Digging for Answers:

The Effects of Mining Pollution Externalities on Local Agricultural Output in Africa

Mira Korb

Click here for latest version.

Abstract

The extraction of critical minerals and metals in Sub-Saharan Africa is expected to intensify in the coming years, potentially displacing existing economic activities. Mining can impact local agriculture in two key ways: through demand shocks, which raise the returns to inputs in non-agricultural sectors, and pollution shocks, which lower returns within agriculture itself. This paper examines the extent to which pollution externalities from industrial mining affect local agricultural output in Sub-Saharan Africa. Combining data on mine geolocations, topographical data, and satellite-based measures of pollution, crop yields, and weather, I identify areas around mines that are disproportionately exposed to pollution but not the economic effects of local demand booms. Leveraging variation around the openings of over 300 mines across the region, I find that air and water pollution externalities account for 44% of the overall reduction in yields caused by industrial mining, with water pollution being over 10 times more influential than air pollution. Furthermore, using machine learning methods, I show that negative pollution impacts on local yields are 2-3 times more severe for mines in countries with weaker governance and regulatory environments.

1 Introduction

A majority of the countries in Sub-Saharan Africa depend heavily on natural resources for government revenues, export earnings and development potential. Of the 48 countries in this region, 26 are classified as "resource-rich" by the International Monetary Fund, meaning that at least 20% of their exports or fiscal revenue derive from natural resources (Cust and Zeufack, 2023). This dependence on natural resource extraction in Africa is anticipated to increase. From electric vehicles to solar panels, the global transition to clean energy has led to a boom in demand for critical minerals and metals. According to the International Energy Agency, demand for cobalt and nickel is projected to double, while lithium demand could increase nearly ten-fold over the next two decades (Chen et al., 2024).

I investigate the extent to which industrial mining affects local agricultural output in Sub-Saharan Africa through pollution externalities versus raising the returns to inputs outside of agriculture. The emergence of new, high-return sectors like mining can attract resources and increase labor demand, fostering economic

development through structural transformation. However, mining also creates local pollution externalities that may lower agricultural productivity, causing reallocation of economic activity without accompanying increases in earnings.

The primary empirical contribution of this paper is to quantify the pollution externalities of mining on agricultural output in Sub-Saharan Africa. I combine a comprehensive geospatial dataset of mines from S&P Global Market Intelligence with topographical data and satellite-based measures of pollution, crop yields, and weather. To separate the effects of pollution from those of local demand booms, I exploit geographic factors—such as topography and wind patterns—that channel pollution in specific directions, unlike labor and land demand shocks, which are not similarly constrained. Using variation in pollution exposure around mine openings across Africa between 2000 and 2022, I apply two spatial difference-in-difference (DID) approaches: (1) comparing upstream and downstream areas to measure water pollution impacts, and (2) comparing areas with high and low downwind exposure for air pollution. In both cases, I define a mine opening as the start of commercial-level production. Local crop yields are proxied using the normalized difference vegetation index (NDVI) detected over cropland pixels.

To assess the impact of water pollution, I compare remotely sensed pollution levels and crop yields between areas located upstream and downstream of mines along rivers, before and after a mine opening. With this approach, I make an additional methodological contribution by using a novel, satellite-based proxy for total suspended solids (TSS) in water - the normalized difference turbidity index (NDTI). While NDTI has previously been used in remote sensing to monitor water quality in regions with limited data (Lacaux et al., 2007; Bid and Siddique, 2019), to the best of my knowledge, I am the first to apply NDTI in a causal inference context for a developing country. To estimate the corresponding effect of mining-induced water pollution on yields, I define upstream and downstream cropland around mines located along rivers, using downstream cropland as the treated group- representing areas that might be affected by the seepage of contaminated water into surrounding soil or the use of polluted river water for irrigation-and upstream cropland as the control.

Similarly, I use localized, daily wind direction data to identify cropland exposed to varying levels of air pollution from mines, proxied by satellite-based measures of aerosol optical density (AOD). I construct a continuous measure of downwind exposure for each of the four 90-degree sides of a circular buffer around a mine, corresponding to the cardinal directions that the wind may blow from the mine. For each side, downwind exposure is defined as the share of days in a month that the wind blows from the mine toward that direction. This setup allows me to identify treated sides—those downwind more frequently—which will be exposed to higher levels of pollutants such as particulate matter or sulfur dioxide from mining activities like drilling, blasting, or smelting. Sides that are downwind less frequently serve as controls. Additionally,

I examine the effects of both contemporaneous and cumulative exposure to air pollution on agricultural yields. Contemporaneous exposure refers to when air pollution in the current season affects yields within that same season, for instance by interfering with plant photosynthesis (Feng and Kobayashi, 2009; Schiferl et al., 2018). Cumulative exposure additionally accounts for the effects of air pollution in previous seasons on current yields, such as atmospheric deposition of pollutants changing soil quality over time (Sanders and Barreca, 2022).

I find evidence that mining activities increase pollution and reduce local yields, with these effects primarily driven by water pollution. In the upstream vs. downstream DID analysis, I observe that downstream water turbidity increases by over 12% around the time of a mine opening, relative to water upstream of a mine. This increase in turbidity coincides with a 3-4% reduction in NDVI for downstream cropland after a mine opens. The magnitude of this NDVI effect corresponds to a yield reduction of approximately 92 kg/hectare for the average smallholder farmer. In contrast, while mining increases air pollution in downwind areas by 2-3%, I find no corresponding decline in yields. Notably, air pollution levels start to rise about three years before the official start of commercial production, suggesting that construction and pre-production activities at the mine site contribute to local pollution. However, even after considering both contemporaneous and cumulative air pollution exposure, I find no significant impact of mining-induced air pollution on yields in areas with greater downwind exposure, at any point in the mine's life cycle.

Next, I benchmark the effects of mining pollution externalities on yields against an estimate of the overall effect of industrial mining on local agricultural output from a DID that compares NDVI in areas near mines to areas slightly further away, before and after a mine opening. On average, mine openings lead to a statistically significant 1-2% decrease in NDVI for areas within 20 kilometers of the mine, with negligible effects detected beyond this distance. Back-of-the-envelope calculations suggest that pollution externalities account for approximately 44% of the total impact of mining on agriculture.

Finally, I make a second empirical contribution by identifying the factors that drive variation in mine-specific treatment effects. My analysis reveals significant heterogeneity in treatment effects across mines, with the standard deviation of mine-specific treatment effects over ten times greater than expected from a homogeneous treatment effect with sampling error. In a novel application of machine learning (ML), I investigate drivers of treatment effect heterogeneity by predicting mine-specific treatment effects given a broad set of characteristics, including governance indicators, measures of local economic activity, environmental conditions, mine type, and commodity extracted. The ML model identifies governance measures - such as the strength of rule of law and regulatory quality - as the most important factors explaining variation in treatment effects. I then compare these ML findings to the results from traditional heterogeneity analyses, which use triple interactions with the dimensions of interest. I demonstrate the value of ML in disciplining

heterogeneity analysis when there are multiple plausible explanations for heterogeneity. Consistent with the ML model, the standard analysis confirms that governance quality moderates mining's impact on local agriculture: mines in poorly governed areas reduce yields by 2-3 times more than those in better-regulated areas. However, the standard analysis also reveals other statistically significant dimensions of heterogeneity that the ML model deems less important.

My paper contributes to three growing literatures. First, it adds to the literature studying the local impacts of natural resource extraction by examining the understudied question of how the pollution from extractive industries affects local agricultural output. The discussion of whether natural resources are a blessing or a curse (Sachs and Warner, 2001) has focused heavily on the effects of natural resource extraction on local economic outcomes, such as sectoral reallocation (Allcott and Keniston, 2018; James, 2015; Kotsadam and Tolonen, 2016), wealth (von der Goltz and Barnwal, 2019) and income (Aragón and Rud, 2013), as well issues of corruption and governance (Caselli and Michaels, 2013; Asher and Novosad, 2023; Martínez, 2023), conflict (Berman et al., 2017) and crime (Axbard et al., 2021). In a developing country context, our understanding of the effects of natural resource extraction on outcomes through the channel of pollution externalities is primarily limited to human health outcomes (von der Goltz and Barnwal, 2019; Benshaul-Tolonen, 2020) and human capital development (Bonilla Mejía, 2020; Rau et al., 2015). With the exception of Aragón and Rud (2016), who study the effects of 12 gold mines in Ghana, to the best of my knowledge there is no other micro-level work quantifying the effects of natural resource extraction on local agricultural outcomes in developing countries. I extend this analysis by studying over 300 commercial mines across Sub-Saharan Africa, isolating the pollution externalities from other local effects caused by mining, and examining sources of heterogeneity.

Second, my paper contributes to the literature on local structural transformation and agriculture, which suggests that industrialization drives structural change primarily through sectoral reallocation within small areas rather than across large distances (Ekert et al., 2023). As Sub-Saharan Africa is home to 30% of the world's critical minerals, policymakers argue that the region should take advantage of rising mineral demand to drive structural transformation (Chen et al., 2024). In fact, it seems that many African governments are already acting upon this rhetoric by providing tax breaks or other incentives to attract investment from mining companies (Coulibaly and Camara, 2022). By isolating the effect of pollution externalities from the effects of raising the returns to inputs in the non-agricultural sector, I am able to speak to the extent to which local reductions in yields are driven by pollution rather than structural transformation. Huang et al. (2023) and Kotsadam and Tolonen (2016) both find evidence that mining leads to structural shifts out of agriculture into more productive sectors, like low-skilled services. My finding that pollution externalities alone account for almost half of the effect of mine openings on local yields suggests that we must be wary

of interpreting observed declines in agriculture as evidence purely of structural transformation, especially in the long run. While long-lasting effects of pollution externalities may persist even after a mine closes, local economic opportunities may not (Black et al., 2005).

Finally, my paper contributes to an emerging literature that uses remotely sensed outcomes to address data gaps in developing countries. Existing literature has used NDVI to proxy for crop yields in settings where plot-level data on self-reported yields or crop cuts are unavailable (Sukhtankar, 2016; Emerick, 2018; Burke and Lobell, 2017). AOD has been used to overcome similar data gaps for ground-based measures of particulate matter and other air pollutants (Gendron-Carrier et al., 2022; Xie and Yuan, 2023; Zou, 2021). By applying the normalized difference turbidity index (NDTI) as a remotely sensed proxy for water quality, I extend this approach to the context of water pollution and causal inference. This innovation can allow researchers to answer questions in developing country settings where ground-measured indicators of water quality are spatially and temporally sparse, or otherwise completely unavailable.

The paper is structured as follows. Section 2 reviews the literature on how mining can generate pollution and the potential impacts of this pollution on local agriculture. Section 3 introduces a theoretical framework that distinguishes the effects of mining on agriculture due to pollution externalities from those driven by input demand shocks. Section 4 provides an overview of the data construction, while Section 5 details the DID methodology used in the analysis. Section 6 presents the main findings, with machine learning and standard heterogeneity analyses covered in Section 7. Finally, Section 8 presents a quantification exercise estimating the proportion of mining's overall effect on yields attributable to pollution, and Section 9 concludes.

2 Background

Mining pollution can affect crop yields through shocks that affect the degree to which land and labor inputs may be used productively. Shocks to land may arise through channels such as worsening soil quality, direct toxicity of certain pollutants to plant health and interference with plant photosynthetic processes. Labor shocks may arise through pollution affecting human health and resulting labor productivity, which has been documented extensively in the air pollution context (Graff Zivin and Neidell, 2012; He et al., 2019) but less so for water pollution (Russ et al., 2022). The following section provides background information on how mining pollution can affect yields through the plant health and soil quality channels, which is are less clear than the labor productivity channel.

2.1 Mining, Air Pollution and Crop Yields: The technologies employed by large-scale mining operations can contribute to air pollution in a variety of ways. First, metal smelting and refining produces

gaseous emissions, such as carbon dioxide, sulfur dioxide, and nitrogen oxide, as well as particulate matter (Dudka and Adriano, 1997). Sulfur dioxide emissions are especially common, as most economically important mineral ores occur as metallic sulfides, which are released into the atmosphere during blasting and smelting operations that expose rocks to the air and other chemicals. Furthermore, smelting and refining operations are major contributors to emissions of heavy metals, such as arsenic, cadmium, copper, lead and zinc. Since most mining operations in developing countries tend to process the raw ores on site before export, the pollution effects of these smelting and refining activities are concentrated in areas local to the mine. Second, activities such as blasting, crushing, stockpiling, loading and transportation of mine rocks also release large amounts of dust (Petavratzi et al., 2005). This dust contributes to fine particulate matter and heavy metal levels in the atmosphere. Finally, there are two main types of mining: underground mining and surface (also known as open pit) mining. While both types can contribute to air pollution, open pit mining is considered more damaging (Sahu et al., 2015).

Mining-induced air pollution can directly interfere with plant health and photosynthetic processes or indirectly affect crops by worsening soil quality. To begin, activities such as drilling, blasting or smelting can release pollutants into the atmosphere that affect plant photosynthesis or other plant growth processes. Whether air pollution has a positive or negative effect on plant health depends on the type and intensity of pollution. Some pollutants, like ozone, consistently reduce yields by limiting photosynthesis (Feng and Kobayashi, 2009). Effects on yields are more ambiguous for other pollutants, like particulate matter (PM), due to competing absorption and scattering effects, which can depend on the intensity of pollution (Behrer and Wang, 2024). While absorption by PM reduces total solar radiation reaching the surface and thus diminishes a plant's photosynthetic capabilities, scattering by PM increases the fraction of diffuse light available, which can be more effectively used by the plant (Schiferl et al., 2018; Burney and Ramanathan, 2014). The effects of nitrogen oxides on crop yields are similarly moderated by atmospheric conditions (Lobell et al., 2022). This suggests that the effect of mining-induced air pollution on local crop health could be highly heterogeneous across mines, agro-ecological zones and climates.

Furthermore, air pollution can lead to long-run deterioration of soil quality through atmospheric deposition of pollutants on the soil over time. As dust, heavy metals and other pollutants are released into the atmosphere from mining activities, they are carried by the wind and deposited on the ground. While some literature suggests that atmospheric deposition could erode soil quality over time (Liu et al., 2023), other literature finds that atmospheric deposition from polluting industries can actually serve as key nutrient inputs for modern high yield crops (Sanders and Barreca, 2022).

2.2 Mining, Water Pollution and Crop Yields: Mines can pollute surface water and groundwater sources through acid mine drainage (AMD), when water high in sulfates and heavy metals from mining activities is released into the soil or neighboring water bodies (Dudka and Adriano, 1997). Water contaminated with AMD can then affect crops either by overflow or seepage from contaminated rivers onto neighboring land, movement of pollutants through groundwater or irrigation. The extent to which polluted water affects crops through each of these channels depends on multiple factors, including how widespread the use of river water for irrigation is or local hydrological characteristics that influence how pollutants move through groundwater (Lapworth et al., 2017). In some cases, AMD can reduce yields substantially. Pot experiments reveal that fields contaminated with AMD have 62% lower grain yields than non-contaminated fields (Choudhury et al., 2017).

Like air pollution, water pollution from mines can have both contemporaneous and cumulative effects on yields. However, unlike air pollution, there is much less temporal variation in which areas are exposed to AMD. While wind direction can vary substantially within a year, blowing air pollutants in different directions around a mine, the direction of water flow is generally consistent over time.

3 Theoretical Framework

I assume that farmers produce an agricultural good with price p using labor and land (L and M), as well as a concave production technology (F) and idiosyncratic productivity (A). Output Y can be expressed as $Y = F(A, L, M)$.

Prior to a mine opening, each farmer is endowed with labor and land (E_L and E_M) that they can use as inputs to their farm. I assume that farmers are net suppliers of labor and land, which is reasonable given that smallholder farmers, who typically use family labor rather than hired labor, make up 60% of the population in Sub-Saharan Africa (Goedde et al., 2019).

The farmer then solves the following profit maximization problem:

$$\begin{aligned} \max_{L,M} \quad & \Pi = pF(A, L, M) + w(E_L - L) + r(E_M - M) \\ \text{s.t.} \quad & L \leq E_L \\ & M \leq E_M \end{aligned} \tag{1}$$

where the optimal choice of output and inputs, $Y^*(A, w, r)$, $L^*(A, w, r)$ and $M^*(A, w, r)$, depend only on productivity and input prices. Given these choices, the farmer earns $\Pi^* = pY^* + w(E_L - L^*) + r(E_M - M^*)$ before a mine opens.

When a mine opens, local factor prices around the mine will rise to $r' > r$ and $w' > w$. Areas that are are unexposed to pollution (e.g upstream or upwind) will experience no corresponding drop in productivity, A. Farmers in unexposed areas choose a new optimum, Y^{**}, L^{**} and M^{**}, earning new profits, $\Pi^{**} = pY^{**} + w(E_L - L^{**}) + r(E_M - M^{**})$. Inputs are reallocated out of agriculture and sold to other sectors, resulting in lower agricultural output but higher profits than before a mine opening.

In contrast, areas exposed to pollution from a mine (e.g. downstream or downwind) will experience the same increases in input prices to w' and r', as well as a drop in productivity to $A' < A$, driven by pollution. Exposed farmers choose a different optimum, Y^{***}, L^{***} and M^{***}, earning profits $\Pi^{***} = pF(A', L^{***}, M^{***}) + w(E_L - L^{***}) + r(E_M - M^{***})$. Lower productivity drives additional reallocation of inputs out of agriculture, so output and profits are lower than those for farmers in unexposed areas after a mine opens.

This implies that $Y^{***} < Y^{**} < Y^{*}$, $\Pi^{***} < \Pi^{**}$ and $\Pi^{**} > \Pi^{*}$. With sufficiently large declines in productivity, the loss in revenues from the sale of the agricultural good will outweigh the gains in earnings from higher wages. Farmers' profits in areas exposed to pollution will be lower than those earned before the mine opened ($\Pi^{***} < \Pi^{*}$), implying that mining pollution causes welfare losses.

4 Data

To investigate the effect of industrial mining on agricultural output in mining communities, I construct a panel of mining areas from 2000-2022. I link mine opening dates and spatial buffers around mine geo-locations, provided by S&P Global Market Intelligence, to remotely sensed measures of pollution, crop yields and weather variables. The unit of observation in my analysis is a mine-side-month, where a "side" refers to part of a buffer around a mine. In this section, I briefly explain the construction of the main variables used in my analysis, with an in-depth discussion left to Appendix 10.2.

4.1 Mines S&P Global Market Intelligence provides the geo-locations of the centroids of mining operations for large-scale mines across the world. Appendix Figure 11 shows the geo-locations of the S&P mines used in my analysis. These centroids are used to construct three different types of spatial buffers, which define treatment and control areas for my analysis. First, to estimate an overall effect of mining activity on yields, I define circular buffers around each mine to distinguish between areas near mines and areas slightly further away. Areas near mines are more likely to experience the effects of both local demand shocks and pollution exposure from mining activity, relative to areas further away. I refer to the buffer of radius 20-kilometers around a mine as the near group, which I assume is the treated area that would be

impacted by both higher returns to inputs outside of agriculture and pollution. I use findings from existing literature to inform the 20-km cutoff for the spatial extent of local markets around mining areas. Empirical evidence on commuting distances in urban and rural Africa suggests that areas of 5, 10 or 15 km are likely integrated markets (Amoh-Gyimah and Aidoo, 2013; Kung et al., 2014; Shafer, 2000). Furthermore, wealth increases (von der Goltz and Barnwal, 2019) and structural shifts out of agriculture into services (Kotsadam and Tolonen, 2016) are concentrated within 20-km of a mine. To demonstrate that pollution effects are concentrated close to mines, I estimate a spatial lag model with AOD as the outcome of interest. Appendix Figure 12 shows that AOD increases after mine openings are largest within 20-km of a mine, with smaller, statistically significant effects persisting up to 60-km away and no statistically distinguishable effects at greater distances.

Second, I buffer river segments near mines to estimate effects of mining-induced water pollution on yields. I use the HydroRIVERS dataset to identify rivers located in Africa. Each mine within 1 km of a river in this network is matched to the closest river segment, then the upstream and downstream river segments from the mine river segment are identified. The upstream and downstream segments are buffered by 1-km on either side of the river to identify land around the rivers that would be affected by water pollution. The choice of a 1-km buffer is based on contaminant transport modeling from mining areas, which detects sulfates in groundwater up to 2.5 km away from the polluting source, with up to 1.5 km of horizontal dispersion (Myers, 2016). The "side" of a mine in the water pollution analysis refers to the upstream or downstream buffer.

Lastly, to estimate effects of mining-induced air pollution on yields I use slices of circular buffers around each mine that indicate wind direction from the mine. I start by defining a buffer of radius 60-kilometers around each mine, chosen based on the AOD spatial lag model shown in Appendix Figure 12. For each mine, the 60-km buffer is divided into four 90-degree slices representing the cardinal directions that the wind can blow from the mine centroid, where each of these slices represents a "side" of a mine in the air pollution analysis. To limit the influence of highly local air pollution effects that would affect all sides of the mine regardless of wind direction, the circle of radius 5-km around the mine centroid is removed from the 60-km buffer, leaving behind a "donut."

Figure 1 illustrates the three different types of treatment and control areas constructed. The left panel illustrates the near versus far buffers used to estimate an overall effect of mining on local yields, where areas within the 20-km buffer are considered treated. The top right panel shows the upstream and downstream buffers used to estimate the water pollution effect, while the bottom right panel displays the wind direction buffers used to estimate the air pollution effect.

In line with S&P, I define the start of mining activity, also referred to as the opening of a mine, as the year that "the mine/plant has been commissioned or has produced its first metal, concentrates, or bulk

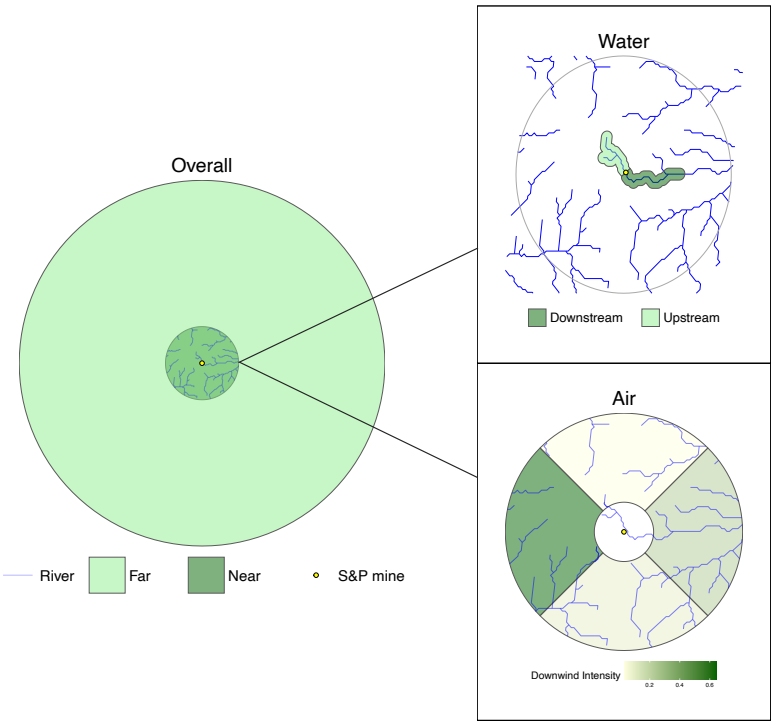

Figure 1: Treatment and Control Buffers

The figure plots the three different types of treatment and control areas used in the analysis. In each sub-figure, the yellow circle indicates the centroid of the mining area. The left panel illustrates the near versus var buffers used to estimate the "overall" effect of mining on local yields. The dark green inner circle has a radius of 20-km and is used to define the "near" group, which is affected by both input demand shocks and pollution externalities (treated), while the light green outer circle acts as a control. The two sub-figures in the right panel zoom into the near group, showing the treatment and control areas for the water and air pollution analyses. The top right panel shows the downstream (treated) and upstream (control) buffers for the river segments around the S&P mine, while the bottom right panel shows the intensity of downwind exposure (treatment) experienced by each of the four cardinal directions around the mine, defined by 90-degree slices of a circle.

commodity at a commercial rate." Other papers have defined mining activity based on years where a mine had non-zero output (von der Goltz and Barnwal, 2019; Benshaul-Tolonen, 2020). Although the S&P data provides yearly production data for some mines, this information is missing or incomplete for many mines in Africa, meaning activity would need to be imputed. As the S&P data on the start year of commercial production is more complete, I opt to define mining activity based on this year instead.

Although the S&P mining data covers mostly large-scale mines operated by multinationals or national governments, my empirical strategy is still able to capture the effects of artisanal, small-scale or illegal mines operating on the fringes of large-scale operations. As the mine buffers extend beyond the footprint of the large-scale mine, they should capture area also affected by informal mining. Additionally, event study estimates shown in Section 6 confirm that increases in remotely-sensed air and water pollution occur around the time of large-scale mine openings in my data, suggesting that informal operations are not polluting prior to the start of large-scale mining activity.

4.2 Wind: Wind data for the air pollution analysis is retrieved from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA2). MERRA2 is a re-analysis product that combines satellite imagery with algorithms and atmospheric models. I define monthly downwind exposure for each of the four 90-degree sides of the mine buffer as the share of days in a month that the wind is blowing from the mine into that side.

4.3 Satellite-based air pollution: Daily measures of aerosol optical density (AOD) are obtained from the Moderate Resolution Imaging Spectroradiometers (MODIS) satellites at a 3-kilometer spatial resolution. In brief, AOD is calculated by comparing the light intensity in a particular band against a reference value and attributing the difference to particulates in the air column. The higher the level of AOD, the higher the level of air pollution as more light is reflected back by particulates in the atmosphere. Several studies have shown that AOD is highly predictive of ground-based measures of PM10 and PM2.5. In their preferred specification, Gendron-Carrier et al. (2022) find that one unit of remotely sensed AOD is associated with about 114 $\mu g/m^3$ of PM10 measured by a ground-based instrument.

Given MODIS data availability, the air pollution analysis covers 2003 - 2017. From daily AOD rasters, I construct mean AOD for each mine-side-month by averaging all pixel-days with non-missing AOD readings that fall within each side of a mine in a given month. I calculate this average monthly AOD only using data from the Aqua satellite, as the Terra satellite suffers from more missing values due to satellite detection errors during my time period of interest.

4.4 Satellite-based water pollution While satellite-based measures have been used extensively to answer causal inference questions related to air pollution (Gendron-Carrier et al., 2022; Gutiérrez and Teshima, 2018; Chen et al., 2022), to the best of my knowledge, no paper is yet to conduct causal inference using remotely-sensed measures of water pollution in developing countries.

In this paper, I use the normalized difference turbidity index (NDTI) as a remotely-sensed measure of water pollution. NDTI uses the spectral reflectance of water pixels to estimate turbidity, which is a measure of water clarity. Similar to AOD, turbidity measures the amount of light scattered by particles in the water column. The higher the level of particles in the water, the more light that will be scattered. High turbidity makes water appear cloudy, muddy or discolored. Turbidity is often used to proxy for total suspended solids (TSS), which are a common indicator of water quality. TSS are particles found in the water column that are larger than 2 microns. Increases in TSS can be driven by natural phenomena, such as stirred bottom sediments from heavy precipitation or algal blooms, as well as industrial runoff or wastewater discharge. When TSS levels cannot be measured, changes in turbidity can be used to reflect changes in TSS concentration in water (Wetzel, 2001).

To measure turbidity, NDTI leverages how electromagnetic reflectance is higher in green spectrum than the red spectrum for clear water (Lacaux et al., 2007; Gardelle et al., 2010). Increased reflectance of the red spectrum relative to green corresponds to an increase in turbidity (Islam and Sado, 2006). NDTI is calculated with the red and green spectral bands according to the following formula:

$$NDTI = \frac{\text{Red} - \text{Green}}{\text{Red} + \text{Green}}$$

Generally, NDTI ranges from -0.2 to greater than +0.25, where lower (negative) values indicate clear water and higher (positive) values indicate turbid water (Sharma et al., 2015).

I use Landsat 7 data from 2000-2022 at a 30-m resolution to calculate daily pixel-level NDTI for river pixels only, then average across pixel-days within the upstream and downstream river segments for each mine. In Appendix Table 6, I show that NDTI is highly predictive of ground-measured TSS from the Global River Water Quality (GRWQ) database. In my preferred specification, a one unit increase in NDTI is associated with a 887 mg/l increase in ground-measured TSS. Additional details on the calculation of NDTI and the ground-truthing exercise are covered in Appendix 10.2.

Importantly, while NDTI is successful at detecting changes in the overall levels of particles in water, it cannot distinguish between types of particles. While some particles may be pollutants from mine waste effluent, such as heavy metals or acidic sulfides, others may be sediment from rock or soil disturbed during the extraction process. Since NDTI cannot distinguish between polluting particles and sediment, it is important

to interpret observed changes in NDTI due to mining as changes in both harmful pollutants and potentially non-harmful sediment.

Satellite-based proxies for pollution have an key advantage in that they allow researchers to answer questions in contexts where ground-based pollution data is unavailable or limited. Importantly, spatial and temporal gaps in ground-based water pollution data can be even larger than those for air pollution in developing countries (Virro et al., 2021). While low cost air quality sensors have been deployed in many African countries, the development and implementation of low cost water quality sensors is in its infancy. These limitations in ground-based water quality monitoring highlight the value of remotely-sensed measures of water pollution, which offer an affordable way to track the quality of African water bodies over time.

4.5 Satellite-based yields: I proxy for crop yields using the normalized difference vegetation index (NDVI), which is strongly correlated with crop productivity and final yields, especially during the growing season (Panek and Gozdowski, 2021). I calculate daily mean NDVI at the pixel level from 2000 - 2022 using the MODIS MCD43A4 Version 6.1, combined Terra and Aqua product at a 500-m resolution. NDVI is calculated from the MODIS data using readings of light reflected in the near-infrared and red spectrum.[1] Given pixel-level data, I construct my final measures of cropland NDVI for the treatment and control areas by averaging NDVI across all pixel days identified as cropland by the Global Food Security Support Analysis Data (GFSAD) within the relevant buffers, for each mine-side-month. I discuss additional details on NDVI construction in Appendix 10.2 and how measurement error in NDVI might affect my estimates in Appendix 10.13.

4.6 Weather Controls: Local weather conditions influence actual yields, as well as the ability for satellites to accurately detect these yields (Gendron-Carrier et al., 2022; Heino et al., 2023). I calculate mine-side-month averages of cloud cover percentage, temperature, precipitation, evapotranspiration, vapor pressure and wet day frequency using the Climatic Research Unit gridded dataset, available at a 0.5 degree resolution from Harris et al. (2014). These monthly averages of key weather variables are the main control variables in my regressions.

4.7 Agricultural Seasons: For each mine, I aggregate daily measures of NDTI, AOD and NDVI to the monthly level from 2000-2022. Each mine-month is then linked to one of the following agricultural seasons: planting, early growing, late growing, harvest and non-farm, to investigate whether observed yield effects differ across seasons. I use the Sacks et al. (2010) global raster dataset on planting and harvesting dates for maize to define area-specific agricultural seasons. This data is available at a 5 arc-minute (approximately

[1] NDVI = $\frac{NIR-RED}{NIR+RED}$

10 km × 10 km) spatial resolution. I focus on maize as it is one of the primary crops grown in Africa and has the most complete crop calendar data. Other important crops like cassava or wheat have missing data across most of Africa and Sacks et al. (2010) caution against using interpolated products. Additional details on construction of agricultural seasons based on the crop calendar data are covered in Appendix 10.2.

5 Methodology

To estimate the impact of mining on yields, I use a difference-in-difference (DID) design that compares areas exposed to the effects of mining to unexposed areas, before and after a mine opening:

$$Y_{smt} = \beta Exposure_{sm} \times Post_{mt} + \alpha_{sm} + \lambda_{mt} + \mathbf{X'_{smt}}\mathbf{\Gamma} + \epsilon_{smt} \tag{2}$$

where s indexes the side of mine m, t indexes the month and outcome Y is either remotely sensed pollution (NTDI, AOD) or yields (NDVI). $Post_{mt}$ is a dummy variable equal to 1 after mine m opened, 0 otherwise. The definition of $Exposure_{sm}$ and side s varies across the three types of analysis: water, air and overall. For water, I define the downstream cropland as exposed and the upstream cropland as unexposed while for air, areas downwind from a mine more frequently are exposed and areas downwind less frequently are unexposed. Finally, for estimation of the overall effect of mining I define areas within 20-km of the mine as exposed to both pollution externalities and the effects of input demand shocks, with areas further away considered unexposed.

I include mine-side fixed effects (α) to control for time invariant unobservables correlated with agricultural output on each side of a mine, such as soil quality, as well as mine-date fixed effects (λ) to control for mine-specific trends in outcomes over time. These fixed effects absorb the main effects of $Exposure$ and $Post$. Given the mine-side and mine-date fixed effects, β is identified by within-month differences in the change in pollution or yields among areas exposed and unexposed to the effects of mine openings. Essentially, β is a weighted average of mine-specific DIDs generated from 40 to 300 mine openings, where the exact number of mines used depends on the type of analysis. To address the influence of weather shocks to yields that vary across sides and time, I control for linear and quadratic terms of mean cloud cover, vapor pressure, temperature, precipitation, evapotranspiration and wet days. Standard errors are clustered at the mine level.

Estimating the causal effect of mining on yields requires that the timing and placement of a mine opening is plausibly exogenous to local changes in NDVI. While the placement of mineral deposits is random, the discovery of these deposits may depend on a variety of non-random factors, including local governance quality, business environments, infrastructure access and input prices (Benshaul-Tolonen, 2020). This should not pose

a threat to identification as upstream (upwind) and downstream (downwind) sides of the same mine should be similarly affected by these non-random factors. Mine-side and mine-year-month fixed effects also help alleviate endogeneity concerns.

Additionally, I estimate event study and season-specific versions of Equation 2 for all three types of analysis. For the event study model, I replace the single $Post_{mt}$ dummy with a series of dummy variables indicating years since the opening of mine m. To isolate season-specific effects on yields, I estimate Equation 2 separately over months in the growing and non-growing seasons.

6 Results

6.1 Mining-induced Water Pollution To estimate the effect of mining-induced water pollution on yields, I use a DID that compares remotely-sensed water turbidity or NDVI between the upstream and downstream sides of a mine, before and after a mine opening. I define $Exposure_{sm}$ in Equation 2 to be a dummy variable equal to 1 if side s is downstream from mine m and equal to 0 if it is upstream. I estimate Equation 2 over a partially balanced panel of 38 mines within 1 km of a river, where non-missing NDVI is observed on both the upstream and downstream sides of the mine, for at least 6 months of every year, for at least 2 years pre- and 2 years post-mine opening.

To demonstrate that mining indeed increases water pollution, I first estimate Equation 2 with the normalized difference turbidity index (NDTI) as the outcome of interest. Since NDTI is calculated over narrowly defined river pixels, monthly NDTI measures are missing more frequently after masking out low quality pixels. As a result, I estimate the event study and DID regressions for turbidity over a subset of the mines used in the NDVI analysis.

Figure 2 shows that on average after a mine opens, downstream river turbidity increases significantly. Larger increases in turbidity during the growing seasons may be explained by increased precipitation causing more runoff from mining areas into rivers. Figure 3 reveals a complementary pattern of reductions in NDVI for downstream cropland after a mine opening, with stronger negative effects during the growing seasons. Prior to a mine opening, we see evidence of parallel trends between upstream and downstream areas for both turbidity and yields.

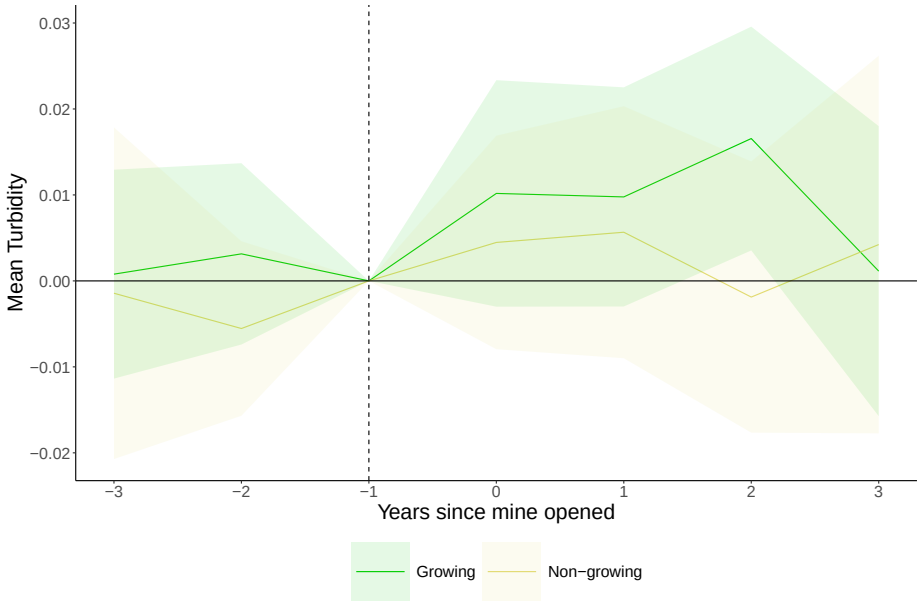

Figure 2: Effect of mining on water turbidity (NDTI)

The figure plots coefficient estimates from two separate event study regressions of mean normalized difference turbidity index on a dummy for whether a side is downstream from a mine interacted with a series of event time dummies for years since a mine first opened. The unit of analysis is a mine-side-month, though event time coefficients are binned into yearly increments. The event study regressions are estimated separately over growing season months (early growing and late growing) and non-growing season months (planting, harvest and non-farm). The shaded bands show point-wise 95 percent confidence intervals for the coefficients of the event-time path of mean NDTI. Coefficients can be interpreted as estimated effects relative to the period one year before the mine opened. Both regressions includes linear and quadratic controls for mean temperature, precipitation, vapor pressure, wet days, evapotranspiration and cloud cover, as well as mine-side and mine-year-month fixed effects. Standard errors are clustered at the mine level. The sample includes the 30 mines for which non-missing NDTI is observed on the upstream and downstream sides for at least 2 months in each year, for at least 2 years pre- and 2 years post-mine opening. Event times less than -2 or greater than 2 are binned into two end points.

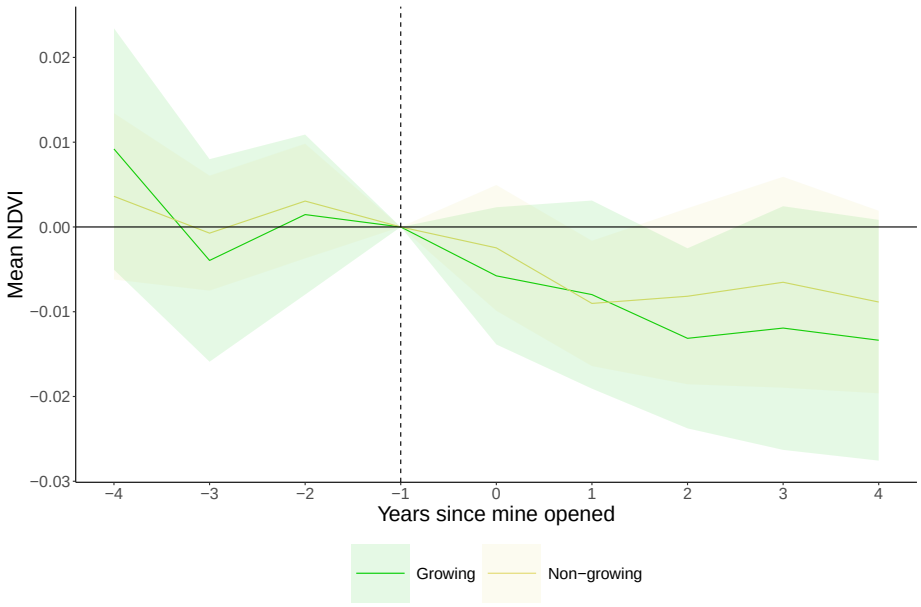

Figure 3: Effect of mining-induced water pollution on yields (NDVI)

The figure plots coefficient estimates from two separate event study regressions of mean NDVI on a dummy for whether a side is downstream from a mine interacted with a series of event time dummies for years since a mine first opened. The unit of analysis is a mine-side-month, though event time coefficients are binned into yearly increments. The event study regressions are estimated separately over growing season months (early growing and late growing) and non-growing season months (planting, harvest and non-farm). The shaded bands show pointwise 95 percent confidence intervals for the coefficients of the event-time path of mean NDVI. Coefficients can be interpreted as estimated effects relative to the period one year before the mine opened. Both regressions includes linear and quadratic controls for mean temperature, precipitation, vapor pressure, wet days, evapotranspiration and cloud cover, as well as mine-side and mine-year-month fixed effects. Standard errors are clustered at the mine level. The sample includes the 38 mines for which non-missing NDVI is observed on the upstream and downstream sides for at least 6 months in each year, for at least 3 years pre- and 3 years post-mine opening. Event times less than -3 or greater than 3 are binned into two end points.

Table 1 presents the main results for remotely sensed turbidity and yields, estimated on all months pooled together, as well as separately by growing and non-growing seasons. Panel A consistently shows that mining increases remotely sensed water turbidity, though these estimates are statistically insignificant, likely due to lack of power. The increase in turbidity is on the magnitude of 12-50% relative to NDTI in the pre-period. I cautiously interpret these results as evidence that mining increases water pollution, though some of this effect may be attributed to increases in non-harmful sediment as well. Using Column 2 of Appendix Table 6 to convert NDTI to total suspended solids suggests that on average, mine openings increase total suspended solids in downstream water by 4.33 - 4.83 mg/l.

The resulting effects of mining-induced water pollution on yields in Panel B show consistent and statistically significant reductions in remotely sensed yields in downstream areas relative to upstream areas, after a mine opens. These estimates correspond to a 3-4% drop in NDVI, with larger effects observed during months in the growing season.

Table 1: Relationship Between Industrial Mine Openings, Remotely-Sensed Water Turbidity and Crop Yields

	(1)	(2)	(3)
	All seasons	Growing	Non-growing
Panel A: Turbidity (NDTI)			
Downstream × Post	0.00488	0.00426	0.00544
	(0.00649)	(0.00761)	(0.00636)
Mines	30	30	30
Observations	11,330	5,304	6,026
Mean NDTI (t-1)	.01	-.01	.04
Panel B: Yields (NDVI)			
Downstream × Post	–0.01404**	–0.01770**	–0.01093**
	(0.00587)	(0.00714)	(0.00531)
Mines	38	38	38
Observations	20,518	9,888	10,630
Mean NDVI (t-1)	.47	.54	.39
Mine-side FE	Yes	Yes	Yes
Mine-year-month FE	Yes	Yes	Yes
Weather	Yes	Yes	Yes

Notes: Each column reports the results of a linear regression. The unit of analysis is a mine-side-month. In Panel A, the dependent variable is the mean normalized difference turbidity index (NDTI) of the river water on the upstream or downstream side of a mine, in a given month. I derive remotely sensed turbidity from Landsat 7. In Panel B, the dependent variable is the mean normalized difference vegetation index (NDVI) of the land within the 1km buffer along the river on either the upstream or downstream side of a mine, in a given month. I derive remotely sensed yields from the MODIS Combined Terra and Aqua product (MCD43A4.061). *Downstream* is equal to 1 if the side is downstream from the mine and 0 if it is upstream from the mine. *Post* is equal to 1 after the mine opened, 0 otherwise. All models include linear and quadratic controls for mean temperature, precipitation, vapor pressure, wet days, evapotranspiration and cloud cover, as well as mine-side and mine-year-month fixed effects. For turbidity, the sample includes the 30 mines for which non-missing NDTI is observed on the upstream and downstream sides for at least 2 months in each year, for at least 2 years pre and 2 years post-mine opening. For yields, the sample includes the 38 mines for which non-missing NDVI is observed on the upstream and downstream sides for at least 6 months in each year, for at least 3 years pre and 3 years post-mine opening. Column 1 reports results estimated by pooling months over all 5 seasons: planting, early growing, late growing, harvest and non-farm. Column 2 reports results estimated only over months in the early growing and late growing seasons and Column 3 reports results estimated only over months in non-growing seasons: planting, harvest and non-farm. Standard errors in parentheses are clustered by mine.

To understand the extent to which water pollution may be affecting local crop yields through irrigation, I estimate Equation 2 using only NDVI over irrigated cropland as the outcome of interest. The GFSAD cropland extent data classifies cropland pixels as rain-fed or irrigated, allowing me to to identify cropland near rivers that may be negatively affected by the use of polluted water during irrigation. Appendix Table 7 shows that estimated reductions in NDVI due to mining are about 25% larger for irrigated downstream cropland compared to non-irrigated areas, suggesting that irrigation with polluted water is indeed an important channel through which mining-induced water pollution may affect local yields.

6.2 Mining-induced Air Pollution To identify areas near mines that are disproportionately exposed to air pollution, I rely on the assumption that wind blows pollutants from the mine towards the downwind side of the mine. I apply high-frequency data on wind direction and the staggered openings of over 100 large-scale mines in two methods that estimate the effect of mining-induced air pollution on local yields: (1) a DID and event study that primarily capture the effects of contemporaneous exposure to air pollution and (2) a distributed lag model that allows for cumulative effects of prolonged exposure to air pollution.

6.2.1 Contemporaneous effect: The model capturing contemporaneous effects of mining-induced air pollution on local yields uses a continuous measure of *Exposure* in Equation 2, defined as the share of days in month t that side s is downwind from mine m. In this way, a one unit increase in *Exposure* corresponds to moving from a side never being downwind from a mine to a side being downwind for the entire month.

Aside from the definition of *Exposure*, the air pollution specification differs from the one used in the water pollution analysis in a few ways. First, unlike direction of water flow, wind direction varies substantially over time, so *Exposure* in Equation 2 varies across mines, sides and time. This means that the *Exposure* main effect is not absorbed by the mine-side and mine-year fixed effects. Additionally, I control for average wind speed experienced by each side of the mine in a given month.

To establish that mining activity increases air pollution, I estimate the air version of Equation 2 using mean AOD as the outcome of interest. Unlike the NDVI data, the AOD data is available only for 2003-2017 and has more missing values due to errors in AOD detection from cloud cover. To ensure a consistent sample of mines between the NDVI and AOD regressions, I use a partially balanced panel of 102 mines, where each mine must be observed on all sides for at least 4 months of every year, for at least 5 years pre- and 5 years post-mine opening. I additionally control for the number of non-missing pixel days used to construct mean monthly AOD to address measurement error in AOD,

The AOD event study in Figure 4 reveals that on average, sides of a mine that are downwind more

frequently experience higher levels of AOD after a mine opening. This increase in AOD begins about 3 years before the mine officially starts commercial production. Prior to this 3-year period before a mine opens, we see evidence of parallel trends between areas with high downwind exposure and areas with low downwind exposure.

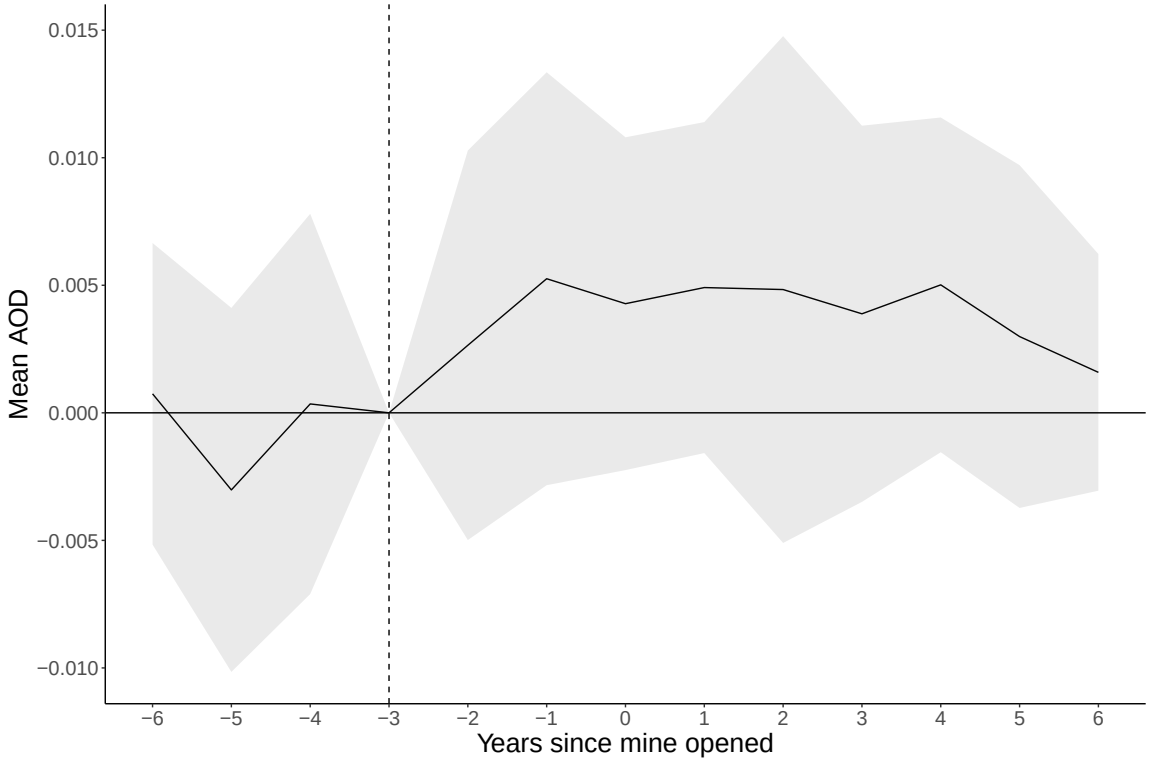

Figure 4: Effect of mining on air pollution (AOD)

The figure plots coefficient estimates from a regression of mean AOD on the share of days in a month that a side is downwind from a mine, interacted with a series of event time dummies for years since a mine first opened. The unit of analysis is a mine-side-month, though event time coefficients are binned into yearly increments. The shaded bands show pointwise 95 percent confidence intervals for the coefficients of the event-time path of mean AOD. Coefficients can be interpreted as estimated effects relative to the period three years before the mine opened. The regression includes linear and quadratic controls for mean temperature, precipitation, vapor pressure, wet days, evapotranspiration and cloud cover, as well as mine-side fixed effects, mine-year-month fixed effects, mean wind speed and a linear control for the number of non-missing pixel days used to calculate mean AOD. The sample includes the 102 mines for which non-missing AOD is observed on all 4 sides of the mine, for at least 4 months in each year, for at least 5 years pre- and 5 years post-mine opening. Event times less than -5 or greater than 5 are binned into two end points.

Why would air pollution occur before a mine officially starts producing? To answer this question, it is helpful to reconceptualize mining activity as a gradual ramp up over time rather than a binary "on/off." Detailed work histories of the mines in the S&P database provide evidence in support of this gradual increase in activity. S&P roughly classifies mining activity into the following phases: discovery, exploration/feasibility study, construction/pre-production, production and closure. The discovery phase describes the time at which mineral potential was first determined in a mining area. Exploration and feasibility studies encompass activi-

ties such as drilling and testing, which are used to estimate the quality and quantity of reserves. Construction and pre-production include activities related to the building of infrastructure needed for large-scale mineral extraction (e.g. roads, electricity grid, buildings). Production refers to industrial-level extraction and processing of raw materials. Closure refers to the permanent shut-down of a mine and the associated reclamation activities that are supposed to return the land to its previous state.

It is not unreasonable to observe increased air pollution in downwind areas prior to the official start of mine production. Literature estimating emissions rates from different activities at a mine site suggests that air pollutants are most likely to be released during the exploration, construction/pre-production and production phases (Patra et al., 2016). In particular, the use of heavy machinery or vehicles during exploration, as well as the clearing of overburden or construction may contribute to local air pollution levels before the official start of production.

To better understand the polluting effects of the different phases of a mine's life cycle, I use satellite data to estimate a structural break in the mean of remotely sensed nighttime light intensity for each mine that opened between 2003 and 2012 using the methods of Andrews (1993) and Andrews and Ploberger (1994).[2] I hypothesize that the construction of infrastructure in mining areas would dramatically increase nightlights and so interpret the structural break in nightlights as aligning with the construction/pre-production phase. Additional details on the structural breaks estimation procedure are covered in the Appendix 10.12.

For the average mine, I find that a structural break in nightlights occurs about 3 years prior to the S&P start date of commercial production. This is similar to Benshaul-Tolonen (2020), who also uses nightlights to demonstrate an "investment phase" of mining activity occurring 2 years before the start of production. In addition, this finding supports Figure 4, which shows that increases in air pollution begin about 3 years before mines start producing. I cautiously interpret this result as evidence that air pollutants generated during the construction and pre-production phase of a mine are major drivers of local air pollution from mines.

Although we observe increases in air pollution around the time of a mine opening, the event study in Figure 5 reveals no economically or statistically significant effects of contemporaneous exposure to air pollution on NDVI in either the growing or non-growing seasons, at any point in a mine's life cycle.

The event study findings for both AOD and NDVI are supported by the DID estimates in Table 2. After establishing that mining starts to increase air pollution in downwind areas during the investment phase, about 3 years before official production begins, I redefine the *Post* dummy in Equation 2 to be equal to 1 for years after the start of the investment phase. I use this definition of $(Post - 3)$ when estimating the effect

[2]In 2013 there was a switch from the DMSP OLS to the VIIRS instrument as the source of nightlights. Since this switch introduced an artificial break in nightlights in 2013, I only estimate structural breaks for mines that opened prior to 2013.

Figure 5: Effect of contemporaneous mining-induced air pollution on yields (NDVI)

The figure plots coefficient estimates from two separate event-study regressions of mean NDVI on the concurrent share of days in a month that a side is downwind from a mine, interacted with a series of event time dummies for years since a mine first opened. The unit of analysis is a mine-side-month, though event time coefficients are binned into yearly increments. The regressions are estimated separately over growing season months (early and late growing) and non-growing season months (planting, harvest and non-farm). The dotted lines show pointwise 95 percent confidence intervals for the coefficients of the event-time path of mean NDVI. Coefficients can be interpreted as estimated effects relative to the period three years before the mine opened. The regression includes linear and quadratic controls for mean temperature, precipitation, vapor pressure, wet days, evapotranspiration and cloud cover, as well as mine-side, mine-year-month fixed effects and mean wind speed. The sample includes only the 102 mines for which non-missing NDVI and AOD is observed on all 4 sides of the mine, for at least 4 months each year, for at least 5 years pre- and 5 years post-mine opening. Event times less than -5 or greater than 5 are binned into two end points.

of contemporaneous air pollution on remotely sensed yields in mining areas.

Panel A of Table 2 shows that relative to sides a mine that are downwind less frequently, sides of a mine that are downwind more frequently experience about a statistically significant increase in AOD after the "investment" phase of a mine begins. The magnitude of this effect is equivalent to roughly a 2% increase and is larger when focusing only on strongly treated cases: mine-months where the side that is downwind from the mine most frequently experiences at least 20 days of wind exposure from the mine in a month.

In contrast, Panel B shows no economically or statistically significant effect of contemporaneous air pollution on NDVI. To better understand these effects I estimate mine-specific DID estimates for the impact of mining on AOD and NDVI using the same set of mines used in the analysis above. Appendix Figure 18 plots the AOD DIDs against the NDVI DIDs for each mine, revealing a non-linear relationship between AOD and NDVI. While small increases in AOD increase yields, larger increases lead to yield reductions. This aligns with literature documenting that local pollution regimes, such as the extent to which the scattering versus absorbing effect of particulate matter dominates (Schiferl et al., 2018; Burney and Ramanathan, 2014), atmospheric conditions (Lobell et al., 2022) or intensity of pollution (Behrer and Wang, 2024), can moderate air pollution effects on yields.

6.2.2 Cumulative effect: Importantly, the DID and event study designs outlined above are limited in that they primarily estimate the effect of contemporaneous air pollution exposure on yields. This is because the mine-side and mine-date fixed effects ensure a comparison of sides of the same mine, within the same time period, which does not account for pollution exposure in prior time periods. Due to seasonality in winds, the side of a mine that is most frequently downwind varies within the year. This means that the downwind exposure of a given side in one month may not necessarily be similar to that side's exposure in prior months.

To overcome this limitation, I estimate a distributed lag version of Equation 2:

$$Y_{smt} = \sum_{p=0}^{P} \beta_p Exposure_{sm,t-p} + \sum_{p=0}^{P} \delta_p Exposure_{sm,t-p} \times Post_{m,t-p} + \alpha_{sm} + \lambda_{mt} + \mathbf{X'_{smt}\Gamma} + \epsilon_{imt} \quad (3)$$

where $Exposure_{sm,t-p}$ is downwind exposure on side s of mine m, p months before t. Downwind exposure is defined as the number of days a side is downwind from the mine, normalized so that one unit corresponds to being downwind for 30 days. $Post_{m,t-p}$ is an indicator for when production first began at mine m, lagged by p months. By interacting lagged wind exposure with a lagged indicator for mining activity, I ensure that only wind exposure in the months after a mine opening is included in the cumulative effect estimated for

Table 2: Relationship Between Industrial Mine Openings, Remotely-Sensed Air Pollution and Crop Yields

	(1)	(2)	(3)
	All	20+ days downwind	25+ days downwind
Panel A: Air Pollution (AOD)			
Wind × (Post - 3)	0.00403***	0.00443***	0.00601***
	(0.00147)	(0.00158)	(0.00212)
Mines	102	102	71
Observations	72,596	31,384	16,852
Mean AOD (t-3)	.22	.22	.22
Panel B: Yields (NDVI)			
Wind × (Post - 3)	0.00028	0.00061	0.00009
	(0.00146)	(0.00140)	(0.00158)
Mines	102	102	71
Observations	72,596	31,384	16,852
Mean NDVI (t-3)	.49	.49	.49
Mine-side FE	Yes	Yes	Yes
Mine-year-month FE	Yes	Yes	Yes
Weather	Yes	Yes	Yes

Notes: Each column reports the results of a linear regression. The unit of analysis is a mine-side-month. In Panel A, the dependent variable is mean aerosol optical depth (AOD) for one of the four sides (N,S, E, W) of a mine within a 60km buffer, in a given month. I derive AOD from the MODIS Aqua Level 2 Daily product (MYD04_3K). In Panel B, the dependent variable is mean normalized difference vegetation index (NDVI) for one of the four sides of a mine within a 60km buffer, in a given month. I derive remotely sensed yields from the MODIS Combined Terra and Aqua product (MCD43A4.061). *Wind* is defined as the share of days in a month that a side is downwind from the mine. *Post - 3* is equal to 1 after the investment phase of a mine, which occurs 3 years before the mine opened, 0 otherwise. All models include linear and quadratic controls for mean temperature, precipitation, vapor pressure, wet days, evapotranspiration and cloud cover, as well as mine-side fixed effects, mine-year-month fixed effects and mean wind speed. The sample includes the 102 mines for which non-missing AOD and NDVI are observed on all four sides of the mine for at least 4 months in each year, for at least 5 years pre and 5 years post-mine opening. Column 1 reports results from a model estimated over all mine months while Column 2 (3) subsets to only mine-months where the side that is downwind most frequently receives at least 20 (25) days of wind from the mine. Standard errors in parentheses are clustered by mine.

the post period. Aside from including the lagged exposure terms, Equation 3 is identical to the version of Equation 2 used to estimate the effect of contemporaneous air pollution on yields.

Before a mine opening, the cumulative effect of exposure to air pollution during the current month and the P months prior is $\sum_{p=0}^{P} \beta_p$. The cumulative effect of pollution exposure after a mine opening is $\sum_{p=0}^{P} \beta_p + \delta_p$. Thus, we can interpret $\sum_{p=0}^{P} \delta_p$ as the difference in the cumulative effect of exposure to air pollution during the current and P prior months, between the pre- and post-mine opening periods. Standard errors, p-values and confidence intervals for each lag structure are obtained by testing whether $\sum_{p=0}^{P} \delta_p = 0$, for each P. In my main specification, I include lagged months up to 2 years prior (24 months) so estimate 25 total coefficients (1 contemporaneous and 24 lagged).

To estimate the distributed lag model, I define a different sample of mines from the set used in the AOD and contemporaneous yield effect estimation shown in Table 2. First, I expand the set of mines to include those where AOD data may be missing or incomplete for certain months, as estimating the coefficients on the main effects and interaction terms for each of the lagged variables in Equation 3 is computationally demanding. Second, I restrict to only the set of mines where NDVI is observed on all four sides, for all 12 months of every year, for at least 5 years pre- and 5 years post-mine opening. This ensures that NDVI for each mine-side-month is observed for all lags included in the main specification, so that the composition of mines does not vary with the lag structure.

Pooling all seasons together, Figure 6 reports point estimates and 95 percent confidence intervals for the cumulative effect of lagged and concurrent pollution exposure, proxied by days of wind downwind from a mine. Along the horizontal axis, I increase the maximum response delay from $P = 1$ to $P = 24$ months, estimating a separate distributed lag model for each lag increment. On the vertical axis, I report the NDVI effect for a +30 day increase in wind sustained over the P + 1 preceding and concurrent months. For example, at P = 12 lagged months on the horizontal axis, Figure 6 reports the pre-post difference in cumulative effects, $\sum_{p=0}^{12} \delta_p$, of a sustained 30 day increase in wind from a mine on NDVI over the prior 12 months and the current month. For interpretation, I normalize the coefficients using mean NDVI from 3 years before a mine opens. Appendix Table 8 presents the coefficients, standard errors and p-values for each of the lag increments. On average, I find no effect of cumulative mining-induced air pollution on NDVI in downwind areas, both when pooling all months (Figure 6) as well as when estimating the model separately by growing and non-growing seasons (Appendix Figure 14).

Importantly, these average treatment effects for both contemporaneous and cumulative air pollution exposure from mines mask substantial heterogeneity. For instance, Appendix Table 14 Column 2 shows that downwind areas near open pit mines experience a small but statistically significant reductions in NDVI due to contemporaneous mining-induced air pollution exposure. This aligns with literature suggesting that

open pit mines tend to pollute more than underground mines (Sahu et al., 2015). Other dimensions of heterogeneity are discussed in more detail in Section 7.

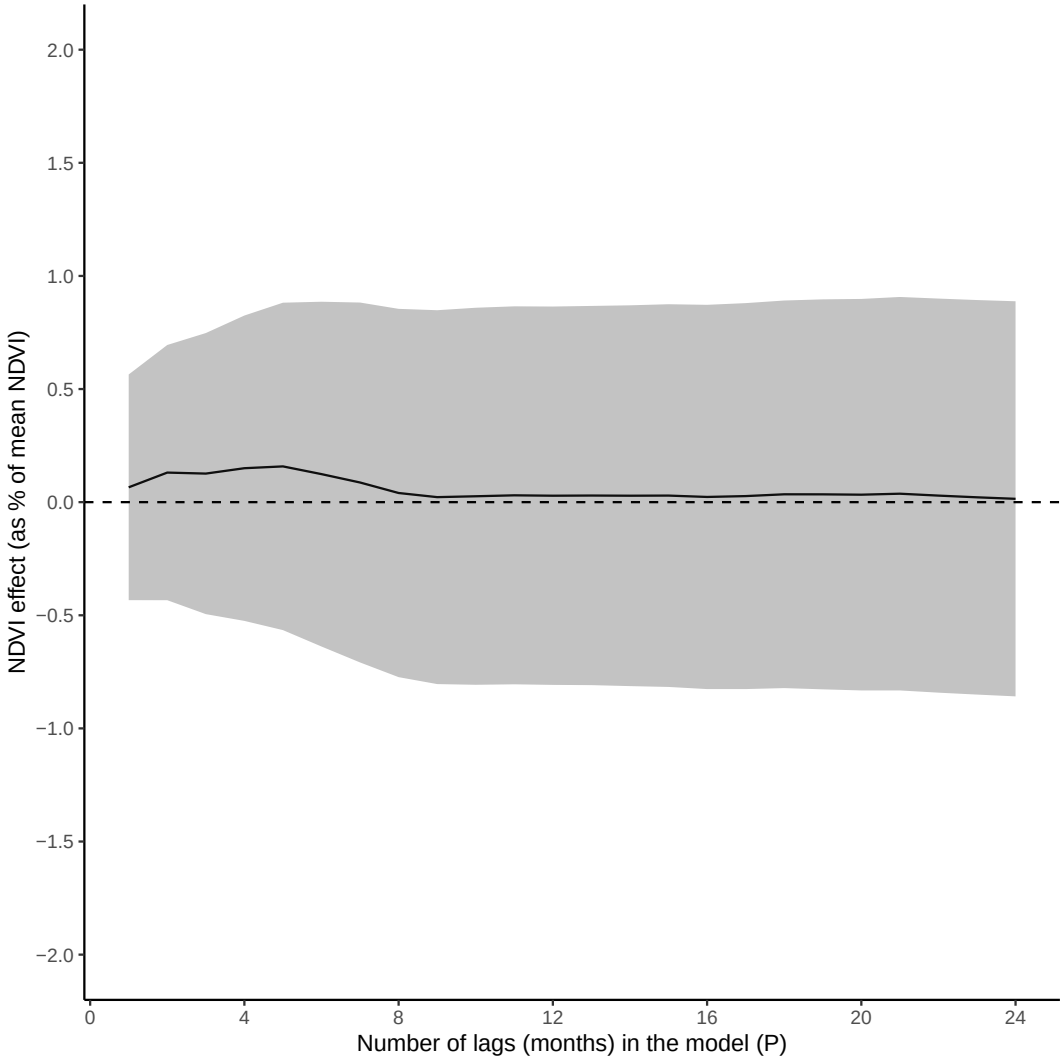

Figure 6: Effect of cumulative exposure to mining-induced air pollution on yields (NDVI)

The figure plots the cumulative impact on mean NDVI of +30 days of downwind exposure to air pollution from a mine in each of the concurrent and previous P months. Each value reports the point estimate and 95 percent CI on $\sum_{p=0}^{P} \delta_p$ from a different distributed lag model, as I increase P along the horizontal axis. $\sum_{p=0}^{P} \delta_p$ can be interpreted as the average difference in cumulative effects of air pollution on NDVI between the pre and post period, where the post period is defined as the time period after a mine opens. The sample includes the set of 219 mines with non-missing NDVI data observed on all 4 sides, for all 12 months in each year, for at least 5 years pre- and 5 years post-mine opening. Each distributed lag model is estimated using mine-side-months observed after the time period 3 years before a mine first opens, to ensure that lagged wind exposure is observed for at least 2 years before the current month. Each distributed lag model is estimated over months in all seasons.

6.3 Overall Effect of Mining on Agriculture: Pollution and Input Demand Shocks

To estimate an overall effect of industrial mining on agricultural output that captures both pollution and input demand shocks, I use the staggered openings of 307 mines across Sub-Saharan Africa in a DID that compares NDVI between areas near mines to areas further away, before and after a mine opening. I define *Exposure* in Equation 2 as a dummy variable equal to 1 for cropland within 20-km of a mine and 0 for the ring between the buffer of 100 km and the buffer of 150 km away from the mine, where these comparison groups are chosen based on a spatial lag model. Appendix Figure 13 shows that the largest reduction in NDVI occurs within 20 kilometers of a mine, with smaller reductions in NDVI occurring between 20-60 kilometers of a mine and no statistically significant effect at distances greater than 100 kilometers.[3]

To demonstrate that the 20-km radius captures the effects of both pollution and higher returns to inputs outside of agriculture, I show that it is consistent with spatial lag estimates for pollution and existing literature on market size. Appendix Figure 12 shows that AOD increases after mine openings are largest within 20-km of a mine, with smaller effects up to 60-km away, and no statistically distinguishable effects at greater distances. Additionally, empirical evidence on commuting distances in urban and rural Africa suggests that areas of 5, 10 or 15 km are likely integrated markets (Amoh-Gyimah and Aidoo, 2013; Kung et al., 2014; Shafer, 2000), while wealth increases (von der Goltz and Barnwal, 2019) and structural shifts out of agriculture into services Kotsadam and Tolonen (2016) due to mining are concentrated within 20-km of the mine.

Table 3 shows that on average, mine openings lead to a statistically significant decrease in NDVI of about 1.5% in areas near mines relative to areas further away, with slightly larger effects in the growing season. These results provide a helpful baseline against which the estimates generated by isolating the pollution effect can be compared to, which I do in Section 8.

7 Heterogeneous Treatment Effects

An advantage of my empirical design is that I observe a treatment and control area for every mine in my analysis, before and after the mine opening date, allowing me to estimate mine-specific DIDs. Each mine-specific DID is estimated by running a separate regression for each mine, where the outcome is NDVI on a side of the mine in a given month, with controls for weather and robust standard errors.[4]

[3] In the spatial lag model, the omitted category is the 100-150km ring.

[4] When estimating the overall specification separately for each mine, only the main effects and interaction terms of Near with Post, along with controls for weather, are included as covariates in the regression. I exclude the side and year-month fixed effects as they are correlated with the Near and Post dummies, respectively. The same holds true for the water pollution specification: I include only the main effects and interaction terms of Downstream with Post, along with controls for weather. For the air pollution specification, I include side fixed effects since downwind intensity varies across sides and time, as well as the main effects and interactions of downwind exposure and the Post dummy. Year-month fixed effects are excluded as they

Table 3: Average effect of mine openings on NDVI - mixing pollution and market channels

	(1) Pooled	(2) Growing only	(3) Non-growing only
Near × Post	-0.00638***	-0.00691***	-0.00589***
	(0.000980)	(0.000990)	(0.00108)
Number of mines	307	307	307
Obs.	168626	80582	88044
Mine-distance group FE	Yes	Yes	Yes
Mine-year-month FE	Yes	Yes	Yes
Weather	Yes	Yes	Yes
Mean NDVI (t-1)	.476	.543	.415

Each column reports the results of a linear regression. The unit of analysis is a mine-distance group-month. The dependent variable is mean NDVI within a given distance group, in a given month. *Near* is equal to 1 for area within the 20km buffer around the mine and equal to 0 for the area in the ring between the buffers of 100km and 150km. *Post* is equal to 1 after the mine opened, 0 otherwise. All models include linear and quadratic controls for mean temperature, precipitation, vapor pressure, wet days, evapotranspiration and cloud cover, as well as mine-distance group and mine-year-month fixed effects. The sample includes the 307 mines for which non-missing NDVI is observed in both distance groups for at least 4 months in each year, for at least 3 years pre- and 3 years post-mine opening. Column 1 reports results estimated by pooling months over all 5 seasons: planting, early growing, late growing, harvest and non-farm. Column 2 reports results estimated only over months in the early growing and late growing seasons and Column 3 reports results estimated only over months in non-growing seasons: planting, harvest and non-farm. Standard errors in parentheses are clustered by mine.

Figure 7 plots the distribution of mine-specific overall, water and air pollution effects on yields, revealing substantial heterogeneity across mines for all three types of analysis. Importantly, the variation in treatment effects observed in Figure 7 is much larger than what would be observed due to sampling variation alone, as shown in Appendix Figure 15. This illustrates the value of further investigation into what might be driving heterogeneity in treatment effects.

The top panel, which shows the distribution of overall effects of mining on local yields, highlights that the majority of treatment effects are negative and concentrated between 0 to -5%. Notably, there are several mines with positive treatment effects, which may be explained by some of the higher income from local demand booms being reinvested in improved inputs or technologies used in agriculture. By focusing on the set of mines closer to the sample used by Aragón and Rud (2016), I am able to recover treatment effects that are much closer in magnitude to the 15% decline in agricultural output that they find. Mine-specific estimates obtained from near versus far DID regressions reveal that treatment effects for all 19 Ghanaian mines in my sample are negative and can be as large as -11%.

7.1 Dimensions explored I investigate heterogeneity across four key dimensions: (1) governance/regulatory environments, (2) local economic factors, (3) mine characteristics and (4) local environmental conditions. Weak local governance and economic indicators have been linked to environmental degradation (Burgess et al., 2012; Cust et al., 2023; Grossman and Krueger, 1995), while pre-existing environmental conditions have been shown to moderate pollution effects (Hutchinson, 1984; Oksanen and Kontunen-Soppela, 2021;

are collinear with the Post dummy.

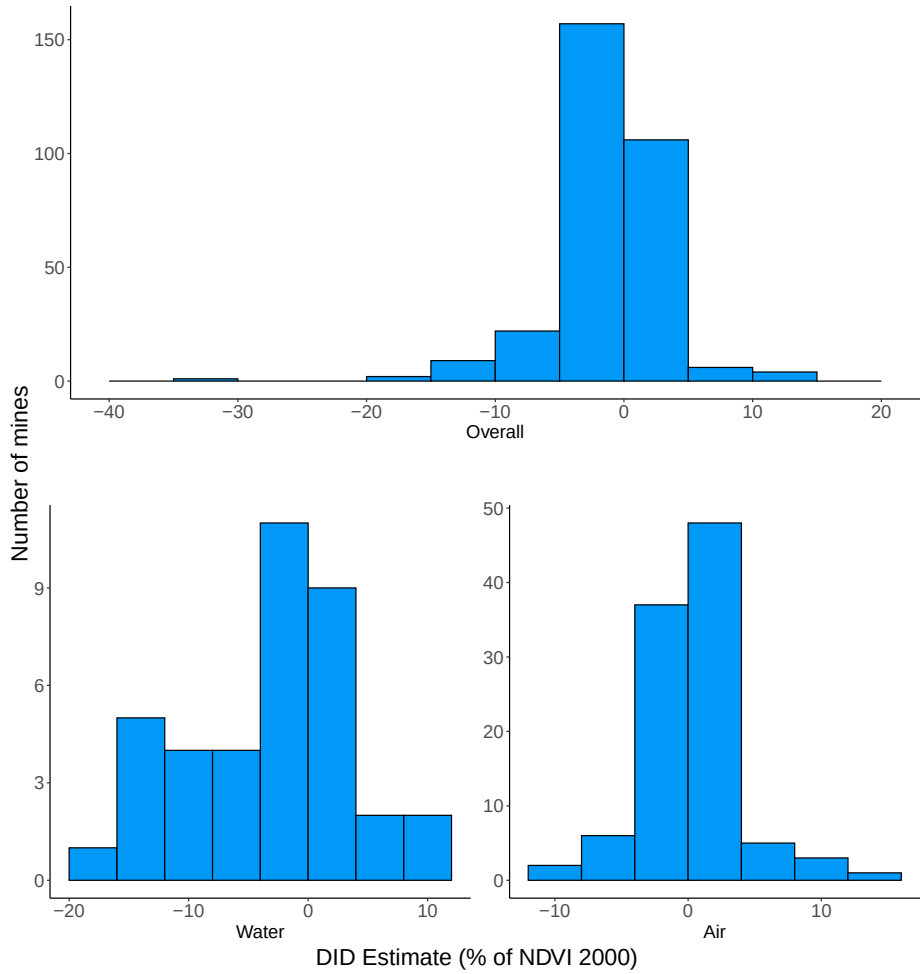

Figure 7: Distribution of Mine-Specific DIDs

The figure plots three histograms showing the distributions of mine-specific DID estimates. Each DID is estimated from a separate regression of a version of Equation 2 for each mine. The top panel shows the distribution of mine-specific near vs. far DID estimates obtained by regressing NDVI for a side (near or far) on the Near dummy interacted with the Post dummy, along with the main effects and controls for weather. The bottom left panel shows the distribution of mine-specific downstream vs. upstream DID estimates obtained by regressing NDVI on a side (downstream or upstream) on the Downstream dummy interacted with the Post dummy, along with the main effects and controls for weather. The bottom right panel shows the distribution of mine-specific continuous wind intensity DID estimates obtained by regressing NDVI on a side (N, S, E, W) on the share of days a side is downwind from the mine in a given month interacted with the Post dummy, the main effects, side fixed effects and weather controls. Across all regressions, I estimate robust standard errors. Each mine-specific DID estimate is normalized by mean NDVI in 2000 for that mine, then multiplied by 100 to represent the magnitude of the effect relative to the mean of the outcome at the start of the panel in percent terms.

Lobell et al., 2022). Similarly, mine characteristics like extraction or processing methods, as well as the type of commodity extracted, have been associated with varying degrees of pollution (Dudka and Adriano, 1997; Sahu et al., 2015; Cox et al., 2022).

First, to explore the influence of governance on the local effects of mines, I use the country-level World Bank World Governance Indicators from 2000, the year prior to the start of my NDVI panel. I focus on five main indicators: control of corruption, government effectiveness, regulatory quality, rule of law, and voice and accountability.[5] Each indicator is based on data from multiple sources, such as surveys, commercial business information providers or non-governmental organizations. The underlying data is aggregated into a score for each of the five indicators, in units of a standard normal distribution from approximately -2.5 to 2.5. Additionally, I examine whether the mine country's membership to the Extractive Industries Transparency Initiative (EITI), designed to promote transparency and accountability in national resource extraction, can explain variation in the local effects of mining.

Second, I examine heterogeneity in effects on NDVI according to the following mine-specific characteristics: extraction method (open pit or non-open pit), commodity type, mine age, distance to the nearest neighboring mine and distance to the nearest town. Finally, I test for heterogeneity by the following economic factors: mineral rents as share of GDP in 2000, GDP in 2000 and population in 2000, as well as the following local environmental conditions: initial air and water pollution exposure (AOD in 2003 and NDTI in 2000), initial cropland productivity (NDVI in 2000) and precipitation in 2000.

7.2 Machine-learning heterogeneity analysis I train a machine learning model to predict mine-specific treatment effects given a large number of mine and country-level characteristics. Though my approach is similar in spirit to the methods of Wager and Athey (2017), who use causal forests to uncover meaningful dimensions of heterogeneity by estimating conditional average treatment effects, it differs in a key way. While I can estimate unit specific treatment effects, Wager and Athey (2017) do not observe control units that can be matched to each treated unit, so use random forests to non-parametrically define groups of treated and control units that are comparable across observables. The exercise I conduct is complementary to their work, as it can be applied to other settings where unit-specific treatment effects can be estimated and treatment effect heterogeneity may be explained by a variety of factors, such that the researcher wishes to demonstrate that statistically significant findings from "standard" heterogeneity analysis are not spurious.

[5]Kaufmann et al. (2010) define each of the indicators as follows: "Control of corruption captures perceptions of the extent to which public power is exercised for private gain, as well as elite or private capture. Government effectiveness captures perceptions of the quality of public services, the quality of policy formulation/implementation and the credibility of government commitment to these policies. Regulatory quality captures perceptions of the government's ability to formulate and implement sound regulations. Rule of law captures perceptions of the extent to which agents have confidence in and follow the rules of society, such as the quality of contract enforcement, property rights, the police and the courts. Voice and accountability captures perceptions of the extent to which a country's citizens can exercise rights to freedom of expression, freedom of association and free media, as well as whether they can participate in free and fair elections."

To find the best performing model, I train a collection of ML models to predict treatment effects based on 8 algorithms, including linear regression models such as OLS, LASSO, Ridge and elastic net, as well as non-parametric models such as random forest, a classification and regression tree (CART) and a bagged CART. The unit of observation is a mine, where the outcome of interest is the mine-specific DID estimate for the overall effect of mining activity on yields. The main predictors include the World Bank governance indicators, a dummy variable for whether the mine is located in an EITI-member country, GDP in 2000 for the mining area, population in 2000 for the mining area, mineral rents as a share of GDP in 2000, dummy variables for mine type, dummy variables for commodity extracted, dummy variables indicating the country that the mine is located in, distance to the nearest mine and distance to the nearest town.[6]

The ML models are applied to the set of 304 mines for which DID estimates are generated and non-missing data for all predictors is available, using 75% of observations for training and the remaining 25% for validation. Model hyperparameters are selected based on a grid search. Out-of-sample accuracy measures are calculated using 5 fold cross-validation of the training dataset. I therefore obtain 25 hyperparameters combinations for each model (except for OLS), totaling 176 different ML models. Out of all models trained, the random forest model is identified as the "best" performing model, defined by lowest out-of-sample root mean-squared error (RMSE).[7] I re-train this random forest model over the entire dataset of DIDs and obtain a measure of variable importance for each predictor.

To estimate variable importance, I construct a measure of "permutation importance," which leverages the out-of-bag samples for each tree. First, the RMSE on the out-of-bag sample is calculated. Then, the values of the predictor of interest in the out-of-bag-sample are randomly shuffled, keeping the values of all other predictors the same. Finally, the decrease in RMSE on the shuffled data is measured. Larger reductions in RMSE after permutation imply that a variable is more "important" in contributing to predictive capacity. I opt for permutation importance over Gini impurity, a common alternative measure of variable importance for random forest models, as impurity-based importance is less reliable in settings where many predictors are dummies.

Figure 8 plots the ten most important variables from the best random forest model, based on permutation importance. One striking result is that all five country-level governance indicators are determined to be highly important. This aligns with a burgeoning literature documenting the role of public governance on local environmental footprints. Lipscomb and Mobarak (2017) demonstrate the importance of governance in explaining the spatial patterns of water pollution in Brazil, Burgess et al. (2012) find that weak governance

[6] For the country, commodity and mine type variables, which are originally categorical, categories accounting for less than 5% of the observations in the sample are binned together into "other."

[7] The best random forest model is grown on 1000 trees, with 1 predictor randomly sampled at each split and a minimum of 15 data points in each node required for additional splitting to occur.

worsens deforestation in Indonesia and Cust et al. (2023) show that forest loss near oil wells is lower in countries with stronger public governance. Additionally, my results fit into a larger economics literature that highlights the importance of well-functioning political institutions for economic development (Acemoglu et al., 2005).

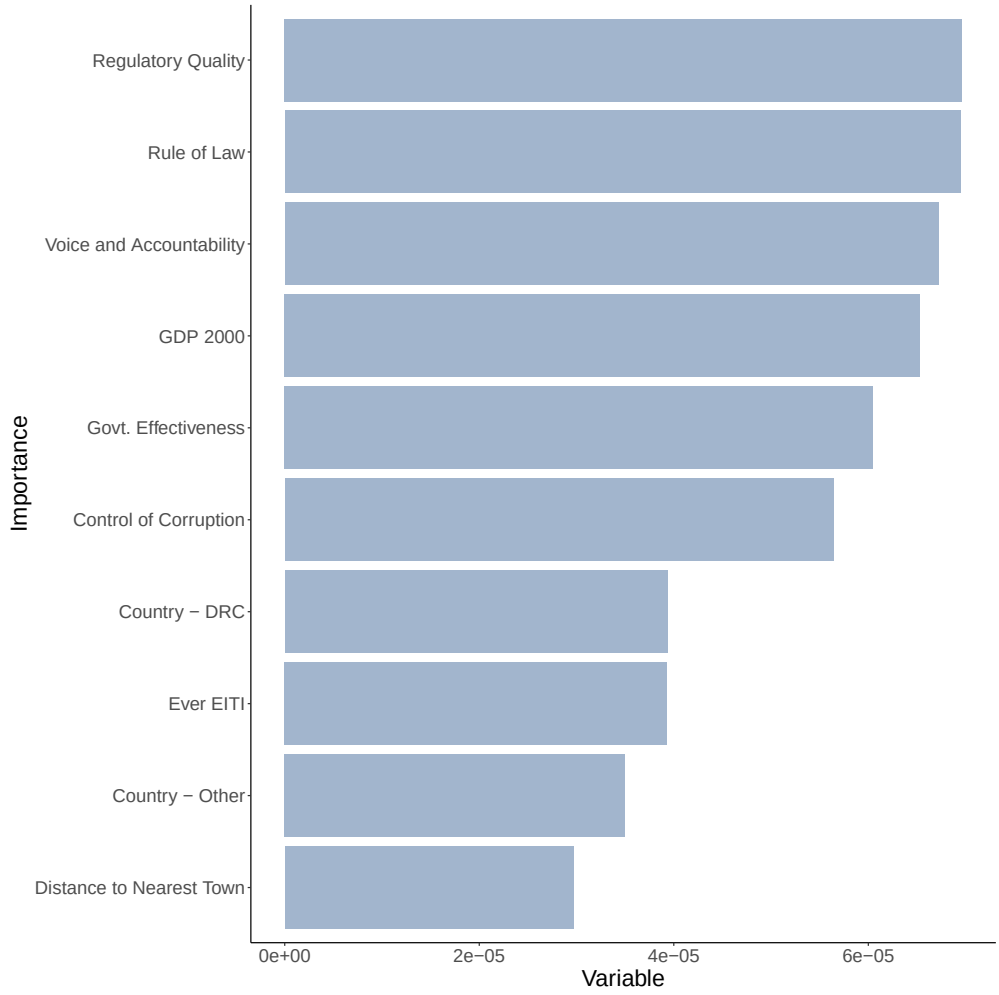

Figure 8: Most Important Variables from Best Random Forest Model

The figure plots the permutation importance measure for each of the top 10 most important variables selected by the best random forest model for predicting mine-specific DID estimates. The unit of observation in the random forest model is a mine, where the outcome variable is a mine-specific near vs. far DID estimate for the effect of mining on NDVI. The set of predictors includes mine type dummies, commodity dummies, the World Bank governance indicators, GDP in 2000, population in 2000, mineral rents as a share of GDP in 2000, precipitation in 2000, distance to the nearest neighboring mine, distance to the nearest town, country dummies, and membership to the EITI. The best random forest model is grown on 1000 trees, with 1 predictor randomly sampled at each split and a minimum of 15 data points in each node required for additional splitting to occur.

7.3 Standard heterogeneity analysis With the ML findings in mind, I then turn to implementing the standard practice for heterogeneity analysis used in the economics literature. To estimate heterogeneity in the overall effect of mining on yields, as well as the effects of water pollution and contemporaneous air pollution, I interact the *Post* and *Exposure* variables from Equation 2 with the dimension of interest. In addition, to test for heterogeneity in the effects of prolonged exposure to air pollution on yields, I estimate Equation 3 separately over different sub-groups defined by the dimension of heterogeneity of interest.

Across the overall, water and air pollution analyses, weak governance and regulatory environments are revealed to be important drivers of the negative effect of mining activity on yields. Table 4 reveals that reductions in NDVI are 2-3 times larger in areas near mines located in countries with below median performance on governance indicators. The triple interactions are negative, sizable in magnitude and statistically significant across all governance indicators, with the largest negative effect detected for mines located in countries that are not members of the EITI.

Table 4: Heterogeneous Treatment Effects - Governance and Regulatory Environments

	Base	Not EITI Member	Control of Corruption	Govt. Effectiveness	Rule of Law	Reg. Quality	Voice and Accountability
Near × Post	-0.00638***	-0.00278***	-0.00405***	-0.00415***	-0.00429***	-0.00426***	-0.00436***
	(0.000980)	(0.000900)	(0.00108)	(0.00110)	(0.00108)	(0.00108)	(0.00111)
Near × Post × Z		-0.00714***	-0.00499**	-0.00472**	-0.00453**	-0.00472**	-0.00414**
		(0.00191)	(0.00198)	(0.00197)	(0.00200)	(0.00202)	(0.00196)
Number of mines	307	307	307	307	307	307	307
Mean NDVI (t-1)	.476	.476	.476	.476	.476	.476	.476

Standard errors in parentheses

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

Similar patterns in heterogeneous treatment effects are uncovered from the air and water regressions. Appendix Table 16 shows that the triple interactions with the governance and EITI indicators from the water pollution regressions are overwhelmingly negative, though not statistically significant, likely due to the small sample size. For air pollution, Appendix Figures 21 - 23, which show the heterogeneous treatment effects from the air pollution distributed lag model, reveal a consistent pattern of reductions in NDVI occurring only for mines with below median government effectiveness, regulatory quality and accountability. Most strikingly, Figure 9, shown below, highlights how statistically significant reductions in NDVI occur for mines located in countries that are not members of the EITI, after about 8 months of exposure to air pollution from mines.

However, the standard heterogeneity analysis identifies several other statistically significant dimensions of heterogeneity that not considered important predictors in the ML model. For instance, I find that exposure to pollution before a mine opens influences the extent to which mining activity impacts local agriculture. Across both the air and water pollution analyses, NDVI reductions are higher for cropland in areas initially exposed

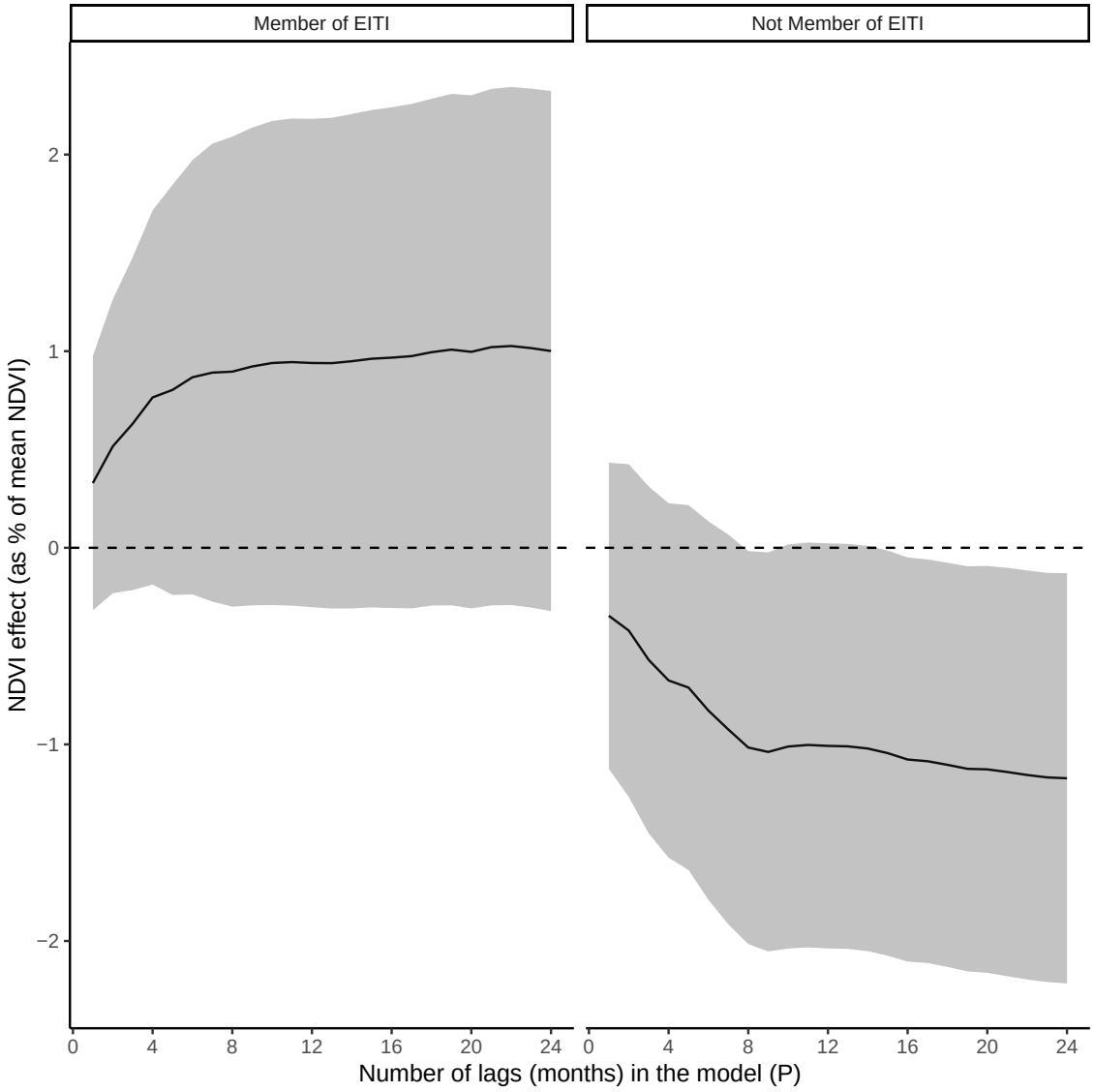

Figure 9: Cumulative effect of mining air pollution on NDVI - heterogeneity by EITI membership

Each panel in the figure plots the cumulative impact on mean NDVI of +30 days of downwind exposure to air pollution from a mine in each of the concurrent and previous P months. The left panel plots the cumulative impact for mines located in countries that are EITI members, while the right panel plots it for mines in non-member countries. Each value reports the point estimate and 95 percent CI on $\sum_{p=0}^{P} \delta_p$ from a different distributed lag model, as I increase P along the horizontal axis. $\sum_{p=0}^{P} \delta_p$ can be interpreted as the average difference in cumulative effects of mining air pollution on NDVI between the pre and post period, where the post period is defined as the time period after a mine opens. The sample includes only the set of mines with non-missing NDVI data observed on all 4 sides, for all 12 months in each year, for at least 5 years pre- and 5 years post-mine opening. Each distributed lag model is estimated using mine-side-months observed after the time period 3 years before a mine first opens, to ensure that lagged wind exposure and NDVI are observed for at least 2 years before the current month.

to lower pollution levels. Appendix Table 14 Column 2 highlights how NDVI reductions are about twice as large in cropland near rivers that had below median levels of initial turbidity, with the triple interaction being statistically significant. Similarly, Appendix Figure 34 illustrates how statistically significant NDVI reductions occur only for cropland initially exposed to lower levels of AOD. These effects could be explained by plant adaptation (Hutchinson, 1984; Oksanen and Kontunen-Soppela, 2021) or changing behaviors of farmers in response to pollution (Magesa et al., 2023; Villamayor-Tomas et al., 2024). Furthermore, while the estimates on the triple interactions with mineral rents as a share of 2000 GDP and distance to the nearest neighboring mine are both large and statistically significant in the standard heterogeneity analysis, both variables are included in the 10 least important variables determined by the ML model (shown in Appendix Figure 16).

This comparison between the ML and the standard heterogeneity analyses is highlighted in Figure 10, which plots the ML importance measures for each dimension of heterogeneity used as a predictor in the ML model on the x-axis and the p-value from the triple interaction with this dimension from the standard heterogeneity analysis on the y-axis. We can observe that only the governance indicators are consistently shown to be statistically significant in conventional heterogeneity analysis, as well as considered highly important by the ML model for explaining variation in the mine-specific treatment effects (bottom-right quadrant). In contrast, several commodity and country dummies have statistically significant triple interactions for the standard analysis, but are deemed less important by the ML model (bottom-left). This exercise highlights the value of ML for disciplining standard heterogeneity analysis in settings where unit-specific treatment effects can be estimated. In Appendix Figure 17, I also show that the main conclusions drawn from the ML analysis discussed above are robust to estimating a classification model, where the outcome is a dummy variable equal to one if the mine-specific DID estimate is negative. While the ranking of the most important variables slightly changes, four of the five country-level governance indicators still rank within the top five most important predictors.

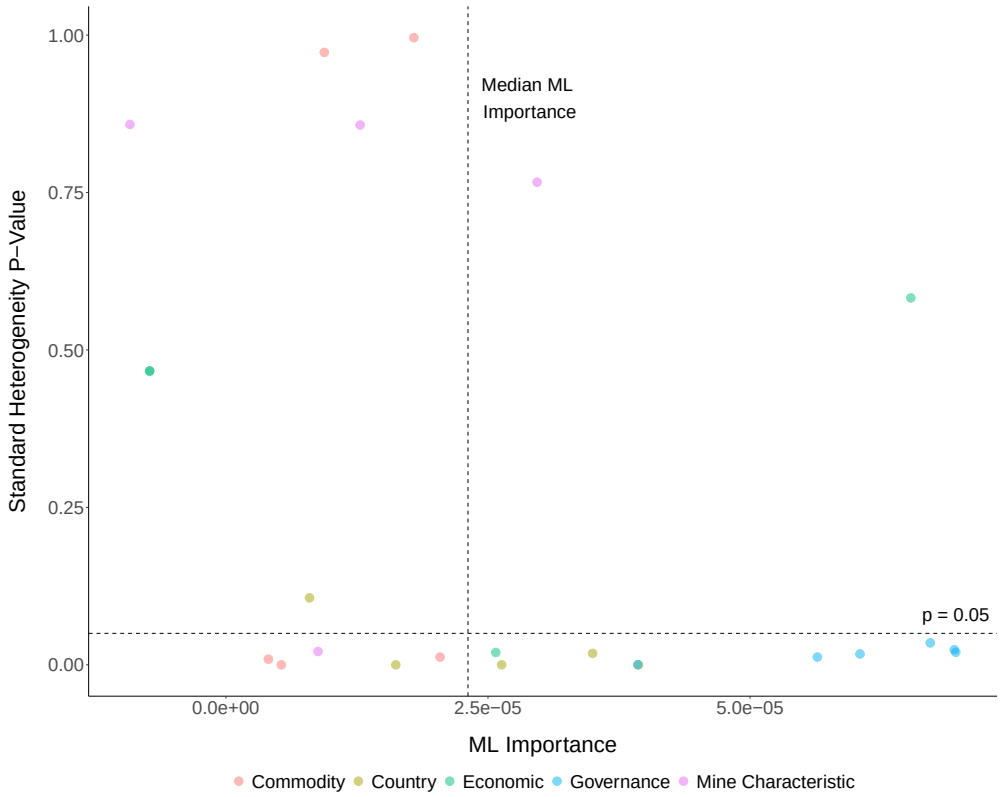

Figure 10: ML versus Standard Heterogeneity Analysis

On the x-axis, the figure plots the measure of ML permutation importance for each predictor used in the best random forest model predicting mine-specific near vs. far DID estimates. The y-axis shows the p-value for the triple interaction of the binarized version of each predictor with the Near and Post dummies in the standard heterogeneity analysis. Each point is color-coded according to the category of predictor it belongs to. The governance category includes the five World Bank Governance Indicators and a dummy variable for membership to the Extractive Industries Transparency Initiative (EITI). The economic category includes GDP in 2000, population in 2000 and mineral rents as a share of GDP in 2000. Mine characteristics include mine type, mineral processing method, distance to the nearest neighboring mine, distance to the nearest town and mine age. The commodity and country categories include dummies for the main commodity extracted at the mine (gold , copper, platinum, diamonds, other) and the country that the mine is located in (Ghana, South Africa, Zambia, DRC, other), respectively.

8 Quantifying Results

In this section, I ask two important questions to quantify the effect of mining on yields. First, what share of the overall effect of industrial mining on local agricultural output can be explained by pollution alone? Second, how can we map the estimated effects in NDVI units to effects on actual yields in kilograms per hectare?

8.1 Estimating Share of Effect on Yields Driven By Pollution I conduct a back-of-the-envelope calculation to estimate the share of the overall effect of mining on yields driven by pollution externalities. I take the ratio of the pollution-only treatment effects estimated with the air and water pollution DIDs, scaled to account for the area in the near buffer exposed to pollution, to the overall effect estimated with the near vs. far DID:

$$\text{Pollution Share} = \frac{\beta_{water} \times A_{downstream} + \beta_{air} \times A_{downwind}}{\beta_{overall}} \tag{4}$$

where $A_{downstream}$ and $A_{downwind}$ give the fraction of the land in the 20-km buffer that are located downstream or downwind from a mine, respectively. β_{water} is the average treatment effect on yields from the water pollution DID, β_{air} the effect from the air pollution DID and $\beta_{overall}$ the overall effect of mining from the near versus far DID.[8] I find that approximately 44% of the overall effect of mining on NDVI can be explained by air and water pollution externalities. When interpreting this share, it is important to keep in mind several caveats. First, the heterogeneity analysis suggests that this share likely varies substantially across mines. Second, wealth increases driven by higher returns to inputs outside of agriculture generated by mining could offset some of the negative impacts of pollution for areas near mines. For example, increased wealth could be invested into inputs, such as fertilizer, or more productive technologies that may counteract the effects of mining-induced pollution on crop health. Alternatively, increased wealth could be used to treat health issues arising from exposure to polluted air or water.

8.2 Translating NDVI Effects to Yield Effects To map NDVI effects to yield effects, I estimate the correlation between actual maize yields from plots of smallholder farmers across Africa and

[8]From the distributed lag results in Table 8, we see that at most NDVI falls by 0.0007 units in downwind areas due to cumulative and contemporaneous mining-induced air pollution, so I use $\beta_{air} = -0.0007$. I scale this estimate by $A_{downwind} = 0.25$, to reflect the fact that on average, 25% of the 20-km near buffer around mine is downwind most frequently in a given month. Table 1 Column 2 shows that on average, NDVI is 0.0177 units lower in downstream areas during the growing seasons, so I use $\beta_{water} = -0.0177$. I re-scale this treatment effect by $A_{downstream} = 0.17$, which is the 25^{th} percentile of the share of the 20 km buffer that is at a lower elevation than the mine, and so could be affected by water flow from the mine, across all mines within 1 km of a river. For $\beta_{overall}$ I use the overall effect of industrial mining on growing season NDVI of -0.00691, from Column 2 of Table 3.

NDVI local to these plots. Data on maize yields, defined in kilograms of maize produced per hectare, are obtained from Aramburu-Merlos et al. (2024). This dataset covers over 13,000 smallholder farms across 7 countries: Burundi, Uganda, Nigeria, Tanzania, Kenya, Rwanda and Zambia. It is a repeated-cross section, where measurements are taken over the 2016-2022 time period from different plots located in similar areas. The data from Aramburu-Merlos et al. (2024) was collected by One Acre Fund, an NGO that provides smallholder farmers with access to credit, training, crop insurance and farming supplies. Maize yields for each farmer at the time of harvest were measured in two randomly spaced boxes of 36 square meters, avoiding field edges. While farm size is not provided in the public data of Aramburu-Merlos et al. (2024), African agriculture is dominated by small farms, typically defined as less than 1 acre (Carletto et al., 2015).

To link plot-level yields to NDVI, I define a grid of 1km x 1km cells over the study areas in Aramburu-Merlos et al. (2024). Each plot is linked to a grid-cell using plot GPS coordinates. I opt for the 1km x 1km grid cell to limit the influence of inaccuracies in plot coordinates (Jin et al., 2017). While I do not observe the same plots over time, most grid cells can be observed for at least 2 different years. For each grid cell, I calculate mean NDVI across pixels falling within the cell for each season from 2016-2022. Mean seasonal NDVI at the cell-year level is then linked to the average of observed yields across all plots in the corresponding cell-year.

I run the following regression to estimate the relationship between remotely-sensed yields and actual yields measured from One Acre Fund cropcuts, separately for each season:

$$Yield_{ct} = \alpha_c + \lambda_t + \beta_{NDVI} NDVI_{ct} + \mathbf{X'_{ct}}\mathbf{\Gamma} + \epsilon_{ct} \tag{5}$$

where $Yield_{ct}$ is mean maize yields in kg/ha across all plots in cell c in year t and $NDVI_{ct}$ is mean NDVI in one of the five seasons. I include grid-cell and year fixed effects, as well as cell-level controls for the average number of maize growing degree days, average precipitation over the maize season and average temperature over the maize season from Aramburu-Merlos et al. (2024), across all plots within a cell-year. I also include cell-level weather variables, such as mean cloud cover and evapotranspiration from Harris et al. (2014). $NDVI_{ct}$ is scaled so that β_{NDVI} reflects the effect of a 0.01 unit increase in NDVI on maize yields. Inclusion of weather controls and fixed effects help limit omitted variable bias due to NDVI detection errors.

Table 5 shows that on average, a 0.01 unit increase in NDVI is associated with a statistically significant 52.10 kg/ha increase in maize yields during the early growing season, with smaller increases in other seasons. This is supported by existing literature, which also finds that NDVI is a strong predictor of yields specifically during the early stages of growing (Panek and Gozdowski, 2021). Linking this result to the overall effect of mining on NDVI of -0.00691 suggests that on average, industrial mining reduces actual yields for a

smallholder farmer by about 36 kilograms per hectare. This corresponds to about a 1.2 % decrease in actual yields relative to the mean of 3000 kg/ha, which is a fifth of the magnitude of the reduction in yields caused by an additional dust storm in Iran (Birjandi-Feriz and Yousefi, 2017) or the reduction in yields caused by the Acid Rain Program in the USA (Sanders and Barreca, 2022). Importantly, these other estimates were generated using administrative data on yields, so may not not suffer from attenuation bias due to measurement error common to remotely sensed variables (Proctor et al., 2023). Furthermore, this 1.2% decrease represents an average effect: yield reductions in areas with poor governance and regulatory quality are over 3 times larger, which would be on par with the magnitudes estimated in other contexts.

Table 5: Relationship between Cell-level NDVI and Cell-level Yields, cell-size = 1000m

	(1)	(2)	(3)	(4)	(5)
	Planting	Early Growing	Late Growing	Harvest	Nonfarm
Mean NDVI	11.83	52.10**	20.25	24.43	9.937
	(16.42)	(22.04)	(21.11)	(17.69)	(14.62)
Year FE	Yes	Yes	Yes	Yes	Yes
Grid cell FE	Yes	Yes	Yes	Yes	Yes
Include weather controls	Yes	Yes	Yes	Yes	Yes
Mean Yields	3005.643	3005.643	3005.643	3005.643	3005.643
Obs.	1127	1153	1153	1141	1153
R-sq	0.631	0.638	0.632	0.634	0.632

Each column reports the results of an OLS regression. The unit of observation is a cell-year. The dependent variable is mean plot-level yields in kg/ha across all plots falling within a cell, in a given year. *Mean NDVI* is the average NDVI in the cell of size 1000m containing the plot, across days in a particular season during the year that maize on the plot was harvested. The columns indicate the season for which NDVI is calculated during the year the plot was harvested. *Mean NDVI* is scaled so that one unit represents a 0.01 increase in NDVI. Each regression includes linear controls for cell-level averages of growing degree days, temperature and precipitation during the maize season across plots falling within the cell, as well as cell-level controls for mean temperature, precipitation, vapor pressure, cloud cover, evapotranspiration and wet days. Year and cell-level fixed effects are included in the regressions.

Table 5 also shows that the NDVI model at the cell-level explains over 50% of the variation in plot-level yields, which is similar to the R^2 found in other work linking the One Acre Fund data to remotely sensed yields (Burke and Lobell, 2017; Jin et al., 2017). Conditional on observing a sufficiently high R^2, in Appendix 10.13 I discuss whether the estimated relationship between NDVI and local yields from Equation 5 may be biased in my setting, even after controlling for fixed effects and weather. Furthermore, Appendix Tables 20 - 22 demonstrate robustness to varying the size of the grid cell used in the quantification exercise, consistently showing that NDVI during the growing season is most strongly correlated with observed plot-level yields.

9 Discussion

This paper quantifies the extent to which industrial mining affects local agricultural output through pollution externalities versus local demand shocks that raise the returns to inputs outside of agriculture. To isolate

the effects of pollution, I link mine geolocations to remotely sensed measures of pollution and crop yields, employing spatial difference-in-difference designs that leverage variation in pollution exposure around mines driven by topography or weather—factors that are plausibly exogenous to local market conditions.

A key contribution of this paper is documenting increases in mining-induced air and water pollution, then estimating the subsequent externalities on local yields. Although mining activities generate meaningful increases water and air pollution in downstream and downwind areas respectively, the corresponding declines in yields are modest. Specifically, mining leads to statistically significant reductions in yields of approximately 3-4% through the water pollution channel, while there is no observable impact on yields from air pollution.

Measurement error in remotely sensed yield data may explain why my estimated treatment effects are smaller than those obtained using self-reported yields from survey data (Aragón and Rud, 2016). Mean-reverting measurement error in remotely sensed dependent variables can attenuate treatment effects (Proctor et al., 2023), suggesting that my estimates of mining pollution externalities may represent a conservative lower bound of the true effects. Unfortunately, the lack of ground-truth data on plot-level yields near mining areas limits my ability to correct for this type of measurement error.

I benchmark my estimates of the effects of mining-induced air and water pollution on yields against an estimate of the overall impact of mining, which includes both pollution externalities and the increased returns to inputs outside of agriculture. I find that pollution externalities account for nearly half of the overall effect on yields, indicating that declines in local agricultural output are not solely driven by positive structural transformation. Furthermore, average effects mask significant heterogeneity across mines, with both machine learning and standard heterogeneity analyses revealing that poor governance and regulatory environments are associated with larger reductions in yields.

My findings raise important distributive concerns regarding the clean energy transition. As more developed countries promote widespread clean energy use to internalize negative pollution externalities within their own borders, they may inadvertently shift the burden of these externalities to less developed countries, where the necessary inputs for these green technologies are sourced. This concern is borne out in my data: the extraction of critical inputs for clean energy negatively impacts agricultural output in African countries, particularly those with weak governance. The persistence of mining pollution externalities, in the absence of sufficiently large agglomeration effects, may lead to a situation where local economic booms from natural resource extraction fade once a mine closes (Black et al., 2005; Kotsadam and Tolonen, 2016), while the effects of mining pollution on local agriculture may linger for decades (Akcil and Koldas, 2006). If pollution has driven the agricultural sector to a lower, less productive equilibrium, a shift back into agriculture following a mine's closure could result in lower living standards for mining communities compared to those existing before the mine opened. Given that many of the poorest households rely on agriculture, policymakers must

consider the justice and equity implications of the clean energy transition going forward.

References

Daron Acemoglu, Simon Johnson, and James Robinson. Institutions as the fundamental cause of long-run growth. *The Handbook of Economic Growth*, 2005. URL https://www.wcfia.harvard.edu/sites/projects.iq.harvard.edu/files/wcfia/files/894_jr_handbook9sj.pdf.

Ata Akcil and Soner Koldas. Acid Mine Drainage (AMD): causes, treatment and case studies. *Journal of Cleaner Production*, 14(12):1139–1145, January 2006. ISSN 0959-6526. doi: 10.1016/j.jclepro.2004.09.006. URL https://www.sciencedirect.com/science/article/pii/S0959652605000600.

Hunt Allcott and Daniel Keniston. Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America. *The Review of Economic Studies*, 85(2):695–731, April 2018. ISSN 0034-6527. doi: 10.1093/restud/rdx042. URL https://doi.org/10.1093/restud/rdx042.

Richard Amoh-Gyimah and Eric Nimako Aidoo. Mode of transport to work by government employees in the Kumasi metropolis, Ghana. *Journal of Transport Geography*, 31:35–43, July 2013. ISSN 0966-6923. doi: 10.1016/j.jtrangeo.2013.05.008. URL https://www.sciencedirect.com/science/article/pii/S0966692313000859.

Donald W. K. Andrews. Tests for Parameter Instability and Structural Change With Unknown Change Point. *Econometrica*, 61(4):821–856, 1993. ISSN 0012-9682. doi: 10.2307/2951764. URL https://www.jstor.org/stable/2951764. Publisher: [Wiley, Econometric Society].

Donald W. K. Andrews and Werner Ploberger. Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative. *Econometrica*, 62(6):1383–1414, 1994. ISSN 0012-9682. doi: 10.2307/2951753. URL https://www.jstor.org/stable/2951753. Publisher: [Wiley, Econometric Society].

Fernando M. Aragón and Juan Pablo Rud. Natural Resources and Local Communities: Evidence from a Peruvian Gold Mine. *American Economic Journal: Economic Policy*, 5(2):1–25, May 2013. ISSN 1945-7731. doi: 10.1257/pol.5.2.1. URL https://www.aeaweb.org/articles?id=10.1257/pol.5.2.1.

Fernando M. Aragón and Juan Pablo Rud. Polluting Industries and Agricultural Productivity: Evidence from Mining in Ghana. *The Economic Journal*, 126(597):1980–2011, 2016. ISSN 1468-0297. doi: 10.1111/ecoj.12244. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ecoj.12244. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecoj.12244.

Fernando Aramburu-Merlos, Fatima A. M. Tenorio, Nester Mashingaidze, Alex Sananka, Stephen Aston, Jonathan J. Ojeda, and Patricio Grassini. Adopting yield-improving practices to meet maize demand in Sub-Saharan Africa without cropland expansion. *Nature Communications*, 15(1):4492, May 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-48859-0. URL https://www.nature.com/articles/s41467-024-48859-0. Publisher: Nature Publishing Group.

Sam Asher and Paul Novosad. Rent-Seeking and Criminal Politicians: Evidence from Mining Booms | The Review of Economics and Statistics | MIT Press. *The Review of Economics and Statistics*, 2023. URL https://direct.mit.edu/rest/article/105/1/20/102838/Rent-Seeking-and-Criminal-Politicians-Evidence.

Sebastian Axbard, Anja Benshaul-Tolonen, and Jonas Poulsen. Natural resource wealth and crime: The role of international price shocks and public policy. *Journal of Environmental Economics and Management*, 110:102527, October 2021. ISSN 0095-0696. doi: 10.1016/j.jeem.2021.102527. URL https://www.sciencedirect.com/science/article/pii/S0095069621000905.

Oscar Barriga-Cabanillas, Mira Korb, and Travis Lybbert. Time is (not) Money: The decay of machine learning models to predict wealth using cellphone data. *Working Paper*, 2024.

A. Patrick Behrer and Sherrie Wang. Current benefits of wildfire smoke for yields in the US Midwest may dissipate by 2050. *Environmental Research Letters*, 19(8):084010, July 2024. ISSN 1748-9326. doi: 10.1088/1748-9326/ad5458. URL https://dx.doi.org/10.1088/1748-9326/ad5458. Publisher: IOP Publishing.

Anja Benshaul-Tolonen. Local Industrial Shocks and Infant Mortality. *The Economic Journal*, n/a(n/a), 2020. ISSN 1468-0297. doi: 10.1111/ecoj.12625. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ecoj.12625. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecoj.12625.

Nicolas Berman, Mathieu Couttenier, Dominic Rohner, and Mathias Thoenig. This Mine is Mine! How Minerals Fuel Conflicts in Africa. *American Economic Review*, 107(6):1564–1610, June 2017. ISSN 0002-8282. doi: 10.1257/aer.20150774. URL https://pubs.aeaweb.org/doi/10.1257/aer.20150774.

Sumanta Bid and Giyasuddin Siddique. Identification of seasonal variation of water turbidity using NDTI method in Panchet Hill Dam, India. *Modeling Earth Systems and Environment*, 5(4):1179–1200, December 2019. ISSN 2363-6211. doi: 10.1007/s40808-019-00609-8. URL https://doi.org/10.1007/s40808-019-00609-8.

Maliheh Birjandi-Feriz and Kowsar Yousefi. When the Dust Settles: Productivity and Economic Losses Following Dust Storms, November 2017. URL https://papers.ssrn.com/abstract=3230265.

Dan Black, Terra McKinnish, and Seth Sanders. The Economic Impact of the Coal Boom and Bust. *The Economic Journal*, 115(503):449–476, April 2005. ISSN 0013-0133. doi: 10.1111/j.1468-0297.2005.00996.x. URL https://doi.org/10.1111/j.1468-0297.2005.00996.x.

Leonardo Bonilla Mejía. Mining and human capital accumulation: Evidence from the Colombian gold rush. *Journal of Development Economics*, 145:102471, June 2020. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2020.102471. URL https://www.sciencedirect.com/science/article/pii/S0304387820300468.

John Bound and Alan B. Krueger. The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right? *Journal of Labor Economics*, 9(1):1–24, 1991. ISSN 0734-306X. URL https://www.jstor.org/stable/2535111. Publisher: [University of Chicago Press, Society of Labor Economists, NORC at the University of Chicago].

Robin Burgess, Matthew Hansen, Benjamin A. Olken, Peter Potapov, and Stefanie Sieber. The Political Economy of Deforestation in the Tropics*. *The Quarterly Journal of Economics*, 127(4):1707–1754, November 2012. ISSN 0033-5533. doi: 10.1093/qje/qjs034. URL https://doi.org/10.1093/qje/qjs034.

Marshall Burke and David B. Lobell. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proceedings of the National Academy of Sciences*, 114(9):2189–2194, February 2017. doi: 10.1073/pnas.1616919114. URL https://www.pnas.org/doi/abs/10.1073/pnas.1616919114. Publisher: Proceedings of the National Academy of Sciences.

Jennifer Burney and V. Ramanathan. Recent climate and air pollution impacts on indian agriculture. *PNAS*, 111, 2014. doi: 10.1073/pnas.1317275111. URL https://www.pnas.org/doi/10.1073/pnas.1317275111.

Calogero Carletto, Sydney Gourlay, and Paul Winters. From Guesstimates to GPStimates: Land Area Measurement and Implications for Agricultural Analysis. *Journal of African Economies*, 24(5):593–628, November 2015. ISSN 0963-8024, 1464-3723. doi: 10.1093/jae/ejv011. URL https://academic.oup.com/jae/article-lookup/doi/10.1093/jae/ejv011.

Raymond Caroll, David Ruppert, Leonard Stefanski, and Ciprian M. Crainiceanu. *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition.* Chapman and Hall/CRC, 2006.

Francesco Caselli and Guy Michaels. Do Oil Windfalls Improve Living Standards? Evidence from Brazil. *American Economic Journal: Applied Economics*, 5(1):208–238, January 2013. ISSN 1945-7782. doi: 10.1257/app.5.1.208. URL https://www.aeaweb.org/articles?id=10.1257/app.5.1.208.

Shuai Chen, Paulina Oliva, and Peng Zhang. The effect of air pollution on migration: Evidence from China. *Journal of Development Economics*, 156:102833, May 2022. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2022.102833. URL https://www.sciencedirect.com/science/article/pii/S0304387822000153.

Wenjie Chen, Paola Ganum, Athene Laws, Hamza Mighri, Balazs Stadler, Nico Valckx, and David Zeledon. Digging for opportunity: Harnessing sub-saharan africa's wealth in critical minerals. *The International Monetary Fund, Regional Economic Outlook: Sub-Saharan Africa*, 2024. URL https://www.imf.org/en/Publications/REO/SSA/Issues/2024/04/19/regional-economic-outlook-for-sub-saharan-africa-april-2024.

Burhan U. Choudhury, Akbar Malang, Richard Webster, Kamal P. Mohapatra, Bibhash C. Verma, Manoj Kumar, Anup Das, Mokidul Islam, and Samarendra Hazarika. Acid drainage from coal mining: Effect on paddy soil and productivity of rice. *The Science of the Total Environment*, 583:344–351, April 2017. ISSN 1879-1026. doi: 10.1016/j.scitotenv.2017.01.074.

Seydou Coulibaly and Abdramane Camara. The end of tax incentives in mining? Tax policy and mining foreign direct investment in Africa. *African Development Review*, 34(S1):S177–S194, 2022. ISSN 1467-8268. doi: 10.1111/1467-8268.12651. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8268.12651. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-8268.12651.

Benjamin Cox, Sally Innis, Nadja C. Kunz, and John Steen. The mining industry as a net beneficiary of a global tax on carbon emissions. *Communications Earth & Environment*, 3(1):1–8, February 2022. ISSN 2662-4435. doi: 10.1038/s43247-022-00346-4. URL https://www.nature.com/articles/s43247-022-00346-4. Publisher: Nature Publishing Group.

James Cust and Albert Zeufack. Africa's Resource Future. 2023. URL https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099080123145011993/p16722906c03ca09409ace06cb32991395b.

James Cust, Torfinn Harding, Hanna Krings, and Alexis Rivera-Ballesteros. Public governance versus corporate governance: Evidence from oil drilling in forests. *Journal of Development Economics*, 163:103070, June 2023. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2023.103070. URL https://www.sciencedirect.com/science/article/pii/S0304387823000251.

Stanislaw Dudka and Domy C. Adriano. Environmental Impacts of Metal Ore Mining and Processing: A Review. *Journal of Environmental Quality*, 26(3):590–602, 1997. ISSN 1537-2537. doi: 10.2134/jeq1997.00472425002600030003x. URL https:

//onlinelibrary.wiley.com/doi/abs/10.2134/jeq1997.00472425002600030003x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2134/jeq1997.00472425002600030003x.

Fabian Ekert, John Juneau, and Michael Peters. Sprouting Cities: How Rural America Industrialized. *AEA Papers and Proceedings*, 113:87–92, May 2023. doi: 10.1257/pandp.20231075. URL https://ideas.repec.org/a/aea/apandp/v113y2023p87-92.html.

Kyle Emerick. Trading frictions in Indian village economies. *Journal of Development Economics*, 132:32–56, May 2018. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2017.12.010. URL https://www.sciencedirect.com/science/article/pii/S0304387817301293.

Zhaozhong Feng and Kazuhiko Kobayashi. Assessing the impacts of current and future concentrations of surface ozone on crop yield with meta-analysis. *Atmospheric Environment*, 43(8):1510–1519, March 2009. ISSN 1352-2310. doi: 10.1016/j.atmosenv.2008.11.033. URL https://www.sciencedirect.com/science/article/pii/S1352231008010972.

J. Gardelle, P. Hiernaux, L. Kergoat, and M. Grippa. Less rain, more water in ponds: a remote sensing study of the dynamics of surface waters from 1950 to present in pastoral sahel (gourma region, mali). *Hydrology and Earth System Sciences*, 14(2):309–324, 2010. doi: 10.5194/hess-14-309-2010. URL https://hess.copernicus.org/articles/14/309/2010/.

Nicolas Gendron-Carrier, Marco Gonzalez-Navarro, Stefano Polloni, and Matthew A. Turner. Subways and Urban Air Pollution. *American Economic Journal: Applied Economics*, 14(1):164–196, January 2022. ISSN 1945-7782. doi: 10.1257/app.20180168. URL https://www.aeaweb.org/articles?id=10.1257/app.20180168.

Lutz Goedde, Amandla Ooko-Ombaka, and Gillian Pais. Winning in Africa's agricultural market, 2019. URL https://www.mckinsey.com/industries/agriculture/our-insights/winning-in-africas-agricultural-market.

Joshua Graff Zivin and Matthew Neidell. The Impact of Pollution on Worker Productivity. *American Economic Review*, 102(7):3652–3673, December 2012. ISSN 0002-8282. doi: 10.1257/aer.102.7.3652. URL https://www.aeaweb.org/articles?id=10.1257/aer.102.7.3652.

Gene M. Grossman and Alan B. Krueger. Economic Growth and the Environment. *The Quarterly Journal of Economics*, 110(2):353–377, 1995. ISSN 0033-5533. doi: 10.2307/2118443. URL https://www.jstor.org/stable/2118443. Publisher: Oxford University Press.

Emilio Gutiérrez and Kensuke Teshima. Abatement expenditures, technology choice, and environmental performance: Evidence from firm responses to import competition in Mexico. *Journal of Development Economics*, 133:264–274, July 2018. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2017.11.004. URL https://www.sciencedirect.com/science/article/pii/S0304387817301037.

I. Harris, P.D Jones, T.J. Osborn, , and D.H. Lister. Updated high-resolution grids of monthly climatic observations—cru ts3.26 dataset. *International Journal of Climatology*, 2014.

Jiaxiu He, Haoming Liu, and Alberto Salvo. Severe Air Pollution and Labor Productivity: Evidence from Industrial Towns in China. *American Economic Review*, 11, 2019. doi: 10.1257/app.20170286. URL https://www.aeaweb.org/articles?id=10.1257/app.20170286.

Matias Heino, Pekka Kinnunen, Weston Anderson, Deepak K. Ray, Michael J. Puma, Olli Varis, Stefan Siebert, and Matti Kummu. Increased probability of hot and dry weather extremes during the growing season threatens global crop yields. *Scientific Reports*, 13(1):3583, March 2023. ISSN 2045-2322. doi: 10.1038/s41598-023-29378-2. URL https://www.nature.com/articles/s41598-023-29378-2. Publisher: Nature Publishing Group.

Qing Huang, Victoria Wenxin Xie, and Wei You. Resource Rents, Urbanization, and Structural Transformation. *SSRN Electronic Journal*, 2023. ISSN 1556-5068. doi: 10.2139/ssrn.4543449. URL https://www.ssrn.com/abstract=4543449.

Thomas C. Hutchinson. Adaptation of Plants to Atmospheric Pollutants. In *Ciba Foundation Symposium 102 - Origins and Development of Adaptation*, pages 52–72. John Wiley & Sons, Ltd, 1984. ISBN 978-0-470-72083-7. doi: 10.1002/9780470720837.ch5. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470720837.ch5. Section: 5 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470720837.ch5.

Md. Monirul Islam and Kimiteru Sado. Analyses of aster and spectroradiometer data with in situ measurements for turbidity and transparency study of lake abashiri. *International journal of geoinformatics*, 2, 2006. URL https://api.semanticscholar.org/CorpusID:126540631.

Alexander James. The resource curse: A statistical mirage? *Journal of Development Economics*, 114:55–63, May 2015. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2014.10.006. URL https://www.sciencedirect.com/science/article/pii/S0304387814001199.

Zhenong Jin, George Azzari, Marshall Burke, Stephen Aston, and David B. Lobell. Mapping Smallholder Yield Heterogeneity at Multiple Scales in Eastern Africa. *Remote Sensing*, 9(9):931, September 2017.

ISSN 2072-4292. doi: 10.3390/rs9090931. URL https://www.mdpi.com/2072-4292/9/9/931. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

Daniel Kaufmann, Aart Kraay, and Massimo Mastruzzi. The Worldwide Governance Indicators: Methodology and Analytical Issues, September 2010. URL https://papers.ssrn.com/abstract=1682130.

Ruth H. Keogh, Pamela A. Shaw, Paul Gustafson, Raymond J. Carroll, Veronika Deffner, Kevin W. Dodd, Helmut Küchenhoff, Janet A. Tooze, Michael P. Wallace, Victor Kipnis, and Laurence S. Freedman. STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 1-Basic theory and simple methods of adjustment. *Statistics in Medicine*, 39(16): 2197–2231, July 2020. ISSN 1097-0258. doi: 10.1002/sim.8532.

Andreas Kotsadam and Anja Tolonen. African Mining, Gender, and Local Employment. *World Development*, 83:325–339, July 2016. ISSN 0305-750X. doi: 10.1016/j.worlddev.2016.01.007. URL https://www.sciencedirect.com/science/article/pii/S0305750X1600005X.

Kevin S. Kung, Kael Greco, Stanislav Sobolevsky, and Carlo Ratti. Exploring Universal Patterns in Human Home-Work Commuting from Mobile Phone Data. *PLOS ONE*, 9(6):e96180, June 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0096180. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0096180. Publisher: Public Library of Science.

J. P. Lacaux, Y. M. Tourre, C. Vignolles, J. A. Ndione, and M. Lafaye. Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. *Remote Sensing of Environment*, 106(1):66–74, January 2007. ISSN 0034-4257. doi: 10.1016/j.rse.2006.07.012. URL https://www.sciencedirect.com/science/article/pii/S0034425706002811.

D. J. Lapworth, D. C. W. Nkhuwa, J. Okotto-Okotto, S. Pedley, M. E. Stuart, M. N. Tijani, and J. Wright. Urban groundwater quality in sub-Saharan Africa: current status and implications for water security and public health. *Hydrogeology Journal*, 25(4):1093–1116, 2017. ISSN 1431-2174. doi: 10.1007/s10040-016-1516-6. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6991975/.

R.C. Levy, S. Mattoo, L.A. Munchak, L.A. Remer, A.M. Sayer, F. Patadia, and N.C. Hsu. The collection 6 modis aerosol products over land and ocean. *Atmospheric Measurement Techniques*, 2013.

Molly Lipscomb and Ahmed Mushfiq Mobarak. Decentralization and Pollution Spillovers: Evidence from the Re-drawing of County Borders in Brazil*. *The Review of Economic Studies*, 84(1):464–502, January 2017. ISSN 0034-6527. doi: 10.1093/restud/rdw023. URL https://doi.org/10.1093/restud/rdw023.

Peng Liu, Qiumei Wu, Wenyou Hu, Kang Tian, Biao Huang, and Yongcun Zhao. Effects of atmospheric deposition on heavy metals accumulation in agricultural soils: Evidence from field monitoring and Pb isotope analysis. *Environmental Pollution*, 330:121740, August 2023. ISSN 0269-7491. doi: 10.1016/j.envpol.2023.121740. URL https://www.sciencedirect.com/science/article/pii/S026974912300742X.

Xiang Liu and Ankur R. Desai. Significant Reductions in Crop Yields From Air Pollution and Heat Stress in the United States. *Earth's Future*, 9(8):e2021EF002000, 2021. ISSN 2328-4277. doi: 10.1029/2021EF002000. URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2021EF002000. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021EF002000.

Davis Lobell, Stefania Di Tomasso, and Jennifer Burney. Globally ubiquitous negative effects of nitrogen dioxide on crop growth. *Science Advances*, 2022. doi: 10.1126/sciadv.abm9909. URL https://www.science.org/doi/10.1126/sciadv.abm9909.

Bahati A. Magesa, Geetha Mohan, Hirotaka Matsuda, Indrek Melts, Mohamed Kefi, and Kensuke Fukushi. Understanding the farmers' choices and adoption of adaptation strategies, and plans to climate change impact in Africa: A systematic review. *Climate Services*, 30:100362, April 2023. ISSN 2405-8807. doi: 10.1016/j.cliser.2023.100362. URL https://www.sciencedirect.com/science/article/pii/S2405880723000237.

Luis R. Martínez. Natural Resource Rents, Local Taxes, and Government Performance: Evidence from Colombia. *The Review of Economics and Statistics*, pages 1–28, April 2023. ISSN 0034-6535. doi: 10.1162/rest_a_01324. URL https://doi.org/10.1162/rest_a_01324.

Tom Myers. Acid mine drainage risks – A modeling approach to siting mine facilities in Northern Minnesota USA. *Journal of Hydrology*, 533:277–290, February 2016. ISSN 00221694. doi: 10.1016/j.jhydrol.2015.12.020. URL https://linkinghub.elsevier.com/retrieve/pii/S0022169415009683.

Elina Oksanen and Sari Kontunen-Soppela. Plants have different strategies to defend against air pollutants. *Current Opinion in Environmental Science & Health*, 19:100222, February 2021. ISSN 2468-5844. doi: 10.1016/j.coesh.2020.10.010. URL https://www.sciencedirect.com/science/article/pii/S2468584420300738.

Ewa Panek and Dariusz Gozdowski. Relationship between MODIS Derived NDVI and Yield of Cereals for Selected European Countries. *Agronomy*, 11(2):340, February 2021. ISSN 2073-4395. doi: 10.3390/agronomy11020340. URL https://www.mdpi.com/2073-4395/11/2/340. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

Aditya Kumar Patra, Sneha Gautam, and Prashant Kumar. Emissions and human health impact of particulate matter from surface mining operation—A review. *Environmental Technology & Innovation*, 5: 233–249, April 2016. ISSN 2352-1864. doi: 10.1016/j.eti.2016.04.002. URL https://www.sciencedirect.com/science/article/pii/S2352186416300153.

E. Petavratzi, S. Kingman, and I. Lowndes. Particulates from mining operations: A review of sources, effects and regulations. *Minerals Engineering*, 18(12):1183–1199, October 2005. ISSN 0892-6875. doi: 10.1016/j.mineng.2005.06.017. URL https://www.sciencedirect.com/science/article/pii/S0892687505002050.

Jonathan Proctor, Tamma Carleton, and Sandy Sum. Parameter Recovery Using Remotely Sensed Variables. *NBER Working Paper*, 2023. URL http://www.nber.org/papers/w30861.

Nathan Ratledge, Gabe Cadamuro, Brandon de la Cuesta, Matthieu Stigler, and Marshall Burke. Using machine learning to assess the livelihood impact of electricity access. *Nature*, 611(7936):491–495, November 2022. ISSN 1476-4687. doi: 10.1038/s41586-022-05322-8. URL https://www.nature.com/articles/s41586-022-05322-8. Publisher: Nature Publishing Group.

Tomás Rau, Sergio Urzúa, and Loreto Reyes. Early Exposure to Hazardous Waste and Academic Achievement: Evidence from a Case of Environmental Negligence. *Journal of the Association of Environmental and Resource Economists*, 2(4):527–563, December 2015. ISSN 2333-5955. doi: 10.1086/683112. URL https://www.journals.uchicago.edu/doi/full/10.1086/683112. Publisher: The University of Chicago Press.

Jason Russ, Esha Zaveri, Sebastien Desbureaux, Richard Damania, and Aude-Sophie Rodella. The impact of water quality on GDP growth: Evidence from around the world. *Water Security*, 17:100130, December 2022. ISSN 2468-3124. doi: 10.1016/j.wasec.2022.100130. URL https://www.sciencedirect.com/science/article/pii/S2468312422000219.

Jeffrey D. Sachs and Andrew M. Warner. The curse of natural resources. *European Economic Review*, 45(4):827–838, May 2001. ISSN 0014-2921. doi: 10.1016/S0014-2921(01)00125-8. URL https://www.sciencedirect.com/science/article/pii/S0014292101001258.

William J. Sacks, Delphine Deryng, Jonathan A. Foley, and Navin Ramankutty. Crop planting dates: an analysis of global patterns. *Global Ecology and Biogeography*, 19(5):607–620, 2010. ISSN 1466-8238. doi: 10.1111/j.1466-8238.2010.00551.x. URL https://onlinelibrary.wiley.com/doi/abs/

10.1111/j.1466-8238.2010.00551.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1466-8238.2010.00551.x.

H. B. Sahu, N. Prakash, and S. Jayanthu. Underground Mining for Meeting Environmental Concerns – A Strategic Approach for Sustainable Mining in Future. *Procedia Earth and Planetary Science*, 11:232–241, January 2015. ISSN 1878-5220. doi: 10.1016/j.proeps.2015.06.030. URL https://www.sciencedirect.com/science/article/pii/S1878522015000818.

Nicholas J. Sanders and Alan I. Barreca. Adaptation to Environmental Change: Agriculture and the Unexpected Incidence of the Acid Rain Program. *American Economic Journal: Economic Policy*, 14(1): 373–401, February 2022. ISSN 1945-7731. doi: 10.1257/pol.20190060. URL https://www.aeaweb.org/articles?id=10.1257/pol.20190060.

Luke D. Schiferl, Colette L. Heald, and David Kelly. Resource and physiological constraints on global crop production enhancements from atmospheric particulate matter and nitrogen deposition. *Biogeosciences*, 15(14):4301–4315, July 2018. ISSN 1726-4170. doi: 10.5194/bg-15-4301-2018. URL https://bg.copernicus.org/articles/15/4301/2018/. Publisher: Copernicus GmbH.

Andreas Shafer. Regularities in Travel Demand: An International Perspective. *Journal of Transportation and Statistics*, 2000. URL https://rosap.ntl.bts.gov/view/dot/4809.

Alka Sharma, Sushma Panigrahy, T. S. Singh, J. G. Patel, and Hemant Tanwar. Wetland information system using remote sensing and gis in himach pradesh, india. *Asian Journal of Geoinformatics*, 14, 2015. URL https://api.semanticscholar.org/CorpusID:128677550.

Sandip Sukhtankar. Does firm ownership structure matter? Evidence from sugar mills in India. *Journal of Development Economics*, 122:46–62, September 2016. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2016.04.002. URL https://www.sciencedirect.com/science/article/pii/S0304387816300232.

Sergio Villamayor-Tomas, Alexander Bisaro, Kevin Moull, Amaia Albizua, Isabel Mank, Jochen Hinkel, Gerald Leppert, and Martin Noltze. Developing countries can adapt to climate change effectively using nature-based solutions. *Communications Earth & Environment*, 5(1):1–11, April 2024. ISSN 2662-4435. doi: 10.1038/s43247-024-01356-0. URL https://www.nature.com/articles/s43247-024-01356-0. Publisher: Nature Publishing Group.

Holger Virro, Giuseppe Amatulli, Alexander Kmoch, Longzhu Shen, and Evelyn Uuemaa. Grqa: Global river water quality archive. *Earth System Science Data*, 13, 2021. doi: 10.5194/essd-13-5483-2021.

Jan von der Goltz and Prabhat Barnwal. Mines: The local wealth and health effects of mineral mining in developing countries. *Journal of Development Economics*, 139:1–16, June 2019. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2018.05.005. URL https://www.sciencedirect.com/science/article/pii/S0304387818304875.

Stefan Wager and Susan Athey. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests, July 2017. URL http://arxiv.org/abs/1510.04342. arXiv:1510.04342.

Robert G. Wetzel. *Limnology: Lake and River Ecosystems*. Elsevier Science, 2001.

Tingting Xie and Ye Yuan. Go with the wind: Spatial impacts of environmental regulations on economic activities in China. *Journal of Development Economics*, 164:103139, September 2023. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2023.103139. URL https://www.sciencedirect.com/science/article/pii/S0304387823000949.

Eric Yongchen Zou. Unwatched Pollution: The Effect of Intermittent Monitoring on Air Quality. *American Economic Review*, 111, 2021. doi: 10.1257/aer.20181346. URL https://www.aeaweb.org/articles?id=10.1257/aer.20181346.

10 Appendix

10.1 Mine Locations Figure 11 plots the set of 307 mines from the S&P dataset used to estimate the overall effect of mining on local yields. Each yellow point indicates the centroid of a mining operation, while the countries shaded in blue are those that contain at least one mine used in the analysis.

Figure 11: Map of mines used in analysis

10.2 Data Construction

10.2.1 Water Pollution I select only mines that are within 1 kilometer of a river. I use the HydroRIVERS data to identify rivers located in Africa. This dataset is a vectorized line network of rivers that have a catchment area of at least 10 km^2, an average river flow of at least 0.1 m^3/sec, or both. Each observation in this dataset is a river segment, where segment end points are defined by splitting rivers or streams at nodes where they fork. Each mine within 1 km of a river in this network is "snapped" to the closest river segment,

then the upstream and downstream river segments from the mine river segment are identified. Mines that are snapped to river segment end points, for which an upstream or downstream segment is missing, are excluded from the analysis. Parts of the upstream or downstream river segments that extend beyond a buffer of radius 10-km from the snapped mine point are removed. The upstream and downstream segments are buffered by 1-kilometer on either side of the line to identify land area around the rivers that would likely be affected by acid mine drainage either through seepage or irrigation. The "side" of a mine in the water pollution analysis refers to the upstream or downstream buffer.

Next, I calculate the satellite-based measure of water pollution, NDTI, over river water pixels. I first identify the precise shape of the upstream and downstream river segments near mines using a water mask based on the normalized difference water index (NDWI). This index is commonly used to detect water bodies in satellite images and is calculated from the green and near-infrared (NIR) bands.[9] I generate the NDWI water mask from a mosaic of cloud-free Sentinel-2 pixels from 2020-2023 at a 15-m resolution for the entire area within the 10-km buffer around the snapped mine point. Pixels identified as water based on a positive NDWI value are marked as river pixels.

Then, I calculate monthly mean NDTI for each river pixel according to the following formula: $NDTI = \frac{\text{Red}-\text{Green}}{\text{Red}+\text{Green}}$. I use Landsat 7 data at a 30-m resolution, as it is the highest resolution data available that covers the time period of 2000-2024.[10] I mask out pixels marked as "low quality" due to cloud cover, shadow or other atmospheric interference. Finally, I average monthly NDTI across water pixels within the upstream or downstream river segments, for each mine, to create mine-month-side level measures of river NDTI.

10.2.2 Air Pollution I rely on global daily rasters of MODIS AOD generated by Gendron-Carrier et al. (2022). The authors provide daily tifs from July 4, 2002 until August 31, 2018, for the Aqua satellite and February 24, 2000 until July 31, 2020 for the Terra satellite. Given MODIS data availability, the panel used in the air pollution analysis covers 2003 - 2017. From the daily AOD rasters, I construct mean AOD for each mine-side-month by averaging all pixel-days with non-missing AOD readings that fall within each side of a mine in a given month. I calculate this average monthly AOD only using data from the Aqua satellite, as the Terra satellite suffers from more missing values due to satellite detection errors during my time period of interest.

The algorithm detecting AOD can perform poorly in certain cases, which generates missing values. Levy et al. (2013) find that the algorithm performs poorly over light surfaces, such as desert or snowy regions. To

[9] $$NDWI = \frac{\text{Green} - \text{NIR}}{\text{Green} + \text{NIR}}$$

[10] The Sentinel-2 satellite only launched in 2015

address this limitation, I exclude mining areas that are adjacent to the desert, such as parts of Burkina Faso and Mali that neighbor the Sahel. In addition, the MODIS instruments can only detect AOD on cloud-free days, so missing data in certain regions or on certain days tend to reflect high levels of cloud cover. This results in seasonality of missingness in the MODIS data; I manage this issue by using mine-year-month fixed effects in my model, which account for mine-specific monthly variations in weather patterns correlated with satellite errors. Furthermore, in regressions with AOD as the outcome variable, I control for the number of non-missing pixels used to construct mean monthly AOD. I opt for a linear control of non-missing pixel count over interpolating missing AOD as interpolation algorithms may introduce measurement error that is difficult to interpret.

To calculate my measure of downwind exposure for each mine-month-side, I use the MERRA-2 daily aggregates of hourly time-averaged U and V wind vector components for heights of 2 meters and 50 meters, at a 0.5 degree $\times$ 0.625 degree resolution. I first determine the direction (in degrees from due North) that the wind is blowing from the mine centroid, on each day from 2000-2022.[11] I then classify this direction into one of the four 90-degree sides defined by the cardinal directions (North, East, South, West). Finally, for each mine-month, I count the number of days that the wind blows from the mine into each of the four sides, as well as determine the average monthly wind speed experienced by each side. Downwind exposure of a given side is defined as the share of days in a month that the side is downwind from the mine and ranges from 0 to 1.

10.2.3 NDVI I proxy for crop yields using the normalized difference vegetation index (NDVI). NDVI is strongly correlated with crop productivity and final yields, especially during the growing season (Panek and Gozdowski, 2021). I construct a measure of pixel-level daily mean NDVI from 2000 - 2022 using the MODIS MCD43A4 Version 6.1, combined Terra and Aqua product at a 500-m resolution. NDVI is calculated from the MODIS data using readings of light reflected in the near-infrared and red spectrum.[12]

When aggregating NDVI from the pixel-day to the mine-side-month level, I first mask out pixels that are not classified as cropland according to the Global Food Security Support Analysis Data (GFSAD) cropland extent map, available at a 1-km resolution. This helps ensure that changes in NDVI reflect changes in crop yields rather than changes in forest cover, for instance due to deforestation that might occur during construction of the mine site.

My measure of cropland NDVI may be affected by different sources of measurement error. To begin, the coarse 1-km resolution of the cropland extent data may constrain the ability to detect cropland of very small

[11] $Direction = 90 - atan2(v, u) \times 180/\pi$
[12] NDVI = $\frac{NIR-RED}{NIR+RED}$

farms. Additionally, NDVI may also suffer from errors due to factors like cloud cover, surface reflectance, canopy thickness, the level of atmospheric aerosols and satellite sensor errors. I discuss how measurement error in NDVI might affect estimates from causal inference in Appendix 10.13. Despite these potential sources of measurement error, Table 5 shows that NDVI is still highly predictive of plot-level yields in Sub-Saharan Africa, though the magnitude of the correlation may be attenuated. I construct my final measures of cropland NDVI for the treatment and control areas by averaging NDVI across all cropland pixel-days within each mine-side-month, within the relevant buffers.

An emerging literature uses machine learning algorithms to predict crop yields based on vegetation indices, like NDVI, and other covariates (Burke and Lobell (2017), Jin et al. (2017)). Unfortunately, most of the mining areas in my analysis are not located near sites where ground truth measures of yields, like cropcuts, are available. Furthermore, high quality ground truth data are not available prior to 2015. Given that machine learning models predict more poorly on observations that are spatially distant from training data samples Proctor et al. (2023) or outside of the time period covered by the training data (Barriga-Cabanillas et al., 2024), I am concerned that using predicted yields as my main outcome of interest would introduce additional measurement error into the estimation of the effect of mining on yields. As a result, I opt to use NDVI as my main measure of agricultural output rather than predicted yields. This allows me to remove prediction error as a source of measurement error, and instead focus solely on understanding how measurement error in the NDVI data itself could be affecting estimated coefficients.

10.2.4 Agricultural Seasons For each mining area, I observe 4-5 agricultural seasons. Like Sacks et al. (2010), I define the seasons for planting and harvest using start and end dates, and define the growing season as the days in between the planting end date and the harvest start date. As the biological sciences literature suggests that pollution can affect crops more strongly during the reproductive phase of the plant, which occurs early in the growing season, I divide this growing season in half to make early and late growing seasons (Liu and Desai, 2021). Finally, I define the non-farming season as the time between the harvest end date of one year and the planting start date of the next year. In mining areas where the harvesting period of one year overlaps with the planting period of the subsequent year, I define the harvest season as ending at the start of next year's planting season. For each year, I assign each mining area the modal season dates across all cells falling within the buffers linked to that mine, where the season dates are consistent across the treatment and control buffers of the same mine.

10.3 NDTI vs. ground-based turbidity measurements To estimate the relationship between NDTI and ground-based turbidity, I use geo-located data on ground-based measures of total suspended solids from the Global River Water Quality (GRWQ) data. I calculate NDTI over water pixels located within a 500-m radius of the GRWQ sampling point, during the month that the sample was collected, again using Landsat 7 data.

Table 6: Relationship between Total Suspended Solids from Ground-based Sampling and Remotely-Sensed Turbidity

	(1)	(2)
Normalized Difference Turbidity Index	940.9***	887.1**
	(341.4)	(364.7)
Number of sampling sites	178	178
Include weather controls	No	Yes
TSS mean	147.012	147.012
NDTI mean	.007	.007

Each column reports the results of a linear regression. The unit of analysis is a sampling site on the date that the water sampling occurred. The dependent variable is the level of total suspended solids identified from ground-based water sampling and testing at the sampling site, on a specific date. The independent variable is the mean normalized difference turbidity index for the river segment falling within a 500 meter buffer around the sampling site, in the month corresponding to the sampling date. All models include linear and quadratic controls for mean temperature, precipitation, vapor pressure, wet days, evapotranspiration and cloud cover. Robust standard errors are reported in parentheses

10.4 AOD Spatial Lag Model To estimate the spatial lag model for AOD, I define concentric rings around a mine point with the following distances: 0-10km, 10-20km, 20-40km, 40-60km, 60-100km, 100-150km, 150-200km. Each of these rings is divided into four 90-degree slices indicating the cardinal directions. For each mine-month, I calculate the number of days the wind is blowing from the mine into a given side, as well as the average wind speed experienced by that side. I estimate a version of Equation 2, with an additional interaction between the post dummy, wind intensity and a dummy for each distance ring, with the 150-200km ring being the ommitted category. Figure 12 plots the coefficients of these triple interactions, which show the DID estimates for the effect of mining on AOD at increasing distances from the mine centroid.

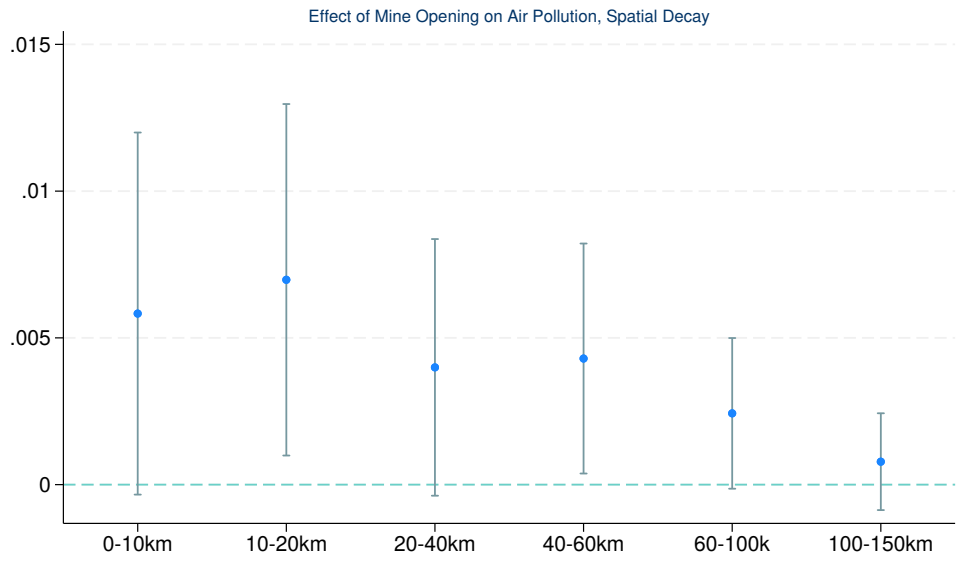

Figure 12: Mean AOD Spatial Decay Model

10.5 NDVI Spatial Lag Model To estimate the spatial lag model for AOD, I define concentric rings around a mine point with the following distances: 0-20km, 20-40km, 40-60km, 60-100km, 100-150km. I estimate a version of Equation 2 that replaces the Near dummy with a series of dummies for each of the distance rings, with the 100-150km ring being the omitted category. Figure 13 plots the coefficients on the interactions between the post dummy and each of the distance ring dummies, showing the DID effect of mining on NDVI at increasing distances from the mine centroid.

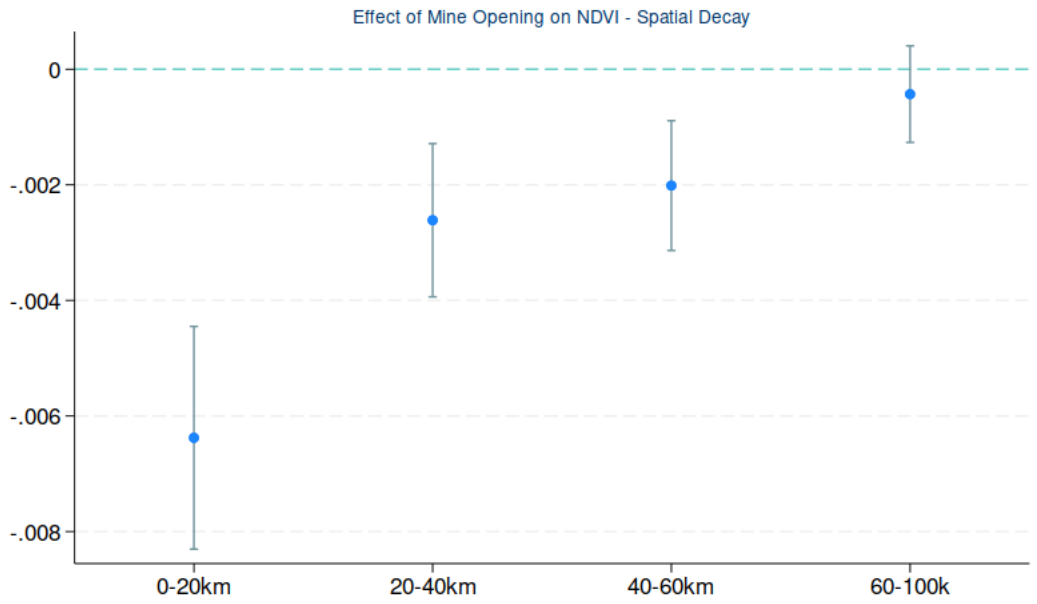

Figure 13: Mean NDVI Spatial Decay Model

10.6 Water Pollution Robustness

10.6.1 Irrigated pixels only The table below reports the results for estimating Equation 2 using mean NDVI calculated over only irrigated cropland pixels within water buffers as the outcome of interest.

Table 7

	Mean NDVI - Irrigated Cropland Only
Downstream × Post	-0.0206
	(0.0233)
Number of mines	4
Mine-side FE	Yes
Mine-year-month FE	Yes
Mean NDVI (t-1)	.46

Standard errors in parentheses

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

10.7 Distributed Lag Results

10.7.1 Average effects: distributed lag as table

Effect of Cumulative Air Pollution Exposure on NDVI

Number of Months Prior	$\sum_p^P \delta_p$	$\frac{\sum_p^P \delta_p}{\text{Mean NDVI}}$	Standard error	p
1	0.0003	0.0656	0.0011	0.7956
2	0.0006	0.1307	0.0012	0.6483
3	0.0005	0.1263	0.0014	0.6890
4	0.0007	0.1501	0.0015	0.6616
5	0.0007	0.1581	0.0016	0.6672
6	0.0005	0.1240	0.0017	0.7488
7	0.0004	0.0869	0.0018	0.8297
8	0.0002	0.0405	0.0018	0.9220
9	0.0001	0.0222	0.0018	0.9579
10	0.0001	0.0260	0.0018	0.9509
11	0.0001	0.0302	0.0018	0.9432
12	0.0001	0.0285	0.0018	0.9464
13	0.0001	0.0293	0.0018	0.9451
14	0.0001	0.0285	0.0019	0.9468
15	0.0001	0.0291	0.0019	0.9460
16	0.0001	0.0231	0.0019	0.9574
17	0.0001	0.0268	0.0019	0.9506
18	0.0002	0.0348	0.0019	0.9363
19	0.0002	0.0345	0.0019	0.9371
20	0.0001	0.0330	0.0019	0.9401
21	0.0002	0.0374	0.0019	0.9326
22	0.0001	0.0290	0.0019	0.9478
23	0.0001	0.0214	0.0019	0.9614
24	0.0001	0.0148	0.0019	0.9734

10.7.2 Distributed lag: growing versus non-growing seasons The figure below plots the distributed lag model estimated separately over months in the growing and non-growing seasons.

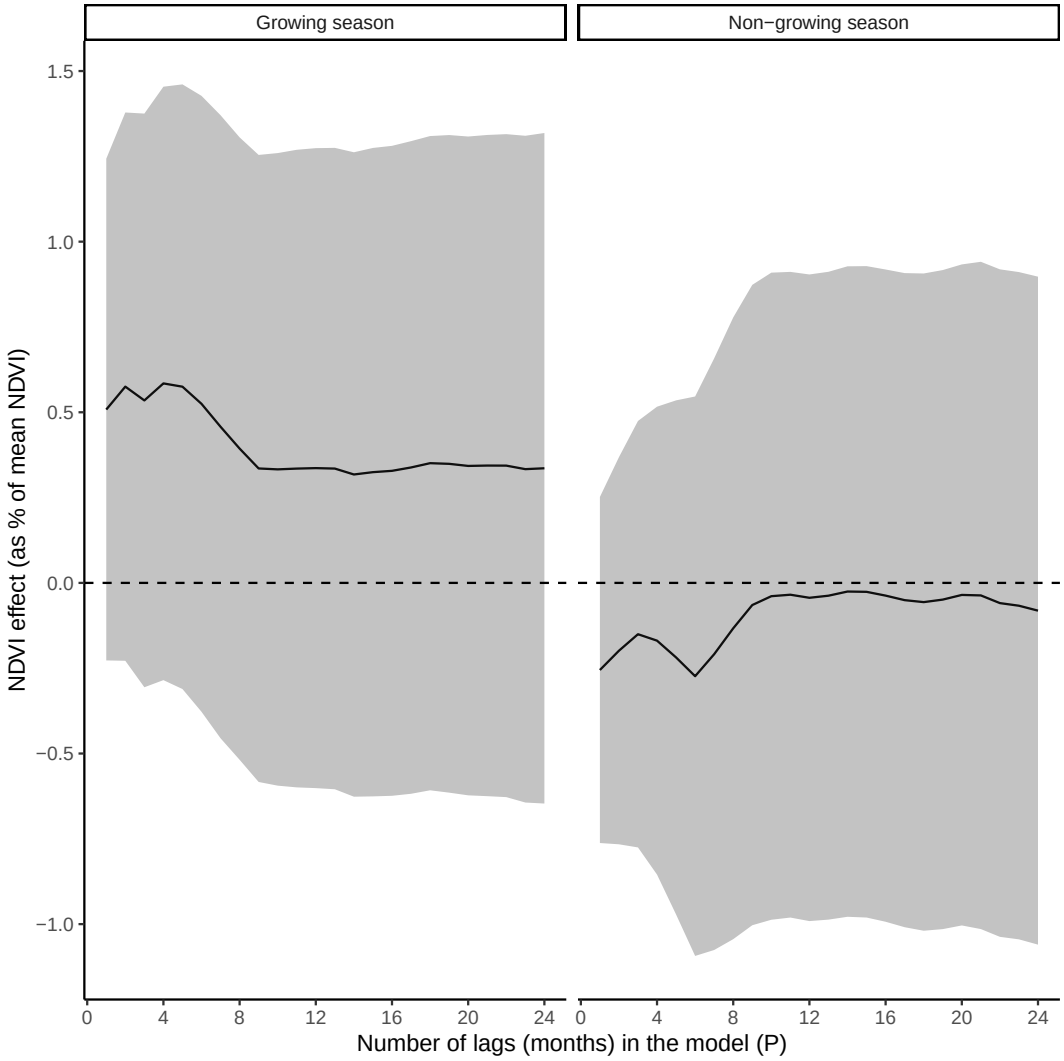

Figure 14: Average effect of cumulative mining-induced air pollution exposure on NDVI - growing vs. non-growing seasons

10.8 Heterogeneity - Near vs. Far

10.8.1 Mine-specific treatment effects The figure below plots a histogram of the mine-specific DID estimates of the overall effect of mining on local yields against a histogram of simulated DID estimates. To generate the observed mine-specific DID estimates, I regress NDVI on the side of a mine (near/far) on a dummy variable for Near interacted with the dummy variable for Post, the main effects and controls for weather. To generate the simulated DID estimates, I make random draws from a normal distribution with a mean equal to the average overall treatment effect, -0.00638, shown in Table 3 Column 1, and standard deviation equal to the standard error of the average overall treatment effect, 0.000980.

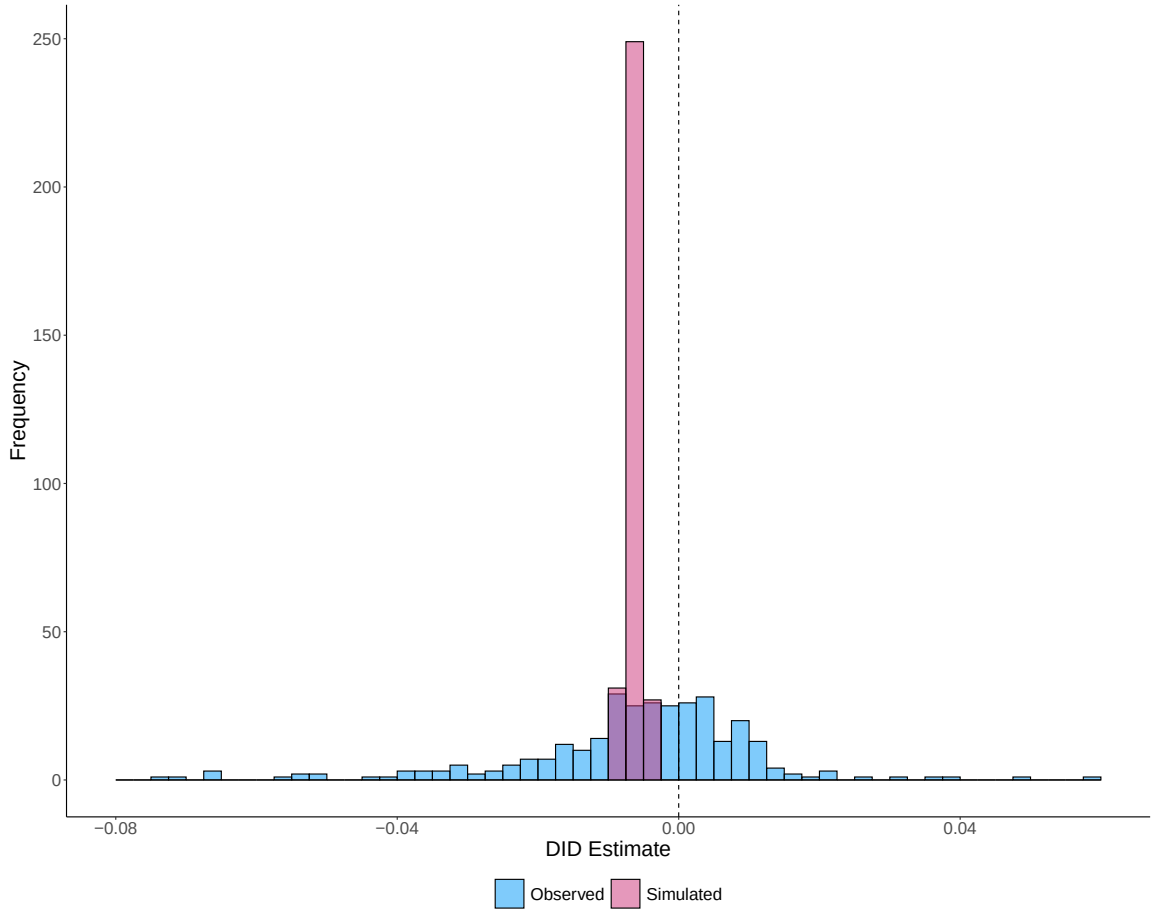

Figure 15: Variation in observed mine-specific DIDs is greater than what we would expect due to sampling variation

10.8.2 Standard heterogeneity analysis The section below reports the results for the standard heterogeneity analysis for the overall effect of mining on yields.

Table 9: Heterogeneous Treatment Effects - Economic Factors

	Base	Above Med. Mineral Rents	Above Med. GDP - 2000	Above Med. Pop - 2000
Near × Post	-0.00638***	-0.00534***	-0.00692***	-0.00709***
	(0.000980)	(0.00108)	(0.00157)	(0.00160)
Near × Post × Z		-0.00568**	0.00107	0.00142
		(0.00242)	(0.00195)	(0.00195)
Number of mines	307	307	307	307
Mean NDVI (t-1)	.476	.476	.476	.476

Standard errors in parentheses

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

Table 10: Heterogeneous Treatment Effects - Mine Characteristics

	Base	Open Pit	Gold	Coal	Near Other Mine	Near Town	Old Mine
Near × Post	-0.00638***	-0.00615***	-0.00639***	-0.00675***	-0.00416***	-0.00611***	-0.00615***
	(0.000980)	(0.00174)	(0.00133)	(0.00113)	(0.00123)	(0.00150)	(0.00159)
Near × Post × Z		-0.000376	0.00000921	0.00247	-0.00449**	-0.000573	-0.000362
		(0.00210)	(0.00178)	(0.00151)	(0.00194)	(0.00193)	(0.00201)
Number of mines	307	307	307	307	307	307	307
Mean NDVI (t-1)	.476	.476	.476	.476	.476	.476	.476

Standard errors in parentheses

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

Table 11: Heterogeneous Treatment Effects - Environmental Conditions

	Base	Above Med. NDVI - 2000	Above Med. Precipitation - 2001	Growing Season
Near × Post	-0.00638***	-0.00526***	-0.00762***	-0.00560***
	(0.000980)	(0.00152)	(0.00163)	(0.00113)
Near × Post × Z		-0.00226	0.00247	-0.00162
		(0.00195)	(0.00194)	(0.00114)
Number of mines	307	307	307	307
Mean NDVI (t-1)	.476	.476	.476	.476

Standard errors in parentheses

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

10.8.3 ML heterogeneity analysis Figure 16 plots the set of the 10 least important variables used in the best performing random forest model, while Figure 17 plots the 10 most important variables from an alternative classification random forest model, trained to predict an indicator for whether the treatment effect is less than zero using the same set of predictors as in the main ML exercise.

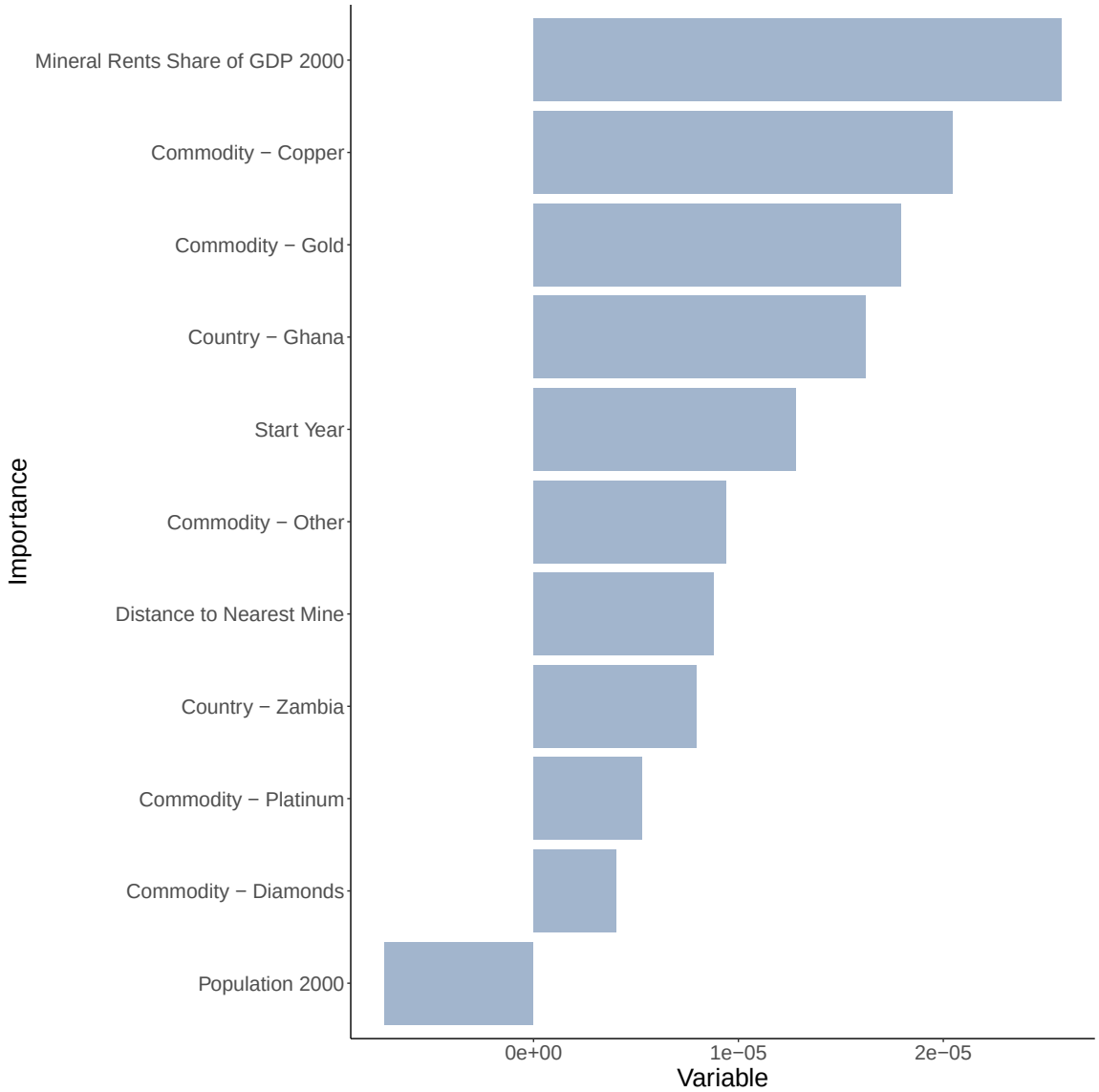

Figure 16: 10 Least Important Variables from ML Model

Figure 18 plots mine-specific DID estimates from estimating the effect of mine openings on remotely sensed air pollution using Equation 2 with AOD as the outcome of interest against mine-specific DID estimates obtained by estimating the same specification using NDVI as the outcome of interest.

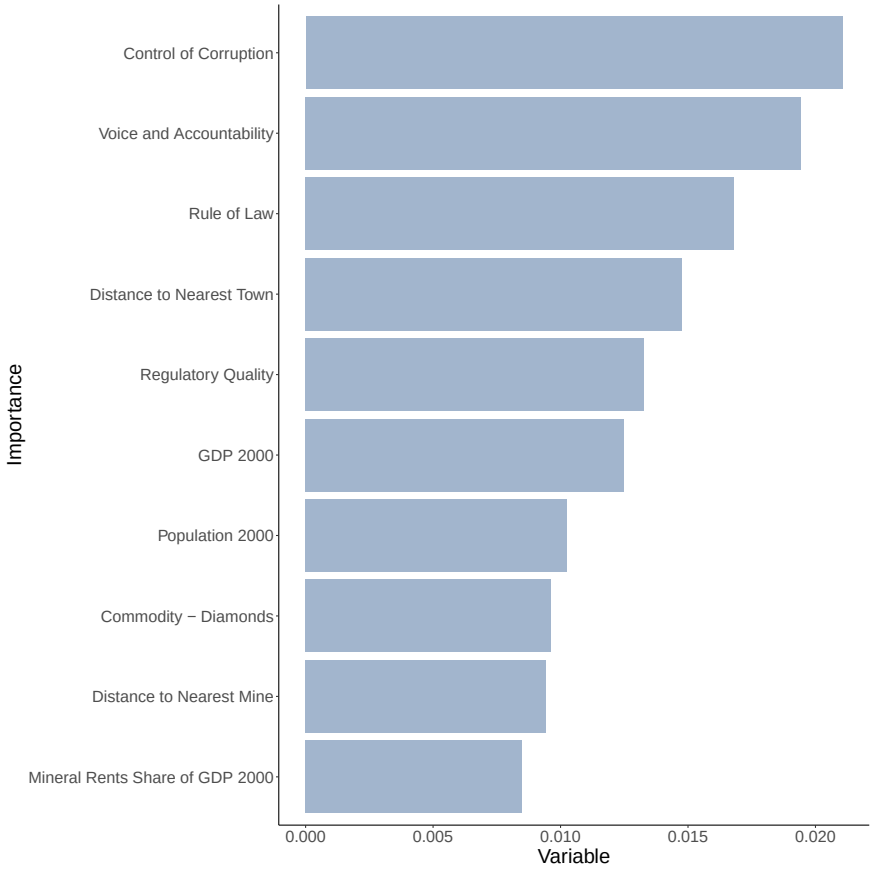

Figure 17: Most Important Variables from Best Random Forest Model: Classification

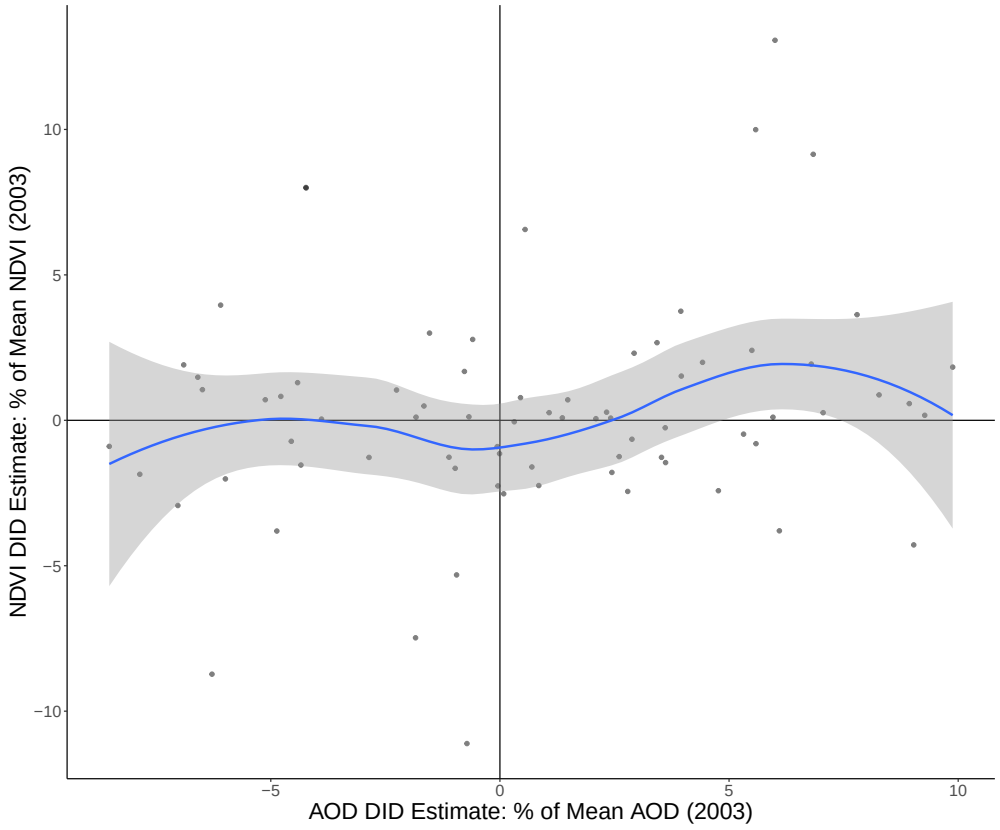

Figure 18: Mine-specific AOD DID estimates vs. NDVI DID estimates

10.9 Heterogeneity - Contemporaneous Air Pollution The tables below report heterogeneous treatment effect estimates using the contemporaneous air pollution specification of Equation 2.

Table 12

	Base	EITI	Control of Corruption	Govt. Effectiveness	Rule of Law	Reg. Quality	Voice and Accountability
Wind	-0.00123	-0.00261	-0.00190	-0.000730	-0.00280	-0.000526	-0.000656
	(0.00224)	(0.00412)	(0.00333)	(0.00382)	(0.00337)	(0.00375)	(0.00328)
Wind × (Post - 3)	0.000277	0.000621	-0.00240	-0.00122	-0.000435	-0.00169	-0.00100
	(0.00146)	(0.00202)	(0.00189)	(0.00166)	(0.00188)	(0.00172)	(0.00152)
Wind × (Post - 3) × Z		-0.000554	0.00402	0.00246	0.00106	0.00321	0.00221
		(0.00284)	(0.00267)	(0.00271)	(0.00271)	(0.00268)	(0.00271)
Number of mines	102	102	102	102	102	102	102
Mean NDVI (t-1)	.492	.492	.492	.492	.492	.492	.492

Standard errors in parentheses

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

Table 13

	Base	Above Med. Mineral Rents	Above Med. GDP - 2000	Above Med. Pop - 2000
Wind	-0.00123	-0.00283	-0.00102	-0.000486
	(0.00224)	(0.00265)	(0.00282)	(0.00341)
Wind × (Post - 3)	0.000277	0.0000825	0.000147	0.00193
	(0.00146)	(0.00175)	(0.00193)	(0.00209)
Wind × (Post - 3) × Z		0.000671	0.000372	-0.00370
		(0.00274)	(0.00285)	(0.00274)
Number of mines	102	102	102	102
Mean NDVI (t-1)	.492	.492	.492	.492

Standard errors in parentheses

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

Table 14

	Base	Open Pit	Gold	Coal	Copper	Near Other Mine	Near Town	Old Mine
Wind	-0.00123	-0.00688	0.00243	-0.00407*	-0.00174	-0.00227	-0.00156	-0.00123
	(0.00224)	(0.00555)	(0.00333)	(0.00235)	(0.00256)	(0.00328)	(0.00329)	(0.00224)
Wind × (Post - 3)	0.000277	0.00779**	0.00125	0.000267	0.00104	0.00229	-0.0000280	0.000277
	(0.00146)	(0.00300)	(0.00204)	(0.00164)	(0.00143)	(0.00186)	(0.00217)	(0.00146)
Wind × (Post - 3) × Z		-0.0104***	-0.00252	-0.000114	-0.00381	-0.00412	0.000698	0
		(0.00338)	(0.00253)	(0.00369)	(0.00439)	(0.00281)	(0.00278)	(.)
Number of mines	102	102	102	102	102	102	102	102
Mean NDVI (t-1)	.492	.492	.492	.492	.492	.492	.492	.492

Standard errors in parentheses

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

Table 15

	Base	Above Med. AOD - 2003	Above Med. NDVI - 2003	Above Med. Precipitation - 2003	Growing Season
Wind	-0.00123	0.000823	-0.00151	-0.00132	-0.00223
	(0.00224)	(0.00463)	(0.00338)	(0.00461)	(0.00283)
Wind × (Post - 3)	0.000277	-0.000160	-0.00162	0.000426	-0.000810
	(0.00146)	(0.00151)	(0.00160)	(0.00216)	(0.00159)
Wind × (Post - 3) × Z		0.000682	0.00298	-0.000231	0.00260
		(0.00265)	(0.00258)	(0.00292)	(0.00218)
Number of mines	102	102	102	102	102
Mean NDVI (t-1)	.492	.492	.492	.492	.492

Standard errors in parentheses

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

10.10 Heterogeneity - Distributed Lag The figures below plot the distributed lag model from Equation 3 estimated separately over different subsets of the data split along various dimensions of heterogeneity.

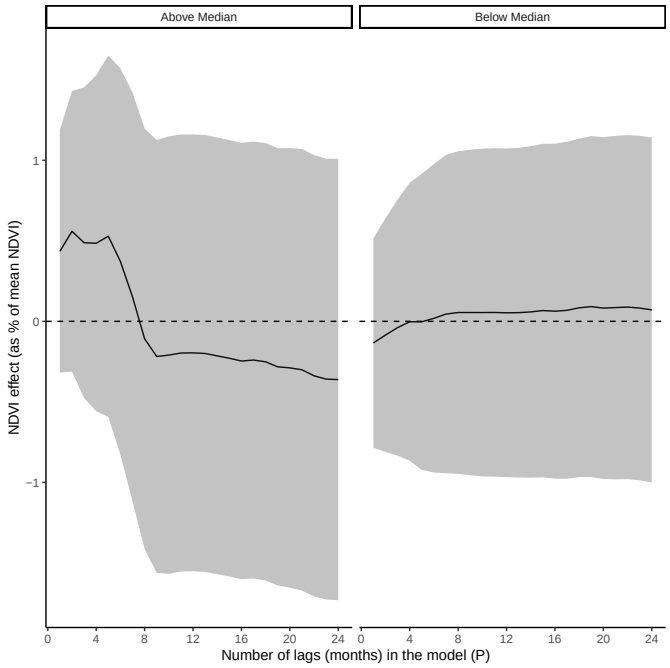

Figure 19: Heterogeneity by control of corruption

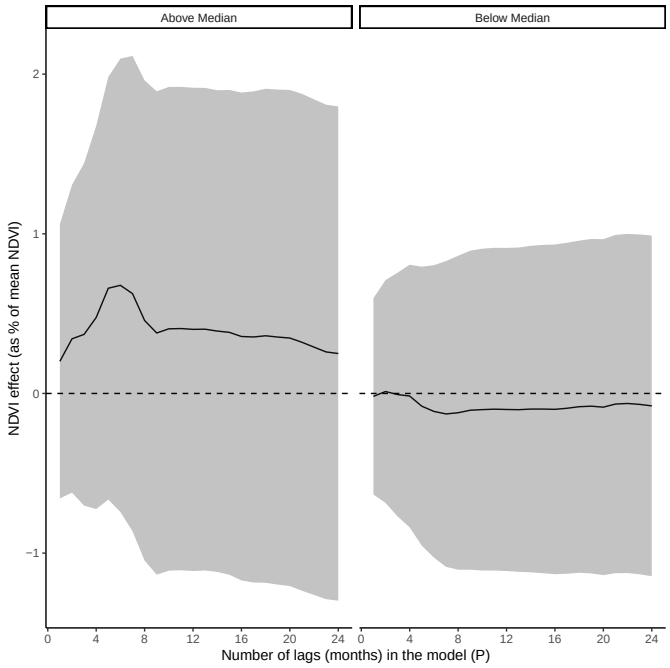

Figure 20: Heterogeneity by rule of law

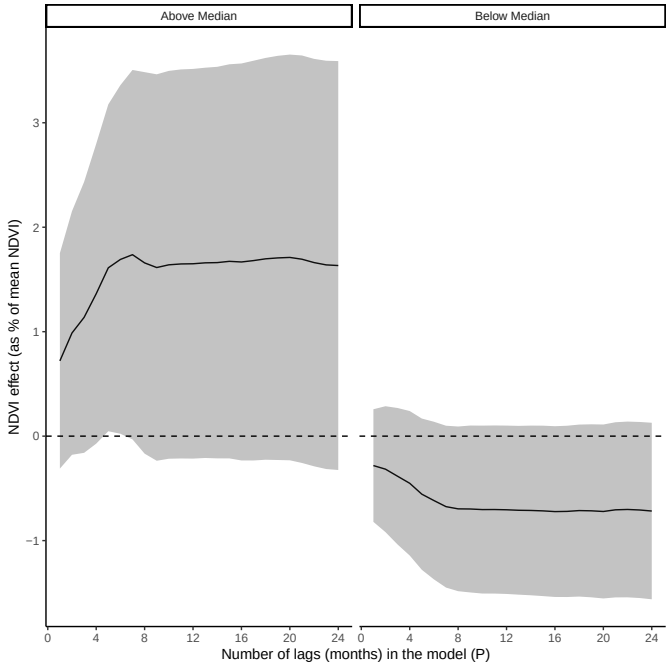

Figure 21: Heterogeneity by government effectiveness

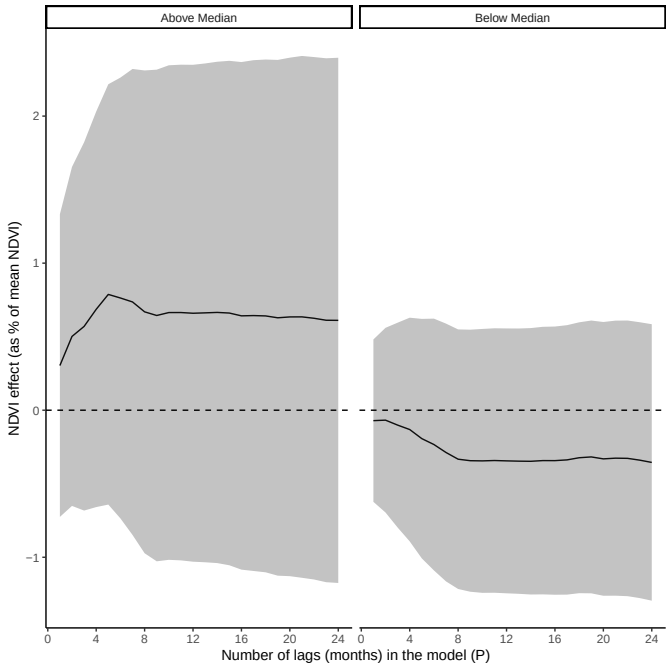

Figure 22: Heterogeneity by regulatory quality

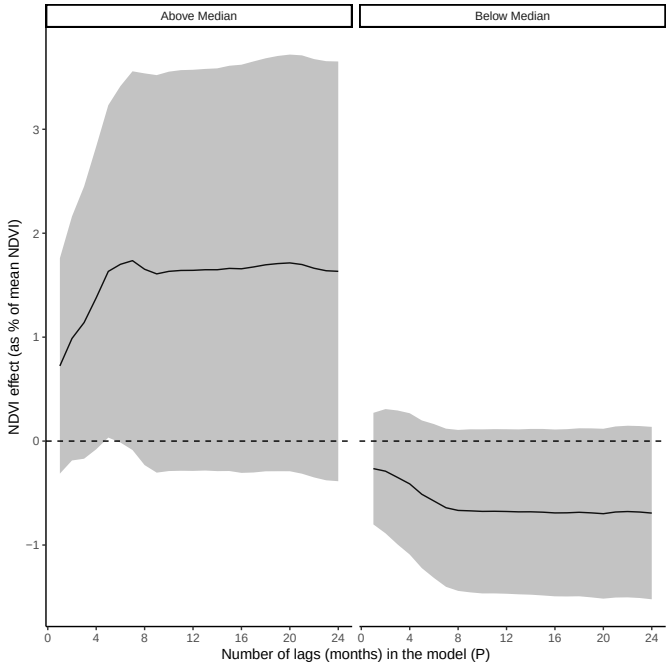

Figure 23: Heterogeneity by voice and accountability

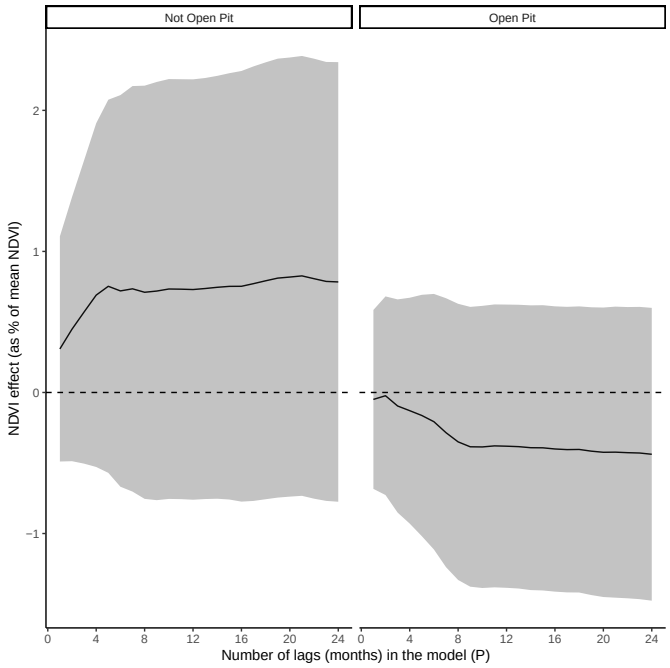

Figure 24: Heterogeneity by mine type

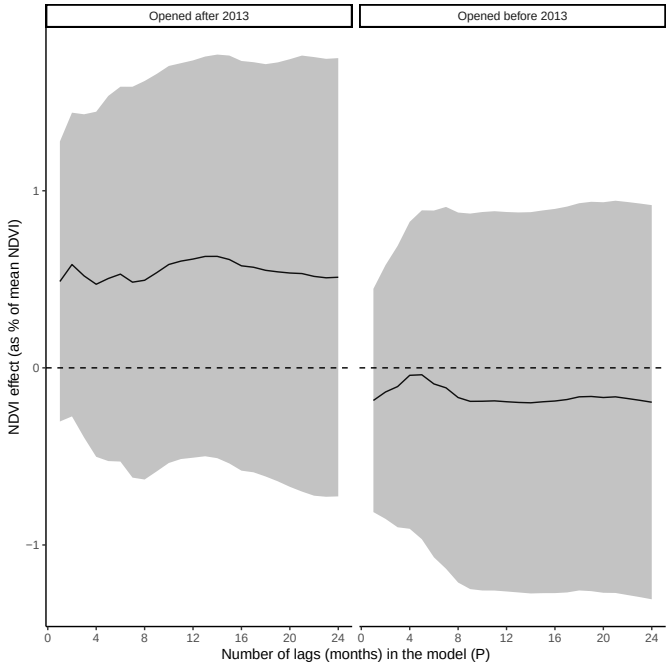

Figure 25: Heterogeneity by mine age

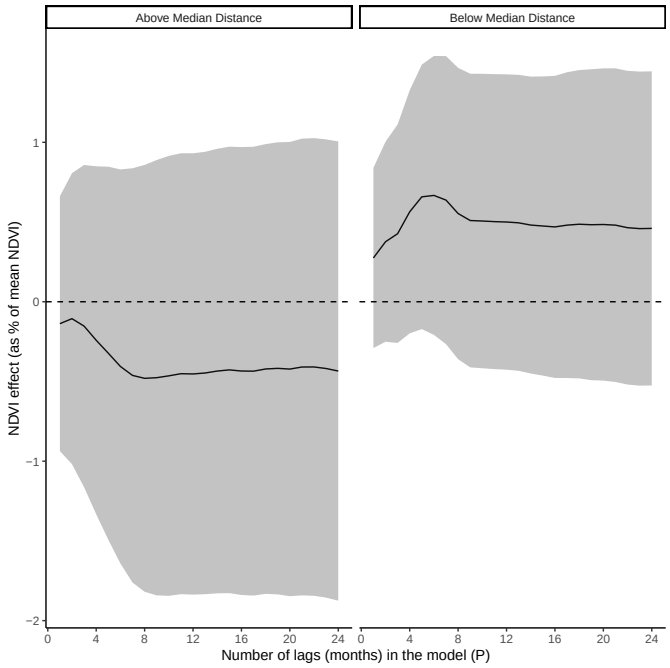

Figure 26: Heterogeneity by distance to nearest mine

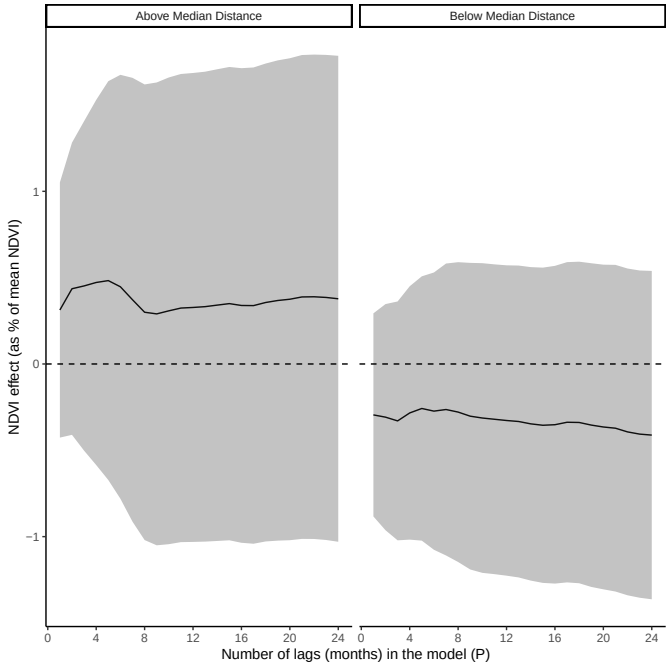

Figure 27: Heterogeneity by distance to nearest town

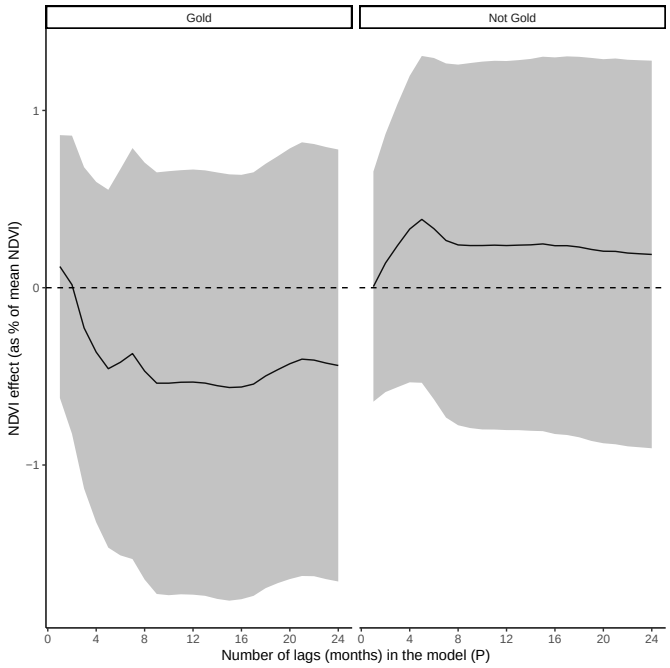

Figure 28: Heterogeneity by commodity: gold

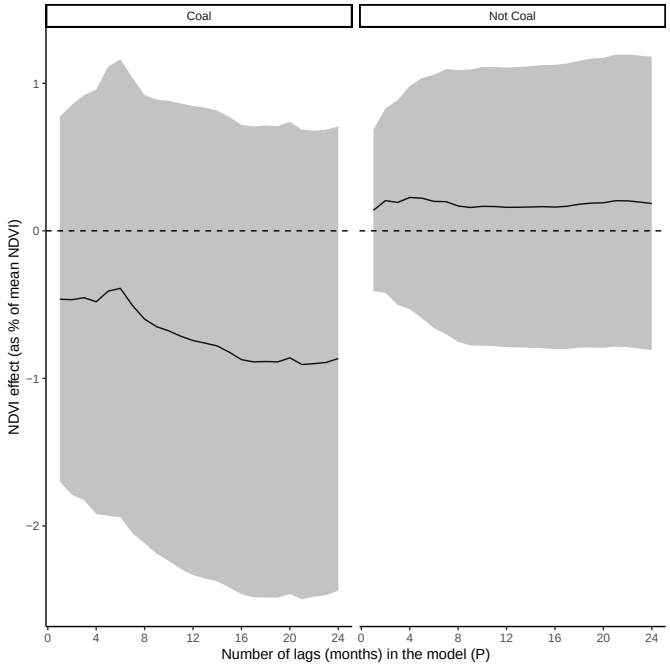

Figure 29: Heterogeneity by commodity: coal

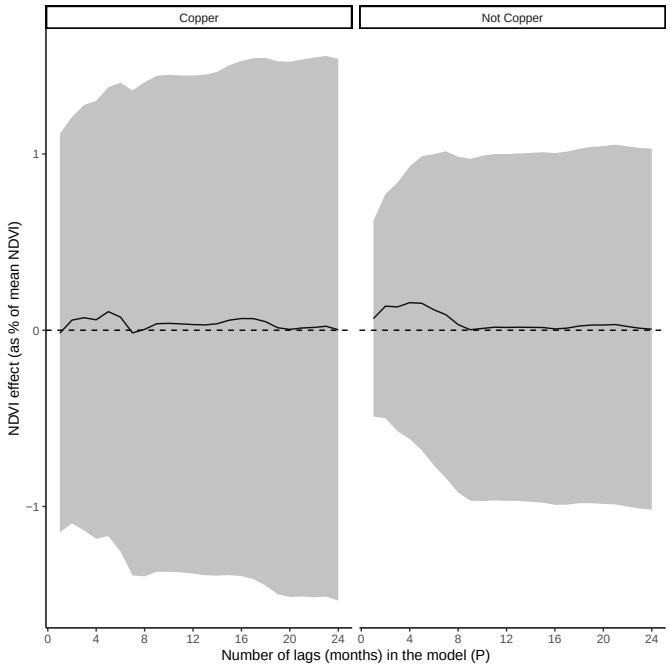

Figure 30: Heterogeneity by commodity: copper

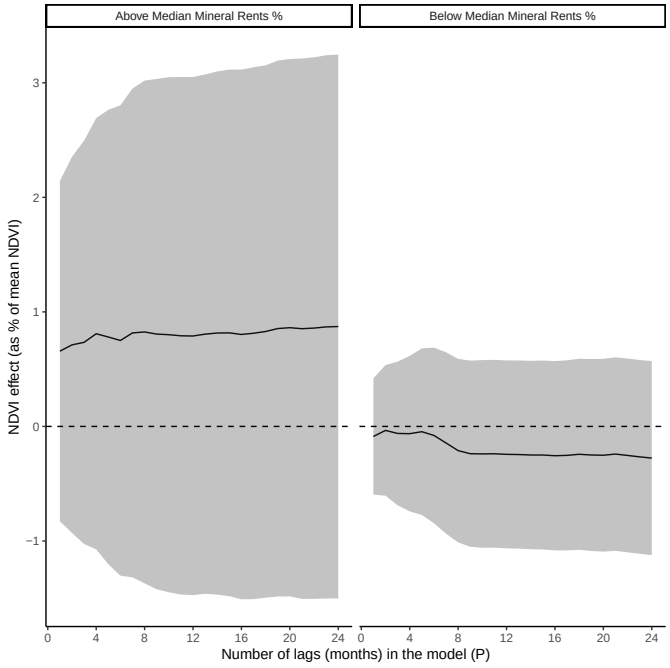

Figure 31: Heterogeneity by mineral rents as share of GDP - 2000

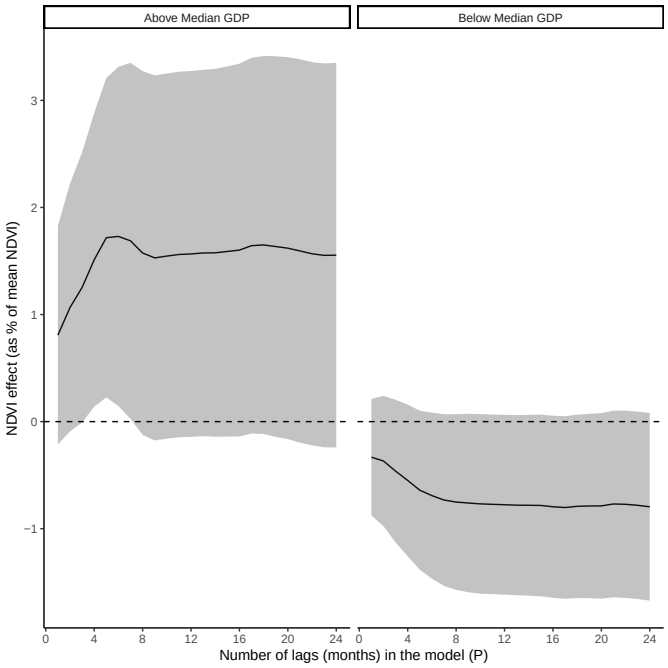

Figure 32: Heterogeneity by GDP - 2000

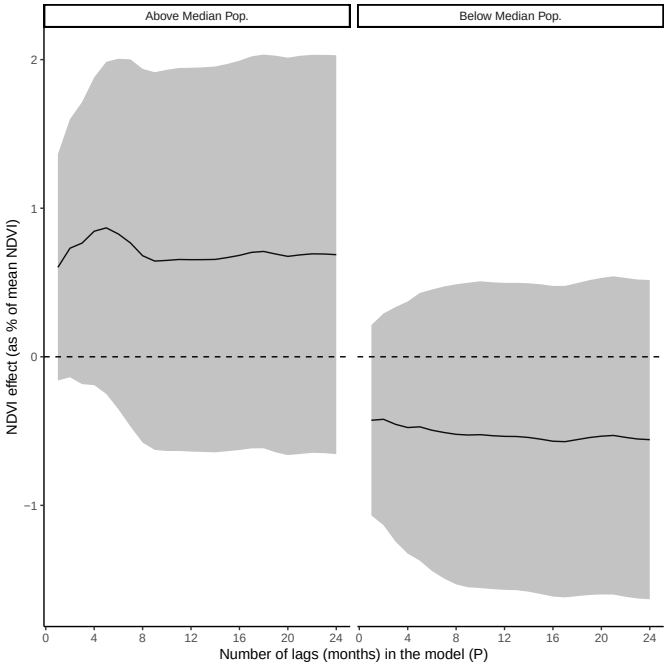

Figure 33: Heterogeneity by population - 2000

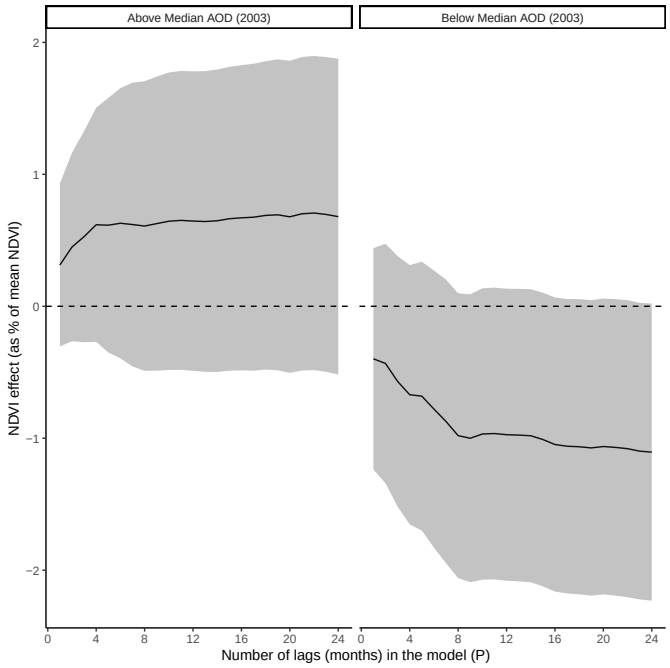

Figure 34: Heterogeneity by initial pollution levels (AOD 2003)

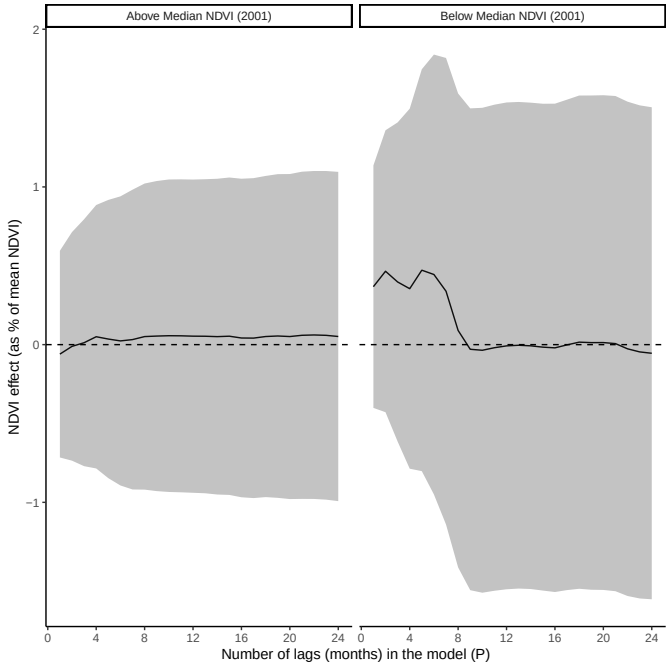

Figure 35: Heterogeneity by initial yields (NDVI 2001)

10.11 Heterogeneity - Water The tables below report the results from estimating Equation 2 for water pollution, with an additional triple interaction for the dimension of heterogeneity of interest.

Table 16

	Base	EITI	Control of Corruption	Govt. Effectiveness	Rule of Law	Reg. Quality	Voice and Accountability
Downstream × Post	-0.0140**	-0.0129	-0.0131**	-0.0164*	-0.0126**	-0.0117*	-0.0149**
	(0.00587)	(0.00911)	(0.00600)	(0.00840)	(0.00487)	(0.00657)	(0.00660)
Downstream × Post × Z		-0.00203	-0.00185	0.00527	-0.00330	-0.00514	0.00170
		(0.0119)	(0.0119)	(0.0115)	(0.0126)	(0.0121)	(0.0119)
Number of mines	38	38	38	38	38	38	38
Mean NDVI (t-1)	.467	.467	.467	.467	.467	.467	.467

Standard errors in parentheses

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

Table 17

	Base	Open Pit	Gold	Coal	Near Other Mine	Near Town	Old Mine
Downstream × Post	-0.0140**	-0.0162	-0.00849	-0.0132*	-0.000109	-0.0189***	-0.0234**
	(0.00587)	(0.0109)	(0.00511)	(0.00706)	(0.00625)	(0.00687)	(0.0100)
Downstream × Post × Z		0.00333	-0.0166	-0.00432	-0.0285**	0.00937	0.0151
		(0.0129)	(0.0147)	(0.00961)	(0.0108)	(0.0115)	(0.0122)
Number of mines	38	38	38	38	38	38	38
Mean NDVI (t-1)	.467	.467	.467	.467	.467	.467	.467

Standard errors in parentheses

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

Table 18

	Base	Above Med. Mineral Rents	Above Med. GDP - 2000	Above Med. Pop - 2000
Downstream × Post	-0.0140**	-0.0180***	-0.0121	-0.0193**
	(0.00587)	(0.00553)	(0.00948)	(0.00878)
Downstream × Post × Z		0.0129	-0.00406	0.0108
		(0.0151)	(0.0115)	(0.0115)
Number of mines	38	38	38	38
Mean NDVI (t-1)	.467	.467	.467	.467

Standard errors in parentheses

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

Table 19

	Base	Above Median Turbidity - 2000	Above Med. NDVI - 2000	Above Med. Precipitation - 2001	Growing Season
Downstream × Post	-0.0140**	-0.0252***	-0.0107**	-0.0214**	-0.0132**
	(0.00587)	(0.00856)	(0.00450)	(0.00975)	(0.00556)
Downstream × Post × Z		0.0199*	-0.00633	0.0146	-0.00185
		(0.0113)	(0.0112)	(0.0116)	(0.00467)
Number of mines	38	38	38	38	38
Mean NDVI (t-1)	.467	.467	.467	.467	.467

Standard errors in parentheses

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

10.12 Structural Breaks Model For each mine, I construct total luminosity for each year from 1992-2012 as the sum of nighttime light intensity across all pixels that fall within the 20-km disk around the mine centroid, in a given year. The nightlight data is provided by the DMSP OLS at a 1-km resolution. However, the DMSP-OLS data is not available to the public after 2013. From 2013 onwards, there was a switch to the Visible Infrared Imaging Radiometer Suite (VIIRS) instrument as the source of nightlights. Given that the spatial and radiometric resolution from VIIRS is higher than DMSP-OLS, Li et al. (2020) created a temporally harmonized global nighttime lights dataset from 1992-2019. Unfortunately, the temporal consistency of their harmonized data does not perform well in areas with low levels of luminosity (pixels with DN values greater than 7). Time series of nighttime light intensity for mining areas reveal a sharp jump in luminosity in 2013, when there was a switch to VIIRS. Given that the average pixel in a mining area has a DN value of 6.5, this suggests that the harmonized data is not temporally consistent for mining areas. As a result, I only estimate a structural break in nightlights for mines that opened before 2013, to avoid confounding due to the switch from DMSP to VIIRS.

I use the methods of Andrews (1993) and Andrews and Ploberger (1994) to detect a structural break in the mean of nightlights for each mine. Given a single proposed break point, the nightlights time series is split into two bins. The estimation method fits a regression of nightlights on an intercept for the data in each bin and calculates an F statistic based on the null hypothesis that the mean in nightlights is the same between the two bins. This step is repeated for all possible break points in the data to yield a time series of F-statistics. The structural break is identified as the year which yields the largest F-statistic. To test whether this structural break is statistically significant, I use the critical values for the F-statistics identified by Andrews (1993) and Andrews and Ploberger (1994), which are computed under the null hypothesis of no structural change such that the asymptotic probability that the supremum of the time series of F-statistics exceeds this critical value is $\alpha = 0.05$. In other words, if the F-statistic of the structural break is higher than the critical value, we can reject the null of no structural breaks in nightlights at the 0.05 level.

I am successfully able to estimate a structural break for 100 mines in my sample that opened prior to 2013. For the median mine, a structural break in nightlights occurs 3 years prior to the listed S&P start date,

with approximately 60% of the mines having the break year between 0 and 3 years prior to the S&P defined date. This finding aligns with that of Benshaul-Tolonen (2020), who uses a local polynomial regression to show that there is a break in the trend for nightlights within 10 kilometers of a mine that occurs roughly 2 years prior to the date of a mine opening. She refers to this break as the start of the "investment" phase in a mine's life cycle. Figure 36 plots the histogram of the difference between the S&P year and the break year.

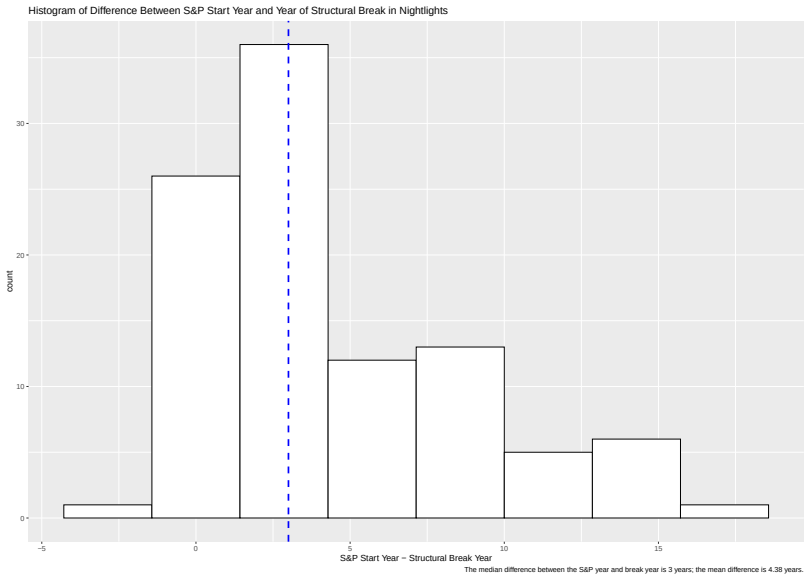

Figure 36: Difference between S&P Year and Structural Break Year

Furthermore, for most mines the year identified as the structural break visually aligns with the timing of a sharp increase in nighttime lights. Figure 37 provides an example from Goedgevonden coal mine in South Africa.

To investigate the robustness of this method for detecting structural breaks, I examine the correlation between the F-statistic from the break year and the start-date "error", defined as the difference between the S&P defined year and the break year. One might be concerned if the method is more likely to reject the null of no structural break in cases where there is a large difference between the S&P defined year and the break year, suggesting that it would not perform well in detecting breaks that occur close to the listed date.

These concerns do not seem to be borne out: in Figure 38 we see that across the domain of the "error", the sup F statistic ranges between 0-150 in most cases. In fact, the few outliers with very high supF statistics occur for cases where the S&P year and the break year are very close.

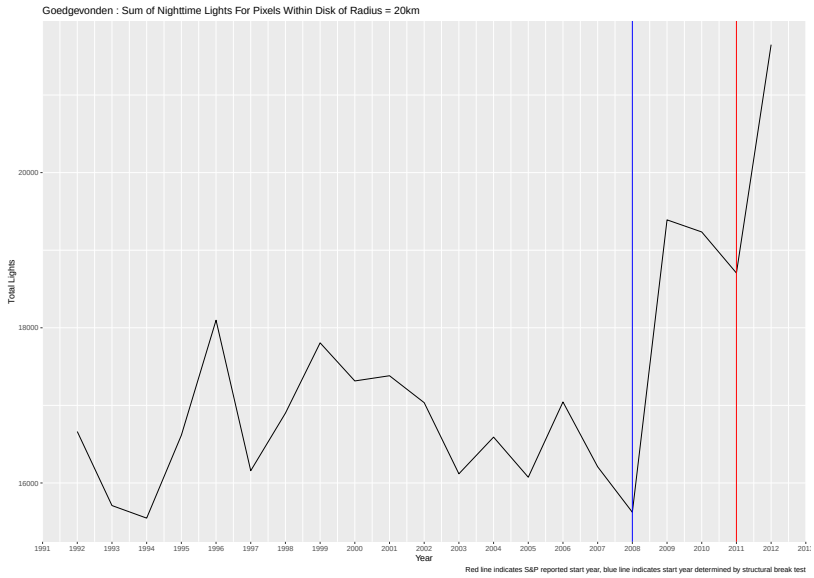

Figure 37: Time series of nightlights for Goedgevonden coal mine, South Africa

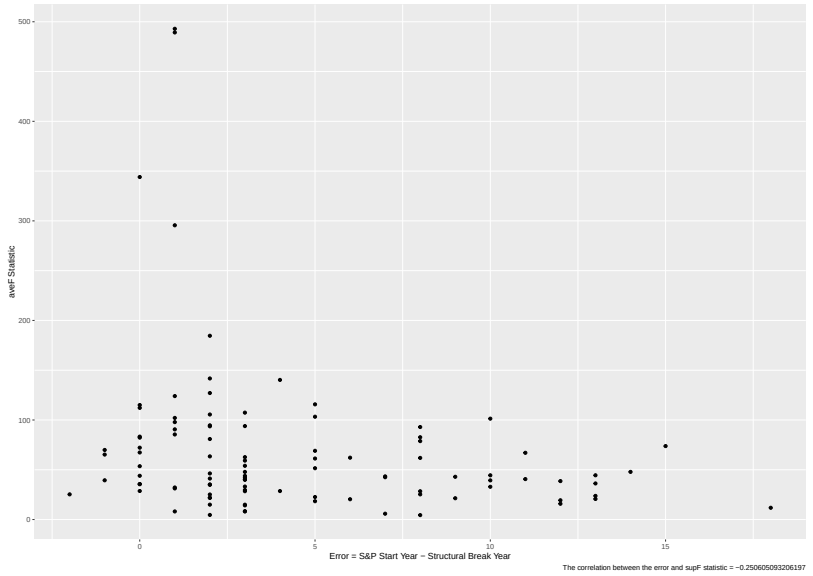

Figure 38: Magnitude of supF statistic vs. the difference between S&P defined start year and break year

10.13 Measurement Error

10.13.1 Effect of Measurement Error on DID Estimates Ideally, I would observe farm-level yields around mining areas over time and estimate the following version of Equation 2

$$Yields_{smt} = \beta_Y Exposure_{sm} \times Post_{mt} + \alpha_{sm} + \lambda_{mt} + \mathbf{X'_{smt}}\mathbf{\Gamma} + \epsilon_{smt} \tag{6}$$

However, these data are not available in Sub-Saharan Africa, so I use NDVI as the remotely sensed proxy for yields, estimating the following specification instead:

$$NDVI_{smt} = \beta_{\tilde{Y}} Exposure_{sm} \times Post_{mt} + \alpha_{sm} + \lambda_{mt} + \mathbf{X'_{smt}}\mathbf{\Gamma} + \epsilon_{smt} \tag{7}$$

Due to measurement error in NDVI, estimation of Equation 7 will recover a different coefficient and standard error for $\hat{\beta}_{\tilde{Y}}$ than those from Equation 7. To understand drivers of measurement error, I specify the following linear measurement error model from Keogh et al. (2020):

$$NDVI_{smt} = \theta + \mu Yield^*_{smt} + v_{smt} \tag{8}$$

where NDVI on side s of mine m at time t is observed, while total yields across all farms on side s of mine m at time t is the true measure of interest and v_{smt} is noise. I assume v to be mean 0 and $cov(Yield^*, v) = 0$.

Defining $T = Exposure \times Post$, estimation of Equation 7 without any corrections to NDVI will uncover the following coefficient:

$$\mathbb{E}(\hat{\beta}_{\tilde{Y}}) = \lambda\beta_Y + \frac{\sigma_{Tv}}{\sigma^2_T} \tag{9}$$

where σ^2_T is the variance in $Exposure \times Post$ and σ_{Tv} is the covariance between the errors in NDVI and the values of $Exposure \times Post$.

With classical measurement error, estimates of the DID regressions using NDVI as the outcome of interest will be unbiased, though standard errors will be larger. Classical measurement error corresponds to the special case of Equation 8 where $\theta = 0$, $\mu = 1$ and $cov(Yield, v) = cov(T, v) = 0$.

However, Proctor et al. (2023) show that assumptions of classical measurement error are unlikely to hold in regressions with remotely sensed dependent variables, which are potentially subject to both mean-reverting and differential measurement error. First, they demonstrate that mean-reverting measurement error (negative correlation between errors in one variable and itself) is common with remotely-sensed measures by

highlighting that extreme values are systematically underestimated in remotely sensed predictions (Bound and Krueger, 1991; Ratledge et al., 2022). Mean-reverting measurement error corresponds to the case of Equation 8 where $\mu < 1$. Second, they demonstrate the presence of differential measurement error, correlations between errors in one variable and levels of another variable, by showing non-zero covariance between errors in one variable and levels of other variables.

If NDVI only suffers from mean-reverting error and no differential measurement error, a model using uncorrected NDVI with measurement error will tend to have attenuated coefficients. Additionally, the back-of-the-envelope calculation conducted in Section 8 will yield consistent estimates for the share of the overall effect attributed to pollution. This is because no differential measurement error implies $\sigma_{Tv} = 0$ in Equation 9 and taking the ratio of two estimates with NDVI as the outcome of interest will cancel out the bias from λ.

While mean-reverting measurement error is most common with remotely-sensed variables, differential measurement error explains most of the coefficient bias in the case of a mismeasured dependent variable. Importantly, both mean-reverting and differential measurement error can also lead to bias in standard errors of coefficients. In general, the reduced variance in the outcome variable in the presence of mean-reverting measurement error in the dependent variable could lower the sum of squared errors and underestimate standard errors on the DID coefficients of my main specification. However, when differential measurement error is also present, the direction of bias in the uncertainty parameters is theoretically ambiguous (Caroll et al., 2006).

Unfortunately, it is infeasible in my setting to implement the multiple imputation correction outlined by Proctor et al. (2023) to address issues of measurement error bias in NDVI. This is because potential data sources for ground-based yields (One Acre Fund data, World Bank LSMS-ISA) have an insufficient number of observations that are observed for at least one year pre and post-mine openings, in both the near and far groups. Proctor et al. (2023) highlight that the calibration set should have at least 500 observations, otherwise bias can be even higher in the corrected model. Furthermore, while plot-level yields from data outside of Sub-Saharan Africa could be used for the calibration set, bias in the corrected model can be higher than in the uncorrected model if calibration locations are greater than 200km away from the locations used in the main analysis Proctor et al. (2023).

10.13.2 Effect of Measurement Error on Quantification Exercise Given plot-level yields, the "ideal" measure of NDVI for each plot, $NDVI_p^*$, would be NDVI based on information detected by the satellite without any errors or interference and calculated over the exact boundaries of plot p. However, the NDVI measure used in Equation 5 differs from $NDVI^*$ in two ways. Dropping the t subscript for simplicity,

I define the observed NDVI measure for plot p, which falls in cell c, as follows:

$$NDVI_{p(c)} = \xi NDVI_p^* + d_p + a_p$$

where d_p corresponds to deviations from "true" plot-level NDVI caused by satellite-detection errors and a_p corresponds to deviations for plot p NDVI from the average of all plots in c. NDVI detection errors, d_p, are driven by a variety of factors, most notably cloud cover, but also surface reflectance, canopy thickness, the level of atmospheric aerosols and sensor errors[13]. These detection errors make NDVI observed from satellite data an imperfect measure of NDVI in the presence of no detection errors. Indeed, low pixel quality due to detection errors can cause errors in vegetation indices to increase by 0.04-0.1. While some types of detection errors, such as satellite sensor errors, are likely uncorrelated with plot-level yields, detection errors driven by cloud cover or aerosol loading could be correlated with plot-level yields through the direct effect of atmospheric determinants on crop health, so d_p could introduce bias from non-classical measurement error.

Additionally, aggregation errors, a_p, are generated by the limited 500m resolution of the MODIS data, which makes it infeasible to obtain precise plot-level NDVI measures.[14] GPS-based measures of plot area from household surveys in four African countries show that over 50% of the fields in these countries are smaller than 1 acre, suggesting that the grid-cells used in my analysis likely cover multiple plots (Carletto et al., 2015). If yields are highly heterogeneous across plots within a grid-cell, averaging NDVI over the grid cells may not only weaken the correlation between grid-cell NDVI and plot-level yields, but also introduce aggregation bias into estimation of Equation 5.

I address measurement erorr in cell-level NDVI as follows. First, I limit potential bias from detection errors by calculating NDVI using only high quality pixels free of clouds and aerosols, as well as controlling for local weather variables such as cloud cover, vapor pressure and precipitation. Second, I estimate a version of Equation 5 with grid cells of varying sizes (500m, 2500m, 5000m), to investigate whether β_{NDVI} changes with the degree of aggregation. The tables below report the results of estimating Equation 5 with grid cells of various sizes.

[13] `https://modis-land.gsfc.nasa.gov/ValStatus.php?ProductID=MOD13`

[14] While higher resolution data is available from sources like Sentinal-2 or PlanetLabs, this data does not go far enough back to cover mines that opened prior to 2014.

Table 20: Relationship between Cell-level NDVI and Cell-level Yields, cell-size = 500m

	(1) Planting	(2) Early Growing	(3) Late Growing	(4) Harvest	(5) Nonfarm
Mean NDVI	45.68	38.76	40.76	30.06	7.269
	(32.85)	(25.83)	(33.81)	(22.08)	(25.08)
Year FE	Yes	Yes	Yes	Yes	Yes
Grid cell FE	Yes	Yes	Yes	Yes	Yes
Include weather controls	Yes	Yes	Yes	Yes	Yes
Mean Yields	3005.643	3005.643	3005.643	3005.643	3005.643
Obs.	340	467	421	524	445
R-sq	0.657	0.673	0.673	0.672	0.650

Each column reports the results of an OLS regression. The unit of observation is a cell-year. The dependent variable is mean plot-level yields in kg/ha across all plots falling within a cell, in a given year. *Mean NDVI* is the average NDVI in the cell of size 500m containing the plot, across days in a particular season during the year that maize on the plot was harvested. The columns indicate the season for which NDVI is calculated during the year the plot was harvested. *Mean NDVI* is scaled so that one unit represents a 0.01 increase in NDVI. Each regression includes linear controls for cell-level averages of growing degree days, temperature and precipitation during the maize season across plots falling within the cell, as well as cell-level controls for mean temperature, precipitation, vapor pressure, cloud cover, evapotranspiration and wet days. Year and cell-level fixed effects are included in the regressions.

Table 21: Relationship between Cell-level NDVI and Cell-level Yields, cell-size = 2500m

	(1) Planting	(2) Early Growing	(3) Late Growing	(4) Harvest	(5) Nonfarm
Mean NDVI	21.96*	41.64**	0.756	29.74**	27.23**
	(12.06)	(18.23)	(16.61)	(12.99)	(10.77)
Year FE	Yes	Yes	Yes	Yes	Yes
Grid cell FE	Yes	Yes	Yes	Yes	Yes
Include weather controls	Yes	Yes	Yes	Yes	Yes
Mean Yields	3005.643	3005.643	3005.643	3005.643	3005.643
Obs.	2038	2070	2071	2057	2069
R-sq	0.588	0.585	0.583	0.585	0.586

Each column reports the results of an OLS regression. The unit of observation is a cell-year. The dependent variable is mean plot-level yields in kg/ha across all plots falling within a cell, in a given year. *Mean NDVI* is the average NDVI in the cell of size 2500m containing the plot, across days in a particular season during the year that maize on the plot was harvested. The columns indicate the season for which NDVI is calculated during the year the plot was harvested. *Mean NDVI* is scaled so that one unit represents a 0.01 increase in NDVI. Each regression includes linear controls for cell-level averages of growing degree days, temperature and precipitation during the maize season across plots falling within the cell, as well as cell-level controls for mean temperature, precipitation, vapor pressure, cloud cover, evapotranspiration and wet days. Year and cell-level fixed effects are included in the regressions.

Table 22: Relationship between Cell-level NDVI and Cell-level Yields, cell-size = 5000m

	(1) Planting	(2) Early Growing	(3) Late Growing	(4) Harvest	(5) Nonfarm
Mean NDVI	11.93	54.57***	18.00	46.21***	34.77***
	(11.11)	(15.96)	(14.47)	(11.52)	(10.03)
Year FE	Yes	Yes	Yes	Yes	Yes
Grid cell FE	Yes	Yes	Yes	Yes	Yes
Include weather controls	Yes	Yes	Yes	Yes	Yes
Mean Yields	3005.643	3005.643	3005.643	3005.643	3005.643
Obs.	2016	2040	2038	2036	2040
R-sq	0.551	0.559	0.552	0.557	0.553

Each column reports the results of an OLS regression. The unit of observation is a cell-year. The dependent variable is mean plot-level yields in kg/ha across all plots falling within a cell, in a given year. *Mean NDVI* is the average NDVI in the cell of size 5000m containing the plot, across days in a particular season during the year that maize on the plot was harvested. The columns indicate the season for which NDVI is calculated during the year the plot was harvested. *Mean NDVI* is scaled so that one unit represents a 0.01 increase in NDVI. Each regression includes linear controls for cell-level averages of growing degree days, temperature and precipitation during the maize season across plots falling within the cell, as well as cell-level controls for mean temperature, precipitation, vapor pressure, cloud cover, evapotranspiration and wet days. Year and cell-level fixed effects are included in the regressions.