Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
**Federal Office of Meteorology and Climatology MeteoSwiss**

# Multimodels on seasonal to multidecadal time-scales

## Potential and limitations

Workshop on representing model uncertainty and error in numerical weather and climate prediction models
ECMWF, Reading, UK, 21 June 2011

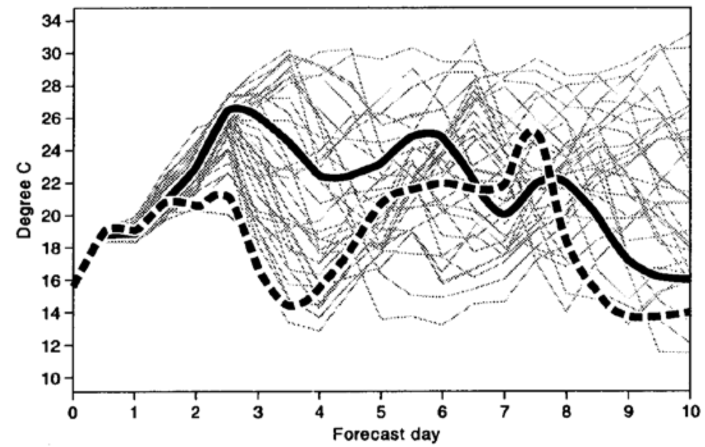
Andreas Weigel

NCCR CLIMATE
Swiss Climate Research

# Motivation

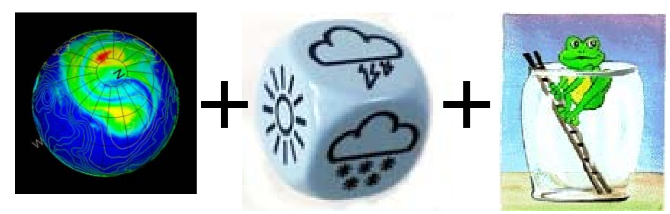
Two kinds of **uncertainties**:

Initialization
Model error




**IC Ensembles**

**Multimodels**



Success of multimodel combination demonstrated in many studies *(e.g. Krishnamurti et al. 1999, Palmer et al. 2004, Weigel et al. 2008)*

*Part 1*

*Multimodels in weather and seasonal forecasting*

# Why it works: deterministic perspective

Let $m_1 \ldots m_n$ be the forecasts stemming from $n$ models,
$M$ be the multi-model mean,
$O$ be the verifying observation:

$$\frac{1}{n}\sum\|m_i - O\|^2 \;=\; \frac{1}{n}\sum\|(m_i - M) - (O - M)\|^2$$

**0**

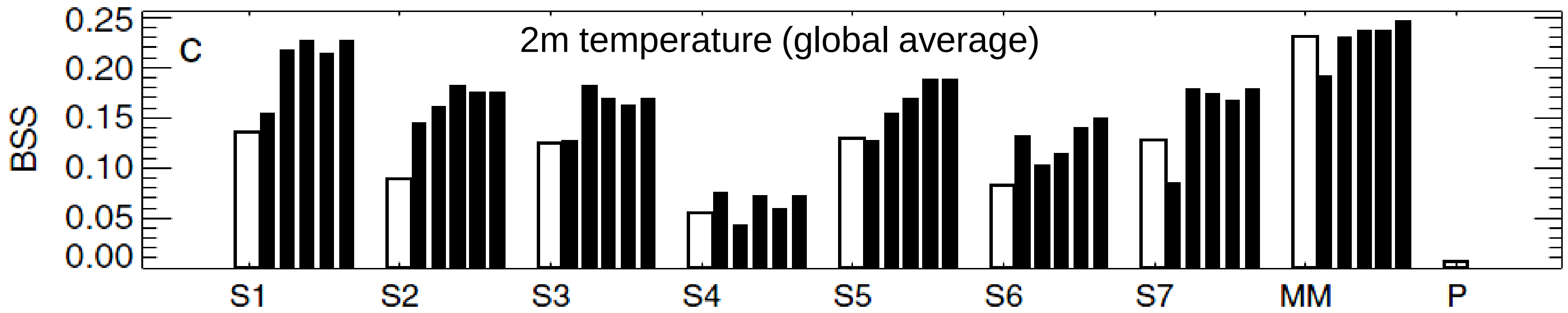
$$=\; \frac{1}{n}\sum\|m_i - M\|^2 - \frac{2}{n}\sum(m_i - M)\cdot(O - M) + \|O - M\|^2$$

$$\boxed{\frac{1}{n}\sum\|m_i - O\|^2 = \frac{1}{n}\sum\|m_i - M\|^2 + \|O - M\|^2}$$

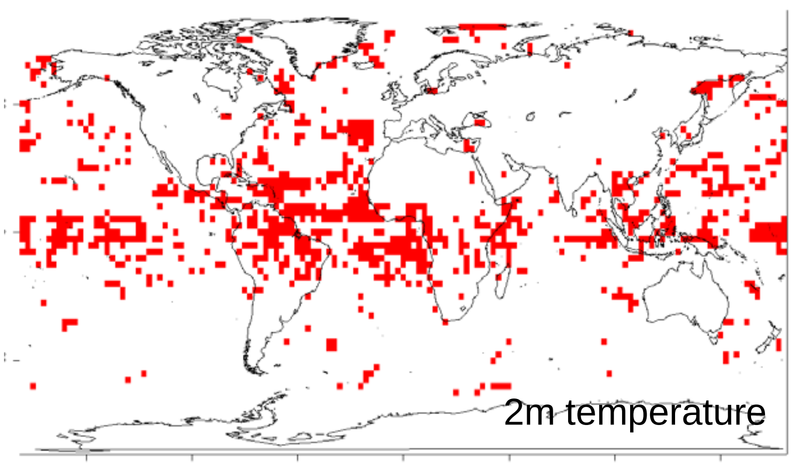
(e.g. *Annan and Hargreaves 2011*)

# Why it works: probabilistic perspective

7 DEMETER models: Multi-model better than any single model




*Doblas-Reyes et al. 2005*

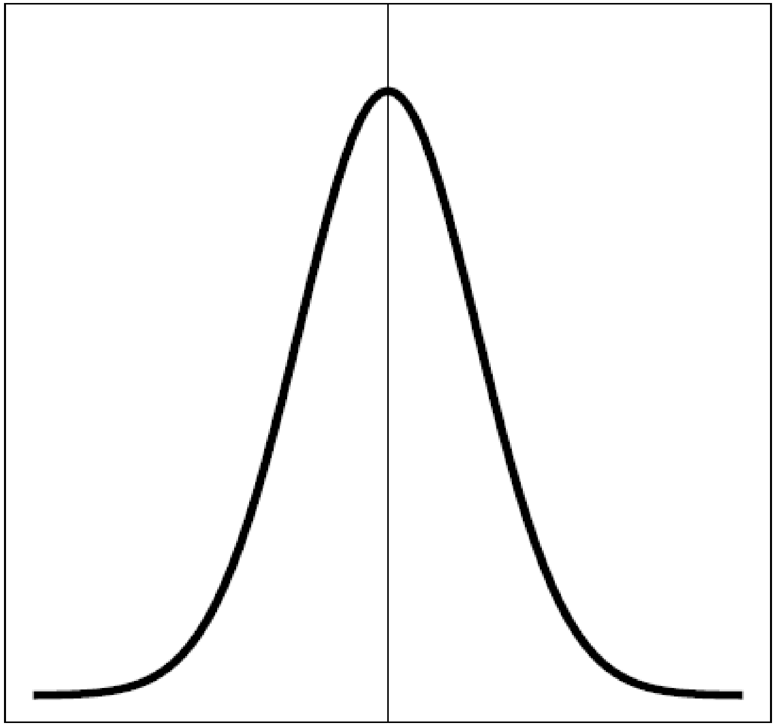



*Weigel et al. 2008*

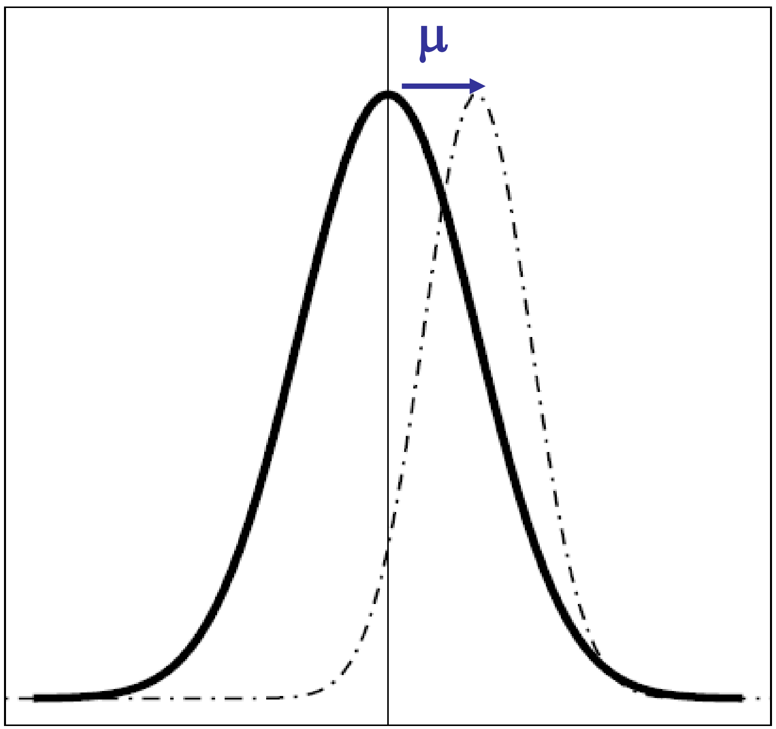
2 DEMETER models

Red points:
Multi-model locally better than any single model

# A conceptual view



# A conceptual view




future observation

$$x = \mu + \epsilon_x$$

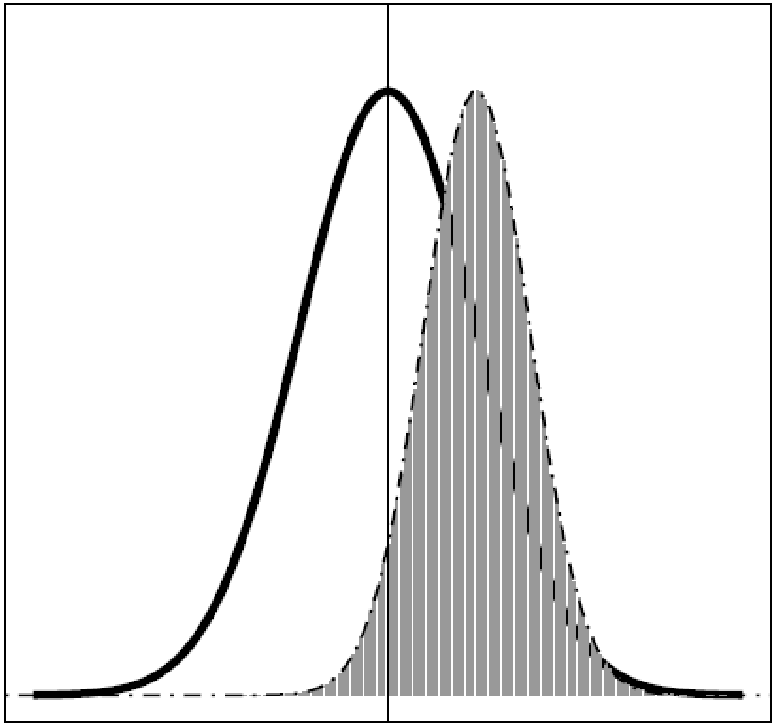
predictable signal

unpredictable noise



NCCR CLIMATE
Swiss Climate Research

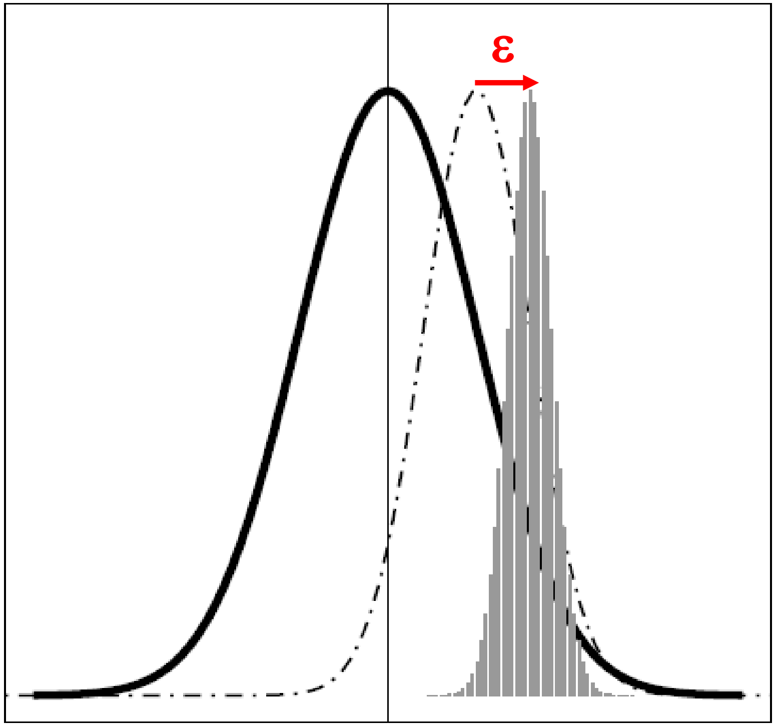
# A conceptual view



$$x = \mu + \epsilon_x$$

$$\begin{pmatrix} f_1 \\ f_2 \\ \cdots \\ f_M \end{pmatrix} = \mu + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdots \\ \epsilon_M \end{pmatrix}$$

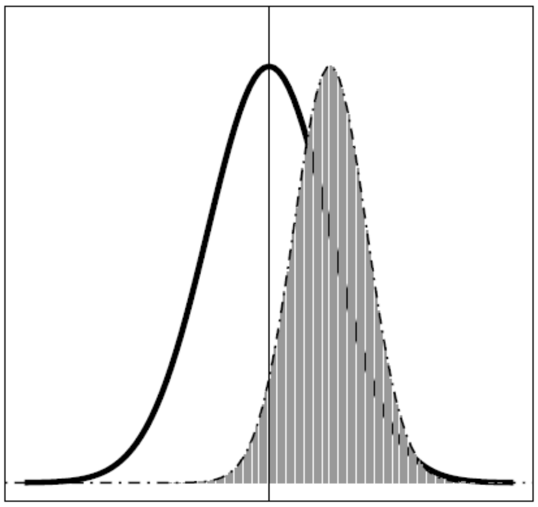
# A conceptual view



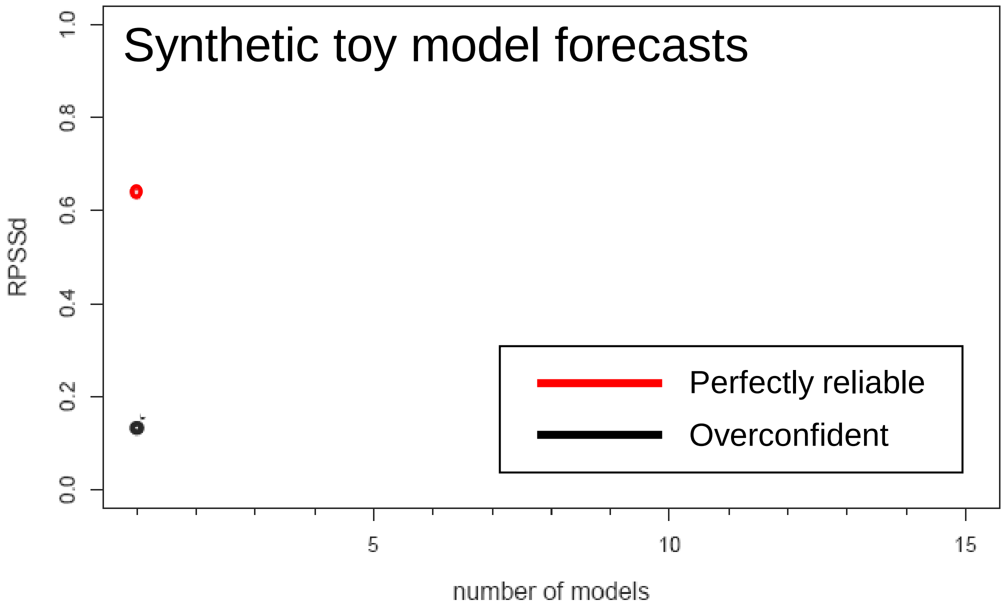

$$x = \mu + \epsilon_x$$

$$\begin{pmatrix} f_1 \\ f_2 \\ \dots \\ f_M \end{pmatrix} = \mu + \epsilon_\beta + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_M \end{pmatrix}$$

# Multi-model and Skill Score



*Weigel et al. 2008*

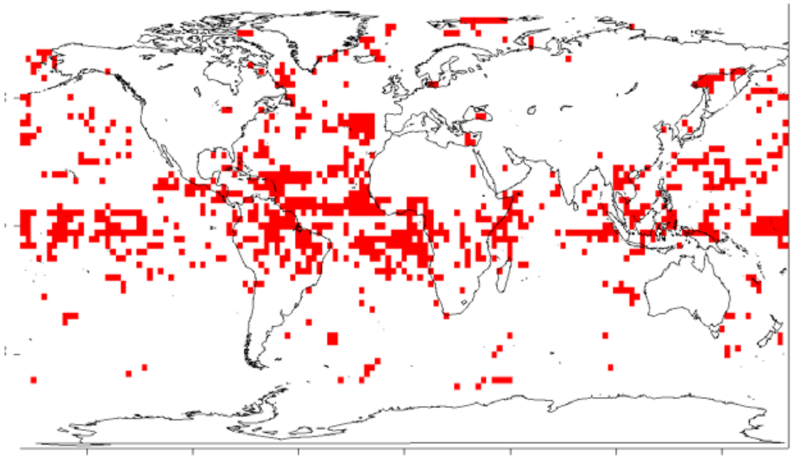



**Can multi-models outperform best single model?**

Yes → if models are overconfident
No → if models are already perfectly reliable

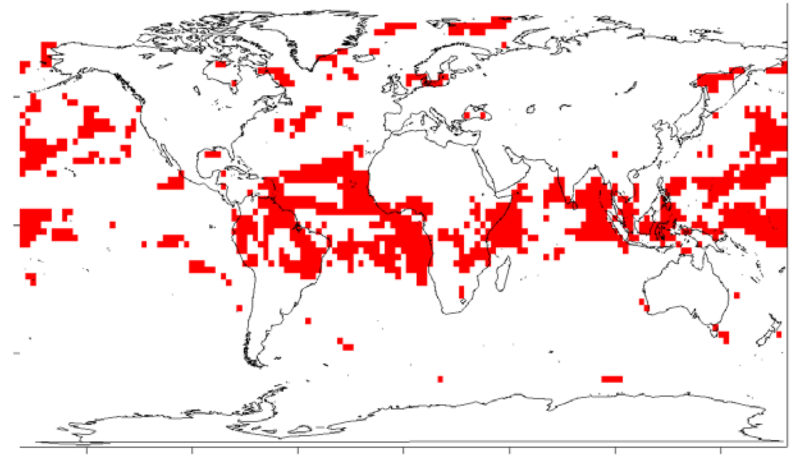
Multi-models do NOT improve physical predictability.

# Application to real data (DEMETER)



**2 DEMETER models**
**2m temperature**

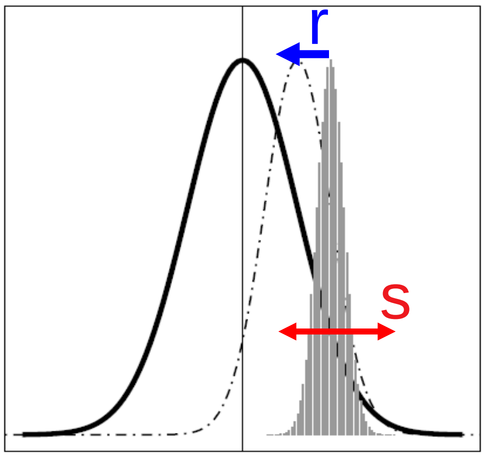
multi-model
better than
best participating model



both models
highly
overconfident

*Weigel et al. 2008*

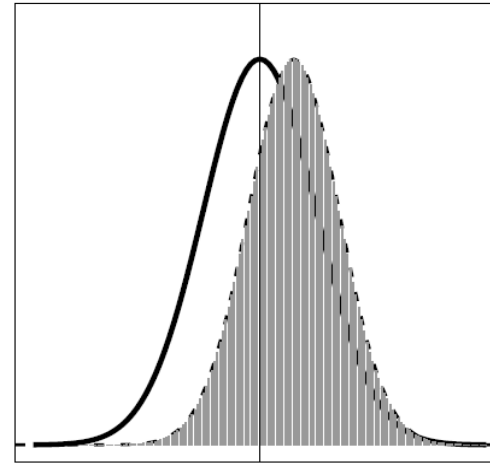
# Recalibration – an alternative strategy?




$$
\begin{aligned}
r &= \frac{\mu_{x'}}{\mu_f} \\
&= \rho_{(\mu_x, \mu_f)} \frac{\sigma_{\mu_x}}{\sigma_{\mu_f}} \\
&= \frac{cov\left(\mu_x, \mu_f\right)}{\sigma_{\mu_x}\sigma_{\mu_f}} \frac{\sigma_{\mu_x}}{\sigma_{\mu_f}} \\
&= \frac{cov\left(\mu_x + \epsilon_x, \mu_f\right)}{\sigma_{\mu_f}} \frac{1}{\sigma_{\mu_f}} \\
&= \rho_{(x, \mu_f)} \frac{\sigma_x}{\sigma_{\mu_f}}
\end{aligned}
$$

for $s$ is obtained by taking the condition that t
recasts is identical to the observed climatology,
on, together with Eq. 6, in Eq. 2 yields:

$$
\begin{aligned}
\sigma_x^2 &= \sigma_{f_i'}^2 \\
&= \sigma_{\mu_f}^2 r^2 + s^2 \sigma_{\epsilon_i}^2 \\
\Rightarrow s &= \frac{\sigma_x}{\sigma_{\epsilon_i}} \sqrt{1 - \rho_{(x, \mu_f)}^2}
\end{aligned}
$$



**Rescale** ensemble mean + **inflate** ensemble spread

$$
\begin{pmatrix} f_1 \\ f_2 \\ \dots \\ f_M \end{pmatrix} = r(\mu + \epsilon_\beta) + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_M \end{pmatrix} \cdot s
$$

*Doblas-Reyes et al. 2005*
*Johnson and Bowler 2009*
*Weigel et al. 2009*

# Recalibration versus multi-models

$$\begin{pmatrix} f_1 \\ f_2 \\ \dots \\ f_M \end{pmatrix} = r(\mu + \epsilon_\beta) + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_M \end{pmatrix} \cdot s$$

$$\begin{pmatrix} f_1 \\ f_2 \\ \dots \\ f_M \end{pmatrix} = \mu \cancel{+ \epsilon_\beta} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_M \end{pmatrix}$$

## Recalibration

Problems:
- Part of $\mu$ is destroyed (usually r<1)
- Sample size needs to be large

## Multimodel

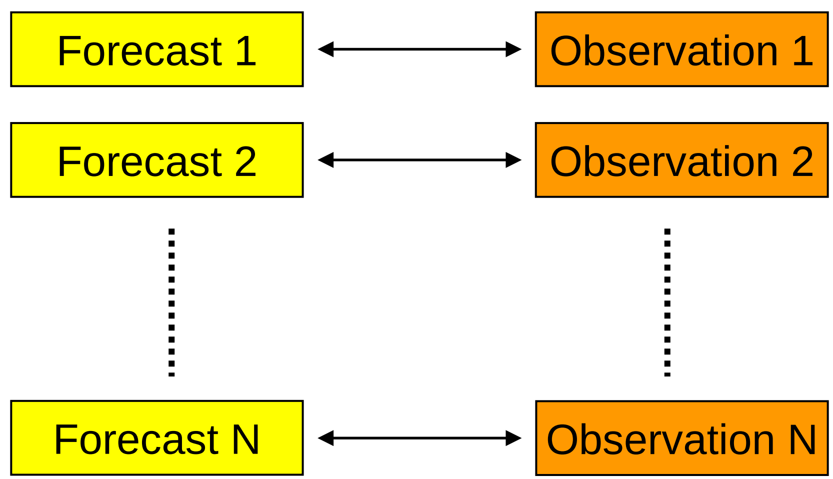
Problems:
- Usually not many models available
- Model errors often dependent
- **Models may differ in $\mu$**

Optimum results may be achieved by combining multi-models and recalibration (e.g. *Stephenson, 2005; Doblas-Reyes et al. 2005*)

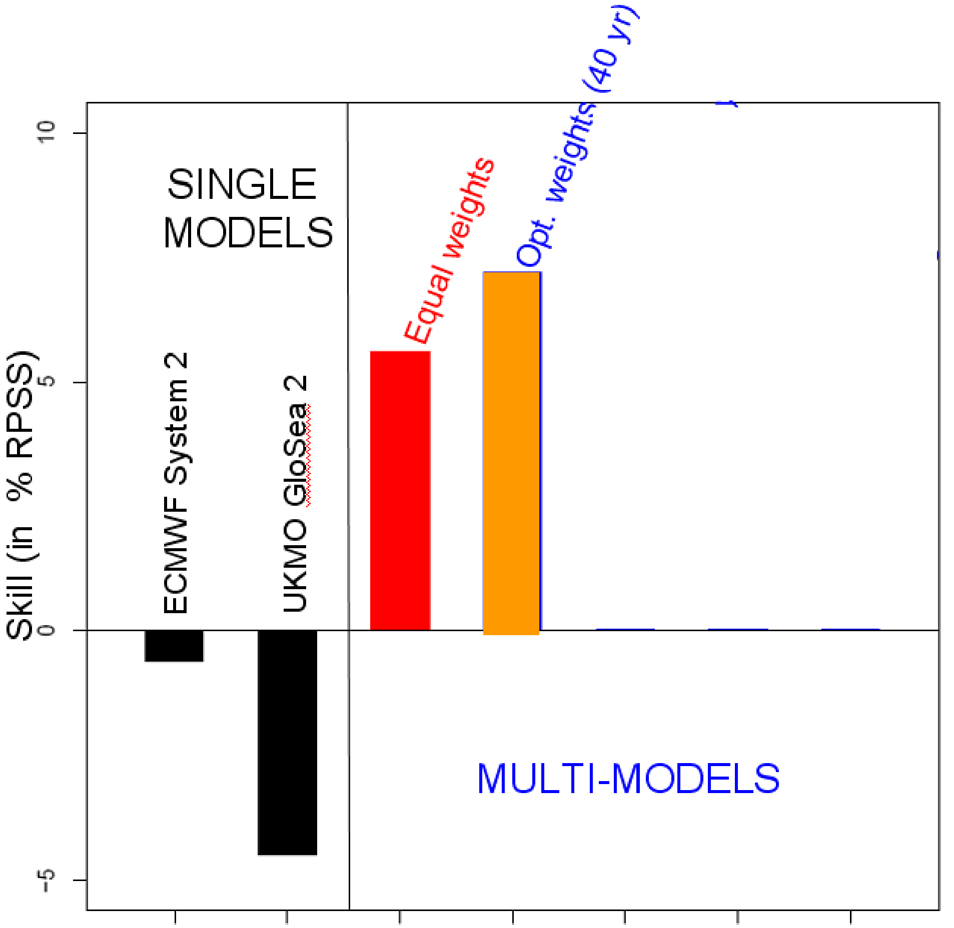
# Model weighting

- Optimization with respect to specific skill or error metrics
- Bayesian approaches with climatology as prior
- Regression approaches
- …

  *(e.g. Rajagopalan et al. 2002, Coelho et al. 2004, DelSole 2007, Raftery et al. 2005, Weigel et al. 2008, and many more)*



# Skill of weighted multi-models




Many samples are required to obtain robust weights

If weights are not robust, more information may be lost than could potentially be gained.

Verification allows to design optimum strategy

*Weigel et al. 2010*

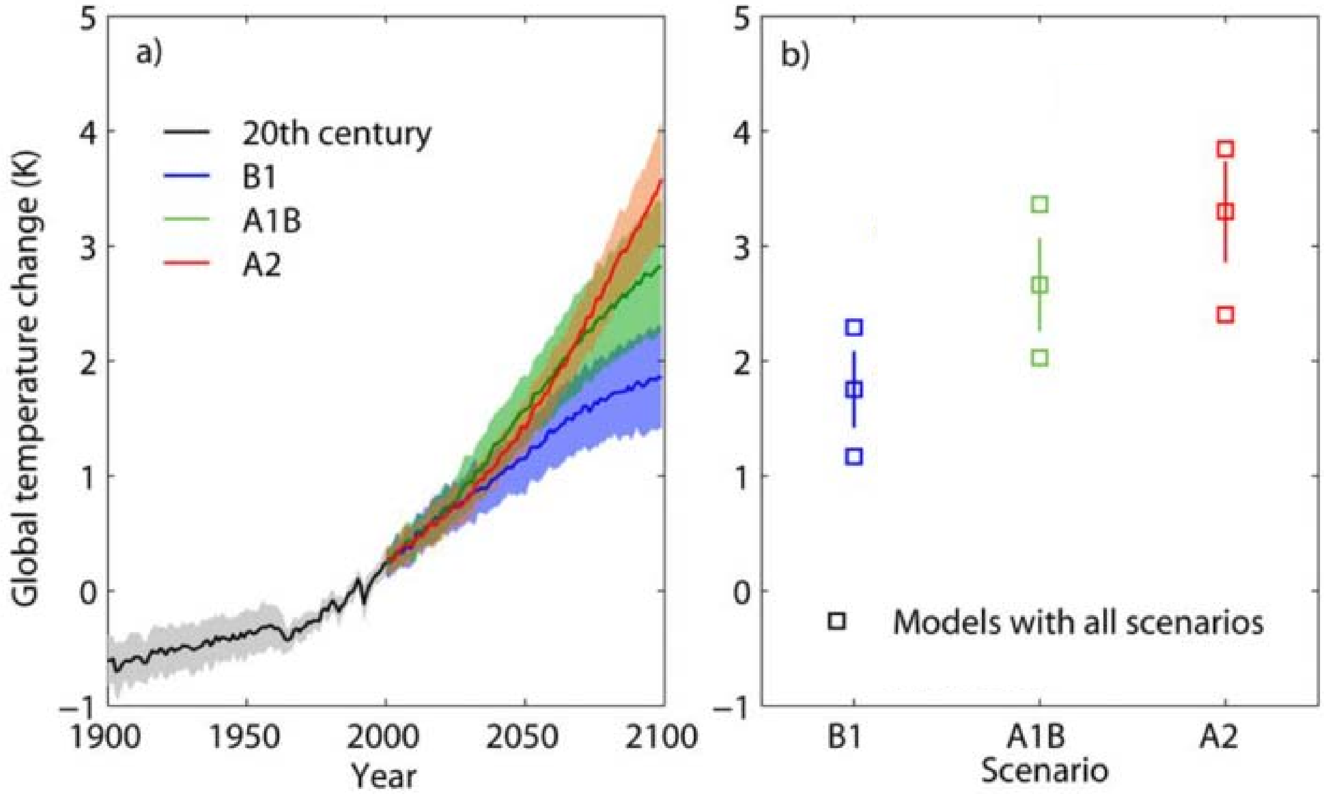


NCCR CLIMATE
Swiss Climate Research

*Part 2*

*Multimodels in multi-decadal climate change projections*

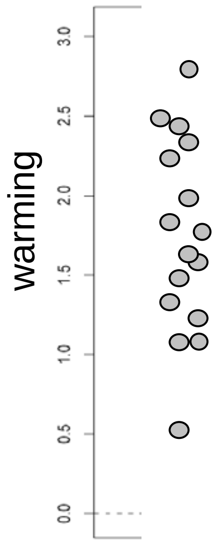
# Example IPCC AR4




*Knutti et al. 2009*

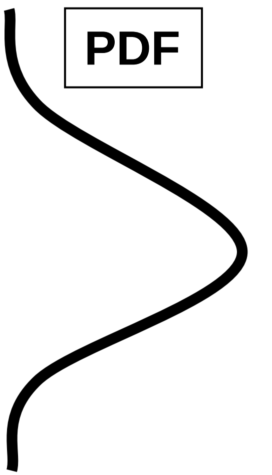
# Multimodel projections

Ensemble of model projections




**?**

Quantifying probabilities

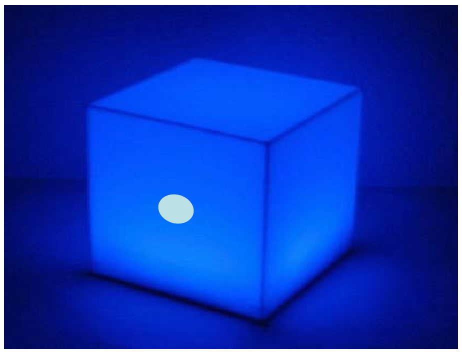


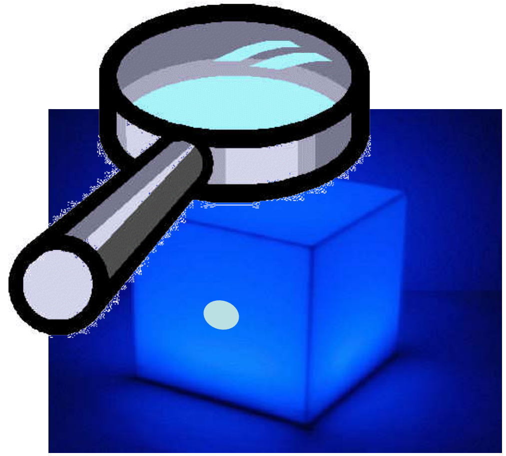

Key questions:

- Credibility of individual model projections?
- Statistical interpretation of ensemble?

# Weighting of long-range climate projections?

What is the probability of throwing a “one” with this dice?





Short-term forecasts:
Many experiments

Climate projections:
One experiment

Make inference on probability by analy-zing characteristics

Need to pick the “right” characteristics

Extrapolation problem:
Is model performance during control period representative for scenario period?

**Model weighting is straight forward on short time-scales, but not in the context of long-range climate projections!**

# Performance based weighting

Example: RCM weighting in ENSEMBLES project

$$W_{\text{PROD}} = \prod_{i=1}^{6} f_i$$

$f_1$: Large-scale **circulation** based on weather regime classification
$f_2$: **Meso-scale** signal based on seasonal mean temp. and precip.
$f_3$: Distribution of **daily and monthly** temp. and precip.
$f_4$: **Extremes** in terms of re-occurrence periods for temp. and precip.
$f_5$: Long-term **trends** in temperature
$f_6$: **Annual cycle** in temperature and precipitation

$n_i$: **Scaling** exponent

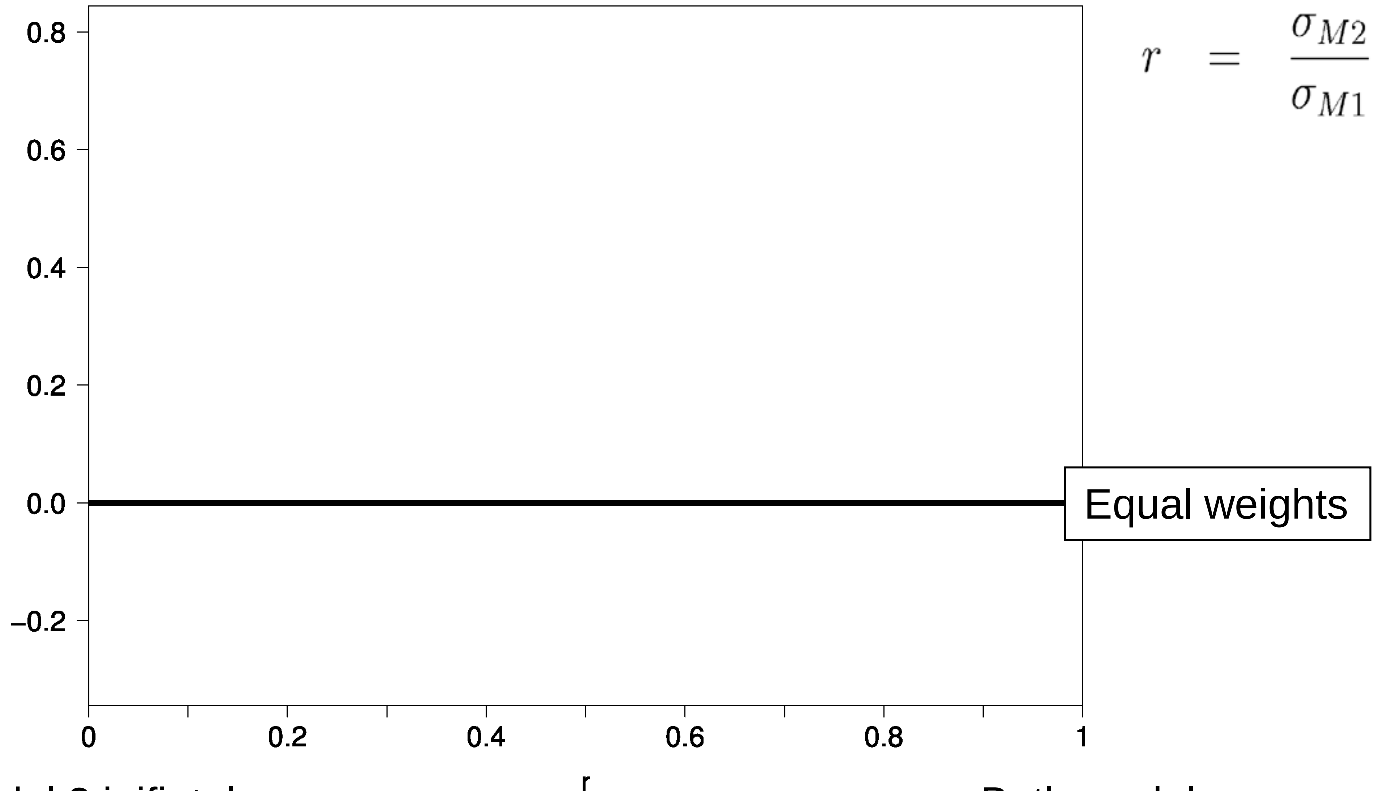
*Christensen et al. 2010*

# Effects of weighting

Increase of error (MSE)

Benchmark

Decrease of error (MSE)




$$r = \frac{\sigma_{M2}}{\sigma_{M1}}$$

Model 2 inifintely better than Model 1

Both models have same skill

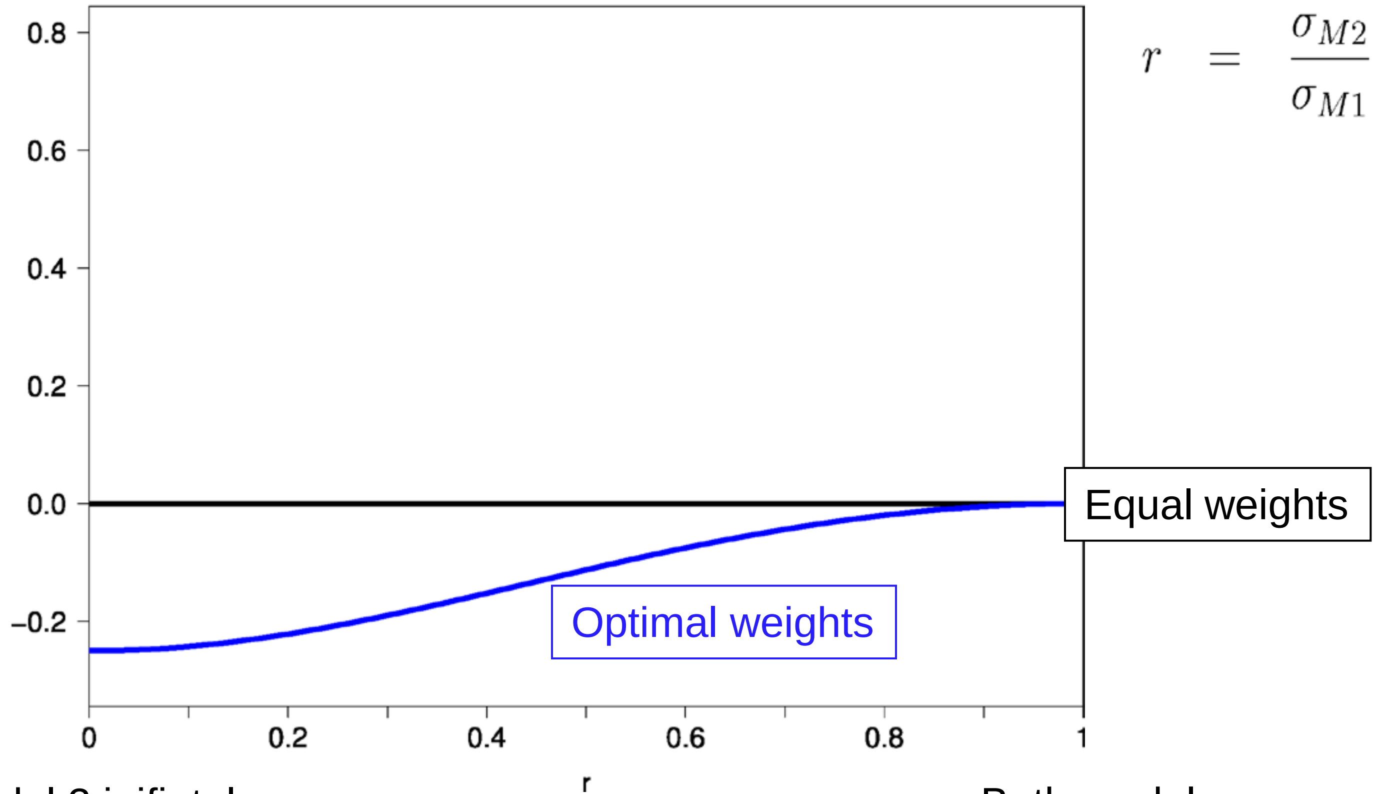
*Weigel et al. 2010*

# Effects of weighting

Increase of error (MSE)

Benchmark

Decrease of error (MSE)




$$r = \frac{\sigma_{M2}}{\sigma_{M1}}$$

Model 2 inifintely better than Model 1

Both models have same skill

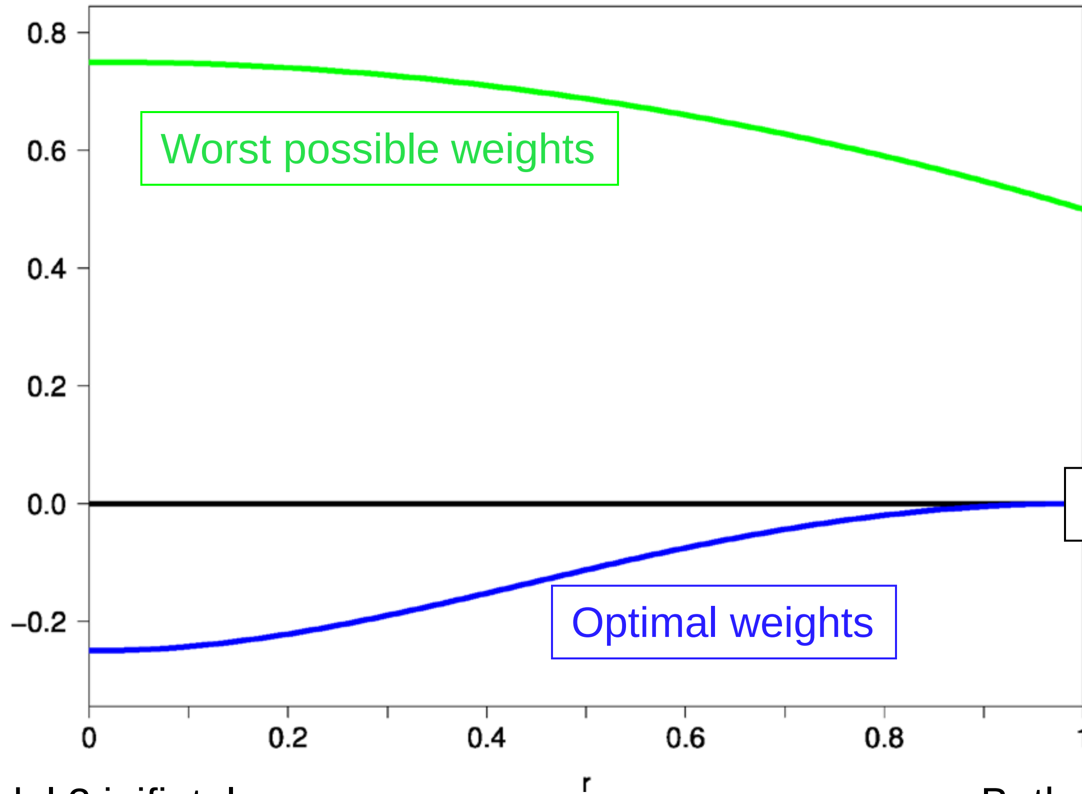
*Weigel et al. 2010*

# Effects of weighting

Increase of error (MSE)

Benchmark

Decrease of error (MSE)




Equal weights

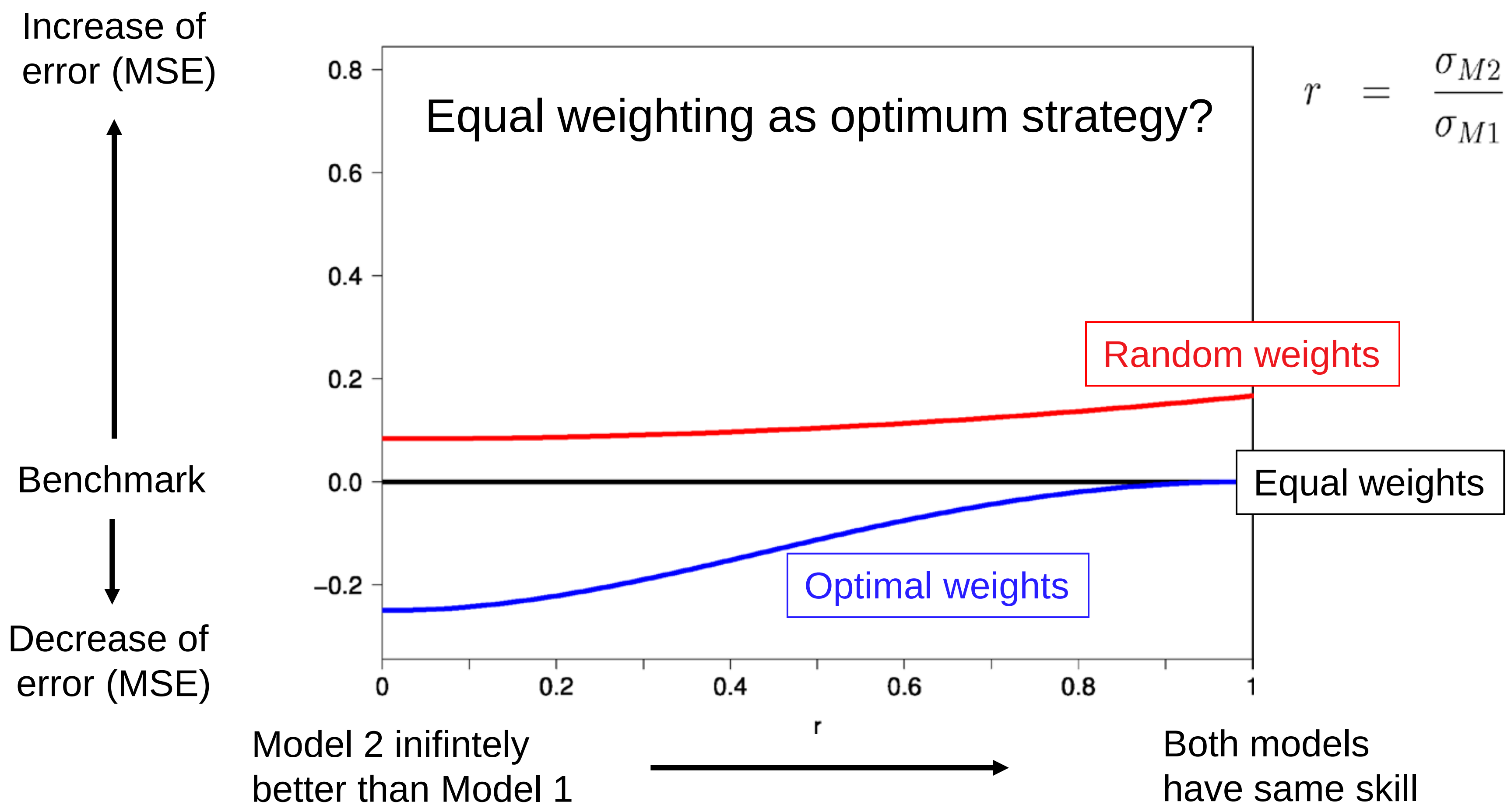
$$r = \frac{\sigma_{M2}}{\sigma_{M1}}$$

Model 2 inifintely better than Model 1

Both models have same skill

*Weigel et al. 2010*

# Effects of weighting



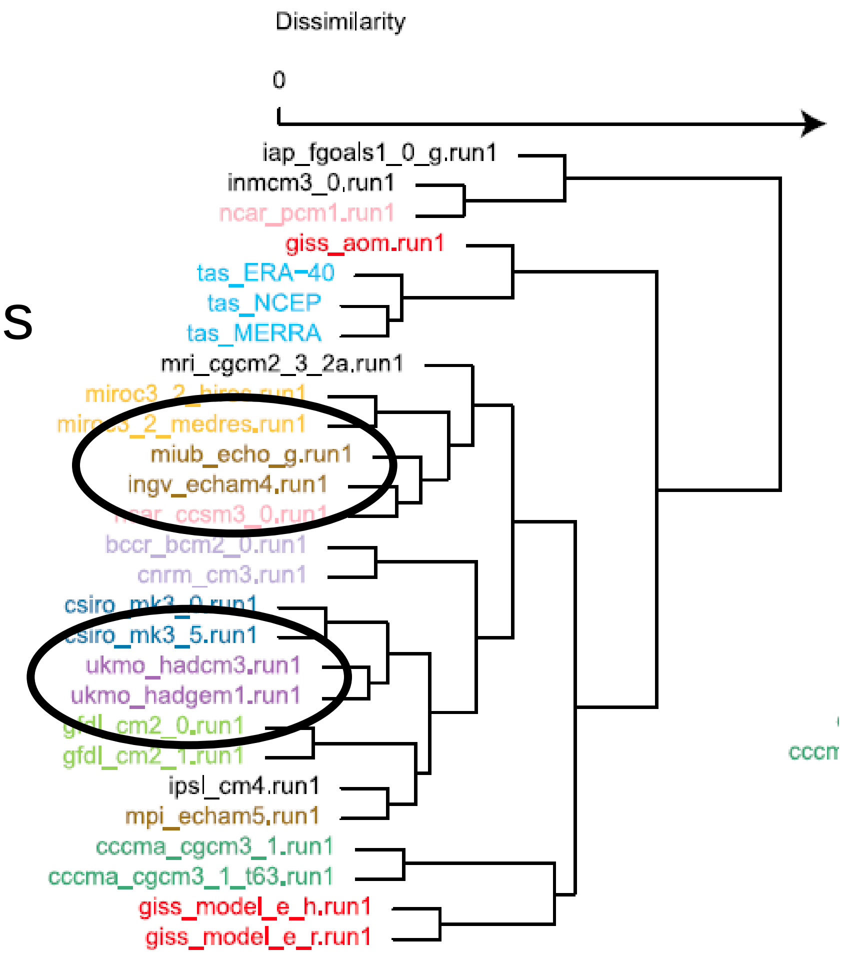

*Weigel et al. 2010*

# Structural similarities

**Surface temperature CMIP3**

“Family tree” of models

Models from same institute and models sharing version of same atmospheric model are shown in same color

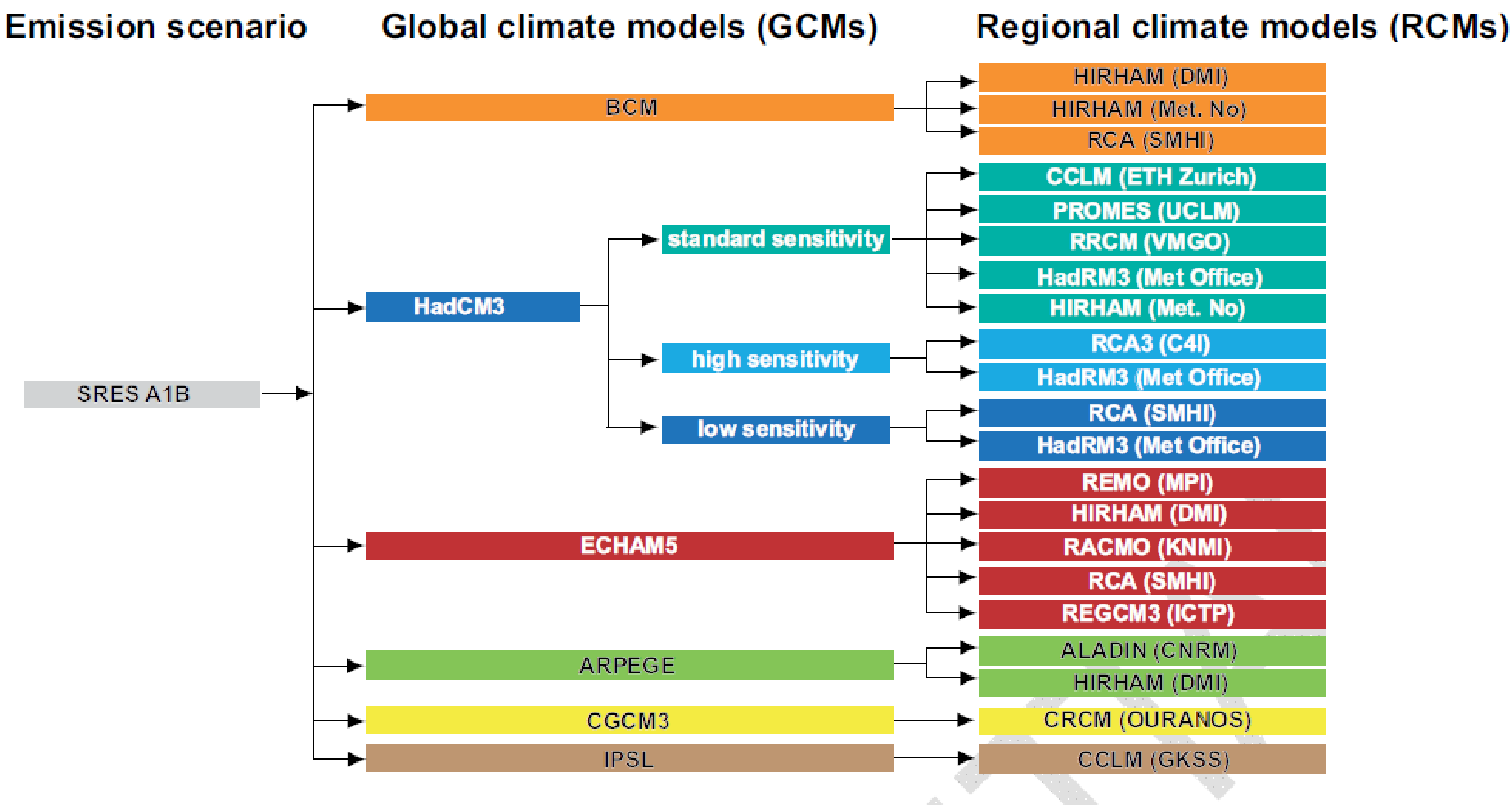



*Masson and Knutti, 2011*



NCCR CLIMATE
Swiss Climate Research

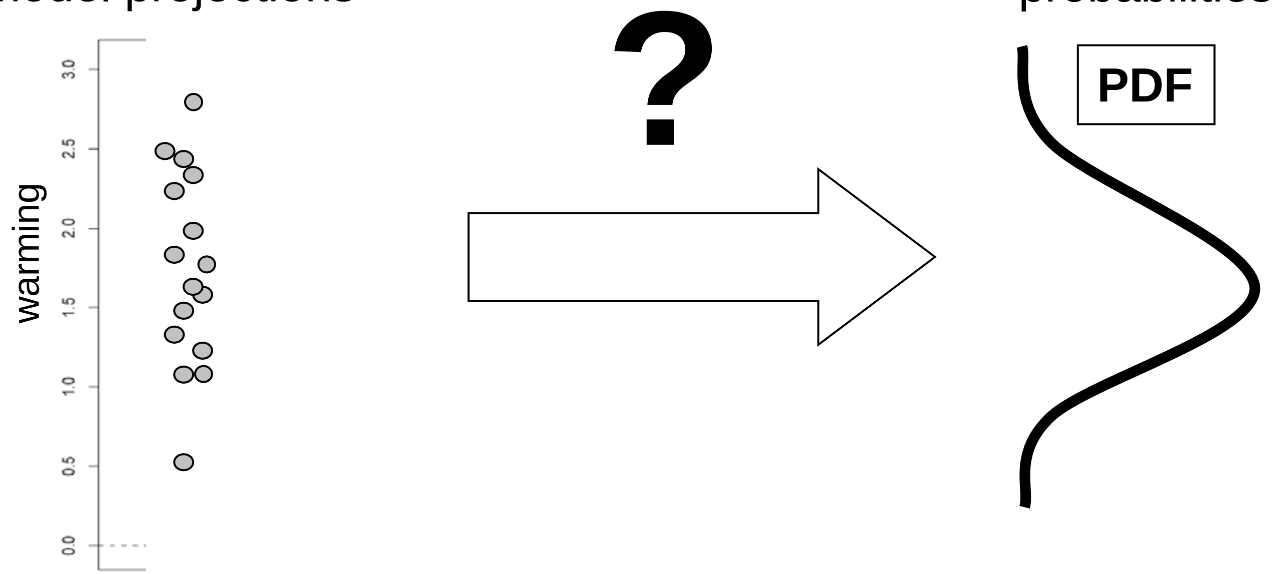
# ENSEMBLES R2TB



# Multimodel projections

Ensemble of model projections

Quantifying probabilities

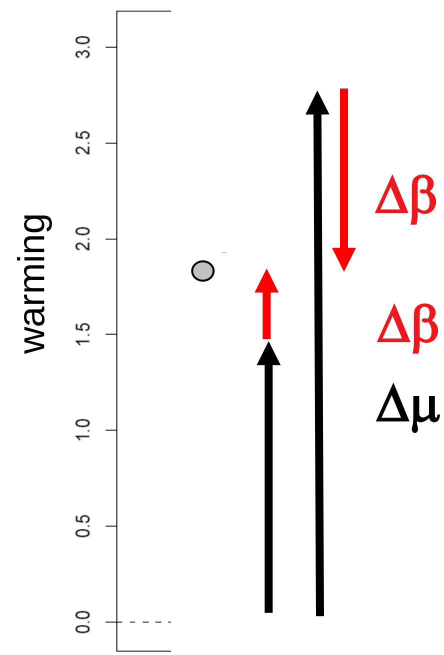



Key questions:

- Credibility of individual model projections?
- Statistical interpretation of ensemble?

# A simple Bayesian model

**Observations CONTROL** $\sim N(\mu, \ldots)$




$\mu$: Climate mean during control period

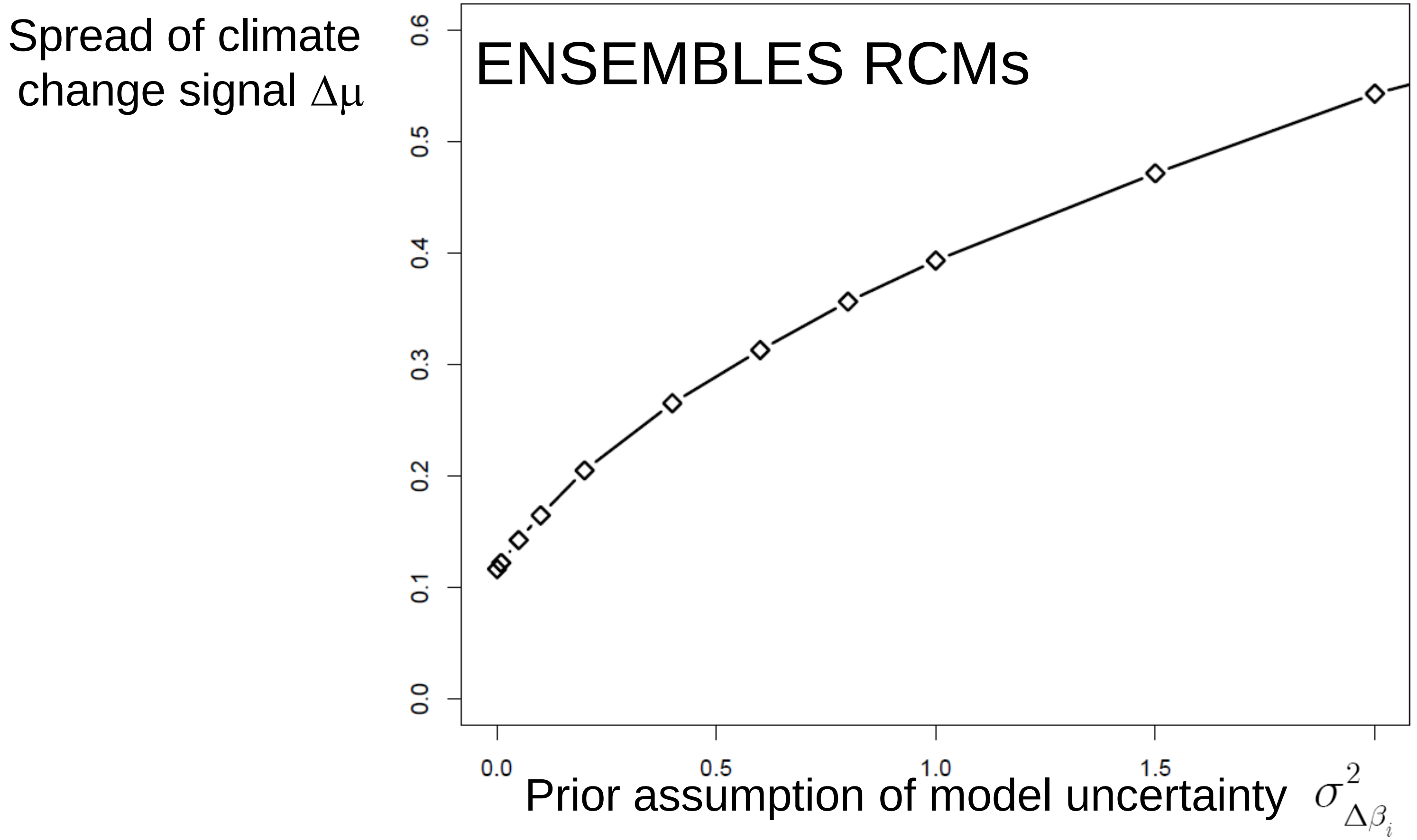
$\beta_i$: Systematic bias of model *i*

$\Delta\mu$: Climate change signal

$\Delta\beta_i$: Projection error of model *i*

*Buser et al. 2009*

# The effect of prior assumptions




Probabilistic interpretation of multi-model climate projections can be very problematic

*A. Fischer et al. 2011, subm*

# Conclusions (1)

- MMEs improve projections both in a deterministic and probabilistic sense
- A key aspect for the success of multimodels is the reduction of overconfidence
- MMEs are ensembles of opportunity
- On short time-scales:
  - Combination strategies can be judged and optimized by verification
  - Model weighting can further improve skill
  - Recalibration can make MMEs reliable

# Conclusions (2)

- As time-scales get longer, the “nature of uncertainties” becomes increasingly Bayesian

- On multi-decadal time-scales:
  - Currently no convincing concept to derive probabilistically meaningful model weights
  - Equal weighting may be better strategy, but also difficult to accomplish
  - “Strictly” probabilistic interpretation highly problematic

- Particularly on long time-scales, approaches which sample uncertainties more systematically may be preferable