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

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

September 2025

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 describe its creation and use ARCD to analyze major trends over 25 years. We then show two applications combining ARCD with complementary data. First, we assess spatial features of camp locations compared to stratified random sites. Camps align with logistical guidelines—flat terrain, moderate vegetation—but are often near borders, protected areas, and far from provincial capitals. Second, we estimate the effect of camp openings on forest and vegetation cover using a differences-in-differences approach. Camp establishment reduces forest cover by 1–2 percentage points within two years, largely due to land clearing for shelter, infrastructure, and roads.

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.

[‡]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 percent of the world's refugees[1] obtain asylum in low- or middle-income countries close to the nation they were forced to leave. In sub-Saharan Africa (SSA), which currently hosts 27 percent 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 are residing 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 a dominant hosting modality. Accommodation in camps is particularly prevalent in SSA, where the majority of refugees (74 percent) 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). Refugee camps 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

[1]The definition of "refugee" we use throughout this paper aligns with UNHCR operational practice with respect to refugee status designation. A refugee is someone who, due to a credible fear of persecution or violence in their home country, has crossed an international border and received protected status. This operational definition extends on the 1967 Protocol on the Status of Refugees, which emphasizes persecution, to include additional and protection regimes. For example, in Africa, the 1969 Organization of African Unity Refugee Convention states that refugee status is also met when an individual is fleeing "events seriously disturbing public order."

[2]Other camp-like situations include residence in 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, schools, etc. (UNHCR & IOM, 2010). In some cases, refugees may create spontaneous or informal camps. Categorical designations are applied differently across settings. In Bangladesh, for example, camps that receive assistance and are managed by the government are nevertheless designated as "makeshift", in contrast to the older "formal" camps (Filipski et al., 2021). More work needs to be done to understand the factors that influence these classifications. For now, we can break down the 75 percent population statistic for SSA in 2023 as follows: 65.1 percent in formal, or planned, camps; 8.7 percent in informal settlements; 0.1 percent in collective centers, and 0.2 percent in transit centers. These figures are based on UNHCR data shared with the study team.

available data product that identifies 469 refugee camps in existence in SSA between 1999 and 2024 and identifies their geographic coordinates, years of operation, and population over time, as well as the event that originally forced the encamped population into exile.

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 towards 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 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).

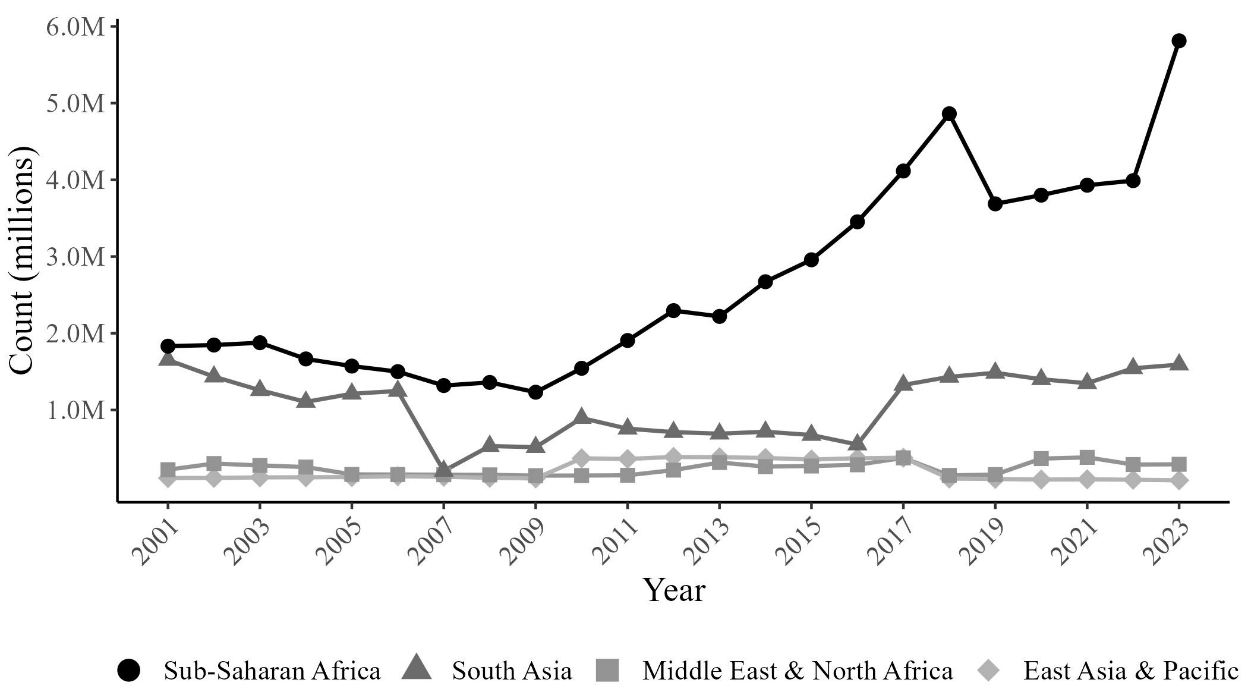
These contributions have greatly improved our understanding of markets, livelihoods, welfare, vulnerability, social interactions, and policies in African encampment settings. But data limitations continue to constrain the discourse. Consequently, evidence tends to be limited to national or sub-national case studies predominately focused on Uganda and Tanzania and almost exclusively studying East Africa.[3] To augment 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). Often, the study data originate with 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

[3]While Uganda is a common choice for researchers (Betts et al., 2017; d'Errico et al., 2022; Kadigo & Maystadt, 2023; Musasizi et al., 2024; Taylor et al., 2016), encampment policies in Uganda are unique, as encampments are large in area and refugee households are allocated plots to cultivate. There are few other camps on the continent that apply this model. In the case of Tanzania, studies tend to focus on camps in Kagera during the Great Lakes Crisis in the mind-1990's (Alix-Garcia & Saah, 2010; Baez, 2011; Maystadt & Duranton, 2019; Maystadt & Verwimp, 2014), and it remains unclear if this case study is relevant to modern encampment and hosting policy or to the contemporary economic and infrastructural development of hosting countries.

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 more recent years, the UNHCR has 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 percent) were in SSA.

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



Notes: Data from UNHCR Global Data Service. "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.

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 primary 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, whether publicly available (Zhou & Shaver, 2021) or proprietary, provide examples of the limitations of the UNHCR data. Two major problems are persistent 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 certain 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, as well as the number of camps accommodated. For example, false positives are very common in the Democratic Republic of Congo.

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. In fact, the UNHCR data only contain fields for when the camp was entered into the camp tracking system and when the entry's status was changed

[4] Despite this urgent humanitarian mandate, the UNHCR has taken strides towards supporting and incorporating research into its work. Since its founding in 2020, The Joint Data Center for Forced Displacement, a collaboration between UNHCR and the World Bank, has greatly augmented researcher access to data and information.

(for example from “active” camp to “inactive” camp or from “active” camp to “refugee site”). In many cases, these data entry dates are entered into the system several years after the documented establishment of the camp.

The aforementioned 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 over the past 25 years in SSA.

In this paper, we review our approach to constructing the ARCD data. Using a diverse set of sources and primarily drawing on the UNHCR People of Concern mapping platform and the UNHCR Statistical Yearbooks, we identify the geo-locations and years of operation 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 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 then use the data to explore and describe major trends in refugee camps in SSA over the past 25 years. Then, we offer an illustration of how researchers can combine the ARCD with other existing data products. First, we derive zonal statistics of refugee camp proximity to features of interest (borders, provincial and national capitals, and protected areas), landscape characteristics (terrain ruggedness, vegetation and forest cover, bioclimatic area), and population activity (nighttime light radiance). We evaluate descriptive statistics by region and compare summary means to two stratified counterfactual samples: a “border comparison” representing non-encampment areas within 30-50 km of a camp, and a “country comparison” drawn from across the host countries. We additionally apply a robust difference-in-differences (DID) estimator (de Chaisemartin & D’Haultfœuille, 2024), estimating the impact of camp openings on forest and vegetation cover of the camp interior.

Our descriptive results show that refugee camps are often close to international borders, particularly in Central Africa. Refugee camps are far from major cities: 39% of the sample is 100 km or more from the nearest capital (provincial or national). Due to border proximity, camps are close to protected areas, especially in East and West Africa. While camp border proximity does not correspond to UNHCR protection recommendations (UNHCR, 2024b), other factors exhibit stronger alignment with UN guidelines. For example, camps tend to have a lower terrain ruggedness score than comparison areas. Two years prior to camp openings, forest cover tends to be lower in camp areas than in comparison areas, and camps are more likely to be in a grasslands bioclimatic region relative to comparison samples. These environmental characteristics are suggestive of camp site selection that seeks to avoid the costs of clearing land or may also reflect a tendency to site camps in land characterized by lower demand.

Our DID results capture a 1.7 percentage point reduction in forest cover during the first seven years a camp is open, representing an 16.2% reduction relative to the baseline mean (10.5%). Event study analysis shows that this reduction occurs during the first two years the camp is operational then stabilizes, leaving a small but nonzero share of forest cover. The trend aligns with logistically motivated land clearing and not continuous biofuel harvesting. Additional heterogeneity analysis reveals that forest cover losses tend to occur when camps have larger populations. It is likely these camps are more densely populated such that settlement planning requires more tree felling to make way for shelters, infrastructure, etc.

The remainder of this paper proceeds as follows. Section 2 describes the underlying sources of the ARCD and how we harmonized the information to cross-validate our listing and generate variables on population size, opening and closing years, and reasons for camp establishment and closure. This section also provides a simple validation check using country-level UNHCR data for encamped populations. Section 3 uses the ARCD to examine the state of refugee camps in SSA and major trends in refugee camps in the region over the past 25 years. Section 4 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 5 concludes.

2 Dataset Construction

Identifying refugee camps across SSA and their years of operation is not a trivial task. The underlying data sources are numerous, yet no single one is perfect and complete. Sources also frequently contradict each other. This section outlines the process through which we built the ARCD and how we prioritized information and made decisions over which information was the most reliable.

The dataset contains five major components: (1) the list of camps, (2) the years the camps were operational, (3) the geolocational coordinates (longitude and latitude) of each camp, (4) the annual population of refugees in each camp, and (5) decisions over which information is prioritized in the presence of contradictions. We additionally identify the events that motivated the opening of each camp, and for camps that closed within the years of the panel, we identify reasons for closure. Below, we discuss the development of each component and present the results of a simple validation exercise.

2.1 List of Camps

We derived the primary list of camps from the UNHCR's Persons of Concern (POC) Map, which provides a mapping of all active and inactive locations for refugees and internally displaced persons (IDP) locations around the world. This map contains the locations for all sites related to refugees and IDPs, not just camps. By extracting the attribute table populating this map, we create a list of all POC sites, both formal and informal, recognized by the UNHCR. We then filter this register to only entries with the "refugee" tag and "Formal Settlement" subtype (the POC Map applies the label "Settlement" to sites that align with the ARCD definition of a refugee camp). This provides us with our initial listing of camps. We also check other subtypes in the POC universe (integrated communities, settlements, transit camps, and districts hosting refugees). We eliminate these other non-camp locations by individually examining each entry to verify if and when UNHCR ever identified it as a camp during its existence.

To supplement the UNHCR POC Map and to further validate that the entries in our listing represent camps, we extract the list of refugee camps from the UNHCR's Statistical Yearbooks from 1999 to 2016 (see citation listing in Section A.1 of the Online Appendix). These Statistical Yearbooks contain detailed lists of each location where people of concern were residing around the world in a given year, as well as the status of their location (i.e. whether they are settled in a planned refugee camp or some other type of location). Although extremely informative, there is reason to view the Yearbooks as incomplete, since they are the result of data gathering within a relatively decentralized administrative system of humanitarian and emergency response. To further supplement the Yearbooks and POC Map, we use news reports, peer-reviewed papers, and reports from NGOs to confirm which locations were indeed camps.

We include refugee sites that are considered to be "refugee settlements" by the host country. Uganda is perhaps the best known case of hosting refugee settlements, but this approach is also found in Kenya, Zambia, Gabon, and other countries. Settlements are often distinguished from camps by their emphasis on "self-reliance" programming that seeks to eliminate refugees' dependence on humanitarian aid (Armstrong, 1988; Chambers, 1982; Harrell-Bond, 1986).[5] Nevertheless, these sites have many of the same characteristics of refugee camps, and in UNHCR documentation like the Statistical Yearbooks they are categorized as camps. As a result, we add them to the dataset, although researchers may wish to account for them separately in their analysis.[6]

2.2 Timing of Camp Operations

Unfortunately, the UNHCR POC Map does not contain reliable information on the start date and closing date of camps. As mentioned above, it includes a field for the date that the site was entered into the UNHCR tracking system, and this does correspond with camp openings for a number of well-known camps. However, we found numerous instances where this date was inaccurate, motivating us to systematically review publicly available information for opening and closing dates.

[5] Self-reliance programming is often unsuccessful in its goal of eliminating aid dependence: see Harrell-Bond (1986) and Verdirame and Pobjoy (2013).

[6] For example, they are often large in land area: Nakivale Settlement in Uganda is 195 km^2 (Bjørkhaug, 2020), far larger than most camps. If the analysis assumes that camps are small in land area, the researcher may choose to omit all these settlements

For this effort, the primary source for site openings is operational documentation from the UNHCR Document Portal and documents from *Refworld*, an online UNHCR database for documents pertaining to law or policy with regards to displaced persons. For a majority of camps, these sources are sufficient for identifying an opening date, either through a camp profile document or press release recorded at the time of the camp's opening. For each camp, both the year from the source and the source itself are archived in the "Opening Year" and "Opening Year Source" columns respectively. This process also helped us eliminate site listings from the Yearbooks or POC Map that turned out to not be refugee camps upon closer inspection. For many non-camp refugee sites, like the "Maryland Communities" in Liberia that host refugees in a more integrated fashion, there is no clear information on an opening date due to the lack of a formal start or end date to the hosting itself. In cases such as these, where the entry was clearly not a camp, the site is moved to the "Non-Camp" Sheet, alongside a source showing what the site was/is instead of a camp.

Likewise, the most common sources for identifying closing dates are either documents from the UNHCR document portal or *Refworld* publications. When a camp opens, there is often an explicit opening date, while for closings, a camp often dwindles away until support for the camp eventually ceases. 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 choose the year repatriation ended for the camp (2007), although camp population data is

[7] In the best case scenario, repatriation occurs when it is safe for refugees to voluntarily return home. But unfortunately, we cannot assume that the cessation of hostilities in the country of origin is a necessary condition for repatriation. See for example Lawrence (2017) for an account of Somali refugees compelled by poor conditions in the Dadaab Camp complex to repatriate, despite ongoing conflict.

contained in the dataset for researchers to make different decisions that they feel are reasonable.

As with the openings, the source for any closing dates is present next to the date applicable in the spreadsheet. While we were able to find opening and closing dates for most of the camps in the sample, there were certain (mostly older) camps that had very little documentation and therefore are missing opening or closing dates. For example, while we could find evidence that Gerihun Camp in Liberia was open before 2000, the documentation before 2000 is so sparse that it is impossible to place an exact opening year. This is also the case for some camps for which we could only find approximate dates, such as some camps in the Sudan that opened in the 1960s.

To supplement this opening and closing information, we provide timing information that is inferred from entry into the UNHCR Statistical Yearbooks. If a camp ceases to be listed in the Yearbooks, one could infer that it has closed. Similarly, after 1999, one can infer that the first year the location is listed in the yearbook is the year it opens. We provide the dates implied by the yearbook entries in this fashion in the dataset for locations that do not have any other available documentation for their operational timing for researchers that are comfortable with this inference. However, we locate these years it in another variable labeled "First Mention in the Yearbook" since it is based on an inference rather than clear statement from a primary source.

2.3 Spatial and Population Data

Data on the longitude and latitude for camps mostly come from the UNHCR POC Map. For the few camps that were missing location information from that map, such as Oru Camp in Nigeria or Nicla Camp in Ivory Coast, we supplemented with data available from UNHCR operational maps that are published intermittently and available on *RefWorld*.

The main resource for camp populations was the UNHCR Statistical Yearbooks. Published every year from 1999 to 2016, the Yearbooks document key statistics on refugee populations, including estimates of camp populations, locations, and statuses. These yearbooks provide a foundational dataset for identifying refugee camps that were active during that time period. 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 underreporting 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 aggregated into “Shagrabs 1, 2 and 3” in 2007, but disaggregated in every year past into the camps themselves. These size and aggregation issues also influence Ugandan settlements. Uganda’s Adjumani District (a first administrative division) contains many settlements, such as Alere 1 and Alere 2. But in each Yearbook, the settlements are all aggregated under the “Adjumani” entry, creating the impression that Adjumani is itself a settlement.

Going through the Yearbooks revealed that in the UNHCR POC map, many refugee camps are coded as being informal refugee sites due to their closing being categorized as a transformation from formal to informal instead of a switch from open to closed. To correct this, for all refugee locations in the Yearbooks not on the list, we use the same resources used in finding the opening and closing dates to check whether they were official refugee camps. For camps that are in the Yearbooks but not coded as formal settlements in the POC map, we extract coordinates as well as opening and closing dates.

For population data after 2016, the most valuable resources are the UNHCR WASH dashboard, which has conducted regular audits of water and sanitation access in major refugee camps, and the UNHCR POC map covering the years 2021 to 2023. The WASH audits, conducted annually and often monthly, include population estimates at the time of each survey. By compiling these audits, we gather over 450 combined years of population data, with the majority coming from 2020 to 2024.

We stress, however, that incorporating these monthly-level data alongside all sources other than the UNHCR Statistical Yearbooks introduces data collection inconsistencies into the dataset that researchers should understand. The Yearbooks consistently record camp populations at the end of each year, whereas the WASH data captures populations at varying points in time throughout the year. This discrepancy means that, for year-over-year comparisons, a population estimate taken in January 2017 might be directly compared to one from December 2016 by a researcher that is unaware of inconsistency in the timing of the observations. While both are technically accurate for their respective years, they may not reflect meaningful changes over time, complicating longitudinal analyses of camp populations. So, for any camps where the WASH data or other sources provided multiple months of observation within the same year, we always take the latter. For each observation, the month is included in the "Month of Observation" column.

For the years and camps with no WASH data, we rely on individual countries' reporting and the UNHCR POC Population Map. The UNHCR POC Population map has been published yearly, starting in 2021, and offers annual end-of-year population counts for most formal and informal refugee settlements and camps. We added this data from 2021-2023 to the ARCD, filling in much of the space in the last three years not covered by the WASH dataset. Since this dataset only covers 2021-2023, the gap between 2016 and 2021 is mostly filled with individual country-level reporting. This leads to considerable variability by country in population data completeness. For example, Cameroon has published population statistics for all its refugee camps every year since 2014, while the DRC has published almost no population information for any of its camps after 2016.

Camps often have a structure of sub-encampments that make up an entire complex. For example, the settlements in Uganda are typically organized into settlements and subzones, where settlements represent broader areas containing multiple smaller subzones. In the context of the Adjumani District, however, these subzones are referred to as settlements, even though they are more comparable in size and structure to the subzones of other settlements. Very large refugee camps may also be organized based on sub-camps. The Dadaab refugee complex in Kenya, for example, consists of several distinct camps. To keep these hierarchical structures clear in the dataset, each camp has both a camp number, indicating what large camp it belongs to, as well as a variable "Sub camp Indicator" that is zero if the camp is the main aggregated camp, and non-zero for each sub camp of the main camp. This allows the researcher to decide how to manage the different types of organizational structures based on the needs of a particular project.

2.4 Cleaning and Prioritization Decisions

As mentioned, the underlying sources for this dataset reflect an operational environment that is uncertain and fluid, as well as a decentralized institutional structure at the UNHCR that is collecting data for operational decision making rather than academic research. This leads to contradictions in sources, such as inaccurate and incomplete lists of refugee camp locations across official documents, as well as contradictions in opening and closing dates.

We also sometimes noticed discrepancies between camp opening and closing information. If this is the case, we prioritized the date that had the most sources to validate it. If each date has the same number of sources, we prioritized the date that came from news reports or academic papers, since these are most likely to be the result of direct inquiry by the reporter or researcher.

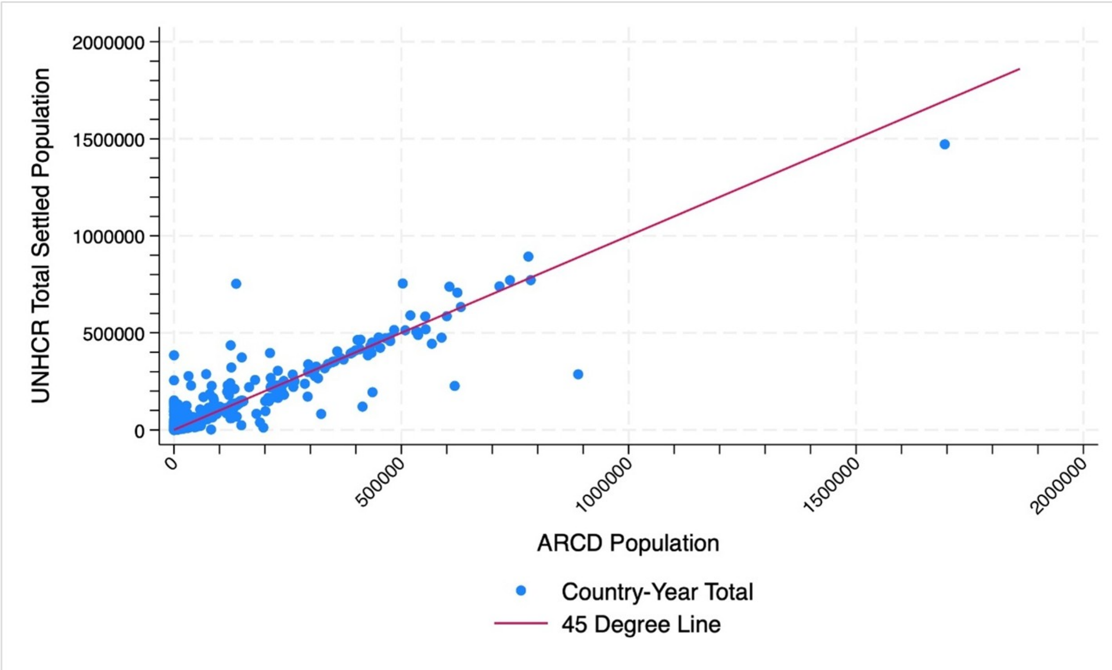
2.5 Opening and Closing Reasons Data

We also include information on the reason for each refugee camp's opening and closing. We do this from careful review of online resources available from the UNHCR, *Refworld*, public media, and other available information. All primary sources are listed in the dataset file. 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 tracking reasons for opening than it is reasons for closing.

2.6 Consistency with UNHCR Country Data: National-Level Validation Exercise

To evaluate the accuracy of our camp population data collection, we compare it to country-level data that is available from the UNHCR Global Data Service on the population in camps over time. To do this, we use the ARCD to construct a country-year panel dataset with a variable aggregating

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.

the camp population information available in the ARCD to the country level. We then examine its correlation with the country-year panel available from the UNHCR. When one is regressed on the other, the coefficient is 0.906 with an R-squared of 0.844. Figure 2 contains the scatter plot 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 counting sites that are classified as "Informal Settlement" in the UNHCR data. While less common, it may also be the case that a population no longer serviced by the UNHCR is nevertheless still included in the data underlying ARCD. For 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. 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.

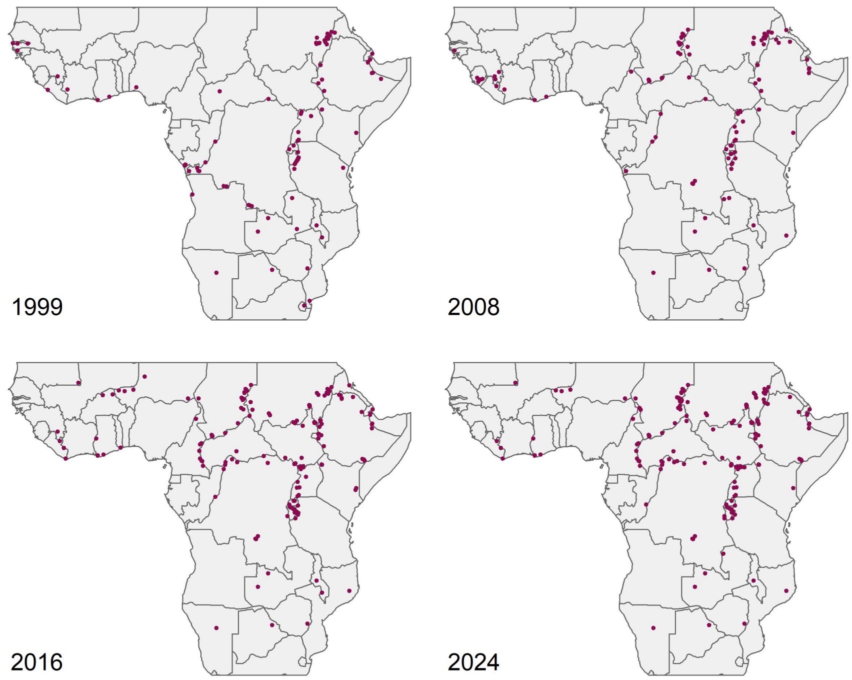
3 Descriptive Properties of ARCD

We next demonstrate the descriptive features of the ARCD data, which demonstrates trends in refugee encampment over time across the sub-continent.

3.1 Spatial Distribution of Camps

For the purpose of illustration, we map the ARCD points across the subcontinent in Figure 3 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 and resolution over time. For example, we observe numerous camps open after 1999 in West Africa in the wake 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, with camps remaining active over the entire period. This is particularly the case in the Great Lakes and Horn of

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.

Africa region.

3.2 Precipitating Events and Closings

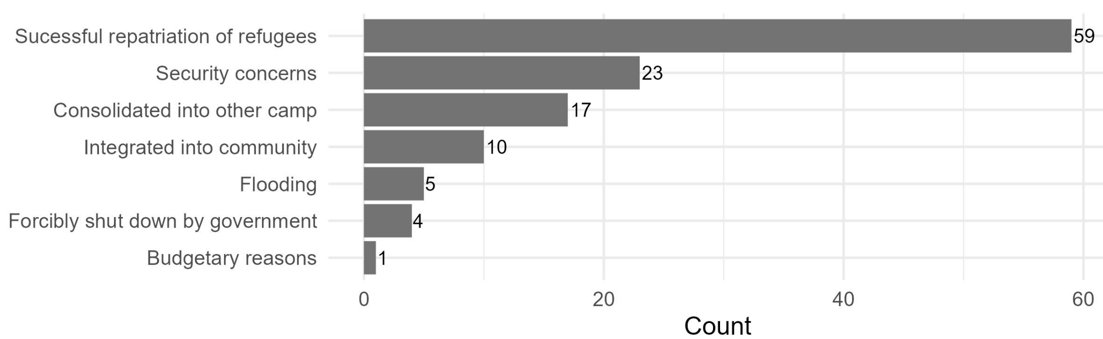
Figure B1 in the Online Appendix presents the counts of ARCD camps by the main precipitating events in the dataset. The two events that led to the first and second largest count of refugee camps were both related to conflicts in Sudan: The South Sudanese Civil War (44 camps) and the Second Sudanese Civil 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 B2 (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 population 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 4 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 either 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 indeed occur, then these events are rarely publicized or noted in a way so that a researcher at a distance can identify. In fact, we were only able to confirm 10 instances in which camps closed as part of a disencampment and integration process.

Figure 4: 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.

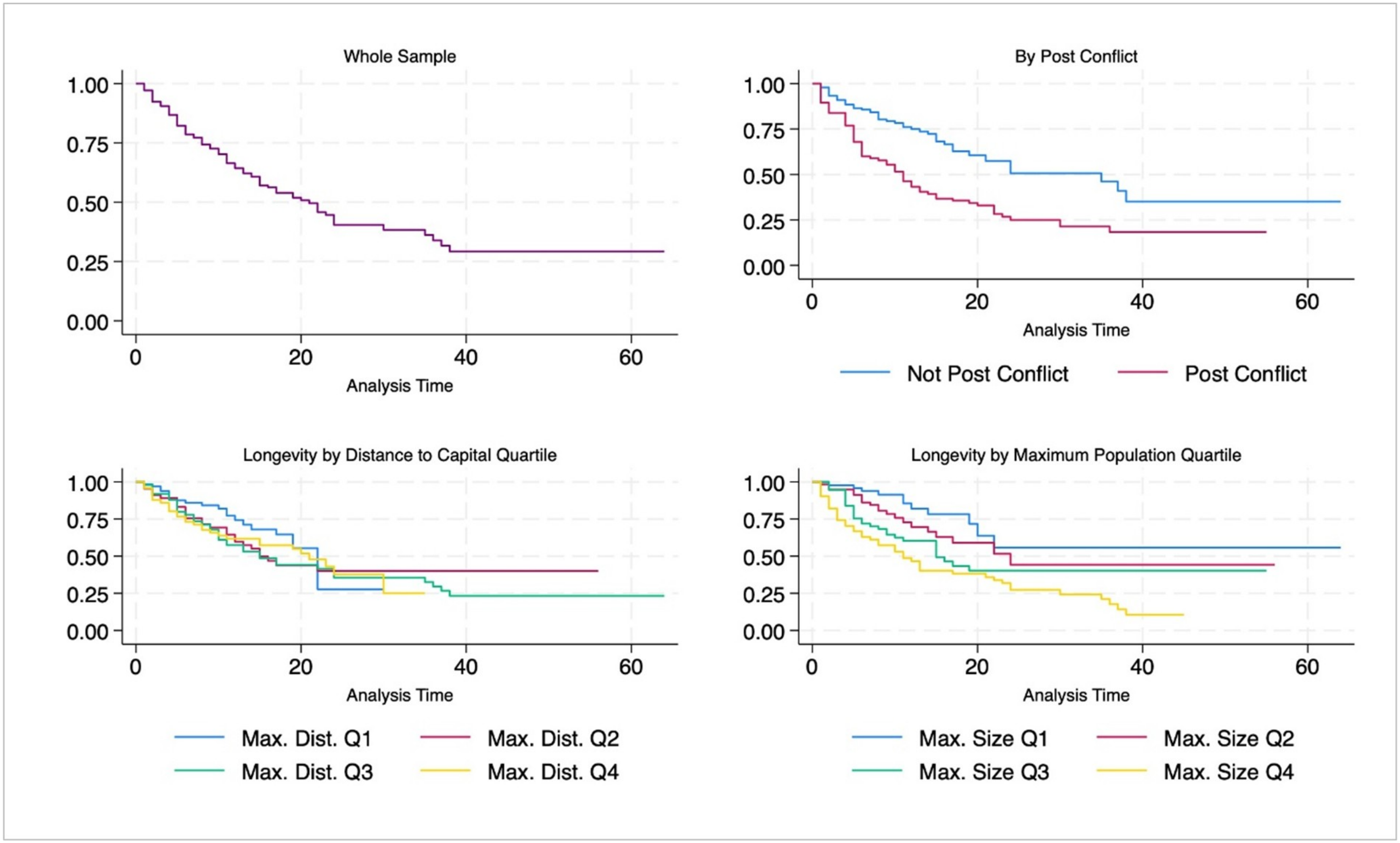
3.3 Camp Longevity Patterns

We use the ARCD to examine the basic patterns in the longevity of refugee camps. The extraordinary duration of certain camps has been noted by many observers over the decades. The Kakuma Camp in Kenya has been open since 1992 and still hosts over 203,000 inhabitants as of 2024. Dukwi in Botswana is even older. It started in 1978 and continues to host almost 800 inhabitants. Camps like Kakuma and Dukwi are common in the region, but 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 potential use of the ARCD. First, we provide basic information on how long refugee camps last in the region. Of the 274 camps in the dataset that came into existence since 1999, 36.9 percent have closed, while the remaining 63.1 percent are still open. Of those that have closed, the average length of time the camps have been open has been 10.21 years, lasting anywhere between one and 38 years. Of those that continue to operate, the average length of time the camp has been in existence (as of 2024) was 16.18 years. This is remarkably close to the controversial 17-year lifespan of camps often cited (Moore, 2017; UNHCR, 2006). Altogether, once a camp opens, we expect a camp to last on average at least 13.98 years.

Next, we use the data to 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 using the Kaplan-Meier (K-M) estimator. 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 to date the starting year of camps for those that do not contain an otherwise verified starting year. Figure 6 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. There is a steady rate of closure from zero until about 23 years, at which point the rate of exit becomes very flat. Slightly before 40 years, the probability of closure becomes entirely flat at about 26 percent. This means that once a camp is established, there is a little over a one-in-four chance that camp will last at least 40 years. There are 12 camps in the data set that have existed 40 years or longer, located in Botswana (1), Sudan (7), Uganda (2), and Zambia (2).

Figure 5: Kaplan-Meier Estimates of Survival Functions by camp attributes



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 are resolved. Distance to capital is based on Euclidean distance measure.

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 20% chance a camps operates for up to 40 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 slight deviation for camps in the first quartile 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. We observe a clear relationship between the size that the camp gets and the length of time the camp stays open, with camps in the first quartile clearly lasting longer than those in the other quartiles.

4 Descriptive and Causal Analyses

This section provides examples of how the ARCD can be combined with existing spatial products to examine camp attributes or impacts. The point locations can be used to infer relevant statistics on distance and proximity to relevant features (ex: borders, urban areas, protected areas); land cover, terrain and biophysical characteristics (ex: forest and vegetation cover, terrain ruggedness); climatic conditions and climate change vulnerability (ex: drought, evapotranspiration, natural hazard risk); population and economic features as visible by satellite (settlement cover, night light radiance), and others. The point location data can also be used in conjunction with other spatially explicit data products on, for example, household consumption and welfare (Demographic and Health Surveys, Living Standard Measurement Surveys), conflict occurrence (Armed Conflict Location and Event Database, Uppsala Conflict Data Program), or refugee hosting policy (Developing World Refugee and Asylum Policy dataset). Because the data include camp years of operations, the researcher can examine baseline characteristics prior to camp establishment and can also examine how geographic and spatial features change over time once camps open.

4.1 Geospatial characteristics of camps

We demonstrate simple descriptive results from a spatial exercise in which we determine several zonal statistics and distance measures to better understand what types of areas are most likely to be selected to host a camp. Throughout this exercise, we restrict the ARCD sample to only geolocated entries that represent the main camp, omitting sub-camps to avoid over-counting: this yields a camp sample of 325 point locations. The sample size is lower when examining time-varying spatial properties, as we are restricted to camps with well-identified open dates that are sufficiently within the temporal scope of the relevant gridded data product. In this case, we use ARCD data on operational years based on verified sources and do not supplement with the inferred years based on the timing of appearances in the Statistical Yearbooks.

We descriptively assess statistical differences in camp attributes relative to other regions based on two comparison samples. For our "border comparison" sample, we randomly select $n = 2,911$ points stratified by ACRD national location distributions where (1) all geolocations must be and from a camp geolocation, (2) geolocations cannot fall in bodies of water, and (3) geolocations cannot fall within protected areas.[8] The border comparison sample provides us with a sense of whether encampment areas locally differ from other nearby regions. This tells us if whether, conditional on the selection of camp locations close to borders (Salemi, 2021), particular areas are more likely to be selected than others. Our second sample ($n = 2,468$) provides us with a "country comparison".[9] This sample is also stratified by the distribution of ACRD host countries and restricts randomly

[8] With very few exceptions (three camps in Ethiopia) all ARCD points are outside of protected areas of Africa. The omission of protected areas constitutes an advancement relative to Coniglio et al. (2023), which is agnostic to protected area tenure and draws comparison locations from within these regions, even though they most likely cannot be settled.

[9] Please note the sample size of the border and country comparison samples. The program written to generate the two comparison samples was designed for a target count of 3,000, a size chosen to increase statistical power while not sacrificing computational efficiency. The routine drops points that fall in bodies of water or protected areas, resulting in marginally lower counts than 3,000. The border sample routine oversamples, so the final count is closer to 3,000.

placed geopoints so they may not fall within water bodies or protected areas. But the country comparison points can be drawn from any location in the country, so the sample shows us how camps vary with respect to country averages.

For our ACRD location and the two comparison samples, we estimate Euclidean proximity to provincial and national capitals based on the GeoNames Gazetteer database. We restrict the distance minimization to within-country matches so that border-proximate camps cannot be matched with a nearby city on the other side of an international boundary. We also measure Euclidean distance to the nearest active terrestrial protected area (national parks, safari parks, wildlife refuges, etc.) using Protected Planet polygon data (NEP-WCMC & IUCN, 2025), as well as interior borders based on publicly available administrative divisions. To briefly gauge time-invariant biological features, we determine each point's terrain ruggedness index (TRI) score by applying the method introduced in Riley et al. (1999) to 30-meter elevation data (ESA, 2024). We also identify the broader bioecological region the camp falls into (forest, grasslands, wetlands, or desert) using the RESOLVE ecoregions polygons (Dinerstein et al., 2017).

We consider time-varying geospatial information by examining vegetation cover and nighttime light radiance. For vegetation cover, we rely on the MODIS Terra Continuous Fields 250m resolution data available 2000-2024 to characterize the percent forest, non-forest vegetation, and non-forested, non-vegetative cover in camp areas by year (DiMiceli et al., 2022). We also use the VIIRS nighttime light annual mean radiance maps 2012-2021 (Elvidge et al., 2021) and select the masked version of each annual composite, which has been cleaned of random noise.[10]

We derive annual zonal statistics of annual percent forest cover, non-forest vegetative cover, and non-forest non-vegetative cover, as well as mean radiance, using a 1 km buffer around the camp centroid. Our objective is to capture a sample of the camp interior. Unlike some past work characterizing refugee camp locations (ex: Coniglio et al. (2023), we do not use a larger buffer area, such as 10km. When our camp identification information is based on a centroid and not a boundary polygon, a buffer considerably larger than the spatial extent of many encampment areas

[10]Random noise in nighttime light data can come from aurora, cloud cover, sunlight contamination, or ephemeral light (ex: cargo ships).

will combine camp interiors and peripheries in a manner endogenous to camp size, resulting in non-random measurement error.[11]

Using these data, we report summary statistics on all time-invariant attributes overall and by camp region (West, Central, or East Africa).[12] We compare these attributes to the two comparison samples and use t-tests to demonstrate any statistically meaningful disparities. Using the time-invariant data, we examine descriptive characteristics two years prior to camp opening and compare these to comparison sample means weighted by country-year representativeness in the relative time sample for camps.

4.1.1 Results

Table 1 provides the time-invariant spatial characteristics of the camp sample and compares them to both the border and country comparison samples, using t-tests to identify significant differences. On average, camps in the ARCD are not far from internal borders (53.5 km) or protected areas (58.7 km) and tend to be far removed from national (481.9 km) and even provincial (101.0 km) capitals. Relative to national averages, camps are significantly closer to protected areas (123.2 km) and, in fact, *closer* to provincial capitals (114.6 km). The marginally greater proximity of camps to provincial capitals may suggest that, on average, camps are sited marginally closer to key distribution points for necessary goods. But we note that the border sample is not statistically different in terms of the average distance to a provincial capital (100.0 km). Taken together, these results suggest that when using Euclidean distance measurements, border proximity results in slightly higher provincial capital proximity, though the mean distance remains large in magnitude. Camp proximity to protected areas also appears to be related to border proximity, as the border comparison sample is equally proximate

[11] It is crucial that researcher makes an explicit decision as to whether they wish to measure the camp interior or exterior area. Camps vary substantially in size: Mahama Camp in Rwanda is 1.6 km^2, while Nakivale Settlement in Uganda is 185 km^2 (Bjørkhaug, 2020; UNHCR, 2024a). A 10 km buffer around Mahama Camp will include Mahama but will mostly capture the surrounding host community. A 10km buffer around the Nakivale centroid will predominately or exclusively intersect Nakivale. These two zonal statistics are measuring different phenomena: the former largely captures the host community, and the latter largely captures the camp area. Clear analysis of camp and host community attributes requires that we examine these areas separately.

[12] Due to low counts ($N = 2$), we omit Southern Africa from the regional disaggregation but include these camps to estimate statistics for the Total sample.

to protected area boundaries (61.1 km).

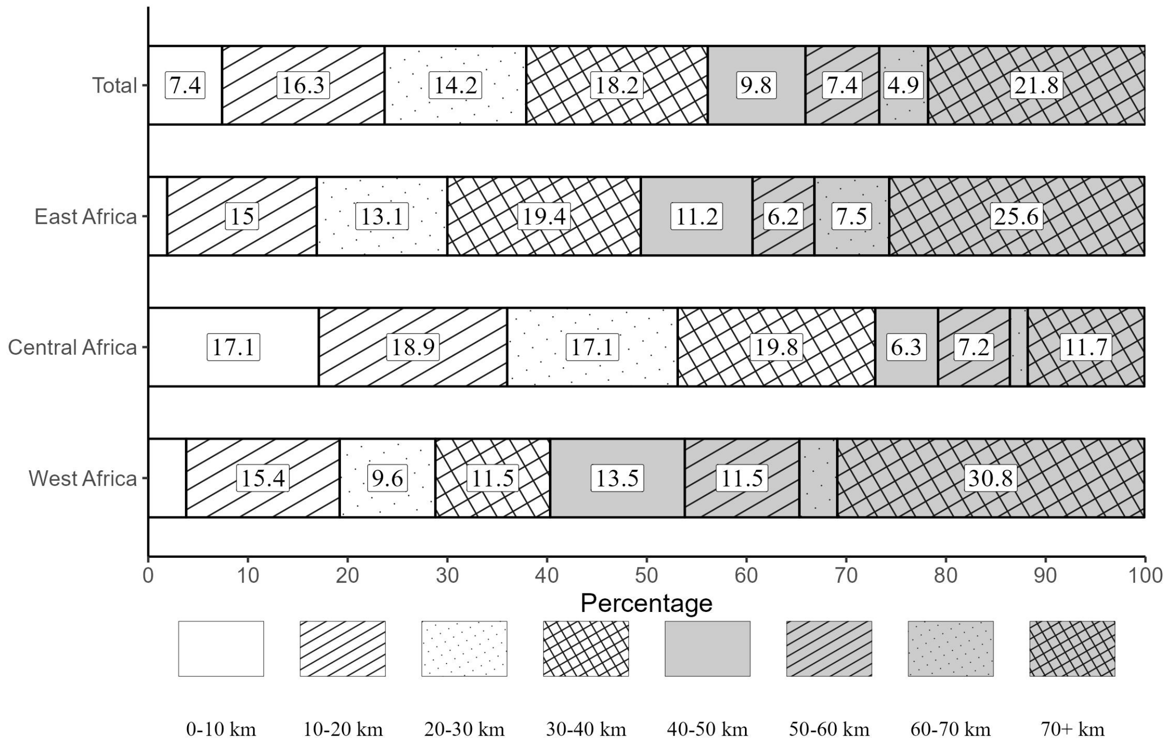
Table 1: Means and t-tests of of refugee camp time-invariant characteristics relative to border and country comparison samples

	ARD		Border Sample		ARD vs.	Country Sample		ARD vs.
	Mean	SD	Mean	SD	BS (Δ)	Mean	SD	CS (Δ)
Euclidean dist.								
Internal border	53.5	54.4	58.3	56.2	-4.8	129.4	118.8	75.9***
Protected area	58.7	66.6	61.1	68.1	-2.4	95.8	123.2	37.1***
Provincial capital	101.0	86.3	100.0	78.8	1.0	136.3	114.6	35.3***
National capital	481.9	331.0	507.1	366.8	-25.2	511.9	360.2	-30.0
Terrain ruggedness								
TRI	4.0	4.1	7.1	9.1	-3.1***	7.2	9.2	-3.2***
Bioclimatic area								
Forests	19.4	39.6	21.3	41.0	-1.9	21.7	41.2	-2.3
Wetlands	1.5	12.3	1.3	11.5	0.2	2.2	14.6	-0.6
Grasslands	75.7	43.0	72.0	44.9	3.7	63.5	48.1	12.2***
Deserts	3.4	18.1	5.3	22.5	-2.0	12.6	33.2	-9.3***

Notes: *, **, and *** indicate significance at the 5, 1, and 0.1 percent levels. ARCD sample restricted to observations with latitude and longitude listings and to main camps, removing sub-camps. Distance to capitals is based on provincial and national capital geopoints included in the GeoNames Gazetteer repository, and matches are restricted to within the same country as the host country. "TRI" stands for the Terrain Ruggedness Index, estimated using 30m digital elevation data following Riley et al. (1999). Values between 0 and 80 represent relatively level terrain, with lower values being more level. Bioclimatic regions aggregate finer-resolution categories: "Forest" includes tropical and subtropical moist and dry broadleaf forests; "Wetlands" includes flooded grasslands and savannahs, and mangroves; "Grasslands" includes montane, tropical, and subtropical grasslands, savannahs, and shrublands; "Deserts" includes deserts and Xeric shrublands.

UNHCR recommends that camps be placed at least 50 km from the border with the refugee sending country to increase refugee protection from cross-border attacks (UNHCR, 2024b). The average distance (53.5 km) hits this target, but with considerable variation (SD = 54.4). Figure 6 displays the distribution of ARCD camp distance to the nearest internal border by region. We find that 17.1 percent of ARCD camps in Central Africa are within 10 km of the border and over one half (53.1%) are within 30 km of the nearest international border. While this is less characteristic of the other regions, a non-negligible share of East African (16.9%) and West African (19.2%) camps

Figure 6: Distribution of camp distances to nearest internal border by region



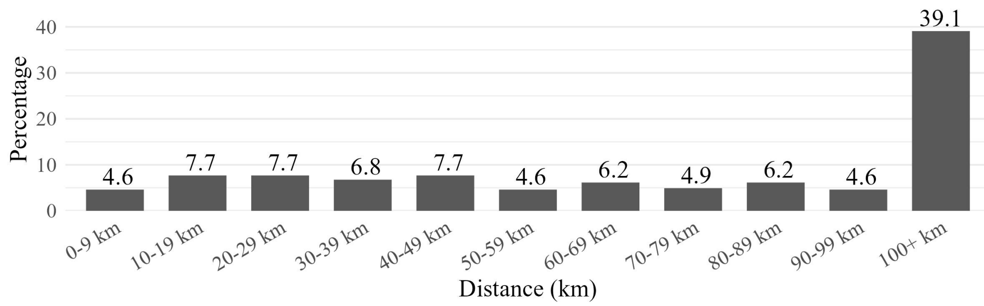
Notes: Camp geolocations based on ARCD and includes all ARCD entries (omitting subcamps) with latitude and longitude. Regional definitions are based on UN classification taxonomy. Southern Africa omitted from regional breakdown due to low count (n=2) but included in the sample used to derive the Total column. National borders restricted to internal borders, and distance measurements based on Euclidean distance.

are within 20 km of the nearest border.

As mentioned, cross-sample comparisons suggest that distance from capital cities (national or provincial) is not unique to camp locations, which speaks to the frequent remoteness in many large-area refugee hosting countries across SSA. But distance to the nearest capital also varies considerably within the sample. Figure 7 provides the overall distribution of ARCD distances to the nearest provincial or national capital. 20% of ARCD locations are within 30 km of the nearest capital city: these camps are predominately located in West Africa (Figure B3, Online Appendix). But across the sample, two out of five camps (39.1%) are 100 km or farther from the nearest capital city. Remoteness from capitals is a especially common for camps in Central and East Africa.

While border proximity speaks to constraints that prevent full compliance with UNHCR

Figure 7: 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.

guidelines, results for terrain ruggedness suggest that the use of relatively flat terrain for camps, another UN recommendation (UNHCR, 2007), is more likely to be realized. On average, camp locations are positioned on marginally flatter terrain than our comparison points near borders or comparison points across countries. This is a clear sign that despite some limitations on optimal siting, logistical considerations do play a role in camp location designation.

Examining bioclimatic areas, we note that the majority (75.7%) of sample camps are in grasslands areas, while a smaller share are located in (dry or wet) broadleaf forests (19.4%). By contrast, few camps are in deserts or wetlands, likely because of the associated logistical challenges related to water and infrastructure. The bioclimatic features of camps appear to match those of other points near the border. But they are statistically dissimilar to the average across host countries. In particular, camps are more likely to be in grasslands and less likely to be in desert areas in comparison to host country national averages.

Time-varying features from the period years prior to camp opening (Table 2) suggest that areas selected for camps exhibit important local and national divergences with respect to vegetation. On average, camp areas have low levels of forest cover (10.5%), with considerably more nonforest vegetative cover (62.2%) 2 years prior to camp opening. Relative to the country comparison sample, forest cover is marginally lower (2.2 percentage points, ppt.) and non-forest vegetation cover is substantially higher (19.5 ppt.). We observe deviations from the border comparison sample at the regional level. Areas later selected for camps in West and Central Africa have on average 10.0 and 3.6 ppt. less forest cover relative to this sample. In Central Africa, camps are sited in areas with significantly more non-forest vegetative cover (67%) relative to the reference border sample (58%), which corresponds to a a lower percentage of non-forest, non-vegetative cover (22.5% vs. 28.4%). Nightime light radiance results suggest this disparity in nonvegetative cover for camps in Central Africa is not explained by population density as proxied through nighttime light: both the ARCD and border comparison samples have equally low average radiances in this region.

Relative to national averages, areas later selected for camp siting have the same mean radiance as a random stratified national sample. The parity speaks to the very limited radiance across countries hosting refugee camps, as nighttime lights will only be high in limited urban land areas. But camp areas are notably unique relative to national averages in terms of their forest and vegetation attributes. As mentioned, areas later selected for camps in West and Central Africa have significantly less forest cover. Overall, and especially in Central and East Africa, camps have significantly more non-forest vegetation and less non-forest, non-vegetative cover relative to the weighted average across host countries.

4.2 Impacts on Forest and Vegetation Cover of Camp Interiors

Our causal assessment uses the panel of zonal statistics collected from the MODIS Terra Continuous Fields data on percent forest, non-forest vegetation, and non-vegetated cover intersecting each camp and comparison point 2000-2024. Our difference-in-difference (DID) approach examines a straightforward 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 in order to make way for the “camp footprint”: settlement blocks, latrines, fire breaks, and necessary infrastructure and buildings (UNHCR, 2005, 2007). We therefore expect camp opening to result in vegetation loss simply due to settlement logistics. But for our purposes, demonstrating this causal effect using the ARCD provides some additional legitimacy of the data,

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

	ARD			Border Comp. (BC)			ARD vs. BC (Δ)	Country Comp. (CC)			ARD vs. CC (Δ)
	Mean	SD	N	Mean	SD	N		Mean	SD	N	
Percent forested											
West	16.1	17.4	31	26.1	20.5	914	-10.0**	24.7	20.2	817	-8.6*
Cent.	10.8	10.5	71	14.4	19.7	6,866	-3.6**	18.0	26.4	5,783	-7.2***
East	8.5	9.0	77	9.4	12.1	10320	-1.0	8.2	12.1	8,816	0.3
Total	10.5	11.4	179	12.1	16.4	18,100	-1.5	12.7	20.0	15,416	-2.2*
Percent non-forest vegetation											
West	50.6	25.6	31	52.8	23.5	914	-2.3	53.8	25.3	817	-3.2
Cent.	67.0	18.7	71	57.5	25.9	6,866	9.5***	37.9	33.7	5,783	29.1***
East	61.4	24.7	77	56.4	27.2	10,320	5.0	45.2	32.4	8,816	16.2***
Total	62.2	23.1	179	56.7	26.6	18,100	5.5**	42.7	32.9	15,416	19.5***
Percent non-forest non-vegetated											
West	33.6	34.7	31	21.4	29.3	914	12.2	21.8	30.4	817	11.8
Cent.	22.5	22.3	71	28.4	29.8	6,866	-5.9*	44.3	42.6	5,783	-21.8***
East	30.4	28.7	77	34.5	31.9	10,320	-4.0	46.9	38.3	8,816	-16.4***
Total	27.6	27.4	179	31.5	31.2	18,100	-3.9	44.8	40.1	15,416	-17.2***
Mean radiance											
Cent.	0.012	0.063	29	0.016	0.243	3,249	-0.004	0.005	0.094	2,712	0.007
East	0.028	0.146	35	0.082	1.337	4,791	-0.055	0.005	0.112	4,260	0.022
Total	0.020	0.114	64	0.057	1.059	8,076	-0.037*	0.005	0.106	7,009	0.015

Notes: *, **, and *** indicate significance of 5, 1, and 0.1 percent levels. ARCD sample restricted to observations with (1) latitude and longitude listings, (2) opening date listings, and (3) main camps, removing sub-camps. Measurements derived using 1 km buffer. Comparison sample means are weighted based on the distribution of country-years present in the camp data subset to camp observation years at $t = -2$. Southern Africa omitted from regional sub-group analysis due to low camp count ($n = 2$) but not omitted from estimates for entire sample ("Total"). West Africa sub-group omitted from mean radiance sub-panel due to low counts of camps that opened after 2014 (two years after first VIIRS map year) but are included in estimates using entire sample. "Cent"=Central Africa.

as it suggests that, on average, the data are not geographically misclassifying camp locations or opening years.

Analytically, it is interesting for us to consider variation in the magnitude of these effects. For this reason, we examine heterogeneity by camp size, classifying camps based on their having a maximum population size of less than 1,000, 1,000 to less than 10,000, 10,000 to less than 30,000, and 30,000 or more persons. We also consider how the treatment effect changes based on the camp's region (East, Central, West) and the bioclimatic area (grasslands or dry/wet broadleaf forests). We report summary statistics overall and by camp characteristic (maximum population, bioclimatic area, and region) for the year prior to camp opening in Appendix Table C1. Sample means a year before camp opening provide a baseline mean to contextualize coefficient estimates.

4.2.1 Method

Our regressions use the primary camp geolocation and operational years, omitting sub-camps as in Section 4.1. We exclusively rely on the border comparison points as our control group, given their similarity to camp areas in terms of remoteness, geography, etc. The sample includes 325 camps and 2,911 border comparison points spanning the years 2000 to 2024. Within this sample, camps are active for an average of 11.4 years and inactive for 12.3 years.

Consider the following DID twoway fixed effects (TWFE) estimator

$$y_{it} = \beta_0 + \beta_1 Open_{it} + \gamma_i + \lambda_t + \varepsilon_{it} \tag{1}$$

Where y_{it} is the percent forest, nonforest vegetation, or nonforest, nonvegetative cover for the 1km buffer area around geopoint i year t. The independent variable of interest is $Open_{it}$, an indicator for whether the camp is operational in year t. This baseline specification includes a unit fixed effect (γ_i) that accounts for time-invariant point characteristics, such as terrain ruggedness, distance to cities, bioclimatic area, etc. We also include year fixed effects that capture time-varying factors across the sub-continent (λ_t). Our error terms are clustered at the host country level.

Rather than estimate Equation (1) with ordinary least squares (OLS), we use the estimator developed by de Chaisemartin and D'Haultfœuille (2024), which avoids the additional untestable assumption of treatment cohort effect homogeneity necessary when estimating two-way fixed effects DID models with OLS (Goodman-Bacon, 2021). And unlike other robust DID estimators for heterogeneous treatment effects, the approach introduced by de Chaisemartin and D'Haultfœuille

(2024) does not require treatment to be absorbing, meaning that camps may open and then close within the panel, as they do in practice.

The method developed by de Chaisemartin and D'Haultfœuille (2024) relies on the estimation of an event study, yielding placebo pre-treatment estimates as well as dynamic treatment effects post-treatment. The former provides a method of assessing evidence of the validity of the parallel trends assumption. Based on the post-treatment average treatment effects, the researcher can also recover the cumulative average treatment effect (CATE), which is the total effect of the treatment over the designated post-treatment period. As part of the estimation strategy, we designate seven placebo years and seven post-treatment years. We assume no anticipation, given the fact operationally, the designation and preparation of a new camp occurs after the onset of a population influx.[13]

4.2.2 Results

Table 3 provides the average treatment effect estimates using all camps as well as heterogeneity analysis by camp maximum population size. Across all camps in the sample, camp opening results in a 1.7 ppt. reduction in forest cover over the seven-year assessment period. These reductions occur within the first two years of camp opening and persist until the end of the event study period (Figure 8). The reduction represents a 16.2% percent decrease relative to the sample mean on the eve of camp designation (10.5%). The event study results offer support of parallel pre-trends, with null placebo estimations and a sudden decline in percent forest cover post-treatment (Figure 8).

The results show that deforested land transitions to having no vegetation: non-vegetated cover increases by 2.4 ppt. This sudden, but non-continuous decline in forest (and conversely, the sudden, but not ongoing increase in nonvegetative cover) suggests that trees are removed to make way for the camp footprint. It is possible that the forest loss is also related to biofuel harvesting. But if fuelwood and charcoal consumption were the primary driver, we would expect the tree losses to be monotonically decreasing until the area is fully cleared.

[13]The UNHCR Handbook for Emergencies explicitly discourages planners from preparing new encampments before populations in need make the decision to cross the border. The organization claims that "an unoccupied, developed site may send the wrong signal and encourage people to cross the border" (UNHCR, 2007).

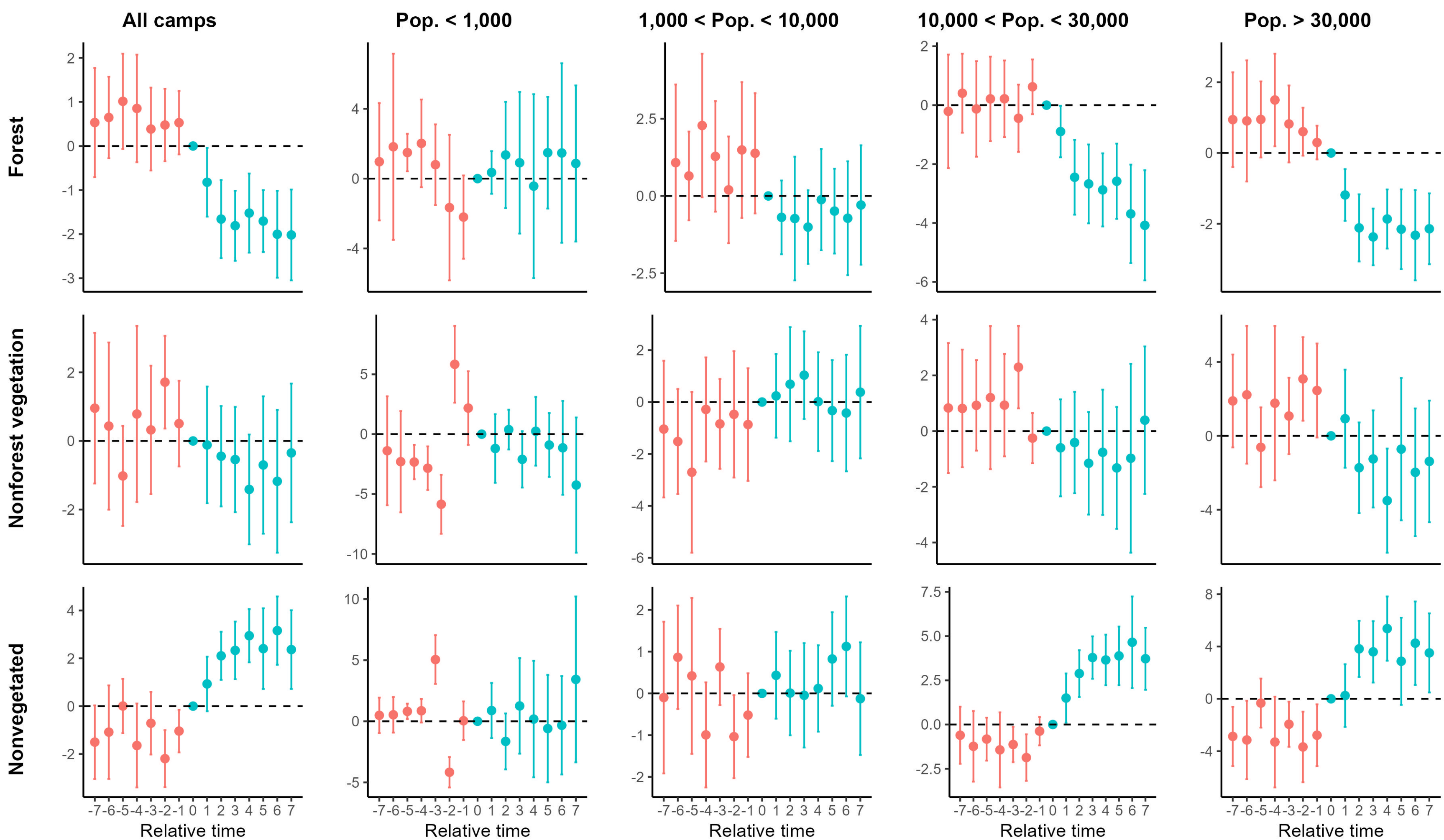
We fail to detect a significant change in non-forest vegetation. Point estimates on non-forest vegetation are not only statistically insignificant, they are also small in magnitude, particularly in relation to the pre-encampment mean (61.9%). There are several reasons why non-forest vegetation may be unaffected by the camp opening. First, the small size of shelters within camps may allow for construction around existing vegetation. Another possibility is that the vegetation is unsuitable for energy production. Ground-level non-tree vegetation such as grass burns quickly and is not well suited for cooking or heating using traditional technologies.

Table 3: Cumulative average treatment effects of camp opening on percent forest, vegetation, and non-vegetated cover by camp maximum population size

	All camps	Pop. < 1k	Pop. $\in [1k, 10k)$	Pop. $\in [10k, 30k)$	Pop. ≥ 30k
	(1)	(2)	(3)	(4)	(5)
Forest (%)	-1.727***	0.945	-0.663	-2.778***	-2.038***
	(0.408)	(1.971)	(0.816)	(0.649)	(0.425)
Vegetation (%)	-0.707	-1.378	0.269	-0.711	-1.367
	(0.861)	(1.451)	(0.987)	(0.998)	(1.423)
Nonvegetated (%)	2.427***	0.468	0.374	3.488***	3.390**
	(0.618)	(2.125)	(0.503)	(0.617)	(1.244)
N	74,178	55,414	71,024	71,022	68,039

Notes: *, **, and *** indicate significance at the 5, 1, and 0.1 percent levels, respectively. Results are from the estimation of Equation 1 using the de Chaisemartin & D'Haultfœuille (2024) difference-in-differences estimator for multiple group-times (`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 $N = 325$. The control group is drawn from areas 30–50 km from the nearest camp geopoint outside protected areas. Random sampling of control group points is stratified by country representation in the ARCD data: $N = 2,911$. Vegetation and forest data are sourced from MODIS Terra Continuous Fields 250m annual data, 2000–2024. Zonal statistics are derived using a 1 km buffer around each treatment and control geopoint. Each coefficient represents the estimated effect on percent forest, percent nonforest vegetation ("Vegetation"), or percent nonforest nonvegetated cover ("Nonvegetated"). Columns 2–5 report estimations based on camps with at least one year of population data, classified by maximum population size. Sample sizes (N) do not add up to 25 observation years by number of observations because our estimations use 7 placebo ("pre") years and 7 post-treatment years; across camps, the available 14-year span varies (e.g., a camp opening in 2003 contributes only 3 placebo years).

Figure 8: Event study estimates of the impact of camp opening on forest and vegetation change in camp interiors by camp maximum population size



Notes: The event study corresponds to the cumulative average treatment effect results reported in Figure 8. Results are estimated from Equation (1) 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) (N = 325). 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 = 2,911). Vegetation and forest data are from MODIS Terra Continuous Fields. Zonal statistics are derived using a 1 km buffer around each treatment and control geopoint. Population ("Pop.") sizes reflect the maximum population size of the camp based on available ARCD data.

Forest losses occur within encampments with a maximum population of 10,000 or more (Table 3 and Figure 8). For these camps, we measure a statistically significant 2.0 to 2.8 ppt. decrease in forest cover. The findings likely reflect congestion in refugee camps with larger populations. Unless the encampment is designed as an agricultural settlement, the land allocated for an encampment is often modest in area. UNHCR strongly encourages planners to maintain existing natural ground cover when possible (UNHCR, 2023). But during a displacement emergency, a camp may need to absorb more people than it was designed to accommodate, and unless more land can be allocated, areas must be cleared of trees to make way for additional households. Across all camp sizes, however, we again fail to detect any changes in non-forest vegetation. Even if the coefficient estimates were significant, they are economically insignificant relative to the sample mean. Instead, forest loss again makes way for non-forest non-vegetation cover, which likely represents the camp footprint.

We also estimate the cumulative average treatment effect and event studies by biome (grasslands or moist/dryleaf forest) and region (West, Central, and East Africa). We report the cumulative average treatment effect results in Table 4 and the event studies in Appendix Figure B4. Forest losses are significantly higher in forest bioclimatic regions, where forest cover a year prior to camp opening is, as expected, higher than in the overall sample (21.9 percent). In these regions, camp opening leads to a cumulative reduction of 3.5 ppt. on average, which represents a 16.0% reduction relative to the mean. In grasslands biomes we also observe at 1.5 ppt. reduction in forest cover. Given pre-encampment percentage forest cover in grasslands areas of 8.1 percent, this constitutes a marginally higher percentage loss relative to the sample mean (18.7 percent). In both biomes, forest loss yet again gives way to nonvegetative cover, and non-forest vegetation appears largely unchanged.

Regionally, the ppt. reduction is largest among camps in West Africa (-3.2 ppt.) followed by Central Africa (-2.3 ppt) and East Africa (-1.7 ppt). The variation of effects across regions largely reflects differences in baseline forest cover levels (Table C1, Online Appendix), with West African camps exhibiting the highest forest cover one year prior to camp opening (14.5%), followed by Central (11.6%) and East Africa (7.9%).

Table 4: Cumulative average treatment effects of camp opening on percent forest, vegetation, and non-vegetated cover by region and bioclimatic area

	Grasslands biome (1)	Wet/dryleaf forest biome (2)	West Africa (3)	Central Africa (4)	East Africa (5)
Forest (%)	-1.507***	-3.481***	-3.154***	-2.326***	-1.727***
	(0.445)	(0.914)	(0.488)	(0.370)	(0.408)
Vegetation (%)	-1.079	0.888	0.448	-0.386	-0.707
	(1.036)	(1.010)	(1.190)	(1.012)	(0.861)
Nonvegetated (%)	2.577***	2.591***	2.704**	2.693***	2.427***
	(0.732)	(0.478)	(0.967)	(0.707)	(0.618)
N	53,043	15,565	9,659	25,722	74,178

Notes: *, **, and *** indicate significance at the 5, 1, and 0.1 percent levels, respectively. Results are from the estimation of Equation (1) 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 $N = 325$. The control group is drawn from areas 30–50 km from the nearest camp geopoint outside protected areas. Random sampling of control group points is stratified by country representation in the ARCD data ($N = 2,911$). Vegetation and forest data are sourced from MODIS Terra. Zonal statistics are derived using a 1 km buffer around each treatment and control geopoint. Each coefficient represents the estimated effect on percent forest, percent nonforest vegetation ("Vegetation"), or percent nonforest nonvegetated cover ("Nonvegetated"). Column 1 is restricted to sample observations (treatment and control) within the biome "Tropical & Subtropical Grasslands, Savannas & Shrublands." Column 2 is restricted to sample observations within either the "Tropical & Subtropical Moist Broadleaf Forests" or the "Tropical & Subtropical Dry Broadleaf Forests" biomes (the latter being rare; this category largely represents rainforest areas). Columns 3–5 separately estimate effects by region of sub-Saharan Africa. Southern Africa is omitted due to low sample count. Sample sizes (N) do not add up to 25 observation years by number of observations because our estimations use 7 placebo ("pre") years and 7 post-treatment years; across camps, the available 14-year span varies (e.g., a camp opening in 2003 contributes only 3 placebo years).

Taken together, the results suggest that across the sub-continent, encampment leads to a statistically and economically significant decrease in forest cover within the camp boundary. This reduction likely speaks to land clearing for logistical reasons, such as settlement blocks, fire breaks, service centers, 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, the largest camps, in terms of population, are associated with the greatest amount of forest loss. This likely reflects the higher population densities of these camps and the need to transition more land from forest cover for logistical purposes.

5 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 exclusively 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 summarizing refugee camps 1999-2024 and addressing questions concerning the timing and location of camps in the first quarter of the century, their sizes, patterns of longevity, and the important historical events that predicated their creation. We additionally offer two illustrative applications to demonstrate potential uses of the ARCD. First, our geospatial dsecriptive assessment shows that areas selected for camps align with some logistical guidelines yet often remain far from provincial capitals, close to borders, and near protected areas. The tendency of keeping camps near border areas appears to account for many of the camp site characteristics that diverge from national averages (such as protected area proximity or higher probability of being in a grasslands bioclimatic areas). But in West and Central Africa, forest and vegetative cover statistically differs from other border regions, reflecting local selection on the basis of logistics or land value.

Second, our causal assessment of encampment and its impacts on the forest and vegetation cover of the camp interiors reveals significant reductions in forest cover. All forest losses occur within the first 1-2 years of camp opening, which is suggestive of land clearing to make way for shelter, roads, schools, and other infrastructural and practical needs. We fail to demonstrate a causal impact of camp openings on vegetation cover, suggesting that camps can be built around existing vegetation and that this vegetation may not be a suitable fuel source. Importantly, the impacts are largest for camps with the largest populations, which likely speaks to higher population densities and less flexibility with siting shelters and infrastructure around existing foliage.

There are numerous opportunities to build on this work. It is possible to expand the dataset to include other known precise sites where refugees are located on the continent, creating opportunities to study other dimensions of refugee hosting beyond the encampment modality. This expansion could 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. Archival data collection could also facilitate greater temporal scope, examining camp operational years prior to 1999. And there are many camp-level variables of interest that could be added to the ARCD related to aid provision, demographics of the refugee population, inter-camp policy variation, etc. 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.

6 Bibliography

References

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

Bjørkhaug, I. (2020). Revisiting the refugee–host relationship in Nakivale Refugee Settlement: A dialogue with the oxford refugee studies centre. *Journal on Migration and Human Security*, *8*(3), 266–281.

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.

DiMiceli, C., Sohlberg, R., & Townshend, J. (2022). Modis/terra vegetation continuous fields yearly l3 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

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

Filipski, M. J., Rosenbach, G., Tiburcio, E., Dorosh, P., & Hoddinott, J. (2021). Refugees who mean business: Economic activities in and around the rohingya settlements the rohingya settlements in bangladesh. *Journal of Refugee Studies*, *34*(1), 1202–1242.

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.

Harrell-Bond, B. E. (1986). *Imposing aid: Emergency assistance to refugees*. Oxford University Press.

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

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., & Vandercasteelen, 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*.

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

Ngayimpenda, 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

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. (2023). *The green companion guide*. UNHCR.

UNHCR. (2024a). *Mahama refugee camp factsheet* (Report). UNHCR. https://data.unhcr.org/en/documents/details/108899

UNHCR. (2024b). *Safe and secure settlements (Emergency Handbook)*. UNHCR.

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

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

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

Supplemental 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 extract information from the following Statistical Yearbooks to create the initial camp listing:

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-overiew

[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

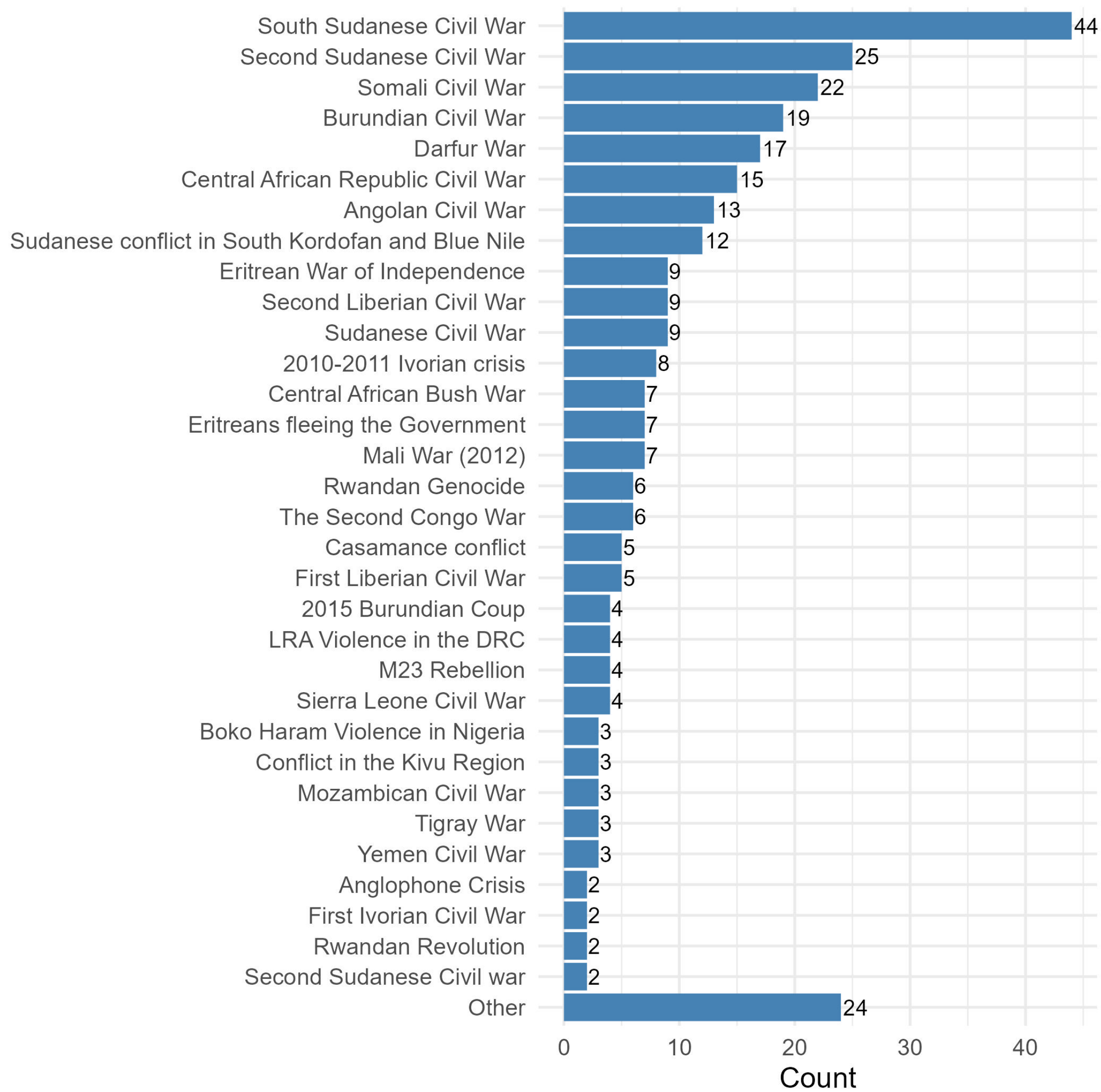
[16] UNHCR. (2015). *UNHCR statistical yearbook 2014, 14th edition* [Statistical Yearbook]. United Nations High Commissioner for Refugees (UNHCR). https://www.unhcr.org/publications/unhcr-statistical-yearbook-2014-14th-edition

[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

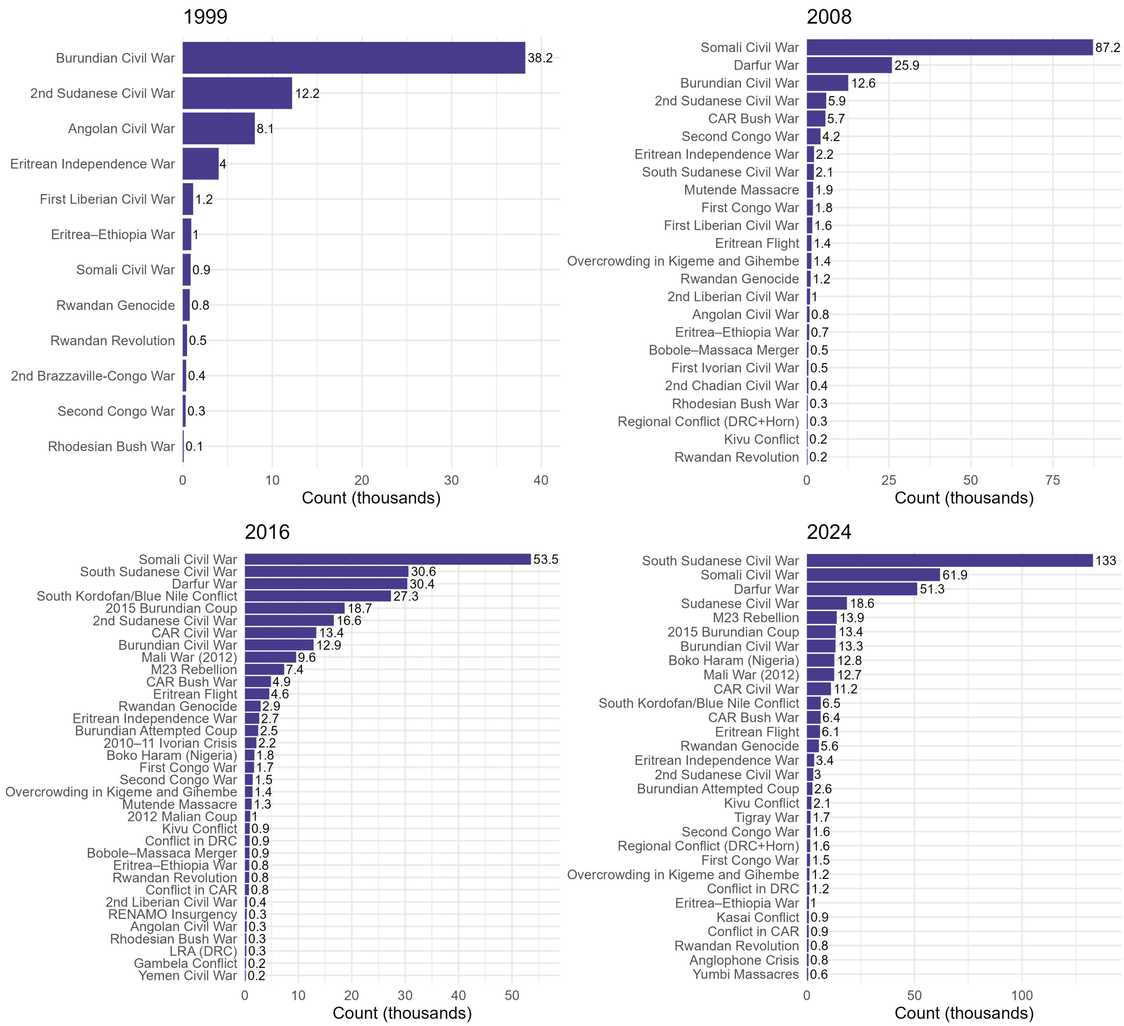
B Additional Figures

Figure B1: Number of camps by precipitating conflict event, ARCD camps that opened 1999 or later



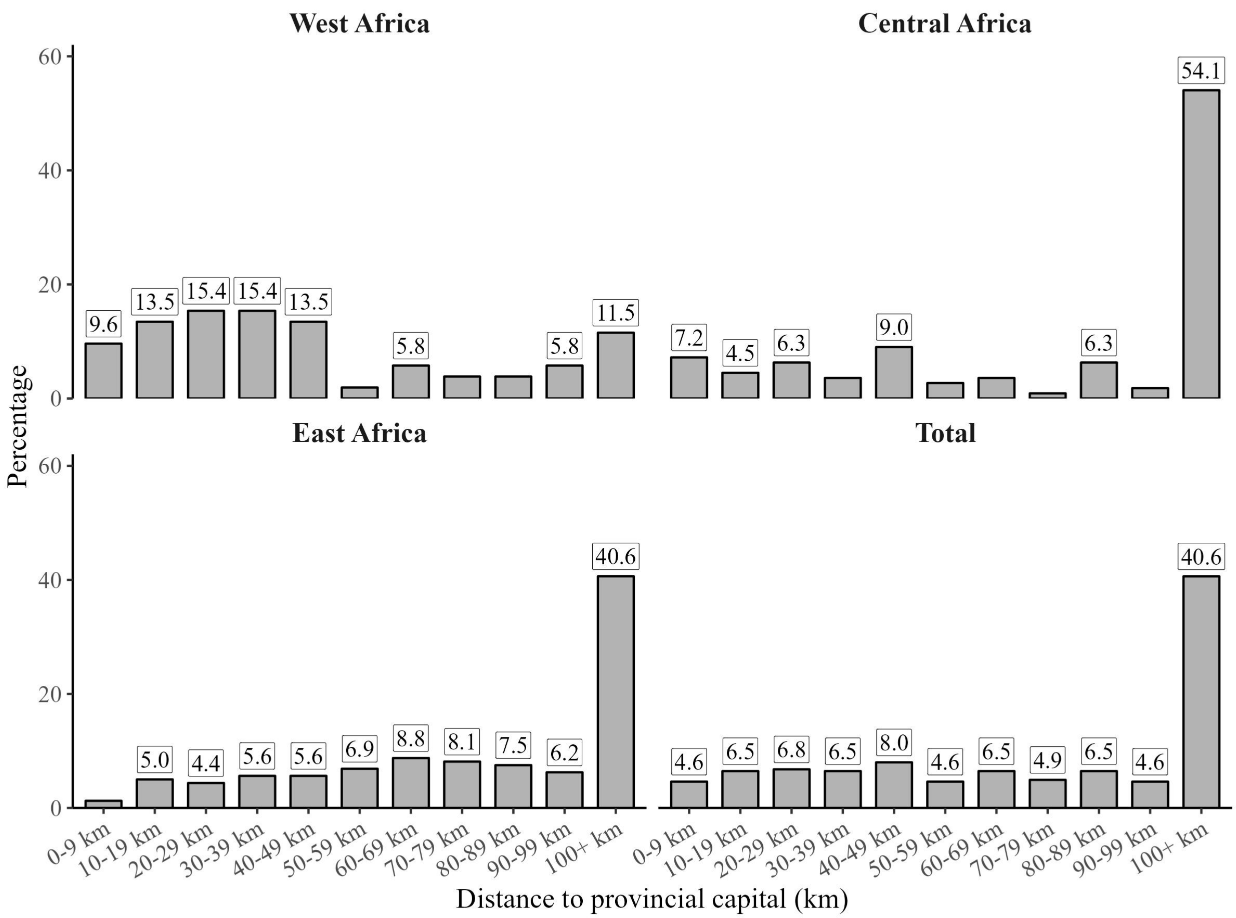
Notes: Counts from ARCD data, including camps and subcamps, with opening date of 1999 or later. "Other" category includes all cases in which a precipitating event was only associated with one ARCD entry opening.

Figure B2: Encamped population count (thousands) by precipitating event and year



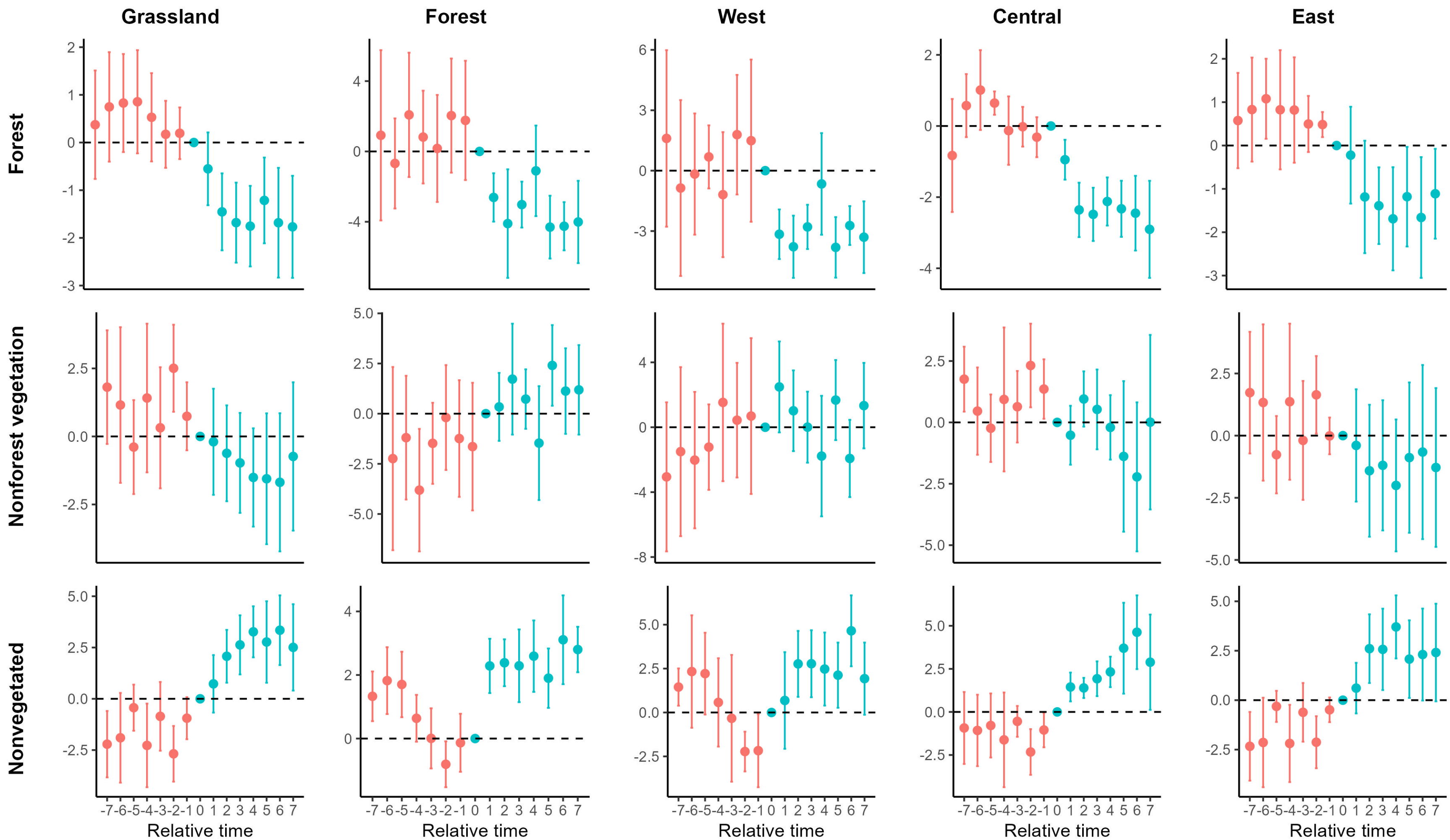
Notes: UNHCR Data from ARCD camps (including sub-camps) with non-missing population counts in reference years. Count represents the number of refugees accommodated in an ARCD camp in the reference year due to the precipitating event that occurred the same year or before.

Figure B3: 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. Southern Africa omitted from regional breakdown due to low count (n=2) but included in the sample used to derive the Total column. Provincial capital locations identified using Geonames Gazetteer.

Figure B4: 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 4 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 $N = 325$. The control group is drawn from areas 30–50 km from the nearest camp geopoint outside protected areas. Random sampling of control points is stratified by country representation in the ARCD data ($N = 2{,}911$). 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 geopoint. "Grasslands" = *Tropical & Subtropical Grasslands, Savannas & Shrublands*. "Forest" = *Tropical & Subtropical Moist Broadleaf Forests* or the *Tropical & Subtropical Dry Broadleaf Forests*. Regions based on UN classification.

C Additional Tables

Table C1: Vegetation Descriptive Statistics by Treatment Status for Entire Sample and for Treated Sample in $t-1$ by Region, Camp Population, and Biome)

Panel A: Treatment and Control Sample, Camps in $t-1$ by Region

			Year prior to camp opening			
	Camps	Control	Camps	West Africa	Central Africa	East Africa
Forest	9.26	16.38	10.49	14.45	11.61	7.85
Nonforest Veg.	61.88	59.96	62.14	55.42	66.17	61.13
Nonforest Nonveg.	29.19	23.99	27.70	30.46	22.55	31.33
N	6,450	72,725	179	31	71	77

Panel B: Camps in $t-1$ by Maximum Population Size and Biome

	Year prior to camp opening					
	Pop $< 1{,}000$	Pop $\in [1k, 10k)$	Pop $\in [10k, 30k)$	Pop $\geq 30k$	Grasslands biome	Forest biome
Forest	13.22	15.58	10.28	6.33	8.09	21.87
Nonforest Veg.	70.80	65.04	65.39	53.47	63.85	68.95
Nonforest Nonveg.	16.35	19.72	24.67	40.49	28.40	9.55
N	5	53	58	53	134	35

Notes: ARCD sample restricted to observations with latitude and longitude listings and to main camps, removing sub-camps. Bioclimatic regions aggregate finer-resolution categories: "Forest" includes tropical and subtropical moist and dry broadleaf forests; "Wetlands" includes flooded grasslands and savannahs, and mangroves; "Grasslands" includes montane, tropical, and subtropical grasslands, savannahs, and shrublands; "Deserts" includes deserts and xeric shrublands. Regional definitions are based on UN classification taxonomy. Zonal statistics from MODIS Terra Continuous Fields data derived using 1 km buffer. Population counts represent maximum population recorded in ARCD during camp's operational years.