# On the value of reforecasts for the TIGGE database

Tom Hamill
NOAA/ESRL/PSD

Renate Hagedorn
European Centre for Medium-Range Weather Forecasts

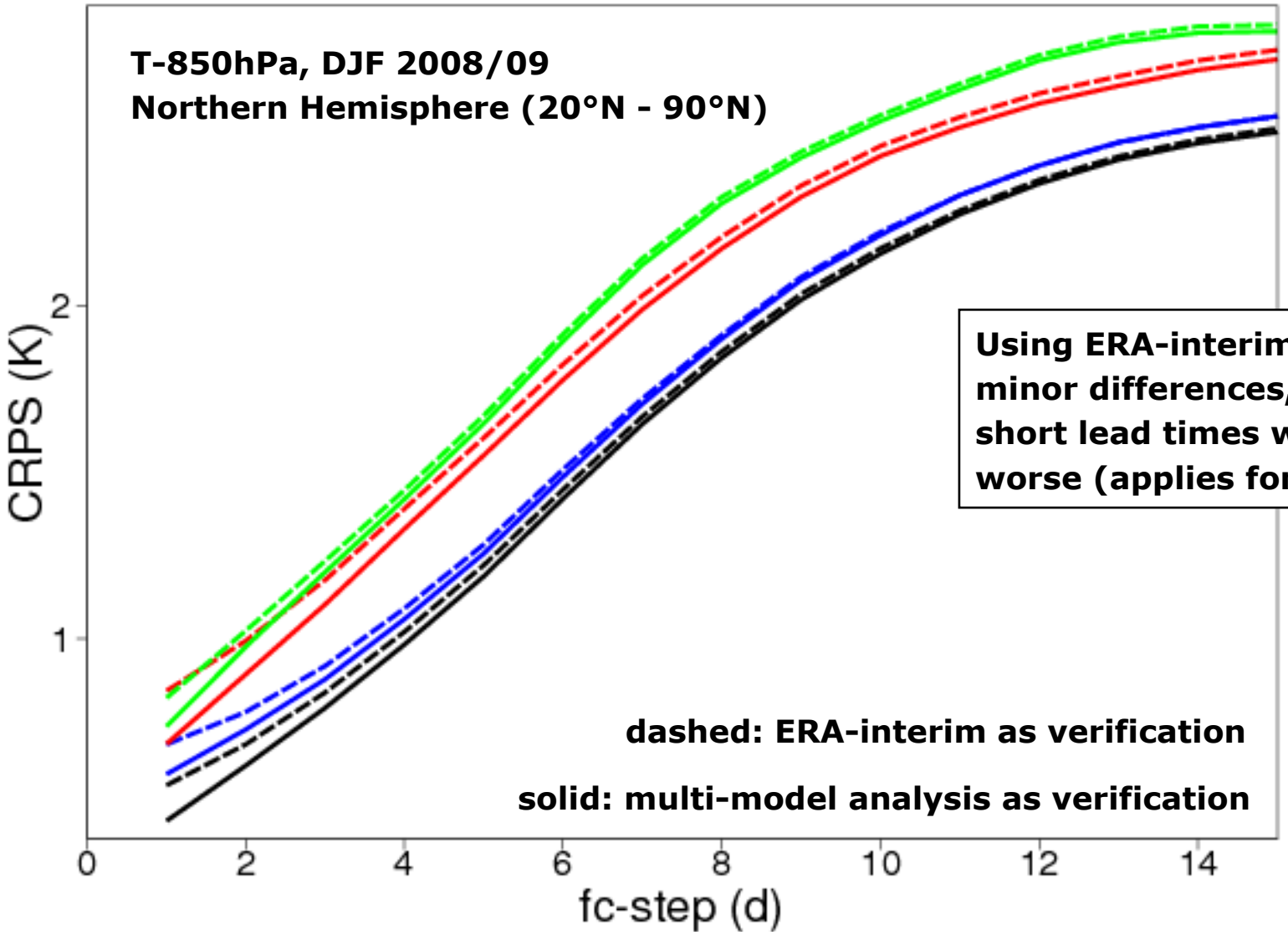
# Motivation

- One goal of TIGGE is to investigate whether multi-model predictions are an improvement to single model forecasts
- The goal of using reforecasts to calibrate single model forecasts is to provide improved predictions
- Questions:
  - What are the relative benefits (costs) of both approaches?
  - What is the mechanism behind the improvements?
  - Which is the “better” approach?

# Possible verification datasets

- If we don't verify against model independent observations we need to agree on a **'fair'** but also **'most useful'** verification dataset
- Use each model's own analysis as verification
  - Multi-model has no "own analysis"
  - Intercomparison of skill scores "difficult" because reference forecast scores differently for different analysis
- Use a multi-model analysis as verification
  - Incorporating less accurate analyses might not necessarily lead to an analysis which is closest to reality
  - Calibration needs a consistent verification dataset used in both training and application phase, MM-analysis not available for reforecast training period
- Use "semi-independent" analysis: ERA-interim
  - Assumed to be as close as possible to reality
  - Available for long period in the past and near real-time
  - For upper air fields in Extra-Tropics close to analyses of best models / MM-analysis
  - For Tropics and near-surface fields use bias-corrected forecasts for 'fair' assessment

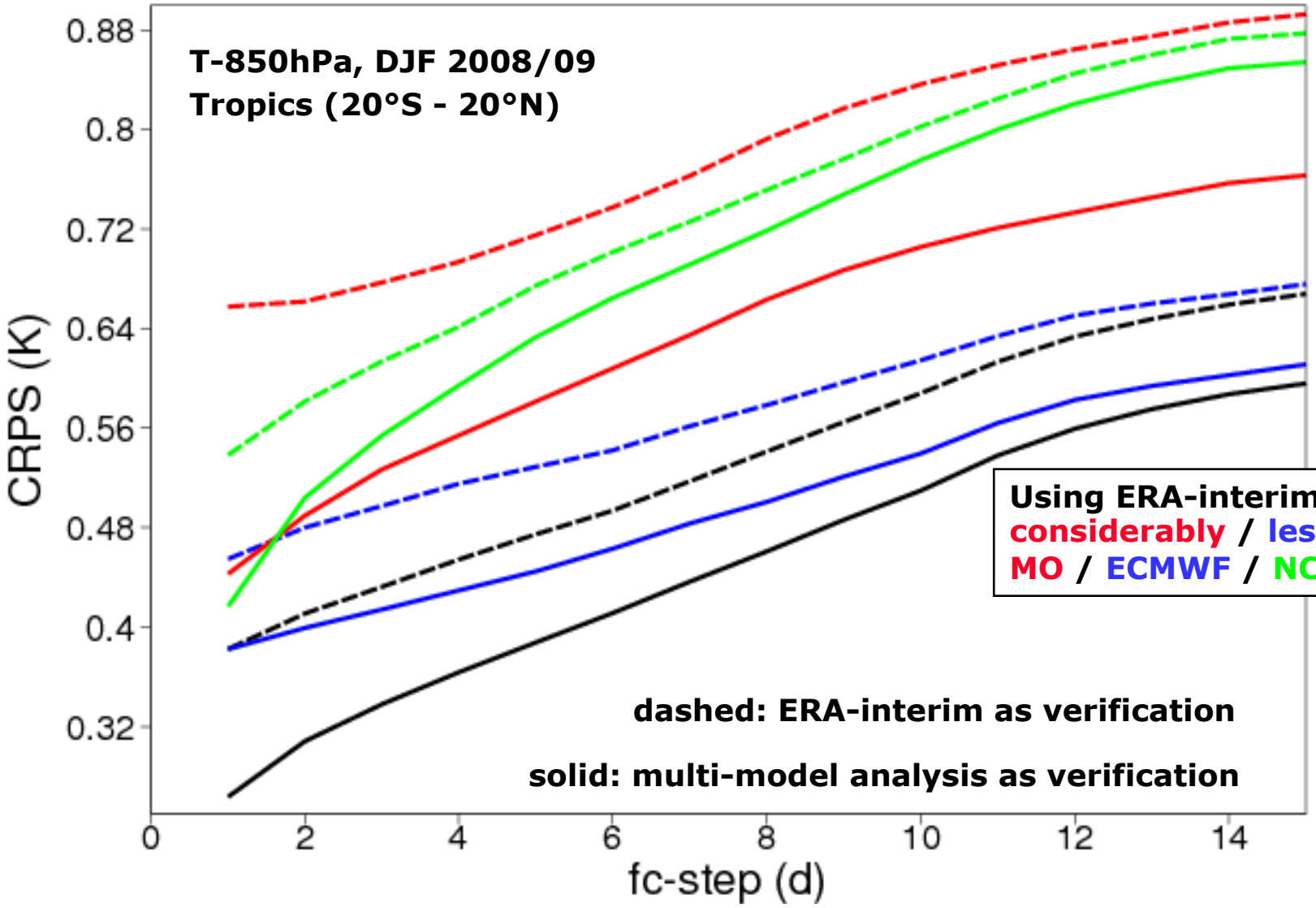
# Choice of analysis: upper air, extra-tropics




**Using ERA-interim leads to only minor differences, except for short lead times when scores get worse (applies for all models)**

**NCEP**
**Met Office**
**ECMWF**
**TIGGE**

# Choice of analysis: upper air, tropics




**Using ERA-interim worsens scores considerably / less / least for MO / ECMWF / NCEP**

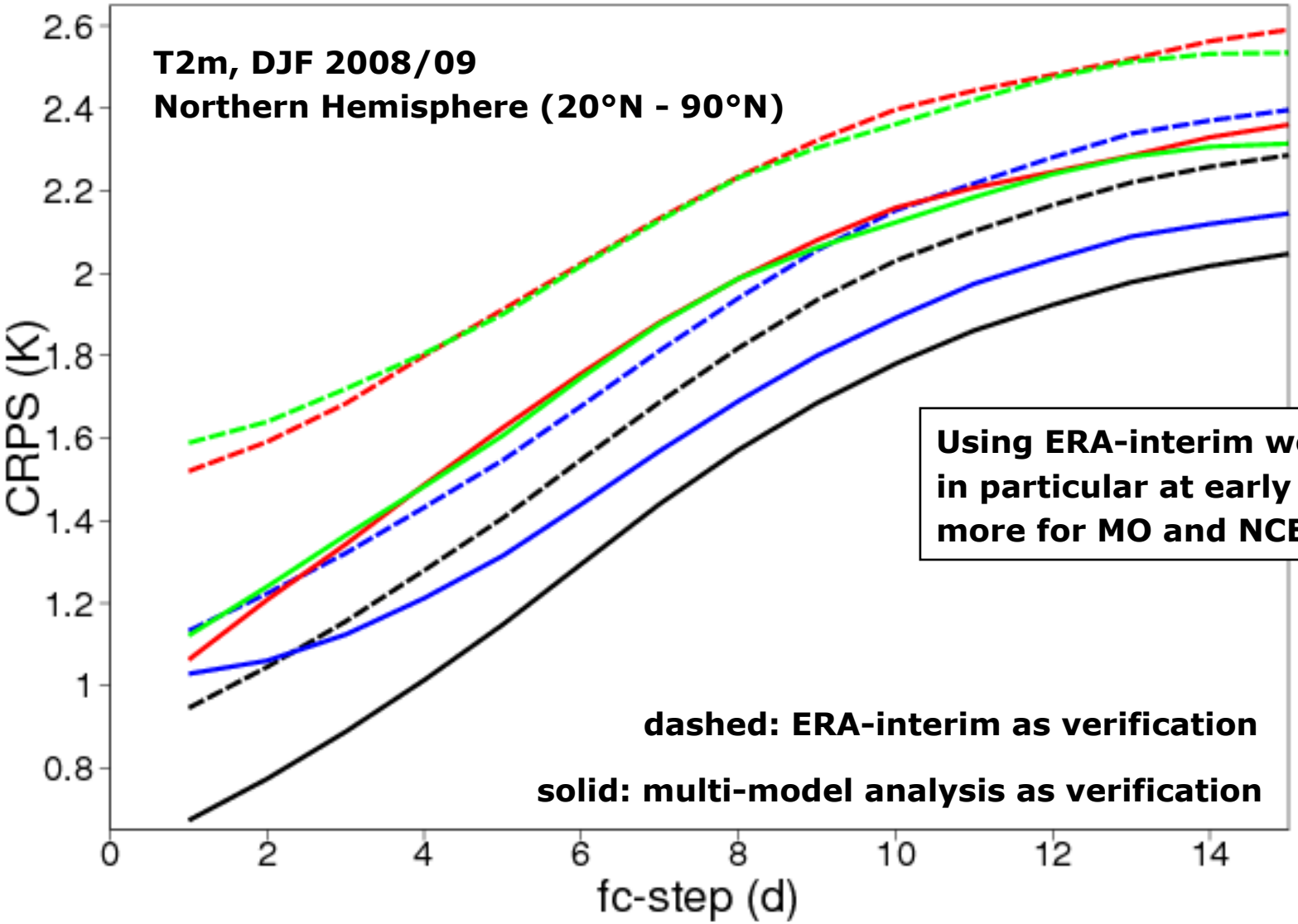
**NCEP**

**Met Office**

**ECMWF**

**TIGGE**

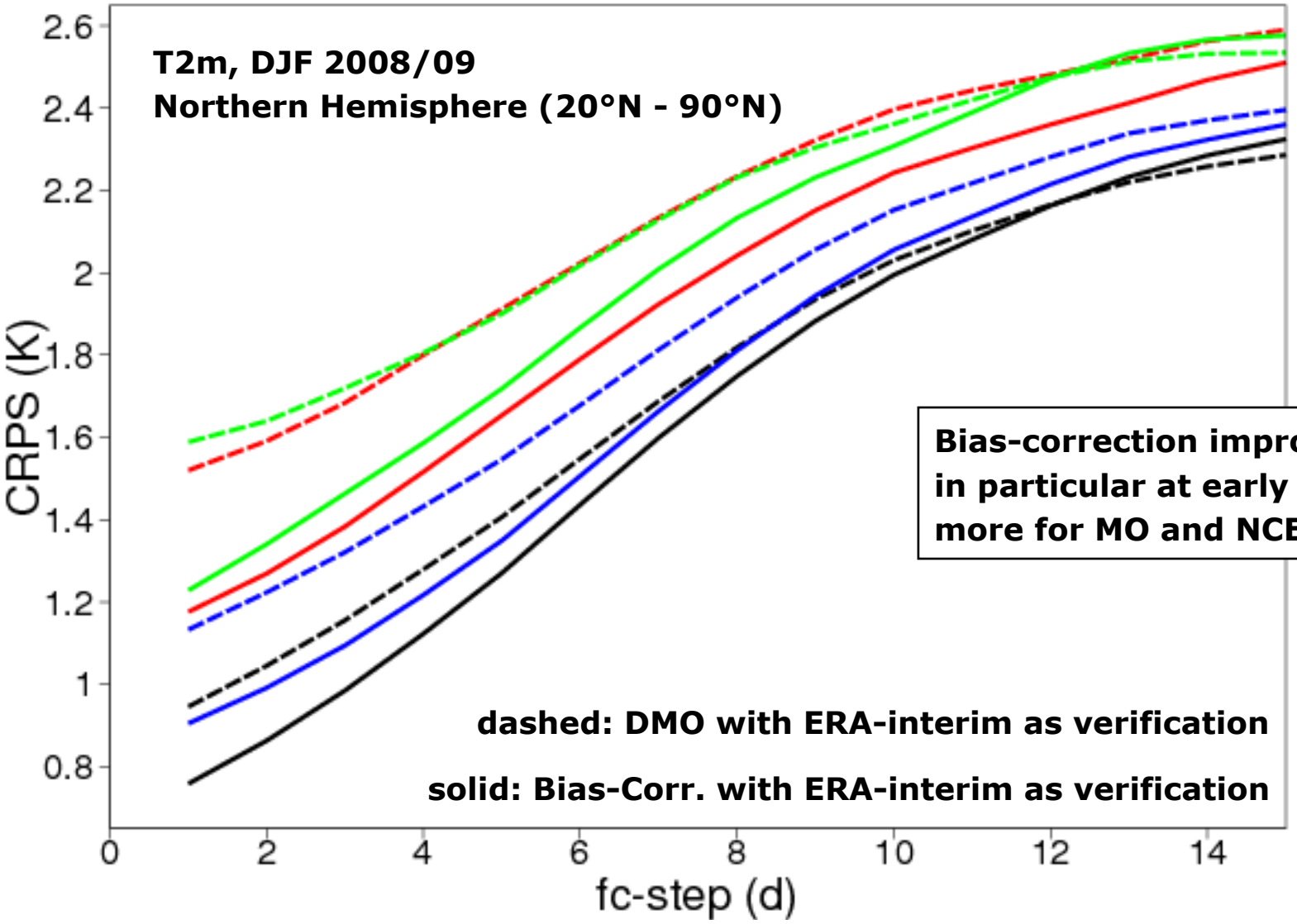
# Choice of analysis: surface




**Using ERA-interim worsens scores, in particular at early lead times, more for MO and NCEP, less for ECMWF**

**NCEP**
**Met Office**
**ECMWF**
**TIGGE**

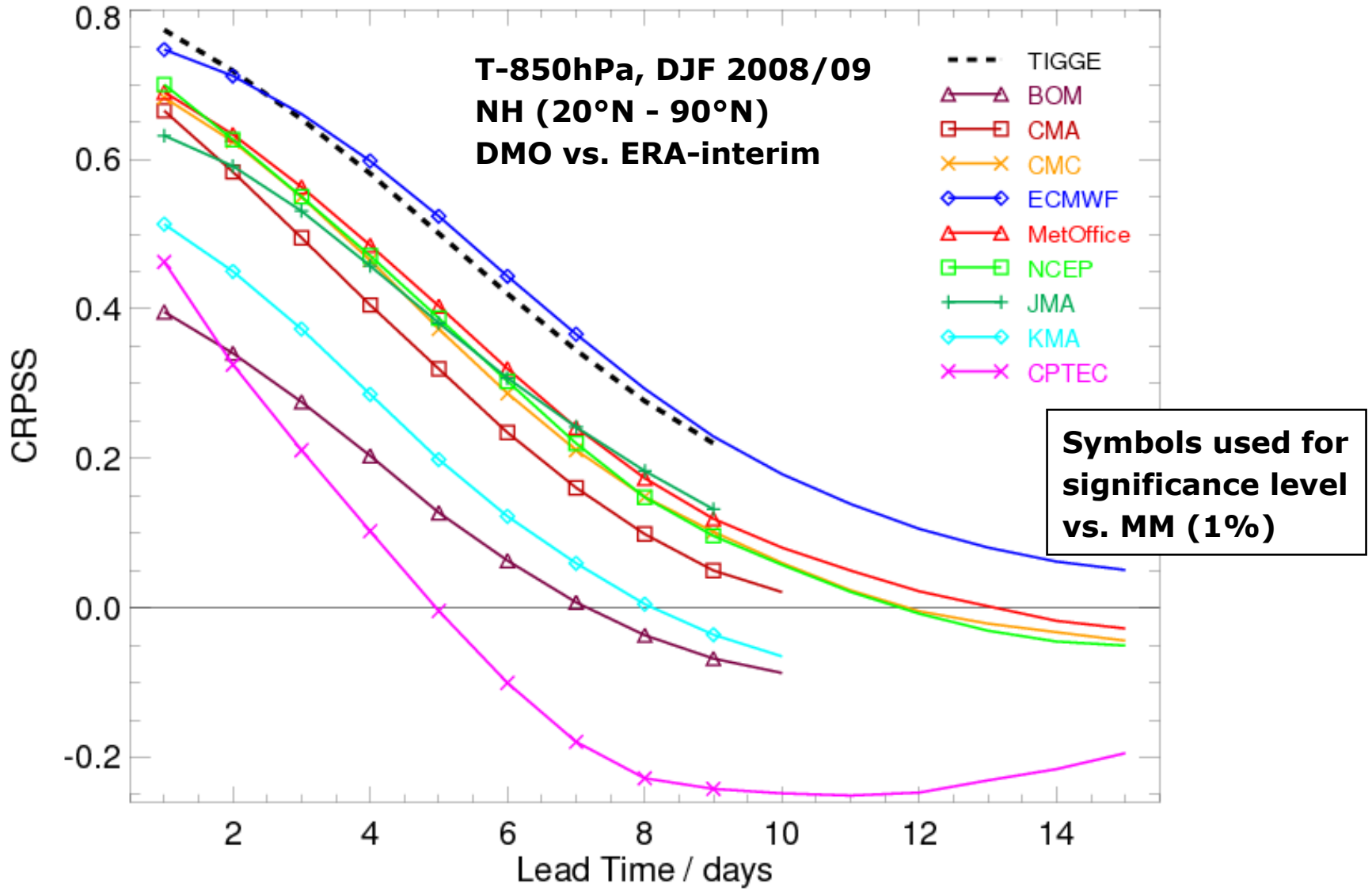
# Choice of analysis: surface, bias-corrected



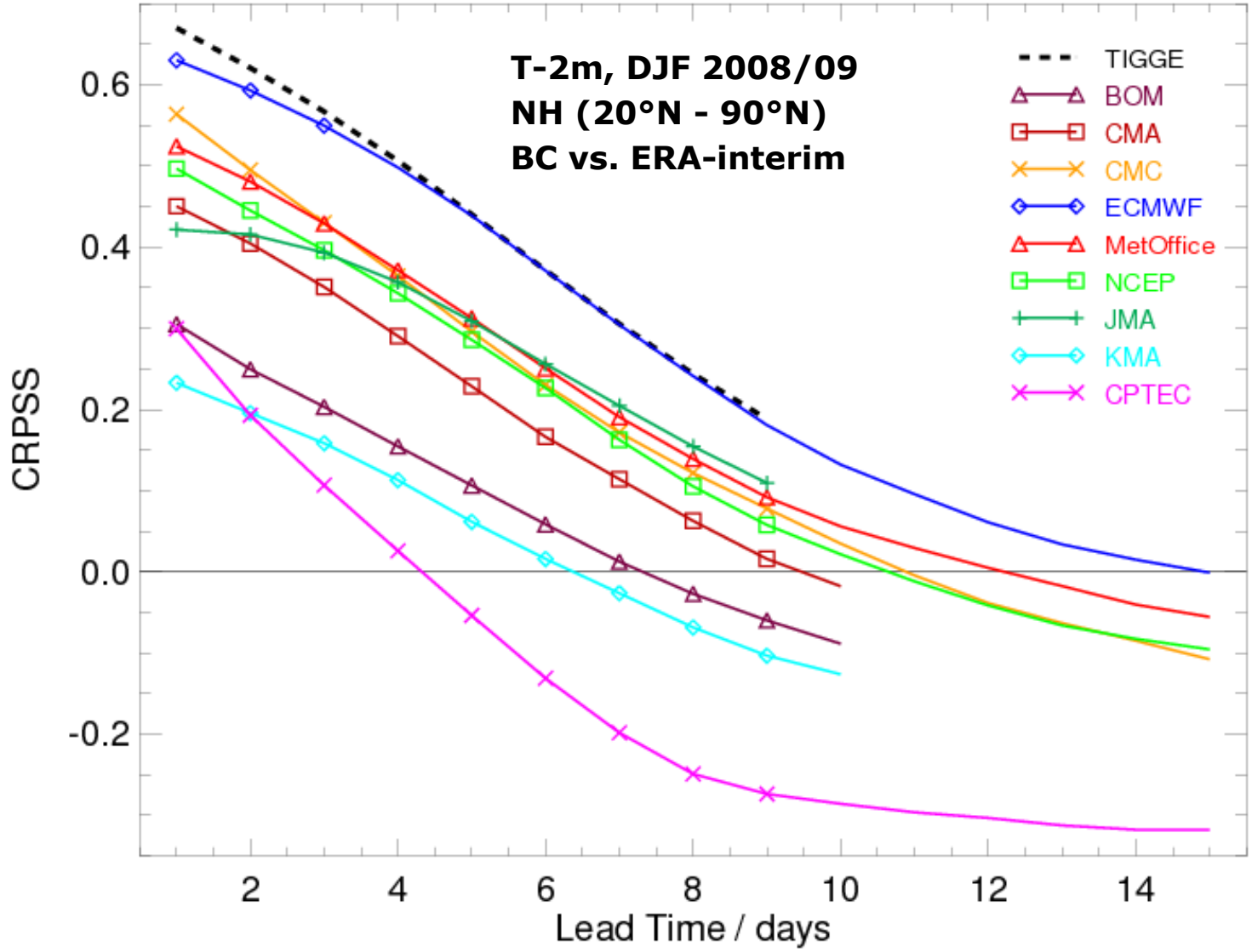

**Bias-correction improves scores, in particular at early lead times, more for MO and NCEP, less for ECMWF**

**NCEP**
**Met Office**
**ECMWF**
**TIGGE**

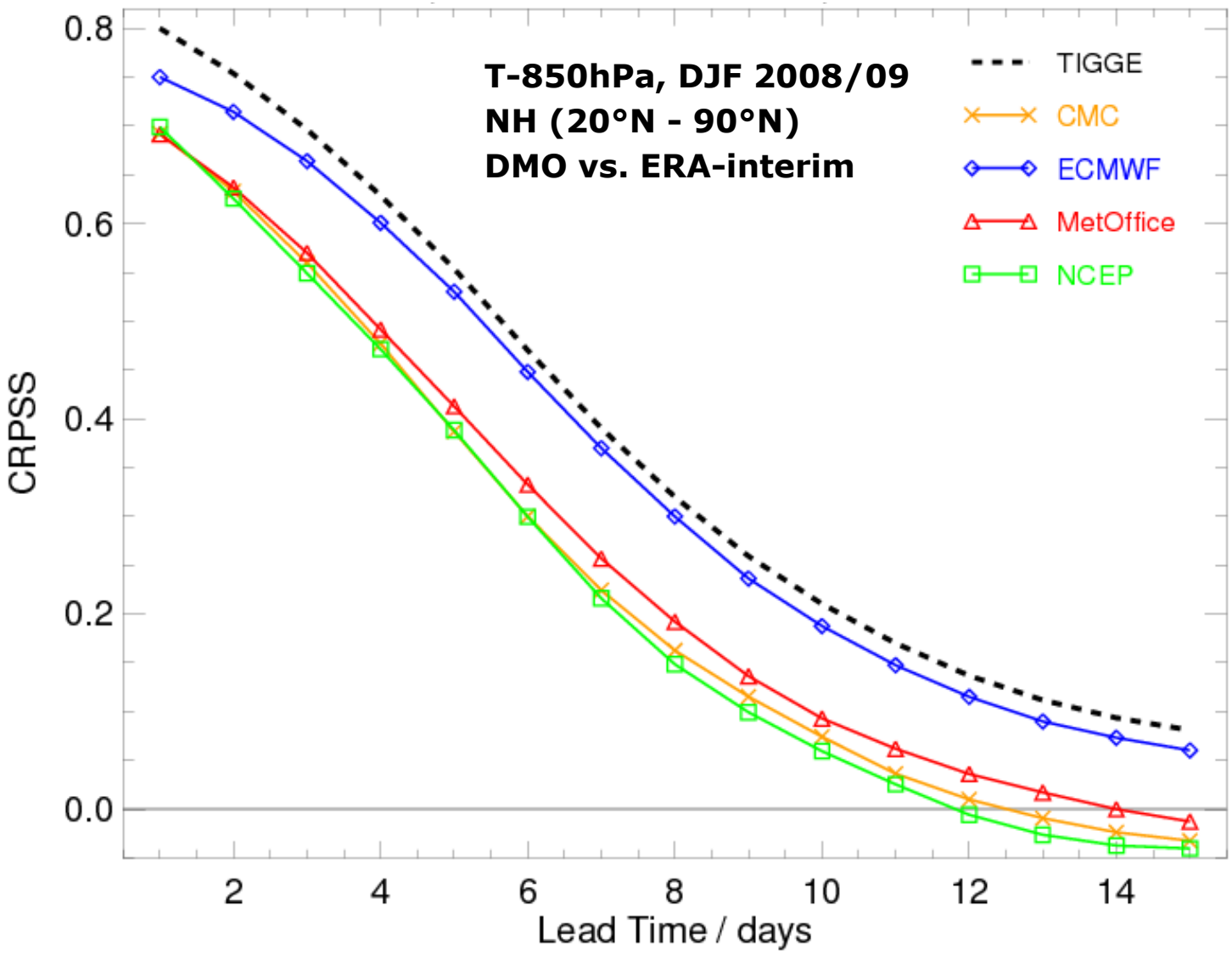
# Comparing 9 TIGGE models & the MM



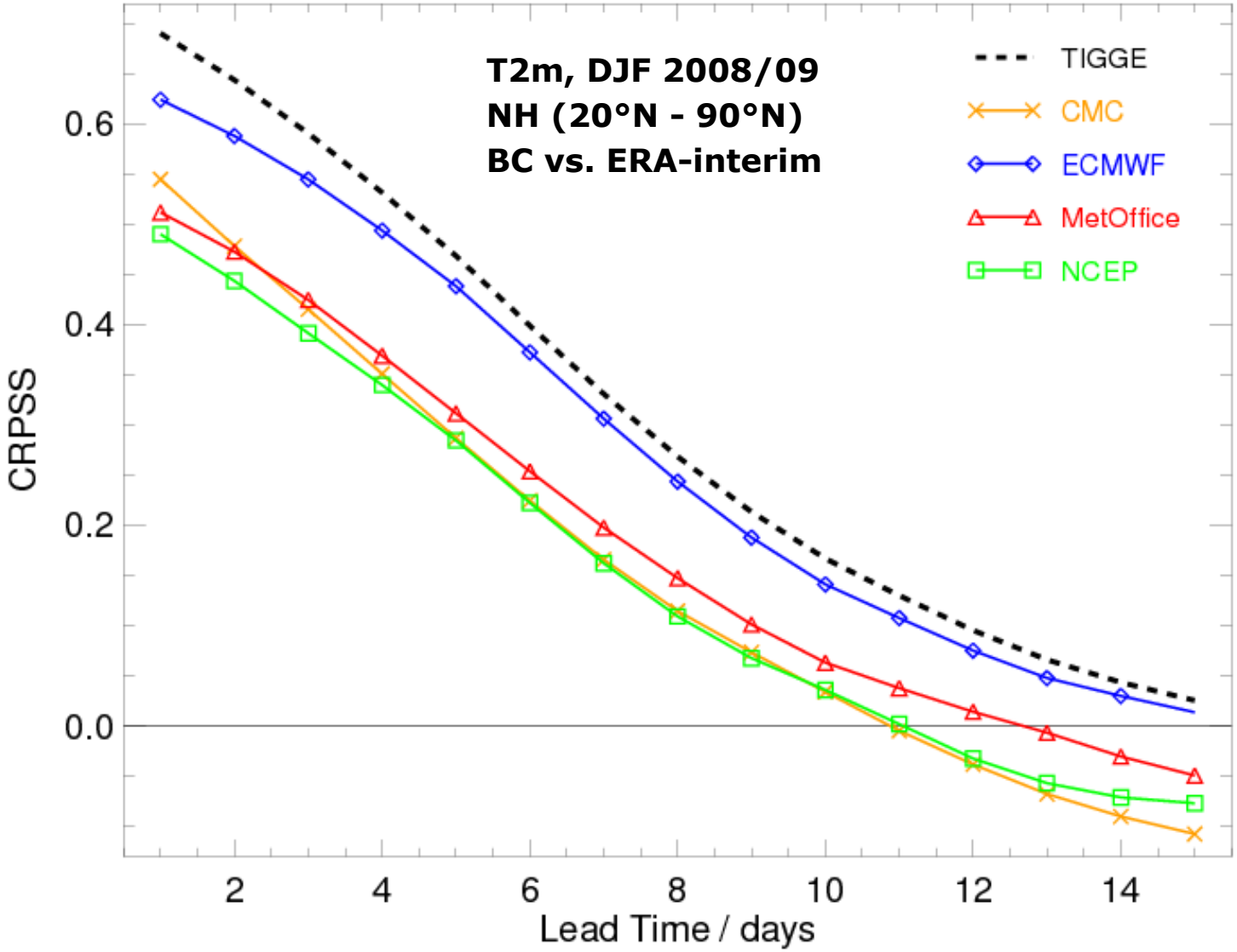
# Comparing 9 TIGGE models & the MM



# Comparing 4 TIGGE models & the MM



# Comparing 4 TIGGE models & the MM



# Calibration using reforecasts

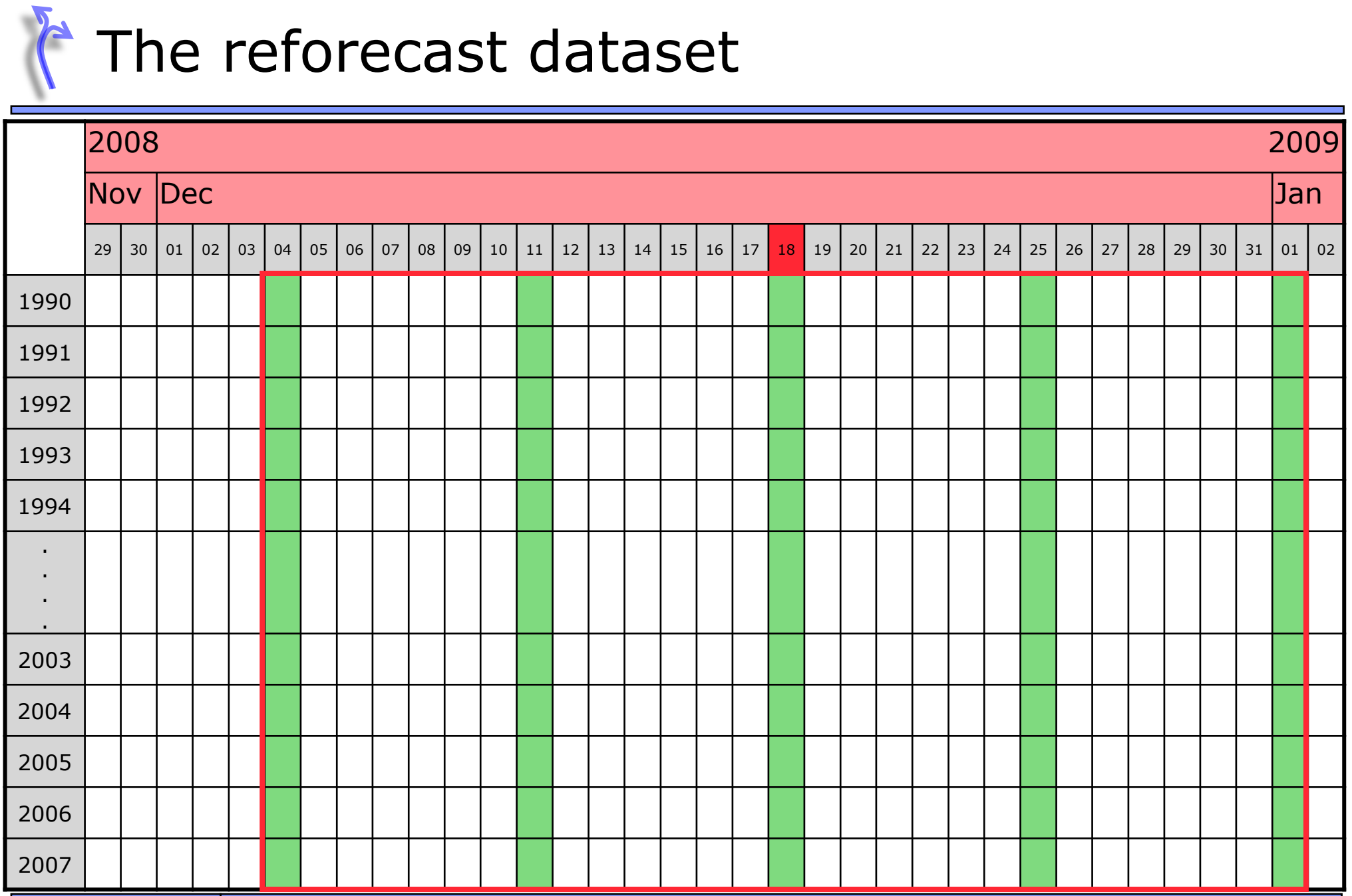
- All calibration methods need a training dataset, containing a number of forecast-observation pairs from the past
- Non-homogeneous Gaussian Regression (NGR) provides a Gaussian PDF based on the ensemble mean and variance of the raw forecast distribution

$$P(v \le q) = \Phi\left[\frac{q - (a + b\overline{x}_{ens})}{\sqrt{c + ds_{ens}^{\ 2}}}\right]$$

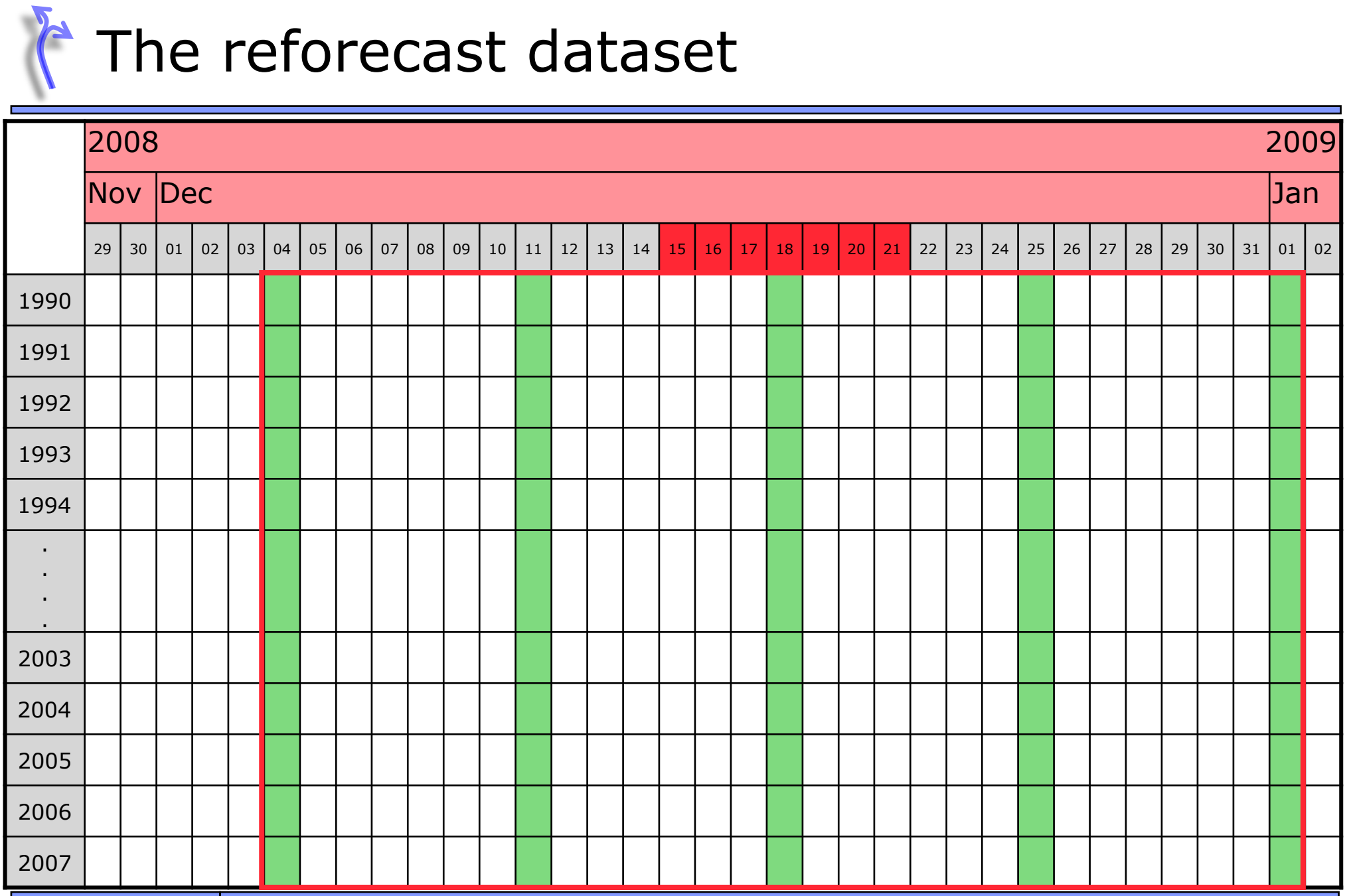
with: $\Phi$ = CDF of standard Gaussian distribution

- Calibration process:
  - Determine optimal calibration coefficients by minimizing CRPS for training dataset
  - Apply 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

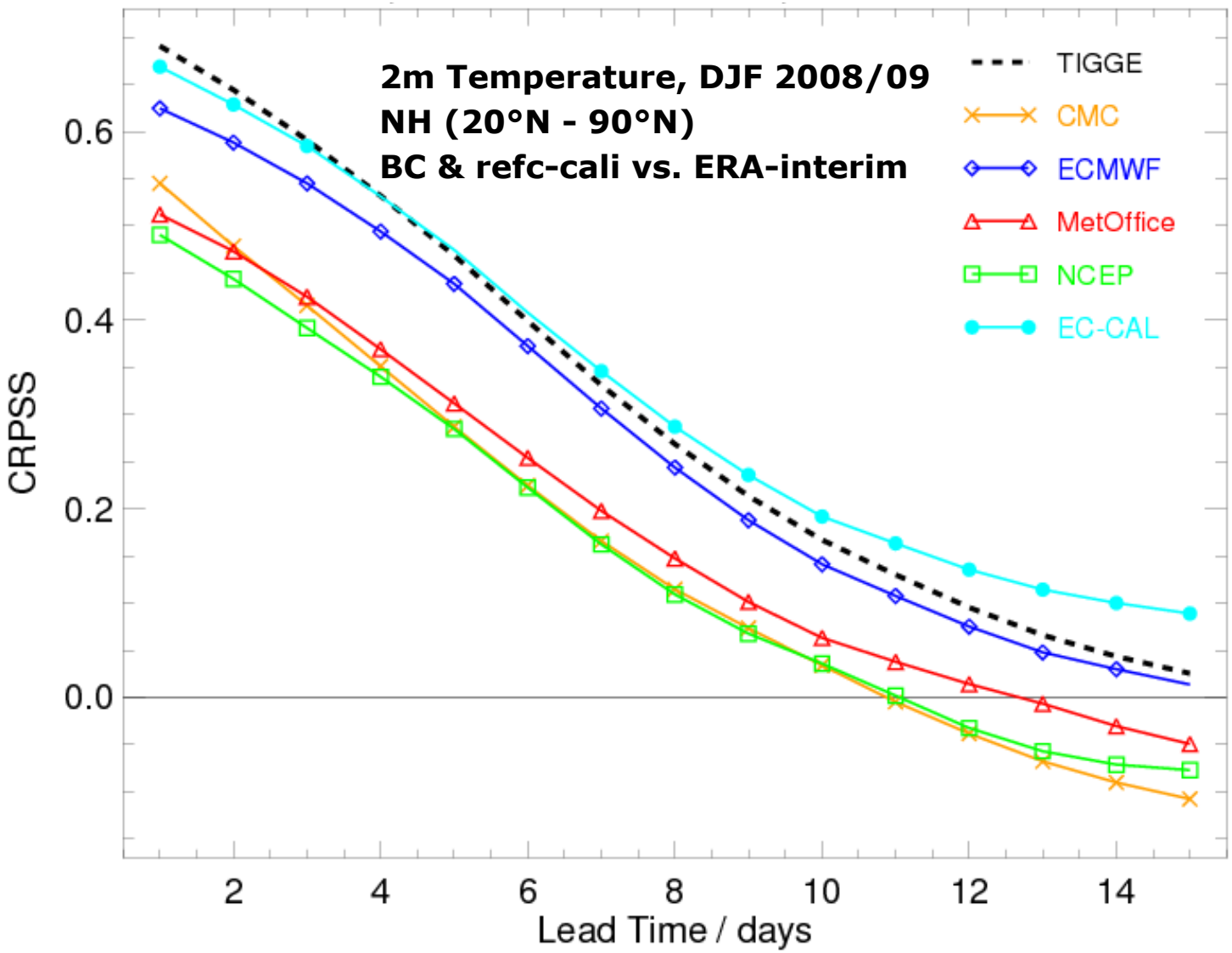
# The reforecast dataset



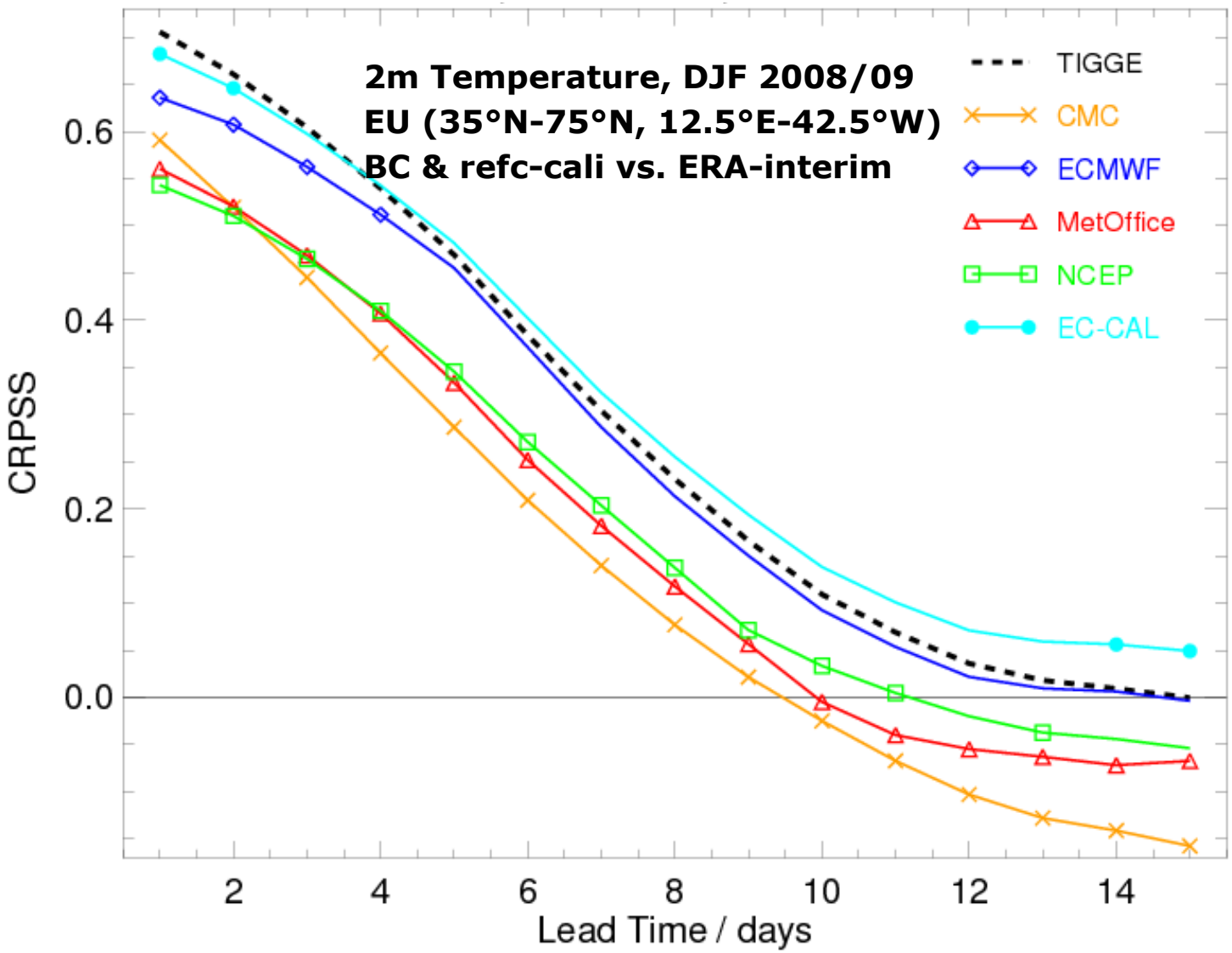
# The reforecast dataset



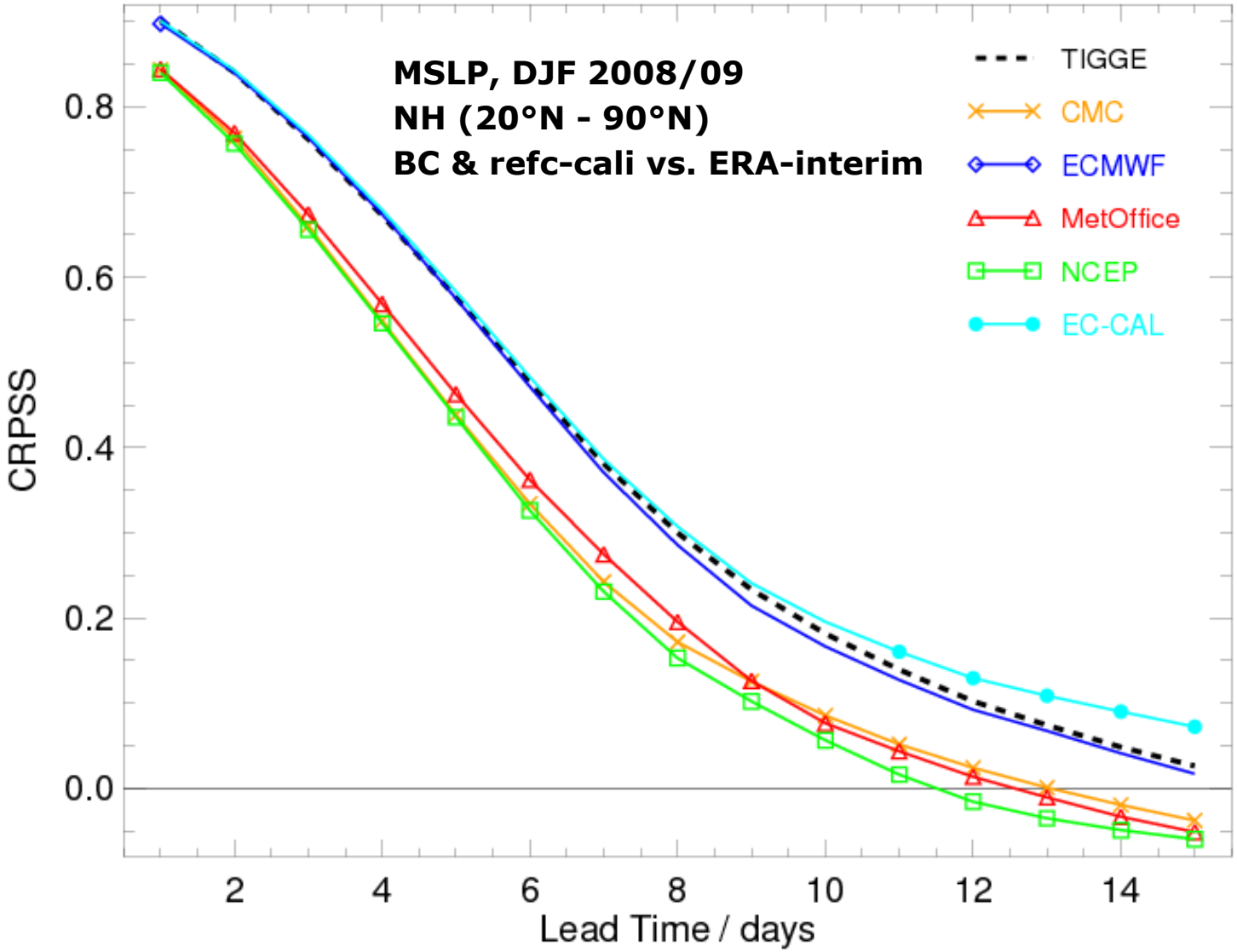
# Comparing 4 TIGGE models, MM, EC-CAL



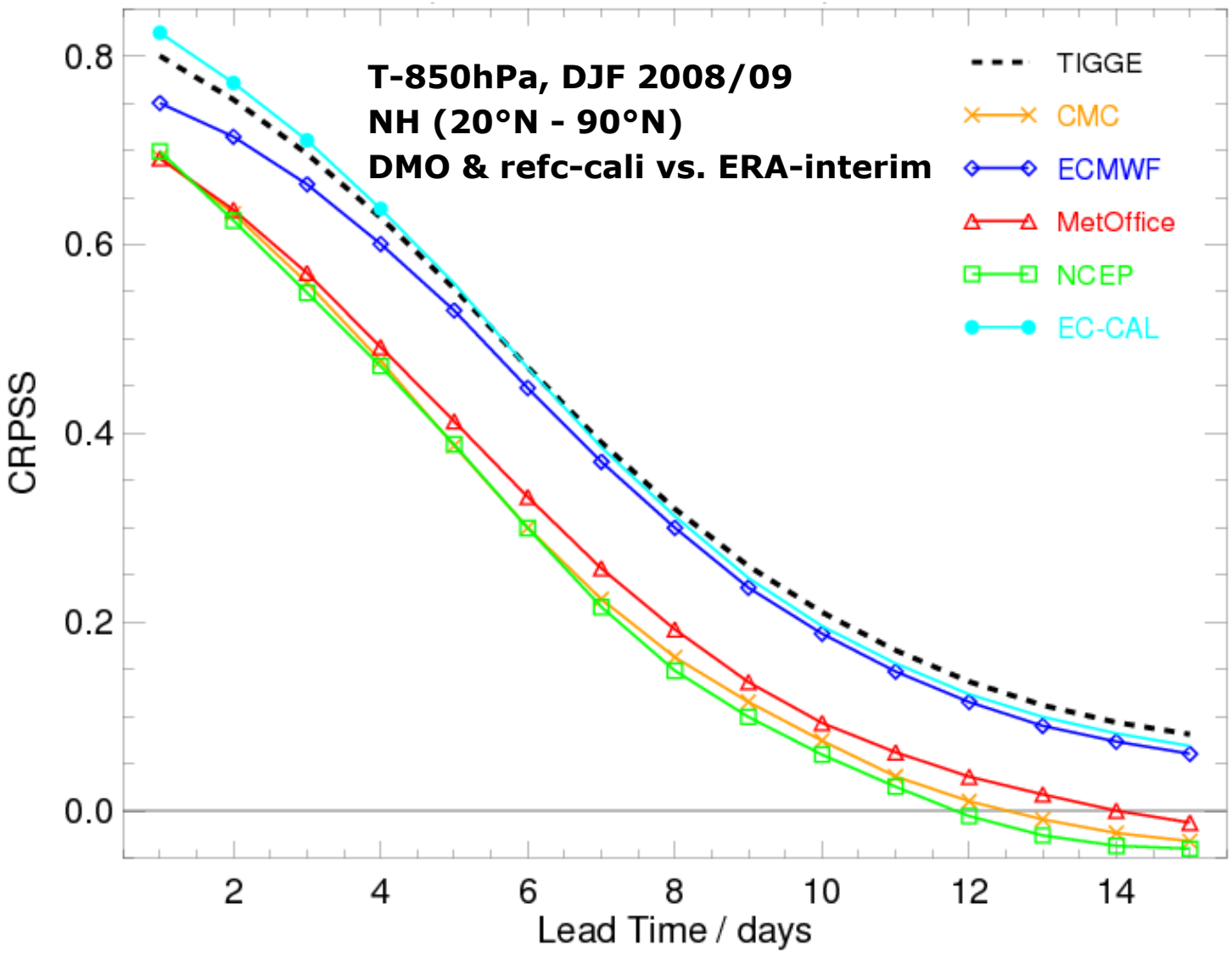
# Comparing 4 TIGGE models, MM, EC-CAL



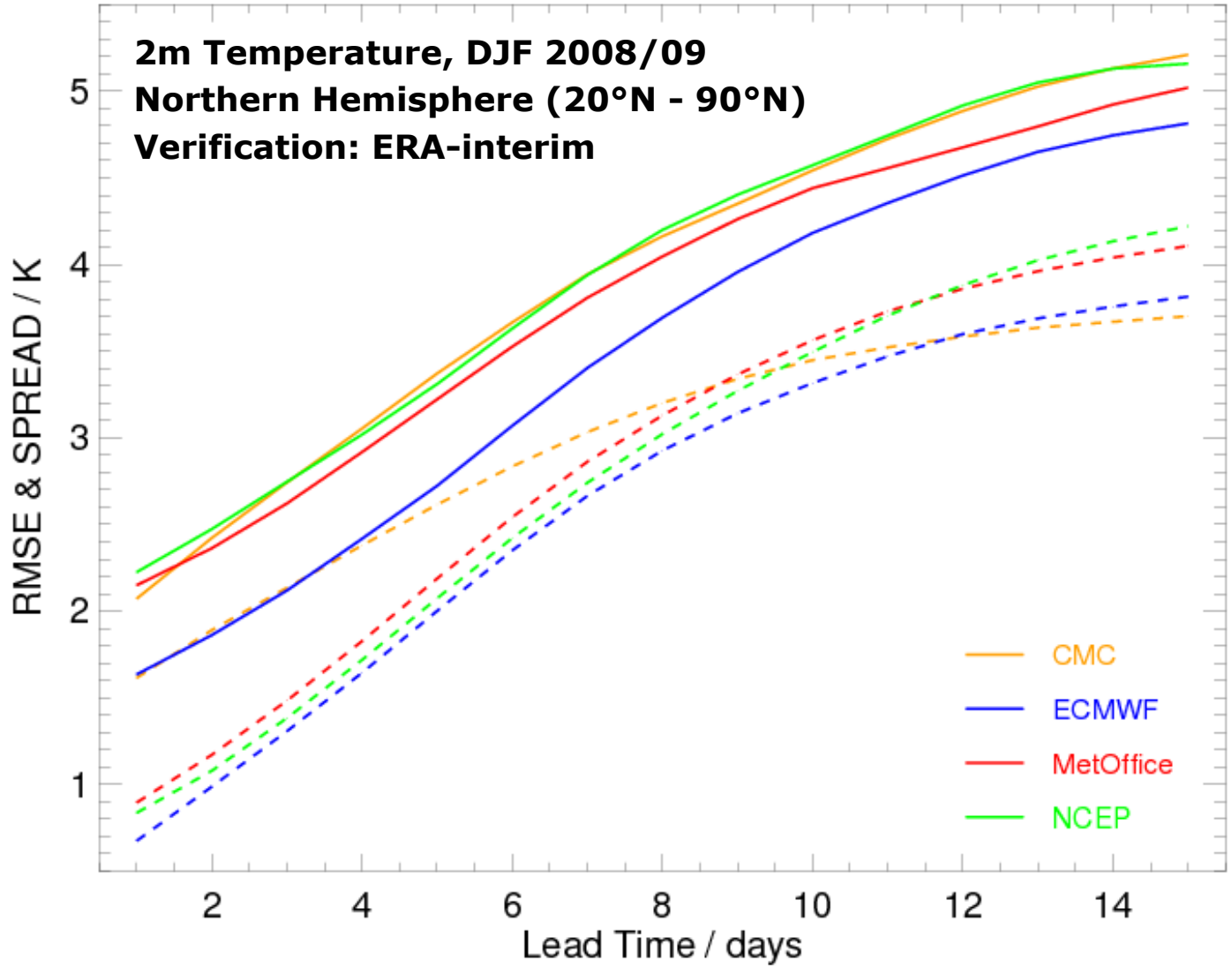
# Comparing 4 TIGGE models, MM, EC-CAL



# Comparing 4 TIGGE models, MM, EC-CAL



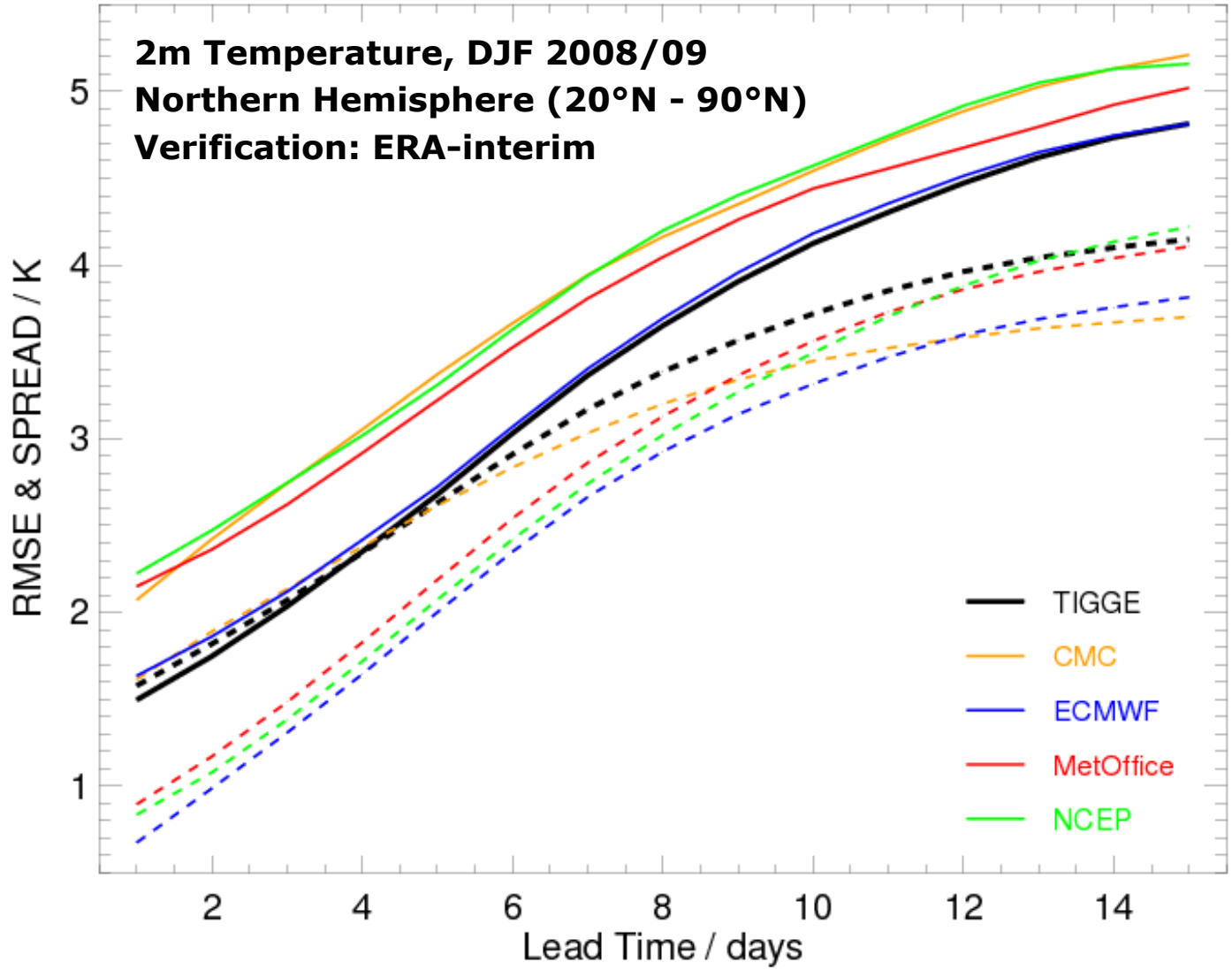
# Mechanism behind improvements




**RMSE (solid)**

**SPREAD (dash)**

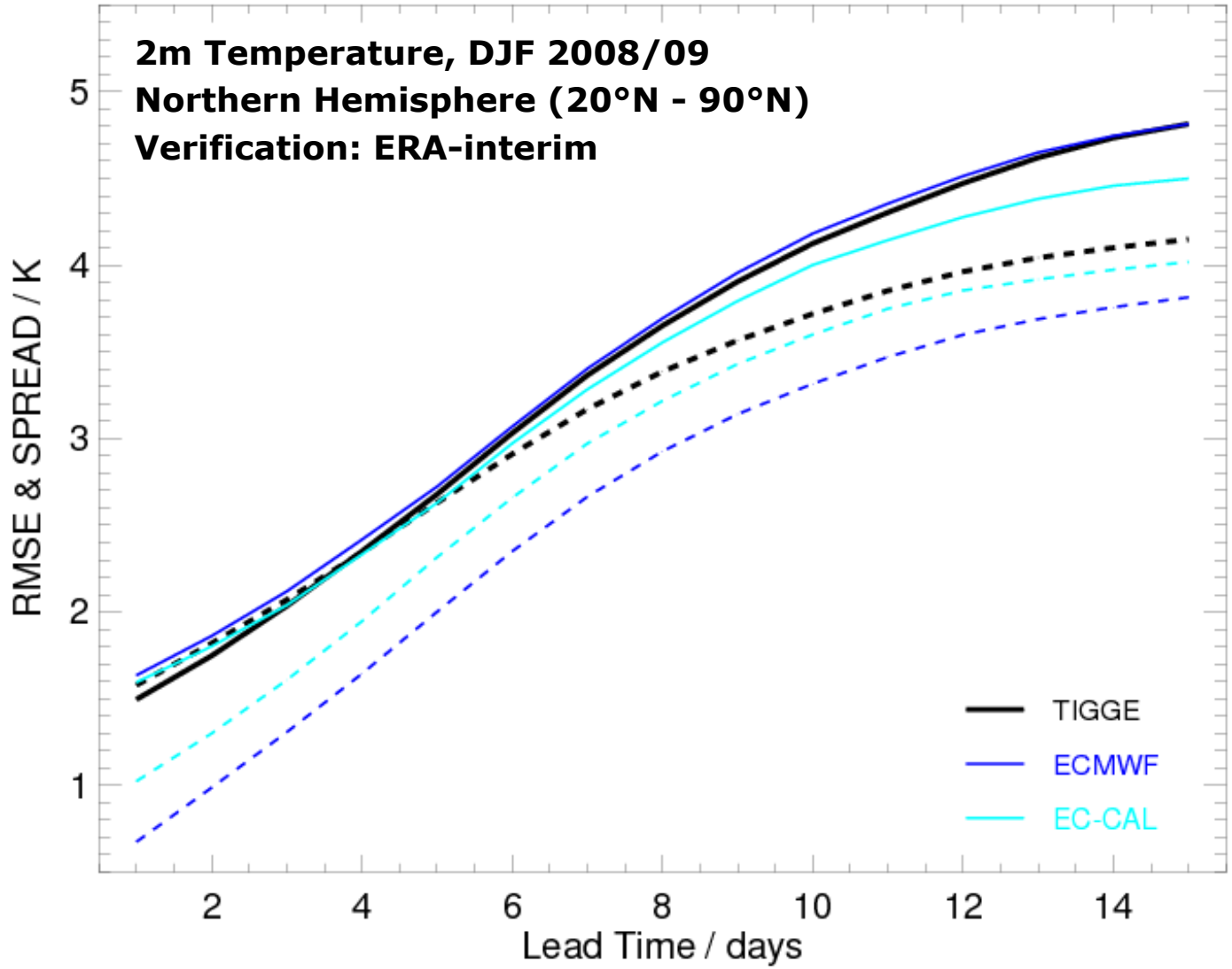
# Mechanism behind improvements




**RMSE (solid)**

**SPREAD (dash)**

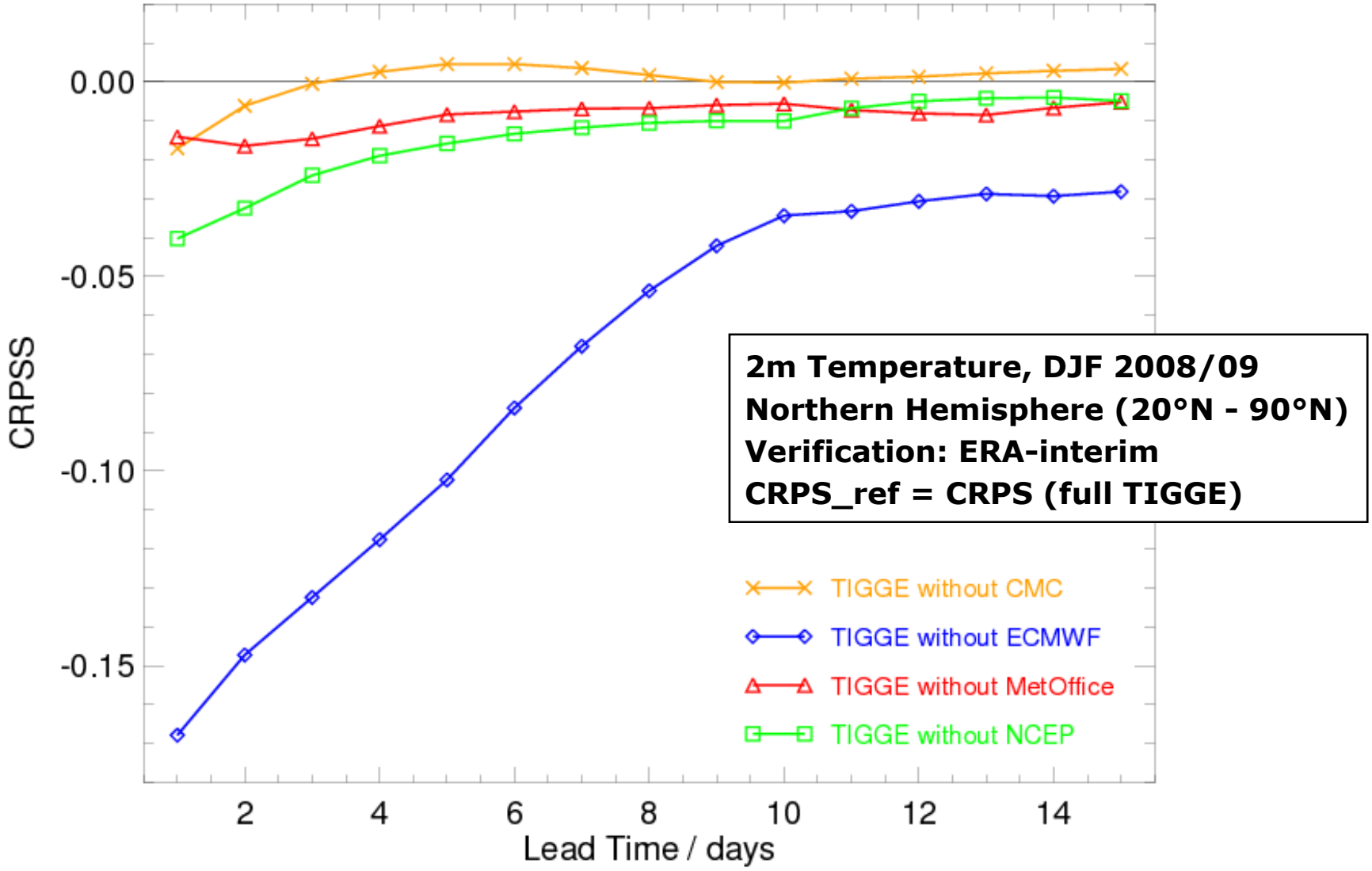
# Mechanism behind improvements



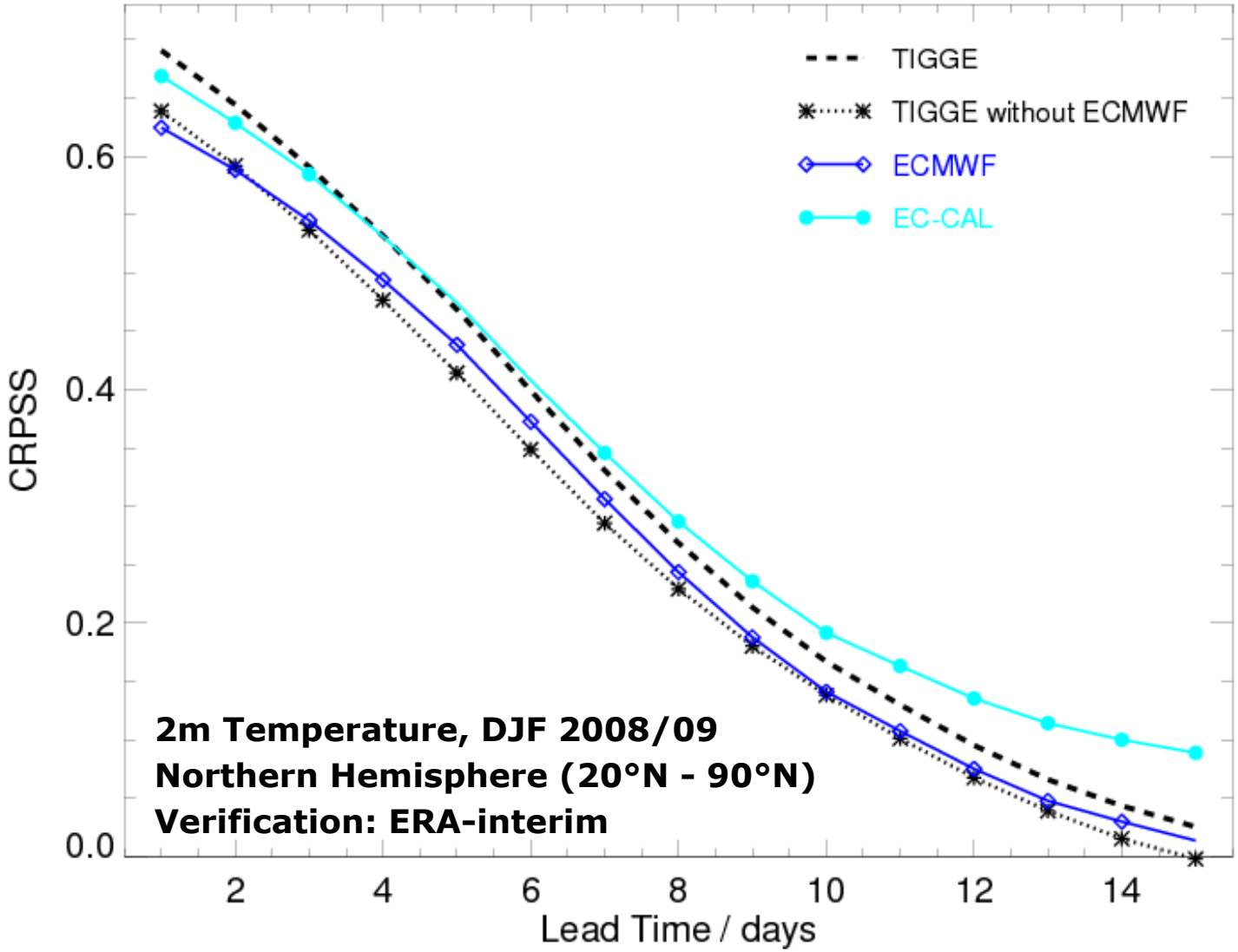

**RMSE (solid)**

**SPREAD (dash)**

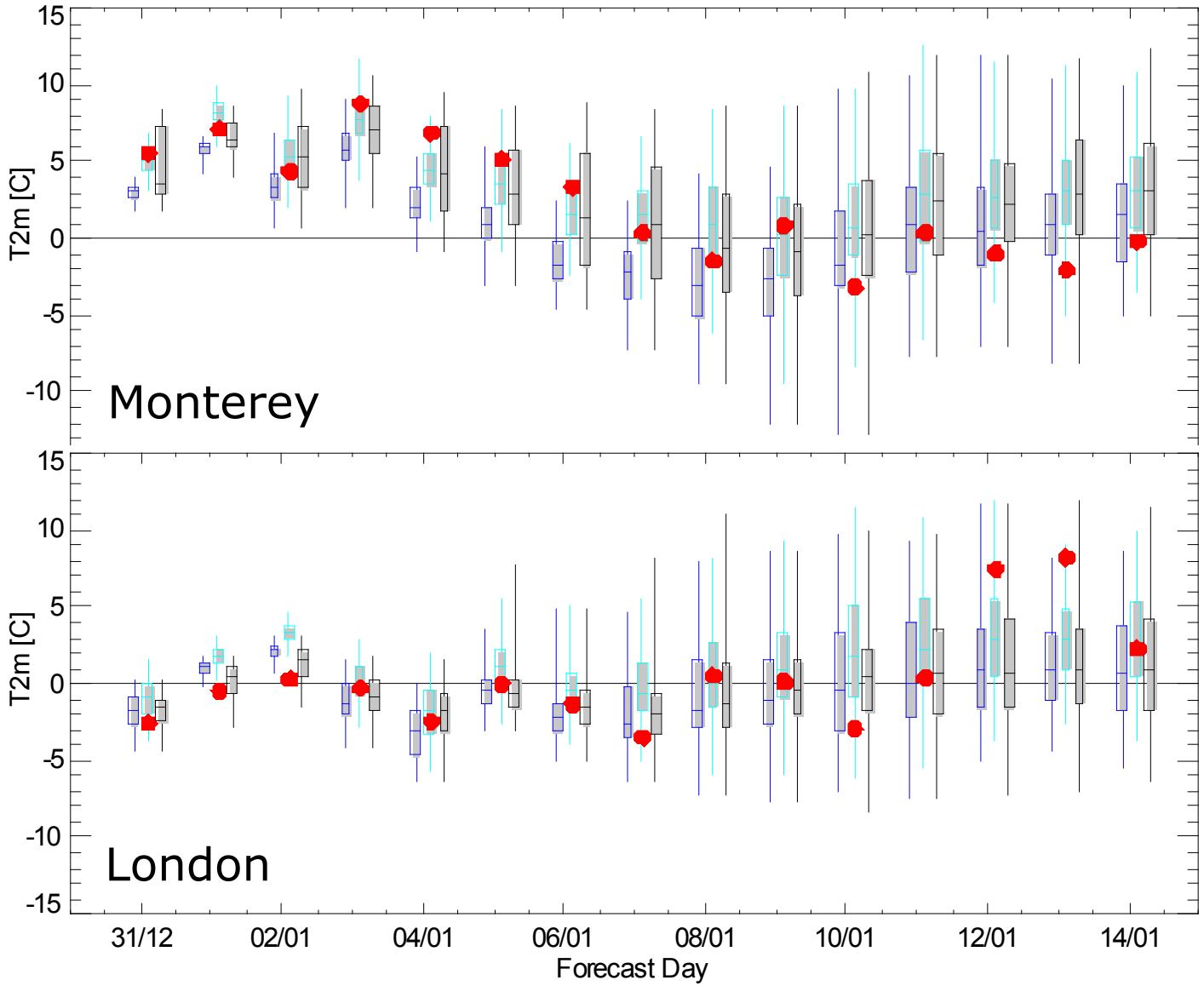
# Reduced TIGGE multi-model



# TIGGE vs. ECMWF vs. EC-CAL



# Impact of calibration & MM in EPSgrams




2m Temperature
FC: 30/12/2008

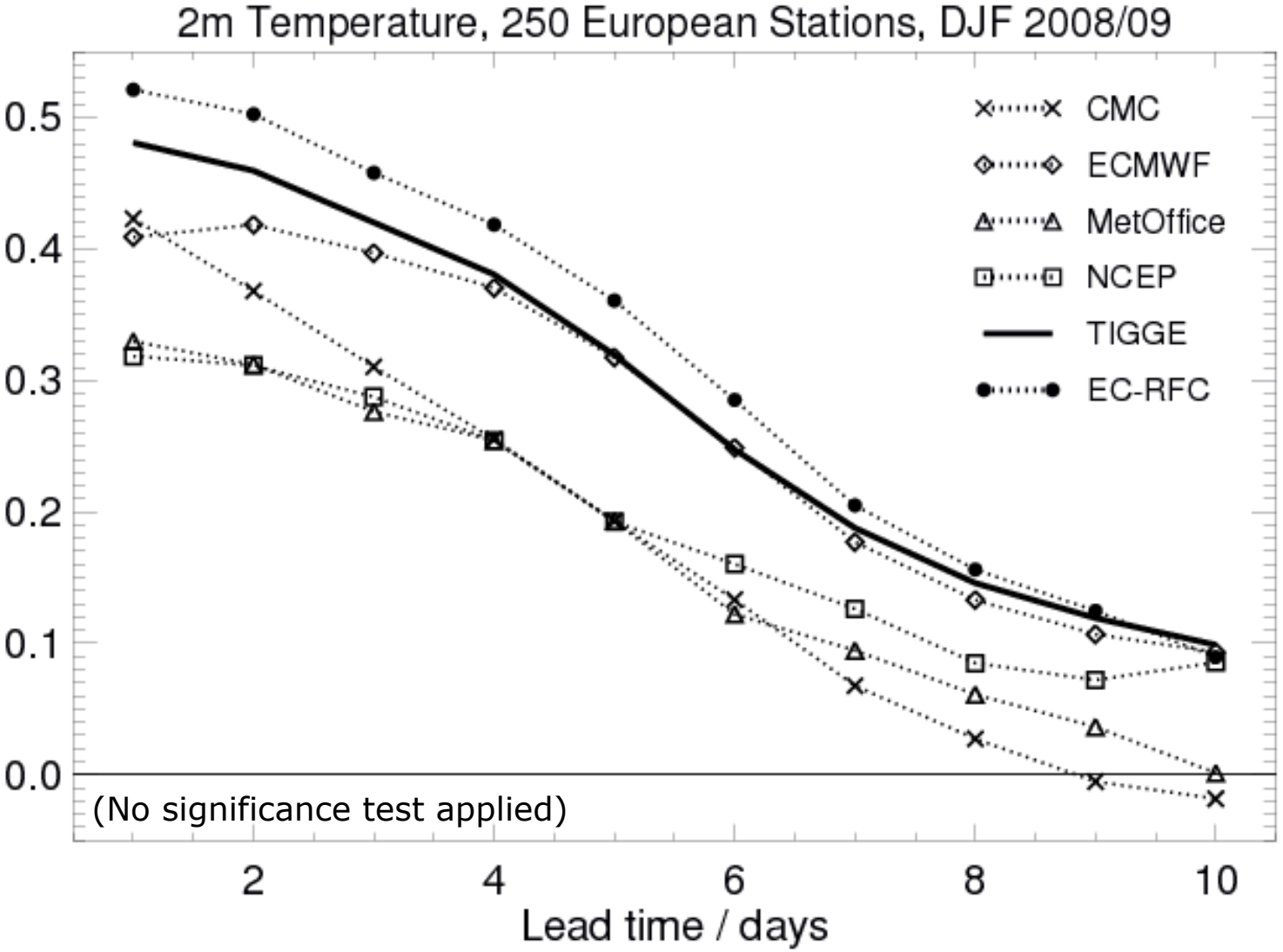
**ECMWF**

**ECMWF-NGR**

**TIGGE**

**Analysis**

# What about station data?



# Relative benefits and costs

| | TIGGE multi-model | NGR Calibration using reforecasts |
|---|---|---|
| Benefits:<br>upper air fields | limited | limited |
| Benefits:<br>surface fields | Improved scores through reduced systematic error and increased spread | Improved scores through reduced systematic error and more appropriate spread |
| Costs:<br>Computational aspects | No extra computer time but data transfer costs | Moderate increase in computing time (~10%), “for free” if reforecasts are produced for other purposes |
| Costs:<br>Logistic aspects | Significantly increased complexity could make system more prone to failures, and timing issues could arise | Slight increase in complexity, e.g. when changing model cycles |

# Summary

- What are the relative benefits (costs) of both approaches?
  - Both multi-model and reforecast calibration approach can improve predictions, in particular for (biased and under-dispersive) near-surface parameters
- What is the mechanism behind the improvements?
  - Both approaches correct similar deficiencies to a similar extent
- Which is the “better” approach?
  - On balance, reforecast calibration seems to be the easier option for a reliable provision of forecasts in an operational environment
  - Both approaches can be useful in achieving the ultimate goal of an optimized, well tuned forecast system