Chemical Data Assimilation and Analog-Based Uncertainty Quantification to Improve Decision-Making in Public Health and Air Quality

Rajesh Kumar
NCAR, Boulder, CO

National Center for Atmospheric Research (NCAR)
Research Application Laboratory (RAL)
National Security Application Program (NSAP)

Outline

- Project team
- Project goal and objectives
- Tasks

Project Team

Principal Investigator:	Rajesh Kumar (NCAR/RAL, Boulder CO) / Luca Delle Monache (Scripps/UCSD)
Co-Principal Investigator:	Gabriele Pfister (NCAR/ACOM, Boulder CO)
Co-Investigators:	Stefano Alessandrini (NCAR/RAL, Boulder CO) Barry Baker ((UMD, College Park, MD) Jamie Bresh (NCAR/MMM, Boulder CO) Irina Djalalova (CU Boulder, Boulder CO) David Edwards (NCAR/ACOM, Boulder, CO) Zhiquan Liu (NCAR/MMM, Boulder CO) Youhua Tang (UMD, College Park, MD)
Collaborators:	Pius Lee (NOAA/ARL, College Park, MD) Pablo Saide (UCLA, Los Angeles, CA) James Wilczak (NOAA/ESRL, Boulder CO)

Project Goal & Objectives

Goal:

- National Oceanic and Atmospheric Administration (NOAA) / National Centers for Environmental Prediction (NCEP) air quality (AQ) forecasting system is a key tool for decision makers across the U.S. to protect the public from poor AQ
- To enhance this decision-making activity this project aims to improve the accuracy of NOAA/NCEP short-term predictions of ground-level ozone (O_3) and particulate matter less than 2.5 µm in diameter ($PM_{2.5}$) and to provide reliable quantification of their uncertainty

Objectives:

① Improve initialization of NOAA/NCEP Community Multiscale AQ (CMAQ) model through chemical data assimilation of satellite retrieval products and in-situ observations with the Community Gridpoint Statistical Interpolation (GSI) system

① Improve CMAQ prediction accuracy and reliably quantify their uncertainty with analog-based post-processing methods

Main Tasks

① Generating the analysis and deterministic forecasts of O_3 and $PM_{2.5}$

① Analog-based methods for deterministic and probabilistic predictions of O_3 and $PM_{2.5}$

② Two-dimensional gridded deterministic and probabilistic predictions

③ Transition to operations of the new AQ forecasting capability

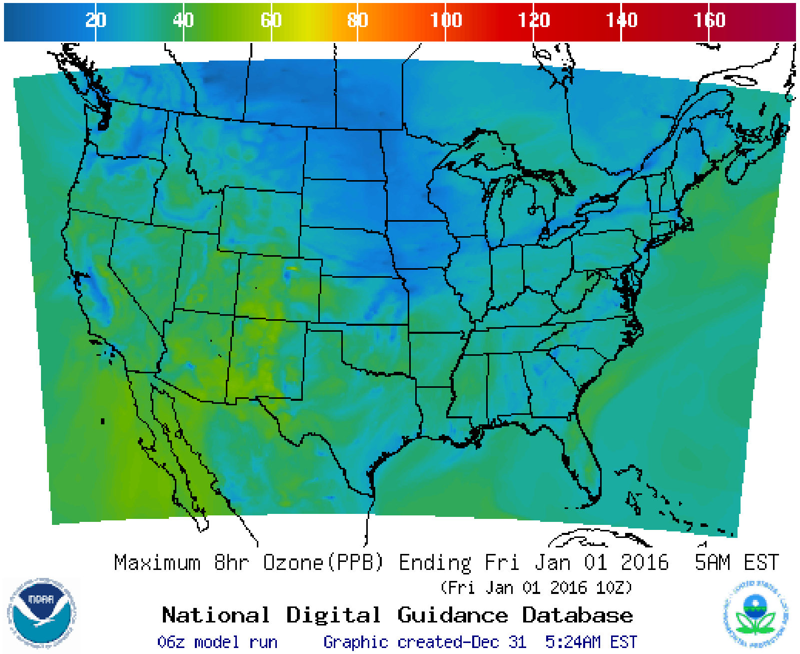
Current Status: Task 1

Generating the analysis and deterministic forecasts of O_3 and $PM_{2.5}$

- Chemical transport modeling and emission processing
- Assimilation with the GSI/CMAQ system of:
  - ✧ Aerosol optical depth from NASA Aqua/Terra Moderate Resolution Imaging Spectroradiometer (MODIS) satellite instruments
  - ✧ Retrieval of carbon monoxide from the NASA/Terra Measurements Of Pollution In The Troposphere (MOPITT)
  - ✧ Surface observations of $PM_{2.5}$ (and possibly of ground-level ozone) from the AIRNow network, the Interagency Monitoring of Protected Visual Environments (IMPROVE) stations, and the Clean Air Status and Trends Network (CASTNET)

CMAQ model configuration

National Air Quality Forecasting Capability Domain

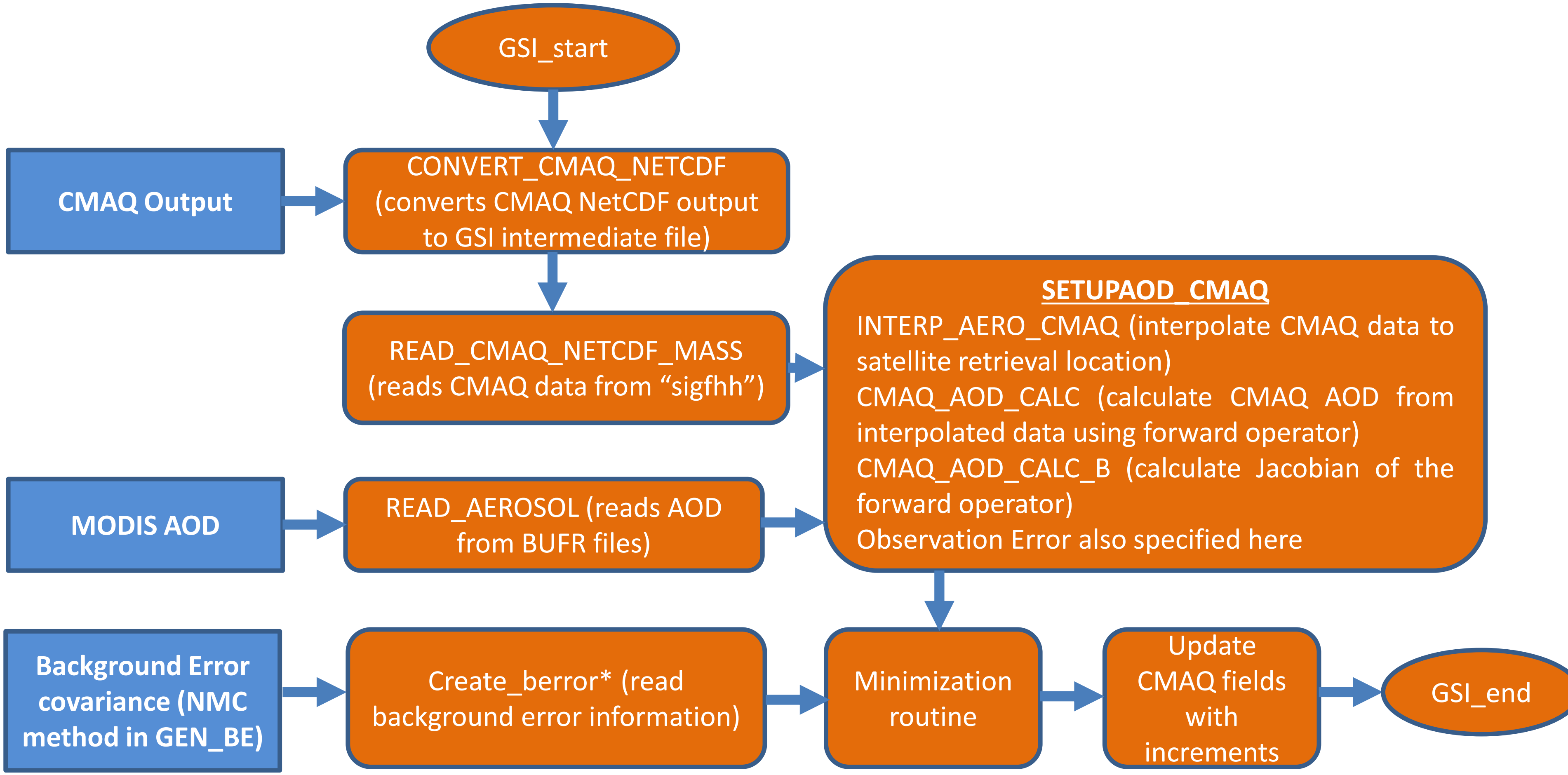


CMAQ version – 5.1
CMAQ resolution – 12 km^2
Emissions – NEI 2011
Biogenic emissions – Online (BEIS)
Photolysis rates –Online
Other configuration option – Consistent with NAQFC
Initial conditions – Use previous CMAQ run
Boundary conditions – constant similar to the NAQFC

Background error generation

Ran 24-h CMAQ forecasts for the FRAPPE period (15 Jul -15 Aug 2014) and fed to GEN_BE to generate BEC matrix

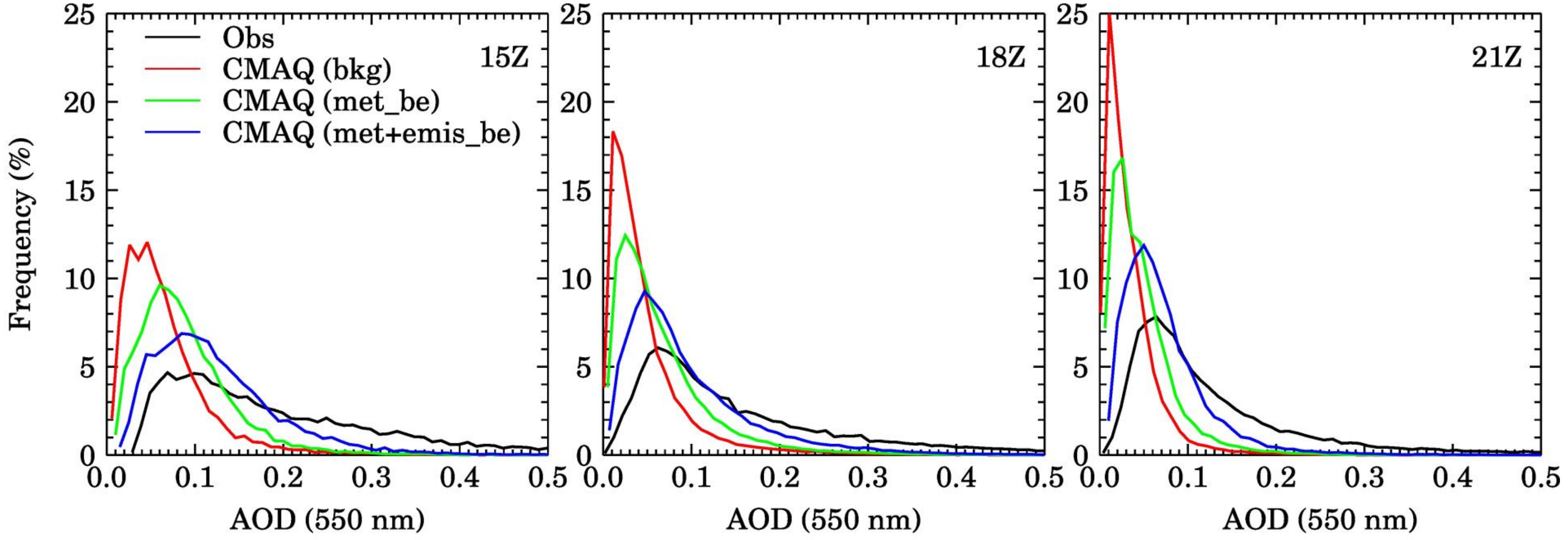
GSI/CMAQ code development



NSAP/RAL/NCAR – National Security Applications Program

CMAQ-MODIS AOD comparison

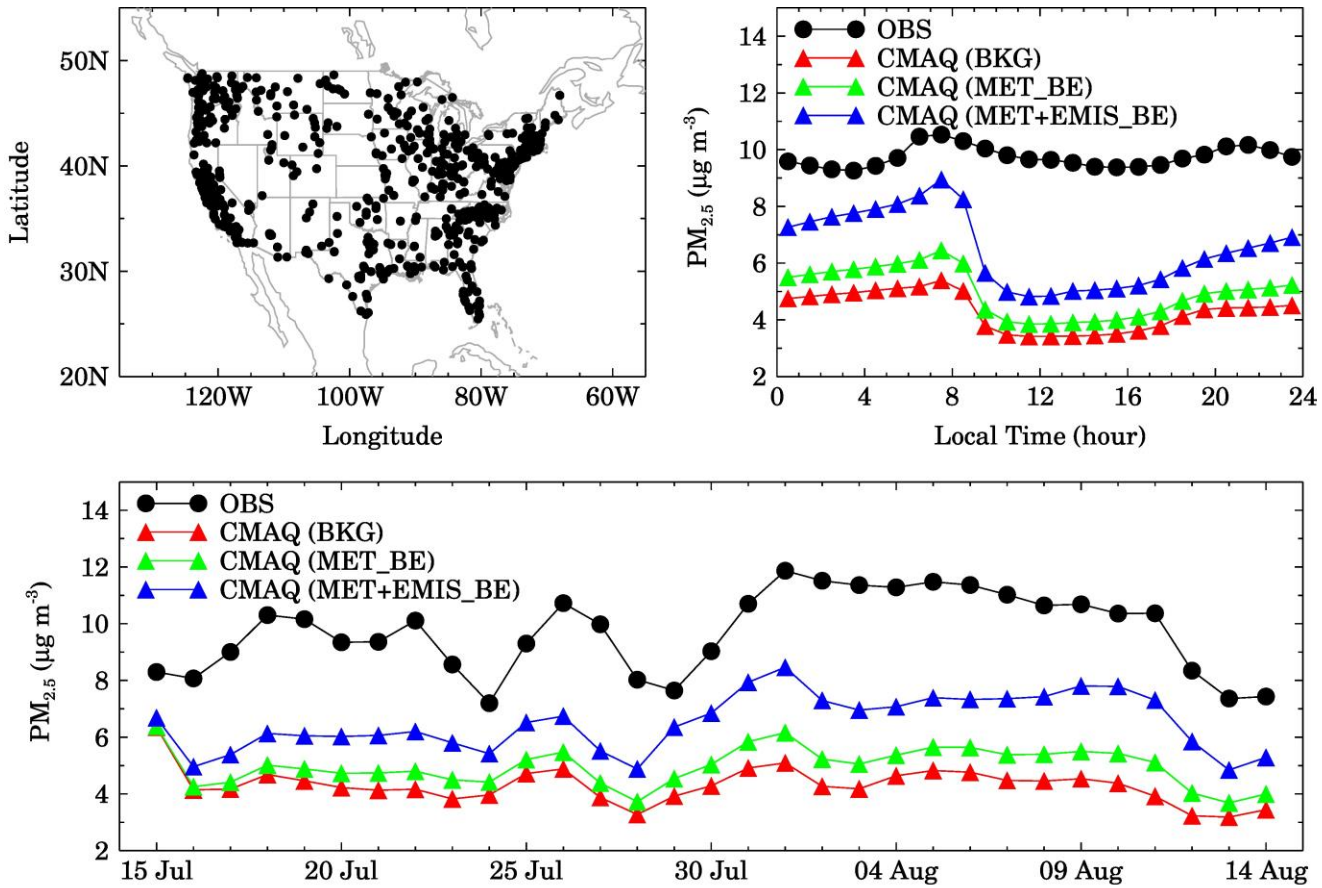
MODIS AOD from NASA neural network (NNR) at 10 km resolution provided by GMAO is used



We can tune the background and observation errors to improve agreement between MODIS and CMAQ AOD (with assimilation) but our objective here is to represent the model errors realistically.

[Kumar et al., revised submission, JGR, 2018]

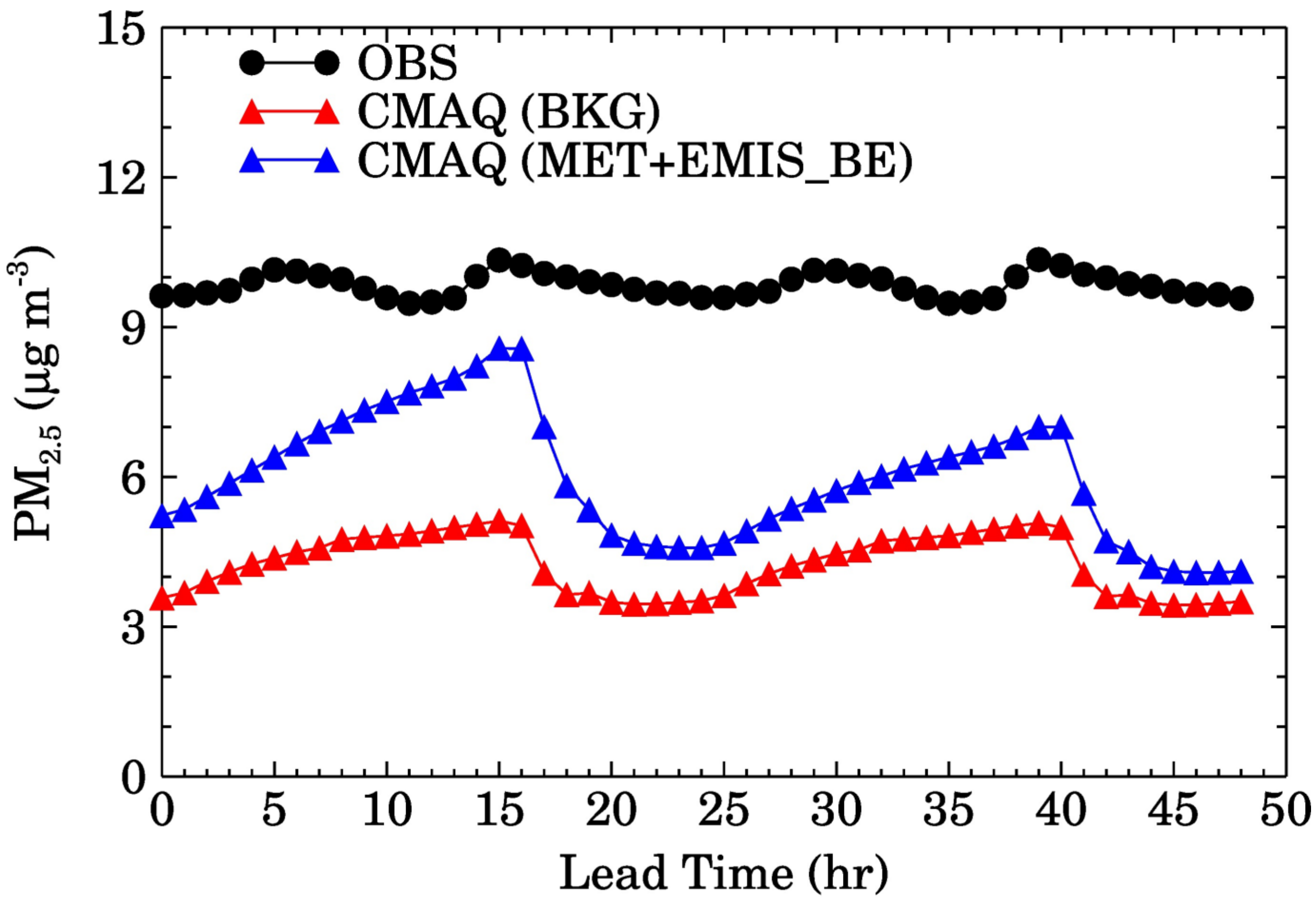
Effect of assimilation on CMAQ $PM_{2.5}$



- The assimilation of MODIS AOD in CMAQ model improves the correlation coefficient between the model and observed $PM_{2.5}$ by ~48-67% and reduces the mean bias by ~20-38%.
- Large improvements are seen at more than 80% of the AirNOW sites.

[Kumar et al., revised submission, JGR, 2018]

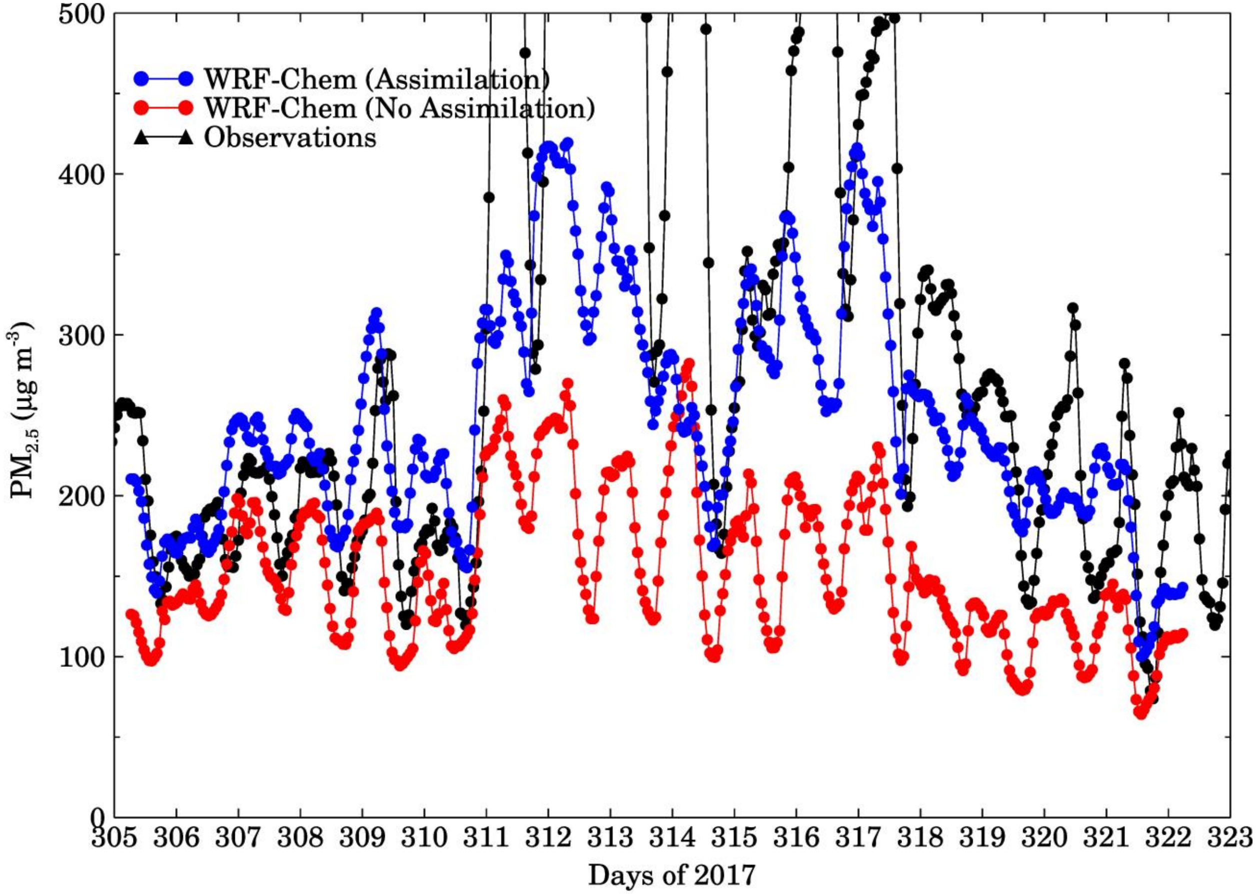
Effect on 48-h $PM_{2.5}$ forecasts



Improving initial conditions with MODIS AOD assimilation improves $PM_{2.5}$ forecasts for more than 24 hours.

[Kumar et al., revised submission, JGR, 2018]

NCAR Improving air quality forecasts in Delhi



Assimilation of satellite AOD retrievals can significantly help developing countries in air quality management.

Current Status: Task 2

Analog-based methods for deterministic and probabilistic predictions of O_3 and $PM_{2.5}$

- Improving deterministic predictions with analog-based post-processing methods
- Providing uncertainty quantification of O_3 and $PM_{2.5}$ predictions, that is crucial information for effective decision-making to protect the public health

References:
Delle Monache et al., *Monthly Weather Review* 2011, 2013; Djalalova et al., *Atmospheric Environment*, 2015

The Analog Ensemble

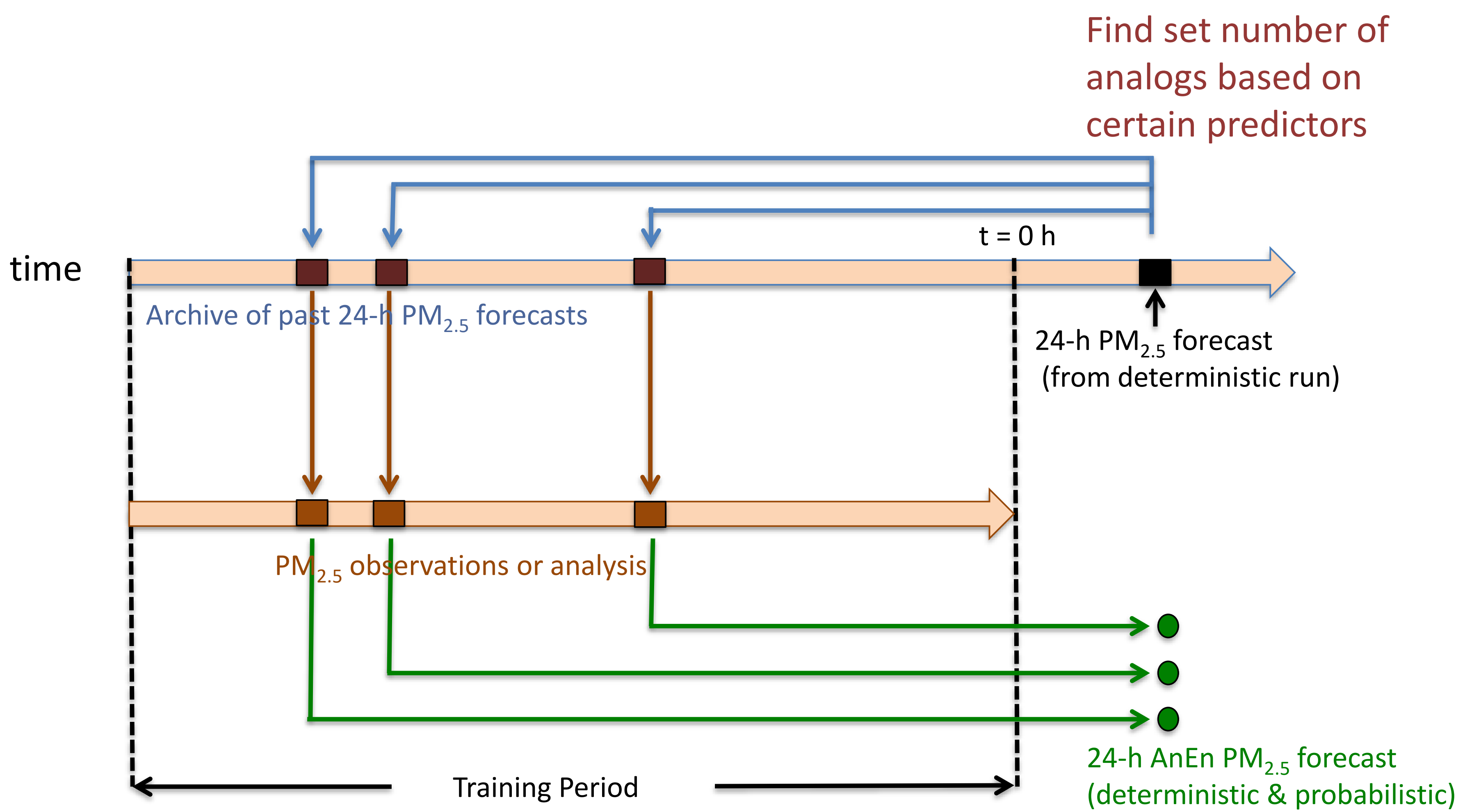
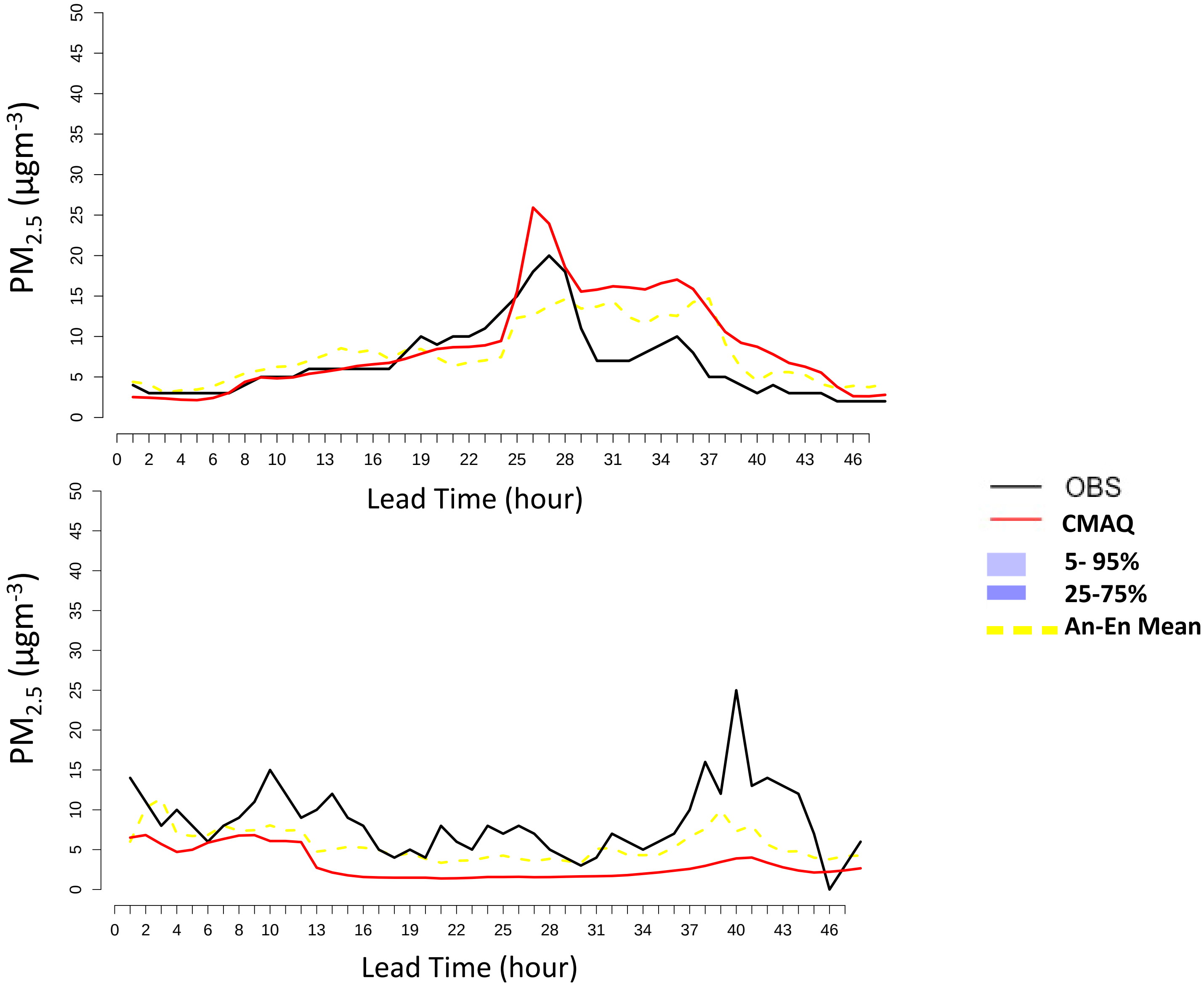


Figure adapted from Delle Monache et al. (2013)

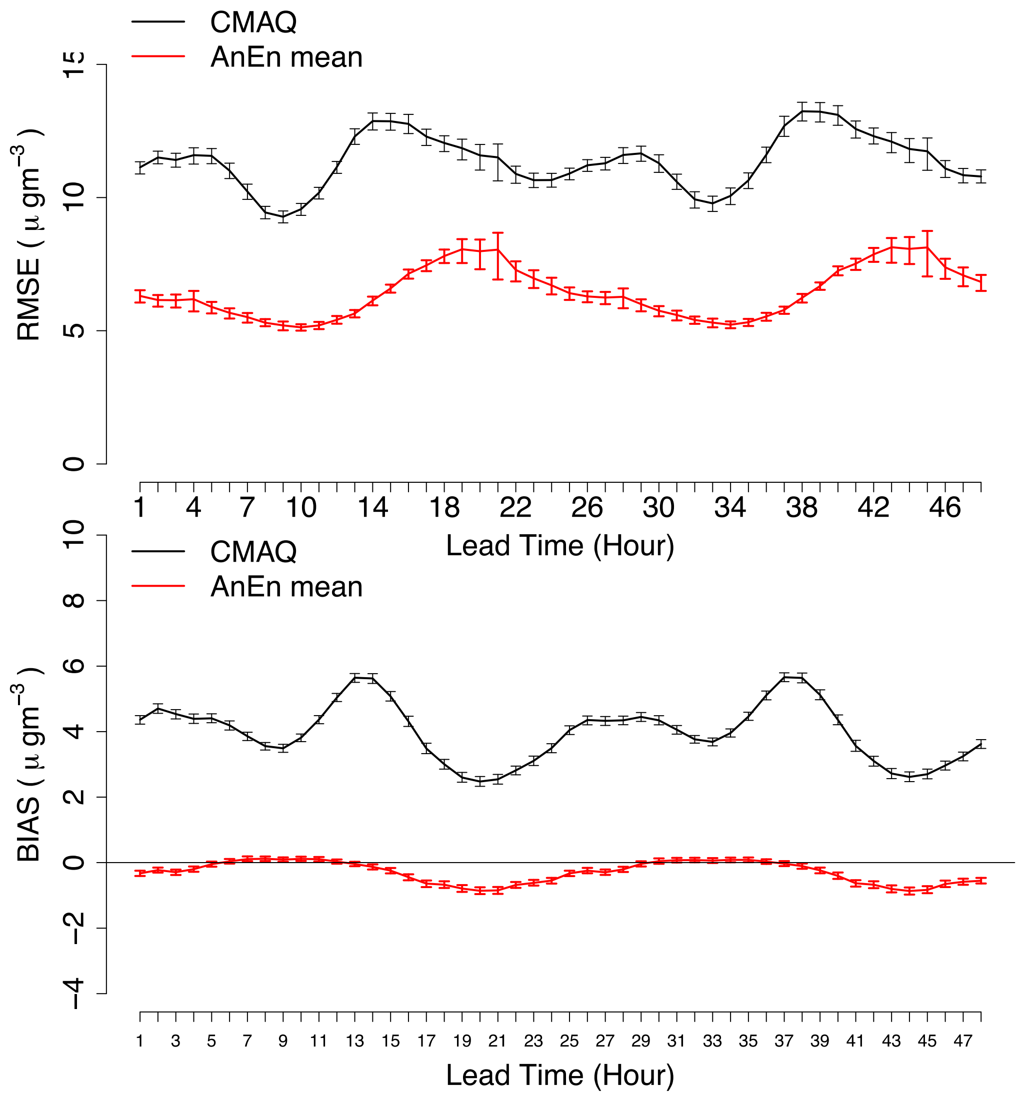
Example of 48-h $PM_{2.5}$ probabilistic predictions



AnEn generates forecast quantiles that provide uncertainty quantification

[Delle Monache et al., revised submission, ACP, 2018]

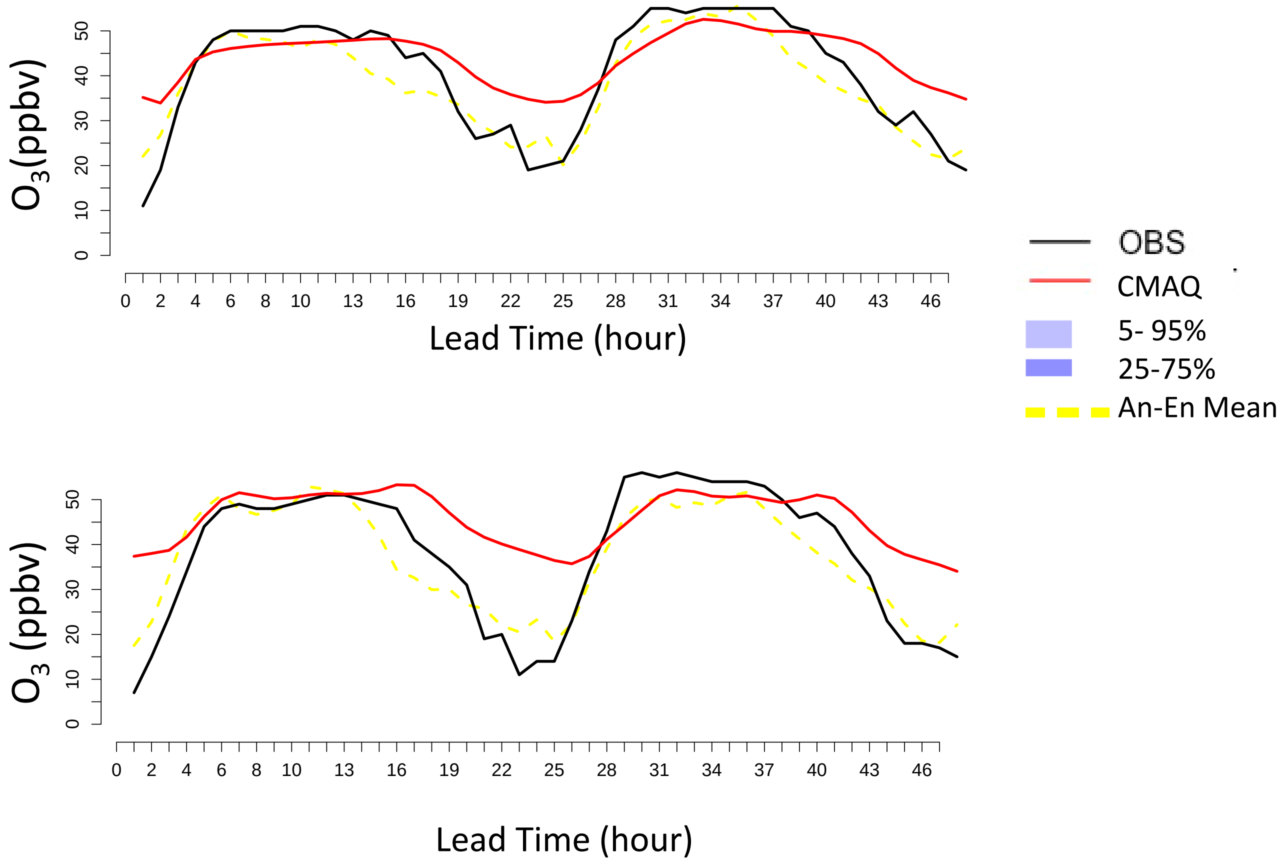
Drastic reduction of CMAQ errors



AnEn error reductions with respect to CMAQ: ~50% RMSE, ~95% BIAS

[Delle Monache et al., revised submission, ACP, 2018]

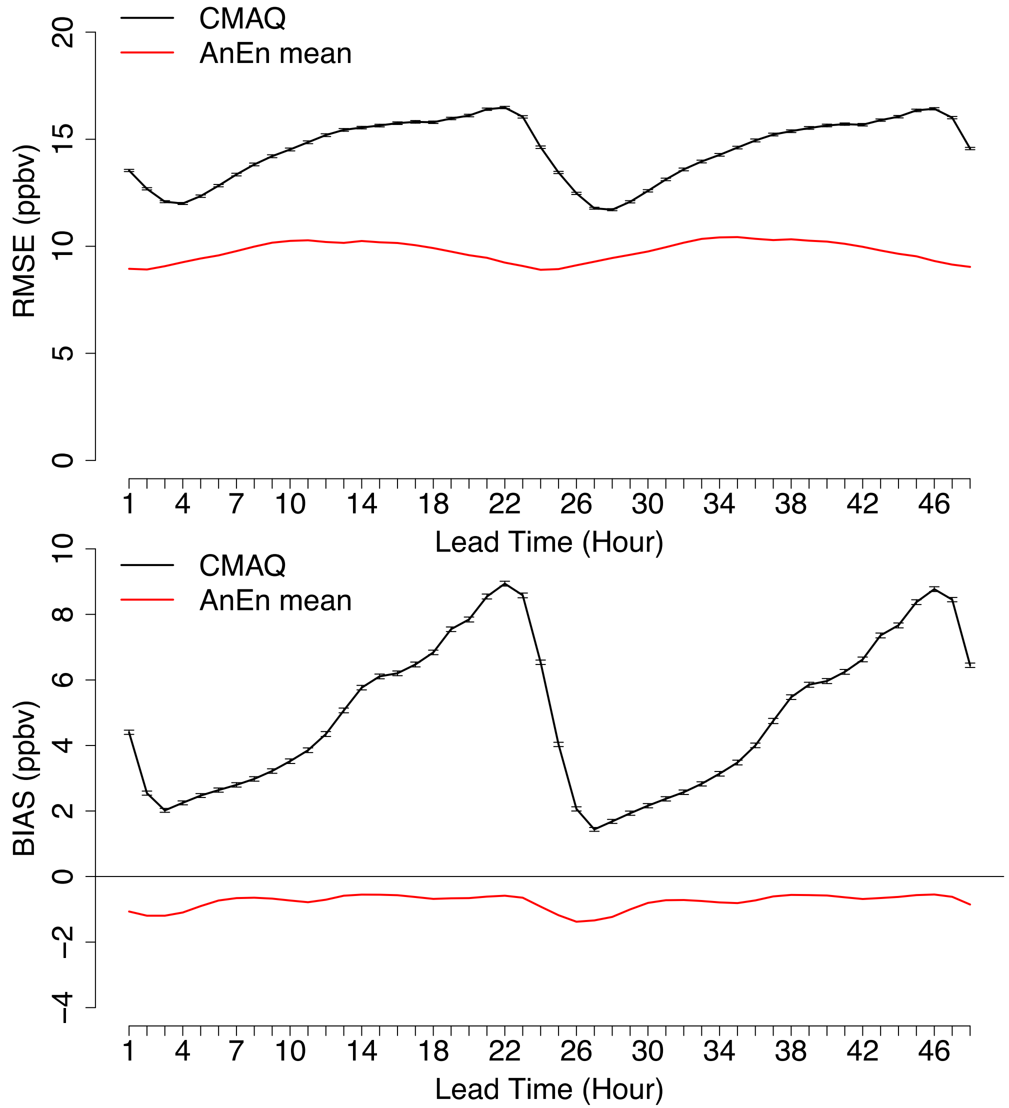
AnEn, CMAQ output



AnEn generates forecast quantiles that provide information about uncertainty quantification

[Delle Monache et al., revised submission, ACP, 2018]

AnEn, CMAQ verification



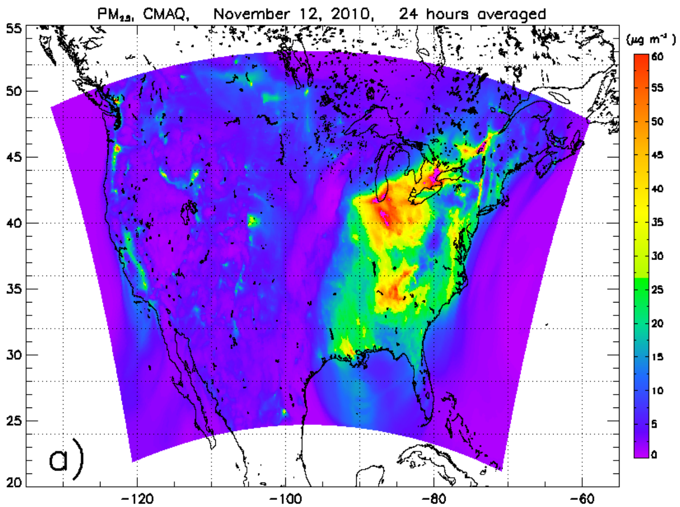
AnEn mean reduces CMAQ's RMSE by ~35%

AnEn mean reduces CMAQ's BIAS by ~85%

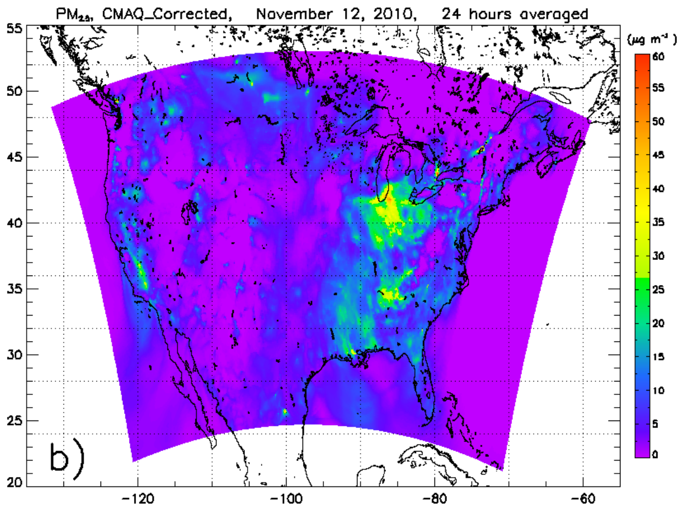
[Delle Monache et al., revised submission, ACP, 2018]

Current Status: Task 3

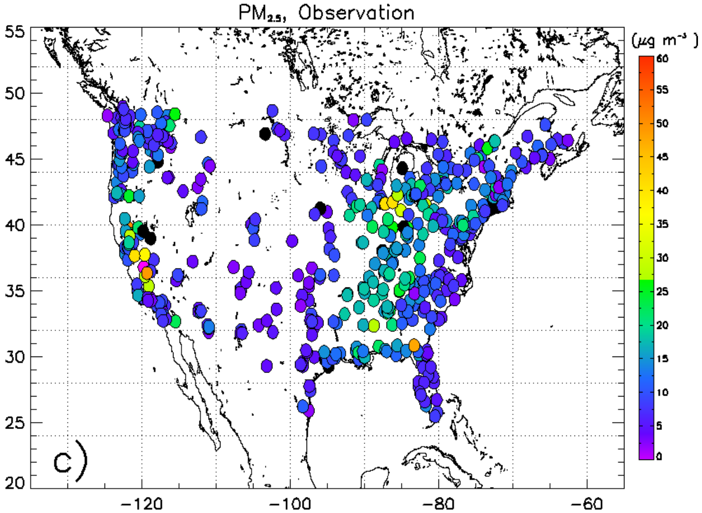
Two-dimensional gridded deterministic and probabilistic predictions



CMAQ



KFAN



OBSERVATION

Reference:
Djalalova et al., *Atm. Envi.*, 2015

Current Status: Task 4

Transition to operations of the new AQ forecasting capability

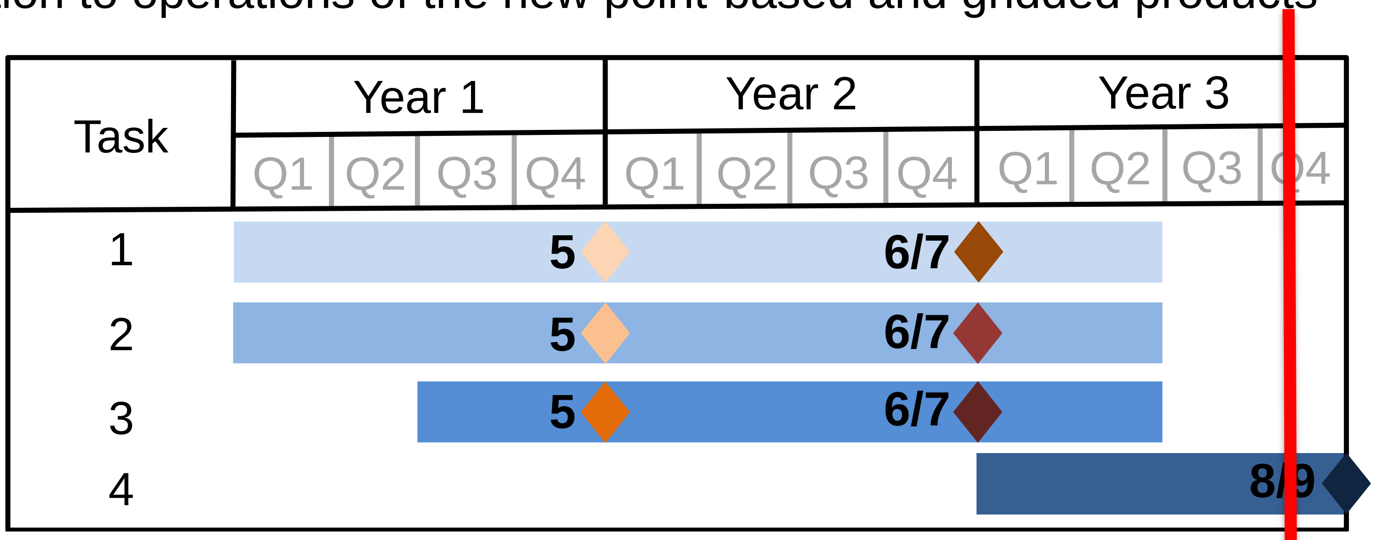
Transition to operations of $PM_{2.5}$ deterministic predictions:

- Chemical data assimilation system transferred to the research team of NAQFC
- Automatic quality control procedures have been developed to eliminate spurious measurement values
- Analog-based method (for now only simple analog ensemble mean)
- Spreading technique to generate gridded maps
- Running operationally since 20 October 2015
- NOAA/NCEP very satisfied with results
- **This subtask is already at ARL 8/9**

Reference: Djalalova et al., *Atmospheric Environment*, 2015

Plans for Coming Year

1. Complete development and test of AOD assimilation system; start assimilation of CO and/or surface $PM_{2.5}$; test new forward operator; manuscript submission
2. Complete tests of point-based O_3 predictions and submit manuscript on analog-based methods for both O_3 and $PM_{2.5}$ of point-based predictions
3. Development of 2D gridded maps for deterministic O_3 and $PM_{2.5}$ ensemble
4. Transition to operations of the new point-based and gridded products



Milestones:
GSI/CMAQ chemical DA, ARL-5 (◆); analog-based methods, ARL-5 (◆);
2D maps, ARL-5 (◆); GSI/CMAQ chemical DA, ARL-6/7 (◆);
analog-based methods, ARL-6/7 (◆); 2D maps, ARL-6/7 (◆); all capabilities, ARL-8/9 (◆)

NCAR

Thanks! Questions?

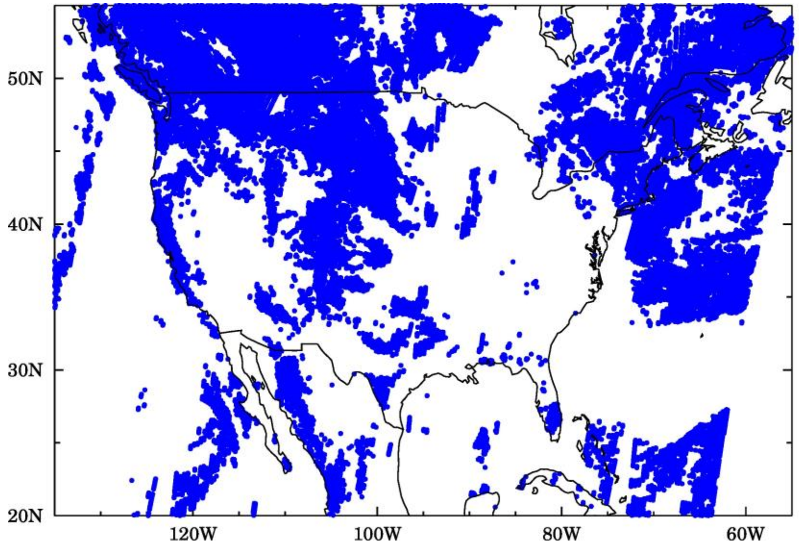
New Task

- Title:
  "Socioeconomic benefits of improved forecasts on decision-making in public health and air quality"
- Period of Performance: 1 October 2016 – 30 September 2017
- Budget: $250K
- Team:
  - Jeffrey Lazo (Lead), Luca Delle Monache − NCAR
  - James Hammitt, Lisa Robinson − Harvard Center for Risk Analysis
  - Lauraine Chestnut, David Mills − Abt Associates

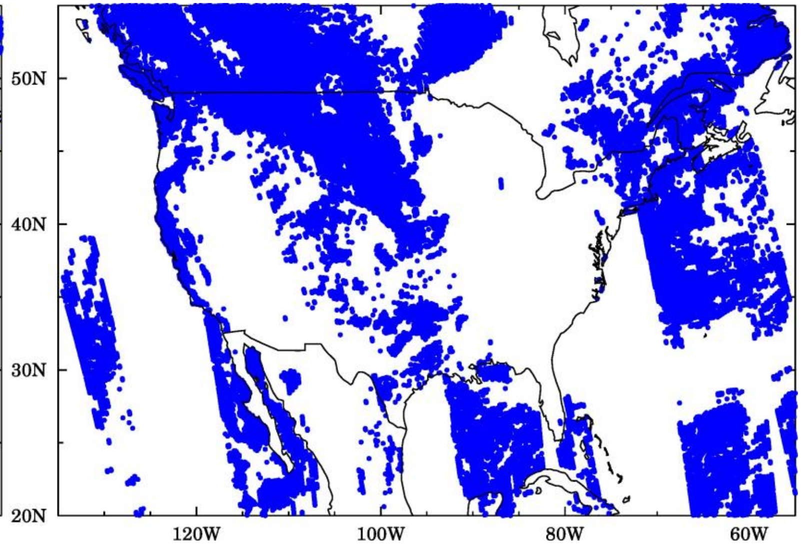
MODIS Observations for Assimilation

MODIS AOD from NASA neural network at 10 km resolution provided by GMAO is used

Typical 1-day coverage MODIS Terra (10:30 am LT)

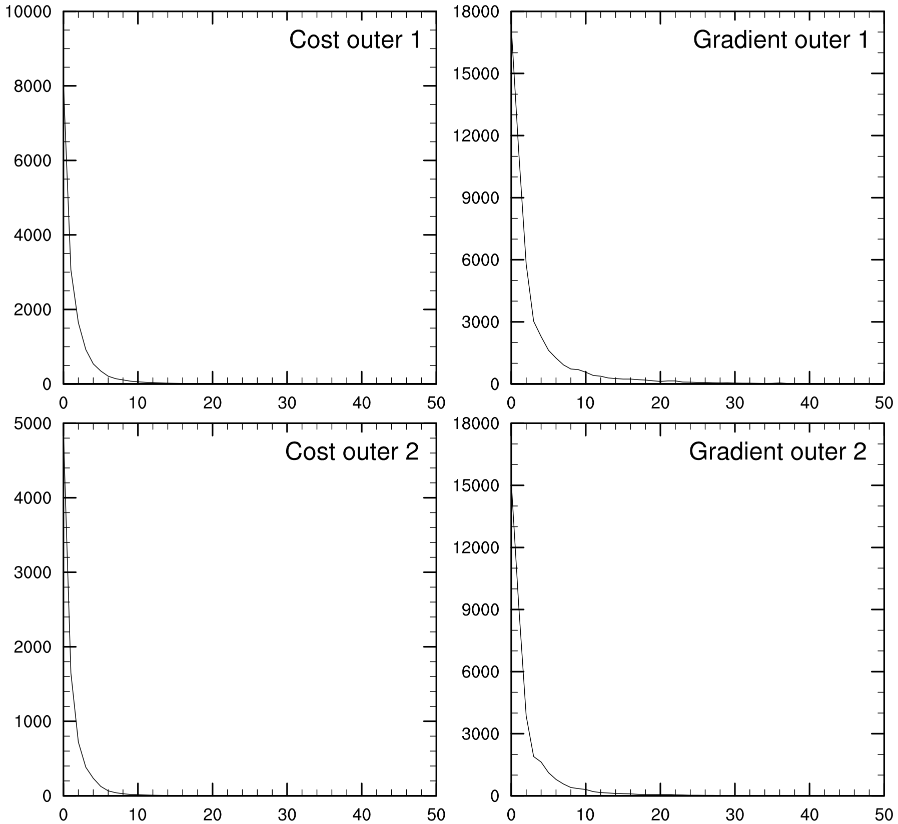


Typical 1-day coverage MODIS Aqua (1:30 pm LT)



August 12, 2014

Cost function and gradient minimization



No. of inner loops

Tangent Linear Test of Forward Operator

The tangent linear code "P" is tested against the forward operator code "Q" using the Taylor-Lagrange formula:

$$\frac{\boldsymbol{Q(C+h*C)-Q(C)}}{\boldsymbol{P\ (h\ *C)}}=\boldsymbol{1}$$

C: CMAQ aerosol chemical composition.
h: perturbation factor (0.1 to 10^{-9}).

h = 0.10E+00 ratio = 0.10000000E+01
h = 0.10E-01 ratio = 0.10000000E+01
h = 0.10E-02 ratio = 0.10000000E+01
h = 0.10E-03 ratio = 0.10000000E+01
h = 0.10E-04 ratio = 0.10000000E+01
h = 0.10E-05 ratio = 0.10000000E+01
h = 0.10E-06 ratio = 0.10000000E+01
h = 0.10E-07 ratio = 0.10000000E+01
h = 0.10E-08 ratio = 0.99999997E+00
h = 0.10E-09 ratio = 0.10000004E+01

Adjoint Test

Adjoint code "P^T" is tested against TL code "P" using the following adjointness relation:

$$\langle \boldsymbol{P}(\boldsymbol{C}+\boldsymbol{h}*\boldsymbol{C}), \boldsymbol{P}(\boldsymbol{C}+\boldsymbol{h}*\boldsymbol{C})\rangle = \left\langle (\boldsymbol{C}+\boldsymbol{h}*\boldsymbol{C}), \boldsymbol{P}^{\boldsymbol{T}}\boldsymbol{P}(\boldsymbol{C}+\boldsymbol{h}*\boldsymbol{C})\right\rangle$$

h = 0.10E+00 LHS = 0.32998862E+00 RHS = 0.32998862E+00
h = 0.10E-01 LHS = 0.32998862E-02 RHS = 0.32998862E-02
h = 0.10E-02 LHS = 0.32998862E-04 RHS = 0.32998862E-04
h = 0.10E-03 LHS = 0.32998862E-06 RHS = 0.32998862E-06
h = 0.10E-04 LHS = 0.32998862E-08 RHS = 0.32998862E-08
h = 0.10E-05 LHS = 0.32998862E-10 RHS = 0.32998862E-10
h = 0.10E-06 LHS = 0.32998862E-12 RHS = 0.32998862E-12
h = 0.10E-07 LHS = 0.32998862E-14 RHS = 0.32998862E-14
h = 0.10E-08 LHS = 0.32998862E-16 RHS = 0.32998862E-16
h = 0.10E-09 LHS = 0.32998862E-18 RHS = 0.32998862E-18

Spreading method

NCAR

- To create graphical images of corrected CMAQ PM2.5 forecasts, the forecasted bias calculated at each AIRNow obs location must be spread to every model gridpoint.

- An iterative objective analyses method is used which starts with a very large radius of influence (R=2000km).

- $Sbias_k$ = $CMAQ\_KFAN_k$ - $CMAQ\_RAW_k$, $Mbias_{i,j} = 0$

- At each grid point the correction value is calculated as

  $C_{i,j} = \frac{1}{n} \Sigma \frac{R*R-d*d}{R*R+d*d} (Sbias_k - Mbias_{i,j})$, $d<R$,where

  R is radius of influence;

  d is the distance from a grid point to the site k inside the circle R;

  $C_{i,j}$ is the correction at a grid point;

  $Mbias_{i,j} = Mbias_{i,j} + C_{i,j}$

  Summation is done over ALL obs sites k inside the circle R.

- 8 passes with R=2000, 1000, 500, 250, 125, 62, 31 & 15 km are used.

NSAP/RAL/NCAR – National Security Applications Program

Spreading technique based on Glahn et al., 2012

NCAR

AnEn has been successfully applied for:

- Short-term predictions of:
  - 10- and 80-m wind speed, 2-m temperature, etc.
    Delle Monache et al. MWR 2011,2013, Junk et al. MZ 2015
  - Wind power
    Alessandrini et al. RE 2015, Davo et al. SE 2016
  - Solar GHI
    Alessandrini et al. SE 2015
  - Load **Alessandrini et al. ICEM 2015**
  - Air quality predictions (ground level ozone, surface $PM_{2.5}$)
    Djalalova et al. AE 2015, Delle Monache et al. JGR 2016
  - Tropical cyclones intensity
    Alessandrini et al. MWR 2016
- Downscaling, resource assessment:
  Vanvyve et al. RE 2015, Zhang et al. AE 2015
  - Wind speed
  - Computationally efficient dynamical downscaling

Available EPA CMAQ O_3 data

- Community Multiscale Air Quality (CMAQ) Modeling System daily run (457 days, 12 UTC), lead time from 0 to 48 hours
- Available variables ($O_{3,}$ wind speed and wind direction, 2-m temperature, cloud fraction)

AnEn configuration ($PM_{2.5}$)

Continuous Ranked Probability Score (CRPS) minimization

Training (1-302 days)

Optimization
CRPS (Nov. 2014)

Test (62 days,
Dec 2014 & Jan 2015)

- 20 historic analog ensemble members
- 4 predictors with different weighting
  ($PM_{2.5}$, wind speed and wind direction, 2-m temperature)
- Analog-predictor weights obtained by an optimization algorithm (minimizing CRPS) over November 2014, performed independently at each station
- Possible weights for each predictor: 1, 0.9, 0.8,…..0.1, 0.

AnEn configuration (O_3)

Continuous Ranked Probability Score (CRPS) minimization

Training (1-304 days)

Test 153 days,
May 2015 - Sep 2015)

- 20 historic analog ensemble members
- 5 predictors with different weighting:

  (O_3, wind speed and wind direction, 2-m temperature, cloud cover)
- Analog-predictor weights assigned are fixed (optimization not done yet)
- Weights assigned: O_3 (0.5), wind speed (0.2), wind dir. (0.1), 2-m temp. (0.1), cloud cover (0.1)

Observation operator design

Aerosol chemical composition to AOD

- Use the parametric expression of Malm and Hand (2007) to calculate AOD from CMAQ aerosol composition.

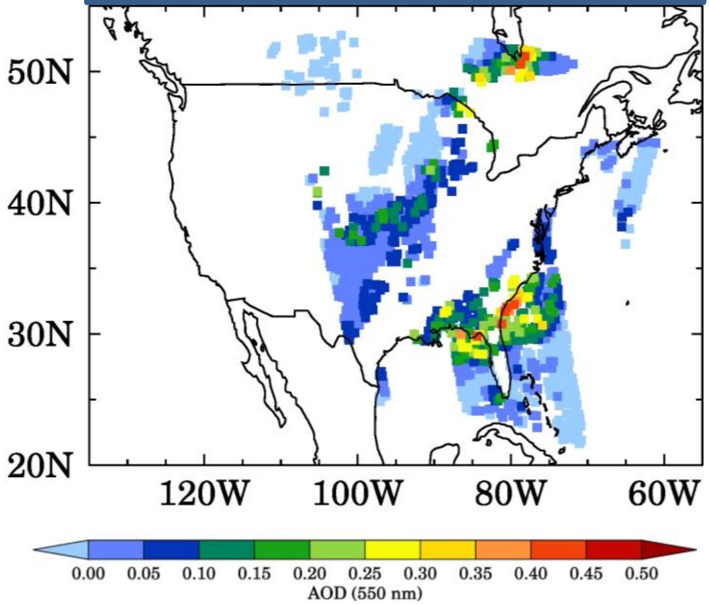
AOD = ∑Ext. Coeff. (i) * dz (i); i = 1, 2,…, N represent CMAQ vertical layers

Ext. Coeff. (1/km) = 0.003 * f(RH) * {[ammonium sulfate] + [ammonium nitrate]} + 0.004 * [organic mass] + 0.010 * [Light absorbing carbon] + 0.001 * [fine soil]+ 0.0006* [coarse mass]

f(RH) is a relative humidity correction factor and is based on a look up table from Malm et al., (1994).

- Advantages: simple to implement, consistent with visibility calculation in CMAQ.

An example of AOD calculation by the forward operator



NOTE: The tangent linear and adjoint of the forward operator has been generated with the automatic differentiation tool TAPENADE, www-tapenade.inria.fr:8080/tapenade/

The Analog Ensemble (1)

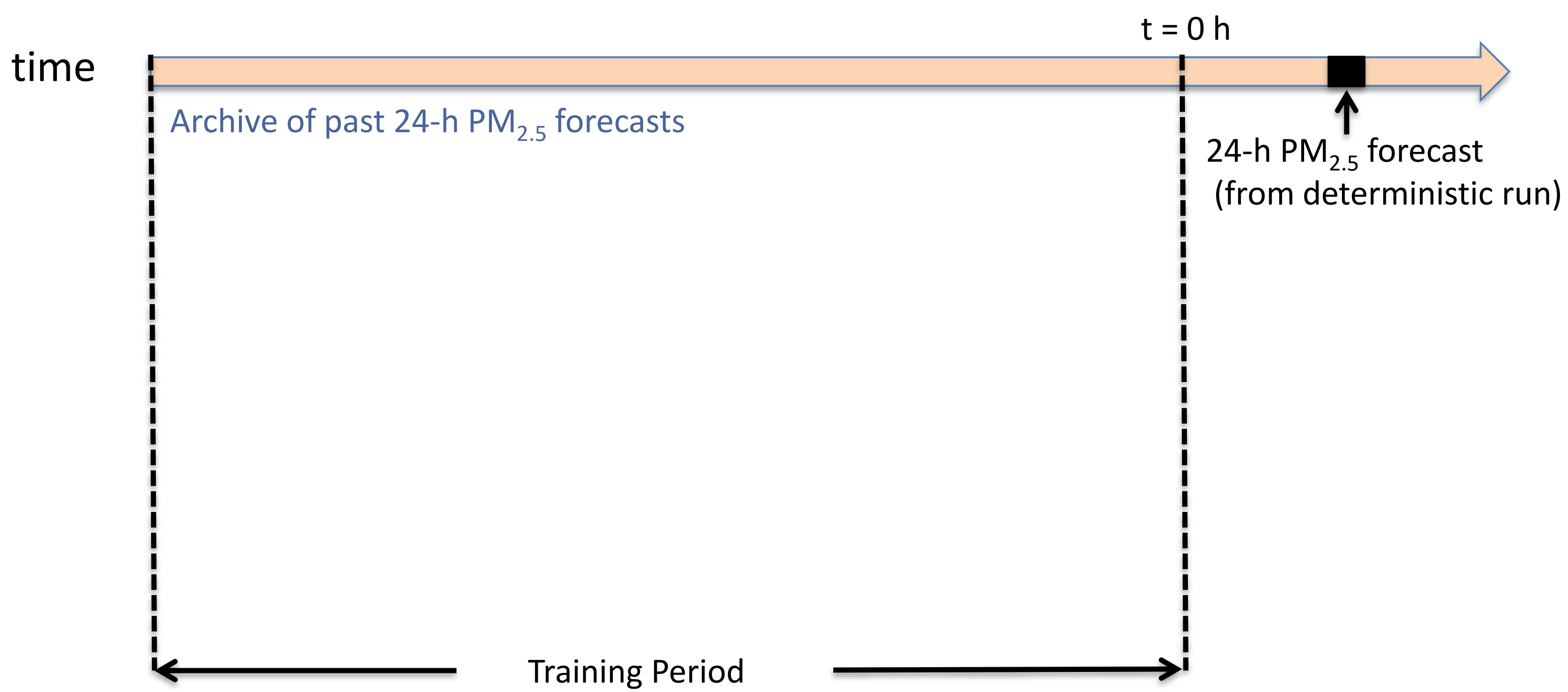


Figure adapted from Delle Monache et al. (2013)

The Analog Ensemble (2)

Find set number of analogs based on certain predictors

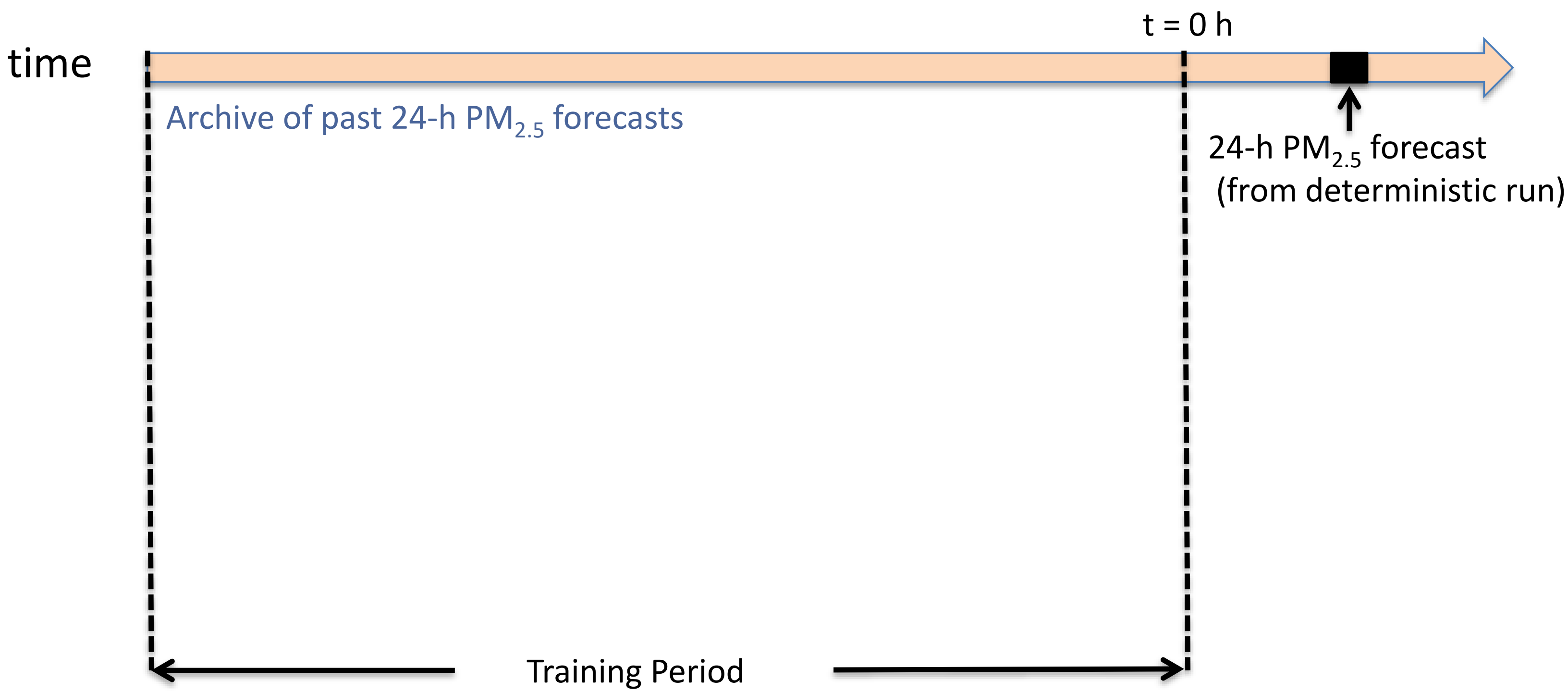


Figure adapted from Delle Monache et al. (2013)

The Analog Ensemble (3)

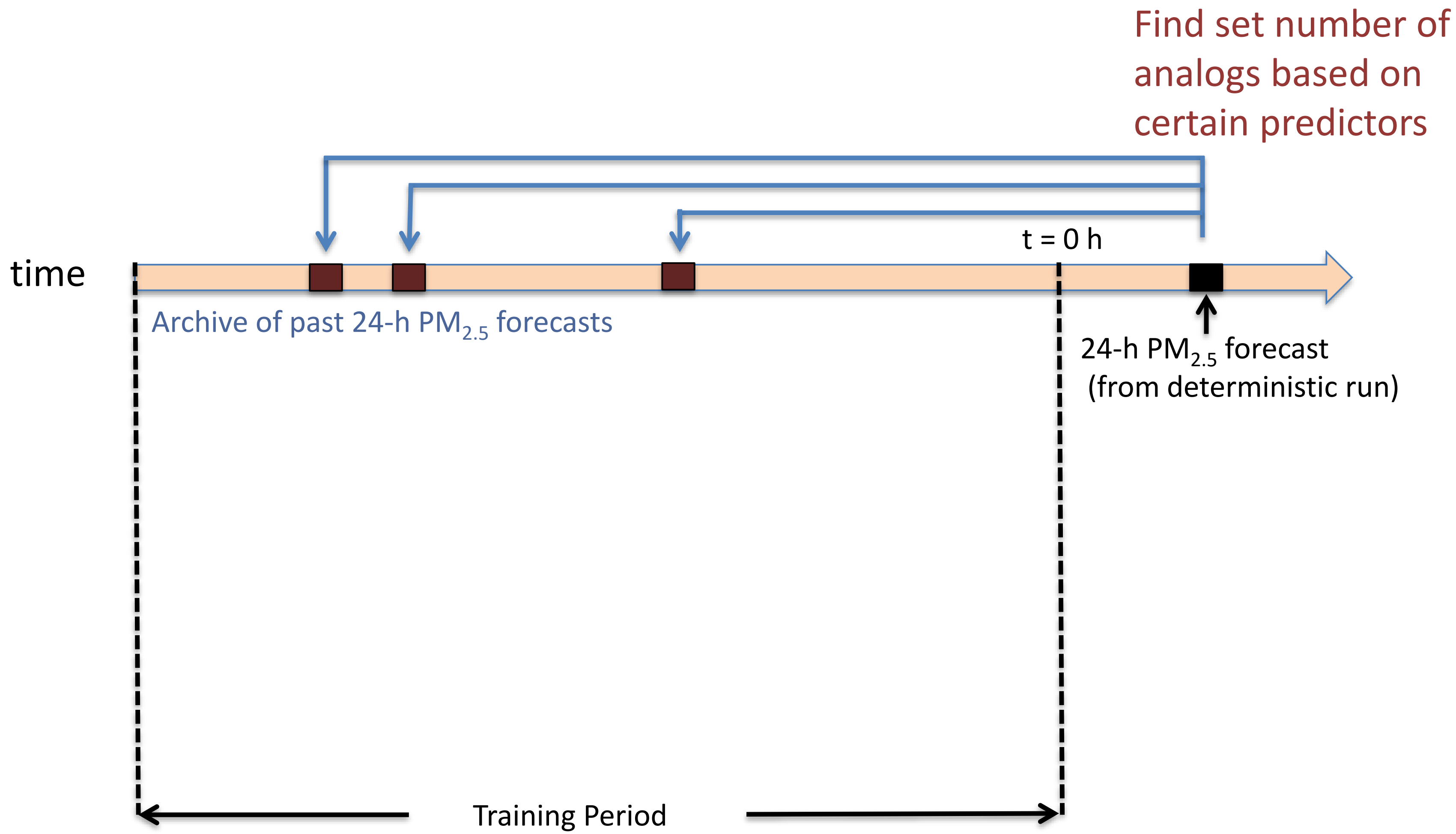


Figure adapted from Delle Monache et al. (2013)

The Analog Ensemble (4)

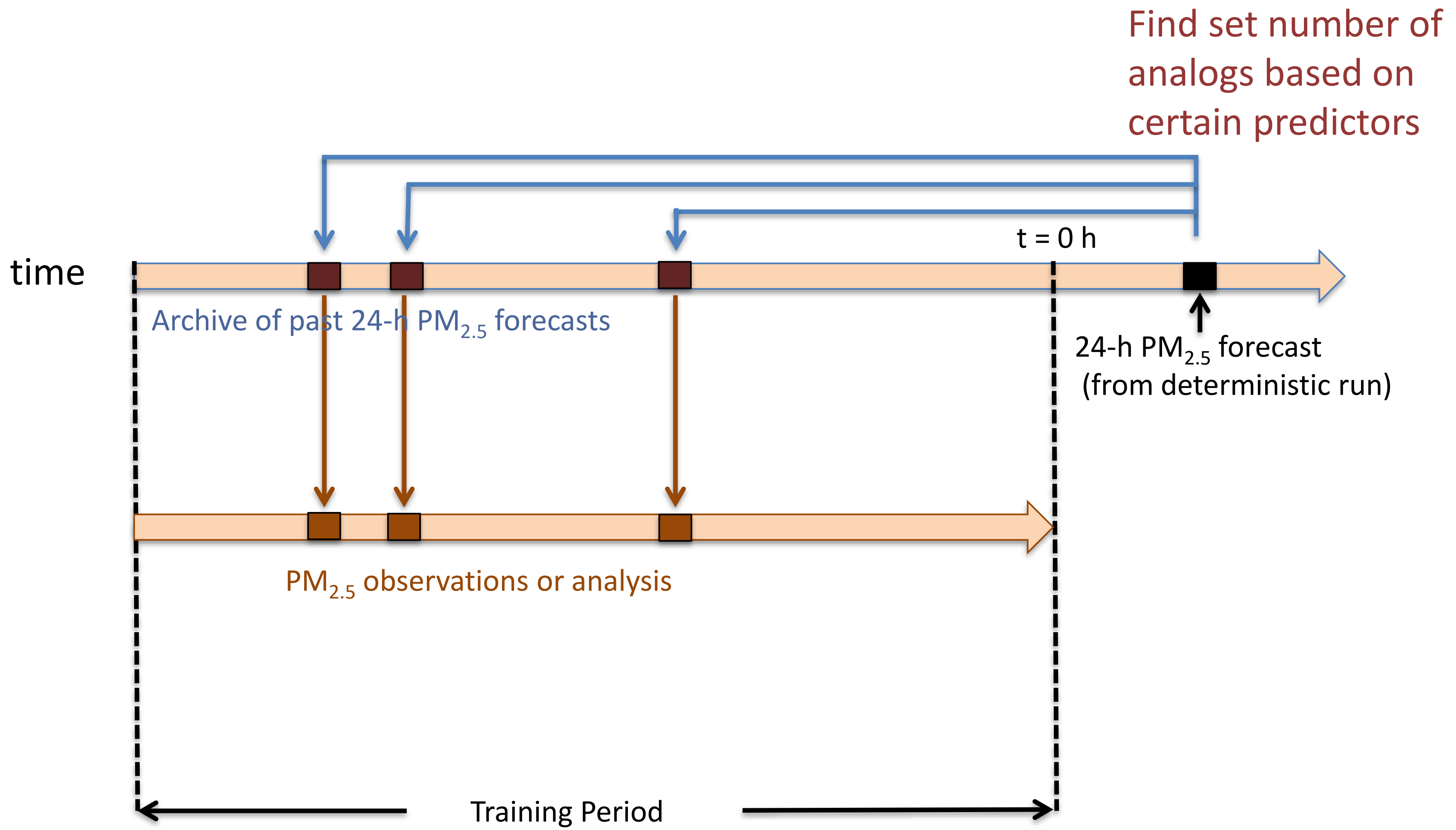


Figure adapted from Delle Monache et al. (2013)

The Analog Ensemble (5)

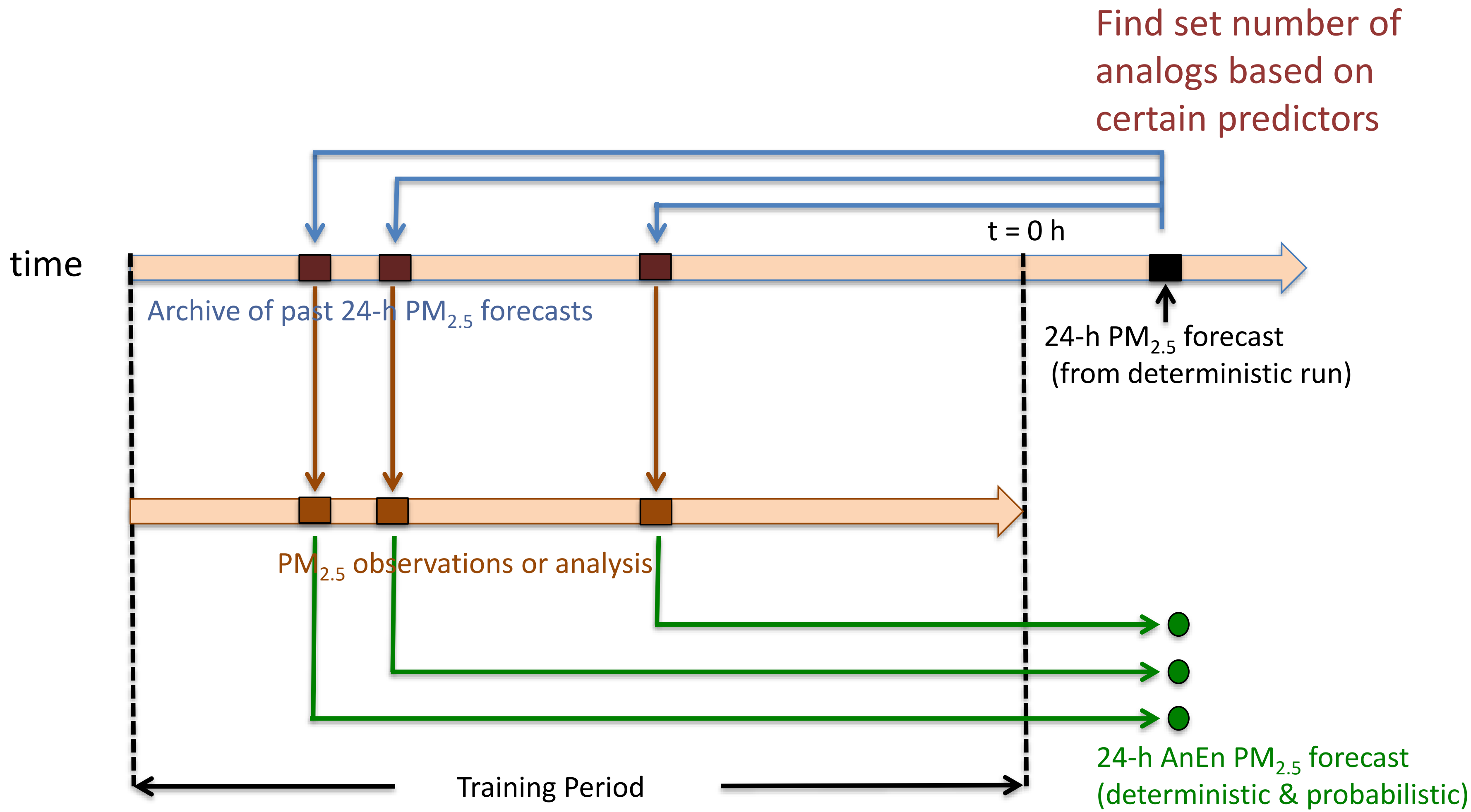
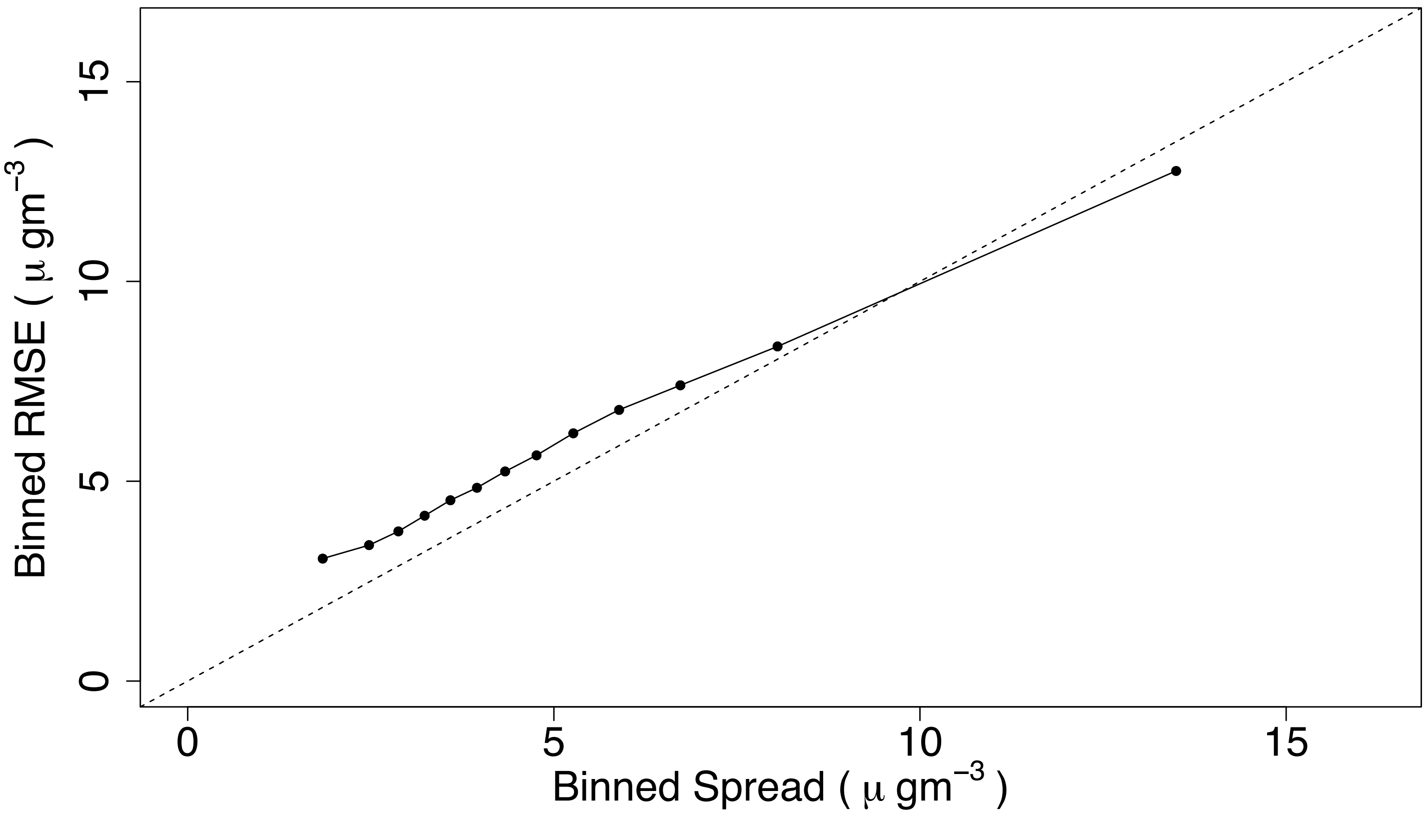


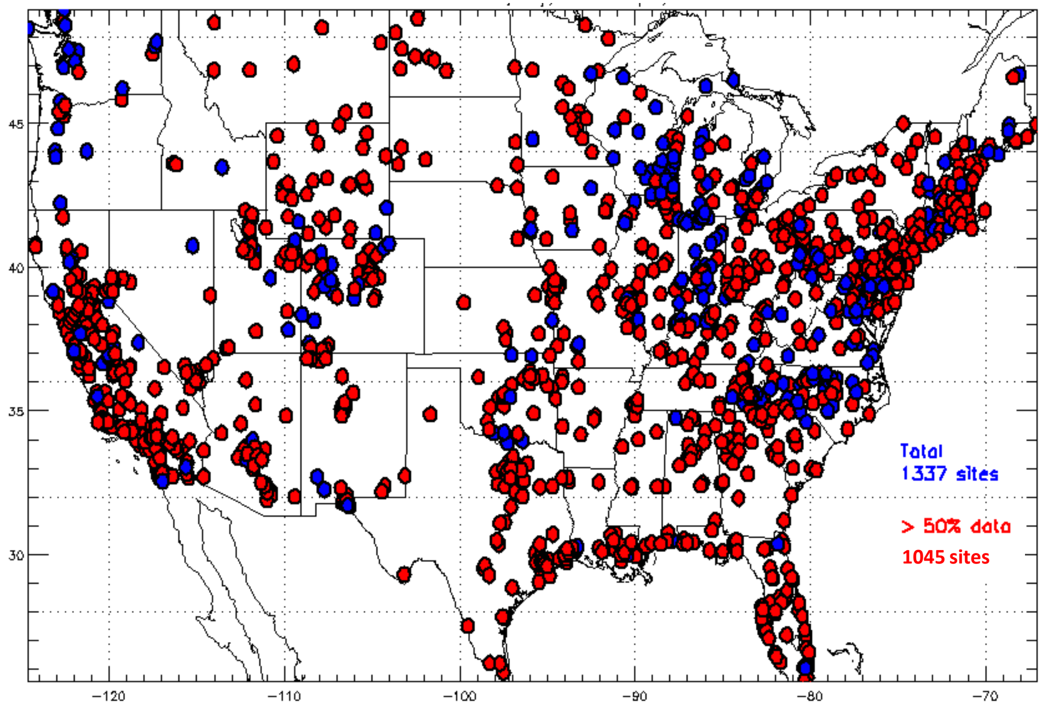
Figure adapted from Delle Monache et al. (2013)

Reliable uncertainty quantification



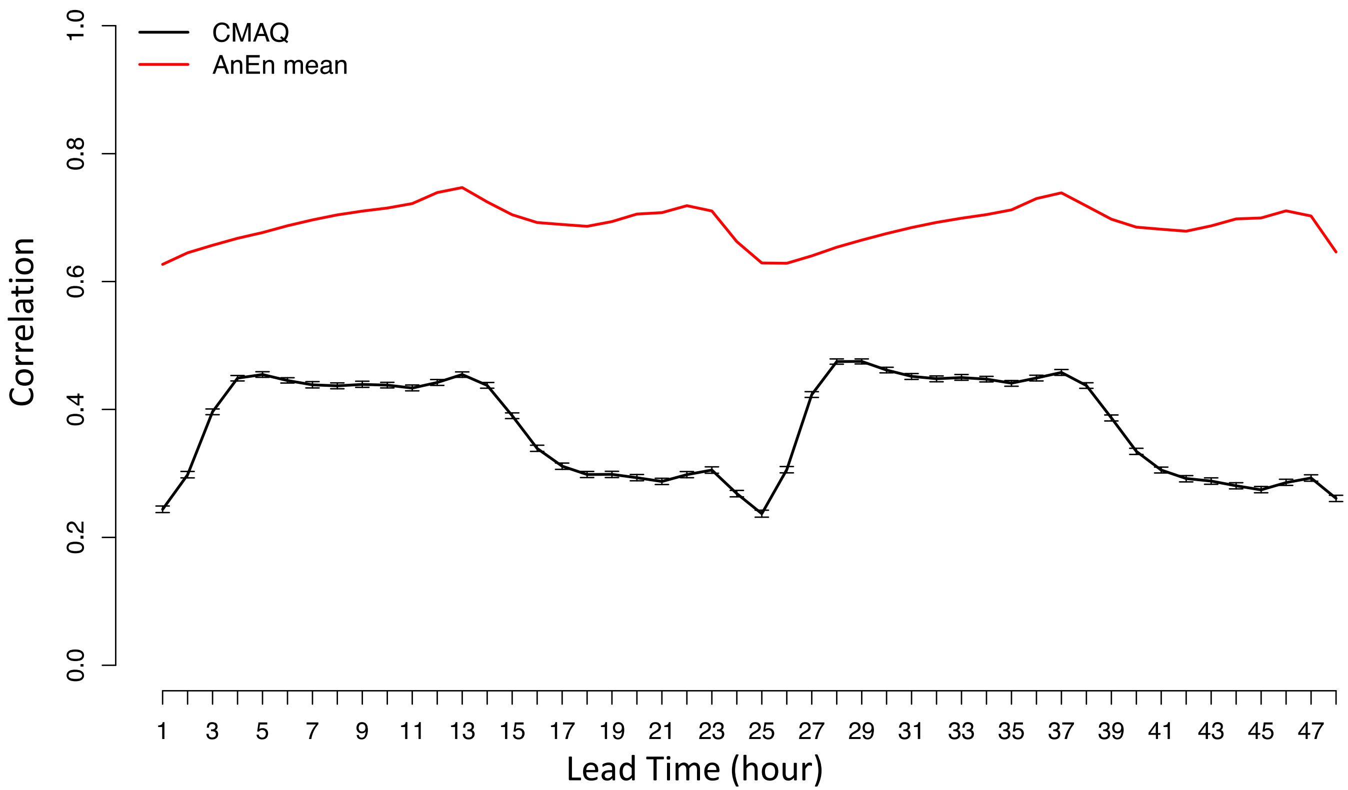
AnEn shows a very good ability to quantify the prediction uncertainty

Probabilistic predictions of O_3 with the analog ensemble



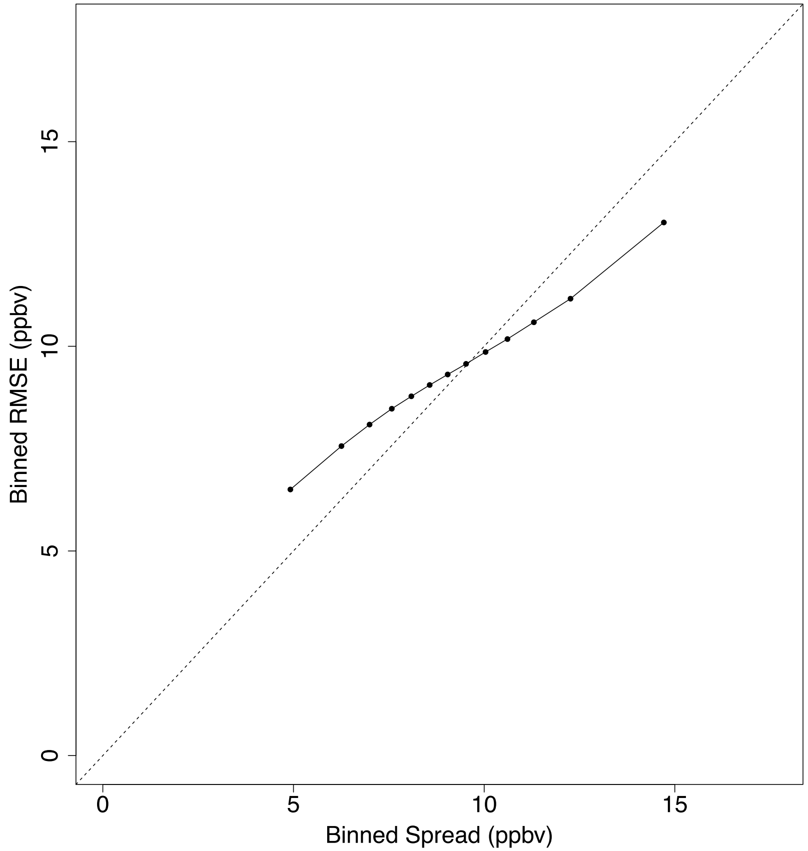
- 1337 AirNow stations with available O_3 measurements
- Hourly concentrations data for the 457-day long period (from 07-01-2014 to 09-30-2015)
- 1045 stations more than 50% of valid data

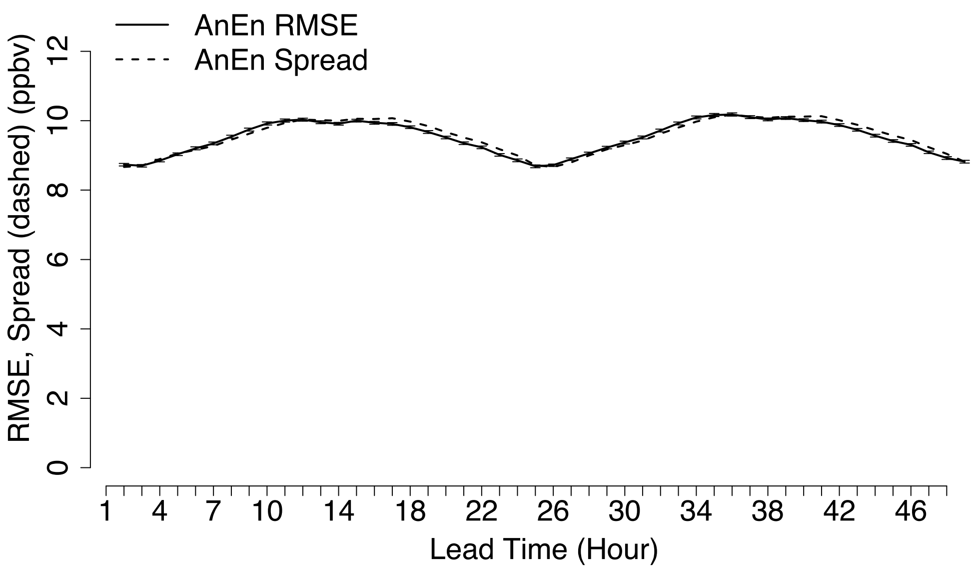
AnEn, CMAQ verification



AnEn mean improves CMAQ's correlation by ~50%

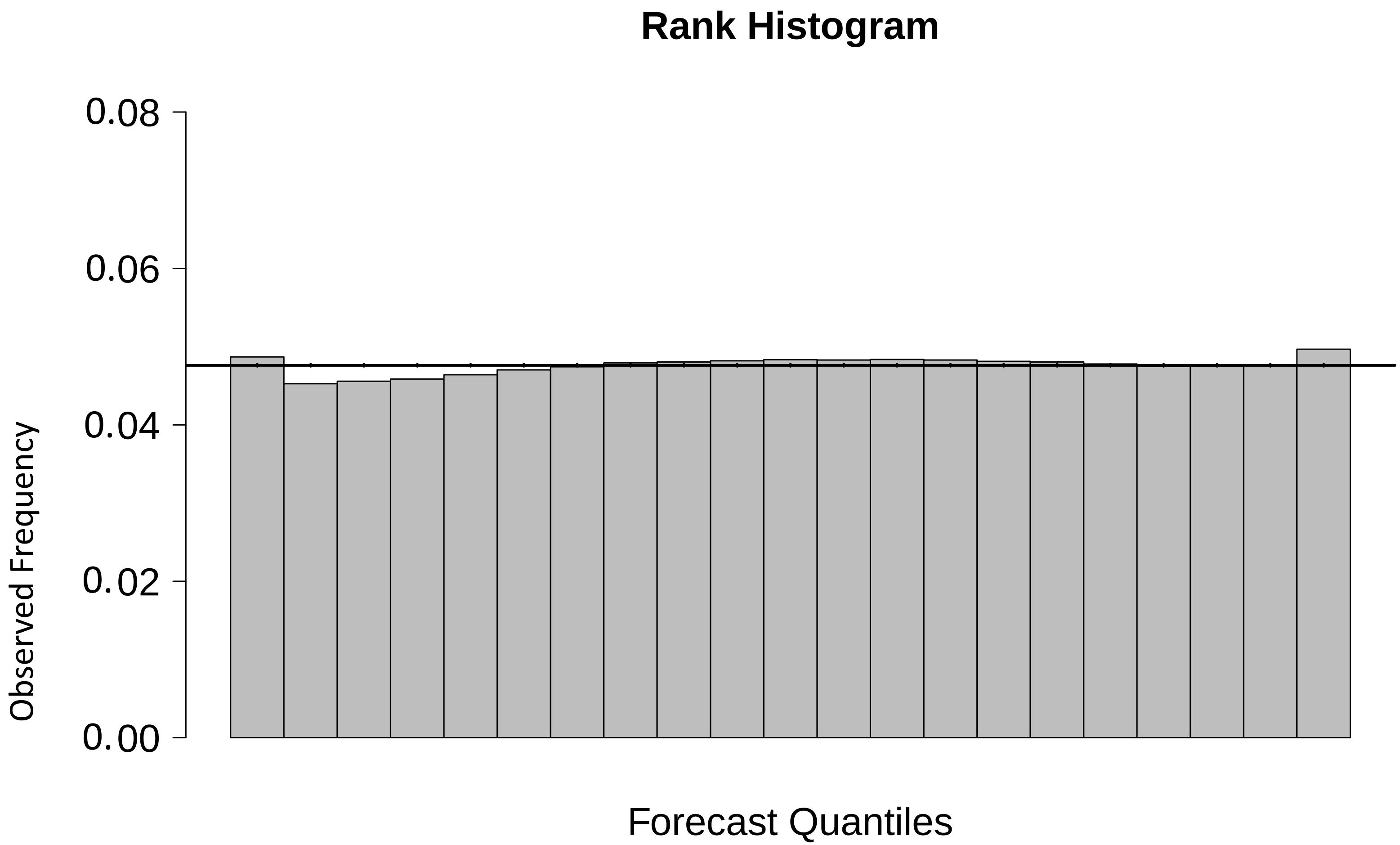
AnEn, verification (statistical consistency)





AnEn shows an excellent spread/RMSE correlation

AnEn, verification (statistical consistency)



AnEn shows a rank histogram close to flat

Probabilistic predictions of $PM_{2.5}$ with the analog ensemble

- 564 AirNow stations with available $PM_{2.5}$ measurements
- Hourly concentrations data for the 396-day long period (from 07-01-2014 to 07-31-2015)
- Average data availability: 86 %