Using Mobile Phone Data to Predict the Spatial Spread of Cholera

Linus Bengtsson, Jean Gaudart, Xin Lu, Sandra Moore, Erik Wetter, Kankoe Sallah, Stanislas Rebaudet, Renaud Piarroux

To cite this version:

Linus Bengtsson, Jean Gaudart, Xin Lu, Sandra Moore, Erik Wetter, et al.. Using Mobile Phone Data to Predict the Spatial Spread of Cholera. Scientific Reports, Nature Publishing Group, 2015, 5 (8923), <10.1038/srep08923>. <hal-01218137>

HAL Id: hal-01218137

https://hal-amu.archives-ouvertes.fr/hal-01218137
Submitted on 20 Oct 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Using Mobile Phone Data to Predict the Spatial Spread of Cholera

Linus Bengtsson1,2, Jean Gaudart3, Xin Lu4,1,2, Sandra Moore5, Erik Wetter2,6, Kankoe Sallah3, Stanislas Rebaudet5 & Renaud Piarroux5

1Department of Public Health Sciences. Karolinska Institutet, Stockholm, Sweden, 2Flowminder Foundation, Stockholm, Sweden, 3Aix-Marseille University, UMR 912 SESTIM (INSERM-IRD-AMU), Marseille, France, 4College of Information System and Management, National University of Defence Technology, Changsha, China, 5Aix-Marseille University, UMR MD 3, Marseille, France, 6Stockholm School of Economics, Stockholm, Sweden.

Effective response to infectious disease epidemics requires focused control measures in areas predicted to be at high risk of new outbreaks. We aimed to test whether mobile operator data could predict the early spatial evolution of the 2010 Haiti cholera epidemic. Daily case data were analysed for 78 study areas from October 16 to December 16, 2010. Movements of 2.9 million anonymous mobile phone SIM cards were used to create a national mobility network. Two gravity models of population mobility were implemented for comparison. Both were optimized based on the complete retrospective epidemic data, available only after the end of the epidemic spread. Risk of an area experiencing an outbreak within seven days showed strong dose-response relationship with the mobile phone-based infectious pressure estimates. The mobile phone-based model performed better (AUC 0.79) than the retrospectively optimized gravity models (AUC 0.66 and 0.74, respectively). Infectious pressure at outbreak onset was significantly correlated with reported cholera cases during the first ten days of the epidemic (p < 0.05). Mobile operator data is a highly promising data source for improving preparedness and response efforts during cholera outbreaks. Findings may be particularly important for containment efforts of emerging infectious diseases, including high-mortality influenza strains.

Re-occurring infectious disease outbreaks due to cholera, measles and other preventable infectious diseases contribute to a major disease burden affecting low- and middle-income countries1,2. Concurrently, outbreaks of new infectious diseases with pandemic potential pose a considerable threat to human life and development3,4. Response to, and ideally containment of, an infectious disease outbreak can be greatly improved if health care response and outbreak control measures can be focused to areas predicted to be at the highest risk of experiencing new outbreaks5,6. Accurate models of the geographic distribution of epidemic risk could significantly enhance the population-level effects of interventions implemented to control the spread of transmittable diseases5. Considerable progress has been made in predicting temporal evolution of epidemics once outbreaks have progressed beyond a small initial group of cases5,6. However, predicting spatial transmission routes of epidemics has proven to be remarkably difficult, due to the importance of rare, long-distance transmission events8, limited data on population mobility, unknown population immunity levels9, low sensitivity and specificity of case reports10 and limited access to accurate and spatiotemporally resolved case data11,12.

Empirical data has provided key insight into the spatial spread of measles in England13 and Niger14 as well as into influenza spread in the USA and Europe15,16. While population mobility plays a key role in such modelling studies10,16, it has not been possible, until now, to study detailed and concurrent data on both population mobility and spatiotemporal distribution of cases. Instead, empirical studies have used either models of population mobility, preferentially gravity models17, or census data on work-home commuting as proxies for total mobility during outbreaks11. Although highly significant correlations exist between these mobility patterns and retrospective data on epidemic spread, large unexplained variations remain10,11,17. It is also not clear how to choose and properly parameterize mobility models across contexts in new outbreaks. This is especially problematic during the critical early outbreak phases, when interventions have the greatest effect, but limited data are available to fit transmission models.

Anonymous mobile operator data may provide a new source of large-scale empirical data on which to build more accurate models of infectious disease spread. Mobile phone operators register the mobile phone tower closest to the mobile user at the time of each call and text message. This allows individual phones to be localized at a resolution equal to the coverage area of the mobile phone tower (typically one to ten km²). In the public health...
field, this data has notably been used as a proxy for nationwide mobility patterns in malaria modelling studies. However, the extent to which this type of data accurately reflects movements of infectious persons and its utility in predicting spatial spread of infectious agents have not been evaluated.

The largest cholera epidemic to strike a single country in recent history was the 2010 Haitian outbreak. The first confirmed cholera case in Haiti developed symptoms on October 14, 2010 in a hamlet 60 km north of the capital Port-au-Prince. The epidemic spread first explosively along the nearby Artibonite river (Fig. 1) and subsequently, during a period of two months, throughout the entire country.

In this study, we utilized data on the movement of 2.9 million anonymized mobile phone SIM cards in Haiti during the early phase of the Haitian cholera epidemic together with highly spatiotemporally resolved case data. We used these data to test the hypothesis that mobile operator data could be used to dynamically predict the spatial evolution of the epidemic from outbreak onset.

**Methods**

**Data collection. Cholera case data.** As soon as the epidemic was recognized, the Haitian government, with support from the US Centers for Disease Control and Prevention, implemented a nationwide monitoring program. Each day, government and non-governmental health facilities in Haiti reported probable cholera cases (ambulatory patients, hospital admissions and deaths) to the Directorate of Health in each of the ten administrative departments (Haitian provinces). Probable cases were defined according to a modified WHO definition as "acute watery diarrhoea, with or without vomiting, in persons of all ages". *Vibrio cholerae* O1 infection was confirmed via bacterial culture for early cases in all departments. We have previously validated case data from the National Cholera Surveillance system by carrying out field investigations, comparing case reports with data available from registers in cholera treatment centres managed by Haitian teams, Doctors without Borders and medical brigades from Cuba.

The daily case reports per health facility enabled us to determine daily case numbers per commune while the epidemic spread throughout the country (October 14 and 64 days onwards). We defined the end of the study period as December 16, when the peak of the epidemic was reached and all but one commune had reported at least one case. In 62 communes out of 140 communes, including the eight communes within the Port-au-Prince metropolitan area, there may have been patients who sought healthcare in neighbouring communes. For all such suspected communes we merged the communes into a single area, thereby creating a total of 78 study areas throughout the country (see S1 for details). One study area was excluded due to absence of mobile network coverage. To predict spatial spread of cholera, we defined a study area to have acquired a novel local outbreak if five or more cases were recorded on any given day, thereby avoiding misclassification of sporadic cases of diarrhoea as new cholera outbreaks. For sensitivity analyses of the outbreak definition, see S5.

**Mobile phone data.** The analysed anonymous mobile phone data consisted of the last outgoing call or text message each day from October 15 to December 19, 2010 for all 2.9 million users belonging to the largest mobile operator, Digicel Haiti. Research on bias in population mobility estimates stemming from differential ownership of mobile phones between socio-economic groups has been evaluated in Kenya, finding only minor bias. Mobile phone mobility patterns based on the mobile phone dataset used in this study have previously been shown to approximate mobility patterns from a representative survey of 2,500 households in the capital Port-au-Prince, Haiti, during the same year (2010). S2 provides further details on the mobile operator dataset.

**Analyses.** We used the mobile phone data to construct a mobility matrix \( M^{\text{phone}} \), with elements \( m_{ij}^{\text{phone}} \), indicating the average daily proportion of mobile phones relocating from study area \( i \) to \( j \), comparing their last registered location on day \( t \) with their last registered location on day \( t-1 \). The mobility network built on the basis of \( M^{\text{phone}} \) displays strong connectivity both between Port-au-Prince and large parts of the country as well as between other urban areas and their surrounding countryside (Fig. 1). We calculated the infectious pressure \( P(t) \), sustained by each study area \( j \) during the period from October 21 (from seven days after the disease onset of the first case in Haiti) to December 16, according to Eq. 1, in which \( c(t) \) is the number of reported cases in study area \( j \) on day \( t \):  

\[
P(t) = \sum_{i=1}^{n} m_{ij}^{\text{phone}} \sum_{k=1}^{7} c(t-k) 
\]

We thus assumed that a) the number of infectious individuals in a study area was proportional to the cumulative number of reported cases in the area during the preceding seven days (approximating the generation time of cholera, see also S5 and below) and b) the proportion of mobile phone movements between study areas was a measure of the movements of infectious cases between study areas.

For comparative purposes, we implemented a gravity model of population mobility. Gravity models have previously been used to model mobility of infectious persons in a large number of studies in Haiti and elsewhere and assume that population mobility between areas depends positively on their population sizes and negatively on the distance between them. We calculated the infectious pressure according to Eq. 1, replacing \( m_{ij}^{\text{phone}} \) by \( m_{ij}^{\text{grav}} \) using the following gravity model:

\[
m_{ij}^{\text{grav}} = \mu_{i} \frac{H_{i} H_{j} e^{-\frac{d_{ij}}{\delta}}}{\sum_{j} H_{j} e^{-\frac{d_{ij}}{\delta}}} 
\]

in which \( \mu_{i} \) is the average daily proportion of the population in area \( i \) that moves out of the area. The remaining ratio is the estimated probability that an individual leaving study area \( i \) goes to study area \( j \). \( H_{i} \) denotes the population in study area \( j \), and \( d_{ij} \) denotes the distance between the population-weighted centroids of study areas \( i \) and \( j \). In the absence of detailed mobility data, values for \( \mu_{i} \) and \( \delta \) (the scaling parameter) in Eq. 2 are unknown and need to be assigned. Parameter values vary however widely between studies. As appropriate values for Haitian mobility are unknown, we chose to optimize the model based on the retrospective case data from the complete study period. Note that this optimisation thus could not have been performed until after the spatial spread of the epidemic was complete. Our comparison model thus performs better than a model that could have been developed during the epidemic. We produced two separate optimisation runs. In the first we optimised the gravity model by choosing values for \( \mu_{i} \) and \( \delta \) (0.154 and 122 respectively), which minimised the residual sum of squares between reported daily cholera cases in each study area and the estimated pressure from the gravity model. In the second we chose parameter values (0.153 and 3.5 respectively) that maximised the area under the curve (AUC), among all possible ROC curves. The ROC curve in this analysis depicts the sensitivity and specificity, using increasing thresholds of infectious pressures, to predict outbreak occurrence (see Results and Fig. 2b). We denote infectious pressures calculated from mobile phone movements by \( P^{\text{phone}} \) and from the optimised gravity models by \( P^{\text{grav}} \) and \( P^{\text{optim}} \), respectively (see S3).

**Results**

**Outbreak risk.** For an efficient response to a developing epidemic, it is important to rapidly focus intervention resources to areas at

---

**Figure 1 | Mobile phone mobility network.** The average absolute number of mobile phones moving between the study areas (October 15 to December 19, 2010). Thicker, bluer lines indicate larger number of travelers. The original outbreak location (Mirebalais), the Artibonite River (dark blue) and Port-au-Prince (PAP) are depicted (visualisation using Gephi and ArcGIS).
highest risk. By utilizing data for all days of the study period, we plotted the proportion of non-infected communes that experienced an outbreak within seven days, for various intervals of infectious pressure, \( P_{phone} \) (Fig. 2a). The risk of a study area experiencing a new outbreak correlated closely with the infectious pressure. Over a pressure level of 22 \( (P_{phone}) \), all areas (six study areas) experienced outbreaks within seven days (see S5 for \( P_{grav1} \), \( P_{grav2} \) and sensitivity analyses).

**Sensitivity and specificity of outbreak predictions.** Building upon this strong correlation between infectious pressure and outbreak risk, we created a binary test to predict an outbreak occurring within the upcoming seven days, based solely on thresholds of infectious pressure. We predicted an outbreak to occur at a given pressure threshold and plotted the corresponding sensitivity and specificity of each threshold (Fig. 2b). We compared the model based on the mobile operator mobility data \( (P_{phone}) \) with the gravity models \( (P_{grav1} \text{ and } P_{grav2}) \), for which \( P_{grav2} \) was optimised specifically to yield the maximum possible area under the curve (AUC) in this analysis. Comparing these ROC curves, the \( P_{phone} \) model clearly performs better than the \( P_{grav1} \) model and slightly better than the \( P_{grav2} \) model, yielding a higher specificity for a given level of sensitivity. Note that both gravity models rely on parameter optimisations that could not have been performed until the epidemic spread was completed. Analyses of ranks of infectious pressure yielded similar results (S4). As the generation time of cholera in Haiti is uncertain and may have deviated from the seven-day period assumed in this study (Eq. 1)\(^{29} \), we additionally calculated infectious pressure based on other time periods (three, five and nine days), which did not alter the results (S5).

**Early outbreak incidence.** In addition to predicting the risk of a new outbreak occurring in an area, an effective health care response requires good estimates of the number of cases that are likely to occur if an outbreak takes place. We may however expect the local evolution of an epidemic, after its start, to be largely dependent on local environmental and behavioural factors. One may thus hypothesize that the infectious pressure leading to the seeding of an outbreak would have little further influence on the number of cases in an area. This did however not seem to be the case.

We correlated, for each newly infected area, the infectious pressure sustained by the area at outbreak onset with the average daily number of cases during the first \( D \) days of the new outbreak. For all values of \( D \) from one to ten days, we found a linear correlation \((r)\) of approximately 0.3 (Fig. 3). Correlation coefficients were significant \((p < 0.05)\) for all periods for the \( P_{phone} \) model and non-significant for all 10 periods for the \( P_{grav1} \) and the \( P_{grav2} \) model (one outlier excluded).

**Discussion**

Our results show that the risk of epidemic onset of cholera in a given area and the initial intensity of local outbreaks could have been anticipated during the early days of the Haitian epidemic using case reports and the mobility patterns of mobile phones. We show that the specificity and sensitivity of predictions of epidemic spread was improved or comparable to currently available optimized mobility models. Most importantly, the predictions based on the mobile operator data did not rely on retrospective optimization of parameter models and could thus be available from the start of an outbreak. This is important as gravity model parameters are highly context specific\(^{33,34} \). These results indicate that outbreak preparedness and response to epidemic agents, such as cholera, can be enhanced. The findings may have particular importance for improving early con-
tainment efforts of emerging infectious diseases, such as high mortal-
ty rates of pandemic influenza, and the response to vaccine-
preventable diseases, such as measles, in low-income settings.

Cholera is known to be disseminated not only via human move-
ment, but also by water and sometimes food contamination17. Even
stronger predictive power may thus be achieved for infectious agents
that exhibit exclusive person-to-person transmission. Furthermore,
in this study, we focused on an extremely large and rapid cholera
epidemic. It is likely that surveillance data would be more accurate
for a disease that exhibits obvious symptoms and longer generation
times, which may reduce reporting bias and delays. In particular,
mobile operator data may represent a powerful tool for the contain-
ment of measles, which exhibits a high mortality rate, is preventable
with vaccine, is readily identifiable based on simple clinical criteria
displays a sufficiently long generation time. Mobile phone-based
connectivity matrices may also be very useful for early containmen-
t of emerging infectious agents, such as the localized appearance of
high-mortality influenza strains with pandemic potential16. Mobile
operator data may be especially advantageous in settings where poor
road quality renders distance a weak measure of connectivity.

Access to operator data is a prerequisite for the future use of
the method. The study demonstrates however that for the purpose
of predicting future outbreaks, mobile phone operators may not need to
provide access to their complete customer databases, but rather only
to aggregated data on mobility between areas. Such aggregated con-
cnectivity matrices can be made available for preparedness purposes
before outbreaks. They can then be repeatedly updated during the
outbreak response to take into account changes in population mobili-
ity, which would not be captured by a gravity model.

In newly infected areas, infectious pressure based on the mobile
phone model (but not gravity models) correlated positively with
initial incidence. The correlation between mobility and initial inci-
dence was unexpected as cholera transmission depends on a number
of local socio-economic and environmental conditions. There may
either be a causal connection between mobility and outbreak size (if
rapid outbreaks were caused by multiple seeding events), alterna-
tively areas between which there is high mobility may have envir-
onmental and behavioural similarities. Although the findings
indicate an important new use of mobility data, a policy relevant tool
for predicting early case numbers thus needs to incorporate addi-
tional variables to strengthen correlations.

Some study limitations should be noted. Reporting errors and
delays are likely to have occurred in the case data and may have
reduced the predictive accuracy for both the phone and gravity mod-
els. This will also be the case in future applications of the method.
The case data in the study should be interpreted as reflecting relative
rather than absolute differences in case load per area, as the
Haitian cholera reporting system excluded asymptomatic and mild
infections, as well as some severely ill patients who did not reach
health facilities18.

Short-range movements may be under-recorded in the mobile
data since they take place over shorter time intervals. However,
weighting short and long-term movements differently did not
change the prediction results (S5). Frequency of mobile phone use
varies throughout the country and differential mobility between
phone and non-phone users may exist. Studies evaluating bias in
mobility estimates based on mobile operator data do however pro-
vide strong support for operator data being the currently best mea-
sure of nationwide mobility patterns19,20,21. Although this study
focuses only on the influence of human mobility, future mobile
phone based models focusing on cholera may benefit from including
data on spatial distributions of access to water and sanitation18, bac-
terial transmission via waterways18, agricultural practices19, differ-
ential population immunity levels22 and interactions between infectiousness and mobility and between infectiousness and phone use.

In summary, the results show that mobile phone mobility patterns
in Haiti during the 2010 cholera outbreak enabled predictions of
epidemic spread, which did not require retrospective optimization of
parameter models and could thus be available at outbreak onset.
The results imply that anonymous mobile phone data may represent
a key data source to both increase our understanding of the mechan-
isms underlying the spatial spread of infectious agents and provide
an important policy-relevant tool for future outbreak preparedness
and response efforts.

References

1. Black, R. E. et al. Global, regional, and national causes of child mortality in 2008: a
systematic analysis. Lancet 375, 1969–1987, doi:10.1016/S0140-6736(10)60549-1
(2010).

2. Murray, C. J. et al. Disability-adjusted life years (DALYs) for 291 diseases and
injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of
Disease Study 2010. The Lancet 380, 2197–2223 (2013).

3. Murray, C. J., Lopez, A. D., Chin, B., Feehan, D. & Hill, K. H. Estimation of
potential global pandemic influenza mortality on the basis of vital registry data
from the 1918–20 pandemic: a quantitative analysis. Lancet 368, 2211–2218,
doi:10.1016/S0140-6736(06)68985-4 (2006).

4. Smith, R. D., Keogh-Brown, M. R., Barnett, T. & Tait, J. The economy-wide impact
of pandemic influenza on the UK: a computable general equilibrium modelling
exercise. Bmj 339, b4571, doi:10.1136/bmj.b4571 (2009).

5. Longini, I. M. Jr. et al. Containing pandemic influenza at the source. Science 309,
1083–1087, doi:10.1126/science.1117177 (2005).

6. Ferguson, N. M. et al. Strategies for containing an emerging influenza pandemic in
Southeast Asia. Nature 437, 209–214, doi:10.1038/nature00407 (2005).

7. Anderson, R. M. & May, R. M. Infectious diseases of humans: dynamics and
control. (Oxford University Press, 1991).

8. Tuite, A. R. et al. Cholera epidemic in Haiti, 2010: using a transmission model to
explain spatial spread of disease and identify optimal control interventions. Am
Intern Med 154, 593–601, doi:10.7326/0003-4819-15-920100530-00334
(2011).

9. Grassly, N. C. & Fraser, C. Mathematical models of infectious disease transmis-
sion. Nature Reviews Microbiology 6, 477–487 (2008).

10. Riley, S. Large-scale spatial–transmission models of infectious disease. Science
316, 1298–1301, doi:10.1126/science.1134695 (2007).

11. Viboud, C. et al. Synchrony, waves, and spatial hierarchies in the spread of
influenza. Science 312, 447–451, doi:10.1126/science.1125237 (2006).

12. Eggo, R. M., Cauchemez, S. & Ferguson, N. M. Spatial dynamics of the 1918
influenza pandemic in England, Wales and the United States. J R Soc Interface 8,
233–243, doi:10.1098/rsif.2010.0216 (2011).

13. Grenfell, B. T., Bjornstad, O. N. & Kappey, J. Travelling waves and spatial
hierarchies in measles epidemics. Nature 414, 716–723, doi:10.1038/414716a
(2001).

14. Bhatti, N. et al. Explaining seasonal fluctuations of measles in Niger using
nighttime lights imagery. Science 334, 1424–1427 (2011).

15. Crepey, P. & Barthelemy, M. Detecting robust patterns in the spread of infectious
diseases: a case study of influenza in the United States and France. Am J Epidemiol
166, 1299–1311, doi:10.1093/aje/kwz266 (2007).

16. Balcan, D. et al. Multiscale mobility networks and the spatial spreading of
infectious diseases. Proceedings of the National Academy of Sciences of the United
States of America 106, 2184–2189, doi:10.1073/pnas.0906910106 (2009).

17. Truscott, J. & Ferguson, N. M. Evaluating the adequacy of gravity models as a
description of human mobility for epidemic modelling. PLoS Comput Biol 8,
e1002699, doi:10.1371/journal.pcbi.1002699 (2012).

18. Crepey, P. & Barthelemy, M. Detecting robust patterns in the spread of epidemics:
a case study of influenza in the United States and France. Am J Epidemiol 166,
1244–1251, doi:10.1093/aje/kwz266 (2007).

19. Viboud, C., Nelson, M. L., Tan, Y. & Holmes, R. C. Contrasting the epidemiological
and evolutionary dynamics of influenza spatial transmission. Philosophical
transactions of the Royal Society of London. Series B, Biological sciences 368,
20120199, doi:10.1098/rstb.2012.0199 (2013).

20. Wesołowski, A. et al. Quantifying the impact of human mobility on malaria.
Science 338, 267–270, doi:10.1126/science.1223467 (2012).

21. Le Menach, A. et al. Travel risk, malaria importation and malaria transmission in
Zanzibar. Sci Rep 1, 93, doi:10.1038/srep00093 (2011).

22. Barzilay, E. I. et al. Cholera surveillance during the Haiti epidemic—the first 2
years. The New England journal of medicine 368, 599–609, doi:10.1056/
NEJMoa1204927 (2013).

23. Piarroux, R. et al. Understanding the cholera epidemic, Haiti. Emerg Infect Dis
17, 1161–1168, doi:10.3201/eid1707.110509 (2011).

24. Frerichs, R. R., Keim, P. S., Barrais, R. & Piarroux, R. Nepalese origin of cholera
epidemic in Haiti. Clin Microbiol Infect 18, E158–E163, doi:10.1111/j.1469-
0691.2012.03841.x (2012).

25. Frerichs, R. R., Piarroux, R., Keim, P. S., Barrais, R. & Piarroux, R. Source attribution
of 2010 cholera epidemic in Haiti. Proceedings of the National Academy of Sciences
of the United States of America 109, E3208, author reply E3209 doi:10.1073/
pnas.1211512109 (2012).
26. Lu, X., Bengtsson, L. & Holme, P. Predictability of population displacement after the 2010 Haiti earthquake. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 11576–11581, doi:10.1073/pnas.1203882109 (2012).

27. Wesolowski, A., Eagle, N., Noor, A. M., Snow, R. W. & Buckee, C. O. The impact of biases in mobile phone ownership on estimates of human mobility. *J R Soc Interface* **10**, 20120986, doi:10.1098/rsif.2012.0986 (2013).

28. Bengtsson, L., Lu, X., Thorson, A., Garfield, R. & von Schreeb, J. 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**, e1001083, doi:10.1371/journal.pmed.1001083 (2011).

29. Mari, L. et al. Modelling cholera epidemics: the role of waterways, human mobility and sanitation. *J R Soc Interface* **9**, 376–388, doi:10.1098/rsif.2011.0304 (2012).

30. Azman, A. S. et al. Urban cholera transmission hotspots and their implications for reactive vaccination: evidence from Bissau city, Guinea bissau. *PLoS Negl Trop Dis* **6**, e1901, doi:10.1371/journal.pntd.0001901 (2012).

31. Rinaldo, A. et al. Reassessment of the 2010–2011 Haiti cholera outbreak and rainfall-driven multiseason projections. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 6602–6607, doi:10.1073/pnas.1203331109 (2012).

32. Bradley, A. P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition* **30**, 1145–1159 (1997).

33. Garske, T. et al. Travel patterns in China. *PLoS One* **6**, e16364, doi:10.1371/journal.pone.0016364 (2011).

34. Alberto Rubio, V. F.-M., Frias-Martinez, E. & Oliver, N. Human Mobility in Advanced and Developing Economies: A Comparative Analysis. *Telefonica Research, Madrid, Spain.*

35. Sack, D. A., Sack, R. B., Nair, G. B. & Siddique, A. K. Cholera. *Lancet* **363**, 223–233 (2004).

36. Jackson, B. R. et al. Seroepidemiologic survey of epidemic cholera in Haiti to assess spectrum of illness and risk factors for severe disease. *The American journal of tropical medicine and hygiene* **89**, 654–664, doi:10.4269/ajtmh.13-0208 (2013).

37. Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A. L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).

38. Waldman, R. J., Mintz, E. D. & Papowitz, H. E. The cure for cholera--improving access to safe water and sanitation. *N Engl J Med* **368**, 592–594, doi:10.1056/NEJMtp1214179 (2013).

39. Gaudart, J. et al. Spatio-temporal dynamics of cholera during the first year of the epidemic in Haiti. *PLoS Negl Trop Dis* **7**, e2145, doi:10.1371/journal.pntd.0002145 (2013).

40. Chao, D. L., Halloran, M. E. & Longini, I. M. Vaccination strategies for epidemic cholera in Haiti with implications for the developing world. *Proceedings of the National Academy of Sciences* **108**, 7081–7085 (2011).

**Acknowledgments**

The authors are grateful to Digicel Haiti for making available the anonymous network data. This work was funded by the Embassy of France, Haiti and the National Natural Science Foundation of China (No. 71301165, No. 71101013). LB and XL acknowledge support from the Swedish Research Council. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Author contributions**

L.B. wrote the main manuscript text and all authors provided edits. X.L., L.B., R.P., K.S. and J.G. performed analyses. E.W., S.M. and S.R. provided additional input and participated in data collection. X.L. prepared the figures.

**Additional information**

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Bengtsson, L. et al. Using Mobile Phone Data to Predict the Spatial Spread of Cholera. *Sci. Rep.* **5**, 8923; DOI:10.1038/srep08923 (2015).