The TIGGE experience:

Does the multi-model concept work for medium-range weather forecasts?

Renate Hagedorn

European Centre for Medium-Range Weather Forecasts

YES…

…BUT

Background and Motivation

- 2000–2003: The DEMETER project assess the value of the multi-model concept for seasonal forecasting
- 2004–2008: ECMWF tests value of reforecasts for its EPS
- 2009–2010: TIGGE multi-model forecasts as new benchmark for the (reforecast calibrated) ECMWF EPS

Does the multi-model concept work for medium-range weather forecasts?

or

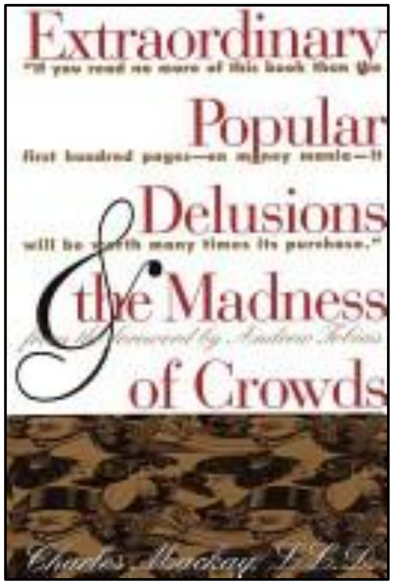
Can the TIGGE multi-model beat the (reforecast-calibrated) ECMWF EPS?

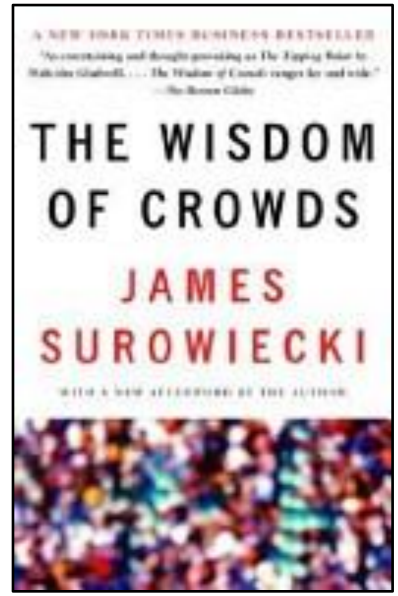
Outline

- Post-processing: Adding value to existing (raw) forecasts
  - The multi-model concept
  - Calibration
- Practical implementation and resulting limitations
- Results:
  - TIGGE multi-model vs. single-model forecasts
  - weighted TIGGE multi-model scores
- Conclusions and future work

The MM-concept

- General principle not confined to weather forecasting
- The question of the "Madness" or "Wisdom" of crowds discussed in





Charles Mackay, 1841 vs. James Surowiecki, 2004

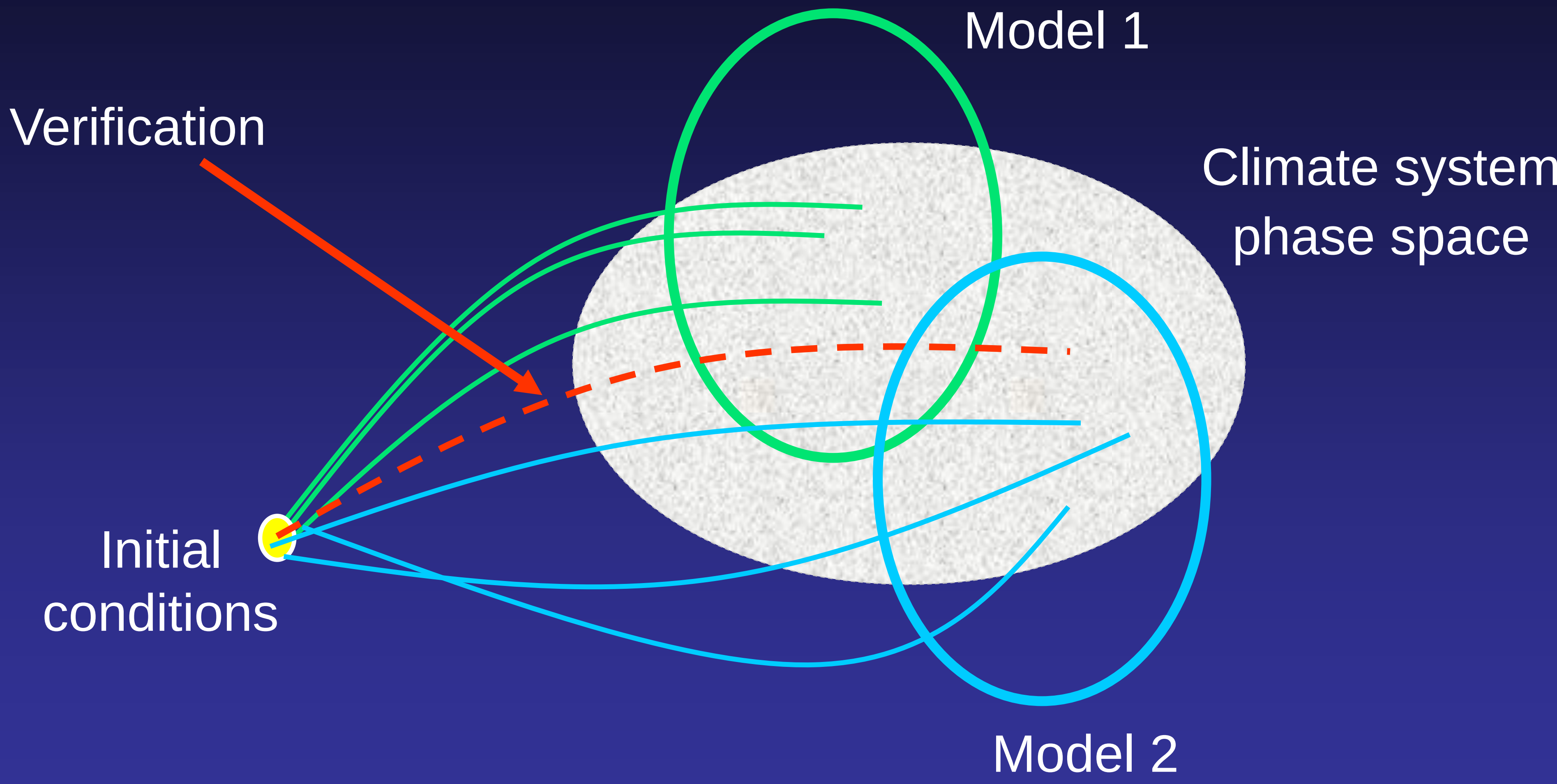
"Men, it has been well said, think in herds; it will be seen that they go mad in herds, while they only recover their senses slowly, and one by one."

Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations

The MM-concept

- Four elements required to form a wise crowd
  - **Diversity of opinion**: Each person should have private information even if it's just an eccentric interpretation of the known facts.
  - **Independence**: People's opinions aren't determined by the opinions of those around them.
  - **Decentralization**: People are able to specialize and draw on local knowledge.
  - **Aggregation**: Some mechanism exists for turning private judgments into a collective decision.

Multi-Model Ensemble Approach



Calibration

- As a simple first order calibration a **Bias Correction (BC)** can be applied:

$$c = \frac{1}{N} \sum_{i=1}^{N} (\bar{e}_i - o_i)$$

with: $\bar{e}_i$ = ensemble mean of the i[th] forecast
o_i = value of i[th] observation
N = number of observation-forecast pairs

- This correction factor is applied to each ensemble member (spread not affected)

- The **Nonhomogeneous Gaussian Regression (NGR)** accounts for existing spread-skill relationships and corrects for spread deficiencies:

$$P(v \leq q) = \Phi\left[\frac{q - (a + b\bar{x}_{ens})}{\sqrt{c + d s_{ens}^{2}}}\right]$$

- The parameter a,b,c,d are fit iteratively by minimizing the CRPS of the training data set

- Calibration provides mean and spread of Gaussian distribution

EC-CAL: combine BC and NGR

- EC-CAL Calibration process:
  - Determine optimal NGR calibration coefficients by minimizing CRPS for training dataset
  - Apply calibration NGR calibration coefficients to determine calibrated PDF from ensemble mean and variance of actual forecast to be calibrated
  - Create calibrated NGR-ensemble with 51 *synthetic* members
  - Combine NGR-ensemble with ‘30-day bias corrected’ forecast ensemble

- This combined BC-NGR calibration improves the pure NGR calibration by
  - partly retaining shape of original PDF
  - Improved bias-correction through higher weighting of bias related to current or most recent weather conditions

Training datasets

- All post-processing methods need a training dataset, containing a number of forecast-observation pairs from the past.
- Post-processing coefficients / weights can be:
  - Fixed: same coefficients are used every day, calculation of coefficients based on long training dataset of operational forecasts (~1-2 years)
  - Varying: coefficients are continuously updated, based on
    i. Set of previous operational forecasts (last 30-45 days)
    ii. Reforecast dataset (~5 week window around target)

The reforecast dataset

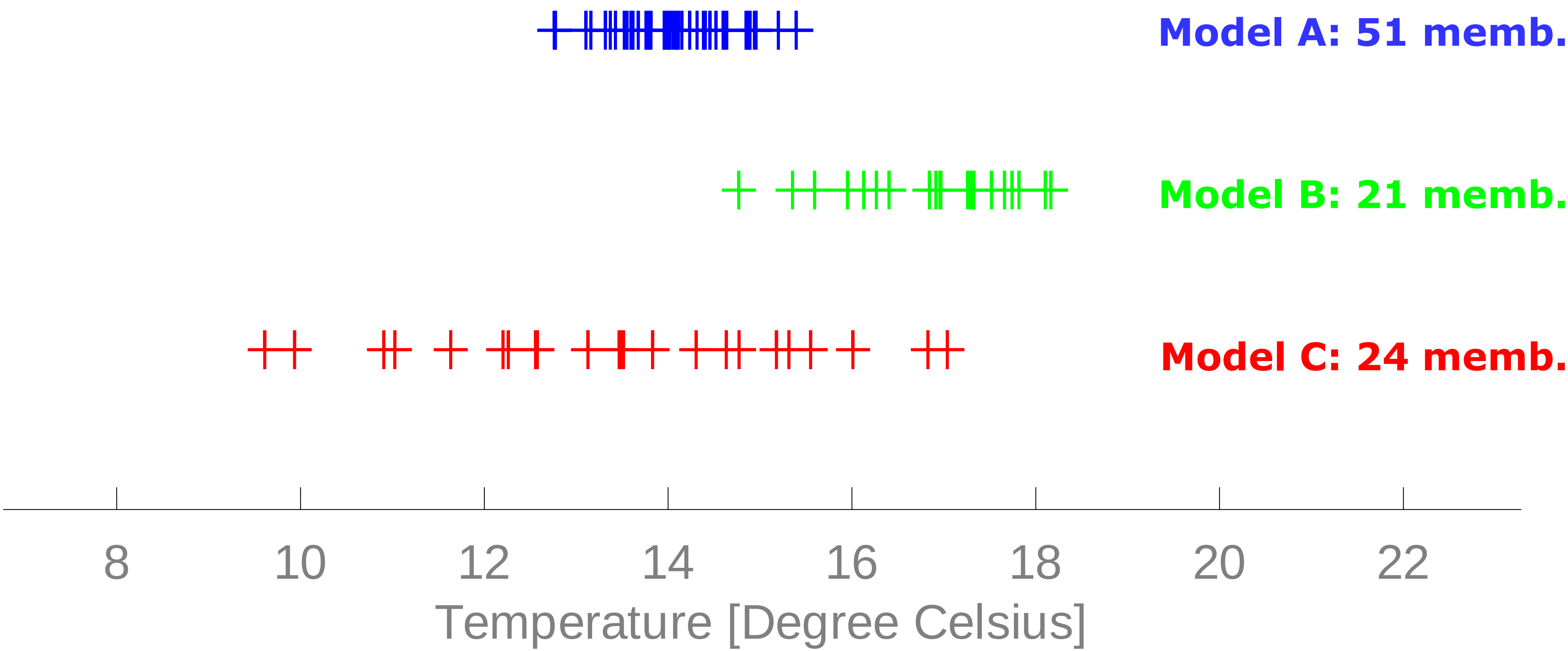
	2010																																		
	Mar		Apr																														May		
	30	31	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	01	02	03
1992																																			
1993																																			
1994																																			
1995																																			
1996																																			
. . . .																																			
2005																																			
2006																																			
2007																																			
2008																																			
2009																																			

Outline

- Post-processing: Adding value to existing (raw) forecasts
  - The multi-model concept
  - Calibration
- Practical implementation and resulting limitations
- Results:
  - TIGGE multi-model vs. single-model forecasts
  - weighted TIGGE multi-model scores
- Conclusions and future work

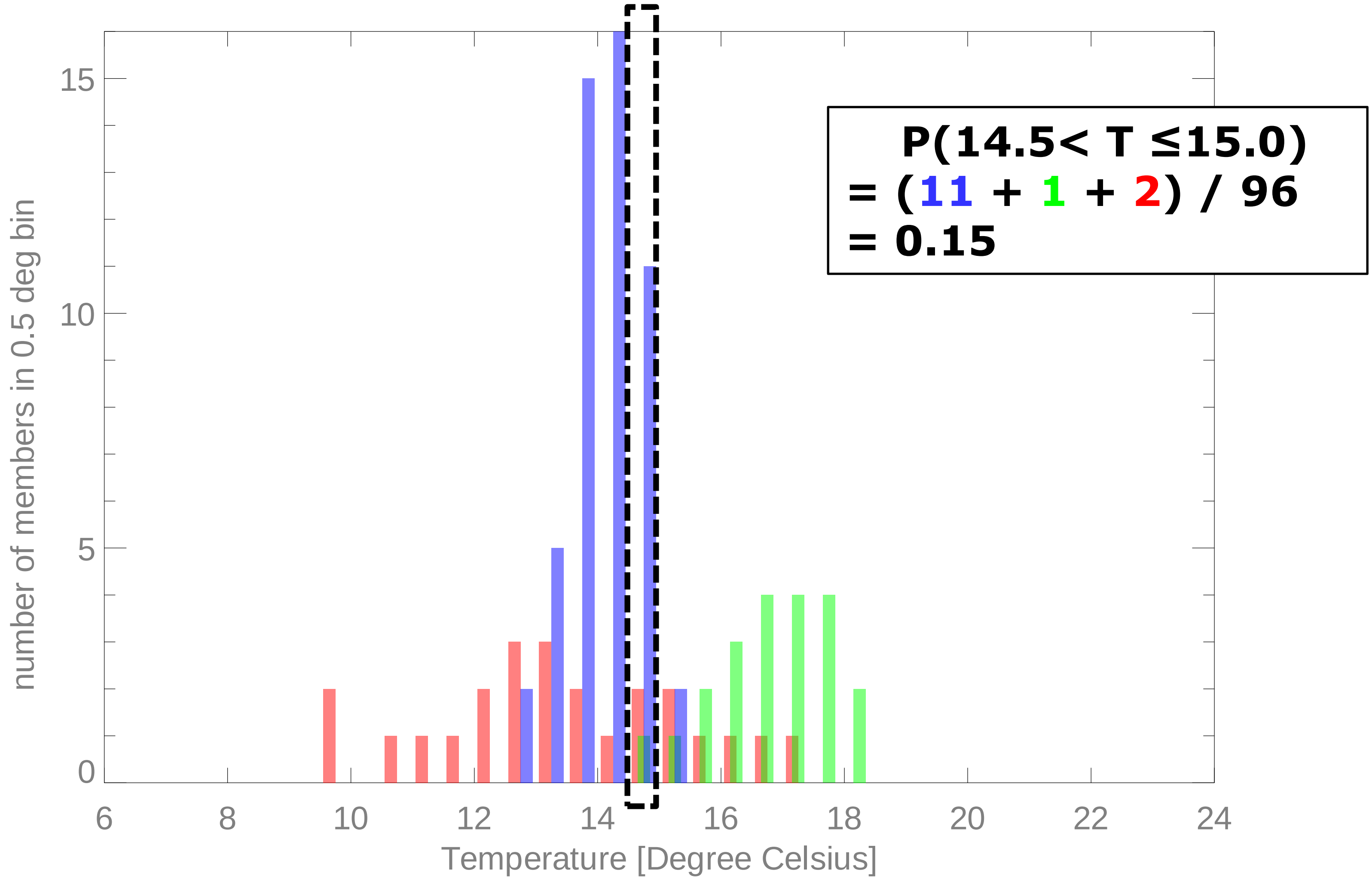
Constructing MM-Forecasts

- Starting with a number of single model ensemble forecasts



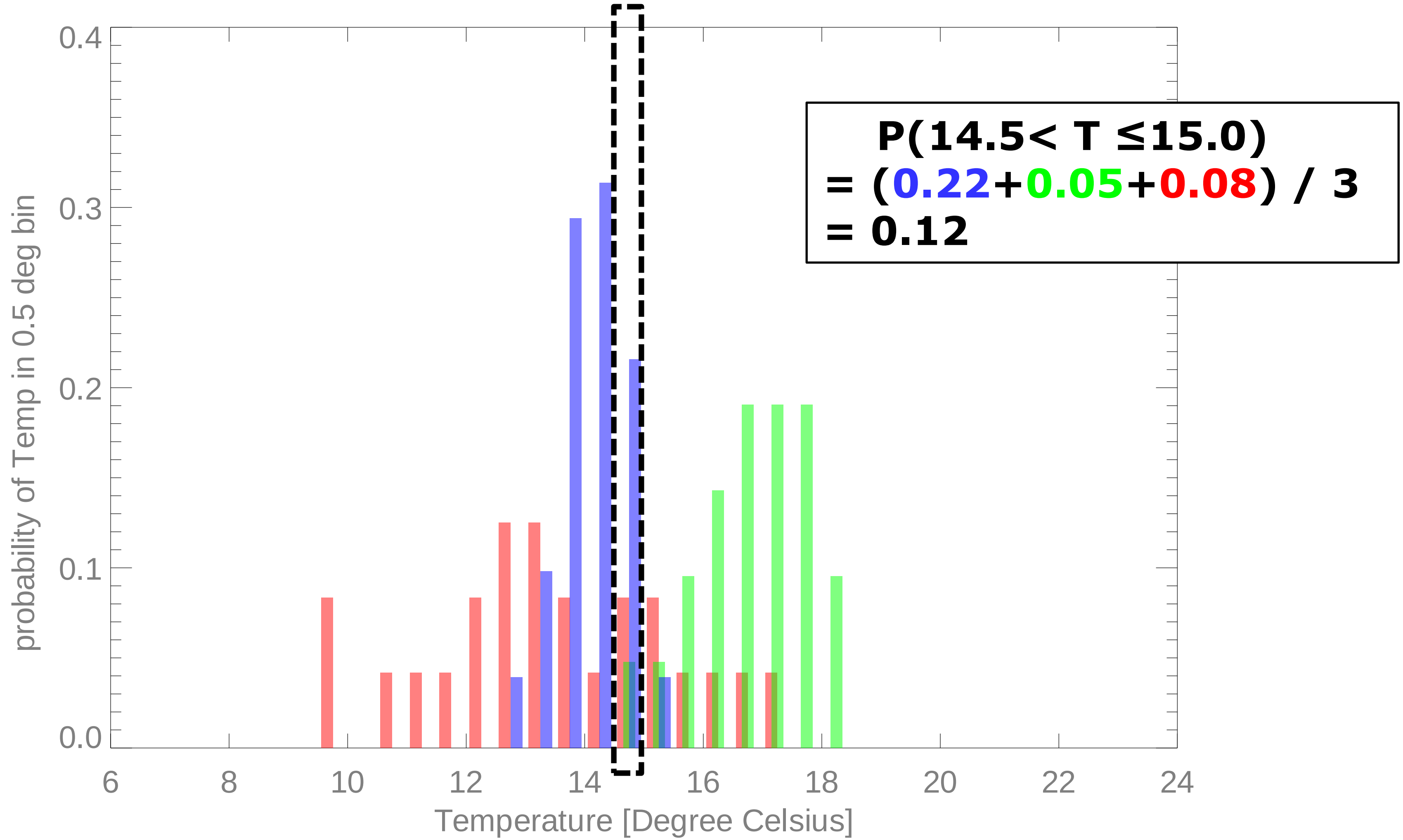
Calculating MM-Probabilities

- Counting ensemble members in discrete bins



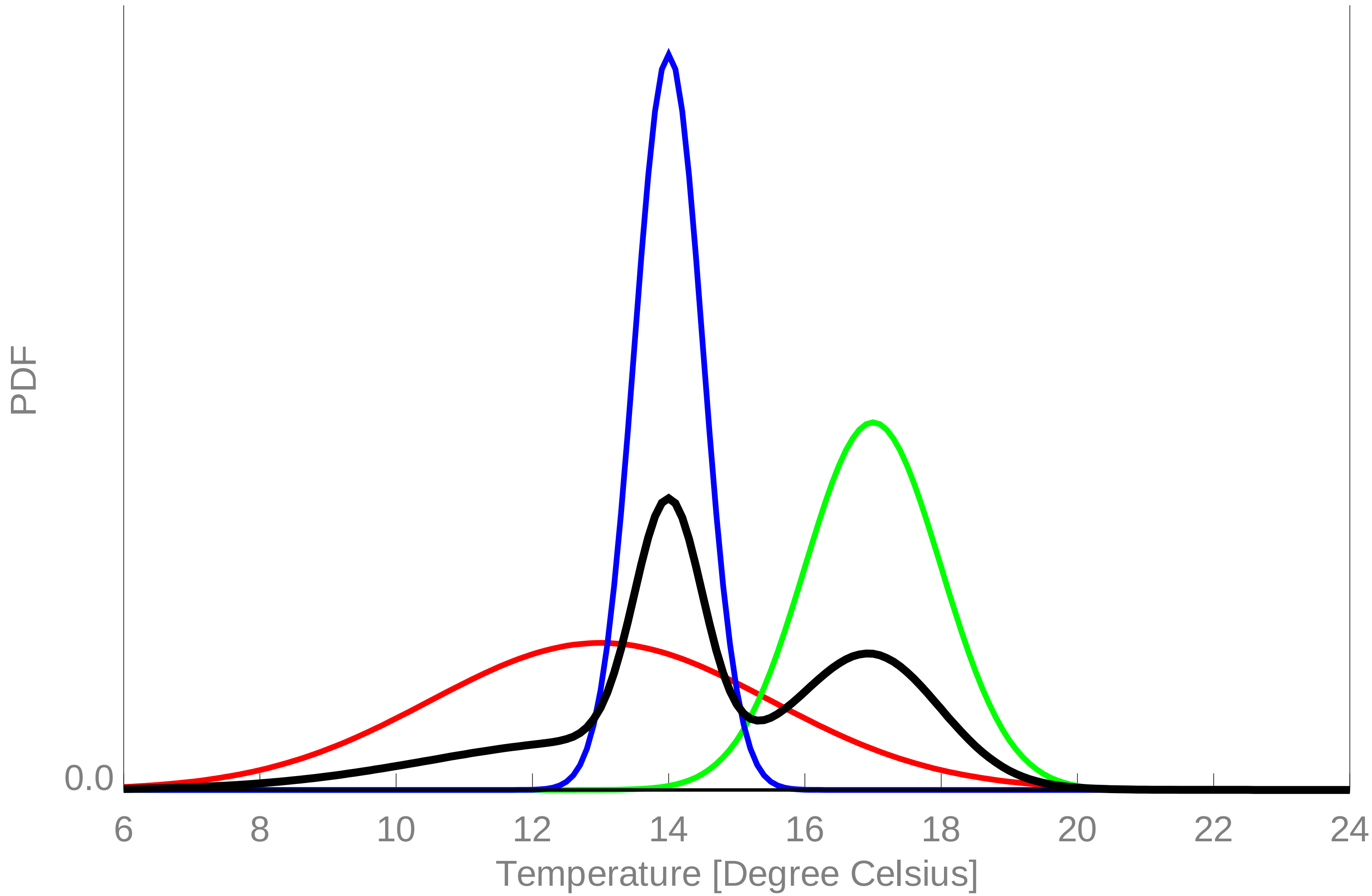
Calculating MM-Probabilities

- Averaging probabilities in discrete bins



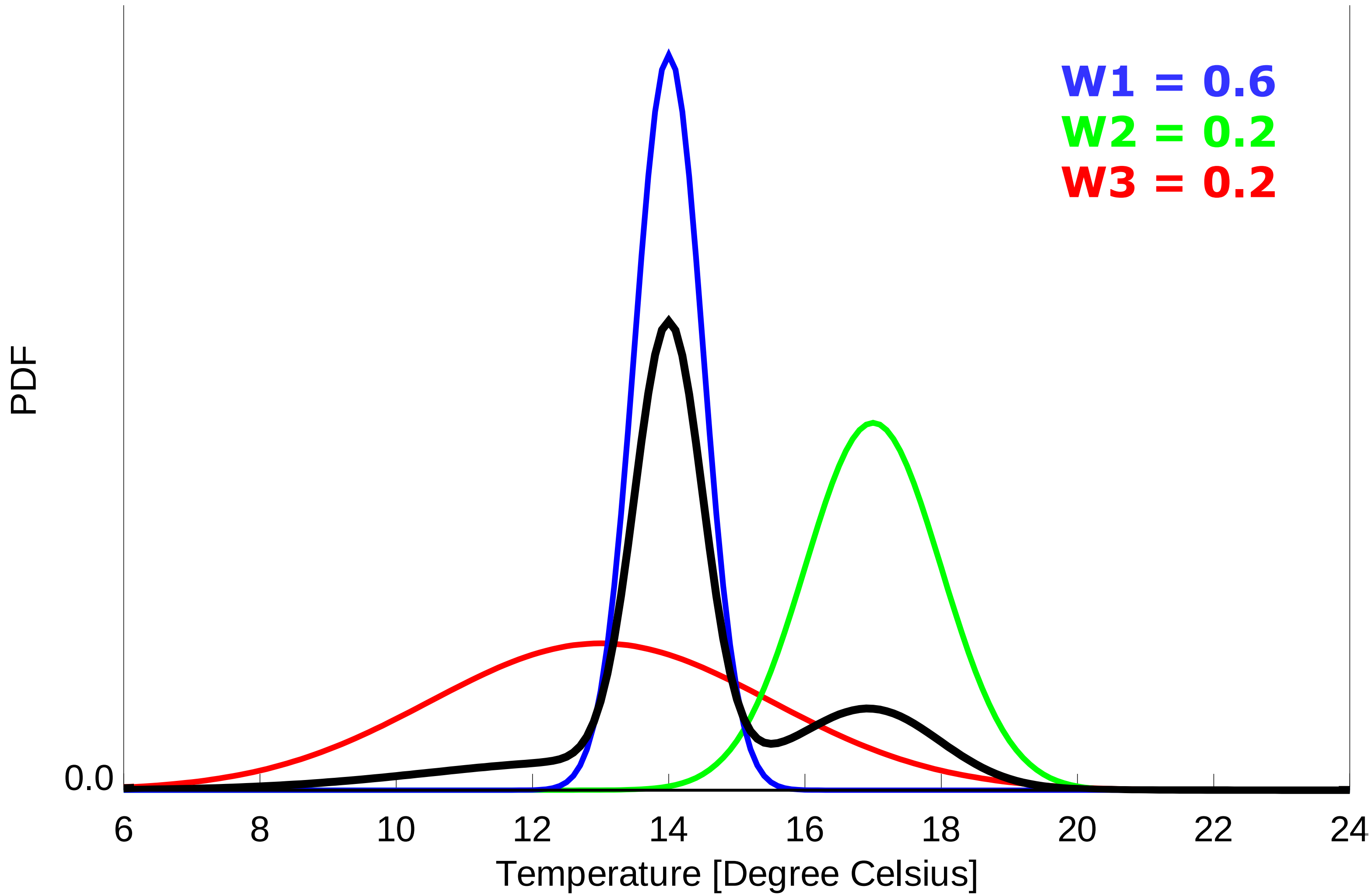
Constructing MM-PDFs

- Averaging probabilities of continuous PDFs



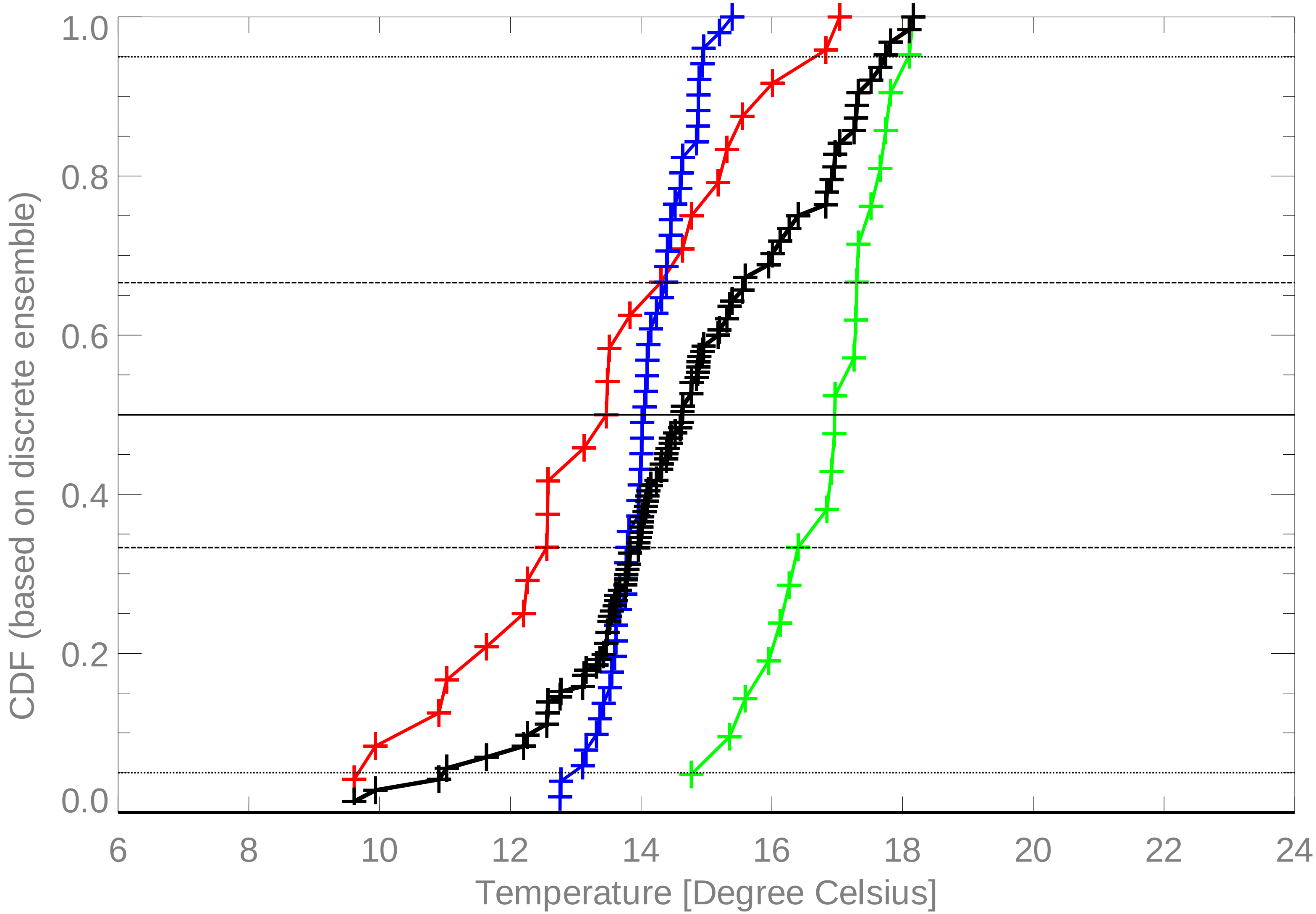
Constructing MM-PDFs

- Averaging probabilities of continuous PDFs with different weights



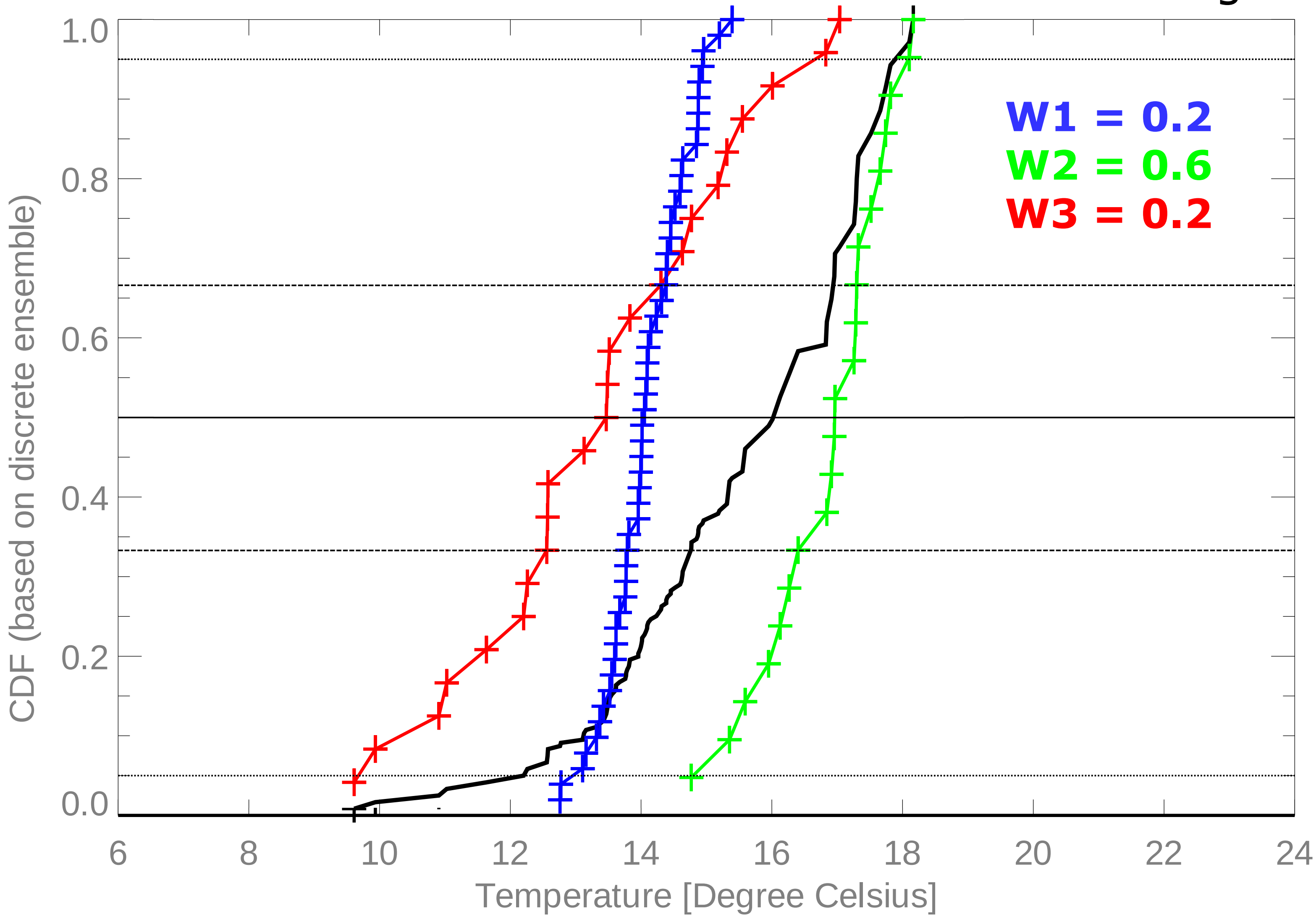
Constructing MM-CDFs

- Construct discrete CDFs from ensemble members



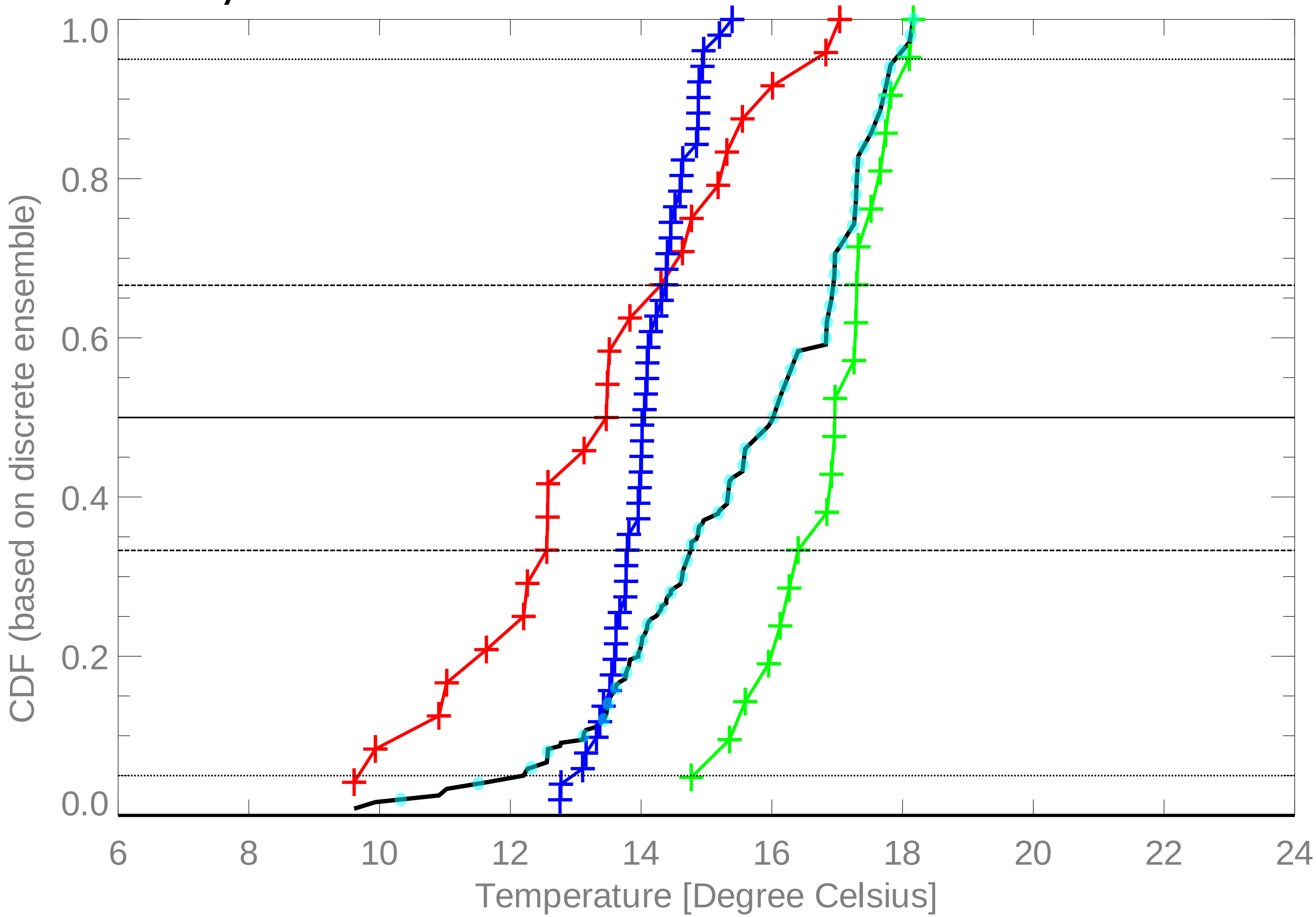
Constructing MM-CDFs

- Construct discrete CDFs from ensembles with different weights

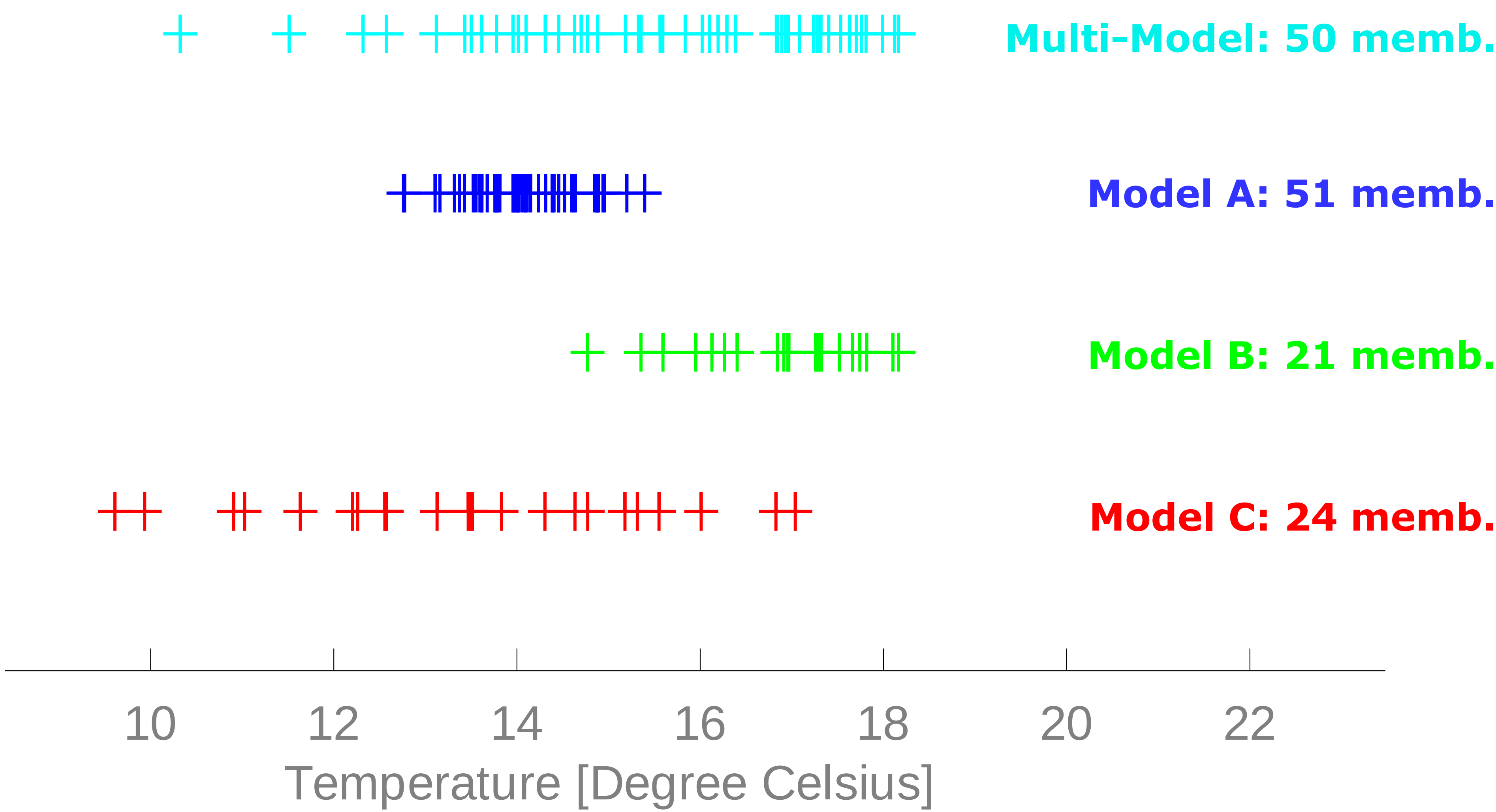


Re-constructing MM-ensemble members

- Re-construct synthetic ensemble members from MM-CDF



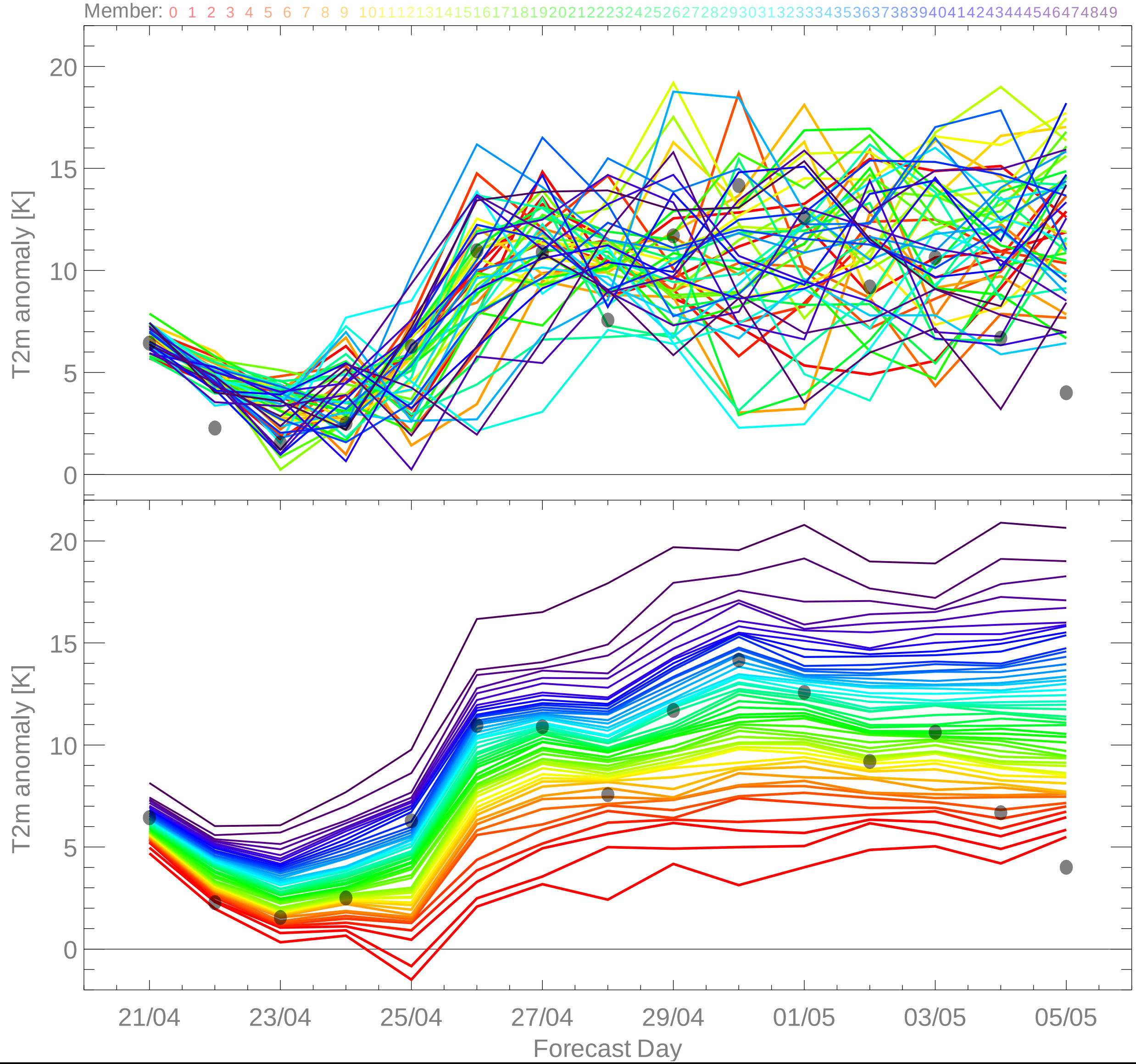
Final weighted MM-ensemble members



Real vs. synthetic ensemble

Bias-corrected ECMWF EPS

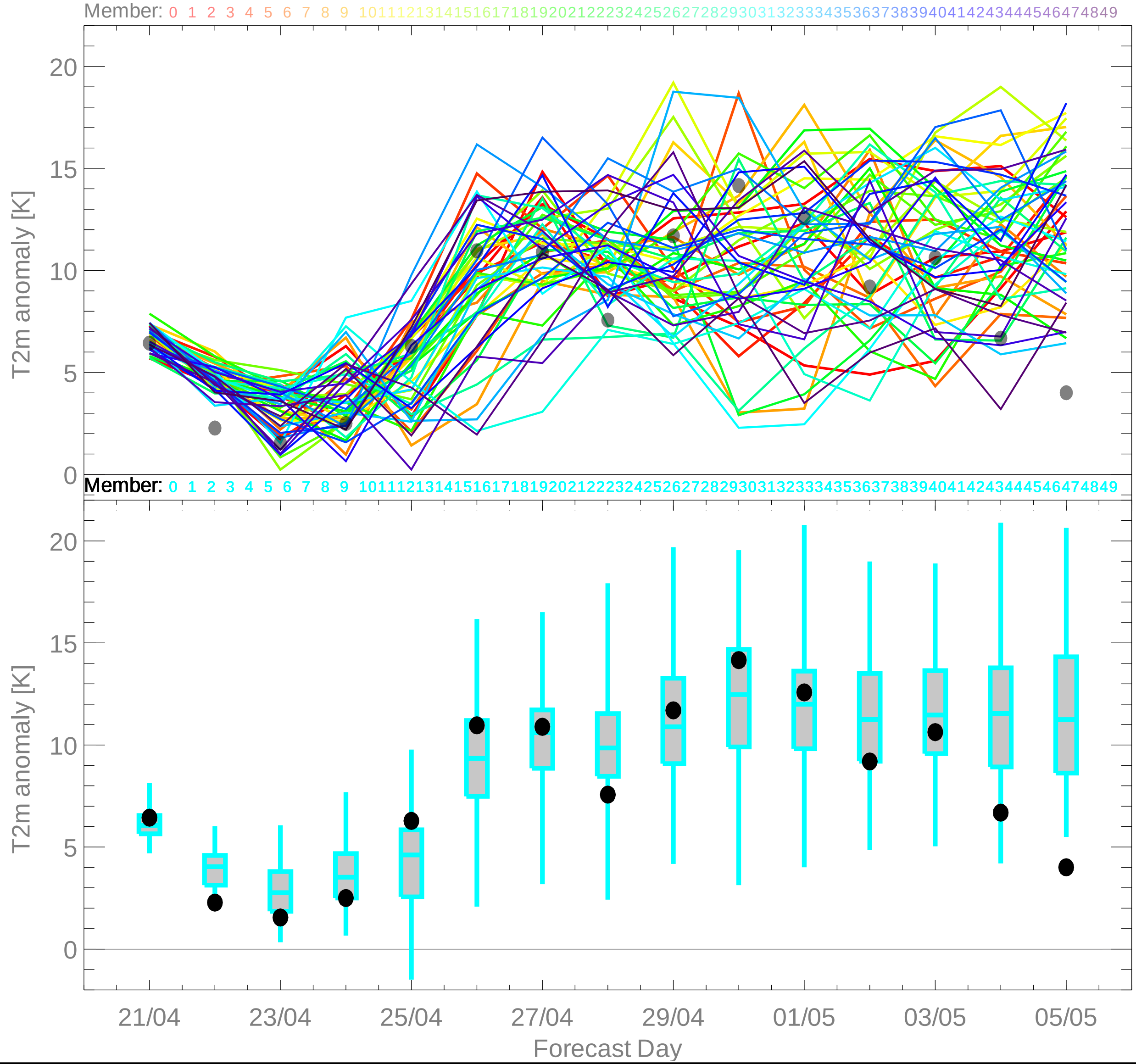
Synthetic TIGGE MM



Real vs. synthetic ensemble

Bias-corrected ECMWF EPS

Synthetic TIGGE MM



Be aware that...

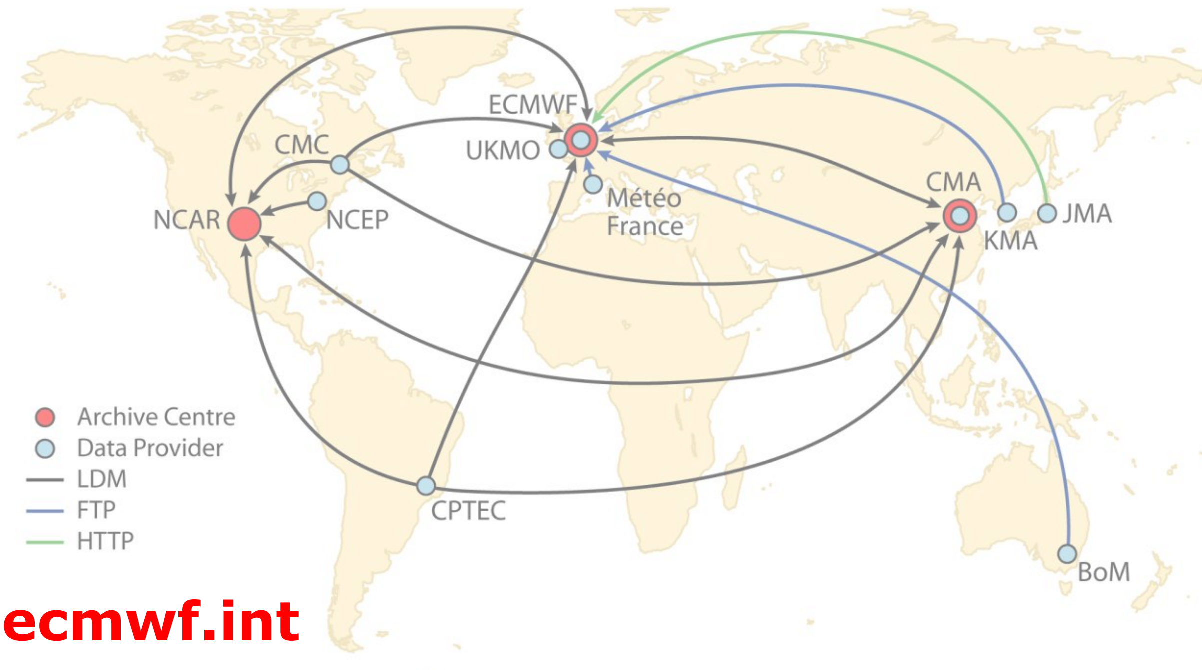
- By creating synthetic ensemble members
  - we lose the temporal/spatial structure of individual forecasts
  - we potentially lose multivariate relations
- Acceptable, if we are interested in local predictions and not scenarios, i.e. if we do not look at temporal evolution, spatial fields, multivariate structures

Outline

- Post-processing: Adding value to existing (raw) forecasts
  - The multi-model concept
  - Calibration
- Practical implementation and resulting limitations
- Results:
  - TIGGE multi-model vs. single-model forecasts
  - weighted TIGGE multi-model scores
- Conclusions and future work

The TIGGE dataset

- THORPEX Interactive Grand Global Ensemble:
  - Global operational ensemble forecasts from 10 centres:

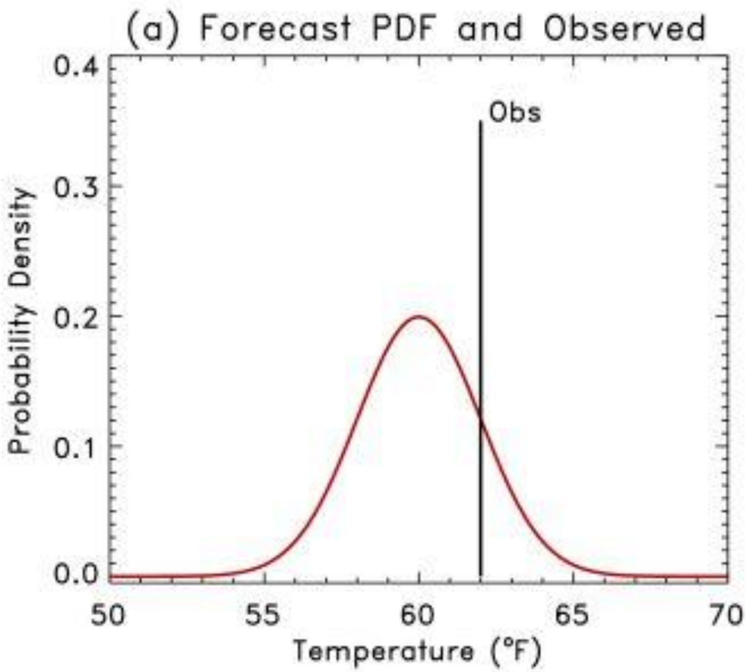


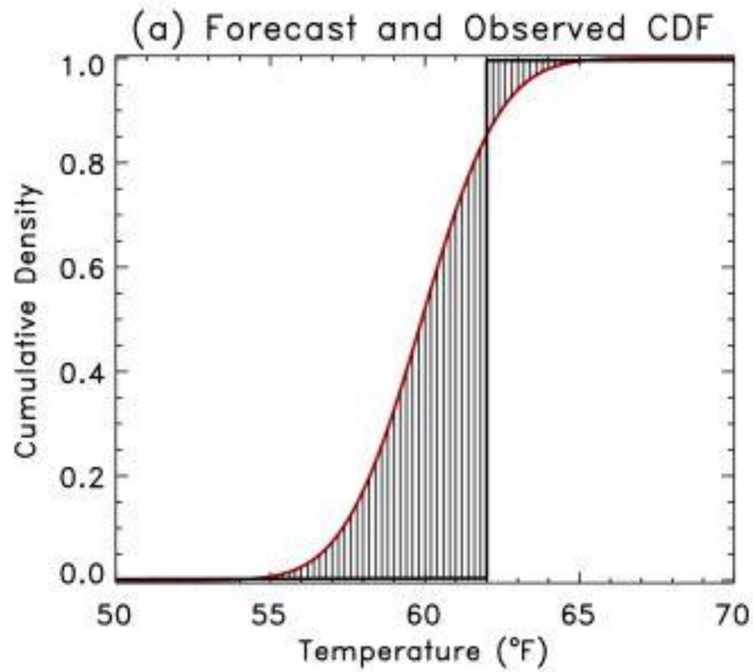
http://tigge.ecmwf.int

  - A large range of variables are available in "near-realtime"
  - Here we consider 2m Temperature forecasts (DJF 2008/09 & MAM 2010)

Continuous Rank Probability Score

$$CRPS = \frac{1}{N}\sum_{i=1}^{N}\int_{x=-\infty}^{x=+\infty}\left(F_i(x)-O_i(x)\right)^2$$





$$CRPSS = 1 - \frac{CRPS_{FC}}{CRPS_{ref}}$$

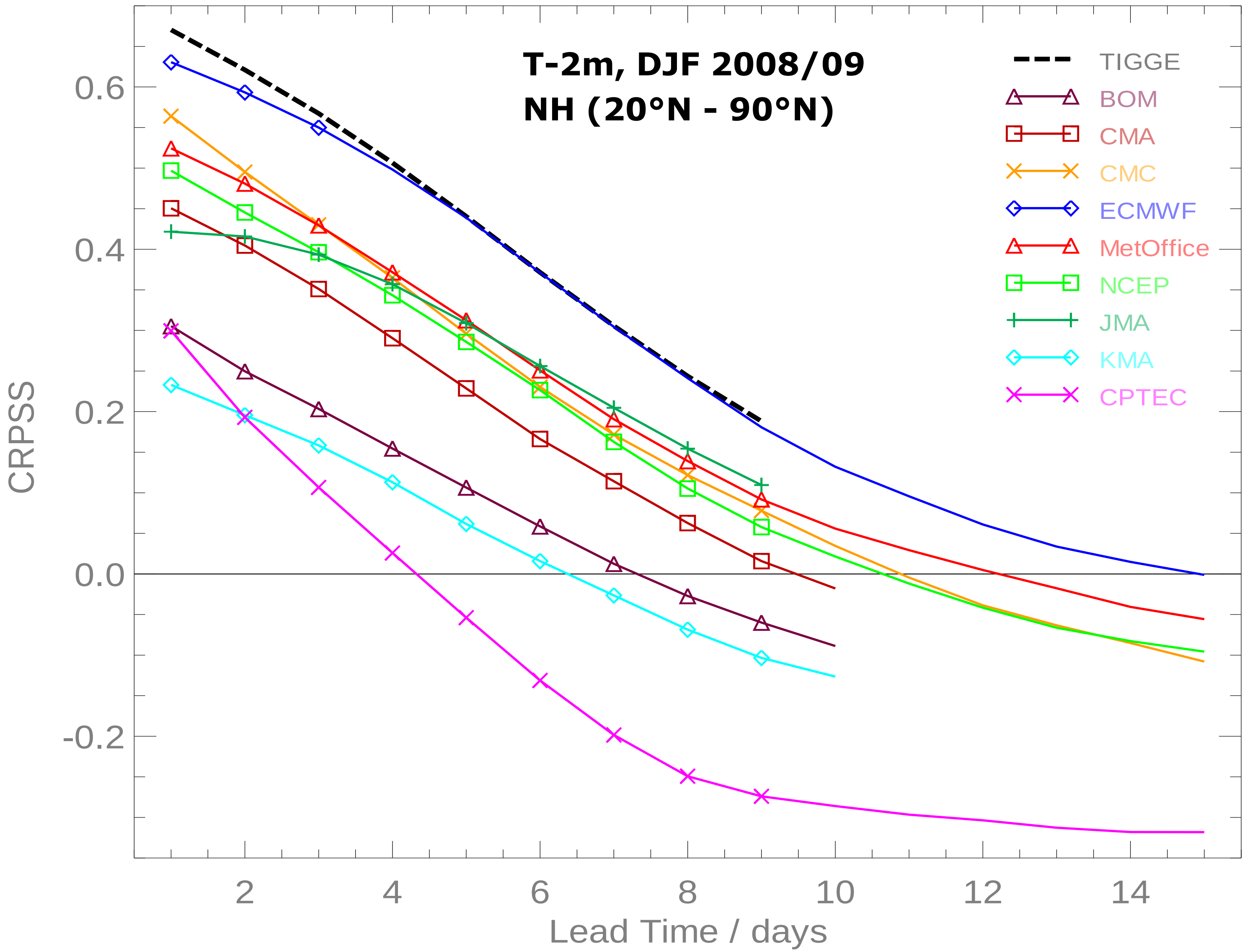
Reminder of main question

Does the multi-model concept work for medium-range weather forecasts?

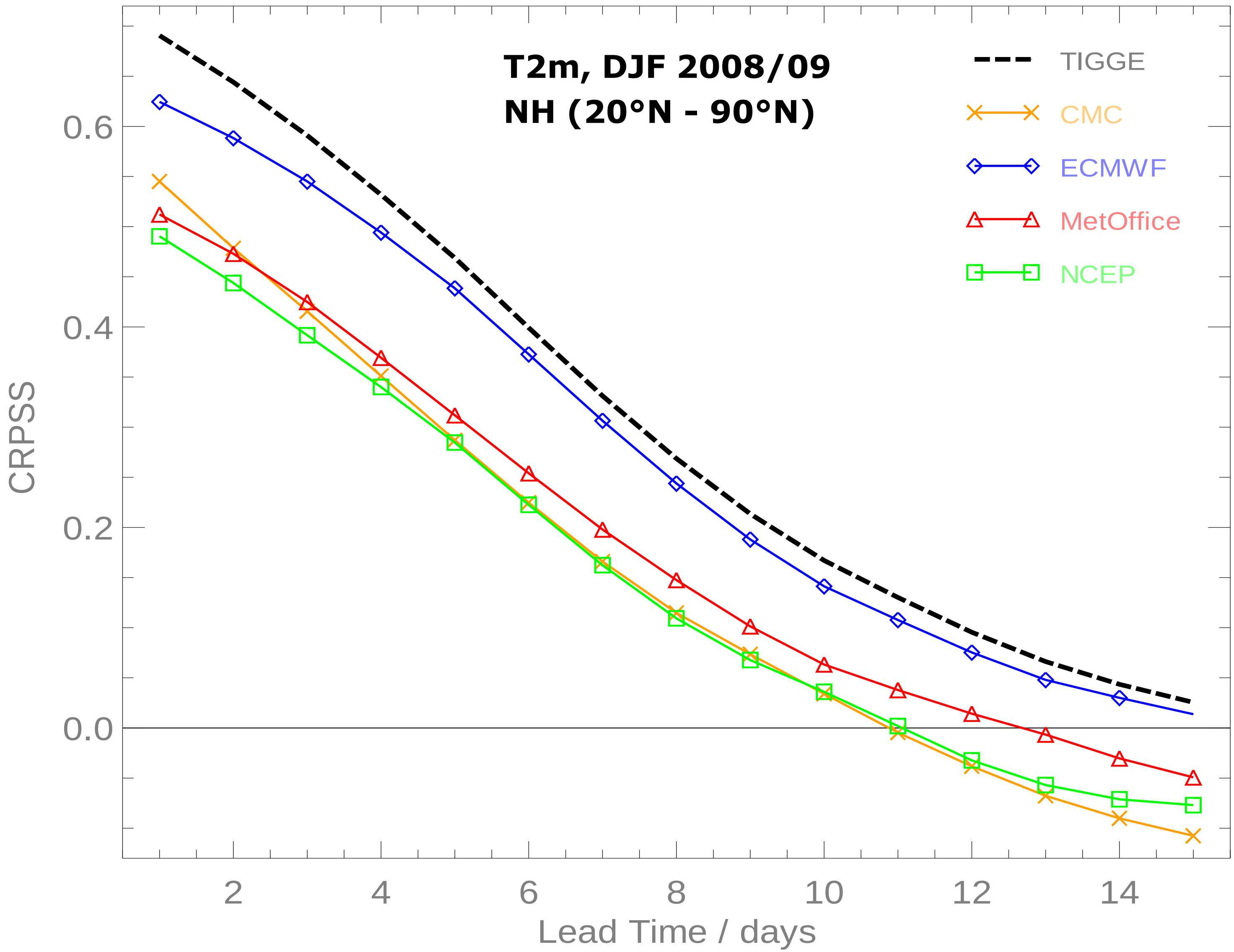
or

Can the TIGGE multi-model beat the (reforecast-calibrated) ECMWF EPS?

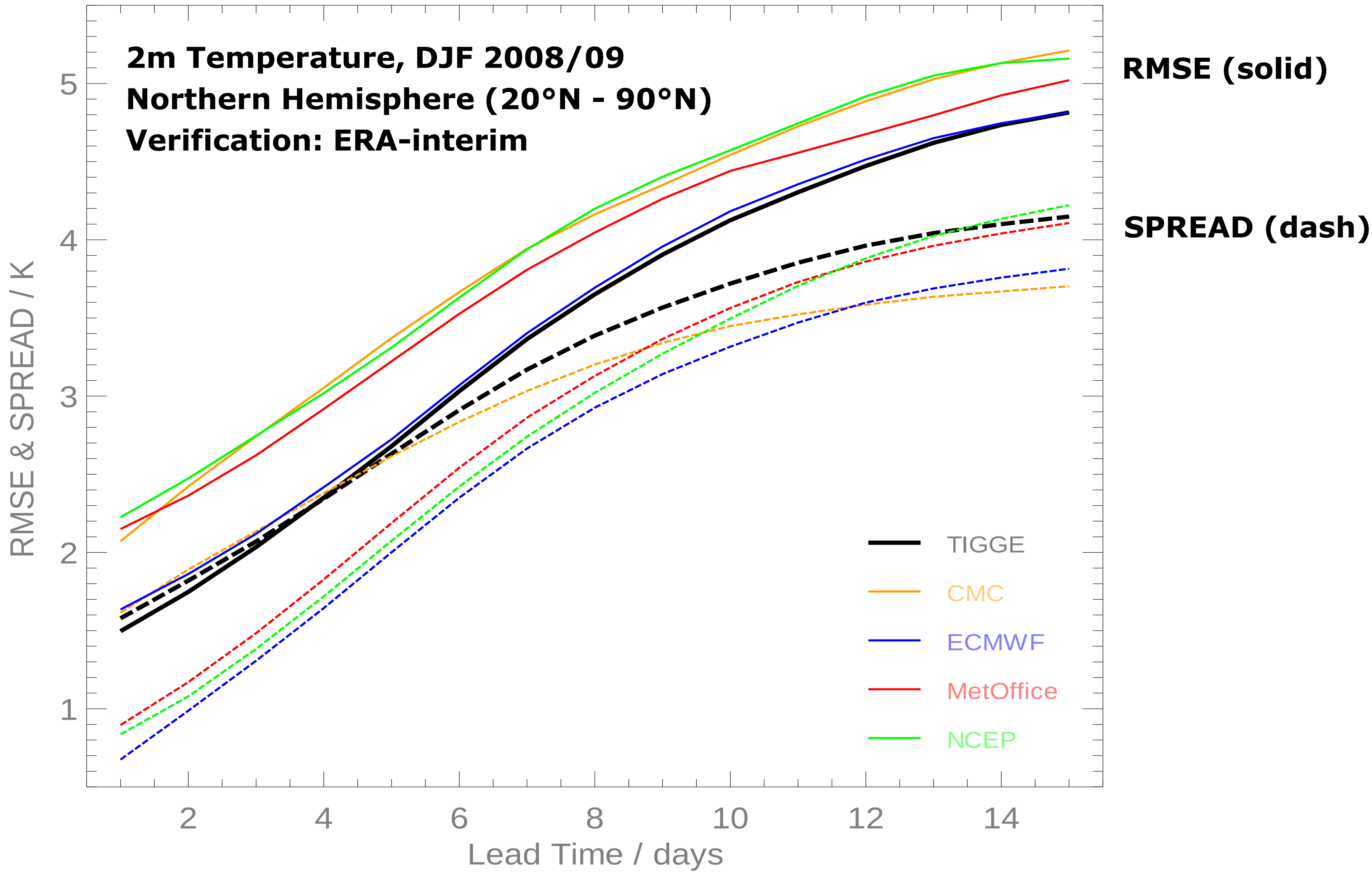
Comparing 9 TIGGE models & the MM



Comparing 4 TIGGE models & the MM



Mechanism behind improvements

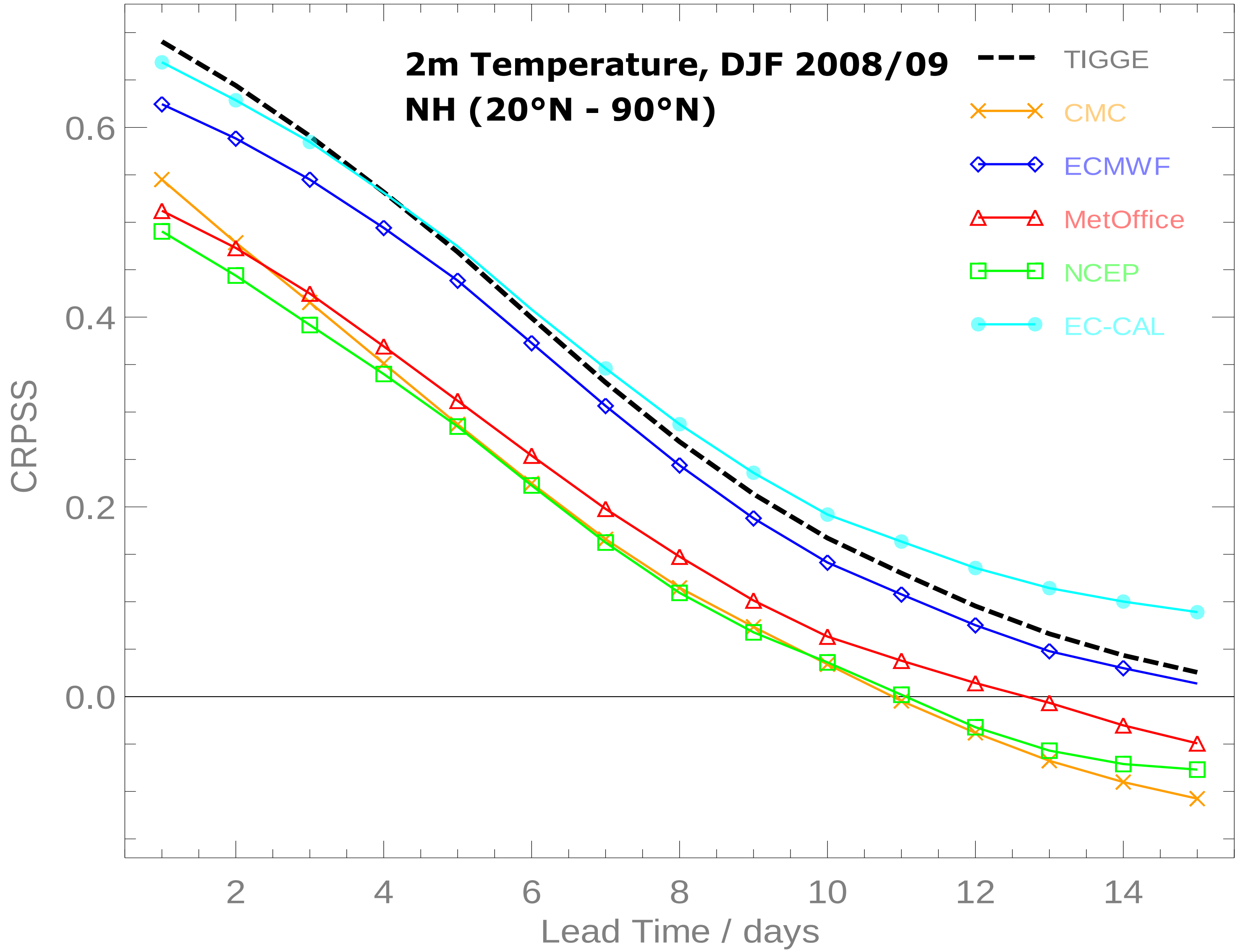


1st answer

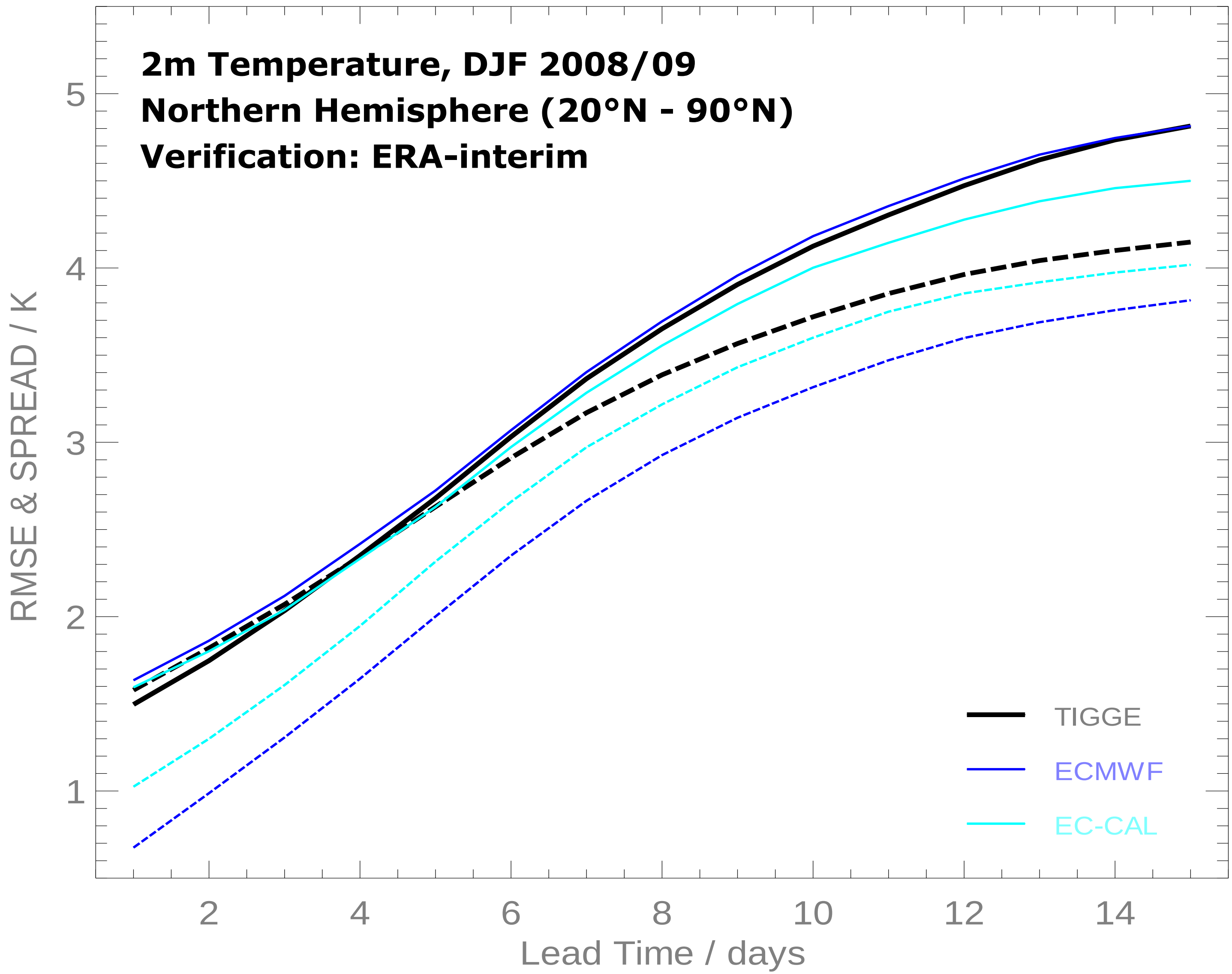
Does the multi-model concept work for medium-range weather forecasts?

Yes, but we are more skilful if we remove the least skilful components from the multi-model

Comparing 4 TIGGE models, MM, EC-CAL



Mechanism behind improvements



RMSE (solid)

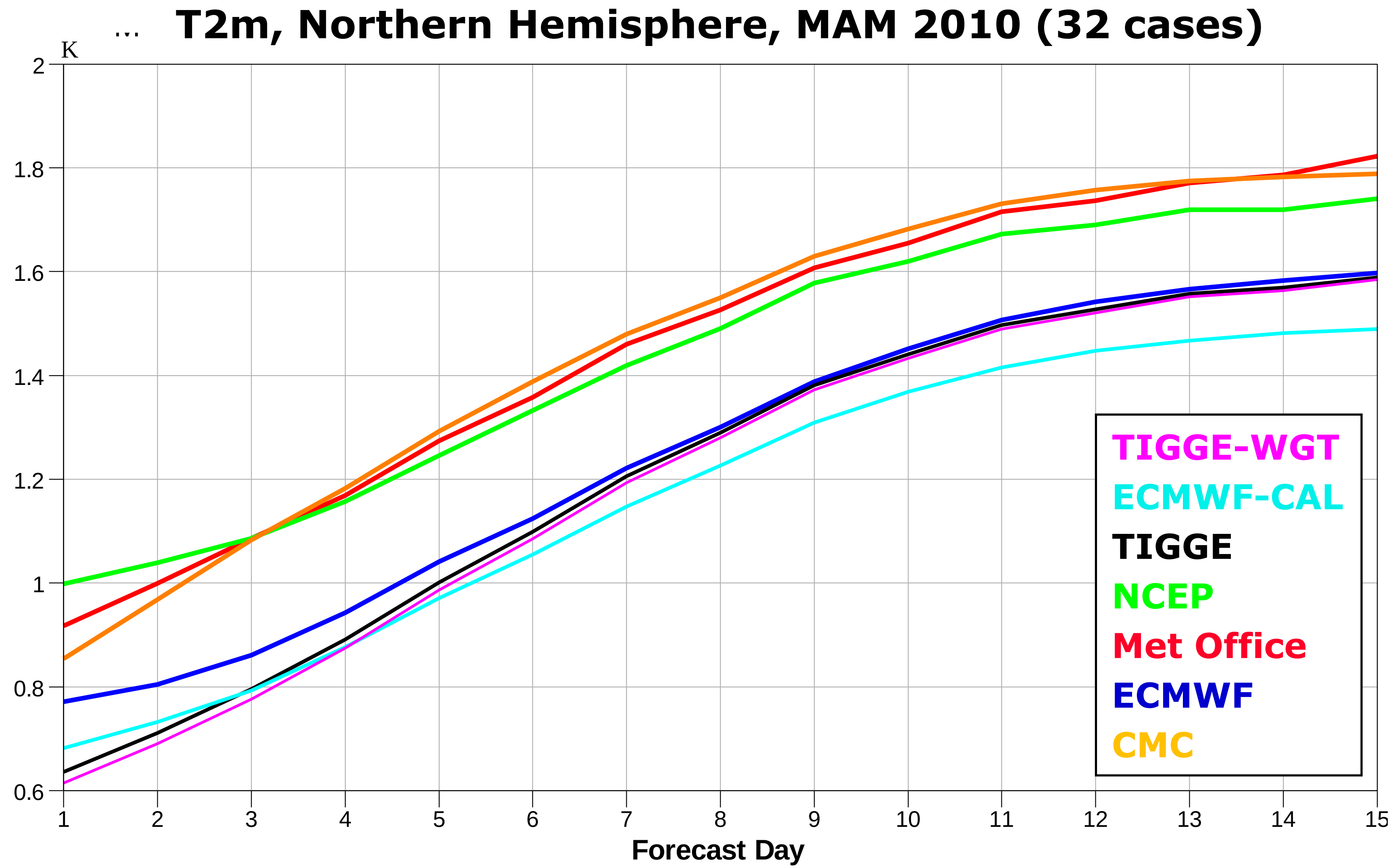
SPREAD (dash)

2nd answer

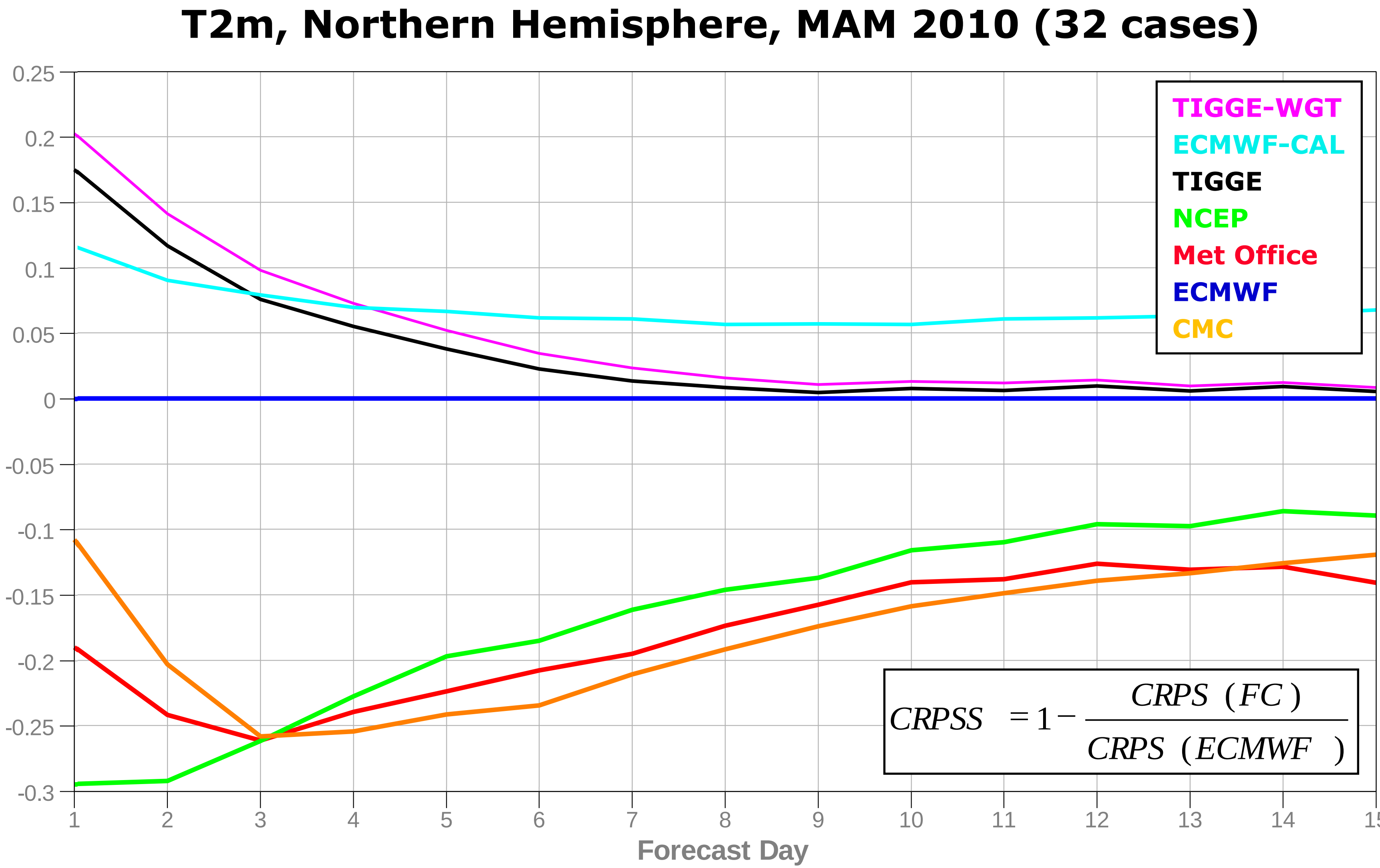
Can the TIGGE multi-model beat the (reforecast-calibrated) ECMWF EPS?

Not really, but maybe if we can improve the multi-model by weighting its components?

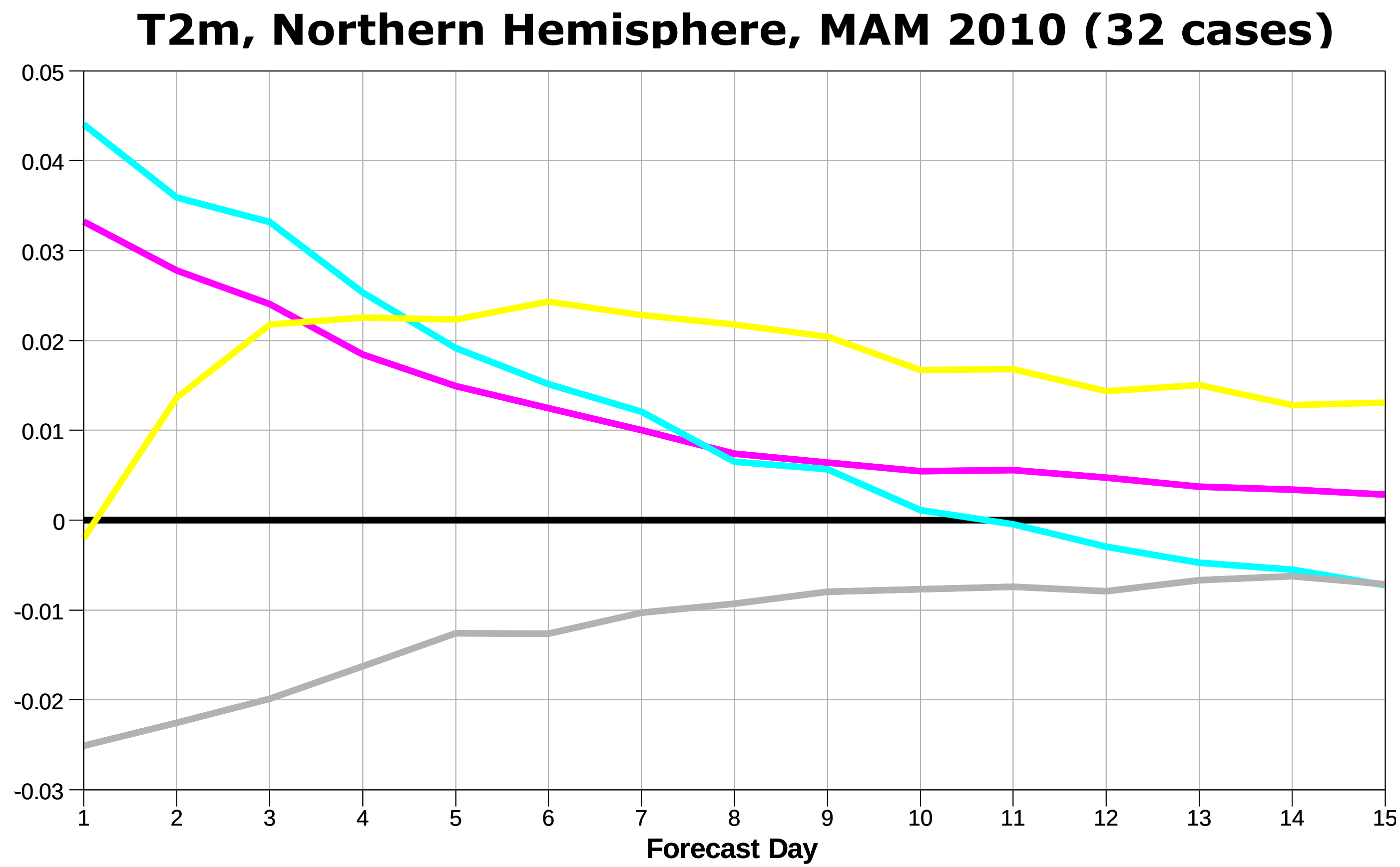
CRPS of 2m Temperature



CRPSS with ECMWF as reference



CRPSS with “equal weight” as reference



W1: ~ 1/MSE

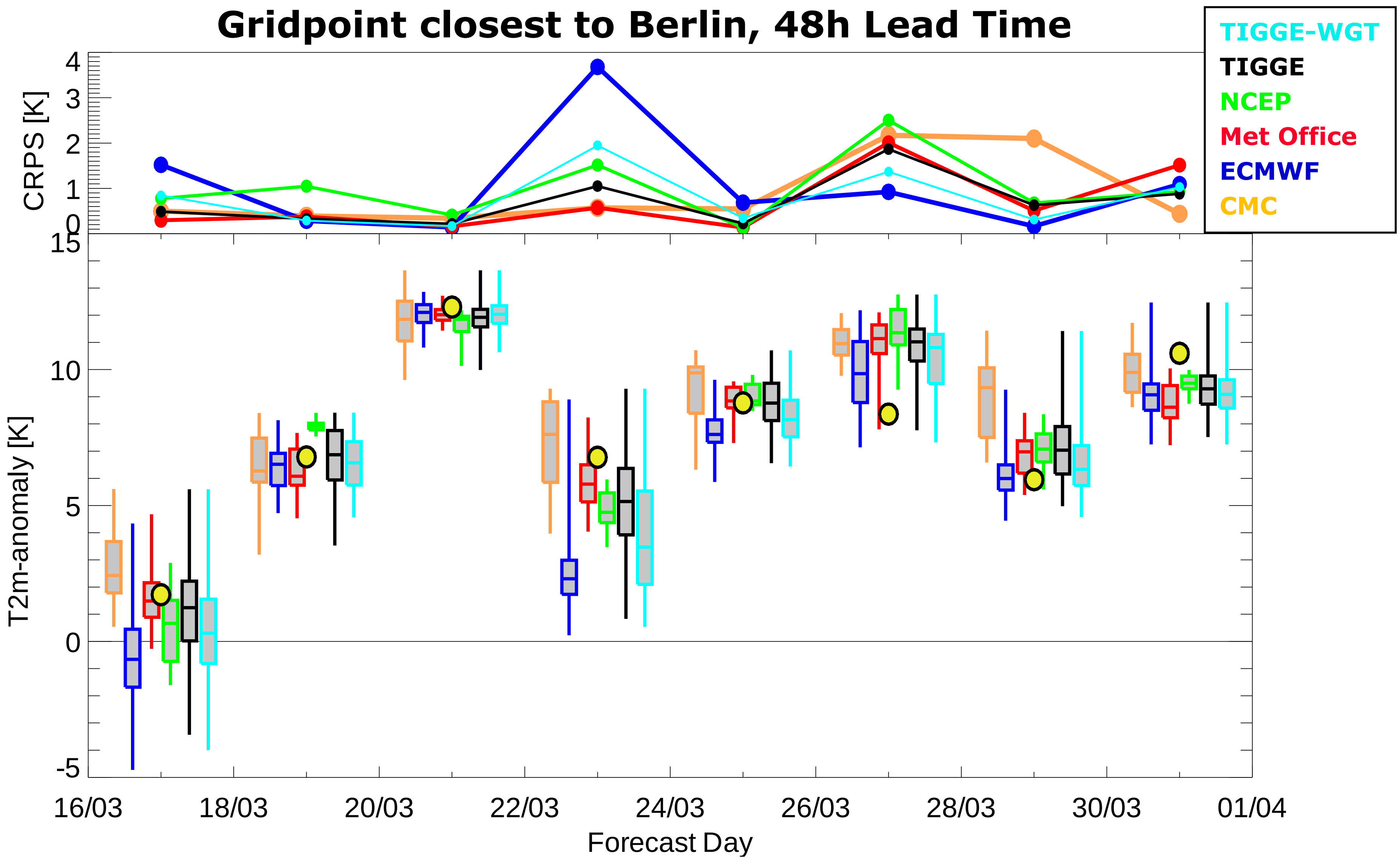
W2: optimized with respect to CRPS

W3: random weights

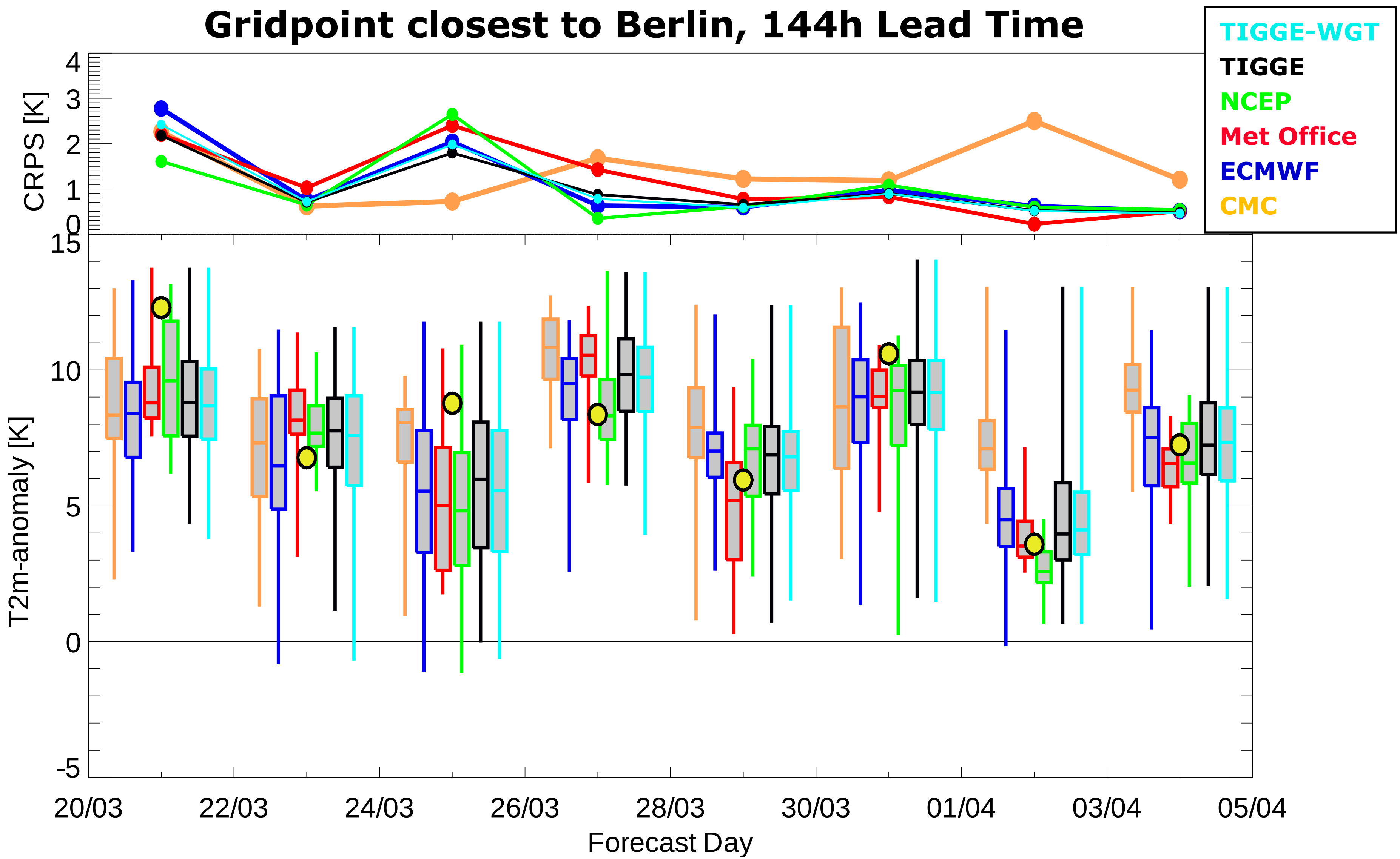
W4: constant weights **(0.1,0.6,0.2,0.1)**

$$CRPSS = 1 - \frac{CRPS\ (TIGGE\_Wx)}{CRPS\ (TIGGE\_EW)}$$

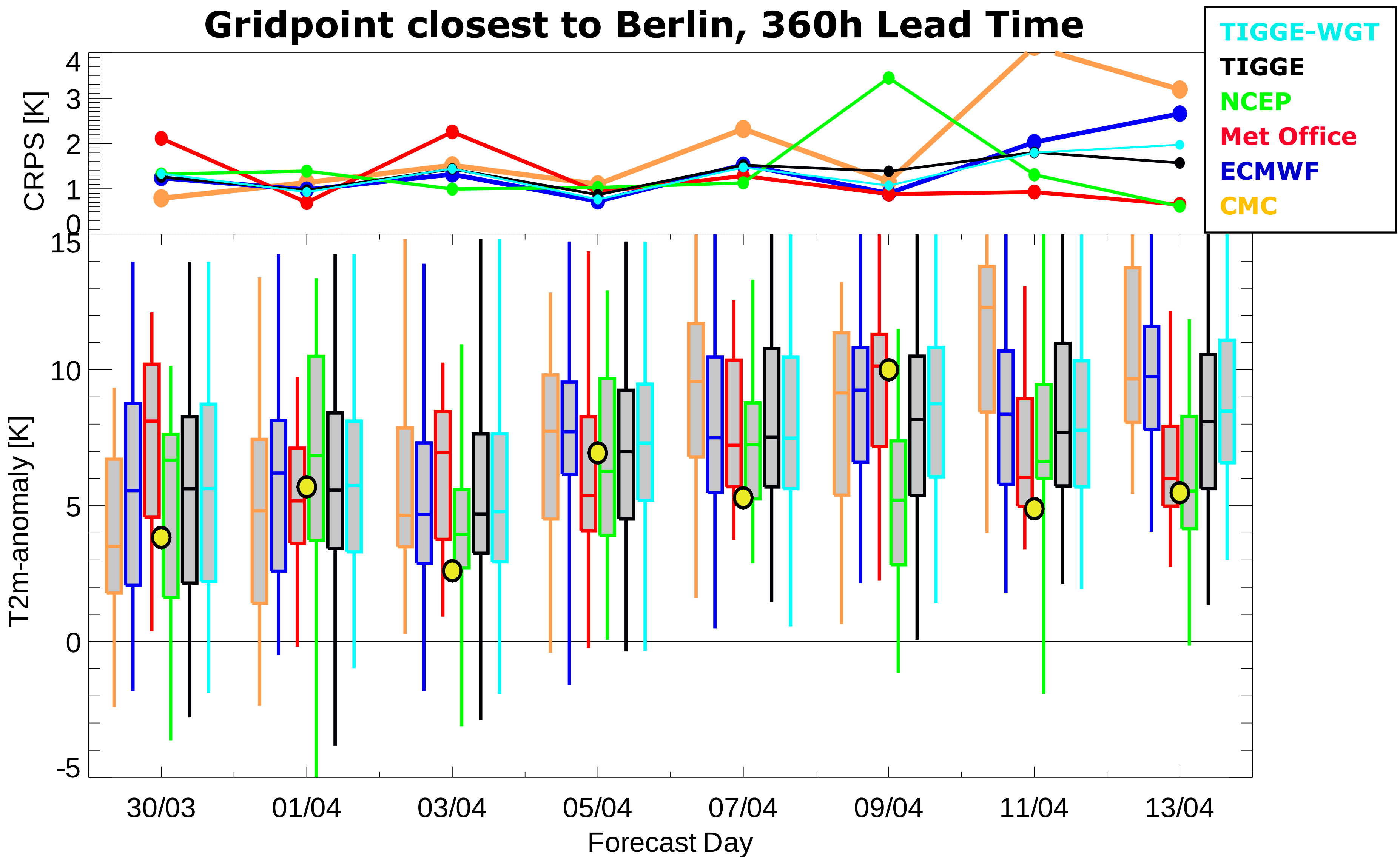
Impact of weighting for individual cases



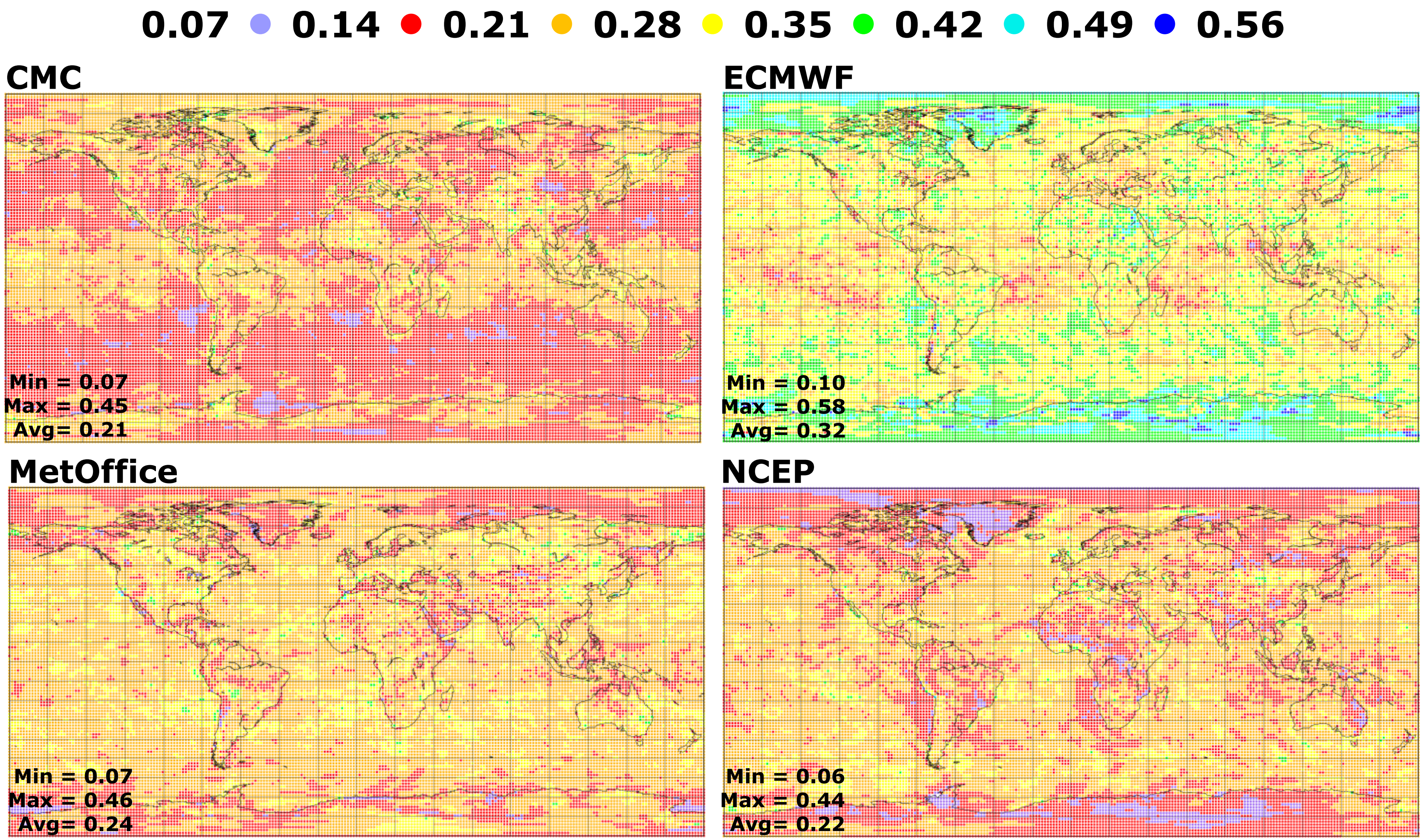
Impact of weighting for individual cases



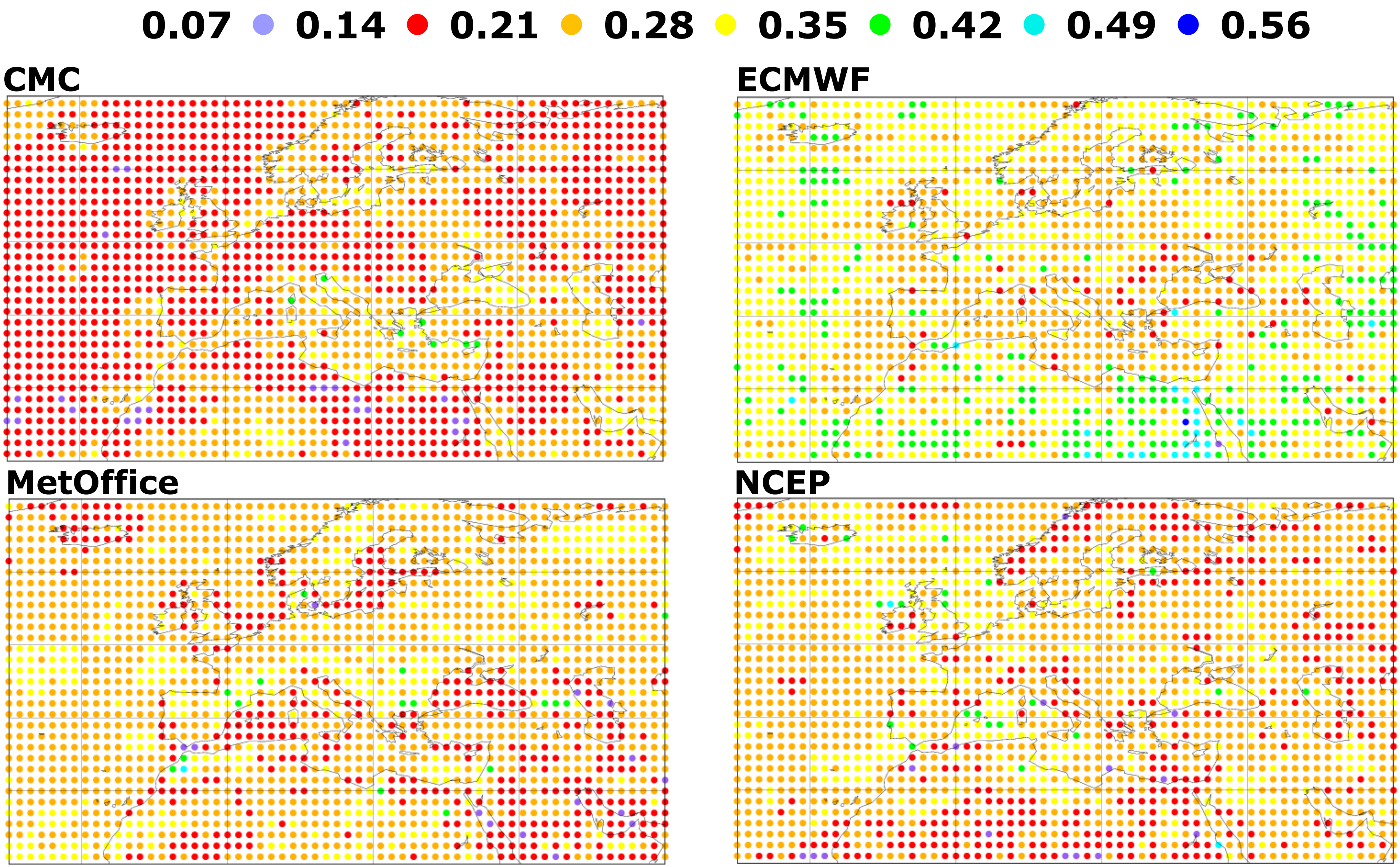
Impact of weighting for individual cases



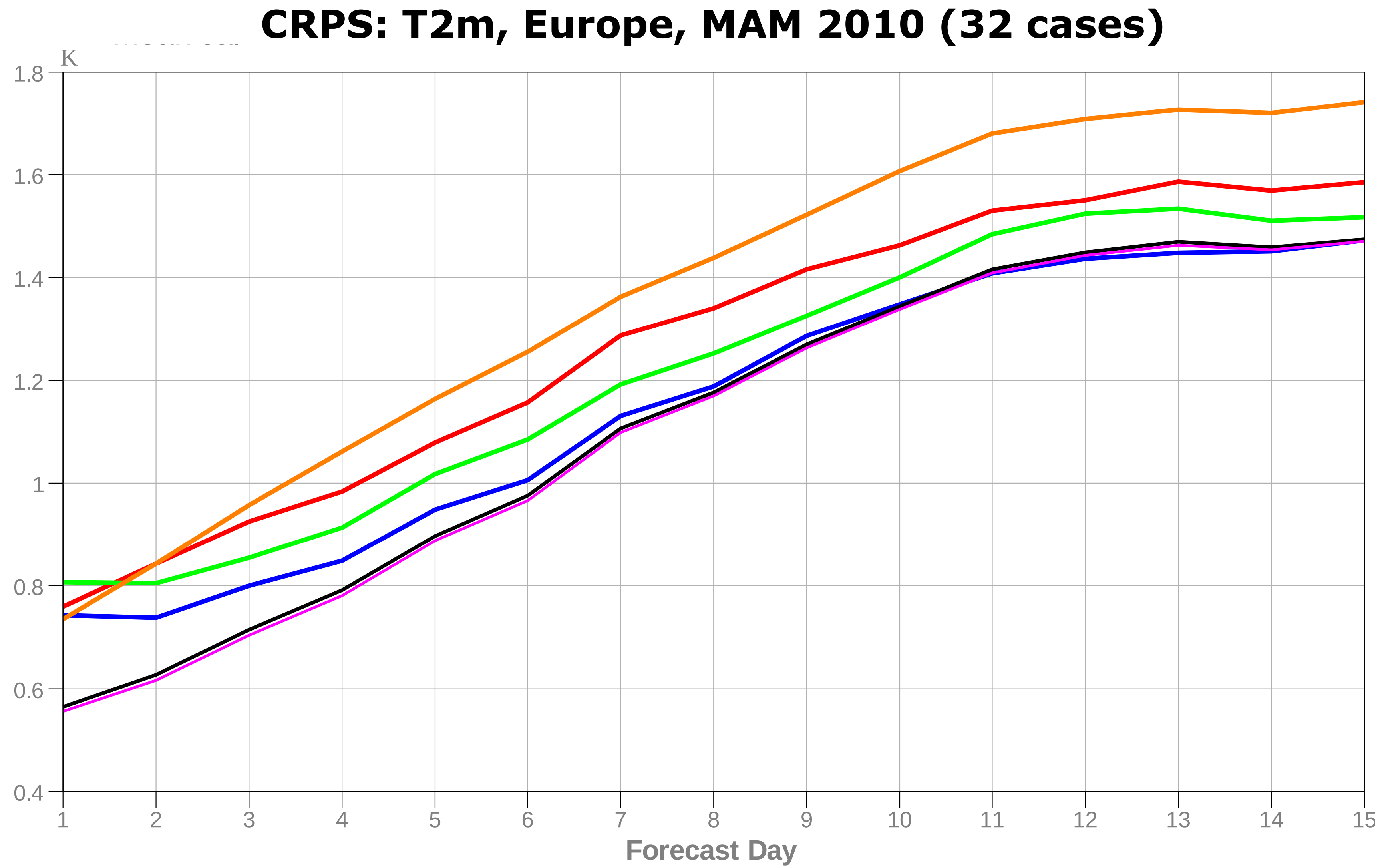
Seasonal average weights @48h



Seasonal average weights @48h

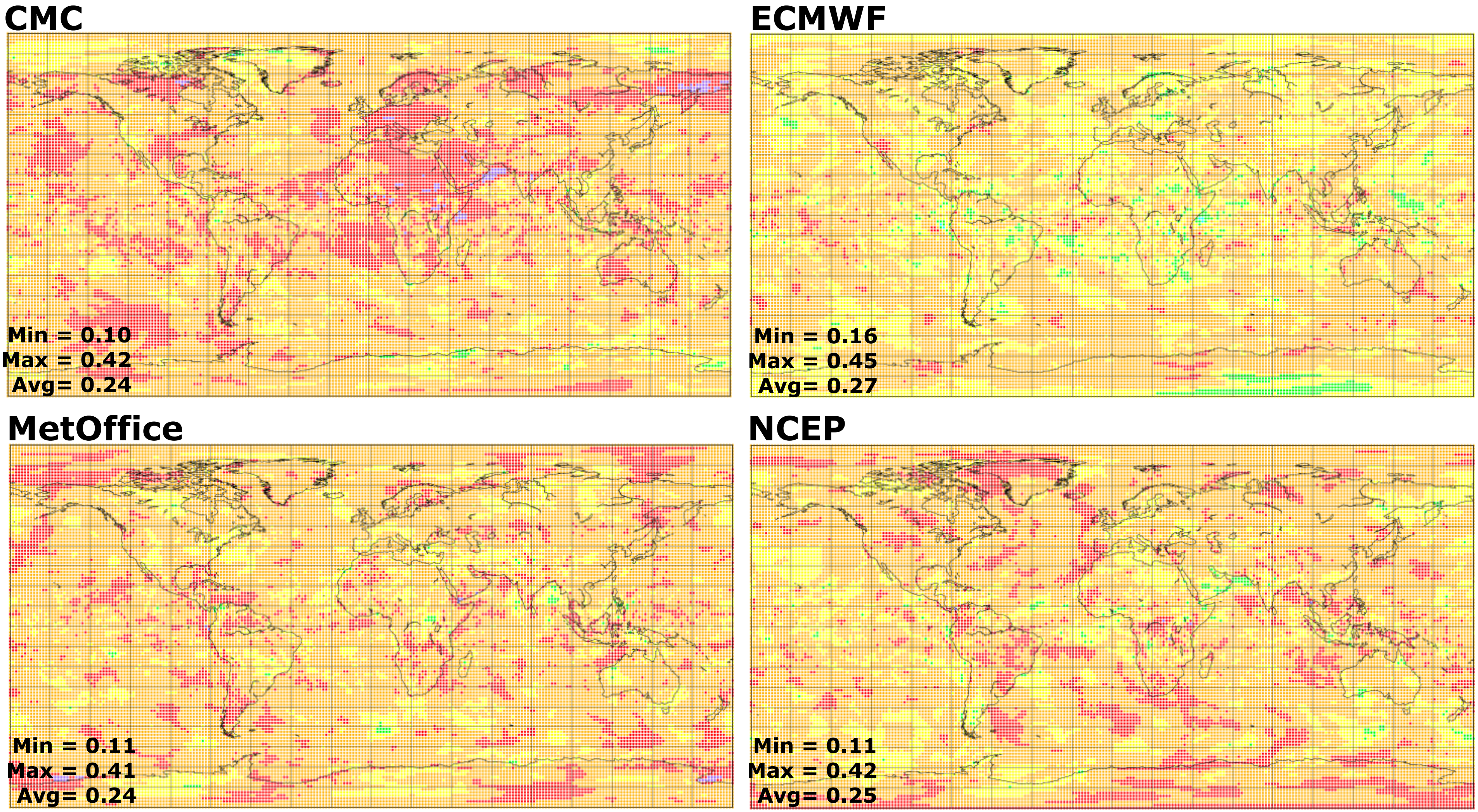


Performance over Europe

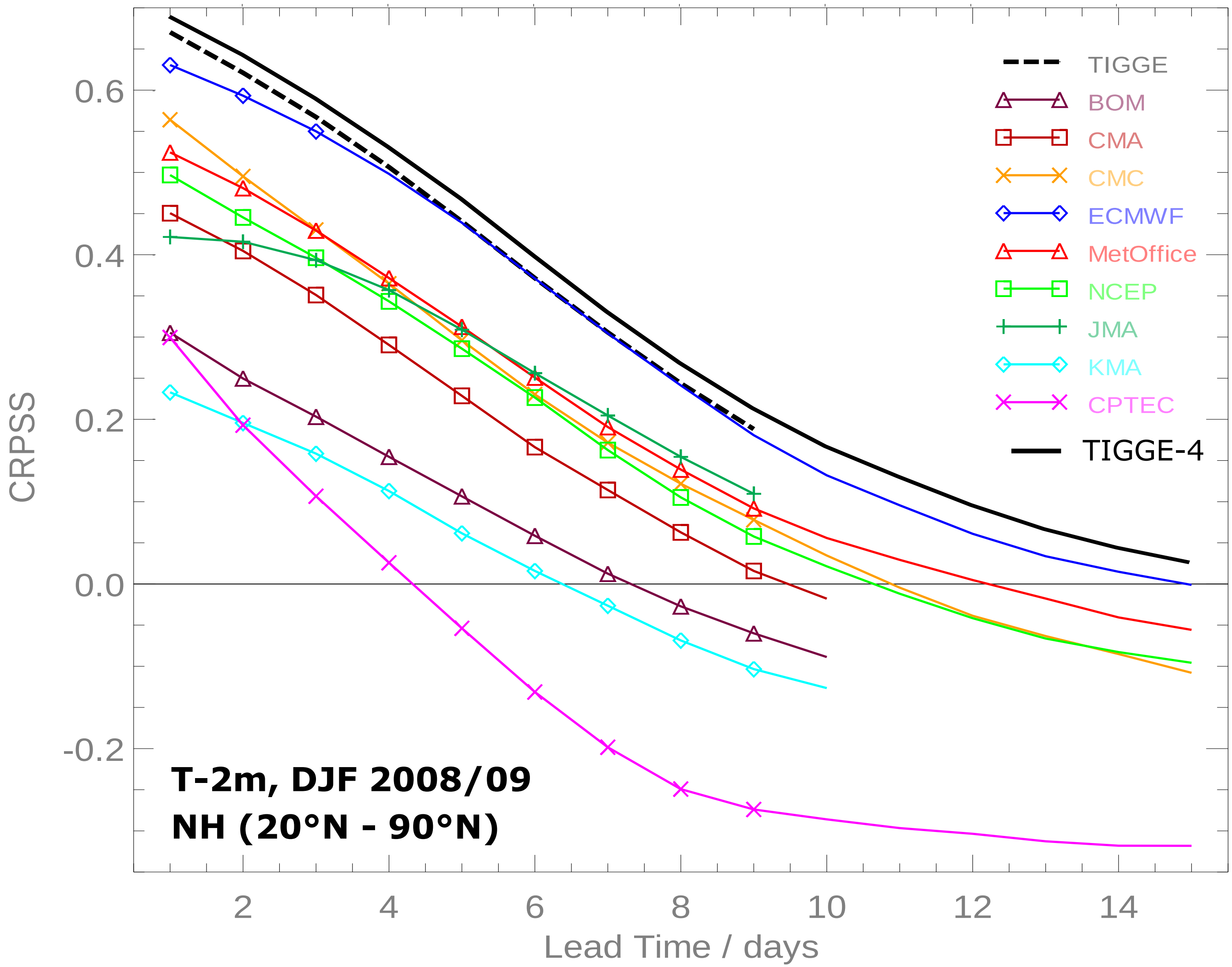


Seasonal average weights @360h

0.07 ● 0.14 ● 0.21 ● 0.28 ● 0.35 ● 0.42 ● 0.49 ● 0.56



Impact of zero weight for 5 out of 9



Conclusions

- MM-concept does work for medium-range weather forecasting
  - it improves on (overconfident) single-model forecasts
  - it is of comparable quality to re-forecast calibrated ECMWF EPS, without the drawback of necessarily being based on synthetic ensembles
- Improvements through:
  - Improved spread-error characteristic
  - Reducing ensemble mean error (MM: early FC range, EC-CAL: longer FC range)
- Weighting MM's individual components leads to only marginal improvements (see also Johnson & Swinbank, 2009)
  - No stable error characteristic can be detected, single-model forecasts too similar
  - Effort of calculating weights not worthwhile, especially considering the drawback of synthetic ensembles
  - This might change when considering other variables like e.g. precipitation

The future...

- Monitor performance of different models to detect if skill of individual models change

- For applications needing local information, add value to forecasts by
  - Applying a suite of post-processing methods to the DMO
  - Working with users, i.e. move on from simple examples and ideal scenarios to "real life" applications and face the reality of forecasters and users

- Where should this work be done? Who should do this work?
  - Directly at the source of the forecasts or individually for/by users?