

A New Expanded Dataset to Study Refugee Camps in Sub-Saharan Africa 1999-2024[†]

Sebastian Anti[‡] Colette Salemi[§] Jonathan Rigberg[¶] Sofia Leopold[¶]
Karishma Silva* Johannes Hoogeveen*

May 2026

ABSTRACT: One in five refugees live in camps or camp-like settings, and three-quarters of encamped refugees are in sub-Saharan Africa. No reliable public data has systematically tracked camp locations, operations, or populations over time. To address this, we introduce the African Refugee Camps Dataset (ARCD), a geospatial panel dataset. We first describe how the dataset was constructed, the sources used to build it, and the structure and content of the resulting data. We then use the data to characterize the refugee camps in the region over the 26 years the ARCD covers. We finish by presenting two applications combining the ARCD with complementary data to illustrate its potential for research. First, we assess spatial features of camp locations compared to stratified random sites. Second, we estimate the effect of camp openings on forest and vegetation cover using a differences-in-differences approach.

JEL CODES: F22, O55, C81, J15, Q56, O18

KEYWORDS: refugee camps, geospatial data, sub-Saharan Africa, spatial analysis, deforestation

[†]This work was supported by the Forced Displacement Trust Fund (MDTF). The MDTF is a World Bank administered global trust fund, intended to promote development responses that support sustainable solutions to forced displacement. The authors gratefully acknowledge their support of this work. We additionally thank the UNHCR Global Data Service team, particularly Pierre Vernier and Volker Schimmel, for providing feedback and for validating our geospatial data.

[‡]Bryn Mawr College

[§]University of Victoria

[¶]Independent Analyst

*World Bank

1 Introduction

In high-income countries, public narratives of forced displacement often emphasize migration to Europe or the United States. In reality, 73% of the world’s refugees¹ obtain asylum in low- or middle-income countries close to the country they were forced to leave. In sub-Saharan Africa (SSA), which currently hosts 27% of the world’s refugees, the refugee population has grown tremendously in the past 15 years, from 2.1 million in 2009 to 8.5 million in 2024. Another common misperception is that most refugees reside outside of camps, settling predominately in urban areas where they live among host communities (The Economist, 2018; Vos & Dempster, 2021). This is the case for many regions (Europe and Latin America, for example), but in other parts of the world, encampment is the dominant hosting modality. Accommodation in camps is particularly prevalent in SSA, where the majority of refugees (74%) are living in refugee camps or other camp-like situations.² These refugee camps are best understood as designated, intentionally created settlements designed and managed by the hosting government and multilateral entities for the purpose of accommodating refugees and offering a range of essential services (UNHCR, 2007). They are often characterized by remoteness, semi-permanence, and aid dependence (Chambers, 1983; Harrell-Bond, 1986; Lawrence, 2017; Smith, 2004).

To enhance our understanding of forced displacement and refugee encampment for African countries, we present the African Refugee Camps Dataset (ARCD), a spatially explicit and publicly available data product that identifies 466 refugee camps in existence in SSA between 1999 and 2024 and identifies their geographic coordinates, years of operation, and numerous characterizing variables. We focus on SSA because the region accounts for the majority of camps and encamped refugees worldwide during this period.

¹We define a “refugee” as someone who, due to a credible fear of persecution or violence in their home country, has crossed an international border and received protected status.

²Other camp-like situations include transit centers, which are temporary sites that serve to document, service, and transfer refugees to camps (UNHCR, 2007), or collective centers, which are best understood as temporary accommodations operating out of existing structures such as hotels or schools (UNHCR & IOM, 2010). In some cases, refugees may create spontaneous or informal settlements. We can break down the 75% population statistic for SSA in 2023 as follows: 65.1% in formal, or planned, camps; 8.7% in informal settlements; 0.1% in collective centers, and 0.2% in transit centers. These figures are based on UNHCR data shared with the study team.

There have been major developments in the quantitative literature on African refugee camps since 2010 (Maystadt et al., 2019; Verme & Schuettler, 2021). A large share of this research focuses on the impact of refugees and refugee camps on host communities and economies (Alix-Garcia & Saah, 2010; Alix-Garcia et al., 2018; Baez, 2011; Betts et al., 2017; d’Errico et al., 2022; Kadigo & Maystadt, 2023; Kreibaum, 2016; Loschmann et al., 2019; Maystadt & Duranton, 2019; Maystadt & Verwimp, 2014; Musasizi et al., 2024; Taylor et al., 2016; Zhou et al., 2023). There has been limited but growing attention to encampment and refugee welfare (Alloush et al., 2017; Das et al., 2025; MacPherson & Sterck, 2021; Owen et al., 2023) and a limited consideration of the impacts of camps on landscapes (Maystadt et al., 2020; Salemi, 2021). The discourse has also recently seen considerable developments on refugee inclusion and refugee-host social cohesion (Baseler et al., 2025; Betts et al., 2023; Bousquet et al., 2025; Fajth et al., 2019; Ghosn et al., 2019).

These contributions have greatly improved our understanding of markets, livelihoods, welfare, vulnerability, social interactions, and policies in African encampment settings. But data limitations continue to constrain the discourse. Consequently, evidence tends to be limited to national or sub-national case studies focused on East African countries, particularly Uganda and Tanzania.³ To address external validity, some studies have pooled data at the sub-continental level (Bertinelli et al., 2025; Coniglio et al., 2023; Maystadt et al., 2020; Salemi, 2021; Zhou & Shaver, 2021). In these studies, the data often originate with the UNHCR and are available only in a proprietary format (Bertinelli et al., 2025; Coniglio et al., 2023; Maystadt et al., 2020).

The need for improved data on the topic is a high priority as researchers advance our understanding of camps and their associated populations of refugee residents and proximate hosts, and as policymakers struggle to navigate the complex set of decisions entailed in refugee hosting policy. Since the 1980s, UNHCR guidelines have stressed that encampment should only be undertaken if there are no other choices available (Harrell-Bond, 1986). In recent years, the UNHCR has

³Uganda is a common focus for researchers (Betts et al., 2017; d’Errico et al., 2022; Kadigo & Maystadt, 2023; Musasizi et al., 2024; Taylor et al., 2016), but encampment policies in Uganda are unique, as encampments are large in area and refugee households are given plots to cultivate. There are few other camps in SSA that apply this model. Studies on Tanzania tend to focus on camps in Kagera during the Great Lakes Crisis in the mid-1990s (Alix-Garcia & Saah, 2010; Baez, 2011; Maystadt & Duranton, 2019; Maystadt & Verwimp, 2014), and it remains unclear if this case is externally valid or relevant to modern encampment and hosting policy.

discouraged encampment through its 2009 *Policy on Refugee Protection and Solutions in Urban Areas* (UNHCR, 2009) and 2014 *Policy on Alternatives to Camps* (UNHCR, 2014; Verdirame & Pobjoy, 2013). Yet the number of refugee camps and size of the encamped population in Africa has increased considerably over time. Figure 1 shows that for at least two decades, SSA has stood out as the major world region in terms of the number of encamped people, and the number of encamped people has grown considerably on the subcontinent. As of 2023, about 7.8 million refugees were living in camp or camp-like situations around the world, and 5.8 million of them (75%) were in SSA.

The issues underlying the UNHCR data that researchers have been using to study refugee camps and their effects are related to the UNHCR's mission and organizational priorities. The multilateral agency is highly decentralized and primarily concerned with managing complex emergency and humanitarian response operations. Its core mandate focuses on legal protection, and it is also responsible for a considerable amount of logistical planning and service provision. Moreover, the UNHCR is often operating in fluid, insecure, and remote environments. Its data production efforts are focused on operational needs, frequently in an emergency setting, and not academic research.⁴ As a result, UNHCR data collection and dissemination activities, while extensive, are inconsistent across country offices and over time, meaning that without further validation, researchers may inadvertently apply data with considerable inaccuracies.

UNHCR sub-continental data used in recent studies provide examples of the limitations of the UNHCR data. Two major problems persist across these sources. First, the camp data that UNHCR shares with researchers often contain numerous false positives. In some cases, towns or other administrative divisions are classified as camps. In some scenarios, GIS cross-checks show that an alleged camp's geopoint is in an uninhabited area with no access road and no visible shelter footprint. False positives may also be due to the conflation of different types of sites holding refugees, such as mistaking a detention center with a camp. These errors may be endogenous to the institutional capacity of a given host country, the political complexity of the host's displacement situation, and the number of camps accommodated.

⁴This has changed in recent years. Since its founding in 2020, the Joint Data Center for Forced Displacement, a collaboration between UNHCR and the World Bank, has greatly improved researcher access to data and information.

Besides doubting the accuracy of the refugee camps known to have existed in that period, close inspections also reveal that UNHCR data often contain inaccurate information for when the refugee camp was established and when it closed. For example, the UNHCR's Persons of Concern maps (UNHCR, 2025), a publicly available source on refugee locations, only record when a camp was entered into the tracking system and when its status changed (for example, from active" to inactive" or to "refugee site"). In many cases, these data entry dates are entered into the system several years after the documented establishment of the camp. These errors in camp operational data across SSA reflect inconsistencies in data quality across countries. This inconsistency likely explains researchers' tendency to concentrate on a limited set of countries or encampment areas. We construct the ARCD to fill the gap and provide reliable, harmonized data on refugee camps from 1999 to 2024 in SSA.

In this paper, we review our approach to constructing the ARCD data. Using a diverse set of sources, we identify the geo-locations, years of operation, and refugee populations for refugee camps for the years 1999 to 2024. To address inconsistencies and missing information in these datasets, we cross-validate entries using sources from news reports, peer-reviewed papers, and reports from international and domestic non-governmental organizations (NGO). Our data collection focuses on formal camps, which are sometimes referred to as "settlements" in specific contexts. We exclude entries classified as informal encampments, transit centers, and collection centers.

We augment the ARCD by classifying operational characteristics, identifying administrative and spatial attributes of camp locations, estimating proximity measures (distances to borders, cities, capitals, ports, and protected areas), and extracting zonal statistics describing the bioecological, climatic, and socioeconomic conditions of camp areas.

We then use the ARCD to examine patterns of refugee encampment in SSA from 1999 to 2024, including the drivers of camp formation, closure, and longevity. We additionally demonstrate the dataset's potential for empirical analysis. First, we compare camp locations against two control samples: a border comparison" of non-encampment areas within 30-50 km of a camp, and a country comparison" drawn from across host countries. Second, we apply a robust difference-in-differences (DID) estimator (de Chaisemartin & D'Haultfœuille, 2024) to estimate the impact of camp openings

on forest and vegetation cover within camp interiors.

Our descriptive statistics show that refugee camps are frequently close to international borders: 19% of ARCD camps (and 34.6% in Central Africa) are within 20 km of an international border. Camps are closer to capitals than national averages suggest, though 43% are still 100 km or more from the nearest capital. Camp areas also tend to have lower terrain ruggedness and less forest cover than comparison areas two years prior to opening, and they are largely sited in grassland bioclimatic regions. In our causal analysis, camp openings reduce forest cover by 2.5 percentage points. This effect is concentrated in the first two years of operation and driven by camps with populations greater than 1,000, which is consistent with logistically motivated land clearing to make way for shelters and infrastructure.

The remainder of this paper proceeds as follows. [Section 2](#) orients readers to the ARCD through specifying the dataset's location and outlining its structure and codebook. [Section 3](#) describes the underlying sources of the ARCD and our process to validate and harmonize that information. [Section 4](#) uses the ARCD to examine the state of refugee camps in SSA and major trends in refugee camps in the region over the past 26 years. [Section 5](#) presents the data, methods, and results of both our descriptive and causal analyses, both of which use the ARCD in combination with complementary data. [Section 6](#) concludes.

2 Dataset Location, Structure, Variables, and Codebook

Location: The ARCD is publicly available on the Harvard Dataverse (permanent link: <https://doi.org/10.7910/DVN/OE1RE8>). This section describes the data repository and its organizational structure, which we reference when discussing our data construction process in [Section 3](#).

Main Files: Our primary data product is `ARCD_long.dta`, a panel dataset that contains our finalized list of camps operational between 1999 and 2024, as well as characterizing variables. The repository includes the data in wide format as an Excel file (`Camps.xlsx`). Definitions of all variables in the `ARCD_long.dta` file along with their sources can be found in the data codebook `codebook.xlsx`

and in the Appendix to this manuscript ([Table A2](#)).

Supplemental Files: We organize the assembly and validation of our data into a set of resources that users can review. First, the file `Non-Camps.xlsx` contains a list of all refugee locations that were part of our initial roster of potential refugee camps that, based on supplementary sources, we determined were not refugee camps. The file includes a source to justify our decision. This file contains geolocations, opening and closing dates, and refugee population information so that the locations can still be used for tailored analysis. The file `Metadata.xlsx` stores the sources used to validate that the location is a camp, determine opening and closing years, identify the reason the camp closed, and impute population counts.⁵ In a third file called `Conflict_Notes.xlsx`, we compile the sources used to identify precipitating conflict events.

Main Variable Descriptions and Structure: The `ARCD_long.dta` includes the point geocoordinates (longitude and latitude) of each camp, the years each camp was operational, the annual population of refugees in each camp, the events that precipitated the opening of each camp and reasons for camp closure, and additional characteristics of the camp location. Below, we provide a brief orientation to the structure and organization of this file along with definitions of some of the main variables necessary for understanding how to use the data.

We structure the `ARCD_long.dta` file as a panel dataset where each record represents the camp location in a particular year. To illustrate the panel structure, [Table 1](#) provides an excerpt of the data limited to a subset of variables for Kenya's Kakuma Camp. The panel is balanced: each camp has an entry for every year from 1999 to 2024, even if camp is not operational in the reference year. We include time-invariant point geocoordinates for each camp (variables `latitude` and `longitude`). We also include a variable flagging whether we were able to validate the point through visual inspection (`Coords_validated`). We include the opening year (`opening_year`) and closing year (`closing_year`) for each camp. In the case of Kakuma, the information on the closing year is missing because Kakuma remains operational. These opening and closing variables

⁵The `Metadata.xlsx` and `Non-Camps.xlsx` files do not have separate codebooks, but a `ReadMe.txt` document at Harvard Dataverse site provides essential information for how to understand these files.

are time-invariant and will take the same value over all records for any particular camp. Using this information, researchers can easily remove records for camps that were not operational in particular years. We include a flag for whether our process determined that there is some doubt whether the location was in fact a refugee camp or some other type of refugee site (*Potentially_Non_Camp*). The *Camp_ID* variable provides a unique identifier for each record in the dataset.

Camps are sometimes organized in a nested structure where a larger camp complex is composed of smaller sub-camps. Sub-camps may also be organized into smaller sub-sub camps. The ARCD contains records for the larger complexes as well as sub-camps and sub-sub-camps. The *Camp_ID* variable identifies the type of record (main vs. sub-camp) and the main camps that correspond to each sub-camp. The first seven digits of the *Camp_ID* variable uniquely identifies the larger camp complex. For example, Kakuma is identified as 2190020. The first three digits (219) correspond to the country location and 0020 distinguish Kakuma from other main camp complexes in Kenya. Digits 8-9 are sub-camp identifiers, which take the value of 00 if the record is not a subcamp (it represents the larger complex, or the camp is too small for such subdivisions), and an integer 01-99 if the record is for a sub-camp. The Kakuma entry represents a main complex, so the concatenated camp and subcamp identifiers results in ID 219002000. By contrast, the first eight digits of the identifier for Kakuma I (a sub-camp of Kakuma) is 219002001 The last digit indicates whether the record is for a sub-sub-camp and will take on the value of 0 if the record is not a sub-sub-camp, and an integer between 1 and 9 if it is for a sub-sub-camp. Adding the digit indicating that Kakuma is not a sub-sub camp, we arrive at the full unique identifier of 2190020000.

The dataset also includes an indicator for whether a given camp opened and closed more than once within the panel (*Reopening_Indicator*). We observe this for 17 camps overall. We include each operational episode as a separate camp, even though the locations are identical. Users who want to combine these entries into one camp can do so using the indicator variable.

The primary dataframe also includes variables documenting camp populations (and the source for these figures), the conflict that precipitated each camp opening, socioeconomic characteristics of the host area, and a range of bioecological and climatic features of the camp location. We

provide a full variable listing with descriptions and underlying data sources in [Table A2](#) and in the accompanying `codebook.xlsx` file in the repository. In the section that follows, we further describe the process of vetting each location and constructing the additional variables.

3 Dataset Construction

We construct the dataset by starting with locational listings derived from the UNHCR Persons of Concern (POC) Map, a publicly accessible database of known past and present refugee locations (UNHCR, 2025), as well as UNHCR’s Statistical Yearbooks, introduced below. We go through each entry, searching publicly available sources to verify its classification as a camp, confirm its coordinates, and establish its period of operation and population over time. We then use the resulting camp roster to construct a harmonized panel dataset combining geospatial, environmental, demographic, and operational information, including zonal statistics, annual population counts, precipitating events, and camp opening and closure dates.

We draw on numerous publicly available sources to validate the locations and define their variables. Because source reliability varies substantially across records, we classify sources using a structured hierarchy based on institutional credibility, consistency across observations, and proximity to the original administrative record. We prioritize sources produced directly by the UNHCR and other UN agencies, which are most closely linked to official camp administration and registration processes. Our ranking of credible sources is as follows: (1) UNHCR Population of Concern data; (2) UNHCR Operational Data Portal; (3) UN-WASH; (4) UNHCR Statistical Yearbooks; (5) other UN sources; (6) government sources; (7) NGO sources; (8) media coverage; (9) academic literature; and (10) other sources. When sources conflict, we prioritize the highest-ranked source unless multiple lower-ranked sources provide stronger corroborating evidence. We include variables in `ARCD_long.dta` that record the source used to determine key classification variables, including whether a location qualifies as a camp, its operational timing, and its population count.

We consider the Population of Concern data, provided to us by UNHCR Global Data Service (2024b), as the highest-quality source. This is proprietary data that lists refugee population counts

by known refugee location for years 2021-2023. We rank it first because it comes from the primary authority on refugee enumeration and represents a recent, systematically collected dataset. Note that in variables listing data sources within the ARCD data, we refer to the Population of Concern data as the UNHCR POC Population Map.

Another vital source of information is the Operational Data Portal (data.unhcr.org), an online repository for all documentation produced by UNHCR field and regional offices, including but not limited to operational maps, updates, reports, and meeting minutes. Because it originates from UNHCR offices and captures reporting that supports field operations, we also consider the primary evidence extracted from the Portal as highly credible.

The UN-WASH Dashboard (Dubey, 2024) tracks water and sanitation indicators in camp areas and includes population counts for its camp observations. Although it originates from the UNHCR and is more recent, we rank this source lower than the Operational Data Portal and the Population of Concern data. This is because coverage is uneven across time and does not capture the full universe of refugee camps in any year. Observations are very sparse 2013–2016, and even in the relatively more comprehensive years that follow, the data represent only a subset of refugee camps rather than complete coverage.

The UNHCR Statistical Yearbooks, published annually from 1999 to 2016 (see the citation listing in [Section A.1](#) of the Online Appendix), rank fourth in our source hierarchy. The Yearbooks provide detailed annual listings of locations where people of concern were residing worldwide, along with information on the status of each location (e.g., whether refugees were residing in a planned camp or another type of site). These records are particularly useful for validating camp status, tracking operational timing, and identifying locations that do not appear in the POC Map. We consider the Statistical Yearbooks to be marginally less reliable than other UNHCR sources due to reporting discrepancies described in [Section 3.4](#).

Our data collection led us to consider other sources, including reports written by other UN entities, NGOs, and governments. While these still originate from stakeholders, we rank them lower than UNHCR documentation since the UNHCR is the primary registering authority in this context.

We rank media and academic sources lowest, largely because the information conveyed targets an audience that is not familiar with operational terminology of refugee-hosting, leading to broad and imprecise usage of the term "refugee camp."

3.1 List of Camps

Locations widely recognized in practice as refugee camps share many common features. They are purpose-built for refugee accommodation, managed by governments and UNHCR, and provide services through NGO partners. Yet drawing a clear boundary between refugee camps and other forms of refugee accommodation is often not straightforward. Below, we describe how we generate a finalized roster of camp locations based on the prevailing terminology used by corroborating sources. Our data repository also includes tabular information on the locations that, under our methodology, are classified as non-camps, so users can bring these locations into the analysis if desired.

Decision Rules: We derive our initial listing from the entries in the UNHCR POC Map (UNHCR, 2025), which includes both active and inactive locations. The POC Map pre-classifies locations using administrative site categories such as formal settlement, informal settlement, collective center, and individual accommodation in communities, rather than using “camp” as a standalone category. Sites considered camps fall in the "formal settlement" category, and we extract all past and present POC map locations tagged as hosting refugees in a "formal settlement". To supplement the UNHCR POC Map initial roster, we also extract the list of refugee hosting locations classified as camps in the Statistical Yearbooks (Section A.1).

We retain the listing in the ARCD roster if at least one credible source clearly and explicitly referenced the location as a camp. We document the entry’s validating source type in the ARCD_Long.dta file using variable Source_Type_Camp_Validation. If we cannot find substantiating evidence that the location is a camp, we move the listing to Non-Camps.xlsx and include notes on our own classification decision.

Refugee locations labeled as “settlements” present a challenge for classification because the distinction between camps and settlements varies across contexts and over time. Historically,

“settlements” were associated with stronger self-reliance programming: refugees were allocated land, encouraged to farm, and were expected to transition away from humanitarian assistance (Armstrong, 1988; Chambers, 1982; Harrell-Bond & Verdirame, 2005; Harrell-Bond, 1986).⁶ In practice, however, some sites that function as refugee camps are also labeled as settlements by governments or agencies because “settlement” carries more positive associations of autonomy and integration than “camp”. This tendency makes it difficult to distinguish substantively different hosting models from differences in terminology alone.

We retain a location in the ARCD roster, even if some sources label it a “settlement,” when at least one high-quality source in our ranking explicitly identifies it as a camp or reports on it in a manner that suggests that it is not operationally distinguished from camps. For example, Kalobeyi in Kenya is often referred to as a settlement, but UNHCR reporting covers it jointly with Kakuma Camp, suggesting operational similarity across the two sites. Uganda’s settlements are referred to as camps in numerous Statistical Yearbooks. We include these locations in the ARCD roster but flag them as potentially representing a different hosting approach using the variable `Settle_model`. If a location is referred to only as a settlement across validating sources and reflects a substantively different operational model, we instead place it in `Non-Camps.xlsx` and classify it as a settlement. For example, Mishamo in Tanzania has consistently been referred to as a settlement by government and UN agencies, and UNHCR Tanzania planning documentation distinguishes Mishamo from its camp-focused refugee response (UNHCR, 2019). We therefore move Mishamo to the `Non-Camps.xlsx` file. This approach provides flexibility for users, who can revise these classifications and adjust the sample according to their preferred definition of a camp, or alternatively focus specifically on locations classified as settlements.

We remove sites from the ARCD dataset if we are unable to find any validating sources confirming that a camp by the listed name ever existed. We also remove sites if sources exclusively refer to the site as an alternative form of refugee accommodation, such as a transit center, informal settlement, or administrative division. We keep a listing of these removed locations in the `Non-Camps.xlsx` file.

⁶Self-reliance programming often falls short of eliminating aid dependence (Harrell-Bond, 1986; Verdirame & Pobjoy, 2013).

In some cases, there is conflicting evidence on the type of site. For example, a location may be listed as a transit center in one source and as a camp in others. In such a case, we include the site in the ARCD camps listing. To signal that we are less confident in its inclusion than with other fully validated listings, we construct a binary variable indicating whether conflicting evidence makes us less confident that the location is indeed a camp (*Potentially_Non_Camp*). We include the link or the reference to the sources that generate the reason for doubt in the dataset.

False negatives may also be present in the POC map: locations classified under alternative site categories (including transit centers, informal settlements, and collective centers) may nonetheless be considered refugee camps by other sources. We therefore systematically review listings in these non-camp categories for SSA locations that may function operationally as camps. When a location is referred to as a refugee camp in relevant documentation, we include it in the final ARCD camp roster. Locations reviewed from these alternative site categories and determined not to be camps are not retained in the *Non-Camps.xlsx* file.

3.2 Camp Coordinates

Decision Rules: The POC Map provides a geolocation (latitude and longitude) point for each camp. For a small number of camps that do not have their longitude and latitude documented in the POC Map, such as the Oru Camp in Nigeria and the Nicla Camp in Ivory Coast, we use geocoordinates available from UNHCR operational maps that are available on *RefWorld.org*.

UNHCR's Global Data Service (GDS) performed a check on all ARCD locations. In late 2025, The GDS team examined every ARCD-listed geopoint against their internal, non-public data and provided corrected locations for 10 ARCD sites. We directly modified the geocoordinate variables for these locations in the ARCD data.

We further validate geolocations using Google Earth historic imagery. These are high-resolution optical satellite maps that can be visually inspected across multiple time periods. For each ARCD geopoint, we visually examine the location using map layers from the camp's operational time period. Grid-like planned infrastructure or structures that appear or disappear rapidly are suggestive

of a camp location rather than the land use change expected from ordinary economic or population growth. If we could identify a piece of large planned infrastructural development like this in the location on Google Earth, we considered the point validated. When there was no such site, we examined the general area around the geolocation, and if we could locate infrastructure resembling a camp, we updated the coordinates in the ARCD along with a note of the change retaining the original coordinates. If we could not find a structure resembling a camp in the general area, we note that the coordinates are not validated and at best an approximation and the exact location of the camp is unknown. In the main dataframe, we encode the variable `Coords_validated` to indicate whether the site location could be confidently corroborated via Google Earth imagery.

Observations and Challenges: For some ARCD locations, the Google Earth historical imagery is sparse, or the camp's operational lifespan is short. Google Earth historical imagery is also sometimes obscured by cloud or canopy cover. When we cannot find a single clear image of the location during its operational period, we cannot fully validate the site.

3.3 Camp Starting and Closing Years

Decision Rules: Data on the camps' operational lifespans represents a major contribution of the ARCD, as reliable and systematic data on when camps opened and closed is largely absent from existing sources. We include opening and close years for an ARCD entry if there is a credible source in the public domain that referenced the starting year or the closing year.

Using the vetted camp listing, we review each entry, searching through credible sources in the public domain that document their operational years. When we encounter contradictory information on closing and opening dates, we use the date with the highest number of corroborating sources. If sources list more than one date, and each of those dates had the same number of corroborating sources, we note in the dataset that sources contradict each other and provide references for those sources. We also include variables in `ARCD_long.dta` listing the highest ranking source affirming the encoded opening and closing dates (`Source_Type_Closing` and `Source_Type_Opening`), in case researchers may see some as less reliable and omit those observations from their analysis.

To supplement this information, we provide an alternative definition of camp operational time periods based on each camps' appearances in the Statistical Yearbooks. To do this, we record the last year a camp appears in the Yearbook series and the first year. This provides a proxy year for camp opening as long as the entry is not listed in the 1999 Yearbook (the the first year of the series). It provides a proxy for closure as long as the camp closed before 2016, the last year of the Yearbook run. We report these years in separate variables (`First_Mention_in_Yearbook` and `Last_Mention_in_Yearbook`).

Observations and Challenges: When a camp opens, there is often an explicit opening date, while for closings, a camp often dwindles away until operational support for the camp eventually ceases, and a public pronouncement or media report about the closure is much less common. While we were able to find opening and closing dates for most of the camps in the sample, there were (mostly older) camps that had little documentation are missing opening or closing dates. For example, while we could find evidence that Gerihun Camp in Liberia was open before 2000, the documentation pre-2000 is so sparse that it is impossible to place an exact opening year.

As we characterized the reasons that ARCD-identified camps closed, we found that the most common cause was sufficient levels of repatriation,⁷ resulting in the camp's population to fall below the point where it makes sense to continue supporting or administering the camp. This is often a gradual process. For example, Molangue Camp in the Central African Republic hosted refugees from the Republic of Congo, who began repatriating in 2004. This process took two years, during which time the population of the camp fell from around 3,000 to a little over 100 people. In 2007, the UN ended its repatriation efforts of Congolese in the Central African Republic, leaving 112 people still living in the camp who, according to UNHCR documentation, remained there until 2009. In cases like this, there are multiple reasonable options for choosing a specific year as the "closing" year. We use 2007, the year repatriation ended, as the camp's closing date. Population data beyond this date is retained in the dataset, allowing researchers to apply alternative closure definitions if

⁷In the best case scenario, repatriation occurs when it is safe for refugees to voluntarily return home, but we cannot assume that the cessation of hostilities in the country of origin is a necessary condition for repatriation. See for example Lawrence (2017).

desired.

3.4 Annual Camp Populations

Decision Rules: We include a population count in the dataset if we could find a credible source in the public domain that referenced the population of the camp in a particular year. We draw on multiple sources for population figures, with the predominant source varying by observation year. For 1999–2016, population data come mainly from the Statistical Yearbooks. For 2017–2020, they are largely drawn from the UNHCR WASH dashboard. For 2021–2023, they are primarily derived from Population of Concern data. Sometimes multiple population counts are available for different dates within the same year. In these cases, we include the population figure that comes from the month closest to the end of the year. We do this because we source a large share of the population information from the Statistical Yearbooks, which report counts based on the end of the year. Aligning our timing conventions with this reporting structure helps ensure consistency across the majority of population observations in the dataset. We also include a variable in `ARCD_long.dta` that identifies the month that the camp-year population measure reflects. This helps researchers avoid interpreting population measures taken at the end of one calendar year and the beginning of the following calendar year as a true year-to-year change.⁸

In cases where we have sources for population information corresponding to the same point in time that contradict each other, we prioritize sources based on the aforementioned ranking. We include a variable in `ARCD_long.dta` that identifies the population information source (`Source_Type_Population`), in case researchers may see some as less reliable and omit those observations from their analysis.

We do not employ any imputation method for missing observations and leave it to researchers to pursue these methods independently if they feel they would be useful.

Observations and Challenges: For years 1999–2016, almost all camp population data comes from

⁸Researchers seeking higher-frequency population data for specific camps can query the UNHCR WASH dashboard (Dubey, 2024), which provides an accessible API and is a population data source that, in some cases, reports more than one count per year.

the Statistical Yearbooks. While they are relatively comprehensive, they have issues both in a lack of consistent scope for identifying camps and inconsistencies in definitions and reporting practices. This leads to gaps in the data where certain camps did not have populations reported, even while operating as UNHCR-administered refugee camps. The most common cause of missing population data was size thresholds for reporting. Each Yearbook has its own protocol for what camps and regions to report, with varying populations required for a camp to be listed depending on the year. This leads to smaller camps between 2011 and 2016 being potentially under-reported. This under-reporting of small camps is exacerbated by the aggregation of camps in the Yearbooks into regions if the camps are sufficiently small. For example, the Shagarab camps in Sudan are reported as a single aggregated observation (“Shagarab 1, 2 and 3”) in the 2007 Yearbook, whereas prior Yearbooks report on the three locations separately.

3.5 Events Precipitating Camp Opening and Reasons for Closure

Decision Rules: Our rule for determining the precipitating event and reason for closure of a camp was whether we could find a credible source in the public domain that referenced the population of the camp and discussed its reasons for forming or disbanding.

Observations and Challenges: While information on the events precipitating the openings of camps is readily available, information on the reasons for closing camps is relatively scarce. As a result, the dataset is much more complete in tracking reasons for opening than it is in tracking reasons for closing.

3.6 Geospatial Characteristics of the Camp Location

Process: We use camp geopoints to generate zonal statistics that characterize camp locations. Appendix [Table A1](#) documents all spatial variables and sources. These variables extend the analytical scope of the ARCD, and we use them to examine sample characteristics in Section 5.

To do this, we combine the camp geopoints with spatial reference data (Africa GeoPortal,

2017; GeoNames, 2026; NEP-WCMC & IUCN, 2025), and calculate the Euclidean distance to the nearest international border, protected area, provincial or national capital, and urban center. We use global accessibility maps (Nelson et al., 2019) to estimate the travel time to the nearest port. We capture the ecological context using biome and ecoregion classifications from Dinerstein et al. (2017).⁹ We also calculate terrain ruggedness using Copernicus elevation data (ESA, 2024) and the Terrain Ruggedness Index (TRI) from Riley et al. (1999). Because the underlying gridded data is coarse, we use point intersection rather than zonal statistics to derive annualized measures of land surface temperature, precipitation, drought exposure (Abatzoglou et al., 2018; Reig-Gracia & Latorre Garcés, 2023), and local GDP (Kummu et al., 2025). We additionally use the point location to determine the administrative area in which the camp is located. Using the Global Administrative Areas database Version 4.1 (GADM, 2024), we determine the first, second, and (when available) third administrative division of the camp location.

We use UNHCR camp-boundary polygons, available for a subset of ARCD locations (UNHCR, 2026), to measure observed camp land area and calibrate the spatial scale of zonal statistics. Calculating the polygon areas and merging them onto the main dataframe (variable `camp_area_km2`), we find that the vast majority (94%) of camps are larger in area than a buffer with a 250 meter radius (see Appendix Section A.3). Ideally, the spatial extent used to generate zonal statistics would scale proportionally with each camp's observed area. But because polygon boundaries are unavailable for many ARCD locations, we instead use a uniform 250 meter buffer around each camp geopoint as a common basis for zonal measures characterizing the camp's interior.

Within this buffer, we compute time-varying zonal statistics describing environmental and socioeconomic conditions. Environmental measures include vegetation composition (forest, non-forest vegetation, and non-vegetated cover), annual forest loss, and the normalized difference in vegetation index (NDVI) (Didan, 2021; DiMiceli et al., 2022; Hansen et al., 2013; Wan et al., 2021). We also measure flood exposure and the share of the buffer area falling into each soil-constraint class using Hengl et al. (2021). We generate socioeconomic zonal measures, including within-buffer

⁹The biome captures the broad bioclimatic zone (e.g., grassland or tropical rainforest), while the ecoregion provides a finer ecological classification within that zone.

total population and mean nighttime light intensity (Elvidge et al., 2021; WorldPop, 2025).

Observations and Challenges: The spatial data used to generate time-varying zonal statistics is not always available for all years in the ARCD panel, with missing coverage especially common in 1999. Another limitation is that our sampling buffer is uniform, but in reality, camps vary in size (Figure A1). When a camp is much larger than the 250 meter buffer used for sampling, the estimated “interior” attributes reflect only the local environment around the representative point rather than the settlement as a whole. Consequently, measures such as forest cover or road access become sensitive to point placement, introducing potential measurement error when substantial within-camp heterogeneity exists. This issue is particularly important for large camps, where environmental and infrastructural conditions may vary considerably across the full settlement extent.

4 Descriptive Properties of the ARCD

In this section, we demonstrate the descriptive features of the ARCD data, showing how the ARCD camp population data reflects UNHCR country-level data and overall trends in refugee encampment over time across the sub-continent.

4.1 Consistency with UNHCR Data: National Level Population Validation

To assess the accuracy of our camp population data, we compare it with country-level proprietary counts of camp populations over time from UNHCR Global Data Service (2024a). We construct comparable country-year totals from the ARCD by aggregating population counts across camps within each country and year, and then we examine the correlation between these aggregates and the corresponding country-year panel from the UNHCR Global Data Service (2024a). When one is regressed on the other, the coefficient is 0.958 with an R-squared of 0.854. Figure 2 contains the scatterplot of the two measures illustrating this strong relationship. In the chart, we plot the 45-degree line, which illustrates where the points would fall if we had a perfect one-to-one matching. The relationship between the two closely follows the 45-degree line, with some noise around it.

While the two measurements do not match perfectly one-to-one, altogether this confirms that our measurements are closely related to the aggregate figures tracked by the UNHCR and suggests confidence in the data.

The gaps between encampment population counts are likely explained by the following factors. The first, and most obvious, reason is that the UNHCR count will be higher than the ARCD-based count whenever camp population counts are not complete for all camps within the country. The case in which the ARCD offers a higher count than the UNHCR is more complex. Spot-checking reveals that this is often due to the underlying ARCD sources classifying sites as camps that are, in the UNHCR GDS data, considered to be “Informal Settlements”. While less common, it may also be the case that a refugee population no longer serviced by the UNHCR is nevertheless still included in the population data underlying ARCD.¹⁰ These discrepancies evoke new questions about the strategies undertaken when reporting camp statistics. Additional study of changes in terminology over time may reveal the role of institutional strength, funding incentives, refugee population identity, and other factors in explaining these incongruities.

4.2 Spatial Distribution of Camps

[Figure 3](#) maps the ARCD by year of activity, including only the main camps and omitting any sub-camp geopoints for visual clarity. The spatial patterns align with our knowledge of conflict dates over time. For example, we observe numerous camps open after 1999 in West Africa in the wake of the Second Liberian Civil War and the First Ivorian Civil War. Camps multiply in Chad in 2008 in response to the Darfur War. Intensifying violence in the Central African Republic results in a ring of camps around the country in host countries such as Cameroon, Chad, and the Democratic Republic of Congo by 2016. Between 2016 and 2024, additional camps for those seeking refuge from Sudan’s most recent Civil War appear in Chad, South Sudan, and Ethiopia. During this period, many camps opened, and a smaller number closed, but in some countries we observe persistence,

¹⁰For example, in 2002, a population of Burundian refugees in Tanzania often referred to as the “old caseload” were living unassisted in previously designated agricultural settlements (Ngayimpenda, 2005). The gap between official UNHCR statistics and ARCD-based counts is roughly the size of this population in 2002. We suspect they may have been counted as part of a camp population in this year.

with camps remaining active over the entire period. This is particularly the case in the Great Lakes and Horn of Africa region.

4.3 Precipitating Events and Closings

Figure C1 in the Online Appendix presents the counts of ARCD camps by the main precipitating events in the dataset. The two events that led to the first and second largest count of refugee camps were both related to conflicts in Sudan: The South Sudanese Civil War (45 camps) and the Second Sudanese Civil War (25 camps). Although Sudan accounts for a large number of refugee camps, conflicts in other regions of the continent additionally lead to many camp openings. This is particularly visible for Central Africa (the Central African Republic Civil War and the Burundian Civil War), the Horn of Africa (the Somali Civil War), and Southern Africa (the Angolan Civil War).

Figure C2 (Online Appendix) illustrates counts of the encamped population by precipitating event in 1999, 2008, 2016, and 2024 according to the ARCD data. In the earliest year (1999), The Burundian Civil War constituted a major contributor to the encamped refugee population in SSA. In later years, the encamped population was predominately displaced by the Somali Civil War and conflicts in Sudan. The former conflict started prior to 1999, and the low encampment population count in the 1999 data associated with Somalia's civil war reflects not only the additional waves of Somali displacement to Kenya in the 21st century, but also the incomplete population characteristics of the ARCD population data for certain country-years. In this case, counts for the Dadaab Camp complex only begin in 2000.

Figure C3 presents the counts of camps by the reasons we could identify for closing. For many of the camps (173) we were not able to identify a specific reason. Of the camps for which we could find information, most are still open. The most common reason for closing in the ARCD is the successful repatriation of refugees. The ARCD data strongly suggests that camp closure is rarely motivated by a deliberate effort to integrate refugees into community settlements. If this does occur, these events are rarely publicized or noted in a way so that a researcher at a distance can identify them. In fact, we were only able to confirm 10 instances in which camps closed as part of a

disencampment and integration process.

4.4 Camp Longevity Patterns

The extraordinary duration of certain camps has been noted by many observers over the decades. For example, the Kakuma Camp in Kenya has been open since 1992 and still hosts over 203,000 inhabitants as of 2023. To date, there has been no systematic quantitative analysis of what drives the lifespan of these camps to understand what qualities are associated with closure. Why do some camps close quickly, while others, like Kakuma, become “accidental cities” (Jansen, 2015)?

While a full survival analysis is beyond the scope of this paper, we offer basic descriptions of longevity to highlight the potential use of the ARCD. Of the 182 camps in the dataset that came into existence since 1999, 31.87% have closed, while the remaining 68.13% are still open. Among camps that have closed, the average length of time the camps have been open has been 5.74 years, lasting anywhere between one and 19 years. Of those that continue to operate, the average length of time the camp has been in existence (as of 2024) was 11.51 years. When we use the entire dataset to calculate this figure (using the Statistical Yearbook mentions to infer openings and closings when otherwise verified information is not available) and including sub-camps, it is 17.01 years. This is almost identical to the controversial 17-year lifespan of camps often cited (Moore, 2017; UNHCR, 2006). Altogether, looking at the post-1999 data, once a camp opens, we expect a camp to last on average at least 9.67 years.

Using the Kaplan-Meier (K-M) approach, we estimate survival functions for camps and examine differences in these survival functions by conflict timing, maximum camp population size and distance to the capital city. For this assessment, we use observations from all camps that came into existence in 1999 or later, and we use the timing of the earliest appearance in the Statistical Yearbooks (other than 1999) to date the starting year of camps for those that do not contain an otherwise verified starting year. We also use the last mention in the Statistical Yearbooks (other than 2016) to date their closing if we do not have a verified closing year. We omit all camps for which we cannot verify any closure at all. [Figure 4](#) contains these estimates. The top left panel contains the

K-M estimate of the survival function for all refugee camps in the data that were established in 1999 or later. The probability of survival is high, above 50%, over the entire domain. The probability of survival declines steadily with the age of the camp until it becomes very flat at about 15 years of age. At 19 years, the probability of survival becomes entirely flat at about 55%. This means that once a camp is established in this period, there is an over 50% chance that camp will last at least 19 years. There are 23 camps that last 19 years or longer in this period: 14 in Chad, two in Burundi, and one each in the Republic of Congo, Djibouti, Eritrea, Mozambique, Namibia, Rwanda, and Uganda.

The top right panel illustrates the K-M survival functions for camps that have experienced the cessation of the conflict that prompted the refugee flow that led to the establishment of the camp versus those that have not. Here we see a clear divergence. It is much more likely for a camp to close after the cessation of the precipitating conflict. However, the survival function for camps post-conflict indicates that the cessation of the precipitating conflict is not sufficient for camp closure in many cases. After the cessation of the conflict, results suggest there remains a 50% chance a camp operates for up to 12 years and beyond. The bottom left panel shows the K-M survival functions for camps divided into quartiles based on how close they are to the capital city in the country in which they are located. We observe a notable deviation for camps in the first quartile (farthest from the capital city) relative to the general trend for the other three quartiles, suggesting that camps that are more remote are more likely to stay open longer. Last, in the bottom right panel we show the K-M survival functions for the camps divided into quartiles based on the maximum size that the camp population reached over its lifespan. There is a clear relationship between the size the camp achieves and the length of time the camp stays open, with camps in the first quartile lasting longer than those in the other quartiles, and the survival functions declining for each successive quartile.

5 Descriptive and Causal Analyses

This section uses the zonal statistics in the ARCD to explore descriptive attributes of camps and demonstrate how the data can support causal empirical analysis. These applications represent only an initial set of use cases, as the ARCD can be combined with a wide range of spatially explicit

datasets to address broader research questions.

5.1 Geospatial characteristics of camps

We restrict the ARCD to geolocated main camp entries, omitting sub-camps to avoid over-counting ($n = 323$). We generate two comparison samples. For our “border comparison” sample, we randomly select $n = 2,937$ points stratified by ARCD national location distributions. All points must be 30-50 km from a camp geolocation, outside water bodies, and outside protected areas. Using this sample, we can assess whether encampment areas differ from nearby border regions. Our “country comparison” sample ($n = 2,354$) is also stratified by host country distribution and excludes water bodies and protected areas. For this sample, however, geopoints can be drawn from any habitable area in the country, resulting in a sample that reflects national averages.¹¹

Table 2 reports camp characteristics alongside both comparison samples, with t-tests identifying significant differences. On average, camps are not far from internal borders (53.0 km) or protected areas (58.5 km) and are far from national (450.2 km) and provincial (100.4 km) capitals. Relative to national averages, camps are significantly closer to protected areas and are notably *closer* to provincial capitals, possibly reflecting the preference to locate camps near key distribution points.

The UNHCR recommends camps be placed at least 50 km from the refugee-sending country’s border (UNHCR, 2024). The average distance (53.0 km) meets this target but with considerable variation ($SD = 53.7$). Figure C4 shows camp distances to the nearest border by region. In Central Africa, 16.5% of camps are within 10 km of the border and nearly half (48.5%) are within 30 km. A non-negligible share of East (33.8%) and West African (29.4%) camps are also within 30 km.

Distance to the nearest capital also varies widely. Figure C5 shows that 19.4% of camps are within 30 km of the nearest capital. These are predominantly located in West Africa (Figure C6). But two in five camps (39%) are 100 km or farther from the nearest capital. Such remote camps are more common in Central and East Africa.

¹¹Both comparison samples target 3,000 points. The country comparison sample is marginally smaller because invalid draws (in water or protected areas) are dropped without replacement, whereas the border comparison sample oversamples candidate points before applying the same exclusions.

Results for terrain ruggedness suggest that UN recommendations on flat terrain (UNHCR, 2007) are consistently realized. Camp locations sit on marginally flatter terrain than either comparison sample, indicating that logistical considerations shape site selection. Most camps (76.2%) are in grasslands, with a smaller share in dry or wet broadleaf forests (19.2%). Relative to the national comparison sample, camps are more likely to be located in grasslands and less likely to be located in forest biomes. However, camps do not differ substantially from the border comparison sample in biome composition, suggesting that this pattern primarily reflects the ecological characteristics of border regions rather than local siting decisions.

We examine time-varying features in the relative period two years prior to camp opening to evaluate the baseline attributes of areas selected for camp siting (Table 2). At baseline, camp areas have low forest cover (10.1%) and considerable non-forest vegetative cover (62.3%). Both measures differ significantly from the border comparison sample. Camp areas have lower forest cover and higher non-forest vegetation cover than nearby non-camp locations. These baseline characteristics reinforce the biome finding. Even within border regions, sites with less canopy cover tend to be selected. This tendency toward less forested areas may reflect feasibility considerations, or simply that these areas face less competing demand for land. Relative to national averages, both forest and non-forest vegetative cover are marginally lower.

Nighttime light radiance is low in the camp and comparison samples, reflecting limited electrification across camp-hosting countries generally. Camp areas have similar mean radiance as compared to national averages, but radiance is significantly lower for camp locations than border comparison points (0.005 versus 0.096). Though both values are low in absolute terms, this difference suggests that camps tend to be sited in areas with less prior settlement, infrastructure, and/or economic activity than other nearby locations.

5.2 Impacts on Forest and Vegetation Cover of Camp Interiors

Our causal assessment uses zonal statistics on the percent forest, nonforest vegetation, and non-vegetative cover, along with a DID estimation strategy, to examine the question: What happens

to different types of vegetation cover within a camp interior once a camp opens? Scholars and practitioners familiar with camp management understand that camps may require land clearing to make way for the “camp footprint” comprised of settlement blocks, latrines, fire breaks, and necessary infrastructure and buildings (UNHCR, 2005, 2007). We therefore expect camp opening to result in vegetation loss due to settlement logistics. For our purposes, demonstrating this causal effect using the ARCD provides some additional validation for the data, as it suggests that, on average, the data are not geographically misclassifying camp locations or opening years.

Our estimates use the primary camp geolocation and operational years, omitting sub-camps as in [Section 5.1](#). We exclusively rely on the border comparison points as our control group, given their similarity to camp areas in terms of remoteness and geography. We rely on a DID identification strategy exploiting trends in forest and vegetation cover before and after camp openings. Opening year varies across camps, and we report relevant summary statistics overall and for the year prior to camp opening in [Table D2](#). Given the potential for bias in empirical settings with heterogeneous treatment timing (Goodman-Bacon, 2021), we rely on the estimator developed by de Chaisemartin and D’Haultfœuille (2024) rather than a standard two-way fixed effects approach. Using this robust estimator, we generate event studies, which we use to derive the cumulative average treatment effect (CATE). See [Section B](#) in the Appendix for full technical details.

[Table 3](#) provides cumulative average treatment effect estimates for all camps and camps by population size, biome, and region. Camp opening reduces forest cover by 2.5 ppt over the seven-year assessment period, a 25.5% decrease relative to the pre-treatment mean (9.8%, [Table D2](#)). Non-forest vegetation declines by 2.3 ppt (3.7% relative to the 62.6% sample mean), and non-vegetated cover increases by 4.8 ppt, consistent with cleared land transitioning to bare ground. Event study results show these losses occur within the first two years of camp opening and stabilize thereafter ([Figure 5](#)), with null placebo estimates supporting parallel pre-trends. This pattern of rapid initial clearing followed by stability points to purposeful land preparation for the camp footprint. Our evidence is less consistent with biofuel harvesting as the primary driver, as fuelwood and charcoal consumption would produce continuous tree loss until the surrounding area is exhausted, rather than the sharp,

short-lived decline we observe.

Heterogeneity by maximum camp population suggests that land cover impacts generally increase with camp size. We detect no effect on forest or vegetative cover for camps with fewer than 1,000 residents. Camps with populations between 1,000 and 9,999 experience a 3.0 ppt decline in forest cover, while camps with 10,000-30,000 residents see a 3.3 ppt reduction in forest cover and a 3.7 ppt decline in vegetative cover. For these camps, the findings likely reflect congestion as camp populations rise. Unless the encampment is designed as an agricultural settlement, the land allocated for a camp is often modest in area. UNHCR strongly encourages planners to maintain existing natural ground cover when possible (UNHCR, 2023). However, during displacement emergencies, camps may absorb more people than they were originally designed to accommodate, and unless additional land can be allocated, trees may need to be cleared to make way for additional households.

For camps with populations above 30,000, forest impacts are somewhat smaller (2.1 ppt), though vegetation loss is larger (4.3 ppt). This smaller forest effect should not be interpreted as evidence that very large camps are less environmentally disruptive. The largest population category is disproportionately composed of camps in Uganda, where camps tend to occupy larger land areas and therefore have lower population densities, as well as camps in comparatively arid settings such as Chad and Ethiopia where baseline forest cover is already limited (Table D2)..

Examining results by biome, we find that forest losses are significantly higher in forest bioclimatic regions, with camp opening associated with a cumulative forest cover reduction of 5.5 ppt. In grassland biomes, we also observe a 2.0 ppt reduction in forest cover. Regionally (Figure C7) the camp-induced forest losses are largest among camps in West Africa (-4.9 ppt) followed by Central Africa (-4.0 ppt) and East Africa (-2.5 ppt). The variation of effects across regions likely reflects differences in baseline forest cover levels (Table D2), with West African camps exhibiting the highest forest cover one year prior to camp opening.

In summary, camp establishment leads to a significant decrease in forest and vegetative cover within the camp boundary. This reduction likely speaks to land clearing for logistical reasons, such as settlement blocks, fire breaks, and infrastructure. Tree loss may reflect biofuel harvesting

independent of planned logistics, though the event study trends do not align with biofuel harvesting patterns, which we would expect to result in continuous tree reductions until all trees are exhausted. Importantly, larger camps are associated with greater overall land cover change, though the forest effects are somewhat smaller for the largest population category, likely reflecting the prominence of lower-density settlement-style camps in Uganda and camps located in more arid environments.

6 Conclusion

We have introduced a new and unique dataset for researchers to advance questions related to refugee settlement and the phenomenon of refugee encampment in the SSA region. While this dataset comes mostly from publicly available sources, we are aware of no previous effort to consolidate and verify the information in order to track refugee camp locations and years of operation in this way. The dataset opens new opportunities to study highly localized effects of refugee camps on many relationships, and explore variation by time, country, or policy framework.

We present descriptive work on refugee camps from 1999-2024, covering the timing and location of camps, their sizes, patterns of longevity, and the historical events that precipitated their creation, and offer two applications demonstrating the ARCD's analytical potential. Our geospatial assessment shows that camp areas align with some logistical guidelines but often remain close to borders and protected areas and far from provincial capitals. Border proximity accounts for many site characteristics that diverge from national averages, though in West and Central Africa, forest and vegetative cover differs from other border regions, reflecting local selection on logistics or land value. Our causal assessment finds that camp openings significantly reduce forest cover within the camp interior, with all losses occurring within the first one to two years of operation, consistent with land clearing for shelter and infrastructure. We find no effect on vegetation cover, suggesting camps are built around existing vegetation. Impacts are largest for camps with bigger populations, likely reflecting higher densities and less flexibility in siting infrastructure around existing foliage.

There are numerous opportunities to build on this work. We intend to expand the dataset to include other world regions. Additional work includes developing a similar dataset on other known

precise non-camp sites where refugees are located, creating opportunities to study other dimensions of refugee hosting beyond the encampment modality. This expansion would include locations of spontaneous self-settlement, locations where refugees have integrated into the local population, as well as other institutionalized refugee locations like transit centers. Additionally, archival data collection could facilitate expanding the dataset's temporal scope, examining camp operational years prior to 1999. Last, there are many camp-level variables that could be added to the ARCD related to aid provision, demographics of the refugee population, and inter-camp policy variation, as well as improved camp footprint estimates that would allow researchers to derive population density measures and better contextualize land-use impacts. While such future directions would be valuable, we believe the ARCD in its current form will enable new research efforts matching the compiled data to household, policy, and geospatial information. Our hope is that the data will enrich the expanding research discourse on the topics of refugee movement, refugee and host welfare, and the impacts of forced displacement in SSA.

7 Bibliography

References

- Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, 5, 170191. <https://doi.org/10.1038/sdata.2017.191>
- Africa GeoPortal. (2017). Africa cities [Accessed February 1, 2025]. <https://www.africageoportal.com/datasets/africa-cities-1/about>
- Alix-Garcia, J., & Saah, D. (2010). The effect of refugee inflows on host communities: Evidence from Tanzania. *The World Bank Economic Review*, 24(1), 148–170. <https://doi.org/10.1093/wber/lhp014>
- Alix-Garcia, J., Walker, S., Bartlett, A., Onder, H., & Sanghi, A. (2018). Do refugee camps help or hurt hosts? the case of Kakuma, Kenya. *Journal of Development Economics*, 130, 66–83. <https://doi.org/10.1016/j.jdeveco.2017.09.005>
- Alloush, M., Taylor, J. E., Gupta, A., Rojas Valdes, R. I., & Gonzalez-Estrada, E. (2017). Economic life in refugee camps. *World Development*, 95, 334–347. <https://doi.org/10.1016/j.worlddev.2017.02.030>
- Armstrong, A. (1988). Aspects of refugee wellbeing in settlement schemes: An examination of the Tanzanian case. *Journal of Refugee Studies*, 1(1), 57–73. <https://doi.org/10.1093/jrs/1.1.57>

- Baez, J. E. (2011). Civil wars beyond their borders: The human capital and health consequences of hosting refugees. *Journal of Development Economics*, 96(2), 391–408. <https://doi.org/10.1016/j.jdeveco.2010.08.011>
- Baseler, T., Ginn, T., Hakiza, R., Ogude-Chambert, H., & Woldemikael, O. (2025). Can redistribution change policy views? aid and attitudes toward refugees. *Journal of Political Economy*. <https://doi.org/10.1086/736209>
- Bertinelli, L., Cömertpay, R., & Maystadt, J.-F. (2025). Ethnic diversity and conflict in sub-saharan Africa: Evidence from refugee-hosting areas. *Journal of Development Economics*, 172, 103393. <https://doi.org/10.1016/j.jdeveco.2024.103393>
- Betts, A., Bloom, L., Kaplan, J., & Omata, N. (2017). *Refugee economies: Forced displacement and development*. Oxford University Press.
- Betts, A., Flinder Stierna, M., Omata, N., & Sterck, O. (2023). Refugees welcome? inter-group interaction and host community attitude formation. *World Development*, 161, 106088. <https://doi.org/10.1016/j.worlddev.2022.106088>
- Bousquet, J., Gasten, A., Kadigo, M. M., Maystadt, J.-F., & Salemi, C. (2025). Does perceived labor market competition increase prejudice between refugees and their local hosts? evidence from Uganda and Ethiopia. *Journal of Development Economics*, 175, 103481. <https://doi.org/10.1016/j.jdeveco.2025.103481>
- Chambers, R. (1982). Rural refugees in Africa: Past experience, future pointers. *Disasters*, 6(1), 21–30. <https://doi.org/10.1111/j.1467-7717.1982.tb00740.x>
- Chambers, R. (1983). *Rural development: Putting the last first*. Routledge.
- Coniglio, N. D., Peragine, V., & Vurchio, D. (2023). The effects of refugees' camps on hosting areas: Social conflicts and economic growth. *World Development*, 168, 106273.
- Das, N., Gupta, A., Mingo, C., & Zhu, H. (2025). Impacts of financial literacy training on refugee youth outcomes. *The Journal of Development Studies*, 1–25. <https://doi.org/10.1080/00220388.2025.2487670>
- de Chaisemartin, C., & D'Haultfœuille, X. (2024). Difference-in-differences estimators of intertemporal treatment effects. *Review of Economics and Statistics*.
- d'Errico, M., Mariani, R. D., Pietrelli, R., & Rosati, F. C. (2022). Refugee-host proximity and market creation in Uganda. *The Journal of Development Studies*, 58(2), 213–233.
- Didan, K. (2021). Mod13q1 modis/terra vegetation indices 16-day 13 global 250m sin grid v061. *NASA EOSDIS Land Processes DAAC*. <https://doi.org/10.5067/MODIS/MOD13Q1.061>
- DiMiceli, C., Sohlberg, R., & Townshend, J. (2022). Modis/terra vegetation continuous fields yearly 13 global 250m sin grid v061 [dataset]. <https://doi.org/10.5067/MODIS/MOD44B.061>
- Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E. C., Jones, B., Barber, C. V., Hayes, R., Kormos, C., Martin, V., Crist, E., et al. (2017). An ecoregion-based approach to protecting half the terrestrial realm. *BioScience*, 67(6), 534–545. <https://doi.org/10.1093/biosci/bix014>
- Dubey, Y. (2024). Unhcrwash: Wash data from refugee camps and settlements (unhcr). <https://doi.org/10.5281/zenodo.14185117>
- Elvidge, C. D., Zhizhin, M., Ghosh, T., Hsu, F.-C., & Taneja, J. (2021). Annual time series of global viirs nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sensing*, 13(5), 922. <https://doi.org/10.3390/rs13050922>
- ESA. (2024). Copernicus global digital elevation model [dataset]. <https://doi.org/10.5069/G9028PQB>

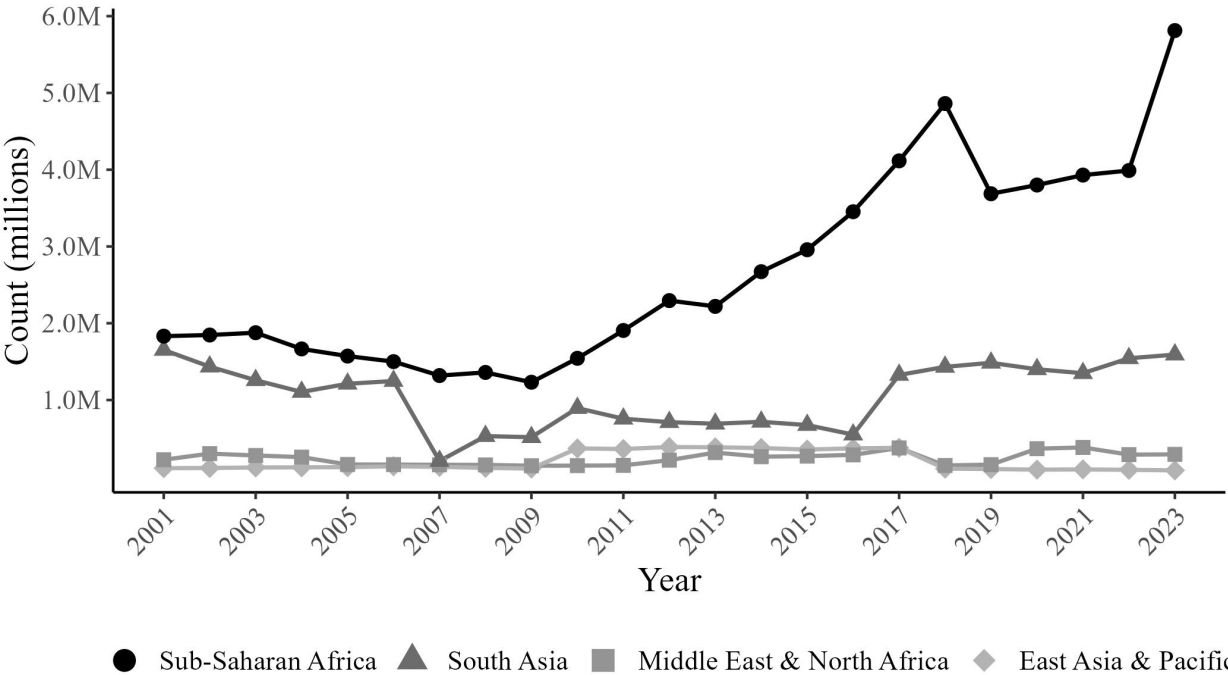
- Fajth, V., Bilgili, Ö., Loschmann, C., & Siegel, M. (2019). How do refugees affect social life in host communities? the case of congolese refugees in Rwanda. *Comparative Migration Studies*, 7(1). <https://doi.org/10.1186/s40878-019-0139-1>
- GADM. (2024). Database of global administrative areas, version 4.1 [Accessed 28 April 2026].
- GeoNames. (2026). Geonames gazetteer [Accessed April 1, 2026]. <https://www.geonames.org/>
- Ghosn, F., Braithwaite, A., & Chu, T. S. (2019). Violence, displacement, contact, and attitudes toward hosting refugees. *Journal of Peace Research*, 56(1), 118–133. <https://doi.org/10.1177/0022343318804581>
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254–277.
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160), 850–853. <https://doi.org/10.1126/science.1244693>
- Harrell-Bond, B., & Verdirame, G. (2005). *Rights in exile: Janus-faced humanitarianism*. Berghahn Books.
- Harrell-Bond, B. E. (1986). *Imposing aid: Emergency assistance to refugees*. Oxford University Press.
- Hengl, T., Miller, M. A. E., Križan, J., et al. (2021). African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. *Scientific Reports*, 11, 6130. <https://doi.org/10.1038/s41598-021-85639-y>
- Jansen, B. (2015). Digging aid: The camp as an option in east and the horn of Africa. *Journal of Refugee Studies*, 29(2), 149–165.
- Kadigo, M. M., & Maystadt, J.-F. (2023). How to cope with a refugee population? evidence from Uganda. *World Development*, 169, 106293. <https://doi.org/10.1016/j.worlddev.2023.106293>
- Kreibaum, M. (2016). Their suffering, our burden? how Congolese refugees affect the Ugandan population. *World Development*, 78, 262–287. <https://doi.org/10.1016/j.worlddev.2015.10.019>
- Kummu, M., Kosonen, M., & Masoumzadeh Sayyar, S. (2025). Downscaled gridded global dataset for gross domestic product (GDP) per capita PPP over 1990–2022. *Scientific Data*, 12(1). <https://doi.org/10.1038/s41597-025-04487-x>
- Lawrence, B. R. (2017). *City of thorns: Nine lives in the world's largest refugee camp*. Picador.
- Loschmann, C., Bilgili, Ö., & Siegel, M. (2019). Considering the benefits of hosting refugees: Evidence of refugee camps influencing local labour market activity and economic welfare in Rwanda. *IZA Journal of Development and Migration*, 9(1), 5. <https://doi.org/10.1186/s40176-018-0138-2>
- MacPherson, C., & Sterck, O. (2021). Empowering refugees through cash and agriculture: A regression discontinuity design. *Journal of Development Economics*, 149, 102614. <https://doi.org/10.1016/j.jdeveco.2020.102614>
- Maystadt, J.-F., & Duranton, G. (2019). The development push of refugees: Evidence from Tanzania. *Journal of Economic Geography*, 19(2), 299–334. <https://doi.org/10.1093/jeg/lby020>
- Maystadt, J.-F., Hirvonen, K., Mabiso, A., & Vandecastelen, J. (2019). Impacts of hosting forced migrants in poor countries. *Annual Review of Resource Economics*, 11(1), 439–459. <https://doi.org/10.1146/annurev-resource-090518-095629>

- Maystadt, J.-F., Mueller, V., Van Den Hoek, J., & Van Weezel, S. (2020). Vegetation changes attributable to refugees in Africa coincide with agricultural deforestation. *Environmental Research Letters*, 15(4), 044008. <https://doi.org/10.1088/1748-9326/ab6d7c>
- Maystadt, J.-F., & Verwimp, P. (2014). Winners and losers among a refugee-hosting population. *Economic Development and Cultural Change*, 62(4), 769–809. <https://doi.org/10.1086/676458>
- Moore, B. (2017). Refugee settlements and sustainable planning. *Forced Migration Review*, (55). <https://ora.ox.ac.uk/objects/uuid:838571e3-ce3d-4ba3-9207-b3fd2fcf3951/files/smg74qn26g>
- Musasizi, J., Arunachalam, D., & Forbes-Mewett, H. (2024). Understanding the dynamics of refugee impact on employment: Evidence from northern Uganda. *International Migration Review*.
- Nelson, A., Weiss, D. J., Van Etten, J., Cattaneo, A., McMenomy, T. S., & Koo, J. (2019). A suite of global accessibility indicators. *Scientific Data*, 6(1), 266. <https://doi.org/10.1038/s41597-019-0265-5>
- NEP-WCMC & IUCN. (2025). *Protected planet: The world database on protected areas (wdpa) and world database on other effective area-based conservation measures [dataset]*. <http://www.protectedplanet.net>
- Ngayimpinda, E. (2005). *Evaluation de la politique sectorielle de rapatriement, de réinsertion et de réhabilitation des sinistrés au Burundi: Rapport définitif*. Observatoire de l'Action Gouvernementale (OAG). <https://repositories.lib.utexas.edu/server/api/core/bitstreams/00da86e0-f94d-47b4-9941-4cad1f302dee/content>
- Owen, M., Kruczkiewicz, A., & Van Den Hoek, J. (2023). Indexing climatic and environmental exposure of refugee camps with a case study in East Africa. *Scientific Reports*, 13(1), 7533. <https://doi.org/10.1038/s41598-023-31140-7>
- Reig-Gracia, F., & Latorre Garcés, B. (2023). *Speibase v.2.9* (Version 2.9). DIGITAL.CSIC. <https://doi.org/10.20350/digitalCSIC/15470>
- Riley, S. J., DeGloria, S. D., & Elliot, R. (1999). A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences*, 5(1–4).
- Salemi, C. (2021). Refugee camps and deforestation in sub-saharan Africa. *Journal of Development Economics*, 152, 102682. <https://doi.org/10.1016/j.jdeveco.2021.102682>
- Smith, M. (2004). Warehousing refugees: A denial of rights, a waste of humanity. *World Refugee Survey*, 38(1).
- Taylor, J. E., Filipski, M. J., Alloush, M., Gupta, A., Rojas Valdes, R. I., & Gonzalez-Estrada, E. (2016). Economic impact of refugees. *Proceedings of the National Academy of Sciences*, 113(27), 7449–7453. <https://doi.org/10.1073/pnas.1604566113>
- The Economist. (2018). *The economist explains: Why most refugees do not live in camps*. The Economist. <https://www.economist.com/the-economist-explains/2018/06/19/why-most-refugees-do-not-live-in-camps>
- UNHCR. (2005). *Environmental guidelines: Main report*. UNHCR. <https://www.gdrc.org/uem/disasters/disenvi/unhcr-eg.pdf>
- UNHCR. (2006). *The state of the world's refugees 2006: Human displacement in the new millennium*. <https://www.unhcr.org/publications/state-worlds-refugees-2006-human-displacement-new-millennium>
- UNHCR. (2007). *UNHCR handbook for emergencies: Third edition*. United Nations High Commissioner for Refugees.

- UNHCR. (2009). *UNHCR policy on refugee protection and solutions in urban areas: September 2009*. UNHCR. <https://academic.oup.com/ijrl/article-lookup/doi/10.1093/ijrl/eep028>
- UNHCR. (2014). *Policy on alternatives to camps*. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/...>
- UNHCR. (2019). *Tanzania country refugee response plan: The integrated response plan for refugees from Burundi and the Democratic Republic of the Congo (January 2019 – December 2020)* (tech. rep.). United Nations High Commissioner for Refugees. <https://reliefweb.int/report/united-republic-tanzania/tanzania-country-refugee-response-plan-integrated-response-plan>
- UNHCR. (2023). *The green companion guide*. UNHCR.
- UNHCR. (2024). *Safe and secure settlements (Emergency Handbook)*. UNHCR.
- UNHCR. (2025). Poc locations [Accessed January 2025]. <https://onemap.un.org/datasets/98a4a069db2a4919bffd2bf6607cb6e>
- UNHCR. (2026). Unhcr population reference areas. <https://data.humdata.org/dataset/unhcr-population-reference-areas>
- UNHCR Global Data Service. (2024a). Refugee population by country, year, and accommodation type [Unpublished dataset shared by the United Nations High Commissioner for Refugees (UNHCR); not publicly available].
- UNHCR Global Data Service. (2024b). Unhcr population of concern (poc) data, 2021–2023 [Unpublished dataset shared by the United Nations High Commissioner for Refugees (UNHCR); not publicly available].
- UNHCR & IOM. (2010). *Collective centre guidelines (Guidelines)* (Prepared by the CCCM Cluster). United Nations High Commissioner for Refugees (UNHCR) and International Organization for Migration (IOM). <https://emergency.unhcr.org/sites/default/files/UNHCR-IOM%2C%20Collective%20Centres%20Guidelines%2C%202010.pdf>
- Verdirame, G., & Pobjoy, J. (2013). The end of refugee camps? In *Ashgate research companion to migration law, theory and policy*. Routledge.
- Verme, P., & Schuettler, K. (2021). The impact of forced displacement on host communities: A review of the empirical literature in economics. *Journal of Development Economics*, 150, 102606. <https://doi.org/10.1016/j.jdeveco.2020.102606>
- Vos, A., & Dempster, H. (2021). *Most refugees live in cities, not camps. our response needs to shift* [Blog post]. Center for Global Development. <https://www.cgdev.org/blog/most-refugees-live-cities-not-camps-our-response-needs-shift>
- Wan, Z., Hook, S., & Hulley, G. (2021). *Modis/terra land surface temperature/emissivity 8-day l3 global 1km sin grid v061*. NASA Land Processes Distributed Active Archive Center. <https://doi.org/10.5067/MODIS/MOD11A2.061>
- WorldPop. (2025). Worldpop project.
- Zhou, Y.-Y., Grossman, G., & Ge, S. (2023). Inclusive refugee-hosting can improve local development and prevent public backlash. *World Development*, 166, 106203. <https://doi.org/10.1016/j.worlddev.2023.106203>
- Zhou, Y.-Y., & Shaver, A. (2021). Reexamining the effect of refugees on civil conflict: A global subnational analysis. *American Political Science Review*, 1–22. <https://doi.org/10.1017/S0003055421000502>

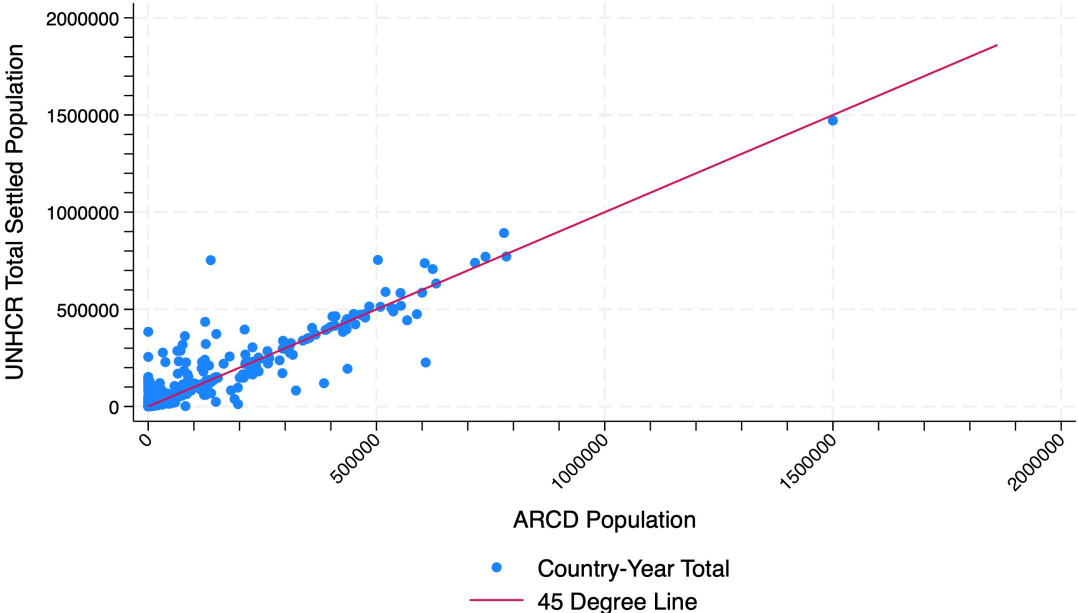
8 Figures

Figure 1: Count of Refugees (millions) Living in Camps or Camp-like Situations by Year and Region



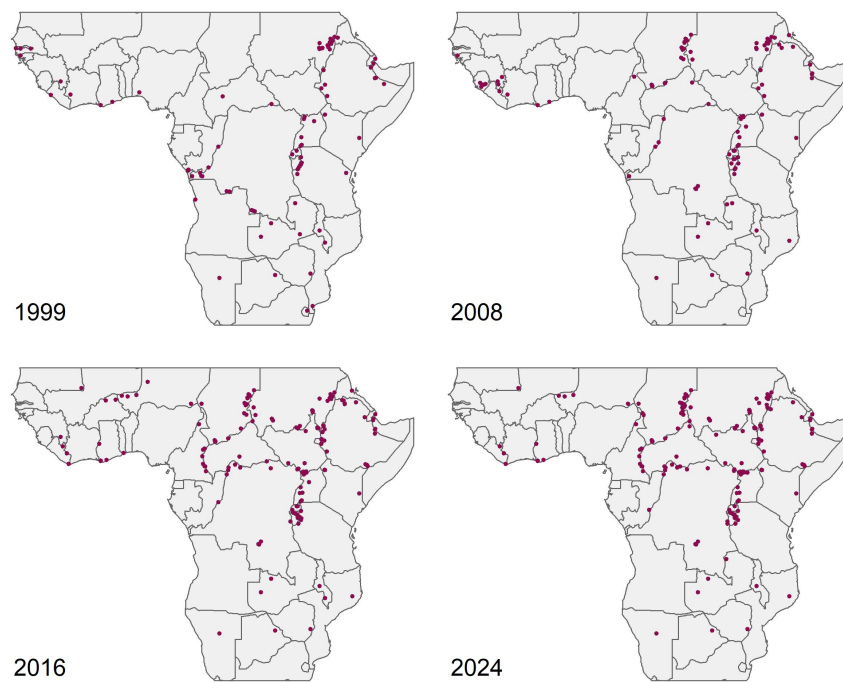
Notes: Proprietary data from UNHCR Global Data Service (2024a). “Camp-like situations” includes refugees living in formal or informal camps, transit centers, and communal centers. Europe and Latin America excluded due to low counts of encamped refugees over time.

Figure 2: Country-year scatterplot comparing population settled in formal refugee camps according to UNHCR data and according to aggregation of ARCD population data



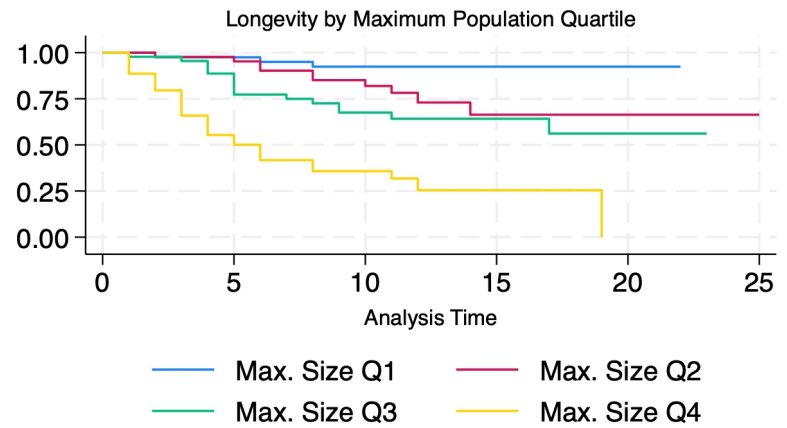
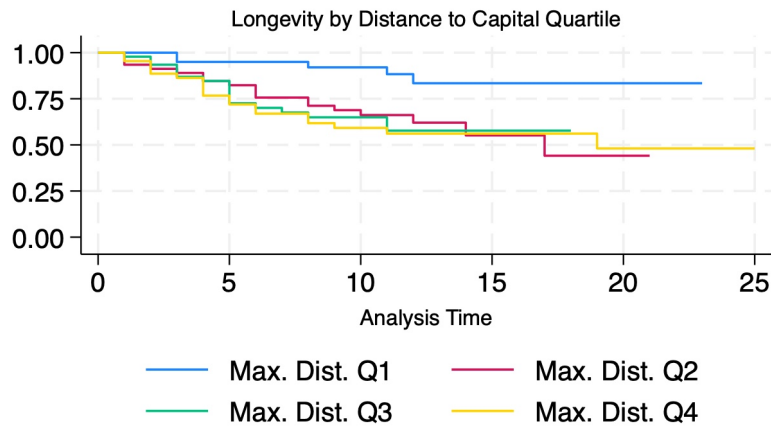
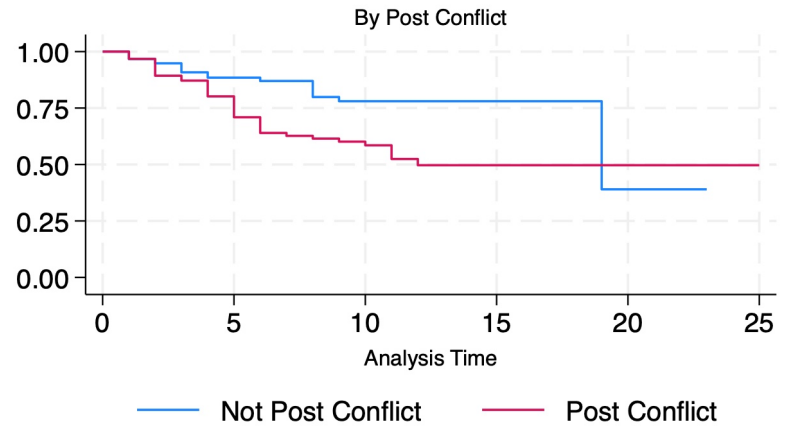
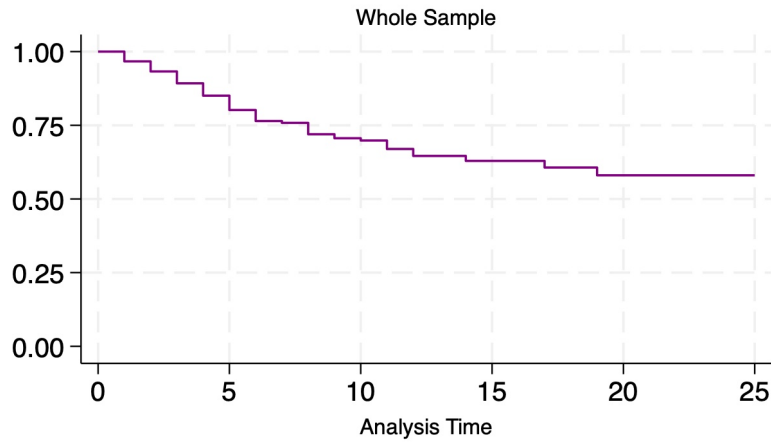
Notes: UNHCR Total Settled Population is a count of refugees considered to be in formal encampments according to UNHCR Global Data Service. This excludes those listed as being in informal settlements, transit centers, collective centers, or non-settled/living in host communities. ARCD data restricted to camp-years with available population data.

Figure 3: Spatial distribution of ARCD geolocations by year



Notes: Maps based on ARCD camp locations, with data restricted to camps with identified geolocations and operational years. Sub-camps of primary camp are removed.

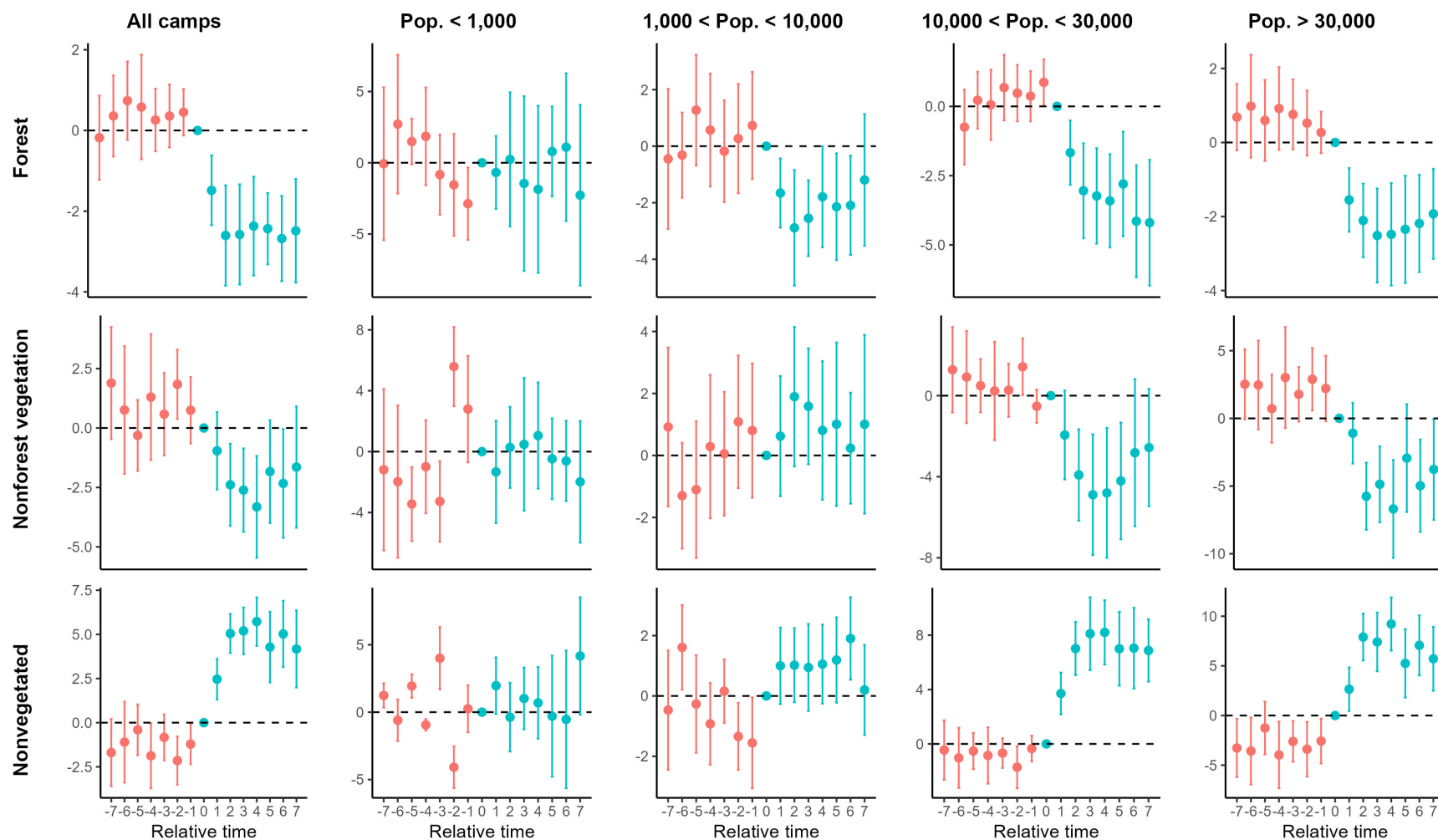
Figure 4: Kaplan-Meier Estimates of Survival Functions by camp attributes



36

Notes: Based on ARCD observations with opening and closing dates. The data used are all camps that came into existence in 1999 or after omitting subcamps. Data used in the bottom right is further limited to camps for which we have at least one year of population information. “Post conflict” means the conflict event generating population flows to the camp is resolved. Distance to capital is based on Euclidean distance measure.

Figure 5: Event study estimates of the impact of camp opening on forest and vegetation change in camp interiors by camp maximum population size



37

Notes: The event study corresponds to the cumulative average treatment effect results reported in Figure 5. Results are estimated using the method developed by de Chaisemartin and D’Haultfœuille (2024). See Section B in Appendix for technical details. The ARCD camp sample is restricted to observations with (1) latitude and longitude listings, (2) opening date listings, and (3) main camps (excluding sub-camps) ($N = 323$). Control group drawn from areas 30-50 km from the nearest camp outside protected areas. Random sampling of control points is stratified by country representation in the ARCD data ($N = 2,354$). Vegetation and forest data are from MODIS Terra Continuous Fields. Zonal statistics are derived using a 1 km buffer around each treatment and control geopot. Population (“Pop.”) sizes reflect the maximum population size of the camp based on available ARCD data.

9 Tables

Table 1: Simplified ARCD Extract for Kakuma Camp (from “ARCD_long.dta”)

Camp_ID	Camp_Name	Year	Country	Latitude	Longitude	Opening_Year	Closing_Year	Reason_For_Closing	First_Conflict_Name	CampPop
2190020000	Kakuma	1999	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	86600
2190020000	Kakuma	2000	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	69749
2190020000	Kakuma	2001	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	84600
2190020000	Kakuma	2002	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	82200
2190020000	Kakuma	2003	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	89378
2190020000	Kakuma	2004	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	86217
2190020000	Kakuma	2005	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	94677
2190020000	Kakuma	2006	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	69749
2190020000	Kakuma	2007	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	62497
2190020000	Kakuma	2008	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	53068
2190020000	Kakuma	2009	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	64791
2190020000	Kakuma	2010	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	69822
2190020000	Kakuma	2011	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	85862
2190020000	Kakuma	2012	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	107205
2190020000	Kakuma	2013	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	128540
2190020000	Kakuma	2014	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	178079
2190020000	Kakuma	2015	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	184550
2190020000	Kakuma	2016	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	154947
2190020000	Kakuma	2017	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	146975
2190020000	Kakuma	2018	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	149076
2190020000	Kakuma	2019	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	193684
2190020000	Kakuma	2020	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	158365
2190020000	Kakuma	2021	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	173742
2190020000	Kakuma	2022	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	197475
2190020000	Kakuma	2023	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	203876
2190020000	Kakuma	2024	Kenya	3.7307854	34.841857	1992	.	Still Open	Second Sudanese Civil War	.

Notes: This table presents a simplified extract from the ARCD dataset for Kakuma refugee camp in Kenya, covering annual observations from 1999 to 2024. The variable CampPop reports the camp population for each year. A period (“.”) indicates missing data in the original dataset, implying that the value was not reported for the CampPop variable and indicating that the camp remains open when in the Closing_Year column. The camp was opened in 1992 and remained operational throughout the observation period, with no recorded closing year. The first associated conflict is identified as the Second Sudanese Civil War. Coordinates indicate the geographic location of the camp and remain constant across years. Most variable columns excluded - see [Table A2](#) for full codebook of variables.

Table 2: Descriptive comparisons of refugee camp characteristics relative to border and country comparison samples

Panel A: Time-invariant camp characteristics											
	ARCD			Border Sample (BS)			ARCD vs.	Country Sample (CS)			ARCD vs.
	Mean	SD	N	Mean	SD	N	BS (Δ)	Mean	SD	N	CS (Δ)
<i>Euclidean distance</i>											
Internal border	53.0	53.7	323	61.9	53.8	2937	-8.9***	118.1	107.1	2354	-65.2***
Protected area	58.5	66.4	323	48.9	65.1	2937	9.6**	76.7	116.8	2354	-18.2***
Provincial capital	100.4	84.3	323	104.1	72.3	2937	-3.7	127.0	100.7	2354	-26.6***
National capital	450.2	330.6	323	417.9	333.5	2937	32.3*	417.3	329.6	2354	33.0 *
<i>Terrain ruggedness</i>											
TRI	4.0	4.0	323	7.3	9.4	2935	-3.4***	7.0	9.1	2351	-3.1***
<i>Bioclimatic area</i>											
Forests	19.2	39.4	323	20.3	40.2	2934	-1.1	19.2	39.4	2352	-0.0
Wetlands	1.6	12.4	323	1.2	11.0	2934	0.3	1.6	11.9	2352	0.1
Grasslands	76.2	42.7	323	75.4	43.1	2934	0.7	69.6	46.0	2352	6.5**
Deserts	3.1	17.4	323	3.1	17.3	2934	-0.0	9.7	29.6	2352	-6.6***
Panel B: 1 km buffer characteristics at $t = -2$											
	ARCD			Border Sample (BS)			ARCD vs.	Country Sample (CS)			ARCD vs.
	Mean	SD	N	Mean	SD	N	BS (Δ)	Mean	SD	N	CS (Δ)
Pct. forested	10.1	11.1	176	13.4	15.7	16,002	-3.3***	14.1	20.1	12,949	-3.9***
Pct. non-forest veg.	62.3	23.8	176	59.2	25.1	16,002	3.1*	45.1	32.3	12,949	17.2***
Pct. non-vegetated	27.9	28.2	176	27.7	29.6	16,002	0.2	41.1	39.4	12,949	-13.2***
Mean radiance	0.005	0.038	64	0.096	1.596	7,557	-0.091***	0.014	0.286	6,256	-0.009

Notes: This table compares the ARCD refugee camp sample to the **Border Sample (BS)** (comparison points located 30-50 km from ARCD camp locations) and the **Country Sample (CS)** (comparison points drawn from across the host country). Both BS and CS exclude locations overlapping protected areas or bodies of water. Δ columns report ARCD minus comparison group means; positive values indicate higher values at camp locations. *, **, and *** indicate significance at the 90, 95, and 99 percent confidence levels based on two-sample t-tests. The ARCD sample is restricted to main camps with valid coordinates; sub-camps are excluded. **Panel A** reports time-invariant characteristics. Euclidean distance variables report Euclidean straight-line distances in kilometers to the nearest feature of the indicated type, derived from GeoNames gazetteer data. TRI is the Terrain Ruggedness Index (Riley et al., 1999); higher values indicate more rugged terrain. Bioclimatic indicators equal one if the geopoint falls within the corresponding WWF biome: “Forests” includes tropical and subtropical moist and dry broadleaf forests; “Wetlands” includes flooded grasslands, savannas, and mangroves; “Grasslands” includes montane, tropical, and subtropical grasslands, savannas, and shrublands; “Deserts” includes deserts and xeric shrublands. **Panel B** reports characteristics measured within a 1 km buffer at $t = -2$, two years prior to camp opening, and additionally requires non-missing opening years. Pct. forested, non-forest veg., and non-vegetated are derived from MODIS Terra Continuous Fields data and sum to approximately 100%. Mean radiance measures nighttime light intensity from VIIRS luminosity data. West Africa is excluded due to insufficient post-2014 camp openings for reliable coverage. In the time-varying panel, comparison sample observations are assigned to match country-specific calendar years in the ARCD sample at $t = -2$.

Table 3: Cumulative average treatment effects of camp opening on land cover outcomes

Panel A: Effects by camp maximum population size					
	All camps	Pop. < 1k	Pop. ∈ [1k, 10k)	Pop. ∈ [10k, 30k)	Pop. ≥ 30k
	(1)	(2)	(3)	(4)	(5)
Forest (%)	-2.501*** (0.556)	-0.739 (2.795)	-2.988** (0.909)	-3.259*** (0.875)	-2.177*** (0.573)
Vegetation (%)	-2.272** (0.974)	-0.437 (1.363)	1.153 (1.115)	-3.676*** (1.303)	-4.326*** (1.449)
Nonvegetated (%)	4.800*** (0.713)	1.157 (1.831)	1.169** (0.585)	6.982*** (1.009)	6.501*** (1.259)
N	74,507	55,717	71,343	71,472	68,449

Panel B: Effects by region and bioclimatic area					
	Grasslands biome	Wet/dryleaf forest biome	West Africa	Central Africa	East Africa
	(1)	(2)	(3)	(4)	(5)
Forest (%)	-1.975*** (0.529)	-5.465*** (0.864)	-4.854*** (0.470)	-3.996*** (0.769)	-2.501*** (0.556)
Vegetation (%)	-3.069*** (1.112)	0.435 (1.110)	-0.621 (1.157)	-1.989 (1.405)	-2.272** (0.974)
Nonvegetated (%)	5.068*** (0.777)	5.061*** (0.693)	5.545*** (1.120)	6.041*** (0.670)	4.800*** (0.713)
N	55,855	14,917	6,879	16,635	74,507

Notes: *, **, and *** indicate significance at the 90, 95, and 99 percent confidence levels. Coefficients report the cumulative average treatment effect (CATE) of camp opening on each land cover outcome, estimated using the de Chaisemartin and D’Haultfœuille (2024) difference-in-differences estimator for staggered and non-absorbing treatments (*did_multiplegt_dyn*), with 7 placebo and 7 post-treatment years. The CATE aggregates dynamic treatment effects over the full post-treatment window; standard errors clustered at country level are in parentheses. The ARCD sample is restricted to main camps with valid coordinates and opening years ($N = 323$); sub-camps are excluded. This control group consists of points located 30-50 km from camp geolocations, outside protected areas, stratified by country ($N = 2,354$). Land cover outcomes (percent forest cover, percent nonforest vegetation, and percent nonvegetated cover) are derived from MODIS Terra Continuous Fields (250m, 2000-2024) using a 250 meter buffer around each geopoint. Reported N reflects panel length after applying the 7-year pre- and post-treatment window; available time spans vary across camps. **Panel A** stratifies by maximum recorded camp population over operational years. **Panel B**, Columns (1)-(2) stratify by biome: Column (1) is restricted to the Tropical & Subtropical Grasslands, Savannas & Shrublands biome; Column (2) to tropical moist and dry broadleaf forests. Columns (3)-(5) report estimates by region following UN classification, with Sudan assigned to East Africa; Southern Africa is omitted due to insufficient sample size.

**A New Expanded Dataset to Study Refugee Camps in Sub-Saharan
Africa 1999-2024**

Supplemental Online Appendix

A Additional Information on ARCD data production

We provide more information on the production of the ARCD dataframe below.

A.1 Statistical Yearbooks

We use the following Statistical Yearbooks for data production.

References

- [1] UNHCR. (2000). *Refugees and others of concern to UNHCR: 1999 statistical overview* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/media/refugees-and-others-concern-unhcr-1999-statistical-overview>
- [2] UNHCR. (2002a). *2001 UNHCR population statistics (provisional)* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/media/2001-unhcr-population-statistics-provisional-table-1-population-concern-unhcr-location-sex>
- [3] UNHCR. (2002b). *Refugees and others of concern to UNHCR: 2000 statistical overview* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/media/refugees-and-others-concern-unhcr-2000-statistical-overview>
- [4] UNHCR. (2002c). *Statistical yearbook 2001: Refugees, asylum-seekers, and other persons of concern—trends in displacement, protection and solutions* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/publications/unhcr-statistical-yearbook-2001>
- [5] UNHCR. (2003). *2002 UNHCR population statistics (provisional)* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/media/2002-unhcr-population-statistics-provisional>
- [6] UNHCR. (2005). *Statistical yearbook 2003: Trends in displacement, protection and solutions* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/publications/unhcr-statistical-yearbook-2003>
- [7] UNHCR. (2006). *Statistical yearbook 2004: Trends in displacement, protection and solutions* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/publications/unhcr-statistical-yearbook-2004>
- [8] UNHCR. (2007a). *Statistical yearbook 2005: Trends in displacement, protection and solutions* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/publications/unhcr-statistical-yearbook-2005>
- [9] UNHCR. (2007b). *Statistical yearbook 2006: Trends in displacement, protection and solutions* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/publications/unhcr-statistical-yearbook-2006>

- [10] UNHCR. (2008). *Statistical yearbook 2007: Trends in displacement, protection and solutions* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/publications/unhcr-statistical-yearbook-2007>
- [11] UNHCR. (2009a). *Statistical yearbook 2008: Trends in displacement, protection and solutions* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/publications/unhcr-statistical-yearbook-2007>
- [12] UNHCR. (2010). *Statistical yearbook 2009: Trends in displacement, protection and solutions* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/publications/unhcr-statistical-yearbook-2009>
- [13] UNHCR. (2011). *UNHCR statistical yearbook 2010, 10th edition* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/media/unhcr-statistical-yearbook-2010-statistical-annex>
- [14] UNHCR. (2013a). *UNHCR statistical yearbook 2011, 11th edition* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/publications/unhcr-statistical-yearbook-2011-11th-edition>
- [15] UNHCR. (2013b). *UNHCR statistical yearbook 2013, 13th edition* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/publications/unhcr-statistical-yearbook-2013-13th-edition>
- [16] UNHCR. (2015). *UNHCR statistical yearbook 2014, 14th edition* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/publications/unhcr-statistical-yearbook-2014-14th-edition>
- [17] UNHCR. (2017a). *UNHCR statistical yearbook 2015* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/sites/default/files/legacy-pdf/59b294387.pdf>
- [18] UNHCR. (2017b). *UNHCR statistical yearbook 2016* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). <https://www.unhcr.org/sites/default/files/legacy-pdf/59b294387.pdf>

A.2 Zonal statistics from secondary spatial data

We include zonal statistics derived from secondary geospatial sources in the ARCD dataset. These data are intended to capture time-invariant and some time-varying characteristics of the camp.

Table A1: Spatial variables and underlying data sources

Variable	Unit	Data source	Years	Resolution	Citation	Buffer?
Cumulative GDP	USD PPP	Gridded GDP	2000, 2005, 2010, 2015, 2020	30 arc-sec.	Kummu et al. (2025)	No
Nighttime lights	Mean radiance	VIIRS V2.1 masked mean layer	2012-2021	15 arc-sec.	Elvidge et al. (2021)	Yes
Cumulative population	Persons	WorldPop	2000-2020	100 m	WorldPop (2025)	Yes
Forest cover	Percent	MODIS Terra Vegetation Continuous Fields	2000-2024	250 m	DiMiceli et al. (2022)	Yes
Nonforest vegetative cover	Percent	MODIS Terra Vegetation Continuous Fields	2000-2024	250 m	DiMiceli et al. (2022)	Yes
Nonforest nonvegetated cover	Percent	MODIS Terra Vegetation Continuous Fields	2000-2024	250 m	DiMiceli et al. (2022)	Yes
Forest loss	Square kilometer	Global Forest Change	2001-2024	30 m	Hansen et al. (2013)	Yes
Annual NDVI summary (from 16-day composites)	Mean NDVI	MOD13Q1.061 Terra Vegetation Indices 16-Day Global	2000-2024	250 m	Didan (2021)	Yes
Annual total precipitation	millimeters	TerraClimate	2000-2024	4,638.3 m	Abatzoglou et al. (2018)	No
Months in drought (PDSI < 0)	Number of months	TerraClimate	2000-2024	4,638.3 m	Abatzoglou et al. (2018)	No
Months in drought according to Standardized Precipitation-Evapotranspiration Index	Number of months	SPEIBase	2000-2023	55,660 m	Reig-Gracia and Latorre Garcés (2023)	No

Notes: Table describes all spatial variables included in the ARCD dataset, along with their underlying sources, years available, and spatial resolution. Column “Buffer” indicates whether the variable is summarized using a 250 meter spatial buffer around the relevant camp location. If “No”, then variable produced based on intersection or proximity to point location. When column “Years” is missing, the data are considered time-invariant. Column “Resolution” is missing if the secondary data source is non-raster (point, polygon, linestring, etc.).

Table A1: Spatial variables and underlying data sources (continued)

Variable	Unit	Data source	Years	Resolution	Citation	Buffer?
Annual mean day land surface temperature	Kelvin	MOD11A2.061 Terra Land Surface Temperature and Emissivity 8-Day Global	2000-2024	1 km	Wan et al. (2021)	Yes
Share of area with i soil constraints, $i \in \{0, 1, 2, 3, 4, \geq 5\}$	Percent	iSDAsoil Fertility Capability Classification	-	30 m	Hengl et al. (2021)	Yes
Minimum Euclidean distance to protected areas	Kilometers	Protected Planet	-	-	NEP-WCMC and IUCN (2025)	No
Minimum Euclidean distance to provincial capital	Kilometers	GeoNames Gazetteer	-	-	GeoNames (2026)	No
Minimum Euclidean distance to national capital	Kilometers	GeoNames Gazetteer	-	-	GeoNames (2026)	No
Minimum Euclidean distance to nearest city	Kilometers	GeoNames Gazetteer(population: > 1,000)	2017	-	Africa GeoPortal (2017)	No
Minimum Euclidean distance to nearest internal international border	Kilometers	Database of Global Administrative Areas	-	-	-	No
Travel time to nearest port, any size	Minutes	Travel time to cities and ports in the year 2015	2015	1 km	Nelson et al. (2019)	No
Biome	-	RESOLVE Ecoregions	-	-	Dinerstein et al. (2017)	No
Ecoregion	-	RESOLVE Ecoregions	-	-	Dinerstein et al. (2017)	No

Notes: Table describes all spatial variables included in the ARCD dataset, along with their underlying sources, years available, and spatial resolution. Column “Buffer” indicates whether the variable is summarized using a 250 meter spatial buffer around the relevant camp location. If “No”, then variable produced based on intersection or proximity to point location. When column “Years” is missing, the data are considered time-invariant. Column “Resolution” is missing if the secondary data source is non-raster (point, polygon, linestring, etc.).

Table A1: Spatial variables and underlying data sources (continued)

Variable	Unit	Data source	Years	Resolution	Citation	Buffer?
Terrain ruggedness index score	TRI	Copernicus Global Digital Elevation data	-	30 m	Riley et al. (1999)	No
First, second, and third administrative division	string	Global Administrative Areas V4.1	-	-	GADM (2024)	No

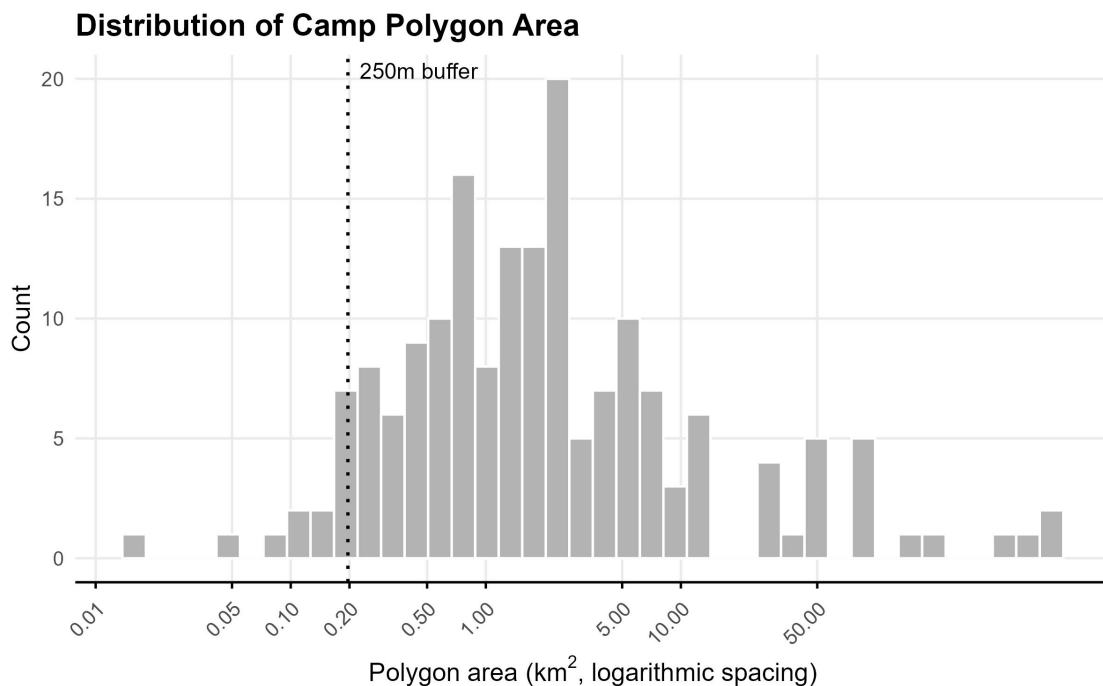
Notes: Table describes all spatial variables included in the ARCD dataset, along with their underlying sources, years available, and spatial resolution. Column “Buffer” indicates whether the variable is summarized using a 250 meter spatial buffer around the relevant camp location. If “No”, then variable produced based on intersection or proximity to point location. When column “Years” is missing, the data are considered time-invariant. Column “Resolution” is missing if the secondary data source is non-raster (point, polygon, linestring, etc.).

A.3 Buffer size selection

We use the UNHCR Population Reference Areas REST API to obtain publicly available data on camp boundaries (stored as polygons). The resulting sample covers 149 of the 465 ARCD camp entries. Whenever an ARCD camp location is present in the Population Reference Areas data, we include the camp area size (in km^2) in the ARCD data (`camp_area_km2`).

With this information, we examine the distribution of camp sizes and determine a buffer size for deriving zonal statistics that characterize the camp interior. Our goal is to select a buffer size that is smaller than the majority of camps but not so small that it approaches the pixel size of the highest-resolution raster inputs. We note that 94% of the sample is larger than a 250 meter buffer (with area 0.196 km^2). Wary of making the buffer any smaller, we use this buffer size, which our evidence suggests will fall in the interior for the majority of our camps.

Figure A1: Distribution of ARCD camp sizes



Notes: Data from 149 ARCD entries that matched with UNHCR Population Reference Areas polygon spatial data. Dotted vertical reference line at 0.196 km^2 , the area size of a 250 meter buffer. Distribution is log scaled, but x axis labels report corresponding numeric values.

A.4 ARCD Codebook

Table A2: ARCD Variable Descriptions

Variable	Type	Description	TV	Source
<i>Panel A: Identifiers and Operational Characteristics</i>				
Camp_ID	Char	Unique camp identifier	No	
Camp_Name	Char	Official camp name	No	
Camp_Alt_Name	Char	Alternative camp name	No	
year	Num	Year of observation	No	
Country	Char	Country name	No	
ISO	Char	ISO 3-letter country code	No	
Country_ID	Num	Numeric country identifier	No	
Camp_No	Num	Country-level camp index number	No	
area_sqkm	Num	Country land area (km ²)	No	Natural Earth ADM0
Region_ID	Cat	Administrative region (1=West; 2=East; 3=Central; 4=South; 5=North)	No	UNSTATS Geoscheme
Reopening_Indicator	Cat	First or second opening (0=First; 1=Second)	No	
Subcamp_ID	Num	Subcamp identifier within main camp	No	
Sub_Sub_ID	Num	Sub-subcamp identifier	No	

Continued on next page

Table A2 - continued

Variable	Type	Description	TV	Source
Potentially_Non_Camp	Cat	Flag for locations potentially not functioning as camps (0=No; 1=Yes)	No	
Settle_model	Cat	Settlement included due to camp-like characteristics (0=No; 1=Yes)	No	
Latitude	Num	Latitude of camp centroid (decimal degrees)	No	UNHCR, validated via Google Earth
Longitude	Num	Longitude of camp centroid (decimal degrees)	No	UNHCR, validated via Google Earth
Coords_validated	Cat	Geocoordinates validated via satellite imagery (0=No; 1=Yes)	No	Google Earth
Source_Type_Camp_Validation	Cat	Source type indicator for validation of location being a camp (1=UNHCR POC Population Map; 2=UNHCR Operational Data Portal; 3=UN Wash; 4=UNHCR Statistical Yearbook(s); 5=Alternative UN Sources; 6=Government Sources; 7=NGO Sources; 8=Media coverage; 9=Academic Literature; 10=Other)	No	ARCD Metadata
Elevation	Num	Elevation (m.a.s.l)	No	AWS Terrain Tiles
<i>Panel B: Operation Dates and Camp History</i>				
Opening_Year	Num	Year camp first opened	No	

Continued on next page

Table A2 - continued

Variable	Type	Description	TV	Source
Closing_Year	Num	Year camp closed	No	
Source_Type_Opening	Cat	Source type indicator for camp opening year (1=UNHCR POC Population Map; 2=UNHCR Operational Data Portal; 3=UN Wash; 4=UNHCR Statistical Yearbook(s); 5=Alternative UN Sources; 6=Government Sources; 7=NGO Sources; 8=Media coverage; 9=Academic Literature; 10=Other)	No	ARCD Metadata
Source_Type_Closing	Cat	Source type indicator for camp closing year (1=UNHCR POC Population Map; 2=UNHCR Operational Data Portal; 3=UN Wash; 4=UNHCR Statistical Yearbook(s); 5=Alternative UN Sources; 6=Government Sources; 7=NGO Sources; 8=Media coverage; 9=Academic Literature; 10=Other)	No	ARCD Metadata
First_Mention_in_Yearbook	Num	First year in UNHCR yearbook	No	UNHCR Statistical Yearbooks
Last_Mention_in_Yearbook	Num	Last year in UNHCR yearbook	No	UNHCR Statistical Yearbooks
Last_Year_with_pop_entry	Num	Last year with CampPop entry, when population numbers are recorded after official camp closing date	No	Calculated from CampPop and Closing_Year

Continued on next page

Table A2 - continued

Variable	Type	Description	TV	Source
Length_of_Opening	Num	Total years camp was open	No	
Reason_For_Closing	Cat	Reason for closure (0=Still Open; 1=Security; 2=Consolidated; 3=Gov. Shut-down; 4=Repatriation; 5=Integration; 6=Flooding; 7=Budgetary)	No	
Open	Cat	Camp open at end of study period (0=No; 1=Yes)	No	
status	Cat	Annual operational status (0=Closed; 1=Open)	Yes	
<i>Panel C: Conflict</i>				
First_Conflict_Name	Char	Name of first associated conflict	No	
Second_Conflict_Name	Char	Name of second associated conflict	No	
First_Conflict_Start_Date	Date	Start date of first conflict	No	
First_Conflict_End_Date	Date	End date of first conflict	No	
Second_Conflict_Start_Date	Date	Start date of second conflict	No	
Second_Conflict_End_Date	Date	End date of second conflict	No	
Year_from_Conflict_End_to_Close	Num	Years between conflict end and camp closure	No	
<i>Panel D: Population</i>				

Continued on next page

Table A2 - continued

Variable	Type	Description	TV	Source
CampPop	Num	Annual camp population	Yes	UNHCR Statistical Yearbooks and additional sources
Month_of_Observation	Char	Month of observation for CampPop figure	Yes	ARCD Metadata
Source_Type_Population	Cat	Source type indicator for CampPop data (1=UNHCR POC Population Map; 2=UNHCR Operational Data Portal; 3=UN Wash; 4=UNHCR Statistical Yearbook(s); 5=Alternative UN Sources; 6=Government Sources; 7=NGO Sources; 8=Media coverage; 9=Academic Literature; 10=Other)	Yes	ARCD Metadata
Max_Population	Num	Maximum observed population	No	
Maximum_Camp_Size	Cat	Size category (0=No data; 1=<1,000; 2=<10,000; 3=<30,000; 4= \geq 30,000)	No	
PopContradiction	Cat	Flag for contradictory population information (0=No; 1=Yes)	No	
camp_area_km2	Num	Camp area (km ²)	No	UNHCR Polygons
<i>Panel E: Administrative Divisions</i>				
ADM1	Char	Subnational administrative boundaries level 1	No	GADM

Continued on next page

Table A2 - continued

Variable	Type	Description	TV	Source
ADM2	Char	Subnational administrative boundaries level 2	No	GADM
ADM3	Char	Subnational administrative boundaries level 3	No	GADM
<i>Panel F: Proximity Measures</i>				
borderDistKM	Num	Distance to nearest international border (km)	No	USDOS LSIB Simple 2017
protectDistKM	Num	Distance to nearest protected area (km)	No	UNEP-WCMC & IUCN, WDPA (2025)
minDistProv	Num	Distance to nearest provincial capital (km)	No	GeoNames cities1000
minDistNat	Num	Distance to national capital (km)	No	GeoNames cities1000
DistClosestCity	Num	Distance to nearest city of 1,000+ (km)	No	GeoNames cities1000
TravelTimePorts	Num	Travel time to nearest port (minutes)	No	Nelson et al. (2019)
<i>Panel G: Socioeconomic Features</i>				
pop	Num	Population within buffer	Yes	WorldPop
gdp_sum	Num	Total GDP of 1km grid cell (PPP)	Yes	Kummu et al. (2025)

Continued on next page

Table A2 - continued

Variable	Type	Description	TV	Source
income	Cat	Host country income classification (1=Low; 2=Lower-middle; 3=Upper-middle)	Yes	World Bank
viirs_mean	Num	Nighttime light radiance (nW/cm ² /sr)	Yes	Elvidge et al. (2017)
<i>Panel H: Bioecological Features</i>				
BIOME	Cat	Biome classification	No	RESOLVE Ecoregions 2017
ECO	Cat	Ecoregion classification	No	RESOLVE Ecoregions 2017
TRI	Num	Terrain Ruggedness Index	No	COPERNICUS DEM
pctFor	Num	Forest cover within buffer (%)	Yes	MOD44B.061
pctNonTreeVeg	Num	Non-tree vegetation cover within buffer (%)	Yes	MOD44B.061
pctNonVeg	Num	Non-vegetated land within buffer (%)	Yes	MOD44B.061
Forest_Loss	Num	Annual forest loss within buffer (km ²)	Yes	Hansen GFC v1.12
NDVI	Num	Normalized Difference Vegetation Index	Yes	MOD13Q1.061
Flood_Exp	Num	Flood exposure within buffer	No	Global Flood Database

Continued on next page

Table A2 - continued

Variable	Type	Description	TV	Source
soil_con_0_pct	Num	Share of buffer with no soil constraints (%)	No	iSDAsoil FCC
soil_con_1_pct	Num	Share of buffer with one soil constraint (%)	No	iSDAsoil FCC
soil_con_2_pct	Num	Share of buffer with two soil constraints (%)	No	iSDAsoil FCC
soil_con_3_pct	Num	Share of buffer with three soil constraints (%)	No	iSDAsoil FCC
soil_con_4_pct	Num	Share of buffer with four soil constraints (%)	No	iSDAsoil FCC
soil_con_5plus_pct	Num	Share of buffer with five or more soil constraints (%)	No	iSDAsoil FCC
<i>Panel I: Climatic Features</i>				
LST	Num	Annual mean land surface temperature (Kelvin)	Yes	MOD11A2.061
Precip	Num	Annual total precipitation (mm)	Yes	TerraClimate
nPDSI	Num	Months of drought (Palmer Drought Severity Index)	Yes	TerraClimate
nSPEI	Num	Months of drought (SPEI)	Yes	SPEIbase v2.10

Notes: This table reports on all variables included in the ARCD dataset. TV = Time-varying. Type abbreviations: Num = Numeric; Cat = Categorical; Char = Character. See `Metadata.txt` dataset for detailed source information on opening/closing years, conflict variables, and population sources. Secondary data used to derive spatial variables (Panels E-H) listed in [Table A1](#).

B Econometric framework for causal analysis

Below we provide greater detail into the empirical specification we use to estimate the effect of camp opening on forest and vegetation cover in camp interiors.

Baseline Specification

Consider the standard two-way fixed effects (TWFE) estimator:

$$y_{it} = \beta_0 + \beta_1 Open_{it} + \gamma_i + \lambda_t + \varepsilon_{it} \quad (1)$$

where y_{it} is the percent forest cover, nonforest vegetation cover, or nonforest non-vegetated cover within a 250 meter buffer around geopoint i in year t . The variable of interest, $Open_{it}$, is a binary indicator equal to one if the camp at location i is operational in year t and zero otherwise. Unit fixed effects γ_i absorb all time-invariant location characteristics, including terrain ruggedness, distance to cities, and bioclimatic zone, while year fixed effects λ_t capture subcontinental time trends common to all units. Standard errors are clustered at the host country level.

B.1 de Chaisemartin-D’Haultfœuille (DCDF) Estimator

Estimating [Equation \(1\)](#) via OLS can be problematic in staggered adoption settings because the TWFE estimator decomposes $\hat{\beta}_{DD}$ into a weighted average of all possible 2×2 DID comparisons between treatment groups defined based on when they are first treated ([Goodman-Bacon, 2021](#)). Some of these comparisons use already-treated units as controls for later-treated units. Because the already-treated unit’s outcome already reflects its own treatment effect, that effect is differenced out of the later-treated unit’s estimate, biasing it toward zero regardless of whether treatment effects are homogeneous across units and time periods.

While this bias is unlikely to be consequential in our setting given the large never-treated comparison sample, we nonetheless adopt the estimator developed by de Chaisemartin and D’Haultfœuille ([2024](#)) to ensure robustness to any residual contamination from staggered treatment

timing. The estimator targets the following period-specific effects. For each $\ell \geq 1$, the parameter of interest is

$$\delta_\ell = \frac{1}{N_\ell} \sum_{g: F_g - 1 + \ell \leq T_g} S_g \delta_{g,\ell}, \quad (2)$$

where F_g denotes the first year of camp operation for unit g , N_ℓ is the number of units for which $\delta_{g,\ell}$ can be estimated, $S_g = 1$ for units whose treatment increases at F_g and -1 for units whose treatment decreases, and $\delta_{g,\ell}$ is the expected difference between unit g 's actual outcome at $F_g - 1 + \ell$ and the counterfactual outcome it would have obtained had its treatment remained at its period-one value. The coefficients δ_ℓ for $\ell < 0$ are placebo estimates of pre-treatment trend differences between camp and control locations; under the parallel trends assumption, these should be jointly indistinguishable from zero. The coefficients δ_ℓ for $\ell \geq 1$ capture how vegetation cover evolves in the years following camp opening.

Critically, the de Chaisemartin and D'Haultfœuille (2024) estimator does not recover each δ_ℓ via OLS. Instead, it constructs each period-specific effect using only clean comparisons: units switching treatment status at F_g against units with the same period-one treatment level whose treatment has not yet changed by $F_g - 1 + \ell$. This avoids the bias from heterogeneous treatment effects across cohorts that afflicts standard TWFE (Goodman-Bacon, 2021). The estimator also does not require treatment to be absorbing, which naturally accommodates camps that open, close, and reopen within the panel.

Cumulative Average Treatment Effect and Event Study Design

From the dynamic estimates $\hat{\delta}_\ell$, we recover the cumulative average treatment effect (CATE), defined as the weighted sum

$$\widehat{\text{CATE}} = \sum_{\ell=1}^L \hat{w}_\ell \hat{\delta}_\ell, \quad (3)$$

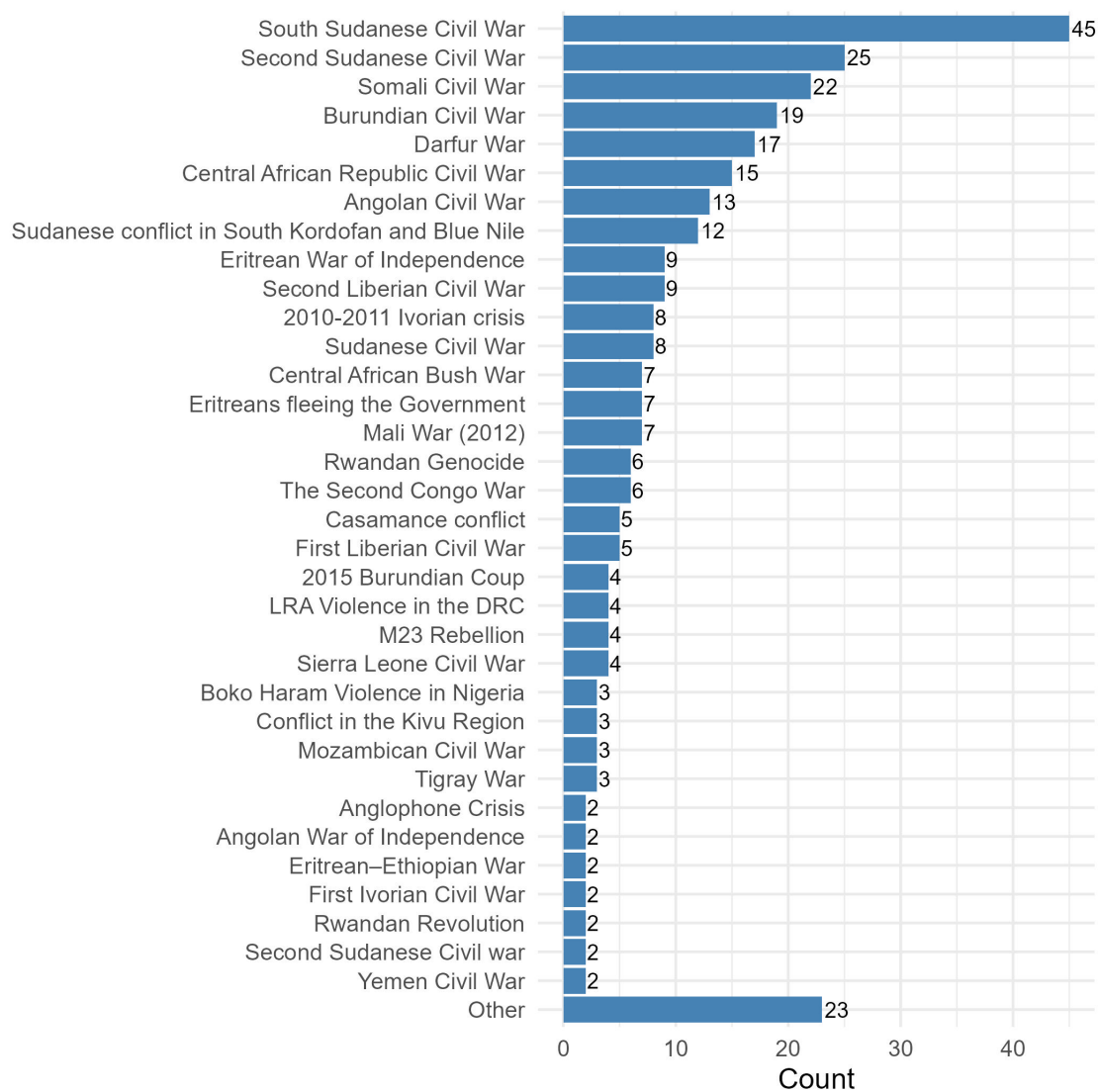
where the weights \hat{w}_ℓ are proportional to the number of switchers contributing to each $\hat{\delta}_\ell$ (de Chaisemartin & D'Haultfœuille, 2024). The CATE summarizes the total accumulated effect of camp

operation on vegetation cover over the post-treatment window. We designate $K = 7$ placebo years and $L = 7$ post-treatment years. We assume no anticipation based on the fact that camp designation and site preparation follow rather than precede population influxes.¹² The placebo coefficients $\hat{\delta}_\ell$ for $\ell \in \{-7, \dots, -1\}$ serve as a direct empirical test of this assumption and of the parallel trends condition more broadly.

¹²The UNHCR Handbook for Emergencies explicitly discourages planners from preparing new encampments before populations in need make the decision to cross the border, noting that “an unoccupied, developed site may send the wrong signal and encourage people to cross the border” (UNHCR, 2007).

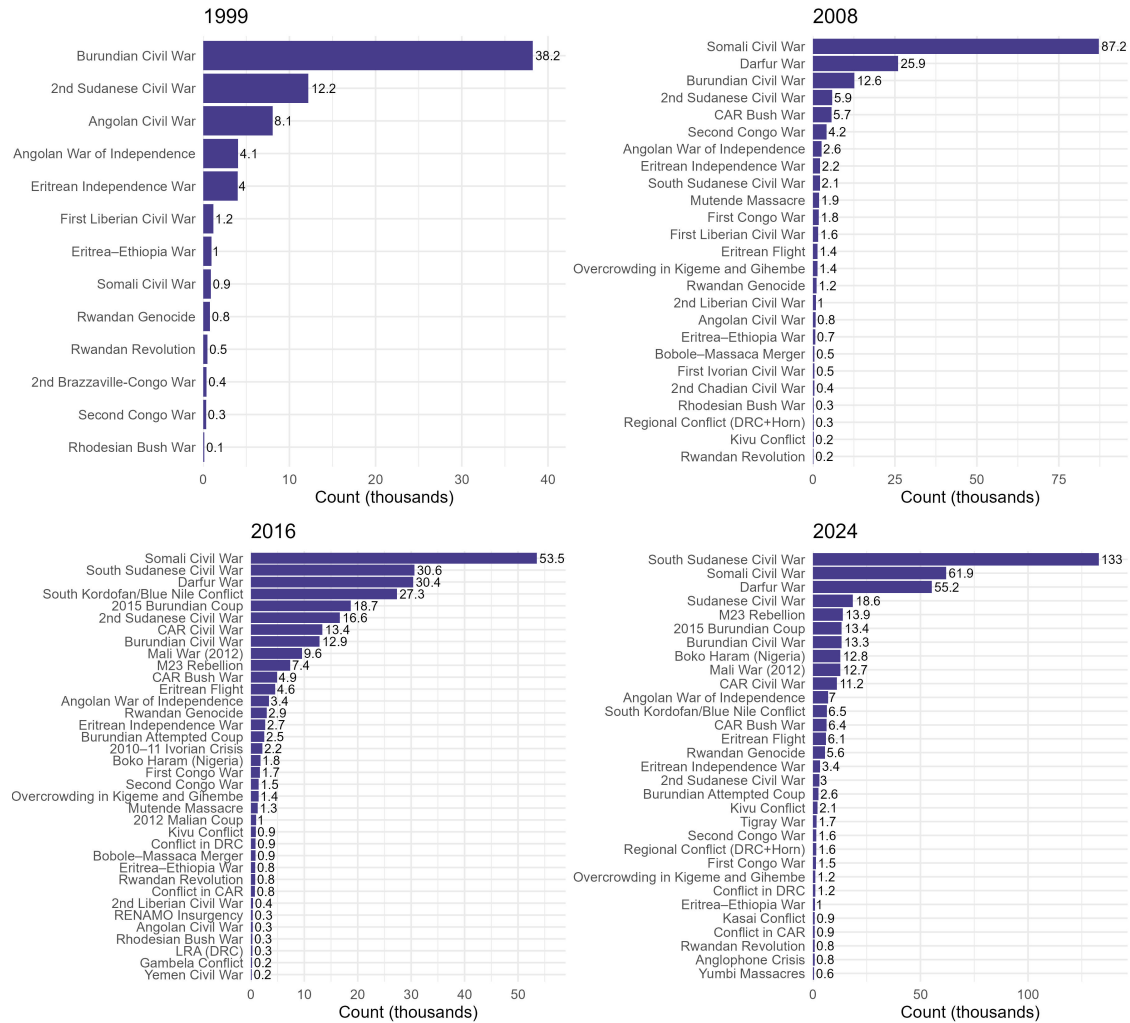
C Additional Figures

Figure C1: Number of camps by precipitating conflict event, ARCD camps that opened 1999 or later



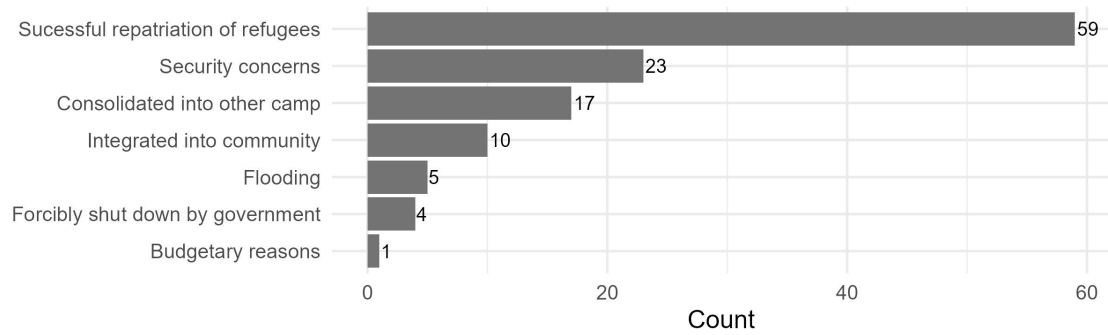
Notes: Counts from ARCD data, including camps and subcamps, with opening date of 1999 or later. “Other” category includes all cases in which a precipitating event was only associated with one ARCD entry opening.

Figure C2: Encamped population count (thousands) by precipitating event and year



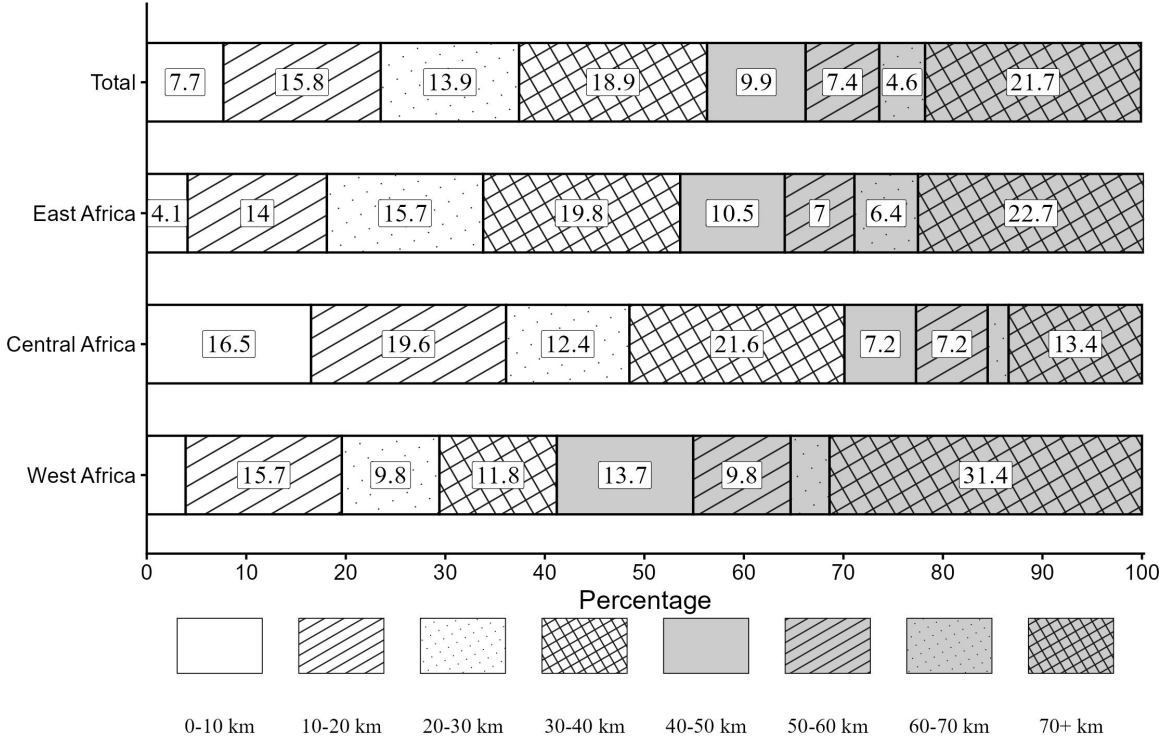
Notes: UNHCR Data from ARCD camps (including sub-camps) with non-missing population counts in reference years. Count represents the number of refugees accommodated in an ARCD camp in the reference year due to the precipitating event that occurred the same year or before.

Figure C3: Number of closed camps by reason for closing



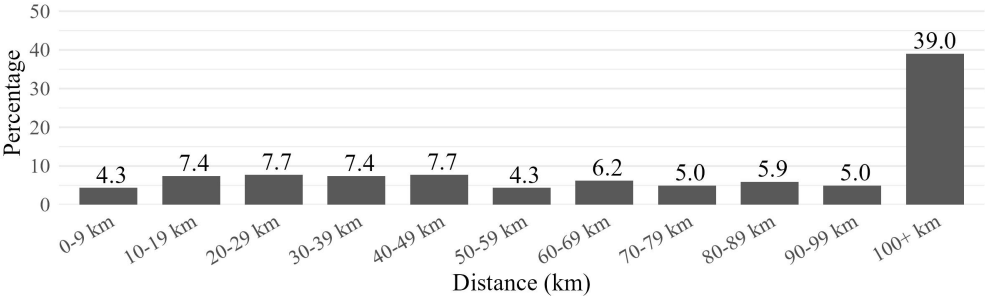
Notes: Counts from ARCD data subset to camps that closed before 2024 and had an identified reason for closure.

Figure C4: Distribution of camp distances to nearest internal border by region



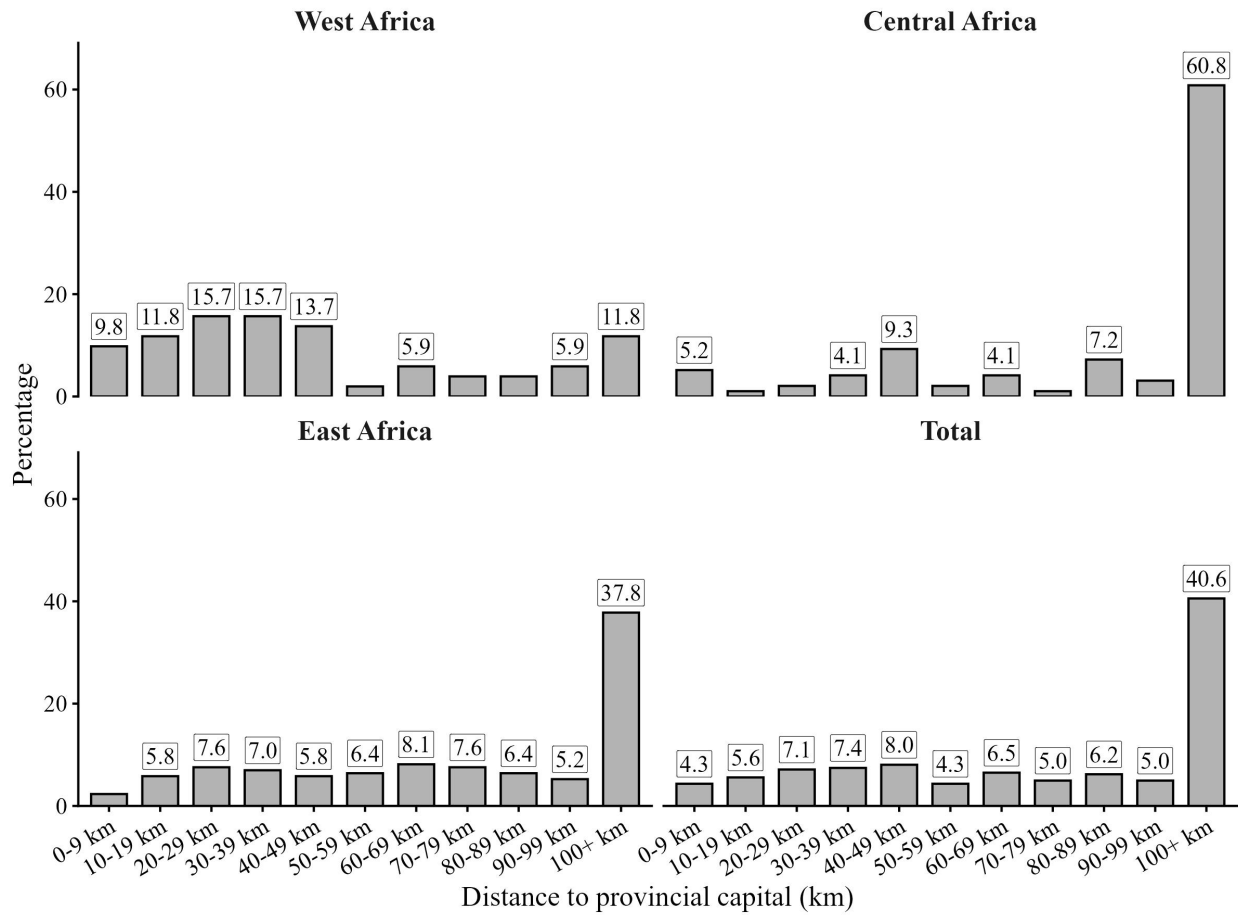
Notes: Camp geolocations based on ARCD and includes all ARCD entries (omitting subcamps) with latitude and longitude. Regional definitions are based on UN classification taxonomy (exception: Sudan is classified as East Africa instead of Northern Africa for this figure). Southern Africa omitted from regional breakdown due to low count (n=3) but included in the sample used to derive the Total column. National borders restricted to internal borders, and distance measurements based on Euclidean distance.

Figure C5: Distribution of ARCD refugee camps by proximity to capital city (provincial or national)



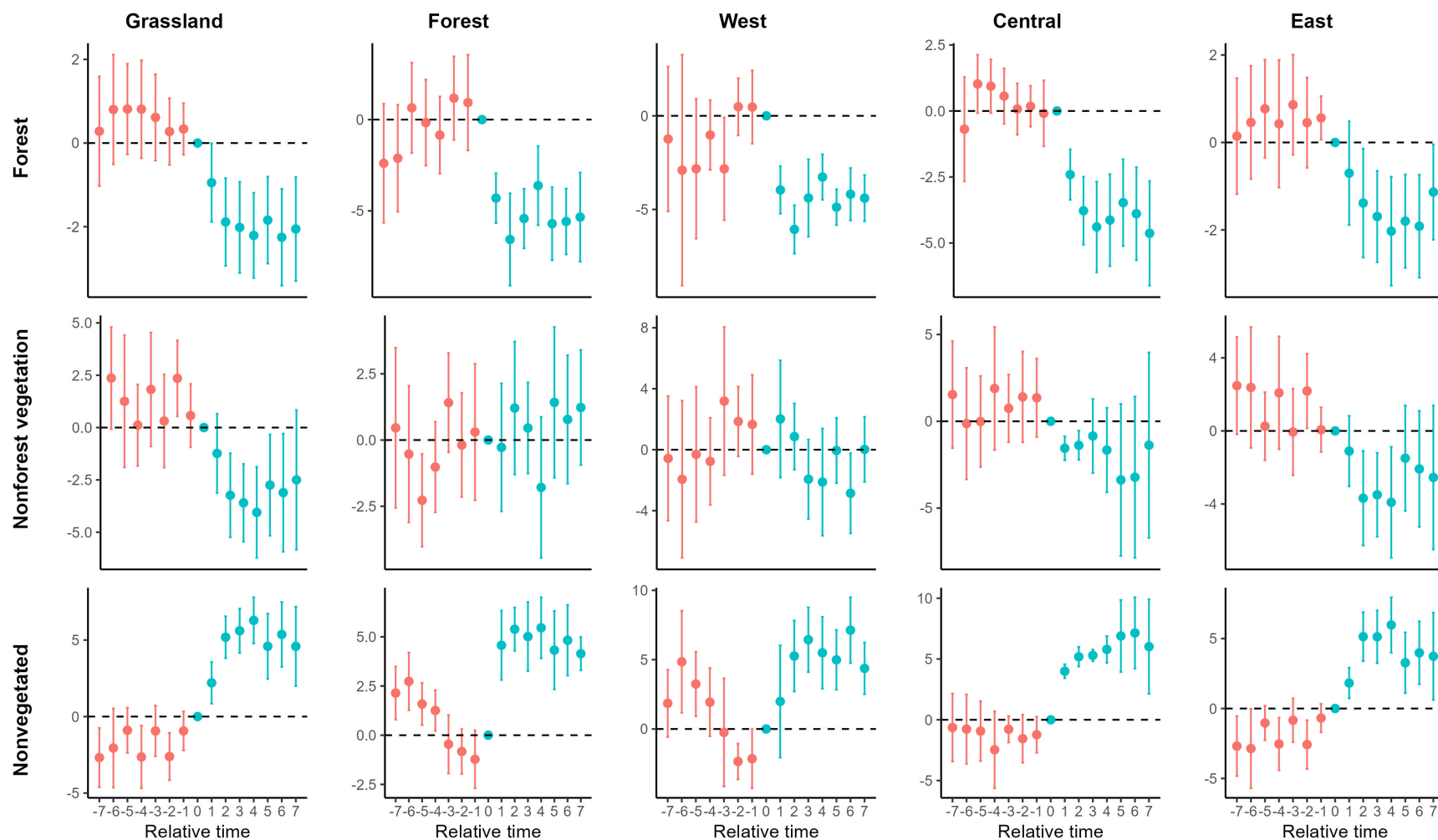
Notes: Minimum distances based on the nearest capital (provincial or national) of the same country as the ARCD camp location. ARCD sample limited to observations with geolocation data. Capital locations from Geonames Gazetteer data. Distance measurements based on Euclidean distance.

Figure C6: Distribution of distance to the nearest provincial capital by region



Notes: Camp geolocations based on ARCD and include all ARCD entries with a latitude and longitude. Regional definitions are based on UN classification taxonomy (exception: Sudan is classified as East Africa instead of Northern Africa for this figure). Southern Africa omitted from regional breakdown due to low count (n=3) but included in the sample used to derive the Total column. Provincial capital locations identified using Geonames Gazetteer.

Figure C7: Event study estimates of the impact of camp opening on forest and vegetation change in camp interiors, by maximum camp population size



Notes: The event study corresponds to the cumulative average treatment effect results reported in Table 3 and using (`did_multiplegt_dyn`). The ARCD camp sample is restricted to observations with (1) latitude and longitude listings, (2) opening date listings, and (3) main camps (excluding sub-camps), yielding a sub-sample of $N = 323$. The control group is drawn from areas 30-50 km from the nearest camp geoint outside protected areas. Random sampling of control points is stratified by country representation in the ARCD data ($N = 2,354$). Vegetation and forest data are sourced from MODIS Terra Continuous Fields 250 m annual data, 2000-2024. Zonal statistics are derived using a 1 km buffer around each treatment and control geoint. “Grasslands” = *Tropical & Subtropical Grasslands, Savannas & Shrublands*. “Forest” = *Tropical & Subtropical Moist Broadleaf Forests* or the *Tropical & Subtropical Dry Broadleaf Forests*. Regions based on UN classification.

D Additional Tables

Table D1: Means and t-tests of characteristics of 1 km buffer around camp geopoint at $t = -2$ years before camp opening relative to comparison samples by region

	ARCD			Border Comp. (BC)			ARCD vs. BC (Δ)	Country Comp. (CC)			ARCD vs. CC (Δ)
	Mean	SD	N	Mean	SD	N		Mean	SD	N	
<i>Percent forested</i>											
West	14.4	15.7	31	25.2	22.3	504	-10.8***	24.1	20.5	450	-9.8***
Cent.	10.5	11.2	60	14.5	19.5	4,172	-4.0***	20.2	28.4	3,467	-9.7***
East	8.7	9.0	85	12.3	12.4	11,326	-3.6***	10.4	12.7	9,032	-1.7*
Total	10.1	11.1	176	13.4	15.7	16,002	-3.3***	14.1	20.1	12,949	-3.9***
<i>Percent non-forest vegetation</i>											
West	51.3	26.9	31	52.5	25.2	504	-1.2	53.7	26.2	450	-2.4
Cent.	65.3	22.2	60	56.4	27.5	4,172	8.9***	36.2	33.4	3,467	29.1***
East	63.4	23.4	85	61.0	23.6	11,326	2.4	49.2	31.1	9,032	14.2***
Total	62.3	23.8	176	59.2	25.1	16,002	3.1*	45.1	32.3	12,949	17.2***
<i>Percent non-forest non-vegetated</i>											
West	34.6	35.7	31	22.6	30.3	504	12.0*	22.5	31.4	450	12.1*
Cent.	24.6	26.3	60	29.4	31.9	4,172	-4.9	43.8	43.0	3,467	-19.2***
East	28.2	27.0	85	27.0	28.3	11,326	1.2	40.7	37.6	9,032	-12.5***
Total	27.9	28.2	176	27.7	29.6	16,002	0.2	41.1	39.4	12,949	-13.2***
<i>Mean radiance</i>											
Cent.	0.012	0.060	26	0.056	1.288	2,239	-0.045	0.005	0.233	1,851	0.007
East	0.000	0.000	37	0.114	1.718	5,289	-0.114***	0.018	0.306	4,379	-0.018***
Total	0.005	0.038	64	0.096	1.596	7,557	-0.091***	0.014	0.286	6,256	-0.009

Notes: This table compares the ARCD sample to the comparison samples, focusing on time-varying characteristics observed prior to camp opening. *, **, and *** indicate differences significant at the 10, 5, and 1 percent levels, respectively, based on two-sample t-tests of mean differences between the ARCD sample and each comparison group. “ARCD vs. BC (Δ)” and “ARCD vs. CC (Δ)” report the difference in means between refugee camp locations and the corresponding comparison samples. All outcomes reflect the mean of a 250 m buffer around each camp geopoint. For ARCD observations, $t = -2$ refers to characteristics measured two years before the recorded camp opening year, so estimates capture pre-establishment conditions rather than conditions after camp settlement. The ARCD sample is restricted to main camps with valid latitude and longitude coordinates and non-missing opening years; sub-camps are excluded. Border comparison (BC) observations are sampled from points located 30–50 km from ARCD camp locations, while country comparison (CC) observations are drawn from the broader national sample. Both BC and CC samples exclude locations overlapping protected areas or bodies of water. Comparison observations are assigned to match the same country-specific calendar years represented by the ARCD sample at $t = -2$, so the temporal distribution of the comparison groups mirrors the pre-opening observation years of camps within each country. Regional definitions follow the UN regional classification, except Sudan is classified as East Africa. Southern Africa is excluded from subgroup analysis because of the small number of camps ($n = 2$), but is included in total estimates. West Africa is excluded from the mean radiance subpanel because too few camps opened after 2014 for reliable VIIRS coverage, but remains included in total estimates. “Cent.” denotes Central Africa.

Table D2: Vegetation Descriptive Statistics by Treatment Status for Entire Sample and for Treated Sample in $t - 1$ by Region, Camp Population, and Biome

Panel A: Sample Characteristics and Pre-Opening Conditions by Treatment Status and Region

	Camps (1)	Control (2)	Year prior to camp opening			
			Camps (3)	West Africa (4)	Central Africa (5)	East Africa (6)
Forest	8.29	16.58	9.82	12.90	11.23	7.71
Nonforest Vegetation	61.23	62.46	62.61	56.53	64.66	63.39
Nonforest Nonvegetated	30.81	21.30	27.87	30.80	24.42	29.24
N	6,400	73,138	176	31	60	85

Panel B: Pre-Opening Land Cover Conditions by Camp Population Size and Biome

	Year prior to camp opening					
	Pop < 1,000 (1)	Pop ∈ [1k, 10k) (2)	Pop ∈ [10k, 30k) (3)	Pop ≥ 30k (4)	Grasslands biome (5)	Forest biome (6)
Forest	14.70	14.80	10.30	6.18	8.60	19.26
Nonforest Veg.	73.43	65.74	66.01	54.38	64.35	70.73
Nonforest Nonveg.	12.26	19.81	24.03	39.73	27.40	10.32
N	6	52	66	58	149	35

Notes: This table presents descriptive statistics to support the causal analysis in [Section 5](#). The ARCD sample is restricted to main camps with valid latitude and longitude coordinates and non-missing opening years; sub-camps are excluded. All land cover values reflect the mean percentage within a 250 m buffer around each camp geopoint, derived from MODIS Terra Continuous Fields satellite data. **Panel A** presents means across two sets of columns. Columns (1) and (2) report means for the full sample by treatment status, averaged across all available years for ARCD camp locations and the control group, respectively. Columns (3)-(6) report means measured exclusively in the year prior to camp opening ($t = -1$), capturing pre-establishment conditions before camp settlement. Column (3) reports overall sample means at $t = -1$; Columns (4)-(6) disaggregate by region. **Panel B** reports means measured exclusively in the year prior to camp opening ($t = -1$) for ARCD camp locations only. Columns (1)-(4) stratify camps by maximum recorded population over all operational years, using the thresholds shown. Columns (5)-(6) stratify camps by biome type. The three land cover categories sum to approximately 100% within each column. Biome categories aggregate finer-resolution classifications of bioclimatic zones: “Forest biome” includes tropical and subtropical moist and dry broadleaf forests; “Grasslands biome” includes montane, tropical, and subtropical grasslands, savannas, and shrublands; “Wetlands” includes flooded grasslands and savannas, and mangroves; “Deserts” includes deserts and xeric shrublands. Regional definitions follow UN classification taxonomy. Population counts represent the maximum population recorded in ARCD during a camp’s operational years.