Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

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

The discrete Brier and ranked probability skill scores

Andreas Weigel, Mark Liniger, Christof Appenzeller

Third International Verification Methods Workshop
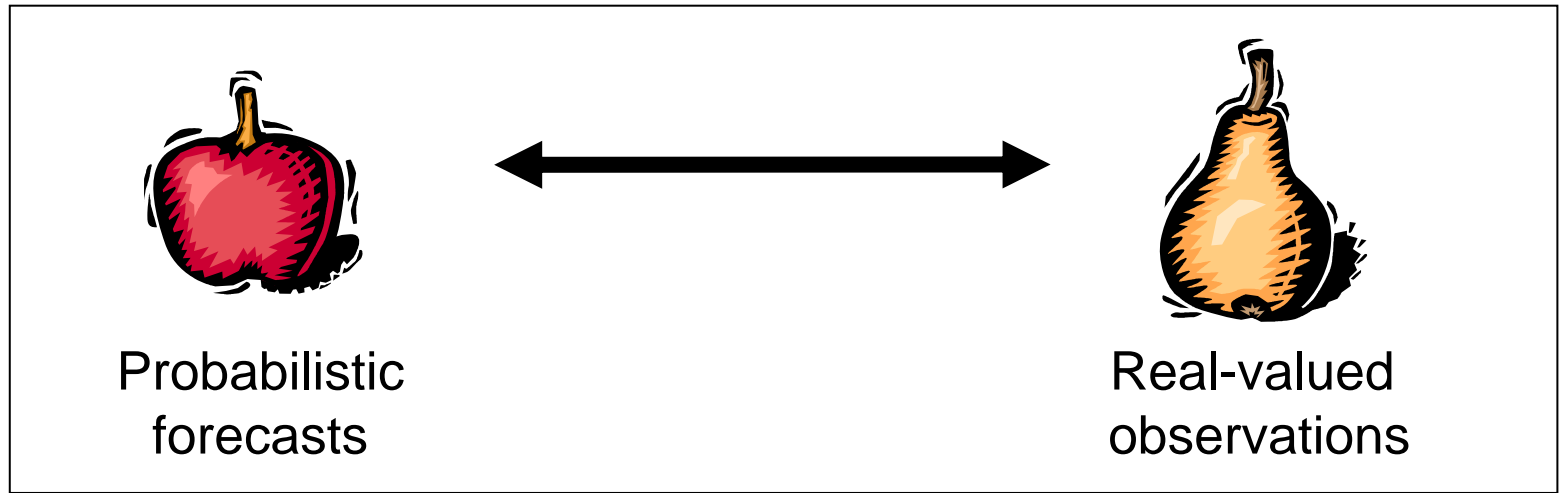ECMWF, Reading, UK

1 February 2007

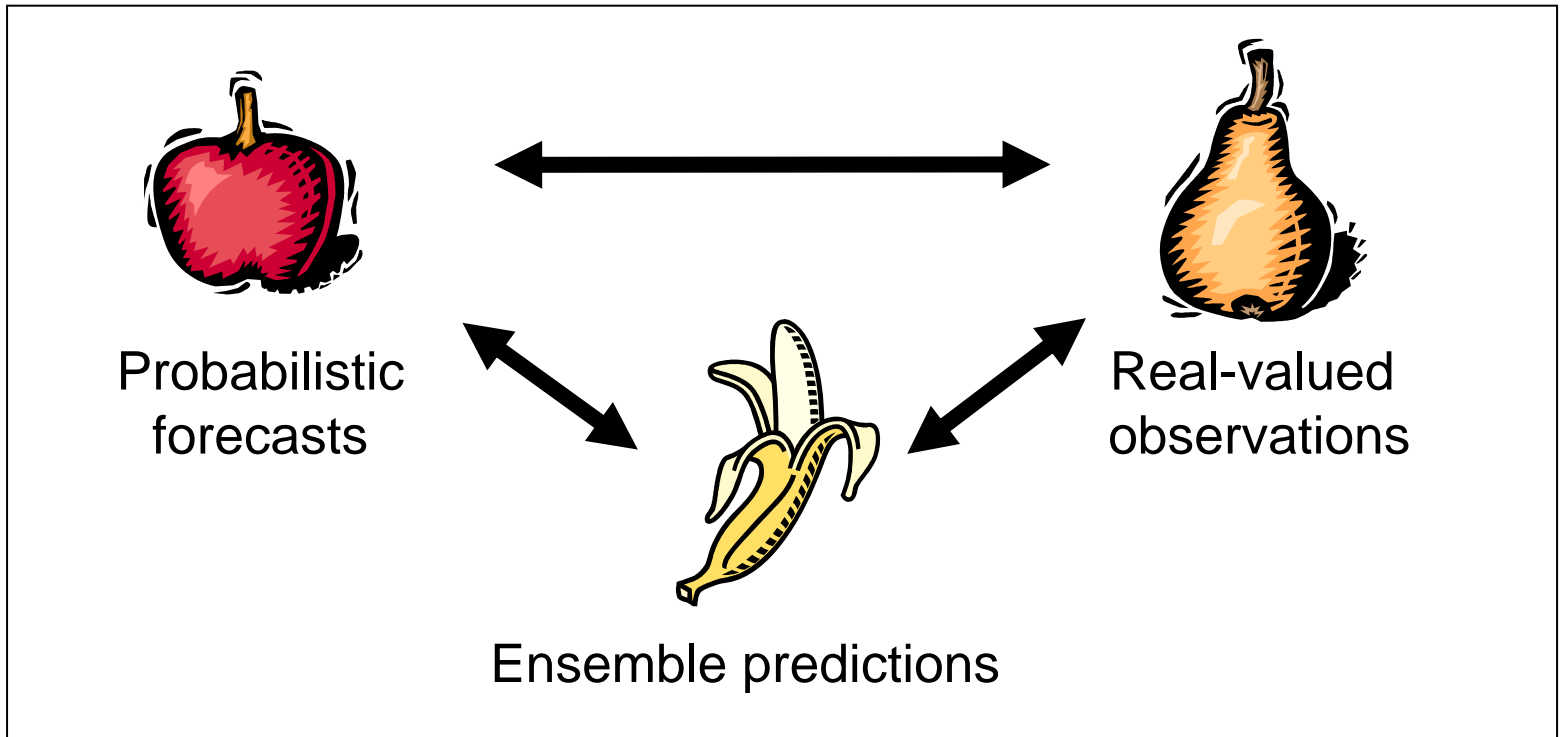NCCR CLIMATE
Swiss Climate Research

Verification of probabilistic forecasts

Probabilistic forecasts

Real-valued observations

Ensemble predictions are not truly probabilistic !!

Verification of probabilistic forecasts

Probabilistic forecasts

Real-valued observations

Ensemble predictions

Ensemble predictions are not truly probabilistic !!

Overview

☐ **The RPSS and RPSS$_D$**

☐ RPSS$_D$ for weighted multi-models

☐ Conclusions

NCCR CLIMATE
Swiss Climate Research

The RPSS
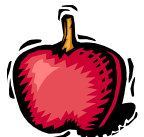## (**R**anked **P**robability **S**kill **S**core)

- Defined over categories (e.g. too cold, normal, too warm)

- Measures degree to which forecasting system outperforms a (typically climatological) reference

$$\text{RPSS} = 1 - \frac{\langle \text{RPS} \rangle}{\langle \text{RPS}_{Cl} \rangle}$$

Deviation of ensemble forecasts from observation
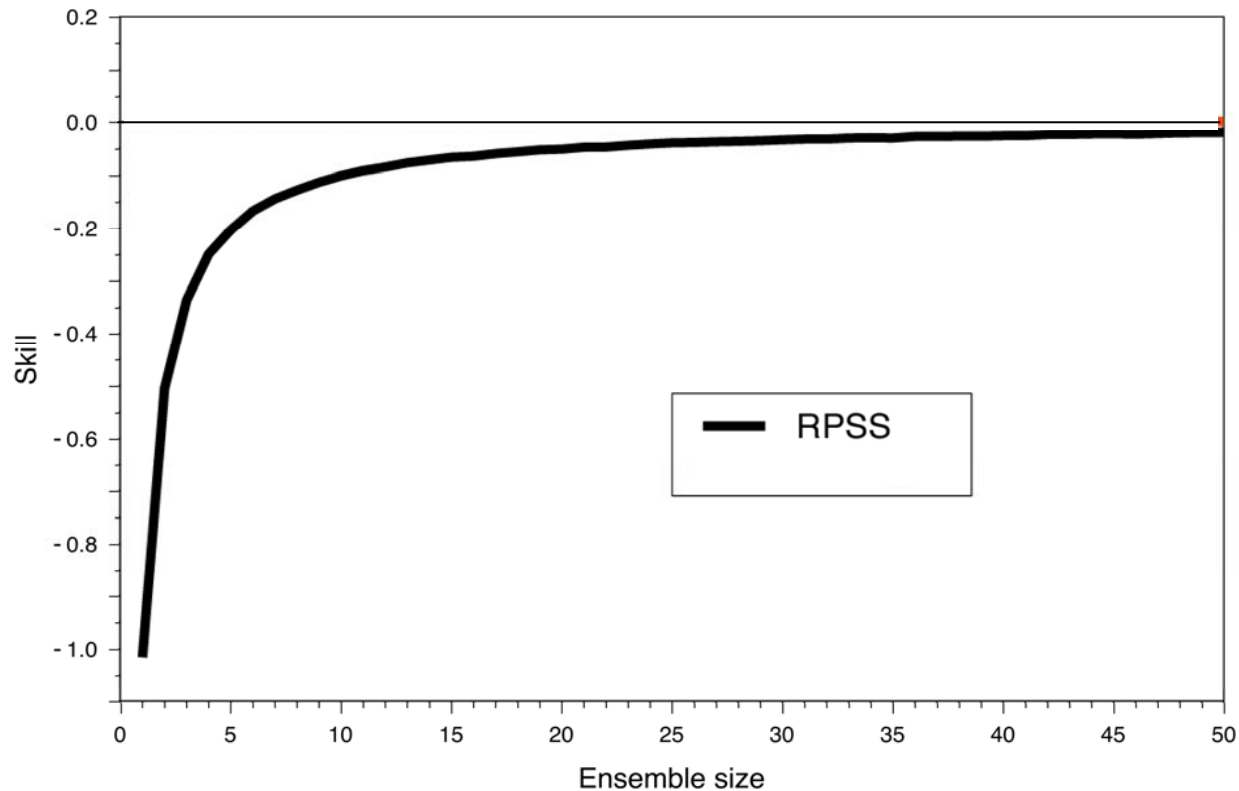
Deviation of climatologic forecasts from observation

The RPSS is negatively biased for small ensemble size !

The RPSS

- ❑ White noise toy model
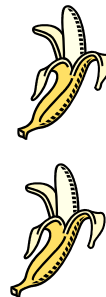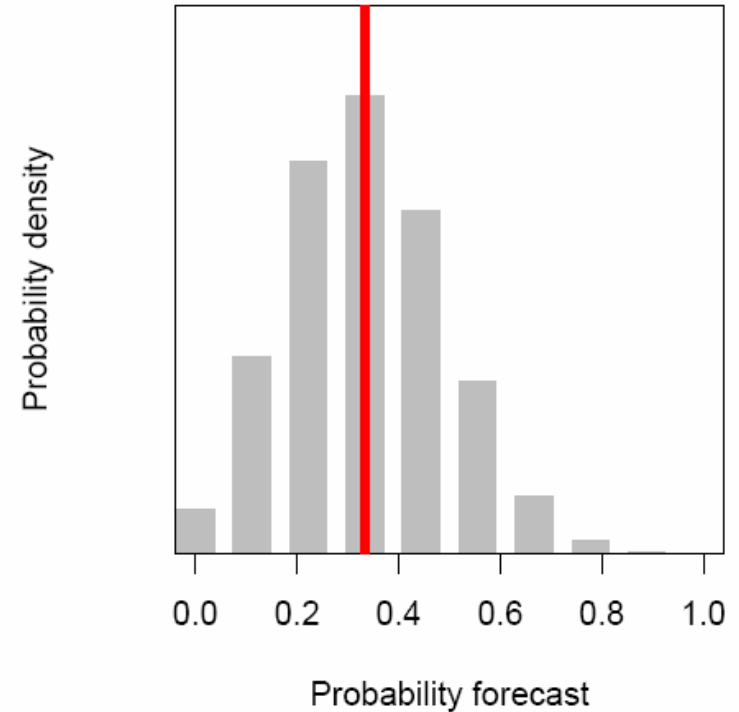- ❑ No skill by construction

The RPSS

$$\text{RPSS} = 1 - \frac{\langle \text{RPS} \rangle}{\langle \text{RPS}_{Cl} \rangle}$$

$$\text{RPSS}_D = 1 - \frac{\langle \text{RPS} \rangle}{\langle \mathcal{E}(\text{RPS}_{\text{ran}}) \rangle}$$

ensemble size: 9

Probability density

0.0 0.2 0.4 0.6 0.8 1.0

Probability forecast

Müller *et al. 2005* -> Monte Carlo approach

The Discrete Brier and Ranked Probability Skill Scores

Andreas Weigel, Verification Workshop, ECMWF, 1 Feb 2007

NCCR CLIMATE
Swiss Climate Research

The debiased RPSS$_D$

Analytical solution   ***Weigel*** *et al. 2007a*

$$\mathrm{RPSS}_D = 1 - \frac{\langle \mathrm{RPS} \rangle}{\langle \mathrm{RPS}_{Cl} \rangle + D}$$

General case

$$D = \frac{1}{M} \cdot \sum_{k=1}^{K} \sum_{i=1}^{k} \left[p_i \cdot \left(1 - p_i - 2 \sum_{j=i+1}^{k} p_j \right) \right]$$

K: Number of forecast categories
p_i: Prob. of *i*-th forecast category
M: Ensemble size

Special case: The debiased BSS_D

Analytical solution

Weigel et al. 2007a

$$BSS_D = 1 - \frac{\langle BS \rangle}{\langle BS_{Cl} \rangle + D}$$
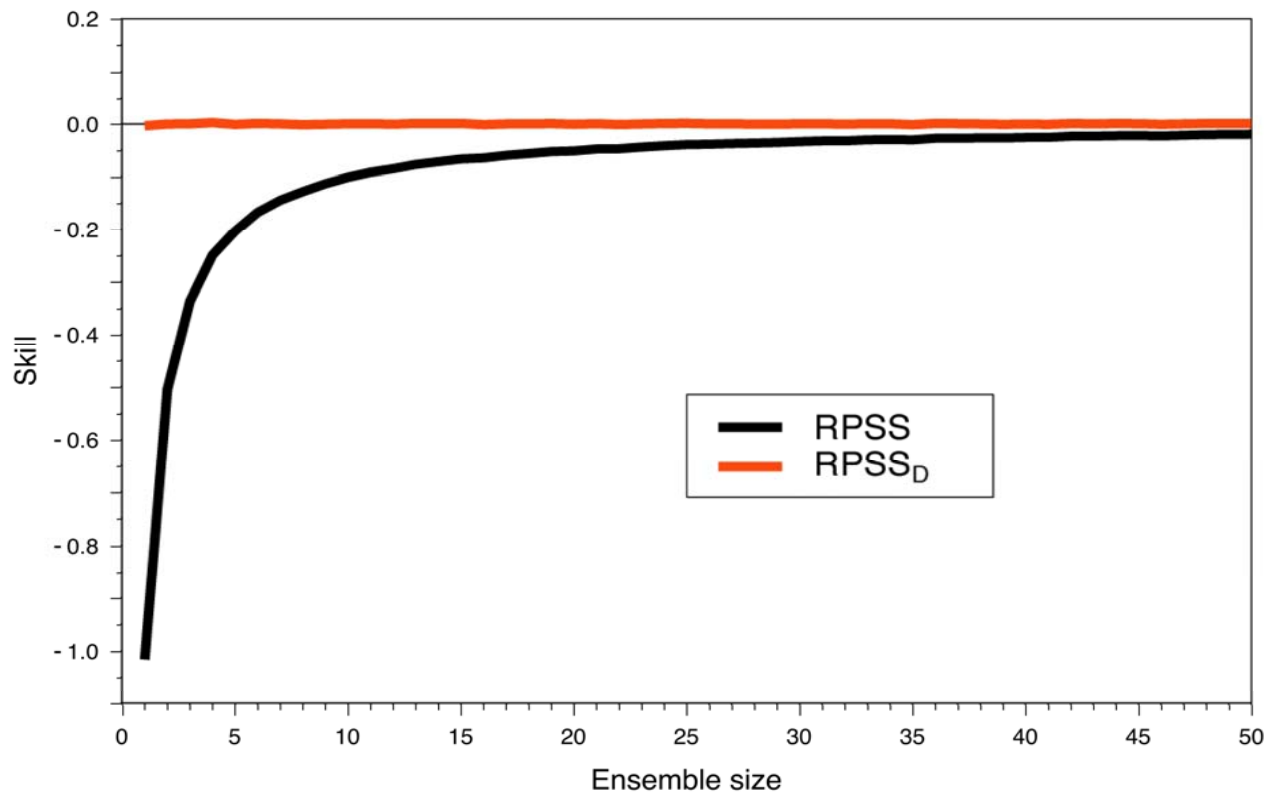
Brier score:

Two categories with prob p and (1-p)

$$D = \frac{1}{M} \cdot p \cdot (1-p)$$

D is *intrinsic (un)reliability* of EPS

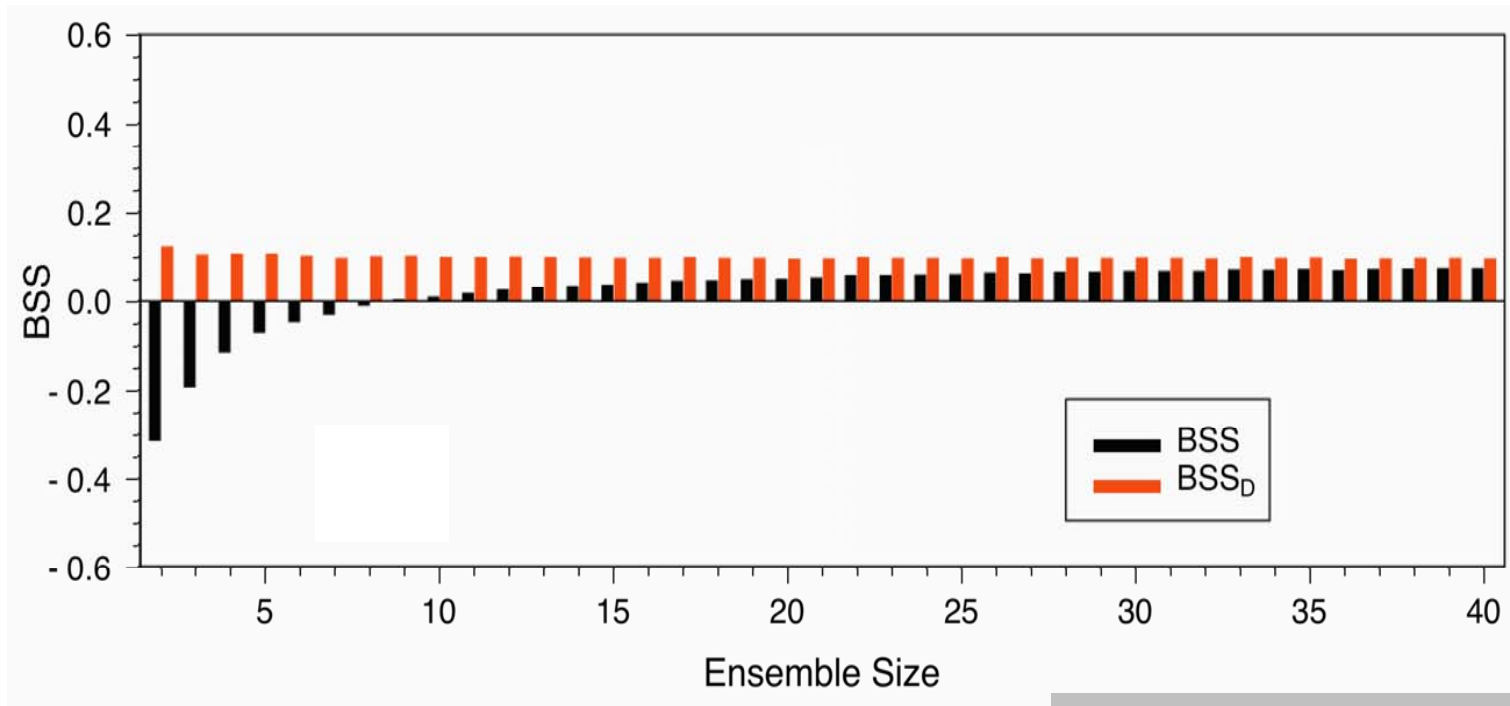NCCR CLIMATE
Swiss Climate Research

The debiased RPSS$_D$

❑ White noise toy model

❑ No skill by construction

The debiased BSS$_D$

❑ ECMWF System 2 seasonal forecasts (T2m)
❑ 2 equiprobable forecast categories

Southern Africa

Overview

❑ The RPSS and $RPSS_D$

❑ **$RPSS_D$ for weighted multi-models**

❑ Conclusions

NCCR CLIMATE
Swiss Climate Research

RPSS$_D$ for multi-models

$$\mathrm{RPSS}_D = 1 - \frac{\langle \mathrm{RPS} \rangle}{\langle \mathrm{RPS}_{Cl} \rangle + D}$$

What is M for multi-models?

$$D = \frac{1}{M} \cdot D_0(p_1, p_2, ..., p_K)$$

N: Number of models
M_n: ensemble size of n-th model
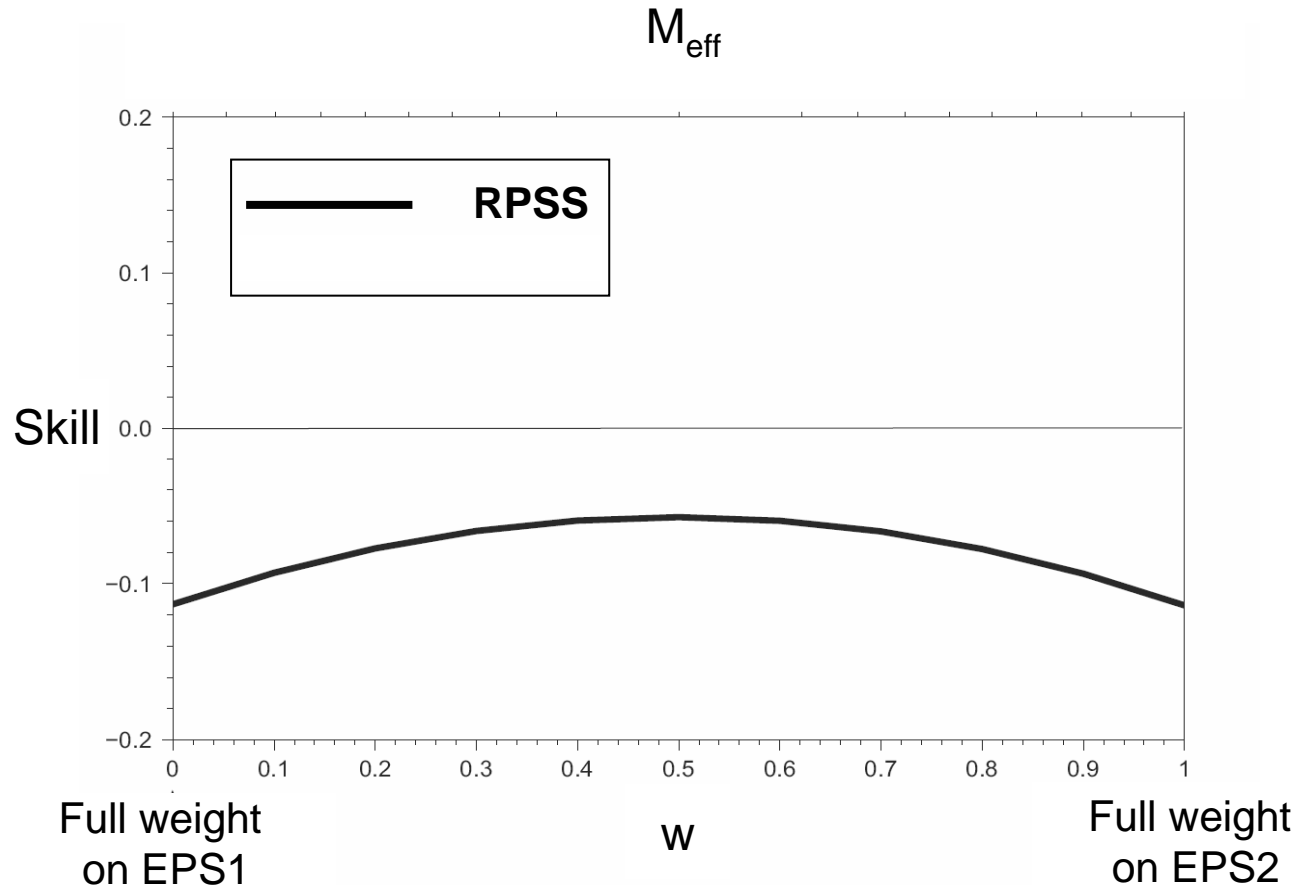w_n: weight of n-th Model

$$M_{eff} = \frac{1}{\sum_{n=1}^{N} \frac{w_n^2}{M_n}}$$
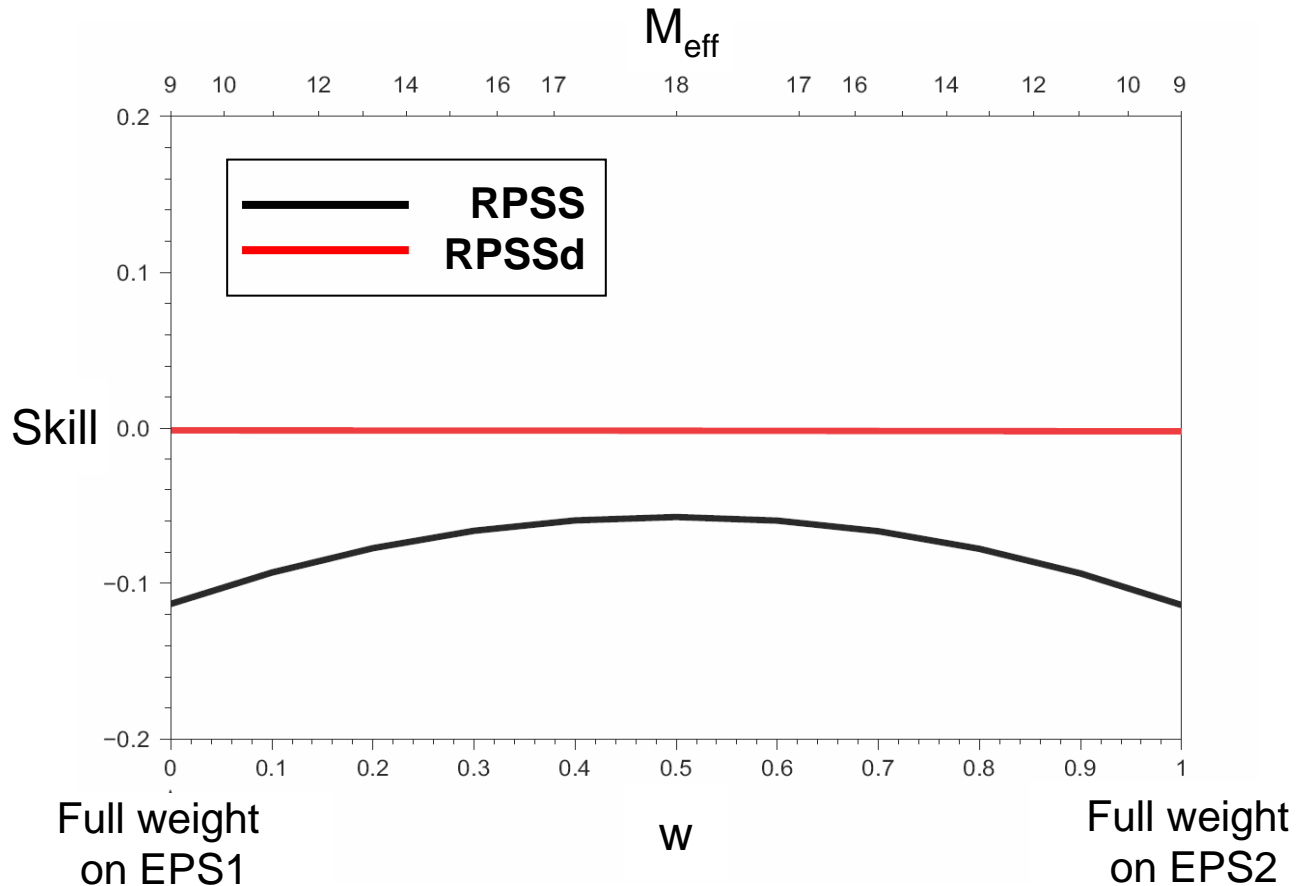
Weigel et al. 2007b

NCCR CLIMATE
Swiss Climate Research

RPSS$_D$ for multi-models

❑ Combining two white noise toymodels (9-member ensembles)
❑ Multi-model with zero skill by construction

M$_{eff}$
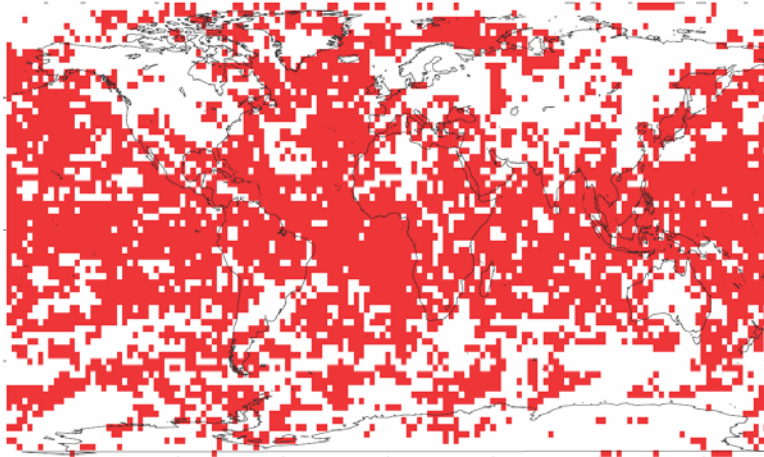
NCCR CLIMATE
Swiss Climate Research

RPSS$_D$ for multi-models

□ Combining two white noise toymodels (9-member ensembles)
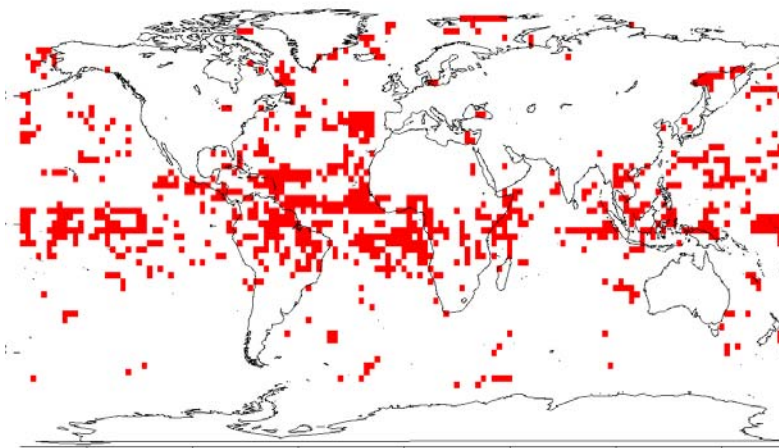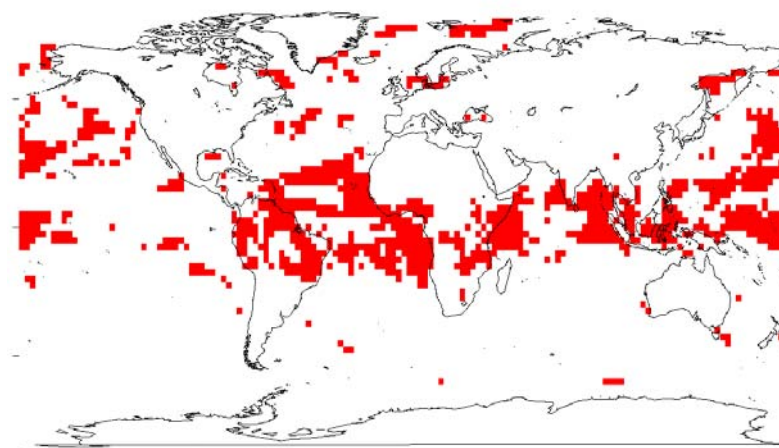□ Multi-model with zero skill by construction

Application (DEMETER data)

multi-model
better than
best participating single model
(measured with **RPSS**)

NCCR CLIMATE
Swiss Climate Research

Application (DEMETER data)

multi-model
better than
best participating single model
(measured with **RPSS$_D$**)

Both participating
single models are
highly under-dispersive

Weigel et al. 2006
(Proc. THORPEX)

The Discrete Brier and Ranked Probability Skill Scores

NCCR CLIMATE
Swiss Climate Research

Conclusions

RPSS (BSS) is negatively biased for small ensemble sizes

Bias can be removed by adding the EPS's
"intrinsic unreliability" to the climatological reference

=> Debiased **RPSS$_D$ (BSS$_D$)**

RPSS$_D$ can be generalized to multi-model ensembles
by introducing an **effective ensemble size**

RPSS measures **actual** skill of **raw** ensemble forecasts

RPSS$_D$ measures **"true"** skill of forecasting system

NCCR CLIMATE
Swiss Climate Research

References

Müller WA et al 2005: A debiased ranked probability skill score to evaluate probabilistic ensemble forecasts with small ensemble sizes. *J.Clim.*, **18**, 1513–1523.

Palmer TN et al 2004: Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bull. Amer. Meteor. Soc.,* **85**, 853–872.

Rajagopalan B et al 2002: Categorical climate forecasts through regularization and optimal combination of multiple GCM ensembles. *Mon. Wea. Rev.*, **130**, 1792-1811

Weigel AP et al 2006: Can multi-model combination really enhance prediction skill of probabilistic ensemble forecasts? *Proc. Second THORPEX International Science Symposium (STISS), Landshut, Germany,* 256-257

Weigel AP et al 2007a: The discrete Brier and ranked probability skill scores. *Mon. Wea. Rev.* **135,** 118–124.

Weigel AP et al 2007b: Generalization of the discrete Brier and ranked probability skill scores for weighted multimodel ensemble forecasts. *Mon. Wea. Rev***. (accepted)**
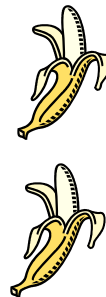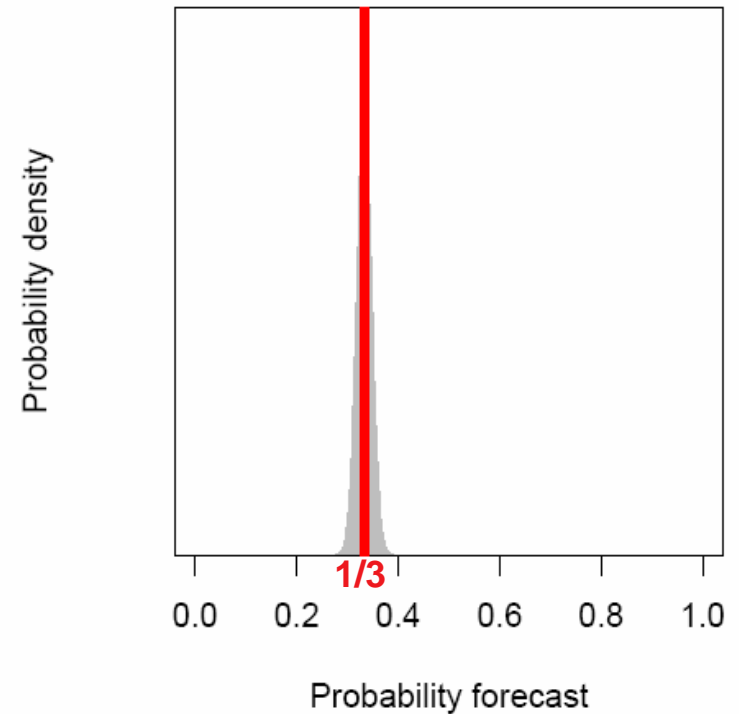
NCCR CLIMATE
Swiss Climate Research

The RPSS

$$RPSS = 1 - \frac{\langle RPS \rangle}{\langle RPS_{Cl} \rangle}$$

$$RPSS_D = 1 - \frac{\langle RPS \rangle}{\langle \mathcal{E}(RPS_{ran}) \rangle}$$

ensemble size: 1000

Probability density

Probability forecast

1/3

0.0   0.2   0.4   0.6   0.8   1.0

The debiased RPSS$_D$

Solution

Müller et al. 2005
Weigel et al. 2007a

$$\mathrm{RPSS}_D = 1 - \frac{\langle \mathrm{RPS} \rangle}{\langle \mathrm{RPS}_{Cl} \rangle + D}$$

Special case 1:

K equiprobable forecast categories

$$D = \frac{1}{M} \cdot \frac{K^2 - 1}{6K}$$

M: Ensemble size

NCCR CLIMATE
Swiss Climate Research

The meaning of *D*

Brier score decomposition:

$$BSS_D = 1 - \frac{\langle BS \rangle}{\langle BS_{Cl} \rangle + D}$$

$$\langle BS \rangle = REL - RES + UNC$$

$$\langle BS_{Cl} \rangle = UNC$$

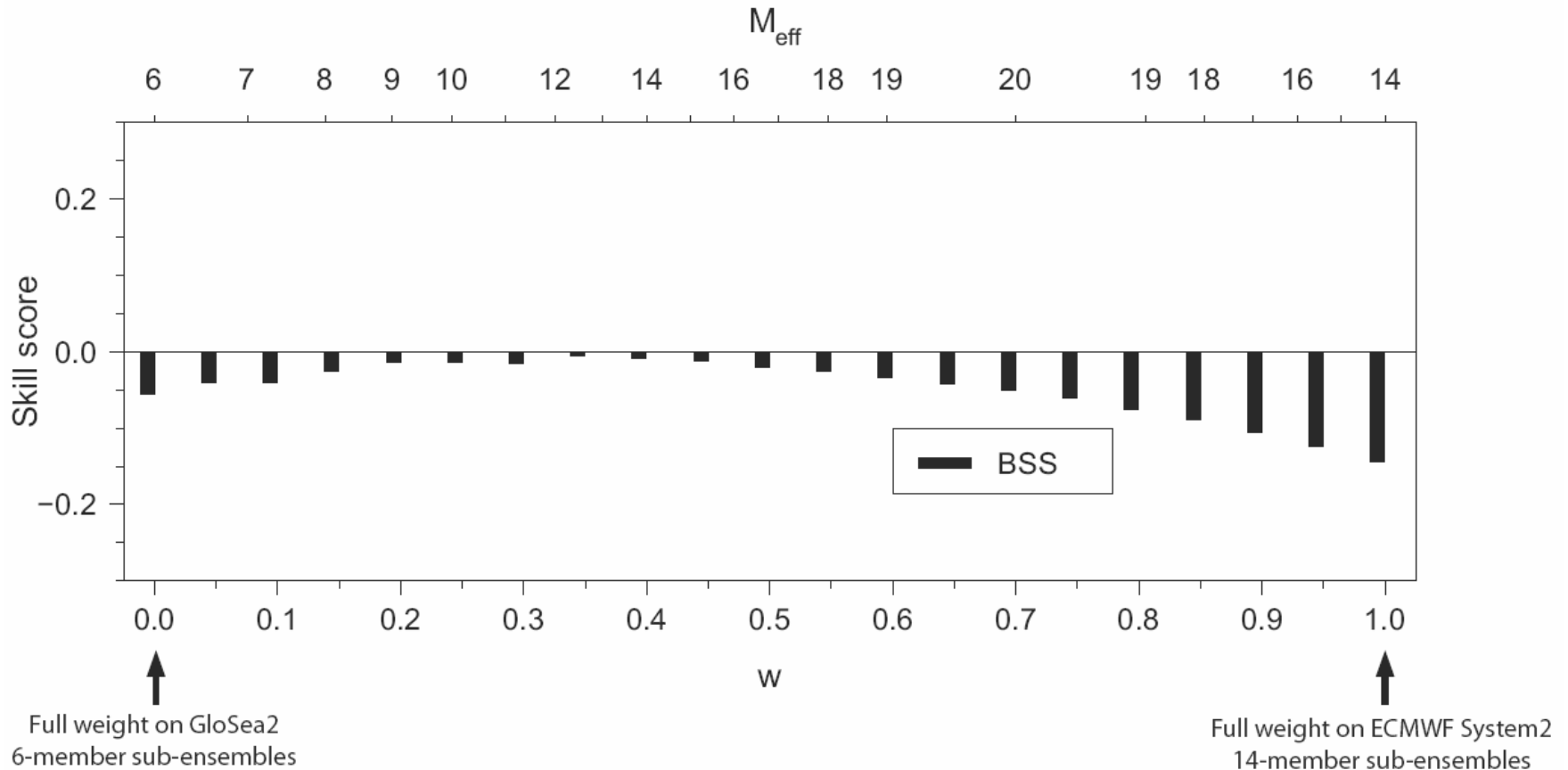$$\langle BS_{ran} \rangle = UNC + REL_{ran}$$

$$D = REL_{ran}$$

"Intrinsic (un)reliability"

NCCR CLIMATE
Swiss Climate Research

The BSS$_D$ for multi-models

- ❑ System 2 forecasts combined with GloSea forecasts (1987-2001)
- ❑ Verification against ERA40
- ❑ T2m for JJA, lead-time 1 month, eastern Europe
- ❑ 2 equiprobable forecast categories (i.e. Brier Score situation)

The BSS$_D$ for multi-models

- System 2 forecasts combined with GloSea forecasts (1987-2001)
- Verification against ERA40
- T2m for JJA, lead-time 1 month, eastern Europe
- 2 equiprobable forecast categories (i.e. Brier Score situation)
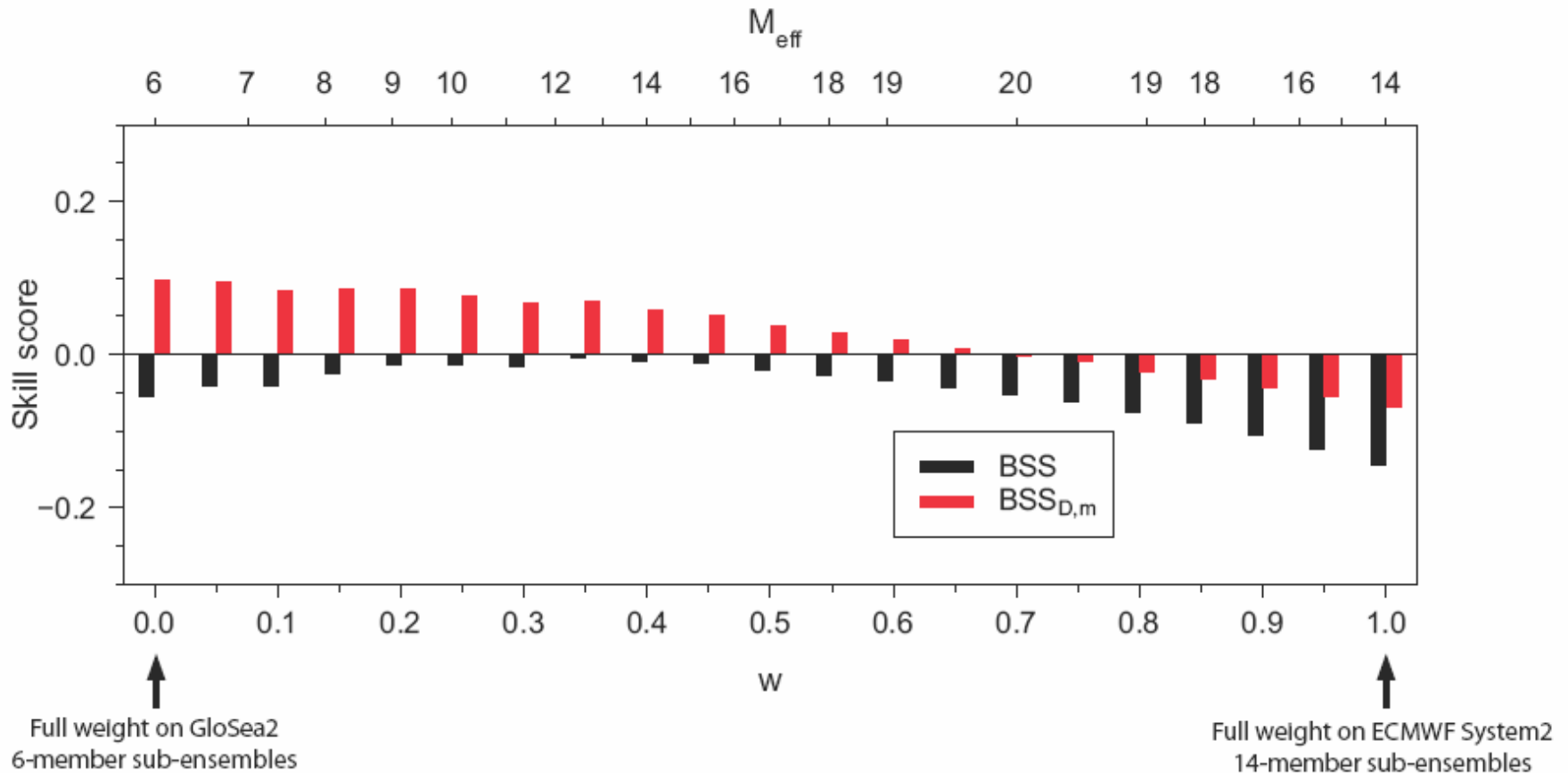
Application (DEMETER data)

Investigate where multi-model ensembles *locally* outperform the participating single models

Consider two models (ECMWF and UKMO) from the DEMETER data set (Palmer et al., 2004)

Seasonal JJA-forecasts of T 2m

Combine them to weighted multi-model ensembles using the method of Rajagopalan et al. (2002)

Verification from against ERA40 data (1960-2001)