Methods for Reducing Biases and Errors in Regional Photochemical Model Outputs for Use in Emission Reduction and Exposure Assessments

P. Steven Porter[1], S. Trivikrama Rao[2], Christian Hogrefe[3], Edith Gego[1], Rohit Mathur[3]

[1]Porter-Gego, Idaho Falls, ID 83401
[2]North Carolina State University, Raleigh, NC 27695
[3]US EPA National Exposure Research Laboratory, Research Triangle Park, NC 27711

Abstract

In the United States, regional-scale photochemical models are being used to design emission control strategies needed to meet the relevant National Ambient Air Quality Standards (NAAQS) within the framework of the attainment demonstration process. Previous studies have shown that the current generation of regional photochemical models can have large biases and errors in simulating absolute levels of pollutant concentrations. Studies have also revealed that regional air quality models were not always accurately reproducing even the relative changes in ozone air quality stemming from changes in emissions. This paper introduces four approaches to adjust for model bias and errors in order to provide greater confidence for their use in estimating future concentrations as well as using modeled pollutant concentrations in exposure assessments. The four methods considered here are a mean and variance (MV) adjustment, temporal component decomposition (TC) adjustment of modeled concentrations, and two variants of cumulative distribution function (CDF) mapping. These methods were compared against each other as well as against unadjusted model concentrations and a version of the relative response approach based on unadjusted model predictions. The analysis uses ozone concentrations simulated by the

Community Multiscale Air Quality (CMAQ) model for the northeastern United States domain for the years 1996 to 2005. Ensuring that base case conditions are adequately represented through the combined use of observations and model simulations is shown to result in improved estimates of future air quality under changing emissions and meteorological conditions.

Corresponding Author: Dr. S.T. Rao (strao@ncsu.edu)

Key Words: Air quality Modeling, Model Evaluation, Bias Adjustment, Emission Control Strategy Assessment, Attainment Demonstration, Exposure Assessment

1. Introduction

Regional-scale photochemical models are useful tools for forecasting (Eder et al., 2010), regulatory decision-making (USEPA, 2014), and exposure assessments (Garcia et al., 2010; Lin et al., 2012). Forecasts of future air quality enable people to take appropriate precautionary measures to reduce their exposure to high pollution levels (Eder et al., 2010). With respect to exposure assessments, it is desirable to relate air quality to human health at unmonitored locations (Garcia et al., 2010; Lin et al., 2012) and to deal with missing data at monitoring sites (i.e., Junninen et al., 2004). Spatially and temporally dense model outputs, adjusted for biases in mean and variability are used for these purposes. Attainment demonstrations use a set of model experiments to assess the magnitude and extent of air pollution and to determine emission reductions needed to meet the National Ambient Air Quality Standards (NAAQS). In particular, models are used to compare past air quality (base case) with model-predicted future conditions. This paper addresses the use of models in attainment demonstrations and exposure assessments.

Before a regional photochemical model is applied in the regulatory setting, USEPA's guidance for the attainment demonstration recommends a thorough evaluation of the model performance for the base case simulation (USEPA, 2014). Sistla et al. (2004), Jones et al. (2005), Pegues et al. (2012), Hogrefe et al. (2012), Kulkarni et al. (2014), and Cohan and Chen (2014) have examined some issues with the use of projected design values in attainment demonstrations. Dennis et al. (2010) provided a framework for performing comprehensive model evaluation, which entails conducting operational (Appel et al., 2007), dynamic (Napelenok et al., 2011), diagnostic (Godowitch et al., 2011), and probabilistic evaluations (Hogrefe and Rao, 2001; Foley et al., 2012; Reich et al., 2013). Also, scientists from North America and Europe have been helping advance the model evaluation framework outlined by Dennis et al. (2010) as part of the Air Quality Modelling Evaluation International Initiative (AQMEII) project (Rao et al., 2011; Galmarini et al., 2012). The AQMEII community is currently focusing on evaluating the performance of coupled meteorology and atmospheric chemistry models (Baklanov et al., 2014) to examine the strengths and limitations of air quality models being used in North America and Europe. In addition, techniques such as Kalman filtering (Delle Monache, 2006; Kang et al., 2008 and 2010) and spectral decomposition approach (Galmarini et al., 2013) have been used to correct for model bias in air quality forecasting.

It is important to promote model evaluations to establish credibility for air quality models so they can be more confidently used for regulatory decisions. However, despite continual improvement in models, discrepancies between model predictions and observations persist, stemming from both reducible and irreducible errors. In addition, whereas modeled concentrations represent volume-average concentrations, observations reflect point measurements at a given location. Also, the stochastic variations affecting the monitored concentrations are not explicitly modeled in the current regional numerical air quality models. Reducible errors (structural and parametric) are attributable to our inadequate understanding of the relevant atmospheric processes, and errors in model input variables (e.g., emissions, meteorology, boundary conditions, physics and chemistry). Irreducible errors arise from our inability to properly characterize the initial state as well as the stochastic nature of the atmosphere. A critical point to bear in mind is that it is difficult, if not impossible, to specify the atmospheric conditions (i.e., emissions, initial and boundary conditions, meteorology, physics, and chemistry) that real-world observations are seeing in all locations at all times by simulating the base case conditions using air quality models. In consequence, modeled pollutant concentrations often are biased (average observed values are not reproduced) (Simon et al., 2012), exhibit less variability than the corresponding observations (i.e., the distribution of modeled values is often more narrow than that of the observations), and show changes in pollutant levels from the base case that differ from observed changes (Gilliland et al., 2008; Godowitch et al., 2010). One way to address these limitations is to constrain the model output by the observations to ensure that observations and modeled values are starting from the same point so the deviations from the base case can be properly evaluated as emissions and/or meteorology are altered. It should be emphasized that this does not imply that scientific advances for better simulating the interactions of pollutants and transport and fate in the atmosphere are not needed.

In an attempt to better meet the needs of regulatory agencies as well as the health sciences community for regional air quality models, we propose four new methods for bringing the statistical properties (i.e., mean, variance, percentiles) of model predictions into closer harmony with pollutant concentrations observed at monitoring sites. Included are adjustments to the concentration time series (matching the mean and variance of model ozone time series to observations), spectral decomposition of time series (matching the low and high frequency variations in model ozone time series to observations), and two variations of cumulative distribution function (CDF) matching (matching sample CDFs of modeled and observed concentrations, disregarding the time sequence). A ten-year long photochemical model simulation for the northeastern USA is used to assess the performance of these methods. The ability of these methods to reproduce different-year ozone concentrations as well as contemporaneous predictions, that is, same-year ozone concentration time series, is evaluated. The new approaches are compared with absolute model projections and with model-based, relative response factors based on unadjusted model predictions.

2. Data and Methods

A. Model Setup

The following is a brief summary of the model set-up used to perform the simulations analyzed in this study. The reader is referred to Hogrefe et al. (2009, 2010) for additional details. The Mesoscale Meteorological Model MM5 (Grell et al., 1994) was used to simulate meteorological conditions for the time period from 1 January 1988 to 31 December 2005. In the current study, we utilize model simulations for the ten year period from 1996 to 2005. The meteorological simulations were performed on two-way nested grids with 36 km and 12 km grid cell sizes covering the northeastern US. Throughout the model simulation, MM5 was nudged towards reanalysis fields from the National Center for Environmental Prediction (NCEP) using four-dimensional data assimilation. All emission processing, including mobile and biogenic sources, was performed within the Sparse Matrix Operator Kernel Emissions (SMOKE) system (Houyoux et al., 2000). Anthropogenic emission inventories for 1988–2005 were compiled from a variety of sources as described in Hogrefe et al. (2009). Various regulatory control programs were implemented for the utility sector (e.g., the acid rain control program, NOx SIP Call) during the period 1995 to 2005. Also, fleet turnover has contributed to large changes in mobile source emissions and control programs were also implemented for a number of other emission sources. The combined effect of all emission control programs was a decrease of domain-wide anthropogenic NO_x and VOC emission by roughly 24% and 28%, respectively during the 1996-2005 period. Biogenic emissions were estimated with the BEIS3.12 model (Pierce et al., 1998) taking into account MM5 temperature, radiation, and precipitation.

Regional air quality simulations were performed with the Community Multiscale Air Quality (CMAQ) model (Byun and Schere, 2006), version 4.6, rather than CMAQ 4.5.1 that was used in Hogrefe et al. (2009) with the same set of meteorological and emission inputs. Air quality model simulations were performed with two one-way nested grids of 36 km and 12 km, corresponding to the MM5 grids except for a ring of buffer cells. The height of the first model layer was set at 38 m. Gas phase chemistry was represented by the CB-IV mechanism (Gery, 1989) while aerosol chemistry was simulated with the "aero3" module. In this study, only results from the 12 km CMAQ simulations were utilized. Chemical boundary conditions for the 36 km grid were extracted from archived monthly-mean fields of global chemistry simulations performed for the 1988–2005 time period with the ECHAM5-MOZART modeling system as part of the RETRO project (RETRO, 2007). The MM5/CMAQ simulations for this paper have been evaluated against observations in Hogrefe et al. (2009, 2010) and Civerolo et al. (2010).

B. Observations

Hourly ozone observations during 1996-2005 were extracted from the USEPA's Air Quality System (AQS) (**http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm).**

From these hourly data, daily maximum 8-hr average ozone concentration time series for each ozone season (May 1 – September 30) and year were developed and screened to ensure that all stations used in this analysis had at least 80% data completeness for any given ozone season and year.

C. Adjustment Methods

(1) RATIO

The USEPA-recommended use of models in an attainment demonstration *(*USEPA, 2014*)* is based on an assumption that the ratio of future case model simulations to base case model simulations averaged over high simulated ozone days is a good predictor of changes in observed ozone design values. Model predictions for the future year are usually a combination of past meteorology and future emission scenarios needed to reduce ozone concentrations in non-attainment areas.

$$DVF = DVC \bullet RRF \quad \textbf{(1)}$$

DVF = <u>D</u>esign <u>V</u>alue for the <u>F</u>uture Year

DVC = <u>D</u>esign <u>V</u>alue for the <u>C</u>urrent Year

RRF = <u>R</u>elative <u>R</u>esponse <u>F</u>actor

As detailed in USEPA (2007), the calculation of the DVC is based on multiple years of observed data while the RRF is calculated from averaged model simulations for both emission scenarios under the same meteorological conditions following a specified set of criteria. *This calculation is based on a combination of past meteorology and future emissions, which cannot be verified because future year meteorology cannot be known and there are no comparable future observations available for model adjustments.* In this study, we examine a simple ratio method (hereafter referred to as the RATIO method) based on a single year of observations and non-averaged model predictions for the base case and future case:

$$\boldsymbol{future\ prediction = observed\ base \times \frac{CMAQ\ future\ year}{CMAQ\ base\ year}} \quad \textbf{(2)}$$

For example, when applying the RATIO method, defined in Eq (2), to project the future year 4^{th} highest daily maximum 8-hr ozone (4HDM8O3) concentration, one would multiply the observed base year 4HDM8O3 concentration with the ratio of CMAQ predicted 4HDM8O3 concentrations for the future and base year. Like the RRF approach described in USEPA (2007), the RATIO approach assumes that model predictions can be used in a relative sense; that is, even though the absolute levels of modeled concentrations may be biased, they can be used to predict the change in pollution levels from a given scenario. In contrast to the approach outlined in USEPA (2007), the CMAQ base year and future year simulations analyzed in this study reflect changes in both emissions and meteorology while in typical attainment demonstrations, the base-year meteorological conditions are used for both the base year and future year when calculating RRFs because the future year meteorology cannot be known. Also, given the uncertainties associated with estimating future year modeled concentrations, the EPA approach is intended to estimate multi-year average design values. The EPA approach does not attempt to precisely predict 4HDM8O3 concentrations. Because meteorology is not held constant in the RATIO approach, it can be verified via a simulation of ozone for the northeastern U.S. for the years 1996 to 2005.

Figure 1 presents an illustrative test of the RATIO approach for an arbitrary base year / future year combination selected from the observational and model database described in the previous section. This figure depicts the observed and CMAQ ratios of future year (2004) to base year (2001) 4HDM8O3 concentrations at all (80% complete) monitoring sites in the modeling domain along with the 1-to-1 line. CMAQ typically underestimates observed changes (Gilliland et al 2008; Godowitch, et al., 2010; Napelenok et al. 2011). The large scatter evident in this figure suggests that it is worth investigating potential alternatives to the RATIO approach that may provide better performance in predicting 4HDM8O3 concentrations into the future.

(2) Mean and Variance matching (MV)

For each ozone season, we adjust the mean and variance of the modeled 8-hour ozone time series (i.e., the daily maximum 8-hr average ozone time series) so that they match those observed:

$$\boldsymbol{future\ prediction} = \quad \textbf{mean}\ \{\boldsymbol{CMAQ\ future}\}\ + \qquad \textbf{(3)}$$

$$\{\boldsymbol{CMAQ\ future - mean\ CMAQ\ future}\} \times \frac{\boldsymbol{\sigma}(\boldsymbol{observed\ base\ year})}{\boldsymbol{\sigma}(\boldsymbol{CMAQ\ base\ year})}$$

$$+\ \ \boldsymbol{mean}\ \{\boldsymbol{observed\ base\ year\ -\ CMAQ\ base\ year}\}$$

Base year adjustment factors (ratio of standard deviations and the mean bias) computed for each ozone season (153 days) are applied to other ozone seasons (**prediction** years). The **prediction** year 4HDM8O3 is then compared with that observed. The rationale for this method is that the distribution of CMAQ ozone is typically narrower than that observed (see Figure 7 in section 3) because modeled values reflect grid volume averages while observations represent discrete point measurements.

(3) Mean and Variance matching with temporal components (TC)

In this approach, mean and variance matching are applied to temporal components of modeled and observed values. Temporal components in this case are high- and low-frequency variation (denoted ‘HF and ‘LF’, respectively). The LF component is the output of a low-pass filter (i.e., the KZ filter with a window size of five (5) days and five (5) iterations, and a 50% cutoff period of about 24 days, Rao and Zurbenko, 1994), while the HF component is the difference between observations and the LF component (and therefore has a mean of zero). The period of separation between the HF and LF components is roughly 24 days.

The adjusted CMAQ values are given by:

$$\boldsymbol{observed\ (base\ year) = observed\ (LF, base\ year) + observed(HF, base\ year)} \qquad \textbf{(4)}$$

$$\boldsymbol{CMAQ = CMAQ(LF) + CMAQ(HF)} \qquad \textbf{(5)}$$

The principle behind this approach is that low- and high-frequency processes are driven by different phenomenon, with high-frequency variation attributable primarily to the synoptic-scale weather (i.e., variations due to fast changing emissions, diurnal forcing, and weather-induced variations embedded in ozone time series data), and low-frequency variation driven by seasonal emissions and trends (Rao et al., 1997). It follows that components with different driving forces should have different adjustments. Therefore, we adjust the mean and variance of each component of the modeled daily maximum 8-hr ozone time series so that they match those observed:

$$\boldsymbol{future\ (LF) =} \quad \textbf{mean}\ \{\boldsymbol{CMAQ\ future\ LF}\} + \qquad \textbf{(6)}$$

$$\{\boldsymbol{CMAQ\ future\ LF - mean\ CMAQ\ future\ LF}\} \times \frac{\boldsymbol{\sigma(observed\ base\ year\ LF)}}{\boldsymbol{\sigma(CMAQ\ base\ yearLF)}}$$

$$+ \ \boldsymbol{mean}\ \{\boldsymbol{observed\ base\ year\ LF -\ CMAQ\ base\ year\ LF}\}$$

$$\boldsymbol{future\ (HF) = CMAQ\ future\ HF \times \frac{\sigma(observed\ base\ year\ HF)}{\sigma(CMAQ\ base\ yearHF)}} \quad \textbf{(7)}$$

$$\boldsymbol{future = future\ (LF) + future\ (HF)} \quad \textbf{(8)}$$

As with the **MV** method, the **TC** method adjusts the entire time series (not just the 4th highest) and assumes base year *variance ratios* and m*ean bias* will be the same in future years for both high- and low- frequency components.

(4) Cumulative Distribution Function matching (CDF1)

Cumulative distribution function matching (CDF matching) is a bias correction technique used in climate data analysis and image processing (see Wang and Chen, 2014 for example). Observed and modeled concentrations are rank-ordered to establish sample CDF's (sample quantiles). The observed quantiles are regressed against those of CMAQ yielding the following:

observed base year quantiles = Ko + K1 ● CMAQ base year quantiles **(9)**

future quantiles = Ko + K1 ● CMAQ future quantiles

The slope and intercept, Ko and K1, rotate and displace the CMAQ CDF such that the root mean squared distance between the modeled and observed quantiles are minimized. The original time order of CMAQ and observations is lost with this approach. As with methods 1 and 2, the adjustment parameters estimated using base year observed and CMAQ information (in this case, Ko and K_1) are applied to future years.

(5) Cumulative Distribution Function matching (CDF2)

In contrast with **CDF1**, the parameters Ko and K1 in equation (11) are applied only to the base year and the future year 4HDM8O3 concentration is estimated by adding the model predicted change in 4HDM8O3 concentrations to the base year adjusted 4HDM8O3 concentration. As with **CDF1**, the time order is lost.

$$\textit{Observed base year quantiles} = Ko + K1 \bullet \textit{CMAQ base year quantiles} \qquad \textbf{(10)}$$

$$\textit{future quantiles} = \{\textit{CMAQ future quantiles} - \textit{CMAQ base year quantiles}\}$$

$$+\{Ko + K1 \bullet \textit{CMAQ base year quantiles}\}$$

3. Results and Discussion

A. Predicting the 4th Highest Daily Maximum 8-hr Ozone Concentration

For every pair of years from 1996-2005, one year served as a base year and the other a prediction year (a total of 90 pairs where the two years are different). For each *base* year-*prediction* year pair, there are approximately 250 sites from which performance metrics were computed, including mean bias (MB), fractional mean bias (FMB), mean absolute bias (MAB), fractional MAB (FMAB), root mean squared error (RMSE), fractional RMSE (FRMSE), square of correlation (R^2) and index of agreement (IA). As noted above, the quantity being evaluated in this analysis is the 4HDM8O3 concentration.

Domain-wide RMSE, MAB, and R values, using 1996 as the base year for predicting the 4HDM8O3 for 1997 to 2005, are presented in Figure 2. There is a considerable amount of year-to-year variation in the RMSE values, ranging from about 6 to 15 ppb. To provide a more comprehensive evaluation of the different approaches, Table 1 provides the performance metrics and their 95% bootstrap confidence intervals for predicting 4HDM8O3 concentrations for all base year-prediction year pairs. From Figure 2, it is evident that no one method is best suited for all the prediction years, but CDF2 often is the best among the six methods examined here for predicting the 4HDM8O3 concentration. Table 1, with respect to the MAB metric on the basis of overlapping confidence intervals, shows that, the MAB follows the order (TC, CDF1, CDF2) < MV < RATIO < CMAQ, whilst for RMSE performance ranks are: (CDF1 , CDF2) , MV , (TC , RATIO , CMAQ). For the other metrics (R2 and IA) none of the methods stand out because all the confidence intervals overlap. To help sort out these relationships, a multiple comparison for all metrics MAB and RMSE (Table 2) also supports the conclusion that CDF1 and CDF2 perform better than RATIO and CMAQ methods.

The fact that the TC method has a higher RMSE, but not MAB, than the other methods both for the 1996 base year (Figure 2) as well as all base year / future year pairs (Table 1) implies the presence of outliers in the projected values that influence the RMSE metric to a greater degree than the MAB metric.

Spatial images of the RMSE values for all base year / prediction year pairs at a given site, displayed in Figure 3, indicate with the large number of red and orange dots that all methods exhibit poorer performance at sites along the northeastern urban corridor (along interstate 95), the northeastern seaboard, and near large water bodies. However, at many locations the four adjustment methods introduced in this study tend to perform better in predicting 4HDM8O3 concentrations than the raw CMAQ output and RATIO approach. Table 3 shows the percentage of pairwise counts for a given metric of 'best' results for all 90 base year / prediction year pairs. Reading across row 1, for example, CMAQ has a lower MAB than RATIO and MV for 36% and 27% of the 90 pairs, respectively. For every metric, CDF2 and CDF1 are better than all other methods for more than 50% of the pairs. CMAQ and RATIO are not better than any of the other methods for more than 50% of the pairs. Thus, judging from the values of MAB, RMSE and R^2, CDF2 appears to be the best method among the four bias and error adjustments methods considered here. However, it is evident from Table 1 that the differences in performance in predicting the 4HDM8O3 values though often statistically significant are often small. For example, RATIO has an RMSE of 9.5 ppb compared with CDF2 of 8.7 ppb.

As noted above, the EPA guidance recommends that a number of the highest ozone days be averaged. Table 4 compares the adjustment methods when used with the highest ozone day up to the average highest 10 days and with the 4th highest. *Averaging brings benefits to all the methods discussed here, but the relative performance of each method is little changed.*

Next, we computed the frequency with which the predicted 4HDM8O3 concentration falls within a specified range of the observed 4HDM8O3 concentrations across all sites and base-year /

prediction-year pairs. This metric can be thought of as the fraction of correct predictions. Results depicted in Figure 4 show that all methods involving bias/error adjustments improve the predictions of future 4HDM8O3 concentrations compared to the raw CMAQ output. Only 25% of predictions fall within a target of 3% of observed 4HDM8O3's. While the CMAQ 4HDM8O3 concentrations are within 10% of the corresponding observations at 60% of the sites, CDF2 is within 10% of the corresponding observations at 70% of the sites. As noted before, not all monitoring sites in the model domain can be considered to be regionally-representative sites since several sites are near urban areas or water bodies where regional scale air quality models such as the 12 km CMAQ simulations used in this study cannot be expected to perform well (Hogrefe et al., 2014).

RMSE values for the daily maximum 8-hr ozone concentrations predicted by the four methods for the upper half of the distribution for the year 2004 using 2001 as the base year, presented in Figure 5, reveal that all methods perform poorly in reproducing the observed concentrations for the extreme values (upper percentiles >90%). Also, there is a large disparity between the changes in absolute levels of CMAQ and observed ozone between 2001 and 2004 for the upper half of the concentration frequency distribution (Fig.6a). The ozone improvement (the difference in ozone concentrations between 2001 and 2004) simulated by CMAQ is much lower than that seen in observations, especially in the upper percentiles (>75%). The difference is highlighted in Fig. 6b which is a histogram of the change in ozone between 2001 and 2004 at the 98th percentile for all monitors in the domain. Napelenok et al (2011) and Kang (2013) speculated that CMAQ underestimation of ozone change is due to uncertainties in the NO_X emission inventory, particularly mobile and area sources (because most point source emissions are measured by continuous emission monitors found at electricity generating facilities).

Both Figures 6a and 6b reveal that all four bias and error adjustment methods tend to reduce the discrepancy between the simulated and observed changes and help to bring the simulated change in ozone between 2001 and 2004 closer to what was observed. It may be, then, that for regulatory assessments (i.e., attainment demonstration), using observations to adjust base case model bias and error leads to better estimates of future year design values than does use of the original model values.

B. Same-year Performance (demonstration of model's usefulness for exposure assessment)

Spatially continuous time-series of air quality surfaces over large geographic regions are desirable for assessments of the effects of chronic human exposure to air pollutants. Photochemical models have been used for this purpose and it has been shown that their use in conjunction with observations is an improvement over the use of observations alone (Garcia et al, 2010). Put simply, model-observation bias is projected to unmonitored sites using geostatistical methods. Modeled values for the gridded model domain are then adjusted using the interpolated bias. Prediction of future (or past) times is usually not of interest (though temporal interpolation may be). In this section, we suggest that model values adjusted via the methods presented here may be an improvement over projecting model bias. A validation of the use of adjusted values for exposure assessments will be addressed in a future paper.

As noted above, model probability distributions tend to be narrower than those observed (smaller standard deviation and range). If modeled ozone values used in a health assessment have a smaller range (or standard deviation) than what would be observed, the estimated health effect will be biased. Many of the methods discussed here could be applied in same-year fashion to model and observed information to help construct ozone fields or any pollutant species with unbiased standard deviation.

As a first step in this direction, in this section, we evaluate 'same-year' performance for ozone time series from raw CMAQ output, modeled ozone that is adjusted for bias and error using the MV and TC methods, and modeled ozone that is adjusted using the CDF1 approach at the discrete locations of the monitoring stations. By same-year it is meant that there is no projection from base-year to prediction-year. Note that no results are computed in this section for the RATIO and CDF2 approaches. For the RATIO approach, the same-year ratio of model predictions shown in Eq (2) is equal to one, thus, predicted values would equal observed values; for the CDF2 approach, the predicted value equals that for the CDF1 approach since the model predicted differences between base-year and prediction-year are zero. The object of investigating same-year performance, as noted above, is to test the potential usefulness of these approaches for creating spatially continuous pollutant concentration time series that could be used for exposure assessment.

Domain-wide statistics for (MAB, standard deviation, and correlation coefficient) for same-year adjusted CMAQ values (annual 4^{th} highest 8-hour daily maximum, 4HDM8O3) predictions were computed for all years from 1996 – 2005. Results, depicted in Figures 7a-c, indicate TC and

CDF1 outperform MV and as expected the raw CMAQ. As noted above, an important criterion for fitting exposure models is that the standard deviation of the adjusted values matches those observed. As expected, standard deviations for the CDF1, MV, and TC methods closely match the observed standard deviation while raw CMAQ underestimated the observed standard deviation (Fig. 7b). That exposure assessment requires unbiased standard deviation estimates was highlighted by NRC (2004). Therefore, a conclusion that can be reached from this study is that raw CMAQ output should not be used for exposure assessments without some type of bias/error-adjustment, and that the CDF1 and TC methods produces surrogate ozone time series that, on average, are very close to the observed ozone time series. While CDF1 also exhibits good performance as measured by these summary statistics, it does not retain the original time order because it is designed to adjust the overall distribution rather than an adjusted concentration time series. Thus, its potential use in health assessment studies would be limited to analyses focused on long-term (chronic) rather than short-term (acute) exposures. The relative performances of the three methods discussed above for reducing the bias in years 2001 and 2004 reveal the superior performance of the CDF1 and TC methods (Figs. 8) across a wide range of percentiles, especially when compared to raw CMAQ output. However, it should also be noted that the performance of all methods investigated here degrades at the very top end of the concentration distribution (>90%). Nevertheless, the bias adjustment methods reduce the magnitude of the error and bias present in the raw CMAQ output.

Since chronic exposure assessments rely on time series of pollutant concentrations, the TC adjustment method appears to be the best approach; this is very useful information because CMAQ output can be more confidently used for filling-in for the days when observations are not available (see Junninen et al., 2004 for examples of data fill-in methods). For example, most $PM_{2.5}$ and speciated $PM_{2.5}$ monitoring sites operate on a 1 in 3 day schedule. For the days with missing data, the TC method can be used for correcting CMAQ output for bias and error so the adjusted model output can be used to generate continuous pollutant time series for exposure assessments by the health community. One could also adjust black carbon (BC) estimates or other species simulated by CMAQ that are of particular interest to health scientists with one of the above two methods to generate daily BC or other pollutant species time series for the whole year for use in health risk assessments. Further steps then would be to spatially-interpolate the adjusted ozone concentrations to the entire model domain using techniques such as those suggested by Garcia et al. (2010), Macmillan (2010), Fuentes and Raftery (2005), and Crooks and Isakov (2013).

4. Summary

Future year air quality predictions can be improved by adjusting for biased concentration estimates while exposure assessments may also benefit from unbiased standard deviation estimates. This paper discussed four methods for reducing the bias and errors in model predicted pollutant concentrations so that the model can be more effectively used for both of these tasks. Four methods were developed and tested using long-term modeling simulations for the northeastern US domain. Three of the methods project base-year CMAQ bias metrics to prediction-years. The projected metrics are (1) mean and variance, (2) mean and variance of temporal components, and (3) regression parameters of a QQ plot. The fourth method uses QQ plot regression parameters only for the base year. The methods were compared against unadjusted model concentrations and model concentrations adjusted with a relative response factor approach, similar in concept to what is typically used in attainment demonstrations to estimate future air quality. Results demonstrate that the adjustment of modeled concentrations so that they more closely resemble observations (i.e. the process of combining observations and model predictions) usually improves the usefulness of CMAQ for both estimating future concentrations stemming emission reduction policies and for conducting exposure assessments. All adjustment methods were demonstrated to improve future year model estimates compared to the absolute modeled projections. The following conclusions can be drawn based on the performances of the methods considered here.

For estimating future year concentrations, the uncertainty in predicted changes in ozone levels can be substantial and depends heavily on the particular locations and base- / prediction-year pairs (recall Figure 1). Uncertainty estimates for predictions are unavailable without validation studies based on multi-year modeling experiments. Methods that use the base year parameters (MV, TC, and CDF1) to adjust prediction year model values are often less effective than methods that do not (CDF2), suggesting that the characteristics of CMAQ's bias are non-stationary. For exposure assessments, MV or TC adjustment methods applied to CMAQ model output provide continuous pollutant time series having standard deviation that is close to that observed.

Acknowledgements

Although this research was funded by the U.S. Environmental Protection Agency and approved for publication, it does not necessarily reflect the Agency's views or policies. Two of the authors (PSP and EG) acknowledge the support from the Coordinating research Council under contract A-89.

References

Appel, K.W., Gilliland, A.B., Sarwar, G. and R.C. Gilliam. 2007. Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: sensitivities impacting model performance part I: ozone. Atmos. Env. 41 (40): 9603-9615.

Baklanov, A., Schlünzen, K., Suppan, P., Baldasano, J., Brunner, D., Aksoyoglu, S., Carmichael, G., Douros, J., Flemming, J., Forkel, R., Galmarini, S., Gauss, M., Grell, G., Hirtl, M., Joffre, S., Jorba, O., Kaas, E., Kaasik, M., Kallos, G., Kong, X., Korsholm, U., Kurganskiy, A., Kushta, J., Lohmann, U., Mahura, A., Manders-Groot, A., Maurizi, A., Moussiopoulos, N., Rao, S. T., Savage, N., Seigneur, C., Sokhi, R. S., Solazzo, E., Solomos, S., Sørensen, B., Tsegas, G., Vignati, E., Vogel, B., and Y. Zhang, Y. 2014. Online coupled regional meteorology chemistry models in Europe: current status and prospects, Atmos. Chem. Phys. 14: 317-398, doi:10.5194/acp-14-317-2014.

Byun, D. W. and K.L. Schere. 2006. Review of the governing equations, computational algorithms, and other components of the Models- 3 Community Multiscale Air Quality (CMAQ) modeling system, Appl. Mech. Rev. 59: 51–77.

Civerolo, K., Hogrefe, C., Zalewsky, E., Hao, W., Sistla, G., Lynn, B., Rosenzweig, C. and P.L. Kinney. 2010. Evaluation of an 18-year CMAQ simulation: Seasonal variations and long-term temporal changes in sulfate and nitrate, Atmos. Environ., http://dx.doi.org/10.1016/j.atmosenv.2010.06.056.

Cohan, D.S. and R. Chen. 2014. Modeled and observed fine particulate matter reductions from state attainment demonstrations, J. Air & Waste Manage. Assoc., 64:9, 995-1002, DOI: 10.1080/10962247.2014.905509

Crooks, J. and V. Isakov. 2013. A wavelet-based approach to blending observations with deterministic computer models to resolve the intraurban air pollution field, J. Air & Waste Manage. Assoc. 1369-1385.

Delle Monache, L., Nipen, T, Deng, X., Zhou, Y. and R. Stull. 2006. Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction, J. Geophys. Res., 111, D05308, doi:10.1029/2005JD006311.

Dennis, R., Fox, T., Fuentes, M., Gilliland, A., Hanna, S., Hogrefe, C., Irwin, J., Rao, S.T., Scheffe, R., Schere, K., Steyn, D., and A. Venkatram. 2010. A framework for evaluating regional-scale numerical photochemical modeling systems. Environ. Fluid Mech. 10 (4): 471-489.

Eder., B., Kang., D., Rao, S.T., Mathur, R., Yu, S., Otte, T, Schere, K., Wayland, R, Jackson, S., Davidson, P., McQueen, J. and G. Bridgers. 2010. Using national air quality forecast guidance to develop local air quality index forecasts, Bull. Amer. Meteor. Soc. 313-326.

Foley, K., Reich, B.J., and S. Napelenok. 2012. Bayesian Analysis of a Reduced-Form Air Quality Model, Environ. Sci. Tech. 46(14): 7604-7611.

Fuentes, M. and A.E. Raftery, 2005. Model Evaluation and Spatial Interpolation by Bayesian Combination of Observations with Outputs from Numerical Models, *Biometrics*. 61: 36-45.

Galmarini, S., Rao, S.T. and D.G. Steyn. 2012. Preface, Atm. Env., Volume 53, June, pp 1-3, http://dx.doi.org/10.1016/j.atmosenv.2012.03.001.

Galmarini, S., Kioutsioukis, I. and E. Solazzo. 2013. *E pluribus unum*: ensemble air quality predictions, Atmos. Chem. Phys. 13: 7153–7182.

Garcia, V.C., Foley, K.M., Gego, E., Holland, D. and S.T. Rao. 2010. A Comparison of Statistical Techniques for Combining Modeled and Observed Concentrations to Create High-Resolution Ozone Air Quality Surfaces, J. Air & Waste Manage. Assoc. 586-595.

Gery, M.W., Whitten, G. Z., Killus, J. P., and M.C. and M.C. Dodge. 1989. A photochemical kinetics mechanism for urban and regional scale computer modeling, J. Geophys. Res. 94: 12925–12956.

Gilliland, A.B., Hogrefe, C., Pinder, R.W., Godowitch, J.M., Foley, K.M. and S.T. Rao. 2008. Dynamic evaluation of regional air quality models: Assessing changes in O3 stemming from changes in emissions and meteorology. Atmospheric Environment. 42: 5110-5123.

Godowitch, J., Pouliot, G.A. and S.T. Rao. 2010. Assessing multi-year changes in modeled and observed urban NOx concentrations from a dynamic model evaluation perspective. Atm. Env. doi:10.1016/j.atmosenv.2010.04.040.

Godowitch, J., Gilliam, R., and S.T. Rao. 2011. Diagnostic evaluation of ozone production and horizontal transport in a regional photochemical air quality modeling system. Atm. Env. 45: 3977–3987.

Grell, G. A., Dudhia, J. and D. Stauffer. 1994. A Description of the Fifth-Generation PennState/NCAR Mesoscale Model (MM5), NCAR Technical Note, NCAR/TN-398 + STR, 128 pp., available at: http://nldr.library.ucar.edu/collections/technotes/asset-000-000-000-214.pdf.

Hirst, D., Storvik, G. and A.R. Syversveen. 2003. A Hierarchical Modeling Approach to Combining Environmental Data at Different Scales; *J. Royal Stat. Soc. Ser. C Appl. Stat.* 52: 377-390.

Hogrefe, C. and S.T. Rao. 2001. Demonstrating Attainment of the Air Quality Standards: Integration of Observations and Model Predictions into the Probabilistic Framework. J. Air & Waste Manage. Assoc. 51:1060-1072

Hogrefe, C., Lynn, B., Goldberg, R., Rosenzweig, C., Zalewsky, E., Hao, W., Doraiswamy, P., Civerolo, K., Ku, J.Y., Sistla, G. and P.L. Kinney. 2009. A Combined Model-Observation Approach to Estimate Historic Gridded Fields of PM2.5 Mass and Species Concentrations, Atm. Env., doi:10.1016/j.atmosenv.2009.02.031

Hogrefe, C., Hao, W., Zalewsky, E., Ku, J.-Y., Lynn, B., Rosenzweig, C., Schultz, , M.G., Rast, S., Newchurch, M., J., Wang, L, Kinney, P.L. and G. Sistla. 2010. An analysis of long-term regional-scale ozone simulations over the Northeastern United States: variability and trends, Atmos. Chem. Phys. 11: 567-582.

Hogrefe, C., Civerolo, K., Hao, W., Ku, J-Y, Zalewsky, E. and G. Sistla. 2012. Rethinking the assessment of photochemical modeling systems in air quality planning applications, Journal of the Air & Waste Management Association, Vol. 58 (8), 1086-1099, DOI:10.3155/1047-3289.58.8.1086.

Hogrefe, C., S. Roselle, R. Mathur, Rao, S.T. and S. Galmarini. 2014. Space-time analysis of the Air Quality Model Evaluation International Initiative (AQMEII) Phase 1 air quality simulations, Journal of the Air & Waste Management Association, Vol. 64, Iss. 4, DOI:10.1080/10962247.2013.811127

Houyoux, M. R., Vukovich, J. M., Coats, C. J., Wheeler, N. J. and P. Kasibhatla. 2000. Emission inventory development and processing for the seasonal model for regional air quality, J. Geophys. Res. 105: 9079–9090.

Jones, J.M., Hogrefe, C., Henry, R.F., Ku, J.Y. and G. Sistla. 2005. An assessment of the sensitivity and reliability of the relative reduction factor approach in the development of 8-hr ozone attainment plans. J. Air & Waste Manage. Assoc. 55(1): 13-19.

Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J. and M.Kolehmainen. 2004. Methods for imputation of missing values in air quality data sets. Atmospheric Environment. 38: 2895–2907

Kang, D., Mathur, R., Rao, S.T. and S. Yu. 2008. Bias adjustment techniques for improving ozone air quality forecasts. J. Geophys. Res. 113.

Kang, D., Mathur, R. and and S.T. Rao. 2010. Real-time bias-adjusted O_3 and $PM_{2.5}$ air quality index forecasts and their performance evaluations over the continental United States, Atm. Env. 2203-2212.

Kang, D., Hogrefe, C., Foley, K., Napelenok, S., Mathur, R. and S.T. Rao. 2013. Application of the Kolmogorov-Zurbenko filter and the decoupled direct 3D method for the dynamic evaluation of a regional air quality model. Atm. Env. 80: 58-69

Kulkarni, S., Kaduwela, A., Avise, J., DaMassa, J. and D. Chau. 2014. An extended approach to calculate the ozone response factors used in attainment demonstration for the national ambient air quality standards. Journal of the Air & Waste Management Association. 64 (10): 1204-1213, DOI:10.1080/10962247.2014.936984.

Lin, S., Jones, R., Pantea, C., Ozkaynak, H., Rao, S.T., Hwang, S. and V. Garcia. 2013. Impact of NOx Emission Reduction Policy on Hospitalizations for Respiratory Disease in New York State. J. of Exposure Science and Environmental Epidemiology. 73-80.

McMillan, N.; Holland, D.M.; Morara, M. and J. Feng. 2010. Combining Numerical Model Output and Particulate Data Using Bayesian Space-Time Modeling; *Environmetrics.* 2: 48-65.

Napelenok, S.L., Foley, K.M., Kang, D., Mathur, R., Pierce, T. and S.T. Rao. 2011. Dynamic evaluation of regional air quality model's response to emissions in the presence of uncertain emission inventories. Atmos. Env. 45: 4091-4098.

NRC. 1994. Science and Judgment in Risk Assessment (NRC Blue Book). Washington, DC: The National Academies Press.

Pegues, A.H., Cohan, D.S., Digar, A., Douglass, C., and R.S. Wilson. 2012. Efficacy of recent state implementation plans for 8-hour ozone. J. Air & Waste Manage. Assoc. 62(2): 252-261.

Pierce, T., Geron, C., Bender, L., Dennis, R., Tonnesen, G. and A. Guenther. 1998. Influence of Increased Isoprene Emissions on Regional Ozone Modeling; J. Geophys. Res. 103: 25611-25629.

Rao, S.T., Zurbenko, I., Neagu, R., Porter, P.S., Ku, J.Y. and R. Henry. 1997. Space and Time Scales in Ambient Ozone Data, Bull. of the Amer. Meteor. Soc. 78(10): 2153-2166.

Rao, S.T. and I.G. Zurbenko. 1994. Detecting and tracking changes in ozone air quality. Journal of the Air and Waste Management Association. 44: 1089–1092.

Rao, S.T., Galmarini, S. and K. Puckett. 2011. Air Quality Model Evaluation International Initiative (AQMEII), Bull. Amer. Meteor. Soc. 23-30.

Reich, B.J., Cooley, D., Foley, K., Napelenok, S., and B.A. Shaby. 2013. Extreme Value Analysis for Evaluating Ozone Control Strategies. Annals of Appl. Stat. 7(2): 739-762

Schulz, M. G., ed. 2010. RETRO final report: REanalysis of the TROpospheric chemical composition over the past 40 years – A long-term global modeling study of tropospheric chemistry, Reports on Earth System Science, edited by: available at: http://retro. enes.org/reports/RETRO Final Report.pdf, last access: 30 August 2010, Max Planck Institute for Meteorology, Hamburg, report no. 48/2007, ISSN 1614-1199, August 2007.

Simon, H., Baker, K.R. and S. Phillips. 2012. Compilation and interpretation of photochemical model performance statistics published between 2006 and 2012, Atmospheric Environment, 61, 124-139, http://dx.doi.org/10.1016/j.atmosenv.2012.07.012

Sistla, G., Hogrefe,C., Hao, W., Ku, J.Y., Zalewsky, E., Henry, R.F., and K. Civerolo. 2004. An operational assessment of the application of the relative reduction factors in the demonstration of the attainment of the 8-hr ozone national ambient air quality standard. J. Air & Waste Manage. Assoc. 54(8): 950-959.

USEPA. 2014. Air Quality Modeling Technical Support Document: Tier 3 Motor Vehicle Emission and Standards; Air Quality Assessment Division, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA-454/R-14-002, 128 pp.

Wang, L. and W. Chen. 2014. Equiratio cumulative distribution function matching as an improvement to the equidistant approach in bias correction of precipitation *Atmos. Sci. Let.* 15: 1–6, Published online 31 July 2013 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/asl2.454.

Legend for Tables

Table 1: Domain-wide mean metric of 4HDM8O3 predictions for all base year – target year pairs. Values in brackets are 95% confidence intervals

Table 2: Multiple comparisons of adjustment methods (95% level)

Table 3: Percent of base- / prediction- year pairs where a given method is better according to a particular metric. Example for MAB: MV is better than CMAQ (73% of pairs) and RATIO (58% of pairs).

Table 4. Effect of averaging the largest values for a given year. Each set of average values begins with the largest observed. For example, '1' means the largest while '3' means the 3 largest.

Legend for Figures

Fig.1: Ratios of future year (2004) to base year (2001) 4th highest daily maximum 8-hr ozone concentrations extracted from the observations and CMAQ output.

Fig. 2: Domain-wide statistics in predicting future 4HDM8O3 concentrations using 1996 as the base year for raw CMAQ output, RATIO approach, and the four new methods introduced in this study. a) RMSE (ppb), b) MAB (ppb), and c) Correlation.

Fig. 3: Spatial image of RSME (in ppb) in predicted 4HDM8O3 concentrations for all base year/prediction year combinations from each method.

Fig. 4: Fraction of sites predicted 4HDM8O3 concentrations fall within a given distance from observations

Fig. 5: RSME (in ppb) for the 2004 daily maximum 8-hr ozone concentrations using 2001 as the base year for the upper half of the distribution of daily maximum 8-hr ozone concentrations.

Fig. 6: Change in daily maximum 8-hr ozone distributions between 2001 base year and 2004 prediction year for each method: a) Change in the upper half of the frequency distribution; b) change in the 98th percentile (changes in the CDF2 method are the same as CMAQ).

Fig. 7: Domain-wide statistics for same year predictions: a) mean absolute bias, b) standard deviation, c) correlation coefficient. Note that no results are shown for the RATIO and CDF2 approaches. For the RATIO approach, the same year ratio of model predictions is equal to one and, thus, the same year predicted value equals the observed value, and for the CDF2 approach, the same year predicted value equals that for the CDF1 approach.

Fig. 8: Mean bias of same-year predictions by percentiles. a) 2001, b) 2004.

Table 1. Domain-wide mean metric of 4HDM8O3 predictions for all base year – target year pairs. Values in brackets are 95% confidence intervals

	Method					
Metric	**RATIO**	**CMAQ**	**MV**	**TC**	**CDF1**	**CDF2**
MAB (ppb)	**7.48** [7.39 7.57]	**8.31** [8.07 8.54]	**7.26** [7.18 7.36]	**7.06** [6.96 7.17]	**7.00** [6.89 7.05]	**6.90** [6.82 6.98]
RMSE (ppb)	**9.49** [9.38 9.60]	**10.35** [10.08 10.62]	**9.25** [9.14 9.37]	**9.81** [9.52 10.17]	**8.91** [8.81 9.02]	**8.71** [8.62 8.81]
FMAB (%)	**8.7** [8.6 8.8]	**9.6** [9.3 9.9]	**8.5** [8.4 8.6]	**8.2** [8.1 8.3]	**8.1** [8.0 9.2]	**8.1** [8.0 8.2]
FRMSE (%)	**11.1** [11.0 11.2]	**11.8** [11.5 12.1]	**10.7** [10.6 10.8]	**11.4** [11.1 11.8]	**10.5** [10.4 10.6]	**10.3** [10.2 10.4]
R2	**0.26** [0.245 0.272]	**0.32** [0.281 0.353]	**0.31** [0.300 0.328]	**0.29** [0.267 0.306]	**0.33** [0.316 0.344]	**0.32** [0.306 0.333]
Index of agreement	**0.78** [0.772 0.785]	**0.77** [0.750 0.780]	**0.81** [0.799 0.811]	**0.79** [0.776 0.799]	**0.82** [0.811 0.821]	**0.81** [0.802 0.813]

Table 2. Multiple comparisons of adjustment methods (95% level)

		MAB	RMSE
RATIO	<	**CMAQ**	**CMAQ, TC**
	>	**MV, TC, CDF1, CDF2**	**CDF1, CDF2**
CMAQ	<	**NONE**	**NONE**
	>	**ALL**	**RATIO, MV, CDF1, CDF2**
MV	<	**RATIO , CMAQ**	**CMAQ, TC**
	>	**TC, CDF1, CDF2**	**CDF2**
TC	<	**RATIO, CMAQ, MV**	**NONE**
	>	**NONE**	**RATIO, MV, CDF1, CDF2**
CDF1	<	**RATIO, CMAQ, MV**	**RATIO, CMAQ, TC**
	>	**NONE**	**NONE**
CDF2	<	**RATIO, CMAQ, MV**	**RATIO, CMAQ, MV, TC**
	>	**NONE**	**NONE**

Table 3. Percent of base- / prediction- year pairs where a given method is better according to a particular metric.
Example for MAB: MV is better than CMAQ (73% of pairs) and RATIO (58% of pairs).

	CMAQ	RATIO	MV	TC	CDF1	CDF2
MAB						
CMAQ		36	27	27	23	17
RATIO	64		42	32	16	17
MV	73	58		46	38	27
TC	73	68	54		43	34
CDF1	77	84	62	57		30
CDF2	83	83	73	66	70	
RMSE						
CMAQ		47	30	46	28	20
RATIO	53		38	44	18	11
MV	70	62		54	37	26
TC	54	56	46		42	32
CDF1	72	82	63	58		31
CDF2	80	89	74	68	69	
R2						
CMAQ		52	38	48	39	19
RATIO	48		34	38	16	04
MV	62	66		46	37	19
TC	52	62	54		42	28
CDF1	61	84	63	58		20
CDF2	81	96	81	72	80	

Table 4. Effect of averaging the largest values for a given year. Each set of average values begins with the largest observed. For example, '1' means the largest while '3' means the 3 largest.

number in average:	1	2	3	4	5	6	7	8	9	10	4th highest
mean absolute bias											
RATIO	12.71	10.75	9.44	8.57	7.94	7.49	7.15	6.88	6.66	6.47	7.48
CMAQ	11.99	10.78	9.90	9.26	8.77	8.37	8.05	7.79	7.55	7.35	8.31
MV	11.07	9.66	8.74	8.15	7.70	7.37	7.11	6.89	6.72	6.56	7.26
TC	10.95	9.46	8.53	7.92	7.45	7.11	6.85	6.63	6.46	6.30	7.06
CDF1	10.96	9.47	8.52	7.89	7.44	7.10	6.83	6.60	6.42	6.25	7.00
CDF2	10.28	8.98	8.16	7.62	7.24	6.96	6.73	6.55	6.39	6.26	6.90
RMSE											
RATIO	16.12	13.62	11.90	10.78	9.99	9.43	9.01	8.68	8.41	8.19	9.49
CMAQ	14.88	13.29	12.11	11.32	10.72	10.26	9.89	9.58	9.31	9.07	10.35
MV	14.05	12.26	11.08	10.31	9.74	9.32	8.99	8.73	8.50	8.31	9.25
TC	14.60	12.80	11.66	10.91	10.35	9.93	9.60	9.33	9.11	8.91	9.81
CDF1	13.90	12.03	10.82	10.02	9.45	9.023	8.69	8.41	8.19	7.99	8.91
CDF2	12.97	11.36	10.31	9.62	9.13	8.77	8.49	8.26	8.07	7.91	8.71
R^2											
RATIO	0.22	0.29	0.35	0.39	0.42	0.45	0.47	0.48	0.49	0.50	0.26
CMAQ	0.31	0.36	0.40	0.42	0.44	0.45	0.46	0.47	0.48	0.48	0.32
MV	0.27	0.34	0.39	0.43	0.46	0.4	0.49	0.50	0.51	0.52	0.31
TC	0.26	0.33	0.38	0.41	0.43	0.45	0.46	0.48	0.48	0.49	0.29
CDF1	0.28	0.36	0.41	0.45	0.48	0.50	0.52	0.53	0.54	0.55	0.33
CDF2	0.29	0.36	0.41	0.44	0.47	0.48	0.50	0.51	0.52	0.52	0.32

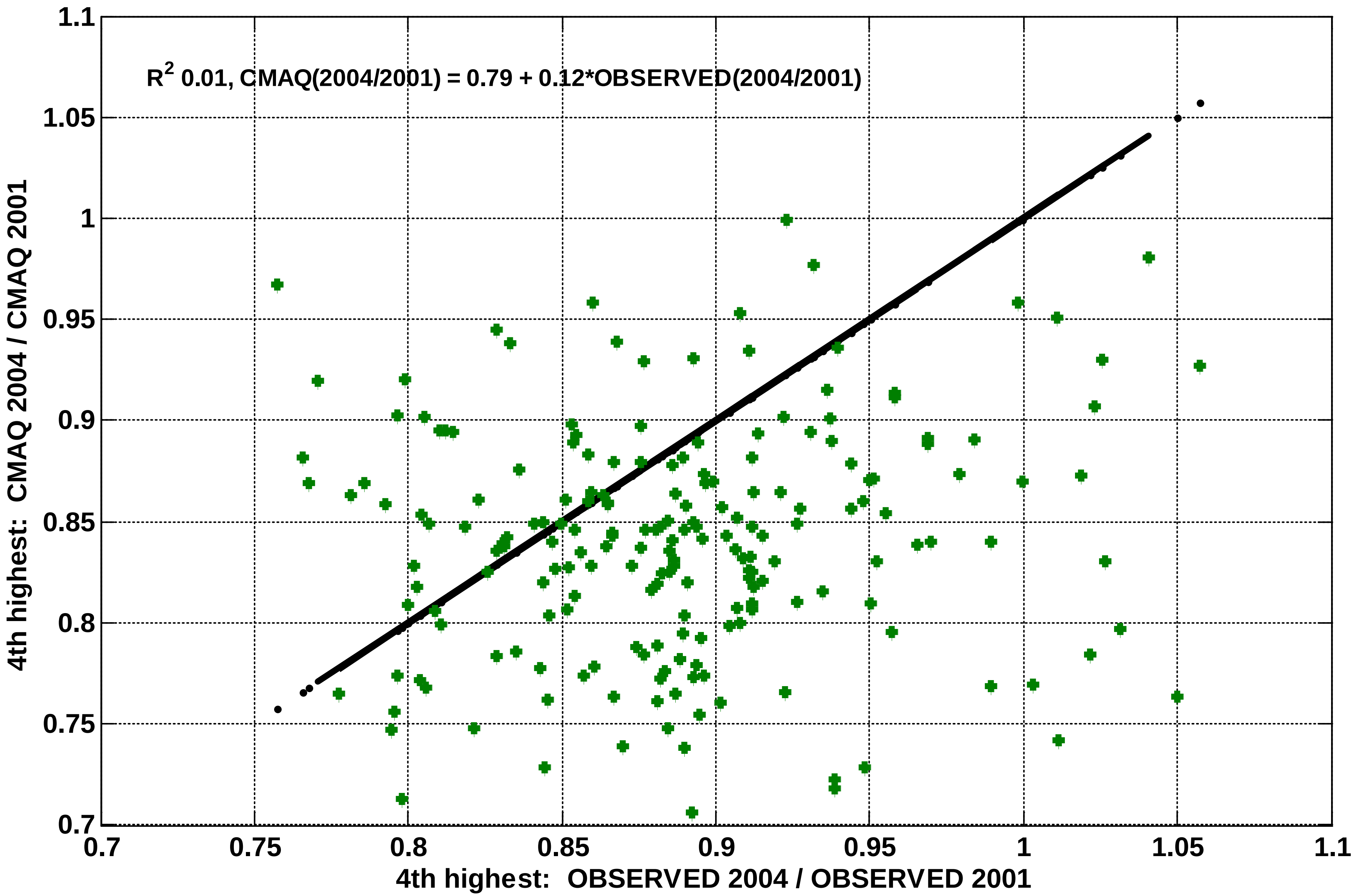

Fig. 1: Ratios of future year (2004) to base year (2001) 4th highest daily maximum 8-hr ozone concentrations extracted from the observations and CMAQ output.

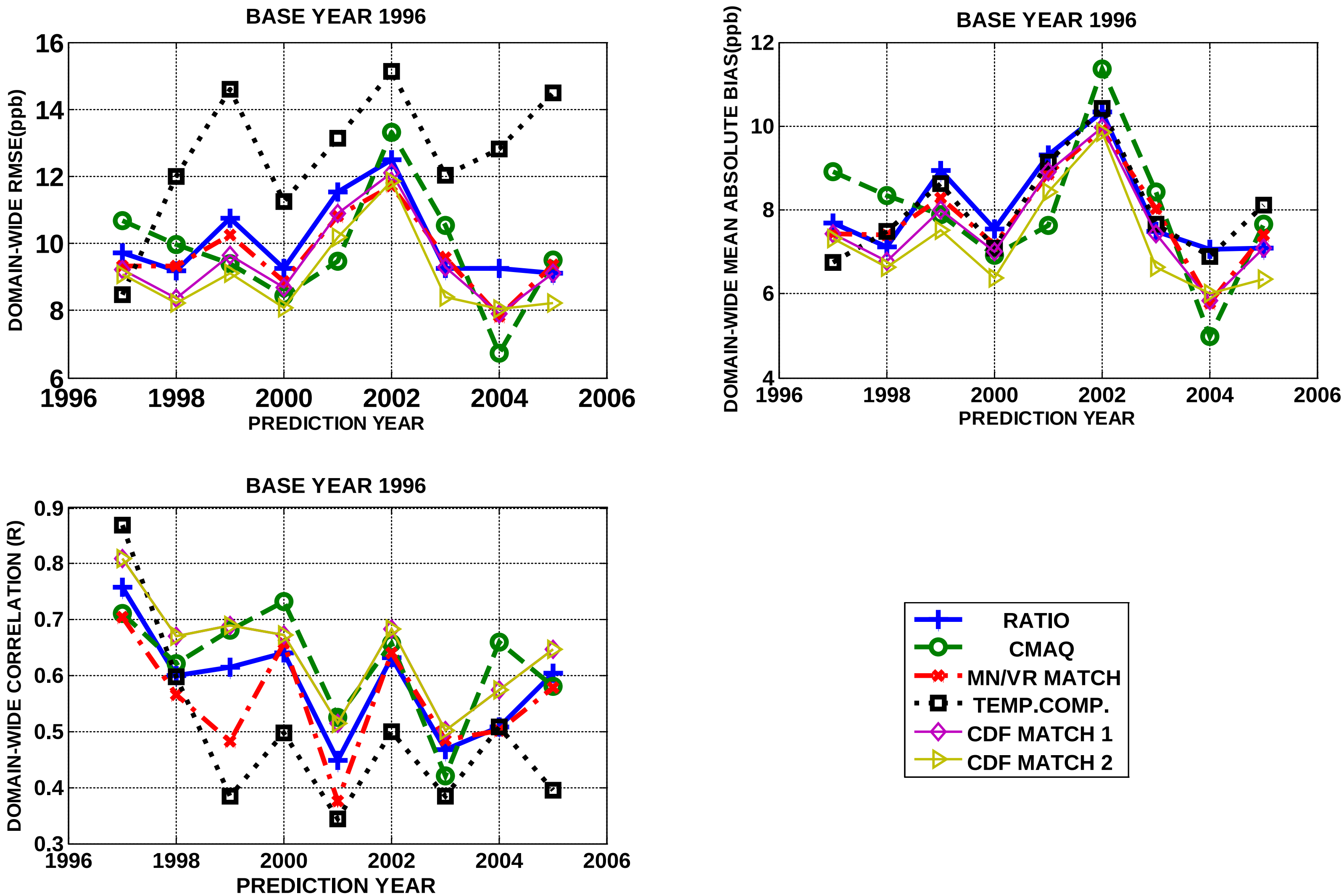

Fig. 2. Domain-wide statistics in predicting future 4HDM8O3 concentrations using 1996 as the base year for raw CMAQ output, RATIO approach, and the four new methods introduced in this study. a) RMSE (ppb), b) MAB (ppb), and c) correlation

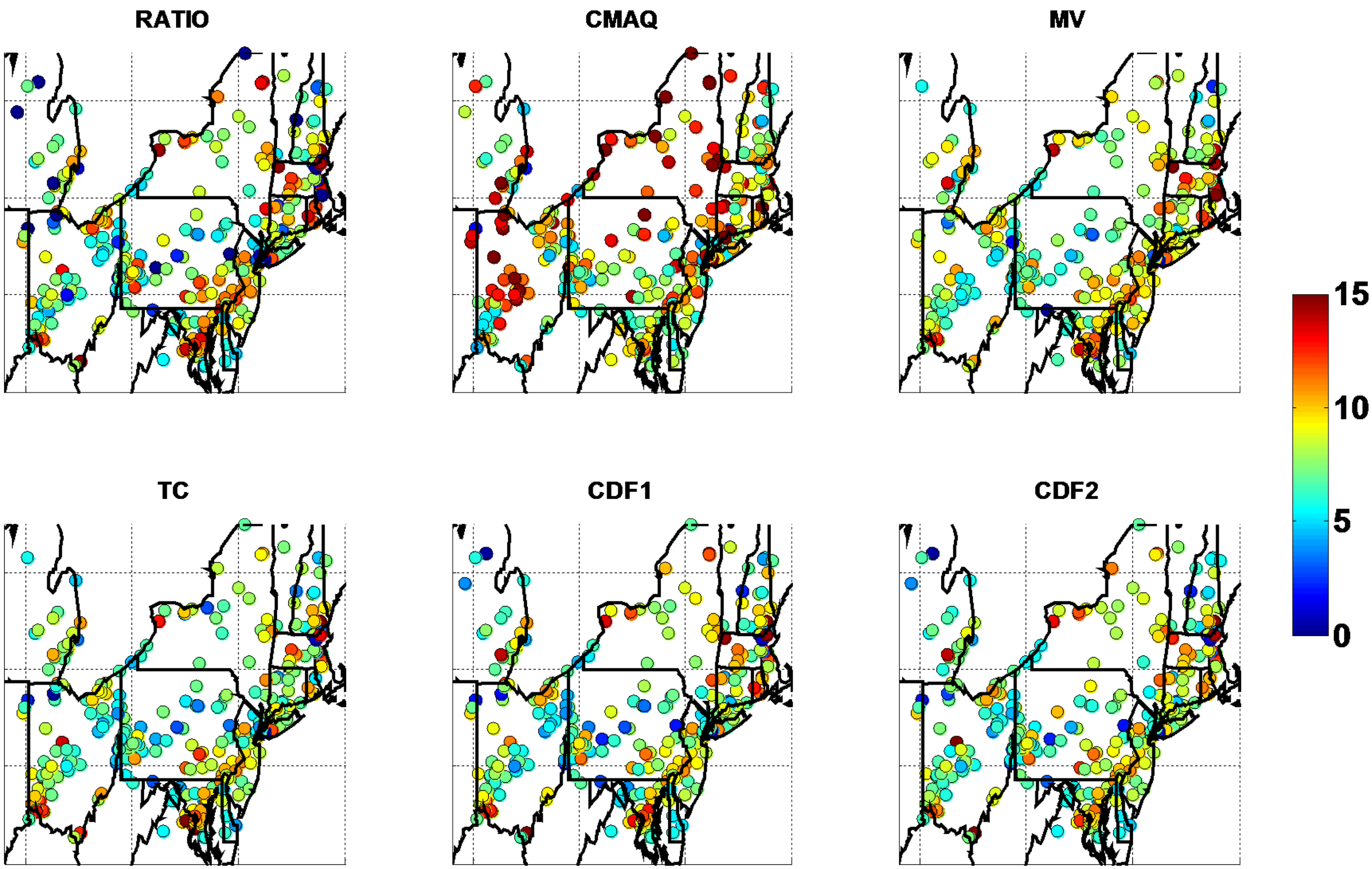

Fig. 3: Spatial image of RSME (in ppb) in predicted 4HDM8O3 concentrations for all base year/prediction year combinations from each method.

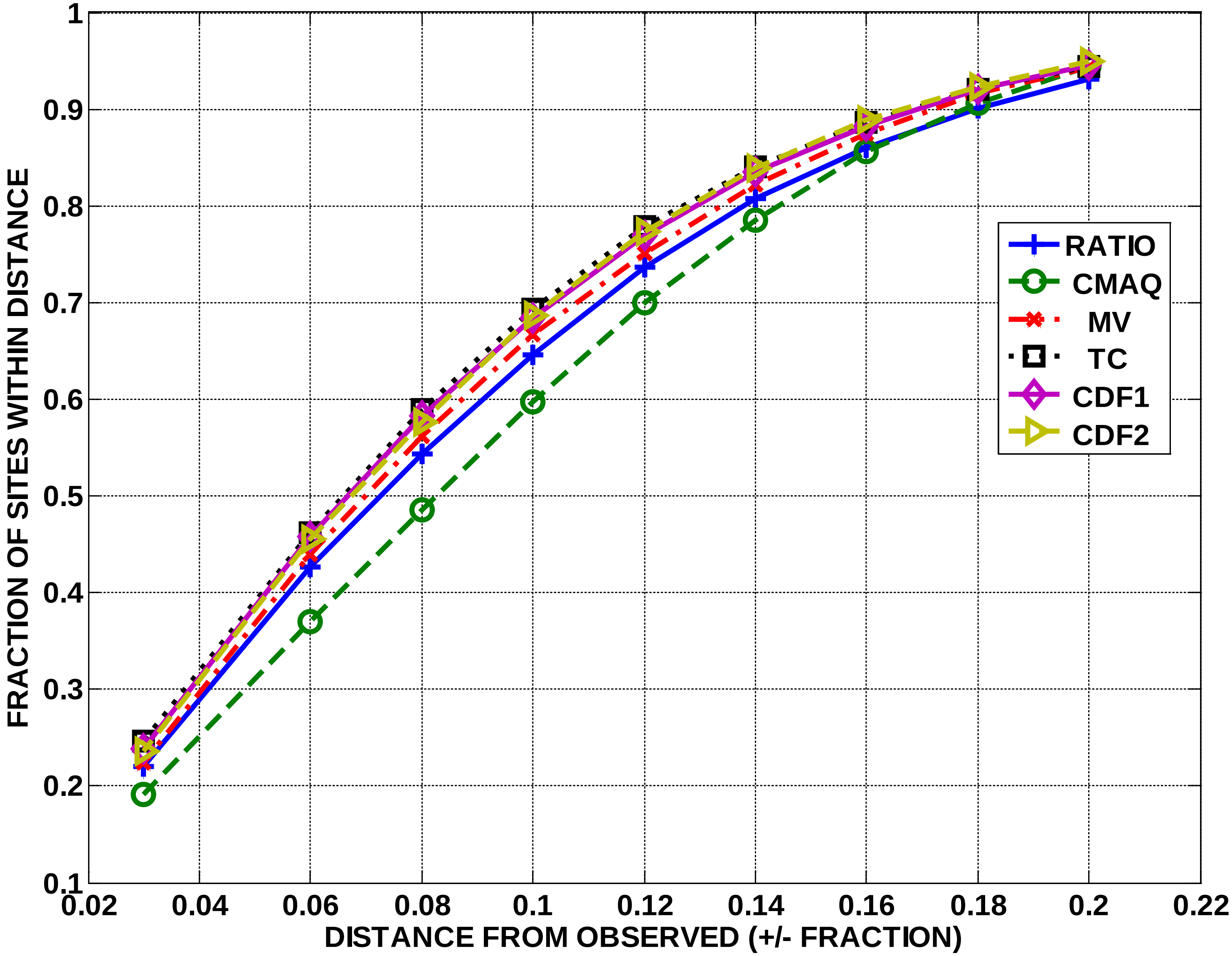

Fig. 4: Fraction of times predicted 4HDM8O3 concentrations fall within a given distance from observations

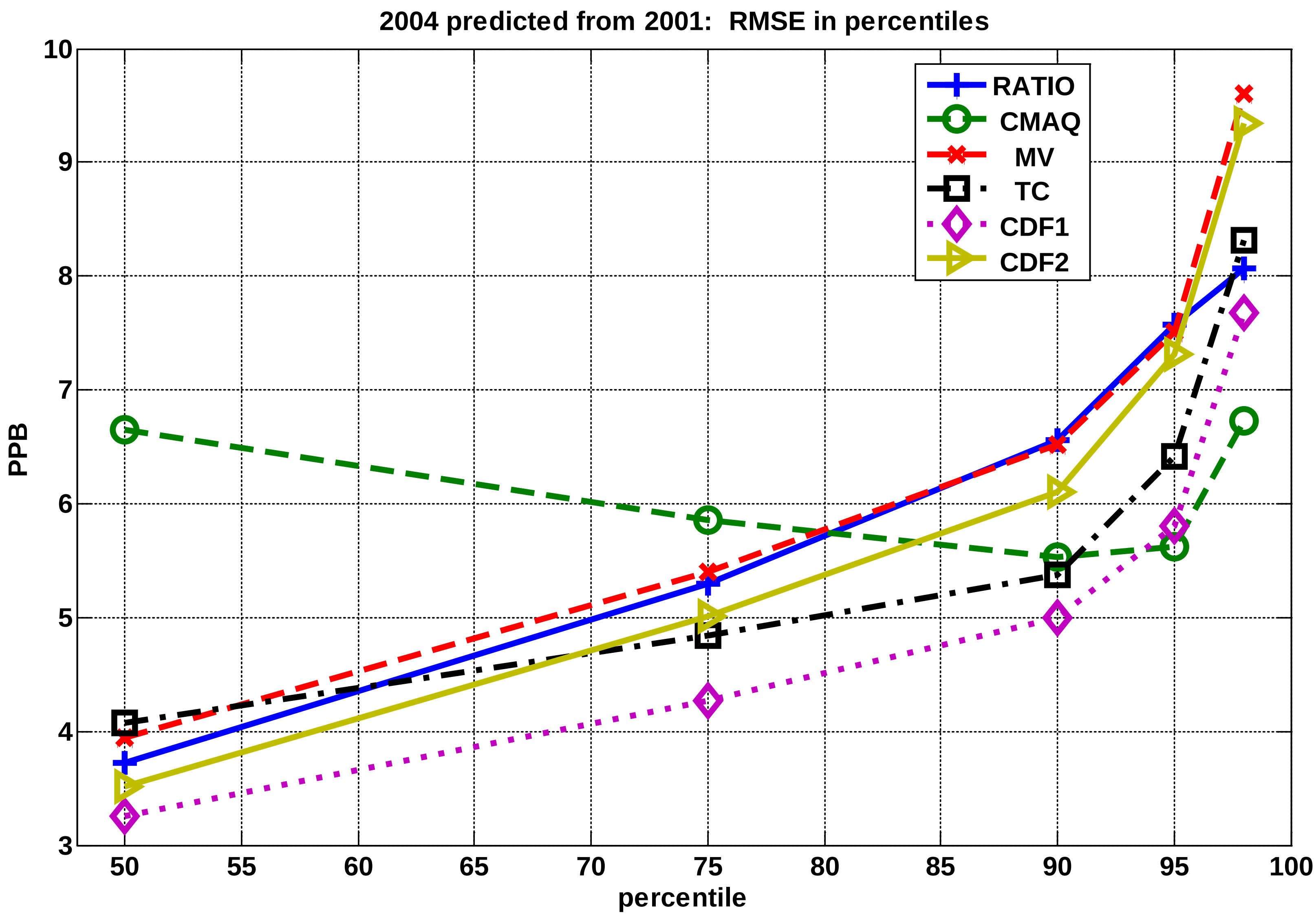

Fig. 5: RSME (in ppb) for the 2004 daily maximum 8-hr ozone concentrations using 2001 as the base year for the upper half of the distribution of daily maximum 8-hr ozone concentrations.

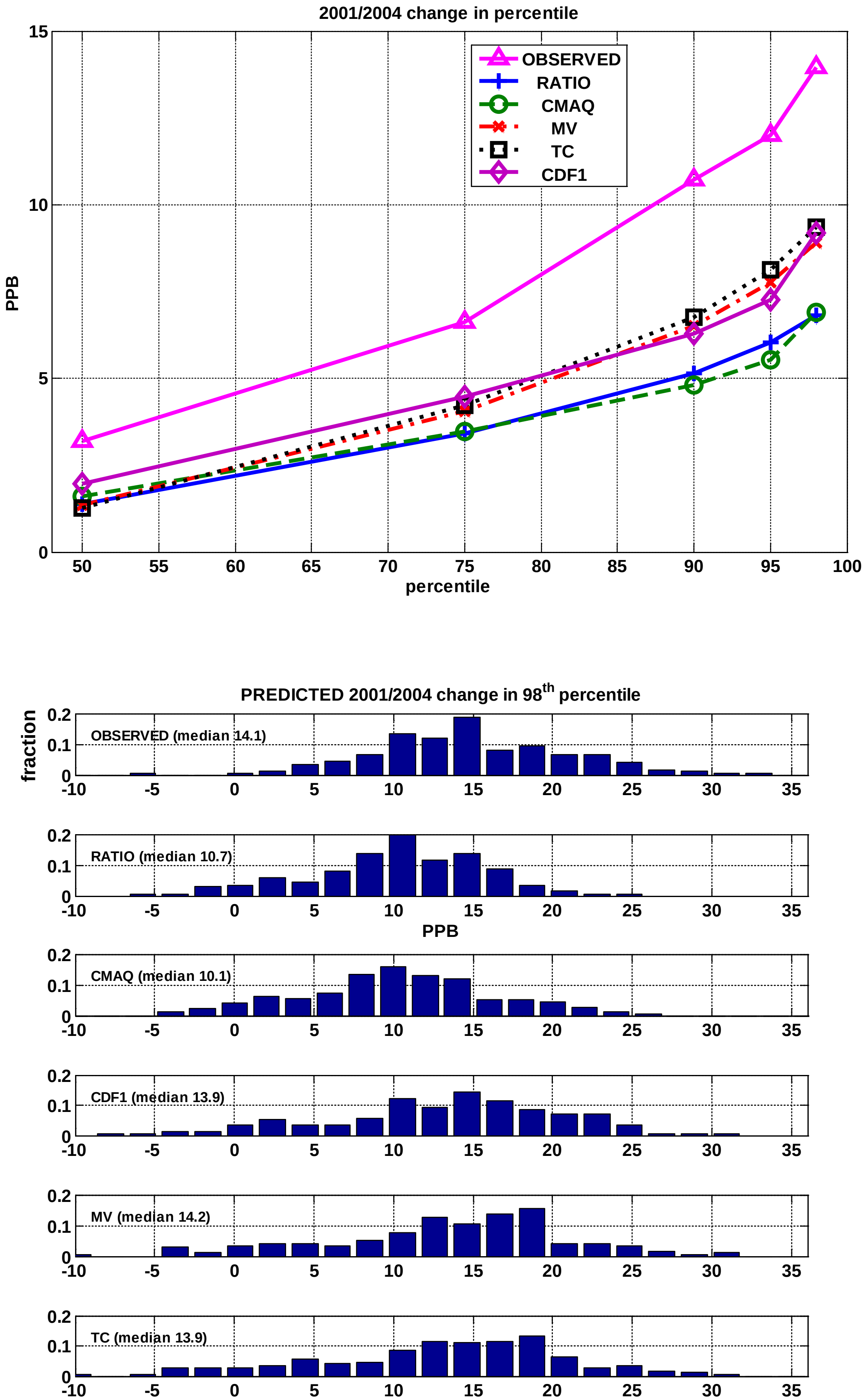

Fig. 6: Change in daily maximum 8-hr ozone distributions between 2001 base year and 2004 prediction year for each method: a) Change in the upper half of the frequency distribution; b) change in the 98th percentile (changes in the CDF2 method are the same as CMAQ).

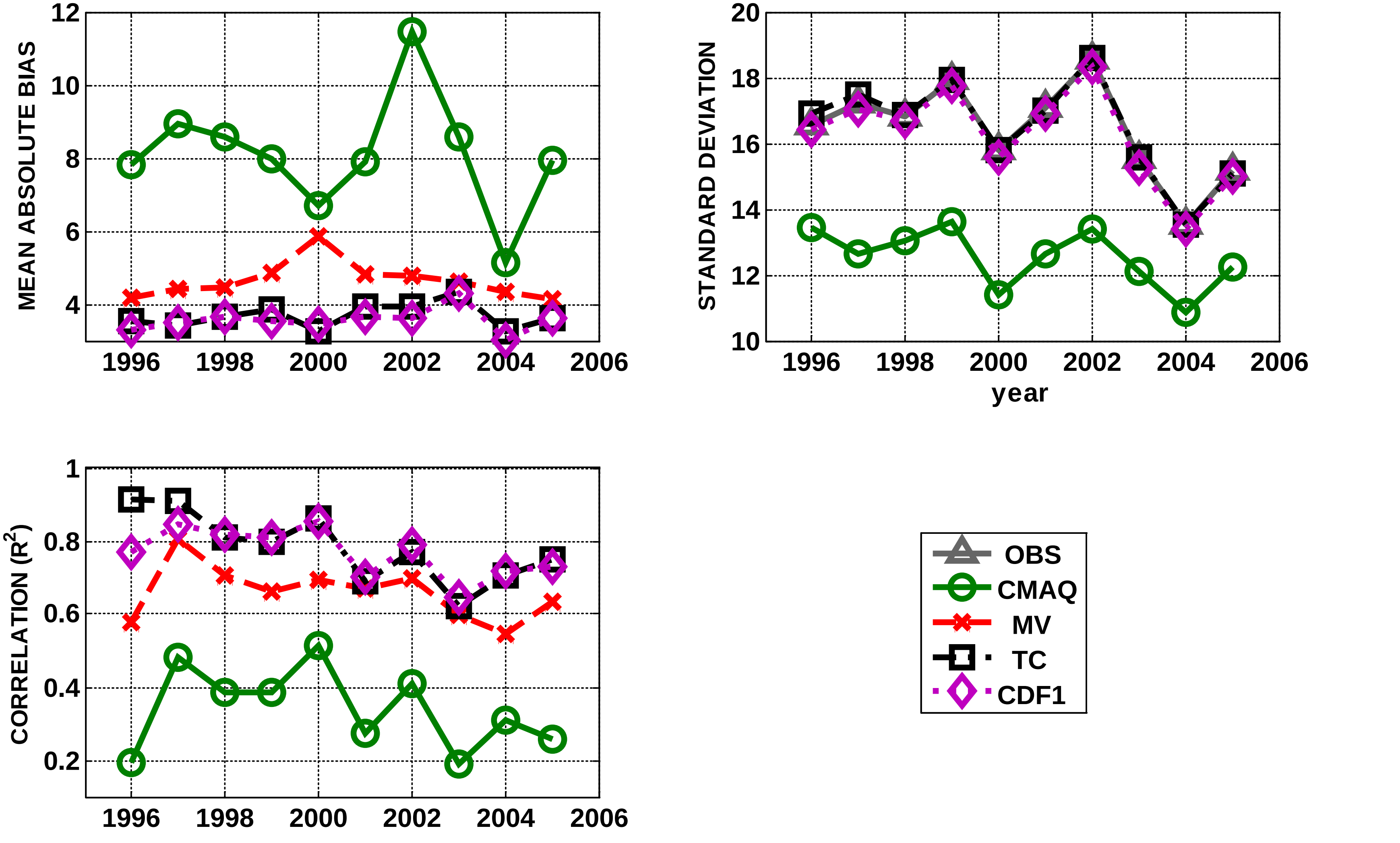

Fig 7: Domain-wide mean metrics for using same-year CMAQ adjustments (methods applied and compared with observations for a single year): a) mean absolute bias of annual 4th highest 8-hour daily maximum (4HDM8O3), b) standard deviation of 8-hour daily maximum , c) correlation coefficient (R^2) between observed and adjusted 4HDM8O3. Note that no results are shown for the RATIO and CDF2 approaches. For the RATIO approach, the same year ratio of model predictions is equal to one and, thus, the same year predicted value equals the observed value, and for the CDF2 approach, the same year predicted value equals that for the CDF1 approach.

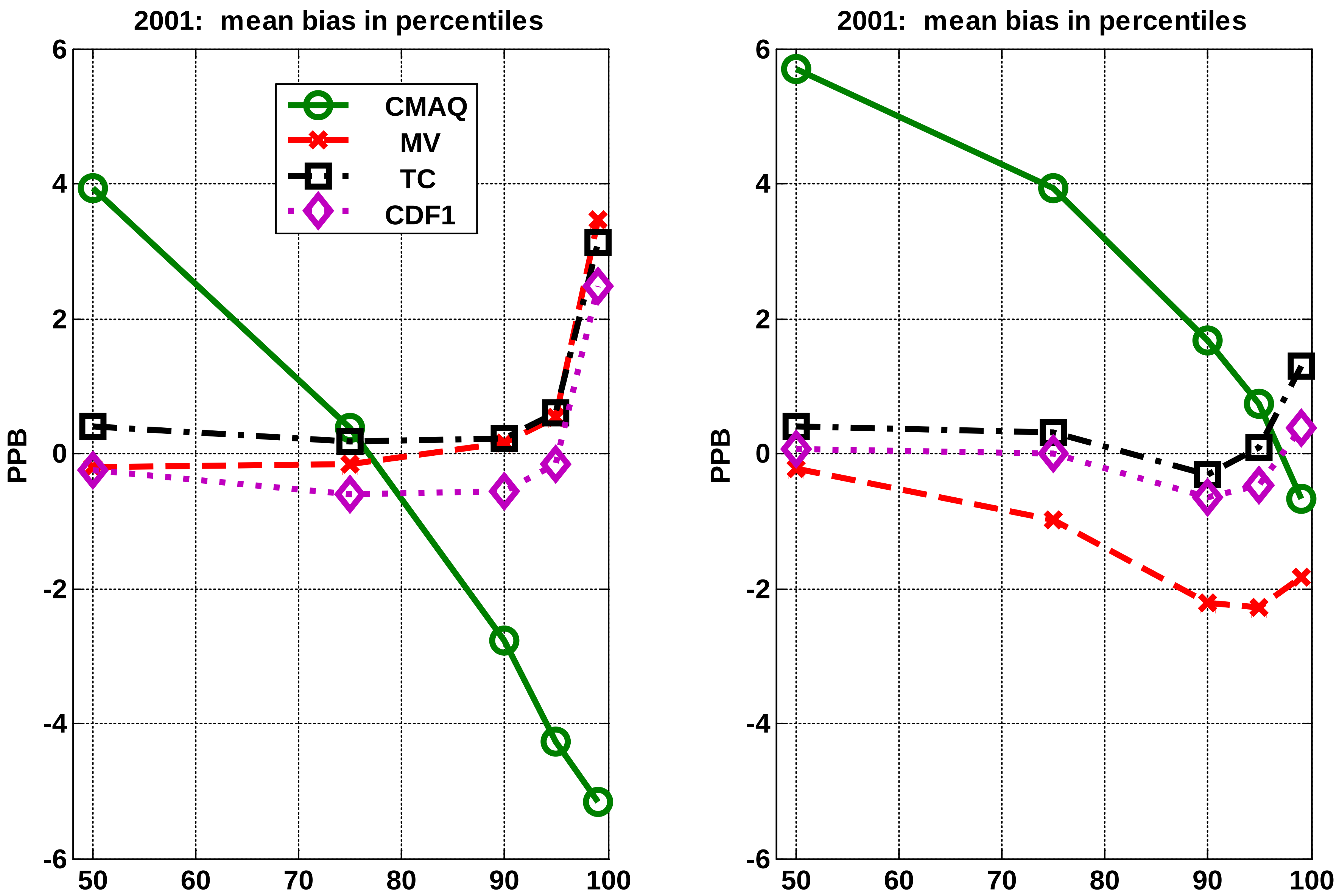

Fig. 8: Mean bias of same-year predictions by percentiles. a) 2001, b) 2004