

# DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

ISSN 1440-771X

## WORKING PAPER SERIES

# Downstream Impacts of Mines On Agriculture in Africa

Lukas Vashold, Gustav Pirich, Maximilian Heinze, Nikolas Kuschnig

Working Paper No. 09/25  
Nov 25

# Downstream Impacts of Mines on Agriculture in Africa

Lukas Vashold<sup>†\*</sup>    Gustav Pirich<sup>†</sup>    Maximilian Heinze<sup>†</sup>  
Nikolas Kuschnig<sup>†‡\*</sup>

<sup>†</sup> *Vienna University of Business and Economics*

<sup>‡</sup> *Monash University*

Mining operations in Africa are expanding rapidly, creating negative externalities that remain poorly understood. In this paper, we provide causal evidence for the impact of water pollution from mines on downstream vegetation and agriculture across the continent. We exploit a natural experiment, where mines cause a discontinuity in water pollution along river networks, to compare vegetation health in upstream and downstream locations. We find that mines significantly reduce peak vegetation downstream by 1.3–1.5%, impairing the productivity of over 74,000 km<sup>2</sup> of croplands. These reductions correspond to annual losses of 91,000–205,000 tons of cereal crops in the immediate vicinity alone, with particularly severe effects in fertile regions and areas where gold mining predominates. Our findings highlight substantial externalities of mining and demonstrate an urgent need for oversight and regulation.

---

\*Correspondence to [nikolas.kuschnig@monash.edu](mailto:nikolas.kuschnig@monash.edu), at 900 Dandenong Road, Caulfield East, Australia (equal contribution first authors). Funding by the Oesterreichische Nationalbank (OeNB anniversary fund, project No. 18799) is gratefully acknowledged.

## 1. Introduction

The mining sector in Africa is experiencing an unprecedented boom, driven by the demands of economic growth and the clean energy transition (Pörtner et al., 2022). This brings economic opportunities but also creates substantial negative externalities. Mining has been linked to pollution, environmental degradation, and biodiversity loss, as well as corruption, conflict, and child labor that undermine the livelihoods of local communities.<sup>1</sup> These externalities are striking and may have far-reaching impacts that could offset the economic benefits of mining, particularly in regions where extraction occurs. Insights into these impacts, however, remain scarce — especially in developing countries where weak institutions struggle to manage the impacts of mining, and external costs often go unmitigated.

In this paper, we provide causal evidence for the impacts of water pollution from mines on vegetation and agriculture in Africa. We exploit a natural experiment for identification, using the discontinuous jump in water pollution at mine locations along directed river networks. This leads to a sharp drop in vegetation health downstream of mines, while upstream areas remain unaffected and can serve as control groups. To implement this regression discontinuity design, we analyze outcomes across fine-grained river basins. We measure vegetation health and agricultural productivity across Africa by using satellite-derived peak vegetation indices combined with land use masks. This approach allows us to estimate the causal effects of water pollution from mines at an unprecedented geographic scale while maintaining rigorous causal identification.

Our research design identifies impacts of mines on vegetation that are transmitted through water.<sup>2</sup> The primary mechanism is water pollution: mines cause acidification, elevated salinity, and heavy metal contamination. These affect vegetation and agricultural productivity by disrupting plant physiology, soil microbiomes, and nutrient uptake. Secondary mechanisms involve farmers' adaptation in response to water pollution, such as relocating or switching crops. Importantly, our research design

---

<sup>1</sup>See, for instance, Aska et al., 2024; Berman, Couttenier, and Girard, 2023; Berman, Couttenier, Rohner, et al., 2017; Giljum et al., 2022; Girard, Molina-Millán, and Vic, 2025; Knutsen et al., 2016; Macklin et al., 2023; Santana et al., 2020.

<sup>2</sup>Our identification strategy has notable parallels in the literature. Most closely related, Gittard and Hu (2024) investigate the impacts of water pollution from industrial mining on health outcomes in Africa. Sigman (2002) and Lipscomb and Mobarak (2017) document water pollution spillovers across and within national borders, while Dias, Rocha, and Soares (2023) and E. Strobl and R. O. Strobl (2011) use similar river-based discontinuities to study agricultural impacts in Brazil and Africa.

controls for other impacts of mining — such as air pollution or economic effects — that do not follow river flows, allowing us to directly isolate the water channel.

The main contribution of our paper is the causal estimate of water-mediated mining impacts on vegetation and agriculture — an important externality that has previously been neglected. Earlier research has investigated various other impacts of mining<sup>3</sup> and many studies document pollution from mine sites.<sup>4</sup> While hydrological studies suggest that extensive areas and millions of livelihoods are affected by mining-induced river pollution (Macklin et al., 2023), causal evidence on this channel remains scarce and limited in scope. This paper addresses this gap by providing causal evidence for negative effects on vegetation across diverse mining contexts throughout Africa.

Our study contributes to several related literatures. Research on water pollution in developing contexts has typically focused on drinking water (e.g., Olmstead, 2010). We extend this literature by highlighting that clean water functions as a broad ecosystem service, supporting vegetation health and agricultural productivity (Keeler et al., 2012; Van Vliet, Flörke, and Wada, 2017). Our work also contributes to the growing body of research using remote-sensing methods to assess mining impacts in a data-scarce environment (Giljum et al., 2022; Maus and Werner, 2024), demonstrating how satellite data can enable causal analysis where traditional data sources are unavailable. Additionally, our findings relate to global environmental justice, as mineral extraction in lower-income African countries often serves demand from higher-income nations while imposing local environmental costs (see, e.g., Banzhaf, Ma, and Timmins, 2019).

We find that mines have substantial negative effects on downstream vegetation. Overall, peak vegetation indices decline by approximately 1.3 percent, while croplands experience reductions of about 1.5 percent. These effects are economically meaningful given their geographic scope — even the immediately impacted area, within 33 km downstream of mines, stretches over 280,000 km<sup>2</sup>, including 74,000 km<sup>2</sup> of croplands. We estimate that these reductions cause annual losses of 91,000–205,000 tons of cereal crops. Moreover, there is suggestive evidence that effects may extend much further downstream. Analysis of impact heterogeneity reveals that the largest negative effects occur in fertile regions and areas dominated by gold mining.

---

<sup>3</sup>Aragón and Rud, 2015; Aska et al., 2024; Berman, Couttenier, Rohner, et al., 2017; Gittard and Hu, 2024; Goltz and Barnwal, 2019; Ofosu et al., 2020.

<sup>4</sup>Awotwi et al., 2021; Du et al., 2024; Duncan, 2020; Mulenga et al., 2024; Wu et al., 2023.

Our findings have important practical implications. Agriculture plays a vital role in local livelihoods and economies across Africa, and results show that large, fertile, populated areas are affected by water pollution from mining. These impacts can cause economic and nutritional disruptions, potentially exacerbating food insecurity challenges across the continent. From a policy perspective, our study highlights inadequate oversight of the mining industry and the scarcity of data to assess their impacts.<sup>5</sup> Remote-sensing approaches can help address data gaps efficiently but rely on conventional data for calibration and are constrained in the types of effects that can be detected.

The remainder of this paper proceeds as follows. In the next section, we provide background and intuition for our analysis. In [Section 3](#), we describe our data and empirical strategy. [Section 4](#) presents the main results, heterogeneity analysis, and robustness checks. We discuss our results and their implications in [Section 5](#) and conclude in [Section 6](#).

## **2. Background and Intuition**

This section introduces key components of our study and explains the intuition behind our research design. We examine how mining operations affect vegetation and agriculture through water pollution, describe our unit of observation, river basins, and the satellite-derived vegetation indices.

### **2.1. Mines and mining in Africa**

Mineral extraction is a growing sector in many African economies, presenting both opportunities and challenges. While mining can generate employment and stimulate local economic development (Bazillier and Girard, 2020; Goltz and Barnwal, 2019; Gräser, 2024; Ofosu et al., 2020), it also brings substantial risks. Resource wealth may crowd out other industries, drive corruption and conflict (Berman, Couttenier, and Girard, 2023; Berman, Couttenier, Rohner, et al., 2017; Knutsen et al., 2016), and mining drives rapid environmental changes (Aska et al., 2024; Barenblitt et al., 2021; Giljum et al., 2022; Girard, Molina-Millán, and Vic, 2025) and pollution with considerable impacts on people and their environments (Awotwi et al., 2021; Macklin et al., 2023).

---

<sup>5</sup>Also see Auld, Betsill, and VanDeveer, 2018; Jones et al., 2024; Maus and Werner, 2024.

Vegetation and agriculture are affected through multiple channels, with water pollution playing a particularly important role (Santana et al., 2020). Mining requires copious amounts of water for processing and disturbs orders of magnitude more material than the metal eventually extracted. This creates erosion and generates waste materials, including sediment, rock, and heavy metals, that are stored as ‘tailings’. These tailings, oxidized by air and weathered by rain, steadily cause pollution of water resources as they feed into rivers (Schwarzenbach et al., 2010) and are well-known ecological risks (Wu et al., 2023). Combined with the release of toxic chemicals used for processing,<sup>6</sup> mining can induce water pollution from heavy metals (Frossard et al., 2018), increase salinity (Russ et al., 2020; Zörb, Geilfus, and Dietz, 2019), and cause acidification (Du et al., 2024) — all of which impact vegetation health via plant physiology and growth.<sup>7</sup> These impacts can persist for hundreds of years, degrade water quality, and devastate terrestrial and aquatic ecosystems (Du et al., 2024; D. B. Johnson and Hallberg, 2005).

Mines also affect vegetation in ways other than water pollution. Air pollution, particularly from coal mining and processing of metals in smelters, is a prominent example (Fugiel et al., 2017; Miao, Huang, and Song, 2017; Pandey, Agrawal, and Singh, 2014). For Ghana, Aragón and Rud (2015) find that air pollution from gold mining in Ghana reduced total factor productivity in the agricultural sector by 40%, mostly through direct effects on crop health by air and the resulting soil pollution. Socioeconomic factors constitute another pathway through which mines affect agriculture and vegetation. Corruption, conflict, and weak institutions have been linked to mining and are detrimental to agricultural productivity (Wuepper et al., 2023). Pollution also has negative impacts on labor supply (Hanna and Oliva, 2015) and productivity (Graff Zivin and Neidell, 2012), human capital accumulation (Currie et al., 2009), and mining industries affect local income and employment (see, e.g., Bazillier and Girard, 2020; Gräser, 2024; Kotsadam and Tolonen, 2016). These impacts may, in turn, affect agricultural productivity.

The risks of mining are particularly acute in Africa, where oversight is often lacking (Macklin et al., 2023). International regulations for the mining industry have been

---

<sup>6</sup>Sodium cyanide and mercury, for example, are heavily used in the (artisanal) extraction of gold (Duncan, 2020; Malone et al., 2023; Verbrugge, Lanzano, and Libassi, 2021) and can cause long-term alterations of sediment chemistry.

<sup>7</sup>See [Appendix B1](#) for a literature overview on the effects of water pollution on plant health.

scarce,<sup>8</sup> while the prevalent small-scale, artisanal mines often operate outside regulatory frameworks. Countries in Africa often lack the robust institutional frameworks necessary to monitor and regulate expanding mining operations and face severe challenges in managing detrimental externalities. Concerns are further emphasized by the heavy reliance on subsistence agriculture, looming food insecurity, and the potential long-term consequences of environmental degradation (see, e.g., Audry et al., 2004; D. B. Johnson and Hallberg, 2005).

## **2.2. River basins and water streams**

Our unit of observation is the river basin, defined as the area drained by a river and its tributaries. Basin boundaries are defined by topographic features, like ridges and hills, ensuring that all surface waters converge to a single point.<sup>9</sup> This definition can be applied recursively, and basins at local to continental scales can be nested in a hierarchical system. Our analysis uses the finest scale of basins (in the dataset of Lehner and Grill, 2013), illustrated in the left panel of Figure 1. Their small size conveys localized impacts while ensuring that relevant factors vary smoothly and discontinuities can be attributed to mine sites themselves. Coupled with the unidirectional flow of water down river networks, this allows us to implement our design: downstream basins may be affected by contaminated water, but upstream basins remain unaffected and can serve as controls.

## **2.3. Agricultural productivity from space**

We use satellite-derived vegetation indices as proxies for vegetation health and agricultural productivity. In our context, where smallholder farming predominates, traditional datasets based on surveys or censuses face considerable challenges and lack consistent coverage across African countries. We construct peak vegetation indices, which correlate strongly with gross primary production of vegetation (D. M. Johnson, 2016; Shi et al., 2017) and are frequently used in similar analyses (see, e.g., Asher and Novosad, 2020; Boudot-Reddy and Butler, 2025; Wuepper et al., 2023), using

---

<sup>8</sup>A recent exception is the 2020 Global Industry Standard on Tailings Management (GISTM; UNEP, 2023), which was established to prevent future tailings dam failures after the catastrophic dam collapse at Vale's Corrego de Feijao mine in Brazil.

<sup>9</sup>Two prominent examples of large basins are the Mississippi River, which drains most of the mainland United States into the Gulf of Mexico, and the Congo Basin, which covers most of the Democratic Republic of the Congo (DRC) and drains into the Atlantic.

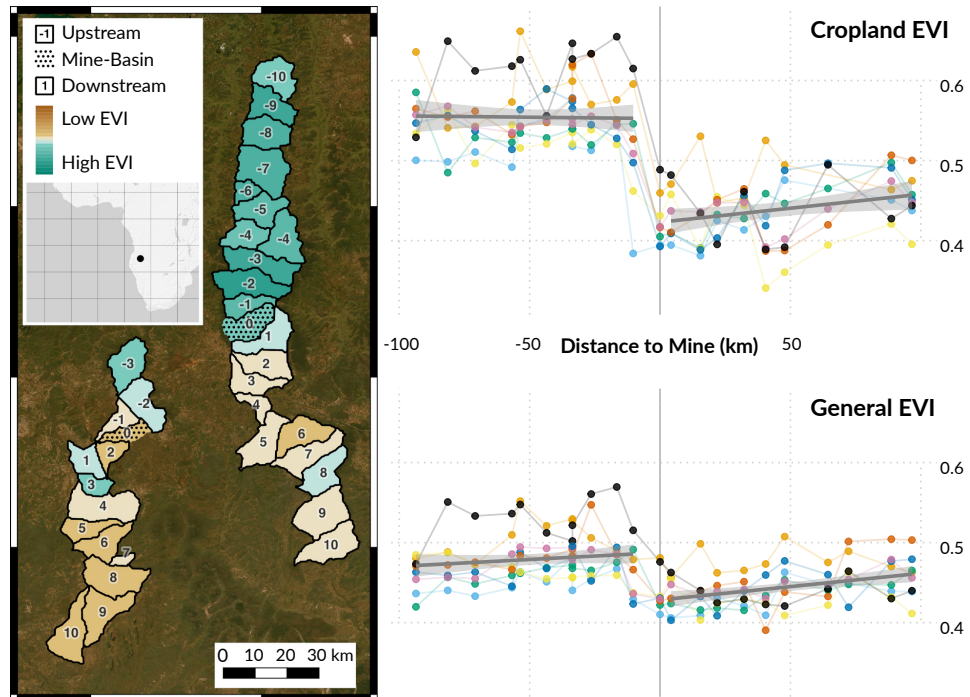


FIGURE 1: Example of two Angolan mine sites (dotted, and labeled with '0') and their upstream and downstream basin systems (left), as well as measurements of the Enhanced Vegetation Index (EVI) for croplands and general vegetation over the years 2016, 2017, 2018, 2019, 2020, 2021, 2022, and 2023 along the larger basin system (right).

vegetation and cropland masks. These consistently capture outcomes of interest with high granularity across time and space. An example of both masked measurements is provided in [Figure 1](#), demonstrating how the indices capture vegetation patterns up- and downstream of a mining site.

## 2.4. Research design

Our empirical strategy integrates mining sites, river basins, and remotely sensed vegetation indices to identify causal effects of mines on downstream agricultural productivity and vegetation health across the African continent. [Figure 2](#) illustrates our approach. We identify basins containing mine sites and compare vegetation in basins located upstream and downstream within river networks. This allows us to isolate the directed river-mediated impact by controlling for undirected effects of mines, thereby identifying the causal effect mediated by water flows.

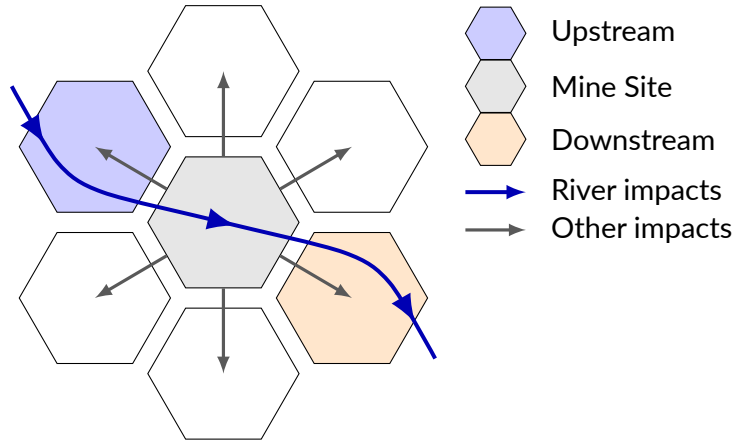


FIGURE 2: Illustration of the research design. The comparison of up- and downstream basins enables the identification of mine impacts that are mediated by the river.

Our strategy relies on the assumption that the observed discontinuities along rivers stem from mines themselves. This is plausible when mine locations are quasi-random rather than driven by factors related to vegetation or agriculture. The location of mines is mainly determined by accessible mineral deposits. Potential confounders exist (e.g., strategic placement near transportation) but are unlikely to confound our estimates because their impacts would need to align with water flow direction and affect mine placement at the basin level.

We primarily focus on the immediate vicinity of mines, where basin characteristics are most comparable and our research design delivers strong causal identification. The reach of impacts over greater distances presents additional challenges, as fundamental differences may emerge from geographical patterns. For instance, other pollution sources are more likely located downstream where river transport is economically attractive. Such factors are unlikely to affect the discontinuity that we exploit for identification but could confound our assessment of impact decay.

The research design captures mining impacts on vegetation mediated through water systems, offering a focused but partial view of externalities. Direct impacts on-site and impacts that are not directed along the river (such as dust deposition, employment effects shifting labor, or income effects allowing for fertilizer use) are not captured by our design. Instead, mediators for the effect of interest include both direct and indirect effects of water pollution. While downstream vegetation is immediately affected, human responses to these impacts, resulting soil pollution, and other second-order effects are also reflected in our estimates.

### 3. Materials and Methods

This section describes the dataset and methods used. First, we describe the basin-level dataset and how it is constructed, and elaborate on individual variables included. Then, we present our empirical strategy.

#### 3.1. Dataset

Our dataset is a balanced panel of  $N = 14,334$  river basins in Africa, observed over eight years from 2016–2023, delineated by Lehner and Grill (2013).<sup>10</sup> For our analysis, we classify basins into three categories:

1. mine-basins, with a mine site in their catchment area,
2. downstream basins, which are downstream of *any* mine-basin,
3. upstream basins, which are upstream of *all* connected mine-basins.

---

<sup>10</sup>The dataset has a hierarchical structure with twelve nested levels, where basins at each level are of comparable size. We use the finest Level 12, with 241,026 basins on the African continent that cover an average ( $\{5, 50, 95\}^{\text{th}}$  percentile) area of 124.4 (11.6, 131.4, 218.9) km<sup>2</sup>, and feature an average diameter of 12.6 km and elevation differences of eleven meters.

We consider up- and downstream chains of basins (departing from mine-basins) with a maximum order of ten which lie at an average river distance of 105 km.

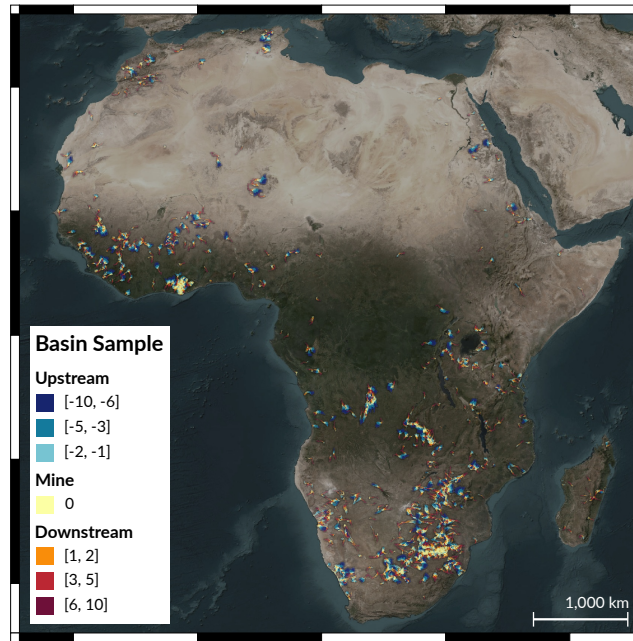


FIGURE 3: Basins in the sample and their treatment status. Basemap imagery provided by Esri, Maxar, Earthstar Geographics, and the GIS User Community.

We observe a total of 1,900 mine-basins, 6,307 upstream and 6,127 downstream basins that are visualized in Figure 3. (For a summary of orders and distance of basins, see Table E3 in the Appendix.) Their average size is 120 km<sup>2</sup>, for a total area of 1,701,343 km<sup>2</sup>. Each basin is unique but may be related to multiple mine-basins. In this case, we associate the basin with the nearest mine-basin. The next (i.e., order one) downstream basin is unique for each basin, but there may be multiple upstream basins of any given order. Not all mine-basins have a full set of up- or downstream basins, and the number decreases with order (see Figure D7 for an overview). More details on the basin data are provided in Appendix A.

**Summary statistics** Table 1 presents summary statistics for the most important variables used.<sup>11</sup> This includes vegetation indices, geographical information on elevation and slope, meteorological information on temperature and precipitation, as well as

<sup>11</sup>See Table E4 in the Appendix for summary statistics that are split by treatment status.

population and accessibility to urban areas. Variables are mapped from a raster level to the level of basins, with the exact procedures described below.

TABLE 1: Summary statistics for the basin-level dataset.

| Variable           | Unit            | $NT$    | Mean  | St. Dev. | Min.   | Max.      |
|--------------------|-----------------|---------|-------|----------|--------|-----------|
| Peak Vegetation    | Index $[-1, 1]$ | 110,576 | 0.428 | 0.154    | 0.016  | 0.993     |
| Mean Vegetation    | Index $[-1, 1]$ | 110,576 | 0.279 | 0.112    | -0.021 | 0.578     |
| Peak Cropland Veg. | Index $[-1, 1]$ | 93,036  | 0.464 | 0.133    | -0.068 | 0.978     |
| Mean Cropland Veg. | Index $[-1, 1]$ | 93,036  | 0.298 | 0.101    | -0.104 | 0.601     |
| Elevation          | Meters          | 110,568 | 820.4 | 481.1    | -118.3 | 3,059.7   |
| Slope              | Degrees         | 110,568 | 2.23  | 2.34     | 0.0    | 20.9      |
| Max. Temperature   | Degree Celsius  | 110,572 | 34.3  | 3.9      | 15.6   | 48.8      |
| Precipitation      | Millimeters     | 110,576 | 901.8 | 595.2    | 0.64   | 4,456.7   |
| Population         | Capita          | 110,576 | 8,471 | 37,716   | 0.0    | 1,396,921 |
| Accessibility      | Minutes         | 110,528 | 164.3 | 179.1    | 1.0    | 2,659.9   |

### 3.1.1. Vegetation and agriculture

To measure agricultural productivity and vegetation health, we create peak vegetation indices for areas with general and cropland-specific vegetation. We base our measures on the remotely sensed Enhanced Vegetation Index (EVI), which we combine with appropriate land use masks for vegetation and agriculture.<sup>12</sup>

We process the raw EVI data (Didan, 2015, available every 16 days at a 250 m resolution) by (1) applying quality filters to minimize the impact of clouds, (2) masking with yearly land use from the ESA CCI, (Defourny et al., 2024, available at a 300 m resolution) to identify relevant areas for the two outcome measures, (3) aggregating the mean EVI at 16-day intervals per basin, and (4) computing the annual maximum EVI as a peak vegetation index. The resulting peak indices correlate strongly with measures of gross primary production of vegetation (see Shi et al., 2017, for an assessment) and crop yields (see, e.g., Azzari, Jain, and David B. Lobell, 2017; D. M. Johnson, 2016). We consider several alternatives, including different land use masks (Digital Earth Africa, 2022) and aggregation schemes, as robustness checks.

<sup>12</sup>The EVI's advantages (see Zeng et al., 2022, for a recent review) include high spatial granularity, frequent availability, and consistency over time and space. Moreover, it is less susceptible to atmospheric distortions and background noise than the commonly used Normalized Difference Vegetation Index (NDVI) and can accurately convey variation in low as well as high biomass conditions (Huete et al., 2002).

### **3.1.2. Treatment**

Our treatment is the incidence of a local mining site, derived from Maus, Giljum, et al. (2022). Their dataset integrates multiple sources, including the SNL Metals & Mining database, and provides comprehensive coverage of both industrial and artisanal mining sites. It offers two key advantages: (1) locations are verified through satellite imagery (2) mining areas (including features such as tailings dams, waste rock dumps, water ponds, and processing plants) are delineated within 10-kilometer buffers around known locations. This way, smaller artisanal mining operations that often develop near larger industrial sites are captured, and potential attenuation from missing or misclassified mine locations is reduced. The dataset lacks detailed mine characteristics and temporal information beyond the 2019 anchor, which we address via supplementary sources (inter alia, Jasansky et al., 2023; Padilla et al., 2021; S&P Global Market Intelligence, 2025; Sepin, Vashold, and Kuschnig, 2025).

We use data on mining sites to construct our treatment variable as the distance to the nearest upstream mine-basin. For our primary specification, we consider the ordinal position of basins relative to the mine-basin as the treatment variable; as an alternative, we use different specifications based on river distance. We work with basins up to order ten but focus our analysis on the immediate vicinity (i.e., the first three basins in either direction) for comparability. The mine-basin itself can be considered treated, but on-site effects are not identified by our research design and thus not interpreted by us.

### **3.1.3. Covariates**

We consider several covariates that could confound our analysis or provide additional insights. Our selection focuses on geographical characteristics that relate directly to river basins and variables with potential imbalances across treatment and control groups.

**Geographic controls** For elevation and slope — key variables that define water flows and basin boundaries — we use high-resolution gridded data (Amatulli et al., 2018), aggregating from 30 arcsecond resolution (approximately 802–926 meters) to basin-level averages. We determine the primary soil type for each basin using 250-meter resolution data from the SoilGrids project (Hengl et al., 2017), to account

for soil characteristics that influence vegetation health and may vary systematically along river networks.

**Climate and socioeconomic controls** These factors become important at larger distances from mine sites, where treatment and control groups may diverge systematically. For the climate, we use precipitation data from CHIRPS (Dinku et al., 2018; Funk et al., 2015), and temperature data from TerraClimate (Abatzoglou et al., 2018), with meteorological data from CRU serving as an alternative for robustness checks (Harris et al., 2020). Both primary sources provide monthly measurements at 4–5 kilometer resolution. We use annual precipitation totals and maximum monthly temperatures, following established practices in the literature. For socioeconomic conditions, we include total population (WorldPop, 2018) and average accessibility, measured by the travel time to the nearest city (Weiss et al., 2018), for each basin. Both are aggregated from a 1-kilometer resolution using values from 2015 to control for initial conditions.

**Other data** We also investigate relations to the number of violent events from ACLED (Raleigh et al., 2010), average annual air pollutant concentrations (from Shen et al., 2024), nighttime lights (Elvidge, Baugh, et al., 2017), rates of forest loss (Hansen et al., 2013), and water pollution (United Nations Environment Programme, 2025).

### 3.1.4. Heterogeneity

We investigate heterogeneity of impacts across four main dimensions. First, we consider the Ecoregion classification (Dinerstein et al., 2017), which divides regional ecosystems with different types of agriculture and vegetation. We also consider country groups based on their primary crops and varying crop calendar cycles following the US Department of Agriculture (USDA).<sup>13</sup> Second, we proxy for mining intensity via the total mine area per mine-basin, and use the development of mining areas over time to proxy for activity.<sup>14</sup> Third, we investigate impacts by mineral type, collating various

---

<sup>13</sup>The classification (available at [ipad.fas.usda.gov](https://ipad.fas.usda.gov)) divides the continent into North Africa, Southern Africa, East Africa, and West Africa; due to limited observations, we merge North and East Africa.

<sup>14</sup>For the latter, we rely on a longitudinal extension of the mine dataset by Sepin, Vashold, and Kuschnig (2025), who automatically delineate mining sites over time using a segmentation model and high-resolution satellite imagery. The temporal information allows us to identify active mines based on their growth and assess their specific impacts.

data sources to predict the most likely commodities mined in each basin (detailed in [Appendix C2](#)). Lastly, we examine heterogeneity across agricultural suitability based on the Global Agro-Ecological Zones (GAEZ) modeling framework (Fischer et al., 2021). For each basin, we calculate the average suitability for thirty crop types and categorize basins into high, medium, and low productivity based on the suitability of the best-suited crops.

### 3.2. Empirical strategy

We employ a regression discontinuity design (RDD) to identify the causal effects of mines on vegetation and agricultural productivity downstream. Our identification strategy exploits the natural discontinuity in the exposure to pollution along river networks. Basins downstream of mines are exposed to contaminated water, while upstream basins remain unaffected and can serve as controls.

Our application differs from standard RDDs in minor but notable ways. Our running variable — the distance to the mine-basin — has a natural direction and coincides with the treatment intensity. This directional feature strengthens our identification strategy, while the latter presents an opportunity to investigate impact decay. We also observe outcomes at the mine-basin, which is immediately impacted by contaminated water but also by third effects of the mine. These on-site observations provide useful information and suggestive evidence for other impacts of mine sites, but they are not identified by our research design and, hence, not interpreted.

Let  $x$  denote the *directed distance* from the mine along the river, where we have  $x = 0$  at the mine-basin and  $x < 0$  before it. Following the notation of Gelman and Imbens (2019), we can express the treatment effect as

$$\tau(x) = \mathbb{E} [y_i(1) - y_i(0) \mid x_i = x],$$

where  $y_i(1), y_i(0)$  are potential outcomes of observation  $i$  under treatment and control conditions. We exploit variation in  $x$  for identification and estimate the average treatment effect using the difference in conditional means between downstream and upstream locations:

$$\tau(x)_{ATE} = \mathbb{E} [y_i^{\text{obs}} \mid x_i = x, x > 0] - \mathbb{E} [y_i^{\text{obs}} \mid x_i = x, x < 0].$$

Our estimation equation is

$$y_{imt} = \beta' F(x_i) + \theta' W_{it} + \mu_m + \psi_t + \varepsilon_{imt}, \quad (1)$$

where  $y_{imt}$  represents vegetation health in basin  $i$  near mine  $m$  at time  $t$ , and  $x_i$  is the running distance. The vector  $W_{it}$  contains basin characteristics,<sup>15</sup> while  $\mu_m$  and  $\psi_t$  represent mine- and year-fixed effects. The error term  $\varepsilon_{imt}$  is clustered by mine. We operationalize distance via the function  $F(\cdot)$  that returns a vector separating upstream, downstream, and mine-basins.

### 3.2.1. Specification

Our preferred specification uses the basin order as distance, i.e.,  $x_i \in \{-10, \dots, 10\}$ , and operationalizes it via indicators. This approach highlights basin-to-basin discrepancies and remains agnostic about functional forms, instead relying on local information at the level of our observations. Specifically, we let  $F(\cdot)$  return indicators

$$f(x)_j = \mathbb{I}(x = j) \text{ for } j \in \{-10, \dots, -2, 0, 1, \dots, 10\}.$$

Here, we omit the first upstream basin (order  $-1$ ) as the reference category. Our design identifies differences between estimates at equivalent (absolute) distance (e.g.,  $\tau(8) = \beta_8 - \beta_{-8}$ ), which we report as effects of interest. For ease of interpretation, we also report pooled estimates comparing the first three basins.<sup>16</sup>

To assess the decay of impacts systematically, we specify an alternative model using the river distance in kilometers as the running variable. We operationalize this as

$$F(x) = \exp\{-\gamma \times |x|\} + \mathbb{I}(x = 0) + \exp\{-\delta \times x\} \times \mathbb{I}(x > 0),$$

where  $+$  separates variables and the parameters  $\gamma, \delta$  capture the rates of exponential impact decay. The exponential form assumes that impact decay is proportional to the impact level, a common assumption in analyses of water pollution. Compared to linear and linear-quadratic functional forms (which we also consider), this functional form allocates more influence to observations at shorter distances and not vice versa.

<sup>15</sup>Of the main set of covariates, only the meteorological variables are time-varying; the other variables are time-invariant (geophysical) or fixed at pre-period levels (socioeconomic).

<sup>16</sup>We let  $f(x)_{\mathcal{J}} = \mathbb{I}(x \in \mathcal{J})$  for  $\mathcal{J} \in \{-10, \dots, -4, 0, \{1, 2, 3\}, 4, \dots, 10\}$ , and focus on  $\beta_{\{1,2,3\}}$ .

### **3.2.2. Identifying assumption**

Our key identifying assumption posits no other discontinuous changes along the river that impact vegetation at the mine-basin. This is supported by the nature of our observations. River basins are defined by natural geographic features that vary smoothly over space, and the granular basin level (Level 12) minimizes systematic differences between adjacent basins.

To validate our approach, we conduct several robustness checks. We control for extensive basin characteristics (geophysical, meteorological, and socioeconomic), use these variables as outcomes to test for spurious discontinuities, and implement covariate matching to ensure that up- and downstream basins are comparable. We also vary the sample considered by excluding mine-basins, restricting to first-order basins, subsetting to narrower periods, and randomizing the treatment assignments as a placebo.

## **4. Results**

Our analysis reveals that mines significantly reduce vegetation health in downstream areas through water networks. We begin by focusing on effects in the immediate vicinity of mines, i.e., downstream basins up to order three, for which causal identification is strongest. In [Section 4.1](#), we investigate potential heterogeneities in these impacts to better understand the underlying mechanisms. Next, we extrapolate beyond the immediate vicinity and investigate the reach and decay of effects in [Section 4.2](#). Finally, we assess the robustness of our estimates in [Section 4.3](#).

[Table 2](#) presents estimates of the impact of mines on general (left columns) and cropland-specific (right columns) vegetation downstream, measured via a peak vegetation index, based on the EVI (Enhanced Vegetation Index). For each outcome, we present estimates from both a plain specification without covariates and a fully saturated specification with controls. The reported estimates represent the causally identified average treatment effect (ATE) for downstream areas, i.e., the difference in vegetation health between downstream (treated) and upstream (control) basins at equivalent distances from mine-basins.

The upper panel of [Table 2](#) shows results for individual basins up to order five, while the lower panel pools the first three downstream basins to provide a more statistically powerful and readily interpretable estimate of the immediate effect. Note that we do

not report and interpret estimates for the mine-basin itself, as they are not identified by our research design. Complete results, with estimates for all basin orders and covariates, are available in [Table E5](#) in the Appendix.

The treatment indicators show statistically significant negative effects on vegetation downstream of mines compared to vegetation upstream for both indices of general and cropland-specific vegetation. The (gradual) inclusion of covariates changes estimates only marginally (see [Table E5](#) for extended results), suggesting that findings are not driven by confounding factors. Estimates suggest that the effect permeates the immediate vicinity and may extend beyond this area with a similar magnitude, although the precision and statistical significance decline with distance.

For general vegetation, the estimated ATEs for the first three basins downstream of the mine are  $-0.0057$  and  $-0.0056$  (with and without covariates) and highly significant ( $p < 0.01$ ). This is mirrored by estimates for individual basins, where the ATEs for the first three basins are significant ( $p < 0.05$ ) and range from  $-0.0043$  to  $-0.0087$ . Estimates for the subsequent basins, including the fourth and fifth downstream basins, are consistently negative but show a drop of statistical significance and a minor drop in magnitude. The reductions in the immediate vicinity correspond to a 1.28–1.35% decrease relative to the sample mean. The impacted area, i.e., vegetation in the first three downstream basins, stretches across 255,000 km<sup>2</sup>.

For croplands, the estimated ATEs are slightly larger. For the first three basins downstream of the mine, estimates are  $-0.0064$  and  $-0.0068$  (with and without covariates) and significant ( $p < 0.05$ ). Individual-level estimates for these basins range from  $-0.0050$  to  $-0.0099$  and are significant ( $p < 0.05$ ) with one exception. Results for higher-order basins are similar to those for maximum vegetation, but the drops in magnitude and precision of estimates is larger. Compared to the sample mean, these estimates imply an index reduction of 1.38–1.47% over an affected cropland area of 74,000 km<sup>2</sup>.

## **Contextualizing impacts**

These effect sizes align with other studies using vegetation indices to approximate agricultural productivity across various contexts, which we discuss in [Section 5](#). To translate these reductions in peak vegetation indices into agricultural terms, we correlated our measure with direct measures of agricultural productivity from LSMS-ISA and AReNA-DHS (see [Appendix B2](#)). Based on the AReNA-DHS data, being imme-

TABLE 2: Main estimation results.

| <i>Outcome</i><br>(Specification)              | <i>Peak Vegetation</i><br>(Plain)      (Full) |                        | <i>Peak Cropland Veg.</i><br>(Plain)      (Full) |                        |
|------------------------------------------------|-----------------------------------------------|------------------------|--------------------------------------------------|------------------------|
| <b>Individual Order</b>                        |                                               |                        |                                                  |                        |
| Downstream (1 <sup>st</sup> )                  | -0.0045***<br>(0.0017)                        | -0.0043**<br>(0.0018)  | -0.0051**<br>(0.0025)                            | -0.0050**<br>(0.0025)  |
| Downstream (2 <sup>nd</sup> )                  | -0.0049**<br>(0.0022)                         | -0.0048**<br>(0.0024)  | -0.0058*<br>(0.0031)                             | -0.0067**<br>(0.0032)  |
| Downstream (3 <sup>rd</sup> )                  | -0.0085***<br>(0.0028)                        | -0.0087***<br>(0.0029) | -0.0088**<br>(0.0037)                            | -0.0099***<br>(0.0038) |
| Downstream (4 <sup>th</sup> )                  | -0.0049*<br>(0.0030)                          | -0.0062*<br>(0.0033)   | -0.0029<br>(0.0038)                              | -0.0044<br>(0.0040)    |
| Downstream (5 <sup>th</sup> )                  | -0.0034<br>(0.0033)                           | -0.0053<br>(0.0037)    | 0.0007<br>(0.0042)                               | -0.0016<br>(0.0045)    |
| <b>Pooled Order</b>                            |                                               |                        |                                                  |                        |
| Downstream (1 <sup>st</sup> –3 <sup>rd</sup> ) | -0.0057***<br>(0.0018)                        | -0.0056***<br>(0.0020) | -0.0064**<br>(0.0025)                            | -0.0068***<br>(0.0026) |
| <i>Fit statistics</i>                          |                                               |                        |                                                  |                        |
| Observations                                   | 110,576                                       | 110,524                | 93,036                                           | 93,000                 |
| # upstream basins                              | 6,084                                         | 6,084                  | 5,232                                            | 5,232                  |
| # downstream basins                            | 5,899                                         | 5,896                  | 4,769                                            | 4,767                  |
| R <sup>2</sup>                                 | 0.903                                         | 0.908                  | 0.816                                            | 0.822                  |
| <i>Controls</i>                                |                                               |                        |                                                  |                        |
| Geophysical                                    | No                                            | Yes                    | No                                               | Yes                    |
| Meteorological                                 | No                                            | Yes                    | No                                               | Yes                    |
| Socioeconomic                                  | No                                            | Yes                    | No                                               | Yes                    |
| <i>Fixed-effects</i>                           |                                               |                        |                                                  |                        |
| Year (2016–2023)                               | Yes                                           | Yes                    | Yes                                              | Yes                    |
| Mine                                           | Yes                                           | Yes                    | Yes                                              | Yes                    |

**Note:** The table reports estimates of the average treatment effects by equidistant basin order based on Equation (1). The left columns hold results for the overall peak vegetation index; the right columns for the cropland-specific peak vegetation index. The first and third columns include no covariates, whereas columns two and four include the full set of control variables. Estimates in the upper panel correspond to the average effect at individual orders, while the lower panel reports a pooled estimate for the three basins that are immediately adjacent to the mine-basin. All specifications include mine and year fixed effects. Standard errors are (in brackets) and clustered at the mine level; significance levels are indicated as \*\*\*: 0.01, \*\*: 0.05, \*: 0.1.

diately downstream of a mine is associated with a 0.613% decrease in cereal yields, while the LSMS-ISA data indicates a 1.102% decrease.

The scale of these impacts becomes apparent when considering the affected area. The first three downstream basins cover 280,000 km<sup>2</sup>, of which 74,000 are croplands — an area equivalent to Ghana’s total cropland, falling between the cropland areas of the United Kingdom and Malaysia. Using average yields from both datasets and scaling up for all affected croplands, the estimated yield reductions translate into annual production losses of 91,000–205,000 tons of cereals. To contextualize these losses, the World Food Programme distributed approximately 1.7 million tons of food aid to over 38 million recipients across African countries in 2023. Our estimates of agricultural production losses from water pollution downstream of mines — covering just the immediate vicinity of mines — correspond to 5.4 to 12.1% of this major humanitarian food aid program.

#### **4.1. Heterogeneity**

Next, we investigate how our results vary along several dimensions to identify where mining impacts on vegetation and agriculture are most pronounced. We examine heterogeneity related to the characteristics of basins (biome, regions, and suitability for crop cultivation) as well as mine characteristics (total mining area, activity, and commodity type).

[Figure 4](#) provides an overview of the results. We present specifications with full controls and pooled estimates for the first three basins, allowing for full heterogeneity by re-estimating with different subsets of the sample. Estimation results are available in [Tables E7 to E8](#) in the Appendix.

##### **4.1.1. Environmental heterogeneity**

We first analyze spatial heterogeneities of basins, which are visualized in the upper half of [Figure 4](#). To check for differences across biomes, we assign each basin to one of three broad ecological groups — grasslands, forests, and deserts — that are aggregated from the more granular ecoregions of Dinerstein et al. (2017). We find negative effects for both the grassland and forest biomes. Grasslands constitute 69% of the vegetation sample and 74% of the cropland sample, while forests make up 15% and 17%. The effect sizes are comparable to the baseline, although impacts in

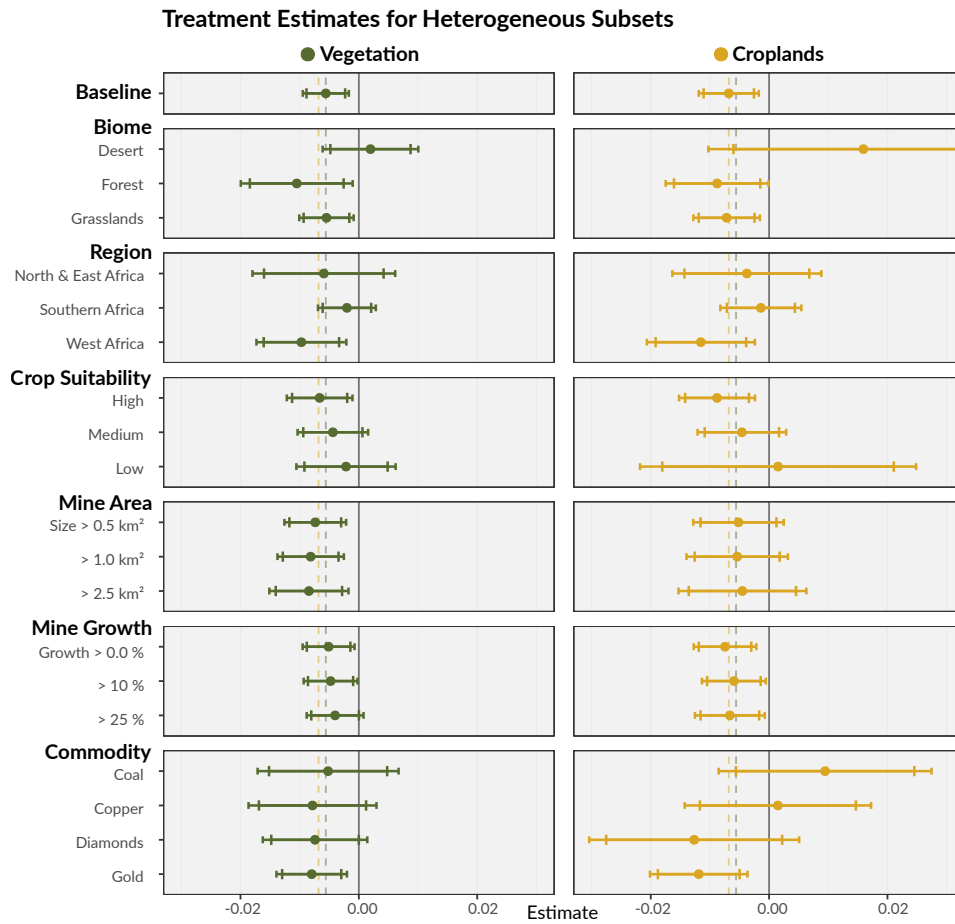


FIGURE 4: Dots indicate the average treatment effect in the first three basins; whiskered lines indicate 90% and 95% confidence intervals. Dashed vertical lines indicate point estimates for baseline specifications.

forest biomes appear somewhat larger. In desert biomes (representing 15% and 8% of the sample), we find no significant effect on either the cropland or overall peak vegetation, likely reflecting limited vegetation to be impacted in these regions.

For regional differences, we separate samples based on the classification system of the USDA.<sup>17</sup> We find substantial effects in West Africa, which represents 29% of the vegetation sample and 26% of the cropland sample. In Southern Africa (61% and 64%) or North & East Africa (10%), we find no significant effects. This regional disparity likely relates to mining practices and commodity types — artisanal gold mining is particularly common in West Africa. Additionally, our identification strategy may reach its limits in the case of South Africa, where a vast cluster of mines permeates entire basin systems, complicating the isolation of downstream effects.

The agricultural potential of affected regions captures another important dimension of impacts. We follow the GAEZ framework in categorizing basins into high, medium and low crop suitability groups, based on the maximum suitability across 30 crops. We find significant negative effects on basins with high crop suitability, which comprise 36% of the vegetation and 39% of the cropland sample. The point estimates for this sample exceed the baseline for both the general and cropland-specific vegetation. In contrast, estimates for medium (43% and 47%) and low suitability (22% and 13%) are smaller and statistically insignificant. This pattern is concerning because more than 35,000 km<sup>2</sup> (over a third) of highly suitable land in the immediately affected basins is actively used for crop production, raising concerns for subsistence farming and food security.

#### **4.1.2. Mine heterogeneity**

Turning to mine characteristics, we first examine whether mine size influences downstream impacts. We progressively restrict the sample to mine-basins with mining areas of at least 0.5 km<sup>2</sup>, 1 km<sup>2</sup>, and 2.5 km<sup>2</sup>, which reduces the sample to around 51%, 40%, and 25% of its original size. For overall vegetation, we observe an increase in effect size as mining area increases. For the 2.5 km<sup>2</sup> cutoff, the estimate is 50% higher compared to the baseline. This result aligns with expectations, as larger mines

---

<sup>17</sup>The system seeks to reflect differences in crop types and cycles. The regions are West Africa (with the Burkina Faso, Guinea, Mali, Ghana, Nigeria, and Niger as the most represented countries with over 250 basins), Southern Africa (with South Africa, Zimbabwe, Namibia, Tanzania, Botswana, Angola, the DRC, Zambia, and Mozambique), and North & East Africa (where only Morocco exceeds 250 basins), which we pool.

typically produce more discharge material and cause stronger contamination of water systems. For croplands, however, we find imprecise estimates with no discernible trend in magnitudes.

For mining activity, we use growth as a proxy and analyze subsets of active mines that exhibited (i) any growth, (ii) at least 10% growth, or (iii) at least 25% growth over the observed period. Our results show no substantial heterogeneity along this dimension for either natural vegetation or the cropland vegetation. Effect estimates remain stable, although precision decreases with smaller sample sizes (approximately 67%, 63%, and 55% of the baseline). This suggests that our dataset of mine locations, which is anchored in 2019, adequately captures the relevant impacts of mining sites regardless of their status. The lack of heterogeneity may also reflect offsetting factors: expanding mines might implement better precautions, while inactive, poorly maintained sites could still produce considerable pollution.

Finally, we examine how impacts vary by commodity types. We focus on gold, diamonds, copper, and coal — commodities that are well-represented in the data and which can reasonably be isolated from other commodities. Results reveal a strong negative effect for gold mining, which we isolated in 26% of the vegetation and 29% of the cropland sample. Compared to the baseline, point estimates for gold mining are 40% larger for vegetation overall and 75% larger for croplands. Among the other commodities — diamonds (13% of the sample), copper (6%), and coal (8%) — only diamond mining shows marginally significant negative effects on overall vegetation.

## **4.2. Impact decay**

We have established significant and economically meaningful downstream impacts of mining in the immediate vicinity. An important question remains, however: How far downstream do these effects persist? The spatial extent of effects is crucial for understanding the overall impact, gauging the size of the externality, and designing effective mitigation strategies. While critically important, this analysis extends beyond our identification strategy — at greater distances, basins may diverge in characteristics and results cannot be interpreted as strictly causal. Our analysis here provides suggestive evidence and highlights the potential overall stream-mediated impact of mining.

The range of impacts is related to the types of mediators that it arises from. Vegetation is affected by various types of water pollution from mining, including heavy

metals, acidification, and salinity (see [Appendix B1](#) for further details). While contaminants gradually disperse along streams, buildup and tipping points could lead to more persistent impacts. Acidification, for instance, is initially triggered by pollution from mines but is sustained by extremophile microbes. We empirically investigate the decay of impacts by (1) extrapolating our analysis beyond the immediate vicinity, and (2) imposing an exponential decay functional form to estimate the speed of decay.

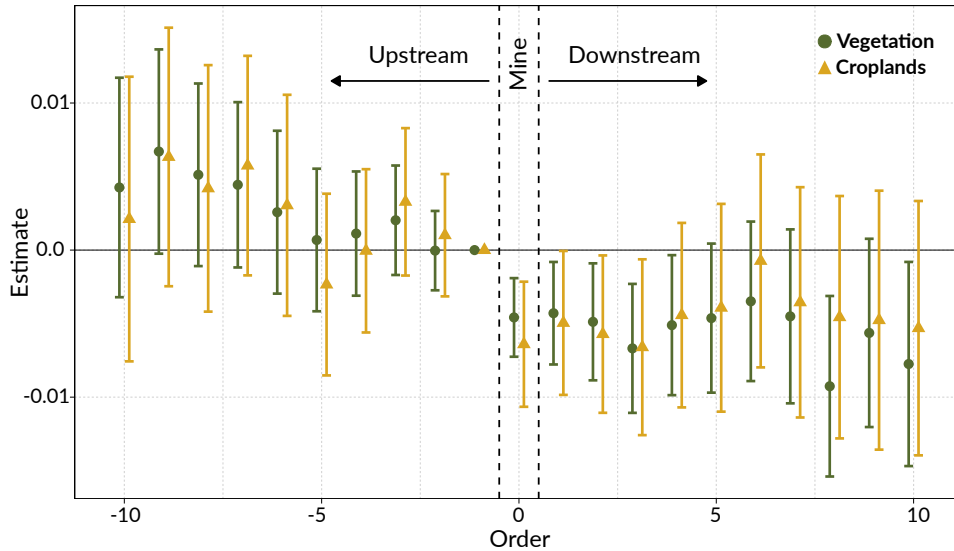


FIGURE 5: Estimated order coefficients for all up- and downstream basins (with the mine-basin in the center) for the overall peak vegetation index (green circle) and the cropland-specific index (yellow triangles) with full covariates. Whiskers represent 95%-confidence intervals.

Figure 5 visualizes estimates along basin orders, comparing all basins to the one immediately upstream of the mine (order -1).<sup>18</sup> If mines had no impacts, we would expect flat estimates across all basin orders. Non-directed mine impacts (e.g., from air pollution) would appear as a V-shape, where the right ‘wing’ of the V would be shifted by downstream effects and rotated if these effects decay with distance.

The figure reveals three notable patterns. First, negative impacts persist beyond the immediate vicinity of mines; effects appear to increase in size after the sixth basin. Second, trajectories differ slightly between general vegetation and croplands. Croplands suffer slightly larger impacts near mines but show more decay at larger

<sup>18</sup>The visualization differs from the coefficients in [Table 2](#) and [Table E5](#) (which report differences between equidistant up- and downstream basins, and not deviations from the first upstream basin) to better convey changes from basin to basin.

distances. Impacts on general vegetation are more persistent, yielding significant negative estimates even at large distances. Third, there is an indication of a V-shaped pattern upstream, with the right side being rotated and shifted downward.

To further investigate decay patterns, we apply an exponential distance decay function to the downstream distance, which aligns with hydrological models of contaminant dissipation. The detailed results appear in [Appendix C1.1](#). In addition to the strength of the effect, we estimate the rate of decay and allow it to vary from meter to kilometer scales. We find minimal decay over our sample range, and overall impacts that are comparable to the immediate vicinity of mines. The effect on overall vegetation downstream begins at  $-0.0062$  and halves after 281 km, which extends beyond the support of our sample. For croplands, the effect begins at  $-0.0068$  and decays by half at a distance of 72 km, diminishing to 10% of its initial value at 234 km.

These findings highlight that vast areas downstream of mines may be impacted by their water pollution, extending considerably beyond the immediate vicinity examined in our main analysis. Isolating the nature of these effects and identifying specific mechanisms driving them is an important avenue for future research.

### 4.3. Robustness

To ensure the validity of our main findings, we conduct a comprehensive set of robustness checks that address potential concerns about our empirical strategy. We organize these robustness checks into six categories: (i) included covariates, (ii) outcome measures, (iii) sample definition, (iv) unobserved factors, (v) covariate balance, and (vi) placebo tests. [Figure 6](#) visualizes estimates under various checks; the complete results appear in the Appendix.<sup>19</sup>

**Covariates** First, we examine whether our results are sensitive to the choice of control variables. We test four variations: (1) adding controls for air pollution based on PM2.5 concentrations, (2) controlling for conflict intensity, (3) using distance to coast as a geographic control, and (4) using alternative meteorological data for rainfall and temperature. None of these modifications substantially alter our baseline findings, suggesting that our results are not driven by omitted variables related to these factors.

---

<sup>19</sup>Further details are provided in [Appendix C3](#); [Tables E9](#) to [E12](#) report the full estimates.

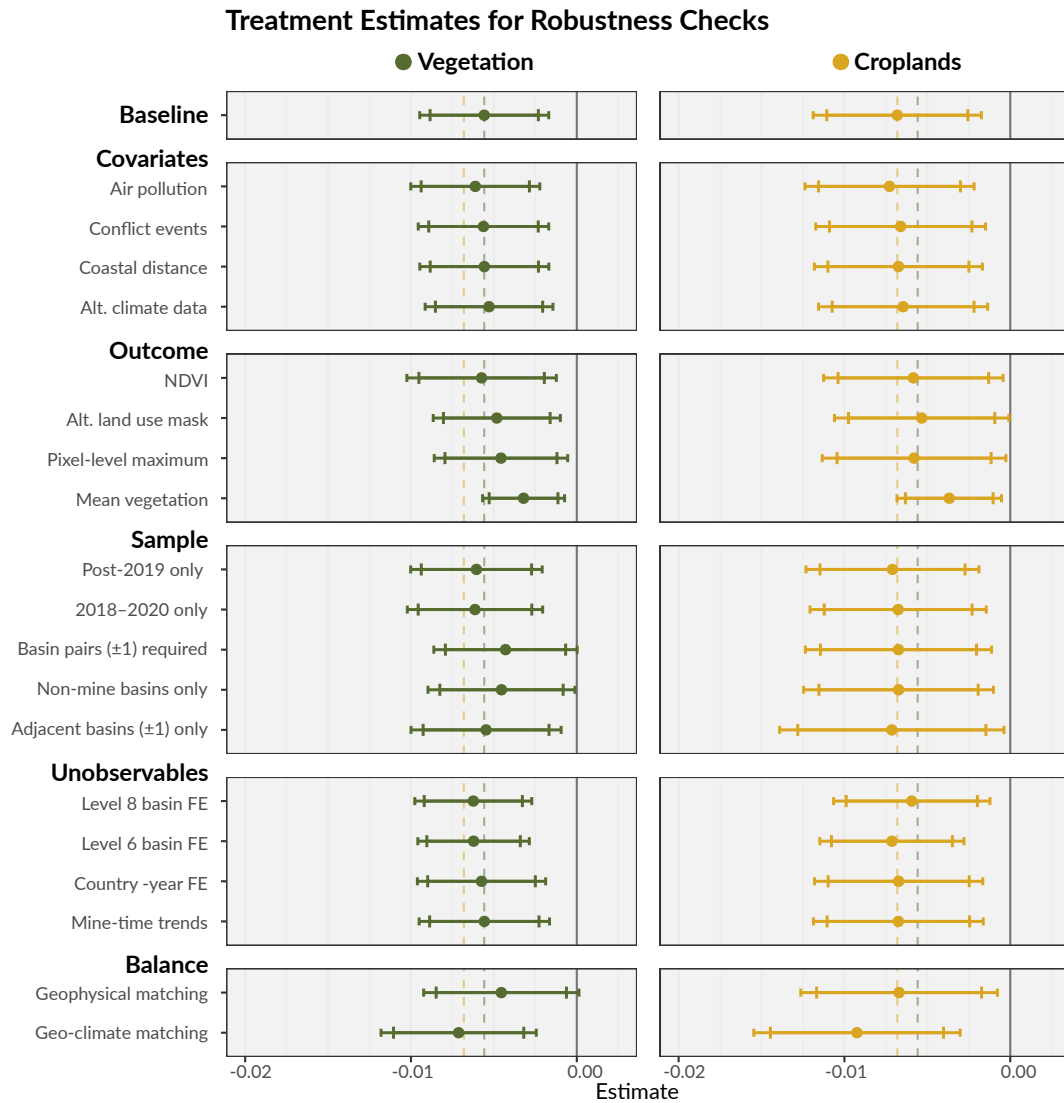


FIGURE 6: Dots indicate the average treatment effect in the first three basins; whiskered lines indicate 90% and 95% confidence intervals. Dashed vertical lines indicate point estimates for baseline specifications.

**Outcome measures** Next, we assess whether our results depend on our specific choice of outcome variables. We (1) substitute the EVI with the Normalized Difference Vegetation Index (NDVI), and (2) use alternative land use masks for vegetation and cropland-specific vegetation index.<sup>20</sup> We also (3) construct the peak EVI to reflect pixel-level peaks (accounting for varying seasonality) before aggregation to the basin level, and (4) use the mean annual EVI to reflect vegetation over the full year, instead of the peak value. These variations change the magnitude of estimates due to different sample moments, but the qualitative results remain consistent, showing significant negative effects of mines on downstream vegetation.

**Sample definition** We also verify the robustness of our results across different sample restrictions. First, we focus on periods around the 2019 satellite imagery used in mine delineation: (1) 2019 and after, and (2) a narrower 2018–2020 window. Second, we modify the sample by (3) only including basin systems with at least one up- and downstream basin, (4) excluding mine-basins entirely, and (5) analyzing only immediate first-order basins of mine-basins that have at least one up- and downstream basin to ensure we recover the intended variation. Finally, we also consider a specification where we average outcomes and covariates over all years (see Column 2 of [Table E11](#)). While the precision of estimates decreases with smaller samples, our key finding remains robust.

**Unobserved factors** We further account for potentially confounding, unobservable factors through alternative fixed effects structures. We introduce hierarchical fixed effects at the (1) Level 6 and (2) Level 8 super-basins, which contain an average of 67 and 6 of our Level 12 basins. We also (3) add country-by-year fixed effects to control for changing national conditions, and (4) include mine-specific linear time trends. In addition, we vary the clustering of standard errors to the mine and year level, the mine-by-year level, and the individual basin level (see Columns 6–8 of [Table E12](#)). These specifications yield consistent results, indicating that our findings are not driven by unobserved spatial or temporal factors.

**Covariate balance** To address potential imbalance between treatment and control groups, we use coarsened exact matching (Iacus, King, and Porro, 2012). We match

---

<sup>20</sup>We use a narrower version for the vegetation mask that excludes sparse and flooded vegetation, and an Africa-specific cropland mask that is anchored in 2019 (Digital Earth Africa, 2022).

upstream and downstream basins on (1) elevation, slope, and soil type in a basic specification, then add (2) temperature, and precipitation in an extended version. The matched samples achieve excellent balance on observables (see [Figure D12](#) and [Figure D13](#) in the Appendix), and the resulting estimates are significant and of similar magnitude. This further corroborates the validity of our identification strategy.

**Placebos** Finally, we conduct two additional validation exercises. First, we randomly reassign the treatment status across basins and re-estimate a total of 5,000 times, finding no evidence of the impacts observed in our actual analysis (see [Appendix C3.2](#) for details, and [Figure D11](#) for results). Second, we use our control variables as well as nighttime light emissions and forest loss as placebo outcomes. We find no structural differences between up- and downstream basins (see [Appendix C3.3](#) and [Figure D10](#) for details), confirming that our identification strategy isolates mining impacts rather than pre-existing environmental or social differences.

## 5. Discussion

Our results provide strong causal evidence for water-mediated effects of mines on vegetation and agriculture. We find reductions of peak vegetation indices of 1.28–1.35% for all vegetation and 1.38–1.47% for croplands specifically. Below, we discuss the mechanisms behind these impacts, how to interpret and gauge their size, which limitations to consider, and what our results imply for policy and future research.

### 5.1. Mechanisms

Our research design identifies mining impacts on vegetation that are specifically mediated along water streams, separating them from other effects of mining. The main mechanisms explaining our findings are water pollution and its second-order effects, such as human responses to it.

Water pollution emerges as the primary mechanism behind our observed effects. As a major and well-documented externality of mining, it provides the most direct explanation for reduced vegetation health downstream of mines. While our design provides only indirect evidence for specific mechanisms, we analyzed available water quality data (from United Nations Environment Programme, 2025) to confirm the role of water pollution (see [Appendix B3](#)).

The data reveal markedly lower water quality in mine-basins and downstream areas compared to upstream basins. Key indicators are consistent with mine-induced pollution: electrical conductivity (measuring dissolved salts and minerals), sodium discharge and adsorption, and sulfate concentrations that can cause acidification are increased downstream. However, these water quality data have important limitations: samples are exclusively from South Africa, were obtained at irregular intervals using different methods, and measurements of heavy metals are lacking. More comprehensive data are needed to assess water quality for the entire continent and differentiate between pollutants, for instance, from mercury- or cyanide-based gold extraction (see Verbrugge, Lanzano, and Libassi, 2021).

Our results could also be influenced by secondary mechanisms in response to water pollution, such as soil pollution, water scarcity, and human impacts. Water systems and soils are closely interlinked (Cheng et al., 2021), and riverine pollution affects both groundwater and soils, impacting the rainfed agriculture that is common in Africa (see [Appendix B1.1](#) for more details). Similarly, mines and water pollution (as well as increased evapotranspiration and percolation) creates water scarcity that may exacerbate the overall downstream impacts. By contrast, the net effects of human responses are ambiguous, as locals may be affected and respond in ways that could amplify or attenuate the direct impacts of mines.

Evidence for air pollution suggests complex human impact patterns. While plant growth is primarily affected directly, pollution also impacts farmer health and consequently labor supply, productivity, income, and land allocation (see Fugiel et al., 2017; Kotsadam and Tolonen, 2016; Miao, Huang, and Song, 2017). To adapt to water pollution, farmers may shift towards more resilient crops (that tolerate, e.g., higher salinity levels) and attenuate overall impacts or respond by farming less intensively or out-migrating, amplifying impacts (note that Aragón and Rud, 2013, 2015, do not find evidence for crop diversification or migration in more focused studies).<sup>21</sup> When water sources are contaminated, agricultural practices could also be modified by reducing irrigation — which is uncommon in the region — or deciding against adopting irrigation systems in the first place.

---

<sup>21</sup>While our design only reflects the net effect, we expect the role of adaptation to be constrained and attenuate overall impacts. A rational response to environmental change would be to mitigate it where possible, but other scenarios are thinkable. If, for instance, impacts drive downstream farmers to abandon their croplands and migrate, the overall effect will be amplified by the decrease in labor supply. If their destination were upstream — which is unlikely, given the other undirected impacts of mines — impacts would be exacerbated further.

## 5.2. Interpretation

Our analysis reveals mines' considerable downstream impacts on vegetation and agriculture. These effects are mediated by water pollution and adaptation responses, making them likely to accumulate and persist for many years. We find that downstream of mines, the peak cropland vegetation index is reduced by 1.38–1.47% of the sample mean.

These effect sizes align with several other studies that use vegetation indices to approximate agricultural productivity. To provide context for our findings, we briefly summarize related results based on similar satellite-derived measures. Using peak annual EVI, Wuepper et al. (2023) find a 2.2% change per step on a ten-step institutional quality index over five years. In India, Asher and Novosad (2020) estimate a 3.5% increase from new road construction, while Ghosh and Vats (2022) find that a 10% increase in guaranteed income beneficiaries increases the same measure by 8.1%. Information and communication technologies show similar effects: Gupta, Ponticelli, and Tesei (2024) find that districts in India with one-standard-deviation higher shares of non-state language speakers generate 0.6% lower yields. Agarwal et al. (2024) find that access to 4G internet increased EVI-implied yields by 12.5% after six years.

Studies examining water and agricultural practices provide more direct comparisons to our findings. Asher, Campion, et al. (2022) estimate that villages with access to irrigation canals show 1.7% and 7.1% higher EVI-proxied yields in the monsoon and winter seasons, respectively. Boudot-Reddy and Butler (2025) find that one-standard-deviation increases in groundwater extraction increase yields by 8.6% in winter. Deines, S. Wang, and David B Lobell (2019) find that long-term conservation tillage increases satellite-derived yields by 0.7 to 3.3% by preventing soil degradation. By comparison, Jain et al. (2019) find a 4.5% yield gain from more effective fertilizer application.

We translated our results into agricultural terms by correlating our measure with LSMS-ISA and AReNA-DHS data. For the immediately affected 74,000 km<sup>2</sup> of small-holder croplands that are up to around 33 km downstream of mine-basins, we estimated annual losses of 91,000–205,000 tons of cereals. These losses correspond to 5.4 to 12.1% of the food aid distributed annually by the World Food Programme.

When interpreting these results, two important caveats apply. First, they reflect only one specific pathway among the many externalities of mines. Second, they might be partly offset by redistributing the benefits from mining. Mining also affects agriculture,

vegetation, and livelihoods via air pollution and other non-directed factors that our research design does not capture. This includes positive effects, such as increased local incomes and wealth ownership (Aragón and Rud, 2015; Goltz and Barnwal, 2019) and alternative employment opportunities (Gräser, 2024; Kotsadam and Tolonen, 2016). Meanwhile, water pollution from mines creates numerous additional negative effects beyond agriculture, including water scarcity, biodiversity loss, and detrimental impacts on human health and cognition (Gittard and Hu, 2024; Van Vliet, Flörke, and Wada, 2017; Vörösmarty et al., 2010).

### 5.3. Limitations

Our analysis faces three key limitations that warrant discussion. These are related to uncertainty in the outcome measures, incomplete mining data, and limited mechanistic evidence.

**Outcome measures** We rely primarily on remotely sensed vegetation indices, which introduces measurement noise that affects our estimates. While these indices offer consistent and highly granular measures of vegetation health that enable our analysis, they have practical and inherent constraints. Our peak vegetation indices perform well as a general proxy, but cannot provide a complete picture of complex vegetation patterns. No single variable can capture the heterogeneity of impacts across different crop types, soil conditions, and climatic zones (Bolton and Friedl, 2013), and data for calibration is limited. When we relate our outcome measure to local agricultural survey data, this introduces additional noise and likely produces attenuated estimates. Future research with a narrower geographical focus could address this limitation by combining remote sensing with more detailed ground-truth agricultural data.

**Incomplete mining data** While we use a comprehensive dataset of mine sites with verified locations and surroundings, these data have important gaps. Critical information about mining sites and operations — including their status, activity, production methods, and commodities mined — is not directly available. We partially address this limitation by assessing mining area growth over time and adding commodity information through supplementary sources. However, the data does not distinguish between artisanal and industrial mining, which differ substantially in management practices,

processing methods, and environmental impacts. While geological bedrock characteristics can be used to proxy for artisanal gold mining presence (Girard, Molina-Millán, and Vic, 2025; Girard, Vic, and Molina-Millán, 2022), the frequent co-location of artisanal and industrial mines means that our research design cannot cleanly disentangle effects. Conducting more targeted analyses that integrate detailed information about mining operations represents an important avenue for future research.

**Mechanistic evidence** Our study provides suggestive rather than definitive evidence for the exact mechanisms behind the water-mediated effects. Water quality and pollution data are scarce, geographically patchy, and methodologically inconsistent across regions (Jones et al., 2024). We investigate heterogeneity across commodity types, regions, and mine growth patterns to infer potential mechanisms but cannot provide direct evidence linking specific mining activities to vegetation impacts. Additionally, while our research design identifies local treatment effects, it cannot establish the full geographic reach of these impacts. More in-depth analyses of mechanisms and the reach of effects are left to future work.

## 5.4. Outlook

Our results have important implications for future research and policy. Two key priorities emerge: overcoming the data scarcity that constrains both scientific understanding and policy responses, and implementing targeted interventions to mitigate mining's external costs and distribute its benefits.

A common theme across our analysis is the scarcity and relatively poor quality of available data. Addressing this gap is paramount, as improved information can advance research and enable policymakers to design effective mitigation strategies. Agricultural statistics remain challenging to collect for smallholder and subsistence farming, limiting our understanding of farm-level impacts. Information on mining sites is inadequate and poorly accessible (Maus and Werner, 2024), especially for artisanal mines where data dissemination must safeguard against potential human rights abuses. Water pollution data are similarly essential but scarce (Jones et al., 2024).

Remote sensing technologies offer considerable potential to address these data gaps. For water pollution, satellite imagery can detect large spills and tailings dam failures (Rudorff et al., 2018; Ruppen, Runnalls, et al., 2023), while refined approaches

may help identify heavy metal contamination (Swain and B. Sahoo, 2017). Automated monitoring of mine sites using machine learning methods (Sepin, Vashold, and Kuschnig, 2025) could reduce laborious mapping tasks. However, these remote sensing approaches depend on high-quality ground-based data for calibration. Where institutional capacities for producing such data are lacking, community-based initiatives, for instance, for water monitoring, could serve as an inexpensive and effective complement (Ruppen, Chituri, et al., 2021).

Policymakers should address mining's external costs through targeted interventions. Mining concessions should explicitly recognize externalities and factor in impacts on local economies, vegetation, and food security. Our results provide concrete evidence for including water-mediated impacts in concession allocation decisions. Formalizing artisanal mining through official titling could increase miners' incentives to invest in precautionary equipment, reducing impacts at the source with benefits for human health and the environment.

Supranational interventions are essential for effectively addressing mining externalities by shifting regulatory burdens from mineral-producing countries towards all who benefit from the extracted resources. The Global Industry Standard on Tailings Management (GISTM) presents an important first step for international governance of industrial mines. Our findings highlight that persistent water pollution, and not just catastrophic failures of tailings dams, requires attention. The Minamata Convention targets the use of mercury, which is particularly damaging and still used in processing gold — the commodity for which we find the strongest downstream impact. Supply chain measures could also help reduce minerals sourced through harmful practices, following models like the European Union's Deforestation Regulation (EUDR) that applies to companies selling products on the EU market.

## **6. Conclusion**

In this paper, we identified the causal effects of mining on agricultural productivity mediated by water. Using a quasi-experimental research design, we exploited the discontinuity created by mine sites along directed river networks for identification. We compared differences in agricultural productivity — measured through remotely sensed peak vegetation indices — upstream and downstream of mine locations. Our main specification revealed reductions in peak vegetation of 1.28–1.35% for general

vegetation and 1.38–1.47% for croplands up to approximately 33 km downstream of mines. This corresponds to annual losses of 91,000 to 205,000 tons of cereals across 74,000 km<sup>2</sup> of immediately affected croplands.

Our estimates capture only one component of the total external costs of mining, and for agriculture specifically. Our research design does not reflect impacts that are not mediated along river networks, such as effects on air pollution or income. Additionally, our indirect measurement of agricultural productivity may attenuate estimates of the water-mediated effects. Despite these limitations, our findings contribute important evidence to discussions about resource extraction in Africa, particularly in regions with weak environmental governance.

The documented effects highlight the urgent need for interventions that reduce mines' negative impacts on water systems and distribute their costs and benefits. Industrial and informal mining operations should be required to implement proper containment and reduce or substitute particularly harmful methods. Enhanced monitoring of mines and water quality is necessary to address the data limitations we encountered and provide a basis to understand impacts and guide effective policies.

This study opens several promising avenues for future research. While a comprehensive Africa-wide mine dataset enables broad analysis, it lacks detailed information on individual mine characteristics. Future studies that incorporate data on containment facilities, mine types, and specific pollutants would allow for more precise impact analyses. Similarly, spatial data on crop distributions could enable detailed investigations of which crop types are most susceptible to mining-induced pollution and inform adaptation strategies. Alternative research approaches might address questions our design could not fully answer, such as disentangling the impacts of industrial versus artisanal mines or providing stronger evidence on effect decay. Such research would further strengthen the evidence base for targeted interventions that help balance the economic benefits of mining with its environmental and human costs.

## References

Abatzoglou, John T. et al. (Jan. 2018). "TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015". In: *Scientific Data* 5.170191, pp. 1–12. ISSN: 2052-4463. DOI: [10.1038/sdata.2017.191](https://doi.org/10.1038/sdata.2017.191) (cited page 13).

- Agarwal, Sumit et al. (2024). “Bridging the information gap: Sowing the seeds of productivity with high-speed 4G internet”. In: *Working Paper* 2024.16. DOI: [10.2139/ssrn.4805486](https://doi.org/10.2139/ssrn.4805486) (cited page 29).
- Amatulli, G. et al. (2018). “A suite of global, cross-scale topographic variables for environmental and biodiversity modeling”. In: *Scientific Data* 5.180040, pp. 1–15. ISSN: 2052-4463. DOI: [10.1038/sdata.2018.40](https://doi.org/10.1038/sdata.2018.40) (cited page 12).
- Aragón, Fernando M. and Juan Pablo Rud (2013). “Natural resources and local communities: Evidence from a Peruvian gold mine”. In: *American Economic Journal: Economic Policy* 5.2, pp. 1–25. ISSN: 1945-7731. DOI: [10.1257/pol.5.2.1](https://doi.org/10.1257/pol.5.2.1) (cited page 28).
- (Sept. 2015). “Polluting industries and agricultural productivity: Evidence from mining in Ghana”. In: *The Economic Journal* 126.597, pp. 1980–2011. ISSN: 0013-0133. DOI: [10.1111/econj.12244](https://doi.org/10.1111/econj.12244) (cited pages 3, 5, 28, 30).
- Asher, Sam, Alison Campion, et al. (2022). *The long-run development impacts of agricultural productivity gains: Evidence from irrigation canals in India*. [Online; accessed 20. Aug. 2025] (cited page 29).
- Asher, Sam and Paul Novosad (Mar. 2020). “Rural roads and local economic development”. In: *American Economic Review* 110.3, pp. 797–823. ISSN: 0002-8282. DOI: [10.1257/aer.20180268](https://doi.org/10.1257/aer.20180268) (cited pages 6, 29).
- Aska, Bora et al. (Jan. 2024). “Biodiversity conservation threatened by global mining wastes”. In: *Nature Sustainability* 7, pp. 23–30. ISSN: 2398-9629. DOI: [10.1038/s41893-023-01251-0](https://doi.org/10.1038/s41893-023-01251-0) (cited pages 2–4).
- Audry, Stéphane et al. (Dec. 2004). “Fifty-year sedimentary record of heavy metal pollution (Cd, Zn, Cu, Pb) in the Lot river reservoirs”. In: *Environmental Pollution* 132.3, pp. 413–426. ISSN: 0269-7491. DOI: [10.1016/j.envpol.2004.05.025](https://doi.org/10.1016/j.envpol.2004.05.025) (cited page 6).
- Auld, Graeme, Michele Betsill, and Stacy D. VanDeveer (Oct. 2018). “Transnational governance for mining and the mineral lifecycle”. In: *Annual Review of Environment and Resources* 43, pp. 425–453. DOI: [10.1146/annurev-environ-102017-030223](https://doi.org/10.1146/annurev-environ-102017-030223) (cited page 4).
- Awotwi, Alfred et al. (2021). “Impact of post-reclamation of soil by large-scale, small-scale and illegal mining on water balance components and sediment yield: Pra River Basin case study”. In: *Soil and Tillage Research* 211, p. 105026. ISSN: 0167-1987. DOI: [10.1016/j.still.2021.105026](https://doi.org/10.1016/j.still.2021.105026) (cited pages 3, 4).
- Ayers, R. S. and D. W. Westcott (1985). *Water quality for agriculture*. Irrigation and Drainage Paper 29. Rome, Italy: Food and Agriculture Organization of the United Nations (cited page xiii).
- Azzari, George, Meha Jain, and David B. Lobell (Dec. 2017). “Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries”. In: *Remote Sensing of Environment* 202, pp. 129–141. ISSN: 0034-4257. DOI: [10.1016/j.rse.2017.04.014](https://doi.org/10.1016/j.rse.2017.04.014) (cited page 11).

- Banzhaf, Spencer, Lala Ma, and Christopher Timmins (Feb. 2019). “Environmental justice: The economics of race, place, and pollution”. In: *Journal of Economic Perspectives* 33.1, pp. 185–208. ISSN: 0895-3309. DOI: [10.1257/jep.33.1.185](https://doi.org/10.1257/jep.33.1.185) (cited page 3).
- Barenblitt, Abigail et al. (2021). “The large footprint of small-scale artisanal gold mining in Ghana”. In: *Science of The Total Environment* 781, p. 146644. ISSN: 0048-9697. DOI: [10.1016/j.scitotenv.2021.146644](https://doi.org/10.1016/j.scitotenv.2021.146644) (cited page 4).
- Bazillier, Remi and Victoire Girard (Mar. 2020). “The gold digger and the machine. Evidence on the distributive effect of the artisanal and industrial gold rushes in Burkina Faso”. In: *Journal of Development Economics* 143, p. 102411. ISSN: 0304-3878. DOI: [10.1016/j.jdeveco.2019.102411](https://doi.org/10.1016/j.jdeveco.2019.102411) (cited pages 4, 5).
- Bentze, Thomas Patrick and Philip Randolph Wollburg (2024). *A longitudinal cross-country dataset on agricultural productivity and welfare in sub-Saharan Africa*. Policy Research Working Paper 10976. Washington, D.C.: World Bank Group. DOI: [10.1596/1813-9450-10976](https://doi.org/10.1596/1813-9450-10976) (cited page ix).
- Berman, Nicolas, Mathieu Couttenier, and Victoire Girard (2023). “Mineral resources and the salience of ethnic identities”. In: *Economic Journal* 133.653, pp. 1705–1737. ISSN: 0013-0133. DOI: [10.1093/ej/uead018](https://doi.org/10.1093/ej/uead018) (cited pages 2, 4).
- Berman, Nicolas, Mathieu Couttenier, Dominic Rohner, et al. (2017). “This mine is mine! How minerals fuel conflicts in Africa”. In: *American Economic Review* 107.6, pp. 1564–1610. ISSN: 0002-8282. DOI: [10.1257/aer.20150774](https://doi.org/10.1257/aer.20150774) (cited pages 2–4).
- Bolton, Douglas K. and Mark A. Friedl (May 2013). “Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics”. In: *Agricultural and Forest Meteorology* 173, pp. 74–84. ISSN: 0168-1923. DOI: [10.1016/j.agrformet.2013.01.007](https://doi.org/10.1016/j.agrformet.2013.01.007) (cited page 30).
- Boudot-Reddy, Camille and Andr Butler (Aug. 2025). “Watering the seeds of the rural economy: Evidence from groundwater irrigation in India”. In: *World Bank Economic Review* 39.3, pp. 571–591. ISSN: 0258-6770. DOI: [10.1093/wber/lhae041](https://doi.org/10.1093/wber/lhae041) (cited pages 6, 29).
- Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik (Nov. 2019). *A practical introduction to regression discontinuity designs: Foundations*. Cambridge University Press. ISBN: 9781108710206. DOI: [10.1017/9781108684606](https://doi.org/10.1017/9781108684606) (cited pages xvii, xxxvii).
- Cheng, Kun et al. (2021). “The role of soils in regulation of freshwater and coastal water quality”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 376.1834, p. 20200176. ISSN: 0962-8436. DOI: [10.1098/rstb.2020.0176](https://doi.org/10.1098/rstb.2020.0176) (cited pages 28, vii).
- Currie, Janet et al. (2009). “Does pollution increase school absences?” In: *Review of Economics and Statistics* 91.4, pp. 682–694. ISSN: 0034-6535. DOI: [10.1162/rest.91.4.682](https://doi.org/10.1162/rest.91.4.682) (cited page 5).
- Defourny, Pierre et al. (2024). *Product user Guide and specification CDR and ICDR Sentinel-3 Land Cover (v2.1.1)*. Tech. rep. UCLouvain / European Space Agency. DOI: [10.24381/cds.006f2c9a](https://doi.org/10.24381/cds.006f2c9a) (cited pages 11, xxxiii).

- Deines, Jillian M, Sherrie Wang, and David B Lobell (Dec. 2019). “Satellites reveal a small positive yield effect from conservation tillage across the US Corn Belt”. In: *Environmental Research Letters* 14.12, p. 124038. ISSN: 1748-9326. DOI: [10.1088/1748-9326/ab503b](https://doi.org/10.1088/1748-9326/ab503b) (cited page 29).
- Dias, Mateus, Rudi Rocha, and Rodrigo R. Soares (Feb. 2023). “Down the river: Glyphosate use in agriculture and birth outcomes of surrounding populations”. In: *Review of Economic Studies* 90.6, pp. 2943–2981. ISSN: 1467-937X. DOI: [10.1093/restud/rdad011](https://doi.org/10.1093/restud/rdad011) (cited page 2).
- Didan, Kamel (2015). *MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006*. DOI: [10.5067/MODIS/MOD13Q1.006](https://doi.org/10.5067/MODIS/MOD13Q1.006) (cited page 11).
- Digital Earth Africa (2022). *Cropland extent maps for Africa*. Tech. rep. Dataset. DE Africa Services (cited pages 11, 26, xxxiii).
- Dinerstein, E. et al. (2017). “An ecoregion-based approach to protecting half the terrestrial realm”. In: *Bioscience* 67.6, pp. 534–545. ISSN: 0006-3568. DOI: [10.1093/biosci/bix014](https://doi.org/10.1093/biosci/bix014). eprint: [28608869](https://doi.org/10.1093/biosci/bix014) (cited pages 13, 19).
- Dinku, Tufa et al. (Nov. 2018). “Validation of the CHIRPS satellite rainfall estimates over eastern Africa”. In: *Quarterly Journal of the Royal Meteorological Society* 144.S1, pp. 292–312. ISSN: 0035-9009. DOI: [10.1002/qj.3244](https://doi.org/10.1002/qj.3244) (cited page 13).
- Dryden, Rachel et al. (Nov. 2021). “Do we prioritize floodplains for development and farming? Mapping global dependence and exposure to inundation”. In: *Global Environmental Change* 71, p. 102370. ISSN: 0959-3780. DOI: [10.1016/j.gloenvcha.2021.102370](https://doi.org/10.1016/j.gloenvcha.2021.102370) (cited page vii).
- Du, Longxu et al. (2024). “Heterogeneous impact of soil acidification on crop yield reduction and its regulatory variables: A global meta-analysis”. In: *Field Crops Research* 319, p. 109643. ISSN: 0378-4290. DOI: [10.1016/j.fcr.2024.109643](https://doi.org/10.1016/j.fcr.2024.109643) (cited pages 3, 5, viii).
- Duncan, Albert Ebo (2020). “The dangerous couple: Illegal mining and water pollution—a case study in Fena river in the Ashanti region of Ghana”. In: *Journal of Chemistry* 2020, pp. 1–9. ISSN: 2090-9071. DOI: [10.1155/2020/2378560](https://doi.org/10.1155/2020/2378560) (cited pages 3, 5, vi).
- Duncan, Albert Ebo, Nanne de Vries, and Kwabena Biritwum Nyarko (2018). “Assessment of heavy metal pollution in the sediments of the river Pra and its tributaries”. In: *Water, Air, & Soil Pollution* 229.8. ISSN: 1573-2932. DOI: [10.1007/s11270-018-3899-6](https://doi.org/10.1007/s11270-018-3899-6) (cited page vi).
- Elvidge, Christopher, Kimberly Baugh, et al. (June 2017). “VIIRS night-time lights”. In: *International Journal of Remote Sensing* 38.21, pp. 5860–5879. ISSN: 1366-5901. DOI: [10.1080/01431161.2017.1342050](https://doi.org/10.1080/01431161.2017.1342050) (cited pages 13, xxi, xxv).
- Elvidge, Christopher, Mikhail Zhizhin, et al. (Mar. 2021). “Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019”. In: *Remote Sensing* 13.5, p. 922. ISSN: 2072-4292. DOI: [10.3390/rs13050922](https://doi.org/10.3390/rs13050922) (cited pages xxi, xxv).

- Fischer, G. et al. (2021). *Global Agro-Ecological Zones (GAEZ v4) model documentation*. Tech. rep. Food, Agriculture Organization (FAO) of the United Nations, and International Institute for Applied System (IIASA), p. 303 (cited page 14).
- Frossard, Aline et al. (May 2018). “Long- and short-term effects of mercury pollution on the soil microbiome”. In: *Soil Biology and Biochemistry* 120, pp. 191–199. ISSN: 0038-0717. DOI: [10.1016/j.soilbio.2018.01.028](https://doi.org/10.1016/j.soilbio.2018.01.028) (cited pages 5, vii).
- Fugiel, Agata et al. (Feb. 2017). “Environmental impact and damage categories caused by air pollution emissions from mining and quarrying sectors of European countries”. In: *Journal of Cleaner Production* 143, pp. 159–168. ISSN: 0959-6526. DOI: [10.1016/j.jclepro.2016.12.136](https://doi.org/10.1016/j.jclepro.2016.12.136) (cited pages 5, 28).
- Funk, Chris et al. (Dec. 2015). “The climate hazards infrared precipitation with stations— A new environmental record for monitoring extremes”. In: *Scientific Data* 2.150066, pp. 1–21. ISSN: 2052-4463. DOI: [10.1038/sdata.2015.66](https://doi.org/10.1038/sdata.2015.66) (cited page 13).
- Gelman, A. and G. Imbens (2019). “Why high-order polynomials should not be used in regression discontinuity designs”. In: *Journal of Business & Economic Statistics*. DOI: [10.1080/07350015.2017.1366909](https://doi.org/10.1080/07350015.2017.1366909) (cited pages 14, xvii).
- Ghosh, Pulak and Nishant Vats (Nov. 2022). *Safety nets, credit, and investment: Evidence from a guaranteed income program*. [Online; accessed 20. Aug. 2025]. DOI: [10.2139/ssrn.4265112](https://doi.org/10.2139/ssrn.4265112) (cited page 29).
- Giljum, Stefan et al. (2022). “A pantropical assessment of deforestation caused by industrial mining”. In: *Proceedings of the National Academy of Sciences* 119.38, e2118273119. DOI: [10.1073/pnas.2118273119](https://doi.org/10.1073/pnas.2118273119) (cited pages 2–4).
- Girard, Victoire, Teresa Molina-Millán, and Guillaume Vic (July 2025). “Artisanal mining in Africa. Green for Gold?” In: *Economic Journal*, ueaf058. ISSN: 0013-0133. DOI: [10.1093/ej/ueaf058](https://doi.org/10.1093/ej/ueaf058) (cited pages 2, 4, 31).
- Girard, Victoire, Guillaume Vic, and Teresa Molina-Millán (Mar. 2022). *Gold-suitable geological layers for the African continent*. Version V4. DOI: [10.60580/novasbe.28957058](https://doi.org/10.60580/novasbe.28957058) (cited page 31).
- Gittard, Mlanie and Irène Hu (2024). *MiningLeaks: Water pollution and child mortality in Africa*. [Online; accessed 13. Feb. 2025] (cited pages 2, 3, 30).
- Goltz, Jan von der and Prabhat Barnwal (June 2019). “Mines: The local wealth and health effects of mineral mining in developing countries”. In: *Journal of Development Economics* 139, pp. 1–16. ISSN: 0304-3878. DOI: [10.1016/j.jdeveco.2018.05.005](https://doi.org/10.1016/j.jdeveco.2018.05.005) (cited pages 3, 4, 30).
- Graff Zivin, Joshua and Matthew Neidell (2012). “The impact of pollution on worker productivity”. In: *American Economic Review* 102.7, pp. 3652–73. ISSN: 0002-8282. DOI: [10.1257/aer.102.7.3652](https://doi.org/10.1257/aer.102.7.3652) (cited page 5).

- Gräser, Melanie (2024). “Industrial versus artisanal mining: The effects on local employment in Liberia”. In: *Journal of Rural Studies* 111, p. 103389. ISSN: 0743-0167. DOI: [10.1016/j.jrurstud.2024.103389](https://doi.org/10.1016/j.jrurstud.2024.103389) (cited pages 4, 5, 30).
- Gupta, Apoorv, Jacopo Ponticelli, and Andrea Tesei (Sept. 2024). “Language barriers, technology adoption and productivity: evidence from agriculture in India”. In: *Review of Economics and Statistics*, pp. 1–28. DOI: [10.1162/rest\\_a\\_01501](https://doi.org/10.1162/rest_a_01501) (cited page 29).
- Haghighizadeh, Atoosa et al. (June 2024). “Comprehensive analysis of heavy metal soil contamination in mining environments: Impacts, monitoring techniques, and remediation strategies”. In: *Arabian Journal of Chemistry* 17.6, p. 105777. ISSN: 1878-5352. DOI: [10.1016/j.arabjc.2024.105777](https://doi.org/10.1016/j.arabjc.2024.105777) (cited page vii).
- Hanna, Rema and Paulina Oliva (2015). “The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City”. In: *Journal of Public Economics* 122, pp. 68–79. ISSN: 0047-2727. DOI: [10.1016/j.jpubeco.2014.10.004](https://doi.org/10.1016/j.jpubeco.2014.10.004) (cited page 5).
- Hansen, M. C. et al. (Nov. 2013). “High-resolution global maps of 21st-century forest cover change”. In: *Science* 342.6160, pp. 850–853. ISSN: 1095-9203. DOI: [10.1126/science.1244693](https://doi.org/10.1126/science.1244693) (cited pages 13, xxi, xxv).
- Harris, Ian et al. (Apr. 2020). “Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset”. In: *Scientific Data* 7.109, pp. 1–18. ISSN: 2052-4463. DOI: [10.1038/s41597-020-0453-3](https://doi.org/10.1038/s41597-020-0453-3) (cited pages 13, xxxiii).
- Hengl, T. et al. (2017). “SoilGrids250m: Global gridded soil information based on machine learning”. In: *PLOS ONE* 12.2, e0169748. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0169748](https://doi.org/10.1371/journal.pone.0169748) (cited page 12).
- Huete, A. et al. (Nov. 2002). “Overview of the radiometric and biophysical performance of the MODIS vegetation indices”. In: *Remote Sensing of Environment* 83.1, pp. 195–213. ISSN: 0034-4257. DOI: [10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2) (cited page 11).
- Iacus, S. M., G. King, and G. Porro (2012). “Causal inference without balance checking: Coarsened Exact Matching”. In: *Political Analysis* 20.1, pp. 1–24. ISSN: 1047-1987. DOI: [10.1093/pan/mpr013](https://doi.org/10.1093/pan/mpr013) (cited pages 26, xix).
- IFPRI (2020). *AReNA's DHS-GIS database*. Version V1. DOI: [10.7910/DVN/OQIPRW](https://doi.org/10.7910/DVN/OQIPRW) (cited page ix).
- Imbens, G. and K. Kalyanaraman (2012). “Optimal bandwidth choice for the regression discontinuity estimator”. In: *Review of Economic Studies* 79.3, pp. 933–959. ISSN: 0034-6527. DOI: [10.1093/restud/rdr043](https://doi.org/10.1093/restud/rdr043) (cited page xvii).
- Jain, Meha et al. (Oct. 2019). “The impact of agricultural interventions can be doubled by using satellite data”. In: *Nature Sustainability* 2.10, pp. 931–934. ISSN: 2398-9629. DOI: [10.1038/s41893-019-0396-x](https://doi.org/10.1038/s41893-019-0396-x) (cited page 29).
- Jasansky, Simon et al. (2023). “An open database on global coal and metal mine production”. In: *Scientific Data* 10.52, pp. 1–12. ISSN: 2052-4463. DOI: [10.1038/s41597-023-01965-y](https://doi.org/10.1038/s41597-023-01965-y) (cited pages 12, xviii).

- Johnson, D. Barrie and Kevin B. Hallberg (2005). “Acid mine drainage remediation options: A review”. In: *Science of The Total Environment* 338.1, pp. 3–14. ISSN: 0048-9697. DOI: [10.1016/j.scitotenv.2004.09.002](https://doi.org/10.1016/j.scitotenv.2004.09.002) (cited pages 5, 6).
- Johnson, David M. (Oct. 2016). “A comprehensive assessment of the correlations between field crop yields and commonly used MODIS products”. In: *International Journal of Applied Earth Observation and Geoinformation* 52, pp. 65–81. ISSN: 1569-8432. DOI: [10.1016/j.jag.2016.05.010](https://doi.org/10.1016/j.jag.2016.05.010) (cited pages 6, 11).
- Jones, Edward R et al. (2024). “Blind spots in global water quality monitoring”. In: *Environmental Research Letters* 19.9, p. 091001. ISSN: 1748-9326. DOI: [10.1088/1748-9326/ad6919](https://doi.org/10.1088/1748-9326/ad6919) (cited pages 4, 31, xii).
- Keeler, B. L. et al. (2012). “Linking water quality and well-being for improved assessment and valuation of ecosystem services”. In: *Proceedings of the National Academy of Sciences* 109.45, pp. 18619–18624. DOI: [10.1073/pnas.1215991109](https://doi.org/10.1073/pnas.1215991109) (cited page 3).
- Knutsen, Carl Henrik et al. (Sept. 2016). “Mining and local corruption in Africa”. In: *American Journal of Political Science* 61.2, pp. 320–334. ISSN: 1540-5907. DOI: [10.1111/ajps.12268](https://doi.org/10.1111/ajps.12268) (cited pages 2, 4).
- Kotsadam, Andreas and Anja Tolonen (July 2016). “African mining, gender, and local employment”. In: *World Development* 83, pp. 325–339. ISSN: 0305-750X. DOI: [10.1016/j.worlddev.2016.01.007](https://doi.org/10.1016/j.worlddev.2016.01.007) (cited pages 5, 28, 30).
- Lehner, Bernhard and Günther Grill (Apr. 2013). “Global river hydrography and network routing: Baseline data and new approaches to study the world’s large river systems”. In: *Hydrological Processes* 27.15, pp. 2171–2186. ISSN: 1099-1085. DOI: [10.1002/hyp.9740](https://doi.org/10.1002/hyp.9740) (cited pages 6, 9, i).
- Lipscomb, Molly and Ahmed Mushfiq Mobarak (Jan. 2017). “Decentralization and pollution spillovers: Evidence from the re-drawing of county borders in Brazil”. In: *Review of Economic Studies* 84.1, pp. 464–502. ISSN: 0034-6527. DOI: [10.1093/restud/rdw023](https://doi.org/10.1093/restud/rdw023) (cited page 2).
- Macklin, M. G. et al. (2023). “Impacts of metal mining on river systems: A global assessment”. In: *Science* 381.6664, pp. 1345–1350. ISSN: 0036-8075. DOI: [10.1126/science.adg6704](https://doi.org/10.1126/science.adg6704) (cited pages 2–5, vii, xv, xvi, xviii).
- Malone, Aaron et al. (2023). “Transitional dynamics from mercury to cyanide-based processing in artisanal and small-scale gold mining: Social, economic, geochemical, and environmental considerations”. In: *Science of The Total Environment* 898, p. 165492. ISSN: 0048-9697. DOI: [10.1016/j.scitotenv.2023.165492](https://doi.org/10.1016/j.scitotenv.2023.165492) (cited page 5).
- Maus, Victor, Stefan Giljum, et al. (2022). “An update on global mining land use”. In: *Scientific data* 9.1, pp. 1–11 (cited pages 12, iii–v).
- Maus, Victor and Tim Werner (2024). “Impacts for half of the world’s mining areas are undocumented”. In: *Nature* 625, pp. 26–29. DOI: [10.1038/d41586-023-04090-3](https://doi.org/10.1038/d41586-023-04090-3) (cited pages 3, 4, 31).

- Miao, Weijie, Xin Huang, and Yu Song (June 2017). “An economic assessment of the health effects and crop yield losses caused by air pollution in mainland China”. In: *Journal of Environmental Sciences* 56, pp. 102–113. ISSN: 1001-0742. DOI: [10.1016/j.jes.2016.08.024](https://doi.org/10.1016/j.jes.2016.08.024) (cited pages 5, 28).
- Mulenga, Mary et al. (Mar. 2024). “Aquatic mercury pollution from artisanal and small-scale gold mining in sub-Saharan Africa: status, impacts, and interventions”. In: *Water* 16.5, p. 756. ISSN: 2073-4441. DOI: [10.3390/w16050756](https://doi.org/10.3390/w16050756) (cited page 3).
- Neath, Andrew A. and Joseph E. Cavanaugh (Mar. 2012). “The Bayesian information criterion: Background, derivation, and applications”. In: *WIREs Computational Statistics* 4.2, pp. 199–203. ISSN: 1939-5108. DOI: [10.1002/wics.199](https://doi.org/10.1002/wics.199) (cited page xvi).
- Nordstrom, Darrell Kirk et al. (Jan. 2000). “Negative pH and extremely acidic mine waters from Iron Mountain, California”. In: *Environmental Science & Technology* 34.2, pp. 254–258. ISSN: 0013-936X. DOI: [10.1021/es990646v](https://doi.org/10.1021/es990646v) (cited page viii).
- Ofosu, George et al. (2020). “Socio-economic and environmental implications of artisanal and small-scale mining (ASM) on agriculture and livelihoods”. In: *Environmental Science & Policy* 106, pp. 210–220. ISSN: 1462-9011. DOI: [10.1016/j.envsci.2020.02.005](https://doi.org/10.1016/j.envsci.2020.02.005) (cited pages 3, 4).
- Olmstead, S. M. (2010). “The economics of water quality”. In: *Review of Environmental Economics and Policy*. DOI: [10.1093/reep/rep016](https://doi.org/10.1093/reep/rep016) (cited page 3).
- Padilla, A.J. et al. (2021). *Compilation of geospatial data (GIS) for the mineral industries and related infrastructure of Africa*. data release. U.S. Geological Survey. DOI: [10.5066/P97EQWXP](https://doi.org/10.5066/P97EQWXP) (cited pages 12, xviii).
- Page, Kathryn L. et al. (May 2021). “Review of crop-specific tolerance limits to acidity, salinity, and sodicity for seventeen cereal, pulse, and oilseed crops common to rainfed subtropical cropping systems”. In: *Land Degradation & Development* 32.8, pp. 2459–2480. ISSN: 1085-3278. DOI: [10.1002/ldr.3915](https://doi.org/10.1002/ldr.3915) (cited page viii).
- Palumbo-Roe, Barbara, Joanna Wragg, and Vanessa J. Banks (July 2012). “Lead mobilisation in the hyporheic zone and river bank sediments of a contaminated stream: Contribution to diffuse pollution”. In: *Journal of Soils and Sediments* 12.10, pp. 1633–1640. ISSN: 1614-7480. DOI: [10.1007/s11368-012-0552-7](https://doi.org/10.1007/s11368-012-0552-7) (cited page vii).
- Pandey, Bhanu, Madhoolika Agrawal, and Siddharth Singh (Jan. 2014). “Assessment of air pollution around coal mining area: Emphasizing on spatial distributions, seasonal variations and heavy metals, using cluster and principal component analysis”. In: *Atmospheric Pollution Research* 5.1, pp. 79–86. ISSN: 1309-1042. DOI: [10.5094/APR.2014.010](https://doi.org/10.5094/APR.2014.010) (cited page 5).
- Parida, Asish Kumar and Anath Bandhu Das (2005). “Salt tolerance and salinity effects on plants: A review”. In: *Ecotoxicology and Environmental Safety* 60.3, pp. 324–349. ISSN: 0147-6513. DOI: [10.1016/j.ecoenv.2004.06.010](https://doi.org/10.1016/j.ecoenv.2004.06.010) (cited page viii).

- Pörtner, Hans-Otto et al. (Aug. 2022). *Climate Change 2022: Impacts, Adaptation and Vulnerability*. [Online; accessed 15. Aug. 2022]. Cambridge: Cambridge University Press (cited page 2).
- Raleigh, Clionadh et al. (Sept. 2010). “Introducing ACLED: An Armed Conflict Location and Event Dataset”. In: *Journal of Peace Research* 47.5, pp. 651–660. ISSN: 0022-3433. DOI: [10.1177/0022343310378914](https://doi.org/10.1177/0022343310378914) (cited pages 13, xxxiii).
- Rudorff, Natalia et al. (Nov. 2018). “Remote sensing monitoring of the impact of a major mining wastewater disaster on the turbidity of the Doce River plume off the eastern Brazilian coast”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 145, pp. 349–361. ISSN: 0924-2716. DOI: [10.1016/j.isprsjprs.2018.02.013](https://doi.org/10.1016/j.isprsjprs.2018.02.013) (cited page 31).
- Ruppen, Désirée, Owen A. Chituri, et al. (Dec. 2021). “Community-based monitoring detects sources and risks of mining-related water pollution in Zimbabwe”. In: *Frontiers in Environmental Science* 9. ISSN: 2296-665X. DOI: [10.3389/fenvs.2021.754540](https://doi.org/10.3389/fenvs.2021.754540) (cited pages 32, vi).
- Ruppen, Désirée, James Runnalls, et al. (Apr. 2023). “Optical remote sensing of large-scale water pollution in Angola and DR Congo caused by the Catoca mine tailings spill”. In: *International Journal of Applied Earth Observation and Geoinformation* 118, p. 103237. ISSN: 1569-8432. DOI: [10.1016/j.jag.2023.103237](https://doi.org/10.1016/j.jag.2023.103237) (cited page 31).
- Russ, Jason et al. (2020). *Salt of the Earth: Quantifying the impact of water salinity on global agricultural productivity*. Policy Research Working Paper 9144. Public Disclosure Authorized. World Bank (cited pages 5, viii).
- S&P Global Market Intelligence (2025). *Metals & Mining Database* (cited pages 12, iii).
- Sahoo, Satyabrata and Somnath Khaoash (Aug. 2020). “Impact assessment of coal mining on groundwater chemistry and its quality from Brajrajnagar coal mining area using indexing models”. In: *Journal of Geochemical Exploration* 215, p. 106559. ISSN: 0375-6742. DOI: [10.1016/j.gexplo.2020.106559](https://doi.org/10.1016/j.gexplo.2020.106559) (cited page vi).
- Sall, Mamadou et al. (Nov. 2020). “Water Constraints and Flood-Recession Agriculture in the Senegal River Valley”. In: *Atmosphere* 11.11, p. 1192. ISSN: 2073-4433. DOI: [10.3390/atmos11111192](https://doi.org/10.3390/atmos11111192) (cited page vii).
- Santana, Caroline S. et al. (2020). “Assessment of water resources pollution associated with mining activity in a semi-arid region”. In: *Journal of Environmental Management* 273, p. 111148. ISSN: 0301-4797. DOI: [10.1016/j.jenvman.2020.111148](https://doi.org/10.1016/j.jenvman.2020.111148) (cited pages 2, 5, vi).
- Schwarzenbach, Ren P. et al. (Nov. 2010). “Global water pollution and human health”. In: *Annual Review of Environment and Resources* Volume 35, 2010, pp. 109–136. DOI: [10.1146/annurev-environ-100809-125342](https://doi.org/10.1146/annurev-environ-100809-125342) (cited page 5).
- Sepin, Philipp, Lukas Vashold, and Nikolas Kuschig (2025). “Mapping Mining Areas in the Tropics from 2016–2024”. In: *Nature Sustainability*. forthcoming (cited pages 12, 13, 32, xxxii).

- Shen, Siyuan et al. (May 2024). “Enhancing global estimation of fine particulate matter concentrations by including geophysical a priori information in deep learning”. In: *ACS ES&T Air* 1.5, pp. 332–345. DOI: [10.1021/acsestair.3c00054](https://doi.org/10.1021/acsestair.3c00054) (cited pages 13, xxxiii).
- Shi, Hao et al. (Jan. 2017). “Assessing the ability of MODIS EVI to estimate terrestrial ecosystem gross primary production of multiple land cover types”. In: *Ecological Indicators* 72, pp. 153–164. ISSN: 1470-160X. DOI: [10.1016/j.ecolind.2016.08.022](https://doi.org/10.1016/j.ecolind.2016.08.022) (cited pages 6, 11).
- Sigman, H. (2002). “International spillovers and water quality in rivers: Do countries free ride?” In: *American Economic Review* 92.4, pp. 1152–1159. ISSN: 0002-8282. DOI: [10.1257/00028280260344687](https://doi.org/10.1257/00028280260344687) (cited page 2).
- Steenbergen, F van et al. (2010). *Guidelines on spate irrigation*. Tech. rep. Rome, Italy: Food and Agricultural Organization of the United Nations (cited page vii).
- Strobl, Eric and Robert O. Strobl (Nov. 2011). “The distributional impact of large dams: Evidence from cropland productivity in Africa”. In: *Journal of Development Economics* 96.2, pp. 432–450. ISSN: 0304-3878. DOI: [10.1016/j.jdeveco.2010.08.005](https://doi.org/10.1016/j.jdeveco.2010.08.005) (cited page 2).
- Swain, Ratnakar and Bhabagrahi Sahoo (May 2017). “Mapping of heavy metal pollution in river water at daily time-scale using spatio-temporal fusion of MODIS-aqua and Landsat satellite imageries”. In: *Journal of Environmental Management* 192, pp. 1–14. ISSN: 0301-4797. DOI: [10.1016/j.jenvman.2017.01.034](https://doi.org/10.1016/j.jenvman.2017.01.034) (cited page 32).
- UNEP (2023). *Global industry standard on tailings management*. Tech. rep. United Nations Environment Programme (cited page 6).
- United Nations Environment Programme (2025). *GEMStat database of the global environment monitoring system for freshwater (GEMS/Water) Programme*. Accessed 21 February 2025. Available upon request from GEMS/Water Data Centre: [gemstat.org](https://gemstat.org) (cited pages 13, 27, xi).
- Van Vliet, Michelle TH, Martina Flörke, and Yoshihide Wada (2017). “Quality matters for water scarcity”. In: *Nature Geoscience* 10.11, pp. 800–802. DOI: [10.1038/ngeo3047](https://doi.org/10.1038/ngeo3047) (cited pages 3, 30).
- Verbrugge, Boris, Cristiano Lanzano, and Matthew Libassi (2021). “The cyanide revolution: Efficiency gains and exclusion in artisanal- and small-scale gold mining”. In: *Geoforum* 126, pp. 267–276. ISSN: 0016-7185. DOI: [10.1016/j.geoforum.2021.07.030](https://doi.org/10.1016/j.geoforum.2021.07.030) (cited pages 5, 28).
- Verdin, K. L. and J. P. Verdin (May 1999). “A topological system for delineation and codification of the Earth’s river basins”. In: *Journal of Hydrology* 218.1, pp. 1–12. ISSN: 0022-1694. DOI: [10.1016/S0022-1694\(99\)00011-6](https://doi.org/10.1016/S0022-1694(99)00011-6) (cited page ii).
- Vörösmarty, Charles J et al. (2010). “Global threats to human water security and river biodiversity”. In: *Nature* 467.7315, pp. 555–561. DOI: [10.1038/nature09440](https://doi.org/10.1038/nature09440) (cited page 30).
- Wang, Fei et al. (Mar. 2024). “The leaching behavior of heavy metal from contaminated mining soil: The effect of rainfall conditions and the impact on surrounding agricultural

- lands”. In: *Science of The Total Environment* 914, p. 169877. ISSN: 0048-9697. DOI: [10.1016/j.scitotenv.2024.169877](https://doi.org/10.1016/j.scitotenv.2024.169877) (cited page vii).
- Weiss, D. J. et al. (2018). “A global map of travel time to cities to assess inequalities in accessibility in 2015”. In: *Nature* 553, pp. 333–336. ISSN: 1476-4687. DOI: [10.1038/nature25181](https://doi.org/10.1038/nature25181) (cited page 13).
- WorldPop (2018). *Global 1km Population*. Dataset. DOI: [10.5258/SOTON/WP00647](https://doi.org/10.5258/SOTON/WP00647) (cited page 13).
- Wu, Lijun et al. (2023). “Metal-mining-induced sediment pollution presents a potential ecological risk and threat to human health across China: A meta-analysis”. In: *Journal of Environmental Management* 329, p. 117058. ISSN: 0301-4797. DOI: [10.1016/j.jenvman.2022.117058](https://doi.org/10.1016/j.jenvman.2022.117058) (cited pages 3, 5).
- Wuepper, David et al. (July 2023). “Institutions and global crop yields”. In: *NBER Working Paper*. DOI: [10.3386/w31426](https://doi.org/10.3386/w31426) (cited pages 5, 6, 29).
- Zeng, Yelu et al. (July 2022). “Optical vegetation indices for monitoring terrestrial ecosystems globally”. In: *Nature Reviews Earth & Environment* 3, pp. 477–493. ISSN: 2662-138X. DOI: [10.1038/s43017-022-00298-5](https://doi.org/10.1038/s43017-022-00298-5) (cited page 11).
- Zörb, C., C.-M. Geilfus, and K.-J. Dietz (2019). “Salinity and crop yield”. In: *Plant Biology* 21.S1, pp. 31–38. ISSN: 1435-8603. DOI: [10.1111/plb.12884](https://doi.org/10.1111/plb.12884) (cited pages 5, viii).

## A. Basins and Mines

Here, we describe the basin and mine datasets, and how we integrate them to create our sample of interest.

### A1. The HydroBASINS dataset

Our analysis uses river basins (watersheds) from the HydroBASINS dataset (Lehner and Grill, 2013) as units of observation.<sup>22</sup> This dataset divides all land mass on earth into twelve levels of nested river basins, offering increasing granularity at each level.

**Hierarchical basin structure** The hierarchy begins at the least granular Level 1, where each continent forms a single basin. Level 2 divides continents into nine similarly sized units, while Level 3 delineates major river systems. Each subsequent level follows hydrological principles to create up to nine sub-basins within each higher-level basin. This nesting continues through Level 12, our chosen level of analysis, where basins across Africa average 124.4 km<sup>2</sup> in area.

#### A1.1. Basin delineation

The delineation process begins with the smallest sub-basins, which are aggregated into larger units. In the ideal case, this leads to a unitary Level 3 basin that contains a large river system. This system consists of a main stream flowing from its source to either the sea or an inland sink. Along its course, the main stream is joined by multiple tributary streams. Each such confluence presents an opportunity to delineate two distinct basins: one encompassing the tributaries' catchment area, and another for the main stream.

To maintain accuracy, while avoiding excessive fragmentation, the authors maintain thresholds before a confluence is used to delimit two basins:

- tributary streams must have a catchment area of (i.e., drain) at least 100 km<sup>2</sup> to form their own tributary sub-basin,
- areas between qualifying tributary streams from inter-basins,

---

<sup>22</sup>We use the version of the dataset that specifically accounts for the position of lakes, delineating lake-adjacent basins similarly to coastal basins.

- if the catchment area of tributary streams or inter-basins exceeds 250 km<sup>2</sup>, they are subdivided using artificial break points.

The resulting boundaries are hydrologically determined, and independent of political borders, which facilitates our analysis.

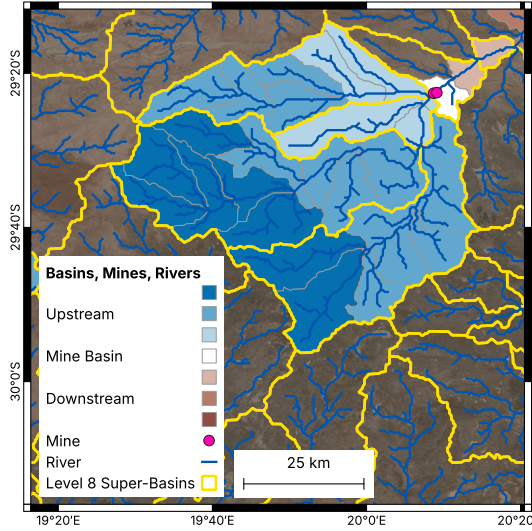


FIGURE A1: Two forked upstream (Level 12) basins join into a single (mine) basin further downstream. The superimposed yellow lines indicate Level 8 basins; these contain varying numbers of sub-basins (due to a level-skipping mechanism), and clearly divide tributary and main basins. The blue lines, which represent river streams, provide additional intuition for the basin topology.

**Basin aggregation and coding** The dataset employs a modified version of the Pfafstetter coding system (K. L. Verdin and J. P. Verdin, 1999) to aggregate basins into higher-level units. Each super-basin contains a maximum of nine sub-basins: (a) four tributary basins, and (b) five sections of the main stream (as defined by tributary confluences or the artificial breaking points). The HydroBASINS dataset departs from the plain Pfafstetter coding in two notable ways:

1. it allows level-skipping when basin areas at a given level deviate significantly from their peers,
2. it permits super-basins to contain fewer than nine sub-basins.

An example of the level-skipping mechanism is visible in Figure A1. Additional adjustments of the HydroBASINS dataset concern endorheic (closed) basins and islands. Islands are grouped with their associated continents at Level 1 and then manually grouped or separated at Levels 2 and 3. At subsequent levels, basins are nested in an island-specific Pfafstetter chain. Endorheic basins are contained entirely

within one super-basin, but do not drain into that larger basin. For consistency, these basins are linked to main streams via virtual links that do not correspond to actual flows. We sever these virtual links for our analysis.

## A2. The mine dataset

Our information on mine locations comes from Maus, Giljum, et al. (2022), who developed a comprehensive dataset of mining areas by expanding on the Metals and Minerals database (S&P Global Market Intelligence, 2025). While the SNL database contains information on approximately 35,000 industrial mines globally, Maus, Giljum, et al. (2022) enhanced this coverage with additional sources and visual inspection of satellite imagery. Entries are generated by inspecting a 10 km buffer area around recorded mine locations for signs of mining operations, which are then delineated. This means that both active and abandoned industrial mines, as well as nearby artisanal and small-scale mine sites, which often continue after industrial operations cease, are covered. The resulting dataset contains 45,000 mine polygons across the globe, around 5,000 of which are located in Africa. Their geographic distribution can be seen in Figure A2.

## A3. Integration of mine and basin data

The described datasets allow us to identify comparable areas that are affected and unaffected by the stream impacts of mines. Specifically, we construct two chains of basins in relation to each mine basin:

- *Downstream basin chains*: we follow the variable indicating the next basin until the sea, a sink, or a distance threshold (ten basins in the main specification) is reached.
- *Upstream basin chains*: we recursively track basins that reference the current basin as the next one, until either an end or a threshold is reached.

As a result of this process, downstream chains follow a single path, while upstream basins may fan out.<sup>23</sup>

---

<sup>23</sup>This is because river bifurcations are rare (and usually non-permanent), while confluences are abundant. While theoretically possible, downstream bifurcations are ruled out by the HydroBASINS dataset's structure.

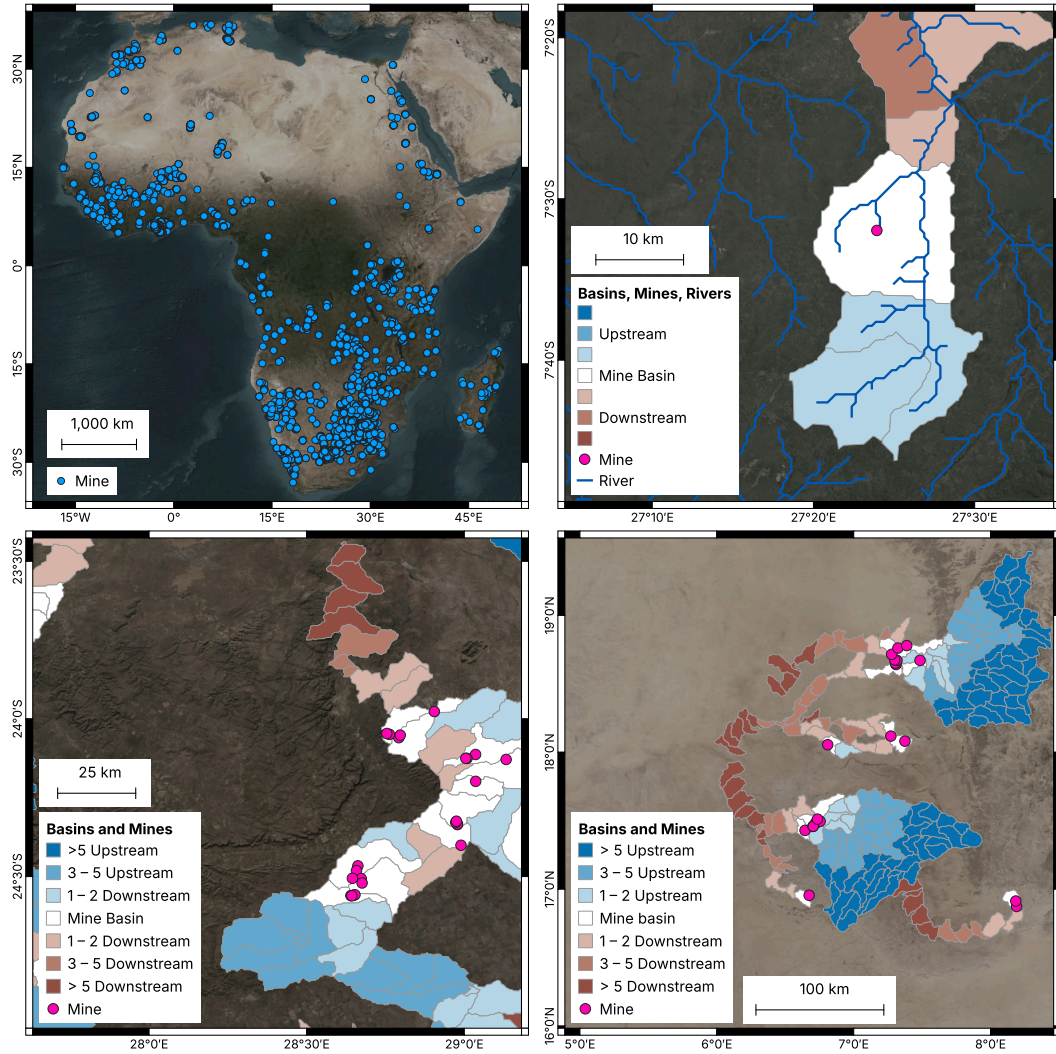


FIGURE A2: Distribution of the mine polygons in Maus, Giljum, et al. (2022) (top-left). Illustration of a mine that is entirely contained within one basin (top-right), mine clusters that reach across multiple basins (bottom-left), and a number of interconnected and closely adjacent basin chains (bottom-right).

**The mine basin** Mine basins are assigned by intersecting the basins with the centroids of the mine polygons by Maus, Giljum, et al. (2022). Two examples are shown in Figure A2, where mines are represented by a pink dot within white mine basins. On the left panel, we can see that the basin contains both areas that are downstream of the mine, indicated by the superimposed river stream, and ones that are upstream. By contrast, the downstream (upstream) basins, marked in brown (blue), and upstream basins contain only areas that are (are not) affected by water flows passing the mine.

### **A3.1. Treatment assignment**

Mines are usually spatially clustered, and, as illustrated in the right panel of Figure A2, our dataset is no exception. This complicates the assignment of the treatment (control) status based on the classification of up- and downstream basin chains, since a single basin may appear in multiple chains. We apply the following rules:

- Basins that only appear in upstream chains are designated as upstream, i.e., control.
- Basins that appear in any downstream chain are designated as downstream, i.e., treated.

This coding affects, for example, the two basins directly to the south of the large mine in the right panel of Figure A2. Even though they are upstream of the larger mine, we designate them as downstream basins because they are downstream of (and thus affected by) a set of smaller mines south of the larger mine. They thus cannot be treated as upstream, that is, unaffected by a mine.

The described treatment assignment hinges on the length of basin chains, which we set at a maximum order of ten. A greater threshold would result in long downstream chains that reach into the control areas of distant mines, while a smaller threshold relies on the quick dissipation of impacts or runs the risk of contamination. An illustration is provided in Figure A2, where the downstream chain, which originates from the mine cluster in the southeastern (bottom-right) corner, reaches into the upstream area of a group of mines at the center of the map.

## **B. Mining, Water Pollution, and Vegetation**

In this section, we investigate the mechanisms behind our estimated causal effect, linking Mines–Rivers–Yields. We begin by reviewing the literature on how mining affects water and impairs plant growth. Next, we correlate our satellite-derived productivity measure (peak vegetation) with regional agricultural production data to quantify impacts in practical terms. Finally, we analyze water quality measurements to provide direct evidence of pollution as the primary mediator of mining’s effects on vegetation.

### **B1. Mine impacts on water quality and vegetation**

Extensive research documents the degradation of water quality downstream of mining operations. Pollution affects both surface water and groundwater and has significant implications for agricultural productivity and ecosystem health.

Regional studies provide compelling evidence of mining’s water quality impacts. In Zimbabwe’s Deka River, Ruppen, Chituri, et al. (2021) documented a severe water quality decline downstream of mining discharge points, with manganese concentrations reaching 70 times the safe limit and elevated levels of nickel, arsenic, and salinity. Similarly, Duncan (2020) and Duncan, Vries, and Nyarko (2018) found elevated concentrations of nickel, chromium, cadmium, and lead in Ghana’s Pra and Fena rivers and their tributaries. They identify illegal mining activities as a key source of these pollutants. In India’s Brajrajnagar coal mining area, S. Sahoo and Khaoash (2020) found that 15% of groundwater samples were of poor quality, with 43% requiring special treatment before agricultural use due to heavy metal contamination. In the Brazilian Jacaré and Contas rivers, Santana et al. (2020) identified toxic levels of cadmium, lead, and uranium in water, while chromium, copper, nickel, and vanadium in sediments exceeded international safety guidelines.

#### **B1.1. Pathways of contamination**

Mining-induced pollution spreads through water systems and affects vegetation through contaminated water and soils that disrupt plant physiology and reduce agricultural productivity. While vegetation that is irrigated directly with polluted river water faces the most immediate impacts, rainfed areas near contaminated streams may suffer significant effects as pollutants disperse through multiple pathways.

Three primary mechanisms distribute mining pollutants beyond immediate river channels. First, river floodplains extend well beyond original stream boundaries, and recurrent floods deposit pollutants across vast non-irrigated cropland areas (Macklin et al., 2023). Second, riverine streams are in close exchange with surrounding areas, leaching and seeping into nearby soils (Palumbo-Roe, Wragg, and Banks, 2012). Contaminated soils then spread to nearby farmlands through rain (see F. Wang et al., 2024). Third, mining pollutants infiltrate shallow groundwater and travel through subsurface soils. Since groundwater gradients often align with surface water patterns, dry periods enable capillary rise that draws contaminants back into top-level soils (Haghighizadeh et al., 2024).

In addition, farming directly on floodplains and the redirection of floods for sediment deposition is common across Africa. While only around 6.5% of the area in Africa is considered as part of floodplains, more than 16% of croplands are located in such areas and cropland density is disproportionately higher in (seasonal) inundation zones (Dryden et al., 2021). In addition, flood recession farming is common in West Africa (see, e.g., Sall et al., 2020), whereas spate farming (using flashfloods for sediment extraction and deposition) is common in North and Eastern Africa as well as more sporadically in other regions throughout Africa (Steenbergen et al., 2010). These complex interconnections between water systems, nearby soils (Cheng et al., 2021) and farming practices explain how mining-induced river pollution transmits to non-irrigated areas downstream of mines.

## **B1.2. Water pollution and plant growth**

Mining operations produce several forms of water pollution that impair plant growth and development. These pollutants — primarily heavy metals, acidic drainage, and dissolved salts — significantly reduce agricultural productivity through multiple physiological pathways.

**Heavy metals** Heavy metal contamination from mining represents a major threat to plant health and crop yields, and acts by disrupting essential metabolic processes, damaging cellular structures and the soil microbiome. Frossard et al. (2018) show that mercury concentrations exceeding 2 mg/kg disrupt the soil microbiome, reducing nitrogen fixation and phosphorus solubilization by up to 20%. This disruption correlates with approximately 25% yield losses in maize — a staple crop across much

of Africa. Similarly, lead concentrations above 50 mg/kg decrease chlorophyll content by approximately 15% and inhibit root development, resulting in maize yield reductions of 30–35%. These effects are even more pronounced in leafy vegetables, which show particularly rapid uptake of contaminants. The results of heavy metal pollution include chlorosis, where leaves yellowing due to insufficient chlorophyll production, and even cell death (necrosis).

**Acid mine drainage** Acid mine drainage represents one of the most severe forms of water pollution from mining. When sulfide minerals in mine waste (such as pyrite) are exposed to oxygen and water, they form sulfuric acid and dramatically lower the pH level in affected watersheds. Extremophile microbes can contribute to this process and can sustain it for decades or centuries (see, e.g. Nordstrom et al., 2000). The acidification has profound effects on plant physiology and growth, and acts by disrupting nutrient uptake of calcium and magnesium, reducing the availability of nitrogen and phosphorus in soil, and increasing solubility of toxic metals like aluminum and manganese. The impact on crop yields is substantial and varies by crop type. Du et al. (2024) found an average yield reduction of 13.7% overall, with more severe impacts on vegetables (33%), and significant effects on staple crops like maize and wheat (18%). These yield losses result from reduced root growth, impaired nutrient uptake, and cellular damage from the compounded toxicity of heavy metals.

**Salinity** Water salinity, often measured through Electrical Conductivity (EC), represents another significant mining-related pollutant that affects plant growth and crop yields (Russ et al., 2020). Salinity stress impacts plants through osmotic effects and nutrient imbalances (Parida and Das, 2005). At the cellular level, high salinity disrupts membrane integrity, inhibits enzyme activity, and compromises plants' ability to detoxify. These physiological disruptions manifest as visible growth impairments. Research shows that even a modest 1,000  $\mu\text{S}/\text{cm}$  (microsiemens per centimeter) increase in irrigation water EC can reduce biomass accumulation in maize by approximately 2%. The effect intensifies at higher salinity levels — when salinity increases from moderate (3,000  $\mu\text{S}/\text{cm}$ ) to high levels (6,000  $\mu\text{S}/\text{cm}$ ), maize grain yields may decline by 33%. Some studies indicate potential yield losses exceeding 50% when EC values surpass 8,000  $\mu\text{S}/\text{cm}$  (Zörb, Geilfus, and Dietz, 2019), with varying impacts by crop type (Page et al., 2021).

## B2. Vegetation indices and crop yields

We map satellite-derived vegetation to agricultural outcomes using two independent micro datasets: (i) the Advancing Research on Nutrition and Agriculture (AReNA) DHS–GIS Database of IFPRI (IFPRI, 2020), and (ii) the World Bank Living Standards Measurement Study–Integrated Surveys on Agriculture (LSMS–ISA) (Bentze and Wollburg, 2024). The AReNA DHS–GIS data cover downscaled yields and production values for various crops at survey sites across 34 African countries (about 45,000 observations, 2001–2018). The LSMS–ISA data cover eight African countries (Ethiopia, Malawi, Mali, Niger, Nigeria, Tanzania, Uganda; about 93,700 observations for all plots and 54,100 observations for plots that contain cereal crops, 2008–2020).

AReNA-DHS includes agricultural-production statistics by applying a downscaled model (SPAM) that integrates multiple input layers, producing estimates for the farmers captured in the survey. LSMS-ISA, by contrast, reports crop-production figures obtained directly from field surveys. While AReNA-DHS covers a broader set of countries from our sample, LSMS-ISA provides real on-the-ground survey data. For each DHS cluster and LSMS–ISA enumeration area, we draw 6.2 km radius disks (matching our average basin size) and extract the maximum annual EVI in the survey year. We relate the *log* of agricultural outcomes (winsorized at the 1% level) to this peak EVI to obtain semi-elasticities (percent changes in outcomes per 1.0 EVI unit). For both AReNA-DHS and LSMS–ISA we include wave fixed effects as well as crop fixed effects for LSMS–ISA.

We focus on: (1) *physical yield* (kg/ha), (2) *financial yield* (USD/ha), and (3) *production value* (USD; total value). To aid interpretation, we summarize effects per a 0.1 increase in EVI.

Across both datasets, higher vegetation is strongly associated with higher agricultural performance (Tables B1 and B2). For instance, using the AReNA DHS-GIS data, a 0.1 unit increase in EVI is associated with a 9.4% rise in cereal yields (Table B1, column 5) and a 10.0% increase in the financial yield for all crops (column 3). The associations are even stronger in the LSMS-ISA dataset, where the same EVI increase corresponds to a 17.7% rise in cereal yields (Table B2, column 3) and a 13.2% increase in all-crop financial yield (column 2).

TABLE B1: Maximum annual EVI and agricultural production — AReNA DHS–GIS

| Outcome:<br>Model:    | ln(All Crops, Value)<br>(1) | ln(All Crops, Yield)<br>(2) | ln(All Crops, FY)<br>(3) | ln(Cereals, Value)<br>(4) | ln(Cereals, Yield)<br>(5) | ln(Cereals, FY)<br>(6) |
|-----------------------|-----------------------------|-----------------------------|--------------------------|---------------------------|---------------------------|------------------------|
| <i>Variables</i>      |                             |                             |                          |                           |                           |                        |
| Max. Cropland EVI     | 3.398***<br>(0.4230)        | 0.7286***<br>(0.2376)       | 0.9519***<br>(0.1828)    | 2.489***<br>(0.9150)      | 0.8995***<br>(0.1586)     | 0.5589**<br>(0.2704)   |
| <i>Fixed effects</i>  |                             |                             |                          |                           |                           |                        |
| Wave                  | Yes                         | Yes                         | Yes                      | Yes                       | Yes                       | Yes                    |
| <i>Fit statistics</i> |                             |                             |                          |                           |                           |                        |
| Observations          | 44,682                      | 44,380                      | 44,380                   | 44,682                    | 44,682                    | 44,171                 |
| R <sup>2</sup>        | 0.65336                     | 0.54083                     | 0.35656                  | 0.50120                   | 0.60944                   | 0.32956                |
| Within R <sup>2</sup> | 0.08225                     | 0.01013                     | 0.00717                  | 0.02177                   | 0.02195                   | 0.00153                |

Clustered (wave) standard errors in parentheses. Significance: \*\*\* 0.01, \*\* 0.05, \* 0.1.

TABLE B2: Effect of Maximum EVI on Crop Outcomes — LSMS–ISA

| Outcome:<br>Model:    | ln(All Crops, Yield)<br>(1) | ln(All Crops, USD/ha)<br>(2) | ln(Cereals, Yield)<br>(3) | ln(Cereals, USD/ha)<br>(4) |
|-----------------------|-----------------------------|------------------------------|---------------------------|----------------------------|
| <i>Variables</i>      |                             |                              |                           |                            |
| Max. Cropland EVI     | 1.866***<br>(0.5007)        | 1.242**<br>(0.5567)          | 1.627***<br>(0.5047)      | 1.064**<br>(0.5089)        |
| <i>Fixed effects</i>  |                             |                              |                           |                            |
| Country–Year          | Yes                         | Yes                          | Yes                       | Yes                        |
| Crop Type             | Yes                         | Yes                          | Yes                       | Yes                        |
| <i>Fit statistics</i> |                             |                              |                           |                            |
| Observations          | 93,725                      | 93,113                       | 54,066                    | 53,971                     |
| R <sup>2</sup>        | 0.30034                     | 0.42089                      | 0.34286                   | 0.44959                    |
| Within R <sup>2</sup> | 0.00768                     | 0.00286                      | 0.00705                   | 0.00253                    |

Clustered (country–year) standard errors in parentheses. Significance: \*\*\* 0.01, \*\* 0.05, \* 0.1.

### **B2.1. Translating mining impacts on EVI into agricultural losses**

To estimate the aggregate economic impact, we calculate the total physical production loss. In our main analysis we find that mining activity lowers the maximum annual cropland EVI in the first three downstream basins by about 0.0068 units (see [Table 2](#)). We map this decline to the crop-specific semi-elasticities obtained from the two datasets. Using the AReNA DHS–GIS coefficients, our main results translate into a 0.496% decrease in crop yields for all crops and a 0.613% decrease in cereal yields. The elasticities derived from LSMS–ISA are larger: they imply a 1.263% reduction in crop yields for all crops and a 1.102% drop in cereal-crop yields.

The total cropland area in the first three downstream basins amounts to slightly more than 74,000 km<sup>2</sup> (or 7.4 million hectares). Using the AReNA-DHS sample (mean cereal yield  $\approx$  2,000 kg/ha) the implied annual shortfall is 91,200 t of cereals ( $2,000 \text{ kg/ha} \times 0.00613 \text{ (yield reduction)} \times 7.4 \text{ million ha}$ ). The LSMS-ISA data has a higher average yield (mean yield 2,497 kg/ha) and a higher elasticity of EVI to crop yields. This implies a larger reduction in cereal crop production than AReNA-DHS of about 205,408 t ( $2,497 \text{ kg/ha} \times 0.01102 \text{ (yield reduction)} \times 7.4 \text{ million ha}$ ).

To express the revenue effects in monetary terms, we multiply the average revenue per hectare, the proportional loss, and the affected area of 7.4 million hectares. In the AReNA DHS–GIS sample, mean revenues are  $\approx$  USD 1,200/ha for cereals and  $\approx$  USD 2,100/ha for all crops. Applying the estimated declines of 0.381% for cereal revenues and 0.649% for all-crop revenues yields annual losses of approximately USD 34.3 million for cereals and USD 102.1 million for all crops. In the LSMS–ISA data, lower average revenues of USD 795/ha for cereals and USD 1,651/ha for all crops are offset by larger declines of 0.725% and 0.845%, respectively. The resulting income losses amount to approximately USD 42.8 million for cereals and USD 103.7 million for all crops.

### **B3. Water pollution measurements**

To provide direct evidence for water pollution as the primary mechanism linking mining to reduced agricultural productivity, we analyzed water quality measurements from the United Nations Environment Programme ([2025](#)) database. This dataset contains water samples voluntarily provided by countries and organizations from

their monitoring networks. We included all samples collected between 2016 and 2024 within our study region, though availability was limited to South Africa.<sup>24</sup>

Figure B3 presents six key water quality parameters measured in upstream, mine, and downstream basins:

1. **Calcium discharge** (Ca-Dis) indicates increased mineral content in water.
2. **Electrical conductivity** (EC) measures salinity (dissolved salts and ions), which inhibits plant growth.
3. **Sodium discharge** (Na-Dis) damages soil structure and impairs water uptake.
4. **pH** measures water acidity or alkalinity.
5. **Sodium adsorption ratio** (SAR) indicates the potential for soil structure deterioration.
6. **Sulfate discharge** ( $\text{SO}_4^{2-}$ ) can cause acidification and indicates mining-related pollution.

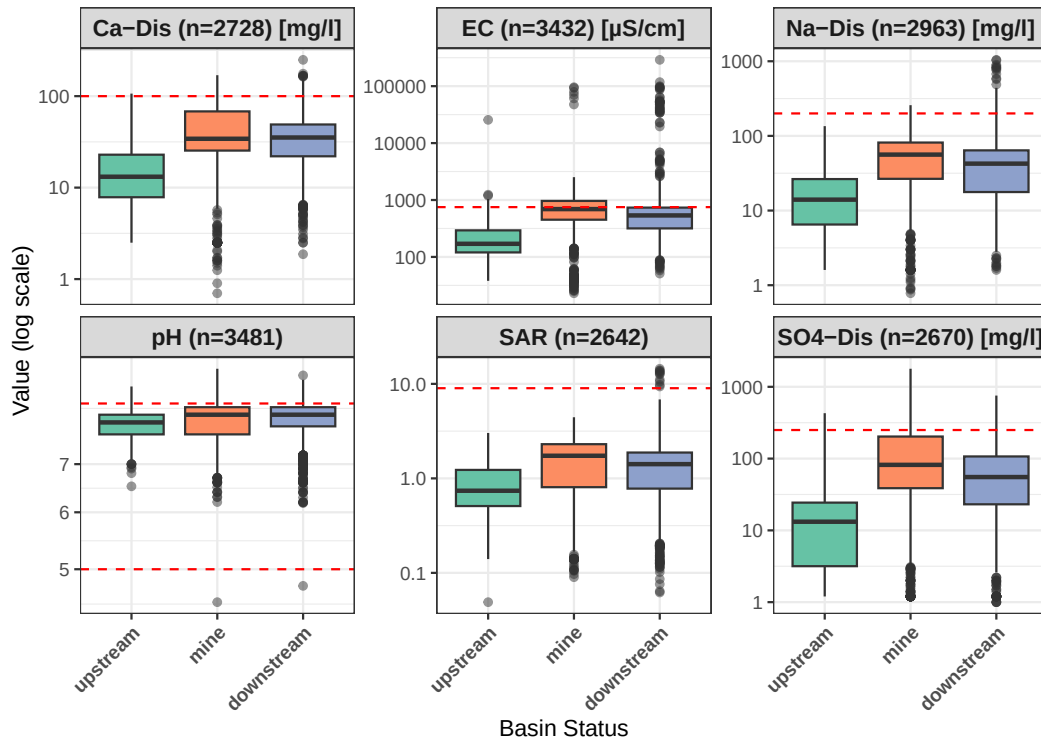
The data provide compelling evidence of water pollution in mine basins and downstream areas compared to upstream sites. Approximately 50% of samples from mine basins exceed the 750  $\mu\text{S}/\text{cm}$  threshold for EC. Moreover, elevated median values for electrical conductivity, calcium discharge, and sodium discharge in downstream basins indicate that mining operations are driving an accumulation of dissolved salts. Acidity levels (pH) remain relatively stable, although the elevated sodium absorption ratio (SAR) reveals a risk for developing sodic soil conditions, and higher sulfate concentrations downstream may indicate mining-induced acidification.

These water quality measurements, though limited in scope, align with our hypothesis that water pollution is the primary mechanism through which mining operations reduce agricultural productivity in downstream areas. Several limitations persist, however: (a) measurements were only available from South Africa, covering a particular portion of our study area, (b) direct measurements of critical heavy metals (lead, mercury, arsenic, cadmium) were unavailable, (c) temporal granularity is lacking, prohibiting in-depth analysis that accounts for seasonal patterns, (d) measurements are inconsistent and were collected using various methods, potentially limiting comparability. Despite these limitations, the available data support our mechanism hypothesis

---

<sup>24</sup>The availability of water quality data is a particular problem in Africa (Jones et al., 2024).

FIGURE B3: Water quality indicators for up-, mine, and downstream basins.



**Notes:** This figure presents key water quality parameters monitored across different basins in South Africa between 2016-2024. Electrical conductivity (EC) estimates the overall dissolved ionic content and measures water's salinity. Sulfate discharge (SO4-DIS) can turn water acidic. The pH values reflect the water's degree of acidity or alkalinity. The sodium adsorption ratio (SAR) assesses the potential for sodium to affect soil structure during irrigation. Dashed red lines indicate levels that are potentially harmful for plant growth (Ayers and Westcot, 1985).

and demonstrate a clear pattern of elevated pollution levels in mine and downstream basins compared to upstream control areas.

## C. Additional Results and Methods

In this section, we report additional results that complement our main analysis. First, we explore alternative specifications for extrapolating mining impacts beyond the immediate vicinity, using exponential-decay and linear-quadratic distance-based models. Second, we describe our approach to adding commodity information to our dataset. Lastly, we detail additional robustness checks, including our matching approach, treatment randomization, and placebo outcomes.

### C1. Distance-based specifications

Our main analysis uses the basin order to estimate mining impacts on vegetation. Here, we complement this approach with distance-based specifications to help quantify how impacts develop over the course of rivers.

#### C1.1. Exponential decay model

The transport of pollutants from mining operations via rivers is the primary transmission channel to downstream basins. Hydrological studies indicate that over 90% of pollutants from mining are sediment-associated and transported 10–100 kilometers from their discharge point (see Macklin et al., 2023). Theory and empirical evidence (see, inter alia, references in Macklin et al., 2023) suggest that concentrations decay non-linearly.

We therefore employ an exponential decay model, described in the main text, to characterize impact patterns over longer distances. Note, however, that our original research design is focused on the immediate vicinity. At greater distances, up- and downstream basins are no longer directly comparable as structural differences mount; hence, estimates from this analysis cannot be interpreted as causal under the same weak conditions as our main results.

Since the decay parameters  $\delta, \gamma$  are not known a priori, we use a (Bayesian) model-averaging approach to (i) estimate their values while (ii) conveying uncertainty around them. We consider a grid of values between  $[0.001, 1]$ , accommodating rapid and slow decay patterns at either side of the cutoff. At the lower (higher) bound, the exponential decay acts along meters (kilometer) of river distance. For each combination  $j$  of decay parameters, we estimate the model(s) and compute the Bayesian

information criterion (BIC) to quantify model fit. We approximate posterior probabilities for each model via  $p(\delta_j, \gamma_j | \mathcal{D}) = \exp\{-\text{BIC}_j/2\} / \sum_j \exp\{-\text{BIC}_j/2\}$  (see e.g. Neath and Cavanaugh, 2012) and a prior. Instead of imposing a flat prior for the decay parameters, we use a moderately informative one, with  $\delta_j, \gamma_j \sim \text{Be}(5.6, 1.4)$ . This implies that impacts decay quickly (at a mean value of 0.8), and represents a conservative prior for our analysis. This procedure allows us to report posterior means (corresponding to regularized maximum likelihood estimates) and use the posterior distribution of parameters to express uncertainty around them.

Our results are available in the lower panel of [Table E13](#), and reveal surprisingly slow rates of decay. Initial impacts (at hypothetical zero distance) range from  $-0.0060$  to  $-0.0093$ , while the average decay parameters range from  $0.035$  to  $0.002$ , suggesting very slow rates of decay. A re-analysis with flat priors (mirroring maximum likelihood estimates) diverges towards a flat downstream indicator, with estimates tending towards zero. [Figure C4](#) visualizes the speed of decay along the river network for the fully saturated specifications. Posterior means indicate that impacts on vegetation halve at a distance of 281 km, while cropland impacts halve at 72 km. This implies minimal decay over the sample, where non-mine basins lie at a mean (median) river distance of 51.5 (45.8) km, with a maximum of 216 km. These impact ranges reach beyond typical detection ranges of pollutants from hydrological studies (see, e.g., Macklin et al., 2023).

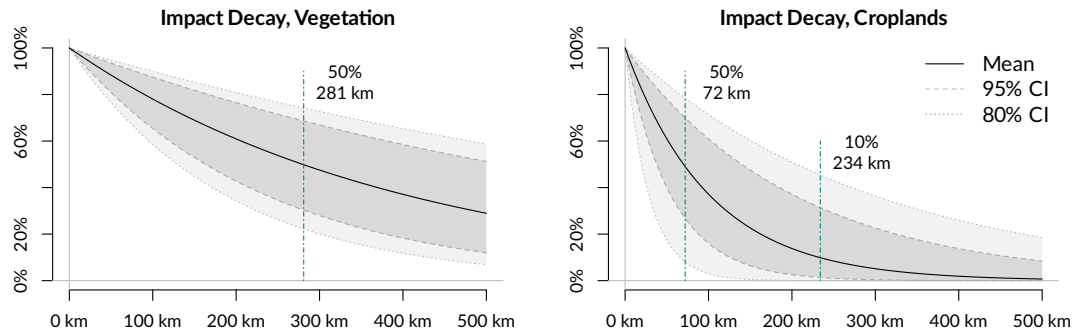


FIGURE C4: Impact decay over the river distance, assuming an exponential decay function. The solid black line denotes the mean effect; the shaded area between the dotted (dashed) lines denotes the 95% (80%) credible interval. The vertical lines denote the distance where the average impact is reduced to 50% and 10%.

### C1.2. Linear-quadratic distance specifications

In addition to the exponential decay model, we also employ polynomial specifications using river distance in kilometers as the running variable. These specifications use the following operationalization of the running distance:

$$F(x) = \left(|x| + x^2\right) \times \mathbb{I}(x < 0) + \mathbb{I}(x = 0) + \left(1 + x + x^2\right) \times \mathbb{I}(x > 0),$$

where  $x$  denotes the river distance relative to the mine, and we consider polynomials up to order two (avoiding issues discussed in Gelman and Imbens, 2019).

Overall, we detect a negative effect of being downstream to a mine across these specifications. Results are reported in Table E13. The linear distance decay specification indicates that downstream basins have a  $-0.0034$  to  $-0.005$  lower annual maximum EVI, though not all estimates are statistically significant. For the quadratic specification, we find statistically significant effects across all models, with downstream coefficients ranging from  $-0.0050$  to  $-0.0056$  for the general vegetation EVI and from  $-0.072$  and  $-0.0077$  for the cropland-specific EVI. All estimates are significant at the 5% level. These magnitudes closely align with our main basin-order specification for basins near the discontinuity.

**Robustness of linear-quadratic distances** We also investigate the robustness of the linear-quadratic specifications, although we note that distance-based polynomials poorly fit the impact of interest and should be interpreted with care. Most notably, this concerns the choice of bandwidth around the cutoff, which limits the influence of distant observations and is natively addressed by our chosen order and exponential specifications. Here, we follow the recent literature on inference in regression-discontinuity designs with continuous running variables (see Cattaneo, Idrobo, and Titiunik, 2019). We use a data-driven bandwidth selection procedure, a weighting scheme for observations that are closer to the cutoff, and separately fitted local polynomials for untreated and treated units (following Imbens and Kalyanaraman, 2012).<sup>25</sup> We follow the set of practices as outlined by Cattaneo, Idrobo, and Titiunik (2019) and employ a triangular kernel, which gives observations closer to the

<sup>25</sup>Note that our main specification includes observations up to the order 10. We implicitly employ a uniform kernel (weighing all observations equally), but separately estimate local treatment effects for treated and untreated units at each order.

cutoff a greater weight, and choose the bandwidth by minimizing the mean squared prediction error.

The results of this exercise are presented in [Table E14](#). For the conventional estimates, the optimal bandwidth ranges from 18.4 to 41.9 km across specifications. The bias-corrected estimates use wider bandwidths between 42.1 and 78.5 km. These bandwidths align with hydrological studies showing elevated toxic pollutants 10–80 km downstream of mines.<sup>26</sup> Despite the narrower bandwidths than for our main specification, the estimated effects remain consistent. With full controls, the local average treatment effects range from  $-0.0028$  to  $-0.0061$  for general vegetation and from  $-0.0045$  to  $-0.0083$  for cropland vegetation. These estimates remain statistically significant after bias correction.

## C2. Commodity type prediction

The environmental impacts of mining vary by the type of mineral being extracted, as different commodities require distinct extraction processes, chemicals, and produce different waste profiles. To investigate this heterogeneity, we developed a methodical approach to extend our dataset with information on commodity types present.

First, we compiled commodity information for relevant mine sites from multiple sources: the SNL Mines and Metals database, the Global Energy Monitor database, data by the US Geological Survey (Padilla et al., 2021), and company reports (Jasansky et al., 2023). After collection, we standardize commodity classifications across sources by harmonizing different naming conventions and variants of the same minerals. Then, we use Gaussian process regression with a Gaussian kernel to predict commodity probabilities for mine sites based on their coordinates. This allows us to predict the probabilities of different commodity types occurring at each mine location in our dataset.

[Figure D8](#) illustrates the predicted spatial distribution of four common commodities (gold, copper, coal, and diamonds) across our study area. While the visualization shows broad patterns, the actual predictions operate at the much finer spatial scale of our training data, allowing us to differentiate between neighboring mining sites. However, commodities often co-occur and their environmental impacts may interact in complex ways, we focus our heterogeneity analysis on these four major commodities

---

<sup>26</sup>Macklin et al. (2023), e.g., find elevated levels of toxic pollutants like zinc, lead, and arsenic between 10 and 80 km downstream of mines.

that we can isolate in mine basins. This allows us to investigate commodity-specific effects in a straightforward way, leaving extensions to future research.

### **C3. Further robustness checks**

This section describes and details robustness checks conducted to validate our main findings. We employed three approaches — matching methods to achieve covariate balance, randomization of treatment assignment, and placebo outcome tests. These checks strengthen confidence in the causal impacts of mining operations and reduced vegetation health downstream.

#### **C3.1. Balance and matching**

The characteristics of river basins may differ systematically between upstream (control) and downstream (treatment) basins. Potential imbalances are likely related to the nature of basins and river streams, with elevation and its correlates playing a central role (see [Appendix A](#) for more details). Our research design is not invalidated by such imbalances, but would suffer from decreased precision and higher dependence on exact model specifications. To counteract this, we use coarsened exact matching (Iacus, King, and Porro, 2012) to achieve covariate balance among groups in a flexible non-parametric fashion.

We implement two matching strategies with increasing stringency. First, we matched solely on elevation and slope — the characteristics that are most closely related to basin systems. The upper panels of [Figure D12](#) show an imbalance of these covariates in the unadjusted sample, which is purged by the matching procedure. Second, we expanded our matching criteria to include meteorological conditions (rainfall and maximum temperature) and soil type. The lower panels of [Figure D12](#) show that the pre-matching imbalances in temperature and rainfall were effectively eliminated in the adjusted sample. [Figure D13](#) shows negligible absolute standardized mean differences, confirming the success of the matching procedure.

Using the weights derived from these matching procedure, we re-estimate the treatment effect for the first three downstream basins. Results are reported in [Figure 6](#) and columns (7) and (8) of [Table E11](#). Both matching approaches yield estimates that are qualitatively similar to our main results, confirming the negative effect of mining on vegetation health and agricultural productivity. The effects were somewhat

stronger when matching on the full set of covariates, though not statistically different from our main results.

### **C3.2. Randomized treatment**

To further validate our identification strategy, we conducted a randomization exercise by shuffling the treatment status of basins. We randomly reassigned the downstream status of basins by changing the sign of the running variable, maintaining the overall balance between upstream and downstream locations. We preserve the status of the mine basin, which is not identified by our procedure and causally interpreted by us.

[Figure D11](#) presents results from 5,000 iterations of this randomization procedure. For the first three downstream basins, the estimated coefficients are centered near zero for both outcomes, with and without covariates. Our point estimates from the main analysis, indicated by red crosses, fall well outside these randomized distributions. This randomization exercise suggests that our findings are not artifacts of random variation but reflect genuine treatment effects.

### **C3.3. Placebo outcomes**

In another validation exercise, we examined whether discontinuities exist at the mine basin for covariates that should not be directly affected by mining's water-mediated impacts. [Figure D9](#) shows results from estimating our main specification using each of the main covariates as an outcome variable.

Population and temperature show no statistically significant discontinuities at the mine basin. Slope and elevation exhibit expected systematic trends moving from upstream to downstream areas (consistent with the nature of basins) but display no apparent discontinuity at the mine location. Accessibility to cities follows a U-shaped pattern, indicating that mines tend to be situated closer to population centers, which is also reflected in the population spike for mine basins. None of the covariates displays the distinctive pattern observed for vegetation in our main analysis.

We also conduct this exercise with river distances as running variables, and find similar results that are reported in [Table E15](#). We find no statistically significant discontinuities for slope, temperature, precipitation, or accessibility. We find moderate discontinuities for population, for which the estimate is only significant at the 10% level, and for elevation. The former might suggest migration as a relevant type of

adaptation, while the latter is expected due to the nature of our basin-level dataset. Coupled with the qualitatively unchanged estimates when accounting for these covariates (in various ways), this validation exercise soothes concerns of confounding at the mine basin.

We investigate further potential threats to identification related to differential industrial development and land use changes in upstream and downstream basins in [Figure D10](#). There, we consider nighttime lights (by Elvidge, Baugh, et al., [2017](#); Elvidge, Zhizhin, et al., [2021](#)) to proxy local industrial development and income, and forest loss (by Hansen et al., [2013](#)) as the most prominent land use change. There is a sharp increase in nighttime lights at the mine basin, likely stemming from mine operations and related infrastructure, but no statistically significant differences for up- and downstream basins. Similarly, we see higher forest loss rates within mine basins, consistent with land clearing for the mine and related infrastructure, but no significant effects for either up- or downstream basins.

## D. Additional Figures

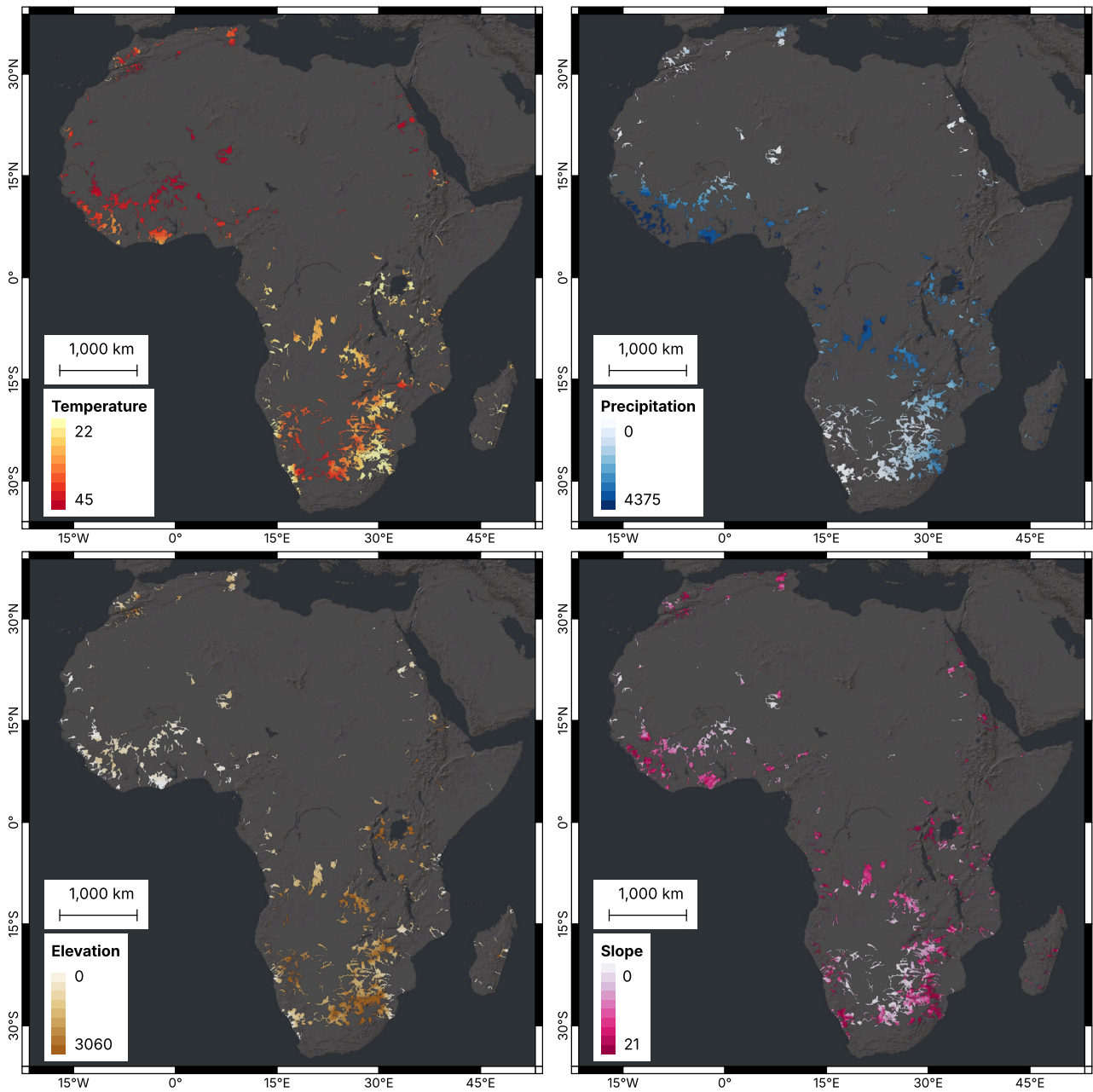


FIGURE D5: Maximum 2023 monthly temperature per basin in degrees Celsius (top-left), 2023 accumulated precipitation per basin in millimeters (top-right), average elevation in meters per basin (bottom-left), and average slope in degrees per basin (bottom-right). Basemap imagery provided by Esri, Maxar, Earthstar Geographics, and the GIS User Community.

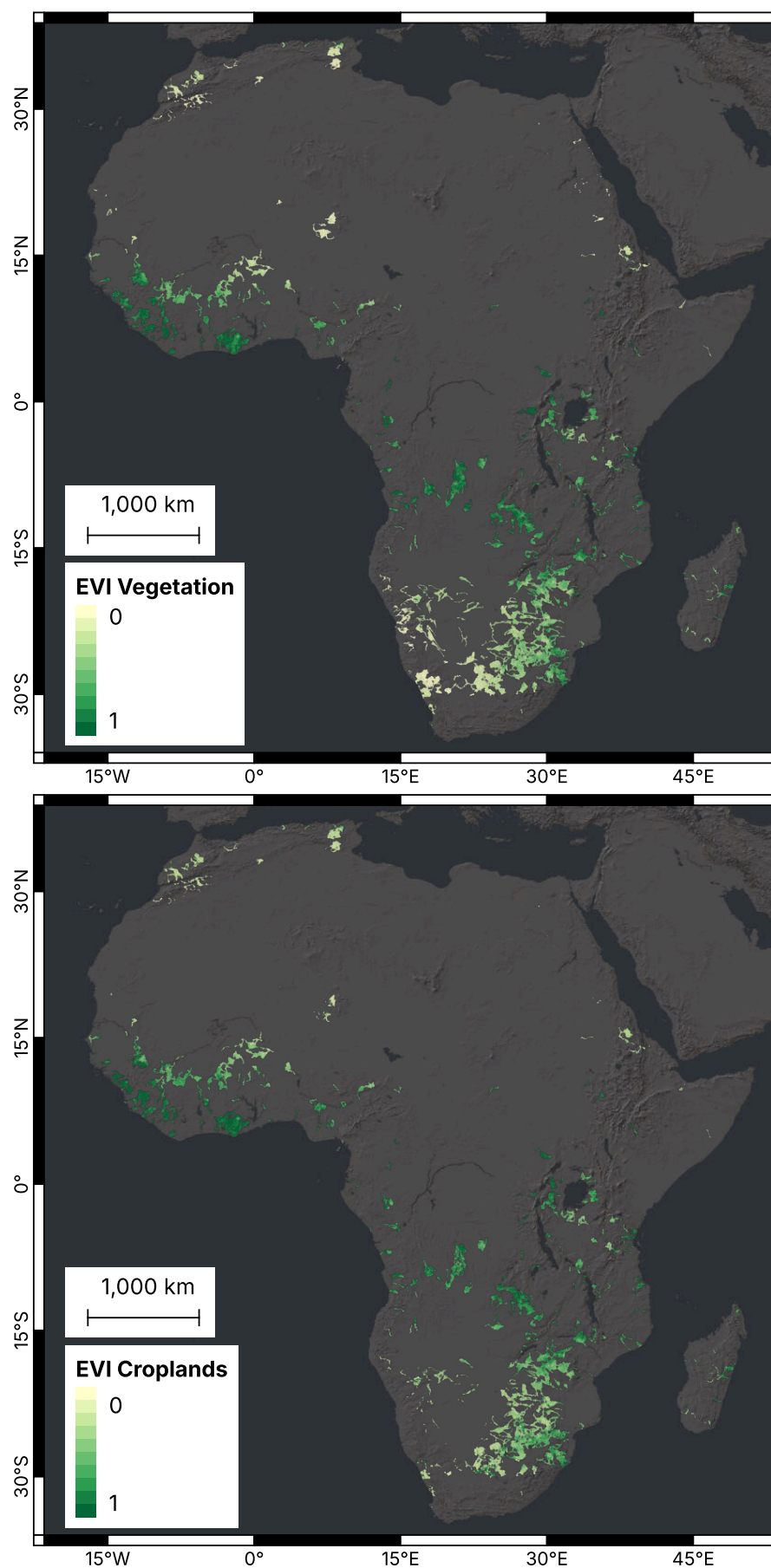


FIGURE D6: Maximum 2023 EVI per basin after applying the CCI vegetation mask (left) and the CCI cropland mask (right). Basemap imagery provided by Esri, Maxar, Earthstar Geographics, and the GIS User Community.

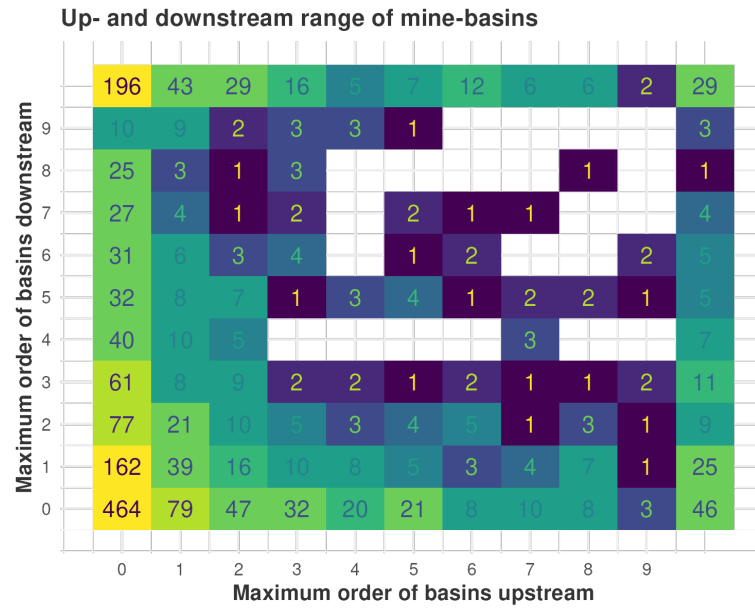


FIGURE D7: Number of mine-basin systems that have a maximum order of upstream basins (on the  $x$ -axis) and downstream basins (on the  $y$ -axis) in the dataset. Numbers add up to total number of mine-basin systems (1,900).

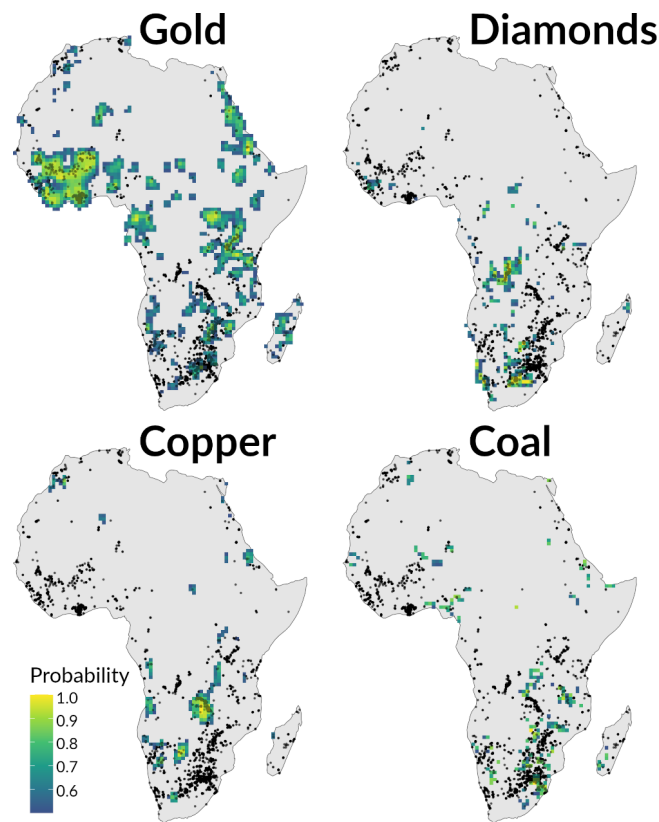


FIGURE D8: Heatmap indicating the results of the commodity prediction and point locations of training data. Note that the zoomed-out level of the heatmap averages over local nuances and does not accurately convey results for individual mines.

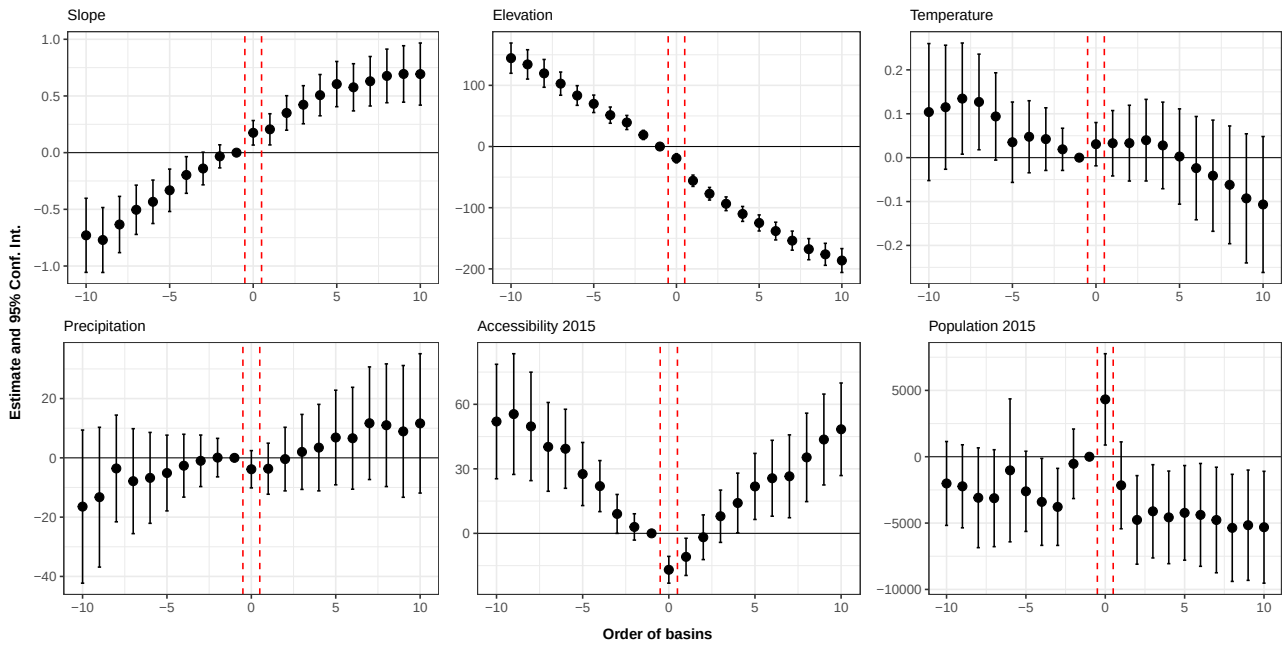


FIGURE D9: Order estimates when using elevation, slope, temperature, precipitation, accessibility to cities, and population as placebo outcomes.

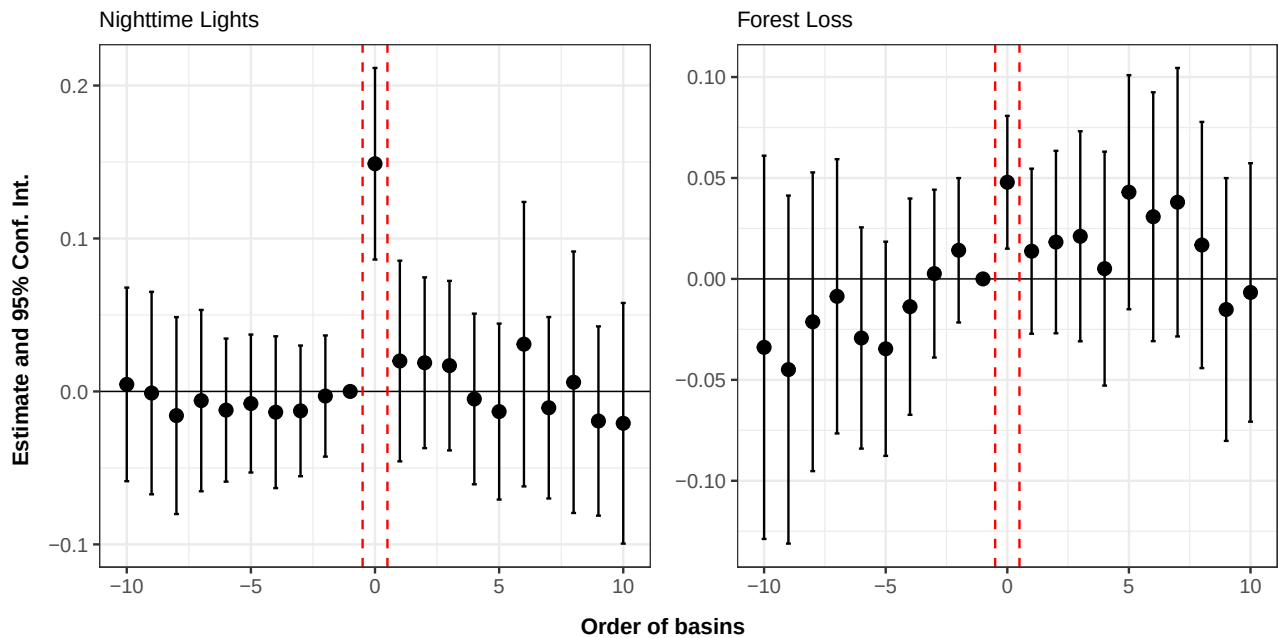


FIGURE D10: Main order estimates with covariates when using nighttime lights and deforestation as main outcome. Nighttime lights are sourced from (Elvidge, Baugh, et al., 2017; Elvidge, Zhizhin, et al., 2021) and deforestation (i.e., forest loss per square meter) (Hansen et al., 2013). Nighttime lights represent average pixel intensity per basin.

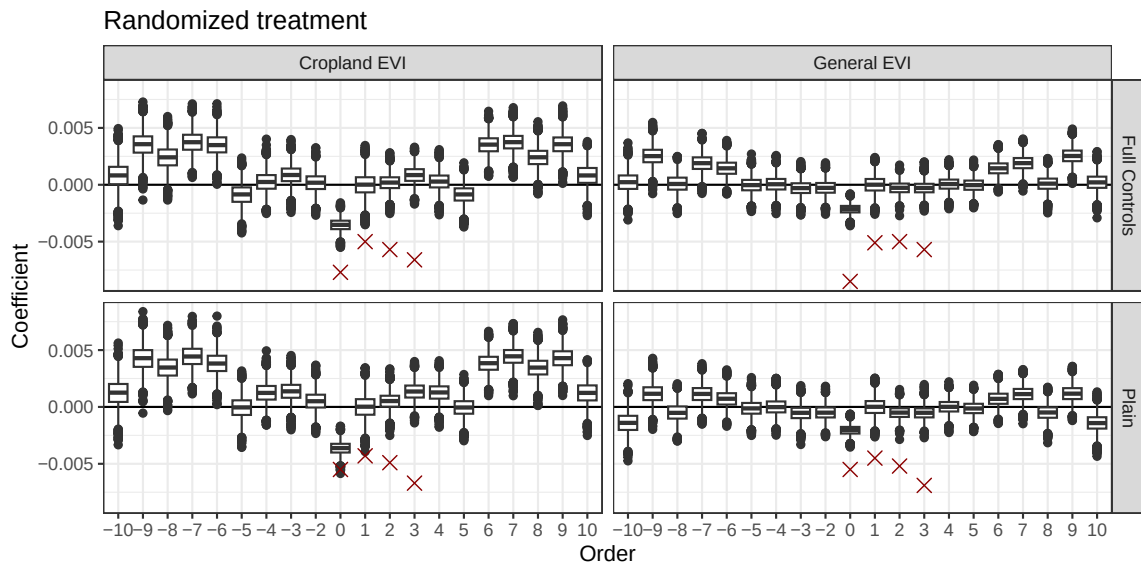


FIGURE D11: Estimation results when the treatment status (i.e., whether basins are down- or upstream) is randomized (5,000 runs, balance between statuses is kept). The red crosses indicate estimation results for the main specification.

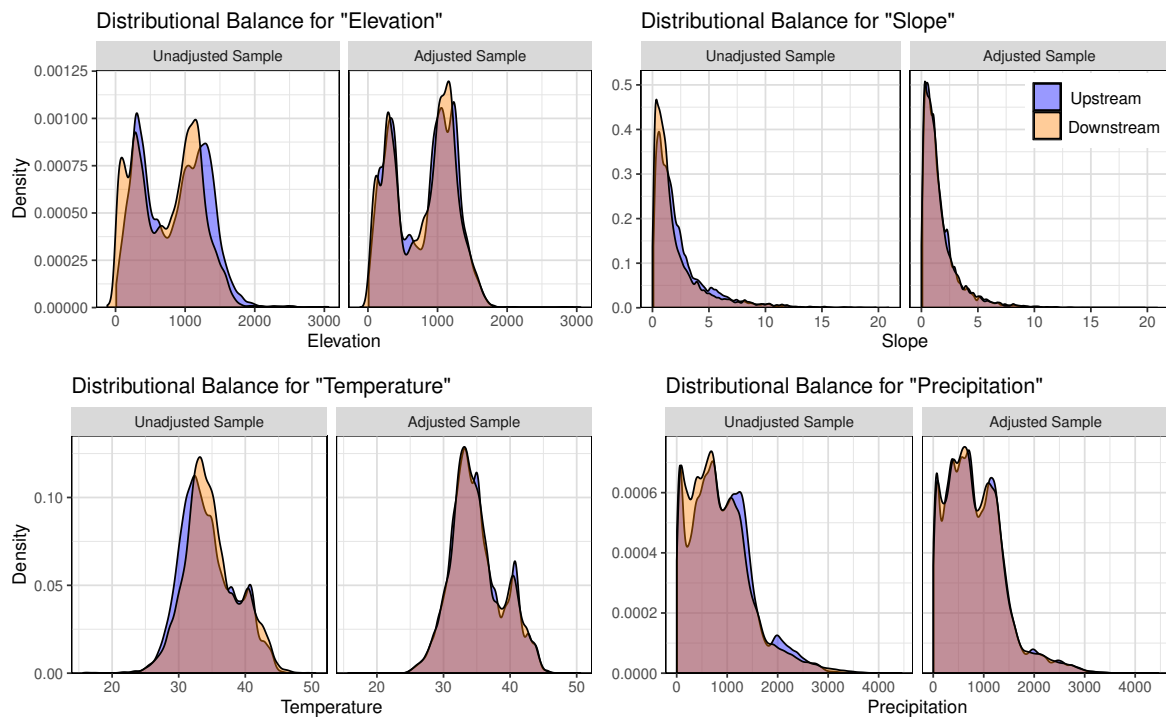


FIGURE D12: Balance of elevation, slope, temperature, and precipitation before and after matching. (Soilgrids not pictured.)

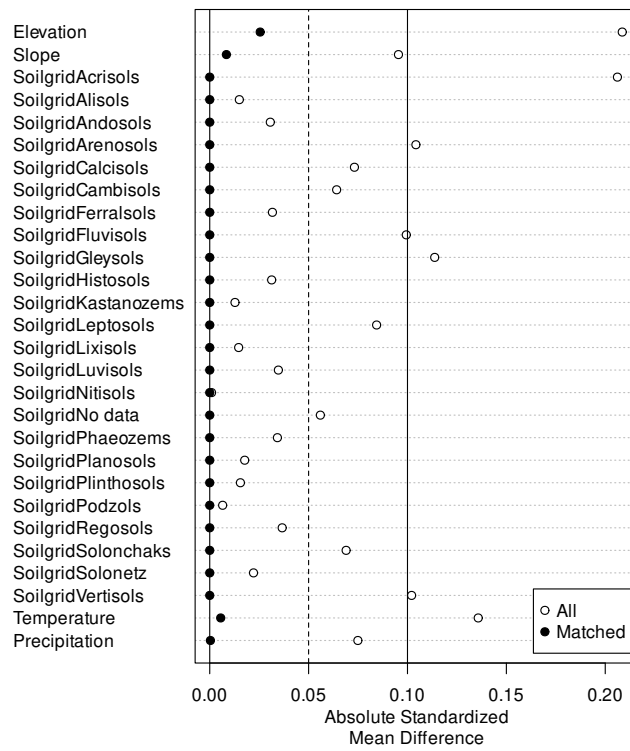


FIGURE D13: Standardized mean difference before and after matching.

## E. Additional Tables

TABLE E3: Number and average distance (km) of basins by order.

| Order | Upstream |          | Downstream |          |
|-------|----------|----------|------------|----------|
|       | <i>N</i> | Distance | <i>N</i>   | Distance |
| 0     | (1900)   | (0.0)    | (1900)     | (0.0)    |
| 1     | 847      | 13.9     | 1162       | 11.1     |
| 2     | 781      | 24.5     | 882        | 22.0     |
| 3     | 722      | 35.0     | 743        | 32.7     |
| 4     | 698      | 44.9     | 643        | 43.3     |
| 5     | 653      | 55.3     | 578        | 54.0     |
| 6     | 576      | 66.3     | 512        | 64.3     |
| 7     | 562      | 75.8     | 458        | 74.1     |
| 8     | 522      | 86.5     | 416        | 84.4     |
| 9     | 494      | 95.8     | 382        | 95.0     |
| 10    | 452      | 104.2    | 351        | 104.7    |

TABLE E4: Summary statistics split by status.

| Upstream Basins     |                     |        |            |            |          |               |
|---------------------|---------------------|--------|------------|------------|----------|---------------|
| Variable            | Unit of Measurement | NT     | Mean       | St. Dev.   | Min.     | Max.          |
| Max. Vegetation EVI | Index [−1, 1]       | 48,666 | 0.433      | 0.156      | 0.028    | 0.983         |
| Mean Vegetation EVI | Index [−1, 1]       | 48,666 | 0.285      | 0.114      | 0.016    | 0.578         |
| Max. Cropland EVI   | Index [−1, 1]       | 41,820 | 0.465      | 0.136      | 0.070    | 0.978         |
| Mean Cropland EVI   | Index [−1, 1]       | 41,820 | 0.300      | 0.104      | 0.030    | 0.601         |
| Elevation           | Meters              | 48,666 | 862.441    | 472.216    | 10.526   | 3,059.727     |
| Slope               | Degrees             | 48,666 | 2.321      | 2.272      | 0.086    | 20.913        |
| Max. Temperature    | Degree Celsius      | 48,666 | 34.194     | 3.958      | 15.633   | 46.146        |
| Precipitation       | Millimeter          | 48,666 | 923.300    | 588.526    | 5.744    | 3,625.230     |
| Population          | Unit                | 48,666 | 7,018.943  | 29,066.720 | 0.000    | 1,396,921.000 |
| Accessibility       | Minutes             | 48,666 | 176.564    | 194.345    | 3.474    | 2,197.584     |
| Mine Basins         |                     |        |            |            |          |               |
| Variable            | Unit of Measurement | NT     | Mean       | St. Dev.   | Min.     | Max.          |
| Max. Vegetation EVI | Index [−1, 1]       | 14,730 | 0.431      | 0.151      | 0.039    | 0.917         |
| Mean Vegetation EVI | Index [−1, 1]       | 14,730 | 0.281      | 0.112      | 0.034    | 0.563         |
| Max. Cropland EVI   | Index [−1, 1]       | 13,089 | 0.468      | 0.129      | 0.072    | 0.917         |
| Mean Cropland EVI   | Index [−1, 1]       | 13,089 | 0.299      | 0.103      | 0.059    | 0.568         |
| Elevation           | Meters              | 14,730 | 873.544    | 527.752    | 10.217   | 3,047.148     |
| Slope               | Degrees             | 14,730 | 2.325      | 2.195      | 0.018    | 20.456        |
| Max. Temperature    | Degree Celsius      | 14,726 | 33.357     | 3.904      | 15.592   | 46.525        |
| Precipitation       | Millimeter          | 14,730 | 919.593    | 594.956    | 0.640    | 4,204.642     |
| Population          | Unit                | 14,730 | 21,797.630 | 78,642.490 | 0.000    | 1,244,492.000 |
| Accessibility       | Minutes             | 14,706 | 117.981    | 124.096    | 1.739    | 1,271.511     |
| Downstream Basins   |                     |        |            |            |          |               |
| Variable            | Unit of Measurement | NT     | Mean       | St. Dev.   | Min.     | Max.          |
| Max. Vegetation EVI | Index [−1, 1]       | 47,180 | 0.422      | 0.153      | 0.016    | 0.993         |
| Mean Vegetation EVI | Index [−1, 1]       | 47,180 | 0.273      | 0.108      | −0.021   | 0.559         |
| Max. Cropland EVI   | Index [−1, 1]       | 38,127 | 0.461      | 0.130      | −0.068   | 0.958         |
| Mean Cropland EVI   | Index [−1, 1]       | 38,127 | 0.294      | 0.098      | −0.104   | 0.597         |
| Elevation           | Meters              | 47,172 | 760.527    | 468.287    | −118.349 | 2,949.539     |
| Slope               | Degrees             | 47,172 | 2.123      | 2.446      | 0.000    | 19.798        |
| Max. Temperature    | Degree Celsius      | 47,180 | 34.709     | 3.875      | 16.590   | 48.845        |
| Precipitation       | Millimeter          | 47,180 | 874.144    | 601.039    | 3.149    | 4,456.690     |
| Population          | Unit                | 47,180 | 5,808.941  | 21,609.500 | 0.000    | 667,053.000   |
| Accessibility       | Minutes             | 47,156 | 166.076    | 174.843    | 1.002    | 2,659.925     |

## E1. Main results

TABLE E5: Main estimation results, order specification.

| Dependent Variables:<br>Model: | Maximum Vegetation EVI |                        |                        |                        | Maximum Croplands EVI  |                        |                        |                        |
|--------------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|                                | (1)                    | (2)                    | (3)                    | (4)                    | (5)                    | (6)                    | (7)                    | (8)                    |
| Mine Basin                     | -0.0049***<br>(0.0014) | -0.0053***<br>(0.0014) | -0.0048***<br>(0.0014) | -0.0046***<br>(0.0014) | -0.0068***<br>(0.0022) | -0.0073***<br>(0.0022) | -0.0070***<br>(0.0022) | -0.0064***<br>(0.0022) |
| Downstream, 1                  | -0.0045***<br>(0.0017) | -0.0047***<br>(0.0018) | -0.0041**<br>(0.0018)  | -0.0043**<br>(0.0018)  | -0.0051**<br>(0.0025)  | -0.0051**<br>(0.0025)  | -0.0049**<br>(0.0025)  | -0.0050**<br>(0.0025)  |
| Downstream, 2                  | -0.0049**<br>(0.0022)  | -0.0048**<br>(0.0024)  | -0.0045*<br>(0.0024)   | -0.0048**<br>(0.0024)  | -0.0058*<br>(0.0031)   | -0.0061*<br>(0.0032)   | -0.0064*<br>(0.0032)   | -0.0067**<br>(0.0032)  |
| Downstream, 3                  | -0.0085***<br>(0.0028) | -0.0086***<br>(0.0029) | -0.0087***<br>(0.0029) | -0.0087***<br>(0.0029) | -0.0088**<br>(0.0037)  | -0.0092**<br>(0.0038)  | -0.0098**<br>(0.0038)  | -0.0099***<br>(0.0038) |
| Downstream, 4                  | -0.0049*<br>(0.0030)   | -0.0057*<br>(0.0032)   | -0.0061*<br>(0.0033)   | -0.0062*<br>(0.0033)   | -0.0029<br>(0.0038)    | -0.0034<br>(0.0039)    | -0.0042<br>(0.0039)    | -0.0044<br>(0.0040)    |
| Downstream, 5                  | -0.0034<br>(0.0033)    | -0.0043<br>(0.0036)    | -0.0053<br>(0.0037)    | -0.0053<br>(0.0037)    | 0.0007<br>(0.0042)     | 0.0003<br>(0.0044)     | -0.0015<br>(0.0045)    | -0.0016<br>(0.0045)    |
| Downstream, 6                  | -0.0027<br>(0.0034)    | -0.0040<br>(0.0039)    | -0.0057<br>(0.0040)    | -0.0061<br>(0.0040)    | -0.0004<br>(0.0046)    | -0.0010<br>(0.0051)    | -0.0035<br>(0.0052)    | -0.0038<br>(0.0052)    |
| Downstream, 7                  | -0.0053<br>(0.0037)    | -0.0061<br>(0.0042)    | -0.0087**<br>(0.0043)  | -0.0089**<br>(0.0043)  | -0.0051<br>(0.0048)    | -0.0057<br>(0.0054)    | -0.0092*<br>(0.0055)   | -0.0093*<br>(0.0055)   |
| Downstream, 8                  | -0.0095**<br>(0.0041)  | -0.0115**<br>(0.0045)  | -0.0141***<br>(0.0046) | -0.0144***<br>(0.0046) | -0.0047<br>(0.0053)    | -0.0056<br>(0.0059)    | -0.0085<br>(0.0060)    | -0.0088<br>(0.0060)    |
| Downstream, 9                  | -0.0066<br>(0.0045)    | -0.0089*<br>(0.0049)   | -0.0120**<br>(0.0050)  | -0.0123**<br>(0.0050)  | -0.0060<br>(0.0056)    | -0.0070<br>(0.0063)    | -0.0108*<br>(0.0064)   | -0.0111*<br>(0.0064)   |
| Downstream, 10                 | -0.0063<br>(0.0049)    | -0.0083<br>(0.0053)    | -0.0118**<br>(0.0054)  | -0.0120**<br>(0.0054)  | -0.0023<br>(0.0060)    | -0.0030<br>(0.0068)    | -0.0071<br>(0.0069)    | -0.0074<br>(0.0069)    |
| Elevation                      |                        | -0.0014**<br>(0.0006)  | -0.0050***<br>(0.0007) | -0.0050***<br>(0.0007) |                        | -0.0010<br>(0.0008)    | -0.0049***<br>(0.0008) | -0.0049***<br>(0.0008) |
| Slope                          |                        | 0.0029***<br>(0.0005)  | 0.0024***<br>(0.0005)  | 0.0025***<br>(0.0005)  |                        | 0.0027***<br>(0.0006)  | 0.0023***<br>(0.0007)  | 0.0022***<br>(0.0006)  |
| Yearly Max. Temperature        |                        |                        | -0.0064***<br>(0.0006) | -0.0064***<br>(0.0006) |                        |                        | -0.0066***<br>(0.0007) | -0.0067***<br>(0.0007) |
| Yearly Precipitation           |                        |                        | 0.0037***<br>(0.0003)  | 0.0037***<br>(0.0003)  |                        |                        | 0.0033***<br>(0.0003)  | 0.0033***<br>(0.0003)  |
| Accessibility in 2015          |                        |                        |                        | -0.0014**<br>(0.0007)  |                        |                        |                        | 0.0011<br>(0.0018)     |
| Population in 2015             |                        |                        |                        | -0.0078***<br>(0.0019) |                        |                        |                        | -0.0069***<br>(0.0023) |
| Soil Type included             |                        | Yes                    | Yes                    | Yes                    |                        | Yes                    | Yes                    | Yes                    |
| Sample Mean (Order 1)          | 0.4259                 | 0.4259                 | 0.4259                 | 0.4259                 | 0.4647                 | 0.4647                 | 0.4647                 | 0.4647                 |
| Relative Effect (Order 1)      | -1.056                 | -1.098                 | -0.9731                | -1.008                 | -1.093                 | -1.098                 | -1.063                 | -1.066                 |
| <i>Fixed-effects</i>           |                        |                        |                        |                        |                        |                        |                        |                        |
| Year                           | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    |
| Mine                           | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    |
| <i>Fit statistics</i>          |                        |                        |                        |                        |                        |                        |                        |                        |
| Observations                   | 110,576                | 110,568                | 110,564                | 110,524                | 93,036                 | 93,036                 | 93,032                 | 93,000                 |
| Number of upstream basins      | 6,084                  | 6,084                  | 6,084                  | 6,084                  | 5,232                  | 5,232                  | 5,232                  | 5,232                  |
| Number of downstream basins    | 5,899                  | 5,898                  | 5,898                  | 5,896                  | 4,769                  | 4,769                  | 4,769                  | 4,767                  |
| R <sup>2</sup>                 | 0.90287                | 0.90457                | 0.90767                | 0.90788                | 0.81624                | 0.81763                | 0.82152                | 0.82179                |
| Within R <sup>2</sup>          | 0.00175                | 0.01894                | 0.05094                | 0.05285                | 0.00216                | 0.00973                | 0.03121                | 0.03245                |

Clustered (mine basin) standard-errors in parentheses

Significance: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

**Note:** Table shows results for estimation of Equation (1), with distance included as measured by the ordering of basins with respect to the mine basin. Columns (1)–(4) hold results from models for the overall EVI as proxy measure for vegetative health within basins, columns (5)–(8) for the cropland-specific EVI as proxy measure for agricultural productivity. Models in columns (1) and (5) include no additional covariates, models (2) and (6) control for geophysical variables (elevation, slope, and soil), models (3) and (7) additionally control for meteorological (yearly sum of precipitation and yearly maximum temperature), and models (4) and (8) additionally control for socioeconomic (accessibility to city in minutes and total population in 2015) conditions. All models include mine and year fixed effects. Standard errors are clustered at the mine basin system level.

TABLE E6: Main estimation results, pooled order specification.

| Dependent Variables:<br>Model: | <i>Maximum Vegetation EVI</i> |                        |                        |                        | <i>Maximum Croplands EVI</i> |                        |                        |                        |
|--------------------------------|-------------------------------|------------------------|------------------------|------------------------|------------------------------|------------------------|------------------------|------------------------|
|                                | (1)                           | (2)                    | (3)                    | (4)                    | (5)                          | (6)                    | (7)                    | (8)                    |
| Mine Basin                     | -0.0053***<br>(0.0015)        | -0.0057***<br>(0.0015) | -0.0054***<br>(0.0015) | -0.0051***<br>(0.0015) | -0.0079***<br>(0.0022)       | -0.0084***<br>(0.0022) | -0.0083***<br>(0.0022) | -0.0076***<br>(0.0022) |
| Downstream, 1–3                | -0.0057***<br>(0.0018)        | -0.0057***<br>(0.0020) | -0.0054***<br>(0.0020) | -0.0056***<br>(0.0020) | -0.0064**<br>(0.0025)        | -0.0065**<br>(0.0026)  | -0.0067***<br>(0.0026) | -0.0068***<br>(0.0026) |
| Elevation                      |                               | -0.0013**<br>(0.0006)  | -0.0050***<br>(0.0007) | -0.0050***<br>(0.0007) |                              | -0.0009<br>(0.0008)    | -0.0049***<br>(0.0008) | -0.0049***<br>(0.0008) |
| Slope                          |                               | 0.0029***<br>(0.0005)  | 0.0024***<br>(0.0005)  | 0.0024***<br>(0.0005)  |                              | 0.0027***<br>(0.0006)  | 0.0023***<br>(0.0007)  | 0.0022***<br>(0.0006)  |
| Yearly Max. Temperature        |                               |                        | -0.0064***<br>(0.0006) | -0.0064***<br>(0.0006) |                              |                        | -0.0066***<br>(0.0007) | -0.0067***<br>(0.0007) |
| Yearly Precipitation           |                               |                        | 0.0037***<br>(0.0003)  | 0.0037***<br>(0.0003)  |                              |                        | 0.0033***<br>(0.0003)  | 0.0033***<br>(0.0003)  |
| Accessibility in 2015          |                               |                        |                        | -0.0015**<br>(0.0007)  |                              |                        |                        | 0.0011<br>(0.0018)     |
| Population in 2015             |                               |                        |                        | -0.0078***<br>(0.0019) |                              |                        |                        | -0.0069***<br>(0.0023) |
| Soil Type included             |                               | Yes                    | Yes                    | Yes                    |                              | Yes                    | Yes                    | Yes                    |
| Sample Mean (Order 1–3)        | 0.4234                        | 0.4234                 | 0.4234                 | 0.4234                 | 0.4623                       | 0.4623                 | 0.4623                 | 0.4623                 |
| Relative Effect (Order 1–3 )   | -1.348                        | -1.348                 | -1.283                 | -1.319                 | -1.378                       | -1.410                 | -1.447                 | -1.473                 |
| <i>Fixed-effects</i>           |                               |                        |                        |                        |                              |                        |                        |                        |
| Year                           | Yes                           | Yes                    | Yes                    | Yes                    | Yes                          | Yes                    | Yes                    | Yes                    |
| Mine                           | Yes                           | Yes                    | Yes                    | Yes                    | Yes                          | Yes                    | Yes                    | Yes                    |
| <i>Fit statistics</i>          |                               |                        |                        |                        |                              |                        |                        |                        |
| Observations                   | 110,576                       | 110,568                | 110,564                | 110,524                | 93,036                       | 93,036                 | 93,032                 | 93,000                 |
| Number of upstream basins      | 6,084                         | 6,084                  | 6,084                  | 6,084                  | 5,232                        | 5,232                  | 5,232                  | 5,232                  |
| Number of downstream basins    | 5,899                         | 5,898                  | 5,898                  | 5,896                  | 4,769                        | 4,769                  | 4,769                  | 4,767                  |
| R <sup>2</sup>                 | 0.90286                       | 0.90456                | 0.90765                | 0.90787                | 0.81622                      | 0.81761                | 0.82150                | 0.82177                |
| Within R <sup>2</sup>          | 0.00163                       | 0.01882                | 0.05080                | 0.05272                | 0.00208                      | 0.00964                | 0.03109                | 0.03234                |

Clustered (mine basin) standard-errors in parentheses

Significance: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

**Note:** Table shows results for estimation of Equation (1), with distance included as measured by the ordering of basins with respect to the mine basin and pooled for the first three basins immediately up- or downstream of the mine basin. Columns (1)–(4) hold results from models for the overall EVI as proxy measure for vegetative health within basins, columns (5)–(8) for the cropland-specific EVI as proxy measure for agricultural productivity. Models in columns (1) and (5) include no additional covariates, models (2) and (6) control for geophysical variables (elevation, slope, and soil), models (3) and (7) additionally control for meteorological (yearly sum of precipitation and yearly maximum temperature), and models (4) and (8) additionally control for socioeconomic (accessibility to city in minutes and total population in 2015) conditions. All models include mine and year fixed effects. Standard errors are clustered at the mine basin system level.

## E2. Heterogeneity analysis

TABLE E7: Estimation results for heterogeneity analysis: Spatial heterogeneity

| Biome                         |                        | Biome              |                       |                       | Region              |                     |                       | Crop Suitability       |                     |                     |
|-------------------------------|------------------------|--------------------|-----------------------|-----------------------|---------------------|---------------------|-----------------------|------------------------|---------------------|---------------------|
|                               |                        | Deserts            | Forests               | Grasslands            | N. & E. Africa      | S. Africa           | W. Africa             | High                   | Medium              | Low                 |
| Model:                        | (1)                    | (2)                | (3)                   | (4)                   | (5)                 | (6)                 | (7)                   | (8)                    | (9)                 | (10)                |
| <i>Maximum Vegetation EVI</i> |                        |                    |                       |                       |                     |                     |                       |                        |                     |                     |
| Downstream, 1–3               | -0.0056***<br>(0.0020) | 0.0020<br>(0.0041) | -0.0105**<br>(0.0048) | -0.0055**<br>(0.0023) | -0.0059<br>(0.0061) | -0.0020<br>(0.0025) | -0.0097**<br>(0.0039) | -0.0066**<br>(0.0028)  | -0.0044<br>(0.0030) | -0.0022<br>(0.0043) |
| <i>Fit statistics</i>         |                        |                    |                       |                       |                     |                     |                       |                        |                     |                     |
| Observations                  | 110,524                | 16,988             | 16,838                | 76,698                | 10,104              | 71,481              | 28,939                | 39,232                 | 47,088              | 24,204              |
| Number of upstream basins     | 6,084                  | 903                | 878                   | 4,303                 | 560                 | 3,840               | 1,684                 | 2,099                  | 2,770               | 1,215               |
| Number of downstream basins   | 5,896                  | 950                | 872                   | 4,074                 | 565                 | 3,844               | 1,487                 | 2,111                  | 2,345               | 1,440               |
| R <sup>2</sup>                | 0.90787                | 0.85191            | 0.92357               | 0.82853               | 0.91358             | 0.89851             | 0.89497               | 0.77085                | 0.84687             | 0.87793             |
| Within R <sup>2</sup>         | 0.05272                | 0.09738            | 0.07345               | 0.05223               | 0.09214             | 0.05905             | 0.03862               | 0.03421                | 0.06666             | 0.09580             |
| <i>Maximum Cropland EVI</i>   |                        |                    |                       |                       |                     |                     |                       |                        |                     |                     |
| Downstream, 1–3               | -0.0068***<br>(0.0026) | 0.0159<br>(0.0133) | -0.0088**<br>(0.0044) | -0.0072**<br>(0.0029) | -0.0038<br>(0.0064) | -0.0014<br>(0.0035) | -0.0115**<br>(0.0046) | -0.0088***<br>(0.0033) | -0.0046<br>(0.0038) | 0.0015<br>(0.0119)  |
| <i>Fit statistics</i>         |                        |                    |                       |                       |                     |                     |                       |                        |                     |                     |
| Observations                  | 93,000                 | 7,856              | 15,885                | 69,259                | 9,028               | 56,946              | 27,026                | 36,611                 | 43,628              | 12,761              |
| Number of upstream basins     | 5,232                  | 449                | 830                   | 3,953                 | 519                 | 3,146               | 1,567                 | 1,980                  | 2,580               | 672                 |
| Number of downstream basins   | 4,767                  | 392                | 807                   | 3,568                 | 486                 | 2,902               | 1,379                 | 1,924                  | 2,134               | 709                 |
| R <sup>2</sup>                | 0.82177                | 0.70207            | 0.91138               | 0.73855               | 0.89173             | 0.75552             | 0.82927               | 0.72121                | 0.80964             | 0.79248             |
| Within R <sup>2</sup>         | 0.03234                | 0.08679            | 0.05972               | 0.03288               | 0.07170             | 0.03574             | 0.02975               | 0.02746                | 0.04722             | 0.05867             |
| <i>Fixed-effects</i>          |                        |                    |                       |                       |                     |                     |                       |                        |                     |                     |
| Year                          | Yes                    | Yes                | Yes                   | Yes                   | Yes                 | Yes                 | Yes                   | Yes                    | Yes                 | Yes                 |
| Mine                          | Yes                    | Yes                | Yes                   | Yes                   | Yes                 | Yes                 | Yes                   | Yes                    | Yes                 | Yes                 |

Clustered (mine basin) standard-errors in parentheses

Significance: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

**Note:** Table shows results for estimation of Equation (1), with distance included as measured by the ordering of basins with respect to the mine basin, with the overall EVI as outcome in the upper panel and the cropland-specific EVI as outcome in the lower panel. Model in column (1) reports results for the full sample, models in columns (2)–(4) for sample splits by primary biome of mine basin system, and models in columns (5)–(7) for sample splits by regions based on the USDA crop classifications, models in columns (8)–(11) for sample splits by crop suitability based on the GAEZ methodology. All specifications include the full set of controls and mine and year fixed effects. Standard errors are clustered at the mine basin system level.

TABLE E8: Estimation results for heterogeneity analysis: Mine characteristics

| Model:                        |                        | Mine Size              |                        |                       | Mine Growth            |                       |                       | Commodity Growth    |                     |                      |                        |
|-------------------------------|------------------------|------------------------|------------------------|-----------------------|------------------------|-----------------------|-----------------------|---------------------|---------------------|----------------------|------------------------|
|                               |                        | > 0.5km <sup>2</sup>   | > 1km <sup>2</sup>     | > 2.5km <sup>2</sup>  | > 0%                   | > 10%                 | > 25%                 | Coal                | Copper              | Diamonds             | Gold                   |
| Model:                        | (1)                    | (2)                    | (3)                    | (4)                   | (5)                    | (6)                   | (7)                   | (8)                 | (9)                 | (10)                 | (11)                   |
| <i>Maximum Vegetation EVI</i> |                        |                        |                        |                       |                        |                       |                       |                     |                     |                      |                        |
| Downstream, 1–3               | -0.0056***<br>(0.0020) | -0.0074***<br>(0.0027) | -0.0081***<br>(0.0029) | -0.0084**<br>(0.0034) | -0.0051**<br>(0.0022)  | -0.0048**<br>(0.0023) | -0.0040<br>(0.0024)   | -0.0052<br>(0.0060) | -0.0078<br>(0.0055) | -0.0074*<br>(0.0045) | -0.0080***<br>(0.0030) |
| <i>Fit statistics</i>         |                        |                        |                        |                       |                        |                       |                       |                     |                     |                      |                        |
| Observations                  | 110,524                | 55,627                 | 43,787                 | 28,265                | 74,180                 | 69,644                | 60,396                | 9,096               | 6,496               | 14,596               | 29,087                 |
| Number of upstream basins     | 6,084                  | 2,850                  | 2,220                  | 1,438                 | 4,248                  | 4,031                 | 3,365                 | 430                 | 283                 | 935                  | 1,584                  |
| Number of downstream basins   | 5,896                  | 2,966                  | 2,326                  | 1,467                 | 3,988                  | 3,757                 | 3,417                 | 508                 | 394                 | 639                  | 1,555                  |
| R <sup>2</sup>                | 0.90787                | 0.91446                | 0.91331                | 0.91714               | 0.89661                | 0.89568               | 0.89697               | 0.77295             | 0.90098             | 0.93217              | 0.87340                |
| Within R <sup>2</sup>         | 0.05272                | 0.05403                | 0.05783                | 0.06594               | 0.05456                | 0.05365               | 0.04939               | 0.10382             | 0.09247             | 0.11833              | 0.03965                |
| <i>Maximum Cropland EVI</i>   |                        |                        |                        |                       |                        |                       |                       |                     |                     |                      |                        |
| Downstream, 1–3               | -0.0068***<br>(0.0026) | -0.0052<br>(0.0039)    | -0.0054<br>(0.0044)    | -0.0045<br>(0.0055)   | -0.0074***<br>(0.0027) | -0.0059**<br>(0.0027) | -0.0066**<br>(0.0030) | 0.0095<br>(0.0091)  | 0.0015<br>(0.0080)  | -0.0127<br>(0.0090)  | -0.0119***<br>(0.0042) |
| <i>Fit statistics</i>         |                        |                        |                        |                       |                        |                       |                       |                     |                     |                      |                        |
| Observations                  | 93,000                 | 48,325                 | 38,247                 | 24,391                | 63,329                 | 59,266                | 51,873                | 8,252               | 5,138               | 11,081               | 27,375                 |
| Number of upstream basins     | 5,232                  | 2,547                  | 1,979                  | 1,275                 | 3,738                  | 3,530                 | 2,973                 | 400                 | 243                 | 732                  | 1,498                  |
| Number of downstream basins   | 4,767                  | 2,464                  | 1,955                  | 1,199                 | 3,236                  | 3,049                 | 2,816                 | 447                 | 280                 | 461                  | 1,440                  |
| R <sup>2</sup>                | 0.82177                | 0.84026                | 0.84695                | 0.84598               | 0.79461                | 0.79109               | 0.79097               | 0.67233             | 0.83484             | 0.76409              | 0.82100                |
| Within R <sup>2</sup>         | 0.03234                | 0.03431                | 0.03966                | 0.05226               | 0.03433                | 0.03215               | 0.03298               | 0.07971             | 0.07025             | 0.04274              | 0.04576                |
| <i>Fixed-effects</i>          |                        |                        |                        |                       |                        |                       |                       |                     |                     |                      |                        |
| Year                          | Yes                    | Yes                    | Yes                    | Yes                   | Yes                    | Yes                   | Yes                   | Yes                 | Yes                 | Yes                  | Yes                    |
| Mine                          | Yes                    | Yes                    | Yes                    | Yes                   | Yes                    | Yes                   | Yes                   | Yes                 | Yes                 | Yes                  | Yes                    |

Clustered (mine basin) standard-errors in parentheses

Significance: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

**Note:** Table shows results for estimation of Equation (1), with distance included as measured by the ordering of basins with respect to the mine basin, with the overall EVI as outcome in the upper panel and the cropland-specific EVI as outcome in the lower panel. Model in column (1) reports results for the baseline specification, models in columns (2)–(4) for subsets of mine basins with increasing total area of mined area, models in columns (5)–(7) for subsets of mine basins with increasing growth in mined area in the period from 2017 to 2023 based on Sepin, Vashold, and Kuschni, 2025, models (8)–(11) report results for subsets of mines split by the primary exclusive commodity mined within them. All specifications include the full set of controls and mine and year fixed effects. Standard errors are clustered at the mine basin system level.

## E3. Robustness analysis

TABLE E9: Estimation results for alternative/additional controls

| Dependent Variables:<br>Model: | Maximum Vegetation EVI |                        |                        |                        |                        | Maximum Cropland EVI   |                        |                       |                        |                       |
|--------------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|-----------------------|------------------------|-----------------------|
|                                | (1)                    | (2)                    | (3)                    | (4)                    | (5)                    | (6)                    | (7)                    | (8)                   | (9)                    | (10)                  |
| <i>Variables</i>               |                        |                        |                        |                        |                        |                        |                        |                       |                        |                       |
| Downstream, 1–3                | -0.0056***<br>(0.0020) | -0.0061***<br>(0.0020) | -0.0056***<br>(0.0020) | -0.0056***<br>(0.0020) | -0.0053***<br>(0.0020) | -0.0068***<br>(0.0026) | -0.0073***<br>(0.0026) | -0.0066**<br>(0.0026) | -0.0067***<br>(0.0026) | -0.0065**<br>(0.0026) |
| <i>Fixed-effects</i>           |                        |                        |                        |                        |                        |                        |                        |                       |                        |                       |
| Year                           | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    | Yes                   | Yes                    | Yes                   |
| Mine                           | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    | Yes                   | Yes                    | Yes                   |
| <i>Fit statistics</i>          |                        |                        |                        |                        |                        |                        |                        |                       |                        |                       |
| Observations                   | 110,524                | 110,528                | 96,702                 | 110,524                | 110,524                | 93,000                 | 93,004                 | 81,324                | 93,000                 | 93,000                |
| Number of upstream basins      | 6,084                  | 6,084                  | 6,084                  | 6,084                  | 6,084                  | 5,232                  | 5,232                  | 5,232                 | 5,232                  | 5,232                 |
| Number of downstream basins    | 5,896                  | 5,896                  | 5,896                  | 5,896                  | 5,896                  | 4,767                  | 4,767                  | 4,767                 | 4,767                  | 4,767                 |
| R <sup>2</sup>                 | 0.90787                | 0.90670                | 0.90691                | 0.90787                | 0.90806                | 0.82177                | 0.82060                | 0.82156               | 0.82182                | 0.82187               |
| Within R <sup>2</sup>          | 0.05272                | 0.04056                | 0.05345                | 0.05272                | 0.05476                | 0.03234                | 0.02565                | 0.03443               | 0.03260                | 0.03287               |

Clustered (mine basin) standard-errors in parentheses

Significance: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

**Note:** Table shows results for estimation of Equation (1), with distance included as measured by the ordering of basins with respect to the mine basin and pooled for the first three basins immediately up- or downstream of the mine basin. Columns (1)–(5) hold results from models for the vegetation EVI as proxy measure for vegetative health within basins, columns (6)–(10) for the cropland-specific EVI as proxy measure for agricultural productivity. Models in columns (1) and (6) are the baseline specification, models in columns (2) and (7) use alternative meteorological variables from the Climatic Research Unit Harris et al., 2020 for precipitation and maximum temperature, models in columns (3) and (8) include an additional control for yearly average concentrations of particulate matter with a diameter of 2.5 $\mu$ g within basins taken from Shen et al., 2024, models in columns (4) and (9) include an additional control for violent events within basins taken from Raleigh et al. (2010), models in columns (5) and (10) include an additional control for distance to coast. All specifications include the full set of controls and mine and year fixed effects. Standard errors are clustered at the mine basin system level.

TABLE E10: Estimation results for varying outcome variables

| Dependent Variables:<br>Model: | Vegetation             |                       |                       |                       |                       | Croplands              |                       |                       |                       |                       |
|--------------------------------|------------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
|                                | (1)                    | (2)                   | (3)                   | (4)                   | (5)                   | (6)                    | (7)                   | (8)                   | (9)                   | (10)                  |
| <i>Variables</i>               |                        |                       |                       |                       |                       |                        |                       |                       |                       |                       |
| Downstream, 1–3                | -0.0056***<br>(0.0020) | -0.0057**<br>(0.0023) | -0.0048**<br>(0.0020) | -0.0046**<br>(0.0021) | -0.0032**<br>(0.0013) | -0.0068***<br>(0.0026) | -0.0058**<br>(0.0028) | -0.0053**<br>(0.0027) | -0.0058**<br>(0.0028) | -0.0037**<br>(0.0016) |
| <i>Fixed-effects</i>           |                        |                       |                       |                       |                       |                        |                       |                       |                       |                       |
| Year                           | Yes                    | Yes                   | Yes                   | Yes                   | Yes                   | Yes                    | Yes                   | Yes                   | Yes                   | Yes                   |
| Mine                           | Yes                    | Yes                   | Yes                   | Yes                   | Yes                   | Yes                    | Yes                   | Yes                   | Yes                   | Yes                   |
| <i>Fit statistics</i>          |                        |                       |                       |                       |                       |                        |                       |                       |                       |                       |
| Observations                   | 110,524                | 110,524               | 109,505               | 110,500               | 110,524               | 93,000                 | 93,000                | 94,596                | 92,719                | 93,000                |
| Number of upstream basins      | 6,084                  | 6,084                 | 6,039                 | 6,083                 | 6,084                 | 5,232                  | 5,232                 | 5,201                 | 5,216                 | 5,232                 |
| Number of downstream basins    | 5,896                  | 5,896                 | 5,825                 | 5,894                 | 5,896                 | 4,767                  | 4,767                 | 4,959                 | 4,750                 | 4,767                 |
| R <sup>2</sup>                 | 0.90787                | 0.93543               | 0.90344               | 0.91885               | 0.95522               | 0.82177                | 0.86923               | 0.78699               | 0.82381               | 0.91360               |
| Within R <sup>2</sup>          | 0.05272                | 0.08588               | 0.05400               | 0.06762               | 0.14092               | 0.03234                | 0.05135               | 0.03110               | 0.03510               | 0.08378               |

Clustered (mine basin) standard-errors in parentheses

Significance: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

**Note:** Table shows results for estimation of Equation (1), with distance included as measured by the ordering of basins with respect to the mine basin and pooled for the first three basins immediately up- or downstream of the mine basin. Columns (1)–(5) hold results from models for vegetation-specific measures as proxy measure for vegetative health within basins, columns (6)–(10) for cropland-specific ones as proxy measure for agricultural productivity. Models in columns (1) and (5) are the baseline specification for the overall maximum EVI and the cropland-specific maximum EVI, respectively. Models in columns (2) and (7) use the respective NDVI instead of the EVI, models in columns (3) and (8) report results for a narrower version of the vegetation mask by ESA (Defourny et al., 2024) and a cropland mask by Digital Earth Africa, 2022, respectively. Models in columns (4) and (9) use the average of the pixel-specific annual maximum EVI per basin, models in columns (5) and (10) use the yearly mean of the EVI as outcome instead of the maximum. All specifications include the full set of controls and mine and year fixed effects. Standard errors are clustered at the mine basin system level.

TABLE E11: Estimation results for varying sample definition

| Model:                        | (1)                    | (2)                    | (3)                    | (4)                    | (5)                   | (6)                   | (7)                   | (8)                   | (9)                    |
|-------------------------------|------------------------|------------------------|------------------------|------------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| <i>Maximum Vegetation EVI</i> |                        |                        |                        |                        |                       |                       |                       |                       |                        |
| Downstream, 1–3               | -0.0056***<br>(0.0020) | -0.0057***<br>(0.0020) | -0.0061***<br>(0.0020) | -0.0061***<br>(0.0021) | -0.0043*<br>(0.0022)  | -0.0045**<br>(0.0023) | -0.0055**<br>(0.0023) | -0.0046*<br>(0.0024)  | -0.0071***<br>(0.0024) |
| <i>Fit statistics</i>         |                        |                        |                        |                        |                       |                       |                       |                       |                        |
| Observations                  | 110,524                | 13,823                 | 69,107                 | 41,449                 | 58,545                | 95,822                | 8,152                 | 94,902                | 72,009                 |
| Number of upstream basins     | 6,084                  | 6,085                  | 6,084                  | 6,084                  | 4,159                 | 6,084                 | 533                   | 6,020                 | 4,672                  |
| Number of downstream basins   | 5,896                  | 5,899                  | 5,896                  | 5,896                  | 2,543                 | 5,896                 | 479                   | 5,845                 | 4,304                  |
| R <sup>2</sup>                | 0.90787                | 0.96239                | 0.91075                | 0.92199                | 0.89783               | 0.90765               | 0.92270               | 0.90891               | 0.90252                |
| Within R <sup>2</sup>         | 0.05272                | 0.06779                | 0.04812                | 0.04801                | 0.05360               | 0.05214               | 0.06617               | 0.05269               | 0.05244                |
| <i>Maximum Cropland EVI</i>   |                        |                        |                        |                        |                       |                       |                       |                       |                        |
| Downstream, 1–3               | -0.0068***<br>(0.0026) | -0.0069***<br>(0.0026) | -0.0071***<br>(0.0027) | -0.0068**<br>(0.0027)  | -0.0073**<br>(0.0029) | -0.0067**<br>(0.0029) | -0.0070**<br>(0.0035) | -0.0067**<br>(0.0030) | -0.0092***<br>(0.0032) |
| <i>Fit statistics</i>         |                        |                        |                        |                        |                       |                       |                       |                       |                        |
| Observations                  | 93,000                 | 11,683                 | 58,288                 | 34,883                 | 49,798                | 79,931                | 6,967                 | 79,123                | 60,357                 |
| Number of upstream basins     | 5,232                  | 5,255                  | 5,232                  | 5,232                  | 3,496                 | 5,232                 | 437                   | 5,175                 | 4,040                  |
| Number of downstream basins   | 4,767                  | 4,786                  | 4,767                  | 4,767                  | 2,066                 | 4,767                 | 399                   | 4,723                 | 3,477                  |
| R <sup>2</sup>                | 0.82177                | 0.90331                | 0.82521                | 0.83763                | 0.80996               | 0.82360               | 0.86559               | 0.82186               | 0.81508                |
| Within R <sup>2</sup>         | 0.03234                | 0.03192                | 0.02663                | 0.02700                | 0.03324               | 0.03211               | 0.06258               | 0.03282               | 0.03575                |
| <i>Fixed-effects</i>          |                        |                        |                        |                        |                       |                       |                       |                       |                        |
| Year                          | Yes                    | No                     | Yes                    | Yes                    | Yes                   | Yes                   | Yes                   | Yes                   | Yes                    |
| Mine                          | Yes                    | Yes                    | Yes                    | Yes                    | Yes                   | Yes                   | Yes                   | Yes                   | Yes                    |

Clustered (mine basin) standard-errors in parentheses

Significance: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

**Note:** Table shows results for estimation of Equation (1), with distance included as measured by the ordering of basins with respect to the mine basin and pooled for the first three basins immediately up- or downstream of the mine basin. The upper panel holds results from models for the vegetation-specific EVI as proxy measure for vegetative health within basins, the lower panel for the cropland-specific EVI as proxy measure for agricultural productivity. Models in column (1) are the baseline specifications, models in column (2) use outcomes and covariates averaged over years, where applicable, models in column (3) restrict the sample to data from 2019 and onwards, models in column (4) restrict the sample to data from the period 2018 to 2020. Models in column (5) only include basin systems with at least one up- and downstream basin, models in column (6) exclude the mine basin itself, models in column (7) include only basins of order  $\pm 1$  of mine-basin systems with at least one up- and downstream basin and exclude the mine-basin. Models in columns (8) and (9) report results from matching procedures. Models in column (8) matches on geophysical variables (elevation, slope, and soil), models in column (9) in addition matches on meteorological variables (precipitation and maximum temperature) and soil type. All specifications include the full set of controls and mine and year fixed effects, except for models in column (2). Standard errors are clustered at the mine basin system level.

TABLE E12: Estimation results for varying unobservables

| Model:                        | (1)                    | (2)                    | (3)                    | (4)                    | (5)                    | (6)                   | (7)                    | (8)                    |
|-------------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|-----------------------|------------------------|------------------------|
| <i>Maximum Vegetation EVI</i> |                        |                        |                        |                        |                        |                       |                        |                        |
| Downstream, 1–3               | -0.0056***<br>(0.0020) | -0.0062***<br>(0.0018) | -0.0062***<br>(0.0017) | -0.0057***<br>(0.0020) | -0.0056***<br>(0.0020) | -0.0056**<br>(0.0020) | -0.0056***<br>(0.0009) | -0.0056***<br>(0.0011) |
| <i>Fit statistics</i>         |                        |                        |                        |                        |                        |                       |                        |                        |
| Observations                  | 110,524                | 110,524                | 110,524                | 110,524                | 110,524                | 110,524               | 110,524                | 110,524                |
| Number of upstream basins     | 6,084                  | 6,084                  | 6,084                  | 6,084                  | 6,084                  | 6,084                 | 6,084                  | 6,084                  |
| Number of downstream basins   | 5,896                  | 5,896                  | 5,896                  | 5,896                  | 5,896                  | 5,896                 | 5,896                  | 5,896                  |
| R <sup>2</sup>                | 0.90787                | 0.90277                | 0.88482                | 0.91812                | 0.91341                | 0.90787               | 0.90787                | 0.90787                |
| Within R <sup>2</sup>         | 0.05272                | 0.05884                | 0.06696                | 0.03154                | 0.10975                | 0.05272               | 0.05272                | 0.05272                |
| <i>Maximum Cropland EVI</i>   |                        |                        |                        |                        |                        |                       |                        |                        |
| Downstream, 1–3               | -0.0068***<br>(0.0026) | -0.0059**<br>(0.0024)  | -0.0071***<br>(0.0022) | -0.0067***<br>(0.0026) | -0.0067***<br>(0.0026) | -0.0068**<br>(0.0026) | -0.0068***<br>(0.0011) | -0.0068***<br>(0.0016) |
| <i>Fit statistics</i>         |                        |                        |                        |                        |                        |                       |                        |                        |
| Observations                  | 93,000                 | 93,000                 | 93,000                 | 93,000                 | 93,000                 | 93,000                | 93,000                 | 93,000                 |
| Number of upstream basins     | 5,232                  | 5,232                  | 5,232                  | 5,232                  | 5,232                  | 5,232                 | 5,232                  | 5,232                  |
| Number of downstream basins   | 4,767                  | 4,767                  | 4,767                  | 4,767                  | 4,767                  | 4,767                 | 4,767                  | 4,767                  |
| R <sup>2</sup>                | 0.82177                | 0.80802                | 0.78259                | 0.83552                | 0.82970                | 0.82177               | 0.82177                | 0.82177                |
| Within R <sup>2</sup>         | 0.03234                | 0.03504                | 0.05020                | 0.01668                | 0.07542                | 0.03234               | 0.03234                | 0.03234                |
| <i>Fixed-effects</i>          |                        |                        |                        |                        |                        |                       |                        |                        |
| Year                          | Yes                    | Yes                    | Yes                    | No                     | Yes                    | Yes                   | Yes                    | Yes                    |
| Mine                          | Yes                    | No                     | No                     | Yes                    | Yes                    | Yes                   | Yes                    | Yes                    |
| Pfafstetter basin level 8     | No                     | Yes                    | No                     | No                     | No                     | No                    | No                     | No                     |
| Pfafstetter basin level 8     | No                     | No                     | Yes                    | No                     | No                     | No                    | No                     | No                     |
| Country-by-year               | No                     | No                     | No                     | Yes                    | No                     | No                    | No                     | No                     |
| Mine-specific time trends     | No                     | No                     | Yes                    | No                     | Yes                    | No                    | No                     | No                     |

Clustered standard-errors in parentheses

Significance: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1

**Note:** Table shows results for estimation of Equation (1), with distance included as measured by the ordering of basins with respect to the mine basin and pooled for the first three basins immediately up- or downstream of the mine basin. The upper panel holds results from models for the vegetation-specific EVI as proxy measure for vegetative health within basins, the lower panel for the cropland-specific EVI as proxy measure for agricultural productivity. Models in column (1) are the baseline specification with mine fixed effects. Models in column (2) use fixed effects at Pfafstetter level 8 basins, models in columns (3) fixed effects at Pfafstetter level 6 basins. Models in column (4) report results using country-by-year fixed effects, models in column (5) report results including mine-specific linear time trends. Models in column (6) cluster standard errors at mine and year level (two-way clustering), models in column (7) cluster at mine-by-year level, and models in column (8) cluster at the individual basin level. All specifications include the full set of controls. Standard errors are clustered at the mine basin system level, unless indicated otherwise.

## E4. Distance-based specifications

TABLE E13: Main estimation results specification using distance as running variable.

| Dependent Variables:<br>Model:             | Maximum Vegetation EVI |                        |                        |                        | Maximum Croplands EVI  |                        |                       |                       |
|--------------------------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|-----------------------|-----------------------|
|                                            | (1)                    | (2)                    | (3)                    | (4)                    | (5)                    | (6)                    | (7)                   | (8)                   |
| <i>Linear distance</i>                     |                        |                        |                        |                        |                        |                        |                       |                       |
| Downstream                                 | -0.0050**<br>(0.0023)  | -0.0045**<br>(0.0022)  | -0.0033<br>(0.0022)    | -0.0034<br>(0.0022)    | -0.0050*<br>(0.0029)   | -0.0049*<br>(0.0029)   | -0.0041<br>(0.0029)   | -0.0042<br>(0.0029)   |
| Downstream × Distance                      | -0.0008<br>(0.0047)    | -0.0036<br>(0.0054)    | -0.0083<br>(0.0054)    | -0.0085<br>(0.0053)    | 0.0015<br>(0.0059)     | -0.0004<br>(0.0069)    | -0.0059<br>(0.0070)   | -0.0060<br>(0.0069)   |
| Distance                                   | 0.0008<br>(0.0039)     | 0.0033<br>(0.0041)     | 0.0056<br>(0.0041)     | 0.0062<br>(0.0040)     | 0.0028<br>(0.0050)     | 0.0041<br>(0.0055)     | 0.0063<br>(0.0054)    | 0.0057<br>(0.0053)    |
| <i>Fit statistics</i>                      |                        |                        |                        |                        |                        |                        |                       |                       |
| Observations                               | 110,576                | 110,568                | 110,564                | 110,524                | 93,036                 | 93,036                 | 93,032                | 93,000                |
| Number of upstream basins                  | 6,084                  | 6,084                  | 6,084                  | 6,084                  | 5,232                  | 5,232                  | 5,232                 | 5,232                 |
| Number of downstream basins                | 5,899                  | 5,898                  | 5,898                  | 5,896                  | 4,769                  | 4,769                  | 4,769                 | 4,767                 |
| R <sup>2</sup>                             | 0.90282                | 0.90452                | 0.90762                | 0.90783                | 0.81609                | 0.81748                | 0.82138               | 0.82165               |
| Within R <sup>2</sup>                      | 0.00124                | 0.01844                | 0.05044                | 0.05234                | 0.00136                | 0.00892                | 0.03044               | 0.03169               |
| <i>Linear-quadratic distance</i>           |                        |                        |                        |                        |                        |                        |                       |                       |
| Downstream                                 | -0.0056**<br>(0.0027)  | -0.0055**<br>(0.0026)  | -0.0050**<br>(0.0025)  | -0.0052**<br>(0.0025)  | -0.0077**<br>(0.0035)  | -0.0076**<br>(0.0036)  | -0.0072**<br>(0.0035) | -0.0073**<br>(0.0035) |
| Downstream × Distance                      | 0.0026<br>(0.0117)     | 0.0020<br>(0.0112)     | 0.0006<br>(0.0117)     | 0.0005<br>(0.0115)     | 0.0159<br>(0.0154)     | 0.0136<br>(0.0155)     | 0.0108<br>(0.0155)    | 0.0105<br>(0.0155)    |
| Downstream × Distance <sup>2</sup>         | -0.0030<br>(0.0085)    | -0.0047<br>(0.0080)    | -0.0073<br>(0.0081)    | -0.0073<br>(0.0080)    | -0.0120<br>(0.0120)    | -0.0117<br>(0.0120)    | -0.0138<br>(0.0118)   | -0.0137<br>(0.0118)   |
| Distance                                   | 0.0040<br>(0.0091)     | 0.0039<br>(0.0086)     | 0.0033<br>(0.0089)     | 0.0036<br>(0.0086)     | -0.0004<br>(0.0117)    | 0.0001<br>(0.0119)     | 0.0012<br>(0.0118)    | -0.0001<br>(0.0120)   |
| Distance <sup>2</sup>                      | -0.0024<br>(0.0063)    | -0.0005<br>(0.0059)    | 0.0018<br>(0.0060)     | 0.0020<br>(0.0059)     | 0.0025<br>(0.0093)     | 0.0032<br>(0.0094)     | 0.0041<br>(0.0091)    | 0.0046<br>(0.0092)    |
| <i>Fit statistics</i>                      |                        |                        |                        |                        |                        |                        |                       |                       |
| Observations                               | 110,576                | 110,568                | 110,564                | 110,524                | 93,036                 | 93,036                 | 93,032                | 93,000                |
| Number of upstream basins                  | 6,084                  | 6,084                  | 6,084                  | 6,084                  | 5,232                  | 5,232                  | 5,232                 | 5,232                 |
| Number of downstream basins                | 5,899                  | 5,898                  | 5,898                  | 5,896                  | 4,769                  | 4,769                  | 4,769                 | 4,767                 |
| R <sup>2</sup>                             | 0.90283                | 0.90453                | 0.90762                | 0.90784                | 0.81612                | 0.81751                | 0.82142               | 0.82168               |
| Within R <sup>2</sup>                      | 0.00134                | 0.01851                | 0.05053                | 0.05243                | 0.00152                | 0.00906                | 0.03062               | 0.03186               |
| <i>Exponential decay</i>                   |                        |                        |                        |                        |                        |                        |                       |                       |
|                                            | $\delta = 0.0046$      | $\delta = 0.0058$      | $\delta = 0.0023$      | $\delta = 0.0025$      | $\delta = 0.0350$      | $\delta = 0.0350$      | $\delta = 0.0200$     | $\delta = 0.0099$     |
| exp $-\delta \times$ Distance × Downstream | -0.0062***<br>(0.0023) | -0.0062***<br>(0.0023) | -0.0060***<br>(0.0023) | -0.0062***<br>(0.0023) | -0.0093***<br>(0.0034) | -0.0091***<br>(0.0033) | -0.0074**<br>(0.0029) | -0.0068**<br>(0.0029) |
| <i>Fit statistics</i>                      |                        |                        |                        |                        |                        |                        |                       |                       |
| Observations                               | 110,576                | 110,568                | 110,564                | 110,524                | 93,036                 | 93,036                 | 93,032                | 93,000                |
| Number of upstream basins                  | 6,084                  | 6,084                  | 6,084                  | 6,084                  | 5,232                  | 5,232                  | 5,232                 | 5,232                 |
| Number of downstream basins                | 5,899                  | 5,898                  | 5,898                  | 5,896                  | 4,769                  | 4,769                  | 4,769                 | 4,767                 |
| R <sup>2</sup>                             | 0.90282                | 0.90451                | 0.90757                | 0.90778                | 0.81609                | 0.81747                | 0.82132               | 0.82159               |
| Within R <sup>2</sup>                      | 0.00124                | 0.01833                | 0.05000                | 0.05185                | 0.00137                | 0.00887                | 0.03007               | 0.03138               |
| <i>Fixed-effects</i>                       |                        |                        |                        |                        |                        |                        |                       |                       |
| Year                                       | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    | Yes                   | Yes                   |
| Mine                                       | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    | Yes                    | Yes                   | Yes                   |

Clustered (mine basin) standard-errors in parentheses

Significance: \*\*\*, 0.01, \*\*, 0.05, \*, 0.1

**Note:** Table shows results for estimation of Equation (1), with distance included as measured in kilometer along the river network. Columns (1)–(4) hold results from models for the overall EVI as proxy measure for vegetative health within basins, columns (5)–(8) for the cropland-specific EVI as proxy measure for agricultural productivity. Models in columns (1) and (5) include no additional covariates, models (2) and (6) control for geophysical variables (elevation, slope, and soil), models (3) and (7) additionally control for meteorological (yearly sum of precipitation and yearly maximum temperature), and models (4) and (8) additionally control for socioeconomic (accessibility to city in minutes and total population in 2015) conditions. All models include mine and year fixed effects. Standard errors are clustered at the mine basin system level.

Table E14: Distance specification using automatic bandwidth selection

|                                 | Max EVI               |                        | Max C EVI             |                        |
|---------------------------------|-----------------------|------------------------|-----------------------|------------------------|
|                                 | No Controls           |                        |                       |                        |
| Conventional                    | -0.0035**<br>(0.0017) | -0.0063***<br>(0.0021) | -0.0047**<br>(0.0024) | -0.0078**<br>(0.0031)  |
| Bias-Corrected                  | -0.0042**<br>(0.0017) | -0.0069***<br>(0.0021) | -0.0056**<br>(0.0024) | -0.0085***<br>(0.0031) |
| Bandwidth (conv)                | 18.5                  | 31.1                   | 18.8                  | 32.3                   |
| Observations (conv)             | 33407                 | 48169                  | 29225                 | 42708                  |
| Number upstream basins (conv)   | 1663                  | 820                    | 1825                  | 930                    |
| Number downstream basins (conv) | 2043                  | 1198                   | 2356                  | 1405                   |
| Bandwidth (bias)                | 42.1                  | 62.3                   | 43.5                  | 64.2                   |
| Observations (bias)             | 59945                 | 78314                  | 52514                 | 68045                  |
| Number upstream basins (bias)   | 3428                  | 2316                   | 3821                  | 2575                   |
| Number downstream basins (bias) | 3449                  | 2618                   | 4129                  | 3078                   |
|                                 | With full Controls    |                        |                       |                        |
| Conventional                    | -0.0028*<br>(0.0017)  | -0.0055***<br>(0.0021) | -0.0045*<br>(0.0023)  | -0.0075**<br>(0.0031)  |
| Bias-Corrected                  | -0.0034**<br>(0.0017) | -0.0061***<br>(0.0021) | -0.0054**<br>(0.0023) | -0.0083***<br>(0.0031) |
| Bandwidth (conv)                | 19                    | 32.3                   | 19                    | 33.1                   |
| Observations (conv)             | 34107                 | 49372                  | 29437                 | 43370                  |
| Number upstream basins (conv)   | 1953                  | 968                    | 1906                  | 978                    |
| Number downstream basins (conv) | 2490                  | 1437                   | 2429                  | 1448                   |
| Bandwidth (bias)                | 42.6                  | 65.6                   | 44                    | 65.4                   |
| Observations (bias)             | 60325                 | 80601                  | 52942                 | 68732                  |
| Number upstream basins (bias)   | 3967                  | 2702                   | 3977                  | 2607                   |
| Number downstream basins (bias) | 4255                  | 3165                   | 4262                  | 3097                   |
| Settings                        |                       |                        |                       |                        |
| Kernel                          | Triangular            | Triangular             | Triangular            | Triangular             |
| BW Criterion                    | mserd                 | mserd                  | mserd                 | mserd                  |
| Polynomial                      | Linear                | Quadratic              | Linear                | Quadratic              |

*Clustered (Mine) standard-errors in parentheses*

*Signif. Codes: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1*

**Note:** Table shows results for estimation of Equation (1), with distance as measured in kilometers along the river network used as the running variable, using practices suggested in Cattaneo, Idrobo, and Titiunik, 2019 for automatic bandwidth selection using a triangular Kernel and the mean squared error distance as selection criterion, and bias correction. Models in the upper panel include no covariates, models in the lower panel include the full set of controls. Models in columns (1) and (2) report results using the overall EVI as outcome, models in columns (3) and (4) for the cropland-specific EVI. Models (1) and (3) fit a linear polynomial of the distance measure at each side of the cutoff, models in columns (2) and (4) a quadratic polynomial. All specifications include mine and year fixed effects. Standard errors are clustered at the mine basin system level.

TABLE E15: Estimation results using covariates as placebo outcomes for squared distance specification

| Dependent Variables:<br>Model:     | Slope<br>(1)        | Elevation<br>(2)     | Temperature<br>(3) | Precipitation<br>(4) | Accessibility<br>(5) | Population<br>(6) |
|------------------------------------|---------------------|----------------------|--------------------|----------------------|----------------------|-------------------|
| <i>Variables</i>                   |                     |                      |                    |                      |                      |                   |
| Downstream                         | -0.009<br>(0.095)   | -0.234***<br>(0.071) | 0.067<br>(0.047)   | 0.014<br>(0.107)     | -0.036<br>(0.069)    | -2.84*<br>(1.62)  |
| Distance × Downstream              | 2.02***<br>(0.435)  | -3.12***<br>(0.368)  | -0.261<br>(0.255)  | -0.068<br>(0.635)    | -0.103<br>(0.385)    | 2.18<br>(5.41)    |
| Distance <sup>2</sup> × Downstream | -0.479<br>(0.380)   | -0.103<br>(0.309)    | -0.014<br>(0.217)  | 0.374<br>(0.540)     | 0.027<br>(0.291)     | -1.35<br>(3.39)   |
| Distance                           | -1.05***<br>(0.328) | 1.70***<br>(0.290)   | 0.188<br>(0.206)   | 0.400<br>(0.575)     | 0.850***<br>(0.275)  | -7.26<br>(4.44)   |
| Distance <sup>2</sup>              | 0.107<br>(0.291)    | 0.077<br>(0.228)     | -0.034<br>(0.185)  | -0.440<br>(0.531)    | -0.146<br>(0.204)    | 4.38<br>(2.77)    |
| <i>Fixed-effects</i>               |                     |                      |                    |                      |                      |                   |
| Year                               | Yes                 | Yes                  | Yes                | Yes                  | Yes                  | Yes               |
| Mine                               | Yes                 | Yes                  | Yes                | Yes                  | Yes                  | Yes               |
| <i>Fit statistics</i>              |                     |                      |                    |                      |                      |                   |
| Observations                       | 114,508             | 114,508              | 114,508            | 114,508              | 114,508              | 114,508           |
| Number of upstream basins          | 6,307               | 6,307                | 6,307              | 6,307                | 6,307                | 6,307             |
| Number of downstream basins        | 6,111               | 6,111                | 6,111              | 6,111                | 6,111                | 6,111             |
| R <sup>2</sup>                     | 0.75464             | 0.97225              | 0.95948            | 0.94458              | 0.88907              | 0.59487           |
| Within R <sup>2</sup>              | 0.19301             | 0.63133              | 0.40933            | 0.08597              | 0.05976              | 0.01816           |

*Clustered (mine basin) standard-errors in parentheses*

*Significance: \*\*\*: 0.01, \*\*: 0.05, \*: 0.1*

**Note:** Table shows results for estimation of Equation (1), with distance included as measured by the ordering of basins with respect to the mine basin using the additionally used covariates as placebo outcomes for the full sample. All specifications control for the remaining covariates except the one used as placebo outcome, as well as mine and year fixed effects. Standard errors are clustered at the mine basin system level.