

# Air pollution and agricultural productivity in a developing country\*

Joshua D. Merfeld<sup>†</sup>

2023-07-03

## Abstract

I document externalities in the agricultural sector of India due to air pollution. Using variation in pollution induced by changes in wind, I show that higher levels of pollution lead to decreased agricultural productivity, with large changes in productivity being common. Combining wind direction with the rollout of coal plants, results indicate pollution from coal plants has a larger effect on agricultural productivity than other types of pollution. Given that the agricultural sector is a refuge for the poor in many developing countries, these results suggest that the effects of pollution may hit the poorest particularly hard.

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

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

---

\*I thank Ariel Ortiz-Bobea, Marc Bellemare, and David Sungho Park for comments and suggestions that greatly improved the quality of the paper. The usual disclaimer applies.

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

# 1 Introduction

While reliance on coal is declining in much of the world, it continues to be an important source of energy in many countries. In 2021, energy generation from coal reached an all-time high,<sup>1</sup> driven in part by rapid population growth in developing countries, including India. From 1990 to 2010, India's population grew by more than 40 percent, leading to a large increase in demand for power.<sup>2</sup> 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 more than doubled, from 42.4 gigawatts to more than 100 gigawatts.<sup>3</sup>

While there are clear 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, 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>

In addition to pollution from coal, pollution more generally is also increasing in much of the world. Carbon dioxide equivalent per capita has almost double over the last 50 years in India, more than doubled in Indonesia, and almost tripled in China.<sup>6</sup> Of course, much of this increase is due to the development process itself; as countries develop, their economies transform, with more people moving out of agriculture and into manufacturing and services (Herrendorf, Rogerson, and Valentinyi 2014). Despite this development, agriculture remains a key sector in many developing countries, employing a large proportion of the poor (Dercon 2009). This also means agriculture is a key sector when it comes to poverty reduction (Christiaensen, Demery, and Kuhl 2011). Given the large effects of pollution on health and previously documented impacts on agriculture, this raises an important question: do rising pollution levels in developing countries threaten agricultural productivity?

In this paper, I study the effects of pollution on agricultural productivity in India. Previous work has

<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)

<sup>6</sup><https://edgar.jrc.ec.europa.eu>

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 pollution more broadly, before turning to a specific source of pollution: coal plants. I show that air pollution leads to lower agricultural productivity in India and that pollution from coal plants is particularly harmful to agriculture.

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 I use here, coal plants are constructed in areas with higher population and higher literacy. 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 identify areas of high pollution emission using the location of coal plants. While coal plants may add emissions, I show that they are located in areas with relatively higher levels of pollution, in general. Using these locations as point sources of pollution, I use plausibly exogenous changes in wind direction – in a similar spirit to Deryugina et al. (2019) – to measure the effects of pollution on key agricultural 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 2000s to the 2010s, with differences across outcomes depending on data availability. I then 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 later in the paper. Using daily wind data and these spatial variables, I create an identifier separately for every village 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. More days being

located downwind from a high-pollution location leads to substantially higher average concentrations of particulate matter 2.5 (PM 2.5).<sup>7</sup>

Using wind direction as an instrument for particulate matter, an increase of PM 2.5 by one percent leads to a large decrease in agricultural productivity, at around 0.44 percent in the most conservative specification. The IV estimates indicate that the effect of pollution is more negative than naïve OLS estimates, which is consistent with the evidence on the construction of coal plants; higher levels of pollution are correlated with higher levels of agricultural productivity, a relationship that the IV strategy adjusts downwards. Robustness checks confirm these main results. Randomization inference shows that the estimates are robust to alternative measures of uncertainty, while leads of pollution do not affect current agricultural productivity.

Heterogeneity analysis shows that the negative effects of pollution are larger in villages that experience higher median exposure to pollution and in villages with higher initial yields. Finally, given how many poor farmers rely on rainfall and how weather shocks increase across the globe, I also look at possible multiplicative effects of shocks by interacting the pollution variables with a rainfall variable. Higher exposure along with less rainfall increases the negative effect of either shock individually, meaning these shocks are compounding. Given concerns regarding agricultural productivity and climate change, the rise of pollution in India may be particularly problematic for poor farmers.

The final set of analyses tries to disentangle the effects of pollution from the effects of coal pollution, specifically, by using wind direction and the rollout of coal plants across the country. Combining instrumental variables with a differences-in-differences strategy, I show that agricultural productivity has an elasticity of around -0.22 with respect to non-coal pollution. The opening of a coal plant, however, doubles the elasticity, to more than -0.44, indicating that coal plant emissions are particularly harmful to agricultural productivity, at least in India. 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.

An important caveat to these results is that I cannot disentangle exactly which pollutants drive results. While I use PM 2.5 as a measure of overall exposure to pollution, effects could be driven by any pollutants carried by the wind. This leaves open an important avenue of future research regarding the effects of different types of pollutants on agricultural productivity.

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,

---

<sup>7</sup>PM 2.5 is defined as particulate matter with a diameter of less than 2.5 micrometers.

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 in at least a subset of months of the agricultural season. Several papers are closely related to this one. Aragón and Rud (2016) show that gold mines in Ghana have substantial negative impacts on agricultural productivity in the surrounding areas. Burney and Ramanathan (2014) model the effects of climate change and pollutants on agricultural productivity in India. Finally, Sanders and Barreca (2022) actually find that coal emissions can be good for agriculture in the United States, due to the presence of sulfur, which can increase agricultural productivity. This paper differs from the previous literature in three important ways. First, I explore effects of pollution more generally as well as from coal plants, which are much more ubiquitous than gold mines across the world. Second, I use a different identification strategy from most other papers, relying on wind blowing pollution from coal plants. Finally, agriculture in India and the United States can be quite different, with the latter being much more mechanized. This means effects of pollution could differ substantially.

The rest of the paper is organized as follows. I first discuss data and then methodology, in two separate sections. The methodology section discusses the empirical strategy and validates the wind direction variable using data on particulate matter. I present the main results in section 4 before concluding in section 5.

## 2 Data

The main goal of this paper is to examine whether exposure to pollution and emissions from coal plants, in particular, affect agricultural productivity. I use several sources of data, which I describe here and list in Table A1 of the appendix.

The first set of data lists the location of coal plant across the globe. This data comes from Global Energy Monitor<sup>8</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.

---

<sup>8</sup>[globalenergymonitor.org/projects/global-coal-plant-tracker/](http://globalenergymonitor.org/projects/global-coal-plant-tracker/)

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, especially since common sources of measurement error are unlikely to be correlated with wind direction. The authors define the monsoon season as June to October and the winter season as November to March. I focus on only the monsoon season and keep the authors' definition when matching across data below. This data is also publicly available from Nature's data-sharing website.<sup>9</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>10</sup> I aggregate the agricultural productivity data to the village level by extracting mean productivity for each feature in the shapefile. I do this separately for the monsoon season 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>11</sup> For each day, I document whether the wind is blowing towards each village centroid<sup>12</sup> and then temporally aggregate this daily data depending on the temporal

<sup>9</sup>[springernature.figshare.com](http://springernature.figshare.com)

<sup>10</sup>[www.devdatalab.org/shrug\\_download/](http://www.devdatalab.org/shrug_download/)

<sup>11</sup>[climatedataguide.ucar.edu/](http://climatedataguide.ucar.edu/)

<sup>12</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.

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.

### 3 Methods

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

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

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 (log) concentration of particulate matter 2.5,  $rain_{it}$  is total rainfall during the season, and  $temp_{it}$  is average temperature during the season. In different specifications, I use different forms for  $f()$ ; the simplest specification uses just rainfall and temperature, but in more flexible specifications I turn rainfall into a binned variable – based on a z score – for each month of the monsoon season and include all of the bins along with interactions with temperature and squared temperature in each month.

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 officials purposefully place coal plants in poorer areas for political reasons, then the estimates might be downwards biased.

What are the characteristics of areas near the construction of new coal plants? Table A2 in the appendix presents four separate regressions, using state-level fixed effects due to the spatial variation in coal plants. 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.

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 and literacy both remain strong predictors of the building of a new coal plant.

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 well as average pollution during the monsoon season for 1998 through 2000 as additional predictors. 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. Pollution, on the other hand, is not significant but is positive in both columns (and is also logged, just like productivity).

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:

$$y_{it} = \alpha_i + \gamma_t + \beta \text{wind}_{it} + f(\text{rain}_{it}, \text{temp}_{it}) + \varepsilon_{it}, \quad (2)$$

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 – primarily agricultural productivity in this paper – except through exposure to pollution.

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, especially the most flexible specifications.

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 1 of the appendix. The outcome variable is particulate matter – specifically, PM2.5, which is particulate matter no larger than 2.5 micrometers in diameter – which comes from Hammer et al. (2020). PM2.5 is one of the harmful byproducts from coal plants, along with sulfur dioxide, different types of nitrogen oxides, and mercury.<sup>13</sup> The main results look at productivity during the main monsoon season, so in Table 1, I aggregate wind exposure to the season.

Both the particulate matter data and the wind data is at the village level, which means the regressions in Table 1 are at the village/season level. As such, the geographic fixed effects are village fixed effects and the temporal fixed effects are season/year 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. Columns one through four include more and more flexible specifications for the weather variable. The last column includes the most flexible specification, which turns rainfall into a binned variable for each month of the season and interacts these bins with temperature and temperature squared, also in each month of the season.

Across all columns, the coefficient on wind direction is quite consistent, ranging from 0.0083 to 0.0086. These coefficients are also very precisely estimated, leading to t-statistics of more than 30 in all columns. Wind direction is clearly a strong predictor of pollution exposure across the season, even conditioning on flexible specifications for weather.

---

<sup>13</sup>[www.epa.gov/airmarkets/power-plants-and-neighboring-communities](http://www.epa.gov/airmarkets/power-plants-and-neighboring-communities)

Table 1: Wind direction and pollution during the monsoon season

	(1)	(2)	(3)	(4)
wind	0.0083*** (0.0002)	0.0085*** (0.0002)	0.0083*** (0.0002)	0.0086*** (0.0002)
weather	No	Yes	Yes	Yes
weather (expanded)	No	No	Yes	No
weather (expanded, bins)	No	No	No	Yes
<b>fixed effects:</b>				
village	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes
F	1,236	1,266	1,205	1,264
observations	1,266,922	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 average pollution during the monsoon season. \* p<0.10 \*\* p<0.05 \*\*\* p<0.01

To put context behind pollution numbers, 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.

### 3.1 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:

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

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

Table 2: Particulate matter (PM 2.5) and agricultural productivity

	(1)	(2)	(3)
particulate matter (log PM 2.5)	-0.042*** (0.003)	-0.045*** (0.003)	-0.046*** (0.003)
weather	No	Yes	Yes
weather (expanded, bins)	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

indicator, or  $wind_{it} \times open_{it}$ .

## 4 Results

I first present the naïve results for agricultural productivity, regression productivity on pollution. The outcome of all the regressions in this section is the log of agricultural land productivity, which is defined as tons per hectare. These 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, and the pollution variable is defined as average pollution during the entire season at the village centroid.

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 the most flexible weather specification. There is a consistent, negative effect of wind exposure on agricultural productivity. Since both variables are logged, the coefficients are similar to elasticities, indicating that a one-percent increase in pollution leads to a 0.045 percent decrease in agricultural productivity, on average, across the three columns.

Recall, however, the results from above: coal plants were placed in places with higher average agricultural productivity. This means that the simple regressions in Table 2 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.

Table 3: Pollution and agricultural productivity, IV estimates

	(1)	(2)	(3)	(4)
particulate matter (log PM 2.5)	-0.941*** (0.078)	-0.769*** (0.075)	-0.621*** (0.080)	-0.437*** (0.075)
weather	No	Yes	Yes	Yes
weather (expanded)	No	No	Yes	No
weather (expanded, bins)	No	No	No	Yes
<b>fixed effects:</b>				
village	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes
F	1,236	1,266	1,205	1,264
observations	1,266,922	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

I present the IV results in Table 3, with the first-stage F-statistic at 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. Lee et al. (2022). Also note that the coefficients are quite different from those in Table 1 because the unit of analysis in Table 3 is the season, while in Table 1 the unit of analysis is the month.

The second-stage results first indicate that PM2.5 is a very strong predictor of agricultural productivity. Using the results in column four – again, the most flexible weather specification – a one-percent increase in particulate matter leads to a 0.44 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 4-to-7 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 3 are always more negative than those in Table 2, and 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.

#### 4.1 Leads for wind and pollution

In this section, I report the results of a robustness check to help validate the use of wind as an instrument for pollution. Specifically, I use leads of wind and particulate matter and show that these do not predict current

Table 4: Agricultural productivity and pollution leads

	(1)	(2)
particulate matter (one-year lead)	-0.073 (0.066)	
particulate matter (two-year lead)		-0.079 (0.051)
<b>fixed effects:</b>		
village	Yes	Yes
year	Yes	Yes
F	603	746
observations	1,161,265	1,055,562

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

agricultural productivity. I present these results in 4.

The first column looks at the effects of a one-year lead, while the second column looks at the effects of a two-year lead. In both cases, the coefficient is much smaller and insignificant than the main results in 3. Although the coefficients are negative, they are less than 20 percent the magnitude of the previous results. In addition, the first-stage F-statistic shows that the instrument continues to be strong, despite the loss of observations when using leads.

## 4.2 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 5, 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 opening.

Columns one through three of Table 5 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-tests indicate that the instruments are strong for both, including the interaction term. Second, the first row shows the effect of pollution on agricultural productivity prior to the opening of a coal plant; the coefficient is negative, indicating negative effects of non-coal pollution. 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, in addition to the overall negative effect of pollution.

Table 5: Agricultural productivity and pollution source

	(1)	(2)
PM 2.5	-0.228*** (0.066)	-0.223*** (0.065)
PM 2.5 times coal	-0.250*** (0.024)	-0.220*** (0.025)
weather	Yes	No
weather (expanded, bins)	No	Yes
<b>fixed effects:</b>		
year	Yes	Yes
village	Yes	Yes
F (pm)	769	772
F (pm times open)	231	205
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

Importantly, since these results come from an IV, both coefficients denote the change in agricultural yield due to an increase in pollution by the same units. In other words, this is not identifying differences in overall pollution, *per se*, but rather differences in the effects of an increase of pollution by the same percentage. These different effects could be due to, among other things, differences in emission patterns for different kinds of plants and factories; coal plants, for example, are responsible for much more mercury than other kinds of polluting industries.<sup>14</sup>

## 5 Conclusion

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.

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 document relatively large, local externalities in the agricultural sector in a developing country. Since the sector is a refuge for some of the poorest citizens in the world, the results raise important questions

<sup>14</sup>[https://www.epa.gov/sites/default/files/2016-12/documents/nei2014v1\\_tsd.pdf](https://www.epa.gov/sites/default/files/2016-12/documents/nei2014v1_tsd.pdf)

about the current speed of the energy transition as it relates to sustainable development and poverty reduction, more generally. While we often frame the debate around greenhouse gasses in terms of climate change, there are important local reasons to reduce pollution emissions, as well. The technology for greener energy exists; at this point, it is simply a question of policies (Markard 2018; Blazquez, Fuentes, and Manzano 2020).

These results, of course, do not address possible positive effects of the building of additional power plants, which can lead to increased electricity access, or shifting more of the population into non-agricultural sectors, which can raise standards of living. As such, the main conclusion is that the location of pollution emissions – especially from coal plants – 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.
- 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.
- Blazquez, Jorge, Rolando Fuentes, and Baltasar Manzano. 2020. "On Some Economic Principles of the Energy Transition." *Energy Policy* 147: 111807.
- 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.
- Christiaensen, Luc, Lionel Demery, and Jesper Kuhl. 2011. "The (evolving) role of agriculture in poverty reduction — An empirical perspective." *Journal of Development Economics* 96 (2): 239–54.
- Dercon, Stefan. 2009. "Rural poverty: Old challenges in new contexts." *The World Bank Research Observer* 24 (1): 1–28.
- 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.
- Herrendorf, Berthold, Richard Rogerson, and Akos Valentinyi. 2014. "Growth and Structural Transformation." *Handbook of Economic Growth* 2: 855–941.
- IEA. 2022. "Coal-fired electricity." International Energy Association.
- 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, 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.
- Markard, Jochen. 2018. “The Next Phase of the Energy Transition and Its Implications for Research and Policy.” *Nature Energy* 3 (8): 628–33.
- 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.
- 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.
- Sanders, Nicholas J, and Alan I Barreca. 2022. “Adaptation to Environmental Change: Agriculture and the Unexpected Incidence of the Acid Rain Program.” *American Economic Journal: Economic Policy* 14 (1): 373–401.
- 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
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)

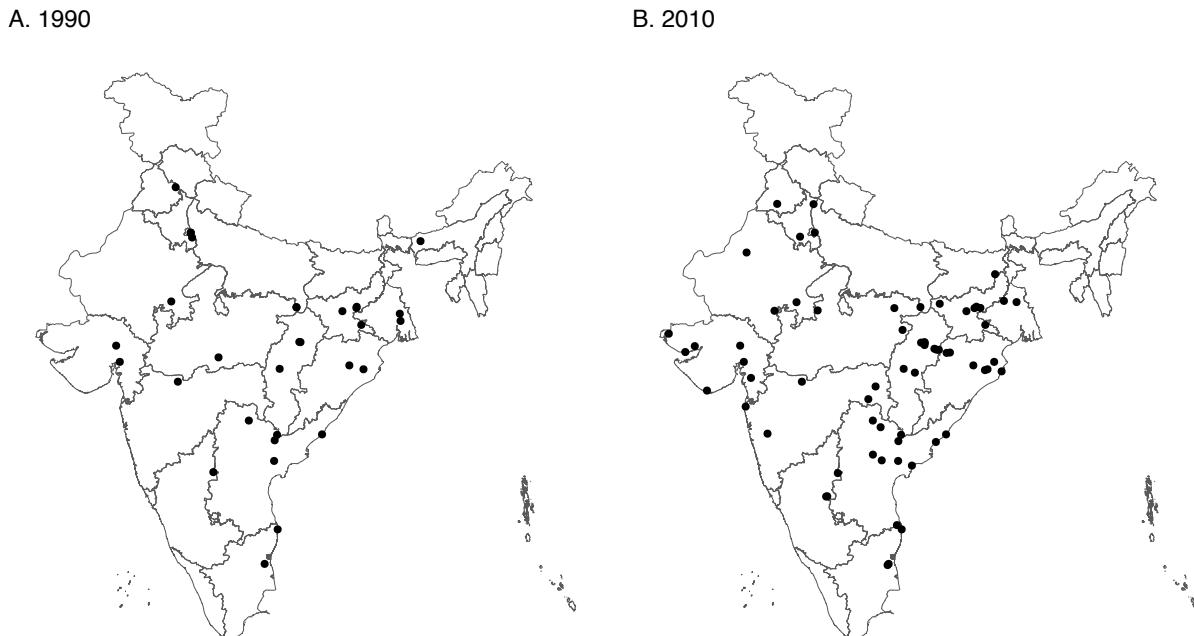
Table A2: Local characteristics and the construction of coal plants

	1991 census		2001 census	
	1991	2001	2001	2011
Population (log)	0.009 (0.006)	0.007* (0.004)	0.013* (0.006)	0.005 (0.004)
Area (log)	-0.007* (0.004)	-0.004 (0.005)	-0.007* (0.004)	-0.004 (0.003)
Literacy (prop)	0.080** (0.029)	0.037 (0.028)	0.073* (0.039)	0.036 (0.026)
Ag. productivity			0.014** (0.005)	0.028*** (0.009)
Pollution (log PM 2.5)			0.047 (0.045)	0.014 (0.018)
Sub-sample	all	no plant	all	no plant
Observations	283,021	264,265	508,784	469,004

Note: State fixed effects are included in regressions. Standard errors are in parentheses and clustered at the state level. 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.

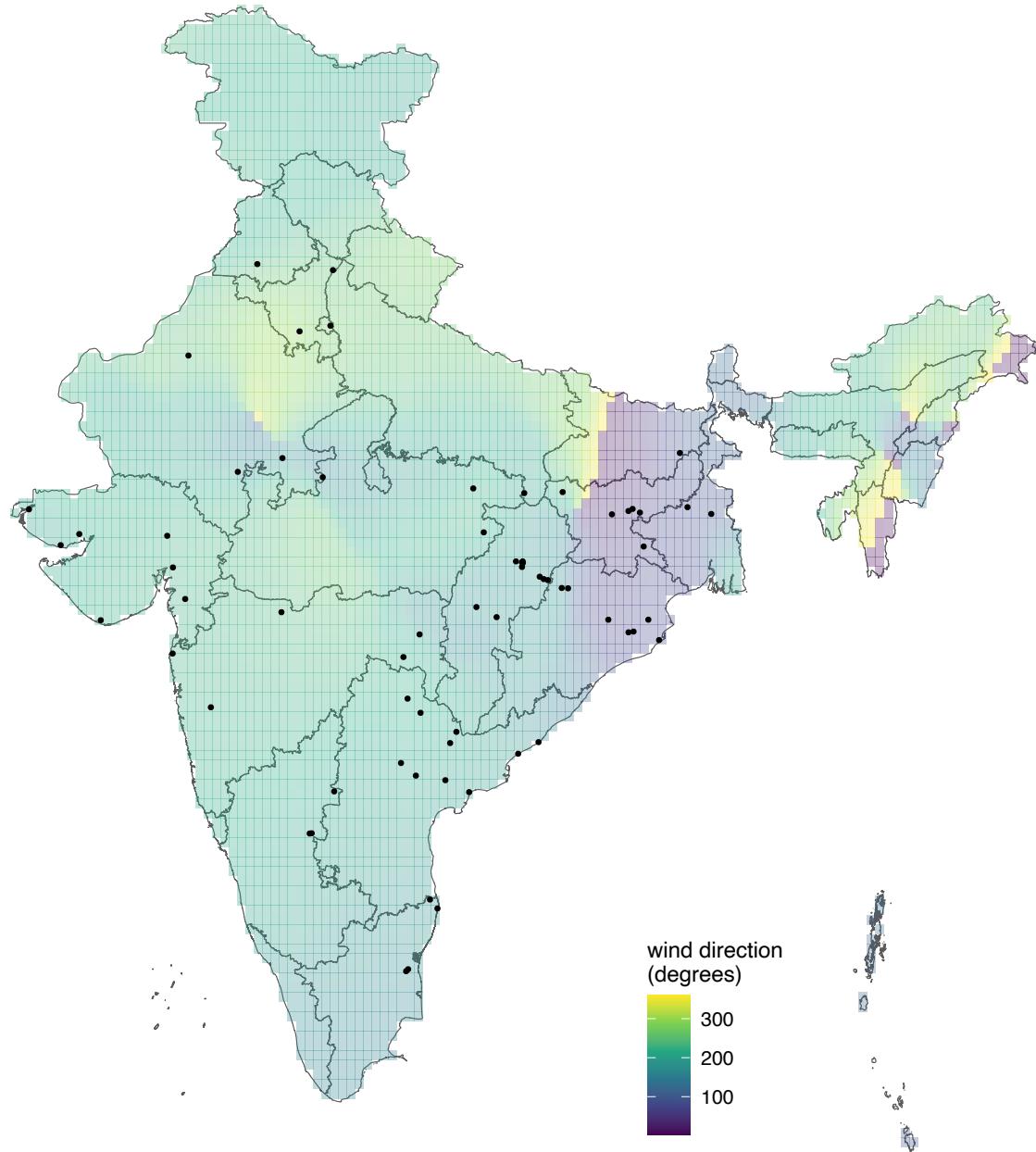
\* p<0.10 \*\* p<0.05 \*\*\* p<0.01

Figure A1: Coal plants in India from 1990 to 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 example (2010-01-01)



Note: The 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.