Large-Scale Irrigation and Water Use Inequality in South Africa

Tereza Varejkova*

May 15, 2025

Click Here for Latest Version

Abstract

Large-scale surface irrigation is essential in keeping agriculture viable in regions prone to drought but it has the potential to exacerbate inequality due to the uneven distribution of its benefits and costs. I investigate this issue in the context of South Africa by estimating heterogeneous effects of irrigation canals on crop productivity and agricultural land expansion by type of farmer. To estimate these effects, I use remote sensing measures of crop yields and a novel land cover classification dataset in a spatial regression discontinuity framework with relative elevation to the nearest canal as the running variable. Areas below the canals serve as the treated group, while areas above serve as the control. The findings show that commercial farmers below canals benefit in terms of higher maize and wheat yields and expand their area under production. Census data further reveal that these expanding commercial farms create employment opportunities for the rural poor. In contrast, subsistence farmers below canals experience lower yields relative to those above and do not expand their cultivated area. Despite the unequal distribution of benefits, a cost-benefit analysis demonstrates that large-scale irrigation infrastructure remains a cost-effective investment.

Keywords: Irrigation, Regression Discontinuity, Land Use, Remote Sensing
JEL codes: O13, Q15, Q24, Q25

*American University, School of International Service. E-mail: tvarejko@umd.edu. I am extremely grateful to my advisors Sebastian Galiani, Guido Kuersteiner, and Maureen Cropper for their guidance and support. I am also grateful to Taryn Dinkelman, Ken Leonard, Cory Smith, Mamidipudi Ramakrishna Sharan, Louis Preonas, and Mahima Vasishth for invaluable feedback and suggestions that I received during my presentations at outside seminars. Finally, I would like to thank Alexia Lochmann, Felipe Kup Barbieri de Matos, Rashid Hassan, Marion Delport, and Carin Bosman for helping me with access to the data and providing me with useful insights into the South African agricultural and water sectors.

1 Introduction

Large-scale surface irrigation infrastructure is a major investment for economic development that requires substantial government expenditure. Support for building large dams and canals has fluctuated over time. Numerous construction projects were initiated in developing countries during the 20th century, but since the early 2000s, concerns about their social and environmental impacts have led to an increased caution (WCD, 2000). Critics of large dams often point out the high construction and maintenance costs, as well as the unequal distribution of benefits. In contrast, proponents emphasize key advantages, such as increases in agricultural productivity that can trigger structural transformation through inter-sectoral linkages (Lewis, 1954; Johnston and Mellor, 1961; Ranis and Fei, 1961) and a potential for climate change adaptation as rainfall patterns become more variable (World Bank, 2010; IPCC, 2018). Today, large dam projects are experiencing a renewed interest from international donors (Dillon and Fishman, 2019). Understanding the long-run impacts of this infrastructure on different communities is therefore a timely and critical concern.

This paper contributes to the ongoing debate on whether large-scale surface irrigation infrastructure constitutes a worthwhile public investment for economic development. An interesting and yet unexplored question is what type of farmers benefit from irrigation dams. Is irrigation access equally available to all farmers downstream of these dams, or do resource-poor farmers face barriers to adoption? Specifically, this paper answers the following questions: Do the effects of proximity to irrigation infrastructure differ between commercial and subsistence farmers? And if so, is large-scale irrigation infrastructure still cost-effective despite potential negative externalities?

South Africa provides a suitable context for studying large-scale irrigation technology due to its well-developed system of dams and canals. This infrastructure is critical as 65% of the country does not receive enough rainfall for rainfed agriculture to be viable. Another important aspect of the South African agriculture is its dual structure. On the one hand, commercial farms generate large revenues and employ around 750,000 paid workers. On the other hand, the approximately two million subsistence farmers represent the poorest and most vulnerable segment of the population.[1] Subsistence farmers are predominantly concentrated in the former homelands—areas where the black population was forcibly relocated to during apartheid. At that time, subsistence farmers were denied land and water rights. Although the 1998 National Water Act, a post-apartheid water policy, aimed to

[1]The statistic on the paid workforce on commercial farms comes from the 2017 Census of Commercial Agriculture (Statistics SA, 2020). The statistic on the number of subsistence farmers comes from Oqubay, Tregenna and Valodia (2020), p.205.

redress historical inequalities by providing secure irrigation water rights to disadvantaged farmers, it remains unclear whether subsistence farmers have been able to harness benefits from large-scale surface irrigation under the new regime. To understand the current distributional impacts of irrigation infrastructure, it is therefore crucial to consider the crop yields and land use patterns of commercial and subsistence farmers separately.

My identification strategy leverages the gravity-based nature of surface irrigation within a regression discontinuity (RD) framework. Fields that are topographically below a canal can be irrigated more easily, as water naturally flows downhill. In contrast, fields above a canal require significant amounts of electrical power to pump water uphill, leading to a sharp increase in irrigation costs at the threshold of zero relative elevation to the canal. For fields above the canal, these costs also rise with the vertical distance from the canal, as pumping water higher requires more electricity. This discontinuity in costs creates a corresponding discontinuity in the probability of land being irrigated: land below a canal is more likely to be irrigated than land above it. This motivates the use of a fuzzy RD design, with relative elevation to the nearest canal as the running variable. Areas below a canal are considered treated, while those above are defined as control. A similar empirical approach has been applied in the context of India (Asher et al., 2023).

The RD strategy requires a unit of analysis that is highly spatially disaggregated. For each irrigation canal, it is crucial to assign several units to both the treatment and control groups, with these units being in close proximity to ensure comparability. To achieve this for agricultural outcomes, I compile a fine spatial dataset consisting of over 29 million observations from various sources. This dataset merges remotely-sensed data on crop productivity, weather conditions, and terrain characteristics with a 73-class land cover classification dataset. By combining this data with the coordinates and elevation of 144 irrigation canals, I can assign treatment status at the field or even sub-field level. Moreover, when analyzing crop yields, I restrict the sample only to grid cells classified as growing annual crops, thus purging my estimates from noise caused by non-crop vegetation or perennial crops. For socio-economic outcomes, I merge census data available at the small area level with an elevation raster dataset and canal coordinates, yielding a sample of 3,035 observations.[2] Finally, I overlay both datasets with the boundaries of the former homelands, enabling the estimation of heterogeneous treatment effects by type of farmer, as subsistence farmers are located predominantly in the former homelands.[3]

[2]The South Africa 2011 Census at the small area level covers 100% of the population and contains 84,907 census units. Small area falls between enumeration area and sub-place levels.

[3]Combining the homelands boundaries with detailed land use data, I find that subsistence farmers are almost exclusively concentrated in the former homelands whereas commercial farms are located in non-homelands.

I find that irrigation canals boost agricultural production both at the intensive margin (by increasing crop yields) and the extensive margin (by expanding the cultivated area). However, only commercial farmers in non-homelands benefit from better access to irrigation. In areas below canals, wheat yields increase by 3.9% and maize yields by 2.3%. In contrast, in the former homelands, wheat yields decrease by 2.9% and maize yields by 2.6%, though the decline in wheat yields is not statistically significant. At the extensive margin, downstream areas in non-homelands are 4.26 percentage points (pp) more likely to be agricultural, a 20% increase relative to the mean in the control group. This increase is driven by the presence of annual crops (up by 4.53 pp or 27%), particularly irrigated annual crops (up by 2.27 pp or 54%). There is no noticeable effect on fallow land or perennial crops. In the former homelands, areas below canals show no increase in the cultivation of annual or perennial crops, whether irrigated or rainfed. However, there is a 0.03 pp (37%) reduction in fallow land. Overall, the evidence indicates that subsistence farmers do not benefit from large-scale irrigation technology to the same extent as commercial farmers do.

One potential concern with the finding that irrigation increases the area under production is the possibility that canals were placed to maximize the amount of arable land downstream. If this were the case, the observed extensive margin effects could reflect preexisting differences rather than the impact of canal construction. To address this concern, I examine the expansion of agricultural land between 1990 and 2020 as an additional outcome variable. This period is significant due to a key historical event that likely caused major land use changes—–the end of apartheid in 1994. Under apartheid, water policy was characterized by riparian rights. Only landowners whose property bordered a body of water had the right to extract the resource and use it. As a result, new farmers struggled to enter the agricultural sector unless their land included direct access to a water source. In contrast, the post-apartheid regime introduced the 1998 National Water Act, which established water use licensing and democratized access to irrigation water. With the new water rights, areas below canals should have experienced faster agricultural land expansion compared to upstream areas, as it is cheaper to bring irrigation water downstream. This is exactly what I find. In non-homelands, the conversion rate of land for agricultural purposes between 1990 and 2020 was 28% (0.01 pp) higher in downstream areas relative to the upstream ones.[4] In the former homelands, however, the effect is smaller in magnitude and not statistically significant.

To determine whether agricultural productivity gains translated to welfare improve-

[4] In this analysis, 1990 represents the "before" period, and 2020 represents the "after" period. These dates were chosen based on the availability of a dataset documenting land use changes during this time.

ments, I examine the impact of irrigation canals and the resulting expansion of commercial farms on rural employment. Using census data and applying the same RD strategy, I estimate the impact of irrigation canals on local labor markets. The results show that areas downstream of canals have a 4.9 percentage point (pp) lower unemployment rate, representing a 20% reduction relative to the control group. Additionally, the proportion of households with zero income decreases by 2.7 pp (19%) in downstream areas.[5] However, these positive effects are limited to non-homelands. In the former homelands, the treatment effects are smaller and statistically insignificant.

The main results for crop yields remain robust when extending the analysis over a longer time period. While the primary specification uses data for a single year due to the limitations of the land cover classification dataset, I replicate the analysis using a less detailed but temporally more extensive cropland map (2000—2019). The same pattern emerges: commercial farmers benefit from irrigation, whereas subsistence farmers in the former homelands do not. These findings are also robust to adjustments such as varying the vertical and horizontal distance bandwidths to the nearest canal, excluding a donut hole, and controlling for a host of soil quality variables.

Finally, I perform a cost-benefit analysis of constructing a new large dam for irrigation purposes. For cost estimates, I rely on a feasibility study commissioned by the government to address water supply needs in Limpopo province. As for the benefits, I use my RD estimates on the effects of irrigation on crop yields and land use changes from the earlier analysis. I assume that the command area of the new dam would include both commercial and subsistence farmers' land, in proportions consistent with my remote sensing dataset. In the baseline scenario, using a 3% discount rate, I find a benefit-cost ratio of 1.62, suggesting that large irrigation dams are a cost-effective investment, despite the negative externalities imposed on subsistence farmers. I also test the sensitivity of this analysis to a higher discount rate (4%) and to a scenario in which the irrigation dam benefits only existing agricultural land. The latter scenario might arise if environmental or conservation regulations prevent uncultivated land from being converted to cropland despite new irrigation access. With the higher discount rate, the benefit-cost ratio remains comfortably above one. However, in the scenario where irrigation benefits only existing land, the ratio falls below one, indicating that the investment would not recover its costs.

The cost-benefit analysis suggests two implications for policymakers. First, subsistence farmers should be adequately compensated for any negative impacts from the construction of irrigation dams in their vicinity. Second, policymakers must carefully consider whether

[5] Although I lack data on household members' specific occupations, subsistence farmers who grow food for their own consumption would be recorded as households with zero income in the census.

a new dam would allow for the cultivation of previously uncultivated land. If the dam only benefits existing fields, the investment may not be justified.

This paper makes two contributions to the literature. First, I use detailed spatial data to estimate the long-run effects of irrigation canals in South Africa, adding to the body of work on the impacts of large-scale irrigation in both developed and developing countries (Duflo and Pande, 2007; Zaveri, Russ and Damania, 2020; Dillon and Fishman, 2019; Hornbeck and Keskin, 2014; Strobl and Strobl, 2011; Olmstead and Sigman, 2015; Blanc and Strobl, 2014; Mettetal, 2019). Much of the existing research relies on instrumental variable (IV) methods that exploit the exogeneity of geographic features to predict locations suitable for dam construction. These studies have found that irrigation increases crop yields, reduces dependency on rainfall, and alleviates poverty. More recently, regression discontinuity (RD) designs have been introduced as a new empirical approach for assessing the impact of irrigation infrastructure. Typically, the running variable in these studies is the distance to the command area boundary (Blakeslee et al., 2023; Jones et al., 2022; Hagerty, 2022), with Asher et al. (2023) being an exception by using relative elevation to the nearest canal. Conditional on demonstrating balance in geophysical covariates on either side of the RD threshold to eliminate concerns of endogenous placement, the RD approach can provide more credible estimates of local treatment effects since it relies on weaker assumptions than IV. I contribute to this growing literature by (1) applying the RD method in a new context—South Africa, (2) examining land use outcomes derived from a novel land cover classification dataset, (3) analyzing distributional impacts based on the type of farmer, and (4) conducting a comprehensive cost-benefit analysis of constructing a new large irrigation dam taking into account the distributional impacts. In particular, the last point has been identified as a critical research gap in the literature on hydrological investments (Dillon and Fishman, 2019).

Second, this paper contributes to the literature on the persistence of place-based inequalities, particularly in light of the fact that subsistence farmers are concentrated almost exclusively in the former homelands. In South Africa, these areas continue to lag behind the rest of the country in terms of labor market outcomes and basic infrastructure (Lochmann, 2022) and individuals relocated to these regions during childhood exhibit lower educational attainment (Carrillo, Charris and Iglesias, 2023). This paper adds to the evidence on the enduring legacy of apartheid by demonstrating that farmers in the former homelands are unable to benefit from large-scale irrigation infrastructure, despite being in close geographic proximity.

The paper is organized as follows. In Section 2, I provide some background information on irrigation infrastructure and water policy in South Africa. In Section 3, I describe the data. Section 4 presents the empirical strategy and discusses the identifying assumptions. In Section 5, I present the main results. In Section 6, I outline a cost-benefit analysis, and in Section 7, I conclude.

2 Background

2.1 Water scarcity and surface irrigation system

South Africa is severely affected by water scarcity, due to its low average annual rainfall. The country receives only about 450 mm per year, much less than the world average of 860 mm (WRC, 2024). Furthermore, climate change is expected to exacerbate the situation by making droughts in southern Africa more frequent and severe (Chikoore and Jury, 2021).

Despite water scarcity, South Africa has developed a robust agricultural sector, largely thanks to irrigation that consumes 60% of available water. Irrigation systems now cover 1.3 million hectares, accounting for 7.2% of arable land, and contribute to 30% of the country's crop production (Dennis and Nell, 2002). Irrigation systems can be classified into three groups based on their components, cost, and performance: (1) flood irrigation, where water flows over the soil by gravity (e.g., basin and furrow systems), (2) mobile systems, which irrigate as they move across fields (e.g., center-pivot and travelling-gun systems), and (3) static systems, which remain stationary during irrigation, including sprinklers and micro-irrigation systems like drip and mini-sprinklers (Reinders, 2011). Flood irrigation, though the least efficient and most prone to water loss, is the most affordable option. In contrast, the mobile and static systems, such as center pivot irrigation, are far more efficient but also the most capital-intensive method (Lichtenberg, 1989).

Irrigation in South Africa primarily relies on surface water.[6] Surface water is released from large dams or weirs in rivers and is distributed via an extensive network of primary and secondary canals.[7] There are four levels of water management infrastructure (Reinders et al., 2010). The first level is the "water source," which stores water in a dam or reservoir. The second level is the "bulk conveyance system," which transports water from the dam to several irrigation schemes via a canal. Alternatively, irrigation water can be extracted directly from a river. The third level is the "irrigation scheme" infrastructure, designed to deliver water from the main canal to the boundaries of participating farms. This

[6] Groundwater irrigation accounts for only 1% of cultivated land (Altchenko and Villholth, 2015). Given that 7.2% of cultivated land is irrigated overall, this implies that about 14% of irrigated crops rely on groundwater, while the remaining 86% depend on surface water.

[7] Figure A–1 in the Appendix illustrates the number of large dams constructed per decade. The first large dam dedicated to irrigation was built in 1913, with the majority of dams completed by the early 1990s. Since then, only six large dams have been constructed, all completed in the 2000s.

includes on-scheme dams, canals, and pipes. The final level is the "irrigation farm" itself, where on-farm dams, canals, and irrigation systems (such as center pivots, sprinklers, and micro-sprays) are used to deliver water directly to the crops.

In this paper, when I refer to irrigation canals, I am specifically referring to the second level of water management infrastructure, which marks the beginning of the canal system. At this point, water is released from a dam or weir into the main canal, which then carries it to irrigation schemes along its path. The distribution of water is controlled and monitored by water management authorities, with each user entitled to the amount of water for which they are registered with the Department of Water and Sanitation. To understand these water allocations, it is important to consider the history of water rights in South Africa.

2.2 History of water policy in South Africa

Water policy in South Africa has evolved over time, drawing its principles from Roman, Dutch, and later English legal traditions (Perret, 2002). The first notable legislation was the 1910 Irrigation Conservation of Water Act, which established the riparian principle as the central feature. Under this principle, landowners whose properties bordered a water source had the right to extract water, with minimal state intervention.

The second major water legislation, the 1956 Water Act, reaffirmed the riparian principle. This system disproportionately benefited large-scale commercial farmers, primarily white landowners, who had unrestricted access to water. Under apartheid, non-white populations were largely excluded from land ownership and forcibly relocated to homelands, or Bantustans, where water access was limited.

With the end of apartheid, the 1998 National Water Act was introduced, discarding the riparian system and establishing the government as the custodian of water resources. This legislation formed new institutions for decentralized water management. The Department of Water and Sanitation was tasked with managing water resources, but most responsibilities were supposed to be delegated to nineteen Catchment Management Agencies. Due to bureaucratic delays, only two agencies are fully operational today.[8] Additionally, Water User Associations were created as cooperative groups of water users, allowing communities to pool financial resources to manage irrigation more efficiently.

The primary goal of the 1998 National Water Act was to redress historical racial inequalities and ensure equitable water distribution to benefit all citizens. The Act intro-

[8]The only two operating agencies are the Breede-Gouritz and the Inkomati-Usutu Catchment Management Agencies. In the remaining parts of the country, the regional offices of the Department of Water and Sanitation take on water management responsibilities.

duced four types of water rights. First, water licences represent a modern approach to water policy. They are granted by the Department of Water and Sanitation based on applications by individual users. Second, users who could demonstrate that their water usage was lawful during a two-year period before the Act's promulgation are allowed to retain their rights under the existing lawful water use provision. Commonly referred to as the "sunset clauses," this category has faced criticism for effectively maintaining the status quo, making it difficult to redistribute water rights and move away from the inequalities entrenched during apartheid.[9] Third, water use can be authorized under Schedule 1, which covers reasonable domestic water use that has minimal impact on water resources. Lastly, general authorizations permit water use without a formal license, within certain limits and conditions, often benefiting smallholder farmers with a relatively small impact on water resources. Water use is also subject to payment of water tariffs. These tariffs serve to recover operation and maintenance costs of water infrastructure and to manage demand for water under water scarcity. Commercial farmers were required to start paying these fees immediately after the law promulgation, while subsistence farmers were phased into the system over a period of five years, with potential support from government subsidy programs such as the resource-poor farmers subsidy (Speelman et al., 2009).

2.3 Smallholder irrigation schemes

In South Africa, several institutions and policies have been implemented to help smallholder farmers benefit from irrigation. One key initiative has been the establishment of smallholder irrigation schemes, which aim to improve rural livelihoods by providing access to water for agriculture. However, some agronomists argue that the development goals of these schemes have largely remained unfulfilled due to poor farming practices, such as inadequate weed and pest control and failure to increase cropping intensity (Fanadzo, Chiduza and Mnkeni, 2018).

Currently, there are around 330 smallholder irrigation schemes located mostly in the former homelands, covering 50,000 hectares and supporting an estimated 200,000 to 250,000 smallholder farmers. The history of smallholder irrigation development in South Africa can be divided into four distinct periods. First, during the 19th century, the early smallholder irrigation development consisted in colonists transferring irrigation technology to locals, but most such projects had ceased to exist by the end of the century. During the smallholder canal scheme era (1930-–1960), the focus was on providing black families

[9] For instance, see the meeting proceedings from the Water and Sanitation Committee of the National Assembly available at https://pmg.org.za/committee-meeting/23583/. [Accessed on 10 September 2024.] The perceived status quo was also the reason behind the proposal of racial quotas for water allocations by the Minister of Water and Sanitation on 19 May 2023. See, for instance, Bega (8 June 2023).

in the homelands with livelihoods based on farming. Most schemes constructed concrete weirs to divert water from rivers, and some large dams were also built. Next, during the homeland era (1960-–1990), the South African government funded large irrigation schemes with the goal of economic development in the homelands. Over time, this infrastructure proved too costly to maintain and fell into disrepair. Finally, the irrigation management transfer and revitalization era started in the early 1990s with the end of the apartheid regime. Government involvement has decreased, leading to a shift towards local management of irrigation through catchment management agencies, creation of water users associations, and transfers of water rights. The main focus is on repairing the existing infrastructure, devising better maintenance strategies, and training farmers on best cultivation practices.

In summary, there are three key takeaways from the background on irrigated agriculture in South Africa. First, irrigation plays a critical role for the country's agricultural productivity. Second, subsistence farmers were not entirely neglected in the development of irrigated agriculture. Smallholder irrigation schemes were implemented even during the apartheid era, although, the anecdotal evidence suggests that these schemes did not reach their potential. Moreover, subsistence farmers lacked secure water rights. Third, a major institutional shift occurred in the late 1990s following the end of apartheid. New water policy democratized access to water and there was a shift to a more decentralized management of water infrastructure, prioritizing improved maintenance and a stronger emphasis on smallholder participation in irrigation. The central question that remains is how these institutional reforms have impacted both commercial and subsistence farmers in terms of their ability to benefit from irrigation resources.

3 Data and summary statistics

This section first describes the unique geospatial dataset that I assemble, including remotely sensed measures and census variables. Then I proceed to describe the South African water users database used to estimate elasticities of water demand. Finally, I show summary statistics and establish some key facts.

3.1 Data sources

Agricultural outcomes. I estimate the effect of canals on two types of agricultural outcomes: crop productivity and land use. For both types I use remotely sensed measures. The main advantage of remote sensing data is high granularity (Donaldson and Storeygard, 2016), which allows for a precise definition of treated and untreated areas. Agricultural surveys are a more traditional source of data on agricultural outcomes. There are two main issues that prevent me from using this data source in my analysis. First, due to data privacy concerns the smallest geographical unit in publicly released datasets is often a province or a district, which would be too coarse for the spatial RD.[10] Second, agricultural surveys often have a limited sample size and even if I had access to GPS coordinates of each observation, it would be unlikely to have enough units above and below the canals in my analysis. A disadvantage of using remotely sensed data, as opposed to an agricultural survey, is the inability to account for variables that would allow for a more comprehensive analysis, such as use of complementary inputs, costs of labor, or plot area.

Agricultural yields are proxied by Enhanced Vegetation Index (EVI) derived from Landsat 8 satellite imagery[11] which is available at a 30 meters resolution. EVI is calculated from the Near-Infra-Red (NIR), Red and Blue image bands of each scene and ranges in value from -1 to 1, where negative values indicate non-vegetated areas such as water or barren land and positive values indicate dense vegetation. EVI improves on the Normalized Difference Vegetation Index (NDVI), which is another commonly used satellite-derived proxy for agricultural yields, by adjusting for canopy background and reducing atmosphere influences (Huete et al., 2002). It is therefore more suitable for areas with high amounts of biomass, such as irrigated land. Both EVI and NDVI have been shown to be reliable proxies for agricultural yields in various geographical contexts (Lobell et al., 2020; Burke and Lobell, 2017; Asher and Novosad, 2020). South Africa has two agricultural seasons and I define each season's measure of productivity as the maximum value of EVI over the growing and harvesting stages.[12] The main crop of the summer season is maize which is grown and harvested from mid-November until mid-May. The main crop of the winter season is wheat which is grown and harvested from mid-June until the end of November.[13] South Africa also has two distinct climates in terms of when most of the rainfall occurs. West of the country is characterized by winter rainfall with smaller amounts of total precipitation, whereas the East is characterized by summer rainfall with larger amounts of precipitation.[14] This translates into distinct growing seasons for maize in the East and in the West. To ensure comparability of climate relevant to agricultural production, I focus

[10]For instance, the 2017 Census of Commercial Agriculture (Statistics SA, 2020) is available only at the province level. South Africa's Census of Agricultural Households is available at the level of district municipality.

[11]Landsat 8 Collection 1 Tier 1 8-Day EVI Composite. Courtesy of the U.S. Geological Survey.

[12]I am not concerned about contamination by non-crop vegetation because I restrict the analysis to agricultural land by applying a crop mask as discussed in the following paragraph.

[13]See Figure A–5 in the Appendix.

[14]Climate Change Knowledge Portal. South Africa. Current Climate. Climatology. https://climateknowledgeportal.worldbank.org/country/south-africa/climate-data-historical. [Accessed on 31 October 2023.]

my analysis on the Eastern part of the country.[15]

Data on land use comes from the South Africa National Land Cover 2018 (SANLC 2018) and the South Africa National Land Cover 1990/2020 Change (SANLC 1990/2020) released by the South African Department of Rural Development and Land Reform (DRDLR). SANLC 2018 is a raster dataset available at the resolution of 20 meters, generated from automated mapping models using Sentinel 2 satellite imagery for the period of 1 January 2018 to 31 December 2018. SANLC 1990/2020 is based on a comparison of SANLC 1990 and SANLC 2020 datasets.[16] I use the land cover data in two ways. First, I create a crop mask and apply it to the data when analyzing agricultural productivity. This allows me to eliminate concerns about contamination of the productivity measure by non-crop vegetation. Second, I generate key outcomes to study the effect on agricultural production at the extensive margin and on land use change (expansion of land farmed by commercial farmers versus subsistence farmers).

To get a better understanding of the data, Figure A–5 in the Appendix provides a visualization. A raw satellite image is compared with two of the main datasets — SANLC 2018 with different colors representing different land cover classification categories and the Landsat 8 EVI Composite with darker green color representing higher values of the index. In particular, one can notice from the satellite image the location of a dam with the starting point of a canal and several circles around it. These circles are pivot-irrigated fields.[17] In the SANLC 2018 data, these circles are classified as pivot-irrigated annual crops and visualized in dark brown color. Finally, the pivot-irrigated circular fields are also identified in the Landsat 8 EVI Composite and are assigned higher values of the index.

Treatment status. To determine the treatment status I collect GPS coordinates of 144 canals from the South Africa's Department of Water and Sanitation[18] and elevation data from ALOS Digital Surface Model (Tadono et al., 2014), available at a 30 meters resolution. First, I determine the elevation relative to the sea level for each of the 144 irrigation canals. Next, I determine the elevation of each grid cell in my dataset relative to the nearest canal by taking a difference between the elevation of the canal and the elevation of the given grid cell. Areas with positive relative elevation are defined as treated, whereas areas with

[15]I focus on areas that lie east of the 25^{th} meridian. This includes the provinces Limpopo, Mpumalanga, KwaZulu-Natal, Free State, Gauteng, Northwest, and parts of Northern Cape and Eastern Cape.

[16]The 72-class SANLC 1990 was generated from the Landsat 5 imagery for the period of 1989–1991 and is available at the resolution of 30 meters. The 73-class SANLC 2020 was generated from the Sentinel 2 imagery for the period of 1 January 2020 to 31 December 2020 and is available at the resolution of 20 meters.

[17]Circular fields are characteristic of the pivot irrigation technology. A pivot is located in the center and rotates the irrigation equipment (an arm with sprinklers attached to it) around the field.

[18]As explained in the previous footnote, I only focus on canals that lie east of the 25^{th} meridian.

negative relative elevation are considered as control.

Covariates. I gather several geophysical covariates to perform balance tests and verify the validity of RD estimates. I also include these covariates into my preferred specification in order to improve precision. I extract daily precipitation values from the Climate Hazards Center InfraRed Precipitation with Station (CHIRPS) dataset (Funk et al., 2015). The data is available at the resolution of 5,566 meters. I construct a precipitation measure as the total annual precipitation in each cell and then I average this value over the years 2014–2018. Next, I extract the daily land surface temperature at the resolution of 1,000 meters from the MODIS Terra Land Surface Temperature dataset (Wan, Hook and Hulley, 2021). I construct a temperature measure as the monthly maximum temperature averaged over the the years of 2014–2018. The terrain ruggedness index (TRI) comes from Nunn and Puga (2012). It is a measure of topographic variability within a given area that quantifies the variation in elevation between neighboring cells in a digital elevation model. Finally, I calculate distance to the nearest river using WWF HydroSHEDS dataset (Grill et al., 2019).

Homelands. A shapefile containing the boundaries of former homelands has been obtained from the Department of Agriculture, Land Reform, and Rural Development (DALRRD). Notably, some irrigation canals are situated either within these former homelands or just outside their boundaries, which could extend water to the agricultural fields located in the homelands. Figure 1 depicts how the former homelands and the irrigation canals in my sample intersect.

South Africa 2011 Census. For the analysis of socio-economic outcomes, I use the 100% sample of the South Africa 2011 Census available at the small area level.[19]

3.2 Summary statistics

The unit of analysis is a 30 meters by 30 meters grid cell. The grid cells are obtained by converting a geospatial raster dataset with all the above described variables into a delimited text format. The conversion is limited to areas within 10 km of each canal. This results in 56,465,817 observations. The regression discontinuity (RD) sample further restricts the dataset to observations that are within 50 meters of relative elevation to the nearest canal. This results in 29,049,656 observations, out of which 6,795,394 grid cells are classified as agricultural land. Summary statistics for the full sample and the RD sample are presented in Table 1.

Restricting the sample to a narrow bandwidth of 50 meters relative elevation does not result in large differences compared to the full sample, although the land that lies within

[19]A small area is an aggregation of enumeration areas (EAs) and it is the most granular unit made available by the South African statistical office.

50 meters of relative elevation to the nearest canal is more likely to be an agricultural field.[20] Furthermore, columns (2)–(5) present summary statistics disaggregated by the treatment and former homelands status. Treated areas (those lying below canals) have significantly higher values of EVI, they are more likely to consist of agricultural land and grow commercial irrigated annual crops. They are less likely to grow commercial rainfed annuals, be cultivated by subsistence farmers, or lie fallow. There are also significant differences in terms of geophysical covariates. Treated areas have slightly lower maximum monthly temperature (36.2 degrees versus 35.4 degrees) and slightly higher mean annual precipitation (523 mm versus 552 mm). They are also on average closer to the nearest canal (5,250 m versus 5,059 m), and to the nearest river (1,110 m versus 836 m). Although these differences are significant, they do not pose a concern for the RD analysis since the RD assumptions demand an absence of discontinuity at the threshold. In other words, as the bandwidth narrows, these differences should disappear. The test of these assumptions is presented in Table 2.

Former homelands cover 9% of the RD sample. They tend to be slightly more agricultural than non-homelands (23.2% versus 24.4%), but consist almost exclusively of land farmed by subsistence farmers (68%) or fallow land (30%). Only 1% of homelands agricultural land is exploited by commercial farmers. In terms of geophysical characteristics, homelands tend to have higher maximum monthly temperature (35.9 degrees versus 36.5 degrees), higher mean annual rainfall (528 mm versus 579 mm), they are on average further away from the canals (5,129 m versus 6,154 m) and slightly closer to rivers (1,035 m versus 1,28 m). They also lie at a much lower elevation above the sea level (1,109 m versus 846 m).

4 Empirical Strategy

4.1 Previous literature

The main challenge in estimating the impact of irrigation canals is their endogenous placement. For instance, in order to maximize the return on investment, irrigation dams (and the corresponding network of irrigation canals) might be placed in areas with better agricultural potential. Alternatively, irrigation dams might be constructed in politically favored places, which could be correlated with the provision of other goods and services correlated with agricultural yields. A related issue is the lack of historical data for the period before canals were constructed, which prevents the use of panel data methods. Most canals were

[20] I also examine sensitivity of the main results to changing the horizontal bandwidth (10 km) and the vertical bandwidth (50 m).

built in the 1960s and 1970s and the Landsat 8 EVI measures are available only from 1984.

The previous literature employed two main strategies to deal with this endogeneity issue. Early studies relied on an instrumental variable approach. Duflo and Pande (2007) use river gradient which reflects the geographic suitability to predict the distribution of irrigation dams across districts in India. A similar approach was previously employed also in the context of Sub-Saharan Africa. In addition to the river gradient, Strobl and Strobl (2011) distinguish between ephemeral and perennial rivers to predict the distribution of dams across the continent. Ephemeral rivers are considered less suitable for dam construction.

The second empirical strategy is the regression discontinuity (RD) design, which has become feasible only recently due to the availability of high-resolution data on crop yields. In this approach, the running variable is either the distance to the command area boundary (Blakeslee et al., 2023; Jones et al., 2022; Hagerty, 2022) or the relative elevation to the nearest canal (Asher et al., 2023). A key advantage of RD is that it provides causal estimates under weaker assumptions compared to the IV approach. The main assumption for the internal validity of RD estimates is the continuity at the threshold of all the observable and unobservable characteristics that could be correlated with the outcome. In contrast, the IV method requires the assumptions of relevance, exogeneity, and the exclusion restriction, which are more likely to be violated. For example, the instrument used in Duflo and Pande (2007) and subsequent studies——river gradient—— might influence yields directly and not only through the channel of increasing the probability of dam presence. Lower river gradient makes the area more suitable for construction of a dam. However, lower river gradient is also mechanically correlated with the land gradient, which makes the area more suitable for agriculture and also leads to higher yields. This potential correlation would violate the exclusion restriction. Additionally, the IV approach estimates only the local average treatment effect (LATE) for compliers, tat is, the impact of dams built due to favorable geophysical conditions. It does not provide insights into the effects of dams built for other reasons, making the RD approach more appealing in this context. It should be noted, however, that RD also produces local treatment effects, as the estimates are valid only for observations within the selected bandwidth of the running variable.

In this paper I use an RD design with relative elevation to the nearest canal as the running variable. A visual representation of the elevation-based RD design is shown in Figure A–2 in the Appendix. My preferred specification consists of regressing outcome variables on a treatment indicator and a linear function of the running variable while allowing for different slopes in the treatment and the control groups.

4.2 Regression Discontinuity Design

My empirical strategy relies on the fact that surface irrigation is gravity-based. Water from either large irrigation dams or from weirs in rivers is distributed through a canal that carries water toward irrigation schemes located in downhill areas.[21] It is technically possible to pump water uphill from the main canals, and therefore, the running variable does not perfectly determine the treatment. However, land below canals still has a higher probability of being irrigated, which allows for a fuzzy RD design.

The main fuzzy RD specification estimates the local average treatment effect (LATE) of irrigation canals on the agricultural productivity and land use outcomes for areas just below the canals. Following (Imbens and Lemieux, 2008; Gelman and Imbens, 2019), I regress each outcome on the treatment indicator (whether an observation lies below the canal) while controlling linearly for the running variable (relative elevation to the nearest canal) separately on each side of the threshold:

$$\begin{aligned} Y_{ic} =& \beta_0 + \beta_1 Treat_{ic} + \beta_2 Rel\_Elev_{ic} + \beta_3 Rel\_Elev_{ic} \times Treat_{ic} \\ &+ \beta_4 X_{ic} + \mu_c + \epsilon_{ic} \end{aligned} \tag{1}$$

Y_{ic} is the outcome of interest in grid cell i and near canal c. $Treat_{ic}$ is the treatment indicator that is equal to one when a grid cell lies below a canal, which is determined based on its elevation relative to the nearest canal. The running variable Rel_Elev_{ic} is calculated as the elevation of the nearest canal minus the elevation of a given grid cell.[22] Observations with positive values of Rel_Elev_{ic} lie below a canal (treatment group) and observations with negative values lie above a canal (control group). I interact the running variable Rel_Elev_{ic} with the treatment dummy to allow for different slopes at each side of the threshold.

X_{ic} refers to three geophysical covariates that I control for in order to improve precision of the RD estimates: average maximum monthly temperature, average total annual precipitation, terrain ruggedness index, distance to the nearest canal, and distance to the nearest river. Finally, I include district fixed effects μ_c to account for unobserved, time-invariant characteristics that may vary across canals, such as differences in agricultural

[21] See Section 2 for detailed information on surface irrigation infrastructure.

[22] Elevation of each grid cell is determined using the ALOS Digital Surface Model data. The dataset is available at the resolution of 30 meters, which is the same as the resolution of the EVI outcome variable and similar to the resolution of the land use outcome variables (20 meters). The values from the ALOS raster data are thus simply matched to the raster data containing the outcomes. This considerably improves accuracy of the analysis. For instance, Asher et al. (2023) need to determine the elevation of each polygon (village) in their dataset. They do so by taking the 5th percentile of the elevation distribution of the pixels constituting the polygon. However, it is likely that some parts of these polygons (villages) lie above the canal even though they are coded as treated.

policy, institutions or other infrastructure, that may affect the outcome variable.

Standard errors are clustered at the canal level since it is the unit at which the treatment is assigned (Abadie et al., 2023). This also allows for arbitrary spatial correlation in places around each canal. To ensure comparability of treatment and control units, I restrict the sample to grid cells that lie within 10 kilometers of distance and 50 meters of relative elevation to the nearest canal.

4.3 Identifying Assumptions

To interpret the RD estimates as causal effects, two key assumptions must be met. First, the treatment must be at least partially determined by the running variable. In other words, land below a canal should have a higher likelihood of being irrigated, while land above a canal should have a lower likelihood. Figure A–3 in the Appendix illustrates the probability of land being used for irrigated annual crops as a function of the running variable in non-homelands. Unfortunately, a similar figure cannot be generated for subsistence farmers in the former homelands because the land cover classification dataset used does not differentiate between irrigated and non-irrigated subsistence land.

The relationship between the probability of irrigation and the running variable shows no sharp discontinuity at the threshold. This pattern reflects that commercial farmers can pump irrigation water uphill from the canals. For areas above a canal, the cost of pumping increases with vertical distance, as indicated by the rising probability of irrigation as relative elevation approaches the zero threshold. In contrast, for areas below a canal, the absence of pumping costs results in a flat relationship between relative elevation and the probability of irrigation. Thus, this RD specification follows a fuzzy design.

Second, all the observable and unobservable variables that also determine the outcome must be continuous at the threshold. In other words, there should be no sudden discontinuity in any characteristics (weather variables, topography, etc.) between places that are just above and just below a canal. The only source of discontinuity must be access to irrigation through the canals. Note that the purpose is to estimate the *long-run* effect of irrigation canals. In the long-run, farmers might adjust their input use to take into account the increase in water supply due to irrigation (Hagerty, 2022). Therefore, a discontinuity at the threshold in, for instance, the use of fertilizer would not be a threat to identification, but rather one of the components of the long-run effect. In order to better articulate this concept, suppose that crop yields $Y(W, I)$ are a function of water supply W and other inputs I (fertilizer, pesticides, labor, etc.). We assume that in the long-run, inputs are adjusted to the level of water supply. The effect of water supply on crop yields is then given by the total derivative of $Y(W, I)$ with respect to W which consists of a direct effect and an indirect effect:

$$\frac{dY(W,I)}{dW} = \underbrace{\frac{\partial Y(W,I)}{\partial W}}_{\text{direct effect}} + \underbrace{\frac{\partial Y}{\partial I}\frac{dI(W)}{dW}}_{\text{indirect effect}} \tag{2}$$

In this paper, I estimate the total effect which includes both the direct and indirect effects.

To verify the smoothness assumption, I examine five geophysical characteristics (average maximum monthly temperature, average total annual precipitation, terrain ruggedness index, distance to the nearest canal, and distance to the nearest river) and three variables related to soil quality (proportion of sand, silt, and clay particles). Figure 2 presents binned scatter plots for the first five variables, residualized on the canal fixed effects. Table 2 reports statistical tests of the presence of discontinuity at the threshold. The estimates are derived by estimating Equation 1 on the sample of grid cells within 10 kilometers of a canal and within 50 meters of relative elevation. I perform these tests on both agricultural and full samples. In both samples and for both homelands and non-homelands, there are no large and significant discontinuities in terms of ruggedness, temperature, precipitation, and soil quality. However, for the two distance variables — distance to the nearest canal and distance to the nearest river — I find significant discontinuities. To address this, the main specification controls for distance to the nearest canal, distance to the nearest river, and other weather and terrain covariates. Soil quality variables are controlled for in robustness checks due to a high number of missing values.

5 Results

5.1 Agricultural seasons in South Africa

South Africa has two distinct agricultural seasons during which winter and summer crops are grown. Winter crops are typically planted during the autumn months of April and May, and then harvested during the winter and early spring months of September to November. Wheat is the most important winter crop with a production of 2,263,000 tonnes during the 2021/2022 season (SAGIS, Monthly Producer Deliveries). Most of the production is centered in Western Cape.

Summer crops are typically planted in the early summer months of October to December, and are harvested during the late summer and early autumn months of February to April. The most important summer crop is maize with a production of 15,810,000 tonnes during the 2021/2022 season (SAGIS, Monthly Producer Deliveries). Maize is produced in almost all parts of South Africa, but the majority of production is concentrated in the Free State, Mpumalanga, North West, and KwaZulu-Natal provinces. The planting, growing, and harvesting months are summarized in Figure A–5 in the Appendix. I use the temporal ranges of growing and harvesting seasons of wheat and maize (east) to construct the EVI measures of agricultural productivity for the wheat and maize seasons.

5.2 The intensive margin effects: canals increase crop yields in non-homelands

I first examine the effects of irrigation canals on the intensive margin of agricultural production. The results of estimating equation 1 are shown in Table 3. The regressions are estimated separately for non-homelands (Panel A) and the former homelands (Panel B). I restrict the analysis to only agricultural grid cells to ensure that results are not contaminated by non-crop vegetation.

In non-homelands, I find that irrigation canals increase agricultural productivity by 3.9% during the wheat season and by 2.3% during the maize season (columns (1) and (2)). Note that these are reduced-form estimates that are not scaled by the change in probability of irrigated land at the threshold.

In the former homelands, where mostly subsistence farmers operate (as established in Table 1), irrigation canals do not have the same effects. For both the maize and wheat seasons, the estimates are negative, though the wheat season results are not statistically significant. There are two possible explanations for this outcome. Either subsistence farmers are entirely excluded from water institutions and therefore lack access to water from large-scale irrigation schemes, or they are granted water use licenses but are unable to demand sufficient quantities due to market failures that hinder their ability to make profitable investments.

5.3 The extensive margin effects: canals increase the area under production in non-homelands

Next, I consider how irrigation canals affect the extensive margin of agricultural production. In other words, do irrigation canals increase cultivated area? To answer this question, I estimate equation 1 on the full RD sample (*not* restricting to agricultural land only). Again, the regression is estimated separately for non-homelands (Panel A) and the former homelands (Panel B).

The results are reported in Table 3. In non-homelands, areas below canals have a 21% higher probability of being cultivated with annual crops (an increase of 3.6 pp relative to the mean in the control group of 16.9%). On the other hand, the land below canals in the former homelands is not more likely to be cultivated with annuals, although it is

47% less likely to lie fallow (a decrease of 4.2 pp relative to the mean in the control group of 9%). There is no statistically significant effect on the likelihood of fallow land in non-homelands. These findings are consistent with the previously discussed crop yield effects. If subsistence farmers located downstream from irrigation canals do not benefit from higher yields, then they are less motivated to expand area under production than commercial farmers in non-homelands.

Interestingly, the presence of canals does not increase the probability of cultivation of perennial crops, such as fruit trees, vines, and sugarcane, that tend to be water-intensive.

5.4 The extensive margin effect is explained by the post-apartheid land use changes

The problem with the results in the previous section is that the higher probability of cultivated land could simply reflect a selection bias. Large dams and canals could have been built so as to benefit existing fields and to maximize the amount of area cultivated downstream. Ideally, I would like to examine the RD effects before a dam was built to establish that the probability of annual crops being grown is the same above and below the future dam. However, I do not have land use data for the period before most large-scale irrigation projects were completed.[23]

An alternative strategy is to consider the change in land use outcomes between 1990 and 2020 which is derived from the SANLC 1990/2020 Change dataset. Before the end of apartheid in 1994 water access was based on riparian rights, which means that only landowners could extract water from their property. The new post-apartheid water legislation aimed to democratize access to water with a particular goal of redressing past racial inequalities and introduced a system of water licensing. Therefore, one would expect an expansion of cultivated land below canals as new farmers can now acquire water rights.

Table 4 shows the RD effects on the expansion of different type of farming between 1990 and 2020, separately for homelands and non-homelands. The effects are estimated for four outcomes: expansion of agricultural land with annual crops, expansion of commercial pivot irrigated land, expansion of commercial non-pivot irrigated or non-irrigated land[24], and expansion of subsistence land. In non-homelands, treated areas below a canal experienced a 20% higher expansion of agricultural land with annual crops than control areas above a canal (an increase of 0.7 pp relative to the mean in the control group of 3.5%). On the other hand, there was no such expansion in the former homelands. Irriga-

[23]Most large irrigation dams were completed by the 1960s and the 1970s. See Figure A–1 in the Appendix.

[24]These two categories of commercial land are lumped together because SANLC 1990 data cannot distinguish between irrigated and non-irrigated land without the circular shape of the fields. As explained earlier, center pivot irrigation results in easily distinguishable circular patterns.

tion canals, therefore, contribute to the expansion of the commercial agriculture in South Africa but do not affect subsistence farmers.

5.5 Robustness checks

I perform several robustness checks to test the sensitivity of my main results to different empirical choices.

Extending the analysis to multiple years. A potential concern is the "temporal" external validity of the effects on crop yields, as I only use data for one year. It is possible that the negative effects of irrigation canals on subsistence farmers would not be observed in other years. In particular, 2018 was an unusual year due to a country-wide drought, and it is plausible that subsistence farmers situated below canals are negatively impacted only during droughts.[25] To address this issue, I extend the analysis to multiple years.

Rather than relying solely on SANLC 2018, which covers only one year, I generate a new crop mask based on the Global Cropland Maps produced by the University of Maryland Global Land Analysis & Discovery (UMD GLAD) for the period 2000-–2019 (Potapov et al., 2022). I utilize all five UMD GLAD datasets, each covering a four-year period.[26] When I compare the crop masks generated by UMD GLAD 2019 and SANLC 2018, I find them to be mostly consistent.[27] I estimate an equation similar to 1, with the main difference being that the observations are now indexed by year and year fixed effects are included in the regressions.

The results in Table A–1 indicate that the negative effect of irrigation canals in the former homelands persists even after extending the analysis to multiple years. In non-homelands, irrigation canals continue to have a positive effect on yields, although the magnitude is smaller and not statistically significant for maize. This is likely because the UMD GLAD crop mask includes perennials, which attenuates the main effect since there is

[25]During droughts, there is typically less water available in the irrigation system. Commercial farmers might be prioritized over subsistence farmers for access to the limited water supplies, which could lead to lower yields for subsistence farmers below canals compared to those above canals, who usually do not have access to any water from the canal networks.

[26]I use UMD GLAD 2000, 2004, 2007, 2011, 2015, and 2019. For example, UMD GLAD 2015 classifies a grid cell as cropland if active crops were detected during the years 2012—2015. Potapov et al. (2022) define cropland as "land used for annual and perennial herbaceous crops for human consumption, forage (including hay), and biofuel." (p.19)

[27]In non-homelands, 87% of grid cells classified as cropland by UMD GLAD 2019 are also classified as agricultural in my previous analysis using SANLC 2018. In homelands, only 37% of grid cells classified as cropland by UMD GLAD 2019 are also classified as agricultural in my previous analysis. This discrepancy is mainly due to the presence of sugarcane, which is correctly detected as cropland by UMD GLAD but is not included in my previous analysis, where I focus only on annual crops. For a more detailed comparison of the SANLC and UMD GLAD datasets, Table A–2 in the Appendix provides a breakdown of NLC categories for both homeland and non-homeland detected cropland.

no seasonal variation in the biomass of perennial crops.

Different vertical bandwidths. The results in Tables A–3 and A–4 show robustness to different vertical bandwidths of four main outcomes (log wheat EVI, log maize EVI, the probability of land being agricultural, and the probability of annual crops being grown). To ensure comparability of treated and untreated land, my preferred specification restricts the sample to grid cells that are within 50 m of relative elevation to the nearest canal.

In Tables A–3 and A–4 I change the vertical bandwidth to 25 m, 75 m, and 100 m. The wider bandwidths lead to RD estimates that are larger in magnitude and highly statistically significant. In the former homelands, the 25 m bandwidth also generates estimates consistent with the previous results. On the other hand, in non-homelands, the 25 m bandwidth generates smaller and statistically insignificant estimates. This can be explained byt two factors. First, the 25 m vertical bandwidth leads to a smaller sample size. Second, it is possible that a majority of commercial farmers above canals up to 25 m of relative elevation are able to pump water uphill from the distribution point to their fields, and so there is a small difference in the probability of land being irrigated on both sides of the threshold.

Different horizontal bandwidths. Tables A–5 and A–6 display robustness to decreasing the horizontal distance to the nearest canal. After restricting the distance to 2,500 m, 5,000 m, and 7,500 m, the results remain similar to those of the preferred specification, although the crop yield effects lose statistical significance in the former homelands, which is most likely due to the lower sample sizes.

Excluding a donut hole. As previously mentioned, commercial farmers located just above irrigation canals can pump water uphill from the distribution point to their fields. A more appropriate empirical approach might therefore be a donut RD, where observations immediately around the cutoff are excluded (Barreca et al., 2011). Additionally, the donut RD approach could help mitigate measurement error bias that might otherwise attenuate my main results.[28] The donut RD results reported in Tables A–7 and A–8. As the size of the donut hole increases, the RD estimates also increase in magnitude, but the main finding remains consistent: commercial farmers in non-homelands benefit from higher yields and land expansion, whereas subsistence farmers do not. While the larger estimates may be due to the reduction of measurement error and the consequent attenuation bias, recent literature suggests that donut RD estimates could be biased under the standard RD assumptions (Noack and Rothe, 2023). Therefore, I present these results as robustness checks.

[28]The measurement error could come from an imprecise determination of elevation of grid cells. Grid cells that have approximately the same elevation as the nearest canal might be assigned to the wrong treatment status.

Controlling for soil quality and different level of FE. Finally, I check that the main results are not biased by the endogenous placement of canals so as to benefit soil of better (or worse) quality. I already established that the soil quality variables are continuous around the cutoff (Table 2) but I also include these as additional covariates in my RD regressions. In Tables A–9 and A–10, I control for additional soil quality indicators, including the proportion of sand, silt, and clay particles in the fine earth fraction, which determine the type of soil, as well as soil carbon and nitrogen content.[29] The results I qualitatively unchanged. The reason why I do not include soil quality variables in the main specification is that there are missing values for some grid cells, which would exclude these observations from the analysis. I also present findings from models that exclude any covariates and from those that adjust the level of fixed effects from canals to local municipalities.

5.6 Local labor markets: canals decrease unemployment in non-homelands

In the previous sub-sections I established that irrigation canals increase crop yields and area under production for commercial farmers located in non-homelands. But do the benefits from irrigation extend beyond the population of large landowners? The expansion of commercial farms could benefit the rural poor through job creation and thus contribute to a decrease in rural poverty.

To understand the impact of irrigation canals on local labor markets, I employ a spatial RD strategy similar to the one used earlier. The main difference lies in the unit of analysis. While the remotely sensed measures of crop yields and land use allowed for a more granular analysis, the outcome variables derived from the census are available at a coarser level. This presents a challenge in determining the treatment status of each census enumeration unit (small area), as these units are larger than the 30 m by 30 m grid cells used previously. To define treatment status, I follow the approach of Asher et al. (2023), using the fifth percentile of the elevation distribution within each small area and comparing it to the elevation of the nearest canal. I then estimate the following equation:

$$\begin{aligned} Y_{ac} =& \alpha_0 + \alpha_1 Treat_{ac} + \alpha_2 Rel\_Elev_{ac} + \alpha_3 Rel\_Elev_{ac} \times Treat_{ac} \\ &+ \mu_c + \epsilon_{ac} \end{aligned} \tag{3}$$

where Y_{ac} is the outcome variable averaged over households living in small area a near canal c, $Treat_{ac}$ is the treatment indicator equal to one if the fifth percentile of the

[29] I include the sand, silt, and clay particles and the carbon and nitrogen content as two separate sets of soil control variables since the latter are potentially endogenous to the presence of irrigation and might bias the results.

distribution of elevation grid cells constituting a small area is less than the elevation of the nearest canal, and Rel_Elev_{ac} is the relative elevation to the nearest canal. As previously, the regression equation includes canal fixed effects and standard errors are clustered at the canal level. The sample is restricted to small areas that lie within 10 km of horizontal distance and 50 m of vertical distance to the nearest canal. α_3 is the coefficient of interest.

Table 5 shows the RD estimates of the effect of irrigation canals on employment and household income. In non-homelands, the unemployment rates decrease by 4.9 pp (20% relative to the mean of the control group) and the fraction of households with zero income decreases by 2.7 percentage points (19% relative to the mean of the control group). However, the results are significant only in non-homelands. In the former homelands, unemployment and the proportion of households with zero income is also lower (by 5% and 6% respectively), but the effects are much smaller in magnitude and not statistically significant. Furthermore, irrigation canals have an imprecisely estimated positive effect on the average annual household income per capita[30] and the average annual household income per capita among the bottom 10% of the income distribution.

These results suggest that, at least in non-homelands, commercial farms do generate new employment opportunities for rural poor, and thus, it can be argued that large-scale irrigation infrastructure in South Africa contributes to the decrease in poverty. This result is in line with previous findings in India (Duflo and Pande, 2007), Mali (Dillon, 2011), and Sri Lanka (Sellamuttu et al., 2014). The link between employment and poverty is not straightforward since the quality of jobs matters. If the jobs generated by the irrigation dams are stable and well-paying, this is more likely to lead to a significant reduction in poverty. However, if the jobs are seasonal and low-paying, the impact on poverty might be less pronounced. Nonetheless, employment is a major determinant of income, and a reduction in unemployment generally leads to higher household incomes, especially in rural areas where job opportunities might be limited. Increased employment can therefore contribute to poverty reduction, as more people have access to wages. This claim can also be supported by the decrease in the proportion of households with zero income.

Poverty is one dimension of welfare that allows us to understand the economic and social impact of large-scale irrigation infrastructure. To further understand their overall welfare impacts, I also conduct a cost-benefit analysis (Section 6). While the reduction in unemployment and its potential link to poverty alleviation is an important indicator of the benefits of dams, a cost-benefit analysis provides a more comprehensive evaluation by considering direct economic benefits (higher agricultural productivity and increased area under production) and direct costs of construction. This method can help provide a clearer

[30] The average is at the census small area level

picture of whether the benefits of these irrigation projects outweigh their costs.

6 Cost-benefit analysis

High costs and long periods of construction often deter investments into large irrigation dams. A natural question that arises is whether the delayed benefits justify the large upfront expenses.

The goal of this section is to compare the costs of building an irrigation dam to the returns from such an investment. To define benefits, I take advantage of the RD estimates of the effect of dams on annual crop yields. In non-homelands, I found that irrigation dams increase wheat yields by 3.9% on average, and maize yields by 2.3%. This constitutes an intent-to-treat (ITT) effect. To derive the average treatment effect on the treated (ATT), I scale the ITT by the "first stage", that is, by the increase of probability in land being irrigated below a canal. I find this first stage effect to be 8.9% (Figure 1). Therefore, the ATT effects are 43% for wheat and 25% for maize. In the former homelands, I found that irrigation dams decrease maize yields by 2.6%. The effect on wheat yields is not statistically significant and so I consider it to be 0%.[31] As costs, I use estimates from a feasibility study performed by engineers and commissioned by the South African government to evaluate possible solutions to water supply shortages in the Limpopo province.

The feasibility study evaluates several options and recommends the best one based on the cost of one cubic meter of water supplied. The recommended option is the Paswane Dam on Mutshindudi River. I choose to perform the cost-benefit analysis based on the Paswane Dam and its associated costs since it is likely that the government would choose the option that allows them to supply one cubic meter for the lowest cost. Implementing this infrastructure requires two types of costs: capital costs and operation and maintenance (O&M) costs. Capital costs are further sub-divided into the cost of building the dam, professional fees, social and environmental costs, cost of infrastructure replacement, and cost of land acquisition and relocation. Capital costs are incurred over the duration of the construction, which is ten years in case of the Paswane Dam.[32] O&M costs accrue every year since the dam becomes operational (year 10) until the end of the life of the project

[31] Note that the land cover classification data set does not distinguish between irrigated and non-irrigated subsistence land. I will therefore use the ITT instead of ATT to capture the (dis)benefits that accrue to subssitence farmers.

[32] I assume that professional fees, social and environmental costs and cost of land acquisition and relocation are paid immediately in the first period, one third of the dam cost is paid in year 3 and the remaining two thirds are paid in year 10, and the infrastructure replacement cost is split in half and paid 15 years and 30 years after the dam becomes operational. I make these assumptions to match the estimates of discounted capital costs from the feasibility study as closely as possible. The authors of the study do not specify how these costs are spread over the construction period.

(45 years later). The detailed costs discounted at 3% are shown in Table 6.

The estimation of benefits relies on several assumptions. First, RD identifies only local treatment effects. For the purpose of this cost-benefit analysis I assume that treatment effects extend to all treated areas. Second, I assume there are no general equilibrium effects. For example, changes in crop prices due to an increased production are not allowed. Third, I assume that the Paswane dam is built for irrigation purposes. In practice, a dam can also supply water to industry, but I do not have estimates of treatment effects on non-agricultural sectors. Therefore, net benefits should capture an increase in the value of agricultural production that is causally attributable to an irrigation dam. Finally, I assume that irrigation water from the dam is used to grow annual crops. This is a more conservative scenario since annual crops are relatively low-value compared to other potential uses of irrigation water (vegetables, vines, nut trees). This assumption is also in line with this paper's finding that irrigation dams lead to an expansion of land dedicated to growing annual crops, but not perennial crops.

The Paswane dam would be able to supply 43 million cubic meters per year. I further assume that only 99% of this volume would be used for irrigation, whereas 1% would be for domestic use purposes. This split is the same as for other existing irrigation dams in the area.[33] Furthermore, from the water user database provided by the DWS, I found that the average irrigation water use in the Limpopo province is 14,215 cubic meters per hectare. This implies that the Paswane Dam could irrigate 2,995 hectares of agricultural land. This land would likely be split between commercial and subsistence farmers. I assume that 12% of land is cultivated by subsistence farmers and the remaining 88% by commercial farmers.[34] Finally, I assume that the land used for growing annuals is split between white maize (51%), yellow maize (36%), and wheat (13%). These proportions are based on the national average from the 2017 Census of Commercial Agriculture.

Next, I quantify the additional production in tons for commercial farmers. According to the 2017 Census of Commercial Agriculture, the average yield of rainfed white maize production is 4.54 t/ha. With an ATT of 25%, the additional yield that is attributable to irrigation on existing fields is 1.07 t/ha. Similarly, the additional yields for yellow maize and wheat are 1.26 t/ha and 1.03 t/ha respectively. I also found that irrigation dams lead to an increase in area under production. In such cases, the additional yields attributable to irrigation on new fields is 5.31 t/ha for white maize, 6.26 t/ha for yellow maize, and 3.46 t/ha for wheat.[35] Note that the expansion of area under production is not captured

[33] Notably the Albasini Dam and the Tzaneen Dam.

[34] This is the average split across the 143 canals in my main analysis sample, that is, within 10 km of distance from each point.

[35] 4.54 t/ha + 1.07 t/ha for white maize. 5 t/ha + 1.26 t/ha for yellow maize. 2.43 t/ha + 1.03 t/ha for

by the ATT effect on crop yields since I apply a crop mask and, therefore, I only compare productivity on agricultural fields.[36] To take into account the fact that irrigation dams lead to an expansion of the area under production, I use my RD estimate of the treatment effect at the extensive margin (increase of the probability of land being cultivated by 26.8%) and assume that a part of the land cultivated by commercial farmers existed before the dam construction, whereas the other part is newly established fields.[37]

In my previous analysis, I find that downstream subsistence farmers do not benefit from irrigation dams and even achieve lower maize yields. This translates into a loss of 0.11 t/ha for white maize and 0.13 t/ha for yellow maize. The RD effect on wheat yields is not statistically significant, therefore, I assume that the effect on wheat production is null. I also found no effect on the extensive margin for subsistence farmers, and thus, all of the land cultivated by subsistence farmers is considered to be existing fields.

Finally, to value the additional production resulting from a new irrigation dam, I collect domestic future prices of maize and wheat from the South Africa Futures Exchange (SAFEX).[38] I convert the prices to the 2014 South African Rand using the Consumer Price Index and take the average over the past five years (2019–2023).[39] The additional crop production is then valued at 4,916 R/t for white maize, 4,931 R/t for yellow maize, and 8,021 R/t for wheat. The benefits are discounted using a 3% discount rate and assuming that the production starts after the dam becomes operational (year 10) and lasts for 45 years.

Table 7 reports the benefit-cost ratios for the baseline scenario described above and two different counterfactual scenarios. The purpose of these counterfactual scenarios is to assess the sensitivity of the cost benefit analysis to changes in some key assumption.

In the baseline scenario, the project exhibits a benefit-cost ratio of 1.62, implying that large irrigation infrastructure is a cost-effective investment for economic development. The result holds under a more conservative discounting. The benefit-cost ratio in counterfactual 1 with a discount rate of 4% remains well above 1. I also examine the sensitivity to the absence of the extensive margin effect. In counterfactual 2, I assume that the dam benefits only existing agricultural land and no new fields are set up. In this scenario, the

wheat.

[36]If I did not apply the crop mask, then the estimated treatment effect would also capture the fact that areas below canals are more likely to be agricultural fields since the productivity measure would average over cultivated and uncultivated land.

[37]Specifically, I assume that existing fields comprise 2,068 ha and the new fields comprise 554 ha, an increase of 26.8% relative to the existing fields.

[38]Historic domestic future prices are made available by the South African Grain Information Service (SAGIS): at https://www.sagis.org.za/safex_historic.html. [Accessed on 25 July 2024.]

[39]The CPI is available at https://www.statssa.gov.za/publications/P0141/CPIHistory.pdf. [Accessed on 30 July 2024.]

benefit-cost ratio falls below 1. In other words, if the dam benefits only existing fields, the investment may not be justified. Policymakers should consider this margin when assessing cost-effectiveness of large irrigation dams.

7 Conclusion

The findings of this paper offer insights into the distributional impacts of large-scale irrigation infrastructure on different types of farmers in South Africa. The results suggest that while irrigation canals significantly boost agricultural productivity in commercial farming areas, subsistence farmers in the former homelands do not experience the same benefits. This disparity is evident in both intensive and extensive margins of agricultural production. Although canals lead to higher yields and expanded cultivation for commercial farms downstream, subsistence farmers see a decline in yields. These findings raise questions about the effectiveness of the 1998 National Water Act in redressing past racial inequalities in water access — one of the stated goals of this post-apartheid water policy.

The cost-benefit analysis shows that large-scale irrigation projects can be a cost-effective investment despite the negative externalities. The benefit-cost ratios are larger than one under the discount rates of 3% and 4%. However, the negative externalities imposed on subsistence farmers in the former homelands warrant attention. Policymakers should ensure that the benefits of such infrastructure projects are shared equitably, either through water tariff subsidies or through encouraging income diversification. Furthermore, policymakers should consider whether new irrigation dams are built only to boost productivity on existing agricultural land, or whether they allow for conversion of uncultivated land. The results of the cost-benefit analysis suggest that the investment might not be cost-effective if it benefits only established farms.

Further research could investigate the role of commercial farming expansion in driving structural transformation, the process by which an economy shifts from being agriculture-based to one centered on manufacturing and services. Some argue that the commercialization of agriculture can stimulate industrialization in agrarian economies (Suri and Udry, 2022). For example, a shift from subsistence farming, where smallholder farmers grow crops for own consumption, to commercial farming, where large-scale farmers produce crops for sale, can generate sufficient and affordable food surpluses to support the labor force in the emerging manufacturing sector. Additionally, smallholder farms often face inefficiencies and low productivity. Collier and Dercon (2014) suggest that substantial changes in the organization of Africa's agricultural sector are essential for economic development, and these changes may conflict with the common focus on improving smallholder productivity as a means of reducing poverty. Commercialization, along with increasing farm sizes, could be a necessary step.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. 2023. "When Should You Adjust Standard Errors for Clustering?*." *The Quarterly Journal of Economics*, 138(1): 1–35.

Altchenko, Y., and K. G. Villholth. 2015. "Mapping irrigation potential from renewable groundwater in Africa – a quantitative hydrological approach." *Hydrology and Earth System Sciences*, 19(2): 1055–1067.

Asher, Sam, Alison Campion, Douglas Gollin, and Paul Novosad. 2023. "The Long-Run Development Impacts of Agricultural Productivity Gains: Evidence from Irrigation Canals in India." *Working Paper*.

Asher, Sam, and Paul Novosad. 2020. "Rural Roads and Local Economic Development." *American Economic Review*, 110(3): 797–823.

Barreca, Alan I., Melanie Guldi, Jason M. Lindo, and Glen R. Waddell. 2011. "Saving Babies? Revisiting the effect of very low birth weight classification." *The Quarterly Journal of Economics*, 126(4): 2117–2123.

Bega, Sheree. 8 June 2023. "Race-based water licence allocations draw ire of farmers." *Mail & Guardian*. https://mg.co.za/news/2023-06-08-race-based-water-licence-allocations-draw-ire-of-farmers/#:~:text=South%20Africa's%20biggest%20federation%20of,food%20security%20at%20further%20risk%E2%80%9D.

Blakeslee, David, Aaditya Dar, Ram Fishman, Samreen Malik, Heitor S. Pellegrina, and Karan Singh Bagavathinathan. 2023. "Irrigation and the spatial pattern of local economic development in India." *Journal of Development Economics*, 161: 102997.

Blanc, Elodie, and Eric Strobl. 2014. "Is Small Better? A Comparison of the Effect of Large and Small Dams on Cropland Productivity in South Africa." *The World Bank Economic Review*, 28(3): 545–576.

Burke, Marshall, and David B. Lobell. 2017. "Satellite-based assessment of yield variation and its determinants in smallholder African systems." *Proceedings of the National Academy of Sciences*, 114(9): 2189–2194.

Carrillo, Bladimir, Carlos Charris, and Wilman Iglesias. 2023. "Moved to Poverty? A Legacy of the Apartheid Experiment in South Africa." *American Economic Journal: Economic Policy*, 15(4): 183–221.

Cattaneo, Matias D., Richard K. Crump, Max H. Farrell, and Yingjie Feng. 2024. "On Binscatter." *American Economic Review*, 114(5): 1488–1514.

Chikoore, Hector, and Mark R. Jury. 2021. "South African drought, deconstructed." *Weather and Climate Extremes*, 33: 100334.

Collier, Paul, and Stefan Dercon. 2014. "African Agriculture in 50Years: Smallholders in a Rapidly Changing World?" *World Development*, 63: 92–101. Economic Transformation in Africa.

Dennis, H. J., and W. T. Nell. 2002. "Precision irrigation in South Africa." *Paper prepared for presentation at the 13th International Farm Management Congress, Wageningen, The Netherlands, July 7-12, 2002*.

Dillon, Andrew. 2011. "The Effect of Irrigation on Poverty Reduction, Asset Accumulation, and Informal Insurance: Evidence from Northern Mali." *World Development*, 39(12): 2165–2175.

Dillon, Andrew, and Ram Fishman. 2019. "Dams: Effects of Hydrological Infrastructure on Development." *Annual Review of Resource Economics*, 11(1): 125–148.

Donaldson, Dave, and Adam Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives*, 30(4): 171–98.

Duflo, Esther, and Rohini Pande. 2007. "Dams." *The Quarterly Journal of Economics*, 122(2): 601–646.

Fanadzo, M., C. Chiduza, and M.S. Mnkeni. 2018. "Overview of smallholder irrigation schemes in South Africa: Relationship between farmer crop management practices and performance." *International Journal of Irrigation and Water Management*, 5(12): 1–10.

Funk, Chris, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, James Rowland, Laura Harrison, and Andrew Hoell and Joel Michaelsen. 2015. "The climate hazards infrared precipitation with stations-a new environmental record for monitoring extremes." *Scientific Data 2, 150066*.

Gelman, Andrew, and Guido Imbens. 2019. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics*, 37(3): 447–456.

Grill, G., B. Lehner, M. Thieme, B. Geenen, D. Tickner, F. Antonelli, S. Babu, P. Borrelli, L. Cheng, H. Crochetiere, and H.E. Macedo. 2019. "Mapping the world's free-flowing rivers." *Nature*, 569(7755): 215.

Hagerty, Nick. 2022. "Adaptation to Surface Water Scarcity in Irrigated Agriculture." *Working Paper*.

Hornbeck, Richard, and Pinar Keskin. 2014. "The Historically Evolving Impact of the Ogallala Aquifer: Agricultural Adaptation to Groundwater and Drought." *American Economic Journal: Applied Economics*, 6(1): 190–219.

Huete, A, K Didan, T Miura, E.P Rodriguez, X Gao, and L.G Ferreira. 2002. "Overview of the radiometric and biophysical performance of the MODIS vegetation indices." *Remote Sensing of Environment*, 83(1): 195–213. The Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring.

Imbens, Guido W., and Thomas Lemieux. 2008. “Regression discontinuity designs: A guide to practice.” *Journal of Econometrics*, 142(2): 615–635. The regression discontinuity design: Theory and applications.

IPCC. 2018. “Global warming of 1.5 °C.” IPCC (Intergovernmental Panel on Climate Change).

Johnston, Bruce F., and John W. Mellor. 1961. “The Role of Agriculture in Economic Development.” *The American Economic Review*, 51(4): 566–593.

Jones, Maria, Florence Kondylis, John Loeser, and Jeremy Magruder. 2022. “Factor Market Failures and the Adoption of Irrigation in Rwanda.” *American Economic Review*, 112(7): 2316–52.

Lewis, W. A. 1954. “Economic Development with Unlimited Supplies of Labour.” *The Manchester School*.

Lichtenberg, Erik. 1989. “Land Quality, Irrigation Development, and Cropping Patterns in the Northern High Plains.” *American Journal of Agricultural Economics*, 71(1): 187–194.

Lobell, David B, George Azzari, Marshall Burke, Sydney Gourlay, Zhenong Jin, Talip Kilic, and Siobhan Murray. 2020. “Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis.” *American Journal of Agricultural Economics*, 102(1): 202–219.

Lochmann, Alexia. 2022. “Diagnosing Drivers of Spatial Exclusion: Places, People, and Policies in South Africa’s Former Homelands.”

Mettetal, Elizabeth. 2019. “Irrigation dams, water and infant mortality: Evidence from South Africa.” *Journal of Development Economics*, 138: 17–40.

Noack, Claudia, and Christoph Rothe. 2023. “Donut Regression Discontinuity Designs.” *Unpublished*.

Nunn, Nathan, and Diego Puga. 2012. “Ruggedness: The Blessing of Bad Geography in Africa.” *Review of Economics and Statistics*, 94(1): 20–36.

Olmstead, Sheila M., and Hilary Sigman. 2015. “Droughts, dams, and economic activity.” *Working Paper*.

Oqubay, Arkebe, Fiona Tregenna, and Imraan Valodia, ed. 2020. *The Oxford Handbook of the South African Economy.* Oxford University Press.

Perret, Sylvain R. 2002. “Water policies and smallholding irrigation schemes in South Africa: a history and new institutional challenges.” *Water Policy*, 4(3): 283–300.

Potapov, Peter, Svetlana Turubanova, Matthew C. Hansen, Alexandra Tyukavina, Viviana Zalles, Ahmad Khan, Xiao-Peng Song, Amy Pickens, Quan Shen, and Jocelyn Cortez. 2022. “lobal maps of cropland extent and change show accelerated cropland expansion in the twenty-first century.” *Nature Food*, 3: 19–28.

Ranis, Gustav, and John C. H. Fei. 1961. "A Theory of Economic Development." *The American Economic Review*, 51(4): 533–565.

Reinders, FB, I Van der Stoep, NL Lecler, KR Greaves, JT Vahrmeijer, N Benade, FJ Du Plessis, PS Van Heerden, JM Steyn, B Grove, A Jumman, and G Ascough. 2010. "Standards and Guidelines for Improved Efficiency of Irrigation Water Use from Dam Wall Release to Root Zone Application: Main Report." Water Research Commission.

Reinders, Felix B. 2011. "Irrigation methods for efficient water application: 40 years of South African research excellence." *Water SA*, 37(5).

Sellamuttu, Sonali Senaratna, Takeshi Aida, Ryuji Kasahara, Yasuyuki Sawada, and Deeptha Wijerathna. 2014. "How Access to Irrigation Influences Poverty and Livelihoods: A Case Study from Sri Lanka." *The Journal of Development Studies*, 50(5): 748–768.

Speelman, Stijn, Jeroen Buysse, Stefano Farolfi, Aymen Frija, Marijke D'Haese, and Luc D'Haese. 2009. "Estimating the impacts of water pricing on smallholder irrigators in North West Province, South Africa." *Agricultural Water Management*, 96(11): 1560–1566.

Statistics SA, Pretoria: Statistics South Africa. 2020. "Census of commercial agriculture 2017."

Strobl, Eric, and Robert O. Strobl. 2011. "The distributional impact of large dams: Evidence from cropland productivity in Africa." *Journal of Development Economics*, 96(2): 432–450.

Suri, Tavneet, and Christopher Udry. 2022. "Agricultural Technology in Africa." *Journal of Economic Perspectives*, 36(1): 33–56.

Tadono, T., H. Ishida, F. Oda, S. Naito, K. Minakawa, and H. Iwamoto. 2014. "Precise Global DEM Generation By ALOS PRISM." *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-4: 71–76.

Wan, Z., S. Hook, and G. Hulley. 2021. "MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V061." Distributed by NASA EOSDIS Land Processes Distributed Active Archive Center.

WCD, World Commission on Dams. 2000. "Dams and Development: A New Framework for Decision-Making." London: EarthScan Publications.

World Bank. 2010. "World Development Report 2010: Development and Climate Change. Chapter 3: Managing Land and Water to Feed Nine Billion People and Protect Natural Systems." World Bank.

WRC. 2024. "Exploring South African Irrigation History." Water Research Commission.

Zaveri, Esha, Jason Russ, and Richard Damania. 2020. "Rainfall anomalies are a significant driver of cropland expansion." *Proceedings of the National Academy of Sciences*, 117(19): 10225–10233.

Tables

Table 1: Summary statistics

	Full sample mean	RD sample: Control mean	Treatment vs. control diff.	Non-homeland mean	Homeland vs. non-homeland diff.
As share of total area:					
Agri land (share of total area)	0.167	0.215	0.061***	0.233	0.008***
			(0.000)		(0.000)
As share of agri land:					
Commercial pivot irrigated	0.158	0.150	0.093***	0.203	-0.201***
			(0.000)		(0.001)
Commercial non-pivot irrigated	0.032	0.028	0.029***	0.042	-0.042***
			(0.000)		(0.000)
Commercial rainfed	0.51	0.530	-0.044***	0.566	-0.557***
			(0.000)		(0.001)
Subsistence	0.089	0.074	-0.019***	0.004	0.681***
			(0.000)		(0.000)
Fallow land	0.212	0.218	-0.059***	0.185	0.119***
			(0.000)		(0.001)
Geo. Controls					
Terrain Ruggedness Index	2.058	1.202	0.039***	1.184	0.333***
			(0.000)		(0.001)
Max. monthly temperature	35.2	36.2	-0.769***	35.9	0.616***
			(0.001)		(0.002)
Mean annual precipitation	565	523	29.3***	528	51.9***
			(0.063)		(0.101)
Distance to nearest canal	5,834	5,250	-191***	5,097	1,035***
			(1.029)		(1.653)
Ditance to nearest river	1,329	1,110	-264***	1,029	-13.4***
			(0.326)		(0.532)
Number of obs.	56,465,817	20,008,184	29,049,656	26,428,908	29,049,656
Ag. Productivity					
Wheat season EVI (log)	-1.190	-1.273	0.226***	-1.173	-0.182***
			(0.000)		(0.001)
Maize season EVI (log)	-0.539	-0.590	0.140***	-0.516	-0.245***
			(0.000)		(0.000)
Number of obs.	9,427,582	4,303,419	6,795,394	6,163,588	6,795,394
Census variables					
% unemployed	0.282	0.292	-0.038***	0.244	0.122***
			(0.009)		(0.008)
% hhs with zero income	0.145	0.151	-0.023***	0.139	0.022***
			(0.004)		(0.004)
Avg hh income (pc)	33,177	33,625	-1,722	40,957	-25,410***
			(1,751)		(1,601)
Avg hh income (pc) - bottom 10%	1,159	1,066	357***	1,439	-914***
			(115)		(108)
Number of obs.	4,746	2,247	3,035	2,106	3,035

Note: This table shows summary statistics for the main outcomes and the control variables for different samples of the data. Column (1) includes all the grid cells within 10km distance of the canals. Columns (2) to (5) include the RD sample of grid cells that are $\leq$ 50m of relative elevation to the nearest canal. Column (2) shows the mean of the control group (grid cells above the canals) and column (3) shows the result of the t-test of a difference between treatment and control means (treatment minus control) with standard errors in the parentheses. Column (4) shows the mean of the non-homeland areas and column (5) shows the result of the t-test of a difference between non-homeland and homeland means (homeland minus non-homeland) with standard errors in the parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2: Balance in geophysical characteristics

	(1) TRI	(2) Max monthly temp. (°C)	(3) Annual precip. (mm)	(4) Distance to canal (m)	(5) Distance to river (m)	(6) Proportion of sand part. (%)	(7) Proportion of silt part. (%)	(8) Proportion of clay part. (%)
Panel I. All grid cells								
Panel A. Non-homelands								
Below canal	-0.022	0.040	0.406	344***	48.3**	0.104	0.010	-0.115
	(0.023)	(0.081)	(0.834)	(94.9)	(23.3)	(0.115)	(0.073)	(0.089)
Control mean	1.18	36.1	519	5,149	1,111	59.4	16.7	23.9
R2	0.438	0.692	0.966	0.213	0.191	0.892	0.896	0.835
Clusters	139	139	139	139	139	139	139	139
N	26,425,820	26,425,820	26,425,820	26,425,820	26,425,820	25,442,685	25,442,685	25,442,685
Panel B. Homelands								
Below canal	0.018	0.538*	3.43	659***	195***	-0.467	0.067	0.401
	(0.067)	(0.311)	(4.30)	(226)	(71.1)	(0.580)	(0.159)	(0.502)
Control mean	1.39	37.0	564	6,196	1,105	56.3	18.0	25.7
R2	0.627	0.744	0.940	0.132	0.260	0.894	0.911	0.783
Clusters	34	34	34	34	34	34	34	34
N	2,619,889	2,619,889	2,619,889	2,619,889	2,619,889	2,416,827	2,416,827	2,416,827
Panel II. Agri. grid cells								
Panel A. Non-homelands								
Below canal	-0.001	0.011	1.93*	325**	59.2*	-0.040	0.181	-0.142
	(0.028)	(0.070)	(1.16)	(145)	(34.6)	(0.158)	(0.115)	(0.128)
Control mean	0.885	35.8	549	5,208	1,154	59.6	15.7	24.7
R2	0.455	0.675	0.968	0.320	0.282	0.899	0.911	0.807
Clusters	137	137	137	137	137	137	137	137
N	6,163,457	6,163,457	6,163,457	6,163,457	6,163,457	6,146,023	6,146,023	6,146,023
Panel B. Homelands								
Below canal	-0.142**	0.294	2.84	778*	347**	-0.489	-0.141	0.629
	(0.066)	(0.210)	(3.49)	(382)	(161)	(0.376)	(0.320)	(0.462)
Control mean	1.04	37.7	539.3	6,521	1,234	57.9	17.0	25.2
R2	0.604	0.849	0.951	0.142	0.300	0.916	0.929	0.827
Clusters	28	28	28	28	28	28	28	28
N	631,761	631,761	631,761	631,761	631,761	620,484	620,484	620,484

Note. This table reports regression discontinuity estimates of the coefficients on the treatment indicator obtained by estimating equation 1 and omitting the control variables. Standard errors are clustered at the canal level. The sample is restricted to grid cells within 10 km of distance and $\leq$ 50 m of relative elevation to the nearest canal. "TRI" is the terrain ruggedness index, a topographic measure that captures variability of elevation of a given area and is derived from Nunn and Puga (2012). "Max monthly temp. (°C)" is calculated as an average over the maximum temperatures of each month in the period of 2014–2018 and is derived from MODIS Terra Land Surface Temperature dataset. "Annual precip. (mm)" is calculated as an average of total annual rainfall over the period of 2014–2018 and is derived from the CHIRPS dataset (Funk et al., 2015). "Distance to canals (m)" is the distance of the grid cell to the nearest canal. "Distance to river (m)" is the distance of the grid cell to the nearest river (the river network comes from the HydroSHEDS dataset). "Proportion of sand/silt/clay particles (%)" is the percentage of the fine earth fraction made up of sand/silt/clay particles (< 0.002 mm). * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 3: Regression discontinuity results for agricultural and land use outcomes

	Intensive margin		Extensive margin				
	Wheat EVI (log) (1)	Maize EVI (log) (2)	Agricultural land (3)	Annual crops (4)	Irrigated annuals (5)	Fallow land (6)	Perennial crops (7)
Panel A. Non-homelands							
Below canal	0.039***	0.023**	0.043***	0.045***	0.023***	-0.003	0.005
	(0.013)	(0.010)	(0.013)	(0.013)	(0.007)	(0.003)	(0.004)
Control mean	-1.255	-0.562	0.212	0.169	0.042	0.044	0.013
R2	0.626	0.413	0.161	0.161	0.157	0.042	0.255
Clusters	137	137	139	139	139	139	139
N	6,163,451	6,163,457	26,425,820	26,425,820	26,425,820	26,425,820	26,425,820
Panel B. Homelands							
Below canal	-0.029	-0.026**	0.011	0.040	0.002	-0.033*	0.002
	(0.028)	(0.010)	(0.036)	(0.030)	(0.002)	(0.018)	(0.006)
Control mean	-1.416	-0.820	0.244	0.161	0.000	0.090	0.029
R2	0.778	0.641	0.121	0.119	0.007	0.067	0.181
Clusters	28	28	34	34	34	34	34
N	631,758	631,761	2,619,889	2,619,889	2,619,889	2,619,889	2,619,889

Note. This table reports regression discontinuity estimates of the coefficients on the treatment indicator obtained by estimating equation 1 using the SANLC 2018 dataset and the Enhanced Vegetation Index (EVI) derived from the Landsat 8 Collection 1 Tier 1 8-Day EVI Composite. Panel A shows the results for non-homelands and Panel B for the former homelands. Standard errors are clustered at the canal level. The sample is restricted to grid cells within 10 km of distance and $\leq$ 50m of relative elevation to the nearest canal. In columns (1) and (2), the sample includes only grid cells classified as agricultural land in SANLC 2018 (categories 38,39,40,41,43,44,45 which comprise annual crops and corresponding fallow land). In columns (3)–(7), all grid cells are included. In columns (1) and (2), the dependent variable is EVI which is a remote sensing measure generated from the Near-IR, Red and Blue bands of satellite imagery, and ranges in value from -1 to 1. I extract the maximum EVI over the growing and harvesting seasons of wheat and maize respectively. In columns (3)–(7), the dependent variable is a dummy equal to 1 if land is classified as belonging to a given category in SANLC 2018. "Agricultural land" is equal to 1 if the grid cell is classified as agricultural (cat. 38,39,40,41,43,44,45 which comprise annual crops and corresponding fallow land), and 0 otherwise. "Annual crops" is equal to 1 if the grid cell is classified growing annual crops (cat. 38,39,40,41 which comprise commercial and subsistence annual crops). "Irrigated annuals" is equal to 1 if the grid cell is classified as commercial annuals pivot-irrigated or commercial annuals non-pivot-irrigated (cat. 38 and 39). "Fallow land" is equal to 1 if the grid cell is classified as fallow land (cat. 43,44,45, not including fallow trees). "Perennial crops" is equal to 1 if the grid cell is classified as growing perennial crops (cat. 32,33,34,36,37 which comprise orchards, sugarcane, and vines) and 0 otherwise. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 4: Regression discontinuity results for land use change

	Expansion between 1990 and 2020 of:			
	Agricultural land (1)	Commercial pivot irr. (2)	Commercial non-pivot (3)	Subsistence (4)
Panel A. Non-homelands				
Below canal	0.010**	0.015**	0.004**	0.000
	(0.005)	(0.006)	(0.002)	(0.000)
Control mean	0.035	0.033	0.020	0.000
R2	0.039	0.122	0.014	0.005
Clusters	139	139	139	139
N	26,425,820	26,425,820	26,425,820	26,425,820
Panel B. Homelands				
Below canal	0.006	-0.000	0.002	0.006
	(0.005)	(0.000)	(0.005)	(0.005)
Control mean	0.045	0.001	0.026	0.031
R2	0.129	0.018	0.229	0.075
Clusters	34	34	34	34
N	2,619,889	2,619,889	2,619,889	2,619,889

Note. This table reports regression discontinuity estimates of the coefficients on the treatment indicator obtained by estimating equation 1 using the SANLC 1990/2020 Change dataset. Panel A shows the results for non-homelands and Panel B for the former homelands. Standard errors are clustered at the canal level. The sample is restricted to grid cells within 10 km of distance and $\leq$ 50m of relative elevation to the nearest canal. The sample includes both agricultural and non-agricultural grid cells. "Agricultural land" "Any annuals" is equal to 1 if the grid cell is classified as having annual crops in the SANLC 2018 data but not in the SANLC 1990 data, and 0 otherwise. "Commercial pivot irr." is equal to 1 if the grid cell is classified as commercial pivot-irrigated in the SANLC 2018 data but not in the SANLC 1990 data, and 0 otherwise. "Commercial non-pivot" is equal to 1 if the grid cell is classified as commercial non-pivot irrigated or commercial rainfed in the SANLC 2018 data but not in the SANLC 1990 data and 0 otherwise. "Subsistence" is equal to 1 if the grid cell is classified as subsistence land in the SANLC 2018 data but not in the SANLC 1990 data and 0 otherwise. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 5: RD effects of canals on local labor markets

	Unemployed (1)	% HHs with zero inc. (2)	HH inc. pc (3)	HH inc. pc (bottom 10%) (4)
Panel A. Non-homelands				
Below canal	-0.049*	-0.027**	5980	206.5
	(0.028)	(0.011)	(10625)	(547.7)
Control mean	0.250	0.138	45501	1548
R2	0.323	0.191	0.414	0.179
Clusters	110	110	110	110
N	2,103	2,106	2,103	2,103
Panel B. Homelands				
Below canal	-0.020	-0.010	1141	398.7
	(0.039)	(0.015)	(3568)	(287)
Control mean	0.377	0.157	16082	553
R2	0.125	0.108	0.306	0.122
Clusters	42	42	42	42
N	927	928	928	928

Note. This table reports regression discontinuity estimates of the coefficients on the treatment indicator obtained by estimating equation 3. Panel A shows the results for non-homelands and Panel B for the former homelands Standard errors are clustered at the canal level. The sample is restricted to census small areas that lie within 10 km of distance and $\leq$ 50m of relative elevation to the nearest canal. "Unemployed" is the small-area-level share of individuals participating in the labor market who are unemployed. "% HHs with zero inc." is the small-area-level share of households who declare having no income. "HH inc. pc" is the small-area-level average annual household income per capita. "HH inc. pc (bottom 10%)" is the average annual household income per capita, where the small-area-level average is calculated only among the bottom decile of the income distribution. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 6: CBA: Breakdown of costs

	Amount (in 2014 Rand)	Discounted amount (in 2014 Rand, total)
Capital costs		
Total	395,419,349	327,221,685
Dam	328,506,564	
Professional fees	39,420,788	
Social & Environmental	3,285,066	
Cost of Infrastructure Replacement	18,416,570	
Cost of Land Acquisition and Relocation	5,790,362	
O&M costs	821,266	15,895,856

Source: Own calculations based on *Department of Water Affairs, South Africa, 2014. Development of a reconciliation strategy for the Luvuvhu and Letaba Water Supply System: Water Supply Schemes, Social and Environmental Aspects Report.* Discount rate is 3%.

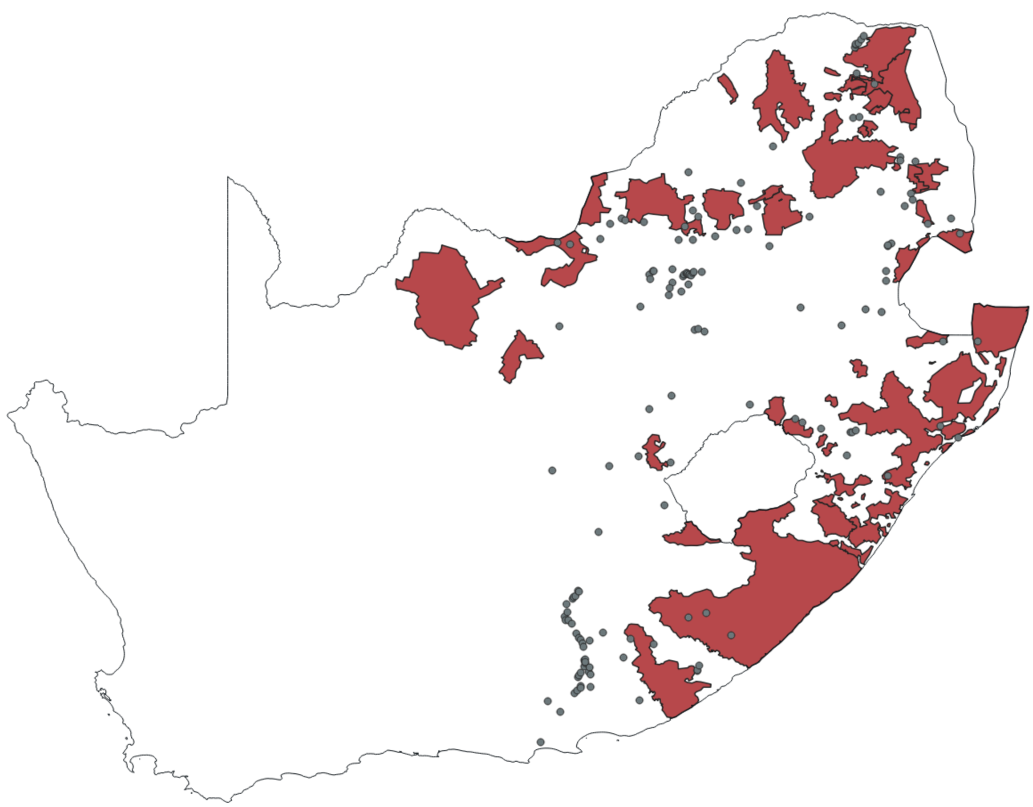
Table 7: CBA: Benefit-cost ratios

Scenario	Discount rate	Discounted benefits (in 2014 Rand)	Discounted costs (in 2014 Rand)	Benefit-cost ratio
Baseline	3%	537,468,348	332,039,758	1.62
Counterfactual 1	4%	420,413,737	321,851,166	1.31
Counterfactual 2	3%	301,065,675	332,039,758	0.91

Source: Own calculations based on *Department of Water Affairs, South Africa, 2014. Development of a reconciliation strategy for the Luvuvhu and Letaba Water Supply System: Water Supply Schemes, Social and Environmental Aspects Report.* Baseline scenario uses RD effects estimated in this paper as benefits. Counterfactual 1 uses 4% discount rate. Counterfactual 2 assumes no extensive margin effect. Counterfactual 3 assumes that subsistence farmers benefit at the intensive margin. Counterfactual 4 assumes the subsistence farmers benefit both at the intensive and the extensive margin.

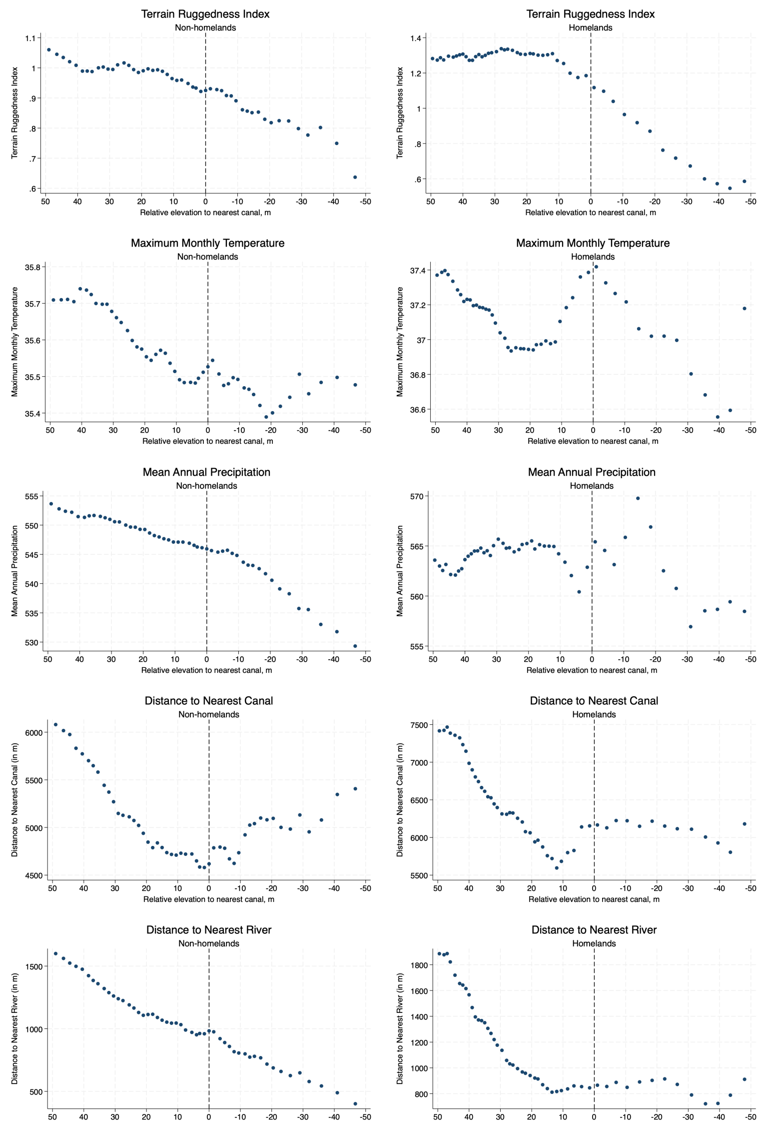
Figures

Figure 1: Spatial distribution of the irrigation canals and the former homelands



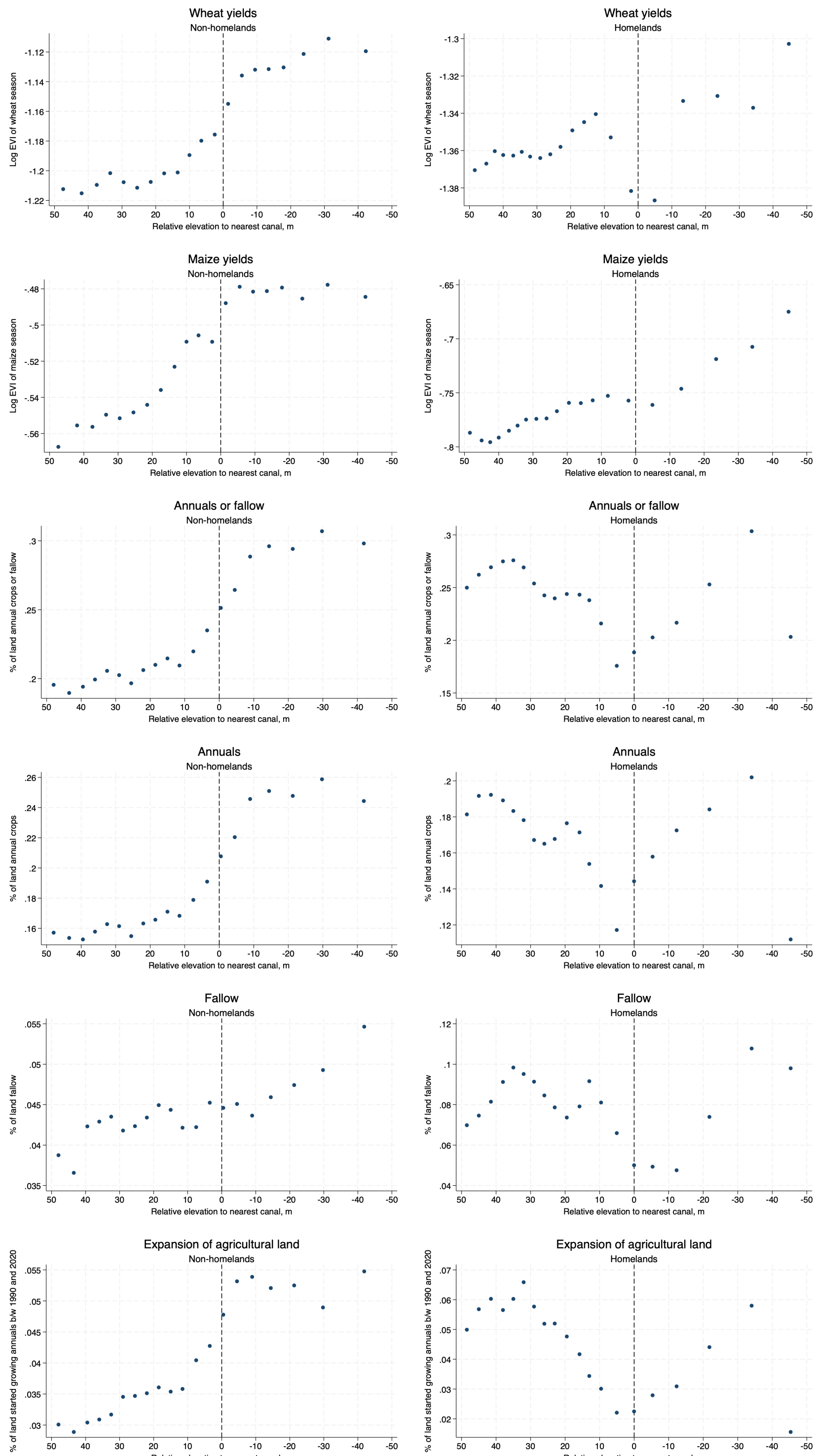
Note. Map shows the location of the former homelands (in red) and the location of the irrigation canals (grey circles).

Figure 2: Continuity through threshold of geophysical characteristics



Note. Each panel plots the average outcome within each of the 50 quantile bins of relative elevation to the nearest canal, separately for non-homelands (left-hand side) and homelands (right-hand side). The data is residualized on canal fixed effects using the `binsreg` package (Cattaneo et al., 2024). The positive values to the left of 0 represent the control units (above a canal) and the negative values to the right of 0 represent the treatment units (below a canal). Fitted linear regression lines of the underlying data are plotted separately for each side of the threshold. The sample is restricted to grid cells within 10 km of distance and $\leq$ 50 m of relative elevation from the nearest canal. Crop mask is applied.

Figure 3: Regression discontinuity binned scatterplots for agricultural and land use outcomes



Note. Each panel plots the average outcome within each of the 20 quantile bins of relative elevation to the nearest canal, separately for non-homelands (left-hand side) and homelands (right-hand side). The data is residualized on canal fixed effects and the weather and terrain covariates using the `binsreg` package (Cattaneo et al., 2024). The positive values to the left of 0 represent the control units (above a canal) and the negative values to the right of 0 represent the treatment units (below a canal). Crop mask is applied only for wheat and maize yields. The corresponding RD estimates are reported in Tables 3 and 4.

Appendix

Table A–1: Robustness of the intensive margin results: extending to multiple years

	Intensive margin (2000-2019)	
	Wheat EVI (log) (1)	Maize EVI (log) (2)
Panel A. Non-homelands		
Below canal	0.022*	0.009
	(0.013)	(0.006)
Control mean	-1.495	-0.778
R2	0.578	0.331
Canal FE	Yes	Yes
Year FE	Yes	Yes
Clusters	137	137
N	63,065,671	63,410,102
Panel B. Homelands		
Below canal	-0.058*	-0.036***
	(0.034)	(0.010)
Control mean	-1.489	-0.894
R2	0.615	0.451
Canal FE	✓	✓
Year FE	✓	✓
Clusters	27	27
N	1,466,412	1,464,739

Note. This table reports regression discontinuity estimates of the coefficients on the treatment indicator obtained by estimating equation 1 using the Enhanced Vegetation Index (EVI) derived from the Landsat 8 Collection 1 Tier 1 8-Day EVI Composite and the UMD GLAD Global Cropland Map (Potapov et al., 2022). Regressions include year fixed effects. The sample includes only grid cells classified as cropland by the UMD GLAD Global Cropland Map (annuals, perennials, forage, biofuel). Enhanced Vegetation Index (EVI) is a remote sensing measure that is generated from the Near-IR, Red and Blue bands of each satellite image, and ranges in value from -1 to 1. It is derived from the Landsat 8 Collection 1 Tier 1 8-Day EVI Composite. I extract the maximum EVI over the growing and harvesting seasons of wheat and maize respectively. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A–2: Comparison of UMD GLAD 2019 and SANLC 2018

	UMD GLAD 2019 cropland by NLC category
Panel A. Non-homelands	
Commercial Annuals Crops Rain-Fed / Dryland / Non-Irrigated	49%
Commercial Annuals Pivot Irrigated	32%
Commercial Annuals Non-Pivot Irrigated	6%
Natural Grassland	4%
Cultivated Commercial Sugarcane Non-Pivot	2%
Panel B. Homelands	
Subsistence / Small-Scale Annual Crops	32%
Cultivated Emerging Farmer Sugarcane Non-Pivot	25%
Open Woodland	12%
Cultivated Commercial Sugarcane Non-Pivot	4%
Residential Formal (Bush)	4%

Note. The table compares the grid cells defined as cropland in the UMD GLAD 2019 global cropland map with the SANLC 2018 land cover classes. Panel A represents cropland within non-homeland areas, while Panel B shows cropland within homeland areas. Only the five most frequent land cover categories within each area are reported. Percentages indicate the proportion of UMD GLAD 2019 cropland grid cells that fall into each SANLC 2018 land cover category within the respective regions. Only five most frequent categories for both homelands and non-homelands are included.

Table A–3: Robustness of the intensive margin results: different vertical bandwidths

	Wheat EVI (log)				Maize EVI (log)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Non-homelands								
Below canal	0.021	0.039***	0.044***	0.052***	0.010	0.023**	0.032***	0.035***
	(0.013)	(0.012)	(0.013)	(0.013)	(0.009)	(0.010)	(0.011)	(0.012)
Control mean	-1.204	-1.255	-1.272	-1.271	-0.525	-0.562	-0.576	-0.579
R2	0.621	0.625	0.625	0.623	0.391	0.410	0.420	0.421
Clusters	137	137	137	138	137	137	137	138
Vertical bandwidth	25	50	75	100	25	50	75	100
N	3,792,405	6,163,451	7,342,030	7,848,264	3,792,407	6,163,457	7,342,038	7,848,272
Panel B. Homelands								
Below canal	-0.037	-0.029	-0.067**	-0.057**	-0.018**	-0.026**	-0.025*	-0.008
	(0.022)	(0.028)	(0.030)	(0.024)	(0.006)	(0.010)	(0.012)	(0.015)
Control mean	-1.397	-1.416	-1.414	-1.399	-0.794	-0.820	-0.817	-0.805
R2	0.765	0.778	0.777	0.770	0.658	0.641	0.637	0.634
Clusters	26	28	31	32	26	28	31	32
Vertical bandwidth	25	50	75	100	25	50	75	100
N	279,188	631,758	831,456	908,112	279,188	631,761	831,459	908,115

Note. See notes for Table 3. The robustness of the intensive margin results is shown for vertical bandwidths of 25 m, 75 m, and 100 m. The preferred specification vertical bandwidth (50 m) is shown for comparison.

Table A–4: Robustness of the extensive margin results: different vertical bandwidths

	Agricultural land				Annual crops			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Non-homelands								
Below canal	0.020*	0.043***	0.053***	0.053***	0.020*	0.045***	0.054***	0.053***
	(0.011)	(0.013)	(0.013)	(0.014)	(0.011)	(0.013)	(0.014)	(0.014)
Control mean	0.227	0.212	0.200	0.189	0.184	0.169	0.158	0.149
R2	0.162	0.161	0.166	0.170	0.164	0.161	0.160	0.163
Clusters	138	139	139	139	138	139	139	139
Vertical bandwidth	25	50	75	100	25	50	75	100
N	15,201,625	26,425,820	33,427,200	37,590,184	15,201,625	26,425,820	33,427,200	37,590,184
Panel B. Homelands								
Below canal	0.001	0.011	-0.008	-0.034	0.025	0.040	0.017	-0.003
	(0.021)	(0.036)	(0.036)	(0.035)	(0.021)	(0.030)	(0.028)	(0.027)
Control mean	0.212	0.244	0.242	0.228	0.139	0.161	0.161	0.152
R2	0.115	0.121	0.135	0.142	0.120	0.119	0.127	0.133
Clusters	33.000	34.000	36.000	37.000	33.000	34.000	36.000	37.000
Vertical bandwidth	25	50	75	100	25	50	75	100
N	1,313,787	2,619,889	3,494,244	3,961,235	1,313,787	2,619,889	3,494,244	3,961,235

Note. See notes for Table 3. The robustness of the extensive margin results is shown for vertical bandwidths of 25 m, 75 m, and 100 m. The preferred specification vertical bandwidth (50 m) is shown for comparison.

Table A–5: Robustness of the intensive margin results: different horizontal bandwidths

	Wheat EVI (log)				Maize EVI (log)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Non-homelands								
Below canal	0.042**	0.036**	0.041***	0.039***	0.002	0.017	0.024**	0.023**
	(0.017)	(0.018)	(0.014)	(0.013)	(0.025)	(0.016)	(0.012)	(0.010)
Control mean	-1.082	-1.158	-1.221	-1.255	-0.506	-0.518	-0.543	-0.562
R2	0.714	0.652	0.632	0.625	0.501	0.444	0.414	0.410
Clusters	130	132	137	137	130	132	137	137
Horizontal bandwidth	2500	5000	7500	10000	2500	5000	7500	10000
N	1,256,639	3,088,758	4,784,964	6,163,451	1,256,639	3,088,761	4,784,969	6,163,457
Panel B. Homelands								
Below canal	-0.047	-0.021	-0.013	-0.029	-0.064	-0.013	-0.014	-0.026**
	(0.044)	(0.033)	(0.036)	(0.028)	(0.038)	(0.027)	(0.013)	(0.010)
Control mean	-1.516	-1.510	-1.443	-1.416	-0.889	-0.885	-0.830	-0.820
R2	0.826	0.775	0.750	0.778	0.716	0.670	0.647	0.641
Clusters	17	20	27	28	17	20	27	28
Horizontal bandwidth	2500	5000	7500	10000	2500	5000	7500	10000
N	40,292	181,745	395,794	631,758	40,292	181,748	395,797	631,761

Note. See notes for Table 3. The robustness of the intensive margin results is shown for horizontal bandwidths (the distance to the nearest canal) of 2500 m, 5000 m, and 7500 m. The preferred specification horizontal bandwidth (10000 m) is shown for comparison.

Table A–6: Robustness of the extensive margin results: different horizontal bandwidths

	Agricultural land				Annual crops			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Non-homelands								
Below canal	0.023	0.037**	0.048***	0.043***	0.023	0.037**	0.051***	0.045***
	(0.020)	(0.015)	(0.014)	(0.013)	(0.020)	(0.015)	(0.014)	(0.013)
Control mean	0.221	0.203	0.209	0.212	0.183	0.167	0.168	0.169
R2	0.157	0.156	0.161	0.161	0.161	0.159	0.161	0.161
Clusters	134	134	138	139	134	134	138	139
Horizontal bandwidth	2500	5000	7500	10000	2500	5000	7500	10000
N	5,019,861	13,351,772	20,579,592	26,425,820	5,019,861	13,351,772	20,579,592	26,425,820
Panel B. Homelands								
Below canal	-0.017	0.016	0.019	0.011	-0.024	0.030	0.045	0.040
	(0.048)	(0.039)	(0.035)	(0.036)	(0.040)	(0.035)	(0.029)	(0.030)
Control mean	0.164	0.213	0.225	0.244	0.132	0.156	0.151	0.161
R2	0.198	0.145	0.119	0.121	0.195	0.139	0.121	0.119
Clusters	21	26	33	34	21	26	33	34
Horizontal bandwidth	2500	5000	7500	10000	2500	5000	7500	10000
N	238,859	848,240	1,717,153	2,619,889	238,859	848,240	1,717,153	2,619,889

Note. See notes for Table 3. The robustness of the extensive margin results is shown for horizontal bandwidths (the distance to the nearest canal) of 2500 m, 5000 m, and 7500 m. The preferred specification horizontal bandwidth (10000 m) is shown for comparison.

Table A–7: Robustness of the intensive margin results: donut hole

	Wheat EVI (log)				Maize EVI (log)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Non-homelands								
Below canal	0.047***	0.055***	0.061***	0.066***	0.026**	0.027*	0.029*	0.030
	(0.016)	(0.018)	(0.020)	(0.022)	(0.012)	(0.014)	(0.016)	(0.018)
Control mean	-1.262	-1.267	-1.272	-1.278	-0.568	-0.571	-0.576	-0.581
R2	0.627	0.630	0.631	0.633	0.414	0.417	0.421	0.425
Clusters	137	137	137	137	137	137	137	137
Donut size	2	4	6	8	2	4	6	8
N	5,754,599	5,434,303	5,108,714	4,783,933	5,754,605	5,434,307	5,108,718	4,783,937
Panel B. Homelands								
Below canal	-0.028	-0.026	-0.020	-0.013	-0.030**	-0.032**	-0.032*	-0.034*
	(0.032)	(0.033)	(0.034)	(0.034)	(0.012)	(0.013)	(0.016)	(0.019)
Control mean	-1.417	-1.417	-1.418	-1.419	-0.822	-0.824	-0.825	-0.827
R2	0.779	0.779	0.780	0.781	0.642	0.643	0.643	0.643
Clusters	28	28	28	28	28	28	28	28
Donut size	2	4	6	8	2	4	6	8
N	608,750	590,832	571,526	551,885	608,753	590,835	571,529	551,888

Note. See notes for Table 3. The robustness of the intensive margin results is shown for donut holes (in the relative elevation to the nearest canal) of 2 m, 4 m, 6 m, and 8 m.

Table A–8: Robustness of the extensive margin results: donut hole

	Agricultural land				Annual crops			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Non-homelands								
Below canal	0.054***	0.065***	0.073***	0.075***	0.057***	0.067***	0.076***	0.079***
	(0.015)	(0.017)	(0.020)	(0.022)	(0.015)	(0.017)	(0.019)	(0.021)
Control mean	0.209	0.208	0.206	0.205	0.166	0.165	0.163	0.162
R2	0.162	0.163	0.165	0.167	0.161	0.163	0.164	0.166
Clusters	139	139	139	139	139	139	139	139
Donut size	2	4	6	8	2	4	6	8
N	24,842,043	23,590,233	22,331,548	21,091,141	24,842,043	23,590,233	22,331,548	21,091,141
Panel B. Homelands								
Below canal	0.016	0.017	0.019	0.027	0.050	0.056	0.060	0.069
	(0.043)	(0.048)	(0.055)	(0.062)	(0.035)	(0.039)	(0.044)	(0.051)
Control mean	0.247	0.249	0.252	0.253	0.163	0.164	0.166	0.167
R2	0.119	0.119	0.119	0.121	0.119	0.120	0.121	0.122
Clusters	34	34	34	34	34	34	34	34
Donut size	2	4	6	8	2	4	6	8
N	2,494,570	2,390,651	2,287,231	2,191,794	2,494,570	2,390,651	2,287,231	2,191,794

Note. See notes for Table 3. The robustness of the extensive margin results is shown for donut holes (in the relative elevation to the nearest canal) of 2 m, 4 m, 6 m, and 8 m.

Table A–9: Robustness of the intensive margin results: control variables and fixed effects

	Wheat EVI (log)				Maize EVI (log)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Non-homelands								
Below canal	0.036**	0.037***	0.037***	0.036***	0.025**	0.022**	0.022**	0.021*
	(0.014)	(0.012)	(0.012)	(0.013)	(0.010)	(0.010)	(0.010)	(0.011)
Control mean	-1.255	-1.255	-1.255	-1.255	-0.562	-0.562	-0.562	-0.562
R2	0.604	0.626	0.632	0.611	0.389	0.413	0.429	0.395
Clusters	137	137	137	137	137	137	137	137
Weather & terrain controls		✓	✓	✓		✓	✓	✓
Soil quality controls		✓	✓			✓	✓	
Control for carbon & nitrogen			✓				✓	
Canal FE		✓	✓			✓	✓	
District FE				✓				✓
N	6,163,582	6,146,017	6,139,658	6,163,451	6,163,588	6,146,023	6,139,664	6,163,457
Panel B. Homelands								
Below canal	-0.023	-0.031	-0.035	-0.029	-0.020*	-0.027**	-0.031***	-0.029***
	(0.034)	(0.026)	(0.026)	(0.028)	(0.012)	(0.011)	(0.011)	(0.009)
Control mean	-1.416	-1.416	-1.416	-1.416	-0.820	-0.820	-0.820	-0.820
R2	0.768	0.784	0.786	0.776	0.636	0.648	0.650	0.636
Clusters	28	28	28	28	28	28	28	28
Weather & terrain controls		✓	✓	✓		✓	✓	✓
Soil quality controls		✓	✓			✓	✓	
Control for carbon & nitrogen			✓				✓	
Canal FE		✓	✓			✓	✓	
Munic. FE				✓				✓
N	631,803	620,484	618,006	631,758	631,806	620,484	618,006	631,761

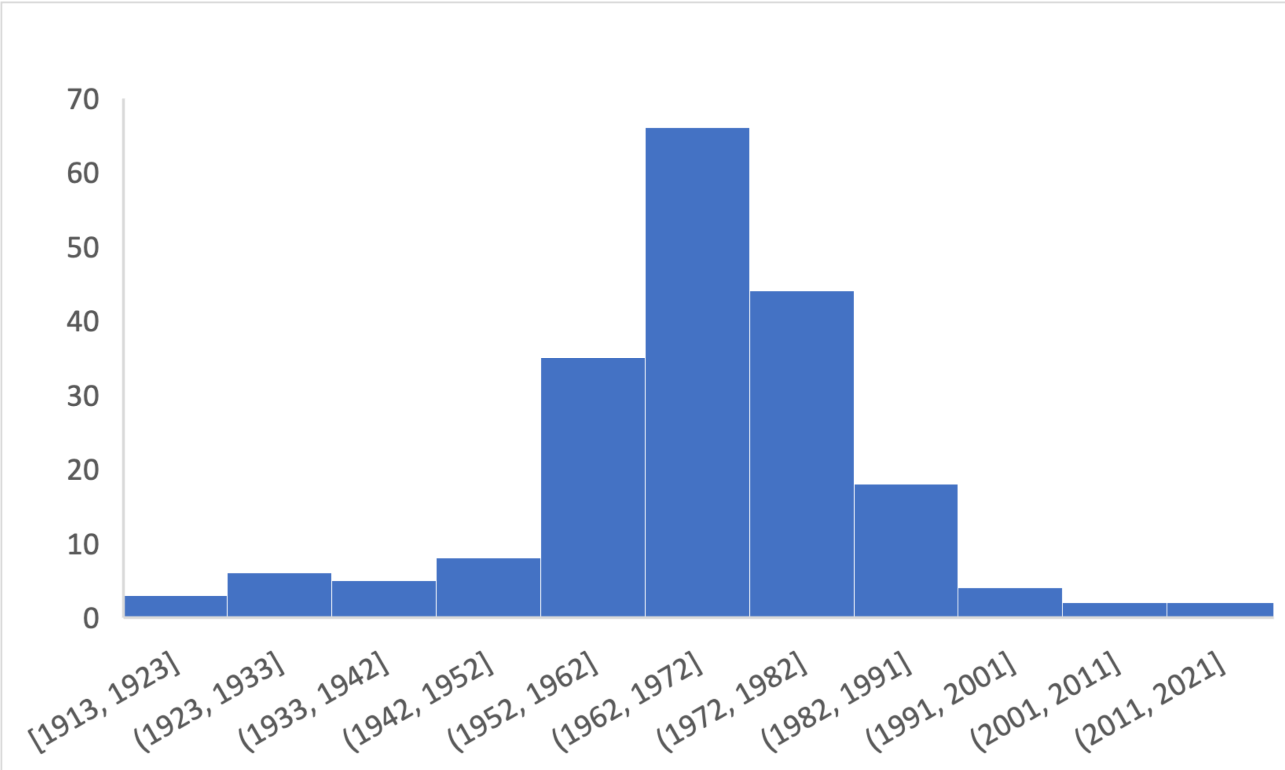
Note. See notes for Table 3. The robustness of the intensive margin results is shown for different sets of control variables and different level of fixed effects. Columns (1) and (5) show results of RD regressions without including the covariates from the preferred specification (weather & terrain controls). Columns (2) and (6) show results of RD regressions with the covariates from the preferred specification and soil quality controls. Columns (3) and (7) show results of RD regressions with the covariates from the preferred specification, soil quality controls, and carbon and nitrogen controls. Columns (4) and (8) show the results of RD regressions with the covariates from the preferred specification and local municipality FE replacing the canal FE.

Table A–10: Robustness of the extensive margin results: control variables and fixed effects

	Agricultural land				Annual crops			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Non-homelands								
Below canal	0.046***	0.034***	0.029**	0.045***	0.048***	0.036***	0.032***	0.048***
	(0.013)	(0.012)	(0.012)	(0.014)	(0.013)	(0.012)	(0.011)	(0.013)
Control mean	0.212	0.212	0.212	0.212	0.169	0.169	0.169	0.169
R2	0.147	0.175	0.194	0.133	0.144	0.175	0.197	0.132
Clusters	139.000	139.000	139.000	139.000	139.000	139.000	139.000	139.000
Weather & terrain controls		✓	✓	✓		✓	✓	✓
Soil quality controls		✓	✓			✓	✓	
Control for carbon & nitrogen			✓				✓	
Canal FE	✓	✓	✓		✓	✓	✓	
Munic. FE				✓				✓
N	26,428,908	25,442,685	25,346,761	26,420,063	26,428,908	25,442,685	25,346,761	26,420,063
Panel B. Homelands								
Below canal	0.036	-0.028	-0.027	0.008	0.053	0.009	0.010	0.038
	(0.039)	(0.031)	(0.031)	(0.035)	(0.033)	(0.027)	(0.027)	(0.029)
Control mean	0.244	0.244	0.244	0.244	0.161	0.161	0.161	0.161
R2	0.096	0.129	0.130	0.112	0.103	0.129	0.129	0.112
Clusters	34.000	34.000	34.000	34.000	34.000	34.000	34.000	34.000
Weather & terrain controls		✓	✓	✓		✓	✓	✓
Soil quality controls		✓	✓			✓	✓	
Control for carbon & nitrogen			✓				✓	
Canal FE	✓	✓	✓		✓	✓	✓	
Munic. FE				✓				✓
N	2,620,748	2,416,827	2,398,606	2,619,578	2,620,748	2,416,827	2,398,606	2,619,578

Note. See notes for Table 3. The robustness of the extensive margin results is shown for different sets of control variables and different level of fixed effects. Columns (1) and (5) show results of RD regressions without including the covariates from the preferred specification (weather & terrain controls). Columns (2) and (6) show results of RD regressions with the covariates from the preferred specification and soil quality controls. Columns (3) and (7) show results of RD regressions with the covariates from the preferred specification, soil quality controls, and carbon and nitrogen controls. Columns (4) and (8) show the results of RD regressions with the covariates from the preferred specification and local municipality FE replacing the canal FE.

Figure A–1: Histogram of years of construction of large irrigation dams



The data comes from AQUASTAT, the geo-referenced database on dams in Africa. Only dams with the listed purpose of irrigation are included. There are 200 irrigation dams and 550 dams in total.

Figure A–2: Schematic representation of elevation-based RDD

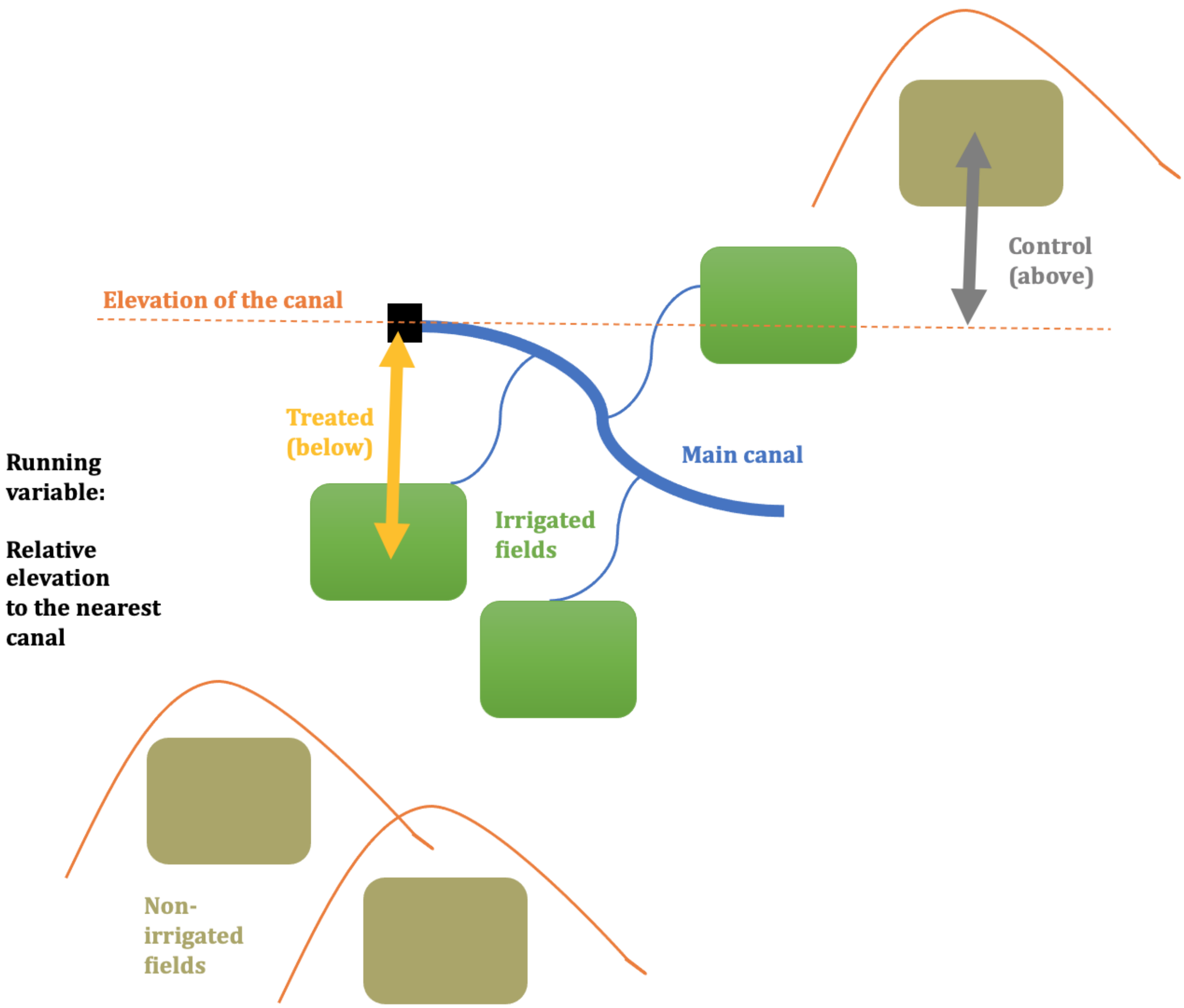
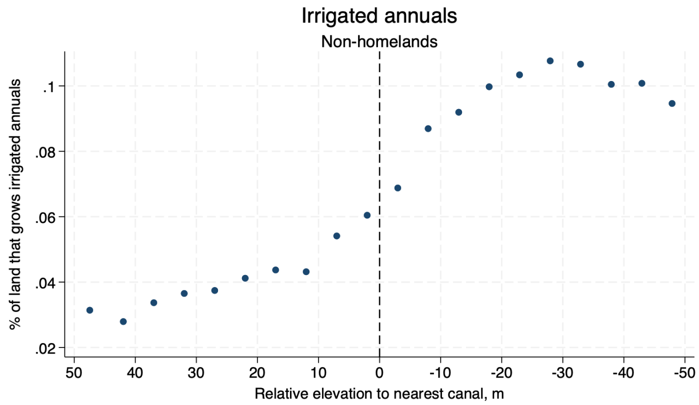
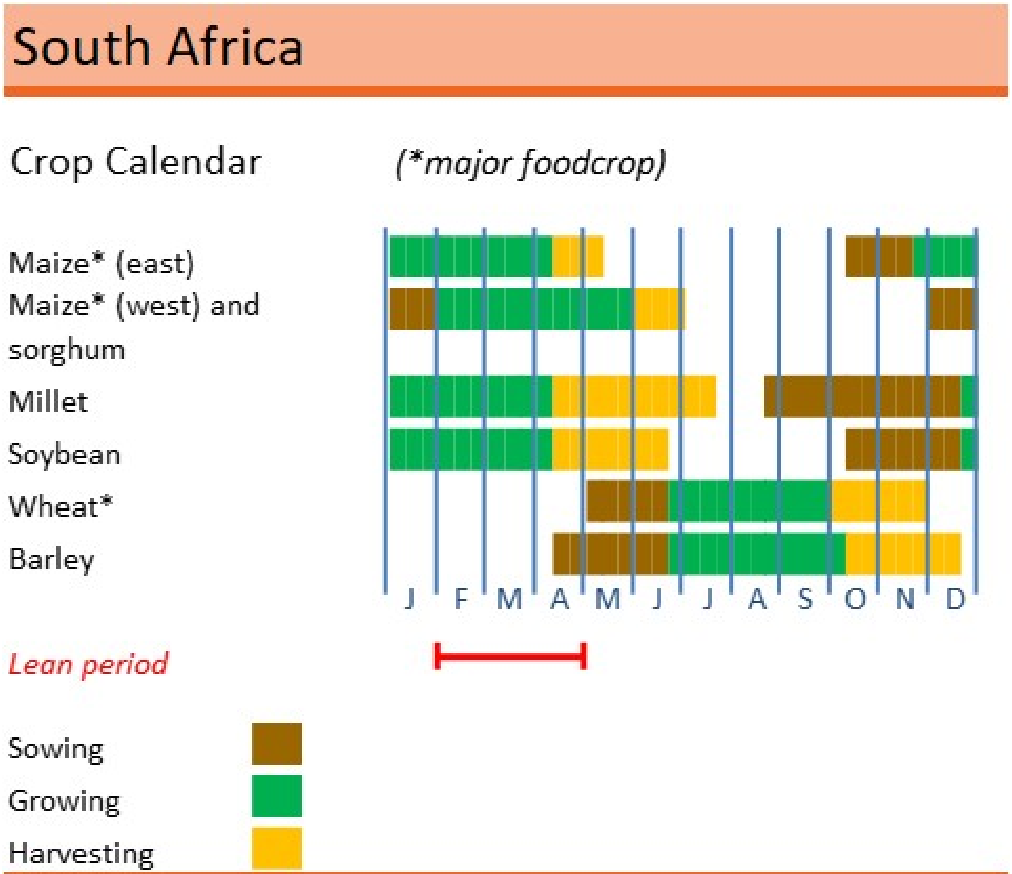


Figure A–3: Probability of land growing irrigated annuals



Note. The graph plots the average outcome within each of the 20 bins (equally spaced) of relative elevation to the nearest canal, only for non-homelands. The data is residualized on canal fixed effects and weather and terrain covariates using the `binsreg` package (Cattaneo et al., 2024). The positive values to the left of 0 represent the control units (above a canal) and the negative values to the right of 0 represent the treatment units (below a canal). Crop mask is applied. The corresponding RD estimate is reported in Table 3, column (5).

Figure A–4: FAO Crop Calendar

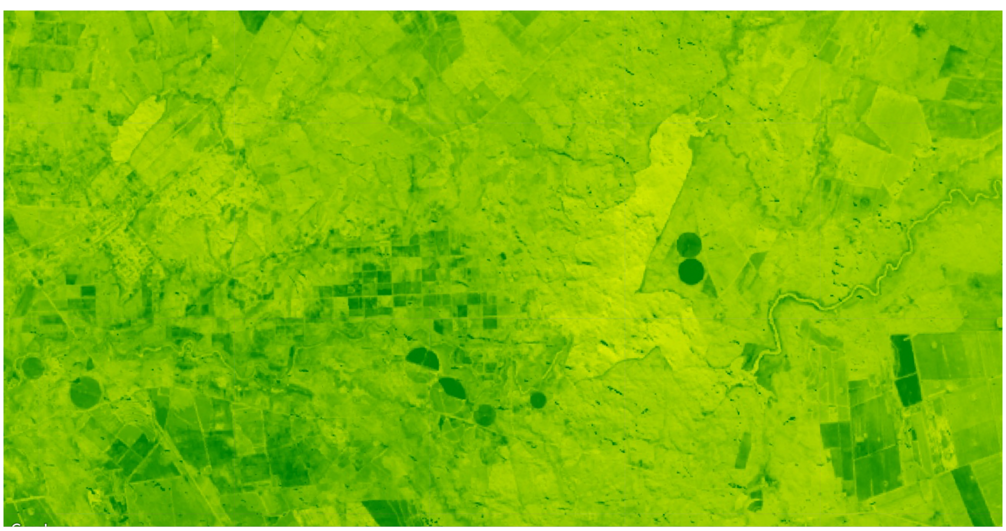


Source: `https://www.fao.org/giews/countrybrief/country.jsp?code=ZAF`.

Figure A–5: Visualization of the main outcome variables

(a) Raw satellite image

(b) SANLC 2018



(c) Landsat 8 EVI Composite