Introducing the African Refugee Dataset 1999-2016

Sebastian Anti[1], Colette Salemi[2], and Kaitlyn Wilson[3]

Department of Applied Economics, University of Minnesota - Twin Cities

February 2, 2020

ABSTRACT: To date, there has been no publicly available systematic data collection and consolidation of refugee camp locations and periods of operation on a broad scale. Here, we introduce the African Refugee Dataset (ARD), the first geospatial dataset of this kind. We present and describe the creation of the dataset. We then explore the characteristics of the refugee camp locations by matching the locations with other geospatial datasets. We show the location of refugee camps are, on average, more rugged and dryer than other areas in their general region. They are also closer to population centers and closer to conflict zones. Taken together, this reflects the prior qualitative understanding about the characteristics of the locations of refugee camps and provides confidence in the new dataset.

Keywords: Refugees, Forced Migration, Data, Geospatial

[1] Corresponding author. Affiliation: Department of Applied Economics, University of Minnesota – Twin Cities. Address: Sebastian Anti, University of Minnesota – Twin Cities, Dept. of Applied Economics, Ruttan Hall, Office 252, 1994 Buford Avenue, St. Paul, MN 55108, Phone: 413.362.2282, Email: sebastian.anti@gmail.com. Funding statement: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Declarations of interest: None.

[2] Affiliation: Department of Applied Economics, University of Minnesota – Twin Cities. Address: Colette Salemi, University of Minnesota – Twin Cities, Dept. of Applied Economics, Ruttan Hall, Office 213, 1994 Buford Avenue, St. Paul, MN 55108, Email: salem043@umn.edu.Funding statement: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Declarations of interest: None.

[3] Affiliation: College of Liberal Arts and Sciences, University of Minnesota – Twin Cities. Funding statement: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Declarations of interest: None.

SECTION 1: INTRODUCTION

The global count of forcibly displaced people is at an all-time high. The United Nations High Commissioner for Refugees (UNHCR) currently tallies 29.4 million refugees in the world,[4] a figure that greatly surpasses the previous peaks in refugee population counts from the early 1990s.[5] Although the global refugee population has been growing rapidly since 2012, the severity of forced displacement today came to the forefront of global discourse in 2015, when asylum-seeking groups began arriving in Europe in large numbers through unofficial channels. As wealthier countries responded to unauthorized asylum-seeker inflows, a political and economic debate emerged that shares many of the characteristics of the scholarly discussion on voluntary migration (Borjas 1987; Card 1990). At the heart of this debate is the question of whether refugees have a negative effect on the people and places where they reside, or (to borrow from frequently invoked tropes) whether refugee population influxes represent a "blessing" or a "burden", a "crisis" or an "opportunity".

Despite the recent media attention on refugee migration to Europe, the majority of refugees today reside in low- or middle-income countries close to their country of origin. In the case of Africa, which currently hosts about one-quarter of the world's refugees, the refugee population has grown tremendously in the past 15 years, from a count of less than 5 million to 21.3 million people in 2019 (Maystadt et al. 2000). In order to enhance our understanding of the impacts of this forced displacement for African countries, we have built the Africa Refugee Dataset (ARD), a dataset that identifies 423 refugee camps in sub-Saharan Africa (SSA) between 1999 and 2016 and provides their geographic coordinates and years of operation. In this paper, we introduce and

[4] We define a refugee as a person who has fled their home country due to fear of persecution on the basis of race, ethnicity, sex, membership of a particular social group, or political opinion. The count of 29.4 million persons does not distinguish between those who received protected status from UNHCR (refugees) and those who have not yet obtained this status (asylum-seekers) (UNHCR 2019).

[5] UNHCR data counted 17-18 million refugees in the years 1991-1993. This count does not distinguish between refugees and asylum-seekers and represents a previous peak in refugee numbers (UNHCR 2019).

explore this dataset. We first provide an overview of our approach to constructing the data. Using a diverse set of sources and primarily drawing on the UNHCR Statistical Yearbooks and the UNHCR's People of Concern mapping platform, we identify the geo-locations and years of operation for refugee camps for the years 2000 to 2016. Because of inconsistencies and missing information in these datasets, we cross-validate entries using sources from newspapers, peer reviewed papers, and NGO reports. Finally, we explore the economic and environmental characteristics of the refugee camp locations using available GIS data. We are particularly interested in seeing whether these geographic locations are significantly dissimilar from characteristics of the broader region in which we find refugee camps. Such dissimilarities would speak to the various processes that influence the location of refugee residence in the host country, a topic that remains under-examined.

We believe that the ARD will enable empirical studies of refugee camps that address current gaps in the literature on forced displacement at a time when the topic is receiving increasing attention. For decades, social scientists have contributed to our understanding of forced displacement under the subject area of Refugees and Forced Migration Studies, but this work has been predominately qualitative (Fiddian-Qasmiyeh et al. 2014). There have however been several important quantitative efforts over the past years (Verme and Schuettler 2019). A considerable share of this work focuses on the impact of refugees in African host countries, especially in terms of wages, employment, prices, and economic activity (Alix-Garcia et al. 2018; Ruiz and Vargas-Silva 2016, 2015), consumption, prices, and wellbeing (Alix-Garcia and Saah 2010; Kreibaum 2016; Maystadt and Duranton 2019; Maystadt and Verwimp 2014), as well as health, nutrition, environmental, and education outcomes (Baez 2011; Maystadt et al. 2020).

The recent scholarship exhibits strong empirical rigor, but the external validity of the findings remains somewhat in question.[6] With few exceptions (Maystadt et al. 2020), almost all peer reviewed papers have been case studies looking at a particular country or a particular camp. For Africa, the host country scope has been primarily limited to Tanzania, Kenya, Uganda, and Rwanda (Verme and Schuettler 2019). While these countries have hosted large refugee populations in recent decades, they represent only a small share of the 35 countries that had at least one refugee camp between 2000 and 2016 on the continent. Moreover, these case studies examine not only recent displacement-related outcomes, but also include cases as early as the 1990s. Considering the evolution of refugee response planning, it is unclear whether results are comparable over time.[7]

Refugee camp locations are not exogenous, as considerations related to land availability, proximity to nearest border (Maystadt and Verwimp 2014), and prospective host community benefits heavily inform what land host governments choose for camp creation. Our exploration of the ARD for this paper provides some important descriptive insights with corresponding t-tests that interrogate the unique characteristics of refugee camp locations as compared to a comparison sample. In terms of geography and environment, we find that refugee camp locations tend to be more rugged and receive less rainfall. While they are usually close to borders, refugee camps are also relatively closer to an urban center. As expected, the distance between a camp and a conflict hotspot is significantly shorter than the distance between the comparison points and the nearest conflict.

[6] For thorough reviews of the literature on the impacts of forced displacement, see Maystadt et al. 2000, Ruiz and Vargas-Silva 2013, and Verme and Schuettler 2019. Verwimp and Maystadt (2015) also provides a contextual overview on refugees and host communities in SSA.

[7] In particular, we have seen an increase in humanitarian investments in host communities, often referred to as "resilience" programming. We also find more emphasis on "inclusive" programs that foster refugee and host sharing of assistance-related resources, such as boreholes, schools, and medical clinics, so that the local population stands to gain from hosting.

We believe that the ARD will strengthen the discourse surrounding refugees and the impacts of forced displacement across international borders by allowing for stronger examinations into heterogeneous outcomes and the mechanisms behind these discrepancies. The remainder of our paper is as follows. In Section 2, we outline the methodological approach to building our dataset. Section 3 describes the characteristics of the resulting dataset and how it aligns with known historical trends in SSA. Section 4 compares the characteristics of these camps to a random selection of locations within the general regions of camps, which provides us with a glimpse of how camp locations compare to a random selection process in the same general region. We share our concluding remarks in Section 5.

SECTION 2: AFRICAN REFUGEE DATASET COLLECTION

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 contradict each other routinely. This section outlines the process through which we built the ARD and how we prioritized information and made decisions over which information was the most reliable.

The dataset contains four major components: (1) the list of camps, (2) the years the camps were operational, (3) the geolocational coordinates (longitude and latitude) of each camp, and (4) decisions over which information is prioritized in the presence of contradictions. We discuss the development of each below.

2.1 List of Camps

We extracted the list of refugee camps from three sources: the UNHCR's Statistical Yearbooks from 1999 to 2016 (UNHCR 2002a, 2013, 2010, 2002b, 2004, 2003, 2012, 2005, 2000, 2015a,

2006, 2016, 2008, 2007, 2009, 2014, 2011), The UNHCR's Persons of Concern (POC) Map from 2018 (UNHCR 2018), and UNHCR operational maps published sporadically from 2000 to 2016, and now hosted on UNHCR's *refworld.org* website. We first collected every location listed as having a refugee camp in SSA in the Statistical Yearbooks from 1999 to 2016. 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, informal settlement, or some other type of location). Although extremely informative, there is reason to view these yearbooks as incomplete, since they are the result of data gathering within a relatively decentralized administrative system over humanitarian and emergency response in which protecting a vulnerable population (not the collection of data) is the primary objective. To address this, we supplement this information with several other sources.

We cross-referenced the list that we derived from the Statistical Yearbooks with the sites listed in the POC Map. The POC map was published by UNHCR in 2018 and documents all locations where people of concern were residing in that year. We added locations in the POC Map that were not mentioned in the Statistical Yearbooks for us to check later.

We then listed all the locations noted as refugee camps in the UNHCR's operational maps. The operational maps are a large collection of maps the UNHCR has published since the early 2000s. They are an extremely rich source of data on the locations and timing of camps over this period, although these maps taken collectively are not a complete record of camps over this period since they do not systematically cover the entire region or time period of interest. For example, maps exist for Ghana in 2002, 2005-2008, stop for several years, and then become available again in 2011 and in 2014. Ivory Coast by comparison has maps for every year from 2007 until 2013

except for 2012, but no maps before or after that period. Using these sources, we again added any locations that were not contained in the first two sources.

In this process, we noticed that although the Statistical Yearbooks appear to be the most complete, they often mislabeled locations as camps when in reality the entry was for a general administrative area. For example, Montserrado is mentioned as a refugee camp in Liberia over a long stretch of time in the Statistical Yearbooks. But research into an actual camp by this name using the UNHCR operational maps and news reports revealed no specific camp location by that name. Instead, we found that a county in Liberia by that name did receive a large refugee influx of refugees at some point. Given the absence of primary or secondary evidence of a camp, we believe that the refugee population in Montserrado was never organized in a camp and that the Statistical Yearbooks mislabeled this out-of-camp population as the residents of a camp. We identified these types of entries and removed them from the analysis because we could not confirm that a UNHCR-affiliated camp actually existed in these cases.

2.2 Times of Operation

We determined the years the camps were operational using the following strategy. First, if the camp is documented in the Statistical Yearbooks, we simply used the year the entry first appeared in the yearbooks and the year the entry last appeared in the yearbooks. Early Statistical Yearbooks notably display more inconsistencies and mislabeling so we are worried that the UNHCR's data gathering process was less reliable in its early years. If the entry was from an early statistical yearbook, around 1999 to 2003, we checked secondary sources such as news reports, academic papers, the World Food Program's (WFP) Refugee Geonode dataset, and UNHCR reports to determine when the camp opened, out of concern that the camp had been in operation before the data gathering process for the Statistical Yearbooks.

For those locations found in the POC map or UNHCR operational maps that were not identified in the UNHCR Statistical Yearbooks, we relied entirely on secondary sources to date the opening and closure of the camps. After close inspection, we chose not to use the POC maps to validate years of camp opening and closing. POC dataset contains start dates for its entries, and we confirmed with the UNHCR that these dates reflect the year the location entered into the UNHCR's administrative system. Upon further inspection, we found this data to be highly unreliable. Spot-checking revealed that for some countries the dates were relatively accurate (for example, Cameroon and Ethiopia), and totally inaccurate for others. As a result, we decided not to use any date information from the POC dataset.

Instead, we use a variety of sources to validate camp opening and closing dates. Many of these were helpful publications from the UNHCR such as "Camp Reports", which are short published reports updating stakeholders on the status of a camp. Often these reports include a brief section on the history of the camp including its opening date. Other useful resources include media reporting and NGO reporting on various events of interest in relation to a camp. All sources that we consulted in order to determine the best approximate starting and ending dates for the camps are listed in the ARD data file.

2.3 Spatial Data

Spatial data for ARD came from three primary sources. First, the POC map contains the locations of camps that were active in 2018. For any camps operating between 1999 and 2016 still active in 2018, we took the geographic data from this resource. Second, the World Food Program (WFP) Refugee Geonode dataset also contains longitude and latitude for a number of the camps we identified, and this allowed us to fill in much of the spatial information excluded from the POC

map. Still, the WFP data is not a comprehensive resource in that it does not contain many of the camp locations in the Statistical Yearbooks or in the POC map.

For the remaining camps, we used Google Earth and *geonames.com* searches to locate the camps based on their name. Some camps (such as Nyarugusu in Tanzania) are specifically geoidentified in Google Earth. For the camps that Google Earth does not uniquely recognize, we assume that the camp name originates from the name of the adjacent village. We acknowledge that this approach is imperfect, as a camp's closest village may actually be very distant, or the camp is named after a physical feature, such as a river that shares its name with a distant village. To reduce this type of measurement error, we attempted to verify the locations of these camps using the UNHCR operational maps previously mentioned, or some other resource that indicated the camp's position. An example of this would be the "Camp Reports" we mention above, which often have a small map indicating the location of the camp. For any camp that still had an unidentified geolocation after this process, we used a non-UN source to locate, such as a media report, with a map. These sources are clearly chronicled in the ARD data.

2.4 Cleaning and Prioritization Decisions

As mentioned, the underlying sources for this dataset reflect an operational environment that is uncertain and fluid. The gaps and inconsistencies we found also speak to an institutional structure for data-gathering that is highly decentralized. 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. To address the mislabeling of camps in the dataset, we carefully examined the information in each entry to verify that there was indeed a camp corresponding to the entry. We commonly found entries listed as camps that, upon further inspection, actually corresponded with a regional administrative unit, as in the aforementioned

Montserrado example. Another entry error we frequently found was the labeling of a transit center as a camp. Transit centers are a different type of refugee location[8] and are not the focus of this current data collection effort, so we remove these entries from the dataset.[9]

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. If there is no other source for the opening and closing date of a camp, we use the information derived from the UNHCR Statistical Yearbooks.

SECTION 3: AFRICAN REFUGEE DATASET ATTRIBUTES

Overall, the dataset contains precise geolocations and years of operation for 423 refugee camps active between 1999 and 2016. The three countries with the greatest number of refugee camps were DRC (52 camps), Ethiopia (49 camps), and Sudan (46 camps).

Figure 1 shows the aggregate trends in the number of refugee camps in the region over the period covered by the ARD. Overall, it is striking that there was never less than 130 operational refugee camps in the region over this time. But the trend also reveals two distinct peaks. The highest refugee camp counts were in 2002, coinciding with the later years of the Liberian Civil War (ended in 2003), Sierra Leone Civil War (ended in 2002) and the Second Congo War (ended in 2003). The other peak year was 2014, which coincides to an escalation in Boko Haram's militant

[8] In many countries, refugees are held in transit centers after crossing the border and declaring asylum while longer-term accommodations are arranged.
[9] Future versions of the ARD will include other refugee locations beyond camps: transit centers would be included in such a later version.

activity in Nigeria starting in 2008 and the civil war in the Central African Republic (CAR), starting in 2012.

Figure 2 displays all the camps operational between 1999 and 2006 that appear in the ARD over the entire region of SSA. The geographic distribution displayed by this map corresponds with our knowledge of conflict in the region over the time period covered in the data. In particular, we note a heavy concentration of camps in the Great Lakes region, which includes Rwanda, Burundi, Uganda, and the eastern area of the Democratic Republic of Congo (DRC). We also see a heavy concentration of camps in Liberia and Sierra Leone in the data.

Figure 3 shows the camps in the region active in the years 1999, 2004, 2009, and 2016 respectively. This provides a sense of how the distribution and count of camps is changing over time. Figure 3 also corresponds with our understanding of the timing of conflict events in SSA. The heavy concentration in Liberia and Sierra Leone is present in the early 2000s, during a period of civil war in both countries. By 2016, the number of refugee camps in this area of West Africa notably declined. This trend is similar to that of Eastern Chad, which has no refugee camps prior to the start of the War in Darfur, but then shows a heavy concentration after the start of that conflict in 2003. Similarly, countries in the Sahel region, such as Niger, Burkina Faso, and Cameroon, have little camp activity in the early years of our study period. But in 2016, we find considerable numbers of refugee camps. These more recent camps were likely opened to accommodate refugees fleeing civil war in the CAR, as well as violence in Nigeria and deteriorating conditions in Mali.

It is immediately apparent that the camps are concentrated around borders of various countries, which corresponds to past observations from Maystadt and Verwimp (2014). This is probably the result of several driving factors depending on the country. In relatively stronger states capable of orchestrating the refugee response, host governments often select locations within close

proximity to the border with the refugee sending country in order to increase the probability of voluntary return, reduce transportation costs, and increase the cultural proximity between host communities and refugees. Host communities residing near borders are often rural and under supported by the central government, so the creation of a camp close to these peripheral villages effectively directs large humanitarian efforts towards these populations. Given that refugee response planning has grown increasingly inclusive, this means newfound investments in local infrastructure shared with refugees (schools, health clinics) and improved access to clean water, healthcare, and other services. Hence, border area camps may serve host governments well because of the corresponding allocation of new resources to these oft-neglected areas. Some countries, such as Kenya, also see the locating of camps in peripheral border areas as a security measure. Since *prima facie* refugee registration is rapid, and since some militant groups recruit from refugee camps, Kenya believes it can mitigate the risk of al-Shabab-led violent attacks (in its main commercial hubs, at least) by maximizing the distance between camps and major cities.

We currently know less about the dynamics of emergency response planning and refugee camp creation in weaker states with high tenure insecurity, such as the DRC. While stronger states have, for decades, redirected refugee arrivals to designated areas, we acknowledge the possibility that this need not be the case in the absence of a sufficient state apparatus. In these contexts, we expect that either UNHCR is orchestrating the movement of refugees to designated camp areas, or that UNHCR takes the area that refugees congregate after crossing the border as the location for a future camp. The former seems questionable given the challenge of establishing rights to land, while the latter seems logistically difficult. We hope to fill in this missing gap in our understanding as we further pursue our research agenda.

SECTION 4: CHARACTERISTICS OF REFUGEE CAMP LOCATIONS

4.1 Method and Data

It is common to find media representations to the undesirability of the locations of refugee camps due to their perceived remoteness and lack of access to resources (Citations). But to our knowledge, there have been no large-scale quantitative descriptive study of the geographical characteristics of refugee camp locations. To begin strengthening our understanding of what camp locations are like, we use the ARD and multiple geospatial datasets to examine the camps' economic, agricultural, and environmental features, as well as their proximity to active conflicts.

The depiction of refugee camps in the popular press implies that camps tend to be in rather desolate areas (for example Calvert 2005; Kushner and Naughton 2017). But refugee camps are components of complex humanitarian missions requiring operating environments with a certain sufficiency in terms of access to resources, logistics, and supply lines. Although host governments usually have final say in where a camp is placed, there is a recognition among all parties that a location will not work if it has, for example, no access to water and no roads for transporting water in. This could mean that refugee camps are actually located in relatively more desirable locations. Indeed, the UNHCR's own handbook for humanitarian emergencies, which lays out an idealized decision making process for refugee camp establishment for policymakers, makes it clear that all else equal, refugee camps should be located in places that are not far from population centers, have sufficient road access, are located on relatively flat terrain with agricultural potential, and have adequate access to water. (UNHCR 2015b).

For this descriptive analysis, we exploit the widespread availability of remote sensing-derived data for SSA. The data come from many sources. We use a measure of caloric potential available from Galor and Özak (2015) as a proxy for the food growing potential of the region

around camps. Data from the US Geological Survey (USGS) provides us with measures of the ruggedness and average slope of the terrain.[10] We also use data from the US National Atmospheric and Oceanic Administration (NOAA) to measure nighttime lights, which is a measure of economic activity and population density (Alix-Garcia et al. 2018). With HarvestChoice (2015) raster data, we measure how close the camp locations are to population centers with 100,000 and 250,000 residents, respectively. For data on precipitation and temperature, we use remote sensing data from Hijmans et al. (2005). Additional details on each of these datasets are in Table 1, which inventories these sources and describes the measures we construct.

Along with this data, we also construct our own conflict measure. Based on the Armed Conflict Location and Event Data (ACLED) (Raleigh et al. 2010), we identify conflict hotspots on the continent. ACLED contains geolocated points of conflict in SSA starting in the late 1990s through the present. For each year covered by the data, we construct concave hulls around the points by first constructing buffers around the individual points, and then dissolving the overlapping borders of the buffers. Finally, we measured the size of each of these hulls and selected the largest 10 percent of these shapes for each year. We then create a point, we call a "conflict node", at the geometric center of each resulting hull to provide an approximate point for the center of large clusters of conflict events across the region. These conflict nodes coincide with the locations of the known conflicts on the continent over this period of time, with points clustering in Liberia, Sierra Leone, Darfur, the eastern DRC, Rwanda, Burundi, Uganda, and the CAR. We then measure the linear distance (in kilometers) from each refugee camp to its closest conflict node. Figure 4 illustrates these measures.

[10] Terrain ruggedness provides a proxy for agricultural potential and suitability for building structures and cultivating fields.

To measure the geographical environment around a camp, we construct buffers with a 20 km radius around each camp location. The total area a in square kilometers covered by each buffer is therefore 1,256.64 square kilometers. We clip these buffer zones for areas over stable surface water, so our buffer areas are not of identical shape. We then calculate an average of the pixel values within that buffer zone using the various raster files we have assembled.

We also use simple t-tests to examine whether the sites selected for camps are significantly different from a random selection of areas further from the camps. To do this, we use the ARD data to estimate the general region of SSA where refugee camps are found. We construct this region by generating buffers around the ARD camp points with 300-kilometer radii. We then clip these buffers by the African land shapefile so that the region we are creating does not cover areas of water. We then dissolve the borders of these buffers to create a contiguous polygon indicative of the general region of the continent where refugee camps are located. Figure 5 shows the sub-region of SSA that this covers. We then generate 423 random points within this area to serve as the refugee camp locations chosen by a totally random site selection process.

4.2 Results

Table 2 contains the measurements of the geographic characteristics of the refugee camps in the ARD and compares them to the outcomes for a random set of points taken from the general region of the camps. Compared to the general region, camps are in areas that are more rugged and have a greater average slope, although the difference is only statistically significant at the 90 percent level. Rugged areas with higher average slopes are generally less hospitable to agriculture (Nunn and Puga 2012). This result, combined with the highly significant observation that camp locations receive much less precipitation than non-camp locations, suggest that stakeholders are choosing areas that are less hospitable to agriculture than the average location in the region. Surprisingly,

we find no statistically significant difference between refugee camp locations and the comparison locations in terms of their caloric potential. Perhaps the camps are located in areas that are *perceived* as being less valuable for agricultural purposes, despite comparable caloric potential.

The camps are also in areas that are significantly closer to population centers than other locations in the general region. This likely reflects of the stakeholder imperative to situate camps near infrastructure and markets that can facilitate the servicing of the camps' logistical needs. We find, on average, a 160-kilometer difference in the distance between a camp and a conflict hotspot versus the distance between the comparison points and the nearest hotspot. This difference is highly statistically significant and indicates that on average, the camp locations are much closer to conflict hotspots on the continent than other areas of the general region.

SECTION 5: CONCLUSION

We have introduced a new and unique dataset for researchers to advance questions related to refugee settlement and the phenomenon of refugee camps 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 up new and exciting opportunities to study highly localized effects of refugee camps on many relationships on which researchers can only presently speculate and to explore variation by time, country, or policy framework.

We also offer some basic summary statistics of the characteristics of the locations that contain refugee camps and compare them to a random subset of points in adjacent regions. We do so in order to evaluate whether the refugee camps are in less hospitable locations, which has often

been the depiction in the qualitative literature and popular media. We find that the data reflect this image to a considerable extent.

As of this writing we are planning several extensions to this dataset. The next version of ARD will use the Statistical Yearbooks to include measures of the population size of the camps. Additionally, we plan to expand the dataset to include other known precise sites where refugees are located on the continent. This will include informal locations where refugees have integrated into the local population, as well as other institutionalized refugee locations like transit centers and settlements. Together this will enable a broader and deeper research agenda on the topic of refugees, their movements, and the impacts of forced displacement.

SECTION 6: SOURCES

Alix-Garcia, Jennifer, and David Saah. 2010. "The Effect of Refugee Inflows on Host Communities : Evidence from Tanzania Stable URL : Https://Www.Jstor.Org/Stable/40647424 REFERENCES Linked References Are Available on JSTOR for This Article : Communities : Evidence from Tanzania The Effect of Refugee." *World Bank Economic Review* 24 (1): 148–170.

Alix-Garcia, Jennifer, Sarah Walker, Anne Bartlett, Harun Onder, and Apurva Sanghi. 2018. "Do Refugee Camps Help or Hurt Hosts? The Case of Kakuma, Kenya." *Journal of Development Economics* 130. Elsevier B.V.

Anti, Sebastian, and Colette Salemi. 2020. "When the Camp Comes to Town: African Refugee Camps and Host Outcomes." *Working Paper*.

Baez, Javier E. 2011. "Civil Wars beyond Their Borders: The Human Capital and Health Consequences of Hosting Refugees." *Journal of Development Economics* 96 (2). Elsevier B.V.: 391–408.

Borjas, George J. 1987. "Immigrants, Minorities, and Labor Market Competition." *Industrial An* 40 (3).

Calvert, Scott. 2005. "Desolate UN Refugee Camp Shelters Africans from Strife, Many Are Resigned to Spend Their Lives Stuck in Botswana." *SFGate*.

Card, David. 1990. "The Impact of the Mariel Boatlift on the Miami Labor Market." *Industrial and Labor Relations Review* 43 (2): 245.

Galor, Oded, and Ömer Özak. 2015. "Land Productivity and Economic Development: Caloric Suitability vs. Agricultural Suitability."

Hijmans, Robert J., Susan E. Cameron, Juan L. Parra, Peter G. Jones, and Andy Jarvis. 2005. "Very High Resolution Interpolated Climate Surfaces for Global Land Areas." *Royal Meterological Society*.

Kreibaum, Merle. 2016. "Their Suffering , Our Burden? How Congolese Refugees Affect the Ugandan Population." *World Development* 78. Elsevier Ltd: 262–287.

Kushner, Jacob, and Jake Naughton. 2017. "Letter from Kakuma." *The Nation*.

Maystadt, Jean-François, and Gilles Duranton. 2019. "The Development Push of Refugees : Evidence from Tanzania." *Journal of Economic Geography* 19: 299–334.

Maystadt, Jean-françois, Kalle Hirvonen, Athur Mabiso, and Joachim Vandercasteelen. 2000. "Impacts of Hosting Forced Migrants in Poor Countries." *Annual Review of Resource Economics*.

Maystadt, Jean-François, Valerie Mueller, Jamon Van Den Hock, and Stijn van Weezel. 2020. "Vegetation Changes Attrituabe to Refugees in Africa Coincide with Agricultural." *Environmental Research Letters* In press.

Maystadt, Jean-François, and Philip Verwimp. 2014. "Winners and Losers among a Refugee-Hosting Population." *Economic Development and Cultural Change* 62 (4): 769–809.

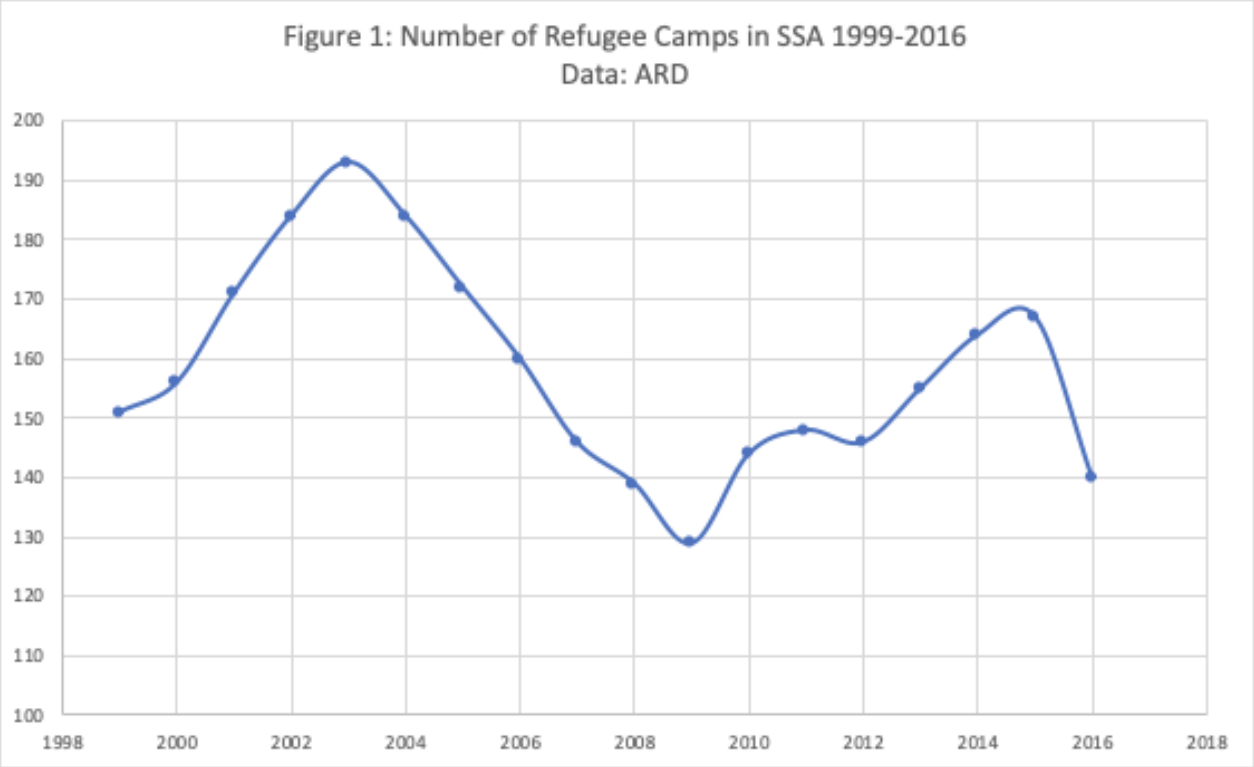
Nunn, N, and D Puga. 2012. "Ruggedness: The Blessing of Bad Geography in Africa." *The Review of Economics and Statistics* 94 (1): 20-36.

Raleigh, Clionadh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. 2010. "Introducing ACLED: An Armed Conflict Location and Event Dataset." *Journal of Peace Research* 47 (5): 651–660.

Ruiz, Isabel, and Carlos Vargas-Silva. 2013. "The Economics of Forced Migration." *Journal of Development Studies* 49 (6).

———. 2015. “The Labor Market Impacts of Forced Migration.” *American Economic Review Papers and Proceedings* 105 (5): 581–586.
———. 2016. “The Labour Market Consequences of Hosting Refugees.” *Journal of Economic Geography* 16.
UNHCR. 2000. “Refugees and Others of Concern to UNHCR: 1999 Statistical Overview.”
———. 2002a. “Refugees and Others of Concern to UNHCR, 2000 Statistical Overview.” Geneva.
———. 2002b. “Statistical Yearbook 2001: Refugees, Asylum-Seekers, and Other Persons of Concern - Trends in Displacement, Protection and Solutions.”
———. 2003. “Statistical Yearbook 2003: Trends in Displacement, Protection and Solutions.”
———. 2004. “Statistical Yearbook 2004: Trends in Displacement, Protection, and Solutions.”
———. 2005. “Statistical Yearbook 2005: Trends in Displacement, Protection and Solutions.”
———. 2006. “Statistical Yearbook 2006: Trends in Displacement, Protection and Solutions.”
———. 2007. “Statistical Yearbook 2007: Trends in Displacement, Protection and Solutions.”
———. 2008. “Statistical Yearbook 2008: Trends in Displacement, Protection, and Solutions.”
———. 2009. “Statistical Yearbook 2009: Trends in Displacement, Protection and Solutions.”
———. 2010. “Statistical Yearbook: Ten Years of Statistics.”
———. 2011. “Statistical Yearbook.”
———. 2012. “UNHCR Statistical Yearbook 2012.”
———. 2013. “Statistical Yearbook: 25 Years of Global Forced Displacement.”
———. 2014. “UNHCR Statistical Yearbook 2014.” *Statistical Yearbook*: 92.
———. 2015a. “Statistical Yearbook 2015.”
———. 2015b. “Camp Planning Standards (Planned Settlements).” In *UNHCR Emergency Handbook*, 4th ed.
———. 2016. “Statistical Yearbook 2016.”
———. 2018. “UNHCR Presence and People of Concern ’s Locations.” *UNHCR Maps*. http://unhcr.maps.arcgis.com/apps/webappviewer/index.html?id=2028db44801d43fe8eb49321eea19285.
———. 2019. “Figures at a Glance.” *About Us*. Retrieved August 16, 2019. https://www.unhcr.org/figures-at-a-glance.html.
Verme, Paolo, and Kirsten Schuettler. 2019. “The Impact of Forced Displacement on Host Communities: A Review of the Empirical Literature in Economics.” *World Bank Policy Research Paper 8727*.
Verwimp, Philip, and Jean-François Maystadt. 2015. “Forced Displacement and Refugees in Sub-Saharan Africa An Economic Inquiry.” *World Bank Policy Research Working Paper 7517*.

SECTION 7: CHARTS AND FIGURES



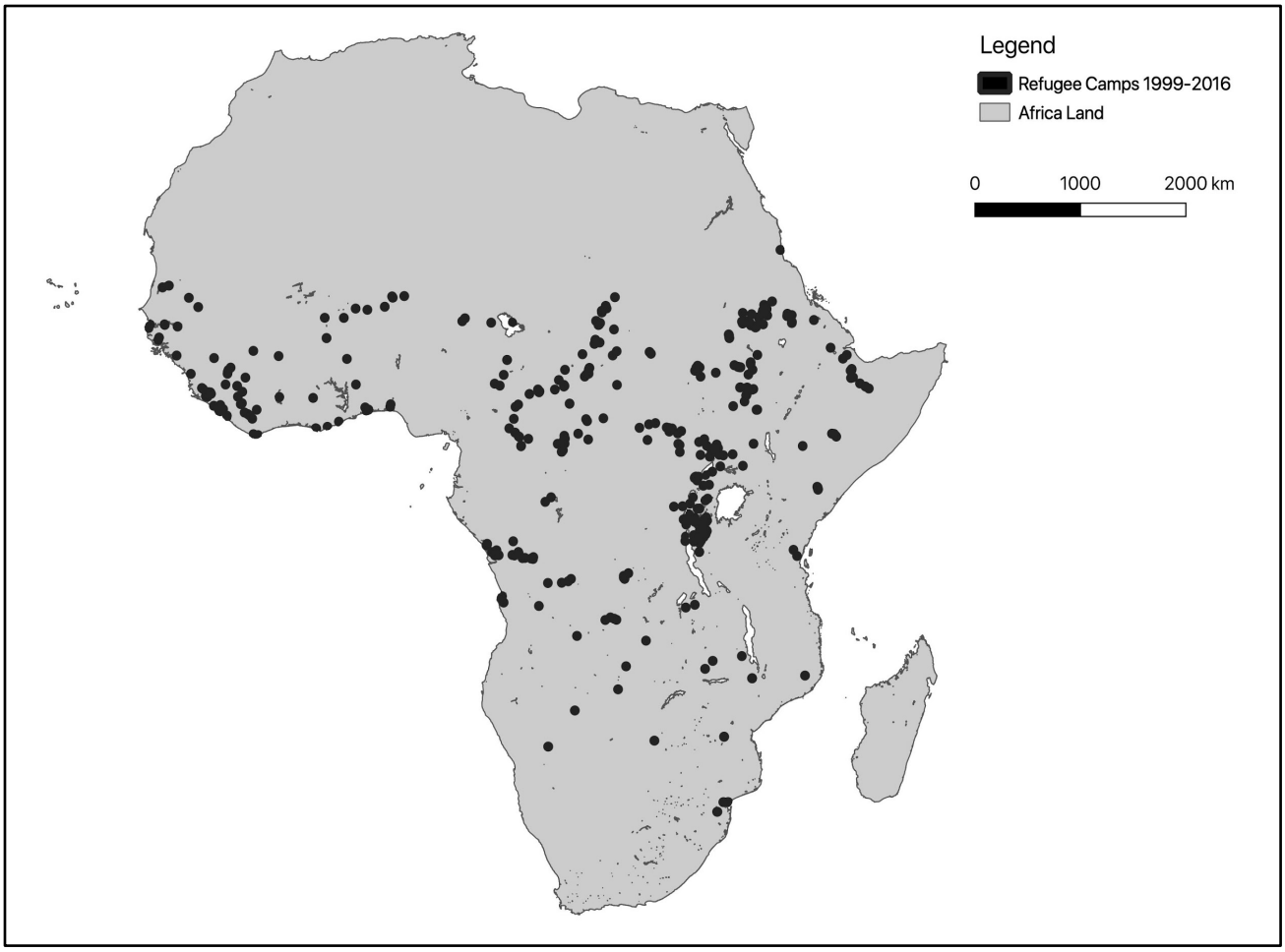


Figure 2: locations of all refugee camps in ARD 1999-2016. Data: ARD

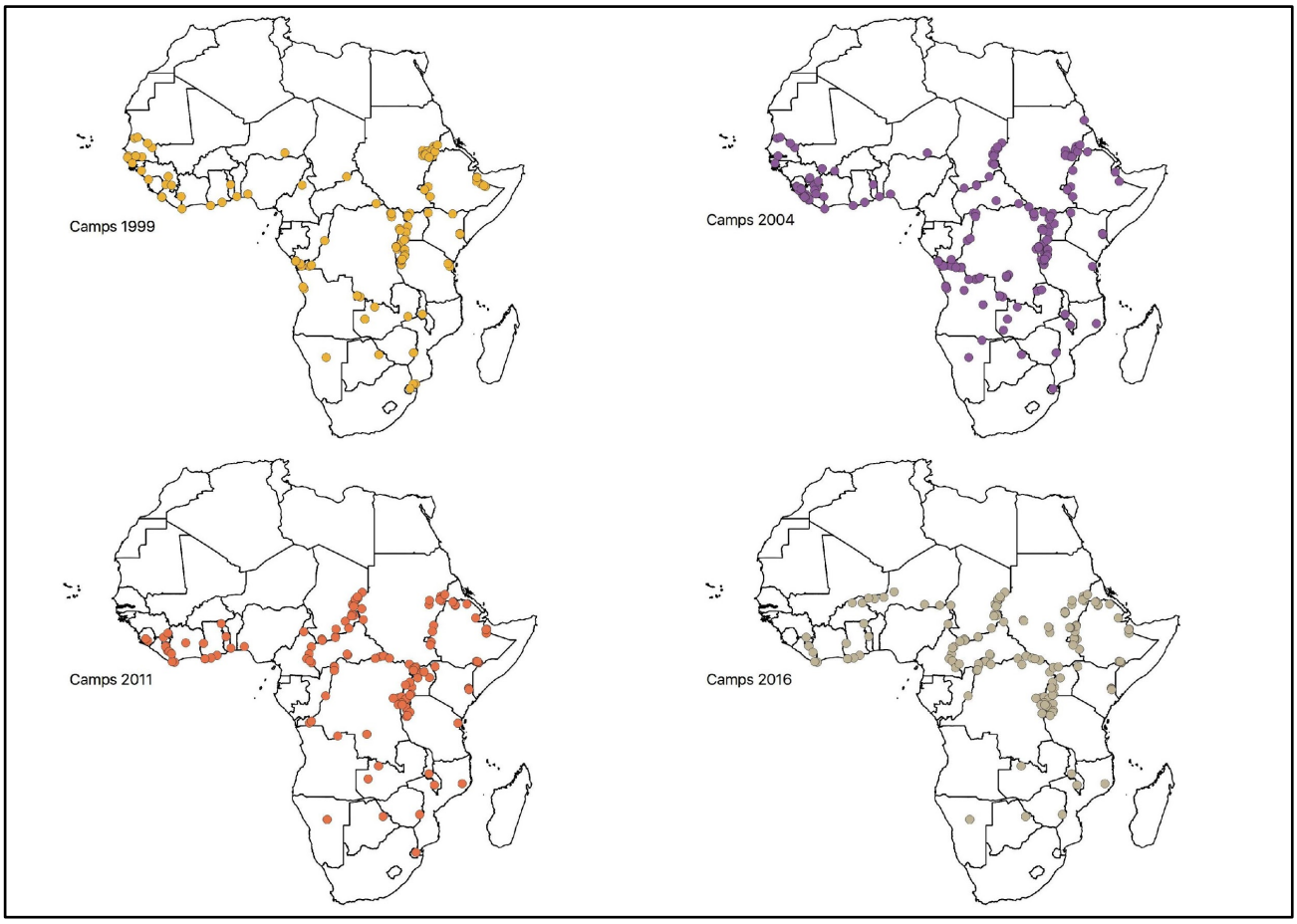


Figure 3: ARD refugee camps over time. Data: ARD. Note: This is the same as Figure 1 in Anti and Salemi (2020).

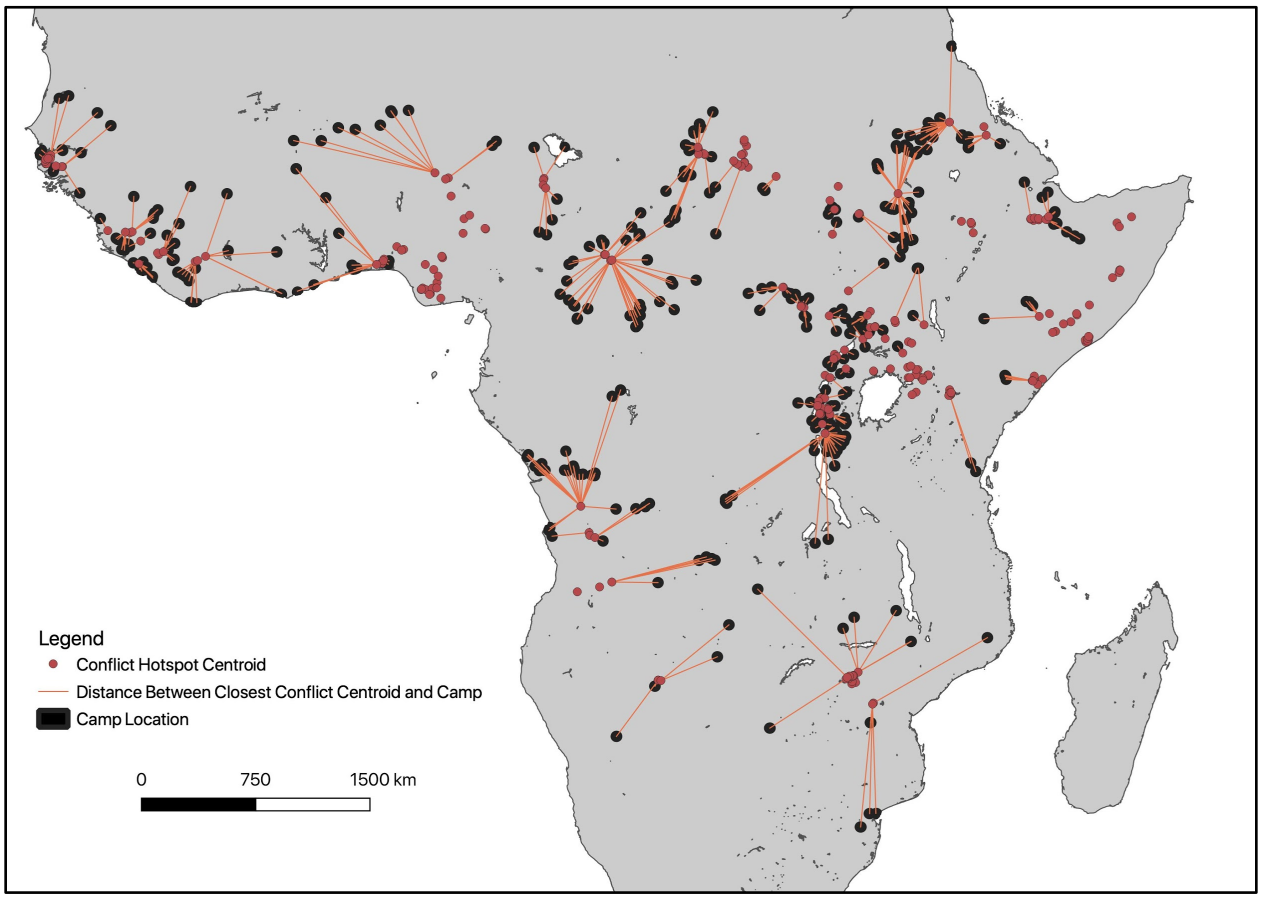


Figure 4: Distance between refugee camps and closest conflict hotspot centroid. Data: ARD and ACLED

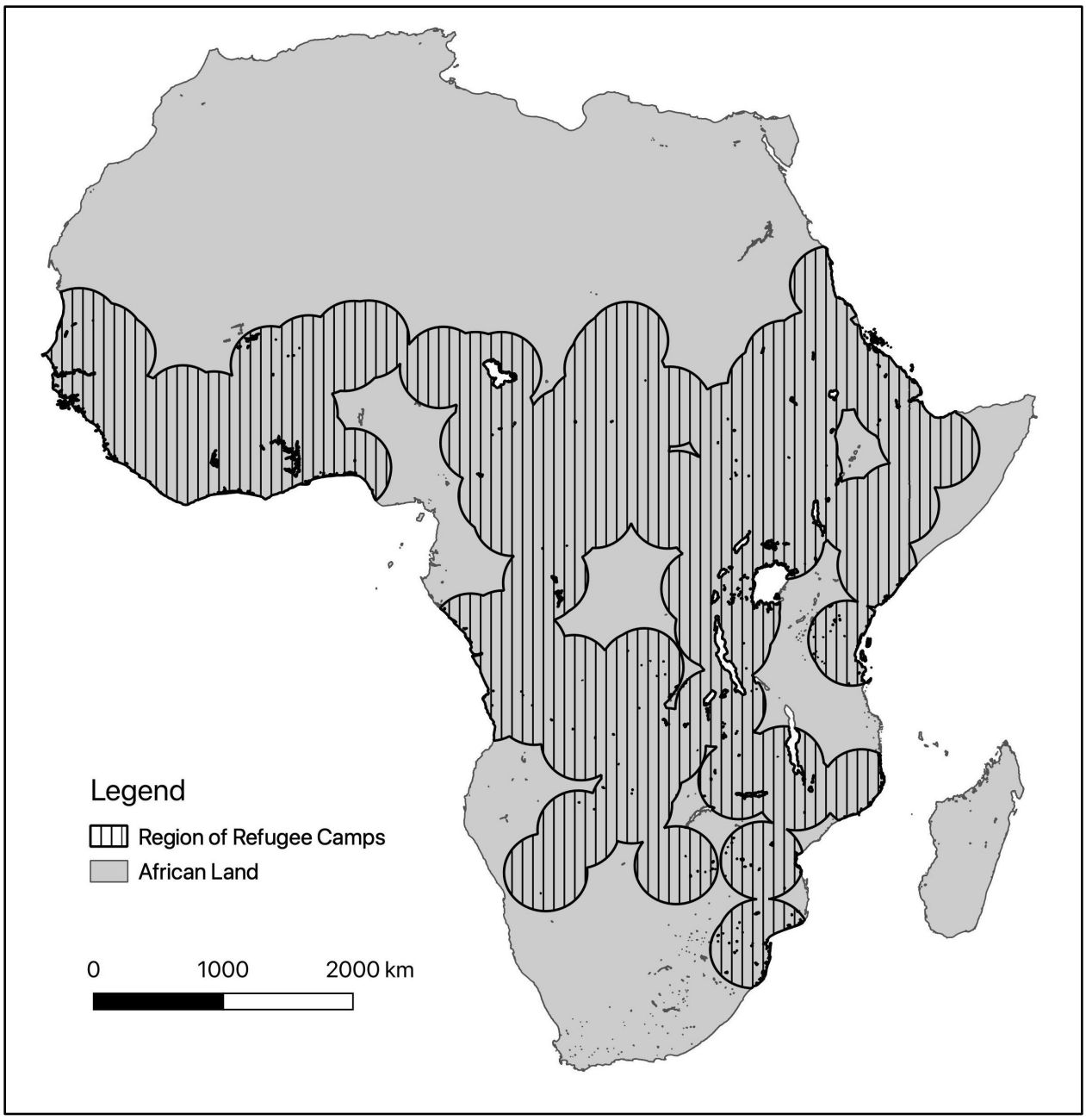


Figure 5: regions within 300 km of a refugee camp in SSA. Data: ARD

SECTION 8: TABLES

Table 1: Spatial Data Definitions and Sources

Measure	Definition	Data Source
Calorie Potential Index		Galor and Özak (2015)
Stable Nighttime Lights		NOAA
Distance to Conflict Hotspot		Raleigh et al. (2010) (ACLED data)
Ruggedness		USGS, Nunn and Puga (2012)
Rainfall and Temperature		Hijmans et al. (2005) (WorldClim data)
Surface Water Access (Used for clipping buffers only)		JRC Global Surface Water Mapping, Maximum Water Extent Layer
Time to Closest Town >100k Pop,		HarvestChoice (2015)
Time to Closest Town >250k Pop		HarvestChoice (2015)
Proximity to active agricultural areas		NASA Global Agricultural Lands

Table 2: Statistical Analysis of Camp Geography

Variable	Camp Average	Random Site Average	Difference
Average Caloric Potential Index	9,145.35	8,862.172	283.179 (262.474)
Average Ruggedness (Index)	69,111.9	58,031.79	11,080.1* (6,224.325)
Average Slope (Index)	1,840.332	1,552.295	288.037* (165.966)
Average Rainfall (Inches)	45.118	64.747	-19.628*** (5.089)
Average Temperature (Celcius)	24.028	23.749	0.278 (0.177)
Average Time to Closest Town (pop. >100,000) (hours)	6.59	10.5	-3.909*** (0.443)
Average Time to Closest Town (pop. >250,000) (hours)	8.872	12.831	-3.959*** (0.492)
Average Stable Night Lights (2000)	0.330	0.277	0.053 (.1)
Distance to Closest Conflict Hub (km)	212.936	372.994	-160.057*** (13.945)
Observations	423	423	846

Notes: *** p<0.01, ** p<0.05, * p<0.1 correspond to t-test for whether difference ≠ 0.
Standard errors are in parentheses.