Science Behind the National Blend of Models Temperature Elements

Bruce Veenhuis
NWS/OSTI/MDL
04/15/2015

MDL Collaborators:
Tabitha Huntemann
David Rudack
Daniel Plumb
Dana Strom
Geoff Wagner
Christina Finan
John L. Wagner

National Blend of Models (NBM) Project Goals & Requirements

- Objective
  - Improve quality and consistency of the NWS National Digital Forecast Database (NDFD)
- Project Goals
  - Through an integrated and structured approach
    - Develop a set of foundational gridded guidance products for the NDFD weather elements based on NWS and non-NWS model information
    - Create a methodology for a national blend ("best") from multiple models, beginning with the Day 3-8 time frame and extensible to a full set of deterministic and probabilistic products covering days 1-10
  - Project Requirements:
    - NWS Enterprise Solution
      - Nationally uniform product with spatial and temporal consistency
      - Extensible methodologies (models, elements, lead times…)
    - Meet R2O criteria
      - Implementable and Sustainable
    - No degradation of service

Slide courtesy: Kathryn Gilbert & David Myrick
An Introduction to the National Blend of Global Models Project
VLab Forum – Feb. 18, 2015

Comparison of Blends

MDL Blend	WPC Blend	CR Super Blend
Statistically derived weights based on recent verification	Expert weights determined by verification. Forecasters may adjust weights.	Expert weights determined by verification.

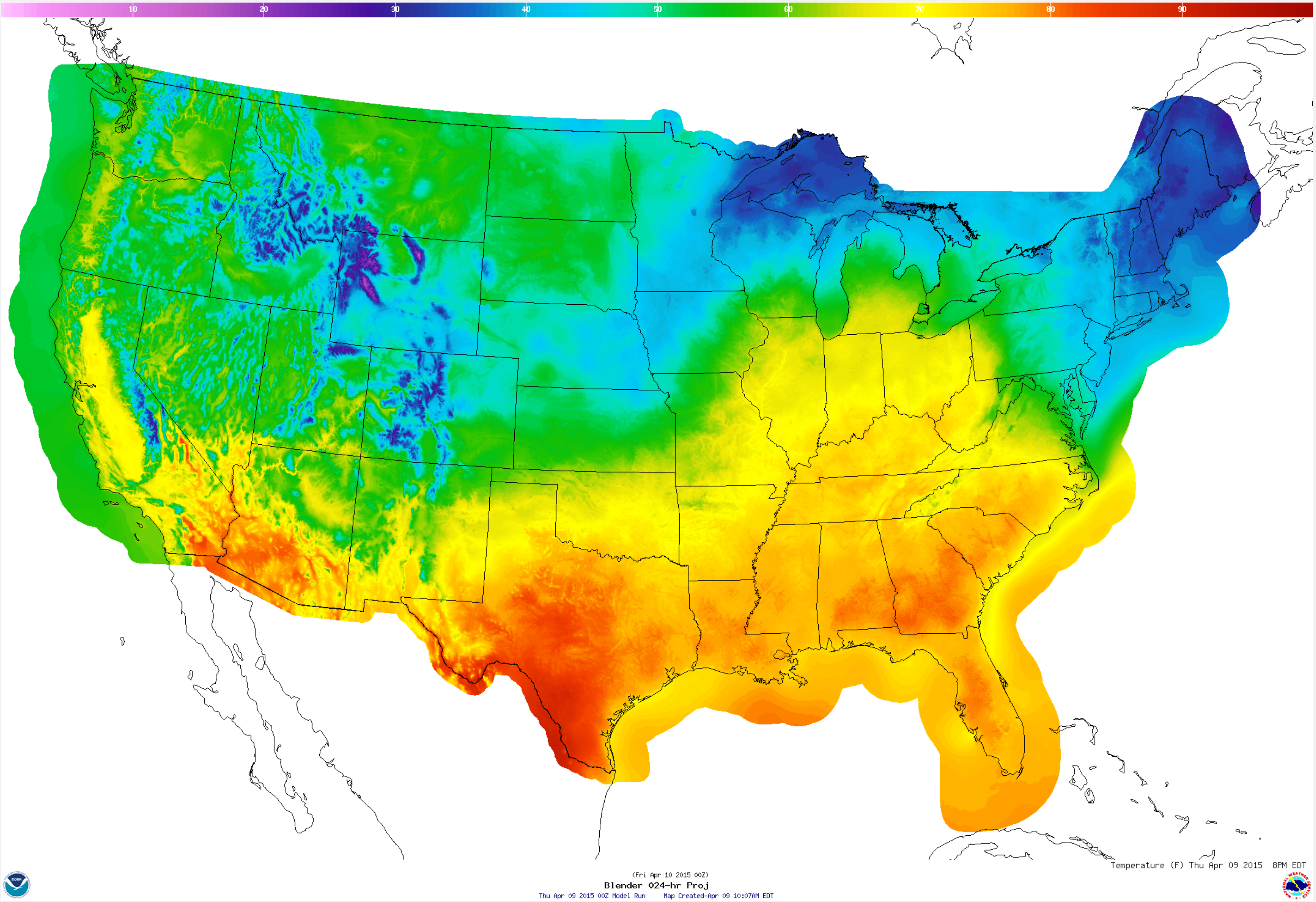
Overview

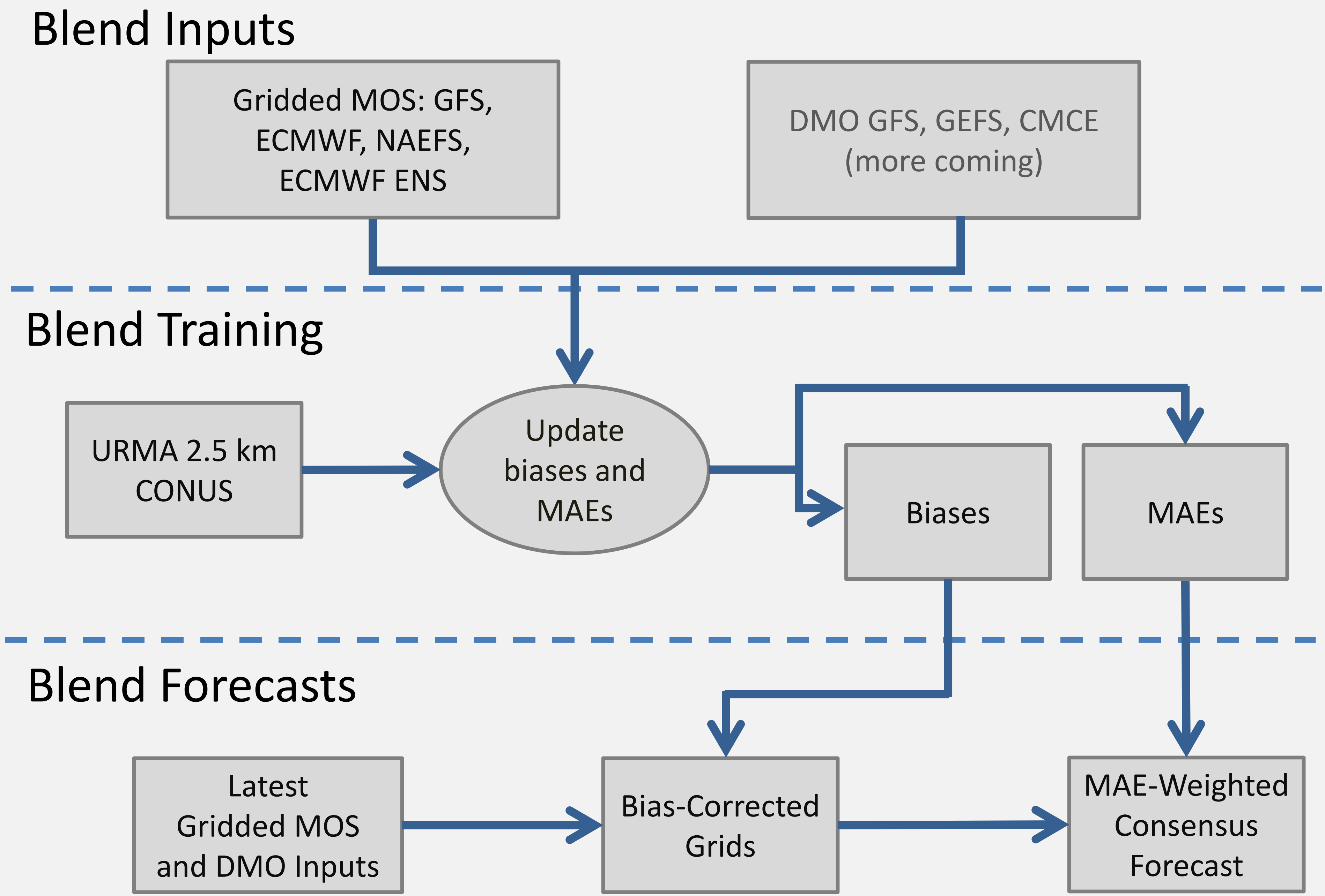
- Explanation of current Blend prototype
- Scientific reasoning for current configuration
- Verification Results

Part 1: Overview

- The National Blend of Models (NBM) combines forecasts from numerical weather prediction models to produce bias-corrected and statistically downscaled guidance on the 2.5 km NDFD grid
- Here we outline the methodology for 2-m temperature, 2-m dewpoint, daytime maximum temperature and nighttime minimum temperature
- Each input is bias-corrected relative to a common high-resolution analysis
- The bias-corrected components are blended using a MAE-based weighting technique

Blend: 09 April 2015, 24-hr 2-m Temperature Forecast




Blend Inputs
Gridded MOS: GFS, ECMWF, NAEFS, ECMWF ENS
DMO GFS, GEFS, CMCE (more coming)
Blend Training
URMA 2.5 km CONUS
Update biases and MAEs
Biases
MAEs
Blend Forecasts
Latest Gridded MOS and DMO Inputs
Bias-Corrected Grids
MAE-Weighted Consensus Forecast

Part 1: Bias-Correction

- Track the bias of each model using an Exponentially Weighted Moving Average (EWMA; Roberts 1959 also called "decaying average" Cui et al. 2012)

$$B_t = (1-\alpha)B_{t-1} + \alpha(FCST_{t-1} - OBS_{t-1})$$

B = Bias   α = "Decaying Weight"   OBS = Observation   FCST = Forecast

- Bias-correction is performed separately for each grid point, projection, and element
- Used to create bias-corrected forecast grids

$$BCFCST_t = FCST_t - B_t$$

Part 1: MAE-based Weighting

- Track the MAE of each bias-corrected component using an EWMA

$$MAE_t = (1-\alpha)MAE_{t-1} + \alpha|BCFCST_{t-1} - OBS_{t-1}|$$

MAE = Mean Absolute Error   BCFCST = Bias-corrected Forecast   α = "Decaying Weight"
OBS = Observation

- Separate MAE estimates for each grid point, projection, and element

Part 1: MAE-based Weighting (cont.)

- MAE-based weighting scheme (Woodcock and Engel, 2005)

$$w_m = a_m^{-1}\left(\sum_{k=1}^{K} a_k^{-1}\right)^{-1}$$

- Where w_m is the weight for member m, a_m is the most recent MAE_t for member m, and K is the total number of models being blended

Part 1: MAE-based Weighting (cont.)

$$w_m = a_m^{-1} \left(\sum_{k=1}^{K} a_k^{-1} \right)^{-1}$$

Part 1: MAE-based Weighting (cont.)

Example with 3 models:
MAE_1=2
MAE_2=3
MAE_3=4

$$w_m = a_m^{-1}\left(\sum_{k=1}^{K} a_k^{-1}\right)^{-1}$$

Weight for model 1…

$$w_1 = \frac{\left(\frac{1}{MAE_1}\right)}{\frac{1}{MAE_1}+\frac{1}{MAE_2}+\frac{1}{MAE_3}}$$

Part 1: MAE-based Weighting (cont.)

$$w_m = a_m^{-1}\left(\sum_{k=1}^{K} a_k^{-1}\right)^{-1}$$

Example with 3 models:
MAE_1=2
MAE_2=3
MAE_3=4

Weight for model 1…

$$w_1 = \frac{\left(\frac{1}{MAE_1}\right)}{\frac{1}{MAE_1}+\frac{1}{MAE_2}+\frac{1}{MAE_3}}$$

$$w_1 = \frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{3}+\frac{1}{4}} = 0.46$$

Part 1: MAE-based Weighting (cont.)

Example with 3 models:
MAE_1=2
MAE_2=3
MAE_3=4

$$w_m = a_m^{-1}\left(\sum_{k=1}^{K} a_k^{-1}\right)^{-1}$$

Weight for model 1…

$$w_1 = \frac{\left(\frac{1}{MAE_1}\right)}{\frac{1}{MAE_1}+\frac{1}{MAE_2}+\frac{1}{MAE_3}}$$

$$w_1 = \frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{3}+\frac{1}{4}} = 0.46$$

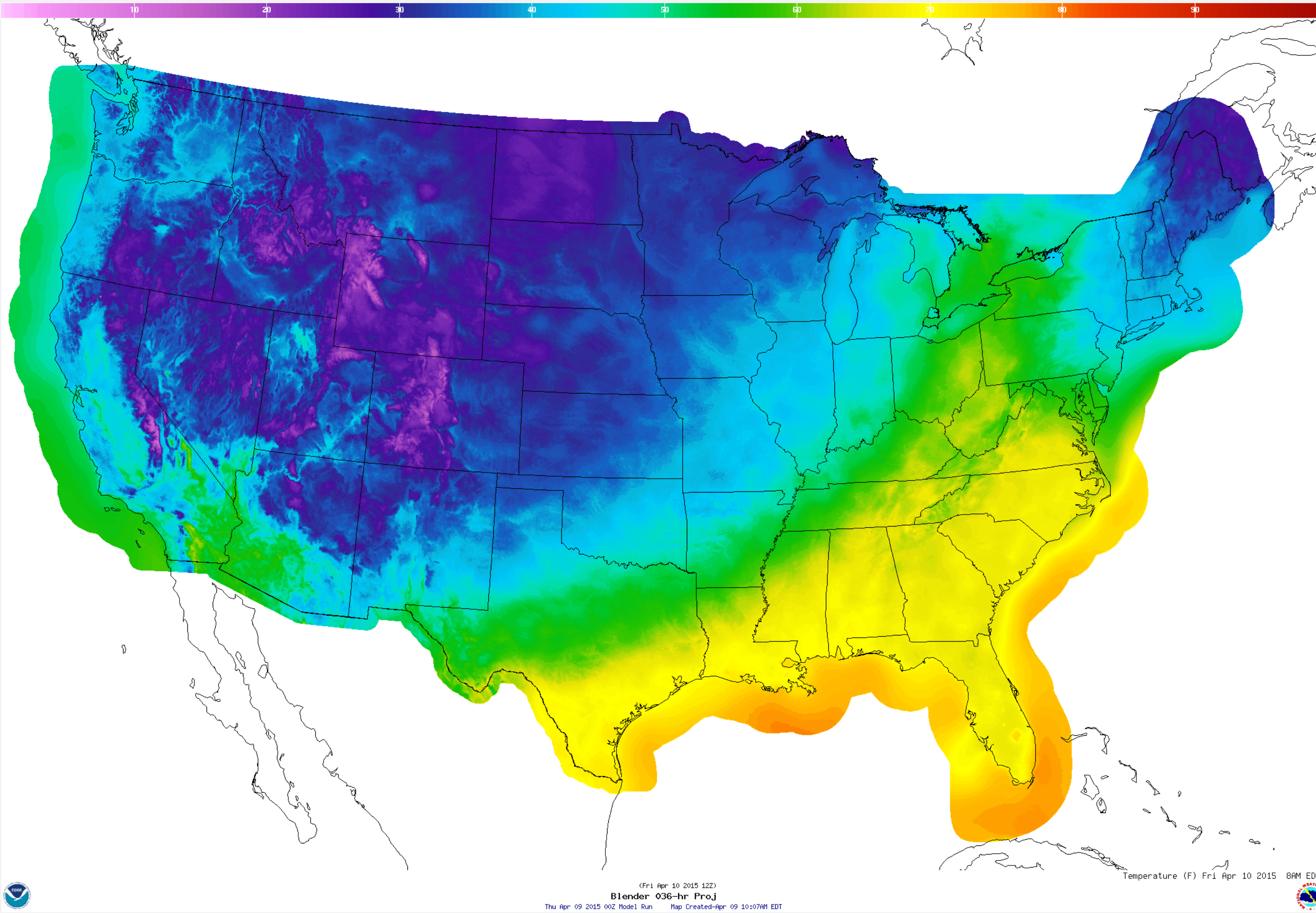
Repeat for remaining two models…

Part 1: MAE-based Weighting (cont.)

$$FCST_{Blend} = \sum_{m=1}^{M} w_m BCFCST_m$$

- Where w_m is the weight for member *m*, $BCFCST_m$ is the bias-corrected forecast for member *m*, and *M* is the total number of models being blended

Blend: 09 April 2015, 36-hr 2-m Temperature Forecast



Part 1: MAE-based Weighting (cont.)

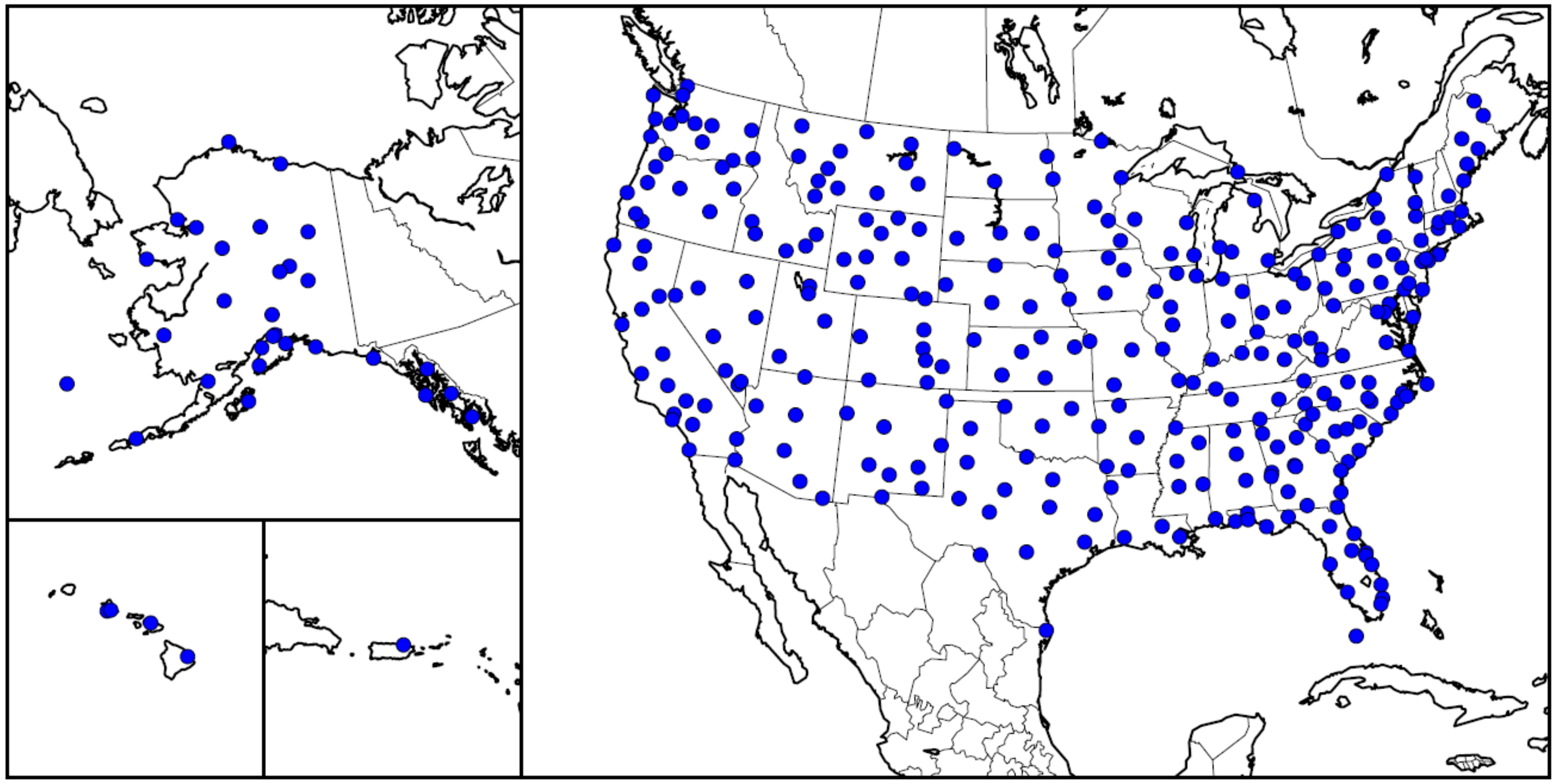
- Pros: Simple computations, straightforward to implement, reasonable results, easy to handle missing model forecasts
- Cons: Does not adjust for error correlation among models

Part 2: Reasoning Behind Blend

Part 2: Reasoning Behind Blend

- Provide justification for Blend configuration backed by verification results
- Before implementing the prototype we tested various weighting techniques using a station-based dataset
  - Direct model output (DMO) 2-m temperature from ECMWF Deterministic, GFS, GEFS, CMCE, and NAM (projections $\leq$ 84-hrs)
  - DMO interpolated to stations and bias-corrected relative to the station-based observations using an EWMA
  - Results for 1 Oct. 2008 – 30 Sept. 2012

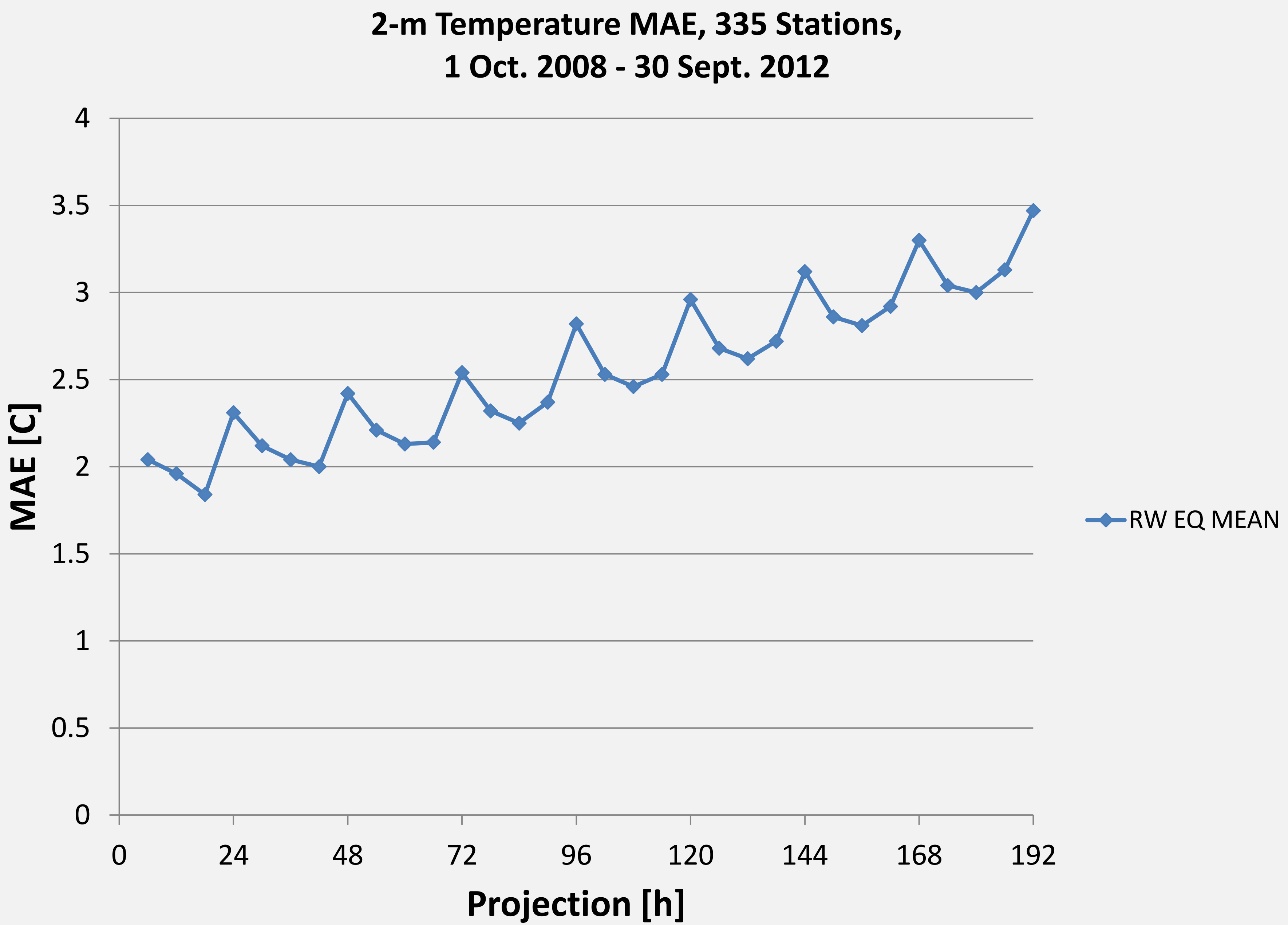
Part 2: 335 Stations

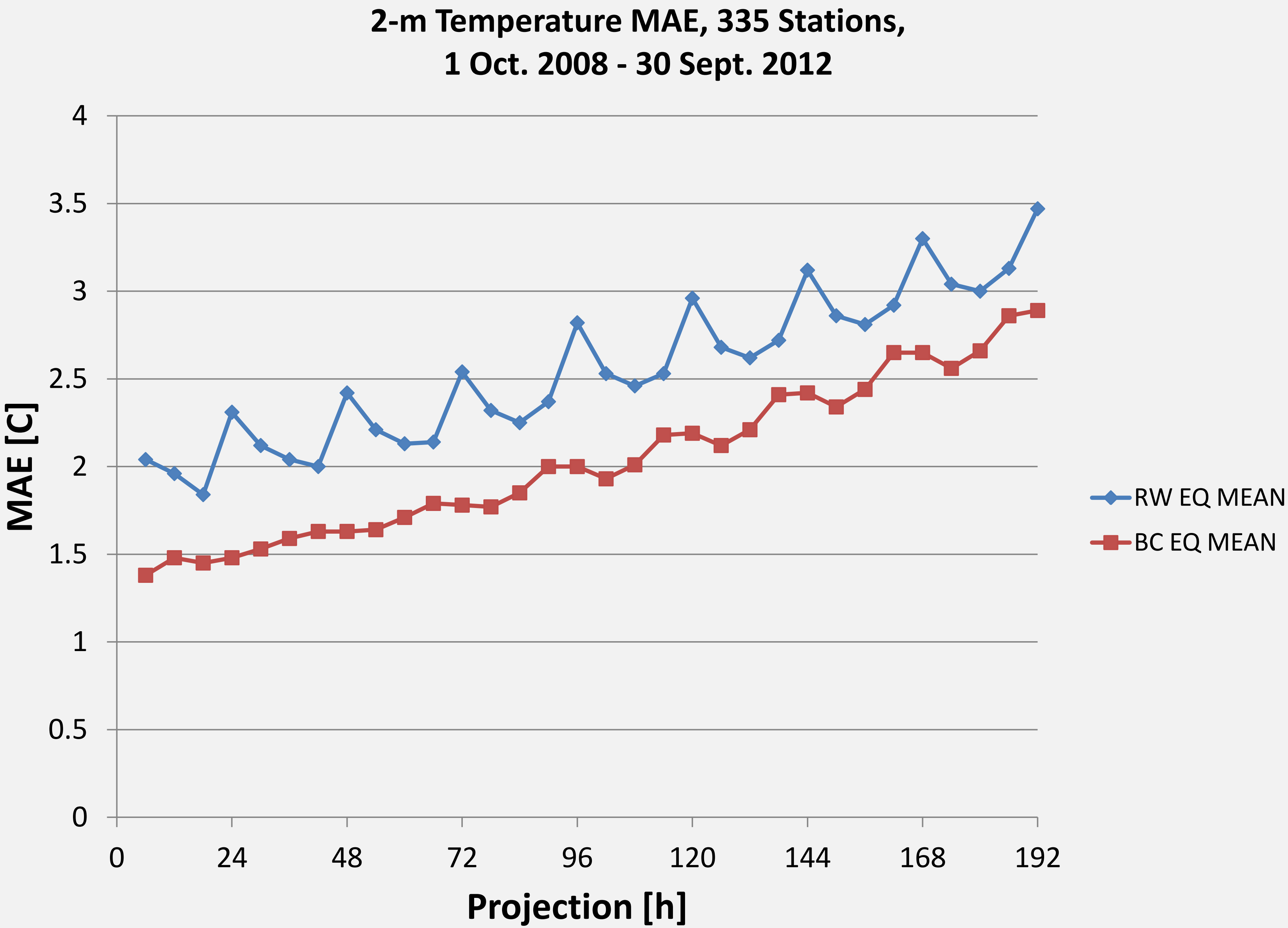


Part 2: Candidate Techniques

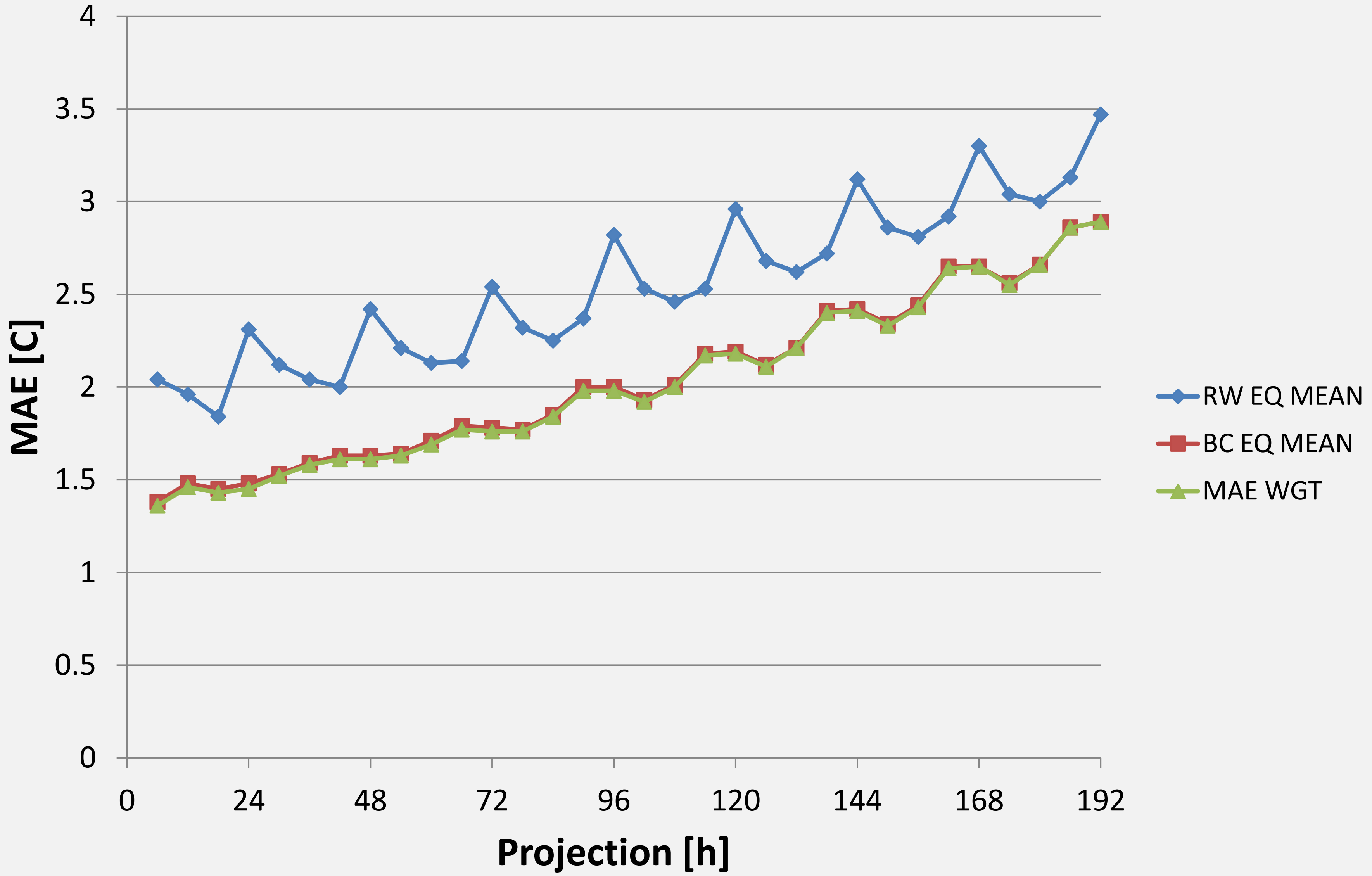
- Equal Weights
- MAE and RMSE-based weights
  - Woodcock and Engel (2005)
- Ridge Regression
  - Peña and van den Dool (2008)
- Bayesian Model Averaging (BMA)
  - Raftery et al. (2005), Veenhuis (2014)

Increasing Complexity

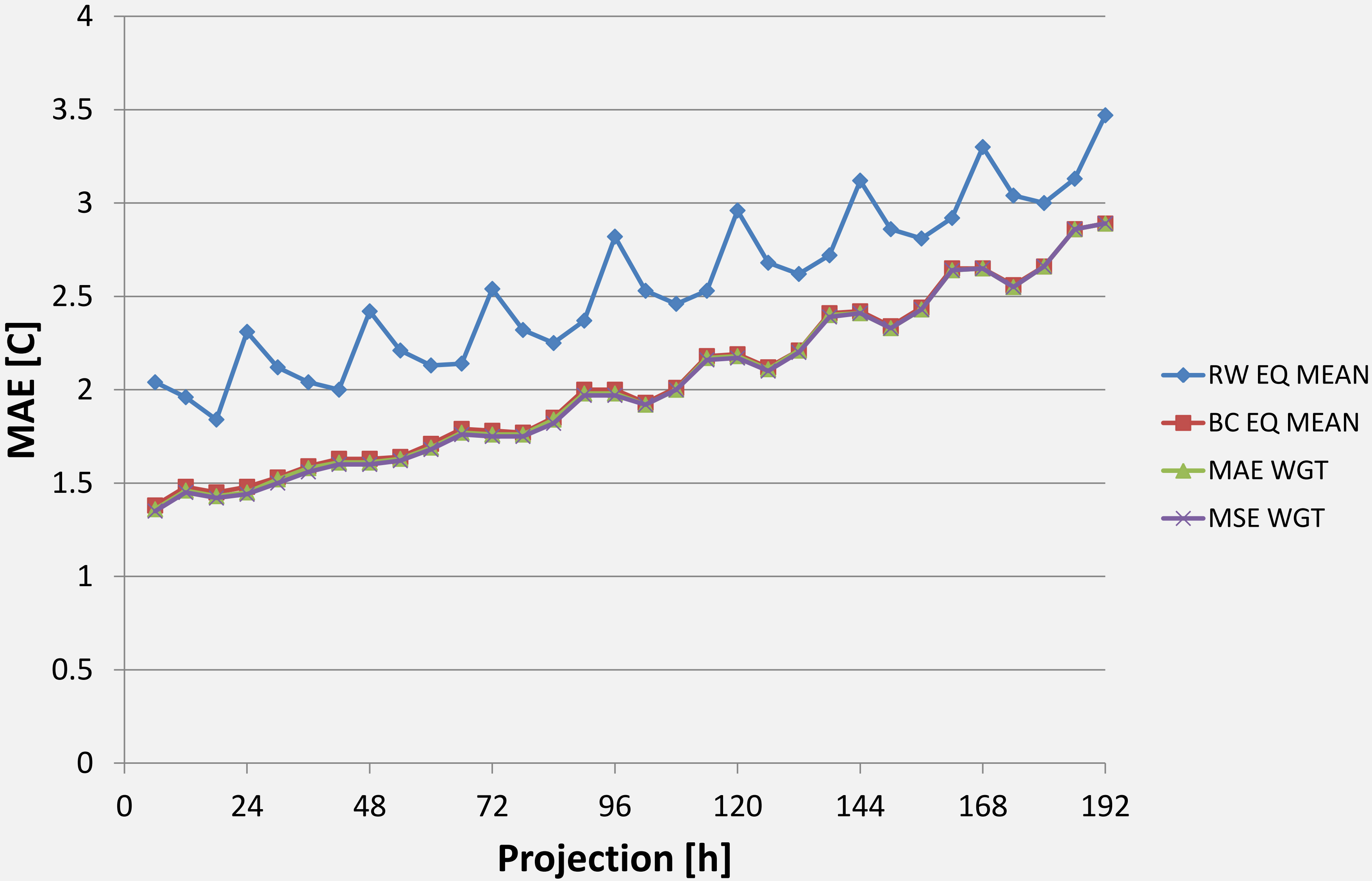

2-m Temperature MAE, 335 Stations,
1 Oct. 2008 - 30 Sept. 2012
MAE [C]
4
3.5
3
2.5
2
1.5
1
0.5
0
0
24
48
72
96
120
144
168
192
Projection [h]
RW EQ MEAN

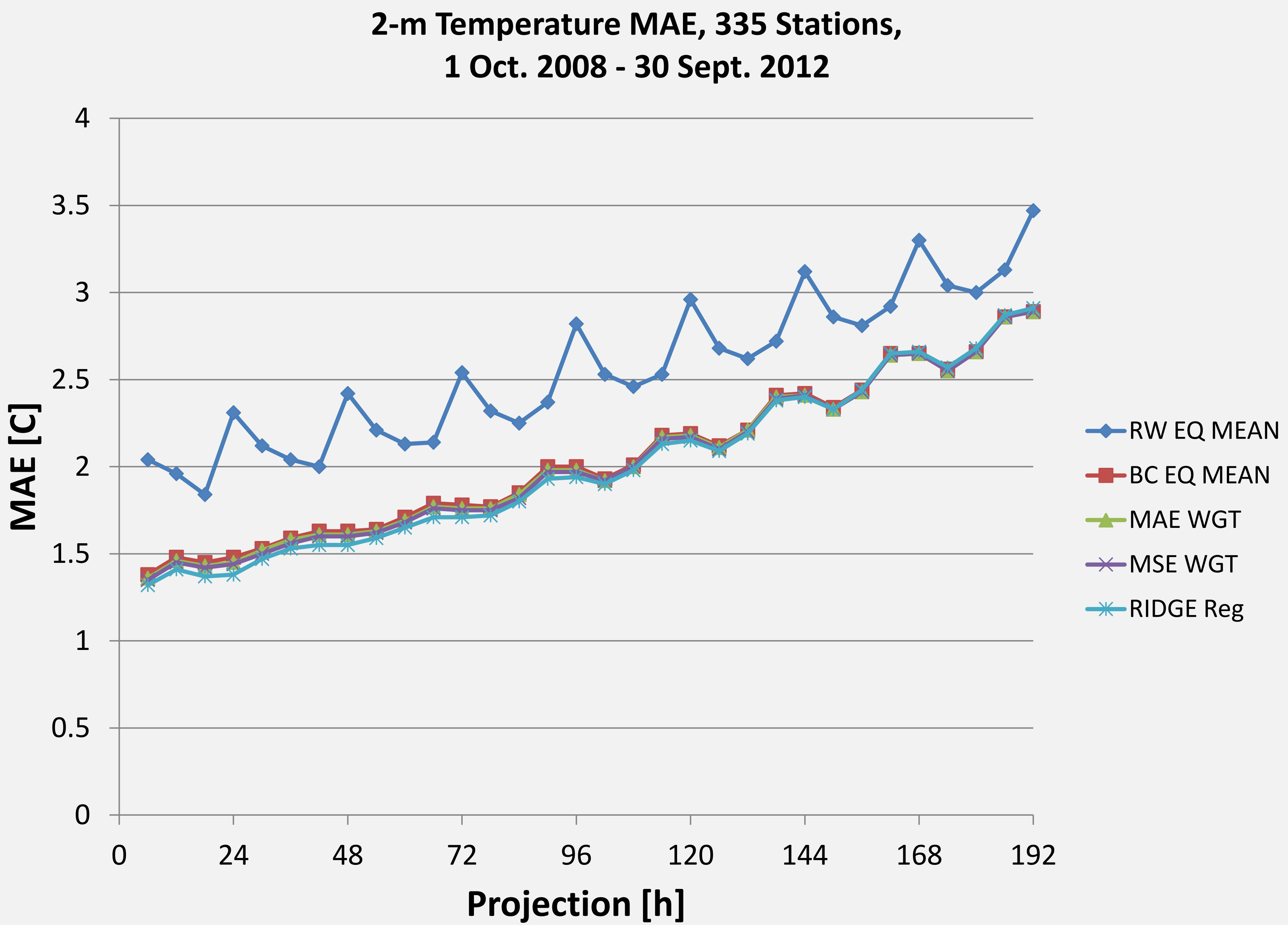

2-m Temperature MAE, 335 Stations,
1 Oct. 2008 - 30 Sept. 2012
MAE [C]
4
3.5
3
2.5
2
1.5
1
0.5
0
0
24
48
72
96
120
144
168
192
Projection [h]
RW EQ MEAN
BC EQ MEAN

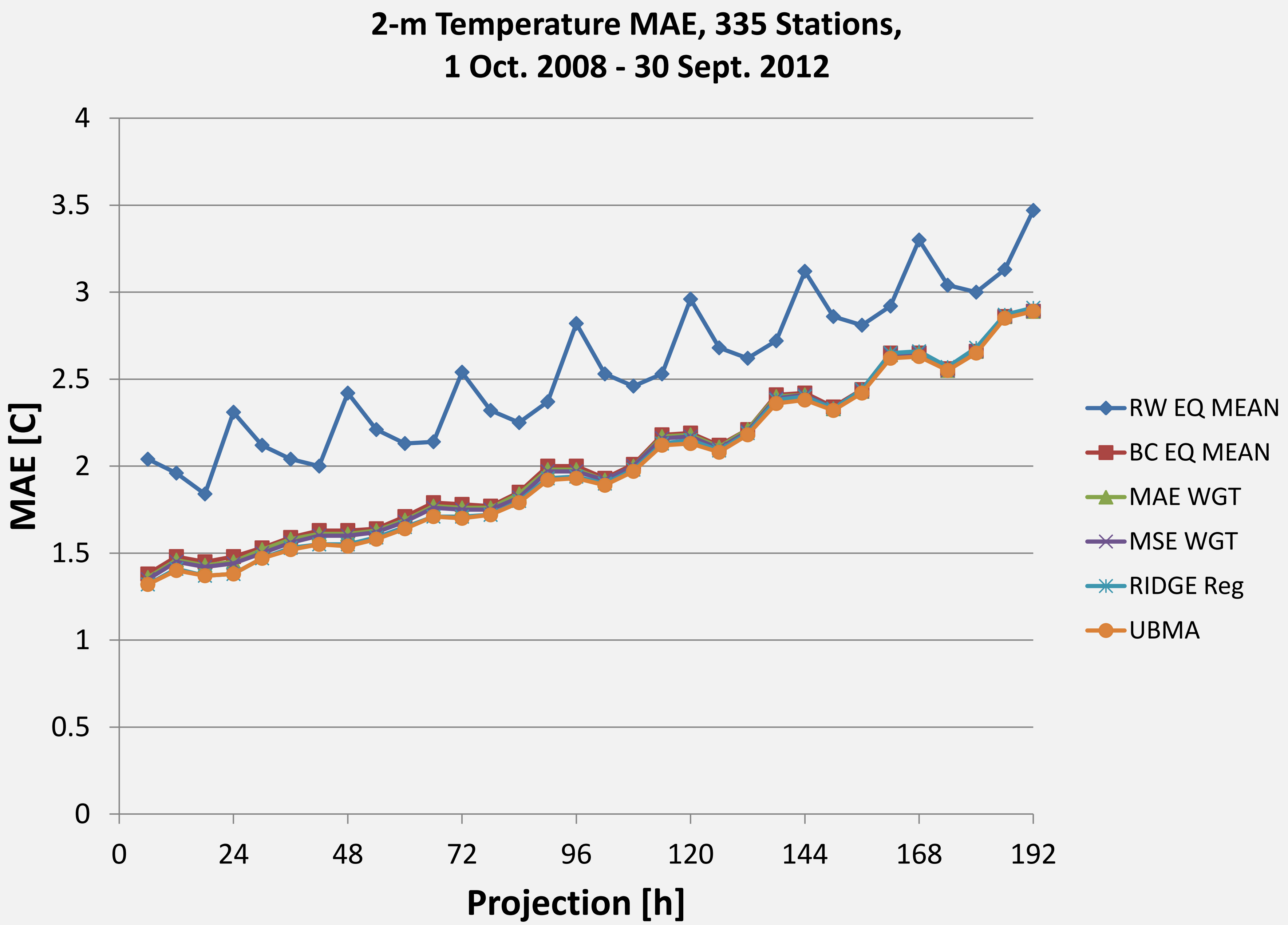
2-m Temperature MAE, 335 Stations, 1 Oct. 2008 - 30 Sept. 2012



2-m Temperature MAE, 335 Stations, 1 Oct. 2008 - 30 Sept. 2012




2-m Temperature MAE, 335 Stations,
1 Oct. 2008 - 30 Sept. 2012
MAE [C]
0
0.5
1
1.5
2
2.5
3
3.5
4
0
24
48
72
96
120
144
168
192
Projection [h]
RW EQ MEAN
BC EQ MEAN
MAE WGT
MSE WGT
RIDGE Reg


2-m Temperature MAE, 335 Stations,
1 Oct. 2008 - 30 Sept. 2012
MAE [C]
Projection [h]
0
0.5
1
1.5
2
2.5
3
3.5
4
0
24
48
72
96
120
144
168
192
RW EQ MEAN
BC EQ MEAN
MAE WGT
MSE WGT
RIDGE Reg
UBMA

Part 2: Summary

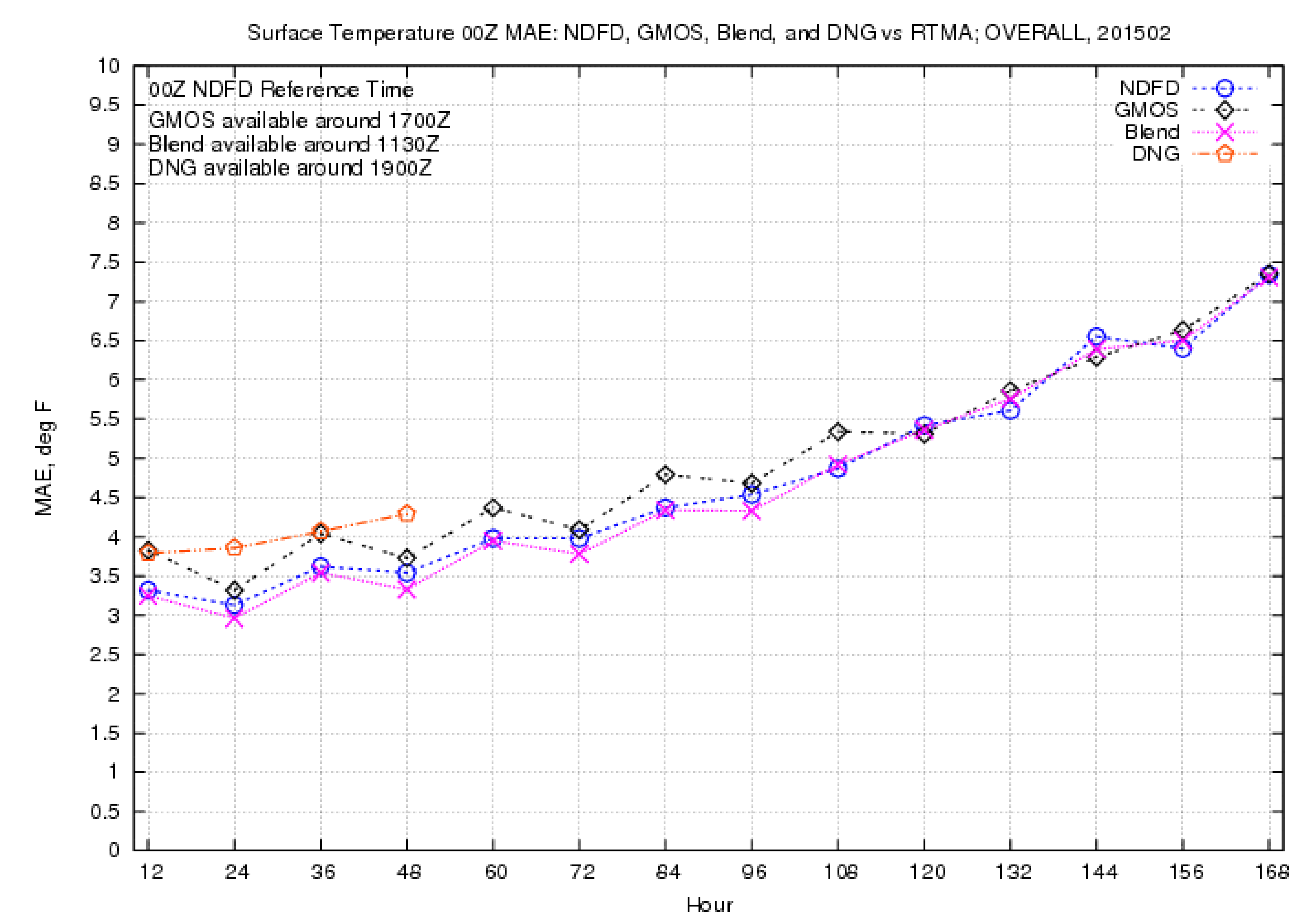
- MAE-weighted Blend performed well for 2-m temperature
- Increasing complexity yielded diminishing returns
- MAE-based weighting scheme is robust and easiest to implement operationally
- Can set a competitive benchmark for future improvements

Part 3: Blend Verification

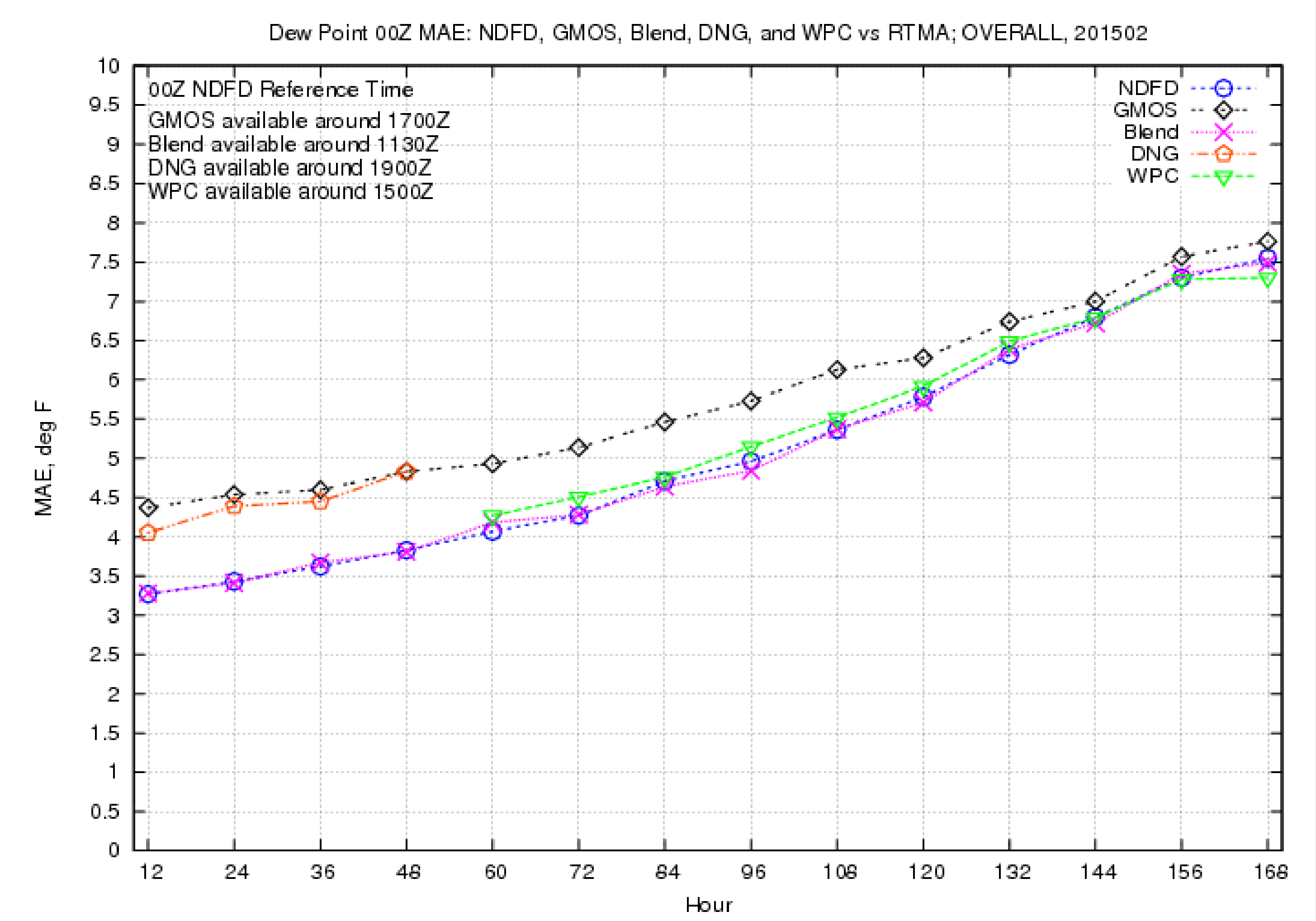
Part 3: Blend Verification

- Verification results for the Blend prototype
- Gridded Verification Relative to RTMA
  - Courtesy of the NBM Verification Team (Tabitha Huntemann)
- Point Verification (Blend vs. ECMWF GMOS)
  - Courtesy of David Rudack
- Examples of MAE-based weights at specific points

Part 3: February 2015
2-m Temperature

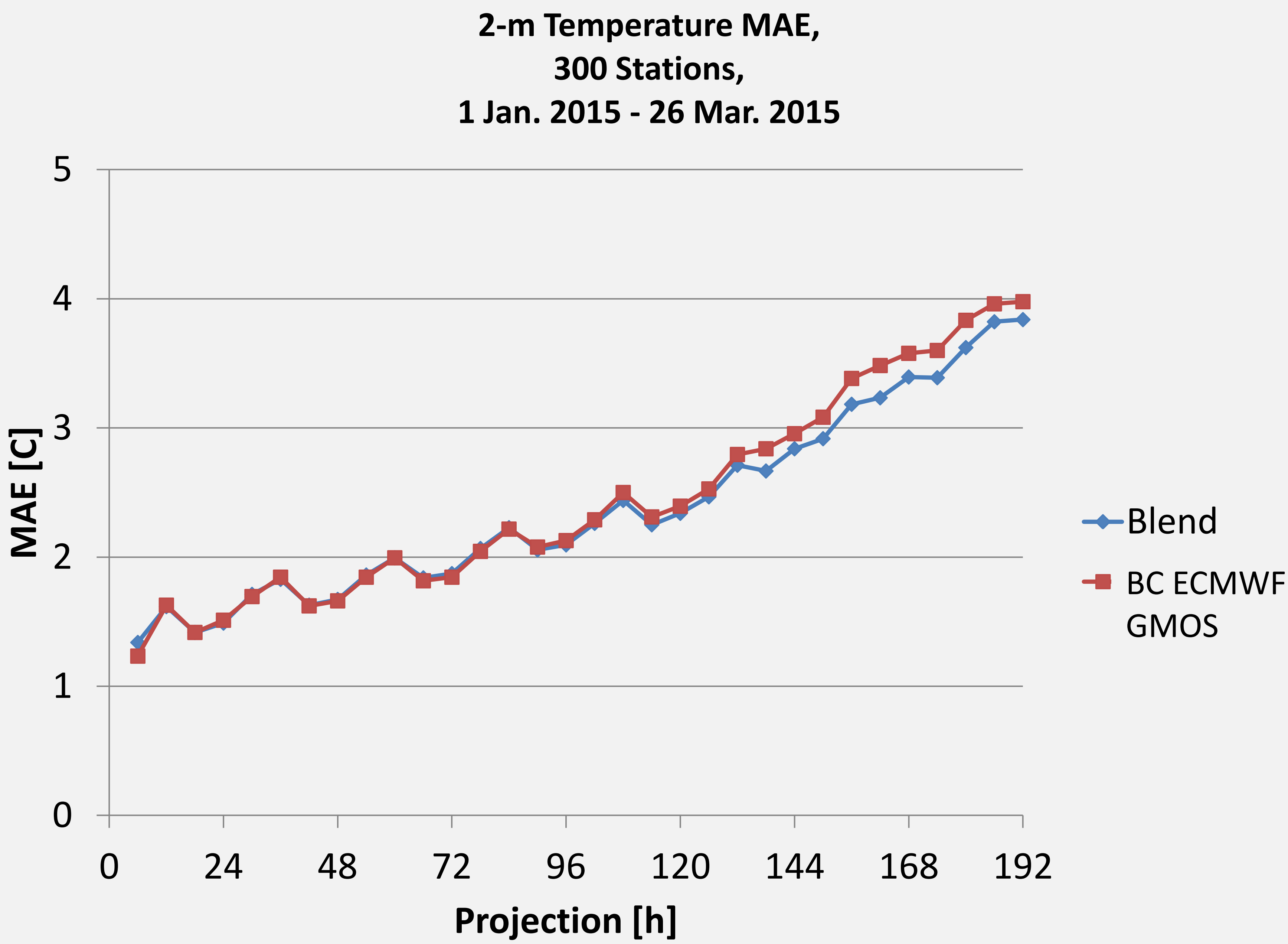


Part 3: February 2015
2-m Dewpoint

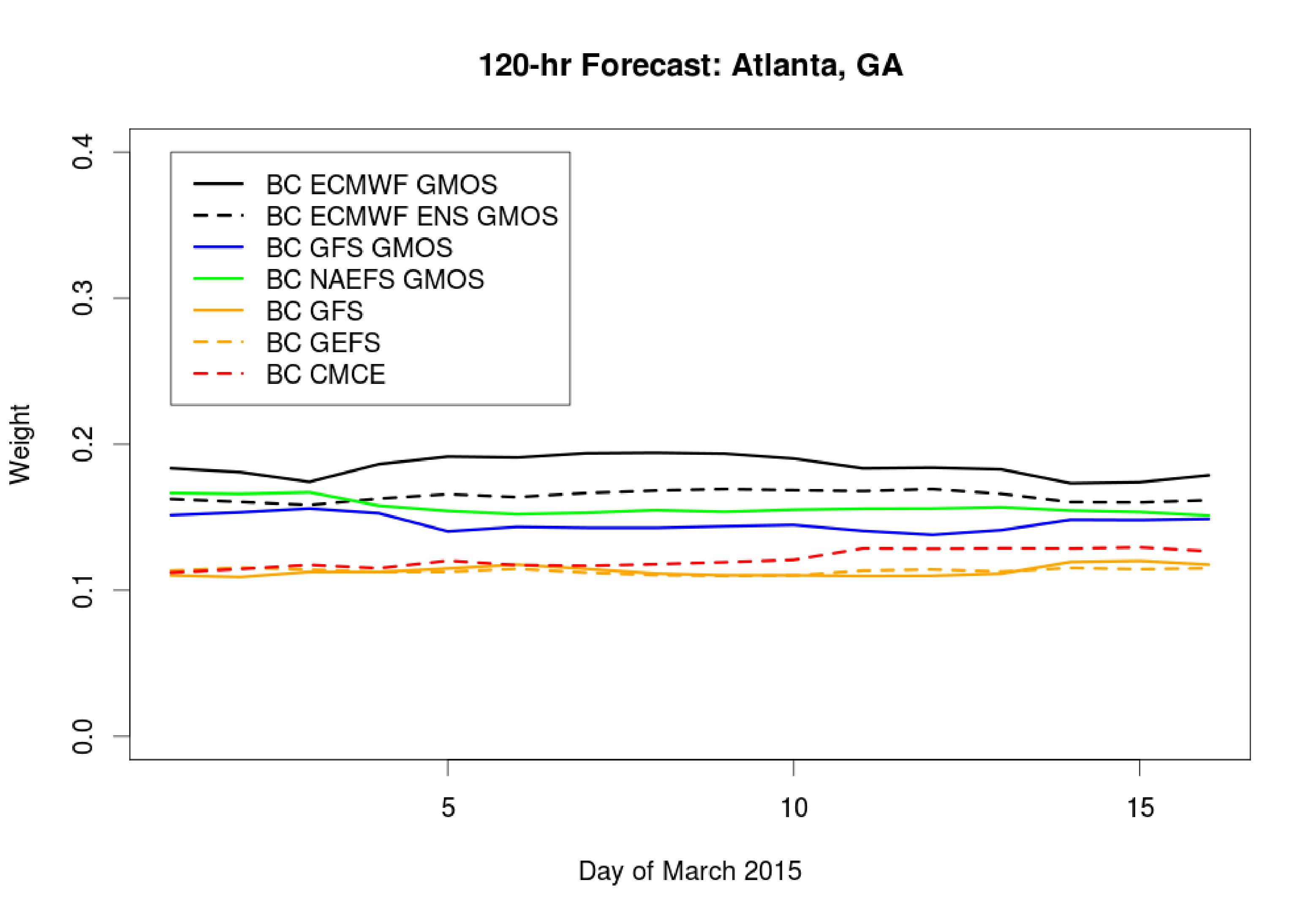


Part 3: Blend vs. ECMWF GMOS

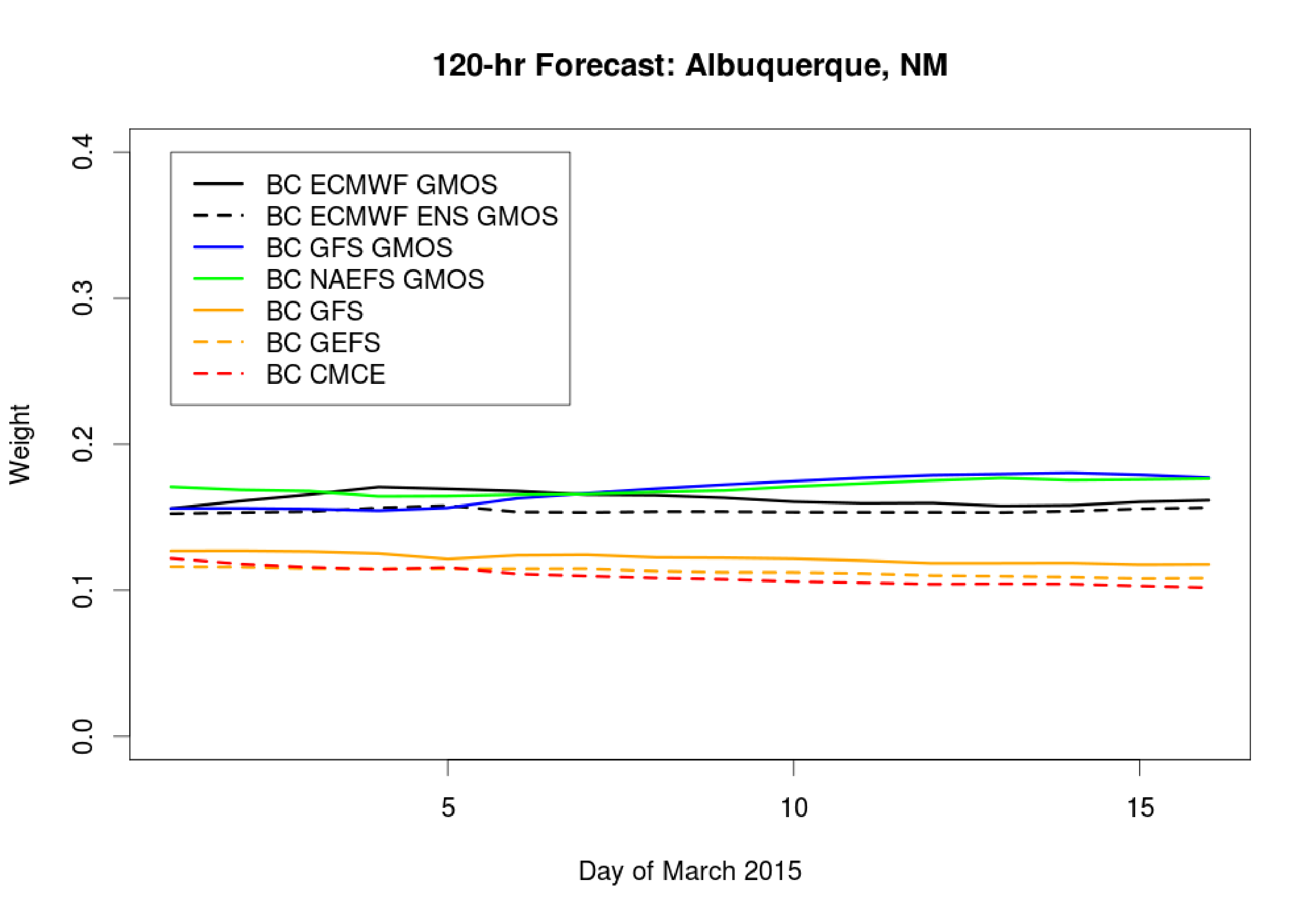
- Is the Blend more skillful than the single best component?
- Interpolated Blend forecast grids to stations and verified relative to station-based observations
- Compared with bias-corrected ECMWF GMOS grids interpolated to stations
- 2-m Temperature
- 1 Jan. 2015 – 26 March 2015


2-m Temperature MAE,
300 Stations,
1 Jan. 2015 - 26 Mar. 2015
MAE [C]
0
1
2
3
4
5
0
24
48
72
96
120
144
168
192
Projection [h]
Blend
BC ECMWF GMOS

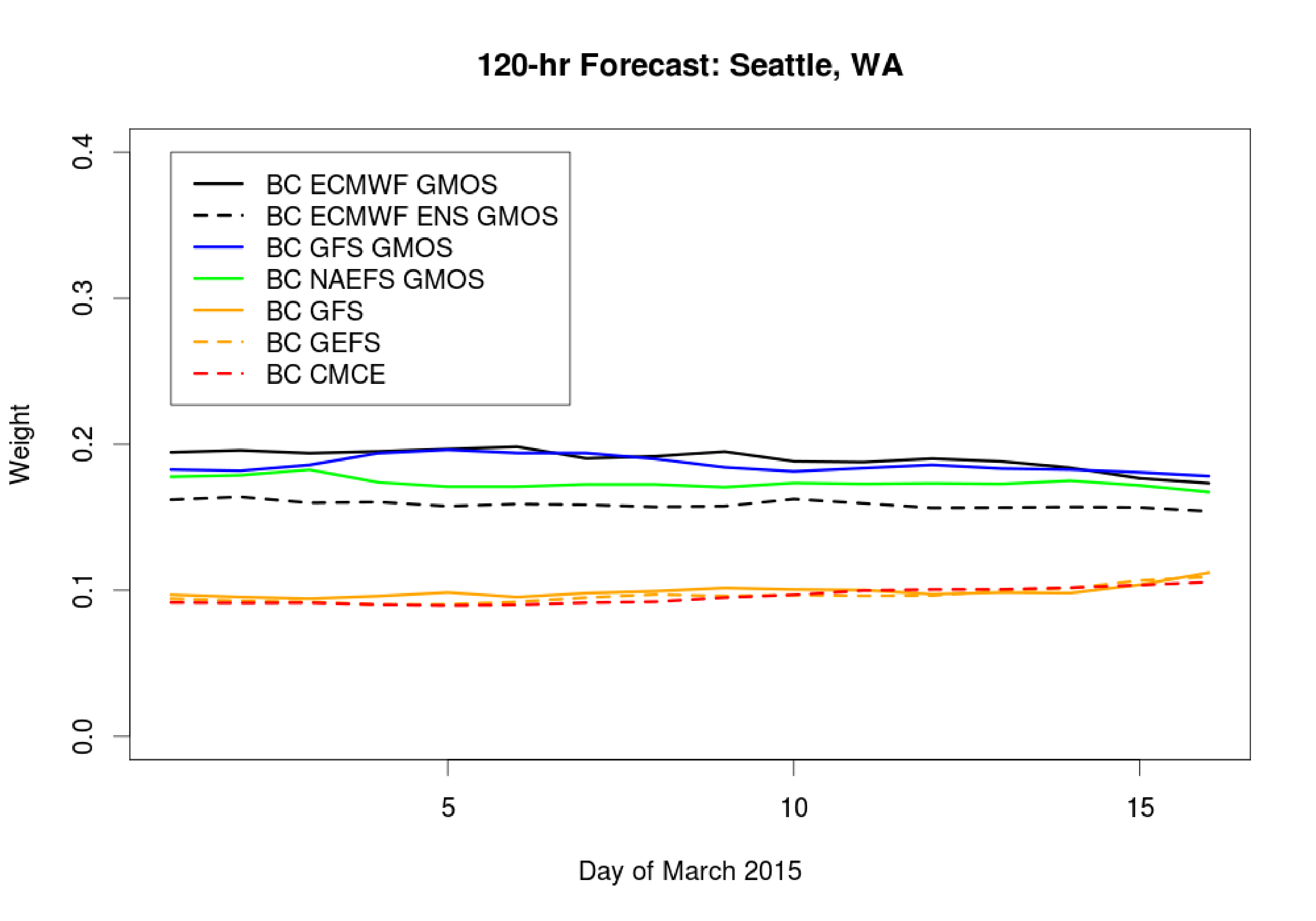
Part 3: Example Weights 2-m Temperature



Part 3: Example Weights 2-m Temperature



Part 3: Example Weights 2-m Temperature



Summary

- Prototype Blend is created by weighting the bias-corrected components using an MAE-based weighting scheme
- MAE-weighted Blend is more skillful than the equally-weighted Blend
- Plan to use technique outlined here for 2-m temperature, 2-m dewpoint, daytime maximum temperature and nighttime minimum temperature

Future Work

- National Blend of Models (NBM) prototype temperature and dewpoint grids are being produced on the development WCOSS platform
- Daytime Maximum and Nighttime minimum grids will be added soon.
- Blend Version 1 scheduled for operational implementation in December 2015
  - CONUS Domain
  - 2-m Temp, Dew, Max T, Min T, AppT, RH, POP 12, sky cover, wind speed and direction

References

- Cui, B., Z. Toth, Y. Zhu, and D. Hou, 2012: Bias Correction for Global Ensemble Forecast. *Wea. Forecasting*, **27**, 396–410.
- Peña, M., and H.M. van den Dool, 2008: Consolidation of Multimodel Forecasts by Ridge Regression: Application to Pacific Sea Surface Temperature. *J. Climate*, **21**, 6521–6538.
- Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174.
- Roberts, S. V., 1959: Control chart tests based on geometric moving averages. *Technometrics* ,**1**,239-250.
- Veenhuis, B. A., 2014: A practical model blending technique based on Bayesian model averaging. *Preprints, 22st Conference on Probability and Statistics in the Atmospheric Sciences*, Atlanta, Amer. Meteor. Soc.
- Woodcock, F., and C. Engel, 2005: Operational consensus forecasts. *Wea. Forecasting*, **20**, 101–111.