

RESEARCH ARTICLE

Open Access



# Mapping refugee populations at high resolution by unlocking humanitarian administrative data

Edith Darin<sup>1,2\*</sup> , Ahmadou Hamady Dicko<sup>3</sup>, Hisham Galal<sup>4</sup>, Rebeca Moreno Jimenez<sup>5</sup>, Hyunju Park<sup>6</sup>, Andrew J. Tatem<sup>1</sup> and Sarchil Qader<sup>1,7</sup>

**Background** Informing local decision-making, improving service delivery and designing household surveys require having access to high-spatial resolution mapping of the targeted population. However, this detailed spatial information remains unavailable for specific population subgroups, such as refugees, a vulnerable group that would significantly benefit from focused interventions. Given the continuous increase in the number of refugees, reaching an all-time high of 35.3 million people in 2022, it is imperative to develop models that can accurately inform about their spatial locations, enabling better and more tailored assistance.

**Methods** We leverage routinely collected registration data on refugees and combine it with high-resolution population maps, satellite imagery derived settlement maps and other spatial covariates to disaggregate observed refugee totals into 100-m grid cells. We suggest a deterministic grid cell allocation inside monitored refugee sites based on building count and a random-forest-derived grid cell allocation outside refugee sites based on geolocating the textual geographic information in the refugee register and on high-resolution population mapping. We test the method in Cameroon using the registration database monitored by the United Nations High Commissioner for Refugees.

**Results** Using OpenStreetMap, 83% of the manually inputted information in the registration database could be geolocated. The building footprint layer derived from satellite imagery by Ecopia AI offers extensive coverage within monitored refugee sites, although manual digitization was still required in rapidly evolving settings. The high-resolution mapping of refugees on a 100-m grid basis provides an unparalleled level of spatial detail, enabling valuable geospatial insights for informed local decision-making.

**Conclusions** Gathering information on forcibly displaced persons in sparse data-setting environment can quickly become very costly. Therefore, it is critical to gain the most knowledge from operational data that is frequently collected, such as registration databases. Integrating it with ancillary information derived from satellite imagery paves the way for obtaining more timely and spatially precise information to better deliver services and enhance sampling frame for target data collection exercises that further improves the quality of information on people in need.

**Keywords** Gridded population, Refugee, Administrative data, Building footprint, Satellite imagery

\*Correspondence:

Edith Darin

[edith.darin@demography.ox.ac.uk](mailto:edith.darin@demography.ox.ac.uk)

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Background

High-spatial resolution population mapping is of paramount importance in providing support for localized decision-making within various domains of public interest. Examples of such domains include urban planning (Thakuria 2017; , Chen 2021), environmental hazard risk management (Thomas et al. 2023; Wu 2018; Licínio 2013) and public health (Augusto 2021; Linard et al. 2012; Linard et al. 2010), where optimal resource allocation among different geographical areas is a critical concern. This significance is further underscored in the context of populations impacted by crises. In 2022, the United Nations High Commissioner for Refugees (UNHCR) documented 29.4 million refugees under UNHCR mandate and an additional 5.9 million refugees under United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA) mandate. The world has never had more refugees than today with the largest yearly increase since the start of global refugee statistics (+36%), and yet we know little about the precise location of their resettlement (UNHCR 2023a). Nonetheless, detailed spatial information not only enhances the efficiency of distributing humanitarian aid but also facilitates strategic decision-making regarding the placement of crucial facilities and helps guide and target collection of new and improved data.

So far, spatial demography scholarships have mainly focused on increasing the spatial resolution of population figures derived from census data that is providing spatially detailed estimates of the long-term stable resident population. Significant modelling efforts have been made using remotely sensed data (like night lights or land cover) and other geographic data (like roads networks or infrastructure maps) to disaggregate coarse census totals into small grid cells (Stevens 2015; Tatem 2017; Leyk 2019). Since Tobler's "World population in a grid of spherical quadrilaterals" (Tobler et al. 1997) and Liverman et al.'s "People and Pixels" (Liverman 1998), the benefits of gridded population data have been acknowledged for the spatial insights given by detailed mapping. Another strength of this data format is its analytical flexibility, whereby population totals can be aggregated to any spatial unit (Leyk 2019). If the first developed methods (Tobler et al. 1997; Liverman 1998) consisted in distributing evenly population across grid cells within census unit, the availability of finer remote-sensing product as well as advancement in statistical modelling has led to refine the spatial allocation of population using ancillary datasets. This approach, known as dasymetric mapping (Semenov-Tian-Shansky 1928), has been mainly applied for enhancing the spatial distribution of total population count, sometimes including age and sex

subgroups (Alegana 2015; Pezzulo 2017). The purpose of this research is to examine if the dasymetric mapping method could be extended for mapping subpopulations in a crisis context, more specifically for refugees in the context of large-scale international forced migration, and to understand what humanitarian insights could be gained from adopting a gridded map of refugees.

Remote-sensing-based methods have already been previously leveraged in crisis context to provide rapid estimation of forcedly displaced population sizes based on the manual enumeration of makeshift structures from satellite imagery (Checchi et al. 2013). Building upon the success of this approach, subsequent research has focused on refining feature detection algorithms tailored to the mapping of refugee settlements (Logar 2020; Quinn 2018). However, to our knowledge, it remains yet to harness those refined spatial data for mapping a countrywide refugee population at a fine spatial scale. Such a resource not only facilitates policy micro-planning (Augusto 2021) but also inform the design of survey sampling focused on vulnerable population (Thomson et al. 2020) and participates more generally to the toolbox of crisis analytics at the postcrisis response stage (Qadir 2016).

A second previously untapped data source for refugee mapping explored in this research is humanitarian administrative data. Refugees have been for long a blind spot of comprehensive active data collection exercise such as national census, even in major refugee hosting countries (Carr-Hill 2013), not only because it is harder to capture a moving population but also for the political reasons deprived neighbourhoods can be missed on the map (Mahabir 2016). In parallel though, large-scale passive data collection such as the digital registration of all individuals in need has now become a cornerstone of humanitarian data management systems. An example of such an initiative is the proGres database developed by the United Nations High Commissioner for Refugees (UNHCR), which records all refugees defined as "individuals who are outside their country of origin and who are unable or unwilling to return there owing to serious threats to life, physical integrity or freedom resulting from generalised violence or events seriously disturbing public order" (UNHCR 2019). The aim of the registration is to provide access to rights, effective protection and assistance to those in need (Hovy 2018). While the primary purpose of this administrative registry is operational, it also serves as an indirect valuable source of demographic information on the refugee population and its dynamic, for example to compile mortality statistics (Masquelier 2017). However, this administrative data contains also precise subregional geographic

information that is yet to be exploited for evidence-based decision-making.

Countrywide detailed refugee mapping exercise is particularly important for refugees who reside outside of monitored camps, as locating and effectively assisting them present greater challenges. We therefore propose a model that maps the registered refugee population by converting the information in the register database to geographical locations and combining it with geospatial layers derived from high-resolution remote-sensing data. The area of focus is Cameroon, a country located in a region facing a multifaceted humanitarian and protection crisis caused by conflict, intercommunal violence, and the impacts of climate change. As a consequence, Cameroon hosts 471,386 refugees as of April 2023 (17th refugee host country worldwide, 7th in Africa) (UNHCR 2023). The modelling output is a countrywide gridded map of the refugee population at 100-m resolution. The novelty of the proposed high-resolution mapping is twofold. First, the extent of the register database combined with remote-sensing-derived covariates enables a countrywide mapping at an unprecedented spatial scale and extent. Second, by informing about the relative spatial concentration of refugees compared to the resident population, it paves a way for the inclusion of the unregistered refugee population in surveys and policy designs, provided their spatial allocation follows that of the registered refugee population.

Alongside this research, we also provide an assessment in a context of limited data of a freely available gazetteer to geolocate textual geographic data and of seven settlement maps to gain a better understanding of the quality of automated mapping of settlement layout in refugee sites from satellite imagery. Ultimately, we explore two analyses exclusively enabled by the gridded map format, making a valuable contribution to a data-driven crisis response.

### Data and methods

We demonstrate how to leverage diverse data sources, satellite imagery and modelling techniques to provide nuanced insights into refugee spatial dispersion. We first explain why we choose the Cameroon setting, then enumerate the different data sources explored and their global availability and finally present our modelling framework and potential for replicability.

#### Location: the refugee context in Cameroon

Cameroon is confronting a multifaceted humanitarian and protection crisis caused by conflict, intercommunal violence and the impacts of climate change (International Rescue Committee 2019). The location of refugees follows roughly this schema: refugees from Nigeria who

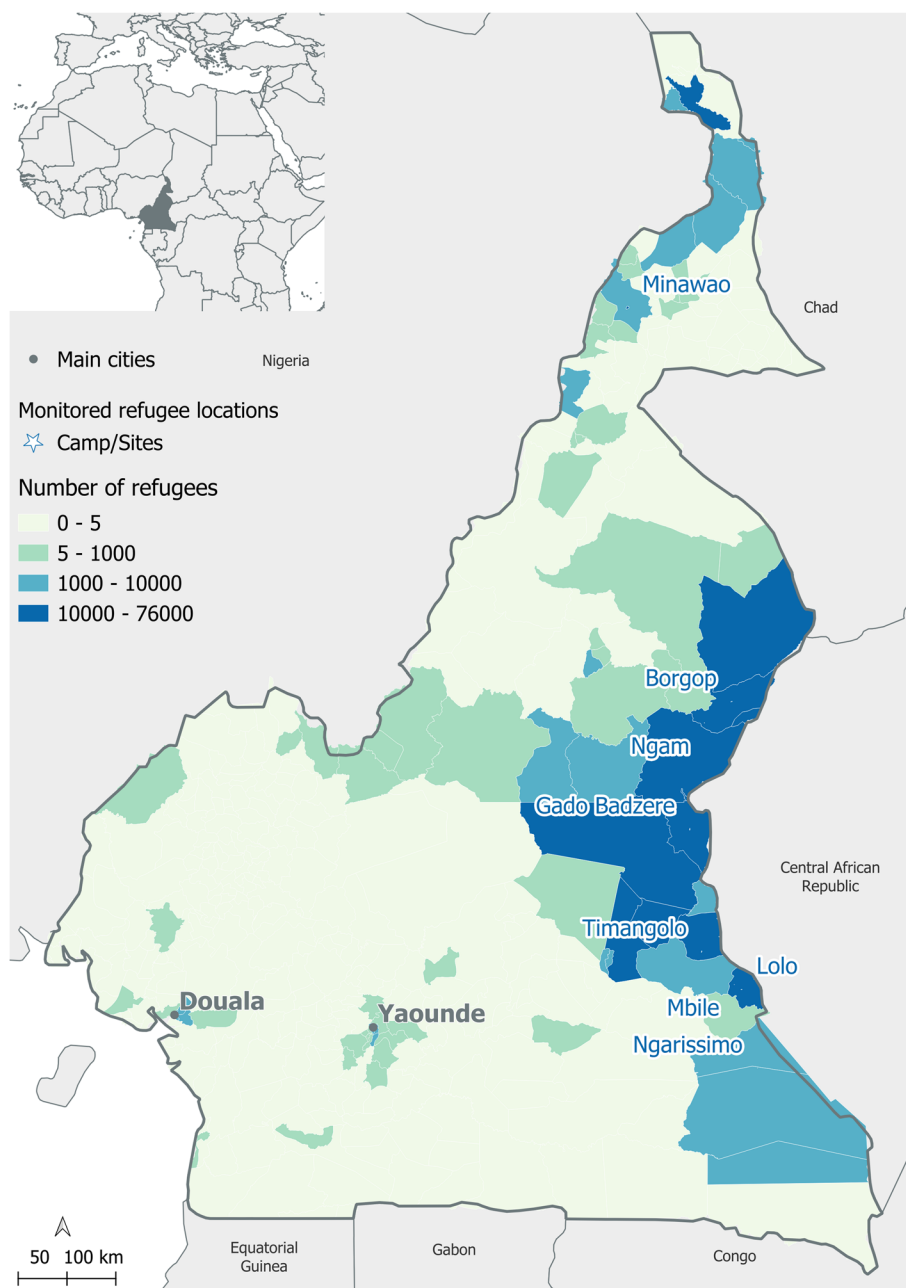
fled Boko Haram violence are situated in the Far North Region, and refugees from Central African Republic (CAR) are located in the East region (see Fig.1) (International Rescue Committee 2019). Some refugees from CAR have been displaced for over a decade, and the political and security situation in CAR has not improved sufficiently to warrant their return (International Rescue Committee 2019). UNHCR monitors seven refugee sites along the border with CAR (see Fig.1). Nigerians began seeking refuge in Cameroon in 2012 (International Rescue Committee 2019). In response, Cameroon established a camp in Minawao to accommodate up to 20,000 refugees, but the camp's capacity was nearly exceeded by 2014 (International Rescue Committee 2019). This is the only refugee site titled camp in Cameroon. By the end of 2015, violence along the Cameroon-Nigeria border displaced more than 90,000 Cameroonians and refugees who had settled in these areas, and this pressure at the border never really stopped (International Rescue Committee 2019). At the end of March 2023, Cameroon had 471,386 refugees, including roughly 345,000 Central African Republic refugees and 121,000 Nigerians, and more than one million internally displaced people (UNHCR 2024). The degree of integration between refugees in Cameroon and host communities is highly contextual. A majority (81%) of CAR refugees reside alongside host communities, rather than in designated refugee settlements. Conversely, in the Far North Region, where Nigerian refugees constitute a smaller proportion of the total population, over half (64%) reside in the Minawao camp. The remaining refugee population coming from other countries like Tchad (around 4000 people) or Niger (around 3000 people) reside all outside of designated areas (United 2024).

#### Data sources

Data are involved at three stages of the high-resolution refugee mapping pipeline: first to quantify the population to map which is in this case the comprehensive refugee listing obtained through administrative records, second to geolocate the refugee population through a settlement map and a gridded population map and third to model its spatial variation through multiple spatial covariates.

#### Refugee population

The data has been retrieved in April 2023 from the Progres version 4 registration database, UNHCR's corporate registration, identity and case management tool that holds the individual data of forcibly displaced and stateless persons (UNHCR 2018). First rolled out in 2018, it is built on Microsoft customer relationship management and is the backbone of multiple interoperable tools and application ranging from the rapid off-line inputting of



**Fig. 1** Number of refugee per province in Cameroon as of April 2023. Source: proGres registration database, 2023

refugees information when they seek assistance to the biometric checks during food distributions (UNHCR 2019). The information recorded covers not only basic demographic characteristics but also identification of specific needs and recording of events such as resettlement acceptance or voluntary departure for all countries where UNHCR operates. Considered as the most up-to-date list on refugees in a country, the proGres database (UNHCR 2018) covers refugee cases both inside

monitored camps and outside of camps and is regularly used as a sampling frame to conduct surveys and to provide statistics on the current size of the refugee population. The aggregated statistics from the database are openly available on UNHCR website (UNHCR 2022) and through a R package (Galal 2023), but accessing the individual records requires special authorisation. In countries where government also undertake refugee registration, UNHCR database can contain only a fraction of the

refugees, but this is not the case in Cameroon where it can be considered as a refugee census. Out of the 471,386 refugees recorded in Cameroon, 337,223 refugees were reported as living outside of UNHCR-monitored refugee camps and sites. The registration database (UNHCR 2018) provided us with refugee count at province level (administrative level 3) for Cameroon and partial information on refugee precise locations based on textual input.

#### **Gazetteer**

To link the textual information inputted manually by UNHCR staff member in the proGres registration database (UNHCR 2018) with spatial coordinates, we used a spatial gazetteer, which is a database that links place name and location. For that purpose, we leveraged OpenStreetMap database (OpenStreetMap 2023), with which we could retrieve 37,531 point locations and their associated names by querying all the point features, as of June 2023, with as key “place” and as value “all” using the QuickOSM plug-in from QGIS (Herbreteau 2018).

#### **Settlement maps**

To map refugee population inside monitored sites, we assessed the accuracy of settlement maps that were extracted from different satellite imagery by five different institutions. Those five settlement data sources differ in their format (building polygons or gridded settled area), their spatial resolution and their temporal resolution as summarised in Table 1.

#### **Base high-resolution population**

We used as base layer for the refugee population located outside of monitored refugee sites a 100-m gridded population layer, product of a collaboration between the Cameroon National Statistics Office and the WorldPop Research Group produced from household survey dating from 2021 and 2022 and building

footprints dating from 2021. The method used was a Bayesian hierarchical geostatistical model (United Nations High Commissioner for Refugees Cameroon 2023). This is the gridded population map that has been derived from the most recent population and geospatial data rather than the standard outdated WorldPop gridded population that despite being labelled 2020 disaggregates 2020 projections from the 2005 old census with spatial covariates dated from 2015 (Bondarenko 2020).

#### **Covariates**

The dasymetric mapping consists in informing the disaggregation of population totals into grid cells by leveraging ancillary covariates that are linked with the spatial distribution of people at local level.

We therefore gathered 20 covariates that map drivers of the spatial allocation of refugees inside Cameroon:

- Covariates linked to infrastructure are derived from OpenStreetMap: Distance to health, education, local roads, major roads, marketplaces, places of worship, road intersections (OpenStreetMap 2023) and from NASA Visible Infrared Imaging Radiometer Suite: mean intensity of night lights in 2022 (Elvidge et al. 2017)
- Covariates linked to insecurity are derived from the Armed Conflict Location and Event Database (Raleigh et al. 2010): distance to conflict locations in 2019, 2020, 2021 and 2022 and from the UNHCR sites location (OpenStreetMap 2023): distance to monitored refugee sites
- A covariate linked to the level of population density is derived from the previously described WorldPop gridded population dataset (Nnanatu 2024): The sum of population counts in a 1-km window

**Table 1** Attributes of the settlement maps assessed for refugee mapping inside sites

Name	Source	Date selected	Type	Spatial resolution	Coverage	Temporal resolution
World settlement footprint	German Space Agency (Marconcini 2021)	2019	Raster	10 m	Global	4 years
Global human settlement layer	Joint Research Centre — European Commission (Schiavina et al. 2022)	2020	Raster	100 m	Global	~5 years
Microsoft building footprints	Microsoft (Bing 2022)	2022	Polygon		Global	Ad hoc
Google building footprints (v1)	Google (Sirko 2021)	2021	Polygon		Africa, South Asia and Latin America	Ad hoc
Ecopia building footprints	Ecopia AI Maxar Technologies (Ecopia.AI 2019)	2021	Polygon		Africa	Ad hoc

- Covariates linked to the morphology of settlement are derived from the Ecopia building footprint: The mean and coefficient of variation of the area and perimeter of buildings contained in each grid cell, the date of the satellite imagery used, and the classification in urban and rural of the buildings (Dooley 2020).

**Mapping refugee distributions with high-spatial resolution**

The goal is to provide a workflow for mapping with high-resolution refugee across Cameroon. We have access to reliable refugee totals for the 8 monitored sites and for the 360 provinces (administrative level 3). To gain spatial details, we aim at disaggregating those totals into a fine resolution grid following the format of gridded population which consists in estimating population counts over a complete partitioning of the country of interest into same-sized small grid cells.

As summarised in Fig. 2, we developed two different methods to disaggregate refugee totals into grid cells depending on if the recorded refugee location was inside or outside of the UNHCR-led sites.

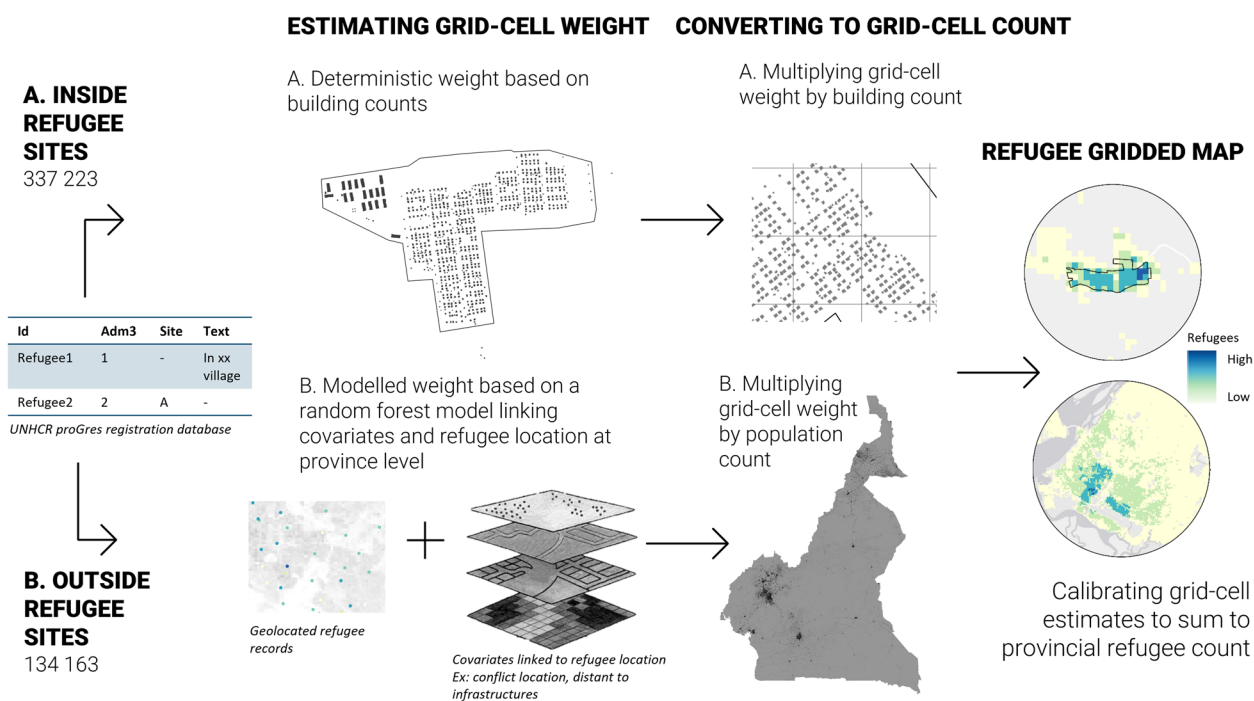
**Mapping refugees inside UNHCR-led refugee sites**

The mapping of refugees inside UNHCR-led refugee sites consists in spatially disaggregating the refugee totals observed for each site into all the grid cells covered by

the site. To do so, we first checked the site boundaries against different satellite imagery (Google, ESRI and Bing basemap visualised in April 2023) to update them in case of new evolution. We assessed the different settlement maps to select the most accurate in terms of number of buildings delineated and extent of settlement mapped. We manually digitised structures that could be seen from satellite imagery but were not in the satellite-imagery derived building footprint products. We overlaid a 100-m grid over the sites and computed the number of structure footprints in each grid cell. We disaggregated the total number of refugees registered in each site into the 100-m grid cells by using the number of structures as weights, adopting thus a deterministic dasymetric allocation of refugees (see Fig. 2). Since the method of the disaggregation of refugee totals inside site relies on the number of structures in each grid cell, the manual digitisation can be restricted to dropping a point for each structure.

**Mapping refugee outside UNHCR-led refugee sites**

The mapping of refugees outside sites consists in disaggregating at grid-cell level the refugee count observed at administrative level 3 as provided by the proGres database (UNHCR 2018). There are two primary steps for this approach: (1) geolocating the records from the administrative register and (2) combining the geolocation with



**Fig. 2** Grid-based mapping of refugees: a two-stage workflow

other spatial covariates to model the spatial allocation of the entire refugee population.

To geolocate the records from the registration database, we cleaned the text recorded in the freely inputted field about refugee location by removing Arabic character, upper case and trailing spaces and replacing internal whitespace and apostrophe by hyphen. We then merged the refugee locations with the point locations from OpenStreetMap that went through the same cleaning process. The point location is not necessarily the exact location of the text field, even less the actual residence of each refugee, but more an indication of the area they are likely to have stayed. We therefore converted the refugee point location layer to a continuous spatial indicator by creating a buffer for each refugee point based on the size of the refugee population assigned to this location. We then computed a distance layer from all grid cells to the buffered refugee location in order to avoid a strict subjective spatial cut-off. It is not possible to use the geolocation of refugee directly because of the following: (1) the point location provided by OpenStreetMap may not always correspond exactly with the location provided in the registry and (2) that would entail removing the refugees who could not be located with OpenStreetMap.

To address the second step that consists in modelling the refugee count at grid-cell level, we adapted a similar procedure as developed by Stevens (Stevens 2015) and widely applied by the WorldPop Research Group (Tatem 2017) for disaggregating total population count into grid cells:

1. We select gridded covariates that are related to refugee locations (see “Data and methods” section), in addition to the layer geolocating refugees administrative records as describe in previous paragraph.
2. We estimate through a random-forest model the relationship at administrative level 3 between the selected gridded covariates and the log of the total number of refugees divided by the total number of people, which is the number of refugees per host communities.
3. We predict the number of refugees per host communities using the fitted model for every grid cell contained in Cameroon and located outside of UNHCR-led refugee sites.
4. We multiply the estimated number of refugees per host communities by the number of host communities to obtain the estimated number of refugees for every grid cell.
5. To ensure that the sum of the grid cells with refugees is exactly equal to the refugee totals reported at administrative level 3, we calibrate the number of ref-

ugees per grid cell with the reported totals at administrative level 3.

Consistent with the foundational studies cited earlier (Wu 2018; Licínio 2013), our approach employs a random forest model to examine the relationship between population count and spatial covariates at administrative level 3. This methodological decision is informed by Metzger’s findings (Metzger 2022), which demonstrate that when population count spatial units are sufficiently granular (i.e. at the administrative level 3, same as our study’s spatial resolution), random forest models achieve similar performance to convolutional neural network models.

## Results

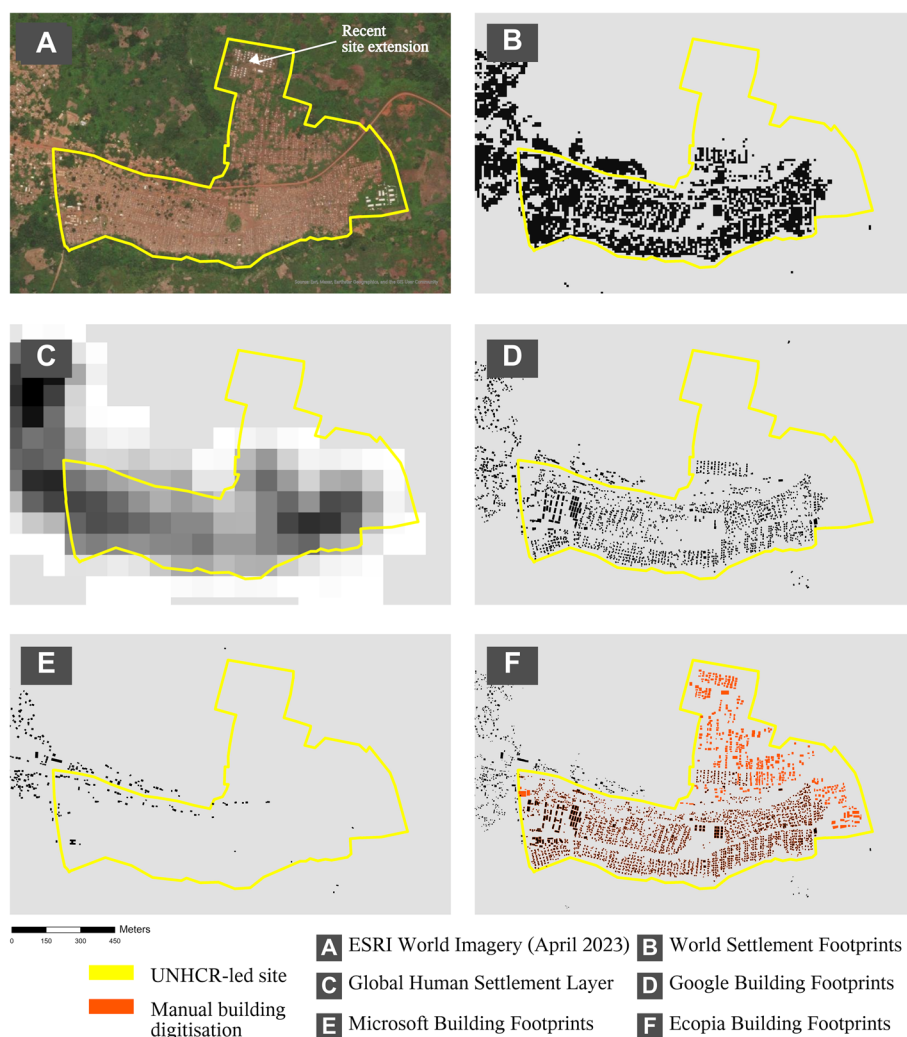
The analysis were performed within the R software framework (RStudio 2019), using the tidyverse suite (Wickham 2017), the randomForest (Cutler 2018), terra (Hijmans 2023) and sf (Pebesma 2018) package. Maps were drawn using the QGIS software (QGIS 2023).

### Mapping refugees inside of UNHCR-led refugee sites

Disaggregating refugee totals inside the UNHCR-monitored refugee sites depends strictly on the quality of the detected structures from satellite imagery. Figure 3 shows a visual assessment of the different feature extraction products and provides a clear example of the limitations of each settlement map in detecting structures specific to refugee sites. The main one relates to leveraging outdated raw satellite image: all the five maps have missed the recent extension in the north of the site. Microsoft building footprints layer despite providing detailed information on building delineation massively under-detects structures. We conclude that, across the eight sites, the best accurate data source for obtaining a comprehensive image of the distribution of buildings within UNHCR-led refugee facilities is by adding a manual delineation of newer structures to the most spatially detailed layer that is Ecopia AI as shown in Fig. 3.

### Mapping refugees outside of UNHCR-led refugee sites

The first step to model refugees outside UNHCR-led refugee sites is to convert the text fields describing spatially relevant information into a spatial mapping format. The procedure adopted led to 83% of the refugee population to be associated to a point location. As displayed in Fig. 4, Touboro is the province (administrative level 3) with the most refugees not being mapped (10,981 refugees or 20% of the total refugee population in the province), followed by Mora (4620 refugees or 50% of the total refugee population), Meiganga (3844 refugees or 15% of the total refugee population), Batouri (3649 refugees or 38% of the



**Fig. 3** Example of a settlement map assessment in a UNHCR-led site

total refugee population) and Makary (2915 refugees or 21% of the total refugee population). We then converted the geolocated records as described in the “Data and methods” section and represented an example in Fig. 4.

The second step to disaggregate refugee totals outside of monitored sites involves the use of a random-forest model. This model was able to explain 62% of the total variance in the proportion of refugees per host communities at the administrative level 3. While this goodness-of-fit is respectable, it is poorer than the disaggregation models for total population count regularly produced by the WorldPop research team, which can explain up to 99% of the variance in population count (Bondarenko 2020). One reason for this lower performance is that in our model, the spatial distribution of the resident population is already accounted for as we are modelling the number of refugees per host communities. The model’s aim, therefore, is to explain the spatial variation specific

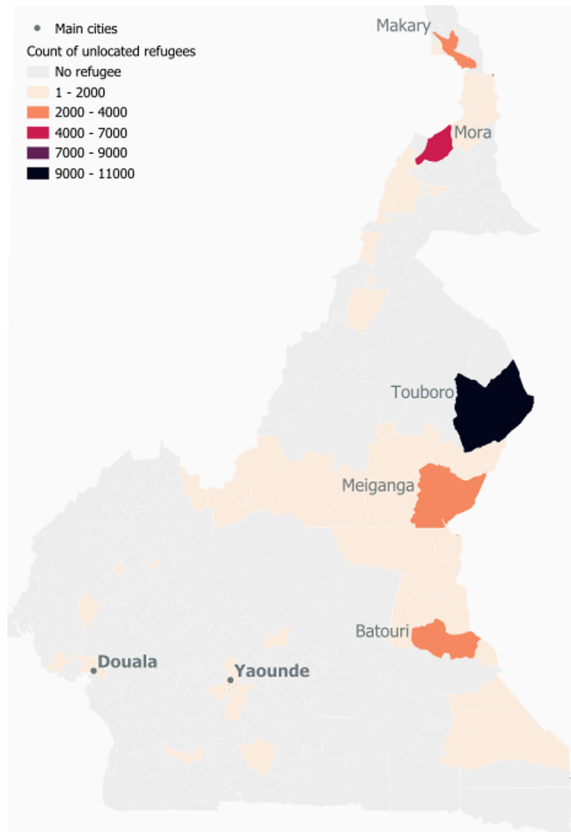
to the refugee population which is harder to capture. When we do not differentiate the host population and the refugee population, which is using as target variable the number of refugees per hectare rather than per host communities, the model fit jumps to 74% of explained variance which means that 12% of the variance is explained by the spatial distribution of the resident population.

Figure 5 illustrates the covariates that are the most important to model refugee spatial distribution among host communities. The top 2 are, as expected, the one geolocating refugees from the administrative register and the distance to refugee sites. If removed, they reduce by 28.7% and 18.5% the goodness of fit of the model. The following set of covariates that reduce by 10.8 to 11.4% the goodness of fit of the model are related to the morphology of the settlement layout (the mean perimeter and surface of the buildings) and the infrastructure (the distance to local roads). We can notice as well that recent



## Records geolocation

Province of the 54,612 refugees that could not be located



## Covariate processing

Transforming the geolocated records to a spatial covariate

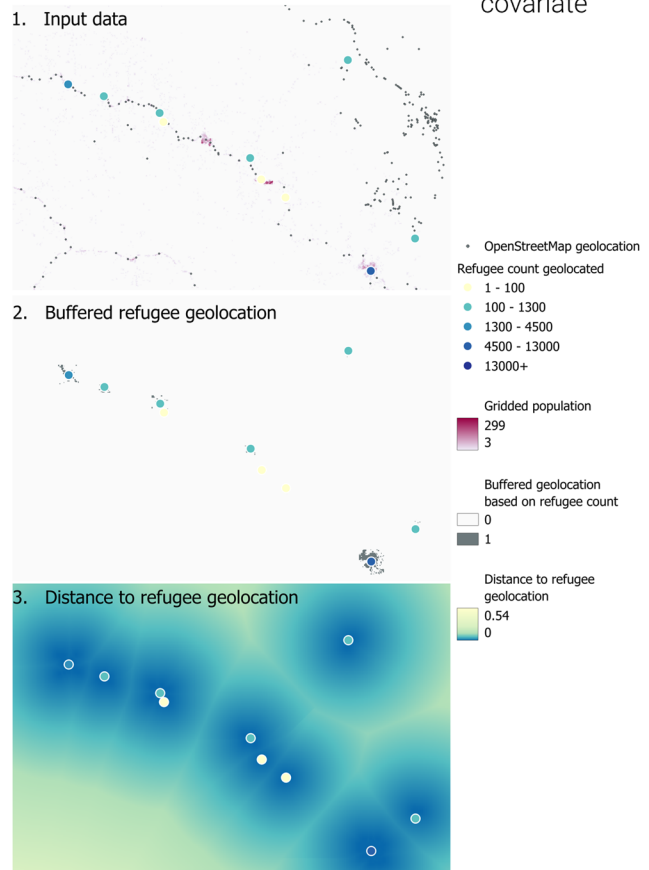


Fig. 4 Geolocating proGres refugee records

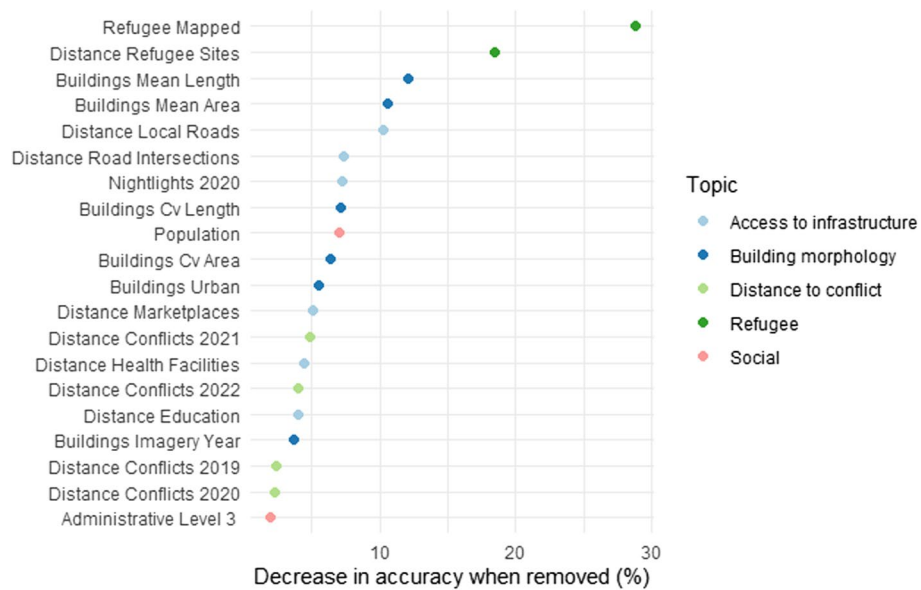
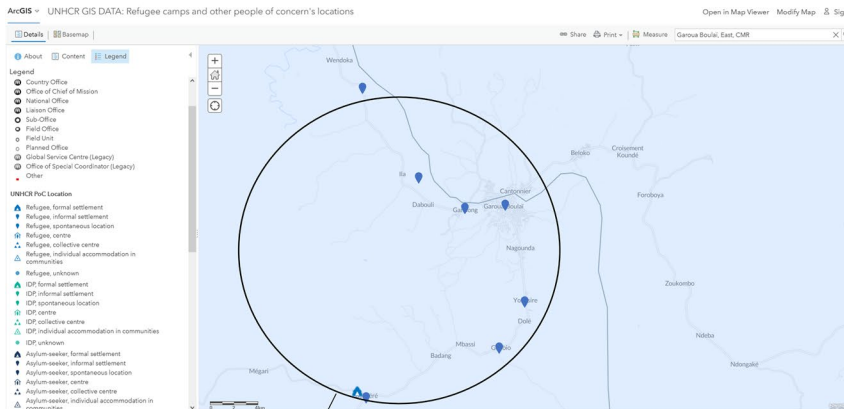
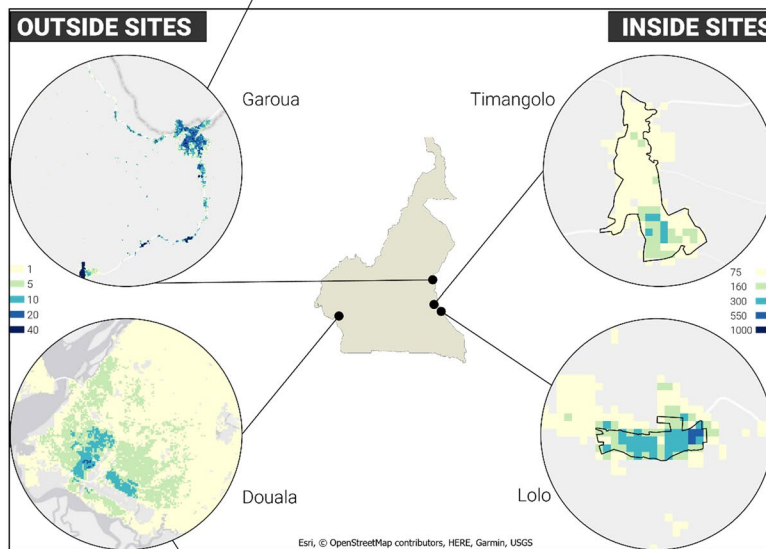


Fig. 5 Covariate importance in modelling refugees. CV stands for coefficient of variation

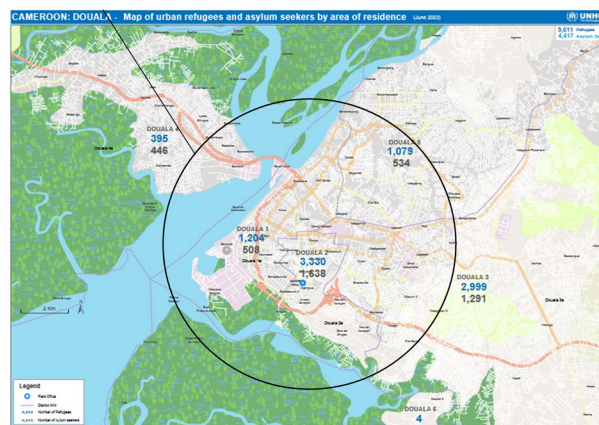
1. UNHCR Interactive Web Map



2. Dasymetric Mapping Result



3. UNHCR Standard Report



**Fig. 6** Comparison of the high-resolution mapping with UNHCR conventional refugee mapping report (United Nations High Commissioner for Refugees Cameroon 2023) and UNHCR interactive mapping (UNHCR 2024)

conflict locations (2022 and 2021) are more important than past conflict locations (2018 and 2019).

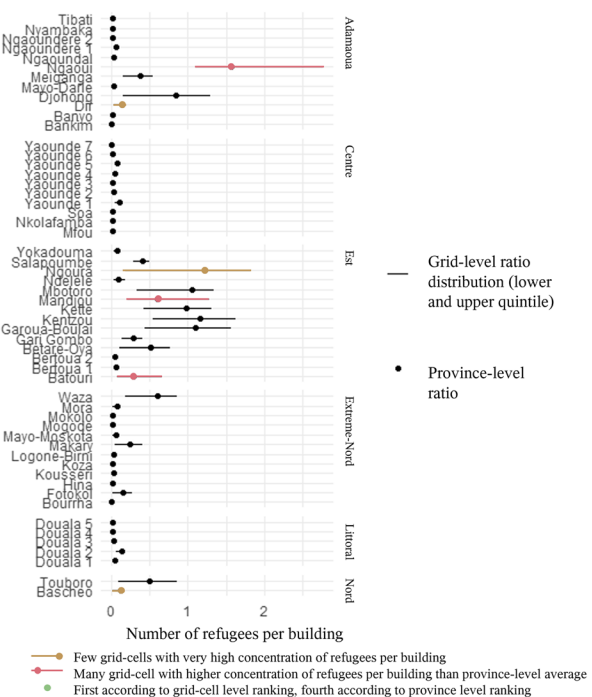
**Mapping refugees across Cameroon**

The final step consists in combining the gridded refugee population inside and outside of monitored sites as represented in Fig. 6 to obtain a comprehensive gridded mapping product of refugees at 100-m spatial resolution. Compared to UNHCR standard static map (United 2023) and UNHCR interactive web-based map (UNHCR 2024) as displayed in Fig. 6, we see a clear gain in visualising spatial variation of refugee locations. For the interactive map, if the precision of the geographical location of refugee is impressive, it is no linked with any extent of refugee settlement or refugee count which drastically reduces the amount of information that can be derived on the refugee population in Cameroon. For the standard static map, first, printing numbers on top of map is not intuitive for comparing count unlike a continuous colour map. Secondly, the UNHCR map displays refugee counts at administrative 3 level (similar as the ones used as input for our fine resolution map) which obfuscates variations at lower spatial scale. For example, the high-resolution map is able to highlight subsections of administrative 3 unit that are less likely to host refugee as pictured in the eastern part of Douala 1. Ultimately, this map comparison effectively illustrates the modifiable

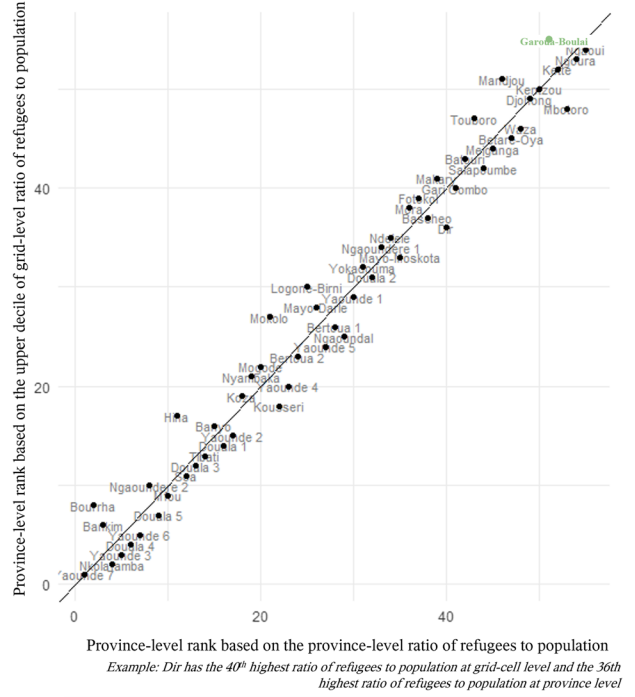
areal unit problem (MAUP) encountered when providing aggregates at administrative level. Indeed, as per administrative-level figures, Douala 2 hosts over twice the number of refugees as Douala 1. However, the high-resolution geographical mapping shows that these refugees are predominantly congregated around the airport, which straddles both Douala 1 and Douala 2. The high-resolution map thus provides a more nuanced understanding of the spatial distribution of refugees, key when planning the spatial allocation of resources to support the humanitarian response.

If Fig. 6 illustrates the visual gains resulting from the fine-scale mapping of the refugee population, Fig. 7 showcases two instances of its analytical gains specifically focusing on refugees located outside UNHCR-monitored sites, where less information is traditionally available. The left panel summarises insights gained by combining the gridded refugee map with Ecopia building footprints layer, allowing the computation of the ratio of refugee per detected structures. This highlights areas in need of infrastructures tailored to the refugee population. If the analysis was already possible at the province level as displayed by the black points on Fig. 6, computing the ratio at grid-cell level shows that the province-level ratio obfuscates large grid-cell level variations. Two scenarios explain the discrepancy between province-level and grid-cell level refugee-to-structure ratio. First, when

**Assessing the ratio of refugees to buildings**



**Assessing the ratio of refugees to host communities**



**Fig. 7** Gain for crisis analytics of spatially disaggregating the refugee subpopulation

the province-level ratio is closer to the upper bound of the grid-cell level range (e.g. in Dir, Ngoura or Bascheo, in light brown on Fig. 7), few small areas in the province have very high refugee concentration that is driving the province-level ratio. Second, when the province ratio is closer to the lower bound of the grid-cell level range (e.g. in Ngaoui, Mandjou or Batouri, in pink on Fig. 7), extended areas in the province have high refugee concentration that is not reflected by the province assessment leading thus to underestimating its vulnerability. The second panel of Fig. 7 articulates the refugee map with the resident population map to discern the pressure of refugee settlements on local host communities outside of monitored sites. It illustrates how a vulnerability ranking based on province-level assessment can differ from a ranking that focuses on the most vulnerable areas within the province. For example, Garoua-Boulai that presents the highest ratio refugees to residents at grid-cell level ranks only fourth lowest per its province-level ratio. Such a divergence has then an impact on the prioritisation of the delivery of assistance in a resource-limited setting.

## Discussion

In the context of large-scale international forced displacement, administrative records and satellite imagery-derived products represent valuable untapped geographic information to address the specific needs of massive refugee population. The proposed high-resolution dasymetric mapping resulting in the disaggregation of refugee totals into grid cell-based maps holds significant potential for harnessing these new data sources to produce countrywide maps of refugees at high-spatial resolution and unlock data-driven-based decision-making processes for local communities.

### Data promise: a constantly updated information source with low cost for repurposing

The initial data assessment enabled by our modelling pipeline involved visually comparing several settlement maps in their ability to identify a very specific settlement type: structures in monitored refugee sites. Our findings first showed the importance of using up-to-date satellite imagery in a context of rapid changes. Secondly, Microsoft building footprints layer, either because of the date of the imagery used or the algorithm implemented, was shown to massively under-detect structures in refugee sites which is not surprising as the coverage limitation of the Microsoft layer has already been reported in the literature (Chamberlain 2024). Lastly, we confirmed the primacy of Ecopia AI fine-grained feature extraction product layer following findings for refugee settlements in Uganda when compared with grid cell-based settlement map (Van Den Hoek 2021).

From this mapping exercise based on novel data from registration systems, we learned that the spatial textual information inputted manually in the refugee register database can be successfully linked with geocoded information. Such information are coined as explicit geo-text data provided in an unstructured form that requires a specific modelling pipeline called geoparsing to generate the correct spatial footprint, involving a spatial gazetteer (Hu 2018). To increase the accuracy of the refugee register geoparsing, the standardisation of the spatial free text needs to improve with more robust data collection system that agrees on naming standards and spatial resolution of the textual information (currently, the textual spatial information in the proGres data system ranges from “Cameroon” to “behind the tree”) (Leidner 2021). Our assessment highlighted the top five provinces (Touboro, Mora, Meiganga, Batouri) that have the poorest match between the register and the spatial gazetteer due to either missing coordinates from OpenStreetMap or lack of standardised naming convention and thus highlighted places where a collaboration between local mapping and local refugee recording team could lead to further cross-fertilisation of the register and the spatial data systems. Geoparsing has broader impact in terms of data management system. For example, the registration database could enable storing alongside the records their spatial information in a spatial format to geocode in real-time refugee residence and allow direct processing and modelling pipeline (Leidner 2021).

This exercise has demonstrated that registration systems can contain a wealth of operations data that can be repurposed efficiently. Refugee administrative data exemplifies the “data revolution” in demography where routinely collected operational data is repurposed for research (Kashyap 2021). The proGres database can be considered as a “ready-made” data source (Salganik 2019), not designed for mapping, which nonetheless offers valuable insights into refugee population sizes and distributions (Masquelier 2017). Other similar undertakings in crisis setting have been to monitor population displacement from social media in Ukraine (Leasure 2023) or through mobile phone data analytics in Turkey (Salah 2019). More broadly, there has been a growing awareness inside refugee-monitoring institutions on the wealth of data routinely generated through operations that could be leveraged to provide socio-demographic insights provided that centralised and standardised data systems are being developed (Ladek 2019).

### Model promise

Dasymetric mapping (as displayed in Fig. 5) that is a non-uniform spatial cartography has a long history as a technique for refining the spatial representation of population

by excluding unpopulated areas from the map. It has been however so far restricted to total population count, sometimes broken down by age and sex (Alegana 2015; Pezzulo 2017; Szarka 2022), or in data-rich context, to map transient population (Martin 2015) or race (Depsky 2022). To our knowledge, however, no population dasy-metric mapping research has focused on the refugee's subgroup in particular, and none has been using register data. Indeed, all high-resolution gridded cartography of social-, economic- or health-related indicators efforts could be previously done only from sampled data by spatial interpolation, for example for vaccination coverage map (Utazi 2019) or directly derived from satellite imagery with the creation of custom indices, for example the ratio between people and night lights to map poverty with the lack of ground-truth validation (Elvidge 2009). In this research in contrast, the administrative data provides us with comprehensive data on refugee totals for the entire country and free text information on refugee residence that supports the geolocalisation of the administrative records. We demonstrated that the dasymetric mapping approach, originally developed for mapping census populations, can be effectively applied to specific subgroups derived from register databases. This is contingent upon the availability of reliable subgroup population totals and the ability to correlate their spatial distribution with high-resolution spatial covariates. Furthermore, the dasymetric approach allowed us to produce for the first time a high-resolution grid-based allocation of refugees across the country which offers a visualisation of the spatial distribution of refugees and more specifically the areas where they concentrate. This spatial format is key for optimising the local delivery of primary services including food, medical support and housing infrastructures. As demonstrated on Fig. 6, estimating the refugee population at grid level reveals local variations that are masked by the admin 3 aggregates. The gridded map of refugees provides a more accurate depiction of their needs and can identify local vulnerability hotspots that would otherwise go unnoticed. Moreover, it facilitates the more effective spatial allocation of humanitarian aid at the subprovincial level. It has also implication when designing household and needs assessment surveys. Indeed, sampling hard-to-reach population is a complex issue tackled mainly by purposeful interviewee selection or network-based sampling methods such as snowball sampling or respondent-driven sampling (MacDonald 2015). By estimating a gridded map of refugee, we can instead adopt survey sampling methods developed for gridded population (Thomson et al. 2020; Qader 2019). Lastly, although the gridded map is based on refugee totals and locations derived from the UNHCR-registered refugee population, the spatial pattern and hotspots area

can also help reach out the non-registered population provided they follow similar settling behaviour.

#### **Model shortcomings**

The random forest approach that we used to link the spatial variation of refugee totals with remote-sensing-derived covariates showed a poorer fit (62% of variance explained) than what is usually obtained when mapping the general resident population (often above 99% variance explained). A first explanation for this poorer fit is that we model a different type of variable that is the number of refugees per inhabitant rather than the number of people per hectare, such that the variations of refugee location that are similar to the country's population are already accounted for. The remaining of the unexplained variance has the following roots: (1) the available spatial covariates are not able to capture all the spatial variations of refugee location, and (2) the sample size available to train the model is smaller than when modelling an entire country's population, as there are only 146 administrative level 3 units in Cameroon that currently have refugees. It was expected that the spatial variation of refugee location, especially self-settled, is harder to capture than the general population because of higher mobility patterns, smaller sizes and more driven by social features that are difficult to capture by remote-sensing approaches, such as the presence of community ties. However, it would be interesting to see if including covariates mapping the operational delivery of services specifically targeting vulnerable populations — which were not available to us — could help increase the explained variance. A last source explaining the poorer fit of the model lies in the quality of the input variable, that is the register, which might vary according to locations. It is nonetheless impossible to assess it with the current data available.

#### **Output technical shortcomings**

We designed a method that is conservative and allocates refugees to all populated grid cells to avoid removing potential refugee location from the map. This approach though results in stretching out the spatial distribution of refugee populations producing a continuous mapping, with many grid cells estimating a very low refugee count (83% have less than one refugee allocated). In addition, the choice for a continuous mapping does not work well when refugee totals are low and administrative area is large. More generally, the map is very dependent on the accuracy of two data inputs: the gridded population map and the refugee registration data. The gridded resident population influences the refugee map through two mechanisms. First, the gridded population defines the spatial extent of the refugee mapping that is the set of grid cells considered as populated, which is dictated itself by the settlement map.

New extent of settlement not detected from the satellite imagery will be missed by the population map and thus by the refugee map if it happens outside of the monitored sites that can be visually inspected. Rapid urban extension nonetheless can be closely linked with the influx of refugee population as seen in the urban development of Amman following the influx of Palestinian refugees (Alnsour 2016). This issue is not prevalent in Cameroon as refugees tend to gather in the centre of the main cities rather than in the outskirts (see the example for Douala in Fig. 6) (United 2023). It is nonetheless possible to integrate in the refugee population model the mapping of rapid settlement expansion as undertaken in North Jordan facing the Syrian refugee influx (Shatnawi 2020). The population map serves also two crucial functions as a denominator: firstly, at the administrative unit level, where it is used to calculate refugee density before fitting the refugee population model, and, secondly, at the grid cell level, where it is utilised to estimate the refugee population. Conversely, inaccuracies in the numerator that is the refugee totals derived from the administrative register will have a direct impact on the accuracy of the dasymetric model if the ratio of the number of refugees per inhabitant is not correct. The quality of the refugee register influences the mapping through a second channel: the accuracy of the recorded geographic information from which is derived the main spatial covariate driving the spatial disaggregation of refugee count.

### Output political shortcomings

Lastly, it is essential to emphasise the sensitivity of the resulting outcome. A map presenting an estimated refugee count within each 100-m grid cell of a country entails significant risks. Firstly, the statistical underpinnings of the modelling do not substantiate precise population size claims at this scale; rather, they provide insight into the relative spatial distribution of the refugee population. Secondly, the map could be utilised in policies that may adversely affect refugee populations. One technical approach involves aggregating the 100-m grid cell counts to create a map with reduced spatial resolution and converting the counts into a bespoke index. An institutional approach entails imposing restrictions on access rights to ensure responsible usage as mentioned in Article 42 of the UNHCR general data protection policy: “UNHCR shall take appropriate measures to identify and assess the potential risks, harms and benefits of automated decision-making and to prevent or mitigate any risks or harm identified for the data subjects” (UNHCR 2022).

### Conclusions

The study shows that the current demographic data revolution has unlocked the possibility for high-resolution mapping of refugee populations through the integration

of administrative data, gridded population data and settlement maps within a grid-based dasymetric mapping framework. The map can then be leveraged to design surveys on refugee living conditions, optimise operations and more broadly inform on areas of need for humanitarian relief due to influxes of incoming refugees.

### Acknowledgements

The authors express their sincere gratitude to the management offices of the WorldPop Research Group and the United Nations High Commissioner for Refugees (UNHCR) for their invaluable support throughout the project. Special appreciation is extended to the dedicated team of UNHCR staff members responsible for meticulously recording the individual information of refugees. Their diligence and dedication significantly contributed to the availability and accuracy of the data essential for this study. We gratefully acknowledge the resources provided by the Leverhulme Centre for Demographic Science and the International Max Planck Research School for Population, Health and Data Science (IMPRS-PHDS). We also thank the reviewers and the editor for their contributions, which significantly improved the paper.

### Authors' contributions

ED prepared and analysed the data, designed and implemented the modelling pipeline and led the writing. AHD prepared the refugee data. AHD, GH, RMK, HP and SQ designed the aims and objectives of the study and supported the interpretation of the results. AJT provided guidance on the broader implications of the study. All authors read, commented and approved the final manuscript.

### Funding

Endorsed team of the Data Innovation Fund 2022–2023 which is a collaborative initiative between UNHCR's Innovation Service, Global Data Service (GDS) and Division of Information and Telecommunications (DIST) aimed to empower UNHCR teams to address challenges and create opportunities to enhance UNHCR's work in data analytics in responsible and appropriate ways.

### Availability of data and materials

The data that support the findings of this study are available from the United Nations High Commission for Refugees, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of the United Nations High Commission for Refugees.

### Declarations

#### Ethics approval and consent to participate

Ethical approval for this study was obtained from the University of Southampton Faculty Ethics Committee under submission ID 80630.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>WorldPop Research Group, University of Southampton, Southampton, UK. <sup>2</sup>Leverhulme Centre for Demographic Science, University of Oxford, Oxford, UK. <sup>3</sup>United Nations High Commissioner for Refugees Regional Bureau for West and Central Africa, United Nations, Dakar, Senegal. <sup>4</sup>United Nations High Commissioner for Refugees Regional Bureau for the Americas, Panama City, Panama. <sup>5</sup>United Nations High Commissioner for Refugees Innovation, United Nations, Geneva, Switzerland. <sup>6</sup>United Nations High Commissioner for Refugees Regional Bureau for Southern Africa, Pretoria, South Africa. <sup>7</sup>Natural Resources Department, College of Agricultural Engineering Sciences, University of Sulaymaniyah, Sulaymaniyah, Kurdistan Region, Iraq.

Received: 22 January 2024 Accepted: 26 August 2024

Published online: 05 October 2024

## References

- Alegana VA, Atkinson PM, Pezzulo C, Sorichetta A, Weiss D, Bird T et al (2015) Fine resolution mapping of population age-structures for health and development applications. *J R Soc Interface* 12(105):20150073. Available from: <https://royalsocietypublishing.org/doi/full/10.1098/rsif.2015.0073>
- Alnsour JA. Managing urban growth in the city of Amman, Jordan. *Cities*. 2016;50:93–9. Available from: <https://www.sciencedirect.com/science/article/pii/S0264275115001225>
- Augusto Hernandez Rocha T, Grapiuna de Almeida D, Shankar Kozhumam A, Cristina da Silva N, Bárbara Abreu Fonseca Thomaz E, Christine de Sousa Queiroz R et al (2021) Microplanning for designing vaccination campaigns in low-resource settings: a geospatial artificial intelligence-based framework. *Vaccine* 39(42):6276–82. Available from: <https://www.sciencedirect.com/science/article/pii/S0264410X2101197X>
- Bing Maps Team. Microsoft has released new and updated building footprints. Microsoft Bing Blogs. 2022. Available from: <https://blogs.bing.com/maps/2022-01/New-and-updated-Building-Footprints>
- Bondarenko M, Kerr D, Sorichetta A, Tatem A (2020) Census/projection-disaggregated gridded population datasets for 51 countries across sub-Saharan Africa in 2020 using building footprints. Univ Southampton, Southampton
- Carr-Hill R. Missing millions and measuring development progress. *World Dev*. 2013;46:30–44. Available from: <https://www.sciencedirect.com/science/article/pii/S0305750X13000053>
- Chamberlain HR, Darin E, Wole AA, Jochem WC, Lazar AN, Tatem AJ (2024) Building footprint data for countries in Africa: to what extent are existing data products comparable? *Comput Environ Urban Syst*. 110:102104. <https://doi.org/10.1016/j.compenvurbys.2024.102104>
- Checchi F, Stewart BT, Palmer JJ, Grundy C (2013) Validity and feasibility of a satellite imagery-based method for rapid estimation of displaced populations. *Int J Health Geogr* 12(1):4. <https://doi.org/10.1186/1476-072X-12-4>
- Chen W, Cheng L, Chen X, Chen J, Cao M (2021) Measuring accessibility to health care services for older bus passengers: a finer spatial resolution. *J Transp Geogr* 93:103068. Available from: <https://www.sciencedirect.com/science/article/pii/S0966692321001216>
- Cutler F original by LB and A, Wiener R port by AL and M. randomForest: Breiman and Cutler's random forests for classification and regression. 2018. Available from: <https://CRAN.R-project.org/package=randomForest>
- Depsky NJ, Cushing L, Morello-Frosch R (2022) High-resolution gridded estimates of population sociodemographics from the 2020 census in California. *PLoS ONE* 17(7):e0270746. Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0270746>
- Dooley CA, Tatem AJ. Gridded maps of building patterns throughout sub-Saharan Africa, version 1.0. WorldPop Research Group, University of Southampton; 2020. <https://doi.org/10.5258/SOTON/WP00666>
- Ecopia AI, Maxar Technologies. Digitize Africa data. 2019. Available from: <https://digitizeafrica.ai>
- Elvidge CD, Baugh K, Zhizhin M, Hsu FC, Ghosh T (2017) VIIRS night-time lights. *Int J Remote Sens* 38(21):5860–5879
- Elvidge CD, Sutton PC, Ghosh T, Tuttle BT, Baugh KE, Bhaduri B et al (2009) A global poverty map derived from satellite data. *Comput Geosci* 35(8):1652–1660. Available from: <https://www.sciencedirect.com/science/article/pii/S0098300409001253>
- Galal H, Dicko A, UNHCR (2023) refugees: UNHCR refugee population statistics database. Available from: <https://www.cran.r-project.org/web/packages/refugees/index.html>
- Herbreteau V, Révillion C, Trimaille E. GeoHealth and QuickOSM, two QGIS plugins for health applications. In: QGIS and Generic Tools. John Wiley & Sons, Ltd; 2018. p. 257–86. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119457091.ch7>
- Hijmans RJ. terra: spatial data analysis. 2023. Available from: <https://CRAN.R-project.org/package=terra>
- Hovy B (2018) Registration – a sine qua non for refugee protection. In: Hugo G, Abbasi-Shavazi MJ, Kraly EP (eds) *Demography of refugee and forced migration*. Springer International Publishing, Cham, pp 39–55. (International Studies in Population). Available from: [https://doi.org/10.1007/978-3-319-67147-5\\_3](https://doi.org/10.1007/978-3-319-67147-5_3)
- Hu Y (2018) Geo-text data and data-driven geospatial semantics. *Geogr Compass* 12(11):e12404. <https://doi.org/10.1111/gec3.12404>
- International Rescue Committee. New responses to the refugee crisis: promises and challenges in Cameroon. International Rescue Committee; 2019 Jun. (New Response to Protracted Refugee Crisis in Cameroon).
- Kashyap R (2021) Has demography witnessed a data revolution? Promises and pitfalls of a changing data ecosystem. *Popul Stud*. 75(sup1):47–75. Available from: <https://doi.org/10.1080/00324728.2021.1969031>
- Ladek S, Zamora NA, Cameron S, Green S, Procter C. Evaluation of UNHCR's data use and information management approaches. 2019
- Leasure DR, Kashyap R, Rampazzo F, Dooley CA, Elbers B, Bondarenko M, Verhagen M, Frey A, Yan J, Akimova ET, Fatehikia M, Trigwell R, Tatem AJ, Weber I, Mills MC (2023) Nowcasting Daily Population Displacement in Ukraine through Social Media Advertising Data. *Popul Dev Rev* 49:231–254. <https://doi.org/10.1111/padr.12558>
- Leidner JL (2021) A survey of textual data and geospatial technology. In: Werner M, Chiang YY (eds) *Handbook of big geospatial data*. Springer International Publishing, Cham, pp 429–457. Available from: [https://doi.org/10.1007/978-3-030-55462-0\\_16](https://doi.org/10.1007/978-3-030-55462-0_16)
- Leyk S, Gaughan AE, Adamo SB, de Sherbinin A, Balk D, Freire S et al (2019) The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use. *Earth Syst Sci Data* 11(3):1385–1409. Available from: <https://essd.copernicus.org/articles/11/1385/2019/>
- Licínio MV, Freitas AC, Evangelista H, Costa-Gonçalves A, Miranda M, Alencar AS (2013) A high spatial resolution outdoor dose rate map of the Rio de Janeiro city, Brasil, risk assessment and urbanization effects. *J Environ Radioact* 126:32–39. Available from: <https://www.sciencedirect.com/science/article/pii/S0265931X13001616>
- Linard C, Tatem AJ (2012) Large-scale spatial population databases in infectious disease research. *Int J Health Geogr* 11(1):7. <https://doi.org/10.1186/1476-072X-11-7>
- Linard C, Alegana VA, Noor AM, Snow RW, Tatem AJ (2010) A high resolution spatial population database of Somalia for disease risk mapping. *Int J Health Geogr* 9(1):45. <https://doi.org/10.1186/1476-072X-9-45>
- Liverman DM, Moran EF, Rindfuss RR, Stern PC (1998) *People and pixels: linking remote sensing and social science*. National Academies Press, Washington DC, p 276
- Logar T, Bullock J, Nemni E, Bromley L, Quinn JA, Luengo-Oroz M (2020) PulseSatellite: a tool using human-AI feedback loops for satellite image analysis in humanitarian contexts. *Proc AAAI Conf Artif Intell* 34(09):13628–9. Available from: <https://ojs.aaai.org/index.php/AAAI/article/view/7101>
- MacDonald AL (2015) Review of selected surveys of refugee population. In Turkey: United Nations High Commissioner for Refugees (UNHCR), Geneva
- Mahabir R, Crooks A, Croitoru A, Agouris P. The study of slums as social and physical constructs: challenges and emerging research opportunities. *Reg Stud Reg Sci*. 2016; Available from: <https://rsa.tandfonline.com/doi/abs/10.1080/21681376.2016.1229130>
- Marconcini M, Marconcini AM, Esch T, Gorelick N (2021) Understanding current trends in global urbanisation - the world settlement footprint suite. *GIForum* 9:33–38. [https://doi.org/10.1553/giscience2021\\_01\\_s33](https://doi.org/10.1553/giscience2021_01_s33)
- Martin D, Cockings S, Leung S. Developing a flexible framework for spatiotemporal population modeling. *Ann Assoc Am Geogr*. 2015;105(4):754–72. Available from: <https://www.jstor.org/stable/24537868>
- Masquelier B, Silva R (2017) Assessing UNHCR registration data as a source of mortality statistics for conflict-affected populations: a case study in Yemen. p 20
- Metzger N, Vargas-Muñoz JE, Daudt RC, Kellenberger B, Whelan TTT, Offli F et al (2022) Fine-grained population mapping from coarse census counts and open geodata. *Sci Rep*. 12(1):20085. Available from: <https://www.nature.com/articles/s41598-022-24495-w>
- Nnanatu C, Yankey O, Abbott T, Gadiaga A, Lazar A, Darin É, Tatem A (2024) Modelled gridded population estimates for Cameroon 2022. Version 1.0. University of Southampton. <https://doi.org/10.5258/SOTON/WP00784>. [Dataset]
- OpenStreetMap contributors. 2023. Planet dump retrieved from <https://planet.osm.org>
- Pebesma EJ (2018) Simple features for R: standardized support for spatial vector data. *R J* 10(1):439
- Pezzulo C, Hornby GM, Sorichetta A, Gaughan AE, Linard C, Bird TJ et al (2017) Sub-national mapping of population pyramids and dependency ratios in Africa and Asia. *Sci Data* 4(1):170089. Available from: <https://www.nature.com/articles/sdata201789>
- Qader S, Lefebvre V, Ninneman A, Himelein K, Pape UJ, Bengtsson L et al (2019) A novel approach to the automatic designation of predefined census enumeration areas and population sampling frames: a case study in Somalia. The World Bank, Washington DC

- Qadir J, Ali A, ur Rasool R, Zwitter A, Sathiseelan A, Crowcroft J (2016) Crisis analytics: big data-driven crisis response. *J Int Humanit Action* 1(1):12. Available from: <https://doi.org/10.1186/s41018-016-0013-9>
- QGIS Development Team. QGIS Geographic Information System. QGIS Association; 2023. Available from: <https://www.qgis.org>
- Quinn JA, Nyhan MM, Navarro C, Coluccia D, Bromley L, Luengo-Oroz M. Humanitarian applications of machine learning with remote-sensing data: review and case study in refugee settlement mapping. *Philos Trans R Soc Math Phys Eng Sci*. 2018;376(2128):20170363. Available from: <https://royalsocietypublishing.org/doi/full/10.1098/rsta.2017.0363>
- Raleigh C, Linke A, Hegre H, Karlsen J (2010) Introducing ACLED: an armed conflict location and event dataset: special data feature. *J Peace Res* 47(5):651–660
- RStudio Team (2019) RStudio: integrated development environment for R. RStudio, Inc., Boston. Available from: <https://www.rstudio.com/>
- Salah AA, Pentland A, Lepri B, Letouzé E (eds) (2019) Guide to mobile data analytics in refugee scenarios: the 'data for refugees challenge' study. Springer International Publishing, Cham. Available from: <http://link.springer.com/10.1007/978-3-030-12554-7>
- Salganik MJ (2019) Bit by bit: social research in the digital age. Princeton University Press, Princeton
- Schiavina M, Melchiorri M, Pesaresi M, Politis P, Freire S, Maffeni L et al (2022) GHSL data package 2022: public release GHS P2022. Publications Office of the European Union, Luxembourg (Available at: <https>)
- Semenov-Tian-Shansky B (1928) Russia: territory and population: a perspective on the, (1926) Census. *Geogr Rev* 18(4):616–640. <https://doi.org/10.2307/207951>
- Shatnawi N, Weidner U, Hinz S (2020) Monitoring urban expansion as a result of refugee fluxes in North Jordan using remote sensing techniques. *J Urban Plan Dev* 146(3):04020026. Available from: <https://ascelibrary.org/doi/10.1061/%28ASCE%29UP.1943-5444.0000584>
- Sirko W, Kashubin S, Ritter M, Annkah A, Bouchareb YSE, Dauphin Y, et al. Continental-scale building detection from high resolution satellite imagery. arXiv; 2021. Available from: <https://arxiv.org/abs/2107.12283>
- Stevens FR, Gaughan AE, Linard C, Tatem AJ (2015) Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLOS ONE* 10(2):e0107042. Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0107042>
- Szarka N, Biljecki F (2022) Population estimation beyond counts—inferring demographic characteristics. *PLOS ONE* 17(4):0266484. Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0266484>
- Tatem AJ (2017) WorldPop, open data for spatial demography. *Sci Data* 4(1):170004. Available from: <https://www.nature.com/articles/sdata20174>
- Thakuriah P (Vonun, Tilahun NY, Zellner M. Big data and urban informatics: innovations and challenges to urban planning and knowledge discovery. In: Thakuriah P (Vonun, Tilahun N, Zellner M, editors. *Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics*. Cham: Springer International Publishing; 2017. p. 11–45. Available from: [https://doi.org/10.1007/978-3-319-40902-3\\_2](https://doi.org/10.1007/978-3-319-40902-3_2)
- Tobler W, Deichmann U, Gottsegen J, Maloy K (1997) World population in a grid of spherical quadrilaterals. *Int J Popul Geogr* 3(3):203–225
- Thomas BEO, Roger J, Gunnell Y, Ashraf S (2023) A method for evaluating population and infrastructure exposed to natural hazards: tests and results for two recent Tonga tsunamis. *Geoenvironmental Disasters* 10(1):4. <https://doi.org/10.1186/s40677-023-00235-8>
- Thomson DR, Rhoda DA, Tatem AJ, Castro MC (2020) Gridded population survey sampling: a systematic scoping review of the field and strategic research agenda. *Int J Health Geogr* 19(1):34
- United Nations High Commissioner for Refugees (2019) Guidance on registration and identity management. Available from: <https://www.unhcr.org/registration-guidance/>.
- United Nations High Commissioner for Refugees (2018) ProGres in partnership. United Nations High Commissioner for Refugees. Available from: <https://www.unhcr.org/registration-guidance/chapter3/registration-tools/>
- United Nations High Commissioner for Refugees Cameroon. Cameroun: Statistiques des personnes déplacées de force (2024) Apr. Available from: <https://data.unhcr.org/en/documents/details/108673>
- United Nations High Commissioner for Refugees. UNHCR. 2023. Refugee data finder. Available from: <https://www.unhcr.org/refugee-statistics/>
- United Nations High Commissioner for Refugees (2024) UNHCR GIS data: refugee camps and other people of concern's locations. 2024. Available at: <https://www.arcgis.com/home/webmap/viewer.html?webmap=24cad2271eaf4219832bf82da5803193>. Accessed 1 June 2024.
- United Nations High Commissioner for Refugees Cameroon (2023) Douala - Map of urban refugees and asylum seekers by area of residence, UNHCR Operational Data Portal (ODP). Available at: <https://data.unhcr.org/en/documents/details/101789>. Accessed 17 August 2023
- United Nations High Commissioner for Refugees (2023a) Global Trends Report 2022. Statistics and Demographics Section; 2023 p. 48. Available at: <https://www.unhcr.org/global-trends-report-2022>. Accessed 6 Oct 2023
- United Nations High Commissioner for Refugees. General policy on personal data protection and privacy. 2022. Available from: <https://www.refworld.org/docid/63d3bdf94.html>
- Utazi CE, Thorley J, Alegana VA, Ferrari MJ, Takahashi S, Metcalf CJE et al (2019) Mapping vaccination coverage to explore the effects of delivery mechanisms and inform vaccination strategies. *Nat Commun* 10(1):1633. Available from: <https://www.nature.com/articles/s41467-019-09611-1>
- Van Den Hoek J, Friedrich HK (2021) 'Satellite-Based Human Settlement Datasets Inadequately Detect Refugee Settlements: A Critical Assessment at Thirty Refugee Settlements in Uganda'. *Remote Sensing*. 13(18):3574. <https://doi.org/10.3390/rs13183574>
- Wickham H (2017) The tidyverse. R Package Ver 1(1):1
- Wu J, Li Y, Li N, Shi P (2018) Development of an asset value map for disaster risk assessment in China by spatial disaggregation using ancillary remote sensing data. *Risk Anal* 38(1):17–30. Available from: <https://doi.org/10.1111/risa.12806>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.