Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks

Flavio Finger[a], Tina Genolet[a], Lorenzo Mari[b], Guillaume Constantin de Magny[c], Noël Magloire Manga[d], Andrea Rinaldo[a,e,1], and Enrico Bertuzzo[a,1]

[a]Laboratory of Ecohydrology, École Polytechnique Fédérale Lausanne, 1015 Lausanne, Switzerland; [b]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milano, Italy; [c]Maladies Infectieuses et Vecteurs: Ecologie, Génétique, Evolution et Contrôle, Institute of Research for Development, 64501 Montpellier, France; [d]Service des Maladies Infectieuses et Tropicales de l'Hôpital de la Paix, Unité de Formation et de Recherche en Sciences de la Santé, Université Assane Seck de Ziguinchor, 27000 Ziguinchor, Senegal; and [e]Dipartimento dell'Ingegneria Civile, Edile ed Ambientale, Università di Padova, 35131 Padova, Italy

Contributed by Andrea Rinaldo, April 5, 2016 (sent for review November 12, 2015; reviewed by Vittoria Colizza and Aaron A. King)

The spatiotemporal evolution of human mobility and the related fluctuations of population density are known to be key drivers of the dynamics of infectious disease outbreaks. These factors are particularly relevant in the case of mass gatherings, which may act as hotspots of disease transmission and spread. Understanding these dynamics, however, is usually limited by the lack of accurate data, especially in developing countries. Mobile phone call data provide a new, first-order source of information that allows the tracking of the evolution of mobility fluxes with high resolution in space and time. Here, we analyze a dataset of mobile phone records of ~150,000 users in Senegal to extract human mobility fluxes and directly incorporate them into a spatially explicit, dynamic epidemiological framework. Our model, which also takes into account other drivers of disease transmission such as rainfall, is applied to the 2005 cholera outbreak in Senegal, which totaled more than 30,000 reported cases. Our findings highlight the major influence that a mass gathering, which took place during the initial phase of the outbreak, had on the course of the epidemic. Such an effect could not be explained by classic, static approaches describing human mobility. Model results also show how concentrated efforts toward disease control in a transmission hotspot could have an important effect on the large-scale progression of an outbreak.

mobile phone call data | cholera epidemics | spatially explicit epidemiological models | waterborne disease

Human mobility is undisputedly one of the main spreading mechanisms of infectious diseases. Understanding the propagation of an epidemic in a population at any spatial scale of analysis inevitably calls for the understanding of the underlying mobility patterns (1–6). Researchers have commonly focused on infectious diseases transmitted through direct contact between persons (e.g., refs. 1–4). The key role of human mobility has only recently been acknowledged also for water-related diseases (where transmission is mediated by water, which influences the habitat's suitability for the pathogen and/or its possible intermediate hosts), as highlighted by the development and widespread application of spatially explicit epidemiological models (7–10). Such models translate our comprehension of the mechanisms driving disease transmission [such as rainfall (10)] and spread [such as hydrologic transport of pathogens (8, 11) besides human mobility] into a simplified mathematical form. They may be used not only to predict the spatiotemporal pattern of the spread of a disease (12–14) but also to test alternative model implementations (15), or to evaluate the effects of various interventions on disease dynamics (16–18).

To include population movement in epidemiological models, researchers often rely on approaches such as gravity (e.g., ref. 19) or radiation (20) models, where the fluxes between any two sites are expressed as a function of their relative distance and the embedded population distribution. Such models have primarily been developed and tested for countries in the western world, where transportation networks are dense and efficient, supraregional travel is cheap, and regular commuting patterns are predominant. Lack of data has so far frustrated a thorough validation of such models in the developing world, where mobility drivers and patterns may be different than those of western countries. In some applications, the absence of information about mobility fluxes has been circumvented by inferring the parameters of the mobility model directly from epidemiological data (9, 10, 17). This, however, contributes to increasing uncertainty in model identification because many different factors concur in the spreading of an epidemic. Another important shortcoming of current mobility models is their inability to adapt to seasonal and subseasonal changes in mobility patterns.

With the increasing diffusion of mobile phones, which have become very widely used even in developing countries (21, 22), a new source of information about human mobility has emerged. Each time a phone emits or receives a call or text message, the antenna that the cell phone is logged into is registered by the service provider, along with the time of the event (23). It is thus possible to track the movement of cell phone users as they advance from antenna to antenna. Suitably aggregated and properly anonymized to prevent privacy issues (24), a sample of this data can be used to estimate fluxes of people between areas in a region by assigning a set of antennas to each geographical area in the study domain (e.g., based on administrative boundaries). The resolution in time can be as high as the typical frequency of calls allows, whereas the spatial resolution is limited only by the typical distance between two antennas (23). Using mobile phone records of a sufficiently large number of users, one can thus estimate human mobility fluxes with high accuracy, including spatiotemporal

Significance

Big data and, in particular, mobile phone data are expected to revolutionize epidemiology, yet their full potential is still untapped. Here, we take a significant step forward by developing an epidemiological model that accounts for the spatiotemporal patterns of human mobility derived by directly tracking properly anonymized mobile phone users. Such data allow us to investigate, with an unprecedented level of detail, the effect that mass gatherings can have on the spreading of waterborne diseases like cholera. Identifying and understanding transmission hotspots opens the way to the implementation of novel disease control strategies.

Author contributions: F.F., L.M., A.R., and E.B. designed research; F.F., T.G., and E.B. performed research; F.F., G.C.d.M., and N.M.M. analyzed data; and F.F., L.M., A.R., and E.B. wrote the paper.

Reviewers: V.C., INSERM; and A.A.K., University of Michigan.

The authors declare no conflict of interest.

[1] To whom correspondence may be addressed. Email: andrea.rinaldo@epfl.ch or enrico.bertuzzo@epfl.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1522305113/-/DCSupplemental.

variability across a variety of scales (24), and without resorting to any particular model.

A number of recent studies focus on the use of mobile phone data to extract human mobility patterns in developing countries at different scales in space and time (25–27). Others compare the movement patterns extracted from mobile records to traditional data sources such as censuses (28) and surveys (29). Several studies deal with the comparison with human mobility models (21, 30). In the context of infectious disease spread in developing countries, this new source of information enables previously unseen kinds of analyses. Examples are the derivation of magnitude and destination of population fluxes following a sudden outbreak (25, 31), and the quantification of the importance of human mobility and its seasonal variations on the spread of disease in terms of increased outbreak risk in and infectious pressure on connected areas (5, 30, 32–34).

Mass gatherings, such as pilgrimages, sport events, or music festivals, can be critical in the spread of infectious diseases via various transmission routes (35, 36). When it comes to orofecally transmitted diseases, such as shigellosis (37) or cholera (38, 39), insufficient safe drinking water supply and sanitary infrastructure related to overcrowding are often the main causes of local disease outbreaks and subsequent spread by homecoming infected attendees. To model the effect of mass gatherings, one needs to account for the spatiotemporal dynamics of human mobility and the associated short-term fluctuations of population distribution. Mobility models and static data sources, such as censuses or surveys, are therefore unsuitable. Conversely, mobile phone records contain all required information at the desired timescales and thus represent an excellent new data source for epidemiological models.

Here, we study the cholera epidemic that spread throughout Senegal in 2005. A distinctive feature of this outbreak was its sudden flare. It started from the order of magnitude of hundreds of cases per week during the first 3 mo of the year, localized in the region of Diourbel and surroundings, and abruptly jumped to thousands of cases at the end of March, rapidly spreading to 10 out of 11 regions of the country, with over 27,000 reported cases (Table 1). Anecdotal evidence (38, 40, 41) suggests that this first peak was related to a religious pilgrimage, the Grand Magal de Touba (GMdT), that took place in late March when an estimated 3 million pilgrims traveled to Touba in the region of Diourbel. During later stages, the outbreak evolved, showing distinct dynamics in different regions of the country, rainfall and the associated floods being important drivers, especially in the capital city of Dakar (39).

Here we develop a spatially explicit, fully mechanistic model for the 2005 Senegal cholera outbreak, based on previous work (10, 14, 42). In addition to human mobility, we take into account rainfall as an important driver of disease transmission (10, 39), and we incorporate the effect of overcrowding by assuming an increase in exposure and contamination rates caused by unusually high density of people, and the related pressure on water and sanitation infrastructures (*Materials and Methods*). Daily population fluxes between the 123 arrondissements of Senegal (Fig. S1*A*) are estimated from a dataset of roughly 150,000 randomly selected mobile phone users tracked during the entire year 2013 (*Materials and Methods* and ref. 43). We specifically aim at testing the role played by human mobility and mass gatherings in the spread of a cholera epidemic, with implications for disease control.

Results

Fig 1 shows the evolution of the estimated number of mobile people (i.e., people having left their home arrondissement on a given day) throughout the year 2013. Seasonal fluctuations, weekly patterns, and sudden peaks can clearly be identified. The latter correspond to mass gatherings, most notably the GMdT [which took place twice in 2013 (*Materials and Methods*)], and during which the number of people traveling outside their home arrondissement almost doubles with respect to an average day. Fig. 1*B* shows the estimated fraction of people present in every arrondissement of Senegal during the GMdT. Major differences can be noted with respect to the yearly average (Fig. 1*C*). People traveled to Touba from all over the country, and the estimated number of people present during the GMdT in the arrondissement where the city is located was nearly 6 times its usual population.

Table 1. Regions of Senegal (as of 2005) with their population (2005 estimates), total number of reported cases during the epidemic, cumulative incidence, and mobile phone sample size (relative to 2013 population)

Region	Population, ×10^6	Cases	Incidence, ‰	Sample size, ‰
Dakar	2.62	6,573	2.51	22.64
Diourbel	1.22	11,772	9.61	4.11
Fatick	0.64	1,928	3.00	4.63
Kaolack	1.06	1,014	0.96	5.19
Kolda	0.89	57	0.06	3.86
Louga	0.68	1,806	2.64	5.43
Matam	0.50	0	0	7.12
Saint-Louis	0.75	1,653	2.20	8.99
Tambacounda	0.58	87	0.15	6.11
Thiès	1.28	2,515	1.97	9.60
Ziguinchor	0.31	124	0.40	9.79

Model results and estimated uncertainties of the best performing model are shown in Fig. 2 (total cases and the regions most severely hit) and Fig. S2 (all regions). The values of the calibrated parameters are reported in Table S1. The model accurately reproduces the important peak of cases in Diourbel coinciding with the GMdT (coefficient of determination between modeled and reported weekly cases $R^2 = 0.78$ in the region of Diourbel) as well as the spread of the disease throughout Senegal by pilgrims returning to their homes. The second peak, most probably related to the rainy season, is also well reproduced ($R^2 = 0.72$ in the region of Dakar). The overall value of R^2, computed using all data points in all regions, is equal to 0.77. Fig. 3 shows the spatial distribution of cases in the country during the GMdT, and during two other key periods of the outbreak according to the reported cases and to our model.

A comparison of different models (*Supporting Information* and Table S2) shows that the ones including both human mobility fluxes between arrondissements and the effect of overcrowding outperform other models. Including either of the two mechanisms individually, however, is not sufficient to reproduce all features of the epidemic correctly (Fig. S3). In addition, a model adopting a calibrated gravity model performs poorly compared with models using mobile phone data to estimate human mobility. The inclusion of rainfall as a driver of the disease enables our model to capture the autumn peak in addition to the one related to the GMdT (Table S2 and Fig. S3). Finally, it appears that both the correction of bias in mobile phone ownership and the calibration of the initial number of infected in Diourbel improve the model performance.

Potential effects of localized interventions in Touba during the GMdT, such as improving sanitation and access to clean drinking water (*Materials and Methods*), are reported in Fig. 2. Under the assumptions of our model, these actions could have led to considerably lower numbers of new cases during the pilgrimage as well as all over the country during later stages of the outbreak (*Supporting Information*). For instance, a reduction of the rates of exposure and contamination by 10% (20%) in Touba during the GMdT could have led to a reduction of the total number of cases of 23% (38%) in Diourbel and 18% (34%) in the whole country (Fig. S4 and Table S3).

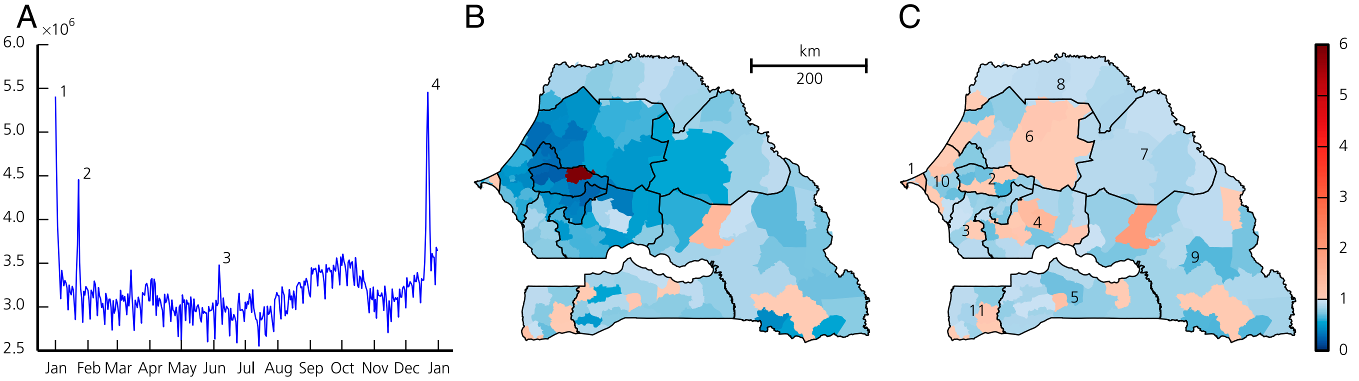

Fig. 1. (*A*) Daily evolution of the total number of moving people (i.e., people leaving their home arrondissement) throughout 2013 estimated from mobile phone records. Numbered peaks correspond to the following mass gatherings: GMdT (1 and 4), Gamou de Tivaouane (2), and Magal de Kazu Rajab (3). (*B* and *C*) Number of people present in each arrondissement on December 22, 2013, during the GMdT (*B*) and averaged over the year (*C*) divided by the number of people living there. Regions (according to the 2005 subdivision; see *Supporting Information*) are numbered as follows: Dakar (1), Diourbel (2), Fatick (3), Kaolack (4), Kolda (5), Louga (6), Matam (7), Saint-Louis (8), Tambacounda (9), Thiès (10), and Ziguinchor (11).

Discussion

The case study of the 2005 Senegal cholera outbreak illustrates the crucial role played by human mobility (and its spatiotemporal variability) in a cholera epidemic whose sudden flare and subsequent spread can be explained by the repercussions of a mass gathering that took place during the initial phase of the outbreak. Indeed, the temporary high density of people in Touba during the pilgrimage and the related pressure on water, sanitation, and health infrastructure are likely to have created favorable conditions for cholera transmission. After the initial peak, homecoming infected pilgrims spread the disease throughout vast parts of the country. No approach to quantify human mobility other than mobile phone data analysis could have provided the required level of detail to capture such phenomena. In addition, the comparison of different models shows that the actual epidemiological dynamic cannot be reproduced accurately without including mobility fluxes and the related effect of overcrowding, nor does the use of a gravity model lead to acceptable results.

The high temporal and spatial resolution of the mobility patterns extracted from mobile phone data allows identification of disease transmission hotspots suggesting intervention strategies to control the evolution of an epidemic, whose expected benefits can be evaluated using epidemiological models. In our case study, concentrated effort to reduce the transmission rate at the mass gathering site, for example, providing safe drinking water or sanitation for a higher number of people, could have had important effects, preventing numerous infections not only locally but throughout the whole country (*Supporting Information*).

Although our model has a high explanatory power at the whole-country scale and in regions with high cumulative incidence, it does not perform equally well in less affected regions. Although the timing of disease introduction and rainfall-related autumn peaks is well captured in all regions, the simulated temporal evolution of the number of cases deviates from the reported numbers of cases, especially in some of the regions less impacted by the disease. Possible reasons for this include higher influence

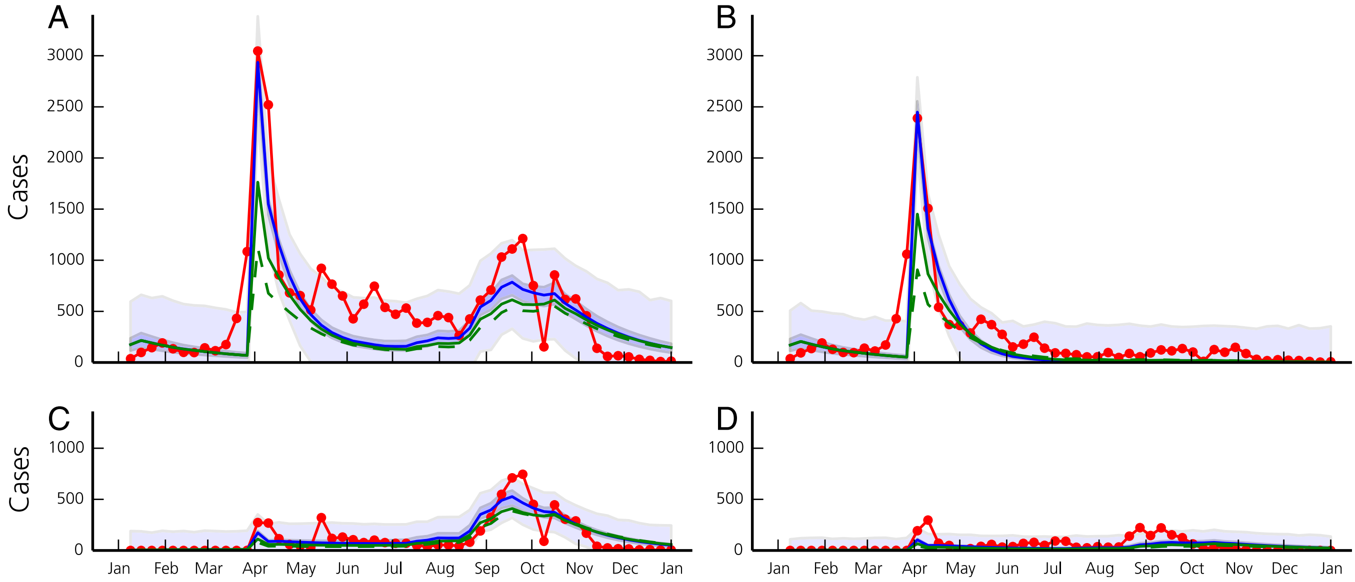

Fig. 2. Reported (red line) and modeled number of new cases per week for the entire country of Senegal (*A*), and for the regions of Diourbel (*B*), Dakar (*C*), and Thiès (*D*). Blue lines correspond to runs of the model (Eqs. **1–4**) with the best posterior parameter set. Shaded bands correspond the 2.5–97.5 percentiles of the uncertainty related to parameter estimation (dark blue) and of the total uncertainty assuming Gaussian, homoscedastic error (light blue). Modeled cases under the assumption of a 10% (solid green line) and 20% (dashed green line) reduction in transmission in Touba during the GMdT are also shown.

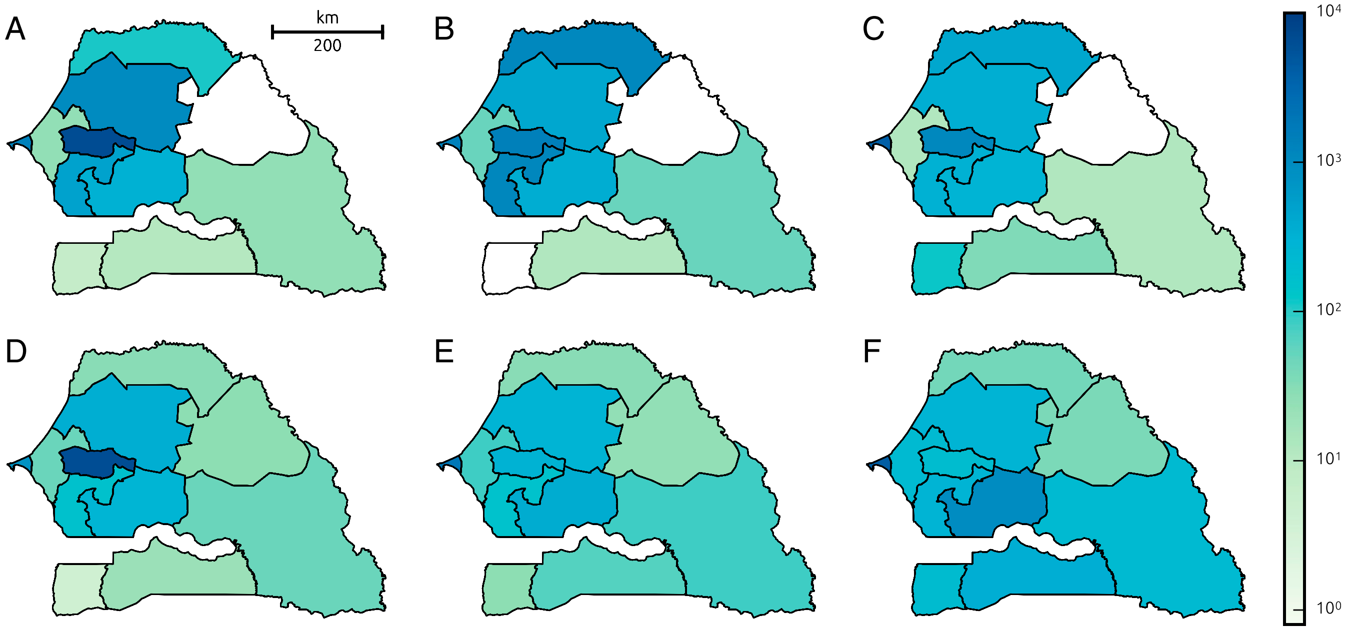

Fig. 3. Spatial distribution of reported (*A–C*) and modeled (*D–F*) cases from March 28 to May 29, the first weeks after the GMdT (*A* and *D*), from June 30 to September 4 (*B* and *E*), and from September 5 to December 31 (*C* and *F*).

of demographic stochasticity when the number of infected is low (*Supporting Information*), but also biased case reporting and/or identification (39) in regions with lower numbers of cases or with low population density (e.g., Matam). Also, one should consider that our likelihood formulation emphasizes peak values because it includes a square error term (*Supporting Information*).

Even if mobile phone data provide an excellent source of information about human mobility, several downsides still exist. One of them is the strong assumptions (*Materials and Methods* and ref. 34) made when translating mobile phone records to human mobility patterns, especially considering that they are difficult to validate due to the lack of alternative data sources. Studies comparing different methods and their underlying assumptions would be necessary to determine the sensitivity of the resulting mobility patterns. In addition, a potential source of inaccuracy in the analysis of mobile phone data is the possible presence of a bias in device ownership. A Kenya-based case study (44) has shown that mobile phone owners are more likely to be wealthy, male, and well educated, and that a bias exists between urban and rural populations. Urbanites with higher incomes tend to travel more often and farther, leading to overestimations of frequency and distance of trips (45). In our study, this effect was at least partially addressed by the introduction of a parameter (*Materials and Methods*) accounting for the underrepresentation of people staying at their home node. The values taken by this parameter during calibration might indeed indicate the presence of a bias, but might also be due to the fact that long-distance human mobility has played a major role in the propagation of the outbreak only during the pilgrimage, whereas local factors, such as precipitation and flooding, might have been more important in later stages. Additional sources of bias could arise from the fact that not all social classes are equally represented among the pilgrims (46), as well as from the uneven coverage of the mobile phone network between different areas of the country.

The reconstruction of the 2005 mobility matrix from that of 2013 (*Materials and Methods*) is based on the implicit assumption that general mobility patterns on relevant scales did not change significantly between the two years. Although several ways of reconstructing the 2005 mobility matrix have been compared (*Supporting Information*), their validity cannot be verified, due to the lack of alternative data sources. Among numerous factors that might have influenced mobility patterns is the cholera outbreak itself, which might have led to behavioral change of individuals in 2005, in turn affecting the disease dynamics (3, 42, 47).

In conclusion, we demonstrate that mobile phone records allow for an accurate quantification of spatiotemporal fluctuations in human mobility, whether short term, seasonal, or during rare events such as mass gatherings. The resulting mobility patterns allow for a deeper understanding of epidemiological dynamics. Inclusion in epidemiological models is straightforward and may lead to higher accuracy with respect to other approaches, as human movement patterns can be directly derived from data rather than inferred from models (e.g., gravity or radiation).

Materials and Methods

Mobile Phone Data and Inference of Human Mobility Patterns. Human mobility has been estimated from a dataset containing the locations of calls and text messages (hereafter calls) made by 146,352 randomly selected users throughout the year 2013 at arrondissement level (Fig. 1 and Table 1). The dataset has been temporarily released by an important Senegalese mobile phone provider, for the D4D-Senegal challenge (43) and can no longer be legally accessed. A record in the dataset consists of an anonymous user identification, a time stamp, and the arrondissement where the call was made. First the home of each user, e.g., the arrondissement where the most calls were made during night hours (1900 to 0700 hours), was determined. Then, for every day t, the quantity $Q_{ij}(t)$ was computed as the number of calls made while in arrondissement j by users with home node i divided by the total number of calls made by users with home node i. Under the assumption that the number of phone calls made by a user while in arrondissement j is proportional to the time spent there, the value $Q_{ij}(t)$ represents the community-level average fraction of time that users living in arrondissement i spend in arrondissement j during day t. $Q_{ii}(t)$ thus represent the fraction of time spent at the home arrondissement (34). The quantity $Q_{ij}(t)$ is provided (Datasets S1 and S2) to ensure the reproducibility of the results only. For any other use, a request should be submitted to Orange/Sonatel.

As the Islamic calendar is based on a lunar scheme with 354 d per year, the dates of the pilgrimages change within every Gregorian year. The GMdT, for instance, took place twice in 2013, on January 1 and December 22, whereas, in 2005, it was held just once, on March 29. To develop a model for the 2005 cholera outbreak, it was thus necessary to reconstruct the 2005 mobility matrix accordingly. For the purpose of this study, we averaged the human mobility matrix throughout 2013, excluding only the periods of the two occurrences of the GMdT. We used the resulting mobility matrix for all days in

2005, except for the period of the GMdT (March 29 ± 3 d), which, in turn, was assigned the mobility of the December 2013 event. Alternative ways of reconstructing the mobility matrix of 2005 from that of 2013, also accounting for seasonal components in the mobility and/or for other pilgrimages, have been tested but were not retained in model selection (*Supporting Information*, Fig. S1C, and Table S4).

Spatially Explicit Epidemiological Model. The spatially explicit epidemiological model used herein builds on previous work (10, 14, 16, 42). The model domain is the country of Senegal, each arrondissement (Fig. S1*A*, $N = 123$) being a node i with population H_i (*Supporting Information*). The population of each node i is subdivided into three compartments, namely susceptibles S_i, infected I_i, and recovered R_i. Every node is considered to have an ambient bacterial concentration B_i of *Vibrio cholerae*. We thus get the following set of differential equations describing the evolution of $4 \times N$ state variables (terms and parameters of the equations will be explained hereafter):

$$\frac{dS_i}{dt} = \mu(H_i - S_i) - \mathcal{O}_i(t)\mathcal{F}_i(t)S_i + \rho R_i \qquad [1]$$

$$\frac{dI_i}{dt} = \sigma\mathcal{O}_i(t)\mathcal{F}_i(t)S_i - (\gamma + \mu + \alpha)I_i \qquad [2]$$

$$\frac{dR_i}{dt} = \gamma I_i + (1 - \sigma)\beta_i(t)\mathcal{O}_i(t)\mathcal{F}_i(t)S_i - (\rho + \mu)R_i \qquad [3]$$

$$\frac{dB_i}{dt} = -\mu_B B_i + \frac{p}{aH_i}[1 + \lambda J_i(t)]\mathcal{O}_i(t)\mathcal{G}_i(t) \qquad [4]$$

where

$$\mathcal{O}_i(t) = \exp\left(\frac{\omega}{H_i}\sum_{j=1}^{N} M_{ji}(t)H_j\right) \qquad [5]$$

$$\mathcal{F}_i(t) = \beta\sum_{j=1}^{N} M_{ij}(t)\frac{B_j}{K + B_j} \qquad [6]$$

$$\mathcal{G}_i(t) = \sum_{j=1}^{N} M_{ji}(t)I_j. \qquad [7]$$

The population is assumed to be in demographic equilibrium, with per capita birth and natural death rate μ. Equations of different nodes are coupled via the human mobility matrix $M_{ij}(t)$, which is derived matrix $Q_{ij}(t)$ estimated from mobile phone data. To account for a possible underestimation of the number of people staying at their home node due to, e.g., bias in mobile phone ownership (44, 45), we introduce a calibration parameter c that relates the two matrices as follows:

$$M_{ii}(t) = cQ_{ii}(t) \qquad [8]$$

$$M_{ij}(t) = c_i'(t)Q_{ij}(t), \quad j \neq i \qquad [9]$$

$$c_i'(t) = \frac{1 - cQ_{ii}(t)}{\sum_{h \neq i} Q_{ih}(t)}, \qquad [10]$$

where $c_i'(t)$ ensures that rows sum to 1.

Susceptibles living at node i get infected at rate $\mathcal{O}_i(t)\mathcal{F}_i(t)$. $\mathcal{F}_i(t)$ is the rate at which a person living at node i comes into contact with contaminated water at node j during day t and becomes infected depending on the bacterial concentration B_j through a semisaturation function with parameter K and rate of exposure β. $\mathcal{O}_i(t)$ accounts for the effects of the increase in exposure and contamination rate due to the increased population density (overcrowding). This increase is modeled as an exponential function with the exponent composed of parameter ω and the number of people present at the node at time t divided by its actual population. We assume that only a fraction σ of infections are symptomatic. Asymptomatically infected hosts do not significantly contribute to the bacterial load in the environment, nor die of cholera (14, 16, 48), and can thus, for the purpose of the model, be considered recovered immediately. Symptomatically infected people may recover at rate γ or die from cholera-unrelated causes at rate μ or from cholera at rate α, whereas those who recover lose their acquired immunity at rate ρ or die from causes not related to cholera.

Bacteria are shed at rate p by infected $\mathcal{G}_i(t)$ present at node i at time t and reach the local environmental compartment, whose size is proportional to the population H_i with a proportionality constant a. The contamination of the environment is increased by local rainfall $J_i(t)$ via parameter λ (10) (*Supporting Information*), and by overcrowding through the factor $\mathcal{O}_i(t)$. The environmental bacteria population decays with rate μ_B. We define $B_i^* = B_i/K$. Expressing the system of equations in this term, parameters a and K get absorbed in $\theta = p/aK$ so that the number of free parameters is reduced by 2.

The model (Eqs. **1**–**4**) was solved numerically. Model outputs are the number of new cases per arrondissement and week, which are upscaled to the regional level for calibration and comparison with reported data (Table 1 and *Supporting Information*). Six parameters were estimated by using values from the literature, and another six were calibrated (Table S1), including the number of cholera cases present in the region of Diourbel in January 2005 (initial condition). Calibration was done using a method based on Markov chain Monte Carlo (MCMC) (*Supporting Information* and ref. 49).

To determine relevant processes to be included in the model and to find an appropriate compromise between accuracy and model complexity, hereby preventing overfitting, eight candidate models were compared using the Deviance Information Criterion (DIC) (50). Processes and mechanisms tested for their significance are the coupling of the local models in individual arrondissements through human mobility fluxes, the overcrowding effect, the correction of bias in mobile phone ownership, the inclusion of precipitation, and the calibration of the initial number of infected in Diourbel as a parameter. We also include a model that makes use of a gravity model instead of mobile phone data to determine human mobility. The model presented above (Eqs. **1**–**4**) is selected as the best performing candidate. Descriptions of all other candidate models, as well as results of the model comparison, are reported in *Supporting Information*. A discrete, stochastic version of the model has also been implemented to verify the validity of the assumption of continuous variables underlying Eqs. **1**–**4**, which proves reasonable (*Supporting Information* and Fig. S5).

Potential Effects of Local Interventions. To investigate the potential effects of local interventions, we ran our best fit model with 10% and 20% reduction of the rates of exposure to contaminated water and bacterial shedding. Such reductions are assumed to be concentrated in Touba during the GMdT, and could have been achieved by providing additional drinking water and sanitation facilities to the pilgrims (*Supporting Information*).

ACKNOWLEDGMENTS. Anonymous mobile phone data were made temporarily available by Orange and Sonatel within the framework of the D4D-Senegal challenge. The authors thank Dr. Malang Coly from the Office of the WHO representation in Dakar for his help in obtaining the cholera case data, and Prof. Renato Casagrandi and Prof. Marino Gatto (Politecnico di Milano) for useful discussions on the analysis of mobile phone data. F.F., E.B., and A.R. acknowledge support from the Swiss National Science Foundation project "Dynamics and controls of large-scale cholera outbreaks" (DYCHO CR23I2 138104).

1. Colizza V, Barrat A, Barthélemy M, Vespignani A (2006) The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc Natl Acad Sci USA* 103(7):2015–2020.
2. Balcan D, et al. (2009) Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc Natl Acad Sci USA* 106(51):21484–21489.
3. Meloni S, et al. (2011) Modeling human mobility responses to the large-scale spreading of infectious diseases. *Sci Rep* 1:62.
4. Bajardi P, et al. (2011) Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PLoS One* 6(1):e16591.
5. Wesolowski A, et al. (2012) Quantifying the impact of human mobility on malaria. *Science* 338(6104):267–270.
6. Tizzoni M, et al. (2014) On the use of human mobility proxies for modeling epidemics. *PLOS Comput Biol* 10(7):e1003716.
7. Chao DL, Halloran ME, Longini IM, Jr (2011) Vaccination strategies for epidemic cholera in Haiti with implications for the developing world. *Proc Natl Acad Sci USA* 108(17):7081–7085.
8. Gurarie D, Seto EYW (2009) Connectivity sustains disease transmission in environments with low potential for endemicity: Modelling schistosomiasis with hydrologic and social connectivities. *J R Soc Interface* 6(35):495–508.
9. Mari L, et al. (2012) Modelling cholera epidemics: The role of waterways, human mobility and sanitation. *J R Soc Interface* 9(67):376–388.
10. Rinaldo A, et al. (2012) Reassessment of the 2010–2011 Haiti cholera outbreak and rainfall-driven multiseason projections. *Proc Natl Acad Sci USA* 109(17):6602–6607.
11. Bertuzzo E, et al. (2008) On the space-time evolution of a cholera epidemic. *Water Resour Res* 44(1):W01424.
12. Bertuzzo E, et al. (2011) Prediction of the spatial evolution and effects of control measures for the unfolding Haiti cholera outbreak. *Geophys Res Lett* 38(6):L06403.
13. Reiner RC, Jr, et al. (2012) Highly localized sensitivity to climate forcing drives endemic cholera in a megacity. *Proc Natl Acad Sci USA* 109(6):2033–2036.
14. Bertuzzo E, Finger F, Mari L, Gatto M, Rinaldo A (2014) On the probability of extinction of the Haiti cholera epidemic. *Stochastic Environ Res Risk Assess*, 10.1007/s00477-014-0906-3.

15. Mari L, et al. (2015) On the predictive ability of mechanistic models for the Haitian cholera epidemic. *J R Soc Interface* 12(104):20140840.
16. Kühn J, et al. (2014) Glucose- but not rice-based oral rehydration therapy enhances the production of virulence determinants in the human pathogen Vibrio cholerae. *PLoS Negl Trop Dis* 8(12):e3347.
17. Tuite AR, et al. (2011) Cholera epidemic in Haiti, 2010: Using a transmission model to explain spatial spread of disease and identify optimal control interventions. *Ann Intern Med* 154(9):593–601.
18. Azman AS, et al. (2012) Urban cholera transmission hotspots and their implications for reactive vaccination: evidence from Bissau city, Guinea bissau. *PLoS Negl Trop Dis* 6(11):e1901.
19. Erlander S, Stewart NF (1990) *The Gravity Model in Transportation Analysis – Theory and Extensions* (VSP Books, Zeist, The Netherlands).
20. Simini F, González MC, Maritan A, Barabási A-L (2012) A universal model for mobility and migration patterns. *Nature* 484(7392):96–100.
21. Palchykov V, Mitrović M, Jo H-H, Saramäki J, Pan RK (2014) Inferring human mobility using communication patterns. *Sci Rep* 4:6174.
22. Wesolowski A, et al. (2014) Commentary: Containing the ebola outbreak - the potential and challenge of mobile network data. *PLoS Curr* 6:1.
23. Candia J, et al. (2008) Uncovering individual and collective human dynamics from mobile phone records. *J Phys A Math Theor* 41:224015.
24. de Montjoye Y-A, Hidalgo CA, Verleysen M, Blondel VD (2013) Unique in the crowd: The privacy bounds of human mobility. *Sci Rep* 3:1376.
25. Lu X, Bengtsson L, Holme P (2012) Predictability of population displacement after the 2010 Haiti earthquake. *Proc Natl Acad Sci USA* 109(29):11576–11581.
26. Lu X, Wetter E, Bharti N, Tatem AJ, Bengtsson L (2013) Approaching the limit of predictability in human mobility. *Sci Rep* 3:2923.
27. Perkins TA, et al. (2014) Theory and data for simulating fine-scale human movement in an urban environment. *J R Soc Interface* 11(99):20140642.
28. Wesolowski A, et al. (2013) The use of census migration data to approximate human movement patterns across temporal scales. *PLoS One* 8(1):e52971.
29. Wesolowski A, et al. (2014) Quantifying travel behavior for infectious disease research: A comparison of data from surveys and mobile phones. *Sci Rep* 4:5678.
30. Bengtsson L, et al. (2015) Using mobile phone data to predict the spatial spread of cholera. *Sci Rep* 5:8923.
31. Bengtsson L, Lu X, Thorson A, Garfield R, von Schreeb J (2011) Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti. *PLoS Med* 8(8):e1001083.
32. Tatem AJ, et al. (2014) Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. *Malar J* 13:52.
33. Wesolowski A, et al. (2015) Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. *Proc Natl Acad Sci USA* 112(35): 11114–11119.
34. Mari L, et al. (2015) Uncovering the impact of human mobility on schistosomiasis via mobile phone data. *Netmob Conference 2015: Data for Development Challenge Senegal* (Mass Inst Technol, Cambridge, MA), pp 71–97.
35. Abubakar I, et al. (2012) Global perspectives for prevention of infectious diseases associated with mass gatherings. *Lancet Infect Dis* 12(1):66–74.
36. Memish ZA, et al. (2015) Mass gathering and globalization of respiratory pathogens during the 2013 Hajj. *Clin Microbiol Infect* 21(6):571.e1–571.e8.
37. Wharton M, et al. (1990) A large outbreak of antibiotic-resistant shigellosis at a mass gathering. *J Infect Dis* 162(6):1324–1328.
38. World Health Organization (2008) *Cholera Country Profile: Senegal* (World Health Org, Geneva).
39. de Magny GC, et al. (2012) Cholera outbreak in Senegal in 2005: Was climate a factor? *PLoS One* 7(8):e44577.
40. International Federation of Red Cross and Red Crescent Societies (2007) *Senegal: Cholera Final Report* (Int Fed Red Cross Red Cresent Soc, Geneva), DREF Bull 05ME020.
41. Echenberg M (2011) *Africa in the Time of Cholera: A History of Pandemics from 1817 to the Present* (Cambridge Univ Press, Cambridge, UK).
42. Mari L, et al. (2012) On the role of human mobility in the spread of cholera epidemics: Towards an epidemiological movement ecology. *Ecohydrology* 5(5):531–540.
43. de Montjoye Y-A, Smoreda Z, Trinquart R, Ziemlicki C, Blondel VD (2014) D4D-Senegal: The second mobile phone data for development challenge, arXiv:1407:4885.
44. Wesolowski A, Eagle N, Noor AM, Snow RW, Buckee CO (2012) Heterogeneous mobile phone ownership and usage patterns in Kenya. *PLoS One* 7(4):e35319.
45. Wesolowski A, Eagle N, Noor AM, Snow RW, Buckee CO (2013) The impact of biases in mobile phone ownership on estimates of human mobility. *J R Soc Interface* 10(81): 20120986.
46. Boone C (2003) *Political Topographies of the African State: Territorial Authority and Institutional Choice* (Cambridge Univ Press, Cambridge, UK).
47. Funk S, Salathé M, Jansen VAA (2010) Modelling the influence of human behaviour on the spread of infectious diseases: A review. *J R Soc Interface* 7(50):1247–1256.
48. King AA, Ionides EL, Pascual M, Bouma MJ (2008) Inapparent infections and cholera dynamics. *Nature* 454(7206):877–880.
49. Foreman-Mackey D, Hogg DW, Lang D, Goodman J (2013) emcee: the MCMC hammer. *Publ Astron Soc Pac* 125(925):306–312.
50. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002) Bayesian measures of model complexity and fit. *J R Stat Soc Ser A Stat Soc* 64(4):583–639.
51. Linard C, Gilbert M, Snow RW, Noor AM, Tatem AJ (2012) Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS One* 7(2):e31743.
52. Koelle K, Rodó X, Pascual M, Yunus M, Mostafa G (2005) Refractory periods and climate forcing in cholera dynamics. *Nature* 436(7051):696–700.
53. Sorooshian S, Dracup JA (1980) Stochastic parameter estimation procedures for hydrologie rainfall-runoff models: Correlated and heteroscedastic error cases. *Water Resour Res* 16(2):430–442.
54. Gelman A, Carlin J, Stern H, Rubin D (2003) *Bayesian Data Analysis* (CRC Press, Boca Raton, FL), 2nd Ed.
55. Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81(25):2340–2361.
56. Central Intelligence Agency (2013) *The World Factbook 2013-14* (Central Intell Agency, Washington, DC).
57. Codeço CT (2001) Endemic and epidemic dynamics of cholera: The role of the aquatic reservoir. *BMC Infect Dis* 1:1.