

Activity Space Maps: a novel human mobility data set for quantifying time spent at risk

Daniel T. Citron
danielcitron@fb.com
Facebook, Inc.
Menlo Park, CA, USA

Shankar Iyer
shankar94@fb.com
Facebook, Inc.
Menlo Park, CA, USA

Robert C. Reiner
bcreiner@uw.edu
University of Washington, Department of Health Metrics
Sciences
University of Washington, Institute for Health Metrics and
Evaluation
Seattle, WA, USA

David L. Smith
smitdave@uw.edu
University of Washington, Department of Health Metrics
Sciences
University of Washington, Institute for Health Metrics and
Evaluation
Seattle, WA, USA

ABSTRACT

Activity Space Maps are a novel global-scale movement and mobility data set which describes how people distribute their time through geographic space. The maps are intended for use by researchers for the purposes of epidemiological modeling. Activity Space Maps are designed to complement existing digitally-collected mobility data sets by quantifying the amount of time that people spend in different locations. This information is important for estimating the duration of contact with the environment and the potential risk of exposure to disease. More concretely, the type of information contained in Activity Space Maps will make it easier to model the spatial transmission patterns of vector-borne diseases like malaria and Dengue fever. We will discuss the motivation for designing Activity Space Maps, how the maps are generated from mobile phone user app location history data, and discuss an example use case demonstrating how such data may be used together with spatial epidemiological data to advance our understanding of spatial disease patterns and the relationship between travel behaviors and infection risk.

KEYWORDS

human mobility, epidemiology, datasets

1 INTRODUCTION

Data sets which describe human movement patterns are important for understanding epidemiologically relevant interactions and contact patterns. Most disease-causing pathogens are not able to move great distances on their own. Instead, the pathogens can be transported between places by human and animal hosts. Infected people may travel from their home territory to another location and infect others there, causing a new outbreak; or, travelers may contract an illness while away and unintentionally spread it to their

households and neighbors after returning home. Knowing where it is that people go and how they move around in geographical space is key for mapping out spatial patterns of contact and exposure [16, 17, 27].

Until recently, it was very difficult to obtain accurate data on human movement patterns at scale. Medical surveys of patients are subject to selection bias (they only survey people who end up at a healthcare facility) and patient recall bias (may not accurately report their travel history) [13]. Distributing surveys to the general population is expensive to accomplish at scale [4] and similarly affected by recall biases. More recently, it has become possible for researchers to distribute GPS location history trackers to people as a way of recording their movements, but these studies are hard to sustain at a large scale [19]. With the widespread adoption of mobile phones, digitally-collected mobile-phone user location history data offers an unprecedented opportunity to quantify patterns of movement behavior at a broad scale [1, 11, 14, 25]. Effectively, this becomes an observational study similar to the GPS location history tracker studies, but without the need to distribute and maintain devices.

The formatting of digitally-collected location history data matters for its applicability for research. Many early movement data sets have been aggregated and formatted as origin-destination matrices, which count the number of users who travel from one location to another during an observation period [12]. This formatting is useful both for parsimony as well as for disguising individual user trajectories for the sake of preserving privacy. While valuable for understanding population flows, data formatted as origin-destination matrices do not contain information on how much time a traveler spends in any location. For example, the matrix might count one user passing from location A to location B on a particular day, but not know whether the user spent the entire day in location B or whether they immediately traveled from location B to location C. For epidemiological purposes, the duration of time that a user spends in any given location is key for understanding epidemiologically relevant contact and exposure patterns [2, 6, 23]. After all, a person who spends only a day in a setting where they are at risk of contracting malaria is less likely to become sick than if they spend two weeks in that setting.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

KDD Workshop on Data-driven Humanitarian Mapping, 27th ACM SIGKDD Conference, Virtual Conference,

2 DATASET DESCRIPTION

Activity Space Maps are a novel data set developed through Facebook's Data for Good initiative. They are built using de-identified Facebook mobile-app user location history data of users who consented to share their precise location and opted into Facebook's location history collection. The maps aggregate movement trajectories across space and time such that they represent a probability distribution, describing the average amount of time that a user from one location spends visiting other locations. That is to say, what is the probability that an average individual spends in one location, given that their home residence is found in another location. The maps are formatted similarly to the space utilization functions that are found in ecology which are constructed using GPS trackers and aggregated as a way of summarizing how it is that animals move around and use space [9].

Activity Space Maps are generated by collecting and aggregating individual users' location histories. All users included in the pipeline have consented to collection of their precise location and have opted into Facebook's location history collection. The raw source data are location history traces, initially recorded as a series of tuples of the form user, time, location. Initially, locations are binned using Microsoft's Level 12 Bing tile maps, a quadkey map of square regions approximately 9.8 kilometers to a side at the equator [18]. Each user's trajectory is binned into a series of 24-hour periods. Each 24-hour period is denoted by a date stamp and contains one nighttime period, spanning from 8 p.m. on the previous date to 6 a.m. on the given date, and one daytime period, spanning from 8 a.m. to 6 p.m. on the given date. For each date, modal locations for both the daytime and nighttime periods are extracted. At this stage, all users who do not consistently use location history services and all users who appear in the data set fewer than 7 out of the past 28 nights are excluded. The next step is to aggregate the total number of person-periods (i.e., person-days and person-nights) spent by each individual in each location across the last 28 days. This protects users' privacy by removing links between specific locations and specific visit times and instead summarizes each user's space utilization over the previous four weeks. Each user's home location is estimated using the most frequently visited nighttime modal location during the 28-day aggregation period. The assumption here is that, while there are some instances of people spending their nights away from their home residences, the vast majority of people are most likely to sleep at home on most nights.

Lastly, Activity Space Maps are generated by aggregating over all users in each home location, producing a probability distribution showing how the average individual who lives in a particular location distributes their time across geographical space. The resulting dataset quantifies, for each home-visit tile pair, the relative fraction of person-days and person-nights spent in the visit tile location by users who live in the home tile location. In this way, Activity Space Maps quantify typical movement behavior and patterns of users who live in the same location. Figure 1 shows two example plots of Activity Space Maps data collected for a home tile in Rome, Italy. Both the daytime and nighttime spent distributions are shown. The temporal aggregation means that the dataset reflects daily, weekly, or monthly movement patterns such as commuting or short-term

trips. The data set is not intended to capture or reflect longer-term patterns such as seasonal migration or permanent relocation.

Preserving individual user privacy is a key feature of Activity Space Maps and all Facebook Data for Good products. No individual trajectories or records which pinpoint any individual's visit to a location at a particular time are exposed through this data set. All locations which don't meet a population density threshold are excluded from the data set for the sake of concealing any individual trajectories. Following guidance from humanitarian organizations who have partnered with Facebook Data for Good, cross-border movements are excluded from the dataset for the sake of protecting at-risk populations who move in times of crisis.

Activity Space Maps collect data in 151 countries and territories worldwide. In terms of the number of users sampled, Activity Space Maps has the most coverage for countries in Europe, North America, South America, South Asia, and Southeast Asia, but less coverage in other parts of Asia, Africa, and Oceania. Activity Space Maps are recorded on an ongoing weekly basis, beginning in mid-April of 2021. Facebook Data for Good plans to release Activity Space Maps to vetted Data for Good external partners in the second half of 2021.

We must highlight a few properties of how Activity Space Maps are constructed which have the potential to skew or bias the movement patterns which they represent. Primarily, it is likely that the sampled population does not represent an unbiased sample of the overall population [25, 26]. The data only reflect those who have opted into location history services through the Facebook mobile app - it is plausible to think that such individuals may have different movement patterns than those who do not access these features. Furthermore, the data set is only able to represent the movements of people who own mobile phones while they are traveling in areas with access to phone service. For these reasons, we caution users that the data sets are likely to over-represent the movements of people who live in higher-income, urban areas and under-represent the movements of people who live in lower-income, rural areas. Secondly, there is also the potential problem of under-sampling in low population density areas, combined with the privacy-related exclusion of low density areas. For this reason, Activity Space Maps currently are able to reveal the connectivity patterns originating in high-density, urban areas but are less likely to portray connectivity patterns originating in low-density, rural areas. Finding methods to account for and correct potential sources of bias in the data is an ongoing line of research, both within the Facebook Data for Good team as well as within the broader research community [10, 28].

3 APPLICATIONS

There are a variety of potential epidemiological applications for Activity Space Maps. Generally speaking, Activity Space Maps have a potential for improving understanding of epidemiological patterns where movement and travel expose people to hazards in the environment. To give one example, we outline research which combines Activity Space Maps with malaria prevalence maps from the Malaria Atlas Project (MAP) [5, 24] using a mechanistic transmission model to show how travel drives the incidence of malaria among people living in urban areas in sub-Saharan Africa. This line of research draws on earlier analysis performed for Bioko Island in Equatorial

KDD Workshop on Data-driven Humanitarian Mapping, 27th ACM SIGKDD Conference, Virtual Conference, Activity Space Maps: a novel human mobility data set for quantifying time spent at risk

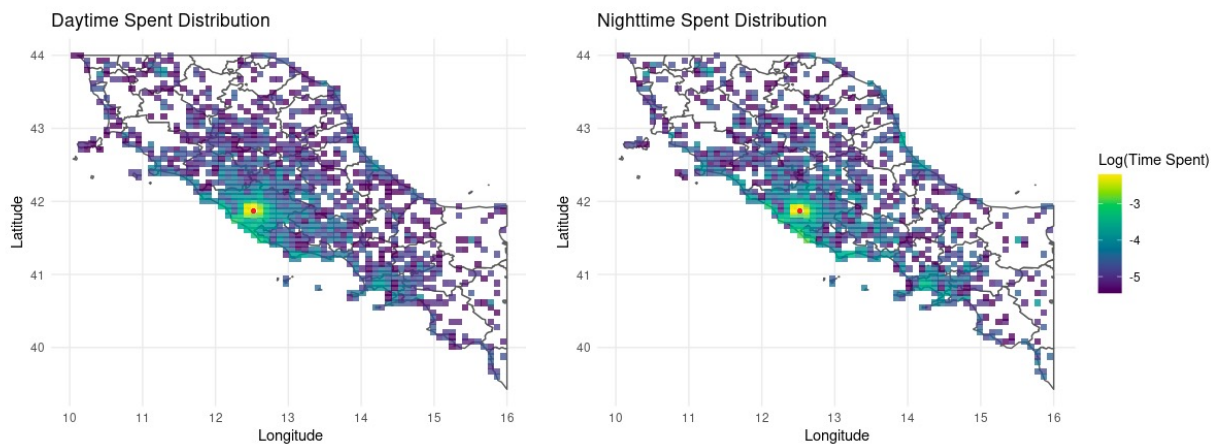


Figure 1: Activity Space Maps example: The data plotted in the two maps were collected for a home tile in Rome, Italy during the 28-day period ending on June 15th, 2021. The map on the left shows the daytime spent distribution and the map on the right shows the nighttime spent distribution. The red dot indicates the home location. The other colors denote the logarithm of the relative probability of spending a day or night in each other location.

Guinea [3, 7, 8], which estimated the fraction of malaria prevalence which may be attributable to exposure to infectious mosquitoes while traveling.

Malaria is a vector-borne parasitic disease which primarily affects people living in tropical climates in sub-Saharan Africa, Asia, and Central and Southern America. The WHO estimates that over 220 million malaria cases and 400,000 deaths due to malaria occurred in 2019 alone, making it one of the most dangerous causes of death and disease [15]. Malaria cases primarily affect children and people living in rural settings. Malaria is spread by *Anopheles* mosquitoes, who pass parasites to human hosts when they blood feed. Because malaria is transmitted by an insect vector, the risk experienced by people depends strongly on whether or not those people spend time in areas where infectious vector mosquitoes live: one will be at risk of becoming infected if one spends time in areas where the local environment supports a viable vector population. It is for this reason that knowing where people spend their time is so important for understanding who is at risk and the patterns of transmission.

The Malaria Atlas Project (MAP) produces raster maps of malaria prevalence, or parasite rate (PR) [5, 24], showing the proportion of infected people in a given location. The maps reveal the geographical heterogeneity of malaria prevalence, which reflects the geographical heterogeneity in the local disease-producing ecology as experienced by human hosts. The PR maps show where it is that infected people are located, but the maps do not show where it is that the people became infected. In many cases, most likely people are becoming infected in their own homes. However, some of those cases are likely due to exposure that occurred during travel. Malaria vector mosquitoes typically do not thrive in urban areas. It is likely that cases of malaria reported in urban areas are attributable to travel.

Bioko Island represents one setting where this is likely to be the case. Since 2004, the Bioko Island Malaria Elimination Program has

implemented an extensive program of interventions against malaria [4]. Despite this, malaria persists among island residents, most of whom live in the urbanized capital city of Malabo. Using travel data collected through surveys in 2015-2018, it was estimated that much of the malaria occurring in Malabo could be attributable to travel to mainland Equatorial Guinea, where there is a high risk of malaria [3, 7, 8]. This is an example of how travel data may be used to discover source-sink disease dynamics, where the high incidence of disease in one location drives the occurrence of disease in another location because of people traveling from one location to another [16, 17].

Activity Space Maps may be used to perform similar analysis, this time using digitally-collected data to infer travel patterns. The leftmost plot in Figure 2 shows the Activity Space Map for individuals living in the highlighted location in Lagos, Nigeria (marked with a red dot). The distribution plotted shows that most time is spent in the area of Lagos, but that some trips are taken to Abuja and the Niger river delta region. The center plot in Figure 2 shows the Malaria Atlas Project's estimated prevalence levels of malaria (PR) in 2015 among children aged 2-10, plotted only in the same areas for which Activity Space Maps also provides data. The maps show how dense urban areas like Lagos and Abuja tend to have lower prevalence, but that there still are some infected people living in these areas.

By combining together the epidemiological prevalence data and the movement data, it is possible to infer where it is likely that people are becoming exposed and infected. The next step of the analysis utilizes on a mechanistic model of malaria transmission [3, 16, 17, 20]. The model relates the prevalence to the force of infection (FOI) experienced by individuals in different locations. The FOI is a way of quantifying risk to individuals, an infection rate which reflects exposure to infectious vectors in the local environment. Within the model, the average individual's overall FOI is expressed as a weighted average of the amount of time spent in each location

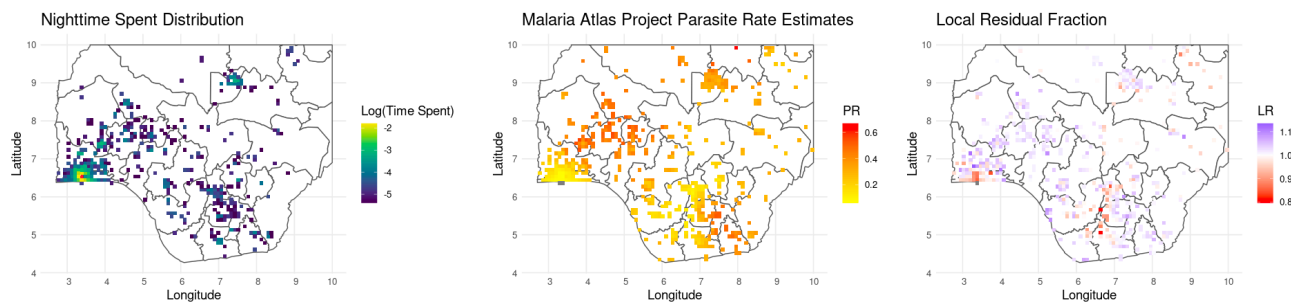


Figure 2: Example analysis: Left plot illustrates the distribution of nighttime periods spent by residents in part of Lagos, Nigeria (red dot indicates home location), with the color scale showing the logarithm of the relative probability of spending a night in each visit location. Center plot shows Malaria Atlas Project *Plasmodium falciparum* Parasite Rate (PR) estimates for children aged 2-10 in Nigeria in 2015. Right plot shows the local residual fraction of prevalence which is attributable to local transmission. This last plot is the result of analysis which uses Activity Space Maps to estimate the influence which travel has on local transmission. The areas marked in red are locations where malaria exposure occurs when people travel to other areas, and the areas marked in purple are locations where malaria exposure mostly occurs locally.

visited, multiplied by the FOI in each location. Beginning with the prevalence map, inverting the model results in an estimate of local FOI in each location on the map.

Knowing the true FOI in each location makes it possible to estimate the exposure rate experienced by each individual in each place they travel to. Activity Space Maps describe how an average individual distributes their time through geographical space, making it possible to derive estimates of where it is that people are exposed to malaria. The rightmost plot in Figure 2 plots Local Residual Fraction, or the fraction of prevalence that is attributable to local transmission at one's home location [3]. The areas marked in red, such as in Lagos, are areas where exposure that occurs while traveling has a significant impact on local malaria prevalence. That is to say, people with malaria may have been exposed elsewhere. The areas marked in purple, such as in the region to the north of Lagos, are areas where local transmission is strong and travel does not have much influence on overall prevalence. This simple preliminary analysis can provide some insights into where risk of infection occurs, and can assist with strategic planning for intervention programs. The results suggest that there are some areas where interventions to decrease local exposure should be prioritized, but also that there are other areas where additional protections for travelers could be helpful.

Replicating the analysis of travel and malaria seen in the Bioko Island study in a large country using traditional survey data would be very costly, but Activity Space Maps has made it possible to perform a spatial analysis of transmission patterns for Nigeria using digitally-collected data. This analysis represents one example of how the movement maps may be applied to improve understanding of spatial patterns of disease transmission. Generally speaking, Activity Space Maps are designed to understand how movement patterns bring people into contact with hazards in the environment. Activity Space Maps may be able to support research on other vector-borne diseases, such as Dengue fever [21, 22]. For water-borne diseases such as cholera, exposure occurs through a different mechanism, but Activity Space Maps may provide insights into identifying which populations spend time in places where they are

at risk. Lastly, there is potential for applications related to directly transmitted pathogens such as SARS-CoV-2 [1, 14], where Activity Space Maps may support existing movement data sources to understand which populations travel to high-transmission locations.

REFERENCES

- [1] Caroline O Buckee, Satchit Balsari, Jennifer Chan, Mercè Crosas, Francesca Dominici, Urs Gasser, Yonatan H Grad, Bryan Grenfell, M Elizabeth Halloran, Moritz UG Kraemer, et al. 2020. Aggregated mobility data could help fight COVID-19. *Science (New York, NY)* 368, 6487 (2020), 145–146.
- [2] Daniel T Citron, Carlos A Guerra, Andrew J Dolgert, Sean L Wu, John M Henry, David L Smith, et al. 2021. Comparing metapopulation dynamics of infectious diseases under different models of human movement. *Proceedings of the National Academy of Sciences* 118, 18 (2021).
- [3] Daniel T Citron, Carlos A Guerra, Guillermo A Garcia, Sean L Wu, Katherine E Battle, Harry S Gibson, and David L Smith. 2021. Quantifying malaria acquired during travel and its role in malaria elimination on Bioko Island. *medRxiv* (2021).
- [4] Jackie Cook, Dianna Hergott, Wonder Phiri, Matilde Riloha Rivas, John Bradley, Luis Segura, Guillermo Garcia, Chris Schwabe, and Immo Kleinschmidt. 2018. Trends in parasite prevalence following 13 years of malaria interventions on Bioko island, Equatorial Guinea: 2004–2016. *Malaria journal* 17, 1 (2018), 1–13.
- [5] Peter W Gething, Anand P Patil, David L Smith, Carlos A Guerra, Iqbal RF Elyazar, Geoffrey L Johnston, Andrew J Tatem, and Simon I Hay. 2011. A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malaria journal* 10, 1 (2011), 1–16.
- [6] John R Giles, Elisabeth zu Erbach-Schoenberg, Andrew J Tatem, Lauren Gardner, Ottar N Bjørnstad, CJE Metcalf, and Amy Wesolowski. 2020. The duration of travel impacts the spatial dynamics of infectious diseases. *Proceedings of the National Academy of Sciences* 117, 36 (2020), 22572–22579.
- [7] Carlos A Guerra, Daniel T Citron, Guillermo A García, and David L Smith. 2019. Characterising malaria connectivity using malaria indicator survey data. *Malaria journal* 18, 1 (2019), 1–12.
- [8] Carlos A Guerra, Su Yun Kang, Daniel T Citron, Dianna EB Hergott, Megan Perry, Jordan Smith, Wonder P Phiri, José O Osá Nfumu, Jeremias N Mba Eyono, Katherine E Battle, et al. 2019. Human mobility patterns and malaria importation on Bioko Island. *Nature communications* 10, 1 (2019), 1–10.
- [9] Mevin B Hooten, Devin S Johnson, Brett T McClintock, and Juan M Morales. 2017. *Animal movement: statistical models for telemetry data*. CRC press.
- [10] MUG Kraemer, N Golding, D Bisanzio, S Bhatt, DM Pigott, SE Ray, OJ Brady, JS Brownstein, NR Faria, DAT Cummings, et al. 2019. Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings. *Scientific reports* 9, 1 (2019), 1–11.
- [11] Moritz UG Kraemer, Adam Sadilek, Qian Zhang, Nahema A Marchal, Gauravi Tuli, Emily L Cohn, Yulin Hswen, T Alex Perkins, David L Smith, Robert C Reiner, et al. 2020. Mapping global variation in human mobility. *Nature Human Behaviour* 4, 8 (2020), 800–810.
- [12] Paige Maas, Shankar Iyer, Andreas Gros, Wonhee Park, Laura McGorman, Chaya Nayak, and P Alex Dow. 2019. Facebook Disaster Maps: Aggregate Insights for

- Crisis Response & Recovery.. In *KDD*, Vol. 19. 3173.
- [13] John M Marshall, Sean L Wu, Samson S Kiware, Micky Ndhlovu, André Lin Ouédraogo, Mahamoudou B Touré, Hugh J Sturrock, Azra C Ghani, Neil M Ferguson, et al. 2018. Mathematical models of human mobility of relevance to malaria transmission in Africa. *Scientific reports* 8, 1 (2018), 1–12.
- [14] Nuria Oliver, Bruno Lepri, Harald Sterly, Renaud Lambiotte, Sébastien Deletaille, Marco De Nadai, Emmanuel Letouzé, Albert Ali Salah, Richard Benjamins, Ciro Cattuto, Vittoria Colizza, Nicolas de Cordes, Samuel P. Fraiberger, Till Koebe, Sune Lehmann, Juan Murillo, Alex Pentland, Phuong N Pham, Frédéric Pivetta, Jari Saramäki, Samuel V. Scarpino, Michele Tizzoni, Stefaan Verhulst, and Patrick Vinck. 2020. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances* 6, 23 (2020). <https://doi.org/10.1126/sciadv.abc0764> arXiv:<https://advances.sciencemag.org/content/6/23/eabc0764.full.pdf>
- [15] World Health Organization. 2020. World malaria report 2020: 20 years of global progress and challenges. (2020).
- [16] Nick W Ruktanonchai, Patrick DeLeenheer, Andrew J Tatem, Victor A Alegana, T Trevor Caughlin, Elisabeth zu Erbach-Schoenberg, Christopher Lourenço, Corrine W Ruktanonchai, and David L Smith. 2016. Identifying malaria transmission foci for elimination using human mobility data. *PLoS computational biology* 12, 4 (2016), e1004846.
- [17] Nick W Ruktanonchai, David L Smith, and Patrick De Leenheer. 2016. Parasite sources and sinks in a patched Ross–Macdonald malaria model with human and mosquito movement: implications for control. *Mathematical biosciences* 279 (2016), 90–101.
- [18] Joe Schwartz. [n.d.]. Bing Tile Maps System. Retrieved June 1, 2021 from <https://docs.microsoft.com/en-us/bingmaps/articles/bing-maps-tile-system>
- [19] Kelly M Searle, Jaiolos Lubinda, Harry Hamapumbu, Timothy M Shields, Frank C Curriero, David L Smith, Philip E Thuma, and William J Moss. 2017. Characterizing and quantifying human movement patterns using GPS data loggers in an area approaching malaria elimination in rural southern Zambia. *Royal Society open science* 4, 5 (2017), 170046.
- [20] David L Smith, Katherine E Battle, Simon I Hay, Christopher M Barker, Thomas W Scott, and F Ellis McKenzie. 2012. Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. *PLoS pathogens* 8, 4 (2012), e1002588.
- [21] K James Soda, Sean M Moore, Guido Espana, Jonathan Bloedow, Benoit Raybaud, Ben Althouse, MA Johansson, Edward Wenger, Philip Welkhoff, TA Perkins, et al. 2018. DTK–Dengue: A new agent-based model of dengue virus transmission dynamics. *bioRxiv* (2018), 376533.
- [22] Steven T Stoddard, Brett M Forshey, Amy C Morrison, Valerie A Paz-Soldan, Gonzalo M Vazquez-Prokopec, Helvio Astete, Robert C Reiner, Stalin Vilcarrromero, John P Elder, Eric S Halsey, et al. 2013. House-to-house human movement drives dengue virus transmission. *Proceedings of the National Academy of Sciences* 110, 3 (2013), 994–999.
- [23] Steven T Stoddard, Amy C Morrison, Gonzalo M Vazquez-Prokopec, Valerie Paz Soldan, Tadeusz J Kochel, Uriel Kitron, John P Elder, and Thomas W Scott. 2009. The role of human movement in the transmission of vector-borne pathogens. *PLoS Negl Trop Dis* 3, 7 (2009), e481.
- [24] Daniel J Weiss, Tim CD Lucas, Michele Nguyen, Anita K Nandi, Donal Bisanzio, Katherine E Battle, Ewan Cameron, Katherine A Twohig, Daniel A Pfeffer, Jennifer A Rozier, et al. 2019. Mapping the global prevalence, incidence, and mortality of *Plasmodium falciparum*, 2000–17: a spatial and temporal modelling study. *The Lancet* 394, 10195 (2019), 322–331.
- [25] Amy Wesolowski, Caroline O Buckee, Kenth Engø-Monsen, and Charlotte Jessica Eland Metcalf. 2016. Connecting mobility to infectious diseases: the promise and limits of mobile phone data. *The Journal of infectious diseases* 214, suppl_4 (2016), S414–S420.
- [26] Amy Wesolowski, Nathan Eagle, Abdisalan M Noor, Robert W Snow, and Caroline O Buckee. 2013. The impact of biases in mobile phone ownership on estimates of human mobility. *Journal of the Royal Society Interface* 10, 81 (2013), 20120986.
- [27] Amy Wesolowski, Nathan Eagle, Andrew J Tatem, David L Smith, Abdisalan M Noor, Robert W Snow, and Caroline O Buckee. 2012. Quantifying the impact of human mobility on malaria. *Science* 338, 6104 (2012), 267–270.
- [28] Amy Wesolowski, Wendy Prudhomme O’Meara, Nathan Eagle, Andrew J Tatem, and Caroline O Buckee. 2015. Evaluating spatial interaction models for regional mobility in sub-Saharan Africa. *PLoS Comput Biol* 11, 7 (2015), e1004267.