

# Pollution, agricultural productivity, and development: Evidence from coal plants in India\*

Joshua D. Merfeld<sup>†</sup>

2023-03-02

## Abstract

Pollution is a growing problem in many developing countries. Using the rollout of coal plants to identify high-pollution locations in India and random variation induced by changes in wind direction, I show that pollution has large, negative impacts on agricultural productivity. Results indicate that some of this impact is driven by changes in labor allocation patterns in response to increased exposure to pollution. I also show that pollution from coal plants, specifically, has a larger effect on agricultural productivity than other types of pollution.

**Keywords:** pollution, productivity, agriculture, labor, India

**JEL Codes:** H40, I15, J22, O13, Q52, Q53

---

\*I thank Marc Bellemare and David Sungho Park for comments and suggestions that greatly improved the quality of the paper. The usual disclaimer applies. All scripts used in this paper are publicly available on my GitHub page: [github.com/JoshMerfeld/pollution\\_development](https://github.com/JoshMerfeld/pollution_development)

<sup>†</sup>KDI School of Public Policy and Management and IZA; merfeld@kdis.ac.kr

# 1 Introduction

Coal is one of the most common energy sources throughout the world. In 2021, energy generation from coal reached an all-time high,<sup>1</sup> driven in part by rapid population growth and rising energy prices. India is no exception to these trends. From 1990 to 2010, India's population grew by more than 40 percent.<sup>2</sup> This population growth led to a large increase in demand for power. While natural gas has become a popular alternative in much of the world, the Indian government met much of the increased demand through the construction of coal power plants. During those same two decades, power generation from all coal units of at least 30 megawatts in the country increased by almost 140 percent, from 42.4 gigawatts to more than 100 gigawatts.<sup>3</sup> There are no doubt positive effects from this increase in power generation; for example, electricity may lead to economic and productivity growth (Dinkelman 2011; Kline and Moretti 2014; Rud 2012; Van de Walle et al. 2017), though electrification alone may not be a sufficient condition (K. Lee, Miguel, and Wolfram 2020).

These improvements do not come without downsides, however, especially when it comes to power derived from coal. Pollutants from coal contribute to climate change and the phase out of coal energy is therefore a major climate goal (IEA 2022). Emissions from the production and burning of coal are also directly harmful to health and the environment, including pollutants like sulfur dioxide, nitrogen oxides, and mercury, to name but a few.<sup>4</sup> While the use of coal had been declining, recent events – including Russia's invasion of Ukraine – have led to an increase in the demand for coal.<sup>5</sup>

I study the effects of emissions from coal power plants on agricultural productivity in India. Previous work has suggested that there may be harmful direct effects of air pollution (Heck et al. 1982; Marshall et al. 1997) and water pollution (Reddy and Behera 2006) on agricultural productivity, including work showing that gold mines – which release large amounts of pollutants into the surrounding environment – lead to lower agricultural productivity in Ghana (Aragón and Rud 2016). These effects can work through the well known effects on human health and productivity, but also directly through land productivity. Black carbon in soot, for example, can directly affect solar radiation reaching plants on the surface (Ramanathan and Carmichael 2008; Burney and Ramanathan 2014). Relative to previous work in economics, this paper tackles the broader case of coal plants, which are a leading source of power throughout the world. I show that air pollution from these plants can lead to lower agricultural productivity and even decreased economic growth.

---

<sup>1</sup>[www.iea.org/reports/coal-fired-electricity](http://www.iea.org/reports/coal-fired-electricity)

<sup>2</sup>[data.worldbank.org](http://data.worldbank.org)

<sup>3</sup>[globalenergymonitor.org/](http://globalenergymonitor.org/)

<sup>4</sup>[www.eia.gov/energyexplained/coal/coal-and-the-environment.php](http://www.eia.gov/energyexplained/coal/coal-and-the-environment.php)

<sup>5</sup>[www.npr.org/2022/08/15/1117560560/a-rising-demand-for-coal-amidst-war-in-ukraine](http://www.npr.org/2022/08/15/1117560560/a-rising-demand-for-coal-amidst-war-in-ukraine)

Identifying the causal effects of pollution is difficult. For starters, the construction of coal plants is likely to be endogenous, though the relationship between construction and economic conditions is not a priori clear. On the one hand, governments may locate new plants in fast-growing areas, while on the other hand, they may decide the exact location of plants based on the political power of local citizens (Kopas et al. 2020). In the case of the data used here, coal plants are constructed in areas with higher population and higher literacy (Table 1). In terms of agriculture, higher productivity predicts more coal plants in a given year as well as the construction of a coal plant over the next decade, indicating possible bias when estimating OLS regressions of agricultural productivity on pollution levels.

To identify the effects of air pollution on agricultural productivity, I instead combine the roll out of coal plants in India with plausibly exogenous changes in wind direction – in a similar spirit to Deryugina et al. (2019) – to measure the effects of pollution from coal plants on key development outcomes. I first plot the location of coal plants in India by year. Using a village-level shapefile, I identify all villages that are located within 30km of a coal plant at any time during the sample period, which comprises the 1990s to the 2010s, with differences depending on data availability.

I use the location of coal plants to identify high-pollution areas and calculate the direction from each coal plant to all villages within 30km. I include all areas, regardless of whether the coal plant has opened yet, to allow me to disaggregate the effects of pollution from coal plants with other kinds of pollution. Using daily wind data and these spatial variables, I create an identifier separately for each of the one-hundred-thousand-plus villages for whether a given village is located downwind from a coal plant on any given day. I then aggregate this exposure variable based on the desired analysis; for agricultural productivity, for example, I create an exposure variable measuring the total number of exposure days in the five-month agricultural season, focusing only on the monsoon (Kharif) season. Consistent with expectations, wind direction is a strong predictor of particulate matter at the monthly level. More days being located downwind from a coal power plant leads to higher average concentrations of particulate matter 2.5.

Using newly available estimates of India-wide agricultural (land) productivity, results show that the total number of days of being downwind from a coal power plant has a negative effect on overall agricultural productivity. Just a single additional day of being downwind from a coal power plant leads to a decrease in predicted yields of 0.6 percent. Combined with the within-village absolute deviation of wind exposure, results indicate that median changes in wind direction are responsible for fluctuations in agricultural productivity of around 1.5 percent.

Using wind direction as an instrument for particulate matter, an increase of PM 2.5 by one percent leads to

an even larger decrease in agricultural productivity, at around 0.73 percent. Given within-village changes across years, pollution could be responsible for swings of 7-to-14 percent in agricultural productivity across years. This IV strategy is similar to that in Deryugina et al. (2019), though they use general “high-pollution” wind directions while I concentrate only on wind directions from the location of a coal power plant. The IV estimates indicate that the effect of pollution is more negative than naïve OLS estimates indicate, which is consistent with the evidence on the construction of coal plants; higher levels of pollution are correlated with higher levels of agricultural productivity and economic growth, a relationship that the IV strategy adjusts downwards.

These effects are apparently larger in villages that experience higher median exposure to pollution and in villages with higher initial yields. I also show that rainfall shocks are compounding; higher exposure along with less rainfall increases the negative effect of either shock individually. As weather shocks become more common across the globe, the rise of pollution in India may be particularly problematic for farmers.

To explore the mechanisms driving this effect, I also use data on labor allocation from the National Sample Survey (NSS). While there do not appear to be large changes in labor allocation patterns when looking at the season as a whole, there is a clear pattern when we break down exposure to pollution by month: higher levels of pollution at the beginning of the monsoon season (June and July) lead to lower levels of farm labor allocation, but no changes in non-farm labor allocation. This matches up well with a similar exercise with agricultural yields: pollution in June is the only month in which there is a negative effect of pollution on agricultural yield. In tandem, these two results point to effects on labor as one mechanism driving the total effects, though I cannot rule out direct effects of pollution on land productivity.

The final set of results tries to disentangle the effects of pollution from the effects of coal pollution, specifically. Combining instrumental variables with a differences-in-differences strategy, I show that agricultural productivity has an elasticity of around -0.15 with respect to non-coal pollution. The opening of a coal plant, however, triples the elasticity, to more than -0.45, indicating that coal plant emissions are particularly harmful to agricultural productivity. This is consistent with coal emissions being quite different from other forms of pollution; for example, coal emissions contain much more mercury than other forms of pollution.

This paper contributes to several strands of literature. First, we already have evidence of effects of pollution on different forms of labor activity. For example, general levels of pollution decrease productivity of call center workers in China (Chang et al. 2019), lead people to perform worse on cognitive functions (Ebenstein, Lavy, and Roth 2016; Wen and Burke 2022), and also lead to lower levels of farm labor productivity in California, driven by changes in ozone concentration (Graff Zivin and Neidell 2012). Also in China, He,

Liu, and Salvo (2019) find small but significant negative effects of pollution on manufacturing productivity, while Chen, Oliva, and Zhang (2022) show pollution leads to labor out migration. Perhaps relevant to my labor allocation results, Hanna and Oliva (2015) show that decreases in pollution lead to increases in labor allocation, while I show the same relationship among the oldest cohort. The papers most related to this one are Aragón and Rud (2016), who show that gold mines in Ghana have substantial negative impacts on agricultural productivity in the surrounding areas, and Burney and Ramanathan (2014), who model the effects of climate change and pollutants on agricultural productivity in India. This paper differs in two important ways. First, I explore effects of pollution from coal plants, which are much more ubiquitous than gold mines across the world. Second, I use a different identification strategy from either paper, relying on wind blowing pollution from coal plants.

I also contribute to the more general literature on the effects of pollution on health. Pollution has substantial negative effects on health in both developing and developed countries (Arceo, Hanna, and Oliva 2016) and we have known of these negative effects for many years (Brunekreef and Holgate 2002; Kampa and Castanas 2008; Pope III and Dockery 2006). Currie et al. (2014) review the literature on the health effects of pollution during adolescence; we have robust evidence of pollution exposure leading to increases in infant mortality, including in developing countries (Heft-Neal et al. 2018, 2019). Since these effects are often driven by heart and lung disease and asthma (Kampa and Castanas 2008), this may explain the labor allocation effects I identify in this paper.

The rest of the paper is organized as follows. I first discuss data and methodology. I break this section up into subsection 2.1, which covers the multiple sources of data used in this paper; subsection 2.2, which discusses the empirical strategy and validates the wind direction variable using data on particulate matter. I present the main results in section 3 before concluding in section 4.

## 2 Data and methods

In this section, I describe the different data sources and the identification strategy. I also spend time validating the use of wind direction as a proxy for particulate matter as well as looking at the correlates of construction of coal power plants.

## 2.1 Data

The main goal of this paper is to examine whether exposure to emissions from coal plants affects agricultural productivity, labor allocation, and economic growth. I use several sources of data, which I describe here and in Table A1 in the appendix.

The first set of data lists the location of coal plant across the globe. This data comes from Global Energy Monitor<sup>6</sup> and lists all units generating at least 30 megawatts of electricity. The data on coal plants includes the year of opening (and, if applicable, the year of retirement), the GPS (latitude/longitude) location of each plant, and the amount of power produced by the units at each plant. For this paper, I do not use the information on the capacity of the plant.

Figure A1 in the appendix displays the location of coal plants in India for two specific years: 1990 and 2010. There is a clear increase in the prevalence of coal plants across the country over the two decades. Additionally, the overall capacity from coal plants increased from 42.4 gigawatts to 100.4 gigawatts, an increase of 136.6 percent in just 20 years. This increase is driven both by the roll out of new plants as well as the construction of new units at existing plants, the latter of which is not shown in Figure A1.

The second dataset includes agricultural productivity for both the monsoon and winter seasons, from 2002 to 2018. To match other data, I only use the data up to 2013. This data comes from Gangopadhyay et al. (2022) and estimates land productivity (i.e. yield, in tons per hectare) for the major crops in India, using satellite-derived data. While there is a wealth of literature on measurement error in land (Desiere and Jolliffe 2018; Abay et al. 2022), the authors use a static cropland mask for the productivity estimates. This means that any measurement error will be constant across time and subsumed by the village fixed effects. While the measure of output is also noisy, the use of satellite-derived data and a static cropland mask reduces concerns about measurement error being responsible for the results below. The authors define the monsoon season as June to October and the winter season as November to March. I keep these definitions when matching across data below. This data is also publicly available from Nature's data-sharing website.<sup>7</sup> Since the main regressions below use village fixed effects, any differences in cropping patterns across areas should not bias estimates on changes, as long as these patterns do not change markedly in response to pollution.

The agricultural productivity data is available as raster files with a resolution of 500m. To aggregate this data up to a useful administrative unit, I use village-level shapefiles provided by Asher et al. (2021) and publicly available on the SHRUG platform.<sup>8</sup> I aggregate the agricultural productivity data to the village

<sup>6</sup>[globalenergymonitor.org/projects/global-coal-plant-tracker/](http://globalenergymonitor.org/projects/global-coal-plant-tracker/)

<sup>7</sup>[springernature.figshare.com](https://springernature.figshare.com)

<sup>8</sup>[www.devdatalab.org/shrug\\_download/](http://www.devdatalab.org/shrug_download/)

level by extracting mean productivity for each feature in the shapefile. I do this separately for each season – monsoon and winter – and each year, from 2002 to 2013.

To measure exposure to pollution from coal plants, I first locate all village centroids located within 30km from a coal plant in a given year. I choose 30km due to previous research on the effects of (air) pollution (Aragón and Rud 2016; Bencko and Symon 1977; Li and Gibson 2014). I then calculate the direction from coal plants to all village centroids within that 30km radius. To define exposure, I then pull daily wind direction data from the National Center for Atmospheric Research.<sup>9</sup> For each day, I document whether the wind is blowing towards each village centroid.<sup>10</sup> I then temporally aggregate this daily data depending on the temporal definition of the corresponding outcome. For example, agricultural productivity is defined across five months (e.g. the monsoon season is from June to October) so I count the total days a given village is exposed to wind blowing from *any* coal plant within 30km during those five months.

The top panel in Figure A2 in the appendix is an example wind direction raster. The raster shows wind direction for the entirety of India on January 1st, 2010, as well as the location of coal power plants on that date. The prevailing winds on the date differ across the country and, though not shown in the figure, across days. This means that the overall exposure for a given area to coal plant emissions changes across time. I also pull data on particulate matter to help validate this proxy for exposure. This data comes from Hammer et al. (2020) and I also aggregate this to the village using the same method detailed above.

To help understand some of the mechanisms driving the relationship between exposure to emissions and agricultural productivity, I also include household survey data. Unfortunately, I am not aware of any publicly available survey data that covers multiple years and has identifiers below the district level. For example, the ARIS-REDS data<sup>11</sup> includes GPS identifiers for one year, but that only allows GPS matching for a total of two years: 1999 and 2007/2008. The DHS, on the other hand, only has state-level identifiers (except for the newest two waves) and is also not a labor survey. Instead, I use the National Sample Survey (NSS) from the Ministry of Statistics and Programme Implementation.

The NSS is a nationally representative survey that was conducted every couple of years until relatively recently. Many rounds of the NSS also included a module on employment that asked about labor allocation over the previous seven days. I use the 61st, 62nd, 64th, 66th, and 68th rounds of the NSS, all of which include the employment module. The publicly available data also includes the date of survey, which allows me to match pollution exposure to the exact day of the survey.

---

<sup>9</sup>[climatedataguide.ucar.edu/](http://climatedataguide.ucar.edu/)

<sup>10</sup>I define “towards” as within five degrees to help take into account that I am using village centroids, using the x and y components of wind speed to calculate angles.

<sup>11</sup>[www.ncaer.org/project/additional-rural-incomes-survey-aris-rural-economic-demographic-survey-reds](http://www.ncaer.org/project/additional-rural-incomes-survey-aris-rural-economic-demographic-survey-reds)

The biggest issue with using the NSS is that it includes only district-level identifiers. The main exposure variables, on the other hand, are defined at a lower level of aggregation, at the village. I choose to aggregate this village-level exposure to the district level as shown in the second panel of Figure A2 – the district is a randomly selected district in the data, Guna district in Madhya Pradesh. I use the area overlap of the villages with the district to define a weighted average of exposure for the district at a given time. I calculate the area of overlap for each village and the district as well as the overall area of the district and use these variables to define weights for individual villages. Any part of the district that does not overlap with one of the villages – the white area in Figure A2 – does not have any exposure and, as such, receives a value of zero for exposure.

## 2.2 Identification

I am primarily interested in the effects of pollution on agricultural productivity. We might consider a regression of the following form:

$$y_{it} = \alpha_i + \gamma_t + \beta PM_{it} + \delta rain_{it} + \psi temp_{it} + \varepsilon_{it}, \quad (1)$$

where  $y_{it}$  is outcome  $y$  for unit  $i$  at time  $t$ ,  $\alpha$  is a geographic fixed effects,  $\gamma_t$  is time fixed effects,  $PM_{it}$  is the concentration of particulate matter 2.5,  $rain_{it}$  is total rainfall during the season, and  $temp_{it}$  is average temperature during the season.

However, even with the fixed effects, this regression likely suffers from endogeneity; places with more pollution may be very different from places with less pollution. It is difficult to sign the direction of the bias a priori, because it depends on how the coal plants are placed. For example, if coal plants are placed in places in need of power – for example, due to rapid growth – then the estimates might be upwards biased. On the other hand, if official purposefully place coal plants in poorer areas for political reasons, then the estimates might be downwards biased.

So what are the characteristics of areas near the construction of new coal plants? Table 1 presents four separate regressions. The first two columns use 1991 census data – downloaded from SHRUG – and the second two columns use 2001 census data, all at the village level. The first column presents results from regressing a dummy for whether there is a coal plant within 30km in 1991 on several census variables. Villages with higher populations and higher literacy in 1991 are more likely to be near a coal plant at the time. Interestingly, however, they are less likely to have a paved road.

Table 1: Local characteristics and the construction of coal plants

	1991 census		2001 census	
	1991	2001	2001	2011
pop (log)	0.010*** (0.000)	0.001*** (0.000)	0.017*** (0.000)	0.004*** (0.000)
area (log)	-0.006*** (0.001)	0.001** (0.000)	-0.015*** (0.000)	-0.001*** (0.000)
literacy (prop)	0.024*** (0.003)	0.017*** (0.002)	0.032*** (0.002)	0.029*** (0.002)
ag productivity			0.026*** (0.001)	0.004*** (0.000)
sub-sample	all	no plant	all	no plant
observations	283,021	264,265	508,784	469,004

Note: Robust standard errors are in parentheses. The outcome in the first column is whether the village is within 30km of a coal plant in 1991. The second column is whether a village in 1991 will have a coal plant in 2001, conditional on not having one in 1991. The last two columns are similarly defined, except using 2001 and 2011 as the years. \* p<0.10 \*\* p<0.05 \*\*\* p<0.01

The second column keeps the 1991 census data but changes the outcome variable to a dummy for whether there *will be* a coal plant constructed within 30km before 2001. This column restricts the sample to only villages that do not already have a coal plant within 30km in 1991. Population is no longer a significant predictor while the coefficients on literacy and the presence of a paved road have flipped.

The last two columns are particularly relevant to the current results. They present similar results to those in the first two columns, but use the 2001 census data and a dummy for the presence of a coal plant in 2001 and 2011 in columns three and four, respectively. Importantly, I include agricultural productivity in the monsoon season of 2002 – the earliest year available – as an additional predictor. In 2001, agricultural productivity is a strong predictor for the presence of a coal plant. Moreover, it is also a significant predictor of the construction of a coal plant between 2001 and 2011 (column four). Given that the coefficient is positive, this indicates that coal plants are more likely to be built in areas with higher agricultural productivity. In other words, any selection in the overall location of coal plants would bias against finding negative effects of pollution on agricultural productivity. However, note that this is not the same as looking at wind direction, though comparing the OLS and IV results below leads to the same conclusion regarding selection.

Given the likely endogeneity, I instead turn to a plausibly exogenous source of variation in pollution: wind direction. While the location of a coal plant might be endogenous, the amount of pollution received by a specific village in a given year can vary randomly based on wind direction. By counting the number of days in a given year that a village is downwind from a polluted area and conditioning on village fixed effects and weather (rainfall and temperature), variation across years is plausibly exogenous.

Consider the reduced form effect of wind exposure on agricultural productivity:

$$y_{it} = \alpha_i + \gamma_t + \beta wind_{it} + \delta rain_{it} + \psi temp_{it} + \varepsilon_{it}, \quad (2)$$

where variables are defined identically to above but with wind as the primary variable of interest instead of pollution. Combining these into a two-stage procedure, we can instrument for pollution using wind direction; small changes in wind direction can lead to large changes in pollution exposure for villages located near point sources of pollution, allowing me to identify the causal effects of pollution by isolating the random component driven by wind direction. This is similar in spirit to Deryugina et al. (2019).

Identification relies on changes in wind direction across time for the same geographic units. In other words, conditional on the fixed effects and weather, I assume changes in wind direction are as good as random and uncorrelated with the outcome – e.g. agricultural productivity – except through exposure to pollution. One major threat to identification is that the building of a coal plant could lead to a spurious correlation between the exposure variable and the outcome. Specifically, since exposure is defined as zero prior to the building of a coal plant, the exposure variable itself might be endogenous even though wind direction is not. To help allay these concerns, I estimate additional regressions restricting the sample to only time periods after which a given location has a coal plant. I also present results using leads to show that future wind direction does not seem to predict current outcomes.

The inclusion of rainfall and temperature is meant to reduce concerns that wind direction may not affect agricultural productivity but, instead, be correlated with it through changes in weather. For example, prevailing winds from one direction may be associated with dry weather, while winds from another direction may be more likely to carry moisture and rain. It is not a priori clear how this would bias the estimates with respect to pollution and coal plants, however. The differences in weather patterns is very context specific and can differ even for villages located near one another. Nonetheless, adding rainfall and temperature should alleviate some of these concerns.

It is worth taking the time to validate the use of wind direction as a measure of exposure to pollution. Consider the data used for the regressions presented in Table A2 of the appendix. The outcome variable is particulate matter – specifically, PM<sub>2.5</sub>, which is particulate matter no larger than 2.5 micrometers in diameter – which comes from Hammer et al. (2020). PM<sub>2.5</sub> is one of the harmful byproducts from coal plants, along with sulfur dioxide, different types of nitrogen oxides, and mercury.<sup>12</sup> This data is available at the monthly level, so in Table A2, I aggregate wind exposure to the monthly level, as well.

---

<sup>12</sup>[www.epa.gov/airmarkets/power-plants-and-neighboring-communities](http://www.epa.gov/airmarkets/power-plants-and-neighboring-communities)

Both the particulate matter data and the wind data is at the village level, which means the regressions in Table A2 are at the village level. As such, the geographic fixed effects are village fixed effects and the temporal fixed effects are month fixed effects. Since the exposure variable is at the village level and we follow villages over time, the standard errors are also clustered at the village level. In columns two and four, I switch the month fixed effects for district-month fixed effects, which considers only villages in the same district in the same month.

In addition to serving as an example, Table A2 also helps validate the use of wind direction as a proxy for exposure to pollution. The first two columns include all available years of data while the last two columns restrict estimation only to the years for which the agricultural productivity regressions are estimated. Across all four columns, the story is the same: when the wind is blowing in the direction of a village from a coal plant within 30 kilometers, estimated PM<sub>2.5</sub> is substantially higher. The regressions indicate that one additional day of wind in the direction of the village increases mean PM<sub>2.5</sub> for the month by between 0.00 and 0.00  $\mu\text{g}/\text{m}^3$ .

To put these numbers into perspective, the World Health Organization's updated guidelines are that average annual exposure should not exceed 15  $\mu\text{g}/\text{m}^3$ . For the month, an increase at the midpoint of the estimated range is equal to an increase of 0.2 percent of the maximum recommended mean concentration from the WHO. This is just a single day of wind in the direction of a village. Since this includes all villages located within 30km of a coal plant, this is evidence that wind direction can have serious repercussions on the health of hundreds of millions of people in India.

### 2.2.1 Identifying effects of pollution from coal plants

I use the location of coal plants to identify locations of high pollution, but there can be other pollution sources located near coal plants, as well. I nonetheless want to identify the effects of pollution coming specifically from coal plants. I do this by combining the two-stage procedure above with a differences-in-differences framework. Specifically, I create a new variable identifying the exact year the coal plant opens in a given area and calculate the change in the effects of pollution on agricultural productivity.

Consider the following regression of interest:

$$y_{it} = \alpha_i + \gamma_t + \omega pm_{it} + \phi open_{it} + \beta pm_{it} \times open_{it} + \delta rain_{it} + \psi temp_{it} + \varepsilon_{it}, \quad (3)$$

where the interaction term  $pm_{it} \times open_{it}$  indicates the change in the effect of pollution following the opening

of the coal plant. Note that this is not the same as saying that pollution increased; instead, it is focusing on the change in the effect of a change in pollution of one unit on the outcome of interest, agricultural yield. I instrument for both  $pm_{it}$  and  $pm_{it} \times open_{it}$  using wind and the interaction between wind and the open indicator, or  $wind_{it} \times open_{it}$ .

## 3 Results

This section presents the main results of this paper. I present two main sets of results: one set looking at agricultural productivity and another looking at labor allocation.

### 3.1 Agricultural productivity

I first present results for agricultural productivity, with the reduced form. The outcome of all the regressions in this section is the log of agricultural land productivity, which is defined as tons per hectare. The first set of results are in Table 2. The unit of analysis is the village-year – from 2002 to 2013 – focusing only on the monsoon (kharif) season. The data used to construct the yield measure is calculated using five months of data, so the wind variable is defined for the same five months and is the sum of days in which the wind is pointed towards the village centroid from a coal plant.

The first column presents the most simple results, with only the wind exposure variable and the fixed effects. The second column adds a rainfall and average temperature variable, while the third column includes higher-order weather variables, as well. While the coefficient decreases slightly in magnitude as we move from the first to the second column and the second to the third column, the decrease is relatively small. There is a consistent, negative effect of wind exposure on agricultural productivity.

To put the size of the coefficient in context, it is worth digging a little deeper into the exposure variable. The mean within-village absolute deviation in wind exposure is approximately 2.42 days, meaning that the average change in agricultural productivity from year-to-year due to nothing but changes in wind patterns carrying particulate matter is around 1.5 percent relative to the mean. This implies that swings in agricultural productivity – due to wind direction – of several percent are probably quite common, since the absolute deviation includes deviations below and above the mean.

Of course, it is not wind, *per se*, that we are interested in. Instead, we are actually interested in better understanding the effects of exposure to pollution on agricultural productivity. Table 3 presents the same regressions as Table ??, with a key difference: instead of wind exposure, it uses logged average concentration

Table 2: Wind direction and agricultural productivity

	(1)	(2)	(3)
wind days	-0.0078*** (0.0007)	-0.0069*** (0.0007)	-0.0062*** (0.0006)
rain (z)		0.0854*** (0.0018)	-0.0424*** (0.0046)
mean temp (10s)		-0.0021*** (0.0001)	0.1321*** (0.0034)
rain square			-0.0010*** (0.0001)
temp square			-0.0002*** (0.0000)
rain times temp			0.0003*** (0.0000)
<b>fixed effects:</b>			
village	Yes	Yes	Yes
year	Yes	Yes	Yes
observations	1,266,922	1,266,922	1,266,922

Note: Standard errors are in parentheses and are clustered at the village level. The outcome in all regressions is estimated agricultural yield, defined as tons per hectare. \* p<0.10 \*\* p<0.05 \*\*\* p<0.01

Table 3: Particulate matter (PM 2.5) and agricultural productivity

	(1)	(2)	(3)
particulate matter (log PM 2.5)	-0.042*** (0.003)	-0.044*** (0.003)	-0.032*** (0.003)
weather	No	Yes	Yes
weather (expanded)	No	No	Yes
<b>fixed effects:</b>			
village	Yes	Yes	Yes
year	Yes	Yes	Yes
observations	1,266,922	1,266,922	1,266,922

Note: Standard errors are in parentheses and are clustered at the village level. The outcome in all regressions is estimated agricultural yield, defined as tons per hectare. \* p<0.10 \*\* p<0.05 \*\*\* p<0.01

of  $\text{PM}_{2.5}$  – from Hammer et al. (2020) – during the season as the key explanatory variable. There is again a consistent, negative relationship between particulate matter and agricultural productivity. Since both variables are logged, the coefficients are similar to elasticities, indicating that a one-percent increase in pollution leads to a 0.03-to-0.04 decrease in agricultural productivity.

Recall, however, the results from above: coal plants were placed in places with higher average agricultural productivity. This means that the simple regressions in 3 might actually be biased upwards: although pollution might negatively affect productivity, pollution is higher in more productive areas due to the endogenous placement of coal plants in the study areas. To recover the actual effect of pollution, we need an instrument. If, conditional on the fixed effects and weather variables, wind direction is exogenous and only affects agricultural productivity through pollution, then it is a valid instrument for pollution.

I present the IV results in Table 4, with the first-stage F-statistic in the bottom. The first-stage results make clear how strong of an instrument wind is; all three F-statistics are above 1,000, meeting even the higher bar for the first-stage F-test argued for by D. S. Lee et al. (2022). Also note that the coefficients are quite different from those in Table A2 because the unit of analysis in Table 4 is the season, while in Table A2 the unit of analysis is the month.

The second-stage results first indicate that  $\text{PM}_{2.5}$  is a very strong predictor of agricultural productivity. Using the results in column three, a one percent increase in particulate matter leads to a 0.73 percent decrease in agricultural productivity. To put this in further perspective, the mean particulate matter concentration is just short of 40 (39.68) in the monsoon sample, while the within-village absolute deviation in particulate matter is 3.98, meaning year-to-year swings of 7-to-14 percent in agricultural productivity due to pollution exposure are common.

The comparison of the results to those in the previous table also confirm the concerns around endogenous placement of coal plants. The coefficients in Table 4 are always more negative than those in Table 3, and sometimes substantially so. In other words, the simple regression of agricultural productivity on pollution results in coefficients on pollution that are upwards biased, since higher levels of pollution occur in more productive areas, on average. The IV strategy corrects this bias and leads to a robust, negative relationship between pollution and agricultural productivity.

## 3.2 Heterogeneity

Having established that exposure to pollution has a negative effect on agricultural productivity, I now move to heterogeneity in these effects. Table 5 presents three different analyses: heterogeneity by median exposure,

Table 4: Pollution and agricultural productivity, IV estimates

	(1)	(2)	(3)
particulate matter	-0.941*** (0.078)	-0.813*** (0.076)	-0.728*** (0.075)
weather	No	Yes	Yes
weather (expanded)	No	No	Yes
<b>fixed effects:</b>			
village	Yes	Yes	Yes
year	Yes	Yes	Yes
F (first stage)	1,235	1,262	1,263
observations	1,266,922	1,266,922	1,266,922

Note: Standard errors are in parentheses and are clustered at the village level. The outcome in all regressions is estimated agricultural yield, defined as tons per hectare. \* p<0.10 \*\* p<0.05 \*\*\* p<0.01

heterogeneity by initial yield, and possible multiplicative effects of shocks. All regressions use the IV strategy detailed above and I include the F statistic from the first stage regressions (of which there are two in the last column) at the bottom of the table. Table ?? in the appendix presents the first-stage results for all columns.

The first two columns split the sample based on median exposure. In other words, for each village, I calculate the median exposure (based on wind) and then use this single value for each village to split the sample into villages above and below median exposure. The coefficient in the first column is 2.3 times larger than the coefficient in the second column, indicating much larger effects of pollution in places where it is more common. One possible explanation is that exposure is cumulative, leading to larger and larger effects. I return to this in column five.

Columns three and four split the sample based on median yield at the very beginning of the sample (the first year available, 2002-2003). Column three presents results for villages above the median while column four presents results for villages below the median. The largest effects are apparently concentrated in the upper half of the distribution, meaning the most productive areas are also most susceptible to the effects of pollution.

Given ongoing changes in the climate, a key question is how welfare will change as shocks become more and more common. The last column looks at possible multiplicative effects of shocks by interacting pollution with rainfall and seeing whether the effects compound one another. Since there are now two endogenous regressors, I instrument for both using wind exposure and an interaction between wind exposure and rainfall.

The coefficient on the level term for pollution is negative. Since rainfall is defined as a z score, this represents the effect of pollution only when rainfall is at its mean value for a given season. The interaction term, on the other hand, is positive. Since higher values of rainfall are generally good, this positive coefficient represents

Table 5: Heterogeneity in the effects of pollution on productivity

	wind		yield		
	>p(50)	<=p(50)	>p(50)	<=p(50)	
PM 2.5	-0.767*** (0.104)	-0.328*** (0.095)	-1.130*** (0.060)	-0.690*** (0.146)	-0.409** (0.160)
PM times rain					0.061*** (0.014)
weather	Yes	Yes	Yes	Yes	Yes
<b>fixed effects:</b>					
village	Yes	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes	Yes
F (1st stage, PM)	684	491	688	484	865
F (1st stage, PM times rain)					1,437
observations	617,804	649,118	634,342	632,580	1,266,922

Note: Standard errors are in parentheses and are clustered at the village level. The outcome in all regressions is estimated agricultural yield, defined as tons per hectare. The first two columns split the sample based on the median exposure (by village). The second two columns split the sample based on initial – the first year available – agricultural productivity. \* p<0.10 \*\* p<0.05 \*\*\* p<0.01

a compounding effect of pollution and rainfall; when rainfall is low (bad), the effects of particulate matter on agricultural productivity is larger (more negative). Another way to think about this, however, is that higher values of rainfall dampen the effects of pollution shocks. This could have very profound effects on welfare, since this essentially increases the variance of agricultural productivity across years (which is already increasing due to climate change).

In the appendix, I present in Table A3 as a robustness check similar regressions but also including wind/pollution leads. The negative effect is only seen with current pollution, not future pollution.

### 3.3 Labor allocation

The previous set of results show a consistent negative effect of wind-driven pollution on agricultural productivity. A key question is the mechanism through which this happens; is this driven by changes in labor allocation/productivity or by changes in land productivity? While I cannot definitely prove the mechanism, I can shed some light on possibilities by looking at labor allocation decisions. Since these shocks generally happen after initial planting decisions, it seems unlikely that changes in land *allocation* could drive the results.

Table 6 presents results for labor allocation over the last seven days using the National Sample Survey and the methodology described above. Since the labor recall is for seven days, labor allocation ranges from zero to seven, as does the wind variable. Recall that the data is at the district level, not the village level. However, it is also more temporally disaggregated, with rolling interviews throughout the year-long survey. Figure A3

Table 6: Wind direction and labor allocation by rural status

	all	all	self	wage	farm	non-farm
<b>panel a: all</b>						
wind	-0.011 (0.031)	-0.009 (0.034)	0.030 (0.039)	-0.049 (0.034)	-0.009 (0.034)	-0.010 (0.033)
DV mean	3.303	3.303	1.862	1.442	0.789	2.514
observations	523,075	522,957	522,957	522,957	522,957	522,957
<b>panel b: rural</b>						
wind	-0.006 (0.044)	-0.034 (0.040)	0.068 (0.043)	-0.102*** (0.035)	-0.024 (0.043)	-0.010 (0.036)
DV mean	3.494	3.494	2.184	1.311	1.271	2.223
observations	307,381	307,325	307,325	307,325	307,325	307,325
controls	No	Yes	Yes	Yes	Yes	Yes
<b>fixed effects:</b>						
district	Yes	Yes	Yes	Yes	Yes	Yes
state-month	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors are in parentheses and are clustered at the district level. Control variables include female, age, age squared, and (years of) education. The outcome in each regression is days worked in the previous seven days. \* p<0.10 \*\* p<0.05 \*\*\* p<0.01

in the appendix shows this temporal distribution across the five waves of the survey. Since pollution is not disaggregated at a daily level, I instead stick to the reduced form results with wind.

The top panel presents results for all respondents during the five-month monsoon season while the bottom panel presents results for respondents in rural areas. Interestingly, we do not see any signs of large, negative effects of wind exposure on labor allocation across any sectors, with the possible exception of wage employment in rural areas. However, the coefficient for self-employment in rural areas is of the same magnitude and *positive*, but imprecisely estimated, making it hard to come to any definitive conclusions, at least in the aggregate.

### 3.4 Month-specific exposure

While overall results for labor allocation are inconclusive, perhaps we can learn more by looking at temporally disaggregated results, focusing on May through December. Although the yield data was created by looking at five specific months of the year – June through October – this is partially because this is the time of year when remote-sensing data is most able to pick up overall changes in crop cover. Preparation, however, starts before that, with operations such as ploughing often taking place prior to June.

Since the NSS data has survey dates and the both the pollution and wind data can be aggregated to the month, we can look at how exposure during different months of the agricultural season (differentially) affect

Table 7: Wind and labor allocation, by month

	farm	non-farm
May	-0.006 (0.068)	-0.039 (0.070)
June	-0.136* (0.080)	0.114 (0.099)
July	-0.285** (0.142)	0.176 (0.134)
Aug.	0.048 (0.086)	-0.010 (0.069)
Sept.	0.076 (0.085)	-0.082 (0.091)
Oct.	0.013 (0.076)	-0.083 (0.059)
Nov.	-0.063 (0.062)	-0.013 (0.075)
controls	Yes	Yes
<b>fixed effects:</b>		
district	Yes	Yes
year-month	Yes	Yes
observations	307,325	307,325

Note: Standard errors are in parentheses and are clustered at the village level. The outcome in column one is . \* p<0.10 \*\* p<0.05 \*\*\* p<0.01

labor allocation and overall agricultural productivity. I begin with labor allocation in Table 7. As before, the exposure measure is just the wind variable (over the last seven days), since I can directly match daily wind direction with the timing of the survey, but I cannot do that for pollution, which is monthly. The first column presents results for farm labor and the second column presents results for non-farm labor.

Higher exposure leads to significantly less farm labor in June and July, but there are no discernible effects in other months of the year. There are likewise no discernible effects on non-farm labor, but some of the coefficients are rather imprecisely estimate. To put the effects for June and July in context and taking the coefficients at face value, an additional day of wind in the last seven days decreases farm labor in June by around 11.2 percent of the mean and in July by around 17.1 percent of the mean.

I perform a similar exercise with yield, the results of which are in Table 8. I include only the five months of the year that are used in the calculation of the yield estimates, or June through October. The results are again from two-stage least squares, with the F-statistics for each month presented at the bottom; monthly wind remains a strong predictor of pollution. Column one presents results without controls for weather while column two presents results with controls for weather; the differences in coefficients make clear how important the inclusion of weather is here.

Table 8: Pollution and productivity, by month

	(1)	(2)
June	-0.325** (0.135)	-0.736*** (0.062)
July	-0.778*** (0.151)	0.051 (0.095)
Aug.	-1.201*** (0.455)	0.087 (0.070)
Sept.	1.270*** (0.427)	-0.140 (0.180)
Oct.	-3.013*** (0.963)	0.334 (0.291)
weather	No	Yes
<b>fixed effects:</b>		
village	Yes	Yes
year	Yes	Yes
F (June)	1,063	972
F (July)	2,090	2,200
F (Aug.)	1,646	2,340
F (Sept.)	2,025	2,024
F (Oct.)	473	402
observations	1,266,922	1,266,922

Note: Standard errors are in parentheses and are clustered at the village level. The outcome in all regressions is estimated agricultural yield, defined as tons per hectare. \* p<0.10 \*\* p<0.05 \*\*\* p<0.01

Focusing on column two, there is only one month where we see a clear effect of pollution on yield: June. Given that we also see negative effects of pollution on farm labor allocation in June, this is suggestive – but not conclusive – evidence that pollution works through labor. Note that this does not rule out pollution also having direct effects on agricultural productivity, however.

A final caveat about these month-specific results is that they do not take into account possible interactions between different months. For example, yield might respond strongly to average pollution in June and July, which would not be captured in Table 8.

### 3.5 Coal plants or overall pollution?

Finally, while I identify high-pollution areas using the location of coal plants, an important question is how much of the overall effect is driven by coal plants and how much is driven by other polluting industries located in areas near coal plants. I try to answer this question in Table 9, where I implement an instrumental variables differences-in-differences, essentially instrumenting for pollution prior to the opening of a coal plant with wind and instrumenting for the interaction term (pollution times plant opening) with wind times plant

Table 9: Agricultural productivity and pollution source

	(1)	(2)	(3)
PM 2.5	-0.299*** (0.066)	-0.107* (0.063)	-0.147** (0.063)
PM 2.5 times coal	-0.328*** (0.026)	-0.321*** (0.024)	-0.304*** (0.023)
weather	No	Yes	Yes
weather (expanded)	No	No	Yes
<b>fixed effects:</b>			
year	Yes	Yes	Yes
village	Yes	Yes	Yes
F (pm)	766	776	788
F (pm times open)	199	222	232
observations	1,266,922	1,266,922	1,266,922

Note: Standard errors are in parentheses and are clustered at the village level. The outcome in all regressions is estimated agricultural yield, defined as tons per hectare. \* p<0.10 \*\* p<0.05 \*\*\* p<0.01

opening.

Columns one through three of Table 9 present specifications with progressively increasing weather controls. Column three presents the most extensive set of controls and I focus on that column here. First, there are now two first stages but the F-test indicates that the instruments are strong for both, including the interaction term ( $F = 232$ ). Second, the first row shows the effect of pollution on agricultural productivity prior to the opening of a coal plant; the coefficient is negative but relatively smaller. Third, the interaction term is strongly negative and more than twice as large in magnitude as the level term, indicating a very clear effect specific to coal plants. This could be due to differences in emission patterns for different kinds of plants and factors; coal plants, for example, are responsible for much more mercury than other kinds of polluting industries.<sup>13</sup>

## 4 Conclusion

Pollution is a growing problem throughout the developing world. India, for example, has some of the worst air quality on earth. The rapid growth of coal power plants has also led to increases in emissions from the burning of coal. In this paper, I show some of the negative effects associated with pollution, in general, and coal pollution, in particular. Using exogenous variation in pollution driven by changes in wind direction, results here show that pollution has a large, negative effect on agricultural productivity in India.

<sup>13</sup>[https://www.epa.gov/sites/default/files/2016-12/documents/nei2014v1\\_tsdpdf](https://www.epa.gov/sites/default/files/2016-12/documents/nei2014v1_tsdpdf)

These negative effects are at least partly due to changes in labor allocation decisions; the largest negative effects on agricultural productivity and on labor allocation happen in the same month of the season: June. Nonetheless, I am unable to rule out direct effects on land productivity with the data available.

Using instrumental variables in a differences-in-differences set up, I also show that the effects of pollution from coal, in particular, are particularly harmful to agricultural productivity. Changes of an equivalent increase in pollution lead to three times larger negative effects on agricultural productivity following the opening of a coal plant, compared to just before.

These results, of course, do not address possible positive effects of the building of additional power plants, which can lead to increased electricity access. As such, the main conclusion is that the location of new power generation facilities may be particularly important for multiple aspects of economic activity, not just health and mortality.

## References

- Abay, Kibrom A, Christopher B Barrett, Talip Kilic, Heather Moylan, John Ilukor, and Wilbert Drazi Vundru. 2022. "Nonclassical Measurement Error and Farmers? Response to Information Reveal Behavioral Anomalies."
- Aragón, Fernando M, and Juan Pablo Rud. 2016. "Polluting industries and agricultural productivity: Evidence from mining in Ghana." *The Economic Journal* 126 (597): 1980–2011.
- Arceo, Eva, Rema Hanna, and Paulina Oliva. 2016. "Does the effect of pollution on infant mortality differ between developing and developed countries? Evidence from Mexico City." *The Economic Journal* 126 (591): 257–80.
- Asher, Sam, Tobias Lunt, Ryu Matsuura, and Paul Novosad. 2021. "Development Research at High Geographic Resolution: An Analysis of Night Lights, Firms, and Poverty in India using the SHRUG Open Data Platform." *The World Bank Economic Review*.
- Bencko, Vladimir, and Karel Symon. 1977. "Health Aspects of Burning Coal with a High Arsenic Content." *Environmental Research* 13 (3): 378–85.
- Brunekreef, Bert, and Stephen T Holgate. 2002. "Air pollution and health." *The Lancet* 360 (9341): 1233–42.
- Burney, Jennifer, and V Ramanathan. 2014. "Recent climate and air pollution impacts on Indian agriculture." *Proceedings of the National Academy of Sciences* 111 (46): 16319–24.
- Chang, Tom Y, Joshua Graff Zivin, Tal Gross, and Matthew Neidell. 2019. "The effect of pollution on worker productivity: evidence from call center workers in China." *American Economic Journal: Applied Economics* 11 (1): 151–72.
- Chen, Shuai, Paulina Oliva, and Peng Zhang. 2022. "The effect of air pollution on migration: evidence from China." *Journal of Development Economics* 156: 102833.
- Currie, Janet, Joshua Graff Zivin, Jamie Mullins, and Matthew Neidell. 2014. "What do we know about short-and long-term effects of early-life exposure to pollution?" *Annual Review Resource Economics* 6 (1): 217–47.
- Deryugina, Tatyana, Garth Heutel, Nolan H Miller, David Molitor, and Julian Reif. 2019. "The mortality and medical costs of air pollution: Evidence from changes in wind direction." *American Economic Review* 109 (12): 4178–4219.
- Desiere, Sam, and Dean Jolliffe. 2018. "Land productivity and plot size: Is measurement error driving the inverse relationship?" *Journal of Development Economics* 130: 84–98.
- Dinkelmann, Taryn. 2011. "The effects of rural electrification on employment: New evidence from South Africa." *American Economic Review* 101 (7): 3078–108.
- Ebenstein, Avraham, Victor Lavy, and Sefi Roth. 2016. "The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution." *American Economic Journal: Applied Economics* 8 (4): 36–65.
- Gangopadhyay, Prasun K, Paresh B Shirasath, Vinay K Dadhwal, and Pramod K Aggarwal. 2022. "A new two-decade (2001–2019) high-resolution agricultural primary productivity dataset for India." *Scientific Data* 9 (1): 1–12.
- Graff Zivin, Joshua, and Matthew Neidell. 2012. "The impact of pollution on worker productivity." *American Economic Review* 102 (7): 3652–73.
- Hammer, Melanie S, Aaron van Donkelaar, Chi Li, Alexei Lyapustin, Andrew M Sayer, N Christina Hsu, Robert C Levy, et al. 2020. "Global Estimates and Long-Term Trends of Fine Particulate Matter Concentrations (1998–2018)." *Environmental Science & Technology* 54 (13): 7879–90.
- Hanna, Rema, and Paulina Oliva. 2015. "The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City." *Journal of Public Economics* 122: 68–79.
- He, Jiaxiu, Haoming Liu, and Alberto Salvo. 2019. "Severe air pollution and labor productivity: Evidence from industrial towns in China." *American Economic Journal: Applied Economics* 11 (1): 173–201.
- Heck, Walter W, OC Taylor, Richard Adams, Gail Bingham, Joseph Miller, Eric Preston, and Leonard Weinstein. 1982. "Assessment of crop loss from ozone." *Journal of the Air Pollution Control Association* 32 (4): 353–61.
- Heft-Neal, Sam, Jennifer Burney, Eran Bendavid, and Marshall Burke. 2018. "Robust relationship between air quality and infant mortality in Africa." *Nature* 559 (7713): 254–58.
- Heft-Neal, Sam, Jennifer Burney, Eran Bendavid, Kara Voss, and Marshall Burke. 2019. "Air pollution and

- infant mortality: Evidence from Saharan dust.” National Bureau of Economic Research.
- IEA. 2022. “Coal-fired electricity.” International Energy Association.
- Kampa, Marilena, and Elias Castanas. 2008. “Human health effects of air pollution.” *Environmental Pollution* 151 (2): 362–67.
- Kline, Patrick, and Enrico Moretti. 2014. “Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority.” *The Quarterly Journal of Economics* 129 (1): 275–331.
- Kopas, Jacob, Erin York, Xiaomeng Jin, SP Harish, Ryan Kennedy, Shiran Victoria Shen, and Johannes Urpelainen. 2020. “Environmental justice in India: incidence of air pollution from coal-fired power plants.” *Ecological Economics* 176: 106711.
- Lee, David S, Justin McCrary, Marcelo J Moreira, and Jack Porter. 2022. “Valid t-ratio Inference for IV.” *American Economic Review* 112 (10): 3260–90.
- Lee, Kenneth, Edward Miguel, and Catherine Wolfram. 2020. “Does Household Electrification Supercharge Economic Development?” *Journal of Economic Perspectives* 34 (1): 122–44.
- Li, Ya-Ru, and Jacqueline MacDonald Gibson. 2014. “Health and air quality benefits of policies to reduce coal-fired power plant emissions: a case study in North Carolina.” *Environmental Science & Technology* 48 (17): 10019–27.
- Marshall, Fiona, Mike Ashmore, Fiona Hinchcliffe, et al. 1997. *A hidden threat to food production: Air pollution and agriculture in the developing world*. International Institute for Environment and Development.
- Pope III, C Arden, and Douglas W Dockery. 2006. “Health effects of fine particulate air pollution: lines that connect.” *Journal of the Air & Waste Management Association* 56 (6): 709–42.
- Ramanathan, Veerabhadran, and Gregory Carmichael. 2008. “Global and regional climate changes due to black carbon.” *Nature Geoscience* 1 (4): 221–27.
- Reddy, V Ratna, and Bhagirath Behera. 2006. “Impact of water pollution on rural communities: An economic analysis.” *Ecological Economics* 58 (3): 520–37.
- Rud, Juan Pablo. 2012. “Electricity provision and industrial development: Evidence from India.” *Journal of Development Economics* 97 (2): 352–67.
- Van de Walle, Dominique, Martin Ravallion, Vibhuti Mendiratta, and Gayatri Koolwal. 2017. “Long-term gains from electrification in rural India.” *The World Bank Economic Review* 31 (2): 385–411.
- Wen, Jeff, and Marshall Burke. 2022. “Lower test scores from wildfire smoke exposure.” *Nature Sustainability* 5 (11): 947–55.

## Appendix A

Table A1: Data sources

	source	geographic coverage	temporal coverage
shapefile	Asher et al. (2021)	India	
coal plants	Global Energy Monitor	global	yearly
wind	NCAR	global	daily
pollution	Hammer et al. (2020)	global	monthly
agriculture	Gangopadhyay et al. (2022)	India	two seasons/year
nightlights	Asher et al. (2021)	India	yearly
weather	TerraClimate	global	monthly
labor	National Sample Survey (NSS)	India	daily interviews

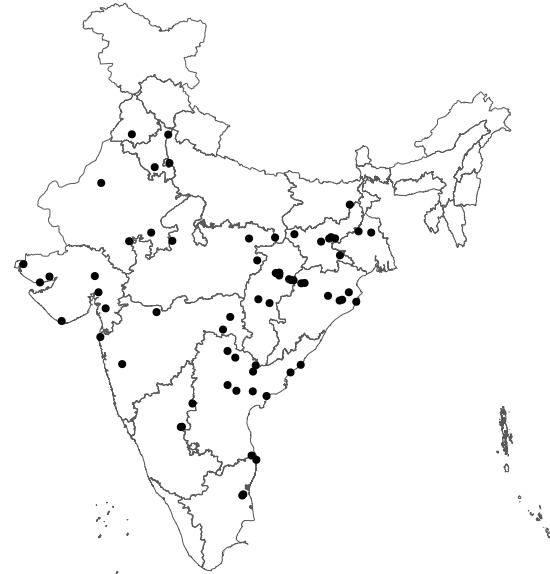
Global Energy Monitor: [globalenergymonitor.org/projects/global-coal-plant-tracker](http://globalenergymonitor.org/projects/global-coal-plant-tracker). NCAR: [climatedataguide.ucar.edu/](http://climatedataguide.ucar.edu/). TerraClimate: [www.climatologylab.org/terraclimate.html](http://www.climatologylab.org/terraclimate.html)

Figure A1: Coal plants in India from 1990 to 2010

Figure A: 1990



Figure B: 2010



Note: The top figure shows the location of coal plants in 1990. The bottom figure shows the location of coal plants in 2010.

Figure A2: Wind direction and aggregation examples (2010-01-01)

Figure A: Wind direction

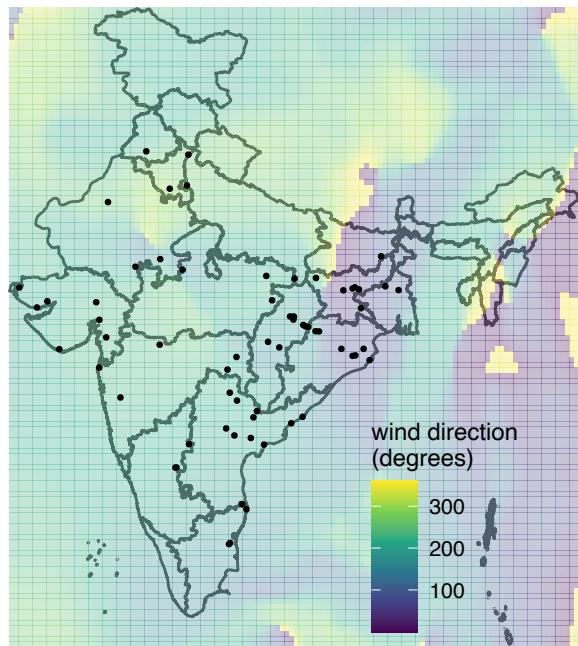
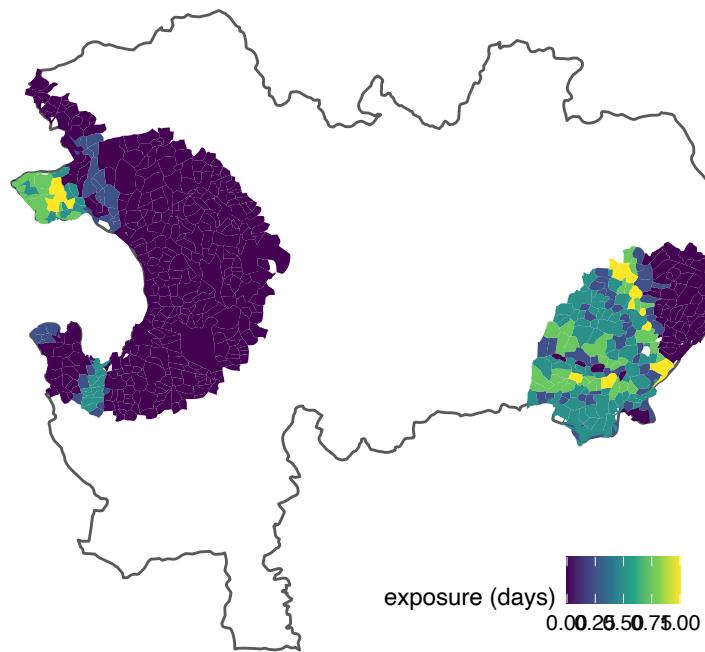


Figure B: Aggregation to district



Note: The top figure shows the average wind direction on January 1st, 2010, with zero degrees indicating north. The points are the location of coal plants on that date. The bottom figure shows the distribution of pollution exposure in a specific district – Guna district in Madhya Pradesh – on the same date.

Table A2: Wind direction and particulate matter

	1998-2015		2002-2013	
	(1)	(2)	(3)	(4)
wind (10s)	0.0023*** (0.0005)	0.0016*** (0.0002)	0.0057*** (0.0005)	0.0015*** (0.0002)
<b>fixed effects:</b>				
village	Yes	Yes	Yes	Yes
month	Yes	No	Yes	No
district-month	No	Yes	No	Yes
observations	24,376,458	24,376,458	16,250,973	16,250,973

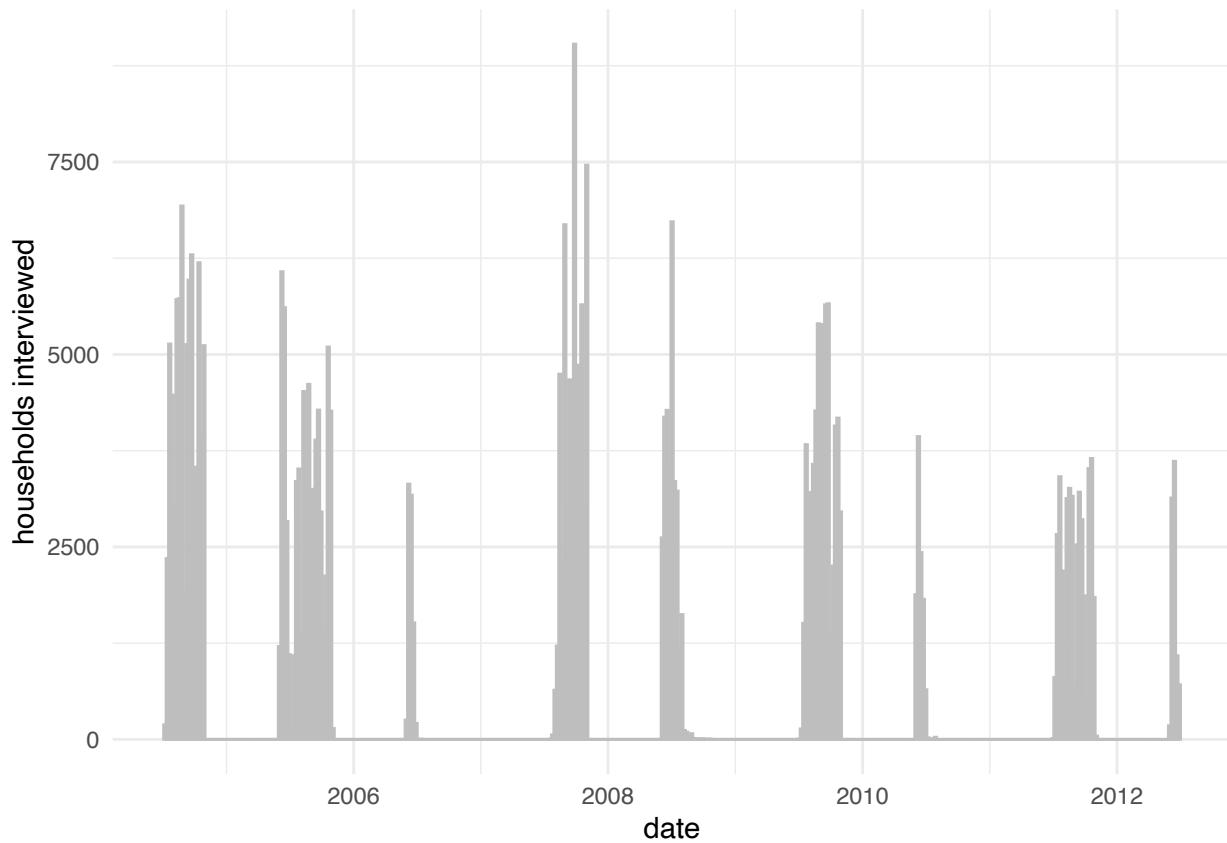
Note: Standard errors are in parentheses and are clustered at the village level. \* p<0.10  
 \*\* p<0.05 \*\*\* p<0.01

Table A3: Agricultural productivity and pollution leads

	(1)	(2)	(3)
PM	-1.1524*** (0.0755)	-1.0915*** (0.0736)	-1.0387*** (0.0731)
PM (lead)	0.0681 (0.0716)	0.2890*** (0.0755)	0.3091*** (0.0727)
weather	No	Yes	Yes
weather (expanded)	No	No	Yes
<b>fixed effects:</b>			
village	Yes	Yes	Yes
year	Yes	Yes	Yes
F	974	993	1,014
F (lead)	612	585	571
observations	1,161,265	1,161,265	1,161,265

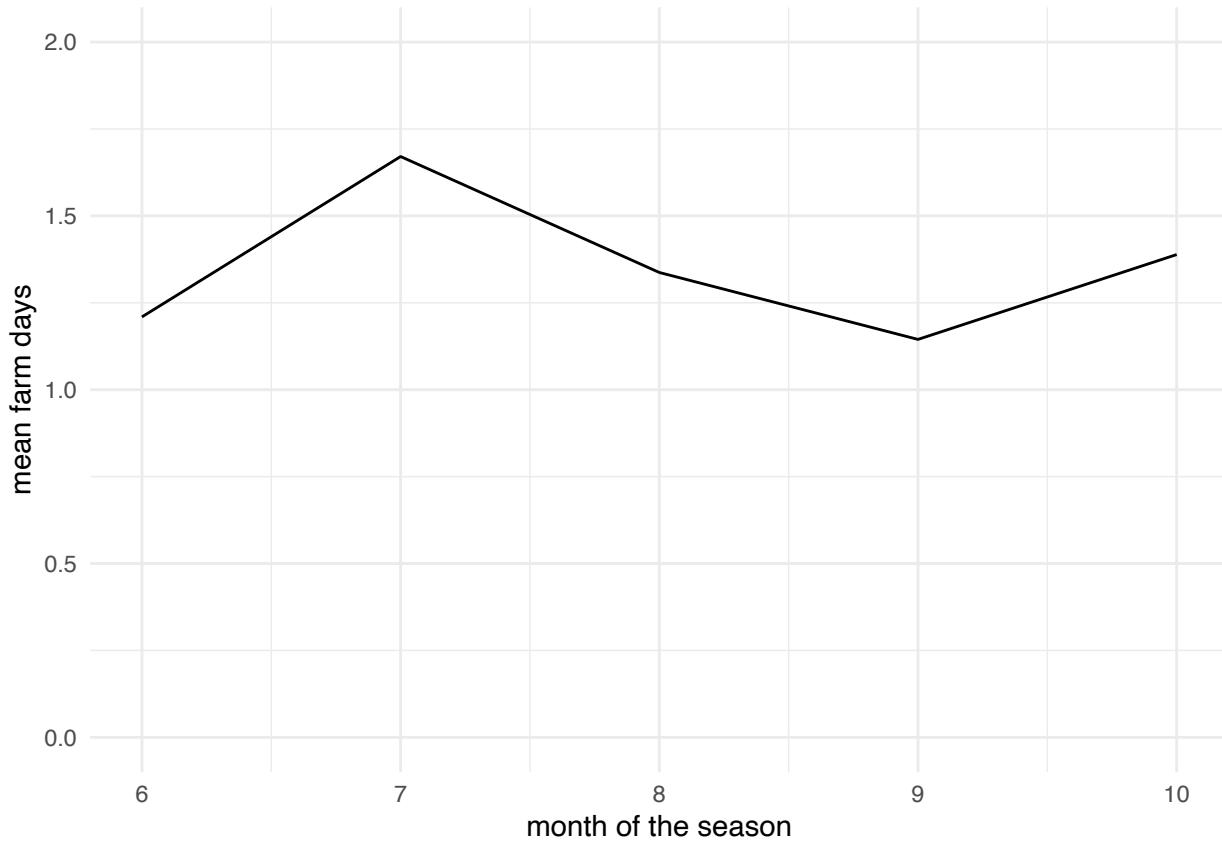
Note: Standard errors are in parentheses and are clustered at the village level. The outcome in all regressions estimated agricultural yield (tons per hectare). \* p<0.10 \*\* p<0.05 \*\*\* p<0.01

Figure A3: Timing of household surveys for the National Sample Survey



Note: The histogram shows the distribution of survey dates for the four waves of the National Sample Survey (NSS) for the five waves used in this paper.

Figure A4: Average farm days by season, rural only



Note: The line graph shows the mean number of farm days worked per week based on the month of the relevant agricultural season. The monsoon season is defined as June (month 1) to October (month 5) while the rainy season is defined as November (month 1) to March (month 5) of the following year. The data is from the NSS.