

Rachel Inafuku  
Timothy J. Halliday  
Lester Lusher  
Aureo de Paula

25/07

Working paper

# **Vog: using volcanic eruptions to estimate the impact of air pollution on student test scores**

# Vog: Using Volcanic Eruptions to Estimate the Impact of Air Pollution on Student Test Scores \*

Rachel Inafuku  
UHERO

Timothy J. Halliday<sup>†</sup>  
University of Hawai‘i at Mānoa  
UHERO  
IZA

Lester Lusher  
University of Pittsburgh  
IZA

Aureo de Paula  
University College London

February 20, 2025

## Abstract

We pair variation stemming from volcanic eruptions with the census of Hawai‘i’s public schools student test scores to estimate the impact of  $PM_{2.5}$  and  $SO_2$  on student performance. Increased particulate pollution decreases test scores. These results are concentrated among schools with the highest long-term average levels of pollution. The effects of  $PM_{2.5}$  are larger for the poorest pupils by a factor of at least three. We demonstrate that poor air quality disproportionately impacts the human capital accumulation of economically disadvantaged children.

Key Words: Vog, Particulates, Test Scores, Kriging, Environmental Justice

JEL Classification: I22, I24, Q52

Word Count: 10491

---

\*We thank seminar participants at the University of Hawai‘i Applied Micro Workshop, the Society of Labor Economists Annual Meetings in Philadelphia, and the Annual Workshop on Econometrics and Health Economics in Manila for useful comments. We are especially indebted to Hawai‘i Data eXchange Partnership for allowing us to access the data and for strong support of this research. This research is supported by the Institute of Education Sciences, U.S. Department of Education, through Grant 84.372A (Statewide Longitudinal Data Systems) to the Hawai‘i State Department of Education. We thank Kristin Butcher, Andrew Foster, Timo Hener, Scott Imberman, Adriana Lleras-Muney, Hannes Schwandt, and Hitoshi Shigeoka for valuable feedback.

<sup>†</sup>Corresponding Author. Address: 2424 Maile Way; 533 Saunders Hall; Honolulu, HI 96822. e-mail: halliday@hawaii.edu

# 1 Introduction

Researchers and policymakers have increasingly sought to understand the consequences of poor air quality. An abundance of evidence suggests pollution can have severe adverse effects on health, fertility, and mortality outcomes.<sup>1</sup> A smaller yet growing set of studies has identified labor productivity losses, where pollution harms workers across both physically demanding (e.g. fruit picking ([Graff Zivin and Neidell, 2012](#)) and pear-packing ([Chang et al., 2016](#))) and mentally demanding (e.g. baseball umpiring ([Archsmith et al., 2018](#))) occupations.<sup>2</sup>

In this study, we investigate how air quality affects student performance on standardized tests. Student scores not only are a marker of cognitive performance but they also often have long term consequences. A better understanding of this topic, therefore, has implications for how air quality impacts human capital acquisition and subsequent labor market outcomes and we add to the literature on the topic. There is some evidence of air pollution sharply reducing cognitive performance ([Zhang et al., 2018](#)). There is also a growing number of studies on the effects of specific pollutants (primarily particulates) on student test scores. Work from [Ebenstein et al. \(2016\)](#) uses Israeli data from 2000 to 2002 to find drops in high school exit exam test scores and worsened longer run outcomes in response to poor air quality. Similar results come from [Amanzadeh et al. \(2020\)](#) who utilize data from Iran. [Carneiro et al. \(2021\)](#) show that higher concentrations of particulates result in lower scores on college entry examinations in Brazil. Using data from the US, [Marcotte \(2017\)](#) found decreased performance among kindergartners on testing days with both worse pollen and fine airborne particulate matter. [Pham and Roach \(2023\)](#) utilize natural variation across the US to investigate how fine particulate pollu-

---

<sup>1</sup>Several studies have found a positive association between pollution and fertility abnormalities ([Nieuwenhuisen et al., 2014](#); [Slama et al., 2013](#); [Perin et al., 2010](#)). See [Carré et al. \(2017\)](#) for a review of the literature. [Burnett et al. \(1999\)](#) and [Koken et al. \(2003\)](#) find increases in air pollution leads to an increase in cardiorespiratory hospitalizations. [Linares et al. \(2010\)](#) find children who attend schools closer to major air pollution sources were more likely to develop respiratory and lung abnormalities. [Di et al. \(2017\)](#) finds increases in pollution (even at levels below the national standard) were associated with an increase in mortality for US Medicare beneficiaries, especially amongst racial minority groups.

<sup>2</sup>Other studies examining labor productivity loss include [He et al. \(2019\)](#) (textile industry workers), [Chang et al. \(2019\)](#) (call center workers), and [Lichter et al. \(2017\)](#) (professional soccer players). For a comprehensive review of the literature on “non-health” effects of air pollution, we refer the reader to [Aguilar-Gomez et al. \(2022\)](#).

tion harms third through eighth grader achievement.<sup>3</sup> Outside student performance outcomes, related studies from [Currie et al. \(2009\)](#), [Liu and Salvo \(2018\)](#), [Chen et al. \(2018\)](#), and [Komisarow and Pakhtigian \(2022\)](#) find increased student absences in response to poor air quality. A related literature looks at the effects of various “human made” pollutants and interventions on student outcomes. For example, [Stafford \(2015\)](#) investigates how school renovations, such as mold remediation and ventilation improvements, impacted student test scores in Texas. [Austin et al. \(2019\)](#) identify the effects of retrofitting school busses to reduce emissions on student test scores in Georgia. [Persico and Venator \(2021\)](#) investigate how the introduction and disappearance of a local Toxic Release Inventory affects student test scores in Florida. Also using data from Florida, [Heissel et al. \(2022\)](#) identify the effects of traffic on student test scores and other shorter run outcomes. [Gilraine and Zheng \(2022\)](#) identify the effects of installing air filters in classrooms in Los Angeles. Finally, [Duque and Gilraine \(2022\)](#) investigate how the presence of coal power plants affect math test scores in North Carolina, and similarly, [Jacqz \(2022\)](#) identify the effects of toxic chemical releases on student test score performance (ten years later).

An important feature of our study is that we pay close attention to how poor air quality affects poorer students within schools. It has long been understood that air pollution disproportionately impacts the poor and disadvantaged minorities in the United States despite recent progress ([Currie et al., 2020](#)). Importantly, work that demonstrates disproportionate effects of poor air quality on test scores by socioeconomic status is limited. Moreover, the results that we do have often conflict, leaving the issue still up for debate. [Ebenstein et al. \(2016\)](#) finds that Israeli students of lower socioeconomic status experience larger declines in test scores due to higher pollution. They attribute this finding to higher rates of asthma among those of lower socioeconomic classes.<sup>4</sup> However, in the US, seminal work by [Case et al. \(2002\)](#) shows that asthma is more prevalent among children with richer parents in the United States. On the other

---

<sup>3</sup>At the post-secondary level, [Bedi et al. \(2021\)](#) investigate the impacts of  $PM_{2.5}$  on grammatical reasoning tests of university students in Brazil, while [Yao et al. \(2023\)](#) look at university student test scores from China’s College English Test.

<sup>4</sup>[Marcotte \(2017\)](#) also shows that the effects of particulate pollution are largest for asthmatic students.

hand, [Heissel et al. \(2022\)](#) show that economically disadvantaged students (as proxied by eligibility for federal programs such as free and reduced lunch) experience smaller impacts on test scores compared to their more advantaged peers; however, they also experience more absences and behavioral issues. Finally, [Wen and Burke \(2022\)](#) do document socioeconomic gradients in the effects of particulates on test scores using exogenous variation in particulates using school-level proxies for socioeconomic status.

The context of our study is the islands of O‘ahu and Hawai‘i in the state of Hawai‘i. This is a particularly advantageous setting for several reasons. One is Hawai‘i’s rich, plausibly exogenous variation in air quality. Another is that despite its reputation for moderate climate, Hawai‘i can claim ten of the world’s fourteen classifications for climate zones (microclimates) - the only place in the world with such diversity in one small area.<sup>5</sup>

Hawai‘i provides a unique and powerful opportunity to estimate the effects of two pollutants, particulate matter ( $PM_{2.5}$ ) and sulfur dioxide ( $SO_2$ ), on cognitive performance. We do so using  $SO_2$  emissions from Kilauea volcano which is located on the island of Hawai‘i. These gaseous emissions eventually form particulate matter in the form of sulfate aerosols. This pollution is called vog and is similar to smog pollution in many cities. Because this species of particulates is high in sulfuric acid, they resemble particulates from sources that produce sulfate aerosols such as coal-fired power plants.<sup>6</sup> Importantly, 8% of the world’s population faces potential risks from volcanic eruptions and so, our estimates will have a direct bearing on these other settings ([Choumert-Nkolo et al., 2021](#)).

The emission of  $SO_2$  from the Kilauea volcano represents a rare case of truly unpredictable variation in air pollution in the United States. Based on local wind conditions and whether the

---

<sup>5</sup>Source: [Hawai‘i Magazine](#), (accessed 16 Sep. 2020).

<sup>6</sup>See [Halliday et al. \(2019\)](#) for further detail on the similarities and differences between vog and other man-made pollutants. In short, smog that is composed of sulfate aerosols (such as smog near coal-fired power plants) is very similar to Kilauea vog. Still, there are differences based on particle size, shape, chemistry, and absorption of co-pollutant, though this is true of smog pollution across the US. For example, the smog in Pennsylvania cannot be directly compared to smog in New York or California, as there are differences in the physical and chemical properties of particulate pollution across regions and sources. Particulates in vog are likely to be more acidic than typical city smog, though the health implication for this is ambiguous. In general, Hawai‘i has much lower NOx, ozone, CO, soot, and volatile organics than other cities as well.

volcano is emitting, the air quality of Hawai‘i can shift from hazardous to pristine in a matter of hours across differing parts of the islands. Previous research has leveraged this high frequency variation on a day-to-day basis to find increased emergency room admissions due to respiratory reasons on days with higher pollution levels (Halliday et al., 2019).<sup>7</sup>

An additional advantage of our setting is that average pollution levels are far below Environmental Protection Agency (EPA) ambient air quality standards throughout most of the state. Identifying and understanding the effects of pollution at lower baseline levels is important as this can help to inform and potentially update EPA standards. Moreover, lower pollution levels also better reflect modal households in the US. While prior literature has focused entirely on air quality within higher-baseline polluted environments, average pollutant levels in our study are comparable to pollution within the US. In 2021, the US average seasonally-weighted concentration of particulate matter ( $PM_{2.5}$ ) pollution was 8.47 micrograms per cubic meter ( $\mu g/m^3$ ).<sup>8,9</sup> In our sample, Hawai‘i island saw similar mean  $PM_{2.5}$  levels with an average of  $8.08 \mu g/m^3$  while mean levels on O‘ahu are slightly lower at  $4.35 \mu g/m^3$ .

We pair this variation in particulates with the census of public school student test scores in the State of Hawai‘i. These data were obtained from the Hawai‘i Data eXchange Partnership (DXP), a collaboration between the University of Hawai‘i, the Executive Office of Early Learning, and the State of Hawai‘i Department of Education. Because we have a census, we will have enough power to detect even small impacts of air pollution. The data track students from elementary to middle and high-school from 2015 through 2018. Math and English literacy assessments are given in grades three through eight, and again in grade eleven. In total, the data include nearly 450,000 student-test-year observations. These data allow us to estimate day-of measures of air quality on student performance across varying ages, assessment types, and air quality conditions.

---

<sup>7</sup>Halliday et al. (2019) articulate the numerous advantages of using variation in vog to study the impact of pollution. For example, vog is emitted naturally, whereas the majority of the literature relies on variation in human activity (e.g. from cars, airplanes, factories) which may plausibly suffer from endogeneity biases. Another advantage comes from temporal variation: vog can vary on a day-to-day basis, whereas most other types of pollutants are highly serially correlated.

<sup>8</sup>For example, in Ebenstein et al. (2016) the average level of  $PM_{2.5}$  on student test days was  $21.05 \mu g/m^3$ .

<sup>9</sup>Source: US Environmental Protection Agency, (accessed 25 Apr. 2021).

An important feature of this study is that we employ three separate techniques to predict pollution exposure at a school. The first of these, Kriging, comes from geostatistics ([Cressie, 1990](#); [Montero et al., 2015](#)). The Kriging procedure leverages information on the spatial correlation in pollution as well as the distance between the relative locations of the pollution monitoring stations and the schools. We also exploit the fact that the presence of northeasterly winds, or trade winds, affects the spatial distribution of pollution in Hawai‘i. In general, trade winds lower pollution levels throughout most of the archipelago. While Kriging is common in the geostatistics literature, it is not common in environmental economics.<sup>10</sup> We also employ predictors that use inverse distance weighting and uniform weighting for monitoring stations that are near a school which are more common in economics. We show that Kriging works well near Kilauea on the island of Hawai‘i, particularly for  $SO_2$ , but performs poorly on the island of O‘ahu.

For the full sample of students, we estimate a small but statistically significant impact of particulates on student test scores, but these estimates mask a large degree of heterogeneity across a number of domains. Overall, we find that a one standard deviation increase in  $PM_{2.5}$  reduces test scores by 0.32-0.55 percent of a standard deviation. We then find that the effects are significantly tied to levels of pollution exposure within schools. For schools in the top quintile of long-term pollution exposure (as proxied by school-average pollution exposure over the duration of the sample), a one standard deviation increase in  $PM_{2.5}$  lowers test scores by between 0.53 and 0.72 percent of a standard deviation with most of the estimates closer to the upper range of this interval. In addition, formal tests indicate that the effects in the bottom 80% of the long-term pollution distribution are null. Moreover, when we examine the possibility of non-linearities in the relationship between test scores and pollution exposure, we find that these effects are concentrated in the upper quintiles of daily exposure. All told, this suggests that the damages from  $PM_{2.5}$  increase precipitously with both average long-term and daily exposure.

In the results that we report in this paper, we provide no evidence that  $SO_2$  adversely impacts student learning. However, we are careful to point out that  $SO_2$  could very well have impacts

---

<sup>10</sup>[Lleras-Muney \(2010\)](#) is the only economics study that we know of that employs the technique.

(particularly near Kilauea) but our data and research design may not be able to detect it.

We find that the effects of pollution are particularly concentrated among economically disadvantaged students. Poorer pupils experience at least three times the effect of  $PM_{2.5}$  when compared to their more advantaged counterparts. Interestingly, we find little difference in the effects of particulates across schools by the fraction of the school's students who were disadvantaged, suggesting that the economically disadvantaged student gap is not driven by differences in school resources (e.g. classroom air conditioning). We also note that this is less of an issue in Hawai'i as there is one statewide school district in which schools are not funded by local property taxes. So, we conclude that disadvantaged students *within* the same school are significantly more harmed by pollution than their more advantaged counterparts.

Importantly, [Wen and Burke \(2022\)](#) also document similar socioeconomic gradients in the effects of particulates on test scores. There are some important similarities and differences between our gradient estimates and theirs. They employ a similar research design to our own by studying the impact of cumulative exposure during the year on test scores using a school fixed effects design. They find that an additional  $10 \mu g/m^3$  of cumulative  $PM_{2.5}$  from forest fires leads to decreases in average test scores by 0.029 of a standard deviation. This is comparable to our short-term exposure estimates (see Table 4).<sup>11</sup> Their results also suggest a higher gradient for economically disadvantaged students in schools with a high percentage of non-whites than our estimates. Their estimates of the gradient for economically disadvantaged students in schools with a low percentage of non-whites are nonetheless lower than ours. We should note that while we are able to record whether each student is economically disadvantaged, [Wen and Burke \(2022\)](#) use a measure that is aggregated at the school level. We are thus able to offer more precise estimates.<sup>12</sup> Finally, when we use measures of disadvantage that are aggregated at the school-level (akin to what is used by [Wen and Burke \(2022\)](#)), we do not find any gradients.

---

<sup>11</sup>This refers to our estimates in Table 4 once appropriate adjustments are made. (As indicated in section 3.5, we scale our coefficient estimates up by 100 so our impact is reported per additional  $100 \mu g/m^3$  instead of  $10 \mu g/m^3$ .)

<sup>12</sup>Once adequate adjustments are made to reconcile our units of measure to changes in test scores standard deviations in response to  $10 \mu g/m^3$  increases in  $PM_{2.5}$ , our standard errors are smaller.



## 2 Data and Background

### 2.1 Student Test Scores from Hawai‘i

The context of our study is the state of Hawai‘i. US News ranks the pre-K through 12 education system in Hawaii at #33, nearby Rhode Island, Washington, North Dakota, and Arkansas.<sup>13</sup> Most statistics, however, suggest that the public schooling system in Hawai‘i severely lags behind other US states. Hawai‘i spends less on public education than any other state,<sup>14</sup> and more directly related to our study, around 3,400 classrooms in Hawai‘i still lack air conditioning.<sup>15</sup> Consequently, more students in Hawai‘i enroll in private schools compared to any other state (19% of students in Hawai‘i vs. 10% of students in the US).<sup>16</sup> Overall, Hawai‘i ranks 47th in the US in transition and completion rates from ninth grade to college.<sup>17</sup>

We collect student test score data from the Hawai‘i Data eXchange Partnership (DXP), a collaboration between five of Hawai‘i’s state agencies (Department of Health, Department of Labor and Industrial Relations, Department of Education, Department of Human Services, and University of Hawai‘i). The data from the DXP consists of all students in Hawai‘i’s public school system spanning elementary through secondary education. The data include education performance measures as well as demographic characteristics of the student. Test score data come from the Smarter Balanced Assessment (SBA). The SBA is an annual assessment of college and career readiness that includes modules on math and English literacy. It has been administered to students in grades three through eight and grade ten since 2015.<sup>18</sup> We standardize test scores to a mean of zero and a standard deviation of one at the grade-module-year level. Our data on

---

<sup>13</sup>Source: [US News](#), retrieved May 13, 2023.

<sup>14</sup>Source: [Hawai‘i Department of Business, Economic Development and Tourism \(DBEDT\)](#), retrieved May 13, 2023.

<sup>15</sup>Source: [The 74](#), retrieved May 13, 2023.

<sup>16</sup>Source: [Private School Review](#), retrieved May 13, 2023.

<sup>17</sup>Source: [The National Center for Higher Education Management Systems](#), retrieved May 13, 2023.

<sup>18</sup>Prior to 2015, the DOE administered the Hawai‘i State Reading and Math Assessment (HSA) to measure student performance. The HSA was administered to students in grades three through eight and ten. Though our data includes test scores from the HSA, we do not have data on test dates, and thus we strictly focus on utilizing data from the SBA.

student test scores span the years 2015-2018.

Though the SBA is mandatory for all public school students, test dates are unique at the student level. Each school year, the DOE provides a one to three month testing window within which schools are required to administer the modules. Each school is then individually responsible for determining the exact date that students take their assessments. Importantly, schools determine their exam dates at the beginning of the school year, well before the school can forecast potential weather conditions or vog levels on the exam date. Schools typically have students within the same grade take the same module on the same date, though larger schools often space students within the same grade across multiple testing dates. The two modules (math and reading) are always taken on separate dates. In some circumstances, school faculty are authorized to have some students take the exams earlier or later than their peers. Since students with cognitive disabilities are subject to alternative assessments, we drop them from our sample.

## 2.2 Air Quality Measurements

We employ data on particulates ( $PM_{2.5}$ ) and sulfur dioxide ( $SO_2$ ) obtained from the State of Hawai‘i Department of Health (DOH).<sup>19</sup> Particulates are measured in micrograms per cubic meter ( $\mu g/m^3$ ).  $PM_{2.5}$  measures particulates that are 2.5 micrometers in diameter or smaller.  $SO_2$  is measured in parts per billion (ppb). The DOH reports measures of each pollutant at hourly frequencies. For our analysis, we aggregate the pollutant measures from each DOH monitoring station to 24-hour averages and merge these data with the DXP data using the date that students took their assessments.

## 2.3 Summary Statistics

In Figure 1, we present histograms of the month that math and reading exams were taken at the student-year level. Most students take both of their assessments in April, but some do take

---

<sup>19</sup>We do not use data on  $PM_{10}$  as the state only had three stations monitoring it.

Table 1: Summary Statistics (Hawai‘i Data eXchange Partnership Student Measures)

	Mean	Std. Dev.
<b>Panel A: Student-Year Level Statistics</b>		
Economically Disadvantaged	0.50	0.50
Received English Language Services	0.06	0.24
<b>Panel B: Student Level Statistics</b>		
Female	0.50	0.50
Asian (Non-Filipino)	0.16	0.37
Filipino	0.24	0.43
Native Hawaiian	0.24	0.43
Pacific Islander	0.09	0.29
White	0.19	0.39
Other Ethnicity	0.08	0.27
Unique Individuals	116,397	
Schools	230	
Years	2015 - 2018	

Notes: Data on student summary statistics comes from the Hawai‘i Data eXchange Partnership for the years 2015-2018. Economically disadvantaged students refer to those who qualify for federal programs such as free and reduced lunch. Those who received English language services are students who enrolled in the State of Hawai‘i Department of Education’s English Learner Program for the academic school year.

them in March and May. Exams are not taken in any other months.

We discuss other summary statistics from the DXP student data in Table 1. Half of student-years in the public school system come from economically disadvantaged families and around six percent received English language services. The table also reveals Hawai‘i’s ethnically diverse population. Nearly a quarter of students identify as Native Hawaiian and just under another quarter identify as Filipino. Another 18% of students identify as non-Filipino Asian, 10% as Pacific Islander, 18% as White, and 8% identify as another ethnicity. The data includes 116,374 unique individuals enrolled across 230 schools.

Summary statistics for pollution are presented in Table 2 by monitoring station.<sup>20</sup> Overall,  $PM_{2.5}$  averages are relatively similar across the islands of O‘ahu, Maui, and Kauai. Hawai‘i island sees slightly higher levels of  $PM_{2.5}$  in certain areas due to Kilauea’s volcanic activity. The

<sup>20</sup>Hawai‘i Island also has a monitoring station at Kamehameha Schools Hawai‘i that began collecting data on  $PM_{2.5}$  and  $SO_2$  in 2019. Our sample of test scores ends in 2018 so we do not use this information to construct pollution predictions. However, we do use it to construct the Kriging weights and for our cross-validation exercise.

Figure 1: Month of Math and Reading Exam

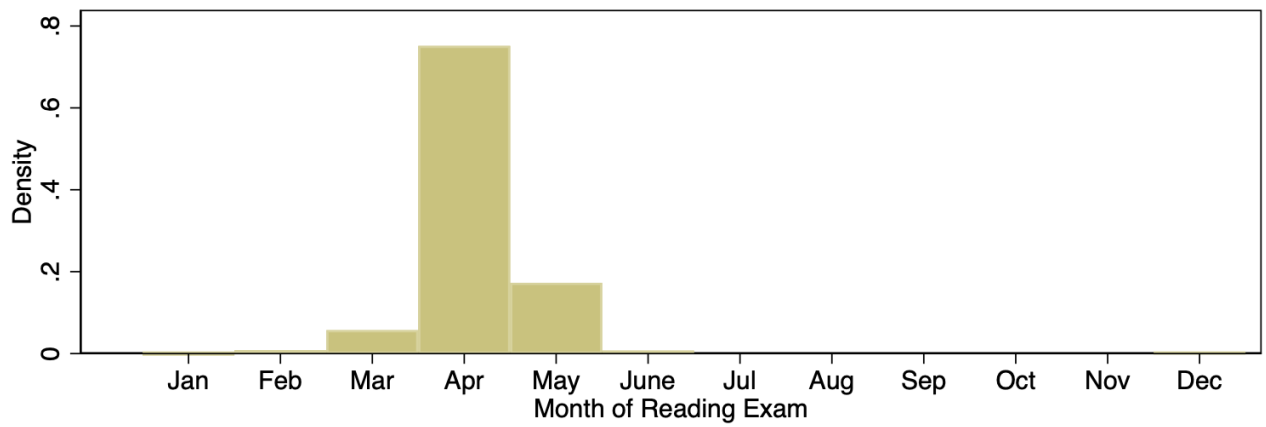
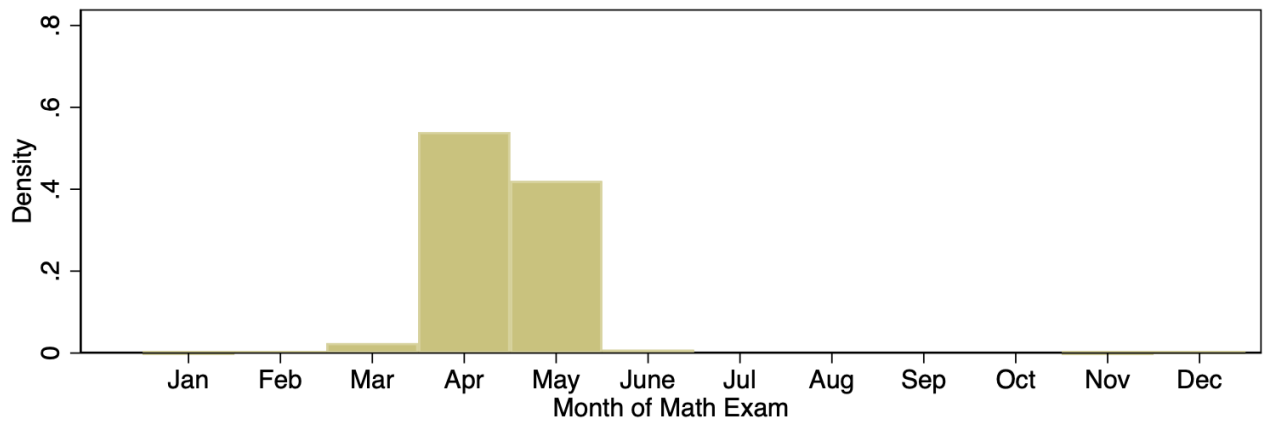


Table 2: Summary Statistics (Pollutant Measures)

Station	$PM_{2.5}$		$SO_2$	
	Mean	Std. Dev.	Mean	Std. Dev.
<i>Hawai‘i Island</i>				
Hilo	8.41	5.62	3.87	8.16
Kona	11.12	4.42	3.25	1.96
Mountain View	3.65	3.18	1.75	2.43
Ocean View	12.60	4.79	18.60	16.06
Pahala	4.77	2.61	28.94	18.89
Hawai‘i Island Average	8.08	2.80	12.09	6.57
<i>O‘ahu</i>				
Honolulu	3.66	2.41	0.38	0.48
Kapolei	4.90	2.01	0.14	0.34
Pearl City	4.05	2.04		
Sand Island	4.92	1.90		
O‘ahu Average	4.35	1.73	0.26	0.34
Sample Average	6.69	1.98	9.04	4.92

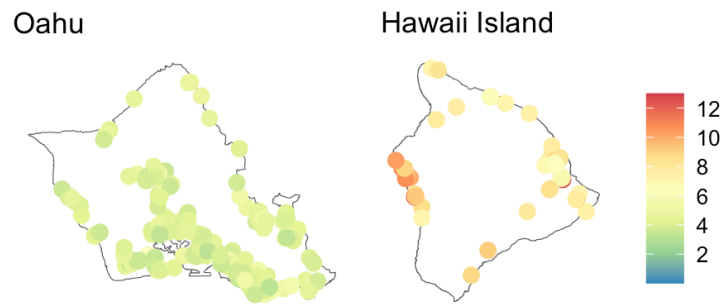
Notes: Data on pollutant measures come from the State of Hawai‘i Department of Health for the years 2015-2018. Measures of  $PM_{2.5}$  and  $SO_2$  are reported for each pollutant monitoring station. The particulate  $PM_{2.5}$  is reported in  $\mu g/m^3$  and  $SO_2$  is reported in  $ppb$ .

Pahala monitoring station is located less than 20 miles south of the Kilauea volcano. Because of the Pahala monitoring station’s close proximity to the volcano’s active vents, average levels of  $SO_2$  in Pahala are nearly three times the state average. For the full sample, the average  $PM_{2.5}$  is  $6.69 \mu g/m^3$  (with a standard deviation of 1.98) and the average  $SO_2$  is  $9.04 ppb$  (with a standard deviation of 4.92).

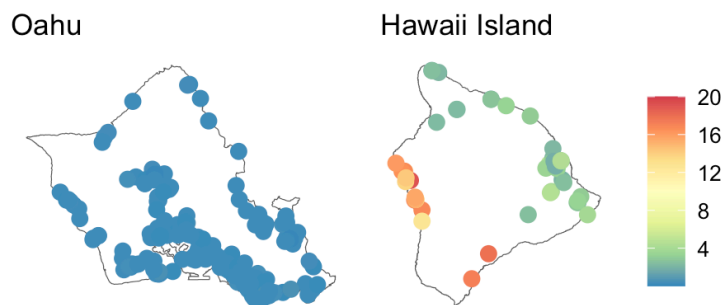
In Figure 2, we display heat maps of average pollution exposure on Oahu and Hawai‘i island across all testing days. In both maps, we employ uniform weights (see Section 3.1). The figure shows considerable variation across islands with the island of Hawai‘i having substantially higher levels of pollution. Within Oahu, there is not a tremendous amount of variation in pollution levels. However, within Hawai‘i island, there is considerable variation with the western coast experiencing substantially elevated levels of pollution. Finally, there is considerably more variation in  $SO_2$  both within and across islands with Oahu having essentially no  $SO_2$  pollution.

Figure 2: Heat Maps of Average  $PM_{2.5}$  and  $SO_2$  Exposure

$PM_{2.5}$  Exposure



$SO_2$  Exposure



Notes: Each map depicts the average exposure on test dates of either  $PM_{2.5}$  or  $SO_2$  for each school. School-level exposure was computed using uniform weights.

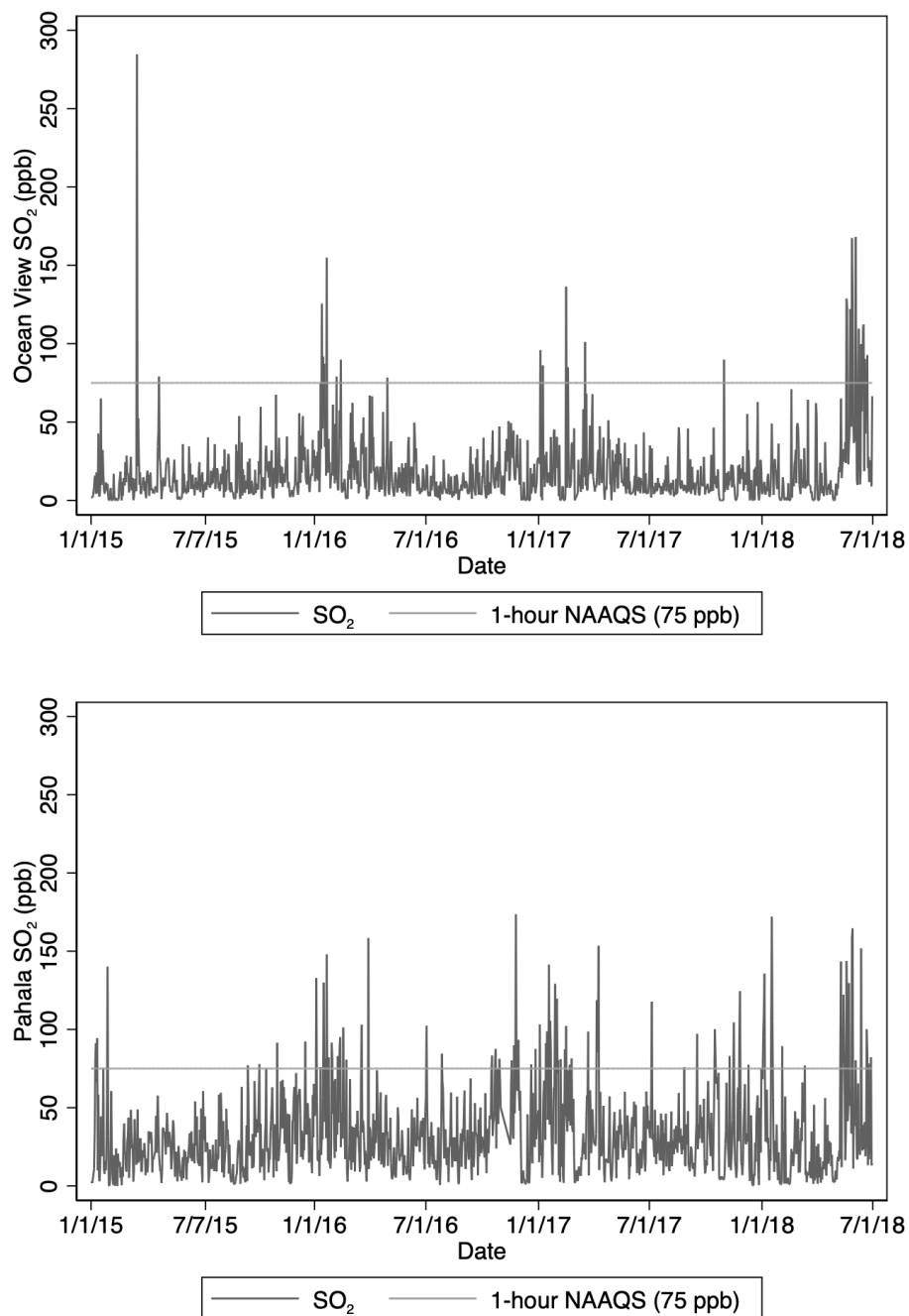
## 2.4 Emissions from Kilauea: Jan 2015 - Jun 2018

In Figure 3, we display time series plots of  $SO_2$  levels near Kilauea from two adjacent monitoring stations: Ocean View and Pahala. We display daily pollution levels over our sample period which spans January 2015 to June 2018. In each figure, we display the one hour National Ambient Air Quality Standards (NAAQS) level for  $SO_2$  from the EPA of 75 ppb. Note that we display one *day* averages and so any levels above the NAAQS line reflect particularly poor air quality.

While the figure does indicate a fairly steady emission of  $SO_2$  over the sample period, it also shows three specific events with increased volcanic activity ([United States Geological Survey, 2018](#); [National Park Service, 2018](#)). First, in April 2015, the lava lake at the summit spilled onto the floor of Halema'uma'u Crater several times causing increased emissions of  $SO_2$ . The lake level abruptly dropped by May 10. Next, the lava lake rose at the start of 2016 and again subsequently declined. Finally, the summit of Pu'u'o'o collapsed on April 30, 2018 causing lava to drain away from the summit from May through August. This lava was subsequently released by a magnitude 6.9 earthquake which caused a fissure near a residential area. Lava then covered 13.7 square miles of land, destroyed more than 600 homes in the Puna district, and added 875 acres of land to the island ([National Park Service, 2018](#)).

Although Hawai'i island tends to see a lot of volcanic activity throughout the year, the 2018 eruption was a highly uncommon event. Prior to this, the last eruption that caused residential damage to that degree took place in 1983 and destroyed 215 homes, much fewer than the 2018 eruption. To this end, we are less concerned about the potential effects of stress related to home displacement due to the volcano on test scores. However, as a robustness check, we later exclude schools located in the area that experienced property damage from our analysis for May through the end of our sample in 2018.

Figure 3: Average Daily  $SO_2$  Levels Near Kilauea



Notes: Each figure displays time series plots of daily averages of  $SO_2$  levels at the Ocean View and Pahala monitoring stations over the period Jan 2015 to Jun 2018. Testing windows vary based on the type of student schedule at each school. Although exact test dates within a window differ annually, during the 2024-2025 school year, testing windows spanned mid-February to late May, mid-November to mid-December, or mid-March to mid-June. See <https://www.hawaiipublicschools.org/TeachingAndLearning/Testing/StateAssessment/Pages/home.aspx> for more information.



## 3 Research Design

### 3.1 Measuring Pollution at Schools

We employ three methods to predict pollution exposure at schools which we subscript with  $s$ . Denoting the measured pollution at monitoring station  $m$  on day  $t$  with  $\Pi_{mt}$ , we can write all three predictors using the generic form

$$\hat{P}_{st} = \sum_{m \in N(s)} \lambda_{sm}(X_t) \Pi_{mt} \quad (1)$$

where  $N(s)$  is a neighborhood of school  $s$  and  $X_t$  is a set of time varying regressors which can possibly be empty.<sup>21</sup> We restrict the weights to sum to unity so that  $\sum_{m \in N(s)} \lambda_{sm}(X_t) = 1$ . Negative weights, however, are permitted. Under stationarity conditions, the means of the pollution measures from each station and school (if they existed) will be the same. Under these conditions, predictors of this form are unbiased.

The first two methods employ inverse distance weights (ID-weights) and uniform weights (U-weights). ID-weights are computed as

$$\lambda_{sm} = \frac{d_{sm}^{-1}}{\sum_{n \in N(s)} d_{sn}^{-1}}$$

where  $d_{sm}$  is the distance between school  $s$  and monitoring station  $m$ . ID-weights are very common in environmental economics. U-weights are computed as

$$\lambda_{sm} = \frac{1}{\#N(s)}.$$

U-weights simply deliver the arithmetic mean using local monitoring stations.

---

<sup>21</sup>In fact, Kriging is a “universal” predictor in that it uses all available measurements to make predictions. However, in non-stationary environments such as this, it is advisable to use more local measures. With this in mind, we employ three neighborhoods: O’ahu, south Hawai‘i (includes Kilauea), and north Hawai‘i (the rest of the island of Hawai‘i).

The third method we employ uses Kriging, which is a common technique in geostatistics (Cressie, 1990). We call these weights K-weights. These weights have nice theoretical properties and can be easily modified to accommodate covariates. With the exception of Lleras-Muney (2010), we know of no other economists who have employed K-weights in the pollution literature. One of the contributions of the paper, therefore, is to investigate in an empirical application how K-weights perform when compared to their better known cousins: ID- and U-weights.

The K-weights that we compute use pollution measurements from monitoring stations in conjunction with variation in Hawai‘i’s trade wind patterns to predict pollution at each of Hawai‘i’s schools. Theoretically, Kriging delivers the best linear unbiased predictor (BLUP) of unobserved pollution at each school (Cressie, 1990).<sup>22</sup> Normally, Kriging weights depend solely on the spatial correlations of pollution measurements across monitoring stations. However, we extend the procedure so that we can incorporate external variables (wind direction in our case) to generate more accurate predictions. This addresses a common critique of ID- and U-weights, namely, that they do not address the spatial correlation in pollution nor can they cannot easily accommodate covariates such as wind direction or relative location.

The K-weights that we employ depend on the distance between school  $s$  and monitoring station  $m$  (previously notated as  $d_{sm}$ ), the relative location of the monitoring station vis-a-vis the school ( $l_{sm}$ ), and the wind direction on that day ( $NE_t$ ). We denote the weight given to monitoring station  $m$  to predict pollution at school  $s$  as  $\lambda(d_{sm}, l_{sm}, NE_t) \equiv \lambda_{sm}(NE_t)$ . In the spirit of Halliday et al. (2019), we employ the variable  $NE_t$ , a binary variable indicating that the winds on that day were northeasterly. As previously discussed, such winds are called “trade winds” and tend to improve air quality throughout the state.

While K-weights must sum to one, they can be negative or greater than unity. This allows the predictions to take on a value outside of the simplex generated by the pollution measurements. In principle, this is a positive feature of Kriging - not a deficiency. To see this, we note that the

---

<sup>22</sup>In simulation studies and under proper conditions, it also has also been shown superior to other commonly used prediction methods such as inverse distance weighting (Zimmerman et al., 1999).

monitoring stations on the island of O’ahu are all in urban Honolulu on the southern shore of the island (see Figure 4). However, many schools on this island are in rural parts of the island and/or on the northern facing shores placing them outside of the simplex generated by the monitoring stations. Because the weights are not constrained to be between zero and one, the predictions at these rural schools can be smaller (or larger than) than *all* of the  $\Pi_{mt}$  used to construct  $\hat{P}_{st}$ . Other common predictors used in this literature such as nearest neighbor or inverse distance weighting do not share this property. On the other hand, schools located far away from monitoring stations may lead to extreme values for weights (despite the fact that they sum to unity) which may not be desirable in certain applications.

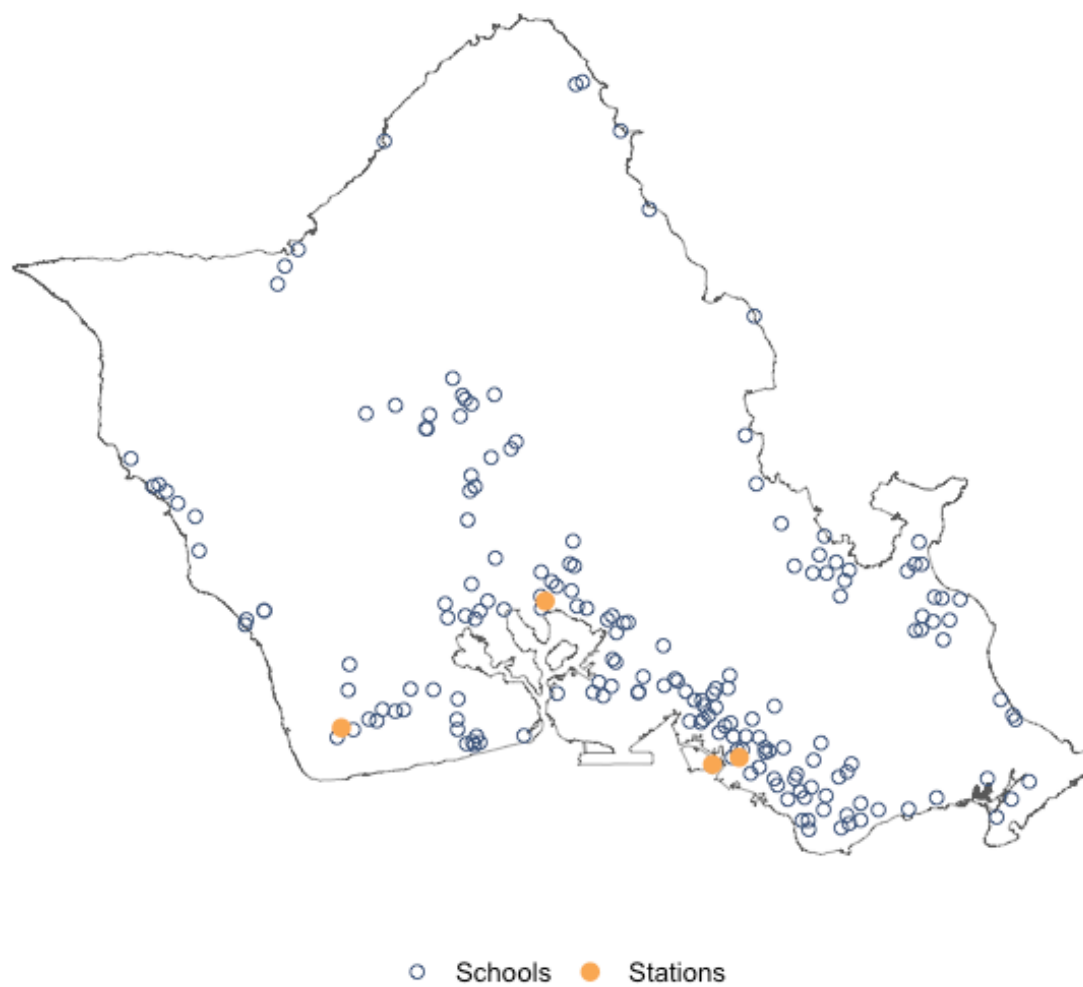
To illustrate the relative locations of the schools and the monitoring stations, we present Figures 4 and 5. These figures plot each school’s location and each monitoring station’s location (denoted by the solid red circles) on the islands of O’ahu and Hawai‘i, respectively. Schools on O’ahu are empty circles. Schools on south Hawai‘i are empty triangle whereas those on north Hawai‘i are crosses. These correspond to the three neighborhoods (denoted  $N(s)$  in equation (1)) for which we compute the ID-, U-, and K-weights. The figure illustrates that the schools in our sample are often located well outside of the simplex generated by the monitoring stations.

We now briefly discuss the GMM procedure that we use to estimate the Kriging weights. A detailed treatment of this can be found in Appendix A.1. Denoting  $M(s) \equiv \#N(s)$ , we define

$$\lambda_s(b) \equiv (\lambda_{s1}(b), \dots, \lambda_{sM(s)}(b), \alpha_b)'$$

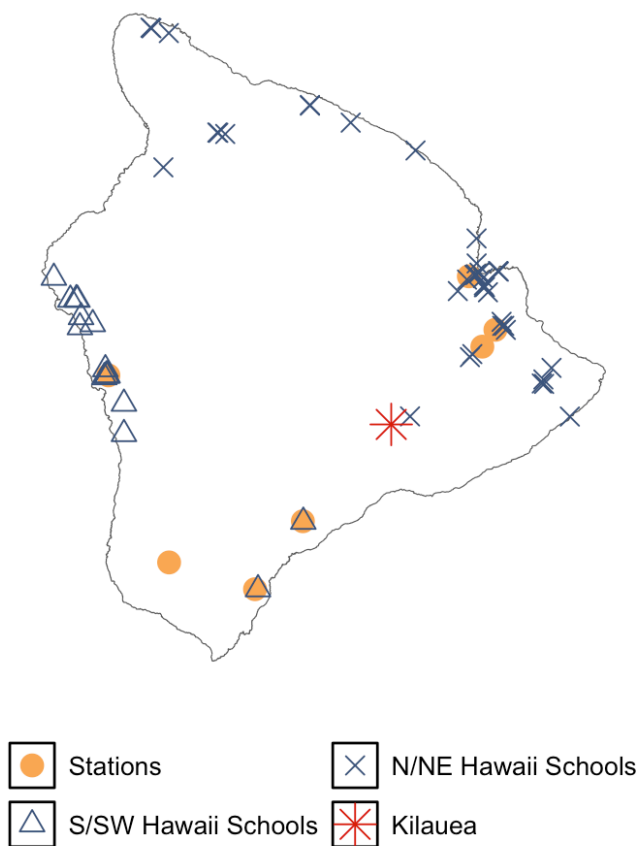
which is a vector that includes the Kriging weights,  $\lambda_{sm}(b)$  for  $b \in \{0, 1\}$  ( $b$  is an indicator for northeasterly winds), and the Lagrangian multiplier on the constraint that the weights must sum to one denoted by  $\alpha_b$  for  $b \in \{0, 1\}$ . The weights depend on the semivariogram between stations  $m$  and  $n$  on trade wind and non-trade wind days. The semivariogram is one minus the spatial correlation between the two locations and equals zero at a given location *i.e.* when  $m = n$ . We denote this measure by  $\gamma_{mn}(b)$ . The Kriging weights can then be derived as  $\lambda_s(b) = \Gamma(b)^{-1}\Gamma_s(b)$

Figure 4: Schools and Monitoring Stations on O'ahu



Notes: This figure displays all pollutant monitoring stations and schools located on the island of O'ahu.

Figure 5: Schools and Monitoring Stations on Hawai‘i Island



Notes: This figure displays all pollutant monitoring stations and schools located on the island of Hawai‘i. We also display the location of Kilauea volcano.

where

$$\Gamma(b) \equiv \begin{bmatrix} \gamma_{11}(b) & \dots & \gamma_{1M(s)}(b) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{M(s)1}(b) & \dots & \gamma_{M(s)M(s)}(b) & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}$$

and

$$\Gamma_s(b) \equiv \begin{bmatrix} \gamma_{1s}(b) \\ \vdots \\ \gamma_{M(s)s}(b) \\ 1 \end{bmatrix}.$$

Hence, the task of computing the Kriging weights hinges on computing the semivariogram for  $NE_t = 1$  and  $NE_t = 0$ . We refer the reader to the appendix for details on the derivation and, specifically, what optimization problem delivers these weights.

To compute the semi-variograms in  $\Gamma(b)$  and, especially,  $\Gamma_s(b)$  (which requires out-of-sample prediction), we postulate a parametric model indexed by a vector. To estimate this parameter, first, we compute the empirical semivariograms for trade wind ( $NE_t = 1$ ) and non-trade wind days ( $NE_t = 0$ ). Next, we model the spatial correlation between stations  $m$  and  $n$  as

$$1 - \gamma_{mn}(NE_t) = \exp \left( -d_{mn} \times \left( \phi NE_t + \sum_{j \in \mathcal{L}} (\delta_j 1_{mn}(j) + \beta_j \times 1_{mn}(j) \times NE_t) \right) \right)$$

where  $\mathcal{L} \equiv \{NE, SE, SW, NW\}$  (which collects the relative location variables) and  $1_{mn}(j)$  is an indicator for the location of  $n$  relative to  $m$  where  $m$  is held fixed. This functional form allows the spatial correlation: (1) to decline with the distance between locations; (2) to decline when trade winds are blowing; (3) to depend on relative locations according to wind direction. The semivariogram is zero when  $d_{mn} = 0$ . We then note that estimation of the parametric model can proceed using a simple Poisson regression or Generalized Linear Model (with a log link function) package in Stata or R and is relatively easy to implement.

## 3.2 Covariogram Estimates

We plot the covariograms for  $PM_{2.5}$  and  $SO_2$  in Figures A1 and A2. Each figure has four plots corresponding to the relative locations of the monitoring stations: NE, SE, SW, and NW. For any pair of stations,  $(m, n)$ , the location corresponds to  $m$  relative to  $n$ . We also plot the covariograms for trade and non-wind days in each plot. Each dot in these plots is a pair of stations.<sup>23</sup>

In Figure A1, we see a large degree of spatial correlation in  $PM_{2.5}$ . In the diagonal plots corresponding to relative locations NE and SW, we see that the spatial correlation remains above 0.2 for up to a distance of 60 miles. In the off-diagonal plots corresponding to NW and SE, the spatial correlations remain well above zero when there are no trade winds, but they are substantially smaller when the trade winds are blowing. Overall, this figure strongly indicates that trade winds result in lower spatial correlations in  $PM_{2.5}$ .

Figure A2 shows a much more muted degree of spatial correlation for  $SO_2$ . When the monitoring stations are either to the northeast or southeast, the figure shows that there is essentially no spatial correlation in  $SO_2$ . When the stations are to the northwest or the southwest, there appears to be moderate spatial correlation through about 100 miles and trade winds modestly dampen it. The more modest degree of spatial correlation in  $SO_2$  makes it more difficult to predict  $SO_2$  exposure at schools. All told, the spatial correlations in  $PM_{2.5}$  are more informative than they are in  $SO_2$ . This most likely underscores our subsequent difficulties predicting  $SO_2$  using any of our predictors.

---

<sup>23</sup>The GMM procedure that we use only employs the bottom triangle of the covariance matrix since the information in the upper triangle is redundant. This implies, however, that the samples used in the off-diagonal plots in the top right and bottom left plots of Figures A1 and A2 are different and so the estimations are also different. For example, for a pair of stations  $(m, n)$ , if  $m$  is northeast of  $n$  then  $1_{mn}(NE) = 1$  and  $1_{mn}(SW) = 0$ . On the other hand, for  $(n, m)$ , we would have  $1_{mn}(NE) = 0$  and  $1_{mn}(SW) = 1$ . The same is true for the plots on the diagonal of the figure (not the covariance matrix). This is highly technical point but some readers may have wondered why these plots are different. This is why.

### 3.3 Pollution Predictions: Summary Statistics

In Table 3, we provide the means and standard deviations of the predictions of pollution at each school on days with and without trade winds using K-, ID-, and U-weights.<sup>24</sup> In Section A.2, we conduct a simple cross-validation exercise. We provide descriptive statistics for the three neighborhoods for which we computed the weights: O’ahu, south/southeast Hawai‘i, and north/northeast Hawai‘i. Using K-weights, the table shows that on trade wind days levels of  $PM_{2.5}$  are lower on O’ahu and north Hawai‘i across all three of our pollutant measure methods. However, we do not see lower particulate levels on trade wind days on south Hawai‘i. Presumably, the reason for this is that many of the schools on south Hawai‘i are located to the west of Kilauea, and the trade winds blow  $PM_{2.5}$  towards those schools. Finally, we see a similar pattern for  $SO_2$  with lower levels on trade wind days on O’ahu and north Hawai‘i.

We note that the standard deviations of the K-weighted predictors of  $PM_{2.5}$  are higher on O’ahu than they are for the ID- and U-weighted predictors. This is particularly the case on trade wind days on O’ahu for which the standard deviation is  $6.12 \mu g/m^3$  using K-weights whereas it is 1.63 and  $1.56 \mu g/m^3$  using ID- and U-weights, respectively. This might reflect that the K-weights are larger in absolute value for some of the more remote schools on O’ahu located far away from the monitoring stations, thereby, necessitating a large degree of extrapolation. On the other hand, the standard deviations of the ID- and U-weighted predictors of  $PM_{2.5}$  on south and north Hawai‘i are more similar.

However, the standard deviations of the K-weighted means of  $SO_2$  are fairly comparable to those using ID- and U-weights on O’ahu but substantially *lower* on South and North Hawai‘i. Their comparability on O’ahu presumably reflects the low spatial correlation in  $SO_2$  on O’ahu. On the other hand, the standard deviations of predictors are substantially lower when employing K-weights throughout the island of Hawai‘i than they are for either the ID- and U-weighted

---

<sup>24</sup>We replaced the prediction with a zero when the Kriging prediction was negative. We also replaced all predictions above the 99th percentile with a missing value. In the raw pollution data, 22.10% of  $PM_{2.5}$  predictions were negative and 1.08% of  $SO_2$  predictions were negative. These numbers might seem large but we note per Halliday et al. (2019) many days in those data essentially had very low pollution levels.



predictors. In general, we find that the K-weights weights for  $SO_2$  are less likely to be negative or greater than one (presumably due to lower spatial correlation) than they are for  $PM_{2.5}$ . This probably underlies the lower standard deviations of the K-weight predictors for  $SO_2$ .

Table 3: Summary Statistics (Pollutant Measures) by Tradewind Status

	O'ahu		S/SW Hawai'i		N/NE Hawai'i	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
<i>No Tradewinds</i>						
$PM_{25}$ (kriging)	4.20	3.11	7.71	6.79	7.07	5.21
$PM_{25}$ (inverse distance)	3.97	2.50	7.72	6.74	7.08	5.30
$PM_{25}$ (uniform weights)	3.97	2.46	7.53	6.23	7.08	5.32
<i>Tradewinds</i>						
$PM_{25}$ (kriging)	2.94	6.12	8.87	6.44	5.65	4.01
$PM_{25}$ (inverse distance)	3.19	1.63	9.11	7.15	5.60	4.00
$PM_{25}$ (uniform weights)	3.23	1.56	8.80	6.84	5.55	3.98
<i>No Tradewinds</i>						
$SO_2$ (kriging)	0.80	0.78	4.96	4.52	3.55	4.03
$SO_2$ (inverse distance)	0.81	0.79	8.33	11.52	5.76	10.52
$SO_2$ (uniform weights)	0.81	0.77	13.06	13.92	5.83	10.38
<i>Tradewinds</i>						
$SO_2$ (kriging)	0.60	0.63	6.79	4.94	1.36	0.98
$SO_2$ (inverse distance)	0.60	0.64	11.43	13.83	1.52	2.16
$SO_2$ (uniform weights)	0.61	0.63	18.03	14.67	1.51	2.11

Notes: Data on pollutant measures come from the State of Hawai'i Department of Health. Measures of  $PM_{2.5}$  and  $SO_2$  are reported for O'ahu, North/Northeast Hawai'i island, and South/Southwest Hawai'i for days with and without tradewinds (northeasterly winds). The particulate  $PM_{2.5}$  is reported in  $\mu g/m^3$  and  $SO_2$  is reported in  $ppb$ . O'ahu stations include Honolulu, Kapolei, Pearl City, and Sand Island stations. South/Southwest Hawai'i stations include Honaunau, Kailua Kona, Kona, Naalehu, Ocean View, and Pahala stations. North/Northeast Hawai'i stations include Hilo, Keaau, and Mountain View stations.

### 3.4 Estimation Equation

To identify the impact of pollution on student cognitive functioning, we employ each of our predictors of pollution exposure at a given school and estimate a linear regression via OLS of

student standardized test scores onto pollution exposure while adjusting for school or individual fixed effects, seasonality, and student demographic characteristics. Specifically, we estimate the model:

$$Y_{igset} = \beta \hat{P}_{st} + \gamma X_{ige} + \sigma_s + \mu_m + \theta_y + v_{igset} \quad (2)$$

where  $Y_{igset}$  is the standardized test score (i.e. the raw score minus its means divided by its standard deviation) of student  $i$  enrolled in grade  $g$  at school  $s$  taking exam  $e$  (math or English) on day  $t$ . Our main variable of interest,  $\hat{P}_{st}$ , is a prediction of exposure to  $PM_{2.5}$  or  $SO_2$  at school  $s$  on day  $t$  discussed above. We scale all estimates of  $\beta$  up by 100 in order to make the estimates more readable by providing more significant figures. This and the fact that the dependent variable is a z-score implies that the interpretation of  $\beta$  is that a one unit increase in the pollutant increases test scores by  $\beta$  % of a standard deviation. The vector  $X_{ige}$  contains time-varying student characteristics such as indicators for economic disadvantage status and reciprocity of English language services, time-invariant student characteristics such as indicators for gender and ethnicity, and an indicator for the student's grade and the type of exam.<sup>25</sup> We also control for air temperature as [Park et al. \(2020\)](#) has shown that temperature strongly correlates with learning outcomes. We include school fixed effects, denoted by  $\sigma_s$ , in order to control for variation at the school level. The terms,  $\mu_m$  and  $\theta_y$ , are month and academic year fixed effects respectively and  $v_{igset}$  is the error term. Standard errors are clustered by school. Identification of  $\beta$  in equation (2) comes from plausibly exogenous variation in  $\hat{P}_{st}$  *within* schools, across time.<sup>26</sup>

While we do not have any information on where particular students live, the majority of

---

<sup>25</sup>Controlling for the economic status of each students' family is particularly important in this specification because the DOE grants geographic exceptions (GE) to students who wish to enroll in a school outside of their district of residence under several qualifying circumstances (e.g. parents are faculty at the receiving school, a program is offered at the receiving school but not at the student's district school, etc). Because of the GE, students who reside in areas with relatively lower average household incomes may attend schools in districts that have higher average household incomes and thus better education programs.

<sup>26</sup>As an additional robustness check, we replace all time-invariant student controls with student fixed effects, which account for all unobserved time-invariant differences across students (e.g. innate ability). Our preferred model avoids student fixed effects since the time frame of our study is limited to several years (2015 to 2018) which, therefore, creates some considerable power issues.

students live in the area in which their school is located. According to the State of Hawai‘i Department of Education, in the 2022-2023 school year, 12% of all public school students received a Geographic Exception and were enrolled in a school other than their home school. To this end, the vast majority of students live near their schools - especially at the elementary level. Accordingly, it is not unreasonable to assume that exposure at a school location is a close proxy for student exposure.

We conclude with a few remarks about the calculation of standard errors in the presence of a generated regressor ( $\hat{P}_{st}$  in our case). With a generated regressor, standard errors should account for sampling uncertainty in  $\hat{P}_{st}$ . In a standard situation in which the regressor is generated from the same sample that is used in the second stage estimation, bootstrapping the generated regressor and then bootstrapping the second stage coefficient estimates provides a common solution.

However, two points make this solution less viable in our scenario. First, the Kriging procedure that is used to generate the regressor takes 20-30 minutes (depending on the pollutant) on a fast machine. This implies that a single standard error with 100 replications could take over a day to compute using standard bootstrapping procedures.<sup>27</sup> Second, as is common in the literature on the impacts of pollution, the generated regressor comes from a separate sample with a separate sampling scheme than the primary estimation sample. Accordingly, the asymptotic distribution computed in Appendix 6A of [Wooldridge \(2010\)](#) for linear models with generated regressors does not apply as these calculations presume a single sample.<sup>28</sup>

---

<sup>27</sup>We do note, however, that there are faster alternatives for extremum estimations that could be considered ([Andrews, 2002](#)).

<sup>28</sup>We do not account for the possibility of spatial correlation in our bootstrapped standard errors, though we cluster them at the school level. It has nonetheless been noticed that “[f]or longitudinal data the [*spatially*] corrected standard errors tend to be markedly smaller than the clustered ones that are routinely calculated. The reason is that fixed effects have already soaked up a good deal of the spatio-temporal structure of residuals, with the result that clustering corrections tend to be too aggressive at clearing out whatever correlation remains.” (([Kelly, 2020](#))).

## 4 Results

### 4.1 Baseline Results

Before we discuss our baseline estimates, in Section A.3, we conduct a simple balance test to ensure that our pollution measures are plausibly exogenous. The bulk of our exogenous controls are indeed unrelated to both  $PM_{2.5}$  and  $SO_2$ . We do see that air temperature is negatively associated with  $PM_{2.5}$  and positively associated with  $SO_2$ . Perhaps, this underscores the need to adjust for air temperature in our estimations.

We report our first set of OLS estimations in Table 4. The table contains three panels corresponding to K-weighted predictions (Panel A), ID-weight predictions (Panel B), and U-weighted predictions (Panel C). Column (1) presents our results for the effect of standardized  $PM_{2.5}$  levels on student test scores, as estimated in equation (2). Looking at the Kriging-based estimate in Panel A, we estimate a drop in student standardized test scores on days with higher  $PM_{2.5}$  levels that is significant at the 5% level. The effect is small; a one unit increase in  $PM_{2.5}$  leads to a 0.16 percent of a standard deviation drop in student test scores (recall that our pollutant effects are scaled up by a factor of 100). With a full sample standard deviation of 1.98 (see Table 2), a one standard deviation increase in  $PM_{2.5}$  corresponds to a 0.32 percent drop in student test scores, on average. Our estimates using inverse distance and uniform weighting schemes in Panels B and C are larger: -0.271 and -0.278, respectively. Both estimates are significant at the 1% level. Scaling each of the estimates up by the standard deviation of  $PM_{2.5}$ , we estimate declines in test scores of 0.54 and 0.55 of percent a standard deviation, respectively. In the second column, we replace student controls with student fixed effects. We see that the estimate is no longer significant using the K-weights. However, the results using ID- and U-weights are robust to the inclusion of student fixed effects. In fact, both estimates only change slightly once student fixed effects are included. Both Panels B and C indicate that a one standard deviation increase in  $PM_{2.5}$  corresponds to a 0.45 percent of a standard deviation decline in test scores. In the final

two columns, we do not see any impact of  $SO_2$  on test scores.

The fact that the impact of  $PM_{2.5}$  is robust to the inclusion of student fixed effects should assuage a large number of identification concerns. This strongly indicates that the effects of poor air quality on student performance are not driven by any differential selection out of test-taking on high pollution days.

While many of these effects are small or null, there are some important considerations. First, these weaker effects might mask important underlying heterogeneity by geography, the average level of pollution, and SES within schools. We will explore each of these in the next section. Second, there is a tendency for estimates to be larger for schools that are located near monitoring stations suggesting the results in this table may suffer from some attenuation bias from measurement errors.

To better understand this second point, we estimated the same models as in Table 4 restricting the sample to schools that are within two, five, and ten miles of a monitoring station. The idea of this exercise is that there will be less measurement error in exposure measures for schools that are closer to monitoring stations. We report the results in Table A3.<sup>29</sup>

We see that restricting the sample to schools that are nearest to monitoring stations increases the magnitudes of the estimates substantially. This is particularly true when we employ Kriging. For example, in Table 4, the point estimate on  $PM_{2.5}$  using Kriging with school fixed effects is -0.160 and significant at the 5% level. On the other hand, in Table A3, the corresponding point estimate for schools within two miles of a monitoring stations is -0.428 which is significant at the 1% level. The estimates for the effects for  $PM_{2.5}$  using K-weights decline monotonically as we include schools within five miles (-0.155) and ten miles (-0.126) of the nearest station. One interpretation of this is that Kriging works well relatively close to the monitoring stations but performs worse as the amount of required extrapolation increases. Finally, it is also worth

---

<sup>29</sup>In Table A2, we report the averages of demographic variables with two, five, and ten miles of a monitoring station. It is true that there are some demographic differences as we move away from the monitoring stations. However, we are already adjusting for these covariates and so we do not believe that this poses an identification issue.

noting that the standard errors for all estimates using K-weights are lower than the corresponding estimates using inverse distance or uniform weights when using  $PM_{2.5}$ .

We see similar patterns for inverse distance and uniform weights in Table A3. Looking at the estimates of the effects of  $PM_{2.5}$  with school fixed effects, the point estimates using inverse distance and uniform weights are -0.577 and -0.514 when restricting the sample to schools within two miles of the nearest monitoring station. These estimates are about twice as large as the corresponding estimates from Table 4: -0.271 and -0.278. However, this distance restriction makes a bigger difference when employing Kriging than the other two predictors. In addition, we also see that the magnitudes of the point estimates decline as we add schools that are further away from the monitoring stations. However, the degree of this decline is not as large when using inverse distance and uniform weighting than when using Kriging. While speculative, one interpretation of this is that extrapolation poses a greater problem for Kriging than the other predictors.

## 4.2 Non-linearities in the Effects of Particulates on Test Scores

We now investigate non-linearities in the relationship between particulate exposure and student test scores. To do this, we estimate a variant of our main regression equation in which we estimate separate coefficients on  $PM_{2.5}$  by quintiles of daily exposure. Specifically, we define the subset of the support of *daily*  $PM_{2.5}$  containing the  $q^{th}$  quintile as  $Q_q(PM_{2.5})$ . We estimate the following model (with some abuse of notation):

$$Y_{igset} = \sum_q 1\left(\hat{P}_{st} \in Q_q(PM_{2.5})\right) \times \hat{P}_{st} \times \beta_q + \gamma X_{ige} + \sigma_s + \mu_m + \theta_y + v_{igset}. \quad (3)$$

The parameters,  $\beta_q$ , identify the effects of daily levels of  $PM_{2.5}$  at each quintile of its daily distribution.

We report the results in Table 5 both with school and individual fixed effects for each of our three pollution measures for a total of six separate regression estimates. While not com-

Table 4: Effect of Pollution on Math and Reading Scores, OLS Estimates

	$PM_{2.5}$		$SO_2$	
	(1)	(2)	(3)	(4)
<i>Panel A: Kriging</i>				
Pollutant	-0.160** (0.073)	-0.081 (0.073)	-0.185 (0.187)	-0.070 (0.190)
$R^2$	0.267	0.851	0.266	0.850
No. Students	92,168	92,168	115,722	115,722
<i>Panel B: Inverse Distance</i>				
Pollutant	-0.271*** (0.093)	-0.227** (0.103)	-0.074 (0.073)	-0.025 (0.071)
$R^2$	0.268	0.850	0.267	0.850
No. Students	97,317	97,317	116,241	116,241
<i>Panel C: Uniform Weights</i>				
Pollutant	-0.278*** (0.091)	-0.227** (0.101)	-0.070 (0.068)	-0.063 (0.065)
$R^2$	0.268	0.850	0.267	0.850
No. Students	97,317	97,317	116,241	116,241
Number of Schools	230	230	230	230
School FE	X	X	X	X
Month FE	X	X	X	X
Year FE	X	X	X	X
Individual FE		X		X

Notes: Standard errors are clustered by school. Control variables include gender, economically disadvantaged students, English language service recipients, exam subject, air temperature, grade level and ethnicity. All estimations control for school, month and academic year fixed effects. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

pletely stark, there is a tendency for the most precise (and hence most significant) estimates to be in the highest quintile of exposure. We do note that some other point estimates at lower quintiles have larger point estimates but we are careful to point out that these estimates often have extremely large standard errors. One estimate in the second column notwithstanding, the point estimates for the highest quintiles tend to range between 0.19 and 0.40. These are larger (albeit more imprecise) effects than the estimates in Table 4 that do not consider non-linearities. Finally, towards the bottom of the table, we report  $F$ -statistics of the null that coefficients on the interactions for the bottom four quintiles are zero along with their  $p$ -values. We fail to reject this null in half of the estimations in columns two through four. All told, the results in this table provide weak evidence that effects of particulates on student test scores are most pronounced at the highest levels of exposure.



Table 5: Daily Average  $PM_{2.5}$  Exposure Effects

	Kriging		Inverse Dist.		Uniform	
	(1)	(2)	(3)	(4)	(5)	(6)
$PM_{2.5}$ Quintile 1 $\times PM_{2.5}$	-0.114 (0.652)	0.539 (0.612)	-1.453* (0.750)	0.072 (0.653)	0.072 (0.899)	-0.684 (0.814)
$PM_{2.5}$ Quintile 2 $\times PM_{2.5}$	-0.619 (0.387)	-0.202 (0.353)	-0.732 (0.534)	0.163 (0.434)	0.046 (0.566)	-0.626 (0.521)
$PM_{2.5}$ Quintile 3 $\times PM_{2.5}$	-0.397 (0.261)	0.038 (0.280)	-0.725* (0.394)	-0.141 (0.353)	0.244 (0.381)	-0.288 (0.361)
$PM_{2.5}$ Quintile 4 $\times PM_{2.5}$	0.179 (0.184)	0.212 (0.172)	-0.690** (0.310)	-0.282 (0.325)	-0.385 (0.243)	-0.587*** (0.225)
$PM_{2.5}$ Quintile 5 $\times PM_{2.5}$	-0.204*** (0.078)	-0.095 (0.071)	-0.396*** (0.122)	-0.189* (0.113)	-0.248* (0.139)	-0.263** (0.131)
$R^2$	0.267	0.851	0.268	0.850	0.268	0.850
F-test	2.834	1.153	1.419	0.677	2.567	4.749
P-value	0.025	0.332	0.228	0.608	0.039	0.001
Observations	301535	296008	317780	317156	317780	317156
Month FE	X	X	X	X	X	X
Academic Year FE	X	X	X	X	X	X
School FE	X	X	X	X	X	X
Individual FE		X		X		X

Notes: Standard errors are clustered by school. Control variables include gender, economically disadvantaged students, English language service recipients, exam subject, air temperature, grade level and ethnicity. All estimations control for school, month and academic year fixed effects. The F-statistic tests the null that coefficients on the bottom four  $PM_{2.5}$  quintile interactions are equal to zero. The p-value for this test is displayed below. Daily average  $PM_{2.5}$  quintiles were calculated by taking the daily average  $PM_{2.5}$  level over the entire sample period. Observations are at the student-test-year level. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

### 4.3 Impacts by Long-Term Pollution Levels

Are the effects of  $PM_{2.5}$  on student cognitive performance in fact larger when mean exposure is higher? A key advantage of the Hawai'i context (as indicated by Figure 2 and Table 3) comes from its rich variation in pollution levels across schools stemming from their location on each island and their proximity to the Kilauea volcano. This allows us to test for possible nonlinear effects in how pollutants affect student performance, particularly across schools with differing average levels of exposure.

To investigate this, we estimate an equation akin to equation (3) in which we replace the indicators of quintiles of daily exposure with quintiles for long-term exposure. We measure long-term exposure with the average of  $PM_{2.5}$  over the sample period at a particular school. We denote this by  $\overline{PM_{2.5,s}}$ . Similar to above, we define the subset of the support of long-term  $PM_{2.5}$  containing the  $q^{th}$  quintile as  $Q_q(\overline{PM_{2.5}})$ . Then we estimate the following model (again with some abuse of notation):

$$Y_{igset} = \sum_q 1(\overline{PM_{2.5,s}} \in Q_q(\overline{PM_{2.5}})) \times \hat{P}_{st} \times \beta_q + \gamma X_{ige} + \sigma_s + \mu_m + \theta_y + v_{igset}. \quad (4)$$

Now, the parameters,  $\beta_q$ , identify the effects of daily levels of  $PM_{2.5}$  at each quintile of the distribution of long-term pollution.

We present the results in Table 6. In four out of the six specifications that we estimated, the only significant estimate was the interaction with the quintile for the schools with the worst air quality during the sample period. These effects tend to be larger than corresponding estimates in Table 4 that do not account for such heterogeneity. For example, the estimates of  $\beta_5$  with school fixed effects in Table 6 are -0.364 (K-weights), -0.270 (ID-weights), and -0.310 (U-weights), whereas the corresponding estimates from Table 4 are -0.160 (K-weights), -0.271 (ID-weights), and -0.278 (U-weights). We note that the estimates for the most polluted schools are significant even after the inclusion of individual fixed effects for two out of three specifications. Finally,

we conducted an  $F$ -test of the null that the interactions with the bottom four quintiles were zero. We fail to reject this null for all six estimations. This provides evidence that the effects of  $PM_{2.5}$  on test scores are strongly conditioned by long-term pollution exposure and most pronounced at the highest quintile of exposure.

Table 6: School Average  $PM_{2.5}$  Exposure Effects

	Kriging		Inverse Dist.		Uniform	
	(1)	(2)	(3)	(4)	(5)	(6)
$PM_{2.5}$ Quintile 1 $\times PM_{2.5}$	0.027 (0.274)	0.086 (0.305)	-0.692 (0.469)	-0.136 (0.623)	-0.413 (0.489)	0.202 (0.661)
$PM_{2.5}$ Quintile 2 $\times PM_{2.5}$	0.195 (0.220)	0.266 (0.291)	0.359 (0.487)	0.129 (0.573)	-0.142 (0.451)	-0.126 (0.571)
$PM_{2.5}$ Quintile 3 $\times PM_{2.5}$	-0.217 (0.184)	-0.045 (0.148)	0.154 (0.378)	-0.246 (0.356)	0.334 (0.444)	-0.153 (0.374)
$PM_{2.5}$ Quintile 4 $\times PM_{2.5}$	-0.110 (0.129)	-0.126 (0.110)	-0.446* (0.243)	-0.622*** (0.238)	-0.325 (0.309)	-0.516* (0.304)
$PM_{2.5}$ Quintile 5 $\times PM_{2.5}$	-0.364*** (0.101)	-0.208* (0.107)	-0.270** (0.110)	-0.149 (0.112)	-0.310*** (0.095)	-0.187* (0.102)
$R^2$	0.267	0.851	0.268	0.850	0.268	0.850
F-test	0.768	0.623	1.685	1.938	0.621	0.796
P-value	0.547	0.646	0.154	0.105	0.648	0.529
Observations	301535	296008	317780	317156	317780	317156
Month FE	X	X	X	X	X	X
Academic Year FE	X	X	X	X	X	X
School FE	X	X	X	X	X	X
Individual FE		X		X		X

Notes: Standard errors are clustered by school. Control variables include gender, economically disadvantaged students, English language service recipients, exam subject, air temperature, grade level and ethnicity. All estimations control for school, month and academic year fixed effects. The F-statistic tests the null that coefficients on the bottom four  $PM_{2.5}$  quintile interactions are equal to zero. The p-value for this test is displayed below. School average  $PM_{2.5}$  quintiles were calculated by taking the average  $PM_{2.5}$  level by school over the entire sample period. Observations are at the student-test-year level. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## 4.4 Heterogeneity by Economic Disadvantage Status

Does poor air quality have larger effects on the most disadvantaged pupils within a school? Case et al. (2002) show that children from poorer backgrounds are at higher risk of developing a host of health problems than better off children. This suggests that more well off children will be in better health which could confer more resiliency when combating the pernicious effects of air pollution. In this sense, air pollution could exacerbate pre-existing inequities within schools.

To investigate this, in Table 7, we estimate a variation of equation (2) while including an interaction term between the pollutant level and an indicator for whether the pupil was eligible for the free and reduced school lunch program - a proxy for student economic disadvantage. Importantly, we employ the SES indicator from baseline defined as the first year that we observe the pupil in the sample. While it is true that the SES indicator is highly persistent across years, using baseline SES status does assuage concerns that changes in SES status across years are endogenous.

We find the the effects of pollution are larger for economically disadvantaged pupils. In Panel A of the table, we display the estimates of the effects of particulates and we observe drastically different effects by economic status. For example, using K-weighted predictors in the first column, the interaction between  $PM_{2.5}$  and the disadvantaged indicator is -0.297 and highly significant, whereas the direct effect is near zero and not significant. This implies that the the harmful effects from  $PM_{2.5}$  for disadvantaged students are magnitudes greater than their effects for their more advantaged counterparts. We see similar results using inverse distance and uniform weights.<sup>30</sup> In Panel B, we see that the interactions between  $SO_2$  and the SES indicator are similar in magnitude to those for  $PM_{2.5}$  when using K and ID weights, although the estimate is only marginally significant in the latter case.

The magnitudes of these effects on cognitive performance of disadvantaged students are not trivial. Once again, using the descriptive statistics from Table 2, we calculate that a one standard

---

<sup>30</sup>We compared the sum of the interaction and the direct effect of particulates to the direct effect by itself.

Table 7: Effect of Pollution on Exam Scores for Economically Disadvantaged Students Using Baseline Economic Disadvantage Status

	Kriging (1)	Inverse Distance (2)	Uniform Weights (3)
Economically Disadvantaged * $PM_{2.5}$	-0.297** (0.114)	-0.373** (0.178)	-0.345* (0.183)
$PM_{2.5}$	0.010 (0.092)	-0.023 (0.134)	-0.051 (0.133)
Economically Disadvantaged	-29.896*** (1.138)	-29.235*** (1.319)	-29.399*** (1.353)
$R^2$	0.268	0.268	0.268
Economically Disadvantaged * $SO_2$	-0.274 (0.247)	-0.270* (0.152)	-0.086 (0.107)
$SO_2$	-0.021 (0.230)	0.108 (0.145)	-0.017 (0.109)
Economically Disadvantaged	-30.819*** (1.067)	-30.843*** (1.022)	-31.073*** (1.052)
$R^2$	0.266	0.268	0.268

Notes: Standard errors are clustered by school. Control variables include gender, English language service recipients, exam subject, economic disadvantage status, air temperature, grade level and ethnicity. All estimations control for school, month and academic year fixed effects. Baseline economic disadvantage status is determined by taking each students socioeconomic status at the first year that they enter the data. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

deviation increase in  $PM_{2.5}$  decreases test scores for disadvantaged pupils by 0.39 percent of a standard deviation using our Kriging weights. A similar calculation indicates declines of 0.78 percent using inverse distance and uniform weights.

Could these larger impacts of pollution for disadvantaged students be driven by selection? For example, there could be a potential correlation between where disadvantaged students enroll and school characteristics including the school's location, its average pollution level, or the school's potential resources to combat the harmful effects of pollutants (e.g. air conditioning). We do not believe that this is the case for the simple reason that the estimates in Table 7 all include school fixed effects.

However, we can also offer an alternative test of this possibility to eliminate any lingering

doubts. In Table A5, we estimate a similar specification as in equation (4) in which we interact  $PM_{2.5}$  with dummies for being in a quintile for the percent of economically disadvantaged students by school. This will address whether or not the effects in Table 7 are driven to some degree by school-level resources. If this were the case, we would expect to see the largest effects at the highest quintiles. However, we do not see this. Instead, we see that the largest impacts of particulates occur towards the middle of the SES distribution. In other words, disadvantaged students do not appear to be especially harmed by pollutants due to their school’s location or school resources. Rather, it appears as if poorer pupils are more adversely impacted by pollution than richer pupils conditional on school resources.

## 4.5 Effects by Gender

In Appendix Table A6, we test for potential heterogeneous responses to pollution by gender by estimating the effects separately for male and female students. Generally, we find that pollution harms the performance of both genders, but has a stronger negative effect on male students for  $PM_{2.5}$ . The magnitude of the difference is sensitive to the model considered, where K-weighted predictors detect little difference across the genders, while uniform weights suggest male students see drops in performance about twice the size in magnitude as female students in response to the same change in  $PM_{2.5}$ . In our context, like in many other K-12 contexts, girls significantly outperform boys; our results suggest that increasing  $PM_{2.5}$  levels may further widen the gender performance gap in primary and secondary schools. The results for  $SO_2$ , on the other hand, suggest no heterogeneous effects by gender.

## 4.6 Robustness Checks

### Additional Weather Controls

In Table A7, we explore the robustness of our core findings in Table 4 to a number of concerns. For the sake of comparison, the first two columns of Table A7 reproduce columns one

and two of Table 4. In the third column of the table, we add indicators for temperature deciles from the test day. All three point estimates in that column remain unaffected. In the fourth column, we include indicators for deciles of average temperature from the previous year. This does attenuate all of the estimates by roughly 33%. The estimate that employ K-weights is now -0.108 but no longer significant. However, the estimates using ID-weights and U-weights are -0.193 and -0.197, respectively, and each retains significance at the 10% level.

In the last column, we include island/date fixed effects. Note that we only include month and year fixed effects in our baseline specification and so this is a much more stringent control. It only relies on variation in pollution within an island and within the same day. We suspect that a lot of this residual variation is measurement error. We see that the three estimates in this column are negative, but have much larger standard errors. None are significant at conventional levels.

Next, in Table A8, we replicate the same specifications from Table A7 but now we add indicator variables for rainfall deciles during the day of the test. The specifications in this table are otherwise identical to those in Table A7. Overall, our findings are robust to adjusting for contemporaneous rainfall in this way and, if anything, the point estimates tend to increase in magnitude.

### Using Satellite Pollution Data

We now explore the robustness of our core findings to using pollution data culled from satellites. The data come from the Modern Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2).<sup>31</sup> It is important to highlight that the  $SO_2$  data is reported in terms of mass concentrations, with units expressed as  $kg/m^3$ . Since these measurements are not directly convertible to parts per billion (ppb), the corresponding coefficients differ from our

---

<sup>31</sup>MERRA-2 is a global atmospheric reanalysis generated by NASA's Global Modeling and Assimilation Office. The MERRA-2 data consists of global, hourly surface  $PM_{2.5}$  and  $SO_2$  mass concentrations, derived using a machine learning approach with a convolutional neural network. Observations are reported in  $\mu g/m^3$  for  $PM_{2.5}$  and  $kg/m^3$  for  $SO_2$ . For this study, we aggregate the particulate data to the daily frequency by averaging hourly pollution measurements and estimate  $PM_{2.5}$  and  $SO_2$  levels at each school by extracting data from the MERRA-2 dataset based on the school's latitude and longitude coordinates.

estimates utilizing data from monitoring stations.

We present four estimates in Table A9 for  $PM_{2.5}$  and  $SO_2$ , incorporating both school and individual fixed effects. The primary takeaway from this table is that none of the estimated coefficients are statistically significant. While the estimates for  $PM_{2.5}$  are negative, their magnitudes are substantially attenuated compared to those reported in Table 4. Furthermore, the estimates for the effects of  $SO_2$  derived from satellite data are not only insignificant but also exhibit the opposite sign. We suspect that this discrepancy arises due to a considerable degree of measurement error in the satellite-based pollution data relative to the data obtained from ground-based monitoring stations.

To substantiate this claim, we present the correlations between satellite and ground-based station data at the school complex level in Table A10.<sup>32</sup> For  $PM_{2.5}$ , the correlations on Oahu range from 0.322 at Roosevelt to 0.767 at Kalaheo. On the island of Hawai‘i, they vary from -0.284 at Hilo to 0.346 at Konawaena. The discrepancies are even more pronounced for  $SO_2$ . On Oahu, the correlations range from 0.196 at Kalani to 0.486 at Radford, while on the island of Hawai‘i, they range from -0.007 at Konawaena to 0.320 at Hilo.

These correlations are far from unity, which would be expected if both measures served as accurate proxies for air quality. In fact, some correlations are even negative. However, this does not necessarily indicate whether satellite or station data is more accurate; rather, it suggests a substantial degree of measurement error in at least one of the pollution metrics. That said, we posit that station data is likely to exhibit less measurement error than satellite data due to its closer proximity to the actual pollution sources.

### **Test Date Delays and Home Displacement**

Finally, in Sections A.4 and A.5, we explore the robustness of our findings to two factors. First, we investigate the role that delayed test dates play in potentially biasing our estimates. Second, we examine whether home displacement can explain any of our effects. Overall, these concerns do not impact our results.

---

<sup>32</sup>A school complex consists of a high school along with its feeder elementary and middle schools.



## 5 Conclusion

Using variation in air quality in the Hawaiian islands due to volcanic activity, we estimate the impacts of  $PM_{2.5}$  and  $SO_2$  on student performance. Because of the state's normally pristine air quality conditions, variation in pollutants is primarily determined by wind direction and volcanic emissions from Kilauea volcano on the island of Hawai'i. Exploiting this variation in pollution across the state, we find that worsening air quality decreases student exam scores. Specifically, we find that a standard deviation increase in  $PM_{2.5}$  decreases test scores by about 0.32 to 0.55 percent of a standard deviation on average.

We also find that schools with higher average pollution levels tend to see worse effects of air quality on student performance. For the schools with the worst air quality, a one standard deviation increase in  $PM_{2.5}$  lowers test scores by between 0.53 and 0.72 percent of a standard deviation. Most of the estimates are closer to the upper range of this interval. Formal tests indicate that the effects in the bottom 80% of the long-term pollution distribution are null.

Why would the effects of particulate pollution be greater in schools with higher average levels of pollution? Our view is that students attending these schools experience more chronic exposure to pollution. This prolonged exposure may weaken their ability to cope with additional environmental insults compared to students with less chronic exposure.

Finally, the negative effects of pollution on student performance are much larger for poorer students. The effects of  $PM_{2.5}$  on test scores for economically disadvantaged pupils are at least three times larger than they are for their more advantaged peers. These effects are not driven by school level characteristics but are instead a result of student level differences within schools. This is in line with previous literature which shows that poorer children are subject to worse health outcomes (Case et al., 2002) which may imply greater susceptibility to environmental insults. In summary, we show that poor students face additional obstacles in accumulating human capital when air quality is poor relative to those who are more financially stable.

## References

- Aguilar-Gomez, S., H. Dwyer, J. S. G. Zivin, and M. J. Neidell (2022): “This is Air: The” Non-Health” Effects of Air Pollution,” Tech. rep., National Bureau of Economic Research.
- Amanzadeh, N., M. Vesal, and S. F. F. Ardestani (2020): “The impact of short-term exposure to ambient air pollution on test scores in Iran,” *Population and Environment*, 41, 253–285.
- Andrews, D. W. (2002): “Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators,” *Econometrica*, 70, 119–162.
- Archsmith, J., A. Heyes, and S. Saberian (2018): “Air quality and error quantity: Pollution and performance in a high-skilled, quality-focused occupation,” *Journal of the Association of Environmental and Resource Economists*, 5, 827–863.
- Austin, W., G. Heutel, and D. Kreisman (2019): “School bus emissions, student health and academic performance,” *Economics of Education Review*, 70, 109–126.
- Bedi, A. S., M. Y. Nakaguma, B. J. Restrepo, and M. Rieger (2021): “Particle pollution and cognition: Evidence from sensitive cognitive tests in Brazil,” *Journal of the Association of Environmental and Resource Economists*, 8, 443–474.
- Burnett, R. T., M. Smith-Doiron, D. Stieb, S. Cakmak, and J. R. Brook (1999): “Effects of particulate and gaseous air pollution on cardiorespiratory hospitalizations,” *Archives of Environmental Health: An International Journal*, 54, 130–139.
- Carneiro, J., M. A. Cole, and E. Strobl (2021): “The effects of air pollution on students’ cognitive performance: Evidence from brazilian university entrance tests,” *Journal of the Association of Environmental and Resource Economists*, 8, 1051–1077.
- Carré, J., N. Gatimel, J. Moreau, J. Parinaud, and R. Léandri (2017): “Does air pollution play a role in infertility?: a systematic review,” *Environmental Health*, 16, 1–16.
- Case, A., D. Lubotsky, and C. Paxson (2002): “Economic status and health in childhood: The origins of the gradient,” *American Economic Review*, 92, 1308–1334.
- Chang, T., J. Graff Zivin, T. Gross, and M. Neidell (2016): “Particulate pollution and the productivity of pear packers,” *American Economic Journal: Economic Policy*, 8, 141–69.
- Chang, T. Y., J. Graff Zivin, T. Gross, and M. Neidell (2019): “The effect of pollution on worker productivity: evidence from call center workers in China,” *American Economic Journal: Applied Economics*, 11, 151–72.
- Chen, S., C. Guo, and X. Huang (2018): “Air pollution, student health, and school absences: Evidence from China,” *Journal of Environmental Economics and Management*, 92, 465–497.
- Choumert-Nkolo, J., A. Lamour, and P. Phélinas (2021): “The economics of volcanoes,” *Economics of Disasters and Climate Change*, 1–23.

- Cressie, N. (1990): “The origins of kriging,” *Mathematical geology*, 22, 239–252.
- Currie, J., E. A. Hanushek, E. M. Kahn, M. Neidell, and S. G. Rivkin (2009): “Does pollution increase school absences?” *The Review of Economics and Statistics*, 91, 682–694.
- Currie, J., J. Voorheis, and R. Walker (2020): “What caused racial disparities in particulate exposure to fall? New evidence from the Clean Air Act and satellite-based measures of air quality,” Tech. rep., National Bureau of Economic Research.
- Di, Q., Y. Wang, A. Zanobetti, Y. Wang, P. Koutrakis, C. Choirat, F. Dominici, and J. D. Schwartz (2017): “Air pollution and mortality in the Medicare population,” *New England Journal of Medicine*, 376, 2513–2522.
- Duque, V. and M. Gilraine (2022): “Coal use, air pollution, and student performance,” *Journal of Public Economics*, 213, 104712.
- Ebenstein, A., V. Lavy, and S. Roth (2016): “The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution,” *American Economic Journal: Applied Economics*, 8, 36–65.
- Gilraine, M. and A. Zheng (2022): “Air pollution and student performance in the US,” Tech. rep., National Bureau of Economic Research.
- Graff Zivin, J. and M. Neidell (2012): “The impact of pollution on worker productivity,” *American Economic Review*, 102, 3652–73.
- Halliday, T. J., J. Lynham, and Á. de Paula (2019): “Vog: using volcanic eruptions to estimate the health costs of particulates,” *The Economic Journal*, 129, 1782–1816.
- He, J., H. Liu, and A. Salvo (2019): “Severe air pollution and labor productivity: Evidence from industrial towns in China,” *American Economic Journal: Applied Economics*, 11, 173–201.
- Heissel, J. A., C. Persico, and D. Simon (2022): “Does Pollution Drive Achievement? The Effect of Traffic Pollution on Academic Performance,” *Journal of Human Resources*, 1218–9903R2.
- Jacqz, I. (2022): “Toxic test scores: The impact of chemical releases on standardized test performance within US schools,” *Journal of Environmental Economics and Management*, 115, 102628.
- Kelly, M. (2020): “Understanding Persistence,” *CEPR Discussion Paper No. DP15246*.
- Koken, P. J., W. T. Piver, F. Ye, A. Elixhauser, L. M. Olsen, and C. J. Portier (2003): “Temperature, air pollution, and hospitalization for cardiovascular diseases among elderly people in Denver,” *Environmental health perspectives*, 111, 1312–1317.
- Komisarow, S. and E. L. Pakhtigian (2022): “Are power plant closures a breath of fresh air? Local air quality and school absences,” *Journal of Environmental Economics and Management*, 112, 102569.

- Lichter, A., N. Pestel, and E. Sommer (2017): “Productivity effects of air pollution: Evidence from professional soccer,” *Labour Economics*, 48, 54–66.
- Linares, B., J. M. Guizar, N. Amador, A. Garcia, V. Miranda, J. R. Perez, and R. Chapela (2010): “Impact of air pollution on pulmonary function and respiratory symptoms in children. Longitudinal repeated-measures study,” *BMC Pulmonary Medicine*, 10, 1–9.
- Liu, H. and A. Salvo (2018): “Severe air pollution and child absences when schools and parents respond,” *Journal of Environmental Economics and Management*, 92, 300–330.
- Lleras-Muney, A. (2010): “The needs of the army using compulsory relocation in the military to estimate the effect of air pollutants on children’s health,” *Journal of Human Resources*, 45, 549–590.
- Marcotte, D. E. (2017): “Something in the air? Air quality and children’s educational outcomes,” *Economics of Education Review*, 56, 141–151.
- Montero, J.-M., G. Fernández-Avilés, and J. Mateu (2015): *Spatial and spatio-temporal geo-statistical modeling and kriging*, vol. 998, John Wiley & Sons.
- National Park Service (2018): “2018 Eruption and Summit Collapse,” <https://www.nps.gov/havo/learn/nature/2018-eruption.htm>, accessed: Dec 9 2022.
- Nieuwenhuijsen, M. J., X. Basagaña, P. Dadvand, D. Martinez, M. Cirach, R. Beelen, and B. Jacquemin (2014): “Air pollution and human fertility rates,” *Environment international*, 70, 9–14.
- Park, R. J., J. Goodman, M. Hurwitz, and J. Smith (2020): “Heat and learning,” *American Economic Journal: Economic Policy*, 12, 306–339.
- Perin, P. M., M. Maluf, C. E. Czeresnia, D. A. N. F. Januário, and P. H. N. Saldiva (2010): “Impact of short-term preconceptional exposure to particulate air pollution on treatment outcome in couples undergoing in vitro fertilization and embryo transfer (IVF/ET),” *Journal of assisted reproduction and genetics*, 27, 371–382.
- Persico, C. L. and J. Venator (2021): “The effects of local industrial pollution on students and schools,” *Journal of Human Resources*, 56, 406–445.
- Pham, L. and T. Roach (2023): “Particulate pollution and learning,” *Economics of Education Review*, 92, 102344.
- Slama, R., S. Bottagisi, I. Solansky, J. Lepeule, L. Giorgis-Allemand, and R. Sram (2013): “Short-term impact of atmospheric pollution on fecundability,” *Epidemiology*, 871–879.
- Stafford, T. M. (2015): “Indoor air quality and academic performance,” *Journal of Environmental Economics and Management*, 70, 34–50.
- United States Geological Survey (2018): “Chronology of Kīlauea’s summit eruption, 2008–2018,” <https://www.usgs.gov/volcanoes/kilauea/2008-2018>, accessed: Dec 9 2022.

- Wen, J. and M. Burke (2022): “Lower test scores from wildfire smoke exposure,” *Nature Sustainability*, 5, 947–955.
- Wooldridge, J. M. (2010): *Econometric analysis of cross section and panel data*, MIT press.
- Yao, Z., W. Zhang, X. Ji, and W. Weng (2023): “Short-term exposure to Air Pollution and Cognitive Performance: New Evidence from China’s College English Test,” *Environmental and Resource Economics*.
- Zhang, X., X. Chen, and X. Zhang (2018): “The impact of exposure to air pollution on cognitive performance,” *Proceedings of the National Academy of Sciences*, 115, 9193–9197.
- Zimmerman, D., C. Pavlik, A. Ruggles, and M. P. Armstrong (1999): “An experimental comparison of ordinary and universal kriging and inverse distance weighting,” *Mathematical Geology*, 31, 375–390.

## A.1 Technical Details of the Kriging Procedure

To fix ideas, we let  $s \in \{1, \dots, S\}$  denote the school, and  $t \in \{1, \dots, T\}$  denote the time periods, and  $m \in N(s)$  denote the monitoring station where  $N(s)$  is the neighborhood of school  $s$ . We consider three neighborhoods: the island of O’ahu, the southwestern part of Hawai‘i (that is most exposed to Kilauea’s emissions), and the remainder of Hawai‘i. These are depicted in Figures 4 and 5. We denote the pollution measurement at a given monitoring station on a particular day as  $\Pi_{mt}$ . The predicted exposure is then

$$\hat{P}_{st} = \sum_{m \in N(s)} \lambda_{sm}(NE_t) \Pi_{mt}$$

where the Kriging weights are  $\lambda_{sm}(NE_t) \equiv \lambda(d_{sm}, l_{sm}, NE_t)$ . Once again, bear in mind that  $NE_t \in \{0, 1\}$ .

The weights are chosen to guarantee that the predictions are unbiased and that the prediction error has minimum variance. Unbiasedness requires that the weights sum to unity. If we let  $\Pi_{st}$  represent the true pollution measurement at school  $s$ , both criteria can formally be written as

$$\begin{aligned} & \min_{\{\lambda_{sm}(b)\}_{m \in N(s)}} V \left( \hat{P}_{st} - \Pi_{st} \right) \\ & \text{subject to } \sum_{m \in N(s)} \lambda_{sm}(b) = 1 \end{aligned}$$

This minimization problem is solved twice: once for trade wind days ( $b = 1$ ) and once for non-trade wind days ( $b = 0$ ). This delivers two sets of weights which depend on the prevailing winds for that day.

Following [Montero et al. \(2015\)](#) (see p. 86) and making some local stationarity assumptions,

the first order conditions that guarantee these criteria are

$$\sum_{m \in N(s)} \lambda_{sm}(b) \gamma_{nm}(b) + \alpha_b = \gamma_{ns}(d_b) \text{ for } n \in N(s) \quad (\text{A.1})$$

$$\sum_{m \in N(s)} \lambda_{sm}(b) = 1 \quad (\text{A.2})$$

where both conditions hold for  $b \in \{0, 1\}$  and  $\alpha_b$  is the Lagrangian multiplier on the constraint in A.2 which guarantees the unbiasedness of the prediction. The object,  $\gamma_{nm}(b)$ , is the semi-variogram between locations  $m$  and  $n$  when  $NE_t = b$ . For each school in  $N(s)$ , equations A.1 and A.2 constitute a set of  $\#N(s) + 1$  equations in as many unknowns. So, we then have  $\#N(s) \times (\#N(s) + 1)$  equations in total. If we index the monitoring stations in  $N(s)$  from one to  $M(s) \equiv \#N(s)$  (with some abuse of notation) and define

$$\lambda_s(b) \equiv (\lambda_{s1}(b), \dots, \lambda_{sM(s)}(b), \alpha_b)'$$

then the Kriging weights are  $\lambda_s(b) = \Gamma(b)^{-1} \Gamma_s(b)$  where

$$\Gamma(b) \equiv \begin{bmatrix} \gamma_{11}(b) & \dots & \gamma_{1M(s)}(b) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{M(s)1}(b) & \dots & \gamma_{M(s)M(s)}(b) & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}$$

and

$$\Gamma_s(b) \equiv \begin{bmatrix} \gamma_{1s}(b) \\ \vdots \\ \gamma_{M(s)s}(b) \\ 1 \end{bmatrix}.$$

Hence, the task of computing the Kriging weights is reduced to computing the semivariogram

for  $NE_t = 1$  and  $NE_t = 0$ .

To compute the semivariogram, we postulate a functional form for the semivariogram. We assume that

$$\gamma_{mn}(NE_t) = 1 - \exp \left( -d_{mn} \times \left( \phi NE_t + \sum_{j \in \mathcal{L}} (\delta_j 1_{mn}(j) + \beta_j \times 1_{mn}(j) \times NE_t) \right) \right) \quad (\text{A.3})$$

where  $\mathcal{L} \equiv \{NE, SE, SW, NW\}$  (which collects the relative location variables) and  $1_{mn}(j)$  is an indicator for the location of  $n$  relative to  $m$  where  $m$  is held fixed. Note that distance enters multiplicatively to ensure that the semivariogram is zero when  $d_{mn} = 0$  and, by construction,  $\gamma_{mn}(NE_t) \in [0, 1]$ . This assumption reduces the spatial covariance structure to a smaller number of parameters which allows us to make extrapolations and interpolations needed to construct  $\Gamma_s(b)$ . We collect these parameters in the vector  $\theta$ .

We estimate these parameters using GMM. We let  $\widetilde{\gamma}_{mn}(b)$  denote the *empirical* semivariogram for the monitoring station pair  $(m, n)$  for  $b \in \{0, 1\}$ . Similarly,  $\widetilde{\Gamma}(b)$  is the matrix that collects the empirical semivariograms. Then for each pair  $(m, n)$ , we can compare two collections of moment conditions

$$q_b(\theta) = \text{vec} \left( \text{lower triangle} \left( \widetilde{\Gamma}(b) - \Gamma(b; \theta) \right) \right)$$

for  $b \in \{0, 1\}$ . We further collect these in the  $(M(s)+1) \times M(s) \times 2$  vector  $q(\theta) \equiv (q_0(\theta)', q_1(\theta)')'$ .

We estimate  $\theta$  by minimizing  $q(\theta)'q(\theta)$ .

Importantly,  $\theta$  is easy to estimate. First, compute the empirical semivariograms for trade wind ( $NE_t = 1$ ) and non-trade wind days ( $NE_t = 0$ ). Second, we re-write [A.3](#) as

$$1 - \gamma_{mn}(NE_t) = \exp \left( -d_{mn} \times \left( \phi NE_t + \sum_{j \in \mathcal{L}} (\delta_j 1_{mn}(j) + \beta_j \times 1_{mn}(j) \times NE_t) \right) \right)$$

and we then note that estimation can proceed by using a simple Poisson regression package in



Stata or R. Once  $\theta$  is estimated, we can then estimate the Kriging weights  $\lambda_s(b)$ .

## A.2 Cross-Validation Exercise

We now conduct a simple cross-validation exercise to investigate the accuracy of our three predictors. For the monitoring stations in each of our three groups (O’ahu, south Hawai‘i, and north Hawai‘i), we exclude information from one station then use the remaining stations in the group to predict its pollution levels. We then compare the actual pollution levels that were measured by the station with the predicted pollution levels.

To gauge the accuracy of the prediction, we compute two sets of statistics. The first is the  $R^2$  from a regression of the actual pollution measure onto the predicted pollution measure. The second is the Mean Squared Error (MSE) for the prediction i.e.  $\frac{1}{T} \sum_{t=1}^T (P_{mt} - \widehat{P}_{mt})^2$ . Note that the MSE that we compute is *not* the sum of squared residuals from the regression in which we compute the  $R^2$  and so the usual identity will *not* hold.

In Figures A3 and A4, we display the  $R^2$  and MSE for each station. In both of these figures, we depict the statistic in question for each group and each monitoring station in that group. We do so for  $PM_{2.5}$  in the left panel and  $SO_2$  in the right panel of each figure. Accordingly, Figures A3 and A4 each have six plots corresponding to three regional groupings and two pollutants. Each of these statistics is informative of the accuracy of the predictors but the MSE will tend to be larger when the object that is being predicted is more volatile.

Looking at the  $R^2$  of the predictions of  $PM_{2.5}$  in the left panel of Figure A3, we see considerable variation in the quality of the predictions depending on the grouping and the type of weighting (e.g. K-weights, ID-weights, or U-weights). In many cases, Kriging does not perform well for  $PM_{2.5}$  which is interesting since Kriging (under appropriate conditions) delivers the BLUP. On O’ahu, we can predict roughly between 20 and 60% of the variation in  $PM_{2.5}$ . K-weights perform particularly poorly in Pearl City and, especially, Kapolei which are further away from the urban core of Honolulu. This might suggest that schools that are further away

from monitoring stations also have poorer predictions perhaps due to abnormally small or large weights. In south Hawai‘i, our predictions are more accurate with  $R^2$ ’s sometimes exceeding 60% (e.g. Kona and Ocean View). On the whole, K-weights perform better on south Hawai‘i perhaps due to the nature of its geography. The predictions are the lowest in the north Hawai‘i region with  $R^2$ ’s between 10 and 40%. As with south Hawai‘i, K-weights perform substantially better in north Hawai‘i than they do on O‘ahu. Importantly, ID-weight appear to offer little advantage over U-weights or simple local raw means.

The estimated  $R^2$ ’s for  $SO_2$  predictions in Figure A3 tell a very different story than they do for  $PM_{2.5}$ . First, in contrast to the case of  $PM_{2.5}$ , K-weighted predictions perform much better for  $SO_2$ . Particularly, we see that Kriging based predictors dominate for two of three monitoring stations on south Hawai‘i and one of two monitoring stations on north Hawai‘i. Also, Kriging never performs as poorly as it did for  $PM_{2.5}$  for some of the monitoring stations on O‘ahu e.g. Kapolei. Second, the  $R^2$ ’s for the  $SO_2$  predictions on O‘ahu are zero. This reflects two factors. First, there are only two monitoring stations for  $SO_2$  on O‘ahu. Accordingly, the prediction of  $SO_2$  levels at one station is just the level at the other station. Second, as reflected in Figure A2, the spatial correlations in  $SO_2$  are effectively zero except when one station is to the northwest of the other station and relatively close (e.g. within 50 miles). The monitoring station at Honolulu is due east of the Kapolei station and about 15 miles separate them. Taken together, this suggests that the  $SO_2$  predictions on O‘ahu could be quite poor which should be borne in mind as we proceed. Finally, as was the case with  $PM_{2.5}$ , ID-weights offer no ostensible advantage to using U-weights when predicting  $SO_2$ .

The calculations of the MSEs displayed in Figure A4 reveal several interesting patterns. We still see that K-weights perform well on north Hawai‘i for both  $PM_{2.5}$  and  $SO_2$  and on south Hawai‘i for  $SO_2$ . However, K-weights still do not perform well on O‘ahu when compared to ID- and U-weights, particularly, at the more remote stations of Kapolei and Pearl City. We also observe that all of the MSE’s (i.e. for each weighting scheme) are *very* high on north Hawai‘i and, especially on south Hawai‘i. This is particularly the case for  $SO_2$  levels near Kilauea. This

most likely reflected the high volatility of  $SO_2$  near Kilauea.

### A.3 Balance Test

Though the Hawai‘i context likely provides exogenous variation for the identification of the effects of pollutants on test scores, we can still test whether there are observable differences in student characteristics correlated with pollutant exposures. In Appendix Table A1, we regress our pollution measures on the full vector of student and exam characteristics using all three of our weighting schemes. Overall, we find little to no evidence of correlations between observable characteristics and pollutant levels. Excluding the estimates of the grade fixed effects in the second part of the table, the only significant coefficient in either column is for receipt of English language services which appears to predict  $PM_{2.5}$  levels. However, this is the only significant variable across all columns. We do see that the grade indicators predict  $SO_2$  levels and  $PM_{2.5}$  levels for 5th grade using our inverse distance and uniform weighting schemes, but this is easily dealt with by the inclusion of grade fixed effects in the estimations. Finally, we note that an  $F$ -test that all of the covariates in the estimates are zero resoundingly fails to reject the null. All told, we suspect that this significant estimate in the table is the consequence of Type I error. Thus, we conclude that the variation in pollutants has no systematic relationships with observable confounders.

### A.4 Test Date Delays

One potential threat to the validity of our results is that although schools dictate test dates, students might stay home from school on high pollution days, taking their make-up exam on a later day when the air quality is better. Importantly, this mechanism would suggest a *positive* bias as pupils ostensibly should be performing better on the delayed date with lower pollution levels. As such, this vignette does not pose much of a threat to our findings as it suggest that we

are underestimating the true effect of pollution on test scores.

In Figure A5, we present histograms of test dates for a sample of six schools. We note that there are 230 schools in our sample spanning seven grades. With the exception of one school (Robert Louis Stevenson Middle), the distribution of test dates tends to be concentrated around a single mode. James B Campbell High School (in the top left of the figure) does show slightly more dispersion but it is the largest school in the State of Hawaii with over 3000 pupils.

In Table A11, we estimate the same models as in Table 4 but include a binary control variable for the student delaying their test date. Throughout, we only employ school fixed effects. This indicator variable is turned on when the pupil's test date is either above the modal test date (odd columns) or the median test date (even columns). The modes and medians were computed within the school/grade/exam (math or english).

The estimates in Table A11 are larger in magnitude than those in Table 4. Prima facie, this appears to lend credence to the story in which students delay their exams on high pollution days and this enables them to obtain higher scores when they do take their exams. However, the indicator for delayed exam is *negative* and significant. This most likely reflects that students who are absent from school more often tend to perform worse - a negative selection effect. Accordingly, in order to obtain a positive bias from the delayed test indicator, it must be the case that students are more likely to delay their test when pollution is lower (not higher).<sup>33</sup> All told, delaying tests in the face of poor air quality is neither a plausible nor even possible explanation for our finding that poorer air quality causes students to perform worse on standardized tests.

## A.5 Home Displacement Effects on Test Scores

While volcanic activity is a recurring phenomenon on Hawai'i Island, the year 2018 witnessed a significant and disruptive eruption, leading to the destruction of over 600 homes, with

---

<sup>33</sup>We do not report these results, but this is indeed the case. We do not have student absenteeism data at the daily level, so we cannot make statements about the effects of pollution on absenteeism on the day of the exam.

nearly 300 being primary residences, and the displacement of approximately 2,000 individuals. It is possible that the adverse impact of pollution on test scores may be instead driven by the heightened stress and mental health challenges faced by residents who underwent displacement due to the volcanic activity.

To address this potential threat to our analysis, we exclude schools that were located within the Puna district, which saw property damage from the 2018 volcano for the months of May (when the eruptions started) through the end of our sample period. The results are displayed in Table [A12](#) of the appendix. Even after excluding Puna schools, our findings remain similar to our main results in Table [4](#). Therefore, we do not believe the negative effect of pollution on test scores were driven by the stress of displaced residents.

## A.6 Additional Tables and Figures

Table A1: Balance Test

	Kriging		Inverse Distance		Uniform Weights	
	(1) $PM_{2.5}$	(2) $SO_2$	(3) $PM_{2.5}$	(4) $SO_2$	(5) $PM_{2.5}$	(6) $SO_2$
Female	-0.008 (0.014)	-0.001 (0.004)	-0.003 (0.009)	0.004 (0.008)	-0.004 (0.009)	0.014 (0.011)
Economically Disadvantaged	-0.028 (0.030)	0.006 (0.008)	-0.034 (0.025)	0.015 (0.018)	-0.026 (0.024)	0.018 (0.021)
Received English Language Services	0.085* (0.049)	-0.004 (0.013)	0.006 (0.027)	0.020 (0.029)	0.022 (0.026)	0.025 (0.029)
Math Exam	0.031 (0.222)	0.040 (0.049)	-0.071 (0.117)	0.115 (0.100)	-0.102 (0.112)	0.001 (0.128)
Air Temperature	-0.422*** (0.062)	0.032* (0.017)	-0.157*** (0.045)	0.102*** (0.036)	-0.150*** (0.043)	0.140*** (0.047)
Filipino	-0.017 (0.025)	-0.001 (0.005)	0.008 (0.014)	-0.013 (0.012)	0.003 (0.013)	0.004 (0.012)
Native Hawaiian	0.027 (0.030)	-0.002 (0.009)	0.010 (0.021)	0.012 (0.016)	0.004 (0.021)	0.017 (0.020)
Pacific Islander	-0.007 (0.031)	-0.003 (0.010)	0.014 (0.021)	-0.020 (0.024)	0.011 (0.021)	-0.010 (0.024)
White	-0.010 (0.024)	-0.013* (0.007)	0.018 (0.015)	-0.003 (0.014)	0.016 (0.016)	0.004 (0.015)
Other Ethnicity	-0.021 (0.030)	-0.010 (0.010)	0.010 (0.020)	0.002 (0.017)	0.006	0.004

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Notes: Standard errors clustered by school. Regressors include female, economic disadvantage status, non-native English speaking, math exam, air temperature, grade level and ethnicity. All estimations control for school, month and academic year fixed effects. The F-test tests for whether the covariates in each model are jointly equal to zero.

Table A1: Balance Test Continued

	Kriging		Inverse Distance		Uniform Weights	
	(1)	(2)	(3)	(4)	(5)	(6)
	$PM_{2.5}$	$SO_2$	$PM_{2.5}$	$SO_2$	$PM_{2.5}$	$SO_2$
4th Grade	0.186 (0.188)	-0.135 (0.083)	-0.041 (0.091)	-0.221* (0.124)	-0.034 (0.090)	-0.404** (0.183)
5th Grade	-0.013 (0.155)	-0.058 (0.049)	-0.164** (0.068)	-0.119* (0.069)	-0.157** (0.067)	-0.071 (0.088)
6th Grade	0.037 (0.200)	-0.116** (0.051)	-0.047 (0.087)	-0.380*** (0.125)	-0.051 (0.088)	-0.436*** (0.131)
7th Grade	-0.180 (0.222)	-0.169** (0.070)	-0.047 (0.131)	-0.385** (0.151)	-0.012 (0.123)	-0.462** (0.179)
8th Grade	0.019 (0.219)	-0.188*** (0.070)	-0.055 (0.122)	-0.465*** (0.142)	-0.015 (0.121)	-0.433** (0.200)
11th Grade	-0.246 (0.283)	0.010 (0.173)	-0.358* (0.215)	-0.794 (0.564)	-0.383	-0.504
Month FE	X	X	X	X	X	X
Academic Year FE	X	X	X	X	X	X
School FE	X	X	X	X	X	X
$R^2$	0.315	0.672	0.398	0.545	0.377	0.623
F-test	1.019	1.159	1.043	1.249	1.103	1.222
p-value	0.498	0.814	0.354	0.596	0.766	0.470

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Notes: Standard errors clustered by school. Regressors include female, economic disadvantage status, non-native English speaking, math exam, air temperature, grade level and ethnicity. All estimations control for school, month and academic year fixed effects. The F-test tests for whether the covariates in each model are jointly equal to zero.

Table A2: Summary Statistics (Hawai‘i Data eXchange Partnership Student Measures) by Distance from Schools to Nearest Monitoring Station

	<i>&lt; 2mi.</i>		<i>≥ 2mi., &lt; 5mi.</i>		<i>≥ 5mi., &lt; 10mi.</i>		<i>&gt; 10mi.</i>	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Economically Disadvantaged	0.55	0.19	0.48	0.23	0.48	0.22	0.59	0.20
Received English Language Services	0.09	0.09	0.07	0.09	0.04	0.04	0.05	0.06
Female	0.50	0.03	0.50	0.03	0.50	0.04	0.50	0.03
Asian (Non-Filipino)	0.25	0.19	0.19	0.19	0.16	0.17	0.07	0.09
Filipino	0.24	0.19	0.23	0.19	0.15	0.13	0.13	0.10
Native Hawaiian	0.22	0.18	0.19	0.20	0.32	0.23	0.39	0.17
Pacific Islander	0.13	0.10	0.12	0.15	0.06	0.05	0.09	0.10
White	0.10	0.08	0.18	0.16	0.23	0.17	0.26	0.18
Other Ethnicity	0.06	0.05	0.08	0.08	0.08	0.07	0.06	0.02
Number of Schools	43		79		64		44	

Notes: Data on student summary statistics comes from the Hawai‘i Data eXchange Partnership for the years 2015-2018. Economically disadvantaged students refer to those who qualify for federal programs such as free and reduced lunch. Those who received English language services are students who enrolled in the State of Hawai‘i Department of Education’s English Learner Program for the academic school year.



Table A3: Effect of Pollution on Test Scores: Restricting the Sample Based on Distance to Nearest Air Monitoring Station

	(1) 2 mi	(2) 2 mi	(3) 5 mi	(4) 5 mi	(5) 10 mi	(6) 10 mi
<i>Panel A: Kriging</i>						
$PM_{2.5}$	-0.428*** (0.105)		-0.155* (0.083)		-0.126 (0.080)	
$SO_2$		-0.932* (0.461)		-0.879* (0.498)		-0.549* (0.319)
$R^2$	0.296	0.281	0.266	0.263	0.260	0.262
<i>Panel B: Inverse Distance</i>						
$PM_{2.5}$	-0.577*** (0.186)		-0.469*** (0.126)		-0.366*** (0.108)	
$SO_2$		-0.080 (0.097)		0.006 (0.079)		-0.027 (0.078)
$R^2$	0.293	0.288	0.265	0.265	0.260	0.263
<i>Panel C: Uniform Weights</i>						
$PM_{2.5}$	-0.514*** (0.154)		-0.484*** (0.120)		-0.374*** (0.104)	
$SO_2$		-0.083 (0.081)		-0.056 (0.086)		-0.059 (0.076)
$R^2$	0.293	0.288	0.265	0.265	0.260	0.263

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Notes: Standard errors clustered by school. Regressors include female, economic disadvantage status, non-native English speaking, math exam, air temperature, grade level and ethnicity. All estimations control for school, month and academic year fixed effects. Estimates in columns 1 and 4 include schools that are within 2 miles of a nearby pollution monitoring station. Estimates in columns 2 and 5 include schools that are within 5 miles of a nearby pollution monitoring station. Estimates in columns 3 and 6 include schools that are within 10 miles of a nearby pollution monitoring station.

Table A4: Effects of Pollutants on Exam Scores for Students by Region with Individual Fixed Effects

	O‘ahu		S/SW Hawai‘i		N/NE Hawai‘i	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: Kriging</i>						
$PM_{2.5}$	-0.004 (0.094)		-0.002 (0.163)		0.018 (0.178)	
$SO_2$		1.940 (2.060)		0.096 (0.172)		0.043 (0.320)
$R^2$	0.858	0.856	0.843	0.842	0.823	0.824
<i>Panel B: Inverse Distance</i>						
$PM_{2.5}$	-0.261 (0.250)		0.037 (0.171)		0.007 (0.180)	
$SO_2$		1.040 (1.710)		0.0169 (0.053)		0.041 (0.119)
$R^2$	0.856	0.856	0.844	0.844	0.823	0.823
<i>Panel C: Uniform Weights</i>						
$PM_{2.5}$	-0.298 (0.269)		0.088 (0.154)		0.003 (0.186)	
$SO_2$		1.150 (1.700)		-0.007 (0.052)		0.033 (0.154)
$R^2$	0.856	0.856	0.844	0.844	0.823	0.823
Number of Schools	176	176	16	16	38	38

Notes: Standard errors are clustered by school. Control variables include economic disadvantage status, English language service recipients, exam subject, air temperature, and grade level. All estimations control for school, month, academic year, and individual fixed effects. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table A5: Percent of Economically Disadvantaged Students Within Schools Exposure Effects

	Kriging		Inverse Dist		Uniform	
	(1)	(2)	(3)	(4)	(5)	(6)
Quintile 1 $\times PM_{2.5}$	-0.002 (0.196)	-0.082 (0.146)	0.066 (0.344)	-0.552 (0.480)	-0.020 (0.378)	-0.561 (0.535)
Quintile 2 $\times PM_{2.5}$	0.091 (0.119)	0.080 (0.151)	-0.007 (0.271)	-0.141 (0.305)	-0.022 (0.275)	-0.098 (0.323)
Quintile 3 $\times PM_{2.5}$	-0.335** (0.138)	-0.201 (0.154)	-0.564** (0.219)	-0.395* (0.223)	-0.600** (0.237)	-0.468** (0.230)
Quintile 4 $\times PM_{2.5}$	-0.303* (0.170)	-0.051 (0.162)	-0.243 (0.161)	-0.167 (0.156)	-0.230 (0.147)	-0.164 (0.144)
Quintile 5 $\times PM_{2.5}$	-0.166 (0.156)	-0.133 (0.166)	-0.301* (0.163)	-0.125 (0.177)	-0.283* (0.160)	-0.097 (0.165)
Month FE	X	X	X	X	X	X
Academic Year FE	X	X	X	X	X	X
School FE	X	X	X	X	X	X
Individual FE		X		X		X
$R^2$	0.267	0.851	0.268	0.850	0.268	0.850

Notes: Standard errors are clustered by school. Control variables include gender, economic disadvantage status, English language service recipients, exam subject, air temperature, grade level and ethnicity. All estimations control for school, month and academic year fixed effects. The F-statistic tests the null that coefficients on the bottom four  $PM_{2.5}$  quintile interactions are equal to zero. The p-value for this test is displayed below. Daily average  $PM_{2.5}$  quintiles were calculated by taking the daily average  $PM_{2.5}$  level over the entire sample period. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table A6: Effect of Pollution on Exam Scores by Gender

	$PM_{2.5}$		$SO_2$	
	Male	Female	Male	Female
<i>Panel A: Kriging</i>				
Pollutant	−0.182** (0.083)	−0.141* (0.079)	−0.131 (0.203)	−0.214 (0.211)
$R^2$	0.265	0.273	0.264	0.272
<i>Panel B: Inverse Distance</i>				
Pollutant	−0.358*** (0.107)	−0.191* (0.113)	−0.058 (0.098)	−0.062 (0.085)
$R^2$	0.266	0.273	0.266	0.273
<i>Panel C: Uniform Weights</i>				
Pollutant	−0.388*** (0.110)	−0.179 (0.109)	−0.053 (0.076)	−0.070 (0.079)
$R^2$	0.266	0.273	0.266	0.273
Number of Schools	230	230	230	230
School FE	X	X	X	X
Month FE	X	X	X	X
Year FE	X	X	X	X

Notes: Standard errors are clustered by school. Control variables include economic disadvantage status, English language service recipients, exam subject, air temperature, grade level and ethnicity. All estimations control for school, month and academic year fixed effects. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table A7: Effect of Pollution on Math and Reading Scores with Additional Controls

	(1)	(2)	(3)	(4)	(5)
<i>Panel A: Kriging</i>					
$PM_{2.5}$	-0.160** (0.073)	-0.081 (0.073)	-0.177** (0.079)	-0.108 (0.076)	-0.173 (0.118)
$R^2$	0.267	0.851	0.267	0.267	0.272
<i>Panel B: Inverse Distance</i>					
$PM_{2.5}$	-0.271*** (0.093)	-0.227** (0.103)	-0.284*** (0.097)	-0.193* (0.108)	-0.075 (0.176)
$R^2$	0.268	0.850	0.268	0.268	0.272
<i>Panel C: Uniform Weights</i>					
$PM_{2.5}$	-0.278*** (0.091)	-0.227** (0.101)	-0.293*** (0.095)	-0.197* (0.105)	-0.120 (0.184)
$R^2$	0.268	0.850	0.268	0.268	0.272
Number of Schools	230	230	230	230	230
School FE	X	X	X	X	X
Month FE	X	X	X	X	X
Year FE	X	X	X	X	X
Individual FE		X			
Temperature Quintiles			X		
Prev. 12-month Temp Quintiles				X	
Island $\times$ Date FE					X

Notes: Standard errors are clustered by school. Control variables include gender, economically disadvantaged students, English language service recipients, exam subject, air temperature, grade level and ethnicity. All estimations control for school, month and academic year fixed effects. Column 3 includes temperature quintile dummy variables that bin temperature by island-date. Column 4 includes temperature quintile dummy variables based on the previous 12-month average temperature by island-date \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table A8: Effect of Pollution on Math and Reading Scores with Rainfall Quintile Controls

	(1)	(2)	(3)	(4)	(5)
<i>Panel A: Kriging</i>					
$PM_{2.5}$	-0.150* (0.090)	-0.064 (0.094)	-0.171* (0.098)	-0.110 (0.092)	-0.179 (0.129)
$R^2$	0.265	0.857	0.266	0.265	0.271
<i>Panel B: Inverse Distance</i>					
$PM_{2.5}$	-0.405*** (0.124)	-0.347** (0.148)	-0.420*** (0.131)	-0.374** (0.147)	-0.080 (0.197)
$R^2$	0.265	0.856	0.266	0.266	0.271
<i>Panel C: Uniform Weights</i>					
$PM_{2.5}$	-0.390*** (0.116)	-0.333** (0.145)	-0.413*** (0.123)	-0.360*** (0.137)	-0.107 (0.216)
$R^2$	0.265	0.856	0.266	0.265	0.271
Number of Schools	204	197	204	204	204
School FE	X	X	X	X	X
Month FE	X	X	X	X	X
Year FE	X	X	X	X	X
Individual FE		X			
Temperature Quintiles			X		
Prev. 12-month Temp Quintiles				X	
Island $\times$ Date FE					X
Rainfall Quintiles	X	X	X	X	X

Notes: Standard errors are clustered by school. Control variables include gender, economically disadvantaged students, English language service recipients, exam subject, air temperature, grade level, ethnicity, and quintiles for rainfall by island-date. All estimations control for school, month and academic year fixed effects. Rainfall for each school was estimated by assigning the total daily rainfall in millimeters from the nearest monitoring station to the school. Column 3 includes temperature quintile dummy variables that bin temperature by island-date. Column 4 includes temperature quintile dummy variables based on the previous 12-month average temperature by island-date \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table A9: Effect of Pollution on Math and Reading Scores Using Satellite Data

	(1) $PM_{2.5}$	(2) $PM_{2.5}$	(3) $SO_2$	(4) $SO_2$
Pollutant	-0.025 (0.179)	-0.072 (0.155)	6.6e+08 (1.1e+09)	1.4e+09 (1.3e+09)
$R^2$	0.267	0.850	0.267	0.850
Number of Schools	230	230	230	230
School FE	X	X	X	X
Month FE	X	X	X	X
Year FE	X	X	X	X
Individual FE		X		X

Notes: Standard errors are clustered by school. Control variables include gender, economically disadvantaged students, English language service recipients, exam subject, air temperature, grade level and ethnicity. All estimations control for school, month and academic year fixed effects. The particulates,  $PM_{2.5}$  and  $SO_2$ , are reported in  $\mu g/m^3$  and  $kg/m^3$ , respectively. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table A10: Correlations Between MERRA Satellite Data and Hawaii Department of Health Pollution Data by School Complex

Hawai'i Complex	$PM_{2.5}$ Corr.	$SO_2$ Corr.	O'ahu Complex	$PM_{2.5}$ Corr.	$SO_2$ Corr.
Hilo	-0.284	0.320	Aiea	0.370	0.337
Kau	0.312	0.194	Campbell	0.610	0.254
Keaau	-0.213	0.111	Castle	0.496	0.462
Kealakehe	0.109	-0.003	Farrington	0.480	0.368
Konawaena	0.346	-0.007	Kahuku	0.584	0.263
Laupahoehoe	0.017	0.098	Kailua	0.484	0.335
Pahoa	-0.259	0.233	Kaimuki	0.571	0.235
Waiakea	0.052	0.284	Kaiser	0.706	0.074
			Kalaheo	0.767	0.236
			Kalani	0.424	0.196
			Kapolei	0.419	0.301
			Leilehua	0.429	0.274
			McKinley	0.432	0.209
			Mililani	0.280	0.351
			Moanalua	0.549	0.286
			Nanakuli	0.552	0.485
			Pearl City	0.591	0.301
			Radford	0.558	0.486
			Roosevelt	0.322	0.381
			Waialua	0.599	0.281
			Waianae	0.534	0.480
			Waipahu	0.511	0.397

Notes: Data on pollutant measures come from the State of Hawai'i Department of Health and the Environmental Protection Agency (EPA) Air Quality System data. Correlations between measures of  $PM_{2.5}$  and  $SO_2$  are reported for each school complex on O'ahu and Hawai'i island. A school complex consists of a neighborhood's high school and the elementary and middle schools that feed into it. There are fewer complexes on Hawai'i island due to the island's much smaller population compared to O'ahu.



Table A11: OLS Estimates With Delayed Test Date Control

	$PM_{2.5}$		$SO_2$	
	Modal Date (1)	Median Date (2)	Modal Date (3)	Median Date (4)
<i>Panel A: Kriging</i>				
Pollutant	-0.170** (0.076)	-0.180** (0.076)	-0.116 (0.173)	-0.139 (0.175)
Delayed Test	-10.805*** (1.049)	-10.109*** (1.325)	-10.686*** (1.125)	-10.006*** (1.349)
$R^2$	0.269	0.263	0.261	0.261
<i>Panel B: Inverse Distance</i>				
Pollutant	-0.325*** (0.090)	-0.330*** (0.091)	-0.073 (0.072)	-0.080 (0.072)
Delayed Test	-10.845*** (1.098)	-10.197*** (1.327)	-10.764*** (1.117)	-10.098*** (1.336)
$R^2$	0.269	0.263	0.263	0.263
<i>Panel C: Uniform Weights</i>				
Pollutant	-0.341*** (0.090)	-0.343*** (0.090)	-0.053 (0.068)	-0.056 (0.068)
Delayed Test	-10.861*** (1.098)	-10.209*** (1.327)	-10.759*** (1.117)	-10.089*** (1.336)
$R^2$	0.269	0.263	0.263	0.263
Number of Schools	230	230	230	230
School FE	X	X	X	X
Month FE	X	X	X	X
Year FE	X	X	X	X

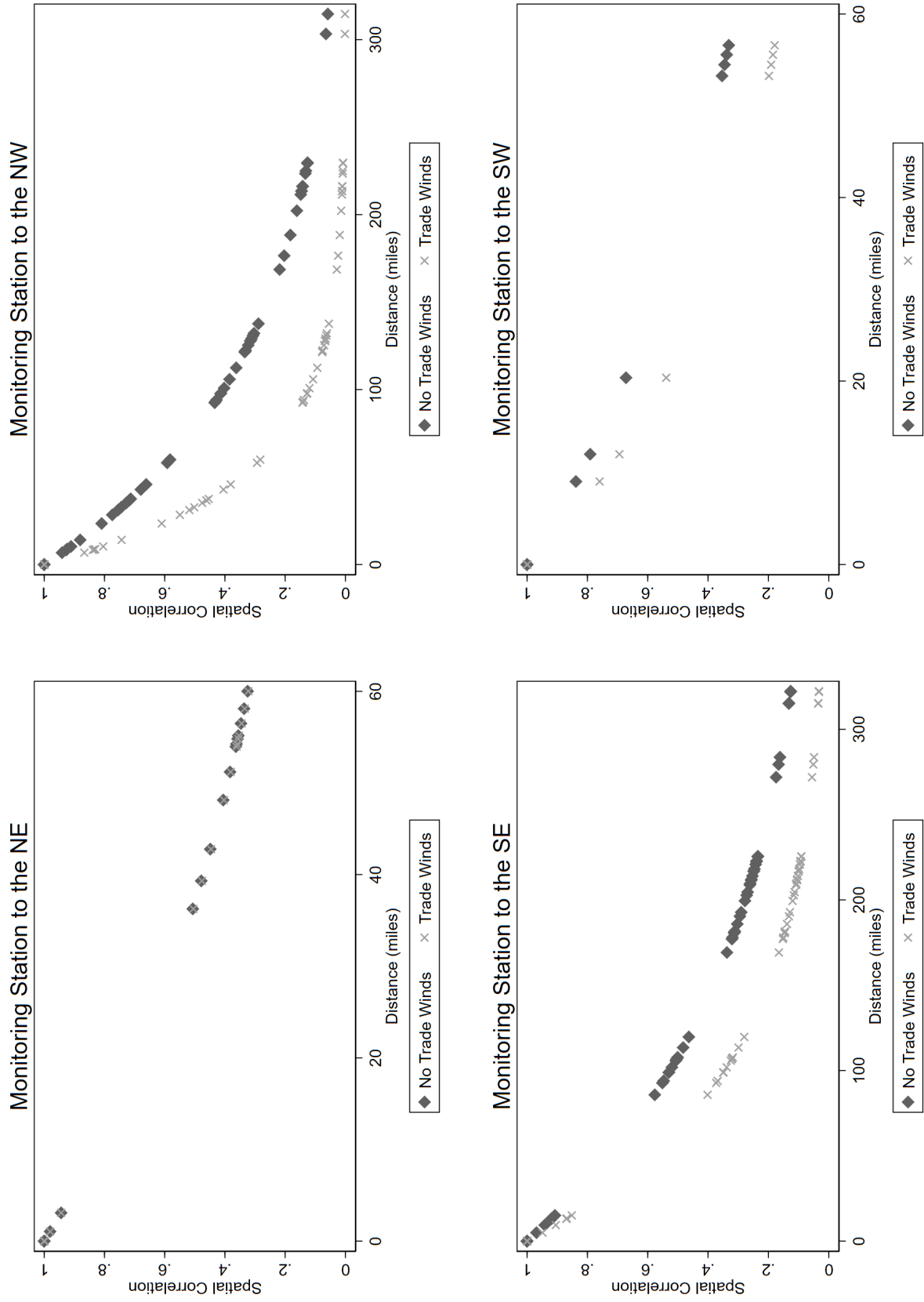
Notes: Standard errors are clustered by school. Control variables include a dummy equal to 1 if a student has a delayed exam date and 0 otherwise, gender, economic disadvantage status, English language service recipients, exam subject, air temperature, grade level and ethnicity. All estimations control for school, month and academic year fixed effects. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table A12: Effects Excluding May through August 2018 Eruption-Affected Schools

	$PM_{2.5}$		$SO_2$	
	(1)	(2)	(3)	(4)
<i>Panel A: Kriging</i>				
Pollutant	−0.163** (0.073)	−0.080 (0.074)	−0.233 (0.185)	−0.106 (0.192)
$R^2$	0.267	0.851	0.266	0.850
<i>Panel B: Inverse Distance</i>				
Pollutant	−0.278*** (0.094)	−0.221** (0.105)	−0.081 (0.073)	−0.032 (0.071)
$R^2$	0.268	0.850	0.268	0.850
<i>Panel C: Uniform Weights</i>				
Pollutant	−0.287*** (0.092)	−0.222** (0.103)	−0.074 (0.068)	−0.069 (0.066)
$R^2$	0.268	0.850	0.268	0.850
Number of Schools	230	230	230	230
School FE	X	X	X	X
Month FE	X	X	X	X
Year FE	X	X	X	X
Individual FE		X		X

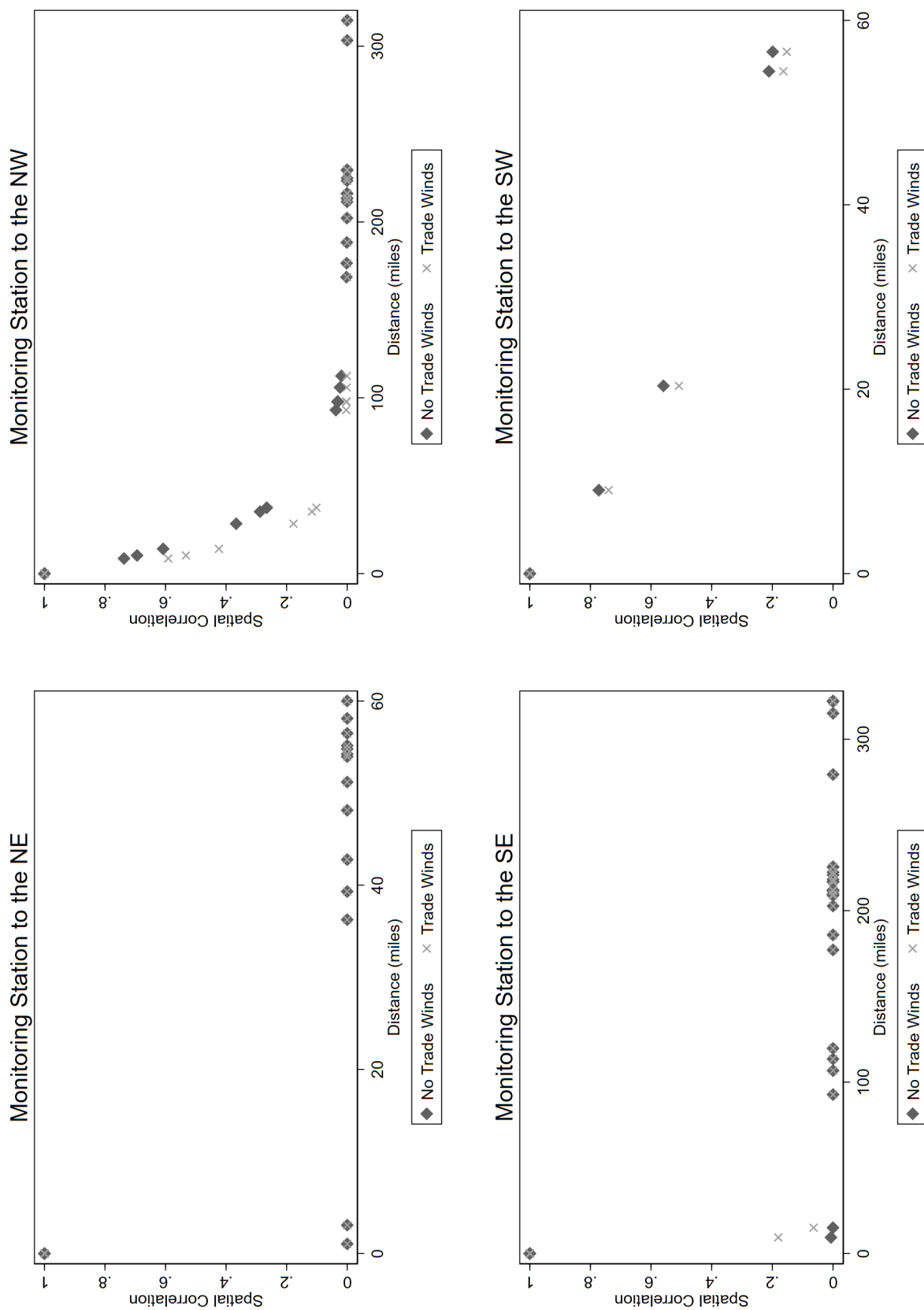
Notes: Standard errors are clustered by school. Control variables include gender, economic disadvantage status, English language service recipients, exam subject, air temperature, grade level and ethnicity. All estimations control for school, month and academic year fixed effects. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Figure A1: Covariograms for  $PM_{2.5}$



Notes: Displays estimated covariograms by the relative locations of the monitoring stations and by the direction of the trade winds.

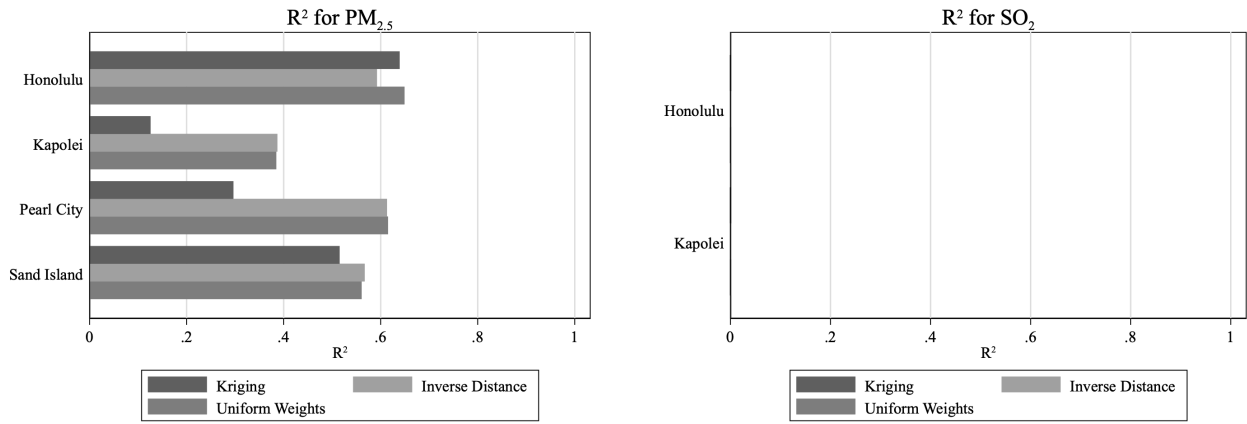
Figure A2: Covariograms for  $SO_2$



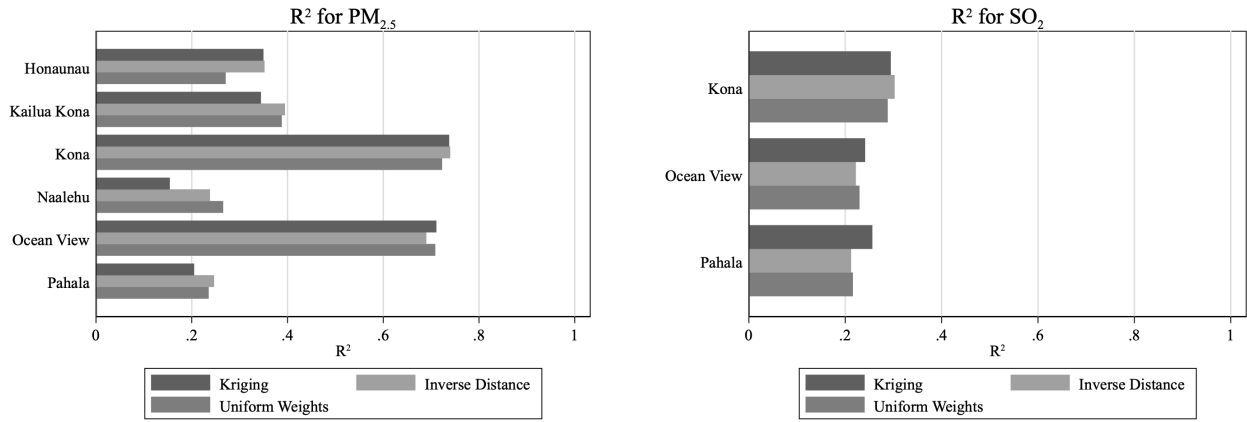
Notes: Per Figure A1

Figure A3: Cross-Validation:  $R^2$  of Predictions

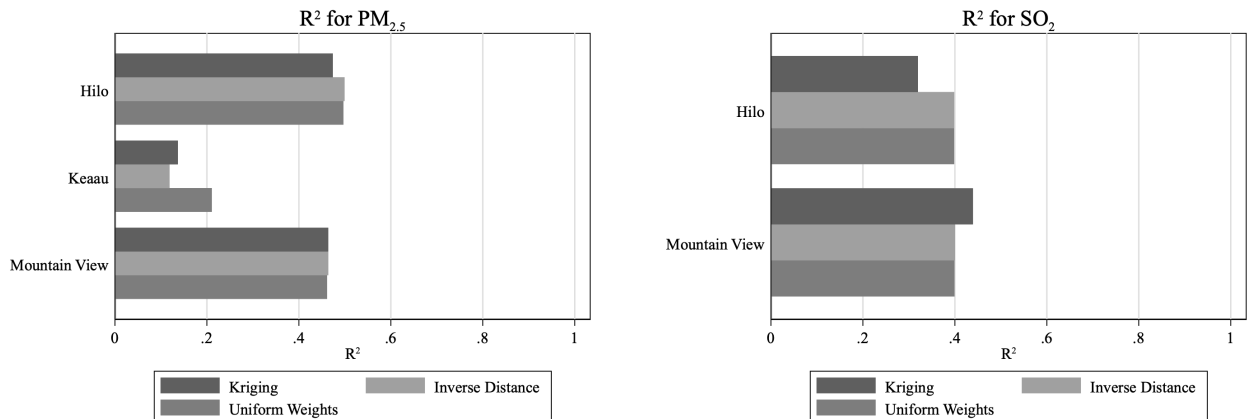
O'ahu



South/Southwest Hawai'i



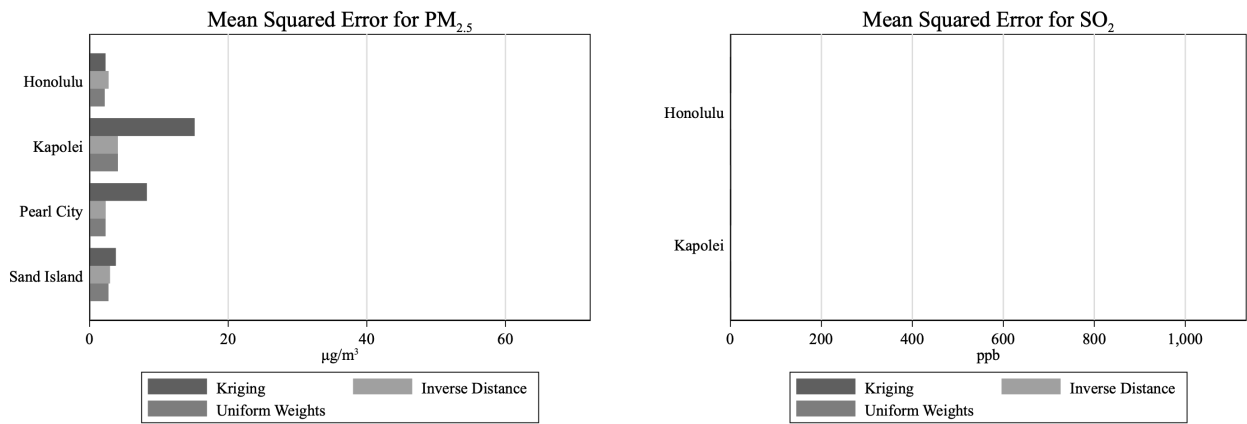
North/Northeast Hawai'i



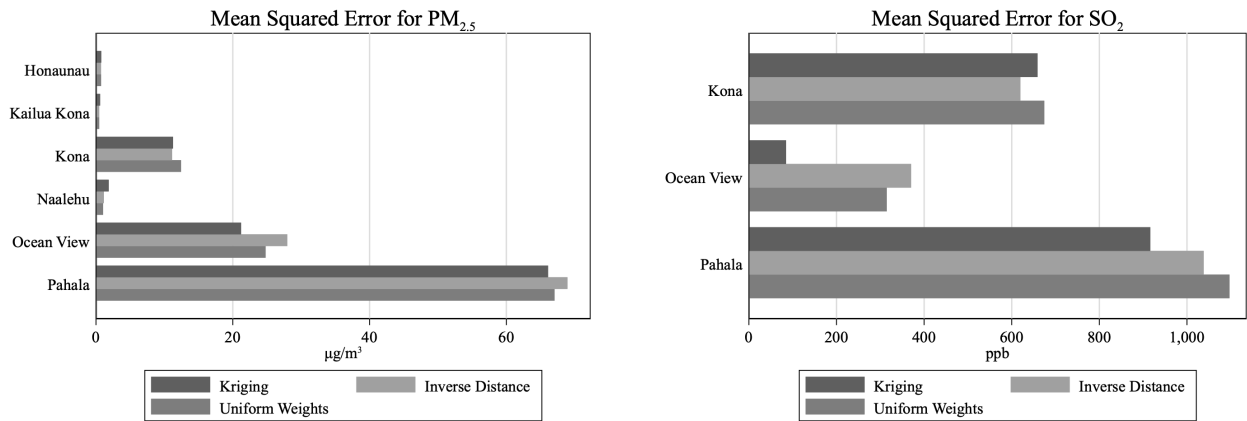
Notes: This figure displays the  $R^2$  of regressions of pollution measures at a given monitoring in a given group onto its prediction when the monitoring station in question is excluded from the prediction.

Figure A4: Cross-Validation: Mean-Squared Error of Predictions

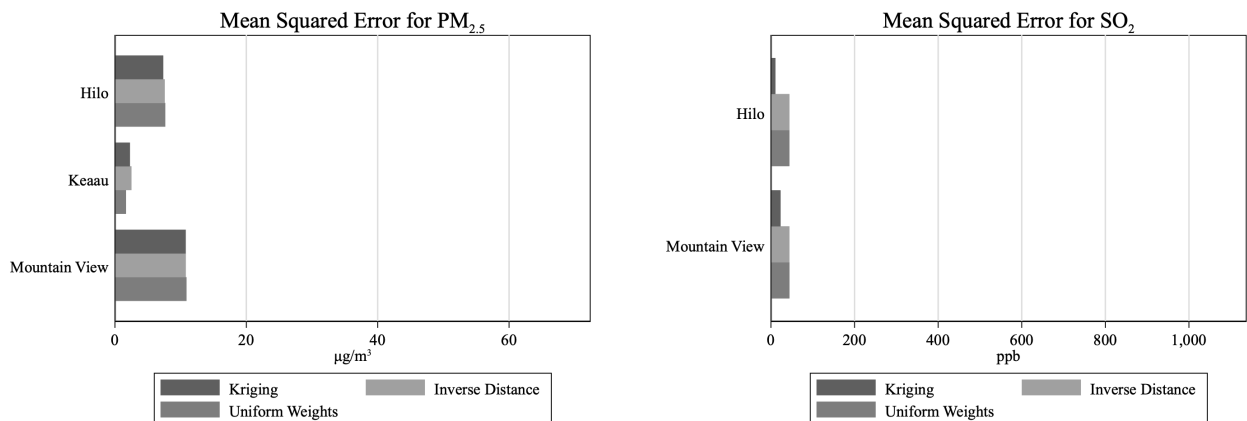
### O'ahu



### South/Southwest Hawai'i



### North/Northeast Hawai'i



Notes: This figure displays the Mean-Squared Errors of pollution predictions from the cross-validation exercise.

Figure A5: Math SBA Exam Dates by School, 2015

