Modeled and observed fine particulate matter reductions from state attainment demonstrations

Daniel S. Cohan and Ran Chen

Department of Civil and Environmental Engineering, Rice University, Houston, TX

ABOUT THE AUTHORS

Daniel Cohan is an Associate Professor in the Department of Civil and Environmental Engineering (CEE) at Rice University. Ran Chen completed a B.S. in Civil Engineering and a B.A. in Policy Studies at Rice in 2013.

ABSTRACT

States rely upon photochemical models to predict the impacts of air quality attainment strategies, but the performance of those predictions is rarely evaluated retrospectively. State implementation plans (SIPs) developed to attain the 1997 U.S. standard for fine particulate matter (PM_{2.5}, denoting particles smaller than 2.5 microns in diameter) by 2009 provide the first opportunity to assess modeled predictions of PM_{2.5} reductions at the state level. The SIPs were the first to rely upon a speciated modeled attainment test methodology recommended by the U.S. Environmental Protection Agency to predict PM_{2.5} concentrations and attainment status. Of the 23 eastern U.S. regions considered here, all but one achieved the 15 µg/m³ standard by 2009, and the other achieved it the following year, with downward trends sustained in subsequent years. The attainment tests predicted 2009 PM_{2.5} design values at individual monitors with a mean bias of 0.38 µg/m³ and mean error of 0.68 µg/m³, and were 95% accurate in predicting whether a monitor would achieve the standard. All of the errors were false alarms, in which the monitor observed attainment after a modeled prediction of an exceedance; in these cases, the states used

weight of evidence determinations to argue that attainment was likely. Overall, $PM_{2.5}$ concentrations at monitors in the SIP regions declined by $2.6 \mu\text{g}/\text{m}^3$ from 2000-2004 to 2007-2009, compared with $1.6 \mu\text{g}/\text{m}^3$ in eastern U.S. regions originally designated as attainment. Air quality improvements tended to be largest at monitors that were initially the most polluted.

IMPLICATIONS

As states prepare to develop plans for attaining a more stringent standard for fine particulate matter, this retrospective analysis documents substantial and sustained air quality improvements achieved under the previous standard. Significantly larger air quality improvements in regions initially designated nonattainment of the 1997 standard indicate that this status prompted heightened control efforts. The speciated modeled attainment test is found to be accurate and slightly conservative in predicting particulate concentrations for the cases considered here, providing confidence for its use in upcoming attainment plans.

INTRODUCTION

The Clean Air Act authorizes the U.S. Environmental Protection Agency (EPA) to set National Ambient Air Quality Standards (NAAQS) for six criteria pollutants, yet it is the states' responsibility to attain those standards. Attainment demonstrations, submitted by states to US EPA as revisions to state implementation plans (SIPs), set forth control measures for reducing emissions as well as modeling to demonstrate that the measures will be sufficient to attain the NAAQS by the targeted attainment date. The effectiveness of state-level attainment efforts is a

key consideration in overall assessments of air quality management in the United States

(Anderson, 2006; Fine and Owen, 2005; NRC, 2004).

Particulate matter was historically regulated based on total suspended particulate or other broad measures of particle mass. Epidemiological evidence showing particles smaller than 2.5 microns in diameter ($PM_{2.5}$) to be especially damaging to human health (Dockery et al., 1993; Pope and Dockery, 2006) prompted EPA to develop standards specifically limiting mass concentrations in this size range. Though promulgated in 1997, the standards were not implemented until 2002 following lengthy litigation (Bachmann, 2007). The 1997 standards limited $PM_{2.5}$ to $15 \mu\text{g}/\text{m}^3$ on an annual basis (subsequently lowered to $12 \mu\text{g}/\text{m}^3$ in 2012) and $65 \mu\text{g}/\text{m}^3$ on a 98th percentile daily basis (lowered to $35 \mu\text{g}/\text{m}^3$ in 2006). Each standard is evaluated based on 3-year averages known as design values (DVs). EPA designated areas as nonattainment for the annual and/or 24-hour $PM_{2.5}$ NAAQS based on air quality monitoring data from 2001 to 2003. Most nonattainment areas were required to achieve attainment based on their 2009 DV (i.e., 2007-2009 data), and thus their attainment demonstrations can now be evaluated by ambient observations.

The attainment demonstrations that aimed to attain the $PM_{2.5}$ standard by 2009 were the first to use EPA's recommended speciated modeled attainment test (US-EPA, 2007) to predict whether $PM_{2.5}$ attainment would be achieved. Like the modeled attainment test for ozone, this approach simulates the ratio between pollution under target year and base year emission inventories for identical historical meteorological conditions. That ratio, known as a response factor (RRF), is then multiplied by observed base period pollution levels to predict target year DVs. However, since $PM_{2.5}$ is composed of multiple chemical compounds, its test requires that $PM_{2.5}$ first be

speciated into seven components, with separate RRFs computed and applied to each. Since the majority of regulatory PM_{2.5} monitors do not speciate composition, speciation is typically inferred from nearby monitors that provide those measurements.

Very few studies have reviewed the content of attainment demonstrations or evaluated their performance in improving air quality and achieving ambient standards. Several studies have documented substantial reductions in PM observed in various portions of the U.S. over the past decade (Blanchard *et al.*, 2013; Coen *et al.*, 2013; Hand *et al.*, 2013) but did not link those improvements to state-level efforts. Fine and Owen (2005) reviewed the SIP process and the roles of modeling and public participation. A comprehensive review of air quality management in the United States (NRC, 2004) assessed the role of attainment demonstrations and the relative contributions of state and federal measures.

Two studies have assessed the modeling approach to predicting attainment of ozone standards (Jones *et al.*, 2005; Sistla *et al.*, 2004). Pegues *et al.* (2012) evaluated the performance of SIPs in achieving attainment of the 1997 ozone standard, and compared rates of air quality improvement in regions that were and were not required to develop an attainment demonstration. However, the methodology of the speciated modeled attainment test for PM_{2.5} and its performance in predictions of pollutant levels have not been evaluated previously in the peer-reviewed literature.

Here, we build from the approach of Pegues *et al.* to assess attainment demonstrations developed to meet EPA's 1997 annual PM_{2.5} standard by 2009. Regulatory model predictions of future air quality are compared to observations to assess their performance in predicting future attainment and air quality improvements. Trends in reducing PM_{2.5} are compared for regions that did and

did not submit an attainment demonstration. This provides circumstantial evidence of whether nonattainment designation and the development of an attainment plan prompted accelerated efforts to improve air quality.

METHODS

Attainment Demonstrations Considered

Online searches and inquiries to state air quality agencies and regional EPA offices were conducted to obtain all available attainment demonstrations that sought attainment of the 1997 annual PM_{2.5} standard by 2009 (Table 1). The analysis covers 23 nonattainment regions, some of which span multiple states that each developed attainment demonstrations. When a nonattainment region is under the jurisdiction of more than one state, attainment demonstrations from all of these states are used, with modeling results for each monitor taken from the corresponding state to the extent possible. In some cases (e.g., the New Jersey SIP for Philadelphia), a multi-state region necessitates submittal of a document even by a state whose monitors within that region meet the standard. Although Table 1 separates results by state in multi-state regions, it is the highest DV across each region that ultimately determines attainment status.

Some regions were excluded in this study, as they achieved the standard early and thus did not submit full attainment demonstrations. We were unable to obtain attainment demonstrations from Illinois, and thus use the Missouri attainment demonstration for all St. Louis monitors. Given our

focus on the 2009 attainment year, regions that requested an extension to 2014 (Liberty-Clairton, PA, and Los Angeles-South Coast and San Joaquin Valley, CA) were also excluded. However, the attainment demonstration for Birmingham, AL, is included, since it initially sought attainment by 2009 but requested a 2012 target after modeling showed more time was needed; the region was subsequently redesignated to attainment status after its 2010 observed DV attained the standard.

The attainment demonstrations considered here were submitted between 2008 and 2010 (Table 1), due to the delays in implementing the $PM_{2.5}$ NAAQS following litigation. Thus, observations counting toward the targeted 2009 DV were already underway when the attainment demonstrations were submitted. Nevertheless, the modeling and emission inventory development reported in the attainment demonstrations were undertaken years earlier and thus can still be viewed as predictions of year 2009 conditions.

The EPA guidance document for attainment demonstrations recommends the speciated modeled attainment test as the primary predictor of whether an attainment plan is sufficient to attain the $PM_{2.5}$ NAAQS by the targeted date (US-EPA, 2007). The test is based on the application of a photochemical model to predict trends in pollutant levels between the base and targeted time periods. All of the attainment demonstrations considered here adopted the recommended test and used either the Community Multiscale Air Quality (CMAQ) model (Byun and Schere, 2006) or the Comprehensive Air Quality Model with Extension (CAMx) (ENVIRON, 2010) (Table 1). CAMx was used primarily in Midwest states which shared modeling development through the Lake Michigan Air Directors Consortium, whereas CMAQ was used in most other states.

The attainment demonstrations used future year (i.e., 2009) and base year (i.e., 2002) emission inventories as inputs to simulations with identical historical meteorology (typically from 2002) to compute the ratios (RRFs) between future and base year concentrations for each speciated component of PM_{2.5} at each monitor. The RRFs were then multiplied by the observed base design value (DVB) concentrations on a speciated basis to predict the future design value (DVF) of PM_{2.5} at each monitor. To demonstrate attainment for a region, all monitors must be projected to meet the NAAQS limit by the targeted year (in this case, 15 µg/m³ by 2009). The guidance document recommends that DVBs be calculated by averaging three successive 3-year averages of PM_{2.5} observations; thus, the 2002, 2003, and 2004 DVs (i.e., 2000-2004 data) were used to establish DVBs in the attainment demonstrations.

Predictions of 2009 DVs were taken from the modeled attainment demonstrations, and compared to observations at the monitor with the corresponding EPA Air Quality System (AQS) ID number. Modeling results were considered only for monitors within the state issuing the attainment demonstration, with the exceptions that the Missouri demonstration was used for monitors in the Illinois portion of St. Louis, and a joint Washington, DC, region demonstration was used for monitors in the Maryland and Virginia portions of the Washington region.

Beyond the modeled attainment demonstrations, we also qualitatively considered supplemental weight of evidence determinations provided in the documents where applicable. The EPA guidance document encourages states to use supplemental analyses such as additional modeling and emission and air quality trend analyses to complement the primary modeled attainment test,

especially if the primary modeled results are near the NAAQS threshold (a maximum DVF between 14.5 to 15.5 $\mu\text{g}/\text{m}^3$).

Observational Data

Observational data were obtained from a July 26, 2012, data query of the AQS database. A subsequent data query on July 10, 2013, was used to extend the data to 2012 DVs. $\text{PM}_{2.5}$ annual DVs represent the 3-year average of the annual mean PM concentrations at each monitor. Following the protocol of the SIP attainment demonstrations, baseline conditions were defined based on the average of three DVs ending in 2002, 2003, and 2004. This results in a weighted average of observational data from 2000-2004, with greatest weight on the year 2002.

Performance of each attainment demonstration was assessed based on the 2009 DV (i.e., observational data from 2007-2009), corresponding to the targeted attainment year. AQS defines data completeness based on a monitor recording valid measurements for at least 75% of days in each quarter, or a minimum of 11 days for monitors exceeding 12.0 $\mu\text{g}/\text{m}^3$. Monitors with incomplete data for any year 2007-2009 or identified in AQS as non-representative micro-scale sites are excluded from the analysis. Monitors with complete data for 2007-2009, but with incomplete data for the 2000-2004 base period, were included in the assessment of model predictions of 2009 DVs, but excluded from assessments of $\text{PM}_{2.5}$ reductions across these periods.

RESULTS

Predicted vs. Observed Attainment Status

We first compare SIP model predictions of $PM_{2.5}$ with data observed at monitors to evaluate their accuracy in predicting 2009 DVs and attainment (Figure 1). The model predictions are taken from the 2009 DVFs reported by modeled attainment tests. Observed 2009 DVs in Figure 1 are based on observations of $PM_{2.5}$ from 2007-2009 at each monitor. The vertical line across $15 \mu\text{g}/\text{m}^3$ in Figure 1 divides monitors that detect attainment from those that did not, whereas the horizontal line divides predicted attainment from predicted nonattainment.

All but one of the 117 monitors considered in Figure 1 did in fact observe a 2009 DV of $15 \mu\text{g}/\text{m}^3$ or less, with a single Birmingham monitor observing $15.1 \mu\text{g}/\text{m}^3$ and one Cincinnati monitor exactly meeting the limit. Subsequent AQS data obtained after our initial analysis (new data query July 10, 2013) show that 2012 DVs (based on 2010-2012 observations) at all of the Figure 1 monitors, including the highest monitors in Birmingham (13.0) and Cincinnati (13.4), were below $15 \mu\text{g}/\text{m}^3$. The standard modeled attainment tests correctly predicted attainment for 110 monitors, and nonattainment at a Birmingham monitor, representing a 94.8% success rate for predictions. “False alarms,” in which the modeled attainment test predicted an exceedance but attainment was observed, occurred at monitors in Atlanta, Birmingham, Cleveland, Detroit (2 monitors), and St. Louis. No false negatives were observed. Thus, overall, the standard modeled attainment test was highly accurate in predicting attainment status, and tended to be conservative rather than optimistic when predictions were errant.

In cases where nonattainment was predicted by the modeled attainment test, states adopted alternate approaches to complying with the regulation. The predictions of exceedances at the Birmingham monitors led Alabama to seek attainment by 2012. However, actual PM_{2.5} levels in 2009 were lower than predicted, and attainment was subsequently achieved in 2010.

In the other states that modeled nonattainment, the attainment demonstrations made weight of evidence arguments to claim that attainment would be achieved by 2009 despite the results of the test. For example, the Michigan attainment demonstration for Detroit noted downward trends in observations and additional control measures near the most polluted monitors that it expected would lead to attainment. Similarly, the Ohio attainment demonstration argued that localized controls not fully captured by the air quality model would lead to attainment. The weight of evidence arguments were subsequently borne out by observations, as attainment was achieved in each of these regions by 2009.

PM_{2.5} continued to decline after the 2009 target year, with DVs for the attainment demonstration regions (defined by the single monitor with highest DV, even for interstate regions) falling from an average 13.7 µg/m³ in 2009 to 12.2 µg/m³ in 2012 (Table 1). The number of Table 1 regions exceeding the new 12 µg/m³ standard fell from 21 to 13. Thus, the air quality improvements achieved by 2009 were not merely short-term or meteorology-induced occurrences, but instead part of a sustained trend of declining PM_{2.5} that is bringing most of the eastern U.S. in or near attainment of the new standard.

Predicted vs. Observed Design Values

The modeled attainment test was also slightly conservative on a numerical basis in predicting design values. The models overpredicted actual 2009 DVs by an average of $0.38 \mu\text{g}/\text{m}^3$, with mean error of $0.68 \mu\text{g}/\text{m}^3$. Overpredictions occurred at more than twice as many monitors (81; points above the solid diagonal line in Figure 1) as underpredictions (32) or exact matches (4). At only four monitors was the underprediction larger than $1 \mu\text{g}/\text{m}^3$, whereas overpredictions of $1 \mu\text{g}/\text{m}^3$ or more occurred at 25 monitors. Correlation between predicted and observed 2009 DVs was high, with an R^2 of 0.74. Mean bias was slightly higher for predictions made by CMAQ ($0.42 \mu\text{g}/\text{m}^3$) than for those made by CAMx ($0.22 \mu\text{g}/\text{m}^3$). However, the difference is not statistically significant ($p=0.38$ for two-tailed t-test) and may reflect the fact that CAMx was primarily used in Midwest states and CMAQ elsewhere (Table 1).

Figure 2 shows improvements in $\text{PM}_{2.5}$ from the base (2000-2004) to target (2007-2009) period on a modeled and observed basis. This evaluation considers a more limited subset of 78 monitors for which complete data were available for both periods and for which the attainment demonstrations provided predicted DVs. All of the monitors in the regions that developed attainment demonstrations observed improvements of at least $1 \mu\text{g}/\text{m}^3$, and 20 monitors observed improvements of at least $3 \mu\text{g}/\text{m}^3$. The models underpredicted observed improvements by an average of $0.37 \mu\text{g}/\text{m}^3$, with greatest underpredictions in Detroit, Birmingham, and Atlanta.

An important aspect of model performance is its prediction of the highest DV in a region for the target year, since it is the monitor with the highest value that determines attainment. Excluding regions in which the monitor with the highest base and predicted DV was shut or had incomplete

data for the 2009 DV period (Atlanta, Chattanooga, and New York, NY) and considering only the single highest DV in interstate regions, the modeled attainment test overpredicted the highest observed DV within that region by an average of $0.47 \mu\text{g}/\text{m}^3$, with a mean error of $0.72 \mu\text{g}/\text{m}^3$. All predictions were within $1.7 \mu\text{g}/\text{m}^3$ of the actual value. As noted earlier, the modeled attainment test predicted nonattainment in four regions where attainment was in fact observed, and correctly predicted the status of the remaining regions. One of the false alarm regions, Atlanta, may have been influenced by the decommissioning of the monitor with the highest base and predicted DV.

Improvements in Nonattainment and Attainment Regions

An important question in assessing the value of the SIP process is whether nonattainment designation and the subsequent development of state and local control measures for attainment plans spur accelerated improvements in air quality. During the period of interest, substantial emission reductions were achieved through national and interstate efforts to control NO_x and SO_2 from point sources, and tighter emissions standards for vehicles and fuels (US-EPA, 2003,2005,2012). State-level legislation approved prior to PM attainment planning, such as the North Carolina Clean Smokestacks Act (Schwarz, 2005), also contributed to improvements. All of these measures, together with SIP-motivated state and local emissions controls, contributed to substantial reductions in all of the primary and secondary sources of PM except for ammonia (Xing *et al.*, 2013). The relative contribution of state and local measures prompted by attainment planning can be investigated by comparing the air quality improvements in the regions that developed attainment demonstrations with those that occurred elsewhere.

We expand our consideration of EPA AQS data beyond the regions identified in Table 1 to include all monitors with complete data for the base (2000-2004) and target (2007-2009) periods, excluding any sites designated as non-representative. Monitors are classified based on whether they were originally designated as attainment or nonattainment of the 1997 PM_{2.5} standard. Nonattainment region monitors are further divided into (1) regions that developed an attainment plan seeking attainment by 2009 (Table 1), (2) regions that attained the standard prior to submitting an attainment demonstration, and (3) regions that are seeking to develop an attainment plan for a later attainment date. Results for these categories are summarized in Table 2 and displayed in Figure 3.

Monitors in nonattainment regions achieved greater improvements in PM_{2.5} from 2000-2004 to 2007-2009 than monitors in regions that originally attained the 15 µg/m³ standard (Table 2). For monitors in the nonattainment regions that submitted attainment demonstrations (Table 1), improvements averaged 2.6 µg/m³ (17%), compared to 1.5 µg/m³ (13%) in regions that attained the standard from the beginning, or 1.6 µg/m³ (13%) if only eastern U.S. attainment monitors are considered (defined as Missouri and eastward, since Missouri is the westernmost state in Table 1, and excluding Puerto Rico). The greater improvement in attainment demonstration regions than at other eastern U.S. monitors that had never been designated nonattainment was highly statistically significant ($p < 10^{-18}$ for improvements expressed on a µg/m³ basis, and $p \sim 1.5 \times 10^{-8}$ for improvements expressed on a percentage basis). Air quality improvements in the regions were widespread, as all monitors achieved at least 1 µg/m³ of reduction. By contrast, 10% of monitors in the originally attaining regions (though only 6% of those within the eastern U.S.) experienced no change or increases in PM.

Some of the larger improvement in nonattainment regions than in attainment regions reflects a general trend of greater improvements at monitors that were more polluted in earlier years. The correlation between initial air quality and the magnitude of improvements parallels the findings of Pegues et al. (2012), who showed that ozone and NO₂ reductions were greatest in regions where pollution levels were initially highest. The effect of nonattainment status can be isolated by considering monitors with similar initial DVs. Among monitors with 2000-2004 PM_{2.5} DVs of 12-15 µg/m³, average improvements were larger in the nonattainment regions that submitted attainment demonstrations (2.3 µg/m³) than in the attainment regions within the eastern U.S. (1.9 µg/m³). The difference was statistically significant ($p=2.9 \times 10^{-5}$ for two-tailed t-test).

Nonattainment regions that did not write an attainment demonstration due to observing clean data prior to submission achieved slightly larger average improvements than the regions that submitted attainment demonstrations (Table 2). This is to be expected, since to achieve redesignation these regions inherently had observed improvements in air quality, which may or may not have resulted from control measures developed when an attainment demonstration was anticipated. Even faster air quality improvements were observed in regions that had not yet written an attainment demonstration as of 2009 due to a later attainment date. All but one of these monitors is in California, which has taken aggressive actions to reduce emissions of PM_{2.5} and its precursors (CARB, 2007).

One hypothesis to be tested is whether PM reductions in each nonattainment region occurred fastest at the most polluted monitor, as localized control measures near that monitor may have been prioritized to achieve attainment. The PM_{2.5} standard assigns attainment status and an

overall design value to each region based on the monitor with the highest design value, which creates added impetus to achieve reductions at that monitor. Many of the attainment demonstrations noted specific control measures targeting controls near the monitor modeled to have the highest 2009 DV. For example, the Michigan attainment demonstration for Detroit noted controls at steel producers, an oil refinery, and a rail yard adjacent to the Dearborn monitor which had the highest $PM_{2.5}$ in that region. Similarly, the Alabama attainment demonstration for Birmingham targeted controls at steel producers and pipe companies near its two highest monitors.

Considering only the monitor with the highest predicted 2009 DV in each region (selecting only one in multi-state regions) and excluding monitors with incomplete data in the base or target period, 18 monitors were available for analysis. At these monitors, the average improvement in $PM_{2.5}$ from 2000-2004 to 2007-2009 was $3.1 \mu\text{g}/\text{m}^3$, compared to $2.5 \mu\text{g}/\text{m}^3$ at other monitors in the Table 1 regions (Table 2). This could reflect the success of localized controls in achieving additional improvements at the most problematic monitors. However, the stronger reduction at peak monitors is also part of the general trend of greatest air quality improvements where the initial concentrations were the highest, even among monitors in attainment regions that did not prepare an attainment demonstration (Figure 3).

DISCUSSION

All of the attainment demonstrations considered here used the EPA recommended speciated modeled attainment test as the primary predictor of $PM_{2.5}$ levels in the target year. These

attainment demonstrations represent the first opportunity to evaluate the performance of the speciated $PM_{2.5}$ test in regulatory practice, since the test was set forth in an EPA guidance document in 2007. The modeled attainment test for $PM_{2.5}$ faces greater challenges than the similar one for ozone evaluated by Pegues et al. (2012), since it must consider PM on a speciated basis in predicting future concentrations. As noted earlier, the test for PM faces the added complication that speciation is not directly measured at most monitors and must be inferred from nearby speciation monitors.

Despite these challenges, the attainment test achieved accurate to somewhat conservative performance in predicting 2009 $PM_{2.5}$ DVs, with low mean bias and error both for individual monitors and for the highest value in each region. The test also achieved a 94.8% success rate in predicting whether monitors would exceed the standard. At monitors where the test predicted false alarms, weight of evidence determinations subsequently proved correct in predicting that attainment would in fact be achieved.

The meaningfulness of these results in evaluating the speciated modeled attainment test and associated models may be questioned given that the attainment demonstrations were issued around the same time (2008-2010; Table 1) as the observations were being observed that determined the target DVs (2007-2009). However, most of the emission inventory development and modeling was conducted years before the attainment demonstrations were submitted to EPA, and the modeling computed reduction factors relative to a 2002 base emissions period. The modeled tests also relied upon base rather than target period meteorology, and did not account for the financial recession that may have affected emissions during the final years. Though

observations within the target period were considered in some of the weight of evidence determinations, they did not factor directly into the speciated modeled attainment tests. In sum, the modeled tests achieved strong performance in spite of considerable challenges, even if the time horizon of predictions was short.

Regions issuing attainment demonstrations to attain the $15 \mu\text{g}/\text{m}^3$ $\text{PM}_{2.5}$ air quality standard by 2009 were remarkably successful in achieving that goal. All of these regions attained the standard by 2009, and Birmingham attained a year later after originally requesting a three year extension. Average air quality improvements at monitors in these regions was substantial, $2.6 \mu\text{g}/\text{m}^3$, given the short time between the base (2000-2004) and target (2007-2009) years. Although $\text{PM}_{2.5}$ reductions were slightly greater at peak monitors that drive attainment status, improvements were widespread across the regions that submitted attainment demonstrations. Given the correlations between incremental changes in $\text{PM}_{2.5}$ and mortality rates (Pope and Dockery, 2006), these improvements likely yielded substantial public health benefits.

The role of nonattainment status and the SIP process relative to national trends in driving these air quality improvements is difficult to quantify, since $\text{PM}_{2.5}$ levels also declined in most regions that did not develop attainment demonstrations. The fastest air quality improvements occurred in nonattainment regions with later attainment dates than the regions considered in Table 1, indicating an aggressive pursuit of control measures before those attainment demonstrations are submitted. Monitors in nonattainment regions that submitted attainment demonstrations achieved nearly twice as much $\text{PM}_{2.5}$ reduction as monitors in attainment regions. Some of this difference reflects more polluted conditions in earlier years at monitors in regions that submitted attainment

demonstrations. However, even on a percentage basis, monitors in the Table 1 regions observed PM reductions at significantly faster rates than those in attainment regions. Among monitors with similar initial DVs, moderately faster improvements occurred in the nonattainment regions than in attainment regions. Overall, the results are consistent with a hypothesis that national controls drove the majority of PM_{2.5} reductions, but that the SIP process yielded substantial improvements as well. Additional insight could be gained in future research by considering relative trends in PM_{2.5} on a speciated basis, utilizing data from the minority of monitors that reports speciated data.

The results presented here provide a retrospective evaluation of the performance of modeling and control measures for attaining the 1997 15 µg/m³ standard, as attention now turns to attaining the more stringent 12 µg/m³ standard issued in 2012. EPA plans to designate nonattainment regions in 2014, and allow states until 2020-2025 to attain the standard. The results presented here provide confidence in the continued use of the speciated modeled attainment test to predict future PM_{2.5} levels. The weight of evidence determinations proved to be a helpful supplement to the attainment test in the attainment demonstrations considered here, as they allowed states to document localized controls and other considerations that counteracted false alarm predictions of nonattainment. Caution is still needed in reviewing weight of evidence determinations, since for both ozone and PM_{2.5} they have almost uniformly been used to suggest pollution levels will be lower than modeled; as documented by Pegues *et al.* (2012), this led to false predictions of attainment in some ozone attainment demonstrations. Nevertheless, the cases considered here paint an overall picture of success in achieving the previous PM_{2.5} standard, and substantial progress toward achieving the new one.

ACKNOWLEDGMENTS

We appreciate the assistance of numerous state agencies in providing state implementation plans. The work was funded by National Science Foundation CAREER Award Grant 087386 and the NASA Air Quality Applied Sciences Team.

REFERENCES

Anderson, J. (2006). "Revisiting the SIP process: Finding a better approach to cleaner air." State Bar of Texas Environmental Law Journal **36**(4): 213-222.

Bachmann, J. (2007). "Will the circle be unbroken: A history of the US national ambient air quality standards." Journal of the Air & Waste Management Association **57**(6): 652-697.
doi:10.3155/1047-3289.57.6.652

Blanchard, C. L., G. M. Hidy, S. Tanenbaum, E. S. Edgerton and B. E. Hartsell (2013). "The Southeastern Aerosol Research and Characterization (SEARCH) study: Temporal trends in gas and PM concentrations and composition, 1999-2010." Journal of the Air & Waste Management Association **63**(3): 247-259. doi:10.1080/10962247.2012.748523

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." Applied Mechanics Review **59**: 51-77. doi:10.1115/1.2128636

CARB (2007). 2007 State Strategy for California's State Implementation Plan (SIP) for Federal PM_{2.5} and 8-Hour Ozone Standards. C. A. R. Board. Sacramento, CA.

Coen, M. C., et al. (2013). "Aerosol decadal trends - Part 1: In-situ optical measurements at GAW and IMPROVE stations." *Atmospheric Chemistry and Physics* **13**(2): 869-894.
doi:10.5194/acp-13-869-2013

Dockery, D. W., C. A. Pope, X. P. Xu, J. D. Spengler, J. H. Ware, M. E. Fay, B. G. Ferris and F. E. Speizer (1993). "An association between air pollution and mortality in 6 United States cities." *New England Journal of Medicine* **329**(24): 1753-1759. doi:10.1056/nejm199312093292401

ENVIRON (2010). User's Guide - Comprehensive Air Quality Model with Extensions Version 5.30.

Fine, J. D. and D. Owen (2005). "Technocracy and democracy: Conflicts between models and participation in environmental law and planning." *Hastings Law Journal* **56**(5): 901-+.

Hand, J. L., B. A. Schichtel, W. C. Malm and N. H. Frank (2013). "Spatial and Temporal Trends in PM_{2.5} Organic and Elemental Carbon across the United States." *Advances in Meteorology*.
doi:10.1155/2013/367674

Jones, J. M., C. Hogrefe, R. F. Henry, J. Y. Ku 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." *Journal of the Air & Waste Management Association* **55**(1): 13-19.
doi:10.1080/10473289.2005.10464601

NRC (2004). *Air Quality Management in the United States*. Washington, DC, The National Academies Press.

Pegues, A. H., D. S. Cohan, A. Digar, C. Douglass and R. S. Wilson (2012). "Efficacy of recent state implementation plans for 8-hour ozone." *Journal of the Air & Waste Management Association* **62**(2): 252-261. doi:10.1080/10473289.2011.646049

Pope, C. A. and D. W. Dockery (2006). "Health effects of fine particulate air pollution: Lines that connect." *Journal of the Air & Waste Management Association* **56**(6): 709-742. doi:10.1080/10473289.2006.10464485

Schwarz, P. M. (2005). "Multipollutant efficiency standards for electricity production." *Contemporary Economic Policy* **23**(3): 341-356. doi:10.1093/cep/byi026

Sistla, G., C. Hogrefe, W. Hao, J. Y. Ku, E. Zalewsky, R. F. Henry and K. Civerolo (2004). "An operational assessment of the application of the relative reduction factors in the demonstration of attainment of the 8-hr ozone national ambient air quality standard." *Journal of the Air & Waste Management Association* **54**(8): 950-959. doi:10.1080/10473289.2004.10470964

US-EPA (2003). NO_x budget trading programs for state implementation plans, U.S. Code of Federal Regulations: 665-707.

US-EPA (2005). Rule to reduce interstate transport of fine particulate matter and ozone (Clean Air Interstate Rule), Federal Register: 25162-25404.

US-EPA (2007). Guidance on the use of models and other analyses for demonstrating attainment of air quality goals for ozone, PM_{2.5}, and regional haze.

US-EPA (2012). Our Nation's Air -- Status and Trends Through 2010.

Xing, J., J. Pleim, R. Mathur, G. Pouliot, C. Hogrefe, C. M. Gan and C. Wei (2013). "Historical gaseous and primary aerosol emissions in the United States from 1990 to 2010." *Atmospheric Chemistry and Physics* **13**(15): 7531-7549. doi:10.5194/acp-13-7531-2013

Table 1. Attainment demonstrations for PM_{2.5} considered in this study.

Region	SIP State	Submit Date	Model Used	Highest Base DV^a	Highest Predicted 2009 DV^b	Highest Observed 2009 DV^c	Highest Observed 2012 DV^g
Atlanta	GA	Mar-10	CMAQ	18.3	16.6	13.5 ^d	12.9
Baltimore	MD	Mar-08	CMAQ	16.6	14.4	12.9	11.1
Birmingham	AL	May-09	CMAQ	18.4	16.5	15.1	13.0
Charleston	WV	Nov-09	CMAQ	17.1	14.8	14.4	11.9
Chattanooga	GA	Oct-09	CMAQ	15.7	13.9	NA ^d	10.0
	TN	Apr-08	CMAQ	16.2	14.5 ^e	11.6	11.2
Cincinnati-Hamilton	KY	Nov-08	CAMx	14.9	12.7	12.3	10.6
	OH	Jul-08	CAMx	17.8	14.7	15.0	13.

Cleveland	OH	Jul-08	CAMx	18.4	15.2	14.4	13.0
Detroit	MI	May-08	CAMx	19.3	15.8	14.1	11.5
Greensboro-Winston Salem	NC	Aug-09	CMAQ	16.0	13.4	13.0	11.1
Hickory-Morganton	NC	Aug-09	CMAQ	15.6	13.1	12.5	10.3
Huntington-Ashland	KY	Nov-08	CAMx	14.9	12.8	12.4	10.6
	WV	May-09	CMAQ	16.6	14.1	14.3	11.6
Indianapolis	IN	May-08	CAMx	16.5	13.2	14.3	12.7
Knoxville	TN	Apr-08	CMAQ	16.7	14.9	13.7	12.2
Louisville	KY	Nov-08	CAMx	16.6	14.8	13.6	12.3
	IN	Jun-08	CAMx	16.3	13.6	14.7	13.2
Macon	GA	Jun-09	CMAQ	15.7	14.3	13.5	13.1

Martinsburg- Hagerstown	MD	Feb-08	CMAQ	14.3	11.8	11.5	11.6
New York	CT	Nov-08	CMAQ	13.7	11.7	11.4	9.4
	NJ	Jun-08	CMAQ	16.6	14.3	13.1	11.2
	NY	Oct-09	CMAQ	17.2	15.2	13.9 ^d	11.8
Philadelphia- Wilmington	NJ	Jun-08	CMAQ	14.1	12.4	11.7	9.5
	PA	Apr-10	CMAQ	16.1	13.5	13.7	13.4
Pittsburgh	PA	Nov-09	CMAQ	16.7	14.3	14.3	12.6
Rome	GA	Oct-09	CMAQ	15.7	14.0	13.3	12.1
St. Louis ^f	MO-IL	Sep-09	CMAQ	17.3	15.5	14.1	13.5
Steubenville	OH	Jul-08	CAMx	17.9	13.5	14.2	12.7

Washington	DC-MD-VA	Mar-08	CMAQ	15.8	12.6	12.1	10.5
------------	----------	--------	------	------	------	------	------

^aAverage of the observed 3-year design values ending in 2002, 2003, and 2004, in $\mu\text{g}/\text{m}^3$, at the highest monitor in the region. All observational data are from a 7/26/2012 query of the US EPA Air Quality System, and consider only monitors with complete data.

^bFrom the modeled attainment demonstration for 2009.

^cFrom observational data for 3 years ending in 2009.

^dMonitor with highest base and predicted DV was decommissioned prior to 2009.

^eChattanooga, TN, attainment demonstration was not found. Modeled DV from Knoxville attainment demonstration.

^fIllinois attainment demonstration not found. Modeled DV from Missouri attainment demonstration for monitor in Illinois.

^g From observational data for 3 years ending in 2012, from a 7/10/2013 query of the US EPA Air Quality System

Table 2. Reduction in PM_{2.5} at monitors in regions originally designated as nonattainment or attainment of the annual standard.

	Number of Monitors	Average Base DV^a	Average (stdev) Reduction^b
Nonattainment regions			
Attainment demonstration to attain by 2009 ^c	102	15.3	2.6 (0.7)
Monitor with highest predicted DVF ^d	18	16.7	3.1 (0.7)
Other monitors	84	15.0	2.5 (0.7)
Post-2009 target	24	20.3	4.5 (3.5)
Attained before submitting a demonstration	30	15.3	2.8 (0.6)
Attainment regions	287	11.3	1.5 (1.0)

Missouri and eastward ^e	189	11.9	1.6 (0.9)
West of Missouri	96	10.0	1.1 (1.0)

^aAverage of the observed 3-year design values ending in 2002, 2003, and 2004, in $\mu\text{g}/\text{m}^3$.

^bReduction in $\text{PM}_{2.5}$ from base period to 2007-2009 design value.

^cIncludes Birmingham, Alabama, which sought a later attainment date after modeling exceedance values in its 2009 attainment demonstration.

^dMonitor for which modeled attainment test predicted highest 2009 DV in each attainment demonstration.

^eExcludes two monitors in Puerto Rico. Includes Minnesota, Iowa, Arkansas and Louisiana.

Figure 1. Annual PM_{2.5} design values predicted by attainment demonstrations and observed at monitors in 2007–2009.

Figure 2. Predicted and observed reductions in PM_{2.5} design values from 2000-2004 to 2007-2009.

Figure 3. Observed reductions in PM_{2.5} design values from 2000-2004 to 2007-2009, grouped by the nonattainment and SIP status of the region.