

Mines → Rivers → Yields

Downstream Impacts of Mines on Agriculture in Africa

Lukas Vashold^{†*} Gustav Pirich[†] Maximilian Heinze[†]
Nikolas Kuschnig^{†*}

[†] *Vienna University of Business and Economics*

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 vegetation and agriculture across the continent. We exploit a natural experiment, where mines cause a discontinuity in water pollution along river networks, comparing vegetation health in upstream and downstream locations. We find that mines significantly reduce peak vegetation indices downstream by 1.3–1.5%, and impair the productivity of over 74,000 km² of croplands. Impacts may reach farther downstream, and are particularly strong in fertile regions and areas where gold mining predominates. Our findings highlight substantial externalities of mining and an urgent need for enhanced regulation and oversight to mitigate and monitor them.

*Correspondence to nikolas.kuschnig@wu.ac.at, at Welthandelsplatz 1, 1020 Vienna, Austria (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 requirements of the transition to cleaner energy [Pörtner et al., 2022]. While this can bring economic opportunities, mining also creates many negative externalities. It has been linked to pollution, environmental change, and biodiversity loss, as well as corruption, conflict, and child labor that undermine livelihoods.¹ These externalities are striking and may have far-reaching impacts that could offset the benefits of minerals. However, reliable and generalizable insights into these impacts remain scarce — particularly in a development context where fledgling institutions struggle to internalize the costs of mining, and negative impacts 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, where mine locations see a discontinuous jump in water pollution that follows river networks. Mines cause a sharp drop in vegetation health downstream, for which regions upstream can serve as a control group. To implement this research design, we use fine-grained river basins as our unit of analysis. We overcome data limitations by using satellite-derived peak vegetation indices with appropriate land use masks to measure vegetation health and agricultural productivity across the continent. By combining this information with a comprehensive dataset of mining sites, we can estimate the causal effects of water pollution from mines across Africa with unprecedented scope.

Our research design identifies impacts of mines on vegetation that are mediated by water.² The primary mechanism is water pollution; mines are known to cause acidification, elevated salinity, and heavy metal contamination. These pollutants affect vegetation and agricultural productivity by disrupting plant physiology, soil microbiomes, and nutrient uptake. Secondary mechanisms include adaptation be-

¹See, for instance, Aska et al. [2024], Berman et al. [2017, 2023], Giljum et al. [2022], Girard et al. [2024], Knutsen et al., Macklin et al. [2023], Santana et al. [2020]

²Our design has several notable parallels in the literature. Most closely related, Gittard and Hu [2024] investigate the impact of industrial mining on health outcomes in Africa, with impacts transmitted via water pollution. Sigman [2002] finds increased water pollution upstream of international borders, providing evidence of free-riding at a national level, while Lipscomb and Mobarak [2017] show that this problem also exists within countries. Dias et al. [2023] exploit the discontinuity to identify the impacts of glyphosate use in the cultivation of soybean in Brazil on birth outcomes, and Strobl and Strobl [2011] assess the impact of dams on agricultural productivity in Africa.

haviors — farmers switching crops or relocating in response to water pollution. The many other effects of mining, such as air pollution or economic effects, that are not directed along river flows are netted out by our research design.

The main contribution of our paper is the robust causal estimation of mining impacts on vegetation health that are mediated by water pollution — one important and previously neglected externality of mining. While hydrological studies suggest that vast areas and millions of livelihoods are affected by river pollution from mining [Macklin et al., 2023], causal evidence remains scarce and limited in scope. Previous research has investigated various other impacts of mining,³ and many studies document pollution from mine sites.⁴ The link between mining, water pollution, and vegetation health, however, remains underexplored at larger scales. Our study closes this gap by contributing strong causal evidence for negative effects across diverse mining operations throughout Africa.

Our study also speaks to several related strands of the literature. We add to earlier research on water pollution in a development context, typically focused on drinking water [e.g. Olmstead, 2010], by highlighting that water quality serves as a broad ecosystem service — not only an end in itself but also a means to other services [Keeler et al., 2012]. Our work complements research on looming water scarcity that acknowledges pollution concerns [e.g. Van Vliet et al., 2017, Jones et al., 2023], as well as recent studies that use remote-sensing methods to assess the impacts of mining [e.g. Giljum et al., 2022] in a data-scarce environment [see Maus and Werner, 2024]. To the extent that higher-income countries drive the demand for minerals and outsource polluting industries to lower-income countries, our results can be seen in the context of global environmental justice [see, e.g., Banzhaf et al., 2019].

We find that mines have considerable negative effects on vegetation downstream. Overall, peak vegetation is reduced by approximately 1.3 percent at the mean; for croplands specifically, peak vegetation is reduced by about 1.4 percent at the mean. These effects are economically meaningful — the immediately impacted area (up to around 33 km downstream) alone stretches over 280,000 km², of which 74,000 are croplands — corresponding to the total cropland area of Ghana. Conservative estimates place the resulting loss of cereal crops at 91,000 tons annually, although suggestive evidence indicates that effects persist much further downstream. We

³Aragón and Rud [2015], Aska et al. [2024], Berman et al. [2017], Gittard and Hu [2024], Ofofu et al. [2020], von der Goltz and Barnwal [2019]

⁴Awotwi et al. [2021], Du et al. [2024], Duncan [2020], Mulenga et al. [2024], Wu et al. [2023]

additionally investigate impact heterogeneities to identify important mechanisms, and find the largest negative effects in fertile regions and ones where gold mining predominates.

Our findings have important practical implications — agriculture plays a vital role in local livelihoods and economies, and large, fertile, populated areas are affected by mines. Water pollution from mines can cause great economic and nutritional disruptions, affecting the already severe and worsening food insecurity in Africa. From a policy perspective, our study highlights the severe lack of accountability mechanisms for the mining industry and the scarcity of data to assess their impacts.⁵ As a result, current research is limited in scope and imperfect proxies constrain the types of effects that can be detected. Remote-sensing approaches can help bridge this gap efficiently, but rely on conventional high-quality data for calibration.

The remainder of this paper is structured as follows. In the next section, we provide background and intuition for the key components of our analysis. In [Section 3](#), we describe the specific data and empirical strategy that we use to implement our research design. [Section 4](#), presents the main results, potential heterogeneities, and robustness checks for our findings. Finally, 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, using river basins as natural experimental units and satellite-derived indices to measure outcomes. This approach allows us to isolate causal effects by comparing areas upstream and downstream of mines.

2.1. Mines and mining in Africa

Mining on the African continent has experienced considerable growth, largely driven by the increasing global demand for minerals and metals [[International Council on Mining and Metals, 2022](#)]. This expansion presents both opportunities and challenges for African economies. While mining can generate employment and stimulate local economic development [[Bazillier and Girard, 2020](#), [Gräser, 2024](#), [von der Goltz](#)

⁵Also see [Auld et al. \[2018\]](#), [Maus and Werner \[2024\]](#), [Jones et al. \[2024\]](#)

and Barnwal, 2019, Ofosu et al., 2020], it also brings substantial risks. Resource wealth may crowd out other industries, drive corruption and conflict [Berman et al., 2017, 2023, Knutsen et al.], and mining drives rapid environmental changes [Aska et al., 2024, Barenblitt et al., 2021, Giljum et al., 2022, Girard et al., 2024] and creates pollution with considerable impacts on people and their environments [Awotwi et al., 2021, Macklin et al., 2023].

Mines affect vegetation and agriculture through multiple channels, with water pollution playing a particularly important role [Santana et al., 2020]. Mining disturbs orders of magnitude more material than the metal eventually extracted, creating erosion, and generating waste materials. These waste products include sediment, rock, and (heavy) metals, and are stored as ‘tailings’, which are large deposits of byproducts. Tailings and toxic chemicals used in processing are well-known ecological risks [see, e.g., Wu et al., 2023, for a meta-analysis of sediment pollution in China]. Sulfides cause acid mine drainage, leading to acidification that can persist for hundreds of years, degrades water quality, and devastates aquatic ecosystems [Du et al., 2024, Johnson and Hallberg, 2005]. Sodium cyanide and (in artisanal mining) mercury are used for the extraction of gold [Duncan, 2020, Malone et al., 2023, Verbrugge et al., 2021], reduce biodiversity, disrupt nutrient cycles, and cause long-term alterations of sediment chemistry.

Water plays a central role in the environmental impacts of mining. Operations require copious amounts of water for processing, which is often returned to the environment contaminated with chemicals, heavy metals, and sediment. Tailings are oxidized by air and weathered by rain, which steadily causes pollution of water resources as they feed into rivers [Schwarzenbach et al., 2010]. The resulting pollution from heavy metals [Frossard et al., 2018], increased salinity [Russ et al., 2020, Zörb et al., 2019], and acidification [Du et al., 2024] impact vegetation health via plant physiology and growth, by disrupting nutrient uptake, and by impairing soil microbiomes. (See Appendix B1 for a literature overview on the effects of water pollution and plant health].

Beyond water pollution, mines also affect vegetation through air pollution — particularly from coal mining and processing of metals in smelters [Fugiel et al., 2017, Miao et al., 2017, Pandey et al., 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.

These risks of mining are particularly acute in Africa, where mines often lack effective oversight [Macklin et al., 2023]. The mining industry and its large industrial mining sites have been slow to adopt global regulations,⁶ while the prevalent small-scale, artisanal mines often operate outside regulatory frameworks entirely. As a result, artisanal miners often employ particularly dangerous processing methods, such as the use of mercury for gold mining [Barenblitt et al., 2021], and have little incentive to manage tailings in sustainable ways.

The combination of expanding mining operations and weak institutional oversight endangers both the environment and local populations. Countries that lack robust institutional frameworks to monitor and regulate mining face severe challenges in managing its detrimental externalities. This situation is especially concerning given 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, Johnson and Hallberg, 2005].

2.2. River basins and water streams

River basins are an ideal unit of observation for studying water-mediated impacts of mining due to their natural hydrological properties. A river basin is defined as the area that is drained by a river and its tributaries. Following the prevailing elevation and slope, all surface waters in the area converge to a single point.⁷ Basin boundaries are defined by topological features such as ridges and hills, creating

⁶A recent exception is the 2020 Global Industry Standard on Tailings Management [GISTM; UNEP, 2023], which was established to prevent future tailing dam failures after the catastrophic dam collapse at Vale’s Corrego de Feijao mine in Brazil.

⁷Two prominent examples 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.

naturally distinct hydrological units. This definition can be applied recursively, ranging from continental to local scales and giving rise to a hierarchical system of nested basins, in which larger basins are divided into smaller sub-basins that merge at river confluences [Lehner and Grill, 2013].

The most important property of river basins for our analysis is the unidirectional flow of surface water.⁸ Water flows downstream, following elevation gradients and carrying pollutants with it. This creates a natural discontinuity at the mine site: downstream basins may be affected by contaminated water, while upstream basins remain unaffected and can serve as controls. A basin’s distance to the mine site reflects the intensity of the treatment, as well as the degree to which treated and control units are comparable.

We use fine-scale river basins, illustrated in the left panel of Figure 1, as our units of observation. These basins have an average area of 124.4 km², diameter of 12 km, and elevation differences of eleven meters, providing sufficient resolution to detect localized impacts of mining. Their relatively small size ensures that relevant factors vary smoothly between adjacent basins, suggesting that discontinuities in vegetation and agricultural productivity can be attributed to mine sites themselves. The boundaries of river basins are determined by natural topographic features, relative to which exact mine locations can reasonably be considered as random. While rivers often serve as natural borders in political and administrative contexts, the basin-level is rarely considered outside specialized hydrological studies.

2.3. Agricultural productivity from space

Measuring the impacts of mining on agricultural productivity in Africa presents significant challenges that we address using remote sensing data. For our analysis, we need a temporally and spatially consistent measure of agricultural production. Traditional methods like surveys or censuses face several challenges with consistency, availability, and accuracy in the context of our study. The continent is characterized by diverse climates, cropping systems, and agricultural practices. Smallholder farming plays an important role, agriculture is predominantly rainfed, and conventional agricultural statistics are scarcely available.

⁸River basins correlate strongly with groundwater systems, but only capture the flow of surface water.

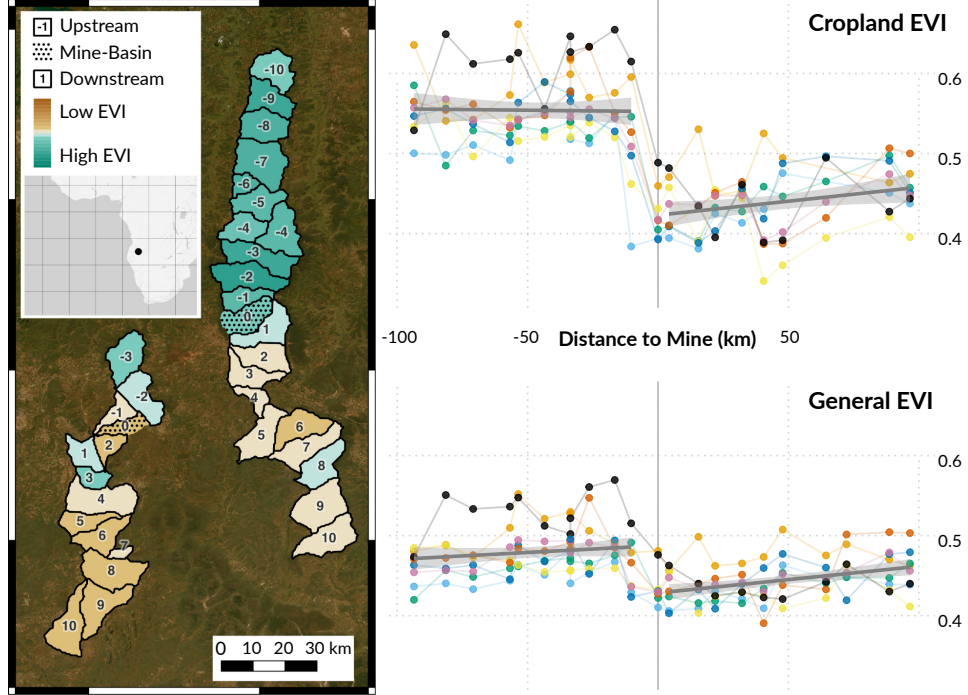


FIGURE 1: Example of two Angolan mine sites (dotted, and labeled with '0') and their upstream and downstream basin systems (left), and measurements of the Enhanced Vegetation Index (EVI) for croplands and general vegetation over the years 2016, 2017, 2018, 2019, 2020, 2021, 2022, and 2023 for the right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

To overcome these data limitations, we employ satellite-derived vegetation indices as proxies for vegetation health and agricultural productivity. This provides us with consistent measurements for the entire continent, with high granularity across time and space. Specifically, we use the Enhanced Vegetation Index (EVI) to construct a peak vegetation index, which correlates strongly with measures of gross primary production of vegetation and agriculture [Shi et al., 2017, Johnson, 2016], and is frequently used in similar analyses [see, e.g., Asher and Novosad, 2020, Agarwal et al., 2024, Wuepper et al., 2023]. The EVI offers several advantages over alternative vegetation indices [see Zeng et al., 2022, for a recent review] that make it suitable for our application.⁹

To distinguish between agricultural areas, general vegetation, and other land use types not relevant for our analysis, we apply vegetation and cropland land use masks to the EVI. An example of both EVI measurements along a basin system is provided in Figure 1, which demonstrates how the indices capture vegetation patterns up- and downstream of a mining site. By combining these satellite-derived measures with our basin-level approach, we can identify granular discontinuities in vegetation health and agricultural productivity, while maintaining consistent measurement across the diverse environmental and institutional conditions throughout Africa.

2.4. Research design

Our empirical strategy integrates these three components — mining sites, river basins, and remotely sensed vegetation — to identify causal effects of mining activity on agricultural productivity and vegetation health downstream across the African continent. Figure 2 illustrates our approach.

First, we identify basins that contain mine sites. Mines impact their surrounding environment in various ways and create a sharp discontinuity in vegetation patterns at the basin boundaries. Second, we identify basins located up- and downstream of the mine basin within the river network. Third, we isolate the river-mediated impact that follows the directed flow of water along this network. We assess the vegetation downstream and contrast it with vegetation upstream, allowing us to control for

⁹Compared to the commonly used Normalized Difference Vegetation Index (NDVI), the EVI reduces the influence of atmospheric distortions and background noise caused by canopy and soil [Gao et al., 2000]. Moreover, the EVI does not suffer from the NDVI's scaling problems, and can accurately convey variation in low as well as high biomass conditions where the NDVI tends to saturate [Huete et al., 2002].

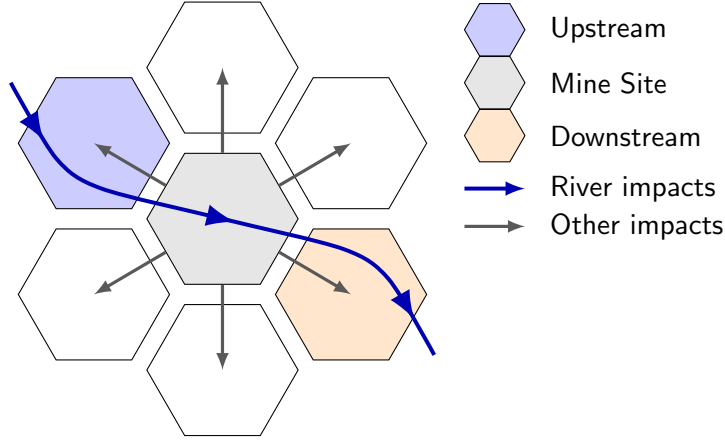


FIGURE 2: Illustration of the research design. The comparison of up- and downstream basins enables the identification of mine impacts that are mediated by the river.

general effects of mines (e.g., air pollution or economic effects) and identify the causal effect that is mediated by water.

Our strategy relies on the key assumption that the observed discontinuity along the river stems from the mine itself. This is plausible when mine sites are located quasi-randomly, and not driven by factors related to vegetation or agricultural potential. Mine locations are primarily determined by the location of accessible mineral deposits. While potential confounders exist, e.g., due to strategic placement near transportation infrastructure, these are unlikely to confound the effect of interest for two reasons. First, their impacts would need to align with the direction of water flow, and second, they would have to affect the mine location at the fine-scale basin level.

The immediate vicinity of mine basins provides strong causal identification, but estimating the reach and decay of impacts over greater distance presents additional challenges. In areas close to mines, basin characteristics remain balanced due to the granular level of observations. At greater distances, however, fundamental differences may emerge due to broader geographical patterns, and minor differences may begin to accumulate in impact. Other sources of water pollution, for example, are slightly more likely to be located downstream, as many mines are located at higher altitudes and river transport is economically attractive. While these factors arguably do not affect the discontinuity on-site, they could confound our assessment of impact decay.

To address these challenges while providing suggestive evidence for the extent of impacts, we divide our analysis in two parts. First, we focus on effects in the immediate vicinity, which can be cleanly causally identified in a model-agnostic fashion. Second, we impose a specific model on the impact of mines to investigate the potential range and decay of impacts over greater distances while accounting for potential confounders. This dual approach allows us to maintain rigorous causal identification while also exploring broader spatial patterns of mining impacts.

Mediators and limitations Our research design only captures impacts of mining on vegetation that are mediated by rivers, providing a focused but partial view of mining’s total external effects. Direct impacts on-site, such as land clearing, and impacts that are not directed along the river, such as dust deposition on plants or employment effects, are not captured by our estimates. This limitation helps us isolate a specific and possibly far-reaching impact channel — water pollution — but means that our results represent a lower bound for the total impact of mining.

The effect of water pollution that we identify in our study may be mediated by several factors that influence our interpretation of estimates. Most importantly, human adaptation to impacts is reflected in our estimates. If farmers relocate in response to pollution, this could amplify the measured impact. Conversely, other adaptations such as changing crops or farming practices might attenuate impacts. Another important mediator is soil pollution, which particularly affects the prevalent rainfed agriculture in Africa. As water pollutants settle in the soil and groundwater, they can impact plant growth in areas that are not directly irrigated from the polluted river. This means our estimates likely capture both direct water effects on irrigated agriculture and indirect soil-mediated effects on rainfed farming.

These mediating factors help contextualize our findings and suggest directions for future research. Studies of effective adaptation behavior, in particular, may be helpful to mitigate negative impacts. While our approach cannot disentangle these mechanisms, it provides robust causal evidence for the overall impact of mining on downstream vegetation and agricultural productivity — one important and neglected instance of the many externalities of mining.

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 that we observed over eight years (from 2016–2023). These basins, drawn from the HydroBASINS dataset [see [Lehner and Grill, 2013](#)], are delineated using information on water body locations, elevation, terrain slope, and stream gradients. The dataset has a hierarchical structure with twelve nested levels, where basins at each level are of comparable size, of which we use the most fine-scale Level 12.¹⁰ The key feature of river basins is their directional water flow — water only moves downstream, meaning that upstream basins remain unaffected by conditions further down the stream. 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.

For this classification, we consider up- and downstream chains of basins (departing from mine-basins) with a maximum order of ten (or an average river distance of 105 km).¹¹

We observe a total of 1,900 mine-basins, 6,307 upstream, and 6,127 downstream basins, which are visualized in [Figure 3](#). (For a summary of orders and distance, see [Table E2](#) in the Appendix.) Their average size is 120 km², for a total area of 1,701,343 km². Each basin in our dataset is unique, although they 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. This is because streams

¹⁰At Level 12, there are 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) km².

¹¹Our analysis is focused on the immediate vicinity, and more restrictive sample subsets are considered as robustness checks for our results.

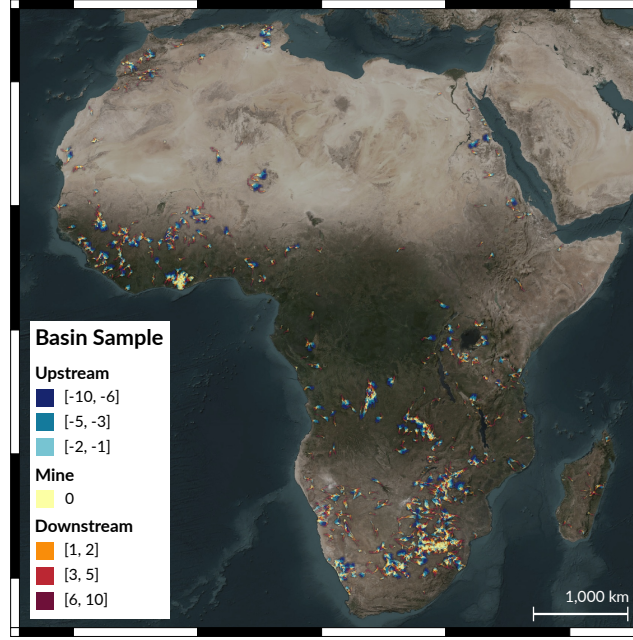


FIGURE 3: Basins in the sample and their treatment status. Basemap imagery provided by Esri, Maxar, Earthstar Geographics, and the GIS User Community.

can join, but do not split in the direction of their flow (compare the basins labeled ‘-4’ in Figure 1). Not all mine-basins have a full set of up- or downstream basins, and the number generally decreases with order. When mine-basins are located in or near the top or bottom branch of a river network, or if another mine-basin is situated up- or downstream of the mine-basin, there will be fewer relevant basins up- and downstream. More details on the basin data are provided in Appendix A.

Summary statistics Table 1 presents summary statistics for the most important variables used. This includes vegetation indices, geographical information on elevation and slope, meteorological information on temperature and precipitation, as well as 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 shows that vegetation indices using the cropland mask are slightly higher than for the general vegetation mask. Since not all basins contain croplands, the sample that we use to assess impacts on agricultural productivity is smaller. The variables exhibit strong variation across our sample, including, e.g., high and low altitude basins as well as densely populated and unpopulated ones. The characteristics of up- and

downstream basins are well-balanced across covariates.¹² Downstream basins exhibit slightly higher precipitation, and are more populated, but less accessible, and lie at lower altitudes. These minuscule differences across our treatment and control groups alleviate concerns of potential non-comparability.

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

Variable	Unit	<i>NT</i>	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 rely on the remotely sensed Enhanced Vegetation Index (EVI). The EVI offers important advantages over traditional crop statistics for our analysis. It offers the high spatial granularity needed to measure localized impacts, is frequently available over the investigated period, and it is consistent over time and space, avoiding cross-country differences in calculation and reporting.

We process the raw EVI data [by [Didan, 2015](#), ; available every 16 days at a 250 m resolution] in several steps to create our outcome measures. First, we apply quality filters to minimize the impact of cloud cover. Then, we use yearly land use masks [from the European Space Agency’s (ESA) Climate Change Initiative (CCI), by [Defourny et al., 2024](#), ; available at a 300 m resolution] to identify relevant areas with croplands and vegetation, while excluding irrelevant features such as water bodies. For each basin, we aggregate the mean EVI over 16-day periods and compute the annual maximum EVI, producing a peak vegetation index. This peak

¹²See [Table E3](#) in the Appendix for summary statistics that are split by treatment status, [Figure D11](#) for a visualization of sample distributions of geophysical and meteorological variables. [Appendix C3.1](#) includes further information about balance, and [Figure D12](#) shows absolute standardized mean differences for all variables.

index correlates strongly with measures of gross primary production of vegetation [see Shi et al., 2017, for an assessment] and crop yields [see, e.g., Azzari et al., 2017, Johnson, 2016].

We apply two different land use masks to create two distinct outcome measures. For agricultural productivity, we focus on all forms of croplands as well as mosaics between croplands and natural vegetation, which are common classifications for the small-scale farming that is common in Africa. For general vegetation health, we additionally consider land use classes comprising trees, shrubland, grassland as well as sparse and flooded natural vegetation. As a robustness check, we consider an alternative cropland mask that is specifically targeted to the African continent [Digital Earth Africa, 2022]. This mask is only available for 2019 (which introduces noise), but yields results that are qualitatively and quantitatively similar.

While the EVI offers several advantages over alternative vegetation indices [see Zeng et al., 2022, for a recent review], we validate our findings using the Normalized Difference Vegetation Index (NDVI) as an alternative outcome measure and find comparable results. We also link our EVI-based measures to agricultural yields and production values from AReNA-DHS [IFPRI, 2020] for additional validation in our setting and to gauge the quantity of impacts measured. While the EVI provides consistent measurements at a fine scale, it has limitations: it does not capture specific crops or yields, displaying heterogeneous correlations, complicating interpretation, and introducing noise into the analysis.

We consider a number of robustness checks to limit the impacts of our specific choice of outcome variables. This includes the aforementioned alternative information sources, [NDVI, AReNA-DHS, and Digital Earth Africa, 2022] as well as alternative aggregation procedures and land use masks. For one, we use the annual mean EVI instead of the maximum, to capture the full year. We also compute the pixel-level maximum before aggregating to better capture vegetation with different peak timings. Finally, we test a narrower version of the vegetation mask that excludes sparse and flooded natural vegetation. As we show in Section 4.3, our results remain robust across these variations.

3.1.2. Treatment

Our treatment is derived from the spatial location of mining sites, which we obtain from Maus et al. [2022]. Their dataset integrates multiple sources, including the SNL

Metals & Mining database, and provides comprehensive coverage of both industrial and artisanal mining operations. This source offers two key advantages over other databases. First, it addresses the noise in commonly used point location data by verifying actual mine locations through satellite imagery. Second, [Maus et al. \[2022\]](#) scan for and delineate mining areas (including features such as tailings dams, waste rock dumps, water ponds, and processing plants) within a 10-kilometer buffer around known locations, allowing them to capture smaller artisanal mining operations that often develop near larger industrial sites. This helps minimize potential attenuation of our estimates that could occur from missing or misclassified mine locations.

While the dataset by [Maus et al. \[2022\]](#) provides extensive spatial coverage, it lacks mine characteristics and temporal information beyond being anchored in 2019. This limitation, however, is less concerning given the persistence of mines and their impacts. Pollution from mines may persist long after operations have ceased [e.g. [Audry et al., 2004](#)], and even ‘inactive’ mines continue to affect local environments through illegal artisanal mining and ongoing pollution from tailings [[Macklin et al., 2023](#)]. To still address these limitations in our analysis and investigate potential heterogeneities, we (a) evaluate time-based subsets, (b) differentiate between active and inactive sites using longitudinal mine site data [from [Sepin et al., 2025](#)], and (c) incorporate information on commodity types from supplementary sources.¹³

To construct our treatment variable, we first identify mine-basins by intersecting the centroids of mining sites from [Maus et al. \[2022\]](#) with hydrological basins from [Lehner and Grill \[2013\]](#). We then classify basins as upstream or downstream relative to these mine-basins (as described above), with downstream basins forming our treatment group and upstream basins serving as controls. While mine-basins themselves are also treated, the water-mediated effects within them are not identified by our research design, and not interpreted by us. We consider basins up to order ten, but focus our analysis on the first three basins in either direction, as they share characteristics and would be most strongly impacted by the treatment.

We measure treatment intensity in terms of the distance to the mine-basin. For our primary specification, we consider the ordinal position of basins relative to the mine-basin as the treatment variable. This agnostic approach avoids common spec-

¹³Specifically, the [Global Energy Monitor](#) database, as well as [S&P Global Market Intelligence \[2025\]](#), [Jasansky et al. \[2023\]](#), [Padilla et al. \[2021\]](#)

ification issues in regression discontinuity designs,¹⁴ while sacrificing degrees of freedom. Additionally, we consider different specifications based on river distance. Estimates from an exponential decay specification that we use to extrapolate insights, and alternative linear-quadratic distance specifications, including established robust methods, appear in [Appendix C3](#).

3.1.3. Other factors

We consider several covariates that could confound our analysis of mining impacts or add more in-depth insights to it. Our selection of control variables is particularly focused on geographical characteristics that relate directly to river basin delineation, and ones that could create imbalances between treatment and control groups.

For the key geographic variables of elevation and slope, which define the flow along basins as well as their boundaries, we utilize high-resolution gridded data [[Amatulli et al., 2018](#)]. We aggregate this data from its original 30 arcsecond resolution (approximately 802–926 meters in our study area) to the basin level by averaging. Soil characteristics also influence vegetation health and may vary systematically with elevation. To account for this, we determine the primary soil class in each basin based on data from the SoilGrids project [[Hengl et al., 2017](#)], which has a 250-meter resolution.

Climate and socioeconomic factors become increasingly important at larger distances from mine sites, where treatment and control groups may diverge systematically. As climate controls, we use precipitation data from the Climate Hazards Center InfraRed Precipitation with Station (CHIRPS) dataset [[Funk et al., 2015](#)], which is particularly accurate for Africa [[Dinku et al., 2018](#)], as well as temperature data from TerraClimate [[Abatzoglou et al., 2018](#)]. Both datasets provide monthly measurements at resolutions of approximately 4 and 5 kilometers. As an alternative, we also consider meteorological data from the Climatic Research Unit [[Harris et al., 2020](#)]. Following established practices in research of vegetation dynamics, we use annual precipitation totals and maximum monthly temperatures in our analysis.

For socioeconomic controls, we include the total population [from [WorldPop](#)] and average accessibility, measured by the travel time to the nearest city [[Weiss et al., 2018](#)], per basin. Both datasets provide information at a 1-kilometer resolution and

¹⁴For instance, the choice of appropriate polynomials or bandwidth for continuous running variables; see [Cattaneo et al. \[2019\]](#), [Gelman and Imbens \[2019\]](#)

are aggregated to the basin level. We use values from 2015 to control for initial conditions in our study period. To further isolate the effect of interest from possible mediators, we also consider the number of violent events from the Armed Conflict Location and Event Database [ACLED; [Raleigh et al., 2010](#)] and average annual air pollutant concentrations [from [Shen et al., 2024](#)].

Heterogeneity The impacts of mines may vary across different environmental conditions and mine characteristics. We investigate this by examining heterogeneity across four main dimensions.

First, we assess spatial heterogeneities of basins. We differentiate them by biomes using the Ecoregion classification of [Dinerstein et al. \[2017\]](#), which divides ecosystems of regional extent that capture different types of agriculture and vegetation. Furthermore, we consider country groups based on their primary crops and varying crop calendar cycles, based on the regional classification of the US Department of Agriculture (USDA).¹⁵ These environmental categorizations help us understand how mining impacts might differ across diverse ecological contexts.

Second, we analyze how the intensity of mining operations and the activity of mining sites relates to their impacts. We proxy for intensity via the total mine area per mine-basin, and use the development of mining areas over time to proxy for activity. For the latter, we rely on a longitudinal extension of the mine dataset by [Sepin et al. \[2025\]](#), 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.

Third, we investigate whether impacts vary by the types of minerals being extracted, as different commodities use different processes, chemicals and techniques, and coincide with different waste metals. For this analysis, we collate various data sources and predict the most likely commodities mined in each basin using Gaussian process regression (which is detailed in [Appendix C2](#)). This helps us identify particularly important types of mining sites.

Lastly, we examine how effects differ based on land suitability for agriculture. Using information from the Global Agro-Ecological Zones (GAEZ) modeling frame-

¹⁵The classification (available at 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.

work [Fischer et al., 2021], we construct a composite measure of crop suitability for each basin. We calculate average suitabilities for thirty crop types per basin, and select the maximum value that represents the suitability for the best-suited crop. Then, we categorize basins into high, medium, and low productivity classes, allowing us to assess which agricultural areas are most affected by mining.

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 discontinuity itself. While parts of the mine basin are impacted by contaminated water, others are not, and other third effects of the mine may be present without a suitable control. Impacts on-site provide useful information and suggestive evidence for further impacts of mine sites, but are not identified by our strategy.

Let x denote the directed distance from the mine, 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

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

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

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

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 in basin i near mine m at time t , and x_i is the running distance. The vector W_{it} contains basin characteristics,¹⁶ while μ_m and ψ_t represent mine- and year-fixed effects. The error term ε_{imt} is clustered by mine. We operationalize distance via the function $F(\cdot)$, which returns a vector, separating upstream, downstream, and mine basins.

Main 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

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

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 effect of interest. For ease of interpretation, we also report pooled estimates comparing the first three basins.¹⁷

River distance specification 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

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

where $+$ separates variables, and the parameters γ, δ 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

¹⁶Of 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).

¹⁷We let $f(x)_{\mathcal{J}} = \mathbf{I}(x \in \mathcal{J})$ for $\mathcal{J} \in \{-10, \dots, -4, 0, \{1, 2, 3\}, 4, \dots, 10\}$, and focus on $\beta_{\{1,2,3\}}$.

form allocates more leverage, and thus higher influence, to observations at shorter distances and not vice versa.

3.2.1. Identifying assumption

Our key identifying assumption posits no other discontinuous changes impacting vegetation at the mine basin. While not directly testable, several factors support the plausibility of this assumption. River basins are delineated by natural geographic features that vary smoothly across space. Our use of the most granular basin level (Level 12) minimizes systematic differences between adjacent basins.

To validate our results, we conduct a battery of robustness checks to address potential threats to identification. First, we control for an extensive set of basin characteristics: geophysical features (elevation, slope, distance to coast, soil composition), meteorological conditions (temperature, precipitation), and socioeconomic indicators (population, accessibility, conflict). Second, we use these variables as placebo outcomes and look for discontinuities at mine basins, which would violate our identifying assumption. Third, we implement a matching procedure based on these covariates to ensure that up- and downstream basins are comparable across relevant dimensions. This reduces model dependency and strengthens our argument for causal effects.

As an additional set of robustness checks, we vary the sample definition to ensure that our results are not driven by our particular sample. Our main sample is already limited to the vicinity of mines (the maximum order of ten corresponds to an average distance of 100 km), and we emphasize estimates for their immediate surroundings. We also restrict this sample in additional ways. We (a) exclude the mine basin itself, (b) only retain first order basins, and (c) only consider basin chains with at least one up- and one downstream basin. To address temporal uncertainty stemming from the mine dataset, we consider two subsamples: (d) post-2019, and (e) from 2018–2020. Lastly, we repeatedly randomize the treatment assignment and compare estimates with our sample. The consistency of results across these variations supports the robustness of our findings.

These complementary approaches address different potential threats to identification. The covariate controls and matching handle potential confounding variables, the placebo tests check for other discontinuities, and the sample variations ensure our results are not artifacts of particular specification choices. Together, they pro-

vide strong support for our identification strategy, suggesting that estimates convey the water-mediated impact of mines on vegetation.

4. Results

Our analysis reveals that mines significantly reduce vegetation health in downstream areas through water pollution. 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), 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. Complete results, with estimates for all basin orders and covariates, are available in [Table E4](#) in the Appendix.

The treatment indicators show statistically significant negative effects on vegetation downstream of mines, for both indices of general and cropland-specific vegetation. The (gradual) inclusion of covariates changes estimates only marginally (see [Table E4](#) 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 them with similar magnitude, though declining statistical significance.

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

TABLE 2: Main estimation results.

<i>Outcome</i> (Specification)	<i>Peak Vegetation</i> (Plain)	<i>Peak Vegetation</i> (Full)	<i>Peak Cropland Veg.</i> (Plain)	<i>Peak Cropland Veg.</i> (Full)
Individual Order				
Downstream (1 st)	-0.0045*** (0.0017)	-0.0043** (0.0018)	-0.0051** (0.0025)	-0.0050** (0.0025)
Downstream (2 nd)	-0.0049** (0.0022)	-0.0048** (0.0024)	-0.0058* (0.0031)	-0.0067** (0.0032)
Downstream (3 rd)	-0.0085*** (0.0028)	-0.0087*** (0.0029)	-0.0088** (0.0037)	-0.0099*** (0.0038)
Downstream (4 th)	-0.0049* (0.0030)	-0.0062* (0.0033)	-0.0029 (0.0038)	-0.0044 (0.0040)
Downstream (5 th)	-0.0034 (0.0033)	-0.0053 (0.0037)	0.0007 (0.0042)	-0.0016 (0.0045)
<i>Fit statistics</i>				
Observations	110,576	110,524	93,036	93,000
R ²	0.903	0.908	0.816	0.822
Pooled Order				
Downstream (1 st –3 rd)	-0.0057*** (0.0018)	-0.0056*** (0.0020)	-0.0064** (0.0025)	-0.0068*** (0.0026)
<i>Fit statistics</i>				
Observations	110,576	110,524	93,036	93,000
R ²	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 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 column 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.

ATE 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 basin, 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 km².

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 highly significant ($p < 0.01$). 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 drop in magnitude is larger and there is one instance of an (insignificant) positive estimate. Compared to the sample mean, these estimates imply an index reduction of 1.38–1.47% over an affected cropland area of 74,000 km².

Contextualizing impacts To translate these peak vegetation index reductions into agricultural terms, we correlated our measure with direct measures of agricultural productivity (from AReNA-DHS, see [Appendix B2](#)). We estimate that being immediately downstream of a mine is associated with a 0.57–0.61% decrease in cereal yields, a 1.59–1.70% decrease in the value of cereal production, and a 2.16–2.31% decrease in the value of overall crop production. The scale of these impacts becomes apparent when considering the affected area — the first three downstream basins cover 280,000 km², 74,000 of which are croplands. This corresponds to the total cropland area of Ghana, falling just between the cropland areas of the United Kingdom and Malaysia.

Using an average cereals yield of two tons per hectare (based on AReNA-DHS), this corresponds to an annual loss of 91,000 tons of cereals for consumption. For additional context, the World Food Programme (WFP), the largest humanitarian organization globally, distributed approximately 1.7 million tons of food to over 38 million recipients in African countries in 2023. Our conservative estimate of the agricultural production loss caused by water pollution from mines, just in their immediate vicinity, represents about 5.4% of this major 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 Table E6 and Table E7 in the Appendix.

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 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.¹⁸ 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.

¹⁸The 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.

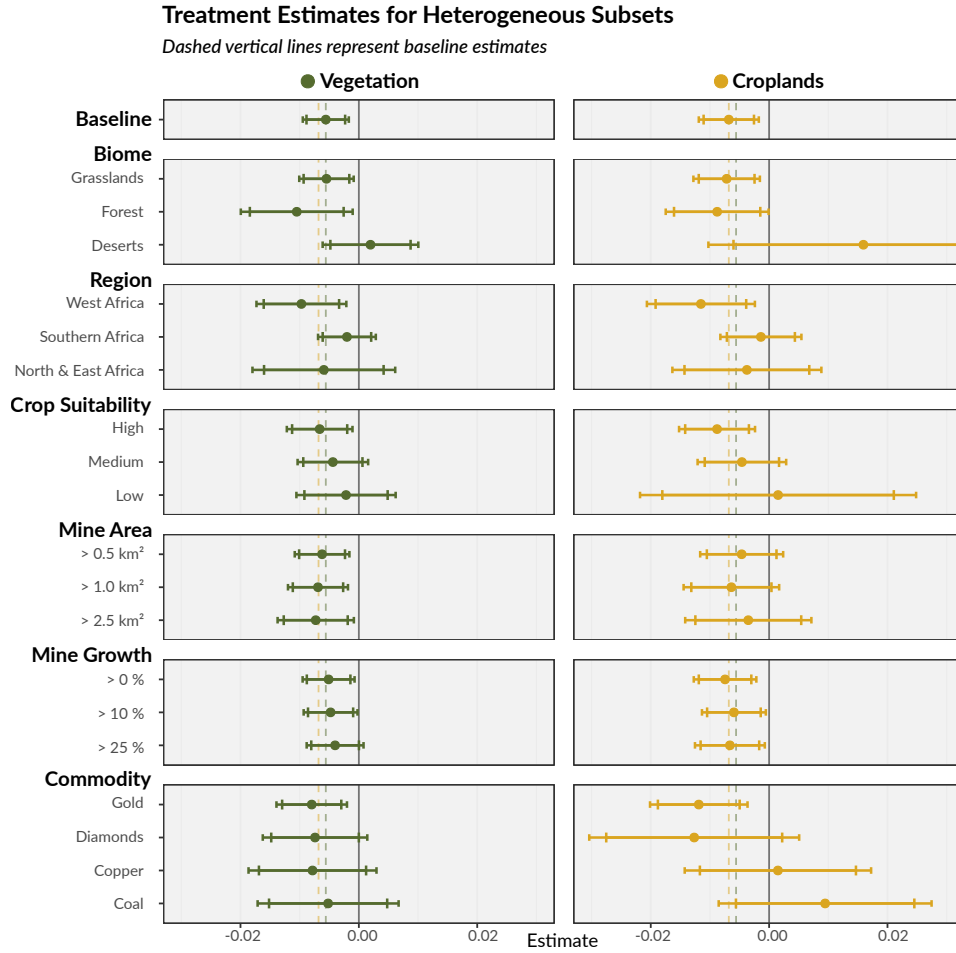


FIGURE 4: Dots indicate the average treatment effect in the first three basins; whiskered lines indicate 90% and 95% confidence intervals.

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. These estimates 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 km² (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.

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 km², 1 km², and 2.5 km², which reduces the sample to around 51%, 40%, and 25% of its original size. For overall vegetation, we observe a clear increase in effect size as mining area increases. For the 2.5 km² cutoff, the estimate is 30% higher compared to the baseline. This result aligns with expectations, as larger mines 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. We interpret this as suggestive evidence for adaptation — farmers may respond to large, salient mine sites by relocating or changing crops.

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 produce considerable pollution

Finally, we examine how impacts vary by commodity types. Since our primary dataset lacks commodity information, we compile data from related sources to identify the most likely commodity (or mix) at each mine site (see [Appendix C2](#) for details). We focus on four main commodities that are well-represented in the data and for which we can reasonably exclude interference by other commodities: gold, diamonds, copper, and coal. The results reveal a particularly 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. This finding aligns with earlier research identifying gold mining as especially damaging to agriculture and the environment. Gold extraction uses toxic chemicals like mercury and sodium cyanide for extraction,¹⁹ and artisanal gold miners rarely have the capacity or incentives to mitigate their impacts. Among other commodities — diamonds (13% of the sample), copper (6%), and coal (8%) — only diamond mining shows marginally significant negative effects on overall vegetation.

These heterogeneity findings show that the impacts of mining are not uniform, but vary substantially across environmental conditions and mining practices. The concentration of effects in agriculturally suitable areas, in West Africa, and for gold mining highlights priorities for interventions and alleys for future research.

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 to understand 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 distance basins 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

¹⁹[Duncan \[2020\]](#), [Malone et al. \[2023\]](#), [Verbrugge et al. \[2021\]](#)

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.

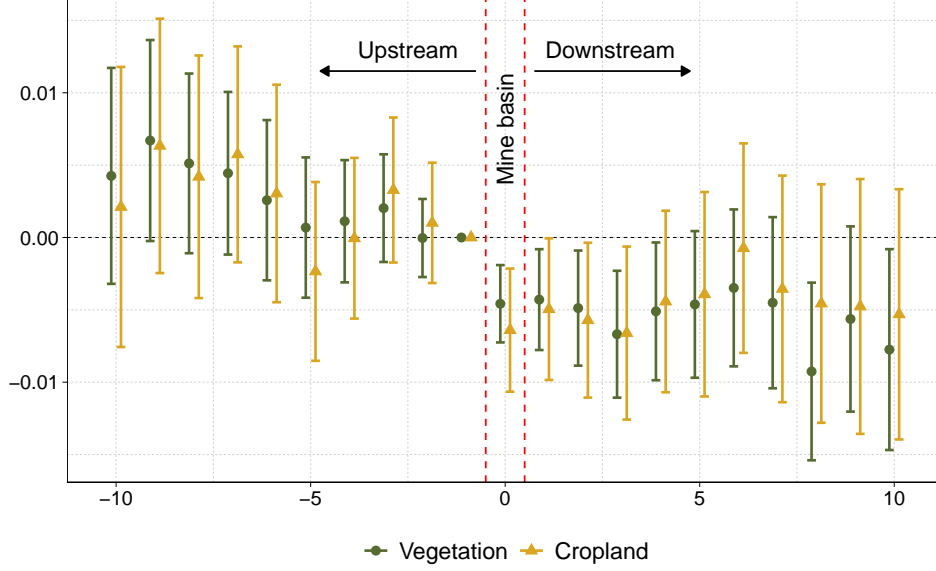


FIGURE 5: Estimated order coefficients for all up- and downstream basins (with the mine-basin in the center) for the overall peak vegetation index (green circle) and the cropland-specific index (yellow triangles) with full covariates. Whiskers represent 95%-confidence intervals.

Figure 5 visualizes estimates along basin orders, comparing all basins to the one immediately upstream of the mine (order -1).²⁰ 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

²⁰The visualization differs from the coefficients in Table 2 and Table E4 (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. This discrepancy may reflect adaptation behavior by farmers, a result that is supported by our heterogeneity analysis. 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 [also see [Macklin et al., 2023](#)]. 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 these various checks; the complete results appear in the Appendix.²¹

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

²¹Further details are provided in [Appendix C3.1](#) and [Appendix C3.2](#); [Tables E8 to E11](#) report the full estimates.

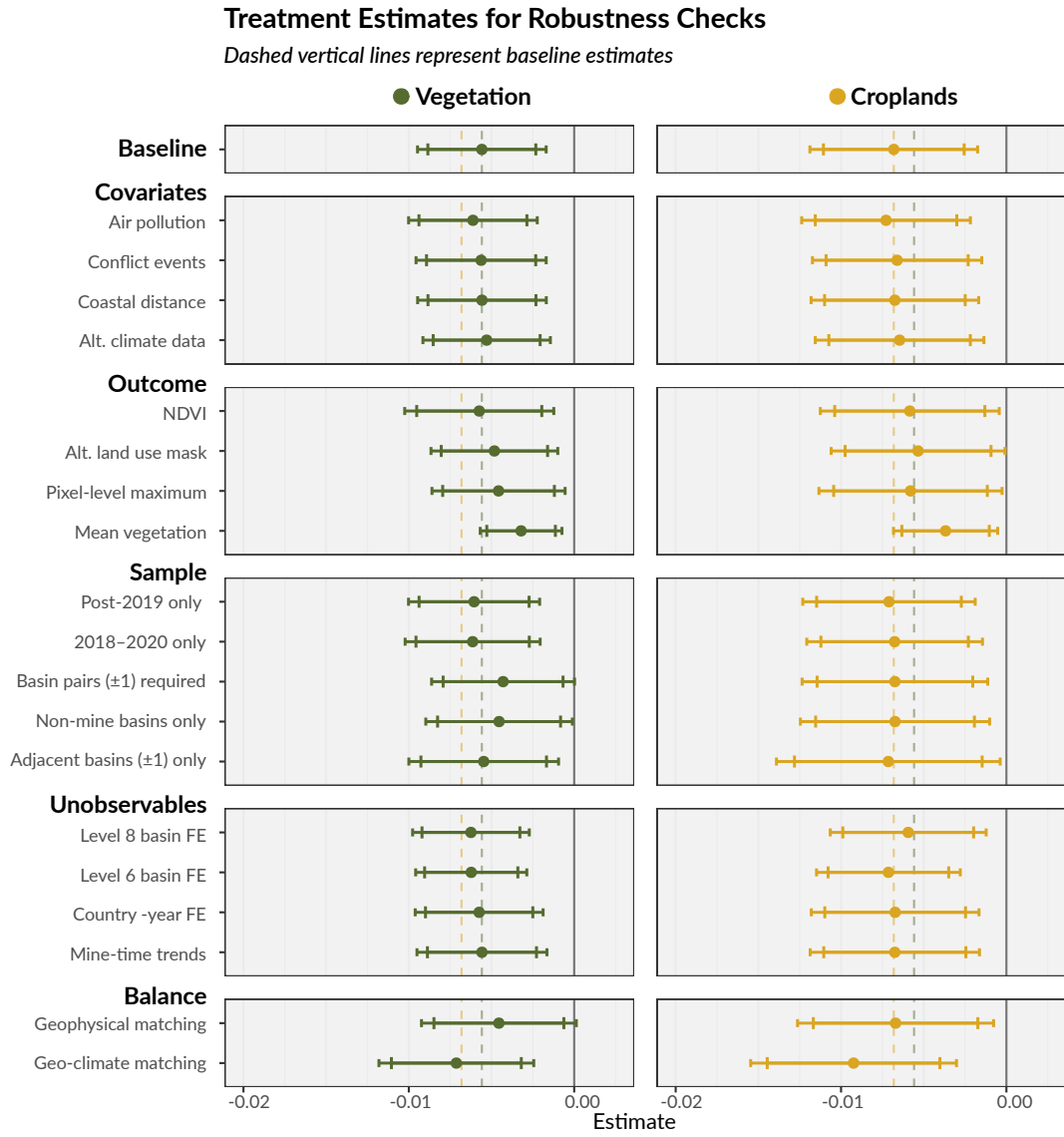


FIGURE 6: Dots indicate the average treatment effect in the first three basins; whiskered lines indicate 90% and 95% confidence intervals.

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.

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.²² 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) the post-2019 period only, and (2) a narrower 2018–2020 window. Second, we modify the spatial 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. 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 additional 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. 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 et al., 2012]. We match upstream

²²We 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].

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 D11](#) and [Figure D12](#) in the Appendix), and the resulting estimates are significant and of larger 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 D10](#) for results). Second, we use our control variables as placebo outcomes and find no discontinuities at mine locations (see [Appendix C3.3](#) for details), confirming that our identification strategy isolates mining impacts rather than pre-existing environmental or social differences.

Collectively, these robustness checks demonstrate that our findings of negative mining impacts on downstream vegetation and agriculture remain consistent for diverse specifications. This consistency strengthens confidence in our causal interpretation and the reliability of our results.

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. In this section, we discuss the mechanisms behind these impacts, how to interpret them, which limitations to be aware of, 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 by water streams, separating them from other effects of mining. Two main mechanisms help explain our findings — water pollution and human responses to it.

Water pollution

Water pollution emerges as the dominant and evident mechanism behind our observed effects. It is a major and well-documented externality of mining, and Sub-Saharan Africa has been projected to become a global hotspot for it [Jones et al., 2023]. Our main results only provide indirect evidence for this mechanism, so we additionally analyze available water quality data [from United Nations Environment Programme, 2025] to help confirm the role of water pollution in Appendix B3. The data shows that water quality is markedly lower in mine basins and downstream regions as compared to upstream basins. Key indicators — including electrical conductivity (measuring salinity), sodium discharge and adsorption, and sulfate concentrations (which can cause acidification) — show patterns that are consistent with mine-induced pollution.

Water quality data supports our interpretation, but comes with important limitations. Available samples from our observation period are exclusively from South Africa [water quality monitoring is notoriously sparse in Africa; see Jones et al., 2024], were obtained at irregular intervals, and analyzed using different methods. Critically, they lack measurements of heavy metals, which are among the most harmful mining pollutants, and specific chemicals used in ore processing. This prevents us from directly differentiating between, e.g., mercury-based or cyanide-based processing in gold mining, which could affect both the magnitude and reach of environmental impacts [Verbrugge et al., 2021].

Adaptation responses

Human adaptation represents a secondary mechanism influencing our results. Farmers may respond to water pollution in ways that could amplify or attenuate our measured impacts on vegetation and agriculture.

Evidence from related literature on air pollution suggests complex response patterns. While pollution primarily affects plant growth directly, it also affects the health, and, in turn, labor supply and productivity of farmers.²³ These impacts can further affect income, the allocation of land and labor, and other relevant factors. The final result of these effects is ambiguous, but their magnitude is likely

²³Aragón and Rud [2015], Fugiel et al. [2017], Kotsadam and Tolonen [2016], Miao et al. [2017]

constrained in our context, where little fertilizer is used, and significant market frictions exist that deter migration [Chen et al., 2022].

In the case of directed water pollution, the adaptation of farmers plays an important, immediate role. Farmers could shift towards cultivating crops with greater resilience to water pollution — substituting, for instance, lentil and maize for barley and millet, which tolerate higher levels of salinity [Page et al., 2021]. Agricultural practices present another opportunity for adaptation. Importantly, farmers may reduce their reliance on irrigation (or decide against transitioning to it in the first place) if water is contaminated.²⁴ We do not directly capture adaptation behaviors, but they represent important avenues for future research and opportunities for mitigation.

5.2. Interpretation

Our analysis reveals mines’ considerable downstream impacts on vegetation and agriculture, which are mediated by water pollution and adaptation, and thus likely to accumulate and persist for many years. We find that the peak cropland vegetation index is reduced by 1.38–1.47% of the sample mean. This effect size aligns with other studies that use vegetation indices to approximate agricultural productivity. For instance, Wuepper and Finger [2022] also use the peak annual EVI to proxy for yields in 1 km² grid-cells and find a 2.2% change in yields per step on their ten-step index of institutions. Similarly, Asher and Novosad [2020] proxy agricultural yields via the EVI, and find a 1.7 percent increase from the construction of new roads.

Translating these results into agricultural terms, we estimate that being immediately downstream of a mine is associated with at least a 0.57–0.61% decrease in cereal yields, a 1.59–1.70% decrease in the value of cereal production, and a 2.16–2.31% decrease in the value of overall crop production. More than 74,000 km² of smallholder croplands are immediately affected (i.e., up to around 33 km downstream of mine basins). This implies annual losses of over 91,000 tons of harvested cereal crops, although our evidence suggests that effects may reach much farther downstream, affecting considerably larger areas.

Importantly, our results only reflect one specific pathway among the many externalities of mines. Agriculture and vegetation are also affected by other mining-related

²⁴While rainfed agriculture is not unaffected (water pollution affects groundwater and soil), impacts are likely attenuated through the water cycle.

factors, such as air pollution [Fugiel et al., 2017], which our research design does not capture. Meanwhile, water pollution from mines creates numerous other negative effects, including water scarcity, biodiversity loss, and detrimental impacts on human health and cognition [Gittard and Hu, 2024, Van Vliet et al., 2017, Vörösmarty et al., 2010].

While our findings must be understood within this wider context, we have reason to believe that our estimates provide a lower bound for the specific water-mediated pathway. First, our results reflect post-adaptation outcomes, incorporating mitigating behaviors that farmers adopt in response to pollution. Second, the limited use of irrigation means that estimates are attenuated and the full extent of impacts does not emerge in practice. Third, our analysis is only focused on the immediate vicinity of mines, where our research design allows for strong causal claims. Finally, various sources of noise in our data represent a major limitation that likely leads to conservative estimates of true impacts.

5.3. Limitations

Our analysis faces limitations that warrant discussion. Excess uncertainty from noisy measurements is an important example, with reliance on remotely sensed vegetation indices as a first source. While they offer consistent and highly granular measures of vegetation health that enable our analysis, they come with inherent constraints. Our peak vegetation indices perform well as a proxy, but do not capture all relevant phases of crop development. Robustness checks with alternative measures soothe pressing concerns, but cannot convey a full picture. Singular metrics cannot fully capture the heterogeneity of impacts across different crop types, soil types, and climatic conditions [Bolton and Friedl, 2013], and there is limited data for calibration. While we relate our outcome measure to local agricultural data, this introduces additional noise and likely results in attenuated estimates. Future research with a narrower geographical focus could address this limitation by combining remote sensing with more accurate and detailed agricultural data.

A second limitation concerns the information available about mining operations. While we use a comprehensive dataset of mine sites with verified locations and their surroundings (within a ten-kilometer radius), these data have constraints [see Maus et al., 2022, Maus and Werner, 2024]. Much relevant information on mining sites — including their status, activity, production methods, and commodities mined —

is not directly available. We address this limitation partially by assessing mining area growth over time and adding commodity information through supplementary sources. However, we cannot disentangle artisanal and industrial mining operations, which likely differ in management, processing, and impacts [Girard et al., 2024], and lack detailed information to narrow down relevant mechanisms.

Lastly, our study only offers suggestive evidence for specific mechanisms and the total range and impact of observed effects. Data on water quality and pollution are scarce, and the available information is geographically patchy and inconsistent. We investigate heterogeneity for different subsets, e.g., based on commodity types, location, and mine growth, to infer potential mechanisms, but do not provide direct evidence for specific mining activities impacting vegetation. Meanwhile, our research design provides the conditions for identifying local treatment effects, but cannot provide causal evidence for their reach. More in-depth analysis of mechanisms and the reach of effects is left to future work.

5.4. Outlook

Our results have important implications for future research and policy. Two key goals are (1) overcoming data scarcity, which constrains both scientific understanding and policy responses, and (2) implementing targeted interventions at various governance levels to mitigate the external costs of mining.

Data scarcity

A common theme across our analysis is the scarcity and poor quality of available data. Addressing this gap is paramount, as improved information can advance research and enable local policymakers to design effective mitigation strategies. Agricultural statistics are challenging to collect for smallholder and subsistence farming, which limits our understanding of impacts at the farm level. Information on mining sites remains inadequate and poorly accessible [Maus and Werner, 2024], especially for artisanal mines.²⁵ Pollution data are essential to assess impacts,

²⁵For artisanal mines in particular, data needs to be disseminated in a way that safeguards against potential human rights abuses. Two examples for such data repositories are the International Peace Information Service that collects data on artisanal mining in the Central African Republic, Tanzania, Zimbabwe, and parts of the DRC, as well as the Revenue Development Foundation that supports governments in managing their natural resources and collates data on mining sites in several West African countries.

but similarly scarce [Jones et al., 2024]. Where institutional capacities are lacking, community-based water monitoring could serve as an inexpensive and effective complement [Ruppen et al., 2021].

Remote sensing technologies play a crucial role, with considerable future potential to further refine and extend new and existing data. For water pollution, satellite imagery has already detected large spills or tailings dam failures [Rudorff et al., 2018, Ruppen et al., 2023], and refined approaches may help identify heavy metal pollution [Swain and Sahoo, 2017]. The monitoring of mines represents another application [Maus et al., 2022] with considerable potential for automating laborious tasks. For instance, Sepin et al. [2025] use machine learning methods to automatically map the evolution of mining sites over time. Vegetation indices already provide valuable proxies for agricultural productivity — future work could integrate high-quality yield data into ready-made analytical products. Beyond satellite-based imaging, unmanned aerial vehicles have developed rapidly, and can offer higher-resolution data for monitoring critical locations.

Interventions

Policymakers can and should also address the external costs of mining today. Mining concessions should explicitly recognize these externalities, and should factor in their impacts on agriculture, which is vital for local economies and food security. Our results provide concrete evidence for including water-mediated impacts in the allocation of concessions. The formalization of artisanal mining, for example through official titling, could increase miners’ incentives to invest in precautionary equipment. This could reduce impacts at the source — with economic as well as health benefits for humans and their environment.

Supranational interventions are essential in effectively addressing mining externalities. These can help shift regulatory and monitoring burdens from mineral-producing countries towards all that benefit from the extracted resources. For industrial mines, the Global Industry Standard on Tailings Management (GISTM) presents an important first step for international governance. Our findings highlight that not only catastrophic failures but also persistent water pollution requires attention. Similarly, the Minamata Convention on Mercury targets mercury pollution, which is still used in artisanal gold mining — the commodity we found to have the strongest

downstream impact. Going forward, supply chain measures could also help reduce the prevalence of minerals sourced in harmful ways.²⁶

6. Conclusion

In this paper, we identified the causal effects of mining on agricultural productivity mediated by water pollution. In a quasi-experimental research design, we used the discontinuity from mine sites along a directed network of river basins for identification. We compared differences in agricultural productivity, based on a remotely sensed peak vegetation index, up- and downstream of mine locations. Our main specification revealed a reduction of peak vegetation 1.28–1.35% for general vegetation and 1.38–1.47% for croplands immediately downstream of mines. This corresponds to annual losses of 91,000 tons of cereals across 74,000 km² of affected croplands.

Our results can be contextualized both quantitatively and in terms of policy. Crucially, our estimates only reflect a specific part of the total external costs of mining for agriculture. Our research design does not capture impacts that are not mediated alongside rivers, such as air pollution or local labor markets. Furthermore, our estimates of the water-mediated effect may be attenuated by our indirect measurement of agricultural productivity. Despite these limitations, our findings inform the discussion about resource extraction in Africa, particularly in regions with weak environmental governance. The documented effects highlight the need for interventions that reduce negative impacts of mining on water systems. Proper containment facilities, for instance, should be required for industrial mining operations but also for the informal mining sector, and especially in gold-mining regions, for which we found particularly strong impacts. Enhanced monitoring of mines and surface water quality is necessary to address 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 the mine dataset we use enables a comprehensive analysis across Africa, it lacks detailed information on individual mine characteristics. Future studies that incorporate data on containment facilities, mine types, and pollutants produced would allow for more

²⁶The European Union’s Deforestation Regulation (EUDR), which aims to prevent the conversion of forests towards agricultural land, is a notable and recent example. It covers seven agricultural commodities that are implicated in deforestation, and applies to companies that place these products on the EU market.

precise analyses of impacts. Similarly, spatial data on crop distributions could enable detailed analyses of which crop types are most susceptible to mining-induced pollution, and could help inform adaptation strategies. Different research approaches might also address questions our design could not fully answer, such as disentangling the impacts of industrial and artisanal mines or providing stronger evidence on how effects decay with distance. Such research would further strengthen the evidence base for targeted interventions that balance the economic benefits of mining with its environmental and human costs.

References

- Hans-Otto Pörtner, Debra C. Roberts, Melinda M.B. Tignor, Elvira Poloczanska, Katja Mintenbeck, Andrés Alegría, Marlies Craig, Stefanie Langsdorf, Sina Löschke, Vincent Möller, Andrew Okem, and Bardhyl Rama. *Climate Change 2022: Impacts, Adaptation and Vulnerability*. Cambridge University Press, Cambridge, August 2022. URL <https://www.ipcc.ch/report/ar6/wg2>. [Online; accessed 15. Aug. 2022].
- Bora Aska, Daniel M. Franks, Martin Stringer, and Laura J. Sonter. Biodiversity conservation threatened by global mining wastes. *Nature Sustainability*, 7:23–30, January 2024. ISSN 2398-9629. doi: 10.1038/s41893-023-01251-0.
- Nicolas Berman, Mathieu Couttenier, Dominic Rohner, and Mathias Thoenig. This mine is mine! how minerals fuel conflicts in Africa. *American Economic Review*, 107(6):1564–1610, 2017. ISSN 0002-8282. doi: 10.1257/aer.20150774.
- Nicolas Berman, Mathieu Couttenier, and Victoire Girard. Mineral resources and the salience of ethnic identities. *Economic Journal*, 133(653):1705–1737, 2023. ISSN 0013-0133. doi: 10.1093/ej/uead018.
- Stefan Giljum, Victor Maus, Nikolas Kuschnig, Sebastian Luckeneder, Michael Tost, Laura J. Sonter, and Anthony J. Bebbington. A pantropical assessment of deforestation caused by industrial mining. *Proceedings of the National Academy of Sciences*, 119(38):e2118273119, 2022. doi: 10.1073/pnas.2118273119.
- Victoire Girard, Teresa Molina-Millán, and Guillaume Vic. Artisanal mining in Africa. Green for gold? *Working Paper*, 2024.

- Carl Henrik Knutsen, Andreas Kotsadam, Eivind Hammersmark Olsen, and Tore Wig. Mining and local corruption in Africa. 61(2):320–334. ISSN 1540-5907. doi: 10.1111/ajps.12268.
- M. G. Macklin, C. J. Thomas, A. Mudbhatal, P. A. Brewer, K. A. Hudson-Edwards, J. Lewin, P. Scussolini, D. Eilander, A. Lechner, J. Owen, G. Bird, D. Kemp, and K. R. Mangalaa. Impacts of metal mining on river systems: A global assessment. *Science*, 381(6664):1345–1350, 2023. ISSN 0036-8075. doi: 10.1126/science.adg6704.
- Caroline S. Santana, Diango M. Montalván Olivares, Vinnícius H. C. Silva, Francisco H. M. Luzardo, Fermin G. Velasco, and Raildo M. de Jesus. Assessment of water resources pollution associated with mining activity in a semi-arid region. *Journal of Environmental Management*, 273:111148, 2020. ISSN 0301-4797. doi: 10.1016/j.jenvman.2020.111148.
- Mélanie Gittard and Irène Hu. MiningLeaks Water Pollution and Child Mortality in Africa, 2024. URL <https://shs.hal.science/halshs-04685390v1>. [Online; accessed 13. Feb. 2025].
- H. Sigman. International spillovers and water quality in rivers: Do countries free ride? *American Economic Review*, 92(4):1152–1159, 2002. ISSN 0002-8282. doi: 10.1257/00028280260344687.
- Molly Lipscomb and Ahmed Mushfiq Mobarak. Decentralization and pollution spillovers: Evidence from the re-drawing of county borders in Brazil. *Review of Economic Studies*, 84(1):464–502, January 2017. ISSN 0034-6527. doi: 10.1093/restud/rdw023.
- Mateus Dias, Rudi Rocha, and Rodrigo R. Soares. Down the river: Glyphosate use in agriculture and birth outcomes of surrounding populations. *Review of Economic Studies*, 90(6):2943–2981, February 2023. ISSN 1467-937X. doi: 10.1093/restud/rdad011.
- Eric Strobl and Robert O. Strobl. The distributional impact of large dams: Evidence from cropland productivity in Africa. *Journal of Development Economics*, 96(2): 432–450, November 2011. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2010.08.005.

- Fernando M. Aragón and Juan Pablo Rud. Polluting industries and agricultural productivity: Evidence from mining in Ghana. *The Economic Journal*, 126(597): 1980–2011, September 2015. ISSN 0013-0133. doi: 10.1111/ecoj.12244.
- George Ofori, Andreas Dittmann, David Sarpong, and David Botchie. Socio-economic and environmental implications of artisanal and small-scale mining (asm) on agriculture and livelihoods. *Environmental Science & Policy*, 106: 210–220, 2020. ISSN 1462-9011. doi: 10.1016/j.envsci.2020.02.005.
- Jan von der Goltz and Prabhat Barnwal. Mines: The local wealth and health effects of mineral mining in developing countries. *Journal of Development Economics*, 139:1–16, June 2019. ISSN 0304-3878. doi: 10.1016/j.jdeveco.2018.05.005.
- Alfred Awotwi, Geophrey K. Anornu, Jonathan Arthur Quaye-Ballard, Thompson Annor, Isaac Kwadwo Nti, Samuel N. Odai, Emmanuel Arhin, and Charles Gyamfi. 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. *Soil and Tillage Research*, 211:105026, 2021. ISSN 0167-1987. doi: 10.1016/j.still.2021.105026.
- Longxu Du, Zhiyu Zhang, Yanqiu Chen, Yue Wang, Chengxiang Zhou, Huaiyu Yang, and Wei Zhang. Heterogeneous impact of soil acidification on crop yield reduction and its regulatory variables: A global meta-analysis. *Field Crops Research*, 319: 109643, 2024. ISSN 0378-4290. doi: 10.1016/j.fcr.2024.109643.
- Albert Ebo Duncan. The dangerous couple: Illegal mining and water pollution—a case study in Fena river in the Ashanti region of Ghana. *Journal of Chemistry*, 2020:1–9, 2020. ISSN 2090-9071. doi: 10.1155/2020/2378560. URL <http://dx.doi.org/10.1155/2020/2378560>.
- Mary Mulenga, Kennedy O. Ouma, Concillia Monde, and Stephen Syampungani. Aquatic mercury pollution from artisanal and small-scale gold mining in Sub-Saharan Africa: status, impacts, and interventions. *Water*, 16(5):756, March 2024. ISSN 2073-4441. doi: 10.3390/w16050756. URL <http://dx.doi.org/10.3390/w16050756>.
- Lijun Wu, Weifeng Yue, Jin Wu, Changming Cao, Hong Liu, and Yanguo Teng. Metal-mining-induced sediment pollution presents a potential ecological risk and

- threat to human health across China: A meta-analysis. *Journal of Environmental Management*, 329:117058, 2023. ISSN 0301-4797. doi: 10.1016/j.jenvman.2022.117058.
- S. M. Olmstead. The economics of water quality. *Review of Environmental Economics and Policy*, 2010. doi: 10.1093/reep/rep016.
- B. L. Keeler, S. Polasky, K. A. Brauman, K. A. Johnson, J. C. Finlay, A. O’Neill, K. Kovacs, and B. Dalzell. Linking water quality and well-being for improved assessment and valuation of ecosystem services. *Proceedings of the National Academy of Sciences*, 109(45):18619–18624, 2012. doi: 10.1073/pnas.1215991109.
- Michelle TH Van Vliet, Martina Flörke, and Yoshihide Wada. Quality matters for water scarcity. *Nature Geoscience*, 10(11):800–802, 2017. doi: 10.1038/ngeo3047.
- Edward R Jones, Marc FP Bierkens, Peter JTM van Puijenbroek, Ludovicus (Rens) PH van Beek, Niko Wanders, Edwin H Sutanudjaja, and Michelle TH van Vliet. Sub-Saharan Africa will increasingly become the dominant hotspot of surface water pollution. *Nature Water*, 1(7):602–613, 2023. doi: 10.1038/s44221-023-00105-5.
- Victor Maus and Tim Werner. Impacts for half of the world’s mining areas are undocumented. *Nature*, 625:26–29, 2024. doi: 10.1038/d41586-023-04090-3.
- Spencer Banzhaf, Lala Ma, and Christopher Timmins. Environmental justice: The economics of race, place, and pollution. *Journal of Economic Perspectives*, 33(1):185–208, February 2019. ISSN 0895-3309. doi: 10.1257/jep.33.1.185. URL <http://dx.doi.org/10.1257/jep.33.1.185>.
- Graeme Auld, Michele Betsill, and Stacy D. VanDeveer. Transnational governance for mining and the mineral lifecycle. *Annual Review of Environment and Resources*, 43:425–453, October 2018. doi: 10.1146/annurev-environ-102017-030223.
- Edward R Jones, Duncan J Graham, Ann van Griensven, Martina Flörke, and Michelle T H van Vliet. Blind spots in global water quality monitoring. *Environmental Research Letters*, 19(9):091001, 2024. ISSN 1748-9326. doi: 10.1088/1748-9326/ad6919. URL <http://dx.doi.org/10.1088/1748-9326/ad6919>.

- International Council on Mining and Metals. Mining contribution index (mci), 2022. URL https://www.icmm.com/website/publications/pdfs/social-performance/2022/research_mci-6-ed.pdf?cb=16134.
- Remi Bazillier and Victoire Girard. The gold digger and the machine. evidence on the distributive effect of the artisanal and industrial gold rushes in Burkina Faso. *Journal of Development Economics*, 143:102411, March 2020. ISSN 0304-3878. doi: 10.1016/j.jdevco.2019.102411.
- Melanie Gräser. Industrial versus artisanal mining: The effects on local employment in Liberia. *Journal of Rural Studies*, 111:103389, 2024. ISSN 0743-0167. doi: 10.1016/j.jrurstud.2024.103389.
- Abigail Barenblitt, Amanda Payton, David Lagomasino, Lola Fatoyinbo, Kofi Asare, Kenneth Aidoo, Hugo Pigott, Charles Kofi Som, Laurent Smeets, Omar Seidu, and Danielle Wood. The large footprint of small-scale artisanal gold mining in Ghana. *Science of The Total Environment*, 781:146644, 2021. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2021.146644.
- D. Barrie Johnson and Kevin B. Hallberg. Acid mine drainage remediation options: A review. *Science of The Total Environment*, 338(1):3–14, 2005. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2004.09.002.
- Aaron Malone, Linda Figueroa, Weishi Wang, Nicole M. Smith, James F. Ranville, David C. Vuono, Francisco D. Alejo Zapata, Lino Morales Paredes, Jonathan O. Sharp, and Christopher Bellona. Transitional dynamics from mercury to cyanide-based processing in artisanal and small-scale gold mining: Social, economic, geochemical, and environmental considerations. *Science of The Total Environment*, 898:165492, 2023. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2023.165492.
- Boris Verbrugge, Cristiano Lanzano, and Matthew Libassi. The cyanide revolution: Efficiency gains and exclusion in artisanal- and small-scale gold mining. *Geoforum*, 126:267–276, 2021. ISSN 0016-7185. doi: 10.1016/j.geoforum.2021.07.030.
- René P. Schwarzenbach, Thomas Egli, Thomas B. Hofstetter, Urs von Gunten, and Bernhard Wehrli. Global water pollution and human health. *Annual Review of*

Environment and Resources, (Volume 35, 2010):109–136, November 2010. doi: 10.1146/annurev-environ-100809-125342.

Aline Frossard, Jonathan Donhauser, Adrien Mestrot, Sebastien Gyga, Erland Bååth, and Beat Frey. Long- and short-term effects of mercury pollution on the soil microbiome. *Soil Biology and Biochemistry*, 120:191–199, May 2018. ISSN 0038-0717. doi: 10.1016/j.soilbio.2018.01.028. URL <http://dx.doi.org/10.1016/j.soilbio.2018.01.028>.

Jason Russ, Esha Zaveri, Richard Damania, Sébastien Desbureaux, Jorge Ecurra, and Aude-Sophie Rodella. Salt of the Earth: Quantifying the impact of water salinity on global agricultural productivity. Policy Research Working Paper 9144, World Bank, 2020. URL <https://openknowledge.worldbank.org/server/api/core/bitstreams/78b9fa85-83e6-549b-9c2c-d33099c0bc50/content>. Public Disclosure Authorized.

C. Zörb, C.-M. Geilfus, and K.-J. Dietz. Salinity and crop yield. *Plant Biology*, 21 (S1):31–38, 2019. ISSN 1435-8603. doi: 10.1111/plb.12884.

Agata Fugiel, Dorota Burchart-Korol, Krystyna Czaplicka-Kolarz, and Adam Smoliński. Environmental impact and damage categories caused by air pollution emissions from mining and quarrying sectors of European countries. *Journal of Cleaner Production*, 143:159–168, February 2017. ISSN 0959-6526. doi: 10.1016/j.jclepro.2016.12.136.

Weijie Miao, Xin Huang, and Yu Song. An economic assessment of the health effects and crop yield losses caused by air pollution in mainland china. *Journal of Environmental Sciences*, 56:102–113, June 2017. ISSN 1001-0742. doi: 10.1016/j.jes.2016.08.024.

Bhanu Pandey, Madhoolika Agrawal, and Siddharth Singh. Assessment of air pollution around coal mining area: Emphasizing on spatial distributions, seasonal variations and heavy metals, using cluster and principal component analysis. *Atmospheric Pollution Research*, 5(1):79–86, January 2014. ISSN 1309-1042. doi: 10.5094/APR.2014.010.

- David Wuepper, Haoyu Wang, Wolfram Schlenker, Meha Jain, and Robert Finger. Institutions and global crop yields. *NBER Working Paper*, July 2023. doi: 10.3386/w31426.
- Rema Hanna and Paulina Oliva. The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City. *Journal of Public Economics*, 122: 68–79, 2015. ISSN 0047-2727. doi: 10.1016/j.jpubeco.2014.10.004.
- Joshua Graff Zivin and Matthew Neidell. The impact of pollution on worker productivity. *American Economic Review*, 102(7):3652–73, 2012. ISSN 0002-8282. doi: 10.1257/aer.102.7.3652.
- Janet Currie, Eric A. Hanushek, E. Megan Kahn, Matthew Neidell, and Steven G. Rivkin. Does pollution increase school absences? *Review of Economics and Statistics*, 91(4):682–694, 2009. ISSN 0034-6535. doi: 10.1162/rest.91.4.682.
- Andreas Kotsadam and Anja Tolonen. African mining, gender, and local employment. *World Development*, 83:325–339, July 2016. ISSN 0305-750X. doi: 10.1016/j.worlddev.2016.01.007.
- UNEP. Global industry standard on tailings management. Technical report, United Nations Environment Programme, 2023. URL <https://www.unep.org/resources/report/global-industry-standard-tailings-management>.
- Stéphane Audry, Jörg Schäfer, Gérard Blanc, and Jean-Marie Jouanneau. Fifty-year sedimentary record of heavy metal pollution (Cd, Zn, Cu, Pb) in the Lot river reservoirs. *Environmental Pollution*, 132(3):413–426, December 2004. ISSN 0269-7491. doi: 10.1016/j.envpol.2004.05.025.
- Bernhard Lehner and Günther Grill. Global river hydrography and network routing: Baseline data and new approaches to study the world’s large river systems. *Hydrological Processes*, 27(15):2171—2186, April 2013. ISSN 1099-1085. doi: 10.1002/hyp.9740.
- Hao Shi, Longhui Li, Derek Eamus, Alfredo Huete, James Cleverly, Xin Tian, Qiang Yu, Shaoqiang Wang, Leonardo Montagnani, Vincenzo Magliulo, Eyal Rotenberg, Marian Pavelka, and Arnaud Carrara. Assessing the ability of MODIS EVI to estimate terrestrial ecosystem gross primary production of multiple land cover

- types. *Ecological Indicators*, 72:153—164, January 2017. ISSN 1470-160X. doi: 10.1016/j.ecolind.2016.08.022.
- David M. Johnson. A comprehensive assessment of the correlations between field crop yields and commonly used MODIS products. *International Journal of Applied Earth Observation and Geoinformation*, 52:65—81, October 2016. ISSN 1569-8432. doi: 10.1016/j.jag.2016.05.010.
- Sam Asher and Paul Novosad. Rural roads and local economic development. *American Economic Review*, 110(3):797—823, March 2020. ISSN 0002-8282. doi: 10.1257/aer.20180268.
- Sumit Agarwal, Mohit Desai, Pulak Ghosh, and Nishant Vats. Bridging the information gap: Sowing the seeds of productivity with high-speed 4G internet. *Working Paper*, 2024(16), 2024. doi: 10.2139/ssrn.4805486.
- Yelu Zeng, Dalei Hao, Alfredo Huete, Benjamin Dechant, Joe Berry, Jing M. Chen, Joanna Joiner, Christian Frankenberg, Ben Bond-Lamberty, Youngryel Ryu, Jingfeng Xiao, Ghassem R. Asrar, and Min Chen. Optical vegetation indices for monitoring terrestrial ecosystems globally. *Nature Reviews Earth & Environment*, 3:477–493, July 2022. ISSN 2662-138X. doi: 10.1038/s43017-022-00298-5.
- Xiang Gao, Alfredo R. Huete, Wenge Ni, and Tomoaki Miura. Optical–biophysical relationships of vegetation spectra without background contamination. *Remote Sensing of Environment*, 74(3):609–620, December 2000. ISSN 0034-4257. doi: 10.1016/S0034-4257(00)00150-4.
- A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1):195–213, November 2002. ISSN 0034-4257. doi: 10.1016/S0034-4257(02)00096-2.
- Kamel Didan. MOD13Q1 MODIS/terra vegetation indices 16-day L3 global 250m SIN grid v006, 2015. URL <https://lpdaac.usgs.gov/products/mod13q1v006/>.
- Pierre Defourny, Celine Lamarche, Audric Bos, Carsten Brockmann, Martin Boettcher, and Grit Kirches. Product user guide and specification CDR and ICDR Sentinel-3 Land Cover (v2.1.1). Technical report, UCLouvain / European Space Agency, 2024.

- George Azzari, Meha Jain, and David B. Lobell. Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. *Remote Sensing of Environment*, 202:129—141, December 2017. ISSN 0034-4257. doi: 10.1016/j.rse.2017.04.014.
- Digital Earth Africa. Cropland extent maps for Africa. Technical report, DE Africa Services, 2022. URL https://docs.digitalearthafrika.org/en/latest/data_specs/Cropland_extent_specs.html. Dataset.
- IFPRI. ARENA’s DHS-GIS database, 2020.
- Victor Maus, Stefan Giljum, Dieison M da Silva, Jakob Gutschlhofer, Robson P da Rosa, Sebastian Luckeneder, Sidnei LB Gass, Mirko Lieber, and Ian McCallum. An update on global mining land use. *Scientific data*, 9(1):1–11, 2022.
- Philipp Sepin, Lukas Vashold, and Nikolas Kuschnig. Mapping mining areas in the tropics from 2016–2024, 2025. URL https://www.kuschnig.eu/files/wp_mapping-mines_wip.pdf.
- S&P Global Market Intelligence. Metals & mining database, 2025. URL <https://www.spglobal.com/marketintelligence/en/campaigns/metals-mining>.
- Simon Jasansky, Mirko Lieber, Stefan Giljum, and Victor Maus. An open database on global coal and metal mine production. *Scientific Data*, 10(52):1–12, 2023. ISSN 2052-4463. doi: 10.1038/s41597-023-01965-y.
- A.J. Padilla, D. Otarod, S.W. Deloach-Overton, R.F. Kemna, P.A. Freeman, E.R. Wolfe, L.R. Bird, A.L. Gulley, M.H. Trippi, C.L. Dicken, J.M. Hammarstrom, and A.S. Brioché. Compilation of geospatial data (GIS) for the mineral industries and related infrastructure of Africa. data release, U.S. Geological Survey, 2021.
- Matias D. Cattaneo, Nicolás Idrobo, and Rocío Titiunik. *A practical introduction to regression discontinuity designs: Foundations*. Cambridge University Press, November 2019. ISBN 9781108710206. doi: 10.1017/9781108684606.
- A. Gelman and G. Imbens. Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 2019. doi: 10.1080/07350015.2017.1366909.

- G. Amatulli, S. Domisch, M.-N. Tuanmu, B. Parmentier, A. Ranipeta, J. Malczyk, and W. Jetz. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific Data*, 5(180040):1–15, 2018. ISSN 2052-4463. doi: 10.1038/sdata.2018.40.
- T. Hengl, J. M. de Jesus, G. B. M. Heuvelink, M. R. Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, M. N. Wright, X. Geng, B. Bauer-Marschallinger, M. A. Guevara, R. Vargas, R. A. MacMillan, N. H. Batjes, J. G. B. Leenaars, E. Ribeiro, I. Wheeler, S. Mantel, and B. Kempen. SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, 12(2):e0169748, 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0169748.
- Chris Funk, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shrad-dhanand Shukla, Gregory Husak, James Rowland, Laura Harrison, Andrew Hoell, and Joel Michaelsen. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, 2(150066): 1–21, December 2015. ISSN 2052-4463. doi: 10.1038/sdata.2015.66.
- Tufa Dinku, Chris Funk, Pete Peterson, Ross Maidment, Tsegaye Tadesse, Hussein Gadain, and Pietro Ceccato. Validation of the CHIRPS satellite rainfall estimates over eastern Africa. *Quarterly Journal of the Royal Meteorological Society*, 144 (S1):292–312, November 2018. ISSN 0035-9009. doi: 10.1002/qj.3244.
- John T. Abatzoglou, Solomon Z. Dobrowski, Sean A. Parks, and Katherine C. Hegewisch. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, 5(170191):1–12, January 2018. ISSN 2052-4463. doi: 10.1038/sdata.2017.191.
- Ian Harris, Timothy J. Osborn, Phil Jones, and David Lister. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, 7(109):1–18, April 2020. ISSN 2052-4463. doi: 10.1038/s41597-020-0453-3.
- WorldPop. Global 1km population. Dataset.
- D. J. Weiss, A. Nelson, H. S. Gibson, W. Temperley, S. Peedell, A. Lieber, M. Hancher, E. Poyart, S. Belchior, N. Fullman, B. Mappin, U. Dalrymple, J. Rozier, T. C. D. Lucas, R. E. Howes, L. S. Tusting, S. Y. Kang, E. Cameron, D. Bisanzio, K. E. Battle, S. Bhatt, and P. W. Gething. A global map of travel

- time to cities to assess inequalities in accessibility in 2015. *Nature*, 553:333–336, 2018. ISSN 1476-4687. doi: 10.1038/nature25181.
- Clionadh Raleigh, Rew Linke, Håvard Hegre, and Joakim Karlsen. Introducing ACLED: An Armed Conflict Location and Event Dataset. *Journal of Peace Research*, 47(5):651–660, September 2010. ISSN 0022-3433. doi: 10.1177/0022343310378914.
- Siyuan Shen, Chi Li, Aaron van Donkelaar, Nathan Jacobs, Chenguang Wang, and Randall V. Martin. Enhancing Global Estimation of Fine Particulate Matter Concentrations by Including Geophysical a Priori Information in Deep Learning. *ACS ES&T Air*, 1(5):332–345, May 2024. doi: 10.1021/acsestair.3c00054.
- E. Dinerstein, D. Olson, A. Joshi, C. Vynne, N. D. Burgess, E. Wikramanayake, N. Hahn, S. Palminteri, P. Hedao, R. Noss, M. Hansen, H. Locke, E. C. Ellis, B. Jones, C. V. Barber, R. Hayes, C. Kormos, V. Martin, E. Crist, W. Sechrest, L. Price, J. E. M. Baillie, D. Weeden, K. Suckling, C. Davis, N. Sizer, R. Moore, D. Thau, T. Birch, P. Potapov, S. Turubanova, A. Tyukavina, N. de Souza, L. Pintea, J. C. Brito, O. A. Llewellyn, A. G. Miller, A. Patzelt, S. A. Ghazanfar, J. Timberlake, H. Klöser, Y. Shennan-Farpón, R. Kindt, J.-P. B. Lillesø, P. van Breugel, L. Graudal, M. Voge, K. F. Al-Shammari, and M. Saleem. An ecoregion-based approach to protecting half the terrestrial realm. *Bioscience*, 67(6):534–545, 2017. ISSN 0006-3568. doi: 10.1093/biosci/bix014.
- G. Fischer, F.O. Nachtergaele, H.T. van Velthuisen, F. Chiozza, G. Franceschini, M. Henry, D. Muchoney, and S. Tramberend. Global Agro-Ecological Zones (GAEZ v4) model documentation. Technical report, Food and Agriculture Organization (FAO) of the United Nations and International Institute for Applied System (IIASA), 2021.
- S. M. Iacus, G. King, and G. Porro. Causal inference without balance checking: Coarsened Exact Matching. *Political Analysis*, 20(1):1–24, 2012. ISSN 1047-1987. doi: 10.1093/pan/mpr013.
- United Nations Environment Programme. GEMStat database of the global environment monitoring system for freshwater (GEMS/water) Programme, 2025. Accessed 21 February 2025. Available upon request from GEMS/Water Data Centre: gemstat.org.

- Chaoran Chen, Diego Restuccia, and Raül Santaaulàlia-Llopis. The effects of land markets on resource allocation and agricultural productivity. *Review of Economic Dynamics*, 45:41—54, July 2022. ISSN 1094-2025. doi: 10.1016/j.red.2021.04.006.
- Kathryn L. Page, Yash P. Dang, Cristina Martinez, Ram C. Dalal, J. Bernhard Wehr, Peter M. Kopittke, Thomas G. Orton, and Neal W. Menzies. Review of crop-specific tolerance limits to acidity, salinity, and sodicity for seventeen cereal, pulse, and oilseed crops common to rainfed subtropical cropping systems. *Land Degradation & Development*, 32(8):2459–2480, May 2021. ISSN 1085-3278. doi: 10.1002/ldr.3915.
- David Wuepper and Robert Finger. Regression discontinuity designs in agricultural and environmental economics. *European Review of Agricultural Economics*, 50(1): 1—28, October 2022. ISSN 1464-3618. doi: 10.1093/erae/jbac023.
- Charles J Vörösmarty, Peter B McIntyre, Mark O Gessner, David Dudgeon, Alexander Prusevich, Pamela Green, Stanley Glidden, Stuart E Bunn, Caroline A Sullivan, C Reidy Liermann, et al. Global threats to human water security and river biodiversity. *Nature*, 467(7315):555–561, 2010. doi: 10.1038/nature09440.
- Douglas K. Bolton and Mark A. Friedl. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology*, 173:74—84, May 2013. ISSN 0168-1923. doi: 10.1016/j.agrformet.2013.01.007.
- Désirée Ruppen, Owen A. Chituri, Maideyi L. Meck, Numa Pfenninger, and Bernhard Wehrli. Community-based monitoring detects sources and risks of mining-related water pollution in Zimbabwe. *Frontiers in Environmental Science*, 9, December 2021. ISSN 2296-665X. doi: 10.3389/fenvs.2021.754540. URL <http://dx.doi.org/10.3389/fenvs.2021.754540>.
- Natalia Rudorff, Conrado M. Rudorff, Milton Kampel, and Gustavo Ortiz. 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. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145:349–361, November 2018. ISSN 0924-2716. doi: 10.1016/j.isprsjprs.2018.02.013.

- Désirée Ruppen, James Runnalls, Raphael M. Tshimanga, Bernhard Wehrli, and Daniel Odermatt. Optical remote sensing of large-scale water pollution in Angola and DR Congo caused by the Catoca mine tailings spill. *International Journal of Applied Earth Observation and Geoinformation*, 118:103237, April 2023. ISSN 1569-8432. doi: 10.1016/j.jag.2023.103237.
- Ratnakar Swain and Bhabagrahi Sahoo. Mapping of heavy metal pollution in river water at daily time-scale using spatio-temporal fusion of MODIS-aqua and Landsat satellite imageries. *Journal of Environmental Management*, 192:1–14, May 2017. ISSN 0301-4797. doi: 10.1016/j.jenvman.2017.01.034.
- K. L. Verdin and J. P. Verdin. A topological system for delineation and codification of the Earth’s river basins. *Journal of Hydrology*, 218(1):1–12, May 1999. ISSN 0022-1694. doi: 10.1016/S0022-1694(99)00011-6.
- Albert Ebo Duncan, Nanne de Vries, and Kwabena Biritwum Nyarko. Assessment of heavy metal pollution in the sediments of the river Pra and its tributaries. *Water, Air, & Soil Pollution*, 229(8), 2018. ISSN 1573-2932. doi: 10.1007/s11270-018-3899-6.
- Satyabrata Sahoo and Somnath Khaoash. Impact assessment of coal mining on groundwater chemistry and its quality from brajrajnagar coal mining area using indexing models. *Journal of Geochemical Exploration*, 215:106559, August 2020. ISSN 0375-6742. doi: 10.1016/j.gexplo.2020.106559. URL <http://dx.doi.org/10.1016/j.gexplo.2020.106559>.
- Darrell Kirk Nordstrom, Charles N. Alpers, Carol J. Ptacek, and David W. Blowes. Negative ph and extremely acidic mine waters from Iron Mountain, California. *Environmental Science & Technology*, 34(2):254–258, January 2000. ISSN 0013-936X. doi: 10.1021/es990646v.
- Asish Kumar Parida and Anath Bandhu Das. Salt tolerance and salinity effects on plants: A review. *Ecotoxicology and Environmental Safety*, 60(3):324–349, 2005. ISSN 0147-6513. doi: 10.1016/j.ecoenv.2004.06.010.
- R. S. Ayers and D. W. Westcot. *Water quality for agriculture*. Irrigation and Drainage Paper 29. Food and Agriculture Organization of the United Nations, Rome, Italy, 1985.

- Andrew A. Neath and Joseph E. Cavanaugh. The bayesian information criterion: Background, derivation, and applications. *WIREs Computational Statistics*, 4(2): 199–203, March 2012. ISSN 1939-5108. doi: 10.1002/wics.199.
- G. Imbens and K. Kalyanaraman. Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies*, 79(3):933–959, 2012. ISSN 0034-6527. doi: 10.1093/restud/rdr043.

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.²⁷ 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 km² 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 km² to form their own tributary sub-basin,
- areas between qualifying tributary streams from inter-basins,

²⁷We 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 km², they are subdivided using artificial break points.

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

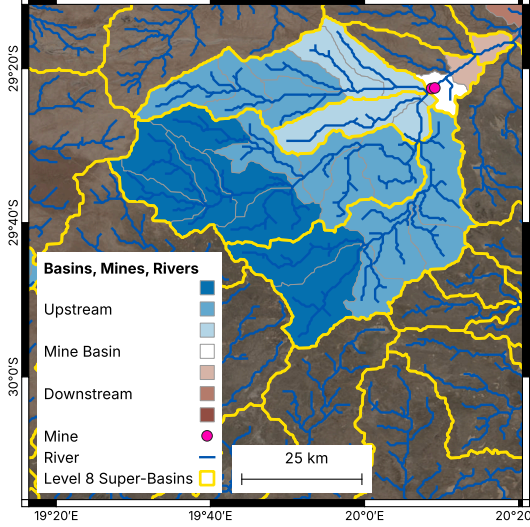


FIGURE A1: Two forked upstream (Level 12) basins join into a single (mine) basin further downstream. The superimposed yellow lines indicate Level 8 basins; these contain varying numbers of sub-basins (due to a level-skipping mechanism), and clearly divide tributary and main basins. The blue lines, which 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 [Verdin and 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 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 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.²⁸

²⁸This 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.

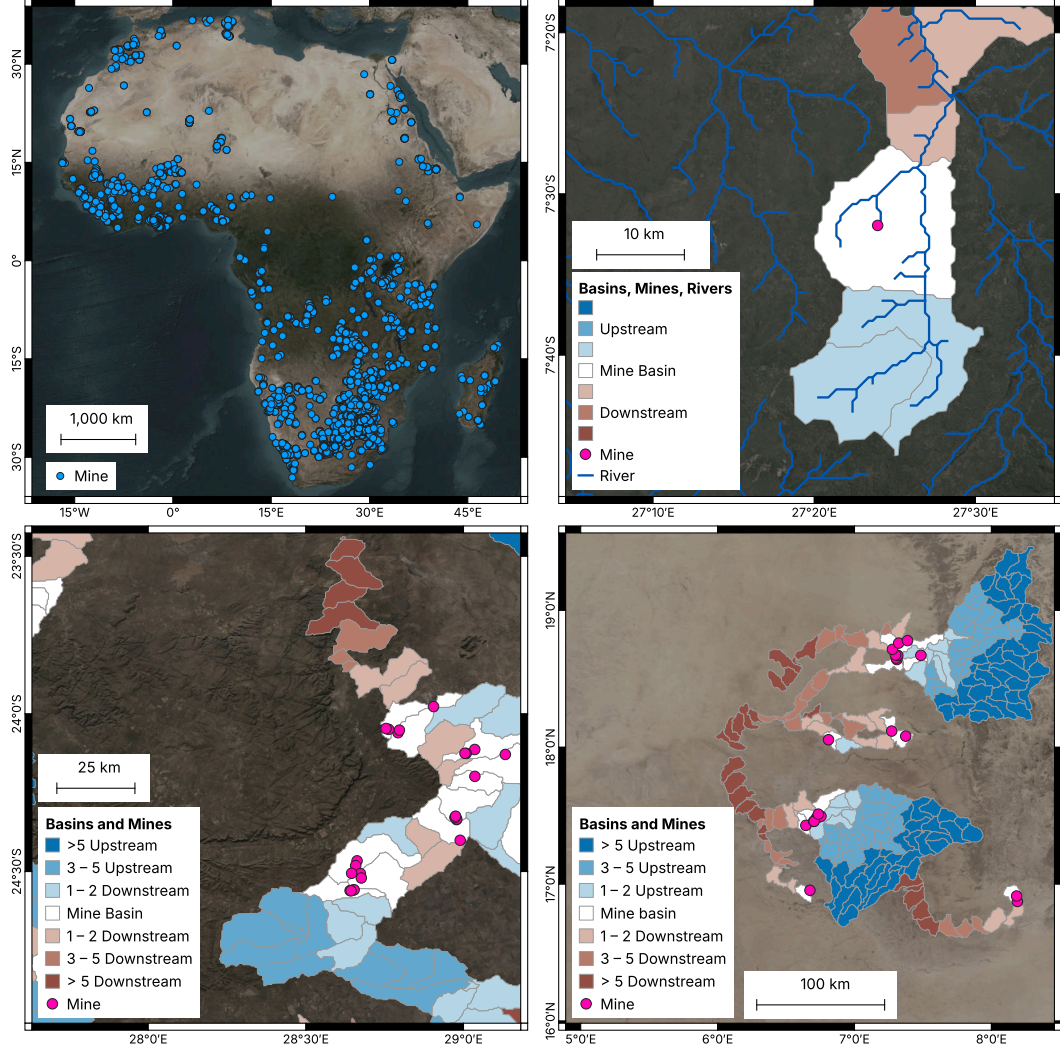


FIGURE A2: Distribution of the mine polygons in [Maus et al. \[2022\]](#) (top-left). Illustration of a mine that is entirely contained within one basin (top-right), mine clusters that reach across multiple basins (bottom-left), and a number of interconnected and closely adjacent basin chains (bottom-right).

The mine basin Mine basins are assigned by intersecting the basins with the centroids of the mine polygons by Maus 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 quality and plant growth. Then, we correlate our satellite-derived productivity measure (the maximum annual EVI) with agricultural production data in the region, allowing us to quantify economic impacts. 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

Numerous studies document the degradation of water quality downstream of mining operations. This pollution affects both surface water and groundwater, and has significant implications for agriculture and ecosystem health.

For Zimbabwe’s Deka River, [Ruppen et al. \[2021\]](#) documented a 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 et al. \[2018\]](#), [Duncan \[2020\]](#) 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, [Sahoo and Khaoash \[2020\]](#) found that 15% of groundwater samples in the Brajrajnagar coal mining area were of poor quality, with 43% requiring special treatment before agricultural use due to elevated heavy metal concentrations. In Brazil, [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 in the Jacaré and Contas rivers. Collectively, these studies confirm mining as a major contributor to hazardous water pollution that affects both human and agricultural systems.

B1.1. 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 cen-

timeter) 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 et al., 2019], with varying impacts by crop type [Page et al., 2021].

B2. Vegetation indices and crop yields

To translate our satellite-derived vegetation measurements to meaningful agricultural impacts, we correlated our outcome with downscaled crop production statistics from the Advancing Research on Nutrition and Agriculture (AReNA) Demographic and Health Surveys (DHS)-GIS Database of the International Food Policy Research Institute [IFPRI, 2020]. This dataset contains information on yields and production values for various crops at survey sites across 34 African countries, comprising approximately 45,000 observations collected between 2001 and 2018.

We focused on three key metrics: (1) physical yield (kg/hectare), (2) total production value (USD), and (3) financial yield (USD/hectare). For each DHS cluster, we created 6.2 km radius disks (matching our average basin size) and extracted the maximum annual EVI for the survey year. We then related the logarithm of agricultural metrics (winsorized at the 1% level) to this EVI value to determine semi-elasticities.

Table B1 reports these relationships. Our proxy of agricultural vegetation correlates strongly with all agricultural production measures. A 0.1 unit increase in maximum annual EVI is associated with: a 9.0% increase in cereal yields, a 5.6% increase in the financial yield of cereal crops, a 24.9% increase in the value of cereal production, a 9.6% increase in the financial yield of all crops, and a 34.0% increase in the value of overall crop production.

B2.1. Impact of mining on agricultural production

Using these correlations, we can translate our estimated mining impacts on vegetation into quantifiable agricultural losses. Our analysis found a 0.0064–0.0068 unit decrease in the maximum annual EVI on croplands for the three basins immediately downstream of the mine basin (see Table 2), depending on the included set of covariates. Applying the correlations from Table B1, this EVI reduction implies: a

TABLE B1: Maximum annual EVI and agricultural production

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

Clustered (wave) standard-errors in parentheses

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

0.57–0.61% decrease in cereal yields, a 0.36–0.38% decrease in the financial yield of cereal crops, a 1.59–1.70% decrease in the value of cereal production, a 0.61–0.65% decrease in the financial yield of all crops and a 2.16–2.31% decrease in the value of overall crop production.

To estimate the aggregate economic impact, we calculate the total physical production loss. The total cropland area in the first three downstream basins amounts to slightly more than 74,000 km² (or 7.4 million hectares). The winsorized (1% level) AReNA DHS-GIS data suggests an average yield of 2,000 kg/hectare, roughly falling in line with FAO values. Coupled with our estimate, we reach a total annual production loss of 91,100 tons of cereal (2,000 kg/hectare × 0.006 reduction × 7.4 million hectares).

Next, we quantify the aggregate loss in financial terms. Average values for financial yields from the winsorized (1% level) AReNA DHS-GIS dataset are around (i) 1,200 USD/hectare for cereal crops and (ii) around 2,100 USD/hectare for overall crop production. Coupled with the average reductions in financial yields of 0.36–0.38% for cereal crops and 0.61–0.65% for overall crop production, and the above area for crop cultivation, we estimate total financial losses at roughly 34 million USD for cereal crops (1,200 USD/hectare × 0.0038 reduction × 7.4 million) and roughly 102 million USD for overall crop production (2,100 USD/hectare × 0.0065 reduction × 7.4 million).

Importantly, these estimated losses occur annually, persist over time, do not reflect negative impacts of adaptation, only reflect water-mediated impacts on the immediate vicinity of mines, and affect some of the most destitute regions of the

world. Water pollution from mines can continue for decades after operations cease [Macklin et al., 2023], and economic impacts accumulate over long periods.

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.²⁹

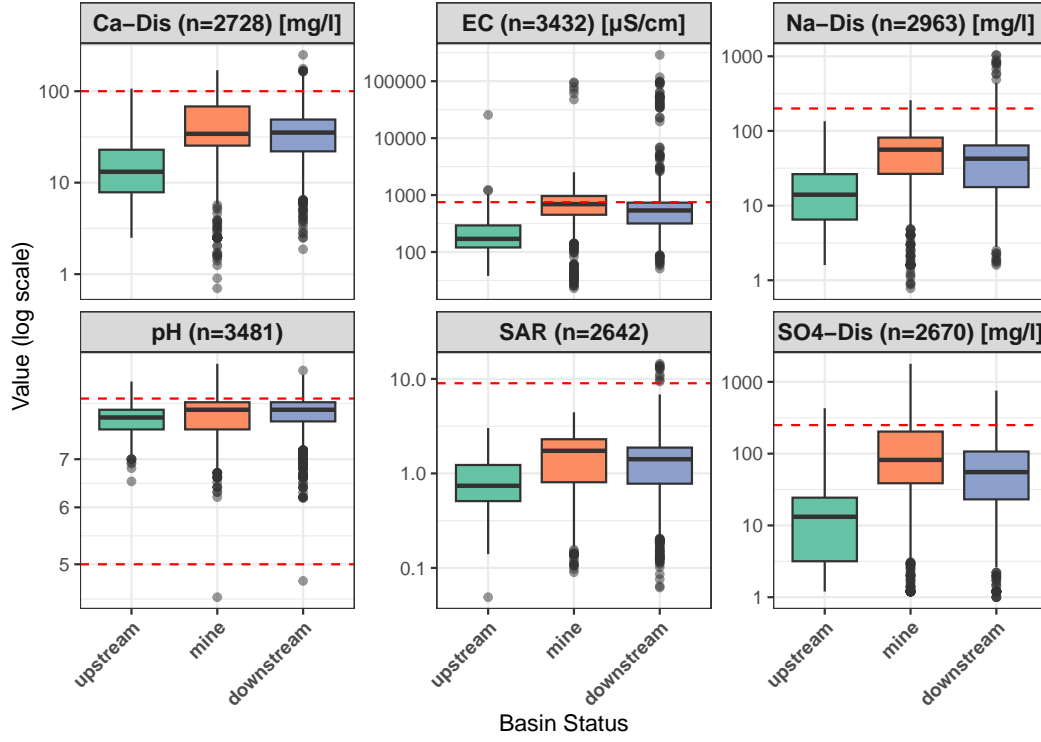
[Figure B3](#) presents six key water quality parameters measured in upstream, mine, and downstream basins:

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** (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.

²⁹The 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].

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 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 δ, γ 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 E12](#), 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](#)].

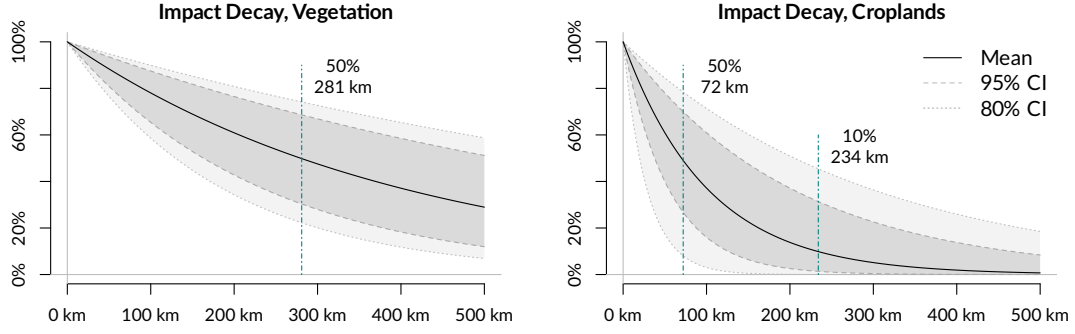


FIGURE C4: Impact decay over the river distance, assuming an exponential decay function. The solid black line denotes the mean effect; the shaded area between the dotted (dashed) lines denotes the 95% (80%) credible interval. The vertical lines denote the distance where the average impact is reduced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

$$F(x) = (|x| + x^2) \times I(x < 0) + I(x = 0) + (1 + x + x^2) \times I(x > 0),$$

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 E12](#). 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 et al., 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](#)].³⁰ We follow the set of practices as outlined by [Cattaneo et al. \[2019\]](#) and employ a triangular kernel, which gives observations closer to the cutoff a greater weight, and chose the bandwidth by minimizing the mean squared prediction error.

The results of this exercise are presented in [Table E13](#). 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.³¹ 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 min-

³⁰Note 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.

³¹[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.

erals. 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 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 et al., 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 D11 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 D11](#) show that the pre-matching imbalances in temperature and rainfall were effectively eliminated in the adjusted sample. [Figure D12](#) 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 E10](#). 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 D10](#) 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 variables 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 E14](#). 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.

D. Additional Figures

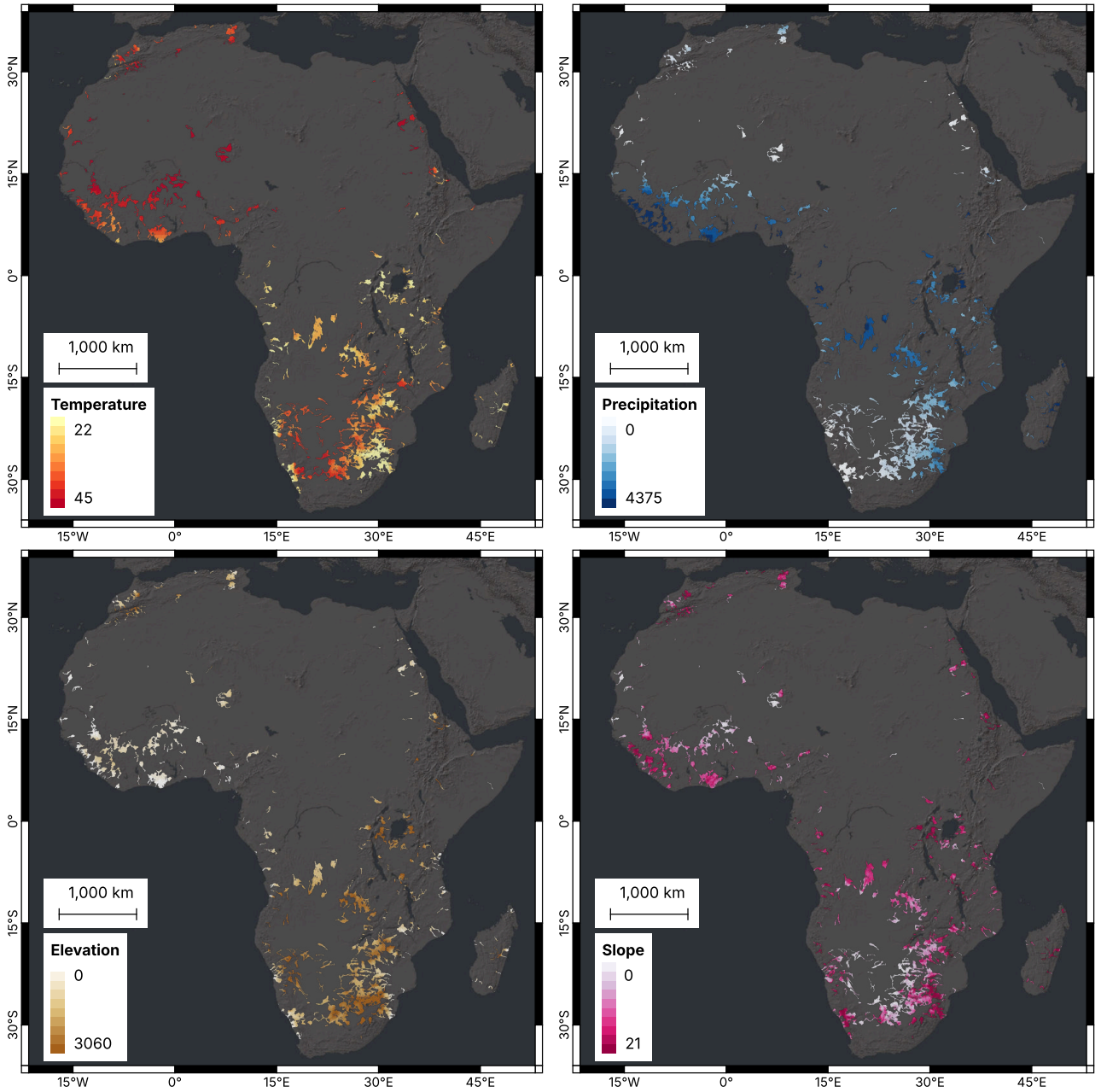


FIGURE D5: Maximum 2023 monthly temperature per basin in degrees Celsius (top-left), 2023 accumulated precipitation per basin in millimeters (top-right), average elevation in meters per basin (bottom-left), and average slope in degrees per basin (bottom-right). Basemap imagery provided by Esri, Maxar,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

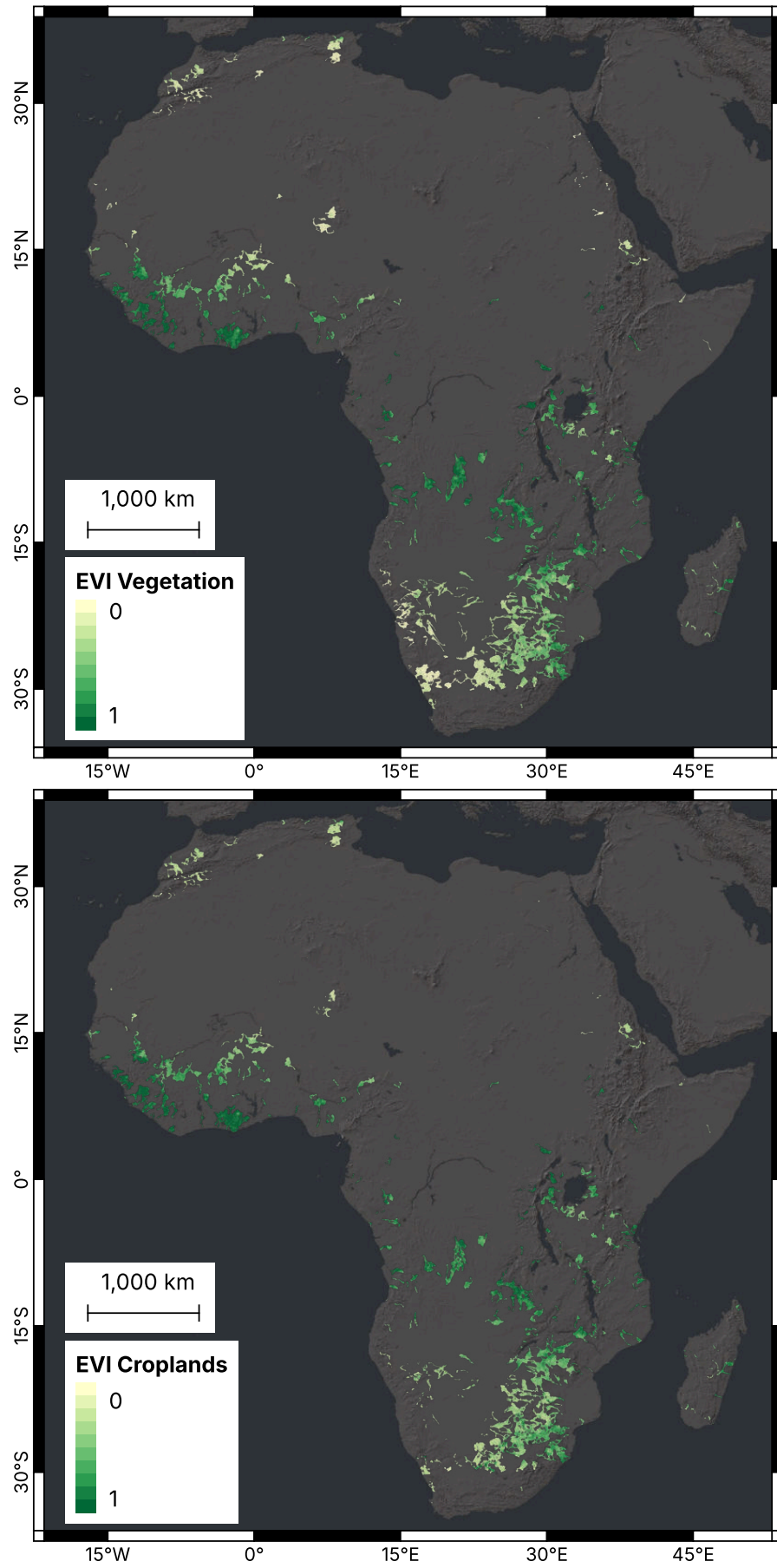


FIGURE D6: Maximum 2023 EVI per basin after applying the CCI vegetation mask (left) and the CCI cropland mask (right). Basemap imagery provided by Esri, Maxar, Earthstar Geographics, and the GIS User Community.

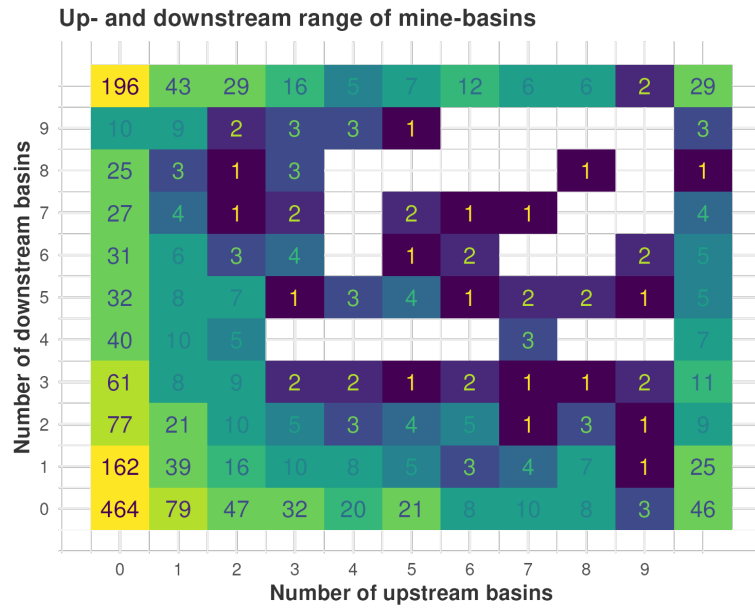


FIGURE D7: Number of mine-basins with Y upstream and X downstream basins in the dataset.

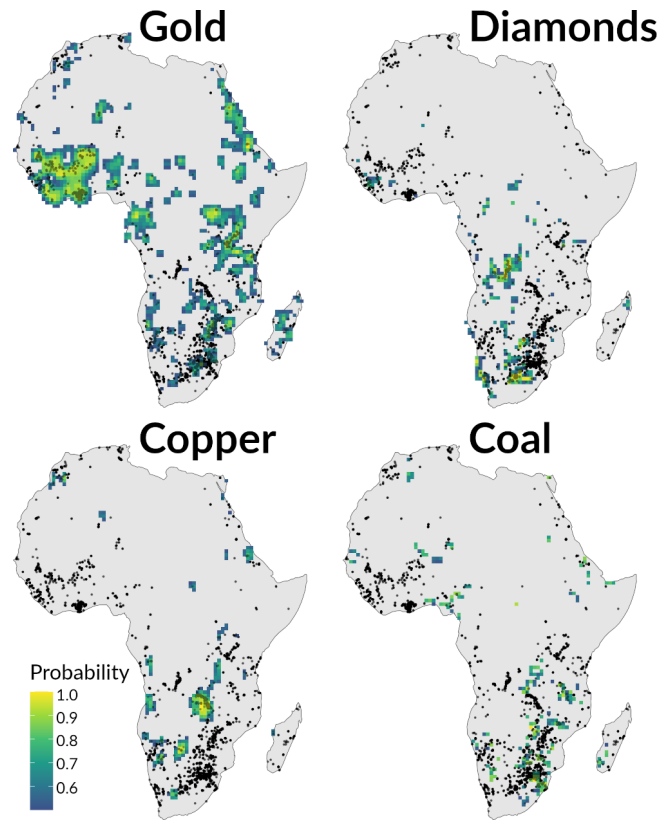


FIGURE D8: Heatmap indicating the results of the commodity prediction and point locations of training data. Note that the zoomed-out level of the heatmap averages over local nuances and does not accurately convey results for individual mines.

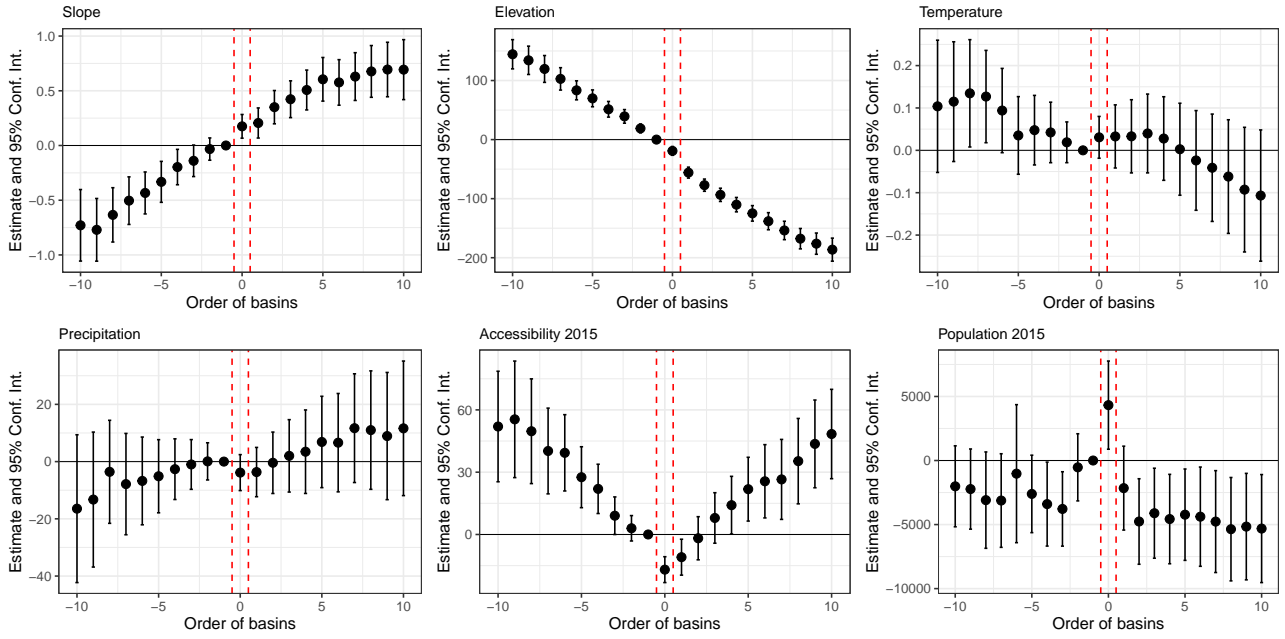


FIGURE D9: Order estimates when using elevation, slope, temperature, precipitation, accessibility to cities, and population as placebo outcomes.

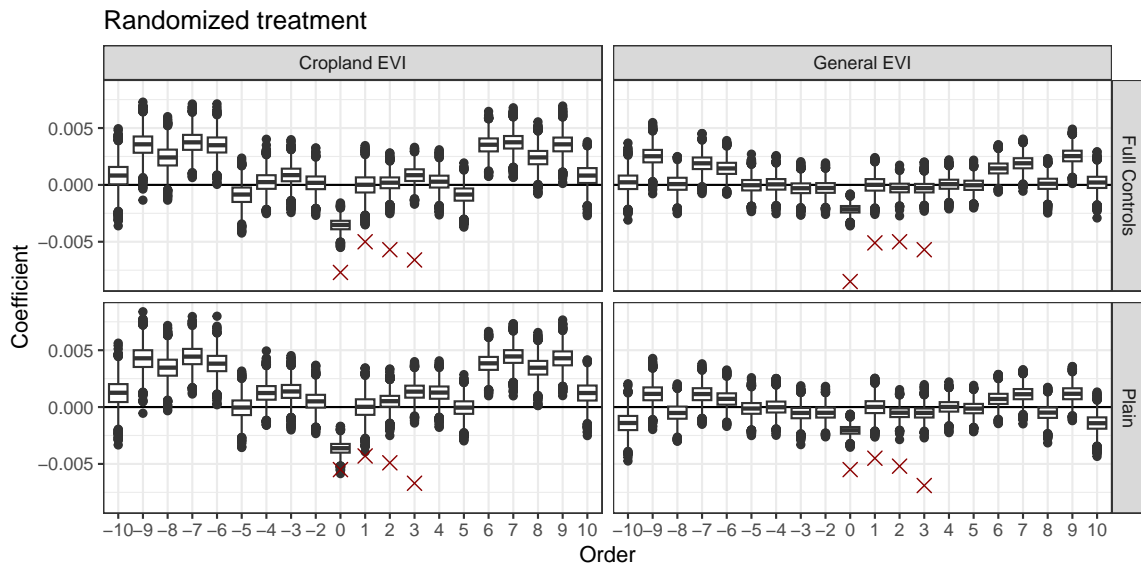


FIGURE D10: Estimation results when the treatment status (i.e., whether basins are down- or upstream) is randomized (5,000 runs, balance between statuses is kept). The red crosses indicate estimation results for the main specification.

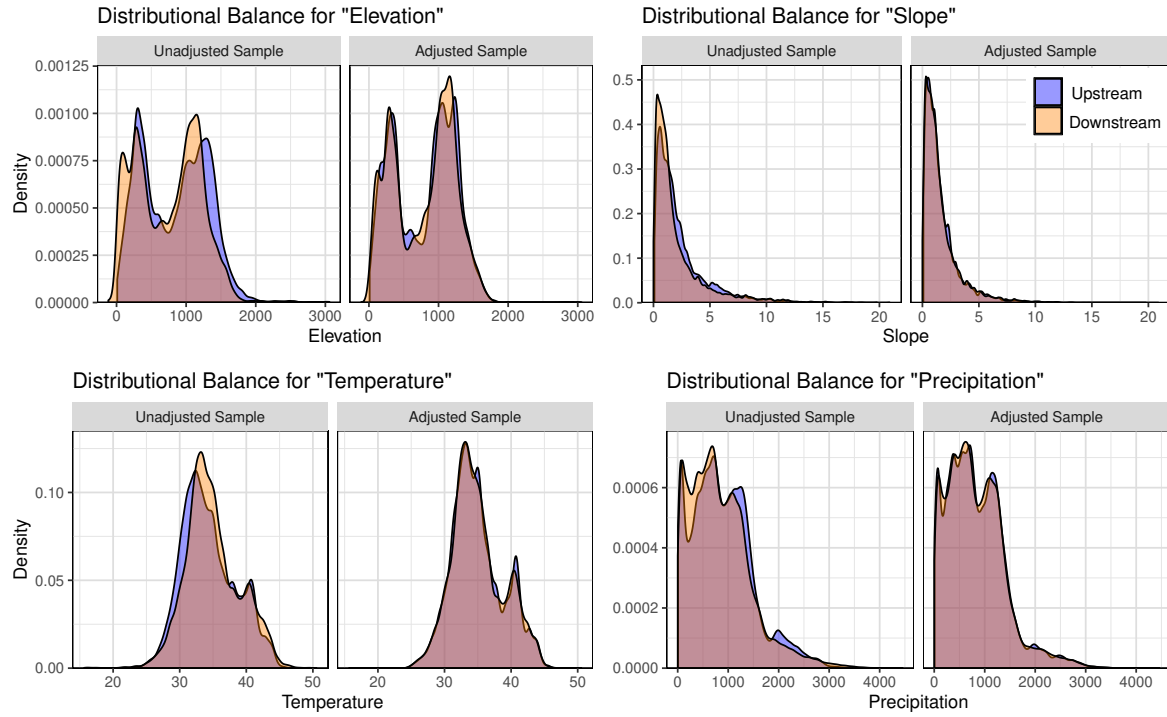


FIGURE D11: Balance of elevation, slope, temperature, and precipitation before and after matching. (Soilgrids not pictured.)

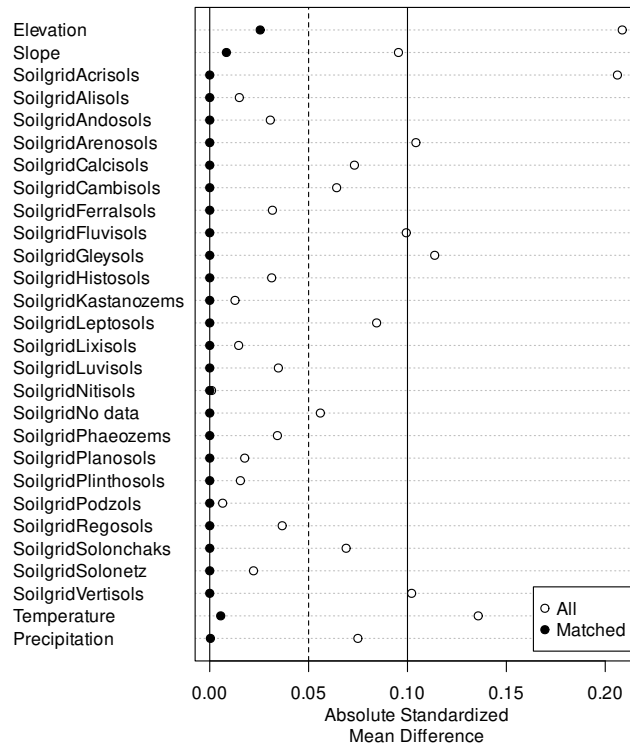


FIGURE D12: Standardized mean difference before and after matching.

E. Additional Tables

TABLE E2: Number and average distance (km) of basins by order.

Order	Upstream		Downstream	
	N	Distance	N	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 E3: Summary statistics split by status.

Upstream Basins						
Variable	Unit of Measurement	<i>NT</i>	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	<i>NT</i>	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	<i>NT</i>	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 E4: Main estimation results, order specification.

Dependent Variables: Model:	<i>Maximum Vegetation EVI</i>				<i>Maximum Croplands EVI</i>			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mine Basin	-0.0049*** (0.0014)	-0.0053*** (0.0014)	-0.0048*** (0.0014)	-0.0046*** (0.0014)	-0.0068*** (0.0022)	-0.0073*** (0.0022)	-0.0070*** (0.0022)	-0.0064*** (0.0022)
Downstream, 1	-0.0045*** (0.0017)	-0.0047*** (0.0018)	-0.0041** (0.0018)	-0.0043** (0.0018)	-0.0051** (0.0025)	-0.0051** (0.0025)	-0.0049** (0.0025)	-0.0050** (0.0025)
Downstream, 2	-0.0049** (0.0022)	-0.0048** (0.0024)	-0.0045* (0.0024)	-0.0048** (0.0024)	-0.0058* (0.0031)	-0.0061* (0.0032)	-0.0064* (0.0032)	-0.0067** (0.0032)
Downstream, 3	-0.0085*** (0.0028)	-0.0086*** (0.0029)	-0.0087*** (0.0029)	-0.0087*** (0.0029)	-0.0088** (0.0037)	-0.0092** (0.0038)	-0.0098** (0.0038)	-0.0099*** (0.0038)
Downstream, 4	-0.0049* (0.0030)	-0.0057* (0.0032)	-0.0061* (0.0033)	-0.0062* (0.0033)	-0.0029 (0.0038)	-0.0034 (0.0039)	-0.0042 (0.0039)	-0.0044 (0.0040)
Downstream, 5	-0.0034 (0.0033)	-0.0043 (0.0036)	-0.0053 (0.0037)	-0.0053 (0.0037)	0.0007 (0.0042)	0.0003 (0.0044)	-0.0015 (0.0045)	-0.0016 (0.0045)
Downstream, 6	-0.0027 (0.0034)	-0.0040 (0.0039)	-0.0057 (0.0040)	-0.0061 (0.0040)	-0.0004 (0.0046)	-0.0010 (0.0051)	-0.0035 (0.0052)	-0.0038 (0.0052)
Downstream, 7	-0.0053 (0.0037)	-0.0061 (0.0042)	-0.0087** (0.0043)	-0.0089** (0.0043)	-0.0051 (0.0048)	-0.0057 (0.0054)	-0.0092* (0.0055)	-0.0093* (0.0055)
Downstream, 8	-0.0095** (0.0041)	-0.0115** (0.0045)	-0.0141*** (0.0046)	-0.0144*** (0.0046)	-0.0047 (0.0053)	-0.0056 (0.0059)	-0.0085 (0.0060)	-0.0088 (0.0060)
Downstream, 9	-0.0066 (0.0045)	-0.0089* (0.0049)	-0.0120** (0.0050)	-0.0123** (0.0050)	-0.0060 (0.0056)	-0.0070 (0.0063)	-0.0108* (0.0064)	-0.0111* (0.0064)
Downstream, 10	-0.0063 (0.0049)	-0.0083 (0.0053)	-0.0118** (0.0054)	-0.0120** (0.0054)	-0.0023 (0.0060)	-0.0030 (0.0068)	-0.0071 (0.0069)	-0.0074 (0.0069)
Elevation		$-1.36 \times 10^{-5**}$ (6.33×10^{-6})	$-5.04 \times 10^{-5***}$ (6.85×10^{-6})	$-5.04 \times 10^{-5***}$ (6.85×10^{-6})		-9.82×10^{-6} (7.7×10^{-6})	$-4.93 \times 10^{-5***}$ (7.88×10^{-6})	$-4.93 \times 10^{-5***}$ (7.9×10^{-6})
Slope		0.0029*** (0.0005)	0.0024*** (0.0005)	0.0025*** (0.0005)		0.0027*** (0.0006)	0.0023*** (0.0007)	0.0022*** (0.0006)
Soil Type included		Yes	Yes	Yes		Yes	Yes	Yes
Yearly Max. Temperature			-0.0064*** (0.0006)	-0.0064*** (0.0006)			-0.0066*** (0.0007)	-0.0067*** (0.0007)
Yearly Precipitation			$3.69 \times 10^{-5***}$ (3.23×10^{-6})	$3.69 \times 10^{-5***}$ (3.21×10^{-6})			$3.3 \times 10^{-5***}$ (3.44×10^{-6})	$3.3 \times 10^{-5***}$ (3.44×10^{-6})
Accessibility in 2015				$-1.44 \times 10^{-5***}$ (7.17×10^{-6})				1.12×10^{-5} (1.76×10^{-5})
Population in 2015				$-7.78 \times 10^{-8***}$ (1.93×10^{-8})				$-6.85 \times 10^{-8***}$ (2.33×10^{-8})
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
R ²	0.90287	0.90457	0.90767	0.90788	0.81624	0.81763	0.82152	0.82179
Within R ²	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 E5: Main estimation results, pooled order specification.

Dependent Variables: Model:	<i>Maximum Vegetation EVI</i>				<i>Maximum Croplands EVI</i>			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mine Basin	-0.0053*** (0.0015)	-0.0057*** (0.0015)	-0.0054*** (0.0015)	-0.0051*** (0.0015)	-0.0079*** (0.0022)	-0.0084*** (0.0022)	-0.0083*** (0.0022)	-0.0076*** (0.0022)
Downstream, 1–3	-0.0057*** (0.0018)	-0.0057*** (0.0020)	-0.0054*** (0.0020)	-0.0056*** (0.0020)	-0.0064** (0.0025)	-0.0065** (0.0026)	-0.0067*** (0.0026)	-0.0068*** (0.0026)
Elevation		$-1.31 \times 10^{-5**}$ (6.26×10^{-6})	$-4.97 \times 10^{-5***}$ (6.78×10^{-6})	$-4.97 \times 10^{-5***}$ (6.79×10^{-6})		-9.24×10^{-6} (7.63×10^{-6})	$-4.86 \times 10^{-5***}$ (7.8×10^{-6})	$-4.85 \times 10^{-5***}$ (7.81×10^{-6})
Slope		0.0029*** (0.0005)	0.0024*** (0.0005)	0.0024*** (0.0005)		0.0027*** (0.0006)	0.0023*** (0.0007)	0.0022*** (0.0006)
Soil Type included		Yes	Yes	Yes		Yes	Yes	Yes
Yearly Max. Temperature			-0.0064*** (0.0006)	-0.0064*** (0.0006)			-0.0066*** (0.0007)	-0.0067*** (0.0007)
Yearly Precipitation			$3.69 \times 10^{-5***}$ (3.23×10^{-6})	$3.69 \times 10^{-5***}$ (3.22×10^{-6})			$3.29 \times 10^{-5***}$ (3.45×10^{-6})	$3.29 \times 10^{-5***}$ (3.45×10^{-6})
Accessibility in 2015				$-1.45 \times 10^{-5**}$ (7.12×10^{-6})				1.12×10^{-5} (1.75×10^{-5})
Population in 2015				$-7.8 \times 10^{-8***}$ (1.93×10^{-8})				$-6.89 \times 10^{-8***}$ (2.33×10^{-8})
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
R ²	0.90286	0.90456	0.90765	0.90787	0.81622	0.81761	0.82150	0.82177
Within R ²	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 E6: Estimation results for heterogeneity analysis: Mine characteristics

Mine Size	> 0.5km ²	> 1km ²	> 2.5km ²								
Mine Growth					> 0%	> 10%	> 25%				
Mine Commodity								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*** (0.0020)	-0.0062*** (0.0024)	-0.0069*** (0.0026)	-0.0073** (0.0033)	-0.0051** (0.0022)	-0.0048** (0.0023)	-0.0040 (0.0024)	-0.0052 (0.0060)	-0.0078 (0.0055)	-0.0074* (0.0045)	-0.0080*** (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
R ²	0.90787	0.91447	0.91335	0.91718	0.89661	0.89568	0.89697	0.77295	0.90098	0.93217	0.87340
Within R ²	0.05272	0.05421	0.05831	0.06647	0.05456	0.05365	0.04939	0.10382	0.09247	0.11833	0.03965
<i>Maximum Cropland EVI</i>											
Downstream, 1–3	-0.0068*** (0.0026)	-0.0046 (0.0036)	-0.0064 (0.0041)	-0.0035 (0.0054)	-0.0074*** (0.0027)	-0.0059** (0.0027)	-0.0066** (0.0030)	0.0095 (0.0091)	0.0015 (0.0080)	-0.0127 (0.0090)	-0.0119*** (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
R ²	0.82177	0.84026	0.84698	0.84604	0.79461	0.79109	0.79097	0.67233	0.83484	0.76409	0.82100
Within R ²	0.03234	0.03432	0.03980	0.05264	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 et al. [2025]. Models (8)–(11) report results for subsets of mines split by the primary commodity mined within them, not considering by-products. 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 E7: Estimation results for heterogeneity analysis: Spatial heterogeneity

Biome	Deserts	Forests	Grasslands							
Region				N. & E. Africa	S. Africa	W. Africa				
Crop Suitability							High	Medium	Low	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<i>Maximum Vegetation EVI</i>										
Downstream, 1–3	-0.0056*** (0.0020)	0.0020 (0.0041)	-0.0105** (0.0048)	-0.0055** (0.0023)	-0.0059 (0.0061)	-0.0020 (0.0025)	-0.0097** (0.0039)	-0.0066** (0.0028)	-0.0044 (0.0030)	-0.0022 (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
R ²	0.90787	0.85191	0.92357	0.82853	0.91358	0.89851	0.89497	0.77085	0.84687	0.87793
Within R ²	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*** (0.0026)	0.0159 (0.0133)	-0.0088** (0.0044)	-0.0072** (0.0029)	-0.0038 (0.0064)	-0.0014 (0.0035)	-0.0115** (0.0046)	-0.0088*** (0.0033)	-0.0046 (0.0038)	0.0015 (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
R ²	0.82177	0.70207	0.91138	0.73855	0.89173	0.75552	0.82927	0.72121	0.80964	0.79248
Within R ²	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.

E3. Robustness analysis

TABLE E8: Estimation results for alternative/additional controls

Dependent Variables:	<i>Maximum Vegetation EVI</i>					<i>Maximum Cropland EVI</i>				
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<i>Variables</i>										
Downstream, 1–3	-0.0056*** (0.0020)	-0.0061*** (0.0020)	-0.0056*** (0.0020)	-0.0056*** (0.0020)	-0.0053*** (0.0020)	-0.0068*** (0.0026)	-0.0073*** (0.0026)	-0.0066** (0.0026)	-0.0067*** (0.0026)	-0.0065** (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
R ²	0.90787	0.90670	0.90691	0.90787	0.90806	0.82177	0.82060	0.82156	0.82182	0.82187
Within R ²	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 μ 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 E9: Estimation results for varying outcome variables

Dependent Variables:	<i>Vegetation</i>					<i>Croplands</i>				
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<i>Variables</i>										
Downstream, 1–3	-0.0056*** (0.0020)	-0.0057** (0.0023)	-0.0048** (0.0020)	-0.0046** (0.0021)	-0.0032** (0.0013)	-0.0068*** (0.0026)	-0.0058** (0.0028)	-0.0053** (0.0027)	-0.0058** (0.0028)	-0.0037** (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
R ²	0.90787	0.93543	0.90344	0.91885	0.95522	0.82177	0.86923	0.78699	0.82381	0.91360
Within R ²	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 E10: Estimation results for varying sample definition

Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Maximum Vegetation EVI</i>								
Downstream, 1–3	-0.0056*** (0.0020)	-0.0061*** (0.0020)	-0.0061*** (0.0021)	-0.0043* (0.0022)	-0.0045** (0.0023)	-0.0055** (0.0023)	-0.0046* (0.0024)	-0.0071*** (0.0024)
<i>Fit statistics</i>								
Observations	110,524	69,107	41,449	58,545	95,822	8,152	94,902	72,009
R ²	0.90787	0.91075	0.92199	0.89783	0.90765	0.92270	0.90891	0.90252
Within R ²	0.05272	0.04812	0.04801	0.05360	0.05214	0.06617	0.05269	0.05244
<i>Maximum Cropland EVI</i>								
Downstream, 1–3	-0.0068*** (0.0026)	-0.0071*** (0.0027)	-0.0068** (0.0027)	-0.0067** (0.0029)	-0.0067** (0.0029)	-0.0071** (0.0034)	-0.0067** (0.0030)	-0.0092*** (0.0032)
<i>Fit statistics</i>								
Observations	93,000	58,288	34,883	49,798	79,931	6,967	79,123	60,357
R ²	0.82177	0.82521	0.83763	0.80996	0.82360	0.86559	0.82186	0.81508
Within R ²	0.03234	0.02663	0.02700	0.03324	0.03211	0.06258	0.03282	0.03575
<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 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 columns (2) restrict the sample to data from 2019 and onwards, models in column (3) restrict the sample to data from the period 2018 to 2020. Models in column (4) only include basin systems with at least one up- and downstream basin, models in column (5) exclude the mine basin itself, models in column (6) include only basins of order ± 1 of mine-basin systems with at least one up- and downstream basin and exclude the mine-basin. Models in columns (7) and (8) report results from matching procedures. Models in column (7) matches on geophysical variables (elevation, slope, and soil), models in column (8) 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. Standard errors are clustered at the mine basin system level.

TABLE E11: Estimation results for varying fixed effects

Dependent Variables: 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*** (0.0020)	-0.0062*** (0.0018)	-0.0062*** (0.0017)	-0.0057*** (0.0020)	-0.0056*** (0.0020)	-0.0068*** (0.0026)	-0.0059** (0.0024)	-0.0071*** (0.0022)	-0.0067*** (0.0026)	-0.0067*** (0.0026)
<i>Fixed-effects</i>										
Year	Yes	Yes	Yes		Yes	Yes	Yes	Yes		Yes
Mine	Yes			Yes	Yes	Yes			Yes	Yes
Pfaffstetter basin level 8		Yes					Yes			
Pfaffstetter basin level 6			Yes					Yes		
Country-by-year				Yes					Yes	
Mine-specific time trends					Yes					Yes
<i>Fit statistics</i>										
Observations	110,524	110,524	110,524	110,524	110,524	93,000	93,000	93,000	93,000	93,000
R ²	0.90787	0.90277	0.88482	0.91812	0.91341	0.82177	0.80802	0.78259	0.83552	0.82970
Within R ²	0.05272	0.05884	0.06696	0.03154	0.10975	0.03234	0.03504	0.05020	0.01668	0.07542

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-specific 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 with mine fixed effects. Models in columns (2) and (7) use fixed effects at Pfaffstetter level 8 basins, models in columns (3) and (8) fixed effects at Pfaffstetter level 6 basins. Models in columns (4) and (9) report results using country-by-year fixed effects, models in columns (5) and (10) report results including mine-specific linear time trends. All specifications include the full set of controls. Standard errors are clustered at the mine basin system level.

E4. Distance-based specifications

TABLE E12: Main estimation results specification using distance as running variable.

Dependent Variables: Model:	Maximum Vegetation EVI				Maximum Croplands EVI			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Linear distance</i>								
Downstream	-0.0050** (0.0023)	-0.0045** (0.0022)	-0.0033 (0.0022)	-0.0034 (0.0022)	-0.0050* (0.0029)	-0.0049* (0.0029)	-0.0041 (0.0029)	-0.0042 (0.0029)
Downstream × Distance	-7.57×10^{-6} (4.69×10^{-5})	-3.59×10^{-5} (5.36×10^{-5})	-8.32×10^{-5} (5.38×10^{-5})	-8.47×10^{-5} (5.32×10^{-5})	1.47×10^{-5} (5.85×10^{-5})	-4.19×10^{-6} (6.91×10^{-5})	-5.85×10^{-5} (6.96×10^{-5})	-5.96×10^{-5} (6.94×10^{-5})
Distance	7.75×10^{-6} (3.91×10^{-5})	3.26×10^{-5} (4.13×10^{-5})	5.61×10^{-5} (4.12×10^{-5})	6.18×10^{-5} (4.04×10^{-5})	2.75×10^{-5} (4.97×10^{-5})	4.08×10^{-5} (5.45×10^{-5})	6.32×10^{-5} (5.45×10^{-5})	5.66×10^{-5} (5.3×10^{-5})
<i>Fit statistics</i>								
Observations	110,576	110,568	110,564	110,524	93,036	93,036	93,032	93,000
R ²	0.90282	0.90452	0.90762	0.90783	0.81609	0.81748	0.82138	0.82165
<i>Linear-quadratic distance</i>								
Downstream	-0.0056** (0.0027)	-0.0055** (0.0026)	-0.0050** (0.0025)	-0.0052** (0.0025)	-0.0077** (0.0035)	-0.0076** (0.0036)	-0.0072** (0.0035)	-0.0073** (0.0035)
Downstream × Distance	2.64×10^{-5} (0.0001)	2.01×10^{-5} (0.0001)	5.75×10^{-6} (0.0001)	5.45×10^{-6} (0.0001)	0.0002 (0.0002)	0.0001 (0.0002)	0.0001 (0.0002)	0.0001 (0.0002)
Downstream × Distance ²	-3.04×10^{-7} (8.52×10^{-7})	-4.7×10^{-7} (8×10^{-7})	-7.27×10^{-7} (8.09×10^{-7})	-7.35×10^{-7} (7.99×10^{-7})	-1.2×10^{-6} (1.2×10^{-6})	-1.17×10^{-6} (1.2×10^{-6})	-1.38×10^{-6} (1.18×10^{-6})	-1.36×10^{-6} (1.18×10^{-6})
Distance	3.97×10^{-5} (9.08×10^{-5})	3.93×10^{-5} (8.61×10^{-5})	3.33×10^{-5} (8.93×10^{-5})	3.64×10^{-5} (8.63×10^{-5})	-4.23×10^{-6} (0.0001)	1.24×10^{-6} (0.0001)	1.17×10^{-5} (0.0001)	-1.21×10^{-6} (0.0001)
Distance ²	-2.43×10^{-7} (6.32×10^{-7})	-5.04×10^{-8} (5.85×10^{-7})	1.76×10^{-7} (6.05×10^{-7})	1.97×10^{-7} (5.91×10^{-7})	2.55×10^{-7} (9.26×10^{-7})	3.18×10^{-7} (9.37×10^{-7})	4.13×10^{-7} (9.11×10^{-7})	4.64×10^{-7} (9.2×10^{-7})
<i>Fit statistics</i>								
Observations	110,576	110,568	110,564	110,524	93,036	93,036	93,032	93,000
R ²	0.90283	0.90453	0.90762	0.90784	0.81612	0.81751	0.82142	0.82168
<i>Exponential decay</i>								
	$\delta = 0.005$	$\delta = 0.006$	$\delta = 0.002$	$\delta = 0.002$	$\delta = 0.035$	$\delta = 0.035$	$\delta = 0.020$	$\delta = 0.010$
$\exp -\delta \times \text{Distance} \times \text{Downstream}$	-0.0062*** (0.0023)	-0.0062*** (0.0023)	-0.0060*** (0.0023)	-0.0062*** (0.0023)	-0.0093*** (0.0034)	-0.0091*** (0.0033)	-0.0074** (0.0029)	-0.0068** (0.0029)
<i>Fit statistics</i>								
Observations	110,576	110,568	110,564	110,524	93,036	93,036	93,032	93,000
R ²	0.901147	0.902842	0.905958	0.906169	0.812592	0.813949	0.817862	0.818141
<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 E13: Distance specification using automatic bandwidth selection

	Max EVI		Max C EVI	
	No Controls			
Conventional	-0.0048*** (0.0017)	-0.0081*** (0.0021)	-0.0088*** (0.0024)	-0.0118*** (0.0027)
Bias-Corrected	-0.0055*** (0.0017)	-0.0086*** (0.0021)	-0.0098*** (0.0024)	-0.0120*** (0.0027)
Observations	33063	50360	30239	51370
Bandwidth (conv)	18.4	33.2	19.6	41.9
Bandwidth (bias)	42.1	67.5	47.3	78.5
	With full Controls			
Conventional	-0.0028* (0.0017)	-0.0055*** (0.0021)	-0.0045* (0.0023)	-0.0075** (0.0031)
Bias-Corrected	-0.0034** (0.0017)	-0.0061*** (0.0021)	-0.0054** (0.0023)	-0.0083*** (0.0031)
Observations	34107	49372	29437	43370
Bandwidth (conv)	19	32.3	19	33.1
Bandwidth (bias)	42.6	65.6	44	65.4
<i>Settings</i>				
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 kilometer along the river network used as the running variable, using practices suggested in Cattaneo et al. [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 E14: Estimation results using covariates as placebo outcomes for squared distance specification

Dependent Variables: Model:	Slope (1)	Elevation (2)	Temperature (3)	Precipitation (4)	Accessibility (5)	Population (6)
<i>Variables</i>						
Downstream	-0.009 (0.095)	-23.4*** (7.08)	0.067 (0.047)	1.39 (10.7)	-3.63 (6.90)	-2,838.9* (1,622.7)
Distance \times Downstream	0.020*** (0.004)	-3.12*** (0.368)	-0.003 (0.003)	-0.068 (0.635)	-0.103 (0.385)	21.8 (54.1)
Distance ² \times Downstream	-4.79×10^{-5} (3.8×10^{-5})	-0.001 (0.003)	-1.36×10^{-6} (2.17×10^{-5})	0.004 (0.005)	0.0003 (0.003)	-0.135 (0.339)
Distance	-0.011*** (0.003)	1.70*** (0.290)	0.002 (0.002)	0.400 (0.575)	0.850*** (0.275)	-72.6 (44.4)
Distance ²	1.07×10^{-5} (2.91×10^{-5})	0.0008 (0.002)	-3.44×10^{-6} (1.85×10^{-5})	-0.004 (0.005)	-0.001 (0.002)	0.438 (0.277)
<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
R ²	0.75464	0.97225	0.95948	0.94458	0.88907	0.59487
Within R ²	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.