UNIVERSITY OF BERGEN

Challenges in bias correction of climate projections

Asgeir Sorteberg

Geophysical Institute, UiB
Bjerknes Centre for Climate Research, UiB

What are the challenges?

Examples from Norway

Are Bias Corrections a good or bad thing?

Yes or No …

Why bias correct?

Due to imperfect model representations which seriously hampers the quality of impact models.

What is a bias correction?

Should ideally correct the discrepancy between a model and reality on the scales resolved by the model

**Joint Working Group on Forecast Verification Research, (JWGFVR)*

What is downscaling?

Downscaling attempts to resolve the scale discrepancy between the resolutions required for impact assessment and the models resolution.

Bias correction or downscaling?

In practical use

The distinction between bias correction and downscaling is unclear

Many bias correction attempts also include a downscaling component

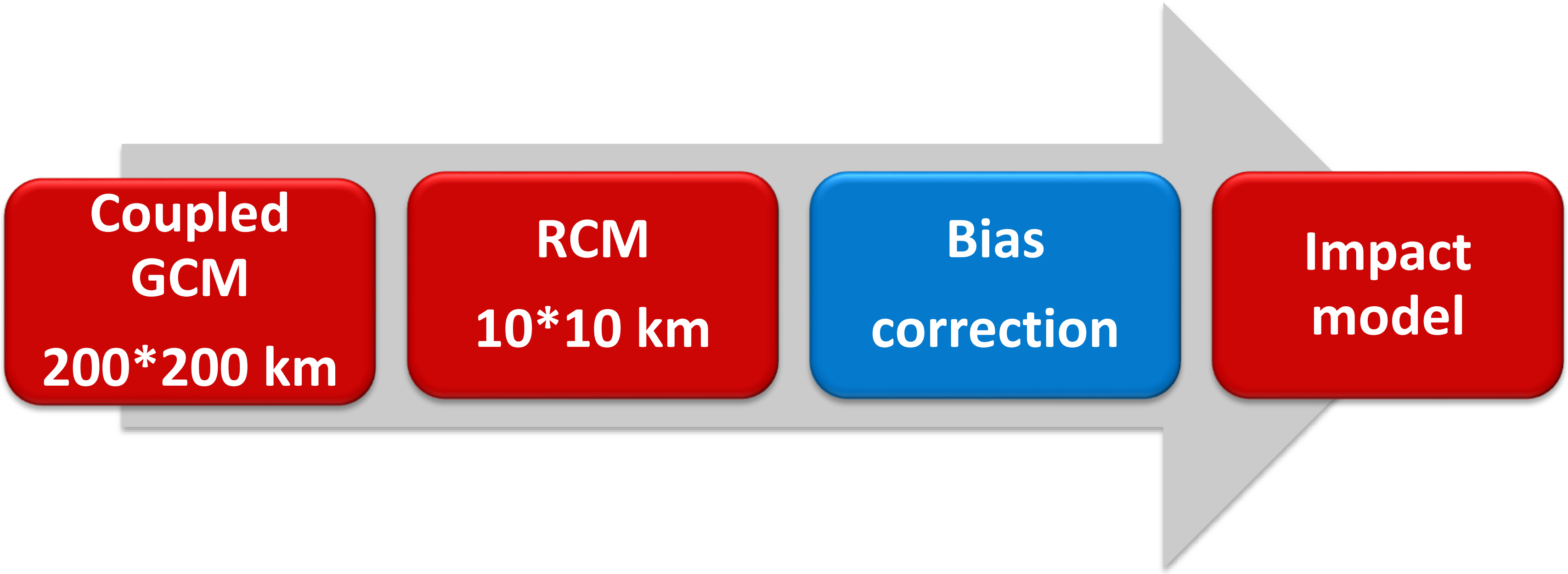
- ✓ Correction of coarse resolution data to point locations
- ✓ Correction of coarse resolution data to a finer grid

Methods for reducing the biases

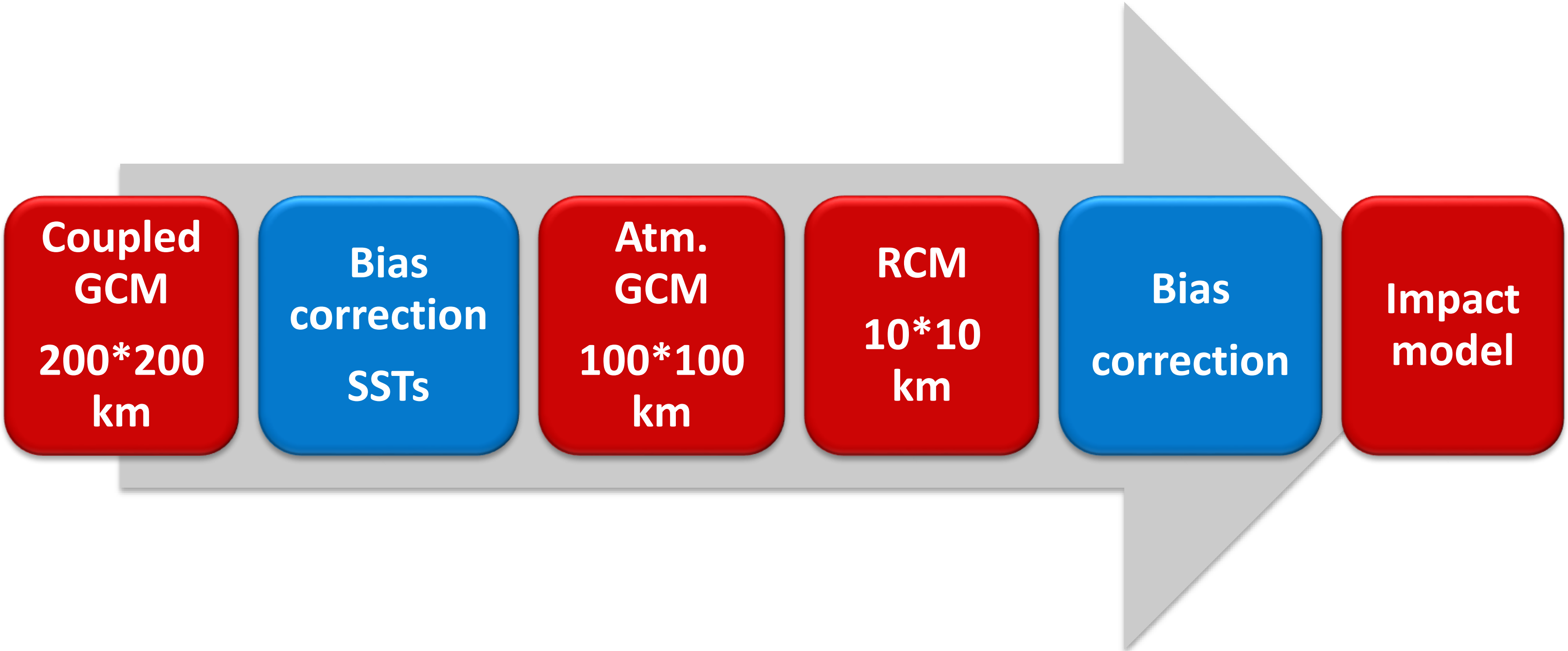
1. *Reduce the bias* by improving the models
2. Multi-model ensembles - averaging over the ensemble tend to *reduce the bias* compared to single model approaches and it provides an uncertainty estimate
3. **Model output statistics (MOS) - *masking the model biases* in a post processing step.**

A fourth method Empirical-statistical downscaling where the intention is to establish links between observed large scale predictors and observed local scale predictands rather than correcting model errors

The climate modelling chain



The climate modelling chain



Bias correction methods

In practice:
select your favorite transfer function $f\left(x_{obs}, x_{org}\right)$
and do the correction

$$x_{corr} = x_{org} \cdot f\left(x_{obs}, x_{org}\right)$$

$$x_{corr} = x_{org} + f\left(x_{obs}, x_{org}\right)$$

Note
most bias corrections assume that the transfer function is time-independent and can be used for the future

Method	Description	Comments	
Delta change approach	RCM-simulated future change signals (anomalies) superimposed on observational time series	**Same correction factor for all events**	Freq. and intensities not corr. separately
Linear scaling	Adjusts RCM time series with correction values based on the differences between mean observed values and RCM simulation		
Local intensity scaling	combines a precipitation threshold with the linear scaling		Freq. and intensities corr. separately
Power trans-formation	Combines the correction of the coefficient of variation (CV) with linear scaling.	**Events adjusted non-linearly.**	Freq. and intensities partly corr.
Distribution mapping	Matches selected distribution functions of observations and RCM-simulated climate values, plus a precipitation threshold		Freq. and intensities corr. separately

What is the problem?

- ✓ Suitability of observationally based estimates
- ✓ Misrepresentation of event size
- ✓ Misrepresentation of temporal and spatial correlations
- ✓ Stationarity and the choice of length of the control period
- ✓ Selection of timescale for bias correction
- ✓ Physical Consistency
- ✓ Introduction of artificial climate change signals

Suitability of observationally based estimates

How close are the observationally based estimates to the truth?

- ✓ non-representativeness of the underlying observations
  - ➢ bias in location
  - ➢ inhomogeneities in the observed dataset
- ✓ instrument limitations
  - ➢ undercatch of precipitation
- ✓ assumptions made in the interpolation procedure.
  - ➢ Changes in the observational network

Suitability of observationally based estimates

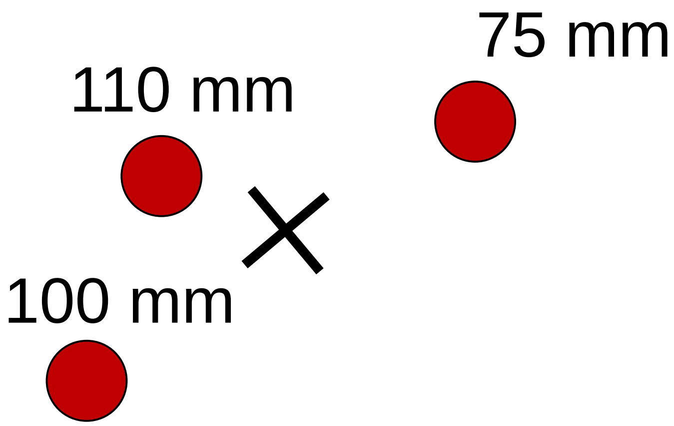
Changes in the observational network may introduces systematic changes in variability

Suitability of observationally based estimates

Example: interpolation

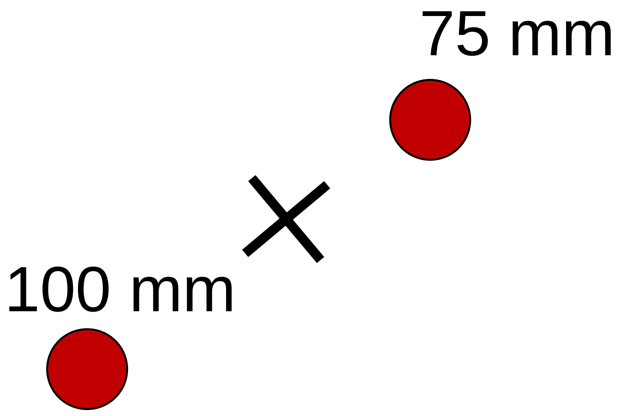
$$x_{\text{interp}} = \frac{x_1 \cdot {}^{1}\!/_{d_1} + x_2 \cdot {}^{1}\!/_{d_2} + x_3 \cdot {}^{1}\!/_{d_3}}{\left({}^{1}\!/_{d_1} + {}^{1}\!/_{d_2} + {}^{1}\!/_{d_3} \right)}$$

Event 1



$$x_{\text{interp}} = 98.8\text{mm}$$

Event 2



$$x_{\text{interp}} = 87.5\text{mm}$$

Suitability of observationally based estimates

Climate studies need tailored variance adjusted observational products

This requires different interpolation strategies than the *“best-estimate of the day”* interpolations used for operational applications

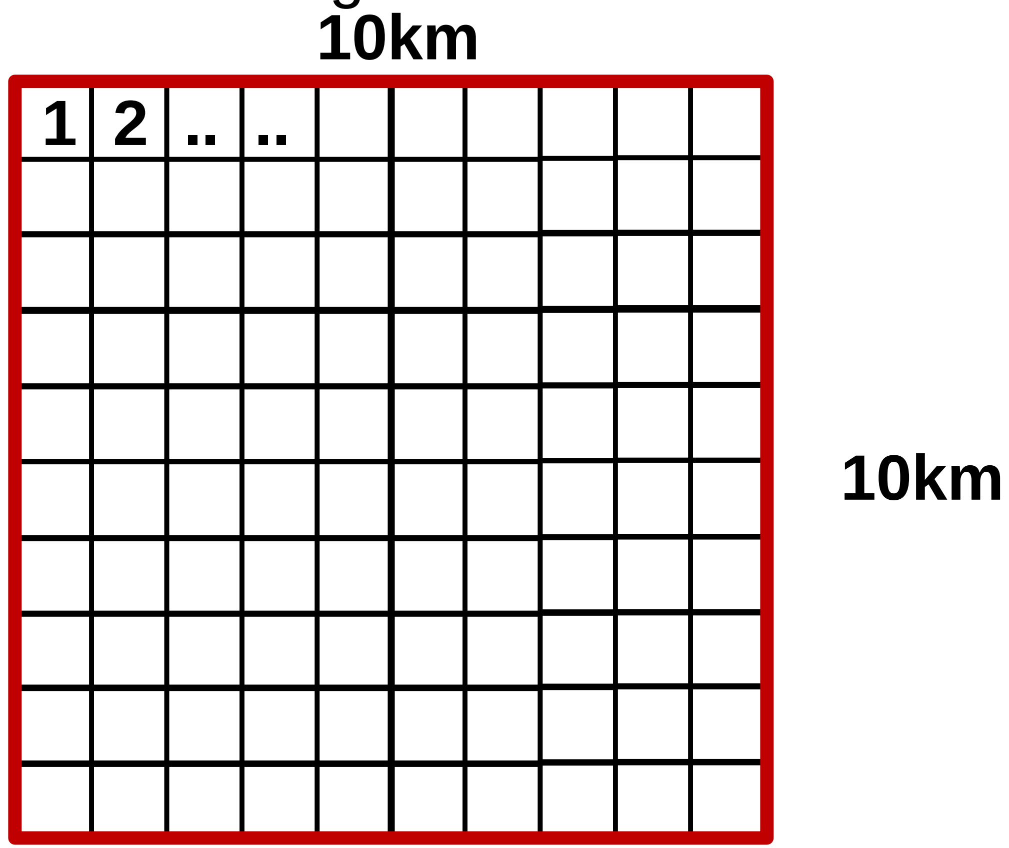
The bias correction will never be better than the quality of the observational estimate

Spatial and temporal correlations

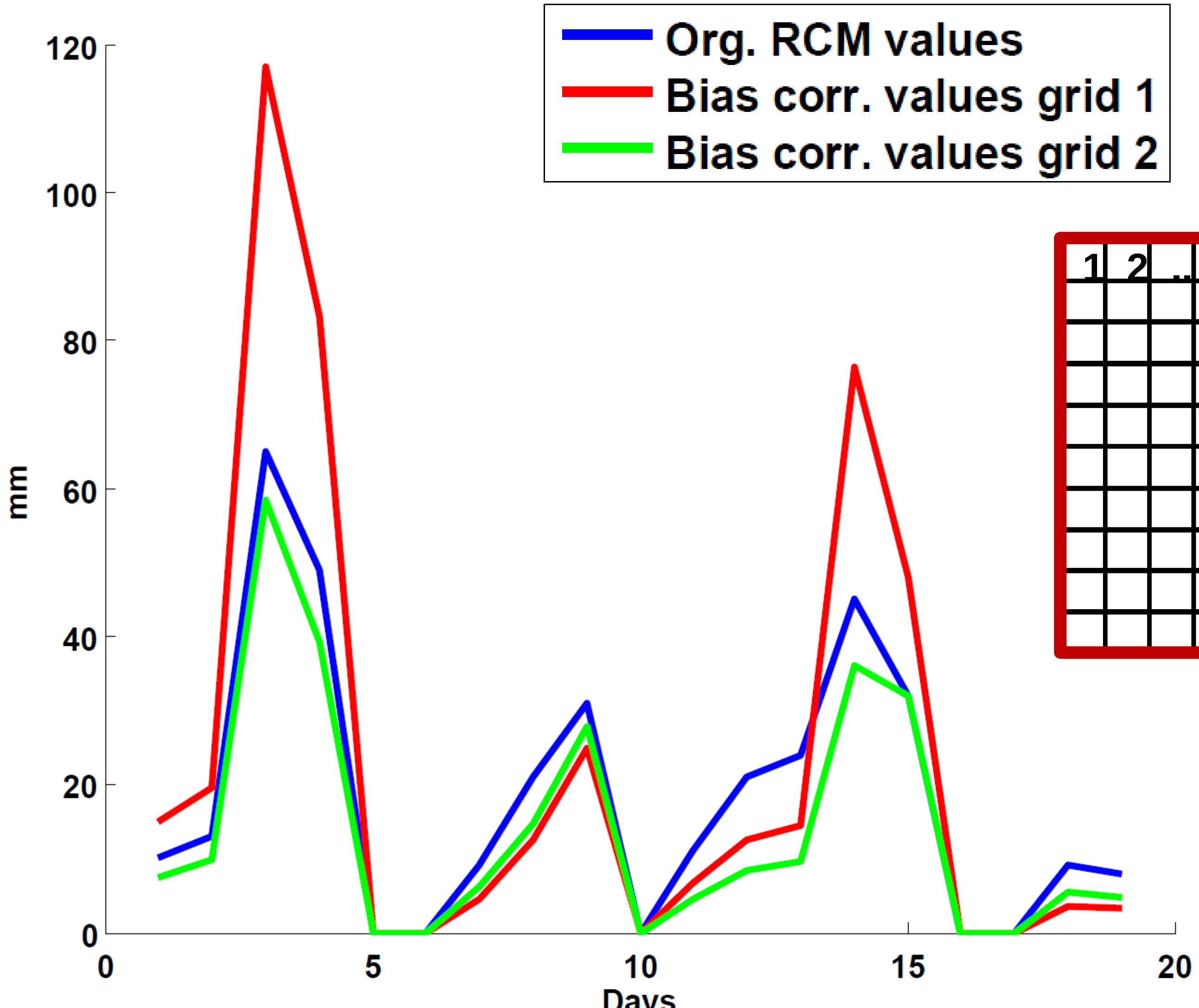
Most bias correction methods are not designed for correcting errors in spatial or temporal correlations

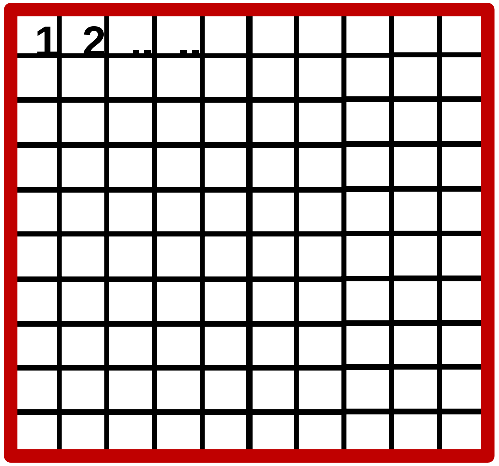
Model grid

Observational based grid



Spatial and temporal correlations





Spatial and temporal correlations

Even if the distributions looks reasonable in the different points.

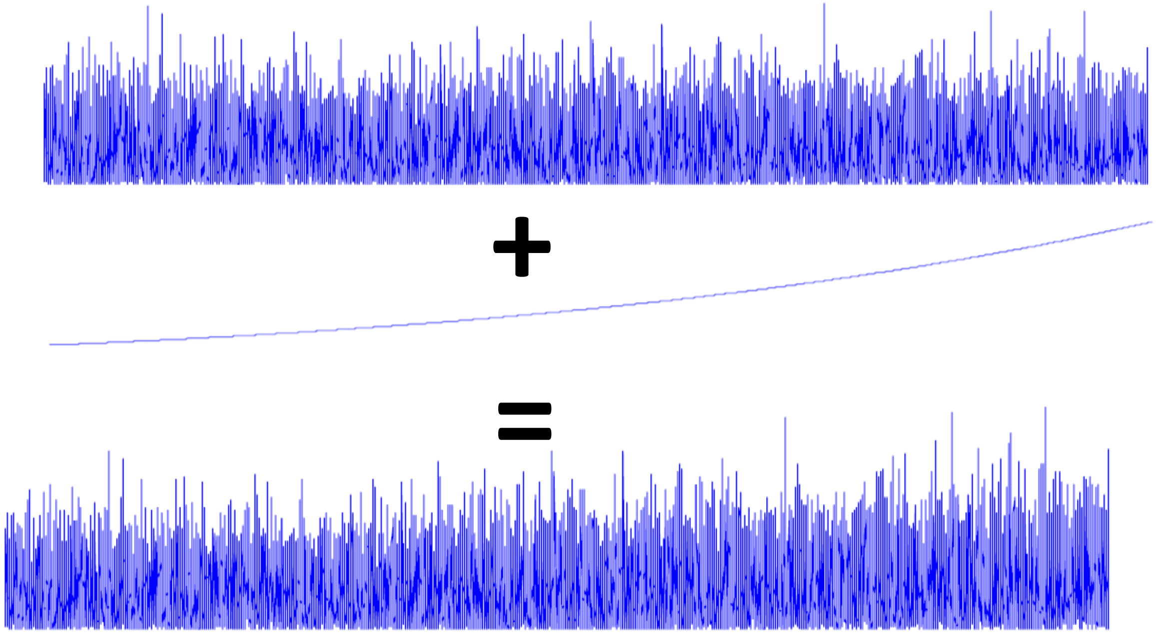
- ✓ Area-mean dry vs wet days gets overcorrected
- ✓ Area-mean extremes are often overestimated

Point based bias corrections need to be evaluated on catchment scale

Stationarity and length of the control period

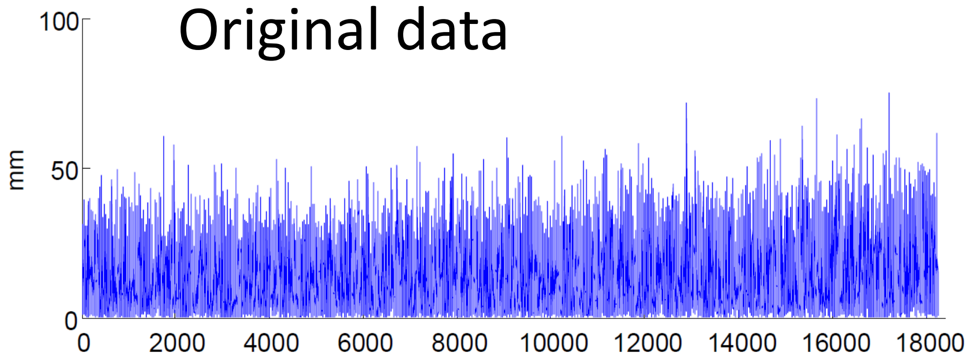
If the data is not stationary the decadal and longer variability may be hampered by the bias correction

Example: random values with a trend

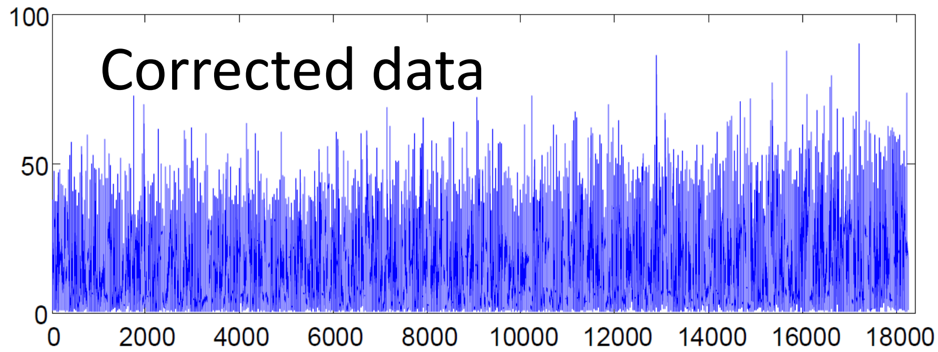


Stationarity and length of the control period

BC gives different corrections for different values
Simple: 20% increase for values higher than mean and 20% reduction for values lower than mean



Trend 3.95 mm/50 yrs
29%/50 yrs



Trend 5.0 mm/50 yrs
35%/50 yrs

Physical Consistency

Most bias corrections are done separately for different parameters

Links and feedbacks between the meteorological parameters are often broken with bias corrections.

The loss of physical consistency may have severe effects on things like snowmelt and evapotranspiration.

If your application has strong sensitivities to the combined effects of temperature and precipitation (or other climate variables) the lack of physical consistency may be a major issue!

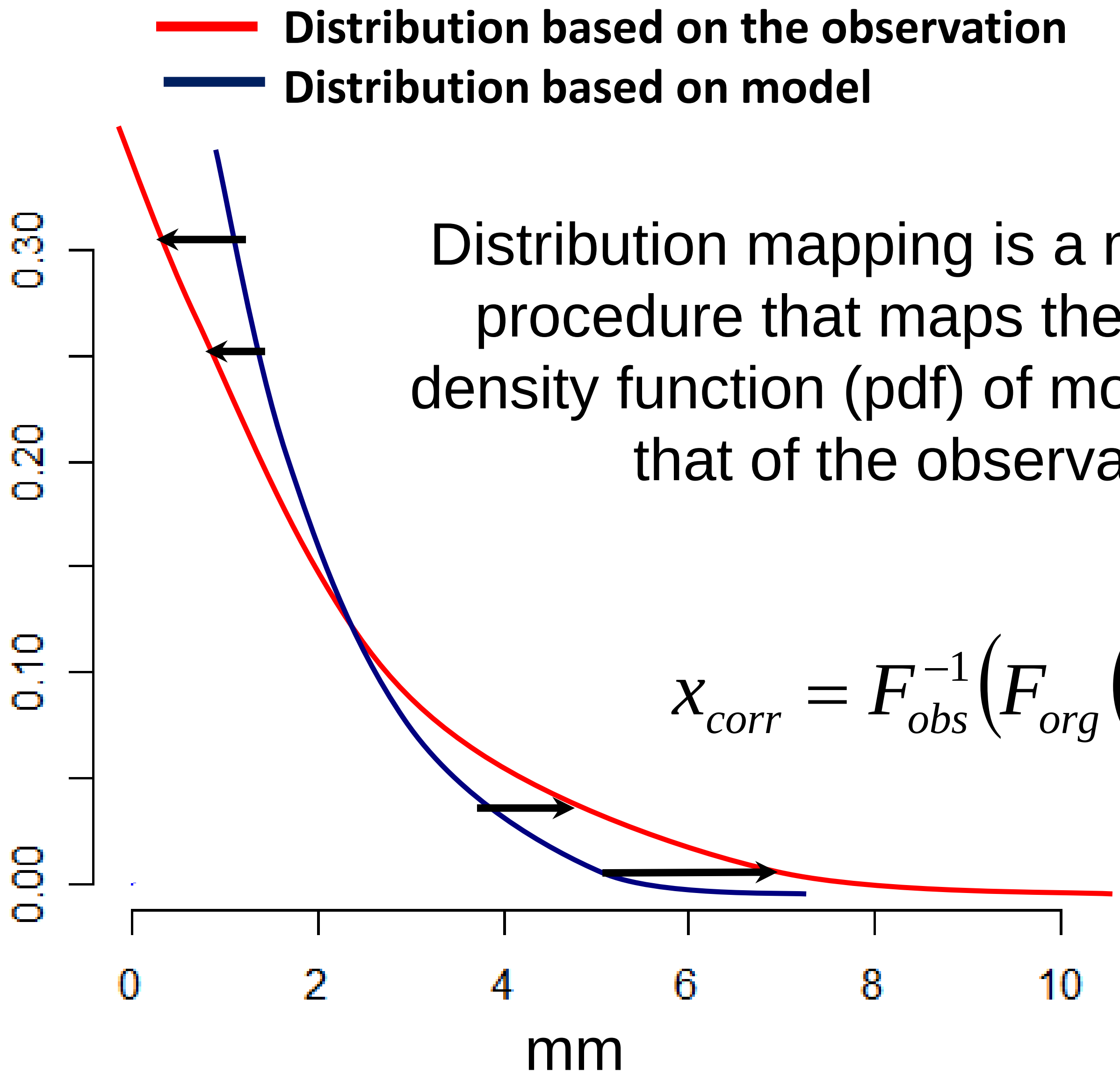
NCCS report no. 1/2014

Evaluation of distribution mapping based bias correction methods

Asgeir Sorteberg[1,2], Ingjerd Haddeland[3], Jan Erik Haugen[4], Stefan Sobolowski[5, 2] and Wai K. Wong[3]

Method	Description	Comments	
Delta change approach	RCM-simulated future change signals (anomalies) superimposed on observational time series	Same correction factor for all events	Freq. and intensities not corr. separately
Linear scaling	Adjusts RCM time series with correction values based on the differences between mean observed values and RCM simulation		
Local intensity scaling	combines a precipitation threshold with the linear scaling		Freq. and intensities corr. separately
Power trans-formation	non-linear correction. Combines the correction of the coefficient of variation (CV) with linear scaling.	Events adjusted non-linearly.	Freq. and intensities partly corr.
Distribution mapping	Matches selected distribution functions of observations and RCM-simulated climate values, plus a precipitation threshold	Events adjusted non-linearly	Freq. and intensities corr. separately

Distribution mapping



Distribution mapping is a mathematical procedure that maps the probability density function (pdf) of model data onto that of the observations

$$x_{corr} = F_{obs}^{-1}\left(F_{org}\left(x_{org}\right)\right)$$

Distribution mapping evaluation

- ✓ Parameter: Precipitation
- ✓ 6 different bias corrections based on different varieties of distribution mapping
- ✓ Daily corrections
- ✓ Evaluated 3 point locations (Bergen, Oslo, Tromsø)
- ✓ 39 validation measures (24 for current climate 15 for climate change)
- ✓ Mean absolute error of monthly averages used as validation criteria

$$MAE = abs\left(\frac{x_{corr} - x_{obs}}{x_{obs}}\right)$$

How did we validate the corrections?

Ability to reproduce point measures of today's climate

Ability to reproduce the original climate change signal.

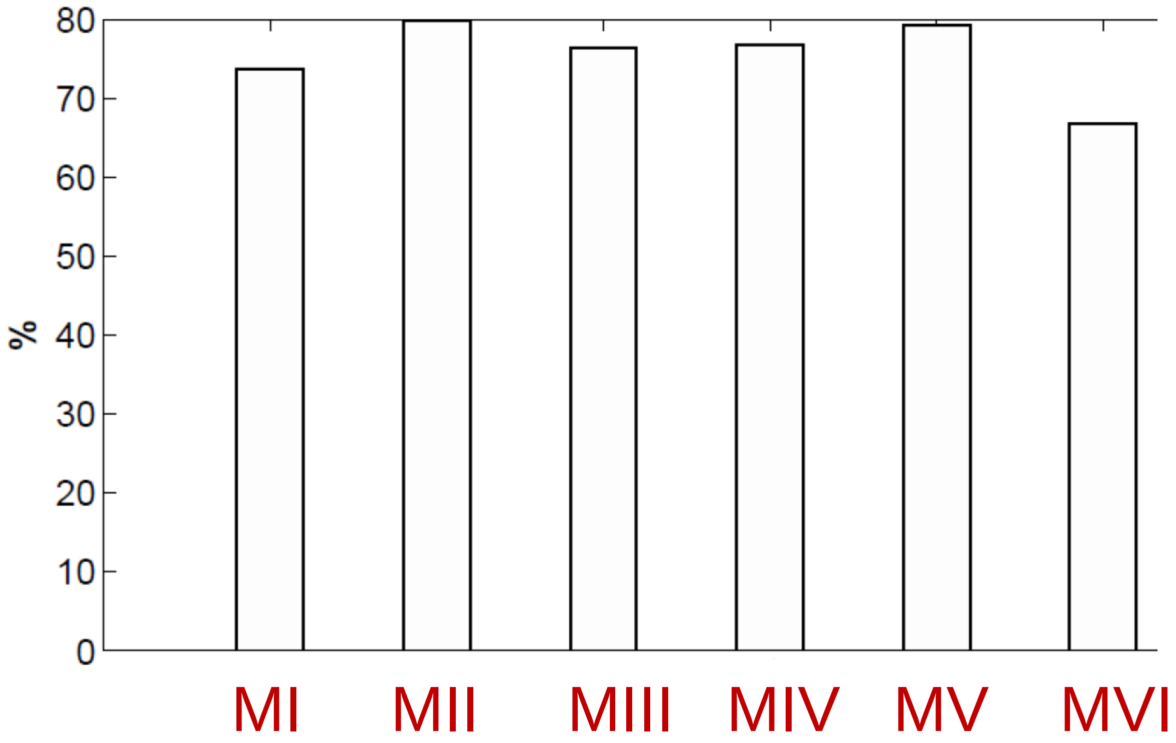
Validation measures

Ability to reproduce today's climate

Wet day frequency (days)	Mean length of wet spell (days)	Maximum length of wet spell (days)	Mean length of dry spell (days)	Maximum length of dry spell (days)
Mean 1-day precipitation (mm)	Mean 1-day intensity on wet days (mm)	25 percentile 1-day intensity on wet days (mm)	75 percentile 1-day intensity on wet days (mm)	95 percentile 1-day intensity on wet days (mm)
99 percentile 1-day intensity on wet days (mm)	99.5 percentile 1-day intensity on wet days (mm)	Maximum 1-day precipitation (mm)	Variance 1-day precipitation (mm^2)	1-day lagged autocorrelation in 1-day precipitation
Accum. 10-day precipitation (mm)	25 percentile 10-day accum. precipitation (mm)	75 percentile 10-day accum. precipitation (mm)	95 percentile 10-day accum. precipitation (mm)	99 percentile 10-day accum. precipitation (mm)
99.5 percentile 10-day accum. precipitation (mm)	Maximum 10-day accum. precipitation (mm)	Variance 10-day precipitation (mm^2)	5-day lagged autocorrelation in 10-day accum. precipitation	

Did the 24 validation measures improved with bias correction?

Percentage of measures that improved



75-80% of the validation measures improved with BC

20-25% of the validation measures became worse ….

What is the typical error after correction?

Validation measure	MAE MI
Wet day frequency	±0.4%
Mean 1-day precip intensity	±0.5%
99.5 precentile 1-day precip intensity	±16.9%
1-day variance	±18.5%
Mean dry spell length	±10.1%
Mean wet spell length	±10.9%
99.5 precentile 10-day precip	±19.0%
Mean over all 24 measures	**±13.4%**

Expect the monthly corrected values to be 10-16% away from the observed (sometimes underestimations and sometimes overestimations).

Ability to reproduce the original climate change signal

What should be reproduced?

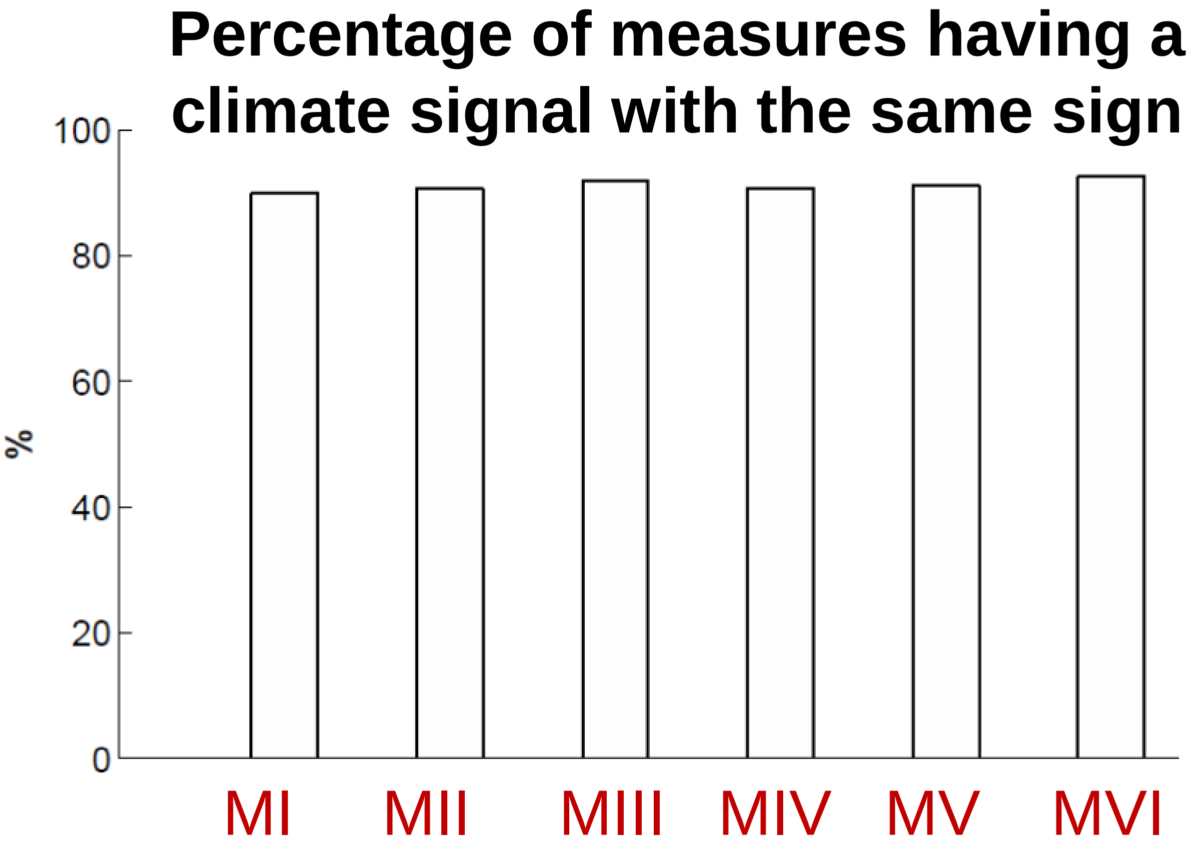
The relative changes

Validation measures

Ability to reproduce the original climate change signal.

1-day mean precipitation (%)	1-day mean intensity on wet days (%)	25 percentile 1-day intensity on wet days (%)	75 percentile 1-day intensity on wet days (%)	95 percentile 1-day intensity on wet days (%)
99 percentile 1-day intensity on wet days (%)	99.5 percentile 1-day intensity on wet days (%)	max 1-day intensity on wet days (%)	10-day mean accum. precipitation (%)	25 percentile 10-day accum. Precipitation (%)
75 percentile 10-day accum. Precipitation (%)	95 percentile 10-day accum. precipitation (%)	99 percentile 10-day accum. pecipitation (%)	99.5 percentile 10-day accum. precipitation (%)	max 10-day accum. precipitation (%)

Same sign after correction



9 of 10 climate change signals have the same sign after the correction

we should expect 1 of 12 months to change sign after the bias correction is performed.

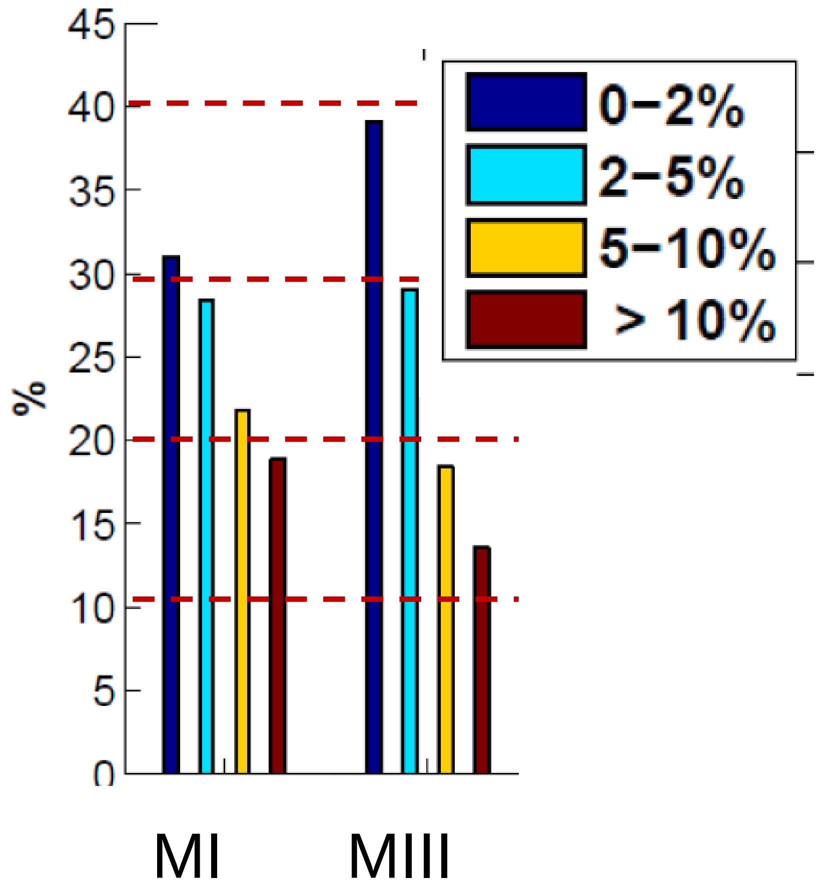
How well do the correction conserve the climate signal?

Validation measure	Org	MI
Mean 1-day precip	10%	±4.0%
25 precentile 1-day precip intensity	13.5%	±16.4%
99.5 precentile 1-day precip intensity	15.1%	±6.6%
Mean over all measures	**14.6%**	**±6.5%**

The average deviation of the monthly change from the true signal after correction was an artificial climate change signal in the order of 1/3 to 1/2 of the original signal.

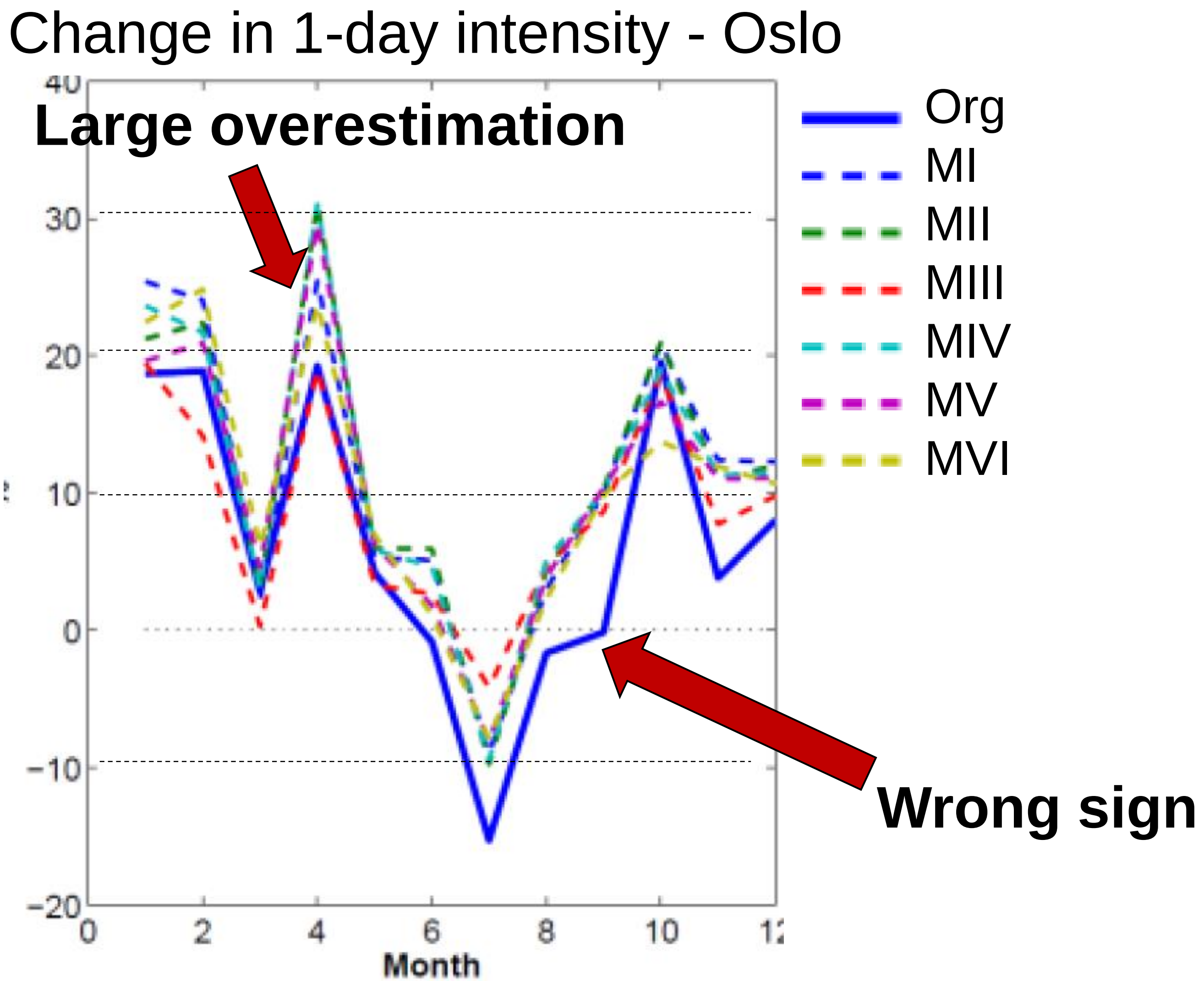
How well do the correction conserve the climate signal?

Percentage deviation from the original signal

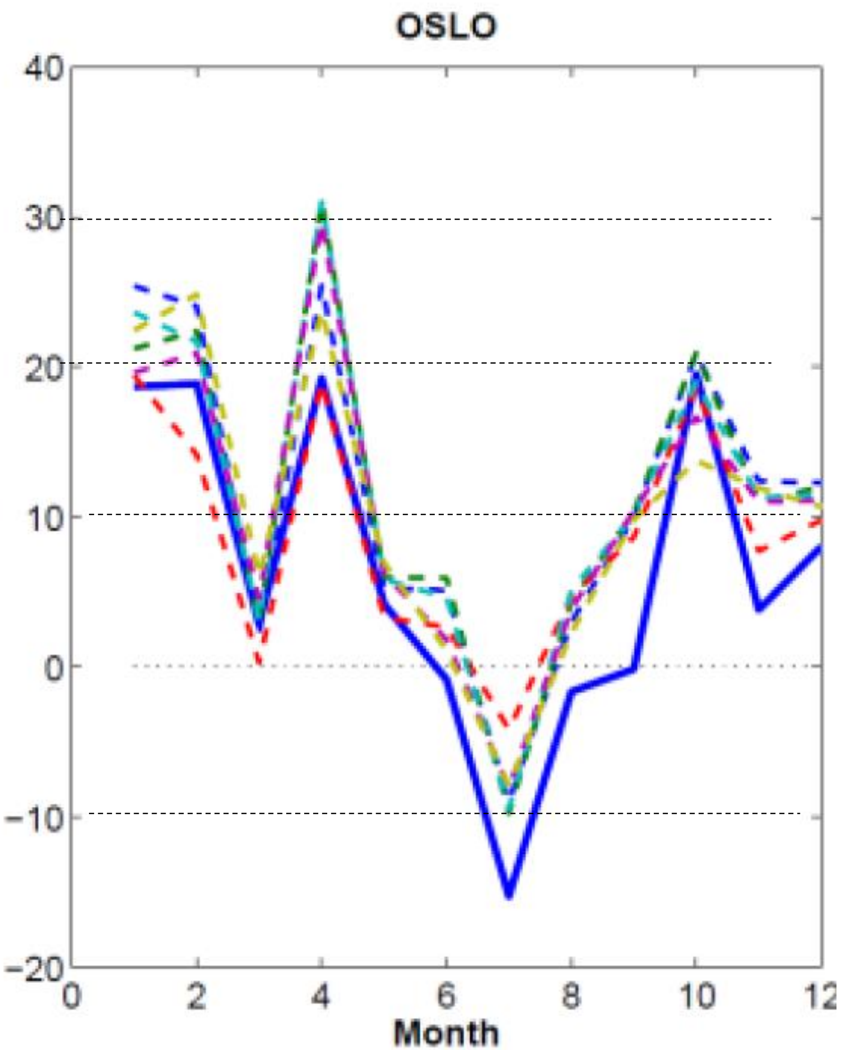


- ✓ 28-39% (approx. 4 to 5 of 12 months) within +-2% of the original climate change signal
- ✓ 13-19% (approx. 1 to 2 of 12 months). of the climate change signals deviating more than 10% from the original climate change signal

How well do the correction conserve the climate signal?



Are there systematic changes in the climate change signal?



Positive changes overestimated
Negative underestimated …

Avg. positive changes ‘
17.1% → 19.4%
Avg. negative changes
-9.3% → -6.9%

Are there systematic changes in the climate change signal?

Positive changes overestimated, negative underestimated

Why?

- ✓ Low and high amounts are corrected differently. If the distribution between low and high amounts change in the future this will lead to changes in the bias corrected climate signal

- ✓ Wet day frequency corrections leads to changes in the bias corrected climate signal if corrected number of wet days are different in historical and future simulations

Are there systematic changes in the climate change signal?

Example with 5 values:

Do a wet frequency correction so values less than 1 are set to 0 to get the correct frequency.

Data	Values	Change
Original data		
Historical	[0.95 3.0 5.0 8.0 11.0]	**+10%**
Future	[1.05 3.3 5.5 8.8 12.1]	
wet frequency corrected data		
Historical	[0.00 3.0 5.0 8.0 11.0]	**+13.9%**
Future	[1.05 3.3 5.5 8.8 12.1]	

Conclusion

Bias corrections are extremely difficult .....

Has to be tailor made for the intended purpose

The loss of physical consistency between parameters may introduce unphysical impacts

Watch out for artificial climate signals ....

What can we do?

Provide all results of any impact study for both *bias corrected AND non-corrected input*, for the historical future simulations

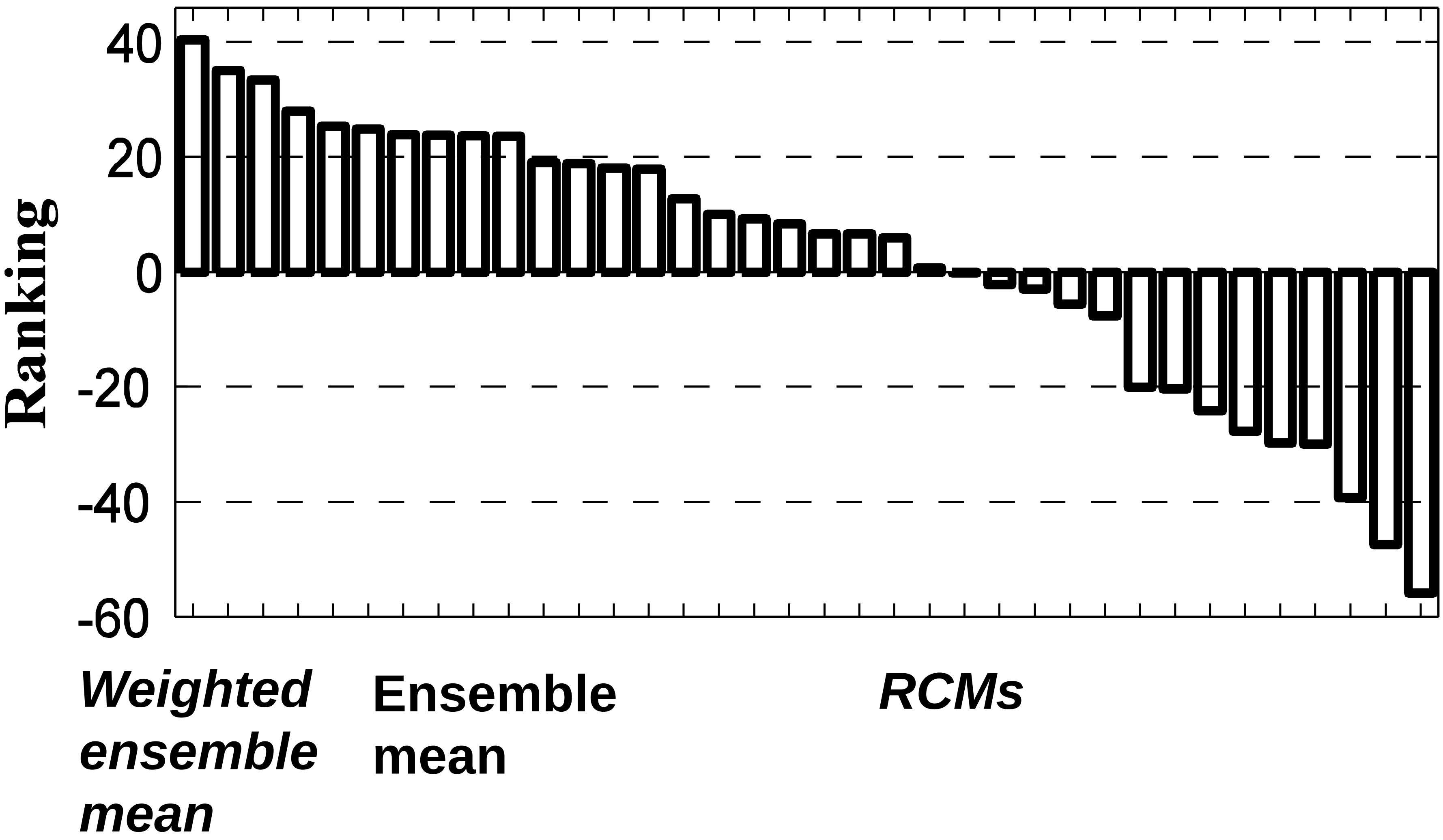
Do a proper evaluation of the bias correction method

Improve the bias correction methods

Use multi-model ensembles

multi-model ensembles has been shown to often outperform 'best-model' approaches

Multi-model ensembles



What is the problem?

Misrepresentation of event size

Is model data representing the scales of the events correctly?

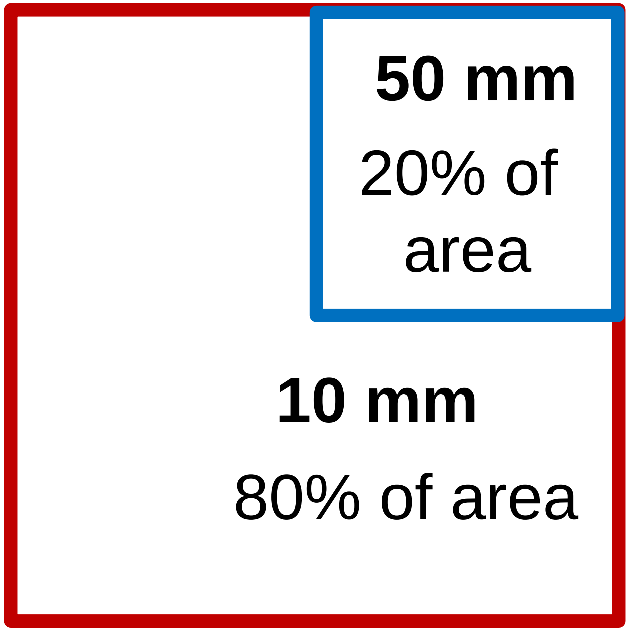
Misrepresentation of event size

Example

overestimations in the extent of convective cells
(minimum one grid cell),

Model grid

Convective region

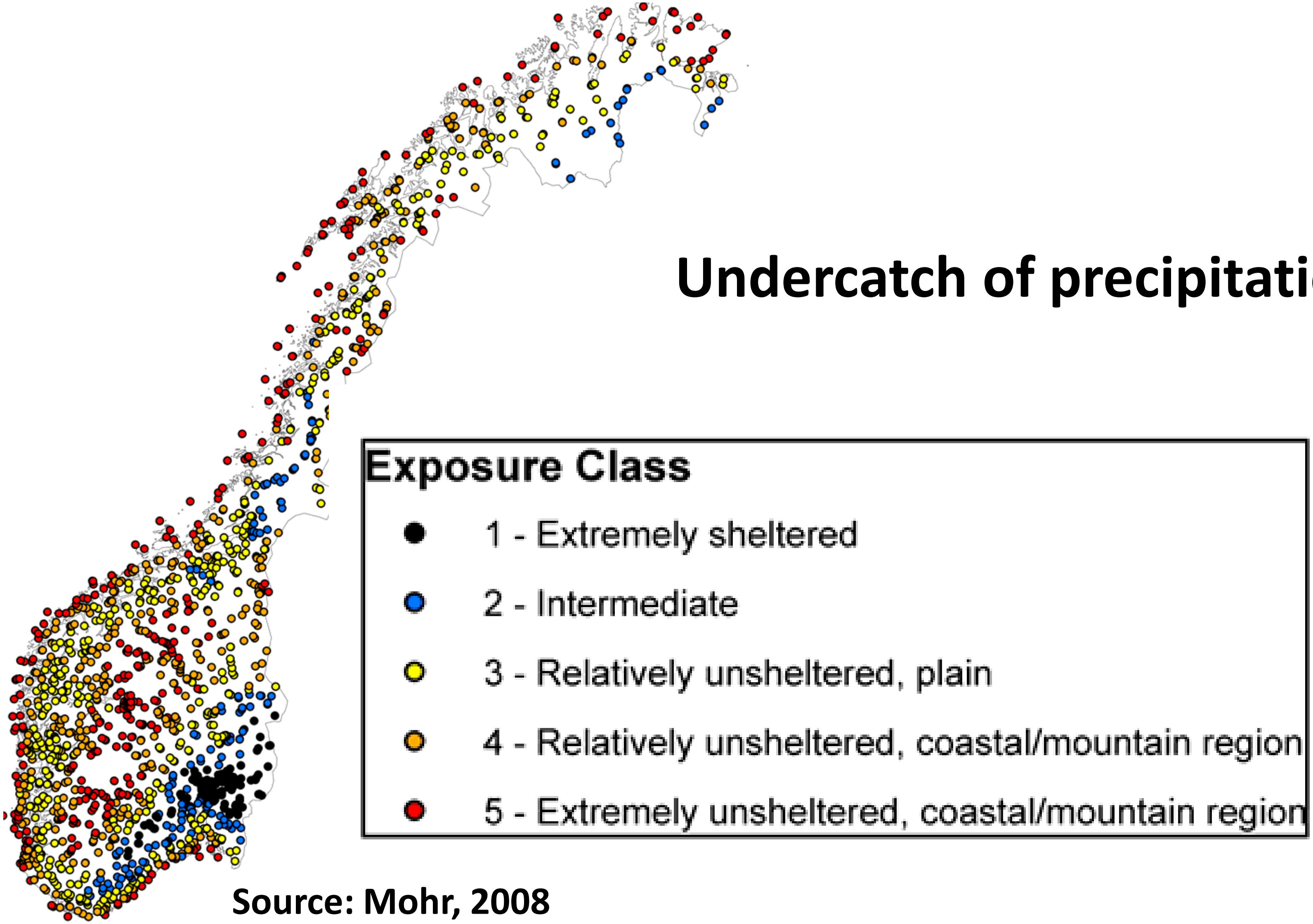


$$P_{\mathrm{i}} = \frac{P_{conv} \cdot A_{conv} + P_{non-conv} \cdot A_{non-conv}}{\left(A_{conv} + A_{non-conv}\right)}$$

$$P_{\mathrm{i}} = 18\mathrm{mm}$$

Suitability of observationally based estimates

Undercatch of precipitation



Source: Mohr, 2008

Multi-model ensembles vs ‘best-model’ approach

“*when many RCMs are used in a coordinated way, ... the ensemble mean nearly always is in better agreement with observed climatology than any individual model*.” (Jacob et al., 2007)

multi-model ensembles often outperform a ‘best-model’ approach, if the single-model ensembles are overconfident. The reason is that multi-model combination reduces overconfidence, i.e. ensemble spread is widened while average ensemble-mean error is reduced.

meaning that ensemble spread is too narrow

Multi-model ensembles

The control simulation of the different models was validated against daily data from the observational network. Only stations which contained enough data to make annual means for at least 25 of the 30 years in the period 1961-1990 was selected (a total of 417 stations).

The bias in seasonal precipitation for each region and each season was calculated and the performance of the models ranked from 1 to 33. The total ranking was taken as the mean rank.

Figure .. show the relative ranking of the different models compared to the mean of the models.