

# Spatially explicit modelling of cholera epidemics

THÈSE N° 7565 (2017)

PRÉSENTÉE LE 3 MARS 2017

À LA FACULTÉ DE L'ENVIRONNEMENT NATUREL, ARCHITECTURAL ET CONSTRUIT  
LABORATOIRE D'ÉCOHYDROLOGIE  
PROGRAMME DOCTORAL EN GÉNIE CIVIL ET ENVIRONNEMENT

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Flavio FINGER

acceptée sur proposition du jury:

Prof. T. I. Battin, président du jury  
Prof. A. Rinaldo, Prof. E. Bertuzzo, directeurs de thèse  
Prof. W. J. Edmunds, rapporteur  
Dr A. Azman, rapporteur  
Prof. M. Salathé, rapporteur



ÉCOLE POLYTECHNIQUE  
FÉDÉRALE DE LAUSANNE

Suisse  
2017



Dene wos guet geit  
Giengs besser  
Giengs dene besser  
Wos weniger guet geit  
Was aber nid geit  
Ohni dass's dene  
Weniger guet geit  
Wos guet geit

Drum geit weni  
Für dass es dene  
Besser geit  
Wos weniger guet geit  
Und drum geits o  
Dene nid besser  
Wos guet geit

— Mani Matter



# Acknowledgements

First and foremost I wish to thank Andrea Rinaldo and Enrico Bertuzzo for their enthusiasm, continuous encouragement and eager support during my PhD. The time spent at the Laboratory of Ecohydrology has had a great impact on my personal and professional development and has shaped my understanding on how research should be conducted. I am particularly grateful to Andrea and Enrico for the time we spent discussing the full details of every aspect of my work, while they ensured that I never lost the sight of the big picture. I also want to thank them for providing a great working environment, plenty of liberties and a great deal of trust.

My sincere thanks go to Lorenzo Mari and Damiano Pasetto for their comprehensive advice and insightful answers given to me at innumerable occasions, to Andrew Azman for his guidance and support and for helping me to put my work in to the perspective of real-world applications, and to all other researchers I collaborated with during the last years. Thank you to all the people who provided me with the data essential to every modeler, in particular to Guillaume Constantin de Magny, Noël Magloire Manga, Thomas Esch, Didier Bompangue, Wim Thierry, the teams from Médecins sans Frontières, Epicentre, the WHO national office in Dakar and the Ministry of Health of Chad.

I gratefully acknowledge the funding from the Swiss National Science Foundation (SNF/FNS). I also want to give thanks to all my colleagues and friends, in particular to Paolo Benettin for being a great office mate and for all the deep conversations about serious and not so serious subjects we shared, to Pierre Queloz and Bernard Sperandio for innumerable discussions about snow and weather conditions, to Luca Carraro and Ana Clara Santos for enduring the latter, to Andrea Giometto for giving me a glimpse of Italian culinary delicacies, to Silvia Zaoli for her cheerful nature and her love of broccoli, to Javier Perez-Saez for his great motivation, his pragmatism and his delicate humor and to Jonathan Giezendanner for his uncountable jokes and for reminding me to have lunch every day. Special thanks go to Anna Rothenbühler for leading the way through the administrative jungle of academia. Thank you to all the above and also to Tristan Brauchli, Francesco Carrara, Nathalie Ceperley, Lorenzo Gorla, Allyn Knox, Théophile Mandé, Giulia Ruggeri, Bettina Schaepli and many others for the great ambiance, innumerable beers at Sat, and friendships that go beyond the boundaries of EPFL.

Finally, a big thank you goes to my family, who always supported and encouraged me, and to my friends from outside EPFL for keeping me connected to the world out of academia.

*Lausanne, 28 November 2016*

F. F.



## Summary

Understanding the epidemiology of cholera, when and where it occurs and how it spreads, is key to its prevention and control. Models can help to apprehend cholera outbreaks by providing insight into critical epidemiological processes, and may be used to evaluate alternative intervention strategies or to predict the future course of epidemics. This thesis aims at advancing the evolution of spatially explicit epidemiological models of cholera outbreaks through methodological developments and practical applications.

Over 160 years after John Snow first analyzed the spatial pattern of cholera cases in London and identified water as its pathway of contagion, the disease remains a major public health threat in many regions around the globe. It causes an estimated number of 2.86 (1.30 – 4.00) million cases and 95 000 (21 000 – 143 000) deaths in 69 endemic countries every year.

A set of metapopulation and individual-based, mechanistic and semi-mechanistic epidemiological models has been developed to tackle epidemiological questions at the country, sub-national and city scale. The models explicitly take into account the spatial variability of epidemiological processes such as the spread of the disease through hydrological connectivity and human mobility, or the high resolution spatiotemporal clustering of cases. A method to extract large-scale mobility fluxes from mobile phone call records and directly incorporate them into a model has also been established. Different environmental drivers of cholera epidemics have been taken into account. The models have been applied to recent cholera outbreaks in Haiti, Senegal, Chad and the Democratic Republic of the Congo.

Results highlight the important part played by human mobility in the spreading of the disease and the influence of rainfall and other climatic variables as drivers of disease dynamics in several settings. Applications demonstrate how models can inform epidemiological policy and show the effect of alternative intervention strategies on the course of an epidemic. The evaluation of the preventive allocation of oral cholera vaccine, antibiotics and/or water, sanitation and hygiene interventions within a given radius around reported cases in densely populated areas shows that such interventions are effective and efficient alternatives to mass intervention campaigns. Moreover, an alternative type of oral rehydration solution proves to have a significant effect on the course of a simulated epidemic.

This thesis concludes that the explicit treatment of spatial heterogeneity at an appropriate scale is crucial to reproduce real-world dynamics of cholera outbreaks. It highlights how suitable models can address relevant questions about the dynamics of the disease, provide insights into ongoing epidemics, may aid emergency management and complement current epidemiological practice.

## Summary

---

**Key words:** cholera, model, spatially explicit, spatial heterogeneity, scales, infectious disease, epidemiology, interventions, treatment, prediction, public health, human mobility



## Résumé

La compréhension des mécanismes épidémiologiques du choléra, quand et où il apparaît et comment il se propage, est primordial afin de pouvoir planifier des mesures de prévention de manière ciblée et contrôler la maladie. La modélisation permet d'aider à saisir les processus épidémiologiques essentiels et peut être utilisée afin d'évaluer l'efficacité de différents scénarii d'intervention ou de produire des prévisions à court terme sur le développement et la propagation d'épidémies. Cette thèse prétend à contribuer à l'avancée de la modélisation distribuée du choléra à travers des développements méthodologiques ainsi que par des applications pratiques.

Plus de 160 ans après que John Snow ait analysé en pionnier la distribution spatiale des cas de choléra à Londres et identifié l'eau en tant que principal vecteur de contagion, la maladie reste une préoccupation primaire en santé publique dans beaucoup de pays tout autour du monde. Parmi les 69 pays où la maladie est endémique, on dénombre quelques 2.86 (1.30 – 4.00) millions de cas d'infection dont 95 000 (21 000 – 143 000) mortelles.

Un jeu de modèles, basés sur les principes de métapopulation ou sur l'individu, a été développé dans ce travail afin de répondre aux questions épidémiologiques à l'échelle de pays, de régions ou de villes. Les modèles prennent en compte la variabilité spatiale des processus épidémiologiques tels que la propagation de la maladie à travers le réseau hydrologique ou les flux de mobilité humaine, ou les phénomènes spatiotemporels d'accumulations multiples de cas (clustering). Une nouvelle méthode permettant d'extraire les flux de mobilité humaine à large échelle à partir de données d'appels téléphonique sur le réseau mobile a été développée, et intégrée dans les modèles. Des applications à des épidémies récentes en Haïti, au Sénégal, au Tchad et en République Démocratique du Congo sont également présentées.

Les résultats relèvent le rôle important de la mobilité humaine dans la propagation de la maladie, ainsi que l'influence des précipitations et autres variables climatiques sur la dynamique de propagation dans différentes situations. Les applications démontrent comment les modèles peuvent être utilisés pour extraire des informations qui influencent la pratique épidémiologique et permettent d'appréhender les effets de différentes stratégies de lutte au cours d'épidémies. On peut par exemple citer l'évaluation de stratégies de lutte préventive utilisant des vaccins ou des antibiotiques, des interventions assurant l'accès à l'eau potable ou aux installations sanitaires dans un rayon défini autour de nouveau cas dans des zones densément peuplées. Les bénéfices significatifs d'un nouveau type de solution de réhydratation administrée par voie orale ont également été démontrés dans le cas d'une épidémie simulée. Cette thèse conclue que le traitement explicite de l'hétérogénéité spatiale à une échelle ap-

## Résumé

---

propriété est cruciale afin de reproduire les dynamiques d'épidémies de choléra de manière adéquate. Elle relève comment des modèles appropriés permettent de répondre à des questions sur la dynamique de la maladie, à amener à mieux comprendre les épidémies en cours, à aider à la gestion d'urgence et à compléter les pratiques épidémiologiques classiques.

**Mots clés :** choléra, modélisation distribuée, hétérogénéité spatiale, échelles, maladies infectieuses, épidémiologie, interventions, traitement, prédiction, santé publique, mobilité humaine

# Zusammenfassung

Das Verständnis der Epidemiologie der Infektionskrankheit Cholera, d.h. wo und wann sie auftritt und wie sie sich ausbreitet, ist für die Prävention und die Bekämpfung von Ausbrüchen von grosser Wichtigkeit. Mathematische Modelle können zu einem erweiterten Verständnis entscheidender epidemiologischer Prozesse führen und somit dazu beitragen Epidemien besser zu verstehen. Solche Modelle können des Weiteren dazu genutzt werden, verschiedene Bekämpfungsstrategien abzuwägen oder gar den zukünftigen Verlauf von Epidemien vorherzusagen. Das erklärte Ziel dieser Dissertation ist es, die Entwicklung von räumlich verteilten epidemiologischen Modellen von Choleraausbrüchen mittels der Erweiterung bestehender Methoden sowie praktischer Anwendungen voranzutreiben.

Mehr als 160 Jahre nachdem John Snow das räumliche Auftreten von Cholerafällen in London analysiert und erstmals (Trink-)Wasser als den primären Ansteckungsweg identifiziert hat, stellt die Krankheit noch immer eine wesentliche Gefahr für die öffentliche Gesundheit vieler Länder und Regionen rund um den Globus dar. Jährlich erkranken geschätzt 2.86 (1.30 – 4.00) Millionen Personen in den 69 Ländern in denen Cholera endemisch ist. In 95 000 (21 000 – 143 000) Fällen ist der Verlauf tödlich.

Zur Beantwortung epidemiologischer Fragen verschiedener räumlicher Grössenordnungen, von national über regional bis urban, wurde ein Satz mathematischer, auf Metapopulationen oder Individuen basierender, mechanistischer und semi-mechanistischer Modelle entwickelt. Die Modelle schliessen die skalierbare räumliche Heterogenität epidemiologischer Prozesse mit ein, insbesondere die Ausbreitung der Krankheit über das hydrologische Netzwerk und durch die Mobilität der Bevölkerung sowie die hochaufgelöste räumliche und zeitliche Häufung von Cholerafällen. Zudem wurde eine Methode zur Herleitung von Bewegungen der Bevölkerung anhand von Mobiltelefonanrufrufen zur direkten Anwendung in epidemiologischen Modellen entwickelt. Des Weiteren sind verschiedene Umwelt- und Klimafaktoren, die die Ausbreitung von Epidemien beeinflussen, miteinbezogen worden. Die Modelle kamen in kürzlich aufgetretenen Choleraausbrüchen in Haiti, im Senegal, in Chad und in der Demokratischen Republik Kongo zur Anwendung.

Die Resultate verdeutlichen den grossen Einfluss der Mobilität der Bevölkerung, des Niederschlags und anderen klimatologischen Faktoren auf die Ausbreitung der Krankheit. Konkrete Anwendungen zeigen auf, wie solche Modelle zur Verbesserung epidemiologischer Grundsätze und Richtlinien beitragen können und welche Auswirkungen von neuartigen Bekämpfungsstrategien zu erwarten sind. Ein Beispiel stellt die Beurteilung präventiver Strategien dar, im Rahmen derer die Bevölkerung in urbanen, dicht besiedelten Gebieten in einem gewissen

## Zusammenfassung

---

Radius um gemeldete Cholerafälle zu impfen, mit Antibiotika zu behandeln oder mit Trinkwasser, Sanitäreinrichtungen und hygienischer Grundbildung zu versorgen sind. Diese Strategien haben sich als effektiver und effizienter im Vergleich mit Massenimpfungen und andere Masseninterventionen herausgestellt. Des Weiteren wurden die Vorteile einer neuen Art oraler Rehydrationslösung untersucht, wobei sich gezeigt hat, dass diese einen bedeutenden Einfluss auf den Verlauf simulierter Epidemien hat.

Abschliessend führt diese Dissertation aus, dass die räumliche Heterogenität in Modellen ausdrücklich und im richtigen Masstab miteinzuziehen ist, um den realen Verlauf von Choleraausbrüchen nachvollziehen zu können. Sie hebt hervor, inwiefern passende Modelle relevante Fragen zur Dynamik der Krankheit beantworten, Einblicke in andauernde Epidemien verschaffen, dem Notfallmanagement Hilfestellung leisten und die aktuelle epidemiologische Praxis komplementieren können.

**Stichwörter:** Cholera, Modellierung, verteilte Modelle, Skalierbarkeit, räumliche Heterogenität, Infektionskrankheiten, Epidemiologie, Bekämpfung, Behandlung, Vorhersagen, Gesundheitswesen, Bewegungsprofile

# Sommario

La corretta comprensione di quando e dove un'epidemia di colera si verifica, assieme al modo in cui si diffonde, sono elementi essenziali per la sua prevenzione e controllo. L'utilizzo di modelli permette di facilitare il processo di comprensione delle epidemie, fornendo una stima dei processi chiave, consentendo la valutazione delle strategie di intervento e fornendo la previsione di scenari futuri. Questa tesi mira a far progredire l'evoluzione dei modelli spazialmente espliciti di diffusione del colera, attraverso sia sviluppi metodologici che applicazioni pratiche.

Più di 160 anni dopo gli studi di John Snow che per primo analizzò lo sviluppo spaziale dei casi di colera a Londra, identificando l'acqua come principale percorso di contagio, la malattia rimane una grave minaccia per la salute pubblica in molte regioni del mondo. Essa provoca un numero stimato di 2,86 (1,30 – 4,00) milioni di casi e 95 000 (21 000 – 143 000) morti in 69 paesi endemici ogni anno.

Un insieme di modelli epidemiologici di tipo 'metapopulation' e 'individual-based', meccanicistici e semi-meccanicistici viene qui sviluppato per affrontare problemi epidemiologici a scala nazionale, regionale e municipale. I modelli tengono esplicitamente conto della variabilità spaziale dei processi, come la diffusione della malattia attraverso le reti idrologiche o la mobilità umana, o il raggruppamento spazio-temporale dei casi. Inoltre, è stato sviluppato un metodo per estrarre i flussi di mobilità su larga scala a partire da dati di telefonia mobile, con la possibilità di includerli direttamente all'interno dei modelli. Diversi controlli ambientali sulle epidemie sono stati presi in considerazione e i modelli sono stati applicati a recenti epidemie di colera ad Haiti, in Senegal, Ciad e nella Repubblica Democratica del Congo.

I risultati mettono in evidenza l'importante ruolo svolto dalla mobilità umana nella diffusione della malattia e l'influenza delle precipitazioni e altre variabili climatiche come promotori delle dinamiche di contagio. Le applicazioni mostrano come i modelli possono aiutare le politiche in tema epidemiologico e mostrare l'effetto di diverse strategie di intervento sul corso di un'epidemia. Tra gli esempi, viene valutata la ripartizione preventiva del vaccino per via orale, degli antibiotici e/o dell'acqua, dei servizi igienico-sanitari e degli interventi di igiene mirati alle aree densamente popolate, che si dimostrano essere una valida alternativa alle campagne di intervento di massa. Inoltre, vengono illustrati i vantaggi di un tipo alternativo di soluzione reidratante orale, che si è rivelato avere un effetto significativo sull'evoluzione dell'epidemia secondo le simulazioni virtuali.

Questa tesi conclude che il trattamento esplicito delle variabilità spaziali ad una scala adeguata è fondamentale per riprodurre le effettive dinamiche delle epidemie di colera. I risultati

## Sommario

---

evidenziano come un modello appropriato possa affrontare problemi di rilievo sulla dinamica della malattia, fornire informazione sulle epidemie in corso, essere di aiuto nella gestione delle emergenze e complessivamente completare la pratica epidemiologica attualmente in uso.

**Parole chiave:** colera, modelli, spazialmente esplicito, eterogeneità spaziale, malattia infettiva, epidemiologia, interventi, trattamento, previsione, salute pubblica, mobilità umana

# Contents

<b>Acknowledgements</b>	<b>i</b>
<b>Summary</b>	<b>iii</b>
<b>Résumé</b>	<b>v</b>
<b>Zusammenfassung</b>	<b>vii</b>
<b>Sommario</b>	<b>ix</b>
<b>Contents</b>	<b>xi</b>
<b>List of figures</b>	<b>xv</b>
<b>List of tables</b>	<b>xvii</b>
<b>Introduction</b>	<b>1</b>
Cholera . . . . .	1
<i>Vibrio Cholerae</i> and the human host . . . . .	1
Epidemiology . . . . .	3
Treatment and prevention . . . . .	5
Modeling cholera epidemics . . . . .	6
Epidemiological cholera models . . . . .	6
Spatially explicit models . . . . .	8
Environmental drivers . . . . .	10
Spatially explicit models of cholera epidemics and their application . . . . .	11
<b>1 Modeling the Haiti cholera epidemic: the effect of rice-based oral rehydration solution (ORS)</b>	<b>13</b>
1.1 Introduction . . . . .	14
1.2 Methods . . . . .	15
1.2.1 Model setup and data . . . . .	15
1.2.2 Spatially explicit epidemiological model . . . . .	17
1.2.3 Modeling the effect of rice-based ORS in the Haiti epidemic . . . . .	19
1.2.4 Initial condition . . . . .	20
1.2.5 Model calibration and parameter estimation . . . . .	21

## Contents

---

1.3	Results . . . . .	21
1.3.1	Sensitivity to the (a)symptomatic contamination probability . . . . .	23
1.4	Discussion . . . . .	25
1.5	Other studies based on the same model setup . . . . .	28
<b>2</b>	<b>Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks</b>	<b>31</b>
2.1	Introduction . . . . .	32
2.2	Methods . . . . .	35
2.2.1	Study domain and administrative subdivision of Senegal . . . . .	35
2.2.2	Data . . . . .	36
2.2.3	Inference of human mobility patterns . . . . .	37
2.2.4	Spatially explicit epidemiological model . . . . .	38
2.2.5	Initial conditions . . . . .	40
2.2.6	Parameter estimation . . . . .	40
2.2.7	Model selection . . . . .	42
2.2.8	Potential effects of local interventions . . . . .	44
2.3	Results . . . . .	44
2.3.1	Model selection . . . . .	45
2.3.2	Impact of reduced transmission during the GMdT . . . . .	47
2.4	Discussion . . . . .	50
<b>3</b>	<b>The impact of case-centered interventions in response to cholera outbreaks: a modeling study</b>	<b>55</b>
3.1	Introduction . . . . .	56
3.2	Materials and Methods . . . . .	58
3.2.1	Case study and data . . . . .	58
3.2.2	Spatial setup and population distribution . . . . .	58
3.2.3	Rainfall data . . . . .	60
3.2.4	Quantification of spatiotemporal clustering . . . . .	60
3.2.5	Epidemiological model . . . . .	61
3.2.6	Initial conditions . . . . .	64
3.2.7	Calibration . . . . .	64
3.2.8	Simulation . . . . .	65
3.2.9	Intervention strategies . . . . .	66
3.2.10	Allocation strategies . . . . .	69
3.3	Results . . . . .	71
3.3.1	Calibration . . . . .	71
3.3.2	Individual case-centered Interventions . . . . .	71
3.3.3	Combined case-centered Interventions . . . . .	74
3.3.4	Distance range of case-centered interventions . . . . .	74
3.3.5	Efficiency compared to other modes of intervention allocation . . . . .	75
3.3.6	Additional scenarios . . . . .	77



<b>4 Cholera in the Lake Kivu region (DRC): Integrating remote sensing and spatially explicit epidemiological modeling</b>	<b>87</b>
4.1 Introduction . . . . .	88
4.2 Case study . . . . .	89
4.2.1 Spatial setting . . . . .	89
4.2.2 Pathogen transport . . . . .	89
4.2.3 Climate . . . . .	89
4.2.4 Data . . . . .	91
4.3 Model . . . . .	93
4.3.1 Inhomogeneous Markov chain model for endemic cholera . . . . .	93
4.3.2 Model calibration and validation . . . . .	96
4.4 Results . . . . .	97
4.5 Discussion . . . . .	99
<b>Conclusions and perspectives</b>	<b>103</b>
<b>A Appendix</b>	<b>105</b>
A.1 Supplementary Table to Chapter 3 . . . . .	105
A.2 Supplementary Table to Chapter 4 . . . . .	112
<b>Bibliography</b>	<b>115</b>
<b>Curriculum Vitae</b>	<b>137</b>



# List of Figures

1	<i>Vibrio cholerae</i> . . . . .	2
2	Estimated annual number of cholera cases in endemic countries . . . . .	3
3	Processes involved in model development and application. . . . .	7
1.1	Geography of Haiti and model setup. . . . .	16
1.2	Schematic representation of the model. . . . .	18
1.3	Evolution of the epidemic . . . . .	23
1.4	Evolution of the epidemic per department . . . . .	25
1.5	Reduction of the number of cases when reducing shedding rate and disease duration. . . . .	26
1.6	Evolution of modeled susceptibles and symptomatically infected. . . . .	26
1.7	Evolution of the epidemic using different values for $q_A/q_S$ . . . . .	27
1.8	Total cases up to December 2011 using different values for $q_A/q_S$ . . . . .	28
2.1	Population density . . . . .	36
2.2	Daily precipitation depth. . . . .	37
2.3	Mobile phone derived mobility estimates . . . . .	38
2.4	Reconstruction of the 2005 mobility matrix. . . . .	43
2.5	Reported and modeled cases over time. . . . .	45
2.6	Spatial distribution of reported and modeled cases. . . . .	47
2.7	Reported and modeled cases in all departments. . . . .	48
2.8	Model performance using different mechanisms. . . . .	49
2.9	Total averted cases overall Senegal with hypothetical interventions. . . . .	51
3.1	Map of cholera cases. . . . .	59
3.2	Built-up density. . . . .	59
3.3	Daily precipitation depth in N'Djamena from April 2011 to April 2012. . . . .	60
3.4	Comparison of relative risk $\tau$ evaluated in 15m distance ranges for different time ranges. . . . .	61
3.5	Schematic representation of the epidemiological model and evolution of the infectious state of inhabitants of a neighborhood. . . . .	62
3.6	Calibration results. . . . .	65
3.7	Distributions of intervention parameters. . . . .	66
3.8	Number of people living within a given range in N'Djamena. . . . .	70

## List of Figures

---

3.9	Circles of given radius (in m) as implemented in the regular model grid with. . .	70
3.10	Posterior parameters. . . . .	72
3.11	Comparison of the simulated evolution of the epidemics with and without case-centered interventions. . . . .	72
3.12	Outcome of the three main interventions with case-centered allocation in a 100 m radius. . . . .	74
3.13	Case-centered intervention within a radius of 100 m combining antibiotics and OCV . . . . .	75
3.14	Simulated evolution of the epidemics without intervention and with case-centered allocation of combination of the three main intervention types within a 100 m radius starting around the epidemic peak. . . . .	76
3.15	Outcome of combinations the three main intervention types with case-centered allocation in a 100 m radius. . . . .	77
3.16	Intervention outcomes as a function of distance in case-centered allocations. . .	78
3.17	Efficiency (persons to treat per case averted) of different modes of intervention allocation. . . . .	79
3.18	Outcome of the three main intervention types in a mass intervention campaign randomly targeting 70% of the population. . . . .	79
3.19	Outcome of the three main intervention types with random allocation. . . . .	80
3.20	Intervention outcomes of case-centered allocation of OCV within a varying radius and antibiotics within a radius of 15 m. . . . .	81
3.21	Comparison between single and repeated antibiotics administration through case-centered allocation in a 100 m radius. . . . .	82
3.22	Comparison between a 7- and 2-days lag between the administration of a single-dose of OCV and the onset of protection in a case-centered allocation framework in a 30 m radius. . . . .	83
4.1	Study domain: the Lake Kivua area. . . . .	90
4.2	Cholera incidence data. . . . .	92
4.3	Environmental drivers. . . . .	93
4.4	Simulations from the best performing model. . . . .	100

# List of Tables

1	Serogroups, serotypes, biotypes and cholera toxin (CT) production by the species <i>V. cholerae</i> . (Adapted from <i>Kaper et al.</i> [1995] and <i>Nelson et al.</i> [2009].) . . . . .	1
2	Summary of a selection of recently published applied cholera models. (Table adapted from <i>Finger et al.</i> [2014]). . . . .	9
1.1	Model parameters . . . . .	22
1.2	Number of calibrated parameters and AIC scores. . . . .	25
2.1	Regions of Senegal, population and number of cases. . . . .	35
2.2	Fixed (top) and calibrated (bottom) parameters of the best performing model (A in Table 2.3). For the latter the 95% confidence intervals of the posterior distribution are also shown. . . . .	41
2.3	Model comparison using different mechanisms. . . . .	46
2.4	Recalibration of model A using different mobility matrices (Section 2.2.7 and Figure 2.4). . . . .	50
2.5	Total averted cases overall Senegal with hypothetical interventions. . . . .	52
3.1	Effects of the three types of interventions as implemented in the model. . . . .	67
3.2	Distribution of the delay between the reporting of the initial case in a cluster and the deployment of an intervention team. . . . .	70
4.1	Results of the fitting procedure. . . . .	98
4.2	Parameter sets corresponding to the best ranked models . . . . .	99
4.3	Log-likelihood values obtained during validation by comparing one, two or three months in advance simulations with reported incidence. . . . .	101
A.1	Summary of all intervention scenarios. . . . .	105
A.2	Results of the fitting procedure ordered by increasing AIC score. . . . .	112



# Introduction

## Cholera

### *Vibrio Cholerae* and the human host

Cholera is a waterborne infectious disease caused by the bacterium *Vibrio cholerae* (Figure 1) and may lead to severe diarrhea and serious dehydration [e.g. *Kaper et al.*, 1995; *Nelson et al.*, 2009]. The species *V. cholerae* is subdivided into serogroups, serological subtypes and biotypes (Table 1). Strains that have the potential to cause epidemic cholera and thus are of public health significance belong to serogroups O1 or O139 and produce cholera toxin (CT), the main virulence factor [*Kaper et al.*, 1995; *Nelson et al.*, 2009]. They are thus said to be toxigenic. Non-toxigenic strains of *V. cholerae*, mostly belonging to other serogroups, may cause disease, but are not associated with epidemic diarrhea [*Kaper et al.*, 1995]. Toxigenic and non-toxigenic strains can be found in the environment and belong to the bacterial flora of estuaries (see below), with non-toxigenic strains being more abundant [*Colwell and Spira*, 1992; *Colwell et al.*, 1977; *Gil et al.*, 2004; *Hill et al.*, 2011; *Huq et al.*, 1983; *Kaper et al.*, 1995; *Roberts et al.*, 1982; *Tamplin et al.*, 1990].

The ingestion of an infectious dose of *V. Cholerae* ( $10^8$  to  $10^{11}$  pathogens in healthy North American volunteers [*Nelson et al.*, 2009]) can lead to a variety of symptoms, with severity ranging from asymptomatic colonization to *cholera gravis*, marked by severe diarrhea, vomiting, fluid loss and dehydration, which may result in death [*Kaper et al.*, 1995; *Nelson et al.*,

Table 1 – Serogroups, serotypes, biotypes and cholera toxin (CT) production by the species *V. cholerae*. (Adapted from *Kaper et al.* [1995] and *Nelson et al.* [2009].)

Serogroup	CT production	Epidemic spread	Serological subtypes (no.)	Biotypes (no.)
O1	Yes	Yes	3 (Inaba, Ogawa, Hikojima)	2 (classical, El Tor)
O2–O138	No	No	None	1
O139	Yes	Yes	None	1
O140–O200	No	No	None	1

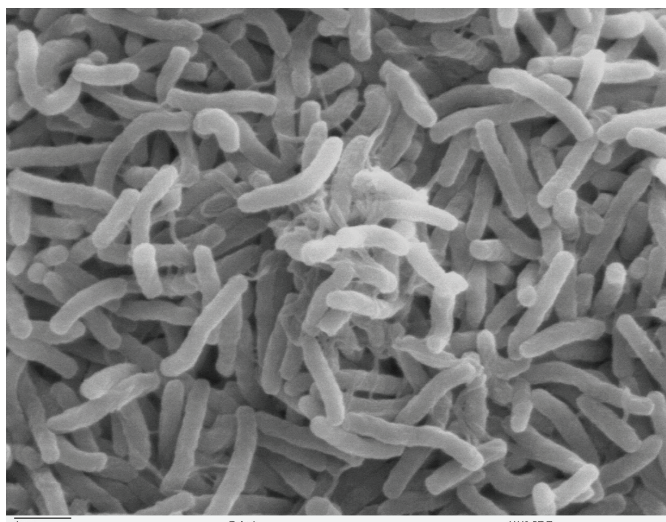


Figure 1 – Scanning electron microscope image of *Vibrio cholerae* bacteria. (Ronald Taylor, Tom Kirn, Louisa Howard. <http://remf.dartmouth.edu/imagesindex.html> (accessed on December 6, 2016). Image in public domain.)

2009]. Only a minority of infections are severely symptomatic (i.e. lead to *cholera gravis*), with the proportion varying between strains [Kaper *et al.*, 1995], age and endemicity of the disease [Nelson *et al.*, 2009]. Estimates of the proportion of asymptomatic cases are highly uncertain and range from 3 to 100 asymptomatic per symptomatic infection [Kaper *et al.*, 1995; King *et al.*, 2008; McCormack *et al.*, 1969; Van de Linde and Forbes, 1965]. The incubation period is estimated to be around 1.4 days (90% Confidence Interval (CI) 0.5 to 4.4), depending on the strain [Azman *et al.*, 2013] and the ingested bacterial dose [Hornick *et al.*, 1971; Kaper *et al.*, 1995]. Stool shed by severely symptomatic individuals typically contains  $10^{10}$  to  $10^{12}$  vibrios per liter [Nelson *et al.*, 2009] and shedding can last for 1-2 weeks [Kaper *et al.*, 1995]. Mildly and asymptotically infected individuals shed significantly less and for a shorter duration [Nelson *et al.*, 2009]. Bacteria can then infect a new host via the fecal-oral route, mainly mediated by contaminated water, or may enter the aquatic environment. A hyperinfectious state, which causes freshly shed vibrios to be more virulent, is believed to exist, its epidemiological relevance is subject to current research [Butler *et al.*, 2006; Merrell *et al.*, 2002; Nelson *et al.*, 2009]. *V. Cholerae* can persist in aquatic environments and can associate with different forms of plankton and biofilms [Faruque *et al.*, 2006; Huq *et al.*, 1983; Islam *et al.*, 2007; Nelson *et al.*, 2009; Tamplin *et al.*, 1990]. It is known that periodic introduction of aquatic *V. cholerae* into human populations may cause outbreaks [Kaper *et al.*, 1995; Nelson *et al.*, 2009; Rebaudet *et al.*, 2013a; Vezzulli *et al.*, 2010], long-term persistence of the disease in certain areas might however also result from infected humans as a reservoir [Bompangue *et al.*, 2011; Rebaudet *et al.*, 2013c]. Severely symptomatic patients acquire a high level of immunity to subsequent infections, with protection possibly lasting for 3 years or more [Kaper *et al.*, 1995; Levine *et al.*, 1981; Nelson *et al.*, 2009], whereas mild and inapparent infections may also confer immunity,



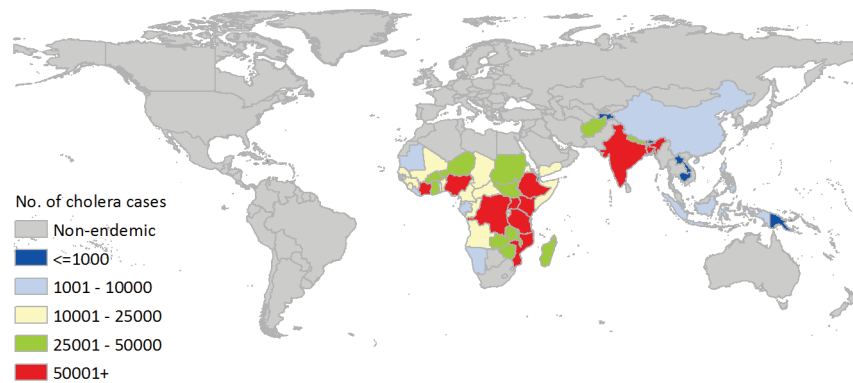


Figure 2 – Annual number of cholera cases in endemic countries. (reproduced from *Ali et al.* [2015] with permission). Note that Haiti is hardly visible due to its size, but the estimated number of annual cholera cases there is above 50 000.

which is thought to be much shorter [*King et al.*, 2008].

## Epidemiology

In 2015, a total of 172 454 cholera cases were reported in 42 countries, 41% of which in Africa, 37% in Asia and 21% in Hispaniola (Central American island) [*World Health Organization*, 2016a]. In the same period, 1304 deaths were reported [*World Health Organization*, 2016a]. Those numbers only include cases that have been reported by national surveillance systems, actual figures are believed to be orders of magnitude higher. The worldwide annual number of cholera cases has been estimated to lie between 1.3 and 4 million in 69 cholera-endemic countries, with an estimated number of deaths between 21 000 and 143 000 (Figure 2) [*Ali et al.*, 2015]. Areas with a particularly high burden of cholera are Sub-Saharan Africa, the Indian subcontinent, and Haiti [*Ali et al.*, 2015; *World Health Organization*, 2016a].

Cholera is endemic in countries adjacent to the Bay of Bengal, and particularly in Bangladesh, where it occurs periodically with a clear seasonal pattern [*Glass et al.*, 1982; *Kaper et al.*, 1995]. *V. cholerae* is believed to be autochthonous in the estuarine, brackish-water ecosystems in this region [*Colwell and Spira*, 1992], the environmental reservoir playing a part in the seasonal revamping of epidemics and possibly even contributing to sustain transmission during epidemics [*Colwell*, 1996; *King et al.*, 2008; *Vezzulli et al.*, 2010]. Other major factors leading to seasonal patterns in the region are thought to be environmental drivers of cholera transmission, such as rainfall, floods, river discharge and monsoon [e.g. *Bertuzzo et al.*, 2012; *Bouma and Pascual*, 2001; *Koelle et al.*, 2005; *Pascual et al.*, 2002]. Endemicity of cholera in the region leads to high immunity within the population, which results in an age distribution of cases skewed towards young ages [*Kaper et al.*, 1995; *King et al.*, 2008].

Since the 1800s cholera spread around the globe in several pandemics, all originating on the

## Introduction

---

Indian subcontinent and in particular the delta of the Ganges, Meghna and Brahmaputra rivers in modern Bangladesh [Kaper *et al.*, 1995]. Recent phylogenetic analyses have shown that the current (seventh) pandemic took its origin in Bangladesh in 1961 from where it reached South-East Asia, the Middle-East and Africa, and finally spread to South-America, probably through infected travelers [Chun *et al.*, 2009; Kaper *et al.*, 1995; Mutreja *et al.*, 2011; Valia *et al.*, 2013]. This pandemic is caused by the El Tor biotype of serogroup O1, but variants combining features of classical and El Tor strains have also been observed (Table 1) [Longini *et al.*, 2002; Nelson *et al.*, 2009; Valia *et al.*, 2013].

The current cholera pandemic first got introduced to Africa in 1970. Since then, several re-introduction events have happened [Kaper *et al.*, 1995; Mutreja *et al.*, 2011]. It has become endemic in more than 40 countries in Sub-Saharan Africa [Ali *et al.*, 2015] and is present in coastal as well as in inland regions, exhibiting complex dynamics with inland regions responsible for three quarters of cases between 2009 and 2011 [Rebaudet *et al.*, 2013a, b]. In coastal areas, estuaries, lagoons and mangrove forests seem to play a major part in the disease dynamics, together with large coastal cities [Rebaudet *et al.*, 2013a], whereas in inland regions many major outbreaks have happened in regions adjacent to lakes and rivers [Bompangue *et al.*, 2008; Rebaudet *et al.*, 2013b]. Although many cholera outbreaks have been associated to coastal and lakeside areas, the persistence of toxigenic *V. cholerae* in the environment of the African continent outside outbreak periods has not been proven, which increases the plausibility of the alternative hypothesis of the human host as the principal reservoir [Rebaudet *et al.*, 2013a, b]. Sanitary condition and access to clean drinking water have been shown to be key factors in both settings. Climate driven seasonality and interannual fluctuations have been observed in both, coastal and inland Africa. Disease spread within and among countries is thought to happen mostly through human movement [Rebaudet *et al.*, 2013a, b].

Cholera is known to cause explosive outbreaks when reaching new geographic areas inhabited by cholera-naïve populations, i.e. populations with no prior exposure or acquired immunity to the pathogen, such as Peru in the early 1990s [Kaper *et al.*, 1995; Levine, 1991; Ries *et al.*, 1992; Swerdlow *et al.*, 1992]. In Haiti, the disease has been introduced in October 2010 [Chin *et al.*, 2011; Frerichs *et al.*, 2012; Hendriksen *et al.*, 2011; Piarroux, 2011], leading to one of the largest cholera outbreaks in recent years. The outbreak started in the valley of the Artibonite river, before expanding to the entire country within months (Section 1.1). Since then, outbreaks alternating with lull phases have been recorded every year [Rebaudet *et al.*, 2013c]. Rainfall has been identified as an important driver [Eisenberg *et al.*, 2013; Gaudart *et al.*, 2013]. A major role of environmental reservoirs in persistence of the disease has, to date, not been confirmed, even if environmental sampling campaigns have led to increasing numbers of isolates of toxigenic *V. cholerae* over time [Alam *et al.*, 2015, 2016, 2014; Azarian *et al.*, 2016; Morris, 2016; Rebaudet *et al.*, 2013c]. As of September 25, 2016, the Ministry of Public Health and Population had reported more than 790 000 cases since the introduction [Ministère de la santé publique et de la population (MSPP), 2016], which corresponds to an average cumulative attack rate of around 8%. With more than 9400 deaths, the resulting cumulative case fatality rate is 1.2%. In October and November 2016, an important outbreak has been recorded after

Haiti had been severely hit by hurricane Matthew, with a total of 8916 cases according to the Panamerican Health Organization ([http://ais.paho.org/hip/viz/ed\\_haiticoleracases.asp](http://ais.paho.org/hip/viz/ed_haiticoleracases.asp), accessed on December 8, 2016) [Ivers, 2016a].

### Treatment and prevention

The general treatment of cholera patients consists in replacing lost fluids and electrolytes through the administration of intravenous or oral rehydration solutions (ORS). The majority of patients, those with mild or moderate symptoms, are generally treated using so-called oral rehydration therapy (ORT). For the latter purpose, the administered solution contains a mixture of several compounds that were designated oral rehydration salts, including sodium, chloride, and potassium ions as well as a carbon source (e.g. glucose) [Kaper *et al.*, 1995; Nelson *et al.*, 2009]. More severe cases can be treated with antimicrobial agents, which shorten the duration of the diarrhea as well as the bacterial shedding, and reduce the severity of the disease [Kaper *et al.*, 1995; Leibovici-Weissman *et al.*, 2014; Nelson *et al.*, 2009, 2011; Reveiz *et al.*, 2011].

As for any disease spread via the fecal-oral route, providing access to clean drinking water, appropriate sanitation facilities as well as education about hygiene practices are key to sustainable cholera prevention. The number and diversity of possible water, sanitation and hygiene (WaSH) interventions is large and ranges from teaching people to wash their hands with soap over the provision of sewage treatment facilities to providing water supply infrastructure or promoting point-of-use water treatment [Fewtrell and Colford, 2004; Fewtrell *et al.*, 2005]. The choice of the most appropriate type of intervention for a given situation is complex and depends on many factors, such as e.g. available resources and time, the socio-economical situation of the target population or the expected short- and long-term benefits. Their exploration is beyond the scope of this thesis. Reactive campaigns, i.e. with the goal to halt or slow down ongoing cholera outbreaks, in most cases with little resources available, are often limited to providing drinking water or means to make available water safe to drink, e.g. through chlorination, as well as educating people about hygiene behavior and the ways cholera spreads [Farmer *et al.*, 2011; Tappero and Tauxe, 2011; Walton and Ivers, 2011].

Recent developments of new oral cholera vaccines (OCV) added another, complementary and cost-effective option to the quiver of cholera prevention [Desai *et al.*, 2016; Mogasale *et al.*, 2016]. They can be used in preemptive (i.e. before an outbreak) or reactive (i.e. after the start of an outbreak) vaccination campaigns [Azman *et al.*, 2015]. According to vaccine trials in Kolkata, India, they provide a vaccine efficacy higher than 65% during at least 5 years [Bhattacharya *et al.*, 2013; Sur *et al.*, 2009, 2011] in addition to significant herd immunity [Ali *et al.*, 2013b]. Current OCVs have been licensed for two-dose schedules, the two doses being administered with an interval of two weeks. The administration of a single-dose, however, in addition to doubling the number of people who can be targeted with a given number of vaccines, has been shown to have logistical advantages [Azman *et al.*, 2015]. In vaccine trials, a

## Introduction

---

single-dose of OCV has been shown to provide a vaccine efficiency comparable to two doses up to at least 6 months after administration [Azman *et al.*, 2016; Qadri *et al.*, 2016]. In 2013, a global stockpile of OCV has been established in order to make the vaccine more easily available [Desai *et al.*, 2016; Martin *et al.*, 2012]. The stockpile was successfully used for the first time in South Sudan in 2014 [Abubakar *et al.*, 2015], and subsequently, among other outbreaks, in Haiti in 2016 [Ivers, 2016b; World Health Organization, 2016b].

## Modeling cholera epidemics

Epidemiological models of infectious diseases can be defined as a simplified formulation of epidemiological processes in mathematical terms, with the goal to simulate epidemics and to gain insights about their real-world counterparts. Simulating epidemics by the means of a model, at the disadvantage that processes must be simplified, can have several benefits. Other than in the real-world, all processes and variables can be observed in time and space, which allows to gain critical insights into the course of an epidemic. Simulations can be repeated under the same or different conditions and/or with different parameters, aiming at identifying the influence of incorporated stochasticity or altered conditions. In practical terms, the possible goals of epidemiological models can be formulated as follows: 1) understanding the influence of different epidemiological processes on the model outcome, usually the number of cases over time, and thus determining which processes are relevant in a particular case, 2) understanding the way altered conditions, such as those created by various interventions, influence epidemics, 3) predicting the future course of epidemics (Figure 3). Model parameters have to be calibrated in order to match the desired epidemiological characteristics. This is often done using data from past outbreaks. Usually models are validated to ensure that they do not only reproduce the particular epidemiological datasets which they have been calibrated to, but rather represents a generalization of epidemiological processes with wider applicability. The three processes of model formulation, calibration and validation can be repeated until the predictive power of the model during validation is satisfactory and thus the model formulation can be considered to reproduce the relevant epidemiological processes (Figure 3).

## Epidemiological cholera models

In the literature, two main modeling approaches have been followed. The first consists of predictive empirical models relying on environmental drivers which influence the ecology of *V. cholerae* [Bouma and Pascual, 2001; Lipp *et al.*, 2002; Matsuda *et al.*, 2008; Pascual *et al.*, 2002; Ruiz-Moreno *et al.*, 2007], often using remotely-sensed information [Akanda *et al.*, 2009, 2013; de Magny *et al.*, 2008; Ford *et al.*, 2009; Jutla *et al.*, 2010, 2013a, b; Leckebusch and Abdussalam, 2015; Lobitz *et al.*, 2000; Ngwa *et al.*, 2016]. Such methods, suited in particular to regions where cholera is endemic, but applied to predict other infectious disease outbreaks as well [Ford *et al.*, 2009], have been shown to relate changes in climatological and biological variables to interannual and annual cyclic patterns of infections [de Magny *et al.*, 2008; Emch *et al.*,

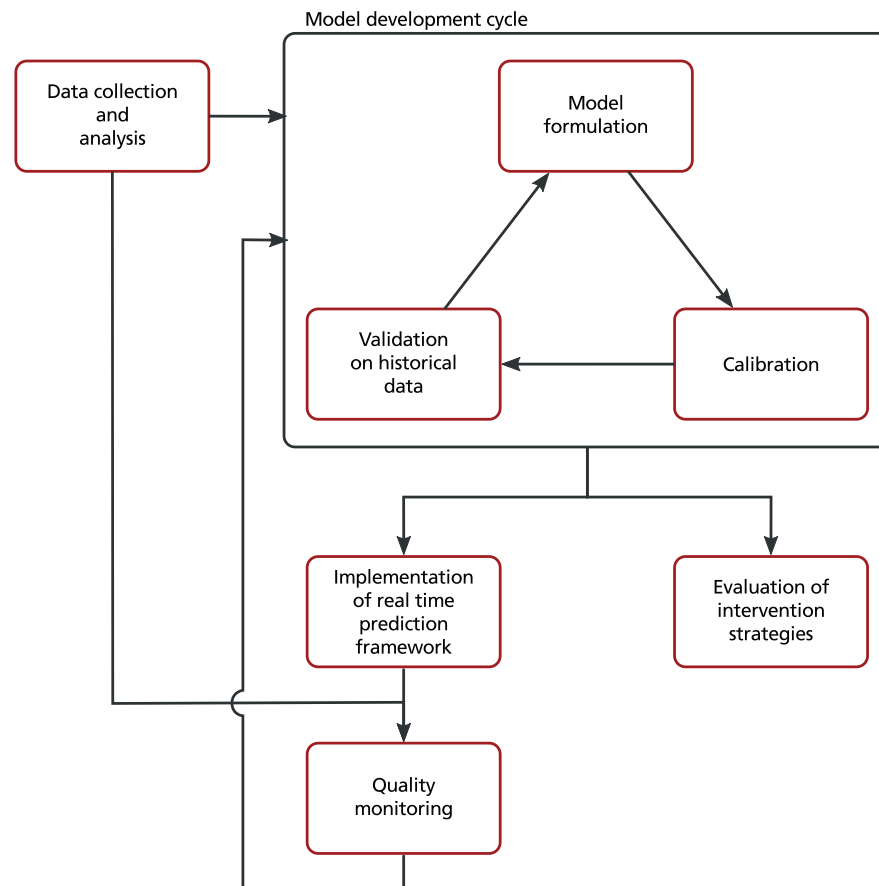


Figure 3 – Processes involved in model development and application.

2008; Jutla *et al.*, 2013b; Matsuda *et al.*, 2008]. For cholera, such variables often consist of chlorophyll *a* (a proxy for plankton concentration, which in turn may promote growth of *V. cholerae* (see above), sea surface temperature (SST), sea surface height, precipitation, air temperature and other covariates of global and local climatic conditions (see below). Predictive empirical models have contributed new epidemiological perspectives, however, their use for the understanding, prediction and control of waterborne disease outbreaks is challenged by the fact that they do not explain the underlying ecological and epidemiological mechanisms but merely represent a probabilistic link between observed environmental variables and reported cases.<sup>1</sup>

Another type of approach relies on mechanistic models of disease spread, either deterministic or stochastic. They often rely on metacommunity based approaches, usually called SIR-models, where a population is subdivided into compartments keeping track of the number of susceptibles, infected and recovered over time by the means of a system of differential

<sup>1</sup>This paragraph has been adapted from Finger *et al.* [2014].

## Introduction

---

equations [Anderson and May, 1992]. Other compartments, for example one describing the bacterial concentration in the local environment [Codeço, 2001], can be added. Note that the addition of a bacterial compartment does not necessarily mean that cholera transmission has to exclusively follow an environmental route through natural water bodies in the sense of Colwell [1996], as a short environmental route through the contamination of stored water, food, hands, surfaces etc. always exists for fecal-oral diseases [Morris, 2016]. In so-called individual- or agent-based models, instead, the epidemiological status (e.g. susceptible, infected, recovered) of every individual is followed over time [Arifin, 2016]. In addition, models can be categorized by the way they address spatial heterogeneity (see below). They can give key insights into the course of an ongoing epidemic [Bertuzzo et al., 2016], explain processes influencing past outbreaks [Ali et al., 2013a; Azman et al., 2012; Tien et al., 2011], provide predictive frameworks [Abrams et al., 2013; Koepke et al., 2016; Pasetto et al., 2016] and potentially aid emergency management in allocating health care resources, also by anticipating the impact of alternative interventions [Azman et al., 2015; Kim et al., 2016; Lewnard et al., 2016]. A selection of recent applied cholera models and their main characteristics is presented in Table 2.<sup>2</sup>

### Spatially explicit models

In mechanistic approaches, modeling an entire country as a single spatial unit means implicitly assuming homogeneous mixing of the whole population, i.e. every person can be infected by every other person in the country with equal probability, which may lead to important shortcomings and invalidate model results. In addition, key factors such as the epidemiological status of the population, environmental conditions and climatic drivers may be heterogeneously distributed in space [Gatto et al., 2012, 2013; Grad et al., 2012].

A spatially explicit approach, made possible by the now widespread access to geographically distributed data of reported cases, population distribution, transportation infrastructure, environmental and hydrological drivers can overcome those issues by incorporating their heterogeneity in the model and, in addition, address the spatiotemporal evolution of disease propagation. In particular, spatial phenomena are bound to become fundamental when the disease propagation through networks of human mobility or hydrologic connectivity is to be described. In metacommunity models this can be achieved by subdividing a population into compartments by geographical location and linking those compartments according to the networks. For human mobility, this can be done by allowing the exchange of susceptibles and infected between locations, for example through a gravity model [Erlander and Stewart, 1990; Mari et al., 2012a; Rinaldo et al., 2012]. Hydrologic connectivity can be incorporated by allowing *V. cholerae* of the bacterial compartment to be propagated to downstream communities [Bertuzzo et al., 2008; Rinaldo et al., 2012]. Whereas the optimal spatial scale for a model should ideally be selected to match the heterogeneity of underlying processes (e.g. neighborhoods, cities, districts, departments, countries, regions), in practice the choice is

---

<sup>2</sup>This paragraph has been adapted from Finger et al. [2014].

Table 2 – Summary of a selection of recently published applied cholera models. (Table adapted from *Finger et al.* [2014]).

	Region	Empirical	Mechanistic	Implicit <sup>c</sup>	Explicit <sup>c</sup>	Data-driven <sup>d</sup>	Simulated <sup>d</sup>	None <sup>d</sup>
<i>Lobitz et al.</i> [2000]	BoB	×		×		×		
<i>Pascual et al.</i> [2000]	BoB <sup>a</sup>	×		×			×	
<i>Koelle et al.</i> [2005]	BoB <sup>a</sup>		×	×			×	
<i>Longini et al.</i> [2007]	BoB <sup>a</sup>		×	×				×
<i>de Magny et al.</i> [2008]	BoB <sup>a</sup>	×		×		×		
<i>Bertuzzo et al.</i> [2008]	S. Africa		×		×		×	
<i>Fernández et al.</i> [2009]	Zambia	×		×		×		
<i>King et al.</i> [2008]	BoB <sup>a</sup>		×	×			×	
<i>Matsuda et al.</i> [2008]	BoB <sup>a</sup>	×		×		×		
<i>Pascual et al.</i> [2008]	BoB <sup>a</sup>		×	×			×	
<i>Akanda et al.</i> [2009]	BoB <sup>a</sup>	×		×		×		
<i>Islam et al.</i> [2009]	BoB <sup>a</sup>	×		×		×		
<i>Andrews and Basu</i> [2011]	Haiti		×	×				×
<i>Bertuzzo et al.</i> [2011]	Haiti		×		×			×
<i>Chao et al.</i> [2011]	Haiti		×		×			×
<i>Mukandavire et al.</i> [2011]	Zimbabwe		×	×				×
<i>Reyburn et al.</i> [2011]	Zanzibar	×		×		×		
<i>Tuite et al.</i> [2011]	Haiti		×		×			×
<i>Tien et al.</i> [2011]	Great Britain		×	×			×	
<i>Mari et al.</i> [2012b]	S. Africa		×		×			×
<i>Gatto et al.</i> [2012]	Haiti		×		×		×	
<i>Rinaldo et al.</i> [2012]	Haiti		×		×	×		
<i>Reiner et al.</i> [2012]	BoB <sup>a</sup>		×	×			×	
<i>Azman et al.</i> [2012]	Guinea-Bissau		×		×			×
<i>Eisenberg et al.</i> [2013]	Haiti	×	×		×			×
<i>Jutla et al.</i> [2013b]	BoB <sup>a</sup>	×		×		×		
<i>Jutla et al.</i> [2013a]	BoB <sup>a</sup>	×		×		×		
<i>Mukandavire et al.</i> [2013]	Haiti		×	×				×
<i>Righetto et al.</i> [2013]	Haiti		×		×	×		
<i>Akanda et al.</i> [2013]	BoB <sup>a</sup>	×		×		×		
<i>Abrams et al.</i> [2013]	Haiti		×		×			×
<i>Ali et al.</i> [2013a]	BoB <sup>a</sup>	×		×		×		
<i>Sardar et al.</i> [2013]	Zimbabwe		×	×				×
<i>Kühn et al.</i> [2014]	Haiti		×		×	×		

	Region	Empirical	Mechanistic	Implicit <sup>c</sup>	Explicit <sup>c</sup>	Data-driven <sup>d</sup>	Simulated <sup>d</sup>	None <sup>d</sup>
<i>Finger et al.</i> [2014]	DRC <sup>b</sup>		×		×	×		
<i>Mari et al.</i> [2015a]	Haiti		×		×	×		
<i>Leckebusch and Abdussalam</i> [2015]	Nigeria	×		×		×		
<i>Kirpich et al.</i> [2015]	Haiti		×	×		×		
<i>Azman et al.</i> [2015]	(multiple)		×	×				×
<i>Ngwa et al.</i> [2016]	Cameroon	×		×		×		
<i>Pasetto et al.</i> [2016]	Haiti		×		×	×		
<i>Finger et al.</i> [2016]	Senegal		×		×	×		
<i>Bertuzzo et al.</i> [2016]	Haiti		×		×	×		
<i>Koepke et al.</i> [2016]	BoB <sup>a</sup>		×	×		×		
<i>Lewnard et al.</i> [2016]	Haiti		×		×			×
<i>Baracchini et al.</i> [2016]	BoB <sup>a</sup>		×		×	×		

<sup>a</sup> Bay of Bengal

<sup>b</sup> Democratic Republic of the Congo

<sup>c</sup> Attributes refer to the treatment of space by the models. Explicit treatment of space means that the model incorporates terms for the spatial spread of disease vehiculed by human mobility, hydrologic connectivity, etc.

<sup>d</sup> Attributes describing the treatment of environmental forcings. Simulated refers to cyclicity based on seasonal and interannual patterns.

often limited by the spatial resolution of the available data. Data from different sources may have to be aggregated or disaggregated in order to match the resolution of the model. <sup>3</sup>

### Environmental drivers

Climate variables that have been related to cholera can be categorized into global and local, where global variables may mediate local ones [*de Magny et al.*, 2006]. Global scale climate phenomena, such as the El Niño Southern Oscillation (ENSO), have been reported to influence cholera dynamics mostly in endemic regions, such as the Bay of Bengal [*Colwell*, 1996; *Koelle et al.*, 2005; *Lipp et al.*, 2002; *Pascual et al.*, 2000, 2008; *Rodó et al.*, 2002], but also in African countries [*Bompangue et al.*, 2011; *de Magny et al.*, 2006; *Olago et al.*, 2007; *Rebaudet et al.*, 2013a, b]. However, the causative mechanistic links between ENSO and disease dynamics remains controversial to date. Some authors pointed out correlations between cholera incidence and suitably delayed phyto- and zooplankton abundances in coastal waters, which in turn are driven by local climate anomalies, mediated by ENSO [*Bompangue et al.*, 2011; *de Magny*

<sup>3</sup>This paragraph has been adapted from *Finger et al.* [2014].



*et al.*, 2008; *Ford et al.*, 2009; *Jutla et al.*, 2010, 2013a; *Mishra et al.*, 2011]. This hypothesis thus places great importance on the role of aquatic environmental reservoirs in maintaining the disease. Others argue that the local climate anomalies caused by ENSO may as well influence disease dynamics via other pathways, such as precipitation, droughts, water salinity, or human behavior and population dynamics [*Pascual et al.*, 2002; *Rebaudet et al.*, 2013a; *Rodó et al.*, 2002].<sup>4</sup>

While the role of local climatic conditions, rainfall, air temperature and sea surface temperature (SST) on patterns of cholera transmission has long been studied, especially in empirical frameworks [*Altizer et al.*, 2006; *de Magny et al.*, 2008; *Koelle et al.*, 2005; *Leckebusch and Abdussalam*, 2015; *Lipp et al.*, 2002; *Ngwa et al.*, 2016], spatially explicit mechanistic models of cholera epidemics have incorporated hydroclimatological drivers only more recently, most notably in models used to study the course of the Haitian epidemic, starting from the very first months after its outbreak in late 2010 and following disease resurgence (May 2011) in connection with unusually intense tropical rains [*Eisenberg et al.*, 2013; *Gaudart et al.*, 2013; *Righetto et al.*, 2013; *Rinaldo et al.*, 2012]. Possible mechanisms of enhanced cholera spread due to heavy rains include increased bacterial concentration in drinking water due to failure of sanitation systems, washout of open-air defecation sites [*Gaudart et al.*, 2013; *Rinaldo et al.*, 2012], or the modification of human water sources and human behavior [*Gaudart et al.*, 2013].<sup>4</sup>

## Spatially explicit models of cholera epidemics and their application

As highlighted in the previous section, models can help to understand epidemiological processes and may be used to evaluate alternative intervention strategies or to predict the future course of epidemics. This thesis aims at advancing the evolution of spatially explicit epidemiological models of cholera outbreaks through methodological developments and practical applications. It is organized in four chapters, each of them describing an epidemiological model best suited for the application to a practical epidemiological problem in a given region.

In Chapter 1, a spatially explicit metapopulation model is developed and applied to the first year of the cholera outbreak in Haiti. The model includes additional epidemiological processes, such as the differentiation between symptomatic and asymptomatic infection. It is used to evaluate the potential population-level impact of changing the type of oral rehydration solution employed, which has previously been shown to have a significant effect at the individual level.

In Chapter 2, the problem of getting accurate information about large-scale population movements for a spatially explicit metapopulation model is tackled via a new approach consisting in directly incorporating human mobility estimates derived from mobile phone call records. The method is applied to a recent outbreak in Senegal, where cholera spread from a single

---

<sup>4</sup>This paragraph has been adapted from *Finger et al.* [2014].

## Introduction

---

district to the entire country within a few days through important population movements before and after a mass gathering.

In Chapter 3, a stochastic, agent-based, spatially explicit model at the scale of a single city (N'Djamena, Chad) is developed. Every individual is assigned a position within the model space according to the population distribution. The model is calibrated to match the spatiotemporal clustering of cases repeatedly observed in cholera outbreaks. Subsequently, it is used to evaluate the benefits of novel, case-centered intervention strategies, which aim to provide preventive interventions, such as WaSH, antibiotics or oral cholera vaccine, to people living within a given radius around reported cases.

Chapter 4 aims at evaluating the drivers of seasonality of cholera in the Lake Kivu Region, Democratic Republic of the Congo, by using a semi-mechanistic model which, in addition to environmental forcings such as precipitation, chlorophyll *a* concentration in the lake or ENSO, incorporates a mechanistic description of human mobility.

A set of conclusions and perspectives for further research in this field closes the thesis.

# 1 Modeling the Haiti cholera epidemic: the effect of rice-based oral rehydration solution (ORS)

*The main analysis presented in this chapter has been published in PLOS Neglected Tropical Diseases as a part of Kühn et al. [2014]. The article consists of two main parts: an experimental analysis of the molecular mechanisms related to the performance of rice-based ORS, and a modeling analysis to estimate the effect of this alternative treatment on large-scale outbreaks. The latter was designed by Flavio Finger, Enrico Bertuzzo, Lorenzo Mari, Marino Gatto and Andrea Rinaldo and led by Flavio Finger, who implemented and calibrated the model, and produced and described the results. The same model setup has also been used in other publications co-authored by Flavio Finger [Bertuzzo et al., 2016; Mari et al., 2015a; Pasetto et al., 2016], from which certain text passages of this chapter have been adapted.*

Kühn, J., F Finger, E. Bertuzzo, S. Borgeaud, M. Gatto, A. Rinaldo, and M. Blokesch, Glucose- but not rice-based oral rehydration therapy enhances the production of virulence determinants in the human pathogen *Vibrio cholerae*, *PLoS Neglected Tropical Diseases*, 8(12), e3347, doi: 10.1371/journal.pntd.0003347, 2014

## Overview



**Country:** Haiti

**Study domain:** country

**Surface:** 27750 km<sup>2</sup>

**Population:** 10911819 (2015)

**Cholera:** Outbreaks have been occurring every year since the introduction in 2010.

**Period studied:** October 2010 to December 2011

**Number of reported cases:** 520000

## Chapter 1. Modeling the Haiti cholera epidemic: the effect of rice-based oral rehydration solution (ORS)

---

### 1.1 Introduction

Starting in October 2010, only months after a disastrous earthquake, Haiti was hit by the largest cholera epidemic in recent history. The outbreak was most likely triggered by the importation of a toxigenic strain of *V. cholerae* [Chin *et al.*, 2011; Frerichs *et al.*, 2012; Hendriksen *et al.*, 2011; Piarroux, 2011], flared up in the Artibonite valley, home to one of the largest rivers of the country, and spread to the entire Haitian territory within less than two months. After one year, almost 500 000 cases and over 6200 deaths had been reported [Barzilay *et al.*, 2013; Gaudart *et al.*, 2013].

The general treatment of cholera patients is based on a so-called oral rehydration therapy (ORT), which is a cost-effective and easily applicable method to replace lost fluids and electrolytes. For the latter purpose, the administered solution contains a mixture of several compounds that were designated oral rehydration salts, including sodium, chloride, and potassium ions as well as glucose. Indeed, glucose is the most commonly added carbohydrate because it stimulates water absorption in the small intestine [Farthing, 2002]. However, field studies have shown that ORS might be improved by alternative carbon sources [Atia and Buchman, 2009]. Moreover, a meta-analysis comparing the treatment with standard, glucose-versus rice-based ORS illustrated the beneficial effects of the latter composition, such as reduced episodes of vomiting, a decrease of the stool volume, and a shortened recovery time [Dutta *et al.*, 2000; Gore *et al.*, 1992; Guarino *et al.*, 2001; Molla *et al.*, 1985].

This chapter was written as a retrospective analysis of the potential impact on the course of the Haitian epidemic of the main findings of Kühn *et al.* [2014], which unravels the molecular mechanisms underlying the better performance of rice-based ORS. Results show that glucose leads to an increased expression of the major virulence genes in the pathogen and, accordingly, to an enhanced production of cholera toxin during *in vitro* experimentation. Because the cholera toxin is primarily responsible for the severe symptoms that are associated with the disease, findings highlight the negative effects of glucose-based ORT.

To understand whether the advantages that rice-based ORS has on individual patients can also be beneficial in limiting the transmission of the disease in the whole community, a spatially explicit model has been developed and applied to the first year of the epidemic in Haiti. The novel approach has been developed along the lines of previous work [Bertuzzo *et al.*, 2008; Rinaldo *et al.*, 2012]. The model consists of 365 nodes corresponding to Haitian watersheds. The epidemiological state (susceptible, infected and recovered) of the population of each node is followed over time. Nodes are connected through hydrological connectivity and human mobility. Infected individuals contribute to the local environmental bacterial concentration, where *V. cholerae* survive for a certain duration and may infect susceptible individuals. The pathogens can also be subject to hydrological transport to downstream nodes. Moreover, rainfall increases the local bacterial concentration through deterioration of sanitary conditions [Gaudart *et al.*, 2013].

To predict the effect of using rice-based instead of standard (glucose-based) ORS, the disease

duration and the bacterial shedding rate of the symptomatic infected have been reduced by ten percent each, an assumption consistent with clinical results [Dutta *et al.*, 2000; Guarino *et al.*, 2001; Molla *et al.*, 1985] as well as with the findings of *in-vitro* studies [Kühn *et al.*, 2014]. The reduction was only applied after an initial period of 30 days, which is assumed to be necessary to switch from standard to rice-based ORS.

## 1.2 Methods

### 1.2.1 Model setup and data

The model domain is the Haitian mainland, subdivided into 365 hydrological subunits (watersheds) with an average surface of 76 km<sup>2</sup>, each one corresponding to a node in our model (Figure 1.1B). The subdivision has been derived from a digital terrain model (DTM) (Data available from the U.S. Geological Survey, <http://earthexplorer.usgs.gov/>) using established hydrological methods [Band, 1986; Montgomery and Dietrich, 1988, 1992; Rodríguez-Iturbe and Rinaldo, 2001; Tarboton, 1997] (Figure 1.1A). The procedure consists of the determination of the unique steepest descent flow path from every pixel of the DTM to the sea. Pixels draining to one and the same outlet belong to the same river basin. As this leads to a very heterogeneous watershed size distribution, the larger basins had to be split into smaller units according to catchment divides, whereas the coastal (smaller) watersheds needed to be aggregated.

The use of hydrologically defined units allowed for a straightforward identification of the hydrological connection from each watershed to its unique downstream neighbor (or to the ocean, for coastal watersheds). The hydrologic connectivity matrix  $P_{ij}$ , which contains information about which watershed drains into which one, follows directly (Section 1.2.2).

Differently from previous applications [Rinaldo *et al.*, 2012], where euclidean distance was used, in this study nodes are connected through the actual road network (Figure 1.1C, OpenStreetMap contributors, available online at <http://www.openstreetmap.org> under the Open Database License), from which the shortest distance between each pair has been computed [Dijkstra, 1959] and taken as input to build the distance matrix  $Q_{ij}$ , necessary to compute human mobility (Equation 1.3).

The population of each subunit (Figure 1.1E) was derived from the remotely sensed population distribution [Oak Ridge National Laboratory, 2011], which has been updated in order to correspond to the most recent population estimates [Mukandavire *et al.*, 2013].

Daily precipitation fields were obtained from a remotely sensed dataset by the National Aeronautics and Space Agency (NASA), which has a spatial resolution of 0.25° latitude and longitude [Huffman *et al.*, 2010] (Figure 1.1F).

**Chapter 1. Modeling the Haiti cholera epidemic: the effect of rice-based oral rehydration solution (ORS)**

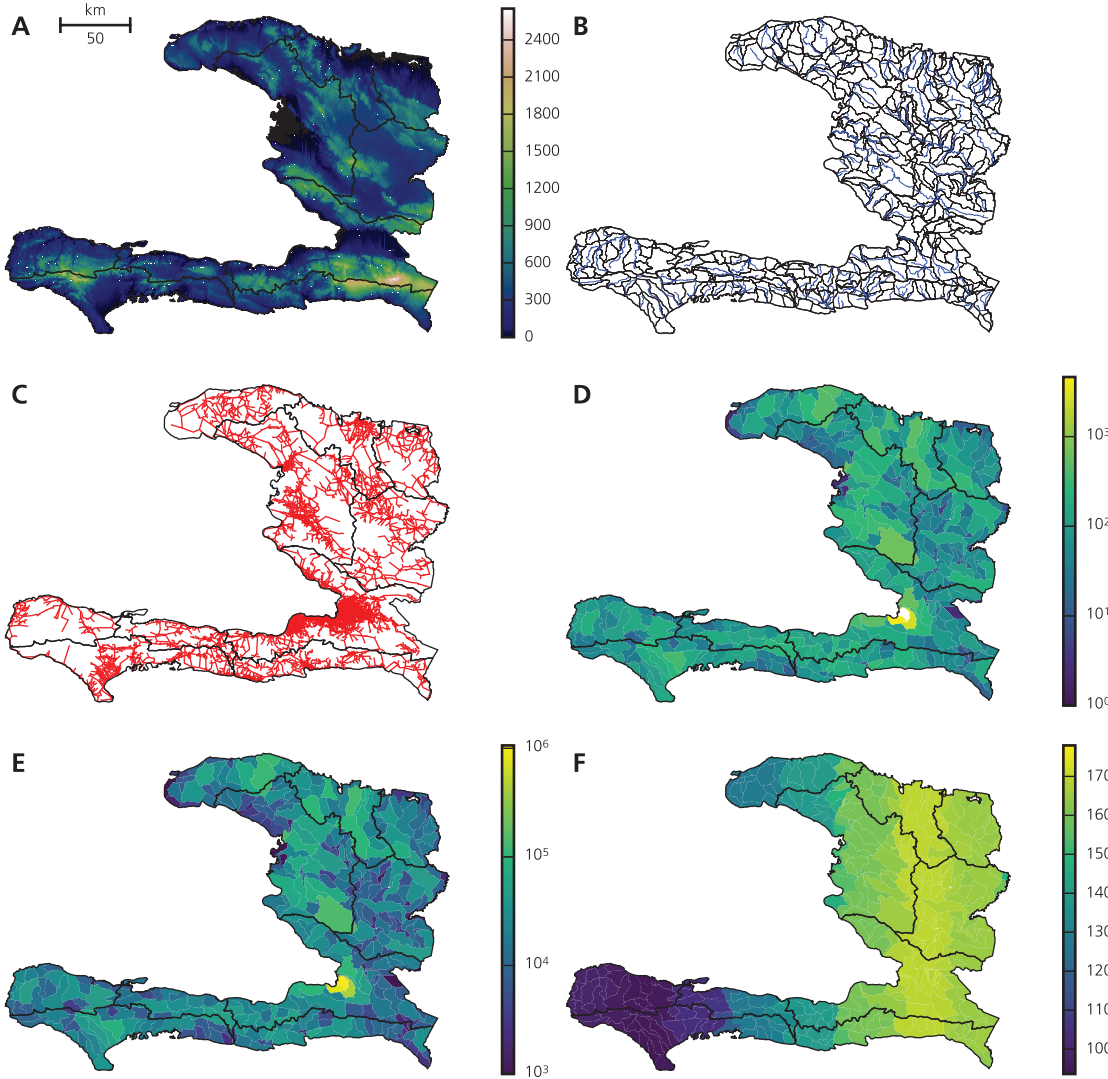


Figure 1.1 – DTM of the Haitian mainland (A) used to extract the 365 model nodes (e.g. watersheds) (B, principal rivers also shown). Road network (C) used to derive the distance matrix between model nodes. Estimates of daily human mobility (number of persons) from the Haitian capital Port-au-Prince to other model nodes according to the calibrated gravity model (Section 1.2.2) (D). Population of each model node (E). Rainfall depth (mm/day) in each watershed on November 5, 2010 (day when hurricane Thomas passed close to the island) (F).

### 1.2.2 Spatially explicit epidemiological model

The spatially explicit epidemiological model builds on the one presented by *Rinaldo et al.* [2012]. In the latter, the difference between symptomatic and asymptomatic infections are not explicitly considered, with parameters such as the rate of infection and the rate of loss of acquired immunity representing an average value between symptomatic and asymptomatic infections. For this study, the specific application to treatment, i.e. ORS, which is only allocated to reported and thus symptomatic cases, makes it necessary to separate the two groups, allowing to apply specific sets of parameters to each of them.

The epidemiological dynamics were modeled using a system of differential equations, which take into account hydrological pathogen transport, human mobility as well as precipitation (Figure 1.2).

$$\begin{aligned}
\frac{dS_i}{dt} &= \mu(H_i - S_i) - \mathcal{F}_i(t)S_i + \rho R_i \\
\frac{dI_{S,i}}{dt} &= \sigma \mathcal{F}_i(t)S_i - (\gamma + \mu + \alpha) I_{S,i} \\
\frac{dI_{A,i}}{dt} &= (1 - \sigma) \mathcal{F}_i(t)S_i - (\varepsilon\gamma + \mu) I_{A,i} \\
\frac{dR_i}{dt} &= \gamma(I_{S,i} + \varepsilon I_{A,i}) - (\rho + \mu) R_i \\
\frac{dB_i}{dt} &= -\mu_B B_i - l \left( B_i - \sum_{j=1}^n P_{ji} \frac{W_j}{W_i} B_j \right) \\
&\quad + \frac{1}{W_i} [1 + \lambda J_i(t)] (p_S q_S \mathcal{G}_{S,i}(t) + p_A q_A \mathcal{G}_{A,i}(t))
\end{aligned} \tag{1.1}$$

Individuals living at node  $i$  can be susceptible ( $S_i$ ), symptomatic infected ( $I_{S,i}$ ), asymptomatic infected ( $I_{A,i}$ ) or recovered ( $R_i$ ).  $B_i$  represents the bacterial concentration in the water reservoir of the node. The population is assumed to be in demographic equilibrium, with newborns considered susceptible.  $\mu$  is the mortality rate (unrelated to cholera) and  $H_i$  the total population size.

The force of infection, i.e. the rate at which susceptible individuals get infected, is:

$$\mathcal{F}_i(t) = \beta \left[ (1 - m) \frac{B_i}{(K + B_i)} + m \sum_{j=1}^n Q_{ij} \frac{B_j}{(K + B_j)} \right] \tag{1.2}$$

where  $\beta$  stands for the exposure rate.  $B_i/(K + B_i)$  is the probability to get infected after exposure,  $K$  being a half-saturation constant [*Capasso and Paveri-Fontana, 1979; Codeço, 2001*].

The fraction of the population traveling to other nodes is given by the parameter  $m$ . The

## Chapter 1. Modeling the Haiti cholera epidemic: the effect of rice-based oral rehydration solution (ORS)

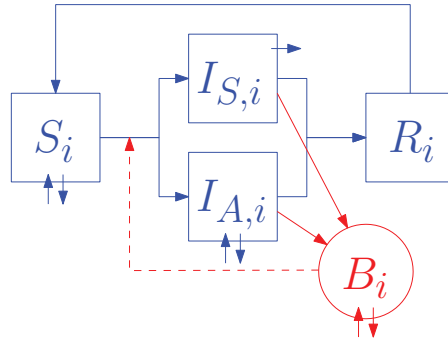


Figure 1.2 – Schematic representation of the model.  $S_i$  stands for susceptible individuals at node  $i$ ,  $I_{S,i}$  and  $I_{A,i}$  for symptomatic and asymptomatic infected respectively,  $R_i$  for recovered and  $B_i$  for bacterial concentration. Blue and red solid arrows indicate fluxes of individuals and bacteria respectively, whereas the red dashed arrow indicates that the bacterial concentration governs infection.  $\updownarrow$  stands for human mobility between nodes and  $\updownarrow$  for hydrological transport of *V. cholerae*.  $\rightarrow$  represents mortality due to cholera. Natural mortality is not shown.

probability to get infected depends on the bacterial concentration of the environmental water reservoir at the home node  $i$  for individuals who are not traveling, and at a remote node  $j$  for the rest. The probability  $Q_{ij}$  to travel from node  $i$  to node  $j$  is based on a gravity-like pattern, used to model short-term displacements and thus does not change the actual population of every node [Erlander and Stewart, 1990]:

$$Q_{ij} = \frac{H_j e^{-d_{ij}/D}}{\sum_{k \neq i}^n H_k e^{-d_{ik}/D}} \quad (1.3)$$

where the attractiveness of node  $j$  is governed by its population size, whereas deterrence exponentially increases with distance  $d_{ij}$  between nodes  $i$  and  $j$  according to a shape factor  $D$ .

Infected individuals show symptoms with probability  $\sigma$ . They may recover with rate  $\gamma$  or die from cholera or other causes with rate  $\alpha$  or  $\mu$  respectively. Asymptomatic infected individuals do not die from cholera and recover faster by a factor  $\varepsilon$  [Nelson et al., 2009]. Recovered individuals lose their acquired immunity with rate  $\rho$ , or die at rate  $\mu$ .

The bacterial concentration at each node is a product of the the bacterial shedding rate  $p$  of the infected individuals and the probability  $q$  that the freshly shed bacteria reach the environmental water reservoir, divided by the volume of the latter. The reservoir is to be seen as a conceptual element uniting different possible transmission routes, such as water, food, surfaces etc., rather than an actual, well designed body of water (see Introduction). Its volume is taken to be directly proportional to the population size at the node  $W_i = aH_i$  [Bertuzzo et al., 2008]. Parameters  $p$  and  $q$  take different values for symptomatic (S) and asymptomatic (A) individuals because people without symptom shed less bacteria [Nelson et al., 2009] and because of different sanitation conditions. See section 1.3.1 for a thorough discussion of these aspects. Symptomatic individuals are assumed to stay within their home node, whereas the



movement of asymptomatics is modeled similarly to that of susceptibles, i.e. via equation (1.2). Asymptomatic people leave their node  $j$  with probability  $m$  and reach node  $i$  based on the gravity factor  $Q_{ji}$ . We thus have, in (1.1):

$$\begin{aligned} \mathcal{G}_{S,i} &= I_{S,i} \\ \mathcal{G}_{A,i} &= (1 - m) I_{A,i} + m \sum_{j=1}^n Q_{ji} I_{A,j} \end{aligned} \tag{1.4}$$

We further consider that rainfall events cause a deterioration of sanitary conditions and a higher probability of contamination of water reservoirs [Gaudart *et al.*, 2013]. The shedding term in equation (1.1) thus increases with precipitation intensity  $J_i(t)$  according to factor  $\lambda$  [Rinaldo *et al.*, 2012]. In addition, *V. cholerae* pathogens are subject to natural mortality at rate  $\mu_B$  and hydrological transport at rate  $l$ . In our case the probability  $P_{ij}$  that pathogens travel from node  $i$  to  $j$  is equal to one if  $j$  is the downstream nearest neighbor of  $i$ , and zero otherwise.

### 1.2.3 Modeling the effect of rice-based ORS in the Haiti epidemic

Our model can be used to estimate the impact of the two presumed effects of changing glucose-to rice-based ORS [Dutta *et al.*, 2000; Gore *et al.*, 1992; Guarino *et al.*, 2001; Molla *et al.*, 1985], a reduction of disease duration and stool volume of infected individuals, on the overall dynamics of the epidemic in space and time. To include these effects into the model, we assume that the reported cases correspond to the symptomatic infected, and that all of them got at least basic treatment with ORS (in a hospital, at home or at a so-called rehydration point [Walton and Ivers, 2011]). In addition, we assume that the reduction in stool volume goes along with a reduction in the number of *V. cholerae* shed per unit time and that a reduced disease duration also shortens the period of bacterial shedding. The effects can then be integrated into the model by reducing the infectious period as well as the bacterial shedding rate for symptomatic individuals. Two additional adjustable parameters ( $\tau_1$  and  $\tau_2$ ) have thus been added to the model:

## Chapter 1. Modeling the Haiti cholera epidemic: the effect of rice-based oral rehydration solution (ORS)

---

$$\begin{aligned}
\frac{dS_i}{dt} &= \mu(H_i - S_i) - \mathcal{F}_i(t)S_i + \rho R_i \\
\frac{dI_{S,i}}{dt} &= \sigma \mathcal{F}_i(t)S_i - \left( \left( \frac{1}{1-\tau_1} \right) \gamma + \mu + \alpha \right) I_{S,i} \\
\frac{dI_{A,i}}{dt} &= (1-\sigma) \mathcal{F}_i(t)S_i - (\varepsilon\gamma + \mu) I_{A,i} \\
\frac{dR_i}{dt} &= \gamma \left( \left( \frac{1}{1-\tau_1} \right) I_{S,i} + \varepsilon I_{A,i} \right) - (\rho + \mu) R_i \\
\frac{dB_i}{dt} &= -\mu_B B_i - l \left( B_i - \sum_{j=1}^n P_{ji} \frac{W_j}{W_i} B_j \right) \\
&\quad + \frac{p_S q_S}{W_i} [1 + \lambda J_i(t)] \left( (1-\tau_2) \mathcal{G}_{S,i}(t) + \chi \mathcal{G}_{A,i}(t) \right),
\end{aligned} \tag{1.5}$$

where  $\chi = p_A q_A / p_S q_S$  (Section 1.3.1).  $\tau_1$  and  $\tau_2$  can be used to adapt the recovery rate as well as the bacterial shedding rate of the symptomatic only.  $\tau_1$  is the reduction in disease duration whereas  $\tau_2$  is the reduction in bacterial shedding rate, both expressed as a fraction of one. For calibration against observed data, both parameters have been fixed to zero (no effect), thus the equations were equivalent to (1.1). To show the influence of a given treatment effect on the overall disease dynamics,  $\tau_1$  and  $\tau_2$  were set to appropriate values after an initial phase of 30 days we assume necessary to initiate systematic treatment with rice-based ORS. The calibrated parameters were kept fixed.

### 1.2.4 Initial condition

At the beginning of the epidemic the entire population is assumed to be susceptible ( $S_i(t_0) = H_i$ ), an assumption suitable for the Haitian epidemic, as no cholera had been recorded for the previous decades [Bertuzzo *et al.*, 2011; Enserink, 2010; Piarroux, 2011; Rinaldo *et al.*, 2012; Sack, 2011; Walton and Ivers, 2011]. As an initial condition a number of infected individuals was introduced to selected nodes in the Centre and the Artibonite departments according to a detailed report about the state of the epidemic on October 20, 2010 [Piarroux, 2011]. Additionally an initial equilibrium bacterial concentration (1.6) was imposed to the same nodes. We further assume that no recovered individuals were present at the beginning of the epidemic ( $R_i(t = t_0) = 0$  for every node  $i$ ).

$$B_i(t = t_0) = \frac{(p_S q_S I_{S,i}(t_0) + p_A q_A I_{A,i}(t_0))}{W_i \mu_B} \tag{1.6}$$

### 1.2.5 Model calibration and parameter estimation

In order to reduce the number of unknown parameters, we introduced the dimensionless bacterial concentration  $\mathcal{B}_i = B_i/K$  along with the aggregated contamination rate for symptomatics  $\theta = p_S q_S / (cK)$ .

Parameters that could not be derived from previous work or from the literature were calibrated using data from daily epidemiological reports available on the website of the Ministry of Public Health (Ministère de la Santé Publique et de la Population) of Haiti [Barzilay *et al.*, 2013]. The calibration period starts at the beginning of the epidemic (20 October 2010) and ends in December 2011. As the case-data are freely available at department level only, our model outputs need to be upscaled for comparison. We assume that the number of reported cases corresponds to the number of newly infected symptomatic individuals. In order to derive the number of reported cases from the model, one thus needs to solve the following equation for  $C_i$  (the modeled cumulative reported cases):

$$\frac{dC_i}{dt} = \sigma \mathcal{F}_i(t) S_i \quad (1.7)$$

The sum of squared residuals between the reported cases and the model output computed according to (1.7) was used as the objective function to minimize, which is equivalent to maximizing the model likelihood while assuming normal, homoscedastic residuals [Sorooshian and Dracup, 1980]. For calibration we relied on a Markov Chain Monte Carlo (MCMC) approach with several chains (DREAM<sub>ZS</sub>, Vrugt *et al.* [2008, 2009]). See Table 1.1 for other parameter values and references.

## 1.3 Results

*In vitro* experiments (Section 1.1 and Kühn *et al.* [2014]) suggest a strong reduction of the amount of cholera toxin when using rice-based ORS, which matches the observed reduction of shedding rates from field studies [Dutta *et al.*, 2000; Gore *et al.*, 1992; Guarino *et al.*, 2001; Molla *et al.*, 1985]. This reduction can be as high as 50% (especially compared to glucose-based ORS, which was recommended by the WHO before 2002 (WHO-ORS)). The reduction of diarrheal duration is usually no larger than 30%. If one assumes a 10% reduction of both, duration and shedding rate, for rice-based compared to glucose-based ORT (for HYPO-ORS, Dutta *et al.* [2000]), the model predicts a considerable decrease in disease incidence over the entire country (Figure 1.3) as well as in individual departments (Figure 1.4). Indeed, the total number of cholera cases within the first 14 months of the epidemic would be reduced from 520,000 cases (as reported by the Haitian Ministry of Health, our model predicts 535,000 cases) to 375,000 cases (i.e. 30% (95% CI 22% – 39%) less total cases until the end of 2011) according to the model. More importantly, if these parameters could be reduced by 15%, then the

## Chapter 1. Modeling the Haiti cholera epidemic: the effect of rice-based oral rehydration solution (ORS)

Table 1.1 – Model parameters with their values and references. Parameters in the upper part of the table have been taken from literature or estimated. Calibrated parameters (with 95% confidence intervals) are shown in the lower part.

Parameter	Units	Value	References
$\beta$	$\text{d}^{-1}$	1.0	<i>Codeço</i> [2001]; <i>Rinaldo et al.</i> [2012]; <i>Tuite et al.</i> [2011]
$\alpha$	$\text{d}^{-1}$	$4.0 \times 10^{-3}$	<i>Pan American Health Organization</i> [2011]; <i>Rinaldo et al.</i> [2012]
$\mu$	$\text{d}^{-1}$	$1/(61 \times 365)$	<i>Central Intelligence Agency</i> [2009]; <i>Rinaldo et al.</i> [2012]
$\gamma$	$\text{d}^{-1}$	0.2	<i>Andrews and Basu</i> [2011]; <i>Bertuzzo et al.</i> [2008]; <i>Codeço</i> [2001]
$\mu_B$	$\text{d}^{-1}$	0.2	<i>Bertuzzo et al.</i> [2008]; <i>Codeço</i> [2001]; <i>Mari et al.</i> [2012b]
$\rho$	$\text{d}^{-1}$	$1/(3 \times 365)$	<i>Koelle et al.</i> [2005]; <i>Rinaldo et al.</i> [2012]
$\varepsilon$	–	5	<i>Nelson et al.</i> [2009]
$p_A/p_S$	–	$10^{-3}$	<i>Andrews and Basu</i> [2011]; <i>Kaper et al.</i> [1995]; <i>Nelson et al.</i> [2009]
$q_A/q_S$	–	200 ( $10^{-1}$ to $10^4$ )	(see section 1.3.1 and Figure 1.8)
$\theta$	–	0.55 [0.49 0.63]	–
$l$	$\text{d}^{-1}$	0.20 [0.14 0.24]	–
$m$	–	0.037 [0.026 0.050]	–
$D$	km	343 [221 398]	–
$\lambda$	$\text{d mm}^{-1}$	0.081 [0.073 0.094]	–
$\sigma$	–	0.10 [0.10 0.11]	–

total number of cholera cases would drop by 59% (95% CI 47% – 67%), and if the parameters were reduced by 20%, the number of cases would even decrease by 74% (95% CI 71% – 76%) (Figure 1.5). The ranges of variation shown reflect the 2.5 – 97.5 percentiles of the uncertainty related to parameter estimation. Such behavior (i.e. the more than doubled reduction of total infections owing to a 10% – 20% reduction in bacterial shedding and disease duration) is typical of nonlinear epidemiological dynamics [*Gatto et al.*, 2012]. Interestingly, owing to a higher number of susceptibles (Figure 1.6), a larger number of cholera cases would have been predicted for November 2011, one year after the initial onset of the outbreak, triggered by important rainfall events. However, such a one-year time span would have allowed other intervention strategies to be put into place, which could potentially avoid later cholera case peaks e.g., by reducing exposure rates via improved sanitary conditions.

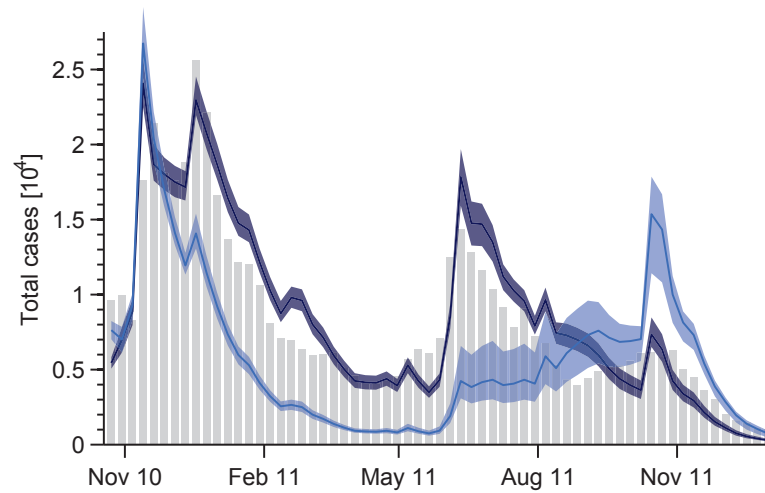


Figure 1.3 – Evolution of the epidemic: sum of weekly cases in the entire Haiti over time. Observed cases (gray bars), the calibrated model (dark blue) and model runs with a ten percent reduction of the symptomatic shedding rate as well as duration of the infectious period due to rice-based ORT (light blue). Shaded areas show the 95% credibility intervals resulting from parameter uncertainty, solid lines show the median trajectory. The replacement of glucose-based with rice-based ORS was assumed to take place 30 days after the onset of the epidemic.

### 1.3.1 Sensitivity to the (a)symptomatic contamination probability

While the relative value of asymptomatics versus symptomatics shedding rate ( $p_A/p_S$ ) can be found in the literature [Andrews and Basu, 2011; Kaper et al., 1995; Nelson et al., 2009], the relative probability that asymptomatics contaminate the local environment ( $q_A/q_S$ ) is difficult to estimate. Assuming that a high fraction of symptomatics is admitted to hospitals or treatment centers, this ratio indeed depends on the sanitary conditions in health care facilities compared to regular households. One would expect sanitation to be a key issue in hospitals. However, little functional sewer systems exist in Haiti [Farmer et al., 2011]. In addition, during the peak phases of the epidemics, health care facilities were subject to over-occupancy [Walton and Ivers, 2011]. Furthermore, during cholera outbreaks, symptomatic patients are released as soon as their condition starts improving (less than three liters of stool in six hours [Bauernfeind et al., 2004]), even if they might still shed *V. cholerae*. Therefore, depending on the relative impact of the various effects stated above, the ratio  $q_A/q_S$  can take a broad range of values. By adding it to the calibration procedure as a parameter we were not able to identify its value uniquely due to a very flat posterior distribution caused by a high correlation with other parameters (e.g.  $\theta$ ). In addition, the Akaike information criterion (AIC) [Akaike, 1974; Burnham and Anderson, 2002] indicates that the improvements of the model fit by adding this additional parameter are not significant (see table 1.2). We thus decided not to calibrate  $q_A/q_S$ . To assess the sensitivity of the remaining calibration parameters as well as the model outputs with respect to changes of the ratio  $q_A/q_S$  we calibrated the model with a range

# Chapter 1. Modeling the Haiti cholera epidemic: the effect of rice-based oral rehydration solution (ORS)

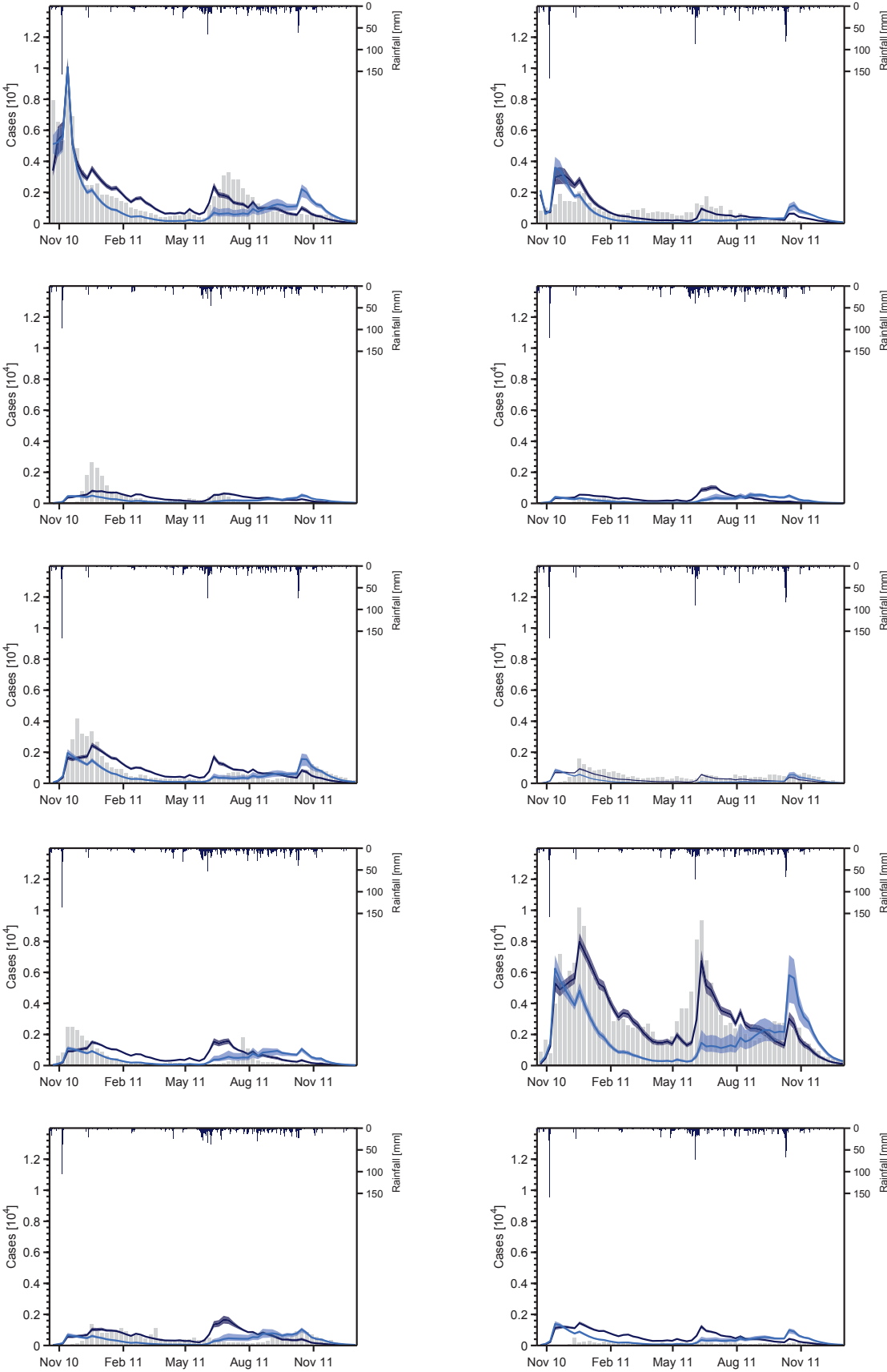


Figure 1.4 – Evolution of the epidemic: sum of weekly cases in the Haitian departments over time. Observed cases (gray bars), the calibrated model (dark blue) and model runs with a ten percent reduction of the symptomatic shedding rate as well as duration of the infectious period due to rice-based ORT (light blue). Shaded areas show the 95% credibility intervals resulting from parameter uncertainty, solid lines show the median trajectory. The replacement of glucose-based with rice-based ORS was assumed to take place 30 days after the onset of the epidemic. Departments are (from left to right, top to bottom): Artibonite, Centre, Grande Anse, Nippes, Nord, Nord-Est, Nord-Ouest, Ouest, Sud, and Sud-Est. Daily rainfall in each department is also shown (top of each panel).

Table 1.2 – Number of calibrated parameters and AIC scores for calibration with and without  $q_A/q_S$  (Section 1.3.1). The last column shows the Akaike difference, which must be  $> 4$  for significance.

Model	Parameters	AIC	$\Delta_{AIC}$
calibration with $q_A/q_S = 200$	6	12205	–
calibrating also $q_A/q_S$	7	12217	12

of different values (0.1 to 10000, see Figure 1.7). We were able to calibrate the model almost equally well independently of the value of  $q_A/q_S$ , while the values of the calibrated parameters varied.  $\theta$  is particularly sensitive to changes in  $q_A/q_S$  because both parameters directly affect the bacterial shedding and they thus compensate. Figure 1.8 shows the important influence  $q_A/q_S$  has on the result of a ten percent reduction of the bacterial shedding rate as well as the disease duration (using parameters  $\tau_1$  and  $\tau_2$  according to Section 1.2.3). This is because  $q_A/q_S$  directly acts on the relative contribution of symptomatics versus asymptomatics to the environmental bacterial concentration, whereas only symptomatic are treated with ORS. It can further be seen that only a very high value of  $q_A/q_S$  (above 200) reduce the impact of rice-based ORS significantly. Based on the above analysis, it can, however, be considered highly unlikely that this fraction takes even higher values under Haitian conditions, and that assuming a value of  $q_A/q_S = 200$  constitutes a conservative assumption.

## 1.4 Discussion

Model results suggest that observed reductions of shedding rate and disease duration by changing the type of ORS [Dutta *et al.*, 2000; Gore *et al.*, 1992; Guarino *et al.*, 2001; Kühn *et al.*, 2014; Molla *et al.*, 1985] used during the Haitian cholera epidemic could have led to a considerable reduction in the number of cases in all Haitian departments. The reduction in the number of cases immediately follows the switch to rice-based ORS after an initial period of 30-days. The second, rainfall-related peak of the epidemic in November 2011, however, is predicted to be much larger in all departments when changing ORS, owing to the higher number of susceptibles present.

**Chapter 1. Modeling the Haiti cholera epidemic: the effect of rice-based oral rehydration solution (ORS)**

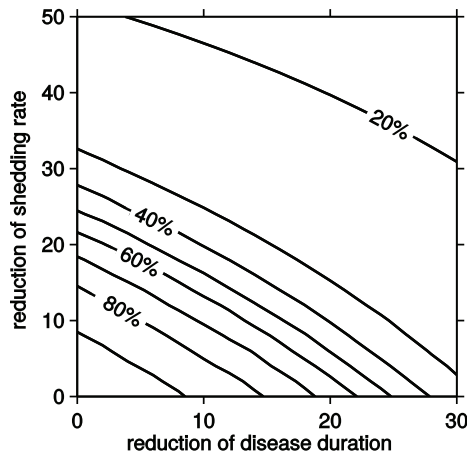


Figure 1.5 – Reduction of the number of cases when reducing shedding rate and disease duration. Contour plot of the percentage of total cases in the whole country until December 2011, predicted by the model when applying variable reductions of the shedding rate (y-axis) and disease duration (x-axis), compared to the cases predicted by the calibrated model.

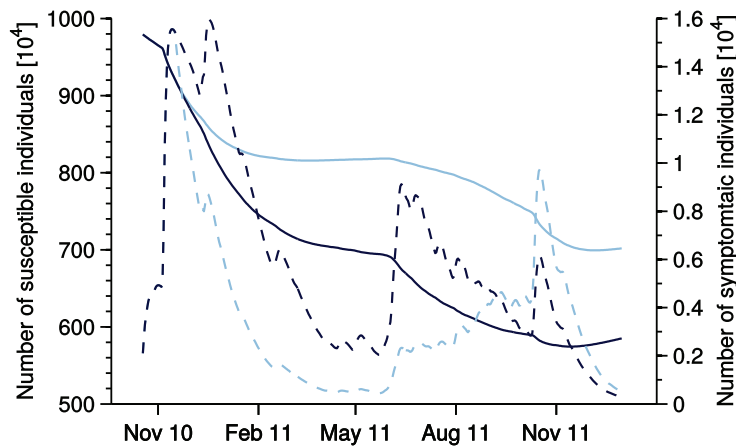


Figure 1.6 – Evolution of modeled susceptibles and symptomatically infected. The solid lines show the modeled evolution of the number of total susceptibles over time as calibrated (dark blue) and with a 10% reduction in bacterial shedding rate as well as disease duration (light blue). Dashed lines: idem for the number of symptomatically infected. Trajectories shown correspond to the best performing parameter set. Note the higher number of susceptibles in fall 2011 after introducing the reductions, which leads to the more pronounced peak of total infections in November 2011.



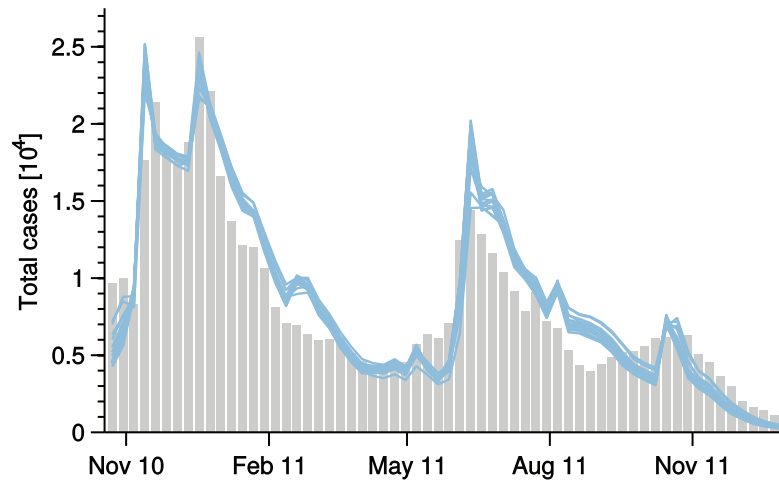


Figure 1.7 – Evolution of the epidemic. Observed cases (gray bars) and model calibrated with different values of  $q_A/q_S$  (blue lines, see Figure 1.8).

The number of model nodes and their size is determined by the procedure used to derive the hydrologic subdivision from the DTM (Section 1.2.1). The optimal spatial aggregation of the model might, however, depend upon factors such as the population distribution and the available data and needs to be investigated further. In addition, the model incorporates two mechanisms through which the cholera epidemic spreads among watersheds, hydrologic dispersal as well as human mobility. Their relative importance on the disease dynamics, which possibly varies with time and between different phases of the epidemic, is unknown and should be subject to in-depth analysis.

While the model fits the sum of cases over the whole country well, the epidemic curves in certain departments are followed less accurately, particularly where the number of cases is low (Figure 1.4). This is due to the fact that the same parameter sets are applied to all model nodes and that the likelihood formulation emphasizes higher values. Departments with a higher number of cases thus have a higher influence on the calibration results. Our main conclusions, however, should not be affected by this limitation, as the reduction of the number of cases in the model results can be observed throughout the country, with the highest reduction in departments with high numbers of cases, well fit by the model. Future work should investigate spatially explicit parameterizations of models in presence of heterogeneous distributions of cases.

The proportion of averted cases by changing ORS, which is the main result of this chapter, depends on the environmental bacterial concentration and on the way this concentration influences the force of infection (Section 1.2.2). Whereas the reduction in the concentration of bacteria present in the environment due to reduced shedding and reduced disease duration has been estimated using literature values and by performing a sensitivity analysis for effects that are difficult to quantify (Section 1.3.1), the way in which this bacterial concentration influences the force of infection has remained the same. The nonlinearity observed in the

## Chapter 1. Modeling the Haiti cholera epidemic: the effect of rice-based oral rehydration solution (ORS)

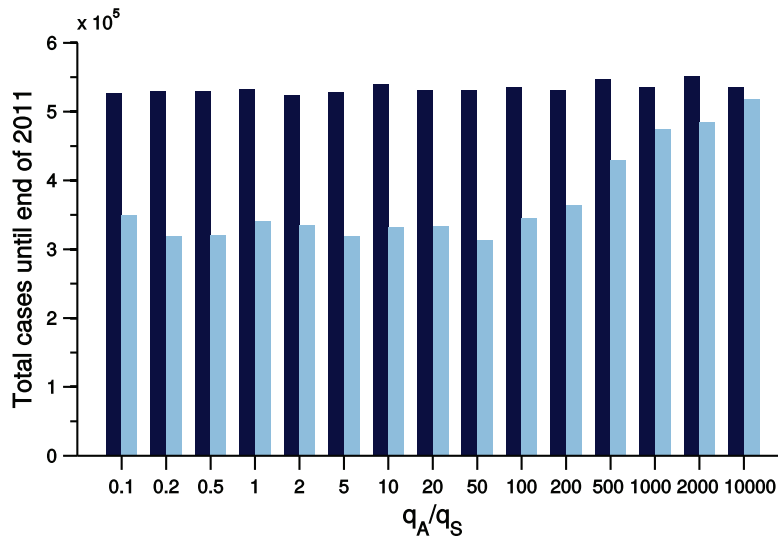


Figure 1.8 – Total cases up to December 2011 according to the model calibrated with different values for  $q_A/q_S$  (dark blue). Light blue bars show model runs using the exact same parameters except for the addition of a ten percent decrease of shedding rate and a ten percent shortening of the disease duration (Section 1.3).

number of averted cases depending on the reduced bacterial load (Section 1.3) might result from the logistic function linking the bacterial concentration to the force of infection [Codeço, 2001; Grad *et al.*, 2012]. Whereas we advocate that the principal conclusion of this chapter, namely an important impact of changes in bacterial shedding rate and duration on the number of cases, is robust to changes in this formulation, the sensitivity of numerical results to such changes remains unknown.

### 1.5 Other studies based on the same model setup

Since the end of 2011, when the principal study presented in this chapter ends, outbreaks alternating with lull phases have been recorded in Haiti every year [Rebaudet *et al.*, 2013c]. As of September 25, 2016, the Ministry of Public Health and Population has reported more than 790 000 cases [Ministère de la santé publique et de la population (MSPP), 2016], which correspond to an average cumulative attack rate of around 8%. With more than 9400 deaths, the resulting cumulative case fatality rate is 1.2%.

The model setup described previously has been used in several other studies over the past few years, notably one addressing the probability of extinction of the Haitian cholera epidemic by the means of a stochastic formulation of the model [Bertuzzo *et al.*, 2016]. Another study investigated the benefits of using a spatially explicit model formulation depending on the available data by assessing the predictive ability of models using different spatial aggregation levels, among them the one presented in this chapter, and different ways of parameterizing the model [Mari *et al.*, 2015a]. Pasetto *et al.* [2016] showed how the model presented above, in

### **1.5. Other studies based on the same model setup**

---

conjunction with data assimilation tools such as the ensemble Kalman filter [Evensen, 2009] and suitable rainfall forecasts, can be used to issue short- to mid-term forecasts of the number of cholera cases in Haitian regions.



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

*This chapter has been published in the Proceedings of the National Academy of Sciences of the United States of America [Finger et al., 2016]. In close collaboration with Lorenzo Mari, Enrico Bertuzzo and Andrea Rinaldo, Flavio Finger has formulated the research questions, designed the study and adapted the model formulation from previous work by his colleagues and himself. The mobile phone data has been analyzed by Lorenzo Mari, Enrico Bertuzzo and Flavio Finger, the final extraction of mobility patterns has been performed by Lorenzo Mari. Flavio Finger has supervised the work of a master student (Tina Genolet) who implemented the first version of the model and performed a preliminary analysis with parts of the data. Flavio Finger has re-implemented an adapted, improved version of the model and calibrated it to case-data provided by Guillaume Constantin de Magny and Noël Magloire Manga. He has produced and analyzed the results, performed all additional analyses and written the draft of the article. Lorenzo Mari, Enrico Bertuzzo and Andrea Rinaldo have contributed to the writing. All authors have contributed to the interpretation of the results.*

Finger, F., T. Genolet, L. Mari, G. C. de Magny, N. M. Manga, A. Rinaldo, and E. Bertuzzo, Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks, *Proceedings of the National Academy of Sciences of the United States of America*, 113(23), 6421–6426, doi: 10.1073/pnas.1522305113, 2016

## Chapter 2. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks

---

### Overview



**Country:** Senegal  
**Study domain:** country  
**Surface:** 196712 km<sup>2</sup>  
**Population:** 13508715 (2013)  
**Cholera:** Outbreaks occurring at irregular intervals  
**Period studied:** January 2011 to January 2012  
**Number of reported cases:** 27000

### Abstract

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 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 provides a new, first-order source of information which 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 approximately 150000 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 27000 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 effect could not be explained by classic, static approaches describing human mobility. Model results also show how concentrated efforts towards disease control in a transmission hotspot could have an important effect on the large-scale progression of an outbreak.

### 2.1 Introduction

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 [Bajardi *et al.*, 2011; Balcan *et al.*, 2009; Colizza *et al.*, 2006; Meloni *et al.*, 2011; Tizzoni *et al.*, 2014; Wesolowski *et al.*, 2012b]. Researchers have commonly focused on infectious diseases transmitted through direct contact between persons (e.g. [Bajardi *et al.*, 2011; Balcan *et al.*, 2009; Colizza *et al.*, 2006; Meloni *et al.*, 2011]). 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 [Chao *et al.*, 2011; Gurarie and Seto, 2009; Mari *et al.*, 2012b; Rinaldo *et al.*, 2012]. Such models translate our comprehension of the mechanisms driving disease transmission (such as rainfall [Rinaldo *et al.*, 2012]) and spread (such as hydrologic transport of pathogens [Bertuzzo *et al.*, 2008; Gurarie and Seto, 2009], 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 [Bertuzzo *et al.*, 2011, 2016; Reiner *et al.*, 2012], but also to test alternative model implementations [Mari *et al.*, 2015a], or to evaluate the effects of various interventions on disease dynamics [Azman *et al.*, 2012; Kühn *et al.*, 2014; Tuite *et al.*, 2011].

To include population movement in epidemiological models, researchers often rely on approaches such as the gravity (e.g. [Erlander and Stewart, 1990]) or radiation [Simini *et al.*, 2012] 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 with respect to 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 [Mari *et al.*, 2012b; Rinaldo *et al.*, 2012; Tuite *et al.*, 2011]. 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 sub-seasonal changes in mobility patterns.

With the increasing diffusion of mobile phones, which have become very widely used even in developing countries [Palchykov *et al.*, 2014; Wesolowski *et al.*, 2014a], a new source of information about human mobility has emerged. Each time a phone emits or receives a call or text message, the antenna which the cell phone is logged in to is registered by the service provider along with the time of the event [Candia *et al.*, 2008]. 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 [de Montjoye *et al.*, 2013], 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 [Candia *et al.*, 2008]. Using mobile phone records of a sufficiently large number of users, one can thus estimate human mobility fluxes with high accuracy, including spatiotemporal variability across a variety of scales [de Montjoye *et al.*, 2013] and without resorting to any particular model.

## Chapter 2. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks

---

A number of recent studies focuses on the use of mobile phone data to extract human mobility patterns in developing countries at different scales in space and time [Lu *et al.*, 2012, 2013; Perkins *et al.*, 2014]. Others compare the movement patterns extracted from mobile records to traditional data sources such as censuses [Wesolowski *et al.*, 2013a] and surveys [Wesolowski *et al.*, 2014b]. Several studies deal with the comparison to human mobility models [Bengtsson *et al.*, 2015; Palchykov *et al.*, 2014]. 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 [Bengtsson *et al.*, 2011; Lu *et al.*, 2012], 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 [Bengtsson *et al.*, 2015; Mari *et al.*, 2015b; Tatem *et al.*, 2014; Wesolowski *et al.*, 2012b, 2015].

Mass gatherings, such as pilgrimages, sport events, or music festivals, can be critical in the spread of infectious diseases following various transmission routes [Abubakar *et al.*, 2012; Memish *et al.*, 2015]. When it comes to orofecally transmitted diseases, such as shigellosis [Wharton *et al.*, 1990] or cholera [de Magny *et al.*, 2012; World Health Organization, 2008], 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 three months 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 2.1). Anecdotal evidence [Echenberg, 2011; International Federation of Red Cross and Red Crescent Societies, 2007; World Health Organization, 2008] 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 [de Magny *et al.*, 2012].

We develop a spatially explicit, fully mechanistic model for the 2005 cholera outbreak in Senegal, based on previous work [Bertuzzo *et al.*, 2016; Mari *et al.*, 2012a; Rinaldo *et al.*, 2012]. In addition to human mobility, we take into account rainfall as an important driver of disease transmission [de Magny *et al.*, 2012; Rinaldo *et al.*, 2012] and we incorporate the effect of over-



Table 2.1 – Regions of Senegal (as of 2005) with their population (2005 estimates), the total number of reported cases during the epidemic, cumulative incidence and the mobile phone sample size (relative to 2013 population).

Region	Population ( $\times 10^6$ )	Cases	Incidence (‰)	Sample size (‰)
Dakar	2.62	6573	2.51	22.64
Diourbel	1.22	11772	9.61	4.11
Fatick	0.64	1928	3.00	4.63
Kaolack	1.06	1014	0.96	5.19
Kolda	0.89	57	0.06	3.86
Louga	0.68	1806	2.64	5.43
Matam	0.50	0	0	7.12
Saint-Louis	0.75	1653	2.20	8.99
Tambacounda	0.58	87	0.15	6.11
Thiès	1.28	2515	1.97	9.60
Ziguinchor	0.31	124	0.40	9.79

crowding 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 (Section 2.2). Daily population fluxes between the 123 arrondissements of Senegal are estimated from a dataset of roughly 150 000 randomly selected mobile phone users tracked during the entire year 2013 (Section 2.2, [*de Montjoye et al.*, 2014]). 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.

## 2.2 Methods

### 2.2.1 Study domain and administrative subdivision of Senegal

The domain of our study is the country of Senegal, subdivided into 123 arrondissements as of 2013 (Figure 2.1). The administrative subdivision of the country has changed in 2008, in particular the number of regions changed from 11 to 14. Epidemiological data refers to the regions as of 2005. To upscale the model output from the 2013 arrondissement scale to that of the epidemiological data, each 2013 arrondissement was assigned to a 2005 region. For 2013 arrondissements belonging to more than one 2005 region, cases were assigned proportionally to the population living in each region.

## Chapter 2. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks

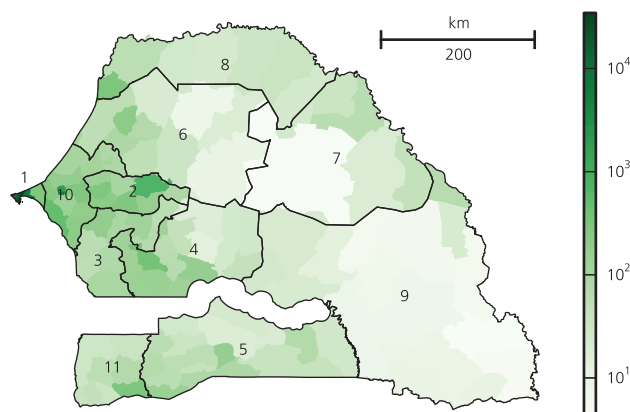


Figure 2.1 – Population density (people per km<sup>2</sup>) per arrondissement in Senegal (2010). Regions (according to the 2005 subdivision) are numbered from 1-11: Dakar, Diourbel, Fatick, Kaolack, Kolda, Louga, Matam, Saint-Louis, Tambacounda, Thiès, Ziguinchor.

### 2.2.2 Data

#### Mobile phone records

Mobile phone call records belong to a dataset which has been released by Orange/Sonatel, an important mobile phone provider in Senegal, for the D4D-Senegal challenge (<http://d4d.orange.com>, accessed on November 10, 2015) [*de Montjoye et al.*, 2014]. The dataset used herein has been coarse-grained by the provider from antenna to arrondissement level (Section 2.2.1) and contains the arrondissement where 146352 randomly selected users were located while making calls or sending text messages throughout the year 2013.

#### Population

Spatially distributed population estimates for the year 2010 with a resolution of approximately 100 m have been obtained from AfriPop (<http://www.afripop.org>, accessed on November 14, 2014) [*Linard et al.*, 2012], and spatially aggregated to the 123 arrondissements of Senegal (Figure 2.1). As the total population of Senegal has increased by 15% between 2005 and 2010, an average growth rate per region has been computed using official data from the Agence Nationale de la Statistique et de la Démographie (<http://donnees.ansd.sn>, accessed on November 14, 2014), and the population in each arrondissement adapted accordingly.

#### Cholera cases

Reported cholera case data were obtained from the website of the Senegalese Ministry of Health [*de Magny et al.*, 2012] and from the WHO national office in Dakar.

## Precipitation

Daily remotely acquired precipitation estimates (CPC/Famine Early Warning System Daily Estimates) for the year 2005 with a resolution of approximately  $0.1^\circ$  have been obtained from the National Oceanic and Atmospheric Administration (NOAA) (<http://www.cpc.ncep.noaa.gov/penalty/exhyphenpenaltyproducts/penalty/exhyphenpenaltyfews/penalty/exhyphenpenaltyrfe.shtml>, accessed on October 14, 2015). They have been spatially averaged over each of the 123 arrondissements (Figure 2.2).

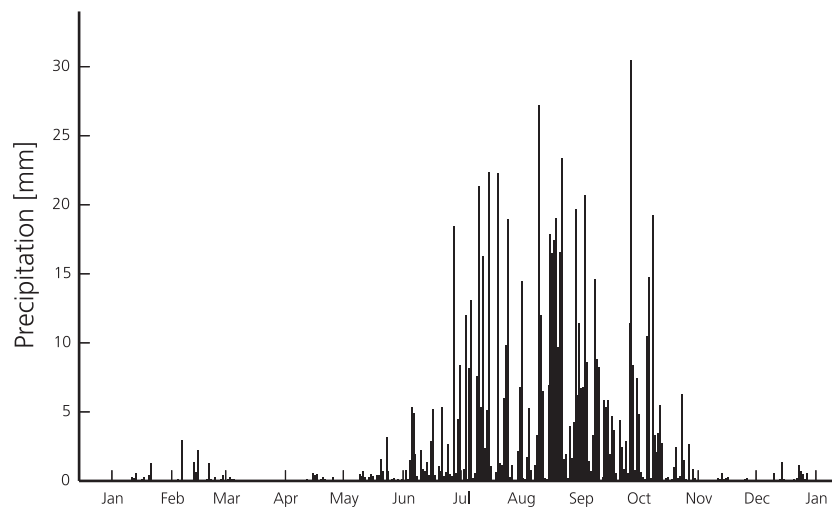


Figure 2.2 – Daily precipitation depth in 2005 averaged over all arrondissements.

### 2.2.3 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 146352 randomly selected users throughout the year 2013 at arrondissement level (Section 2.2.2, Table 2.1). A record in the dataset consists of an anonymous user id, 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 (7 pm to 7 am), 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 assumptions 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  $i$  spent in arrondissement  $j$  during day  $t$ .  $Q_{ii}(t)$  thus represent the fraction of time spent at the home arrondissement [Mari *et al.*, 2015b].

As the Islamic calendar is based on a lunar scheme with 354 days per year, the dates of the

## Chapter 2. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks

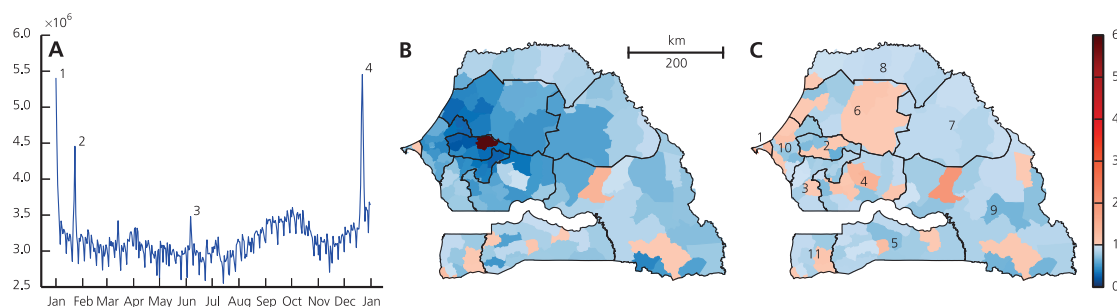


Figure 2.3 – (A) Daily evolution of the total number of moving people (i.e. people leaving their home arrondissement) throughout 2013 estimated from mobile phone records. Numbered peaks correspond to the following mass gatherings: Grand Magal de Touba (1 and 4), Gamou de Tivaouane (2), Magal de Kazu Rajab (3). (B and C) Number of people present in each arrondissement on December 22, 2013, during the Grand Magal de Touba (B) and averaged over the year (C) divided by the number of people living there. Regions (according to the 2005 subdivision, see 2.2.1) are numbered from 1-11: Dakar, Diourbel, Fatick, Kaolack, Kolda, Louga, Matam, Saint-Louis, Tambacounda, Thiès, Ziguinchor.

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  $29 \pm 3$  days), 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 (Section 2.2.7, Figure 2.4 and Table 2.4).

### 2.2.4 Spatially explicit epidemiological model

The spatially explicit epidemiological model used herein builds on previous works [Bertuzzo *et al.*, 2016; Kühn *et al.*, 2014; Mari *et al.*, 2012a; Rinaldo *et al.*, 2012]. In addition to the different formulation of human mobility and the addition of the overcrowding effect (see below), the main difference to the model presented in Chapter 1 is that here asymptotically infected are assumed not to contribute significantly to the bacterial load and can thus be considered recovered [Kaper *et al.*, 1995; King *et al.*, 2008; Nelson *et al.*, 2009]. Furthermore, the spreading of the disease is thought to be caused mainly by human mobility and hydrologic connectivity can thus be neglected. The model domain is the country of Senegal, each arrondissement (Figure 2.1A,  $N = 123$ ) being a node  $i$  with population  $H_i$  (Section 2.2.2). 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):

$$\begin{aligned}
 \frac{dS_i}{dt} &= \mu(H_i - S_i) - \mathcal{O}_i(t)\mathcal{F}_i(t)S_i + \rho R_i \\
 \frac{dI_i}{dt} &= \sigma\mathcal{O}_i(t)\mathcal{F}_i(t)S_i - (\gamma + \mu + \alpha)I_i \\
 \frac{dR_i}{dt} &= \gamma I_i + (1 - \sigma)\beta_i(t)\mathcal{O}_i(t)\mathcal{F}_i(t)S_i - (\rho + \mu)R_i \\
 \frac{dB_i}{dt} &= -\mu_B B_i + \frac{p}{aH_i} [1 + \lambda J_i(t)]\mathcal{O}_i(t)\mathcal{G}_i(t)
 \end{aligned} \tag{2.1}$$

where

$$\mathcal{O}_i(t) = \exp\left(\frac{\omega}{H_i} \sum_{j=1}^N M_{ji}(t)H_j\right) \tag{2.2}$$

$$\mathcal{F}_i(t) = \beta \sum_{j=1}^N M_{ij}(t) \frac{B_j}{K + B_j} \tag{2.3}$$

$$\mathcal{G}_i(t) = \sum_{j=1}^N M_{ji}(t)I_j. \tag{2.4}$$

The population is assumed to be in demographic equilibrium, with per capita birth and natural death rate  $\mu$ . Equations of different nodes are coupled via the human mobility matrix  $M_{ij}(t)$ , which is derived from 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 [Wesolowski *et al.*, 2012a, 2013b], we introduce a calibration parameter  $c$  which relates the two matrices as follows:

$$M_{ii}(t) = cQ_{ii}(t) \tag{2.5}$$

$$M_{ij}(t) = c'_i(t)Q_{ij}(t), \quad j \neq i \tag{2.6}$$

$$c'_i(t) = \frac{1 - cQ_{ii}(t)}{\sum_{h \neq i} Q_{ih}(t)}, \tag{2.7}$$

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

## Chapter 2. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks

---

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 semi-saturation function with parameter  $K$  and rate of exposure  $\beta$ .  $\mathcal{O}_i(t)$  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  $\omega$  and the number of people present at the node at time  $t$  divided by its actual population. We assume that only a fraction  $\sigma$  of infections are symptomatic. Asymptomatically infected hosts do not significantly contribute to the bacterial load in the environment nor die of cholera [Bertuzzo *et al.*, 2016; King *et al.*, 2008; Kühn *et al.*, 2014], and can thus, for the purpose of the model, be considered recovered immediately. Symptomatically infected people may recover at rate  $\gamma$  or die from cholera-unrelated causes at rate  $\mu$  or from cholera at rate  $\alpha$ , whereas recovered lose their acquired immunity at rate  $\rho$  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  $\lambda$  [Rinaldo *et al.*, 2012] (Section 2.2.2), and by overcrowding through the factor  $\mathcal{O}_i(t)$ . The environmental bacteria population decays with rate  $\mu_B$ . We define  $B_i^* = B_i/K$ . Expressing the system of equations in this term, parameters  $a$  and  $K$  gets absorbed in  $\theta = p/aK$  so that the number of free parameters is reduced by 2.

### 2.2.5 Initial conditions

The initial conditions characterize the epidemiological state of the population at the beginning of January 2005. An initial number of cases was assigned to each arrondissement in Diourbel, region where the first cases were reported, which was either manually fixed (1 case per arrondissement) or calibrated (see Section 2.2.6 and Table 2.3). The rest of the population is assumed to be susceptible. We consider that there is no initial immunity, because the last major cholera epidemic in Senegal had occurred in 1996 [World Health Organization, 2008] and thus the period of time between the two events is much longer than reported immunity duration in endemic settings [Koelle *et al.*, 2005]. The initial bacterial concentration is assumed to be in equilibrium with the initial number of infected in absence of mobility:  $B_{i,0}^* = I_{i,0}\theta/(\mu_B H_i)$  [Kühn *et al.*, 2014; Rinaldo *et al.*, 2012].

### 2.2.6 Parameter estimation

While some parameters were assigned using values from the literature (Table 2.2), others (number depending on the model, Section 2.2.7 and Table 2.3) were calibrated, including the initial number of cases in the region of Diourbel, equally distributed among arrondissements. Model calibration was performed using a parallel implementation of the Markov chain Monte Carlo (MCMC) method called EMCEE PT sampler [Foreman-Mackey *et al.*, 2013], which

allows exchange of information among walkers. To explore the largest possible portion of the parameter space a total of 300 walkers running at 3 different “temperatures”, which set the probability of accepting jumps to less favorable regions, and starting from the region of a well performing hand-tuned parameter set were used. We employed wide uniform priors (Table 2.2). The walkers were run up to visual convergence (5000 – 8000 iterations) and all but the last 1000 iterations were discarded as burn-in.

Table 2.2 – Fixed (top) and calibrated (bottom) parameters of the best performing model (A in Table 2.3). For the latter the 95% confidence intervals of the posterior distribution are also shown.

Parameter	Units	Prior	Value	Reference
$\gamma$	day <sup>-1</sup>		0.2	<i>Rinaldo et al.</i> [2012]
$\mu$	day <sup>-1</sup>		1/(61 · 365)	<i>Central Intelligence Agency</i> [2013]
$\alpha$	day <sup>-1</sup>		0.004	<i>Bertuzzo et al.</i> [2011]; <i>Rinaldo et al.</i> [2012]
$\mu_B$	day <sup>-1</sup>		0.2	<i>Bertuzzo et al.</i> [2011]; <i>Rinaldo et al.</i> [2012]
$\rho$	day <sup>-1</sup>		1/600	<i>Koelle et al.</i> [2005]
$\beta$	day <sup>-1</sup>		1	<i>Bertuzzo et al.</i> [2008]; <i>Codeço</i> [2001]; <i>Rinaldo et al.</i> [2012]
$\theta$	day <sup>-1</sup>	[0 2]	0.34 [0.28 0.41]	
$\lambda$	mm <sup>-1</sup>	[0 1]	0.049 [0.040 0.061]	
$\sigma$	–	[0 0.5]	0.019 [0.016 0.021]	
$\omega$	–	[0 2]	0.86 [0.824 0.889]	
$c$	–	[1 2]	1.40 [1.375 1.410]	
$I_0^a$	–	[1 500]	301 [204 416]	

<sup>a</sup> Initial number of infected in the region of Diourbel equally distributed among arrondissements

The models were evaluated against reported numbers of cases in all 11 regions. Weekly cumulative cases  $C_i$  were computed from the model using the following equation:

$$C_i(\tau) = \sigma \int_{\tau-\Delta t}^{\tau} \mathcal{O}_i(t) \mathcal{F}_i(t) S_i dt$$

where  $\tau$  corresponds to the end of the week and  $\Delta t$  is 1 week. The results were then upscaled from the arrondissement to the regional scale for comparison with reported cases. Model likelihood was computed assuming mutually independent, homoscedastic and normally distributed residuals [*Sorooshian and Dracup*, 1980] across regions.

## Chapter 2. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks

---

### 2.2.7 Model selection

To determine relevant processes to be included in the model and to find an appropriate compromise between accuracy and model complexity, hereby preventing overfitting, candidate models were compared using the Deviance Information Criterion (DIC) [Spiegelhalter *et al.*, 2002] as well as the coefficient of determination. DIC, which allows for the ranking of different models while preventing overfitting, is straightforward to compute from the output of our Bayesian calibration procedure, as it is based on the likelihood values of the posterior distribution. 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 the gravity model instead of mobile phone data to determine human mobility. Model A (described in Section 2.2.4 and Equation 2.1) was selected as the best performing candidate. Descriptions of all other candidate models, as well as results of the model comparison, are reported in Section hereafter.

#### Processes

We evaluated models which consider mobile phone data to be unbiased ( $c = 1$ ) (model B) or with a fixed initial condition (one initial case in each arrondissement of Diourbel, model C). In model D the mobile phone data are used to determine temporal variations in population distribution due to human mobility, and thus to account for overcrowding, whereas the mobility fluxes between individual arrondissements are not considered. Model E includes mobility fluxes but not the overcrowding. Model F does not take into account human mobility at all. The absence of fluxes in models D and F leads to a *de facto* uncoupling of the local models, which makes it necessary to calibrate the initial number of cases and equally distributing them among arrondissements. Model G does not take precipitation into account, and Model H adapts the gravity model (see Chapter 1 for implementation) instead of mobile phone data to determine the human mobility.

#### Alternative ways of reconstructing the mobility of 2005

The mobility matrix extracted from the mobile phone records contains not only information about mobility during exceptional events such as the Grand Magal de Touba (GMdT) or other pilgrimages (Figure 2.3), but also about seasonal and sub-seasonal variations of mobility. To exploit this information, and to test if its use leads to improvements in model performance with respect to our baseline mobility matrix presented in Section 2.2.3, we compared 5 alternative mobility matrices by incorporating them into our best performing model (A) and recalibrating:

- I. The baseline mobility matrix, as presented in Section 2.2.3. We averaged the human mobility matrix throughout 2013, excluding only the periods of the two occurrences of



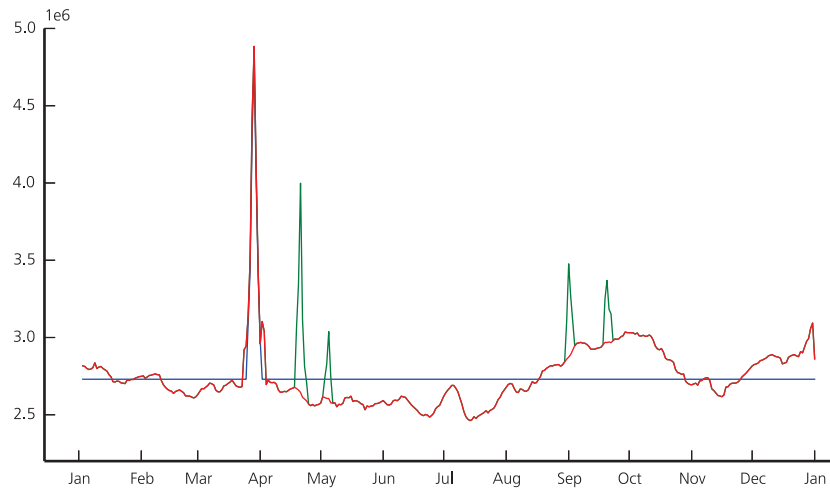


Figure 2.4 – Evolution of the total number of moving people (i.e. people leaving their home arrondissement) throughout 2005 estimated using methods I (blue), II (red) and III (green) (Section 2.2.7). The first spike, present in all scenarios, corresponds to the GMdT. The four spikes present only in scenario 3 correspond to the following events (chronological order): Gamou de Tivaouane, Magal de Porokhane, Magal de Kazu Rajab, Magal de Darou Mouhty.

the GMdT. We used the resulting mobility matrix for all days in 2005, except for the period of the GMdT ( $\pm 3$  days), which in turn was assigned the mobility of the GMdT that had taken place in December 2013.

II. The purpose of this mobility matrix is to test if seasonal and sub-seasonal variations of mobility other than the GMdT should also be considered in our model (instead of assuming constant mobility throughout the year, except from the GMdT). We thus first extracted the seasonal signal, defined as the mobility matrix excluding the effect of the GMdT as well as other important and clearly identifiable mobility pulses. We followed the following procedure:

- Exclusion of both editions of the GMdT and replacement by the average mobility of the previous and following weeks.
- Idem with four other clearly identifiable mobility pulses caused by the following events: Gamou de Tivaouane, Magal de Porokhane, Magal de Kazu Rajab, Magal de Darou Mouhty.
- Idem with four irregularities present in the mobility matrix, identified by visual inspection, which might correspond to cell phone network breakdowns or electric power cuts.
- Application of a 7 day moving average to smooth out the weekly cycle and get a purely seasonal signal.

## Chapter 2. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks

---

- Determine individual contribution of the GMdT by subtracting the seasonal signal from the original mobility matrix during the period of the event.

The contribution of the GMdT in December 2013 was then added to the seasonal signal during the period of the GMdT 2005 to obtain a mobility matrix for the entire year 2005.

III. As 2., except that, in addition to the GMdT, we also added the contribution of four other events (Gamou de Tivaouane, Magal de Porokhane, Magal de Kazu Rajab, Magal de Darou Mouhty) to the seasonal signal.

IV. As 1., but without considering the GMdT, e.g. constant mobility throughout the year.

V. As 2., but without considering the GMdT, e.g. considering the seasonal variation of mobility only.

Variants IV and V have been included to evaluate if the mobility during the GMdT is essential for our model to perform well. A comparison of the countrywide number of mobile people every day according to variants I to III is shown in Figure 2.4.

### 2.2.8 Potential effects of local interventions

To investigate the potential effects of local interventions, we run 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.

## 2.3 Results

Figure 2.3 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, Section 2.2.3), and during which the number of people traveling outside their home arrondissement almost doubles with respect to an average day. Figure 2.3B 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 (Figure 2.3C). 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 six times its usual population.

Model results and estimated uncertainties of the best performing candidate, model A, are shown in Figure 2.5 (total cases and the regions most severely hit) and 2.7 (all regions). The values of the calibrated parameters are reported in Table 2.2. 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

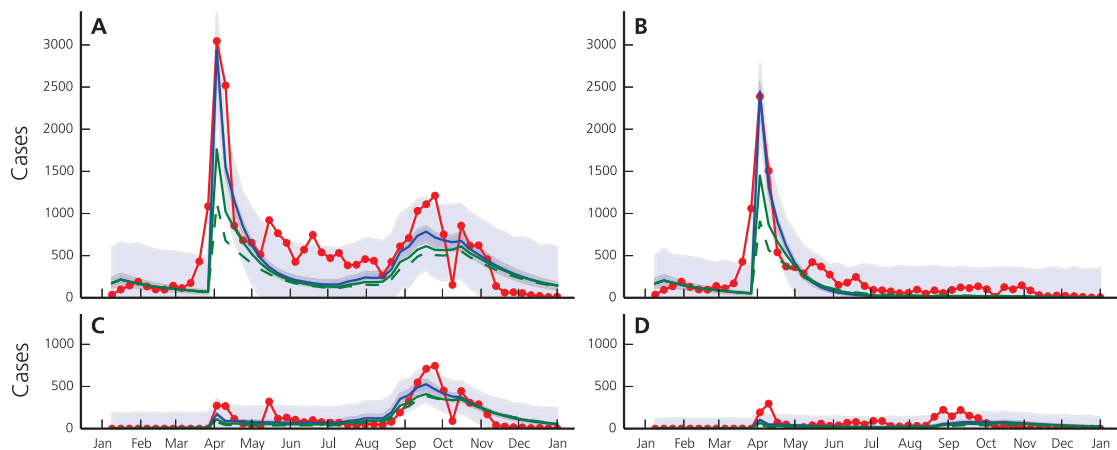


Figure 2.5 – Reported (red line) and modeled number of new cases per week for the entire country of Senegal (A), and for the regions of Diourbel (B), Dakar (C) and Thiès (D). Blue lines correspond to runs of model 2.1 with the best posterior parameter set. Shaded bands correspond to the 2.5–97.5% percentiles 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 Grand Magal de Touba are also shown.

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. Figure 2.6 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.

### 2.3.1 Model selection

Results show that models including human mobility (to estimate fluxes between arrondissements and/or overcrowding effect) clearly outperform model F, which does not account for those effects (Figure 2.8 Table 2.3). The gravity model does not provide an appropriate description of human mobility for the case of this study. Indeed, model H provides a reasonable fit for the region with the highest number of cases, however, lacking a proper description of spatiotemporal variations of human mobility, it does not correctly capture the spread of the disease to other regions. This also leads to convergence problems and unrealistic posterior parameter values. The overcrowding effect alone leads to a model performing relatively well (D), which, however, does not correctly reproduce the spread of the epidemic, and which is outcompeted by models accounting also for human mobility fluxes between the arrondissements (A and B). The bias correction of mobility data leads to a slight improvement in model performance, as does the calibration of the initial number of infected in Diourbel. Interesting insight is provided by results of model G, implying that the overall results can still be reason-

## Chapter 2. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks

Table 2.3 – Comparison of models including different mechanisms (see Section 2.2.7) using the Deviance Information Criterion (DIC) as well as the coefficient of determination  $R^2$ , computed including weekly case data from all or from one selected region.

Model	Mobility fluxes <sup>a</sup>	Overcrowding <sup>a</sup>	Precipitation	Bias correction <sup>b</sup>	IC <sup>c</sup>	Parameters	Log-likelihood <sup>d</sup>	DIC	$R^2$ overall	$R^2$ Dakar	$R^2$ Diourbel	$R^2$ Thiès
A	+	+	+	+	c	6	-3256	6533	0.77	0.72	0.78	0.20
B	+	+	+	-	c	5	-3328	6669	0.71	0.68	0.69	0.41
C	+	+	+	+	d	5	-3279	6573	0.75	0.73	0.75	0.04
D	-	+	+	+	a	6	-3308	6631	0.72	0.71	0.75	-0.47
E	+	-	+	+	c	5	-3595	7204	0.25	0.10	0.08	-0.02
F	-	-	+	-	a	4	-3641	7295	0.12	0.60	-0.19	0.31
G	+	+	-	+	c	5	-3302	6615	0.73	0.22	0.81	0.13
H <sup>e</sup>	+	+	+	-	c	7	-3459	6943	0.54	-0.50	0.73	-0.50

<sup>a</sup> Note that human mobility fluxes and overcrowding both depend on human mobility estimates, but can be taken into account separately. See Section 2.2.7 for more details.

<sup>b</sup> Absence of bias correction with  $c = 1$ .

<sup>c</sup> Initial number of infected. c: calibrated (only Diourbel), a: calibrated (all arrondissements), d: fixed (only Diourbel).

<sup>d</sup> Highest log-likelihood value in the posterior sample.

<sup>e</sup> For model H, human mobility has been determined using a gravity model (e.g. *Rinaldo et al.* [2012]) instead of deriving it from mobile phone data.

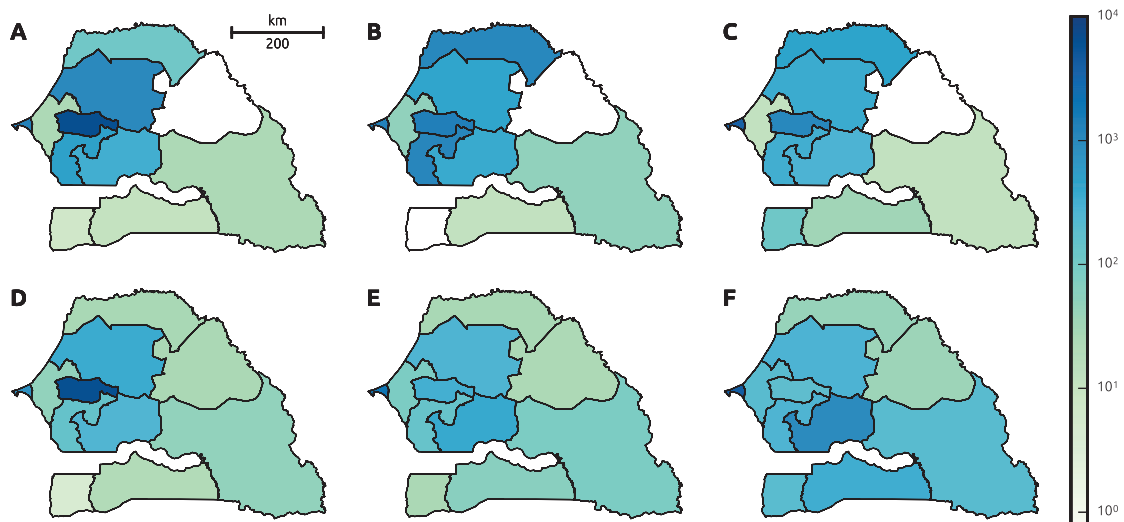


Figure 2.6 – Spatial distribution of reported (A,B and C) and modeled (D,E and F) cases from March 28 to May 29, the first weeks after the Grand Magal de Touba (A and D), from June 30 to September 4 (B and E), and from September 5 to December 31 (C and F).

ably good without rainfall, but that its addition is necessary to be able to capture the autumn peak in Dakar (among other regions), previously associated to rainfall [*de Magny et al.*, 2012].

The comparison of model performance under different assumptions about mobility shows that the inclusion of the GMdT in the mobility matrix is crucial for the model to perform well, but that including the baseline seasonality as well as additional but smaller mass gatherings decreases the model's ability to reproduce the data (Table 2.4). This might be due to the fact that the seasonality in 2005 was different from the one in 2013, or that mobility was of high importance only during the GMdT but not during the rest of the year.

### 2.3.2 Impact of reduced transmission during the GMdT

We tested several scenarios to quantify the influence of control measures that could possibly have attenuated the 2005 cholera epidemic. Modeling results suggest Touba as a promising focal point for actions aimed at containing disease spread. Therefore, we focused our attention on interventions localized (in space and/or) time around the GMdT. We assume that by providing additional sanitation facilities and clean drinking water, a reduction of disease transmission through a reduced bacterial shedding rate (parameter  $\theta$ ), also accounting for a reduced contamination of environmental water bodies with fecal matter [*Kühn et al.*, 2014], and through a reduced rate of exposure to contaminated water (parameter  $\beta$ ) can be achieved. We run our best performing model reducing both relevant parameters by a varying percentage in Touba either only during the GMdT ( $\pm 10$  days) or throughout the year. According to our model, the number of averted cases increases with the duration of the interventions not only in Diourbel but throughout Senegal (Figure 2.9 and Table 2.4). When reducing exposure and

**Chapter 2. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks**

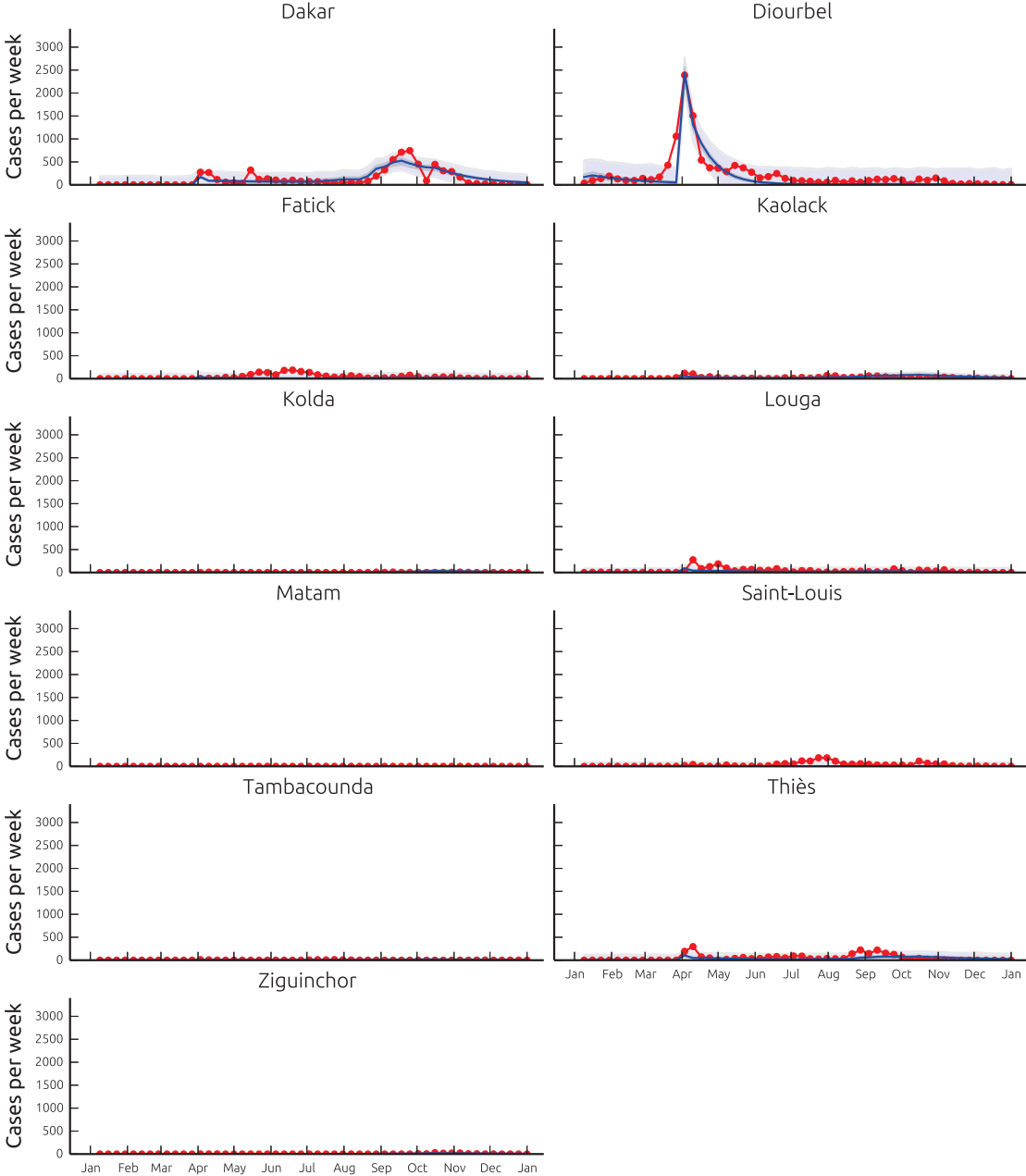


Figure 2.7 – Reported (red line) and modeled number of cases per week (blue line), resulting from model A run with the best parameter set (Table 2.2), in the 11 regions of Senegal. Shaded bands show the 2.5–97.5% percentile bounds of the uncertainty related to parameter estimation (dark blue) and of the total uncertainty assuming Gaussian, homoscedastic error (light blue).

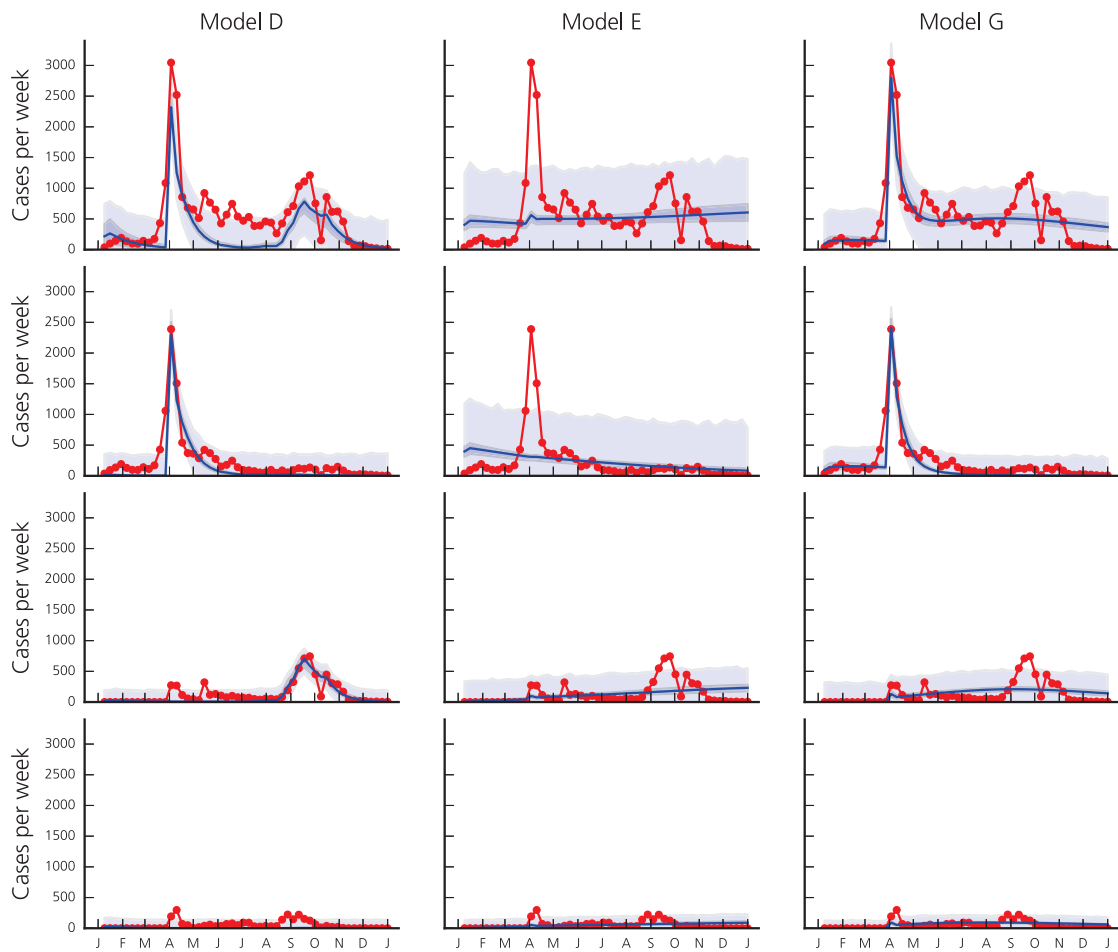


Figure 2.8 – Reported (red line) and modeled number of new cases per week for the entire country of Senegal (first row), and for the regions of Diourbel (second row), Dakar (third row) and Thiès (last row) according to models D (first column, not including mobility fluxes), E (second column, not including the overcrowding effect) and G (third column, not including rainfall). Blue lines correspond to model runs with the best posterior parameter set. Shaded bands shown 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).

## Chapter 2. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks

Table 2.4 – Recalibration of model A using different mobility matrices (Section 2.2.7 and Figure 2.4).

Mobility matrix	Log-likelihood <sup>a</sup>	DIC	R <sup>2</sup> overall	R <sup>2</sup> Dakar	R <sup>2</sup> Diourbel	R <sup>2</sup> Thiès
I <sup>b,c</sup>	-3256	6533	0.77	0.72	0.78	0.20
II <sup>c,d</sup>	-3306	6632	0.73	0.38	0.79	-0.25
III <sup>c,d,e</sup>	-3267	6559	0.76	0.74	0.77	0.15
IV <sup>b</sup>	-3474	6974	0.51	-0.50	0.66	-0.51
V <sup>d</sup>	-3579	7229	0.30	-0.49	0.32	-0.51

<sup>a</sup> Highest Log-likelihood value in the posterior sample.

<sup>b</sup> constant mobility

<sup>c</sup> GMdT

<sup>d</sup> seasonal variations of mobility

<sup>e</sup> other events

contamination only during the GMdT the number of averted cases grows less rapidly than when applying the reductions throughout the year, which might be the result of less cases and a smaller bacterial concentration in Touba just before the GMdT.

## 2.4 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 using a gravity model.

The high temporal and spatial resolution of the mobility patterns extracted from mobile phone data allows to identify 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 by providing safe drinking water or sanitation for a higher number of people, could have had important effects, preventing numerous infections



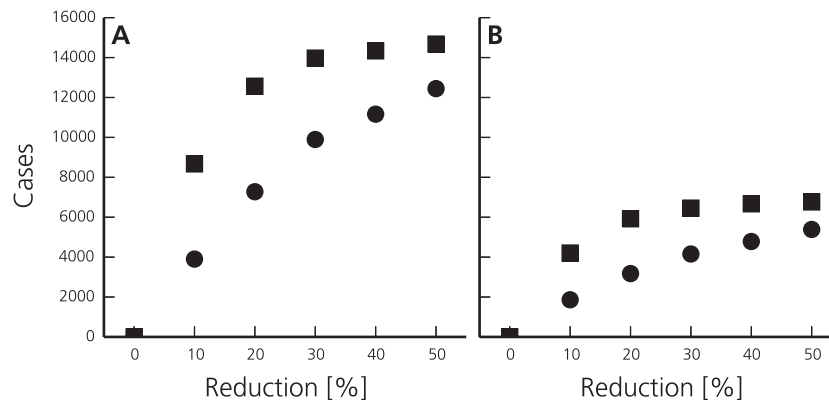


Figure 2.9 – Modeled number of total averted cases in 2005 overall Senegal (A) and in the region of Diourbel (B) when reducing the rates of exposure to contaminated water and bacterial shedding by a varying percentage in Touba during the GMdT only ( $\pm 10$  days, circles) and throughout the year (squares).

not only locally but throughout the whole country.

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. While 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 of demographic stochasticity when the number of infected is low, but also biased case reporting and/or identification [*de Magny et al.*, 2012] 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 (Section 2.2.6).

Even if mobile phone data provides an excellent source of information about human mobility, several downsides still exist. One of them is the strong assumptions (Section 2.2.3 and *Mari et al.* [2015b]) 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 [*Wesolowski et al.*, 2012a] 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 income tend to travel more often and further, leading to overestimations of frequency and distance of trips

## Chapter 2. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks

Table 2.5 – Estimated (according to model A) number of cases and percentage of averted cases in all regions in 2005 when reducing the rates of exposure to contaminated water and bacterial shedding in Touba by 10% or 20% during the GMdT or during the entire year.

Region	Modeled cases	10% during GMdT [%]	20% during GMdT [%]	10% entire year [%]	20% entire year [%]
Dakar	7062	12	21	20	34
Diourbel	8276	23	38	51	72
Fatick	579	19	41	47	68
Kaolack	1609	17	41	50	71
Kolda	435	16	47	54	74
Louga	940	25	47	53	73
Matam	90	28	48	55	74
Saint-Louis	104	28	47	53	69
Tambacounda	341	24	47	54	71
Thiès	1802	21	42	45	65
Ziguinchor	229	3	37	46	62
<b>Senegal</b>	<b>21467</b>	<b>18</b>	<b>34</b>	<b>40</b>	<b>59</b>

[Wesolowski *et al.*, 2013b]. In our study, this effect was at least partially addressed by the introduction of a parameter (Section 2.2.4) accounting for the underrepresentation of people staying at their home node in the mobility fluxes extracted from mobile phone call records. 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 [Boone, 2003], 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 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 (Section 2.2.7), 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 behavioural change of individuals in 2005, in turn affecting the disease dynamics [Funk *et al.*, 2010; Mari *et al.*, 2012a; Meloni *et al.*, 2011].

In conclusion, we demonstrate that mobile phone records allow for an unprecedentedly 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

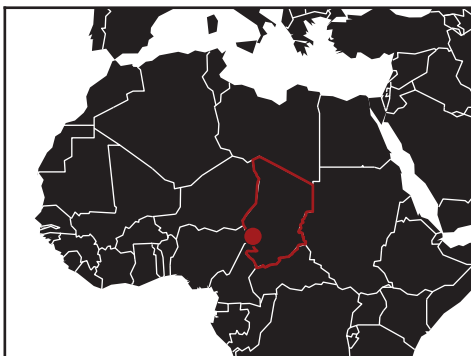
models is straightforward and may lead to higher accuracy with respect to other approaches, as human movement patterns can be directly derived from data rather than inferred from models (e.g. gravity or radiation).



### 3 The impact of case-centered interventions in response to cholera outbreaks: a modeling study

*This chapter is currently in preparation to be published as an article. The study has been designed by Flavio Finger, Enrico Bertuzzo and Andrew Azman<sup>1,2</sup>, with the help of Andrea Rinaldo, Francisco Luquero<sup>3,4</sup> and Justin Lessler<sup>1</sup>. Data has been collected and provided by MSF<sup>2</sup>, Epicentre<sup>4</sup> and the Ministry of Public Health of the Republic of Chad. Flavio Finger has implemented and calibrated the model, run all simulations and produced the results. The results have been interpreted by Flavio Finger, Andrew Azman and Francisco Luquero. Flavio Finger has written the draft of the article in close collaboration with Andrew Azman, all other authors have contributed to the writing.*

#### Overview



**Country:** Chad

**Study domain:** N'Djamena (capital)

**Surface:** 225 km<sup>2</sup>

**Population:** 993 492 (2013)

**Cholera:** Outbreaks occurring at irregular intervals.

**Period studied:** April 2011 to April 2012

**Number of reported cases:** 4352

<sup>1</sup>Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America

<sup>2</sup>Médecins sans Frontières, Geneva, Switzerland

<sup>3</sup>Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America

<sup>4</sup>Epicentre, Paris, France

### **Chapter 3. The impact of case-centered interventions in response to cholera outbreaks: a modeling study**

---

#### **Abstract**

Cholera cases tend to cluster in space and time, with the risk of another person becoming infected increasing the closer she/he lives to a case and the closer to the time when the case becomes infectious. Therefore, interventions, including improved water, sanitation and hygiene, oral cholera vaccine, and prophylactic antibiotics, may most efficiently curb epidemics when applied in proximity to reported cases, leading to fewer cases and deaths and ultimately save scarce public health resources during an epidemic. Here, we investigate the effectiveness and efficiency of case-centered interventions using a model fit to detailed spatiotemporal cholera data from Chad. We explore key determinants of impact, including the mix of different interventions, 'ring' size, and timing of interventions to provide practical guidance on how these interventions might be used.

We developed a spatially explicit, individual-based stochastic transmission model and calibrated it to both the epidemic curve and a random set of 1585 spatial coordinates of case households from a large 2011 cholera outbreak in N'Djamena, Chad, using an approach based on Approximate Bayesian Computation. We simulated 1000 epidemics, and explored the impact of intervention scenarios using different combinations of antibiotics, oral cholera vaccine and/or water, sanitation and hygiene measures. The benefits of interventions were compared with regard to the number of averted cases and the resource utilization. We show that, compared to allocating interventions to a large fraction of the population in mass campaigns, case-centered interventions are more resource-efficient. For a campaign using oral cholera vaccine and starting at the epidemic peak, case-centered interventions at a radius of 100 m are 64 [95% credible interval (CI) 11 – 518] times more efficient than a mass intervention campaign in terms of the number of people targeted per case averted. Combining several types of case-centered interventions even increases their benefits. Further, results show that the earlier during an epidemic interventions are started, the higher is the number of averted cases, and the more resource-efficient is the intervention. Optimal ranges to apply case-centered interventions depend on the type of intervention. In the case of antibiotics, the optimal distance is around 30 m – 45 m, whereas for oral cholera vaccine and water, sanitation and hygiene interventions, the intervention impact increases with distance and saturates around 70 m – 100 m.

We show that case-centered interventions are an effective and efficient way to fight cholera epidemics and present a complementary approach to mass intervention campaigns. The former imply heavier logistics than the latter but offer a more resource-efficient way to achieve similar results.

#### **3.1 Introduction**

Cholera continues to be a major public health threat in developing countries, with over 170 000 cases and 1300 deaths reported globally in 2015, the largest proportion occurring in Africa

[World Health Organization, 2016a]. The true number of cases is likely orders of magnitude higher due to under-reporting [Ali *et al.*, 2015]. Many cities in Sub-Saharan Africa are regularly ravaged by cholera outbreaks [Rebaudet *et al.*, 2013a, b], causing disruption and thwarting societal and economic development. These cities may act as sub-national, national and international hubs of disease spread due to regular travel and migration [Finger *et al.*, 2016] and quickly quelling cholera outbreaks in these areas may avert substantial cases both within the cities and afar.

The mainstay of cholera prevention and control has been improved access to safe water, sanitation and hygiene (WaSH) and case management. WaSH interventions include a heterogeneous mix of interventions ranging from provision of safe water through infrastructure or point-of-use water treatment tools to latrine building and hygiene behavior change measures. Antibiotics have been used to shorten the duration of shedding in cholera cases and, in some instances, to provide short-term prophylaxis for their household contacts [Guévert *et al.*, 2007; Leibovici-Weissman *et al.*, 2014]. Recently, oral cholera vaccines (OCVs) have been added to this arsenal, with the increased availability as a result of the global cholera vaccine stockpile and the addition of a new affordable WHO-prequalified vaccine [Desai *et al.*, 2016; Martin *et al.*, 2012]. OCVs have been shown to be safe, immunogenic and protective, with two-doses (the standard regimen) lasting up to five years [Bhattacharya *et al.*, 2013] and a single-dose at least 6-months [Azman *et al.*, 2016; Qadri *et al.*, 2016], a similar time scale to many cholera epidemics.

These tools are used either preventively [Abubakar *et al.*, 2015] in areas deemed at high risk for cholera transmission, or reactively, in response to a cholera outbreak [Azman *et al.*, 2016; Ciglenecki *et al.*, 2013; Ivers *et al.*, 2015]. Typically, control measures are given to the population at-large within areas of high risk through mass campaigns, although targeted interventions to households or neighborhoods, including delivery of antibiotics and WaSH [Farmer *et al.*, 2011; Guévert *et al.*, 2007; Piarroux *et al.*, 2009; Reveiz *et al.*, 2011], are common and part of the national control policy in a number of countries (e.g., South Sudan, Chad, Kenya).

Spatiotemporal clustering of cholera cases, at distances ranging from tens to hundreds of meters, has been observed in many cholera outbreaks in endemic and epidemic areas [Ali *et al.*, 2016; Blackburn *et al.*, 2014; Carrel *et al.*, 2009; Debes *et al.*, 2016a; Luquero *et al.*, 2011; Snow, 1855; You *et al.*, 2013]. Previous analyses have shown that cases were significantly clustered up to a radius of 320 m within the first five days of a case presenting for care during the 2011 epidemic in N'Djamena, Chad [Azman *et al.*, in prep.]. In the past, this clustering has been attributed to common risk factors in those living close to one another in addition to the risk of transmission often being higher the closer one lives to an infected individual. Intervention strategies targeting disease hotspots [Azman and Lessler, 2015], particularly vulnerable neighborhoods and camps [Abubakar *et al.*, 2015] and other communities are known and have been successfully applied in the past. Limited literature exists, however, on case-centered reactive intervention campaigns, which take advantage of the inherent spatiotemporal clustering of cholera cases by preventively targeting people living within a given

### **Chapter 3. The impact of case-centered interventions in response to cholera outbreaks: a modeling study**

---

distance around reported cases. Such strategies could not only present efficient alternatives to reactive mass intervention campaigns in outbreak situations, where resources may be limited or their availability delayed, but may also be used as a complementary approach to the latter when only low numbers of cases are present, such as during initial phases of epidemics or to shorten the tail of a declining epidemic.

Here, we use a spatially explicit, stochastic, individual-based model fit to detailed spatiotemporal data from a 2011 cholera epidemic in Chad to evaluate the potential impact of case-centered reactive interventions. Specifically, we explore whether OCV, WaSH and/or prophylactic antibiotics (combined or separate) have the potential to curb epidemics, and how mixes of interventions and intervention timing may influence the ultimate impact of these control measures. We provide practical suggestions on how to optimize case-centered interventions and compare their efficiency to more traditional mass-campaigns to help decision makers weigh the costs and benefits of these different approaches.

## **3.2 Materials and Methods**

### **3.2.1 Case study and data**

During the 2011 cholera epidemic in N'Djamena, Chad, field staff from Médecins sans Frontières (MSF) collected the household coordinates of all suspected cholera case presenting at the main cholera treatment center/unit, by visiting people at their home starting on June 22 (Figure 3.1). In early October, when the caseload began to rapidly increase, household coordinates were collected for every third patient, through the end of the epidemic in December. The resulting dataset, combining the overall epidemic curve of suspected cholera cases with GPS coordinates of patients homes, has been described previously [Azman *et al.*, in prep.].

### **3.2.2 Spatial setup and population distribution**

The domain of our model is the city of N'Djamena, Chad, subdivided into regular grid cells (30 m by 30 m). The remotely sensed built-up density (Figure 3.2) [Esch *et al.*, 2012, 2013] was used as a proxy for the small scale spatial population density.

Every inhabitant ( $N = 993\,492$ ) was randomly assigned to a grid cell with a probability proportional to the estimated average built-up density of the cell. The euclidean distances between the centers of the two cells were used to compute the value of the infection kernel (3.4) between persons living in distinct grid cells were taken to be  $\frac{1}{d}$ . A distance of 10 m was assumed between two persons living in the same grid cell.



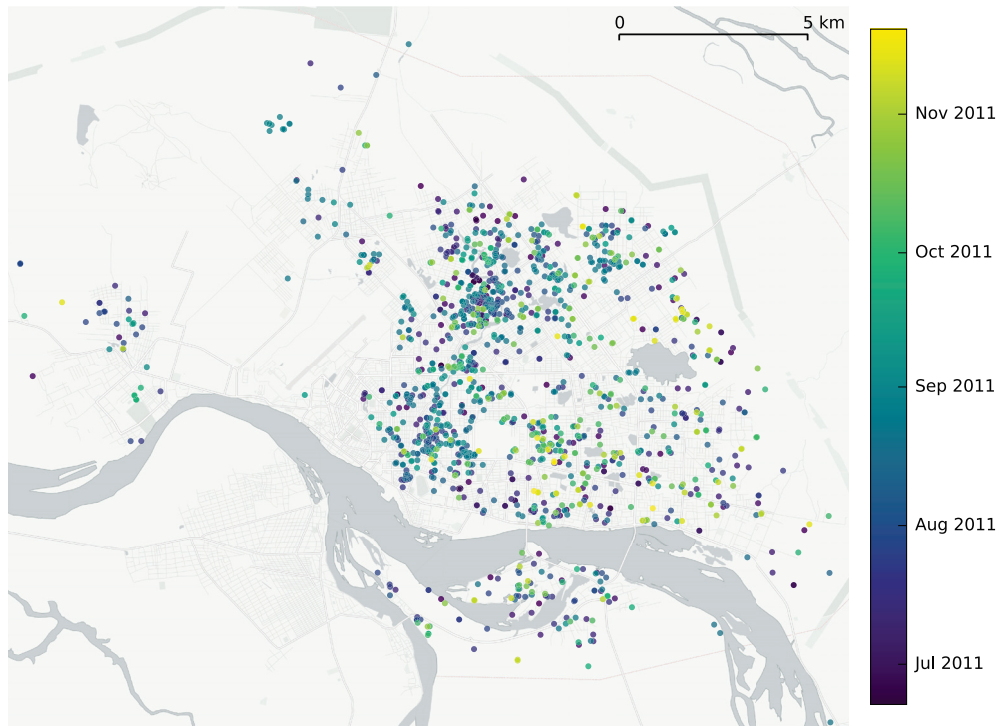


Figure 3.1 – Map of the city of N'Djamena with the locations of cases with available GPS coordinates by time of reporting. (Background map: Tiles by CartoDB, under CC BY 3.0. Data by OpenStreetMap, under ODbL.)

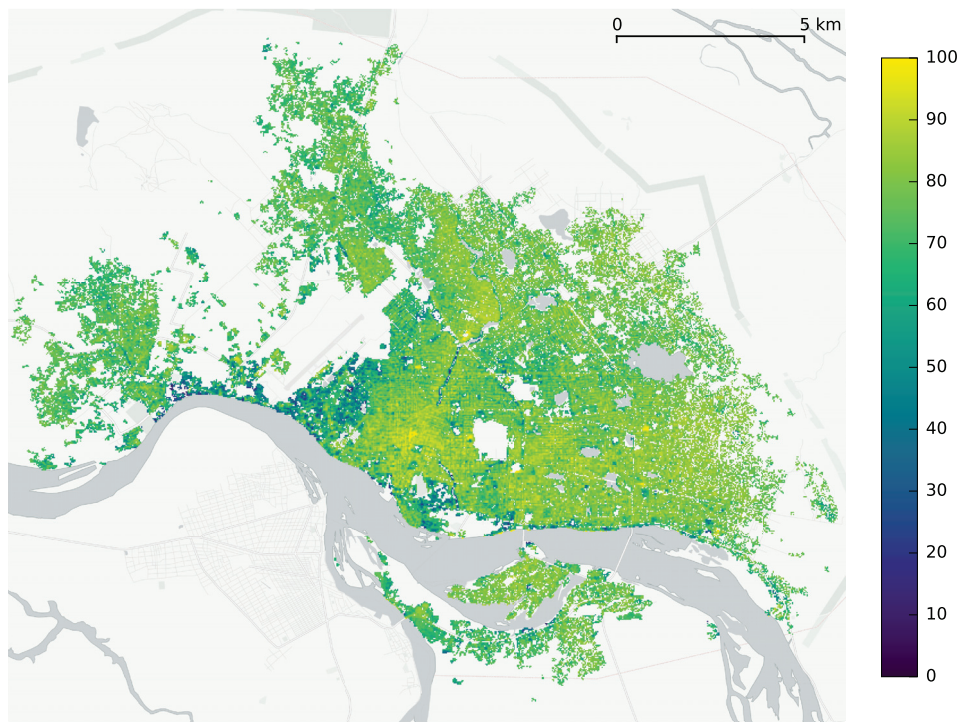


Figure 3.2 – Proxy for population density. Map of the city of N'Djamena with built-up density (in percent) [Esch *et al.*, 2012, 2013] of each 30 m by 30 m grid cell. Values equal to 0 or located outside the city boundary are transparent. (Background map: Tiles by CartoDB, under CC BY 3.0. Data by OpenStreetMap, under ODbL.)

### Chapter 3. The impact of case-centered interventions in response to cholera outbreaks: a modeling study

---

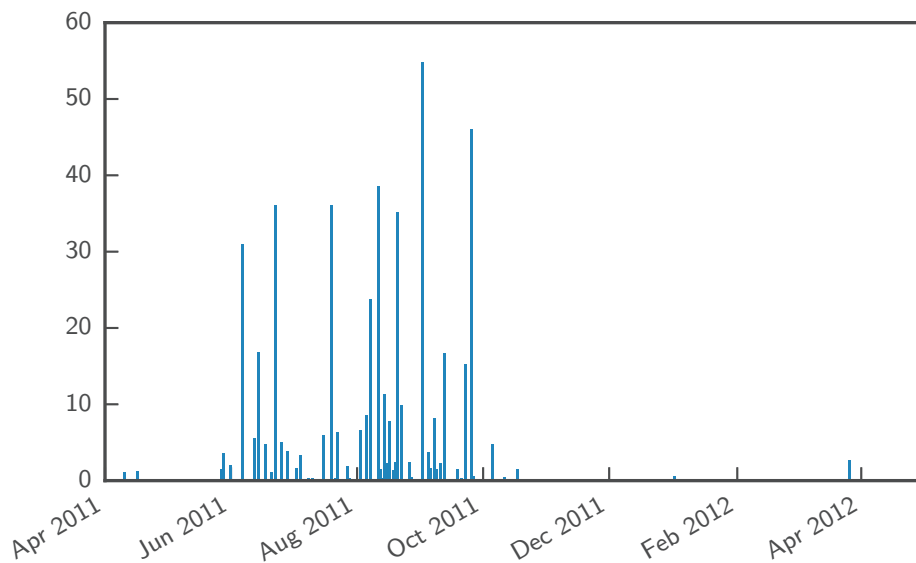


Figure 3.3 – Daily precipitation depth in N’Djamena from April 2011 to April 2012.

#### 3.2.3 Rainfall data

Daily precipitation (Figure 3.3) was obtained from the NASA TRMM Version 7 Daily Precipitation Estimates [Huffman *et al.*, 2010]<sup>5</sup>.

#### 3.2.4 Quantification of spatiotemporal clustering

We use the  $\tau$  statistic introduced by Lessler *et al.* [2016], which has been shown to be a good measure of global spatial clustering for epidemiological applications [Grabowski *et al.*, 2014; Salje *et al.*, 2012, 2016b], to quantify spatiotemporal clustering of cholera cases. Here, we consider two cases to be potentially transmission related (e.g. to potentially share a recent common ancestor) if they occurred within a time interval of 0 – 4 days from each other (using upper bounds between 2 and 6 days gives similar results, see Figure 3.4).  $\tau$  is then defined as the relative risk that a person in a given distance range  $[d_1, d_2]$  from a disease case also becomes a case that is potentially transmission related (i.e. to become infected within a time interval of 0 – 4 days), compared to the risk of any person in the population becoming a potentially transmission related case.  $\hat{\tau}(d_1, d_2)$  can be computed by dividing the estimated odds ratio  $\hat{\theta}(d_1, d_2)$  of the number of potentially transmission related cases against non-transmission related cases within  $[d_1, d_2]$  by the same odds ratio computed for the whole domain  $\theta(0, \infty)$

<sup>5</sup>[http://iridl.ldeo.columbia.edu/SOURCES/.NASA/.GES-DAAC/.TRMM\\_L3/.TRMM\\_3B42/.v7/.daily/.precipitation/X/15.0/15.25/RANGEEDGES/Y/12/12.25/RANGEEDGES/T/\(01%20Apr%202011\)\(01%20May%202012\)RANGEEDGES/](http://iridl.ldeo.columbia.edu/SOURCES/.NASA/.GES-DAAC/.TRMM_L3/.TRMM_3B42/.v7/.daily/.precipitation/X/15.0/15.25/RANGEEDGES/Y/12/12.25/RANGEEDGES/T/(01%20Apr%202011)(01%20May%202012)RANGEEDGES/), accessed on June 29, 2016

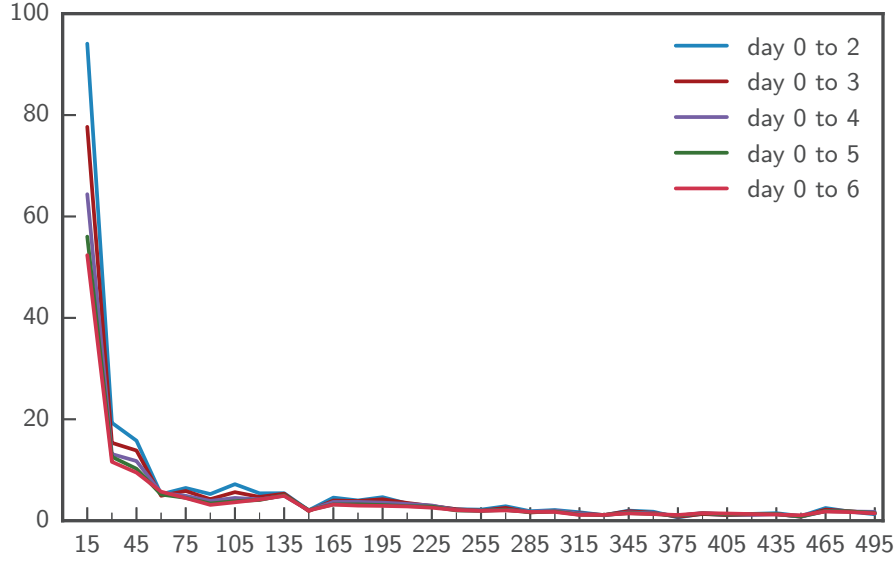


Figure 3.4 – Comparison of relative risk  $\tau$  evaluated in 15m distance ranges for different time ranges.

[Lessler *et al.*, 2016]:

$$\hat{\theta}(d_1, d_2) = \frac{\sum_i \sum_j \mathbf{I}_1(i, j)}{\sum_i \sum_j \mathbf{I}_2(i, j)} \quad (3.1)$$

$$\hat{\tau}(d_1, d_2) = \frac{\hat{\theta}(d_1, d_2)}{\hat{\theta}(0, \infty)}, \quad (3.2)$$

where  $\mathbf{I}_1(i, j)$  denotes an indicator function which is equal to one if cases  $i$  and  $j$  are within the distance range  $[d_1, d_2]$  from each other and within the time interval of 0 – 4 days, and zero otherwise.  $\mathbf{I}_2(i, j)$  is an indicator function which is equal to one if cases  $i$  and  $j$  are within the distance range  $[d_1, d_2]$  from each other but not transmission related (i.e. with the time interval between the cases longer than 4 days), and zero otherwise.

### 3.2.5 Epidemiological model

We employ a spatially explicit, individual-based stochastic epidemiological model (see e.g. *Keeling and Rohani [2008]*) with a timestep  $\Delta t$  of 1 day. The  $N$  individuals in the model space are assigned random positions according to the population distribution (Section 3.2.2) and can either be susceptible ( $S$ ), exposed ( $E$ ), infected ( $I$ ) or recovered ( $R$ ). Every individuals state is

### Chapter 3. The impact of case-centered interventions in response to cholera outbreaks: a modeling study

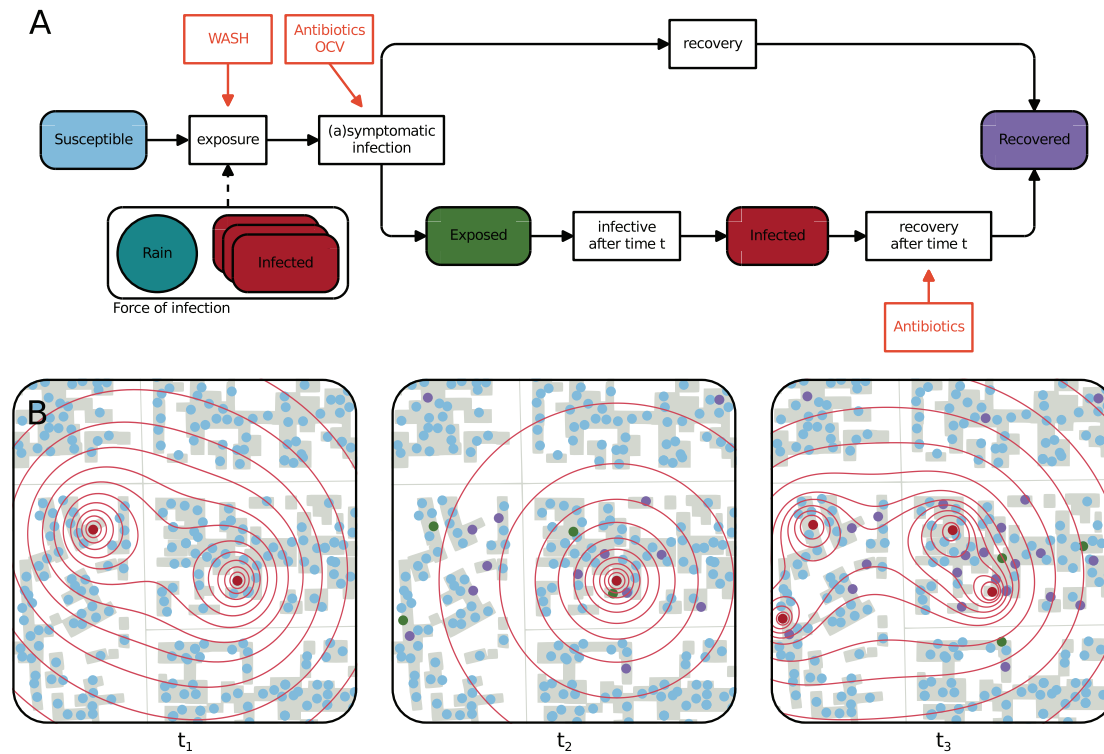


Figure 3.5 – Schematic representation of the epidemiological model and evolution of the infectious state of inhabitants of a neighborhood. A: Flow chart of the model representing the different epidemiological states a person can be in and the processes that lead to a change of state. The force of infection acting on a susceptible depends on the number of infected and the distance to each of them as well as on rainfall during the last 10 days. Orange boxes represent pathways through which interventions (antibiotics, OCV and WaSH) influence the processes in the model. B: Schematic representation of the evolution of the epidemiological state of the inhabitants of a neighborhood in N’Djamena during three time-steps. The closer susceptible people (blue) live to an infected (red), the higher the force of infection (red contours) they face. Susceptibles can get symptomatically infected, which means that they get exposed (green) for a given duration before becoming symptomatically infected (red) and thus contribute to the force of infection, or asymptotically infected, in which case they are assumed to recover (purple). Infected recover after a given duration. Between time  $t_1$  and  $t_2$  one infected recovered and four susceptibles got exposed. At  $t_3$ , the only infected at  $t_2$  has recovered and all exposed have become symptomatic.

tracked during the outbreak (Figure 3.5A). Demographic processes, like births and deaths, are assumed to be negligible during the short time course of the outbreak. A power-law-shaped, isotropic transmission kernel originating from the position of every infected accounts for the spread of the disease (Figure 3.5B):

$$K(d) = cd^{-a} \quad (3.3)$$

$$c = \left( \frac{1}{N} \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N d_{i,j}^{-a} \right)^{-1}, \quad (3.4)$$

where  $d$  is the distance from the infected.  $a$  is a calibration parameter which affects the shape of the kernel and  $c$  a normalization constant which corresponds to the inverse of the spatial mean of the kernel in the case where every individual is infected, and ensures that the values of the kernel are comparable for different parameter values. The force of infection  $F_i$  affecting a susceptible  $i$  depends on his position in the model space relative to infected individuals and can be computed by taking the sum of the kernels originating from all infected evaluated at the position of  $i$  and multiplying it with individual exposure parameters, as well as a term accounting for rainfall, which has been shown to be an important environmental driver of cholera epidemics in several settings [*de Magny et al.*, 2012; *Finger et al.*, 2016; *Gaudart et al.*, 2013; *Ngwa et al.*, 2016; *Rinaldo et al.*, 2012]:

$$F_i(t) = \beta_i (1 + \lambda r(t)) \sum_{\substack{j=1 \\ j \neq i}}^N \mathbf{I}_I(j) K(d_{i,j}), \quad (3.5)$$

where  $\beta_i$  is an individual exposure parameter,  $\mathbf{I}_I(j)$  is an indicator function whose value is equal to 1 if individual  $j$  is infected and 0 otherwise.  $\lambda$  is a parameter that multiplies daily precipitation  $r(t)$ .

Exposure events are assumed to follow a Poisson process with rate  $F_i(t)$ . The interarrival time of exposure events follows a Poisson process. The resulting probability of susceptible  $i$  being exposed during  $\Delta t$  is given in (3.6). A fraction  $\sigma_i$  of infections is symptomatic. Asymptomatic individuals recover immediately after exposure, their contribution to the environmental bacterial concentration is assumed to be negligible [*Kaper et al.*, 1995; *King et al.*, 2008; *Nelson et al.*, 2009]. Symptomatic individuals stay in the exposed state for a time  $t_E$  and in the infected state for a time  $t_I$ , which are drawn from gamma distributions according to (3.7) [*Azman et al.*, 2013] and (3.8) [*Kaper et al.*, 1995].

$$P(S_i \rightarrow E_i) = (1 - e^{-F_i(t)\Delta t}) \quad (3.6)$$

$$t_E \sim \Gamma(2, 0.5) \quad (3.7)$$

$$t_I \sim \Gamma(10, 0.5). \quad (3.8)$$

### Chapter 3. The impact of case-centered interventions in response to cholera outbreaks: a modeling study

---

A timestep  $t$  of a model simulation thus consists of the following steps:

1. update states of individuals from  $E$  to  $I$  or from  $I$  to  $R$  if their  $t_E$  or  $t_I$  is reached
2. compute  $F_i(t)$  (3.5) at the position of every susceptible  $i$
3. use (3.6) to determine susceptibles who get exposed by drawing a uniform random number  $p_i$  for each of them. If  $p_i < P(S_i \rightarrow E_i)$  the individual gets exposed
4. for every exposed individual, draw a random number  $q_i$  to determine if he becomes symptomatic or asymptomatic. If  $q_i < \sigma_i$ ,  $i$  is symptomatic the individual goes to the infected class  $I$ , otherwise she/he recovers and goes to class  $R$ .

To accelerate the model runs for big populations and large areas, the model space is a discrete grid (Section 3.2.2) and the convolution between the kernel and the distance matrix is done using the Fast Fourier Transform.

#### 3.2.6 Initial conditions

An initial number of 25 randomly chosen individuals in the model space are set to be symptomatically infected. They are assumed to be infectious for a period  $t_I$  according to (3.8), starting from a point in time between 4 and 1 days before the start of the model. The remaining part of the population is assumed to be susceptible.

#### 3.2.7 Calibration

In the absence of treatments, parameters values are assumed to be the same for all individuals (i.e.  $\beta_i = \beta$  and  $\sigma_i = \sigma$ ). Four free parameters of our model ( $\sigma$ ,  $\beta$ ,  $a$  and  $\lambda$ ) were calibrated to match the characteristics of the real epidemic in N'Djamena.

We employed a Python implementation of the Approximate Bayesian Computation Population Monte Carlo (ABC-PMC) algorithm [Akeret *et al.*, 2015; Beaumont *et al.*, 2009], using a multivariate normal kernel with optimal local covariance matrix [Filippi *et al.*, 2013]. Two summary statistics were used: the sum of squared residuals (3.9) on the reported number of cases (Figure 3.6A) and on the  $\tau$  statistic over 4 different distance ranges (0 m–15 m, 15 m–45 m, 45 m–105 m and 105 m–225 m) and with a time range of 0–4 days (Figure 3.6B). The distance ranges have been chosen to fit the spatial discretization of the model domain (Section 3.2.2). The calibration was run with 512 particles, which were accepted if the summary statistics were under an initial threshold of 100 000 and 1800 respectively. After each calibration step the thresholds were adapted to the 85<sup>th</sup> percentile of the summary statistic values taken by the particles of the previous step [Akeret *et al.*, 2015].

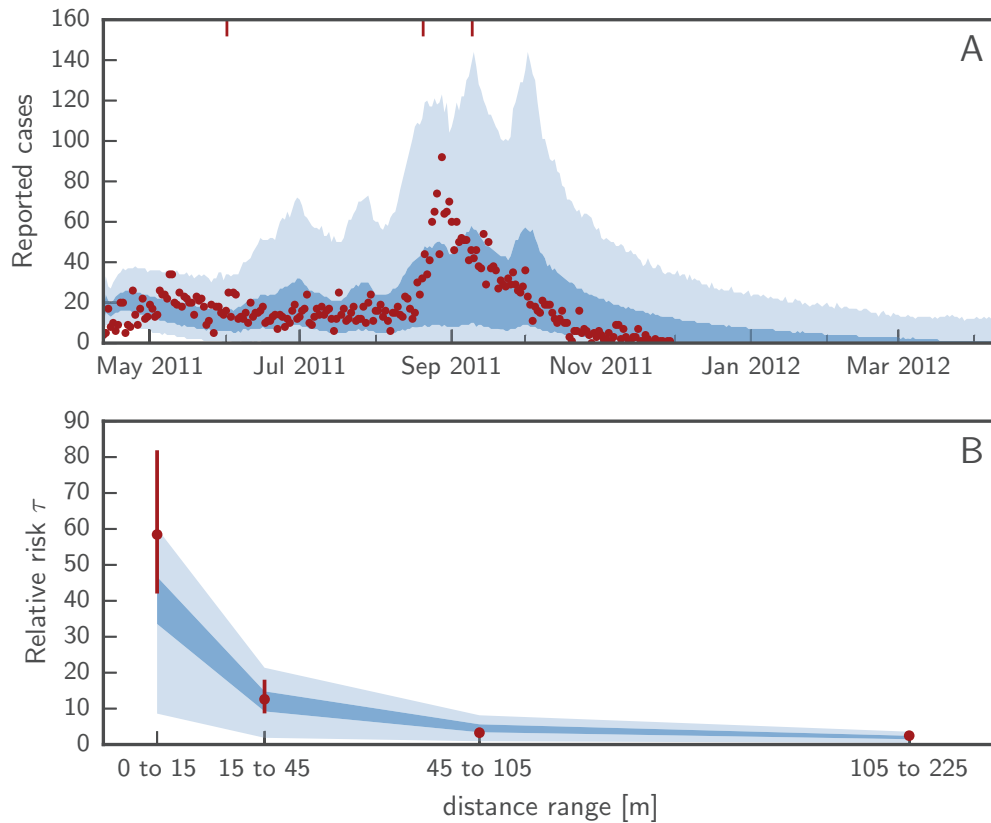


Figure 3.6 – Calibration results. Daily new cholera cases (A) and relative risk ( $\tau$  statistic) for days 0 – 4 (B) computed from reported cases (red) and according to the calibrated model (blue). The shaded areas represent the interquartile range (dark blue) and the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles (light blue) of the posterior computed from 1000 samples. 95% confidence intervals on the  $\tau$ -statistic represent the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles of the bootstrapped estimate distribution with 1000 iterations. Red ticks at the top axis of (A) represent the three times when interventions start.

$$ssr(x, \hat{x}) = \sum_{i=1}^N (x_i - \hat{x}_i)^2 \quad (3.9)$$

### 3.2.8 Simulation

To run a simulation, a parameter set is drawn from the posterior distribution with probability proportional to the posterior weight, and used to run the model. All results and figures presented are derived from a set of 1000 simulation runs. Note that the outcome of two model runs with identical parameter sets may differ because of stochastic processes. Simulations are

### Chapter 3. The impact of case-centered interventions in response to cholera outbreaks: a modeling study

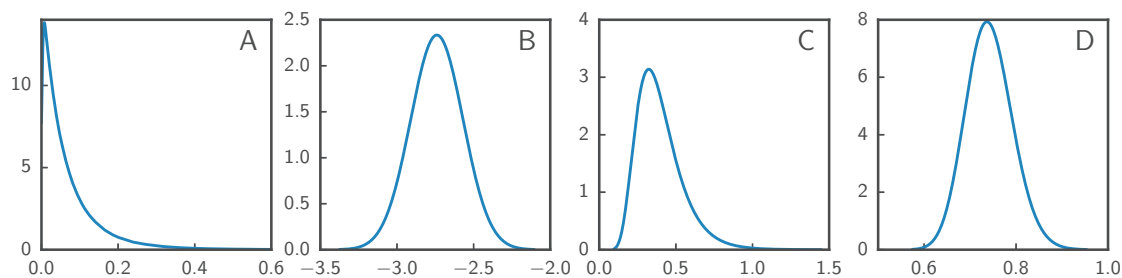


Figure 3.7 – Distributions of intervention parameters. Intervention parameters related to antibiotics (reduction of symptomatic fraction (A) and reduction of duration of shedding (B)), OCV (reduction of symptomatic fraction (C)) and WaSH (reduction of exposure (D)). A has been obtained from *Lewnard et al.* [2016] directly, whereas B (normal), C (log-normal) and D (log-normal) have been fit to the corresponding confidence intervals given in Table 3.1.

run either until the end of the epidemic (no infected or exposed present in the model), or up to a maximal duration of 1 year.

#### 3.2.9 Intervention strategies

We consider three different types of preventive interventions and combinations thereof: the administration of a single-dose of antibiotics, household scale water, sanitation and hygiene (WaSH) measures and the administration of a single-dose of oral cholera vaccine (OCV). The implementation of the effects of those interventions are described below. Values and references are summarized in Table 3.1 and Figure 3.7. In addition to these types of interventions and their possible combinations, we also consider different strategies to select people to benefit from the interventions, and different points in time when the application of interventions start. Appendix A.1 summarizes all interventions tested for this study together with their outcome.

##### Antibiotics

We consider the joint effect of two mechanisms of protection against cholera by antibiotics: a reduced probability of acquiring infection [*Revez et al.*, 2011] and a lower probability to get symptoms if exposed [*Echevarria et al.*, 1995]. As studies have not yet quantified the combined effect [*Revez et al.*, 2011], we follow *Lewnard et al.* [2016] and estimate the joint effect from individual effects by multiplying them. Finally, the joint effect is translated into a reduction in the symptomatic fraction (parameter  $\sigma$ ) by a random factor 0.045 [95% CI 0.001 – 0.296].

Antibiotics have also been found to reduce the duration of bacterial shedding [*Leibovici-Weissman et al.*, 2014; *Lewnard et al.*, 2016], which we model through an additive reduction of  $t_I$  by -2.74 [95% CI -3.07 – -2.40] days.



Table 3.1 – Effects of the three types of interventions as implemented in the model.

	Antibiotics	OCV	WaSH
Reduction of the symptomatic fraction (parameter $\sigma$ ) <sup>a</sup>	0.045 [0.001 – 0.296] <sup>c</sup>	0.37 [0.18 – 0.76] <sup>c</sup>	–
Reduction in exposure (parameter $\beta$ ) <sup>a</sup>	–	–	0.74 [0.65 – 0.85] <sup>c</sup>
Reduction of the infectious period [days] <sup>b</sup>	-2.74 [-3.07 – -2.40] <sup>c</sup>	–	–
Lag	–	7 days <sup>d</sup>	–
Duration of the effect	2 days	– <sup>e</sup>	– <sup>e</sup>
References	<i>Khan et al.</i> [2002]; <i>Qadri et al.</i> [2016] <i>Fewtrell et al.</i> [2005] <i>Leibovici-Weissman et al.</i> [2014]; <i>Lewnard et al.</i> [2016]; <i>Reveiz et al.</i> [2011]		

<sup>a</sup> Reduction factors of symptomatic fraction and exposure are multiplied with the corresponding parameter.

<sup>b</sup> The reduction of the infectious period is subtracted from the value without intervention.

<sup>c</sup> 95% confidence intervals are given in brackets.

<sup>d</sup> To account for a possible earlier onset of OCV protection [*Azman et al.*, 2016], a shorter lag of 2 days is tested alternatively.

<sup>e</sup> The effect is assumed to last at least as long as the current epidemic.

### **Chapter 3. The impact of case-centered interventions in response to cholera outbreaks: a modeling study**

---

We estimate that the beneficial effects of antibiotics last for 2 days, as the drug concentration in stools has been shown to be sufficient to eliminate *V. Cholerae* during this period of time after the administration of a single-dose of Azithromycin in a clinical trial [Khan *et al.*, 2002].

#### **OCV**

The administration of a single-dose of OCV affects the chances of an exposed individual  $i$  to become symptomatic (e.g. to get severe cholera) through the reduction of parameter  $\sigma_i$  by a multiplication with 0.37 [95% CI 0.18 – 0.76], which corresponds to one minus the vaccine efficacy reported by Qadri *et al.* [2016] for severe cholera episodes. As beneficial effects of the vaccine have only been confirmed after a lag of 7 days [Qadri *et al.*, 2016], we assume that the vaccine takes effect only 1 week after administration. To account for the fact (at least partial) protection by OCV might actually occur much earlier [Azman *et al.*, 2016], we also included a scenario where this lag time is shortened to 2 days.

#### **WaSH**

WaSH interventions reduce the probability of individuals to get exposed to an infectious dose of *V. Cholerae*. In the model, this is achieved via the reduction of the exposure parameter  $\beta_i$  for targeted individuals  $i$  by a multiplication with a factor 0.74 [95% CI 0.65 – 0.85], reported by Fewtrell *et al.* [2005] for water quality interventions in (peri-)urban settings.

#### **Combined interventions**

We also consider combinations between the three main type of interventions (Appendix A.1). This is achieved by simultaneously applying the estimated effects of several interventions to the targeted population.

#### **Duration of intervention effects**

Whereas the duration of the effects of antibiotics is short (2 days, see above), we consider that protection from OCV and WaSH lasts (at least) until the end of the current epidemic. This implies that people who benefited from OCV or WaSH interventions once don't need to be treated again. In the case of antibiotics, as the effect vanishes rapidly, we consider two scenarios, one in which every person can get antibiotics only once during the epidemic, and one in which a single person can be allocated antibiotics several times, with a minimal interval of 2 weeks.

### Uncertainty of intervention effects

To propagate uncertainty regarding intervention effects we use the distributions shown in Figure 3.7, which have been obtained by fitting normal or log-normal distributions to the reported 95% confidence intervals shown in Table 3.1 or directly from the cited references. During the simulation, a different set of reduction parameters is drawn for every person treated.

### Intervention timing

We consider three different scenarios as to when interventions start (Figure 3.6):

**early** (day 50, May 31, 2011) interventions are launched during the flat phase early in the epidemic,

**peak** (day 130, August 19) around the peak of the epidemic,

**late** (day 150, September 8) after the epidemic peak, during the recession phase.

During simulation, we assume that intervention scenarios are only started when at least 10 new cases were reported during the week before the start date.

#### 3.2.10 Allocation strategies

##### Case-centered allocation

In addition to the different kinds of interventions we consider different intervention strategies. The first strategy takes advantage of the clustering of cases. It consists in targeting people with an increased risk of getting exposed to *V. cholerae* because they are living within a given distance (in time and space) to a known case. Every time a case gets reported (e.g. when a person in the model changes from the exposed (*E*) to the symptomatically infected state (*I*)), people who live within a distance of 100 m of the reported case's home are targeted by the intervention. In N'Djamena a cluster of this radius typically consists of 100 – 500 people (Figure 3.8). We also evaluate the effect of reducing this radius to 70 m, 45 m, 30 m and 15 m, measured within the gridded model space (Figure 3.9). To account for the fact that an intervention team visiting the target area will not be able to reach all inhabitants, because they might be absent or might not agree to receive preventive treatment or not comply with WaSH measures, we consider that a random sample of 70% of the people who live within the designated area can be effectively reached. In addition, we account for the delay between the reporting of the initial case and the deployment of an intervention team to the corresponding cluster by drawing it from a distribution given in Table 3.2, considering that all clusters can be targeted within 7 days counting from the reporting of the initial case (day 0), with the mode on day 2.

**Chapter 3. The impact of case-centered interventions in response to cholera outbreaks: a modeling study**

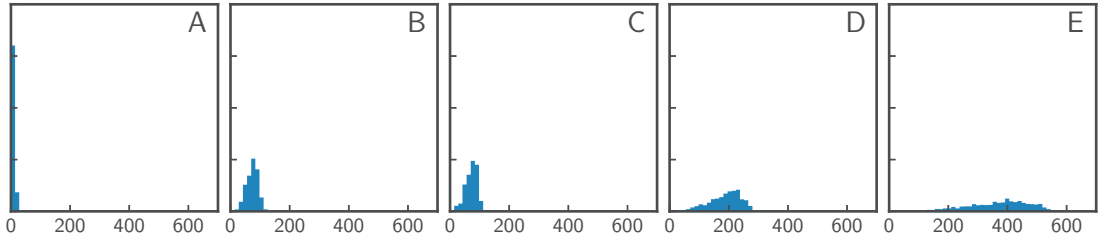


Figure 3.8 – Histograms of the number of people living within a circle with radius 15 m (A), 30 m (B), 45 m (C), 70 m (D) and 100 m (E) in N’Djamena obtained by sampling the population distribution at 1000 random points.

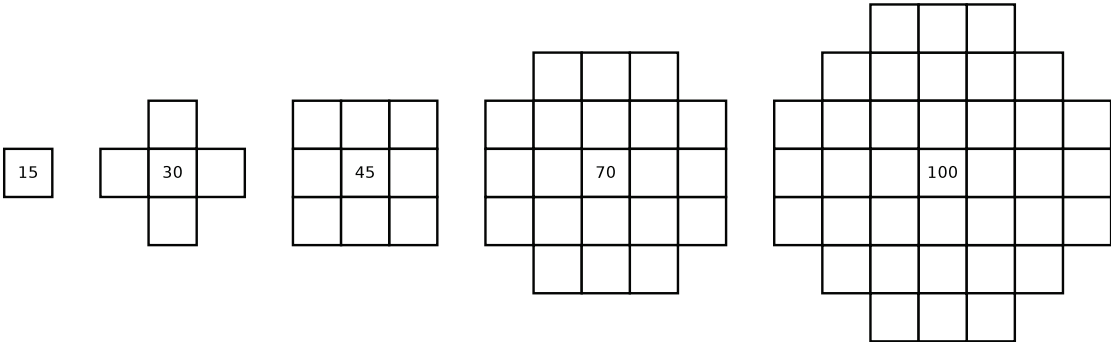


Figure 3.9 – Circles of given radius (in m) as implemented in the regular model grid with. The side length of each square is 30 m.

We assume that interventions continue until the end of the epidemic (i.e. no more reported cases) or the maximal duration of the simulation (Section 3.2.8).

**Random allocation to the same number of people**

To estimate the comparative advantage of case-centered interventions, we also simulated a second strategy, where we allocated the interventions at the same starting time as we would have in the targeted interventions and to the same total number of people for each simulation run, though randomly in space. We consider that all interventions can be administered within 14 days from the start date.

Table 3.2 – Distribution of the delay between the reporting of the initial case in a cluster and the deployment of an intervention team.

Day	0	1	2	3	4	5	6
Probability	0.05	0.23	0.35	0.23	0.1	0.02	0.02

### Random allocation to 70% of the population

To estimate the effect of randomly allocated interventions to a high number of people (i.e. mass interventions), we simulated a strategy where we allocated the interventions at the same time as we would have in the targeted interventions, but randomly to 70% of all people living in the city. We consider that all interventions can be administered within 14 days from the start date.

## 3.3 Results

### 3.3.1 Calibration

The model was able to reproduce the key characteristics of the epidemic, i.e. the evolution of new cases over time and the spatiotemporal clustering of cases ( $\tau$ , Figure 3.6). The calibration was stopped after 14 steps because the posterior had reached a stable state (Figure 3.10).

### 3.3.2 Individual case-centered Interventions

All three primary interventions - WaSH, OCV and antibiotics - achieve a rapid decrease in the incidence of new cases when targeted to individuals living within 100 m of a suspected case. Antibiotics lead to the sharpest short-term drop due to the immediate high degree of protection, while the onset of protection by OCV is subject to a lag and the degree of protection offered by WaSH is lower relative to the other types of intervention (Figure 3.11). Because protection by OCV and WaSH is assumed to last for the whole epidemic, their overall benefit is higher, leading to the extinction of the simulated epidemics within 33 [95% CI 12 – 121] days (OCV) to 150 [95% CI 11 – 289] days (WaSH) when applying interventions early in the epidemic; 39 [95% CI 13 – 85] days (OCV) to 102 [95% CI 18 – 216] days (WaSH) when applying interventions around the epidemic peak; and 35 [95% CI 13 – 70] days (OCV) to 87 [95% CI 115 – 208] days (WaSH) when applying interventions after the epidemic peak. This effect of short versus long-lasting protection is also visible (Figure 3.11) when comparing the number of protected people who received antibiotics, which decreases after an initial increase, with the ones who received OCV or WaSH, where the level of protection saturates after an initial increase. Similar qualitative results were seen with people living at different distances being targeted (Table A.1).

The total number of averted cases varies and depends on the actual number of cases in the epidemic without intervention as well as on the timing and type of intervention (Figure 3.12A). Regardless of the intervention type, the sooner an intervention starts, the more cases are averted. Interventions that avert more cases in a shorter period of time and bringing the epidemic to a halt quickly require less resources (people and clusters targeted) (Figures 3.12B and 3.12C).

**Chapter 3. The impact of case-centered interventions in response to cholera outbreaks: a modeling study**

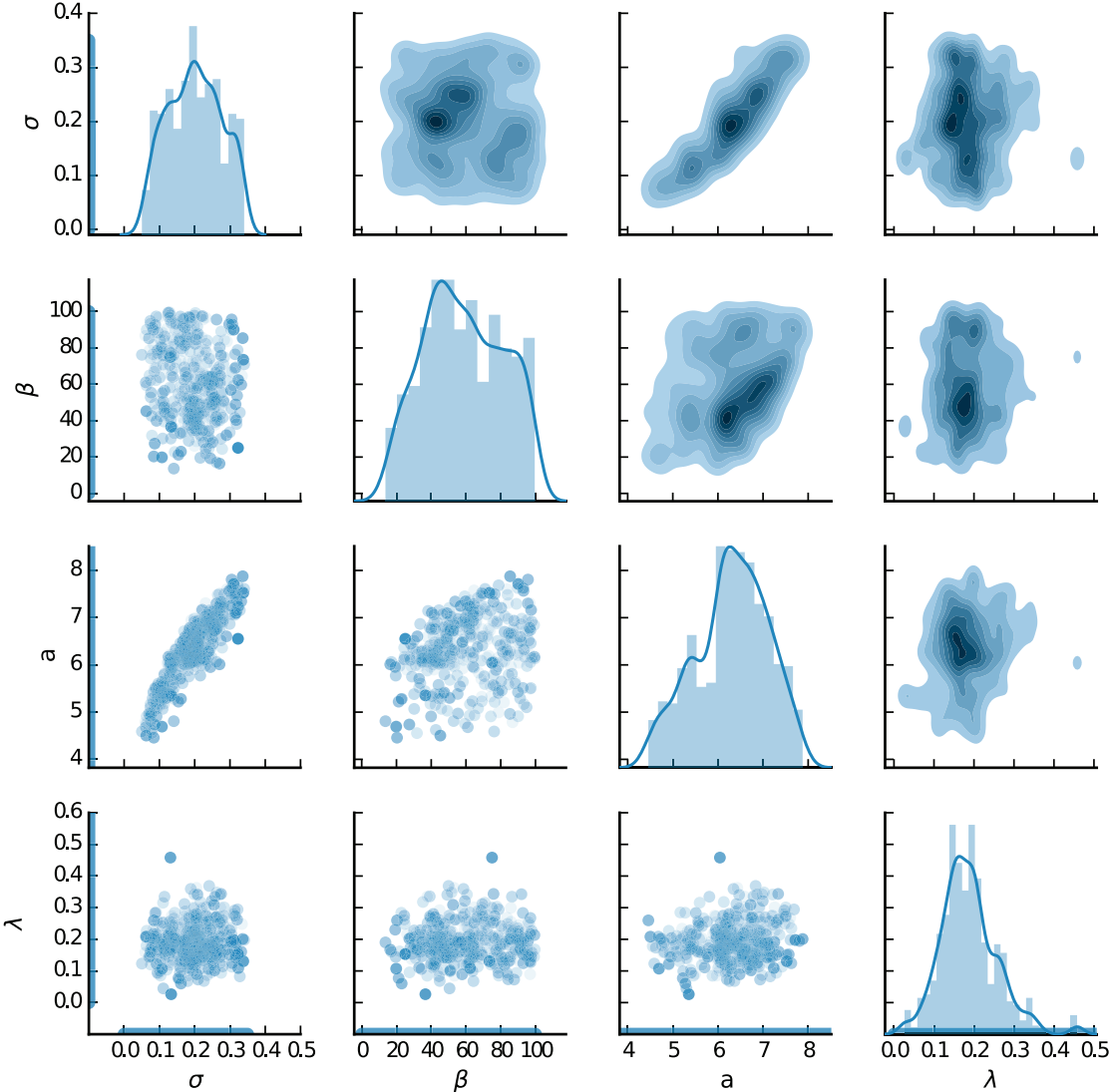
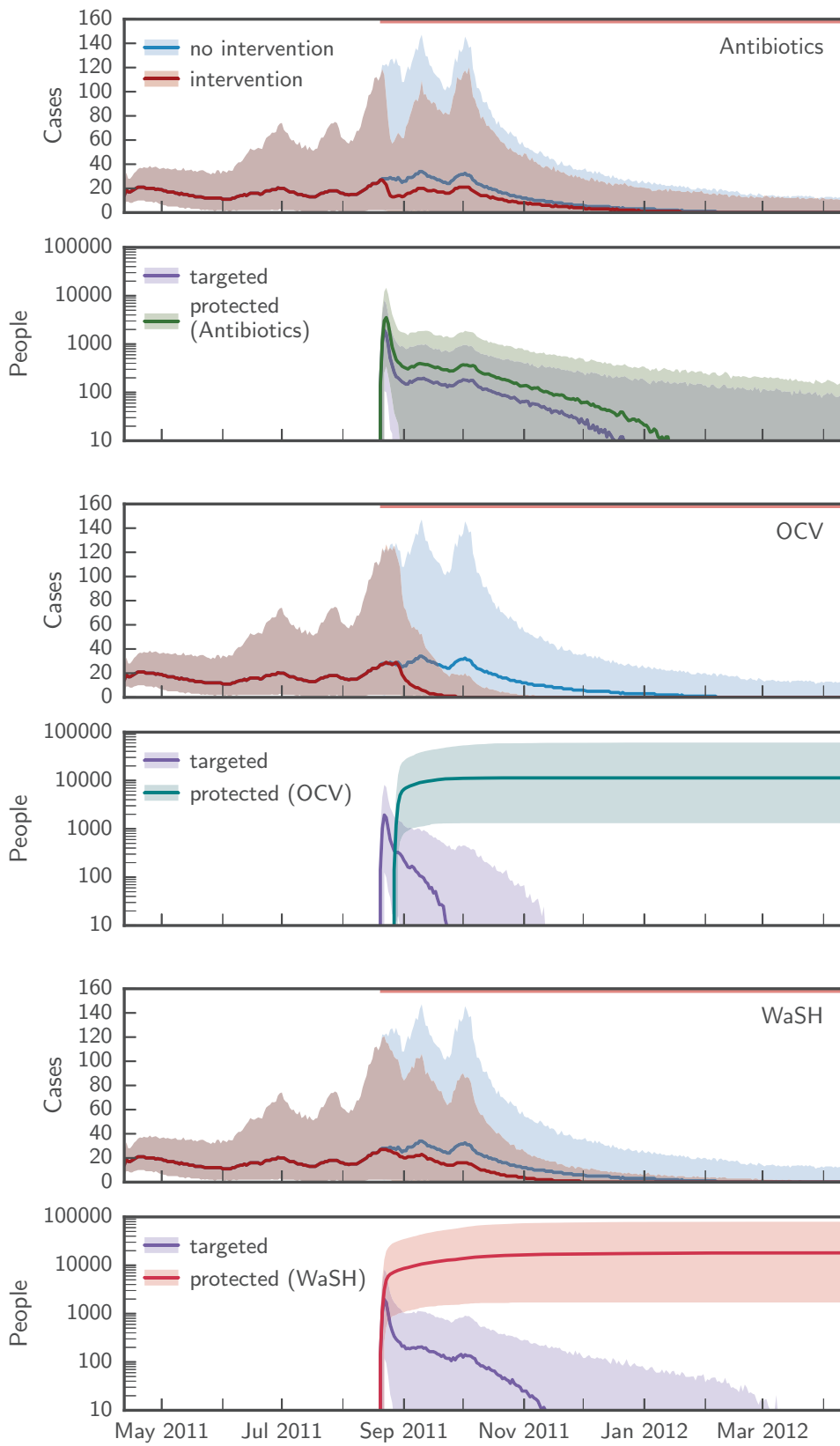


Figure 3.10 – Marginal posterior parameter distributions computed from 1000 samples. Blue shaded ranges along the axes show the intervals within which parameters were allowed to vary during calibration.

Figure 3.11 – Comparison of the simulated evolution of the epidemics with and without case-centered interventions. Upper panels show the simulated evolution of the epidemics without intervention and with case-centered allocation of antibiotics, OCV and WaSH within a 100 m radius starting at the epidemic peak. Lower panels show the corresponding number of people targeted during each time-step and the number of people protected by each intervention. Solid lines designate the median over all simulations, shaded areas the 2.5th and 97.5th percentiles. The red bars at the top of the axes mark the period during which interventions were applied.



### Chapter 3. The impact of case-centered interventions in response to cholera outbreaks: a modeling study

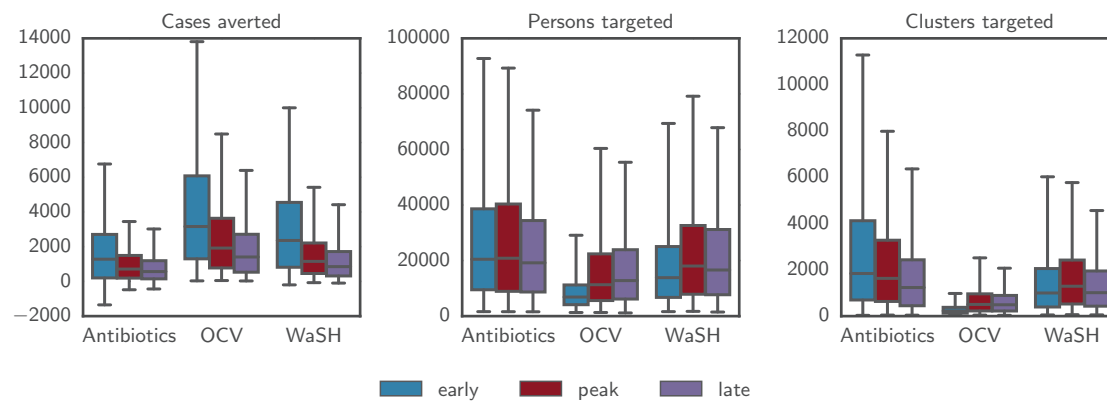


Figure 3.12 – Boxplots of the number of averted cases, the number of targeted persons and the number of targeted clusters predicted by the model for the three main intervention types with case-centered allocation in a 100 m radius starting at three different times. Whiskers mark the 2.5th and 97.5th percentiles. Negative numbers of averted cases, such as visible for antibiotics, are due to stochastic effects which arise when an intervention alters the course of a particular epidemic without halting it and leads to a higher number of cases at a later point in time.

#### 3.3.3 Combined case-centered Interventions

Combining different types of case-centered interventions leads to merging their main advantages, e.g. the immediate effect of antibiotics and the long-term protection of OCV, which can result in an even higher number of averted cases in even shorter time. Combined interventions with OCV and antibiotics in a radius of 100 m leads to the extinction of 97.5% of the simulated epidemics within only 64, 65 and 57 days when starting to intervene early, around the peak or late during the epidemic respectively (Figure 3.13). Combining WASH and OCV within a radius of 100 m, 97.5% of the epidemics vanish within 67, 70 and 57 days when starting interventions early, around the peak and late. For the combinations of antibiotics and WaSH, 97.5% of the epidemics end 273, 215 and 204 days, respectively, after the start of the intervention. Combining all three types of interventions results in epidemic extinction by 49, 58 and 48 days from the start of the intervention in 97.5% of the simulations (Figure 3.14). The faster a combination of interventions leads to the extinction of epidemics, the higher is the number of averted cases and the lower are the numbers of people and clusters targeted, even if the differences are only small for combinations of interventions that lead to a fast epidemic extinction (Figure 3.15).

#### 3.3.4 Distance range of case-centered interventions

For OCV and WaSH, the number of averted cases steadily increases with the radius of case-centered allocations, showing signs of saturation between 70 m and 100 m. In the case of antibiotics, however, the curves of averted cases are hump-shaped, with the maximum at 30 m, roughly twice as high as at 100 m, and similar to OCV at 100 m (Figure 3.16). This effect results from the short duration of the preventive protection offered by antibiotics, in combination



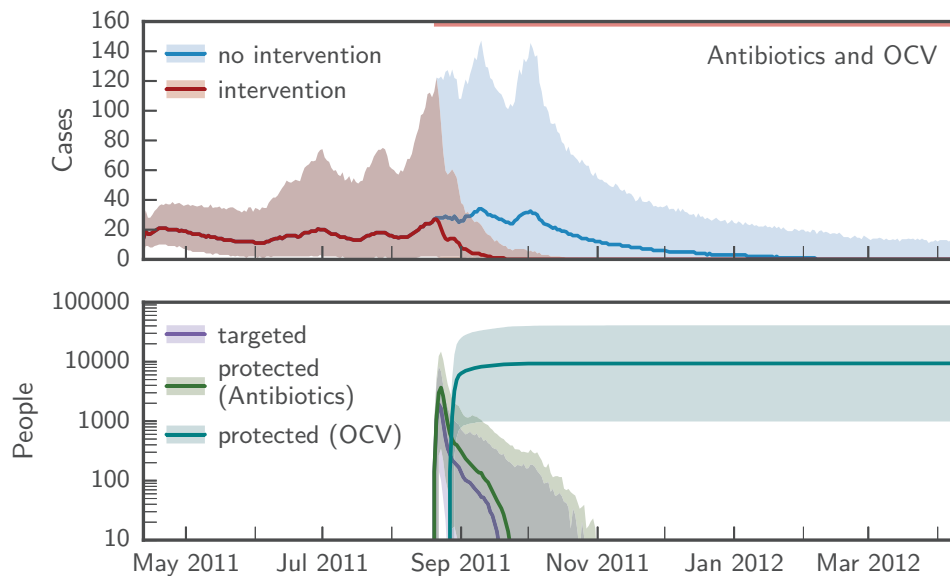


Figure 3.13 – Simulated evolution of the epidemics without intervention (blue) and with simultaneous Case-centered allocation (red) of antibiotics and OCV within a 100 m radius starting around the epidemic peak. The lower panel shows the number of people targeted during each time-step, and the number of people protected by antibiotics and OCV respectively. Solid lines show the median over all simulations, shaded areas the 2.5th and 97.5th percentiles. The red bar at the top of the figure marks the period during which the intervention was applied.

with the limitation that every person can only be targeted once. In addition, the epidemic wave arrives at distances farther from the primary case after the protective effect has vanished. The number of clusters targeted decreases with increasing ability to rapidly stop epidemics. The number of persons targeted over different radii is governed by two contrasting effects, a decrease with better performing interventions and an increase with larger cluster radii.

### 3.3.5 Efficiency compared to other modes of intervention allocation

The efficiency of a type of intervention was computed as the number of people targeted per case averted (Figure 3.17). Only model runs with a positive number of averted cases were considered. The most efficient type of case-centered intervention at a radius of 100 m is OCV, with an efficiency of 2 [95% CI 0.7 – 42] people to target per case averted when starting interventions early, 6 [95% CI 2 – 53] when starting around the peak, and 9 [95% CI 4 – 60] when starting late. For WaSH, the 97.5% CI of efficiency goes from 1 – 71 people to target per averted case when starting interventions early, 4 – 119 when starting around the peak and 5 – 170 when starting late. For antibiotics the 97.5% CI ranges from 2 – 168 when starting early, from 5 – 365 when starting around the peak and from 6 – 360 when starting late.

Mass intervention strategies, where a large proportion (i.e. 70% in our case) of the population is targeted in a short period of time can achieve similar numbers of averted cases (Figure 3.18),

**Chapter 3. The impact of case-centered interventions in response to cholera outbreaks: a modeling study**

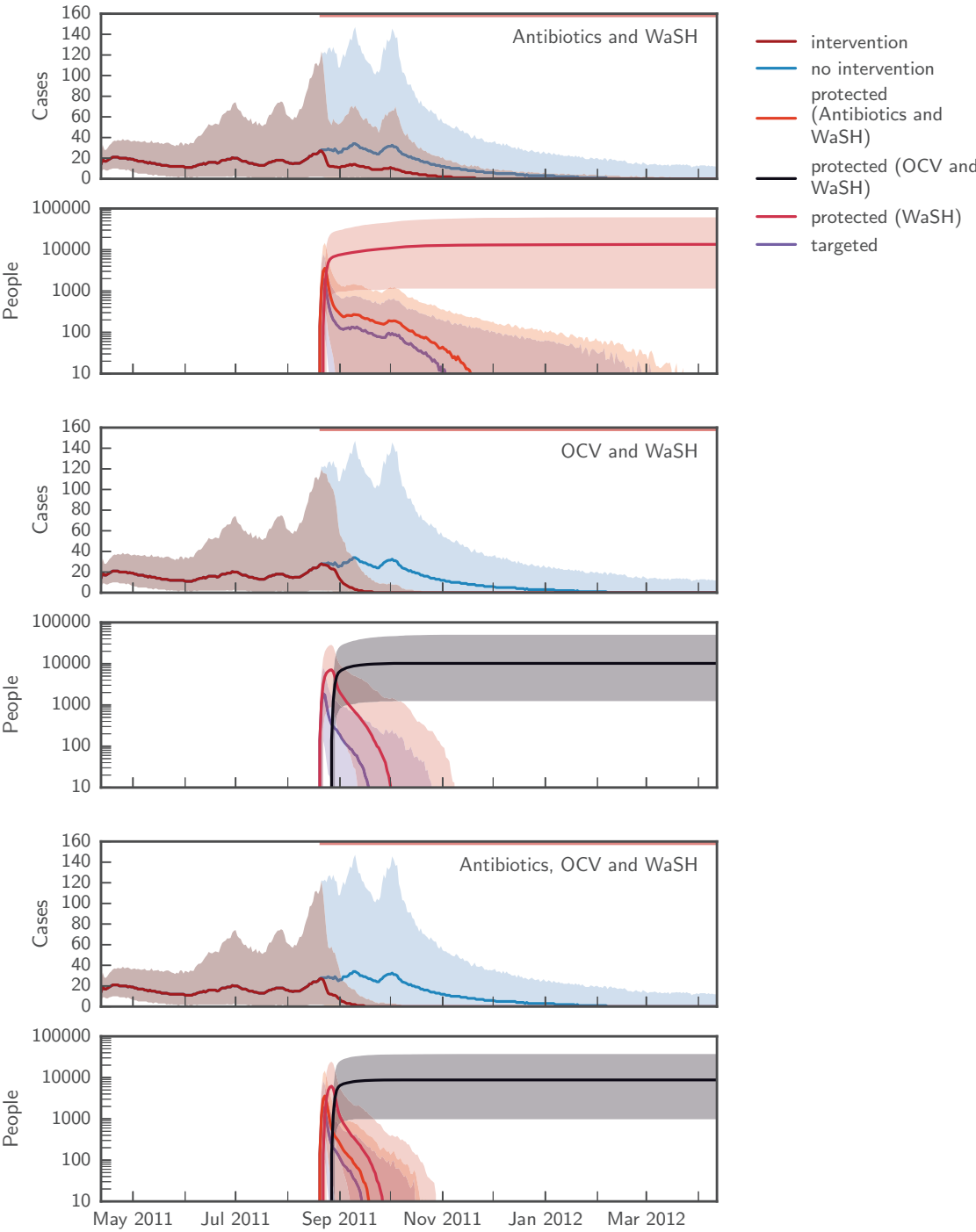


Figure 3.14 – Simulated evolution of the epidemics without intervention and with case-centered allocation of combination of the three main intervention types within a 100 m radius starting around the epidemic peak. The lower panels show the number of people targeted during each time-step, and the number of people protected by to the interventions. Solid lines show the median over all simulations, shaded areas the 2.5th and 97.5th percentiles. The red bar at the top of each panel marks the period during which interventions are applied.

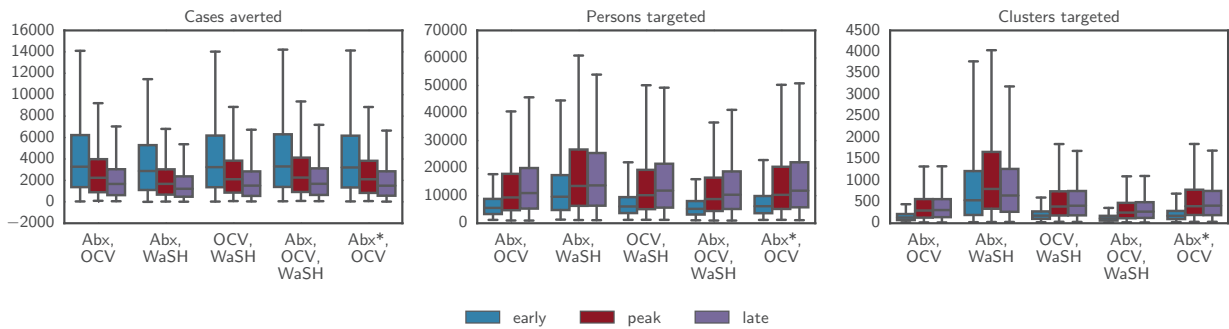


Figure 3.15 – Boxplots of the number of averted cases, the number of targeted persons and the number of targeted clusters predicted by the model for combinations of the three main intervention types with case-centered allocation in a 100 m radius starting at three different times. Abx stands for Antibiotics. Abx\* stands for administering antibiotics only within a range of 15 m, while OCV is administered within the whole cluster. The range between the whiskers comprises 95% of the values from all simulations.

but typically require hundreds to tens of thousands of doses to avert a single case (Figure 3.17). For an intervention campaign starting around the epidemic peak, case-centered interventions within a radius of 100 m are 64 times [95% CI 11 – 518] more efficient than a mass intervention campaign using OCV and 46 times [95% CI 8 – 727] using WaSH.

Random allocation of the same number of doses as with case-centered interventions throughout the entire model domain does not effectively stop epidemics nor reliably avert significant numbers of cases (Figure 3.19), with almost 50% of the simulated epidemics with interventions showing no improvement or even a higher number of cases than those without the interventions. When taking into account simulated epidemics with a positive number of averted cases, case-centered interventions starting around the epidemic peak at a 100 m buffer from a primary case lead to 4-fold [95% CI 1 – 48] higher reduction in cases using OCV and 3 times [95% CI 1 – 69] using WaSH than their non-targeted counterparts (Figure 3.17).

### 3.3.6 Additional scenarios

Despite the high efficiency of the case-centered allocation of antibiotics, especially at smaller radii or in combination with OCV, the administration of such drugs to a high number of people unavoidably raises concerns about the development of antimicrobial resistances. As an alternative, we consider scenarios where antibiotics are administered only to household members and closest neighbors of cholera patients, combined with the allocation of OCV in a larger radius, limiting the number people getting antibiotics and thus the, perhaps unjustified, risk of widespread antimicrobial resistance. We evaluate the administration of antibiotics within the same model cell (i.e. a radius of 15 m) as the reported case and the simultaneous allocation of OCV within a larger radius, varying between 30 and 100 m, thus combining the advantage of the rapid onset of protection by antibiotics at short distances with the long lasting protective effect of OCV. More cases are averted for this strategy than for for OCV

**Chapter 3. The impact of case-centered interventions in response to cholera outbreaks: a modeling study**

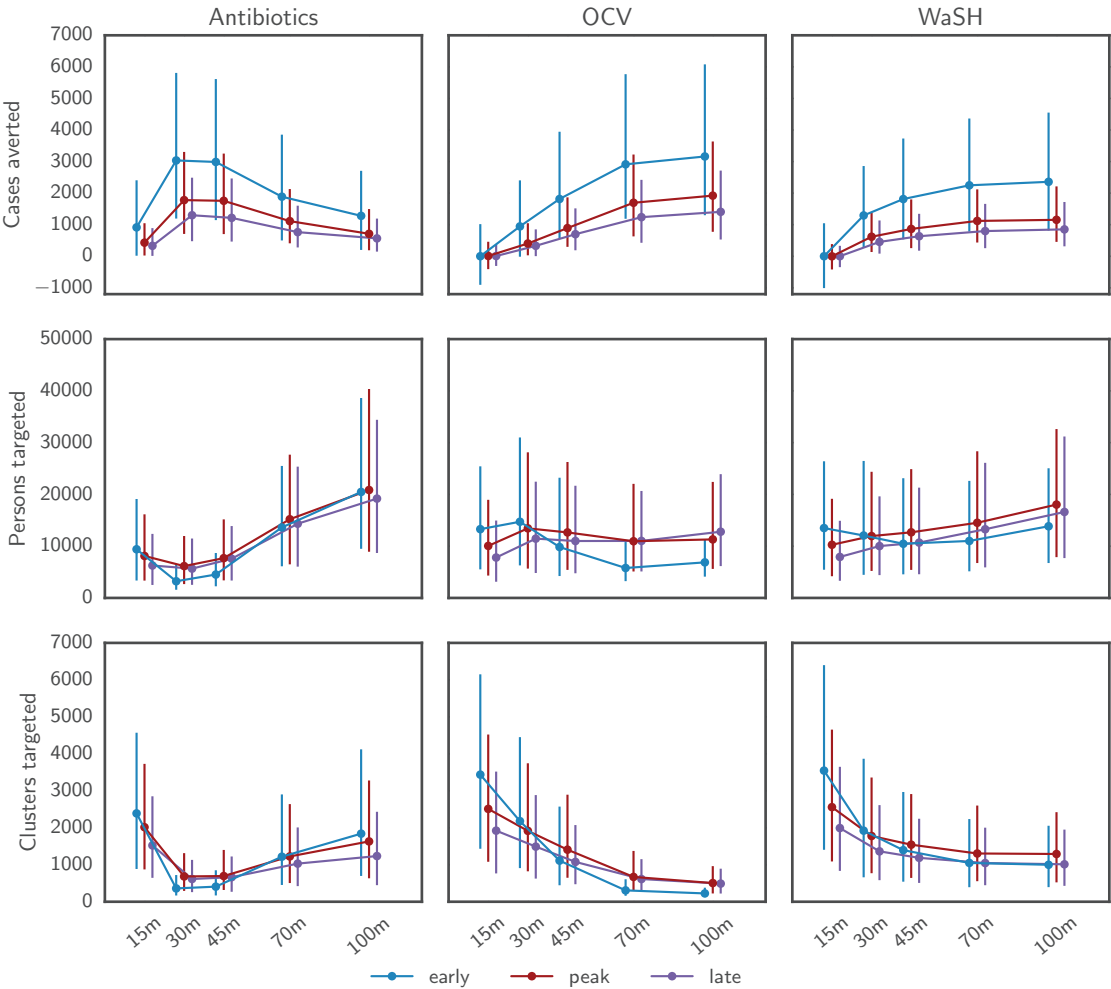


Figure 3.16 – Intervention outcomes as a function of distance in case-centered allocations. The numbers of averted cases, targeted persons and targeted clusters predicted by the model for the three main intervention types with case-centered allocation and variable radius, starting at three different times. The error bars cover the range between the 25<sup>th</sup> and the 75<sup>th</sup> quantile over all simulations.

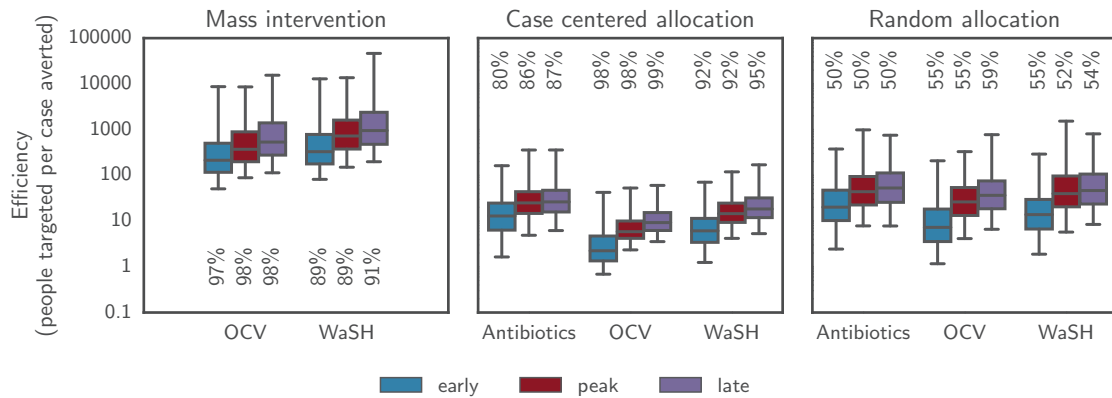


Figure 3.17 – Efficiency (persons to treat per case averted) of different modes of intervention allocation. Boxplots show the efficiency, computed as the number of persons targeted per case averted, of three modes of intervention allocation for Antibiotics, OCV and WaSH. Mass allocation of antibiotics was not considered. Case-centered allocation refers to a radius of 100 m. Random allocation means randomly targeting the same number of people as in case-centered allocation. The range between the whiskers comprises 95% of the values from all simulations. Only model runs with a positive number of cases averted were considered. Numbers within the axis show the percentage of such runs among all simulations.

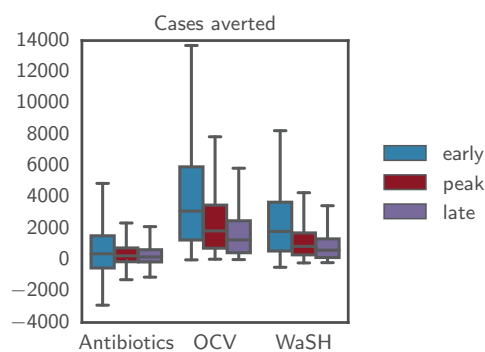


Figure 3.18 – Outcome of the three main intervention types in a mass intervention campaign randomly targeting 70% of the population. Boxplots of the number of averted cases predicted by the model. The range between the whiskers comprises 95% of the values from all simulations.

### Chapter 3. The impact of case-centered interventions in response to cholera outbreaks: a modeling study

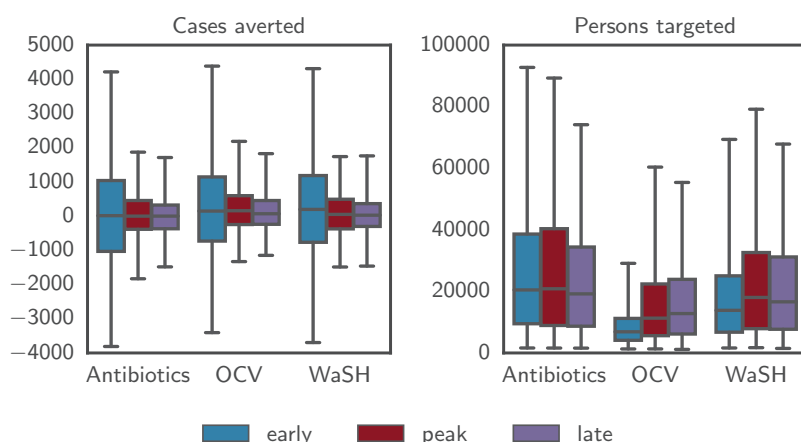


Figure 3.19 – Outcome of the three main intervention types with random allocation. Boxplots of the number of averted cases and the number of targeted persons predicted by the model. The number of targeted persons has been fixed to the same values for case-centered allocation with a 100 m radius (Figure 3.12). The range between the whiskers comprises 95% of the values from all simulations.

alone, particularly at low and intermediate radii (Figure 3.20 and Table A.1). Adding the administration of antibiotics within the first 15 m around cases to an OCV campaign at a radius of 45m starting early, the number of cases averted increases 1.13-fold [95% CI -3.89 – 3.95]. The effect is less pronounced at larger radii because OCV alone already leads to high numbers of averted cases in a radius of 100 m when starting early 1.02-fold [95% CI 0.92 – 1.37] increase of averted cases). With respect to the combined administration of OCV and antibiotics at a radius of 100 m starting early, reducing the radius for antibiotics to 15 m only leads to a 1.01-fold [95% CI 0.87 – 1.20] decrease in the number of averted cases.

Restricting the number of doses of antibiotics a person can receive to one during the whole study period, combined with the short duration of protection offered, limits the efficiency of case-centered interventions with antibiotics. In alternative scenario, we assume that a person can get several doses of antibiotics, with a minimal interval of 2 weeks, if he/she lives within the intervention radius of several cases reported at different times. By allowing for repeated targeting of a single person, we see an improved efficacy of case-centered allocation of antibiotics (Table A.1), even at high radii, such as 100 m, starting around the epidemic peak, the averted cases increasing by a factor of 1.4 [95% CI -8.7 – 12.4], at the cost of a 1.4-fold [95% CI 0.70 – 2.5] higher number of people to target (i.e. doses to administrate, Figure 3.21).

Regardless of the fact that the level of protection after a single-dose of OCV has been assessed starting after a lag of 7-days only [Qadri *et al.*, 2016], some evidence suggests that the onset of (at least partial) protection may occur much earlier [Azman *et al.*, 2016]. If OCV recipients were indeed protected two days after vaccination, we find a much higher efficiency of case-centered

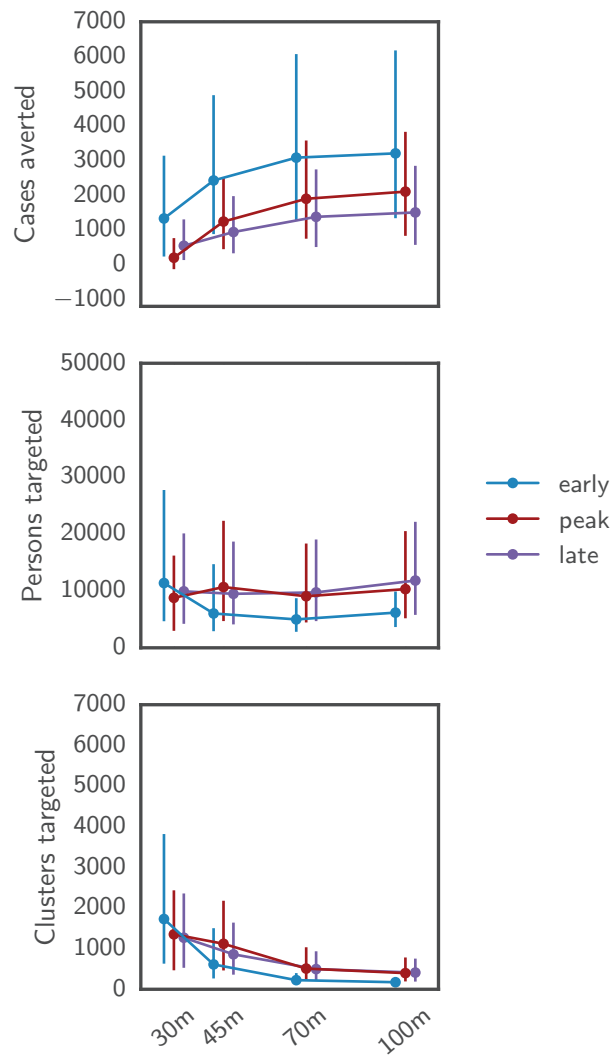


Figure 3.20 – Intervention outcomes of case-centered allocation of OCV within a varying radius combined with the allocation of antibiotics within a radius of 15 m. The number of averted cases, the number of targeted persons and the number of targeted clusters predicted by the model for the three main intervention types with case-centered allocation and variable radius, starting at three different times. The error bars mark the interquartile range over all simulations.

### Chapter 3. The impact of case-centered interventions in response to cholera outbreaks: a modeling study

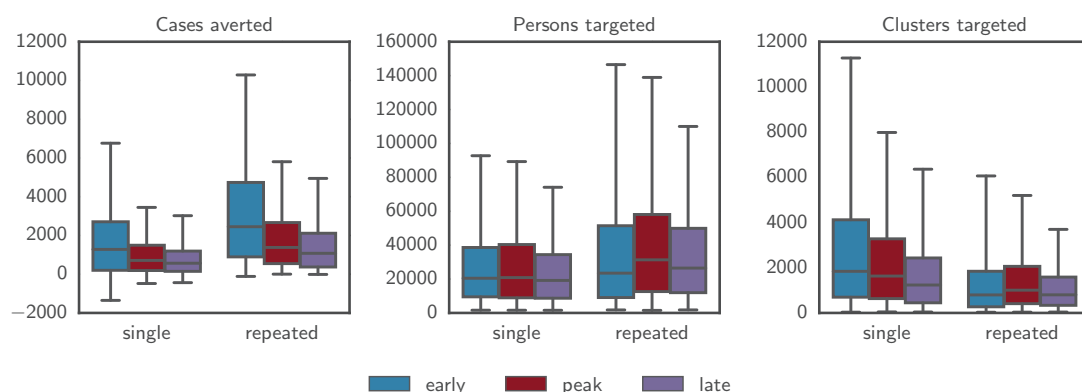


Figure 3.21 – Comparison between single and repeated antibiotics administration through case-centered allocation in a 100 m radius. Boxplots of the number of averted cases, the number of targeted persons and the number of targeted clusters predicted by the model for two different strategies of allocating antibiotics in a case-centered allocation in a 100 m radius starting at three different times. The range between the whiskers comprises 95% of the values from all simulations. Single allocation designates the standard mode where every person can receive a single-dose of antibiotics only once during the epidemic. Repeated allocation designates a mode where one and the same person can get antibiotics several times, with a minimal interval of 2 weeks, if he/she lives within the intervention radius of several cases.

OCV allocations, even within small radii (Figure 3.22 and Table A.1). Within a radius of 30 m and when starting interventions around the epidemic peak, this leads to a 1.8-fold [95% CI -12.0 – 24.7] increase of the number of averted cases. Within a radius of 45 m the increase in the number of averted cases is 1.7-fold [95% CI -9.3 – 8.4], within a radius of 70 m 1.2-fold [95% CI 0.94 – 2.6] and within a radius of 100 m 1.1-fold [95% CI 1.0 – 1.9].

## Discussion

Using a micro-simulation model calibrated to real-world data, we illustrated that case-centered interventions with antibiotics, OCV and/or WaSH can efficiently and effectively mitigate the impact of cholera epidemics in urban areas. Among the three interventions types, OCV most effectively stops epidemics (e.g. 97.5% of all simulated epidemics were stopped within 39 [95% CI 13 – 85] days when intervention start around the epidemic peak), whereas antibiotics have the most important short-term impact. Combinations of the three types of interventions can be used to reduce even more cases and deaths, although adding a third intervention to any two delivers only marginal gains. Further, we found that case-centered interventions, which require tens to hundreds of people treated per case averted, are by far more resource efficient than mass intervention campaigns, which typically require several hundred to hundreds of thousands people to be treated per case averted. The optimal buffer around a case depends on the type of intervention. For antibiotics, which offer a limited duration of preventive protection, the optimal distance is around 30 m – 45 m, whereas for OCV and WaSH, which offer



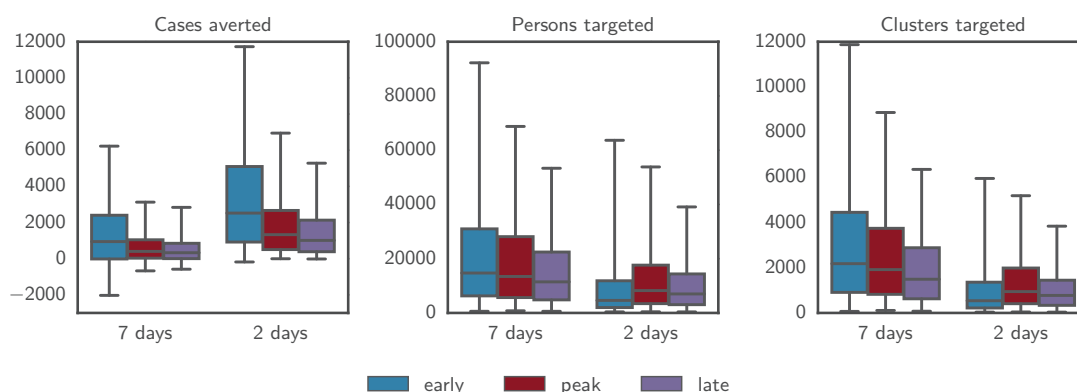


Figure 3.22 – Comparison between a 7- and 2-days lag between the administration of a single-dose of OCV and the onset of protection in a case-centered allocation framework in a 30 m radius. The range between the whiskers comprises 95% of the values from all simulations.

protection lasting several months at least, the intervention impact increases with distance and saturates around 70 m – 100 m.

The practice of visiting households of (or around) cases is not new to public health [Hitchings *et al.*, 2016; Kucharski *et al.*, 2016; Merler *et al.*, 2016] nor cholera [Guévert *et al.*, 2007]. In a number of countries, it is standard practice to visit case households to provide health/hygiene education for household members of cases, soap and sometimes water and/or latrine disinfectants and antibiotics. Timely case-centered interventions, such as those proposed in this study, depend on a well performing epidemiological surveillance, are logistically complex to implement and require well-trained staff. First, the rapid detection and confirmation of cases is key, so that true cholera cases are correctly prioritized. The use of cholera rapid tests may provide one such avenue given that traditional diagnostics (culture) requires days to complete for some patients [Debes *et al.*, 2016b]. Other challenges to implementation, which two studies have shown feasible [Guévert *et al.*, 2007; Parker *et al.*, in prep.], include finding the case's household and negotiating with local leaders to efficiently deliver the interventions.

While effective and efficient, it remains unclear when case-centered approaches should be used, particularly in contrast to mass campaigns, the current standard for outbreak response. As shown, early initiated case-centered interventions can have profound impacts on the epidemic trajectory. If resources, like OCV, are limited as they are today, case-centered interventions may be the most appropriate strategy to target those at highest risk. When only a few cases are detected, Ministries of Health may want to initiate case-centered interventions to efficiently quell the epidemic with supplies already in the country, while making contingency plans for mass interventions if the epidemic continues to grow. Finally, these strategies may be used late in epidemics or epidemic seasons, possibly after a mass campaign, to quickly stop the often protracted tail of the epidemic [Parker *et al.*, in prep.; Rebaudet *et al.*, 2013c]. Case-centered approaches with WaSH interventions, using larger radii than in this paper, are

### Chapter 3. The impact of case-centered interventions in response to cholera outbreaks: a modeling study

---

currently being deployed to fight the cholera epidemic in Haiti, although their effectiveness in reducing cases and deaths remains unknown [Santa-Olalla *et al.*, 2013].

While our study used a rigorous approach to calibrate the models and capture uncertainty in both the epidemic process and the intervention effects, it comes with a number of limitations, which may narrow the inference and generalizability of the results.

The highly stochastic nature of the processes involved, resulting from the number of possible epidemic trees within a population close to 1 million, together with the fact that the observed epidemic represents only one realization thereof, make finding an exact model fit to the epidemic curve difficult. In addition, calibrating on two distinct criteria simultaneously (the epidemic curve and the  $\tau$  values) inherently introduces a trade-off between the two. A combination of those two effects ultimately leads to the high variability observed in the calibrated epidemic curve (Figure 3.6A). Moreover, data on reported cases used in this study represent suspected cholera cases, only some of which have been confirmed [Azman *et al.*, in prep.].

The quantification of the effects and mechanisms of each intervention were based on limited data and ultimately depend on large clinical trials across multiple epidemiological and environmental settings. Effect estimates assumed for antibiotics are taken from meta-analyses and large studies representative of the current state of evidence. However, the variability in the effect estimate is high, both resulting from the limited data for prophylactic antibiotic use and from the diversity of different drugs used [Leibovici-Weissman *et al.*, 2014]. For OCV, there exists only one clinical trial estimating the single-dose efficacy in Bangladesh over a six-month period, where cholera is hyper-endemic [Qadri *et al.*, 2016], and one observational study in South Sudan measuring short-term protection in the first 2-months after vaccination [Azman *et al.*, 2016]. In the case of WaSH interventions, such trials are particularly laborious and difficult to implement because other than just measuring benefits, the behavior of study populations has to be followed over long periods of time. Study participants might, in addition, alter their behavior when being observed, leading to a biased results. The diversity of possible WaSH measures and protocols used adds additional variability among different studies. All those factors lead to high heterogeneity and uncertainties in the estimated beneficial effects of WaSH interventions, some of which may not have been fully captured in our analyses [Fewtrell and Colford, 2004; Fewtrell *et al.*, 2005]. Finally we assumed that WaSH interventions maintained their effectiveness throughout the study period, which may not be true, given that behavior change has been shown to wane quite quickly in different situations.

Although our results are based on a single case study, the underlying property they rely on is the spatiotemporal clustering of cholera cases in urban areas, which has been shown to exist in similar extents in epidemic and endemic settings around the globe [Ali *et al.*, 2016; Azman *et al.*, in prep.; Blackburn *et al.*, 2014; Carrel *et al.*, 2009; Debes *et al.*, 2016a, b; Luquero *et al.*, 2011; Snow, 1855; You *et al.*, 2013]. We thus advocate that case-centered intervention strategies can be a promising approach to control cholera epidemics in urban settings. Optimal

intervention ranges in other cities may, however, depend on population density, mobility patterns of inhabitants and other factors that influence the transmission routes and spread of cholera within a population. However, the optimal radii for interventions with antibiotics will presumably remain lower in comparison to OCV and WaSH, as it is a result of the delay to onset of protection and duration of protection of each, which is unlikely to vary considerably across settings.

Although for distance ranges above 15 m the modeled relative risk  $\tau$  matches the observed data well, the values at very short ranges, i.e. between 0 m and 15 m, are subject to underestimation (Figure 3.6). This is potentially due to the limited resolution of the model, and the fact that the population distribution used does not include the structure of households and close neighborhoods, which have been shown to have significant impacts on transmission for other diseases [Salje *et al.*, 2016a]. Those limitations may, however, only lead to underestimation of the effect of clustered interventions, as one would expect the results to show even stronger support for case-centered interventions with additional short-range clustering.

The purpose of this study being the evaluation of the impact of interventions on the course of a single outbreak, long-term effects of interventions potentially influencing future epidemics have not been evaluated. Given that the effectiveness of OCV and WaSH interventions likely wane differently over time, it is possible that there are different optimal mixes of interventions depending on the time-scale of interest. One might, however, expect the number of susceptible individuals to remain smaller for a duration exceeding the study period in areas where OCV has been used, even if more clinical trials are necessary to evaluate the duration of the protective effect of single-dose OCV application.

The way global spatial clustering is measured in this study solely depends on the number of cases within certain distance and time ranges from other cases. We do not dispose of any information from which transmission chains could be derived. The spatial clustering which we thus observe and employ in our model results from a combination of the transmission kernel of the disease with exogenous risk factors and heterogeneous distribution of disease susceptibility within the population [Lessler *et al.*, 2016; Salje *et al.*, 2012]. Interventions such as the ones proposed in our study, however, rely on the combined spatiotemporal clustering of cases, which is thus more relevant than the ability to reconstruct individual transmission chains [Lessler *et al.*, 2016].

Mobility has been shown to have a major influence on the transmission patterns of cholera epidemics [e.g. Finger *et al.*, 2016] at the scales of regions and countries, and is also expected to impact the way cholera spreads within cities [Azman and Lessler, 2015]. In our model, alongside with other mechanisms leading to the spread of cholera, mobility is incorporated into an isotropic kernel which does not depend on the position of its origin within the study domain. Incorporating more information about actual distances and frequencies of travel of people living in different parts of the city could lead to models fitting the actual epidemiological data better and the disease spread being more representative of reality.

### **Chapter 3. The impact of case-centered interventions in response to cholera outbreaks: a modeling study**

---

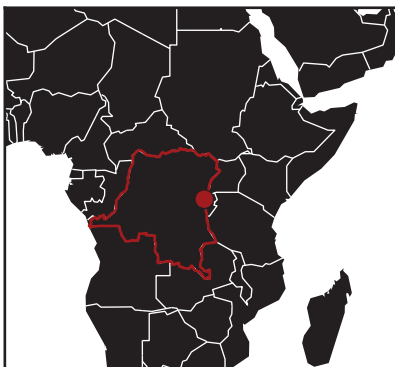
The results of this study suggest that case-centered intervention strategies may be an effective and resource-efficient approach reducing morbidity and mortality in cholera epidemics. While one intervention may perform better than another in different settings, our results suggest that combining any two of the interventions can quickly bring epidemics to an end. More work is needed to understand how and when to best use this approach in outbreaks across different settings, taking into account both human-resource capacity and supply availability.

## 4 Cholera in the Lake Kivu region (DRC): Integrating remote sensing and spatially explicit epidemiological modeling

*This chapter has been published in Water Resources Research [Finger et al., 2014]. The study has been designed by Flavio Finger, Allyn Knox, Enrico Bertuzzo, Lorenzo Mari, Didier Bompangue, Marino Gatto and Andrea Rinaldo, and led by Flavio Finger. The draft of the article has been written by Flavio Finger, all authors have contributed to the final version.*

Finger, F, A. Knox, E. Bertuzzo, L. Mari, D. Bompangue, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo, Cholera in the Lake Kivu region (DRC): Integrating remote sensing and spatially explicit epidemiological modeling, *Water Resources Research*, pp. 5624–5637, doi: 10.1002/2014WR015521, 2014

### Overview



**Country:** Democratic Republic of the Congo

**Study domain:** Lake Kivu area

**Surface:** 3981 km<sup>2</sup>

**Population:** 1 800 000

**Cholera:** Outbreaks occurring every year.

**Period studied:** 2004 – 2011

**Number of reported cases:** 45 000

## **Abstract**

Mathematical models of cholera dynamics can not only help identifying environmental drivers and processes that influence disease transmission, but may also represent valuable tools for the prediction of the epidemiological patterns in time and space as well as for the allocation of health care resources. Cholera outbreaks have been reported in the Democratic Republic of the Congo since the 1970s. They have been ravaging the shore of Lake Kivu in the east of the country repeatedly during the last decades. Here, we employ a spatially explicit, inhomogeneous Markov chain model to describe cholera incidence in eight health zones on the shore of the lake. Remotely sensed datasets of chlorophyll *a* concentration in the lake, precipitation and indices of global climate anomalies are used as environmental drivers in addition to baseline seasonality. In addition, the effect of human mobility is modelled mechanistically. We test several models on a multi-year dataset of reported cholera cases. The best fourteen models, accounting for different environmental drivers and selected using the Akaike information criterion, are formally compared via proper cross-validation. Among these, the one accounting for seasonality, El Niño Southern Oscillation, precipitation and human mobility outperforms the others in cross-validation. Some drivers (such as human mobility and rainfall) are retained only by a few models, possibly indicating that the mechanisms through which they influence cholera dynamics in the Lake Kivu area will have to be investigated further.

## **4.1 Introduction**

The risk, loss and social disruption brought by cholera outbreaks can hardly be overestimated and the global relevance of preventive assessments and controls of cholera spreading is manifest. The recent epidemics in Haiti, the Congo river basin, Cuba, Sierra Leone and the Sahel region [Al-Tawfiq and Memish, 2012; Bompangue et al., 2011; Fernández et al., 2009; Gaudart et al., 2013; Kelvin, 2011] witness the ongoing, widespread inadequacy of reliable drinking water supply and sanitation infrastructure all over the developing world. As a result, cholera remains a major cause of morbidity and mortality in developing countries even to date, despite all public health policies and humanitarian efforts deployed worldwide. As an example, according to the World Health Organization, as much as 85% increase in the number of reported cholera cases has been observed globally in 2011 relative to 2010, with 58 countries involved and a total of 589854 yearly cases leading to an overall case fatality rate of 1.3% [World Health Organization, 2012].

Here we use a semi-mechanistic, spatially explicit modeling framework to describe cholera dynamics around Lake Kivu, Democratic Republic of the Congo (DRC). Our approach builds on the multidimensional inhomogeneous Markov chain (MDIMC) method proposed by Reiner et al. [2012]. This method requires the discretization of the variable to be modeled (i.e. cholera incidence) into a finite number of states, and applies a semi-mechanistic description of the transitions between discrete dynamical states. Transition probabilities vary in time as they account for environmental drivers (estimated through remotely sensed datasets) and human

mobility patterns. Routinely collected surveillance data have been used to construct epidemic curves of cholera cases and map the spatio-temporal evolution of the disease [Bompangue *et al.*, 2009]. Datasets of precipitation, chlorophyll *a* concentration in Lake Kivu as well as indices of global climate phenomena are used as model input, together with a mechanistic description of human mobility among the health zones adjacent to the lake. Specifically, the MDIMC model is fed with all possible combinations of environmental drivers, with variable lags. The performances of different model settings are compared using formal model selection techniques in order to draw conclusions about the relative importance of environmental drivers for the proliferation of cholera in the study area. Furthermore, cross-validation is applied to assess the possibility of predictive modelling of epidemiological dynamics based on environmental data.

## 4.2 Case study

### 4.2.1 Spatial setting

Lake Kivu is situated in eastern DRC on the border with Rwanda (Figure 4.1). In this study we concentrate on eight health zones (or their aggregations) located on the Congolese shore, which include the two major cities of Goma and Bukavu, at the northern and southern ends of the lake respectively. Areas further from the lake are not considered due to their low number of cholera cases and limited population. The total population size of the study area is of about 1.8 millions.

### 4.2.2 Pathogen transport

The Lake Kivu catchment consists of numerous small subcatchments along steep slopes leading down to the lake and a northern region characterized by porous volcanic soils that allow for little (to no) surface runoff. Therefore we assume that the hydrological transport of the pathogen (*sensu Bertuzzo et al.* [2010] and *Rinaldo et al.* [2012]) is negligible at the regional scale. However, rainfall can facilitate local pathways of transmission and/or amplify contamination through failure of inappropriate sanitation systems. The health zones in our model are connected through human mobility fluxes, simulated here by a gravity model. Given the stark difference in sociopolitical stability between the eastern DRC and neighboring Rwanda, and the low number of cases reported in Rwanda during the study period, we assume fluxes between the two countries to be negligible.

### 4.2.3 Climate

Local climate in the study area is characterized by a rainy season from October to May [Bompangue *et al.*, 2009; Plisnier *et al.*, 2000], which is interrupted by a short dry period early in the year. The annual precipitation corresponds to around 1200mm. Monthly av-

**Chapter 4. Cholera in the Lake Kivu region (DRC): Integrating remote sensing and spatially explicit epidemiological modeling**

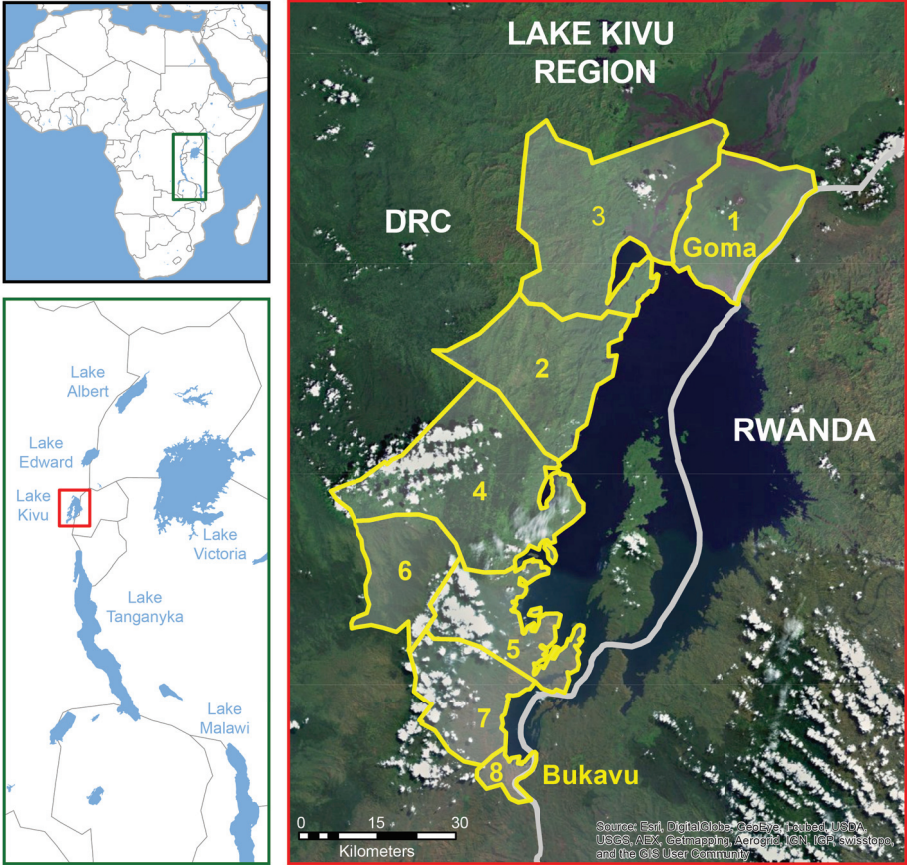


Figure 4.1 – Study area. Lake Kivu is one of the African Great Lakes (a), located to the north of Lake Tanganyika and west of Lake Victoria (b), at the border between the DRC and Rwanda (c). The eight lakeside health zones included in this study (1: Goma, 2: Minova, 3: Kirotshe, 4: Kalehe, 5: Katana, 6: Miti Murhesa, 7: Kabare, 8: Bukavu) as well as Goma and Bukavu, the two main cities on its shore, are shown in panel (c).



erage temperatures are fairly constant, close to 20°C throughout the year (available from: <http://en.climate-data.org/location/1074/>). *Plisnier et al.* [2000] reported highly complex and spatially differentiated effects of ENSO on local climate, such as a positive correlation of ENSO with rainfall, air pressure and temperature in the area. According to *Stager et al.* [2007] ENSO also influences lake levels in the African Great Lakes Region. *Marchant et al.* [2007] found that the the Indian Ocean Dipole (IOD), a cyclic climate phenomenon independent of ENSO, influences the rainfall in East Africa too, in accordance with *Becker et al.* [2010], who revealed an influence of ENSO and IOD on the total water storage in the area. The exact mechanisms leading to all the above so-called teleconnections are yet to be determined [*Marchant et al.*, 2007; *Plisnier et al.*, 2000].

### 4.2.4 Data

#### Cholera

Weekly cholera incidence data (2004–2011) were made available through the work of *Bompangue et al.* [2009]. Briefly, data were collected from registries at each Cholera Treatment Center, aggregated weekly and by health zone, and reported to the Ministry of Health, where they were preserved in electronic or paper format [*Bompangue*, 2009; *Bompangue et al.*, 2008, 2009, 2012, 2011; *Piarroux and Bompangue*, 2007; *Piarroux et al.*, 2009]. For the purpose of this study we aggregated the data to obtain monthly numbers of cases for the eight lakeside health zones described above (Figure 4.2).

The time scale of this study has been chosen to be monthly because of the level of noise and the number of missing values in both reported cholera cases and remotely sensed Chlorophyll *a* concentrations. Note for instance that in order to get a high-quality time series of Chlorophyll *a* data, every time-step must contain a certain number of cloudless days, which is sometimes difficult to enforce during the rainy season in the study region.

#### Demography

A remotely sensed dataset of the estimated 2010 population distribution (available from <http://www.worldpop.org.uk>) was used to approximate the population of each health zone, serving as a base to compute monthly cholera incidence (reported cases divided by population abundance of each health zone).

#### Plankton

The optimization of remotely-sensed plankton biomass estimates for Lake Kivu, described in *Knox et al.* [2014], enabled the selection of the plankton biomass proxy best-suited for this study. Here, we use a chlorophyll *a* database generated with the OC3 bio-optical algorithm and a coastal atmospheric correction model with 90% relative humidity, spanning the years

**Chapter 4. Cholera in the Lake Kivu region (DRC): Integrating remote sensing and spatially explicit epidemiological modeling**

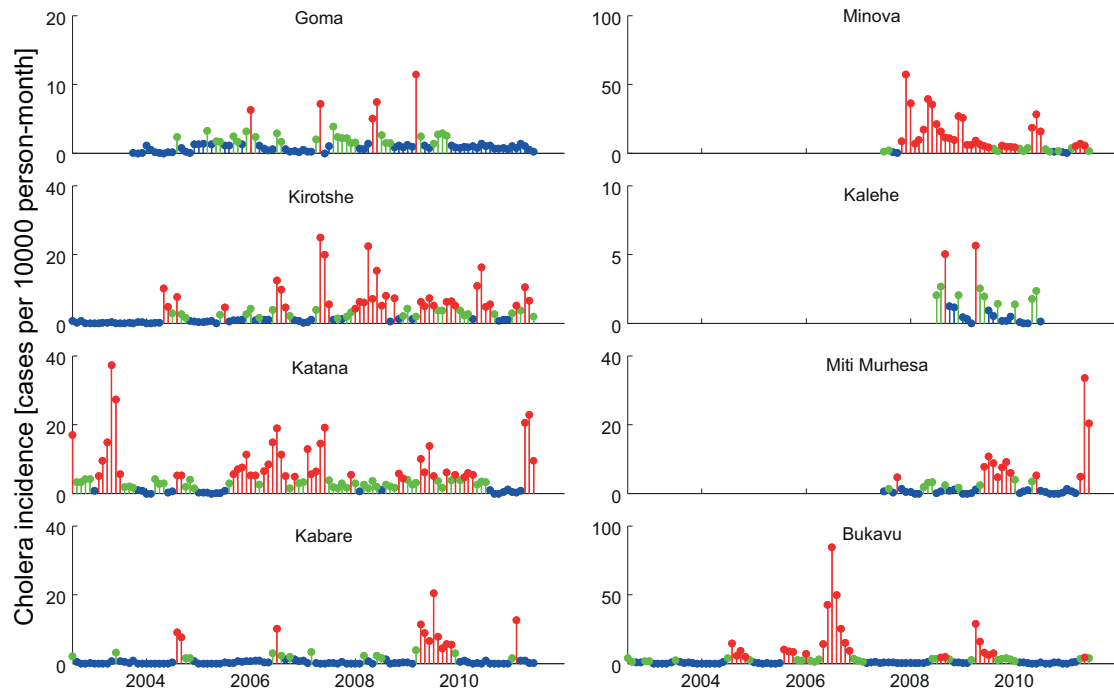


Figure 4.2 – Cholera incidence data for the eight health zones considered. Low (state 1), mild (state 2) and high incidence (state 3) are denoted by respectively blue, green and red color. Note that in some health zones data are available only during parts of the study period.

2002–2012 (Figure 4.3). Daily data were spatially averaged across the entire lake, and monthly averages were created by weighting each day by the number of data yielding pixels.

**Precipitation**

Daily precipitation fields were obtained from a remotely sensed dataset by the National Aeronautics and Space Agency (NASA) [Huffman *et al.*, 2010]. The resolution of the dataset is 0.25 degrees of latitude and longitude. Precipitation estimates were then projected to each health zone and aggregated monthly (Figure 4.3). The projection was done by assigning the corresponding precipitation value to each cell in a rasterized version of the health zones delimitation and subsequently taking the mean over each health zone.

**ENSO and IOD**

In order to account for possible relations between global climate anomalies and the dynamics of the disease in the study region, as reported by Bompangue *et al.* [2011], we included two additional climatic drivers in our study (Figure 4.3). SST anomaly from the Niño 3.4 region made available by the National Oceanic and Atmospheric Administration (NOAA, available online at <http://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices>), was used as index for ENSO. For IOD we used the so-called Dipole Mode Index (DMI) [Saji *et al.*, 1999],

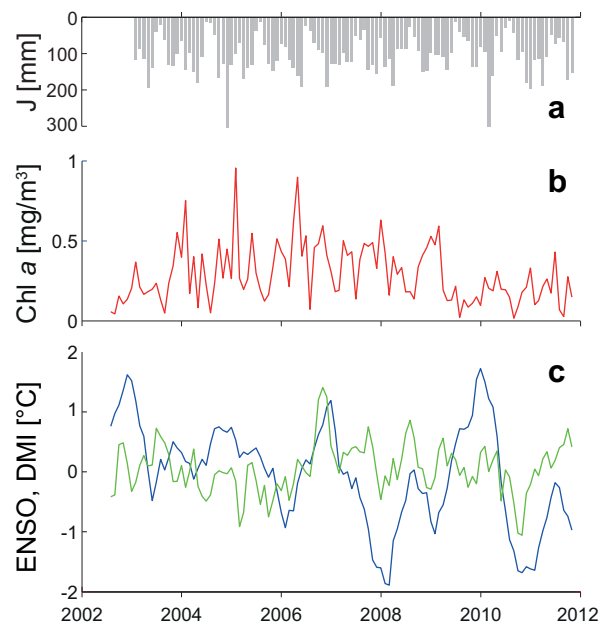


Figure 4.3 – Environmental drivers. (a) Total monthly precipitation averaged over all considered health zones. (b) Remotely sensed, spatially averaged chlorophyll *a* concentration in Lake Kivu. (c) Global climate anomalies. ENSO SST anomaly in the Niño 3.4 region (blue) and IOD Dipole Mode Index (green). Note that the chlorophyll *a* and climate anomalies start 6 months before cholera incidence data in order to allow for lags in the model.

a measure of the SST gradient between two regions in the Indian Ocean (available online at <http://www.jamstec.go.jp/frcgc/research/d1/iod/DATA/dmi.monthly.ascii>).

In order to quantify correlations between the different environmental drivers (precipitation, chlorophyll *a*, ENSO and IOD), we computed their cross-correlation functions. Precipitation and chlorophyll *a* show low significant correlations ( $r \approx 0.2$ ) between lags 0 and 2 months, whereas DMI and ENSO are weakly correlated at lags around 1 month ( $r \approx 0.2$ ) as well as anti-correlated at higher lags. All other combinations do not show significant correlations ( $p > 0.05$ ).

## 4.3 Model

### 4.3.1 Inhomogeneous Markov chain model for endemic cholera

The theoretical framework adopted here builds on a previous deterministic modeling approach (MDIMC) for endemic cholera developed by *Reiner et al.* [2012] that is based on finite-state Markov chain modeling. This approach requires cholera incidence data to be categorized into discrete states. It assigns a probability to the transitions between epidemic states in a given spatial setting and with a defined time-step. The Markov chain model can be made inhomogeneous by allowing transition probabilities to depend on temporal and spatial

## Chapter 4. Cholera in the Lake Kivu region (DRC): Integrating remote sensing and spatially explicit epidemiological modeling

---

environmental drivers, namely precipitation, chlorophyll  $a$  concentration in the lake, ENSO and IOD, as well as on spatial interactions induced by human mobility.

Following *Reiner et al.* [2012] we categorize cholera incidence into three discrete states, namely low (state 1), mild (state 2) and high (state 3). We select the monthly incidence thresholds for the definition of the three discrete states such that half of the records fall in the low incidence category and the remaining data are equally partitioned into the two other categories. Therefore the thresholds correspond to the 50<sup>th</sup> (0.0133% incidence) and the 75<sup>th</sup> (0.0415%) percentiles of the monthly incidence data, respectively.

Figure 4.2 shows cholera incidence in the eight health zones considered, and the category each data point belongs to. Low (state 1), mild (state 2) and high (state 3) incidence are denoted by blue, green and red color respectively. The categorization of the data into finite classes causes the merging of the events in the tail of the incidence distribution with less severe ones. Although some information is therefore discarded, the model focuses on levels of variation that are relevant to public health.

We first define a baseline homogeneous Markov chain model that does not account for environmental drivers and spatial interactions. Let  $X_{k,t} = 1, 2, 3$  be the state of health zone  $k$  at time  $t$ . According to this model, the generic transition  $(X_{k,t} = i) \rightarrow (X_{k,t+1} = j)$  occurs, at any time  $t$ , with probability  $p_{i,j}$ . The baseline model is therefore completely defined by the following transition probability matrix  $\mathcal{P}$ :

$$\mathcal{P} = \begin{bmatrix} p_{1,1} & (1 - p_{1,1} - p_{1,3}) & p_{1,3} \\ p_{2,1} & (1 - p_{2,1} - p_{2,3}) & p_{2,3} \\ p_{3,1} & (1 - p_{3,1} - p_{3,3}) & p_{3,3} \end{bmatrix} \quad (4.1)$$

where the probabilities of transition to state 2 are expressed so as to enforce  $\mathcal{P}$  to be a stochastic matrix (i.e. row-wise sums equal to 1).

We further assume that environmental drivers and spatial interactions among neighboring human communities can modify baseline transition probabilities (4.1). In particular, we hypothesize that these external drivers can modify, through a multiplicative factor, the probabilities of transition from low/mild cholera incidence to a worse state ( $p_{i,j}$ ,  $i < j$ ), as well as the probability of remaining in the highest incidence state ( $p_{3,3}$ ). The resulting transition probabilities  $p'_{i,j,k,t}$  are thus site- and time-specific, as both environmental drivers and cholera incidence patterns vary in space and time. We adopt the following formulation:

$$p'_{i,j,k,t} = p_{i,j} (1 + f_{k,t}^{rain}) (1 + f_t^{chl}) (1 + f_{k,t}^{ENSO}) \\ (1 + f_{k,t}^{IOD}) (1 + f_{k,t}^{mob}) (1 + f_t^{sea}) \quad (4.2)$$

for  $i < j$  or  $i = j = 3$ .

The remaining probabilities are adjusted, proportionally to their baseline values, to ensure

that matrix  $\mathcal{P}'_{k,t} = (p'_{i,j,k,t})$  is stochastic, i.e.:

$$\begin{aligned} p'_{1,1,k,t} &= 1 - p'_{1,2,k,t} - p'_{1,3,k,t} \\ p'_{i,1,k,t} &= (1 - p'_{i,3,k,t}) \frac{p_{i,1}}{p_{i,1} + p_{i,2}} \text{ for } i = 2, 3 \\ p'_{i,2,k,t} &= (1 - p'_{i,3,k,t}) \frac{p_{i,2}}{p_{i,1} + p_{i,2}} \text{ for } i = 2, 3. \end{aligned} \quad (4.3)$$

The term  $f_{k,t}^{rain}$  is assumed to be linearly dependent on the actual precipitation intensity  $f_{k,t}^{rain} = \alpha J_k(t)$ , where  $J_k(t)$  is the mean normalized monthly precipitation of health zone  $k$  during month  $t$ . Precipitation data have been normalized to span the range  $[0, 1]$ . Therefore, to enforce  $p'_{i,j,k,t} > 0$  we impose the constraint  $\alpha \geq -1$ . As an example, if  $\alpha > 0$ , rainfall enhances cholera transmission and therefore all the probabilities to make a transition to a higher incidence state (or to stay at the highest) increase. Consequently, all the other transition probabilities decrease. Analogously, the potential effect of chlorophyll  $a$  concentration on cholera transmission is modelled as  $f_t^{chl} = \beta C(t - t_C)$  ( $\beta \geq -1$ ), where  $C(t)$  is the mean normalized (i.e. rescaled in the range  $[0, 1]$ ) monthly chlorophyll  $a$  concentration of month  $t$ . The lag  $t_C$  is introduced to possibly account for a delay between the dynamics of phyto- and/or zoo-plankton and favorable conditions for bacteria survival in the lake. To account for the possible enhancing effect of climatic drivers on disease dynamics we model the terms  $f_t^{ENSO}$  and  $f_t^{IOD}$  equivalently to  $f_t^{chl}$ , i.e.  $f_t^{ENSO} = \phi ENSO(t - t_{ENSO})$  and  $f_t^{IOD} = \psi IOD(t - t_{IOD})$ , where  $ENSO(t)$  is the normalized SST anomaly in the Niño 3.4 region during month  $t$ ,  $IOD(t)$  is the normalized DMI (Section 4.2.4) during month  $t$ ,  $t_{ENSO}$  and  $t_{IOD}$  are time lags and  $\phi \geq -1$  as well as  $\psi \geq -1$  are proportionality constants.

Cholera transmission in a health zone can also be enhanced by the mobility of people toward health zones with ongoing outbreaks. This potential effect is accounted for in equation (4.2) by the term  $f_{k,t}^{mob}$ , which reads

$$f_{k,t}^{mob} = \gamma \sum_{z \neq k} Q_{kz} X_{z,t}^v,$$

where  $Q_{kz}$  is the probability that a traveller from zone  $k$  visits zone  $z$ , and  $\gamma$  and  $v$  are two positive parameters. We model human mobility through a gravity model [Erlander and Stewart, 1990]. Accordingly, connection probabilities are defined as

$$Q_{kz} = \frac{H_z e^{-d_{kz}/D}}{\sum_{n \neq k} H_n e^{-d_{kn}/D}},$$

where the attractiveness factor of zone  $z$  depends on its population size  $H_z$ , while the deterrence factor is assumed to be dependent on the distance  $d_{kz}$  between the two communities and represented by an exponential kernel (with shape factor  $D$ ). Distances between health zones are measured along the road network.

## Chapter 4. Cholera in the Lake Kivu region (DRC): Integrating remote sensing and spatially explicit epidemiological modeling

---

Finally the term  $f_{k,t}^{sea}$  in equation (4.2) accounts for the seasonality possibly induced by drivers other than those explicitly considered above. Baseline seasonality is modelled through a simple sinusoidal function:

$$f_t^{sea} = \delta \left( 1 + \sin \left( 2\pi \frac{t - t_s}{12} \right) \right),$$

where  $\delta \geq -1$  and  $t_s$  is the lag of seasonality.

### 4.3.2 Model calibration and validation

We consider all the  $2^6 = 64$  model combinations obtained by accounting for or neglecting the effects of rainfall, chlorophyll  $a$ , human mobility, ENSO, IOD and baseline seasonality. Models are fitted by maximizing their likelihood. In the most complex setting we must optimize the values of 17 parameters, including the set of lags (if applicable) up to 6 months that produces the best fit to data. Under the Markovian assumption of the model, the transition from one month to the next is independent of all other transitions. Therefore, likelihood can be defined as the product of the probabilities of the transitions actually observed for each month. We use the simplex search algorithm proposed by *Nelder and Mead* [1965] to maximize the log-likelihood. We enforce the constraint that each transition probability must be between 0 and 1 by a barrier method, i.e. we set likelihood to 0 whenever a transition probability falls outside these limits [*Reiner et al.*, 2012]. Because the Nelder-Mead method can only ensure the identification of local stationary points of the considered objective function, the optimization algorithm is run 100 times with different initial starting points to better approximate the global maximum of the likelihood function. The best model is then selected out of all candidate model combinations through the Akaike information criterion (AIC) which evaluates model performance and discounts for complexity.

The Markovian nature of the model also allows to implement a simulation algorithm. Let us consider a generic health zone  $k$  with discretized cholera incidence  $i$  at time  $t$ , i.e.  $X_{k,t} = i$ . Knowing the epidemic state of the other health zones and the magnitude of the environmental forcing at the same time  $t$ , it is possible to compute the transition probabilities  $p'_{i,j,k,t}$  for  $j = 1, 2, 3$  through equations (4.2) and (4.3). A random variable  $U$ , uniformly distributed in the  $[0, 1]$  interval, is drawn to determine which transition occurs. If  $U < p'_{i,1,k,t}$ , the considered health zone transitions to the low incidence state in the next month, i.e.  $X_{k,t+1} = 1$ . Otherwise, if  $U < p'_{i,1,k,t} + p'_{i,2,k,t}$ , a transition to a mild cholera state occurs, i.e.  $X_{k,t+1} = 2$ . In the remaining case the transition is to a high incidence state, i.e.  $X_{k,t+1} = 3$ . Repeating this procedure for all the health zones gives a 1-month time-step simulation. The simulated state can be used to advance the chain for another time-step and so on to simulate the model for any number of time-steps.

To evaluate the predictive ability of the different models, we perform a validation analysis. Specifically, we perform leave-one-out cross-validation, i.e. we remove one month of data

for all the health zones and recalibrate the model being tested using the remaining data. Starting from the state of the system observed in the month before the one removed, we simulate the model for one time-step, using the newly calibrated parameter set, and compare model prediction to the removed data. The accuracy of the different models in validation is estimated by computing the likelihood of the observed state. To that end, we infer the probability distribution of the predicted state performing 10000 independent simulations. This procedure is then sequentially repeated removing, once at a time, all the monthly data points available. We also perform a validation analysis removing 2 and 3 contiguous months of data at a time. In this case the chain is advanced for 2 and 3 time-steps, respectively. Performance is evaluated through the likelihood of the state observed in the latest month removed. Note that likelihood values evaluated at different lags or between validation and calibration runs cannot be compared because of different numbers of data points.

## 4.4 Results

Table 4.1 shows the results of the calibration procedure described above. The 20 best combinations of model components are shown, ranked according to their AIC score. All 64 possible combinations are shown in Appendix A.2. The best ranked model accounts for the effects of SST anomalies and seasonality only (parameters are shown in Table 4.2). However, models number 2 to 14 have an AIC score close to that of the top-ranked candidate ( $\Delta\text{AIC} < 4$ ) and thus cannot be safely discarded [Burnham and Anderson, 2002]. All these models include seasonality. To test the significance of the individual components of models 1 to 14 (alternative hypothesis) against the model including seasonality only (number 8, null hypothesis) we employ a likelihood ratio test. Improvements in likelihood for models 1 and 2 are significant, and so are the effects of ENSO alone, as well as ENSO combined with precipitation ( $p < 0.05$ ). Improvements in likelihood for models 4 to 7 and 9 to 14 are not significant at  $p = 0.05$ .

Model validation is performed using the 14 models retained in model selection. Table 4.3 shows log-likelihood values obtained by applying cross-validation at lags of one, two and three months. Model 13 has the highest likelihood values for all lags. Figure 4.4 shows the validation of model 13 at respective lags of one, two and three months. Note the decreasing accuracy of the median as predictor of cholera incidence as well as the higher uncertainty of the simulations as the lag increases. A less formal but more intuitive measure of model accuracy is the fraction of times in which the model correctly predicts the observed cholera incidence state. If we assume the mode of the distribution over 10000 runs as the best predictor, model 13 predicts 68% of the state correctly at lag one month, 62% at lag two months and 59% at lag three months.

In addition to the results reported above, we also tested the effect of adding the water surface temperature of Lake Kivu [MacCallum and Merchant, 2012; Thiery et al., 2014a, b] as a further explanatory variable. No significant improvements were found (result not shown for brevity).

**Chapter 4. Cholera in the Lake Kivu region (DRC): Integrating remote sensing and spatially explicit epidemiological modeling**

Table 4.1 – Results of the fitting procedure ordered by increasing AIC score (first 20 lines).<sup>a</sup>

Model	Seasonality	Mobility	Precipitation	Chlorophyll <i>a</i>	ENSO	IOD	Degrees of freedom <sup>b</sup>	log(Likelihood)	AIC	ΔAIC
1	+	-	-	-	+	-	11	-470.02	962.03	0
2	+	-	+	-	+	-	12	-469.22	962.45	0.4
3	+	-	-	-	+	+	13	-468.89	963.78	1.6
4	+	-	-	+	+	-	13	-468.95	963.89	1.9
5	+	-	-	-	-	+	11	-471.03	964.05	2.0
6	+	-	+	-	+	+	14	-468.07	964.15	2.1
7	+	-	-	+	-	+	13	-469.26	964.52	2.5
8	+	-	-	-	-	-	9	-473.27	964.53	2.5
9	+	-	+	+	+	-	14	-468.37	964.74	2.7
10	+	-	-	+	+	+	15	-467.38	964.76	2.7
11	+	-	+	-	-	+	12	-470.6	965.2	3.2
12	+	-	+	-	-	-	10	-472.84	965.67	3.6
13	+	+	+	-	+	-	15	-467.84	965.68	3.6
14	+	-	+	+	+	+	16	-466.9	965.79	3.8
15	+	-	+	+	-	+	14	-469.16	966.31	4.3
16	+	+	+	-	-	-	13	-470.27	966.53	4.5
17	+	-	-	+	-	-	11	-472.36	966.72	4.7
18	+	+	-	-	+	-	14	-469.75	967.51	5.5
19	+	+	-	-	-	-	12	-471.85	967.71	5.7
20	+	-	+	+	-	-	12	-471.95	967.89	5.9

<sup>a</sup> See Appendix A.2 for all 64 lines.

<sup>b</sup> Number of parameters plus one (residual variance)



Table 4.2 – Parameter sets corresponding to the best ranked models in calibration and cross-validation (respectively models 1 and 13 in Table 4.1).

	<b>Model 1</b>	<b>Model 13</b>
$p_{1,1}$	0.8841	0.9203
$p_{2,1}$	0.3822	0.4021
$p_{3,1}$	0.1151	0.1394
$p_{1,3}$	0.02575	0.01879
$p_{2,3}$	0.1366	0.09161
$p_{3,3}$	0.3736	0.2446
$\delta$	0.4029	0.3702
$t_s$	6.358	5.755
$\gamma$		0.2697
$D$		30.53
$\nu$		0.7578
$\alpha$		0.5598
$\phi$	0.5014	0.3642
$t_{ENSO}$	0	0

## 4.5 Discussion

In this work we have applied a MDIMC-based approach to model cholera dynamics in eight health zones in the Lake Kivu region (DRC). The framework chosen allows for a mechanistic description of processes such as human mobility or the enhancing effect of rainfall on disease transmission, as well as for an explicit treatment of space. Its discrete nature allows to characterize spatiotemporal cholera dynamics robustly, even if the reported case-data available present high uncertainties because of over- and under-reporting and missing records. This robustness is especially important in endemic regions such as the eastern DRC, where incidence is generally lower than in epidemic settings. Conversely, classical SIR-type models (like, e.g., the one applied in *Rinaldo et al.* [2012] to describe the Haiti cholera epidemic) rely on detailed epidemiological datasets for parameter estimation and are thus very difficult to apply to the current case at a fine spatial resolution because of the low signal to noise ratio.

Several models were retained during model selection. All of them account for seasonality, which is thus found to be an important factor to explain endemic cholera transmission in the study area. The model that performed best according to AIC accounts for the effect of ENSO in addition to seasonality. During validation, though, a more complex model, including also the effects of mobility and precipitation, proved to perform best. This might indicate that higher complexity in this case does not lead to overfitting but to improved predictive abilities.

## Chapter 4. Cholera in the Lake Kivu region (DRC): Integrating remote sensing and spatially explicit epidemiological modeling

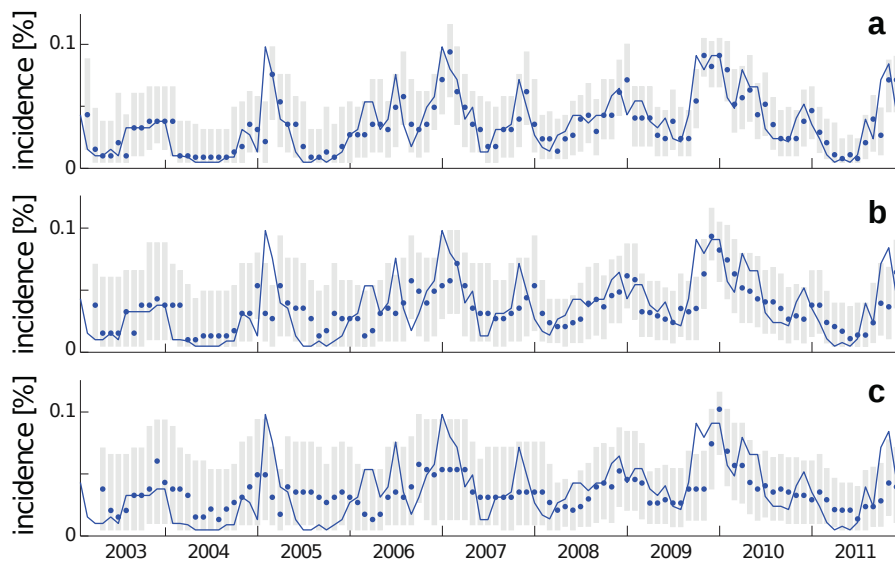


Figure 4.4 – Simulations of model 13, which accounts for mobility, precipitation and SST anomaly in addition to seasonality, at one (a), two (b) and three (c) months lags using the cross-validation procedure described in Section 4.3.2. The blue line shows the average over the states of the eight health zones, where each state is represented by its mean incidence. The blue dots show the median of the simulated values and the grey bars the 5<sup>th</sup>–95<sup>th</sup> percentile range over 10000 simulations.

Global climate anomalies (ENSO and IOD) seem to be the most important environmental factors, as they appear in all but two of the best performing models. Because of the known interactions between these anomalies and local climate [Marchant *et al.*, 2007; Plisnier *et al.*, 2000; Stager *et al.*, 2007] this is not surprising. The effect of precipitation alone did not prove significant compared to a model with seasonality only. However, its combination with ENSO and IOD is significant, presumably because of rainfall deteriorating sanitary conditions.

Chlorophyll *a* concentration in Lake Kivu as a driver of cholera dynamics was retained only by 5 of the models selected in calibration. Apart from the interaction between *V. cholerae* and plankton, possibly more complex than presumed here, significant correlations might be clouded by other environmental factors such as precipitation, or by the fact that the estimates of chlorophyll *a* concentrations used here are of insufficient accuracy and/or spatial detail. The subtleties involved in the remote sensing of chlorophyll *a* concentrations in lakes indeed deserve further investigations [Knox *et al.*, 2014]. However, we maintain that the search for remotely sensed proxies for parameters of mechanistic epidemiological models is an important field of study towards a new concept of mathematical epidemiology.

The effect of human mobility has been retained only by one of the models selected in calibration, which may be an artifact of our decision to limit the study to the lakeside region of Lake Kivu, with only two distant population centers separated by mostly rural areas. Indeed, the most important mobility patterns identified in the eastern DRC are between the lakeside

Table 4.3 – Log-likelihood values obtained during validation by comparing one, two or three months in advance simulations with reported incidence.

Lag (months)	1	2	3
Model <sup>a</sup> 1	-482.35	-559.28	-583.37
Model 2	-482.79	-558.34	-581.98
Model 3	-476.72	-552.91	-579.46
Model 4	-472.56	-549.97	-573.83
Model 5	-472.78	-548.69	-575.05
Model 6	-470.53	-546.38	-571.45
Model 7	-479.03	-556.25	-583.36
Model 8	-483.40	-563.09	-591.90
Model 9	-477.77	-554.54	-578.10
Model 10	-473.55	-549.05	-575.05
Model 11	-475.61	-552.38	-579.58
Model 12	-485.75	-564.61	-592.76
Model 13	-469.67	-543.63	-569.04
Model 14	-471.28	-547.75	-571.63

<sup>a</sup> Model numbering corresponds to the rank obtained according to Table 4.1.

regions and non-lakeside regions [Bompangue *et al.*, 2009] and are thus not accounted for in this study. In addition, since the description of mobility in the model includes three parameters, it is penalized in model selection relative to other model components. Note however that the best performing model in validation accounts for human mobility as well, which thus proves to be important for epidemiological projections.

All models including seasonality have shown to perform better than models not including it. This can be interpreted as a clear indication that other environmental and/or social factors, which have not been explicitly accounted for, might play a major role in the dynamics of the disease in the study area. In particular, the influence of population movements due to war, civil unrest or seasonal migrations [Bompangue *et al.*, 2009], particularly from and to other endemic areas in the country, would merit further investigation.

We have been able to show that cholera incidence in the region is influenced by global (ENSO, IOD) and local (rainfall) climatic variables. Thus it seems clear that climatic and environmental conditions play an important role in the disease dynamics. Our results do not support the hypothesis of phyto- and zooplankton being a major factor for persistence and proliferation of the disease in this area, as indicated by the lack of any significant effect of chlorophyll

#### **Chapter 4. Cholera in the Lake Kivu region (DRC): Integrating remote sensing and spatially explicit epidemiological modeling**

---

*a.* Alternative explanations exist, such as the influence of climate on human behaviour and metapopulation dynamics [Bompangue *et al.*, 2011; Pascual *et al.*, 2002; Rebaudet *et al.*, 2013a; Rodó *et al.*, 2002].

Note that, although the framework applied here allows for prediction of epidemic state a few months in advance, our results did not identify specific environmental drivers with long lead times. Thus, in order to use the model for the purpose of longer-term predictions, projections of environmental drivers have to be made first.

Our results provide further evidence that different geographic and social contexts call for different dominant infection mechanisms, and hence for proxies and modeling approaches shifting on a case-dependent basis. As an example, chlorophyll *a* acts as a suitable proxy in Bangladesh, whereas rainfall is the most important environmental driver in Haiti. The proposed modeling framework is flexible and capable of selecting the dominant infection mechanisms. Thus it can easily be exported to the study of other regions.

## Conclusions and perspectives

In this thesis, a set of metapopulation and individual-based epidemiological models, either mechanistic or semi-mechanistic, has been developed. They have been applied to cholera outbreaks at spatial scales ranging from a city to entire countries. The models explicitly take into account the spatial heterogeneity and temporal variability of epidemiological processes, such as the spread of the disease through hydrological connectivity or human mobility, the micro-scale spatiotemporal clustering of cases in densely populated areas, or the dependence on environmental drivers. The models have been applied to recent cholera outbreaks in Haiti, Senegal, Chad and the Democratic Republic of the Congo, and have been used to tackle real-world epidemiological questions.

Different problem scales in space and time call for different levels of abstraction of epidemiological processes. The spatiotemporal propagation of epidemics has been shown to be realistically reproduced by a detailed and accurate description of hydrological connectivity and human mobility (e.g. from mobile phone call records) at the country scale. At the scale of a single neighborhood, many processes contribute to the spread of the disease, human mobility being just one of them. A description of the infectious pressure by an isotropic spatial kernel around cases has been shown to be an appropriate choice at this scale.

Rainfall and other climatic variables, along with the floods and inundations they may cause, have been shown to be key drivers of cholera in several settings. They need to be taken into account along with their spatial heterogeneity in order to be able to explain highs and lows of transmission and thus to allow models to accurately reproduce decrease and revamping of epidemics.

Applications have demonstrated how models can inform epidemiological policy and show the effect of alternative intervention strategies on the course of epidemics. Using an individual-based stochastic model, the preventive allocation of oral cholera vaccine, antibiotics and/or water, sanitation and hygiene interventions within a given radius around reported cases in densely populated areas have been shown to be effective and efficient alternatives to mass intervention campaigns. Further, using a spatially explicit metapopulation model, an alternative type of oral rehydration solution has been shown to have a significant, large-scale effect on the course of a simulated epidemic.

## Conclusions and perspectives

---

The spatial resolution of the metapopulation models used in this thesis could even be refined in cases where the necessary data is available. This may allow for a more detailed description of local processes by the model, but also require a higher resolution of case data to calibrate on, which are likely to be subject to higher noise and stochasticity the model would have to address. In addition, a spatially explicit parametrization might be necessary in order to account for the heterogeneity of epidemiological processes. It is, however, well known that increasing a model's complexity does not necessarily increase its predictive power, even if it may fit calibration data better. There is no exception when it comes to refining the spatial resolution of a model, critical analysis of the predictive power of more refined models would thus be necessary in order to compare its performance to the one of coarse-grained models.

Even with detailed knowledge regarding epidemiological processes and the use of accurate data at the appropriate scale, models always depend on a set of simplifying assumptions which allow to translate real-world processes into mathematical terms. It lies within the responsibility of the modeler to be aware of those assumptions and of the limitations and uncertainties they cause regarding model results, especially in real-world applications such as the ones shown in this thesis. Model uncertainties have to be explored and suitably conveyed, especially when communicating results to non-modelers. The usefulness of model results depends on their robustness to changes in the underlying assumptions and on the way uncertainties influence the strength of a models conclusions.

Phylogenetic data of pathogen evolution during epidemics and outbreaks has become increasingly used in epidemiological models of infectious diseases. This data allows to directly observe the transmission chain of pathogens, which can reveal important epidemiological processes, e.g. who infected whom at the local scale, or from which geographical area the pathogen got imported to which other area. This data thus constitutes a rich source of information, and may, in the case of cholera, allow to improve the knowledge about the spread of epidemics and the way it is represented in mathematical models. Such data and suitable models may even provide insights into one of the most controversial topics in cholera epidemiology, the one of its long-term persistence and possible environmental reservoirs.

In recent years, decisive progress has been achieved in the domain of epidemiological modeling. Current developments, notably using methods described in this thesis, have shown that real-time predictions of the unfolding of cholera epidemics in space and time is within reach. Models are not meant to replace classical epidemiology, field observations and clinical trials, but rather to provide additional, objective and efficient tools and insights to allow experts to take the right management decisions. The work presented here highlights how suitable models based on accurate data can be used to objectively evaluate different intervention strategies, inform epidemiological policy and may, within the near future, be routinely used to inform experts about the possible course of ongoing epidemics and the impact of epidemic management in real-time.

# A Appendix

## A.1 Supplementary Table to Chapter 3

Table A.1 – Table showing all intervention scenarios considered, specifically showing the type of intervention, the timing and the type of allocation together with median values (2.5% and 97.5% percentiles in brackets) of the number of targeted clusters, the total number of targeted people and the total number of averted cases computed using 1000 model simulations.

<sup>a</sup> Every person can get antibiotics only once during the epidemic.

<sup>b</sup> Every person can get antibiotics several times with a minimum delay of 2 weeks between two administrations.

	Cases averted	People targeted	Clusters targeted	Late	Peak	Early	Entire population	Random	Clusters 15m	Clusters 30m	Clusters 45m	Clusters 70m	Clusters 100m	WASH	OCV (short lag)	OCV	Antibiotics <sup>b</sup>	Antibiotics <sup>a</sup>
1	1280 [-1376, 6772]	20474 [1585, 92962]	1841 [32, 11286]			×							×					×
2	711 [-484, 3448]	20847 [1563, 89378]	1631 [45, 8006]		×								×					×
3	566 [-442, 3034]	19188 [1469, 74292]	1235 [41, 6391]	×									×					×
4	2450 [-132, 10322]	23475 [1756, 150035]	797 [30, 6068]			×							×				×	
5	1380 [1, 5903]	31325 [1499, 139131]	1007 [34, 5210]										×				×	
6	1078 [-15, 4967]	26469 [1785, 110673]	801 [39, 3701]	×									×				×	
7	3161 [30, 13934]	6870 [1297, 29164]	225 [28, 984]										×				×	
8	1920 [43, 8526]	11308 [1308, 60874]	505 [45, 2518]		×								×				×	
9	1407 [24, 6411]	12791 [1141, 55775]	490 [35, 2075]			×							×				×	
10	3297 [45, 14207]	5403 [1096, 17540]	125 [19, 406]										×		×			
11	2205 [85, 9272]	9270 [1144, 41517]	295 [28, 1243]		×								×					
12	1649 [49, 7112]	10820 [1080, 43311]	310 [20, 1263]										×					
13	2358 [-225, 10010]	13844 [1601, 70673]	999 [51, 6026]										×					
14	1154 [-72, 5436]	18034 [1700, 79445]	1291 [63, 5804]		×								×					
15	851 [-97, 4461]	16607 [1437, 68868]	1013 [50, 4566]			×							×					
16	1887 [-974, 9033]	13623 [934, 60624]	1218 [29, 8295]										×					×
17	1113 [-245, 5010]	15202 [1125, 60884]	1225 [42, 6223]		×								×					×
18	764 [-192, 3893]	14332 [823, 54722]	1031 [31, 4982]										×					×
19	2497 [-58, 10566]	17625 [996, 92364]	819 [27, 5291]			×							×					×
20	1389 [-21, 5959]	21551 [1052, 95489]	1033 [33, 5316]		×								×					×



	Cases averted	People targeted	Clusters targeted	Late	Peak	Early	Entire population	Random	Clusters 15m	Clusters 30m	Clusters 45m	Clusters 70m	Clusters 100m	WASH	OCV (short lag)	OCV	Antibiotics <sup>b</sup>	Antibiotics <sup>a</sup>
21	1045 [-25, 4873]	19198 [1045, 78365]	865 [27, 3788]	×								×					×	
22	2912 [10, 13268]	5783 [874, 55967]	307 [30, 2952]			×						×				×		
23	1694 [14, 7730]	10953 [1051, 65069]	671 [58, 3894]		×							×				×		
24	1238 [-3, 5941]	11014 [900, 53146]	614 [39, 2874]			×						×				×		
25	3241 [43, 14252]	4061 [705, 16403]	139 [18, 540]									×			×			
26	2165 [89, 9211]	7483 [821, 36785]	326 [31, 1559]		×							×			×			
27	1563 [44, 7047]	8538 [747, 37135]	350 [25, 1434]									×			×			
28	2247 [-281, 10194]	11006 [1160, 73869]	1049 [49, 6485]			×						×		×	×			
29	1119 [-108, 5011]	14521 [1124, 67190]	1305 [63, 5687]		×							×						
30	798 [-68, 4374]	13268 [1067, 61656]	1046 [57, 4664]									×						
31	2987 [8, 12452]	4521 [504, 25894]	408 [26, 2416]			×						×						×
32	1756 [39, 7444]	7686 [568, 38324]	695 [39, 3381]		×							×						×
33	1216 [4, 5738]	7573 [443, 31077]	664 [29, 2758]									×						×
34	2934 [11, 12860]	5650 [490, 29305]	395 [26, 2167]			×						×					×	
35	1768 [21, 7780]	9451 [591, 45869]	690 [36, 3379]		×							×					×	
36	1270 [11, 5855]	9304 [540, 38308]	640 [31, 2693]									×					×	
37	1814 [-866, 9638]	9835 [639, 90494]	1112 [47, 8681]			×						×					×	
38	893 [-244, 4999]	12657 [1011, 70928]	1408 [80, 6678]									×					×	
39	698 [-202, 4181]	10995 [656, 53954]	1077 [51, 5096]		×							×					×	
40	3002 [20, 13793]	3388 [412, 43473]	249 [22, 2828]			×						×					×	
41	1841 [32, 8280]	6947 [573, 47694]	555 [37, 3377]									×					×	

	Cases averted	People targeted	Clusters targeted	Late	Peak	Early	Entire population	Random	Clusters 15m	Clusters 30m	Clusters 45m	Clusters 70m	Clusters 100m	WASH	OCV (short lag)	OCV	Antibiotics <sup>b</sup>	Antibiotics <sup>a</sup>
42	1369 [7, 6254]	6849 [498, 37048]	494 [27, 2459]	×														
43	1812 [-723, 8719]	10463 [577, 77064]	1395 [41, 8194]			×					×			×				
44	865 [-214, 4374]	12676 [637, 66584]	1543 [71, 7104]		×						×			×				
45	634 [-203, 3733]	10689 [723, 53261]	1187 [70, 5281]	×							×			×				
46	3035 [15, 13289]	3223 [374, 25734]	360 [29, 2175]			×				×								×
47	1778 [42, 7659]	6151 [412, 33533]	684 [37, 3364]		×					×								×
48	1299 [10, 6067]	5660 [392, 29575]	615 [34, 2723]	×						×								×
49	2958 [0, 13412]	3575 [331, 26768]	359 [25, 2254]			×				×							×	
50	1773 [30, 7681]	6700 [441, 37917]	671 [35, 3242]		×					×							×	
51	1286 [10, 5959]	6310 [385, 30108]	591 [29, 2550]	×						×							×	
52	945 [-2049, 6280]	14715 [570, 92375]	2180 [65, 11904]			×				×						×		
53	404 [-678, 3140]	13442 [721, 69200]	1917 [104, 8882]		×					×						×		
54	330 [-581, 2884]	11456 [598, 53477]	1491 [73, 6366]	×						×						×		
55	2522 [-195, 11826]	4622 [380, 63804]	540 [34, 5975]			×				×								
56	1329 [-4, 6964]	8239 [424, 54034]	944 [42, 5295]		×					×						×		
57	1012 [-17, 5293]	7006 [382, 39473]	775 [34, 3962]	×						×						×		
58	1293 [-1108, 6987]	12057 [547, 81687]	1922 [62, 10240]			×				×				×				
59	618 [-530, 3582]	11952 [673, 64496]	1778 [68, 7881]		×					×				×				
60	455 [-407, 3126]	10019 [478, 48993]	1363 [56, 5934]	×						×				×				
61	915 [-2117, 6322]	9401 [279, 51162]	2389 [65, 11387]			×												×
62	431 [-762, 3196]	8107 [322, 38173]	2015 [81, 8761]		×													×

	Cases averted	People targeted	Clusters targeted	Late	Peak	Early	Entire population	Random	Clusters 15m	Clusters 30m	Clusters 45m	Clusters 70m	Clusters 100m	WASH	OCV (short lag)	OCV	Antibiotics <sup>b</sup>	Antibiotics <sup>a</sup>
63	330 [-713, 2865]	6265 [291, 28811]	1524 [67, 6613]	×					×									×
64	923 [-2110, 6137]	9305 [223, 49901]	2336 [53, 11087]		×				×								×	
65	402 [-740, 3083]	8100 [428, 39141]	2003 [110, 9328]			×			×								×	
66	315 [-602, 2598]	6206 [288, 30154]	1537 [77, 6687]	×					×								×	
67	1 [-3722, 3997]	13287 [452, 63520]	3437 [112, 15443]		×				×							×		
68	9 [-1931, 1681]	10049 [516, 42628]	2511 [148, 11062]						×							×		
69	7 [-1345, 1513]	7804 [332, 34311]	1922 [81, 8295]	×					×							×		
70	17 [-4055, 3809]	13778 [306, 64621]	3551 [79, 14855]						×						×			
71	2 [-1882, 1668]	10319 [388, 44873]	2514 [102, 11110]		×				×						×			
72	4 [-1534, 1698]	7860 [401, 34882]	1947 [102, 8134]	×					×									
73	0 [-4073, 4182]	13496 [233, 64426]	3546 [62, 15306]			×			×					×				
74	-1 [-1752, 1667]	10268 [519, 44058]	2558 [124, 10495]		×				×					×				
75	1 [-1628, 1461]	7920 [305, 34753]	1992 [78, 8400]	×					×					×				
76	3277 [35, 14252]	5567 [1212, 17861]	133 [21, 444]										×			×		×
77	2245 [92, 9254]	9364 [986, 40909]	296 [19, 1327]		×								×			×		×
78	1669 [46, 7051]	10957 [970, 47046]	308 [18, 1332]	×									×			×		×
79	2876 [-13, 11565]	9619 [1308, 44802]	536 [24, 3834]										×			×		×
80	1676 [22, 6810]	13531 [1150, 61129]	800 [29, 4085]		×								×			×		×
81	1218 [11, 5494]	13739 [1085, 54961]	647 [29, 3203]			×							×			×		×
82	3237 [22, 14170]	6088 [1221, 22146]	177 [24, 614]										×			×		×
83	2108 [64, 8879]	10176 [1240, 50224]	390 [33, 1866]		×								×			×		×

	Cases averted	People targeted	Clusters targeted	Late	Peak	Early	Entire population	Random	Clusters 15m	Clusters 30m	Clusters 45m	Clusters 70m	Clusters 100m	WASH	OCV (short lag)	OCV	Antibiotics <sup>b</sup>	Antibiotics <sup>a</sup>
84	1513 [32, 6739]	11850 [1225, 49898]	409 [31, 1711]	×												×		
85	3302 [51, 14308]	5227 [1050, 16132]	112 [16, 358]			×							×	×	×	×	×	×
86	2264 [80, 9380]	8774 [979, 37226]	248 [22, 1109]		×													×
87	1681 [57, 7302]	10347 [968, 41397]	270 [19, 1115]	×														×
88	3112 [-4, 13715]	695445 [695445, 695445]	-			×	×											
89	1814 [-488, 8254]	695445 [695445, 695445]	-			×	×											
90	3208 [37, 14195]	6182 [1190, 23294]	173 [25, 692]			×			×	×			×				×	*
91	2105 [58, 8863]	10311 [1210, 50667]	397 [33, 1857]		×				×	×			×				×	*
92	1506 [21, 6713]	11817 [1108, 50888]	410 [34, 1701]	×					×	×			×				×	*
93	3085 [28, 13788]	4990 [850, 31990]	224 [30, 1439]			×			×	×			×				×	*
94	1899 [55, 8484]	9056 [956, 55766]	509 [39, 2726]		×				×	×			×				×	*
95	1379 [14, 6318]	9735 [858, 45672]	496 [32, 2193]	×					×	×			×				×	*
96	2427 [-261, 11517]	6032 [695, 78221]	611 [48, 6467]			×			×	×			×				×	*
97	1243 [-60, 6402]	10674 [849, 64102]	1119 [70, 5633]		×				×	×			×				×	*
98	941 [-30, 4951]	9486 [691, 50970]	860 [49, 4114]	×					×	×			×				×	*
99	1332 [-1439, 7884]	11377 [615, 88610]	1728 [72, 9648]			×			×	×			×				×	*
100	201 [-400, 2540]	8784 [920, 39246]	1352 [122, 6523]		×				×	×			×				×	*
101	545 [-355, 3661]	9923 [529, 49880]	1268 [53, 5722]	×					×	×			×				×	*
102	13 [-3814, 4216]	20474 [1585, 92962]	-			×											×	×
103	0 [-1852, 1885]	20847 [1563, 89378]	-		×												×	×
104	0 [-1487, 1731]	19188 [1469, 74292]	-	×													×	×

	Cases averted	People targeted	Clusters targeted	Late	Peak	Early	Entire population	Random	Clusters 15m	Clusters 30m	Clusters 45m	Clusters 70m	Clusters 100m	WASH	OCV (short lag)	OCV	Antibiotics <sup>b</sup>	Antibiotics <sup>a</sup>
105	1 [-3938, 4052]	23475 [1756, 150035]				×		×									×	
106	10 [-1725, 1702]	31325 [1499, 139131]			×			×									×	
107	13 [-1521, 1551]	26469 [1785, 110673]				×		×									×	
108	148 [-3422, 4448]	6870 [1297, 29164]						×								×		
109	157 [-1342, 2187]	11308 [1308, 60874]			×			×								×		
110	69 [-1158, 1834]	12791 [1141, 55775]						×								×		
111	146 [-3582, 5003]	5403 [1096, 17540]						×							×			
112	108 [-1365, 2209]	9270 [1144, 41517]			×			×							×			
113	93 [-1326, 1628]	10820 [1080, 43311]						×										
114	196 [-3761, 4312]	13844 [1601, 70673]				×		×										
115	50 [-1493, 1745]	18034 [1700, 79445]						×										
116	27 [-1488, 1786]	16607 [1437, 68868]						×										

## A.2 Supplementary Table to Chapter 4

Table A.2 – Results of the fitting procedure ordered by increasing AIC score.

Model	Seasonality	Mobility	Precipitation	Chlorophyll $a$	ENSO	IOD	Degrees of freedom <sup>a</sup>	log(Likelihood)	AIC	$\Delta$ AIC
1	+	-	-	-	+	-	11	-470.02	962.03	0
2	+	-	+	-	+	-	12	-469.22	962.45	0.4
3	+	-	-	-	+	+	13	-468.89	963.78	1.6
4	+	-	-	+	+	-	13	-468.95	963.89	1.9
5	+	-	-	-	-	+	11	-471.03	964.05	2.0
6	+	-	+	-	+	+	14	-468.07	964.15	2.1
7	+	-	-	+	-	+	13	-469.26	964.52	2.5
8	+	-	-	-	-	-	9	-473.27	964.53	2.5
9	+	-	+	+	+	-	14	-468.37	964.74	2.7
10	+	-	-	+	+	+	15	-467.38	964.76	2.7
11	+	-	+	-	-	+	12	-470.6	965.2	3.2
12	+	-	+	-	-	-	10	-472.84	965.67	3.6
13	+	+	+	-	+	-	15	-467.84	965.68	3.6
14	+	-	+	+	+	+	16	-466.9	965.79	3.8
15	+	-	+	+	-	+	14	-469.16	966.31	4.3
16	+	+	+	-	-	-	13	-470.27	966.53	4.5
17	+	-	-	+	-	-	11	-472.36	966.72	4.7
18	+	+	-	-	+	-	14	-469.75	967.51	5.5
19	+	+	-	-	-	-	12	-471.85	967.71	5.7
20	+	-	+	+	-	-	12	-471.95	967.89	5.9
21	+	+	-	-	-	+	14	-470.34	968.68	6.7
22	+	+	+	-	-	+	15	-469.39	968.78	6.7
23	+	+	+	-	+	+	17	-467.57	969.13	7.1
24	+	+	+	+	+	-	17	-467.58	969.16	7.1
25	+	+	-	+	+	-	16	-468.7	969.4	7.4
26	+	+	-	-	+	+	16	-468.76	969.52	7.5
27	+	+	+	+	-	-	15	-470.11	970.22	8.2
28	+	+	-	+	-	+	16	-469.19	970.38	8.4
29	+	+	-	+	-	-	14	-471.26	970.52	8.5
30	-	-	-	+	-	+	11	-474.37	970.73	8.7
31	+	+	-	+	+	+	18	-467.39	970.79	8.8
32	+	+	+	+	-	+	17	-468.65	971.29	9.3

Model	Seasonality	Mobility	Precipitation	Chlorophyll <i>a</i>	ENSO	IOD	Degrees of freedom <sup>a</sup>	log(Likelihood)	AIC	ΔAIC
33	+	+	+	+	+	+	19	-466.67	971.33	9.3
34	-	-	+	+	-	+	12	-474.26	972.53	10.5
35	-	-	-	-	-	+	9	-478.68	975.36	13.3
36	-	-	+	-	-	+	10	-477.95	975.9	13.9
37	-	-	+	-	+	-	10	-478.06	976.12	14.1
38	-	-	-	-	+	-	9	-479.06	976.13	14.1
39	-	-	-	+	+	+	13	-475.28	976.56	14.5
40	-	+	-	+	-	+	14	-474.28	976.57	14.5
41	-	-	+	-	+	+	12	-476.34	976.69	14.7
42	-	-	-	+	+	-	11	-477.38	976.77	14.7
43	-	-	-	-	+	+	11	-477.41	976.81	14.8
44	-	-	-	-	-	-	7	-481.42	976.84	14.8
45	-	-	+	+	+	+	14	-474.49	976.97	14.9
46	-	+	+	+	+	-	15	-473.51	977.01	15.0
47	-	+	+	+	-	+	15	-473.67	977.33	15.3
48	-	-	+	-	-	-	8	-480.7	977.4	15.4
49	-	+	-	+	+	-	14	-474.77	977.53	15.5
50	-	-	+	+	+	-	12	-476.92	977.83	15.8
51	-	+	-	-	-	+	12	-477.1	978.2	16.2
52	-	-	-	+	-	-	9	-480.16	978.33	16.3
53	-	+	+	-	-	-	11	-478.27	978.54	16.5
54	-	+	+	-	-	+	13	-476.4	978.79	16.8
55	-	-	+	+	-	-	10	-479.62	979.24	17.2
56	-	+	+	+	-	-	13	-476.63	979.25	17.2
57	-	+	-	-	-	-	10	-479.65	979.29	17.3
58	-	+	-	+	-	-	12	-477.96	979.93	17.9
59	-	+	+	-	+	-	13	-477.04	980.09	18.0
60	-	+	-	-	+	-	12	-478.47	980.95	18.9
61	-	+	+	-	+	+	15	-475.75	981.49	19.5
62	-	+	-	-	+	+	14	-476.92	981.84	19.8
63	-	+	+	+	+	+	17	-473.94	981.88	19.9
64	-	+	-	+	+	+	16	-475.19	982.38	20.3

<sup>a</sup> Number of parameters plus one (residual variance)





## Bibliography

- Abrams, J. Y., J. R. Copeland, R. V. Tauxe, K. A. Date, E. D. Belay, R. K. Mody, and E. D. Mintz, Real-time modelling used for outbreak management during a cholera epidemic, Haiti, 2010–2011, *Epidemiology & Infection*, 141(6), 1276–1285, doi: 10.1017/S0950268812001793, 2013.
- Abubakar, A., et al., The First Use of the Global Oral Cholera Vaccine Emergency Stockpile: Lessons from South Sudan, *PLOS Medicine*, 12(11), e1001901, doi: 10.1371/journal.pmed.1001901, 2015.
- Abubakar, I., P. Gautret, G. W. Brunette, L. Blumberg, D. Johnson, G. Poucherol, Z. A. Memish, M. Barbeschi, and A. S. Khan, Global perspectives for prevention of infectious diseases associated with mass gatherings, *The Lancet Infectious Diseases*, 12(1), 66–74, doi: 10.1016/S1473-3099(11)70246-8, 2012.
- Akaike, H., A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19(6), 716–723, doi: 10.1109/TAC.1974.1100705, 1974.
- Akanda, A. S., S. Jutla, and S. Islam, Dual peak cholera transmission in Bengal Delta: a hydro-climatological explanation, *Geophysical Research Letters*, 36, L19,401, 2009.
- Akanda, A. S., A. S. Jutla, D. M. Gute, R. B. Sack, M. Alam, A. Huq, R. R. Colwell, and S. Islam, Population Vulnerability to Biannual Cholera Outbreaks and Associated Macro-Scale Drivers in the Bengal Delta, *The American Journal of Tropical Medicine and Hygiene*, 89(5), 950–959, doi: 10.4269/ajtmh.12-0492, 2013.
- Akeret, J., A. Refregier, A. Amara, S. Seehars, and C. Hasner, Approximate Bayesian computation for forward modeling in cosmology, *Journal of Cosmology and Astroparticle Physics*, 2015(08), 043, doi: 10.1088/1475-7516/2015/08/043, 2015.
- Al-Tawfiq, J., and Z. A. Memish, The Hajj: updated health hazards and current recommendations for 2012, *Eurosurveillance*, 17(41), 6–10, 2012.
- Alam, M. T., T. A. Weppelmann, I. Longini, V. M. B. D. Rochars, J. G. M. Jr, and A. Ali, Increased Isolation Frequency of Toxigenic *Vibrio cholerae* O1 from Environmental Monitoring Sites in Haiti, *PLOS ONE*, 10(4), e0124098, doi: 10.1371/journal.pone.0124098, 2015.

- Alam, M. T., S. S. Ray, C. N. Chun, Z. G. Chowdhury, M. H. Rashid, V. E. M. B. D. Rochars, and A. Ali, Major Shift of Toxigenic *V. cholerae* O1 from Ogawa to Inaba Serotype Isolated from Clinical and Environmental Samples in Haiti, *PLOS Neglected Tropical Diseases*, 10(10), e0005,045, doi: 10.1371/journal.pntd.0005045, 2016.
- Alam, M. T., et al., Monitoring Water Sources for Environmental Reservoirs of Toxigenic *Vibrio cholerae* O1, Haiti, *Emerging Infectious Diseases*, 20(3), doi: 10.3201/eid2003.131293, 2014.
- Ali, M., D. R. Kim, M. Yunus, and M. Emch, Time Series Analysis of Cholera in Matlab, Bangladesh, during 1988-2001, *Journal of Health, Population, and Nutrition*, 31(1), 11–19, 2013a.
- Ali, M., A. R. Nelson, A. L. Lopez, and D. A. Sack, Updated Global Burden of Cholera in Endemic Countries, *PLOS Negl Trop Dis*, 9(6), e0003,832, doi: 10.1371/journal.pntd.0003832, 2015.
- Ali, M., et al., Herd protection by a bivalent-killed-whole-cell oral cholera vaccine in the slums of Kolkata, India, *Clinical Infectious Diseases*, p. cit009, doi: 10.1093/cid/cit009, 2013b.
- Ali, M., et al., Potential for Controlling Cholera Using a Ring Vaccination Strategy: Re-analysis of Data from a Cluster-Randomized Clinical Trial, *PLOS Med*, 13(9), e1002,120, doi: 10.1371/journal.pmed.1002120, 2016.
- Altizer, S., A. Dobson, P. Hosseini, P. Hudson, M. Pascual, and P. Rohani, Seasonality and the dynamics of infectious diseases, *Ecology Letters*, 9(4), 467–484, doi: 10.1111/j.1461-0248.2005.00879.x, 2006.
- Anderson, R. M., and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, 1992.
- Andrews, J. R., and S. Basu, The transmission dynamics and control of cholera in Haiti: An epidemic model, *The Lancet*, 377(9773), 1248–1255, doi: 10.1016/S0140-6736(11)60273-0, 2011.
- Arifin, S. M. N. M., *Spatial Agent-Based Simulation Modeling in Public Health*, Wiley, 2016.
- Atia, A. N., and A. L. Buchman, Oral Rehydration Solutions in Non-Cholera Diarrhea: A Review, *The American Journal of Gastroenterology*, 104(10), 2596–2604, doi: 10.1038/ajg.2009.329, 2009.
- Azarian, T., et al., Non-toxigenic environmental *Vibrio cholerae* O1 strain from Haiti provides evidence of pre-pandemic cholera in Hispaniola, *Scientific Reports*, 6, 36,115, doi: 10.1038/srep36115, 2016.
- Azman, A. S., and J. Lessler, Reactive vaccination in the presence of disease hotspots, *Proceedings of the Royal Society B: Biological Sciences*, 282(1798), 20141,341, doi: 10.1098/rspb.2014.1341, 2015.

- Azman, A. S., F. J. Luquero, A. Rodrigues, P. P. Palma, R. F. Grais, C. N. Banga, B. T. Grenfell, and J. Lessler, Urban cholera transmission hotspots and their implications for reactive vaccination: Evidence from Bissau City, Guinea Bissau, *PLOS Neglected Tropical Diseases*, 6(11), e1901, doi: 10.1371/journal.pntd.0001901, 2012.
- Azman, A. S., K. E. Rudolph, D. A. T. Cummings, and J. Lessler, The incubation period of cholera: A systematic review, *Journal of Infection*, 66(5), 432–438, doi: 10.1016/j.jinf.2012.11.013, 2013.
- Azman, A. S., F. J. Luquero, I. Ciglenecki, R. F. Grais, D. A. Sack, and J. Lessler, The Impact of a One-Dose versus Two-Dose Oral Cholera Vaccine Regimen in Outbreak Settings: A Modeling Study, *PLOS Medicine*, 12(8), e1001867, doi: 10.1371/journal.pmed.1001867, 2015.
- Azman, A. S., et al., Effectiveness of one dose of oral cholera vaccine in response to an outbreak: A case-cohort study, *The Lancet Global Health*, 4(11), e856–e863, doi: 10.1016/S2214-109X(16)30211-X, 2016.
- Azman, A. S., et al., Micro-hotspots of Cholera Risk in Urban Epidemics: Rationale for Case-Centered Cholera Control, in prep.
- Bajardi, P., C. Poletto, J. J. Ramasco, M. Tizzoni, V. Colizza, and A. Vespignani, Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic, *PLoS ONE*, 6(1), e16591, doi: 10.1371/journal.pone.0016591, 2011.
- Balcan, D., V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, and A. Vespignani, Multiscale mobility networks and the spatial spreading of infectious diseases, *Proceedings of the National Academy of Sciences of the United States of America*, 106(51), 21,484–21,489, doi: 10.1073/pnas.0906910106, 2009.
- Band, L. E., Topographic Partition of Watersheds with Digital Elevation Models, *Water Resources Research*, 22(1), 15–24, doi: 10.1029/WR022i001p00015, 1986.
- Baracchini, T., A. A. King, M. J. Bouma, X. Rodó, E. Bertuzzo, and M. Pascual, Seasonality in cholera dynamics: A rainfall-driven model explains the wide range of patterns in endemic areas, *Advances in Water Resources*, doi: 10.1016/j.advwatres.2016.11.012, 2016.
- Barzilay, E. J., et al., Cholera Surveillance during the Haiti Epidemic — The First 2 Years, *New England Journal of Medicine*, doi: 10.1056/NEJMoa1204927, 2013.
- Bauernfeind, A., A. Croisier, J.-F. Fesselet, M. van Herp, E. Le Saoût, J. Mc Cluskey, and W. Tuynman, *Cholera Guidelines*, 2 ed., Médecins Sans Frontières, Paris, 2004.
- Beaumont, M. A., J.-M. Cornuet, J.-M. Marin, and C. P. Robert, Adaptive approximate Bayesian computation, *Biometrika*, p. asp052, doi: 10.1093/biomet/asp052, 2009.
- Becker, M., W. Llovel, A. Cazenave, A. Güntner, and J.-F. Crétaux, Recent hydrological behavior of the East African Great Lakes region inferred from grace, satellite altimetry and rainfall

- observations, *Comptes Rendus Geoscience*, 342(3), 223–233, doi: 10.1016/j.crte.2009.12.010, 2010.
- Bengtsson, L., X. Lu, A. Thorson, R. Garfield, and J. von Schreeb, Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti, *PLoS Medicine*, 8(8), e1001083, doi: 10.1371/journal.pmed.1001083, 2011.
- Bengtsson, L., J. Gaudart, X. Lu, S. Moore, E. Wetter, K. Sallah, S. Rebaudet, and R. Piarroux, Using mobile phone data to predict the spatial spread of cholera, *Scientific Reports*, 5, 8923, doi: 10.1038/srep08923, 2015.
- Bertuzzo, E., S. Azaele, A. Maritan, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo, On the space-time evolution of a cholera epidemic, *Water Resources Research*, 44(1), W01424, doi: 10.1029/2007WR006211, 2008.
- Bertuzzo, E., R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo, On spatially explicit models of cholera epidemics, *Journal of The Royal Society Interface*, 7(43), 321–333, doi: 10.1098/rsif.2009.0204, 2010.
- Bertuzzo, E., L. Mari, L. Righetto, M. Gatto, R. Casagrandi, M. Blokesch, I. Rodriguez-Iturbe, and A. Rinaldo, Prediction of the spatial evolution and effects of control measures for the unfolding haiti cholera outbreak, *Geophysical Research Letters*, 38, L06403, doi: 10.1029/2011GL046823, 2011.
- Bertuzzo, E., L. Mari, L. Righetto, M. Gatto, R. Casagrandi, I. Rodriguez-Iturbe, and A. Rinaldo, Hydroclimatology of dual-peak cholera epidemics: inferences from a spatially explicit model, *Geophysical Research Letters*, 39, L05403, 2012.
- Bertuzzo, E., F. Finger, L. Mari, M. Gatto, and A. Rinaldo, On the probability of extinction of the Haiti cholera epidemic, *Stochastic Environmental Research and Risk Assessment*, 30(8), 2043–2055, doi: 10.1007/s00477-014-0906-3, 2016.
- Bhattacharya, S. K., et al., 5 year efficacy of a bivalent killed whole-cell oral cholera vaccine in Kolkata, India: A cluster-randomised, double-blind, placebo-controlled trial, *The Lancet Infectious Diseases*, 13(12), 1050–1056, doi: 10.1016/S1473-3099(13)70273-1, 2013.
- Blackburn, J. K., et al., Household-Level Spatiotemporal Patterns of Incidence of Cholera, Haiti, 2011, *Emerging Infectious Diseases*, 20(9), 1516–1519, doi: 10.3201/eid2009.131882, 2014.
- Bompangue, D., Dynamique des epidemies de cholera dans la region des Grands Lacs Africains: cas de la République Démocratique du Congo, Ph.D. thesis, University of Franche-Compté, 2009.
- Bompangue, D., P. Giraudoux, P. Handschumacher, M. Piarroux, B. Sudre, M. Ekwanzala, I. Kebela, and R. Piarroux, Lakes as sources of cholera outbreaks, Democratic Republic of Congo, *Emerging Infectious Diseases*, 14, 798–800, 2008.

- Bompangue, D., P. Giraudoux, M. Piarroux, G. Mutombo, R. Shamavu, B. Sudre, A. Mutombo, V. Mondonge, and R. Piarroux, Cholera epidemics, war and disasters around Goma and Lake Kivu: An eight-year study, *PLoS Neglected Tropical Diseases*, 3, e436, 2009.
- Bompangue, D., S. M. Vesenbeckh, P. Giraudoux, M. Castro, J. J. Muyembe, B. K. Ilunga, and M. Murray, Cholera ante portas – the re-emergence of cholera in Kinshasa after a ten-year hiatus, *PLoS Currents Disasters*, 4, 2012.
- Bompangue, D., et al., Dynamics of cholera outbreaks in Great Lakes region of Africa, 1978–2008, *Emerging Infectious Diseases*, 17, 2026–2036, 2011.
- Boone, C., *Political Topographies of the African State: Territorial Authority and Institutional Choice*, Cambridge University Press, Cambridge, UK, 2003.
- Bouma, M. J., and M. Pascual, Seasonal and interannual cycles of endemic cholera in Bengal 1891–1940 in relation to climate and geography, *Hydrobiologia*, 460, 147–156, 2001.
- Burnham, K. P., and D. R. Anderson, *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*, Springer, 2002.
- Butler, S. M., E. J. Nelson, N. Chowdhury, S. M. Faruque, S. B. Calderwood, and A. Camilli, Cholera stool bacteria repress chemotaxis to increase infectivity, *Molecular Microbiology*, 60(2), 417–426, doi: 10.1111/j.1365-2958.2006.05096.x, 2006.
- Candia, J., M. C. González, P. Wang, T. Schoenharl, G. Madey, and A.-L. Barabási, Uncovering individual and collective human dynamics from mobile phone records, *Journal of Physics A: Mathematical and Theoretical*, 41(22), 224,015, doi: 10.1088/1751-8113/41/22/224015, 2008.
- Capasso, and Paveri-Fontana, A mathematical model for the 1973 cholera epidemic in the European Mediterranean region., *Revue d'épidémiologie et de santé publique*, 27(2), 121–132, 1979.
- Carrel, M., M. Emch, P. K. Streatfield, and M. Yunus, Spatio-temporal clustering of cholera: The impact of flood control in Matlab, Bangladesh, 1983–2003, *Health & Place*, 15(3), 771–782, doi: 10.1016/j.healthplace.2008.12.008, 2009.
- Central Intelligence Agency, *2009 CIA's World Factbook*, 2009.
- Central Intelligence Agency, *The World Factbook 2013-14*, Central Intelligence Agency, Washington, DC, 2013.
- Chao, D. L., M. E. Halloran, and I. M. Longini, Vaccination strategies for epidemic cholera in haiti with implications for the developing world, *Proceedings of the National Academy of Sciences of the United States of America*, 108(17), 7081–7085, doi: 10.1073/pnas.1102149108, 2011.

- Chin, C.-S., et al., The Origin of the Haitian Cholera Outbreak Strain, *New England Journal of Medicine*, 364(1), 33–42, doi: 10.1056/NEJMoa1012928, 2011.
- Chun, J., et al., Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*, *Proceedings of the National Academy of Sciences of the United States of America*, 106(36), 15,442–15,447, doi: 10.1073/pnas.0907787106, 2009.
- Ciglenecki, I., K. Sakoba, F. J. Luquero, M. Heile, C. Itama, M. Mengel, R. F. Grais, F. Verhoustraeten, and D. Legros, Feasibility of Mass Vaccination Campaign with Oral Cholera Vaccines in Response to an Outbreak in Guinea, *PLOS Medicine*, 10(9), e1001512, doi: 10.1371/journal.pmed.1001512, 2013.
- Codeço, C. T., Endemic and epidemic dynamics of cholera: The role of the aquatic reservoir, *BMC Infectious Diseases*, 1(1), 1, doi: 10.1186/1471-2334-1-1, 2001.
- Colizza, V., A. Barrat, M. Barthélemy, and A. Vespignani, The role of the airline transportation network in the prediction and predictability of global epidemics, *Proceedings of the National Academy of Sciences of the United States of America*, 103(7), 2015–2020, doi: 10.1073/pnas.0510525103, 2006.
- Colwell, R. R., Global climate and infectious disease: The cholera paradigm, *Science*, 274(5295), 2025–2031, 1996.
- Colwell, R. R., and W. M. Spira, The Ecology of *Vibrio cholerae*, in *Cholera*, edited by D. Barua and W. B. G. III, Current Topics in Infectious Disease, pp. 107–127, Springer US, doi: 10.1007/978-1-4757-9688-9\_6, 1992.
- Colwell, R. R., J. Kaper, and S. W. Joseph, *Vibrio cholerae*, *Vibrio parahaemolyticus*, and Other Vibrios: Occurrence and Distribution in Chesapeake Bay, *Science*, 198(4315), 394–396, 1977.
- de Magny, G. C., B. Cazelles, and J.-F. Guégan, Cholera threat to humans in Ghana is influenced by both global and regional climatic variability, *EcoHealth*, 3(4), 223–231, doi: 10.1007/s10393-006-0061-5, 2006.
- de Magny, G. C., et al., Environmental signatures associated with cholera epidemics, *Proceedings of the National Academy of Sciences of the United States of America*, 105(46), 17,676–17,681, doi: 10.1073/pnas.0809654105, 2008.
- de Magny, G. C., et al., Cholera outbreak in Senegal in 2005: Was climate a factor?, *PLoS ONE*, 7(8), e44,577, doi: 10.1371/journal.pone.0044577, 2012.
- de Montjoye, Y.-A., C. A. Hidalgo, M. Verleysen, and V. D. Blondel, Unique in the Crowd: The privacy bounds of human mobility, *Scientific Reports*, 3, 1376, doi: 10.1038/srep01376, 2013.
- de Montjoye, Y.-A., Z. Smoreda, R. Trinquart, C. Ziemlicki, and V. D. Blondel, D4d-Senegal: The second mobile phone data for development challenge, *Tech. rep.*, 2014.

- Debes, A. K., M. Ali, A. S. Azman, M. Yunus, and D. A. Sack, Cholera cases cluster in time and space in Matlab, Bangladesh: Implications for targeted preventive interventions, *International Journal of Epidemiology*, p. dyw267, doi: 10.1093/ije/dyw267, 2016a.
- Debes, A. K., et al., Evaluation in Cameroon of a Novel, Simplified Methodology to Assist Molecular Microbiological Analysis of *V. cholerae* in Resource-Limited Settings, *PLOS Neglected Tropical Diseases*, 10(1), e0004307, doi: 10.1371/journal.pntd.0004307, 2016b.
- Desai, S. N., L. Pezzoli, S. Martin, A. Costa, C. Rodriguez, D. Legros, and W. Perea, A second affordable oral cholera vaccine: Implications for the global vaccine stockpile, *The Lancet Global Health*, 4(4), e223–e224, doi: 10.1016/S2214-109X(16)00037-1, 2016.
- Dijkstra, E. W., A note on two problems in connexion with graphs, *Numerische Mathematik*, 1(1), 269–271, doi: 10.1007/BF01386390, 1959.
- Dutta, D., M. Bhattacharya, A. Deb, D. Sarkar, A. Chatterjee, A. Biswas, K. Chatterjee, G. Nair, and S. Bhattacharya, Evaluation of oral hypo-osmolar glucose-based and rice-based oral rehydration solutions in the treatment of cholera in children, *Acta Paediatrica*, 89(7), 787–790, doi: 10.1111/j.1651-2227.2000.tb00386.x, 2000.
- Echenberg, M., *Africa in the Time of Cholera: A History of Pandemics from 1817 to the Present*, Cambridge University Press, Cambridge, United Kingdom, 2011.
- Echevarria, J., C. Seas, C. Carrillo, R. Mostorino, R. Ruiz, and E. Gotuzzo, Efficacy and Tolerability of Ciprofloxacin Prophylaxis in Adult Household Contacts of Patients with Cholera, *Clinical Infectious Diseases*, 20(6), 1480–1484, doi: 10.1093/clinids/20.6.1480, 1995.
- Eisenberg, M. C., G. Kujbida, A. R. Tuite, D. N. Fisman, and J. H. Tien, Examining rainfall and cholera dynamics in Haiti using statistical and dynamic modeling approaches, *Epidemics*, 5, 197–207, 2013.
- Emch, M., C. Feldacker, M. Yunus, P. K. Streatfield, V. D. Thiem, D. G. Canh, and M. Ali, Local environmental predictors of cholera in Bangladesh and Vietnam, *American Journal of Tropical Medicine and Hygiene*, 78, 823–832, 2008.
- Enserink, M., Haiti's Outbreak Is Latest in Cholera's New Global Assault, *Science*, 330(6005), 738–739, doi: 10.1126/science.330.6005.738, 2010.
- Erlander, S., and N. F. Stewart, *The Gravity Model in Transportation Analysis – Theory and Extensions*, VSP Books, Zeist, The Netherlands, 1990.
- Esch, T., H. Taubenböck, A. Roth, W. Heldens, A. Felbier, M. Thiel, M. Schmidt, A. Müller, and S. Dech, TanDEM-X mission – new perspectives for the inventory and monitoring of global settlement patterns, *Journal of Applied Remote Sensing*, 6(1), 061,702–1, doi: 10.1117/1.JRS.6.061702, 2012.

- Esch, T., et al., Urban Footprint Processor – Fully Automated Processing Chain Generating Settlement Masks From Global Data of the TanDEM-X Mission, *IEEE Geoscience and Remote Sensing Letters*, 10(6), 1617–1621, doi: 10.1109/LGRS.2013.2272953, 2013.
- Evensen, G., The ensemble Kalman filter for combined state and parameter estimation, *IEEE Control Systems Magazine*, 29(3), 83–104, doi: 10.1109/MCS.2009.932223, 2009.
- Farmer, P., et al., Meeting Cholera’s Challenge to Haiti and the World: A Joint Statement on Cholera Prevention and Care, *PLOS Neglected Tropical Diseases*, 5(5), e1145, doi: 10.1371/journal.pntd.0001145, 2011.
- Farthing, M. J., Oral rehydration: an evolving solution, *Journal of pediatric gastroenterology and nutrition*, 34, S64–S67, 2002.
- Faruque, S. M., K. Biswas, S. M. N. Udden, Q. S. Ahmad, D. A. Sack, G. B. Nair, and J. J. Mekalanos, Transmissibility of cholera: In vivo-formed biofilms and their relationship to infectivity and persistence in the environment, *Proceedings of the National Academy of Sciences of the United States of America*, 103(16), 6350–6355, doi: 10.1073/pnas.0601277103, 2006.
- Fernández, M. A. L., A. Bauernfeind, J. D. Jiménez, C. L. Gil, N. E. Omeiri, and D. H. Guibert, Influence of temperature and rainfall on the evolution of cholera epidemics in Lusaka, Zambia, 2003–2006: analysis of a time series, *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 103(2), 137–143, doi: 10.1016/j.trstmh.2008.07.017, 2009.
- Fewtrell, L., and J. M. J. Colford, Water, sanitation and hygiene: Interventions and diarrhoea: A systematic review and meta-analysis, *Tech. Rep. 34960*, The World Bank, 2004.
- Fewtrell, L., R. B. Kaufmann, D. Kay, W. Enanoria, L. Haller, and J. M. Colford Jr, Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: A systematic review and meta-analysis, *The Lancet Infectious Diseases*, 5(1), 42–52, doi: 10.1016/S1473-3099(04)01253-8, 2005.
- Filippi, S., C. P. Barnes, J. Cornebise, and M. P. Stumpf, On optimality of kernels for approximate Bayesian computation using sequential Monte Carlo, *Statistical Applications in Genetics and Molecular Biology*, 12(1), 87–107, doi: 10.1515/sagmb-2012-0069, 2013.
- Finger, F., A. Knox, E. Bertuzzo, L. Mari, D. Bompangue, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo, Cholera in the Lake Kivu region (DRC): Integrating remote sensing and spatially explicit epidemiological modeling, *Water Resources Research*, pp. 5624–5637, doi: 10.1002/2014WR015521, 2014.
- Finger, F., T. Genolet, L. Mari, G. C. de Magny, N. M. Manga, A. Rinaldo, and E. Bertuzzo, Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks, *Proceedings of the National Academy of Sciences of the United States of America*, 113(23), 6421–6426, doi: 10.1073/pnas.1522305113, 2016.



- Ford, T., R. R. Colwell, J. B. Rose, S. S. Morse, D. J. Rogers, and T. L. Yates, Using satellite images of environmental changes to predict infectious disease outbreaks, *Emerging Infectious Diseases*, 15(9), doi: 10.3201/Eid1509.081334, 2009.
- Foreman-Mackey, D., D. W. Hogg, D. Lang, and J. Goodman, Emcee: The MCMC hammer, *Publications of the Astronomical Society of the Pacific*, 125(925), 306–312, doi: 10.1086/670067, 2013.
- Frerichs, R. R., P. Keim, R. Barraï, and R. Piarroux, Nepalese origin of cholera epidemic in Haiti, *Clinical Microbiology and Infection*, 18(6), E158–E163, doi: 10.1111/j.1469-0691.2012.03841.x, 2012.
- Funk, S., M. Salathe, and V. A. A. Jansen, Modelling the influence of human behaviour on the spread of infectious diseases: A review, *Journal of The Royal Society Interface*, 7(50), 1247–1256, doi: 10.1098/rsif.2010.0142, 2010.
- Gatto, M., L. Mari, E. Bertuzzo, R. Casagrandi, L. Righetto, I. Rodriguez-Iturbe, and A. Rinaldo, Generalized reproduction numbers and the prediction of patterns in waterborne disease, *Proceedings of the National Academy of Sciences of the United States of America*, 109(48), 19,703–19,708, doi: 10.1073/pnas.1217567109, 2012.
- Gatto, M., L. Mari, E. Bertuzzo, R. Casagrandi, L. Righetto, I. Rodriguez-Iturbe, and A. Rinaldo, Spatially explicit conditions for waterborne pathogen invasion, *The American Naturalist*, 182(3), 328–346, doi: 10.1086/671258, 2013.
- Gaudart, J., S. Rebaudet, R. Barraï, J. Boncy, B. Faucher, M. Piarroux, R. Magloire, G. Thimothe, and R. Piarroux, Spatio-temporal dynamics of cholera during the first year of the epidemic in Haiti, *PLOS Neglected Tropical Diseases*, 7(4), e2145, doi: 10.1371/journal.pntd.0002145, 2013.
- Gil, A., et al., Occurrence And Distribution Of *Vibrio cholerae* In The Coastal Environment Of Peru, *Environmenta Microbiology*, 6(7), 699–706, doi: 10.1111/J.1462-2920.2004.00601.X, 2004.
- Glass, R. I., S. Becker, M. I. Huq, B. J. Stoll, M. Khan, M. H. Merson, J. V. Lee, and R. E. Black, Endemic cholera in rural bangladesh, 1966–1980, *American Journal of Epidemiology*, 116(6), 959–970, 1982.
- Gore, S. M., O. Fontaine, and N. F. Pierce, Impact of rice based oral rehydration solution on stool output and duration of diarrhoea: Meta-analysis of 13 clinical trials., *BMJ : British Medical Journal*, 304(6822), 287–291, 1992.
- Grabowski, M. K., et al., The Role of Viral Introductions in Sustaining Community-Based HIV Epidemics in Rural Uganda: Evidence from Spatial Clustering, Phylogenetics, and Egocentric Transmission Models, *PLOS Medicine*, 11(3), e1001,610, doi: 10.1371/journal.pmed.1001610, 2014.

- Grad, Y. H., J. C. Miller, and M. Lipsitch, Cholera modeling: Challenges to quantitative analysis and predicting the impact of interventions, *Epidemiology*, 23(4), 523–530, doi: 10.1097/EDE.0b013e3182572581, 2012.
- Guarino, A., F. Albano, S. Guandalini, and Working Group on Acute Gastroenteritis, Oral Rehydration: Toward a Real Solution, *Journal of Pediatric Gastroenterology and Nutrition*, 33, 2001.
- Guévert, É., J. Noeske, J. Sollé, A. Mouangue, and J.-M. Bikoti, Antibioprophylaxie ciblée à large échelle au cours de l'épidémie de choléra de Douala en 2004, *Cahiers d'études et de recherches francophones / Santé*, 17(2), 63–68, doi: 10.1684/san.2007.0072, 2007.
- Gurarie, D., and E. Y. W. Seto, Connectivity sustains disease transmission in environments with low potential for endemicity: Modelling schistosomiasis with hydrologic and social connectivities, *Journal of the Royal Society Interface*, 6(35), 495–508, doi: 10.1098/rsif.2008.0265, 2009.
- Hendriksen, R. S., et al., Population Genetics of *Vibrio cholerae* from Nepal in 2010: Evidence on the Origin of the Haitian Outbreak, *mBio*, 2(4), e00157–11, doi: 10.1128/mBio.00157-11, 2011.
- Hill, V. R., et al., Toxigenic *Vibrio cholerae* O1 in Water and Seafood, Haiti, *Emerging Infectious Diseases*, 17(11), 2147–2150, doi: 10.3201/eid1711.110748, 2011.
- Hitchings, M. T., R. F. Grais, and M. Lipsitch, Using simulation to aid trial design: Ring-vaccination trials, *bioRxiv*, p. 071498, doi: 10.1101/071498, 2016.
- Hornick, R. B., S. I. Music, R. Wenzel, R. Cash, J. P. Libonati, M. J. Snyder, and T. E. Woodward, The Broad Street pump revisited: Response of volunteers to ingested cholera vibrios., *Bulletin of the New York Academy of Medicine*, 47(10), 1181–1191, 1971.
- Huffman, G. J., R. F. Adler, D. T. Bolvin, and E. J. Nelkin, The TRMM multi-satellite precipitation analysis (TMPA), in *Satellite Rainfall Applications for Surface Hydrology*, edited by M. Gebremichael and F. Hossain, pp. 3–22, Springer Netherlands, 2010.
- Huq, A., E. B. Small, P. A. West, M. I. Huq, R. Rahman, and R. R. Colwell, Ecological relationships between *Vibrio cholerae* and planktonic crustacean copepods., *Applied and Environmental Microbiology*, 45(1), 275–283, 1983.
- International Federation of Red Cross and Red Crescent Societies, Senegal: Cholera final report Dref bulletin no. 05ME020, *Tech. rep.*, 2007.
- Islam, M. S., M. I. K. Jahid, M. M. Rahman, M. Z. Rahman, M. S. Islam, M. S. Kabir, D. A. Sack, and G. K. Schoolnik, Biofilm Acts as a Microenvironment for Plankton-Associated *Vibrio cholerae* in the Aquatic Environment of Bangladesh, *Microbiology and Immunology*, 51(4), 369–379, doi: 10.1111/j.1348-0421.2007.tb03924.x, 2007.

- Islam, M. S., et al., Effects of local climate variability on transmission dynamics of cholera in Matlab, Bangladesh, *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 103(11), 1165–1170, doi: 10.1016/j.trstmh.2009.04.016, 2009.
- Ivers, L. C., New strategies for cholera control, *The Lancet Global Health*, 4(11), e771–e772, doi: 10.1016/S2214-109X(16)30257-1, 2016a.
- Ivers, L. C., Eliminating Cholera Transmission in Haiti, *New England Journal of Medicine*, doi: 10.1056/NEJMp1614104, 2016b.
- Ivers, L. C., et al., Effectiveness of reactive oral cholera vaccination in rural Haiti: A case-control study and bias-indicator analysis, *The Lancet Global Health*, 3(3), e162–e168, doi: 10.1016/S2214-109X(14)70368-7, 2015.
- Jutla, A., A. S. Akanda, and S. Islam, Tracking cholera in coastal regions using satellite observations, *Journal Of The American Water Resources Association*, 46(4), 651–662, doi: 10.1111.J.1752-1688.2010.00448.X, 2010.
- Jutla, A., A. S. Akanda, A. Huq, A. Syed, G. Faruque, R. Colwell, and S. Islam, A water marker monitored by satellites to predict seasonal endemic cholera, *Remote Sensing Letters*, 4(8), 822–831, doi: 10.1080/2150704X.2013.802097, 2013a.
- Jutla, A., A. S. Akanda, and S. Islam, A framework for predicting endemic cholera using satellite derived environmental determinants, *Environmental Modeling & Software*, 47, 148–158, doi: 10.1016/j.envsoft.2013.05.008, 2013b.
- Kaper, J. B., J. G. Morris, and M. M. Levine, Cholera, *Clinical Microbiology Reviews*, 8(1), 48–86, 1995.
- Keeling, M. J., and P. Rohani, *Modeling Infectious Diseases in Humans and Animals*, Princeton University Press, 2008.
- Kelvin, A. A., Cholera outbreak in the Republic of Congo, the Democratic Republic of Congo, and cholera worldwide, *Journal of Infection in Developing Countries*, 5, 137–143, 2011.
- Khan, W. A., D. Saha, A. Rahman, M. A. Salam, J. Bogaerts, and M. L. Bennis, Comparison of single-dose azithromycin and 12-dose, 3-day erythromycin for childhood cholera: A randomised, double-blind trial, *The Lancet*, 360(9347), 1722–1727, doi: 10.1016/S0140-6736(02)11680-1, 2002.
- Kim, J.-H., V. Mogasale, C. Burgess, and T. F. Wierzbza, Impact of oral cholera vaccines in cholera-endemic countries: A mathematical modeling study, *Vaccine*, 34(18), 2113–2120, doi: 10.1016/j.vaccine.2016.03.004, 2016.
- King, A. A., E. L. Ionides, M. Pascual, and M. J. Bouma, Inapparent infections and cholera dynamics, *Nature*, 454(7206), 877–880, doi: 10.1038/nature07084, 2008.

- Kirpich, A., T. A. Weppelmann, Y. Yang, A. Ali, J. G. M. Jr, and I. M. Longini, Cholera Transmission in Ouest Department of Haiti: Dynamic Modeling and the Future of the Epidemic, *PLOS Neglected Tropical Diseases*, 9(10), e0004153, doi: 10.1371/journal.pntd.0004153, 2015.
- Knox, A., E. Bertuzzo, L. Mari, D. Odermatt, E. Verrecchia, and A. Rinaldo, Optimizing a remotely sensed proxy for plankton biomass in Lake Kivu, *International Journal of Remote Sensing*, 35(13), 5219–5238, doi: 10.1080/01431161.2014.939782, 2014.
- Koelle, K., X. Rodó, M. Pascual, M. Yunus, and G. Mostafa, Refractory periods and climate forcing in cholera dynamics, *Nature*, 436(7051), 696–700, doi: 10.1038/nature03820, 2005.
- Koepke, A. A., I. M. Longini, Jr., M. E. Halloran, J. Wakefield, and V. N. Minin, Predictive modeling of cholera outbreaks in Bangladesh, *The Annals of Applied Statistics*, 10(2), 575–595, doi: 10.1214/16-AOAS908, 2016.
- Kucharski, A. J., R. M. Eggo, C. H. Watson, A. Camacho, S. Funk, and W. J. Edmunds, Effectiveness of Ring Vaccination as Control Strategy for Ebola Virus Disease, *Emerging Infectious Diseases*, 22(1), 105–108, doi: 10.3201/eid2201.151410, 2016.
- Kühn, J., F. Finger, E. Bertuzzo, S. Borgeaud, M. Gatto, A. Rinaldo, and M. Blokesch, Glucose- but not rice-based oral rehydration therapy enhances the production of virulence determinants in the human pathogen *Vibrio cholerae*, *PLoS Neglected Tropical Diseases*, 8(12), e3347, doi: 10.1371/journal.pntd.0003347, 2014.
- Leckebusch, G. C., and A. F. Abdussalam, Climate and socioeconomic influences on interannual variability of cholera in Nigeria, *Health & Place*, 34, 107–117, doi: 10.1016/j.healthplace.2015.04.006, 2015.
- Leibovici-Weissman, Y., A. Neuberger, R. Bitterman, D. Sinclair, M. A. Salam, and M. Paul, Antimicrobial drugs for treating cholera, in *Cochrane Database of Systematic Reviews*, John Wiley & Sons, Ltd, 2014.
- Lessler, J., H. Salje, M. K. Grabowski, and D. A. T. Cummings, Measuring Spatial Dependence for Infectious Disease Epidemiology, *PLOS ONE*, 11(5), e0155249, doi: 10.1371/journal.pone.0155249, 2016.
- Levine, M., South America: The return of cholera, *The Lancet*, 338(8758), 45–46, doi: 10.1016/0140-6736(91)90024-J, 1991.
- Levine, M. M., R. E. Black, M. L. Clements, L. Cisneros, D. R. Nalin, and C. R. Young, Duration of Infection-Derived Immunity to Cholera, *Journal of Infectious Diseases*, 143(6), 818–820, doi: 10.1093/infdis/143.6.818, 1981.
- Lewnard, J. A., M. Antillón, G. Gonsalves, A. M. Miller, A. I. Ko, and V. E. Pitzer, Strategies to Prevent Cholera Introduction during International Personnel Deployments: A Computational Modeling Analysis Based on the 2010 Haiti Outbreak, *PLOS Medicine*, 13(1), e1001947, doi: 10.1371/journal.pmed.1001947, 2016.

- Linard, C., M. Gilbert, R. W. Snow, A. M. Noor, and A. J. Tatem, Population distribution, settlement patterns and accessibility across Africa in 2010, *PLoS ONE*, 7(2), e31,743, doi: 10.1371/journal.pone.0031743, 2012.
- Lipp, E. K., A. Huq, and R. R. Colwell, Effects of global climate on infectious disease: the cholera model, *Clinical Microbiology Reviews*, 15(4), 757–770, doi: 10.1128/CMR.15.4.757-770.2002, 2002.
- Lobitz, B., L. Beck, A. Huq, B. Wood, G. Fuchs, A. S. G. Faruque, and R. R. Colwell, Climate and infectious disease: Use of remote sensing for detection of *Vibrio cholerae* by indirect measurement, *Proceedings of the National Academy of Sciences USA*, 97, 1438–1443, 2000.
- Longini, I. M., M. Yunus, K. Zaman, A. K. Siddique, R. B. Sack, and A. Nizam, Epidemic and endemic cholera trends over a 33-year period in Bangladesh, *Journal of Infectious Diseases*, 186(2), 246–251, wOS:000176668500014, 2002.
- Longini, I. M., A. Nizam, M. Ali, M. Yunus, N. Shenvi, and J. D. Clemens, Controlling Endemic Cholera with Oral Vaccines, *PLOS Medicine*, 4(11), e336, doi: 10.1371/journal.pmed.0040336, 2007.
- Lu, X., L. Bengtsson, and P. Holme, Predictability of population displacement after the 2010 Haiti earthquake, *Proceedings of the National Academy of Sciences of the United States of America*, 109(29), 11,576–11,581, doi: 10.1073/pnas.1203882109, 2012.
- Lu, X., E. Wetter, N. Bharti, A. J. Tatem, and L. Bengtsson, Approaching the limit of predictability in human mobility, *Scientific Reports*, 3, 2923, doi: 10.1038/srep02923, 2013.
- Luquero, F. J., C. N. Banga, D. Remartínez, P. P. Palma, E. Baron, and R. F. Grais, Cholera Epidemic in Guinea-Bissau (2008): The Importance of “Place”, *PLoS ONE*, 6(5), e19,005, doi: 10.1371/journal.pone.0019005, 2011.
- MacCallum, S. N., and C. J. Merchant, Surface water temperature observations of large lakes by optimal estimation, *Canadian Journal of Remote Sensing*, 38(01), 25–45, doi: 10.5589/m12-010, 2012.
- Marchant, R., C. Mumbi, S. Behera, and T. Yamagata, The indian ocean dipole – the unsung driver of climatic variability in east africa, *African Journal of Ecology*, 45(1), 4–16, doi: 10.1111/j.1365-2028.2006.00707.x, 2007.
- Mari, L., E. Bertuzzo, L. Righetto, R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo, On the role of human mobility in the spread of cholera epidemics: towards an epidemiological movement ecology, *Ecohydrology*, 5(5), 531–540, doi: 10.1002/eco.262, 2012a.
- Mari, L., E. Bertuzzo, L. Righetto, R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo, Modelling cholera epidemics: the role of waterways, human mobility and sanitation, *Journal of the Royal Society Interface*, 9(67), 376–388, doi: 10.1098/rsif.2011.0304, 2012b.

- Mari, L., E. Bertuzzo, F. Finger, R. Casagrandi, M. Gatto, and A. Rinaldo, On the predictive ability of mechanistic models for the Haitian cholera epidemic, *Journal of The Royal Society Interface*, 12(104), 20140,840, doi: 10.1098/rsif.2014.0840, 2015a.
- Mari, L., R. Casagrandi, M. Ciddio, S. H. Sokolow, G. De Leo, and M. Gatto, Uncovering the impact of human mobility on schistosomiasis via mobile phone data, pp. 71–97, Netmob Conference 2015, Data for Development Challenge Senegal, Massachusetts Institute of Technology, Cambridge, USA, 2015b.
- Martin, S., A. Costa, and W. Perea, Stockpiling oral cholera vaccine, *Bulletin of the World Health Organization*, 90(10), 714–714, doi: 10.2471/BLT.12.112433, 2012.
- Matsuda, E., S. Ishimura, Y. Wagatsuma, T. Higashi, T. Hayashi, A. S. G. Faruque, D. A. Sack, and M. Nishibuchi, Prediction of epidemic cholera due to *Vibrio cholerae* O1 in children younger than 10 years using climate data in Bangladesh, *Epidemiology and Infection*, 136, 73–79, 2008.
- Mccormack, W. M., M. S. Islam, and M. Fahimuddin, A Community Study of Inapparent Cholera Infections, *American Journal of Epidemiology*, 89(6), 658–664, 1969.
- Meloni, S., N. Perra, A. Arenas, S. Gómez, Y. Moreno, and A. Vespignani, Modeling human mobility responses to the large-scale spreading of infectious diseases, *Scientific Reports*, 1, 62, doi: 10.1038/srep00062, 2011.
- Memish, Z. A., et al., Mass gathering and globalization of respiratory pathogens during the 2013 Hajj, *Clinical Microbiology and Infection*, 21(6), 571.e1–571.e8, doi: 10.1016/j.cmi.2015.02.008, 2015.
- Merler, S., et al., Containing Ebola at the Source with Ring Vaccination, *PLOS Neglected Tropical Diseases*, 10(11), e0005,093, doi: 10.1371/journal.pntd.0005093, 2016.
- Merrell, D. S., S. M. Butler, F. Qadri, N. A. Dolganov, A. Alam, M. B. Cohen, S. B. Calderwood, G. K. Schoolnik, and A. Camilli, Host-induced epidemic spread of the cholera bacterium, *Nature*, 417(6889), 642–645, doi: 10.1038/nature00778, 2002.
- Ministère de la santé publique et de la population (MSPP), Rapport de cas, *Tech. Rep. 25 septembre 2016*, Port-au-Prince, Haiti, 2016.
- Mishra, A., N. Taneja, and M. Sharma, Environmental and epidemiological surveillance of *Vibrio cholerae* in a cholera-endemic region in India with freshwater environs, *Journal Of Applied Microbiology*, 112, 225–237, 2011.
- Mogasale, V., E. Ramani, H. Wee, and J. H. Kim, Oral Cholera Vaccination Delivery Cost in Low- and Middle-Income Countries: An Analysis Based on Systematic Review, *PLOS Neglected Tropical Diseases*, 10(12), e0005,124, doi: 10.1371/journal.pntd.0005124, 2016.

- Molla, A. M., S. M. Ahmed, and W. B. Greenough, Rice-based oral rehydration solution decreases the stool volume in acute diarrhoea., *Bulletin of the World Health Organization*, 63(4), 751–756, 1985.
- Montgomery, D. R., and W. E. Dietrich, Where do channels begin?, *Nature*, 336(6196), 232–234, doi: 10.1038/336232a0, 1988.
- Montgomery, D. R., and W. E. Dietrich, Channel Initiation and the Problem of Landscape Scale, *Science*, 255(5046), 826–830, doi: 10.2307/2876490, 1992.
- Morris, J. G., Deadly River: Cholera and Cover-Up in Post-Earthquake Haiti, *Emerging Infectious Diseases*, 22(11), 2029–2030, doi: 10.3201/eid2211.161215, 2016.
- Mukandavire, Z., S. Liao, J. Wang, H. Gaff, D. L. Smith, and J. G. Morris Jr, Estimating the reproductive numbers for the 2008–2009 cholera outbreaks in Zimbabwe, *Proceedings of the National Academy of Sciences of the United States of America*, 108, 8767–8772, 2011.
- Mukandavire, Z., D. L. Smith, and J. G. Morris Jr, Cholera in Haiti: Reproductive numbers and vaccination coverage estimates, *Scientific Reports*, 3, doi: 10.1038/srep00997, 2013.
- Mutreja, A., et al., Evidence for several waves of global transmission in the seventh cholera pandemic, *Nature*, 477(7365), 462–465, doi: 10.1038/nature10392, 2011.
- Nelder, J. A., and R. Mead, A simplex method for function minimization, *Computer Journal*, 7, 308–313, 1965.
- Nelson, E. J., J. B. Harris, J. Glenn Morris, S. B. Calderwood, and A. Camilli, Cholera transmission: The host, pathogen and bacteriophage dynamic, *Nature Reviews Microbiology*, 7(10), 693–702, doi: 10.1038/nrmicro2204, 2009.
- Nelson, E. J., D. S. Nelson, M. A. Salam, and D. A. Sack, Antibiotics for Both Moderate and Severe Cholera., *New England Journal of Medicine*, 364(1), 5–7, doi: 10.1056/NEJMp1013771, 2011.
- Ngwa, M. C., et al., Cholera in Cameroon, 2000-2012: Spatial and Temporal Analysis at the Operational (Health District) and Sub Climate Levels, *PLOS Neglected Tropical Diseases*, 10(11), e0005105, doi: 10.1371/journal.pntd.0005105, 2016.
- Oak Ridge National Laboratory, LandScan Haiti Population Data, 2011.
- Olago, D., M. Marchall, and S. O. Wandiga, Climatic, socio-economic, and health factors affecting human vulnerability to cholera in the Lake Victoria basin, East Africa, *Ambio*, 36, 350–358, 2007.
- Palchykov, V., M. Mitrović, H.-H. Jo, J. Saramäki, and R. K. Pan, Inferring human mobility using communication patterns, *Scientific Reports*, 4, 6174, doi: 10.1038/srep06174, 2014.
- Pan American Health Organization, *Haiti Cholera Outbreak Data.*, 2011.

- Parker, L., et al., Neighborhood-targeted and case-centered use of a single dose of oral cholera vaccine in an urban setting: Feasibility and vaccine coverage, in prep.
- Pascual, M., X. Rodó, S. P. Ellner, R. R. Colwell, and M. J. Bouma, Cholera dynamics and El Niño Southern Oscillation, *Science*, 289, 1766–1769, 2000.
- Pascual, M., M. J. Bouma, and A. P. Dobson, Cholera and climate: revisiting the quantitative evidence, *Microbes and Infection*, 4(2), 237–245, doi: 10.1016/S1286-4579(01)01533-7, 2002.
- Pascual, M., L. F. Chaves, B. Cash, X. Rod, and M. Yunus, Predicting endemic cholera: the role of climate variability and disease dynamics, *Climate Research*, 36(2), 131–140, doi: 10.3354/cr00730, 2008.
- Pasetto, D., F. Finger, A. Rinaldo, and E. Bertuzzo, Real-time projections of cholera outbreaks through data assimilation and rainfall forecasting, *Advances in Water Resources*, doi: 10.1016/j.advwatres.2016.10.004, 2016.
- Perkins, T. A., et al., Theory and data for simulating fine-scale human movement in an urban environment, *Journal of the Royal Society Interface*, 11(99), 20140,642, doi: 10.1098/rsif.2014.0642, 2014.
- Piarroux, R., Understanding the Cholera Epidemic, Haiti, *Emerging Infectious Diseases*, 17(7), 1161–1168, doi: 10.3201/eid1707.110059, 2011.
- Piarroux, R., and D. Bompangue, Needs for an integrative approach of epidemics: the example of cholera, *Encyclopedia Of Infectious Diseases: Modern Methodologies*, John Wiley And Sons, Inc., 2007.
- Piarroux, R., D. Bompangue, P.-Y. Oger, A. Boinet, F. Haaser, and T. Vandeveldel, From research to field action: Example of the fight against cholera in the Democratic Republic of Congo, *Field Actions Science Reports. The journal of field actions*, (Vol. 2), 2009.
- Plisnier, P. D., S. Serneels, and E. F. Lambin, Impact of ENSO on East African ecosystems: a multivariate analysis based on climate and remote sensing data, *Global Ecology and Biogeography*, 9(6), 481–497, doi: 10.1046/j.1365-2699.2000.00208.x, 2000.
- Qadri, F., et al., Efficacy of a Single-Dose, Inactivated Oral Cholera Vaccine in Bangladesh, *New England Journal of Medicine*, 374(18), 1723–1732, doi: 10.1056/NEJMoa1510330, 2016.
- Rebaudet, S., B. Sudre, B. Faucher, and R. Piarroux, Cholera in coastal Africa: A systematic review of its heterogeneous environmental determinants, *Journal of Infectious Diseases*, 208(suppl 1), S98–S106, doi: 10.1093/infdis/jit202, 2013a.
- Rebaudet, S., B. Sudre, B. Faucher, and R. Piarroux, Environmental Determinants of Cholera Outbreaks in Inland Africa: A Systematic Review of Main Transmission Foci and Propagation Routes, *Journal of Infectious Diseases*, 208(suppl 1), S46–S54, doi: 10.1093/infdis/jit195, 2013b.



- Rebaudet, S., et al., The dry season in Haiti: A window of opportunity to eliminate cholera, *PLoS Currents*, doi: 10.1371/currents.outbreaks.2193a0ec4401d9526203af12e5024ddc, 2013c.
- Reiner, R. C., A. A. King, M. Emch, M. Yunus, A. S. G. Faruque, and M. Pascual, Highly localized sensitivity to climate forcing drives endemic cholera in a megacity, *Proceedings of the National Academy of Sciences of the United States of America*, 109(6), 2033–2036, doi: 10.1073/pnas.1108438109, 2012.
- Revez, L., E. Chapman, P. Ramon-Pardo, T. P. Koehlmoos, L. G. Cuervo, S. Aldighieri, and A. Chambliss, Chemoprophylaxis in Contacts of Patients with Cholera: Systematic Review and Meta-Analysis, *PLOS ONE*, 6(11), e27,060, doi: 10.1371/journal.pone.0027060, 2011.
- Reyburn, R., D. R. Kim, M. Emch, A. Khatib, L. von Seidlein, and M. Ali, Climate variability and the outbreaks of cholera in Zanzibar, East Africa: A time series analysis, *American Journal of Tropical Medicine and Hygiene*, 84, 862–869, 2011.
- Ries, A. A., et al., Cholera in Piura, Peru: A Modern Urban Epidemic, *Journal of Infectious Diseases*, 166(6), 1429–1433, doi: 10.1093/infdis/166.6.1429, 1992.
- Righetto, L., E. Bertuzzo, L. Mari, E. Schild, R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo, Rainfall mediations in the spreading of epidemic cholera, *Advances in Water Resources*, 60, 34–46, doi: 10.1016/j.advwatres.2013.07.006, 2013.
- 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(17), 6602–6607, doi: 10.1073/pnas.1203333109, 2012.
- Roberts, N. C., R. J. Siebeling, J. B. Kaper, and J. H. B. Bradford, Vibrios in the Louisiana Gulf Coast Environment, *Microbial Ecology*, 8(4), 299–312, 1982.
- Rodríguez-Iturbe, I., and A. Rinaldo, *Fractal River Basins: Chance and Self-Organization*, Cambridge University Press, 2001.
- Rodó, X., M. Pascual, G. Fuchs, and A. S. G. Faruque, Enso and cholera: A nonstationary link related to climate change?, *Proceedings of the National Academy of Sciences*, 99(20), 12,901–12,906, doi: 10.1073/pnas.182203999, 2002.
- Ruiz-Moreno, D., M. Pascual, M. Bouma, A. Dobson, and B. Cash, Cholera seasonality in Madras: dual role for rainfall in endemic and epidemic regions, *EcoHealth*, 4, 52–62, 2007.
- Sack, D. A., How many cholera deaths can be averted in Haiti?, *The Lancet*, 377(9773), 1214–1216, doi: 10.1016/S0140-6736(11)60356-5, 2011.
- Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, A dipole mode in the tropical indian ocean, *Nature*, 401(6751), 360–363, doi: 10.1038/43854, 1999.

- Salje, H., J. Lessler, K. K. Paul, A. S. Azman, M. W. Rahman, M. Rahman, D. Cummings, E. S. Gurley, and S. Cauchemez, How social structures, space, and behaviors shape the spread of infectious diseases using chikungunya as a case study, *Proceedings of the National Academy of Sciences of the United States of America*, p. 201611391, doi: 10.1073/pnas.1611391113, 2016a.
- Salje, H., et al., Revealing the microscale spatial signature of dengue transmission and immunity in an urban population, *Proceedings of the National Academy of Sciences of the United States of America*, 109(24), 9535–9538, doi: 10.1073/pnas.1120621109, 2012.
- Salje, H., et al., Reconstruction of 60 Years of Chikungunya Epidemiology in the Philippines Demonstrates Episodic and Focal Transmission, *Journal of Infectious Diseases*, 213(4), 604–610, doi: 10.1093/infdis/jiv470, 2016b.
- Santa-Olalla, P., et al., Implementation of an Alert and Response System in Haiti during the Early Stage of the Response to the Cholera Epidemic, *The American Journal of Tropical Medicine and Hygiene*, 89(4), 688–697, doi: 10.4269/ajtmh.13-0267, 2013.
- Sardar, T., S. Mukhopadhyay, A. R. Bhowmick, and J. Chattopadhyay, An optimal cost effectiveness study on Zimbabwe cholera seasonal data from 2008–2011, *PLoS ONE*, 8, e81,231, 2013.
- Simini, F., M. C. González, A. Maritan, and A.-L. Barabási, A universal model for mobility and migration patterns, *Nature*, 484(7392), 96–100, doi: 10.1038/nature10856, 2012.
- Snow, J., *On the Mode of Communication of Cholera*, John Churchill, 1855.
- Sorooshian, S., and J. A. Dracup, Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resources Research*, 16(2), 430–442, doi: 10.1029/WR016i002p00430, 1980.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde, Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society*, 64(4), 583–639, doi: 10.1111/1467-9868.00353, 2002.
- Stager, J. C., A. Ruzmaikin, D. Conway, P. Verburg, and P. J. Mason, Sunspots, El Niño, and the levels of Lake Victoria, East Africa, *Journal of Geophysical Research: Atmospheres*, 112(D15), D15,106, 2007.
- Sur, D., et al., Efficacy and safety of a modified killed-whole-cell oral cholera vaccine in India: An interim analysis of a cluster-randomised, double-blind, placebo-controlled trial, *The Lancet*, 374(9702), 1694–1702, doi: 10.1016/S0140-6736(09)61297-6, 2009.
- Sur, D., et al., Efficacy of a Low-Cost, Inactivated Whole-Cell Oral Cholera Vaccine: Results from 3 Years of Follow-Up of a Randomized, Controlled Trial, *PLoS Neglected Tropical Diseases*, 5(10), doi: 10.1371/journal.pntd.0001289, 2011.

- Swerdlow, D. L., et al., Waterborne transmission of epidemic cholera in Trujillo, Peru: Lessons for a continent at risk, *The Lancet*, 340(8810), 28–32, doi: 10.1016/0140-6736(92)92432-F, 1992.
- Tamplin, M. L., A. L. Gauzens, A. Huq, D. A. Sack, and R. R. Colwell, Attachment of *Vibrio cholerae* serogroup O1 to zooplankton and phytoplankton of Bangladesh waters., *Applied and Environmental Microbiology*, 56(6), 1977–1980, 1990.
- Tappero, J., and R. Tauxe, Lessons Learned during Public Health Response to Cholera Epidemic in Haiti and the Dominican Republic, *Emerging Infectious Diseases*, 17(11), doi: 10.3201/eid1711.110827, 2011.
- Tarboton, D. G., A new method for the determination of flow directions and upslope areas in grid digital elevation models, *Water Resources Research*, 33(2), 309–319, doi: 10.1029/96WR03137, 1997.
- Tatem, A. J., et al., Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning, *Malaria Journal*, 13(1), 52, doi: 10.1186/1475-2875-13-52, 2014.
- Thiery, W., A. Martynov, F. Darchambeau, J.-P. Descy, P.-D. Plisnier, L. Sushama, and N. P. M. van Lipzig, Understanding the performance of the FLake model over two African Great Lakes, *Geoscientific Model Development*, 7(1), 317–337, doi: 10.5194/gmd-7-317-2014, 2014a.
- Thiery, W., et al., LakeMIP Kivu: evaluating the representation of a large, deep tropical lake by a set of one-dimensional lake models, *Tellus A*, 66(0), doi: 10.3402/tellusa.v66.21390, 2014b.
- Tien, J. H., H. N. Poinar, D. N. Fisman, and D. J. D. Earn, Herald waves of cholera in nineteenth century London, *Journal of The Royal Society Interface*, 8(58), 756–760, doi: 10.1098/rsif.2010.0494, 2011.
- Tizzoni, M., P. Bajardi, A. Decuyper, G. Kon Kam King, C. M. Schneider, V. Blondel, Z. Smoreda, M. C. González, and V. Colizza, On the use of human mobility proxies for modeling epidemics, *PLoS Computational Biology*, 10(7), e1003716, doi: 10.1371/journal.pcbi.1003716, 2014.
- Tuite, A. R., J. Tien, M. Eisenberg, D. J. Earn, J. Ma, and D. N. Fisman, Cholera epidemic in Haiti, 2010: using a transmission model to explain spatial spread of disease and identify optimal control interventions, *Annals of Internal Medicine*, 154(9), 593–601, doi: 10.1059/0003-4819-154-9-201105030-00334, 2011.
- Valia, R., E. Taviani, M. Spagnoletti, D. Ceccarelli, P. Cappuccinelli, and M. M. Colombo, *Vibrio cholerae* O1 epidemic variants in Angola: A retrospective study between 1992 and 2006, *Aquatic Microbiology*, 4, 354, doi: 10.3389/fmicb.2013.00354, 2013.
- Van de Linde, P. A. M., and G. I. Forbes, Observations on the spread of cholera in Hong Kong, 1961-63, *Bulletin of the World Health Organization*, 32(4), 515, 1965.

- Vezzulli, L., C. Pruzzo, A. Huq, and R. R. Colwell, Environmental reservoirs of *Vibrio cholerae* and their role in cholera, *Environmental Microbiology Reports*, 2, 27–33, 2010.
- Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson, Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resources Research*, 44(12), doi: 10.1029/2007WR006720, 2008.
- Vrugt, J. A., C. ter Braak, C. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon, Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling, *International Journal of Nonlinear Sciences and Numerical Simulation*, 10(3), doi: 10.1515/IJNSNS.2009.10.3.273, 2009.
- Walton, D. A., and L. C. Ivers, Responding to Cholera in Post-Earthquake Haiti, *New England Journal of Medicine*, 364(1), 3–5, doi: 10.1056/NEJMp1012997, 2011.
- Wesolowski, A., N. Eagle, A. M. Noor, R. W. Snow, and C. O. Buckee, Heterogeneous mobile phone ownership and usage patterns in Kenya, *PLoS ONE*, 7(4), e35,319, doi: 10.1371/journal.pone.0035319, 2012a.
- Wesolowski, A., N. Eagle, A. J. Tatem, D. L. Smith, A. M. Noor, R. W. Snow, and C. O. Buckee, Quantifying the impact of human mobility on malaria, *Science*, 338(6104), 267–270, doi: 10.1126/science.1223467, 2012b.
- Wesolowski, A., C. O. Buckee, D. K. Pindolia, N. Eagle, D. L. Smith, A. J. Garcia, and A. J. Tatem, The use of census migration data to approximate human movement patterns across temporal scales, *PLoS ONE*, 8(1), e52,971, doi: 10.1371/journal.pone.0052971, 2013a.
- Wesolowski, A., N. Eagle, A. M. Noor, R. W. Snow, and C. O. Buckee, The impact of biases in mobile phone ownership on estimates of human mobility, *Journal of the Royal Society Interface*, 10(81), 20120,986, doi: 10.1098/rsif.2012.0986, 2013b.
- Wesolowski, A., C. O. Buckee, L. Bengtsson, E. Wetter, X. Lu, and A. J. Tatem, Commentary: Containing the ebola outbreak - the potential and challenge of mobile network data, *PLoS Currents*, 1, doi: 10.1371/currents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e, 2014a.
- Wesolowski, A., C. J. E. Metcalf, N. Eagle, J. Kombich, B. T. Grenfell, O. N. Bjørnstad, J. Lessler, A. J. Tatem, and C. O. Buckee, Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data, *Proceedings of the National Academy of Sciences of the United States of America*, 112(35), 11,114–11,119, doi: 10.1073/pnas.1423542112, 2015.
- Wesolowski, A., et al., Quantifying travel behavior for infectious disease research: A comparison of data from surveys and mobile phones, *Scientific Reports*, 4, 5678, doi: 10.1038/srep05678, 2014b.
- Wharton, M., et al., A large outbreak of antibiotic-resistant shigellosis at a mass gathering, *Journal of Infectious Diseases*, 162(6), 1324–1328, doi: 10.1093/infdis/162.6.1324, 1990.

World Health Organization, Cholera country profile: Senegal, *Tech. rep.*, 2008.

World Health Organization, Cholera, 2011, *Weekly epidemiological record*, 87, 289–304, 2012.

World Health Organization, Weekly Epidemiological Record, vol. 91, 38, *Weekly Epidemiological Record*, 91(38), 432–440, 2016a.

World Health Organization, WHO | Cholera vaccination campaign for Haitians hardest hit by Hurricane Matthew, <http://www.who.int/features/2016/haiti-cholera-vaccination/en/>, 2016b.

You, Y. A., et al., Risk Map of Cholera Infection for Vaccine Deployment: The Eastern Kolkata Case, *PLOS ONE*, 8(8), e71,173, doi: 10.1371/journal.pone.0071173, 2013.



Avenue de l'Université 20  
1005 Lausanne  
Switzerland  
flavio.finger@epfl.ch  
people.epfl.ch/flavio.finger  
researchgate.net/profile/Flavio\_Finger

# Flavio Finger

Nationality: Swiss  
Date of birth: 17 August 1987

## Research interests

Flavio's research focuses on spatially explicit modeling of epidemics, mainly applied to waterborne diseases in developing and emerging countries. He is particularly interested in the links between environmental drivers and disease outbreaks, and in the processes that lead to disease spread over short to long distance ranges, such as human mobility and transport in hydrologic networks. His research aims to reveal the underlying mechanisms of disease spread, but also to provide tools to inform decision makers and public health officials. Other research interests and fields of expertise range from topics such as water and sanitation in developing countries over hydrology to model development and parameter inference.

## Education

- 2013 – 2017 **PhD candidate**, *Laboratory of Ecohydrology, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland*, Supervisors: Prof. Andrea Rinaldo and Dr. Enrico Bertuzzo, Title: Spatially explicit modeling of cholera epidemics.
- 2010–2012 **MSc in Environmental Science and Engineering**, *EPFL, Switzerland*, Specialization in Natural Water, Soil and Ecosystems Engineering.
- 2009–2010 **Academic exchange**, *Indian Institute of Technology Delhi, India*.
- 2007–2010 **BSc in Environmental Science and Engineering**, *EPFL, Switzerland*.

## Experience

- 2012 and 2014 **Stiftung Umwelteinsatz Schweiz**, *Switzerland*, Several months of community service in lieu of compulsory military service: construction of dry stone walls.
- 2012 **e-dric.ch ingénieurs conseils**, *Le Mont-sur-Lausanne, Switzerland*, Master thesis: Hydrological modeling of the Aubonne karst watershed. Development and calibration of a mechanistic rainfall-runoff model.
- 2011 **urbaMond**, *Dakar, Senegal*, Internship: Management of water pollution and inundations in suburban Dakar. Water sampling and analysis, comparison of different sanitation and remediation scenarios.
- 2010 **Department of Nature and Environment of the Canton of Graubünden**, *Switzerland*, Several months of community service in lieu of compulsory military service: management and control of invasive plant species.

## Languages

- German **mother tongue**
- English **fluent written and spoken**, Cambridge Certificate in Advanced English
- French **fluent written and spoken**
- Italian **beginner**

---

## Teaching

- 2013 – 2016 **Hydrology for engineers**, Main teaching assistant. Preparation of exercises, teaching of exercise sessions, course organization, preparation and correction of exams.
- 2011 **Quantitative methods I**, Teaching assistant.
- 2011 **Calculus I**, Teaching assistant.

---

## Computer skills

- Programming Python, Matlab, R, SQL, L<sup>A</sup>T<sub>E</sub>X, Bash, Git
- Other Linux, QGIS, ArcGIS, Office, Graphics

---

## Interests

- Sports skiing, ski mountaineering, rock climbing, hiking
- Traveling discovering cultures, landscapes and nature
- Campus life member of several students associations
- Other beekeeping

---

## Publications

### Journal articles

E. Bertuzzo, F. Finger, L. Mari, M. Gatto, and A. Rinaldo, “On the probability of extinction of the Haiti cholera epidemic,” *Stochastic Environmental Research and Risk Assessment*, pp. 1–13, 2014.

F. Finger, T. Genolet, L. Mari, G. C. de Magny, N. M. Manga, A. Rinaldo, and E. Bertuzzo, “Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 23, pp. 6421–6426, 2016.

F. Finger, A. Knox, E. Bertuzzo, L. Mari, D. Bompangue, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo, “Cholera in the Lake Kivu region (DRC): Integrating remote sensing and spatially explicit epidemiological modeling,” *Water Resources Research*, vol. 50, no. 7, pp. 5624–5637, 2014.

J. Kühn, F. Finger, E. Bertuzzo, S. Borgeaud, M. Gatto, A. Rinaldo, and M. Blokesch, “Glucose- but Not Rice-Based Oral Rehydration Therapy Enhances the Production of Virulence Determinants in the Human Pathogen *Vibrio cholerae*,” *PLoS Neglected Tropical Diseases*, vol. 8, no. 12, p. e3347, 2014.

L. Mari, E. Bertuzzo, F. Finger, R. Casagrandi, M. Gatto, and A. Rinaldo, “On the predictive ability of mechanistic models for the Haitian cholera epidemic,” *Journal of the Royal Society, Interface*, vol. 12, no. 104, 2015.

D. Pasetto, F. Finger, A. Rinaldo, and E. Bertuzzo, “Real-time projections of cholera outbreaks through data assimilation and rainfall forecasting,” *Advances in Water Resources*, 2016.

### Conferences

F. Finger, “Modeling the spread of cholera using human mobility estimates derived from mobile phone records,” in *Swiss Meeting for Infectious Disease Dynamics, Geneva, Switzerland*, 2016.

F. Finger, T. Genolet, L. Mari, G. C. de Magny, A. Rinaldo, and E. Bertuzzo, “Modeling the spread of cholera using human mobility estimates derived from mobile phone records,” in *Epidemics V, Clearwater Beach, Florida*, 2015.



F. Finger, “Modèles spatiaux de la propagation du choléra en utilisant l’approche SIRB,” in *Méthodes Quantitatives et du Traitement de l’Information Médicale (QuantIM)*, Webinar, 2015.

F. Finger, A. Knox, E. Bertuzzo, L. Mari, D. Bompangue, M. Gatto, and A. Rinaldo, “Integrating remote sensing and spatially explicit epidemiological modeling,” in *EGU General Assembly, Vienna, Austria*, 2015.

F. Finger, B. Schaeffli, E. Bertuzzo, L. Mari, and A. Rinaldo, “Parameter and uncertainty estimation for mechanistic, spatially explicit epidemiological models,” in *EGU General Assembly, Vienna, Austria*, 2014.

F. Finger, E. Bertuzzo, L. Mari, A. C. Knox, M. Gatto, and A. Rinaldo, “Spatially explicit modelling of cholera epidemics,” in *AGU General Assembly, San Francisco*, 2013.