Epidemics 42 (2023) 100666

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Gaps in mobility data and implications for modelling epidemic spread: A scoping review and simulation study

Jack Wardle [a,1], Sangeeta Bhatia [a,1], Moritz U.G. Kraemer [b], Pierre Nouvellet [a,c], Anne Cori [a,*]

[a] MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, UK
[b] Department of Biology, University of Oxford, Oxford, UK
[c] School of Life Sciences, University of Sussex, Brighton, UK

ARTICLE INFO

Keywords:
Infectious disease spread
Mathematical modelling
Human mobility data
Gravity model
Radiation model

ABSTRACT

Reliable estimates of human mobility are important for understanding the spatial spread of infectious diseases and the effective targeting of control measures. However, when modelling infectious disease dynamics, data on human mobility at an appropriate temporal or spatial resolution are not always available, leading to the common use of model-derived mobility proxies. In this study we reviewed the different data sources and mobility models that have been used to characterise human movement in Africa. We then conducted a simulation study to better understand the implications of using human mobility proxies when predicting the spatial spread and dynamics of infectious diseases.

We found major gaps in the availability of empirical measures of human mobility in Africa, leading to mobility proxies being used in place of data. Empirical data on subnational mobility were only available for 17/54 countries, and in most instances, these data characterised long-term movement patterns, which were unsuitable for modelling the spread of pathogens with short generation times (time between infection of a case and their infector). Results from our simulation study demonstrated that using mobility proxies can have a substantial impact on the predicted epidemic dynamics, with complex and non-intuitive biases. In particular, the predicted times and order of epidemic invasion, and the time of epidemic peak in different locations can be underestimated or overestimated, depending on the types of proxies used and the country of interest.

Our work underscores the need for regularly updated empirical measures of population movement within and between countries to aid the prevention and control of infectious disease outbreaks. At the same time, there is a need to establish an evidence base to help understand which types of mobility data are most appropriate for describing the spread of emerging infectious diseases in different settings.

1. Introduction

Human movement is a key determinant of the spatial spread of infectious diseases as evidenced by the spread of Ebola (Merler et al., 2015; Yang et al., 2015), yellow fever (Dorigatti et al., 2017; Kraemer et al., 2017), and most recently, SARS-CoV-2 (Tatem et al., 2006; Findlater and Bogoch, 2018). The emergence and subsequent international spread of the latter and its variants of concern (VOCs) highlight the speed and scale at which infectious pathogens can spread around the globe (Li et al., 2021; O'Toole et al., 2021; Viana et al., 2021).

Models of infectious disease dynamics that incorporate human movement have been used to generate insights into the spatial spread of pathogens. They can help identify locations that are most susceptible to receive imported cases (Bogoch et al., 2015; Gilbert et al., 2020; Craig et al., 2020), assess the effectiveness of travel restrictions (Chinazzi et al., 2020; Hollingsworth et al., 2006), and inform the allocation of scarce resources such as vaccines, drugs, personal protective equipment or specialist healthcare staff (Tuite et al., 2011; Wu et al., 2007; Longini et al., 2005; Kucharski et al., 2015). Spatial models of infectious disease spread typically rely on human population movement data. These can be empirical measurements, for example flight passenger statistics, information from national population censuses, or mobile phone call detail records (Tatem, 2014; Kraemer et al., 2016). More recently, a few high resolution datasets based on mobile phone GPS data have been made available to researchers and public health officials and used to analyse within-country movement patterns during the COVID-19 pandemic (Jeffrey et al., 2020; Ruktanonchai et al., 2020; Ascani et al., 2021).

However, empirical data on human movement are not routinely collected, particularly in low- and middle-income countries (Ramiadantsoa

* Corresponding author.
E-mail address: a.cori@imperial.ac.uk (A. Cori).
[1] Equal contribution.

https://doi.org/10.1016/j.epidem.2023.100666
Received 7 March 2022; Received in revised form 18 November 2022; Accepted 6 January 2023
Available online 12 January 2023

et al., 2022). Moreover, where empirical data are available, these are not always at a spatial and temporal resolution that matches the scale and pace at which pathogens spread (Wesolowski et al., 2016). Although mobile phone usage is increasingly being used to understand movement of populations at a fine resolution, these data are known to be biased in systematic ways by limited smartphone ownership or uneven mobile phone coverage (Wesolowski et al., 2012, 2013b), and typically do not capture international movement. They are also expensive and time-consuming to acquire, limiting their potential use in real-time analysis. Finally, privacy concerns constrain wider sharing of these data (Lai et al., 2019), preventing the reproducibility of analyses based on them.

In the absence of empirical data, infectious disease modelling studies often rely on human movement models such as gravity and radiation models (Kraemer et al., 2019; Zipf, 1946; Simini et al., 2012). These models are sometimes calibrated using data from nearby countries (Kraemer et al., 2017, 2019), long-term movement data from Demographic and Health Surveys (DHS), or observed patterns of disease spread (Bhatia et al., 2021). By definition, these mobility proxies cannot be validated against local human movement data which do not exist. The potential impact of using such proxies to predict the spatial dynamics of infectious diseases has received little attention despite their recurring role in informing policy and resource allocation decisions. Therefore, better understanding both the availability of human mobility data, or lack thereof, and the impact of using mobility proxies on predicting infectious disease spread is key to prioritise future research effort.

The aim of our work was twofold. First, we aimed to characterise the availability of human mobility data that could be informative about the risk of epidemic spatial spread in Africa. We focus on the African continent which over the past decade has experienced numerous large epidemics with significant public health consequences, including multiple epidemics of Ebola (Centers for Disease Control and Prevention, 2022), yellow fever (Africa Centers for Disease Control and Prevention, 2022a), cholera (Africa Centers for Disease Control and Prevention, 2022b), and measles (Africa Centers for Disease Control and Prevention, 2022c). Second, we assessed the extent to which relying on mobility proxies, in the absence of adequate empirical data, would affect model-based predictions of epidemic spread.

We reviewed the different data sources and mobility models that have been used to estimate human movement in Africa. We then conducted a simulation study to better understand the implications of using human mobility proxies when predicting the spatial spread of infectious diseases using dynamic transmission models. In the absence of suitable publicly available mobility data from Africa, and motivated by the frequent use of mobility proxies based on data from other countries in the continent (both of which were findings of our scoping review), our simulation study used high resolution movement data from two Western European countries to explore how proxies based on out-of-country data can impact epidemic models. The simulations were performed with two widely-used forms of human mobility models, the gravity and radiation models.

2. Methods

2.1. Search strategy

We carried out a scoping review of mobility data and models in Africa, adhering to the guidelines established by PRISMA extension for Scoping Reviews (Tricco et al., 2018). We searched PubMED and Web of Science on 29th August 2018 for peer reviewed literature in English on human movement data used in mobility models in Africa. We restricted our search to gravity and radiation models, which are the most commonly used in infectious disease modelling (Viboud et al., 2006; Eggo et al., 2011; Charu et al., 2017; Truscott and Ferguson, 2012; Bharti et al., 2008).

In short, in a gravity model the population movement between two locations is assumed to be proportional to the population sizes of the origin and distance and inversely proportional to the distance between them (Eq. (3)). The radiation model also describes how the movement between two locations may be affected by other surrounding highly populated locations (Eq. (S1)).

The search terms used were: ((gravity model OR radiation model) AND human AND africa), (mobility model AND africa AND human), (gravity model AND africa), (radiation model AND africa AND human). Both abstract and full-text screening were carried out independently by two reviewers and all disagreements were resolved by consensus. Only primary research articles were eligible for inclusion. We excluded studies that were not set in Africa, did not have a spatial component, or did not use data or estimates of human mobility.

Data were systematically extracted from the studies to describe the source (e.g. census), spatial resolution, and temporal resolution of data on human movement. We extracted the location and the time period of the data for each study and whether the data were made available.

We classified as "empirical" human movement data obtained from mobile phone call detail records (CDR), micro-census data from Integrated Public Use Microdata Series (IPUMS), short-term migration data from DHS, and data on long-term migration (Global Bi-lateral Migration Database, hereafter GBMD), or on movement of refugees from the United Nations High Commissioner for Refugees (UNHCR). In addition to these generic data sources, some studies carried out surveys tailored to identify movement patterns relevant to the spread of specific diseases, e.g. overnight stays for malaria. These were also considered as empirical data sources. In contrast to these, a large number of studies used models of human mobility calibrated to data from nearby regions, or to the observed patterns of disease spread. Some studies specifically aimed to generate such mobility estimates for regions where no empirical data were available. Where a study did not rely on empirical measurements to characterise mobility and instead used indirect sources such as disease spread or data from other countries, we refer to it as having used (or produced) "mobility proxies". For those studies, we also extracted details of the mobility models used. For each country included in a study, we considered empirical data (or mobility proxy) for a given spatial resolution and for a specific time period as a discrete data set.

2.2. Simulation study

To understand the implications of using human mobility proxies instead of empirical data when predicting the spatial spread and dynamics of infectious diseases, we carried out a simulation study comparing epidemic outcomes when using empirical movement data versus mobility proxies. Specifically, we focused on subnational mobility proxies generated by calibrating mobility models to data from another country in the same region, an approach that was frequently adopted in the studies identified in the scoping review.

2.2.1. Commuting data

To evaluate the implications of using mobility proxies in modelling epidemic spatial spread requires highly resolved empirical data on movement patterns. With such data, a baseline scenario for disease spread can be simulated using the empirical data and its results compared to counterfactual scenarios based on mobility proxies. Ideally, our simulation study would have used subnational movement data from African countries to understand the extent to which relying on mobility proxies instead of empirical data may affect predictions about epidemic spread in Africa. However, as highlighted by the scoping review, such resolved mobility datasets either did not exist or were not available for use for most African countries. For instance, a CDR dataset that had been released for research (Blondel et al., 2012) and used in multiple studies (Wesolowski et al., 2014; Tompkins and McCreesh, 2016; Mari et al., 2017) was no longer available.

The settings for our simulation study were therefore selected based on locations where empirical mobility data were openly accessible and were not based in Africa. Instead we used available commuting data from France and Portugal. While this approach did not allow us to draw explicit conclusions on the impact of using mobility proxies in epidemic models for African settings, it did highlight general issues with predictions from epidemic models using mobility proxies calibrated with out-of-country data (as well as calibrations to data aggregated at different spatial scales). These valuable insights from the European context are informative on the suitability of using mobility proxies in epidemic models in a more general context.

The movement data used were derived by Tizzoni et al. from mobile phone call data in France and Portugal for 2007 and 2006 respectively (Tizzoni et al., 2014). The commuting patterns were obtained from a large sample of users' CDR by identifying the frequently visited locations (based on proximity to mobile phone towers) of each person in the dataset. The authors assumed that the most visited place corresponded to the location where an individual lives, with the second most visited place being the location where an individual works (Song et al., 2010). This was then used to calculate the total numbers of people in the sample who commute between each of the 323 districts (corresponding to the ADM3 administrative units) in mainland France and the 278 municipalities (the ADM2 administrative units) in mainland Portugal. This approach also gave the number of people that both lived and worked in the same spatial unit.

We combined the commutes between and within administrative units to construct an origin–destination (O-D) movement matrix for each country. To generate population-level estimates of the total movement, we scaled the sampled movement flows between pairs of administrative units. The total number of individuals living in spatial unit i and working in unit j is calculated as:

$$N_{ij}^{obs} = \frac{n_{ij}}{\sum_j n_{ij}} pop_i \tag{1}$$

where n_{ij} is the number of people commuting from i to j in the sample of mobile phone users, $\sum_j n_{ij}$ is the total number of people living in i that were included in the mobile phone data sample and pop_i is the population of location i. The scaling method in Eq. (1) ensures that the sum of each row in the O-D matrix equals the population living in that administrative unit ($\sum_j N_{ij}^{obs} = pop_i$).

2.2.2. Mobility proxies

In order to assess the impact of using mobility proxies in epidemic models, we constructed predicted movement matrices at the same spatial resolution as the observed data in France and Portugal. Movement matrices can encode either the total number of individuals (referred to as O-D matrix) or the probability of moving between locations (rescaled O-D matrix). For the latter, matrix entry p_{ij} is the probability that an individual moves between spatial units i and j. This is the product of the probability that (i) they leave i for any destination ($1 - p_i^{stay}$); and (ii) they move to j given that they have moved out of i (for which we use the notation p_{ij}^{move}). This overall probability can be written as:

$$p_{ij} = \begin{cases} p_i^{stay} & \text{if } i = j \\ (1 - p_i^{stay}) p_{ij}^{move} & \text{if } i \neq j \end{cases} \tag{2}$$

where p_i^{stay} is the probability that a person living in i also works in that same location.

Predictions of the relative flows p_{ij}^{move} were obtained from the gravity model, which posits that the flow of individuals from location i to location j is proportional to:

$$p_{ij}^{move} \propto \frac{N_i^{\alpha} N_j^{\beta}}{d_{ij}^{\gamma}} \tag{3}$$

where N_i and N_j are the populations living in locations i and j, and d_{ij} is the distance between the two locations (Riley et al., 2015; Zipf, 1946). α, β and γ are model parameters. We fitted the gravity model to observed movement data and then used the estimated parameters to predict p_{ij}^{move}.

The gravity model in Eq. (3) cannot be used to estimate p_i^{stay}. Therefore, p_i^{stay} values were taken directly from observed data and used to populate the diagonal elements of the p_{ij} matrix and to scale the off-diagonal elements as shown in Eq. (2).

We multiplied the rescaled O-D matrix by the size of the origin population (pop_i) to obtain the numbers of people predicted to move between pairs of spatial units (O-D matrix):

$$N_{ij}^{pred} = p_{ij} pop_i \tag{4}$$

We generated different mobility proxies N_{ij}^{pred} by varying assumptions about the availability of movement data in a country and about p_i^{stay}. In our central scenario presented here, we assume that movement data are unavailable for a given country and we therefore make predictions based on data from a nearby country (Fig. 1). For Portugal, we generate mobility proxies using data from France, i.e. p_{ij}^{move} estimated by a gravity model fitted to France and p_i^{stay} set to the average observed p_i^{stay} across all French administrative units. The same approach was used to generate mobility proxies for France using data from Portugal.

We explored additional scenarios where p_i^{stay} and p_{ij}^{move} were informed by data from either the correct or the neighbouring country and at various spatial resolutions (e.g. gravity model fitted to ADM2 but used to predict movement at ADM3 in the same country) (Suppl Tab. S1). To explore the robustness of our results to the choice of mobility model, we also considered a scenario using a radiation model to predict p_{ij}^{move}.

2.2.3. Epidemic model

We use a stochastic discrete time SEIR metapopulation model to simulate epidemics in France and Portugal. The subpopulations in the metapopulation model are formed from each combination of home and work location. The size of each subpopulation N_{ij} is the number of individuals who live in i and work in j. N_{ij} is set using either the observed data (N_{ij}^{obs}) or mobility proxy (N_{ij}^{pred}). This means that the home and work locations of individuals are fixed within our model, with movements occurring recurrently. People in each subpopulation fall into one of four compartments denoting their infection states: susceptible (S), exposed (E), infectious (I) or recovered (R).

Each subpopulation has two sets of interactions with other subpopulations. For half of each day (a day being defined as a 24-hour period), we assume homogeneous mixing between all individuals who reside in the same spatial unit (i.e. all people living in i) and for the other half of the day we assume homogeneous mixing between individuals who work in the same spatial unit (all people working in j). These assumptions are in line with previous studies that incorporated commuting data in models of infectious disease spread (Keeling et al., 2010; Tizzoni et al., 2014).

Thus there are two forces of infection acting each day on every subpopulation in our model:

$$\lambda_i^{home} = \beta \frac{\sum_j I_{ij}}{\sum_j N_{ij}}$$
$$\lambda_j^{work} = \beta \frac{\sum_i I_{ij}}{\sum_i N_{ij}} \tag{5}$$

where β is the per capita transmission rate. λ_i^{home} therefore depends on the number of infectious people living in the same spatial unit i ($\sum_j I_{ij}$). Similarly, λ_j^{work} depends on the number of infectious people working in the same spatial unit j ($\sum_i I_{ij}$). We make the simplifying assumption that movement is independent of infection state and that β is the same in all settings.

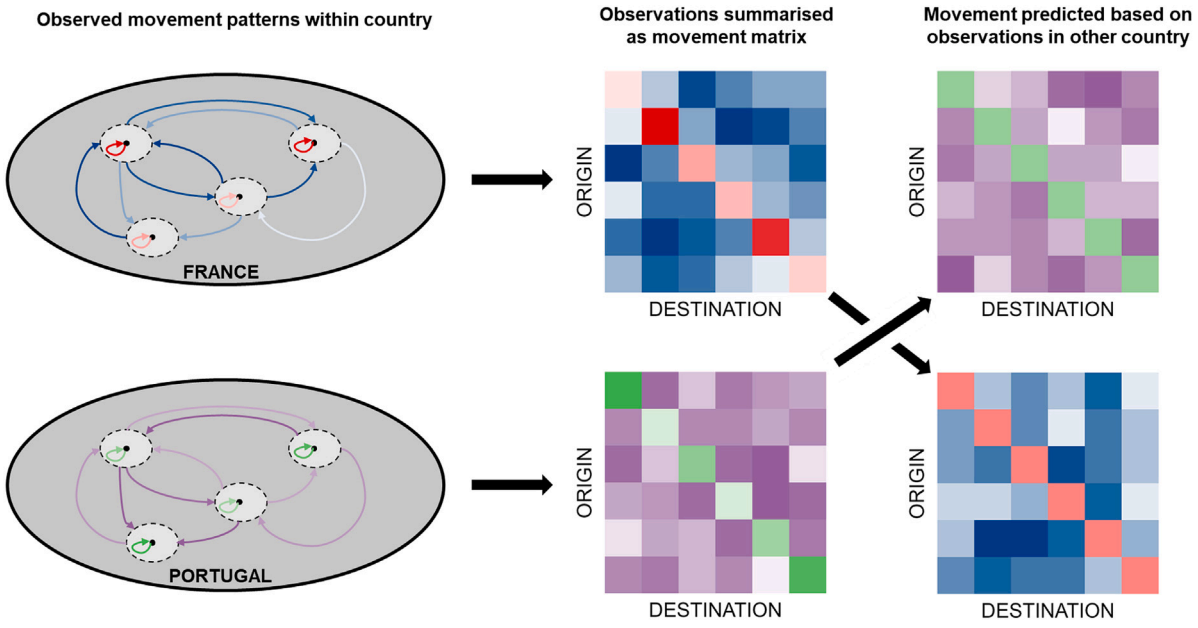

Fig. 1. Illustration of how movement patterns observed in France and Portugal can be used to generate mobility proxies. The ovals on the left represent the movement patterns within each country, with the smaller dashed ovals representing different spatial units within the country. Movement between locations is shown by the blue arrows in France and purple arrows in Portugal. Darker colours indicate more movement. People that move within their own spatial unit are shown with red arrows in France and green arrows in Portugal. These movement patterns can be summarised in a movement matrix (i.e. O-D matrices, middle column) that provide information on movements between each origin–destination pair (including where the destination is the same as the origin, i.e. p_i^{stay}, see methods). The matrices in the right column illustrate how movement patterns in France are predicted in our central scenario based on a combination of a gravity model fitted to Portuguese movement data along with data on the probability that an individual leaves the location they live in (1 - p_i^{stay}). Similarly, we predict Portuguese movement using a gravity model fitted to French movement data and the average p_i^{stay} observed in France. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The transitions between infection compartments at each time step ($\Delta t = 0.1$ day) are modelled stochastically as follows:

$$\begin{aligned} S_{ij} \rightarrow E_{ij} &\sim Binomial\left(S_{ij}, 1-\exp\left(-\lambda \Delta t\right)\right) \\ E_{ij} \rightarrow I_{ij} &\sim Binomial\left(E_{ij}, 1-\exp\left(-\delta \Delta t\right)\right) \\ I_{ij} \rightarrow R_{ij} &\sim Binomial\left(I_{ij}, 1-\exp\left(-\gamma \Delta t\right)\right) \end{aligned} \quad (6)$$

where $1/\delta$ and $1/\gamma$ are the mean latent and infectious periods respectively. λ takes the value λ_i^{home} in the first half of each day when people are at home, and λ_j^{work} in the second half of each day when people are at work.

We simulated and compared epidemics in France and Portugal under different scenarios, where N_{ij} was informed by different data. In the baseline scenario, N_{ij} was set to observed data N_{ij}^{obs}, while in the alternative scenarios N_{ij} was based on mobility proxies N_{ij}^{pred} (Suppl Tab. S1). The epidemics were seeded in either the capital city (Paris and Lisbon) or a less central location (Brest in the northwest of France and Miranda do Douro in the northeast of Portugal). These locations were chosen to explore differences in the dynamics of outbreaks originating in a prominent urban centre and a location on the edge of the commuting network.

For each scenario, we ran 100 simulated epidemics for 300 days, seeding 5 infectious cases of a flu-like pathogen and assuming that the rest of the population is fully susceptible (see Suppl Sec. 2 for further details). We restricted our analysis to epidemics which were seeded successfully (defined as simulations where > 10% of the total population was infected after 300 days).

For each scenario s, we summarised all the successfully seeded epidemic simulations with four metrics, two which characterise scenario s only, and two which compare s to the baseline:

- **Invasion time in** i (t_i^{first}), defined as the median time to the arrival of the first infectious case among those living in location i.
- **Peak time in** i (t_i^{peak}), defined as the median time to the epidemic peak in location i (i.e. time when the number of infectious individuals living in i is at a maximum).
- **Relative error in invasion time in** i (ϵ_i^{first}). This was calculated as the difference between t_i^{first} in scenario s (i.e. using mobility proxy) and in the baseline (using the empirical mobility data), divided by the largest invasion times across all locations in scenario s: $\epsilon_i^{first} = \Delta t_i^{first} / max_j\left(t_j^{first}\right)$.
- **Invasion order similarity** (p_n^{first}), defined as the proportion of the first n locations invaded in the baseline scenario that were also among the first n locations invaded in scenario s. The invasion order was defined based on the median invasion time across all simulations with successfully seeded epidemics, i.e. t_i^{first}.

3. Results

3.1. Scoping review

Of the 471 articles from the initial search, 129 were selected for abstract screening and 30 full-text articles for data extraction, all of which were published between 2007 and 2018. Across the 30 studies, we identified 150 empirical human mobility data sets and 168 mobility proxies.

3.1.1. Empirical data

Empirical data were available for 52 of 54 African countries (Fig. 2). For 51 countries, data on long-term migration patterns between countries were available, from GBMD (n = 42 countries) or data on refugee movement from UNHCR (n = 51). The GBMD provides data on international migrations between all countries for each decade in the period 1960–2000 (Özden et al., 2011). The UNHCR collects data on the annual flows of refugees, asylum seekers and the number of internally displaced persons between a country/territory of origin and asylum (UNHCR, 2022).

Out of 54 countries, we found only 17 with subnational mobility data. These were informed by census (n = 14 countries), mobile phone records (CDR, n = 5, Côte d'Ivoire, Kenya, Namibia, Sierra

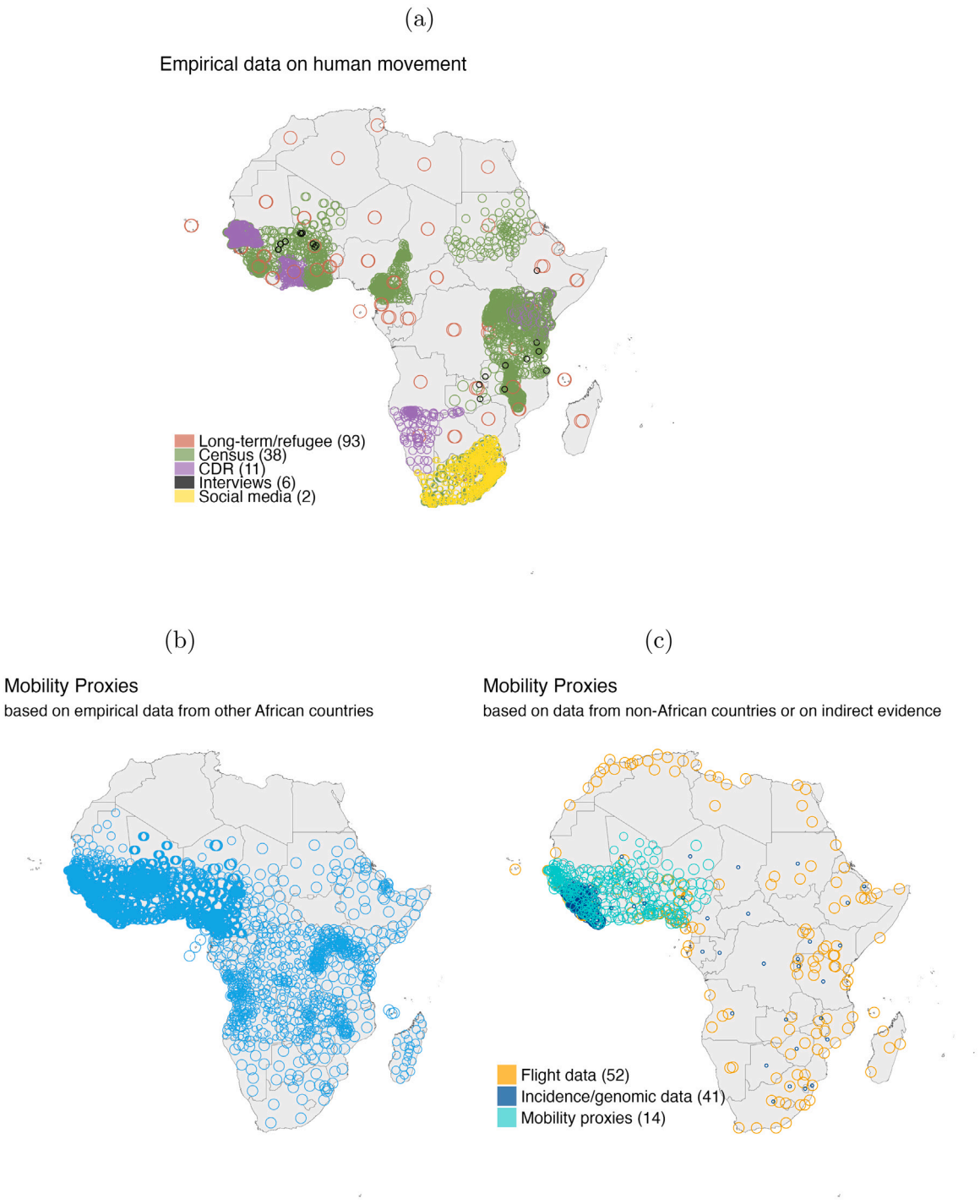

Fig. 2. Human population mobility data and proxies in Africa. We reviewed 30 peer-reviewed articles using human mobility data or proxies in any African country. For each country included in a study, we considered empirical data (or mobility proxy) for a given spatial resolution and for a specific time period as a discrete data set. We identified 150 empirical data sets and 168 mobility proxies. (a) Locations for which empirical data on human mobility were available. Data from Global Bi-Lateral Migration Database on international long-term migration and from the United Nations High Commissioner for Refugees on the movement of refugees between countries (red circles) were used as sources of international mobility within Africa. Census data (including population-level census, census microdata, and health and demographic surveillance surveys) from 14 countries (green circles) were used as empirical sources of subnational mobility at varying spatial resolutions across countries. Call details records (CDR) were used to quantify subnational mobility patterns at a fine temporal and spatial resolution and were available for five countries (purple circles). Tailored interviews at specific sites (black circles) were also used to quantify mobility. Social media usage at a fine spatial resolution (yellow circles) from two time periods were used to characterise mobility patterns in one country. (b) Locations for which human mobility proxies derived from empirical data shown in (a) have been used in the literature. Subnational mobility proxies for 44 countries in Africa (blue circles) were generated using census, CDR, or both, using data from one or more of 14 African countries where these data were available. For 37 countries in Africa, mobility proxies were the only source of information on subnational mobility. (c) Locations for which mobility proxies were derived either from data sources not shown in (a) or indirectly from other mobility proxies. The former were obtained through mobility models calibrated to empirical data from countries outside Africa (orange circles, flight data from Europe, the United States, and Canada), or from indirect evidence of mobility such as disease spread (pink circle). The latter were characterised using mobility proxies from a different country (cyan circles). In all panels, the size of the open circles indicates the spatial scale of the data or proxy, with smaller circles indicating higher spatial resolution. The number of empirical data sets or mobility proxies derived from a given source is indicated in parenthesis in the legends. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Leone, and Senegal), social media records (n = 1) (Dobra et al., 2018) or dedicated surveys (n = 5) (Yukich et al., 2013; Marshall et al., 2018). Censuses include surveys designed to measure changes in socio-demographic trends (such as internal migration) in a country, e.g. recording a change in address at ADM1 level over the last 1, 5, or 15 year period. These data are made available as either aggregate statistics (referred to as census or population-level census) or individual records (referred to as census microdata or individual-level census). Harmonised census microdata across different countries provided by the Integrated Public Use Microdata Series (IPUMS, 2022) were frequently used to quantify subnational mobility. Individual-level census data from specific geographic locations (rather than the entire country) were also available and informed more temporally and spatially resolved movements (Collinson et al., 2014; Dobra et al., 2017; Andrews et al., 2012). CDR provided the most spatially and temporally resolved data and were used for methodological research (Lu et al., 2013; Matamalas et al., 2016; Wesolowski et al., 2013a) as well as research into infectious diseases including Ebola (Peak et al., 2018), malaria (Tompkins and McCreesh, 2016; Ruktanonchai et al., 2016), schistosomiasis (Mari et al., 2017), and cholera (Finger et al., 2016). Similarly, mobility data from social media use (i.e. geolocated tweets) were derived at a high spatial resolution (ADM3 in South Africa) (Dobra et al., 2018). Dedicated surveys collected mobility data relevant to the spread of a specific disease in a specific location (Yukich et al., 2013; Marshall et al., 2018).

3.1.2. Mobility proxies

The empirical data described above were used to derive subnational mobility proxies for 44 of 54 countries in Africa, 37 of which had no empirical data informing subnational mobility (Fig. 2(b)). These proxies were often generated using mobility models fitted to census microdata from one or more countries, or CDR data from other African countries (Sorichetta et al., 2016; Wesolowski et al., 2014; Kraemer et al., 2017). We also identified mobility proxies estimated from flight data from Europe and North America, and subsequently used to inform mobility between 140 cities across 43 countries in Africa (Fig. 2(c)) (Huang et al., 2013). Overall, we identified subnational mobility proxies for 51 of 54 countries.

We found that typically, mobility proxies were not only generated using data from different locations but also extrapolated the mobility patterns from empirical data to earlier or later time periods, sometimes over a decade apart (Ruktanonchai et al., 2016; Finger et al., 2016; Kraemer et al., 2017; Pindolia et al., 2013; Wesolowski et al., 2014).

In addition to those relying on empirical movement data, mobility proxies were also inferred from indirect evidence such as spatio-temporal trends in disease incidence (D'Silva and Eisenberg, 2017; Silal et al., 2015; Tatem et al., 2012) or pathogen genomic information (Gustafson and Proctor, 2017; Dudas et al., 2017) (Fig. 2(c)).

Finally, mobility proxies were sometimes used to model human movement in different countries or at a different spatial scale, e.g. using gravity model parameters fitted to ADM1 unit data to describe movements between ADM2 units, in the same or another country (Kramer et al., 2016).

Typically, mobility proxies quantified movement information as absolute flows over a specified time window (i.e. the number of people moving between a source and destination) (D'Silva and Eisenberg, 2017; Kraemer et al., 2017; Wesolowski et al., 2014; Sorichetta et al., 2016; Huang et al., 2013; Dudas et al., 2017; Silal et al., 2015), or relative flows (i.e. probability of moving from a source to a given destination, conditional on moving out of the source) (Marshall et al., 2018; Tompkins and McCreesh, 2016; Matamalas et al., 2016; Finger et al., 2016). Some studies focusing on disease spread also characterised directly the probability of transmission of a pathogen between locations in a given time unit (Gustafson and Proctor, 2017; Kramer et al., 2016). The focus was therefore to quantify the movement between locations and not the probability of remaining in the same place. However, this assumption was rarely stated explicitly, e.g. clarifying that relative flows are conditional on moving in the first place. In fact, we identified only one study that estimated the probability of not moving out of a spatial unit over an year (i.e. p^{stay}), albeit at a gross temporal scale (Mari et al., 2017)

In studies that used empirical data sources to calibrate mobility models, the underlying data were rarely shared with the publication, even when commercial restrictions did not prevent data sharing (55/150 empirical data sets readily available). Mobility proxies were more often available (125/168) (Huang et al., 2013; Sorichetta et al., 2016; Wesolowski et al., 2014; Kramer et al., 2016).

3.2. Epidemic simulation study

A key finding of our scoping review was the large number of subnational mobility proxies that were based on empirical movement data from outside the country (occurring in 37 countries that did not have subnational empirical data). In this simulation study, we illustrate the potential implications arising from the use of such mobility proxies in epidemic models. We present results for our main scenario in which mobility proxies for France are generated from a gravity model calibrated to Portuguese movement data with p_i^{stay} set as the average across all Portuguese administrative units (and vice versa for Portugal). We compare each summary metric (see Methods) for simulations using mobility proxies against a baseline (using observed mobility data). We also refer to results from some of the supplementary scenarios we explored (Suppl Tab. S1) in which p_i^{stay} and p_{ij}^{move} for the mobility proxy were informed by different combinations of local or out-of-country data. Other scenarios looked at the effects of fitting the gravity model to data from the same country but aggregated at a different spatial scale, as well as a scenario similar to the main one but instead using a radiation model to predict p_{ij}^{move}.

3.2.1. Performance of mobility models

The mobility proxies fitted the observed movement data moderately well in both Portugal and France (R^2 = 0.44 and 0.47 respectively, Fig. 3A, Fig. 4A). In a scenario where the mobility proxy was based on data from the same country, the fit was marginally better (R^2 = 0.48 and 0.55 respectively), and only slightly decreased when using p_{ij}^{move} fitted to the other country and a local p_i^{stay} (Suppl Tab. S4). These results suggest that, in both countries, the mobility model we used (see Eqs. (2)–(4)) is only able to explain about half of the variance in the mobility data, and performance decreases by about 10% when fitted to data from another country.

3.2.2. Impact on epidemic dynamics

Invasion times (t_i^{first}) in simulations using mobility proxies were strongly correlated with those using observed movement, especially for epidemics seeded in well-connected locations (R^2 = 0.89 and 0.85 for Lisboa and Paris respectively, and 0.77 and 0.78 for Miranda do Douro and Brest respectively (Fig. 3B, Fig. 4B)). However invasion times tended to be under-estimated when using mobility proxies, particularly in France and more so in epidemics with peripheral seeding (Fig. 3C, Fig. 4C). In the most extreme scenario (epidemics seeded in Brest), the invasion times were predicted approximately a month too early (Fig. 4B), a considerable mismatch given the short assumed generation time (the time between infection of a case and their infector) of on average 3 days.

Despite this overall trend to under-predict invasion times, locally around peripheral seeding locations, invasion times were paradoxically over-predicted when using mobility proxies.

This led to poor characterisation of the early invasion dynamics, particularly with peripheral seeding (Fig. 3E, Fig. 4E). For epidemics seeded in Brest and Miranda do Douro, only 60% and 55% of the first 20 patches invaded were correctly identified when using mobility

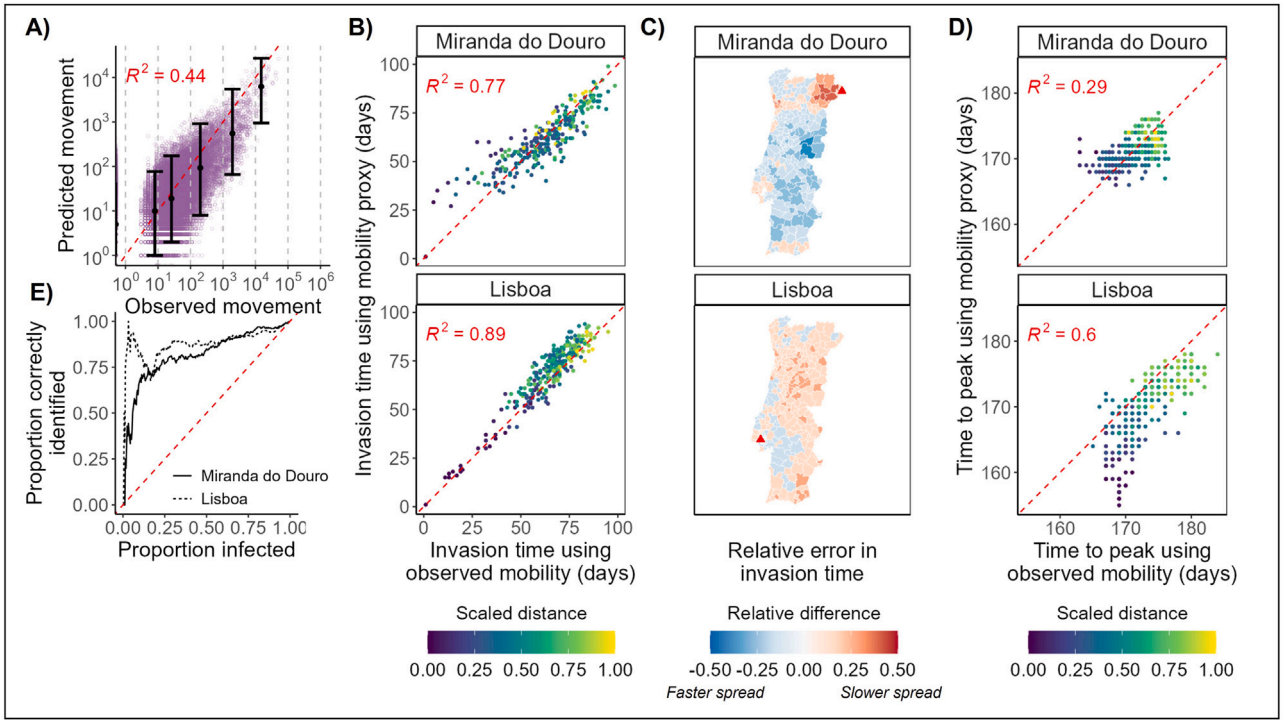

Fig. 3. Mobility predictions and epidemic simulations for Portugal. (A) Predicted numbers moving between Portuguese ADM2 units vs observed numbers from mobile phone call records. Each purple dot represents a pair of ADM2 units. The black dots show the median predicted value (y value) versus the median observed value in each bin (where the bins are marked with grey dashed lines). Error bars show the 2.5% and 97.5% quantiles of the predicted values in that bin. **(B) and (D)** show respectively the median invasion time (t_i^{first}) and the median peak time (t_i^{peak}) amongst residents of each spatial unit (represented by a dot) across all simulations where an epidemic was successfully seeded. Plots compare the times when using observed mobility patterns (x-axis) vs the mobility proxy (y-axis) from our central scenario in the epidemic model (see Methods). The colour represents the normalised distance of each patch from the seed location (shown by the title of each grid), calculated by dividing the distance from the seed by the maximum distance from that seed. The red dashed line is where y = x. **(C)** Maps of the relative error in invasion time in each spatial unit when using the mobility proxy vs observed data. Blue shading indicates the invasion time occurred earlier when using the mobility proxy, while red shading indicates a later invasion time when using mobility proxies. We use a relative scale, calculated as the difference between invasion time using mobility proxy and in the baseline (using the empirical mobility data), divided by the largest invasion times across all locations when using the proxy. The seeding location is marked with a red triangle. **(E)** Invasion order similarity (p_n^{first}). The y-axis shows the proportion of the first n patches invaded in an epidemic model using observed mobility that were also among the first n locations invaded when a mobility proxy was used in the epidemic model. n is the proportion shown on the x-axis. The two lines show different seed locations. The dashed red line shows the expected value if locations are chosen randomly. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

proxies, but this increased to 85% and 90% for epidemics seeded in Paris and Lisboa.

The ability to predict epidemic peak times in each spatial unit (t_i^{peak}) when using mobility proxies was very poor in France (R^2 = 0.05 and 0.01 for Paris and Brest respectively, Fig. 4D) but better in Portugal (R^2 = 0.29 and 0.60 for Miranda do Douro and Lisboa respectively, Fig. 3D). The baseline epidemics seeded in Portugal and in Paris had limited variability in the peak times across administrative units (2–3 weeks). Using mobility proxies reproduced this small amount of variation for Portugal, but not for Paris where a synchronous peak was predicted (1 week). Similarly, although the baseline epidemic seeded in Brest had substantial variability in epidemic peaks (up to 1.5 months apart), again a synchronous peak (1 week) was predicted when using mobility proxies. The sizes of the epidemic peaks were similar across the different movement scenarios.

In both countries, and regardless of the data used to fit the mobility model, the mobility proxies had similar performance in predicting movement data. However, the resulting impact on epidemic dynamics was very different between the two countries. In Portugal, again, epidemic metrics did not differ substantially depending on the data used to fit the mobility model (Suppl Tab. S4, Figs. S5 to S7). In France, using mobility proxies calibrated to Portugal data led to underestimated invasion times, poor invasion order prediction (especially in Brest), and false synchrony in the peaks. However, when using p_i^{stay} from France, all metrics substantially improved, irrespective of the country used to inform p_{ij}^{move} (Suppl Tab. S4, Figs. S1 to S3).

Using the radiation model instead of the gravity model in the main scenario had a mixed impact on epidemic dynamics (Figs. S4 and S8). Overall, the performance was slightly better in France and worse in Portugal. The radiation model improved some metrics, such as the ability to predict early peak timings of the epidemics seeded in Brest, although the later peaks were predicted less well. Use of the radiation model introduced substantial heterogeneity to the Portuguese simulations that was not seen in either the baseline scenario or the additional scenarios using gravity models.

Our results demonstrate that using mobility proxies can have a substantial impact on the predicted epidemic dynamics, with complex and non-intuitive biases, which cannot be predicted when simply comparing the mobility proxies to observed movement data.

4. Discussion

Accounting for over 14% of the world's population (Worldometer, 2022) and approximately 2.4% of the global airline passenger volume (Airports Council International Africa, 2022), Africa is a large source and sink of national and international movement of populations, and is estimated to bear half of the global burden of infectious diseases (Boutayeb, 2010). Our review revealed major gaps in the availability of empirical measurements of human mobility in Africa, leading to mobility proxies being used in place of data. Subnational mobility proxies were commonly derived from out-of-country data. We therefore designed a simulation study using highly resolved mobility

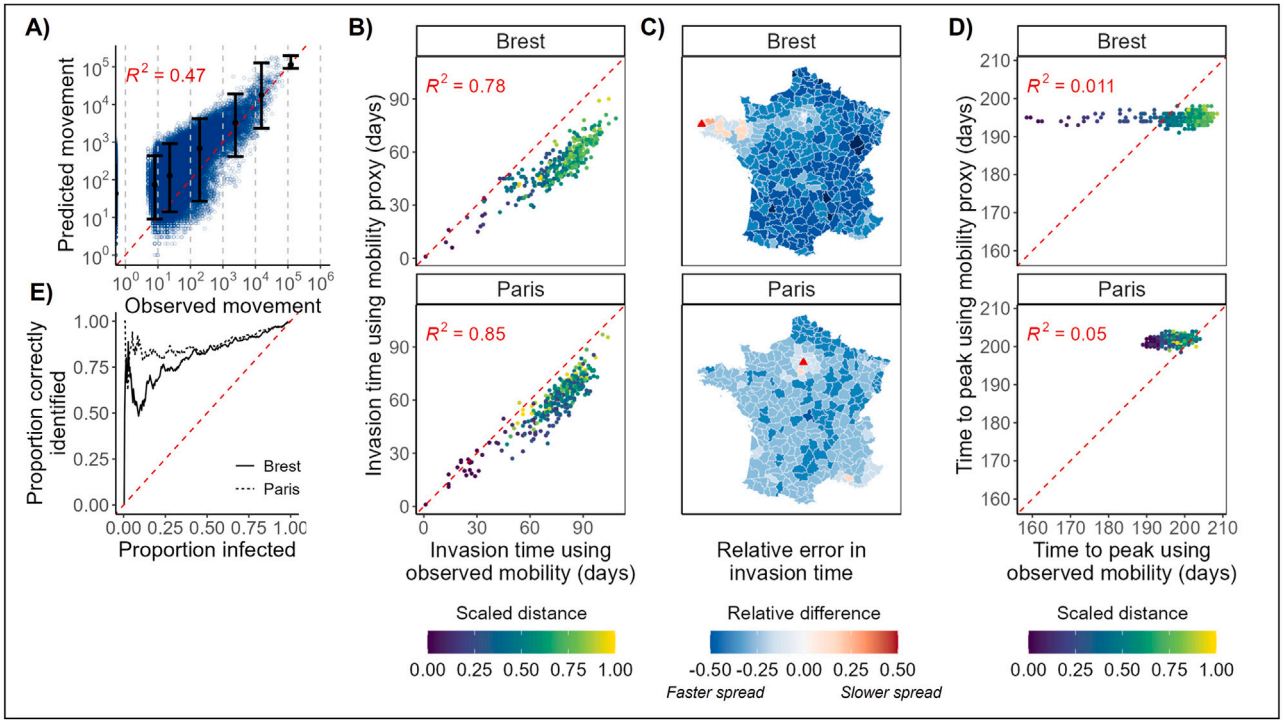

Fig. 4. Mobility predictions and epidemic simulations for France. Figure details as in Fig. 4 except here **(A)** shows predicted numbers moving between French ADM3 units vs observed numbers from mobile phone call records. Each blue dot represents a pair of ADM3 units. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

data to assess the extent to which using mobility proxies based on data from a different country in the same continent may affect predictions of epidemic spread. In the absence of available empirical data from Africa, we used movement data from France and Portugal. Our simulation study showed use of these proxies can impair our understanding of infectious disease dynamics, with biases introduced across a range of scenarios that used the gravity model and a scenario based on the radiation model.

Although we identified empirical mobility data in all 54 African countries (Fig. 2), data informing subnational movement was only available for 17 countries. These were mostly census data and focused on long term movement (1–5 years), which would only be relevant to the spread of pathogens with very slow progression (e.g. HIV). Despite evidence that these long term movements correlate with short-term mobility (Wesolowski et al., 2013a), census data are either not available at all or infrequently updated in many African countries (Fig. 2).

Highly temporally and spatially resolved mobility data sources (UK Department for Transport, 2021; Eurostat, 2022; Transport for London, 2022), including but not limited to mobile phone records (CDR data), are increasingly used to characterise human movement and epidemic spread, most frequently in high income countries (Tizzoni et al., 2014; Pepe et al., 2020; Silm et al., 2021; Chang et al., 2021; Schlosser et al., 2020; Jeffrey et al., 2020). However, such data is sparse for Africa: we only found 5 African countries with available CDR data, which was subsequently used to derive mobility proxies and examine epidemic spread in 17 other African countries, and which were potentially vastly different in size, population, topography and other factors expected to influence movement patterns.

Given the lack of appropriately resolved mobility data, such use of mobility proxies, typically based on data from other countries, is very common in the African context: we identified subnational mobility proxies for 51/54 African countries. Those were sometimes based on flight data from Europe and America, highlighting the scarcity of reliable local movement data. It is worth noting that mobility proxies often focus on describing absolute or relative flows between distinct locations, but rarely attempt to quantify p_i^{stay}, i.e. the number or proportion of people who do not move. This is a critical ingredient to model epidemic spread, where one needs to characterise movement but also the lack thereof.

Our work shows that subnational mobility proxies based on data from nearby countries are imperfect descriptions of empirical mobility, explaining only about half of the variability in the two mobility datasets we considered. Our simulation study demonstrates that this can have substantial and non-intuitive implications on our ability to predict epidemic spread. We focused on the early invasion dynamics and the local peak dynamics, as these metrics would be critical for informing policy, through timely, appropriately scaled and optimally targeted allocation of resources and implementation of control measures (SPI-M, 2018; Danon et al., 2021; Deschepper et al., 2021). The availability of high-resolution data from France and Portugal enabled us to explore the impact of using subnational proxies when modelling epidemic spread in a broader context. We expect that the biases that our simulation study has highlighted could be even more pronounced in African settings because of the heterogeneity in the size of African nations, the population density and other key demographic factors (Meredith et al., 2021; Wesolowski et al., 2015).

These biases may have implications for studies modelling the impact of interventions on infectious disease outbreaks. For example, spatial prioritisation of vaccination in the context of a limited stockpile could be suboptimal if relying on mobility proxies generated from out-of-country movement data, as was done for a yellow fever epidemic in Angola and the Democratic Republic of Congo (Kraemer et al., 2017). Similarly, the potential impact of population mobility reductions on case numbers and the epidemic duration for Ebola in Liberia may have been mis-estimated by using a gravity model fitted to mobile phone data from another country (Valdez et al., 2015; Wesolowski et al., 2014).

In most scenarios explored, predicted invasion times using mobility proxies tended to be earlier than when using empirical mobility data. While this may seem less problematic than the reverse, it could lead

to dismissing preventive interventions, wrongly perceived to take too long to implement given the predicted speed of invasion (e.g. building a new healthcare facility). Although the early invasion order was relatively well characterised for epidemics seeded in the capital cities, this was not the case when the epidemics were seeded in peripheral locations, and could lead to inappropriate spatial targeting of control measures. Heterogeneity in local epidemic peak times was overall well quantified in Portugal but not at all for France. Better characterising the peak heterogeneity may help better coordination of resources among different regions, e.g. allowing movement of medical staff, patients or material depending on the predicted local times of maximal epidemic burden.

These issues were overall more apparent in France than Portugal, and were much improved when using p_i^{stay} from the correct country. Although p_i^{stay} has received little attention by researchers, our work emphasises the importance of collecting data to estimate this parameter in all countries and at a fine spatial resolution, which can be later aggregated if needed. Indeed p_i^{stay} will evidently depend on the size of the spatial unit under consideration. For instance, ADM2 units in Portugal are on average smaller than ADM3 units in France, leading to the aforementioned underestimation of invasion times. Such data collection effort, to characterise non-movement, may be even more critical than collecting data on the destinations of the movements; indeed results from our epidemic simulations were much more sensitive to changes in p_i^{stay} than in p_{ij}^{move}.

In our study, we found that using French data to inform Portugal mobility fared much better than the opposite. Population mobility in a country potentially depends on a large number of factors such as the geography of the country, population density, demographics, and distribution of economic opportunities (Bonifazi and Heins, 2000; Castelli, 2018; Conlan et al., 2021). Understanding the key determinants may allow us to predict to what extent and in which contexts p_i^{stay} and p_{ij}^{move} may be informed by data from other countries. However, this would only be possible if highly resolved mobility data from many countries were available, which would reduce the need for mobility proxies. Although in our simulation study, we generated mobility proxies using data from one other country, we note that some popular published proxies (Sorichetta et al., 2016; Wesolowski et al., 2014) use data from multiple countries. It is unclear what the implications for disease modelling are, and whether such data pooling across countries is more appropriate than using data from a single country.

Our work has some limitations. First, since our search strategy was limited to identifying studies that used movement data in mobility models, we may have overlooked studies that primarily described mobility data and did not use any models. However, we believe it is unlikely that we have missed important data sets, as those would have been used by modelling studies captured in our search.

Second, we carried out the search in 2018. Since then, novel sources of mobility data have emerged specifically in the context of the COVID-19 pandemic. This includes community mobility reports from Google and Apple (Google, 2022; Apple, 2022); however those only measure overall mobility trends which cannot directly be used to inform O-D matrices. Furthermore, many such data are being released specifically to support the COVID-19 response, and cannot replace much needed highly-resolved representative and regularly updated mobility data.

Third, we used gravity and radiation mobility models, and a multipatch compartmental model stratified by home and work location, guided by the available mobility data. More sophisticated models of mobility (Meredith et al., 2021) or disease propagation (Keeling et al., 2010; Van Kerckhove et al., 2013; Haw et al., 2019; Prem et al., 2017) could display more complex dynamics. However our simple and parsimonious approach was sufficient to highlight potential biases in epidemic spread predictions stemming from using mobility proxies instead of empirical data; this would only be magnified by using more sophisticated models.

Finally, one aspect we did not consider, which is rarely acknowledged, is that mobility proxies are outputs of statistical models, and hence carry inherent uncertainty. Such uncertainty is rarely quantified, or reported, and almost never propagated in subsequent analysis, e.g. epidemic models.

Overall, our work underscores the need for regularly updated empirical measurements of population movement within and between countries. Despite potential biases in mobile phone usage due to limited and unevenly distributed smartphone ownership (GSMA, 2021), such data sources remain the most promising to characterise short-term human movement at high resolution. Mobile phone operators could consider periodic release of aggregate data sets to support public health efforts. Even though we did not restrict our search to studies related to infectious diseases, most of the included publications focused on their spread, highlighting the centrality of human movement in infectious disease epidemiology. A few data sources underpinned wide-ranging research on multiple diseases such as cholera, ebola, malaria, and schistosomiasis. This suggests that availability of empirical mobility data could pay large dividends in pandemic preparedness as well as an improved understanding of spread of diseases that are endemic or lead to recurring epidemics (Buckee et al., 2020).

Data availability

All code used in this analysis is available at https://github.com/j-wardle/mobility_africa_models. The simulation study uses data made available at https://doi.org/10.1371/journal.pcbi.1003716.

CRediT authorship contribution statement

Jack Wardle: Study conception and design, Simulation study and analysis, Writing original draft, Review and editing. **Sangeeta Bhatia:** Study conception and design, Scoping review, Data extraction, Simulation study and analysis, Writing original draft, Review and editing. **Moritz U.G. Kraemer:** Study conception and design, Scoping review, Review and editing. **Pierre Nouvellet:** Study conception and design, Scoping review, Review and editing, Supervision. **Anne Cori:** Study conception and design, Scoping review, Review and editing, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: AC has received payment from Pfizer for teaching of mathematical modelling of infectious diseases.

Acknowledgements

We thank Michele Tizzoni and colleagues for publicly sharing data on commuter patterns in France and Portugal, as well as for providing us with additional data on the numbers of people that do not commute. This study is partially funded by the National Institute for Health Research (NIHR) Health Protection Research Unit in Modelling and Health Economics, a partnership between UK Health Security Agency (UKHSA) and Imperial College London, in collaboration with LSHTM (grant code NIHR200908); and acknowledges funding from the MRC Centre for Global Infectious Disease Analysis (reference MR/R015600/1), which is jointly funded by the UK Medical Research Council (MRC) and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement and is also part of the EDCTP2 programme supported by the European Union; and acknowledges funding by Community Jameel. JW acknowledges research funding from the Wellcome Trust (grant 102169/Z/13/Z). SB acknowledges funding from the Wellcome Trust (grant 219415). AC was supported by the Academy of Medical Sciences Springboard scheme, funded by the AMS,

Wellcome Trust, BEIS, the British Heart Foundation and Diabetes UK (grant SBF005/1044).
Disclaimer: The views expressed are those of the author(s) and not necessarily those of the NIHR, UKHSA or the Department of Health and Social Care. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. For the purpose of open access, the author has applied a 'Creative Commons Attribution' (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.epidem.2023.100666. We share two supplementary files:

1. Supplementary information on the methods and results of the simulation study
2. Data extracted from studies included in the scoping review

References

Africa Centers for Disease Control and Prevention, 2022. Yellow fever: Recent outbreaks in Africa, URL https://africacdc.org/disease/yellow-fever/.
Africa Centers for Disease Control and Prevention, 2022. Cholera: Recent outbreaks in Africa. URL https://africacdc.org/disease/cholera/.
Africa Centers for Disease Control and Prevention, 2022. Measles: Recent outbreaks in Africa. URL https://africacdc.org/disease/measles/.
Airports Council International Africa, 2022. Air traffic statistics. URL https://www.aci-africa.aero/data-centre/air-traffic-statistics/.
Andrews, J.R., et al., 2012. Projecting the benefits of antiretroviral therapy for HIV prevention: The impact of population mobility and linkage to care. J. Infect. Dis. 206 (4), 543–551.
Apple, 2022. COVID-19 mobility trends reports. URL https://covid19.apple.com/mobility.
Ascani, A., et al., 2021. Mobility in times of pandemics: Evidence on the spread of COVID19 in Italy's labour market areas. Struct. Change Econ. Dyn. (ISSN: 0954349X) 58, 444–454. http://dx.doi.org/10.1016/j.strueco.2021.06.016.
Bharti, N., et al., 2008. Measles on the edge: Coastal heterogeneities and infection dynamics. PLoS One 3 (4), e1941.
Bhatia, S., et al., 2021. Using digital surveillance tools for near real-time mapping of the risk of infectious disease spread. NPJ Digit. Med. 4 (1), 1–10. http://dx.doi.org/10.1038/s41746-021-00442-3.
Blondel, V.D., et al., 2012. Data for development: The d4d challenge on mobile phone data. arXiv preprint arXiv:1210.0137.
Bogoch, I.I., et al., 2015. Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 West African outbreak. Lancet 385 (9962), 29–35.
Bonifazi, C., Heins, F., 2000. Long-term trends of internal migration in Italy. Int. J. Popul. Geogr. 6 (2), 111–131.
Boutayeb, A., 2010. The impact of infectious diseases on the development of Africa. In: Preedy, V.R., Watson, R.R. (Eds.), Handbook of Disease Burdens and Quality of Life Measures. Springer New York, New York, NY, pp. 1171–1188. http://dx.doi.org/10.1007/978-0-387-78665-0_66, ISBN: 978-0-387-78665-0.
Buckee, C.O., et al., 2020. Aggregated mobility data could help fight COVID-19. Science 368 (6487), 145–146.
Castelli, F., 2018. Drivers of migration: Why do people move? J. Travel Med. 25.
Centers for Disease Control and Prevention, 2022. History of Ebola virus disease (EVD) outbreaks. URL https://www.cdc.gov/vhf/ebola/history/chronology.html.
Chang, S., et al., 2021. Mobility network models of COVID-19 explain inequities and inform reopening. Nature 589 (7840), 82–87.
Charu, V., et al., 2017. Human mobility and the spatial transmission of influenza in the United States. PLoS Comput. Biol. 13 (2), e1005382.
Chinazzi, M., et al., 2020. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 368 (6489), 395. http://dx.doi.org/10.1126/science.aba9757, URL http://science.sciencemag.org/content/368/6489/395.abstract.
Collinson, M.A., et al., 2014. Migration and the epidemiological transition: Insights from the Agincourt sub-district of northeast South Africa. Global Health Action 7 (1), 23514.
Conlan, A.J., et al., 2021. Human mobility data from the BBC pandemic project. medRxiv.
Craig, A., et al., 2020. Risk of COVID-19 importation to the Pacific islands through global air travel. Epidemiol. Infect. 148.
Danon, L., et al., 2021. A spatial model of COVID-19 transmission in England and Wales: Early spread, peak timing and the impact of seasonality. Philos. Trans. R. Soc. B (ISSN: 0962-8436) 376 (1829), 20200272. http://dx.doi.org/10.1098/rstb.2020.0272, ISSN: 1471-2970.
Deschepper, M., et al., 2021. Prediction of hospital bed capacity during the COVID-19 pandemic. BMC Health Serv. Res. (ISSN: 1472-6963) 21 (1), 468. http://dx.doi.org/10.1186/s12913-021-06492-3.
Dobra, A., Mohammadi, R., et al., 2018. Loglinear model selection and human mobility. Ann. Appl. Stat. 12 (2), 815–845.
Dobra, A., et al., 2017. Space-time migration patterns and risk of HIV acquisition in rural South Africa. AIDS (London, England) 31 (1), 137.
Dorigatti, I., et al., 2017. International risk of yellow fever spread from the ongoing outbreak in Brazil, December 2016 to May 2017. Eurosurveillance 22 (28), 30572.
D'Silva, J.P., Eisenberg, M.C., 2017. Modeling spatial invasion of Ebola in West Africa. J. Theor. Biol. 428, 65–75.
Dudas, G., et al., 2017. Virus genomes reveal factors that spread and sustained the Ebola epidemic. Nature 544 (7650), 309.
Eggo, R.M., et al., 2011. Spatial dynamics of the 1918 influenza pandemic in England, Wales and the United States. J. R. Soc. Interface 8 (55), 233–243.
Eurostat, 2022. Eurostat: Your key to European statistics. URL https://ec.europa.eu/eurostat/web/transport/data/database.
Findlater, A., Bogoch, I.I., 2018. Human mobility and the global spread of infectious diseases: A focus on air travel. Trends Parasitol. 34 (9), 772–783. http://dx.doi.org/10.1016/j.pt.2018.07.004, ISSN: 1471-4922.
Finger, F., et al., 2016. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks. Proc. Natl. Acad. Sci. 113 (23), 6421–6426.
Gilbert, M., et al., 2020. Preparedness and vulnerability of African countries against importations of COVID-19: A modelling study. Lancet 395 (10227), 871–877.
Google, 2022. COVID-19 community mobility reports. URL https://www.google.com/covid19/mobility/.
GSMA, 2021. The mobile economy Sub-Saharan Africa 2021, URL https://www.gsma.com/mobileeconomy/wp-content/uploads/2021/09/GSMA_ME_SSA_2021_English_Web_Singles.pdf.
Gustafson, K.B., Proctor, J.L., 2017. Identifying spatio-temporal dynamics of Ebola in Sierra Leone using virus genomes. J. R. Soc. Interface 14 (136), 20170583.
Haw, D.J., et al., 2019. Differential mobility and local variation in infection attack rate. PLoS Comput. Biol. 15 (1), e1006600.
Hollingsworth, T.D., et al., 2006. Will travel restrictions control the international spread of pandemic influenza? Nat. Med. 12 (5), 497–499.
Huang, Z., et al., 2013. An open-access modeled passenger flow matrix for the global air network in 2010. PLoS One 8 (5), e64317.
IPUMS, 2022. Integrated public use microdata series, international: Version 7.3[dataset]. URL https://international.ipums.org/international/index.shtml.
Jeffrey, B., et al., 2020. Anonymised and aggregated crowd level mobility data from mobile phones suggests that initial compliance with COVID-19 social distancing interventions was high and geographically consistent across the UK. Wellcome Open Res. 5, 170. http://dx.doi.org/10.12688/wellcomeopenres.15997.1.
Keeling, M.J., et al., 2010. Individual identity and movement networks for disease metapopulations. Proc. Natl. Acad. Sci. 107 (19), 8866–8870.
Kraemer, M.U.G., et al., 2016. Progress and challenges in infectious disease cartography. Trends Parasitol. (ISSN: 1471-4922) 32 (1), 19–29. http://dx.doi.org/10.1016/j.pt.2015.09.006, URL http://www.sciencedirect.com/science/article/pii/S147149221500207X.
Kraemer, M.U., et al., 2017. Spread of yellow fever virus outbreak in Angola and the democratic Republic of the Congo 2015–16: A modelling study. Lancet Infect. Dis. 17 (3), 330–338.
Kraemer, M.U.G., et al., 2019. Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings. Sci. Rep. (ISSN: 2045-2322) 9 (1), 5151. http://dx.doi.org/10.1038/s41598-019-41192-3.
Kramer, A.M., et al., 2016. Spatial spread of the West Africa Ebola epidemic. R. Soc. Open Sci. 3 (8), 160294.
Kucharski, A.J., et al., 2015. Measuring the impact of Ebola control measures in Sierra Leone. Proc. Natl. Acad. Sci. 112 (46), 14366–14371.
Lai, S., et al., 2019. Measuring mobility, disease connectivity and individual risk: A review of using mobile phone data and mHealth for travel medicine. J. Travel Med. 26 (3), taz019.
Li, J., et al., 2021. The emergence, genomic diversity and global spread of SARS-CoV-2. Nature 1–11.
Longini, I.M., et al., 2005. Containing pandemic influenza at the source. Science 309 (5737), 1083–1087.
Lu, X., et al., 2013. Approaching the limit of predictability in human mobility. Sci. Rep. 3, 2923.
Mari, L., et al., 2017. Big-data-driven modeling unveils country-wide drivers of endemic schistosomiasis. Sci. Rep. 7 (1), 489.
Marshall, J.M., et al., 2018. Mathematical models of human mobility of relevance to malaria transmission in Africa. Sci. Rep. 8 (1), 7713.
Matamalas, J.T., et al., 2016. Assessing reliable human mobility patterns from higher order memory in mobile communications. J. R. Soc. Interface 13 (121), 20160203.
Meredith, H.R., et al., 2021. Characterizing human mobility patterns in rural settings of sub-Saharan Africa. ELife (ISSN: 2050-084X) 10, e68441. http://dx.doi.org/10.7554/eLife.68441.

Merler, S., et al., 2015. Spatiotemporal spread of the 2014 outbreak of Ebola virus disease in Liberia and the effectiveness of non-pharmaceutical interventions: a computational modelling analysis. Lancet Infect. Dis. 15 (2), 204–211.

O'Toole, Á., et al., 2021. Tracking the international spread of SARS-CoV-2 lineages B. 1.1. 7 and B. 1.351/501Y-V2. Wellcome Open Res. 6.

Özden, Ç., et al., 2011. Where on Earth is everybody? The evolution of global bilateral migration 1960–2000. World Bank Econ. Rev. (ISSN: 1564-698X) 25 (1), 12–56. http://dx.doi.org/10.1093/wber/lhr024, ISSN: 0258-6770.

Peak, C.M., et al., 2018. Population mobility reductions associated with travel restrictions during the Ebola epidemic in Sierra Leone: Use of mobile phone data. Int. J. Epidemiol. 47 (5), 1562–1570.

Pepe, E., et al., 2020. COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. Sci. Data 7 (1), 1–7.

Pindolia, D.K., et al., 2013. The demographics of human and malaria movement and migration patterns in East Africa. Malaria J. 12 (1), 397.

Prem, K., et al., 2017. Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLoS Comput. Biol. 13 (9), e1005697.

Ramiadantsoa, T., et al., 2022. Existing human mobility data sources poorly predicted the spatial spread of SARS-CoV-2 in Madagascar. Epidemics 38, 100534.

Riley, S., et al., 2015. Five challenges for spatial epidemic models. Epidemics 10, 68–71.

Ruktanonchai, N.W., et al., 2016. Identifying malaria transmission Foci for elimination using human mobility data. PLoS Comput. Biol. 12 (4), e1004846.

Ruktanonchai, N.W., et al., 2020. Assessing the impact of coordinated COVID-19 exit strategies across Europe. Science 369 (6510), 1465–1470.

Schlosser, F., et al., 2020. COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. Proc. Natl. Acad. Sci. 117 (52), 32883–32890.

Silal, S.P., et al., 2015. Hitting a moving target: A model for malaria elimination in the presence of population movement. PLoS One 10 (12), e0144990.

Silm, S., et al., 2021. Temporary population mobilities between Estonia and Finland based on mobile phone data and the emergence of a cross-border region. Eur. Plan. Stud. 29 (4), 699–719.

Simini, F., et al., 2012. A universal model for mobility and migration patterns. Nature (ISSN: 1476-4687) 484 (7392), 96–100. http://dx.doi.org/10.1038/nature10856.

Song, C., et al., 2010. Limits of predictability in human mobility. Science 327 (5968), 1018–1021.

Sorichetta, A., et al., 2016. Mapping internal connectivity through human migration in malaria endemic countries. Sci. Data 3, 160066.

SPI-M, 2018. SPI-M modelling summary. URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/756738/SPI-M_modelling_summary_final.pdf.

Tatem, A.J., 2014. Mapping population and pathogen movements. Int. Health (ISSN: 1876-3413) 6 (1), 5–11. http://dx.doi.org/10.1093/inthealth/ihu006.

Tatem, A.J., et al., 2006. Global transport networks and infectious disease spread. Adv. Parasitol. 62, 293–343. http://dx.doi.org/10.1016/S0065-308X(05)62009-X, ISSN: 0065-308X.

Tatem, A.J., et al., 2012. Spatial accessibility and the spread of HIV-1 subtypes and recombinants. AIDS 26 (18), 2351–2360.

Tizzoni, M., et al., 2014. On the use of human mobility proxies for modeling epidemics. PLoS Comput. Biol. 10 (7), e1003716. http://dx.doi.org/10.1371/journal.pcbi.1003716.

Tompkins, A.M., McCreesh, N., 2016. Migration statistics relevant for malaria transmission in Senegal derived from mobile phone data and used in an agent-based migration model. Geosp. Health 11 (1 Supp), 408.

Transport for London, 2022. Transport for London: Our open data. URL https://tfl.gov.uk/info-for/open-data-users/our-open-data.

Tricco, A.C., et al., 2018. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. Ann. Internal Med. 169 (7), 467–473. http://dx.doi.org/10.7326/M18-0850.

Truscott, J., Ferguson, N.M., 2012. Evaluating the adequacy of gravity models as a description of human mobility for epidemic modelling.

Tuite, A.R., et al., 2011. Cholera epidemic in Haiti, 2010: Using a transmission model to explain spatial spread of disease and identify optimal control interventions. Ann. Internal Med. 154 (9), 593–601.

UK Department for Transport, 2021. National travel survey: 2020. URL https://www.gov.uk/government/statistics/national-travel-survey-2020/national-travel-survey-2020.

UNHCR, 2022. The UN refugee agency refugee data finder. URL https://www.unhcr.org/refugee-statistics/.

Valdez, L.D., et al., 2015. Predicting the extinction of Ebola spreading in Liberia due to mitigation strategies. Sci. Rep. 5 (1), 1–10.

Van Kerckhove, K., et al., 2013. The impact of illness on social networks: Implications for transmission and control of influenza. Am. J. Epidemiol. 178 (11), 1655–1662.

Viana, R., et al., 2021. Rapid epidemic expansion of the SARS-CoV-2 omicron variant in southern Africa. medRxiv.

Viboud, C., et al., 2006. Synchrony, waves, and spatial hierarchies in the spread of influenza. Science 312 (5772), 447–451.

Wesolowski, A., et al., 2012. Heterogeneous mobile phone ownership and usage patterns in Kenya. PLoS One 7 (4), e35319.

Wesolowski, A., et al., 2013a. The use of census migration data to approximate human movement patterns across temporal scales. PLoS One 8 (1), e52971.

Wesolowski, A., et al., 2013b. The impact of biases in mobile phone ownership on estimates of human mobility. J. R. Soc. Interface 10 (81), 20120986.

Wesolowski, A., et al., 2014. Commentary: Containing the ebola outbreak-the potential and challenge of mobile network data. PLoS Currents 6.

Wesolowski, A., et al., 2015. Evaluating spatial interaction models for regional mobility in sub-Saharan Africa. PLoS Comput. Biol. 11 (7), e1004267.

Wesolowski, A., et al., 2016. Connecting mobility to infectious diseases: The promise and limits of mobile phone data. J. Infect. Dis. 214 (suppl_4), S414–S420.

Worldometer, 2022. Africa population. URL https://www.worldometers.info/world-population/africa-population/.

Wu, J.T., et al., 2007. Spatial considerations for the allocation of pre-pandemic influenza vaccination in the United States. Proc. R. Soc. B: Biol. Sci. 274 (1627), 2811–2817.

Yang, W., et al., 2015. Transmission network of the 2014–2015 Ebola epidemic in Sierra Leone. J. R. Soc. Interface 12 (112), 20150536.

Yukich, J.O., et al., 2013. Travel history and malaria infection risk in a low-transmission setting in Ethiopia: A case control study. Malaria J. 12 (1), 33.

Zipf, G.K., 1946. The P 1 P 2/D hypothesis: On the intercity movement of persons. Am. Sociol. Rev. 11 (6), 677–686.