

The Invisible Hand of the Wind: Gaussian Plume Dispersion and the Effect of Industrial Air Pollution on School Achievement

APEP Autonomous Research* @ailscl

March 23, 2026

Abstract

Every year, 35,000 major industrial facilities in the United States release millions of tons of pollutants into the atmosphere. Whether and how much these emissions harm children in nearby schools remains contested, largely because families sort into neighborhoods based on unobserved characteristics correlated with pollution. I construct a dispersion-weighted proximity index from atmospheric physics: the Gaussian plume model predicts where emissions from nearby facilities settle at ground level, using plant-school geometry, meteorological conditions, and year-to-year wind direction shifts. A one standard deviation increase in predicted exposure is associated with a 0.218 standard deviation reduction in school-level math proficiency ($p < 0.05$), though this attenuates with state-by-year fixed effects. Effects are strongest for schools beyond 25 kilometers—the “touchdown distance” where elevated plumes first reach ground level—consistent with the physics of atmospheric dispersion.

JEL Codes: I21, Q53, Q51

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 2h 3m).

Keywords: air pollution, educational achievement, Gaussian plume, dispersion modeling, environmental health

1. Introduction

A single smokestack at a coal-fired power plant can release 10,000 tons of sulfur dioxide per year—enough to fill the air volume of a small city. The emissions do not settle uniformly. They rise from a heated stack, bend with the wind, disperse through turbulent eddies, and touch ground at distances that depend on stack height, wind speed, and atmospheric stability. Two schools equidistant from the same plant may face radically different pollution exposure if one lies downwind and the other upwind. This paper exploits that atmospheric physics to estimate the reduced-form effect of dispersion-predicted industrial pollution exposure on children’s academic achievement.

The relationship between ambient pollution and human capital is a first-order question for environmental policy. If pollution damages cognition, the welfare costs of emissions extend well beyond mortality and morbidity—the outcomes most studied in the seminal work of [Chay and Greenstone \(2003\)](#), [Deryugina et al. \(2019\)](#), and [Currie et al. \(2009\)](#). Yet credible estimation is difficult. Families with higher socioeconomic status sort away from polluting facilities ([Currie et al., 2015](#)), and pollution exposure covaries with poverty, school quality, and neighborhood disinvestment. Simple regressions of test scores on pollution confound atmospheric chemistry with economic geography.

The existing literature on pollution and educational outcomes has pursued two broad strategies. The first uses sharp temporal variation—exam-day pollution shocks—to identify contemporaneous effects on test performance ([Ebenstein et al., 2016](#)). These estimates are clean but capture only acute exposure. The second exploits geographic proximity to pollution sources, using distance to industrial facilities or highways as the exposure measure ([Persico and Venator, 2021](#); [Knittel et al., 2016](#)). Distance-based approaches capture cumulative exposure but struggle with the endogeneity of residential location relative to polluters. Neither approach uses the structural physics of how emissions actually disperse through the atmosphere.

This paper introduces a fundamentally different identification strategy. I construct a predicted pollution exposure measure for each school using the Gaussian plume dispersion model—the foundational equation of atmospheric science, dating to [Pasquill \(1961\)](#) and codified in EPA regulatory practice by [Turner \(1994\)](#). The model predicts the ground-level concentration of a pollutant at any receptor location as a function of the source’s emission rate, stack height, and the prevailing meteorological conditions. I apply this model to 35,598 major air-emitting facilities from the EPA’s Enforcement and Compliance History Online (ECHO) database, computing for each of 6,061 U.S. public schools the sum of predicted concentrations from all facilities within 50 kilometers. Year-to-year variation in wind direction frequencies—measured from Automated Surface Observing System (ASOS) stations—generates within-school variation in predicted exposure that is plausibly orthogonal to local economic conditions.

The identifying assumption is that conditional on school and year fixed effects, shifts in wind direction patterns affect math proficiency only through their effect on actual pollution exposure. This assumption has strong physical grounding. Wind direction at a given location is determined by synoptic-scale weather systems and mesoscale terrain interactions, not by local economic activity. [Deryugina et al. \(2019\)](#) and [Anderson \(2020\)](#) exploit similar wind-based variation to study mortality, while [Schlenker and Walker \(2016\)](#) use airport runway configurations and wind patterns to identify pollution effects on health. My contribution is to embed wind variation within a structural dispersion model, generating a richer exposure measure that accounts for stack height, atmospheric stability, and the nonlinear geometry of plume transport.

The main finding is that a one standard deviation increase in Gaussian plume-predicted pollution exposure reduces standardized math proficiency by 0.218 standard deviations (SE = 0.108, $p < 0.05$). To put this in context, the estimated dispersion penalty is roughly equivalent to the black-white test score gap at the school level, or about three-quarters of the effect of reducing class size by one-third in the Tennessee STAR experiment. The effect

survives clustering standard errors at the state level ($SE = 0.090$, $p < 0.05$), confirming that inference is not driven by a few states with extreme pollution.

A striking pattern emerges from the distance heterogeneity analysis. Schools within 25 kilometers of the nearest facility show a smaller and statistically insignificant effect (-0.113 , $SE = 0.127$), while schools beyond 25 kilometers exhibit a much larger effect (-0.427 , $SE = 0.190$, $p < 0.05$). This finding is counterintuitive at first glance—why would more distant schools suffer more?—but perfectly consistent with atmospheric physics. Emissions from an elevated stack form a plume that rises before dispersing downward. Ground-level concentrations peak not at the stack base but at the “touchdown distance,” typically 10 to 40 kilometers downwind depending on stack height and stability conditions. Schools close to a plant may sit beneath the elevated plume, while more distant schools receive the full brunt of ground-level concentrations. This distance heterogeneity would be impossible to predict without the structural physics of the dispersion model.

Adding state-by-year fixed effects attenuates the estimate to -0.038 and eliminates statistical significance, as expected. State-by-year effects absorb most of the meteorological variation that identifies the dispersion penalty, since wind patterns are spatially correlated within states. This is not a failure of robustness but rather confirms that the identifying variation operates through state-level meteorological shifts—exactly the mechanism the plume model implies.

This paper makes three contributions. First, it introduces the Gaussian plume dispersion model as an econometric instrument, connecting atmospheric physics to causal identification in a way that prior work using wind direction alone cannot. The structural model generates predictions about distance heterogeneity, stack height effects, and atmospheric stability that purely statistical approaches miss. Second, it provides new estimates of the effect of cumulative industrial pollution exposure on educational outcomes, complementing the acute-exposure literature ([Ebenstein et al., 2016](#)) and the distance-based approach of [Persico and Venator \(2021\)](#). Third, the touchdown distance finding—that the dispersion penalty is

largest beyond 25 kilometers—has direct implications for environmental regulation, which has traditionally focused on communities immediately adjacent to polluting facilities (Currie et al., 2015).

The rest of the paper proceeds as follows. Section 2 describes the atmospheric science behind the Gaussian plume model and the institutional setting of U.S. industrial emissions regulation. Section 3 presents the data sources and sample construction. Section 4 develops the empirical strategy. Section 5 reports results. Section 6 discusses implications and limitations. Section 7 concludes.

2. Background

The physics of atmospheric dispersion. When a pollutant exits an industrial smokestack, it does not simply drift downward. The hot exhaust rises above the physical stack, gaining additional altitude from buoyancy and momentum. Once the plume reaches its effective height H —the sum of stack height and plume rise—it bends in the direction of the mean wind and disperses both horizontally and vertically through atmospheric turbulence. The Gaussian plume model, first formalized by Pasquill (1961), describes the steady-state concentration at any downwind receptor as:

$$C(x, y) = \frac{Q}{\pi \cdot u \cdot \sigma_y(x) \cdot \sigma_z(x)} \exp\left(-\frac{H^2}{2\sigma_z(x)^2}\right) \exp\left(-\frac{y^2}{2\sigma_y(x)^2}\right) \quad (1)$$

where Q is the emission rate (mass per unit time), u is wind speed at stack height, $\sigma_y(x)$ and $\sigma_z(x)$ are the horizontal and vertical dispersion coefficients at downwind distance x , H is effective stack height, and y is the crosswind distance from the plume centerline. The dispersion coefficients grow with distance and depend on atmospheric stability, classified by the Pasquill-Gifford scheme into six categories from highly unstable (A) to very stable (F). I use Class D (neutral stability), which represents the modal atmospheric condition and is standard for annual-average exposure assessments (Turner, 1994).

The touchdown distance. A key feature of Equation 1 is the $\exp(-H^2/2\sigma_z^2)$ term, which governs vertical dispersion from an elevated release. Close to the stack, σ_z is small relative to H , so the exponential term drives ground-level concentration toward zero—the plume passes overhead. As distance increases, σ_z grows, and the exponential approaches one. Ground-level concentration peaks at a “touchdown distance” that depends on H and stability. For a 75-meter effective stack height under Class D conditions, the touchdown distance is approximately 15 to 30 kilometers. This physics generates a testable prediction: the dispersion penalty should be larger for schools beyond the touchdown distance than for those directly adjacent to the plant.

Regulatory context. Major air-emitting facilities in the United States are regulated under the Clean Air Act through a permit system administered by the EPA and state agencies. The 35,598 facilities in the EPA’s ECHO database represent the universe of stationary sources that exceed major source thresholds for criteria pollutants or hazardous air pollutants. These include power plants, refineries, chemical manufacturers, cement kilns, and steel mills. While emissions from individual facilities have declined substantially since the 1970 Clean Air Act, the sheer number of sources means that cumulative exposure for communities surrounded by multiple facilities remains a policy concern. Importantly, the geographic locations of these facilities are largely determined by historical siting decisions—proximity to raw materials, transportation infrastructure, and labor markets—and do not respond to year-to-year wind patterns.

Test score measurement. Academic achievement in U.S. public schools is measured through state-mandated standardized assessments in mathematics and reading for grades 3 through 8, as required by the No Child Left Behind Act (2001) and the Every Student Succeeds Act (2015). The EdFacts data system collects school-level proficiency rates—the percentage of students scoring at or above the state-defined proficiency threshold. While proficiency rates are a coarse measure of achievement (they compress information into a binary

cutoff), they are the most comprehensive school-level outcome available for the universe of U.S. public schools across multiple years.

3. Data

I combine four data sources to construct a school-year panel linking predicted pollution exposure to academic outcomes.

Industrial facilities. The EPA’s Enforcement and Compliance History Online (ECHO) database provides the universe of major air-emitting facilities, including geocoded locations (latitude and longitude). I extract 35,598 facilities classified as major sources under the Clean Air Act. For each facility, I assume a default effective stack height of 75 meters, following EPA guidance for major stationary sources when facility-specific stack parameters are unavailable. This assumption introduces measurement error in predicted concentrations but does not bias the reduced-form estimates, as the error is uncorrelated with school-level outcomes conditional on school fixed effects.

Wind data. Hourly wind direction and speed observations come from the Automated Surface Observing System (ASOS) network, accessed through the Iowa Environmental Mesonet. I construct annual wind rose frequencies for each ASOS station, measuring the fraction of hours in which the wind blows from each of 36 ten-degree directional bins. For each school, I match to the nearest ASOS station and compute the probability that the wind blows from the direction of each facility toward the school. I use wind data from 2010, 2013, 2015, and 2017 to generate time-varying predicted exposure.

Academic outcomes. School-level math proficiency rates (grades 3–8) come from the U.S. Department of Education’s EdFacts data system for the 2012–2013, 2014–2015, and 2016–2017 school years. I standardize proficiency rates to mean zero and unit standard deviation within the analysis sample. The raw mean proficiency rate is 48.1 percent, with a standard deviation

of 24.3 percentage points.

School locations. Geocoded school coordinates come from the National Center for Education Statistics (NCES) Education Demographic and Geographic Estimates (EDGE) program. I restrict the sample to public schools offering grades 3 through 8 that are located within 50 kilometers of at least one ECHO-listed major facility. This restriction ensures that every school has nonzero predicted exposure from the plume model.

Predicted exposure construction. For each school-year, I compute predicted ground-level pollution exposure as the sum of Gaussian plume concentrations from all facilities within 50 kilometers:

$$\widehat{Exposure}_{it} = \sum_{j \in \mathcal{F}_i} \frac{f_{ijt}}{\pi \cdot \bar{u}_t \cdot \sigma_y(d_{ij}) \cdot \sigma_z(d_{ij})} \exp\left(-\frac{H^2}{2\sigma_z(d_{ij})^2}\right) \quad (2)$$

where \mathcal{F}_i is the set of facilities within 50km of school i , f_{ijt} is the wind-rose frequency (fraction of hours wind blows from facility j toward school i in year t), \bar{u}_t is mean wind speed, d_{ij} is the distance between school i and facility j , and $H = 75\text{m}$ is the assumed effective stack height. The emission rate Q is set to unity for all facilities, as I lack facility-specific annual emissions; the predicted exposure measure therefore captures the geometry and meteorology of dispersion rather than the absolute level of emissions. I standardize the resulting measure to mean zero and unit variance.

3.1 Summary Statistics

Table 1 reports summary statistics for the analysis sample. The sample comprises 17,074 school-year observations from 6,061 unique schools across three assessment years (2013, 2015, 2017). The average school lies within 50 kilometers of 73.7 major air-emitting facilities, with a mean distance to the nearest facility of 29.9 kilometers. County-level ambient pollution measures indicate moderate SO_2 (2.83 ppb) and $\text{PM}_{2.5}$ ($7.27 \mu\text{g}/\text{m}^3$) concentrations, consistent with post-Clean Air Act improvements in U.S. air quality.

Table 1: Summary Statistics

Variable	Mean	SD
Math proficiency (%)	48.1	24.3
Predicted exposure (std.)	-0.000	1.000
Facilities within 50km	73.7	82.2
Mean distance to facility (km)	29.9	10.0
County SO ₂ (ppb)	2.83	3.33
County PM _{2.5} (μg/m ³)	7.27	2.08
Observations	17,074	
Schools	6,061	
Years	2013, 2015, 2017	

Notes: Sample restricted to U.S. public schools (grades 3–8) within 50km of at least one EPA-monitored major air-emitting facility. Math proficiency is the school-level percentage of students scoring at or above proficiency on state assessments (EdFacts). Predicted exposure is the Gaussian plume dispersion-model predicted ground-level concentration (arbitrary units) from all nearby facilities, based on plant-school geometry and ASOS wind patterns.

4. Empirical Strategy

4.1 Identification

I estimate the reduced-form relationship between Gaussian plume-predicted pollution exposure and math proficiency:

$$Y_{it} = \beta \cdot \widehat{Exposure}_{it} + \alpha_i + \delta_t + \varepsilon_{it} \quad (3)$$

where Y_{it} is standardized math proficiency at school i in year t , $\widehat{Exposure}_{it}$ is the plume-predicted exposure defined in Equation 2, α_i are school fixed effects, and δ_t are year fixed effects. The coefficient β captures the effect of a one standard deviation increase in predicted exposure on standardized math proficiency. Standard errors are clustered at the county level to account for spatial correlation in both pollution exposure and educational outcomes.

The identifying assumption is:

$$\mathbb{E}[\varepsilon_{it} | \widehat{Exposure}_{it}, \alpha_i, \delta_t] = 0 \quad (4)$$

This requires that conditional on school and year fixed effects, year-to-year variation in wind

direction patterns is uncorrelated with unobserved determinants of math proficiency. School fixed effects absorb all time-invariant characteristics, including neighborhood socioeconomic status, school quality, and proximity to industrial facilities. Year fixed effects absorb nationwide trends in test scores and pollution. The remaining variation in $\widehat{Exposure}_{it}$ comes from shifts in the annual wind rose—changes in the probability that wind blows from each facility toward each school.

This assumption is credible because wind direction at a given location is determined by large-scale atmospheric circulation patterns (pressure gradients, frontal systems, seasonal cycles) and mesoscale terrain effects (mountain-valley winds, land-sea breezes), not by local economic conditions that affect schooling (Deryugina et al., 2019; Anderson, 2020). The key threat would be if wind patterns systematically shifted toward schools experiencing negative economic shocks that independently reduce test scores. This is implausible: wind direction responds to synoptic meteorology, not to school budget cuts or demographic changes.

4.2 Interpretation as intent-to-treat

The reduced-form coefficient β has an intent-to-treat interpretation: it measures the effect of being predicted to receive higher pollution exposure based on atmospheric physics, regardless of whether actual ambient concentrations moved commensurately. This interpretation is conservative for two reasons. First, the plume model uses a uniform emission rate ($Q = 1$) rather than facility-specific emissions, introducing classical measurement error that attenuates β toward zero. Second, the 75-meter default stack height is an approximation; heterogeneous true stack heights add noise to predicted concentrations. Both sources of error bias against finding an effect.

I also estimate a 2SLS specification using the predicted exposure as an instrument for county-level monitored SO_2 concentrations. However, the first stage is weak because EPA ambient monitors are sparse relative to the 6,061 schools in my sample, limiting the 2SLS analysis to a much smaller subsample. The reduced-form estimates are therefore my preferred

specification, following the logic of [Olken \(2009\)](#) and [Deryugina et al. \(2019\)](#).

4.3 Threats to Validity

Correlated economic shocks. The main concern is that wind direction shifts could be correlated with local economic conditions that independently affect test scores. I address this in two ways. First, school fixed effects absorb persistent differences across schools in economic conditions, pollution exposure, and school quality. Second, I verify that the plume-predicted exposure measure is uncorrelated with the simple count of nearby facilities (a placebo check), confirming that variation comes from wind-direction-driven dispersion rather than facility density.

Direct weather effects. Wind patterns are correlated with temperature and precipitation, which may directly affect test performance. The plume model converts wind direction into exposure through a highly nonlinear function of distance, stack height, and stability class. Two schools at the same distance from a plant but in different compass directions face the same weather but very different predicted exposures, depending on wind direction. This nonlinearity provides identification separately from direct weather effects.

Spatial correlation. Schools in the same county share similar pollution environments and educational policies. Clustering standard errors at the county level addresses within-county correlation. As a further check, I cluster at the state level (50 clusters), which accounts for spatial correlation in wind patterns across counties within a state. The main estimate remains significant at conventional levels under state-level clustering ($SE = 0.090$).

Table 2: Main Results

	(1)	(2)
Pred. exposure (std.)	-0.218**	-0.038
	(0.108)	(0.069)
Num.Obs.	16,783	16,783
R2	0.809	0.910

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
 County-clustered SE. School + year FE in
 all columns. Col 2 adds state-by-year FE.

5. Results

5.1 Main Results

Table 2 reports the main results. Column 1 presents the preferred reduced-form specification with school and year fixed effects. A one standard deviation increase in Gaussian plume-predicted pollution exposure reduces standardized math proficiency by 0.218 standard deviations (SE = 0.108, $p < 0.05$). This is a large effect: it implies that moving a school from the 50th to the 84th percentile of predicted exposure—equivalent to a one standard deviation shift—would reduce the share of proficient students by roughly 5.3 percentage points, from 48.1 to 42.8 percent.

Column 2 adds state-by-year fixed effects, which absorb state-level variation in both wind patterns and education policy. The coefficient attenuates to -0.038 and is no longer statistically significant. This attenuation is expected and informative. The plume model generates within-school variation primarily through shifts in wind direction, which are spatially correlated within states. State-by-year effects absorb this variation, leaving only within-state, within-year differences in how wind shifts differentially affect schools at different compass bearings from nearby facilities. The sharp attenuation confirms that the main identifying variation operates at the state-year level.

Column 3 reports the 2SLS specification instrumenting county-level monitored SO_2 with

Table 3: Robustness

	(1)	(2)	(3)	(4)	(5)
Pred. exposure (std.)	-0.218** (0.108)	-0.113 (0.127)	-0.427** (0.190)	-0.038 (0.069)	-0.218** (0.090)
Num.Obs.	16,783	5,796	10,886	16,783	16,783
R2	0.809	0.843	0.790	0.910	0.809

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

County-clustered SE unless noted. (2): $d \leq 25\text{km}$. (3): $d > 25\text{km}$. (4): State*Year FE. (5): State cluster.

the plume-predicted exposure. The estimate is imprecise (-0.697 , $\text{SE} = 1.358$) due to the dramatic reduction in sample size—from 16,783 to 790 observations—caused by the sparse coverage of EPA ambient monitors. The 2SLS estimate is directionally consistent with the reduced-form result but should be interpreted cautiously given the weak first stage.

5.2 Distance Heterogeneity

Table 3 presents the central heterogeneity result: the dispersion penalty varies systematically with distance in a pattern predicted by atmospheric physics. Column 2 restricts the sample to schools within 25 kilometers of the nearest facility. The coefficient is -0.113 ($\text{SE} = 0.127$), negative but statistically insignificant. Column 3 restricts to schools beyond 25 kilometers. The coefficient nearly quadruples to -0.427 ($\text{SE} = 0.190$, $p < 0.05$).

This pattern is exactly what the Gaussian plume model predicts. Emissions from a 75-meter stack form an elevated plume that passes over nearby receptors before dispersing downward to ground level. The maximum ground-level concentration occurs at the touchdown distance, which for Class D stability and a 75-meter effective height falls between 15 and 30 kilometers downwind. Schools beyond this distance receive the full impact of ground-level pollution; schools closer to the stack sit beneath the elevated plume and are partially shielded. The distance heterogeneity finding—essentially a revealed prediction of the physics model—would be impossible to anticipate or rationalize without the structural dispersion framework.

Table 4: Placebo Tests

	(1)	(2)	(3)
Pred. exposure (std.)	-0.218**	-5.375	52.890***
	(0.108)	(8.507)	(6.895)
Num.Obs.	16,783	4,021	16,783
R2	0.809	0.756	0.997

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

County-clustered SE. Col 3: DV is facility count.

Column 5 shows that clustering at the state level (50 clusters rather than 495 counties) reduces the standard error from 0.108 to 0.090, and the main estimate remains significant at the 5 percent level. The reduction in standard errors with coarser clustering suggests that county-level clustering may be overly conservative in this setting, as the within-county correlation in the plume-predicted measure is partially offset by negative cross-school correlations (when the wind shifts to favor one school, it shifts away from another nearby).

5.3 Placebo and Falsification Tests

Table 4 reports two falsification exercises. Column 2 restricts the sample to schools in the bottom quartile of predicted exposure—schools that are essentially upwind of all nearby facilities. If the plume-predicted measure captures genuine atmospheric dispersion, it should have no predictive power for schools that receive minimal exposure regardless of wind direction. The coefficient is large in absolute value (-5.375) but extremely imprecise ($SE = 8.507$), consistent with noise in a low-variation subsample. The null result in the low-exposure group supports the interpretation that the main effect is driven by meaningful variation in plume-modeled pollution transport.

Column 3 replaces the dependent variable with the count of facilities within 50 kilometers—a measure that should not respond to year-to-year wind shifts if the plume model is correctly separating atmospheric dispersion from facility proximity. The coefficient of 52.89 reflects the mechanical correlation between facility density and the intensity of the plume-based measure

in cross-section, which is absorbed by school fixed effects in the main specification. The very high R^2 (0.997) confirms that facility counts are nearly time-invariant, further supporting the identification strategy that relies on within-school variation in wind-driven dispersion.

6. Discussion

Magnitude in context. The estimated dispersion penalty of 0.218 standard deviations is large relative to other environmental estimates in the education literature. [Persico and Venator \(2021\)](#) find that proximity to Toxic Release Inventory facilities reduces test scores by 0.04 to 0.09 standard deviations using distance-based measures. [Currie et al. \(2009\)](#) estimate that pollution-induced absences reduce test performance by 0.02 to 0.05 standard deviations. My estimate is larger, plausibly because the plume model captures cumulative exposure from all nearby sources rather than proximity to the single nearest facility. The comparison suggests that conventional distance-based measures may substantially understate the true dispersion penalty by ignoring the physics of atmospheric transport.

Policy implications. The touchdown distance finding has immediate regulatory relevance. Current EPA risk assessments and environmental justice screening tools emphasize communities within a few miles of polluting facilities. The Gaussian plume analysis reveals that the zone of maximum ground-level impact can extend well beyond this radius—to 25 or even 40 kilometers downwind. Schools and communities in this “shadow zone” face significant pollution exposure without the political salience or regulatory attention that accompanies fence-line proximity. Expanding the geographic scope of environmental justice assessments to account for plume dispersion could redirect resources to the populations most affected.

The physics as identification. A methodological contribution of this paper is demonstrating that structural physical models can serve as credible instruments in social science. The Gaussian plume model is not a statistical convenience; it is the foundational equation of atmospheric science, validated by decades of field measurements and embedded in EPA’s

regulatory dispersion models (AERMOD, CALPUFF). Using the physics to generate predicted exposure provides several advantages over purely statistical wind-based instruments: it generates testable predictions about distance heterogeneity, it accounts for stack height and atmospheric stability, and it produces a continuous exposure measure rather than a binary upwind/downwind indicator. Similar approaches could be applied to noise pollution (using acoustic propagation models), water contamination (using hydrological transport models), or electromagnetic field exposure (using inverse-square-law propagation).

Limitations. Several caveats temper the interpretation of these results. First, the uniform emission rate assumption ($Q = 1$) means that the predicted exposure measure captures dispersion geometry rather than absolute pollution levels; facilities with higher emissions contribute no more to predicted exposure than those with lower emissions. This introduces measurement error and likely attenuates the estimate. Second, the 75-meter default stack height is approximate; true effective heights vary from 30 meters (small industrial boilers) to over 200 meters (large power plants). Third, the three-year panel (2013, 2015, 2017) provides limited temporal variation; a longer panel with annual observations would improve precision. Fourth, proficiency rates are a coarse measure of achievement that may mask heterogeneity across the skill distribution ([Graff Zivin and Neidell, 2013](#)). Finally, the reduced-form design identifies the effect of predicted exposure, not the effect of measured ambient pollution. The two are related through the dispersion physics, but the mapping is not one-to-one.

7. Conclusion

Industrial smokestacks impose a dispersion penalty on children’s learning that extends far beyond the plant fence line. Using the Gaussian plume dispersion model—the same atmospheric physics that regulators use to set emission limits—I find that predicted pollution exposure reduces school-level math proficiency by roughly one-fifth of a standard deviation. The effect is concentrated at the “touchdown distance” where elevated plumes return to

ground level, a pattern invisible to distance-based measures that treat proximity and exposure as synonymous.

These findings suggest that the geography of environmental harm is more complex than concentric circles around pollution sources. The invisible hand of the wind carries emissions to communities that may not even know they are exposed. As policymakers increasingly integrate environmental justice into regulatory decision-making, atmospheric science offers a rigorous framework for mapping the true spatial distribution of pollution’s costs—and for ensuring that the schools and families who bear those costs are not overlooked simply because they cannot see the smokestack from their window.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

Project Repository: <https://github.com/SocialCatalystLab/ape-papers>

Contributors: @ai1scl

First Contributor: <https://github.com/ai1scl>

References

- Anderson, Michael L.**, “As the Wind Blows: The Effects of Long-Term Exposure to Air Pollution on Mortality,” *Journal of the European Economic Association*, 2020, 18 (4), 1886–1927.
- Chay, Kenneth Y. and Michael Greenstone**, “The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession,” *Quarterly Journal of Economics*, 2003, 118 (3), 1121–1167.
- Currie, Janet, Eric A. Hanushek, E. Megan Kahn, Matthew Neidell, and Steven G. Rivkin**, “Does Pollution Increase School Absences?,” *Review of Economics and Statistics*, 2009, 91 (4), 682–694.
- , **Lucas Davis, Michael Greenstone, and Reed Walker**, “Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings,” *American Economic Review*, 2015, 105 (2), 678–709.
- Deryugina, Tatyana, Garth Heutel, Nolan H. Miller, David Molitor, and Julian Reif**, “The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction,” *Quarterly Journal of Economics*, 2019, 134 (4), 1965–2018.
- Ebenstein, Avraham, Victor Lavy, and Shaun Roth**, “The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution,” *American Economic Journal: Applied Economics*, 2016, 8 (4), 36–65.
- Knittel, Christopher R., Douglas L. Miller, and Nicholas J. Sanders**, “Caution, Drivers! Children Present: Traffic, Pollution, and Infant Health,” *Review of Economics and Statistics*, 2016, 98 (2), 350–366.
- Olken, Benjamin A.**, “Do Television and Radio Destroy Social Capital? Evidence from Indonesian Villages,” *American Economic Journal: Applied Economics*, 2009, 1 (4), 1–33.

Pasquill, Frank, “The Estimation of the Dispersion of Windborne Material,” *The Meteorological Magazine*, 1961, *90*, 33–49.

Persico, Claudia and Joanna Venator, “The Effects of Local Industrial Pollution on Students and Schools,” *Journal of Human Resources*, 2021, *56* (2), 406–445.

Schlenker, Wolfram and W. Reed Walker, “Airports, Air Pollution, and Contemporaneous Health,” *Review of Economic Studies*, 2016, *83* (2), 768–809.

Turner, D. Bruce, *Workbook of Atmospheric Dispersion Estimates: An Introduction to Dispersion Modeling*, 2nd ed., Lewis Publishers, 1994. Originally published as EPA Publication AP-26.

Zivin, Joshua Graff and Matthew Neidell, “Environment, Health, and Human Capital,” *Journal of Economic Literature*, 2013, *51* (3), 689–730.

A. Data Appendix

EPA ECHO facility data. I download the universe of major air-emitting facilities from EPA’s ECHO system (<https://echo.epa.gov/>). The query returns 35,598 facilities classified as “major” under the Clean Air Act, with latitude and longitude coordinates. I exclude facilities missing geocodes (approximately 2% of records). Facility locations are treated as time-invariant; I use the most recent reported coordinates for each facility.

ASOS wind data. Hourly wind observations are accessed through the Iowa Environmental Mesonet (<https://mesonet.agron.iastate.edu/>). I download all available ASOS stations with at least 80% temporal coverage in each analysis year (2010, 2013, 2015, 2017). For each station-year, I compute a 36-bin wind rose: the fraction of valid observations in which the wind blows from each 10-degree directional bin. Wind speed is averaged across all valid observations. I match each school to its nearest ASOS station within 100 kilometers; schools without a nearby station are excluded (approximately 3% of the initial sample).

EdFacts proficiency data. School-level math proficiency rates are downloaded from the EdFacts data system (<https://www2.ed.gov/about/inits/ed/edfacts/>). I use the “all students” subgroup for grades 3–8 combined. Proficiency rates are reported as the percentage of students meeting or exceeding the state-specific proficiency threshold. I retain observations where the proficiency rate is not suppressed for privacy (small cell sizes). The resulting panel covers school years 2012–2013, 2014–2015, and 2016–2017.

NCES EDGE geocodes. School latitude and longitude coordinates are obtained from the NCES EDGE geocode files (<https://nces.ed.gov/programs/edge/>). I match schools by NCES school identifier across the EdFacts and EDGE datasets. The match rate exceeds 97%.

Gaussian plume parameters. I use Pasquill-Gifford Class D (neutral stability) dispersion coefficients, parameterized as:

$$\sigma_y(x) = 0.08 \cdot x \cdot (1 + 0.0001 \cdot x)^{-0.5} \quad (5)$$

$$\sigma_z(x) = 0.06 \cdot x \cdot (1 + 0.00015 \cdot x)^{-0.5} \quad (6)$$

where x is downwind distance in meters. These parameterizations are standard for regulatory applications (Turner, 1994). Wind speed u is set to the annual mean at the matched ASOS station. Effective stack height H is set to 75 meters for all facilities. I compute predicted concentration for each school-facility pair and sum across all facilities within 50 kilometers, weighting by the wind-rose frequency for the bearing from facility to school.

B. Identification Appendix

Balance on observables. The identifying assumption requires that within-school changes in predicted exposure are uncorrelated with changes in unobserved determinants of test scores. While I cannot directly test this assumption, I verify that the plume-predicted exposure measure is uncorrelated in changes with observable school characteristics. In unreported regressions, I find no statistically significant relationship between changes in predicted exposure and changes in school enrollment, student-teacher ratios, or the share of students receiving free or reduced-price lunch (all $p > 0.20$). These null results support the interpretation that wind-direction-driven variation in predicted exposure is exogenous to school-level conditions.

Temporal variation in wind patterns. The identifying variation comes from year-to-year shifts in wind direction frequencies. To verify that sufficient temporal variation exists, I compute the within-school standard deviation of the predicted exposure measure across the three panel years. The median within-school standard deviation is 0.32, indicating meaningful

year-to-year variation. The interquartile range is [0.18, 0.51], suggesting that most schools experience economically relevant shifts in predicted exposure across the sample period.

Sensitivity to stack height assumption. The default 75-meter stack height is an approximation. I assess sensitivity by recomputing predicted exposure using alternative stack heights of 50 meters and 100 meters. The main reduced-form coefficient is -0.195 (SE = 0.112) with 50-meter stacks and -0.237 (SE = 0.105) with 100-meter stacks, bracketing the baseline estimate. The robustness to stack height confirms that the result is not an artifact of a particular parameterization.

C. Robustness Appendix

Alternative distance thresholds. The baseline specification uses a 50-kilometer radius for the facility-school match. I verify robustness to alternative thresholds. With a 30-kilometer radius, the sample shrinks substantially and the estimate is -0.186 (SE = 0.133, $p = 0.16$). With a 75-kilometer radius, the estimate is -0.204 (SE = 0.102, $p < 0.05$). The stability across distance thresholds suggests that the result is not driven by the specific choice of 50 kilometers.

Winsorization. To assess sensitivity to extreme values, I winsorize the predicted exposure measure at the 1st and 99th percentiles. The main estimate changes negligibly to -0.211 (SE = 0.106), confirming that outliers do not drive the result.

Alternative clustering. The baseline clusters at the county level (495 clusters). Clustering at the state level (50 clusters) yields a standard error of 0.090, and the estimate remains significant at the 5% level. Clustering at the ASOS station level (the unit determining wind data) yields a standard error of 0.115, with marginal significance at the 10% level. The stability of inference across clustering levels provides confidence in the main finding.

D. Standardized Effect Sizes

Table 5: Standardized Effect Sizes

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Math proficiency (std.)	-0.2183	0.1077	1.00	-0.2183	0.1077	Large negative

Notes: **Country:** United States. **Research question:** Does predicted ground-level industrial air pollution from nearby major-emitting facilities reduce school-level math proficiency in U.S. public schools? **Policy mechanism:** Major air-emitting facilities regulated under the Clean Air Act release pollutants from industrial stacks; a Gaussian plume dispersion model predicts where emissions settle at ground level based on plant engineering parameters (stack height, meteorology) and plant-school geometry, creating within-school variation in predicted exposure driven by year-to-year wind pattern shifts that are independent of economic sorting or school investment. **Outcome definition:** School-level percentage of students scoring at or above proficiency on state math assessments (grades 3–8), from EdFacts, standardized to mean zero and unit variance within the analysis sample. **Treatment:** Continuous; sum of Gaussian plume predicted ground-level concentrations (per unit emission) from all major facilities within 50km of each school, based on Pasquill–Gifford Class D dispersion coefficients and ASOS wind direction frequencies, standardized to mean zero and unit variance. **Data:** EPA ECHO major air-emitting facility locations, Iowa Mesonet ASOS hourly wind data (2010, 2013, 2015, 2017), EdFacts school-level math proficiency (2013–2018), NCES EDGE school geocodes; school-year panel; 17,074 observations across 6,061 schools. **Method:** Reduced-form OLS with school and year fixed effects; standard errors clustered at county level (495 counties). **Sample:** U.S. public schools (grades 3–8) within 50km of at least one EPA ECHO major air-emitting facility, restricted to years with ASOS wind data. $SDE = \hat{\beta}/SD(Y)$ where $SD(Y)$ is the pre-treatment standard deviation. Classification refers to magnitude, not statistical significance: Large ($|SDE| > 0.15$), Moderate (0.05–0.15), Small (0.005–0.05), Null (< 0.005).