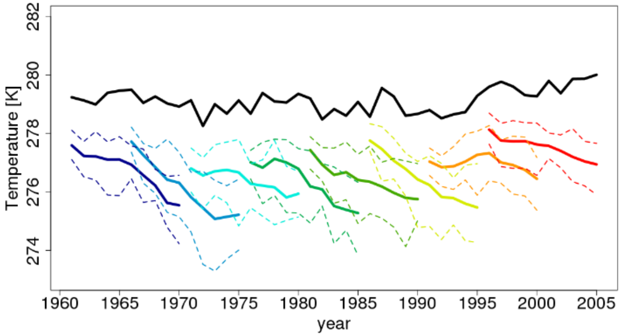
Overview

1	Subsampling
2	Recalibration
3	Statistical Downscaling

Statistical Post-Processing with Recalibration

Recalibration of decadal predictions

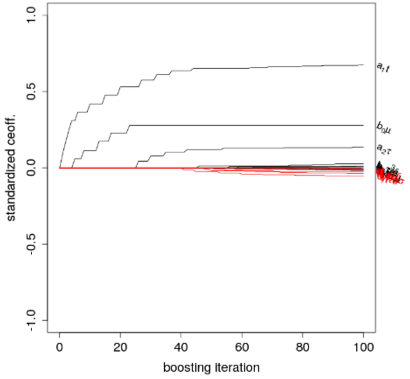
- (decadal) probabilistic forecasts tend to be not reliable
- maximize sharpness without sacrificing reliability
- limited number of hindcasts
- climate trend
- dependence on lead years (drift)



Decadal forecast recalibration strategy (DeFoReSt)

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

With: $\alpha(t,\tau) = \sum_{l=0}^{3}\left(a_{2l} + a_{(2l+1)}t\right)\tau^l$ … and $\beta(t,\tau), \gamma(t,\tau), \delta(t,\tau)$ analogously



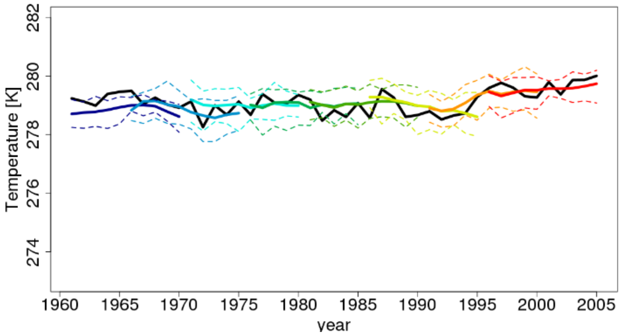
Model selection with boosting

- 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

Statistical Post-Processing with Recalibration

Recalibration of decadal predictions

- (decadal) probabilistic forecasts tend to be not reliable
- maximize sharpness without sacrificing reliability
- limited number of hindcasts
- climate trend
- dependence on lead years (drift)



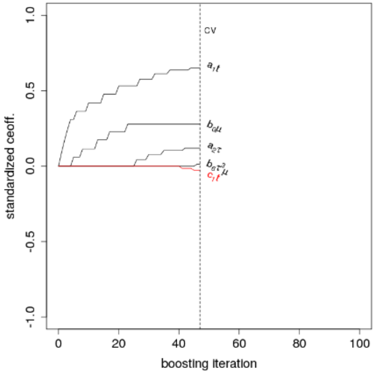
Decadal forecast recalibration strategy (DeFoReSt) (Pasternack et al. 2018, 2021)

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

With: $\alpha(t,\tau) = \sum_{l=0}^{3} \left(a_{2l} + a_{(2l+1)}t\right)\tau^l$ … and $\beta(t,\tau), \gamma(t,\tau), \delta(t,\tau)$ analogously

Model selection with boosting

- 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



Climate predictions @DWD

Subseasonal Predictions

- IFS (extended range)
- 46 days simulations
- 20 years BP
- ensemble members:
  - forecasts: 51
  - hindcasts: 11

~36 km

Seasonal Predictions

- GCFS2.1
- 6 month simulations
- 1990 – today
- ensemble members:
  - forecasts: 50
  - hindcasts: 30

~100 km

Decadal Predictions

- MPI-ESM LR
- 10a simulations
- 1961 – today
- 16 ensemble members

~200 km

Climate predictions @DWD

Subseasonal Predictions
- IFS (extended range)
- 46 days simulations
- 20 years BP
- ensemble members:
  - forecasts: 51
  - hindcasts: 11

~36 km

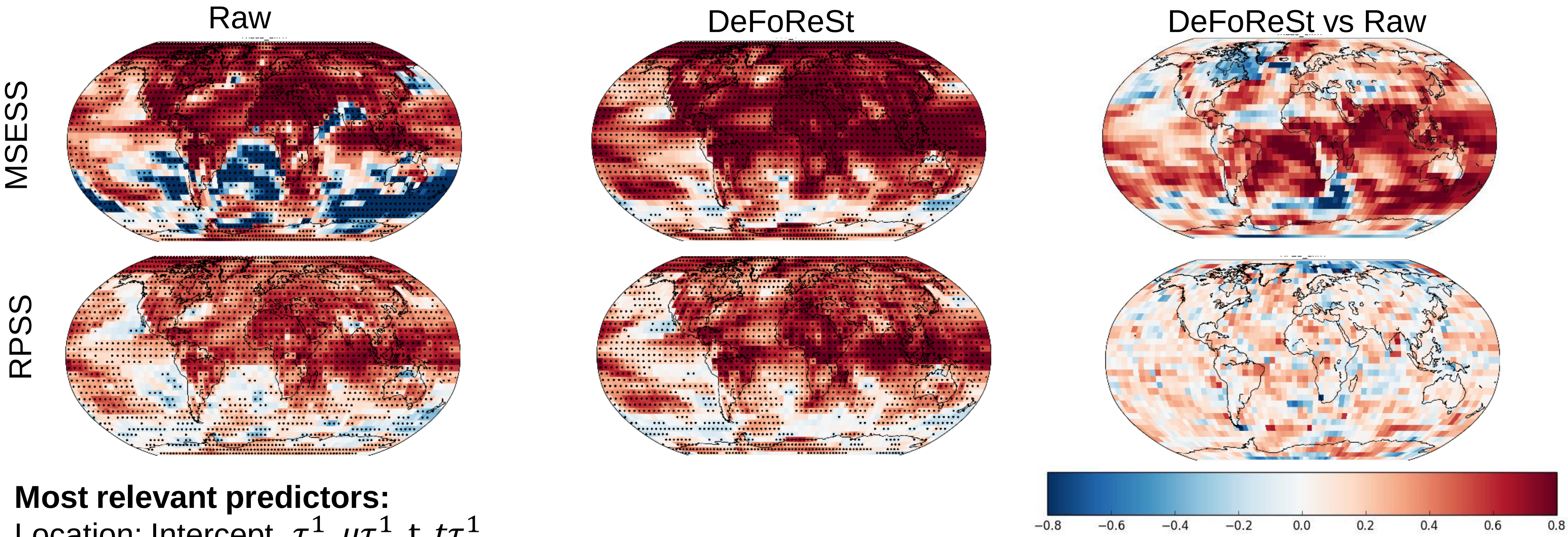
Seasonal Predictions
- GCFS2.1
- 6 month simulations
- 1990 – today
- ensemble members:
  - forecasts: 50
  - hindcasts: 30

~100 km

Decadal Predictions
- MPI-ESM LR
- 10a simulations
- 1961 – today
- 16 ensemble members

~200 km

Decadal Prediction: lead year 1-5 → 2m Temperature



Most relevant predictors:

Location: Intercept, $\tau^1, \mu\tau^1$, t, $t\tau^1$

Scale: Intercept, $\tau^1, \sigma\tau^1, \sigma\tau^2, \sigma\tau^3$

t: Start year
τ: Lead year
μ : Ens. Mean
σ : Ens Spread

Climate predictions @DWD

Subseasonal Predictions

- IFS (extended range)
- 46 days simulations
- 20 years BP
- ensemble members:
  - forecasts: 51
  - hindcasts: 11

~36 km

Seasonal Predictions

- GCFS2.1
- 6 month simulations
- 1990 – today
- ensemble members:
  - forecasts: 50
  - hindcasts: 30

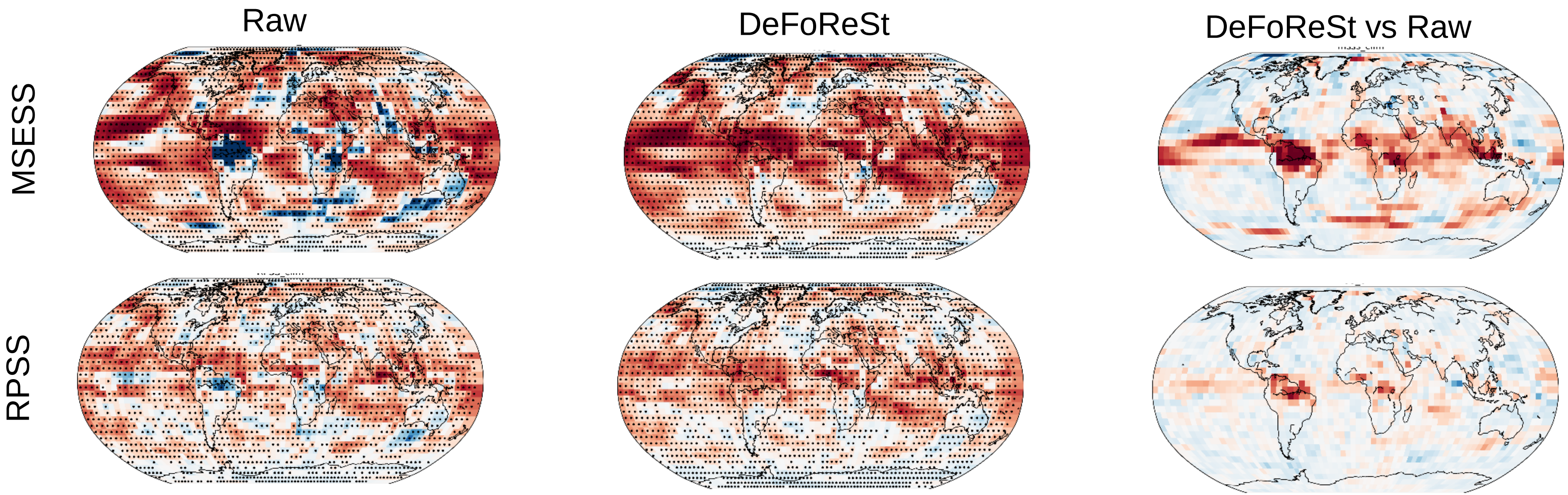
~100 km

Decadal Predictions

- MPI-ESM LR
- 10a simulations
- 1961 – today
- 16 ensemble members

~200 km

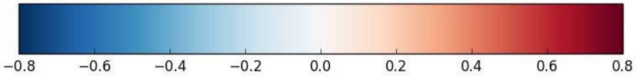
Seasonal Prediction for May: lead month 2-4 → 2m Temperature



Most relevant predictors:

Location: Intercept, $\tau^1, \tau^2, \tau^3, \mu, \mu\tau^1, \mu\tau^2, \mu\tau^3$

Scale: Intercept,$\tau^1, \tau^2, \tau^3, \sigma, \sigma\tau^1, \sigma\tau^2, \sigma\tau^3$



t: Start year μ *: Ens. Mean*

τ*: Lead month* σ *: Ens Spread*

Overview

1 Subsampling

2 Recalibration

3 Statistical Downscaling

Empirical-Statistical downscaling: EPISODES

Target:

→ To provide a solid basement for regional impact modelers and decision makers

→ Climate projections: Enlargement of the EURO-CORDEX ensemble (CMIP5, CMIP6)

→ Climate predictions: Regionalization of global climate projections (decadal, seasonal, sub-seasonal)

Requirements (e.g. for hydrological models):

→ Transient daily data (1951-2100) in high spatial resolution

→ Spatial consistent for at least large river catchments

→ Multivariate consistent

→ No systematic deviations (Bias)

EPISODES

Data (daily values):

- ➔Gridded observations: HYRAS / TRY / COSMO-REA6 /ERA5-Land
- ➔Reanalysis data: NCEP/NCAR
- ➔GCM model output: e.g. CMIP5, CMIP6, GCFS2, IFS, ...

Step 1: Analogue day regression

- ➔Detection of analogue days and regression („Perfect Prog“)
- ➔Based on large scale atmospheric (Geopotential, Temperature und rel. Humidity in 1000, 850, 700 und 500 hPa)

Step 2: Time series construction

- ➔From the results of step 1, a synthetic time series is created in resolution of the used observation data

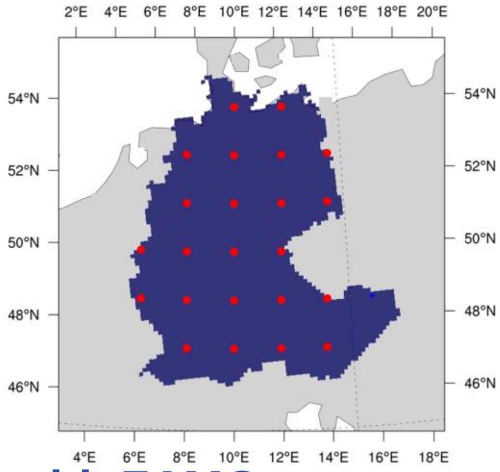
Selector and predictor fields

For each variable and season

Two selector variables:
for detection of the 35 most similar days from the reanalyses for each day of GCM data ("Perfect Prognosis" approach):

➔ Target variables: tas, pr, tasmin, tasmax, hurs.

Predictor variable: for posterior linear regression



From Germany to DACH region – co-operarion with ZAMG (work in progress)

➔ Regionalisation of selector and predictor variables (implemented by ZAMG)

➔ Cross validation for determination of optimal selectors/predictors for DACH region

Step 2: Construction of the daily time series

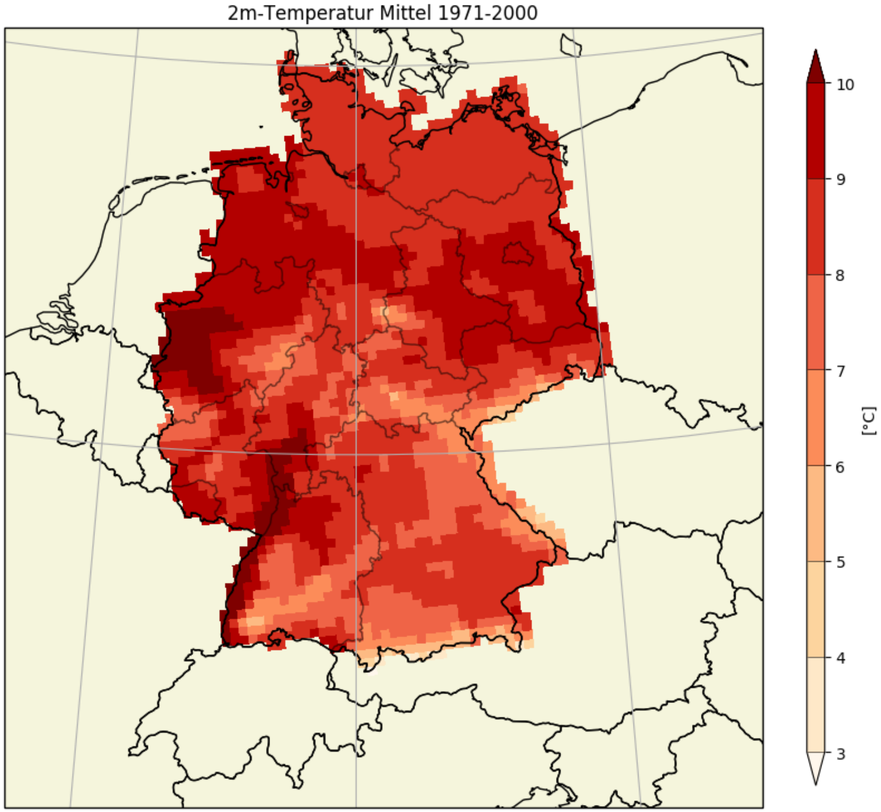
- ➔ Calculation of the short-term variability "V":
  - ➔ GCM: Difference between the results of the regression (step 1) and the low-pass filtered "Climate change guidance G“
  - ➔ Observations: Difference to climatology (C)
  - ➔ Based on the short-term variability (V) of temperature and precipitation, the most similar day from the observations is detected for each day of the GCM data. This day is then used for all variables.
- ➔ The final value is calculated as the sum of $C + G + V_{obs}$:

C : daily climatology (observations)

G : climate change guidance: low-pass filtered results of the regression (step 1)

V_{obs} : short-term variability of the detected day from the observations

Currently used EPISODES model domain



EPISODES output

- Gridded data on HYRAS-5km or CORDEX-EUR11 grid (appr. 12.5km), up to now only grid points within Germany
- Temporal resolution: daily
- Elements:
  - tas – 2m daily mean temperature
  - tasmin – 2m daily minimum temperature
  - tasmax – 2m daily maximum temperature
  - pr – daily precipitation sum
  - hurs – daily mean 2m relative humidity
  - rsds – daily mean downward solar radiation
  - clt – daily mean cluod cover
  - sfcWind – daily mean wind speed (10m)
  - psl – daily mean sea level pressure


Thank you very much
for your attention!