Migration statistics relevant for malaria transmission in Senegal derived from mobile phone data and used in an agent-based migration model

Adrian M. Tompkins *

Abdus Salam International Centre for Theoretical Physics, Italy

Nicky McCreesh

London School of Hygiene and Tropical Medicine, London, U.K.

* *Corresponding author address:* Abdus Salam International Centre for Theoretical Physics (ICTP), Italy

E-mail: tompkins@ictp.it

ABSTRACT

One year of mobile phone location data from Senegal is analysed to determine the characteristics of journeys that result in an overnight stay, and are thus relevant for malaria transmission. Defining the home location of each person as the place of most frequent calls, it is found that approximately 60% of people who spend nights away from home have regular destinations that are repeatedly visited, although only 10% have 3 or more regular destinations. The number of journeys involving overnight stays peaks at a distance of 50 km, although roughly half of such journeys exceed 100 km. Most visits only involve a stay of one or two nights away from home, with just 4% exceeding one week. A new agent-based migration model is introduced, based on a gravity model adapted to represent overnight journeys. Each agent makes journeys involving overnight stays to either regular or random locations, with journey and destination probabilities taken from the mobile phone dataset. Preliminary simulations show that the agent based model can approximately reproduce the patterns of migration involving overnight stays.

Keywords: transport, infectious disease, elimination, surveillance, urbanisation

Introduction

It has long been appreciated that population movements drive the transmission patterns and intensity of many communicable diseases (Garnham 1945; Findlay 1946; Prothero 1961; Marques 1987). Health reports during the early 20th century from the Uganda protectorate often attributed malaria anomalies to population movements that were either sub-national or trans-boundary in nature (Tompkins et al. 2015). In locations where endemicity is spatially heterogeneous, population mobility can also act to transport malaria parasites, and may lead to outbreaks in epidemic zones (Martens and Hall 2000; Wesolowski et al. 2012b). For example, the success of early efforts to control malaria in northern Uganda were hindered by the reintroduction of the parasite through population movements (De Zulueta et al. 1961; Talisuna et al. 2015). A number of studies have attempted to determine the role of migration in malaria transmission (Torres-Sorando and Rodrıguez 1997; Tatem and Smith 2010; Lynch and Roper 2011; Pindolia et al. 2013).

National (internal) population mobility can be classified according to the relevant time-scale: permanent, such as urbanisation trends; long term in response to environmental stress or seasonal work; or regular and cyclic in nature due to economic, social or pastoral reasons (e.g. Todaro 1969; Findley 1994). It is the latter regular mobility of populations on daily to monthly time-scales that is the focus of the present work. Information concerning national migration is limited and subject to considerable uncertainties. Some estimates can nevertheless be obtained from household surveys (Watkins and Fleisher 2002), census reports (Wesolowski et al. 2013), satellite imagery (Bharti et al. 2011), and air and ground transport records (e.g. Griffitts 1933; Tatem et al. 2006). Pindolia et al. (2012) and Tatem (2014) review these data sources and highlight the potential of mobile phone data to supplement their information on national scales. Mobile phone data have been used to assess population cyclic and international mobility in a number of African countries (Tatem et al. 2009; Bengtsson et al. 2011; Blumenstock 2012; Buckee et al. 2013). Gething and Tatem (2011) and Wesolowski et al. (2014) discuss their potential in disaster response situations.

While previous analysis of mobility has shed much light on potential impacts for malaria (Wesolowski et al. 2012b), the analyses are often conducted in terms of total population fluxes between districts, which are then represented by simple diffusion models. In order to be able incorporate information concerning mobility into spatial dynamical malaria modelling frameworks (e.g. Hoshen and Morse 2004; Jones and Morse 2010; Tompkins and Ermert 2013; Tompkins and Di Giuseppe 2015), additional transmission-relevant statistics concerning the journeys are required, in particular whether journeys involve overnight stays.

Key malaria vectors in Africa bite predominantly between the hours of 9pm and 6am (Braack et al. 1994; Pates and Curtis 2005). Short, local journeys that result in no overnight stay are thus unlikely to result in transmission, even if malaria risk can vary rapidly over small spatial scales (Carter et al. 2000; Bousema et al. 2012). The more nights that are spent away, the higher the probability of parasite transport. If a journey is made from a low transmission area to an endemic zone, where the probability of receiving an inoculation per night is β, the probability of that person receiving an inoculation during N nights is simply $1-(1-\beta)^N$. The effect is nonlinear; N people visiting a location for one night are more likely to transport parasites back to the origin than a single person spending N nights (assuming for simplicity that β is equal for all, which Lindsay et al. 1993; Knols et al. 1995; Mukabana et al. 2002, show is is not the case), since in the former case $N\beta$ people become infected, which always exceeds $1-(1-\beta)^N$ for $N \geq 2$. These arguments also apply to the probability of transporting parasites from an endemic area to a low transmission zone. A map of journey densities doesn't reflect the probability of a particular journey being made by an individual, as it has been previously demonstrated that most individuals will regularly visit only a small number of locations(Gonzalez et al. 2008). Urban dwellers may regularly return to their rural origin to visit family, for example. Calculations using diffusive models based on journey density maps will thus overestimate the transport of parasites between low and high endemicity settings.

The first aim of this paper is to conduct an analysis of mobile phone record data for Senegal to determine these transmission-relevant statistics. The strong gradient in malaria transmission intensity in Senegal implies that human mobility may have a significant role in transporting malaria parasites to the northern epidemic-prone districts. Previous studies of the impact of human mobility malaria have used statistical techniques such as diffusion models, (Wesolowski et al. 2012b) or internal migration mapped with gravity models (Garcia et al. 2014). Agent-based models have also been developed for human mobility (Kniveton et al. 2011, 2012; Augustijn-Beckers et al. 2011; Parker and Epstein 2011) and these have the advantage that they permit a specific memory of the history of each agent. The second aim of the research was to develop an agent-based model of population movements, using the mobile phone statistics to set up distributions of regular and random destinations for each agent. The intention is that in the future the population movement model will be coupled to the vector borne disease community model of ICTP (VECTRI) dynamical malaria model (Tompkins and Ermert 2013), to allow the latter to account for the effect of population mobility more accurately. After analyzing the malaria-relevant statistics of the phone data, a preliminary simulation is made with the agent based model to assess its ability to reproduce general patterns of population mobility.

Method

Mobile phone data analysis

The statistics regarding population mobility are derived from a dataset of Orange mobile phone use provided in the 2nd phase of the Data for development (D4D) challenge project (de Montjoye et al. 2014). Earlier research by the data providers has shown that anonymizing mobile data is inadequate to protect identity if records are provided at high spatial resolution or for long periods (de Montjoye et al. 2013). Thus to ensure user privacy, the project provided anomymized data with either short temporally coherent records limited to a 14 day period per individual user tracked, or with a strongly degraded spatial resolution. As an extra safeguard, all selected projects that received D4D data, including the present work, were required to subject their resulting analysis for clearance by the D4D ethical committee.

The degraded spatial resolution dataset that provided continuous tracking information for an entire year for each user at the coarse-scale arrondissement level was used. This dataset was employed in order to be able to identify the arrondissement in which a user's home was located, multi-day journeys and regular destinations of users. One caveat of the analysis is that journey's that occur within an arrondissement are not resolvable, although it is assumed that most of these journeys would not result in an overnight stay, which is the subject of the present analysis, and thus the impact of the degraded spatial resolution is considered minor.

In this analysis, the first 18384 entries were used to summarize journeys that result in an overnight stay. The first step of the analysis was to assign home locations to each individual. This was defined as the place from which a call was made on the greatest number of days. People who lived in or visited locations on the border between arrondissements presented problems for the analysis, as it was difficult to determine where these people lived, and where they spent each night. Locations were therefore recoded, with all observations in locations adjacent to an individual's home location considered to occur in their home location. Each individual's second most commonly visited arrondissement was then identified, and locations adjacent to that recoded. This was then repeated for their third, fourth, etc most commonly visiting locations.

We then attempted to determine where each individual spent each night. Four rules were used. In order of priority, they were:

1. If the last call by an individual on day n, and any call on day n+1, were from location x, then the individual spent night n in location x.

2. If the last call by an individual on any day, and the next call on any day, were from the same location, and there were not more than 60 hours between the two calls, then the individual was considered to have spent any nights between the calls in that location.

3. If the last call on a day occurred after 7pm, then the individual spent the night in the location from which that call was made.

4. If the first call on a day occurred before 7am, then the individual spent the preceding night in the location from which that call was made.

The number and destinations of all trips made by each individual were then identified. Each trip was considered to start when an individual made a call from a location other than their home location, and to end when they next made a call from their home location. The destination of the trip was considered to be the location from which a call was made that was located the furthest distance from the individual's home location. Distances were measured as the Euclidean distance between the centre points of two locations, i.e. measured 'as the crow flies', neglecting the road network in this preliminary analysis (e.g. Shahabi et al. 2003).

The proportion of trips that involved a night away were then calculated. A trip was considered to involve a night away if either the night before or the night of the first day of a trip was considered to have occurred at a location other than the individual's home location. If the location in which an individual spent the night before or the night of the first day of a trip was not known, then that trip was excluded from the analysis.

A number of assumptions were made in the analysis. The ten arrondissements covering the Dakar area were considered to be one location for the purposes of the analysis. In addition, two other small, adjacent arrondissements were considered to be one location. The dataset contained a considerable number of erroneous observations that required removing to avoid biasing the statistics. These included events such as two calls by the same individual from different, non-adjacent arrondissements a short time apart. Implausible calls or groups of calls were removed if they would have required a travel speed of greater than 50km h^{-1}, with distance measured 'as the crow flies' between the closet points of the arrondissements. This may have resulted in some genuine calls being removed if individuals traveled by plane, however the low volume of domestic air travel in Senegal means that the effects of this are likely to have been negligible.

Agent based mobility model

An agent based model for cyclic and permanent migration within developing countries has been constructed. The model presently divides a square domain that includes all Senegal into a regular grid-mesh using a 5km resolution. The population density in each cell is given by the AfriPop dataset (Linard et al. 2012). The model is initialized using 1000 agents in each 5km^2 cell, which is also assigned as the *home* location for the agent. Discounting ocean cells in the simulation domain, the simulation includes a total of 2856000 agents. Each agent thus represents a different number of individuals according to the location. For example, in an urban location with a population density of 2000 km^{-2}, each agent represents 50 people, while in a sparsely populated rural location of 20km^{-2}, each agent represents half an individual. This highlights the fact that individuals per agent is a member of the real set and can take non-integer values. Birth and death of individuals is not accounted for in these simulations, which are of one year in duration.

In addition to its home location, each agent is assigned N regular destination locations at the start of the simulation. The analysis of the D4D dataset indicated that most individuals that regularly travelled to one or more location had just one regular destination, with smaller numbers having two or three regular destinations. The proportion with four or more regular destinations was negligible. Thus, each agent is assigned three regular locations, with a different probability weighting assigned to each regular location to match the mean probabilities derived from the data. Thus a migration event is far more likely to result in a journey taken to regular location 1, with locations 2 and 3 increasingly less likely. In these preliminary tests, all agents have equal probabilities of visiting their regular locations. The number of agents that make n trips to regular locations, and the number of identifiable regular locations per agent per year of simulation, will therefore be Poisson distributed, and under dispersive relative to the phone data.

Each agent is assigned a probability to migrate per day, which is set to give the same mean number of journeys per individual in a year as observed in the D4D dataset (approximately 6.2). In each timestep, a random number is chosen for each agent and used to decide if it will make a journey that will involve an overnight stay. The full decision tree of the model is shown schematically in Fig. 1. For each migration event, a stochastic decision is taken to move to one of the agent's regular locations (assuming $N \geq 1$ for the agent in question), or to a random location. If an agent is in a location away from home, the agent may decide to return to the home location. The probability of returning home p_{home} greatly exceeds that of making a journey ($p_{home} = 0.38$) to ensure that journeys have durations similar to those observed in the phone dataset. Thus the probability of a journey lasting n or more nights is equivalent to $(1-p_{home})^n$. Only approximately 4% of journeys involving an overnight stay have a duration of a week or more.

In order to select the location of the regular and random destinations a probability map is required that specifies the likelihood of migrating from point i to point j, denoted p_{ij}. Simini et al. (2012) recently suggested a modification to the gravity model, to give this probability in terms of the population in the origin and destination location (m_i and m_j) and the total population in a sphere of radiation equal to the distance between the two points (r_{ij}) centred at i and denoted $\hat{m}_{ij}$:

$$p_{ij} = K \frac{m_i m_j}{(m_i + \hat{m}_{ij})(m_i + m_j + \hat{m}_{ij})}. \tag{1}$$

K is the mean migration rate per individual per unit time. While Simini et al. (2012) show how the radiation model addresses some short-comings of the gravity model, initial investigation demonstrated that the standard radiation model gave a poor fit to the data since the $\hat{m}^2$ term in the denominator gives a very strong $O(r^{-4})$ dependency of the migration probability. Indeed, Simini et al. (2012) also pointed out that for the special case of uniform population density, the radiation model reduces to a gravity model with r^{-4} dependency[1].

In this first preliminary investigation, we therefore use the gravity model as in Garcia et al. (2014), while noting that the flexibility of the agent based model easily allows other mobility laws to be substituted later. The gravity law defines the probability p_{ij} of a journey from points i to point j as a function of the population density in each location m_i and m_j and the distance between them:

$$p_{ij} = \frac{K}{\eta} \left(1 - e^{\frac{-r_{ij}}{\tau_{overnight}}}\right) \frac{m_i m_j}{(r_{ij})^{\gamma}} \tag{2}$$

Here the exponent γ is set to 2 in a preliminary simulation, and η is the normalization factor to ensure that the mean probability of a trip is equal to K, which is set to 0.017 day^{-1} from the phone

[1]It is also noted that the comparison of the radiation and gravity models in Simini et al. (2012) does not state how the gravity model was fitted or which power dependency was used for the distance function.

dataset. The additional exponential term in brackets $1 - e^{\frac{-r_{ij}}{\tau_{overnight}}}$ accounts for the fact that short journeys do not result in an overnight stay (see results section), with the decay scale $\tau_{overnight}$ fitted using the D4D phone dataset.

In this preliminary version of the model, the distance between points is simply calculated as the point to point direct distance and does not account for the road network or national boundaries. For example, overland travel between key locations in Casamance such as Ziguinchor and Dakar involves transit through The Gambia, although this is obviously not the case for flights or ferry travel. Also, presently the model does not distinguish The Gambia from Senegal and allows journeys between the two countries, although these are not presented in the analysis.

Results

The probability density function of the number of regular migration destinations that individuals visit within the year of data is given in Fig.2. Approximately 40% of individuals do not make a repeated journey involving an overnight stay within the year. The other 60% of individuals do regularly visit a location away from home, but most only have one or two regular locations. Fewer than 10% of individuals have 3 or more regular locations.

The probability of a journey occurring per day between two points, considering *all* journeys made irrespective of whether they involve an overnight stay, reduces with increasing distance as expected (Fig. 3a), and very few journeys exceed a distance of 200km. The proportion of these journeys that involve an overnight stay is also shown (Fig. 3b), and highlights that over 80% of journeys that exceed 150km involve an overnight stay. The fact that not all journeys at distances exceeding 300km involve a night away is likely to be due to data inconsistencies that were not removed by the quality control algorithm.

Combining these two relationships gives the probability of a trip being made that results in an overnight stay. We applied a best fit exponential function to both of these data (Fig. 3c) which shows that the overnight proportion increases with an e-folding distance of $\tau_{overnight}$=62 km. The resulting relationship shows that the trip distance for which the number of journeys involving an overnight stay is maximum is approximately 56 km. The probability density function (PDF) is wide and positively skewed, and half of journeys involving an overnight stay are for a distance of approximately 100 km or more. It is recalled that these are mean statistics for all origins, and these relationships are likely to be spatially highly heterogeneous depending on the population density (Simini et al. 2012).

The number of journeys made from each arrondissement in Senegal that involve overnight stays, with categories divided into 10% percentiles, is shown in Fig. 4. This map shows that the journey number in general follows the population density, with the exception of the northern-most arrondissements within Saint-Louis that border Mauritania. For these the journey number considerably exceeds those made from other arrondissements of a similar population density and distance from the capital Dakar. Possible reasons for this are outlined in the model discussion below.

Preliminary results of the agent based model in terms of the number of individuals arriving in a location per square km per day are shown, with categories divided into 10% percentiles (Fig. 5). The model predicts the highest journey flux to/from Dakar as expected, but also identifies the high flux from the western, more highly populated arrondissements in the vicinity of Dakar, as well as parts of Casamance.

Discussion

In this study, we have used African mobile phone data to gain a better understanding of short-term population movements relevant to malaria transmission. As most transmission in the region is believed to involve vector species that predominantly bite in the evening/night-time, we focused on journeys that involved overnight stays. We show that the probability of a journey occurring which involves an overnight stay is highest at distances of 56km in Senegal. At higher distances, journeys with or without overnight stays are less frequent, nevertheless but over half the journeys exceed 100 km in range. The analysis also revealed that journeys are often made to regular destinations, but most people have only one or two identifiable regular locations within the year.

One limitation of the data is that phone ownership and financial means to make calls are likely to be higher in people of higher socio-economic status. This means that the assessment of population mobility may have been biased upwards. On the other hand, we will not have detected journeys on which no calls were made, biasing the assessment downwards. These limitations of using mobile phone data are adequately described in literature (e.g. Tatem 2014), and are not discussed in detail here.

Implicit in the analysis is the assumption that phone sharing practices do not overly affect the statistics or their implications for malaria parasite transport. Recently, surveys have been conducted addressing the lack of information concerning phone usage (James and Versteeg 2007), for example in Rwanda (Blumenstock and Eagle 2012) and Kenya (Wesolowski et al. 2012a). Assuming these surveys are representative of the continent, they reveal that phone sharing can be common in Africa. Blumenstock and Eagle (2012) found that 42% of respondents reported that someone else had used their phone in the last day, while 78% reported that someone had used the phone in the last week. Wesolowski et al. (2012a) reported similar sharing rates in Kenya, with strong variability from region to region. To consider how phone sharing might affect the analysis, a distinction needs to be made between household and non-household sharing. Velghe (2012) state that phone sharing is often not a result of handset shortage, but a consequence of inability to purchase credit. Non-household sharing tends to be ad hoc in nature, loaning a handset to neighbours or friend a for a single call. As such, sharing does not involve travel (the handset is generally returned to the owner after the call is made), and this form of sharing will not affect the statistics. Household-level sharing, where a single handset may also be used by a spouse, children or extended family members is more complex. One might reasonably assume that in most cases, travel away from the home involving overnight stays may be mainly conducted by a key family member, or by the family unit as a whole. However, even if different family members make separate journeys taking the mobile device with them, the implication for the analysis presented here is limited, since they return to the same home location. In summary, while phone sharing is a caveat, its impact on the present analysis is likely weak due to the focus on physical movements with the phone involving longer trips with overnight stays.

A district map showing the number of journeys involving overnight stays reveals a rich spatial texture. Journeys are highest where the population density is highest, as expected, with the peak journey density being into and out of the capital, Dakar. There are some departures from expected numbers of journeys, however. For example, the number of journeys made to the northern districts of Dagana and Podor is higher than expected from the population densities there. Although geographically far from Dakar, the northern districts are connected by the recently improved N2 highway that runs through this region. In addition to many of the western counties, the region close to the northern border of Senegal is predominately Wolof, the major ethnic group politically and numerically within Senegal. Thus one may expect enhanced migration to/from these regions relative to other Eastern counties as many Wolof based in the capital or nearby arrondissements may have family ties in the region. In contrast to the north, journeys to the southern region of Casamance, including the districts of Bignona and Sedhiou, are lower than expected from the population density. This is likely to be due partly to the separation of this region from the capital by the Gambia river, increasing effective journey times, while the ongoing conflict in the region until 2014 may also have acted to reduce population mobility.

In terms of the implications for malaria transmission, Senegal is a country with a marked north-south gradient in transmission. Interventions in the north of the country have reduced parasite prevalences to under 5%, while transmission in the south is more intense, with parasite prevalences of around 25%. The National Malaria Control Programme (Programme National de Lutte Contre le Paludisme) presently aims to eliminate transmission in the north over the medium term (e.g. Roll Back Malaria Progress & Impact Series, Focus on Senegal, 2010), a goal that could be rendered more challenging by significant and increasing population movements into the region. While a large number of the journeys are likely to originate from the capital, which has a low malaria burden (Henry et al. 2006), a proportion will originate from higher malaria areas.

This work also introduced a new agent-based migration model, that uses parameter settings derived from the phone dataset. Agent-based models allow individual-level characteristics to be assigned, such as socio-economic attributes that determine the agent's vulnerability to malaria transmission, co-infections, or in this case, information concerning the agents' habitual destinations away from a set home location. The model modifies the probability matrix of destinations in order to include only those journeys likely to result in an overnight stay. A novel aspect of the model is its definition of destinations for each agent as regular or random. This means that, in addition to modelling the overall journey numbers, the model also has the potential to represent the number of destinations per agent, more accurately capturing patterns of population movements.

Preliminary tests of the model show that, despite its present simplicity, it is able to reproduce the overall distribution journeys. That said, the model tends to underestimate the journeys made to distant destinations. In particular, it underestimates population movements in the northern counties, which have a low rural population density. This highlights the need to incorporate ethnic background into the agent based characteristic matrix, in addition to accounting for the transport network. The advantage of the agent-based approach is that the addition of such characteristics is easily accomplished and is memory efficient.

Future developments will include incorporating the transport network in the distance weightings using Dijkstra's algorithm, differentiating major and minor highway routes, and accounting for air-travel possibilities and coastal ferry connections between Dakar and Casamance. The model should also allow variation in the number of destinations, random or regular, visited by agents. The model's agent-based approach will permit developments to account for ethnicity and socio-economic status in the agent characteristics and the mobility probability network. Grid refinements would allow improved resolution of population characteristics by agents in urban areas. Developments are currently underway to incorporate personal wealth and welfare into the decision-process regarding both cyclic and permanent movements; hence the model is referred to as the welfare indexed societal demographic migration model (WISDOM). The overall goal is then to couple WISDOM to the spatially-explicit, climate driven malaria model VECTRI (Tompkins and Ermert 2013), to allow the impact of cyclic population migration on malaria transmission to be assessed in a fully dynamical coupled framework.

Acknowledgments:

This research was supported by the European Commission's Seventh Framework Research Programme through HEALTHY FUTURES grant agreement 266327. Studies and Researches were performed using mobile communication data made available by SONATEL and Orange within the D4D Challenge. An early version of this article was submitted to the D4D project in December 2014 and the use of the mobile phone data passed the D4D challenge project ethics review committee.

References

Augustijn-Beckers, E.-W., J. Flacke, and B. Retsios, 2011: Simulating informal settlement growth in Dar es Salaam, Tanzania: An agent-based housing model. *Computers, Environment and Urban Systems*, **35 (2)**, 93–103.

Bengtsson, L., X. Lu, A. Thorson, R. Garfield, and J. Von Schreeb, 2011: Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. *PLoS medicine*, **8 (8)**, 1128.

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.

Blumenstock, J. E., 2012: Inferring patterns of internal migration from mobile phone call records: Evidence from Rwanda. *Information Technology for Development*, **18 (2)**, 107–125.

Blumenstock, J. E., and N. Eagle, 2012: Divided we call: Disparities in access and use of mobile phones in rwanda. *Information Technologies & International Development*, **8 (2)**, 1–16.

Bousema, T., and Coauthors, 2012: Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS medicine*, **9 (1)**, e1001 165. doi:10.1371/journal.pmed.1001 165.

Braack, L., M. Coetzee, R. Hunt, H. Biggs, A. Cornel, and A. Gericke, 1994: Biting pattern and host-seeking behavior of Anopheles arabiensis (Diptera: Culicidae) in northeastern South Africa. *Journal of medical entomology*, **31 (3)**, 333–339.

Buckee, C. O., A. Wesolowski, N. N. Eagle, E. Hansen, and R. W. Snow, 2013: Mobile phones and malaria: modeling human and parasite travel. *Travel medicine and infectious disease*, **11 (1)**, 15–22.

Carter, R., K. N. Mendis, and D. Roberts, 2000: Spatial targeting of interventions against malaria. *Bull World Health Organ*, **78 (12)**, 1401–1411.

de Montjoye, Y.-A., C. A. Hidalgo, M. Verleysen, and V. D. Blondel, 2013: Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, **3**, doi:10.1038/srep01 376.

de Montjoye, Y.-A., Z. Smoreda, R. Trinquart, C. Ziemlicki, and V. D. Blondel, 2014: D4d-senegal: the second mobile phone data for development challenge. *arXiv preprint arXiv:1407.4885*.

De Zulueta, J., G. Kafuko, J. Cullen, and C. Pedersen, 1961: The results of the first year of a malaria eradication pilot project in Northern Kigezi (Uganda). *East African medical journal*, **38**, 1–26.

Findlay, G. M., 1946: The internal combustion engine and the spread of disease. *Brit Med J*, **2 (4486)**, 979–982.

Findley, S. E., 1994: Does drought increase migration? a study of migration from rural mali during the 1983-1985 drought. *International Migration Review*, 539–553.

Garcia, A. J., D. K. Pindolia, K. K. Lopiano, and A. J. Tatem, 2014: Modeling internal migration flows in sub-saharan africa using census microdata. *Migration Studies*, mnu036.

Garnham, P., 1945: Malaria epidemics at exceptionally high altitudes. *Brit Med J*, **2 (4410)**, 45.

Gething, P. W., and A. J. Tatem, 2011: Can mobile phone data improve emergency response to natural disasters. *PLoS medicine*, **8 (8)**, e1001 085–2pp.

Gonzalez, M. C., C. A. Hidalgo, and A.-L. Barabasi, 2008: Understanding individual human mobility patterns. *Nature*, **453 (7196)**, 779–782.

Griffitts, T., 1933: Air traffic in relation to public health. *Am J Trop Med Hyg*, **1 (3)**, 283–290.

Henry, M., and Coauthors, 2006: Urban malaria in Dakar, Senegal: chemosusceptibility and genetic diversity of Plasmodium falciparum isolates. *Am J Trop Med Hyg*, **75 (1)**, 146–151.

Hoshen, M. B., and A. P. Morse, 2004: A weather-driven model of malaria transmission. *Malar J*, **3**, 32, doi:10.1029/2012GL054 040.

James, J., and M. Versteeg, 2007: Mobile phones in africa: how much do we really know? *Social indicators research*, **84 (1)**, 117–126.

Jones, A. E., and A. P. Morse, 2010: Application and validation of a seasonal ensemble prediction system using a dynamic malaria model. *J Clim*, **23**, 4202–4215.

Kniveton, D., C. Smith, and S. Wood, 2011: Agent-based model simulations of future changes in migration flows for burkina faso. *Global Environmental Change*, **21**, S34–S40.

Kniveton, D. R., C. D. Smith, and R. Black, 2012: Emerging migration flows in a changing climate in dryland africa. *Nature Climate Change*, **2 (6)**, 444–447.

Knols, B. G. J., R. de Jong, and W. Takken, 1995: Differential attractiveness of isolated humans to mosquitoes in Tanzania. *Trans R Soc Trop Med Hyg*, **89**, 604–606.

Linard, C., M. Gilbert, R. W. Snow, A. M. Noor, and A. J. Tatem, 2012: Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS ONE*, **7**, e31 743, doi:10.1371/journal.pone.0031 743.

Lindsay, S. W., J. H. Adiamah, J. E. Miller, R. J. Pleass, and J. R. M. Armstrong, 1993: Variation in attractiveness of human subjects to malaria mosquitoes (Diptera: Culicidae) in The Gambia. *Journal of medical entomology*, **30 (2)**, 368–373.

Lynch, C., and C. Roper, 2011: The transit phase of migration: circulation of malaria and its multidrug-resistant forms in africa. *PLoS medicine*, **8 (5)**, 612.

Marques, A. C., 1987: Human migration and the spread of malaria in brazil. *Parasitology Today*, **3 (6)**, 166–170.

Martens, P., and L. Hall, 2000: Malaria on the move: human population movement and malaria transmission. *Emerg Infect Dis*, **6**, 103–109.

Mukabana, W. R., W. Takken, R. Coe, and B. G. J. Knols, 2002: Host-specific cues cause differential attractiveness of kenyan men to the african malaria vector anopheles gambiae. *Malar J*, **1**, 17, doi:10.1186/1475–2875–1–17.

Parker, J., and J. M. Epstein, 2011: A distributed platform for global-scale agent-based models of disease transmission. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, **22 (1)**, 2.

Pates, H., and C. Curtis, 2005: Mosquito behavior and vector control. *Annu. Rev. Entomol.*, **50**, 53–70.

Pindolia, D. K., A. J. Garcia, Z. Huang, D. L. Smith, V. A. Alegana, A. M. Noor, R. W. Snow, and A. J. Tatem, 2013: The demographics of human and malaria movement and migration patterns in east africa. *Malar J*, **12 (397)**, 10–1186.

Pindolia, D. K., A. J. Garcia, A. Wesolowski, D. L. Smith, C. O. Buckee, A. M. Noor, R. W. Snow, and A. J. Tatem, 2012: Human movement data for malaria control and elimination strategic planning. *Malar J*, **11 (1)**, 205.

Prothero, R. M., 1961: Population movements and problems of malaria eradication in Africa. *Bull World Health Organ*, **24 (4-5)**, 405.

Shahabi, C., M. R. Kolahdouzan, and M. Sharifzadeh, 2003: A road network embedding technique for k-nearest neighbor search in moving object databases. *GeoInformatica*, **7 (3)**, 255–273.

Simini, F., M. C. González, A. Maritan, and A.-L. Barabási, 2012: A universal model for mobility and migration patterns. *Nature*, **484 (7392)**, 96–100.

Talisuna, A. O., A. M. Noor, A. P. Okui, and R. W. Snow, 2015: The past, present and future use of epidemiological intelligence to plan malaria vector control and parasite prevention in Uganda. *Malar J*, **14 (1)**, 158.

Tatem, A. J., 2014: Mapping population and pathogen movements. *Int Health*, **6 (1)**, 5–11.

Tatem, A. J., Y. Qiu, D. L. Smith, O. Sabot, A. S. Ali, B. Moonen, and Coauthors, 2009: The use of mobile phone data for the estimation of the travel patterns and imported plasmodium falciparum rates among zanzibar residents. *Malar J*, **8 (287)**, 10–1186.

Tatem, A. J., D. J. Rogers, and S. I. Hay, 2006: Estimating the malaria risk of african mosquito movement by air travel. *Malar J*, **5 (1)**, 57.

Tatem, A. J., and D. L. Smith, 2010: International population movements and regional plasmodium falciparum malaria elimination strategies. *Proc Nat Acad Sci*, **107**, 12 222–12 227.

Todaro, M. P., 1969: A model of labor migration and urban unemployment in less developed countries. *The American Economic Review*, 138–148.

Tompkins, A. M., and F. Di Giuseppe, 2015: Potential predictability of malaria using ECMWF monthly and seasonal climate forecasts in Africa. *J. Appl. Meteor. Clim*, **54**, 521–540.

Tompkins, A. M., and V. Ermert, 2013: A regional-scale, high resolution dynamical malaria model that accounts for population density, climate and surface hydrology. *Malar J*, **12**, doi:10.1186/1475–2875–12–65.

Tompkins, A. M., L. Larsen, N. McCreesh, and D. M. Taylor, 2015: To what extent does climate explain variations in reported malaria cases in early 20th century Uganda? *Geospat Health*, **10**, submitted.

Torres-Sorando, L., and D. J. Rodrıguez, 1997: Models of spatio-temporal dynamics in malaria. *Ecological modelling*, **104 (2)**, 231–240.

Velghe, F., 2012: Deprivation, distance and connectivity: The adaptation of mobile phone use to a life in wesbank, a post-apartheid township in south africa. *Discourse, Context & Media*, **1 (4)**, 203–216.

Watkins, B., and M. L. Fleisher, 2002: Tracking pastoralist migration: lessons from the ethiopian somali national regional state. *Human organization*, **61 (4)**, 328–338.

Wesolowski, A., C. O. Buckee, L. Bengtsson, E. Wetter, X. Lu, and A. J. Tatem, 2014: Commentary: containing the ebola outbreak-the potential and challenge of mobile network data. *PLoS currents*, **6**.

Wesolowski, A., C. O. Buckee, D. K. Pindolia, N. Eagle, D. L. Smith, A. J. Garcia, and A. J. Tatem, 2013: The use of census migration data to approximate human movement patterns across temporal scales. *PLoS ONE*, **8 (1)**, e52 971.

Wesolowski, A., N. Eagle, A. M. Noor, R. W. Snow, and C. O. Buckee, 2012a: Heterogeneous mobile phone ownership and usage patterns in kenya. *PLoS ONE*, **7 (4)**, e35 319.

Wesolowski, A., N. Eagle, A. J. Tatem, D. L. Smith, A. M. Noor, R. W. Snow, and C. O. Buckee, 2012b: Quantifying the impact of human mobility on malaria. *Science*, **338 (6104)**, 267–270.

LIST OF FIGURES

Fig. 1. Key decision tree for the v1 beta release version of the agent based migration model WISDOM. The dashed lines for the option of the agent returning home are to emphasize that this option is only possible if the agent is away from home when the movement decision is taken. . 24

Fig. 2. Proportional of individuals that have *n* regular cyclic migration locations, where *n* is given on the x axis. The category *3* refers to "3 and over" journeys. 25

Fig. 3. (a) Probability of a trip as a function of distance with an exponential fit shown with the solid line, (b) probability that a trip includes a night away and (c) graph of best fit relationship to overnight stays (right axis) and journey probability as a function of distance (left axis) and combined probability of a trip being made that involves an overnight stay. 26

Fig. 4. Map of total journeys made from each arrondissement divided into ten categories of approximate 10% percentiles. 27

Fig. 5. Preliminary results of the WISDOM model simulations. Units are individuals per square km arriving per month at each location divided into approximate 10 percentile categories. . . . 28

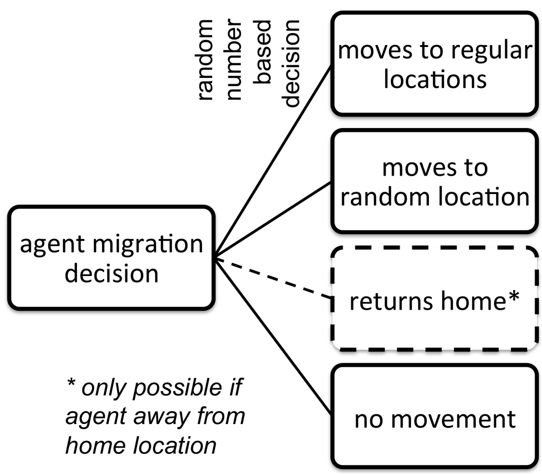


FIG. 1. Key decision tree for the v1 beta release version of the agent based migration model WISDOM. The dashed lines for the option of the agent returning home are to emphasize that this option is only possible if the agent is away from home when the movement decision is taken.

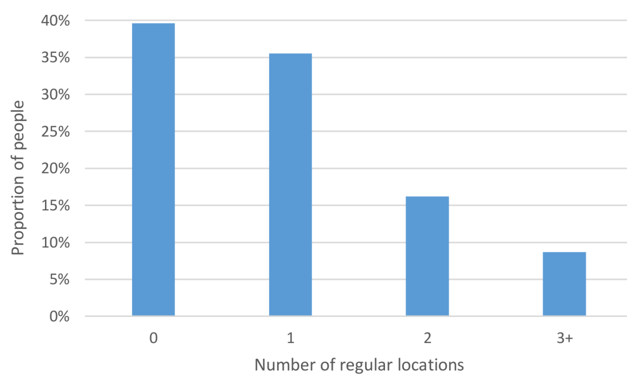


FIG. 2. Proportional of individuals that have n regular cyclic migration locations, where n is given on the x axis. The category *3* refers to “3 and over” journeys.

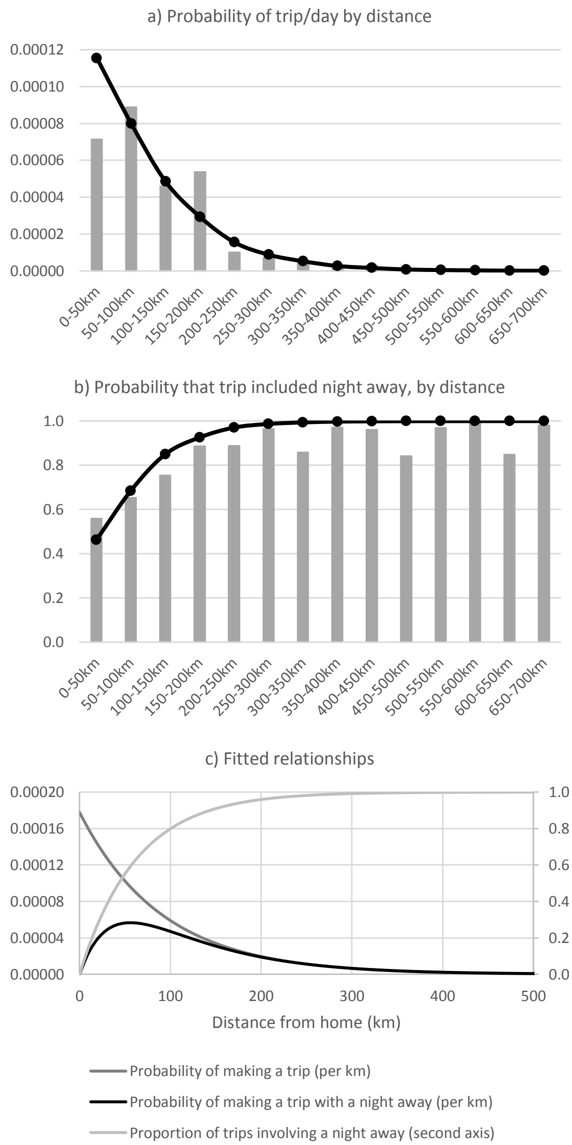


FIG. 3. (a) Probability of a trip as a function of distance with an exponential fit shown with the solid line, (b) probability that a trip includes a night away and (c) graph of best fit relationship to overnight stays (right axis) and journey probability as a function of distance (left axis) and combined probability of a trip being made that involves an overnight stay.

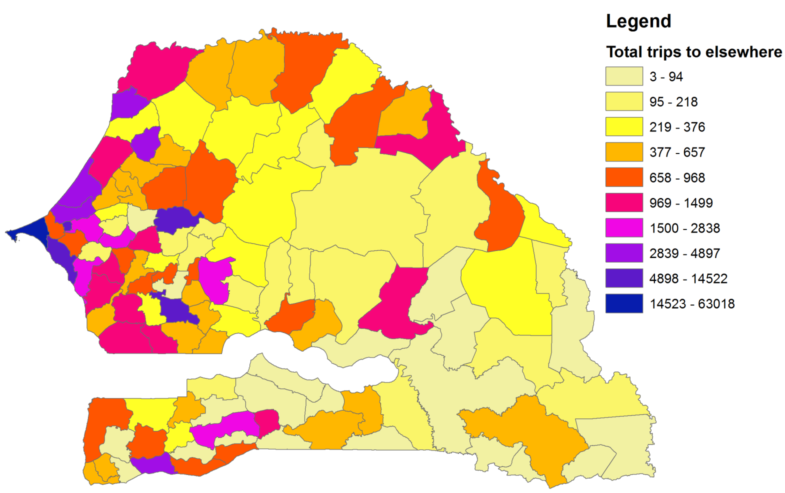
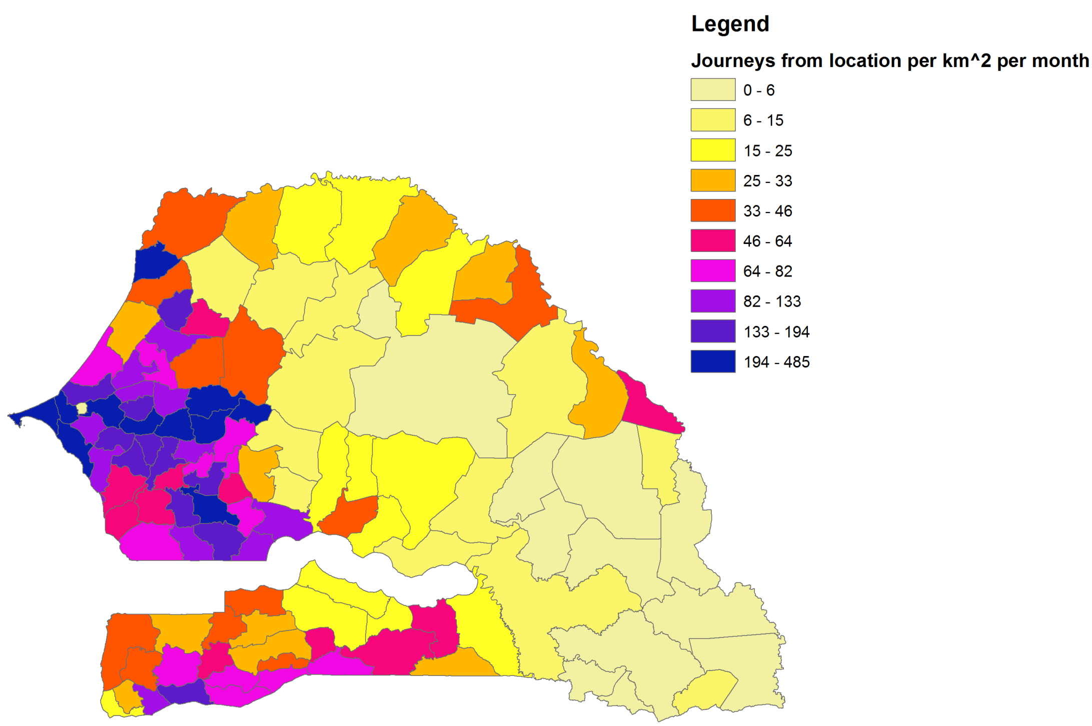


FIG. 4. Map of total journeys made from each arrondissement divided into ten categories of approximate 10% percentiles.



33 FIG. 5. Preliminary results of the WISDOM model simulations. Units are individuals per square km arriving
34 per month at each location divided into approximate 10 percentile categories.