GEOSPATIAL DATA INTEGRATION TO CAPTURE SMALL-AREA POPULATION DYNAMICS

Andrew J. Tatem,[1] Claire A. Dooley,[2] Shengjie Lai,[3] Dorothea Woods,[4] Alex Cunningham[5] and Alessandro Sorichetta[6]

Summary

In this chapter, we highlight the importance of small-area data on population distributions for supporting policymaking. We emphasize how population distributions vary in different ways at different spatial and temporal scales. Various "big" data sets now exist to capture some of these dynamics, each with their own strengths, but also many drawbacks. We discuss how harmonizing and integrating data sets into a common geospatial framework enables the strengths of different data sets representing features of mobility and migration to be brought together, building on each other. We provide an overview of data sets and methods for such integration, then present three illustrative case studies where such integration has been used to support decision-making.

Background

Data on population counts for small areas underlie almost all areas of governance, policymaking and resource allocation. Knowing accurately how many people reside in an area in a certain time period is important for efficient response to natural disasters, ensuring sufficient aid is delivered equitably and that representation in parliaments is fair. Without such data, children can be missed in vaccination programs, election planning becomes challenging, and infrastructure development does not meet needs. However,

[1] Andrew J. Tatem is Spatial Demography and Epidemiology Professor at the University of Southampton. He is Founder and Director of WorldPop (www.worldpop.org), an applied research group focused on the integration of geospatial data for mapping population distributions, demographics and dynamics at small-area scales.

[2] Claire A. Dooley is a quantitative researcher working in the area of demography and health. Her work involves the development and application of spatial methods to understand displacement patterns and estimate health indicators for populations impacted by conflict and protracted crises.

[3] Shengjie Lai has been long engaged in interdisciplinary research on human mobility, environmental changes and epidemics, to provide an improved evidence base for development and infectious disease control decision-making. His work has been published in scientific journals such as *Nature*, *Science*, the *BMJ*, the *Lancet Infectious Diseases* and *Nature Human Behaviour*.

[4] Dorothea Woods is a geographic information systems trainer and analyst who uses a range of methods to add a spatial component and gain new insights into projects in the ecology and social sciences field. Her key interest lies in migration and human mobility, where she works with novel data sets and big data.

[5] Alex Cunningham has a background in geographic information systems and remote sensing, and has spent recent years focused on identifying, processing and generating key statistics from a wide range of spatio-temporal data sets to aid projects exploring seasonal population movements.

[6] Alessandro Sorichetta is Associate Professor in Applied Geospatial Research at WorldPop, University of Southampton. His research in the field of spatial demography focuses on mapping the global distribution of the human population, including its demographic and socioeconomic characteristics, and modelling human migration and mobility.

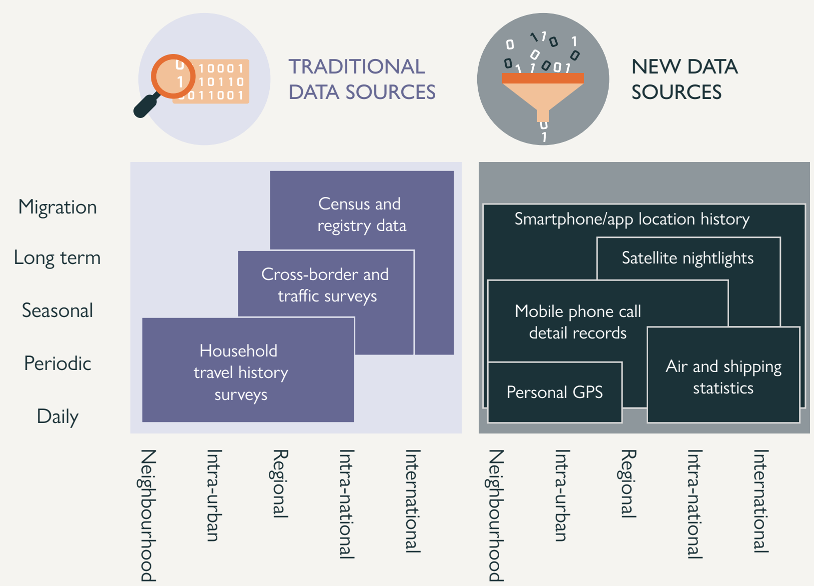
population numbers change substantially over time and space, and regular enumerations at small-area scales can be prohibitively expensive and logistically challenging.

National population and housing censuses typically provide the most comprehensive, complete and accurate source of data on residential population numbers, characteristics and some information on residence changes over the prior 1-to-5-year period. Such data-collection exercises are typically the largest peacetime operations that governments engage in and therefore are normally undertaken just once every decade. In some resource-poor settings, the gaps between such data-collection efforts can be many decades. Between these enumerations, population distributions can change substantially – for example, through daily commutes, seasonal migrations or longer-term displacement. The irregularity of census-based enumeration means that these changes are generally not captured, and estimating them accurately for small areas through standard projection methods is challenging. To capture such population dynamics at small-area levels therefore requires consideration of alternative data sources that can more regularly register changes.

Data sets to support small-area mapping of population dynamics

The past decade has seen the rise of novel, big data sets that capture aspects of population dynamics at a level of richness previously unavailable. Figure 1 highlights the spatial scales and temporal frequencies of population movements that "traditional" and "novel" methods of data collection cover. Each has its own set of weaknesses and biases, and it is often only through linking data sets that some of these weaknesses are mitigated, as they build on the strengths of each other. An overview of these different data sources is provided here; a more detailed description of each can be found in the work of Tatem (2014).

Figure 1. Sources of data for measuring population mobility



Note: Sources of data for measuring population mobility covering temporal (y-axis) and spatial (x-axis) scales of movement, broken down by (a) traditional sources and (b) novel or big data sources. The plot highlights how novel forms of data cover a wider range of spatial scales and temporal frequencies, particularly those derived from mobile phones. (Adapted from: Tatem, 2014)

Traditional sources of data

Figure 1 highlights that methods and data sets for capturing human mobility information from neighbourhood to international scales and from daily to long-term frequencies have been available for decades. However, these methods and the resulting data are typically undertaken only irregularly and can be expensive. National population and housing censuses have been implemented by most countries every decade for half a century, and they often represent the largest and most expensive peacetime operations that governments have undertaken. They typically provide data on changes in residence over the 1-to-5-year period prior to the census. Such data are incredibly valuable for understanding domestic and international migration patterns at small spatial scales and demographics. Their implementation just once a decade (sometimes much less frequent in resource-poor settings) means that data for the intervening years are lacking and little data on shorter-term movements are collected. Registries and administrative data can fill these gaps, but data-collection systems are incomplete or lacking entirely for many countries, particularly across low-income regions such as sub-Saharan Africa.

Surveys continue to play a vital role in our measurement and understanding of population mobility. These can be geographically targeted at specific points of interest, such as border-crossing points, airports or other high-traffic points to obtain data on flows and the characteristics of travellers. Moreover, household surveys often provide a valuable source of population-level information on mobility patterns, together with demographic breakdowns and a range of other factors, including motivation for travel. The major limitation of survey data, however, is the expense in collection, meaning that rich data exist for specific time points and locations, but regular collection over vast geographic areas and populations is difficult. Moreover, designing sampling methods to ensure that migrants, nomads and other mobile populations are well represented can be challenging.

Novel sources of data

A much wider range of spatial scales and temporal frequencies of movements are captured by the novel (or often termed “big”) data sets shown in Figure 1 – such as those from mobile phones, satellite imagery and air-traffic statistics. Some suffer from the same limitations as surveys, particularly the cost and logistics of producing large-area insights over long time periods, such as the use of wearable GPS trackers (Vazquez-Prokopec et al., 2009). However, some benefits are clear, including overcoming some survey limitations, such as recall bias, to provide detailed insights into individual mobility patterns (Floyd et al., 2020; Searle et al., 2017). Air and shipping statistics share some similarities with cross-border and traffic surveys. Still, today big data sets on individual air-travel tickets or GPS tracks of each ship globally are available, forming incredibly rich information on population movements across the world. These register the movements of individuals taking certain forms of transport, but the measurement stops once an individual leaves a port or airport. One form of data that has shown value recently in providing a surrogate for seasonal variations in urban populations in resource-poor settings is night-time satellite images. Across the world, movements can be highly seasonal due to holidays or seasonal labour migration. In the Sahel region of West Africa, this is especially the case in the dry season, when thousands of agricultural labourers move to cities to find alternative work. The switching on of lights and lighting of fires are visible in night-time images and have proven to be a valuable indicator of timings and magnitudes of population mobility and migration in the absence of other data (Bharti et al., 2011, 2016).

The largest area of Figure 1 covered by individual data sources derives from the growing use of mobile devices globally. Call detail records (CDRs) are maintained by mobile phone network operators for billing purposes. They include a phone device ID, the time of communications sent or received, and the cell tower that the communications were routed through. With regular communications to/from a phone and cell tower locations, the approximate movement of the phone can be mapped. Across the millions of phone users, such data have provided a richness of mobility metrics never seen before, covering entire countries at fine spatial and temporal scales. Significant biases exist, with children, the poor, the elderly and some geographical areas often poorly represented, and cross-border measurements typically not possible; but valuable insights can often be obtained despite these (Wesolowski et al., 2013b). CDRs, however, often remain difficult to access, with legally binding agreements with network operators typically required; also, complex data access set-ups are at times necessary due to the sensitivity of such data and the commercial value to the operators. Finally, the world's shift towards smartphones means that GPS location histories obtained by tech companies and app developers are becoming increasingly representative of movement patterns as they capture growing proportions of the population. The value of Google Location History data for capturing mobility across multiple countries over long periods of time has been demonstrated (Ruktanonchai et al., 2018; Kraemer et al., 2020), while similar aggregated and anonymized data sets are being made available by Facebook and Apple to support COVID-19 response efforts.

Geospatial data integration

Each of the data sets outlined above and shown in Figure 1 is typically collected independently, with the comparison and integration with other forms of mobility data often not explicitly planned. Consequently, these data come in all shapes and sizes, making their integration challenging. However, one common feature of all of them is geographical information. The movements captured can generally be mapped to specific administrative units, towns, villages or even precise GPS locations. This provides a framework to overlay and link data sets together, enabling comparisons as well as the development of analyses and models that draw on insights from multiple types of mobility data. Significant effort and care are often required to ensure comparability, but multiple studies have endeavoured to uncover relationships and strengthen mobility insights through linking data sets. These include census migration and CDRs (Wesolowski et al., 2013a), smartphone location history and travel history surveys (Ruktanonchai et al., 2018), and CDRs and travel history surveys (Wesolowski et al., 2014). The following sections dive more deeply into three examples of the integration of geospatially referenced mobility data with other forms of geospatial data to tackle specific policy-relevant challenges: (a) the integration of population displacement surveys with census and satellite data to map contemporary population distributions; (b) linking mobile phone CDRs with disease transmission suitability mapping to prioritize intervention delivery; (c) developing spatial disease spread models driven by smartphone location histories and CDRs to guide policy in outbreak response.

Case studies

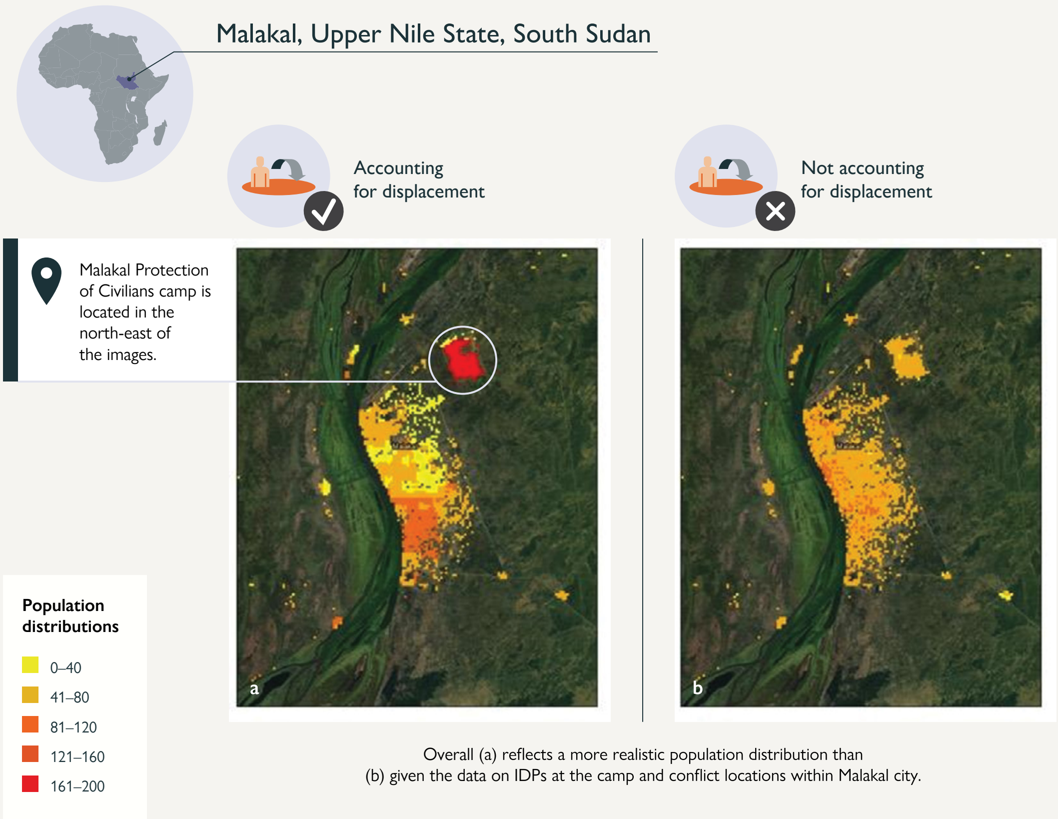
Accounting for displacement in population mapping in South Sudan

South Sudan's last census was conducted in 2008, prior to its independence from the Sudan in 2011. Based on these population counts, South Sudan's National Bureau of Statistics (2015) produces subnational population projections using fertility and mortality rates. However, the population projections do not account for population movement due to the lack of validated migration data. Not accounting for movement is incredibly problematic because since independence, South Sudan has experienced widespread conflict that has led to the displacement of many people across the country as well as into neighbouring territories. Additionally, annual flooding events are common and, depending on their magnitude, often cause displacement. Round 9 of IOM's Displacement Tracking Matrix (DTM) (carried out between July and September 2020) reports that there are approximately 1.6 million internally displaced persons (IDPs) across South Sudan, and the Office of the United Nations High Commissioner for Refugees (UNHCR) reports that there are approximately 2.2 million South Sudanese refugees residing in Uganda (40.4%), the Sudan (33.3%), Ethiopia (16.6%), Kenya (5.6%) and the Democratic Republic of the Congo (4.1%), as of October 2020 (IOM, 2020, 2021; UNHCR, 2020).

To generate an accurate population distribution that accounts for displacement, it is necessary to integrate a number of different displacement-related data sources. In recent years, the spatial coverage of IOM's DTM has expanded, with the most recent rounds including a large proportion of the country. The DTM data contain geolocated destinations of IDPs as well as the county (administrative level 2) from which the majority of IDPs at each destination location have been displaced. The UNHCR's data portal does not provide any information about the subnational place of origin of refugees; however, the Regional Intention Survey includes the approximate proportion of refugees from each state (administrative level 1) within South Sudan based on sampled households across 15 refugee camps: the Democratic Republic of the Congo (3 camps), Ethiopia (3 camps), Kenya (3 camps), Uganda (3 camps), the Sudan (2 camps) and the Central African Republic (1 camp) (UNHCR, 2019). IOM's data on destinations of IDPs can be used directly to map displaced populations and adjust census projections to reflect the corresponding areas of increased population size. The challenge comes when attempting to map the locations of depleted population sizes compared to the census projections. From IOM and UNHCR data on place of origin by administrative units, disaggregation approaches can be applied to infer estimated numbers of people displaced from all areas of the country at a high spatial resolution. For disaggregation to be effective and accurate, appropriate predictors of place of origin are needed. For South Sudan, the key drivers of displacement are conflict and flooding, and therefore variables such as distance to fatal conflict events and precipitation levels during rainy seasons are required. These types of variables have been derived by applying geocomputing methods to spatial data sets such as the Armed Conflict Location and Event Data Project (ACLED) (Raleigh et al., 2010) and WorldClim's monthly weather measurements (Fick and Hijmans, 2017). The resulting disaggregation using these geospatial variables successfully infers larger-scale displacement from locations of intense conflict and flooding.

With estimates of displacement mapped by origin and destination, subnational census projections can be adjusted to produce a more realistic population distribution. The 2019 South Sudan population data set produced by WorldPop (2020) (as part of the GRID3 project)[7] using this approach, integrated with building footprints mapped from recent satellite imagery, provides grid cell-level (~100m x 100m) population estimates that account for displacement (Ecopia AI and Maxar Technologies, 2020; Dooley et al., 2020, 2021a, 2021b). The high spatial resolution of this data set allows the detailed level of information about numbers and locations of IDPs to be explicitly included and not lost through aggregation to larger spatial units, while still providing flexibility to data users to perform aggregations for their own spatial units – e.g. health catchments, settlements. In Figure 2, we highlight the added benefit of including displacement in population estimates by comparing the GRID3 2020 South Sudan estimates to disaggregated census data that do not account for displacement.

Figure 2. Population distribution differences accounting for displacement



Overall (a) reflects a more realistic population distribution than (b) given the data on IDPs at the camp and conflict locations within Malakal city.

Note: A comparison of two population data sets in Malakal, Upper Nile State, South Sudan. The Malakal Protection of Civilians camp is located in the northeast of the images. The maps show disaggregated census projections that (a) account for displacement (Dooley et al., 2021a) and (b) do not account for displacement (Bondarenko et al., 2020). Overall, Figure 2a reflects a more realistic population distribution than Figure 2b, given the data on IDPs at the camp and conflict locations within the city of Malakal. (Service layer credits: Esri, DigitalGlobe, GeoEye-1, Earthstar Geographics, French Space Agency (CNES)/Airbus DS, United States Department of Agriculture, United States Geological Survey, AeroGRID, Institut national de l'information géographique et forestière (IGN)).

These maps are for illustration purposes only. The boundaries and names shown and the designations used on these maps do not imply official endorsement or acceptance by the International Organization for Migration.

[7] More information is available at www.grid3.org.

Mapping population mobility in Namibia to support malaria-elimination efforts

The world has made great strides towards malaria eradication over recent decades. Many countries on the edge of the disease's endemic range have achieved national elimination through a combination of interventions, together with the impact of factors such as urbanization and poverty alleviation. Namibia in Southern Africa is one country that has made significant progress towards malaria elimination over the past two decades, with annual case numbers dropping from a level of hundreds of thousands to around or below the ten-thousand mark.

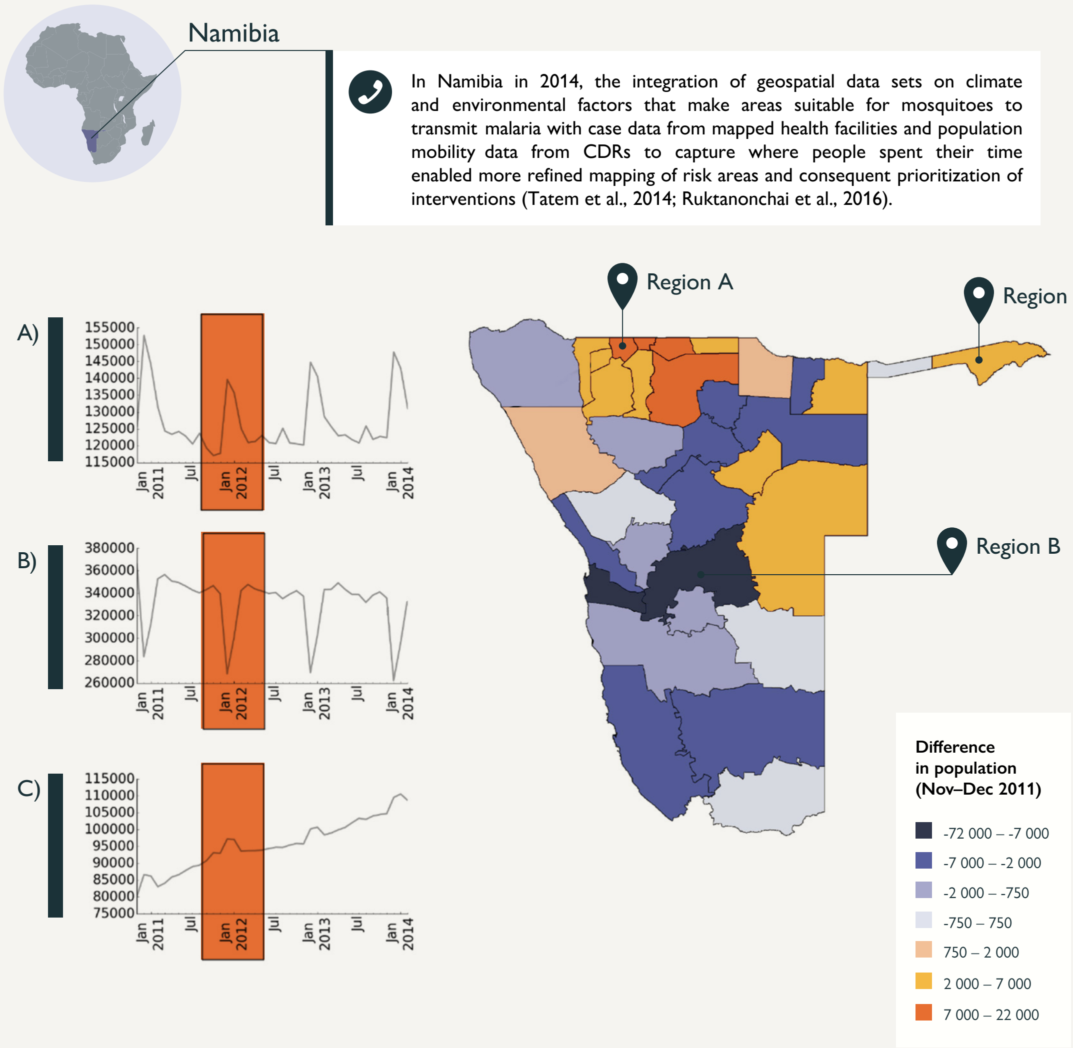
Once malaria case numbers become small and a country is aiming for elimination of the disease, the importation of infections and targeting of interventions become high priorities. Malaria transmission typically occurs in places where densities of mosquitoes are high enough, and where these interact with a population in evenings/night-time, when biting occurs. Surveillance data from health facilities may highlight clusters of cases, but those infections may have resulted from being bitten by infectious mosquitoes far from the facility. Understanding where the transmission took place can help identify hotspots of transmission and, therefore, tailor interventions accordingly to reach elimination. Mobility data can therefore play an important role in understanding one component of this – where people spend their time.

In Namibia in 2014, the integration of geospatial data sets on climate and environmental factors that make areas suitable for mosquitoes to transmit malaria, with case data from mapped health facilities and population mobility data from CDRs, to capture where people spent their time, enabled more refined mapping of risk areas and consequent prioritization of interventions (Tatem et al., 2014; Ruktanonchai et al., 2016). The maps showed that some areas of the country only saw cases because of a high proportion of importations from higher transmission areas. Targeting these higher transmission areas would likely have a disproportionate effect on overall case numbers and be a more effective use of limited resources. The maps were used by the National Vector Borne Disease Control Programme to prioritize bed net distribution to the areas and people most likely important to the transmission cycle.[8]

Following requests from the World Health Organization (WHO) office in Namibia, the value of the CDRs was further demonstrated through integration with census data in providing estimates of monthly changes in population numbers across the country (zu Erbach-Schoenberg et al., 2016; see also Figure 3). When linked with maps of health facility catchments, estimates of seasonal changes and demands for health interventions were obtained, as well as new denominators for health metrics. Finally, the strong correlations between census-based migration flows and phone-based estimates demonstrated the value of CDRs for more regular updates to national migration statistics (Lai et al., 2019).

[8] More information is available at https://dataimpacts.org/project/malaria/.

Figure 3. Mapping seasonal population changes using mobile phone data



In Namibia in 2014, the integration of geospatial data sets on climate and environmental factors that make areas suitable for mosquitoes to transmit malaria with case data from mapped health facilities and population mobility data from CDRs to capture where people spent their time enabled more refined mapping of risk areas and consequent prioritization of interventions (Tatem et al., 2014; Ruktanonchai et al., 2016).

Source: Adapted from: zu Erbach-Schoenberg et al., 2016.

Note: Estimates of seasonal changes in population numbers by region in Namibia based on the integration of census and mobile phone CDRs. The map shows estimated differences in population for health districts between November and December 2011. The inset graphs show predicted population numbers for selected health districts over the January 2011–January 2014 period.

This map is for illustration purposes only. The boundaries and names shown and the designations used on this map do not imply official endorsement or acceptance by the International Organization for Migration.

Incorporating mobility data into COVID-19 transmission modelling

Modern transportation plays a key role in the long-distance and rapid spread of infectious diseases due to the high mobility of human hosts across the globe. In the early stages of the COVID-19 pandemic, there was an urgent need to understand the spread risk and geographic range of COVID-19 transmission, as well as the effectiveness of non-pharmaceutical interventions (NPIs) on COVID-19 containment and mitigation. As it was an emerging disease, vaccines and drugs were not expected to be available in a short period. NPIs – such as case identification and isolation, contact tracing, travel restrictions, physical distancing, face masking, hand-washing, school and workplace closures, and even lockdown (cordon sanitaire) of cities or countries – were implemented across the world. These aimed to reduce transmission, infections and deaths, thereby delaying the epidemic peak, and buying time for health-care preparations and vaccines to be used later on. Prior to COVID-19, however, few studies had systematically investigated the effectiveness of NPIs on various infectious diseases as well as how they should be implemented across space and time to mitigate a pandemic's negative effects and protect vulnerable populations at risk of severe outcomes. The lack of relevant evidence led to delays in COVID-19 containment in the early stages of the outbreak, and uncoordinated interventions also reduced the effectiveness of these measures to prevent resurgences. Aggregated and anonymized location history data from smartphones and mobile phone CDRs, however, have provided a vital basis for studies assessing the effectiveness of NPIs in containing COVID-19 transmission (Lai et al., 2020; Ruktanonchai et al., 2020).

A travel network-based epidemiological model was built to estimate the numbers of susceptible, exposed, infectious and recovered/removed subpopulations per day within defined geographical areas in China, as well as the number of infected travellers moving between each pair of study areas. Based on historical and near real-time anonymized and geographically aggregated human mobility data obtained from smartphone users of Baidu's location-based services, and data on delay from illness onset to reporting of cases across the country, the modelling framework was used to reconstruct the transmission dynamics of COVID-19 across 340 prefecture-level cities in mainland China from 1 December 2019 up until 30 April 2020. Moreover, comparable before-and-after analyses were conducted to quantify the relative effect of the three major groups of NPIs: (a) the restriction of intercity population movement, (b) the identification and isolation of cases, and (c) the reduction of inner-city travel and contact to increase social distance (Lai et al., 2020). Additionally, mobility metrics across Europe during the pandemic were derived from anonymized and aggregated Google Location History data, along with Vodafone CDRs, to illustrate how coordinated exit strategies could delay continental resurgence and limit COVID-19 community transmission (Ruktanonchai et al., 2020).

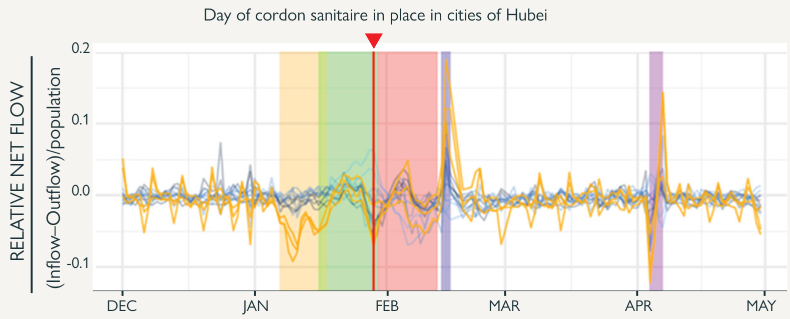
The mobility data were vital to understanding the dynamics of spread, especially in China, where the outbreak coincided with the Lunar New Year holiday (Figure 4), which drives the largest annual movement of people on the planet. The studies estimated that the COVID-19 outbreak in China increased exponentially prior to the Lunar New Year, but the peak of epidemics across the country quickly appeared during the Lunar New Year holiday. Without NPIs, the COVID-19 cases in mainland China would likely have shown a 51-fold (interquartile range 33–71) increase in Wuhan, a 92-fold (58–133) increase in other cities in Hubei, and a 125-fold (77–180) increase in other provinces by 29 February 2020 (Lai et al., 2020). However, the lockdown of Wuhan might not have prevented the seeding of the virus from the city, as the travel ban was put in place at the latter stages of the pre-Lunar New Year population movement out of the city. Nevertheless, if intercity travel restrictions were not implemented, cities and provinces outside of Wuhan would have received more cases from the city, and the affected geographic range would have expanded to the remote western areas of China.

Figure 4. Estimated travel flows in China with and without restrictions at the start of the COVID-19 pandemic

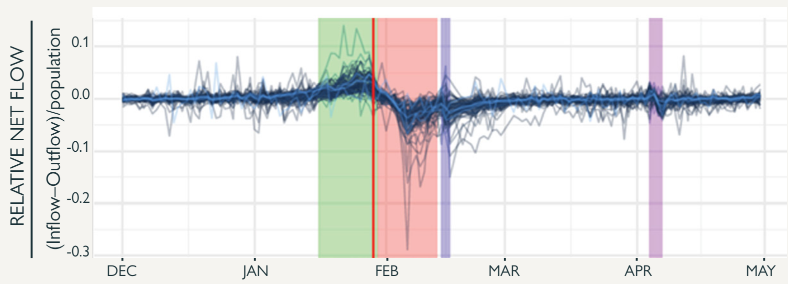
Hubei province, China

The mobility data were vital to understanding the dynamics of spread, especially in China, where the outbreak coincided with the Lunar New Year holiday, which drives the most extensive annual movement of people on the planet. The studies estimated that the COVID-19 outbreak in China increased exponentially prior to the Lunar New Year, but the peak of epidemics across the country quickly appeared during the Lunar New Year holiday.

Wuhan city



Hubei province (excluding Wuhan)



The effectiveness of different interventions varied. Generally, the early detection and isolation of cases were estimated to quickly and substantially prevent more infections than contact reduction and social distancing across the country (5-fold versus 2.6-fold). However, without the intervention of contact reductions, the epidemics would increase exponentially across regions in the longer term. Therefore, combined NPIs achieved the strongest and most rapid effects in terms of COVID-19 outbreak containment. Additionally, these studies also suggested that a resurgent continental epidemic could occur as many as five weeks earlier when well-connected countries with existing stringent interventions end those interventions prematurely. Appropriate coordination of NPIs and vaccine rollouts could significantly improve the likelihood of containing community transmission throughout Europe and prevent resurgences. In particular, synchronizing intermittent lockdowns across Europe meant that half as many lockdown periods were required to end community transmission continent-wide (Ruktanonchai et al., 2020). These studies show that models using mobility data derived from novel, big data sources have significantly improved our understanding of this pandemic and intervention efficacy. A combined, coordinated and timely NPI strategy could substantially reduce

COVID-19 transmission across countries to avoid resurgence, and each study formed part of the evidence base used to guide actions taken by organizations such as the Chinese Center for Disease Control and Prevention (China CDC) and the European Centre for Disease Prevention and Control (ECDC). Given the improving access to timely anonymized population movement data for supporting COVID-19 mitigation across the globe, the potential exists to monitor and assess the effectiveness of NPIs to inform strategies against future COVID-19 waves and potential future pandemics.

Discussion

The value of accurately quantifying population distributions and dynamics at small-area scales is clear. Just focusing on one vital area – as an example, health monitoring and interventions around child mortality in low-income settings – in terms of denominators for health metrics or birth and death registration, it is hard to know whether a national deworming programme for children in Sierra Leone or a vaccination programme for pertussis in Nigeria is reducing mortality when less than 4 per cent of deaths are registered. However, even if 100 per cent of deaths are registered, it remains challenging to implement the programmes and place the number of deaths in context without reliable multi-temporal, disaggregated data on population numbers and distributions, particularly when seasonal dynamics are strong and population groups are highly mobile. Careful surveillance can guide public health in a country and track disease outbreaks that could spread beyond borders. Improving detection and measurement of the numerator without attention to the denominator, however, risks providing an inaccurate picture. The analyses outlined in the Namibia case study showed that improved quantification of seasonal variations in denominator populations changed malaria incidence measures by over 30 per cent (zu Erbach-Schoenberg et al., 2016). Moreover, the reliance on static and ageing figures for denominators leads to the common occurrence of 200 per cent vaccination rates, and/or incidence measures fluctuating by season where population mobility is high (Cutts et al., 2016).

The case studies outlined above highlight how novel and big data sets that capture aspects of human mobility can fill vital data gaps to support decision-making. In each case, the mobility data provide one component of geospatial modelling efforts that draw on multiple sources, building on the strength of each to attempt to overcome gaps, biases and weaknesses. Each data set comes with its own sensitivities, and ethics review board assessments of proposed analyses are vital to ensure the protection of individuals – a single data set may comply with a predetermined standard for anonymization, yet when linked with other geospatially referenced data sets, this can open up risks of reidentification. Understanding and quantifying biases in new forms of data represent an ongoing area of research, and intercomparisons with more traditional forms of data remain vital to understand and account for uncertainties. The new forms of mobility data should not be seen as a replacement for traditional sources – they are more of a complement – and continued investment in census, registry and survey data collection remains vital to anchor insights from big data in reality.

REFERENCES*

Bharti, N., A. Djibo, A.J. Tatem, B.T. Grenfell and M.J. Ferrari
2016 Measuring populations to improve vaccination coverage. *Scientific Reports*, 6.

Bharti, N., A.J. Tatem, M.J. Ferrari, R.F. Grais, A. Djibo and B.T. Grenfell
2011 Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science*, 334(6061):1424–1427.

Bondarenko, M., D. Kerr, A. Sorichetta and A.J. Tatem
2020 Census/projection-disaggregated gridded population datasets, adjusted to match the corresponding UNPD 2020 estimates, for 51 countries across sub-Saharan Africa using building footprints. WorldPop, University of Southampton.

Cutts, F.T., P. Claquin, M.C. Danovaro-Holliday and D.A. Rhoda
2016 Monitoring vaccination coverage: Defining the role of surveys. *Vaccine*, 34(35):4103–4109.

Dooley, C.A., G. Boo, D.R. Leasure and A.J. Tatem
2020 Gridded maps of building patterns throughout sub Saharan Africa, version 1.1. WorldPop, University of Southampton.

Dooley, C.A., W.C. Jochem, D.R. Leasure, A. Sorichetta, A.N. Lazar and A.J. Tatem
2021a South Sudan 2020 gridded population estimates from census projections adjusted for displacement, version 2.0. WorldPop, University of Southampton.

Dooley, C.A., W.C. Jochem, A. Sorichetta, A.N. Lazar and A.J. Tatem
2021b Description of methods for South Sudan 2020 gridded population estimates from census projections adjusted for displacement, version 2.0. WorldPop, University of Southampton.

Ecopia AI and Maxar Technologies
2020 Digitize Africa data.

Fick, S.E. and R.J. Hijmans
2017 WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12):4302–4315.

Floyd, J.R., J. Ogola, E.M. Fèvre, N. Wardrop, A.J. Tatem and N.W. Ruktanonchai
2020 Activity-specific mobility of adults in a rural region of western Kenya. *PeerJ*.

** All hyperlinks were working at the time of writing this report.*

International Organization for Migration (IOM)

2020 South Sudan – Baseline Locations Round 9. Displacement Tracking Matrix (DTM). Available at https://displacement.iom.int/datasets/south-sudan-baseline-locations-round-9.

2021 South Sudan – Baseline Assessment Round 9 – IDP and Returnee. DTM. Available at https://displacement.iom.int/datasets/south-sudan-baseline-assessment-round-9-idp-and-returnee.

Kraemer, M.U.G., A. Sadilek, Q. Zhang, N.A. Marchal, G. Tuli, E.L. Cohn, Y. Hswen, T.A. Perkins, D.L. Smith, R.C. Reiner Jr and J.S. Brownstein

2020 Mapping global variation in human mobility. *Nature Human Behaviour*, 4:800–810.

Lai, S., N.W. Ruktanonchai, L. Zhou, O. Prosper, W. Luo, J.R. Floyd, A. Wesolowski, M. Santillana, C. Zhang, X. Du, H. Yu and A.J. Tatem

2020 Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature*, 585:420–413.

Lai, S., E. zu Erbach-Schoenberg, C. Pezzulo, N.W. Ruktanonchai, A. Sorichetta, J. Steele, T. Li, C.A. Dooley and A.J. Tatem

2019 Exploring the use of mobile phone data for national migration statistics. *Palgrave Communications*, 5.

National Bureau of Statistics

2015 Population projections for South Sudan by county from 2015–2020. Available at www.ssnbss.org/home/document/census/population-projections-for-south-sudan-by-county-from-2015-to-2020/.

Office of the United Nations High Commissioner for Refugees (UNHCR)

2019 Regional Intention Survey of South Sudanese Refugees: Central African Republic, Democratic Republic of the Congo, Ethiopia, Kenya, Sudan, Uganda – June 2019. Available at https://microdata.unhcr.org/index.php/catalog/224/download/674.

2020 Regional overview of the South Sudanese refugee population: 2020 South Sudan Regional RRRP. Available at https://data2.unhcr.org/en/documents/details/79631.

Raleigh, C., A. Linke, H. Hegre and J. Karlsen

2010 Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature. *Journal of Peace Research*, 47(5):651–660.

Ruktanonchai, N.W., P. DeLeenheer, A.J. Tatem, V.A. Alegana, T.T. Caughlin, E. zu Erbach-Schoenberg, C. Lourenço, C.W. Ruktanonchai and D.L. Smith

2016 Identifying malaria transmission foci for elimination using human mobility data. *PLOS Computational Biology*, 12(4).

Ruktanonchai, N.W., J.R. Floyd, S. Lai, C.W. Ruktanonchai, A. Sadilek, P. Rente-Lourenco, X. Ben, A. Carioli, J. Gwinn, J.E. Steele, O. Prosper, A. Schneider, A. Oplinger, P. Eastham and A.J. Tatem

2020 Assessing the impact of coordinated COVID-19 exit strategies across Europe. *Science*, 369(6510):1465–1470.

Ruktanonchai, N.W., C.W. Ruktanonchai, J.R. Floyd and A.J. Tatem

2018 Using Google Location History data to quantify fine-scale human mobility. *International Journal of Health Geographics*, 17.

Searle, K.M., J. Lubinda, H. Hamapumbu, T.M. Shields, F.C. Curriero, D.L. Smith, P.E. Thuma and W.J. Moss

2017 Characterizing and quantifying human movement patterns using GPS data loggers in an area approaching malaria elimination in rural southern Zambia. *Royal Society Open Science*, 4(5).

Tatem, A.J.

2014 Mapping population and pathogen movements. *International Health*, 6(1):5–11.

Tatem, A.J., Z. Huang, C. Narib, U. Kumar, D. Kandula, D.K. Pindolia, D.L. Smith, J.M. Cohen, B. Graupe, P. Uusiku and C. Lourenço

2014 Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. *Malaria Journal*, 13.

Vazquez-Prokopec, G.M., S.T. Stoddard, V. Paz-Soldan, A.C. Morrison, J.P. Elder, T.J. Kochel, T.W. Scott and U. Kitron

2009 Usefulness of commercially available GPS data-loggers for tracking human movement and exposure to dengue virus. *International Journal of Health Geographics*, 8.

Wesolowski, A., C.O. Buckee, D.K. Pindolia, N. Eagle, D.L. Smith, A.J. Garcia and A.J. Tatem

2013a The use of census migration data to approximate human movement patterns across temporal scales. *PLOS ONE*, 8(1).

Wesolowski, A., N. Eagle, A.M. Noor, R.W. Snow and C.O. Buckee

2013b The impact of biases in mobile phone ownership on estimates of human mobility. *Journal of the Royal Society Interface*, 10(81).

Wesolowski, A., G. Stresman, N. Eagle, J. Stevenson, C. Owaga, E. Marube, T. Bousema, C. Drakeley, J. Cox and C.O. Buckee

2014 Quantifying travel behavior for infectious disease research: A comparison of data from surveys and mobile phones. *Scientific Reports*, 4.

WorldPop
2020 South Sudan 2019 gridded population estimates from census projections adjusted for displacement, version 1.0. School of Geography and Environmental Science, University of Southampton. Available at https://eprints.soton.ac.uk/437804/.

zu Erbach-Schoenberg, E., V.A. Alegana, A. Sorichetta, C. Linard, C. Lourenço, N.W. Ruktanonchai, B. Graupe, T.J. Bird, C. Pezzulo, A. Wesolowski and A.J. Tatem
2016 Dynamic denominators: The impact of seasonally varying population numbers on disease incidence estimates. *Population Health Metrics*, 14.

ADDITIONAL READING

Wardrop, N.A., W.C. Jochem, T.J. Bird, H.R. Chamberlain, D. Clarke, D. Kerr, L. Bengtsson, S. Juran, V. Seaman and A.J. Tatem
2018 Spatially disaggregated population estimates in the absence of national population and housing census data. *Proceedings of the National Academy of Sciences of the United States of America*, 115(14):3529–3537.