Spatial Dynamics of the

Severe Acute Respiratory Syndrome (SARS) Epidemic

in Hong Kong in 2003

Wai Yee Lam

Thesis submitted for the Degree of Doctor of Philosophy

Imperial College London

Department of Infectious Disease Epidemiology
School of Public Health

2011

Declaration of Originality

I declare that this thesis is my own work, and I have acknowledged all results and quotations from the published or unpublished work of other people.

Abstract

The Severe Acute Respiratory Syndrome (SARS) epidemic in 2003 was the first infectious disease outbreak caused by a novel pathogen in the twenty-first century. The outbreak in Hong Kong was the second largest worldwide and was characterised by a large proportion of hospital infections and a super-spreading event caused by environmental factors in residential buildings.

Hospitals treating SARS cases were at high risk for transmission. I found that hospital outbreaks triggered community transmission as well as the formation of spatial clusters of community cases. The size of the community outbreak in an area increased with the size of the outbreak in the nearest hospital treating SARS, and an area was more likely to have no community-infected cases if it was far from hospitals treating SARS, or had less hospital-infected cases within the area.

To quantify the transmission between hospital and community, I developed a spatial epidemic-tree-reconstruction method that uses gravity models to spatially define the probability of contact between individuals in the community. From the reconstructed probabilistic infection tree, I estimated that 24% of community transmission was likely to be infected by cases infected in hospitals, with infected patients discharged during their incubation period and hospital visitors the most important drivers of transmission from healthcare settings to the community. Healthcare workers were key drivers of hospital transmission, with the hospital-to-hospital reproduction number, excluding a single hospital super-spreading event, estimated to be 0.8. A typical community-acquired case was estimated to generate 0.6 cases in the community and 0.2 cases in the hospital in which they were subsequently hospitalised.

My findings suggest that hospital infection control could be improved. Restricted hospital visitor policies could have been imposed for longer time during the outbreak and quarantine could be considered for those who recently visited or have been discharged from hospitals treating SARS cases.

Table of Contents

Declaration of Originality 2

Abstract 3

List of Tables 8

List of Figures 9

List of Abbreviations 16

Acknowledgement 18

Chapter 1. Background 19

1.1 Introduction 19

1.2 Literature Review 20

1.2.1 SARS 20

1.2.2 Spatial dynamics of emerging respiratory infectious disease epidemics 27

1.2.3 Mechanistic spatiotemporal models 32

1.3 Study objectives 35

1.4 Outline of the remaining chapters 36

Chapter 2. Data Validation and Summary 38

2.1 Outbreak data 38

2.1.1 Data collection 38

2.1.2 Data validation and reclassification 39

2.1.3 Potential source of infection summary 40

2.2 Geographic Information System (GIS) and demographic data 44

2.3 Hospitalisation data 45

Chapter 3 - Spatial Distribution and Cluster Analysis 47

3.1 Objectives 47
3.2 Method 47
3.2.1 Exploratory analysis 47
3.2.2 Interactions between hospital and community transmission 49
3.2.3 Real-time cluster detection 51
3.3 Results 52
3.3.1 Overall epidemic 52
3.3.2 Hospital transmission 63
3.3.3 Community transmission 72
3.3.4 Interactions between hospital and community transmission 76
3.3.5 Real-time cluster detection 78
3.4 Discussion 84
Chapter 4 - Studying Human Movement Patterns Using Gravity Models 87
4.1 Gravity models 87
4.1.1 Origin 87
4.2 Alternative methods to capture spatial heterogeneity in contact rates 94
4.3 Study objectives 95
4.4 Data 96
4.4.1 Census travel data 96
4.4.2 Travel survey data 97
4.4.3 Adjusted overall census data 98
4.4.4 Home-to-hospital travel data 98
4.4.5 Geo-demographic characteristics 98

4.4.6 Data issues 99
4.5 Preliminary Analysis 100
4.6 Methodology 103
4.6.1 Model formulation - Hong Kong travel patterns 103
4.6.2 Model formulation- Home-to-hospital patterns 107
4.6.3 Model fitting 107
4.6.4 Goodness-of-fit evaluation and model comparison 109
4.7 Results 110
4.7.1 Hong Kong travel patterns 110
4.7.2 Hospital flow patterns 123
4.8 Discussion 125
Chapter 5 - The Spatial Wallinga and Teunis Method 128
5.1 The Wallinga and Teunis method 128
5.1.1 Formulation 128
5.1.2 Alternative approaches 130
5.1.3 Capturing individual heterogeneity 131
5.2 Study objectives 133
5.3 Fitting serial interval data 134
5.3.1 Method 134
5.3.2 Results 136
5.3.3 Discussion 139
5.4 Partial extension of Wallinga and Tenuis's method 141
5.4.1 Method 141

5.4.2 Spatial mixing assumptions 142

5.4.3 Time course of infection assumptions 150

5.4.4 Analysis plan 150

5.4.5 Results 151

5.4.6 Discussion 167

Chapter 6 - Quantifying Hospital and Community Transmissions 170

6.1 Current findings 170

6.2 Study objectives 171

6.3 Method 172

6.3.1 Transmission pathways 172

6.3.2 Mathematical formulation 173

6.4 Results 179

6.5 Discussion 203

Chapter 7 - Discussion 207

7.1 Dynamics of spatial spread and transmission risk factors 208

7.2 Implications for control measures 210

7.3 Spatially explicit infection tree methods 212

7.4 Beyond infection trees - modelling the susceptible population 214

References 217

List of Tables

Table 3.1 Composition of the source of infection (in percentage of total number of cases in the cluster) of the spatiotemporal clusters of all cases (excluding ongoing inpatients) 61

Table 3.2 Percentage of cases captured in clusters by source of infection. 62

Table 3.3 Composition (% total) of the source of infection of the community spatiotemporal clusters. 75

Table 3.4 Comparison of the results of quasi-poisson, negative binomial, zero-inflated poisson / negative binomial, hurdle poisson / negative binomial models on the community case count in TPUs 79

Table 3.5 Comparison of the timeline the Amoy TPU (291) was detected by the prospective and retrospective scan statistics, and the time health officials were alert to and confirmed the outbreak. 80

Table 3.6 Comparison of the prospective and retrospective clusters 83

Table 4.1 Summary of the gravity models fitted to the census and survey data. 106

Table 4.2 Gravity models fitted to the hospital flow data. 107

Table 4.3 Outflow constrained gravity model parameter estimates and the goodness-of-fit statistics. 111

Table 4.4 Parameter estimates and goodness-of-fit statistics of the outflow constrained district-adjusted population gravity models fitted to the survey work and study flow data. 114

Table 4.5 Inflow constrained gravity model parameter estimates and goodness-of-fit statistics. 116

Table 4.6 Unconstrained gravity model parameter estimates and goodness-of-fit statistics 118

Table 4.7 Categorisation of each flow type by the comparison of scaling exponent estimates 119

Table 4.8 Parameter estimates and goodness-of-fit statistics of the gravity models on the healthcare worker, inpatient, visitor and overall flows, using hospital size or number of hospital infection as the attribution factor, and the power-law or exponential distribution as the spatial kernel 124

Table 5.1 Parameter estimates of the infectiousness distribution and the goodness-of-fit of the convoluted serial interval under different incubation period assumptions. 137

Table 5.2 Variants of the partially extended Wallinga and Teunis method studied. 151

Table 6.1 Parameter assumptions applied in the fully extended spatial method. 176

Table 6.2 Summary of the high R_{xy} cases (>3) identified, and details of the matching probable secondary clusters based on the epidemiological linkage data 190

Table 6.3 Index cases of disease clusters with size >1 identifiable from the epidemiological linkage data, and their R estimates by transmission settings from the fully extended method 191

Table 6.4 Results of sensitivity analysis on different assumptions of the parameters and movement mixing. 203

List of Figures

Figure 1.1 Hong Kong in 18 district council districts and the location of the SARS treating hospitals, Amoy Gardens Block E, and Metropole Hotel. 23

Figure 1.2 Population density by tertiary planning units (TPUs) of Hong Kong. 24

Figure 1.3 Epidemic curve (daily incidence) of the SARS outbreak in Hong Kong 25

Figure 2.1 Breakdown (in percentage of total number of hospital cases) of hospital infected cases by hospitals. 41

Figure 2.2 Breakdown (in percentage of total number of community cases excluding Amoy Garden cases) of the potential source of infection of the community infected cases. 41

Figure 2.3 Epidemic curves (daily incidence) of the hospital, community and Amoy Gardens outbreaks. 42

Figure 2.4 Epidemic curves (daily number of cases) of the 3 major hospital outbreaks (PWH, AHN, PMH). 42

Figure 2.5 Composition of hospital infection by infection subgroup. 42

Figure 2.6 Composition of the hospital infection throughout the epidemic. 43

Figure 2.7 Community epidemic curves excluding Amoy Gardens cases 43

Figure 2.8 Composition of the community infection excluding Amoy Gardens cases throughout the epidemic. 44

Figure 3.1 Spatial distribution of SARS cases in Hong Kong (TPU incidence rate per 100,000 population). 53

Figure 3.2 TPU spread by incidence (both in cumulative percentages) for: (i) all cases (whose residential TPUs could be identified), (ii) excluding the Amoy Gardens cases. 54

Figure 3.3 Infected TPU and incidence by source of infection (both in cumulative percentages) by symptom onset day 55

Figure 3.4 Boxplot of the (a) population density, (b) population size of the new TPUs infected by stages 55

Figure 3.5 Scatter plot of the population size and population density by stage of infection. 55

Figure 3.6 (a)-(p): Weekly incidence rate (per 100,000 population) by TPUs (from left top to left bottom, then right top to right bottom). 57

Figure 3.7 Spatiotemporal clusters of all cases excluding ongoing inpatients numbered in descending order of statistical significance (all with p-values < 0.05). 59

Figure 3.8 Composition of the clusters (hospital or community infections). 60

Figure 3.9 Start and end days and the duration (figure in the bar) of the spatiotemporal clusters of all cases (excluding ongoing inpatients) 61

Figure 3.10 Start and end days and the duration (figure in the bar) of the spatiotemporal clusters of hospital cases (excluding ongoing inpatients). 63

Figure 3.11 Spatial distribution of hospital infected SARS cases in Hong Kong (TPU incidence rate per 100,000 population). 64

Figure 3.12 Hospital spatiotemporal clusters (excluding ongoing inpatients) numbered in descending order of statistical significance (all with p-values <0.05). 65

Figure 3.13 Spatial distribution of infected healthcare workers (incidence rate per 100,000 population). . 66

Figure 3.14 Spatial distribution of infected inpatients (incidence rate per 100,000 population). 67

Figure 3.15 Spatial distribution of infected hospital visitors (incidence rate per 100,000 population). 68

Figure 3.16 Spatial distribution of the PWH infected cases (incidence rate per 100,000 population). 69

Figure 3.17 Spatial distribution of the AHN infected cases (incidence rate per 100,000 population). 70

Figure 3.18 Spatial distribution of the PMH infected cases (incidence rate per 100,000 population). 71

Figure 3.19 Distribution of community infected SARS cases in Hong Kong (TPU incidence rate per 100,000 population). 73

Figure 3.20 Community spatiotemporal clusters numbered in descending order of statistical significance (all with p-values <0.05). 74

Figure 3.21 Start and end days and the duration (figure in the bar) of the spatiotemporal clusters of community cases. 75

Figure 3.22 Comparison of hospital and community spatiotemporal clusters. Labels in red: cluster number of hospital clusters, labels in green: cluster number of community clusters. 77

Figure 3.23 Common clusters started as hospital only clusters (TPU 759 only for cluster H9C7) before they became common clusters. Cluster names starting with “H” are hospital clusters, and starting with “C” are community clusters, and the two common clusters are named as “H_C_”. 78

Figure 3.24 Prospective community TPUs detected as clusters compared to TPUs detected retrospectively. 81

Figure 3.25 Start and end days and the duration (figure in the bar) of the prospective spatiotemporal clusters of community cases. Some of the clusters were detected to be significant in multiple periods. TPU marked with * were not detected in the retrospective scan. 82

Figure 4.1 Spatial kernels for gravity models from previous epidemic studies reproduced, with the country and type of flow and the first author of the studies specified. 93

Figure 4.2 Distribution of between-district distance of Hong Kong based on the District Council District division. Measures were grouped in the hierarchy of: (1) Islands: if one of the district measured is the

Islands district; (2) NT (the New Territories): if at least one of the district measured is located in the New Territories; (3) Others: (Hong Kong Island and Kowloon): if the distance measured does not belong to (1) or (2). 99

Figure 4.3 The scatter plots of census and survey flow of the same trip type by outflow and inflow. 101

Figure 4.4 Census outflow of different trip types by population size of the district. 101

Figure 4.5 Left: Inflow by population size of the districts; Right: Inflow to outflow ratio by percentage of within-district flow, by census flow type: (a) work flow, (b) study flow, (c) work-study flow, and (d) adjusted overall flow. 102

Figure 4.6 The outflow (left) and inflow (right) by population size of the home-based others, non-home based and overall trips in the survey data 104

Figure 4.7 (Left) Survey overall outflow by census work outflow; (Right) Survey overall inflow by census work inflow. The straight lines represent the least square regression of the survey overall flow=constant x census flow. 104

Figure 4.8 Comparison of the gravity models using power-law or exponential kernel in estimating the probability distribution of journey distance 110

Figure 4.9 Deviance residual plots of the outflow constrained gravity models by district (left) and of individual observations (right) of the inflow and district-adjusted population models on census work-study flow. 112

Figure 4.10 The normalised power law spatial kernel (in log-log scale) of the inflow and district-adjusted population gravity models on the 5 flow types respectively. 113

Figure 4.11 Spatial kernels derived from the survey work flow and study flow gravity model compared to that of the census work flow and study flow model. 114

Figure 4.12 Spatial kernels derived from the inflow constrained, district-adjusted population size gravity models of different flow types. 115

Figure 4.13 Deviance residual plots of the unconstrained gravity models by district (left) and of individual observations (right) of the marginal flow and district-adjusted population models on census workflow. 117

Figure 4.14 Comparison of the effects of scaling exponents of the high inflow and ordinary districts on the multiplicative flow component (=population$^{\text{scaling exponent}}$ x normalisation constant$^{0.5}$), based on the unconstrained gravity models on census work flow and adjusted census overall flow, assuming the distance effect=1. 119

Figure 4.15 The normalised power law spatial kernel of marginal flow and district-adjusted population gravity models on the five flow types respectively. 121

Figure 4.16 Comparison of the spatial kernel of each flow type derived by inflow constrained, outflow constrained and unconstrained models. 121

Figure 4.17 Comparison of the estimated journal distance distribution under the inflow constrained, outflow constrained and unconstrained models for each flow type. 122

Figure 4.18 Comparison of the observed and estimated probability distribution of journey distance by gravity models using the power law or exponential spatial kernel for the: (a) healthcare worker flow, (b) inpatient flow, (c) visitor flow, and (d) overall flow, using hospital size as the attraction factor. 124

Figure 4.19 Power law spatial kernels derived from the healthcare worker, inpatient, visitor and overall hospital flow gravity models, compared to the census work and work-study spatial kernels from outflow constrained gravity models. 125

Figure 5.1 Timeline of transmission between infector *i* and infectee *j*. S_i and s_j are the time of symptom onset, t_i and t_j are the time of infection, g_{ij} is the generation interval and s_{ij} is the serial interval of the transmission between cases *i* and *j*. 134

Figure 5.2 (a) The convoluted serial interval density using Weibull incubation and infectiousness distribution, compared to the observed and Weibull fitted serial interval distribution. Comparison of: (b) the derived infectiousness density, and (c) the convoluted serial interval density, using Weibull distributed or constant incubation of 2, 3 or 4 days respectively. (d) The number of potential infectors by symptom onset date based on different assumptions to identify potential infectors. 138

Figure 5.3 Schematic presentation of the relationship between potential infectors *i, m, k* to infectee *j* under: (a) Wallinga and Teunis method; (2) Infection driven movement; (c) susceptible driven movement; (d) trip matrix, with $O_{i'}$ and $D_{i'}$ denote the outflow and inflow of district *i'*. 144

Figure 5.4 Schematic presentation of the (a) Wallinga and Teunis method and (b) bilateral movement assumption, with *i* as the infector and *j* as the infectee and in the same settings described in Chapter 6.4.2.1. (c) Relevant components in the trip matrix under the bilateral movement assumption. The colour marks the corresponding potential infector. 147

Figure 5.5 Schematic presentation of the (a) Wallinga and Teunis method and (b) multilateral movement assumption, with *i* as the infector and *j* as the infectee and in the same settings described in Chapter 5.4.2.1. (c) Relevant components in the trip matrix under the multilateral movement assumption. The colour marks the corresponding potential infector. 149

Figure 5.6 Estimated reproduction number *R(t)* using the non-spatial method with the deconvolution of serial interval density of: (i) 2 days incubation and Weibull infectiousness of 18 days, (ii) 4 days incubation and gamma infectiousness of 16 days, compared to the original Wallinga and Teunis method. 152

Figure 5.7 (a) Comparison of *R(i)* and *R(t)* estimates under the fixed and variable infectious period assumption, using the Wallinga and Teunis method with the serial interval distribution. 153

Figure 5.8 The estimated daily and individual reproduction numbers *R(t)* and *R(i)* under different movement assumptions, using the partially extended method with 2-day convolution and variable infectious period and on the adjusted census overall trip patterns. High *R(i)* values are highlighted in green, and the *R(i)* of the index case of the PWH and Amoy Gardens outbreak are identified. 153

Figure 5.9 Box-plots on the difference of R(t) [model of other spatial mixing assumptions - infector driven model] by 3 stages of the epidemic 156

Figure 5.10 The R(t) time series of the inflow constrained model (a) compared to that of the bilateral and distance only model; (b) compared to that of the non-spatial model and the original Wallinga and Teunis method; (c) plotted along the case incidence of the outbreak by source of infection. 157

Figure 5.11 The probability of contact with a specific individual by the district population under the susceptible and infector driven model, with those of the high inflow districts were labelled. 158

Figure 5.12 Percentage of within district transmission in the R(t) of the inflow constrained, bilateral and non-spatial model, and in the number of potential secondary cases estimated by the modified method. 158

Figure 5.13 Difference in estimated district-specific weekly average reproduction number R(w) [infector-driven model - bilateral models]. Blue box: high inflow districts; red box: districts with highest positive *R(w)* difference. Amoy Gardens is located in Kwun Tong district 159

Figure 5.14 The daily reproduction number of inflow constrained model based on gravity models of different travel trip patterns. 160

Figure 5.15 Percentage of within district transmission in *R(t)* and in the number of potential secondary cases under different trip patterns used in the gravity model for the inflow constrained method. 160

Figure 5.16 The estimated weekly average reproduction number by the partially modified model, and the weekly incidence by source of infection. 161

Figure 5.17 The spatial decomposition of (a) The overall case incidence rate, (b) average *R(i)*, and (c) percentage of within district transmission by reproduction number by district council district. 162

Figure 5.18 The spatiotemporal weekly reproduction number *R(w)* by district, and the weekly incidence of the hospital and Amoy outbreaks (from the source of infection data). 163

Figure 5.19 The spatiotemporal within district transmission percentage in *R(w)* by district, and the weekly incidence of the hospital and Amoy outbreaks (from the source of infection data). 164

Figure 5.20 Comparison of the percentage of within district transmission in the reproduction number under the assumptions of 2-day, 4-day and Weibull convolution. 165

Figure 5.21 The difference in daily and individual reproduction nubmers *R(t)* and *R(i)* using the 2-day convolution as baseline compared to: (a) the 4-day convolution model; (b) the Weibull serial interval distribution model. 166

Figure 5.22 Results of the partially modified method using the centroid of the residing district of the cases: (a) the daily and individual reproduction number estimates, (b) proportion of within district transmission. 166

Figure 6.1 Places of exposure of infectious cases by their source of infection and time course of infection. 172

Figure 6.2 Estimates of the daily hospital and community reproduction numbers of the fully extended method. 180

Figure 6.3 Comparison of the overall reproduction number between the fully extended and partially extended method. 180

Figure 6.4 Composition of the reproduction number of the epidemic of: (a) all cases; (b) all cases excluding the PWH and Amoy clusters. 181

Figure 6.5 (a) Weekly reproduction number estimates by transmission settings, and the key infection control measures implemented; (b) composition of weekly reproduction number by transmission setting. 182

Figure 6.6 The overall composition of: (a) hospital transmission and (b) community transmission, by the source of infection of the infector. (c) and (d): same as (a) and (b) but on a weekly basis. 183

Figure 6.7 The average reproduction number of hospital and community transmission by the subgroup of infected cases. 184

Figure 6.8 Weekly reproduction number of hospital and community transmission by infection subgroup. 185

Figure 6.9 (a)-(d): Box plots of individual reproduction number by transmission settings and week of symptom onset. Any case with R_{xy}>3 was marked with a case number (in blanket), and the corresponding probable cluster of secondary infection listed in Table 6.2 (indexed by the case number) 187

Figure 6.10 Percentage of within district transmission in all hospital transmission, community transmission excluding Amoy cases, and community transmission of the Amoy cases. 192

Figure 6.11 Percentage of within-district transmission in the reproduction number by transmission settings. 192

Figure 6.12 Comparison of the within-district transmission by type of infected subjects. 193

Figure 6.13 Overall percentage of within-district transmission by district. 193

Figure 6.14 Comparison of the proportion of within-district transmission over time between the full and partially extended method. 194

Figure 6.15 Disease spread patterns shown as the transmission route of the first case of each district. .. 195

Figure 6.16 Population size of the district (blue gradient colour) and the sequence of district being infected 195

Figure 6.17 Overall reproduction number by district. 196
Figure 6.18 Reproduction number by transmission setting by district. 197
Figure 6.19 Composition of the district reproduction number. 197
Figure 6.20 Spatiotemporal trend of the overall reproduction number by district by week. 198
Figure 6.21 Spatiotemporal trend of the hospital-to-hospital reproduction number R_{hh} by district by week. 199
Figure 6.22 Spatiotemporal trend of the hospital-to-community reproduction number R_{hc} by district by week. 200
Figure 6.23 Spatiotemporal trend of the community-to-community reproduction number R_{cc} by district by week. 201
Figure 6.24 Spatiotemporal trend of the community-to-hospital reproduction number R_{ch} by district by week. 202

List of Abbreviations

AIC	Akaike's information criterion
BIC	Bayesian information criterion
DCD	District council district
FMD	Foot and Mouth Disease
HCW	Healthcare worker
HNB	Hurdle negative binomial
HP	Hurdle Poisson
IP	Inpatient
IP_R	Inpatient discharged during incubation period of SARS
MCMC	Markov Chain Monte Carlo
RJMCMC	Reversible Jump Markov Chain Monte Carlo
R_0	Basic reproduction number
SARS	Severe Acute Respiratory Syndrome
TPU	Tertiary planning unit
WHO	World Health Organisation
ZINB	Zero-inflated negative binomial
ZIP	Zero-inflated Poisson

Hospital names

AHN	Alice Ho Miu Ling Nethersole Hospital
BTH	Baptist Hospital
CMC	Caritas Medical Centre
KWH	Kwong Wah Hospital
NDH	North District Hospital
PMH	Princess Margaret Hospital
PWH	Prince of Wales Hospital
PYN	Pamela Youde Nethersole Eastern Hospital
QEH	Queen Elizabeth Hospital
QMH	Queen Mary Hospital
SH	Shatin Hospital
SPH	St. Paul Hospital
STH	St. Teresa's Hospital
TKO	Tseung Kwan O Hospital
TMH	Tuen Mun Hospital
TPH	Tai Po Hospital
TWA	Tusen Wan Adventist Hospital
UCH	United Christian Hospital
UH	Union Hospital
WTS	Wong Tai Sin Hospital
YCH	Yan Chai Hospital

Acknowledgements

I am extremely grateful to my supervisors Neil Ferguson and Azra Ghani, for their commitment to supervise me, their constructive advice and guidance on my work, and their support and encouragement. I would also like to thank Christl Donnelly for her supervision in the early stage of the study. I am also very grateful to the support and understanding of my family in Hong Kong.

I would also like to thank the School of Public Health, the University of Hong Kong for the data I used in this study, and the Dorothy Hodgkin Postgraduate Award and co-sponsors Medical Research Council and Hutchison Whampoa for providing the funding for my study.

Chapter 1. Background

1.1 Introduction

The Severe Acute Respiratory Syndrome (SARS) epidemic in 2003 was the first infectious disease outbreak caused by a novel pathogen in the twenty-first century, and was not the last. In eight months, it spread to 26 countries, causing 8,096 infections and 774 deaths. Hong Kong was the second region to be affected by SARS after China, and was the second most affected country worldwide, with 1,755 cases and 299 deaths (World Health Organisation (WHO). ; Leung, Hedley et al. 2004). Tackling epidemics caused by new infectious agents is complicated by limited knowledge of the nature of the pathogen, its transmission dynamics and the resulting implications for disease control strategies. Also the lack of a vaccine and limited treatment options restricts control to the use of infection control and public health interventions that reduce contact rates in the population - for instance, social distance measures.

Understanding the spatial dynamics of epidemics is vital because of the localized nature of disease transmission and interactions, and because targeting limited control resources where they will have maximum impact is vital to curtail a fast-moving epidemic. Three aspects are important. Firstly, the spatial distribution of cases captures important facets of the nature of disease transmission and helps to reveal the transmission routes of the emerging epidemic, as in the early example of John Snow in the cholera outbreak investigation (Snow 1854). Secondly, prompt identification of spatial disease clusters is crucial in maximizing the effectiveness of quarantine and other control measures to reduce the morbidity of the epidemic, as in the evacuation of the Amoy Gardens Block E residents during the SARS epidemic in Hong Kong (Severe Acute Respiratory Syndrome Expert Committee 2003). Thirdly, the spatial dynamics of epidemics inform the construction of realistic mathematical models that capture the effects of spatial heterogeneity on the dynamics and control of the epidemic and which can be used to monitor and assess different control options (Ferguson, Donnelly et al. 2001; Keeling, Woolhouse et al. 2001; Ferguson, Cummings et al. 2006; Riley and Ferguson 2006).

Spatial models of disease transmission are increasingly in demand not only because of their capability to capture the complexity of transmission dynamics, but also because spatially heterogeneous interventions are being more frequently considered for cost-effectiveness and logistical reasons, especially during large-scale outbreaks (Wu, Riley et al. 2006). This is facilitated by the improved availability of spatially disaggregated incidence and demographic data and computational facilities (Riley 2007). However, empirical studies on the spatial dynamics of SARS

have been scarce (Lai, Wong et al. 2004; Meng, Wang et al. 2005; Wang, McMichael et al. 2006) with the limited availability of spatial data and lack of data sharing. Models of SARS transmission incorporating space have also been limited (Bauch, Lloyd-Smith et al. 2005), although the need of this kind of model has been highlighted (Dye and Gay 2003).

A special feature of the SARS outbreak was the high proportion of hospital infection in most countries which were heavily affected by SARS (World Health Organisation (WHO) 2004), despite hospital infection control measures having been in place (Bell and In 2004). The significance of hospital infection lies in its potential to generate further transmission not only in the hospitals but also in the community (Lloyd-Smith, Galvani et al. 2003; Varia, Wilson et al. 2003; Wong, Lee et al. 2004), and as the driving force of the spatial spread of the epidemic between districts (Meng, Wang et al. 2005). Therefore the aim of this study is to acquire a better understanding of the transmission dynamics between the hospital and the community, in particular its role in determining spatial spread of the disease.

This following section reviews the key features of SARS and its outbreak in Hong Kong in 2003, followed by discussions of the current findings on the spatial dynamics of emerging respiratory infectious disease epidemics, and of the SARS outbreak in particular. This leads to the research questions for this study at the end of this chapter.

1.2 Literature Review

1.2.1 SARS

1.2.1.1 Transmission dynamics

The SARS coronavirus is believed to be zoonotic in origin, with the civet cat likely to be an important animal reservoir for the virus (Guan, Zheng et al. 2003). The virus is likely to have jumped host from animals to humans in the animal markets in China, especially those in Guangdong province (Webster 2004), where the first human case was identified on 16 November 2002 (Zhong, Zheng et al. 2003). A probable case of SARS was defined by the World Health Organisation (WHO) using three criteria: (1) a suspect case with radiographic evidence of infiltrates consistent with pneumonia or respiratory distress syndrome on chest X-ray; (2) a suspect case of SARS that is positive for SARS coronavirus by one or more assays; (3) a suspect case with autopsy findings consistent with the pathology of respiratory distress syndrome without an identifiable cause (World Health Organisation (WHO) 2003).

The primary mode of SARS transmission in humans is through close contact and respiratory droplets. The mean incubation period was 4.4 days [95% CI: (3.7,5.1)] (Cowling, Muller et al. 2007). The mean serial interval, defined as the time from the onset of symptoms in an index case to the onset of symptoms in a subsequent case infected by the index patient, was 8.4 days (SD 3.8 days) (Lipsitch, Cohen et al. 2003). Common clinical symptoms at hospital admission include high fever, malaise, cough and headache. Most SARS patients received treatment that was based on previous experience with respiratory infections and evolved over the course of the epidemic, with ribavirin and corticosteroids being the most common treatments (Groneberg, Poutonen et al. 2005). Overall the case fatality ratio of SARS was 15% (World Health Organisation (WHO) 2003). The basic reproduction number of the SARS outbreak was around 3 based on estimates from various studies of the epidemic (Bauch, Lloyd-Smith et al. 2005).

The proportion of SARS cases in healthcare workers was high, estimated to be 57% in Vietnam, 43% in Canada, 41% in Singapore and 22% in Hong Kong (World Health Organisation (WHO)). Unrecognised SARS patients in the hospitals were the source that triggered some of the hospital outbreaks in Taiwan (Dwosh, Hong et al. 2003), Toronto (Svoboda, Henry et al. 2004), Beijing (Shen, Ning et al. 2004) and Hong Kong (Wong, Lee et al. 2004). Some of these became super-spreading events (Varia, Wilson et al. 2003; Shen, Ning et al. 2004; Wong, Lee et al. 2004). In some of the hospital outbreaks around 42-57% of cases were in healthcare workers, 14-31% were in hospital visitors, and 12-14% were in inpatients (Varia, Wilson et al. 2003; Chen, Leo et al. 2006). Some of these hospital-acquired cases generated secondary transmission in the community in household settings (Varia, Wilson et al. 2003), especially by hospital visitors (Lau, Lau et al. 2004) but least likely by healthcare workers (Ho, Sung et al. 2003; Goh, Lee et al. 2004; Lau, Lau et al. 2004; Wilson-Clark, Deeks et al. 2006).

Public health measures implemented to control the SARS outbreak by various countries included: (1) contact tracing and isolation of close contacts; (2) decreasing the time from symptom onset to isolation of patients; (3) increasing social distance; (4) enhancement of infection control measures in healthcare and community settings; (5) International travel restriction and entry and exit screening (Bell and In 2004). The timing of the onset of symptoms relative to peak infectivity is likely to have been the most crucial factor in the success of these public health interventions in controlling the SARS epidemic (Fraser, Riley et al. 2004). In SARS patients, viral load peaked between 5 and 10 days after the onset of illness and overt clinical symptoms (Peiris, Chu et al. 2003). Subclinical infection was rare (Leung, Chung et al. 2004; Lai, Keung Ng et al. 2005; Lee, Wong et al. 2006; Leung, Lim et al. 2006). Therefore, there is a period after symptom onset when SARS patients can be isolated and

quarantined before infectiousness increases, and the effectiveness of these measures to interrupt transmission is high.

1.2.1.2 The Hong Kong outbreak

Hong Kong is a former British colony, and since 1997 a Special Administrative Region of the People's Republic of China. Covering Hong Kong Island, the Kowloon Peninsula, the New Territories, the Lantau Island and other surrounding islands (Figure 1.1), it has a land area of 1,100 km^2, and a population of 6.8 million in 2003. It is subdivided into 18 district council districts (Figure 1.1) for administrative purposes, and 292 tertiary planning units (TPU) [Figure 1.2] for town planning purposes. Hong Kong is the fourth most densely populated territory in the world (United Nations 2006), with a population density of about 6,300 people per km^2. With only less than 25% of the territory's landmass developed, Hong Kong has a concentrated high-rise environment. Most of the territory's urban development exists on Kowloon peninsula, along the northern edge of Hong Kong Island, and in scattered settlements throughout the New Territories. 97% of the population live in the urban area, and there is great variation in the population densities of the TPUs (Figure 1.2). Hong Kong has a highly developed transportation network, encompassing both public and private transport. Over 90% of daily journeys are on public transport (Transport Department. The Government of the Hong Kong Special Administrative Region 2005). Situated at the south-eastern tip of China, it borders the city of Shenzhen in Guangdong Province and has close social and economic ties with the region. Cross-boundary traffic by land to and from Mainland China is about 250,000 passengers on weekdays.

The SARS outbreak in Hong Kong was traced to an infected professor from Guangzhou who arrived in Hong Kong in mid February 2003. During his stay at the Metropole Hotel in Kowloon (Figure 1.1), the infection spread to other guests and visitors, who subsequently triggered off a chain of outbreaks in Hong Kong, Singapore, Canada and Vietnam (Leo, Chen et al. 2003; Poutanen, Low et al. 2003; Tsang, Ho et al. 2003). In Hong Kong the SARS outbreak lasted for around 3.5 months, caused 1755 cases and 299 deaths, and the overall incidence rate was 0.026% per person. Figure 1.3 shows the epidemic curve of the SARS epidemic in Hong Kong. The outbreak was characterised by a significant proportion of transmission in hospitals and two super-spreading events – the outbreaks in Prince of Wales Hospital (PWH) in Sha Tin district which started at day 20 (where day 1 of the epidemic is defined as the symptom onset date of the first case in Hong Kong - 15 February 2003) and in the Amoy Gardens - a private housing estate in Kwun Tong district in Kowloon which started at day 28. The epidemic grew significantly with the PWH outbreak and peaked with the community

Figure 1.1 Hong Kong in 18 district council districts and the location of the SARS treating hospitals, Amoy Gardens Block E, and Metropole Hotel.

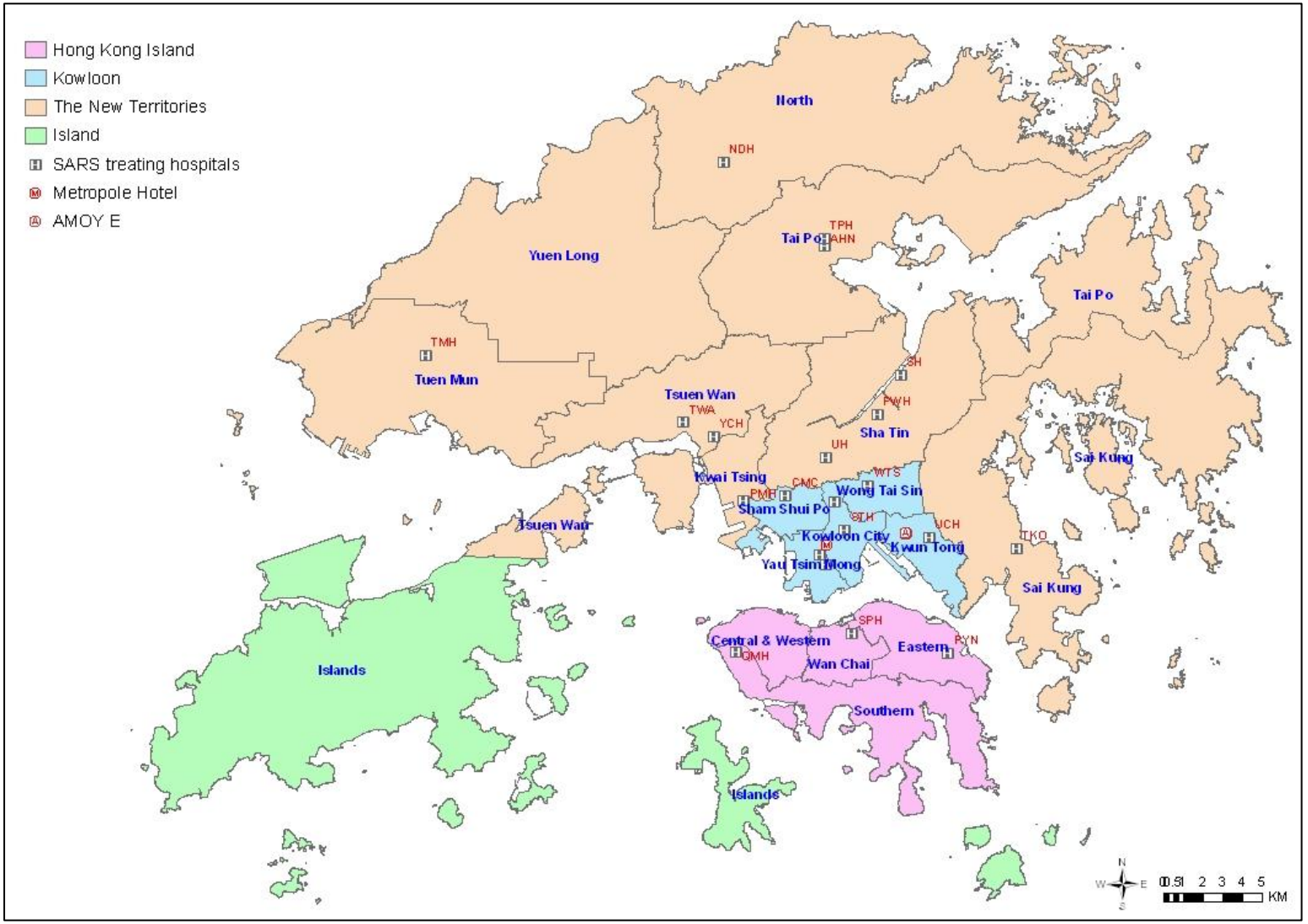


Figure 1.2 Population density by tertiary planning units (TPUs) of Hong Kong.

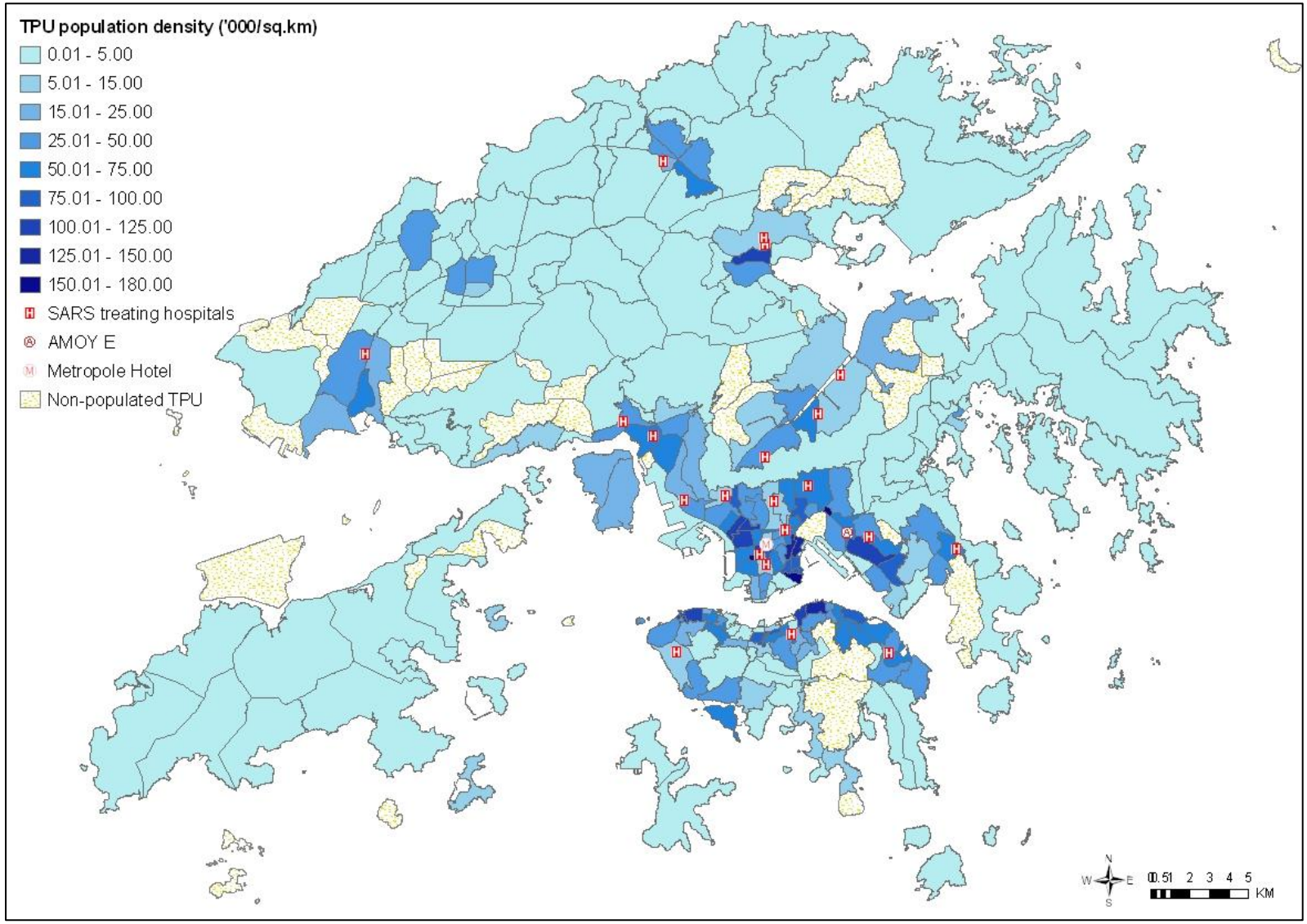
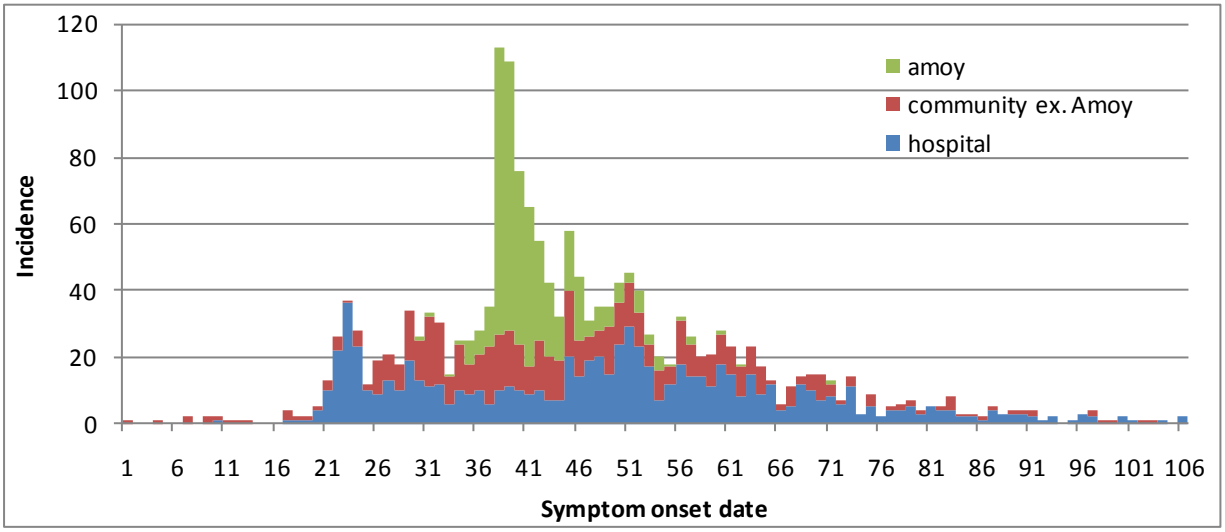


Figure 1.3 Epidemic curve (daily incidence) of the SARS outbreak in Hong Kong. Symptom onset date of the index case of Hong Kong, 15 February 2003, was set as day 1 of the epidemic.



outbreak in Amoy Gardens. After that, it declined slowly in the initial period and accelerated into terminal decline by mid May.

The Prince of Wales Hospital is a 1200-bed public hospital located in the New Territories East region. The outbreak first came to light when there were abnormal patterns of sick leave amongst staff working in ward 8A. At that time knowledge about SARS was limited. The index case of the outbreak was an inpatient in ward 8A who was linked to the professor from Guangzhou through his visit to the Metropole Hotel (Tsang, Ho et al. 2003). In ward 8A there were 123 healthcare workers, inpatients and visitors infected. The outbreak eventually spread to other wards, and by the end of the epidemic the total number of infections in PWH was 243 (data sourced from the data management work of this study, details are given in chapter 2).

Around the time that the PWH outbreak started, there were outbreaks of smaller scale in other healthcare establishments. They were located in: Pamela Youde Nethersole Eastern Hospital (PYN) and St Paul's hospital (SPH) on the Hong Kong Island, Baptist Hospital (BTH) in the Kowloon City district, Queen Elizabeth Hospital (QEH) in the Yau Tsim Mong district, and the clinic of two general practitioners in Yau Tsim Mong district and Tai Po district (Severe Acute Respiratory Syndrome Expert Committee 2003).

Amidst the PWH outbreak and the smaller outbreaks in other healthcare settings was the Amoy Gardens outbreak. The housing estate has an estimated 19,000 residents, and is made of 19 housing blocks (Blocks A – S), and each block has 33 floors, with 8 units on each floor. Units in the same stack on different floors are served by plumbing and drainage piping systems running vertically along the

external wall. Adjacent units are separated by a narrow lightwell, where bathroom windows and exhaust fans face one another. The super-spreading event was first alerted at day 40, when 15 suspected SARS cases from 7 households residing in the Amoy Gardens Block E were admitted to United Christian Hospital (UCH). Most of the cases were from units 7 and 8 of Block E. The index case of the outbreak was an inpatient in the 8A, 8C and 8D wards during the PWH outbreak, and stayed overnight in Amoy Gardens for 2 nights between his hospitalisations. Eventually it emerged that the vertical virus spread was related to the drainage and sewage system, and the running exhaust fan served to transport into the lightwell contaminated droplets from the bathroom evacuation (McKinney, Gong et al. 2006). All Block E residents were then evacuated to government holiday camps for 10 days for quarantine (Severe Acute Respiratory Syndrome Expert Committee 2003). The Amoy Gardens outbreak infected in total 447 cases, including those from other blocks of the Amoy Gardens, the Amoy shopping mall, and the residential buildings and housing estates nearby the Amoy Gardens (Lower and Upper Ngau Tau Kok Estates and Choi Ha Estate) (See Chapter 2 – Data management for the probable source of infection derivation).

The Princess Margaret Hospital (PMH) in Kwai Tsing district is a 1,200-bed acute hospital, and the only hospital in Hong Kong with dedicated infectious disease wards with 86 beds. At day 43 it was assigned as the designated hospital to treat new SARS cases, and throughout the epidemic it treated the highest number (593, 34%) of SARS patients. However, the unexpected surge in cases mainly from the Amoy Gardens outbreak stretched the hospital's capacity to the limit, and eventually hospital transmission increased. Admission of new SARS cases to PMH was suspended at day 56, and new SARS cases were referred to Tuen Mun Hospital (Severe Acute Respiratory Syndrome Expert Committee 2003). Throughout the epidemic, 85 cases were infected in the PMH (Chapter 2).

Alice Ho Miu Ling Nethersole Hospital (AHN) in the Tai Po district is a sister hospital of PWH. The closure of the accident and emergency department in PWH at day 27 because of the 8A ward outbreak increased the workload in AHN, and an outbreak in the hospital was initiated at day 37 when some of the patients admitted with other medical problems later turned out to have SARS (Severe Acute Respiratory Syndrome Expert Committee 2003). There were 136 cases infected in AHN, and it was the second most infected hospital of the outbreak (Chapter 2).

In the declining phase of the epidemic, there were a number of outbreaks in the hospitals and residential buildings in the community. The hospital outbreaks were in Tai Po Hospital (TPH) in Tai Po district, United Christian Hospital (UCH) in Kwun Tong district, Caritas Medical Centre (CMC) in Sham Shui Po district, Tuen Mun Hospital (TMH) in Tuen Mun district, North District Hospital (NDH) in North district, and a second outbreak in Baptist Hospital (BTH). Community outbreaks in residential

buildings were located in Koway Court in Eastern district on the Hong Kong Island, Hing Tung House in Tung Tau Estate situated in the Wong Tai Sin district of Kowloon, and Wing Shui House in Lek Yuen Estate in the Sha Tin district (Severe Acute Respiratory Syndrome Expert Committee 2003).

There were also outbreaks in the community in 8 elderly care homes. Imported cases included those who had visited China or Taiwan, or were infected in flight (Chapter 2).

The key public health control measured implemented at different stages of the Hong Kong epidemic included contact tracing and quarantine of close contacts, suspension of classes in childcare centres, schools and universities, travel advice and body temperature screening at the borders and encouraging enhanced infection control and personal hygiene in the community. Restricted hospital visitor policies were also implemented in some hospitals at the critical period of the epidemic (Severe Acute Respiratory Syndrome Expert Committee 2003).

Compared to the general Hong Kong population, there was an excess of young adults, especially women, infected with SARS, with a significant proportion of them being nurses. There was relative deficit of children and adolescents, and over-representation of elderly men (aged 75+) in the SARS cases. The ratio of women to men in the infected individuals was 5:4. The case fatality ratio was 17.2% (Leung, Hedley et al. 2004).

1.2.2 Spatial dynamics of emerging respiratory infectious disease epidemics

1.2.2.1 Patterns of spatial spread

Previous studies on the spatial patterns of emerging respiratory infectious disease epidemics is very limited due to the lack of disaggregated spatial incidence data available, with the past pandemic influenza probably the only emerging respiratory infectious disease studied in detail. Interpandemic influenza and measles are recurring respiratory infectious diseases and their spatial spread and mechanisms have been extensively studied. Measles is known for its biennial pattern, its spatial diffusion being driven by the birth rate, critical community threshold and vaccination (Finkenstadt and Grenfell 1998; Finkenstadt, Keeling et al. 1998; Grenfell and Bolker 1998), which are not relevant to the epidemiological nature of SARS. Therefore I have focused primarily on the findings from pandemic and interpandemic influenza.

Prior work (Cliff, Haggett et al. 1986) has characterised phenomenologically the spatiotemporal spread of epidemics of respiratory infections as wave-like, hierarchical, or a mixture of both. In a wave-like process, disease spreads from its centre of introduction to the physically nearest neighbouring centres. In a hierarchical diffusion process, disease spreads progressively through the

population hierarchy, typically from large to small centres. A mixed diffusion process contains both the wave-like and hierarchical components of spread, and best described the patterns in the influenza pandemic in 1918-1919 in England and Wales (Smallman-Raynor, Johnson et al. 2002), the interpandemic influenza in the United States between 1972-2002 (Viboud, Bjornstad et al. 2006), and influenza in Iceland from 1945-1970 (Cliff, Haggett et al. 1986). Earlier onset in larger populations were observed (Cliff, Haggett et al. 1986; Viboud, Bjornstad et al. 2006; Chowell, Bettencourt et al. 2008), and wave-like diffusion became more prominent as the pandemic in England and Wales and epidemics in Iceland progressed (Cliff, Haggett et al. 1986; Smallman-Raynor, Johnson et al. 2002).

Underlying these empirically observed spatiotemporal patterns are the mechanisms that drive the geographic dispersal of infection, which ultimately are related to human movement patterns. For pure wave-like diffusion, travel needs to decrease with geographic distance, while hierarchical diffusion requires travel to depend on destination population size.

Gravity models are widely used in human geography to model the relationship between human movement patterns and their key driving factors (Haynes and Fotheringham 1984; Fotheringham and O'Kelly 1988). In studies of epidemics, they are often incorporated in mathematical models to capture both distance and population size dependence in human movement patterns, and can reproduce both wave-like and hierarchical spread (Xia, Bjornstad et al. 2004; Ferguson, Cummings et al. 2006; Viboud, Bjornstad et al. 2006; Balcan, Colizza et al. 2009; Eggo, Cauchemez et al. 2010; Merler and Ajelli 2010). Commuter flows from homes to workplaces are found to be predictive of the spatial patterns of influenza epidemics compared to other types of travel data, and it outperformed models that only assume coupling decays to be a simple function of geographical distance (Viboud, Bjornstad et al. 2006).

Application of gravity coupling in epidemic models are based on the assumption that heterogeneous contact rates between individuals divided into subpopulations are proportional to flow volume between the subpopulations. Flow patterns modelled by gravity models are directional (from origin to destination), therefore under the compartmental model framework spatial mixing assumption of who makes the move is required. Simple unilateral assumption that infectors move to the susceptible population (Xia, Bjornstad et al. 2004; Eggo, Cauchemez et al. 2010), or susceptible move to the infector subpopulation (Merler and Ajelli 2010) are applied in most of the studies. Only one study assumed the more realistic mixing that susceptible and infector both move (Balcan, Colizza et al. 2009). However, it is uncertain how different spatial mixing assumptions affect the model results, as all studies made the assumption without assessment.

In the Beijing SARS outbreak, it was suggested that spatial spread between districts was overall mostly driven by hospital related factors (high numbers of doctors) and also by hierarchical spread (with effects of population density stronger than population size), while wave-like spread dominated at the early phase of the epidemic (Meng, Wang et al. 2005). However, the study used the joint-count statistic (Moran 1948) to measure the spatial autocorrelation of having new cases reported in the districts on the same day, not taking into account the time gap of the serial interval if a case from a district had infected individuals from other district, and the within-district transmission that confounds the result. Establishing the dynamics of spatial spread within and between districts or transmission settings (e.g. hospital and community) is best achieved using infection tree inference (Ferguson, Donnelly et al. 2001; Wallinga and Teunis 2004), where the occurrence of infection between individuals is estimated probabilistically, taking into account the timeline of infection. More details about these methods are discussed in a later section.

Two empirical measures used to characterise spatiotemporal spread are the velocity of the geographical spread of the epidemics (Smallman-Raynor, Johnson et al. 2002; Cliff, Haggett et al. 2008), and the trajectories of the centroid of reported disease incidence (Cliff, Haggett et al. 1986). The speed and trajectory of spread can inform the timing of local epidemic preparation, provided the measures are predictive in nature, and take into account the stochastic property of disease transmission. This requires understanding of the underlying drivers of spatiotemporal spread, otherwise the inference of these measures are restricted to the description of the unique happening of a single epidemic. A powerful tool that can capture and study the complexity of the transmission dynamics and incorporate the spatial components is mathematical modelling of epidemics, and it is discussed further in a later section.

1.2.2.2 Spatial cluster

A spatial cluster can be defined as “a geographically bounded group of occurrences of sufficient size and concentration to be unlikely to have occurred by chance” (Knox 1989). In infectious diseases, spatiotemporal clusters are usually studied, because the relative timing of the occurrence of the cases reflects the possible epidemiological links of the cases.

Two properties of spatial clustering are of particular relevance to infectious disease epidemics – the spatial autocorrelation, and the location of disease clusters. The spatial distribution of the individuals infected with respiratory disease is likely to exhibit spatial autocorrelation. Global clustering tests study the existence of clustering in the patterns and indicate the extent of autocorrelation or spread in the patterns, allowing comparisons among various spatial clusters as in the SARS outbreak in Hong Kong in 2003 (Lai, Wong et al. 2004).

More than a hundred global clustering tests exist (Kulldorff 2006), of which some of the more popular are Moran's I (Moran 1950), Geary's C (Geary 1954), Whittemore's test (Whittemore, Friend et al. 1987), Cuzick–Edward's k-nearest neighbors method (Cuzick and Edwards 1990), Besag–Newell's R (Besag and Newell 1991), Swartz entropy (Swartz 1998), and Tango's maximized excess events test (Tango 2000). Various studies have compared these methods (Oden, Jacquez et al. 1996; Hill, Ding et al. 2000; Kulldorff, Tango et al. 2003; Song and Kulldorff 2003).

Identification of the location of disease clusters facilitates the formulation of control measures in highly affected areas of disease transmission. Cluster detection methods identify locations of excessive cases which are not due to chance. Current methods include Kulldorff's scan statistic in circular or elliptic windows (Kulldorff 1997; Kulldorff, Huang et al. 2006), Tango's flexibly shaped spatial scan statistic (Tango and Takahashi 2005), and the simulated annealing method (Duczmal and Assuncao 2005). The first three methods differ in the shape of the scan window, and the simulated annealing method uses graph-based techniques to detect spatial clusters of arbitrary geometric form. The circular scan statistic is most widely applied because of its power to detect spatial clusters, while being less computer intensive than the other methods. Power evaluations showed that circular scan statistics works very well even for non-circular clusters compared with the elliptic window scan statistic and the simulated annealing method (Duczmal, Kulldorff et al. 2006). Another study found that the circular scan statistic performed well for elliptic shaped clusters, and vice versa (Kulldorff, Huang et al. 2006). The flexibly shaped scan statistic has lower accuracy than the circular scan statistic in detecting circular clusters, and there is a restriction on the maximum cluster size it can detect (Tango and Takahashi 2005).

A study of the spatial distribution of the SARS cases in Hong Kong (Lai, Wong et al. 2004) suggested that cases were clustered throughout the epidemic by the Moran's I (Moran 1950) and the nearest neighbour statistics (Clark and Evans 1954). Spatiotemporal clusters of the cases in the Hong Kong outbreak were identified visually from smoothed maps of the cases, but the precise geographic location of clusters were not identified nor proven significant statistically, and the method of smoothing and the kernel applied were not specified. Comparing the spatial range of the cases of the recognised disease clusters, Amoy Gardens cases were found to be more clustered than the PWH cases, and the descending order of clustering by the types of hospital infection was PWH ward 8A inpatients, PWH healthcare workers, and PWH ward 8A visitors.

The Amoy Gardens outbreak in Hong Kong highlighted the importance of real-time detection of case clusters during an epidemic, so that appropriate control measures can be implemented as early as possible. When the outbreak was first alerted, there were already more than 200 cases, and the

peak of the outbreak had passed. When an isolation order on the residential block was issued five days later, and evacuation of the residents two days after, there were almost 350 cases. Learning from this experience, health officials began to announce daily the names of residential buildings with new cases for public vigilance. A new outbreak data management system (e-SARS, MIIDSS and SARS-CCIS) (described in Section 2.1.1) was subsequently set up, and high risk locations of SARS occurrence were identified by the system (Severe Acute Respiratory Syndrome Expert Committee 2003), though the criteria used to define a high risk location was unclear.

The cluster detection methods mentioned above are usually applied retrospectively - the tests are run on the full dataset to detect the spatial or spatiotemporal clusters during the whole period covered by the dataset. If they are applied to mimic a prospective scan for real-time cluster detection by repeatedly running the tests whenever a dataset is updated with the latest data, statistical inference from such analyses has to be adjusted for the multiple testing of the previous analyses conducted. Kulldorff's prospective scan statistic (Kulldorff 2001) is the prospective version of the Kulldorff's scan statistic (Kulldorff 1997; Kulldorff, Huang et al. 2006) with multiple testing adjustment to detect recently emerged clusters with high excess risks and long-existing clusters with moderately excess risks. Its application as a real-time cluster detection tool in a large-scale outbreak remains to be evaluated.

1.2.2.3 Spatial heterogeneity

Mathematical modelling of epidemics capturing spatial dynamics have been used as the scientific tools of epidemic monitoring and control strategy evaluation during recent emerging outbreaks like the foot and mouth epidemic (Ferguson, Donnelly et al. 2001; Keeling, Woolhouse et al. 2001), the SARS outbreak (Riley, Fraser et al. 2003), and the H1N1 pandemic in 2009 (Fraser, Donnelly et al. 2009; Ghani, Baguelin et al. 2009). Disease transmission and interactions are localized in nature, and are governed by epidemiological, environmental and social factors. Under homogeneous mixing, variations in these factors are averaged out, and each individual has an equal chance of being infected from each infected individual in the population (Anderson and May 1991). This assumption is often applied in standard epidemic models for simplicity and because of inadequate computational resources in the past, at the possible trade-off of missing the key determinants of the observed patterns of the epidemics when the effects of heterogeneities are influential.

Heterogeneities can be of different forms. Spatial heterogeneity refers to the differences between populations or individuals at different geographic locations - which can arise from the underlying difference in the fundamental forces governing the population dynamics, for example different social structure leading to variation in contact rates, or from the emerging dynamic differences, such

as the proportion of the population that are infectious (Keeling and Rohani 2007). I classified spatial heterogeneities into those that arise from the exposure to the infected individuals, and from the susceptibility of the uninfected individuals. Considering the residential location of the individuals as the base location, spatial variation in exposure is primarily determined by the interactions of the movement and contact patterns of the susceptible and infected individuals leading to disease transmission, with reference to their base locations. Gravity models incorporated in metapopulation epidemic models where individuals are classified by spatial location and disease state have proven successful in capturing this heterogeneity as discussed previously. Spatial variation in susceptibility could be related to the geographic or demographic properties of the location where one is based, as well as individual's attributes (e.g. high risk age group). In the SARS outbreak, it is possible that areas where hospitals treating SARS cases are located or nearby, or areas with high proportion of the population with the high risk attribute(s), could be more susceptible. Such hypotheses relating to spatial spread have not been formally investigated.

1.2.3 Mechanistic spatiotemporal models

Previous studies using mechanistic spatiotemporal models to fit epidemic data and capture the effects of spatial heterogeneity can be classified as statistically rigorous and ad-hoc fitting approaches. Statistically rigorous approaches include the work of Ferguson *et al.* (Ferguson, Donnelly et al. 2001), which used an extended infection tree inference method fitted with an expectation-maximisation algorithm to fit a transmission model to data from the 2001 UK Foot and Mouth Disease (FMD) epidemic. In that analysis, the relative risk of transmission was considered to be driven by farm-specific factors of susceptibility and infectivity of disease transmission, and the spatial distance between farms. Spatiotemporal data on the host population was combined with the farm-based case data to estimate spatial transmission coefficients and other epidemiological parameters, as well as generating infection trees.

Wallinga and Teunis (Wallinga and Teunis 2004) formalised the infection tree inference in a non-spatial context, and used the probability distribution of the serial interval of cases (time interval between the symptom onset between the primary and the secondary cases) to infer the infection tree from data on the onset dates of individual SARS cases using a partial likelihood. Compared to Ferguson *et al.*'s method, Wallinga and Teunis method is individual-based, and it assumes all the infected cases of the same symptom onset date to be equal in their susceptibility and infectivity of disease transmission, regardless of their demographic attributes and spatial distance. To infer probabilistic infection trees that take into account the spatial heterogeneities of individuals, the

Wallinga and Teunis method has to be extended spatially to capture the heterogeneous contact rates between individuals based on their reference location.

Other spatially explicit epidemic models includes the work of Gibson (Gibson 1997), which estimated the transmission parameters for a spatiotemporal stochastic model in plant epidemiology using Markov Chain Monte Carlo (MCMC) methods, where the data was the status of the system at individual time points, rather than on individual cases. Chis Ster and colleagues modified Gibson's approach to apply it to data on the location and timing of individual cases of the FMD epidemic, first assuming the epidemic was completely observed (Chis Ster and Ferguson 2007), then allowing for hidden infections using Reversible Jump Markov Chain Monte Carlo (RJMCMC) (Chis Ster, Singh et al. 2009).Xia *et al.* (Xia, Bjornstad et al. 2004) fitted a time series susceptible-infected-recovered (TSIR) model, using a gravity model of coupling between patches, to the measles epidemics in England and Wales. Compared to these methods, the Wallinga and Teunis method is less computationally intensive and does not require the population data which are often not available but only the cases incident data.

Studies using less rigorous fitting methods include the work of Haydon *et al.* (Haydon, Chase-Topping et al. 2003), which constructed the infection trees of the FMD epidemic by adopting rules based on geographic distance between farms to select the index farm from the list of possible candidates generated by the time of infection. Keeling *et al.* (Keeling, Woolhouse et al. 2001) used the contact tracing data to estimate the spatial kernel of transmission for the 2001 FMD epidemic, and fitted the simulation models to data using brute force (generating thousands of simulation runs per parameter point and searching parameter space for the best least square fits with regional incidence trends).

Among the several dozen epidemic models studying the SARS outbreak at the country level, only one contained explicit spatial components (Riley, Fraser et al. 2003) to capture the effects of spatial heterogeneity, demonstrated by 70% of the cases distributed in 3 out of the 18 districts in Hong Kong in the midst of the epidemic. The stochastic metapopulation compartmental model structured the mixing into three levels: within district, between contiguous districts and between more remote districts, with the population within each district classified as susceptible, latent, infectious, hospitalized, recovered or dead. Brute force simulation was used to fit the model to the data. The lack of spatial models despite their importance in studying the SARS outbreak (Bauch, Lloyd-Smith et al. 2005) could be due to the lack of empirical findings of spatial dynamics that inform the model formulation, and the limited availability of spatial data.

The key outcome of the epidemic models is the reproduction number. The basic reproduction number, R_0, also known as the basic reproductive number, rate or ratio, is one of the most important quantities governing the spread and the time-scale of an epidemic. It is defined as the expected number of secondary infections caused by a typical infected case during his or her entire infectious period in an entirely susceptible population. The threshold of R_0=1 has important implications for the development of the epidemic, and therefore the extent of the control measures needed to contain it. With $R_0 < 1$, the epidemic will not be able to sustain itself even without any intervention. Whereas a $R_0 > 1$ implies that the epidemic will grow indefinitely, and the disease transmission will need to be reduced by a factor of $(1-1/R_0)$ to keep it under control, under the mass action assumption of epidemic transmission in a large community. Such reduction can be achieved by reducing social contacts (e.g. by quarantine and social distancing), reducing the susceptibility of individuals (e.g. by vaccination and prophylaxis if they are available), and reducing infectiousness (e.g. by treating the infected individuals).

Factors affecting R_0 are multi-faceted, because R_0 by nature is subject to the biological properties of the underlying pathogen, the health-related factors of the infective and susceptible individuals, and also the behaviour and interactions between the hosts and susceptible. The biological nature of the pathogen determines the natural history of disease progression and provides the possible ranges of the timeline of transmission and virulence of the disease. Host factors like age and co-morbidities, and population structure like demography and contact patterns, affect individual's variation on the timeline of transmission, and efficiency of the transmission between hosts and thus the timescale of transmission.

The effective reproduction number, R_t, is the actual average number of secondary cases per primary case observed in a population following the start of the epidemic. Its value is typically smaller than that of the R_0, because of depletion of the susceptible by the infection, and the impact of control measures. The effective reproduction number by transmission settings, for instance, hospital-to-community transmission, community-to-community transmission etc, informs and facilitates the comparison of the transmission potential of the settings.

Current studies (Lloyd-Smith, Galvani et al. 2003; Cori, Boelle et al. 2009) quantifying the transmission within and between community and hospital of the SARS outbreak have shown opposite results. In one study, healthcare workers were estimated to be the most important driving forces of disease transmission, and they could have generated transmission in both the hospital and community (Lloyd-Smith, Galvani et al. 2003). In the other study, it was estimated that the contact rate of individuals within hospitals was very low. The difference could possibly be attributed to

different model structures, and the more importantly the assumptions on who can infect who in the community and hospital settings. Validity of the latter assumptions is best informed by contact tracing and source of infection data of the epidemic (chapter 2), and methods that capture all the observed transmission routes in the epidemic are needed to quantify the relative role of hospital and community transmission.

1.3 Study objectives

The SARS outbreak in Hong Kong in 2003 comprised transmission in the hospitals and community settings and the environment-driven Amoy Gardens outbreak, and each of these sources exhibited different mechanisms of the spatial spread of the disease. The majority of the Amoy Garden cases were highly concentrated in the surrounding areas of the Amoy Gardens (Lai, Wong et al. 2004) because of environmental drivers of the outbreak. The spatial spread and disease transmission of other community settings could have been driven by the movement patterns of the population, and other factors of spatial heterogeneities arose from the demographic or geographical properties of the subpopulation or spatial location. In contrast, in hospitals where transmission took place between healthcare workers, inpatients and visitors, these infection subgroups could exhibit different levels of disease transmissibility in different transmission settings and hence act as different drivers for spatial spread in the community.

This study has aimed to investigate in a more comprehensive framework the effects of hospital transmission to community transmission, in particular the research question of:

“How did hospital-acquired infections drive onward spread in the community (spatial spread, spatial clustering, and the size of community transmission) and could this be attributed to factors in the hospital (such as type of hospital infections, i.e. healthcare workers, inpatients and visitors) versus properties of the community related to disease transmission (for instance population demographics and geographical properties)?”

This question is best broken down into the following parts for investigation.

I first studied the spatial distribution and identified spatiotemporal clusters of the SARS case using Kulldorff’s scan statistic in the overall, and by hospital and community transmission, to investigate any characteristics of the spatial patterns related to the transmission settings. Then using statistical modelling, I investigated the spatially-stratified risk factors – demographic, geographic and hospital transmission related – associated with the size of community outbreak.

Additionally, given the importance of cluster detection in real-time during an epidemic on control policy formulation, I have evaluated the use of the Kulldorff's prospective scan statistic (Kulldorff 2001) as such a tool for community transmission.

I then developed a spatial Wallinga and Teunis method that captures the spatial heterogeneity and all the observed transmission routes in inferring the infection tree, to quantify the spatial spread within and between areas and the transmission settings (hospital and community). The method also allows me to compare the transmissibility of healthcare workers, inpatients, hospital visitors and also community-acquired cases.

1.4 Outline of the remaining chapters

Chapter 2 describes the datasets used in this study, and the algorithm used to derive the probable source of infection data – the core dataset used in this study, as well as the hospitalisation data that were applied in chapter 6. The probable source of infection data are summarised in the chapter. Knowledge of the transmission routes of the epidemics was also acquired through the data validation steps summarised in that chapter.

In chapter 3 I first conducted exploratory analysis of the spatial distribution and spread of the outbreak, and then applied Kulldorff's scan statistic to identify the spatiotemporal clusters. I also investigated the risk factors associated with community outbreaks using six statistical models of different structures. At the end of the chapter, I evaluated the use of Kulldorff's prospective scan statistic as the real-time tool during the epidemic for spatial cluster detection.

In chapter 4, I reviewed the incorporation of gravity models in epidemic models and derived gravity models that best describe the movement of the Hong Kong population based on different travel data and different forms of the gravity models.

In chapter 5, I developed the partially extended spatial Wallinga and Teunis method incorporating the gravity models I derived in chapter 4, and studied the effects of the modifications and different spatial mixing assumptions that linked the gravity model to the heterogeneous contact rates of individuals on the model results.

In chapter 6, I reviewed the current studies that quantified the relative role of hospital and community transmission, and further modified the spatial Wallinga and Teunis method with spatial details of individual hospitals and community. Based on the inferred infection tree, I estimated the transmissibility by transmission settings and type of hospital and community infection, the extent of

within-district transmission, and derived the mechanism of spatial spread between areas. All these together formed a bigger picture of the spatial dynamics of the outbreak.

In chapter 7, I discuss the results and their implications, their limitations and potential future work.

Chapter 2. Data Validation and Summary

Here I describe how I derived the core dataset for this study – the source of infection data, based on the available data from different sources and summarise the data. I also derived a hospitalisation dataset extracted from three sets of hospitalisation data generated under different criteria. The validated hospitalisation data was used in quantifying the hospital and community transmission in chapter 6. Other datasets used in this study, including the geographic information system (GIS) and demographic data, are also described.

2.1 Outbreak data

2.1.1 Data collection

Collection of good quality outbreak data is of great importance and a pre-requisite for studying the outbreak to understand its dynamics and evaluate control strategies. In reality, given the urgency and uncertainties, especially at the beginning of the outbreak, this may not be easily achieved. In the SARS outbreak in 2003, Hong Kong had a relatively complete dataset compared to other countries (Anderson, Fraser et al. 2004), attributed to remarkable human efforts and the availability of existing hospital and police information systems and technologies. At the beginning of the outbreak, case investigation and contact tracing were carried out on patients reported to have SARS and hospitalized in Hong Kong by the Department of Health - the health advisor of the Hong Kong government responsible for the disease surveillance and control. Clinical and treatment data were collected by the Hospital Authority - a statutory and independent body responsible for the provision of public hospital services and 95% of total inpatient beds in Hong Kong, which has a well developed clinical information system.

Real-time information exchange was not possible with these separate databases until April, when new data management systems (e-SARS, MIIDSS and SARS-CCIS) were set up, and real-time interaction between the data systems became feasible among all relevant parties (Severe Acute Respiratory Syndrome Expert Committee 2003). The e-SARS system is a web-based system capturing the clinical details of SARS patients who were admitted to hospitals, and was administered mainly by the Hospital Authority. The SARS – Case Contact Information System (SARS-CCIS) was managed by the Department of Health for contact tracing and construction of the cluster trees. The Major Incident Investigation and Disaster Support System (MIIDSS) is a crime investigation program used by the Hong Kong Police Force that had been adapted to facilitate quick identification of linkage between cases and contacts and high risk locations of SARS occurrence. The contribution of the e-

SARS, MIIDSS and SARS-CCIS systems to disease control was recognised as the winner of the Stockholm Challenge Award in 2004.

2.1.2 Data validation and reclassification

The data used in this study were derived from the SARSID integrated database, which was coordinated by the Department of Community Medicine, University of Hong Kong on behalf of the Health, Welfare and Food Bureau. It was derived from the Hong Kong Hospital Authority e-SARS system and the Department of Health's master list, which consisted of epidemiological and case contact data. Data in the SARSID database included demographic, clinical and hospitalization data, exposure history, travel history, probable source of infection from the contact tracing questionnaire, and epidemiological cluster defined by the Department of Health based on the contact tracing data they collected. The mechanism used for defining the epidemiological cluster was not available.

On studying the epidemiological cluster variables on the SARSID, and cross-referencing with the epidemiological links of the early cases reported in the official report of the SARS outbreak in Hong Kong (Severe Acute Respiratory Syndrome Expert Committee 2003), it was discovered that cluster variables in the database do not truly reflect the source of infection of the cases, as the indexes of the epidemiological clusters were classified into the clusters that they generated (e.g. index of the Amoy cluster who was infected in hospital was classified into the Amoy cluster, cases infected in household settings by hospital infected indexes were grouped into the hospital clusters). Validation and reclassification of the source of infection data were needed for the purpose of this study.

The reclassification and validation algorithm developed makes use of all the relevant variables in the dataset and the epidemiological links on the SARS report. Subjects on the SARS report are anonymous and matching to the subjects on the SARS database was based on all the demographical and epidemiological data available. For those who were not matched on the SARS report, a hierarchy-based reclassification algorithm was used, with more reliable and informative variables checked first than the other variable to find out the probable source of infection of the cases. For instance, for cases indicated as infected from hospital from the original cluster variable, occupation as healthcare worker had the highest hierarchy to confirm the case was very likely to be infected in hospital and as healthcare worker, followed by hospitalisation record which showed the case was hospitalised before symptom onset (inpatient), and then any record indicating the cases had visited hospital during which an outbreak was ongoing. Based on this algorithm, certain proportion of cases originally classified as hospital infected were found to be more likely to be infected in household settings by an hospital infected index cases.

The reclassification was carried out by writing commands and validated manually when necessary, as some of the codes were not standardised, and some fields were truncated and the complete dataset was not available. All remaining subjects not classified by the algorithm were checked manually record-by-record for the most suitable classification of the source of infection.

Besides the new variable on the possible source of infection, the data reclassification and validation work resulted in more detailed source of infection subgroups, including:

(a) Hospital subgroups of ongoing inpatients (those infected whilst an inpatient for other health conditions and who stayed in hospital throughout for the non-SARS and SARS conditions), inpatients released (those infected while admitted for non-SARS condition but discharged during the SARS incubation period), visitors, healthcare workers, healthcare supporting staff and medical students;

(b) PWH 8A ward outbreak;

(c) Individual elderly care home and the subgroups of residents, visitors and staff;

(d) Household (live in the same household) and close contact subgroups;

(e) Subgroups of outbreaks in residential buildings;

(f) Amoy nearby buildings subgroups.

On the main group level, 239 cases had their new potential source of infection different from the epidemiological cluster variable by the Department of Health. Also coding of the new possible source of infection variable is more standardized and has higher information content.

The new possible source of infection was then sent to the Department of Community Medicine, University of Hong Kong for validation and approval, who had actually conducted an independent reclassification of the source of infection for their own research purposes. 218 mismatches were found between the two versions of the source of infection variables. After looking into and discussing every discrepancy, my version of the source of infection was adopted except one change was made [used in (Kwok, Leung et al. 2007)].

2.1.3 Potential source of infection summary

Based on the new potential source of infection, there were 772 (44.0%) hospital cases and 983 (56.0%) community cases. Excluding the 447 (25.5%) Amoy Gardens cases, hospital cases accounted for 59.0% of the cases and community accounted for 41.0% of the cases. A breakdown of hospital

cases by hospitals is given in Figure 2.1. PWH accounted for 31.5% of cases, together with AHN they accounted for 49.1% of cases, and together with PMH accounted for 60.1% of all hospital cases. Amoy Gardens accounted for 45.5% of the community cases. Composition of community transmission excluding the Amoy Garden cases is given in Figure 2.2. General community (where the specific location of infection could not be identified and the subject had no travel history) accounted for 34.7%, household transmission 18.1% and imported cases 15.9%.

Figure 2.1 Breakdown (in percentage of total number of hospital cases) of hospital infected cases by hospitals.

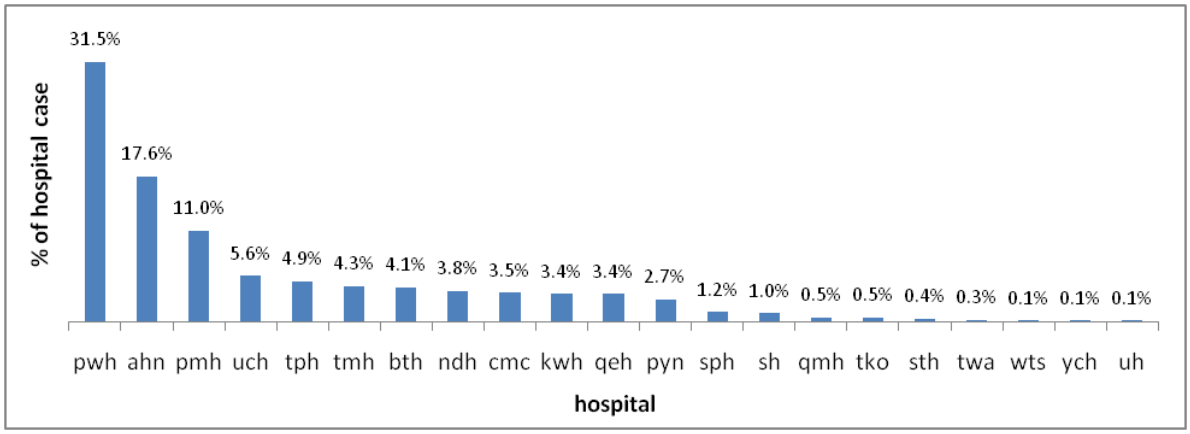
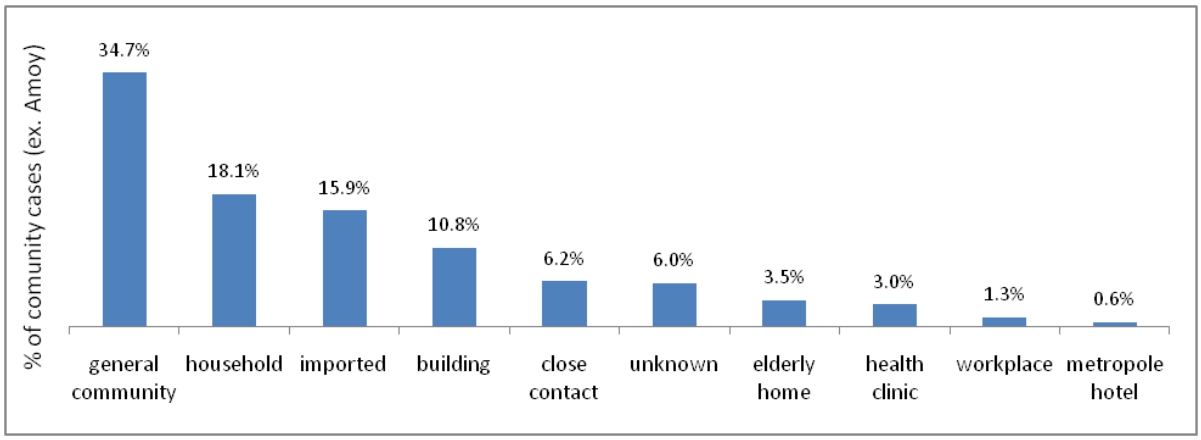


Figure 2.2 Breakdown (in percentage of total number of community cases excluding Amoy Garden cases) of the potential source of infection of the community infected cases.



Epidemic curves by infection source are given in Figure 2.3. The first case onset of the Hong Kong outbreak was on 15 February 2003, and it was set as the first day of the epidemic throughout this study. There was a higher incidence of hospital-based infection than community-based infection throughout most of the epidemic, except from day 31 to 46. The first peak in hospital-based transmission on day 23 was fuelled by the PWH outbreak, and the second peak around day 51 was caused by AHN and PMH (Figure 2.4). Together with UCH, TPH, TMH and BTH which all had more than 30 cases, these 7 hospitals accounted for 79.0% of all hospital transmission. Hospital cases were dominated by healthcare workers who accounted for half of the hospital infection (Figure 2.5),

Figure 2.3 Epidemic curves (daily incidence) of the hospital, community and Amoy Gardens outbreaks. Day 1 refers to 15 February 2003.

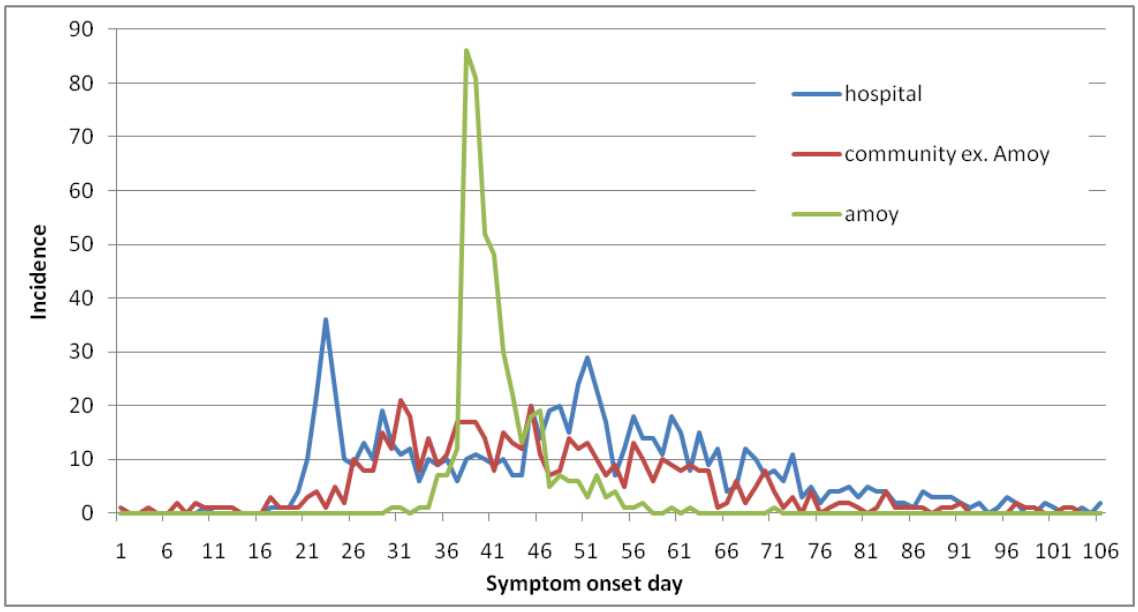


Figure 2.4 Epidemic curves (daily number of cases) of the 3 major hospital outbreaks (PWH, AHN, PMH).

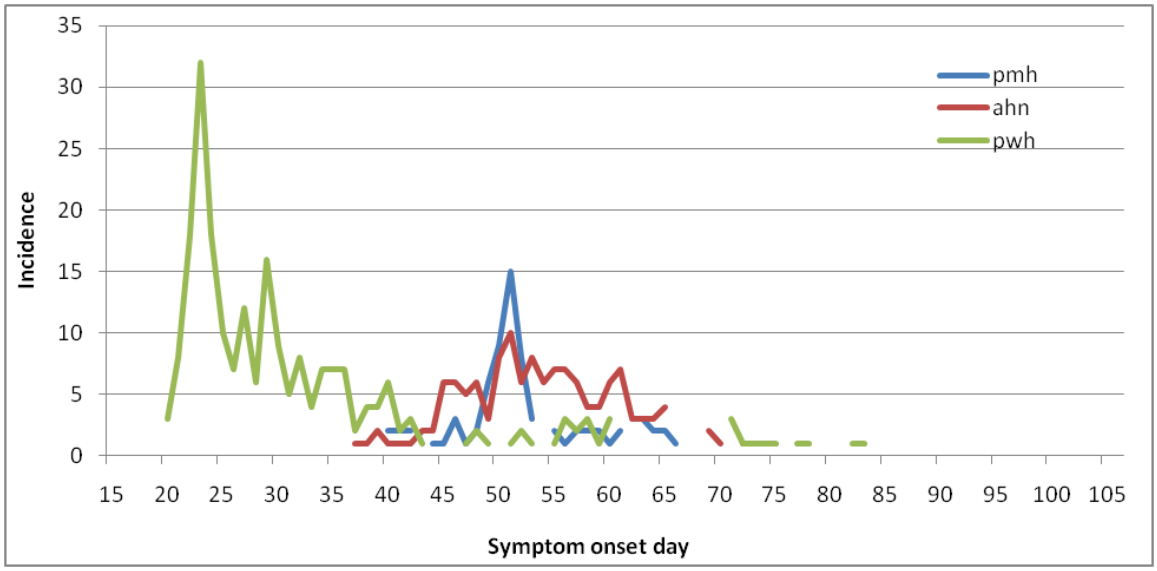
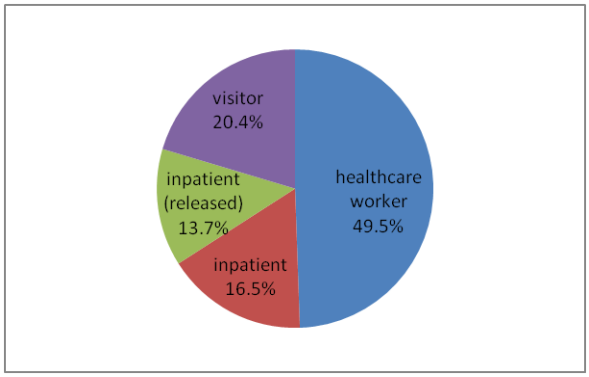


Figure 2.5 Composition of hospital infection by infection subgroup. Inpatient (released): those discharged during SARS incubation. Inpatient: ongoing inpatient.



then inpatients (30.2%, ongoing inpatients: 16.5% and inpatient discharged during SARS incubation: 13.7%), and visitors (20.4%).

Figure 2.6 shows that healthcare workers were the most infected group until around day 57, when the number of inpatient infections started to increase. The diminishing infection of hospital visitors could be due to the restricted visiting policy implemented in the later stage of the epidemic. Within community transmission, excluding Amoy Gardens cases, household and close contact transmission roughly accounted for one-fifth of the cases at the middle phase of the epidemic, and grew in importance in the later stage (Figure 2.7 and Figure 2.8).

Figure 2.6 Composition of the hospital infection throughout the epidemic.

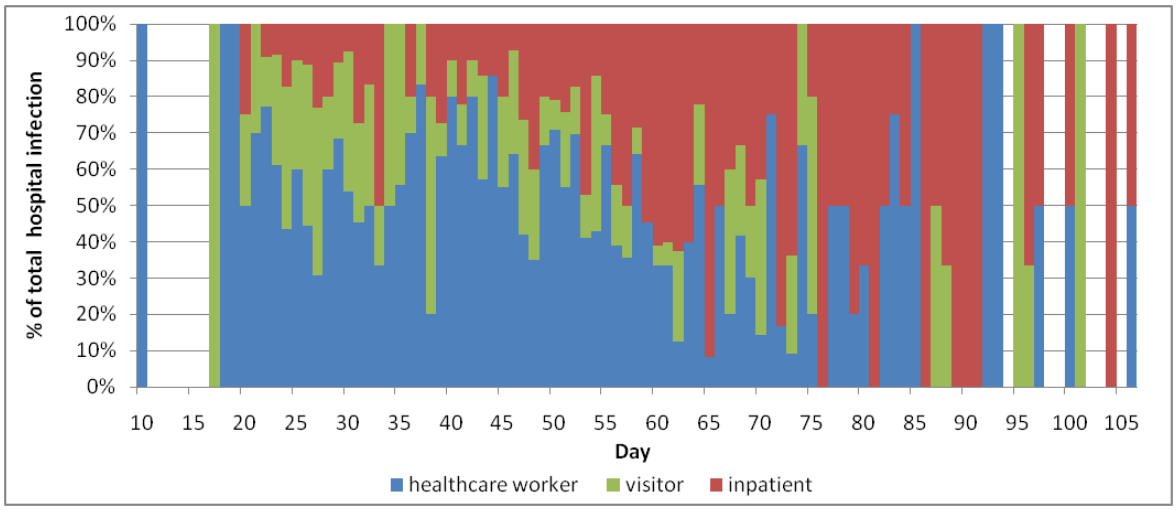


Figure 2.7 Community epidemic curves excluding Amoy Gardens cases. Community (others) includes workplace, elderly home, GP clinic, Metropole Hotel, community and unknown.

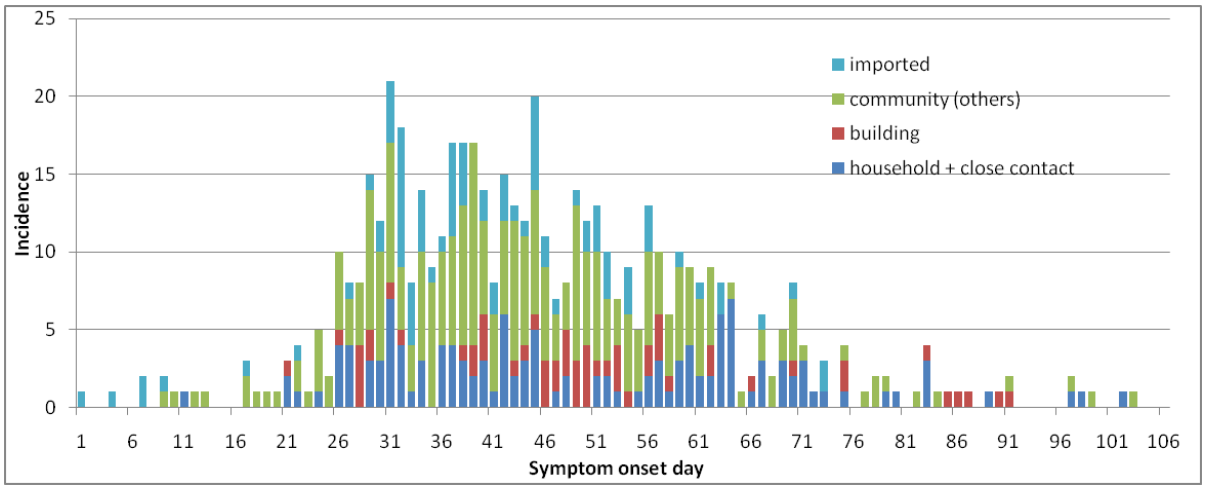
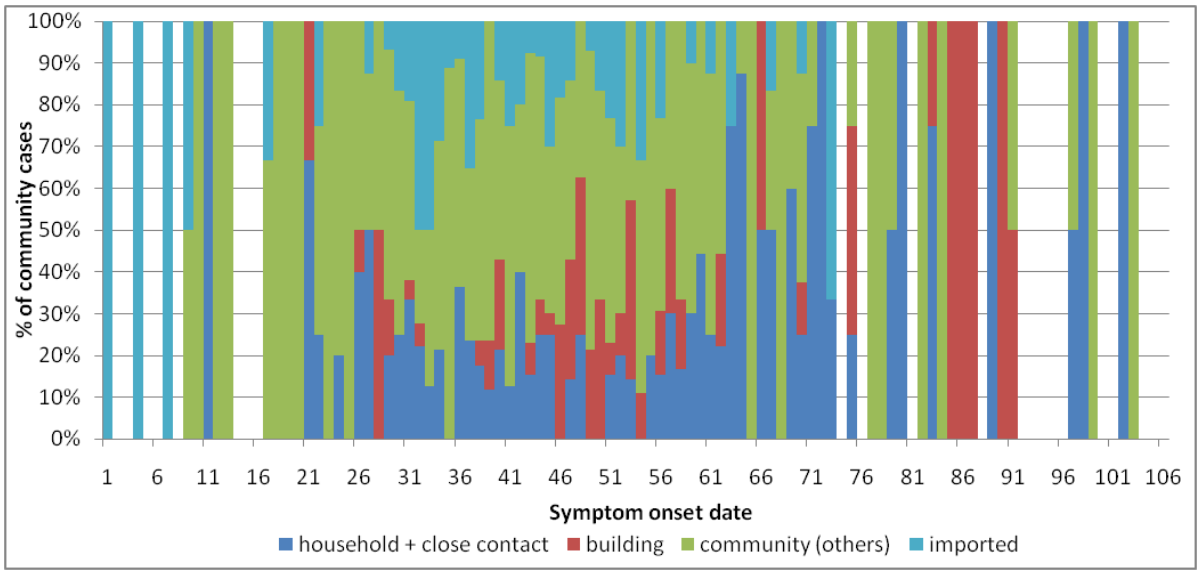


Figure 2.8 Composition of the community infection excluding Amoy Gardens cases throughout the epidemic.



2.2 Geographic Information System (GIS) and demographic data

The GIS data of the residential address of the SARS cases, and the location of the SARS treating hospitals, Amoy Gardens Block E and the Metropole Hotel were provided by Department of Community Medicine, University of Hong Kong. There were 34 (1.9%) missing data on the case GIS data. The digital maps of Hong Kong were obtained from the Hong Kong government. The geographical demarcation system used in the first half of this study was the tertiary planning unit (TPU) in 2001, which divides Hong Kong into 292 TPUs. TPU is for planning purposes and contains areas which did not actually exist yet. A modified TPU map was created by removing the non-existing areas. Centroids of the TPU were projected using ArcGIS software version 9.0. Restricted by the spatial resolution of the trip data available, the second half of the study which modelled the movement patterns of the Hong Kong population using gravity models, and which was subsequently incorporated into the spatial Wallinga and Teunis method in chapters 5 and 6, was based on the District Council District (DCD) - referred as “district” in the rest of the study. There are 18 districts in Hong Kong.

Demographic data for Hong Kong residents by TPU and by district used in this study were obtained from the Hong Kong 2001 population census. When breakdowns of population by age group with small counts were not available, projection was made based on that of the total population distribution.

2.3 Hospitalisation data

There are three sets of hospitalisation data from two datasets. The SARSID integrated dataset contains two sets of hospitalisation data including the name of hospital, admission and discharge dates, collected by the Hospital Authority (HA) and Department of Health (DH) respectively (referred as the HA and DH data). The hospital episode dataset contains the name of hospital, admission and discharge dates of all episodes of hospitalisation related to SARS for the cases (referred as the hospital episode data). It was derived by the University of Hong Kong, inferred by the changes in the hospitalisation data between the regular update of the case data during the epidemic, because the hospitalisation record of HA was not accessible. In the hospital episode data set, some cases have multiple entries because of hospital transfer for treatment or convalescence, or because of changes in hospitalisation protocols during the outbreak (Severe Acute Respiratory Syndrome Expert Committee 2003). In the SARSID dataset both the HA and DH data only have a single entry for each case. The HA data was extracted by health officials during the outbreak from the full hospitalisation data of HA, apparently using the first admission date and the last discharge date of the cases, though the dates might refer to different hospitals. The HA data only contains hospital stays in the public hospitals, which are managed by HA. The DH data was collected using questionnaires administered to the cases or their family if the patient was too ill, and from the follow-up of the cases. It contains data on hospital stays in private hospitals if they were provided by the interviewees. There were inconsistencies among these three sets of data because of the ways they were collected and derived. I designed the following mechanism for data validation and extraction of the three sets of hospitalisation data.

The hospital episode dataset was treated as the base data, because it contains the most detailed information which is not available in the other dataset. On an individual case level, the hospital, dates of admission and discharge were compared among the three sets of data. In dealing with inconsistencies between the datasets, the following guidelines were applied to revise the base data:

1. Entries in private hospital from the DH data were inserted as the first hospitalisation episode. All of these entries have admission and discharge dates before the first public hospital admission in the HA data and the hospital episode data, and the transfer of SARS cases is known to be from private to public hospital only, but not the other way round (Severe Acute Respiratory Syndrome Expert Committee 2003).

2. In dealing with inconsistencies between the DH and HA data on public hospital stays, the HA data is considered to be more reliable because it was collected by their computerised patient record system, while the DH data may subject to reporting errors.

3. For the cases whose episode data is totally missing, the HA data is applied.

4. Hospital episodes in convalescence hospitals and for convalescent purpose were excluded because the cases were no longer infectious during that period.

All three sets of the data take the date of death of the deceased cases as the date of discharge.

Chapter 3 - Spatial Distribution and Cluster Analysis

In the last chapter, I validated the probable source of infection data and other datasets used in this study. In this chapter, I focus on exploring the spatiotemporal patterns and clusters of the epidemic in overall, hospital and community levels, based on the probable source of infection data.

3.1 Objectives

The objectives of this chapter are:

(1) To acquire understanding of the spatial distribution of SARS cases and clustering of these cases
(2) Investigate the potential effects of hospital transmission on community transmission
(3) Evaluate the application of Kulldorff's prospective scan statistic in real-time during an epidemic to detect spatiotemporal clusters in the community

I first conduct an exploratory analysis of the spatial patterns of the outbreak, then detect and investigate the spatial clusters seen in the epidemic, with a focus on hospital and community transmission. Further investigation on the potential interactions between hospital and community infection are performed by statistical modelling methods.

In the evaluation of the prospective scan statistic, I focus only on the community cases because hospital outbreaks are usually detected more easily and promptly by unexpected increase in reported illness of hospital staff, as observed in the SARS outbreak (Severe Acute Respiratory Syndrome Expert Committee 2003). Community outbreaks, on the other hand, usually need epidemiological information to establish any potential connection between the cases. Also formation of localised public health control measures requires the information on the focal point of outbreak and the spatial range affected, which can be attained from the spatial cluster detection method.

3.2 Method

3.2.1 Exploratory analysis

3.2.1.1 *Distribution patterns and spread*

I illustrate the overall spatial patterns and weekly spatiotemporal development of the overall SARS epidemic by maps of incidence rate by TPU (Tertiary planning unit). Trends in the latter are interpreted by disaggregation by transmission source (hospital, community and the Amoy Gardens event) to explore the relationship between epidemic growth and spatial spread. Maps of the spatial

distributions of hospital and community infected incidence rates by TPU are also studied to explore the characteristics of the individual transmission sources. For hospital infection, the spatial distribution of the cases of the key hospital outbreaks – PWH (Prince of Wales Hospital), AHN (Alice Ho Miu Ling Nethersole Hospital) and PMH (Princess Margaret Hospital), and the hospital infection subgroups (healthcare worker, inpatient and visitors) were compared.

I define a TPU as infected if there was at least one SARS case living in the TPU, and the time that a TPU is infected as the symptom onset date of the first case living in that TPU. I then studied the trends in between-TPU spread by time.

3.2.1.2 *Spatiotemporal clusters*

I retrospectively detected the spatiotemporal clusters of the complete epidemic using Kulldorff's spatial scan statistic with a circular window (Kulldorff 1997) through the SaTScan™ software version 7.0 (www.satscan.org). In the space-time scan statistic method, a cylindrical window is scanned across the space of the study region , and time by the height of the cylinder. The window varies in volume during the scan, up to a pre-specified maximum value in percentage of population at risk or Cartesian units for the circular base, and up to a percentage of the study time period for the temporal window. At each location, the numbers of observed and expected observations inside the cylinder are noted. Using the likelihood ratio test, the alternative hypothesis that there is an elevated risk within the cylinder as compared to outside, conditional on the total number of cases observed in the study region is tested, and adjustment of inhomogenous population distribution or other covariates in the expected cases is allowed. Under the Poisson assumption for the number of cases in each location, the likelihood function for a specific cylinder is proportional to:

$$\left(\frac{x}{E(x)}\right)^{x}\left(\frac{N-x}{N-E(x)}\right)^{N-x}$$

Where *N* is the total number of observed cases, *x* is the observed number of cases within the cylinder and *E(x)* is the covariate adjusted expected number of cases within the window under the null hypothesis. Monte Carlo simulation is used to perform the hypothesis testing, and the cylinder with the maximum likelihood is the most likely cluster and the p-value is calculated. Besides the most likely (primary) cluster, secondary clusters are detected and ordered according to their likelihood ratio test statistics.

In this study the spatial scan statistic was applied on the TPU incidence which was geographically referenced by the centroid of the TPU. Cases of ongoing inpatients were excluded in the cluster detection as they had stayed in the hospital throughout the epidemic, making the residential

location a misleading measure of potential infection source. A Poisson distribution was assumed for the number of cases in the TPUs, and the number of expected cases in the TPU was adjusted by the age-sex distribution of the population in the TPU using indirect standardization.

In setting the maximum circle sizes, I used relatively small circles compared to the default settings because large clusters can be captured as a combination of small clusters, but big maximum circle can lead to capturing false positive areas surrounding the truly positive areas. The maximum circle size was set as 5% of the population at risk (ie. 201,074 of 6,702,480 population), compared to the default setting of 50%. The maximum circle radius was set as 2 km, corresponding to a circle area of 12.6 km^2. This was justified given the total size of Hong Kong is 1068 km^2, and 65% of TPUs are smaller than 3 km^2, and 90% smaller than 10 km^2. The maximum time window was 50% of the study period, which is the default setting. This allowed spatiotemporal clusters that lasted for half of the time of the epidemic to be detected. Monte Carlo simulation was repeated 999 times.

The spatiotemporal clusters were identified for the overall epidemic, and then for hospital and community transmission. I also investigated the composition of the source of infection of the cases in the statistically significant clusters (p-value≤ 0.05) in the overall epidemic, and tested if there is any difference in the proportions of cases in different sources of infection being captured in the clusters. This is to test if cases of certain infection sources are more locally clustered than the others.

3.2.2 Interactions between hospital and community transmission

I first explored if the community spatial clusters were related to the location of hospitals with SARS cases, and to the hospital clusters, by comparing the location and timing of the hospital and community spatiotemporal clusters detected to explore any interaction between them.

I then investigated the risk factors that were associated with the size of community infections in the TPUs. The incidence of community cases in the TPUs were modelled with risk factors related to both the local demographic characteristics of the TPUs and the hospital transmissions of TPUs. These included:

(a) population size of the TPU (as the offset variable)
(b) population density of the TPU
(c) age distribution of the TPU (proportion of the age groups of 0-24, 25-49, 50-74, 75+ in the TPU population)
(d) whether there is a SARS treating hospital in the TPU
(e) the geographical distance of the TPU to the nearest SARS treating hospital (derived as the Euclidean distance between the Cartesian co-ordinates of the centroid of the TPU and the hospital)

(f) the number of hospital infection in the nearest SARS treating hospital

(g) the numbers of hospital infected cases living in the TPU

(h) the number of hospital infected healthcare workers living in the TPU

(i) the number of infected inpatients discharged during the incubation period of SARS living in the TPU

(j) the number of infected hospital visitors living in the TPU

(k) the symptom onset day of the first case in the TPU to reflect the control measures implemented by that time

Six statistical models were applied to the model the community case count by TPU, including:

1) quasi-Poisson model
2) negative binomial model
3) zero-inflated Poisson (ZIP) model
4) zero-inflated negative binomial (ZINB) model
5) hurdle Poisson (HP)
6) hurdle negative binomial (HNB) model

The generalized linear models with quasi-Poisson and negative binomial distributions were used to adjust for the overdispersion in the TPU incidence. The zero-inflated (Mullahy 1986; Lambert 1992) and hurdle models (Mullahy 1986) were used to capture the high zero counts in the TPUs incidence, because 61.5% of the populated TPUs had no community cases in the SARS epidemic. Both models are composed of a zero component, denoted as $f_{zero}(0;z,\gamma)$, and a count component, denoted as $f_{count}(y;x,\beta)$, where y is the count variable, x and z are the covariate vectors, and γ and β are the corresponding vector of parameters. The main difference between the two models is how the zero count is modelled. The zero-inflated model can be expressed as:

$$f_{zeroinfl}(y;x,z,\beta,\gamma)=\begin{cases} f_{zero}(0;z,\gamma)+(1-f_{zero}(0;z,\gamma))\,f_{count}(0;x,\beta) & \text{if } y=0 \\ (1-f_{zero}(0;z,\gamma))\,f_{count}(y;x,\beta) & \text{if } y>0 \end{cases}$$

therefore in the public health concept, it can be considered that the zero counts can be from both the at-risk and not-at-risk populations (Rose, Martin et al. 2006). Whereas hurdle models consist of a left truncated count component for the positive counts, and a right-censored zero hurdle component that models the outcome of zero or non-zero. It can be written as:

$$f_{hurdle}(y;x,z,\beta,\gamma)=\begin{cases} f_{zero}(0;z,\gamma) & \text{if } y=0 \\ (1-f_{zero}(0;z,\gamma))\dfrac{f_{count}(y;x,\beta)}{1-f_{count}(0;x,\beta)} & \text{if } y>0 \end{cases}$$

and it can be conceptualized as the zeros are only from the at risk population.

In this study, I used both the Poisson and negative binomial distribution for $f_{count}(y;x,\beta)$, and the logistic regression model for $f_{zero}(0;z,\gamma)$. Model parameters are estimated by maximum likelihood method. Covariates used in the zero and hurdle components of the zero augmented models were the same as in the count model, except the symptom onset date of the first case in the TPU was excluded. Amoy Gardens and imported cases were excluded in the analysis. The models were compared using the Akaike's information criterion (AIC), Bayesian information criterion (BIC), and the Vuong statistic (Vuong 1989).

3.2.3 Real-time cluster detection

I applied the Kulldorff's prospective scan statistic (Kulldorff 2001) to the data available by the start of each day of the epidemic to examine the spatiotemporal clusters detectable at each time point. The prospective scan statistic differs from retrospective scan statistic in the following ways:

(a) Denote the study period covered by the data as (x_1,x_t), and the period scanned by the varying cylindrical window to detect any cluster to be (s_1,s_2). The retrospective scan statistic detects any cluster of time dimension $x_1 \leq s_1 \leq s_2 \leq x_t$, whereas the prospective scan only detects clusters of time dimension $x_1 \leq s_1 \leq s_2 = x_t$. Therefore the latter only identifies clusters with excess risk still present at day x_{t-1}, which can be emerging clusters or long-existing clusters of moderate level of excess risks.

(b) Repeated application of retrospective scan statistic periodically gives rise to multiple comparison issue in the statistical inference, leading to false positive errors. The prospective scan statistic overcomes this by taking the likelihood for the real data set as the maximum over all cylinders with $s_2 = x_{t-1}$ as described above, but in the random data sets, the likelihood is maximized over all cylinders used in previous analyses in addition to the current one, ie. $s_1 \leq s_2 \leq x_{t-1}$.

(c) When the epidemic is still ongoing, prospective scan statistic being run on day k uses the most up-to-date dataset covering the study period (x_1,x_{k-1}). Whereas for the retrospective scan, because of the multiple comparison issue, it is usually run on dataset of completed event, covering the study period (x_1,x_2) where x_2 is the last day of the epidemic, or the defined end day of the study and $x_2 \leq x_{k-1}$. Because both scan statistics are conditional on the corresponding total number of cases, the thresholds used to define "excess risk" in the two methods are different.

I applied the same settings of maximum circle size and temporal window as in the retrospective scan, and the inference was set to adjust for the previous prospective scan since the start of the

study period. The tests were run in the SaTScan™ software 102 times (days) to cover the whole period of the epidemic.

Results of the prospective scans were compared to those detected from the data on the community cases of the full epidemic using the retrospective basis mentioned in Section 3.2.1.2. Though they are not directly comparable because the scan statistic is conditional on the total number of cases observed, it would still be insightful to compare between the two when interpreted with caution. The comparison was made on the TPU level as the prospective clusters changed over time, making comparison on the cluster level difficult. Information on the high risk locations identified by the outbreak data management system (e-SARS, MIIDSS and SARS-CCIS) were not openly accessible, therefore it was not possible to compare these with the identified clusters. The cluster involving the TPU where Amoy Gardens is located was studied further, and compared to the timeline health officials in Hong Kong was alert to the outbreak.

3.3 Results

In this section, I first present the results of the overall epidemic, and then on the hospital and community transmission. After that I will present the findings on the interactions between hospital and community transmission, and the evalualtion of real-time cluster detection method.

3.3.1 Overall epidemic

3.3.1.1 Spatiotemporal distribution and TPU spread

The Hong Kong SARS cases were distributed in 143 TPUs (57.7%), and the TPU incidence rates varied greatly (Figure 3.1) compared to the average Hong Kong incidence rate of 26.2 per 10^5 population. Three TPUs were of extremely high incidence rate, located in the Sai Kung district (TPU 833), Yau Tsim Mong district (TPU 211), and Kwun Tong district (TPU 291) respectively. Except TPU 291 where the Amoy Gardens is located, the high incidence rates in other two coastal TPUs were due to small population sizes (TPU 833: 2 cases in 25 people; TPU 211: 3 cases in 14135 people).

Early cases contributed significantly to spread between the TPUs (Figure 3.2), with the first 20% cases contributing 71.0% of the total TPU spread (65.7% excluding the Amoy Gardens cases), the first 30% of the cases 78.5% of the TPU spread (75.8% excluding the Amoy Gardens cases), and the first 40% of cases 79.7% of the TPU spread (77.5% excluding the Amoy Gardens cases).

Figure 3.3 shows the trend of TPU spread by symptom onset date throughout the epidemic, which could be classified into 4 stages according the rate of TPU spread. Stage 1 was the initial growth of the epidemic which occurred before the PWH outbreak happened, defined as between day 1 to day

Figure 3.1 Spatial distribution of SARS cases in Hong Kong (TPU incidence rate per 100,000 population).

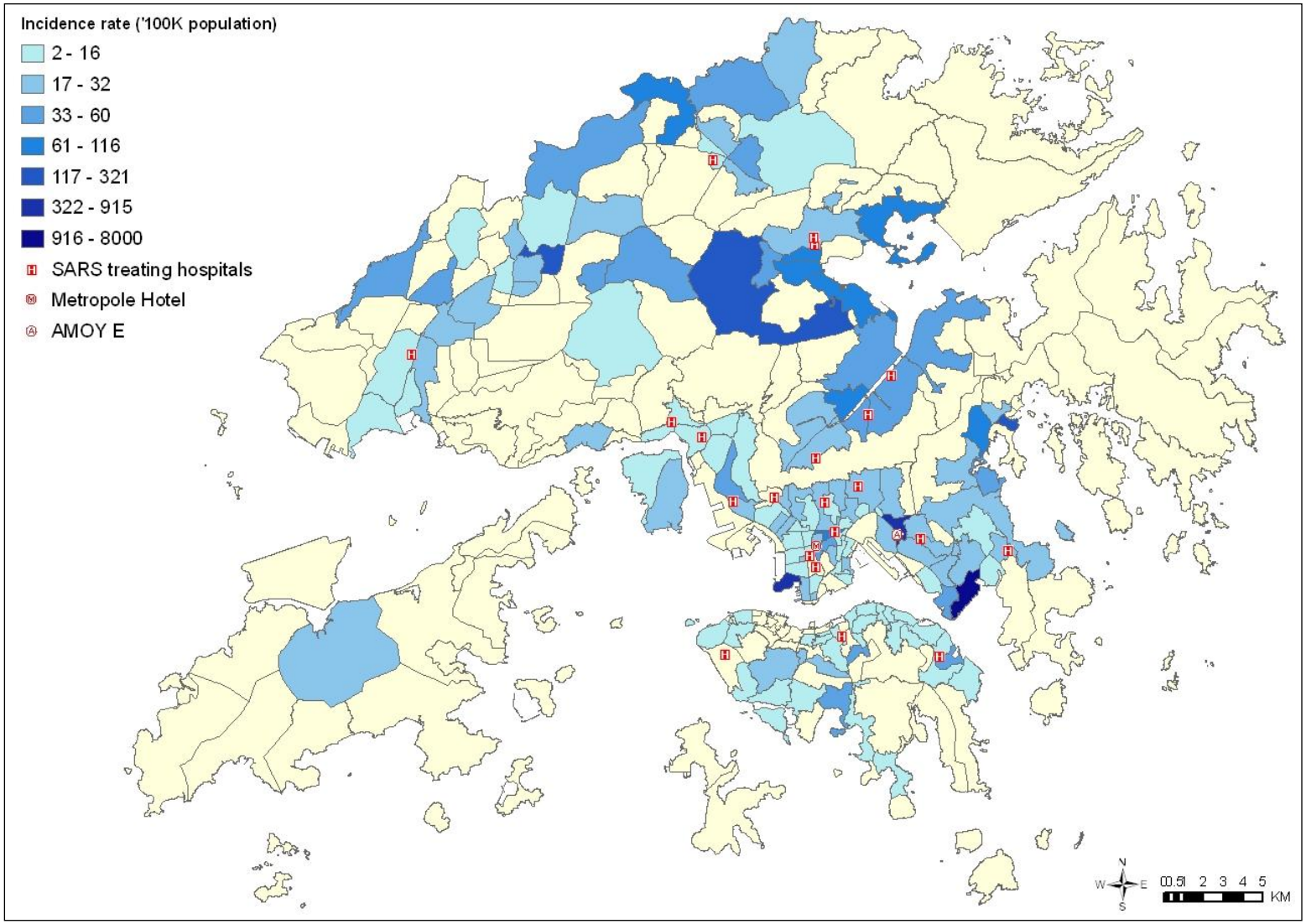
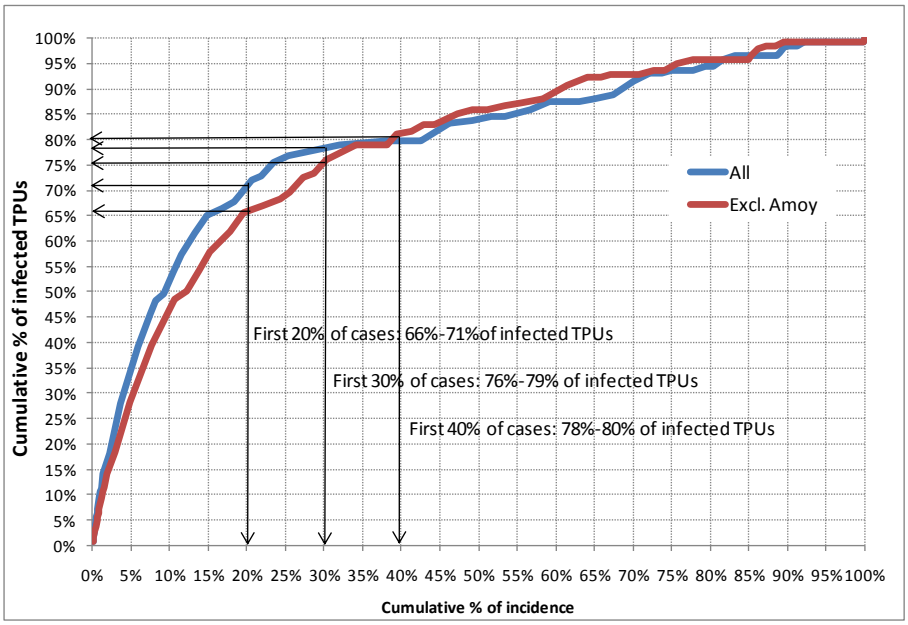


Figure 3.2 TPU spread by incidence (both in cumulative percentages) for: (i) all cases (whose residential TPUs could be identified), (ii) excluding the Amoy Gardens cases.



19, during which the first 1.1% of total cases had symptoms onset and first 11.1% of the TPUs were infected. Stage 2 was the period of rapid growth in TPU spread between day 20 and 33, coinciding with the prime period of the PWH outbreak, and accounted for the next 18.0% of the cases and 58.0% of the TPU infected; stage 3 was characterised as the decelerated TPU spread between day 34 and 60, which captured the peak of the Amoy Gardens outbreak, and accounted for the next 64.2% of incidence and 27.3% of the TPU spread. Stage 4 was the phase of relatively stagnant TPU spread between day 61 and 106, corresponding to the last 15.4% of the cases and 3.5% of the TPU spread. TPUs infected in the later stages of the epidemic had significantly lower mean population densities and also population size [both ordinary least square linear regression p-values=0.010, Figure 3.4 (a-b)], and uninfected TPUs tend to have small population size and low population densities (Figure 3.5), suggesting some form of the hierarchical spread of the epidemic.

The spatiotemporal development of the epidemic is illustrated by the weekly TPU incidence rates in Figure 3.6 (a-c), during the first 3 weeks of the epidemic, which corresponded largely to stage 1 of the TPU spread, the cases were scattered in different TPUs over the territory. In the following 2 weeks [Figure 3.6(d)-(e), week 4 and 5], which was mainly under the stage 2 of the TPU spread, there were sharp increases in both the incidence and TPU spread fuelled by the PWH outbreak (Figure 2.4) and then the increase in community transmissions (Figure 2.3), and the cases were more clustered in distribution than the previous weeks. Week 6 [Figure 3.6(f)] was under the stage 3 of the TPU spread, during which the Amoy Gardens outbreak reached its peak, and there were more community than hospital cases (Figure 2.3) as the PWH outbreak was subsiding (Figure 2.4). Cases

Figure 3.3 Infected TPU and incidence by source of infection (both in cumulative percentages) by symptom onset day.

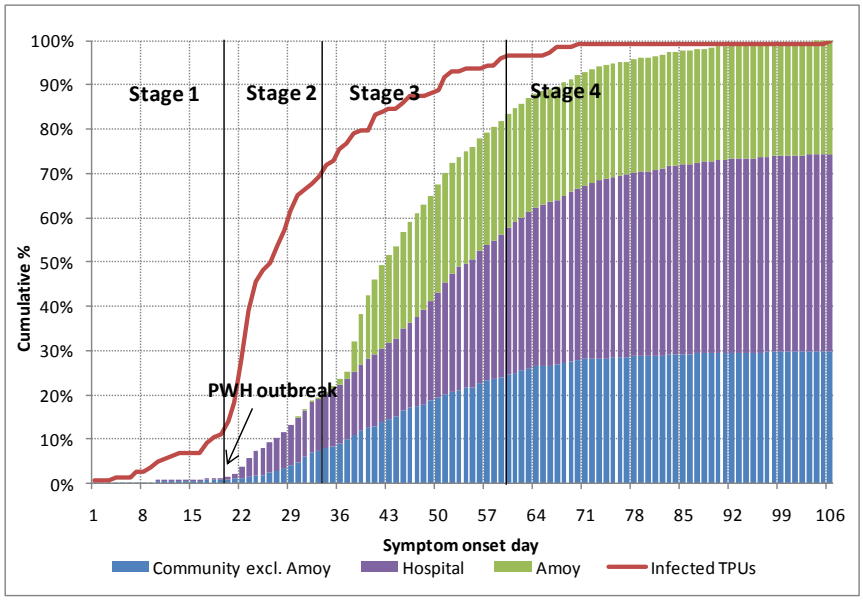


Figure 3.4 Boxplot of the (a) population density, (b) population size of the new TPUs infected by stages. Black cross: mean; red line inside the box: median.

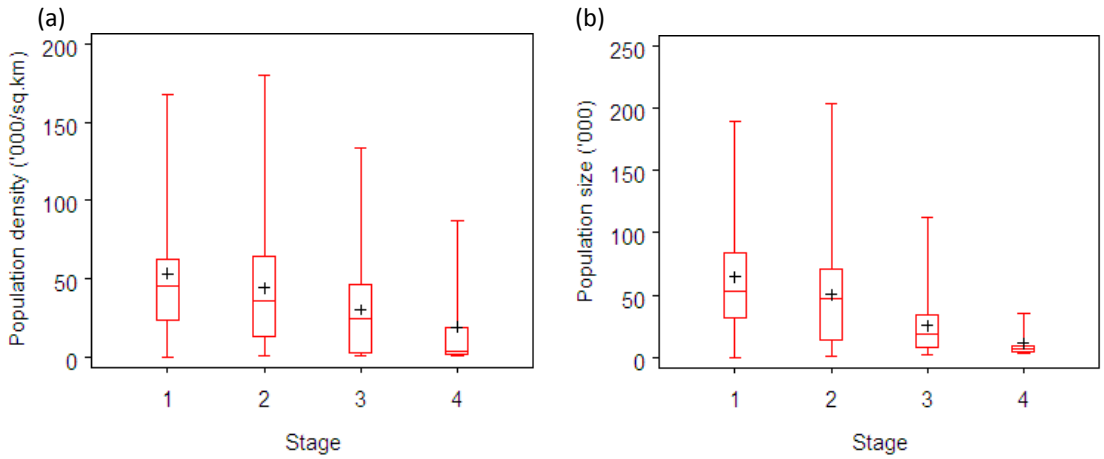
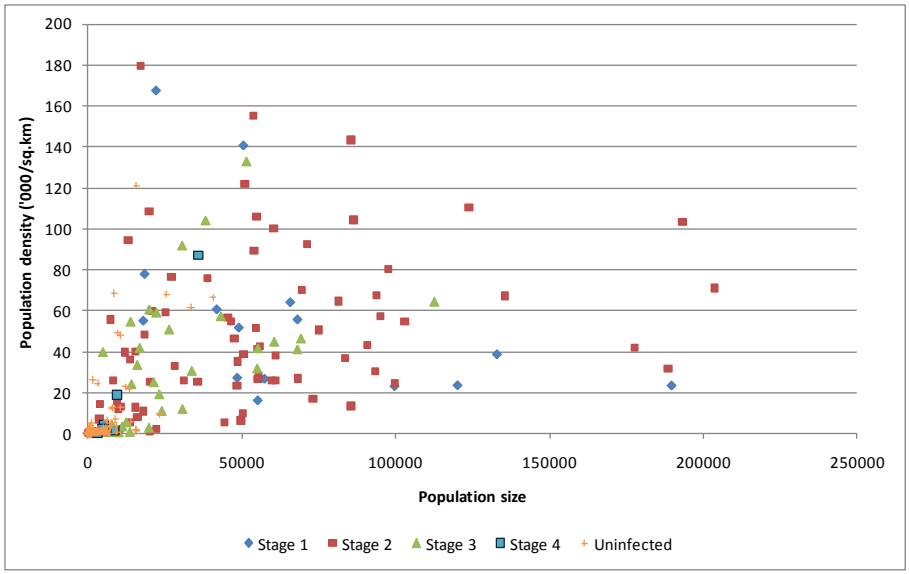


Figure 3.5 Scatter plot of the population size and population density by stage of infection.



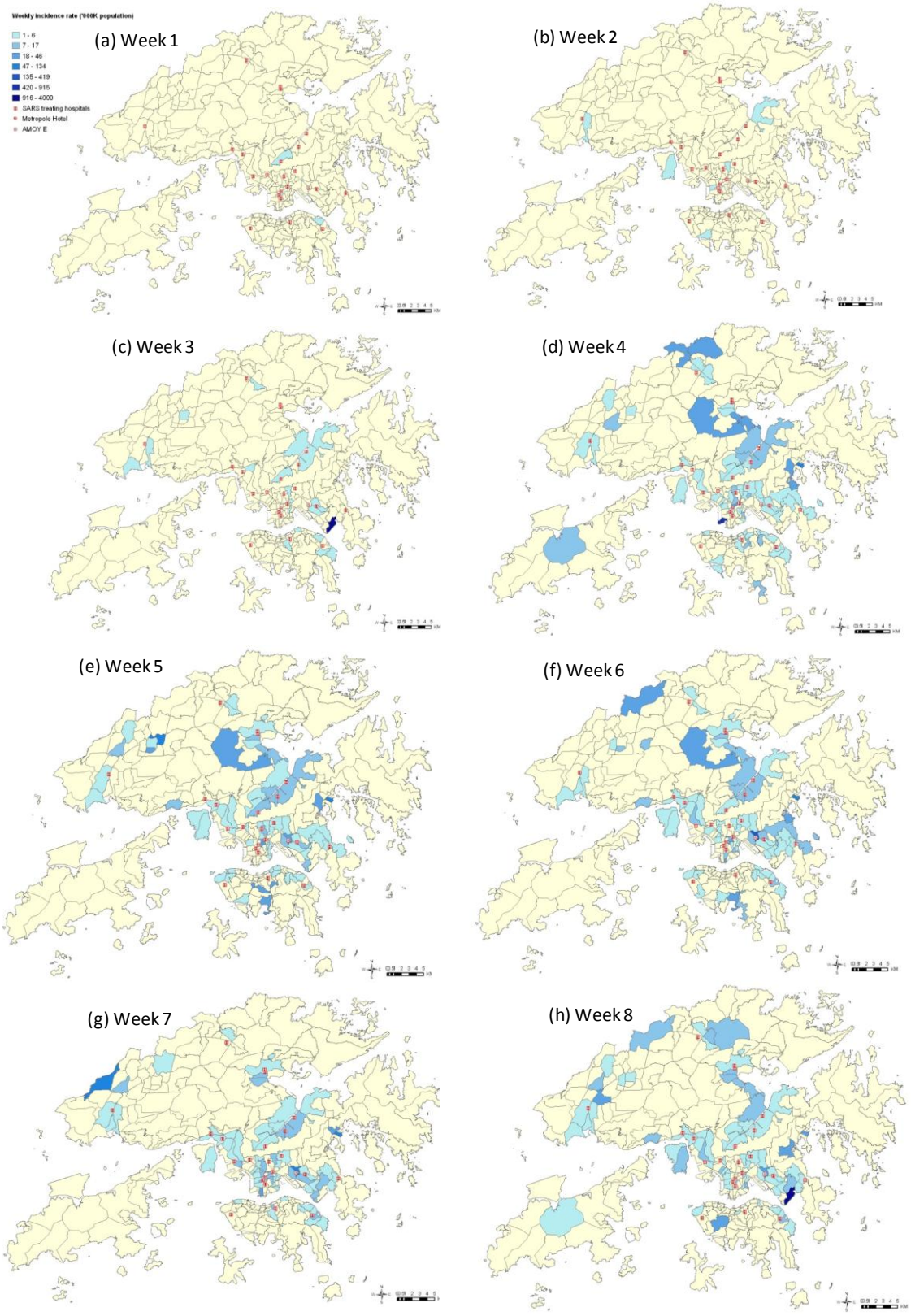
were located in some of the previously infected TPUs, and also in TPUs contiguous to the previously infected TPUs. The next 2 weeks were still under stage 3 (Figure 3.6 (g)-(h), week 7 and 8), during which the AHN, PMH and UCH outbreaks peaked, while the Amoy Gardens outbreaks was subsiding (Figure 2.3), and some TPUs in the Tai Po and North districts were infected. The following 2 weeks (Figure 3.6(i)-(j), week 9 and 10) was the peaks of the TPH and TMH outbreaks, and new TPU infections were found in the Yuen Long district, while considerable cases appeared in the Tai Po district. Stage 4 of the TPU spread commenced during this period, and decrease in weekly incidence and TPU spread was reflected in Figure 3.6(i). The next 6 weeks (Figure 3.6 (k)-(p), week 11-16) showed the ending of the epidemic (Figure 2.3) and case distribution was scattered.

3.3.1.2 Spatiotemporal clusters

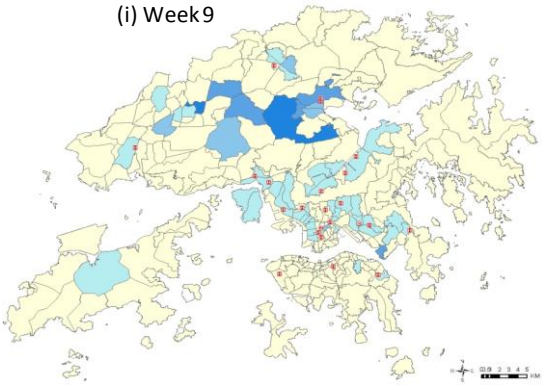
Twelve statistically significant spatiotemporal clusters based on all SARS cases (excluding ongoing inpatients) were identified and mapped in Figure 3.7. The clusters captured 48.8% of the cases where residential TPUs were identifiable. The TPU where the Amoy Gardens is located was identified as the primary cluster (cluster 1, p-value=0.001). For the other 11 secondary clusters (p-values ranged from 0.001 to 0.009), those surrounding the PWH and SH (clusters 3, 5, 6, 11, 12) formed an area of clusters, and cluster 2 was near AHN and TPH, and almost all of them had a high proportion of hospital infections of 58.8%-72.7%, except cluster 11 at 44.0% (Figure 3.8 and Table 3.1). Five other clusters (clusters 1, 4, 7, 9, 10) had a high proportion of community infection cases, and cluster 10 was adjacent to the Amoy primary cluster.

The temporal information on the clusters is given in Figure 3.9. 75% of the clusters started within 8 days from day 20, which was the initial period of the PWH outbreak. Only clusters 1, 9 and 10 started later that this period. Cluster 2, which is near the AHN and TPH, was the longest cluster in duration which lasted for 51 days. Cluster 8, a pure hospital infection cluster, was the shortest cluster which lasted for only 1 day. Table 3.1 shows the composition of potential source of infection of each cluster. Cluster 2 had 37.9% of AHN and 11.4% of PWH cases in the cluster, and this explained the long duration of the cluster given the timing of the PWH and AHN outbreaks (Figure 2.4). Clusters 3, 5, 6, 8, 11 and 12 had high PWH cases of 47.7%, 52.6%, 100.0%, 28.0% and 29.4% in the clusters respectively. Cluster 7 had 19.2% of the cases infected from PWH, and 23.1% from the general community where specific source could not be identified. Cluster 1 and 10 had 98.4% and 39.3% of Amoy Gardens cases, and clusters 4 and 9 had 91.7% and 70.6% of the cases sourced from outbreaks in buildings in the clusters.

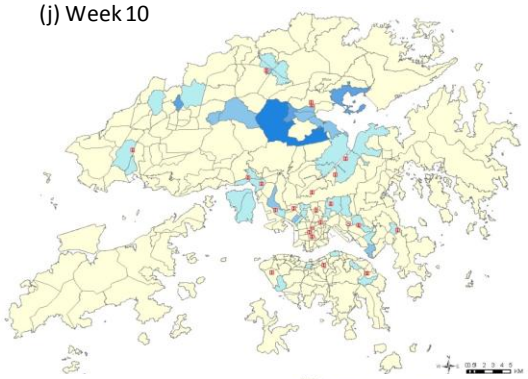
Figure 3.6 (a)-(p): Weekly incidence rate (per 100,000 population) by TPUs (from left top to left bottom, then right top to right bottom).



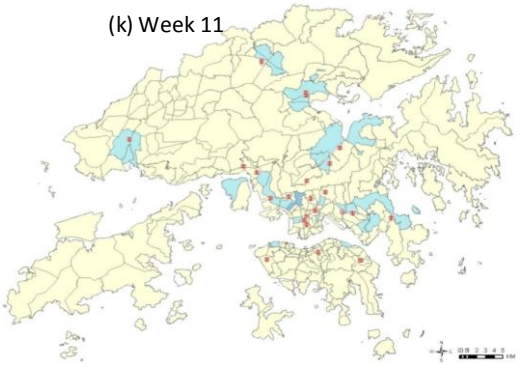
(i) Week 9



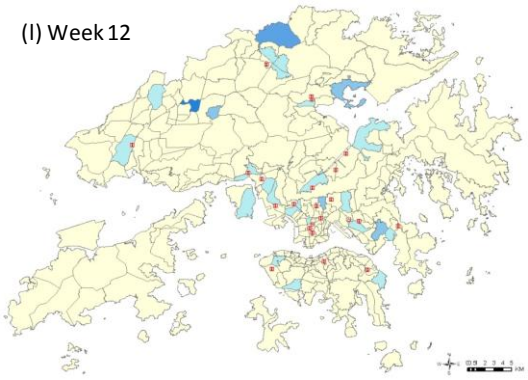
(j) Week 10



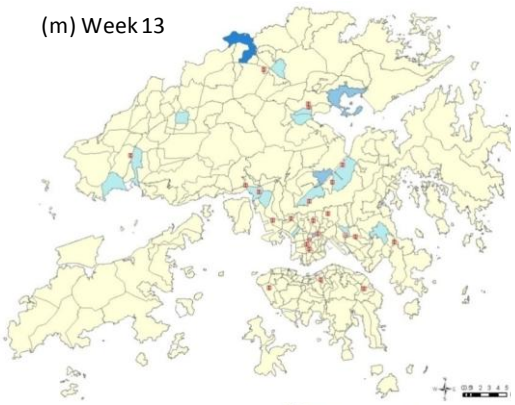
(k) Week 11



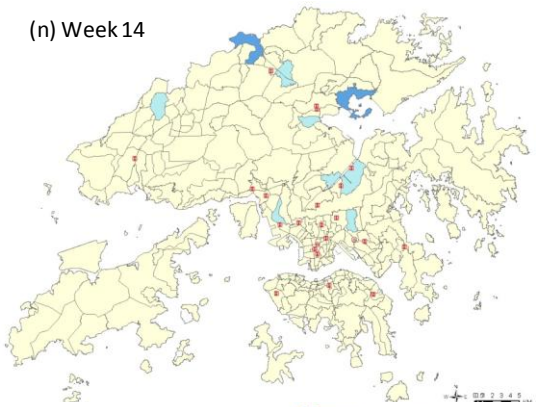
(l) Week 12



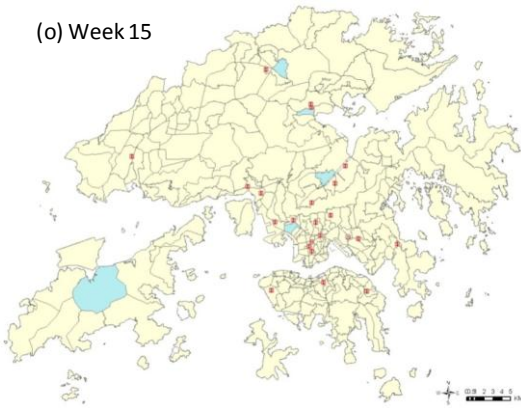
(m) Week 13



(n) Week 14



(o) Week 15



(p) Week 16

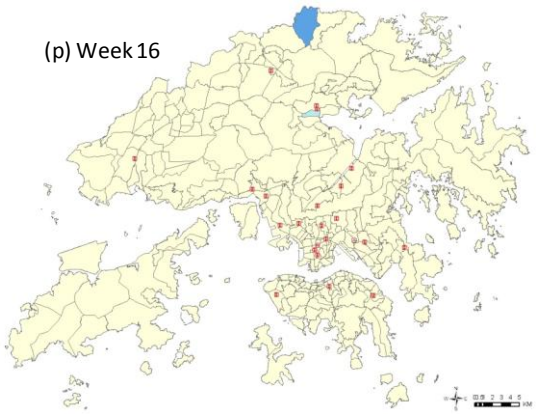


Figure 3.7 Spatiotemporal clusters of all cases excluding ongoing inpatients numbered in descending order of statistical significance (all with p-values < 0.05). The number of infections in the SARS treating hospitals was shown as circles (in proportion to the number of infections).

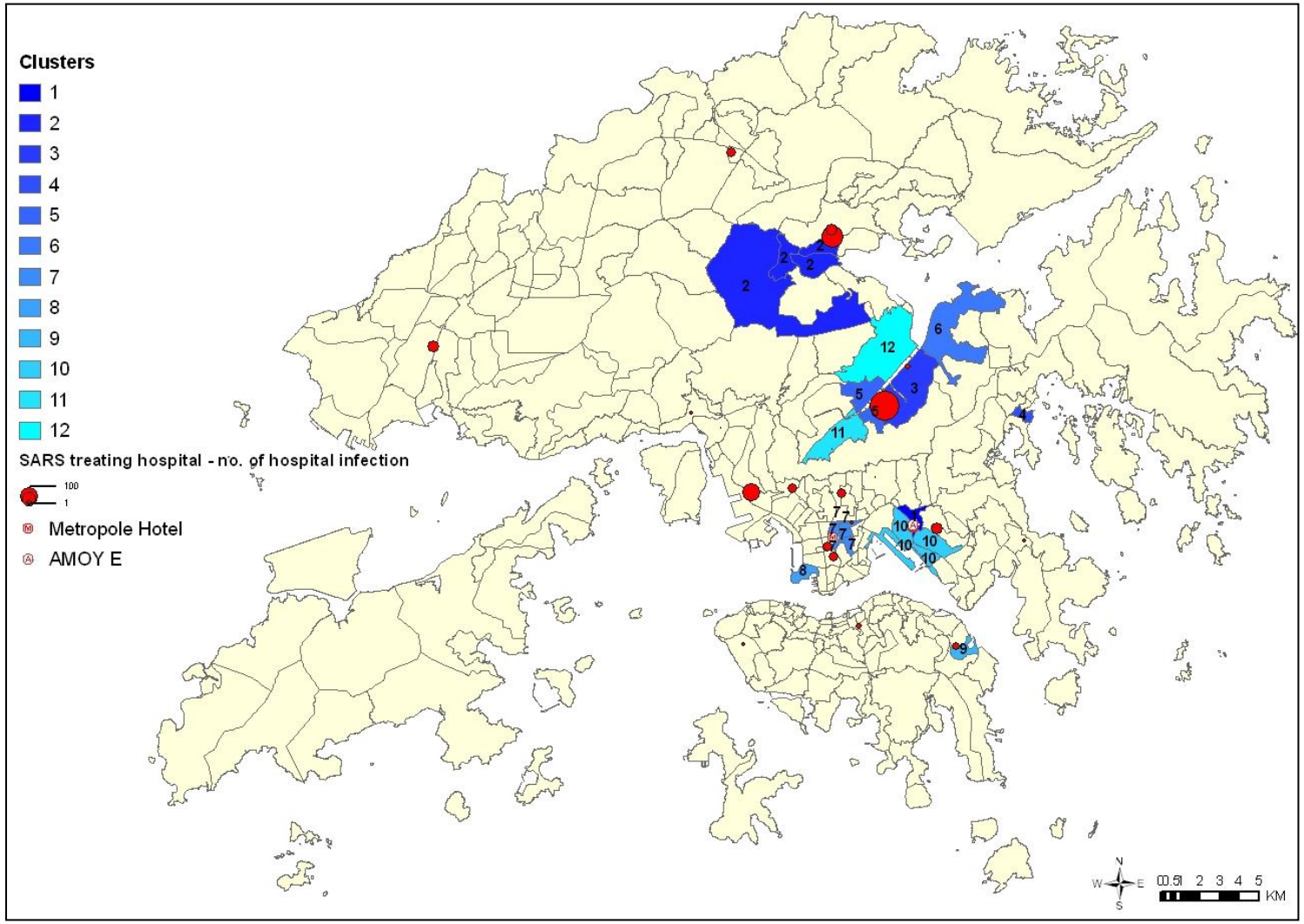


Figure 3.8 Composition of the clusters (hospital or community infections).

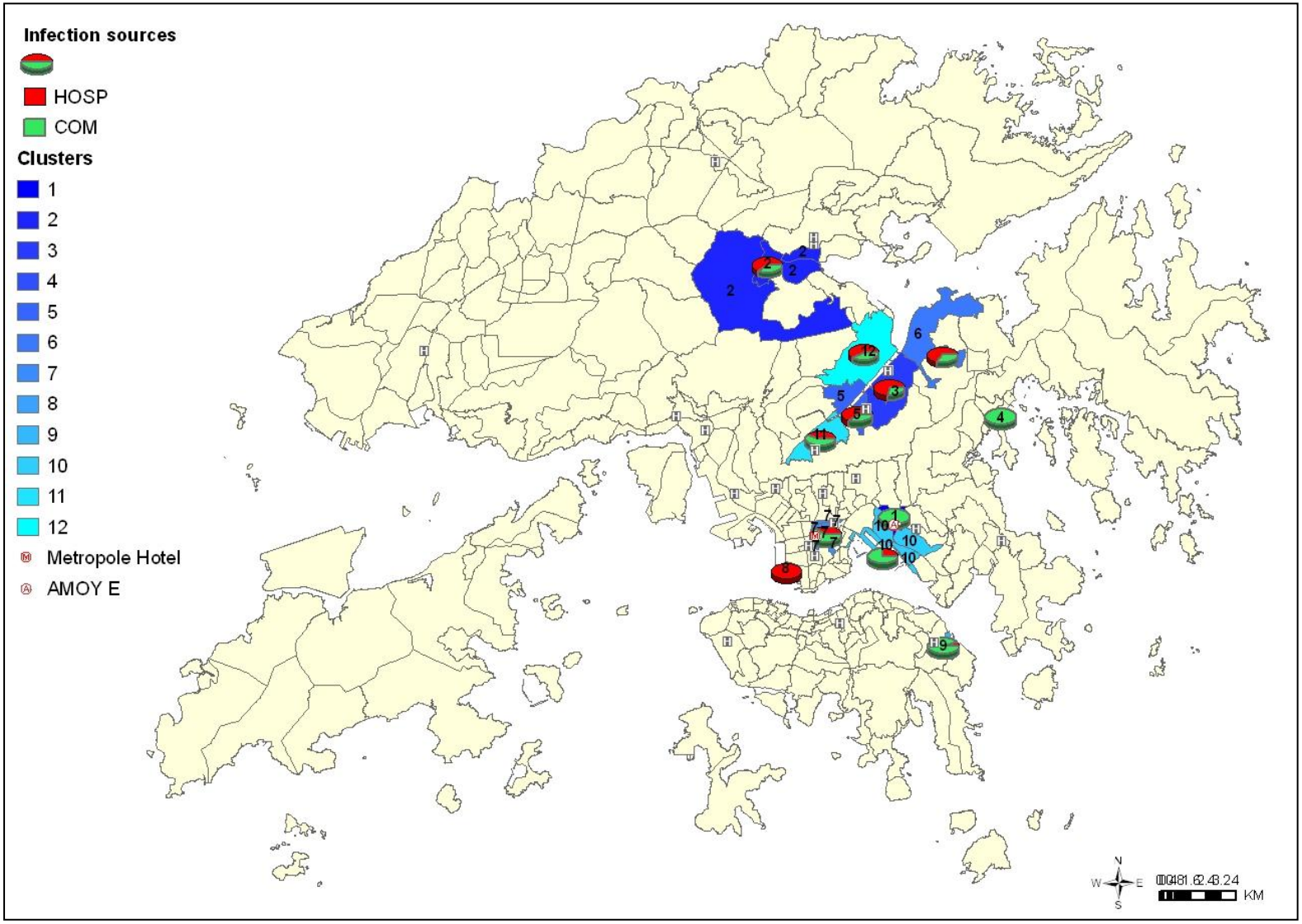


Figure 3.9 Start and end days and the duration (figure in the bar) of the spatiotemporal clusters of all cases (excluding ongoing inpatients).

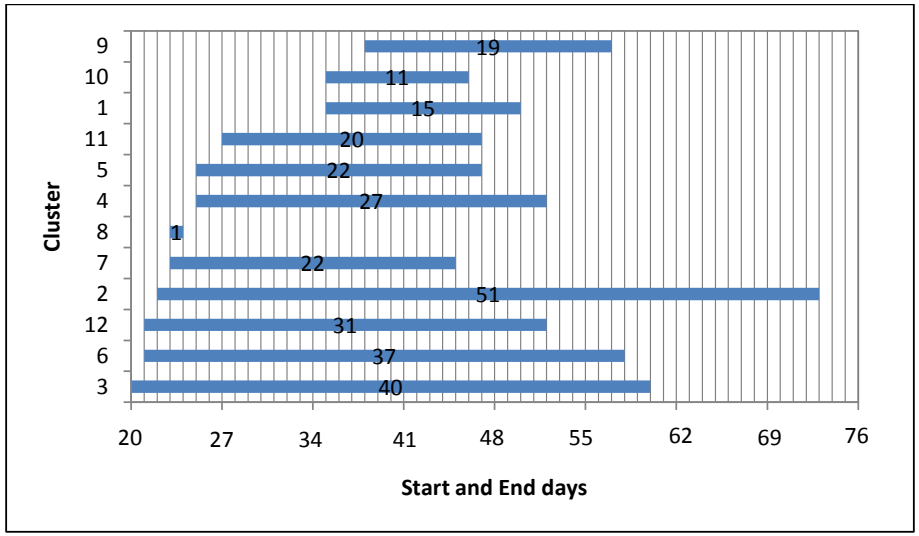


Table 3.1 Composition of the source of infection (in percentage of total number of cases in the cluster) of the spatiotemporal clusters of all cases (excluding ongoing inpatients). (Figures greater than 25% are in bold).

	hospital (%)													community (%)									
cluster	ahn	bth	cmc	ndh	pmh	pwh	pyn	qeh	sh	sph	tph	uch	uh	subtotal	household	close contact	elderly home	GP clinics	community	building	unknown	imported	Amoy	subtotal
1						0.26		0.26				0.26		0.78	0.52				0.26			0.00	**98.43**	99.22
2	**37.88**	1.52	0.76		2.27	11.36				0.76	6.06	0.76		61.36	12.88	2.27	4.55	1.52	11.36	2.27	0.76	2.27	0.76	38.64
3	15.91	2.27		2.27		**47.73**						2.27	2.27	72.73	11.36		9.09	2.27			4.55	0.00		27.27
4														0.00					8.33	**91.67**		0.00		100.00
5	7.89					**52.63**		2.63						63.16	18.42			2.63		5.26	2.63	2.63	5.26	36.84
6	12.73	1.82			3.64	**41.82**	1.82	1.82	1.82			1.82		67.27	9.09	7.27			5.45	5.45		3.64	1.82	32.73
7		11.54				19.23		3.85				3.85		38.46	7.69				23.08		11.54	11.54	7.69	61.54
8						**100.00**								100.00								0.00		0.00
9							5.88							5.88					5.88	**70.59**		0.00	17.65	94.12
10						7.14		3.57				14.29		25.00					14.29	3.57	3.57	14.29	**39.29**	75.00
11	12.00	4.00				**28.00**								44.00	24.00	4.00	4.00	4.00	8.00		4.00	8.00		56.00
12	11.76					**29.41**		5.88		11.76				58.82	5.88	17.65			11.76			5.88		41.18

Table 3.2 shows the percentage of cases in each source of infection captured in the spatiotemporal clusters. Amoy Gardens cases had the highest proportion of cases in cluster (89.8%), followed by AHN cases (72.7%), elderly home cases (57.9%), buildings (56.1%), TPH (47.1%), household (46.4%), and PWH (45.7%) cases. 57.6% of the community cases including Amoy Gardens cases were captured in the clusters, which is significantly higher than that of the hospital cases (35.2%, p-value

Table 3.2 Percentage of cases captured in clusters by source of infection.

Sources of infection	Description	% of cases in clusters	number of cases in the source
amoy	Amoy Gardens	89.8%	442
ahn	Alice Ho Miu Ling Nethersole Hospital	72.7%	99
elderly homes	Elderly care homes	57.9%	19
buildings	Residential buildings	56.1%	57
tph	Tai Po Hospital	47.1%	17
household	Household	46.4%	97
pwh	Prince of Wales Hospital	45.7%	223
unknown	Unknown	37.5%	24
close contact	Close contact	33.3%	33
sph	St. Paul Hospital	33.3%	9
GP clinics	General practitioner clinics	31.3%	16
bth	Baptist Hospital	25.0%	32
sh	Shatin Hospital	25.0%	4
qeh	Queen Elizabeth Hospital	24.0%	25
uch	United Christian Hospital	22.0%	41
community	General community	20.1%	174
imported	Imported	19.0%	84
pyn	Pamela Youde Nethersole Eastern Hospital	10.5%	19
cmc	Caritas Medical Centre	7.7%	13
pmh	Princess Margaret Hospital	6.3%	79
ndh	North District Hospital	5.0%	20
workplace (non-hospital)	Workplace (non-hospital)	0.0%	7
kwh	Kwong Wah Hospital	0.0%	18
metropole hotel	Metropole hotel	0.0%	3
qmh	Queen Mary Hospital	0.0%	4
sth	St. Teresa's Hospital	0.0%	3
tko	Tseung Kwan O Hospital	0.0%	4
tmh	Tuen Mun Hospital	0.0%	27
twa	Tusen Wan Adventist Hospital	0.0%	2
ych	Yan Chai Hospital	0.0%	1
uh	Union Hospital	100.0%	1
hospital total		35.2%	623
community total		57.6%	974
community ex. Amoy		30.8%	532

<0.001). However, on excluding Amoy Gardens cases, there was no significant difference in the proportion of cases in clusters between community (30.8%) and hospital cases (p-value=0.270).

3.3.2 Hospital transmission

3.3.2.1 Spatial distribution and clusters

The spatial distribution of the hospital cases is given in Figure 3.11. Among the top five most infected hospitals, PWH, PMH and UCH had 57.6% to 74.1% of cases in healthcare workers, whereas AHN and TPH had 45.6% to 73.7% of inpatient cases. The two coastal TPUs in Kowloon with the highest incidence rates were due to small population sizes of the TPUs (Figure 3.11 and Figure 3.13). Ten spatiotemporal clusters (p-values ranged from 0.001 to 0.003) of the hospital cases were identified and are displayed in Figure 3.12. The primary cluster (cluster 1) was located around AHN and TPH, and a group of secondary clusters (clusters 2, 3, 4, 8, 9) were located around PWH and SH. These TPUs were also identified as spatial clusters of all SARS cases (Figure 3.7). Cluster 6 was around the PMH, and cluster 7 located in the Tsing Yi Island, which had a population of 193,432 in 55,478 households in 2001 (Census and Statistics Department Hong Kong Special Administrative Region 2001), and both were related to the PMH outbreak (Figure 3.18). Cluster 10 was located between SPH and PYN. Similar to the all cases clusters (Figure 3.9), the majority (80.0%) of the clusters started within one week from day 20, and cluster 1 lasted for the longest (52 days) while cluster 5 only lasted for one day (Figure 3.10). The commencement dates of cluster 6 and 7 matched with the timing of the PMH outbreak (Figure 2.3).

Figure 3.10 Start and end days and the duration (figure in the bar) of the spatiotemporal clusters of hospital cases (excluding ongoing inpatients).

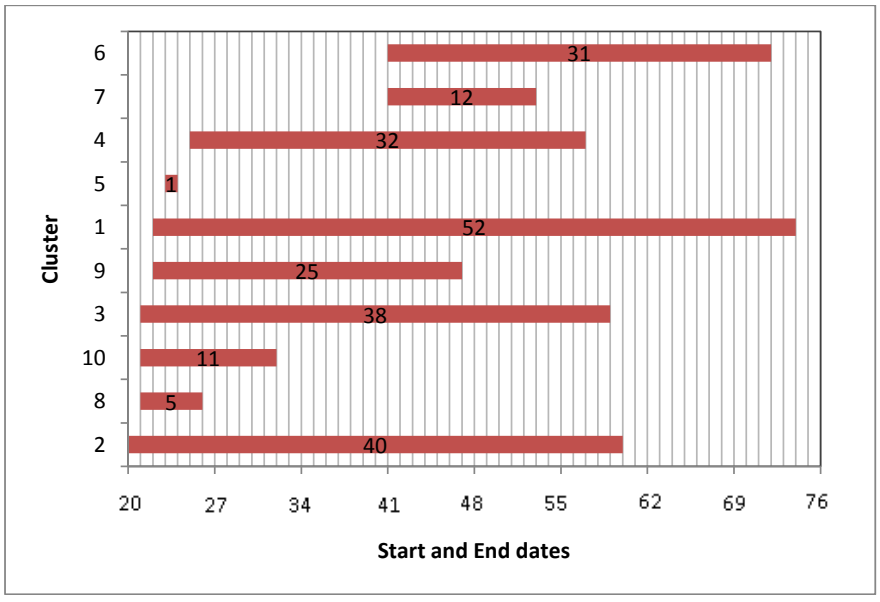


Figure 3.11 Spatial distribution of hospital infected SARS cases in Hong Kong (TPU incidence rate per 100,000 population).

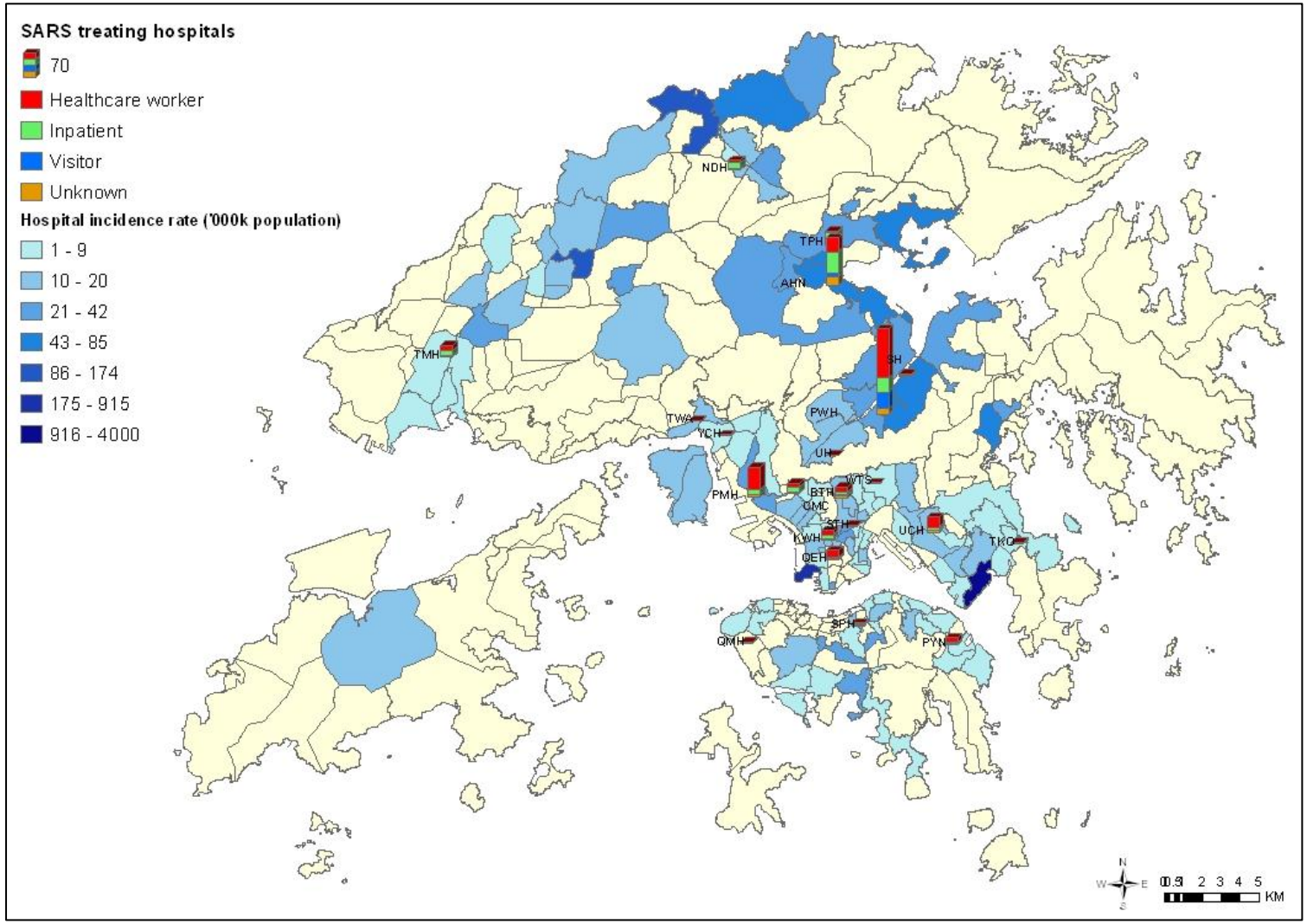


Figure 3.12 Hospital spatiotemporal clusters (excluding ongoing inpatients) numbered in descending order of statistical significance (all with p-values <0.05).

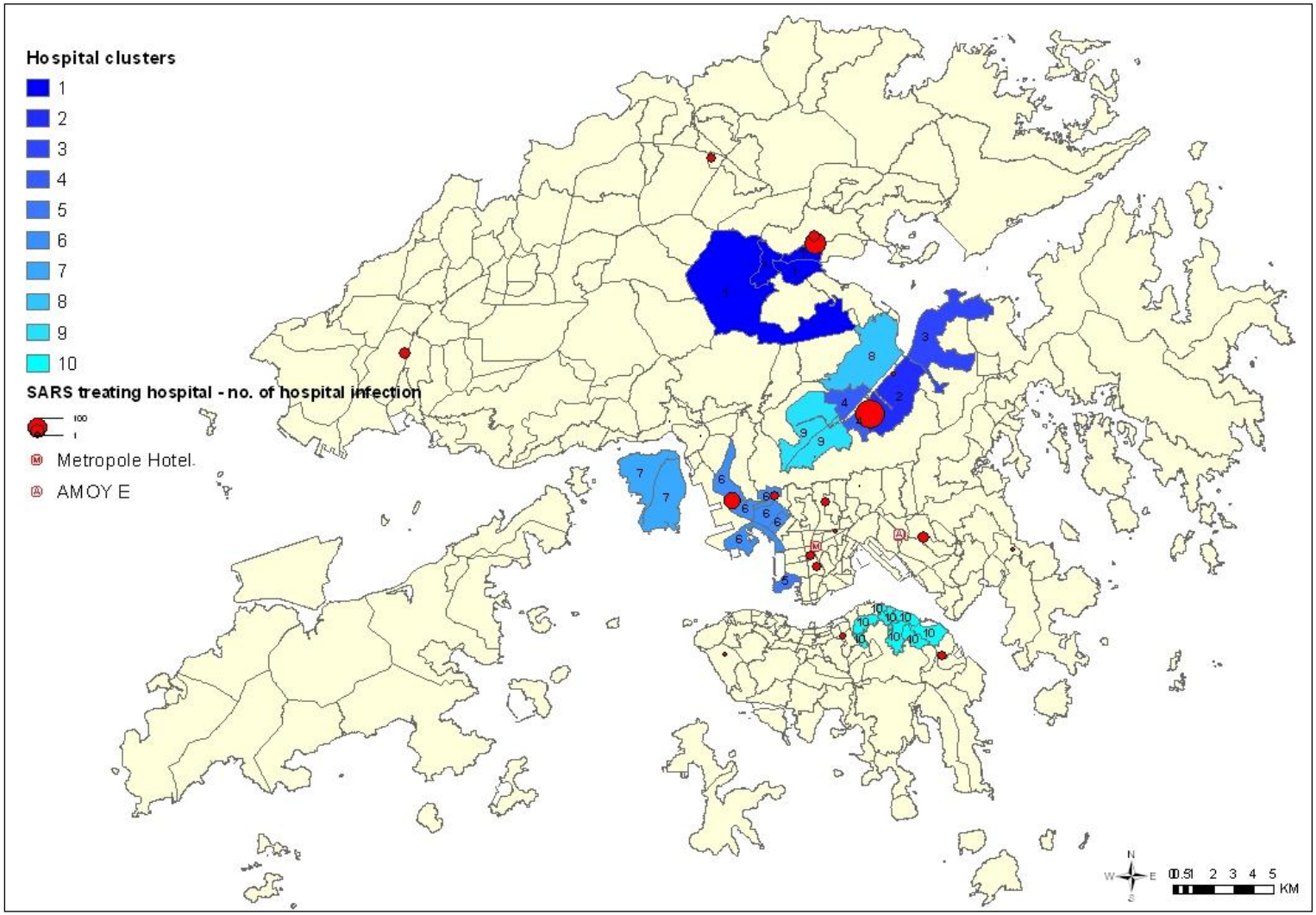


Figure 3.13 Spatial distribution of infected healthcare workers (incidence rate per 100,000 population).

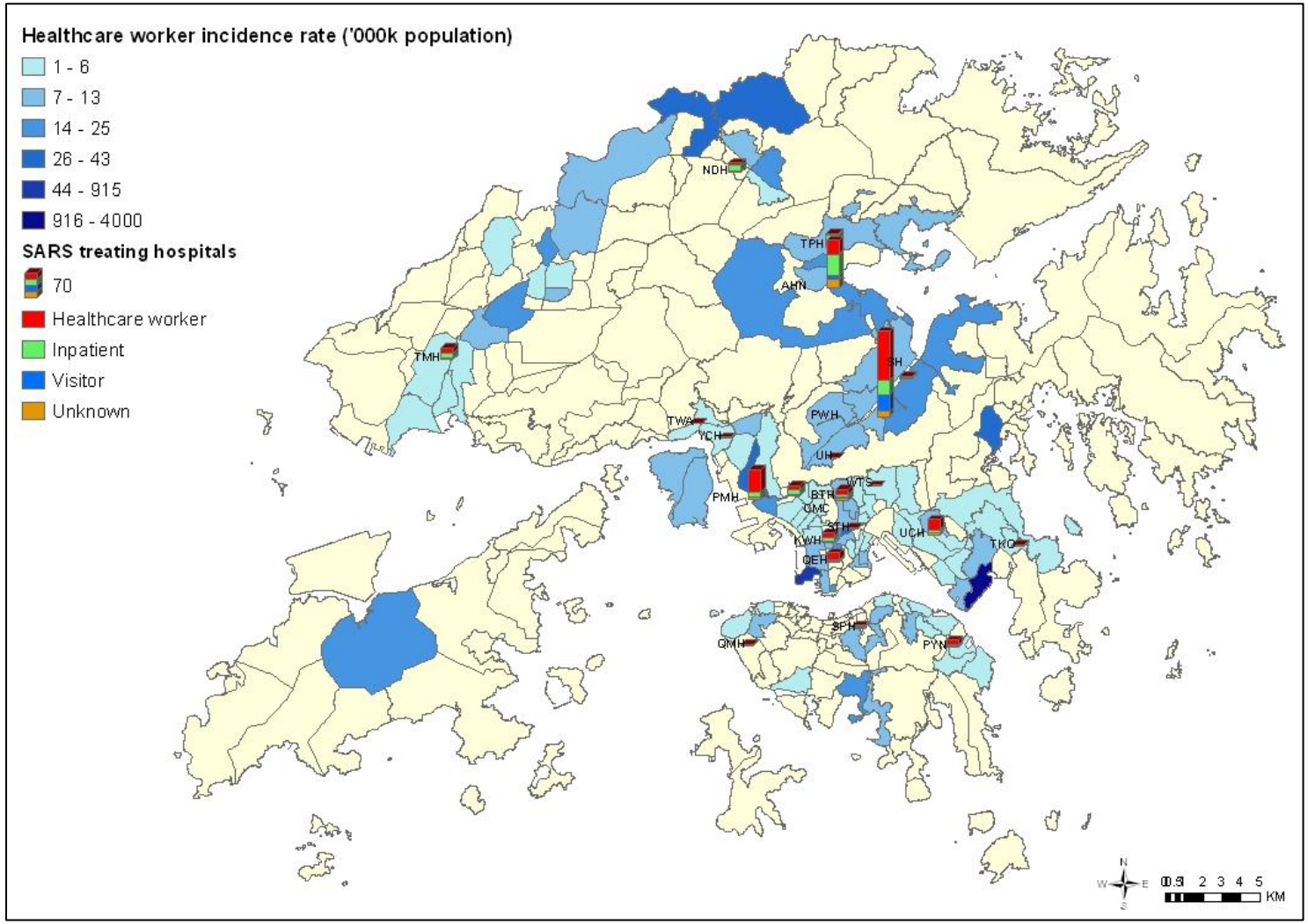


Figure 3.14 Spatial distribution of infected inpatients (incidence rate per 100,000 population).

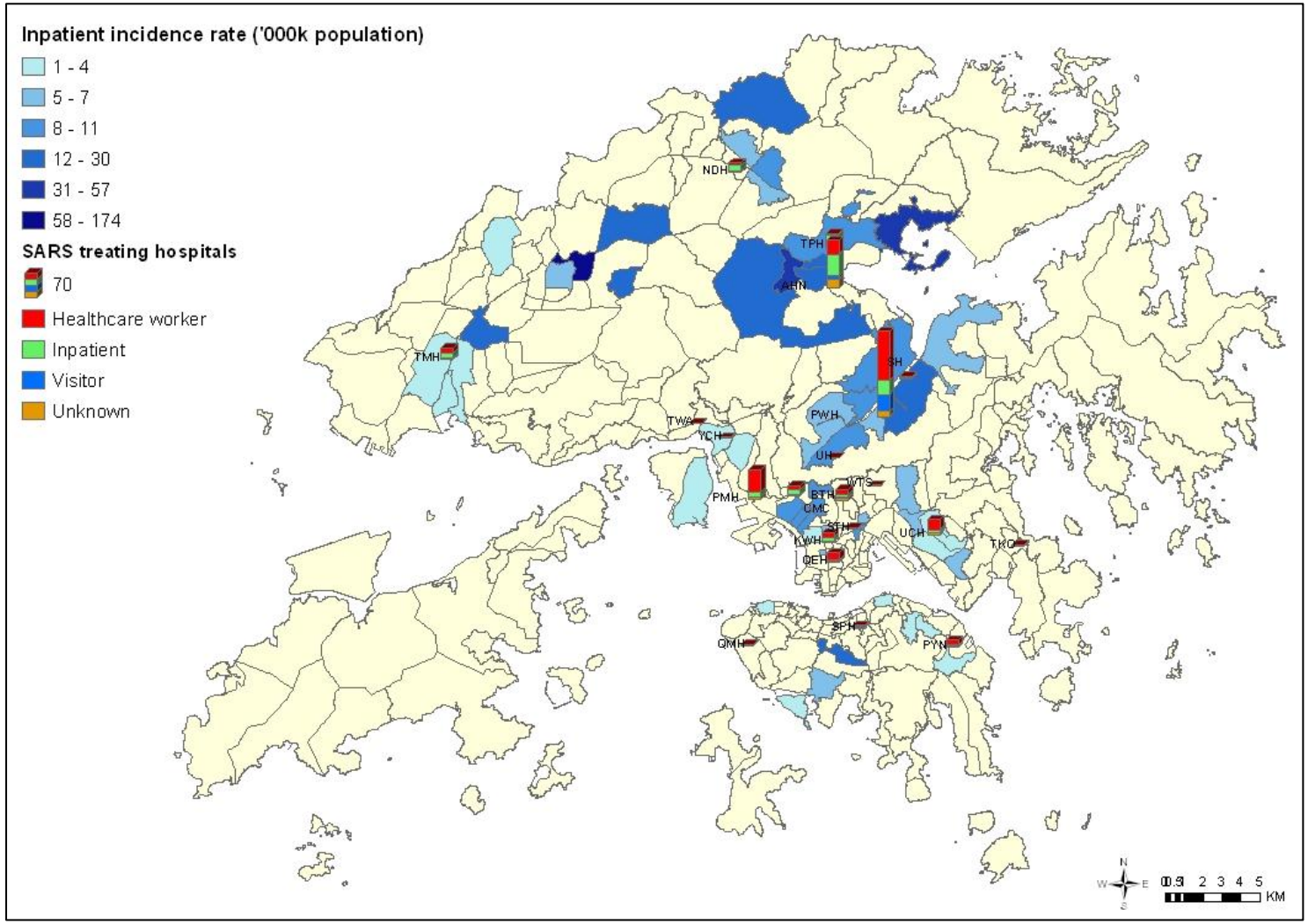


Figure 3.15 Spatial distribution of infected hospital visitors (incidence rate per 100,000 population).

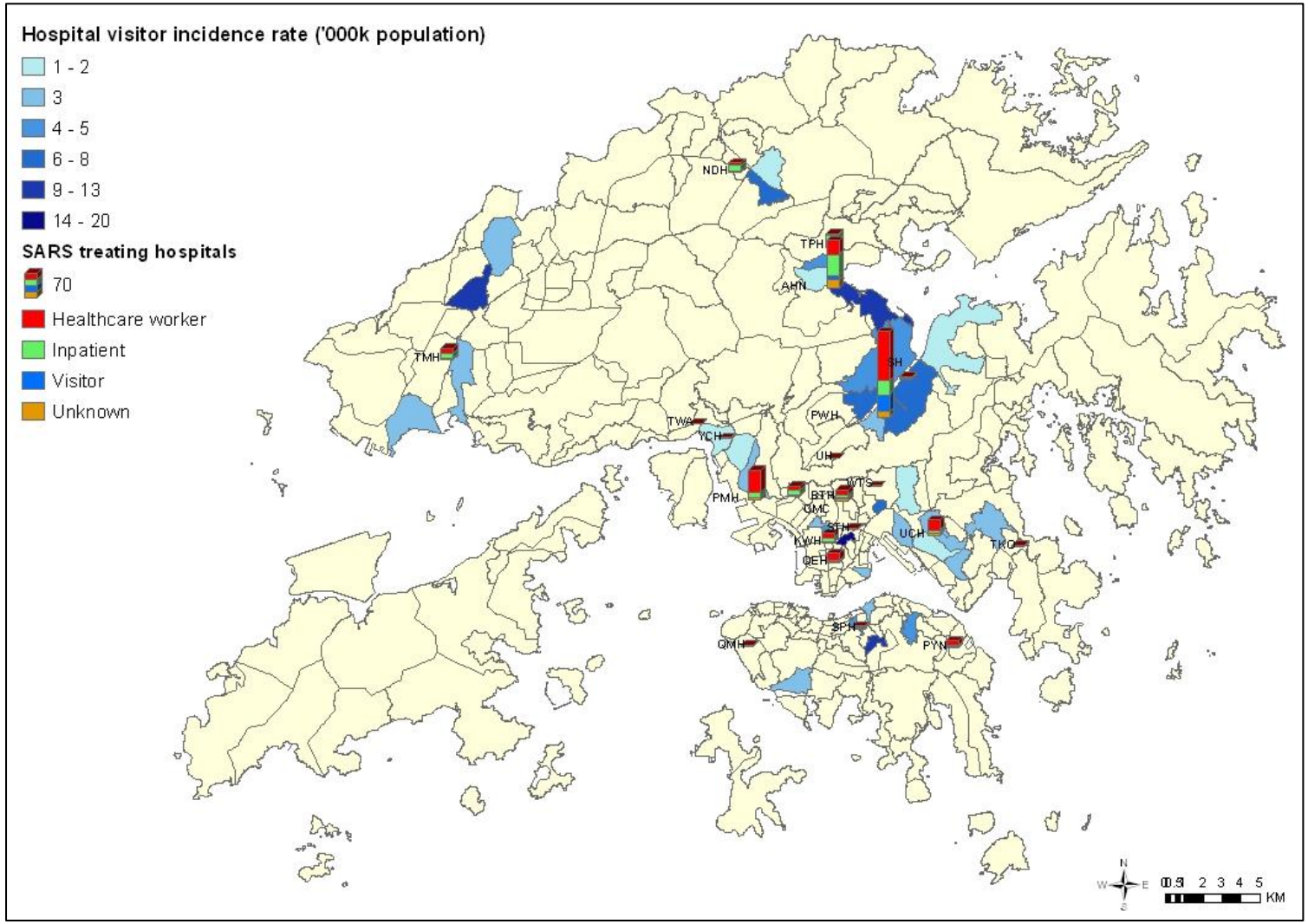


Figure 3.16 Spatial distribution of the PWH infected cases (incidence rate per 100,000 population).

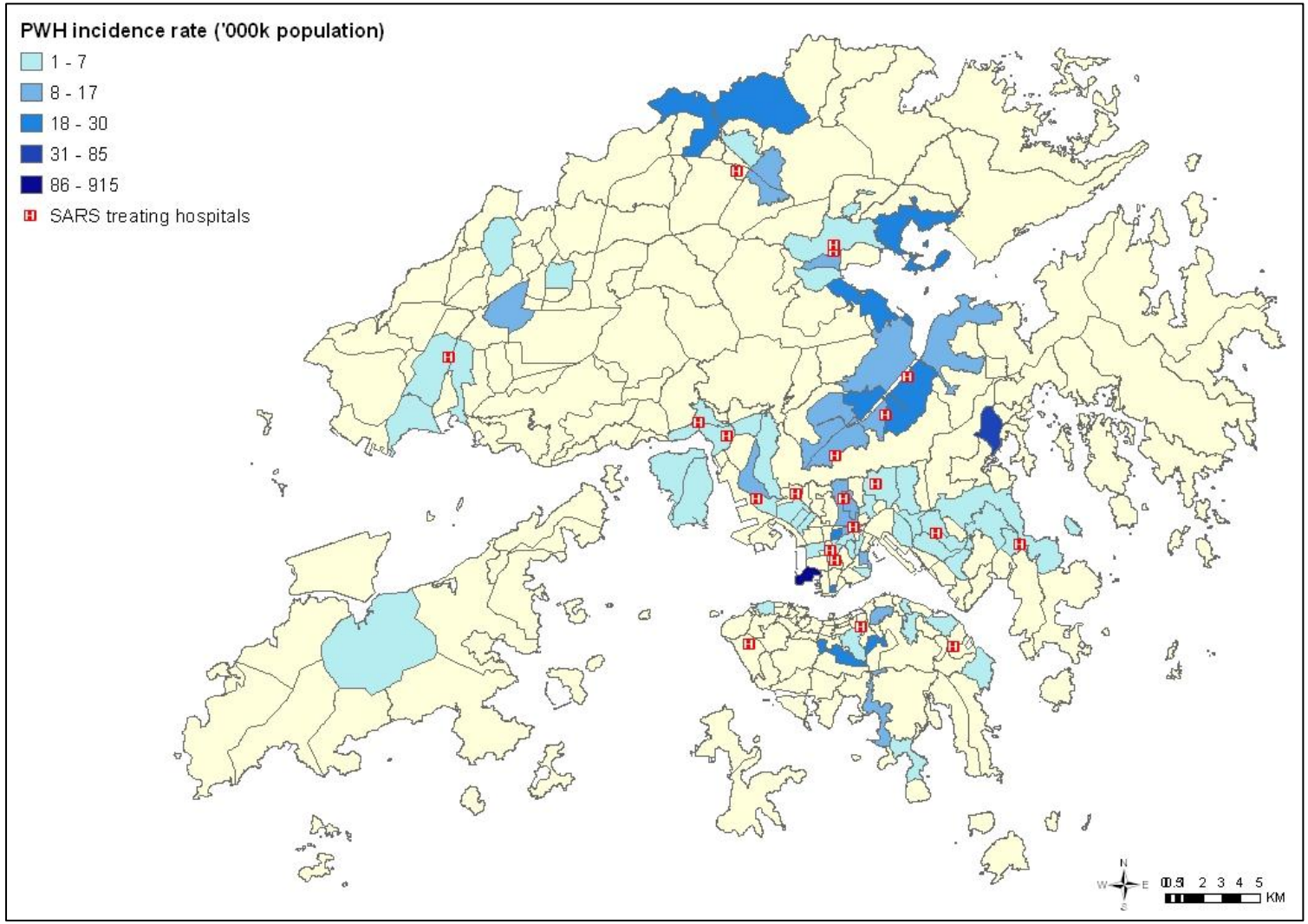


Figure 3.17 Spatial distribution of the AHN infected cases (incidence rate per 100,000 population).

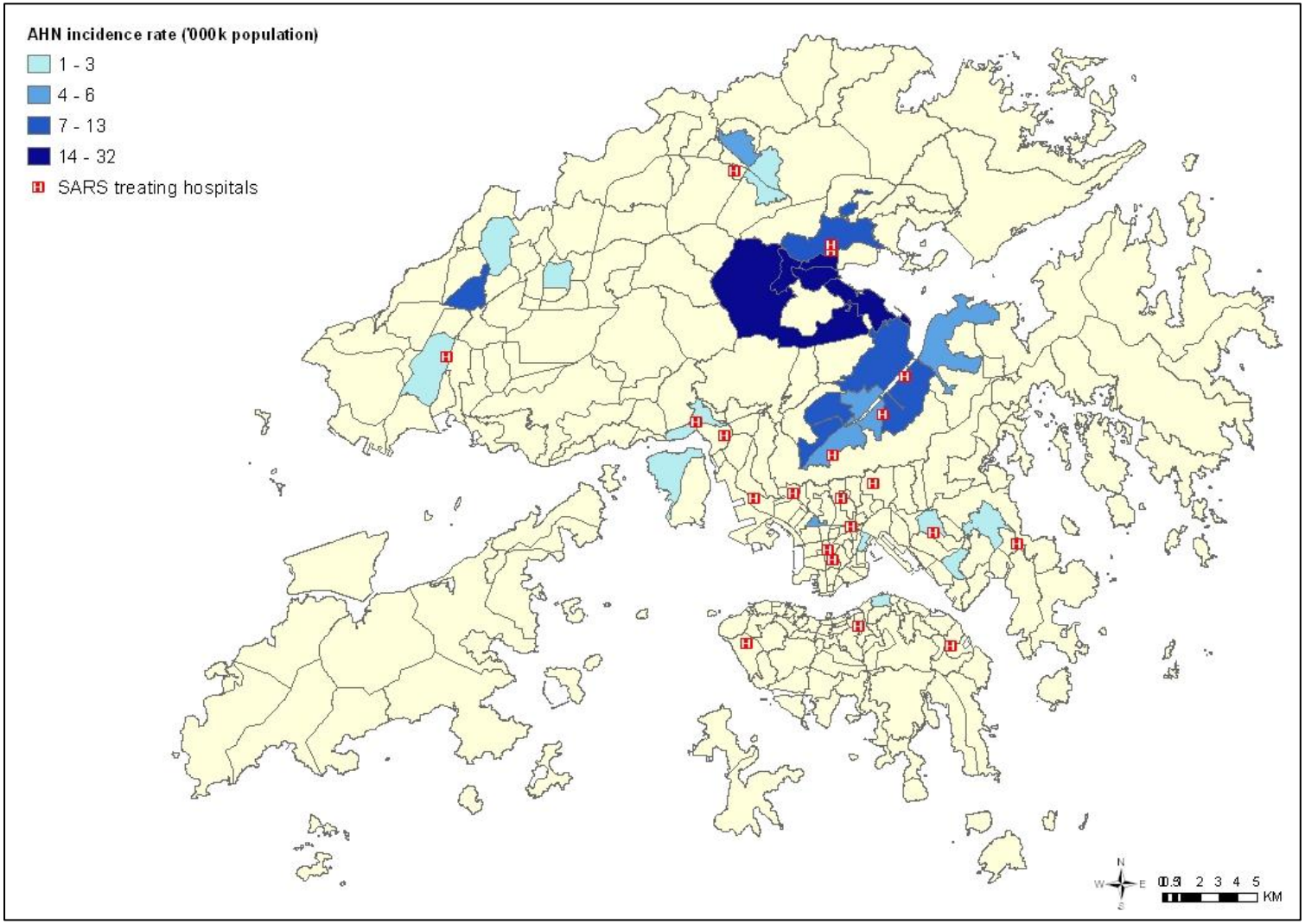
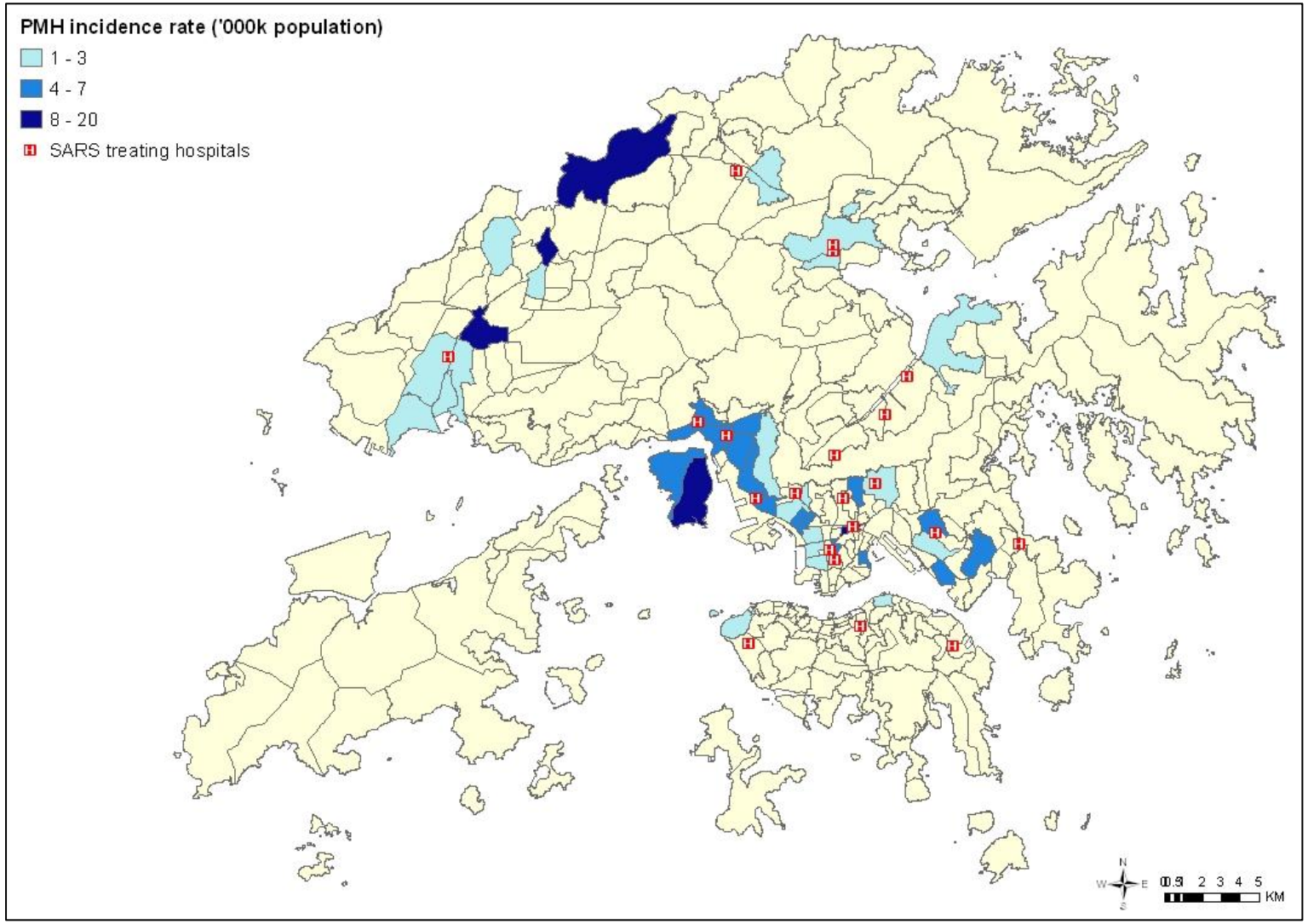


Figure 3.18 Spatial distribution of the PMH infected cases (incidence rate per 100,000 population).



With 49.5% of the hospital infections were in healthcare workers, they were distributed in more TPUs (Figure 3.13), followed by inpatients (30.2%, Figure 3.14) and visitors (9.1%, Figure 3.15), but in terms of the ratio of number of TPUs with the infection subgroup / numbers of cases, healthcare workers had the lowest ratio of 0.27 (103/378), compared to 0.46 (106/233) for the inpatients and 0.49 (34/70) for hospital visitors. However, the relationship between the number of TPUs with the infected subgroup and number of cases was not necessarily linear, as one would expect that as the number of cases increased, the number of TPU remaining uninfected would decrease. In fact, the number of healthcare worker infected was higher than the number of populated TPUs in Hong Kong.

The spatial spread of the cases of individual hospitals varied between the 3 most infected hospitals. PWH spread widely in the territory (Figure 3.16), in contrast to AHN where the cases were relatively concentrated around the Tai Po and Sha Tin districts nearby the hospital (Figure 3.17). PMH cases were slightly more scattered than those of AHN (Figure 3.18).

3.3.3 Community transmission

3.3.3.1 Spatial distribution and clusters

Figure 3.19 shows the distribution of the community cases of the SARS outbreak. The coastal TPU in the south-west of Sai Kung district with the highest incidence rate was due to small population size, and the TPU where the Amoy Gardens is located had incidence rate of 587.6 per 10^5 population. High incident rates were also found in 2 TPUs in the Tai Po and Yuen Long districts, and a TPU in the north of the Sai Kung district. The spatiotemporal clusters of the community cases are displayed in

Figure 3.20, the temporal information of the clusters in Figure 3.21, and the composition of the source of infection in Table 3.3. The TPU where Amoy Gardens is located was the primary cluster (p-value=0.001), dominated by Amoy Gardens cases (99.2%), and lasted for 11 days, the shortest duration among other community clusters. Six secondary clusters were identified (p-values ranged from 0.001 to 0.049), some of them were nearby the sites of major hospital and Amoy Gardens outbreaks. 4 out of the 7 clusters started to be a spatial cluster between day 25 to 29. Cluster 3 near TPH was dominated by household (34.7%) and general community (28.6%) transmission, and it wasthe longest cluster in duration, lasted for 44 days. Cluster 7 near PWH had considerable proportion of household transmission (40.9%). Cluster 6 surrounding the Amoy Gardens had 50.0% of the cases in the cluster infected from the Amoy Gardens outbreak, and was the shortest cluster in duration that lasted for 11 days. Cluster 2 and 4 in Sai Kung district and Eastern district were dominated by outbreaks in residential buildings (91.7% and 63.2% respectively). Cluster 5 in the Kowloon City district had 30.0% of general community and 20.0% of household transmission.

Figure 3.19 Distribution of community infected SARS cases in Hong Kong (TPU incidence rate per 100,000 population).

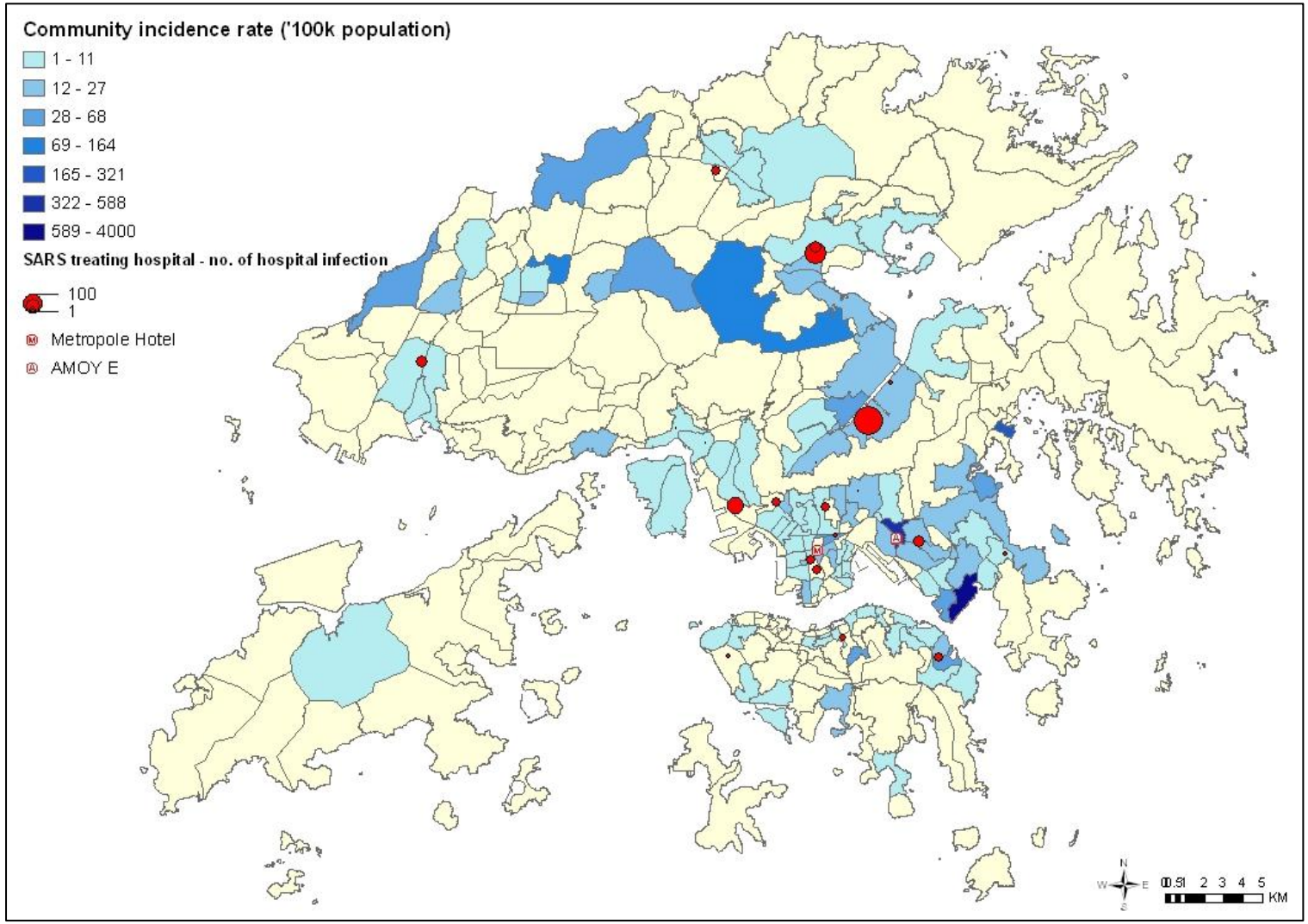


Figure 3.20 Community spatiotemporal clusters numbered in descending order of statistical significance (all with p-values <0.05).

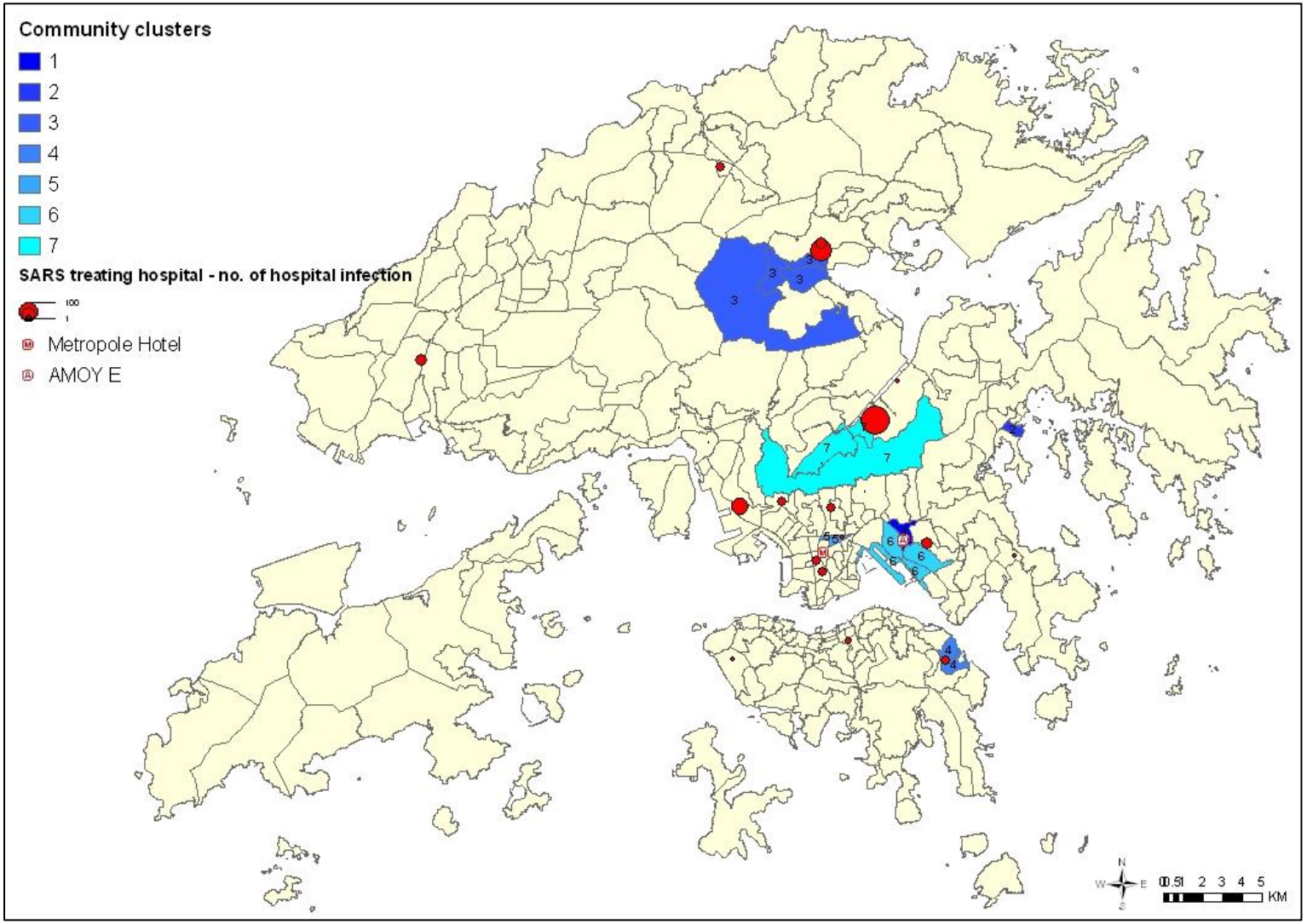


Figure 3.21 Start and end days and the duration (figure in the bar) of the spatiotemporal clusters of community cases.

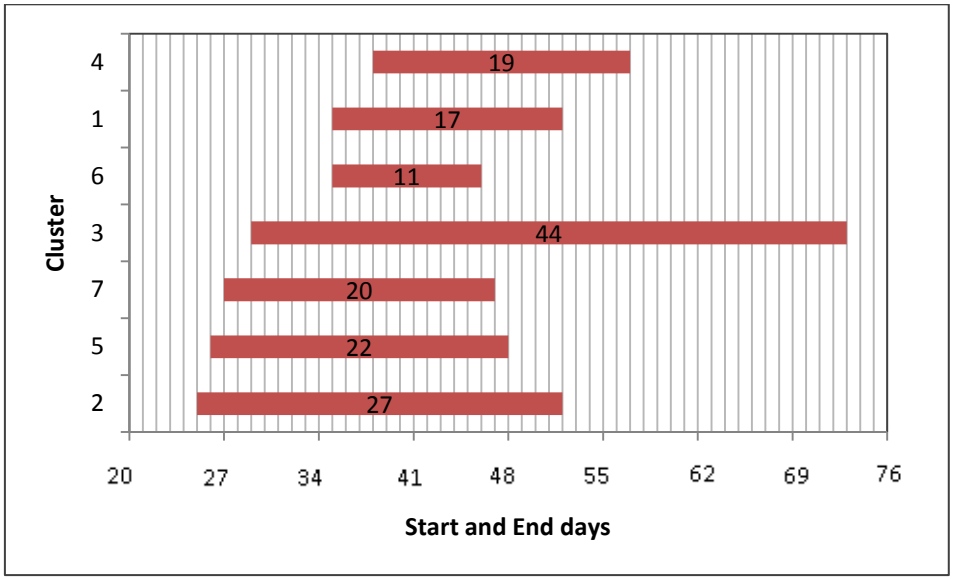


Table 3.3 Composition (% total) of the source of infection of the community spatiotemporal clusters. Figures >25% are in bold.

Cluster	no. of cases in cluster	Source of infection composition (%)								
		amoy	household	building	general community	elderly home	close contact	GP clinic	imported	unknown
1	388	**99.2**	0.5		0.3					
2	12			**91.7**	8.3					
3	49	2.0	**34.7**	6.1	**28.6**	12.2	4.1	4.1	6.1	2.0
4	19	15.8		**63.2**	5.3		5.3		10.5	
5	10	10.0	20.0		**30.0**				10.0	**30.0**
6	22	**50.0**		4.5	18.2				22.7	4.5
7	22	9.1	**40.9**		9.1	4.5	4.5	9.1	13.6	9.1
Total	522	77.2	5.7	5.2	5.0	1.3	0.8	0.8	2.7	1.3

3.3.4 Interactions between hospital and community transmission

3.3.4.1 Comparison of spatiotemporal clusters

Figure 3.22 shows there were two common clusters of the hospital and community cases – adjacent to AHN and PWH respectively, where the latter was a partial overlap (TPU 759 only). Both common clusters first started as hospital clusters for 5 to 7 days, before they also became community clusters. On comparing the timing of the hospital and community clusters (Figure 3.23), the common cluster near AHN ended as hospital and community clusters on the same day, whereas the partial common cluster ended as a hospital cluster 1 day earlier than the community cluster. Hospital clusters in general started earlier than the community clusters, but there was no significant difference in the mean duration of the hospital and community clusters (T-test p-value=0.766).

3.3.4.2 Risk factors for the size of TPU infection

The results of the 6 models on the community case count of the TPUs are shown in Table 3.4. The AIC and BIC showed that model superiority in descending order was ZINB, NB, HNB, ZIP, HP, and Vuong statistic suggested NB, ZINB, HNB were indifferent, and they were better than ZIP, followed by HP. Tests for overdispersion suggested that NB model, compared to quasi-Poisson was justified (p-value <0.001). The HP, HNB and ZINB models accurately predicted the number of zeroes, whereas the other models underestimated the zero counts. The number of cases in the nearest SARS hospital was significant in all the 6 models in predicting the counts and with relatively robust coefficient estimates. In the zero augmented models, the zero and hurdle components of the models predicted the probability of observing a zero count. All the 4 models indicated that increases in geographic distance to the nearest SARS hospital increased the chance of a TPU to be free from community transmission, but the predicted adjusted odds ratio in the ZINB was unreasonably high, at 10.99 (1.40 - 86.41), and that of the HNB, HP and ZIP models were 1.59 (1.07 - 2.38), 1.91 (1.32 – 2.75), and 4.08 (1.54 - 10.80) respectively. Taking into consideration the goodness of fit, accuracy in zero count prediction, and the estimated coefficients, HNB outperformed the other models. The HNB model showed that higher incidence of community transmission in the TPUs was related to a higher percentage of population aged 75+ in the TPU, a lower population density, a greater size of hospital outbreak in the nearest hospital with SARS patients, and earlier first case symptom onset of the TPU. In contrast, a higher chance of a TPU remaining uninfected by SARS was related to a higher percentage individuals of aged 50-74 in the TPU, lower population density, being further away from a hospital with SARS patients, and having less hospital infected subjects living in the TPU. Therefore the effects of hospital to community transmission were confirmed.

Figure 3.22 Comparison of hospital and community spatiotemporal clusters. Labels in red: cluster number of hospital clusters, labels in green: cluster number of community clusters.

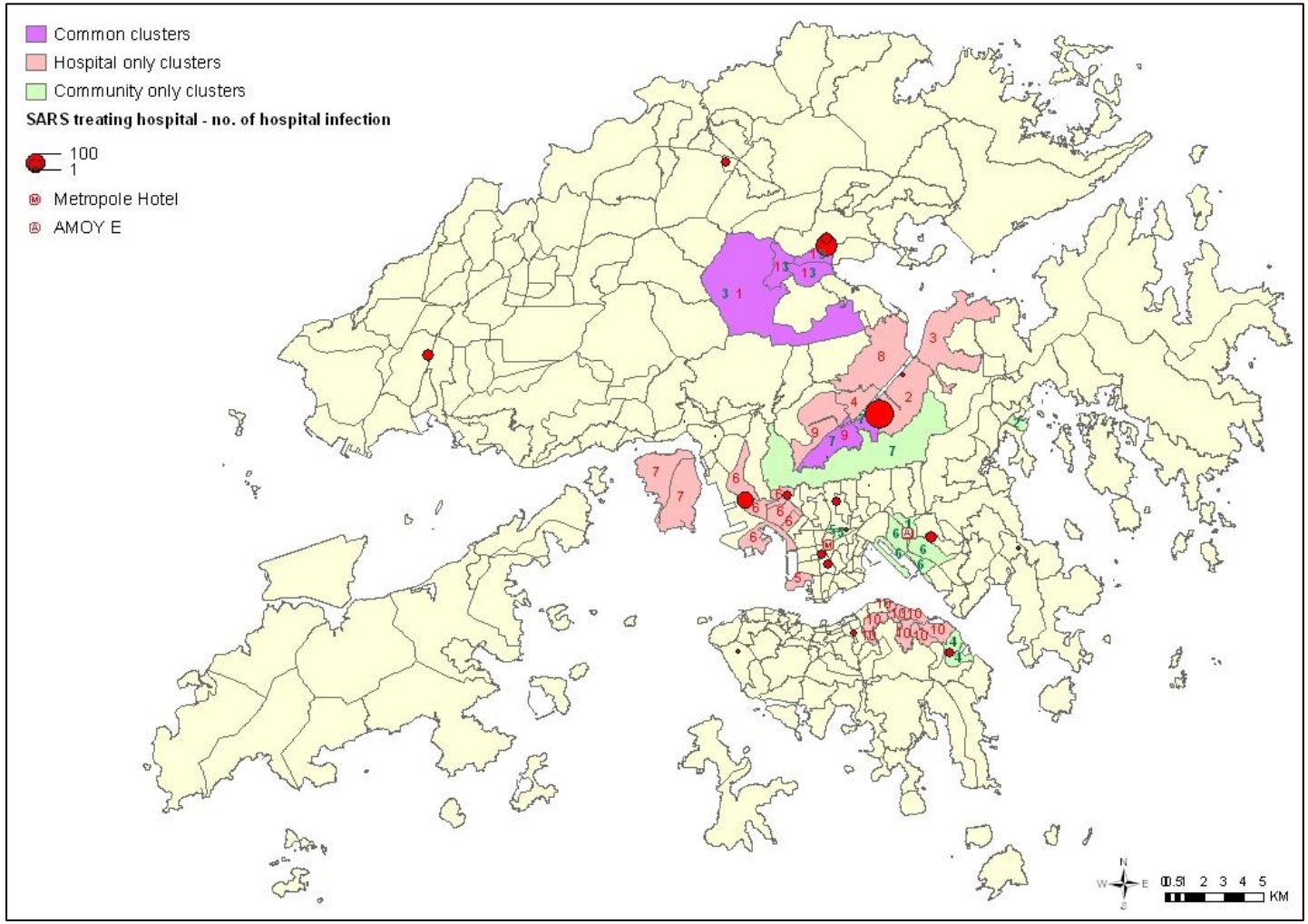
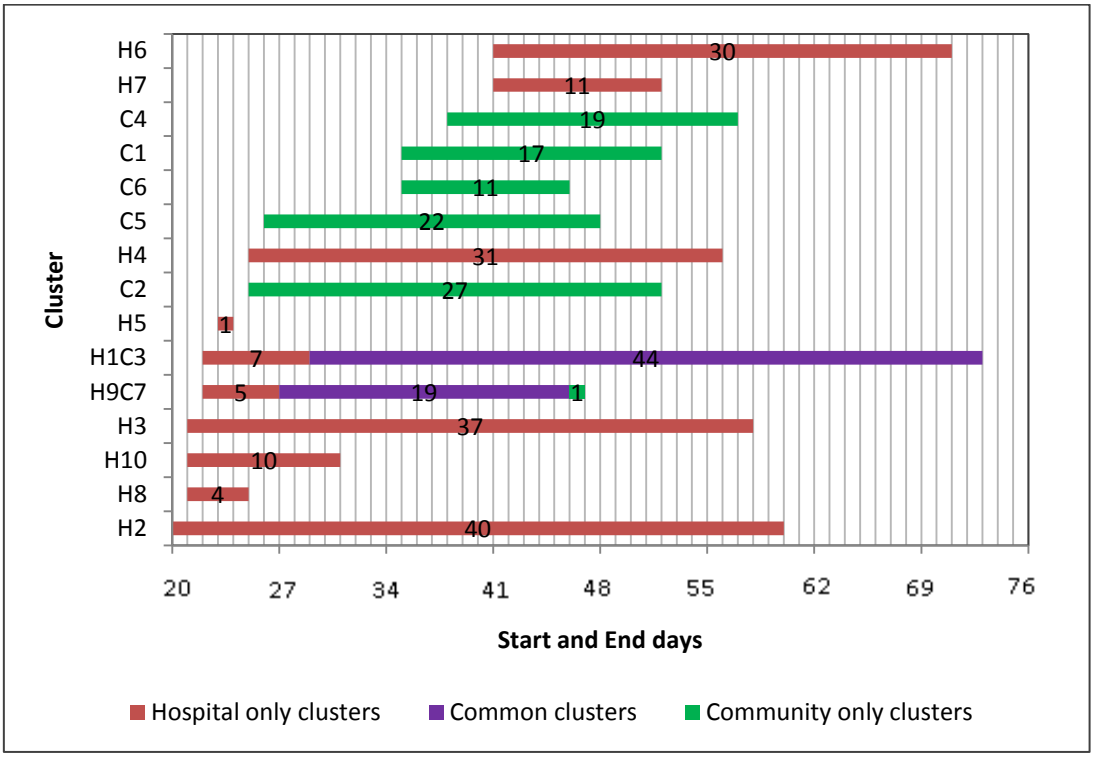


Figure 3.23 Common clusters started as hospital only clusters (TPU 759 only for cluster H9C7) before they became common clusters. Cluster names starting with "H" are hospital clusters, and starting with "C" are community clusters, and the two common clusters are named as "H_C_".



3.3.5 Real-time cluster detection

Figure 3.24 shows the spatial distribution, and Figure 3.25 the temporal information of the 39 TPUs detected as clusters using real-time application of the Kulldorff's prospective scan statistic. The cluster start and end days refers to the days given by the scan statistic that the TPU started or ceased to be a statistical significant cluster. Some TPUs were detected to be a cluster in more than one period of time. Table 3.6 compares the TPUs identified in the prospective and retrospective clusters (note that one must take into account that the scan statistic is conditional on the total number of cases observed when comparing these results). 12 TPUs (30.8%) were commonly identified in both the prospective and retrospective scans as clusters, of which 3 had the same cluster start dates in both scans, but the prospective scan detected later cluster end dates by 7 to 8 days.

Table 3.5 compares the timeline when TPU 291 - where the Amoy Gardens is located, was detected as a cluster. It was first detected as part of a cluster on day 28 along with other TPUs, including TPUs 233, 246, 271, 272, 280, 281, 282, 283, 284, 285, 286, 287, 288 (only 2 of which were identified in

Table 3.4 Comparison of the results of quasi-poisson, negative binomial, zero-inflated poisson / negative binomial, hurdle poisson / negative binomial models on the community case count in TPUs. The zero and hurdle components of the zero-inflated and hurdle models predicted the probability of observing a zero count in the TPU.

	Quasipoisson			Zero-inflated Poisson			Hurdle Poisson			Negative Binomial			Zero-inflated Negative Binomial			Hurdle Negative Binomial		
	Estimate	*SE*	*p-value*	*Estimate*	*SE*	*p-value*	*Estimate*	*SE*	*p-value*	*Estimate*	*SE*	*p-value*	*Estimate*	*SE*	*p-value*	*Estimate*	*SE*	*p-value*
Count model coefficients (Offset variable: ln(population size of TPU))																		
Intercept	-9.053	0.877	<0.001	-9.038	0.433	<0.001	-8.446	0.532	<0.001	-8.400	0.703	<0.001	-7.747	0.833	<0.001	-9.363	0.719	<0.001
No. of cases in the nearest SARS hospital	0.006	0.002	<0.001	0.005	0.001	<0.001	0.006	0.001	<0.001	0.008	0.003	0.003	0.007	0.002	0.006	0.007	0.003	0.044
% of aged 50-74 in TPU	-0.086	0.041	0.035	-0.083	0.022	<0.001	-0.100	0.025	<0.001	-0.069	0.032	0.033	-0.094	0.037	0.011			
% of aged 75+ in TPU	0.074	0.024	<0.001	0.100	0.016	<0.001	0.132	0.018	<0.001							0.088	0.044	0.043
No. of hospital infected cases in the TPU	0.020	0.008	0.017	0.023	0.004	<0.001	0.021	0.005	<0.001									
Population density of the TPU				-0.006	0.002	0.001	-0.008	0.002	<0.001							-0.014	0.004	<0.001
Day of first symptom onset in the TPU							-0.020	0.007	<0.001							-0.036	0.018	0.044
Zero-inflated / zero hurdle model coefficients																		
Intercept				-12.413	4.139	0.003	-7.238	2.135	<0.001				-21.811	9.270	0.019	-5.019	2.396	0.036
ln(distance to the nearest SARS hospital)				1.406	0.496	0.005	0.646	0.187	<0.001				2.397	1.052	0.023	0.466	0.204	0.023
Population density of the TPU							-0.031	0.006	<0.001							-0.023	0.006	<0.001
% of aged 50-74 in TPU							0.167	0.052	<0.001							0.161	0.061	0.009
No. of hospital infected cases in the TPU																-0.488	0.104	<0.001
logLik	NA			-348.651			-373.590			-317.853			-315.598			-316.999		
DF	5			8			11			4			6			11		
AIC	NA			713.302			769.181			643.706			643.197			655.998		
BIC	NA			741.377			807.784			657.744			664.253			694.601		
Vuong's test (superior to)																		
Zero-inflated Poisson				NA		NA	-2.689		0.004	1.363		0.087	1.454		0.073	1.161		0.123
Hurdle Poisson				2.689		0.004	NA		NA	2.595		0.005	2.694		0.004	2.264		0.012
Negative Binomial				-1.363		0.087	-2.595		0.005	NA		NA	0.971		0.166	0.104		0.459
Zero-inflated Negative Binomial				-1.454		0.073	-2.694		0.004	-0.971		0.166	NA		NA	-0.175		0.431
Hurdle Negative Binomial				-1.161		0.123	-2.264		0.012	-0.104		0.459	0.175		0.431	NA		NA
Expected no. of zeroes*	130			145			152			148			152			152		

* Observed number of zeroes is 152

Table 3.5 Comparison of the timeline the Amoy TPU (291) was detected by the prospective and retrospective scan statistics, and the time health officials were alert to and confirmed the outbreak.

Timeline	Day	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
No. of cases	non-Amoy	2	1								1				1	1			
in TPU 291	Amoy			1	1	0	1	1	7	6	11	80	73	44	44	28	17	13	16
	Cumulative Amoy cases			1	2	2	3	4	11	17	28	108	181	225	269	297	314	327	343
Time when cluster	Retrospective																		
detected	Prospective	Part of a bigger cluster								Standalone cluster									
	Health official													Alert					confirm

the retrospective scan), with 9 of the 57 total cases (15.8%) distributed in these TPUs, and 2 in TPU 291. Another case was recorded in TPU 291 on day 29 before the Amoy Gardens outbreak started on day 30. The same cluster remained significant until day 36, when TPU 291 became a primary cluster of its own, by then there were 20 accumulated cases in the TPU, 17 of which were Amoy cases. The retrospective scan detected the Amoy Gardens outbreak on day 35, when there were 11 cumulative Amoy cases. The health officials of Hong Kong become alert of an abnormal outbreak on day 40, when the cumulative cases reached 225 and the peak of the outbreak had past, and issued the isolation order on the Amoy Gardens Block E on day 45, when there was already 343 cases (out of total Amoy cases of 447). The cluster end date detected in the prospective scan was longer than that of the retrospective scan by 49 days.

Other than TPU 291, prospective clusters tended to start later than the retrospective cluster by 2 to 19 days (mean: 5.0 days, standard deviation 6.9 days). Difference of the end dates of the scans ranged from 14 days earlier in the real-time scan to 23 days later (mean: 3.0 days, standard deviation 12.0 days).

27 TPUs (69.2%) of the prospective clusters which were not identified as clusters in the retrospective scan (Table 3.6), and 5 TPUs, which were part of the retrospective cluster 6 (TPUs 247, 280, 294, 295), and part of the cluster 7 (TPU 761), were not detected in the prospective scan as clusters. Three of these TPUs (247, 295, 761) actually had no SARS cases, and they were included in the retrospective scan as part of the cluster formed by more than 1 TPU. The prospective scan seemed to be relatively sensitive at the early stage of the epidemic, during which the number of cases was small. There were 6 TPUs (220, 226, 227, 228, 229, 235) detected as a cluster (p-value=0.037) on day 9 in the prospective scan only, during which 2 cases out of the total 6 cases were distributed in the cluster. In day 24, a prospective only cluster of 4 TPUs (268, 271, 272, 282) was detected (p-value=0.07), where 3 cases out of 29 were distributed in the cluster. This was also illustrated by the cluster including TPU 291 before the Amoy Gardens outbreak as discussed earlier. Large clusters with long duration seemed to dominate the prospective scan, resulting in much longer durations of these clusters (e.g. TPUs 291 and 724) being identified.

Figure 3.24 Prospective community TPUs detected as clusters compared to TPUs detected retrospectively.

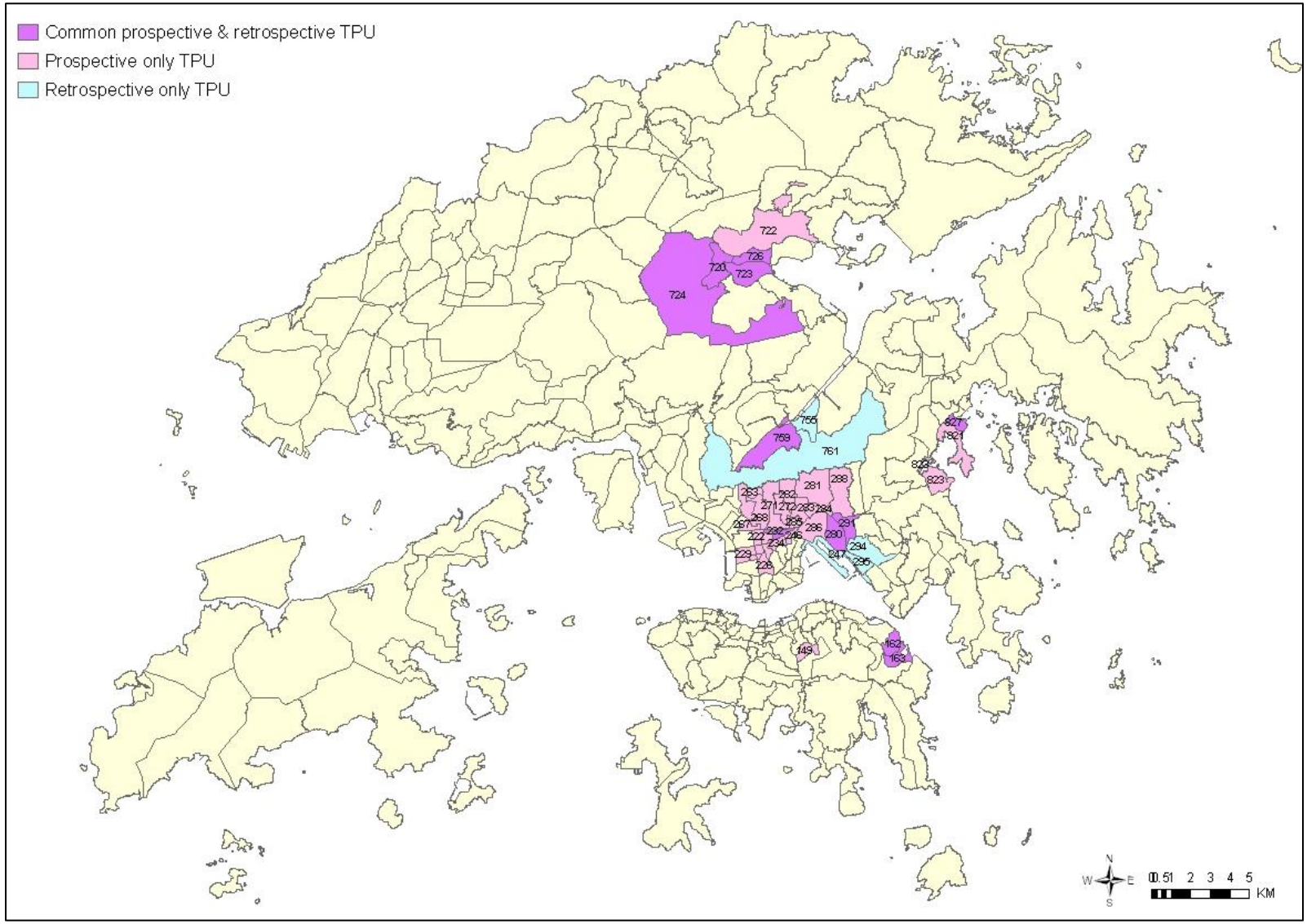


Figure 3.25 Start and end days and the duration (figure in the bar) of the prospective spatiotemporal clusters of community cases. Some of the clusters were detected to be significant in multiple periods. TPU marked with * were not detected in the retrospective scan.

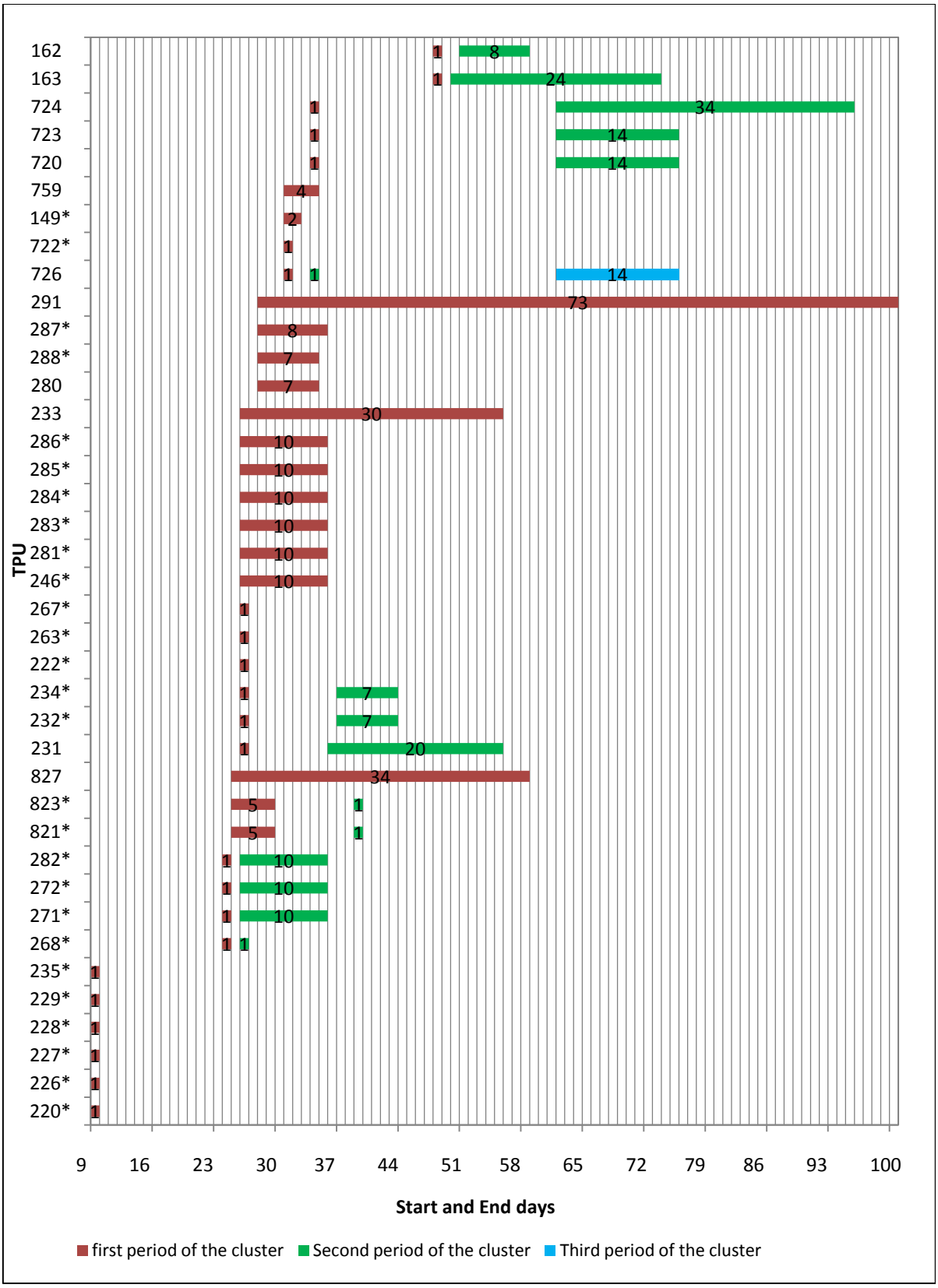


Table 3.6 Comparison of the prospective and retrospective clusters. Prospective clusters with multiple significant periods are shaded in grey.

	Prospective		Retrospective				Difference (day)	
	Cluster start	Cluster end			Cluster start	Cluster end	(Prospective - Retrospective)	
TPU	(day)		Cluster no.	Location	(day)		Cluster start	Cluster end
Match in cluster start date								
827	25	59	2	Sai Kung	25	52		7
231	26	27	5	Kowloon City	26	48		
	36	56						8
233	26	56	5	Kowloon City	26	48		8
Different duration								
280	28	35	6	nearby Amoy	26	48	2	-13
291	28	101	1	Amoy Gardens	35	52	-7	49
726	31	32	3	nearby AHN & TPH	29	73	2	
	34	34						
	62	76						3
759	31	35	7	nearby PWH	27	47	4	-12
720	34	34	3	nearby AHN & TPH	29	73	5	
	62	76						3
723	34	34	3	nearby AHN & TPH	29	73	5	
	62	76						3
724	34	34	3	nearby AHN & TPH	29	73	5	
	62	96						23
162	48	48	4	Eastern	29	73	19	
	51	59						-14
163	48	48	4	Eastern	38	57	10	
	50	74						17
Detected by prospective scan only								
220	9	9						
226	9	9						
227	9	9						
228	9	9						
229	9	9						
235	9	9						
268	24	24						
	26	27						
271	24	24						
	26	36						
272	24	24						
	26	36						
282	24	24						
	26	36						
821	25	30						
	39	39						
823	25	30						
	39	39						
222	26	27						
232	37	44						
	26	27						
234	26	27						
	37	44						
246	26	36						
263	26	27						
267	26	27						
281	26	36						
283	26	36						
284	26	36						
285	26	36						
286	26	36						
287	28	36						
288	28	35						
149	31	33						
722	31	32						
Detected by retrospective scan only								
294			6	nearby Amoy	35	46		
247			6	nearby Amoy	26	46		
295			6	nearby Amoy	26	46		
755			7	nearby PWH	27	47		
761			7	nearby PWH	27	47		

.

3.4 Discussion

The SARS outbreak in Hong Kong was characterised by large proportion of hospital infections, with half of the hospital cases occurring in two hospitals (PWH and AHN). Half of those infected in hospital were healthcare workers. The other large outbreak at Amoy Gardens accounted for one-quarter of all cases and was caused by environmental factors in the affected building. These three large outbreaks certainly have implications for the spatial distribution of cases in the epidemic as a whole. My work has explored the impact of these outbreaks on the spatial disease clusters that are identified, and on the variability in incidence spatially.

I found that the distribution of cases in the SARS outbreak in Hong Kong was spatiotemporally clustered, in agreement with a previous study (Lai, Wong et al. 2004). A considerable proportion of administrative districts remained uninfected, and the incidence rates of the infected areas varied substantially. Half of the total cases were contained in the spatial clusters I identified, and the majority of the clusters were located around major outbreak hospitals including AHN and PWH, and the Amoy Gardens. Before the PWH outbreak, the epidemic was sporadic and the cases were scattered over the territory. The PWH outbreak generated a significant increase in both incidence and spatial spread of SARS, and subsequently the formation of spatiotemporal clusters, the majority of which could be identified within a week of the start of the PWH outbreak.

Hospital and community transmission were inter-related in the SARS outbreak - people infected in hospitals generated secondary cases in the community, both in households and other community settings, and all the infected subjects were hospitalised (sub-clinical infections being negligible in SARS (Leung, Lim et al. 2006)). This made hospitals treating SARS cases places with high transmission risk. In this study I have examined two aspects of the influence of hospitals on the epidemic - disease cluster formation and the size of community transmission.

In the spatial clusters located near hospitals, 27% - 41% of cases were infected in the community. On comparing spatiotemporal clusters identified examining solely hospital or solely community cases the same clusters were identified around AHN and PWH, indicating the importance of the location of major outbreak hospitals on both hospital and community disease clusters. Temporally these common clusters first started as hospital-only clusters and then became simultaneous hospital-transmission and community-transmission clusters, suggesting hospital outbreaks triggered community transmission.

The size of the community outbreak in an area increased with the size of the outbreak in the nearest hospital treating SARS, and an area was more likely to have no community-infected cases if it was far

from hospitals treating SARS, or had less hospital-infected cases within the area. Therefore the distance to hospitals treating SARS patients can be considered as intrinsic factor of spatial heterogeneity of exposure of an area in the transmission of SARS. Population distribution of the elderly was related to the spatial heterogeneity of susceptibility of an area, with a higher proportion of individuals aged 75+ in the area increasing the size of the community outbreak, while a higher proportion of the age group 50-74 increasing the likelihood that an area to had no community-infected cases. This could be related to the over-representation of the aged 75+ and under-representation of the 50-74 age group in the SARS cases compared to the general population (Leung, Hedley et al. 2004).

Given the driving role of hospital transmission in the overall spatial spread of the epidemic, I examined the spatial distribution of the three main types of hospital infections – healthcare workers, inpatients and visitors. Considerable variation was observed in the spatial distribution of the three main outbreak hospitals cases, and in the types of people infected in individual hospitals (i.e. healthcare workers, in-patients, or visitors). The latter is likely to have been related to the implementation of infection control measures in the different hospitals and the number of SARS patients treated. For example, the TPH outbreak was sourced from the transfer of 14 patients who were considered as not infected with SARS from AHN for convalescence (Severe Acute Respiratory Syndrome Expert Committee 2003), and thus seeded a high proportion of inpatients.

The control policy implications from the study are that areas near hospitals that treat SARS patients are of higher risk of forming a disease cluster and inducing further community transmission, and the final size of the community outbreak in the area is related to the size of the nearest hospital outbreak. With the observation that key hospital spatial clusters lead the start of community clusters in a short period of time, close monitoring of the emergence of hospital clusters may provide opportunities to prevent further community transmission and formation of community clusters. Therefore, timely and localised infection control measures in these areas could help to maximize the benefit and cost-effectiveness of the control measures.

Timely implementation of localized control measures partly relies on real-time detection of the emerging spatial disease clusters during the epidemic. My work shows that real-time application of the Kulldorff's prospective scan statistic method would have been effective in identifying some real-time clusters accurately, and would have detected the Amoy Garden cluster earlier than in reality by the health officials during the outbreak, and before peak transmission was reached, though the result was masked by a few early non-Amoy cases.

However, because the method is conditional on the total number of cases in the data, at the initial period of the epidemic when the number of total cases was small, it tended to detect clusters as a group of TPUs where the cases where concentrated. As more cases were records later and with the spatial distribution changed over time, these clusters were no longer significant. Because the total number of cases in the complete dataset was much greater, the retrospective scan did not pick up these TPUs as clusters for that time period. In the public health perspective, when these clusters are detected during the epidemic, it is worth putting them under observation, because it is uncertain if these clusters would grow further. Given that the prospective scan can be run regularly upon new data arrival, any existing cluster which does not grow further will be indicated as insignificant cluster in the later prospective scans.

There were a few TPUs which did not have any SARS cases but detected as part of the retrospective clusters. This justified the much smaller maximum circle size settings applied in the prospective and retrospective scans than the default settings. If the default settings are applied, it is believed more false TPUs would have been included as part of the cluster. The retrospective and prospective scan results verified that using smaller maximum circle in space and time can still detect clusters of large size and long duration as more than one significant cluster. On the contrary, a study on the spatial patterns of Hong Kong H1N1 pandemic that used the default settings of 50% spatial and temporal window as the maximum resulted in some relatively big clusters (Lee and Wong 2010), though it is not directly comparable to my study.

So far my work has focused on cluster formation and the interactions between hospital and community transmission. Between-areas spread and its dynamics are best studied using contact tracing data or methods of infection tree inference. Complete contact tracing data is scarcely available for any large-scale emerging outbreak, and this is the case of the SARS outbreak in Hong Kong in 2003. In chapter 6, I use an infection tree inference method derived by extending the Wallinga and Teunis method spatially to derive the between areas and within area spread of SARS in Hong Kong.

Chapter 4 - Studying Human Movement Patterns Using Gravity Models

In the previous chapter, I studied the dynamics of the spread of the SARS epidemic in and between the hospital and community settings, and suggested a better understanding of the mechanism of spread could be attained by inferring the probabilistic infection tree of the epidemic. Based on my findings that spatial distance from the SARS cases and hospitals treating SARS cases were risk factors of disease transmission, in Chapter 5 I present a modified version of the Wallinga and Teunis (Wallinga and Teunis 2004) method that captures the spatial risk factors in the inference of infection network by using a gravity model to determine spatial connectedness. In this chapter, I derive the gravity model that represents the travel patterns of the Hong Kong population in relation to the SARS transmission. By exploring the relationship between the types of hospital infected subjects and their hospital flow patterns by gravity models, I also determine the extent to which the spatial spread of SARS is driven by spread from hospitals to the community.

I begin the chapter with an introduction to gravity models including a review of its application to studying infectious disease epidemics. This is followed by the research objectives, methods and data used and results.

Throughout I use the term “outflow” of a district to refer to the number of trips made from that district as origin, regardless of the destination, including the within-district flow. Similarly the “inflow” of a district is used to describe the number of trips to that district as the destination, regardless of the origin (which again can be the same as the destination district). The type of flow was classified by the trip purpose. For example, workflow refers to trips between home and workplace, study flow refers to trips between home and place of study, and work-study flow refers to those between home and places of work or study.

4.1 Gravity models

4.1.1 Origin

Gravity models are commonly used to model human spatial interaction behaviour – any movement in space that arises from a human process. They originate from the Newton's law of Universal Gravitation, which states that between two point masses there is a force of attraction with magnitude directly proportional to the product of the two masses, and inversely proportional to the square of the distance between them. It can be expressed as:

$$F = G\frac{m_1 m_2}{d^2}$$

F: magnitude of the gravitational force between the two point masses

G: gravitational constant

m_1, m_2: mass of the two point masses

d: distance between the two point masses

The first conceptual application of Newton's law to human movement was developed by Carey (Carey 1858), but it was almost a century later that a formal postulation was formulated (Stewart 1941) to quantify human movement between regions to be proportional to the populations of the origin and destination, and inversely proportional to the squared distance between the regions. The squared distance relationship was later relaxed and the exponent was calibrated from observations (Huff 1963). In the past few decades, there has been substantial development of the theoretical basis and model formulation for the family of gravity models (Sen and Smith 1995). They have been widely applied to study different types of movement patterns, and are an essential tool in trip distribution modelling and transport planning (Erlander and Stewart 1990), retail planning and modelling different types of human and commodity flows (Haynes and Fotheringham 1984; Fotheringham and O'Kelly 1988).

4.1.2 Generic model formulation

Denote T_{ij} as the flow from origin *i* to destination *j*, *O(i)* as a function of propelling or outflow attribute(s) of the origin *i*, *D(j)* as a function of attracting or inflow attribute(s) of the destination *j*, and $f(d_{ij})$ as a distance deterrence function that reflects the tendency to travel relative to distance, and *c* as the constant. The generic form of gravity model is written as:

$$T_{ij} = cO(i)D(j)f(d_{ij})$$

Gravity models can be inflow constrained, outflow constrained, doubly constrained, or unconstrained (Haynes and Fotheringham 1984; Fotheringham and O'Kelly 1988). The constraint requires the estimated marginal flow (inflow, outflow, or both) of each study region to be equal to that of the observed marginal flow. Usually model fit improves with additional information in the constrained models compared to unconstrained models (Fotheringham and O'Kelly 1988), but the research question under study also determines the appropriate type of model to be applied besides data availability. An outflow constrained model quantifies the relative attractiveness of a region as a destination compared to all other possible destinations to travel from the origin. Similarly, an inflow constrained model quantifies the relative contribution of a region as an origin compared with other possible origins conditional upon the marginal inflow of the destinations. Denote O_i as the observed

marginal outflow for origin *i*, D_j as the observed marginal outflow for destination *j*, A_i and B_j as the balancing factors to ensure the constraints are met. Then these models can be written as:

Outflow constrained model:

$$T_{ij} = A_i O_i D(j) f(d_{ij})$$

$$A_i = \frac{1}{\sum_j D(j) f(d_{ij})}$$

$$\sum_j T_{ij} = O_i \qquad \text{for all i}$$

Inflow constrained model:

$$T_{ij} = B_j D_j O(i) f(d_{ij})$$

$$B_j = \frac{1}{\sum_i O(i) f(d_{ij})}$$

$$\sum_i T_{ij} = D_j \qquad \text{for all j}$$

All the existing applications of gravity models in studying human infectious disease epidemics use the unconstrained model, and they are discussed in the following section.

4.1.3 Application in human infectious disease epidemics

With the spatial spread of human infectious disease ultimately driven by human movement, travel flows between regions modelled by gravity models have been increasingly used to establish the heterogeneous contact rate between individuals in the subpopulations, assuming the contact rate to be proportional to the flow volume. Population sizes and travel distance are the most commonly used attributes in the gravity models that parsimoniously capture the coupling between the patches and are superimposed onto metapopulation epidemic models at various geographical scales.

On the national scale, gravity models were applied in the studies of measles (Xia, Bjornstad et al. 2004) and seasonal influenza (Viboud, Bjornstad et al. 2006) epidemics, and reproduced the regional spatiotemporal patterns which could not be achieved by using a distance function alone. A gravity type of spatial coupling was also used to model the spatial spread of the 1918 influenza pandemic in the UK, but proved less satisfactory as a description of the US pandemic (Eggo, Cauchemez et al. 2010). That study also found that connectivity between cities was density dependent, that is the total connectivity of a city depends on the number of close neighbours. However another study of the epidemic patterns of measles in the coastal regions of UK (Bharti, Xia et al. 2008) using Xia et al.'s model found that the gravity model underestimated the flows to close neighbours, and that having less neighbouring areas does not make a region less connected. A less vigorous gravity model

application was used in another pandemic study (Ciofi degli Atti, Merler et al. 2008)which fitted a gravity model to Italy's categorical work and study trip data (within municipal, outside municipal but within province, outside province but within same region, outside region). However, only the derived spatial kernel from the model was used in the model to determine the distance dependent contact rates.

At a continental scale, influenza pandemic spread in Europe was modelled with an embedded gravity model of the railway transportation data between countries, with the gross domestic product (GDP) of the countries fitting the flows better than population size (Merler and Ajelli 2010). In global metapopulation models, a gravity model of local commuting patterns of 29 countries integrated with the long-range air traffic network data underlies the multi-scaled "Global Epidemic and Mobility model" (GLEaM). This model was used to simulate the spread of epidemics worldwide (Balcan, Colizza et al. 2009), study the early stage of the influenza A/H1N1 pandemic in 2009 (Colizza, Vespignani et al. 2009), and investigate the strain's potential for seasonal transmission (Balcan, Hu et al. 2009).

The gravity models applied in all of these studies were of the general form:

$$T_{j\to k} = k\frac{N_j^{\alpha} N_k^{\beta}}{d_{jk}^{\gamma}}$$

where $T_{j\to k}$ denotes the flow from patch j to k, N_j and N_k are the population size of patch j and k, d_{jk} is the spatial distance between patches *j* and *k,* and α, β, and γ are the scaling factors of the origin and destination population and the travel distance respectively, and k is a constant. In the GLEaM, an exponential distance function $f(d_{jk})=\exp(d_{jk}/r)$ performed better than the power law function $g(d_{jk})=d_{jk}^{\gamma}$ as the kernel, and in Xia *et al*.'s study on measles (Xia, Bjornstad et al. 2004), the predicted number of infectives I_j in the origin patch was used instead, with the gravity model being combined with a time-series susceptible-infected-recovered (TSIR) model.

Different assumptions relating the human movement described by gravity models to disease transmission were implicitly applied in each of these studies. The gravity model is usually embedded in the force of infection term of the epidemic model, with its coupling effects with each region summed up to give the total transmission potential from the heterogeneous level of interaction between the subpopulations. As the compartmental models categorizes individuals into the states of susceptible, infective and others, the nature of gravity model is always directional – from an origin to a destination - and the flow matrices modelled are usually not symmetric. It is therefore essential to decide who travelled from their residing patch and to where, and possibly also where the contact

takes place. A few studies (Xia, Bjornstad et al. 2004; Eggo, Cauchemez et al. 2010) used the "infector driven approach", assuming it is the infected person that travels to other patches and infects the susceptible person there. Thus the total force of infection experienced by a susceptible is proportional to the sum of the flow from other patches to the susceptible's patch. The opposite is the "susceptive driven approach", assuming the susceptible individuals travel to other patches and become infected from the infectious people there (Merler and Ajelli 2010). Therefore the origin patch and the model structure was defined differently under these two approaches, though both of them assumed the movement to be unilateral in nature, with either the susceptible or the infected travelling, and those living in the destination remaining in their patches. However, these assumptions could be an oversimplified depiction of reality.

The framework of capturing "bilateral movement" in epidemic models – that allows for the possibility that the susceptible travels to the infector's patch and becomes infected there (susceptible driven approach), and the infector travels to the susceptible's patch and spreads the infection (infector driven approach), was proposed and developed using the spatial interaction theory (Thomas 1999). It defined the measure of regional attractiveness of region j as a destination for individuals travelling from region i as $T_{i\rightarrow j}=n_j\exp(-rd_{ij})$, where n_j is the population size of j, d_{ij} the distance between regions i and j, and r is the distance decay parameter. Then the probability of a given contact by an individual in i is made in region j, denoted as $p_{i\rightarrow j} = n_j\exp(-rd_{ij})/\sum_j n_j\exp(-rd_{ij})$. Conversely, the probability that an individual in region j travels to region i and makes contact there is $p_{j\rightarrow i} = n_i\exp(-rd_{ij})/\sum_i n_i \exp(-rd_{ij})$. The sum of both terms then forms the contact probability under a bilateral movement assumption. As an alternative to using the spatial interaction model, Keeling and Rohani (Keeling and Rohani 2002) used assumptions about the leaving and returning rate of individuals from their home patches to describe the bilateral movement in a hypothetical study of a population with two patches. However, this movement assumption restricted the location of infection to be at the home patch of either the susceptible or infective.

Sattenspiel and Dietz (Sattenspiel and Dietz 1995) developed a more comprehensive framework that captures the movement of both the susceptible and infective and allows infection between them to occur in any regions they both travel to – referred as "multilateral movement" in this study - using the rates of leaving and returning from home patches. Such information was assumed to be available from other sources like travel data or survey, and the method was first applied to study the influenza epidemic in 1918-1919 in Canada (Sattenspiel and Herring 2003). The GLEaM model was built upon the work of Sattenspiel and Dietz with the commuting flow derived from a gravity model. However, it remains uncertain how the various movement assumptions affect the results of the

epidemic models, or if under certain condition the much simpler unilateral assumption would be sufficient to describe the heterogeneous contact rate.

Gravity coupled epidemic models can either have the parameters of the gravity model derived independently from the origin-destination travel data (Viboud, Bjornstad et al. 2006; Balcan, Colizza et al. 2009; Merler and Ajelli 2010), or simultaneously estimated with the epidemic model parameters as an "endogenous system" (Xia, Bjornstad et al. 2004; Eggo, Cauchemez et al. 2010). Although at the expense of heightened technical challenges and computational complexity, the latter does not rely on the availability of the relevant travel flow data, and can serve as a means to validate the general assumption of gravity coupling. However, when a standard gravity relationship is not adequate to describe the flow dynamics, for instance, if long distance flow beyond several hundred kilometres exhibits different gravity behaviour than the short distance flow (Viboud, Bjornstad et al. 2006; Balcan, Colizza et al. 2009), gravity models fitted as a standalone statistical model allows model adequacy to be diagnosed on a finer level.

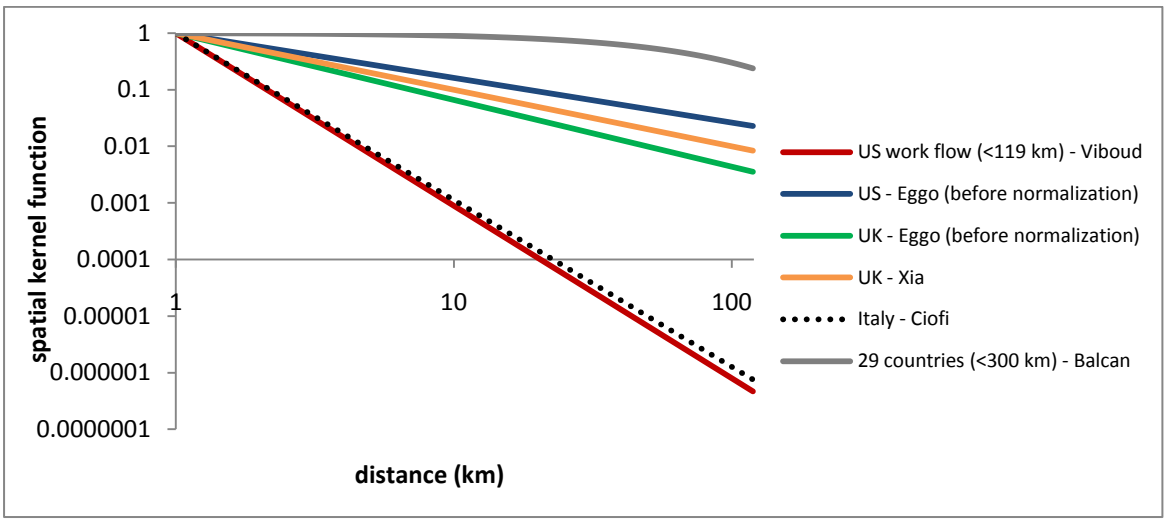
The interpretation of the gravity model parameters was also different under the two methods. In an "endogenous" model, it is the contact rate between individuals in different subpopulations that is directly modelled by the gravity component, whereas the exogenous gravity model represents the factors driving the flow, and the flow is used as the surrogate of the heterogeneous contact rate, assuming they are related proportionally. It remains uncertain how valid this assumption is. Contact surveys (Mossong, Hens et al. 2008) have shown that the number and type of contacts made in different settings vary (e.g. school and work). Therefore contacts generated related to a work trip may be different from that of a study trip.

Independently fitted gravity models have fairly consistent estimates of the origin and destination population exponents (α,β) – (0.3, 0.64) for US work flow of less than 119km (Viboud, Bjornstad et al. 2006), (0.28, 0.66) for work and study flows of Italy (Ciofi degli Atti, Merler et al. 2008), (0.46, 0.64) for commuting patterns of less 300 km in 29 countries. Long distance flow generally had lower α and β values (Viboud, Bjornstad et al. 2006; Balcan, Colizza et al. 2009). Endogenous estimation resulted in the population exponents to be insignificant for US, and at (0, 0.40) for UK (Eggo, Cauchemez et al. 2010). Fixing β=1 resulted in α=1.5 when the number of infective was directly used as the origin attribute (Xia, Bjornstad et al. 2004). This unique finding of the α/β ratio greater than 1 opposed to other studies could be partly attributed to the size of infective group (measles predominantly affecting children). In general, trip flows are more sensitive to the size of the destination population than the origin population with a higher population exponent in most cases, though the increment in both the propulsiveness and attractiveness of the regions decreases with

greater population size. Therefore small populations are disproportionally more important than large populations.

The spatial kernels of the gravity model from the above studies were reproduced based on the kernel parameters available in the studies (Figure 4.1). All studies except Eggo et al. apparently used the non-normalised form of the spatial kernel but with a constant term in the gravity model formulation. Therefore, although model constant or the kernel normalisation factors were not available, reproduced kernels for the same country or those with similar geographic scale can provide general ideas about the shape of the kernels. Eggo et al. normalised the origin population and the spatial distance together for each country, therefore the US kernel before normalisation is not comparable to the UK kernel before normalization. The US short distance work flow kernel was found to be almost identical to Italy's work and study flow kernel, while the two UK kernels for year 1918 and 1940-60s' were close to each other. Balcan et al's common short distance kernel for 29 countries based on contemporary data was much slower decaying than the others. Spatial kernels implicitly capture the transportation infrastructure of the country under study during the study period, and are also affected by the spatial scale of the study units. The common kernel of Balcan et al. was based on commuting data including small countries like Hong Kong with maximum travel distance less than 40 km and large countries like US (>8000 times of the physical size of Hong Kong), each of different geographic division – from 18 districts in Hong Kong, 3141 districts in US, to 36602 districts in France. Therefore the derived kernel has to be applied with caution. The study also found that an exponential kernel fitted the data better than a power law kernel, while all other studies simply assumed the kernel to be distributed as a 1-parameter power law function.

Figure 4.1 Spatial kernels for gravity models from previous epidemic studies reproduced, with the country and type of flow and the first author of the studies specified.



Besides being country specific, the optimal type of flow to be fitted by gravity models for studying epidemic spread should also be disease specific, determined by the dominant population subgroup for disease transmission (Viboud, Bjornstad et al. 2006). Existing gravity models fitted specifically to Hong Kong do not fit into the context of this study, neither were the spatial kernel parameters given (Wong, Tong et al. 1999).

4.2 Alternative methods to capture spatial heterogeneity in contact rates

Spatial distance kernels have also been applied to capture the spatial heterogeneity in contact rates (Ferguson, Cummings et al. 2005; Ciofi degli Atti, Merler et al. 2008). A variant of this is to assume that the average contact rates are dependent on categorised subpopulations using spatial proximity on a coarse resolution (Riley, Fraser et al. 2003), for example, within district, between contiguous districts, between non-contiguous districts. The distance only dependence assumption of contact rate is supported by two prominent studies that generalised human movement patterns as a power law distribution regardless of the geographic distribution of the population, using the banknote circulation (Brockmann, Hufnagel et al. 2006) and mobile phone data (Gonzalez, Hidalgo et al. 2008), though a single trajectory of the bank note circulation could be composed of several journeys by different people, whereas a displacement in the mobile phone data could be a partial or composite journey.

Brockmann et al. also investigated if the circulation patterns could be attributed to spatial inhomogeneities, such that travel patterns varied with population size or density. They compared the short-time trajectories (1-4 days) after initial entry by fitting separate power law distributions to the short ($\leq$10 km) and long (10km+) dispersal and by different population size categories of the entry city, and found that higher populated cities had a higher probability of short distance dispersal, but the long distance kernel exponents were comparable. However, on comparing the composition of the range of dispersal of the bank notes based on another snapshot of the short-time trajectory (<14 days) findings, I found that there were significant difference in the proportions of long range (>800km, chi-square test p-value=0.004) and also short range ($\leq$10km, chi-square test p-value < 0.0001) dispersal among the highlighted cities in the study (Long range: Jacksonville 2.9%, New York 7.4%, Seattle 7.8%; Short range: Jacksonville 71.4%, New York 57.7%, Seattle 52.7%). The population size of the entry cities did not seem to explain the difference (Jacksonville 0.8 million, New York 8.4 million, Seattle 0.6 million (United States Census Bureau 2009)). Unfortunately data were not available to investigate further for any potential role of the population size of the secondary reporting cities, i.e. the “destination” city of the circulation, and thus, the possibility of a gravity type of circulation patterns. However, other potential confounding effects could not be ruled out.

Gonzalez et al.'s mobile phone study finding that the power law distribution of human mobility was a convolution of the probability distribution of individual trajectories and population heterogeneity was an alternative perspective to dissect the underlying mechanism for the observed power law distribution.

Alternatively, instead of using any functional form for the contact rates, some studies used the travel flow data directly to represent the contact rates (Danon, House et al. 2009). Although this avoided any issues related to the validity of gravity model specification and the goodness-of-fit of the model, it lost the additional information and insight on how the contact rates and subsequently the patterns of the epidemic varies with its attributing factors. Comparison with other studies is also made more difficult under this approach.

Different assumptions of human movement patterns result in different patterns of epidemic spread. If human movement is purely distance dependent, a wave-like diffusion pattern with occasional long jumps will occur, and the frequency of long jump occurrence is dependent on the tail of the distance kernel. Under a gravity model, the diffusion pattern is expected to be a mixture of wave-like diffusion and population size hierarchy, as observed in different influenza epidemics (Cliff, Haggett et al. 1986; Smallman-Raynor, Johnson et al. 2002). The assumption of human movement driven by distance alone is not adequate to explain the observed patterns of the SARS outbreak. In the previous chapter I showed that the spatial spread of SARS in Hong Kong was a mixture of a wave-like diffusion and population hierarchy. In Beijing, the spread was driven by spatial contiguity and population density (Meng, Wang et al. 2005). Hospital-related factors also drove the spread in both regions.

4.3 Study objectives

Building on previous studies that used gravity model to capture the heterogeneous contact rates between subpopulations through the underlying factors for the human flow patterns, the first objective of this chapter is to derive gravity models to represent the contact rates relevant to SARS transmission in Hong Kong. These are then incorporated into a modified Wallinga and Teunis method for estimating the reproduction number time series of the Hong Kong SARS outbreak that uses both the spatial and temporal information of the epidemics in Chapter 5. In the next chapter, I also study the effects of trip type and the movement assumptions (susceptive driven, infection driven, bilateral and multilateral movement) on the R_t estimate. Therefore in this chapter I present the derivation for each relevant type of movement data and present parameter estimates for the best fitting gravity model under each model form.

As demonstrated in previous chapters, a high proportion of SARS transmission in Hong Kong occurred in hospital settings, and I found that hospital transmission could trigger community transmission and formation of disease cluster. Therefore my second objective is to use gravity models to investigate whether different types of hospital infected cases (healthcare workers, inpatients and hospital visitors) exhibited different patterns of flow from home to the hospital they were infected with SARS, which would indicate the spatial range of secondary transmission from the hospital to the community. The results are useful in explaining the observed epidemic spread patterns and have implications for disease control.

4.4 Data

This section describes the census, travel survey and the synthetic datasets of travel patterns of the Hong Kong population used to study the first research objective, and the home-to-hospital flow data of the hospital infected SARS cases in Hong Kong used for the second objective. This is followed by a brief description on the geo-demographic characteristics of Hong Kong to provide the background information for this study.

4.4.1 Census travel data

The workflow and study flow data from the 2001 Hong Kong census (Census and Statistics Department Hong Kong Special Administrative Region 2001) were defined spatially by the District Council District (DCD), referred to here as "districts". (Figure 1.1). The work and study flow data were composed of an 18x18 origin-destination flow matrix, with the residing district of the subjects as the origin. For the workflow data, the destination was the district of the workplace of the subject. For people with more than one job, more than one workplace, or who changed workplace regularly, workplace referred to that of their main employment, or with the longest hours spent during the census reference period. For the study flow data, the destination was the district of the school or educational institution the subjects attended for full-time education. As each working person contributed one trip to the workflow data, and each full-time student contributed one trip to the study flow data, the work outflow and study outflow of each district were the working population and full-time student population residing in that district, and the work inflow and study inflow of each district were the number of people working or studying in that district, regardless of where they lived. The sum of the work and study flow gave the work-study flow of the population. Spatial distance between the districts was calculated using the Euclidean distance between the centroid of the districts.

4.4.2 Travel survey data

The travel characteristics survey 2002 in Hong Kong (Transport Department. The Government of the Hong Kong Special Administrative Region. 2002) was conducted in households in area segments selected by random sampling. Travel data on a particular weekday of all household members aged above 3 years old were collected from the 30,005 households selected (1.39% of total households in Hong Kong) by face-to-face interview. The average household size was 3.1 persons. The data were then used to project the travel patterns of the whole population of Hong Kong. The reporting geographic units were 25 districts. The difference of this geographical division compared to the 18 District Council Districts used in the census was mainly in the New Territories and Islands, where the new towns were separated from the rural areas.

Data on all the trips made in a particular weekday were collected, outbound and return trip inclusive. All trips were classified by trip purpose as: (1) home-based work (work flow): trips from home to work place, or vice versa; (2) home-based study (study flow): trips from home to school place for attending lectures or lessons, or vice versa; (3) home-to-others: trips from home to non-work and non-study related places, e.g. shopping places, visiting friends etc., or vice versa; (4) non-home based: trips that did not start or end at home, nor between work places, e.g. from work place or school to shopping or other social places; (5) employer business: trips between work places for work. I refer to the survey data that includes all the flow types above as the "survey overall" flow.

The mean daily number of work trips per working person was 1.40, and study trips per student was 1.13, indicating that a certain proportion of people did not return home directly after work or study, and/or did not start the work and study journey from home. The average daily number of home-based-others trip was 0.57 per person, and non-home-based trip was 0.23 per person. The average daily trip per person is 1.82, with around 15% of the population not making any trip on an average weekday. The composition of the total flow by trip type was: home-based work 37.5%, home-based study 12.8%, home-based others 31.6%, non-home based 12.9%, and employer business 5.2%.

The centroid of the districts was derived by an ad hoc approach for this study, because no data were available for re-generating the map digitally. Although the survey data is a more comprehensive representation of the movement of individuals in the Hong Kong population, its utility was greatly restricted with only the marginal inflow and outflow by district (25x2 matrix) for each trip type reported.

4.4.3 Adjusted overall census data

Contact between individuals could be made without making a trip, for instance, in a household setting, which accounted for 10% of community transmission of SARS in Hong Kong (chapter 2). However, these contacts are likely to be under-represented if the above datasets are used as surrogate of contact rates between individuals, because the travel survey trip data exclude those who did not make any trips during the study timeframe as well as those aged below 3, and the census data exclude those who did not work or study. This was not an important issue in the previous epidemic studies that used gravity models fitted to census data, because they were used only as a proxy for the between-patches contact patterns. But as the modified Wallinga and Teunis method developed in the next chapter requires the actual within-patch contact rates, it is necessary to adjust the trip data. A synthetic trip dataset was therefore created based on the census data, with the estimated population which did not make a trip added as the within-district flow. This resulted in the total movements being the sum of the census work flow, census study flow, and 15% of the population of each district plus those aged below 3 according to the census as the within-district flow. Those who are retired or the unemployed elderly are included in the 15% if they did not make any trip. This synthetic dataset is referred as the "adjusted overall census" data.

4.4.4 Home-to-hospital travel data

The home-to-hospital flow matrix of the healthcare worker, inpatient and hospital visitor respectively captured the flow from the home district of the SARS cases to the hospitals where they were infected. The data were derived from the source of infection data described in chapter 2. Nine hospitals with less than 10 occurrence of SARS infection were excluded in the analysis, leaving the dimension of the flow matrices as 18 (districts) x 12 (hospitals). The overall composition of hospital infected cases were 49.7% healthcare worker, 31.0% inpatient and 19.9% hospital visitors.

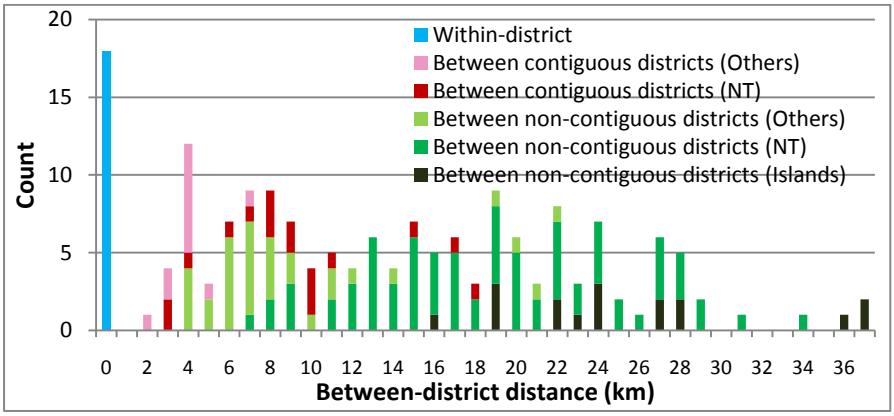
4.4.5 Geo-demographic characteristics

The districts in Hong Kong Island (HKI), Kowloon (KLN) and the New Territories (NT) have marked difference in their geographical and demographic characteristics (Figure 1.1), and their size varies from 7 km^2 to 175 km^2. The smallest and the most densely populated districts are in Kowloon, while the districts in the New Territories districts are highly heterogeneous in population density, with satellite cities surrounded by rural areas. The main commercial and financial areas of Hong Kong are located in the Central and Western, Wanchai and Yau Tsim Mong districts, which are along the harbour between the Hong Kong Island and Kowloon. Wanchai and Yau Tsim Mong are also the most popular and concentrated shopping and entertainment areas. The Islands district is remote and the

least populated district (1.3% of total population), but has disproportionally high travel flow because of the international airport, and several housing estates and tourist spots there, but it is otherwise undeveloped. All districts in Hong Kong are coastal except Wong Tai Sin.

The between-districts distance of Hong Kong spans from 2km to 37 km (mean=15.4km, sd=8.1km, Figure 4.2). The absence or low frequency of distance of between 1 and 3 km is attributed to the physical size of the districts. All distance of 22km and beyond is made from or to at least one of the New Territories districts or the Islands district.

Figure 4.2 Distribution of between-district distance of Hong Kong based on the District Council District division. Measures were grouped in the hierarchy of: (1) Islands: if one of the district measured is the Islands district; (2) NT (the New Territories): if at least one of the district measured is located in the New Territories; (3) Others: (Hong Kong Island and Kowloon): if the distance measured does not belong to (1) or (2).



4.4.6 Data issues

The census flow data have far better data resolution than the travel survey data but are less comprehensive because mainly work and study trips are covered. Home-based other trips accounted for around one-third of the total trips in the travel survey, and 44.4% together with non-home based trips. However, it is not possible to disentangle the zonal trip propulsiveness and attractiveness in the survey data. To illustrate this, consider Hong Kong to be divided into 3 districts. Let C be the matrix of the census work flow c_{ij} from district i to j, S be the matrix of the survey work flow s_{ij} between district i and j, x_{ij} be the proportion of working population living in i and working at j who returns home directly after work. Then theoretically the matrix S can be written in relation to C as:

$$C = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix}$$

$$S = \begin{pmatrix} (1+x_{11})c_{11} & c_{12}+x_{12}c_{21} & c_{13}+x_{13}c_{31} \\ c_{21}+x_{21}c_{12} & (1+x_{22})c_{22} & c_{23}+x_{23}c_{32} \\ c_{31}+x_{31}c_{13} & c_{32}+x_{32}c_{23} & (1+x_{33})c_{33} \end{pmatrix}$$

This shows that S_{ij} does not only carry the information on the propulsion of district *i* and attractiveness of district *j* for work (c_{ij}), but also to some extent the propulsion of district *j* and attractiveness of district *i* (c_{ji}). Therefore, the estimated scaling exponents do not truly reflect the trip attributing factors of the districts, and the extent of offset depends on x_{ij}. Without additional data to disentangle the offsetting effect, the application of the survey data in this study was under caveat.

Given the strengths and weaknesses of both datasets, I compared the census and survey data of the same trip purpose to quantify the potential effects of return flow ("returning effect") in the survey data, and investigated whether the census data could be a good surrogate of the general population movement measured by the survey overall data To begin with, exploratory analysis on the characteristics of the various flow patterns was conducted to facilitate the above investigation and the fitting of gravity models.

4.5 Preliminary Analysis

Marginal flow of the same movement type of the census and survey data were compared by Pearson's correlation coefficient, with the different geographical division of the datasets unified as 14 redefined spatial units. Both the outflow and inflow of the work and study trip respectively were very highly correlated [Work flow: $\rho_{outflow}$=0.995, ρ_{inflow}=0.996; study flow: $\rho_{outflow}$=0.977, ρ_{inflow}=0.907, all p-values<0.0001] (Figure 4.3), suggesting that sampling and projection of the survey data were satisfactory in representing the whole population, and the correlation remained strong despite the returning effect in the survey data.

The outflow of all census trip data were highly correlated with the population size (ρ_{work}=0.81, ρ_{study}=0.74, $\rho_{work\text{-}study}$=0.79, $\rho_{adjusted\ overall}$=0.78, all p-values<0.005), but for inflow the correlation with population size was insignificant for work (ρ=0.24, p-value=0.40) and work-study trips (ρ=0.51, p-value=0.07), and moderately correlated for the study trip (ρ=0.71, p-value=0.004) and adjusted overall trip (ρ=0.64, p-value=0.01). The insignificant correlations were confounded by two different scaling relationships between the inflow and population size, where a few districts showed much higher inflow per population than the other districts ("high inflow districts") [Figure 4.5].

For work flow, the high inflow districts included Wanchai, Central and Western, Yau Tsim Mong and Islands, which are the key financial and commercial centres except Islands. For these districts the inflow was 1.97 to 3.45 times the working population residing in that district. Their within-district work flow (34.7%-49.7%) was also high compared to other districts [Figure 4.4(a)]. For study inflow, the high inflow districts were Kowloon City, Yau Tsim Mong, Central and Western, Wanchai and

Figure 4.3 The scatter plots of census and survey flow of the same trip type by outflow and inflow.

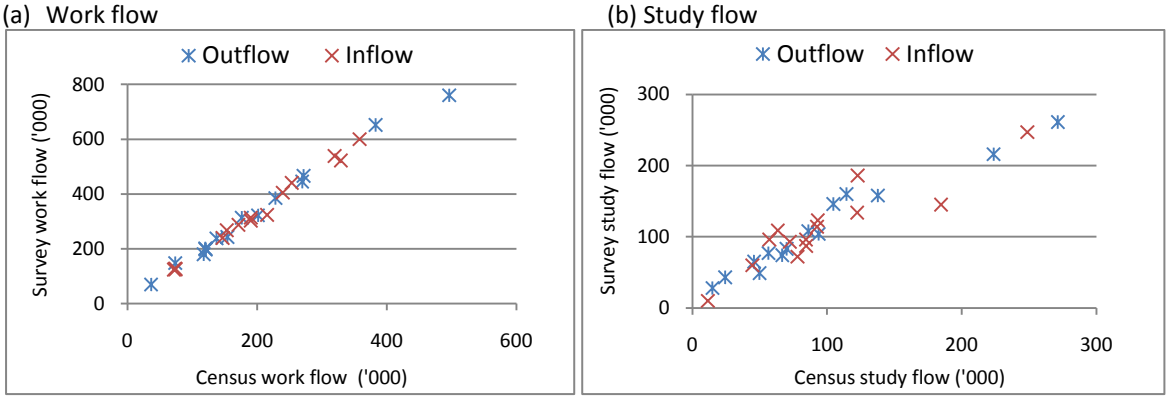
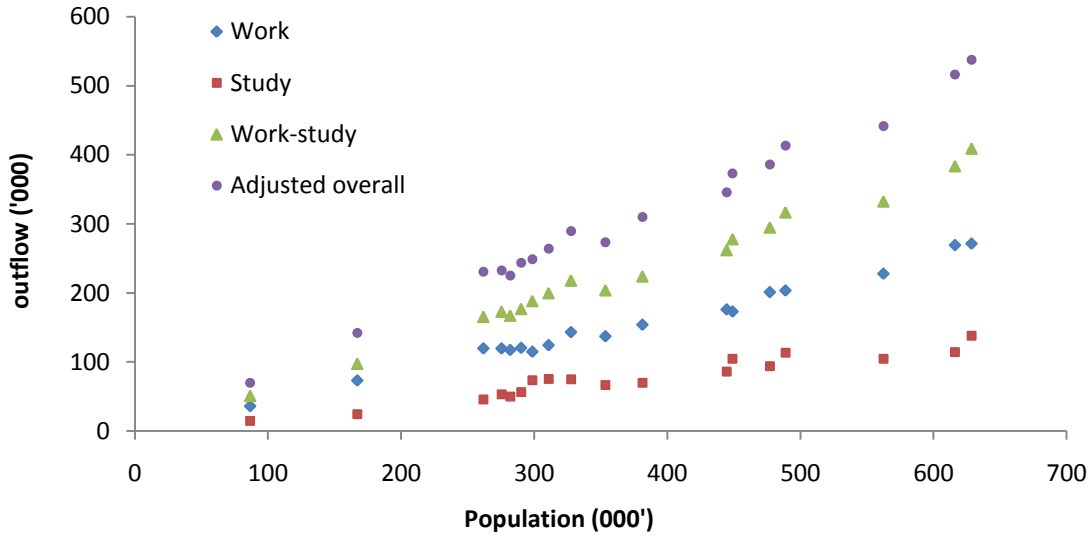


Figure 4.4 Census outflow of different trip types by population size of the district.

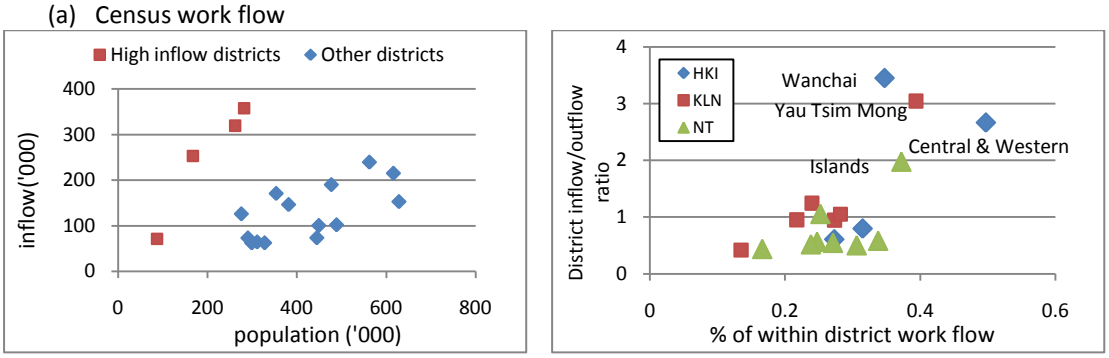


Sham Shui Po, which had their inflow 1.26 to 2.36 times over the student population living in the district, in contrast to a consistent average ratio of 0.85 (sd=0.6) for the rest of the districts [Figure 4.5 (b)]. The majority of the districts in the New Territories had high within-district study flow (63.6%-86.2%), attributed to the satellite cities in the districts which were designed to be self-sufficient in study places (Hui and Lam 2005). Study flow was generally more localised than the work flow (within-district flow: 59.4%-86.2% versus 13.5%-49.7%), as a substantial proportion of work flow was drawn to the key commercial districts in Hong Kong.

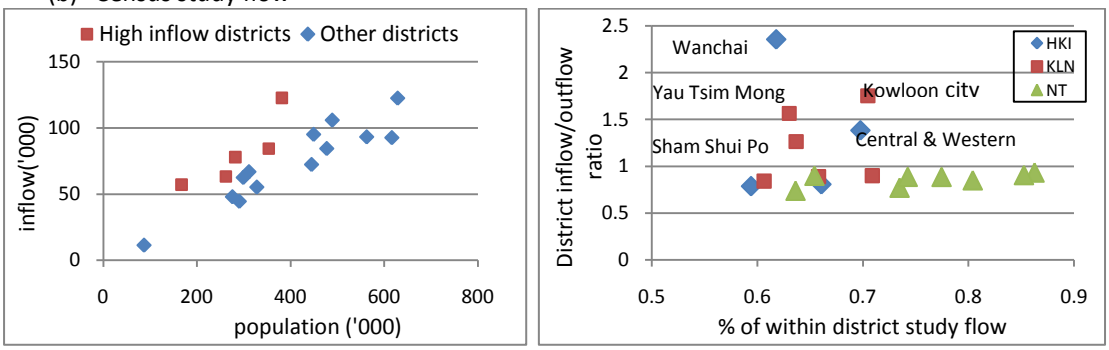
The census work-study flow data had the ratio of work to study trips of 2.05:1 and therefore exhibited mixed characteristics more dominated by workflow, but with an increased proportion of within-district journey. The census adjusted overall data had the within-district flow further increased to between 46.1% and 67.9%. High inflow districts for both trip types were identified.

Figure 4.5 Left: Inflow by population size of the districts; Right: Inflow to outflow ratio by percentage of within-district flow, by census flow type: (a) work flow, (b) study flow, (c) work-study flow, and (d) adjusted overall flow.

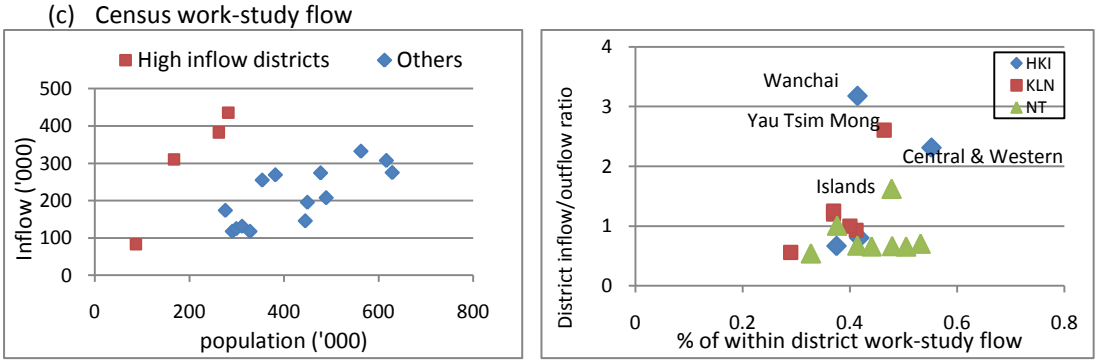
(a) Census work flow



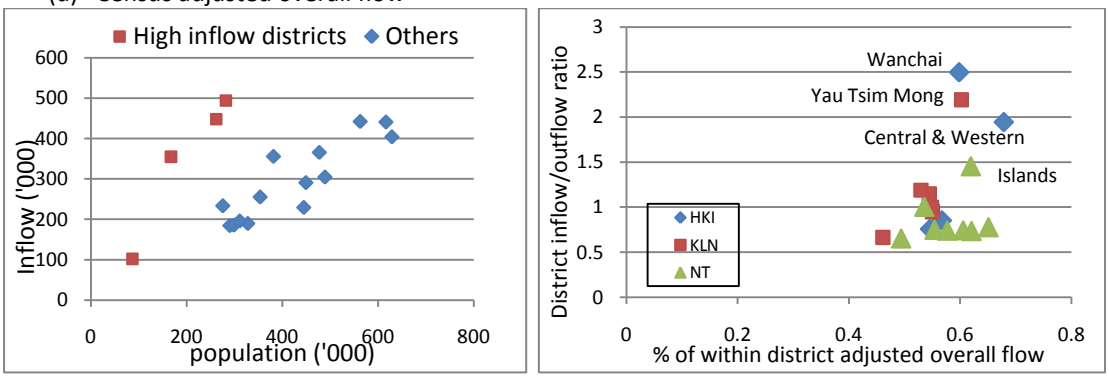
(b) Census study flow



(c) Census work-study flow



(d) Census adjusted overall flow



For the trip types that were studied in the travel survey only, the outflow of home-based others trips was strongly correlated with the population (ρ=0.83, p-value=0.0002), but not the inflow (ρ=0.14, p-value=0.62). Inflow relative to the population size was high in Yau Tsim Mong, Wanchai and Islands, where the first two districts are also the most popular areas for leisure activities and shopping [Figure 4.6 (a-b)]. Non-home based outflow and inflow was not correlated with the population size at all ($\rho_{outflow}$=0.12, p-value=0.69; ρ_{inflow}=0.03, p-value=0.92). Its peak trip making time was between 12pm to 2pm and 5pm to 7pm (Transport Department. The Government of the Hong Kong Special Administrative Region. 2002), and high outflow numbers were observed in Yau Tsim Mong, Central and Western, Wanchai, Kowloon City and Islands, which were also the areas of high work or study inflow [Figure 4.6(c)]. These suggested that a substantial proportion of non-home based trips were made during lunch break or after work or study. The most popular destinations were Yau Tsim Mong and Wanchai [Figure 4.6(d)].

The overall flow in survey data was highly correlated with the population size in its outflow (ρ=0.81, p-value=0.0005), but not correlated in the inflow (ρ=0.24, p-value=0.42). It also had the same districts of high inflow as in the census work and work-study trips [(e-f)]. Its high correlation with the census work flow [$\rho_{outflow}$=0.99, ρ_{inflow}=0.91, all p-values < 0.0001] supports the use of the census work data as a surrogate for the overall movement patterns, and so was the work-study flow data [$\rho_{outflow}$=0.98, ρ_{inflow}=0.87, all p-values <0.0001]. The census adjusted overall trip was slightly less but nevertheless still highly correlated with the survey data [$\rho_{outflow}$=0.98, p-value <0.0001; ρ_{inflow}=0.74, p-value=0.002].

The markedly different characteristics of the high inflow districts compared to other districts implied a dichotomous scaling relationship might be required for the gravity model fitting that used population size as the regional attribute. Alternatively, the marginal flow (inflow or outflow) can be used directly as the regional attributes for that specific flow type, as it captures the features that attract or propel journeys which may not be strongly correlated with population size, for instance, office space.

4.6 Methodology

4.6.1 Model formulation - Hong Kong travel patterns

Outflow constrained, inflow constrained and unconstrained gravity models were fitted to the different flow data. To capture the dichotomous effect of the attribution factors on the flow by districts with high inflow, I used an adjustment scaling effect for modelling both the high within-district and between-district flows of these districts. Denote *T* as the total observed number of trips,

Figure 4.6 The outflow (left) and inflow (right) by population size of the home-based others, non-home based and overall trips in the survey data. The straight lines represent the least square regression of flow=constant x population size.

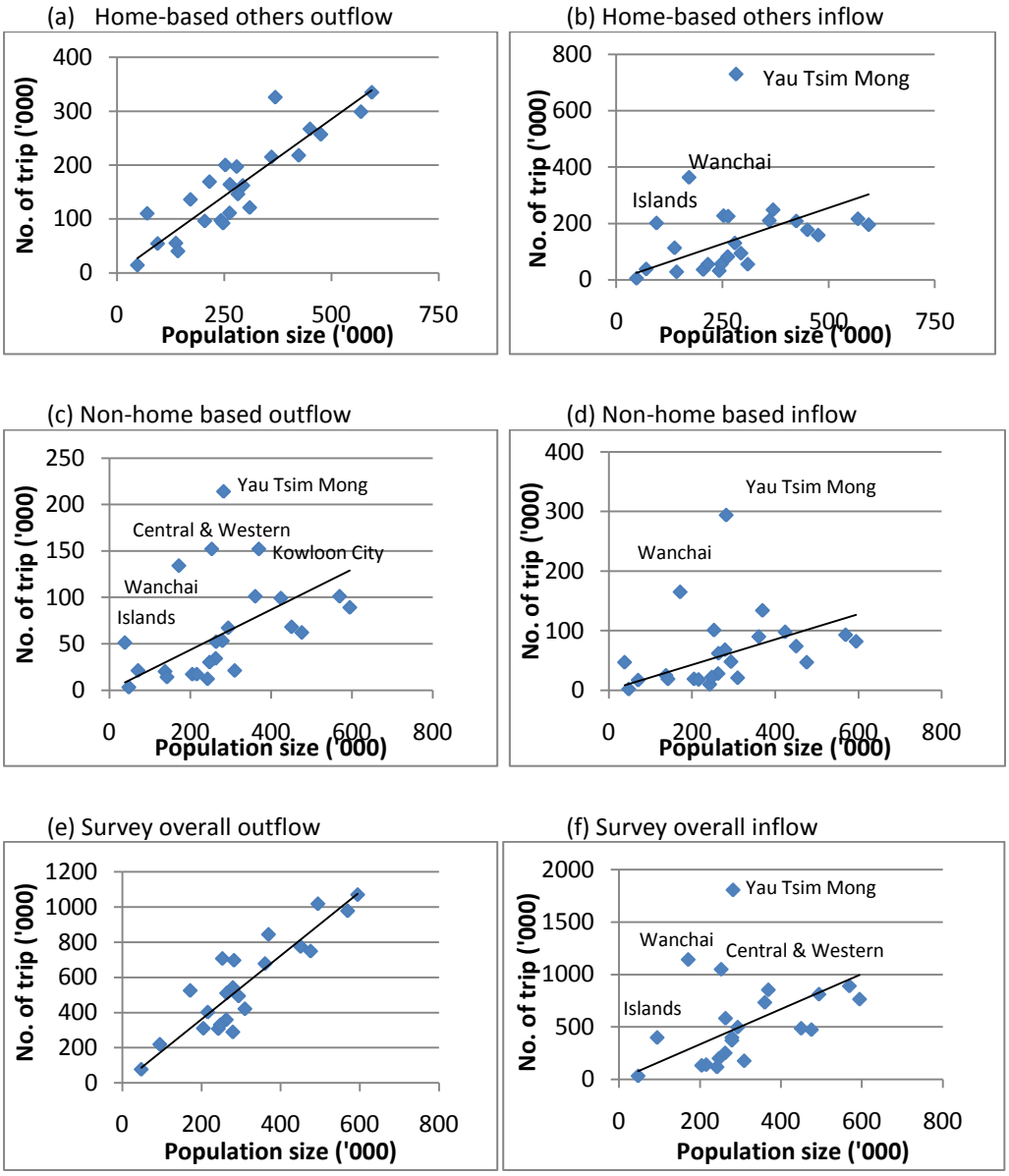
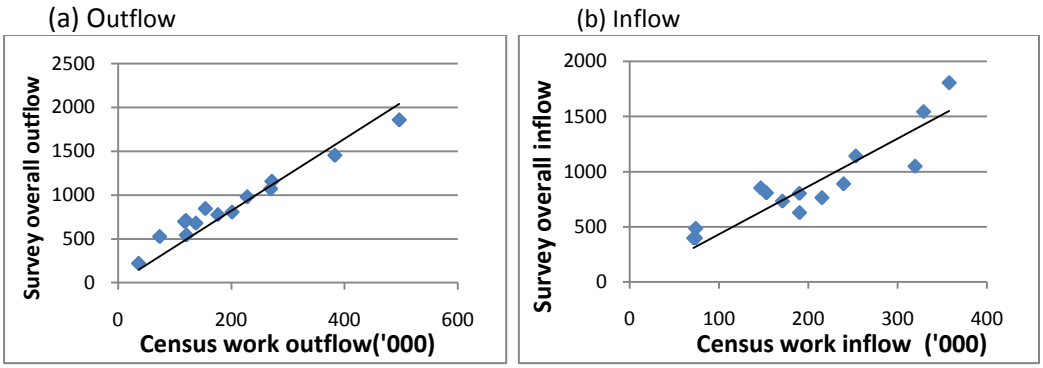


Figure 4.7 (Left) Survey overall outflow by census work outflow; (Right) Survey overall inflow by census work inflow. The straight lines represent the least square regression of the survey overall flow=constant x census flow.



$T_{ij,,}$as the flow from origin *i* to destination *j*, O_i as the observed outflow from region i, D_j as the observed inflow from region j, N_i and N_j as the attributing factor of region i and j respectively, with their effect scaled by non-negative exponent β_1 and α_1, and the scaling adjustment factors β_2 and α_2 respectively, and $f(d_{ij})$ as the distance deterrence function. The gravity model formulations are then written as:

(a) Outflow constrained model

$$T_{ij} = \frac{O_i N_j^{\beta_1+\beta_2} f(d_{ij})}{\sum_j N_j^{\beta_1+\beta_2} f(d_{ij})} \qquad \beta_2 = \begin{cases} \beta' & \text{for districts with high inflow / outflow ratio} \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

$$\sum_j T_{ij} = O_i \qquad \text{for all i}$$

(b) Inflow constrained model

$$T_{ij} = \frac{D_j N_i^{\alpha_1+\alpha_2} f(d_{ij})}{\sum_i N_i^{\alpha_1+\alpha_2} f(d_{ij})} \qquad \alpha_2 = \begin{cases} \alpha' & \text{for districts with high inflow / outflow ratio} \\ 0 & \text{otherwise} \end{cases}$$

$$\sum_i T_{ij} = D_j \qquad \text{for all i}$$

(c) Unconstrained model

$$T_{ij} = cN_i^{\alpha_{u1}+\alpha_{u2}} N_j^{\beta_{u1}+\beta_{u2}} f(d_{ij}) \qquad \alpha_{u2} = \begin{cases} \alpha_u' & \text{for districts with high inflow / outflow ratio} \\ 0 & \text{otherwise} \end{cases}$$

$$\beta_{u2} = \begin{cases} \beta_u' & \text{for districts with high inflow / outflow ratio} \\ 0 & \text{otherwise} \end{cases}$$

$$\sum_{i,j} T_{ij} = T$$

Two measures of N_j were considered: (1) the population size, and (2) inflow (for the outflow constrained model); outflow (for the inflow constrained model); marginal flow (ie. inflow and outflow, for the unconstrained model). The distance deterrence function $f(d_{ij})$ is the normalised function of $k(d_{ij})$ with maximum distance of 50 km, decided based on the physical dimension of Hong Kong. Power-law and exponential functions are the most commonly used functional form of $k(d_{ij})$ in gravity models for human travel (Fotheringham and O'Kelly 1988; Sen and Smith 1995), and with proven success in studying infectious disease transmission of human (Xia, Bjornstad et al. 2004; Viboud, Bjornstad et al. 2006; Bharti, Xia et al. 2008; Balcan, Colizza et al. 2009). In this study, I used a more generic form of the power law function with two parameters, and the exponential function as $k(d_{ij})$, as show below. The better kernel form was decided by comparing the goodness-of-fit statistics of the models.

2-parameter power-law function

$$k(d_{ij}) = (1 + \frac{d_{ij}}{a})^{-b}$$

Exponential function

$$k(d_{ij}) = exp(-ad_{ij})$$

For work flow, work-study flow, adjusted census overall and survey overall flow models, the high inflow to outflow ratio districts included Wanchai, Central & Western, Yau Tsim Mong, Islands. For study flow models, they were Wanchai, Central & Western, Yau Tsim Mong, Kowloon City and Sham Shui Po.

Table 4.1 summarizes the variants of the specification of the gravity models fitted. For the survey overall flow data, only the population size was used as the attraction factor, because the dependent variable was the marginal flow. For clarity, I denoted the gravity model specification in the format of [gravity model structure]–[flow data fitted]–[attraction factor], with the abbreviation of the specification given in Table 4.1 (e.g. U-WS-P stands for unconstrained gravity model on the census work-study flow using population as the attributing factor).

Table 4.1 Summary of the gravity models fitted to the census and survey data.

Nature	Variant
Census data	
Gravity Model structure	1. Outflow constrained (O) 2. Inflow constrained (I) 3. Unconstrained (U)
Flow data T_{ij}	1. Work flow (W) 2. Study flow (S) 3. Work-study flow (WS) 4. Adjusted census overall (AC)
Attributing factor N_j	1. Inflow (I) / Outflow (O) / Marginal flow (M) 2. Population size (P) 3. Population size with high inflow district adjustment (P2)
Spatial kernel functional form $k(d_{ij})$	1. Power law 2. Exponential
Travel survey data (TS)	
Gravity Model structure	1. Outflow constrained (O) 2. Inflow constrained (I) 3. Unconstrained (U)
Flow data T_{ij}	Overall flow (TS)
Attributing factor N_j	1. Population size (P) 2. Population size with high inflow district adjustment (P2)
Spatial kernel functional form $k(d_{ij})$	1. Power law 2. Exponential

4.6.2 Model formulation- Home-to-hospital patterns

Because the spatial distribution of the SARS hospital infected subjects was largely driven by the epidemiological characteristics of the epidemic, I used the outflow constrained gravity model so that conditional on the number (i.e. outflow) of SARS hospital infected cases in each district, I study how the flow patterns were related to travel distance and hospital related factors for each type of hospital infected subjects (healthcare workers, inpatients and hospital visitors) and the overall flow. The gravity model formulation was similar to that of the outflow model in the previous section. Denoting T_{ij} as the home-to-hospital flow, H_j as the attribution factor of hospital j, the formulation can be written as:

$$T_{ij} = A_i O_i H_j^{\beta} f(d_{ij})$$
$$A_i = \frac{1}{\sum_j H_j^{\beta} f(d_{ij})} \qquad (4.2)$$
$$\sum_j T_{ij} = O_i \qquad \text{for all i}$$

The power law and exponential spatial kernel defined previously were applied. Two inflow attribution factors were considered: (1) the hospital size, using the number of hospital beds as the proxy; and (2) the "inflow" of the type of SARS hospital infected subject under study, which essentially captured the hospital and epidemic specific factors for the flow. The derived spatial kernel and the scaling factors for different hospital subject types were compared. Table 4.2 below summarises the specification of the 16 gravity models fitted to study the hospital flow.

Table 4.2 Gravity models fitted to the hospital flow data.

Nature	Variant
Gravity model structure	Outflow constrained
Flow data T_{ij}	1. Healthcare worker flow 2. Inpatient flow 3. Visitor flow 4. Overall flow
Attraction factor H_j	1. Hospital size 2. Inflow
Spatial kernel functional form $k(d_{ij})$	1. Power law 2. Exponential

4.6.3 Model fitting

Because the 2-parameter power law function kernel was used in this study, the gravity model parameters have to be estimated numerically. Assuming the flow is distributed as independent

Poisson or multinomial, maximum likelihood method can be applied for parameter estimation (Fotheringham and O'Kelly 1988; Sen and Smith 1995). Weaker assumptions are made under the multinomial distribution compared to the independent Poisson distribution, though both distributions result in the same maximum likelihood estimates of the gravity model parameters (Sen and Smith 1995). Simulation studies showed that the maximum likelihood estimates of the gravity model parameters remained reliable and robust in small sample size or slight deviation to the independent Poisson assumption (Sen and Matuszewski 1991).

Consider ***T***, the I x J matrix of the flows T_{ij} to be distributed as multinomial (T, ***p***), where T is the total number of the observed flows, and ***p*** is the matrix of $p_{ij,}$ the probability of flows from *i* to *j.* Denote T_{+j} as the marginal inflow, T_{i+} as the marginal outflow, and their vectors as $\mathbf{T_j}$ and $\mathbf{T_i}$ respectively. An outflow constrained gravity model implies that $\mathbf{T_i}$ = (T_{1+}, T_{2+}, …., T_{I+}) is given, and with T being multinomial distributed implies conditional multinomial distribution of $\mathbf{T_j}$ ~ multinomial ($\mathbf{T_i}$, $\mathbf{p_{j/i}}$) for all *j*, where $\mathbf{p_{j/i}}$ is the vector of probability of p_{ij} given p_{i+}, $p_{i+} = T_{i+}/T$ and $\sum_j p_{j/i} = 1$ for all i. The *I x J* matrix of flows is considered as a marginal multinomial for the outflow totals and *i* conditional multinomials for the *i* outflows, with the likelihood function written as:

$$L = \prod_i \left(\frac{T_{i+}!}{\prod_j T_{ij}!} \prod_j p_{j/i}^{\ T_{ij}} \right) \times \left(\frac{T!}{\prod_i T_{i+}!} \prod_i p_{i+}^{\ T_{i+}} \right)$$

It can be shown that

$$\ln L \propto \sum_i \sum_j T_{ij} \ln p_{j/i} \tag{4.3}$$

From the gravity models defined in equation 4.1,

$$\begin{aligned} p_{j/i} = \frac{p_{ij}}{p_{i+}} = \frac{T_{ij}/T}{T_{i+}/T} &= \frac{A_i O_i T_j^{\beta_1+\beta_2} f(d_{ij})}{\sum_j A_i O_i T_j^{\beta_1+\beta_2} f(d_{ij})} \\ &= \frac{T_j^{\beta_1+\beta_2} f(d_{ij})}{\sum_j T_j^{\beta_1+\beta_2} f(d_{ij})} \end{aligned} \tag{4.4}$$

Inserting equation 4.4 into equation 4.3 gives

$$\ln L \propto \sum_i \sum_j T_{ij} \ln \frac{T_j^{\beta_1+\beta_2} f(d_{ij})}{\sum_j T_j^{\beta_1+\beta_2} f(d_{ij})} \tag{4.5}$$

Similarly, it can be derived for the inflow constrained model that:

$$\ln L \propto \sum_i \sum_j T_{ij} \ln \frac{T_i^{\alpha_1+\alpha_2} f(d_{ij})}{\sum_i T_i^{\alpha_1+\alpha_2} f(d_{ij})}$$

For the unconstrained model,

$$L = \frac{T!}{\prod_{i,j} T_{ij}!} \prod_{i,j} p_{ij}^{T_{ij}}$$

and it can be implied that

$$\ln L \propto \sum_{i,j} T_{ij} \ln p_{ij}$$

and

$$\ln L \propto \sum_{i,j} T_{ij} [(\alpha_{u1} + \alpha_{u2}) \ln N_i + (\beta_{u1} + \beta_{u2}) \ln N_j + \ln f(d_{ij})]$$

For the home-to-hospital gravity models, the same method of model fitting was used, and T_j in equation 4.5 was replaced by H_j. The adjusted models had one more parameter γ to be estimated through the likelihood:

$$\ln L \propto \sum_i \sum_j T_{ij} \ln \frac{T_j^{\beta+\gamma} f(d_{ij})}{\sum_j T_j^{\beta+\gamma} f(d_{ij})}$$

The model parameters were estimated numerically using the Generalized Reduced Gradient nonlinear optimization algorithm (Fylstra, Lasdon et al. 1998).

4.6.4 Goodness-of-fit evaluation and model comparison

Evaluation of the goodness-of-fit of the models was performed both on the global and individual observation levels. Models fitted to the same observed flow were compared by deviance and standardized root mean square error statistic (SRMSE, equation 4.6), the latter having been identified as the best measure for spatial interaction model comparison compared to chi-square, R-square, and other less commonly used statistics (Knudsen and Fotheringham 1986). The observed and predicted probability distribution of the journey distance was compared graphically, and deviance residual plots were generated on the district and individual observation levels, categorized into within-district flow and between-district flow residuals.

$$SRMSE = \sqrt{\frac{\sum_i \sum_j (T_{ij} - \hat{T}_{ij})^2}{I \times J}} \Bigg/ \frac{\sum_i \sum_j T_{ij}}{I \times J} \qquad (4.6)$$

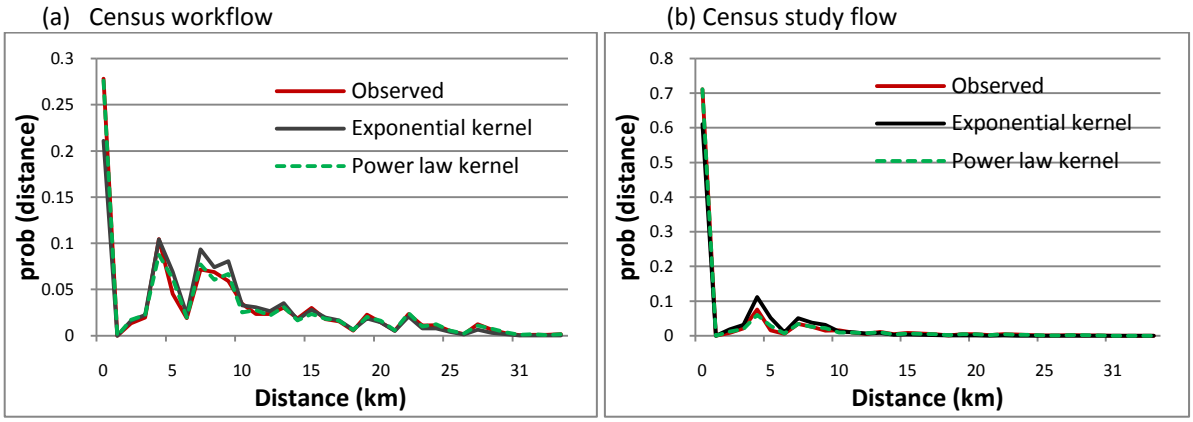
In situations where the difference in deviance was not adequately informative to distinguish between the performance of nested models, Akaike's information criterion (AIC) which penalizes for the number of parameters was used.

4.7 Results

4.7.1 Hong Kong travel patterns

In all gravity models fitted to the movement patterns of the Hong Kong population and regardless of the gravity model type, the 2-parameter power law kernel outperformed the exponential kernel in the goodness-of-fit statistics. The latter tended to greatly underestimate the within-district flows (i.e. zero journey distance), compared to the good fit of power law kernel models to the probability distribution of the journey distance, as shown in Figure 4.8 as examples of illustration. In spite of this, the estimates of the scaling exponents α_1, α_2, β_1 and β_2 were fairly consistent between the power law and exponential kernel models, showing the robustness of the estimates. The following discussion focuses on the power law spatial kernel models only.

Figure 4.8 Comparison of the gravity models using power-law or exponential kernel in estimating the probability distribution of journey distance. Outflow constrained gravity models using district-adjusted population size as the attributing factor were fitted to the census workflow and study flow respective (Models O-W-P2 and O-S-P2).



4.7.1.1 Outflow constrained gravity models

Table 4.3 summarises the parameter estimates and the goodness-of-fit statistics of the outflow constrained gravity models fitted to the five movement datasets. The deviance and SRMSE statistics indicate that the district-adjusted population size model is better than the inflow model for the census study flow. The reverse is suggested for the census work flow and census work-study flow. The adjusted census overall flow was inconclusive, but the residual plots of these three flow types

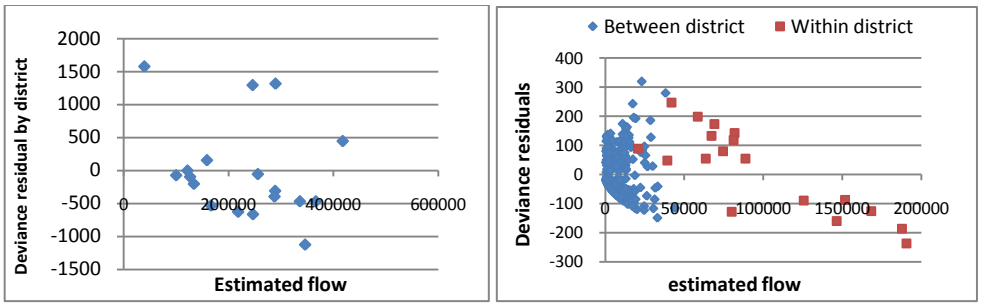
Table 4.3 Outflow constrained gravity model parameter estimates and the goodness-of-fit statistics.

Census work flow (W)			
Attraction factor	Work inflow	Population	Population (adjusted)
Model no.	**O-W-I**	**O-W-P**	**O-W-P2**
β_1	0.965	0.000	0.601
β_2	-	-	0.106
a	3.855	7.635	4.881
b	1.463	1.983	1.627
Log likelihood	-14,465,988	-14,785,300	-14,525,413
Deviance	289,883	928,507	408,732
SRMSE	0.388	0.624	0.507
Census study flow (S)			
Attraction factor	Study inflow	Population	Population (adjusted)
Model no.	**O-S-I**	**O-S-P**	**O-S-P2**
β_1	0.718	0.034	0.619
β_2	-	-	0.075
a	2.208	2.287	1.847
b	2.156	2.198	2.082
Log likelihood	-5,501,432	-5,519,967	-5,476,049
Deviance	225,944	263,015	175,179
SRMSE	0.436	0.492	0.415
Census work & study flow (WS)			
Attraction factor	Work + study inflow	Population	Population (adjusted)
Model no.	**O-WS-I**	**O-WS-P**	**O-WS-P2**
β_1	0.998	0.000	0.536
β_2	-	-	0.092
a	2.180	3.013	2.339
b	1.432	1.600	1.485
Log likelihood	-20,439,354	-20,695,179	-20,459,024
Deviance	485,000	996,649	524,340
SRMSE	0.410	0.600	0.459
Adjusted census overall flow (AC)			
Attraction factor	overall inflow	Population	Population (adjusted)
Model no.	**O-AC-I**	**O-AC-P**	**O-AC-P2**
β_1	0.969	0.000	0.498
β_2	-	-	0.087
a	0.816	0.973	0.752
b	1.177	1.235	1.169
Log likelihood	-25,472,150	-25,667,403	-25,436,898
Deviance	666,374	1,056,879	595,869
SRMSE	0.418	0.502	0.423
Survey overall flow (TS)			
Attraction factor	overall inflow	Population	Population (adjusted)
Model no.		**O-TS-P**	**O-TS-P2**
β_1	n/a	0.062	0.844
β_2		-	0.106
a		1060779.314	6.477
b		229270.104	1.398
Log likelihood		-30,153	-29,050
Deviance		3,011	806
SRMSE		0.562	0.249

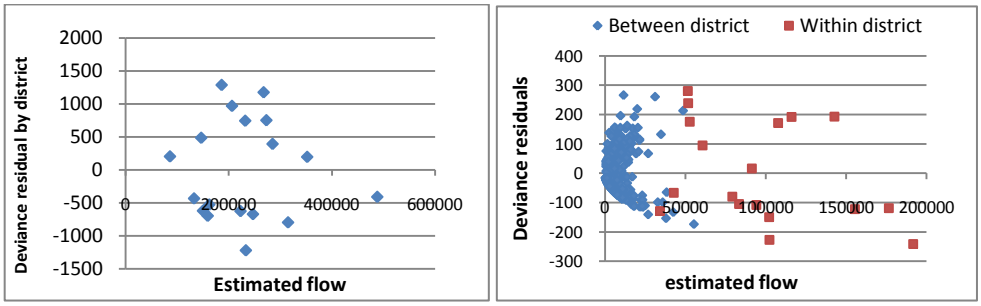
indicates that the inflow models had poor fit. All of the fitted models greatly underestimate the flow of high inflow districts, shown as the few positive deviance residuals by district, and the rest as negative residuals of smaller magnitude [Figure 4.9 (a,left) as an example]. On the other hand, the within-district flow of highly populated districts is overestimated, displayed as the downward trend in these residuals [Figure 4.9 (a,right)]. These were not observed for the district-adjusted population models [Figure 4.9 (b)]. Taking this into account, the district-adjusted population models were preferred for all five types of flow.

Figure 4.9 Deviance residual plots of the outflow constrained gravity models by district (left) and of individual observations (right) of the inflow and district-adjusted population models on census work-study flow.

(a) Inflow effect model (Model O-WS-I)



(b) District-adjusted population size model (Model O-WS-P2)

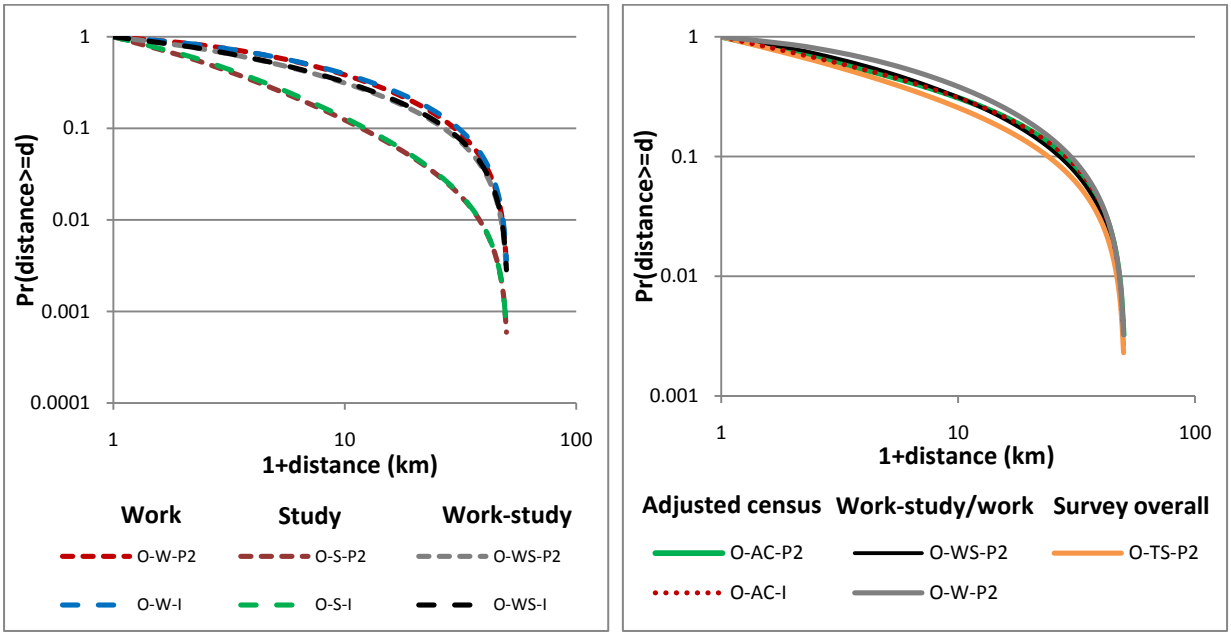


For each census flow type, the normalised spatial kernels derived from the inflow and district-adjusted population gravity models were highly consistent to the extent that they were almost visually indistinguishable [Figure 4.10 (a-b)]. The study flow kernel was the most localised kernel, followed by survey overall, adjusted census overall and census work-study, and the work kernels, and the non-study kernels were close to each other [Figure 4.10(b)]. Distance impendence differed by study trip compared to other trip types as indicated by the kernels. Half of the study trip distance is estimated to be within 2km, i.e. they were all within-district flow (Figure 4.2), consistent with the high within-district study flow observed which captured other trip making factors besides journey distance. For work trips, half of the journeys were estimated to be within 7km – a distance including journeys to non-contiguous districts. For work-study flow and adjusted census overall flow, half of trips were within 5km, and for the survey overall trip, within 4km.

Figure 4.10 The normalised power law spatial kernel (in log-log scale) of the inflow and district-adjusted population gravity models on the 5 flow types respectively.

(a) Census work, study and work-study flow kernel

(b) Adjusted census overall and survey overall kernel (compared to census work and work-study kernel)



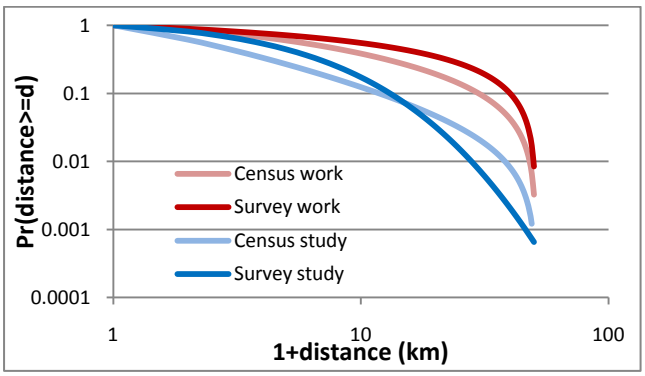
Because of the robustness of the spatial kernel derived between the inflow and population models for each flow type, the difference in the goodness-of-fit between the models fitted to the same flow can be attributed to the attraction factor used and the estimates of the scaling exponents. Inflow as the regional attributing factor generally showed a directly proportional relationship with the flow ($\beta_1 \approx 1$) as expected, except for the study flow, but a single scaling relationship was not adequate to capture the dichotomous mechanism in the flow patterns as discussed. The lower than expected inflow exponent ($\beta_1=0.7$) and the steep spatial kernel of the study flow suggested distance impendence effect was relatively strong in driving the trip patterns compared to other flows. Using population size as the regional attributing factors required a scaling adjustment (β_2) for the high work inflow districts, otherwise the model could not capture any population size effect (Models O-W-P, O-WS-P and O-AC-P), and this was not improved by excluding the remote and lowly-populated Islands district in the analysis. With district adjustment, the common population scaling effect β_1 was estimated at around 0.5-0.6 for all census flow type, around 0.8 for the survey overall flow, and the adjustment effect β_2 was all around 0.08-0.1. On the other hand, census flow of increasingly localised nature tended to have lower β_1 and β_2 estimates [work:(0.60,0.11); work-study: (0.54, 0.09); census adjusted overall: (0.50,0.09)], along with slightly more localised spatial kernel [Figure 4.10(b)]. Adjustment of all four high work inflow districts gave the best model fit for the census work, work-study flow, adjusted census overall flow and survey overall flow, while for study flow the best model had all the five high study inflow districts adjusted.

The potential effects of the return trip in the survey data on the gravity models could be higher scaling exponents, induced from the model comparison of the survey and census work and study flow (β_1 – work: 0.64 vs 0.60, study: 0.75 vs 0.62; β_2 – work: 0.26 vs 0.11, study: 0.14 vs 0.08, Table 4.3 and Table 4.4) and less localised spatial kernels (Figure 4.11). My attempt to fit a gravity model on the survey overall flow that excluded the work and study trips was not successful, possibly because of different underlying dynamics driving the various types of flow in the data, and the low data resolution that only allowed the model fitting on the marginal inflow. However, it could at least be concluded that the high β_1 of 0.84 in the survey overall flow model (with β_2=0.106, same as census work model, Table 4.3) was partly due to the returning effect, though the effects of non-work, non-study and non-home based trips could not be quantified. The same analysis was not conducted as the inflow constrained model, because the outflow of the survey data was more subject to the projection used to generate the data. Nor was it fitted as an unconstrained model since only the marginal flow data were available for model fitting. In general, although high correlation was observed between the census work and work-study flow with the survey overall trip patterns, evidence was lacking to either support or reject that their scaling factors with population size were comparable, and therefore could be a good proxy for the survey overall flow.

Table 4.4 Parameter estimates and goodness-of-fit statistics of the outflow constrained district-adjusted population gravity models fitted to the survey work and study flow data.

Data	Survey work trip	Survey study trip
Parameter estimates		
β_1	0.639	0.750
β_2	0.257	0.143
a	0.751	52.814
b	0.687	12.167
Goodness-of-fit		
Log likelihood	-13,244	-4,589
Deviance	692	79
SRMSE	0.331	0.196

Figure 4.11 Spatial kernels derived from the survey work flow and study flow gravity model compared to that of the census work flow and study flow model.



4.7.1.2 Inflow constrained gravity models

Table 4.5 summarizes the results of the inflow constrained gravity models fitted to the various movement patterns. Consistent with the high correlation observed between outflow and population size (Figure 4.4), the scaling effect α_1 of the population size without district adjustment was around 1 except in the survey overall flow model, probably because journeys not started from home were also included in the survey data. However, adjusting the high inflow districts gave the best model goodness-of-fit in all flow types, and generally more random patterns in the deviance residual plots. Estimates of α_1 varied, from 0.38 for census study flow, to 0.76 for census work flow, and the mixed flow types ranged at 0.6-0.7 and again with the more localised the nature of the flow type having lower scaling exponent. Estimates of α_2 were -0.08 for study flow, and between -0.04 and -0.06 otherwise, suggesting that the high inflow districts were less "competitive" as the trip donor compared per capita, though more competitive as trip attractor based on the results of the outflow constrained models. α_2 was insignificant in the survey overall flow, probably because high inflow districts were also important donors of non-home based trips. Study flow remained more driven by travel distance than the zonal attributes. The same districts were adjusted as in the outflow constrained models, except for the work flow and work-study flow which was better fitted without adjusting for the Islands district.

The spatial kernels derived were also very robust and were also visually indistinguishable from those fitted with outflow as the factor. The rank of descending localisation of the kernels was the almost identical to that of the outflow constrained models, except the survey overall flow was the least localised rather than being the second in the previous ranking by outflow constrained kernels (Figure 4.12). In general, the kernel ranking confirmed the extent of localisation of the census flow fitted in both gravity model settings.

Figure 4.12 Spatial kernels derived from the inflow constrained, district-adjusted population size gravity models of different flow types.

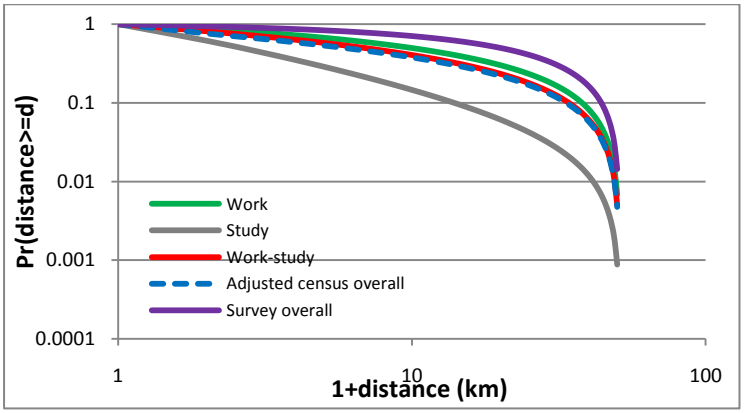


Table 4.5 Inflow constrained gravity model parameter estimates and goodness-of-fit statistics.

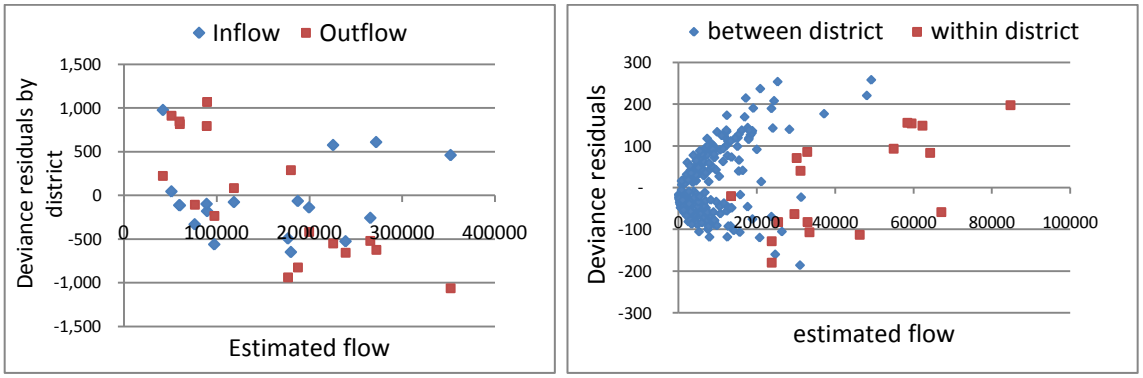
Census work flow			
Attraction factor	Work outflow	Population	Population (adjusted)
Model no.	**I-W-O**	**I-W-P**	**I-W-P2**
α_1	0.991	1.010	0.759
α_2	-	-	-0.043
a	1.081	1.069	1.162
b	0.801	0.802	0.851
Log likelihood	-14,568,824	-14,562,810	-14,534,722
Deviance	495,554	483,526	427,351
SRMSE	0.508	0.505	0.481
Census study flow			
Attraction factor	Study outflow	Population	Population (adjusted)
Model no.	**I-S-O**	**I-S-P**	**I-S-P2**
α_1	1.096	1.034	0.388
α_2	-	-	-0.081
a	1.402	1.369	1.328
b	1.803	1.777	1.808
Log likelihood	-5,519,289	-5,539,079	-5,492,797
Deviance	261,658	301,238	208,674
SRMSE	0.456	0.510	0.368
Census work & study flow			
Attraction factor	Work + study outflow	Population	Population (adjusted)
Model no.	**I-WS-O**	**I-WS-P**	**I-WS-P2**
α_1	1.073	1.049	0.725
α_2	-	-	-0.055
a	0.876	0.803	0.809
b	0.993	0.961	0.984
Log likelihood	-20,563,983	-20,593,038	-20,530,146
Deviance	734,257	792,367	666,583
SRMSE	0.542	0.568	0.500
Adjusted census overall flow			
Attraction factor	overall outflow	Population	Population (adjusted)
Model no.	**I-AC-O**	**I-AC-P**	**I-AC-P2**
α_1	1.095	1.054	0.592
α_2	-	-	-0.062
a	0.386	0.366	0.349
b	0.940	0.926	0.931
Log likelihood	-25,549,138	-25,566,167	-25,483,651
Deviance	820,350	854,408	689,376
SRMSE	0.508	0.518	0.554
Survey overall flow			
Attraction factor	overall outflow	Population	Population (adjusted)
Model no.		**I-TS-P**	**I-TS-P2**
α_1	n/a	0.728	0.647
α_2		-	-0.055
a		0.000	0.001
b		0.219	0.287
Log likelihood		-36,709	-36,697
Deviance		277	253
SRMSE		0.145	0.140

4.7.1.3 Unconstrained gravity models

Results of the unconstrained gravity model fitting are summarized in Table 4.6. The district-adjusted population models were the best model for all flow types, though for the census work flow this was justified by the randomly distributed deviance residuals [Figure 4.13(b)], as the model with marginal flow attribute gave lower model deviance but with trends in the deviance residual plots [Figure 4.13(a)]. Estimates of all scaling exponent α_{u1}, α_{u2}, β_{u1} and β_{u2} were generally comparable to those estimated in the outflow constrained and inflow constrained settings, though to a lesser extent for the overall survey model, where robustness of the model could have been affected by its low data resolution. With the exception of the study flow district-adjusted population model, all marginal flow and district-adjusted population models had $\alpha_{u1} > \beta_{u1}$. For high inflow districts, $(\alpha_{u1}+\alpha_{u2})<(\beta_{u1}+\beta_{u2})$ for work and study flow, and only marginally so for adjusted census overall and survey overall flow, and the opposite was observed for census work-study flow (Table 4.7). On comparing the multiplicative flow component (= population$^{\text{scaling exponent}}$ x model normalisation constant $^{0.5}$) of each attribute (origin / destination, adjusted district / other district), where the origin flow component x destination flow component x spatial kernel = flow from the origin to destination, the key population

Figure 4.13 Deviance residual plots of the unconstrained gravity models by district (left) and of individual observations (right) of the marginal flow and district-adjusted population models on census workflow.

(a) Marginal flow model (Model U-W-M)



(b) District-adjusted population size model (Model U-W-P2)

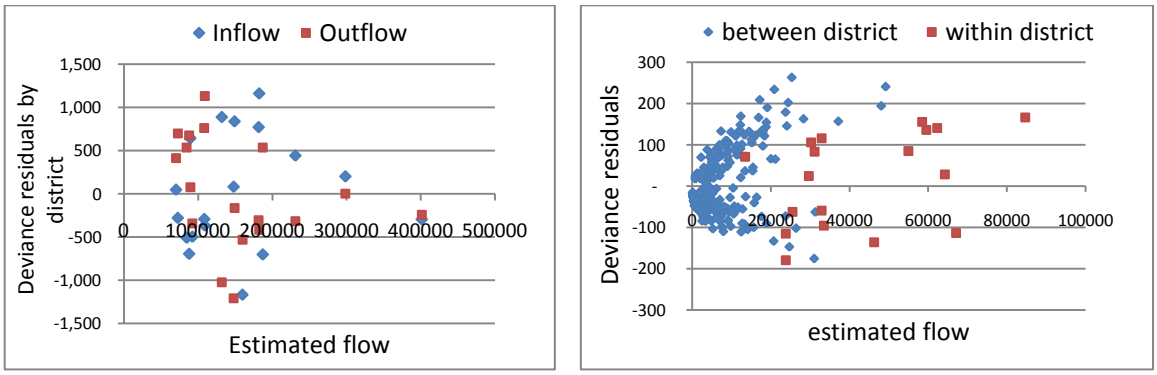


Table 4.6 Unconstrained gravity model parameter estimates and goodness-of-fit statistics.

Census work flow			
Attraction factor	Work outflow & inflows	Population	Population (adjusted)
Model no.	**U-W-M**	**U-W-P**	**U-W-P2**
α_{u1}	0.922	0.960	0.706
α_{u2}	-	-	-0.041
β_{u1}	0.839	0.000	0.606
β_{u2}	-	-	0.104
a	0.656	2.408	1.241
b	0.660	1.100	0.869
Log likelihood	-14,601,623	-14,861,794	-14,611,753
Deviance	561,153	1,081,495	581,412
SRMSE	0.574	0.712	0.533
Census study flow			
Attraction factor	Study outflow & inflow	Population	Population (adjusted)
Model no.	**U-S-M**	**U-S-P**	**U-S-P2**
α_{u1}	0.920	1.003	0.397
α_{u2}	-	-	-0.084
β_{u1}	0.306	0.032	0.548
β_{u2}	-	-	0.067
a	1.373	10.514	1.145
b	1.770	5.611	1.720
Log likelihood	-5,532,508	-5,638,188	-5,511,367
Deviance	288,097	499,456	245,815
SRMSE	0.524	0.922	0.734
Census work & study flow			
Attraction factor	Work-study outflow &	Population	Population (adjusted)
Model no.	**U-WS-M**	**U-WS-P**	**U-WS-P2**
α_{u1}	0.918	0.973	0.742
α_{u2}	-	-	-0.052
β_{u1}	0.708	0.010	0.512
β_{u2}	-	-	0.086
a	0.647	2.282	0.790
b	0.876	1.464	0.974
Log likelihood	-20,632,926	-20,832,279	-20,589,365
Deviance	872,144	1,270,851	785,022
SRMSE	0.633	0.609	0.532
Adjusted census overall flow			
Attraction factor	overall outflow & inflow	Population	Population (adjusted)
Model no.	**U-AC-M**	**U-AC-P**	**U-AC-P2**
α_{u1}	0.823	0.944	0.608
α_{u2}	-	-	-0.060
β_{u1}	0.560	0.000	0.469
β_{u2}	-	-	0.083
a	0.346	0.511	0.329
b	0.895	1.007	0.916
Log likelihood	-25,637,585	-25,747,980	-25,528,309
Deviance	997,244	1,218,033	778,691
SRMSE	0.512	0.464	0.414
Survey overall flow			
Attraction factor	overall outflow & inflow	Population	Population (adjusted)
Model no.		**U-TS-P**	**U-TS-P2**
α_{u1}	n/a	0.482	0.727
α_{u2}		-	-
β_{u1}		0.000	0.509
β_{u2}		-	0.223
a		1444.469	0.295
b		587.536	0.536
Log likelihood		-90,801	-89,481
Deviance		4,746	2,106
SRMSE		0.629	0.353

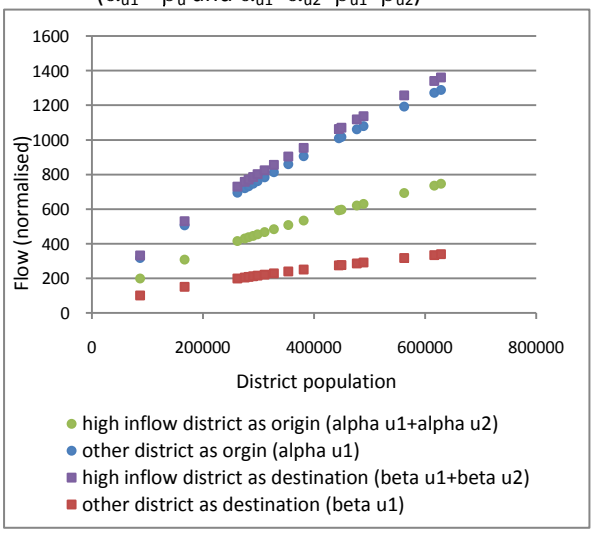
attributes that drove each type of flow (compared on the basis of identical population size) can be determined, assuming constant spatial kernel effect between the regions (Table 4.7,Figure 4.14). The more localised flow types – census work-study and adjusted census overall flow were most driven by other (i.e. non-high inflow district) origin population. Both the work and survey overall flow were most driven by high inflow district destination population and other district origin population. All non-study flow was least sensitive to the population size of the other destination. Study flow was strongly driven by adjusted destination population, but all origins had little impact.

Table 4.7 Categorisation of each flow type by the comparison of scaling exponent estimates. In parenthesis: key driving attributes of the flow in descending order. "Adjusted origin/destination" refers to high inflow district as origin/destination, and "other origin/destination" refers to other districts as origin/destination.

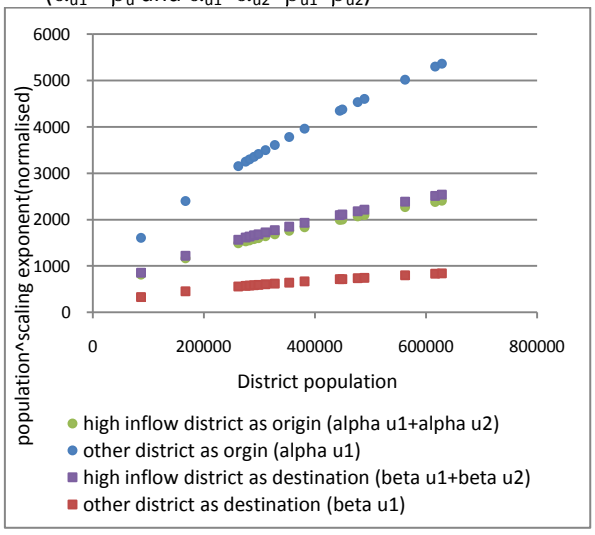
	$\alpha_{u1}+\alpha_{u2}<\beta_{u1}+\beta_{u2}$	$\alpha_{u1}+\alpha_{u2}\approx\beta_{u1}+\beta_{u2}$	$\alpha_{u1}+\alpha_{u2}>\beta_{u1}+\beta_{u2}$
$\alpha_{u1} > \beta_{u1}$	Census work-study *(other origin > adjusted origin > adjusted destination > other destination)*	Adjusted census overall *(other origin > adjusted destination >≈ adjusted origin > other destination)* Survey overall *(adjusted Destination >≈ any origin [$\alpha_{u2}=0$] > other destination)*	Census work *(adjusted destination >≈ other origin > adjusted origin > other destination)*
$\alpha_{u1} < \beta_{u1}$			Census study *(adjusted destination > other destination > other origin > adjusted destination)*

Figure 4.14 Comparison of the effects of scaling exponents of the high inflow and ordinary districts on the multiplicative flow component (=population$^{\text{scaling exponent}}$ x normalisation constant$^{0.5}$), based on the unconstrained gravity models on census work flow and adjusted census overall flow, assuming the distance effect=1.

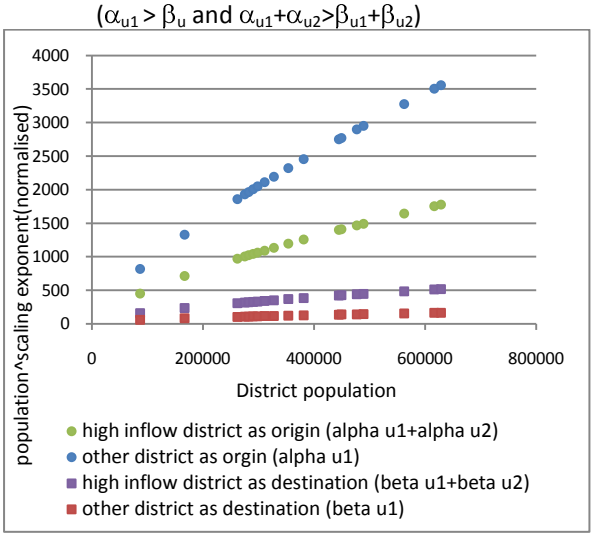
(a) Census work flow ($\alpha_{u1} > \beta_u$ and $\alpha_{u1}+\alpha_{u2}<\beta_{u1}+\beta_{u2}$)



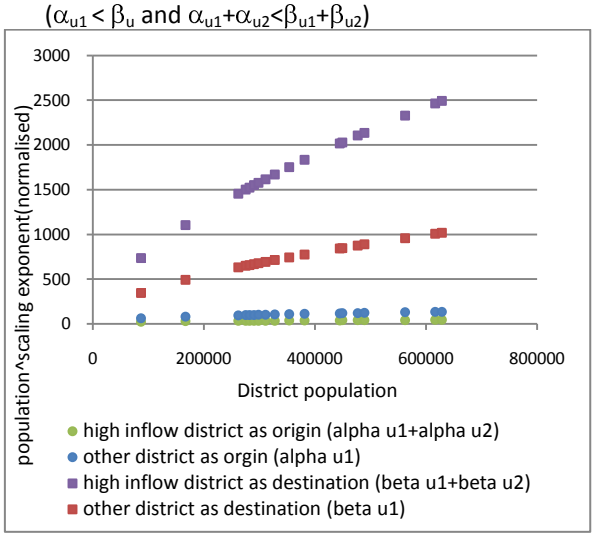
(b) Adjusted census overall flow ($\alpha_{u1} > \beta_u$ and $\alpha_{u1}+\alpha_{u2}\approx\beta_{u1}+\beta_{u2}$)



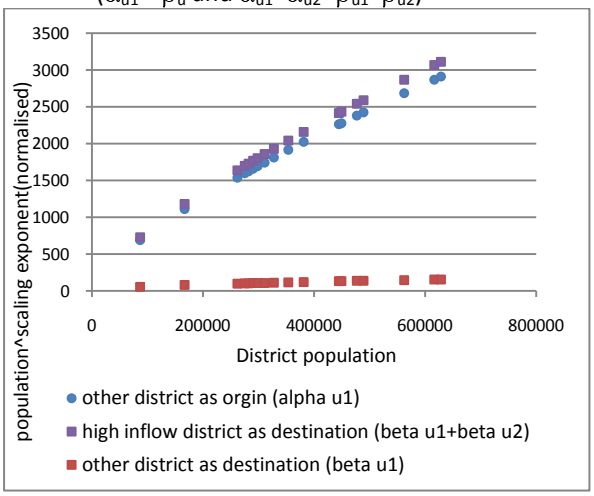
(c) Census work-study flow ($\alpha_{u1} > \beta_u$ and $\alpha_{u1}+\alpha_{u2} > \beta_{u1}+\beta_{u2}$)



(d) Census study flow ($\alpha_{u1} < \beta_u$ and $\alpha_{u1}+\alpha_{u2} < \beta_{u1}+\beta_{u2}$)



(e) Survey overall flow ($\alpha_{u1} > \beta_u$ and $\alpha_{u1}+\alpha_{u2} \approx \beta_{u1}+\beta_{u2}$)

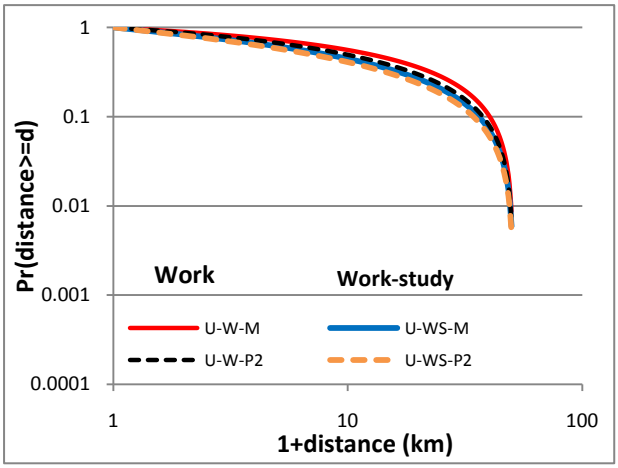


Unconstrained gravity models had slightly less robust spatial kernels for the work and work-study flow on comparing the marginal flow model to the district-adjusted population models [Figure 4.15(a)]. Study flow and adjusted census overall kernels remained visually inseparable under the same comparison. The order of descending localisation of the kernels was identical to that of the inflow constrained models [Figure 4.15(b)].

Comparing the kernels from the three types of gravity model under each flow type (Figure 4.16), kernels from unconstrained models were visually inseparable to those of the inflow constrained models for work, work-study and adjusted census overall flow, and only slightly different for the study and survey overall flow. Kernels from the outflow constrained models were always more localised than the others, and were fairly close to the unconstrained / inflow kernel under the study and adjusted census overall flow, but there was a gap between the survey kernels. Comparing the fitted journey distance distribution (Figure 4.17), unconstrained models for work and work-study

Figure 4.15 The normalised power law spatial kernel of marginal flow and district-adjusted population gravity models on the five flow types respectively.

(a) Census work and work-study flow kernel (marginal flow and population with adjustment model)



(b) Kernel of all 5 flow type (population with adjustment model)

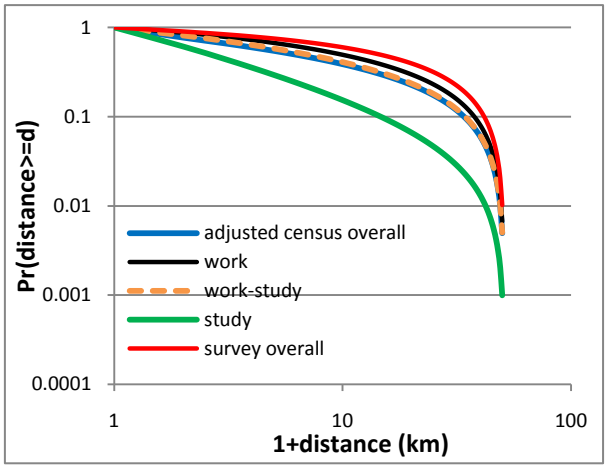
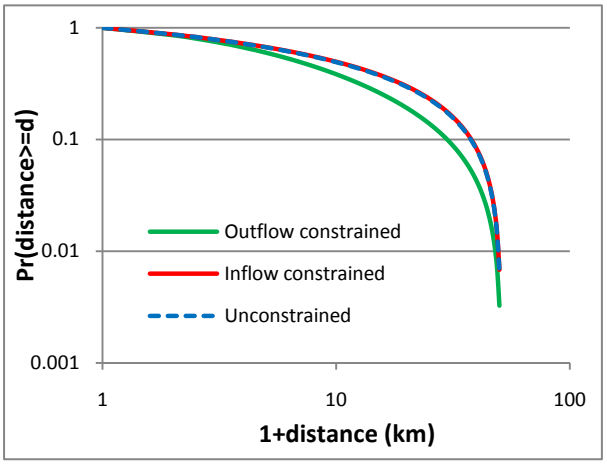
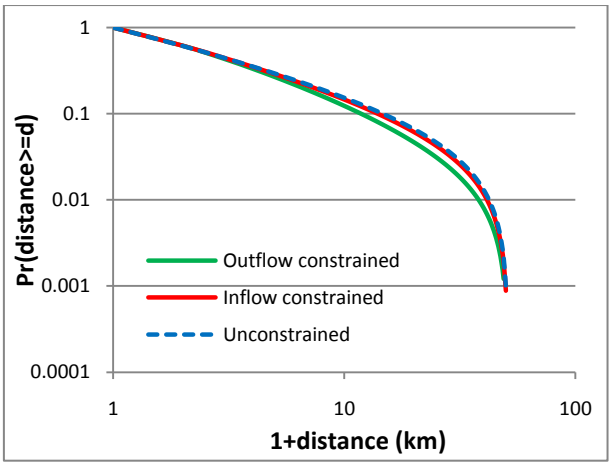


Figure 4.16 Comparison of the spatial kernel of each flow type derived by inflow constrained, outflow constrained and unconstrained models.

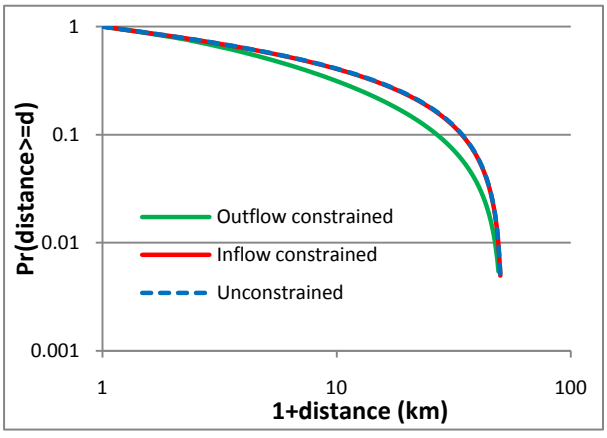
(a) Work flow



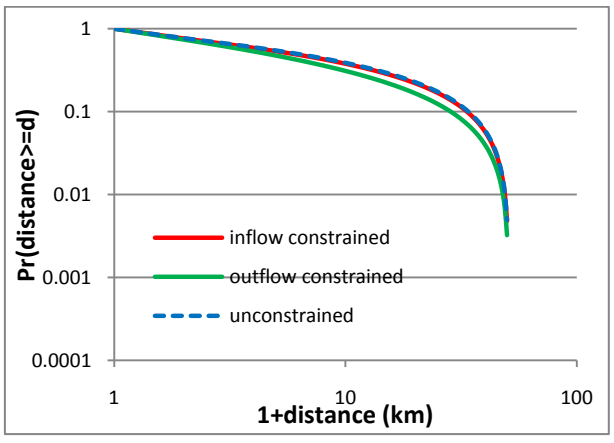
(b) Study flow



(c) Work-study flow



(d) Adjusted census overall flow



(e) Survey overall flow

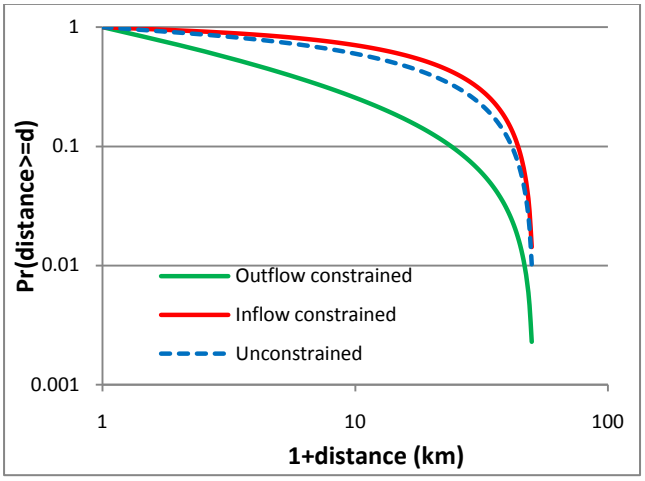
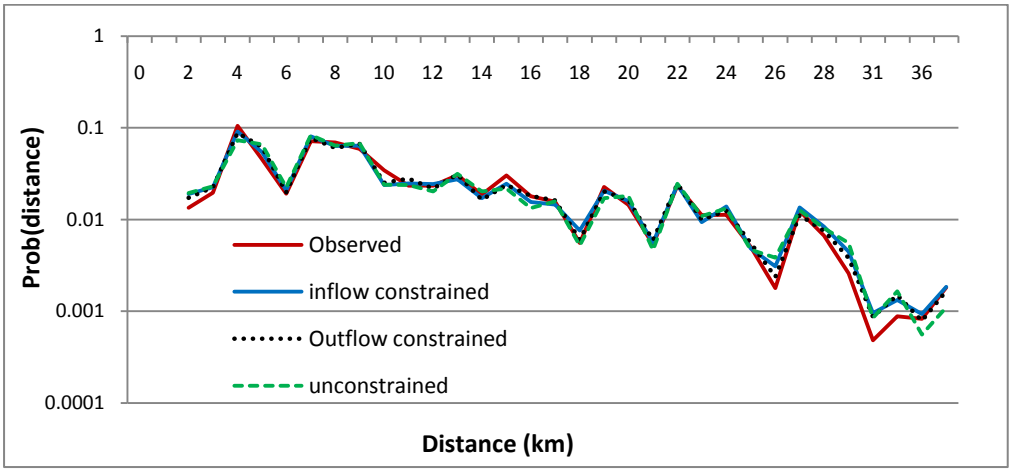
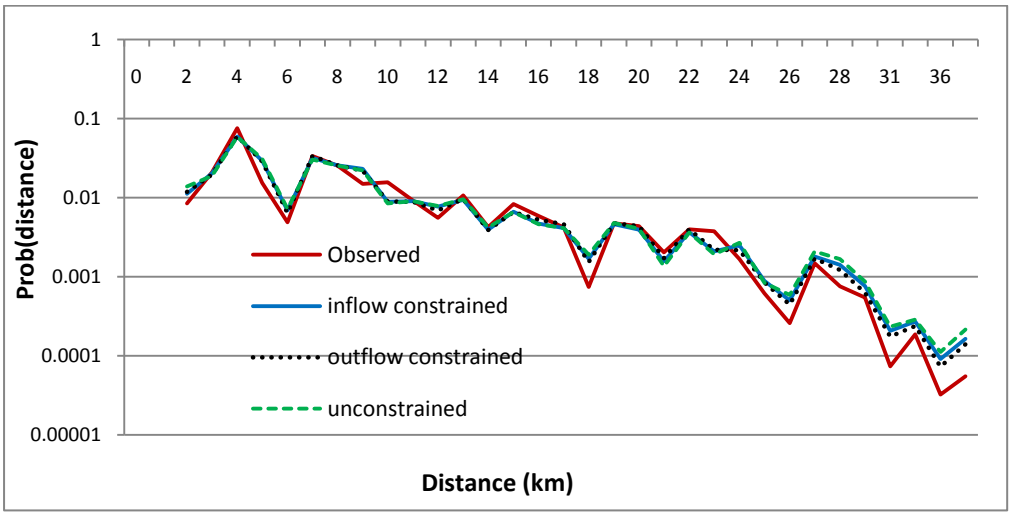


Figure 4.17 Comparison of the estimated journal distance distribution under the inflow constrained, outflow constrained and unconstrained models for each flow type.

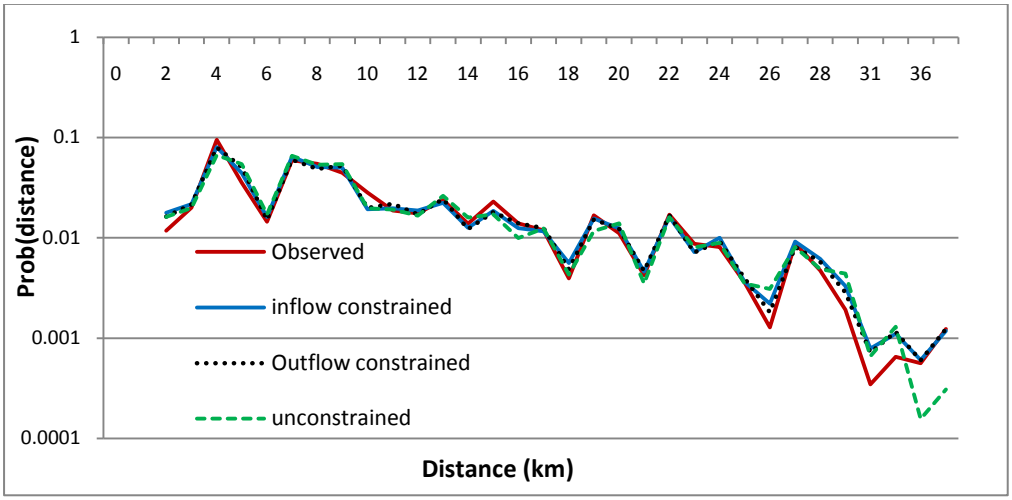
(a) Work flow



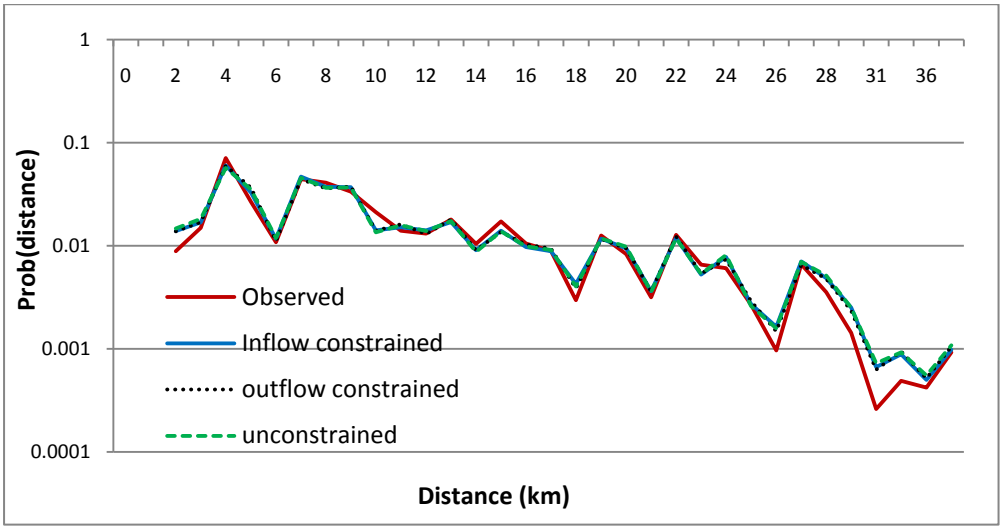
(b) Study flow



(c) Work-study flow



(d) Adjusted census overall flow



flow were inferior to the constrained models, while in general all three types of gravity model fitted equally well for the study and adjusted census overall flow. In general, the performance of the unconstrained gravity models in this study was satisfactory with the robustness of the results not as strong as the constrained models only on some occasions.

4.7.2 Hospital flow patterns

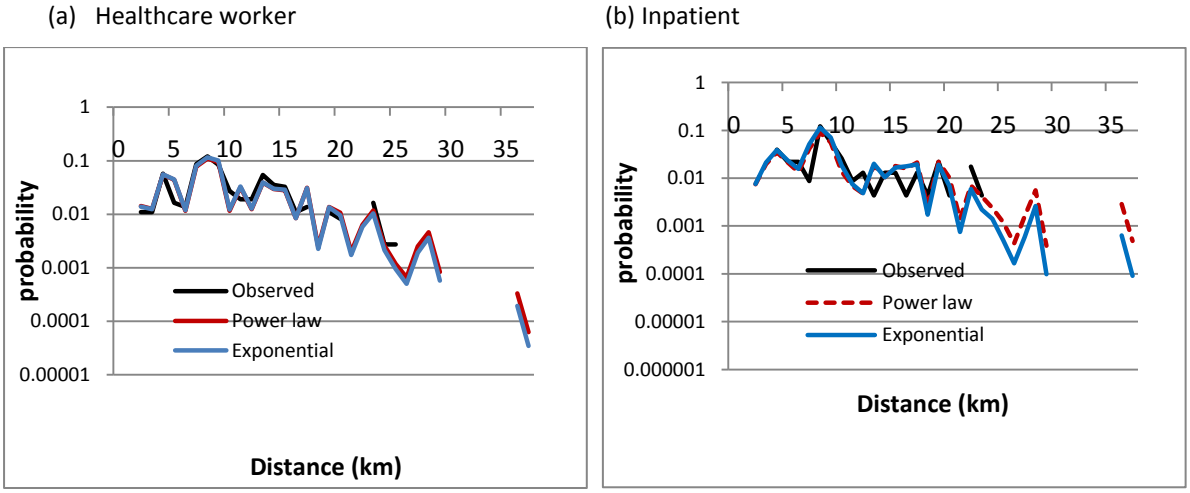
Table 4.8 shows that the home-to hospital flow of inpatient, visitor and the overall flow were not affected by hospital size (models H3-H8), and the increment of healthcare worker's flow diminished with increase in hospital size (β=0.30, models H1-H2). However, using hospital size as the attribution factor was not adequate as the flows to PWH and AHN were highly underestimated from the residual plots. This suggested that the flow of inpatients and visitors were more driven by the epidemic related factors, and using inflow as the attribution factor greatly improved the gravity model fit for all flows (models H9-H16). The scaling component was consistently estimated at 0.9,

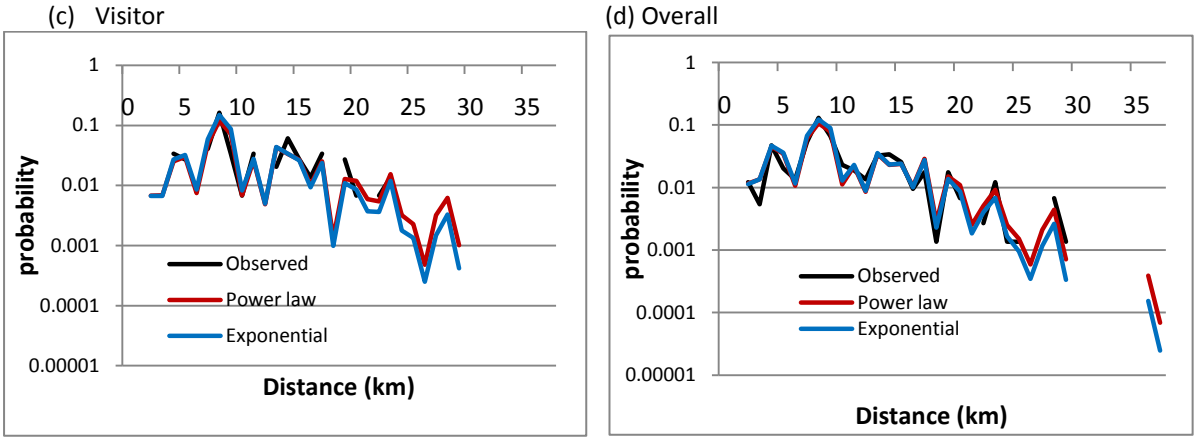
regardless of the flow type and the spatial kernel functional form. The power law models had marginally higher SRMSE in the overall, inpatient and visitor inflow, but were otherwise the better model than the exponential kernel model in other goodness-of-fit statistics and the estimation of journey distance, especially in the within-district flow distribution (Figure 4.18).

Table 4.8 Parameter estimates and goodness-of-fit statistics of the gravity models on the healthcare worker, inpatient, visitor and overall flows, using hospital size or number of hospital infection as the attribution factor, and the power-law or exponential distribution as the spatial kernel. (SRMSE: standardized root mean square error).

	Healthcare worker		**Inpatient**		**Visitor**		**Overall**	
Spatial distance kernel	Power law	Exp	Power law	Exp	Power law	Exp	Power law	Exp
Total observed flows	367		229		147		739	
Attribution factor	Hospital size							
Model	H1	H2	H3	H4	H5	H6	H7	H8
Parameter estimates								
β	0.303	0.300	0.000	0.000	0.000	0.000	0.000	0.000
a	26.882	0.144	5.435	0.210	1.288	0.156	7.241	0.166
b	5.092	-	2.434	-	0.998	-	2.351	-
Goodness-of-fit								
Log likelihood	-1721	-1722	-912	-921	-631	-637	-3288	-3300
Deviance	470	472	288	306	270	281	684	708
AIC	3448	3448	1830	1846	1268	1277	6583	6604
SRMSE	2.627	2.735	2.841	2.973	3.234	3.694	2.358	2.615
Attribution factor	Inflow							
Model	H9	H10	H11	H12	H13	H14	H15	H16
Parameter estimates								
β	0.916	0.912	0.912	0.942	0.894	0.895	0.897	0.896
a	39.860	0.141	9.607	0.202	6.644	0.142	17.224	0.160
b	6.753	-	3.323	-	1.894	-	3.956	-
Goodness-of-fit								
Log likelihood	-1578	-1579	-878	-884	-566	-568	-3092	-3098
Deviance	184	185	220	232	139	143	292	303
AIC	3162	3162	1763	1773	1138	1140	6191	6200
SRMSE	1.027	1.055	2.012	1.926	1.918	1.832	1.186	1.185

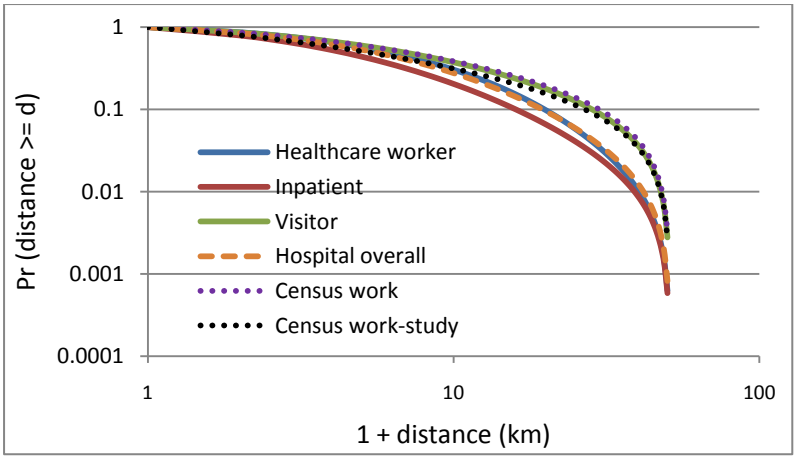
Figure 4.18 Comparison of the observed and estimated probability distribution of journey distance by gravity models using the power law or exponential spatial kernel for the: (a) healthcare worker flow, (b) inpatient flow, (c) visitor flow, and (d) overall flow, using hospital size as the attraction factor.





From the derived normalised power law spatial kernels, inpatients were the most localised compared to healthcare workers, and the latter resided closer to their workplace than the general Hong Kong working population (Figure 4.19). Visitors were the furthest away from the hospital, but their kernel distribution was very close to that of the census work or work-study flow kernels. Half of the inpatients tended to travel within 4km from the hospital they were infected with SARS, and half of the visitors and healthcare workers within 5km.

Figure 4.19 Power law spatial kernels derived from the healthcare worker, inpatient, visitor and overall hospital flow gravity models, compared to the census work and work-study spatial kernels from outflow constrained gravity models.



4.8 Discussion

In this chapter, I studied the characteristics of various trip patterns of the Hong Kong population, and derived gravity models to describe their attributing factors with proven robustness of the parameter estimates and adequacy of the models. Several districts which are the key commercial and leisure areas of Hong Kong demonstrated a different flow dynamics that need a custom-structured gravity model beyond a simple proportional relationship with the marginal flow or the population. It is unlikely that this phenomenon only exists in Hong Kong, but the small physical and population size and the spatial unit of the study could have made these effects more prominent in the gravity model

fitting. Allowing these districts to have a different scaling relationship with the population size better predicted the flow patterns than using marginal flow which directly captures the flow related characteristics of the region.

The scaling effect of population size to work and mixed flow from the census was 0.6-0.7 for destination populations (both normal and high inflow districts), and around 0.5-0.6 for the normal origin population and 0.6-0.7 for high inflow districts origin. The destination scaling effects of the previous studies (mostly ≈0.6-0.7) are comparable to this study, but they had lower origin scaling effects (0.3-0.5), and the ratio of these two exponents indicated that destination population effect is stronger than the origin population effect in driving the trip patterns. The same was observed only in the study flow of this study. Such difference between the studies could be due to the general factors that Hong Kong is more like a city of other countries in terms of its physical and population size, also with its own demo-geographic settings.

To further understand these differences, I looked into each flow type for its key driving attributes. The study flow was markedly different from other flows in that journey distance is a strong trip driver. This conformed to the observation in human geography that public facilities or governmental service establishment (e.g. public schools) are more densely distributed and closer to their users than commercial facilities or non-governmental establishments (e.g. bank, retail stores) (Stephan 1988; Um, Son et al. 2009), which are also locations of retail employment. The key financial and commercial centres in Hong Kong further reduce the distance effect on the commuting patterns, therefore a least localised spatial kernel compared to other flow, but highly driven by high inflow destination population and normal district origin population. The integrated patterns of work and study flow adjusted with those who did not travel were strongly driven by normal district origin population and had a more localised spatial kernel than commuting flow. Thus this model is most appropriate for the heterogeneous contact rate for the modified Wallinga and Teunis method.

The spatial kernels derived in this study are much more localised than the other studies due to the small geographical size of Hong Kong. The work and study trip kernels indicated that if considering the distance impendence effect, half of the travel distance was within 5km – a journey distance of either within-district or between contiguous districts all over Hong Kong, or between non-contiguous districts in Hong Kong Island or Kowloon. Though not directly comparable, a metapopulation epidemic model of the Hong Kong SARS outbreak (Riley, Fraser et al. 2003) estimated the relative contact rates between contiguous district relative to within-district to be 0.57 (95% CI: 0.42, 1), and between non contiguous districts at 0.02 (95% CI: 0, 0.1), suggesting a steep decay of contact rates over distance.

There are a few limitations to applying the derived gravity model into the spatial Wallinga and Teunis method. Although I derived gravity models using the travel survey data which also captured the non-routine trips, which were found by others to affect the timescale and patterns of epidemic spread (Danon, House et al. 2009), the effect of return trips included in the trip data could not be disentangled. Also the spatial resolution of the gravity model restricts the spatial resolution of the division of the SARS cases used in the extended method. Furthermore the distance between the centroid of the districts rather than the actual journey distance was used in the gravity model. For this reason, household transmission is not distinguishable from within-district transmission of non-household settings, and individuals in the same district are therefore assumed to make contacts homogenously. Any potential effect of population density was also not taken into account.

On comparing the distribution of the home-to-hospital flow by types of hospital infected SARS case, I found similar results as in the PWH ward 8A cases (Lai, Wong et al. 2004). Inpatients were the closest, given the geographically localised catchment area of hospitals, and the policies adopted to allocate SARS patients to hospitals during the epidemic. Healthcare workers were closer to their employing hospitals than visitors to those hospitals, and the general population to their workplace. This could be due to the demand of on shift work of the healthcare workers, but there was also a small proportion of more random patterns because of staff deployment from other hospitals during the epidemic (Severe Acute Respiratory Syndrome Expert Committee 2003). Visitors were furthest, though their patterns were expected to be a mixture of that similar to that of the inpatient, contributed by the family and housemaid of the inpatients, and more random in distribution and remote to the hospitals, contributed by relatives and friends. The flow patterns implied that hospital visitors had the longest potential spatial range of secondary infection from hospital to community, whereas the healthcare worker were the key sources for the local transmission and cluster formation in areas with hospitals treating SARS patients. Such findings could provide further insights on the spread patterns and control measures of the epidemics, and the interplay between the hospital and community transmission.

Chapter 5 - The Spatial Wallinga and Teunis Method

In this chapter, I derive a spatially extended version of the Wallinga and Teunis method (referred as "the modified method") that incorporates gravity model based spatial coupling (as described in the previous chapter) to estimate the reproduction number of the SARS outbreak in Hong Kong. I begin with a review of the original Wallinga and Teunis method.

5.1 The Wallinga and Teunis method

5.1.1 Formulation

The Wallinga and Tenuis method (Wallinga and Teunis 2004) considers an epidemic as a branching process connecting infected individuals in a population. It considers all the possible infection trees which might generate an observed set of incident disease cases and estimates the probability of transmission between each potential pair of infected to derive estimates for the individual and daily reproduction numbers. The method makes two basic assumptions: (1) for each non-imported case *j*, there is exactly one infector *i*, who is from the *n* reported cases; and (2) the transmission events from case *i* to case *j*, and from case *i* to any other cases *k* are independent. Given these, the probability that *i* infected *j* is estimated from the serial interval density $w(s|\theta)$, where θ is the vector of the distribution parameters. Serial interval is the time interval between any specific event in the time course of infection of the primary case and the corresponding event of the secondary case (Svensson 2007), and most often the event is chosen as the onset of clinical symptoms. Denoting the cases by $i=1,\ldots n$, and the infection tree by the vector *v*, where *v(i)* is the infector of case *i*, and *q* as the number of imported cases, the probability of case *i* infected case *j* conditional on the observed set of cases is written as:

$$p^{I}_{i->j} = \frac{w(t_i - t_j;\theta)}{\sum_{m=1,m\neq j}^{m=n} w(t_m - t_j;\theta)} \tag{5.1}$$

Hence the probability of infection from case *i* to case *j* is given by the strength of their epidemiological link as estimated from their relative time of symptom onset, normalised by the sum of that of all potential infectors of case *j*.

The use of serial interval density to estimate the probability of infection between cases implicitly assumes that people are only infectious after developing symptoms, and this is a valid assumption for SARS. It also assumes that all people infectious at time *t* could have infected someone infected at that time.

Given the assumptions above, the individual effective reproductive number for case *i*, *R(i)*, is distributed as:

$$R(i) \sim \sum_{k=1}^{k=n-q} Bernoulli(p^{l}_{i->k})$$

The expected value of *R(i)* is therefore written as:

$$E[R(i)] = \sum_{k=1}^{k=n-q} p^{l}_{i->k} = \sum_{j=1}^{j=n-q} \frac{w(t_i - t_j;\theta)}{\sum_{m=1,m\neq j}^{m=n} w(t_m - t_j;\theta)} \quad (5.2)$$

R(t), the reproductive number on day *t*, is the average of *R(i)* of all the n_t cases with symptom onset on date *t*:

$$E[R(t)] = \frac{1}{n_t} \sum_{t_j=t} \sum_{j=1}^{j=n-q} \frac{w(t_i - t_j;\theta)}{\sum_{m=1,m\neq j}^{m=n} w(t_m - t_j;\theta)} \quad (5.3)$$

This method was evaluated on simulated data and its performance was in general satisfactory, though the estimated *R(t)* had a slight tendency to be lower than the actual value by less than 5% (Wallinga and Teunis 2004).

Using partial likelihood for statistical inference has much reduced the computational intensity of the Wallinga and Teunis method. Building on this advantage and extending the method to account for censored cases due to reporting delay, it can evaluate in real-time the transmissibility of the epidemic and the effectiveness of the implemented control measures (Cauchemez, Boelle et al. 2006; Cowling, Lau et al. 2010). When knowledge of the serial interval distribution is not available, the reproduction number and serial interval distribution may be able to be estimated simultaneously (Garske, Clarke et al. 2007), though this can often lead to model over-parameterisation (Griffin, Garske et al. 2010). In general, an epidemic curve only gives information about the real-time growth rate of an epidemic, so that the reproduction number and generation interval estimates are intrinsically correlated (Wallinga and Lipsitch 2007).

In a study of the H1N1 influenza pandemic in UK (Ghani, Baguelin et al. 2009), the Wallinga and Teunis method with censoring was further modified in two ways – incorporation of contact tracing information, and using the convolution of incubation period and infectivity densities instead of the serial interval distribution for the probability of infection. The use of the convoluted density was

supported by the availability of good data on incubation period, and allowed individuals to be infectious before the symptom onset, which is the characteristic of influenza. Lower reproduction number estimates and shorter confidence interval were attained compared with that from a full epidemic model, and this is consistent with the findings from the Wallinga and Teunis's original results with simulated data (Wallinga and Teunis 2004).

5.1.2 Alternative approaches

It is possible to derive the same effective reproduction number formulation [equation 5.3] using a renewal equation method (Fraser 2007). Based on Kermack and McKendrick's time-since-infection transmission model (Kermack and McKendrick 1927), and assuming secondary infections are generated by each infected individual as a Poisson process with rate $\beta(t,\tau)$ (where t is calendar time and τ is time since infection), $\beta(t,\tau)$ was related to the incidence rate $I(t)$, and the individual reproduction number of cases who had symptom onset on day t, denoted as $R^c(t)$, as follows :

$$I(t) = \int_0^\infty \beta(t,\tau) I(t-\tau) d\tau \quad \text{and}$$
$$R^c(t) = \int_0^\infty \beta(t+\tau,\tau) d\tau$$

It was shown that $\beta(t,\tau)$ can be decomposed into $w(\tau)$, the generation time distribution where $\int_0^\infty w(\tau) d\tau = 1$, and $R^I(t)$, the instantaneous reproductive number – the average number of secondary cases that could be generated by a case infected at time t should the conditions remain unchanged, which is a property related to the epidemic at time t rather than to individuals infected at time t as in the $R^c(t)$. Assuming $w(\tau)$ is independent of t, the decomposition is written as:

$$\beta(t,\tau) = R^I(t) w(\tau) \quad \text{and}$$
$$R^I(t) = \int_0^\infty \beta(t,\tau) d\tau$$

By substitution,

$$R^I(t) = \frac{I(t)}{\int_0^\infty I(t-\tau) w(\tau) d\tau} \quad \text{and}$$
$$R^c(t) = \int_0^\infty R^I(t+\tau) w(\tau) d\tau$$

Discretising the incidence and generation time distribution by time intervals t_i and t_{i+1} into I_i and w_i, where $\sum_{i=0}^{n} w_i = 1$,

$$R^{I}(t_i) = \frac{I_i}{\sum_{j=0}^{n} I_{i-j} w_j}$$

$$R^{c}(t_i) = \sum_{j=0}^{n} \frac{I_{i+j} w_j}{\sum_{k=0}^{n} I_{i+j-k} w_k} \qquad (5.4)$$

The $R^c(t)$ formulation is equivalent to equation 5.2, though Wallinga and Teunis used the serial interval density for the probability of transmission. Fraser's method was also extended to capture underreporting during the epidemic (Fraser, Donnelly et al. 2009). There are common assumptions in both approaches: (1) the serial interval and generation interval distribution were assumed unchanged during the epidemic, though Fraser suggested different density functions can be used at different phases of the epidemic; (2). individuals were homogeneous in their mixing, and in their infectivity profile or serial interval distribution; (3) A case was identified as a potential infector of the other cases if the generation or serial interval of the potential infection pairs was within the defined range of the interval. Wallinga and Teunis's approach is more individual based, while Fraser's formulation is calibrated on the daily incidence, though they are technically identical. However, the individual based approach is more suitable to be extended spatially when the spatial location of the potential infection pairs is incorporated.

White and Pagano (White and Pagano 2008) proposed a likelihood-based method based on the branching process concept that used the serial interval distribution to estimate the onset of the secondary cases probabilistically, and by fitting the derived daily incidence to the observed data the daily reproductive number was estimated. The method was also modified to enable the estimation of the serial interval distribution and reproductive number simultaneously, and adjust for censored cases and impute for cases with missing onset dates (White and Pagano 2008; White, Wallinga et al. 2009). It shares the similarities of the Wallinga and Teunis and Fraser's approach above, but is more difficult to be decomposed to allow additional spatial information of the individuals, nor could the infection tree be derived explicitly.

5.1.3 Capturing individual heterogeneity

In the original Wallinga and Teunis method, the only characteristic that differs between cases is the symptom onset date. Cases with the same onset date have exactly the same probability of having infected a specific secondary case regardless of their physical proximity to the infectee, and have the same individual reproductive number. This partly contributes to the *R(t)* estimates from the Wallinga and Teunis method being smoothed, even during the periods of super-spreading events.

Incorporating contact tracing data into the method can improve the accuracy of the reproduction number estimates by incorporating prior data on who infected who. However, in reality, contact tracing data are seldom obtainable from all cases, and in large scale outbreaks, resources required for contact tracing of cases become unaffordable as the epidemic grows (Ghani, Baguelin et al. 2009). On the other hand, only contact with people who are known by the case can be collected in contact tracing, and it is also subject to recall bias. In the Hong Kong SARS outbreak, a high proportion of cases had data on possible epidemiological source of transmission collected, and some of them had specific contacts traced (Severe Acute Respiratory Syndrome Expert Committee 2003). However, still around one-tenth of the cases could not identify any place or person from which infection might have been acquired (source of infection data described in chapter 2).

Spatial information for cases, for instance residential location, is more likely to remain available for identified cases even at the later stages of the epidemic. The prerequisite for infectious disease transmission is that infected and susceptible individuals are in the same place at the same time. Capturing this spatial condition allows more accurate and realistic reconstruction of the infection tree. A study on the UK Foot and Mouth disease epidemic in 2001 (Ferguson, Donnelly et al. 2001) had used a kernel of spatial distance from the infected farm as a multiplicative factor to the farm-specific susceptibility and infectiousness to estimate the relative risk of transmission.

Human movements are less traceable and more complicated that than livestock animals, but gravity models had proved successful in modelling human travel patterns and thus representing the spatial heterogeneity in the probability of contact between individuals based on their spatial location in epidemics.

Past work has explored both infector driven (Xia, Bjornstad et al. 2004; Eggo, Cauchemez et al. 2010) and susceptible driven (Merler and Ajelli 2010) assumptions in linking gravity models to the probability of contact in epidemic models, but the effects of these assumptions were not compared.

The essence of the Wallinga and Teunis method is to identify the potential infectors of cases, and then assign the relative likelihood of infection to all the potential infection pairs. Spatial context has a role to play in both processes. The original method identifies a case as potential infector regardless whether the potential infection pairs would have been in the same location or not. The use of spatial data can refine the list of potential infectors to those who were present at the infection location at the date of infection. The probability of infection between cases can be considered to be composed of two components – probability of contact based on the spatial location of the cases, and the probability of transmission given the timeline of infection. This chapter uses these principles to

derive an extended Wallinga and Teunis method incorporating both spatial and temporal components.

5.2 Study objectives

I consider two levels of spatial extension of the Wallinga and Teunis method. On a coarser spatial resolution, spatial data were used only to estimate the probability of contact between cases, referred as the "partially extended method". On a more refined spatial level that used the source of infection (e.g. hospital or community) and hospitalisation data of the cases, eligible potential infectors were those present at the infection location at the date of infection. The method with this additional spatial extension is referred as the "fully extended method".

Because the occurrence of disease transmission is not observable, the date of infection was inferred from the probability distribution of the incubation period and the symptom onset date to identify potential infectors who fulfilled the spatial criteria. Accordingly the temporal component of the probability of infection were estimated by a convoluted incubation and infectious (or infectivity) distribution instead of the serial interval density. Therefore, the first objective of this chapter is to derive such a deconvoluted distribution for the serial interval density of the SARS outbreak to be applied in both the partially and fully extended methods.

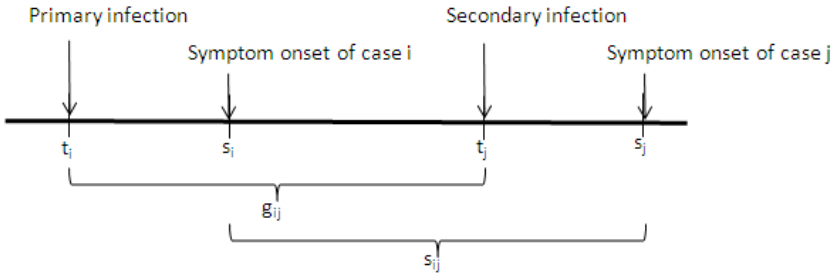
The second objective is to derive the partially extended method and evaluate different spatial mixing assumptions under which gravity models are used to derive the probability of contact. Gravity models predict the flows from a home location to the destination, and therefore give the probability of travel from home to a destination for certain trip purpose. This unilateral direction of movement, regardless of whether one considers the susceptible or infector to be travelling, risks oversimplifying the infection risks associated with travel. Therefore I also considered bilateral (both susceptible and infector could travel to each other's home location) and multilateral (susceptible and infector could travel to any location and meet there) movement, and under different trip types (work, work-study, adjusted overall) and trip data (census and survey), to find the most appropriate proxy for the probability of contact leading to the observed spatiotemporal spread of the SARS epidemic in Hong Kong.

Building on the result of this chapter, the fully extended method is derived in the next chapter to study the hospital and community transmission and their dynamics in the SARS outbreak in Hong Kong. The following section describes the method and discusses the results of deconvoluting the serial interval distribution – the first objective of this chapter.

5.3 Fitting serial interval data

5.3.1 Method

Figure 5.1 Timeline of transmission between infector *i* and infectee *j*. S_i and s_j are the time of symptom onset, t_i and t_j are the time of infection, g_{ij} is the generation interval and s_{ij} is the serial interval of the transmission between cases *i* and *j*.



Deconvolution of the serial interval density with the incubation period distribution to derive the infectiousness density (generation time distribution) was applied in the SARS and H1N1 pandemic (Kuk and Ma 2005; Ghani, Baguelin et al. 2009). Using the observed serial interval data, the unknown density of either the infectiousness or the incubation can be derived based on the decomposition of the timeline of infection illustrated in Figure 5.1. Consider infector *i* and infectee *j* with symptom onset dates s_i and s_j, and time of infections t_i and t_j respectively. Let *w(s)* be the serial interval density, *q(c)* be the incubation period density, and *u(t)* be the infectiousness density. I assume the serial interval distribution to have zero density at negative times given SARS cases started to be infectious only with the onset of symptoms, and that the distribution of incubation period and infectiousness to be independent, the serial interval density can expressed (in discrete time, given the daily interval of SARS case reporting) as:

$$w(s_j - s_i) = \sum_{t_j=s_i}^{s_j} u(t_j - s_i)q(s_j - t_j)$$

Without loss of generality, let s_i=0 and s_{ij}=s_j-s_i, giving:

$$w(s_{ij}) = \sum_{t_j=0}^{s_j} u(t_j)q(s_j - t_j) \qquad (5.5)$$

With the probability distribution of the incubation period and the empirical data of the serial interval reported in previous studies, I derived the infectiousness density *u(t)* using equation 5.5. In the fully and partially extended Wallinga and Teunis methods, given the symptom onset date of case *j* and the incubation period density, I inferred the date of infection of case *j* and the temporal component

of the probability of infection by the potential infector case *i* from the infectiousness density, given the symptom onset date of case *i.*

Assuming the count of the serial interval data by day to be Poisson distributed, I used the maximum likelihood method to fit the convoluted serial interval distribution with a time step of 0.5 days to empirical serial interval data. These data was derived from the Singapore SARS outbreak (*n*=180) by digitising the histogram of Lipsitch *et al.*'s study (Lipsitch, Cohen et al. 2003). The data have mean 8.4 days (SD 3.8 days) and maximum observation at 20 days. Naïvely fitting a Weibull distribution directly to the data (truncating the distribution at 20 days) gave the scale parameter of 9.39 and shape parameter of 2.21.

The goodness-of-fit of the different convoluted serial interval models were compared by their deviance. The following describes the probability distributions explored in modelling the serial interval data:

1. Incubation period distribution
    i) Weibull distribution

    Cowling et al. (Cowling, Muller et al. 2007) compared the goodness of fit of lognormal, Weibull and Gamma distribution to the incubation period distribution of Hong Kong (*n*=168) and Toronto (*n*=149) SARS cases, and concluded the lognormal function fitted best. However, the difference in the Akaike's Information Criterion (AIC) was only marginal for the Hong Kong data (395 vs 396) comparing lognormal and Weibull distributions, while the former had a thicker right tail, with the 99th percentile at 22.3 days (longer than the observed maximum serial interval), compared with 15.9 and 15.0 days for the Weibull and gamma distributions. The gamma distribution used for other Hong Kong SARS studies (Donnelly, Ghani et al. 2003; Leung, Hedley et al. 2004) fitted worst, reinforcing the findings by Farewell *et al.* (Farewell, Herzberg et al. 2005) that various probability distributions should be considered and compared in estimating the incubation period density.

    In this study, I used the Weibull distribution (scale and shape parameters at 4.60 and 1.23) to describe the incubation distribution [Figure 5.2(b)]. The mean was assumed to be 4.3 days (SD 3.5 days), and the distribution was truncated at 14 days (i.e. the 98th percentile of the Weibull incubation period distribution). This maximum period is longer than the 10 days quarantine for suspected cases adopted during the outbreak.

ii) Constant

The use of Weibull incubation period distribution in inferring the date of infection gives a probability distribution of the date of infection and the list of potential infectors. A simplified assumption of constant incubation period gives a specific date of infection. With the mean and the mode of the incubation period were 4.3 and 2 days, I explored and assessed the effects of using constant incubation periods of 2, 3 or 4 days respectively on deriving the infectiousness density.

Assuming a constant incubation assumption, the infectiousness distribution could not simply be considered as an offset of the serial interval distribution, because if

$$q(c) = \begin{cases} 1, & c = k \\ 0, & c \neq k \end{cases}$$

then $t_j = s_{ij} - k$, giving $w(s_{ij}) = u(s_{ij} - k)$, which contradicts the assumption that infectiousness starts with the onset of symptoms for $s_{ij} < k$. Therefore a proper deconvolution of the serial interval was required even under constant incubation assumption.

2. Infectiousness distribution

Both Weibull and gamma distributions were explored to describe the temporal development of infectiousness. Based on the maximum observed serial interval data of 20 days, I assumed the maximum duration of infectiousness to be 18 days when the incubation period was assumed a constant 2 days, 17 days when incubation was 3 days, 16 days when the incubation period was 4 days, and 14 days when using a Weibull incubation distribution assumption truncated at 14 days. In the modified method, this was considered as the maximum duration of infectiousness, and the infectious period was allowed to vary based on the hospitalisation data.

5.3.2 Results

Table 5.1 shows that assuming a constant incubation period of 2 days with Weibull distributed infectiousness gave the best fitting convoluted serial interval model (deviance=20.99), and was comparable to fitting a Weibull serial interval distribution directly to the observed serial interval data (deviance=20.95). A Weibull infectiousness distribution was also preferred when the incubation period was assumed to be Weibull distributed (deviance=25.10), but a gamma function for infectiousness fitted best when a constant incubation period of 3 or 4 days was assumed.

For the rest of this study, the convoluted density of fixed 2-day incubation period and Weibull distributed infectiousness is referred as the "2-day convolution", and the corresponding

infectiousness density as “2-day infectiousness” respectively, and that of the 3 and 4-day incubation with gamma infectiousness distribution are referred as “3-day convolution”, “3-day infectiousness” and “4-day convolution”, “4-day infectiousness” respectively, and that of the Weibull distributed incubation and infectiousness are referred as “Weibull convolution” and “Weibull infectiousness”.

Table 5.1 Parameter estimates of the infectiousness distribution and the goodness-of-fit of the convoluted serial interval under different incubation period assumptions.

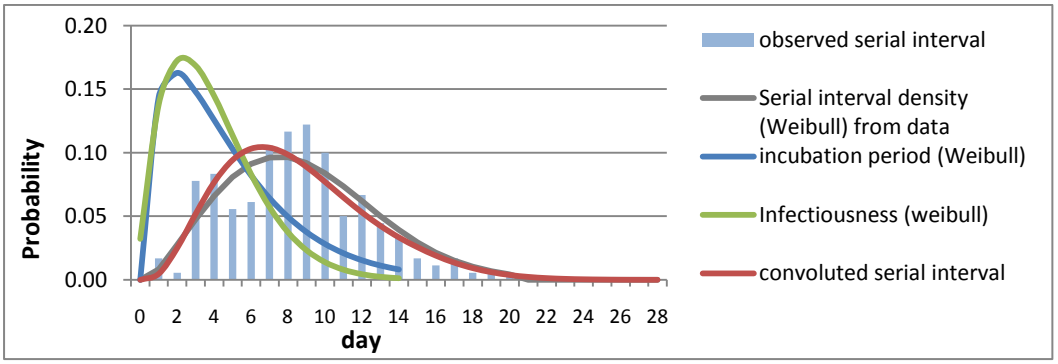
Incubation period distribution	**Weibull**		**Constant (4 days)**		**Constant (3 days)**		**Constant (2 days)**	
Infectiousness distribution	**Weibull**	**Gamma**	**Weibull**	**Gamma**	**Weibull**	**Gamma**	**Weibull**	**Gamma**
Maximum duration (days)	14	14	16	16	17	17	18	18
Parameter estimates								
a (scale parameter)	4.32	1.53	5.38	3.79	6.20	4.57	6.90	3.04
b (shape parameter)	1.59	2.84	1.08	1.46	1.06	1.38	1.56	2.29
Mean	3.87	4.31	4.65	5.06	5.44	5.57	6.06	6.57
Standard deviation	2.49	2.51	3.79	3.70	4.11	4.08	3.82	3.94
Convoluted serial interval								
Log likelihood	263	261	225	233	248	255	265	264
Deviance	25.10	28.68	99.47	85.06	55.18	40.96	20.99	22.46
Mean	8.45	7.91	9.10	9.11	8.62	8.62	8.53	8.56
Standard deviation	3.96	4.00	3.73	3.65	4.10	4.02	3.79	3.92

The 2-day infectiousness had higher mean and mode (4 days) than the Weibull infectiousness [mode 3 days], and was less sharply peaked with a fatter tail [Figure 5.2 (b)]. Although the 4-day infectiousness was closest to the Weibull infectiousness in shape and with the mode a day later, the 4-day convolution gave zero serial interval density for days 1-3 due to the incubation period assumption, and that contributed to a higher and earlier peak of the 4-day convolution than the Weibull convolution [Figure 5.2 (c)]. Besides, being least able to catch the peak of the observed serial interval at day 3 and 4, but peaked at the trough of the observed data at day 5 and 6 had made it the least fitting model to the observed serial interval data.

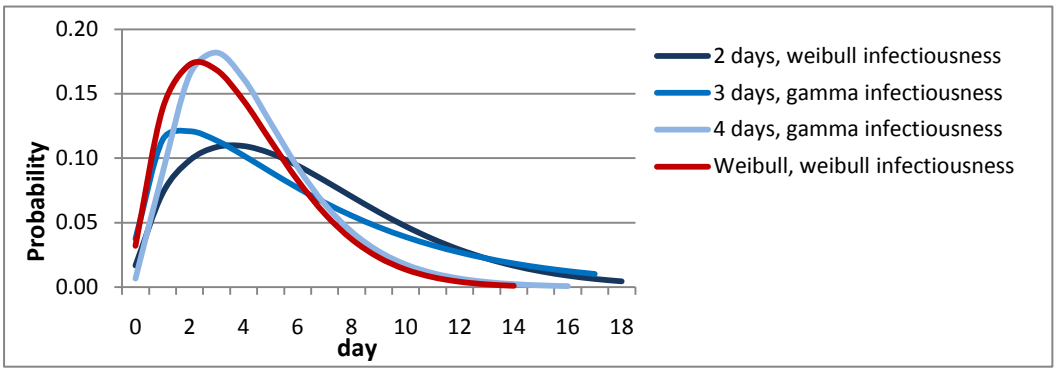
Because of the bimodal shape of the observed serial interval distribution, even the Weibull distribution fitted directly to the data underestimated the two peak probabilities while overestimated the probabilities between the peaks, with a left shift of the single peak to day 7 and 8. The Weibull convolution had a further left shift of the peak to day 6 and 7, and the 2-day convolution to day 6.

Figure 5.2 (a) The convoluted serial interval density using Weibull incubation and infectiousness distribution, compared to the observed and Weibull fitted serial interval distribution. Comparison of: (b) the derived infectiousness density, and (c) the convoluted serial interval density, using Weibull distributed or constant incubation of 2, 3 or 4 days respectively. (d) The number of potential infectors by symptom onset date based on different assumptions to identify potential infectors.

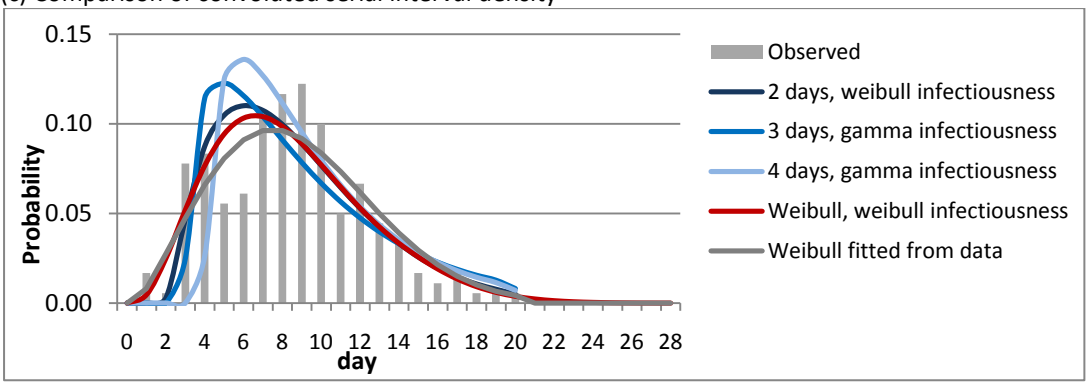
(a) Weibull distributed incubation and infectiousness



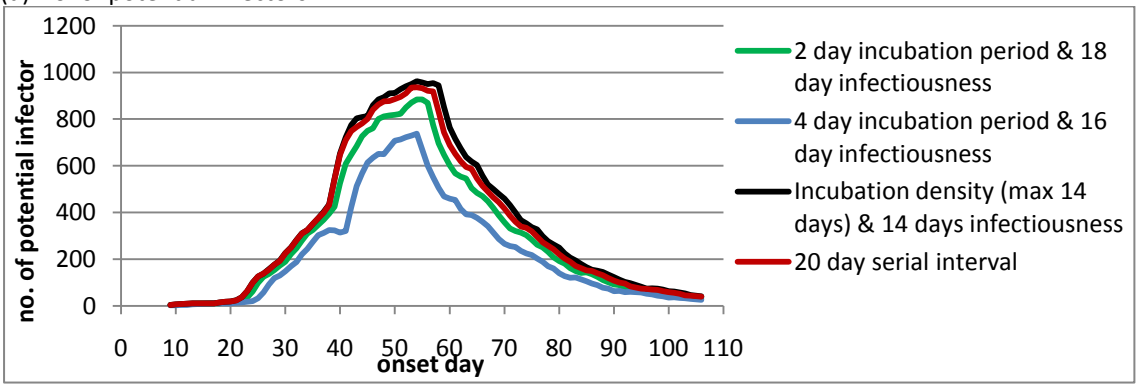
(b) Comparison of infectiousness density



(c) Comparison of convoluted serial interval density



(d) No. of potential infectors



Besides how well the convoluted serial interval fitted to the data, it also matters in estimating *R* how many of the cases are identified as potential infectors through time. The 2-day convolution model implied that potential infectors were those with symptom onset between 2 and 20 days before the symptom onset of the infectee. By comparison, the Weibull convolution implies potential infectors have symptom onset 1 to 28 days earlier before that of the infectee. During the first 3 weeks of the epidemic where the number of cumulative cases was small, the difference in the number of potential infectors for cases with symptom onset during this period is minimal [Figure 5.2 (d)]. However, at the peak of the epidemic, each case had 962 potential infectors under the Weibull convolution. The 2-day convolution gave 80 (SD 48) or 15.7% fewer potential infectors per infectee, and the 4-day convolution had 196 (SD 110) or 37.1% fewer. A broader coverage of potential infectors might reduce the chance of missing out the real infector with long serial interval, but at the cost that the infection might be incorrectly attributed to more cases who were not the true index case. This latter issue means that estimates of *R*(t) and individual reproduction numbers tended to be further smoothed out and less able to reflect the real fluctuations along the epidemic.

Taking into consideration the above results, I used the 2-day convolution model for my baseline analysis in this and later chapters, and (given data suggesting a mean 4 day incubation period) investigated the impact of assuming the 4-day convolution model in sensitivity analyses.

5.3.3 Discussion

Estimation of the profile of infectiousness over time in an infected individual is not straightforward because disease transmission is not observable. Viral load is considered a proxy of infectiousness with the most common assumption being that they are proportional to each other (Anderson, Fraser et al. 2004). For SARS, viral load peaked between 6 to 10 days after symptom onset based on a study of 14 patients (Peiris, Chu et al. 2003). A Bayesian analysis found that a gamma function was the best fitting distributional form (compared with a uniform or triangular distribution), with a peak 9 days after symptom onset (McBryde, Gibson et al. 2006). An unpublished study reported in the WHO meeting during the SARS outbreak suggested peak infectiousness was 7 to 8 days after symptom onset, based on the generation of secondary cases in the Singapore outbreak (Anderson, Fraser et al. 2004). Generally, analyses of viral load data have therefore been consistent with epidemiological data and my analysis presented above.

Another study used the contact tracing data of the Hong Kong SARS outbreak to estimate the daily transmission probability and incubation period simultaneously using a Bayesian discrete-time likelihood model by Markov chain Monte Carlo method. It was estimated that in the household

setting, the probability of transmission was bimodal, peaked at day 2 and 10 after symptom onset (Pitzer, Leung et al. 2007). The authors suggested any discrepancy to the viral shedding findings with their study were due to the infectiousness of index cases and the susceptibility of the contacts. The early peak could be due to infection by highly infectious people or to highly susceptible people, or to caregivers whose exposure was relatively high, or it could also be due to higher compliance with control measures later on (leading to right-censoring of the serial interval distribution), or co-primary infection from outside the household.

However, because the serial interval observed in household outbreaks tends to be shorter than that of community outbreaks (Fine 2003), peak infectiousness estimated from household data could also be earlier than in the non-household settings due to close proximity and more intense exposure. Also contacts identified in contact tracing were likely to be restricted to those whose name was known and could be recalled. Therefore exposure outside household and workplace might have had a lower chance to be included. Pitzer *et al*.'s work excluded non-household contacts which may mean the resulting infectiousness function is biased for use outside the household. On the other hand, the second peak at day 10 was the higher peak, but the number of infectious individual not admitted to hospital 10 days after symptom onset was extremely small (2.5%), and in Pitzer *et al*.'s data only 25% of the household exposure last for more than 5 days.

The idea that there might be a second peak in infectiousness 10 days after symptom onset was further challenged by another study which suggested mean duration of infectiousness was 9.3 days (95% CI 8.6-9.9 days) using a stochastic compartmental model using MCMC methods for parameter estimation (Cori, Boelle et al. 2009). If the second peak of infectiousness at day 10 were ignored, and the peak of 2 day after symptom onset were considered as biased for the general settings, then the true peak infectiousness would be between 3 to at most 9 days after symptom onset. This is consistent with the findings of this study under different assumptions of the incubation period.

Although the Singapore serial interval data used above showed slight evidence of bimodality, the second peak in that data occurred at day 9, which is inconsistent with infectiousness peaking at day 10. Serial interval data collected from contact tracing could be subject to the false assignment of tertiary cases to the index case of the household or disease cluster, resulted in a pseudo-bimodal distribution, and this is a feasible explanation of the observed serial interval data with the time delay between the first and second peak being comparable to the time of the first peak.

Some of the assumptions I make in modelling the serial interval distribution above could be oversimplified, though they are also commonly used in other epidemic studies. The serial interval

and infectiousness distributions were assumed not to change over time, though it was shown that mean serial interval in reality reduced as the epidemic progressed (Lipsitch, Cohen et al. 2003), perhaps due to saturation of infection hazards in exposed individuals (Svensson 2007; Kenah, Lipsitch et al. 2008), but more likely due the effects of more effective control measures such isolation and quarantine. In using serial interval data from the Singapore outbreak and incubation period data from Hong Kong, I implicitly assume distributions did not vary between countries. Without the availability of additional data, it is difficult to resolve these issues.

Lastly, my assumption of a constant incubation period dramatically simplifies the application of Wallinga and Teunis type analyses, but of course is an approximation and is somewhat unrealistic. The impact of incubation period assumptions is assessed in sensitivity analyses.

5.4 Partial extension of Wallinga and Tenuis's method

5.4.1 Method

The second objective of this chapter is to extend the Wallinga and Teunis method to include spatial heterogeneity using the gravity models derived in the previous chapter. The original method was modified in the following ways:

(1) I explicitly modelled the time of infection rather than just model the times of onset of disease. Hence the criterion applied for case *i* to be a potential infector of case *j* was that case *i* had to be infectious at the date of infection of cases *j*.

(2) Instead of just using the serial interval density to estimate the relative likelihood of infection of case *i* to case *j*, I used the product of the probability of contact given the spatial location of the cases *i* and *j*, denoted as $p^c(x_i,x_j;\phi)$ [with x_i being the spatial location of case *i*] and the convolution of the infectivity density $u(t_j\text{-}s_i;\psi)$ and incubation period density $q(s_j\text{-}t_j;\eta)$ [where s_i is the symptom onset date of case *i*, t_i the time of infection, and ϕ, ψ, η are the vectors of parameters of the respective function].

Let

$$g(s_{v(j)},s_j,x_{v(j)},x_j;\phi,\psi,\eta)=(p^c(x_{v(j)},x_j;\phi)\sum_{t_j=s_{v(j)}}^{s_j}u(t_j-s_{v(j)};\psi)q(s_j-t_j;\eta))$$

This modifies the relative probability of case *i* infected case *j* as:

$$p^{I}_{i->j} = \frac{g(s_i, s_j, x_i, x_j; \phi, \psi, \eta)}{\sum_{m=1, m\neq j}^{m=n} g(s_m, s_j, x_m, x_j; \phi, \psi, \eta)} \tag{5.6}$$

$$= \frac{p^c(x_i, x_j; \phi) \sum_{t_j=s_i}^{s_j} u(t_j - s_i; \psi) q(s_j - t_j; \eta)}{\sum_{m=1, m\neq j}^{m=n} \left\{ p^c(x_m, x_j; \phi) \sum_{t_j=s_i}^{s_j} u(t_j - s_m; \psi) q(s_j - t_j; \eta) \right\}}$$

The individual and effective reproduction numbers were then estimated as in the original Wallinga and Teunis method as:

$$E[R(i)] = \sum_{k=1}^{k=n-q} p^{I}_{i->k}$$

$$= \sum_{j=1}^{j=n-q} \frac{p^c(x_i, x_j; \phi) \sum_{t_j=s_i}^{s_j} u(t_j - s_i; \psi) q(s_j - t_j; \eta)}{\sum_{m=1, m\neq j}^{m=n} \left\{ p^c(x_m, x_j; \phi) \sum_{t_j=s_i}^{s_j} u(t_j - s_m; \psi) q(s_j - t_j; \eta) \right\}}$$

And

$$E[R(t)] = \frac{1}{n_t} \sum_{t_j=t} \sum_{j=1}^{j=n-q} \frac{p^c(x_i, x_j; \phi) \sum_{t_j=s_i}^{s_j} u(t_j - s_i; \psi) q(s_j - t_j; \eta)}{\sum_{m=1, m\neq j}^{m=n} \left\{ p^c(x_m, x_j; \phi) \sum_{t_j=s_i}^{s_j} u(t_j - s_m; \psi) q(s_j - t_j; \eta) \right\}}$$

It was noted that the $R(i)$ and $R(t)$ estimates are invariant for any constant multiplicative scaling of $g()$. The formulation of $p^c(x_i,x_j;\phi)$ was derived from the gravity models and varied under different spatial mixing assumptions. These are given in the following section.

5.4.2 Spatial mixing assumptions

Previous epidemic models incorporating gravity model to represent the relative probability of contact between patches have assumed either susceptible-driven or infector-driven spatial movement of infection. In this study, I investigated these two 'unilateral' assumptions together with bilateral and multilateral spatial mixing, deriving expressions for $p^c(x_i,x_j;\phi)$ for each.

5.4.2.1 Unilateral movement

Let T be the origin-destination trip matrix with components $T_{i'j'}$ being the number of trips from district *i'* to *j'*. Denote $p^{T}_{i->j'}$ as the probability of a specific individual in *i* travelling to district *j'*, and $p^{C}_{i->j}$ as the probability of a specific individual in *i* travelling to district *j'* and making contact with a

specific individual in *j'*. Let $N_{i'}$ and $N_{j'}$ be the population sizes of *i'* and *j'* respectively. Under unilateral mixing, depending on who moves, the probability of travel and probability of contact between individuals *i* (from patch *i'*) and *j* (from patch *j'*) can be written as:

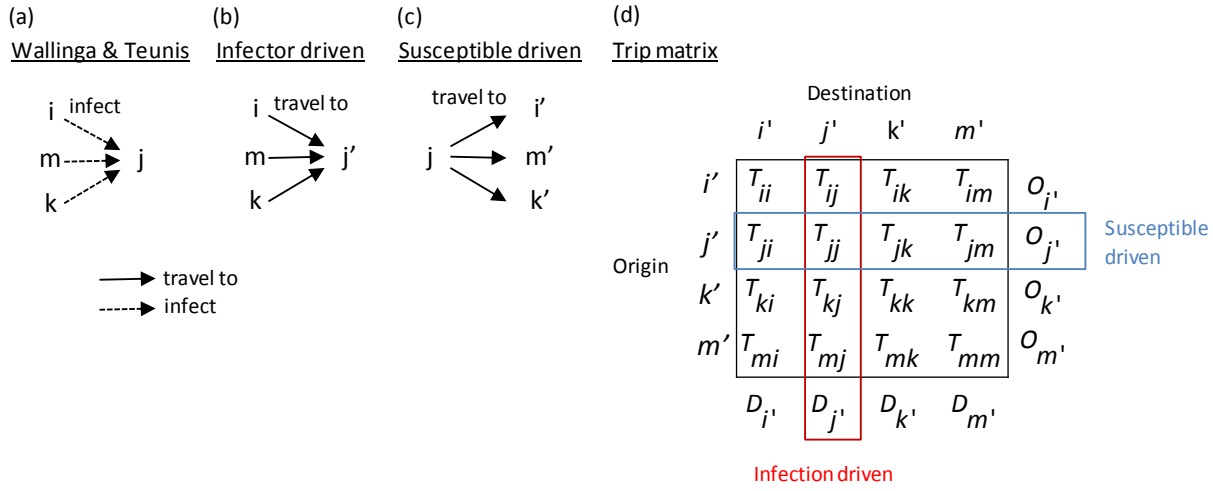
$$p^{T}{}_{i\to j'} = \frac{T_{i'j'}}{N_{i'}}, \quad p^{C}{}_{i\to j} = \frac{T_{i'j'}}{N_{i'}N_{j'}}$$

$$p^{T}{}_{j\to i'} = \frac{T_{j'i'}}{N_{j'}}, \quad p^{C}{}_{j\to i} = \frac{T_{j'i'}}{N_{i'}N_{j'}}$$

To better illustrate how the gravity model was incorporated into the Wallinga and Teunis method assuming unilateral mixing, I considered without loss of generality a population being divided into 4 districts, and with 4 cases *i, j, k, m* living in district *i', j', k'* and *m'* respectively. The Wallinga and Teunis formulation for the probability that *i* infected *j* conditional upon *j* being infected on the observed day of infection is the relative probability (up to an unknown multiplicative scaling constant) that *i* infected *j* divided by the sum of the relative probability of infection from all possible infectors on that day, *i.e.* the denominator is the sum of the likelihood of *j* being infected by *k*, *m* and *i*, as illustrated in Figure 5.3 (a). The unilateral mixing assumption implies either susceptible *j* moved to district *i'* and got infected by *i* , or infector *i* went to district *j'* and infected *j*. Infector driven mixing assumed infectors *i, k, m* moved to *j'* [Figure 5.3 (b)], and the corresponding trip matrix components were $T_{i'j'}$, $T_{k'j'}$, $T_{m'j'}$ [Figure 5.3 (d)]. Susceptible driven mixing assumes susceptible *j* moved to district *i'*, *m'* or *k'* [Figure 5.3 (c)], with the trip matrix components $T_{j'i'}$, $T_{j'k'}$, $T_{j'm'}$ [Figure 5.3 (d)].

In the previous chapter, I derived gravity models for $T_{i'j'}$ using population sizes N_{i}' and N_{j}' and spatial distance between patches. It is necessary to use constrained gravity models when extending the Wallinga and Teunis method. In an outflow constrained model one conditions on the total number of journeys from a patch and partitions these journeys between possible destinations according to their relative attractiveness (as quantified by the gravity model) . This formulation is appropriate for use in a susceptible driven unilateral movement model. An inflow constrained model conditions on the marginal inflow of a district partitions this flow between possible origins according to the relative attractiveness of the destination to those origin patches. This was the appropriate formulation for use with the infector driven model which assumes all possible infectors have to travel to the patch of the susceptible [Figure 5.3 (b,d)]. This simple example excluded the scenario where *j* is infected within its own district just for the purpose of illustration, though of course this possibility is allowed in the model.

Figure 5.3 Schematic presentation of the relationship between potential infectors *i, m, k* to infectee *j* under: (a) Wallinga and Teunis method; (2) Infection driven movement; (c) susceptible driven movement; (d) trip matrix, with $O_{i'}$ and $D_{i'}$ denote the outflow and inflow of district *i'*.



5.4.2.1.1 Susceptible driven movement

To generalise, let the SARS cases be *m*=1,….,*n*, and the Hong Kong population be partitioned by their residential location into *s* = 1,…., *S* districts, and let the population of district *s* be denoted N_s. Denote the residing district of case *m* as *m'*. The outflow constrained models derived in the previous chapter, where O_s is the marginal outflow of district s, are of the form:

$$T_{j'i'} = \frac{O_{j'} N_{i'}^{\beta_1+\beta_2} f(d_{j'i'})}{\sum_{s=1}^{S} N_s^{\beta_1+\beta_2} f(d_{j's})} \qquad \beta_2 = \begin{cases} \beta' & \text{for districts with high inflow / outflow ratio} \\ 0 & \text{otherwise} \end{cases}$$

And $p^T{}_{j\to i'}$ the probability of an individual residing in j' to travel to i' is:

$$p^T{}_{j\to i'} = \frac{T_{j'i'}}{N_{j'}} = \frac{O_{j'} N_{i'}^{\beta_1+\beta_2} f(d_{j'i'})}{N_{j'} \sum_{s=1}^{S} N_s^{\beta_1+\beta_2} f(d_{j's})}$$

The probability of contact $p^C{}_{j\to i}$ between cases *i* and *j* (using the residential location instead of the centroid of i' and j' as it allows higher spatial resolution) is:

$$p^C{}_{j\to i} = \frac{O_{j'} N_{i'}^{\beta_1+\beta_2} f(d_{ji})}{N_{j'} \sum_{s=1}^{S} N_s^{\beta_1+\beta_2} f(d_{js})} \frac{1}{N_{i'}}$$

$$= \frac{O_{j'} N_{i'}^{\beta_1+\beta_2-1} f(d_{ji})}{N_{j'} \sum_{s=1}^{S} N_s^{\beta_1+\beta_2} f(d_{js})}$$

Hence the probability that *i* infects *j* under the susceptible driven assumption, $p^I{}_{s\,i\to j}$, is:

$$p^{I}_{s\ i->j} = \frac{\frac{O_{j'} N_{i'}^{\beta_1+\beta_2-1} f(d_{ji})}{N_{j'} \sum_{s=1}^{S} N_s^{\beta_1+\beta_2} f(d_{js})} \sum_{t_j=s_i}^{s_j} u(t_j - s_i;\psi) q(s_j - t_j;\eta)}{\sum_{m=1,m\neq j}^{m=n} \left\{ \frac{O_{j'} N_{m'}^{\beta_1+\beta_2-1} f(d_{jm})}{N_{j'} \sum_{s=1}^{S} N_s^{\beta_1+\beta_2} f(d_{js})} \sum_{t_j=s_i}^{s_j} u(t_j - s_m;\psi) q(s_j - t_j;\eta) \right\}}$$

$$= \frac{N_{i'}^{\beta_1+\beta_2-1} f(d_{ji}) \sum_{t_j=s_i}^{s_j} u(t_j - s_i;\psi) q(s_j - t_j;\eta)}{\sum_{m=1,m\neq j}^{m=n} \left\{ N_{m'}^{\beta_1+\beta_2-1} f(d_{jm}) \sum_{t_j=s_i}^{s_j} u(t_j - s_m;\psi) q(s_j - t_j;\eta) \right\}}$$

The ratio $O_{j'}/N_{j'}$ cancels from numerator and denominator. The resulting probability of contact component within this probability of infection is the product of the scaled population size of the district and the spatial kernel function.

5.4.2.1.2 Infector driven movement

Now consider an inflow constrained gravity model of the form below, with $D_{i'}$ being the marginal inflow to district *i'*, and α_1, α_2 the power parameters of the gravity model on origin district population size:

$$T_{i'j'} = \frac{D_{j'} N_{i'}^{\alpha_1+\alpha_2} f(d_{i'j'})}{\sum_{s=1}^{S} N_s^{\alpha_1+\alpha_2} f(d_{sj'})} \qquad \alpha_2 = \begin{cases} \alpha' & \text{for districts with high inflow / outflow ratio} \\ 0 & \text{otherwise} \end{cases}$$

The probability of travel $p^{T}_{i\to j'}$ of an individual in *i'* to *j'* is:

$$p^{T}_{i\to j'} = \frac{T_{i'j'}}{N_{i'}} = \frac{D_{j'} N_{i'}^{\alpha_1+\alpha_2-1} f(d_{i'j'})}{\sum_{s=1}^{S} N_s^{\alpha_1+\alpha_2} f(d_{sj'})}$$

And the probability of contact $p^{C}_{i\to j}$ through *i* travel to *j'* and contact *j* is:

$$p^{C}_{i\to j} = \frac{D_{j'} N_{i'}^{\alpha_1+\alpha_2-1} f(d_{ij})}{N_{j'} \sum_{s=1}^{S} N_s^{\alpha_1+\alpha_2} f(d_{sj})}$$

This gives:

$$p^{I}_{I\ i->j} = \frac{\dfrac{D_{j'} N_{i'}^{\alpha_1+\alpha_2-1} f(d_{ij})}{N_{j'} \sum_{s=1}^{S} N_s^{\alpha_1+\alpha_2} f(d_{sj})} \sum_{t_j=s_i}^{s_j} u(t_j - s_i;\psi) q(s_j - t_j;\eta)}{\sum_{m=1,m\neq j}^{m=n} \left\{ \dfrac{D_{j'} N_{m'}^{\alpha_1+\alpha_2-1} f(d_{mj})}{N_{j'} \sum_{s=1}^{S} N_s^{\alpha_1+\alpha_2} f(d_{sj})} \sum_{t_j=s_i}^{s_j} u(t_j - s_m;\psi) q(s_j - t_j;\eta) \right\}}$$

$$= \frac{N_{i'}^{\alpha_1+\alpha_2-1} f(d_{ij}) \sum_{t_j=s_i}^{s_j} u(t_j - s_i;\psi) q(s_j - t_j;\eta)}{\sum_{m=1,m\neq j}^{m=n} \left\{ N_{m'}^{\alpha_1+\alpha_2-1} f(d_{mj}) \sum_{t_j=s_i}^{s_j} u(t_j - s_m;\psi) q(s_j - t_j;\eta) \right\}}$$

This is similar to the comparable expression for the susceptible driven model, with the only difference being that the gravity model power parameters represent the attractiveness of origin patches in one model and that of destination patches in the other. Under the gravity model with α_1=1 and α_2=0 (or the susceptible driven model when β_1=1 and β_2=0), dependence on patch population size is lost and the spatial component only contains the spatial kernel (referred to as the "distance only" model later). Susceptible-driven and Infector-driven models are equivalent in this special case.

5.4.2.1.3 Observed trip patterns

Besides using gravity models, I also used the observed trip matrix directly to derive the probability of contact and incorporate it into the extended Wallinga and Teunis method. The mathematical formulations for the resulted probability of infection estimation are given below, and the probability of infection from case i to case j using the observed trip patterns is denoted as $p_{O\ i\to j}^{I}$. However, spatial resolution on the individuals' location within the district was lost in the partially modified method using the observed trip patterns directly, because the data only give total flows between districts. Therefore the observed trip modified method was used to check the general consistency rather than evaluating the modified methods that used the gravity models.

Under susceptible driven movement we obtain:

$$p_{O\ i->j}^{I} = \frac{\dfrac{T_{j'i'}}{N_{i'} N_{j'}} \sum_{t_j=s_i}^{s_j} u(t_j - s_i;\psi) q(s_j - t_j;\eta)}{\sum_{m=1,m\neq j}^{m=n} \left\{ \dfrac{T_{j'm'}}{N_{m'} N_{j'}} \sum_{t_j=s_i}^{s_j} u(t_j - s_m;\psi) q(s_j - t_j;\eta) \right\}}$$

$$= \frac{\dfrac{T_{j'i'}}{N_{i'}} \sum_{t_j=s_i}^{s_j} u(t_j - s_i;\psi) q(s_j - t_j;\eta)}{\sum_{m=1,m\neq j}^{m=n} \left\{ \dfrac{T_{j'm'}}{N_{m'}} \sum_{t_j=s_i}^{s_j} u(t_j - s_m;\psi) q(s_j - t_j;\eta) \right\}}$$

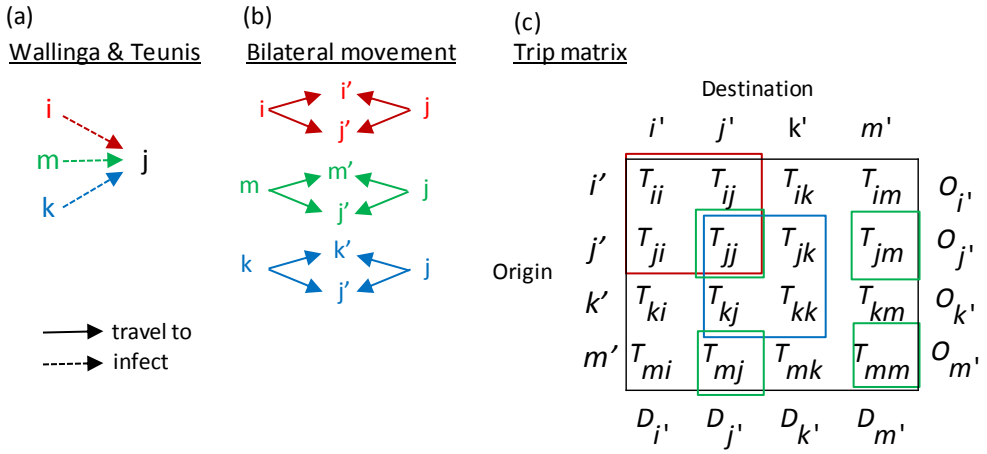
Under infector driven movement the expression is:

$$p_{O\ i->j}{}' = \frac{\frac{T_{i'j'}}{N_{i'}N_{j'}}\sum_{t_j=s_i}^{s_j} u(t_j - s_i;\psi)q(s_j - t_j;\eta)}{\sum_{m=1,m\neq j}^{m=n}\left\{\frac{T_{m'j'}}{N_{m'}N_{j'}}\sum_{t_j=s_i}^{s_j} u(t_j - s_m;\psi)q(s_j - t_j;\eta)\right\}}$$

$$= \frac{\frac{T_{i'j'}}{N_{i'}}\sum_{t_j=s_i}^{s_j} u(t_j - s_i;\psi)q(s_j - t_j;\eta)}{\sum_{m=1,m\neq j}^{m=n}\left\{\frac{T_{m'j'}}{N_{m'}}\sum_{t_j=s_i}^{s_j} u(t_j - s_m;\psi)q(s_j - t_j;\eta)\right\}}$$

5.4.2.2 Bilateral movement

Figure 5.4 Schematic presentation of the (a) Wallinga and Teunis method and (b) bilateral movement assumption, with *i* as the infector and *j* as the infectee and in the same settings described in Chapter 5.4.2.1. (c) Relevant components in the trip matrix under the bilateral movement assumption. The colour marks the corresponding potential infector.



Unilateral mixing is a simplified model of human movement. Bilateral mixing [Figure 5.4 (a-c)] allows susceptible to travel to infectors or vice versa with equal weight. I used an unconstrained gravity model in this case written as:

$$T_{i'j'} = k_u N_{i'}^{\alpha_{u1}+\alpha_{u2}} N_{j'}^{\beta_{u1}+\beta_{u2}} f(d_{i'j'}) \qquad \alpha_{u2} = \begin{cases} \alpha_u' & \text{for districts with high inflow / outflow ratio} \\ 0 & \text{otherwise} \end{cases}$$

$$\beta_{u2} = \begin{cases} \beta_u' & \text{for districts with high inflow / outflow ratio} \\ 0 & \text{otherwise} \end{cases}$$

The probability of contact between infector *i* and infectee *j* in *i'* is the probability of travel of *i* within *i'* and *j* to *i'*, divided by the traveller (i.e. worker or school student) population of *i'*. In the census data where each person contributed one trip to the data, the traveller population of a patch is just the marginal daily inflow into that patch, denoted as $D_{i'}$. Therefore the probability of contact between *i* and *j* under bilateral movement is:

$$p^{C}_{i\leftrightarrow j}=\frac{1}{2}\left(\frac{p^{T}_{i\rightarrow j'}p^{T}_{j\rightarrow j'}}{D_{j'}}+\frac{p^{T}_{j\rightarrow i'}p^{T}_{i\rightarrow i'}}{D_{i'}}\right)$$
$$=\frac{1}{2}\left(\frac{k_u N_{i'}{}^{\alpha_{u1}+\alpha_{u2}}N_{j'}{}^{\beta_{u1}+\beta_{u2}}f(d_{ij})}{N_{i'}}\frac{k_u N_{j'}{}^{\alpha_{u1}+\alpha_{u2}+\beta_{u1}+\beta_{u2}}f(d_{jj})}{N_{j'}}\frac{1}{D_{j'}}+\frac{k_u N_{j'}{}^{\alpha_{u1}+\alpha_{u2}}N_{i'}{}^{\beta_{u1}+\beta_{u2}}f(d_{ij})}{N_{j'}}\frac{k_u N_{i'}{}^{\alpha_{u1}+\alpha_{u2}+\beta_{u1}+\beta_{u2}}f(d_{ii})}{N_{i'}}\frac{1}{D_{i'}}\right)$$
$$=\frac{1}{2}k_u{}^2 f(0)f(d_{ij})(N_{i'}N_{j'})^{\alpha_{u1}+\alpha_{u2}-1}\left(\frac{N_{j'}{}^{2(\beta_{u1}+\beta_{u2})}}{D_{j'}}+\frac{N_{i'}{}^{2(\beta_{u1}+\beta_{u2})}}{D_{i'}}\right)$$

Incorporated into Wallinga and Teunis method, I get:

$$p_{B}{}^{I}{}_{i\rightarrow j}=\frac{\frac{1}{2}k_u{}^2 f(0)f(d_{ij})(N_{i'}N_{j'})^{\alpha_{u1}+\alpha_{u2}-1}\left(\frac{N_{j'}{}^{2(\beta_{u1}+\beta_{u2})}}{D_{j'}}+\frac{N_{i'}{}^{2(\beta_{u1}+\beta_{u2})}}{D_{i'}}\right)\sum_{t_j=s_i}^{s_j}u(t_j-s_i;\psi)q(s_j-t_j;\eta)}{\sum_{m=1,m\neq j}^{m=n}\left\{\frac{1}{2}k_u{}^2 f(0)f(d_{mj})(N_{m'}N_{j'})^{\alpha_{u1}+\alpha_{u2}-1}\left(\frac{N_{j'}{}^{2(\beta_{u1}+\beta_{u2})}}{D_{j'}}+\frac{N_{m'}{}^{2(\beta_{u1}+\beta_{u2})}}{D_{m'}}\right)\sum_{t_j=s_i}^{s_j}u(t_j-s_m;\psi)q(s_j-t_j;\eta)\right\}}$$
$$=\frac{f(d_{ij})N_{i'}{}^{\alpha_{u1}+\alpha_{u2}-1}\left(\frac{N_{j'}{}^{2(\beta_{u1}+\beta_{u2})}}{D_{j'}}+\frac{N_{i'}{}^{2(\beta_{u1}+\beta_{u2})}}{D_{i'}}\right)\sum_{t_j=s_i}^{s_j}u(t_j-s_i;\psi)q(s_j-t_j;\eta)}{\sum_{m=1,m\neq j}^{m=n}\left\{f(d_{mj})N_{m'}{}^{\alpha_{u1}+\alpha_{u2}-1}\left(\frac{N_{j'}{}^{2(\beta_{u1}+\beta_{u2})}}{D_{j'}}+\frac{N_{m'}{}^{2(\beta_{u1}+\beta_{u2})}}{D_{m'}}\right)\sum_{t_j=s_i}^{s_j}u(t_j-s_m;\psi)q(s_j-t_j;\eta)\right\}}$$

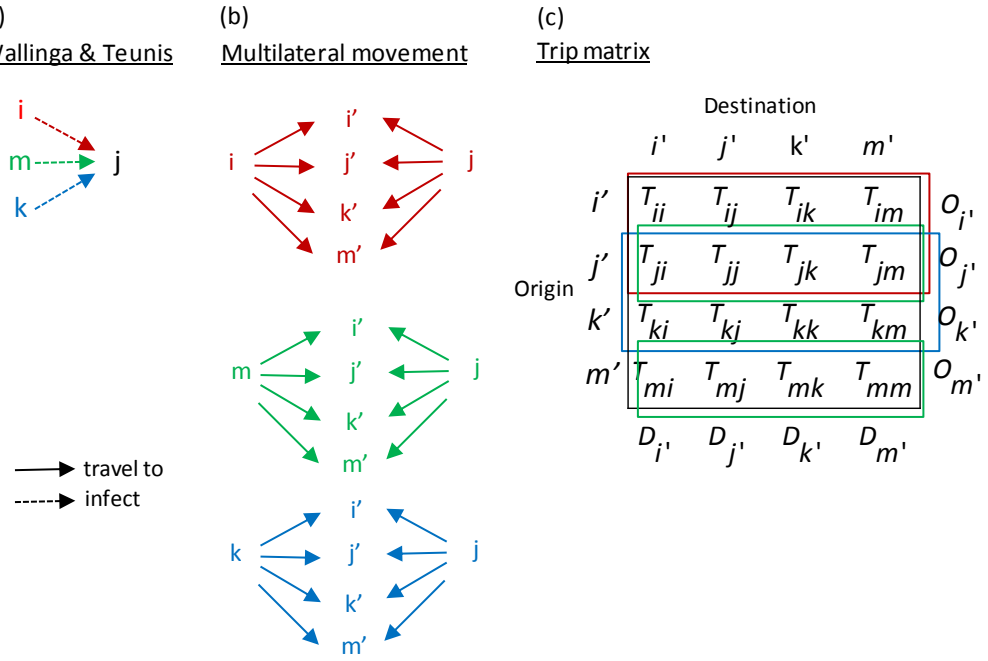
This formulation has both the attractiveness and propulsiveness of the home patch of the potential infectors affecting the probability of infection. Using observed trip patterns instead of the gravity model, the formulation is:

$$p_{B}{}^{I}{}_{i\rightarrow j}=\frac{\left(\frac{T_{ij}T_{jj}}{N_{i'}D_{j'}}+\frac{T_{ji}T_{ii}}{N_{i'}D_{i'}}\right)\sum_{t_j=s_i}^{s_j}u(t_j-s_i;\psi)q(s_j-t_j;\eta)}{\sum_{m=1,m\neq j}^{m=n}\left\{\left(\frac{T_{mj}T_{jj}}{N_{m'}D_{j'}}+\frac{T_{jm}T_{mm}}{N_{m'}D_{m'}}\right)\sum_{t_j=s_i}^{s_j}u(t_j-s_m;\psi)q(s_j-t_j;\eta)\right\}}$$

5.4.2.3 Multilateral movement

The multilateral movement assumption allows individuals *i* and *j* to meet in any patch. Therefore their probability of contact is the sum across all patches of the product of their probability of travelling to the same patch and making contact there [Figure 5.5 (a-c)]. The bilateral model is a subset of the multilateral scenario with the meeting patches being either *i'* or *j'*.

Figure 5.5 Schematic presentation of the (a) Wallinga and Teunis method and (b) multilateral movement assumption, with *i* as the infector and *j* as the infectee and in the same settings described in Chapter 5.4.2.1. (c) Relevant components in the trip matrix under the multilateral movement assumption. The colour marks the corresponding potential infector.



The formulation of the probability of contact under multilateral movement based on the unconstrained gravity model described earlier is:

$$p^{C}_{i\Leftrightarrow j} = \sum_{k=1}^{S} \frac{p^{T}_{i\to k} p^{T}_{j\to k}}{D_{k'}}$$

$$= \sum_{k=1}^{S} \frac{k_u N_{i'}^{\alpha_{u1}+\alpha_{u2}} N_{k'}^{\beta_{u1}+\beta_{u2}} f(d_{ik})}{N_{i'}} \frac{k_u N_{j'}^{\alpha_{u1}+\alpha_{u2}} N_{k'}^{\beta_{u1}+\beta_{u2}} f(d_{jk})}{N_{j'}} \frac{1}{D_{k'}}$$

$$= k_u^{2} (N_{i'} N_{j'})^{(\alpha_{u1}+\alpha_{u2}-1)} \sum_{k=1}^{S} \frac{f(d_{ik}) f(d_{jk}) N_{k'}^{2(\beta_{u1}+\beta_{u2})}}{D_{k'}}$$

Hence the probability that *i* infects *j* is:

$$p^{I}_{M\ i->j} = \frac{k_u^{2} (N_{i'} N_{j'})^{(\alpha_{u1}+\alpha_{u2}-1)} \sum_{k=1}^{S} \frac{f(d_{ik}) f(d_{jk}) N_{k'}^{2(\beta_{u1}+\beta_{u2})}}{D_{k'}} \sum_{t_j=s_i}^{s_j} u(t_j - s_i;\psi) q(s_j - t_j;\eta)}{\sum_{m=1,m\neq j}^{m=n} \left\{ k_u^{2} (N_{m'} N_{j'})^{(\alpha_{u1}+\alpha_{u2}-1)} \sum_{k=1}^{S} \frac{f(d_{mk}) f(d_{jk}) N_{k'}^{2(\beta_{u1}+\beta_{u2})}}{D_{k'}} \sum_{t_j=s_i}^{s_j} u(t_j - s_m;\psi) q(s_j - t_j;\eta) \right\}}$$

$$= \frac{N_{i'}^{\alpha_{u1}+\alpha_{u2}-1} \sum_{k=1}^{S} \frac{f(d_{ik}) f(d_{jk}) N_{k'}^{2(\beta_{u1}+\beta_{u2})}}{D_{k'}} \sum_{t_j=s_i}^{s_j} u(t_j - s_i;\psi) q(s_j - t_j;\eta)}{\sum_{m=1,m\neq j}^{m=n} \left\{ N_{m'}^{\alpha_{u1}+\alpha_{u2}-1} \sum_{k=1}^{S} \frac{f(d_{mk}) f(d_{jk}) N_{k'}^{2(\beta_{u1}+\beta_{u2})}}{D_{k'}} \sum_{t_j=s_i}^{s_j} u(t_j - s_m;\psi) q(s_j - t_j;\eta) \right\}}$$

Using the observed trip data, the equivalent expression is:

$$p_{M\ i->j}^{\ I} = \frac{\sum_{k=1}^{S} \frac{T_{ik}T_{jk}}{N_{i'}D_{k'}} \sum_{t_j=s_i}^{s_j} u(t_j - s_i;\psi)q(s_j - t_j;\eta)}{\sum_{m=1,m\neq j}^{m=n} \left\{ \sum_{k=1}^{S} \frac{T_{mk}T_{jk}}{N_{m'}D_{k'}} \sum_{t_j=s_i}^{s_j} u(t_j - s_m;\psi)q(s_j - t_j;\eta) \right\}}$$

5.4.3 Time course of infection assumptions

5.4.3.1 Variable infectious period

The extended Wallinga and Teunis model I derived assumes all individuals have equal duration of infectiousness. However, hospital discharge dates sometimes allow us to set a more realistic upper bound on the duration of infectiousness. Hence in assigning the end date of infectiousness to a case, I used the earlier of the imputed end of infectiousness (from the convolution model) and the observed hospital discharge date. Although the patients were likely have ceased being infectiousness a few days before discharge, there is a lack of reliable data on this to further improve the estimation of individual's infectious period.

5.4.3.2 Incubation period and infectiousness

The date of infection of the cases was inferred from the incubation period distribution. The 2-day incubation period and Weibull distributed infectiousness with 18 days infectious period (2-day convolution) were applied in this study as the baseline scenario. Infectiousness was assumed to commence on the date of symptom onset. Sensitivity analysis using a 4 days incubation period and gamma infectiousness density with 16 days infectious period (4 day convolution), and the Weibuall serial interval density were conducted.

5.4.4 Analysis plan

I first evaluated the effects of using the decovoluted serial interval and the variable infectious period respectively without incorporating any spatial component (referred as "non-spatial models"), and compared them to the original method. Then the spatially extended method was applied with different spatial mixing assumptions, and travel patterns represented by gravity models (Table 5.2), using the adjusted census overall trip data as the baseline because of its comprehensive coverage. The distance-only model was also fitted to examine the effect of population size dependence in the gravity model. In addition, because spatial data with high resolution may not be available in reality during the epidemic, I also tested the model using the centroid of the residing district of the cases rather than their residential location.

Table 5.2 Variants of the partially extended Wallinga and Teunis method studied.

Spatial mixing assumption	Trip patterns fitted to the gravity model	Probability of transmission based on temporal information
1. Infector driven 2. Susceptible driven 3. Bilateral 4. Multilateral 5. Distance only	1. Adjusted census overall 2. Census work 3. Census work-study 4. Survey overall 5. Observed trip patterns	1. Baseline: 2-day incubation, Weibull infectiousness and variable infectious period of maximum infectious period of 18 days 2. Sensitivity analysis: 4-day incubation, gamma infectiousness and variable infectious period of maximum 16 days 3. Sensitivity analysis: Serial interval density of 20 days

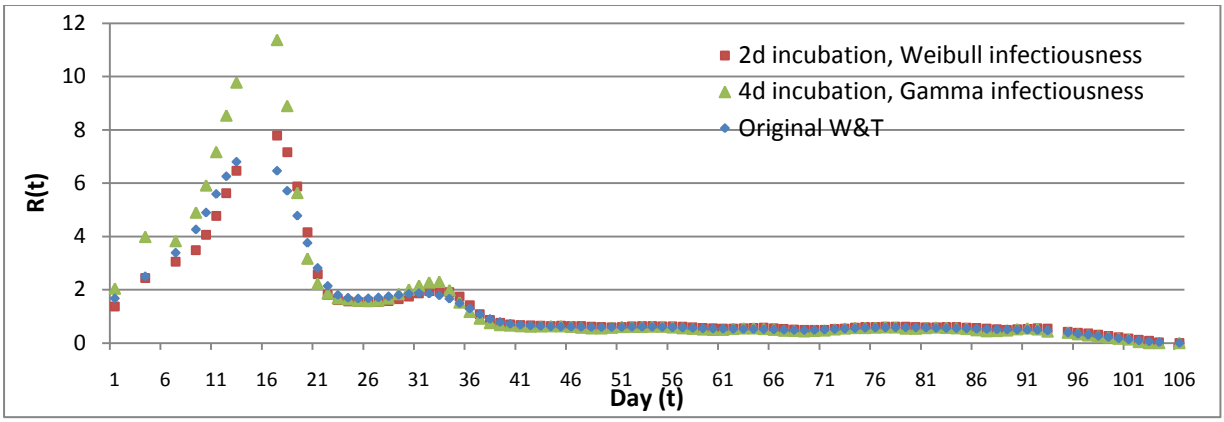
The key outputs of the partially modified method are the daily reproduction numbers $R(t)$, the reproduction number of individual cases $R(i)$, and the percentage of within-district transmission in the estimated reproduction numbers. Evaluation of the method is not straightforward. Use of simulated data can test inference ability of models, but of course does not test whether a model really incorporates the mechanisms underlying the observed spatiotemporal spread of an epidemic. Therefore I compared model results to epidemiological linkage data for the Hong Kong SARS epidemic, and in particular focus on the PWH and Amoy Gardens outbreaks, whose transmission dynamics were more widely studied. These events served as a good test of the ability of the method to capture the full extent of spatial heterogeneity and within-district transmission.

5.4.5 Results

5.4.5.1 Deconvolution model

The non-spatial models using convoluted densities had higher, sharper and later peak of *R(t)* at around the time of the PWH and Amoy Gardens super-spreading events, but otherwise were similar to the results from the original Wallinga and Teunis method (Figure 5.6). The difference observed can be attributed to the characteristics of the convoluted densities compared with that of the original serial interval density [Figure 5.2 (c)]. The earlier peak of the convoluted densities translated into a later peak *R(t)* for these clusters compared with the original method (PWH peak: day 17, i.e. 4 days later; Amoy peak: day 33, i.e. 2 days later). Because *t* in *R(t)* corresponds to the symptom onset date of cases rather than the time the secondary transmission occurs, *R(t)* peaks earlier than the peak case incidence of the observed clusters, at day 23 and 38. This time gap was more realistic in the non-spatial method (5-6 days) than the original method (8-9 days).

Figure 5.6 Estimated reproduction number *R(t)* using the non-spatial method with the deconvolution of serial interval density of: (i) 2 days incubation and Weibull infectiousness of 18 days, (ii) 4 days incubation and gamma infectiousness of 16 days, compared to the original Wallinga and Teunis method.

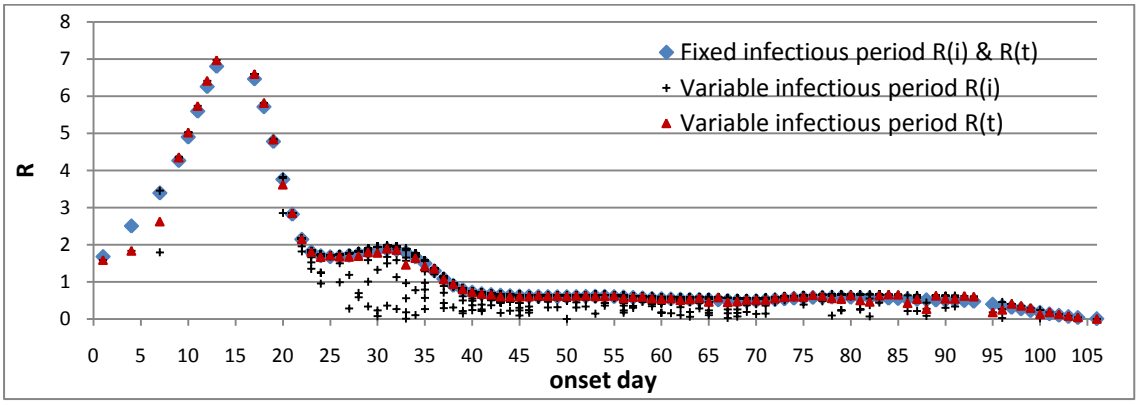


The differences in *R(t)* compared with the original method were larger for the PWH outbreak due to the small and sporadic number of cases (e.g. absence of cases with onset from day 14 to 16) in the initial period of the outbreak, and the number of potential infectors was further reduced under the 4-day incubation period assumption, because individuals with symptom onset 1 to 3 days before the case's onset were not eligible to be infectors. It remained difficult to identify the index cases of the super-spreading events (whose actual onset was day 10 for PWH and day 28 for Amoy) from estimated individual reproduction numbers using only the temporal data in the non-spatial method, because individuals eligible to be infectors on a particular day had the same reproduction number for that date.

5.4.5.2 Variable infectious period

Using the observed hospital discharge date as the upper bound of duration of infectiousness resulted in cases with shorter infectious period to have lower individual *R(i)* (Figure 5.7), but other eligible potential infectors of the same secondary case would correspondingly be attributed a higher individual reproduction number. The overall impact on *R*(*t*) was more significant at the beginning and the end of the epidemic when the case incidence was small, and fluctuated in terms of direction (e.g. higher *R(t)* on day 10 to 13, and lower in day 1, 4 and 7). The difference in *R(i)* and R(t) compared to the original method were mostly of minimal magnitude (25th and 75th percentile, and the maximum: *R(i)*=(0.03,0.06,-1.8), *R(t)*=(-0.02,0.05,-0.8)). The variable infectious period estimated some *R(i)* to be close to 0, reflecting the reality that some cases to have low chance of generating any secondary transmission. The number of potential infectors per onset day reduced on average by 25.4 (SD 20.3, range 0-86) compared to the original method.

Figure 5.7 (a) Comparison of *R(i)* and *R(t)* estimates under the fixed and variable infectious period assumption, using the Wallinga and Teunis method with the serial interval distribution.



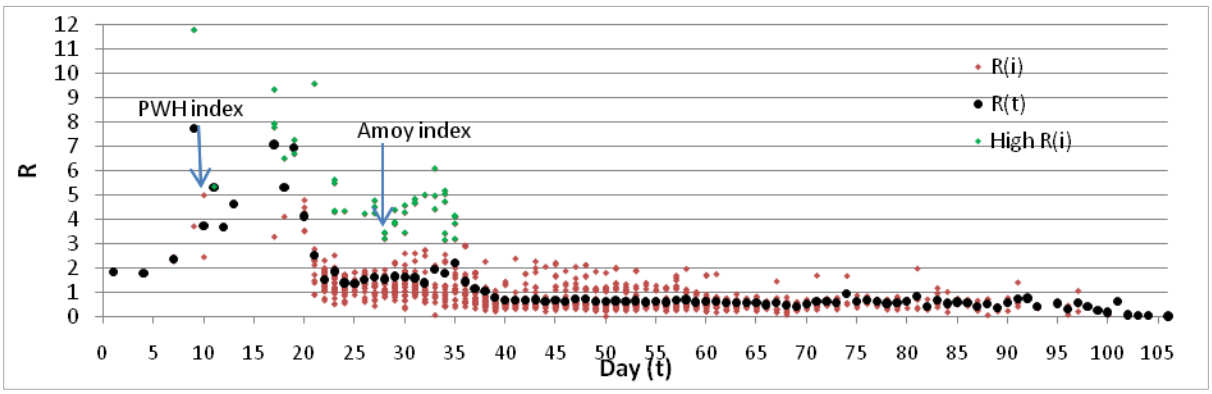
5.4.5.3 Spatial mixing

Figure 5.8 depicts *R(t)* and *R(i)* estimated under different movement assumptions using the spatially extended Wallinga and Teunis method with gravity models fitted to the adjusted census travel data, all with 2-day convolution and variable infectious period. Results are compared with those from the non-spatial model. Defining high *R(i)* as those >5 up to day 20, and >3 afterwards, the susceptible driven model [Figure 5.8(a)] had more cases with high *R(i)* (2.8%) than the infector driven (2.1%), bilateral (1.7%), multilateral (0.5%) and distance only (2.1%) models.

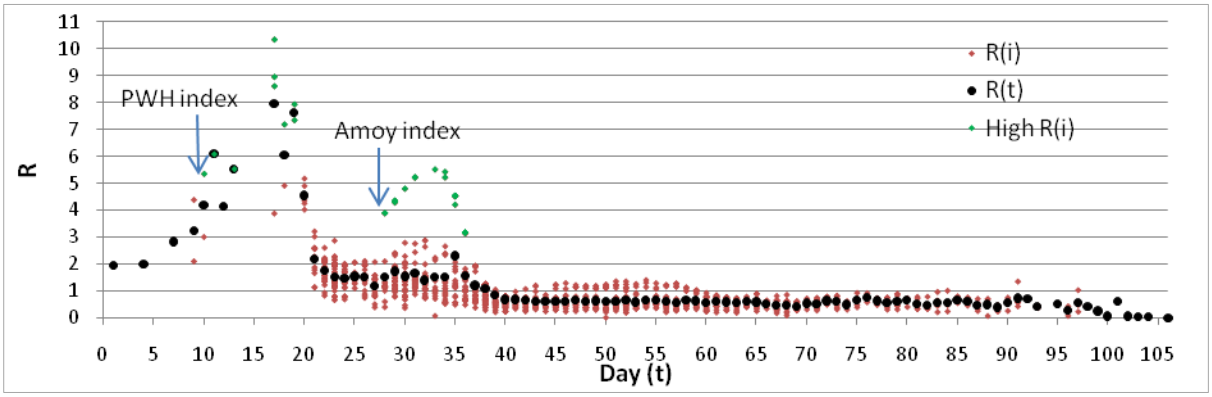
Notably, 48.0% of high *R(i)* cases in the susceptible model resided in the high inflow districts, and none in all other models. This illustrated a tendency of susceptible-driven models to over-represent cases from high inflow districts as having high *R(i).* In reality only 4.4% of all SARS cases resided in high inflow districts, while the model estimated that 92.3% of such cases had high *R(i)*. Similar results were found for the models using the observed trip patterns instead of a gravity model, ruling out the

Figure 5.8 The estimated daily and individual reproduction numbers *R(t)* and *R(i)* under different movement assumptions, using the partially extended method with 2-day convolution and variable infectious period and on the adjusted census overall trip patterns. High *R(i)* values are highlighted in green, and the *R(i)* of the index case of the PWH and Amoy Gardens outbreak are identified.

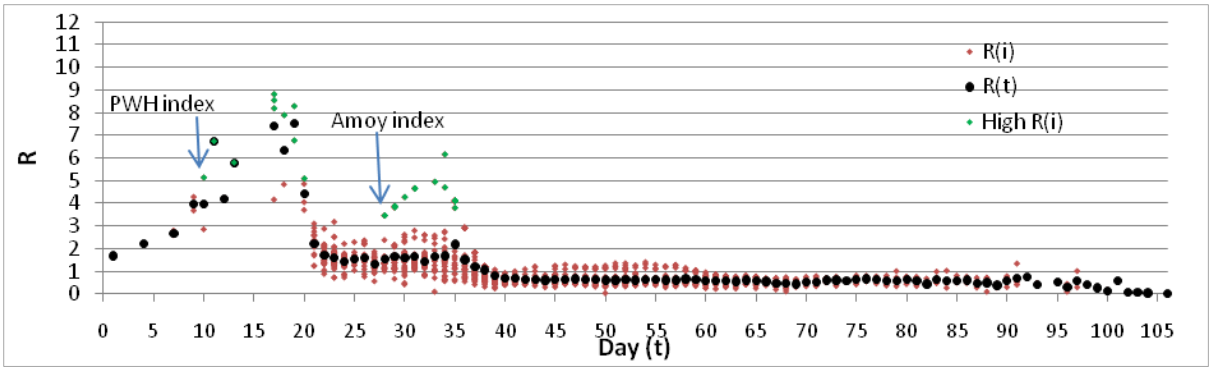
(a) Susceptible driven model



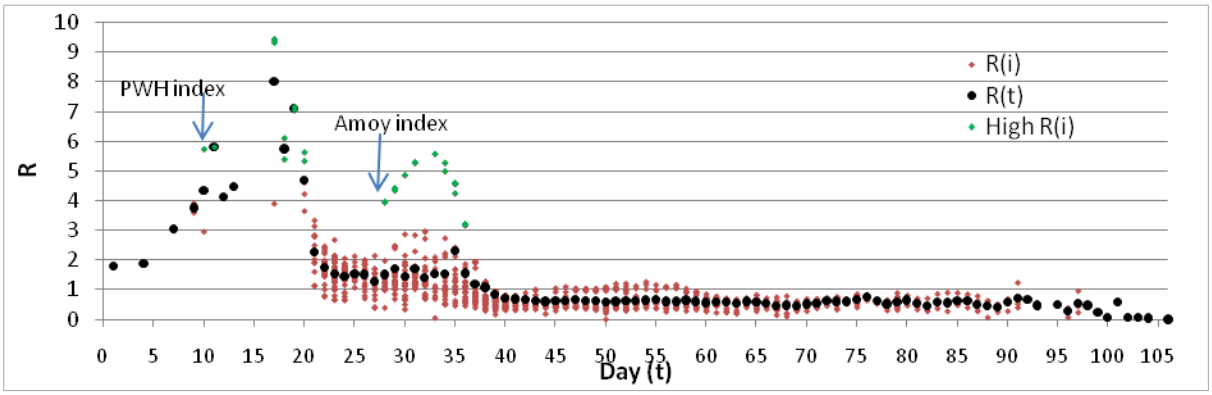
(b) Infector driven model



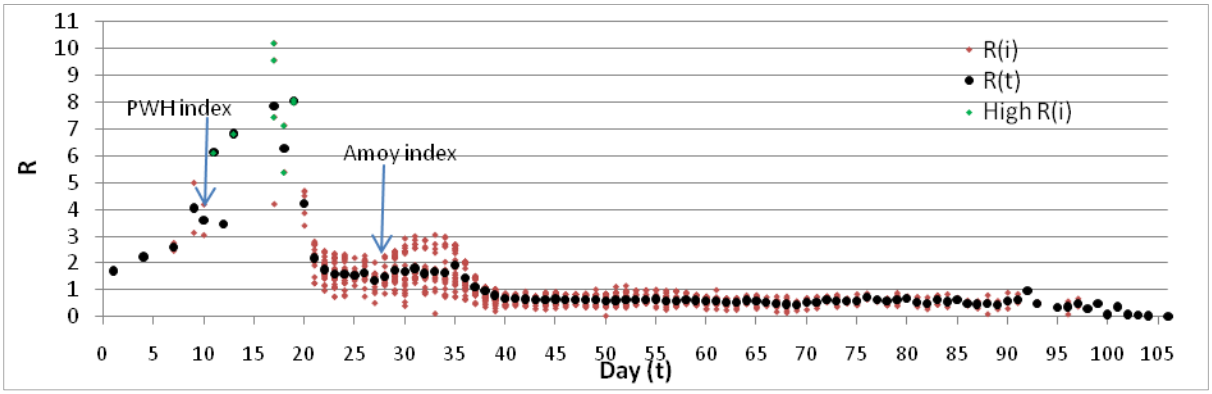
(c) Bilateral model



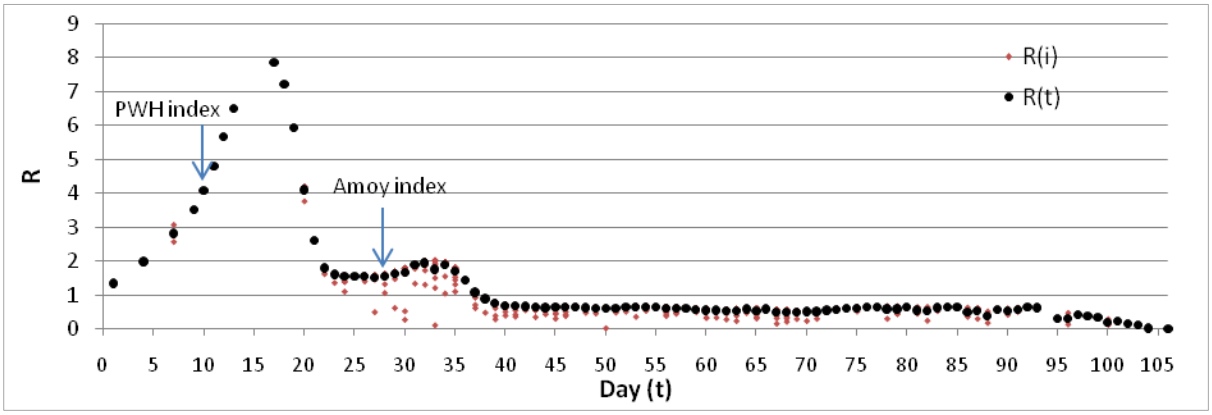
(d) Distance only model (based on the spatial kernel derived from inflow constrained gravity model with α_1=1 and α_2=0)



(e) Multilateral model



(f) Non-spatial model (with 2-day convolution and variable infectious period)



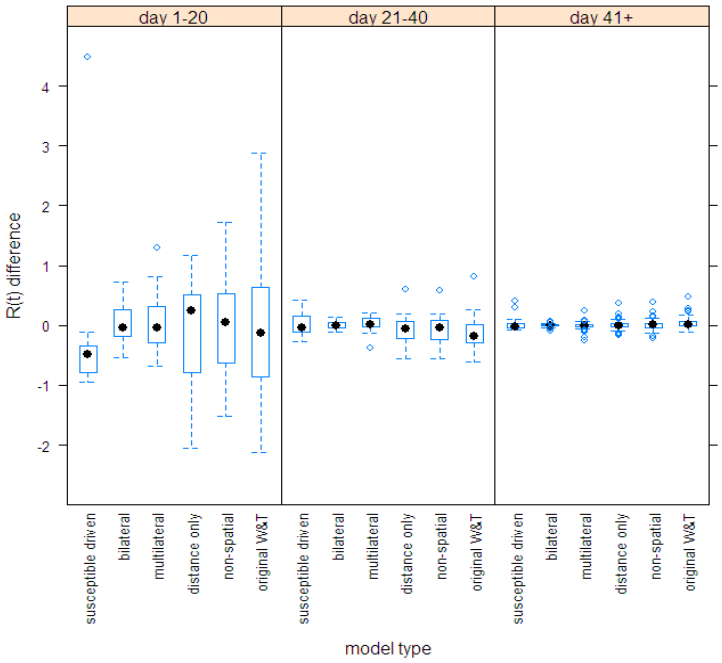
effect being due to poor fitting of the gravity model. This bias tended to prevent the model from identifying the real index cases of the PWH and Amoy outbreaks (e.g. *R(i)*=11.7 for a case onset at day 5, and those high *R(i)* cases on day 22-27), making the model not desirable.

The infector driven, bilateral and distance only models gave similar results to each other [Figure 5.8(b-d)], and all estimated the true index case of the PWH outbreak to have a higher *R(i)* [5.4, 5.0 and 5.7 respectively], compared with the 4.1 obtained from the non-spatial model [Figure 5.8(f)]. Peak *R(i)* for the PWH outbreak was 10.4 (infector driven), 8.8 (bilateral), 9.4 (distance only) and 7.9 (non-spatial) at day 17. All three spatial models better identified the Amoy index case, estimating its *R(i)* as 3.9 (infector driven and distance only) and 3.4 (bilateral), while for the non-spatial model the Amoy index only had an *R(i)* of 1.6 - indistinguishable from many other cases. Peak Amoy *R(i)* reached 5.5 in the infector driven model and 5.6 in the distance only model at day 33, but in the bilateral model the peak *R(i)* of 6.1 at day 44 was from a non-Amoy infected case who lived in the district where the Amoy Gardens is located. From day 41 onwards, the infector driven model estimated a slighter wider range of *R(i)* (including values close to 0) than the bilateral and distance only models, suggesting the infector driven model was able to reproduce more heterogeneity than the other two models. The distance only model gave the least heterogeneity, as expected. All three models estimated two cases at day 91 and 97 to have *R(i)* around or higher than 1 – in reality these were related to the hospital outbreak in NDH. In general, the infector driven model slightly outperformed the bilateral and distance only model, though all of them generated satisfactory and sensible results.

The multilateral model [Figure 5.8 (e)] by allowing the cases to meet in all districts reduced the level of spatial heterogeneity overall, and thus reduced the variation of *R(i)* - most notably during the Amoy outbreak where the maximum *R(i)* was only 3.0. I therefore considered it as inferior to the infector driven, bilateral and distance only models.

The differences in *R(t)* estimates between the infector driven model results and those from other spatial mixing models (Figure 5.9) were greatest at the initial period (day 1-20) when the cumulative number of cases was small (24 cases) and at the beginning of the PWH outbreak. From day 41 onwards when the epidemic size was above 700, the difference was small. The maximum magnitude of *R(t)* difference was 2.9 with the original method, 1.7 with the non-spatial model, 2.0 with the distance only model, 4.5 with the susceptible driven model, and 0.72 with the bilateral model, though the surge in *R(t)* for the peak of Amoy outbreak at day 35 was less remarkable in the bilateral model than the infector driven model [Figure 5.8 (b,c)].

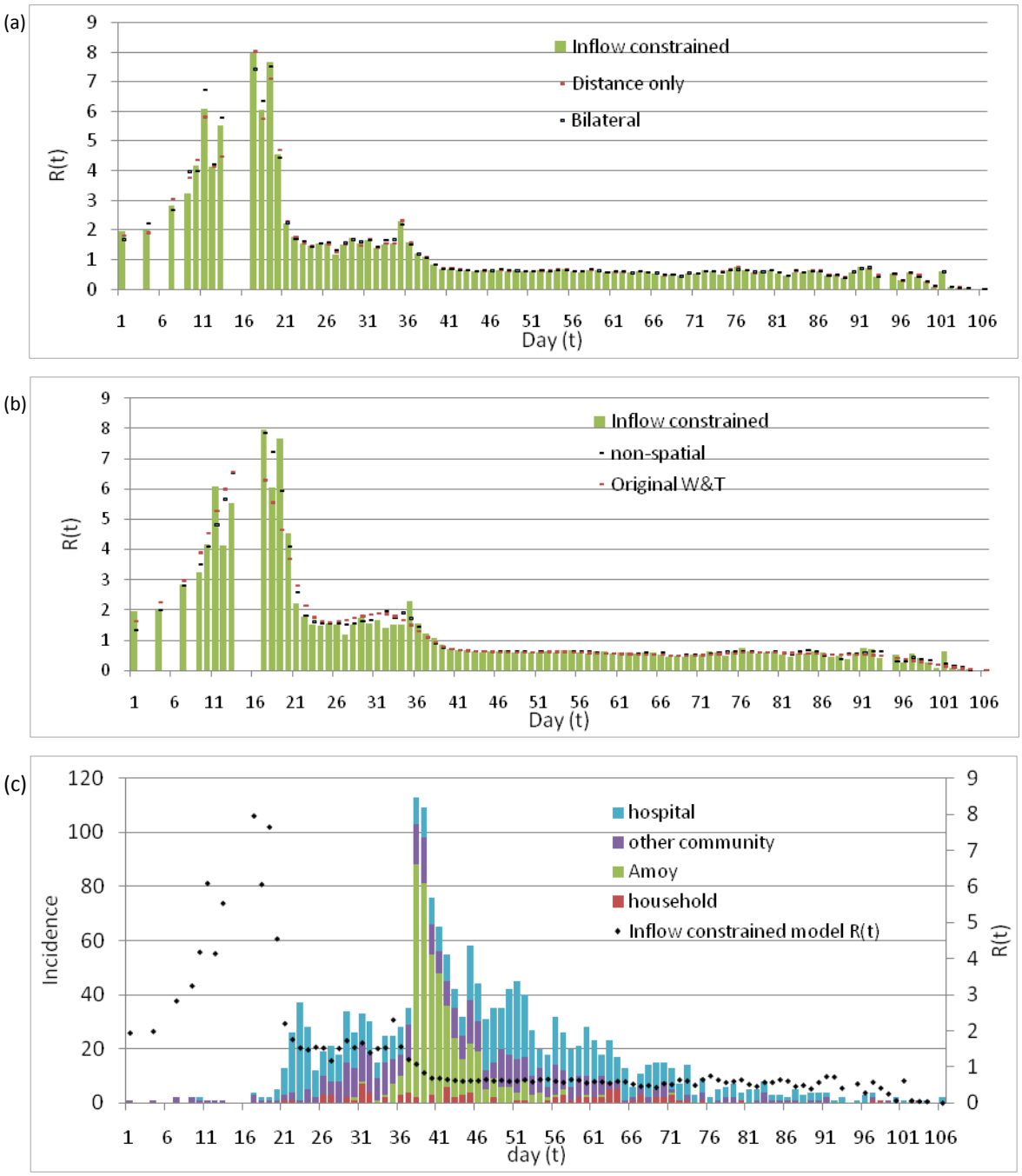
Figure 5.9 Box-plots on the difference of R(t) [model of other spatial mixing assumptions - infector driven model] by 3 stages of the epidemic. The dots represent the median, the top and bottom of the boxes are the 75th and 25th percentile, and the upper and lower cap are the 75th and 25th percentile $\pm$ 1.5 interquartile range, and beyond those are outliers and displayed as circle.



Greater fluctuations in *R(t)* towards the end of the epidemic were predicted in all spatial models, attributed to small case incidence and some local outbreaks [Figure 5.8 (a-e)]. Figure 5.10 shows a clearer comparison of *R(t)* over time. The general shape of the R(t) time-series estimates were very similar among the inflow constrained, bilateral and distance only model [Figure 5.10(a)], and the original method and non-spatial model were over-smoothed [Figure 5.10(b)], especially when compared to the daily incidence of the epidemic [Figure 5.10(c)].

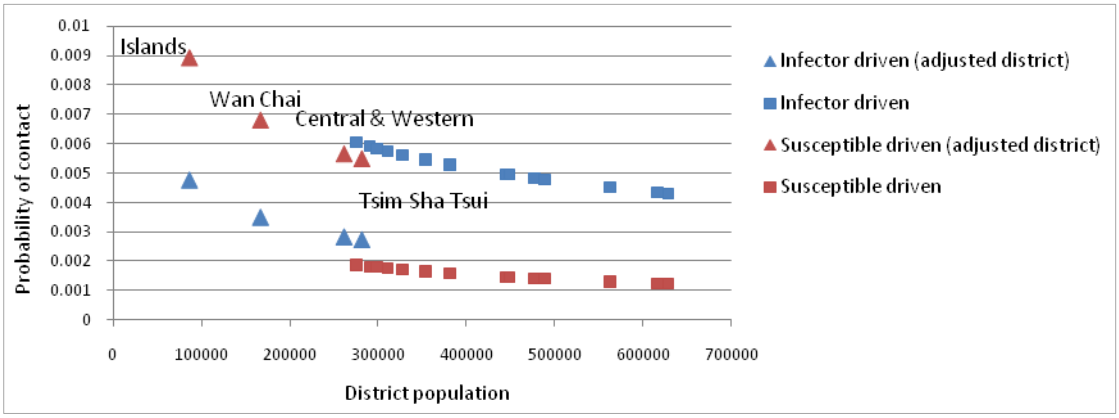
Investigating further the difference between the unilateral movement models, the adjustment exponent of the high inflow districts in the gravity model β_2=0.09 in the susceptible driven model

Figure 5.10 The R(t) time series of the inflow constrained model (a) compared to that of the bilateral and distance only model; (b) compared to that of the non-spatial model and the original Wallinga and Teunis method; (c) plotted along the case incidence of the outbreak by source of infection.



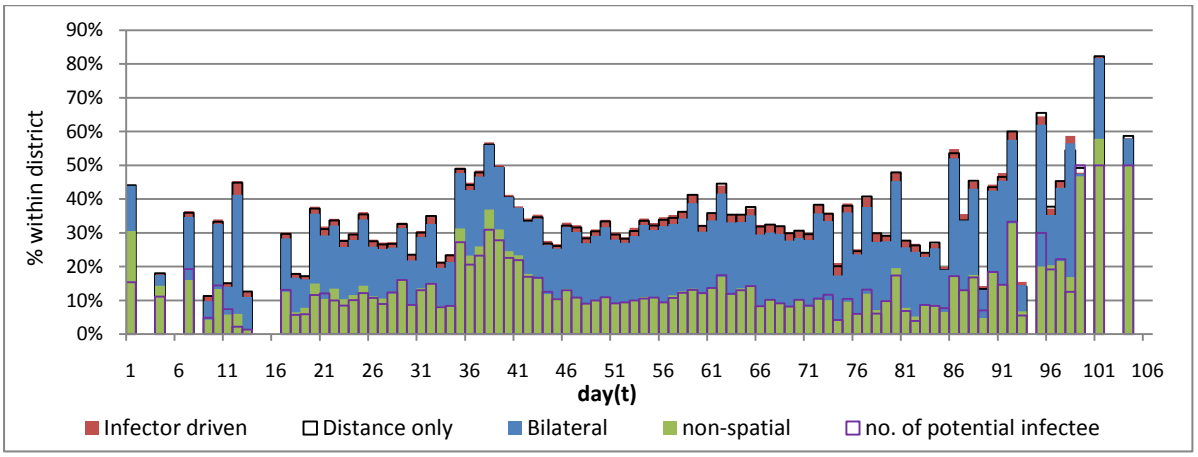
was found to have boosted the probability of contact with a specific individual in these districts (Figure 5.11), while the negative exponent α_2=-0.06 in the infector-driven model correspondingly reduced the probability of contact with a specific individual from these districts, because the inflow constrained gravity model showed that these districts were weak donors of between-district journeys. These districts accounted for 12% of the Hong Kong population, but only for 4.4% of the total SARS cases. Therefore inflow constrained models are more appropriate for representing SARS transmission dynamics.

Figure 5.11 The probability of contact with a specific individual by the district population under the susceptible and infector driven model, with those of the high inflow districts were labelled.



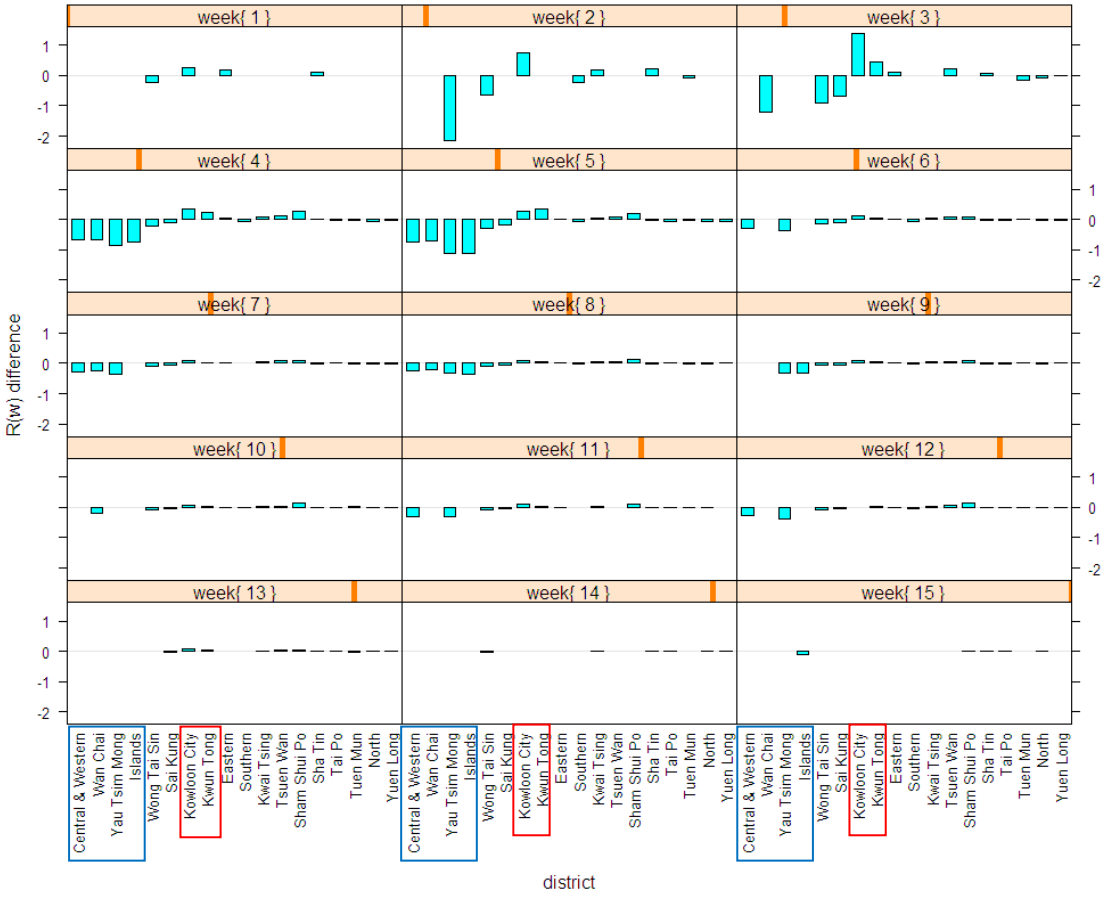
The infector driven model was gave very similar results to the distance-only model in capturing the proportion of within district transmission, though with the population effect in the spatial component (Figure 5.12). The overall proportion of within district transmission was 38.6% under the infector driven model, 38.2% in the distance only model and 35.9% in the bilateral model, all of which were more than double that estimated using the non-spatial model (15.2%), which – as expected - only estimated roughly the same proportion of the number of potential secondary cases in the same district as the index case (15.5%) identified by the incubation period assumption and given the observed set of incident case data. For cases which were associated with the Amoy outbreak the average proportion estimated to have been infected within their residential district was 56.1% for the infector driven model, 53.6% for the bilateral model, and 56.9% for distance only model. However, these figures included hospital transmission by the Amoy cases. We would expect more realistic results the fully extended method in which the hospital transmission was accounted for explicitly (see the next chapter).

Figure 5.12 Percentage of within district transmission in the R(t) of the inflow constrained, bilateral and non-spatial model, and in the number of potential secondary cases estimated by the modified method.



On studying the reproduction number by district (Figure 5.13), the weekly average reproduction number *R(w)* of the high inflow districts were generally higher for the bilateral model than the inflow constrained model, at the expense of lower *R(w)* in a few districts such as Kowloon City and Kwun Tong (where Amoy Gardens located) at week 2 to 5. This could be attributed to the spatial

Figure 5.13 Difference in estimated district-specific weekly average reproduction number R(w) [infector-driven model - bilateral models]. Blue box: high inflow districts; red box: districts with highest positive *R(w)* difference. Amoy Gardens is located in Kwun Tong district



component in the bilateral model which equally weighted the susceptive driven and infector driven type of movement.

Taking the above findings into consideration, I considered the infector driven model to be slightly better than the bilateral model, and thus adopted it as the baseline model. The bilateral model was used for sensitivity analysis to test the robustness of the results.

5.4.5.4 Sensitivity to travel data used

Use of the adjusted census overall trip data enabled the partially modified method to better capture the localised super-spreading events, with higher *R(t)* [Figure 5.14] and *R(i)* estimated for those events. Using the infector driven model, *R(i)* of the index of the PWH and Amoy outbreak under the work-study data based gravity model (see previous chapter) was 0.3 and 0.7 lower than that obtained using the adjusted census overall model. For the work trip data based model these

Figure 5.14 The daily reproduction number of inflow constrained model based on gravity models of different travel trip patterns.

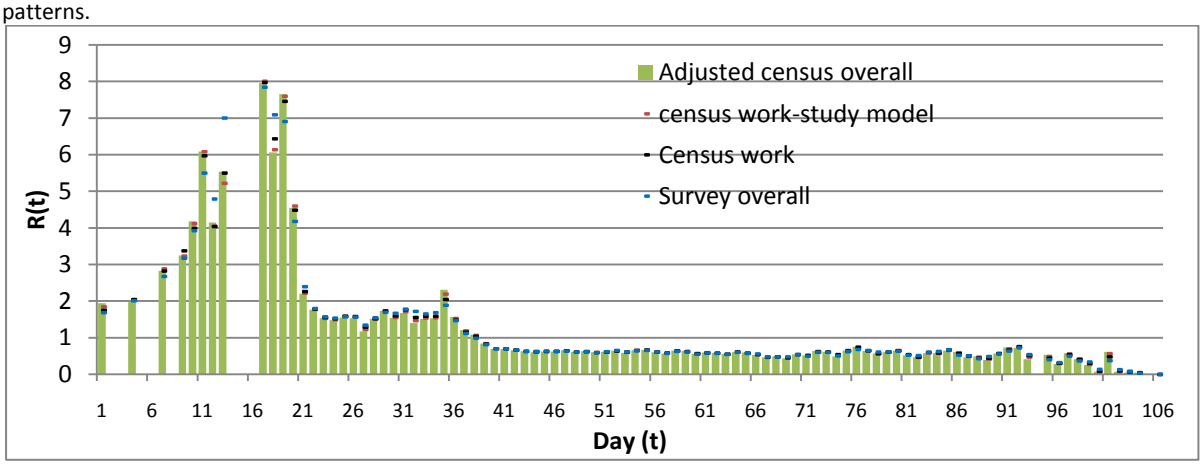
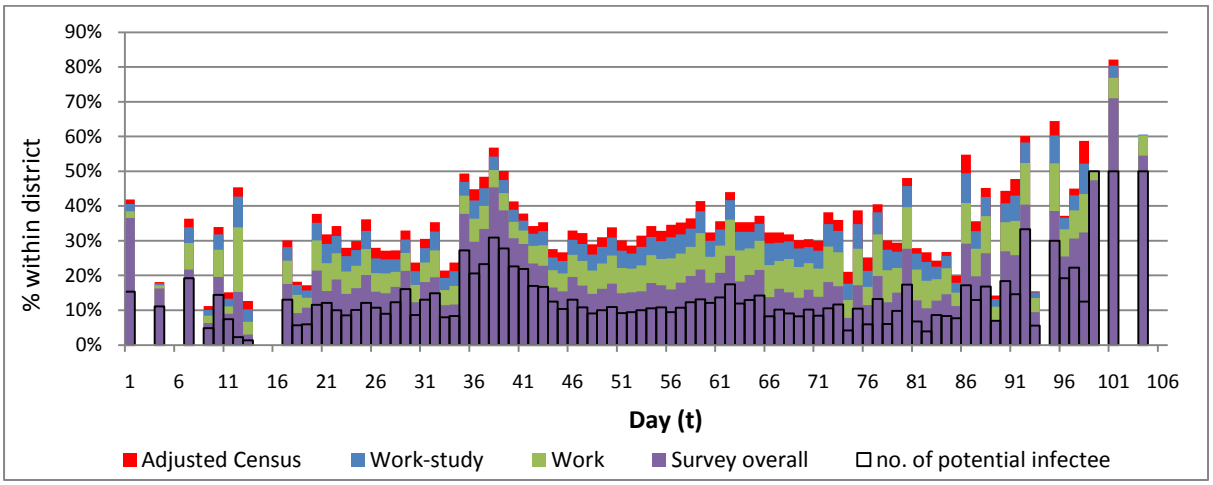


Figure 5.15 Percentage of within district transmission in *R(t)* and in the number of potential secondary cases under different trip patterns used in the gravity model for the inflow constrained method.



estimates were 0.8 and 1.0 lower, and for the gravity model based on the overall survey data, the estimates were 1.4 and 1.6 lower. The same descending order of performance was found in the percentage of within district transmission estimated [Figure 5.15]. Therefore I chose to use the gravity models fitted to the adjusted census trip data for subsequent analyses.

5.4.5.5 Spatiotemporal reproduction number and within district transmission

Based on the infector driven mixing and the adjusted census overall gravity model in the partially modified method, the weekly average reproduction number *R(w)* increased from 2.3 in week 1 to 5.7 in week 3, capturing the effect of the PWH outbreak well (Figure 5.16). The second peak in *R(w)* of 1.7 at week 5 captured the impact of the Amoy Gardens outbreak. From week 6 onwards *R(w)* was below 1. Excluding the 447 and 123 cases from the Amoy and PWH clusters, the initial *R(w)* was estimated as 2.08, and *R(w)* fell below 1 from week 4.

Figure 5.16 The estimated weekly average reproduction number by the partially modified model, and the weekly incidence by source of infection.

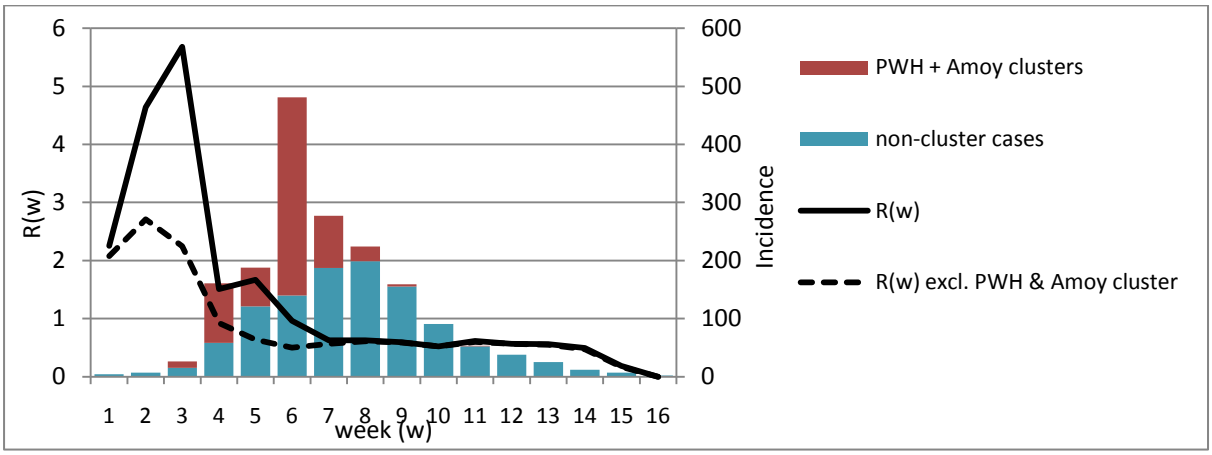


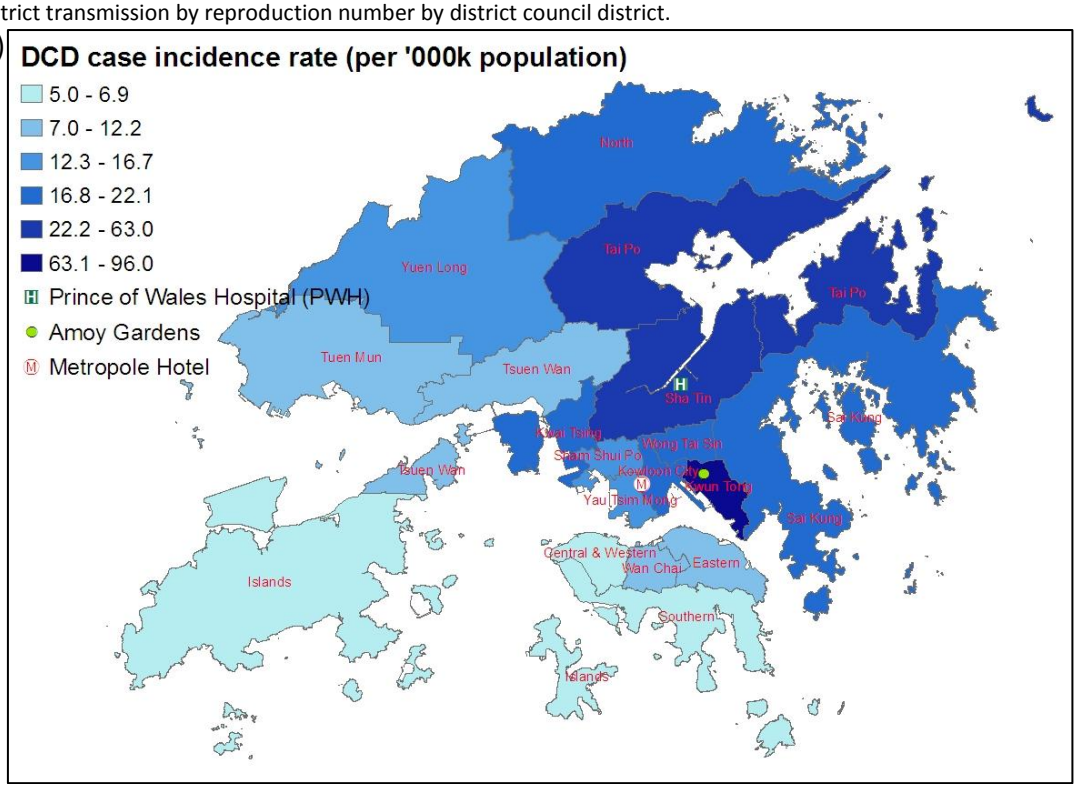
Figure 5.17 (a-c) shows by district the overall incidence rates, average *R(i)* of cases residing in the district, and the overall proportion of within-district transmission by cases living in that district. Figure 5.18 shows district-specific *R(w)* estimates, and Figure 5.19 the weekly percentage of within-district transmission. The four high inflow districts (Wan Chai, Central and Western, Islands and Tsim Sha Tsui) had relatively low *R(i)*, within district transmission and incidence rate. Districts with high *R(i)* (Kowloon City, Wong Tai Sin, Kwun Tong) were those infected in the first week of the epidemic (Figure 5.18) and had high *R(i)* in the next 2 weeks, though the high *R(w)* in Kwun Tong was due to the Amoy outbreak. Districts with high within-district transmission included Sha Tin (where PWH located), Kwun Tong, and Tai Po Figure 5.17 (c), all identified as important spatiotemporal clusters by the Kulldorff's scan statistic in chapter 3.

The first three weeks showed a rapid spread of the epidemic to other districts resulting with moderately high *R(w)* estimates (Figure 5.18) and great variation in the proportion of within-district transmission (Figure 5.19), but in the Sha Tin district the within-district transmission remained high because of the PWH outbreak. At week 4 and 5, all districts were infected but *R(w)* estimates were mostly lower than the previous weeks, with the Kwun Tong district having the highest *R(w)* and

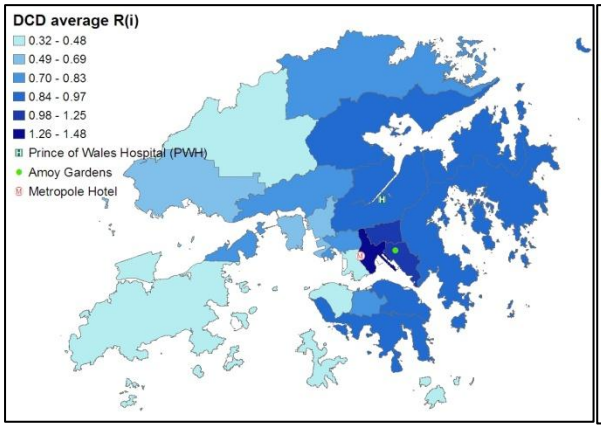
proportion of within district transmission, as early cases of the Amoy cluster had symptom onset during this period and the peak incidence was reached at week 6. The Tai Po district then became the key area of within district transmission till week 9, and from this time onwards till the end of the epidemic the *R(w)* of all districts were <1, during which the North district was alone in sustaining any significant local transmission.

Figure 5.17 The spatial decomposition of (a) The overall case incidence rate, (b) average *R(i)*, and (c) percentage of within district transmission by reproduction number by district council district.

(a)



(b)



(c)

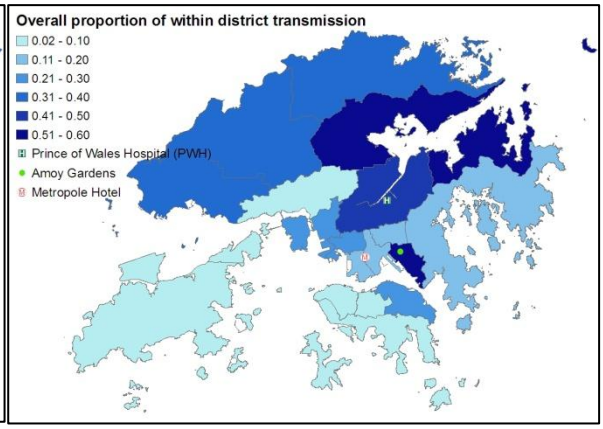


Figure 5.18 The spatiotemporal weekly reproduction number *R(w)* by district, and the weekly incidence of the hospital and Amoy outbreaks (from the source of infection data).

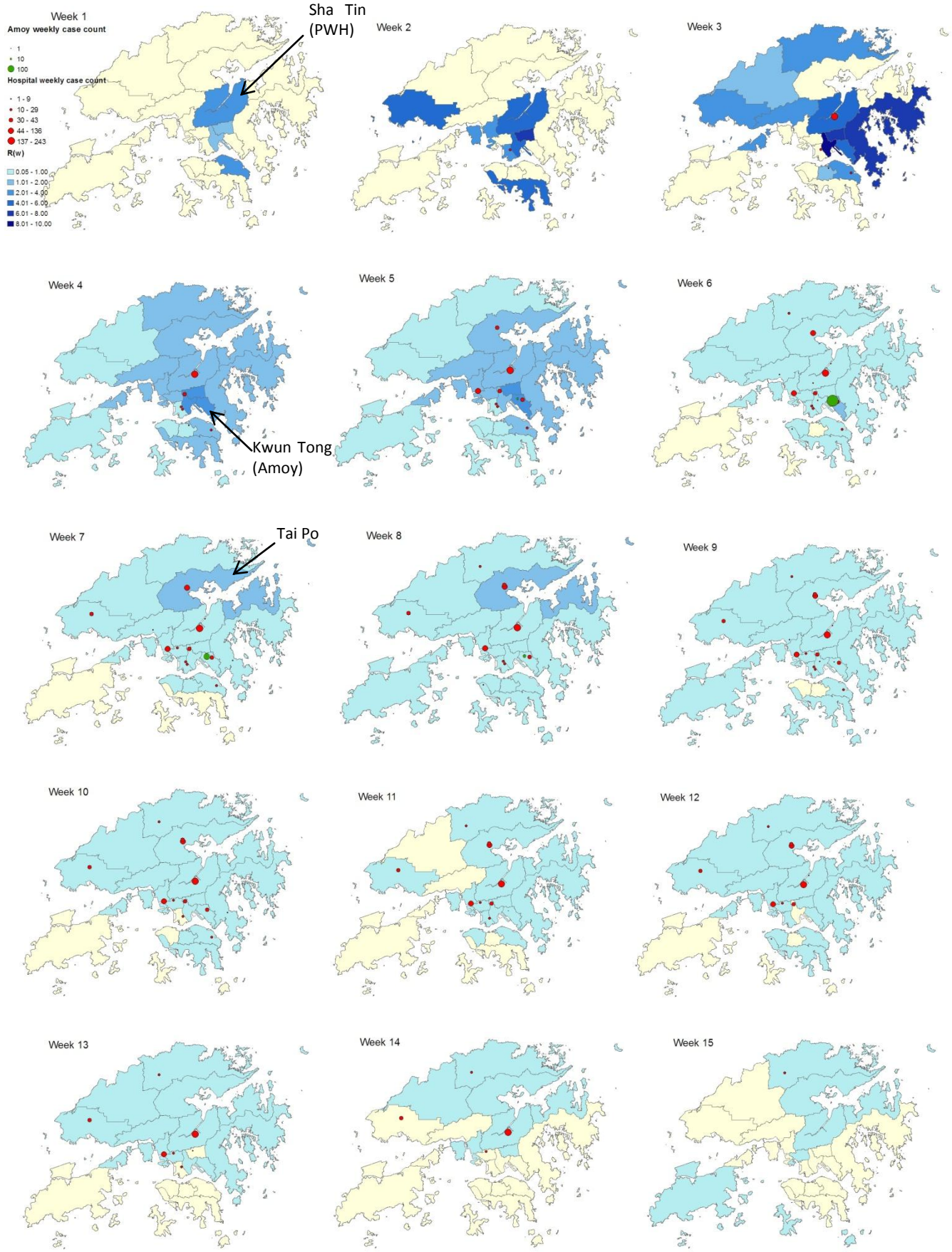
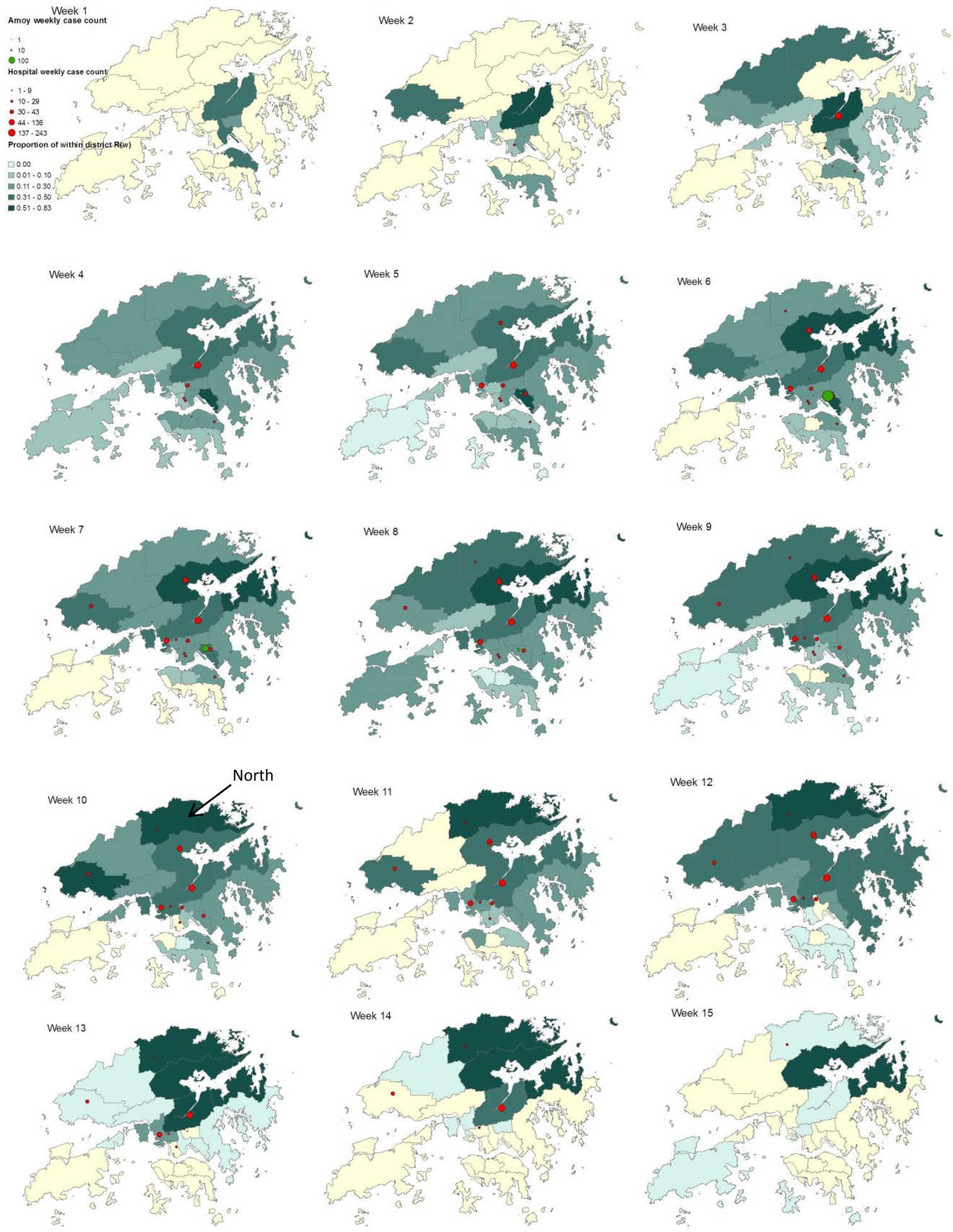


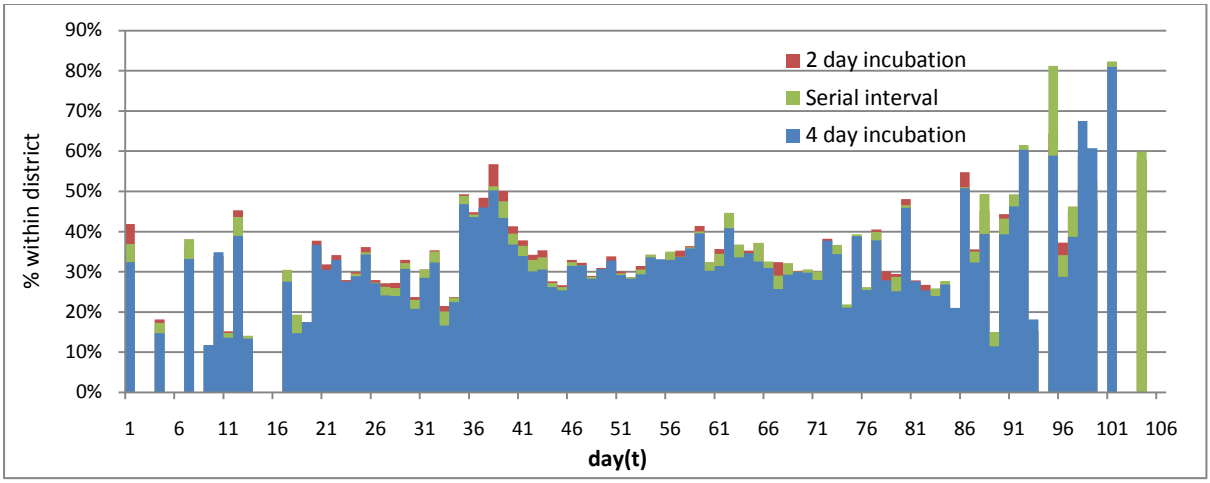
Figure 5.19 The spatiotemporal within district transmission percentage in *R(w)* by district, and the weekly incidence of the hospital and Amoy outbreaks (from the source of infection data).



5.4.5.6 Sensitivity analysis

The 2-day convolution serial interval model generally gave the highest proportion of within-district transmission compared to the 4-day convolution and Weibull serial interval assumptions (Figure 5.20). This could be due to the tendency of household transmission to have shorter serial intervals (Fine 2003), while the 4-day convolution excluded the potential infectors whose onset was less than 3 day before the index cases.

Figure 5.20 Comparison of the percentage of within district transmission in the reproduction number under the assumptions of 2-day, 4-day and Weibull convolution.



As noted previously [Figure 5.2(c)], the non-spatial 4-day convolution showed a shift of the peak R(t) compared to the 2-day convolution model (Figure 5.6). A similar effect was seen for the spatially extended method [Figure 5.21(a)]. Also the higher peak of the 4-day convolution distribution resulted in higher *R(i)* than the 2-day incubation model for the PWH and Amoy index case, by magnitude of 2.9 and 1 respectively. The difference in *R(t)* and *R(i)* between the 2-day convolution and the Weibull serial interval distribution models [Figure 5.21 (b)] were smaller but similar in nature to those beween the 2-day and 4-day convolution models. In general, the *R(i)* and *R(t)* estimates were fairly robust from day 40 onwards under all assumptions on the timeline of the course of infection (Figure 5.22).

Using the residing district centroid rather the residential location of the cases as the spatial reference, the model performance was satisfactory in capturing the effects of spatial heterogeneity, though the Amoy index cases was less easily distinguished by the *R(i)* because there were some cases with increasing *R(i)* from day 24 to 27 [Figure 5.22(a)]. The *R(i)* of the PWH and Amoy index cases were 5.7 and 3.3 respectively. A higher percentage (2.4%) of the cases had high *R(i)* estimates, but none from the high inflow districts. A higher proportion of within district transmission also resulted from the centroid model. This is because for the model which used residential locations,

Figure 5.21 The difference in daily and individual reproduction nubmers *R(t)* and *R(i)* using the 2-day convolution as baseline compared to: (a) the 4-day convolution model; (b) the Weibull serial interval distribution model.

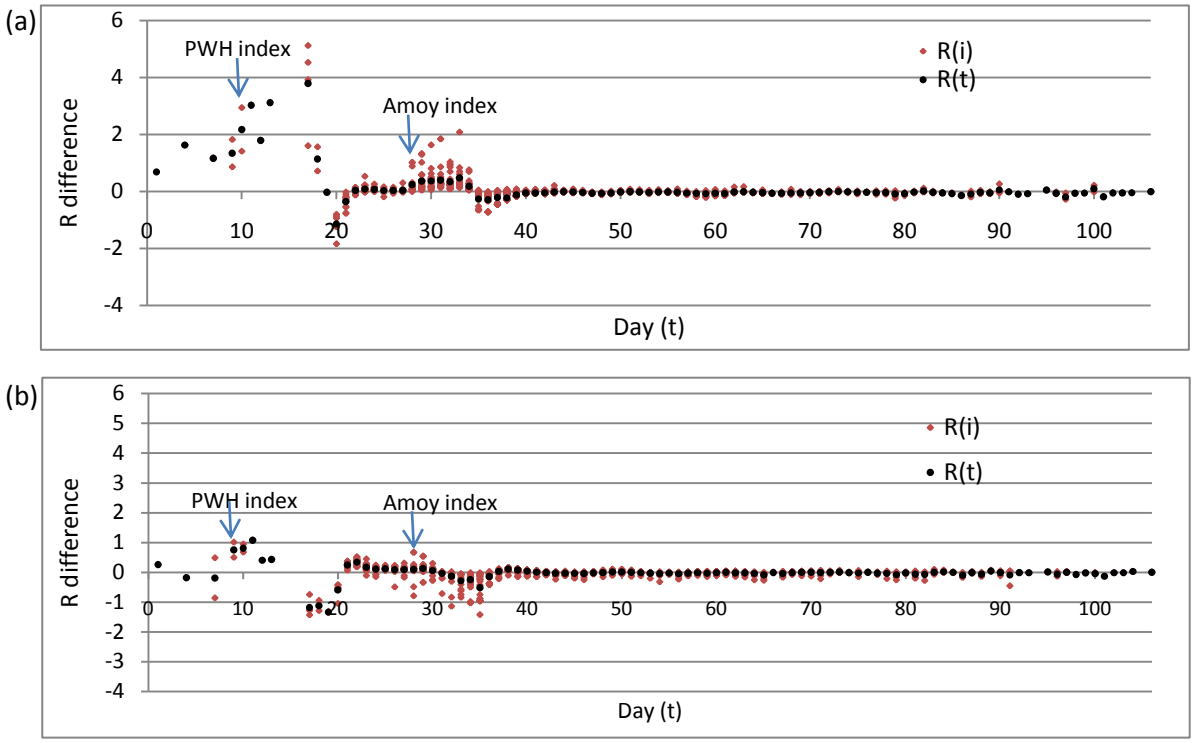
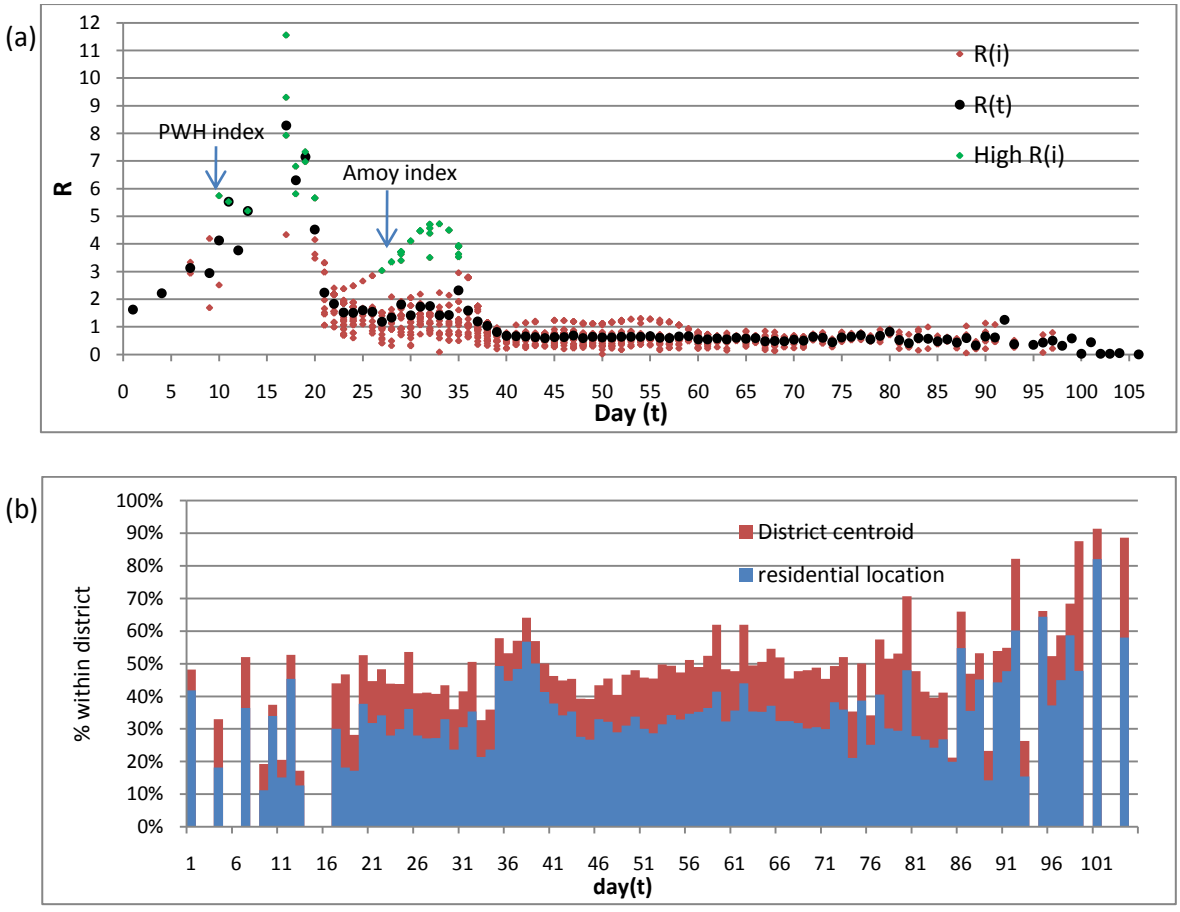


Figure 5.22 Results of the partially modified method using the centroid of the residing district of the cases: (a) the daily and individual reproduction number estimates, (b) proportion of within district transmission.



zero distance only occurred for cases in the same household or building, while when centroid locations were used, all cases residing in the same district had zero spatial separation.

5.4.6 Discussion

In this chapter, I derived a spatially extended Wallinga and Teunis method that used a constant incubation period to infer the date of infection and identify the potential infectors, a Weibull distributed profile of infectiousness over time (fitted to serial interval data), and a spatially-dependent probability of contact derived from gravity models with different spatial mixing assumptions and underlying travel data. I found each of these modifications improved the ability of the model to capture the spatiotemporal heterogeneity of the epidemic, leading to more realistic estimates of the reproduction number at the daily and individual levels.

The modified method identified the index cases of the PWH and Amoy outbreak with a rapid increase in individual reproduction numbers, and more accurately estimated the true extent of these super-spreading events compared with the original method. However, methods of probabilistic infection network inference generally identify the super-spreading events – where cases tended to be close in symptom onset and spatial locations, as high individual reproduction numbers for a small fraction of cases – especially in the early generation of the cluster, rather than as a very high reproduction number of a single case. This also applies to transmission through environmental factors, for instance in the Amoy outbreak.

Using the average reproduction number of the first week of the epidemic to approximate the basic reproduction number, R_0 was estimated at 2.1 with the PWH and Amoy clusters excluded. This is slightly lower than estimates of 2.7 [95% CI: (2.2, 3.7)] estimated by a spatially explicit metapopulation model of the Hong Kong epidemic (Riley, Fraser et al. 2003). This difference may partly be attributed to 100 cases which were only linked to the Amoy cluster after the Riley's study, but in addition Riley's study defined the stage of the epidemic coarsely, and the basic reproduction number estimate corresponded to the first 33 days of the epidemic.

Previous epidemic models (Xia, Bjornstad et al. 2004; Viboud, Bjornstad et al. 2006; Merler and Ajelli 2010) only used the gravity model to reflect the relative contact rate between patches, and the weightings of the within-patch and between-patch contact rates were estimated from the spatiotemporal incidence data through the model. In the modified method derived in this study, gravity models were used for both within and between patches coupling. Travel data that were adjusted to include those who did not travel as remaining in their home patch proved crucial to better capture the movement and contact between individuals in the same district or household,

though closer and longer contact made in the household setting specifically was not explicitly accounted for.

Because trip patterns described by the gravity models are directional, the use of a gravity model to represent contacts in epidemic models requires assumptions on who travels and where individuals meet. Comparing the different assumptions I explored, the infector driven model that assumed the infector travels to the susceptible's patch best reproduced the spatiotemporal spread of the SARS outbreak in Hong Kong. This was largely attributed to the inflow-constrained gravity model it used better capturing the relationship between travel patterns and infectious contact, in particular the correct weighting of distance effect, the population size effect and high inflow districts effect, and how all these were subsequently transformed into within-district and between district contact rates. Key commercial and leisure districts in Hong Kong attract high numbers of visitors from other districts, but the incidence rates in these districts were mostly low. Districts with high incidence rate were those with local outbreaks occurred. This suggested that distance was a stronger driver than the population size and the high inflow districts in the disease spread. The susceptible driven model that used the outflow constrained gravity model had contact rates strongly driven by the high inflow districts, and therefore underestimate the probability of within district transmission. Conversely, infector driven model that used the inflow constrained gravity model with a negative scaling exponent on the high inflow districts discounted the probability of contact with these districts, matching the observed spatial distribution of cases. The bilateral model as a balance of both approaches produced similar results to the infector driven model.

Compared with other models assuming bilateral movements (Thomas 1999), the method developed here allowed the contact probabilities to be scaled with the population size, and scaling effects were allowed to varied for districts with a higher flow as they were the key areas of commercial and leisure activities. Compared with epidemic models that use observed leaving and return rates from and to the patches to capture heterogeneous contact rates (Balcan, Colizza et al. 2009), my approach is somewhat more parsimonious, yet produced substantial spatial and individual-level heterogeneity. As the epidemic grows, the effects of super-spreading events or local outbreaks are averaged out in estimates of the overall population reproduction number, making such events less detectable. Spatially-stratified and individual reproduction number estimates allow such heterogeneity to be revealed.

The spatially extended Wallinga and Teunis method presented here was shown to give satisfactory results even when only spatial data of coarse resolution were available, which is often the case during epidemics. However, the findings on the most appropriate spatial mixing were unique to the

current model structure and the characteristics of the Hong Kong population, therefore maybe not be directly applicable to other studies that incorporate gravity model into metapopulation epidemic models.

Chapter 6 - Quantifying Hospital and Community Transmissions

In this chapter, I study the dynamics of hospital and community transmission of the SARS outbreak in Hong Kong by further modifying the Wallinga and Teunis method to incorporate finer spatial details and to utilise the hospitalisation and source of infection data. A review on the current findings on the role of hospital and community transmission in the SARS outbreak is given below.

6.1 Current findings

Nosocomial transmission of SARS occurred in many countries (World Health Organisation (WHO) 2004). Therefore it is important to capture the transmission dynamics within and between the hospital and the community in epidemic models to give credible results. Nosocomial transmission was considered in one modelling study of SARS transmission in Hong Kong (Riley, Fraser et al. 2003), and two other studies have attempted to quantify the degree of transmission that occurred in the hospital and community settings (Lloyd-Smith, Galvani et al. 2003; Cori, Boelle et al. 2009).

Cori *et al*. (Cori, Boelle et al. 2009) fitted a discrete time stochastic SEIHR model, using the hospitalisation data, to the Hong Kong SARS outbreak. Individuals were categorised as inpatients (those infected whilst hospitalised for other health conditions), healthcare workers and the general population. The date of infection was inferred assuming constant infectiousness between 1 day before to 4 days after symptom onset and a latent period of 5 days. The infectious period was estimated to be 9.3 days. Super-spreading events in the Amoy Gardens and PWH were captured as an additional contribution to the force of infection in the general population and hospitalised patients/health care workers respectively. The start and end dates of these super-spreading events and their size were estimated.

Cori *et al.* found that the estimated effective contact rate over time was a few times higher in the community than in the hospital, and that since late March (roughly day 40 of the epidemic) the hospital effective contact rate had been very close to 0. The average reproduction number (referenced by infection time) of the community infected cases was generally higher than that of the healthcare workers, and inpatients apparently had average reproduction number lower than 0.25 throughout. Under the model, the PWH outbreak corresponded to high reproduction numbers from healthcare worker to community, and community to community, whereas the reproduction numbers for transmission among and between healthcare workers and inpatients did not change much during this period compared to the rest of the epidemic. This is in contrast to another study on the PWH outbreak which found an important contribution of healthcare workers in this super-spreading event (Kwok, Leung et al. 2007).

Lloyd-Smith *et al*. (Lloyd-Smith, Galvani et al. 2003) studied the effect of various infection control measures in reducing transmission within and between the community and hospitals and the scale of the epidemic. They used a discrete time stochastic compartmental model and applied the baseline transmission rate that yielded the basic reproductive number estimates of SARS from previous findings on a hypothetical community with a hospital. Within-hospital transmission was identified as the most important transmission setting in which control measures, especially those targeting healthcare workers, are most effective. Transmission between hospital and community was also found to be important as it had the potential to trigger further transmission within the community. Reduction of contact between off-duty healthcare workers and the community was recommended especially when hospital control measures are not adequate to curtail the outbreak.

A key plausible explanation for the low contribution of healthcare workers in hospital infection in Cori *et al.*'s study compared to the other study is the implicit assumption that infectious healthcare workers, prior to hospital admission for treating SARS, could only infect other healthcare workers in the community and not in the hospital. Therefore the hospital effective contact rate was likely to be underestimated, and hence also the reproduction numbers among and between healthcare workers and inpatients. Lloyd-Smith *et al*.'s study allowed transmission among healthcare workers in the hospital before admission, but ignored transmission between hospital patients and hospital visitors. Hospitals with SARS outbreaks were considered as a single physical unit in both studies. Twenty-one hospitals in Hong Kong experienced the SARS outbreak, and hospitalised patients had little opportunity to contact patients and healthcare workers from other hospitals. Studies with more realistic assumptions about the possible transmission pathways (e.g. healthcare workers could infect other healthcare workers in the hospitals) and the hospital infection settings are therefore needed to better quantify the transmission between and within hospitals and the community.

6.2 Study objectives

I showed in chapter 3 that the spatial distribution of the community infected SARS cases in Hong Kong is related to the hospital outbreaks. Here I quantify disease transmission within and between the hospital and community settings, including their spatiotemporal trends, by capturing the following key elements in the fully extended spatial Wallinga and Teunis method:

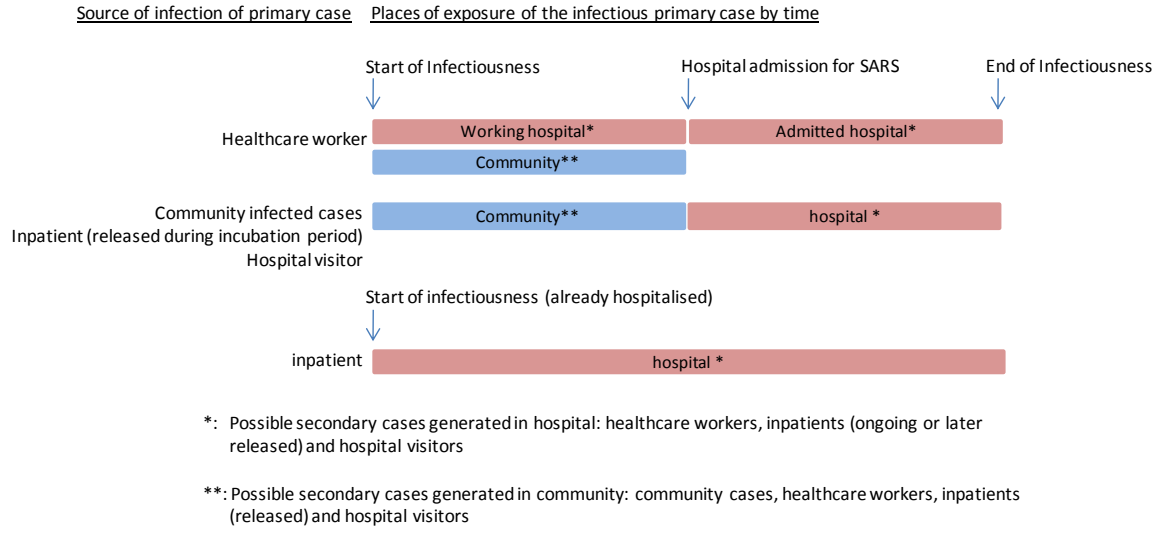
(1) All transmission pathways in the Hong Kong outbreak are accounted for;
(2) Each hospital is treated as a separate spatial unit;
(3) Addition of spatial criterion in identifying potential infectors;

6.3 Method

The fully extended method captures the transmission in the individual hospitals and community explicitly. To do so a potential infector must have been present in the location of infection as well as being infectious on the date of infection of the secondary case. The date of infection and time of infectiousness of the cases were inferred by the incubation period density as in the partially extended method presented in the previous chapter. The infection location of the cases (individual hospital and community) and the infection subgroup (healthcare worker, inpatient, inpatient (released), hospital visitor, community infected) were derived from the source of infection data described in chapter 2. The location of the infectious cases by time (individual hospital and community) was derived using the validated hospitalisation data described in chapter 2. Subject to the transmission pathways defined in the following section, all these together identify the potential infectors of the cases.

6.3.1 Transmission pathways

Figure 6.1 Places of exposure of infectious cases by their source of infection and time course of infection.



From the data validation described in chapter 2, it was possible to define the possible transmission pathways between the subgroups of SARS cases in the Hong Kong outbreak using all available data,including the SARSID integrated database, contact tracing data, the official report on the SARS outbreak which contained more detailed information on the epidemiological linkage of some cases and hence to derive the source of infection of the cases. However, there are cases whose infection source is not definite, because transmission cannot be observed, and information obtainable from the data is sometimes limited. Therefore I allowed healthcare workers, hospital visitors and inpatients (released) to have certain probabilities that they were infected from the community.

Since infectious healthcare workers also spend time in the community before their admission for SARS, they could also generate secondary transmission in the community, along with infectious community infected cases, hospital visitors and inpatients (released) in the community, before being hospitalised for SARS treatment. Therefore healthcare workers could infect other healthcare workers, hospital visitors and inpatients (released) in both the hospital and the community. The possible locations of secondary transmission generated by an infectious individuals based on their source of infection is summarised as follows and depicted in Figure 6.1.

(a). Healthcare worker

Before hospital admission for SARS, infectious healthcare workers could infect the healthcare workers, inpatients, inpatients released during incubation period and hospital visitors in their working hospital, and those in the community including healthcare workers, inpatients (released) and hospital visitors. After hospitalisation for SARS treatment, infectious healthcare workers could infect the healthcare workers, inpatients (ongoing or later released) and visitors in their admitted hospital until the end of their infectious period.

(b). Community infected subject, hospital visitor and inpatient (released)

Infectious community-acquired cases, hospital visitors and inpatients (released) could infect those in the community (including healthcare workers, inpatients (released) and hospital visitors) until they were admitted to hospital for SARS treatment, where they could infect the healthcare workers, inpatients (ongoing or later released) and hospital visitors of the hospital they stayed in until their infectious period ended.

It is assumed that inpatients (released) were not infectious when first discharged from hospital for treating non-SARS condition i.e. the hospitalisation episode when they became infected. Also it was assumed that hospital visitors did not re-visit the hospital that they were infected from their start of infectiousness until their hospital admission (to the same or a different hospital) for SARS treatment.

(c). Inpatient

Inpatients stayed in the hospital continuously until end of infectiousness and could therefore only infect healthcare workers, inpatients (ongoing or later released) and visitors in the hospital(s) they stayed in.

6.3.2 Mathematical formulation

Based on the source of infection data described in chapter 2, let the place of infection of case *j* be $Z(j)$={1, 2,….h_n, h_n+1}, where h_n is the number of hospital with SARS outbreak, 1 to h_n are the index of

the h_n hospitals in Hong Kong with SARS transmission, and h_n+1 is the index for the community. Let *R(j)*={hcw, ip_r, ip, vis, com} represent the infection subgroup for case j defined as one of healthcare worker (hcw), inpatient released during incubation (ip_r), inpatient with ongoing hospitalisation (ip), hospital visitor (vis) and community infected subjects (com). Using the hospitalisation and SARSID integrated dataset, the hospital and community location of case *j* at time *t* is denoted as $L(j,t,\Omega)$, where Ω=1 denotes the hospital location and Ω=2 the community location. $L(j,t,1)$={0, 1,….. h_n}, with 0 denoting that case *j* was not in any hospital at time *t*, and 1 to h_n are the hospital indices as defined in *Z(j)*. $L(j,t,2)$={h_n+1,0}, where h_n+1 is the community index and indicates case *j* was in the community at time *t*, and otherwise 0. Therefore $L(j,t,\Omega)$ can capture the dual location of the healthcare workers before their hospital admission.

Using a constant incubation period *k* implies that the date of infection of case *j*, denoted as $\tilde{t}_j = (s_j - k)$, where s_j is the time of onset of symptoms for case *j*, and s_{ij} is the serial interval between cases *i* and *j*. The probability density function of the serial interval can be obtained as the convolution of the probability density functions of the infectiousness profile (u) and incubation period (q) as:

$$\begin{aligned} w(s_{ij}) &= \sum_{t_j = s_i}^{s_j} u(t_j - s_i) q(s_j - t_j) \\ &= u(s_j - s_i - k) \\ &= u(\tilde{t}_j - s_i) \end{aligned}$$

Therefore for each case *j* there is a set of potential infectors.

Denoting the cases by *i*=1,…*n*, and the infection tree by the vector *v*, where *v(i)* is the infector of case *i*, and *q* as the number of imported cases. Therefore the infectivity of *v(j)* at the time *j* was infected, $\tilde{t}_j$, is given by $u(\tilde{t}_j - s_{v(j)})$. Denote $I_L(v(j), j, \tilde{t}_j)$ as an indicator function of whether *v(j)* was in the same location as *j* at time $\tilde{t}_j$, the likelihood of a network *v* that underlies the observed epidemic curve *t* can be written as:

$$L(v, \psi, \phi; t) = \prod_{j=1}^{j=n-q} u(\tilde{t}_j - s_{v(j)}; \psi) I_L(v(j), j, \tilde{t}_j) [I_h(v(j), j) p_h{}^c(v(j), j) I_{hcw_w}(v(j)) + I_c(v(j), j) p_c{}^c(x_{v(j)}, x_j; \phi) I_{hcw_c}(v(j))]$$

The indicator function of the index and infectee being in the same location is defined as:

$$I_L(v(j), j, \tilde{t}_j) = \begin{cases} 1 & \text{if } L(v(j), \tilde{t}_j, \Omega) = Z(j) \text{ or } L(v(j), \tilde{t}_j, \Omega \mid \Omega = 2) = h_n + 1 \\ 0 & \text{otherwise} \end{cases}$$

The function captures *v(j)* who satisfies at least one of the 3 conditions:

Scenario 1: *v(j)* was at the specific hospital where case *j* was infected at $\tilde{t}_j$

Scenario 2: *v(j)* was in the community at $\tilde{t}_j$, and *j* was community infected

Scenario 3: *v(j)* was in the community at $\tilde{t}_j$, and *j* was hospital infected

The indicator function for hospital transmission

$$I_h(v(j), j) = \begin{cases} \sigma & \text{if } L(v(j), \tilde{t}_j, \Omega \mid \Omega = 1) = Z(j) \\ 0 & \text{otherwise} \end{cases}$$

applies the weighting of hospital acquired infection σ on *v(j)* under scenario 1, because hospital infected cases are considered to have the probability of *(1-σ)* to be community infected. It was assumed σ=99% in the base case model (Table 6.1), and other values of σ were tested in the sensitivity analysis.

The working healthcare worker indicator function in the hospital setting is

$$I_{hcw_w}(v(j)) = \begin{cases} \lambda & \text{if R(v(j))=hcw and } \tilde{t}_j < d_{adm}(v(j)) \\ 1 & \text{otherwise} \end{cases}$$

where $d_{adm}(v(j))$ is the date of hospital admission of case *v(j)*. This function applies the proportion of time λ a healthcare worker spent in the working hospital before being hospitalised. The time spent in the community is therefore *(1-λ)*. λ is assumed to be 80% in the base case scenario (Table 6.1), and the impact of this assumption was assessed in the sensitivity analysis.

The indicator function for community transmission

$$I_c(v(j), j) = \begin{cases} 1 & \text{if } Z(j) = h_n + 1 \\ 1-\sigma & \text{if } Z(j) \neq h_n + 1 \text{ and } L(v(j), \tilde{t}_j, \Omega \mid \Omega = 2) = h_n + 1 \\ 0 & \text{if } L(v(j), \tilde{t}_j, \Omega \mid \Omega = 2) = 0 \end{cases}$$

is used to assign the full probability of infection in community to those under scenario 2, and *(1-σ)* to those under scenario 3 to account for the uncertainty that case *j* could be community infected even though the source of infection data suggested hospital-acquired infection.

The working healthcare worker indicator function in the community

$$I_{hcw_c}(v(j)) = \begin{cases} 1-\lambda & \text{if R(v(j))=hcw and } \tilde{t}_j < d_{adm}(v(j)) \\ 1 & \text{otherwise} \end{cases}$$

captures the proportion of time a healthcare worker spent in the community before hospitalised.

The function $p_h^c(v(j), j)$ denotes the probability of contact between *v(j)* and *j* in the hospital where infection occurred. It is defined as:

$$p_h^c(v(j), j) = \frac{1}{H_{pop}(Z(j))}$$

where $H_{pop}(Z(j))$ gives the population size of the hospital where *Z(j)* was infected. The hospital population is composed of healthcare workers, hospitalised patients and hospital visitors. The number of staff, hospital beds and bed occupancy rate in each public hospital were available in the Statistical Report of the Hospital Authority (Hospital Authority 2003), and they were taken as the population size of healthcare worker and patient (number of beds x occupancy rate) of the hospitals. The ratio of number of hospital visitors per occupied bed (per day) φ was assumed to be 0.8 (Table 6.1), in view of the no visitor policy implemented during part of the outbreak. Different assumptions were tested in the sensitivity analysis. For the 6 private hospitals with SARS transmission, only the number of hospital beds was available from the report. Bed occupancy rate and staff to bed ratio was derived as the average of the comparable hospitals in terms of size and type (e.g. acute or convalescent), because it was observed that large hospitals with more than 1000 beds tend to have higher staff to bed ratios. The number of hospital visitors was derived in the same way as previously described for the public hospitals.

Table 6.1 Parameter assumptions applied in the fully extended spatial method.

Parameters	Symbol	Baseline assumption
Weighting of hospital-acquired infection for hospital infected cases	σ	0.99
hospital visitor per bed ratio	φ	0.8
Working healthcare worker proportion of time in hospital	λ	0.8

The function $p_c^c(x_{v(j)}, x_j; \phi)$ denotes the probability of contact between *v(j)* and *j* in the community. This is derived from the gravity model as in the partial extended method, but normalised by $\tau_{j'}$ such that the probability of travel within district *j'* is scaled to 1, therefore comparable to the hospital infection setting. Under the infector driven mixing model, as described in chapter 5, the normalisation factor $\tau_{j'}$ and $p_c^c(x_{v(j)}, x_j; \phi)$ are derived as:

$$
\begin{aligned}
\tau_{j'} p^{T}{}_{j\to j'} &= 1 \\
\tau_{j'} &= \frac{\sum_{s=1}^{S} N_s{}^{\alpha_1+\alpha_2} f(d_{sj'})}{D_{j'} N_{j'}{}^{\alpha_1+\alpha_2-1} f(0)} \\
p_c{}^{c}(v(j), j) &= \frac{\tau_{j'} p^{T}{}_{v(j)\to j'}}{N_{v(j)'}} \\
&= \frac{\sum_{s=1}^{S} N_s{}^{\alpha_1+\alpha_2} f(d_{sj'})}{D_{j'} N_{j'}{}^{\alpha_1+\alpha_2-1} f(0)} \frac{D_{j'} N_{v(j)'}{}^{\alpha_1+\alpha_2-1} f(d_{v(j)'j'})}{\sum_{s=1}^{S} N_s{}^{\alpha_1+\alpha_2} f(d_{sj'})} \\
&= \frac{N_{v(j)'}{}^{\alpha_1+\alpha_2-1} f(d_{v(j)'j'})}{N_{j'}{}^{\alpha_1+\alpha_2-1} f(0)}
\end{aligned}
\tag{6.1}
$$

Let

$$
\begin{aligned}
g'(s_{v(j)}, s_j, x_{v(j)}, x_j; \psi, \phi) = (u(\tilde{t}_j - s_{v(j)}; \psi) I_L(v(j), j, \tilde{t}_j)[I_h(v(j), j) p_h{}^{c}(v(j), j) I_{hcw_w}(v(j)) \\
+ I_c(v(j), j) p_c{}^{c}(x_{v(j)}, x_j; \phi) I_{hcw_c}(v(j))]
\end{aligned}
$$

Similar to the derivation in the partially extended method on $g(s_{v(j)}, s_j, x_{v(j)}, x_j; \psi, \phi, \eta)$ in equation 5.6, the relative probability that case *i* infected case *j*, can then be written as:

$$
\begin{aligned}
p^{I}{}_{i->j} &= \frac{g'(s_i, s_j, x_i, x_j; \psi, \phi)}{\sum_{m=1, m\neq j}^{m=n} g'(s_i, s_j, x_i, x_j; \psi, \phi)} \\
&= \frac{u(\tilde{t}_j - s_i; \psi) I_L(i, j, \tilde{t}_j)[I_h(i, j) p_h{}^{c}(i, j) I_{hcw_w}(i) + I_c(i, j) p_c{}^{c}(x_i, x_j; \phi) I_{hcw_c}(i)]}{\sum_{m=1, m\neq j}^{m=n} u(\tilde{t}_j - s_m; \psi) I_L(m, j, \tilde{t}_j)[I_h(m, j) p_h{}^{c}(m, j) I_{hcw_w}(m) + I_c(m, j) p_c{}^{c}(x_m, x_j; \phi) I_{hcw_c}(m)]}
\end{aligned}
$$

The probabilities that *i* infected *j* in hospital and in the community respectively are given as:

$$
p_h{}^{I}{}_{i->j} = \frac{u(\tilde{t}_j - s_i; \psi) I_L(i, j, \tilde{t}_j) I_h(i, j) p_h{}^{c}(i, j) I_{hcw_w}(i)}{\sum_{m=1, m\neq j}^{m=n} u(\tilde{t}_j - s_m; \psi) I_L(m, j, \tilde{t}_j)[I_h(m, j) p_h{}^{c}(m, j) I_{hcw_w}(m) + I_c(m, j) p_c{}^{c}(x_m, x_j; \phi) I_{hcw_c}(m)]}
$$

$$
p_c{}^{I}{}_{i->j} = \frac{u(\tilde{t}_j - s_i; \psi) I_L(i, j, \tilde{t}_j) I_c(i, j) p_c{}^{c}(x_i, x_j; \phi) I_{hcw_c}(i)}{\sum_{m=1, m\neq j}^{m=n} u(\tilde{t}_j - s_m; \psi) I_L(m, j, \tilde{t}_j)[I_h(m, j) p_h{}^{c}(m, j) I_{hcw_w}(m) + I_c(m, j) p_c{}^{c}(x_m, x_j; \phi) I_{hcw_c}(m)]}
$$

Therefore the individual reproduction number *R(i)* can be decomposed into the hospital reproduction number $R_H(i)$, the community reproduction number $R_C(i)$. The decomposition also applies in the daily reproduction number *R(t)* to the hospital and community reproduction number $R_H(t)$ and $R_C(t)$. They are formulated as:

$$E[R(i)] = \sum_{k=1}^{k=n-q} p^{I}_{i->k}$$
$$= \sum_{j=1}^{j=n-q} \frac{u(\tilde{t}_j - s_i;\psi) I_L(i,j,\tilde{t}_j)[I_h(i,j) p_h{}^c(i,j) I_{hcw_w}(i) + I_c(i,j) p_c{}^c(x_i,x_j;\phi) I_{hcw_c}(i)]}{\sum_{m=1,m\neq j}^{m=n} u(\tilde{t}_j - s_m;\psi) I_L(m,j,\tilde{t}_j)[I_h(m,j) p_h{}^c(m,j) I_{hcw_w}(m) + I_c(m,j) p_c{}^c(x_m,x_j;\phi) I_{hcw_c}(m)]}$$

$$E[R_H(i)] = \sum_{k=1}^{k=n-q} p_H{}^{I}_{i->k}$$
$$= \sum_{j=1}^{j=n-q} \frac{u(\tilde{t}_j - s_i;\psi) I_L(i,j,\tilde{t}_j) I_h(i,j) p_h{}^c(i,j) I_{hcw_w}(i)}{\sum_{m=1,m\neq j}^{m=n} u(\tilde{t}_j - s_m;\psi) I_L(m,j,\tilde{t}_j)[I_h(m,j) p_h{}^c(m,j) I_{hcw_w}(m) + I_c(m,j) p_c{}^c(x_m,x_j;\phi) I_{hcw_c}(m)]}$$

$$E[R_C(i)] = \sum_{k=1}^{k=n-q} p_C{}^{I}_{i->k}$$
$$= \sum_{j=1}^{j=n-q} \frac{u(\tilde{t}_j - s_i;\psi) I_L(i,j,\tilde{t}_j) I_c(i,j) p_c{}^c(x_i,x_j;\phi) I_{hcw_c}(i)}{\sum_{m=1,m\neq j}^{m=n} u(\tilde{t}_j - s_m;\psi) I_L(m,j,\tilde{t}_j)[I_h(m,j) p_h{}^c(m,j) I_{hcw_w}(m) + I_c(m,j) p_c{}^c(x_m,x_j;\phi) I_{hcw_c}(m)]}$$

And

$$E[R(t)] = \frac{1}{n_t} \sum_{t_j=t} \sum_{j=1}^{j=n-q} \frac{u(\tilde{t}_j - s_i;\psi) I_L(i,j,\tilde{t}_j)[I_h(i,j) p_h{}^c(i,j) I_{hcw_w}(i) + I_c(i,j) p_c{}^c(x_i,x_j;\phi) I_{hcw_c}(i)]}{\sum_{m=1,m\neq j}^{m=n} u(\tilde{t}_j - s_m;\psi) I_L(m,j,\tilde{t}_j)[I_h(m,j) p_h{}^c(m,j) I_{hcw_w}(m) + I_c(m,j) p_c{}^c(x_m,x_j;\phi) I_{hcw_c}(m)]}$$

$$E[R_H(t)] = \frac{1}{n_t} \sum_{t_j=t} \sum_{j=1}^{j=n-q} \frac{u(\tilde{t}_j - s_i;\psi) I_L(i,j,\tilde{t}_j) I_h(i,j) p_h{}^c(i,j) I_{hcw_w}(i)}{\sum_{m=1,m\neq j}^{m=n} u(\tilde{t}_j - s_m;\psi) I_L(m,j,\tilde{t}_j)[I_h(m,j) p_h{}^c(m,j) I_{hcw_w}(m) + I_c(m,j) p_c{}^c(x_m,x_j;\phi) I_{hcw_c}(m)]}$$

$$E[R_C(t)] = \frac{1}{n_t} \sum_{t_j=t} \sum_{j=1}^{j=n-q} \frac{u(\tilde{t}_j - s_i;\psi) I_L(i,j,\tilde{t}_j) I_c(i,j) p_c{}^c(x_i,x_j;\phi) I_{hcw_c}(i)}{\sum_{m=1,m\neq j}^{m=n} u(\tilde{t}_j - s_m;\psi) I_L(m,j,\tilde{t}_j)[I_h(m,j) p_h{}^c(m,j) I_{hcw_w}(m) + I_c(m,j) p_c{}^c(x_m,x_j;\phi) I_{hcw_c}(m)]}$$

I quantified the transmission between and within hospital and community by the reproduction numbers R_{xy}, where *x* denotes the infection place of the case, and *y* denotes the secondary infection generated by the case in location *y*. Therefore R_{hh} measures the infection from hospital to hospital, R_{hc} from hospital to community, R_{cc} from community to community, and R_{hc} from hospital to community, both at the individual and daily level. Because I allowed the hospital infected cases to have a probability of acquiring infection in the community, the probabilities that case *j* is hospital or community infected, denoted as $S_H(j)$ and $S_C(j)$ and $S_H(j) + S_C(j) = 1$, are derived as:

$$S_H(j) = \sum_{k=1}^{k=n} p_h{}^{I}_{k->j}$$
$$= \sum_{i=1}^{i=n} \frac{u(\tilde{t}_j - s_i;\psi) I_L(i,j,\tilde{t}_j) I_h(i,j) p_h{}^c(i,j) I_{hcw_w}(i)}{\sum_{m=1,m\neq j}^{m=n} u(\tilde{t}_j - s_m;\psi) I_L(m,j,\tilde{t}_j)[I_h(m,j) p_h{}^c(m,j) I_{hcw_w}(m) + I_c(m,j) p_c{}^c(x_m,x_j;\phi) I_{hcw_c}(m)]}$$

$$S_C(j) = \sum_{k=1}^{k=n} p_c{}'_{k->j}$$

$$= \sum_{i=1}^{i=n} \frac{u(\tilde{t}_j - s_i;\psi) I_L(i,j,\tilde{t}_j) I_c(i,j) p_c{}^c(x_i, x_j;\phi) I_{hcw_c}(i)}{\sum_{m=1,m\neq j}^{m=n} u(\tilde{t}_j - s_m;\psi) I_L(m,j,\tilde{t}_j)[I_h(m,j) p_h{}^c(m,j) I_{hcw_w}(m) + I_c(m,j) p_c{}^c(x_m, x_j;\phi) I_{hcw_c}(m)]}$$

I further validated the results by identifying cases with high R_{xy} (>3), and considered whether the potential secondary transmission clusters and their index cases estimated here corresponded to the high R_{xy} based on epidemiological linkage retrieved from the SARSID integrated dataset and the official Hong Kong SARS report (Severe Acute Respiratory Syndrome Expert Committee 2003). This shows how the fully extended method captures the transmission dynamics of the important disease clusters, and serves as an evaluation of its performance in identifying the index cases of these clusters correctly. A second evaluation is based upon the index case of disease transmission clusters of size >1 that are identifiable from the SARSID integrated dataset and the official Hong Kong SARS report. Their reported cluster sizes were compared to the estimated R_{xy}, under the caveat that the later generation cases of the clusters had also contributed to further transmission in the cluster and thus the final cluster size.

The extent of local transmission by different settings is assessed by the proportion of within-district transmission in hospital transmission (of all cases), community transmission excluding Amoy cases, and community transmission by Amoy cases. The interpretation of within district transmission in hospital is that the residing district of the index and secondary case in the hospital transmission is the same.

All districts were infected with SARS cases in 4 weeks. To study the spatial patterns of the initial spread, I identified the first case in each district and traced their source of infection. If they were infected from the community, their most probable index case was derived from the fully extended method. Maps of the transmission routes were then generated.

Different assumptions for the hospital visitor to bed ratio φ, the proportion of time working healthcare workers spent in hospitals λ, and the probability of hospital infection σ were tested in the sensitivity analysis. I also studied the impact of using the bilateral movement mixing assumptions.

6.4 Results

Figure 6.2 shows the daily hospital and community reproduction numbers $R_H(t)$ and $R_C(t)$ estimated by the fully extended method under the infector driven spatial mixing assumption. The super-spreading events were well captured. The method accurately captured the onset of the PWH

outbreak at day 10 with a sharp peak $R_H(t)$=6.49, and a high $R_H(t)$=5.12 at day 20 as the cluster grew. At day 28, the day of the Amoy index symptom onset, there was a slight increase in $R_c(t)$, and it gradually reached the peak at day 35 with $R_C(t)$=2.55. The relatively high $R_C(t)$=1.09 at day 85 was from a small outbreak in PWH that further spread to the community, with the small number of incident cases at this time making its effect on the daily estimate of $R_C(t)$ more noticeable. The high estimates of $R_H(t)$ of 1.39 and 3.67 on days 88 and 92 respectively, as well as the high $R_C(t)$ of 3.46 on the same day, were related to an outbreak in NDH with 14 cases that generated further infection in the community. After the PWH outbreak and from day 23 onwards, $R_H(t)$ was below 1 except on the few days mentioned previously. Community transmission was also mostly below 1 from day 38 onwards, after the Amoy outbreak.

Compared to the partially extended method from the previous chapter, the fully extended method has substantially improved the over-smoothing problem in the *R(t)* time-series (Figure 6.3). It performed better than the partially extended model in capturing peak *R(t)* of the PWH outbreak at

Figure 6.2 Estimates of the daily hospital and community reproduction numbers of the fully extended method.

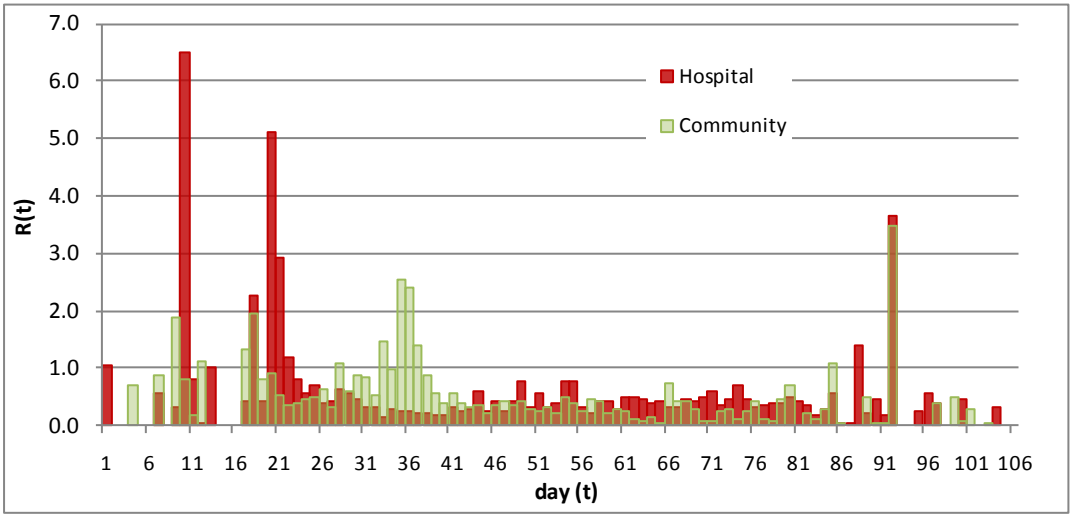
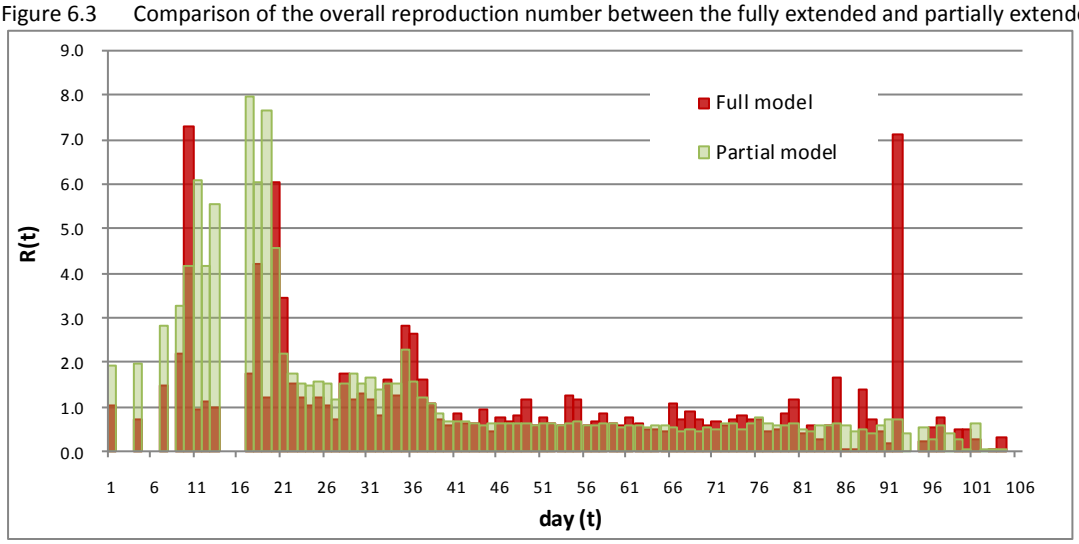


Figure 6.3 Comparison of the overall reproduction number between the fully extended and partially extended method.



its index's symptom onset at day 10 rather than at day 17, estimated higher *R(t)* at the onset day of Amoy's index at day 28, and also detected smaller outbreaks at the late stage of the epidemic with higher *R(t)*. In general, the full model was more sensitive to fluctuations in *R(t)* throughout the epidemic.

Hospital-to-hospital and community-to-community transmission were identified as the main driving forces of the transmission, partly because of the super-spreading events (Figure 6.4). With the Amoy outbreak, the within-community transmission was estimated to account for 42.8% of the total transmission measured by the reproduction number, and with the PWH outbreak the within-hospital transmission was estimated to account for 32.4%. Excluding these two super-spreading events, disease spread between hospital and community (R_{ch}+R_{hc}) was estimated to have increased from 24.9% to 33.4%, and was estimated to be more important than the within-community transmission (27.6%) when the Amoy outbreak was excluded. Within hospital transmission was estimated to be the most important setting where infection occurred. On average, *R(i)*=0.95, R_{hh}=0.31, R_{hc}=0.12, R_{ch}=0.12, R_{cc}=0.41.

Figure 6.4 Composition of the reproduction number of the epidemic of: (a) all cases; (b) all cases excluding the PWH and Amoy clusters.

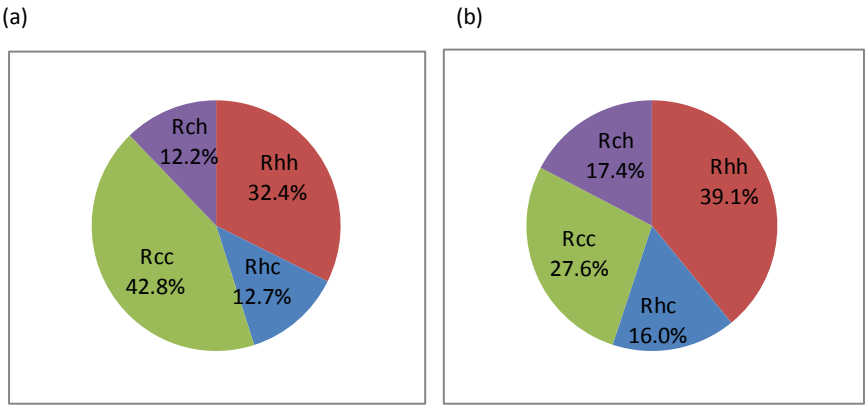
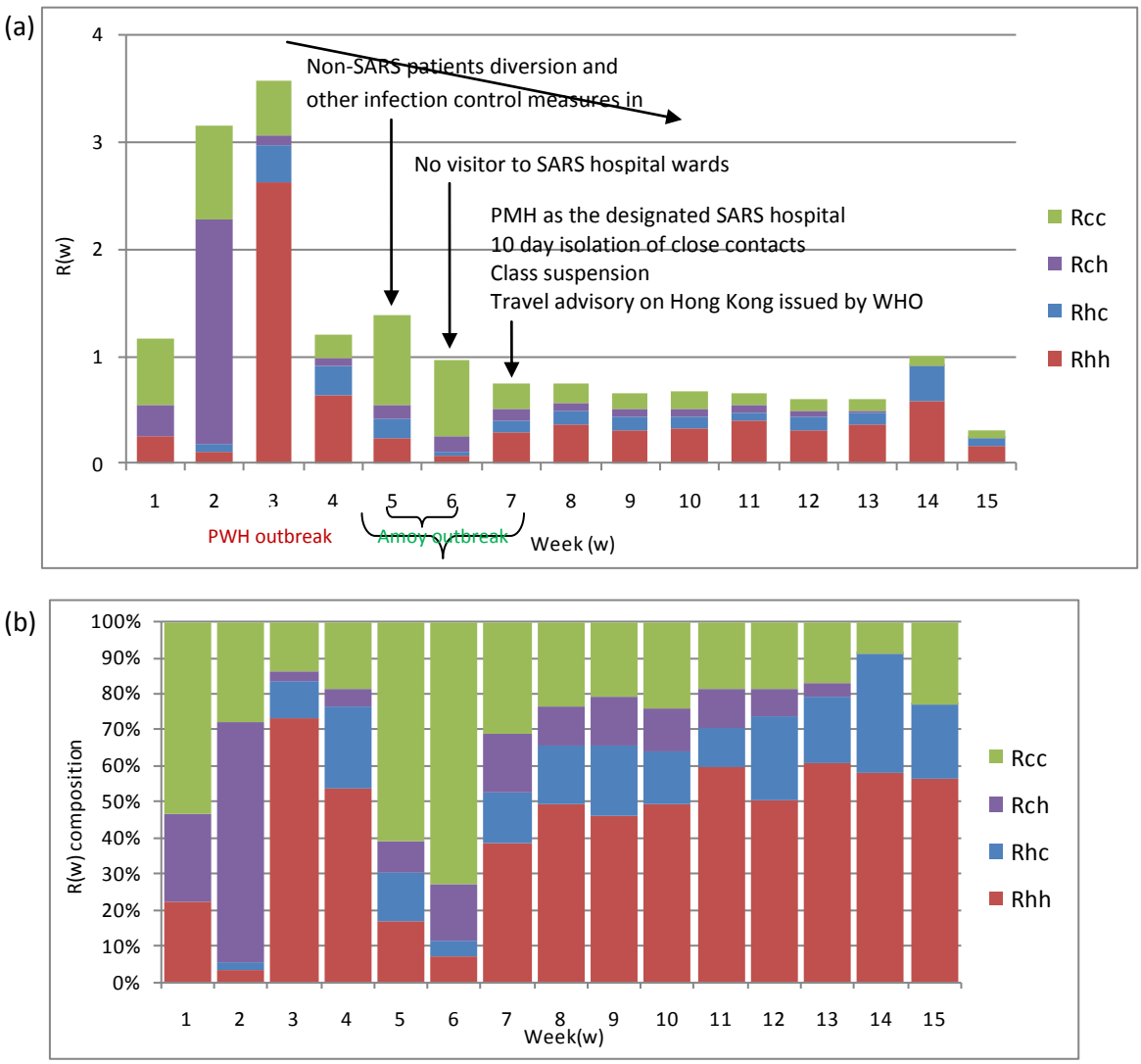


Figure 6.5(a) shows the weekly reproduction number by transmission setting. As the outbreak started from an imported case, disease spread in the first week was initially estimated to be mainly within the community (R_{cc}=0.62). As the infected cases became hospitalised, week 2 estimates showed a strong community-to-hospital transmission [R_{ch}=2.10], eventually leading to the PWH outbreak at week 3 and 4 (R_{hh}=2.62 and 0.64 respectively). In the meantime transmission from hospital back to the community was estimated to have increased (R_{hc}=0.36 and 0.28 respectively in weeks 3 and 4). At week 5, infection control policies were implemented in PWH, and hospital transmission appears to have subsided, but the Amoy outbreak started and drove the within-community transmission during these two weeks (R_{cc}=0.85 and 0.71 respectively in weeks 5 and 6). From week 6 the overall estimated *R(t)* was lower than 1, and more control measures were in place

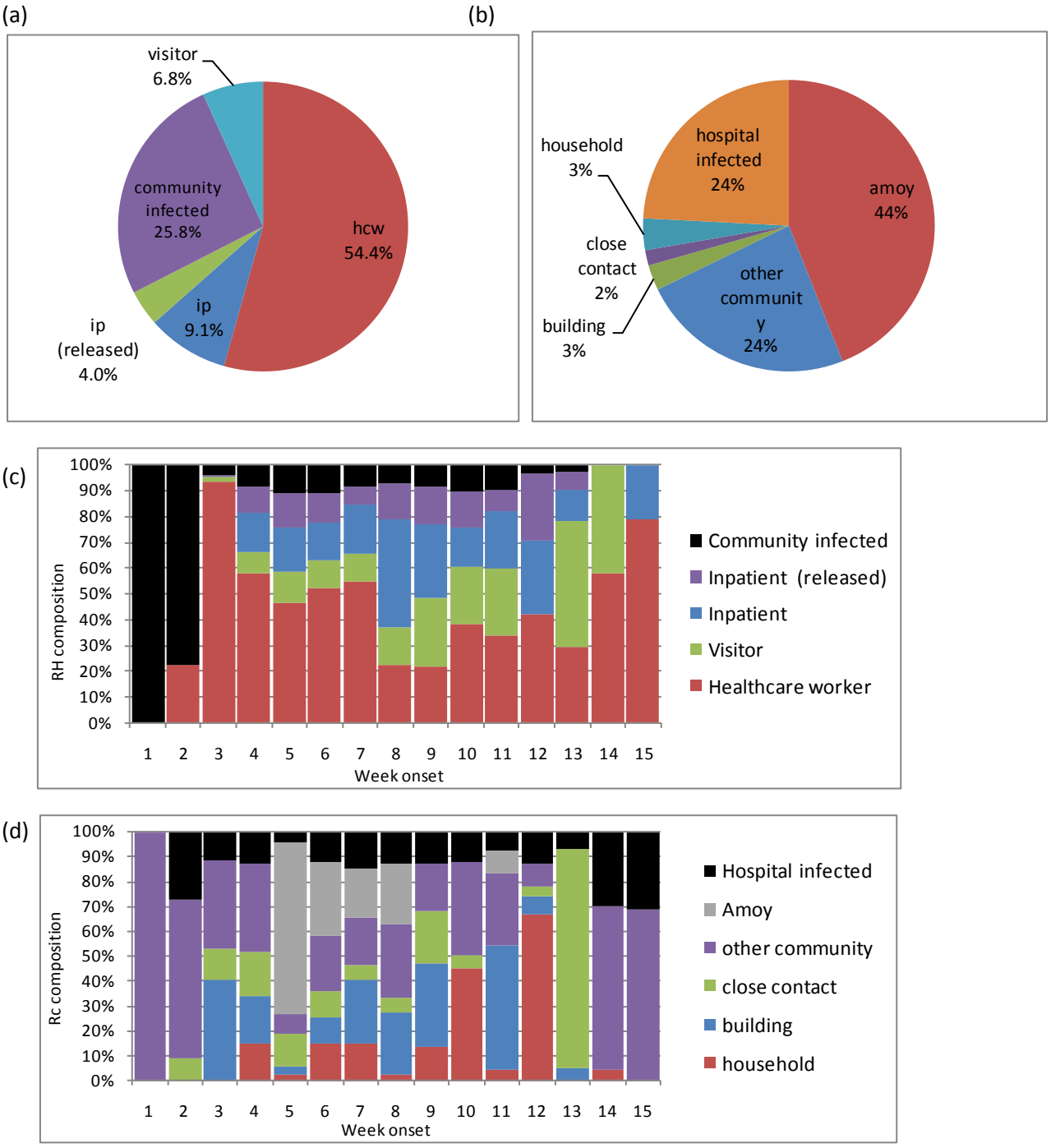
at week 7. From that point onwards within-hospital transmission was estimated to be the main driver of the epidemic, accounting for around half of the transmission [Figure 6.5(b)]. *R(t)* was estimated to have returned to 1 at week 14 driven by transmission from hospital to both hospital and community related to the NDH outbreak, but quickly subsided the week after.

Figure 6.5 (a) Weekly reproduction number estimates by transmission settings, and the key infection control measures implemented; (b) composition of weekly reproduction number by transmission setting.



Around a quarter of hospital infection was estimated to be from cases whose primary infection was in the community [Figure 6.6(a)]. This was mainly driven by its domination at the first 2 weeks of the epidemic, and after that it accounted for no more than 10% of the hospital infection [Figure 6.6(c)]. More than half of the hospital infections were estimated to be from healthcare workers, and they were the key infector for most of the epidemic. Exceptions were during weeks 8 and 9 when hospital infections were driven by inpatients (ongoing and released) largely due to the characteristics of the AHN outbreak that occurred around the time, and in week 13, during which the NDH outbreak occurred, when infection was dominated by visitors.

Figure 6.6 The overall composition of: (a) hospital transmission and (b) community transmission, by the source of infection of the infector. (c) and (d): same as (a) and (b) but on a weekly basis.

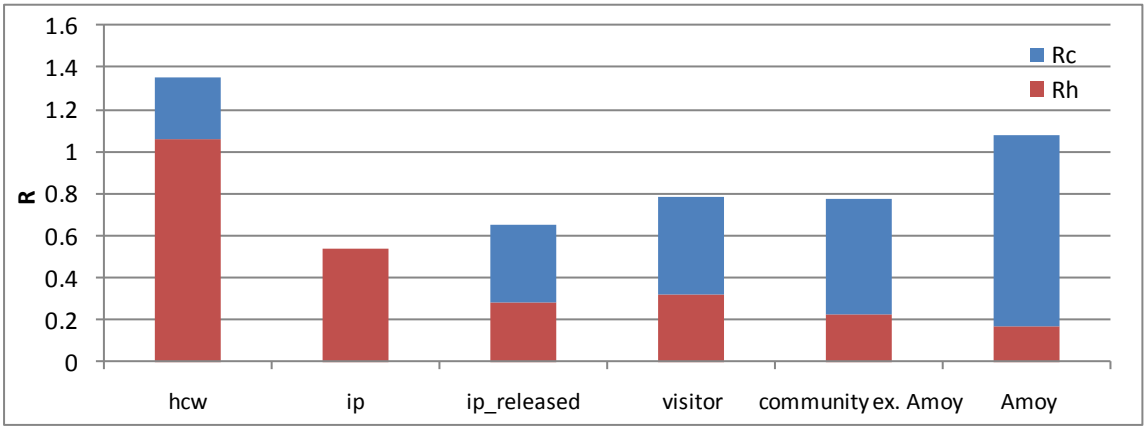


Around a quarter of the community infection was estimated to have been generated by hospital infected cases [Figure 6.6(b)], whilst 44.0% were from Amoy cases. No one type of primary infection source dominated transmission in the community throughout the epidemic. Cases infected from the other community, which included general practitioner clinics, elderly homes, travelling abroad, the Metropole Hotel, and unknown location were estimated to have caused around a quarter of further infections in the community, and they were the estimated to be the key primary infection source

during the first and last 2 weeks of the epidemic [Figure 6.6(d)]. Case infected from an outbreak in residential buildings were estimated to have caused around 40% of the community transmission at week 3 and 11, and household infected cases estimated to have caused over 40% and 60% of the community transmission at week 10 and 12. Cases infected as the close contact of other SARS cases were estimated to have caused over 80% of the further transmission at week 13.

Healthcare workers were estimated to have had the highest overall reproduction number of 1.35 (excluding SSE 1.04), and highest hospital reproduction number R_H=1.06 (excluding SSE 0.79) [Figure 6.7]. Amoy cases on average were estimated to have generated 0.91 secondary cases in the community, and 1.08 secondary cases if including those in the hospital. A typical community case was estimated to have generated 0.77 secondary infection, of which 0.55 were in the community. Hospital visitors were estimated to have generated a very similar level of overall secondary infection (R=0.78, or 0.66 excluding SSE), and had estimated higher transmission potential in the community (R_C=0.46 or 0.37 excluding SSE) than in the hospital (R_H=0.32, or 0.29 excluding SSE). The same is observed in the released inpatients (R=0.65, R_C=0.37, R_H=0.28; excluding SSE R=0.59, R_C=0.32, R_H=0.27). R_H of inpatients was half of the healthcare workers (0.53, or 0.52 excluding SSE).

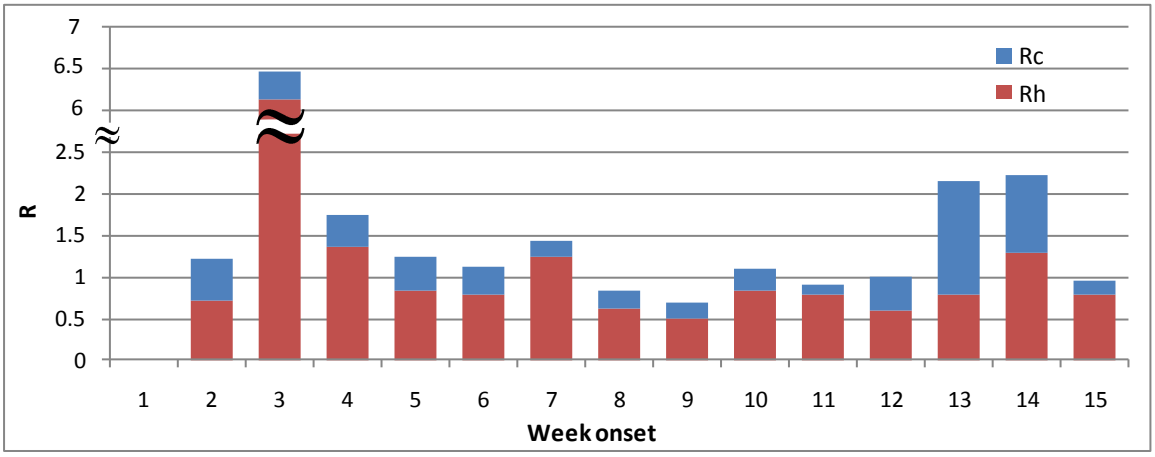
Figure 6.7 The average reproduction number of hospital and community transmission by the subgroup of infected cases.



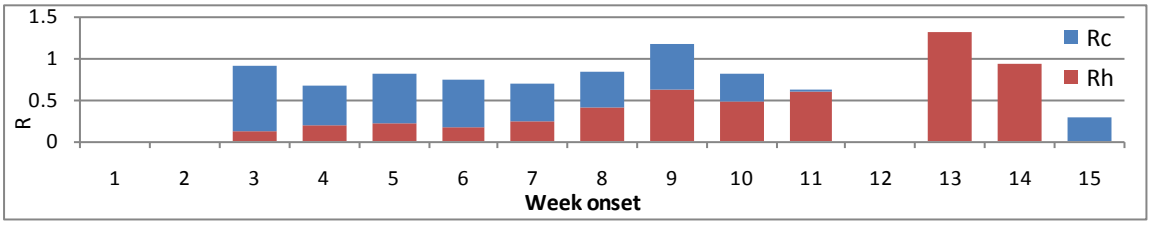
Throughout the epidemic, healthcare workers were estimated to have had an overall reproduction number higher than 1 most of the time (except weeks 8, 9, 11 and 15), and there were a few weeks that the estimated hospital reproduction number was higher than 1 (week 3, 4, 7, 14) [Figure 6.8(a)]. At the peak of PWH outbreak R_H of healthcare worker reached 6.14. Hospital visitors, inpatients (released and ongoing) had estimated reproduction numbers above 1 in a week or two throughout the epidemic [Figure 6.8(b-d)]. Their highest reproduction number was 1.32 for hospital visitors (week 13), 1.92 for released inpatients (week 3), and 1.19 for inpatients (week 8). Community infected cases excluding the Amoy cases had the highest overall estimated reproduction number of

Figure 6.8 Weekly reproduction number of hospital and community transmission by infection subgroup.

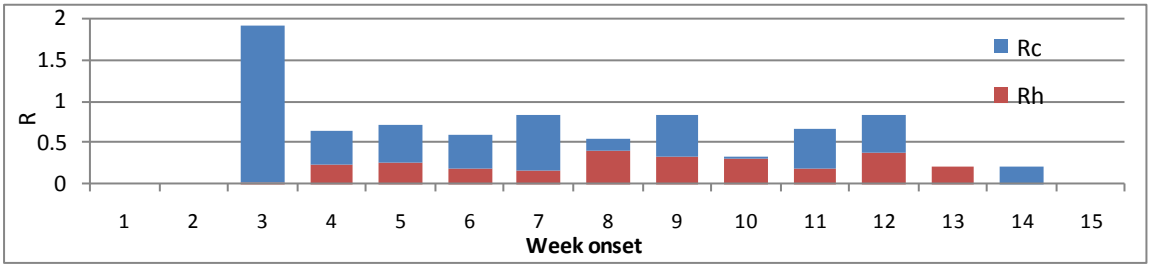
(a) Healthcare worker



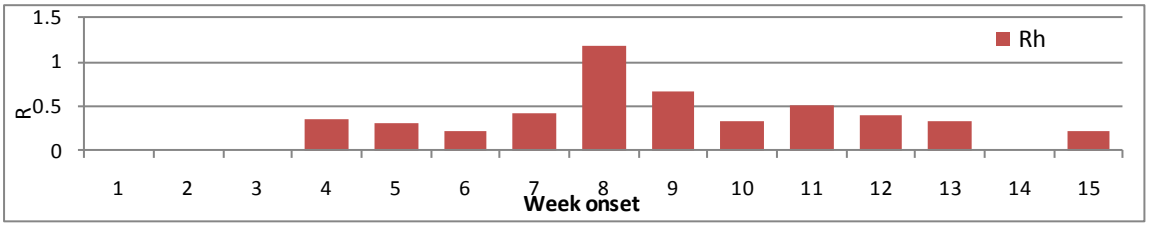
(b) Hospital visitor



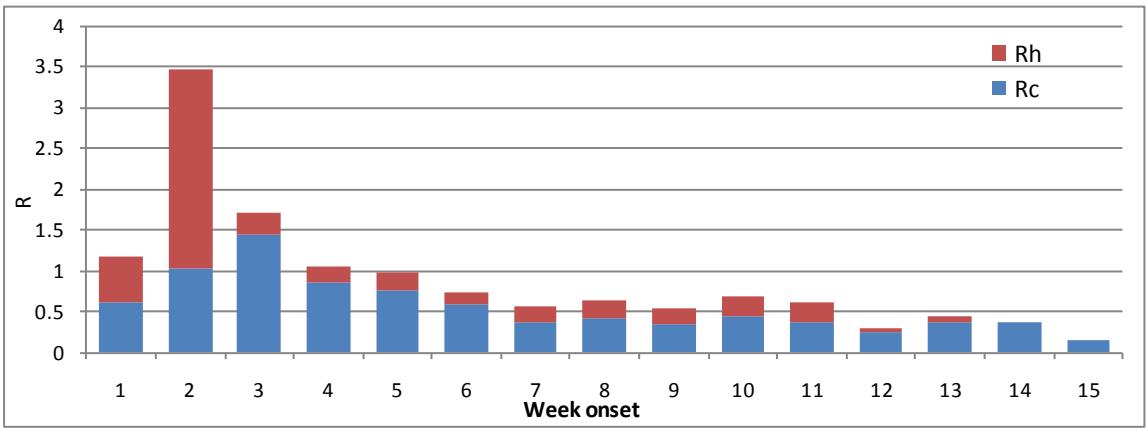
(c) Inpatient (released)



(d) Inpatient



(e) Community infected (excluding Amoy cases)



3.48 at week 2 but was strongly driven by secondary transmission in hospital [Figure 6.8(e)]. Week 3 was the only week their community reproduction number was estimated to be higher than 1 (1.03). From week 5 onwards the overall reproduction number was estimated to be lower than 1 and it gradually decreased for the rest of the epidemic.

Figure 6.9 (a-d) shows the box plots of the estimated individual reproduction numbers by transmission settings by week, and Figure 6.9(e) shows the scatter plot of the overall individual reproduction numbers. Cases with high reproduction number by transmission settings (R_{xy}>3) are marked with the probable secondary transmission cluster identified from epidemiological linkage data and a cases number. Details of the secondary cluster are listed in Table 6.2 under the case number. Table 6.3 lists the index case of identifiable disease clusters of size > 1 based on the epidemiological linkage data, and the reproduction number estimates by transmission settings.

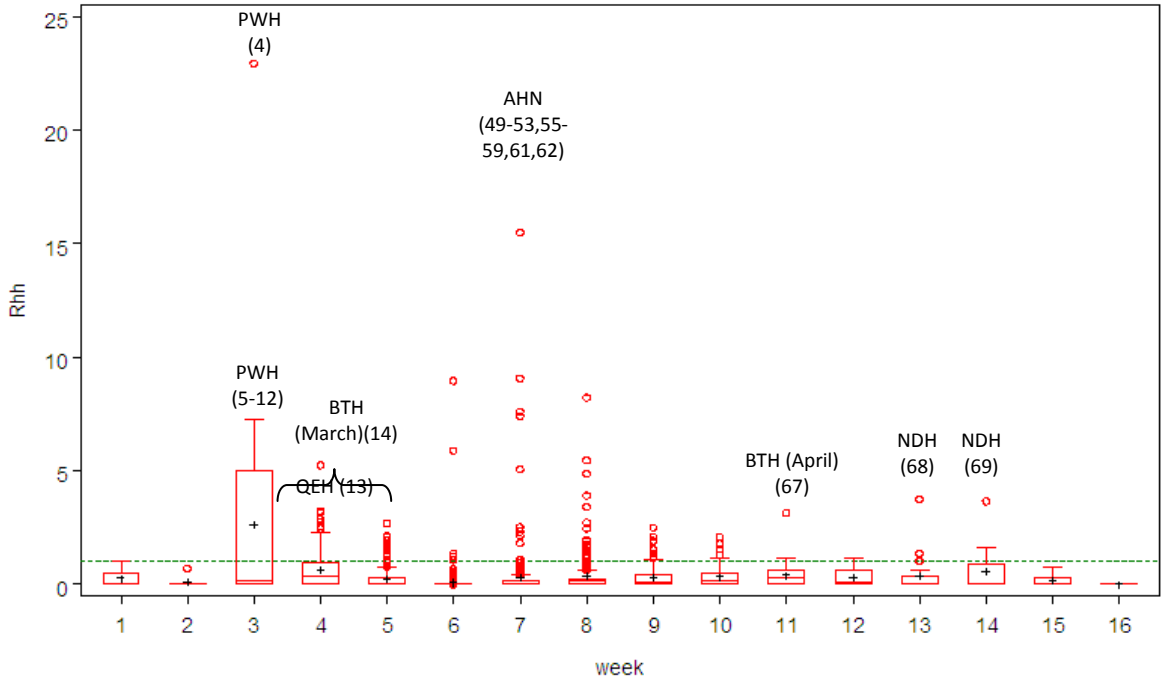
It was known from contact tracing that the index case of the PWH outbreak was infected from the Metropole hotel – where the first occurrence of SARS transmission took place, and the Amoy Garden outbreak was sourced from PWH where the index case had stayed before staying in the Amoy Gardens. The fully extended method identified the index case of the PWH outbreak (case 3) to have a high R_{ch}=12.3, and estimated the first case onset of the cluster who was a healthcare worker (case 4) to have high R_{hh}=22.9, and 7 other healthcare workers to have high R_{hh} between 3.2 and 7.3 (case 5-11). For the Amoy cluster, the method apparently attributed the index to be sourced from community rather than hospital transmission. The real index case had R_{hc}=1.9 (case 76, Table 6.3), but the "pseudo" index case with the same onset day and who lived in the Amoy district had R_{cc}=9.3. Another 25 Amoy district residents also had high R_{cc} between 3.2 and 9.7 (case 15-41).

For the AHN outbreak [Figure 6.9(a)], information from contact tracing was not specific enough to identify the index case (Table 6.2), and there was only low incidence at the beginning before a more established outbreak occurred at day 45. The method estimated that within-hospital and cross-settings transmission were involved, and estimated high R_{hh} from 3.4 to 15.5 for 12 cases including healthcare workers and an inpatient (case 49-53, 55-59, 61-62), high R_{hc} of 3.8 and 4.9 for two cases where the secondary cluster in the community could not be identified, and a high R_{ch} of 3 (case 60).

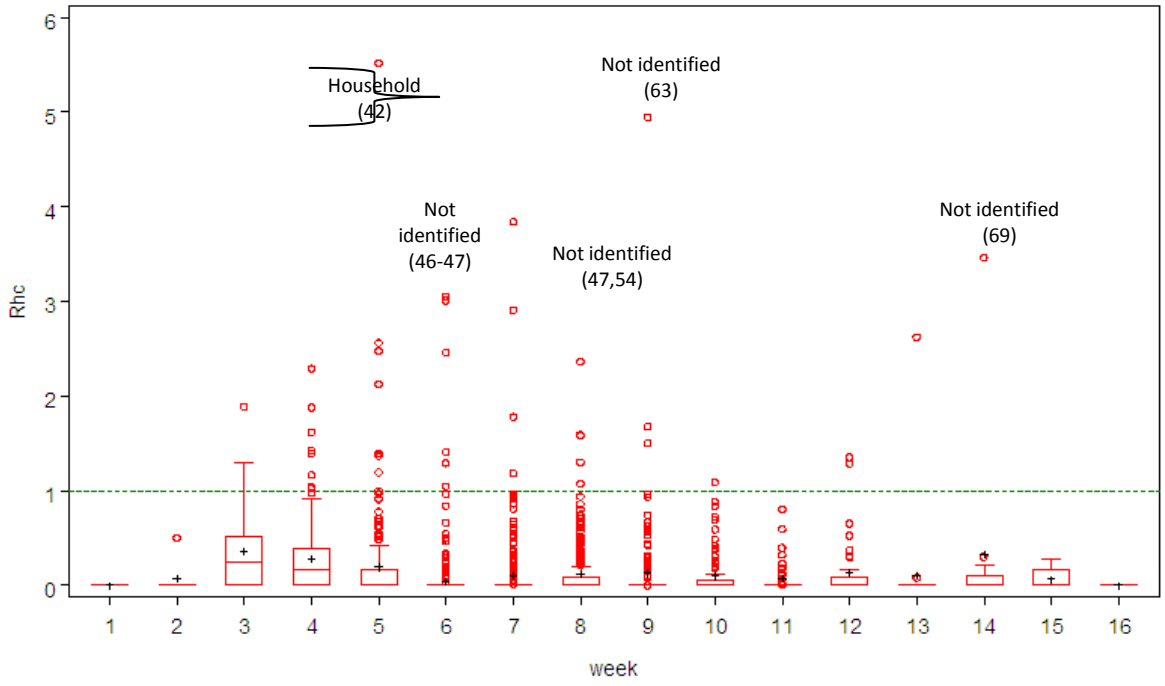
The method also identified some smaller disease clusters in other hospitals (BTH, QEH, NDH) [Figure 6.9(a)] and an elderly home [Figure 6.9(d)]. However, there were 6 cases identified with high R_{cc} (case 1, 2, 43-45, 48) [Figure 6.9(d)] and 4 cases with high R_{hc} (case 47, 63, 69) [Figure 6.9(b)] where the probable secondary cluster(s) were not identifiable from the epidemiological linkage data (Table 6.2). This could be due to community transmissions that were less traceable. Furthermore the

Figure 6.9 (a)-(d): Box plots of individual reproduction number by transmission settings and week of symptom onset. Any case with R_{xy}>3 was marked with a case number (in blanket), and the corresponding probable cluster of secondary infection listed in Table 6.2 (indexed by the case number). Black cross: mean; red horizontal line in box: median; top and bottom of the box: 75th and 25th percentile; upper and lower cap: maximum and minimum observation below 1.5 interquartile range above the 75th percentile and below and 25th percentile; circles: outliners. (e). Scatter plot of individual overall reproduction number R(i) by onset day.

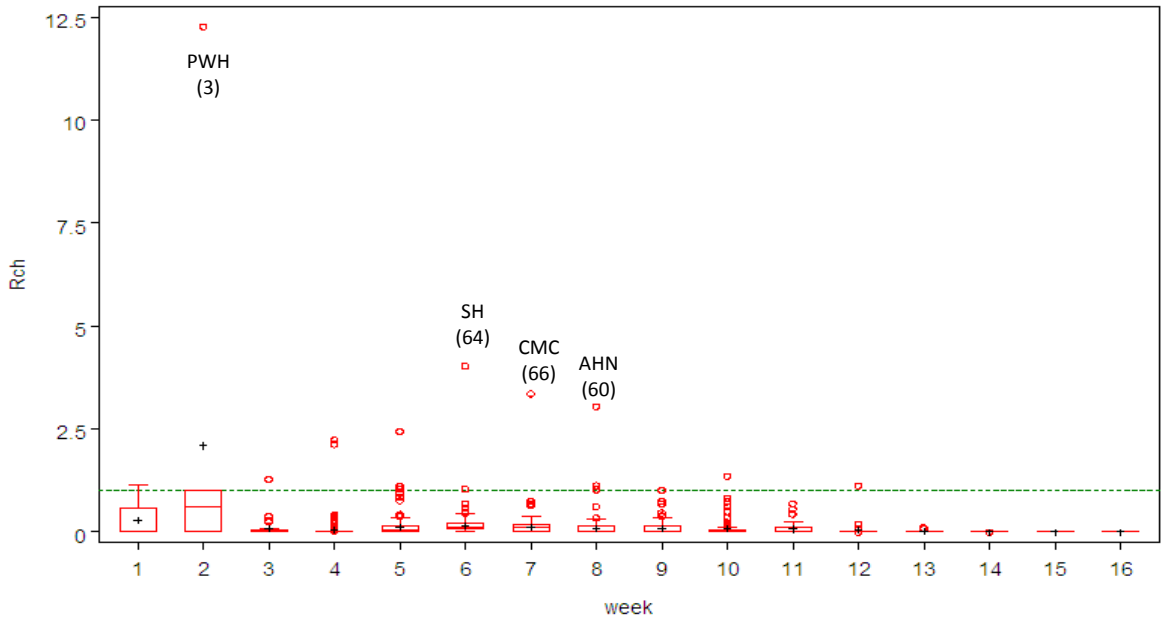
(a) R_{hh}(i)



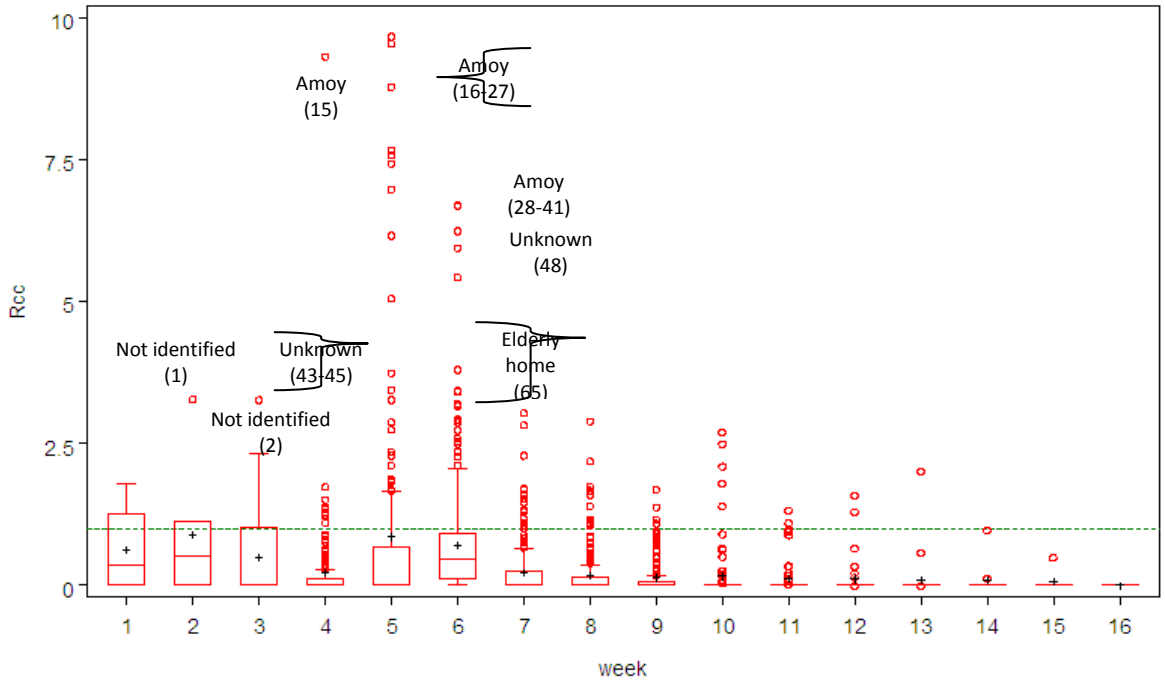
(b) R_{hc}(i)



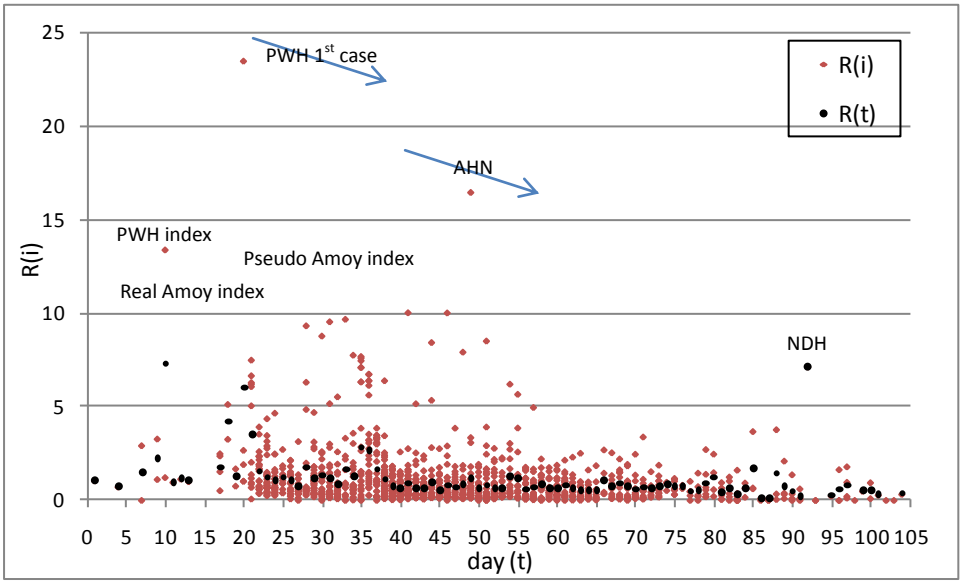
(c) $R_{ch}(i)$


Rch
week
PWH
(3)
SH
(64)
CMC
(66)
AHN
(60)

(d) $R_{cc}(i)$


Rcc
week
Amoy
(15)
Amoy
(16-27)
Amoy
(28-41)
Unknown
(48)
Not identified
(1)
Unknown
(43-45)
Not identified
(2)
Elderly
home
(65)

(e) R(i)



secondary infection might not have occurred as a cluster but individually to several people. In the source of infection data, over 20% of the cases were known to be community infected but without any information about the location of the infection or their potential infector. Comparing the R_{xy} of the index cases of the 38 identifiable clusters to their reported cluster size (Table 6.3), 3 were close to the cluster size (case 76, 82, 83), and others generally had lower corresponding R_{xy} values, which could partly be caused by the infection being assigned to the later generations of the cluster by the epidemic tree reconstruction method.

The scatter plot of the overall individual reproduction number of the fully extended method [Figure 6.9(e)] compared to that of the partial method (Figure 6.16) showed a more random and natural pattern. It also estimated higher *R(i)* for the PWH index (12.3 vs 10.4), and for the "pseudo" Amoy index compared to the correctly identified real index in the partial method (9.3 vs 3.9). The fully extended method also showed sporadic cases with high *R(i)* throughout the epidemic and reflected other disease clusters underway, the wider range of *R(i)* more realistically reflecting the likely variation of *R(i)*.

Overall I estimated that there was 40.7% of the transmission that was within-district over the epidemic. In the community transmission of the Amoy cases it was estimated to be 80.3%, in the community transmission excluding Amoy cases it was estimated to be 33.9%, and in the hospital transmission it was estimated to be 23.8%. The temporal trend of the within-district transmission (Figure 6.10) showed that at the first week of the Amoy outbreak, 87.7% of community transmission of the Amoy cases were estimated to be in the same district, and this decreases gradually, as the Amoy outbreak triggered other outbreaks in nearby housing estates residential buildings, and the

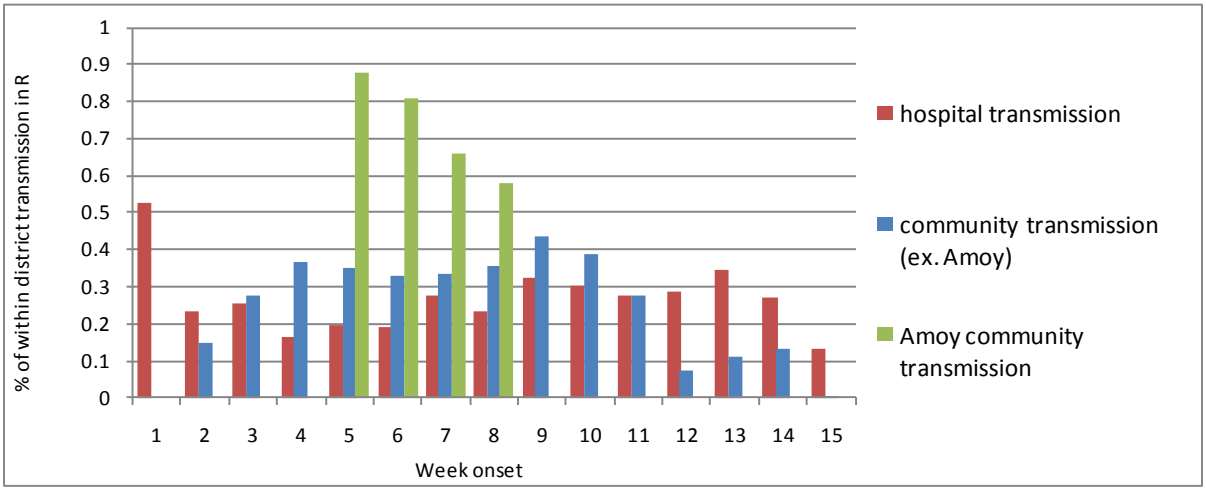
Table 6.2 Summary of the high R_{xy} cases (>3) identified, and details of the matching probable secondary clusters based on the epidemiological linkage data. (#: only R_{xy}>0.01 were displayed; *: from source of infection dataset; **: matched by the epidemiological linkage data. ip: inpatient, hcw: healthcare worker).

cases identified with high R#										The secondary cluster		
Case no	Rhh	Rhc	Rcc	Rch	Ri	onset day (week)	source of infection*	Secondory cluster**	Real index?	Link to cluster (if not the real index)	Real index source of infection*	Cluster 1st onset day	Cluster size
1			3.3		3.3	9 (2)	Metropole	not identified	NA				
2			3.3		3.3	18 (3)	Community	not identified	NA				
3			1.1	12.3	13.4	10 (2)	Metropole	PWH (8A)	Yes			10	123
4	22.9	0.5		0.0	23.5	20 (3)	PWH (8A)	PWH (8A)	No	cluster 1st onset & hcw			
5	5.0	0.1			5.1	21 (3)	PWH (8A)	PWH (8A)	No	hcw			
6	6.0	0.2			6.2	21 (3)	PWH (8A)	PWH (8A)	No	hcw			
7	6.0	0.3		0.0	6.3	21 (3)	PWH (8A)	PWH (8A)	No	hcw			
8	6.0	0.1		0.0	6.1	21 (3)	PWH (8A)	PWH (8A)	No	hcw			
9	6.2	0.5		0.0	6.7	21 (3)	PWH (8A)	PWH (8A)	No	hcw			
10	7.3	0.2			7.5	21 (3)	PWH (8A)	PWH (8A)	No	hcw			
11	3.2	0.8			3.9	22 (4)	PWH (8A)	PWH (8A)	No	hcw			
12	4.5	0.6			5.1	18 (3)	PYN (A5)	PYN (A5)	No	cluster 1st onset	imported	18	9
13	3.3	1.4			4.7	24 (4)	QEH (March)	QEH (March)	No	cluster 1st onset & hcw	imported	24	15
14	5.3	1.0		0.0	6.3	28 (4)	BTH (March)	BTH (March)	No	cluster 1st onset & hcw	close contact	28	20
15			9.3	0.0	9.3	28 (4)	Community	Amoy	No	Amoy district resident	PWH ip	28	447
16			8.8		8.8	30 (5)	Amoy	Amoy	No	Amoy district resident			
17			9.5		9.5	31 (5)	Amoy	Amoy	No	Amoy district resident			
18			5.1	0.1	5.2	31 (5)	Imported	Amoy	No	Amoy district resident			
19			9.7	0.0	9.7	33 (5)	Amoy	Amoy	No	Amoy district resident			
20			7.7	0.1	7.8	34 (5)	Amoy	Amoy	No	Amoy district resident			
21			6.2	0.2	6.3	35 (5)	Amoy	Amoy	No	Amoy district resident			
22			6.2	0.2	6.3	35 (5)	Amoy	Amoy	No	Amoy district resident			
23			7.0	0.1	7.1	35 (5)	Amoy	Amoy	No	Amoy district resident			
24			7.0	0.1	7.1	35 (5)	Amoy	Amoy	No	Amoy district resident			
25			7.4	0.0	7.5	35 (5)	Amoy	Amoy	No	Amoy district resident			
26			7.6	0.0	7.6	35 (5)	Amoy	Amoy	No	Amoy district resident			
27			7.6	0.1	7.7	35 (5)	Amoy	Amoy	No	Amoy district resident			
28			5.4	0.2	5.6	36 (6)	Amoy	Amoy	No	Amoy district resident			
29			5.9	0.2	6.1	36 (6)	Amoy	Amoy	No	Amoy district resident			
30			6.2	0.2	6.4	36 (6)	Amoy	Amoy	No	Amoy district resident			
31			6.2	0.2	6.4	36 (6)	Amoy	Amoy	No	Amoy district resident			
32			6.7	0.1	6.7	36 (6)	Amoy	Amoy	No	Amoy district resident			
33			6.7	0.1	6.7	36 (6)	Amoy	Amoy	No	Amoy district resident			
34			3.2	0.1	3.3	37 (6)	Amoy	Amoy	No	Amoy district resident			
35			3.2	0.2	3.4	37 (6)	Amoy	Amoy	No	Amoy district resident			
36			3.4	0.1	3.5	37 (6)	Amoy	Amoy	No	Amoy district resident			
37			3.4	0.1	3.5	37 (6)	Amoy	Amoy	No	Amoy district resident			
38			3.8	0.1	3.9	37 (6)	Amoy	Amoy	No	Amoy district resident			
39			3.4	0.0	3.5	38 (6)	Amoy	Amoy	No	Amoy district resident			
40			3.2	0.2	3.4	37 (6)	Amoy	Amoy,family cluster	No	Amoy district resident			
41			3.2	0.2	3.3	37 (6)	Amoy	Amoy,family cluster	No	Amoy district resident			
42		5.5			5.5	32 (5)	UCH	Household	Yes	UCH visitor		32	2
43			3.3	0.1	3.3	33 (5)	Close contact	not identified	NA				
44			3.4	0.1	3.6	34 (5)	Community	not identified	NA				
45			3.7	0.1	3.9	35 (5)	Imported	not identified	NA				
46	0.5	3.1			3.5	36 (6)	PWH	PWH	NA	hcw	uncertain	uncertain	uncertain
47	0.2	3.0			3.2	36 (6)	UCH	not identified	NA				
48			3.2	0.0	3.2	36 (6)	Community	not identified	NA				
49	5.9	0.5			6.4	38 (6)	AHN	AHN	No	hcw	community	37	136
50	9.0	1.1			10.0	41 (6)	AHN	AHN	No	hcw			
51	5.1	0.3			5.4	44 (7)	AHN	AHN	No	hcw			
52	7.6	0.8			8.4	44 (7)	AHN	AHN	No	hcw			
53	9.1	0.9			10.0	46 (7)	AHN	AHN	No	hcw			
54	0.0	3.8			3.9	47 (7)	AHN	not identified	NA				
55	7.4	0.5			7.9	48 (7)	AHN	AHN	No	hcw			
56	15.5	1.0			16.5	49 (7)	AHN	AHN	No	hcw			
57	3.9				3.9	51 (8)	AHN	AHN	No	ip			
58	8.2	0.3			8.5	51 (8)	AHN	AHN	No	hcw			
59	5.5	0.7			6.2	54 (8)	AHN	AHN	No	hcw			
60				3.0	3.0	54 (8)	Community	AHN	No	admitted to AHN			
61	3.4	0.5			3.9	55 (8)	AHN	AHN	No	hcw			
62	4.9	0.8			5.7	55 (8)	AHN	AHN	No	hcw			
63	0.0	4.9			5.0	57 (9)	AHN	not identified	NA				
64			1.1	4.0	5.2	42 (6)	SH	SH	Uncertain	SH hcw	uncertain	uncertain	8 max.
65			3.0	0.1	3.1	49 (7)	Elderly home	Elderly home	Probable	1st onset		49	3
66				3.4	3.4	49 (7)	Imported	CMC	No	admitted to CMC	CMC hcw	61	26
67	3.2	0.2			3.4	71 (11)	BTH	BTH (April)	No	cluster 1st onset & hcw	BTH ip	71	8
68	3.8				3.8	88 (13)	NDH	NDH	No	ip	NDH ip	uncertain	29 max.
69	3.7	3.5			7.1	92 (14)	NDH	NDH	No	hcw	NDH ip	uncertain	29 max.
69	3.7	3.5			7.1	92 (14)	NDH	not identified	NA				

Table 6.3 Index cases of disease clusters with size >1 identifiable from the epidemiological linkage data, and their R estimates by transmission settings from the fully extended method. *: from source of infection dataset.

Transmission settings	Case no	Secondary disease cluster*	Cluster index infection source*	index onset day (week)	Cluster size*	Rhh	Rhc	Rcc	Rch	Ri
Hospital to	70	household	PWH hcw	22 (4)	2	0.6				0.6
community	71	household	PWH hcw	23 (4)	3	2.5	0.7			3.2
	72	household	PWH ip	23 (4)	2	0.4				0.4
	73	close contact	PWH ip_released	23 (4)	2	0.3	0.2			0.5
	74	household	PWH visitor	24 (4)	2	0.1	1.0			1.2
	75	household	PYN hcw	24 (4)	2	0.2	0.4			0.6
	76	Amoy	PWH ip_released	28 (4)	447	0.1	1.9			2.0
	76	household	PWH ip_released	28 (4)	2	0.1	1.9			2.0
	77	household	PWH hcw	29 (5)	2	0.8	0.5			1.3
	78	close contact	BTH ip_released	33 (5)	3	0.3	0.4			0.7
	79	household	PWH	35 (5)	2	0.2	1.4			1.5
	80	household	AHN	40 (6)	2	0.2	0.3			0.4
	81	household	PMH	47 (7)	2	0.4	0.6			1.0
	82	household	AHN	48 (7)	2	0.1	1.8			1.9
	83	household	AHN ip_released	57 (9)	3		1.7			1.7
	83	close contact	AHN ip_released	57 (9)	2		1.7			1.7
	84	close contact	UCH ip_released	59 (9)	3	1.2	0.6			1.8
	85	household	AHN	61 (9)	2	0.9	0.2			1.2
	86	household	BTH	64 (10)	2	1.1	0.5			1.6
	87	household	AHN ip_released	65 (10)	2	0.2				0.2
Community to	88	PYN (A5)	imported	7 (1)	9			1.8	1.1	2.9
hospital	3	PWH (8A)	metropole	10 (2)	123			1.1	12.3	13.4
	89	SPH	metropole	13 (2)	9				1.0	1.0
	90	QEH	imported	17 (3)	15			1.2	1.3	2.5
	91	BTH (March)	close contact	24 (4)	20			0.3	0.3	0.7
Community	88	Metropole	imported	1 (1)	3	1.0				1.0
to community	3	close contact	metro	10 (2)	4			1.1	12.3	13.4
	92	close contact	imported	27 (4)	2			0.4	0.1	0.5
	93	household	community	35 (5)	2			2.7	0.1	2.9
	94	household	community	36 (6)	3			0.6	0.1	0.7
	95	household	community	37 (6)	2			1.4	0.1	1.5
	96	Elderly home	Amoy resident	38 (6)	2			0.8	0.2	1.0
	97	household	community	39 (6)	2			0.7	0.1	0.7
	98	household	community	49 (7)	2			0.7	0.1	0.8
	99	household	community	57 (9)	2				0.2	0.2
	100	household	community	58 (9)	2			0.7	0.2	0.9
	101	household	elderly home	60 (9)	2			0.3	0.3	0.6
	102	close contact	household	69 (10)	4			1.8	0.0	1.8

Figure 6.10 Percentage of within district transmission in all hospital transmission, community transmission excluding Amoy cases, and community transmission of the Amoy cases.



Amoy cases were also considered as the infector of other community transmission outside the Amoy cluster. Therefore these figures are expected to be lower than the percentage of the Amoy cases lived in the Kwun Tong district, i.e. 94.2%.

Figure 6.11 shows the percentage of within-district transmission by transmission settings. Within community, 59.8% of the transmission was estimated to be in the same district. Excluding the Amoy cases, hospital-to-community transmission was more spatially localised than the community-to-community transmission (estimated within-district transmission 35.8% vs 32.6%), probably because the former was mainly composed of secondary household transmission of the hospital infected cases, while within-community transmission also consisted of less localised infection, e.g. between close contacts, in health clinics etc. Community-to-hospital and hospital-to-hospital transmission were comparable in the estimated within-district distribution (22.5% vs 24.2%). By the type of infected subjects, the key drivers for within-district transmission from hospital to community were inpatients released during the incubation period (43.6%), and hospital visitors (39.9%) [Figure 6.12]. Healthcare workers had the strongest tendency to infect someone from another districts.

Figure 6.11 Percentage of within-district transmission in the reproduction number by transmission settings.

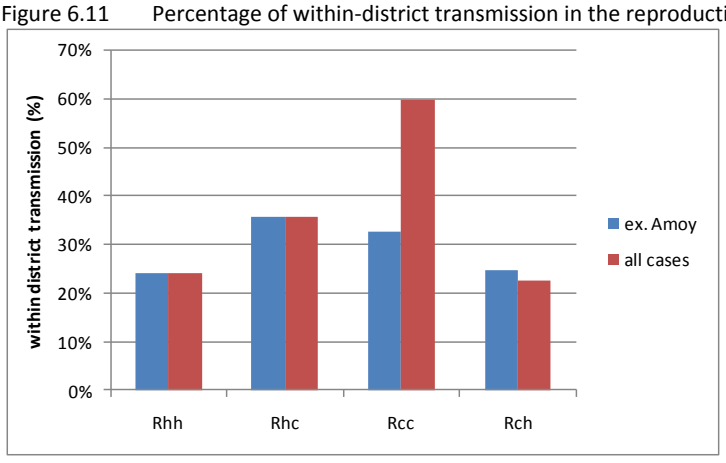
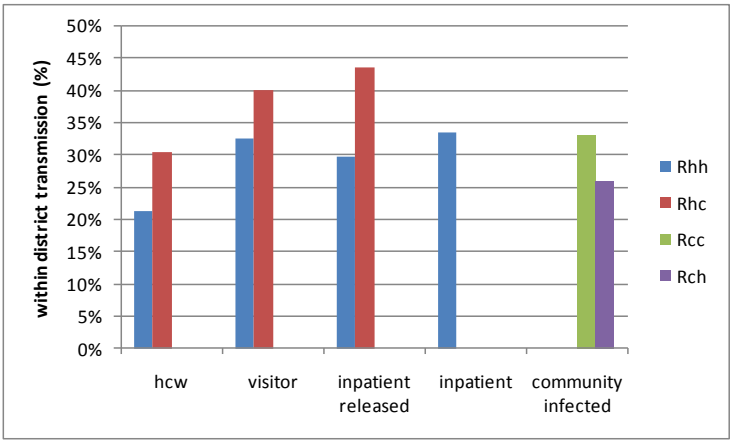


Figure 6.12 Comparison of the within-district transmission by type of infected subjects.



The estimated percentage of within-district transmission by district is shown in Figure 6.13. Kwun Tong district, because of the Amoy outbreak, had high estimated within district transmission of 68.1%, followed by Tai Po district, where AHN is located, at 42.2%, and Sha Tin, where PWH is located, at 32.8%. These are the same locations as those where the key spatial clusters were identified in Chapter 3 (Figure 3.6).

The fully extended method estimated lower within-district transmission during much of the epidemic than the partially extended method (Figure 6.14). This is because the latter does not differentiate the infection location, and therefore captures more cases as within-district than the additional spatial criteria applied in the full method. At the peak of the two super-spreading events, the full method estimated higher within-district transmission.

Figure 6.13 Overall percentage of within-district transmission by district.

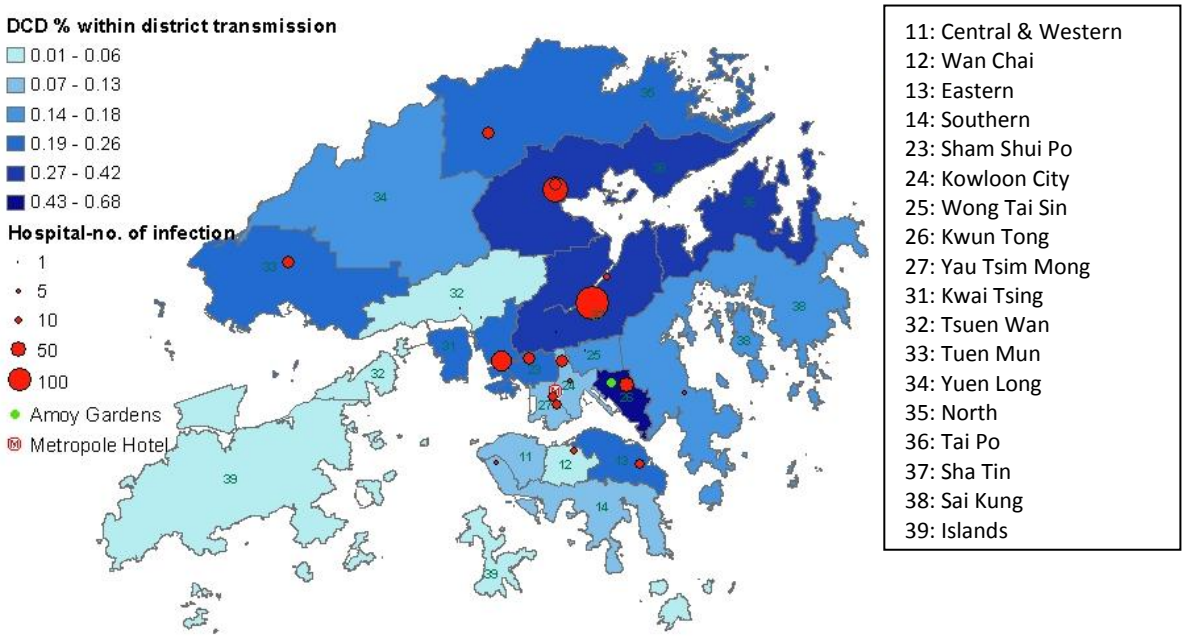
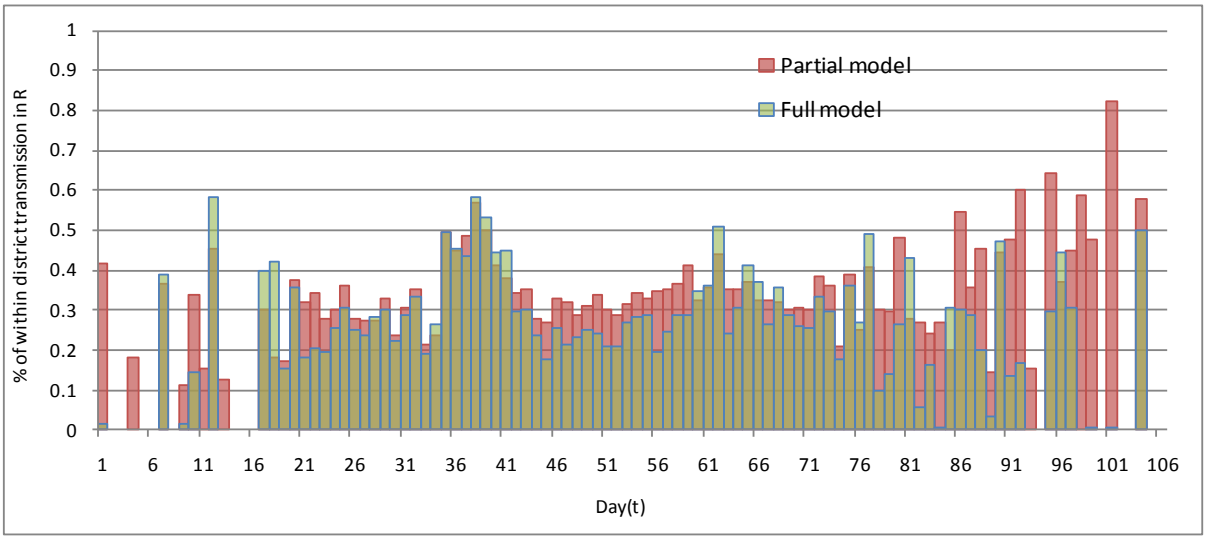


Figure 6.14 Comparison of the proportion of within-district transmission over time between the full and partially extended method.



Using the infection source and most probable index case of the first case in each district, the map of estimated disease spread patterns is given in Figure 6.15, and the sequence of infection and population size of the districts are shown in Figure 6.16. The sequence of district infection was negatively correlated with population size (Spearman's rank correlation coefficient=-0.62, p-value=0.06) and to a lesser extent with population density (Spearman's rank correlation coefficient=-0.54, p-value=0.02). The first 5 districts (27.8%) were estimated to have been infected by imported cases at week 1 and 2, two of which have high population size. Further spread through hospital transmission (KWH) and community transmission from Metropole hotel at week 2 were estimated to be to moderate and low population sized districts which are not adjacent to the infected districts. PWH was estimated to be the main source of infection for week 3 and 4 when the PWH outbreak peaked. In general, and the pattern of spatial spread was driven by PWH as the focal point and also exhibited a mixture of diffusion and population size hierarchy.

Figure 6.17 shows the overall reproduction number by district. Four districts with reproduction number higher than 1 included Kowloon City (1.17), Tai Po (1.12), Kwun Tong (1.07), and Wan Chai (1.02). The variation of R_{hc} and R_{ch} among districts was very small (<0.25), therefore it was mainly the within-hospital and within-community transmission that determine the transmissibility of the districts (Figure 6.18). However, the estimate of the reproduction number by district is affected by the number of cases in the district, and districts like Wan Chai, which only had 16 cases (0.9%) but with cases onset in the early epidemics, therefore tended to have higher reproduction number before the effects of infection control was shown (week 3 R_{hh}=3.40, Figure 6.20 and Figure 6.21). Similarly, Kowloon city was estimated to have accounted for 4.0% of the cases, and had high estimated R_{cc} at week 2 (3.28) and only had moderate to low estimated reproduction numbers

Figure 6.15 Disease spread patterns shown as the transmission route of the first case of each district. Broken black arrow: imported; red arrow: transmitted from hospital infection; green arrow: transmitted from community infection. Blue gradient colour: weekly incidence of the district.

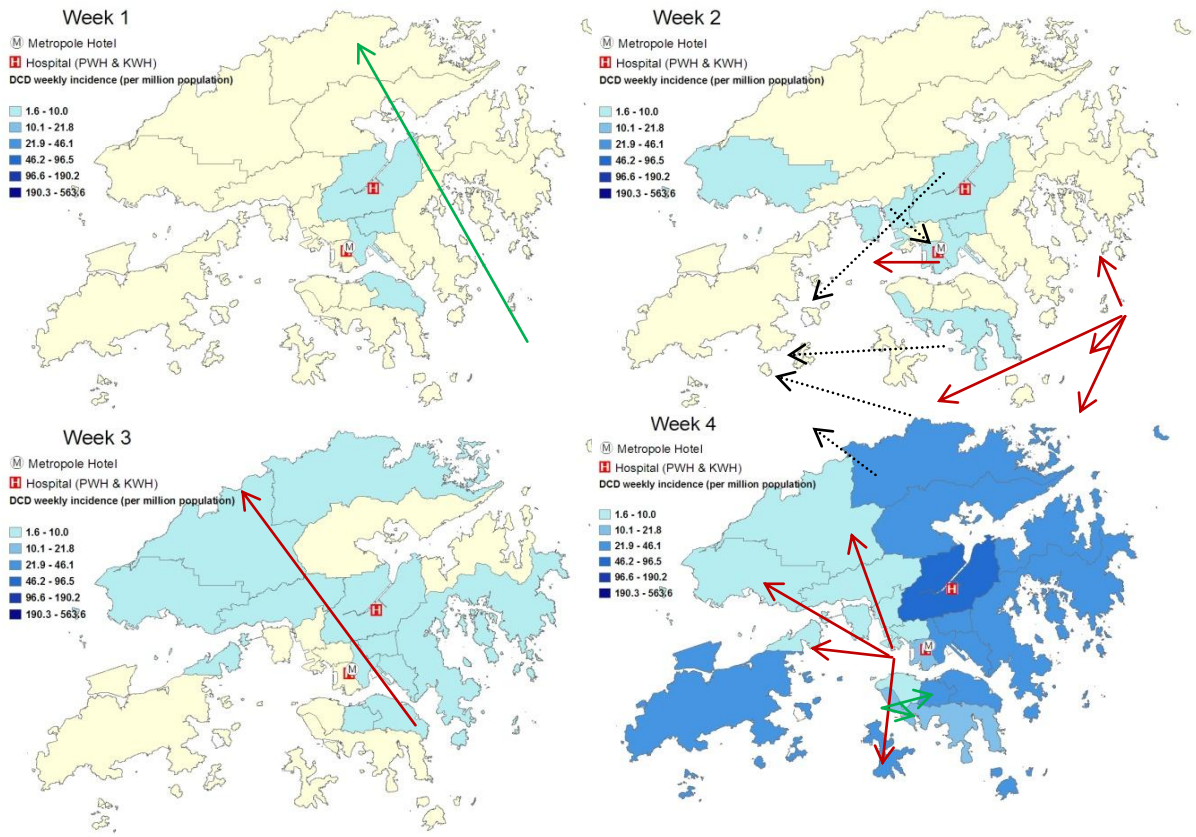


Figure 6.16 Population size of the district (blue gradient colour) and the sequence of district being infected (size of red circle, smaller size earlier infection). Some districts were estimated to have been infected on the same day.

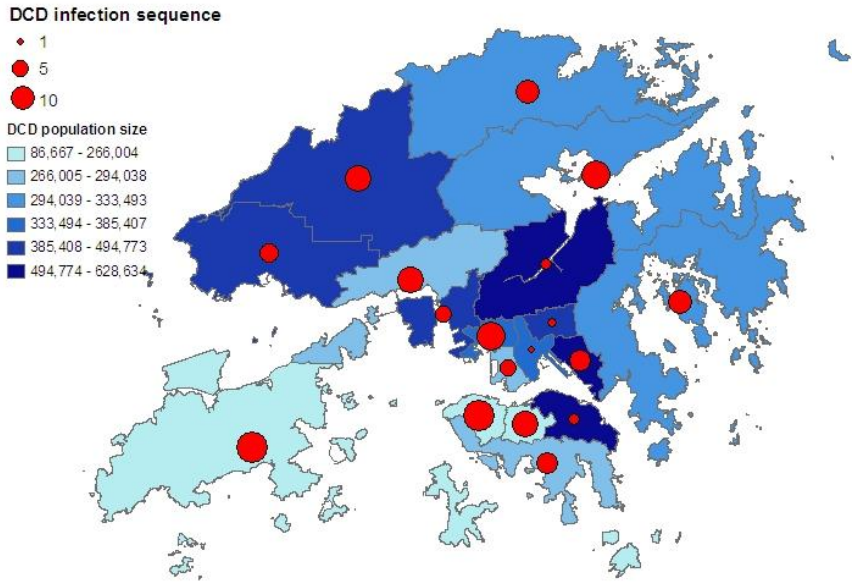
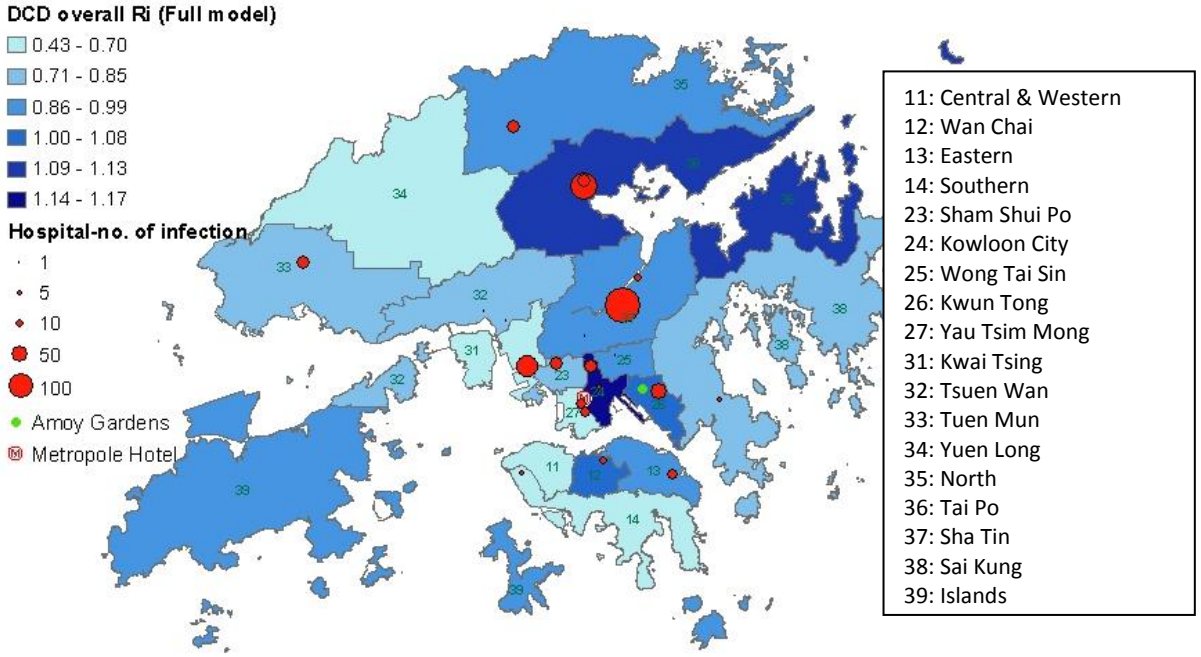


Figure 6.17 Overall reproduction number by district.



afterwards (Figure 6.20 and Figure 6.23). On the other hand, districts with a high number of cases may have individuals with high reproduction numbers but overall a relatively low average, as is the case in Sha Tin (R=0.95). Tai Po and Kwun Tong, which also had high within-district transmission, were identified as potentially sustaining transmission as indicated by the high estimated average reproduction number.

Figure 6.19 shows that in most districts hospital-to-hospital transmission was estimated to be the most important driver of transmission. Kwun Tong, and the adjacent districts Wong Tai Sin, Kowloon City and Sai King were however estimated to have been dominated by within-community transmission. This is consistent with the findings from the spatial cluster detection in Chapter 3.

The spatiotemporal display of the reproduction number by district (Figure 6.20) showed that at week 2, the reproduction number in Sha Tin was as high as 13.39, mainly fuelled by R_{ch} (12.27, Figure 6.24) because of the PWH index case. At week 3 as infection spread from PWH to Yuen Long and Tsuen Wan (Figure 6.15), these districts soon became highly transmissible (R=7.50 and 5.06 respectively), mainly from the hospital-to-hospital infection (Figure 6.21). From week 4 onwards there were usually a few districts, most often the Tai Po district, with reproduction number higher than 1 among other low infection districts. In both the within-community and hospital-to-hospital transmission, I observed that even at the late stage of the epidemic, there were sporadic increases of the reproduction number to greater than 1 (in R_{cc} or R_{hh}, e.g. week 11, 13, 14 in Figure 6.21, week 10 and 13 in Figure 6.23).

Figure 6.18 Reproduction number by transmission setting by district.

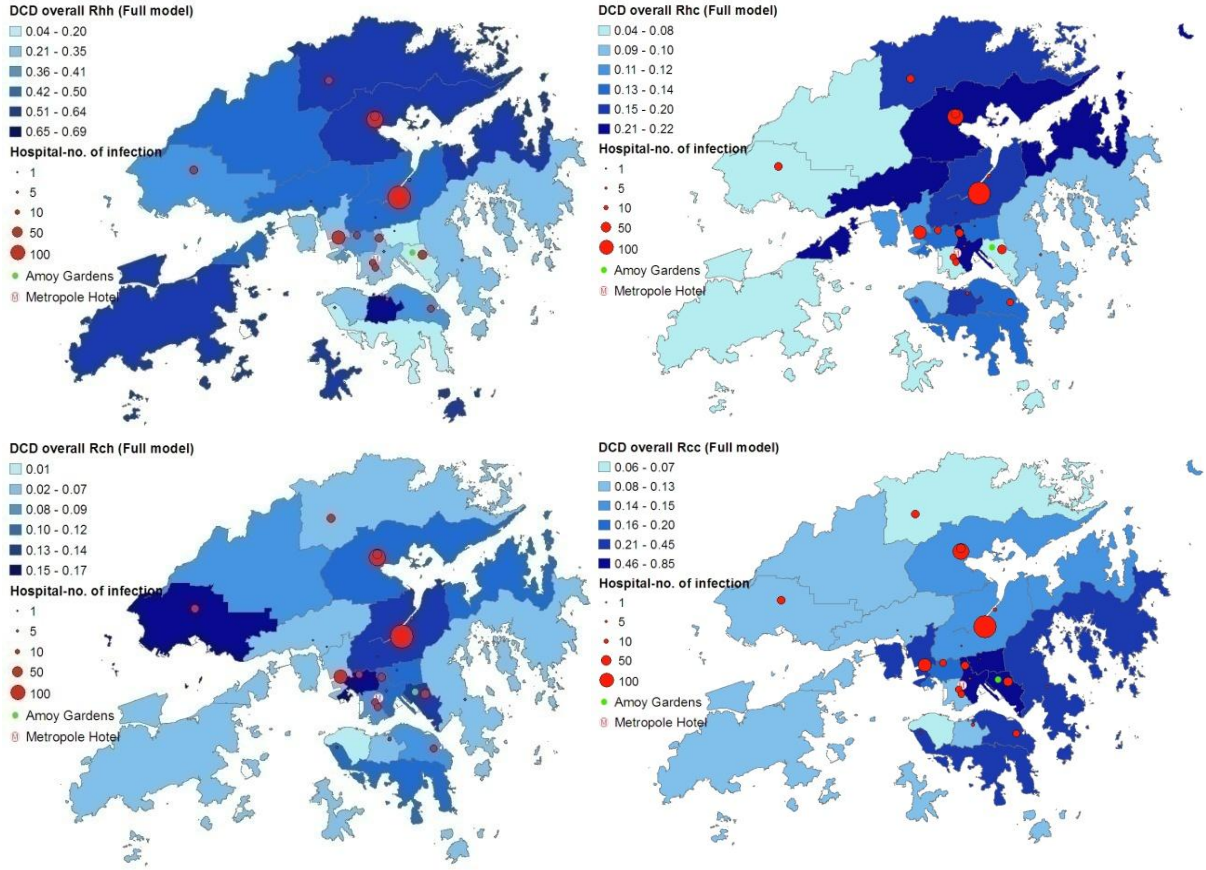


Figure 6.19 Composition of the district reproduction number.

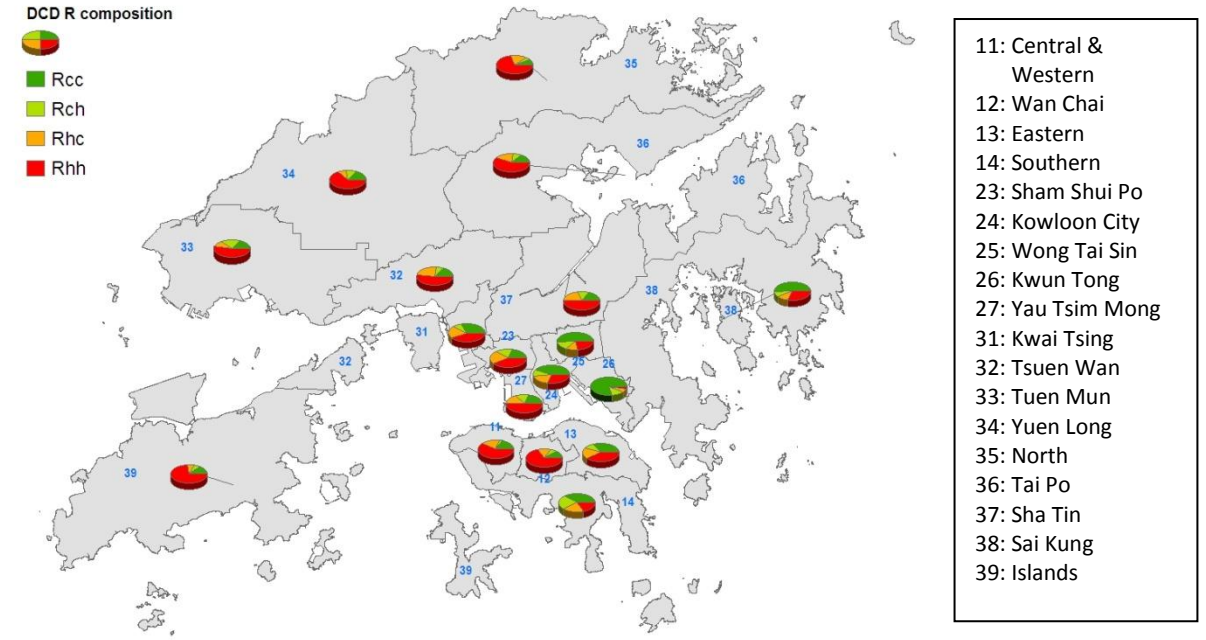


Figure 6.20 Spatiotemporal trend of the overall reproduction number by district by week.

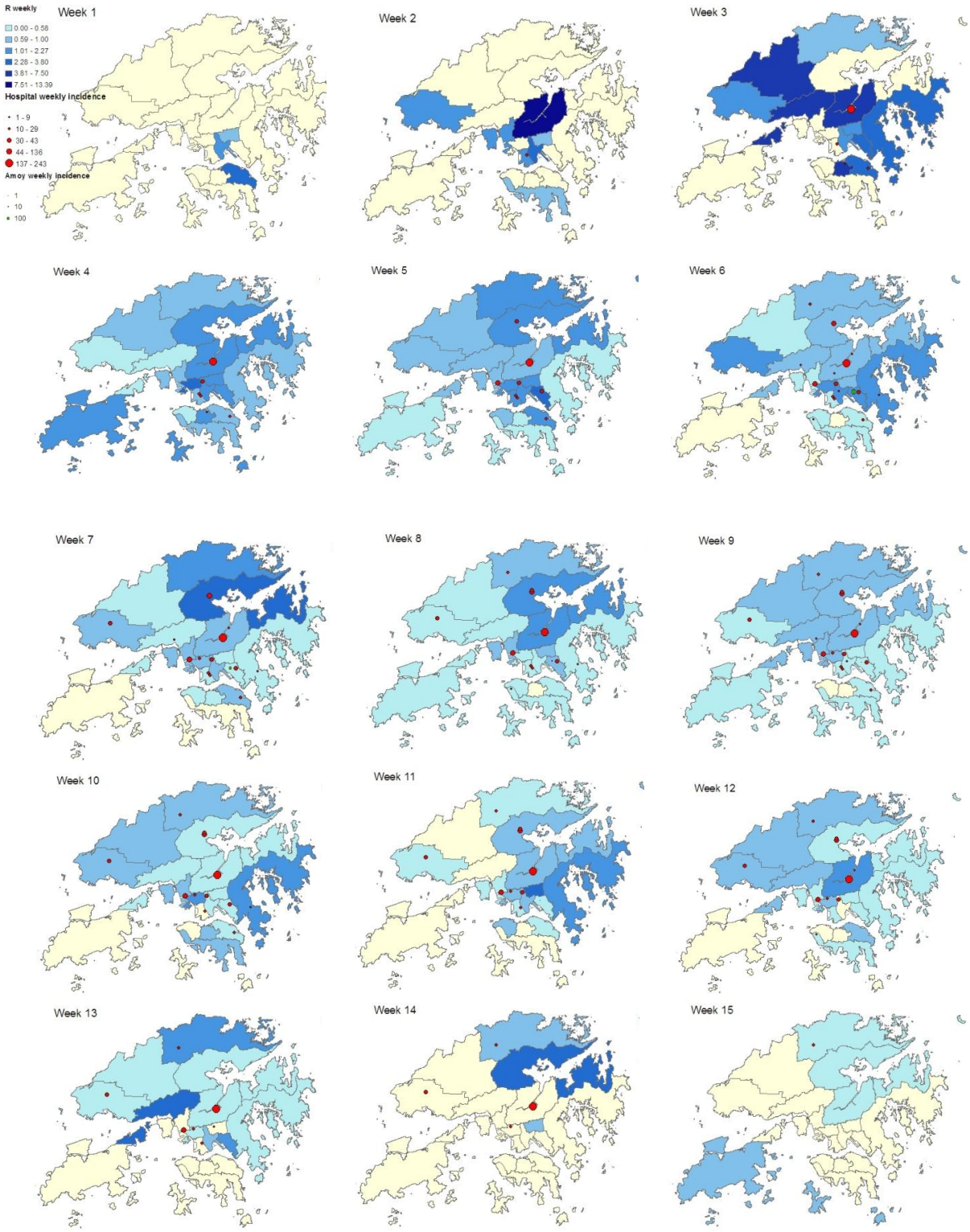


Figure 6.21 Spatiotemporal trend of the hospital-to-hospital reproduction number R_{hh} by district by week.

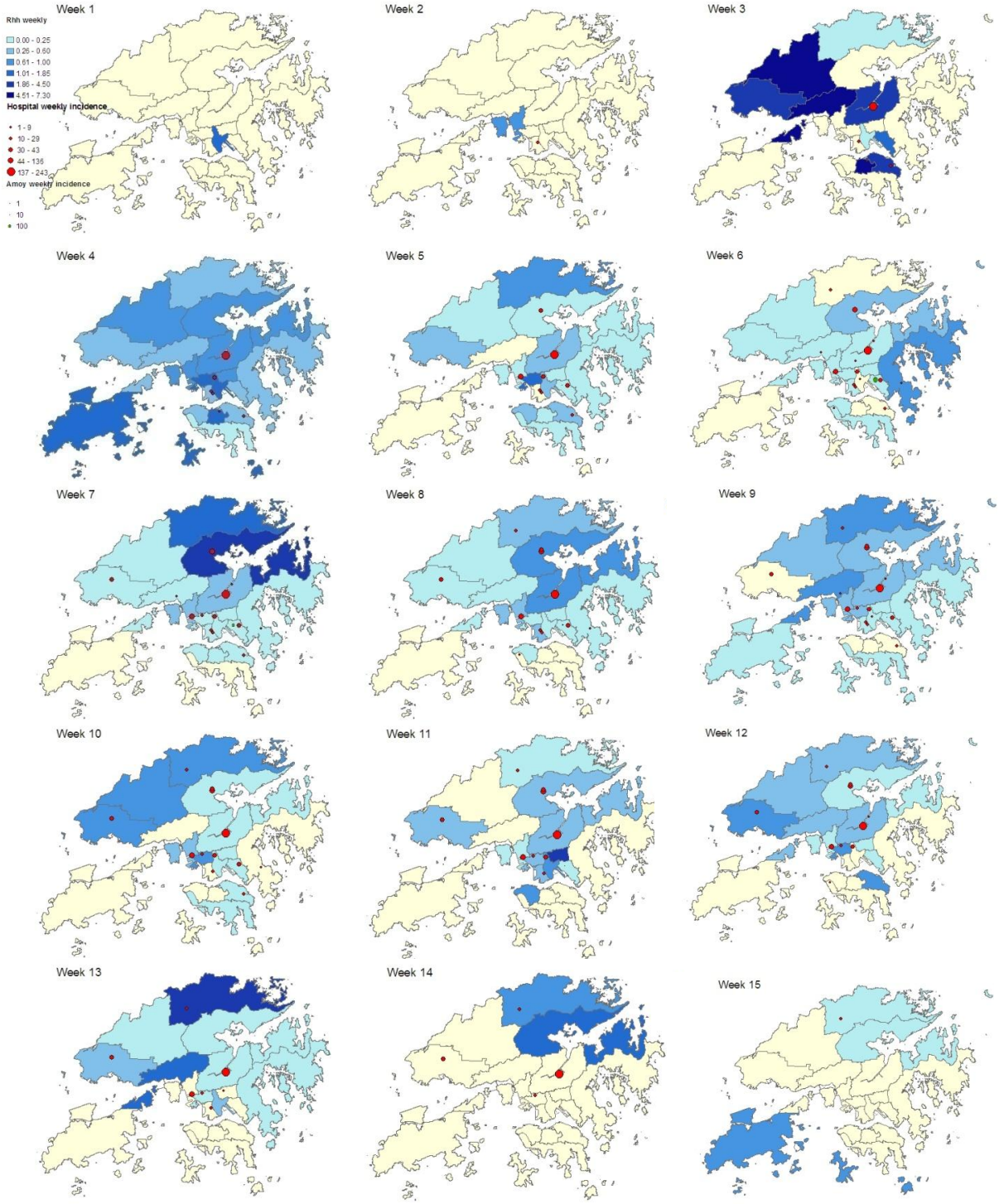


Figure 6.22 Spatiotemporal trend of the hospital-to-community reproduction number R_{hc} by district by week.

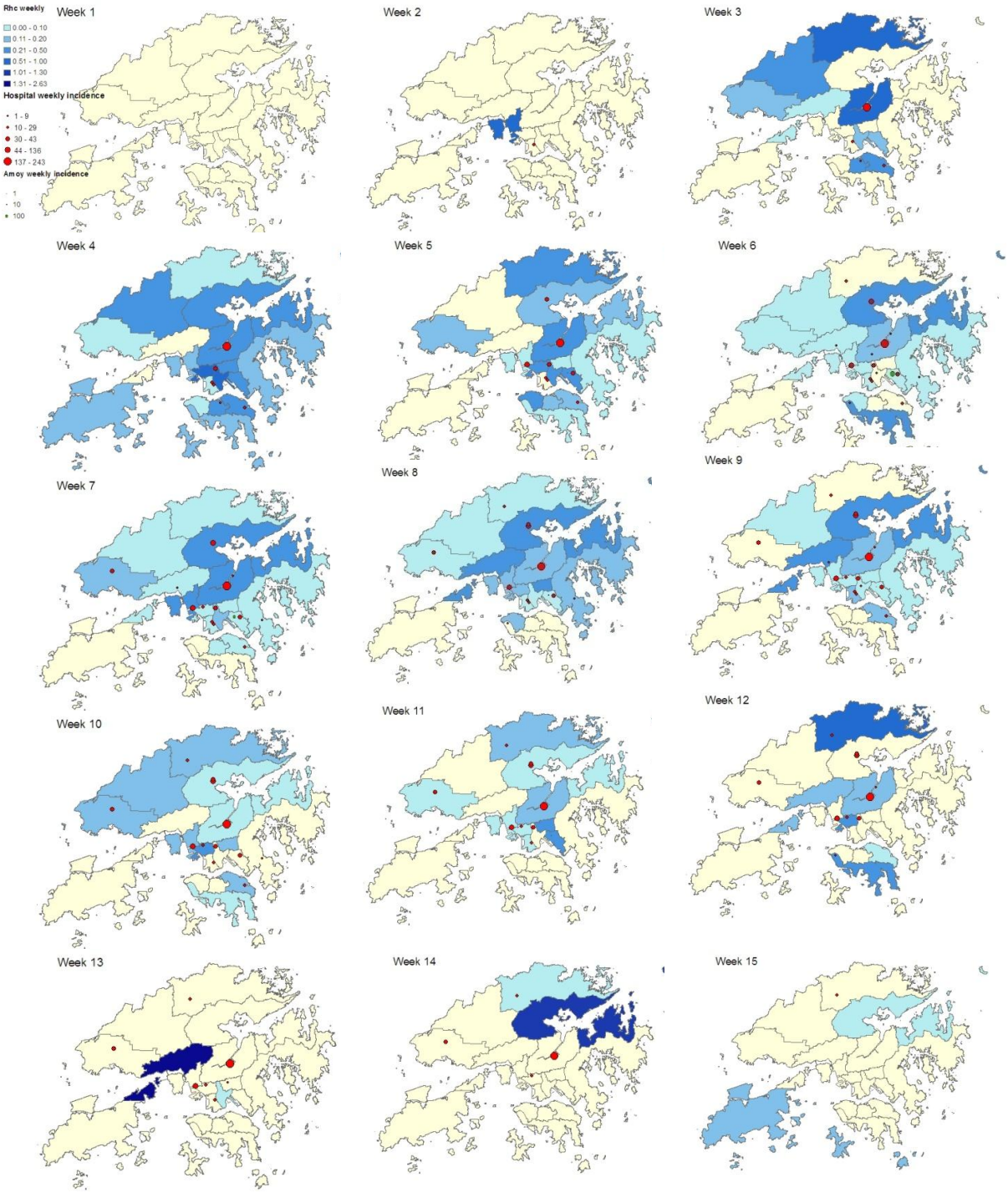


Figure 6.23 Spatiotemporal trend of the community-to-community reproduction number R_{cc} by district by week.

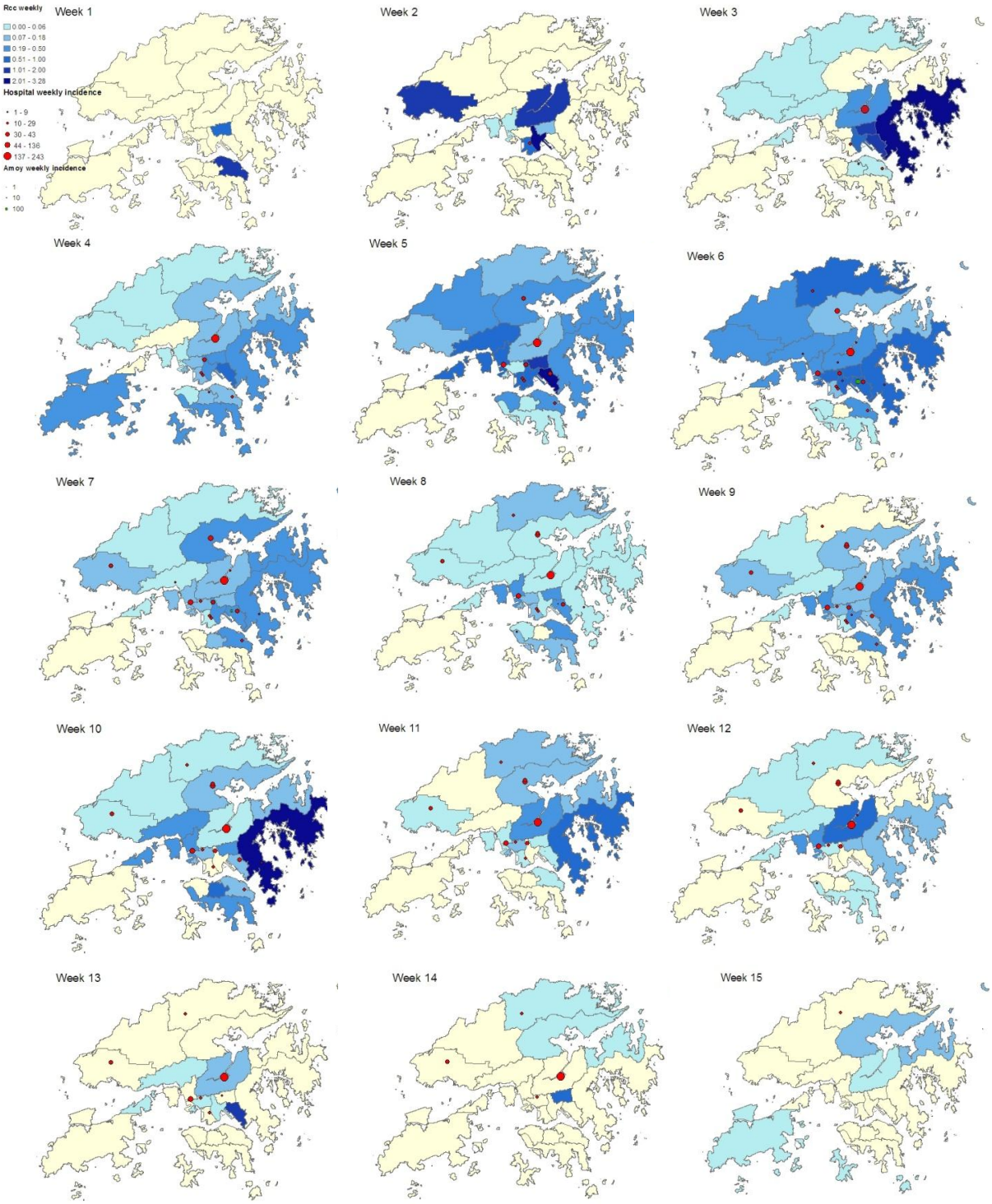


Figure 6.24 Spatiotemporal trend of the community-to-hospital reproduction number R_{ch} by district by week.

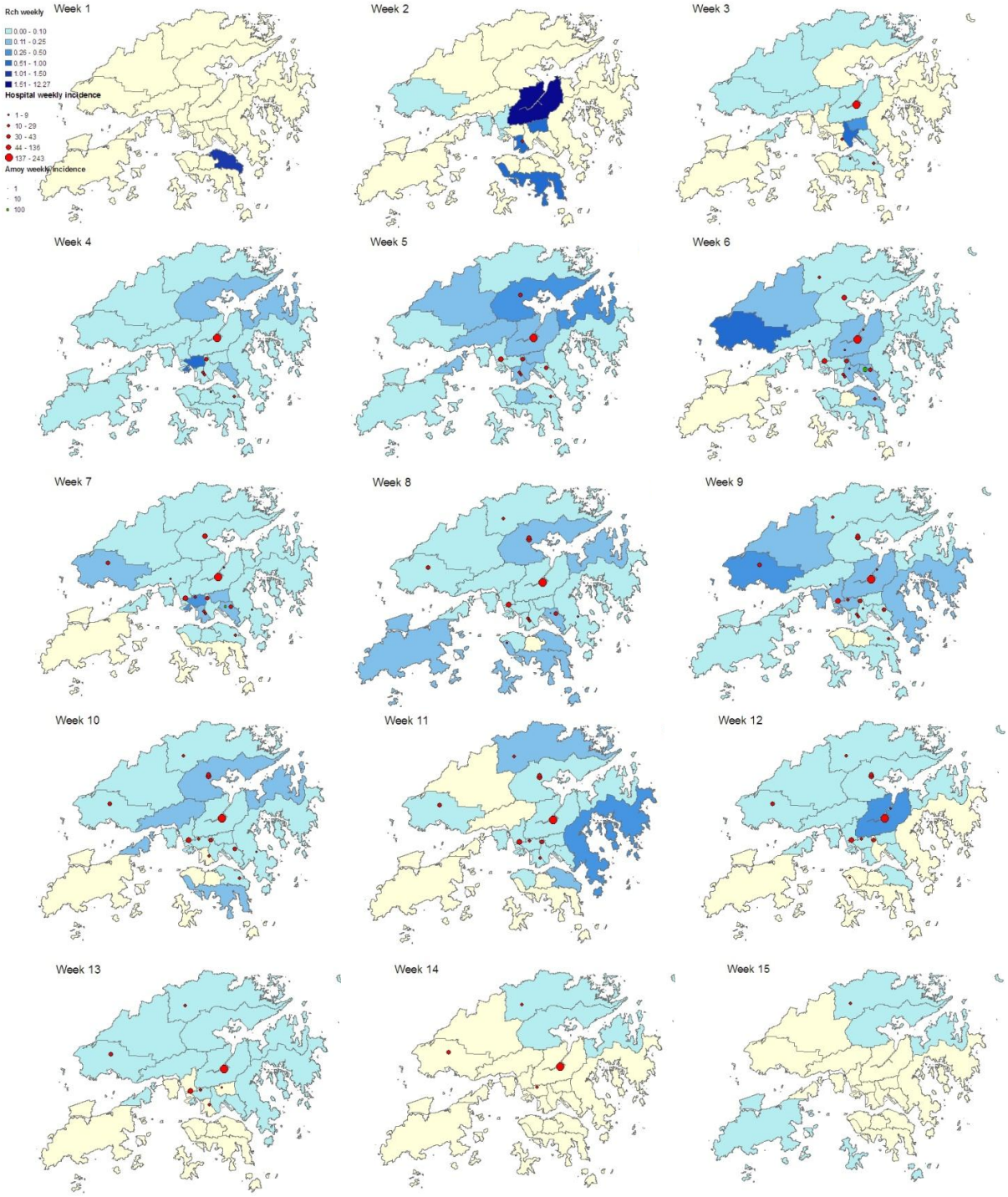


Table 6.4 shows the results of sensitivity analysis to different assumptions about the weighting of hospital-acquired infection for hospital infected cases, hospital visitor per bed ratio, the proportion of time spent in hospital by working healthcare workers, and the use of bilateral movement mixing instead of infector driven mixing. The method was found to be very robust to these assumptions.

Table 6.4 Results of sensitivity analysis on different assumptions of the parameters and movement mixing.

Difference in (alternative assumption - baseline)	**$R_H(t)$**	**$R_C(t)$**	**R(t)**	**$R_H(t)$**	**$R_C(t)$**	**R(t)**
Weighting of hospital-acquired infection for hospital infected cases (baseline σ=0.99)						
Assumption		**$\sigma=0.8$**				
25th percentile	-0.003	0.000	-0.001			
75th percentile	0.000	0.003	0.001			
hospital visitor per bed ratio (baseline $\varphi=0.8$)						
Assumption		**φ=0.5**			**φ=1**	
25th percentile	-0.002	-0.027	-0.035	-0.003	-0.027	-0.034
75th percentile	0.000	0.003	0.003	0.000	0.003	0.003
Working healthcare worker proportion of time in hospital (baseline λ=0.8)						
Assumption		**λ=0.6**			**λ=0.99**	
25th percentile	-0.007	0.000	-0.008	-0.004	0.000	-0.003
75th percentile	0.005	0.000	0.005	0.005	0.000	0.005
Bilateral movement mixing (baseline: infector driven movement)						
25th percentile	0.000	-0.015	-0.015			
75th percentile	0.000	0.007	0.009			

6.5 Discussion

The modified spatial Walling and Teunis method developed and applied here captures spatial heterogeneities to individual hospital level, and in the community to individual residential location level, supported by the source of infection and hospitalisation data. Overall the method appeared to perform well. Variation in an individual's transmissibility was captured in the individual reproduction numbers, overcoming the over-smoothing issues generally observed in the Wallinga and Teunis type method, including the partially extended method developed in last chapter. The evolution of the epidemic across transmission settings and over time were revealed through the estimates of reproduction number by transmission settings, and these estimates at the individual level were indicative of the secondary epidemiological clusters related to the cases, not only of clusters of considerable size, but also smaller clusters and not restricted to those occurred at the beginning or

end stage of the epidemic, suggesting good sensitivity of the method. Higher percentage of within district transmission was also produced with the spatial details captured in the method.

My findings confirmed the importance of transmission between hospital and community (i.e. hospital-to-community and community-to-hospital) in the outbreak. In fact, they triggered the two super-spreading events in the outbreak. This validated my previous findings (chapter 3) that some areas near PWH and AHN started as a hospital spatial cluster, and then subsequently became common clusters of hospital and community transmission. The community-to-hospital transmission was identified most prominently using this method when the secondary cases generated at the first location of SARS infection – the Metropole Hotel – were hospitalised. Hospital transmission then appeared to keep "leaking" back to the community throughout the epidemic. The extent of hospital-to-community transmission could be underestimated in this study because the method attributed the index case of the Amoy outbreak to a community-acquired case living in the Amoy district rather than a released inpatient, though they have the same symptom onset date. However, the index case of PWH outbreak was correctly identified. My findings are also consistent with Lloyd-Smith *et al*.'s hypothetical study (Lloyd-Smith, Galvani et al. 2003) that the transmission between hospital and community accounted for around one-third of total transmission, excluding the super-spreading events.

Hospital-to-hospital transmission was estimated to have been mainly sustained through healthcare workers, who had average hospital reproduction number of slightly over 1, and were also the key contributors after the index case in the PWH outbreak. Compared to inpatients who also spent a long time in the hospital, the duration and type of exposure and contact the healthcare worker made due to their job nature was estimated to have resulted in approximately twice as much transmission from this group compared to inpatients. However, their transmissibility in community was estimated to be lower than hospital visitors and released inpatients. This could due to the higher infection control awareness of the healthcare workers, their lower community contact during the outbreak, for instance, by living in hospital provided quarters (Severe Acute Respiratory Syndrome Expert Committee 2003), and their shorter time interval from symptom onset to hospital admission (Lee, Hui et al. 2003; Poutanen, Low et al. 2003; Tsang, Ho et al. 2003). Therefore, besides targeting hospital infection control at healthcare worker and reducing their mixing with the community (Lloyd-Smith, Galvani et al. 2003), my findings suggest that control measures aiming to reduce the contact between the community and hospital visitors and inpatients discharged from hospital could also be important.

Amoy cases were estimated to have reproduction numbers over 1, and a typical community infected case at around 0.8. Community transmissions were fuelled by different transmission settings (e.g. household, residential building, close contact etc.) at different times. Whilst the community transmission dynamics were more difficult to establish, as unknown contacts or clusters are not traceable, the fully extended method was able to estimate transmission links that were not revealed by existing data, and overcome the limitations in using contact tracing and epidemiological data collection to establish the epidemiological links. In particular, in super-spreading events, conventional reports tend to attribute the majority or the whole cluster of cases to the index case, ignoring the transmission by later generations of the cluster. In contrast, the method presented here tends to capture the transmission cluster or super-spreading events as a portion of cases - especially those with early onset relative to the cluster start, and to identify this subset as having high reproduction numbers.

Using the index case of a district and its source of infection to establish the underlying mechanism of between-district spread, I observed a mixture of spread by population size hierarchy and wave-like diffusion with PWH as the important the focal point. In the composition of transmission type by district, hospital-to-hospital transmission was the most dominant in majority of the districts. The districts of and around Amoy on the other hand generated mostly community-to-community transmission, which suggested the localised and diffusive spatial spread of such transmission type.

The credibility of the results in this study are not only supported by consistent findings from studying the spatial distribution and cluster in chapter 3, but also by the robustness demonstrated to alternative assumptions – including the reduction of the time a healthcare worker spent in the hospital, the increase in the chance that hospital infected cases actually acquired the infection from the community, the increase of hospital visitors who were found to generate more community transmission than in the hospital, and a bilateral spatial mixing.

There are a few limitations in the study. Hospitalisation data from private hospitals were not available, and travel history data were not captured in the method, potentially affecting the correct attribution of sources of infection. Also the initial degree of community transmission could be underestimated because secondary cases infected in the Metropole Hotel by the Hong Kong index case onset in other countries and therefore were not captured in the dataset (Shannon and Willoughby 2004). The use of a gravity model to determine the probability of contact in the community ignored the duration and type of contact made as a result of the journey. Furthermore, in the hospital setting, individuals are assumed to be homogenous despite certain types of contact

potentially being riskier. Lastly, the method is yet to establish clearly the effects of control measures implemented on the containment of the outbreak.

To conclude, the infection network inference method I developed had successfully regenerated the transmission dynamics and mechanisms of spatial spread between community and hospitals. Insights on the important drivers of transmission in different settings were attained, which could inform the formulation of infection control strategies for the future outbreak.

Chapter 7 - Discussion

The importance of capturing spatial heterogeneity in the modelling epidemics is increasingly recognised (Ferguson, Cummings et al. 2006; Colizza, Vespignani et al. 2009). However, spatially stratified data on all cases in an epidemic is rarely available. Resource limitations makes the tracking of cases difficult once epidemic size exceeds a certain point, and sub-clinical infection means the infection network observable from the epidemic curve is incomplete. Both issues affected case-based surveillance in the 2009 H1N1 pandemic in many countries (e.g. the UK and US). Data collected during the SARS outbreak in Hong Kong provides a fairly unique opportunity to study spatial transmission dynamics at the case level with a relatively complete dataset.

The SARS epidemic was characterised by considerable nosocomial transmission in most countries (World Health Organisation (WHO) 2004). In the Hong Kong outbreak, because all identified cases were hospitalised, hospitals became a high-risk context for transmission. Previous studies showed that the distribution of cases in the outbreak was generally clustered spatially, though the degree varied (e.g. the Amoy cases were more localised than the PWH cases) (Lai, Wong et al. 2004). However, considerable uncertainty remains regarding the dynamics of transmission between the hospital and community, and in particular the spatial scale over which transmission occurred. Therefore, my work focussed on investigating the spatial transmission dynamics of SARS, with a particular focus on how hospital-acquired infections drove onward spread in the community, and whether this could be attributed to the type of hospital infection, or community-related demographic or geographic properties.

In examining the spatial dynamics of the epidemic, I used different approaches that required different levels of spatiotemporal data and revealed different aspects of the transmission dynamics. The Kulldorff's scan statistic (Kulldorff 1997) requires only simple spatiotemporal data on incident cases (residential location and symptom onset date), and identified the locations and time of the spatial clusters. Under the same data requirement, the partially extended Wallinga and Teunis method probabilistically infers epidemiological linkage between cases, using a gravity model as a proxy of the probability of contact between individuals. With more detailed spatiotemporal data on cases (including which hospital they visited or were admitted to and when, and their likely source of infection), I reconstructed the dynamics of transmission between and within hospital and community. Overall, my work gives a more complete picture of the complex transmission dynamics seen in Hong Kong in 2003, and the consistency of some findings obtained with different methods strengthened the credibility of the results.

7.1 Dynamics of spatial spread and transmission risk factors

The distribution of cases during the epidemic was heterogeneous, showing significant spatiotemporal clustering. The foci of the key spatiotemporal clusters identified via use of Kulldorff's scan statistic included Amoy Gardens (which accounted for a quarter of the total cases), and the two hospitals with major outbreaks - PWH and AHN, where half of hospital infections occurred. The clusters around these hospitals included a considerable proportion of community-acquired cases. From the reconstructed infection tree derived from the spatial Wallinga and Teunis method I developed, I have argued that this community spread was due to the secondary transmission by cases who were themselves infected in hospital (as inpatients, visitors or health care workers). It was interesting to note that community transmission was more spatially localised than hospital transmission in general (in terms of the distance between residential locations of infector and infectee). Secondary community transmission in these areas eventually developed into more sustained community transmission, making these areas dominate the outbreak for weeks.

Community transmission sourced from hospital-acquired cases also occurred in other hospitals throughout the outbreak after early cases were hospitalised, and I estimated this to have also accounted for a quarter of community transmission. Also community transmission caused by hospital-acquired cases was more likely to be localised in their residential areas than those caused by typical community-acquired cases (i.e. non-Amoy cases). Inpatients who were infected when hospitalised for non-SARS conditions and discharged during the SARS incubation period and hospital visitors were more important than healthcare workers in driving this type of transmission. On average they generated more secondary cases in the community per person, though overall there were more healthcare workers infected than visitors and inpatients.

Results from gravity models suggested that residential locations of hospital visitors tend to be furthest to the hospital where they acquired the infection, while that of the inpatients were closest, suggesting a great spatial range of secondary infection could be generated by hospital-acquired cases in the community. However, community transmission caused by infectious discharged inpatients in the community and hospital visitors were more likely to be localised in their residential areas than those caused by healthcare workers. This could probably be explained by the findings from other empirical studies which showed that hospital visitors were more likely to infect their household members than healthcare workers (Ho, Sung et al. 2003; Goh, Lee et al. 2004; Lau, Lau et al. 2004; Wilson-Clark, Deeks et al. 2006), possibly due to their lower awareness of infection control procedures, whereas some healthcare workers reduced their contact in the community during the epidemic (Severe Acute Respiratory Syndrome Expert Committee 2003).

I estimated that hospital transmission caused by cases who acquired infection in the community accounted for a quarter of hospital infections overall, but this result is heavily biased by the PWH index case (infected in the Metropole Hotel). Hospital transmission was mainly sustained through hospital-acquired cases after the start of the PWH cluster. On average, community-acquired cases generated 0.22 secondary cases when they were hospitalised for SARS treatment.

Half of hospital infection was in healthcare workers based on the tracing data I used on likely source of infection. The spatial Wallinga and Teunis method suggested that healthcare workers were the most important driving force in transmission within hospitals (including for the PWH outbreak post the infections caused by the index case there). Their average hospital reproduction number was higher than 1 at the peak of the PWH, AHN and NDH outbreaks, and even excluding the PWH outbreak it was 0.8 over the epidemic - much higher than that of the inpatients who also had extended stays in the hospital. Healthcare workers were estimated to have generated more than half of all hospital infections (in other healthcare workers, inpatients and hospital visitors). Their relatively high infectivity and susceptibility was likely to be due to their high level of exposure, and more frequent and riskier contact with others due to their job. Other work has shown that the bed location of inpatients in a hospital ward relative to infectious inpatients was a risk factor of being infected by other inpatients, if the possibility that they were infected by healthcare workers was ignored (Yu, Wong et al. 2005).

Transmission within the community occurred in a range of different settings associated with a variety of spatial scales - from localised (e.g. household transmission and outbreak in residential buildings) to more dispersed and random (e.g. social contact, unidentifiable community source). Excluding the super-spreading events in Amoy Gardens, transmission in the community was dominated by different transmission settings at different stage of the epidemic. On average, a typical community-acquired case was estimated to generate 0.55 secondary infections in the community, and an Amoy case 0.91.

Modelling the size of outbreaks with a hurdle negative binomial model suggested that there were two intrinsic risk factors of community transmission for an area – population structure (i.e. age distribution), and distance to the hospital that treated the cases. The latter was also supported by the findings that key outbreak hospitals also triggered formation of community spatial clusters in the surrounding area, and by the fully extended spatial Wallinga and Teunis method which suggested one-third of secondary transmission in the community generated by hospital-acquired cases was highly localised.

Previous work has highlighted a range of possible patterns to spatiotemporal disease spread, including diffusive (generating clear spatial waves), population hierarchy based (city to city then spreading to local commuter belts then rural areas), or a mixture of both (Cliff, Haggett et al. 1986). My findings using the reconstructed infection tree to identify the most probable infector of the first case of each area suggested that hospitals may have acted as hubs for a quasi-hierarchical pattern of spatial spread. PWH was the initial focal point of spatial spread to uninfected areas in the first few weeks of the epidemic, and caused by approximately the first quarter of the cases, with other hospital outbreaks seeding local community outbreaks later. Areas with high population size or densities, or adjacent to an infected area tend to be infected earlier.

My analysis estimated the relative contribution of different transmission settings at different stages of the epidemic. The initial cases of the outbreak (mainly imported from outside Hong Kong) exhibited higher transmissibility in the community, partly due to their late hospitalisation compared with cases identified after SARS had been recognised. The second generation of cases (who were mostly infected in the community) generated large numbers of hospital infection, most notably in a super-spreading event in PWH of more than 100 cases, in which healthcare workers were estimated to contribute substantially in further transmission besides the index cases of PWH. Once infection control measures in PWH (including the transfer of non-SARS patients to other hospitals, a no visitor policy in SARS wards, and other measures targeting healthcare workers) were put in place, transmission within hospital was much reduced. However, infection was brought by the hospital-acquired cases back to the community (mainly causing household transmission), in particular through inpatients released during the incubation period and hospital visitors. One of such incident occurred in Amoy Gardens, which triggered a disease cluster of more than 400 cases, first through environmental contamination and then possibly via secondary transmission to household members and close contacts subsequently. This outbreak eventually spread to the housing estates nearby and caused small local outbreaks there. As community-wide control measures were introduced, transmission in the community subsided, and the effective reproduction number fell to below 1. After that point there were a few outbreaks in hospitals of smaller scale, with healthcare workers being estimated to be the key driver of these outbreaks, albeit briefly.

7.2 Implications for control measures

The rather theoretical study by Lloyd-Smith *et al*. (Lloyd-Smith, Galvani et al. 2003) also identified the importance of healthcare workers in driving the epidemic, though this could have been exaggerated by that study's assumptions that hospital was a single physical unit, and no transmission could have taken place between inpatients and hospital visitors. They suggested

control measures reducing infection of healthcare workers would be most effective, and if that was not adequate to contain the outbreak, mixing of the healthcare workers with the community should be reduced. Correlating my results with the timing of control measures implemented during the epidemic, I find that hospital infection remained the key driver of the epidemic even after the transmission potential of SARS in hospital became clear in the PWH outbreak. Possibly this could be partially explained by the less stringent infection control implemented in hospitals which were thought to be infection free (McDonald, Simor et al. 2004) when infection was in fact latent in the hospital. In addition, compliance of healthcare workers to control measures (e.g. wearing personal protection equipment) was also imperfect (Lau, Fung et al. 2004).

Restricted hospital visitor policies were imposed for a period of time during the outbreak (Severe Acute Respiratory Syndrome Expert Committee 2003). Given the importance of hospital visitors in bringing transmission from hospital to the community, such a policy could have been implemented for a longer period of time, and introduced immediately once hospital transmission is detected. Quarantine could be also considered for those who had recently visited or been discharged from hospitals where infection was detected.

Prior knowledge on the potential high-risk spatial clusters could provide valuable early insights on the transmission dynamics and characteristics of an epidemic (Mammen, Pimgate et al. 2008; Riley 2008). My findings suggest the geographic and demographic properties of areas with intrinsically higher risk of infection. In a general uncontrolled epidemic, the final attack rate in areas with high transmission risk as a result of their demographic characteristics or travel behaviour of their residents is insensitive to the seeding location of the epidemic (Mao and Bian 2010). However for a truncated epidemic such as SARS (where only a small fraction of people were eventually infected), random chance plays a more major role in determining which areas are worst affected – dependent on where the first cases are hospitalised, and at what stage of the epidemic controls start to reduce transmission. Once hospital cases are first detected in a high risk area, community surveillance should be heightened in that area, and community control measures considered.

These suggestions are applicable for any severe emerging respiratory infectious disease which routinely requires hospitalisation of cases and thus makes hospitals a high risk context for transmission. In the H1N1 influenza pandemic in Hong Kong in 2009, around 18% of the confirmed cases were hospitalised for the first 4 months during the pandemic as a precautionary measure (Cowling, Lau et al. 2010), causing sustained transmission in most of the hospitals treating the cases (The University of Hong Kong 2010), despite infection control measures being in place.

The real-time tracking of spatiotemporal clusters of cases of an emerging disease using Kulldorff's prospective scan statistic might be useful to inform health officials of areas which might warrant closer monitoring and more intense application of control measures. Because epidemiological clusters often exhibit spatially clustering, the method has the potential to pick up important transmission events early. Control measures therefore can be formulated and implemented promptly in these areas, and this could make a difference to the time-scale and spread of the epidemic.

7.3 Spatially explicit infection tree methods

The interaction between hospital and community transmission and its effects on the spatial spread of the epidemic can be studied by infection tree inference methods in which hospital and community settings are explicitly incorporated. Current methods of infection tree inference that do not require population data (Wallinga and Teunis 2004; Fraser 2007) assume people with the same symptom onset date are identical and that mixing between individuals is homogenous, causing the reproduction number estimates to be smoothed over time. I chose to extend the Wallinga and Teunis method spatially, as it is formulated on individual case rather than on daily incidence basis as in Fraser's method. The fact that SARS cases tended to only become infectious upon symptom onset makes the Wallinga and Teunis method particularly suitable to study the SARS epidemic, because otherwise the imputation of who infected who would be more complicated due to the added impact of uncertainty in the infection time of the infector as well as the infectee (Chis Ster, Singh et al. 2009). The negligible level of sub-clinical infection (Leung, Chung et al. 2004) seen in all SARS outbreaks in 2003 also implies that the case data can be consider as complete upon which to reconstruct the infection tree.

The first spatially extended method I developed captured the spatial and individual heterogeneity that affect one's likelihood to have infected others. The method uses an infectivity profile to describe the development of infectiousness of an infected individual over time. Combined with an assumed constant incubation period and the use of a gravity model to describe the relative probability of contact as a function of distance between residential locations and their population size, this function is critical in inferring the probability that one case infected another. Therefore I modified the original Wallinga and Teunis method from case-to-case based to infection-to-infection based. Hospital and community transmission were not separated in this model, however.

Gravity models are widely used to model trip patterns in human geography and spatial interaction modelling. The gravity models I derived capture the different flow patterns to the key commercial

and leisure areas compared to other areas with an adjusted scaling component of the population size that significantly improved the model goodness-of-fit. Previous studies (Xia, Bjornstad et al. 2004; Viboud, Bjornstad et al. 2006; Balcan, Colizza et al. 2009) that incorporated gravity coupling in epidemic models used a variety of spatial mixing assumptions without verifying their appropriateness. I compared the spatial patterns of mixing generated by models assuming susceptible move, infectors move, both move to the other's residential location (bilateral movement) or both can meet in any location (multilateral movement), and found that assuming infectors moved to the residence of suceptibles gave the most reasonable results. This was because that version of the gravity model conditioned on the number of inward journeys to each area, and this form of conditioning gave spatial contact patterns that mostly closely reproduced the spatial distribution of the cases with respect to the individual reproduction number compared to the observed distribution and the characteristics of the epidemic. I also found that travel data (e.g. work flow) that do not include people who did not make that type of trip and remained in their residing area (e.g. those retired) could underestimate the within-district contact. Adjusting the trip data to capture this group of people as within their residing area could resolve the problem.

My model produced more realistic estimates of aggregate time-dependent estimates of the effective reproduction number, and of the reproduction number of individual cases, and was able to produce higher individual reproduction number estimates for the index cases of known super-spreading events. Therefore, my method shows that even when limited data are available (e.g. case onset date and residential location), person-to-person variation in transmissibility can be more accurately estimated by a spatially stratified model.

Hospitals are a very different transmission setting from the wider community. Hospitals are spatially confined environments containing limited numbers of people, and contact patterns between individuals vary according to the subgroup one belongs to (healthcare worker, inpatient and visitor). To distinguish transmission in hospitals and the community and the interaction between these settings, I further modified the spatial Wallinga and Teunis model by explicitly including individual hospitals as distinct mixing groups, making use of data collected in the outbreak on which setting particular cases were likely to have been infected in, and which hospitals cases visited, were admitted to or worked in.

The model used the hospitalisation data and the potential source of infection data to infer the location of the cases over time (i.e. which hospital or community area they were in each day). Mixing within individual hospital was assumed homogeneous, and in the community individuals interacted with the same gravity-model derived probability as used in the previous model. The resulting

analysis is the most detailed spatiotemporal study of the Hong Kong SARS epidemic thus far, and is able to reconstruct the spatiotemporal transmission dynamics at a fine-grained level (setting and location).

The method performed well in estimating how transmission changed over time by transmission setting, and in imputing the epidemiological links between cases in large clusters with good sensitivity. It overcomes the over-smoothing problem of non-spatial infection tree methods, and is robust to different parameter assumptions on the spatial mixing, hospital population size and on the probability of alternative infection source having caused a case other than the probable source of infection identified in the input data.

7.4 Beyond infection trees - modelling the susceptible population

The spatially explicit infection tree methods I developed are conditional on the observed case data, the hospitalisation data and the source of infection data. The partial likelihood derived with these constraints simplifies the mathematical formulation and lessens the computational burden of the method. While the use of these data imply a more accurate infection tree inferred for the epidemic, the results are specific to the cases of the observed epidemic and their characteristics, and the conditioning process restricts the generalisation of the results beyond the observed cases and epidemic. Put simply, such methods say little about people who were not infected (the vast bulk of the population) and the risks of infection in different groups. To overcome this, a full epidemic model using the population data and using a full likelihood for the observed epidemic needs to be developed (Chis Ster, Singh et al. 2009; Griffin, Garske et al. 2010), as discussed in more detail later in this section.

My assumption of a constant incubation period also simplified the mathematical formulation of the spatial models. The model could be further extended to capture variability in incubation period, however. With non-constant incubation period distribution, an incubation period could be sampled for each case independently by bootstrapping, thus creating a simulated set of incubation period data for all cases. With replication this generates sets of incubation period samples for the whole set of cases. From each incubation period simulated dataset, the infection times of each case can be imputed, and the existing model rerun to derive an imputed infection tree. Analysis of the results would require examination of the ensemble of infection trees generated.

The spatial resolution of the area units used in my analysis was restricted by the availability of complete trip data for the gravity model. Trip data for finer spatial scales is partially available, and it

is possible that constrained gravity models could be fitted on this data with some modifications of the likelihood function for the parameter estimation.

Real-time monitoring of the epidemic development, facilitates evidence-based control measures formation (Ong, Chen et al. ; Fraser, Donnelly et al. 2009; Cowling, Lau et al. 2010). My spatial model could be modified for real-time application during an epidemic to track reproduction number estimates by area and transmission settings by taking into account the censored cases. A previous study developed such an extension on the Wallinga and Teunis method (Cauchemez, Boelle et al. 2006) by using the probability distribution of the serial interval to predict the number of censored cases during an epidemic over time. For my model, this prediction would need to be stratified by spatial location and transmission setting. Assuming the spatial trend to be independent of the temporal trend, the spatial distribution could be predicted using a mixture of both the distribution of the existing cases, and the trip patterns of the general population. A real-time case dataset adjusted for censored cases using such methods might also be used for the analysis making use of Kulldorff's prospective scan statistic, because currently cases in their incubation period are not captured in the outbreak dataset for cluster detection until their symptom onset, and this indirectly lowers the power to detect clusters early. However, it is likely that confidence bounds around incidence predictions correcting for censoring will become unusually wide if too fine a stratification was used (since case numbers might be very low in single areas or hospitals).

It is possible to apply or further modify spatially and socially stratified Wallinga and Teunis methods for other pathogens and alternative transmission settings (other than the hospital). The models I developed are able to capture cases who belong to more than one spatial or social groups at the same time (for instance, school and community), and allows different mechanisms to assign the probability of contact in the spatial groups. However, in the absence of a complete likelihood or contact tracing data, the relative weighting of different settings to transmission would need to be assumed. The location by time data (analogous to the hospitalisation data) might be simulated if actual data is not available.

As mentioned previously section, the models I developed here condition on the observed cases and their properties, and do not model the susceptible population. Hence they can say little about susceptibility risk factors of individuals or spatial locations, nor properly assess the effectiveness of control measures. It is possible to address these issues with a full, spatially explicit individually based infection model designed for inference, for instance using a modified version of models used for a recent reanalysis of the 2001 Foot and Mouth epidemic in Britain (Chis Ster, Singh et al. 2009). As such models have to be parametric, some aspects of full-likelihood models tend to have to be

simpler than models that solely impute infection trees; most notably, it is usually not possibly to allow transmission coefficients to vary arbitrarily over time. However, models including both susceptible and infected populations can have arbitrarily complex formulations of the likelihood and yet be guaranteed to produce unbiased estimates, while partial likelihood models are only guaranteed to produce unbiased estimates of the full infection tree under a restrictive set of constraints that are unlikely to be fulfilled in reality (Griffin, Garske et al. 2010); for instance, that all cases have the same infectiousness.

In a full epidemic model, a hazard of infection is used to captures the risk of infection based on the location and timing of cases and the transmission setting. Individuals when in the community would most simply be coupled using a gravity model, and in hospital homogeneous mixing would be the simplest assumption, as in the spatial models I developed here. Super-spreading events may require special treatment (Riley, Fraser et al. 2003; Cori, Boelle et al. 2009). Infection time imputation and other parameter estimates would require MCMC methods given the complexity of the model structure. Given the size of the SARS outbreak in Hong Kong, such an analysis would be feasible given current computing resources. Infection trees could be inferred the under such a model, and the predicted spatial distribution of the cases can be compared to the observed patterns in addition to the conventional model comparison methods.

References

Anderson, R. M., C. Fraser, et al. (2004). "Epidemiology, transmission dynamics and control of SARS: the 2002-2003 epidemic." Philosophical Transactions of the Royal Society of London Series B-Biological Sciences **359**(1447): 1091-1105.

Anderson, R. M. and R. M. May (1991). Infectious Diseases of Humans: Dynamics and Control. Oxford, Oxford Science Publications.

Balcan, D., V. Colizza, et al. (2009). "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): 21484-21489.

Balcan, D., H. Hu, et al. (2009). "Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility." BMC Medicine **7**(1): 45.

Bauch, C. T., J. O. Lloyd-Smith, et al. (2005). "Dynamically modeling SARS and other newly emerging respiratory illnesses - Past, present, and future." Epidemiology **16**(6): 791-801.

Bell, D. M. and W. H. O. W. G. P. In (2004). "Public health interventions and SARS spread 2003." Emerging Infectious Diseases **10**(11): 1900-1906.

Besag, J. and J. Newell (1991). "The Detection of Clusters in Rare Diseases." Journal of the Royal Statistical Society Series a-Statistics in Society **154**: 143-155.

Bharti, N., Y. C. Xia, et al. (2008). "Measles on the Edge: Coastal Heterogeneities and Infection Dynamics." PLoS ONE **3**(4).

Brockmann, D., L. Hufnagel, et al. (2006). "The scaling laws of human travel." Nature **439**(7075): 462-465.

Carey, H. C. (1858). Principles of Social Science. Philadelphia, Pennsylvania., J. Lippincott.

Cauchemez, S., P. Y. Boelle, et al. (2006). "Estimating in real time the efficacy of measures to control emerging communicable diseases." American Journal of Epidemiology **164**(6): 591-597.

Cauchemez, S., P. Y. Boelle, et al. (2006). "Real-time estimates in early detection of SARS." Emerging Infectious Diseases **12**(1): 110-113.

Census and Statistics Department Hong Kong Special Administrative Region, P. s. R. o. C. (2001). "2001 Population Census."

Chen, M. I. C., Y. S. Leo, et al. (2006). "The outbreak of SARS at Tan Tock Seng hospital - Relating epidemiology to control." Annals Academy of Medicine Singapore **35**(5): 317-325.

Chis Ster, I. and N. M. Ferguson (2007). "Transmission Parameters of the 2001 Foot and Mouth Epidemic in Great Britain." PLoS ONE **2**(6): e502.

Chis Ster, I., B. K. Singh, et al. (2009). "Epidemiological inference for partially observed epidemics: The example of the 2001 foot and mouth epidemic in Great Britain." Epidemics **1**(1): 21-34.

Chowell, G., L. M. A. Bettencourt, et al. (2008). "The 1918-1919 influenza pandemic in England and Wales: spatial patterns in transmissibility and mortality impact." Proceedings of the Royal Society B-Biological Sciences **275**(1634): 501-509.

Ciofi degli Atti, M. L., S. Merler, et al. (2008). "Mitigation Measures for Pandemic Influenza in Italy: An Individual Based Model Considering Different Scenarios." PLoS ONE **3**(3): e1790.

Clark, P. J. and F. C. Evans (1954). "Distance to nearest neighbor as a measure of spatial relationships in populations." Ecology **35**: 445-453.

Cliff, A. D., P. Haggett, et al. (1986). Spatial Aspects of Influenza Epidemics. London, Pion Limited.

Cliff, A. D., P. Haggett, et al. (2008). "An exploratory method for estimating the changing speed of epidemic waves from historical data." International Journal of Epidemiology **37**(1): 106-112.

Colizza, V., A. Vespignani, et al. (2009). "Estimate of Novel Influenza A/H1N1 cases in Mexico at the early stage of the pandemic with a spatially structured epidemic model." PLoS Curr **1**: RRN1129.

Cori, A., P. Y. Boelle, et al. (2009). "Temporal Variability and Social Heterogeneity in Disease Transmission: The Case of SARS in Hong Kong." Plos Computational Biology **5**(8): -.

Cowling, B. J., M. S. Lau, et al. (2010). "The effective reproduction number of pandemic influenza: prospective estimation." Epidemiology **21**(6): 842-846.
Cowling, B. J., M. P. Muller, et al. (2007). "Alternative methods of estimating an incubation distribution - Examples from severe acute respiratory syndrome." Epidemiology **18**(2): 253-259.
Cuzick, J. and R. Edwards (1990). "Spatial Clustering for Inhomogeneous Populations." Journal of the Royal Statistical Society Series B-Methodological **52**(1): 73-104.
Danon, L., T. House, et al. (2009). "The role of routine versus random movements on the spread of disease in Great Britain." Epidemics **1**(4): 250-258.
Donnelly, C. A., A. C. Ghani, et al. (2003). "Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong." Lancet **361**(9371): 1761-1766.
Duczmal, L. and R. Assuncao (2005). "A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters." Computational Statistics and Data Analysis **15**(2): 1-15.
Duczmal, L., M. Kulldorff, et al. (2006). "Evaluation of spatial scan statistics for irregularly shaped clusters." Journal of Computational and Graphical Statistics **15**(2): 428-442.
Dwosh, H. A., H. H. L. Hong, et al. (2003). "Identification and containment of an outbreak of SARS in a community hospital." Canadian Medical Association Journal **168**(11): 1415-1420.
Dye, C. and N. Gay (2003). "Modeling the SARS epidemic." Science **300**(5627): 1884-1885.
Eggo, R. M., S. Cauchemez, et al. (2010). "Spatial dynamics of the 1918 influenza pandemic in England, Wales and the United States." Journal of the Royal Society Interface.
Erlander, S. and N. F. Stewart (1990). The Gravity Model in Transportation Analysis - Theory and Extensions. Utrecht, The Netherlands, VSP.
Farewell, V. T., A. M. Herzberg, et al. (2005). "SARS incubation and quarantine times: when is an exposed individual known to be disease free?" Statistics in Medicine **24**(22): 3431-3445.
Ferguson, N. M., D. A. T. Cummings, et al. (2005). "Strategies for containing an emerging influenza pandemic in Southeast Asia." Nature **437**(7056): 209-214.
Ferguson, N. M., D. A. T. Cummings, et al. (2006). "Strategies for mitigating an influenza pandemic." Nature **442**(7101): 448-452.
Ferguson, N. M., C. A. Donnelly, et al. (2001). "Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain (vol 413, pg 542, 2001)." Nature **414**(6861): 329-329.
Fine, P. E. M. (2003). "The interval between successive cases of an infectious disease." American Journal of Epidemiology **158**(11): 1039-1047.
Finkenstadt, B. and B. Grenfell (1998). "Empirical determinants of measles metapopulation dynamics in England and Wales." Proceedings of the Royal Society of London Series B-Biological Sciences **265**(1392): 211-220.
Finkenstadt, B., M. Keeling, et al. (1998). "Patterns of density dependence in measles dynamics." Proceedings of the Royal Society of London Series B-Biological Sciences **265**(1398): 753-762.
Fotheringham, A. S. and M. E. O'Kelly (1988). Spatial interaction models: formulations and applications. Dordrecht, Kluwer Academic.
Fraser, C. (2007). "Estimating individual and household reproduction numbers in an emerging epidemic." PLoS ONE **2**(1): e758.
Fraser, C., C. Donnelly, et al. (2009). "Pandemic potential of a strain of influenza A(H1N1): early findings." Science **324**: 1557 - 1561.
Fraser, C., S. Riley, et al. (2004). "Factors that make an infectious disease outbreak controllable." Proceedings of the National Academy of Sciences of the United States of America **101**(16): 6146-6151.
Fylstra, D., L. Lasdon, et al. (1998). "Design and use of the Microsoft Excel Solver." Interfaces **28**(5): 29-55.
Garske, T., P. Clarke, et al. (2007). "The Transmissibility of Highly Pathogenic Avian Influenza in Commercial Poultry in Industrialised Countries." PLoS ONE **2**(4): e349.

Geary, R. C. (1954). "The contiguity ratio and statistical mapping." The Incorporated Statistician **5**: 115-145.

Ghani, A. C., M. Baguelin, et al. (2009). "The Early Transmission Dynamics of H1N1pdm Influenza in the United Kingdom." PLoS Curr Influenza **RRN1130**.

Gibson, G. J. (1997). "Markov chain Monte Carlo methods for fitting spatiotemporal stochastic models in plant epidemiology." Applied Statistics-Journal of the Royal Statistical Society Series C **46**(2): 215-233.

Goh, D. L. M., B. W. Lee, et al. (2004). "Secondary household transmission of SARS, Singapore." Emerging Infectious Diseases **10**(2): 232-234.

Gonzalez, M. C., C. A. Hidalgo, et al. (2008). "Understanding individual human mobility patterns." Nature **453**(7196): 779-782.

Grenfell, B. T. and B. M. Bolker (1998). "Cities and villages: infection hierarchies in a measles metapopulation." Ecology Letters **1**(1): 63-70.

Griffin, J. T., T. Garske, et al. (2010). "Joint estimation of the basic reproduction number and generation time parameters for infectious disease outbreaks." Biostatistics.

Groneberg, D. A., S. M. Poutonen, et al. (2005). "Treatment and vaccines for severe acute respiratory syndrome." Lancet Infectious Diseases **5**(3): 147-155.

Guan, Y., B. J. Zheng, et al. (2003). "Isolation and characterization of viruses related to the SARS coronavirus from animals in Southern China." Science **302**(5643): 276-278.

Haydon, D. T., M. Chase-Topping, et al. (2003). "The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak." Proceedings of the Royal Society of London Series B-Biological Sciences **270**(1511): 121-127.

Haynes, K. E. and A. S. Fotheringham (1984). Gravity and Spatial Interaction Models. USA, Sage.

Hill, E. G., L. Ding, et al. (2000). "A comparison of three tests to detect general clustering of a rare disease in Santa Clara County, California." Statistics in Medicine **19**(10): 1363-1378.

Ho, A. S., J. J. Y. Sung, et al. (2003). "An outbreak of severe acute respiratory syndrome among hospital workers in a community hospital in Hong Kong." Annals of Internal Medicine **139**(7): 564-567.

Hospital Authority, H. (2003). Hospital Authority Statistical Report 2002-2003**:** http://www.ha.org.hk/hesd/nsapi/?MIval=ha_visitor_index&intro=ha%5fview%5ftemplate%26group%23dIFN%26Area%23dPBL.

Huff, D. L. (1963). "A probabilistic analysis of shopping center trade areas." Land Economics **39**: 81-90.

Hui, E. C. M. and M. C. M. Lam (2005). "A study of commuting patterns of new town residents in Hong Kong." Habitat International **29**(3): 421-437.

Keeling, M. and P. Rohani (2002). "Estimating spatial coupling in epidemiological systems: a mechanistic approach." Ecol Lett **5**: 20 - 29.

Keeling, M. J. and P. Rohani (2007). Modeling infectious diseases in humans and animals Princeton University Press.

Keeling, M. J., M. E. J. Woolhouse, et al. (2001). "Dynamics of the 2001 UK foot and mouth epidemic: Stochastic dispersal in a heterogeneous landscape." Science **294**(5543): 813-817.

Kenah, E., M. Lipsitch, et al. (2008). "Generation interval contraction and epidemic data analysis." Mathematical Biosciences **213**(1): 71-79.

Kermack, W. O. and A. G. McKendrick (1927). "A contribution to the mathematical theory of epidemics." Proc Royal Soc London A **115**: 700-721.

Knox, E. G., Ed. (1989). Detection of clusters. Methodology of enquiries into disease clustering. London, Small Area Health Statistics Unit.

Knudsen, D. C. and A. S. Fotheringham (1986). "Matrix Comparison, Goodness-of-Fit, and Spatial Interaction Modeling." International Regional Science Review **10**(2): 127-147.

Kuk, A. Y. C. and S. Ma (2005). "The estimation of SARS incubation distribution from serial interval data using a convolution likelihood." Statistics in Medicine **24**(16): 2525-2537.

Kulldorff, M. (1997). "A spatial scan statistic." Communications in Statistics-Theory and Methods **26**(6): 1481-1496.
Kulldorff, M. (2001). "Prospective time periodic geographical disease surveillance using a scan statistic." Journal of the Royal Statistical Society Series a-Statistics in Society **164**: 61-72.
Kulldorff, M. (2006). "Tests of spatial randomness adjusted for an inhomogeneity: A general framework." Journal of the American Statistical Association **101**(475): 1289-1305.
Kulldorff, M., L. Huang, et al. (2006). "An elliptic spatial scan statistic." Statistics in Medicine **25**(22): 3929-3943.
Kulldorff, M., T. Tango, et al. (2003). "Power comparisons for disease clustering tests." Computational Statistics & Data Analysis **42**(4): 665-684.
Kwok, K. O., G. M. Leung, et al. (2007). "Using models to identify routes of nosocomial infection: a large hospital outbreak of SARS in Hong Kong." Proceedings of the Royal Society B-Biological Sciences **274**(1610): 611-618.
Lai, P. C., C. M. Wong, et al. (2004). "Understanding the spatial clustering of severe acute respiratory syndrome (SARS) in Hong kong." Environmental Health Perspectives **112**(15): 1550-1556.
Lai, T. S., T. Keung Ng, et al. (2005). "Low prevalence of subclinical severe acute respiratory syndrome-associated coronavirus infection among hospital healthcare workers in Hong Kong." Scand J Infect Dis **37**(6-7): 500-503.
Lambert, D. (1992). "Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing." Technometrics **34**(1): 1-14.
Lau, J. T. F., K. S. Fung, et al. (2004). "SARS transmission among hospital workers in Hong Kong." Emerging Infectious Diseases **10**(2): 280-286.
Lau, J. T. F., M. Lau, et al. (2004). "Probable secondary infections in households of SARS patients in Hong Kong." Emerging Infectious Diseases **10**(2): 235-243.
Lee, N., D. Hui, et al. (2003). "A major outbreak of severe acute respiratory syndrome in Hong Kong." New England Journal of Medicine **348**(20): 1986-1994.
Lee, P. P., W. H. Wong, et al. (2006). "Risk-stratified seroprevalence of severe acute respiratory syndrome coronavirus among children in Hong Kong." Pediatrics **117**(6): e1156-1162.
Lee, S. S. and N. S. Wong (2010). "Characterizing the initial diffusion pattern of pandemic (H1N1) 2009 using surveillance data." PLoS Curr **2**: RRN1151.
Leo, Y. S., M. Chen, et al. (2003). "Severe Acute Respiratory Syndrome - Singapore, 2003." MMWR **52**(18): 405-411.
Leung, G. M., P. H. Chung, et al. (2004). "SARS-CoV antibody prevalence in all Hong Kong patient contacts." Emerging Infectious Diseases **10**(9): 1653-1656.
Leung, G. M., A. J. Hedley, et al. (2004). "The epidemiology of severe acute respiratory syndrome in the 2003 Hong Kong epidemic: An analysis of all 1755 patients." Annals of Internal Medicine **141**(9): 662-673.
Leung, G. M., W. W. Lim, et al. (2006). "Seroprevalence of IgG antibodies to SARS-coronavirus in asymptomatic or subclinical population groups." Epidemiology and Infection **134**(2): 211-221.
Lipsitch, M., T. Cohen, et al. (2003). "Transmission dynamics and control of severe acute respiratory syndrome." Science **300**(5627): 1966-1970.
Lloyd-Smith, J. O., A. P. Galvani, et al. (2003). "Curtailing transmission of severe acute respiratory syndrome within a community and its hospital." Proceedings of the Royal Society of London Series B-Biological Sciences **270**(1528): 1979-1989.
Mammen, M. P., C. Pimgate, et al. (2008). "Spatial and temporal clustering of dengue virus transmission in Thai villages." PLoS Med **5**(11): e205.
Mao, L. and L. Bian (2010). "Spatial-temporal transmission of influenza and its health risks in an urbanized area." Computers, Environment and Urban Systems **34**: 204-215.
McBryde, E. S., G. Gibson, et al. (2006). "Bayesian modelling of an epidemic of severe acute respiratory syndrome." Bulletin of Mathematical Biology **68**(4): 889-917.

McDonald, L. C., A. E. Simor, et al. (2004). "SARS in healthcare facilities, Toronto and Taiwan." Emerging Infectious Diseases **10**(5): 777-781.
McKinney, K. R., Y. Y. Gong, et al. (2006). "Environmental transmission of SARS at Amoy Gardens." Journal of Environmental Health **68**(9): 26-30.
Meng, B., J. Wang, et al. (2005). "Understanding the spatial diffusion process of severe acute respiratory syndrome in Beijing." Public Health **119**(12): 1080-1087.
Merler, S. and M. Ajelli (2010). "The role of population heterogeneity and human mobility in the spread of pandemic influenza." Proceedings of the Royal Society B-Biological Sciences **277**(1681): 557-565.
Moran, P. A. P. (1948). "The Interpretation of Statistical Maps." Journal of the Royal Statistical Society. Series B (Methodological) **10**: 243-251.
Moran, P. A. P. (1950). "Notes on continuous stochastic phenomena." Biometrika(37): 17-23.
Mossong, J., N. Hens, et al. (2008). "Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases." PLoS Medicine **5**(3): e74.
Mullahy, J. (1986). "Specification and Testing of Some Modified Count Data Models." Journal of Econometrics **33**(3): 341-365.
Oden, N., G. Jacquez, et al. (1996). "Realistic power simulations compare point- and area-based disease cluster tests." Statistics in Medicine **15**(7-9): 783-806.
Ong, J. B., M. I. Chen, et al. (2010). "Real-time epidemic monitoring and forecasting of H1N1-2009 using influenza-like illness from general practice and family doctor clinics in Singapore." PLoS One **5**(4): e10036.
Peiris, J. S. M., C. M. Chu, et al. (2003). "Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study." Lancet **361**(9371): 1767-1772.
Pitzer, V. E., G. M. Leung, et al. (2007). "Estimating variability in the transmission of severe acute respiratory syndrome to household contacts in Hong Kong, China." American Journal of Epidemiology **166**(3): 355-363.
Poutanen, S. M., D. E. Low, et al. (2003). "Identification of severe acute respiratory syndrome in Canada." New England Journal of Medicine **348**(20): 1995-2005.
Riley, S. (2007). "Large-scale spatial-transmission models of infectious disease." Science **316**(5829): 1298-1301.
Riley, S. (2008). "A prospective study of spatial clusters gives valuable insights into dengue transmission." PLoS Med **5**(11): e220.
Riley, S., C. Fraser, et al. (2003). "Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions." Science **300**(5627): 1961-1966.
Rose, C. E., S. W. Martin, et al. (2006). "On the use of zero-inflated and Hurdle models for modeling vaccine adverse event count data." Journal of Biopharmaceutical Statistics **16**(4): 463-481.
Sattenspiel, L. and K. Dietz (1995). "A structured epidemic model incorporating geographic mobility among regions." Mathematical Biosciences **128**(1-2): 71-91.
Sattenspiel, L. and D. A. Herring (2003). "Simulating the effect of quarantine on the spread of the 1918-19 flu in central Canada." Bulletin of Mathematical Biology **65**(1): 1-26.
Sen, A. and Z. Matuszewski (1991). "Properties of Maximum-Likelihood-Estimates of Gravity Model Parameters." Journal of Regional Science **31**(4): 469-486.
Sen, A. and T. E. Smith (1995). Gravity Models of Spatial Interaction Behavior. Verlag Berlin Heidelberg Germany, Springer.
Severe Acute Respiratory Syndrome Expert Committee, H. K. S. A. R. (2003). SARS in Hong Kong: From Experience to Actions.
Shannon, G. W. and J. Willoughby (2004). "Severe acute respiratory syndrome (SARS) in Asia: a medical geographic perspective." Eurasian Geography and Economics **45**(5): 359-381.
Shen, Z., F. Ning, et al. (2004). "Superspreading SARS events, Beijing, 2003." Emerging Infectious Diseases **10**(2): 256-260.

Smallman-Raynor, M., N. Johnson, et al. (2002). "The spatial anatomy of an epidemic: influenza in London and the county boroughs of England and Wales, 1918-1919." Transactions of the Institute of British Geographers **27**(4): 452-470.

Snow, J. (1854). On the Mode of Communication of Cholera. London, Churchill Livingston.

Song, C. and M. Kulldorff (2003). "Power evaluation of disease clustering tests." International Journal of Health Geographics **2**(9).

Stephan, G. E. (1988). "The Distribution of Service Establishments." Journal of Regional Science **28**(1): 29-40.

Stewart, J. Q. (1941). "An Inverse Distance Variation for Certain Social Influences." Science **93**(2404): 89-90.

Svensson, A. (2007). "A note on generation times in epidemic models." Mathematical Biosciences **208**(1): 300-311.

Svoboda, T., B. Henry, et al. (2004). "Public health measures to control the spread of the severe acute respiratory syndrome during the outbreak in Toronto." N Engl J Med **350**(23): 2352-2361.

Swartz, J. B. (1998). "An entropy-based algorithm for detecting clusters of cases and controls and its comparison with a method using nearest neighbors." Health and Place **4**: 67-77.

Tango, T. (2000). "A test for spatial disease clustering adjusted for multiple testing." Statistics in Medicine **19**(2): 191-204.

Tango, T. and K. Takahashi (2005). "A flexibly shaped spatial scan statistic for detecting clusters." International Journal of Health Geographics **4**(11).

The University of Hong Kong (2010). Influenza A(H1N1) Infodesk-latest development**:** Available at: http://www3.hku.hk/facmed/h1n1/pageC.php. Accessed 7 Jan 2011.

Thomas, R. (1999). "Reproduction rates in multiregion modeling systems for HIV/AIDS." Journal of Regional Science **39**(2): 359-385.

Transport Department. The Government of the Hong Kong Special Administrative Region (2005). "Public Transport Introduction." Available at: http://www.td.gov.hk/transport_in_hong_kong/public_transport/introduction/index.htm. Accessed 3 March 2009.

Transport Department. The Government of the Hong Kong Special Administrative Region. (2002). Travel characteristics survey 2002 final report.

Tsang, K. W., P. L. Ho, et al. (2003). "A cluster of cases of severe acute respiratory syndrome in Hong Kong." New England Journal of Medicine **348**(20): 1977-1985.

Um, J., S. W. Son, et al. (2009). "Scaling laws between population and facility densities." Proceedings of the National Academy of Sciences of the United States of America **106**(34): 14236-14240.

United Nations (2006). World Population Prospects The 2006 Revision. United Nations**:** Available at: http://www.un.org/esa/population/publications/wpp2006/English.pdf. Accessed 2003 March 2009.

United States Census Bureau, P. D. (2009). 2008 Population Estimates.

Varia, M., S. Wilson, et al. (2003). "Investigation of a nosocomial outbreak of severe acute respiratory syndrome (SARS) in Toronto, Canada." Canadian Medical Association Journal **169**(4): 285-292.

Viboud, C., O. N. Bjornstad, et al. (2006). "Synchrony, waves, and spatial hierarchies in the spread of influenza." Science **312**(5772): 447-451.

Vuong, Q. H. (1989). "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." Econometrica **57**(2): 307-333.

Wallinga, J. and M. Lipsitch (2007). "How generation intervals shape the relationship between growth rates and reproductive numbers." Proceedings of the Royal Society B-Biological Sciences **274**(1609): 599-604.

Wallinga, J. and P. Teunis (2004). "Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures." American Journal of Epidemiology **160**(6): 509-516.

Webster, R. G. (2004). "Wet markets - a continuing source of severe acute respiratory syndrome and influenza?" Lancet **363**(9404): 234-236.

White, L. F. and M. Pagano (2008). "A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic." Statistics in Medicine **27**(16): 2999-3016.

White, L. F. and M. Pagano (2008). "Transmissibility of the Influenza Virus in the 1918 Pandemic." PLoS ONE **3**(1): e1498.

White, L. F., J. Wallinga, et al. (2009). "Estimation of the reproductive number and the serial interval in early phase of the 2009 influenza A/H1N1 pandemic in the USA." Influenza and Other Respiratory Viruses **3**(6): 267-276.

Whittemore, A. S., N. Friend, et al. (1987). "A Test to Detect Clusters of Disease." Biometrika **74**(3): 631-635.

Wilson-Clark, S. D., S. L. Deeks, et al. (2006). "Household transmission of SARS, 2003." Canadian Medical Association Journal **175**(10): 1219-1223.

Wong, C. K., C. O. Tong, et al. (1999). "The development and calibration of a Lowry model with multiple market segments." Environment and Planning A **31**(11): 1905-1918.

Wong, T. W., C. K. Lee, et al. (2004). "Cluster of SARS among medical students exposed to single patient, Hong Kong." Emerging Infectious Diseases **10**(2): 269-276.

World Health Organisation (WHO) (2003). "Case Definitions for Surveillance of Severe Acute Respiratory Syndrome (SARS)." Available at: http://www.who.int/csr/sars/casedefinition/en/. Accessed 7 Jan 2011.

World Health Organisation (WHO) (2003). "Consensus document on the epidemiology of severe acute respiratory syndrome (SARS)." Available at: http://www.who.int/csr/sars/en/WHOconsensus.pdf. Accessed 3 March 2009.

World Health Organisation (WHO) (2004). "Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003." Available at: http://www.who.int/csr/sars/country/table2004_2004_2021/en/. Accessed 2003 Jan 2011.

Wu, J. T., S. Riley, et al. (2006). "Reducing the impact of the next influenza pandemic using household-based public health interventions." Plos Medicine **3**(9): 1532-1540.

Xia, Y. C., O. N. Bjornstad, et al. (2004). "Measles metapopulation dynamics: A gravity model for epidemiological coupling and dynamics." American Naturalist **164**(2): 267-281.

Yu, I. T. S., T. W. Wong, et al. (2005). "Temporal-spatial analysis of severe acute respiratory syndrome among hospital inpatients." Clinical Infectious Diseases **40**(9): 1237-1243.

Zhong, N. S., B. J. Zheng, et al. (2003). "Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003." Lancet **362**(9393): 1353-1358.