

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



LSHTM Research Online

Kraemer, MUG; Golding, N; Bisanzio, D; Bhatt, S; Pigott, DM; Ray, SE; Brady, OJ; Brownstein, JS; Faria, NR; Cummings, DAT; +5 more... Pybus, OG; Smith, DL; Tatem, AJ; Hay, SI; Reiner, RC; (2019) Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings. SCIENTIFIC REPORTS, 9 (1). ISSN 2045-2322 DOI: <https://doi.org/10.1038/s41598-019-41192-3>

Downloaded from: <http://researchonline.lshtm.ac.uk/4652710/>

DOI: <https://doi.org/10.1038/s41598-019-41192-3>

Usage Guidelines:

Please refer to usage guidelines at <http://researchonline.lshtm.ac.uk/policies.html> or alternatively contact researchonline@lshtm.ac.uk.

Available under license: <http://creativecommons.org/licenses/by/2.5/>

<https://researchonline.lshtm.ac.uk>

SCIENTIFIC REPORTS



OPEN

Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings

M. U. G. Kraemer^{1,2,3}, N. Golding⁴, D. Bisanzio^{5,6}, S. Bhatt⁷, D. M. Pigott⁸, S. E. Ray⁸, O. J. Brady⁹, J. S. Brownstein^{2,3}, N. R. Faria¹, D. A. T. Cummings^{10,11}, O. G. Pybus¹, D. L. Smith^{8,12}, A. J. Tatem^{13,14}, S. I. Hay⁸ & R. C. Reiner Jr.⁸

Human mobility is an important driver of geographic spread of infectious pathogens. Detailed information about human movements during outbreaks are, however, difficult to obtain and may not be available during future epidemics. The Ebola virus disease (EVD) outbreak in West Africa between 2014–16 demonstrated how quickly pathogens can spread to large urban centers following one cross-species transmission event. Here we describe a flexible transmission model to test the utility of generalised human movement models in estimating EVD cases and spatial spread over the course of the outbreak. A transmission model that includes a general model of human mobility significantly improves prediction of EVD's incidence compared to models without this component. Human movement plays an important role not only to ignite the epidemic in locations previously disease free, but over the course of the entire epidemic. We also demonstrate important differences between countries in population mixing and the improved prediction attributable to movement metrics. Given their relative rareness, locally derived mobility data are unlikely to exist in advance of future epidemics or pandemics. Our findings show that transmission patterns derived from general human movement models can improve forecasts of spatio-temporal transmission patterns in places where local mobility data is unavailable.

The geographic spread of infectious pathogens may be driven by infected individuals travelling between areas of active transmission and disease-free areas¹. Whether the disease is transmitted in a location where an infectious person travels depends on the local characteristics such as population density and contact patterns, among others². The dispersal of a pathogen in space and time is limited structurally by the distribution and nature of transport infrastructure³, which in turn are influenced by economic factors^{4,5}. Dispersal can vary seasonally⁶ due to vacations⁷, growing seasons⁸, and religious events⁹. Previously, human mobility patterns have been inferred from a variety of sources, such as census surveys¹⁰, mobile phone data (CDR)¹¹ or other mobile technologies^{12,13}, but such data are often proprietary, expensive and time consuming to collect and process¹⁴. Hence, during an epidemic it is by no means certain that data on human movements in the outbreak location will be available in order to make predictions of disease spread¹⁵. Therefore we aim to test whether general human movement estimates

¹Department of Zoology, University of Oxford, Oxford, UK. ²Harvard Medical School, Boston, MA, USA. ³Computational Epidemiology Lab, Boston Children's Hospital, Boston, MA, USA. ⁴Department of BioSciences, University of Melbourne, Parkville, VIC, Australia. ⁵RTI International, Washington, D.C., USA. ⁶Epidemiology and Public Health Division, School of Medicine, University of Nottingham, Nottingham, UK. ⁷Imperial College London, London, United Kingdom. ⁸Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA, USA. ⁹Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom. ¹⁰Department of Biology, University of Florida, Gainesville, FL, USA. ¹¹Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA. ¹²Sanaria Institute for Global Health and Tropical Medicine, Rockville, USA. ¹³WorldPop, Department of Geography and Environmental Sciences, University of Southampton, Southampton, UK. ¹⁴Flowminder Foundation, Stockholm, Sweden. Correspondence and requests for materials should be addressed to M.U.G.K. (email: moritz.kraemer@zoo.ox.ac.uk) or S.I.H. (email: sihay@uw.edu) or R.C.R. (email: bcreiner@uw.edu)

can provide insightful predictions of disease invasion in resource-poor settings, including areas where mobility data are often unavailable.

The Ebola virus disease (EVD) epidemic in West Africa caused at least 28,000 infections and resulted in more than 11,000 deaths¹⁶. At the height of the outbreak in late 2014 the geographic extent of transmission was the widest ever recorded for Ebola virus, with cases reported in all districts in Sierra Leone (14/14) and Liberia (15/15) as well as in the majority of districts in Guinea (27/34)¹⁶. Phylogenetic analysis suggests that the outbreak caused by the Makona strain was triggered by a single cross-species transmission event from an animal reservoir near Meliandou, Guinea, with the subsequent outbreak sustained exclusively by human-to-human transmission¹⁷. The rapid geographical expansion of the 2014–2016 epidemic stands in stark contrast to previous outbreaks of EVD¹⁸. It has been hypothesised that the complex interplay between increased urbanisation over recent decades, and increased human mobility through porous borders in West Africa, contributed to the catastrophic nature of this outbreak¹⁹. These changes in human behaviour in part led to the spread of EVD that subsequently overwhelmed the countries' poorly equipped health systems and revealed a lack of coordinated rapid response^{20,21}.

No mobility estimates are available to investigate the spread of EVD in West Africa. To our knowledge no transmission dynamical model has been fitted using mobility data from locations outside the region. Detailed investigations of chains of transmission in Guinea have shown that continued unmonitored re-introductions into large urban centres, and subsequent inter-urban transmission events, led to the extensive geographical spread of the virus²². Such information, however incomplete, pose the question if re-occurring introductions have been the driver of the epidemic, a process observed for other diseases²³. The majority of models attempting to predict the regional spread of EVD were limited to a single country^{24,25} and did not assess important characteristics such as the relative contribution of transmission from each district over time^{26–29}. One study using data from Sierra Leone focussed exclusively on the timing of arrival of the disease but did not include any generalised human mobility models³⁰. Other studies attempted to anticipate the risk of international spread of EVD *via* commercial air travel^{31,32}. Furthermore, phylogenetic studies of EVD in Sierra Leone and Liberia indicate that despite inter-country spread during the early phase (December 2013 to mid-March 2014) of the outbreak, most virus transmission occurred locally during the contracting phase of the outbreak and within national borders^{17,33–35}. Some of these changes may be explained by unofficial border closings, curfews, and restrictions on funeral gatherings³⁶. Local studies investigating the transmission pathways of EVD have suggested that most transmissions resulted from close interactions within hospitals and households³⁷. It has been shown that the spread of EVD follows a gravity type model^{38,39}.

Empirical observations suggest that contacts between infected and susceptible individuals could be more frequent in large, densely-populated urban areas than in smaller communities in rural areas but the impact on disease transmission may vary depending on the pathogen^{2,40}. Recent advances in the availability of high resolution data on human mobility³², new formulations of mathematical models to represent disease-related movement patterns⁴¹, and the integration of such models in disease transmission models provide a comprehensive set of tools to enable detailed investigation of the dynamic drivers of EVD transmission. In novel disease outbreak situations the decision as to where to deploy resources is of crucial importance and therefore understanding where the pathogen may spread next is a key question for policy makers. Detailed and location specific data on human movement is rarely available, so understanding the utility of data that can be obtained readily is of great importance.

Here we use openly available data on human mobility from Europe and Senegal and general movement models, together with the spatial configuration of districts and EVD case counts, in order to investigate the disease dynamics of the 2014–2016 West Africa EVD outbreak. We assess the relative contribution to transmission of a range of openly available mobility metrics, and how these contributions changed over the course of the outbreak. Here we assume that transmission in each district is not independent of transmission that occurs in connected districts. Our analytical framework has been implemented in an open-source software pipeline to enable real-time predictive mapping of EVD using publically available data (<https://github.com/SEEG-Oxford/ebola-spread>). Our software can be rapidly updated, applied to other pathogens, and is flexible enough to be tailored to baseline analyses and predictive mapping of future infectious disease outbreaks.

Methods

Overview. The aim of the study was to show whether re-occurring introductions during the outbreak have sustained the epidemic. Further, we tested whether human movement metrics from other regions can be used to predict the dynamics of EVD. We combined generalised human movement models with parameters inferred from open access mobile phone data (from Europe and Senegal) with a flexible transmission model, to test whether the inclusion of mobility fluxes increased the predictive power of EVD cases in West Africa. In addition we included a model that predicts the arrival of EVD in districts previously unaffected.

Epidemiological data. We obtained weekly case data on numbers of EVD cases (both probable and confirmed) from the World Health Organization (WHO) starting January 2014 ending week 32 (August 3rd to 10th) of 2015 after which only sporadic cases occurred. Details about the Situation Reports used to collate these numbers are described elsewhere⁴².

Mobility models. To test the utility of generalised human movement metrics to explain the dynamics of transmission in West Africa we use predicted human movements between each pair of districts under three distinct mathematical models, each reflecting a different aspect of human mobility (Fig. 1A). We did not have access to human movement data (i.e., counts of people moving between districts) from the three affected countries so the human mobility estimates are reflective of general fluxes of the population between districts. The three models are (i) the gravity model $T_{i,j} = k \frac{N_i^\alpha N_j^\beta}{d_{i,j}^\gamma}$, (ii) the radiation model $T_{i,j} = T_i \frac{N N_j}{(N_i + s_{i,j})(N_i + N_j + s_{i,j})}$, and (iii) an adjacency

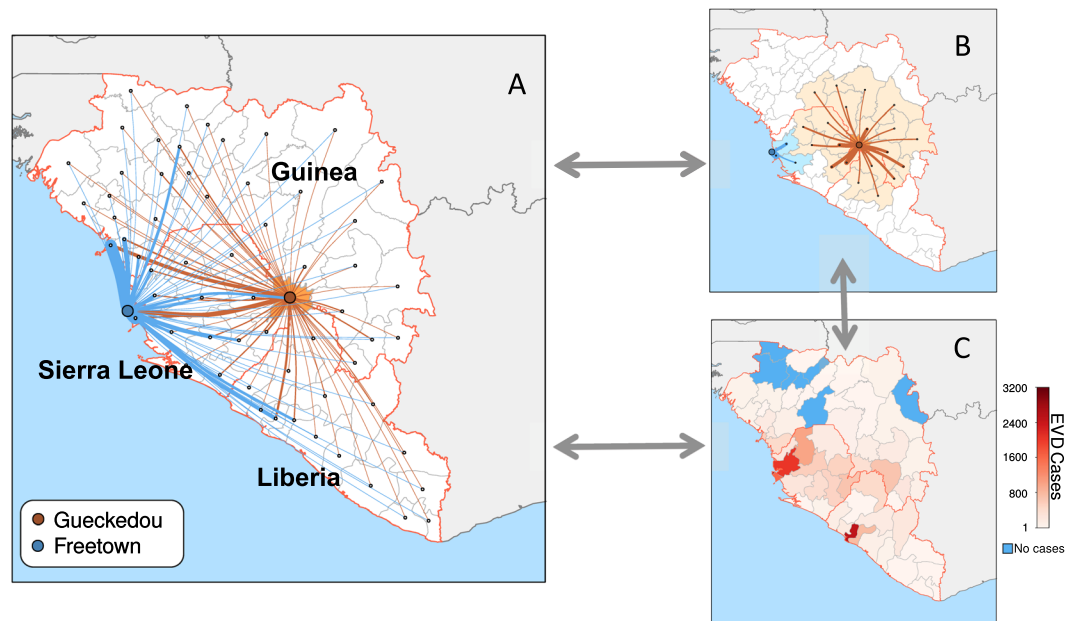


Figure 1. To account for different patterns in movement that might contribute to how the epidemic spread we constructed a comprehensive database that combines different attributes of movement inferred from mobility data in Europe and Senegal which were then predicted to locations in West Africa. **(A)** Shows the connections between Gueckedou, where the outbreak started and all other districts in the region using a gravity model. We further included Freetown to highlight the different strengths of connections that result from the pull of a large city. **(B)** Shows second degree adjacent districts. **(C)** Shows the total distribution of cases as of March 17th, 2016. Blue indicates areas with no cases.

network in which movements are assumed to proceed along edges connecting each of the districts in these countries. In each model T_{ij} represents the total number of individuals moving from district i , to j ; N_i^α is the population size at the origin location; N_j^β is the population size at the destination location; d_{ij}^γ is the distance between them; and s_{ij} is the total population in the radius between i and j . T_i is the total number individuals who make a trip with distance >0 . Parameters k , α , β , and γ are fitted using poisson regression to data in Europe and Senegal. Parameters are shown in Table S7.

The gravity model assumes that relative flow between districts is a log-linear function of the populations of the districts and the distance between them^{41,43}. This model therefore emphasises the attractive power of large population centres. The radiation model also accounts for origin and destination population sizes and distances, but also considers the draw from other populations within the same radius⁴. The radiation model therefore reflects likely patterns of commuting for work, assuming every locality has a competing underlying attractiveness. Adjacency networks encode the number of district borders that an individual would need to cross to move from one district to another. This metric thus reflects the impacts of national and sub-national borders on movements in the region. Each of these models has been shown to be useful depending on the local context to infer regular daily commuting patterns, longer term movements, and general population diffusion processes^{4,44,45}. We used all three metrics, as well as interactions between them to capture possible unexpected effects not described by the metrics alone, as terms in disease spread models^{46–50}. There are a number of other possible movement models such as the Markov model but they often require high-resolution information about the individual trajectories of the users⁵¹.

Mobility data. Both the gravity and radiation models have parameters that determine patterns of movement. These parameters can be deduced by fitting the models to empirical data on human movements. Common data sources for training these models include census commuting or mobile phone data (call detail records)¹⁴. Since no such data were available for the West Africa region, we instead trained models against three high resolution, openly available movement datasets, derived from call detail records, representing movement between districts in France, Spain, Portugal⁴⁸, and one dataset that is currently not openly available from Senegal. We attempt to optimise the parameters of a given movement model based on log likelihoods against the observed data using the optim function and BFGS optimisation method⁵². These movement matrices have been used successfully to predict movements in developing countries⁵³. To show the similarity of these data to high-resolution mobile phone data, we compare our estimates to long-term migration data from the Integrated Public Use Microdata Series (IPUMS, <https://international.ipums.org/international>). These data represent 10% samples of the total population at the individual level from the national censuses conducted in Guinea, Liberia and Sierra Leone in 1996, 2008 and 2004, respectively. Census questions about where respondents lived a year ago were used to quantify migration flows between administrative units in the year before the census. Further details can be found in Wesolowski *et al.*⁴³ and Sorichetta *et al.*⁵⁴. These data have been shown previously to be a good representation of short term movements even though the sample is comparatively small¹⁰. A full list of correlation coefficients is provided in

the Supplementary information. In addition we fitted our movement models to data from a neighbouring country (Senegal) and tested if such data improved our results¹¹.

Mobility metrics. For each district in West Africa ($n = 63$), we determined the total human population size using gridded population estimates and we calculated the distance between the centroids of each pair of districts⁵⁵. Gravity and radiation model parameters were fitted to the empirical data described above and applied to the three core affected countries using the movement R package⁵⁶. National adjacency networks were computed using administrative boundary data from the GADM dataset (<http://www.gadm.org>). This adjacency matrix was then disaggregated into three binary mobility matrices with mobility degrees of one (*i.e.*, districts share a border), two (*i.e.*, districts share a common neighbour), three, and fully connected.

Covariate database. EVD transmission in the core-affected countries is likely to be influenced by human mobility at a variety of spatial and temporal scales, with different aspects of movement varying in importance through the course of the epidemic and among countries. We built a large set of candidate covariates (hereafter used to refer to mobility matrices) to capture spatial interactions⁵⁷. To accomplish this, we considered interactions between the adjacency movement model and both the radiation model and the gravity model. We then applied backwards selection using the “step” function in R⁵⁸ to select a sparse optimal set of features that minimize the generalisation error of the model. Whilst model generalizability and interpretation are generally aided by using a parsimonious set of covariates, this approach reflected our main aim here, which was to maximise the predictive accuracy of the model.

Each of the resulting three fitted mobility models was then used to predict human mobility between all districts in Guinea, Sierra Leone, and Liberia and for each country separately (Fig. 1A). Each of the metrics was then weighted by different sets of adjacency (Fig. 1B) and iterated through each two-week period, depending on the cases in each district (Fig. 1C).

Disease model specifications. To model the effect of human mobility on the geographic spread and rate of transmission of EVD, within and between the three core countries, we used a two-stage model to characterise both geographic expansion (*i.e.* introduction into previously unaffected districts) and the expected secondary cases arising from these introductions.

Invasion model. The invasion model estimates the probability $p_i(t)$ that one or more cases will be identified in previously disease-free district i at time t (with presence or absence of new cases indicated by $Y_i(t)$), as a function of the number of cases $C_{-i}(t-1)$ in all other districts in the previous time point which is chosen to be every two weeks in our analysis, the product of corresponding values of the mobility covariates $x_{j,-i}$ for each covariate j , regression coefficient b_j and a fixed intercept term c . This gives the following logistic regression model:

$$\begin{aligned} Y_i(t) &\sim \text{Bernoulli}(p_i(t)) \\ \log \text{it}(p_i(t)) &= c + \sum_j \sum_{-i} b_j x_{j,-i} C_{-i}(t-1) \end{aligned} \quad (1)$$

Transmission model. For all districts reporting one or more cases, we assume a general bi-weekly transmission model following⁵⁹:

$$I_{t,i} = \beta_{t,i} * \frac{I_{t-1,i}^{\alpha_i}}{N_i} * S_{t-1,i} * \epsilon_{t,i} \quad (2)$$

where $I_{t,i}$ is the number of infected and infectious individuals and $S_{t-1,i}$ the number of susceptible individuals, at time t in district i , N_i is the population of district i and $\beta_{t,i}$ is the covariate-driven mobility rate characterised by a linear combination of the mobility metrics described above (*i.e.*, $\beta_{t,i}$ has the same model structure as the final line in equation (1), although parameters are fitted independently). α_i is a parameter to account for the discretization of a continuous process and can be seen as an approximation of the contact rate of the population in district i which varies between places². The error terms $\epsilon_{t,i}$ are independent, identically log-normally distributed random variables with $N(0, \sigma^2)$. We further assumed all individuals to be initially susceptible to infection. Before including a set of mobility matrices inferred from European and Senegalese movement data we fitted a covariate-free model that assumes no interaction between districts (non-movement).

Calculation of source and sink districts. To pinpoint which district had the highest contribution to transmission in West Africa at each time-step (t , bi-weekly) we calculate the relative weights of each district by converting equation (2) into a linear regression of the form:

$$y_{i,t} = \log \beta_i + \alpha * x_i + \log \epsilon_i \quad (3)$$

with

$$\log \beta_{i,t} = F_1 * X_{1j} + F_2 * X_{2j} + \dots + F_k * X_{kj} \quad (4)$$

and

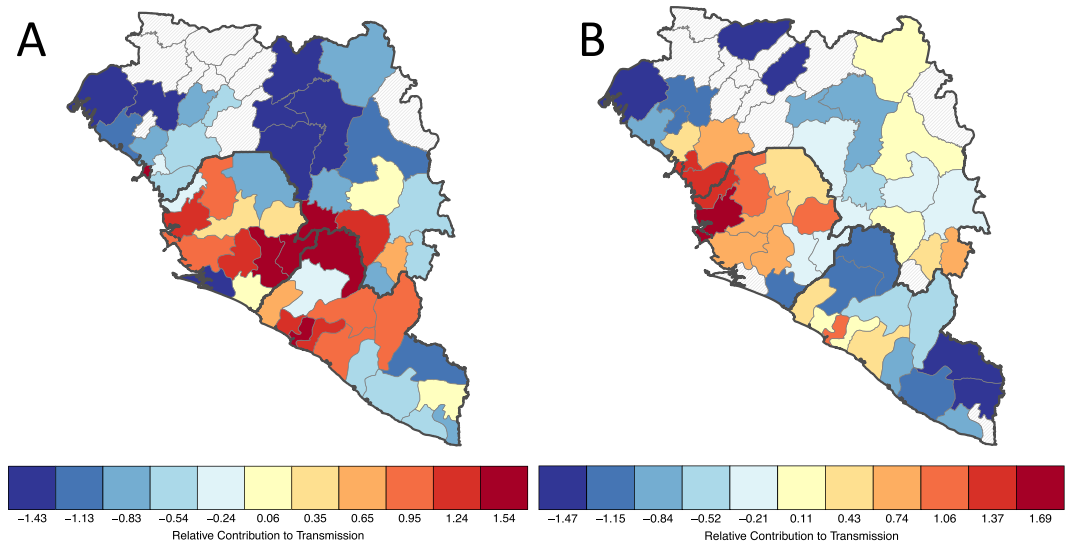


Figure 2. Relative contribution to transmission in the expanding phase of the outbreak in West Africa (week 1–42, panel A) and the second half of the outbreak (42–83, panel B). Red shows sources of transmission measured how much they contributed to transmission elsewhere. Blue shows districts that are contributing less to the spread of EVD.

$$y_{i,t} = \frac{I_{t,i} * N_i}{S_{t-1,i}}; x_i = \log(I_{t-1,i})$$

The F_i are fitted through the regression. For any district i , each X_{ij} is one of k district-specific covariates that combines how many cases there are in all the other districts weighted by a mobility matrix. These district specific covariates are re-calculated at each time step. We can rearrange equation (4) so that $\log \beta_i$ for district i ($i = 1, 2, 3, \dots, 63$) is a function of the number of cases in every other district (x_j):

$$\log \beta_i = \sum_{\substack{j=1 \\ j \neq i}}^{63} \gamma_{ij} x_j \quad (5)$$

Full details of how the relative contributions for each district are derived can be found in the supplementary information (Fig. 2).

Invasion model evaluation. The invasion models estimate the probability of invasion for each district that has not already been invaded (in previous time-step), using the fitted model up to that point. To assess the predictive performance of the invasion models under a realistic real-time scenario, we re-fitted the model for each week of the epidemic, and in each instance we used only the information available from preceding weeks, up to and including the week in question. That week's model was then used to predict the probability of invasion in the following week, two weeks, or in the following month, across all districts currently disease-free. These predictions were then compared with the observed invasions in these districts in the following week. Rather than determining a fixed threshold probability value with which to evaluate the model's predictive power, we computed a receiver operating characteristic (ROC) curve to evaluate predictive power under all possible threshold values. The area under curve (AUC) represents the probability that a district drawn uniformly at random in which an invasion does occur would receive a higher “probability of invasion” than a randomly drawn district in which no invasion occurs. The closer the AUC value is to 1, the better the predictive power of the model.

Transmission model selection. In general $\beta_{i,t}$ terms were fitted entering the covariates linearly, following previous work^{58,60,61}. All model fitting was conducted in R and model selection was conducted using the Akaike Information Criterion (AIC) and Likelihood Ratio Tests (LRT); with AIC used to identify reduced models and LRT to compare final models with simpler nested models³⁰. To arrive at country-specific effects, the transmission model selection was conducted using two approaches. First, a single model including all covariates was fitted to data from all countries and then reduced using backwards selection and AIC following standard procedures (‘step’ function in R). Secondly, the resulting model from step 1 (including only covariates selected) was fitted independently for each country again and reduced using backwards selection and AIC. In general multiple movement matrices were selected, all contributing and moderating the force of infection within and between districts. We evaluated the performance of our model by comparing predicted vs. observed case numbers two weeks ahead (out of sample).

Phases of the epidemic. Both the geographic extent of the outbreak, and the weekly numbers of reported cases reached a peak and then declined as the outbreak was brought under control. The extent of the epidemic

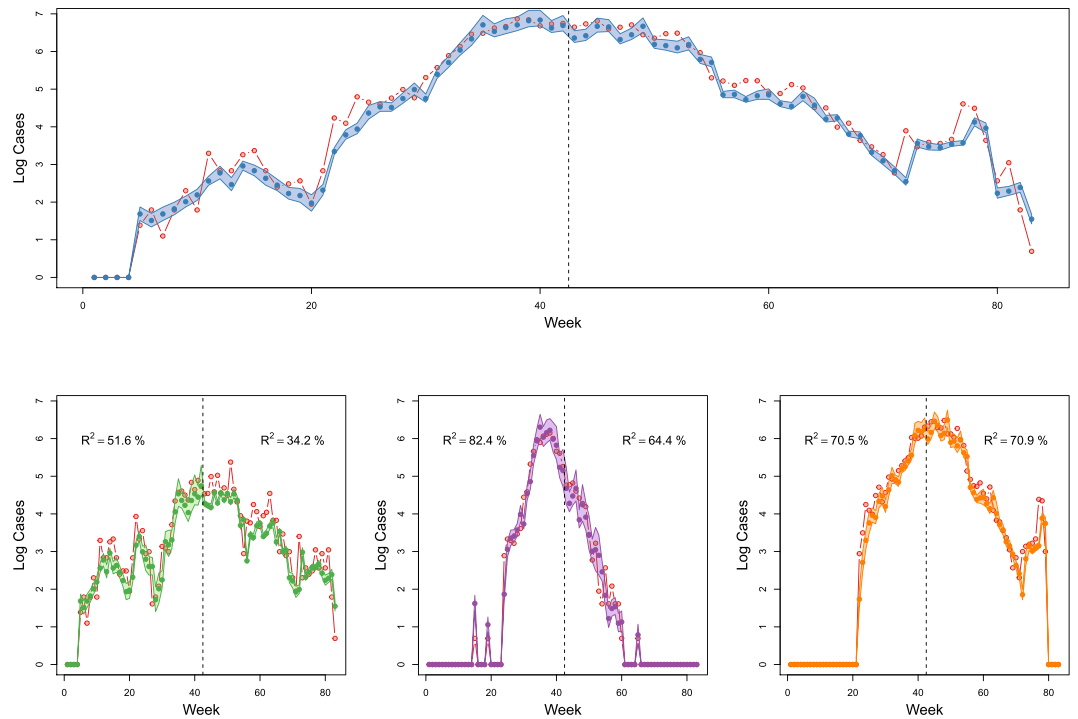


Figure 3. Observed (probable and confirmed) vs. two week ahead predicted transmission in all three core countries (top panel), in Guinea, Liberia, and Sierra Leone for both the expanding and the contracting phase of the epidemic, from left to right respectively. Red lines representing observed cases. 95% CI intervals are given for the predicted cases.

was greatest in week 42 (October 13th to 19th) of 2014, four weeks after the peak in weekly numbers of cases (week 38, September 15th to 21st). In order to account for the impact of human mobility processes contributing to transmission during the epidemic's expansion and contraction phases, we split the data in two and separately analysed the first half of the epidemic (expanding phase) until week 42 of 2014 and the second half of the epidemic (contracting phase) from week 43 to 84.

Model comparisons. To assess the relative performance of a model that is using European cell phone data to fit movement metric parameters, we re-performed the analysis described with a model that utilizes mobile phone data from a neighbouring country, in this case Senegal^{11,62}. We note that such data is often unavailable during outbreaks. This analysis included re-assessing the optimal country-specific mixing coefficient, and fitting country- and district-based transmission models. Model performance was evaluated using AIC, as well as R^2 .

Results

Our analyses found that generalised human movement models derived from data outside the affected geographies can explain a considerable proportion of observed dynamics of the EVD outbreak. For the first half of the outbreak, the covariate-free (no interaction between districts) model fits well to the data (R^2 -adjusted = 0.64, out of sample two week ahead prediction) with a relatively high mixing coefficient of $\alpha = 0.81$ (see² for more discussion of mixing coefficients within TSIR models). Adding a set of human mobility matrices derived from European mobile phone data and applying backward model selection using AIC significantly improved two week ahead predictions ($p < 0.001$, LRT statistic $\chi^2_9 = 38.46$) with moderate change in $\alpha = 0.67$ and minor improvement in R^2 -adjusted (0.67, out of sample two week ahead prediction). The retained covariates are listed in Table S3 and predictions with 95% confidence intervals are shown in Fig. 3.

Further, we showed that there were temporally varying sources and sinks of virus transmission, as well as significant spatio-temporal variation in future risk of invasion to previously unaffected districts during the EVD outbreak (Fig. 2, Video S1). Investigating the relative contribution of different regions to ongoing transmission shows that the geographic focus of the epidemic shifted. During the expanding phase of the epidemic (weeks 1–42) the districts predominantly contributing to transmission were located around the origin of the outbreak in Meliandou, Guinea. Later, transmission shifted to the highly populated corridor along the coast between Conakry, Guinea and Freetown, Sierra Leone (Fig. 3). This indicates clearly that the districts with transmission cannot be seen in isolation over the entire course of the outbreak. A dynamic map showing trends across the whole outbreak is shown in Video S1 and Fig. S6. Our findings also support the hypothesis that the rapid progression of the outbreak in West Africa was preceded by the initial introduction into Kailahun, Sierra Leone and Lofa, Liberia, both adjacent to Gueckedou but under different jurisdictions⁴². Moreover, we show that during the entire outbreak, only Nimba county, Liberia, was a main contributor to transmission that was located along the border to an unaffected country (in this case Côte d'Ivoire) (Fig. 2, Table S1). This result helps to explain why

All countries		Expanding Phase (Week 1–42)	
		Covariate free	With human mobility
	R ²	0.6435	0.668
Guinea		Full model	Country specific model
	R ²	0.385	0.47
	AIC	370.0	357.583
Liberia			
	R ²	0.76	0.81
	AIC	304.8	288.15
Sierra Leone			
	R ²	0.63	0.68
	AIC	350.1	339.3

Table 1. Summary of modelling results (adjusted R² and Akaike Information Criterion) for the covariate free model for all countries and with human mobility. In addition, the country specific results for the full model and country covariates are shown.

transmission was not observed in any of the bordering countries; many of the border districts were sinks of EVD transmission rather than exporters (Video S1, Table S1).

Country specific results. We then tested whether the impact of mobility on transmission varies between countries. This analysis was also undertaken to understand if the outbreak was driven by local rather than between-country movements. For Guinea, the covariate free model fitted poorly (R²-adjusted = 0.384). When adding mobility metrics improved prediction of case numbers in Guinea by 22% (R²-adjusted = 0.42, Table S3, AIC 357.583 and 368.242 respectively, Table 1).

In Liberia, the non-movement covariate-free model fitted well (R²-adjusted = 0.76) and the Liberia-only reduced model that includes mobility matrices improved the fit significantly (R²-adjusted = 0.79, Table S4). Unlike R²-adjusted, AIC showed more separation from the three models with the Liberia-only reduced model (*i.e.*, including human mobility) being strongly preferred (288.15 versus 304.8 for the covariate-free model, Table 1). This indicates a strongly locally driven epidemic structure, in which most of the variation in transmission can be explained by local patterns of human mobility.

In Sierra Leone, the non-movement covariate-free model fitted moderately well (R²-adjusted = 0.63) and again, country specific Sierra Leone-mobility model improved the fit (R²-adjusted = 0.65, Table S5). As with the Liberia models, AIC showed separation for the Sierra Leone-only reduced model being strongly preferred (AIC = 339.3 versus 350.2 for the base model, Table 1).

To further improve our model and account for country-specific differences underlying within-location human behaviour, we allowed the mixing coefficient to be different for each country. This resulted in a relatively low mixing coefficient of 0.54 for Guinea and similar values for Liberia and Sierra Leone (0.70). This more flexible model improves the model fit (medium R² per location: 0.48, 0.81, and 0.67, for Guinea, Liberia and Sierra Leone respectively, Figs S1–S3) compared to a model that assumes mixing to be the same in each country (medium R² per location: 0.42, 0.79, and 0.65).

Sensitivity analyses. To test whether our approach could be utilised in the context of an outbreak in real time, with openly available mobility data, we compared our estimates of numbers of cases to a model that uses mobility metrics derived from a set of mobility data that is not available to the public. Using radiation and gravity model parameters derived from Senegal¹¹ we re-performed the entire transmission model fitting exercise using only this locally-defined mobility matrix. In every circumstance, the model fits and resulting conclusions were very similar. Again, when model selection was allowed to operate independently on the three country-level models, there was a large difference in the resulting R-squared values (0.47, 0.80, 0.65 respectively for Guinea, Liberia and Sierra Leone, Figs S8–S10). For this model, the mixing coefficients were similar to those from the original analysis (0.57, 0.70 and 0.73 respectively for Guinea, Liberia, and Sierra Leone). The “best fit” model when using only locally-defined mobility performed slightly worse than the corresponding “best fit” model from the original analysis (1011.1 versus 1010.7). For the second half of the outbreak, all qualitative conclusions remained identical, as in the original analysis. Visually, the predictive map of the outbreak is also qualitatively indistinguishable from the original analysis (Fig. S11). In all cases, the R² is slightly lower than in the original analysis (Figs S8–S10).

Invasion process. We investigated whether our model, in the absence of local mobility data, has the ability to predict where the disease is most likely to be observed. We find that there was considerable heterogeneity in the invasion process, so models differ from week to week. However, several covariates were retained in the majority of the models. Of the 28 models (from week 10 to week 37 where most of the invasions occurred), the total number of cases in each district was retained in almost all of the final models (26/28). Two of the four covariates that were retained in 25 of the 28 models were not location dependent: (i) the total number of cases in Guinea each week, and (ii) the weighted sum of all cases using the gravity model in West Africa. The other two covariates that were retained in 25 of the 28 models were different for each location: (i) the gravity model weighted sum of all cases that were both in Guinea and direct neighbours (if there were any) and (ii) the radiation model weighted sum of

all cases that were direct neighbours. This confirms that gravity and radiation models, both capturing a different set of patterns of human mobility, are important in understanding the invasion process.

As with the transmission models, the predictive ability of the invasion model varied from country to country (Fig. S4). In general predictions ‘two week ahead’ were difficult to estimate due to the time lag in reporting and underlying uncertainties inherent in reporting of the disease. When the time-steps were increased from one week to one month, however, the invasion model made accurate predictions of the invasion process evaluated as the timing of predicted invasion ($AUC = 0.697$, Fig. S5). This trend was visible for almost the entirety of the outbreak (Fig. S4).

Discussion

Generalized movement models have the ability to improve current models of spread of EVD in West Africa, which is particularly crucial in low-resource settings with limited local mobility data. We further show that mobility plays an important role not only in the process of disease spread (ignition of transmission in new location) but is particularly important over the entire course of the outbreak sustaining transmission. This result has broad implications for modelling of future outbreaks when rapid assessment of spread is of utmost importance and when location-specific mobility data is not immediately available. Our study confirms previous mechanistic models in which the early epidemic spread in Liberia was explained using assumptions about human mobility from an agent-based model²⁵. Our results also corroborate some trends reported in previously studies that analysed genetic data, however we were unable to validate our results with data that specifically disentangles imported vs. local cases. Inspired by genetic data analysis that allow to understand the lineage movements between locations, which is documented to be a big factor in virus spread during the EVD epidemic, we specifically identify that during an outbreak lineage movement occurs over the entire course of the outbreak, not just to initiate transmission. Such results have been documented qualitatively and it has been hypothesised that re-introductions with relatively small clusters sizes were the main driver of spread of the virus. We further show that the transmission sources and sinks change over the course of the outbreak, similar to those documented by genetic data, providing a rationale for use of this model in future outbreak analyses. This stands in contrast to conventional models that assume that once transmission is ignited in a district it is sustained there independently of transmission elsewhere⁶³.

We show that our model can predict the bi-weekly geographic spread of EVD into districts in the three main affected countries with relatively high accuracy using data up until the forecasting period and evaluating the predictions out of our sample ($AUC = 0.697$, Fig. S5). Further, our model is sensitive to change in human population sizes so it can be adjusted as populations grow. In all three countries the centrality in the mobility network of large population hubs (Conakry, Freetown and Monrovia) was identified as facilitating rapid spread (Fig. 2B).

For Liberia and Sierra Leone, intra-country dynamics seem to be more important drivers of transmission dynamics, whereas for Guinea estimates of EVD cases per district were only slightly improved when within-country movements were considered. These findings extend previous work that examined Liberia and Sierra Leone in isolation²⁴. We find that both gravity and radiation human mobility models should be used in conjunction, as there is considerable heterogeneity from week-to-week during the epidemic, which may be explained better by the different types of human movement that the two models capture (e.g. commuting vs. longer distance movements). Interestingly gravity models fitted to different countries have invasive and protective effects when used together in the transmission model. We anticipate the one covariate absorbs the invasive effect and the residual unexplained variation is absorbed by a correlated predictor. This points towards the heterogeneous and complex nature of disease spread and the need for multivariate statistical approaches previously shown to improve predictive accuracy at the expense of interpretability¹. As our models are automated in a near-real time framework, they can be rapidly updated and used to inform public health prevention and response decisions in future outbreak situations¹⁴.

During the course of a catastrophic outbreak, individual's and population everyday behaviours change due to public awareness campaigns, government restrictions, or illness⁶⁴. In this study we identified the relative importance of distinct human mobility measures that changed during the outbreak (Table 1). In the first half of the outbreak, the country-specific mobility metrics were important in governing how the virus initially spread, even in the absence of real-time human movement data. In the second half of the epidemic, adding human movement data did not significantly improve the predictions. Comparing predicted movements fitted to data in Senegal did not improve our predictions. We anticipate that there are fundamental rules of human mobility across geographic settings that can capture the spread process of infectious diseases, even in the absence of direct measures of human mobility. Once the disease had spread into almost all corners of the affected countries the country specific covariates were less important. Instead, overall movement dynamics within and between countries were responsible for improving the models' performance, and differences between countries were less important. This indicates that our dynamic model, in the absence of real-time movement data, was still able to capture the dynamic invasion process and changes in behaviour that may have contributed to the decline in case numbers^{65,66}. Much of this may also be explained by the full deployment of containment activities such as safe burials that prevented onward infection^{21,36}. Given the relatively large and complex model structure we refrained from including results for all possible combinations of mobility metrics.

Limitations. Knowing the actual number of people moving between locations, as opposed to relative flows used in our study may be helpful to gain better insight in the number of importations leading to novel chains of transmission that are spatially distinct. However, we show that our model covariates (i.e. mobility metrics from European mobile phone data) can confidently predict the spread of the pathogen and its dynamics over time, equally well using data from a neighbouring country. Interestingly data from a neighbouring country (Senegal) did not improve predictions of EVD cases in the core affected countries, indicating that there is some

fundamental rules about mobility and its effect on disease spread although economic and social dynamics are very different between European countries and West Africa. However, how different they are has never been quantified. Another limitation of our work is that it does not take into account the effect of interventions that may have had a significant effect of bringing the epidemic to slow down. To fully understand the community- and household-based dynamics that are typical for EVD, information about the specific contact patterns would be of great use to understand specifically the risk of spread after introduction of the virus from another district². Such insights are particularly useful for medical practitioners and public health officials to design appropriate countermeasures. However, our model allowed the flexible integration of spatial differences in mixing, which was most apparent when allowing mixing coefficients to be different between countries. The differences in mixing may explain why EVD was sustained longer in areas with higher mixing coefficients in Sierra Leone and Liberia. Very little is known about the spatio-temporal heterogeneity of reporting during the outbreak which may improve further iterations of our model.

Conclusion

Our results match those of country-specific phylogenetic studies that concluded that virus spread in Sierra Leone primarily occurred within national borders, because adding country-specific covariates for Sierra Leone in our model significantly improved model fit³³. Genomic surveillance, including the use of real-time portable genome sequencing⁶⁷, may be used to extend our modelling approach by helping to identify the origins of an outbreak¹⁷, monitoring the diversity of circulating viruses^{34,35}, characterizing signatures of host adaptation³³, and pinpointing the source and sink locations of circulating strains^{33,34,68–70}. For now, genetic analyses are often limited by sample size, comparatively data release during an outbreak, and heterogeneous spatial coverage⁶³. At this stage, our analyses do not address the EVD outbreak from a genetic perspective, but our results provide a baseline to which genetic results may be directly compared and used. Future work in epidemic prediction would benefit from the joint incorporation of epidemiological, spatial and genetic data^{14,71–73}. It is unfortunate that contemporary mobile phone data for Guinea, Sierra Leone, and Liberia are still unavailable and human movement data from other countries is not openly accessible. Such availability could help analyse the probable nature of pathogen flow prior to an outbreak and thus improve surveillance and containment preparedness plans^{74,75}.

The identification of transmission sources and sinks has broad application in disease control. They can identify where treatment and prevention measures would be best implemented to prevent the rapid geographic spread of a pathogen. This has been shown for other diseases using historical data, but the modelling techniques presented here allow for the application of near real-time data for the control of an ongoing outbreak. Such methodologies have the potential to be used by national and international public health institutions to plan and perform effective control and surveillance systems, with the aim of limiting the geographic extent and burden of future outbreaks in areas with high potential emergence of contagious viral haemorrhagic fevers, as well as other directly transmissible infectious diseases⁷⁶.

Data Availability

Epidemiological data are available from the World Health Organization and mobility data from previous publications cited in the manuscript. Code will be made available after publication.

References

- Kraemer, M. U. G. *et al.* Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015–16: a modelling study. *Lancet Infect. Dis.* **17**, 330–338 (2017).
- Kraemer, M. U. G. *et al.* Big city, small world: density, contact rates, and transmission of dengue across Pakistan. *J. R. Soc. Interface* **12**, 20150468 (2015).
- Uchida, H. & Nelson, A. *Agglomeration index: towards a new measure of urban concentration* (2008).
- Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* **484**, 96–100 (2012).
- Weiss, D. J. *et al.* A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* **553**, 333–336 (2018).
- Wesolowski, A. *et al.* Multinational patterns of seasonal asymmetry in human movement influence infectious disease dynamics. *Nat. Commun.* **8** (2017).
- Deville, P. *et al.* Dynamic population mapping using mobile phone data. *Proc. Natl. Acad. Sci. USA* **111**, 15888–93 (2014).
- Bharti, N. *et al.* Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science* **334**, 1424–7 (2011).
- Lessler, J. *et al.* Estimating potential incidence of MERS CoV associated with Hajj pilgrims to Saudi Arabia, 2014. *PLOS Curr. Outbreaks* November 2 (2014).
- Ruktanonchai, N. W. *et al.* Census-derived migration data as a tool for informing malaria elimination policy. *Malar. J.* **15**, 273 (2016).
- Finger, F. *et al.* Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks. *Proc. Natl. Acad. Sci. USA* **113**, 6421–6 (2016).
- Kraemer, M. U. G. *et al.* Inferences about spatiotemporal variation in dengue virus transmission are sensitive to assumptions about intra-urban human mobility. *EPJ Data Sci.* **7** (2018).
- Vazquez-Prokopec, G. M. *et al.* Using GPS technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment. *PLoS One* **8**, e58802 (2013).
- Kraemer, M. U. G. *et al.* Progress and challenges in infectious disease cartography. *Trends Parasitol.* **32**, 19–29 (2016).
- Halloran, M. E. *et al.* Ebola: Mobility data. *Science* **346**, 9–10 (2014).
- World Health Organization. Ebola Situation Report. 16 March 2016 (2016).
- Gire, S. K. *et al.* Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345**, 1369–1372 (2014).
- Mylne, A. *et al.* A comprehensive database of the geographic spread of past human Ebola outbreaks. *Sci. Data* **1**, 140042 (2014).
- Farrar, J. J. & Piot, P. The Ebola emergency — immediate action, ongoing strategy. *N. Engl. J. Med.* **371**, 1545–6 (2014).
- Gates, B. The next epidemic - lessons from Ebola. *N. Engl. J. Med.* **372**, 1381–1384 (2015).
- Kucharski, A. J. *et al.* Measuring the impact of Ebola control measures in Sierra Leone. *Proc. Natl. Acad. Sci. USA* **112**, 14366–71 (2015).

22. Faye, O. *et al.* Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study. *Lancet Infect. Dis.* **15**, 320–326 (2015).
23. Grubaugh, N. D. *et al.* Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature* **546**, 401–405 (2017).
24. Valdez, L. D., Aragão Rêgo, H. H., Stanley, H. E. & Braunstein, L. A. Predicting the extinction of Ebola spreading in Liberia due to mitigation strategies. *Sci. Rep.* **5**, 12172 (2015).
25. Merler, S. *et al.* 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**, 204–211 (2015).
26. Chretien, J., Riley, S. & George, D. B. Mathematical modeling of the West Africa Ebola epidemic. *Elife* **4**, e09186 (2016).
27. Backer, J. A. & Wallinga, J. Spatiotemporal Analysis of the 2014 Ebola Epidemic in West Africa. *PLoS Comput. Biol.* **12**, e1005210 (2016).
28. Silva, J. P. D. & Eisenberg, M. C. Modeling spatial invasion of Ebola in West Africa. *J. Theor. Biol.* **428**, 65–75 (2017).
29. Ajelli, M. *et al.* Spatiotemporal dynamics of the Ebola epidemic in Guinea and implications for vaccination and disease elimination: a computational modeling analysis. *BMC Med.* **14** (2016).
30. Yang, W. *et al.* Transmission network of the 2014–2015 Ebola epidemic in Sierra Leone. *J. R. Soc. Interface* **12**, 1–9 (2015).
31. Bogoch, I. I. *et al.* Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 west African outbreak. *Lancet* **6736**, 1–7 (2014).
32. Poletti, M. *et al.* Assessing the impact of travel restrictions on international spread of the 2014 West African Ebola epidemic. *Euro Surveill.* **19**, 20936 (2014).
33. Park, D. J. *et al.* Ebola Virus Epidemiology, Transmission, and Evolution during Seven Months in Sierra Leone. *Cell* **161**, 1516–1526 (2015).
34. Ladner, J. T. *et al.* Evolution and Spread of Ebola Virus in Liberia, 2014–2015. *Cell Host Microbe* **18**, 659–669 (2015).
35. Carroll, M. W. *et al.* Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. *Nature* **524**, 97–101 (2015).
36. Fang, L. *et al.* Transmission dynamics of Ebola virus disease and intervention effectiveness in Sierra Leone. *Proc. Natl. Acad. Sci. USA* PNAS Early Edition, <https://doi.org/10.1073/pnas.1518587113> (2016).
37. Lewnard, J. A. *et al.* Dynamics and control of Ebola virus transmission in Montserrado, Liberia: a mathematical modelling analysis. *Lancet Infect. Dis.* **14**, 1189–1195 (2014).
38. Kramer, A. M. *et al.* Spatial spread of the West Africa Ebola epidemic. *R. Soc. Open Sci.* **3**, 160294 (2016).
39. Dudas, G. *et al.* Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature* **544**, 309–315 (2017).
40. Read, J. M. *et al.* China social mixing patterns in rural and urban areas of southern China. *Proc. R. Soc. B* **281**, 20140268 (2014).
41. Wesolowski, A., O'Meara, W. P., Eagle, N., Tatem, A. J. & Buckee, C. O. Evaluating Spatial Interaction Models for Regional Mobility in Sub-Saharan Africa. *PLoS Comput. Biol.* **11**, e1004267 (2015).
42. WHO Ebola Response Team. Ebola Virus Disease in West Africa - The First 9 Months of the Epidemic and Forward Projections. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMoa1411100> (2014).
43. Wesolowski, A. *et al.* Commentary: Containing the Ebola outbreak – the potential and challenge of mobile network data. *PLOS Curr. Outbreaks* Sep **29** (2014).
44. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–42 (2013).
45. Jongejans, E. *et al.* A unifying gravity framework for dispersal. *Theor. Ecol.* **8**, 207–223 (2015).
46. Balcan, D., Colizza, V., Gonc, B. & Hu, H. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. USA* **106**, 21484–89 (2009).
47. Wesolowski, A. *et al.* Quantifying the impact of human mobility on malaria. *Science* **338**, 267–70 (2012).
48. Tizzoni, M. *et al.* On the use of human mobility proxies for modeling epidemics. *PLoS Comput. Biol.* **10**, e1003716 (2014).
49. Simini, F., Maritan, A. & Nédá, Z. Human mobility in a continuum approach. *PLoS One* **8**, e60069 (2013).
50. Reiner, R. C. *et al.* Highly localized sensitivity to climate forcing drives endemic cholera in a megacity. *Proc. Natl. Acad. Sci. USA* **109**, 2033–6 (2012).
51. Perkins, T. A. *et al.* Theory and data for simulating fine-scale human movement in an urban environment. *J. R. Soc. Interface* **11**, 20140642 (2014).
52. Nocedal, J. & Wright, S. J. *Numerical Optimization*, <https://doi.org/10.5194/bgd-11-12733-2014> (Springer, 1999).
53. Nunes, M. R. T. *et al.* Emergence and potential for spread of Chikungunya virus in Brazil. *BMC Med.* **13**, 102 (2015).
54. Soricchetta, A. *et al.* Mapping internal connectivity through human migration in malaria endemic countries. *Sci. Data* **3**, 160066 (2016).
55. WorldPop project. WorldPop. Available at: <http://worldpop.org.uk/>.
56. Golding, N., Schofield, A. & Kraemer, M. U. G. Movement: Functions for the analysis of movement data in disease modelling and mapping. *R Packag. version 0.2* (2015).
57. Blumenstock, J., Cadamuro, G. & On, R. Predicting poverty and wealth from mobile phone metadata. *Science* **350**, 1073–76 (2015).
58. Hastie, T. J. & Tibshirani, R. J. *Generalized additive models*. (CRC Press, 1990).
59. Finkenstädt, B. F. & Grenfell, B. T. Time series modelling of childhood diseases: a dynamical systems approach. *Appl. Stat.* **49**, 187–205 (2000).
60. Dominici, F., McDermott, A., Zeger, S. L. & Samet, J. M. On the use of generalized additive models in time-series studies of air pollution and health. *Am. J. Epidemiol.* **156**, 193–203 (2002).
61. Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73**, 3–36 (2011).
62. de Montjoye, Y.-A., Smoreda, Z., Trinquart, R., Ziemlicki, C. & Blondel, V. D. D4D-Senegal: The Second Mobile Phone Data for Development Challenge. *arXiv arXiv: 1407.4885v2* (2014).
63. Kraemer, M. U. G. *et al.* Reconstruction and prediction of viral disease epidemics. *Epidemiol. Infect.* 1–7, <https://doi.org/10.1017/S0950268818002881> (2018).
64. Fenichel, E. P. *et al.* Adaptive human behavior in epidemiological models. *Proc. Natl. Acad. Sci. USA* **108**, 6306–11 (2011).
65. Rivers, C. Ebola: the power of behaviour change. *Nature* **515**, 492–492 (2014).
66. Funk, S., Gilad, E., Watkins, C. & Jansen, V. A. The spread of awareness and its impact on epidemic outbreaks. *Proc. Natl. Acad. Sci. USA* **106**, 6872–7 (2009).
67. Rambaut, A. Phylogeography of 2014–2015 Ebola virus epidemic. *virological.org* Available at, <http://virological.org/t/phylogeography-of-2014-2015-ebola-virus-epidemic/199>, (Accessed: 15th February 2016).
68. du Plessis, L. & Stadler, T. Getting to the root of epidemic spread with phylogenetic analysis of genomic data. *Trends Microbiol.* **23**, 383–386 (2015).
69. Gardy, J., Loman, N. J. & Rambaut, A. Real-time digital pathogen surveillance — the time is now. *Genome Biol.* **16**, 155 (2015).
70. Faria, N. R. *et al.* Zika virus in the Americas: Early epidemiological and genetic findings. *Science* **352**, 345–349 (2016).
71. Grenfell, B. T. *et al.* Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–32 (2004).
72. Pybus, O. G., Tatem, A. J. & Lemey, P. Virus evolution and transmission in an ever more connected world. *Proc. R. Soc. B Biol. Sci.* **282**, 20142878 (2015).
73. Currie, J., Grenfell, B. & Farrar, J. Beyond Ebola. *Science* **351**, 815–16 (2016).

74. Pigott, D. M. *et al.* Mapping the zoonotic niche of Marburg virus disease in Africa. *Trans. R. Soc. Trop. Med. Hyg.* **109**, 366–378 (2015).
75. Mylne, A. Q. N. *et al.* Mapping the zoonotic niche of Lassa fever in Africa. *Trans. R. Soc. Trop. Med. Hyg.* **109**, 483–492 (2015).
76. Pigott, D. M. *et al.* Mapping the zoonotic niche of Ebola virus disease in Africa. *Elife* **3**, e04395 (2014).

Acknowledgements

MUGK is funded by the Research for Health in Humanitarian Crises (R2HC) programme managed by ELRHA (to SIH). The R2HC programme is funded equally by the Wellcome Trust and the Department of International Development (DFID). MUGK is also supported by The Branco Weiss Fellowship - Society in Science, administered by the ETH Zurich and acknowledges funding from a Training Grant from the National Institute of Child Health and Human Development (T32HD040128) and the National Library of Medicine of the National Institutes of Health (R01LM010812, R01LM011965). NG, DB and DMP received funding from the Bill & Melinda Gates Foundation (#OPP10937011, #OPP1068048). NG was supported by a Discovery Early Career Researcher Award from the Australian Research Council (DE180100635). DLS and AJT are funded by the NIH/National Institute of Allergy and Infectious Diseases (#U10AI089674), and the Bill & Melinda Gates Foundation (AJT #OPP1106427, #1032350, DLS #OPP1110495). AJT is also supported by a Wellcome Trust Sustaining Health Grant (#10688/Z/15/Z). SIH is funded by a Senior Research Fellowship from the Wellcome Trust (#095066), and grants from the Bill & Melinda Gates Foundation (#OPP1119467, #OPP1093011, #OPP1106023 and #OPP1132415). DLS, AJT, SIH and RCR also acknowledge funding support from the RAPIDD programme of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center (FIC), National Institutes of Health (NIH). D.A.T.C. was supported by US NIH MIDAS program (U54-GM088491). OJB was funded by a Sir Henry Wellcome Fellowship funded by the Wellcome Trust (grant number 206471/Z/17/Z). This research received funding from the Oxford Martin School. Funding: Wellcome Trust.

Author Contributions

M.U.G.K., R.C.R., N.G., S.I.H. designed the experiment. D.A.T.C., O.J.B., D.M.P., D.B., S.B., N.R.F., O.G.P., D.L.S., A.J.T., S.E.R., J.S.B., and J.S.B. advised on methods and discussed results. M.U.G.K., R.C.R. and S.I.H. wrote the first draft of the manuscript. All authors read and approved the final version of the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-41192-3>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



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

© The Author(s) 2019