Exploring the performance of bias correction applied outside the calibration period's climate regime

Katharina Klehmet[1] , **Peter Berg[1]** , Pascual Herrera[2] , David Leidinger[3] , Anthony Lemoine[4] , Ernesto Pasten-Zapata[5] , and Rafael Pimentel[2]

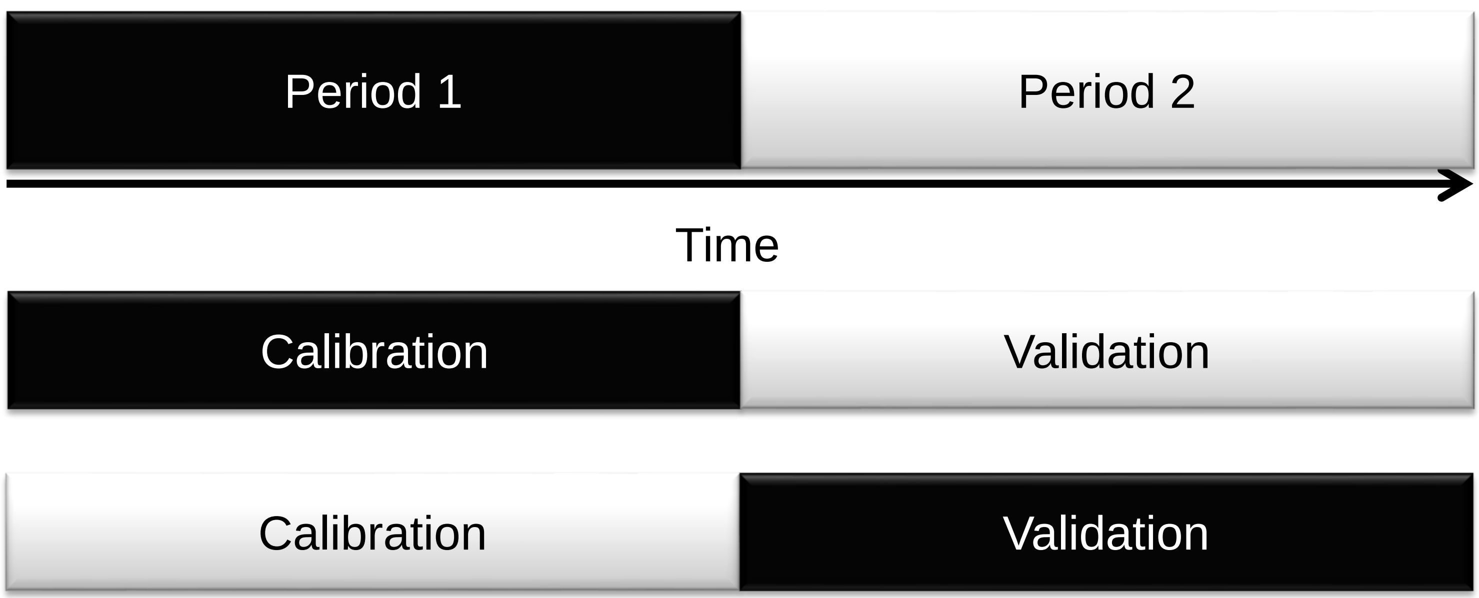
1 Swedish Meteorological and Hydrological Institute (SMHI) , Hydrology Research, Norrköping, Sweden
2Andalusian Institute for Earth System Research, University of Cordoba, Córdoba, Spain
3BOKU, Institut für Meteorologie, Wien, Austria
4Université Paris-Saclay, INRAE, UR HYCAR, Antony, France
5Geological Survey of Denmark and Greenland, Hydrology, Denmark

The basics

- Bias correction is calibrated on some recent period with both observations and model data available
- The future might bring a very different climate compared to the calibration period
- How can we know if the bias correction model holds in the new climate regime?

Split Sample Testing (SST)

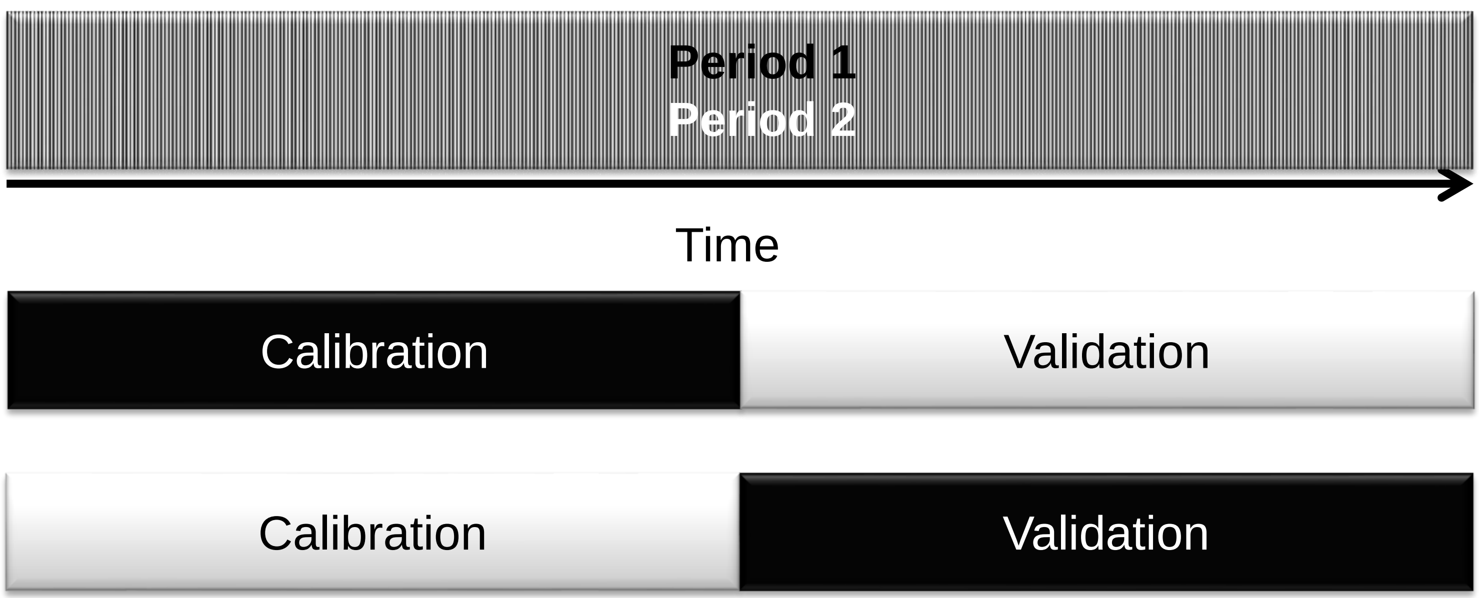
- Split you calibration and validation into two separate samples, e.g. cutting your period into two sub-periods



SMHI

Differential Split Sample Testing (DSST)

- Sort your data after some metric to separate the samples into two distinct climate regimes
- Split you calibration and validation into two separate samples, e.g. cutting your period into two sub-periods



[Klemeš, 1986]

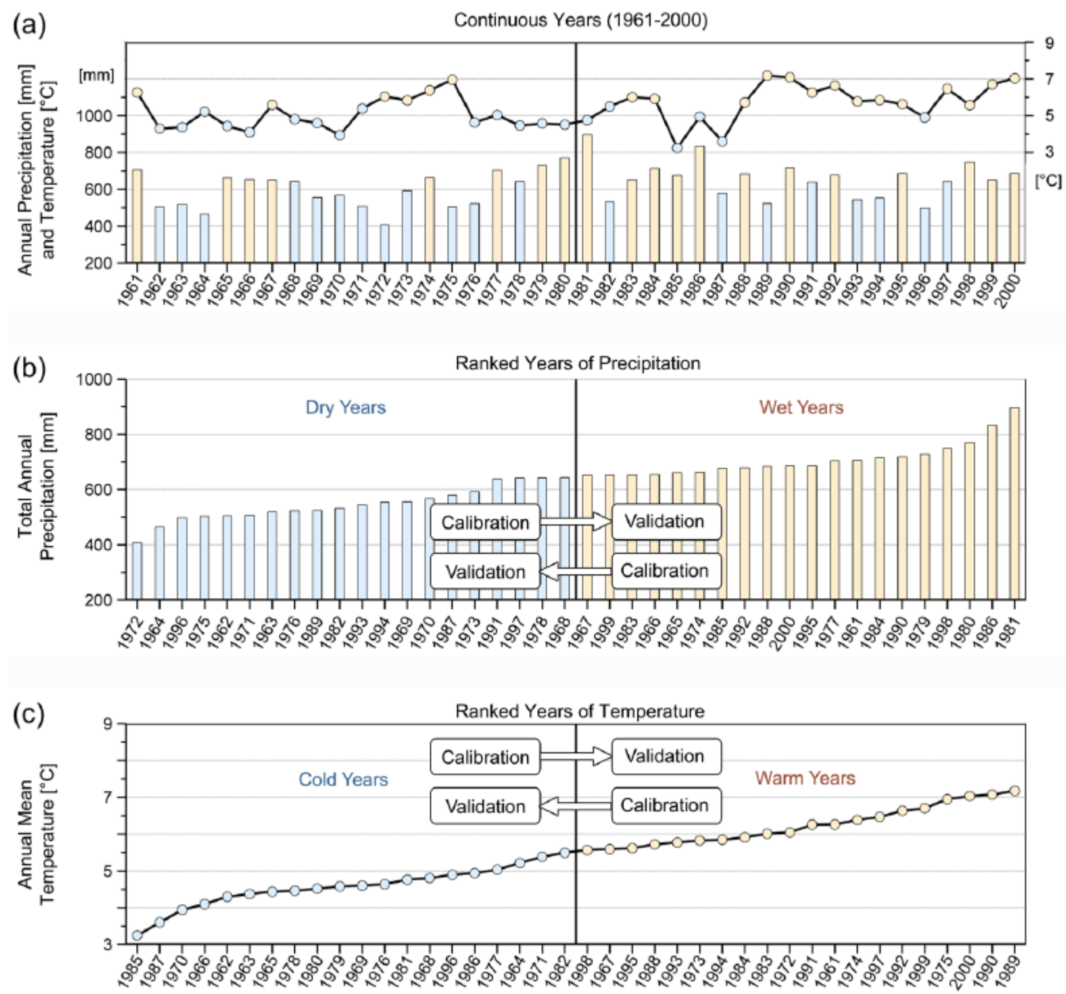
What has been done before, and what is new?

Teutschbein et al. (2013) [TS13] made a DSST experiment using **reanalysis** driven model simulations and separated according to wet and dry years, or cold and warm years.

Wang et al. (2018) explored stationarity tests based on annual average temperature, precipitation, sea surface temperature and the NAO circulation **index**.

Our extension of TS13 is:

- To use pseudo-reality setup, i.e. to use models as reference instead of observations
- To use long time series for sufficient samples
- To investigate multiple metrics



[From TS13]

Five metrics and the DSST setups

DK: mean pr oct—mar, sort wet to dry years

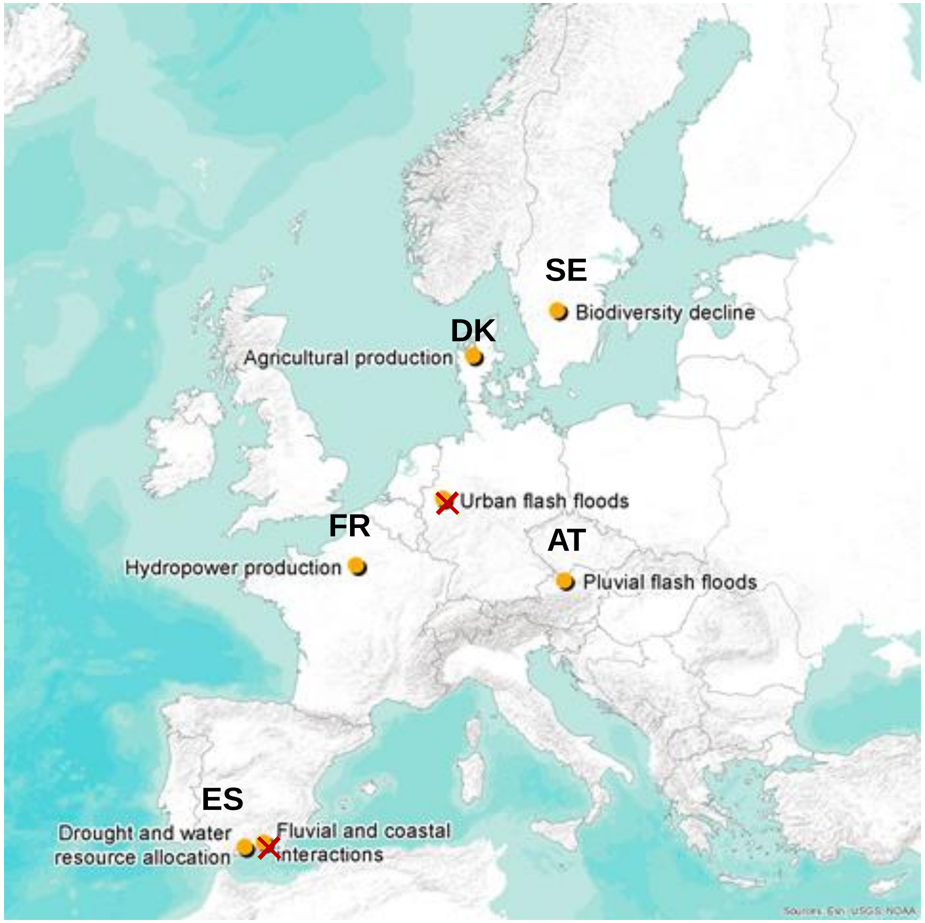
SE: Joint ranking from wet&cold – dry&warm AMJ periods

AT: Rank by JJA mean of heavy pr (>P95 historical).

FR: # of 15-day runmean > P95 hist; hot days

ES: Joint ranking based on annual snowfall amount and number of snowfall days

The aquaclew case studies; https://aquaclew.eu/case-studies/



Five metrics and the DSST setups

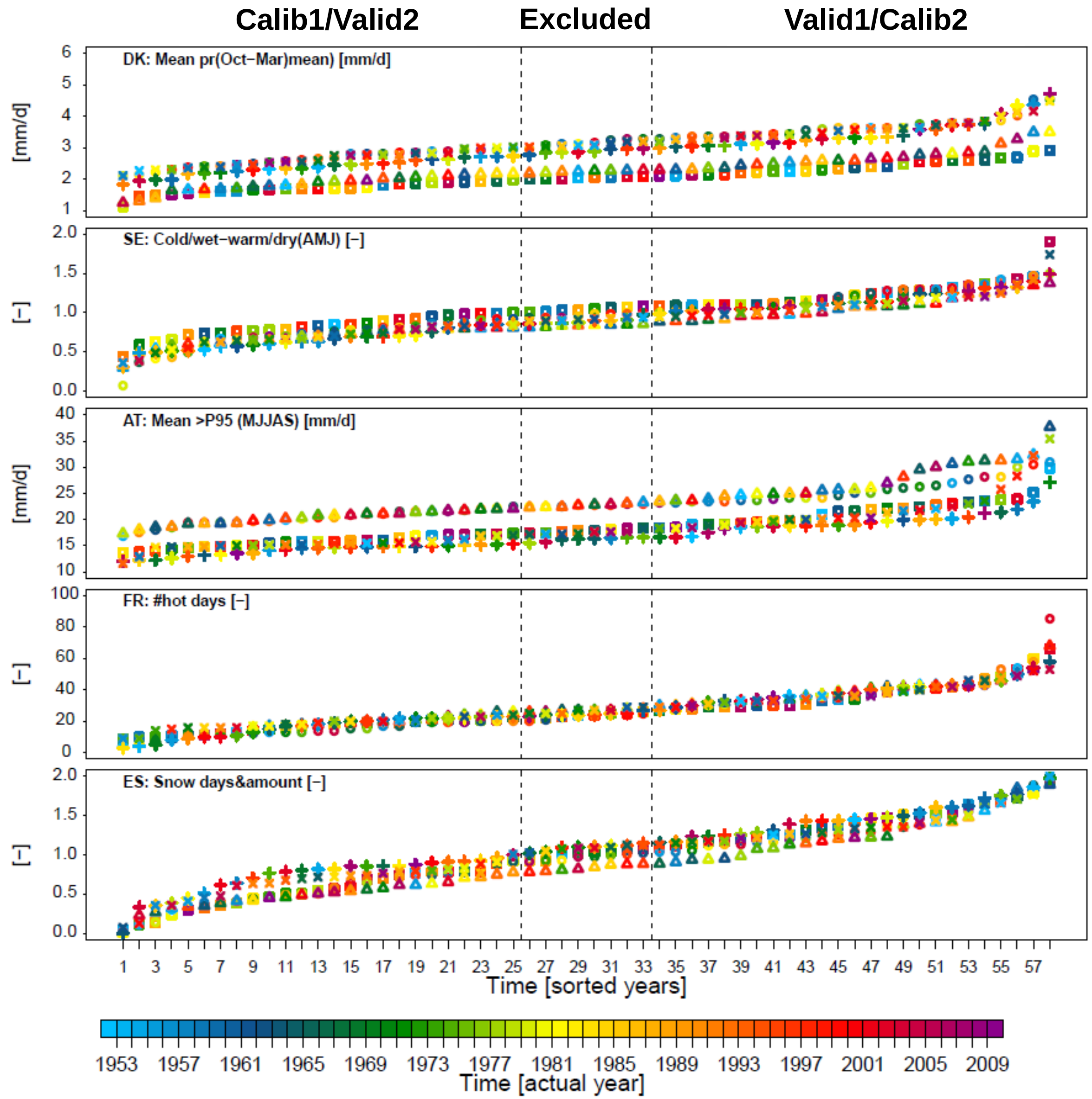
DK: mean pr oct—mar, sort wet to dry years

SE: Joint ranking from wet&cold – dry&warm AMJ periods

AT: Rank by JJA mean of heavy pr (>P95 historical).

FR: # of 15-day runmean > P95 hist; hot days

ES: Joint ranking based on annual snowfall amount and number of snowfall days



Metrics calculated for all models and ranked in increasing order; colours refer to actual model years

Pseudo-reality

In our ensemble of six EURO-CORDEX members, we allow each to act as reference once, to which all models are bias corrected and evalutated.

We then get 30 bias corrected models for each case, and in addition the two split sample tests (increasing or decreasing case from the previous slide).

Model ensemble from Euro-CORDEX 0.11 (12 km)

CNRM-CERFACS-**CNRM-CM5**--CNRM-**ALADIN53**

CNRM-CERFACS-**CNRM-CM5**--RMIB-UGent-**ALARO-0**

ICHEC-**EC-EARTH**--CLMcom-**CCLM4-8-17**

ICHEC-**EC-EARTH**--KNMI-**RACMO22E**

ICHEC-**EC-EARTH**--DMI-**HIRHAM5**

MPI-M-**MPI-ESM-LR**--MPI-CSC-**REMO2009**

Bias correction methods

Quantile mapping type:

MIdAS (MultI-scale bias AdjuStment) – separately correcting 30-day averages and daily data using empricial quantile mapping. (Berg et al., in prep)

DBS90 (Distribution Based Scaling) - Fits a distribution to a theoretical PDF. For precipitation, a double Gamma distribution (90th percentile) is used. For temperature, a normal distribution is used for the correction.

DBS95 (Distribution Based Scaling; UCO): Fits a distribution to the PDF. Precipitation uses a double gamma distribution (95th percentile) and temperature a normal distribution.

Other types:

LS (Linear Scaling): constant factor from comparison of observation and model samples. pr (tas) corrected with multiplicative (additive) term based on difference of long-term monthly means (Lenderik et al., 2007).

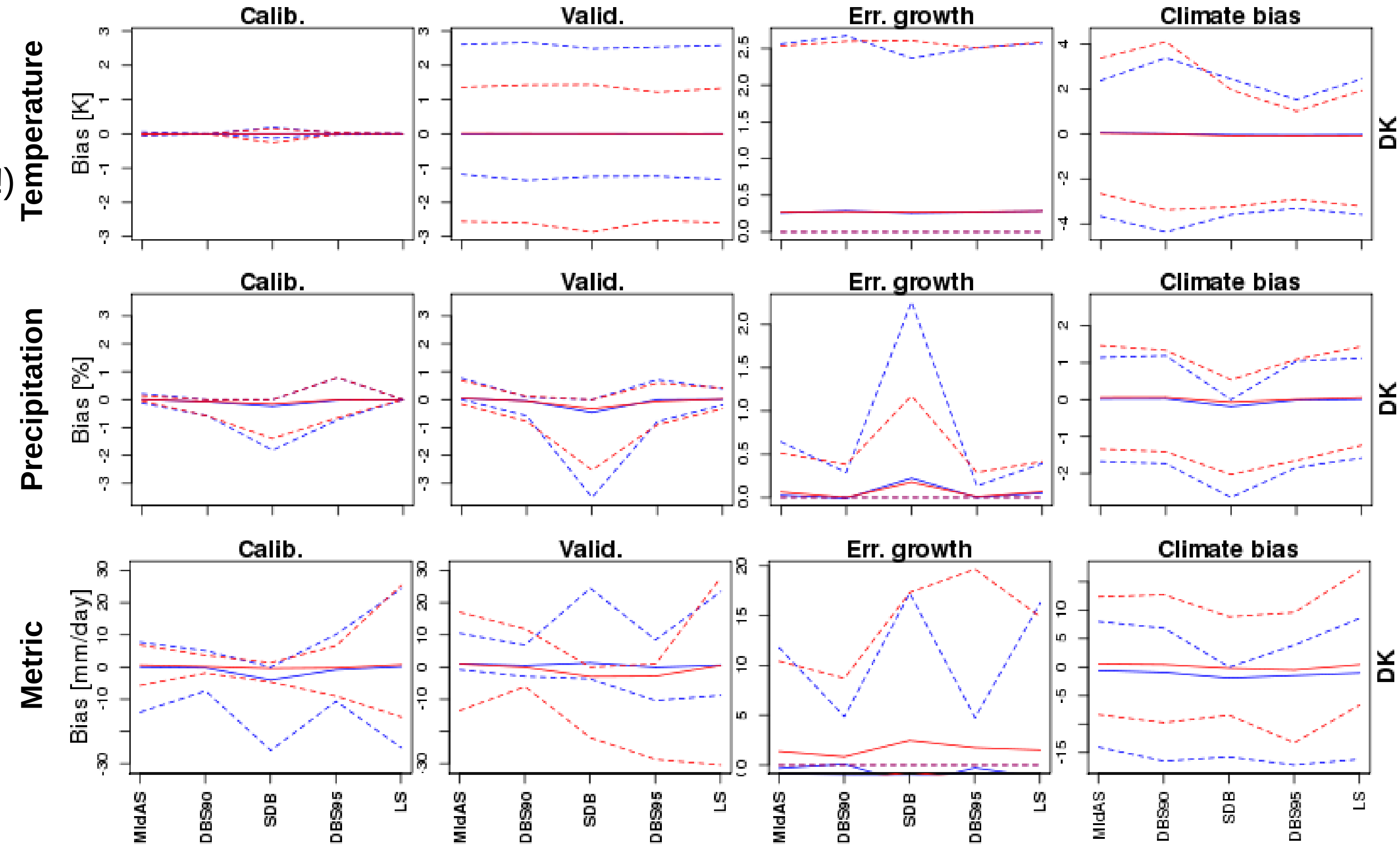
SDM (Scaled Distribution Mapping): scales observed distributions by raw model projected changes in magnitude, rain-day frequency and likelihood of events (return period). No assumption of stationarity. Reference: Switanek et al. (2017).

Example: Performance in DK (wet-dry) case

Temperature
Calib. **OK**
Valid **OK** (but outliers!)
Climate **OK** (but outliers!)

Precipitation
Calib: **OK, except SDM and DBS95**
Valid: reverse error from calib.
Climate: some calib period dependency

Metric
Calib: period dependency
Valid: ~reverse error from calib.
Climate: calib period dependency



Red lines for increasing wetness between calib and valid, blue for decreasing. **Solid** lines for median of all models, **dashed** lines for the single most extreme cases.

Conclusions

More analysis is still to come, but preliminary results show:

- The bias of the model (difference between reference and a particular model) has an impact on the bias correction performance in both the calibration and validation periods. Precipitation is more sensitive than temperature.

- Error growth between calibration and validation periods are mostly acceptable, although some methods have problems with the mean statistic.

- Projected climate change is generally insensitive to the calibration period, but when there is an effect, the DSST approach can give an idea about the impact.