CRC Report No. A-89-2

USING UNCERTAIN REGIONAL AIR QUALITY MODEL OUTPUTS FOR OZONE SOURCE APPORTIONMENT

Deliverable 1C

June 2014

COORDINATING RESEARCH COUNCIL, INC.
5755 NORTH POINT PARKWAY·SUITE 265·ALPHARETTA, GA 30022

The Coordinating Research Council, Inc. (CRC) is a non-profit corporation supported by the petroleum and automotive equipment industries. CRC operates through the committees made up of technical experts from industry and government who voluntarily participate. The four main areas of research within CRC are: air pollution (atmospheric and engineering studies); aviation fuels, lubricants, and equipment performance, heavy-duty vehicle fuels, lubricants, and equipment performance (e.g., diesel trucks); and light-duty vehicle fuels, lubricants, and equipment performance (e.g., passenger cars). CRC's function is to provide the mechanism for joint research conducted by the two industries that will help in determining the optimum combination of petroleum products and automotive equipment. CRC's work is limited to research that is mutually beneficial to the two industries involved, and all information is available to the public.

CRC Project No. A-89

Using Uncertain Regional Air Quality Model Outputs for Ozone Source Apportionment

Written by: P.S. Porter

Diagnostic of Bias Adjustment Methods

A brief diagnostic analysis of the bias-adjustment methods described in the most recent report was carried out. The goal of this analysis is to gain an understanding of bias-adjustment methods that will lead to their improvement.

1. Non-stationary adjustment parameters

Previous results showed that bias-adjusted CMAQ values closely resembled observations for the year that they were developed. For example, with respect to the **CDF1** method, observed and modeled concentrations are rank-ordered to establish empirical CDF's. The **CMAQ** CDF is then modified using linear regression so that it matches the **OBSERVED** CDF as closely as possible:

$$\textit{OBSERVED base-year QUANTILES} = Ko + K1 \bullet \textit{CMAQ base-year QUANTILES} \qquad (10)$$

$$\textit{estimated prediction-year QUANTILES} = Ko + K1 \bullet \textit{CMAQ prediction-year QUANTILES}$$

The base-year parameters, Ko and K1, are applied to prediction years. However, these parameters change from year to year (Figure 1). Bias can result when base-year adjustment parameters are used in a prediction-year.

2. Distributions of thechange in the annual 4th highest

Histograms of the change in the 98th percentile (actually the 4th annual highest value) between 2002 (base-year) and 2005 (prediction-year) (about 250 sites) illustrate the goal of bias adjustment (Figures 2). The distribution of CMAQ (second panel) is more narrow and biased downward relative to that observed (top panel). The distributions of the adjusted values (panels 3-5) more closely resemble the observed distribution. Forecasting is not displayed in this figure; for example, the QQ parameters for 2002 and 2005 were used for the 2002 and 2005 CDF1 values, respectively.

Figure 3 is the same information as Figure 2 except that the 2005 values were derived from 2002 information: the **OBSERVED** 2002 QQ and **CMAQ** 2005 QQ plots were used to create 2005 CDF1 values. Here, the distributions of change in the 98th percentile (4th highest) are about as

wide as observed, but they are all biased downward. The widths of the CMAQ and RRF distributions are roughly the same as in Figure 2.

3. Sign errors

A sign error refers to predicting the wrong direction of prediction-year change. Sign errors lead to severely biased estimates. Figure 4 shows the rate of getting the correct sign for same-year values (same values as Figure 2). CMAQ and adjusted CMAQ values have the correct sign at the 98th percentile for about 70% and 80% of the cases, respectively. When 2005 values are derived from 2002 information, performance of adjustment methods and CMAQ are roughly equal at 70% (Figure 5). Figure 5 also shows that 90th percentile sign errors are smaller than at other percentiles.

4. Percentiles and changes in percentiles

The next few figures examine bias in percentiles and changes in percentiles for CMAQ and the adjusted CMAQ values, beginning with CMAQ (Figures 6 and 7). CMAQ percentiles for 2002 and 2005 are displayed in Figure 6 while changes in percentiles between 2002 and 2005 are shown in Figure 7. CMAQ is biased downward for both years (except for the 50th percentile in 2005). Changes in percentiles are also biased downward (Figure 7). Figures 8 and 9 have the same information for the CDF1 method. Percentiles for the same-year adjustments closely match those observed for both years. When 2002 adjustment parameters are used to predict 2005 values a positive bias is introduced in the 2005 percentiles (purple line in Figure 8) which means a negative bias for the percentile change (Figure 9).

Percentile and percentile change patterns are repeated for the other methods in Figures 10 and 11 (mean/variance (MV)),temporal components (TC) in Figures 12 and 13, RRF in Figures 14 and 15, and CDF2 in Figure 16 (recall that the CDF2 method uses the change in CMAQ directly). The pattern of positive bias at every percentile leading to negative bias with respect to percentile change is present in all methods.

Performance metrics (plotted against percentile in Figures 17-22) provide another illustration of the results discussed above. In these figures there is no forecasting (no use of 2002 adjustment parameters for 2005 values). CMAQ and the MV adjustment have negative bias at all percentiles for the year 2002, while CDF1 and TC are much closer to observations at most percentiles (Figure 17). The CDF1 and TC methods are nearly unbiased at the 95th percentile for 2002 (Figure 17) and 2005 (Figure 18). Figures 19 and 20 show mean absolute bias for 2002 and 2005, repectively. One observes that absolute bias increases with percentile for both years. The root mean square errors (RMSE) patterns for 2002 and 2005 are similar to those for mean absolute bias (Figures 21 and 22, respectively).

Performance metrics for the case of 2002 information being used to forecast 2005 values tell a different story (Figures 23-25). With respect to mean bias, absolute bias and RMSE, the MV method is better at higher percentiles beginning with the 75th.

Figures 26-28 show errors in predicting the change between 2002 and 2005 when 2002 information is applied to 2005. With respect to mean bias (Figure 26), CMAQ and RRF have larger errors than the other three methods (CDF1, TC and MV) at percentiles 75 and greater. Absolute bias and RMSE performances are mixed, with CMAQ doing less well than the other methods at the 75th and 90th percentiles and all mehods with similar performance at the 99th percentile (Figures 27 and 28).

5. Next steps

The next step is to determinc whether prediction-year model information (ozone, emissions and meteorology) can be used to predict sign errors and future year adjustment parameters.

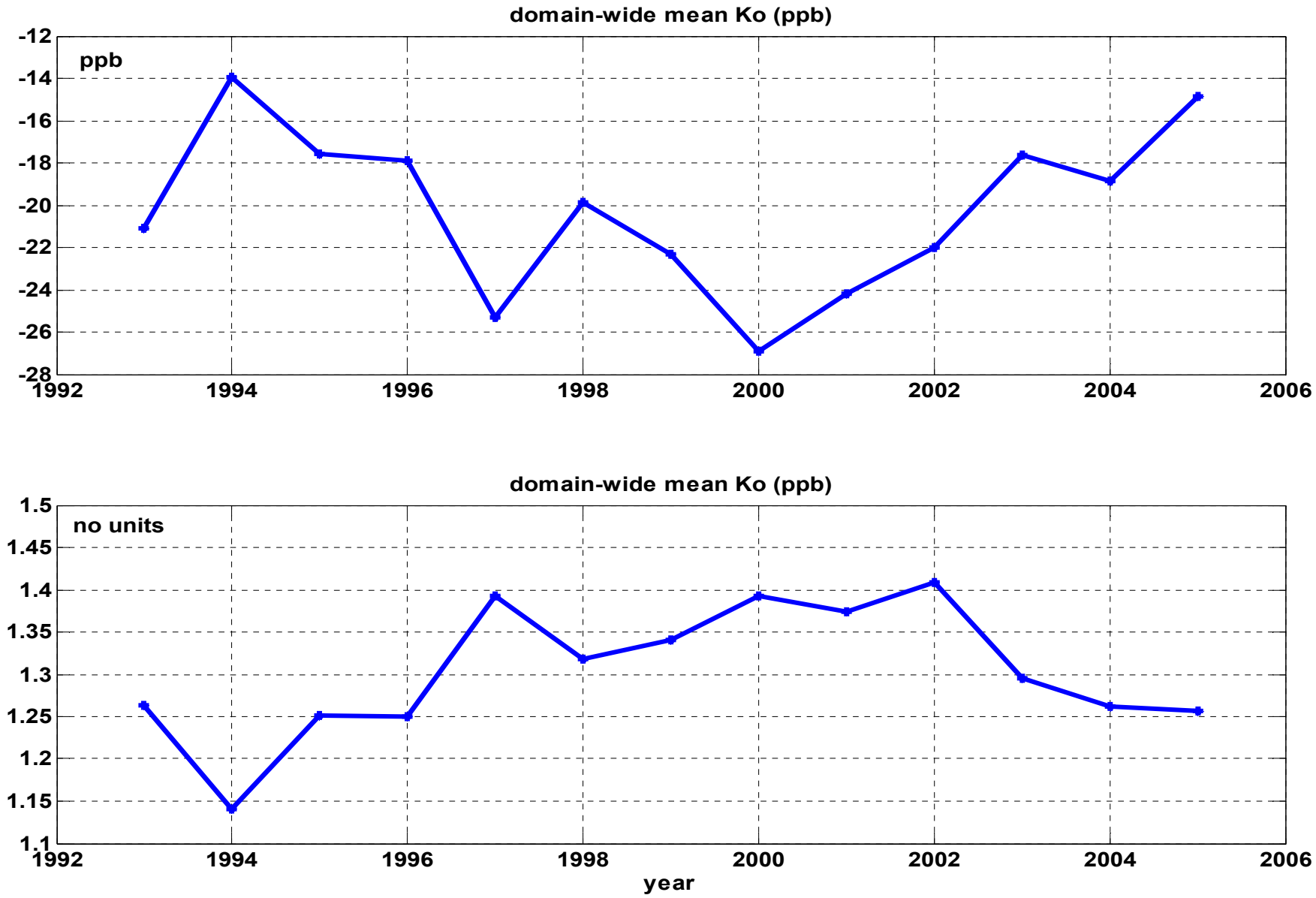

Figure 1. ***CDF1 method: domain-wide mean values for Ko and K1***

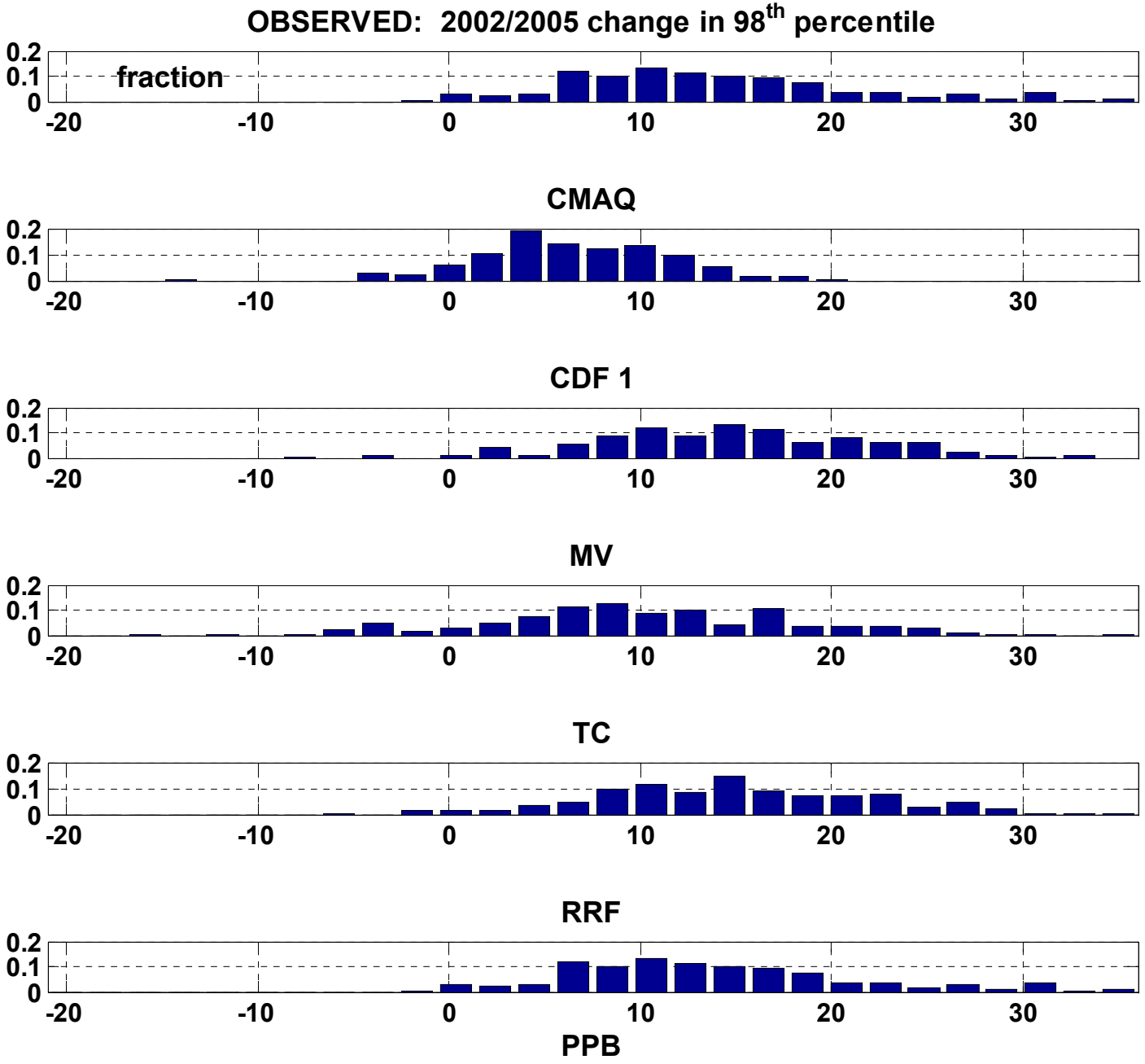

Figure 2. ***Observed*** **distributions of the change in the 98th percentile between base year 2002 and prediction year 2005. CMAQ's distribution is biased downward and more narrow than observed. The other methods have better agreement with observations.**

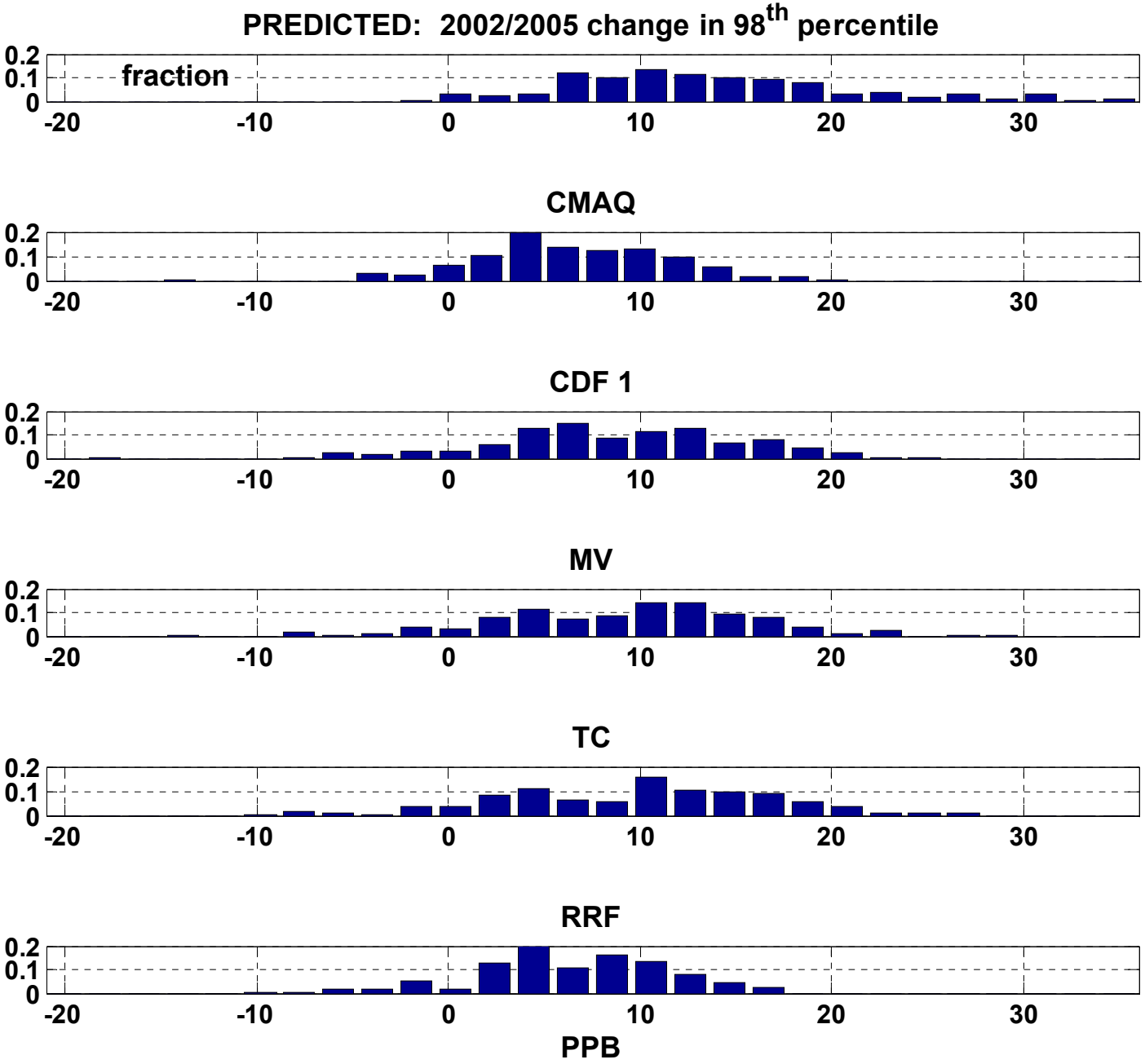

Figure 3. ***Predicted*** **distributions of the change in the 98th percentile between base year 2002 and prediction year 2005. CDF1, MV and TC are more dispersed than CMAQ (closer to the** ***Observed dispersion*****) but all methods are biased low.**

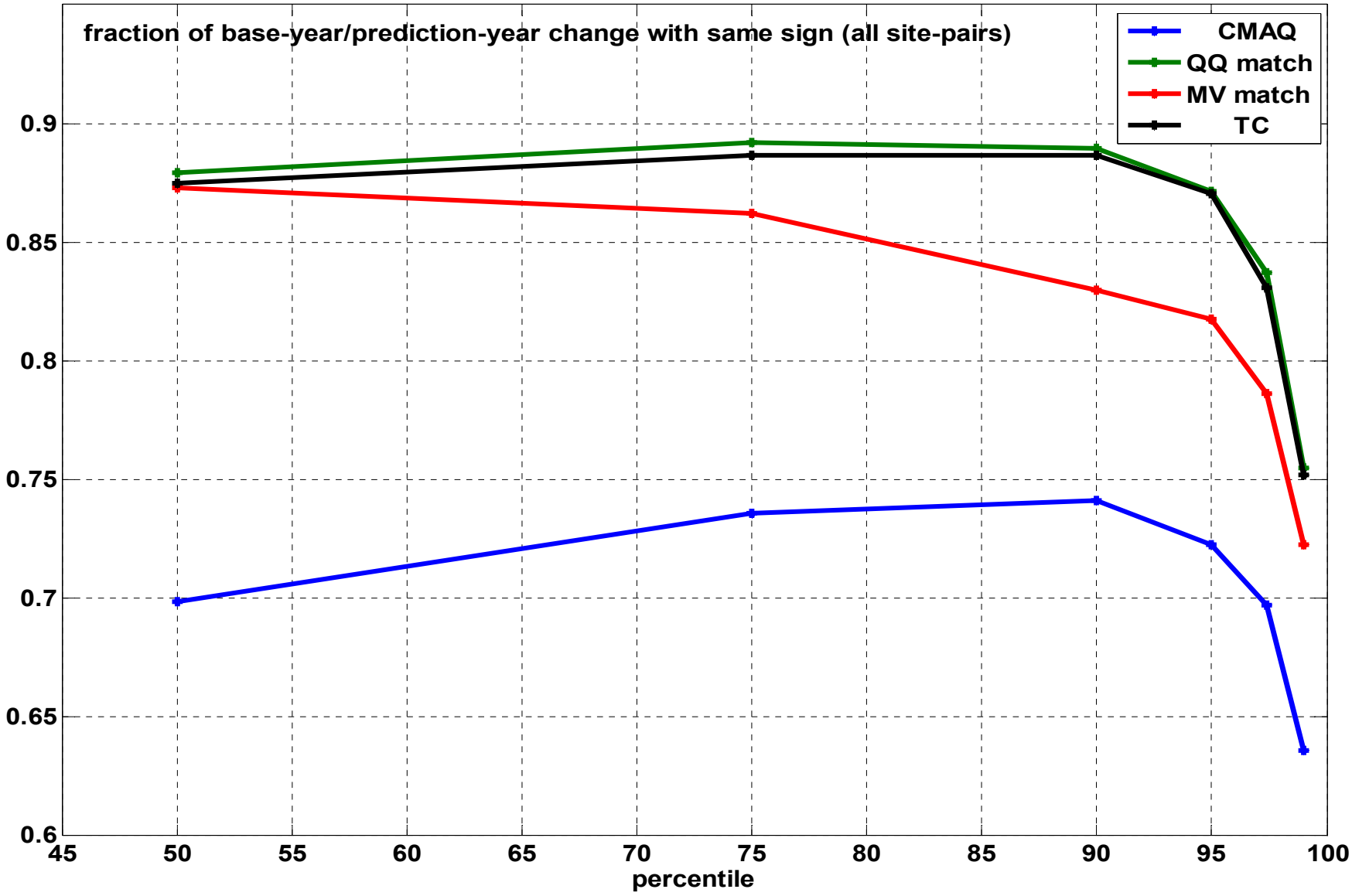

Figure 4. Same-year values: fraction of correct signs for the change from base-year to prediction-year number of pairs about 30,000).

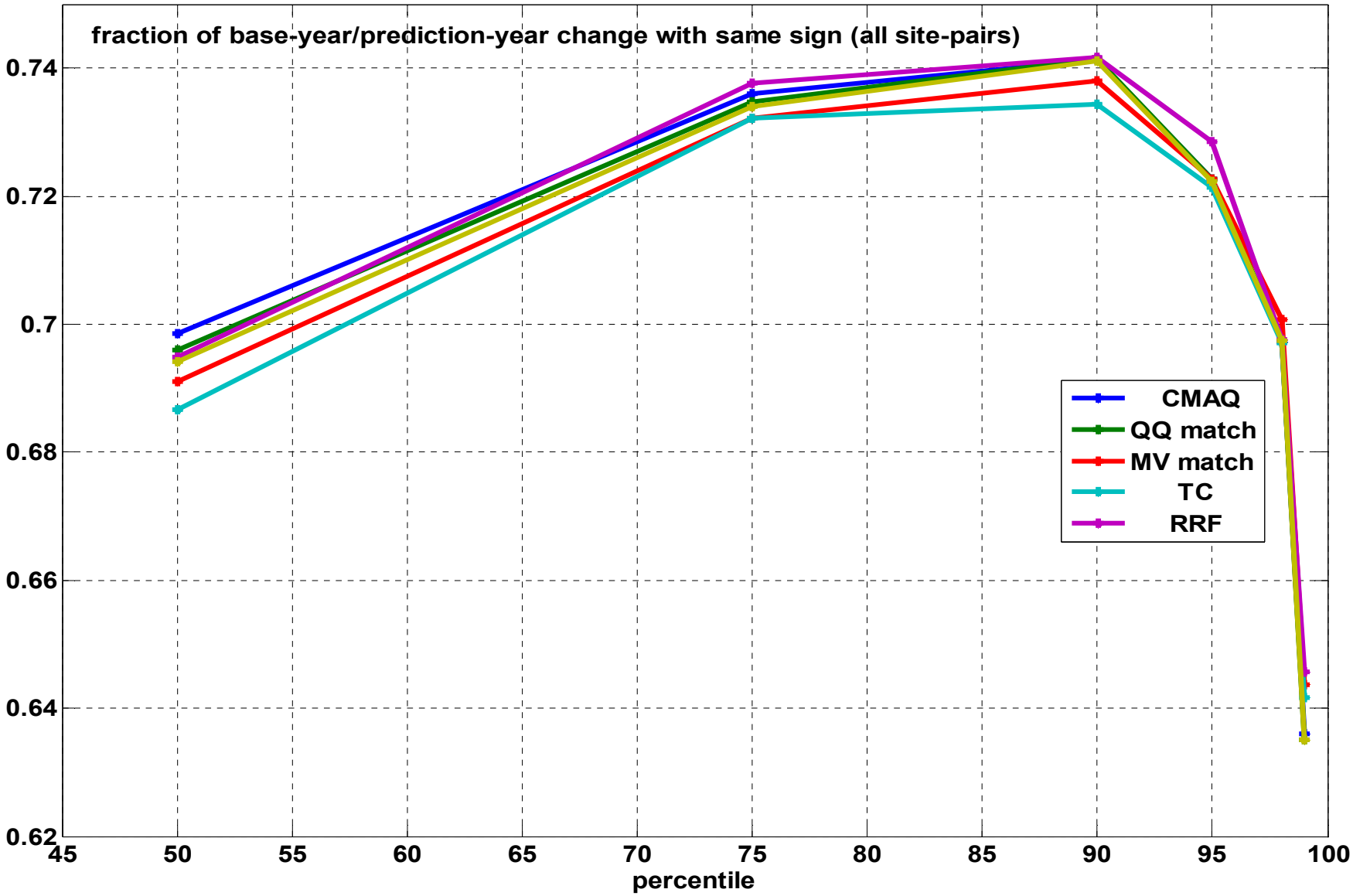

Figure 5. Fraction of correct signs for the change from base-year to prediction-year (number of pairs about 30,000). If the direction of change is incorrect, errors are large compared to when the predicted sign is correct. This aspect of method performance decays rapidly after the 90th percentile.

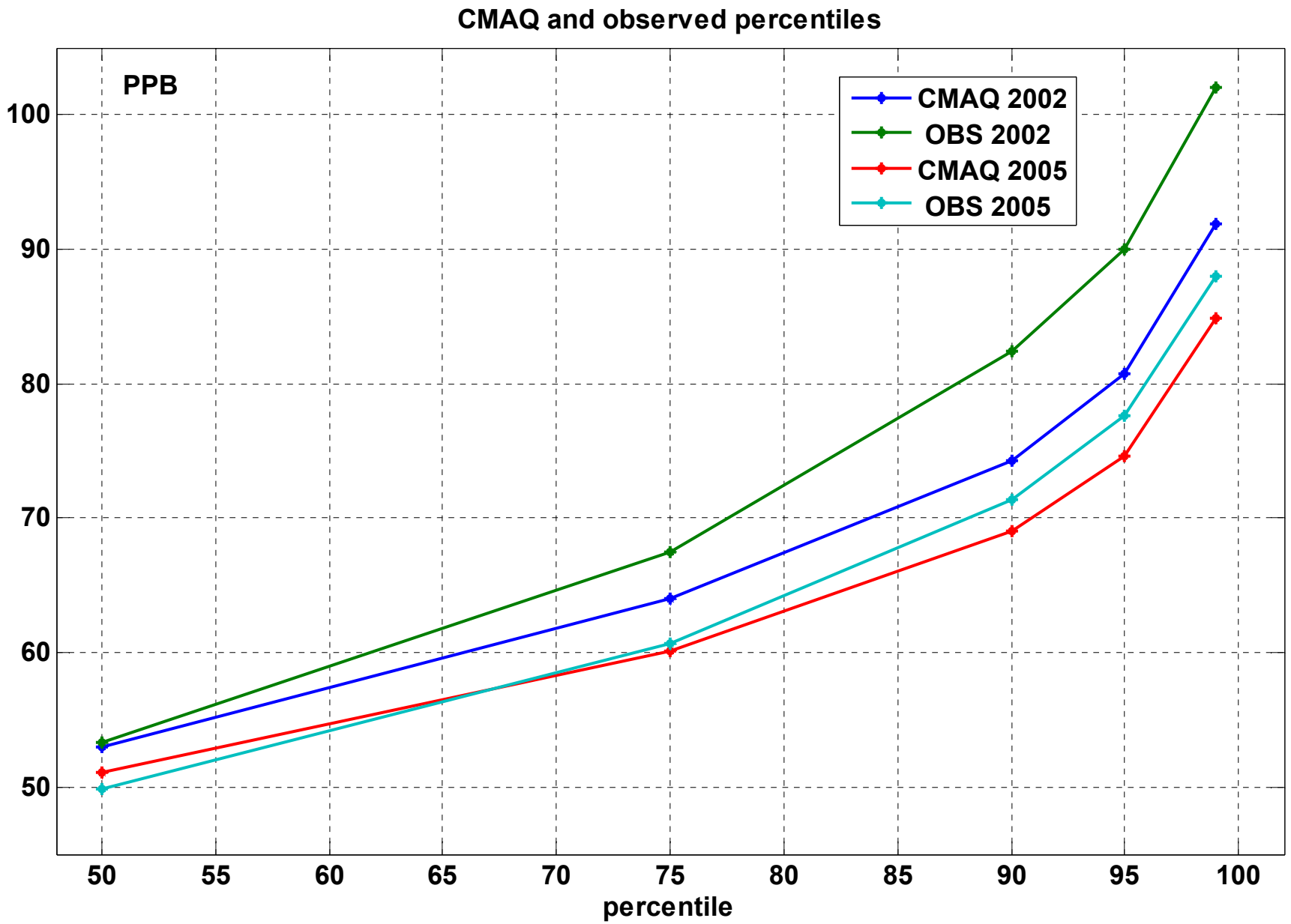

Figure 6. CMAQ and observed mean percentiles for base year 2002 and prediction year 2005. CMAQ is biased low for both years beginning with the 75th percentile.

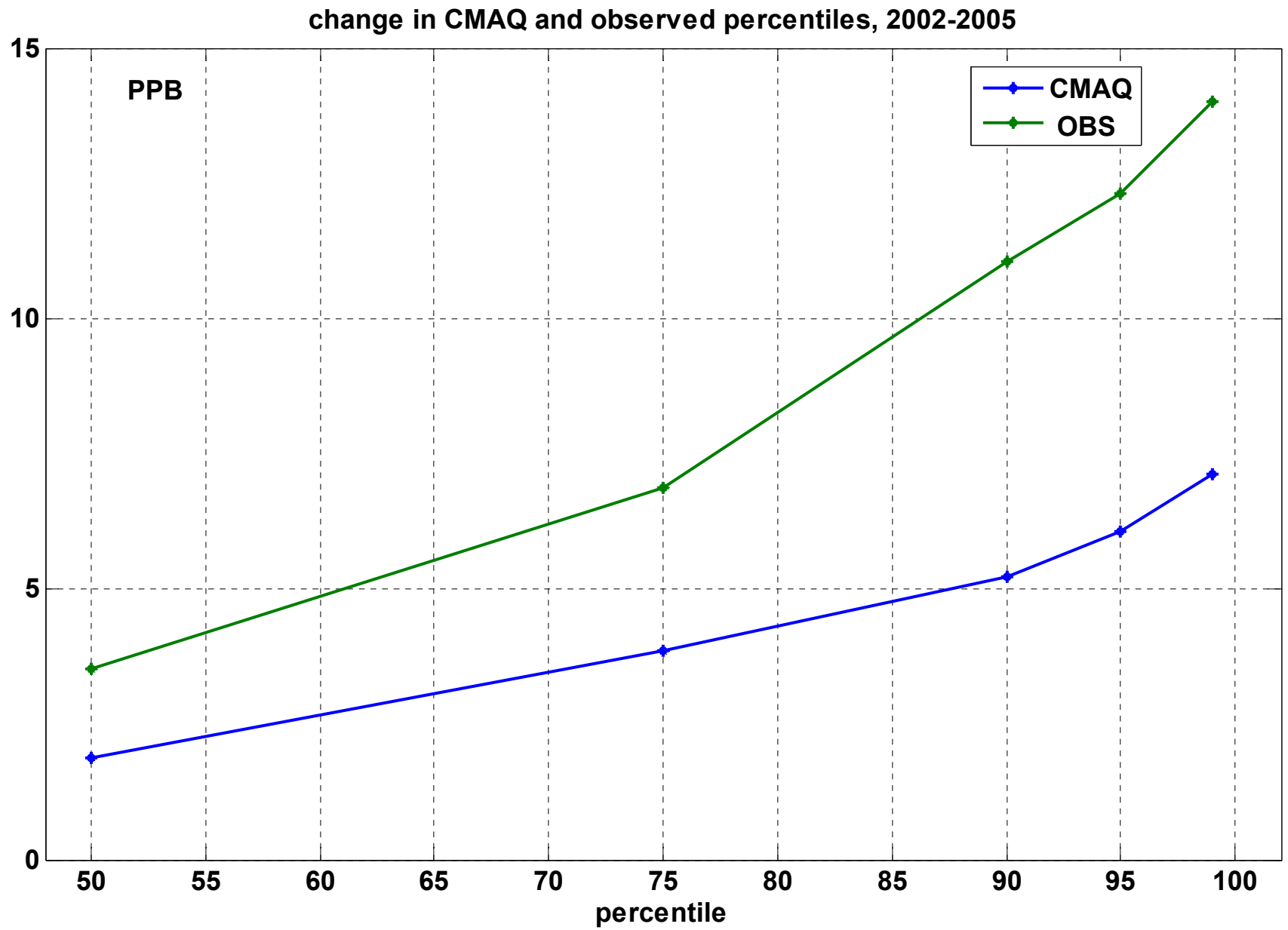

Figure 7. ***Change*** **in CMAQ and observed mean percentiles between the base year 2002 and prediction year 2005. CMAQ change is biased low for all percentiles.**

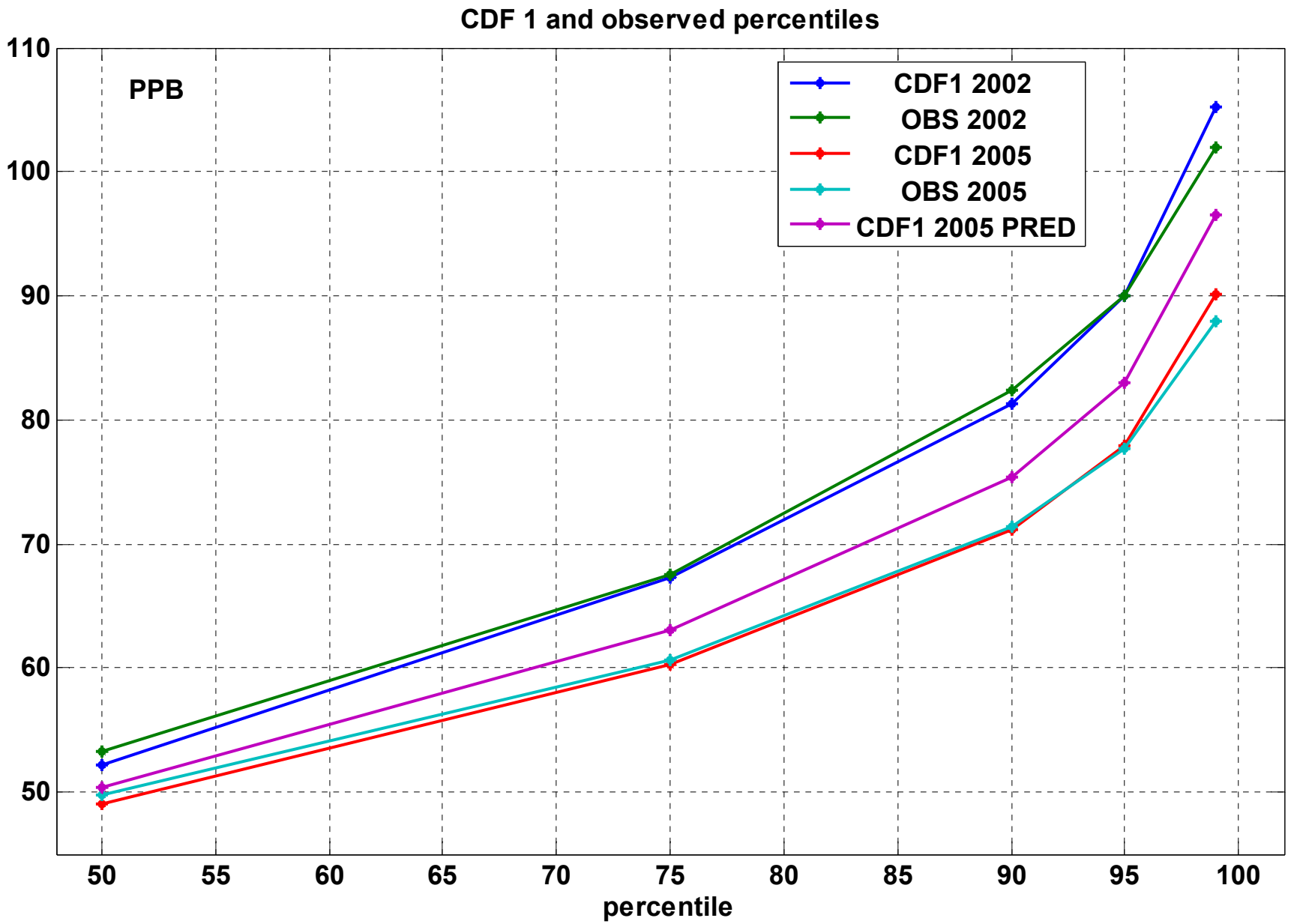

Figure 8. CDF matching (QQ matching) and observed mean percentiles for base year 2002 and prediction year 2005. Bias increases for 2005 when forecast from the 2002 base year (purple line).

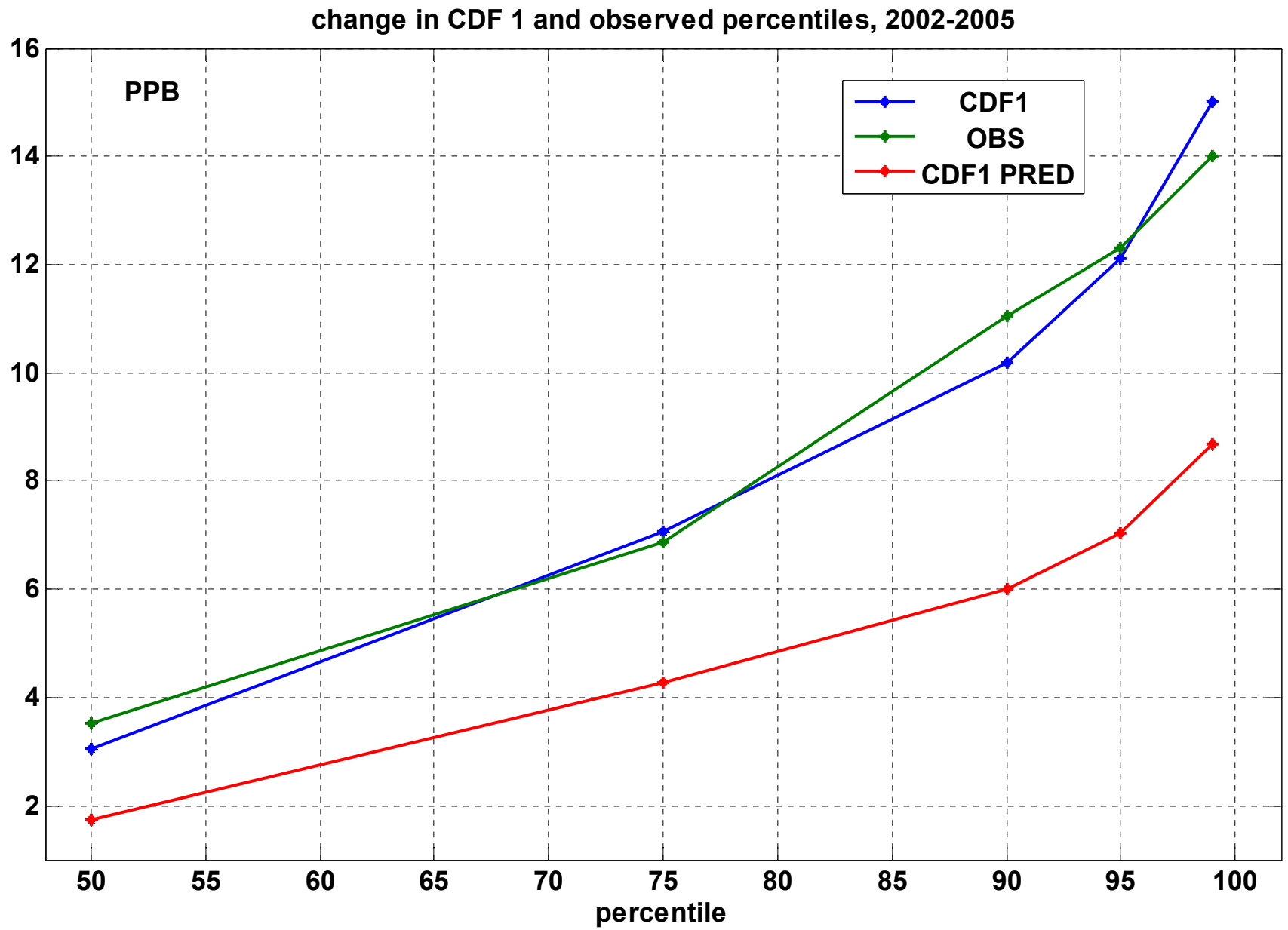

Figure 9. ***Change*** **in CDF matching (QQ matching) and observed in mean percentiles between the base year 2002 and prediction year 2005. CDF matching change is biased low for all percentiles.**

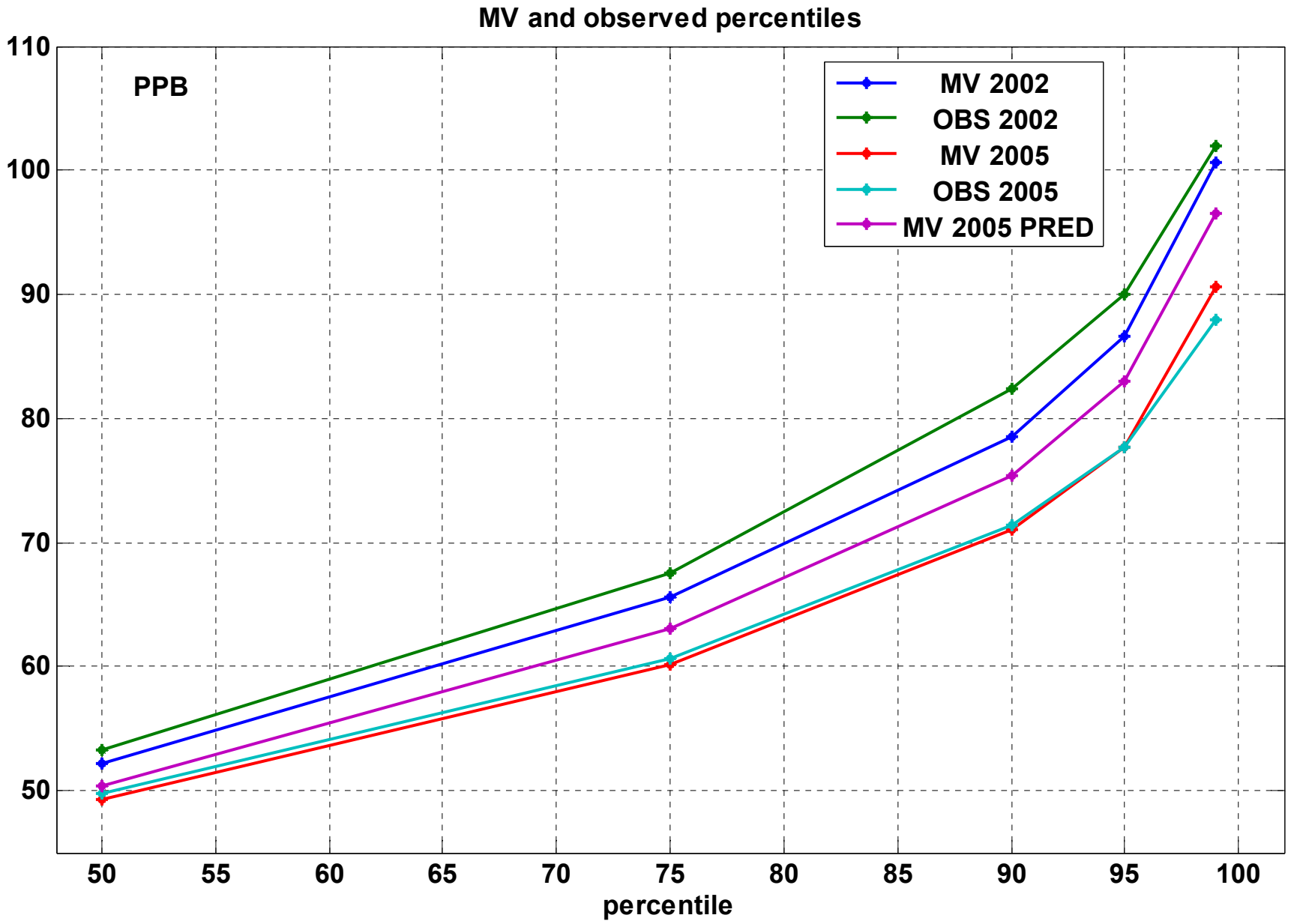

Figure 10. Mean/variance matching and observed mean percentiles for base year 2002 and prediction year 2005. Bias increases for 2005 when forecast from the 2002 base year (purple line).

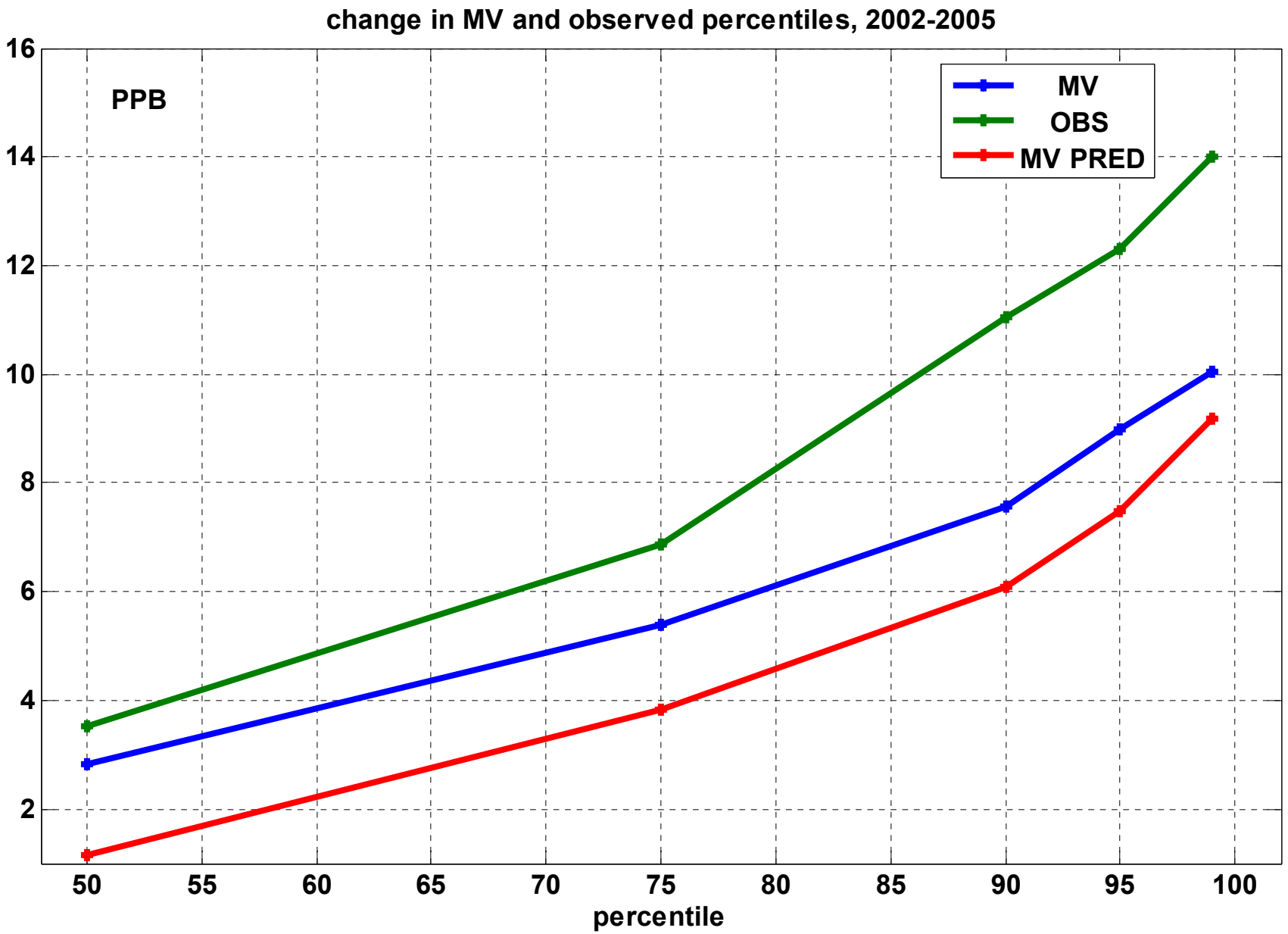

Figure 11. ***Change*** **in Mean/variance matching and observed mean percentiles between the base year 2002 and prediction year 2005. Mean/variance matching change is biased low for all percentiles.**

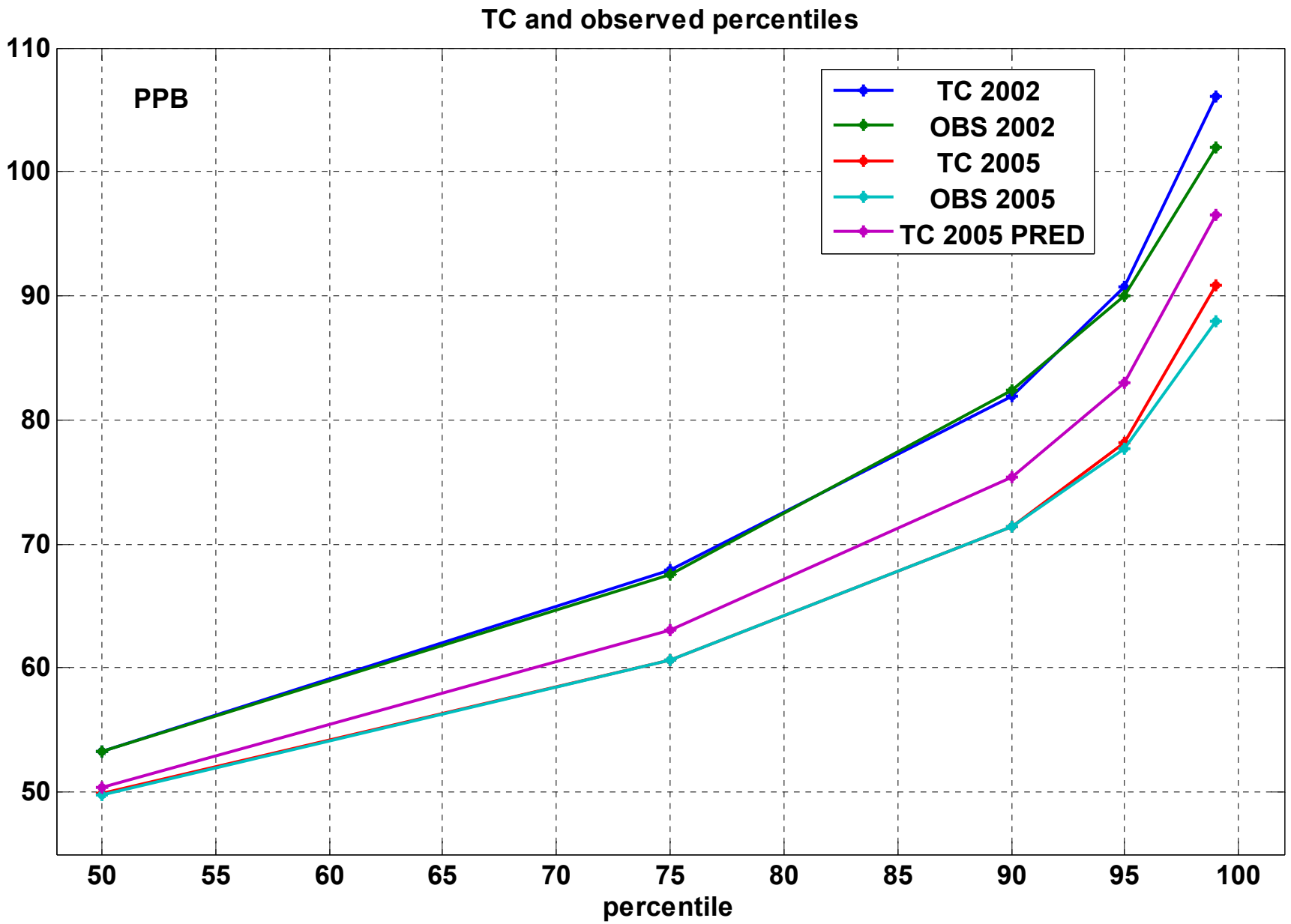

Figure 12. Temporal components and observed mean percentiles for base year 2002 and prediction year 2005. Bias increases for 2005 when forecast from the 2002 base year (purple line).

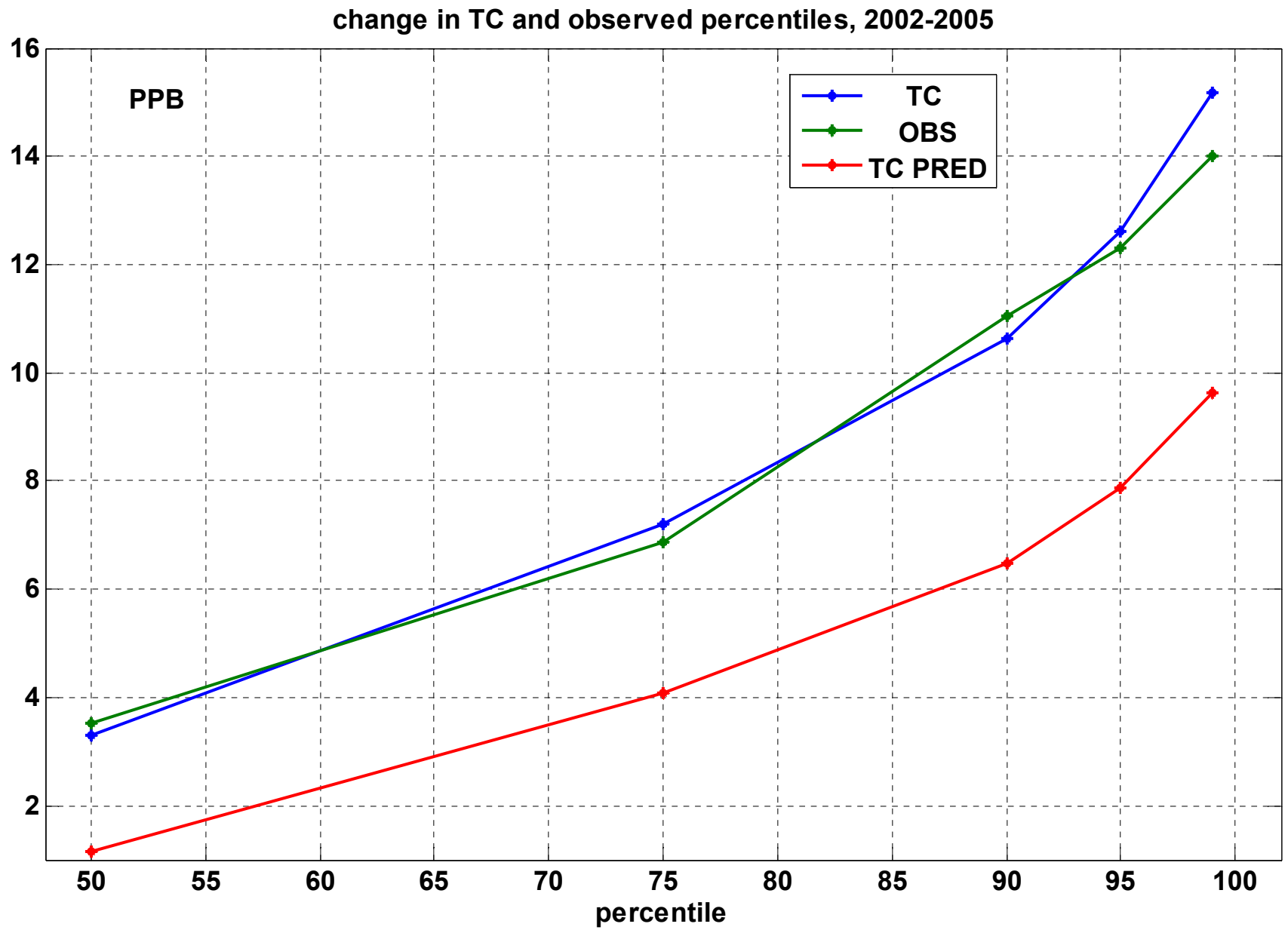

Figure 13. ***Change*** **in temporal components and observed in mean percentiles between the base year 2002 and prediction year 2005. Mean/variance matching** ***change*** **is biased low for all percentiles.**

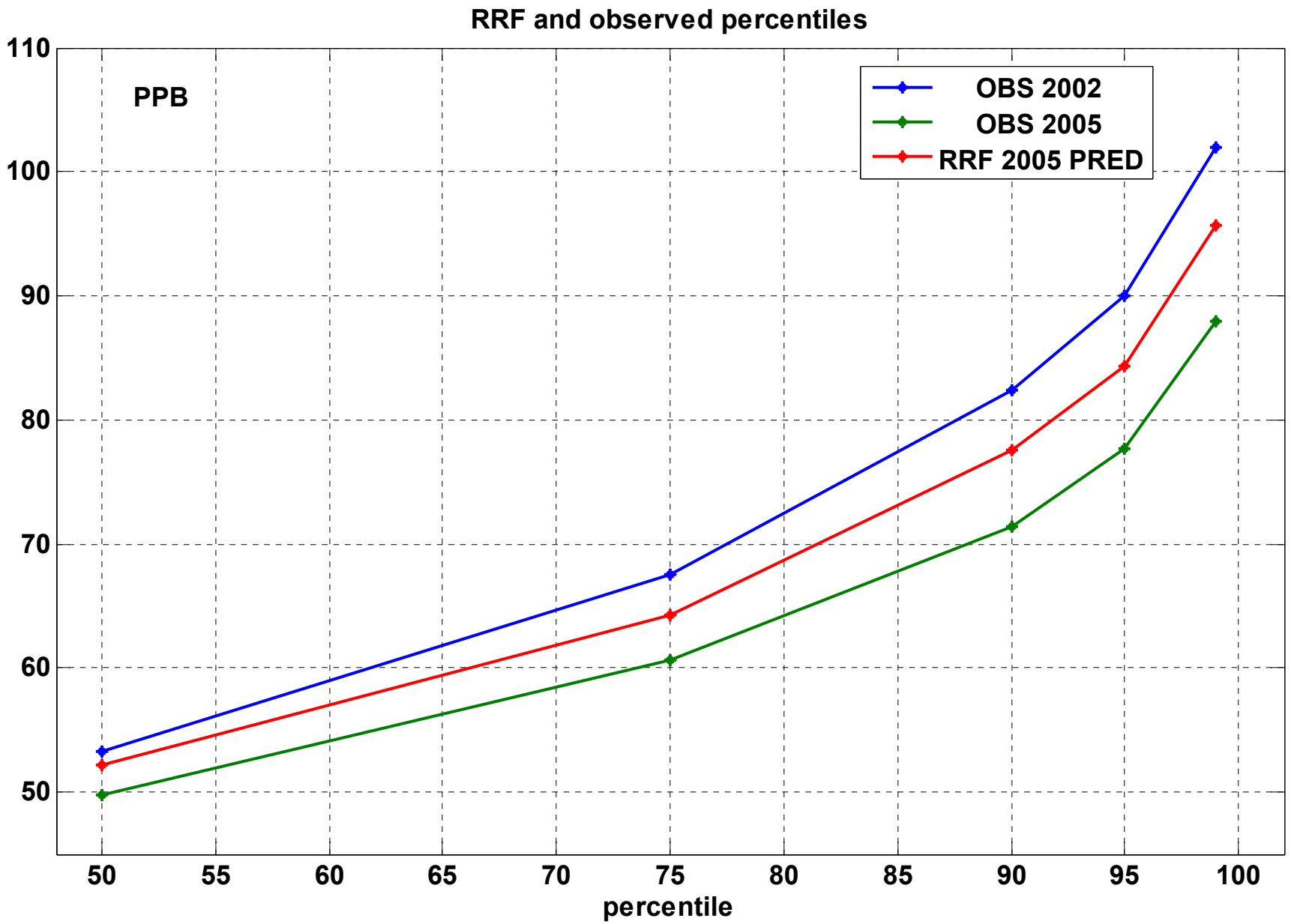

Figure 14. RRF and observed mean percentiles for base year 2002 and prediction year 2005. Bias increases for 2005 when forecast from the 2002 base year (RRF = observed for base year).

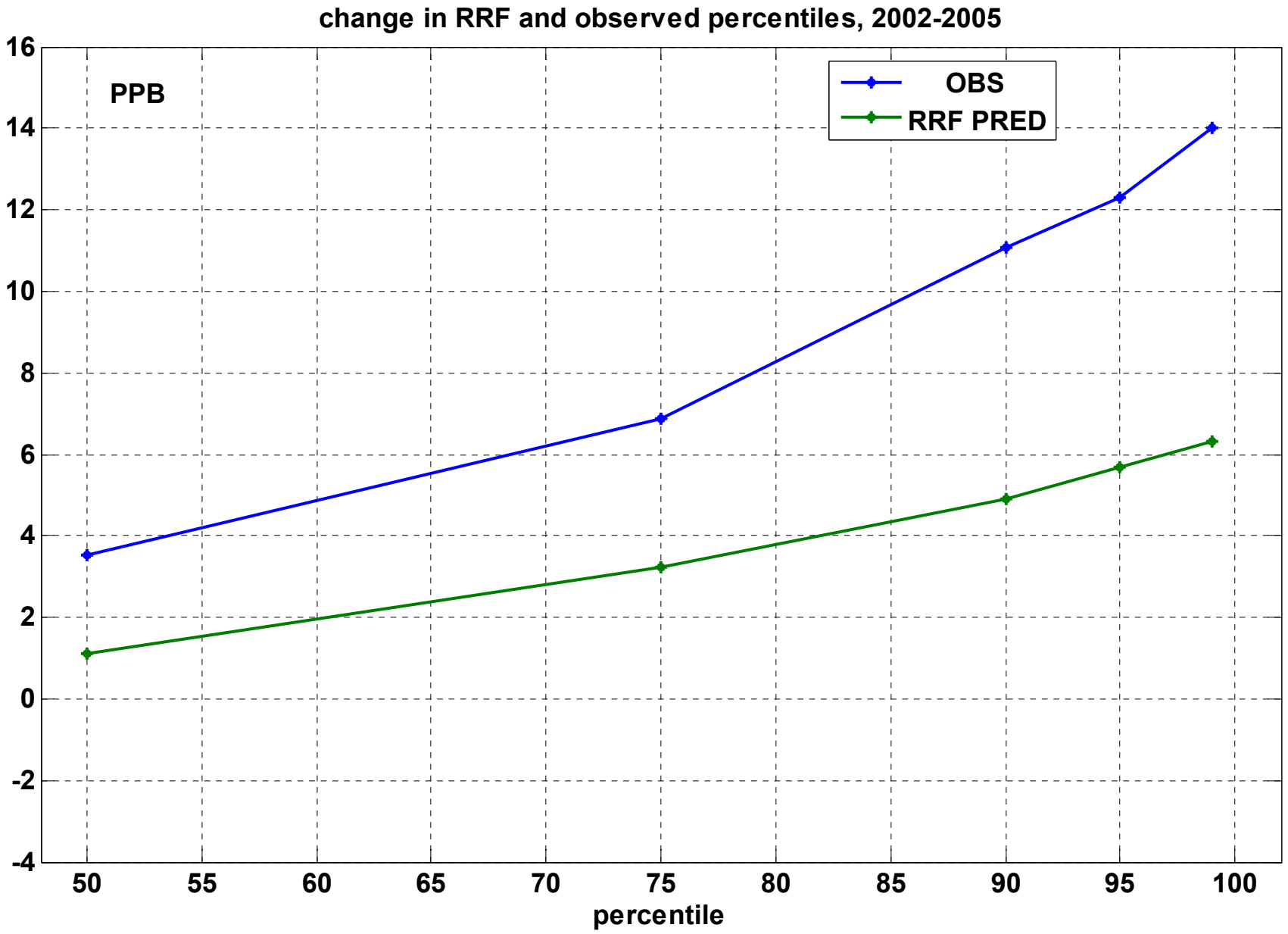

Figure 15. ***Change*** **in RRF and observed mean percentiles between the base year 2002 and prediction year 2005. RRF matching change is biased low for all percentiles.**

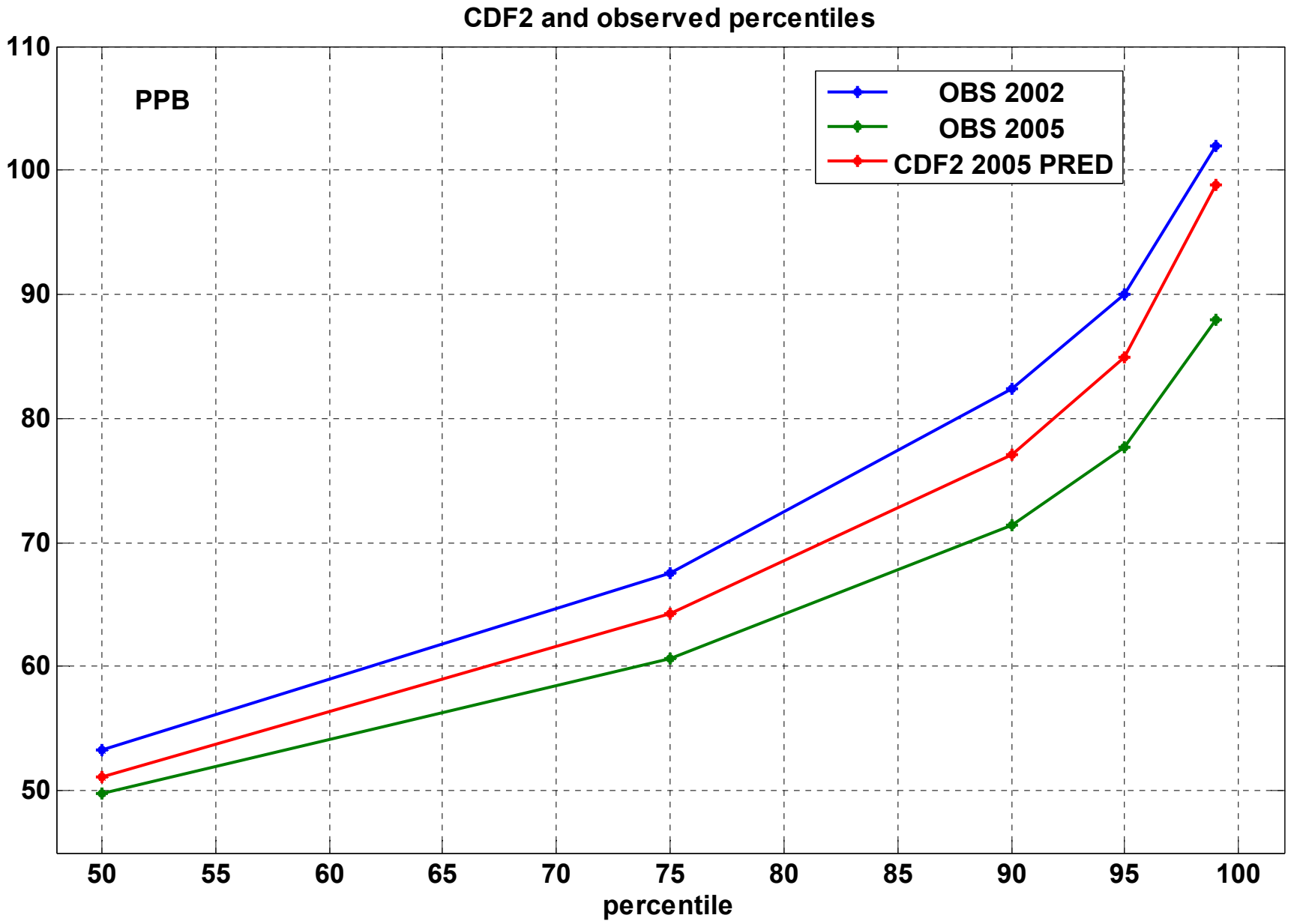

Figure 16. CDF2 and observed mean percentiles for base year 2002 and prediction year 2005. Bias increases for 2005 when forecast from the 2002 base year (RRF = observed for base year).

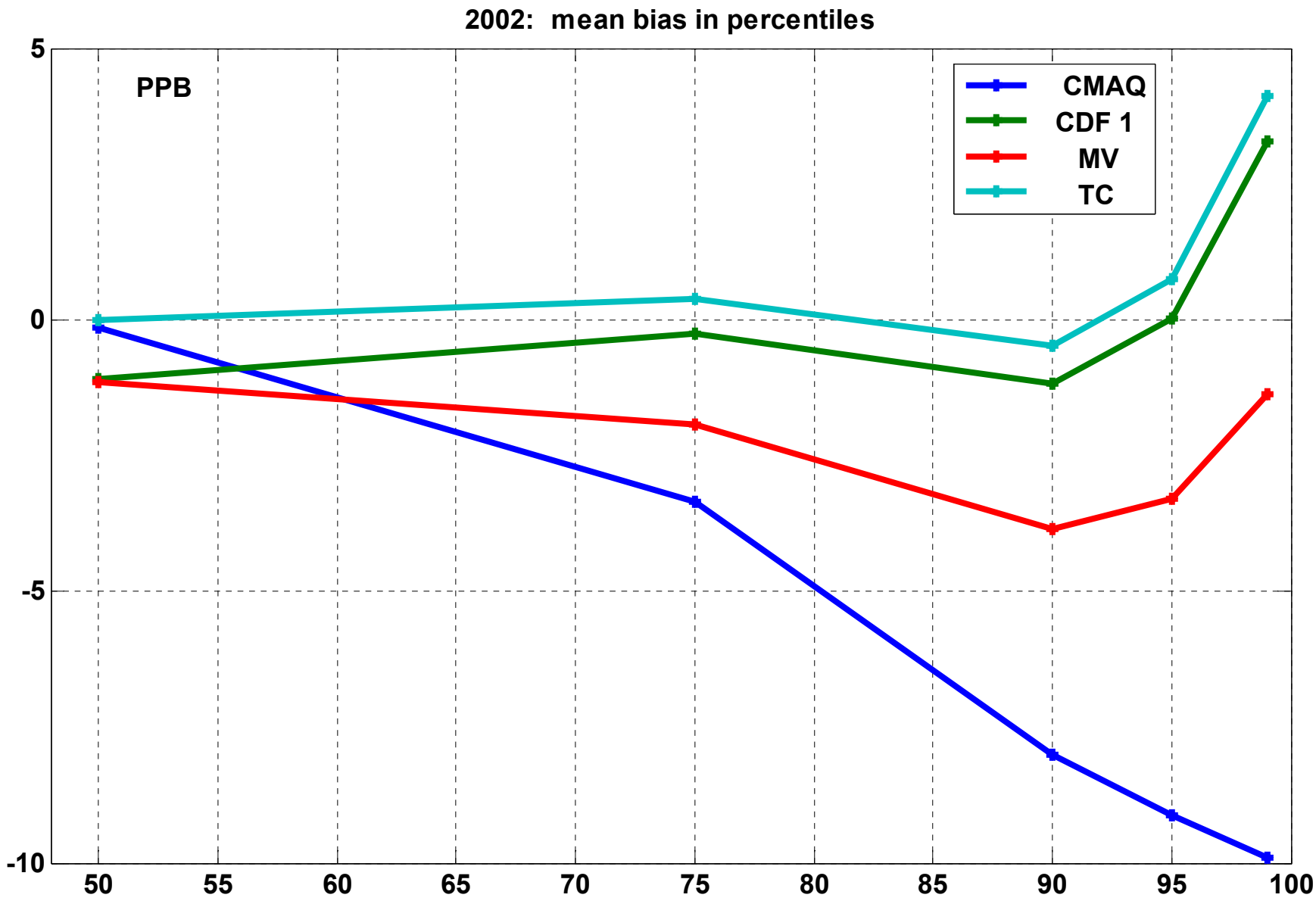

Figure 17. Mean bias vs. percentile for 2002 (method – observed).

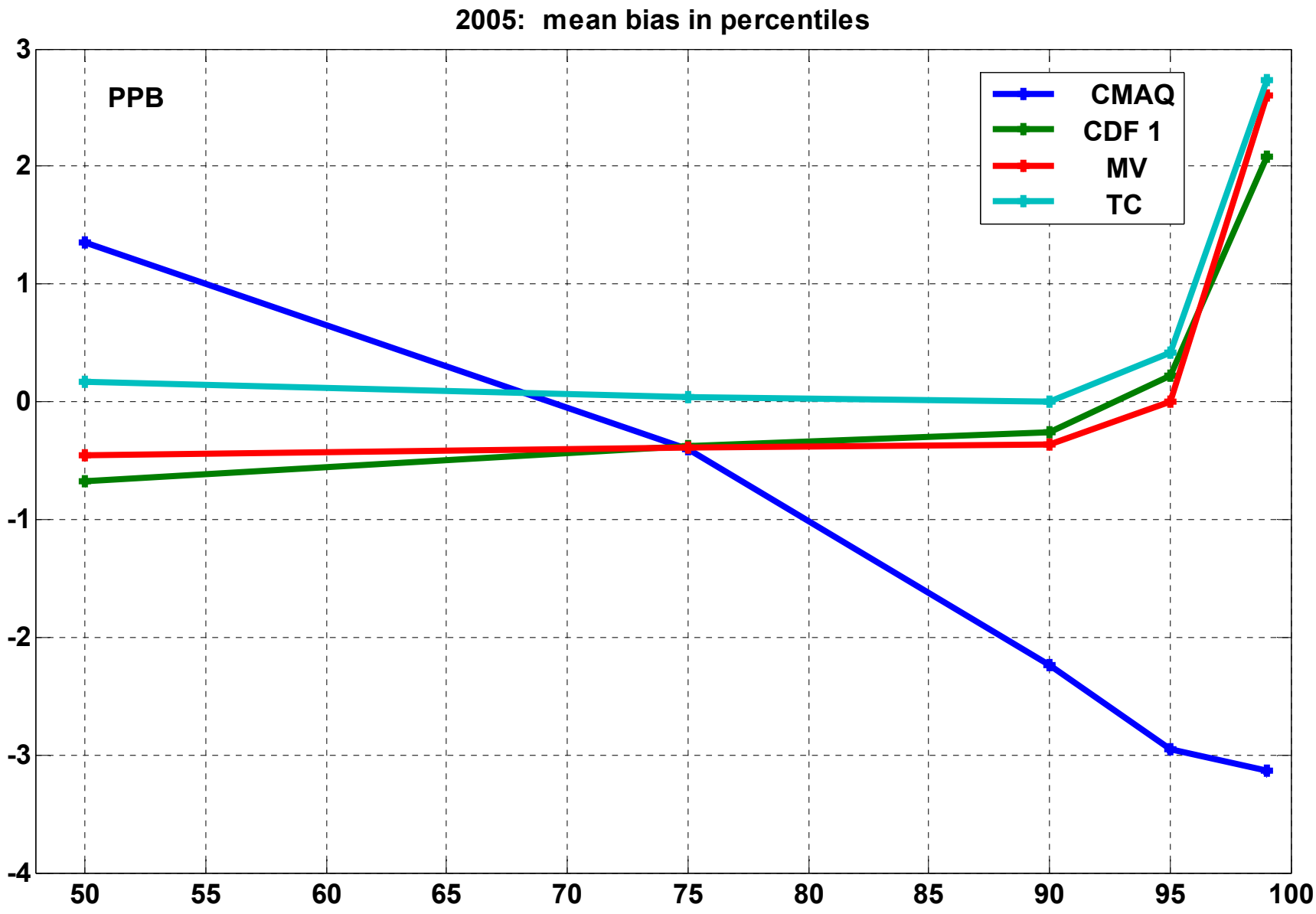

Figure 18. Mean bias vs. percentile for 2005 (method – observed).

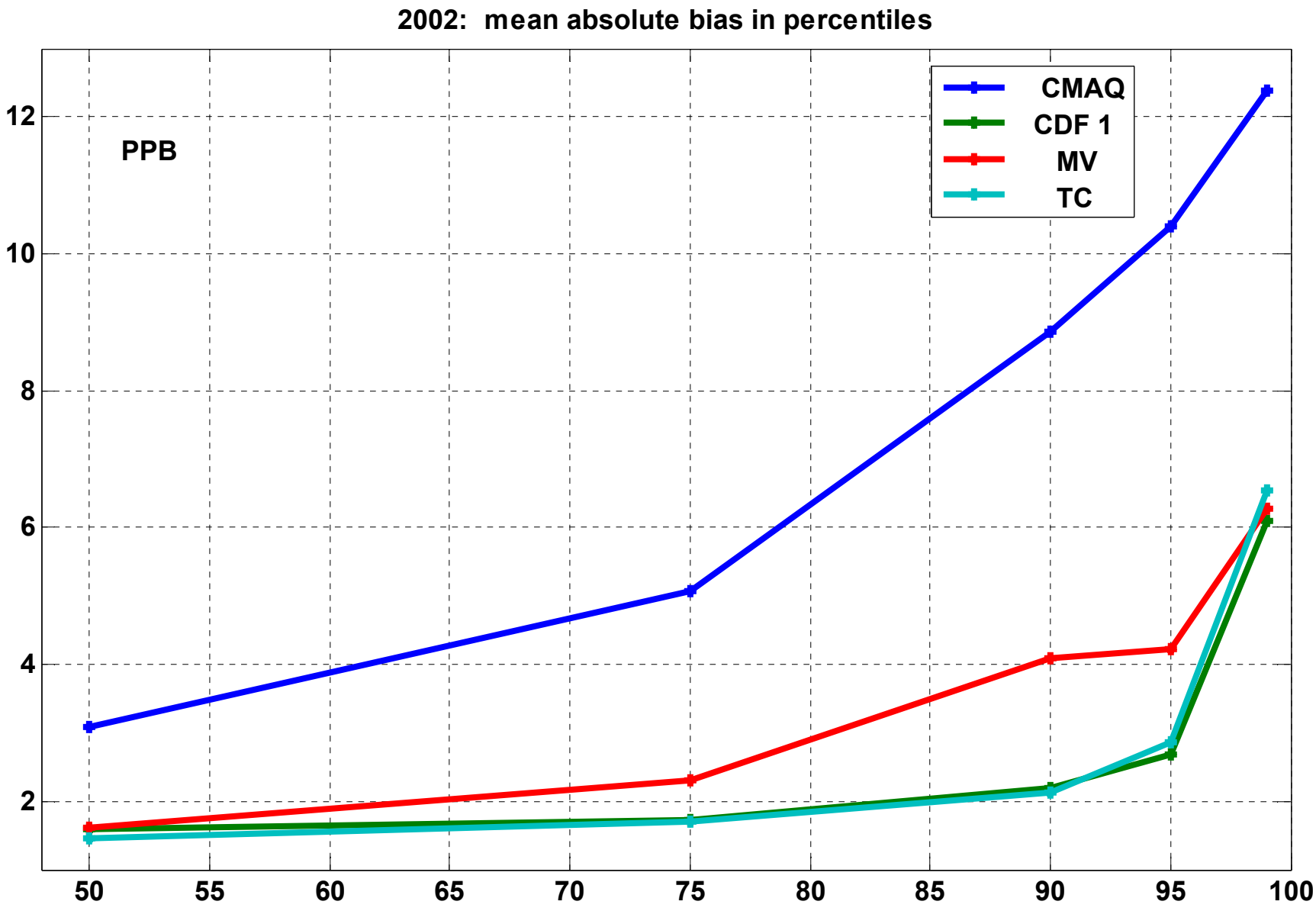

Figure 19. Mean absolute bias vs. percentile for 2002 (method – observed).

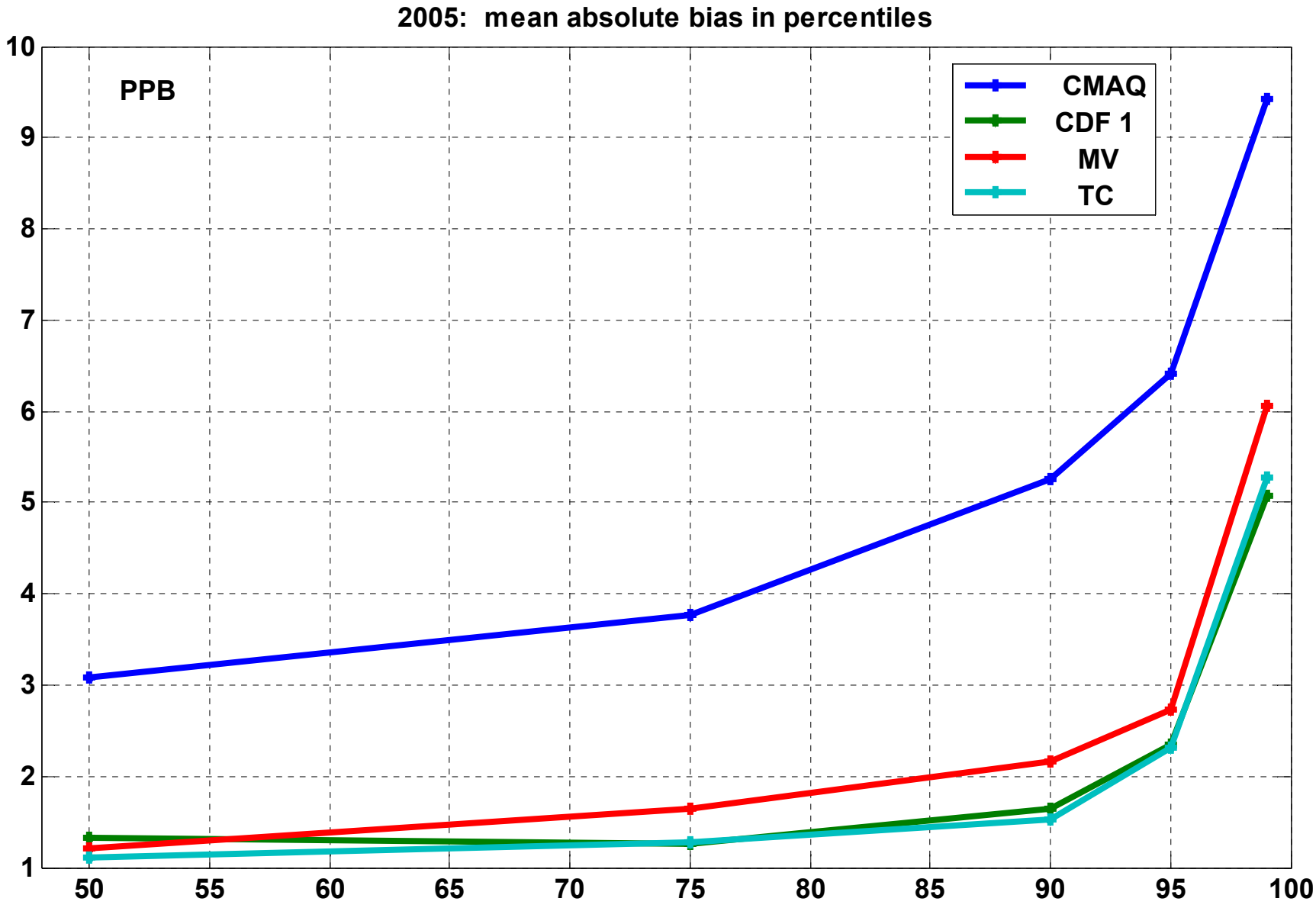

Figure 20. Mean absolute bias vs. percentile for 2005 (method – observed).

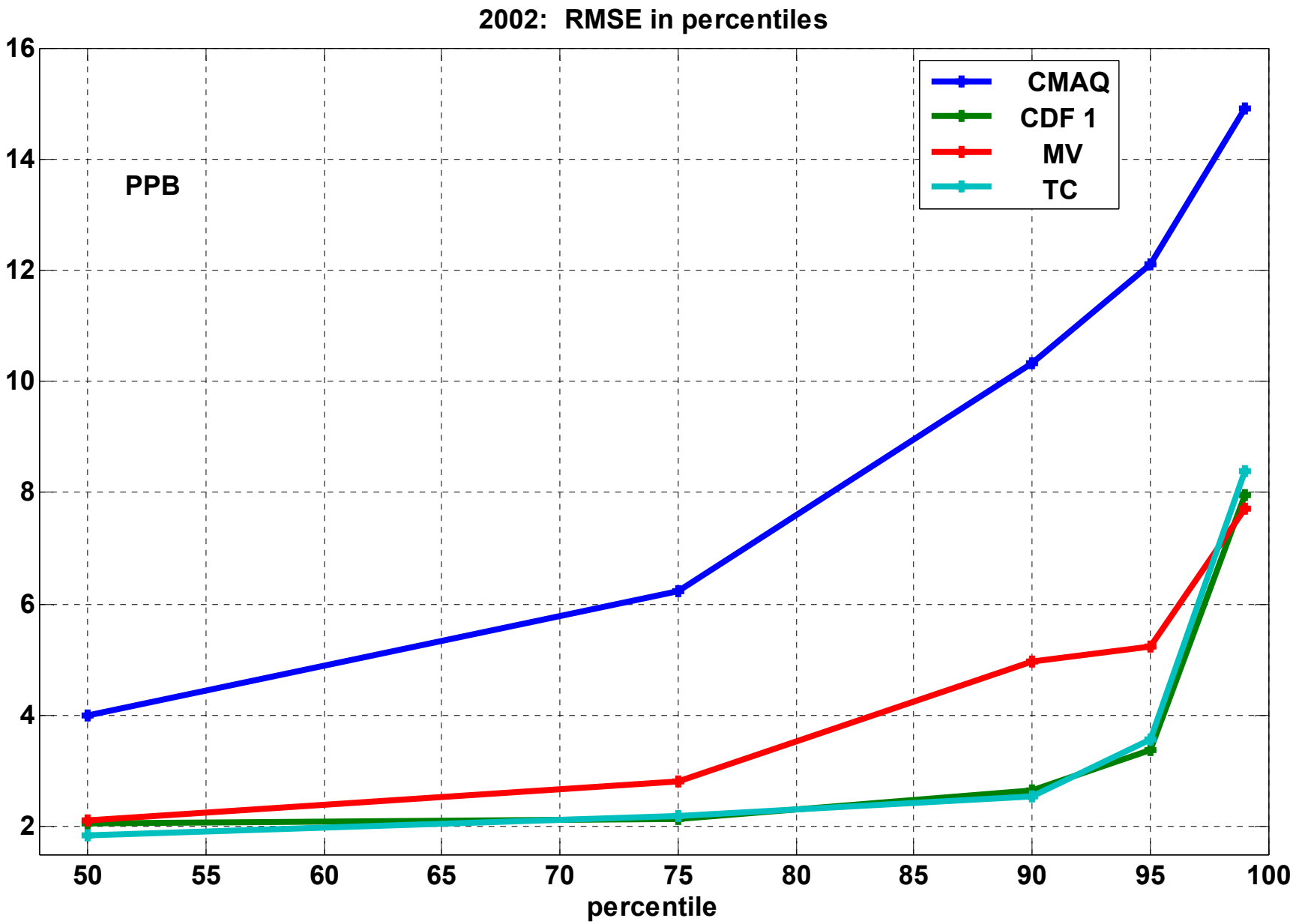

Figure 21. Root mean squared error vs. percentile for 2002 (method – observed).

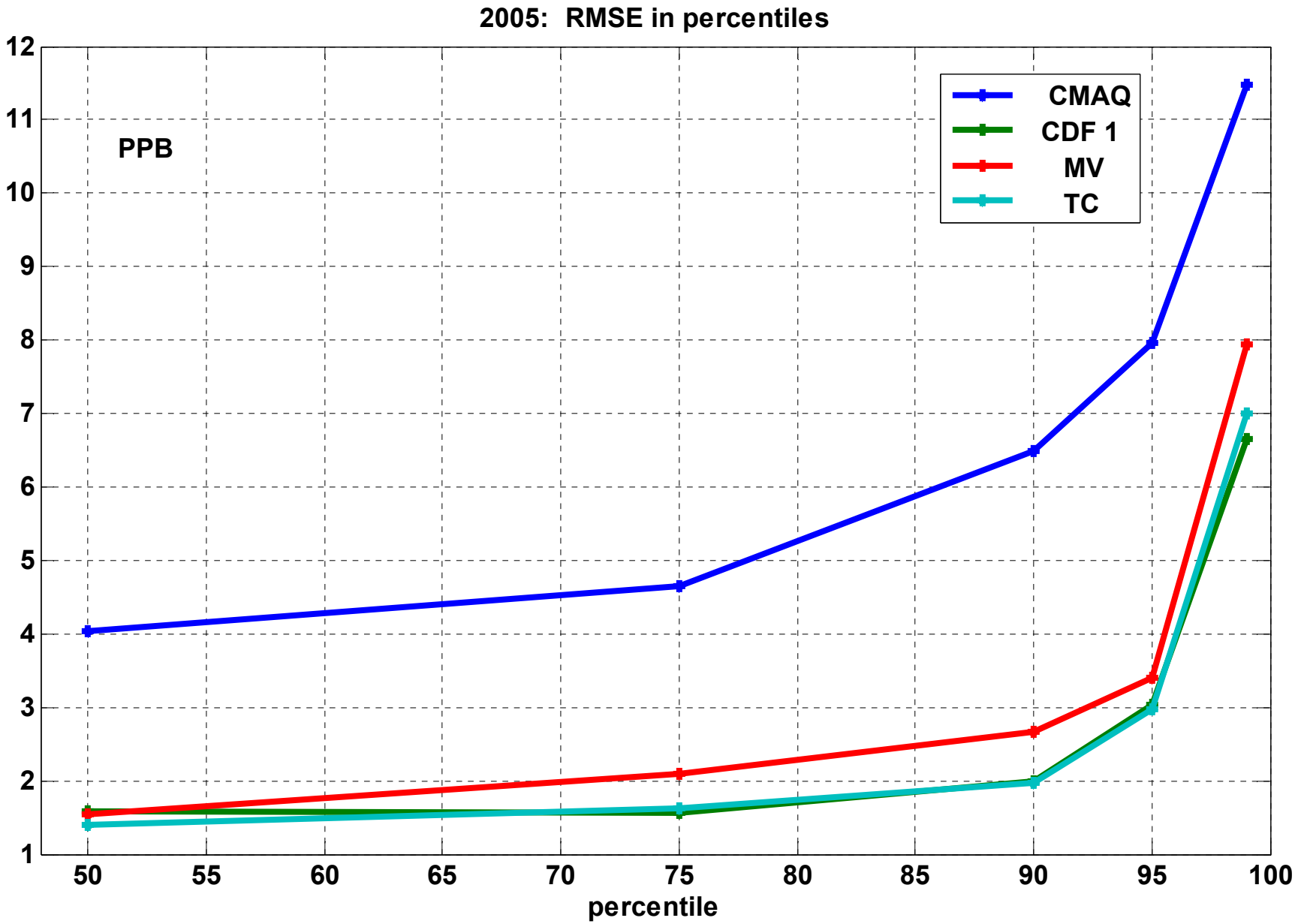

Figure 22. Root mean squared error vs. percentile for 2005 (method – observed).

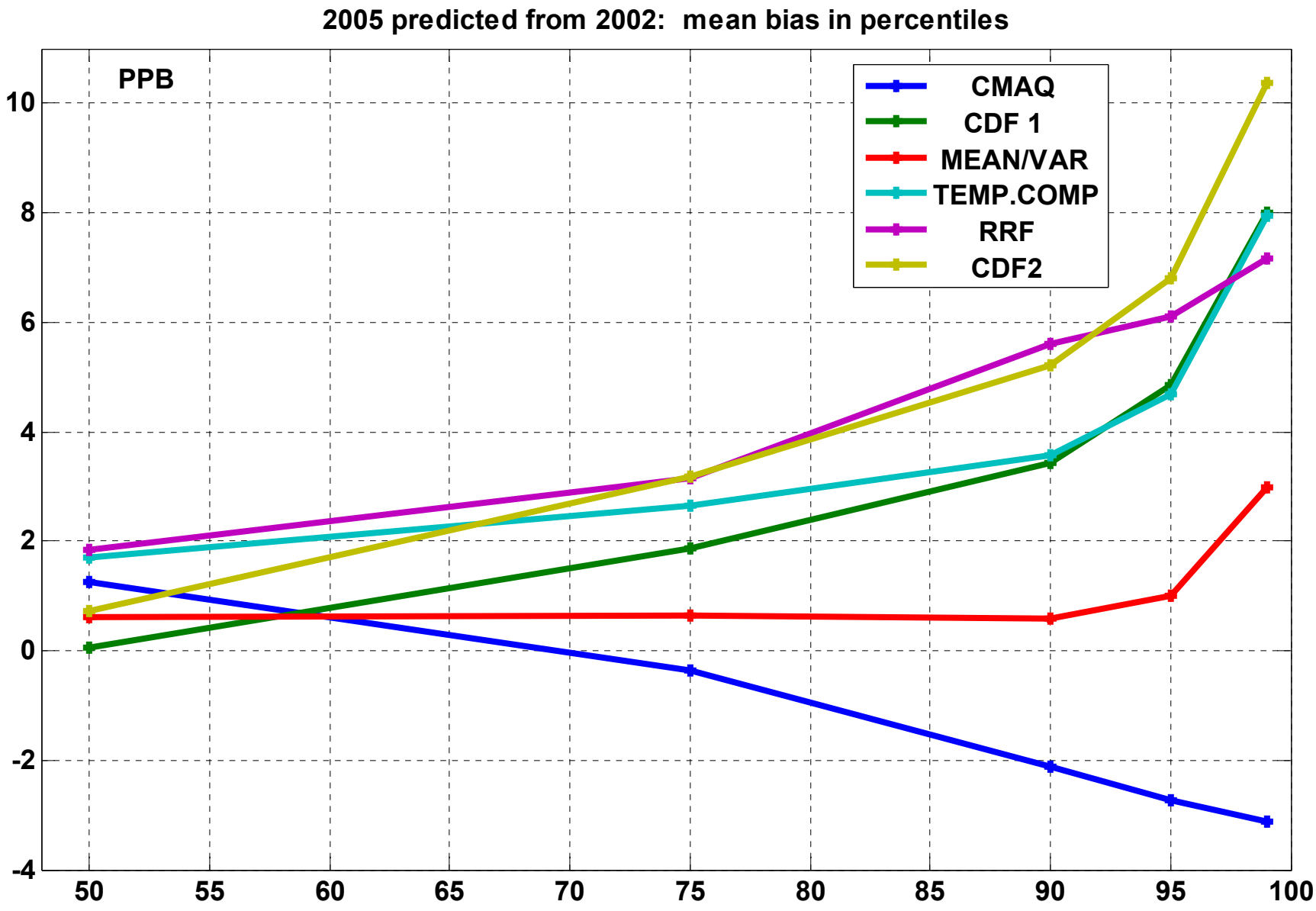

Figure 23. Mean bias vs. percentile for 2005 when predicted from 2002 base year information (method – observed).

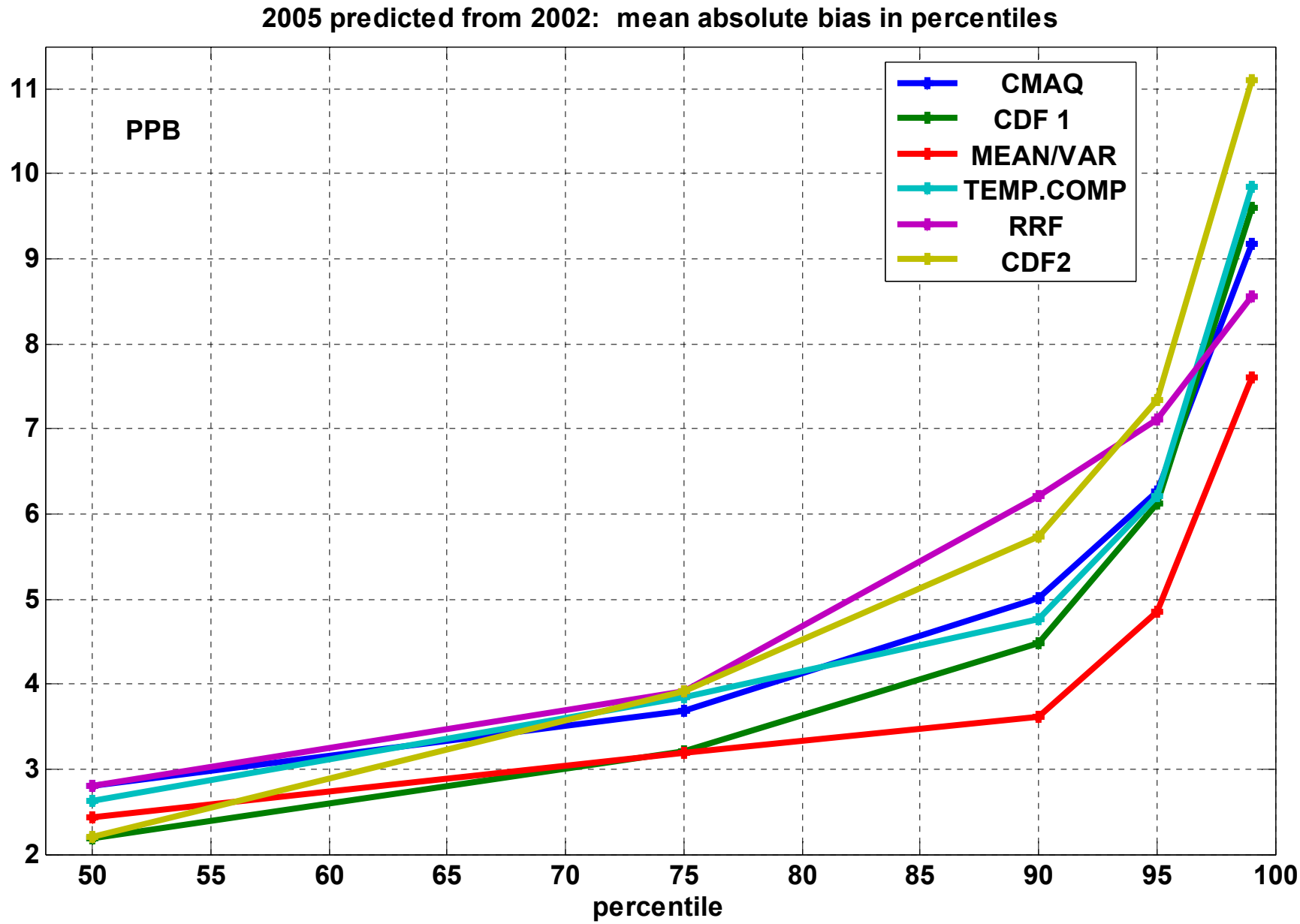

Figure 24. Mean absolute bias vs. percentile for 2005 when predicted from 2002 base year information (method – observed).

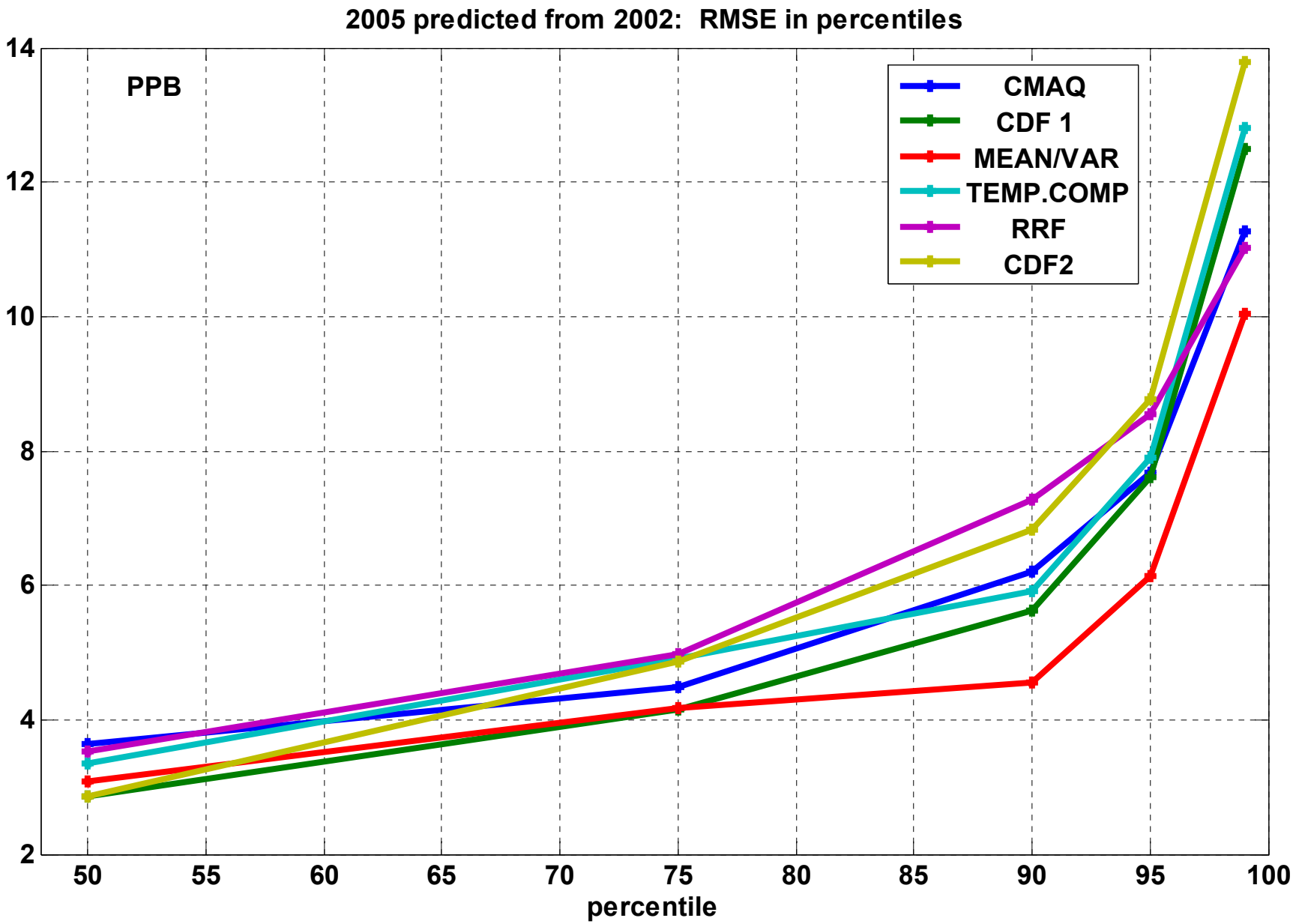

Figure 25. Root mean squared error vs. percentile for 2005 when predicted from 2002 base year information (method – observed).

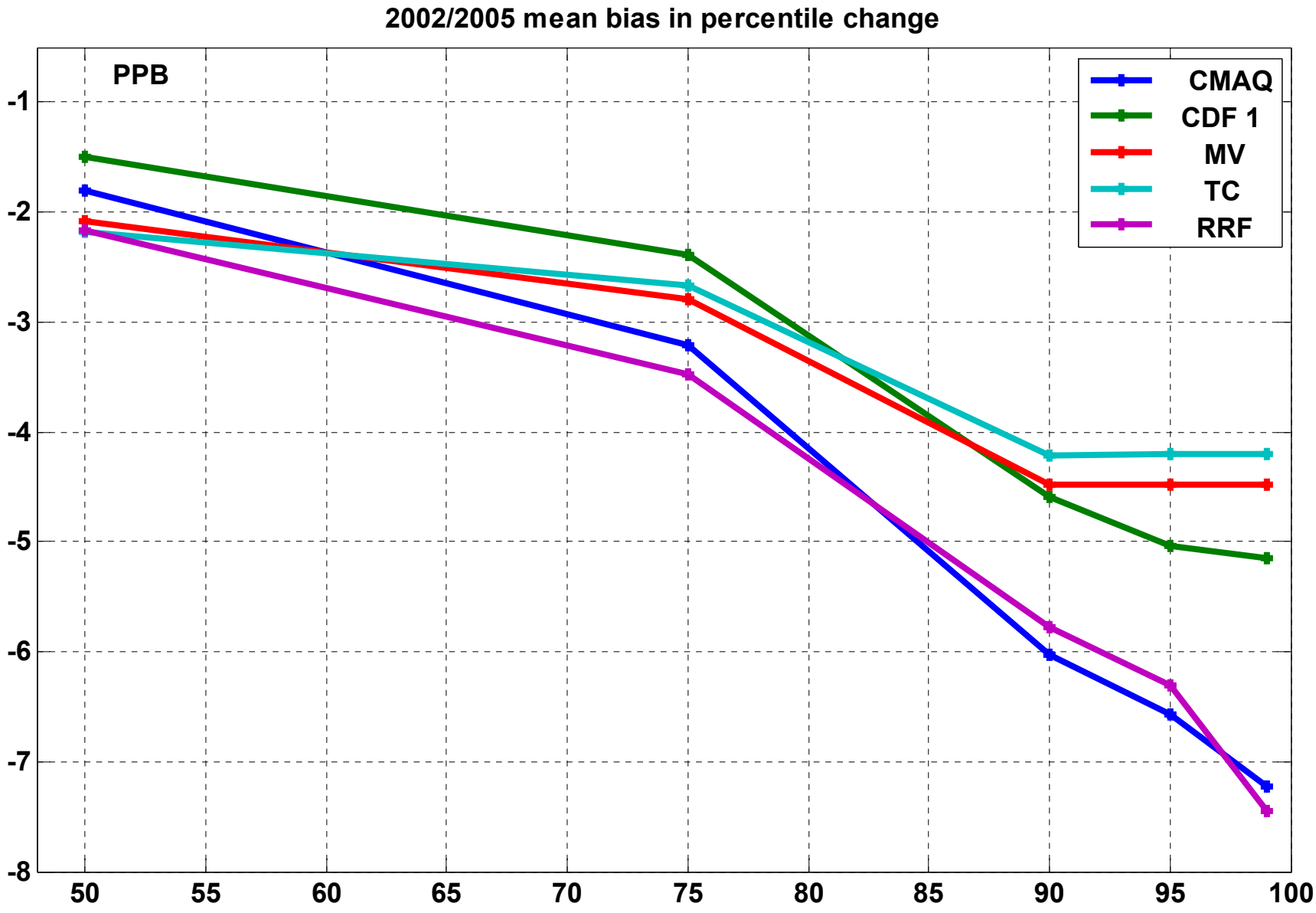

Figure 26. ***Change*** **in percentile: mean bias vs. percentile for 2005 when predicted from 2002 base year information (method – observed).**

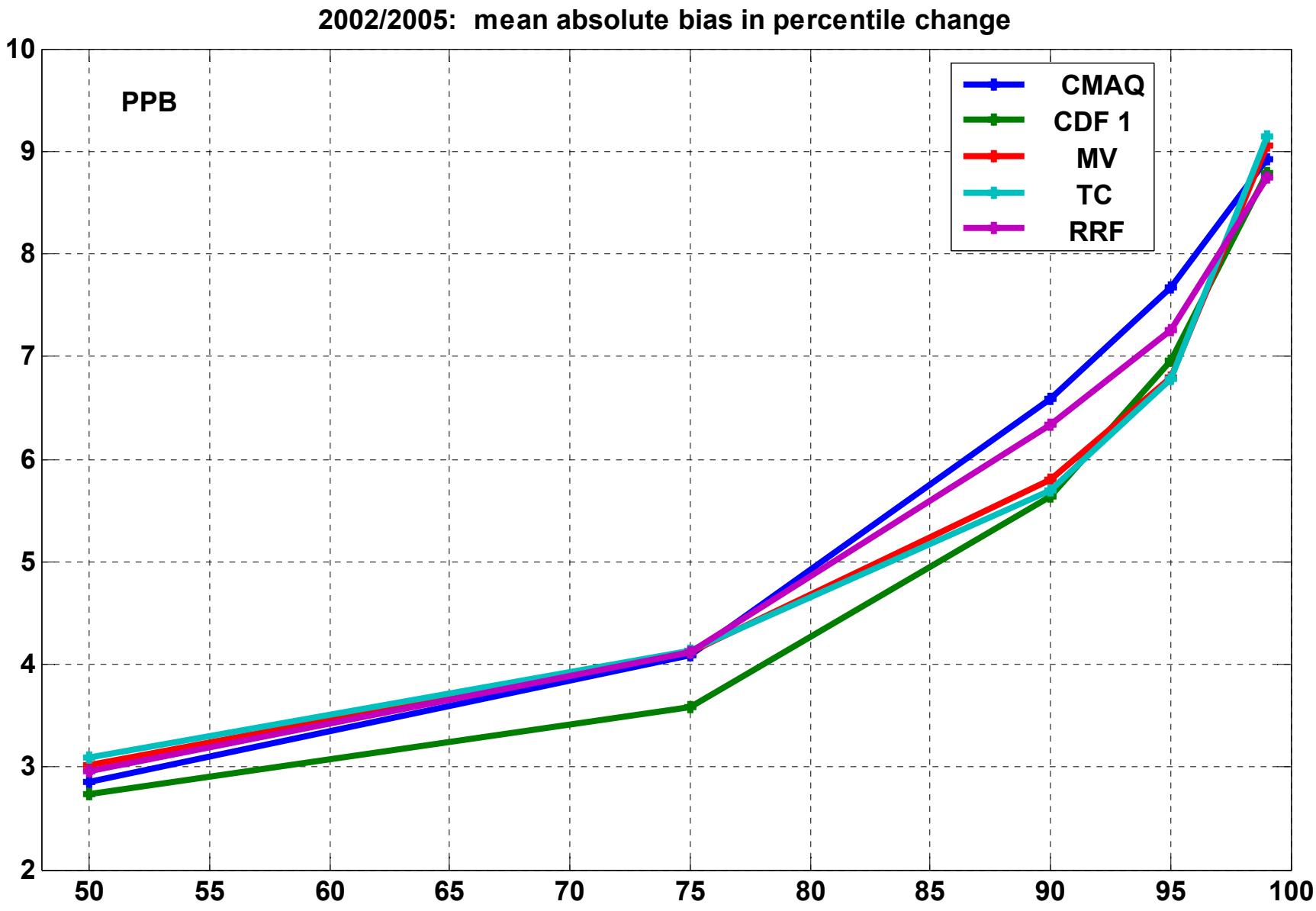

Figure 27. Change in percentile: mean absolute bias vs. percentile for 2005 when predicted from 2002 base year information (method – observed).

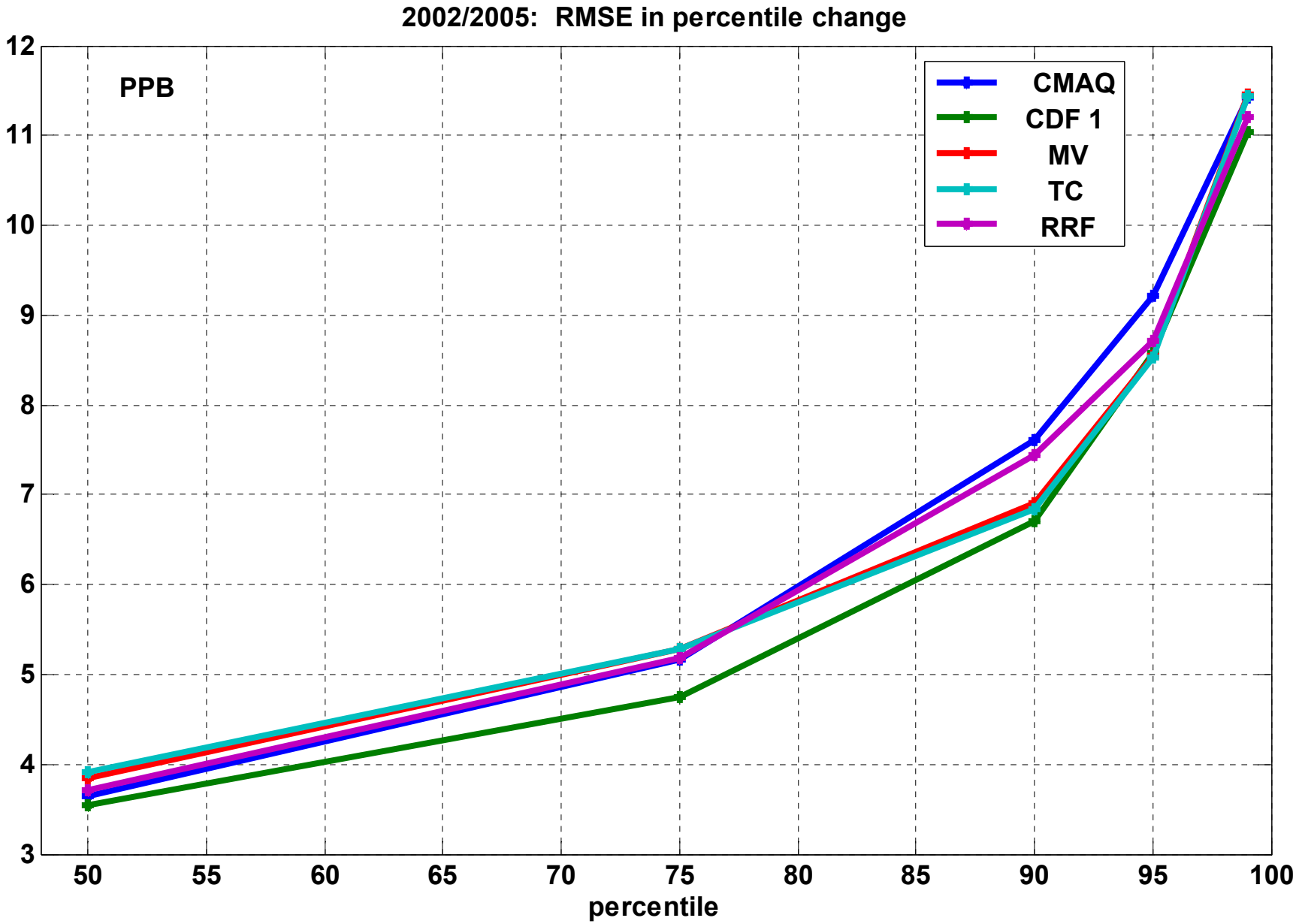

Figure 28. Change in percentile: root mean square error vs. percentile for 2005 when predicted from 2002 base year information (method – observed).