AEA Papers and Proceedings 2019, 109: 283–288
https://doi.org/10.1257/pandp.20191064

Bringing Satellite-Based Air Quality Estimates Down to Earth†

By Meredith Fowlie, Edward Rubin, and Reed Walker*

Particulate matter pollution poses serious health risks—particularly for children, the elderly, and sensitive populations. In the United States, air pollution regulations have increasingly focused on smaller particles, such as those less than 2.5 micrometers ($PM_{2.5}$). These regulations are enforced using ambient air pollution measurements, collected from the EPA's air quality monitoring network.

The network of regulatory-grade monitors is spatially sparse; more than 80 percent of US counties do not contain a $PM_{2.5}$ monitor. Coarse measurements of air pollution concentrations can lead to significant gaps in our understanding of the burden of exposure for certain areas. These gaps have potentially important implications for the design and implementation of existing air quality regulations.

Recent advances in satellite technology, combined with advances in prediction techniques—e.g., machine learning—may relax some of these information constraints. For instance, a growing suite of satellite observations of aerosol optical depth (AOD) makes it possible to estimate ground-level concentrations of $PM_{2.5}$ at fine spatial resolutions (<1km). Social scientists are increasingly using these satellite-based estimates of $PM_{2.5}$ concentrations to analyze the health and economic impacts of ambient pollution exposure (e.g., Sullivan and Krupnick 2018, Voorheis 2016, and Di et al. 2017).

This paper uses two state-of-the-art, satellite-based $PM_{2.5}$ data products (Di et al. 2016, Van Donkelaar et al. 2019) to assess the extent to which the EPA's existing, monitor-based measurements over- or under-estimate true exposure to $PM_{2.5}$ pollution. We show that regulatory-grade monitor measurements fail to capture a significant amount of spatial variation in the satellite-based estimates. Treating satellite-based estimates as truth would imply a substantial number of "policy errors" by the EPA—over-regulating certain areas that are already in compliance with the Clean Air Act (CAA) National Ambient Air Quality Standards (NAAQS) and under-regulating other areas that, according to the satellite-based estimates, are in violation of the standards. Somewhat counterintuitively, we show that recalibrating existing policies to capture more spatially resolved measures of pollution exposure need not improve health outcomes overall.

We also highlight the importance of accounting for prediction error in satellite-based estimates. These highly spatially resolved datasets offer the potential for new and important insights into the distribution and impacts of air quality. However, these data are estimates of the true $PM_{2.5}$ concentration at a location and contain prediction or forecast errors. The forecast errors associated with these satellite-based data products have largely been ignored by the social-science research community, and many of our original conclusions in regards to "policy errors" become substantially more uncertain.

I. Pollution-Concentration Measurement and Estimation

The US EPA directly measures surface $PM_{2.5}$ concentrations using *in situ*, filter-based monitors. Together these monitors form a precise but spatially sparse network of $PM_{2.5}$ measurements that is fairly expensive to maintain. Recent work in atmospheric, computer, and environmental

* Fowlie: University of California Berkeley, 243 University Hall, Berkeley, CA 94720 (email: fowlie@berkeley.edu); Rubin: University of Oregon, Eugene, OR 97403 (email: edwardr@uoregon.edu); Walker: University of California Berkeley, Berkeley, CA 94720 (email: rwalker@berkeley.edu). We thank both Qian Di and Aaron van Donkelaar for sharing their high-resolution, satellite-based PM estimates. We thank Ellen Lin and Catherine Wright for extraordinarily helpful research assistance.

† Go to https://doi.org/10.1257/pandp.20191064 to visit the article page for additional materials and author disclosure statement(s).

sciences offers the potential to extend the spatial coverage of $PM_{2.5}$ measurements.

By combining satellite-based measures of AOD with chemical-transport modeling and land characteristics, researchers are able to estimate ground-level concentrations of $PM_{2.5}$ at high levels of spatial disaggregation. Further, the in situ EPA monitors provide training data for statistical models—mitigating bias and increasing precision in these satellite-based estimates.

We obtained two data products that estimate annual $PM_{2.5}$ concentrations in the continental United States at a high spatial resolution. First, Di et al. (2016) uses a neural network to predict daily $PM_{2.5}$ concentrations at nationwide 1 km $\times$ 1 km grid cells over the period 2000 to 2015. Second, Van Donkelaar et al. (2019) combines satellite remote-sensing data with chemical-transport modeling and geographically weighted regression to predict annual $PM_{2.5}$ concentrations at 1-kilometer resolution 1998–2016. We spatially intersect both sets of data with US Census block-group (CBG) boundary files from the year 2000. Online Appendix Figures 1(a) and 1(b) plot estimated $PM_{2.5}$ concentrations for 2005 by Di et al. (2016) and Van Donkelaar et al. (2019), respectively.

II. Policy Context

The United States' Clean Air Act National Ambient Air Quality Standards (NAAQS) specify maximum allowable concentrations for common air pollutants (e.g., $PM_{2.5}$ and lead). Compliance (*attainment*) within NAAQS is determined using monitor-based *design values*. For $PM_{2.5}$, each EPA monitor is used to construct two design values: a three-year annual average concentration and a three-year average of the annual ninety-eighth percentile of 24-hour concentrations. If either design value exceeds its respective NAAQS $PM_{2.5}$ threshold, the EPA classifies the monitor's jurisdiction (usually its county) as *non-attainment*. Areas that fail to meet these standards must take steps to improve air quality (e.g., mandatory pollution abatement technologies for air pollution point sources).

Our analysis focuses on the 1997 $PM_{2.5}$ NAAQS, which set an annual average standard of 15 $\mu g/m^3$ and a 24-hour standard of 65 $\mu g/m^3$. Following court challenges, these 1997 standards were enacted in 2005. Virtually all non-attainment designations from the 1997 standard occurred due to violations of the annual (versus 24 hour) standard.[1] We use the satellite-based estimates to construct design values for each CBG, and we compare these design values to the de jure, county-level design values (i.e., design values based on the maximum EPA monitor readings within the county).

We first use EPA AQS monitors to construct the three-year annual average design values for all 685 counties that had monitors in 2005. Counties that do not have a monitor are assumed to be in attainment. Next, we use the satellite-based estimates constructed by Di et al. (2016) and Van Donkelaar et al. (2019) to construct the three-year annual average design values for every CBG in 2005. Figure 1 summarizes the relationship between the satellite-based design values and the corresponding monitor-based design values. Figure 1, panel A explores these relationships using the Di et al. (2016) data, whereas Figure 1, panel B plots the monitor versus Van Donkelaar et al. (2019) data. The distribution to the left of each figure shows the extent of variation in satellite-based estimates in counties with no EPA monitor.

These figures illustrate the striking variation in satellite-based measurements for counties that share the same monitor-based, countywide design value. Recall that the monitor-based, countywide design value is the only piece of information that the EPA currently uses to regulate counties under NAAQS. If we assume that these satellite-based estimates are precise and unbiased, these figures suggest that the county-level, monitor-based design values are a very crude proxy for true pollution concentrations in many locations.

However, some of the observed variation in satellite-based estimates likely reflects prediction errors, rather than true variation in underlying $PM_{2.5}$ concentrations. Ideally, our analysis would account for both bias and uncertainty in these estimates. We explore the extent of prediction errors by focusing on the 911 CBGs equipped with an EPA monitor, comparing the satellite-based estimates to the EPA monitor readings for the same area. Online Appendix

[1] In contrast, violations of the current standards (enacted in 2009) were mostly triggered by violations of the 24-hour standard. We cannot construct these design values using annual satellite-based estimates, so we focus on the earlier standard.

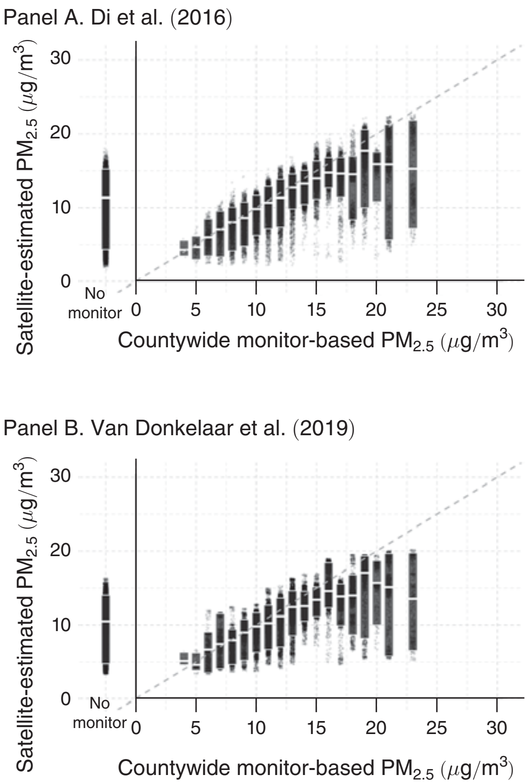

Figure 1. Comparing $PM_{2.5}$ Measurements: Monitor-Based versus Satellite-Based Estimates

Notes: These figures plot the relationship between satellite-based design values and monitor-based design values in 2005. An observation is a census block group. The graphs show the variation in satellite-based design values for each level of monitor design values. The distribution to the left of each figure shows the variation in satellite-based estimates in counties with no EPA monitor.

Source: Di et al. (2016), Van Donkelaar (2019), EPA-AQS

Figures 2(a) and 2(b) provide a sense of the range of satellite-based estimates we observe across CBGs with similar monitor readings. The range of these estimates, particularly at higher measured $PM_{2.5}$ concentrations, is significant.

Regulatory-grade monitors measure pollution concentrations directly and with high precision at a particular location. If we assume that spatial variation within a CBG is minimal, we can interpret the difference between monitor-based design values and the satellite-based design values as prediction errors for the 911 CBGs that have a monitor. However, there are over 215,000 CBGs without a monitor, so we try to forecast the prediction errors for these CBGs "out of sample." We begin by regressing the "in-sample" prediction errors on a set of seven CBG-level observable variables.[2] We use this regression model to predict errors in the satellite-based predictions—for both in-sample (the 911 CBGs that contain a monitor) and out-of-sample predictions (the more than 215,000 CBGs without a monitor).We use the standard error from this regression model to create a 95 percent prediction interval for each CBG pollution estimate.[3] We will use these prediction intervals below to better understand the extent to which our conclusions are sensitive to this measure of satellite-based estimation uncertainty.

III. Non-Attainment Designations, Revisited

We distinguish between two types of attainment designation "errors." A "type 1" error (i.e., false positive) occurs if the three-year annual average of satellite-based estimates of $PM_{2.5}$ concentrations in a CBG falls below the NAAQS standard of 15 $\mu g/m^3$, but the associated county-level, EPA monitor-based design value exceeds this threshold. Conversely, a "type 2" error (i.e., false negative) occurs if the estimated CBG pollution concentration exceeds the regulatory standard, whereas the associated county-level, monitor-based design value does not.

A. *Policy "Errors"*

Panel A of Table 1 summarizes the results of this classification exercise using the Di et al. (2016) satellite data, whereas panel C presents results using Van Donkelaar et al. (2019) $PM_{2.5}$ estimates. We first calculate designation errors assuming that the satellite-based estimates provide an unbiased and precise estimate of true $PM_{2.5}$ concentrations. We then incorporate

[2] The CBG-level explanatory variables in this regression are: the monitor-based $PM_{2.5}$ estimate, total population, the share of the population that is white, the share of the population that is rural, minimum and maximum elevation, and land area.

[3] For this simple thought exercise, we are assuming that the regression error is independent of the explanatory variables normally distributed, with zero mean, and constant variance.

Table 1—Comparing NAAQS Designation: Monitors and Satellite-Based Estimates

	Satellite attain.		Satellite non-attain.	
Monitor designation:	Attainment (1)	Non-Attainment (2)	Attainment (3)	Non-Attainment (4)
Panel A. Population summary (Di et al. 2016)				
Population (millions)	234.3	33.1	5.7	29.9
	(239.7, 111.7)	(29.2, 0.2)	(0.3, 128.3)	(33.8, 62.9)
Population share	77.3%	10.9%	1.9%	9.9%
	(79.1%, 36.9%)	(9.6%, 0.1%)	(0.1%, 42.3%)	(11.2%, 20.8%)
Panel B. Mortality impacts (Di et al. 2016)				
Avoided deaths	4,640	694	116	614
Lower estimate	(4,748, 2,201)	(651, 5)	(8, 2,556)	(657, 1,303)
Avoided deaths	13,489	1,982	335	1,726
Higher estimate	(13,802, 6,448)	(1,868, 14)	(22, 7,376)	(1,840, 3,694)
Panel C. Population summary (Van Donkelaar et al. 2019)				
Population (millions)	238.8	42.3	1.2	20.8
	(240.0, 106.2)	(43.8, 0.2)	(0.0, 133.8)	(19.3, 62.8)
Population share	78.8%	14.0%	0.4%	6.9%
	(79.2%, 35.0%)	(14.5%, 0.1%)	(0.0%, 44.2%)	(6.4%, 20.7%)
Panel D. Mortality impacts (Van Donkelaar et al. 2019)				
Avoided deaths	4,733	883	23	425
Lower estimate	(4,757, 2,080)	(949, 5)	(0, 2,676)	(359, 1,302)
Avoided deaths	13,758	2,532	66	1,175
Higher estimate	(13,824, 6,097)	(2,721, 15)	(0, 7,727)	(987, 3,693)

Notes: These estimates come from comparing satellite-based estimates to EPA AQS monitor data. We spatially intersect the Di et al. (2016) and Van Donkelaar et al. (2019) estimates with census block groups to provide the relevant demographic characteristics and baseline mortality rates. The column NAAQS classifications are based on the 2005 three-year annual design values, calculated at (i) the county level for EPA monitors or (ii) at the census block-group level for the satellite-based estimates. Avoided death estimates come from two concentration-response functions: lower estimate (Krewski et al. 2009) and higher estimate (Lepeule et al. 2012). Numbers in parentheses describe how the point estimate changes when we use the lower (or upper) bound of the 95 percent prediction intervals for the error in the $PM_{2.5}$ estimate.

uncertainty stemming from prediction errors, using the lower and upper bounds of the 95 percent prediction interval (for the predicted error) to compute designation errors. Numbers in parentheses report results using the lower and upper estimates, respectively.

Panels A and C in Table 1 show how populations are distributed across correctly classified and misclassified attainment designations, respectively. Column 1 shows that a majority of the population live in areas that have been correctly designated as in attainment based upon year-2005 design values (satellite-based point estimates imply around 78 percent fall into this category). Column 4 shows that the share of the population living in properly designated non-attainment areas is much smaller. We find type 1 errors (column 2) are much more prevalent than type 2 errors (column 3). Note, 11–14 percent of the population live in areas that are designated as non-attainment using the de jure monitor measurement but are associated with satellite-based estimates of $PM_{2.5}$ concentrations that fall below the NAAQS limits. Only 1–2 percent of the population live in areas that appear to exceed the NAAQS threshold (using either satellite-based data product), but are classified as "attainment" under the de jure, monitor-based NAAQS policy. Estimates in parentheses show how the relative importance of type 1 and type 2 errors is sensitive to the prediction interval bounds we use. Intuitively, when we use the lower bound of the 95 percent prediction interval for the satellite data, we are more likely to see CBGs misclassified as non-attainment based on de jure monitor readings when "true" pollution concentrations, as measured by satellites, meet the standard (i.e., type 1 errors). When we use the upper bound of the 95 percent prediction intervals from the satellite data, we see more CBGs designated as in attainment based on monitor readings when satellite-based estimates exceed the NAAQS threshold (type 2 errors).

B. *Health Implications*

The vast majority of the damages associated with $PM_{2.5}$ exposure are mortality related. Panels B and D of Table 1 use the satellite-based estimates of $PM_{2.5}$ concentrations to estimate the likely health implications of the classification errors we have identified.

To assess the mortality impacts of our findings, we adopt an approach similar to the regulatory impact analyses conducted by the EPA, which is based on estimated concentration-response (or "hazard") functions. These functions relate $PM_{2.5}$ exposure to mortality risk. Importantly, the scientific evidence on health impacts has yet to identify a safe threshold for $PM_{2.5}$ exposure.[4] In contrast, the threshold-based design of NAAQS is most consistent with marginal damages that are low or zero below the threshold and high above. This mismatch between the structure of the NAAQS and the underlying concentration-response relationship has important implications when assessing the health implications of designation errors. In particular, it implies that type 1 errors (i.e., overregulation) generate potentially significant benefits in the form of reduced mortality.

Panels B and D of Table 1 summarize estimated annual mortality benefits associated with a 1 $\mu g/m^3$ reduction in $PM_{2.5}$ concentrations. "Lower" estimates of deaths avoided are based on Krewski et al. (2009). "Higher" estimates are based on Lepeule et al. (2012). See online Appendix A1 for more details. We speculate that moving a county into non-attainment would induce a reduction in annual average concentrations of at least 1 $\mu g/m^3$. To put this assumption in perspective, Sullivan and Krupnick (2018) estimates that a non-attainment classification under the 2012 standard reduced pollution concentrations by more than 2 $\mu g/m^3$.

Satellite-based point estimates imply that the mortality implications of type 1 errors (i.e., reduction in mortality from regulating areas already in compliance) may be much more consequential than the foregone mortality benefits associated with type 2 errors (i.e., the mortality increase associated with failing to regulate areas that are out of compliance). Panel B of Table 1 suggests that when using the higher hazard ratio parameters of Lepeule et al. (2012), 335 deaths resulted from a failure to designate areas exceeding the NAAQS threshold as non-attainment, whereas 1,982 deaths were avoided as a consequence of designating areas that met the standard as non-attainment. The estimates from panel D are qualitatively similar. However, these results are sensitive to which prediction-interval bounds we use. In other words, our estimated prediction errors suggest significant uncertainty underlies these estimated mortality impacts of type 1 and type 2 errors.

IV. Conclusion

Newly available, spatially resolved pollution data present a host of new opportunities—for both research and policy. We use state-of-the-art satellite estimates to assess the extent to which the limited network of EPA monitors leads to over- and/or under-detection of violations of $PM_{2.5}$ standards.

We arrive at the surprising conclusion that using more spatially disaggregated measures of $PM_{2.5}$ concentrations to determine NAAQS attainment need not be welfare improving, relative to the current status quo. The reason is twofold. First, we find that a significant share of the population is living in areas where satellite-based estimates of pollution concentrations fall below the NAAQS threshold, but EPA monitor-based design values exceed the threshold (i.e., these populations received health benefits from "overregulation"). In contrast, the share of the population living in areas where the reverse appears to be true is small. Second, the design of the NAAQS standards poorly approximate the underlying damage function. This implies that marginal benefits from pollution reductions are significant in areas that meet NAAQS standards.

Finally, it is important to recognize that satellite-based estimates of pollution concentrations are not direct measures. Prediction error appears to be economically significant, and the error structure is poorly understood. In general, satellite estimates appear to be biased down at higher $PM_{2.5}$ concentrations. We conclude that further

[4] In fact, there is some evidence that the mortality-related benefits from incremental reductions in $PM_{2.5}$ concentrations may be *higher* at lower concentrations (EPA 2018).

work exploring the precision, bias, and limits of these estimates remains important to understanding the health and policy implications of spatial heterogeneity in pollution exposure.

REFERENCES

Di, Qian, Lingzhen Dai, Yun Wang, Antonella Zanobetti, Christine Choirat, Joel D. Schwartz, and Francesca Dominici. 2017. "Association of Short-Term Exposure to Air Pollution with Mortality in Older Adults." *Journal of the American Medical Association* 318 (24): 2446–56.

Di, Qian, Itai Kloog, Petros Koutrakis, Alexei Lyapustin, Yujie Wang, and Joel Schwartz. 2016. "Assessing PM2.5 Exposures with High Spatiotemporal Resolution across the Continental United States." *Environmental Science and Technology* 50 (9): 4712–21.

Environmental Protection Agency (EPA). 2018. *Integrated Science Assessment for Particulate Matter* (*External Review Draft*). Washington, DC: US Environmental Protection Agency.

Krewski, Daniel, Michael Jerrett, Richard T. Burnett, Renjun Ma, Edward Hughes, Yuanli Shi, Michelle C. Turner, et al. 2009. *Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality*. Boston: Health Effects Institute.

Lepeule, Johanna, Francine Laden, Douglas Dockery, and Joel Schwartz. 2012. "Chronic Exposure to Fine Particles and Mortality: An Extended Follow-Up of the Harvard Six Cities Study from 1974 to 2009." *Environmental Health Perspectives* 120 (7): 965–70.

Sullivan, Daniel M., and Alan Krupnick. 2018. "Using Satellite Data to Fill the Gaps in the US Air Pollution Monitoring Network." Resources for the Future Working Paper 18-21.

Van Donkelaar, Aaron, Randall V. Martin, Chi Li, and Richard T. Burnett. 2019. "Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors." *Environmental Science and Technology* 53 (5): 2595–2611.

Voorheis, John. 2016. "Environmental Justice Viewed from Outer Space: How Does Growing Income Inequality Affect the Distribution of Pollution Exposure?" Unpublished.