The Epidemiology and Control of Human Influenza in Vietnam

Thesis

The Epidemiology and Control of Human Influenza in Vietnam

Pham Quang Thai

Oxford University Clinical Research Unit

Ho Chi Minh City , Vietnam

A thesis submitted to the Open University for the degree of Doctor of Philosophy in the field of Life sciences

April 2014

Declaration by candidate

I have read and understood the University's definition of plagiarism and cheating given in the Research Degrees Handbook. I declare that this thesis is my own work, and that I have acknowledged all results and quotations from the published or unpublished work of other people.

I have read and understood the University's definition and policy on the use of third parties (either paid or unpaid) who have contributed to the preparation of this thesis by providing copy editing and, or, proof reading services. I declare that no changes to the intellectual content or substance of this thesis were made as a result of this advice, and, that I have fully acknowledged all such contributions

Signed: 

Date: 25 April 2014 Full Name: **Pham Quang Thai**

Abstract

Understanding the epidemiology of human influenza in Viet Nam is important for developing local policies and also for understanding the dynamics of influenza in tropical and subtropical southeast Asia. I have analysed an 18 year time-series of influenza-like-illness (ILI) surveillance data, and assessed the relationship of this time-series with climate variables and with sentinel influenza virus surveillance data. I also conducted a study of influenza A/H1N1 transmission within households.

ILI notifications in Viet Nam show a latitudinal gradient, with seasonality in the north but no seasonal pattern observed in low lying areas of central and southern Viet Nam. Seasonality is however observed in the elevated provinces of central Viet Nam, suggesting that the seasonal patterns are driven by climate. Principal component analysis finds that temperature and absolute humidity (AH) are positively correlated and together explain around 59% of total climatic variance, and that there is a strong latitudinal gradient in these variables. Regression tree analysis shows that provinces with strong seasonality of AH have strong ILI seasonality. Although virological surveillance data are limited, increases in ILI notifications are associated with an increase in the proportion of upper respiratory tract swabs that are influenza positive. In a prospective study of H1N1/2009 transmission in a household-based cohort, 11 of 59 household contacts were infected, giving a household secondary infection risk of 18.6% (95%CI 10.7-30.4%), but 5 (45%) did not develop symptoms. Virus genetic sequencing indicated that 10 of the 11 secondary cases (91%) were probably infected within the household rather than from the community.

This research provides new insights into the seasonality and climatic determi-

nants of ILI and influenza epidemiology in Viet Nam, and on the transmission of influenza within households. The findings are valuable for national influenza control policies and also add to the current state of knowledge of influenza epidemiology.

Acknowledgements

Most of my work has been supervised by Professor Peter Horby, who is also my teacher and my friend. I will never be able to thank him enough for his continuous support during my work. Also I would like to thank my co-supervisors Jeremy Farrar, Menno De Jong, Marc Choisy, and Maciej Boni. Their statistical and modelling expertise has taught me a lot, and been essential in helping me complete my studies.

I gratefully acknowledge all staff of the Epidemiology Department and all my colleague at NIHE; none of this would have been possible without their help. I have learnt every day when working with them. Thanks are especially due to Prof. Nguyen Tran Hien, Prof. Tran Nhu Duong, Dr. Vu Dinh Thiem, Prof. Nguyen Thu Yen, Prof. Le Quynh Mai, Dr. Dang Dinh Thoang and all of my colleagues from EPI northern region office and Ha Nam Province Preventive Medicine Center who helped me to finish my field work and gave me meaningful advice for my thesis.

I understand that the research owes much to the hard work of many people, but particular thanks are due to my friend and colleague Annette Fox for so ably managing the extensive laboratory work that was required and for the long nights and weekends she spent in the lab; to my colleagues Hoa, Thanh, Dung, Thuy, Trang and to all my friends at the Oxford University Clinical Research Unit in Hanoi for their friendship and dedication; I will also like to thank the health worker and participants of the Ha Nam cohort for their cooperation.

Thanks also to my two angels Quang Duc and Gia Han, with them time run faster but more meaningful. I always feel peaceful when I am with them.

My final thanks must go to Le Hang for her enduring love, unfailing support,

and endless capacity to listen to my minor work or problem

I dedicate this work to my mother who taught me right from wrong.

October 29th 2014

Table of Contents

List of Figures x

List of Tables xiii

Preface xiv

1 General introduction 1

1.1 Influenza viruses . . . 1

1.2 Influenza disease in humans . . . 3

1.3 Immunity to influenza . . . 4

1.4 Influenza epidemiology in humans . . . 6

1.4.1 Transmission routes. . . . 6

1.4.2 Transmissibility. . . . 6

1.4.3 Epidemic behaviour. . . . 7

1.5 Influenza in Southeast and East Asia . . . 7

1.6 Viet Nam . . . 9

1.7 Influenza in Viet Nam . . . 10

1.8 Influenza control in Viet Nam . . . 11

2 Introduction to the study of the seasonality of influenza 13

2.1 Introduction . . . 13

2.2 Influenza seasonality in temperate areas . . . 14

2.3 Influenza seasonality in tropical and subtropical areas . . . 15

2.4 Possible determinants of the seasonality of respiratory infections . . . 20

2.4.1 Host contact behaviours . . . 20
2.4.2 Host susceptibility . . . 21
2.4.3 Pathogen survival outside of the host . . . 21
2.5 Relative versus absolute humidity . . . 21
2.6 Experimental studies of climatic determinants of influenza survival and transmission . . . 23
2.7 Observational studies of climate variables and influenza transmission . . . 24
2.8 Summary . . . 26
2.9 Methods for assessing the seasonality of influenza . . . 27

3 Spatial patterns of ILI seasonality in Viet Nam 28
3.1 Introduction . . . 28
3.2 Objective . . . 29
3.3 Materials and Methods . . . 29
3.3.1 Data sources . . . 29
3.4 Results . . . 42
3.4.1 Data selection . . . 42
3.4.2 Data transformation, de-trending and normalisation . . . 45
3.4.3 Outliers . . . 45
3.4.4 Longitudinal and elevation effects . . . 47
3.4.5 Latitudinal pattern of ILI notification rates . . . 47
3.4.6 Quantifying the seasonality of ILI using wavelet analysis . . . 47
3.5 Discussion: . . . 51
3.6 Conclusion . . . 52

4 Climate association of influenza-like illness in Viet Nam 53
4.1 Introduction . . . 53
4.2 Materials and Methods . . . 54
4.2.1 Data and transformations . . . 54
4.3 Results . . . 59
4.3.1 Geographic patterns of climate variables . . . 59

4.3.2 Association between ILI seasonality and climatic variables 61
4.3.3 Extrapolation to the global scale 67
4.4 Discussion 67
4.5 Conclusion 74

5 Synchrony of ILI and sentinel virological surveillance data in Viet Nam 75
5.1 Introduction 75
5.2 Materials and methods 76
5.2.1 The Data 76
5.2.2 Analysis 78
5.3 Results 79
5.3.1 Regions Reports and Virological Confirmation Trends. 87
5.3.2 Do the data show influenza seasonality in Viet Nam? 87
5.3.3 Synchrony between ILI consultation and PCR positive rate 88
5.4 Discussion 91
5.5 Conclusion 92

6 Background to household cohort study 96
6.1 The Ha Nam longitudinal community study 96
6.1.1 Subject identification, selection and recruitment 97
6.1.2 Information sheet 99
6.1.3 Consent 99
6.1.4 Ethical review 99
6.1.5 Enrolment 100
6.1.6 Active surveillance for influenza 100
6.1.7 Twice yearly serology 103
6.1.8 Twice yearly re-census 103
6.1.9 Monitoring and quality assurance 103
6.2 Laboratory procedures 103
6.2.1 Detection of influenza infection 103
6.2.2 Detecting antibodies 103

6.2.3 Analysis . . . 104
6.2.4 Pandemic H1N1 sub study . . . 104
6.3 Candidate's role . . . 105

7 Pandemic H1N1/2009 transmission and shedding dynamics in households 106

7.1 Introduction . . . 106
7.2 Material and Method . . . 107
7.2.1 ILI surveillance and sample collection schedule . . . 107
7.2.2 Swab Collection . . . 109
7.2.3 Virology and serology . . . 110
7.2.4 Definitions and Analysis . . . 110
7.3 Results . . . 111
7.3.1 Index case house characteristics . . . 111
7.3.2 Secondary cases . . . 112
7.3.3 Virus RNA shedding and symptom dynamics . . . 115
7.3.4 Risk factors for secondary infection . . . 119
7.4 Discussion . . . 121
7.5 Conclusion: . . . 125

8 General discussion and concluding remarks 126

8.1 Contribution to knowledge on seasonal and pandemic influenza . . . 126
8.2 Further research directions . . . 129

References 130

A The routine surveillance system in Viet Nam 157

B Supplementary for chapter 3 and 4 164

C Ha Nam cohort supplementary 174

D R code 180

E Supplementary research paper **241**

List of Figures

2.1 **Clinical incidence of influenza-like illness in England and Wales.** . 15

2.2 **Weekly number of influenza and pneumonia deaths per 10 million populations from January 1972 to December 1997 in the United States, France, and Australia.** 16

2.3 **World Map of the Köppen-Geiger Climate.** 17

2.4 **Global map of influenza peak timing and epidemic duration.** . . . 20

3.1 **Number of districts per province.** 30

3.2 **Changes to provinces between 1980 and 2010.** 32

3.3 **Map of changes to provinces between 1980 and 2010.** 33

3.4 **Map of provincial population and geographic centroids.** 34

3.5 **Time-series components** 37

3.6 **Wavelet transformation of a time-series** 39

3.7 **Time-dependent spectral analysis of epidemiological time-series with wavelets.** 40

3.8 **Seasonality versus amplitude** 41

3.9 **Raw ILI notification data** 42

3.10 **Selection of optimal time period for analysis** 43

3.11 **Trimmed data** 44

3.12 **Data transformation, detrending and normalisation** 45

3.13 **Removing outliers, detrending, and normalising** 46

3.14 **Heat map of ILI notifications** 48

3.15 **Example of wavelet transform** 49

3.16 Strength of ILI seasonality. 50

4.1 Location of meteorological stations 55
4.2 Sources of data for the NCEP/NCAR project. 57
4.3 Average temperature by time and latitude. 59
4.4 Temporal and spatial patterns of climatic factors. 61
4.5 Variation of climatic variables in Viet Nam over one year 62
4.6 Latitudinal gradient of climate seasonality. 64
4.7 Regression tree analysis of relationship between ILI seasonality and climate factors. 65
4.8 Seasonalities of ILI and absolute humidity. 66
4.9 Lag between ILI and absolute humidity. 67
4.10 Global seasonality of absolute humidity. 68

5.1 Location of national influenza surveillance system sites. 77
5.2 ILI sample collected by age group and year 80
5.3 Proportion of swabs that are influenza positive by time (monthly) 81
5.4 Number of sites and completeness of data collection, by week. . . 82
5.5 PCR positive proportion (>25%) by week for all northern sites. . 83
5.6 PCR positive proportion (>25%) by week for all central and highland sites. 84
5.7 PCR positive proportion (>25%) by week for all southern sites. . 85
5.8 PCR positive proportion (>25%) by week for all region and all site. 86
5.9 Standard Fourier fit to PCR positive rate. 88
5.10 Wavelet analysis for the weekly proportions of specimens positive for influenza. 89
5.11 Phase difference between ILI consultations and PCR positive rate, by age 93
5.12 Phase difference ILI consultations and PCR positive rate after removal of National Paediatric Hospital. 94

5.13 **Graph of phase difference of ILI consultations and PCR positive rate.** . . . 95

6.1 **Position and size of Ha Nam province** . . . 97
6.2 **Schematic diagram of the design of the Ha Nam household cohort study** . . . 98
6.3 **The Ha Nam household characteristics.** . . . 100
6.4 **Training on taking nasal swabs.** . . . 102
6.5 **Working with Village Health Workers in bleeding campaign.** . . . 102

7.1 **The enhanced detection of H1N1 2009 in all cohort members during pandemic time** . . . 109
7.2 **Daily viral loads and symptoms in confirmed H1N1/09 cases from index case households.** . . . 117
7.3 **Daily viral loads and symptoms in confirmed H1N1/09 cases from index case households (continued).** . . . 118

B.1 **Raw data** . . . 165
B.2 Raw data, cont'd. . . . 166
B.3 Raw data, cont'd. . . . 167
B.4 **Transformed data** . . . 168
B.5 Transformed data, cont'd. . . . 169
B.6 Transformed data, cont'd. . . . 170
B.7 **Wavelet transform** . . . 171
B.8 Wavelet transform, cont'd. . . . 172
B.9 Wavelet transform, cont'd. . . . 173

C.1 **The Thanh Ha commune health worker** . . . 174
C.2 **The OUCRU Ha Noi** . . . 175
C.3 **KaplanMeier curves of time until cessation of viral RNA shedding in virologically confirmed cases** . . . 176

List of Tables

5.1 **ILI surveillance data by year** 80

5.2 **Results of tested ILI samples**, by influenza virus subtype from 2006 to 2012 80

5.3 **Distribution of ILI and tested cases by age** 81

5.4 **Monthly averages of influenza isolation proportions and ILI consultation rates for whole country, 2006-2012** 90

6.1 **Characteristics of Participants and Households at Recruitment, Ha Nam, Viet Nam, 2007–2010.** 101

7.1 **Composition of households in the cohort and those with an index case.** 112

7.2 **Comparison of H1N1/2009 envelope gene sequence diversity within households and individuals and between households.** 113

7.3 **Distribution of cases, contacts and secondary cases by age, gender and position in the family.** 115

7.4 **Virus shedding and transmission characteristics.** 116

7.5 **Univariate analysis of factors associated with transmission of H1N1-2009 from index cases to household contacts during the first pandemic wave** 120

7.6 **Risk factors for transmission of H1N1-2009 from index case to household contacts during the first pandemic wave.** 120

Preface

This thesis is divided into two major sections. The first deals with my work to test the hypothesis that there is a spatial pattern of influenza seasonality in Viet Nam and that this pattern in driven by climatic factors. To achieve this I first assessed the seasonal periodicity by province of routine notifications of influenza-like-illness (ILI) collected from 1993 to 2010. The results of this analysis were then compared with the seasonality of a range of climate variables over the same period in order to assess which climate variables are most strongly associated with influenza seasonality in Viet Nam. Finally, I assessed the extent to which ILI notification data matches influenza virological surveillance data collected between 2006 and 2012. The second part of the thesis reports the results of a sub-study of influenza A/H1N1/2009 conducted in a prospective community cohort that was established to provide data on the epidemiology of seasonal influenza in Vietnam. The cohort was running when pandemic influenza A/H1N1/2009 emerged and therefore also provides data on the epidemiology of pandemic influenza in Vietnam. Although the two sections use different data sources and have their own aims, they link together to build a fuller picture of the epidemiology of influenza in Viet Nam and also add to the body of knowledge on the epidemiology of influenza in general.

The results of Chapters 3 and 4 are currently being prepared for publication and the work presented in Chapter 7 was published in early 2014. Additional data and methodological details, including the R script used for much of the analysis, are available in Annexes. My role in the work presented in this thesis included the preparation of research protocols and data collection instruments, the submissions for ethical approvals, implementing the studies and supervising the field work, and managing all the primary and secondary data collection. I

cleaned, prepared and analysed the data personally under the guidance of my supervisors. The laboratory work was conducted by others but I was involved in the planning and design of the laboratory analyses and the interpretation of results. The work done by other colleagues is clearly acknowledged.

All of the work presented in this thesis was conducted under the supervision of and with the permission of the Viet Nam Ministry of Health. The work was conducted under the umbrella of a project agreement between my institute, the National Institute of Hygiene and Epidemiology, and the Oxford University Clinical Research Unit in Hanoi. The research was approved by the Institutional Review Boards of NIHE, the Vietnam Ministry of Health, and the Oxford University Tropical Research Ethics Committee.

Chapter 1

General introduction

1.1 Influenza viruses

Influenza viruses are enveloped, single-stranded, negative-sense RNA viruses (100-120 nm in diameter) of the family Orthomyxoviridae. Influenza viruses are categorised serologically and genetically into three types, named A, B, and C. Influenza B and C viruses are predominantly human pathogens whilst type A naturally infects a wide range of birds and mammals (Webster et al., 1992). Influenza A is the most important of the three influenza types because it regularly causes large epidemics in human populations, and occasionally causes a global outbreak (a pandemic) when a new subtype emerges to which humans are immunologically naïve (Taubenberger and Morens, 2010). Like many RNA viruses, influenza A viruses evolve rapidly, especially in regions (epitopes) of the surface proteins that are recognised by the adaptive immune system, allowing the virus to repeatedly reinfect human populations. This evolution has been termed antigenic drift (Both et al., 1983; Earn et al., 2002; Smith et al., 2004; Webster et al., 1992). Influenza B evolves more slowly and is usually associated with milder disease, but influenza epidemics are sometimes primarily driven by influenza B virus, especially in children (Paul Glezen et al., 2013). Influenza C rarely causes disease in humans and is considered clinically unimportant (Kamps et al., 2006).

Influenza A and B viruses have a segmented genome consisting of eight segments of RNA coding for 11 proteins. On the surface of the influenza A viruses there are three glycoproteins, the haemagglutinin (HA) and neuraminidase (NA), which are the primary immunogenic proteins of influenza A viruses, and the M2 protein. HA mediates binding of the virus to target cells and subsequent entry of the virion into the cells through endocytosis. The head of the HA molecule binds to sialic acid molecules of glycoproteins and glycolipids expressed on host cell membranes, and the binding affinity of HA variants

to different sialic acid motifs is a critical factor determining the host range of influenza A viruses (Matrosovich et al., 2004). Antibodies that bind to the receptor binding site of the HA prevent virions from binding to cell surface receptors and therefore prevent infection. These antibodies are the dominant mechanism of acquired immunity against influenza A and are termed haemagglutination inhibiting (HI) antibodies since they prevent the agglutination of erythrocytes in vitro. The maximum dilution (titer) of serum that inhibits the agglutination of erythrocytes by influenza A viruses is the commonest used method to assess the concentration of virus neutralising antibodies, and is termed the HI assay. NA acts as an enzyme that cleaves the sialic acid from glycoprotein molecules to allow the release of progeny virus from the surface of infected cells and to stop virions binding to one another (McKimm-Breschkin, 2013). Neuraminidase inhibitors are a major class of anti-influenza drug, which act by preventing the release of newly formed virions for the infected cell, and include the drugs oseltamivir, zanamivir, and peramivir. The M2 protein is an ion channel in the viral lipid membrane that allows a change in pH of the inside of the virion once it enters the target cell, leading to uncoating of the virion and release of RNA (Schnell and Chou, 2008). The M2 ion channel is the target of the other major class of anti-influenza drugs, the M2 inhibitors (amantadine and rimantidine). However, the utility of the M2 inhibitors is severely limited by the rapid and widespread development of resistance conferring mutations. The matrix protein (M1) is a structural protein that lies beneath the lipid envelope. The PA, PB1, and PB2, genes code for proteins that are involved in RNA synthesis for progeny viruses during replication inside the host cell. The nucleoprotein (NP) is structurally associated with the viral RNA and is necessary for RNA replication. The non-structural protein (NS1) is involved in the evasion of the host innate immune response, particularly the neutralisation of interferon-induced activities (Hale et al., 2008). Influenza type A is categorised into subtypes based on the genetic and antigenic characteristics of the HA and NA. There are currently 18 identified HA (H 1-18) and 9 NA (NA 1-9) antigenic variants. Only three main subtypes of HA (H1, H3, and H2) and 2 NA subtypes (N1 and N2) are known to have become fully adapted to humans and cause major epidemics. The majority of HA and NA combinations have been identified in aquatic birds, which are the primary natural reservoir of influenza A viruses, and which

are thought to be the source of progenitor viruses or gene segments of pandemic influenza A strains (Alexander, 2007). Zoonotic, non-human-adapted, influenza A viruses, such as influenza (HPAI) subtypes H5N1 and H7N9, occasionally infect humans but, currently, are not able to transmit efficiently between humans (Berg JM, Tymoczko JL, 2002; Gao et al., 2013; Leung et al., 2007; Rott et al., 1996).

1.2 Influenza disease in humans

Influenza is one of the commonest infections of humans. Annually, seasonal influenza viruses are estimated to infect 500-800 million people, resulting in 5 million severe cases and 250,000-500,000 deaths (CDC, 2005). It has been estimated that in children under 5 years of age in 2008 there were 20 million cases of influenza-associated acute lower respiratory tract infection and 1 million severe infections (Nair et al., 2010). In the United States, it is estimated that between 5% and 20% of the population about 50 to 60 million people are infected with influenza each year. Out of those, about 31 million come to see doctor, and 200.000 are admitted to a hospital. This corresponds to an annual incidence of 67/100,000 for influenza hospitalisations. The annual attributable mortality for influenza in the United States is estimated at around 36, (12/100,000 pop.) (Reichert et al., 2004; Wilschut et al., 2006).

During epidemics the overall infection rate can range from 10-20% in the community and can reach up to 50% in closed communities like schools and kindergartens (Heymann and American Public Health Association, 2008). Data from recent prospective cohort studies have found serologically defined infection rates of around 20% per season in the community, of which the majority are subclinical or asymptomatic (Hayward et al., 2013; Horby et al., 2012). However, the standard criteria for defining influenza infection based on serology is the finding of a four-fold or greater rise in HI titer between acute and convalescent serum samples. This criterion is probably too strict for epidemiological studies, since many individuals will have a less than four-fold increase in antibody titers in paired samples, which cannot be fully explained by variability in the assay (Cauchemez et al., 2012). As such, it is likely that the true influenza infection rate per season is often greater

than 20% of the population.

Infection with influenza viruses can cause a broad range of illness, from asymptomatic infection to various, mostly mild respiratory illnesses and, infrequently, fulminant viral pneumonia and/or secondary bacterial pneumonia (Van-Tam and Sellwood, 2010). Whilst the clinical syndrome of influenza is classically associated with upper respiratory tract symptoms accompanied by fever, headache and myalgia; mild and subclinical infection is very common and only a small proportion of all cases of influenza infection meet the classical case definition for Influenza-Like-Illness (ILI) (Hayward et al., 2013; Riley et al., 2011; Thomas, 2014). Disease severity can vary greatly depending on the immunological attributes of the population, the age and health status of individuals, and the pathogenicity of the virus. (Clancy, 2008; Fukuyama and Kawaoka, 2011; Kamps et al., 2006). Clinical attack rates are generally highest in individuals who have low concentrations of antibodies against the HA protein (neutralising antibodies), with a titer of HI antibodies of 1:40 correlating with a 50% protection against clinical influenza (Coudeville et al., 2010). Children generally have the highest clinical attack rate since they are immunologically naïve. However, disease severity is also associated with non-HI related immunity (e.g. cell mediated immunity), host vulnerability (e.g. age and co-morbidities), and the intrinsic virulence of the influenza subtype (e.g. infection with H3N2 viruses is generally accepted to cause more severe disease than infection with H1N1 viruses).

1.3 Immunity to influenza

The innate immune response is the first defence against influenza, with influenza virus infection stimulating the production of interferon and pro-inflammatory cytokines (Couch and Kasel, 1983). However, the NS1 and the PB1 proteins inhibit the production of type I interferon, thereby counteracting a major component of the innate immune system (Kreijtz et al., 2011). Following infection, a major component of acquired immunity is the development of antibodies to the binding domain of the HA protein. These antibodies inhibit the ability of influenza viruses to hemagglutinate red blood cells and are therefore called hemagglutination inhibiting (HI) antibodies. HI antibodies prevent infection and

are subtype specific. However, minor changes in the structure of the binding domain of the HA protein can circumvent HI antibodies. The process of variation in the HA protein is called antigenic drift (see earlier) and is the reason that people can be repeatedly infected with influenza and influenza vaccines need to be updated regularly (Couch and Kasel, 1983; Kreijtz et al., 2011). Antibodies are also produced that are directed against non-HA epitopes (e.g. neuraminidase antibodies) and antibodies have recently been identified that recognise the stem region of the HA protein and can recognise multiple influenza subtypes within the same HA group (Group 1 = H1, H2, H5, H6, H8, H9, H11, H12, H13, H16. Group 2 = H3, H4, H7, H10, H14, H15). Cell mediated immunity also develops following infection, and since it requires the presentation of virus antigens on the surface of T-lymphocytes, it cannot prevent infection but can reduce viral replication and therefore attenuate the duration and severity of infection. Since the internal proteins of influenza viruses are more conserved than the external proteins, cell mediated immunity can cross react to different subtypes of influenza, providing heterosubtypic immunity (Kreijtz et al., 2011).

High risk groups for severe influenza-associated disease are defined by age group or by the presence of certain chronic conditions (Mertz et al., 2013; Wilschut et al., 2006). The elderly (over 65), and especially individuals over the age of 85, are at increased risk for severe complications even in the absence underlying chronic disease. The influenza case fatality rate in the over-65 age group is 11.3 times higher than in the 1-44 year age group, and accounts for approximately 95% of all influenza-attributable deaths (Wilschut et al., 2006; Zaman et al., 2009). The increased vulnerability to severe disease in the elderly is thought to be due to age related declines in the functioning of innate and adaptive immune responses. A range of chronic conditions, such as chronic heart disease, chronic lung disease, and obesity, are risk factors for severe influenza-associated disease; the quality of evidence is however generally low (Mertz et al., 2013). Woman in the last trimester of pregnancy and within the first four weeks of the post-partum period are at increased risk of severe pandemic influenza A/H1N1 (Mertz et al., 2013).

1.4 Influenza epidemiology in humans

1.4.1 Transmission routes.

Influenza viruses can be transmitted from person to person by a variety of routes, including large respiratory droplets, small air-borne particles (aerosols), and direct contact. The relative contribution of each route remains unclear and may vary due to the complex relationship between the environmental conditions, virus survival and transmission, and host susceptibility (Killingley and Nguyen-Van-Tam, 2013). The case for aerosol transmission of influenza has been summarised by Tellier et al. (Tellier, 2009) and may account for up to 50% of transmission events (Cowling et al., 2013). However, a review by Brankston et al. concluded that the majority of transmission occurs over short distances and that long distance aerosol transmission is uncommon (Brankston et al., 2007). Lowen and Palese (Lowen and Palese, 2009) have proposed that the dominant mode of transmission may vary according to climatic conditions, with aerosol transmission predominating in temperate regions, and direct contact predominating in tropical regions. Although this hypothesis is not supported by the findings of Cowling et al. (2013).

1.4.2 Transmissibility.

The fundamental transmissibility of any infectious disease is hard to estimate outside of experimental settings, so the basic reproduction number (R0, the average number of secondary cases generated by one case in an entirely susceptible population) is usually estimated from epidemic dynamic data. The peak R0 for seasonal influenza has been estimated, using data from France, to be between 1.6 and 3, and the waning of immunity to be the in the range of 3-8 years (Truscott et al., 2012). The R0 for pandemic influenza H1N1/2009 has been estimated at between 1.2 and 2.3 (Boëlle et al., 2011). The mean and median serial interval (or generation time) of seasonal influenza has been estimated at 2.6 and 3 days respectively (Suess et al., 2010) and the estimated mean serial interval of pandemic H1N1/2009 was 3 days (95 % CI 2.4-3.6) in a systematic review by Boëlle et al. (2011). Numerous studies have confirmed that children have the highest infection rates (Morgan et al., 2010). The high infection rates due to immunological naivety, and

higher social contact rates amongst school age children means that children play a central role in influenza transmission (Horby et al., 2011). In the early stages of an outbreak, household transmission mainly occurs from children (Morgan et al., 2010) and children are important in sustaining factor community transmission (Sugimoto et al., 2011). The result of community transmission studies by Cowling et al. (2013) suggest that reducing social contact frequency may not prevent household transmission that leads to cycles of household-community transmission.

1.4.3 Epidemic behaviour.

Since influenza is readily transmissible from person to person and has a short serial interval it causes clear epidemics, with a rapid increase in case numbers and a well-defined epidemic curve. The seasonal timing of influenza epidemics is discussed in Chapter 2. As discussed in section 1.2 above, influenza A viruses are antigenically variable, leading to recurring epidemics when new variants emerge, and pandemics when novel subtypes occur.

1.5 Influenza in Southeast and East Asia

Until relatively recently, tropical countries were believed to have a low burden of seasonal influenza, and this perception has contributed to the low levels of utilisation of influenza vaccines in tropical and subtropical regions (Macroepidemiology of Influenza Vaccination (MIV) Study Group, 2005). This was likely due to the paucity of studies conducted in these regions, and the lack of specimen collection in national surveillance systems. South China and south east Asia more generally has been considered an epicentre for the emergence of novel influenza viruses that may cause pandemics (Shortridge and Stuart-Harris, 1982), but most of the scientific investigation has focused on avian influenza viruses. However, recent studies have demonstrated that influenza is a common cause of respiratory illness in tropical countries (Simmerman and Uyeki, 2008) and hospitalisation rates may even exceed those in temperate regions (Chiu et al., 2002). Data from 1982-2004 showed that 22-46% of hospitalised patients were admitted because of respiratory illness, of which influenza was detectable in up to 14% (Simmerman and Uyeki, 2008). In

Thailand from 1993-2002, influenza burden has ranged from 64 to 91 hospitalized cases per 100,000 population per year (Simmerman et al., 2004). Influenza also caused 10.4% of total respiratory illnesses in population, of which, 52% were patients under the age of 15 (236/100,000 pop.) and over the age of 75 (375/100,000 pop). From 2005 to 2008, it is estimated that more than 36,000 patients were admitted to hospitals per year and approximately 300 patients died each year because of influenza in Thailand (Simmerman et al., 2009). However, it is only in-hospital pneumonia deaths, which are likely to represent only a small fraction ($< 10\%$) of total influenza-related deaths. In Indonesia, Influenza A and B have been recorded year-round, with up to 20% of ILI cases testing positive for influenza virus and annual peaks during the rainy season, mainly caused by influenza A (Kosasih et al., 2013). In Lao PDR, Khamphaphongphane et al. (2013) also found influenza activity year-round but with greater transmission during the second half of the year and the virus subtypes changing each year. The aggregated result from ILI and influenza virus surveillance conducted in 14 countries collected by FluNet via the Western Pacific Region of the World Health Organization from 2006-2010 show that influenza is common in all countries but with different patterns in different countries (Members of the Western Pacific Region Global Influenza Surveillance and Response System, 2012). Seasonal cycles were prominent in temperate countries in the northern and southern hemispheres, but less clear patterns were seen in tropical countries (Members of the Western Pacific Region Global Influenza Surveillance and Response System, 2012). There is evidence that a transmission network exists within the Western Pacific Region, with dominant strains in one country later becoming dominant in other countries of the region (Members of the Western Pacific Region Global Influenza Surveillance and Response System, 2012).

In east and south east Asia, Hong Kong and Singapore have the longest running influenza surveillance systems and the most influenza-centered research. Studies in Hong Kong have shown that the hospitalisation rate of influenza is similar to the United States, a representative temperate region (Viboud et al., 2006b; Wong et al., 2004, 2006). As elsewhere, the highest morbidity and mortality is concentrated in children and the elderly (Chiu et al., 2002; Wilschut et al., 2006). A study from Hong Kong, Guangzhou and Singapore found influenza related mortality burden to be slightly higher for A/H1N1

compared to A/H3N2, with Hong Kong having the highest influenza-associated mortality, at 13.4 deaths per 100,000 population (Yang et al., 2011). An influenza study in Singapore by Lee et al. (2009) found that most influenza epidemics between 1950 and 2000 were associated with increases in all cause mortality. Similar results were obtained by Wu et al. (2012) in Hong Kong, with an association between excess deaths and influenza activity: with 95% of the excess deaths occurring in people aged 65 years or more. When pandemic influenza H1N1/2009 emerged, it was initially unclear if the virus was more virulent than seasonal influenza viruses. Studies by Cowling et al. (2010) found that H1N1/2009 was similar to seasonal influenza in terms of household transmission and the secondary household attack rate.

In addition to increased awareness of the importance of influenza as a cause of respiratory illness in east and south east Asia, this region is also of special interest as a source of novel and drifted influenza A viruses. As well as the interest in east and south east Asia as a source of influenza viruses of zoonotic origin (Shortridge and Stuart-Harris, 1982), there has more recently been interest in the region as a source of antigenically drifted influenza A strains, that seed annual epidemics in temperate regions (Nelson et al., 2007; Rambaut et al., 2008; Russell et al., 2008a). However, this hypothesis remains unproven, with more recent analysis proposing a more complex global pattern of influenza virus migration (Bahl et al., 2011) Influenza immunization is not common in the Western Pacific Region, with 30% of 37 surveyed countries having no national influenza immunization policy or recommendations, and only 50% having a well established national immunisation policy (Dwyer et al., 2013). Those countries that did have a publicly funded programme only purchased sufficient vaccine to immunise 25% or less of their population. The evidence base to support decisions about the introduction of influenza vaccines in WHOs Western Pacific Region is limited (Samaan et al., 2013).

1.6 Viet Nam

Viet Nam has a land area of 330,951 km^2, making it the 65th largest in the world. The country is elongated, with a length of 1,650 km, and is situated between 8° and 24°

from south to north, straddling different climate zones. The three main climate types are: northern climate with four distinct seasons, winter temperatures occasionally falling as low as 7°C in Ha Noi and more rain in summer, East Truong Son (mountain) climate with rain in autumn and winter and Southern climate very close to equatorial climates with temperatures rarely dropping below 20°C (Tam et al., 2004). Viet Nam is a narrow country, with elevations ranging from 0 and 3,000 m from east to west. Viet Nam has a long land border, with China to the north and Laos and Cambodia to the west. With a 2013 estimated population of 90 million, it is the third most populous country in Southeast Asia after Indonesia and the Philippines, and 13^{th} in the world. Thirty percent of the Viet Nam population lives in urban areas and the population is concentrated in the agriculturally productive and industrial zones of the Red River Delta in the north and the Mekong River Delta in the south. The age distribution of the population is: 25% aged under 15 years, 68% aged 15-64 years, and 7% aged 65 years or more (GSO, 2012).

Viet Nam has experienced sustained economic growth in the last two decades (average annual growth 7.1%) and recently transitioned from a low-income to a lower-middle income country. The per capita gross national income in 2012 was estimated to be 1,749 USD, and around 60% of the population have unskilled occupations in agriculture, fisheries, and forestry. Around 13% of the population live on less than 1.25 USD per day (2008 Asian Development Bank estimate). After becoming a lower-middle income country, the Vietnamese government's investment in health has remained low at only 5 USD / resident / year: a lower position than other countries in the region (Malaysia, 63 USD, Thailand 44 USD, Laos 8 USD, Indonesia 7 USD). In the developed countries, healthcare costs are approximately $2,000/person/year, and in some cases much higher (data report of the Ministry of Health. Viet Nam). Nevertheless, Viet Nam has good health and development indicators relative to its gross domestic product.

1.7 Influenza in Viet Nam

The clinical syndrome caused by influenza virus, typically referred to as influenza-like illness (ILI), has been one of the many reportable diseases in Viet Nam since 1979. This

routine reporting system only records the aggregated number of syndromic cases of ILI, without retaining a case-based record. Prior to 2004 influenza was a low priority for the Ministry of Health (MOH) and the research community in Viet Nam, with national surveillance limited to routine monthly reporting, in which the disease was not confirmed, typed, or subtyped by PCR or culture. After the SARS outbreak in Ha Noi in 2003 and the highly pathogenic avian influenza H5N1 outbreak in 2004, influenza surveillance has been strengthened in Viet Nam. Prior to 2003, the National Institute of Hygiene and Epidemiology (NIHE) of Viet Nam had collaborated with Nagasaki University of Japan to conduct influenza research in Ha Noi. The outcome of the research from 2001-2003 showed that only 2.5% of ILI patient in Ha Noi tested positive for influenza (Nguyen et al., 2007). In 2006, a sentinel surveillance system for influenza was established in Viet Nam which started with five sites and increased to fifteen sites across the country. The data from 2006 to 2009 showed that, influenza accounted for about 21.9% of total ILI consultation and that about 12.5% of all patient visiting the sentinel sites did so because of ILI. The influenza positive rate of severe pneumonia was 4.8% (Nguyen et al., 2009). More details are provided in chapter 5. In 2007, a household-based cohort study in a semi-rural commune one hour away from Ha Noi was established, and details of the cohort are described in chapter 6 which can be found in Horby et al. (2012).

1.8 Influenza control in Viet Nam

Since the emergence of highly pathogenic avian influenza (HPAI) H5N1 in Viet Nam in 2004, the interest and investment in influenza control in Viet Nam has increased dramatically. Since 2004 the Ministry of Health has organised a steering committee on influenza pandemic preparedness. During the 2009 influenza pandemic, all of the medical and public health systems of Viet Nam joined in this work. Infra-red thermal detection cameras were set up at both major international airports. Only 0.15% of 760.000 screened passengers showed symptoms of ILI and none of them had severe outcomes (Hien et al., 2010). Severe acute respiratory infection (SARI) cases were investigated and quarantined. Specimens were tested and molecularly typed on a rapid schedule to determine if the immigration of

the virus into the country could be stopped. The SARI surveillance system is still functioning but the major concerns now are H5N1 and H7N9. Viet Nam, like many countries, has established a national stockpile of antiviral drugs, and is actively pursuing domestic influenza vaccine production capability. Three projects on the development of domestic influenza vaccine have been established in Viet Nam and they have all made progress with vaccine development. The company Vabiotech has already completed pre-clinical studies, and phase I, II, and III clinical trials for avian influenza A/H5N1 vaccine and is awaiting registration and licensing. Vabiotech are also developing seasonal influenza vaccines using cell based vaccine production (MDCK, microcarrier and PMK cell line). The IVAC, are using a traditional egg based influenza vaccine production system when Polyvac and Pasteur Ho Chi Minh develop A/H1N1/2009 vaccine using Vero cell which is now at phase II. All these influenza vaccine development projects have received support from the Ministry of Science and Technology budget for R&D and/or the Ministry of Health (Ministry of Health, 2013).

Chapter 2

Introduction to the study of the seasonality of influenza

2.1 Introduction

> *Whoever wishes to investigate medicine properly, should proceed thus: in the first place to consider the seasons of the year, and what effects each of them produces for they are not at all alike, but differ much from themselves in regard to their changes.*
>
> *The opening sentence of On Airs, Waters, and Places*
>
> *By Hippocrates, 400 BC*

The term 'seasonal' refers to an event occurring each year at a specific time of the year, and is derived, via the old French term seison, from the Latin Satiō, meaning sowing or planting. The seasonality of human diseases has been recognized since the very beginning of modern medicine and Hippocrates wrote in "Of the Epidemics" that "The greatest and most dangerous disease, and the one that proved fatal to the greatest number, was consumption. With many persons it commenced during the winter" (Hippocrates) Study of the seasonality of respiratory illnesses began even before the influenza virus was identified in 1933 (Britten, 1932; Paul and Freese, 1933; Smith et al., 1933)

Seasonality can be described as a regular and predictable cycle occurring at a frequency of one year or less, and although seasonality is a well-understood concept, for analytic purposes seasonality must be quantitatively defined. There are two separate analytic tasks, detecting seasonality and measuring the characteristics of seasonality. In the past the detection of seasonality relied on subjective impressions, later progressing to plotting data by time in the form of bar or line graphs. Statistical methods for detecting seasonality

include time series analysis (Fourier transformation, wavelet transform)(Broutin et al., 2005; Mi et al., 2005), regression analysis, and mathematician modelling (Dominici et al., 2002; Held and Paul, 2012; Thompson et al., 2006; Yaari et al., 2013). As described in Chapter 3, I chose the wavelet analysis method to detect and measure seasonality.

The last decade has seen a renewed interest in climatic factors that may be associated with influenza outbreaks or epidemics, but the key climatic factors that modulate transmission are still under debate. The association between certain climatic factors and influenza epidemics in temperate countries gives some improved predictability of epidemic onset, but the occurrence of influenza as a winter phenomenon remains unchanged. In the tropics however, the timing of influenza epidemics is much less predictable. Understanding the drivers of influenza dynamics in these regions has recently become a research priority, partially at least because phylogenetic studies have indicated that East Asia may play a critical role in maintaining global influenza circulation (Bedford et al., 2010; Le et al., 2013; Rambaut et al., 2008; Russell et al., 2008b). This chapter is informed by a literature search. The term Influenza, Human [MeSH Terms] AND Periodicity [MeSH Terms] was entered into PubMed (`https://www.ncbi.nlm.nih.gov/pubmed`), and the search was last updated on the 16th March 2014. A total of 2057 records were retrieved and the titles screened for relevance. 69 relevant articles were selected and the full abstracts reviewed. The full text of the most relevant articles were downloaded, where possible.

2.2 Influenza seasonality in temperate areas

The seasonality of ILI and influenza in temperate regions is extremely well established, as demonstrated by long term surveillance data in, for instance, the United Kingdom (see figure 2.1) (Elliot and DM, 2006) the United States (CDC, 2013), France, and Australia (see figure 2.2) (Altizer et al., 2006; Goldstein et al., 2011; Ohmit and Monto, 2006; Viboud et al., 2006b, 2004a). Various longitudinal community based studies of respiratory infections that were conducted from the 1920s (Hagerstown morbidity study) to the early 1980s (Houston and Tecumseh), and which included the New York and Seattle Flu Watch programmes, the Cleveland and Houston Family Studies and the Tecumseh study, have

also demonstrated the seasonality of influenza transmission in temperate regions (Monto, 1994). It is therefore undisputed that in temperate regions that are clear epidemics of influenza each year that predictably occur in the colder winter months, although the precise timing cannot be predicted (Dowell, 2001; Lowen et al., 2008; Park and Glass, 2007; Viboud et al., 2006b).

Out of season circulation of influenza does however occur, and occasionally full scale influenza epidemics can occur outside of the traditional winter periods, most notably when a novel influenza A strain circulated causing a pandemic (Chowell et al., 2011; Jakeman and Sweet, 1996; Kelly et al., 2013; Kohn et al., 1995; Van-Tam and Sellwood, 2010). The reasons for out of season influenza cases and epidemics are an interesting topic, but are outside of the scope of this thesis and will not be discussed further here.

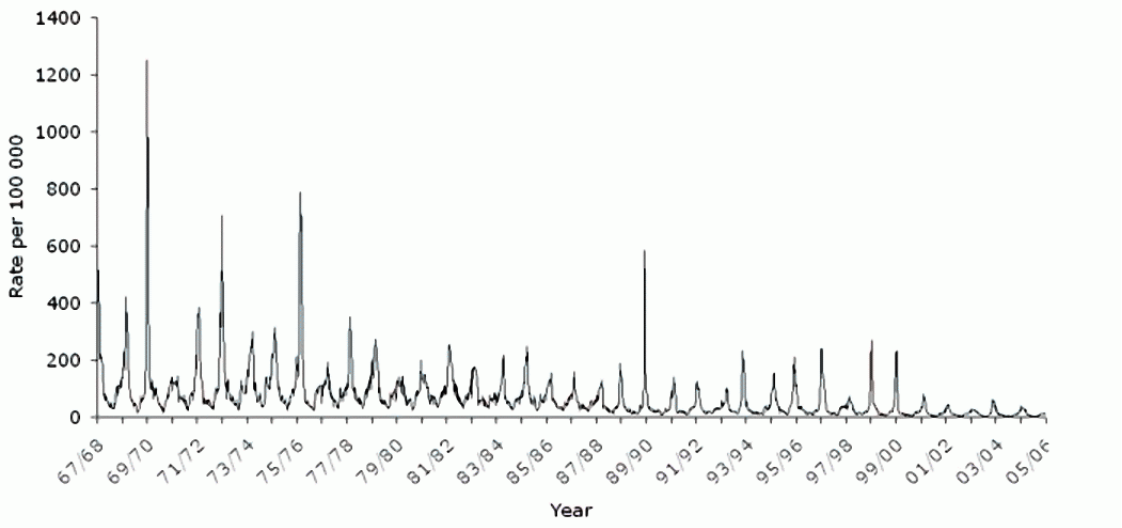


Figure 2.1: Clinical incidence of influenza-like illness in England and Wales.

Weekly incidence rates from 1967 to 2006. Source: (Elliot and DM, 2006)

2.3 Influenza seasonality in tropical and subtropical areas

It should first be noted that the terms tropical and subtropical are principally geographic rather than meteorological terms. The tropics are region of the earth's surface lying between particular circles of latitude. The Tropic of Cancer is defined as the most northerly circle of latitude on the Earth at which the Sun may appear directly overhead

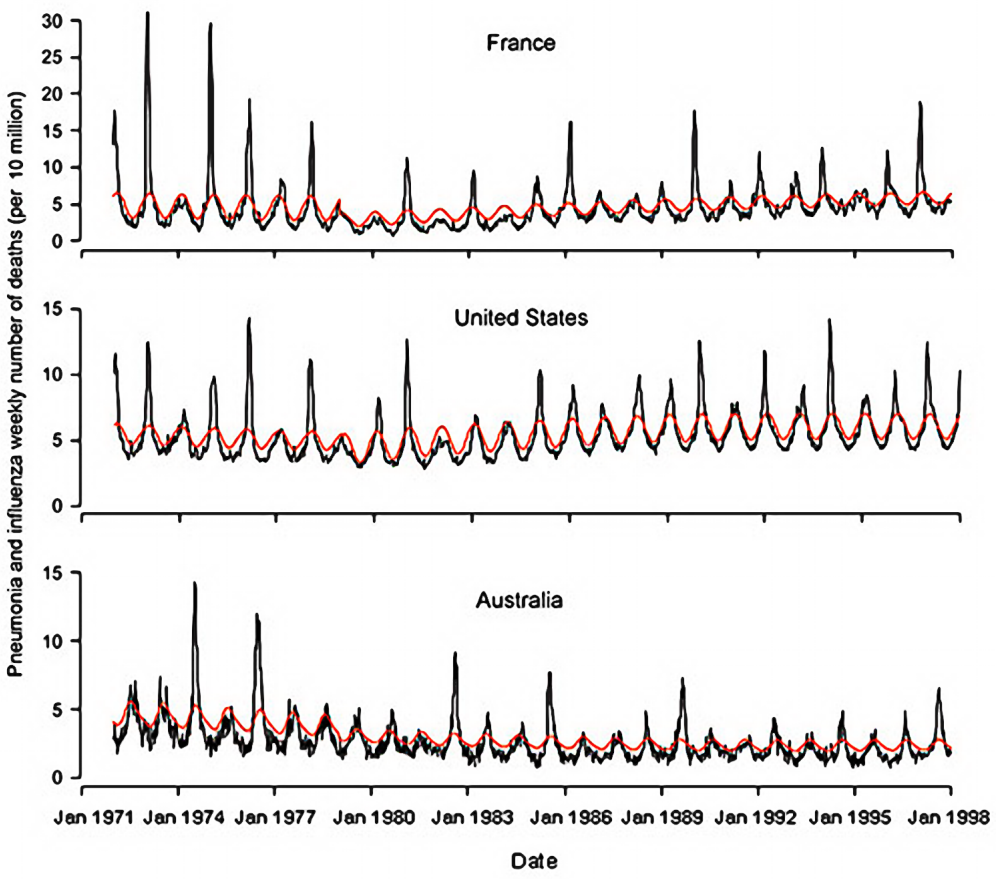


Figure 2.2: Weekly number of influenza and pneumonia deaths per 10 million populations from January 1972 to December 1997 in the United States, France, and Australia (black line). The red line represents the epidemic threshold defined by a seasonal regression. Source: (Viboud et al., 2004b)

at its culmination (approximately 23°27'N latitude). The Tropic of Capricorn is defined as the southernmost circle latitudinal parallel at which the sun may be seen directly overhead (approximately 23°27'S latitude). The subtropics are the areas lying immediately north and south of the tropics, bounded by latitude 35°N and 35°S and the Tropics of Cancer and Capricorn respectively. Whilst the area bounded by the Tropics of Cancer and Capricorn (the 'tropics') typically has year round high temperatures and periods of high precipitation, climate is actually quite variable within the area, being determined by altitude, proximity to water bodies and mountains, prevalent wind directions etc. Definitions of climate zones need to be much more sophisticated than areas bounded simply by latitudinal parallels and Figure 2.3 shows the Koppen-Geiger classification of climate, one of the most widely used climate classifications.

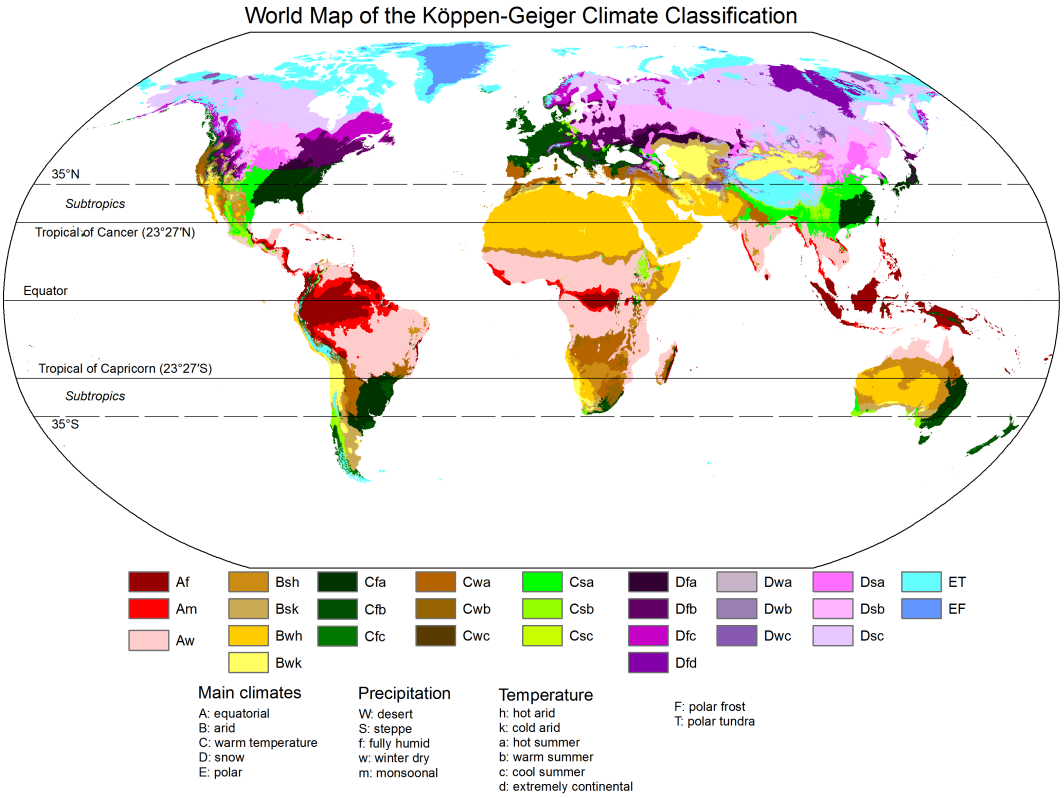


Figure 2.3: World Map of the Köppen-Geiger Climate. Source: (Rubel and Kottek, 2010)

Since traditionally influenza surveillance has not been prioritised in tropical regions to the extent it has in temperate regions, data are somewhat limited, but it is clear that

influenza seasonality is more complex in tropical and subtropical regions than it in temperate regions. In 2003 Shek and Lee summarised seven studies of influenza in tropical countries but six of the seven studies covered five or less years of data, so the validity of inferences on seasonality are limited so as the association with climate factors (Shek and Lee, 2003). Indeed in 2006 Viboud et al. noted that data on influenza burden and seasonality in tropical regions was inadequate and highlighted a need for more studies of influenza in the tropics Viboud et al. (2006a). In 2007, Park and Glass (2007) specifically looked at the seasonality of influenza A in east and southeast Asia and presented data from Japan (1998-2005), Taiwan (1997-99 and 2001-2003), Hong Kong (1997-2005) and Singapore (1972-1986, 1991-1993, 2000-2003). In Japan, which crosses the 35°N parallel that separates the subtropical and temperate regions, influenza A shows clear winter seasonality, with peaks between December and March. The situation in China has been confusing. The cities of Hong Kong and Shenzhen which lie in the subtropical region show seasonality of influenza transmission but contrasting patterns have been reported. The study Shenzhen (which use 5 year data of influenza confirmed) has been reported to have significant annual cycles of influenza and ILI activity in July-August, but also with a smaller second peak in March-April (Cheng et al., 2013). Whereas Hong Kong and Guangzhou, which are situated close to Shenzen to the south and north respectively and share a similar humid subtropical climate, experience dominant winter peaks of influenza activity. Since the data was limited in term of range (time-series of 5 and 3 year respectably) and the method in the latter 2 studies were fitted Poisson regression and modelling, the relationship between climate factors and disease pattern still unclear and need further studies (Yang et al., 2011, 2009). However, influenza throughout China has recently been studied in greater depth. Yu et al. with their 12 year time-series data found that the strength of the annual periodicity increases with increasing latitude (going north), with influenza epidemics peaking in January-February in northern China and April-June in southernmost China. Provinces in between experienced two peaks per year, in January-February and June-August. The tropical and subtropical provinces of China experienced influenza epidemics of longer duration and with a more variable timing of the epidemic peak compared to the northern provinces (Yu et al., 2013). Singapore, which

lies just north of the equator and has a truly tropical climate, has year-round influenza activity with some evidence of two peaks per year, in April-July and November-January (Tang et al., 2012). A systematic review of influenza in Africa by Gessner et al. identified only 4 studies in sub-Saharan Africa that looked at seasonality over several years, and these showed that influenza was seasonal in southern Africa (South Africa, Madagascar, and Zambia) but non-seasonal in tropical Senegal (Gessner et al., 2011). In South America, Brazil is best studied, where seasonality and amplitude of peaks of pneumonia and influenza mortality are highest in the southern areas, and decrease towards the equatorial regions (Alonso et al., 2007; Moura et al., 2009)

Bloom-Feshbach et al. examined the peak timing and duration of influenza epidemic activity (laboratory confirmed cases) in 77 sites in 40 countries spanning temperate, subtropical and tropical areas (Bloom-Feshbach et al., 2013). As expected, influenza epidemic activity was highly concentrated in winter months in the northern and southern temperate regions (modes of February and July respectively). The timing of peak influenza epidemic activity was more diverse in tropical sites, with semi-annual peaks identified in Manila, Philippines; Singapore; Nakhon Phanom and Sa Keao, Thailand; Ha Noi, Viet Nam; and Hong Kong, China (Figure 2.4). The semi-annual peaks tended to occur in winter and summer. The duration of influenza epidemic activity was longer (median 6 months) in tropical areas than in temperate areas (median 4 months). Baumgartner et al. reviewed publicly available data on virologically confirmed influenza infection from 85 countries and found that most of the temperate countries (40/47) and all of the subtropical countries (6/6) studied experienced a single annual influenza epidemic, compared to only 56% (18/32) of tropical countries (Azziz Baumgartner et al., 2012).

In summary, tropical and subtropical areas do experience epidemics of influenza but their timing is more variable than in temperate regions, can occur more than once a year, and the epidemics are often of smaller amplitude and longer duration than in temperate regions. There is also evidence that influenza epidemics occur at specific times of the year in some tropical areas, with annual or semi-annual periodicity (seasonality) then the approach to study seasonality of influenza in tropical and subtropical should not only focus on timing of the epidemic.

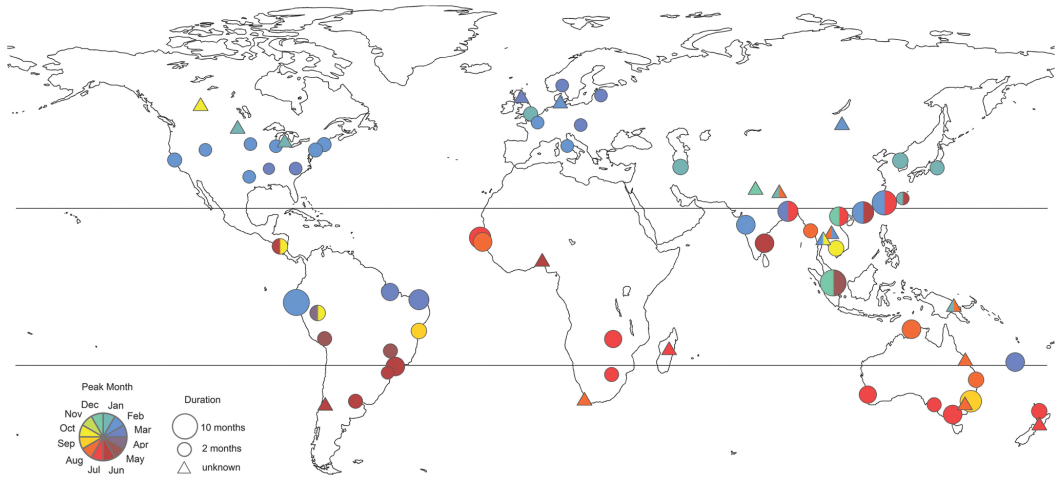


Figure 2.4: Global map of influenza peak timing and epidemic duration (n = 77 locations)
Colours illustrate timing of peak influenza activity, based on the bottom left key, while size of the circles is proportional to epidemic duration. For independent observations for the same location, an average was taken. For studies that did not provide enough information to estimate duration, a triangle is shown. Circles filled out with more than one colour represent locations experiencing semi-annual peaks of virus activity. Source: (Bloom-Feshbach et al., 2013)

2.4 Possible determinants of the seasonality of respiratory infections

Possible drivers of the seasonality of influenza infection have been reviewed in detail by several authors (Altizer et al., 2006; Dowell, 2001; Grassly and Fraser, 2006; Tamerius et al., 2010). In summary, three main classes of seasonally varying factors have been proposed as the drivers of influenza epidemiology. Host contact behaviours; host susceptibility; and pathogen survival outside of the host.

2.4.1 Host contact behaviours

Influenza is transmitted from person by direct contact, large respiratory droplets, by airborne small particles (Tellier, 2009; Teunis et al., 2010). The frequency and duration of contact between infectious and susceptible people may therefore affect the timing of influenza epidemics (Lofgren et al., 2007). It has been proposed that there is more indoor crowding in the winter in temperate regions and in the rainy season in tropical climates, and that this facilitates influenza epidemics (Jordan, 1961; Murray et al., 2012; Willem

et al., 2012). Children are known to play an important role in influenza transmission and the social mixing of children at school is thought to play a role in influenza transmission. Social contact rates of children are lower during holiday periods (Eames et al., 2011; Hens et al., 2009; Jackson et al., 2011) and there is empirical and simulated data to indicate that school closure may reduce influenza transmission (Cauchemez et al., 2008). It is not however established that school term times are an major determinant of the seasonality of influenza (Tamerius et al., 2010). Te Beest et al. did not find that the Christmas school holidays influenced influenza transmission in the Netherlands (te Beest et al., 2013).

2.4.2 Host susceptibility

Seasonal fluctuations in host immune functions have been suggested as drivers of the seasonality of infectious diseases, mediated by concentrations of substances such as cortisone, melatonin, and vitamin D (Dowell, 2001). Vitamin D has acquired the most attention (Cannell et al., 2006, 2008; Grant and Giovannucci, 2009; Juzeniene et al., 2010) Whilst solar radiation is associated with influenza activity, variations in serum vitamin D concentrations and solar radiation are not as good predictors of influenza epidemiology as other climate variables (humidity and temperature) (Shaman et al., 2011a,b, 2010; Tamerius et al., 2013; Yu et al., 2013).

2.4.3 Pathogen survival outside of the host

Influenza can be transmitted by large respiratory droplets, airborne small particles (aerosols) and by direct contact. The duration of survival of influenza viruses in the environment is influenced by environmental conditions such as temperature, humidity and concentration of ultraviolet light. These factors are discussed in greater detail in the following sections.

2.5 Relative versus absolute humidity

The concept of humidity is central to the study of influenza seasonality and therefore a brief introduction is provided here. Water in air can be in an invisible gas phase, known

as water vapour, or in a visible liquid phase as small suspended droplets, visible as mist, steam, or cloud. The higher the temperature of air, the more water vapour the air can hold before the water vapour condenses (becomes liquid phase) i.e. as condensation, mist, cloud, dew or frost. The maximum amount of water vapour a given volume of air can hold at a given temperature before condensation occurs is known as the saturation level.

Absolute (or specific) humidity (AH) is simply the absolute mass of water vapour (in grams) present in a given volume of air and water vapour mixture (g/m^3). AH expressed as g/m^3 is also known as the vapour density. Absolute humidity takes no account of the temperature of the air or the amount of water vapour the air could potentially hold. Absolute humidity can also be expressed as the pressure exerted by the water vapour (in any units of pressure e.g. kiloPascals, millibars, mmHg): this is the vapour pressure. Vapor pressure can be converted to vapour density (g/m^3) by the formula 2.1.

$$AH(g/m^3) = C \times eA/tK \tag{2.1}$$

C = Constant 2.16679 gK/J

eA = Vapour pressure in Pa

tK = Temperature in Kelvin

Relative humidity (RH) is the amount of water vapour in the air relative to the amount of water vapour that the air could maximally hold at a given temperature. Or in other words, RH is the ratio of the current water vapour content of a given volume of air relative to the maximum water vapour content of the volume of air at a given temperature (i.e. when the air is saturated). RH is expressed as a percentage. As temperature decreases, the saturation level reduces, so the relative humidity increases, even though the AH does not change. RH is more closely related to how we feel humidity, since sweat can readily evaporate from our skin if the relative humidity is low but cannot if the relative humidity is high; so we feel warmer in higher relative humidity. RH is therefore more often used in

weather forecasts and is the measure we are sensing when we say it is "humid". AH can be calculated from RH and temperature (see method section of Chapter 4).

2.6 Experimental studies of climatic determinants of influenza survival and transmission

Early experimental studies of the survival and transmission of influenza under varying environmental conditions have been summarised by Tamerius et al. (Tamerius et al., 2010). A common finding in these studies was that influenza virus viability increased with decreasing relative humidity (Harper, 1961; Hemmes et al., 1960; Hood, 1963; Kingdon, 1960; Schaffer et al., 1976; Schulman and Kilbourne, 1963). In 2007 and 2008, Lowen et al. revisited these experiments by studying the aerosol transmission of influenza virus in a guinea pig model under difference environmental settings (relative humidities of 20%, 35%, 50%, 65% and 80%, and temperatures of 5°C, 20°C and 30°C. They found transmission to be most efficient at low relative humidity (20% and 35%) and low temperature (5°C) (Lowen et al., 2007). Outdoor RH is high during the winter in temperate regions, and therefore does not fit the model, but indoor RH is low in the winter in places with indoor heating (see Shaman and Kohn), so Lowen et al. hypothesize that low indoor RH in the winter contributes to the winter peaks of influenza transmission. Since the initial experiments demonstrated no transmission at a temperature of 30°C and RH of 35%, Lowen et al. conducted further experiments at high temperatures and found that at a temperature of 30°C and RHs of 20%, 50%, 65%, and 80%, aerosol transmission was blocked, but contact transmission was not (Lowen et al., 2008). This has led Lowen and Palese to hypothesize that in temperate regions aerosol transmission predominates whilst in tropical regions transmission is by direct contact (Lowen and Palese, 2009). Shaman and Kohn re-examined Lowen's 2007 and 2008 data by calculating AH (vapour pressure) from the RH and temperature, and found that AH was a better predictor of virus transmission than either temperature or RH (Shaman and Kohn, 2009). As AH decreases, transmission increases: an analysis which fits climate data better than Lowen's analysis since outdoor AH (unlike RH) is low in the winter. It is hypothesized that this relationship exists

because low AH causes water to evaporate from expelled droplets, leading the formation of smaller droplets that remain airborne for longer (therefore increasing transmission), and that viruses in airborne particles remain viable for longer at low AH (Shaman and Kohn, 2009). Several authors have explored the mechanism of the relationship between humidity and influenza virus transmission/survival. Yang et al. figure out the important of humidity in aerosol transmission because "it both induces droplet size transformation and affects influenza A virus inactivation rates" (Yang et al., 2011, 2012; Yang and Marr, 2011). A theoretical study by Minhaz Ud-Dean looked at the impact of humidity to the influenza virus envelope and proposed that virus envelop have certain characteristic that determine the interaction with atmospheric under the hosts condition that finally create seasonality in the temperate region. In tropical region, virus can survive longer then create higher risk with aerosol borne epidemic (Minhaz Ud-Dean, 2010)

2.7 Observational studies of climate variables and influenza transmission

The relationship between climate variables and influenza epidemics has been explored through observational studies in many countries. A review by Shek et al. in 2003 suggested that epidemic peaks in Singapore, India, Nigeria, Brazil and Senegal were associated with the rainy season (Shek and Lee, 2003). However, Tang et al. recently summarized the data from Singapore, and concluded that the results were conflicting, with some authors finding no association with climate variables, whilst others report an association with rainfall (Tang et al., 2012). Chan et al. studying 10 years of data on hospital admissions in Hong Kong have reported that influenza A and B activity was associated with cold temperatures and high relative humidity (>70%), this conflicts with the findings of others that low relative humidity is associated with influenza transmission and survival (Chan et al., 2009).

The work of Alonso in Brazil has been an important contribution to the literature, with a travelling wave of influenza from north to south identified, which Alonso et al. have suggested is more likely to be a consequence of climate than population density or

travel, since the north of Brazil is less densely populated than the south, and travel is greatest between the large population centres in the south (Alonso et al., 2007). Other studies that examined the seasonality of influenza in South America include a study in northeast Brazil that identified annual epidemics occurring in the rainy season (Moura et al., 2009), one in French Guiana, which concluded that influenza transmission was associated with high rainfall and low AH (Mahamat et al., 2013), and one in Peru, but this study was only of two years duration (Laguna-Torres et al., 2009).

Studies of climate associations with influenza activity in Africa are limited despite a series of studies in 15 African countries (Katz et al., 2012b; Radin et al., 2012). Whilst studies in Kenya, Uganda, Rwanda have suggested that influenza activity peaks in the rainy season, these studies are of only 2 or three years duration and encompassed the 2009 pandemic, and therefore are not able to identify seasonal trends or climatic predictors with any certainty (Katz et al., 2012a; Lutwama et al., 2012; Nyatanyi et al., 2012).

The association between climate variables and influenza epidemiology in temperate regions has been best studied in the United States. In 2010 Shaman et al. examined the timing of the onset of increased seasonal influenza activity over a 31 year period (1972-2002) and found a strong association with a drop in AH, which was statistically stronger than the association with RH, temperature and sunshine (Shaman et al., 2010). They further found that a mathematical model based on AH alone could accurately reproduce the spatial and temporal epidemiology of seasonal influenza in the United States. Barreca and Shimshank also studied influenza mortality between 1973 and 2002 in the United States and, like Shaman et al., found that AH was the best explanatory variable, after controlling for temperature (Barreca and Shimshack, 2012). In addition however, Barreca and Shimshank identified an AH threshold (6g/kg) above which there was no association between AH and influenza associated mortality. Interestingly, the first assertion that AH may be related to influenza transmission was in 1985 by Shoji (in Japanese), but this paper was largely neglected, and the data were republished by Shoji in 2011 following the work of Shaman on AH (Shoji et al., 2011).

A recent publication has looked at laboratory confirmed influenza infections from 2005 to 2011 across 30 Provinces of China, an area that spans temperate, subtropical and

tropical climates. Yu et al. found no single climate variable explained the patterns of influenza activity across the whole of China. Cold temperatures were associated with a strong annual amplitude of influenza, whereas high rainfall was predictive of the timing of spring epidemics, and low sunshine was associated with semi-annual cycles (Yu et al., 2013).

Tamerius et al. have studied global climatic predictors of influenza seasonality using the same influenza data set reported by Bloom-Feshbach et al., as discussed above (Tamerius et al., 2013); (Bloom-Feshbach et al., 2013). The authors studied the association between the timing (month) of peak influenza activity and monthly outdoor temperature, solar radiation, AH, and rainfall at 78 sites, of which 39% were in the tropics. They did not find a consistent relationship at the global level but identified two distinct patterns: in high latitudes influenza peaks were associated with low temperature, solar radiation and AH, whereas at low latitudes, influenza peaks were associated with high rainfall, AH and RH. They therefore proposed that in temperate regions influenza epidemics occur in months characterized by 'cold-dry' weather (AH $<$ 11-12 g/kg and temp 18-21°C), whereas in the tropical regions influenza epidemics are precipitated by 'humid-rainy' weather (months with average precipitation greater than 150 mm). The relationship between AH and influenza peaks was found to be U-shaped, with the highest probabilities of influenza peaks occurring at low and high values of AH. In general however, the model did not perform well in low latitude regions.

2.8 Summary

In summary, in contrast with the clear pattern of influenza found in temperate areas, knowledge on the seasonality of influenza in tropical and subtropical areas is still limited. The limitations arise from either the length of data or study approach. Recent years have seen great steps in the study of influenza in tropical/subtropical areas especially in China and Hong Kong. This work has found a latitude gradient in influenza dynamics (Yu et al., 2013) and highlighted the role of humidity in influenza transmission activity (Yang et al., 2011, 2009). Other recent studies also describe seasonality in tropical areas which varies

from country to country (Azziz Baumgartner et al., 2012; Bloom-Feshbach et al., 2013). Researchers all over the world also try to explain the possible drivers of seasonality by either using animal model or lab-experiment e.g. (Lowen et al., 2007; Shaman et al., 2010), or time-series analysis with wavelet (Alonso et al., 2007; Viboud et al., 2006b; Yu et al., 2013), or systematic review, reanalysis (Tamerius et al., 2013) which all lead to humidity factor.

2.9 Methods for assessing the seasonality of influenza

As mentioned earlier, many countries in temperate regions have well-structured influenza surveillance data and influenza epidemics mostly occur regularly in well defined seasons. However, in the tropics and subtropics, both ILI notifications and influenza virology surveillance data can exhibit less regular patterns, with the timing of peaks being less predictable (Viboud et al., 2006b). In time-series analysis, a signal whose frequency does not change over time is called *stationary,* and a signal whose frequency does change over time is *non-stationary.* Contemporary statistical techniques such as wavelet analysis may be usefully applied to problems of analysing non-stationary time series. Wavelet analysis was first adopted in the field of climatology (Lau and Weng, 1995; Torrence and Webster, 1999) and has only recently been introduced into the ecology of infectious diseases such as measles (Grenfell et al., 2001), pertussis (Broutin et al., 2005), dengue fever (Cazelles et al., 2005) and influenza (Viboud et al., 2006b). By decomposing a time series into the various time-frequency spaces, wavelet analysis is more suitable for modelling non-stationary seasonality of a time series than traditional Fourier analysis; therefore in this study we used wavelet analysis to model the seasonal variation of ILI rates, climate variables (Chapters 3 and 4) and virology surveillance data (Chapter 5).

Chapter 3

Spatial patterns of ILI seasonality in Viet Nam

3.1 Introduction

In this chapter, I describe my work to examine geographical patterns in the seasonality of ILI notifications in Viet Nam. The primary rationale for this work is to describe the seasonality of ILI in order to inform the design of influenza control programmes in Viet Nam. Viet Nam is working towards the development of domestic influenza vaccine production capacity (see Chapter 1), and knowledge of the periodicity and predictability of influenza transmission is essential for designing an influenza immunization program. Since it is necessary to update the influenza strains included in inactivated influenza vaccines annually because of on-going antigenic changes in circulating influenza viruses (see Chapter 1), program design will need to consider the timing of selection of influenza vaccine strains. This need to annually update influenza vaccines means that strains must be selected every year and the selection process must be timed to permit the vaccine production cycle to be completed in time to vaccinate the target population prior to the period of influenza transmission. Whilst seasonal epidemiological dynamics are common to many vaccine-preventable infectious diseases, the need to update influenza vaccines annually and re-immunise the target population places a unique time constraint on influenza immunisation programs. A secondary objective of this analysis is to assess if there are geographical patterns of seasonality in Viet Nam that might be used to inform our understanding of the climatic determinants of influenza transmission. As will be discussed in Chapter 4, there are substantial differences in climate between North and South Viet Nam, and also an anecdotal belief that the seasonality of influenza epidemics is more marked in the North than in the South of Viet Nam. The first step in exploring the association

between influenza seasonality and climate in Viet Nam is to establish if there are systematic geographic differences in influenza seasonality.

3.2 Objective

The objective of the work presented in this chapter is to describe the spatial patterns of seasonality of ILI notifications in Viet Nam.

3.3 Materials and Methods

3.3.1 Data sources

Influenza-Like-Illness (ILI) notification data

"Influenza syndrome" is one of the 26 notifiable communicable diseases in Viet Nam (see Annex 1 for details of the surveillance system in Viet Nam). The case definition for "influenza syndrome" is*: "Sudden onset of fever:39-40° C with severe headache or body, muscle and joint pain and runny nose/ sore throat/ coughing"* (Ministry of Health, 2010). The definition is a clinical definition with no requirement for laboratory diagnosis. The number of notified cases of 'influenza syndrome' has been monitored by the national surveillance system of the Ministry of Health by month and province since 1979, over which period the case definition has not changed. Throughout this Chapter and Chapter 4, the term Influenza-Like-Illness (ILI) is referring to notifications of 'Influenza Syndrome'. It should be noted that this definition has some difference with the ILI case definition used in the sentinel influenza surveillance system (see chapter 5), which follow the CDC & WHO case definition. The number of notifications of ILI per province per month (the notification rate) is the raw data for the analysis conducted in Chapters 3 and 4. Data were independently double entered into Excel (Microsoft Office 2007) and then cross-checked by "Excel file compare 2.4" (`http://www.formulasoft.com/`) and any discrepancies checked against the original paper records and corrected as necessary. Final corrected datasets were then read into R version 2.14.0 (R Foundation for Statistical Computing, Vienna, Austria). Since reliable annual population estimates by province from 1980 to 2010 are not

available, and since the objective of the analysis was to describe spatial patterns of annual periodicity (seasonality) of ILI notification rates, rather than compare the incidence of ILI notifications per person-time between provinces, we did not calculate notification rates per person-time.

Geographic units of analysis

The province is the first level administrative unit in Viet Nam, below which there are districts, and below them communes. There are roughly 10 (range from 5-27) districts per province and 16 communes per district (range from 1 (an island) to 48) (Figure: 3.1).

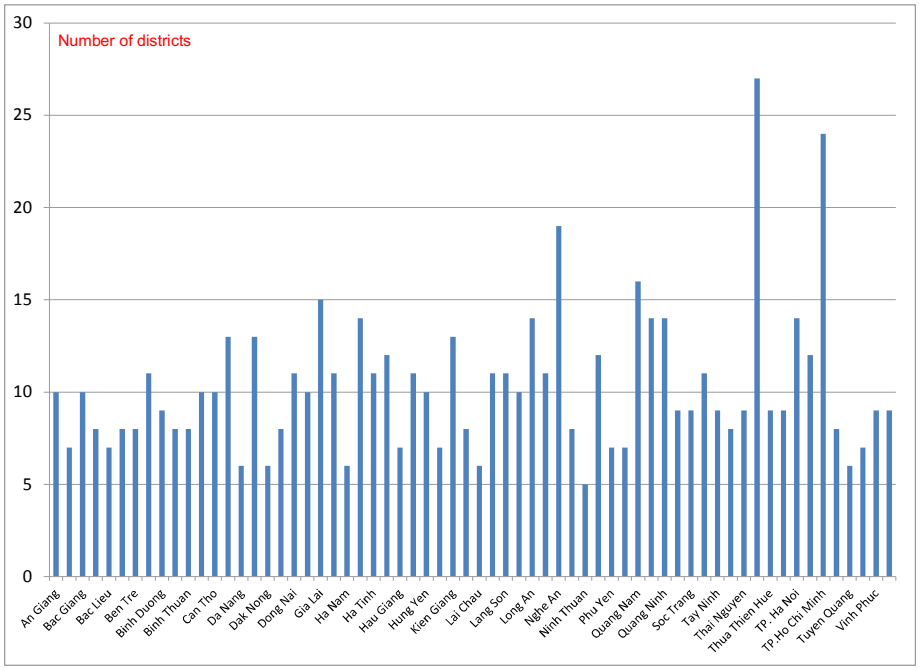


Figure 3.1: Number of districts per province. Each column is one province (63 in total).

The administrative delimitations of provinces in Viet Nam have changed substantially since 1980. A number of provinces have split in to 2 (and one into 3) provinces on the first of January 1990 (3 splits, including the one in three provinces), 1991 (1 split), 1992 (8 splits), 1997 (8 splits), 2004 (3 splits). On the first of August 2008 the province of Ha Tay (originating from the split of the province of Ha Son Binh in 1992) merged with the

province of Ha Noi. This is the sole merging event. The number of provinces in Viet Nam thus increased from 40 to 44 in 1990, to 45 in 1991 to 53 in 1992 to 61 in 1997 and to 64 in 2004, and decreased to 63, the current number in 2008. Figure 3.2 and 3.3 shows the total number of provinces and the merging and splitting of provinces over time. In this figure the year in which a change in province numbers occurred are marked by vertical grey lines. In order to establish a dataset of geographic units that could be analysed over the whole study period, it was necessary to account for the merging and splitting of provinces. For provinces that were merged we aggregate the two time-series into one. For province that split into 2 new smaller provinces, we keep all cases in the province which have higher population and left the other province blank data.

Population centroid of province

Population centroids (the latitude and longitude coordinates that mark the estimated centre of population density) of provinces are required in order to allocate a specific latitude value to each province and, in Chapter 4, to identify the climate station geographically closest to the main population centre. Provincial population centroids were calculated using each communes geographic centroid and population size to generate an average of the commune centroids weighted by the commune population size. Commune level population data were obtained directly from the decennial national Population and Housing Census in 2009 (see detail in central population and housing steering committee (2010)), conducted by the General Statistic Office of Viet Nam (GSO;`http://www.gso.gov.vn`) (GSO, 2012). These population centroids were used as the geographical coordinates of the time series. (see R code in annex for more details of the methodology). The estimated province population centroid was an average of 11.8 km away from the province geographic centroid. This difference was greater in large provinces and provinces with mixed terrain (e.g Nghe An, a coastal province with sparsely populated inland mountainous areas, has a difference of 57.7 km), whereas small and flat provinces have a much smaller difference (e.g Hung Yen and Thai Binh provinces had differences of 1.1 and 1.2 Km respectively) see Figure: 3.4.

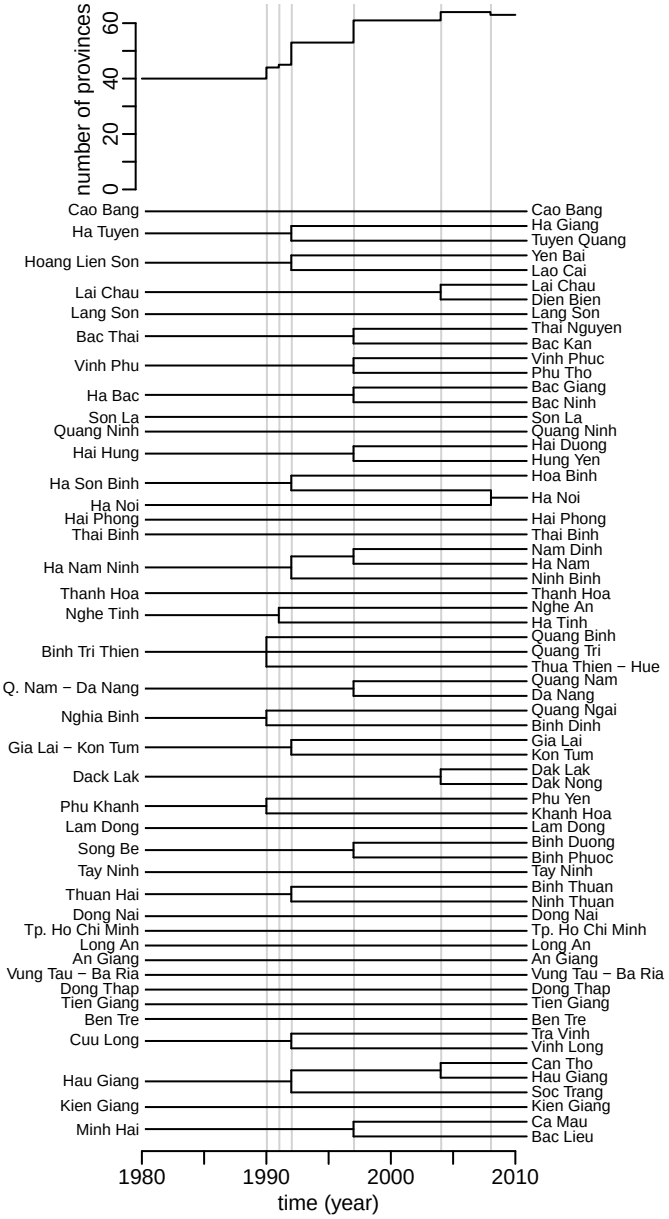


Figure 3.2: Changes to provinces between 1980 and 2010

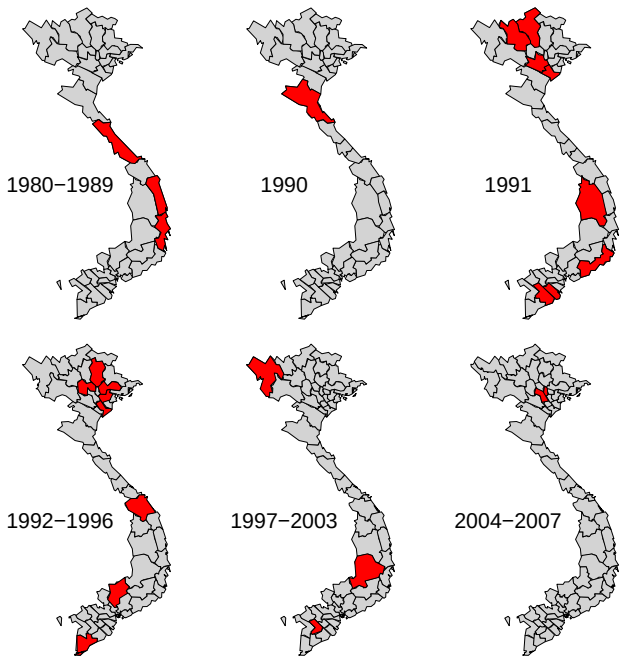


Figure 3.3: Map of changes to provinces between 1980 and 2010.
Provinces that underwent boundary changes in the following time period are marked in red.

The analysis presented in the Chapter is a time series analyses (to characterize periodicity) in different localities (provinces). For such an analysis to be efficient we ideally need long and numerous time series. However, given the history of administrative divisions in Viet Nam (mostly splitting events), the duration and the number of the time series cannot be optimized at the same time: the earlier we start the analysis the smaller the number of provinces but the longer the time series they display. The picture is further complicated by the fact that missing values are more numerous in the earlier years than the recent years. Missing values can be linearly interpolated (linear interpolation uses the equation of a straight line ($y = mx + c$) to fill in missing data points between two known points) but the validity of the method decreases when the number of consecutive missing values to interpolate increases. In order to select the starting year that optimizes the total amount of information in the data set, we plotted the number of notification as a function of the starting date. In doing so, the rule was to discard any time series that had more than a given number of consecutive missing values, this number varying from 0 to the length of the time series (this corresponds to the different lines on figure 3.10 panel B). In practice,

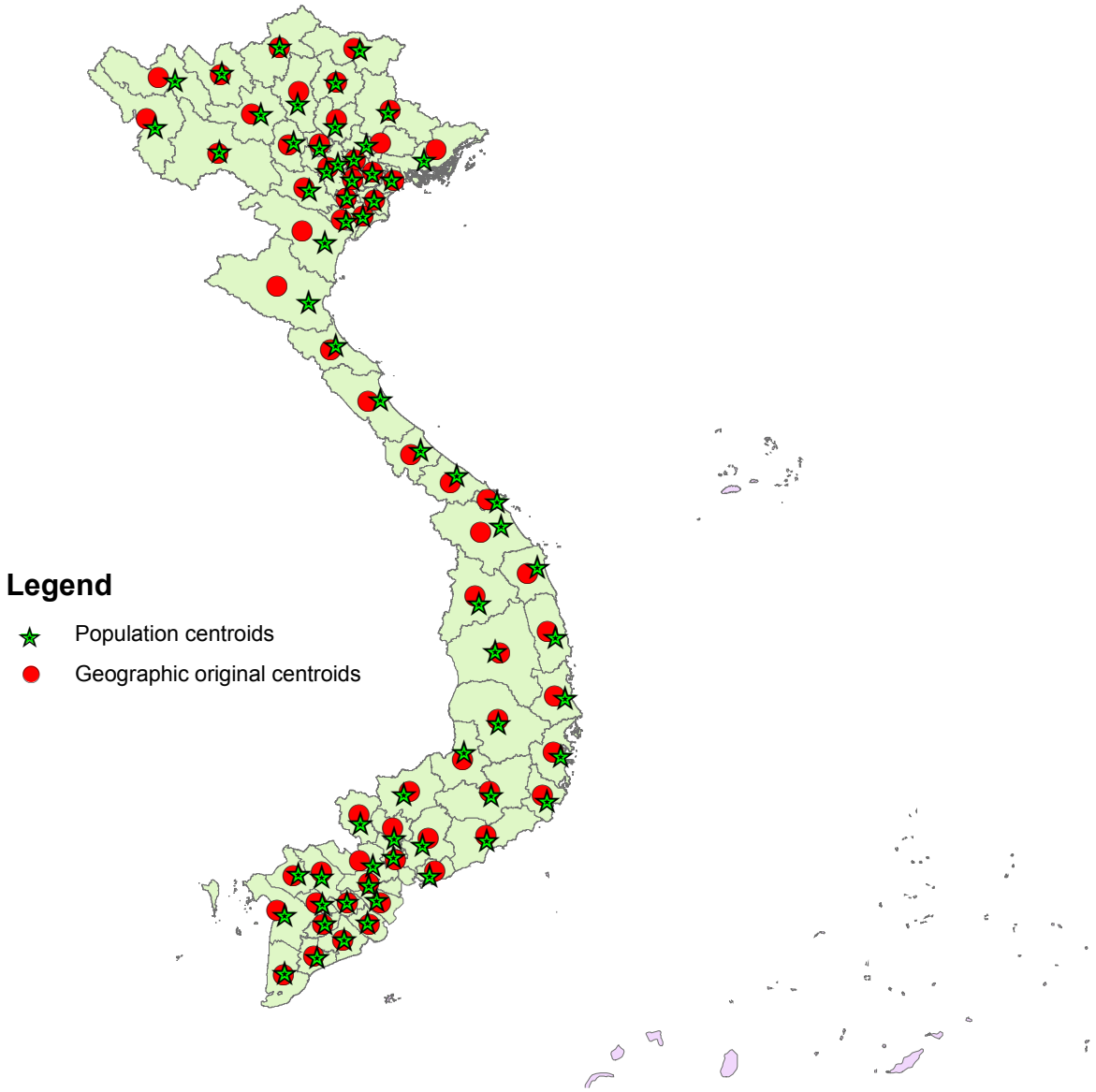


Figure 3.4: Map of provincial population and geographic centroids.
Location of provincial population centroids (stars) versus geographic centroids (circles).

more than 6 consecutive missing values render any interpolation very speculative.

Missing data

Time series with more than 3 consecutive missing values (i.e. one quarter of a year) were discarded from the analysis. The remaining missing values were linearly interpolated using the R function "approx".

Transformation, detrending, centering and reducing

The ILI surveillance data have a number of quality issues due to under- and over-reporting (confusion with other diseases), and changes in the surveillance system effort over time. These issues limit the validity of any quantitative analyses (based on the actual values of the notification). Given these facts, our strategy was to develop qualitative analyses based on qualitative statistics such as periodicities which are less affected by the potential biases mentioned above. The transformation presented here are prerequisite to these qualitative analyses.

The variance of the time series can basically be decomposed between seasonality (i.e. periodicity of one year), longer-term trends, and the remaining that we call 'noise'. Since our analysis focuses on the seasonality, a first step is to remove the long-term trend. Another reason to detrend the time series is that we don't know the cause of it, in particular whether it has real biological value or weather it is simply due to changes in the efficiency of the reporting system over time. The trend was estimated by lowess regression with a smoother parameter of 0.1 (value selected after trying a range of values). The time series were detrended by subtracting their estimated trends.

After detrending, the time series were centered and reduced. Centering refers to the removal of the mean value, whereas reducing refers to the scaling by the standard deviation. Centering and reducing of the time series thus produces so-called z-scores, i.e. time series with means of 0 and standard deviations of 1, thus rendering qualitative comparison of these time series possible. Centering and reducing is also referred to as 'normalization' of the data since the data have a normal distribution after this process.

Finally, the time series were square-root transformed. Population dynamics time series

such as epidemiological ones are typically characterized by a high number of small values and a small number of high values. In order to stabilize the variance (i.e. rebalanced the different values), a common practice is to square-root transform the data (a square-root transformation is a power-transformation that is less stringent that the logarithm one).

Detection and management of outliers

Due to the volume and heterogeneity in the data aggregated for the whole country, it is difficult to detect outliers. However, when the data are viewed by individual province, potential outliers are clearer. Figure 3.13 shows the number of ILI notifications over time from Tra Vinh Province (the raw data from every Province are shown in annex B). Possible outliers are marked in red. These data points may be mistakes in the aggregation of data at the province level not during data entry since data were double-entry-checked. Since such outliers can substantially affect the analyses and in particular the wavelet spectra, we sought to identify and check or discard outliers. To identify outliers in a systematic and reproducible way, we considered, for each province, the square root transformed, (lowess regression with smoothing parameter of 0.1) then scaled (i.e. subtracted by the mean and divided by the standard deviation) Then we calculated the differences between each consecutive values and these differences are expected to roughly follow a normal distribution. And identified as outliers any values below the 1^{st} percentile and above the 99^{th} percentile. Outliers will produce differences that are far away in the tails of the distribution. Note that this automatic procedure successfully identify the outliers. However some points identified by this procedure are evidently not outliers. Then, first, this is still the best way we found to identify outliers, second, we identify very few "false" outliers, and, third, identifying false outliers is not much a problem since outliers values were then checked against with the original hard-copy records or other source (i.e province data record). The value was then corrected if it had been entered incorrectly, or was otherwise discarded. Those discarded was replaced by their linearly interpolated value. False outliers will thus be replaced by a value that is very close to their original value.

Time series components.

A time series can be characterized by 4 main statistics: the mean (M in equation 3.1), the amplitude (A in equation 3.1), the period (T in equation 3.1) and the phase (φ in equation 3.1, see also figure 3.5).

$$I(t) = M + A\cos\left(2\pi\frac{t}{T} + \varphi\right) \tag{3.1}$$

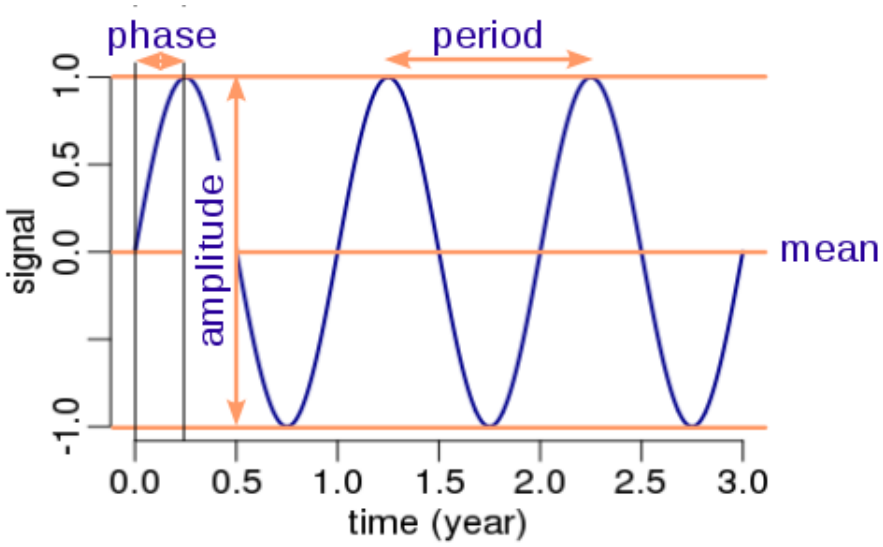


Figure 3.5: Time-series components
Example of sin wave showing relation between difference components described in the text

Less technically and in the context of epidemiology, the mean refers for example to the average incidence per year, the amplitude refers to the magnitude of the epidemics, the period refers to how often epidemics occur (e.g. every year, every two years, etc., or not periodic at all), and the phase refers to the timing of epidemics (i.e. the month of the year the epidemics tend to occur). Note that in the case a time series made of two periodic signals with different periods (e.g. annual plus multi-annual), then each signal has a phase (i.e. a phase refers to a given period).

The coherence between two wavelets is equivalent to a linear correlation measure between them. It is constructed similarly and lives between 0 (no correlation) and 1 (perfect correlation). It informs on when and around what period two time series are better linearly correlated.

Wavelet decomposition.

For the purpose of analysing periodicity we are treating the ILI notification as signal data a time varying continuous quantity. Signal data may appear 'noisy' if it is a composition of multiple random and non-random components. Signal decomposition is the application of methods to filter signal data in order to identify and separate underlying components. One of the most widely used signal decomposition tools is Fourier transformation. Fourier transformation decomposes a signal into its component cycles (a regular oscillation of amplitude) using sinusoidal functions. However, a major constraint of the basic Fourier transformation is that it assumes that the signal is stationary (i.e. statistics such as mean, variance, periodicity are constant in time) and does not allow easy observation of changes over time in the cycle periodicity (the frequency of oscillations). A signal whose frequency does not change over time is called *stationary,* and a signal whose frequency does change over time is *non-stationary.* Figure 3.6 is an example of a non-stationary signal, where the frequency of measles epidemics decreased after the introduction of measles vaccination. Instead of decomposing the signal on sinusoids like Fourier decomposition, wavelet decomposition decomposes a signal as wavelet basis functions. Wavelets are basically sinusoids with an envelope. For example the Morlet wavelet classically used in ecology is a sine wave inserted into a Gaussian envelop (see figure 3.6). Being non-infinite, the decomposition will necessarily be local in time and frequency. A wavelet can be slid to cover all the time periods and stretched and compressed to cover of the period range. This locality in decomposition, make the wavelet decomposition able to deal efficiently with non-stationary time series as often encountered in epidemiology. The method transforms a signal $f(t)$ into a function $W(a,b)$ which illustrates the different frequency components at time (t) according to equation 3.2.

$$W\left(a,b\right)=\frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty}x(t)\varphi^{*}\left(\frac{t-b}{a}\right)dt=\int_{-\infty}^{+\infty}x(t)\varphi_{a,b}^{*}(t)dt \tag{3.2}$$

In which, the * is complex conjugate form. $W\left(a,b\right)$ is wavelet coefficients when "a" is the wavelet frequency (scale) and "b" is the difference time position. x(t) is formula of

time t and φ is scaling function or the wavelet.

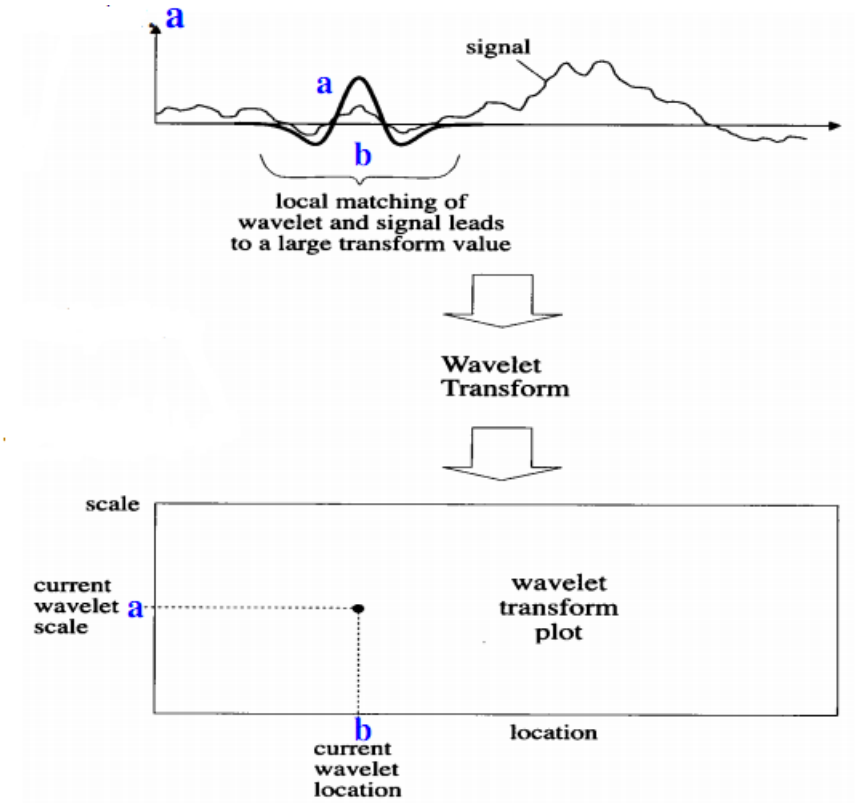


Figure 3.6: Wavelet transformation of a time-series
In which, "a" is frequency resolution (scale factor) and "b" is the shift coefficient. Signal is multiplied with a wavelet function, then analysed the frequency at different times. This figure only show one point of estimation. After the wavelet transform in coordinates W (ai, b) with i =1, 2, 3 ... n we obtain the set of points in rows which show which frequency component appears at the time "*t*". When "b" changes with i =1, 2, 3 ... n we obtain a set of points in columns. These values are expressed in the spectral method with the gradually bold colours depending on the rule and frequency. Identifying the high frequency appearing at a time help us determine the seasonality of epidemic. See Figure 3.7

The local (time specific) values are expressed in colour, where darker colours represent a stronger frequency signal (based on the variance in the signal relative to all the other signals). In wavelet analysis the relative strength of the frequency signal is termed the *power*, and has an arbitrary scale. Since the power is represented as a spectrum of colours, it is termed the *power spectrum*. Frequency signals cannot be detected with certainty at the beginning and end of the time-series since there are insufficient comparison values going beyond the time-series (termed edge-effects), therefore the Wavelet figure draws a "cone of influence" in the region of the wavelet spectrum where edge-effects make the data unreliable (Torrence and Compo, 1998). The longer the time series, the smaller

the proportion of region that is affected by edge effects. As our time series has more than 200 time points, edge-effects are unlikely to affect the interpretation of the data. Wavelet analyses were performed by using the Morlet wavelet, classically used in ecology, with a non-dimensional frequency (Cazelles et al., 2008, 2007). We used the "biwavelet" R package to perform the wavelet analyses (Gouhier and Grinsted, 2013). A major advantage of using the Morlet wavelet is that it is a complex one. The consequence of this property is that the expression of the phase of the time series (i.e. its timing) is straightforward. The use of the Morlet wavelet thus allows to calculate the phase at any point in time and for any frequency of the signal.

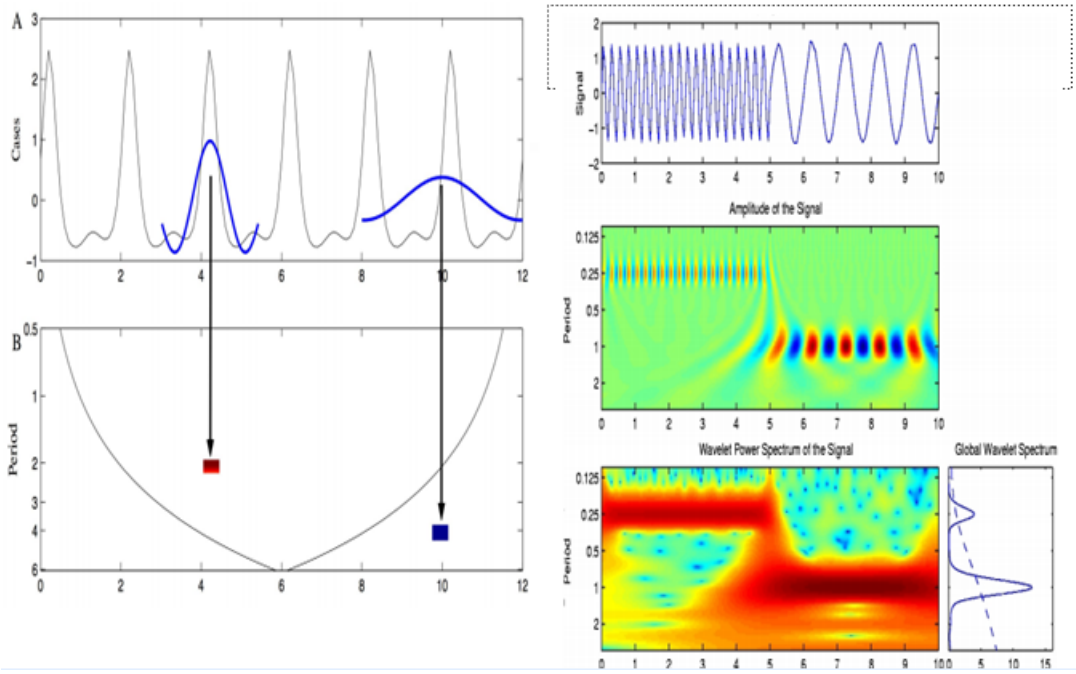


Figure 3.7: Time-dependent spectral analysis of epidemiological time-series with wavelets. Source: (Cazelles et al., 2008, 2007)
Continue from figure 3.6 for one point estimation. If the point has high power (a) then colour code will be red and reverse with blue for low power. For n times of estimation, The wavelet spectrum created where a band of high power (dark red) will indicate the range of coherence or seasonal

To simplify, a local wavelet power spectrum is a matrix with a temporal and a frequency dimensions. If we sum this spectrum over the temporal dimension we get a *global wavelet power spectrum*, which is analogous to a Fourier spectrum. So Fourier decomposition can be considered as a special case of wavelet decomposition. In our analysis, we will use these global wavelet power spectra. In an extreme case, if sum the local wavelet power

spectrum over the frequency dimension, we obtain back the original time series. And if we do this summation only between two given frequency, we are filtering the time series. This wavelet method has been successfully used to analyse a variety of non-stationary infectious diseases time series data, starting with measles (Grenfell et al., 2001). We used wavelet decomposition to explore the periodicity structure of all the times series (from both the local and global power spectra), to filter them around the seasonal component (0.9 - 1.1 year band), and to compute the phase of the filtered time series. The maximum value of the global power spectrum within the annual band (0.9 - 1.1 year) was used as a measure of the strength of seasonality. Seasonality means the temporal structure over the period of one year (as opposed to random noise) and should not be confused with amplitude, as it sometimes is. Figure 3.8 shows that amplitude is a concept totally different from seasonality: within each row the amplitudes are the same and yet the seasonalities are totally different. There is only one other paper that quantifies seasonality of influenza this way (Yu et al., 2013).

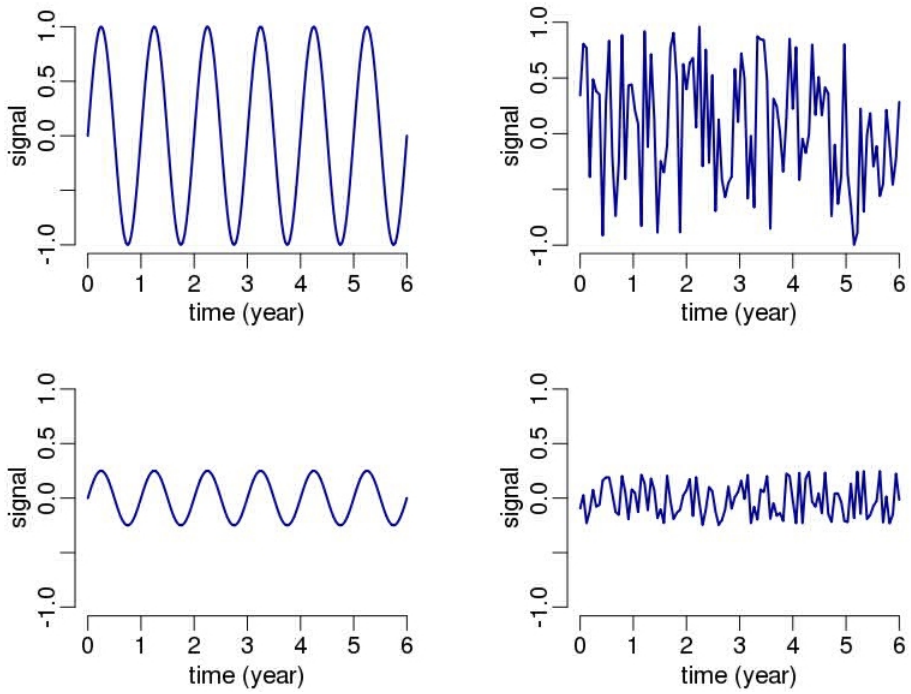


Figure 3.8: Seasonality versus amplitude
The first column shows perfectly seasonal signals (strictly structured in time) and the second column shows totally random signals. The reality is between these two extremes. The first row shows high amplitude signals and second row shows low amplitude ones

3.4 Results

3.4.1 Data selection

Figure 3.9 shows the overall time series of ILI notifications by month from 1980 to 2010. In order to better visualise province specific patterns, the data are represented in Figure 3.9 as a grid with provinces ordered in rows by the latitude of the population centroid of each province, from north at the top to south at the bottom, and the number of notifications each month colour coded from light yellow (lower number of notifications) to dark red (higher number of notifications) with missing data represented by black.

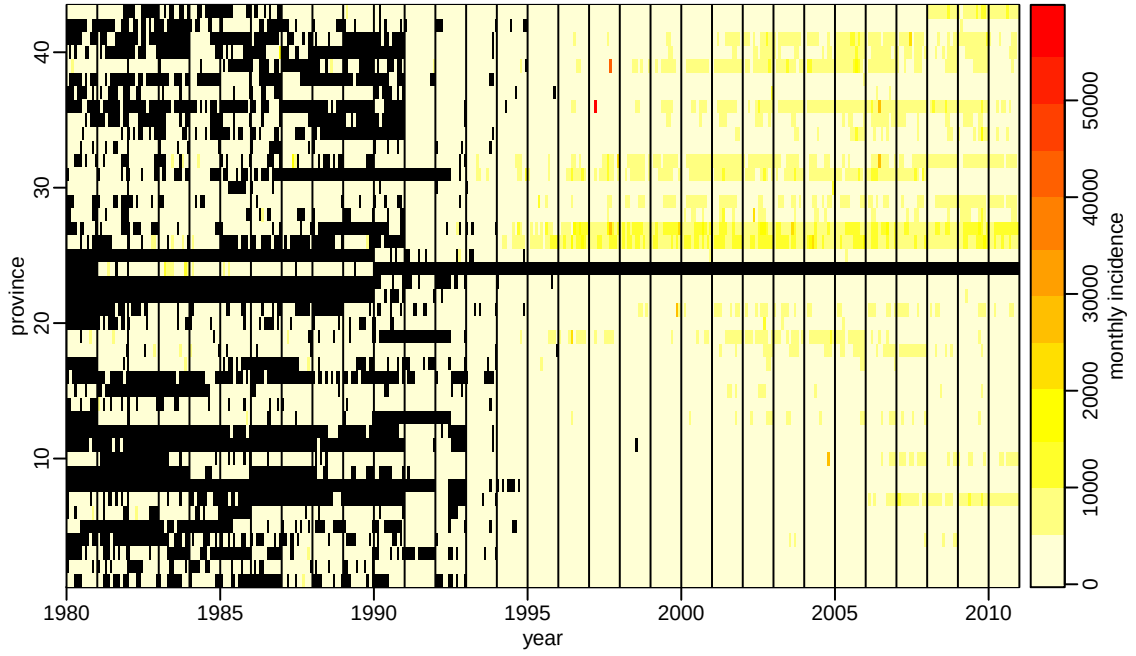


Figure 3.9: Raw ILI notification data
Monthly notifications range from 0 to 60000 per month and are colour coded from yellow (low) to red (high). Black represents missing data.

Since a lot of data are missing early in the time series, we selected the optimal number of provinces and the time frame as described in the methods. Figure 3.10 shows the number of provinces available (A) and the number of non-missing data available (B) as a function of year. The different lines (solid black and solid grey) show the effect of different choices regarding the number of consecutive missing values that are accepted. In both panels the bottom line shows the most stringent choice, i.e. For Panel A the bottom line shows the number of time series that are included over time if any time series with

a single missing value is removed from the data set. The next line shows what happens when time series with 2 or more consecutive missing values are removed, and so on until the upper-most line showing what happens when no time series is removed, regardless of the number of consecutive missing values. In Panel B the bottom line shows the number on non-missing data points included when any time series with a single missing value is removed from the data set. The next line shows what happens when time series with at 2 or more consecutive missing values are removed etc.

The time period that optimizes the amount of information available in the data set, in terms of (i) the number of time series (52 provinces), (ii) their length (216 months), and (iii) the number of non-missing values, is from 1993-2010. Therefore we included time series only from a point where there are no more than three consecutive missing months of notifications. The province definition used in our analysis is thus the one of 1993.

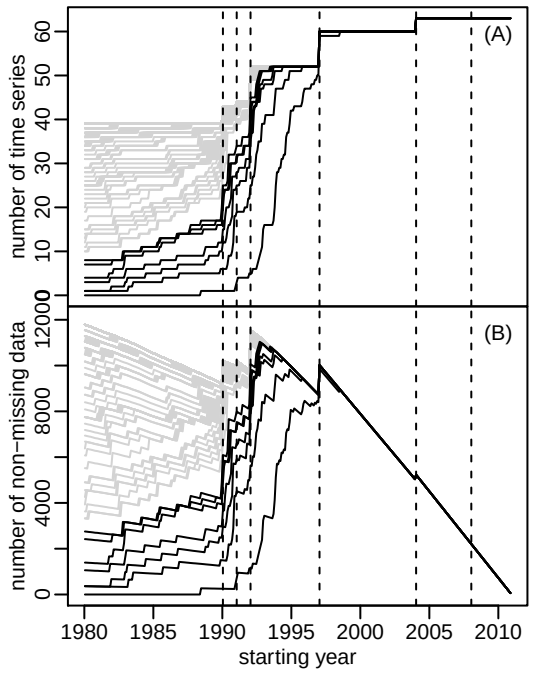


Figure 3.10: Selection of optimal time period for analysis

Effect of the year from which the analysis starts on the number of time series (A) and the number of non-missing data; (B). On each panel, the different lines show the number of consecutive missing values above which we discard the whole time series: from 1 (bottom-most line) to the length of the time series (top-most line). The lines in black highlight the first 6 (i.e. from 1 to 6 consecutive missing values). The vertical dotted lines materialize the province splitting events.

Figure 3.11 shows the time series after selecting the optimal number of provinces and the time period as described above. During the 18-year study period from 1993 to 2010, 26,023,574 cases of ILI were notified, ranging from 320,525 notifications in 1993 to 1,824,195 in 2009.

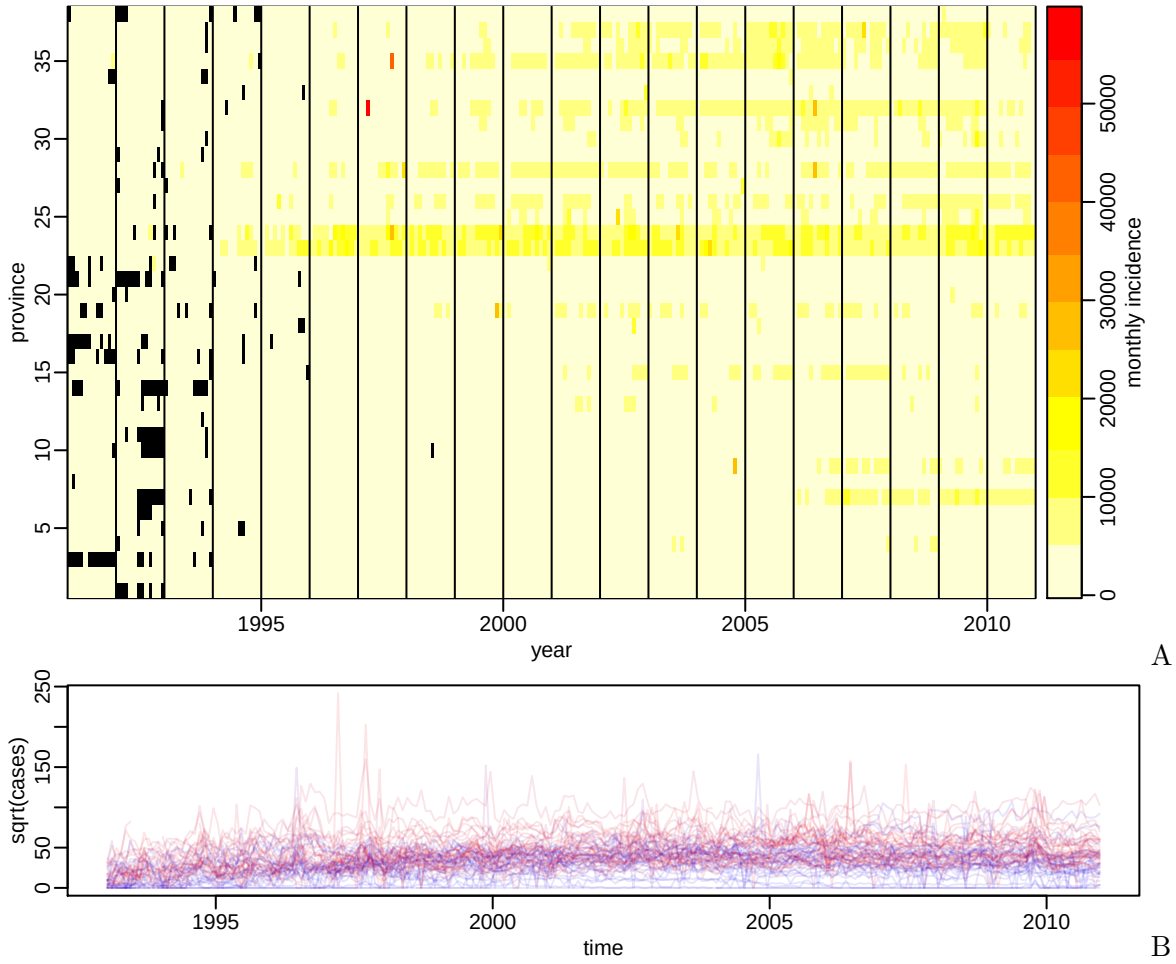


Figure 3.11: Trimmed data
Panel A shows the selected 18 years of data represented as a grid with provinces ordered in rows by the latitude of the population centroid of each province, from north at the top to south at the bottom, and the number of notifications each month colour coded from light yellow (lower number of notifications) to dark red (higher number of notifications) with missing data represented by black. Panel B shows each individual provincial time series colour coded from red (north) to blue (south). This colour code is also apply to other similar figure

3.4.2 Data transformation, de-trending and normalisation

Figures 3.12 show the effect of data transformation, normalisation, and de-trending on the data from one province. The results for each province individually are shown in Annex B.

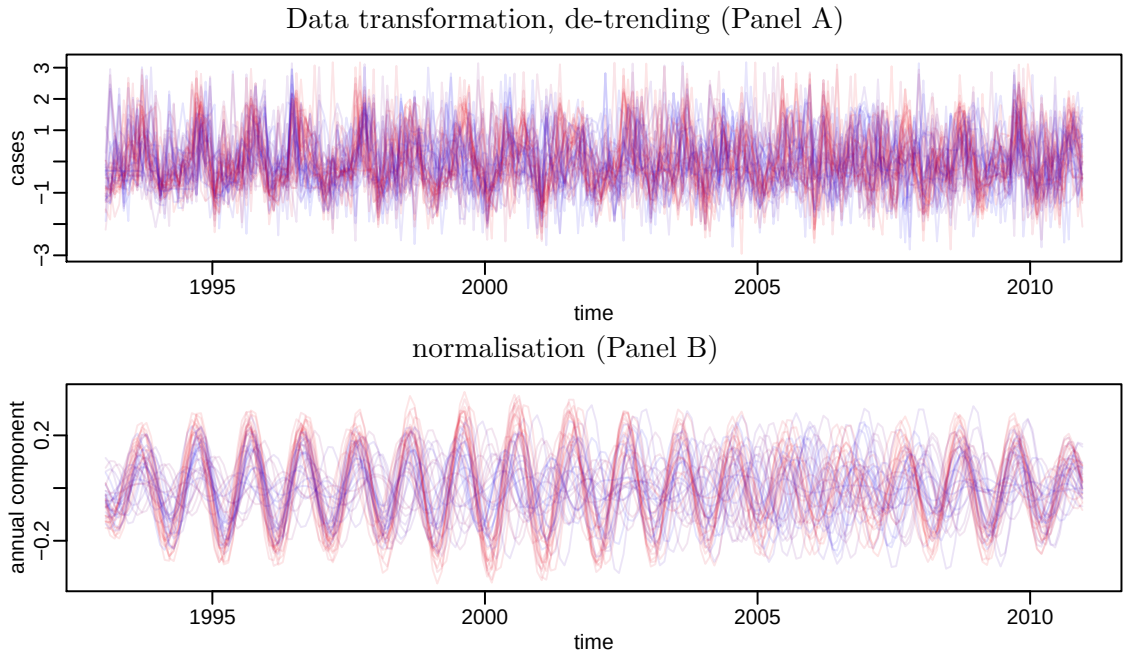


Figure 3.12: Data transformation, detrending and normalisation
Panel A data after detrending and square root transformation. Panel B data after normalisation. Each individual provincial time-series is colour coded

3.4.3 Outliers

One hundred and thirty two outliers (1.17% of all data points) were detected and checked. All outliers were checked against the hard copy data and since all were accurate copies of the available hard copy data, they were discarded and each data point was then interpolated from surrounding data. Figure 3.13 shows the outliers detected for province and the time series after removal and re-interpolation of the outlying data points. See Annex B for full pictures of all outlier of all provinces.

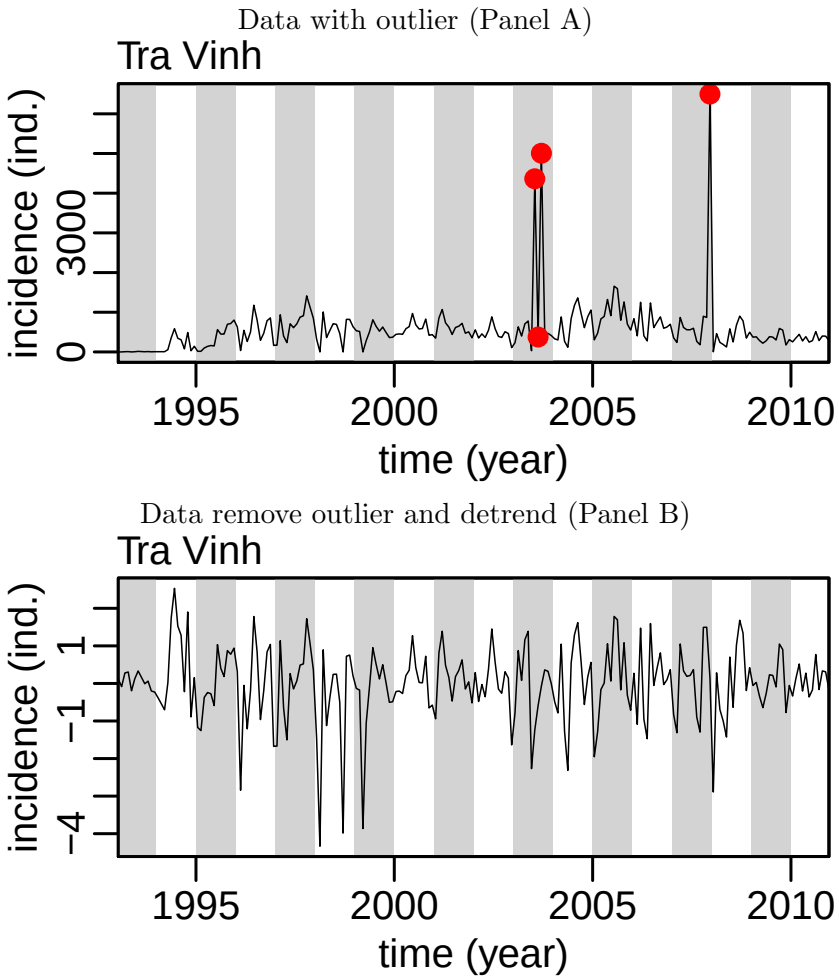


Figure 3.13: Removing outliers, detrending, and normalising

Panel A raw data for Tra Vinh province with outliers marked. Panel B, data from Tra Vinh province after removal of outliers, replacement of outliers by linear interpolation, detrending and normalising *Panel B.*

3.4.4 Longitudinal and elevation effects

Longitude and elevation have no clear effect on ILI (see figure: 3.16. This may be related to the particular, almost linear, shape of the country as mentioned in chapter 1. Viet Nam is long and narrow, with elevations from 0 to 3,000 m running mostly from an east to west direction: latitude and altitude are two gradients that are almost orthogonal (lying at right angles to one another).

3.4.5 Latitudinal pattern of ILI notification rates

Figure 3.14 shows the heat-map of the transformed time series for the 52 provinces ranked from North to South by the latitude of the population centroid of each province. ILI notifications in northern Viet Nam tend to peak in August, September, and October whereas ILI notifications exhibit less variation throughout the year in southern Viet Nam. The heat map therefore suggests that ILI notification is more seasonal in the northern provinces of Viet Nam compared to the southern provinces. This is more formally tested by the wavelet analysis (see also figure 3.15).

3.4.6 Quantifying the seasonality of ILI using wavelet analysis

Wavelet analysis of the dataset shows a clear pattern for most of northern province e.g Hoa Binh in figure 3.15 (A) when a unclear pattern for most of southern province e.g Dong Thap in panel B, and overall pattern in figure 3.16. Both local (see Annex B) and global (figure 3.16) wavelet decompositions for the ILI time series of all the provinces showed a consistent seasonality through time for the northern provinces. The strength of seasonality in each province was quantified by the maximum value of the power between the 0.9 – 1.1 period band. In the rest of the Chapter I will refer to weak or strong seasonality according to this power value. Figure 3.16 (B) shows that the non-northern provinces with high strength of seasonality appear to be located in mountainous areas (figure 3.16 C). This

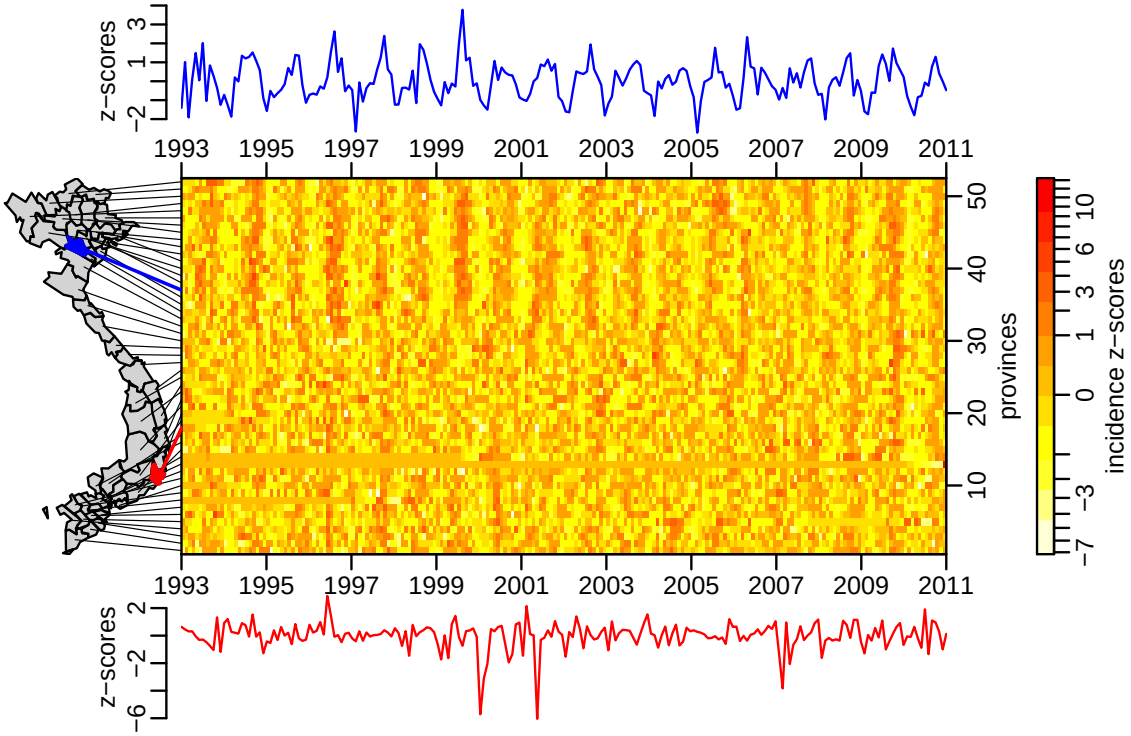


Figure 3.14: Heat map of ILI notifications.

Each row of the matrix corresponds to a province (52 in total) and each column corresponds to a month from January 1993 to December 2010 (216 in total). The colour of each cell shows the value of these notification z-scores, on a square-root scale for better visibility (see the scale on the right). The rows of the matrix are arranged according to the latitude of the population centroid of the province, as can be seen from the map on the left. On this map, each line connecting to the matrix starts from the population centroid of the province. On the top of the matrix is shown the time series of the detrended, centered and reduced time series of ILI raw incidence for the province of Hoa Binh as an example. This province is colourized in blue on the map. The red one at the bottom is for Ninh Thuan province in the south.

suggests a role of climatic factors in explaining the observed latitudinal gradient in the strength of ILI seasonality.

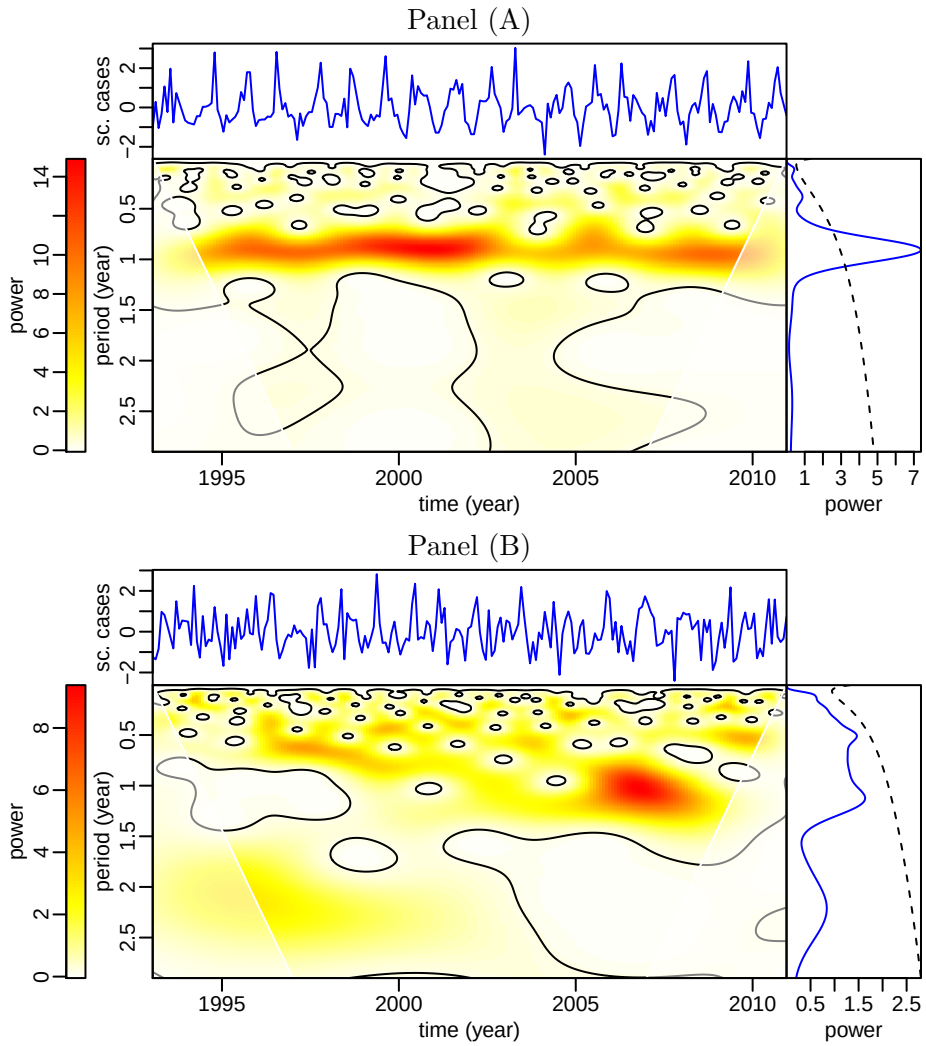


Figure 3.15: Example of wavelet transform

Panel A: Hoa Binh province, Panel B: Dong Thap province.

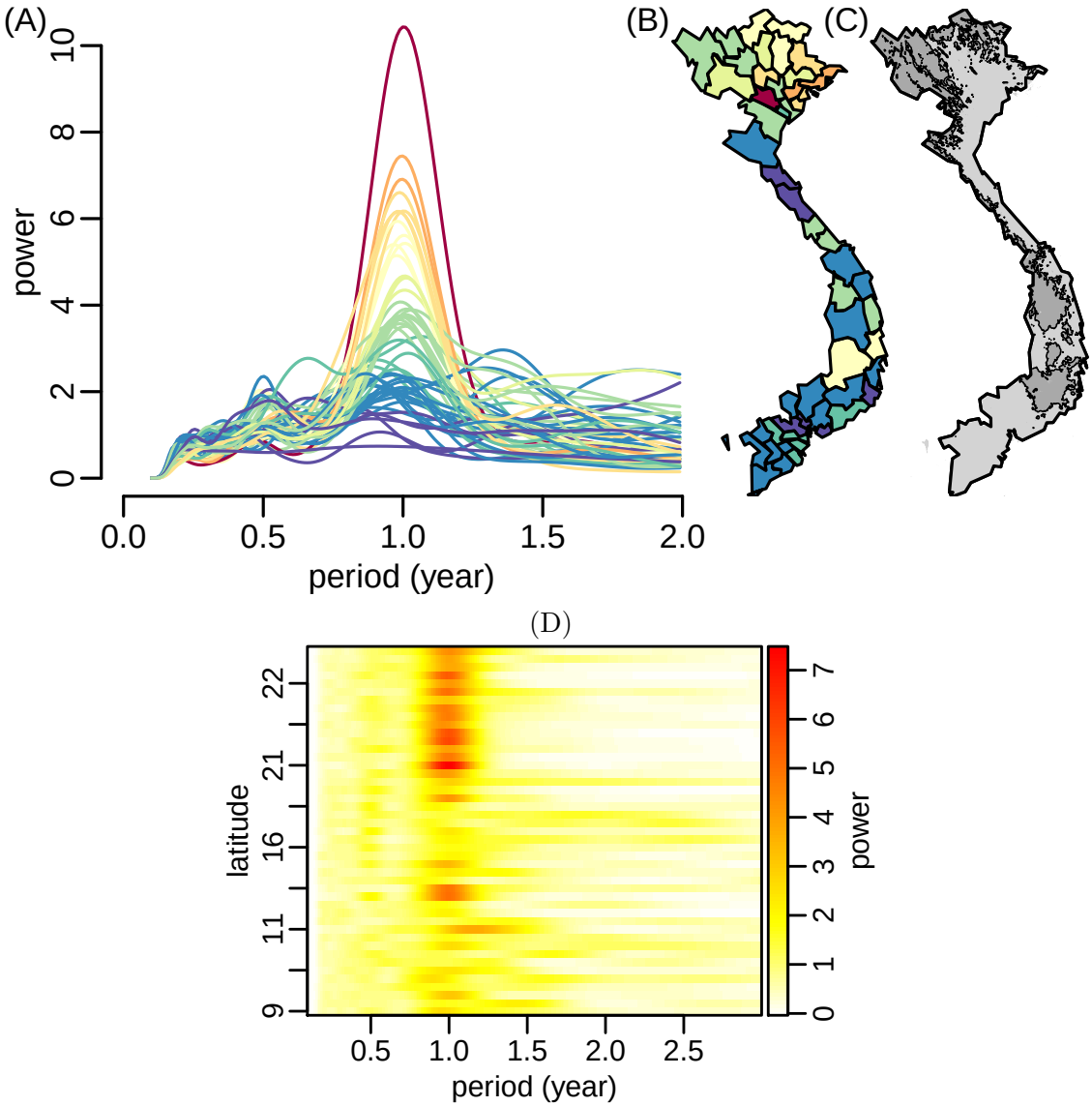


Figure 3.16: Strength of ILI seasonality.

Global wavelet power spectrum for the detrended, centered and reduced time series of ILI raw incidence (A, one curve per province, 52 in total). The colours of the curves vary as the value of the maximum power for a periods between 0.9 and 1.1 year. The same colour code is used to plot the strength of seasonality on the first map (B). The second map (C) shows elevations higher than 500 meters of altitude in dark grey. Panel (D). wavelet transform around the band of 1 year period by latitude of provinces

3.5 Discussion:

During the reshaping and cleaning process, the data were converted from a very noisy dataset to one in which patterns are discernible. This process has now been automated using an R script so that with new data or changes, we will be able to rerun the process effortlessly. During the data selection process, we had to select the optimal cut-off year for data analysis. Data before 1993 were omitted due to a large amount of missing data, and if a time series had more than 3 consecutive missing data points, the earlier data were omitted. However, before trimming the time series I spent much effort searching other sources of data from the provinces themselves and from the Ministry of Health in an attempt to fill in the gaps.

The results of latitude gradient's effect on influenza-like illness force one to think about its implications for influenza control, with particular reference to influenza immunisation. In general, Viet Nam does not have a domestic supply of influenza vaccines for use during seasonal spread or pandemic outbreaks, and a national candidate influenza vaccine supply is still under development (Hoa et al., 2011). Imported influenza vaccine is not widely used in Viet Nam since its cost is high compared to average incomes and the Vietnamese people do not have much awareness of the availability and indications for its use. Globally, influenza vaccine recommendations are provided two times per year, once from March to April for the northern hemisphere season and once from August to September for the southern hemisphere, and these recommendations are used by Viet Nam. The Viet Nam Ministry of Health has produced recommendations on the use of influenza vaccines (Gupta et al., 2012) but the recommendations are not based on a thorough analysis of the local epidemiology of influenza in Viet Nam. Based on our results, it is feasible to provide vaccine before the influenza season in the north of Viet Nam but decisions on the timing of influenza immunisation in southern Viet Nam will be very difficult. One limitation of this work is that the sensitivity and specificity of the ILI notification data as a marker of influenza activity has not been established, and this is the subject of the analysis described in chapter 5.

Our result on the seasonality of ILI incidence throughout the country is in agreement

with what was found in Brazil by Alonso (Alonso et al., 2007), although Alonso et al. studied pneumonia mortality patterns instead of influenza morbidity. Another difference is that Viet Nam is located above the equator whereas Brazil is crossed by the equator. The high power of seasonality found in the northern region of Viet Nam is similar to pattern of timing of influenza epidemics found by Alonso in the southern region of Brazil, which has a similar distance from equator. Alonso et al. (2007) looked at the timing and amplitude of influenza epidemics and, as described in the methods section of this Chapter and in Chapter 2, the amplitude can be affected by increases in awareness in the population or a change in surveillance practices, but the seasonality power component in these cases stays constant. Another advantage of our study is that the number of the spatial units used as well as the duration of our study was greater compares to Alonso et al.

3.6 Conclusion

ILI in the northern region of Viet Nam has a clear seasonal pattern but it contrasts to the southern region where seasonality is not detectable. This pattern correlates with latitude and leads us to hypothesise that the seasonality of ILI in Viet Nam is driven by climate factors, which also change dramatically by latitude in this long and narrow country.

Chapter 4

Climate association of influenza-like illness in Viet Nam

4.1 Introduction

In chapter 3, it was demonstrated that in Viet Nam there is a latitudinal gradient in the seasonality of ILI notifications. Based on prior knowledge (see Chapter 2) it is likely that seasonal changes in climatic variables are drivers of the observed periodicity of ILI in Viet Nam. This hypothesis is supported by our finding that the seasonality of ILI is more marked in areas of higher altitude in the southern areas of Viet Nam compared to low altitude areas in southern Viet Nam. The observed spatial heterogeneity of the strength of seasonality of ILI in Viet Nam offers an opportunity to assess which of the proposed climate drivers of influenza are most strongly associated with this pattern. Potential explanatory factors include absolute humidity, relative humidity, temperature and sunshine (see chapter 2). Viet Nam is an appropriate country in which to test climate-ILI associations since it is a long country spanning many latitudes, with a substantial difference in climate between the north and south, over a small area (330,000 km^2). North Viet Nam experiences clear seasons, with cool winters and hot summers, and obvious intermediate periods of spring and autumn. In contrast, south of Viet Nam is hot all year round, with the seasons defined by the amount of rainfall. In this chapter I present my work to formally test the strength of association between a range of climatic variables and the seasonality of ILI notifications in Viet Nam.

4.2 Materials and Methods

4.2.1 Data and transformations

Population centroids

These were calculated as described in Chapter 3.

Climate data

Monthly average climate variables from January 1993 to December 2010 were obtained from the Viet Nam Institute of Meteorology, Hydrology & Environment (http://www.imh.ac.vn/). These data were originally obtained from the HydroMeteological data centre (http://www.hymetdata.gov.vn/) in a hard copy report format. Data on 8 variables are collected at least once daily (normally the data are recorded 4 time a day or continuously with data logger) on the ground at 250 meteorological stations. About 170 of the meteorological stations are surface stations, of which 122 are permanent. The other stations are specialised stations established for specific objectives e.g. for air traffic control or for agricultural purposes. Data on the exact coordinates of 68 Province stations with a long time series and data on seven variables were obtained from the Viet Nam Institute of Meteorology, Hydrology & Environment. The seven variables were monthly averages of: daily minimum, average, and maximum temperatures (in degrees Celsius), absolute (g/L) and relative (%) humidities. Total time of sunshine (h) and rainfall (mm) are aggregated for the whole month. Where more than one measurement was taken per day, the daily average was calculated. Monthly averages were calculated by averaging daily values i.e average monthly maximum temperature is the average of the maximum daily temperature, and average monthly temperature is the average of the average daily temperature. There was one other variable that we considered but did not obtain which was wind speed, and we believe that this variable does not relate much to ILI time series.

Each province was assigned to one unique climatic station. This station was selected as the one closest to the population centroid of the province (Figure 4.1). These climate datasets were then used for the assessment of the correlation between climatic data and epidemiological data (see below) (see*R code in Annex D for more technical information).*

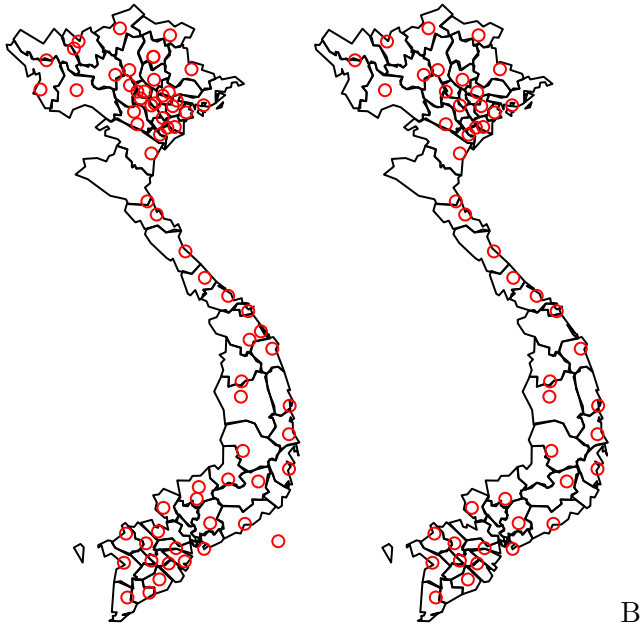


Figure 4.1: Location of meteorological stations
Position of 68 meteorological stations obtained for this study (panel A) and 47 selected for coherence analysis (panel B)

Wavelet decomposition.

We used the same wavelet decomposition methods as described in chapter 3 to explore the periodicity of all the climate time series. See Cazelles et al. (2008, 2007) for more technical details.

Principal component analysis (PCA)

Principal component analysis (PCA) is a statistical technique that is commonly used to identify patterns in data with high dimensions (Ringnér, 2008; Smith, 2002). PCA identifies new variables (Principal Components) that best explain the variability in the outcome measure. The PCA initially selects the single PC that explains the largest proportion of the variance, and then selects the second PC, that is uncorrelated with the first PC, which explains the next largest proportion of variability, and so on. The PCs can then be expressed in terms of the original variables. In our data matrix, the outcome measure is the strength (not the amplitude) of ILI seasonality and the explanatory variables are the strength of seasonality of the seven climate factors; with the strength of seasonality quantified by the maximum value of the global power spectrum within the annual band. We

used the function "prcomp" in R (version 3.0) to perform principal components analysis on the whole set of data matrix. See (Holland, 2008) for more technical details.

Tree regression.

In order to explore which climatic factors best explain the strength of ILI seasonality as defined in chapter 3, we used tree regression. Tree regression is a method to make a prediction model that can be represented as a decision tree (Loh, 2008). In tree regression the data are partitioned at decision nodes such that the variability in the outcome variable is minimised within each partition and maximised between partitions. Each of the terminal nodes or leaves of the tree represents a cell of the partition, and has attached to it a simple model which applies to all the data in that cell only. A point x belongs to a leaf if x falls in the corresponding cell of the partition. To figure out which cell we are in, we start at the root node of the tree, and ask a sequence of questions about the features. The interior nodes are labelled with questions, and the edges or branches between them labelled by the answers. Which question we ask next depends on the answers to previous questions. Since each question refers to only a single attribute, and has a yes or no answer, e.g. below or higher than 15 degree Celsius, out final result is robust to the specific ordering of questions.

In order to account for the maximum number of climatic summary statistics we computed the following statistics for each climatic variable: the mean, the minimum, and maximum values, the range, the number of months below a given threshold (with the threshold varying from the minimum to maximum value in increments of one), and the strength of seasonality calculated from global wavelet power spectrum as explained in chapter 3. This generated 534 climatic summary statistics that were used as explanatory variables in a binary tree regression model with the strength of ILI seasonality as a dependent variable. We used the "tree" R package to perform the tree regressions (Ripley, 2005).

Global extrapolation.

In order to extrapolate climate associations identified within Viet Nam to a global scale, we obtained global climate data from the National Centers for Environmental Prediction (NCEP) / National Centre for Atmospheric Research (NCAR) project (Kalnay et al., 1996). The NCEP/NCAR project produces coarse spatial resolution (2.5 x 2.5 degree) climatic grids globally with a daily temporal resolution. The data (Figure: 4.2 are derived from several different sources and are integrated into a final database.

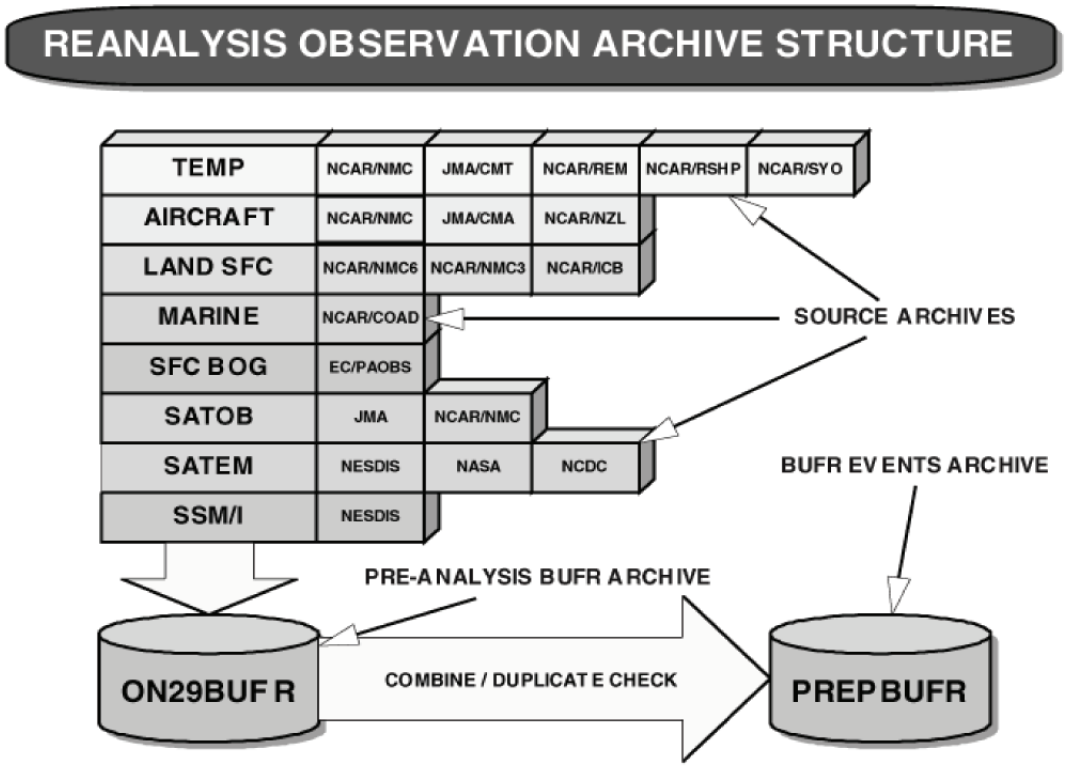


Figure 4.2: Sources of data for the NCEP/NCAR project.
Source: The NCEP/NCAR 40-Year Reanalysis Project (Kalnay et al., 1996)

Monthly mean data were obtained through Daniel Weiss over the 1993-2010 period for 3509 terrestrial pixels. Since data on absolute humidity were not available from NCEP/NCAR, absolute humidity was calculated from the relative humidity and the temperature, as follows:

1. Derive the saturation vapour pressure (eS) (in mb) at temperature tC (in Celsius)

using equation (10) from Bolton (Bolton, 1980).

$$eS = 6.112 \times e^{(17.67 \times tC)/(tC+243.5)} \tag{4.1}$$

2. The actual vapour pressure (eA) (in mb) was calculated by multiplying the rH (in %) by eS from step 1:

$$eA = eS \times \frac{rH}{100} \tag{4.2}$$

3. Absolute humidity (AH) in g/m^3 was then calculated using the following equation:

$$AH =\sim 2.16674 \times \frac{eA}{tK} \tag{4.3}$$

Where the temperature in Kelvin (tK) was derived from Celsius: tK = tC + 273.15.
A constant was derived using the ideal gas law: constant = 18.01528 / 8.31446 = 2.16674 gK/J (where 18.01528 g is the molecular mass of water and 8.31446 J/mol K is the universal gas constant)

On these global climatic variables we computed the same summary statistics as the ones used in the regression tree. We then applied the decision rule of the regression tree selected in Viet Nam to the global scale.

4.3 Results

4.3.1 Geographic patterns of climate variables

Temperature

Patterns of seasonality in temperature are clear in both north and south Viet Nam but the amplitude of the annual fluctuation in mean, maximum and minimum temperature is clearly much greater in north Viet Nam compared to the south. In the north, temperatures are much lower in the winter months than in the north, although summer temperatures are slightly higher in the north, see (Figure: 4.3). The spatial variation in temperature is also shown in figure 4.4 panel A

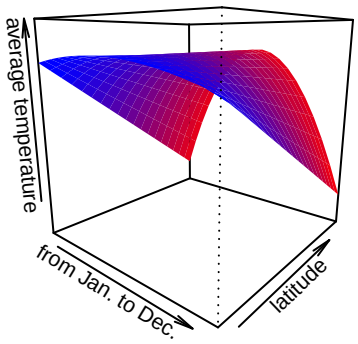


Figure 4.3: Average temperature by time and latitude. Time in months

Sunshine

Sunshine is strongly seasonal throughout the country but the north and south are out of phase, with sunshine peaking in the summer in the north and in the winter in the south. The mean sunshine is also higher in the south, with the minimum value in the south being equal to the maximum value in the north. See panel B of figure 4.4.

Rainfall

Rainfall is seasonal throughout Viet Nam, with the rainfall being greatest in the summer months in both north and south but is greatest in the autumn months in central Viet Nam. In central Viet Nam, the amplitude and the maximum value of the rainfall time

series in much higher than in the north or the south. See panel C of figure 4.4.

Humidity

The signal for relative humidity is quite noisy, with patterns difficult to differentiate, but a suggestion of a biannual cycle in the north, a summer peak in the south, and a summer nadir in central Viet Nam (see figure 4.5). Relative humidity is strongly seasonal in the south, with a peak in summer, but not at all in the north. In the centre, relative humidity looks quite seasonal too with a peak in winter. There is however clear seasonality of absolute humidity throughout Viet Nam, with AH peaking in the summer and lowest in the winter. As is seen with temperature, the amplitude of the annual fluctuation of AH is clearly much greater in north Viet Nam compared to south, with AH levels much lower in the north than the South during the winter, but similarly high levels during the summer. See panel D of figure 4.4.

The climate in Viet Nam is very diverse, as revealed by the variability in seasonality of the different climatic variables, the gradient of seasonality from south to north for temperatures and absolute humidity, the out-of-phase sunshine between the north and the south, and even a more complex pattern for the rainfall (which peaks in summer in the south, in autumn in the centre and back to summer in the north). Such diversity on such a small geographic area is unusual and makes Viet Nam a perfect country to explore the link between infectious diseases epidemiology and climatic drivers. This is reinforced by the fact that other factors that might potentially affect the epidemiology, such as population demographics, are quite homogeneous throughout the country.

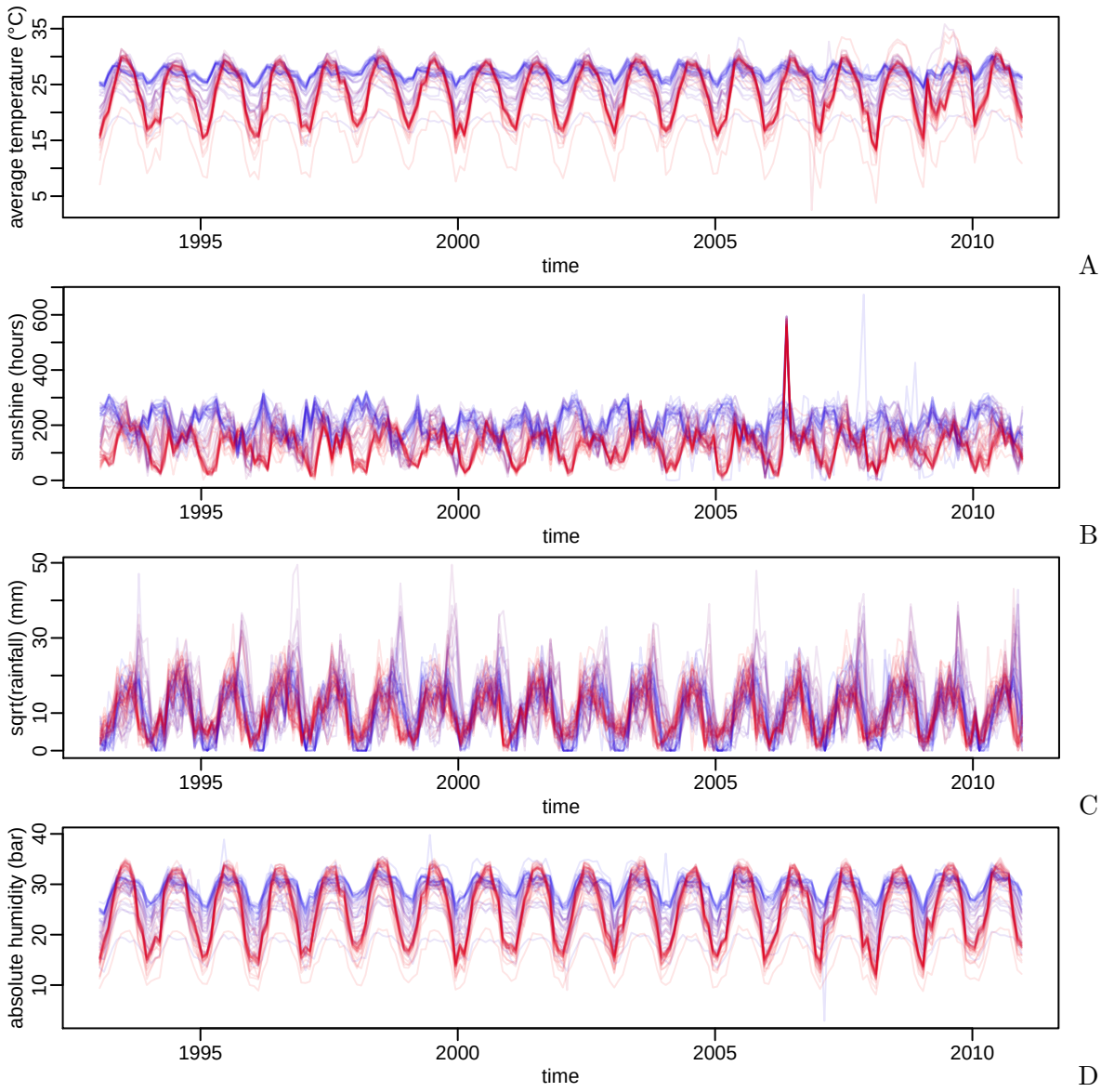


Figure 4.4: Temporal and spatial patterns of climatic factors.
The colours of the lines represents the latitude of the meteorological station, as coded on the map in figure 4.5.

4.3.2 Association between ILI seasonality and climatic variables

Principal components analysis

The principal components analysis (PCA) of the climatic variables reveals that temperature and absolute humidity are positively correlated and together explain around 59% of the total climatic variance (first PC axis, figure 4.6). Furthermore, and not surprisingly, rainfall and relative humidity are positively correlated with each other and both negatively

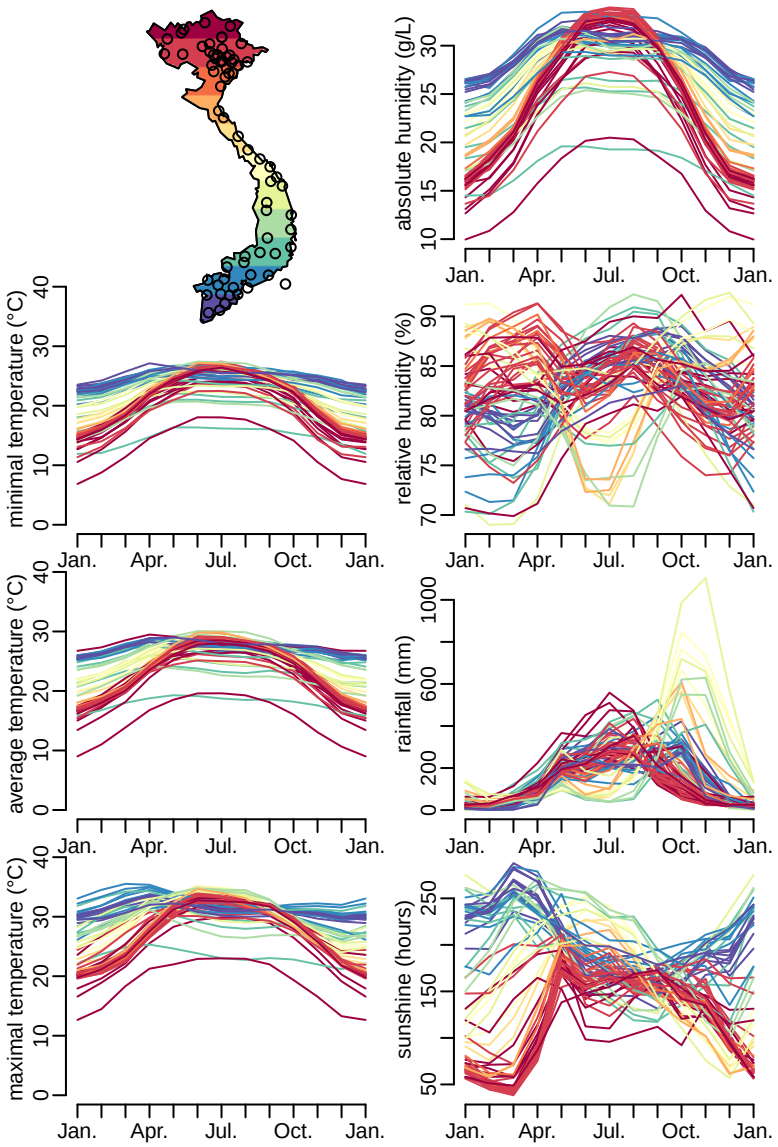


Figure 4.5: Variation of climatic variables in Viet Nam over one year
from month to month: minimum, average and maximum temperatures (°C), absolute and relative humidities (g/L and % respectively), amount of rainfall (mm) and number of hours of sunshine. Each line corresponds to one climatic station (68 in total) and is an average over the 1993-2010 time period. The colours of the lines represent the latitude of the meteorological station, as coded on the map. This latter shows the locations of the climatic stations (open circles)

correlated to sunshine. These three variables together explain 24% of the total climatic variance (second PCA axis). We can see further in panel B that the first PCA component has a variance that increases from south to north. No such trend is observed on the second PCA component (not shown).

Indeed, PCA analysis is a way to characterise the climate from multiple variables to a simple picture. What it shows is that temperature and absolute humidity are closely correlated, and almost independently (orthogonal) from rainfall, relative humidity and sunshine (as expected, rainfall and relative humidity are very correlated and both are negatively correlated with sunshine). This PCA analysis shows that we can summarise the climate very well with temperature (or absolute humidity), (explaining 59% of the total variance), and with a variable reflecting relative humidity, rainfall and sunshine (24% of the total variance). Furthermore, this analysis shows that there is a strong latitudinal gradient in the amplitude of the first PC (more than twice as big in the north as in the south). No clear longitudinal gradient was seen. Further, no clear gradient at all was shown on the second PC. Since we also see a similar latitudinal gradient of seasonality for the ILI epidemiology, it does suggest that there may be a link between variables of the first PC and ILI seasonality. And this is what will be explored by the tree regression analysis.

Regression tree analysis

The regression tree analysis shows that the strength of the seasonality of absolute humidity is the climatic summary statistic that best explains the strength of ILI seasonality (figure 3.16). This result is robust to the number of explanatory variables included into the regression tree analysis, as well as to their order of introduction. The regression tree analysis shows that provinces with weak seasonality in absolute humidity (annual power below a value of 17.60) have weak ILI seasonality, and provinces with strong seasonality of absolute humidity (annual power above 17.60) have strong ILI seasonality. Figure 4.7 shows that the relationship between the two seasonalities (ILI and AH) is non-linear in shape, reinforcing the threshold identified by the tree regression.

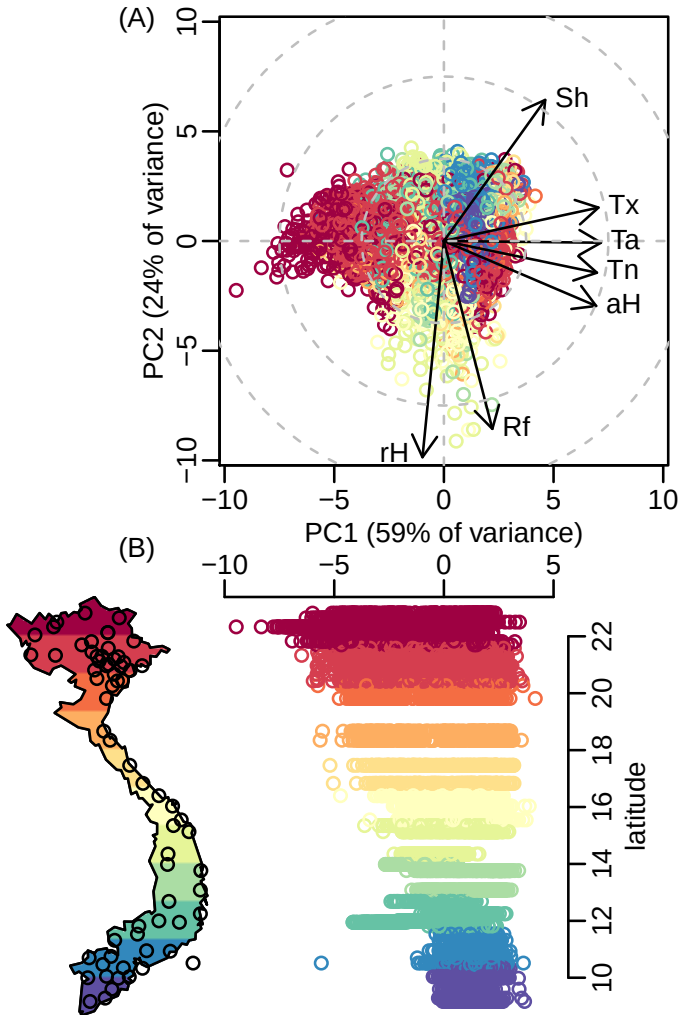


Figure 4.6: Latitudinal gradient of climate seasonality.
(A): principal component analysis (PCA) of the seven climatic variables of 4.5. Sh=sunshine; Tx=Maximum temperature; Ta= Average temperature; Tn=Minimum temperature; Rf=Rainfall; AH=Absolute humidity; rH=Relative humidity. The first two components explain more than 83% of the total variance. Each dot corresponds to a given month, for a given climatic station. The colour of the dots varies according to the latitude of the climatic stations, as shown on the map showing the locations of the climatic stations. (B): relationship between the first component of the PCA (A) and the latitude of the climatic stations. The latitude axis of the plot also corresponds to the map on the left.

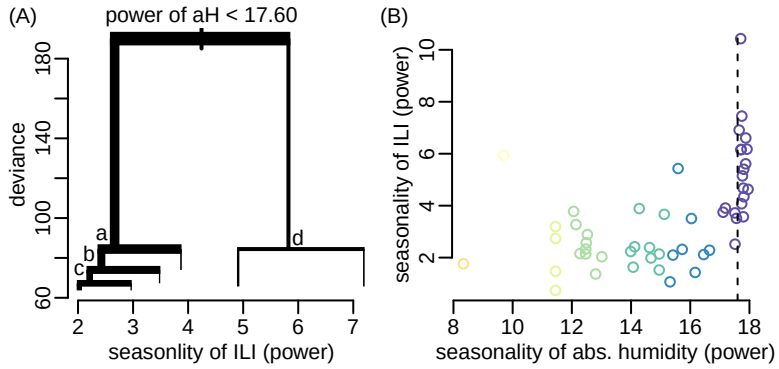


Figure 4.7: Regression tree analysis of relationship between ILI seasonality and climate factors.
(A): regression tree of the maximum power between periods 0.9 and 1.1 year (seasonality of ILI), using, for the explanatory variables, the characteristics of the climatic variables, as described in the materials and methods section. The widths of the segments are proportional to the number of provinces. The first node discriminates the provinces for which the annual power of the absolute humidity is below (left, 36 provinces) and above (right, 16 provinces) a threshold value of 17.60. Node (a) discriminates the provinces for which average relative humidity is below 77% (left) and above (right) a threshold value of 51. Node (b) discriminates the provinces for which the range of the absolute humidity is below (left) and above (right) a threshold value of 14.46 g/L. Node (c) discriminates the provinces for which the annual power of the relative humidity is below (left) and above (right) a threshold value of 11.63. Node (d) discriminates the provinces for which the number of months per year with a maximal temperature above 26°C is below (left) and above (right) a threshold value of 13°C. (B): relationship between the seasonality of ILI and the seasonality of absolute humidity. Each dot is a province (52 in total) and the vertical dash line shows the threshold of the first node on the regression tree (A).

Absolute humidity and ILI: dynamics and phase

The tree regression analysis reveals a strong association between the strength of seasonality of absolute humidity and the strength of seasonality of the ILI time series. Here we look at whether there is a consistent phase difference between absolute humidity and ILI, as has been shown in temperate countries of the world where both ILI and absolute humidity are strongly seasonal. Since a phase and thus a phase difference can be efficiently calculated only when the signal is substantially seasonal, we show the results separately for the provinces with strong and weak seasonality (both in absolute humidity and ILI). Figure 4.8 shows the seasonal dynamics of absolute humidity and ILI for the northern (left) and southern (right) provinces, showing a much stronger seasonality of both absolute humidity and ILI in the north than in the south, in accordance with the result of the

tree regression. Further, what this figure shows is that absolute humidity and ILI seems to be almost in phase. In order to investigate more the link (and possibly the causality) between these two dynamics, we look at the phase difference between the two time series. Figure 4.8 shows that the peak in absolute humidity leads the peak in ILI notification by about one month. Further, as expected, the consistency of this phase difference is much more marked in the provinces where the seasonality of both absolute humidity and ILI are strong.

We use latitude at 19°N to separate north from south. In the north, humidity seems to peak earlier than ILI, but in the south the changes in absolute humidity and ILI are smaller and the peaks are difficult to define (Figure 4.8). Figure 4.9 investigates the phase difference between seasonal components of absolute humidity and ILI. As expected from the fact that both ILI and absolute humidity seasonality are much more pronounced in the north than in the south, the phase differences between the two variables are much more consistent both in time and space in the north (left panel of figure 4.9) than in the south (figure 4.5, right panel of figure 4.9). Despite some temporal and spatial variability on the phase difference in the north, we observe an average lag of one month between absolute humidity and ILI incidence.

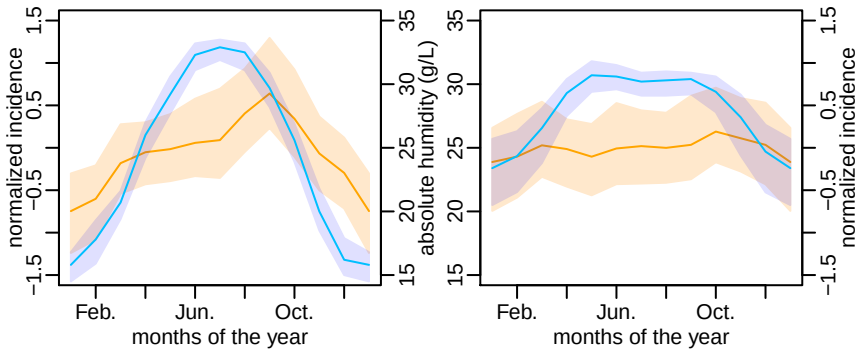


Figure 4.8: Seasonalities of ILI and absolute humidity.
Normalized notification rate (orange) and absolute humidity (blue) are shown for the north (left) and the south (right) provinces of Viet Nam. The limit between north and south is here arbitrarily defined at the latitude of 19°N, but the results are robust to the definition of this limit. The lines represent the median over the 18 years of the studied period and the shaded areas represent the inter-quartile ranges. The numbers of provinces are 21 and 31 for the north and the south, respectively, and the numbers of meteorological stations are 32 and 35 for the north and the south, respectively.

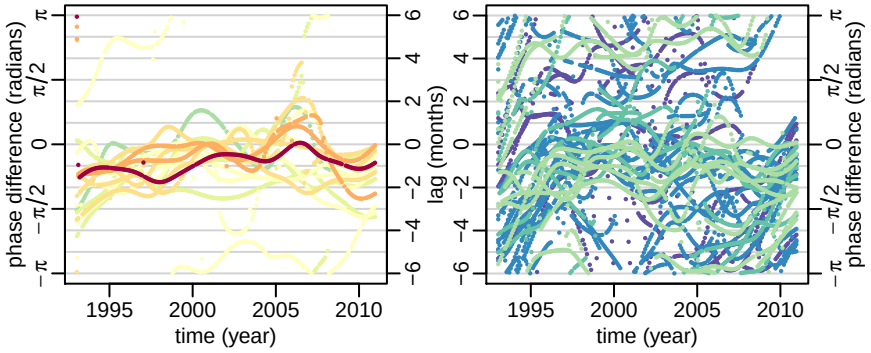


Figure 4.9: Lag between ILI and absolute humidity.
Differences of phase angles between the time series of ILI and absolute humidity, filtered around the 1-year period (a negative difference thus meaning that absolute humidity is ahead of ILI). For better visibility, the left and right panels show the 16 and 36 provinces respectively for which the annual power of ILI are respectively above and below the arbitrary threshold of 4 (see figure 4.5. The colour coding is the same as in figure 4.5)

4.3.3 Extrapolation to the global scale

Results from our analysis on Vietnamese data show that the strength of absolute humidity seasonality is the variable that best explains the strength of ILI seasonality: with strengths of absolute humidity seasonality below and above a threshold power of 17.60 predicting weak and strong ILI seasonality respectively. We retrieved from the NCEP/NCAR project monthly time series of absolute humidity (1993 2010) from all the 3509 2.5 x 2.5 degree terrestrial pixels around the world. For each of these time series we computed the strength of absolute humidity seasonality globally as explained in chapter 3 (figure 4.10 A) from which we predicted the strength of ILI seasonality from the decision rule concluded from the tree regression (figure 4.10 B, strong in green, weak in red).

4.4 Discussion

Our study included time-series from 52 provinces over an 18 year period with a high spatial (province level, *ca* 5000 km^2) and temporal (monthly aggregated data) resolution which to our knowledge makes this the highest resolution of data yet published in our region. Furthermore, Viet Nam has good conditions in which to test hypotheses about potential links between infectious diseases epidemiology and climatic drivers because its high diversity of climate (with two batches of climatic variables orthogonal with each

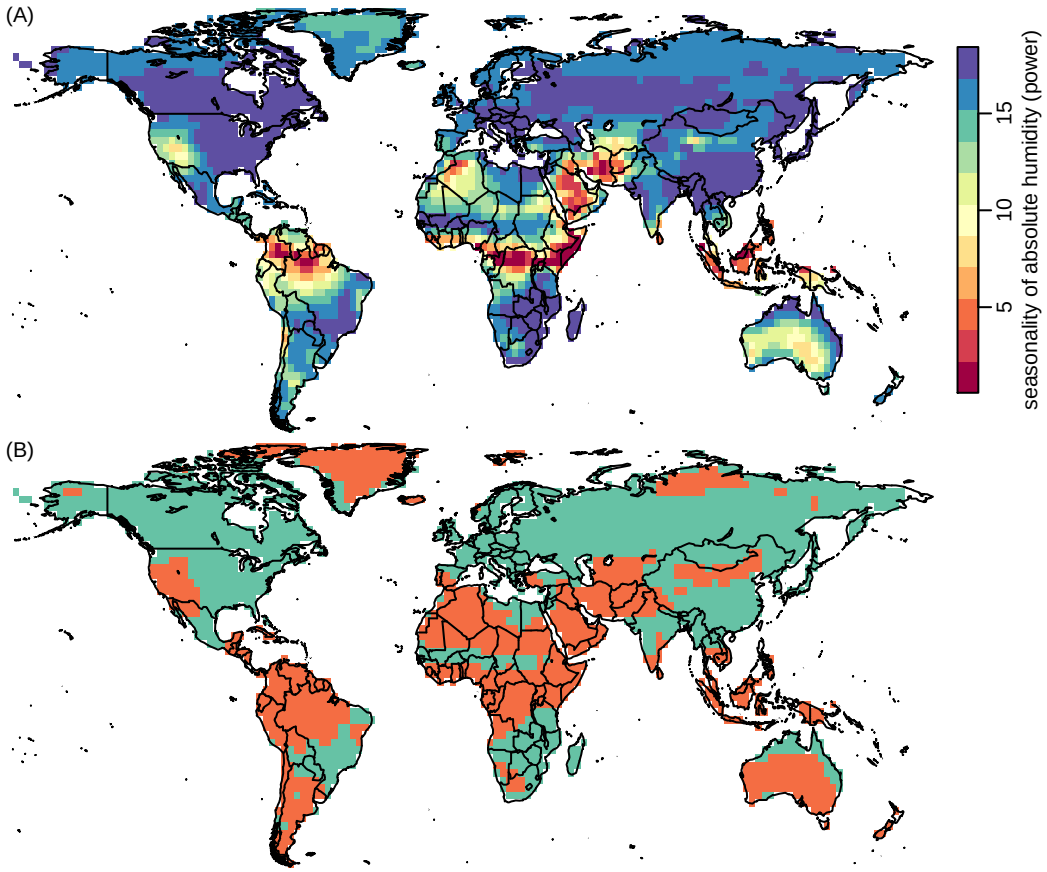


Figure 4.10: Global seasonality of absolute humidity.
(A): seasonality of monthly time series of absolute humidity calculated with a 2.5^o x 2.5^o spatial resolution (see Materials and Methods). Colour code is the same as on figure 4.7B. (B): discrimination between power above (green) and below (orange) the threshold value of 17.60 (see figure 4.7). From figure 4.7, green colour would predict strong seasonal ILI epidemiological dynamics and orange colour would predict non-seasonal ILI epidemiological dynamics.

other) on a small, long but narrow area (see chapter 1). Our study provides an advance on the approach of Shaman et al 2010 who analysed the onset of epidemics in a temperate country (USA) where the influenza dynamics are very seasonal (Shaman et al., 2010). Two studies to date, from Brazil and China, have looked at the strength of seasonality as we did (Alonso et al., 2007; Yu et al., 2013). Both of them found latitudinal gradients of seasonality, but only Yu et al. (2013) tried to explain these gradients with climatic variables. However, the authors did not include absolute humidity in their analysis.

The climatic data of Viet Nam show, on an area of moderate size (330,000 km^2), a great diversity of seasonality for the different climatic variables. Temperatures and absolute humidity peak in the summer time with a gradient of amplitude from the south to the north. Relative humidity has more seasonality in the south than in the north with the peaks in the summer for the south. The gradient of its amplitude is in reverse that increases from the north to the south. The peak in hours of sunshine shifts from the winter time in the south to the summer time in the north. The amplitude of these seasonality are similar in both the north and the south but the average is higher in the south than in the north (peak value in the north is about the value as trough value in the south). Rainfall seasonality is more complicated, with a peak shift from the summer time in the south to the autumn in the centre and back to the summer time in the north. This is an exceptional diversity of both amplitude and timing of the seasonalities of the different climatic variables, which makes Viet Nam an ideal candidate country to test the relationships between climatic variables and infectious diseases transmission. This is confirmed by figure 4.7 which shows that the variability of absolute humidity seasonality observed in Viet Nam is similar to the variability of absolute humidity seasonality seen at a global scale. Such a diversity of climates in a small and highly populated area is an asset compared to world-scale comparative studies, for which factors other than climate (demography, behaviour, etc.) may vary substantially and act as confounding factors.

Our study is an independent epidemiological confirmation of the role of absolute humidity in driving the epidemiology of influenza first identified by Shaman et al. (2010) and Shoji et al. (2011). Actually, our results on the phase of the ILI and absolute humidity (one month delay) do not contradict results of (Shaman et al., 2010) and (Tamerius et al.,

2013) since we are not looking at the same thing, the onset of epidemics. Shaman et al are studying time series in a temperate part of the world where influenza epidemiology is very seasonal and to answer the question "what climatic factors trigger an epidemics every year?". The conclusion of Shaman et al. (2010) is that a drop in absolute humidity triggers epidemics in the temperature region. Tamerius et al. (2013) identifies two types of environmental conditions associated with epidemics: "cold-dry" and "humid-rainy": "For sites where monthly average specific humidity or temperature decreases below thresholds of approximately 11-12 g/kg and 18-21°C during the year, influenza activity peaks during the cold-dry season (i.e., winter) when specific humidity and temperature are at minimal levels. For sites where specific humidity and temperature do not decrease below these thresholds, seasonal influenza activity is more likely to peak in months when average precipitation totals are maximal and greater than 150 mm per month" (Shaman et al., 2010). We are looking at time series in a tropical part of the world where the seasonality of influenza epidemiology appears to be highly variable. The question we ask is "what climatic factors explain the observed gradient of seasonality in influenza epidemiology?" and the result leads to absolute humidity seasonality.

Investigation of the relationships between climatic variables and influenza transmission is usually done by testing each potential climatic variable in turn and comparing the fits. An example is the reanalysis by Shaman and Kohn (2009) of the data of Lowen et al. (2007), thereby proposing absolute humidity as a better predictor of influenza virus transmission than relative humidity. Such investigations involve different tools such as linear regression, generalised linear models, or even mathematical modelling. However, whatever the framework, interactions and colinearities between potential explanatory variables are rarely accounted for. This poses a problem given that the different climatic variables are highly collinear (see for example figure PCA). This issue questions the validity of comparing different studies with different sets of potential explanatory variables and different ways of dealing their interactions and colinearities. As an example, our analysis in Viet Nam is very similar in spirit to the one recently published by Yu et al. (2013) in China. Using generalised linear modelling they show that their observed latitudinal gradient of influenza seasonality is best explained by temperature and hours of sunshine. However,

their analysis did not test the effect of absolute humidity, which makes their result difficult to compare to our study, and also to previous work showing the effect of absolute humidity (e.g. Shaman et al. (2010) in the United States). This calls for the grouping of data from different sources as initiated by Tamerius et al. (2013). However the global study of Tamerius et al. (2013) classified the dynamics of influenza epidemics as annual or biannual, and did not look at the strength of seasonality as we and others have done (Alonso et al., 2007; Yu et al., 2013). Tamerius et al. (2013). Whilst we could compare our extrapolation, approximating biannual epidemics to less seasonal ones, with the analysis of Tamerius et al. (2013), the comparison may be a little stretched, and it would be more robust to calculate the strength of influenza seasonality in Tamerius's database and compare this to our global extrapolations.

The presence of an association between absolute humidity and ILI in the sub-tropics is a novel finding. The tree model that we used to investigate the relationships between ILI seasonality and the climatic variable is a simple binary recursive partitioning method and thus allows the detection of links between a response variable and a number of potential explanatory variables when these latter can be numerous and when the shapes of the relationships can be non-linear and very complicated. These properties allowed us to test as many summary statistics of climatic time series as possible and to detect the most explanatory ones. It also allowed to detect effects that are more threshold effects than continuous ones. For this reason we think that this tree model technique is well suited to the detection of the most relevant explanatory variables. The exact relationship between explanatory variables and response variable can then be refined with other techniques such as generalized linear or non-linear models. Our analysis revealed a very non-linear relationship between the seasonality of ILI and the seasonality of absolute humidity that is almost a threshold effect. Whether this threshold effect emerges from the non-linearity of the transmission process (as in Dushoff et al. (2004)) or rather from the nonlinearity of the effect of absolute humidity on virus transmission (as in Shaman and Kohn (2009)) would require the development and analysis of a mathematical model of influenza transmission in which the effect of the seasonality of the force of infection on the seasonality of the incidence could be investigated controlling for all other factors (such as the amplitude or

variance of the force of infection). This will be the focus of further study. The second unusual observation is that in northern Viet Nam the correlation between ILI and absolute humidity is positive (in that ILI peaks are associated with AH peaks with a lag of one month), which contrasts with the negative correlation observed in temperate regions. Clearly, we need to be careful interpreting this result because as mentioned above, our data are ILI, not true influenza which may contribute only around 20% of the total cases of ILI (see chapter 5 result) and (Nguyen et al., 2009).

A major limitation of our study is the fact that we do not work on influenza data but on ILI data. This is particularly a limitation in the tropical countries where ILI does not always correlate well with influenza (Khamphaphongphane et al., 2013; Nguyen et al., 2009). Thus our results should be interpreted more in terms of respiratory disease than specifically influenza. The next appropriate steps from here to understand the potential climatic drivers of influenza transmission would thus be (i) analyses of confirmed influenza cases as well as cases confirmed to be caused by other respiratory viruses, (ii) an analysis of these different viruses epidemic dynamics to determine if immunological interference plays a role, and (iii) the inclusion of climate variables and school-term into this combined system to determine how strongly these extrinsic factors influence disease dynamics. For the first step I have conducted the analysis presented in chapter 5 to examine the relationship between ILI and influenza activities. A further limitation is that AH is influenced by air pressure, and therefore the altitude of stations may have an affect, for which we were not able to account. Even though we used population centroids to identify the meteorological station closest to the main population centres, it is likely that data from one meteorological station does not fully reflect the actual climate of the province.

For southern Viet Nam and the majority of central Viet Nam, climate-ILI associations cannot be detected because of the weak seasonality of ILI case reporting. In this sense, the ILI pattern in southern/central Viet Nam is similar to that of other tropical regions, which exhibit either multiple peaks per year or unpredictable disease patterns. It is these regions whose influenza dynamics have become of interest over the past decade due to the possibility that low-level but long-term influenza persistence in tropical countries may create optimal conditions for generating immune-escape variants that can spread worldwide

(Adams and McHardy, 2011; Boni et al., 2006; Rambaut et al., 2008; Russell et al., 2008a; Viboud et al., 2006a). Despite the importance of understanding the tropics and their role in global influenza dynamics, high-quality influenza reporting time series in tropical Asia remain rare with the exceptions of Hong Kong (Yu et al., 2013), Singapore (Doraisingham et al., 1988), Thailand (Chittaganpitch et al., 2012) and recently Viet Nam.

Since our study is one of the first to quantify the seasonality of ILI and climate factors, we did not have any prior expectations about the explanatory power of specific factors. The finding of a single strong explanatory variable was surprising, as was the output of our world map, which superficially appears to agree with published data (Bloom-Feshbach et al., 2013; Tamerius et al., 2010, 2013) for many regions. This result also points out that we need to run more investigations to understand the role of AH and other factors on ILI in general and influenza in particular. One may comment when they see the global figure that their country/region has seasonality which it is not shown in the map. However, the data have $2.5^\circ \times 2.5^\circ$ spatial resolution, which may be too coarse for some relatively small areas e.g some coastal area or even a country like Singapore or Brunei (less than $1^\circ \times 1^\circ$ spatial resolution). That may cause some loss of detail since a large area (about 63756 km^2) becomes one small square in the map, the seasonality of a specific region can be diluted by nearby regions if it not large enough. This is probably the case in Southeast Asia that is very insular, as well as in New Zealand or on the coast of Australia. Secondly, aH is just one factor, the other factors that may not show strong effect in Viet Nam setting but may play a certain level of significant in other country. A full model of factors could help explain the difference showed in figure 4.10 and the seasonal observed world wide. One another reason is the effect that may happen when a locality is highly linked (by human movements) to another locality of a very different climatic set-up. In that case, the influenza epidemiology may be very much influenced by distant climates. This is probably what happens on the US west coast and in western African. However, even the fit between our model prediction and actual data is not perfect, what is really surprising is that, for most parts of the world, the fit is doing rather well. We have good agreement at the global scale from just an extrapolation of what is observed in Viet Nam. This was rather unexpected and can probably be explained by the rich diversity of climates in Viet

Nam, that seems to represent a substantial proportion of the world's diversity of climates.

The main finding from our time series analysis on subtropical and tropical Viet Nam is that ILI seasonality is most closely associated with AH seasonality. The strength of AH seasonality seems to correlate well with ILI seasonality in other parts of the world where both climate and influenza data are available, but the most surprising fact about this association, as can be seen in northern Viet Nam, is that AH and ILI can correlate either positively (this study) or negatively (other studies). If positive AH-ILI associations are seen in other tropical or sub-tropical climates, then we may have to revisit our analyses of these climate associations to determine, if absolute humidity is controlled for, which other climatic factors have the strongest effects on ILI of influenza incidence. Additionally, we may want to consider that non-climatic factors may have an important influence on influenza dynamics. This underscores the need for continued future studies on seasonal patterns of influenza transmission in different regions of the world.

Finally, since ILI notification and influenza activity do not always correlate well in the tropics, a critical area of future research must include the study of other respiratory viruses in Viet Nam to understand the varying components that make up the ILI time series signal presented here. In chapter 5, I examine the association between confirmed influenza and ILI notifications.

4.5 Conclusion

Our analysis shows that in Viet Nam the strength of ILI seasonality is best explained by the strength of AH seasonality. The nonlinear relationship between ILI and AH seasonality could be due either to nonlinearity of the transmission dynamics or to a threshold effect of AH on ILI transmissibility. A mathematical model will help to assess which hypothesis is more likely. Validation of the global extrapolation of our results is required, but nevertheless contributes to our under-standing of ILI seasonality around the world.

Chapter 5

Synchrony of ILI and sentinel virological surveillance data in Viet Nam

5.1 Introduction

As mentioned in the previous chapter, influenza burden has been defined differently in temperate and subtropical/tropical regions (Gleeson et al., 2005; Lee et al., 2009; Neuzil et al., 2000; Nicholson et al., 2003; Wenger and Naumova, 2010; Wong et al., 2006; Yang et al., 2008). In Viet Nam, influenza H1N1 was changed from a category B (dangerous) to a category A (very dangerous) infectious disease Ministry of Health (2009), and the action showed the importance of influenza surveillance, which reflects the World Health Organization (WHO) Global Agenda on Influenza (Stohr, 2003). The limitation of the routine ILI surveillance system as described in chapters 2-4 is that it is based on non-specific symptoms, that are common in influenza infection but also infections caused other pathogens. This reduces the reliability of using increased consultation rates for ILI as a signal for influenza epidemics. In addition, it has been reported that none of the ILI symptoms, except for possibly fever, can reliably differentiate influenza infections from those caused by other respiratory pathogens (Navarro-Marí et al., 2005). Sentinel site surveillance with laboratory diagnostics can provide more accurate information about influenza virus activity than syndromic surveillance, and it has been implemented in parallel with ILI surveillance in many regions (Beckett et al., 2004; Jian et al., 2008; Meerhoff et al., 2004; Nguyen et al., 2009; Paget et al., 2007; Torner et al., 2012; Yang et al., 2008; Zaman et al., 2009). Clearly, sentinel surveillance with laboratory confirmation takes more time and requires

a larger budget, which makes it to be restricted in scale with relatively smaller numbers of patients tested, and with the primary aim being the detection of genetic and antigenic changes of circulating strains. Furthermore, in many regions of Viet Nam, there is no laboratory surveillance at all, thus emphasizing the need for validated alternatives. Previous studies on the predictability of influenza activity from clinical surveillance of ILI mainly focused on the sensitivity and specificity of case definitions for ILI in laboratory-confirmed influenza infection (Call et al., 2005; Monto et al., 2000; Ohmit and Monto, 2006; Thursky et al., 2003). One method to test the predictive value of ILI for influenza activity is to examine whether increases in ILI consultation rates in clinics precede or parallel increases of virologically-confirmed influenza virus activity. In temperate regions, these two quantities do correlate positively (Earn et al., 2002). However, in the tropics and subtropics, influenza virological surveillance data seems to exhibit an irregular seasonality, with unpredictable peaks appearing during winter to summer (Viboud et al., 2006b), while ILI rates tend to be constant and sometimes flat. As a result, the same technique that I use in chapter 2 and 3, the wavelet analysis, may be usefully applied to this problem. In this chapter, we use data from the ILI sentinel surveillance system in Viet Nam as described in Chapter 4, and the sentinel surveillance of virologically confirmed influenza activity as described below. I attempt to quantify the synchrony between ILI consultation rates and the virus activity.

5.2 Materials and methods

5.2.1 The Data

The data comes from Viet Nam's sentinel influenza surveillance system, which is a national network of sentinel hospitals. The number of reporting hospitals has however not remained constantly each year. The system started with seven sentinel sites on January 1, 2006 and expanded to 15 sites by July 2007. The sentinel sites are outpatient clinics located at two central referral hospitals in Hanoi (the North) and two in Ho Chi Minh City (the South), two provincial hospitals, seven district hospitals and two urban polyclinics. (Figure: 5.1) (Nguyen et al., 2009). Sites were selected to include adult and paediatric

patient populations in four major geographic (the north: 7 sites; the south: 4 sites; the central coastal: 3 sites; the central highlands: 1 site) and different ecological regions (temperate, tropical, highlands, Mekong River Delta and Red River Delta) of Viet Nam. NIHE and the 3 other regional hygiene and epidemiology and Pasteur institutes in Viet Nam provide epidemiological and laboratory diagnostic support to surveillance activities at sentinel sites within their respective regional jurisdictions.

All specimens from the sentinels are sent to regional laboratories for testing and results are administered overall by the National Institute of Hygiene and Epidemiology (NIHE), MOH, Ha Noi.

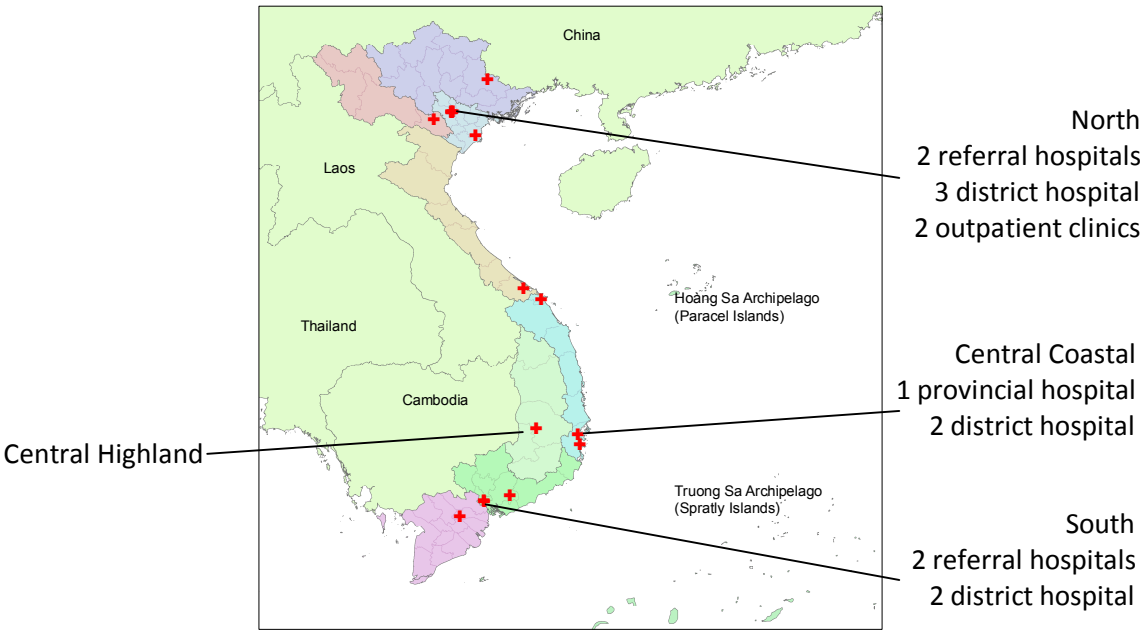


Figure 5.1: Location of national influenza surveillance system sites.

Each sentinel site collected data every weekday on the total number of patient visits and the total number of visits for influenza-like illness ILI, using the WHO case definition of: (1) sudden onset of fever (temperature >38°C) , and (2) either sore throat or cough, and (3) an absence of other diagnoses (WHO, 1999). Demographic and epidemiological data and a throat swab were collected at each site from the first two patients identified each day with ILI and illness onset fewer than three days earlier. Throat swab specimens were stored at each surveillance site at 4°C and sent twice a week to a regional

laboratory, where influenza testing was performed weekly or specimens were stored at −70°C for later testing. All specimens were tested for influenza A, B, H1, H3 and H5 by conventional reverse transcription polymerase chain reaction (RT-PCR) or real time RT-PCR (rt RT-PCR) using primers, probes, and reagents recommended by the CDC and the WHO. Anonymous laboratory testing results and epidemiological data were entered into a database at NIHE. Specimens that were tested positive for influenza by RT-PCR at NIHE and approximately 30% of the influenza-positive specimens from the three other regional laboratories were inoculated onto MDCK cell culture for influenza viral isolation at NIHE's influenza laboratory, a WHO designated National Influenza Centre. Influenza A viral isolates that were not able to be subtyped and a subset of type A and B isolates were sent to the CDC for confirmation and strain characterization. CDC characterized a subset of influenza viruses isolated in Viet Nam by hemagglutination inhibition using ferret antisera.

5.2.2 Analysis

Data were analysed by descriptive statistics using R and MATLAB. The number of total patient visits and ILI cases and percentage of ILI cases testing positive for influenza at each sentinel site were analysed over time. Data were compiled through weekly influenza virus identification in nasal-pharynx and throat systematic sampling (2 first ILI consultations per day per site), and incidence of ILI consultation reporting by the same site. In the analysis, data were presented by individual site but we also merged data from sites in the same region to increase the power to detect patterns. Each sentinel site was assigned a target number of patients to recruit, distributed into five age groups 0-4, 5-14, 15-24, 25-64 and older than 64 years. Since the population structure of subjects recruited is based on quota sampling from the sentinel sites, it does not reflect the real population structure. In the descriptive part of the results, the raw data are presented as well as the results adjusted for age structure. I adjusted the positivity rate for each population age group by multiplying the proportion of samples positive by the proportion of the total population represented by that age group, and then summing the products to derive an overall age-adjusted positivity rate.

We used the wavelet transform as described in chapter 3 and 4. The coherence between two time-series (ILI and positive rate) is used as an exploration of periodic structure of all the times series. High coherence suggests the capability of one time series to predict the other one (Grinsted et al., 2004). We then filtered the time series around the seasonal component (0.9 – 1.1-year band), and computed the phase of the filtered time series. By calculating the phase difference from wavelets calculated for two timeseries, we were able to quantify the lag between the two (Grinsted et al., 2004).

5.3 Results

From January 1, 2006 through December 31, 2012, a total of 3,914,834 patient visits were recorded at the 15 sentinel sites. Of these, 449,907 (11.5%) were patient visits for ILI. Eight percent (37,744) of those with ILI were sampled and tested for influenza; 20.8% (7849) of those tested were positive for influenza by RT-PCR. Fifty one percent of the ILI cases tested for influenza were male. The median age of tested cases was 9 years and the mean age was 17.6 years (range: 1 month–94 years). Sixty one percent of tested cases and 63.8% of influenza-positive patients were among children aged <15 years old (Table 5.3). The highest proportion of ILI cases that tested positive for influenza was among persons aged 5–14 years (29.1%) and the lowest proportion was among persons aged >64 years (11.5%). When adjusted for age of the sampled population compared to the general Vietnamese population, the overall proportion of influenza positives was 20.4%. Influenza viruses circulated throughout Viet Nam during 2006–2012 with more than 10% of ILI cases testing positive each month except for 7 of the 84 months during the surveillance period (Oct 2006, Dec 2006, Sept 2007, Feb 2009, Jan-Feb 2010 and May 2011) (Figure: 5.3). Influenza A viruses, including influenza A subtypes, and influenza B viruses peaked during different periods. After October 2009, A/H1N1 was fully replaced by the A/H1N1/2009 pandemic virus and A/H1N1/2009 began circulating in a manner similar to seasonal influenza (i.e. the previous 1977-lineage A/H1N1).

5. SYNCHRONY OF ILI AND SENTINEL VIROLOGICAL SURVEILLANCE DATA IN VIET NAM

Table 5.1: ILI surveillance data by year

Influenza-like illness (ILI) surveillance	2006	2007	2008	2009	2010	2011	2012
Total sites	7	15	15	13	11	13	13
Total outpatient consultations	688,960	489,373	649,096	774,209	560,247	425,480	327,469
Total ILI patients	123,460	60,806	74,716	88,711	42,139	41,418	18,657
Total ILI samples tested	4,641	6,459	6,994	7,380	4,398	4,444	3,433
Total samples influenza positive (%)	947 (20.4%)	1,170 (18.1%)	1,493 (21.4%)	1,935 (26.2%)	979 (22.3%)	651 (14.7%)	674 (19.6%)

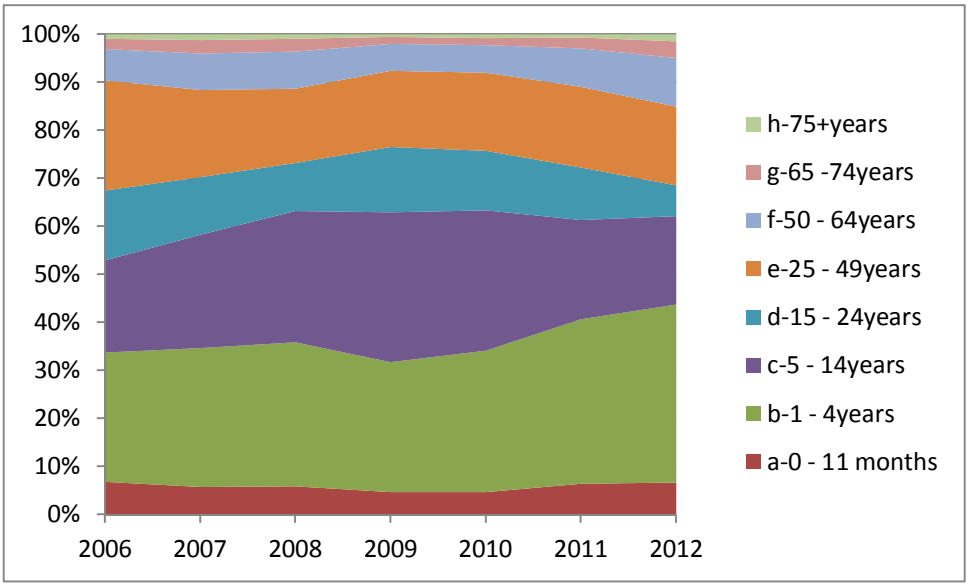


Figure 5.2: ILI sample collected by age group and year

Table 5.2: Results of tested ILI samples, by influenza virus subtype from 2006 to 2012

RT-PCR results	2006	2007	2008	2009	2010	2011	2012	Total (%)
Negative	3,694	5,289	5,501	5,445	3,419	3,793	2,759	32,858 (79.2)
A/H1	574	16	643	63	0	0	0	1296 (3.1)
B	315	274	618	452	505	174	398	2989 (7.2)
A/not sub-typed	2	1	4	0	0	1	0	8 (0.02)
A/H3	56	876	225	717	379	20	267	2844 (6.9)
A/H3&B	0	3	1	7	1	0	0	15 (0.04)
A/H1N1pdm09	0	0	0	696	94	455	9	1466 (3.5)
A/H1N1pdm09 & B	0	0	0	0	0	1	0	5 (0.01)
Total Tested	4,641	6,459	6,994	7,380	4,398	4,444	3,433	44,208 (100)

Table 5.3: Distribution of ILI and tested cases by age

	ILI consultation	ILI tested	% tested	% Positive	Population 2009	Adjusted for population
< 5 year	215,017	13,470	35.68	16.7%	7,034,144	1.37%
5-<15 year	84,101	9,459	25.06	29.1%	13,959,115	4.73%
15-<25 year	42,025	4,405	11.67	23.4%	17,396,769	4.74%
25-<65 year	79,170	9,228	24.45	18.1%	41,942,170	8.86%
65+	29,566	1,187	3.14	11.5%	5,514,799	0.74%
Total	449,879	37,749				20.4%

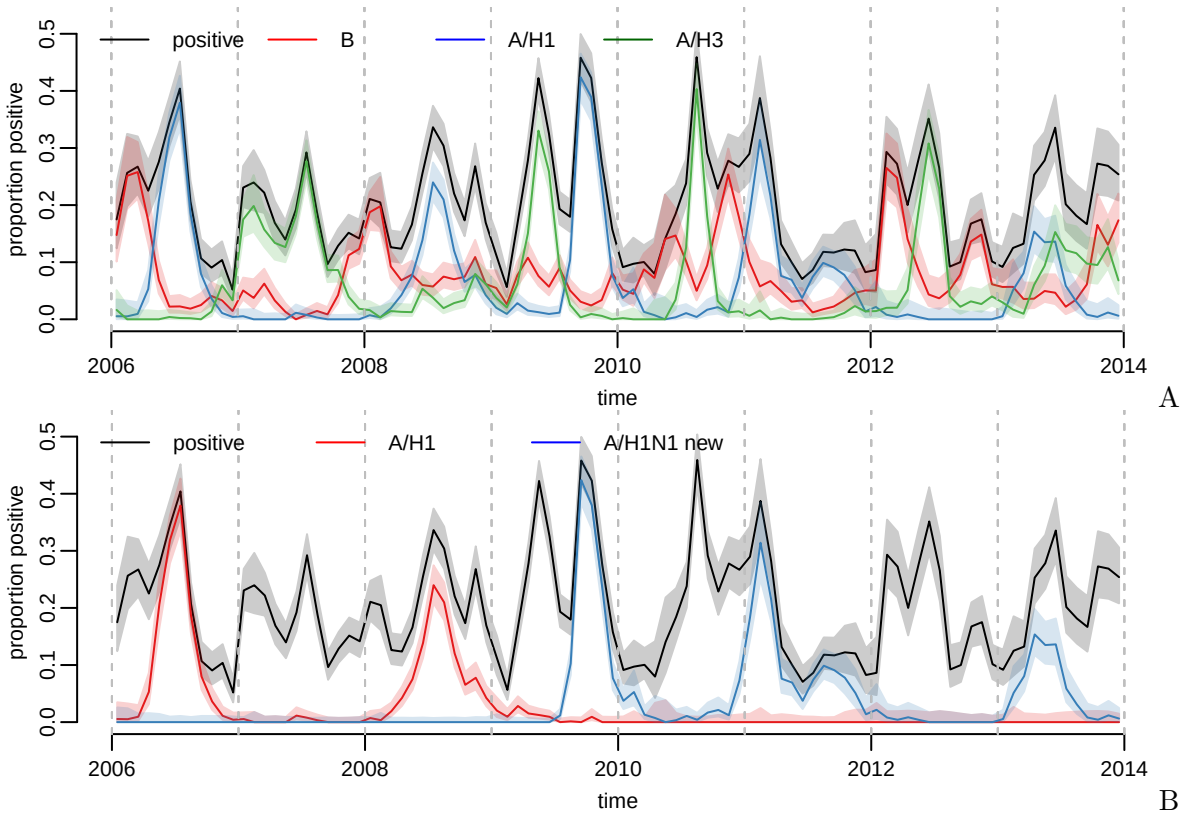


Figure 5.3: Proportion of swabs that are influenza positive by time (monthly)
The opacity around each line is 95% confidence interval of the proportion. In panel A of this figure, the blue colour combines both A/H1N1 and A/H1N1/2009

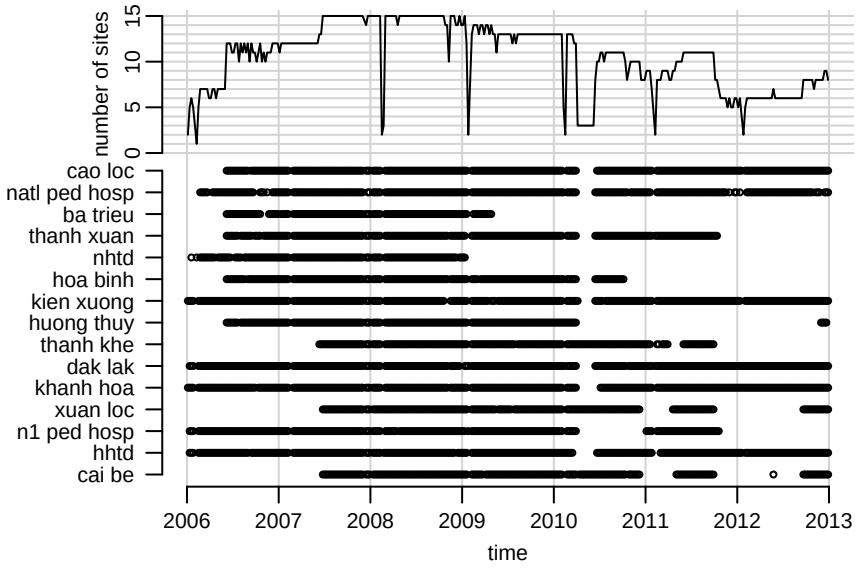


Figure 5.4: Number of sites and completeness of data collection, by week.
Each ∘ is equal to one week of data, the longer the break point the more missing data.

Individual hospital reports, missingness, and virological Confirmation Trends

Due to poor performance, some sites were closed during the project (table 5.1, figure 5.5 to 5.8 and figure 5.3) which makes their time-series less valuable. Missing values also appear during the end and the beginning of the year due to budget constraints (figure 5.4). Despite these limitations in the data, we try to determine whether there is periodic behaviour in the data with a one year cycle. Figures 5.5 to 5.8 show the PCR positive proportion by site and by region. These figures show weekly PCR positivity rates, and only rates greater than 25% are displayed, showing the times when positive rate were higher than the average positive rate. Although highly qualitative, some patterns may be seen in these figures, for each site in the north (with the exception of Kien Xuong), there seems to be higher influenza activity from June to October during the period 2006 to 2008. When H1N1/2009 arrived in the summer of 2009, the picture of influenza activities changed to an earlier onset, which somehow interrupted the previous seasonal pattern. A similar picture, but with less certainty around the peak summer period, is seen in the highland provinces (Buon Ma Thuot), in line with the findings from chapter 3 that highland seasonal ILI patterns are similar to northern seasonal ILI patterns. The situation for sites in central Viet Nam is different from the north and south, as influenza confirmation occurs throughout the year. Southern Viet Nam shows a high proportion of influenza during the

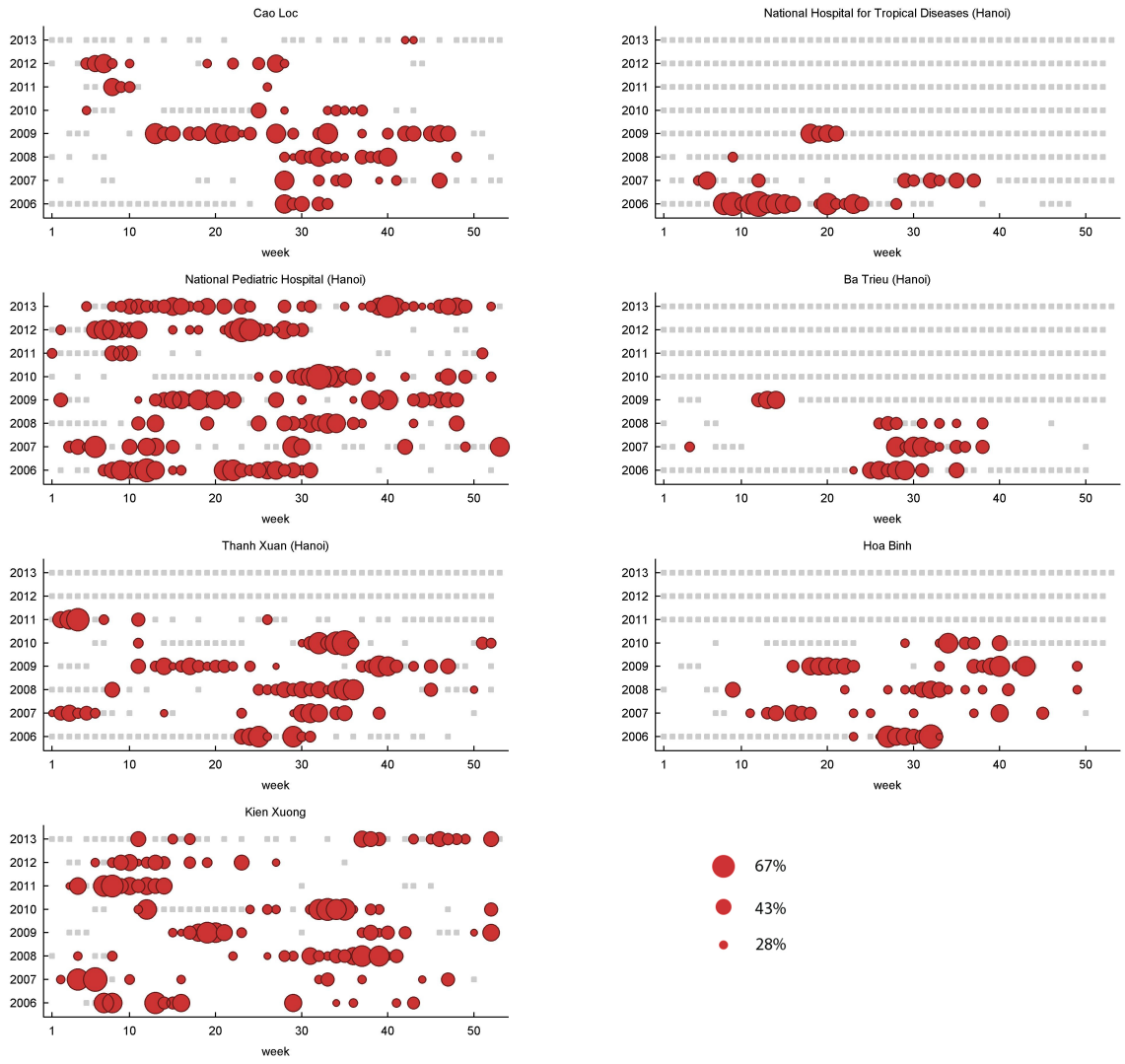


Figure 5.5: PCR positive proportion (>25%) by week for all northern sites. Each ○ is equal to one week of data. The size of the circle represents the proportion of samples positive (only weeks for which the PCR positivity rate is higher than 25% are shown on the graph. If the PCR% is lower than 25% then there is a blank white space in the graph). Grey squares are missing data.

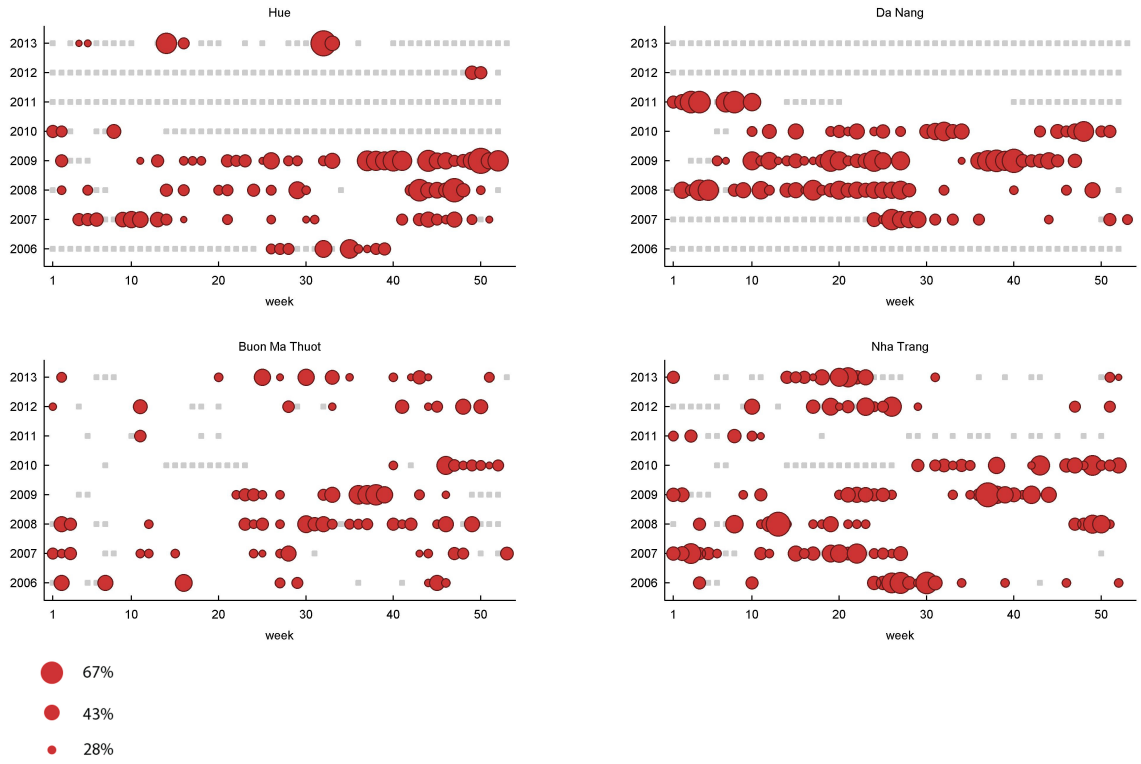


Figure 5.6: PCR positive proportion (>25%) by week for all central sites.
Each ○ is equal to one week of data. The size of the circle represents the proportion of samples positive (only weeks for which the PCR positivity rate is higher than 25% are shown on the graph. If the PCR% is lower than 25% then there is a blank white space in the graph). Grey squares are missing data.

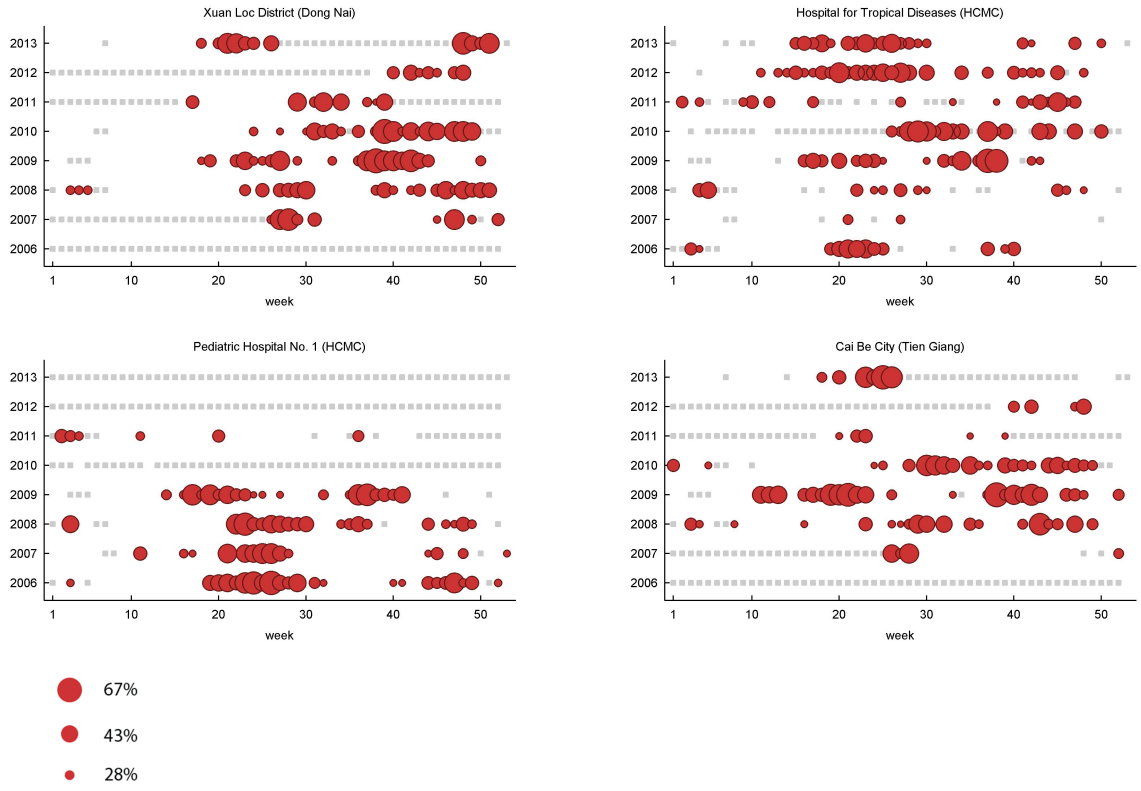


Figure 5.7: PCR positive proportion (>25%) by week for all southern sites. Each ○ is equal to one week of data. The size of the circle represents the proportion of samples positive (only weeks for which the PCR positivity rate is higher than 25% are shown on the graph. If the PCR% is lower than 25% then there is a blank white space in the graph). Grey squares are missing data.

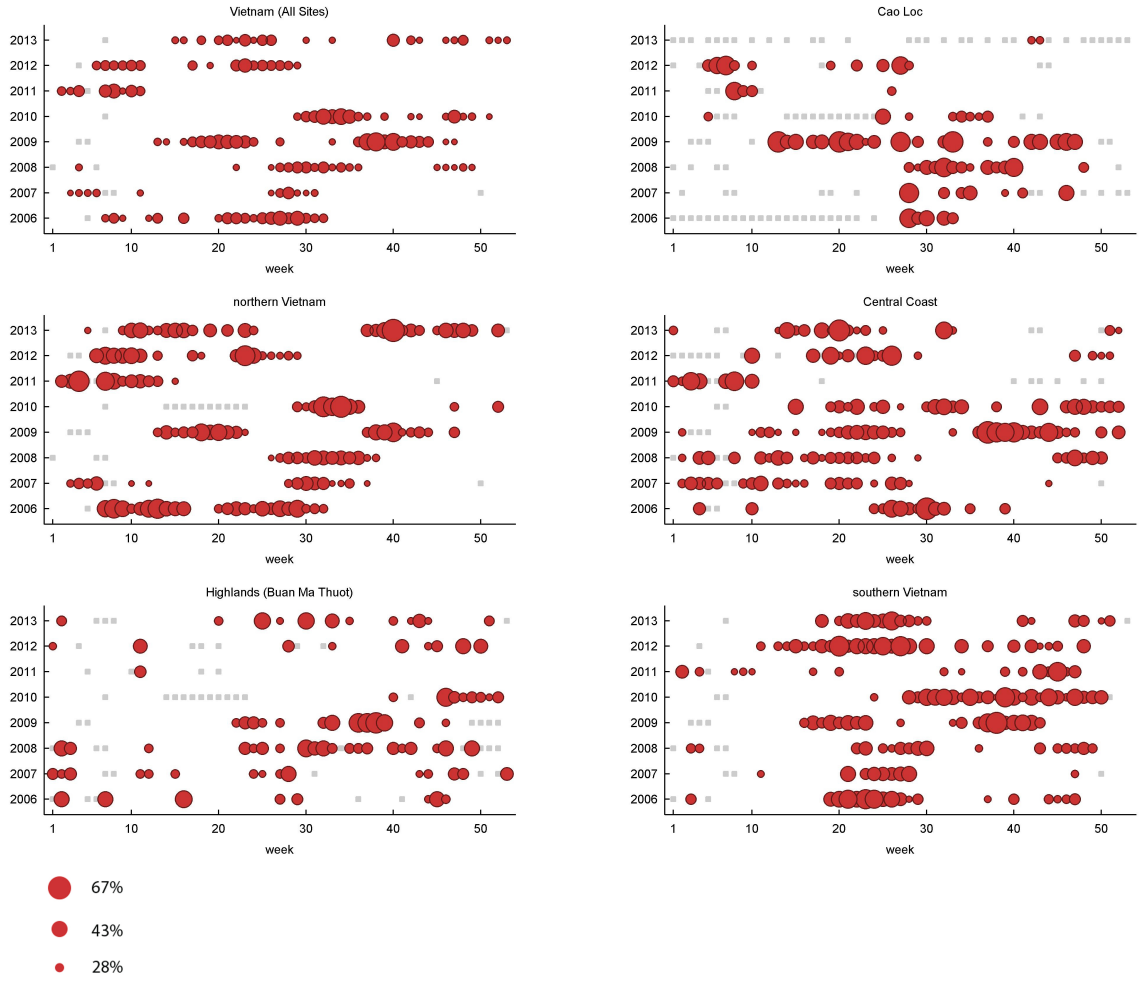


Figure 5.8: PCR positive proportion (>25%) by week for all region and all site. Each ○ is equal to one week of data. The size of the circle represents the proportion of samples positive (only weeks for which the PCR positivity rate is higher than 25% are shown on the graph. If the PCR% is lower than 25% then there is a blank white space in the graph). Grey squares are missing data.

rainy the season, from April to November.

5.3.1 Regions Reports and Virological Confirmation Trends.

Since the results from individual hospital are noisy and based on limited samples, the aggregated data from all sites from one region are a better illustration of regional influenza dynamics. Figure 5.8 shows PCR positive confirmations for the northern, central, and southern regions as well as the whole of Viet Nam. The northern data show a period of more than one year with influenza shifting to later and later epidemic onsets until the pandemic H1N1/2009 arrives, changing the pattern of epidemic dynamics. This is perhaps one of the most important features of pandemic dynamics to study, but with only three years of data prior to the pandemic, we cannot show with certainty that the epidemic dynamics significantly changed. Central Viet Nam and the highlands show influenza activity throughout the year, and it is difficult to see any clear consistent pattern. Southern Viet Nam shows a spring and early-summer epidemic pattern, but it is clear the H1N1/2009 pandemic interrupted the dynamics temporarily. The unusual feature of the dynamics in the south is that they span both the dry and rainy season.

5.3.2 Do the data show influenza seasonality in Viet Nam?

We start with simple Fourier Analysis (see equation 5.1), see figure 5.9.

$$y = c_0 \; + \; \sum_{k=1}^{8} a_k \sin(k\omega t) + b_k \cos(k\omega t) \tag{5.1}$$

After fitting a standard Fourier series to the PCR positivity rates, allowing up to eight terms, we find that the coefficients of the annual component of the Fourier series is close to zero and statistically not significantly different from zero. The bars in this figure show the 95% confidence intervals, and the numbers above the bars show the inferred period (in months) that was closest to one year. Hence, the seasonal picture of influenza dynamics is not straightforward, as the pattern is not as regular as the patterns we see in temperate countries with wintertime influenza. Nevertheless, we can still attempt to link the pattern of influenza positivity to ILI reporting (chapters 3 and 4) as the ILI

reporting was not perfectly seasonal either, especially in southern Viet Nam. In southern Viet Nam in particular, there is a possibility that neither influenza nor ILI are seasonal, but nevertheless that they do correlate with each other.

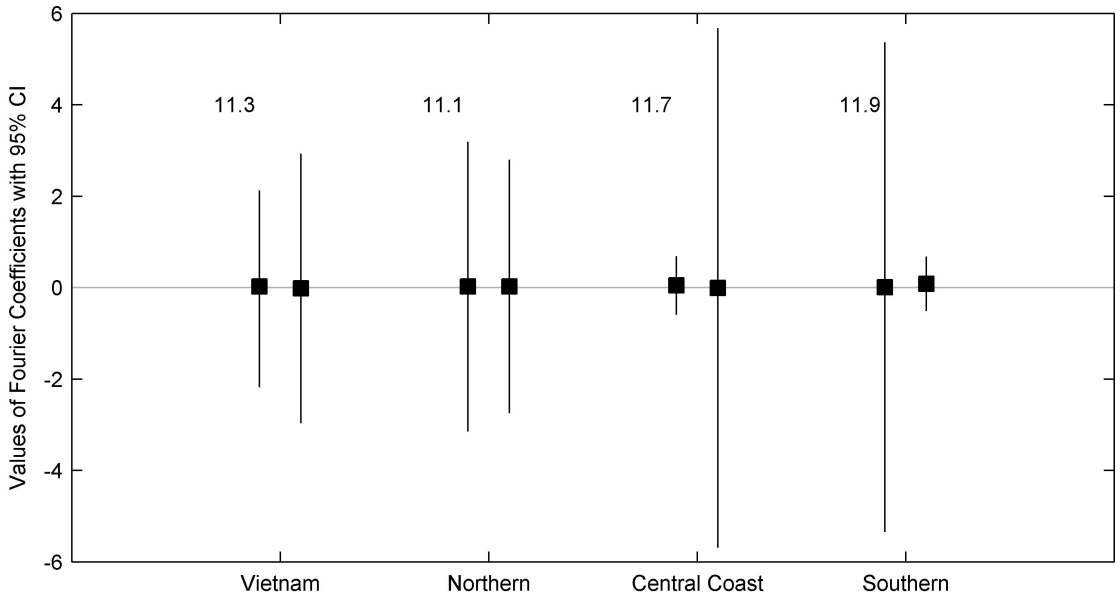
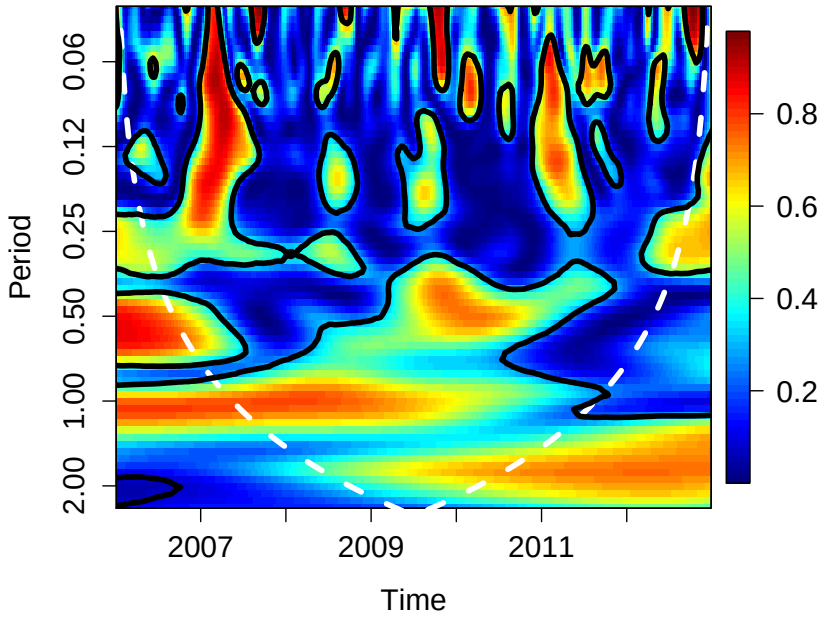


Figure 5.9: Standard Fourier fit to PCR positive rate.
The numbers noted by the text above the bar show the exact values of the period (in months) for the Fourier coefficients of terms that are closest to a frequency of 12 months. The left bars show the a_k coefficient (for sine) and the right bard show the b_k coefficient (for cosine) (see equation 5.1)

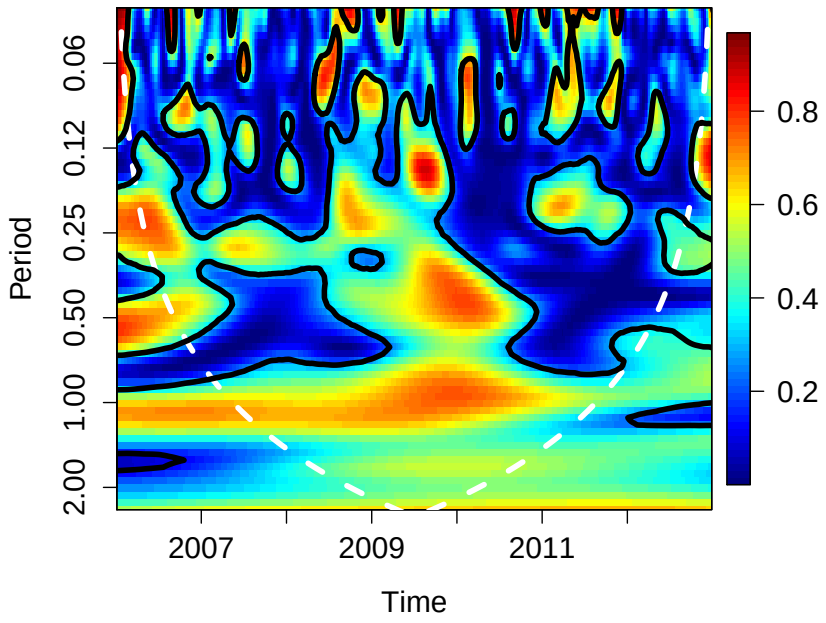
5.3.3 Synchrony between ILI consultation and PCR positive rate

Patterns of coherence were observed between ILI consultation rates and virus activity at both the annual and semi-annual cycles (Figure 5.10). Those patterns were observed in both northern and southern part of Viet Nam. The significant coherence at one year cycle throughout the whole period but from 2010-2012 the semi-annual also have high coherence.

Since the data are noisy with some missingness, as mentioned above, we applied a detrending and transformation similar to what was done in chapter 3 and 4. This gives us a result that ILI and influenza positive rate are coherent at the one-year age-band (Figure 5.10). What we would like to analyse now is the lag between oscillation of ILI and



A



B

Figure 5.10: Wavelet analysis for the weekly proportions of specimens positive for influenza.
Panel A shows the coherence between ILI consultation after transform and detrend and PCR positive rate in the northern sites. Panel B shows the same coherence for the southern sites. The cone of influence (broken white line) indicate the region without edge effects. The power values were coded from dark blue for low power to dark red for high power as shown in the colour bar on the right.

Table 5.4: Monthly averages of influenza isolation proportions and ILI consultation rates for whole country, 2006-2012

month	% positive	% ILI/all consultations
1	18.11	11.90
2	19.33	12.83
3	18.70	14.67
4	18.24	12.98
5	24.10	11.72
6	24.99	11.95
7	25.66	10.74
8	22.60	10.85
9	20.91	10.39
10	19.43	10.90
11	20.42	10.05
12	13.97	10.60

oscillations of influenza activity. Filtering the data using the power band of 0.9-1.1 years, we calculate the phase difference. The result is shown in Figure 5.11.

We see a synchrony between ILI consultations and influenza positivity rates, that is clearest from late 2007. The larger phase difference at the beginning and end of the times series is probably a result of inconsistent data collection and/or drop out of sites. We also conducted phase analysis by age group, by virus subtype and by region. Children aged from 0-15 have higher phase difference at the beginning and end (up to half a year of lag) than adults and age group 65+ have the smallest phase difference. Even if we did transform and detrend data, the large phase difference of small children group make the phase difference in all-age group. Take into account that the young children group mainly come from National Paediatric hospital (account for about 3/4 of total ILI at first year), we decided to remove data from the National Paediatric Hospital in Ha Noi data, where in the first and second year they recorded a very large number of ILI consultations (from 4 clinics in one hospital) which were unrelated to the sample collection site (only one clinic in the hospital). The results are shown in Figure 5.12 and 5.13, which show the relationship more clearly. The average phase difference between ILI consultation and positive proportion is around 2 week (1.6-3.8). The ILI consultations increase first, followed by an increase in the proportion of ILI cases that test influenza positive.

5.4 Discussion

The overall positivity rate for all sites of influenza virus (types A and B) was significantly higher in the age group of 5-14 than in the other age groups (chi-square tests; P<0.001). This was also been observed by other authors Reina et al. (2009); Torner et al. (2012). Furthermore, several studies also found out that school and pre-school children have an important role in transmitting infection to others (Cauchemez et al., 2008). The results are also in agreement with the observed pattern of social contacts Viet Nam, which show that the age group of 5-14 has higher contact activities compared to other groups (Horby et al., 2011).

Unsurprisingly when H1N1/2009 arrived, the north, centre and highland seemed to be affected more by the pandemic's changing the near-term pattern of influenza transmission in those regions; however, in the south, the period of influenza seemed to be unchanged. To understand this behaviour will require a detailed analysis of a much longer time series to determine (1) if there truly is a predictable regional pattern of influenza transmission in Viet Nam, and (2) if these dynamics were potentially interrupted by the 2009 pandemic. The second important question we will need to investigate in the future is the predictability of influenza by ILI trends in Viet Nam. The coherence analysis point out that the oscillating patterns of these trends have some similarity, and as long as the phase difference between ILI and influenza is a constant we will be able to predict one with the other. This result support the study of Yang et al. in Hong Kong (Yang et al., 2008), another subtropical climate that may exhibit similar influenza dynamics to northern Viet Nam.

The major limitation of the data set relates to the short time frame, the limited number of sites, and missing values due to some locations stopping acting as sentinel sites due to budgetary limitations. In the sentinel data, the missing data points cannot be interpolated in the same way as that for the routine data, which makes these data more difficult to interpret. It is necessary to improve the mechanisms and support structures for the influenza sentinel surveillance network in order to improve the quality, coverage and sustainability of the system. A further issue is the oversampling of children, who

are affected by a wide variety of respiratory viruses. In 2006, most of the ILI cases reported from two paediatric hospitals might be caused by other viruses, which also caused syndromes similar to ILI, but were not properly distinguished from ILI during the early stages of sentinel reporting. We consider RSV as the most likely candidate virus that could caused this difference. This issue was also discussed by Yang et al. (2008). After 2006, three out of four data sets obtained from paediatric sites were removed, and we see that the coherence between ILI consultation and influenza positive rate is closer after this period. However, this coherence changes again from the end of 2009 to the middle of 2010, suggesting that H1N1/2009 virus might caused this effect. This result is not surprising as other studies have noted that synchronous dynamics can be interrupted by pandemic-like events that have large effects on host immunity (Finkelman et al., 2007). The H1N1/2009 virus did indeed have a high attack rate and induced significant immunity in the population. We also noted an increase in ILI consultation rates due to the effect of mass media around the time of the 2009 pandemic (Thai et al. unpublished data), which requires further caution in interpretation of the results.

5.5 Conclusion

Our results indicate that ILI notification data in Viet Nam are associated with influenza virus activity but with some uncertainty. It will be important to analyse longer time series and to obtain high-quality reporting data to determine the association between ILI notifications and influenza virus activity more precisely.

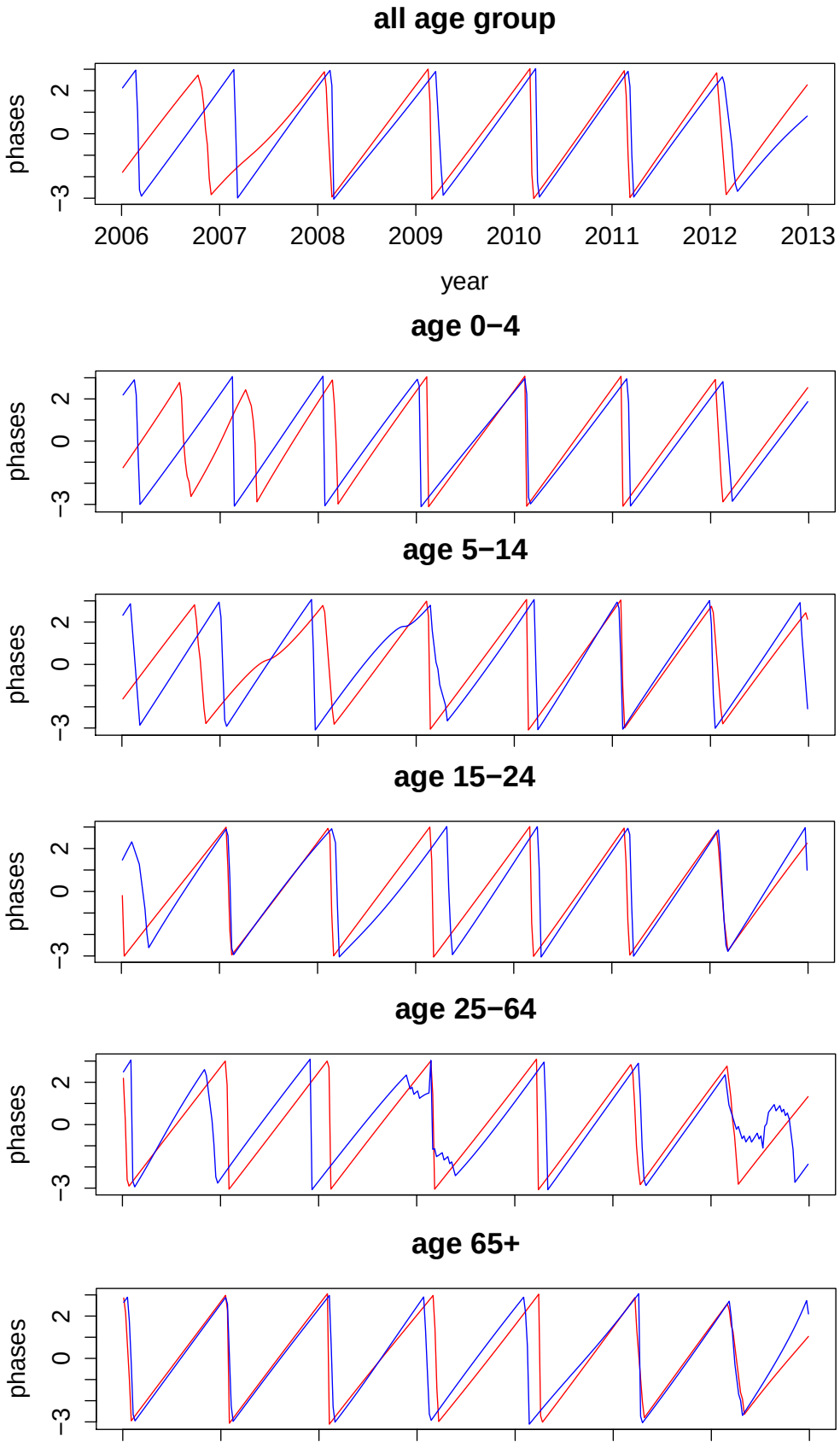
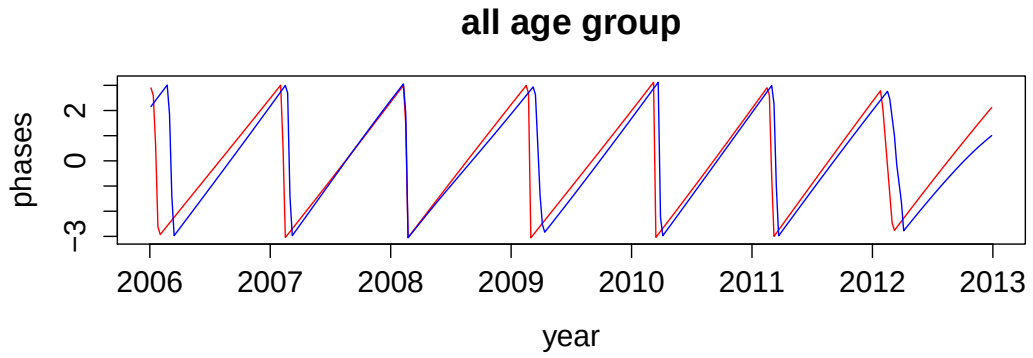
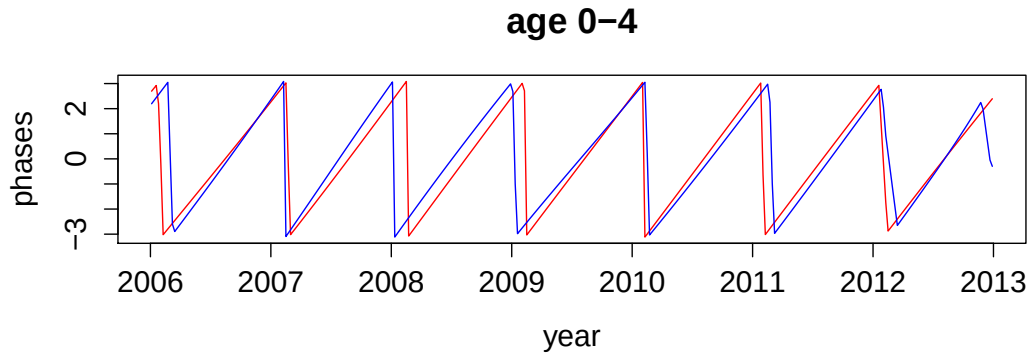


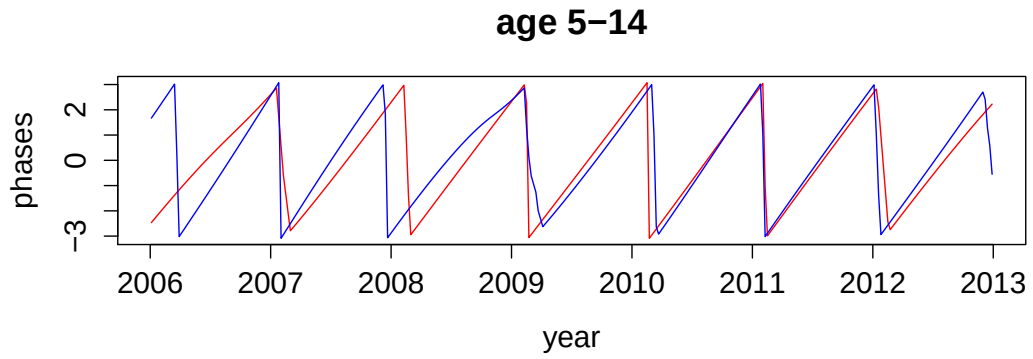
Figure 5.11: Phase difference between ILI consultations and PCR positive rate, by age. The red line is ILI consultations and the blue line is PCR positive proportion, all filtered at power band of 0.9-1.1 year.



A



B



C

Figure 5.12: Phase difference ILI consultations and PCR positive rate after removal of National Paediatric Hospital. The red line is ILI consultations and the blue line is PCR positive proportion, all filtered at power band of 0.9-1.1 year

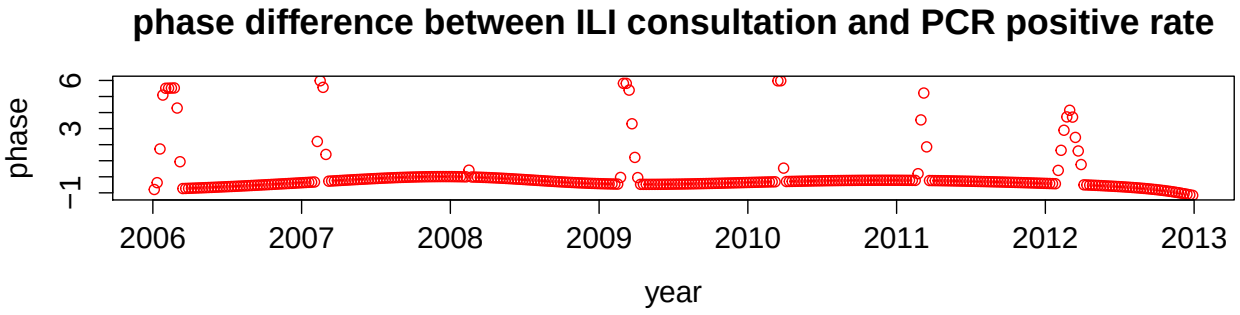


Figure 5.13: Graph of phase difference of ILI consultations and PCR positive rate. Each ∘ is one value compute by take the difference between PCR positive rate phase and ILI consultation phase by subtract directly. The smaller the value the better synchrony between the two variables

Chapter 6

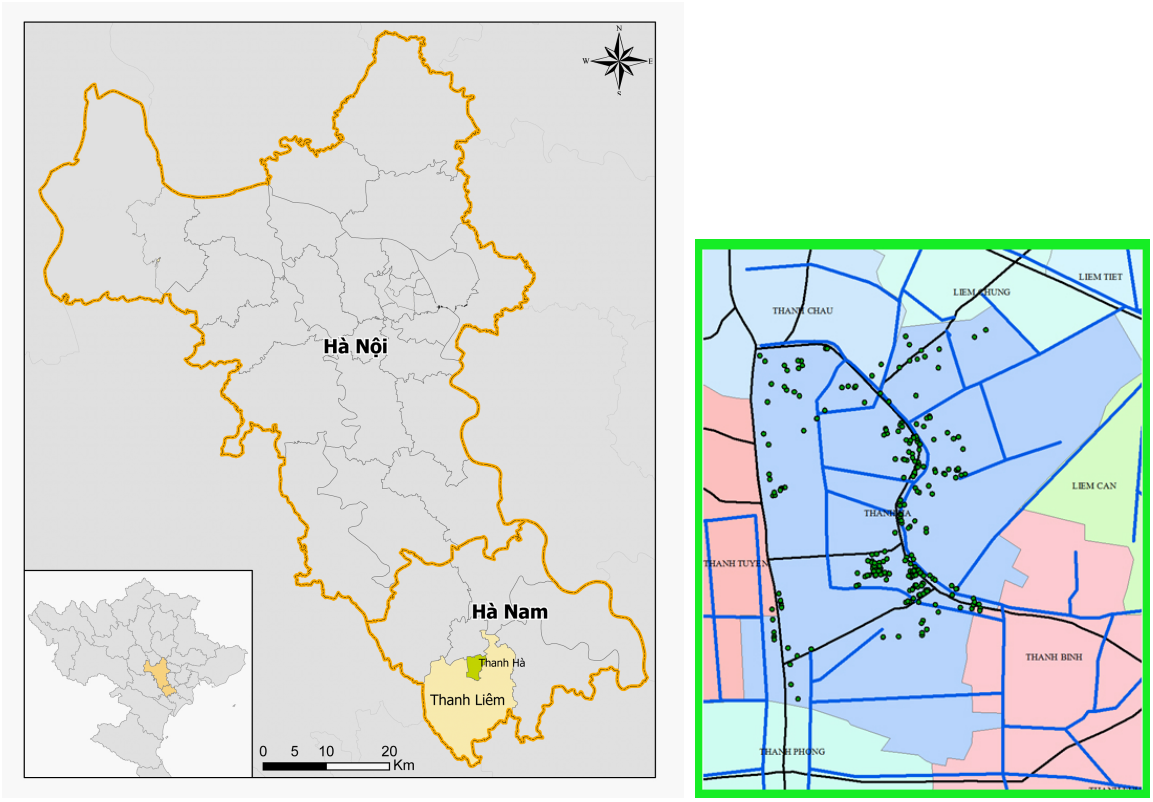
Background to household cohort study

6.1 The Ha Nam longitudinal community study

Reliable estimates of age and risk group specific infection and hospitalization rates are needed in order to inform risk assessments and health policies. However, influenza infection is often mild or even asymptomatic, so cases reported from hospitals are just the tip of an iceberg. In addition, attendances at primary healthcare facilities are also biased, since only a small proportion of all people with influenza infection seek healthcare, and, as shown in chapter 5, only a minority of people seeking healthcare for ILI have an influenza PCR positive respiratory sample. As such, healthcare centered influenza surveillance methods cannot provide a complete picture of the epidemiology of influenza, and community based studies are needed (Garske et al., 2009; Laurie et al., 2013). Studies of respiratory illnesses in the community that have used active surveillance and serology have demonstrated that influenza infection rates are often in the range of 10-20% per season, with only a small proportion developing symptoms Horby (2014); Monto (1994). A recent study from the UK on recent outbreaks of seasonal influenza and the 2009 H1N1 influenza pandemic shows that just 23% of serologically identified infections caused symptoms, and only 17% of people with PCR confirmed influenza infection were ill enough to consult their doctor (Hayward et al., 2013).

In 2007, the NIHE-OUCRU collaboration established a community cohort in Thanh Ha Commune, Thanh Liem District, Ha Nam province of north Viet Nam in order to study the incidence and transmission of seasonal influenza at a household level (Horby et al., 2012, 2011). Thanh Ha is a commune (the third administrative level in Viet Nam after Province and District) that had 7,663 residents living in 2,127 households at the time

of recruitment to the study. The community occupations are a mixture of agriculture and trade, and the site was selected for the practical reasons of 1) being within one day travelling distance of Hanoi (the capital); 2) the Provincial and District health teams were willing to participate in the study; and 3) the Commune had suffered human cases of avian influenza A/H5N1 in 2007 (see figure 6.1).



(a) Ha Noi and Ha Nam, the small green commune is Thanh Ha where we set up the cohort

(b) The household of cohort by GPS location

Figure 6.1: Position and size of Ha Nam cohort. The distance from Ha Nam cohort to centre of Ha Noi is about 70 km

The basic overall design of the study is shown in figure 6.2, and with further details provided below.

6.1.1 Subject identification, selection and recruitment

Agreement to approach the members of Thanh Ha community was reached after discussions with Provincial, District and Commune Preventive Medicine staff and with village leaders, including representatives of the People's Committee, Women's Union, Youth Union and Fatherland Front. An open meeting was also held with all villagers to ex-

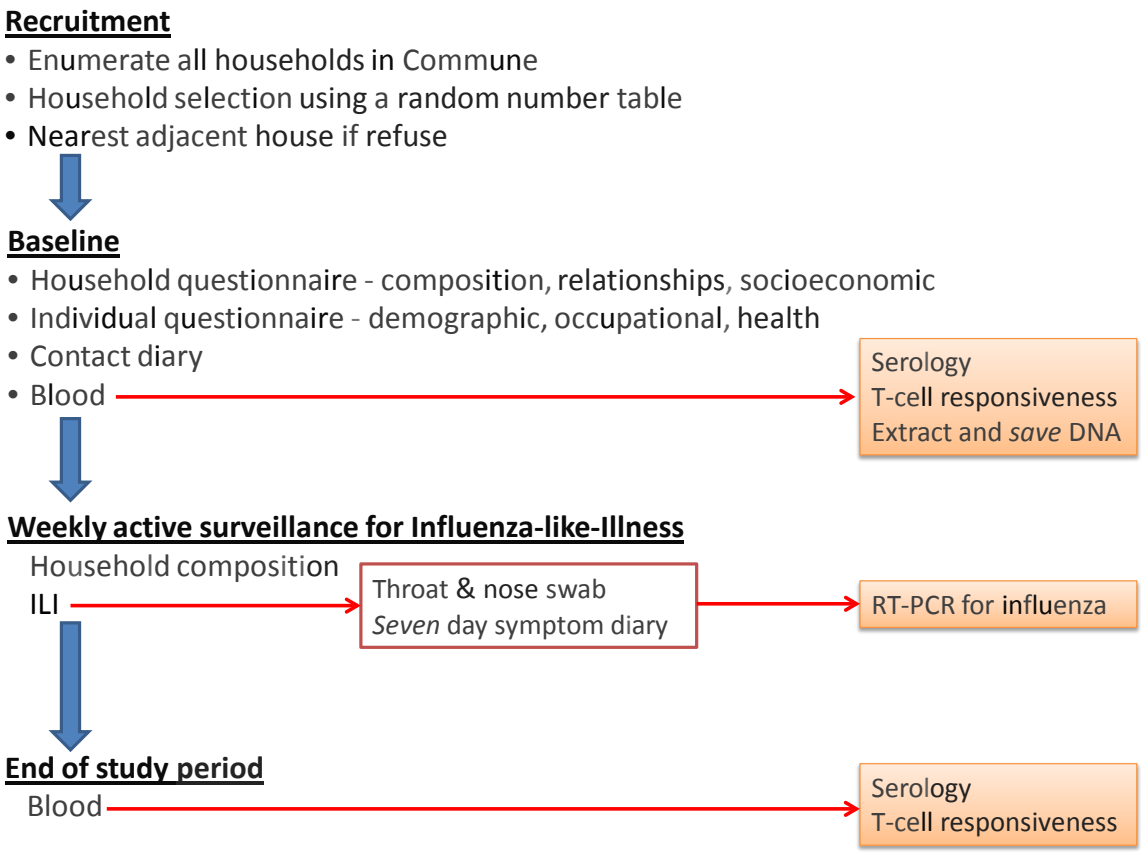


Figure 6.2: Schematic diagram of the design of the Ha Nam household cohort study

plain the study and answer questions. The village population register held by the local authorities was the source document for the selection of households for inclusion in the study. Households were randomly selected from the village population register using a random number table. Those selected were visited by members of the study team, who explained the study objectives, procedures, anticipated benefits and risks and answered any questions. Since the unit of interest is the household, all permanent residents in the household were required to participate. A household was not eligible for inclusion if any members refused to participate. However, children would still be eligible if they decided not to have a blood test as long as other elements of the study were agreed to report for example influenza-like-illness and to provide a nasal swab. Written informed consent was obtained from all participants. If a randomly selected household refused to participate, next nearest household would be approached until a household was successfully recruited.

6.1.2 Information sheet

All participants and individuals asked to provide consent on behalf of another individual (e.g. parents or relatives of deceased adults) were given an information sheet outlining the study objectives, methods and the potential benefits and risks of participation. Children aged 5-17 years were, in the presence of an adult with parental responsibility for the child, provided an age-appropriate information sheet, and had the study explained by a member of the research team.

6.1.3 Consent

Each subject was informed of the aims, methods, anticipated benefits and potential risks of the study in a face-to-face meeting with a member of the study team. Participation is entirely voluntary and all participants had a continuing right to withdraw at any time. Children were defined in this study as persons aged less than 18 years. Children aged 5-17 years were, in the presence of an adult with parental responsibility for the child, provided an age-appropriate information sheet, had the study explained by a member of the research team and were asked to sign an assent form and their parent or legal guardian were be asked to co-sign at the same time. Children aged less than 5 years may enter the study if consent was obtained from a person who had parental responsibility for the child. Adult subjects who were considered not to possess the capacity to fully understand the study and the risks and benefits were not required to sign the consent form. However their legal representative or someone with a qualifying relationship to them were asked to sign the consent on the subject's behalf after the aims of the study were fully explained. A 'qualifying relationship' was defined as spouse, partner, parent, child, brother, sister, grandparent, grandchild, child of a brother or sister, stepfather, stepmother, half-brother, half-sister and friend of long standing. Consent was sought for storage and future testing of biological specimens A copy of the consent form was given to the person who signed it.

6.1.4 Ethical review

The full study protocol was reviewed and approved by the Scientific and Ethical Committees of the National Institute for Hygiene and Epidemiology, Viet Nam and by the

Oxford Tropical Research Ethical Committee, Oxford University, UK.

6.1.5 Enrolment

940 individuals from 270 households were recruited with a study base of 2,127 enumerated households. Figure 6.1b shows the location of all the enrolled households. Figure 6.3 shows the age distribution of the cohort participants compared to that of Ha Nam province and the national rural population. The age distribution in the cohort is significantly different (chi-square tests; both $P<0.001$) due to an over-representation of 10-19 years old and an under-representation of 20-34 years old. The distribution of household sizes of the cohort matches that of the Red River Delta rural population (chi- square goodness of fit test: $P=0.86$). Table 6.1 shows the characteristics of the cohort subject.

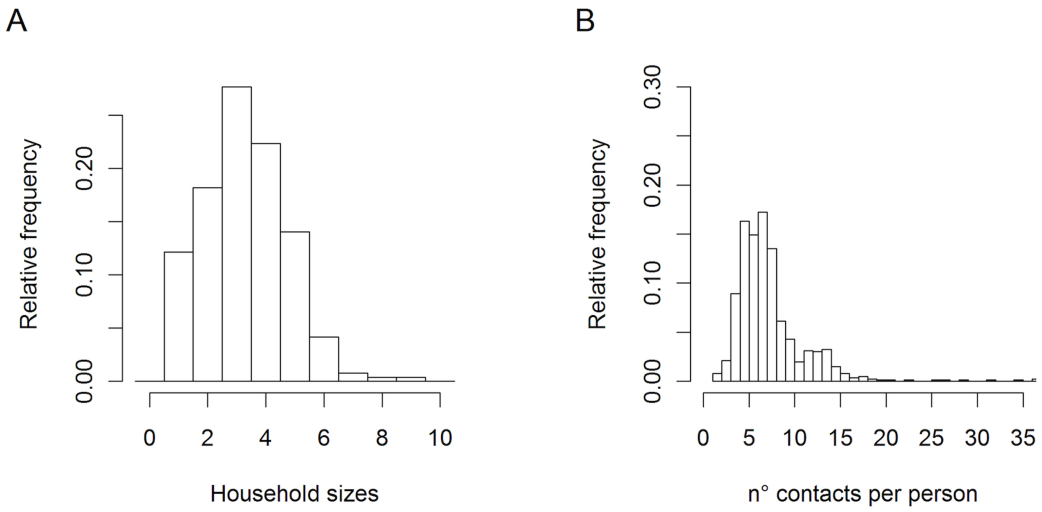


Figure 6.3: The Ha Nam household characteristics.
Household sizes (A) and number of reported contacts per person per day (B). Source: Horby et al. (2011)

6.1.6 Active surveillance for influenza

Households were actively followed up weekly by Village Health Workers (VHW) to detect incident cases of Influenza-Like Illness (ILI) using the WHO and U.S. CDC definition of ILI WHO (1999): an illness with an oral temperature of 38°C or more AND either a cough or a sore throat. Each household was provided with a thermometer for measuring oral temperature. If any household member had an illness meeting the ILI case definition

Table 6.1: Characteristics of Participants and Households at Recruitment, Ha Nam, Viet Nam, 2007–2010. [Source: Horby et al. (2012)]

Characteristic	No. of Participants	Total No. Assessed	%
Entire study population			
Age, years			
0–4	83	929	8.9
5–9	70	929	7.5
10–19	209	929	22.5
20–39	246	929	26.5
40–59	241	929	25.9
=60	80	929	8.6
Female sex	508	932	54.5
Chronic disease[a]	5	869	0.6
Adults (age =18 years)		592	
Caring for children at home or at work			
Never	284	569	49.9
Sometimes	100	569	17.6
Most days	185	569	32.5
Current smoker	107	560	19.1
Cigarettes smoked per day			
=5	49	103	47.6
6–10	45	103	43.7
11–20	9	103	8.7
Households		270	
No. of people in the household			
1	28	270	10.4
2	41	270	15.2
3	65	270	24.1
4	74	270	27.4
5	42	270	15.6
=6	20	270	7.4
Home crowding (>2 people per room)	46	237	19.4
School-aged children in household (5–17 years of age)	156	264	59.1

[a] There were 2 participants with chronic lung disease, 2 with chronic heart disease, and 1 with chronic liver disease.

they would be asked to complete a seven day illness diary and provide a combined nasal and throat swab. During weekly follow up, household composition was re-ascertained. Subjects who are lost to follow up for any reason (e.g. migration, voluntary withdrawal or death) were censored and the period of active observation in days was calculated. If new household members appeared (new born, new married, return from university, return from military), they were informed of the study and asked to join in the same way as existing members including the provision of written consent.

Figure 6.4: Training on taking nasal swabs.

Figure 6.5: Working with Village Health Workers in bleeding campaign.

6.1.7 Twice yearly serology

As discussed in chapter 2, the timing of influenza epidemics in tropical and subtropical regions is variable and may occur more than once per year, therefore the timing of cross sectional bleeding was guided by influenza surveillance data from the cohort itself and from the national sentinel surveillance system described in Chapter 5. All cohort members were requested to provide a 5 ml blood in preservative-free sodium heparin tubes sample twice per year for serological analysis of exposure to influenza.

6.1.8 Twice yearly re-census

In preparation for the six month serological surveys, a re-census of all participating households is conducted.

6.1.9 Monitoring and quality assurance

Weekly team meetings are held every Wednesday between the Field Supervisor and the VHWs to review work, record indicators of VHW activities and to identify problems. Every week a random selection of a 2% sample of all households were selected for re-interview by the field supervisor – to check the completeness and quality of the interviewing (Figure 6.5).

6.2 Laboratory procedures

6.2.1 Detection of influenza infection

Reverse – transcription polymerase chain reaction (RT-PCR) is used to detect influenza RNA in the combined nasal and throat swabs.

6.2.2 Detecting antibodies

Subtype specific antibodies against influenza are detected by haemagglutination inhibition (HI) assay. A fourfold or greater rise in antibody titers is considered to indicate evidence of acute infection occurring sometime between the two samples being taken.

6.2.3 Analysis

Incidence of influenza-like-illness, and symptomatic or asymptomatic influenza infection

Paired samples have been tested for antibodies to recently circulating strains of seasonal influenza by standard HI assay. A person with a fourfold or greater rise in antibody titer or a positive PCR result was considered to have laboratory evidence of influenza infection during the study period. Persons with laboratory evidence of influenza infection and who reported influenza-like-illness (fever and/or cough and/or malaise and/or myalgia) were considered to be symptomatic influenza cases. Persons with laboratory evidence of influenza infection but not reporting influenza-like-illness were considered to be asymptomatic influenza cases. Estimates of incidence rate with 95% confidence intervals for influenza infection were calculated. Subjects who were lost to follow up for any reason (e.g. migration, voluntary withdrawal or death) would be censored and the period of active observation in days was calculated.

6.2.4 Pandemic H1N1 sub study

The Ha Nam cohort has provided valuable information on the burden of seasonal and pandemic influenza burden in Viet Nam (Horby et al., 2012, 2011, 2010). In 2009 the novel influenza subtype A/H1N1/2009 began to circulate globally and seemed likely to displace other influenza sub-types. In the early stages of the circulation of H1N1/2009 there were no robust population based data on the infection rates and the prevalence and clinical relevance of cross-protective immune responses to H1N1/2009. Since all subjects in the cohort had provided written informed consent to supply information, swabs and blood samples for the purposes of studying the transmission of influenza and the immune responses to infection, we had an ideal setting for studying the epidemiology of H1N1/2009. We therefore rapidly developed a study to be conducted within the cohort which is described in detail in Chapter 7. In summary, the primary objective was to estimate the household and community rate of clinical and sub-clinical H1N1/2009 infection and the profile of viral shedding, and to explore how this is influenced by pre-existing humoral and cellular immunity to seasonal influenza. Secondary objectives were 1) To estimate the incubation

period, serial interval, and duration of viral shedding; 2) To track genetic evolution of the virus within individuals and along chains of transmission; 3) To look for evidence of an effect of oseltamivir treatment of index cases on the risk of secondary household cases.

6.3 Candidate's role

I am the principal coordinator of the Ha Nam cohort, and contributed to write the original protocol and case record forms, and prepared paperwork for ethical approval in the UK and Viet Nam. I directly performed or supervised most aspects of the study implementation in Viet Nam, including field staff training, the preparation of Standard Operating Procedures, data management, and finances. I wrote the protocol for the enhanced H1N1/2009 study and performed or supervised all aspects of the study implementation except for the laboratory work. Laboratory assays were conducted by trained laboratory personnel under the supervision of Dr. Annette Fox.

Chapter 7

Pandemic H1N1/2009 transmission and shedding dynamics in households

7.1 Introduction

The infectiousness of influenza cases will depend on how much virus is shed, for how long, and the degree to which symptoms are required for virus to transmitted. The amount of transmission will also depend on contact susceptibility, the frequency and nature of contact between infected and susceptible persons, and infection prevention practices (Donnelly et al., 2011; Horby et al., 2011; Mathews et al., 2007). Quantification of these parameters is needed to develop interventions that control transmission. In particular, the impact of interventions that rely on case finding, such as quarantine and provision of masks and antivirals to contacts, will depend on how much shedding and transmission occur in the absence of symptoms. Other factors such as the duration of shedding in relation to the duration of symptoms inform the duration of intervention required (Donnelly et al., 2011).

Households are important sites of influenza transmission (Ferguson et al., 2006), and provide valuable information about virus transmission and shedding dynamics because contacts of index cases can often be observed before virus shedding and symptoms start. The H1N1/2009 pandemic enabled investigations of transmission when pre-existing immunity was considered to be relatively low. Numerous case ascertainment design studies were conducted whereby households are investigated following passive detection of cases presenting to health care centres (Carcione et al., 2011; Cauchemez et al., 2009; France

et al., 2010; Komiya et al., 2010; Looker et al., 2010; Loustalot et al., 2011; Morgan et al., 2010; Papenburg et al., 2010; Sikora et al., 2010) some of which required laboratory confirmation of secondary infection (Chang et al., 2011; Cowling et al., 2010; Lau et al., 2010; Pebody et al., 2011; Simmerman et al., 2011; Suess et al., 2010, 2012). Estimates of household secondary attack rate (SAR) or secondary infection risk (SIR) ranged from 3 to 38% for twelve studies that collected respiratory specimens (Lau et al., 2012). The factors with the greatest influence on SIR included whether samples were collected to identify asymptomatic infection; whether cases were detected via health systems or during outbreak investigation; and the proportion of index cases that were children. In all but a few studies (Cowling et al., 2010; Lau et al., 2010; Papenburg et al., 2010) some contacts used antiviral prophylaxis, which affects SIR (Calatayud et al., 2010; Carcione et al., 2011; France et al., 2010; Komiya et al., 2010; Pebody et al., 2011; Suess et al., 2010). Few active case finding studies were conducted and these were in school or school camp populations during outbreaks (Calatayud et al., 2010; Loustalot et al., 2011; Sugimoto et al., 2011) and either retrospective citepLoustalot2011, Sugimoto2011 or affected by school closure and prophylaxis citepCalatayud2010. One household cohort study has been reported but used paired pre- and post-season serology to detect infections (Klick et al., 2011).

The current study uses a cohort of initially uninfected households with active case finding. This is considered to be the gold standard design for influenza household studies and should provide a relatively representative and unbiased description of transmission and shedding dynamics (Klick et al., 2012). The participants in this study had been enrolled in the cohort since December 2007 and most had blood samples collected and tested by serology just prior to the pandemic such that prior immune status and susceptibility could be confirmed.

7.2 Material and Method

7.2.1 ILI surveillance and sample collection schedule

The investigations described here were conducted as part of an on-going household-based influenza cohort study that has been described in detail in chapter 6 and in previous

publications (Horby et al., 2012). In advance of the arrival of H1N1/2009 in Viet Nam, systems were established for enhanced detection of H1N1/2009 cases in all cohort members. The standard cohort procedure is to conduct weekly active surveillance for ILI in each household and to take a nose and throat swab from the person with an ILI. All swab samples from the cohort are batched and tested monthly. For the H1N1/2009 sub-study, weekly active surveillance for ILI continued but all household members (not only the person reporting an ILI) were asked to provide a nose and throat swab. The samples were sent the same day (day 1) to NIHE for immediately testing by RT-PCR to provide results within 48 hours. This process was needed in order to implement serial sampling of the index cases from the earliest possible moment and to detect incident cases of infection in household members. It was decided that rapid tests lacked the required sensitivity and therefore RT-PCR was used.

All household members were swabbed daily according to the schedule in figure 7.1. Samples collected on days 1, 5, 10, and 15 were tested to determine the need to continue swabbing and health surveillance. Swabbing was continued for a period of two days beyond each scheduled testing day in order to allow for the lag in return of results. Interim specimens (e.g. days 2, 3, 4, 6, etc.) were stored and tested later as necessary to determine the profile of viral shedding and for the detection of quasi species (a sub population with similar mutations). Health workers examined all persons in suspected and confirmed H1N1/2009 case households, including those without symptoms, each day for up to 15 days during the first pandemic wave (September December 2009). Examinations included collection of nose- and throat- swabs for quantitative RT-PCR and full-genome sequencing; mouth temperature measurement, scored on a 5-tier scale (36-36.9 = 1, 37-37.9 = 2, 38-38.9 = 3, 39-39.9 = 4, $\geq$40 = 5); and evaluation of symptoms (sore throat, nasal congestion, runny nose, sneezing, dry cough, wet cough, headache, diarrhoea, myalgia, fever, and wheeze), which were scored on a 3-tier scale (none = 0, mild = 1, or moderate/ severe = 2). A cough was defined as wet or productive if sputum or material from the bronchi was expectorated. Participants were also asked if they took the day off work because of illness or to care for another household member that was ill, and if they took oseltamivir.

ILI cases were asked to provide a 10 ml blood sample in a heparinized tube on day

12 after illness onset for extraction of peripheral blood mono-nuclear cells (PBMC's). All other household members were asked to provide a 10 ml blood sample in a heparinized tube on day 15 after illness onset in the index case for extraction of PBMC's. These samples were collected in order to document the acute cellular immune responses and to compare the response in symptomatic cases to asymptomatic or sub-clinical infections. The overall aim being to identify cross-reactive cellular immune responses that are associated with attenuation of clinical illness. Blood samples were collected for serology in June 2009 and April 2010.

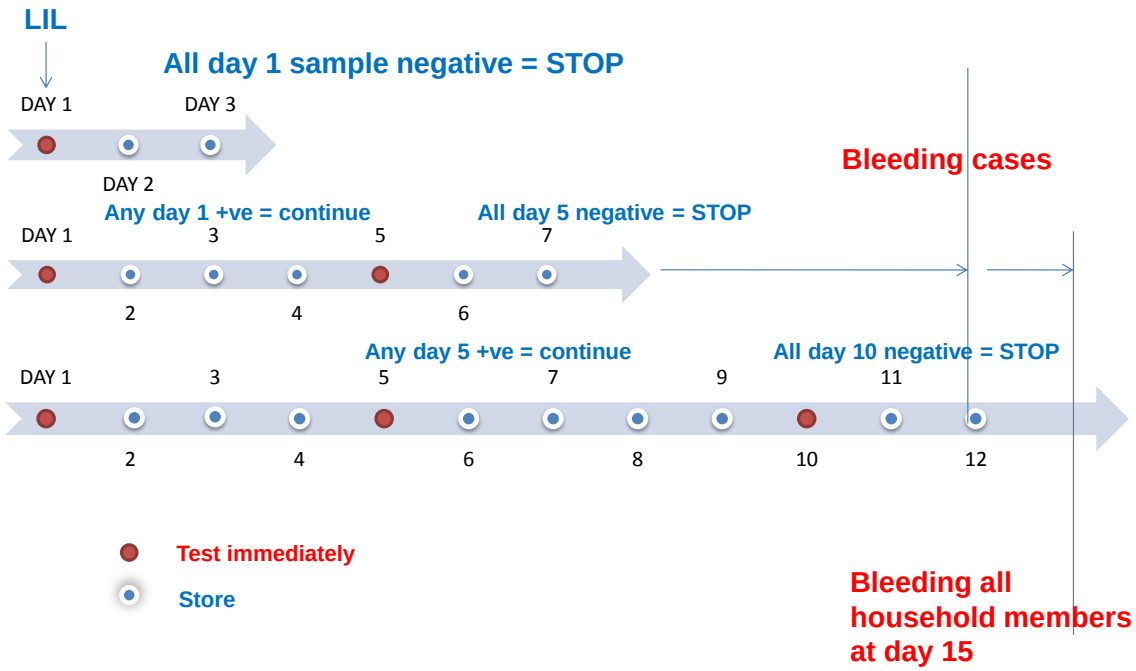


Figure 7.1: Swabbling schedule.
Methods for enhanced detection of H1N1/2009 in all cohort members.

7.2.2 Swab Collection

Separate flocked swabs (Copan, 25125 Brescia, Italy) were used to firmly swab the entire posterior pharynx and tonsillar area and the nasal area at the level of the turbinates. Nasal and throat swabs were combined in 1 tube containing 3 ml or viral transport media, placed on ice and transferred to the laboratory within 24 hours where they were vortexed

before aliquoting and storing the media at -80^{o}C.

7.2.3 Virology and serology

RNA was extracted from swab media using viral RNA extraction kits (Qiagen) and first assessed by real-time reverse-transcriptase polymerase chain reaction (RT-PCR), according to WHO/USCDC protocols (CDC reference no. I-007-05, Accessed November 30, 2009, at http://www.who.int/csr/resources/publications/swineflu/CDCRealtimeRTPCR_SwineH1Assay-2009_20090430.pdf). Daily swabs from participants with confirmed infection were then assessed in a quantitative RT-PCR assay, using primers that amplify a 95 bp section of the M gene: forward 3'-GACAAGACCAATCCTGTCACCTCTG-5', reverse 3'AAGCGTCTACGCTGCAGTCC-5', probe bp 190 5'TTCACGCTCACCGTGCCCAGTGAGC3' (de Jong et al., 2005). The target sequence was cloned and quantified using pico green to prepare a standard curve. Results were expressed as cDNA equivalent copies of viral RNA as determined by comparison to standard curve. The limit of detection was 5 RNA copies/reaction. De novo sequencing using total RNAs extracted from nose and throat swabs with cDNA equivalent influenza RNA copies >30000/ml (˜Ct <33) was performed using 454 and Illumina (Illumina, San Diego, CA, USA) platforms (Gene Bank Accession numbers for consensus sequences will be made available).

Sera were tested in haemagglutination inhibition (HI) assay as previously described (Horby et al., 2012). A reference antigen supplied by WHO (A/California/7/2009(H1N1)-like) was used with turkey erythrocytes. Titers were read as the reciprocal of the highest serum dilution causing complete inhibition of agglutination, partial agglutination was not scored as inhibition of agglutination. If there was no inhibition of HI at the highest serum concentration (1:10 dilution) the titer was designated as 5.

7.2.4 Definitions and Analysis

Infection was defined as a positive influenza RT-PCR result on a nose and/or throat swab regardless of the presence of symptoms. As not being routinely performed on acute sera, serology was not considered in the definition of secondary infection. Nevertheless, seroconversion was reported if there was a 4-fold or greater rise in HI titer between pre-

and post-pandemic sera. Household secondary infection risk (SIR) was calculated as the number of household contacts becoming a case within 1-8 days of symptom onset in the index case divided by the number of household contacts. Serial interval was defined as the number of days between symptom onset in the index case and the first secondary case. Other secondary household cases were only included in the calculation of the serial interval if their symptom onset occurred on the same day as the first secondary case. Children were defined as those up to 15 years of age.

Continuous variables are presented as median and interquartile range and compared using Rank sum test. Chi-squared or Fisher's exact test were used for proportions. All statistical tests were 2 sided, and probability less than 0.05 was considered significant. Univariate and multivariate logistic regression was performed to determine factors associated with H1N1-2009 infection among contacts. Generalised estimating equations (GEE) were used to account for household clustering in the logistic regression model. Predictor variables included the age, sex and position in the house/family (mother, daughter, son, father, other) of the contact and of the index case, number of people in the house and index case viral load, symptom scores and antiviral treatment. Variables with a univariate P value <0.10 were included in multivariate analysis. The Box-Tidwell test was used to assess the assumption of linearity (Cauchemez et al., 2009; Papenburg et al., 2010).

7.3 Results

7.3.1 Index case house characteristics

Index cases were detected in 20 (7.4%) of 270 households (Table 7.1). Two households had two separate index case episodes resulting in 22 index cases. The second episode in each of these two households was excluded from the analysis of transmission. The households contained 81 people including the 22 index cases with the remaining 59 classified as contacts. Households comprising four people were significantly more common than amongst all 270 cohort households (p = 0.009). Accordingly, most index case households comprised nuclear families with similar numbers of mothers, sons and daughters whereas some households lacked fathers. 25% of sons and daughters were older than 15 years. The

median age of people in index case households was 23.3 years (IQR 12.2-39.3) with significantly fewer in the youngest and oldest age categories compared to all 270 households in the cohort. Pre-pandemic blood was collected from 69 (85%) of the index case household members (Table S1). HI titres against A/H1N1/2009 like virus were <10 in all but one individual, who had a titre of 1:20 and was not infected. None reported having received influenza vaccination in the past.

Table 7.1: Composition of households in the cohort and those with an index case.

		All houses n (%)	Index houses n (%)	p value
Houses		270	20	-
People		940	81	-
People per house	1	28 (10.4)	0 (0)	-
	2	41 (15.2)	1 (5)	0.327
	3	65 (24.1)	4 (20)	0.792
	4	74 (27.4)	11 (55)	0.009
	5	42 (15.6)	3 (15)	1
	≥6	20 (7.4)	1 (5)	1
Females		508 (54.5)	42 (51.9)	0.704
Position in the household /family	Mother	250 (26.6)	20 (24.7)	0.756
	Father	207 (22.0)	15 (18.5)	0.496
	Daughter	204 (21.7)	20 (24.7)	0.494
	Son	183 (19.5)	22 (27.2)	0.085
	Other	83 (8.8)	3 (3.7)	0.116
	Unknown	14 (1.5)	1 (1.2)	1
Age	0-4	83 (8.9)	2 (2.5)	0.049
	5-9	70 (7.5)	10 (12.3)	0.107
	10-19	209 (22.5)	25 (30.9)	0.066
	20-39	246 (26.5)	25 (30.9)	0.323
	40-59	241 (25.9)	17 (21.0)	0.386
	≥60	80 (8.6)	1 (1.2)	0.021
	Unknown		1 (1.2)	

7.3.2 Secondary cases

Eleven of 59 contacts were infected, giving a household secondary infection risk (SIR) of 18.6% (95%CI 10.7-30.4%). The secondary cases were from eight (40%) of the index case households. Five households had one secondary case, three households had two and

twelve households had none. Six of the secondary cases were symptomatic giving a household secondary confirmed influenza illness risk of 10.2% (95%CI 4.8-20.5%). Five were asymptomatic, representing 45% of secondary infections. Four asymptomatically infected contacts also had blood collected for serology, of which three seroconverted (Table S1). The asymptomatic case that did not convert was an adult who had a 2-fold rise in titre, and viral RNA detected in swabs on 5 consecutive days. Her two children had virologically confirmed infection and both seroconverted but one was also asymptomatic. Six additional seroconverters were detected among 48 household members whose swabs remained negative during the period of the household transmission study. None of these six seroconverters reported ILI. In total, 69 people from index case households were assessed by serology as well as RT-PCR on swabs. Of these, 39 (56%) had virologically confirmed infection and/or seroconversion during the first pandemic wave (Table S1). Viral sequencing demonstrated that the genetic distance between haemagglutinin and neuraminidase genes of viruses from the same household was around 3-4 times less than between viruses from different households (Table 7.2). Analysis of virus genes indicated that 10 of 11 secondary cases were infected within the household giving an adjusted household SIR of 17.2% (95%CI 9.6-28.9%). One infected household contact, who was the index case's husband, was suspected to have acquired infection in the community because the genetic distance between his virus and the index case's virus (0.002969) was similar to that found between households. Virus from his swabs was more closely related to viruses from another household in the same village.

Table 7.2: Comparison of H1N1/2009 envelope gene sequence diversity within households and individuals and between households.

	Mean p-distance[a] (standard deviation)	
	Haemagglutinin	**Neuraminidase**
Within an individual	0.00007215 (0.000161)	0.00004304 (0.000143)
Within a household[b]	0.000509 (0.001107)	0.000608 (0.001322)
Between households[b]	0.002262 (0.001140)	0.002280 (0.000908)

[a] p-distance is the number of nucleotide substitutions divided by the number of nucleotides calculated using Mega version 5.2. p-distance values were similar to d-distance values, which correct for unmeasured nucleotide changes using the nucleotide substitution

Kimura-2-parameter model.

[b] Only the first time point of each infected participant was used.

Demographic data for index and secondary cases are compared in Table 7.3. Fourteen (64%) of 22 index cases were females and a higher proportion of females than males were index cases. Only one index case was a father whereas around one third each were mothers, daughter or sons. A high proportion of child daughters were index cases (54.5%). Secondary cases comprised fairly even numbers of males and females, and the proportion of male and female contacts with secondary infections was very similar. Similar to index cases, none of the fathers was a secondary case, and the proportion of fathers that was a case was significantly lower than for mothers, daughters and sons. Roughly half of both index and secondary cases were adults although the proportion of children that were cases was high compared to adults. The median age of index (14.9 years, IQR 9.7-36.7) and secondary cases (16.9 years, IQR 9.6-34.6) was lower than for non-infected household members (34.7 years, IQR 13.8-42.5).

Table 7.3: Distribution of cases, contacts and secondary cases by age, gender and position in the family.

	All house members			Contacts
	N	***Any case***[a] **n (%)**	***Index case***[a] **n (%)**	**Secondary case n/N (%)**
Child	30	16 (53.3)	11 (36.7)	5/19 (26.3)
Adult	50	17 (34.0)	11 (22.0)	6/39 (15.4)[b]
Female	42	19 (45.2)	14 (33.3)	5/28 (17.9)
Male	39	14 (35.9)	8 (20.5)	6/31 (19.3)[b]
Mother	20	9 (45.0)	6 (30.0)	3/14 (21.4)
Father	15	1 (6.7)[c]	1 (6.7)	0/14 (0)
Child daughter	11	7 (63.6)	6 (54.5)	1/5 (20.0)
Adult daughter	9	3 (33.3)	2 (22.2)	1/7 (14.3)
Child son	18	9 (50.0)	5 (27.8)	4/13 (30.8)
Adult son	4	3 (75.0)	2 (50.0)	1/2 (50.0)
Other	3	1 (33.3)	0 (0.0)	1/3 (33.3)[b]

[a]The denominator is the number of household members in each category; demographic data was incomplete for 1 household member.

[b]HA and NA gene sequences indicate that one case may have been infected in the community, who was an adult male whose position in the family is other.

[c]The proportion of fathers with virologically-confirmed infection was significantly lower ($X^2 p = 0.021$) compared to mothers (OR 11.45, 95% CI 1.25–104.60), daughters (OR 14.00. 95% CI 1.54–127.62) and sons (OR 16.80, 95% CI 1.87–150.94).

7.3.3 Virus RNA shedding and symptom dynamics

The median serial interval for symptomatic secondary cases was 2 days and ranged from 1 to 3 days (Figure.7.2A, Table 7.4). In households with only asymptomatic secondary cases, viral RNA shedding was detected 1–5 days after symptom onset in the index case (Table 7.4, Figure.7.2A). In 8 secondary cases the first day of viral shedding could be determined absolutely because swabs from preceding days were negative (Figure.7.2A), and in three of the six with symptoms shedding commenced the day before symptoms (Figure.7.2B). The vast majority of cases tested on day 0 through 2 after onset shed viral RNA (Figure.7.2B). Thereafter the proportion that shed virus RNA, and levels shed,

declined. The Kaplan–Meier estimate for median time until viral RNA was undetectable was 7 days (IQR 6-14 days, Figure. S1), and amongst 27 cases in whom the last shedding day could be observed the median viral RNA shedding time was 6 days with no clear difference in shedding times between symptomatic and asymptomatic cases (Table7.4, Figure.7.2A & C). However, both peak and day 2 viral loads were higher in symptomatic compared to asymptomatic cases. In most symptomatic cases viral RNA shedding peaked at around the time that symptoms scores peaked on day 1 and 2 after onset (Figure.7.2B, C & D). Amongst cases that had symptoms there were no clear differences in virus shedding or symptom score between adults and children (Figure.7.3E & F), or between index and secondary cases (Figure.7.2 C & I). However, three secondary cases had only a modest elevation of mouth temperature while the other three had mouth temperatures above 38 °C and classic ILI. None of the symptomatic cases required hospitalization.

Table 7.4: Virus shedding and transmission characteristics.

	Index (n = 18)	**Secondary (n = 6)**	**Asymptomatic (n = 5)**
Serial Interval	NA	1, 1, 2, 2, 3, 3	1, 1, 1, 5
Shedding Days[a]	6.0 (4.0-7.0)	6.5 (6.0-8.8)	6.0 (4.0-7.0)
Peak Log 10 Viral Load	7.0 (6.6-7.4)	7.2 (6.6-7.6)	6.1 (5.0-7.3)
Day 2 Log 10 Viral Load	5.6 (4.6-6.4)	6.4 (4.8-6.6)	4.7 (3.3-5.1)$^{p=0.038}$

Results are presented as median and interquartile range in brackets or as values for individuals.
[a]4 index cases, 1 secondary case and 1 asymptomatic case were excluded because insufficient samples were collected to assess shedding time.

Viet Namese government policy during the first wave of the H1N1/2009 pandemic dictated that all symptomatic cases should be given oral oseltamivir for 5 days. Accordingly 20 cases took oseltamivir for 5 days after symptoms developed, of whom 17 commenced by day 2 after onset (timely) and three commenced 4 days after onset. Participants with asymptomatic infection did not take oseltamivir. Cases that had timely treatment tended to have more severe symptoms and higher viral loads until the day after onset but not thereafter (Figure.7.3G & H). Kaplan–Meier estimates for time until viral RNA shedding ceased were 7 days (IQR 6–7 days) for patients who took timely Oseltamivir and 14 days (IQR 7–14 days) in those who took Oseltamivir late or did not take Oseltamivir ($P < 0.001$,

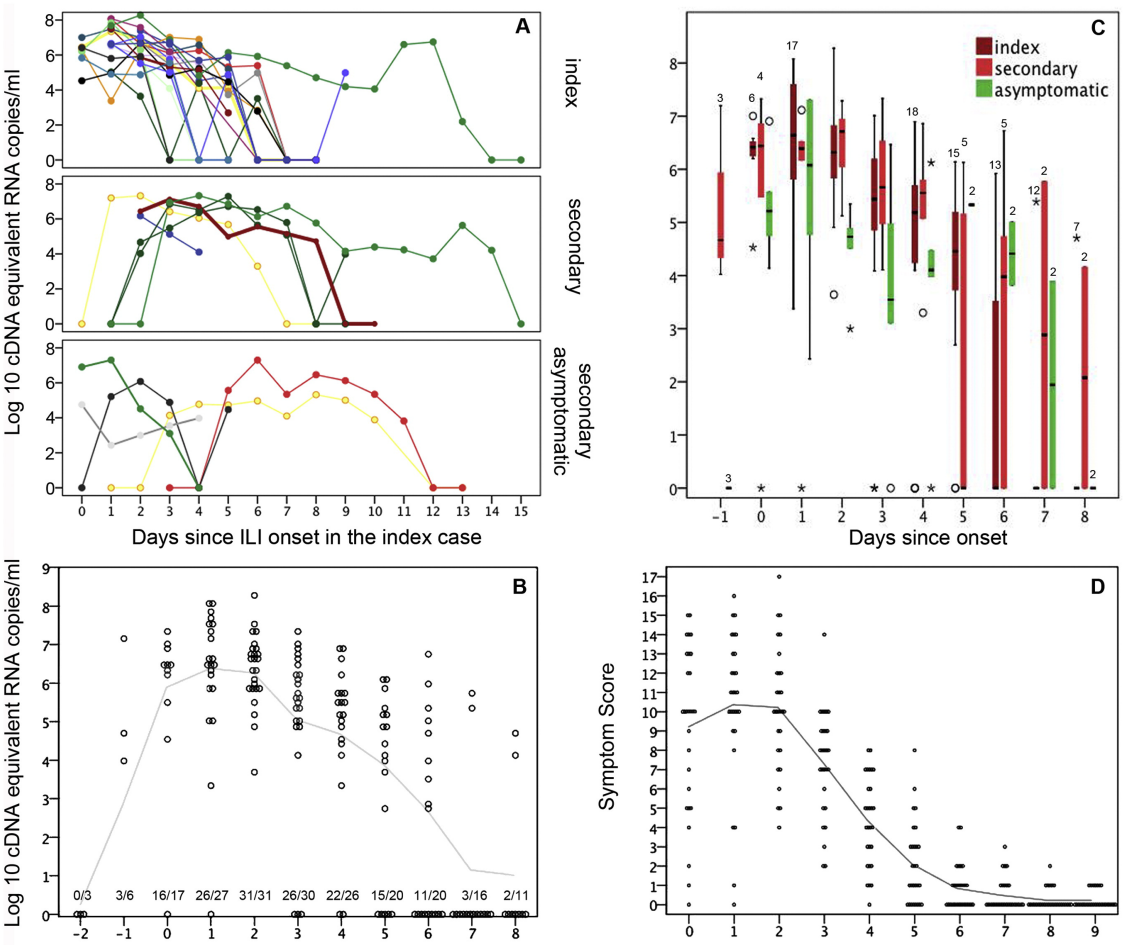


Figure 7.2: Daily viral loads and symptoms in confirmed H1N1/09 cases from index case households.
Panel A shows viral RNA shedding for each individual from index case households with virologically-confirmed infection. Participants from the same household are shown in the same colour and data is shown by day since onset in the index case to indicate the intervals between infections. Panel B shows viral RNA levels by day since onset to demonstrate viral RNA shedding dynamics. Each dot is an individual sample and the line shows the median. Fractions above the x-axis represent the number with detectable viral RNA over the number assessed. Panel C represents daily viral RNA levels for index cases (dark red, $n = 20$), symptomatic secondary cases (red, $n = 6$) and asymptomatic secondary cases (green, $n = 5$). Data is presented as box and whisker plots showing median lines, interquartile ranges (boxes) and ranges (whiskers). All participants in each group were tested except where numbers are shown above each bar. Panels D–I show either viral RNA shedding levels or symptom scores by day of illness for the 28 symptomatic participants. Panel D demonstrates symptom dynamics with dots representing values for individual participants and the line showing the median.

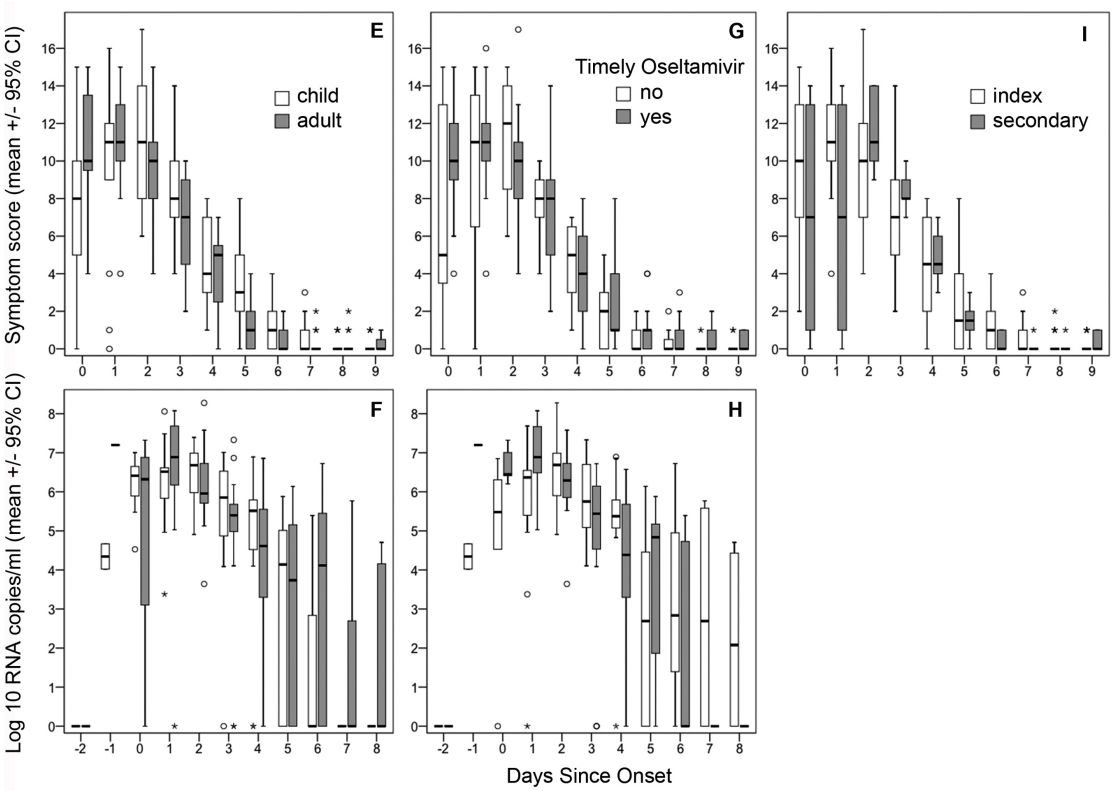


Figure 7.3: Daily viral loads and symptoms in confirmed H1N1/09 cases from index case households (continued).
Panels E and F compare adults and children. Panels G and H compare participants that took Oseltamivir within 48 h of onset versus those who took it later or did not take it. Panel I compares symptoms in index and secondary cases

Figure. S1). Shedding persisted until day 13 after symptom onset in two cases from one household (Figure.7.2A). Both commenced oseltamivir late. These two cases also had the highest wheeze scores, oral temperature was above 38 °C for 5 days, and daily symptom scores were relatively high. Viral sequencing did not reveal any mutations known to be associated with virulence.

7.3.4 Risk factors for secondary infection

Secondary infection of household contacts was associated with index case wet cough score and viral load in univariate analysis, although paradoxically the association with viral load was negative (Table 7.5). Other index case symptoms and index case and contact characteristics were not significant in univariate analysis (Table 7.5), however numbers are small. Although contact age and number of people in the household were not significant in univariate analysis, they were included in multivariate analysis because several other studies demonstrated an association (Carcione et al., 2011; Chang et al., 2011; France et al., 2010). In multivariate analysis (Table 7.6) infection of contacts was positively associated with the index case wet cough score (OR 1.56, 95% CI 1.22–1.99) and negatively associated with number of people in the household (OR 0.20, 95% CI 0.08–0.48). The effect of contact age was small and not significant. The association between index case viral load and contact infection was not maintained in multivariate analysis.

Table 7.5: Univariate analysis of factors associated with transmission of H1N1-2009 from index cases to household contacts during the first pandemic wave

Variable		Contact status			
		Infected (11)	**Not infected (47)**	***OR (CI 95%)***	**p**
Contact	age	16.9 (9.6-34.6)	31.9 (13.9-41.9)	0.96 (0.92-1.01)	0.112
	Female, n/N (%)	5/11 (45)	23/48 (48)	0.91 (0.20-4.05)	0.897
Index	peak Log10 Viral load	6.4 (5.8-7.3)	7.0 (6.7-7.5)	0.33 (0.12-0.86)	0.02
	wet cough score	8 (3-10)	4 (0-7)	1.36 (1.07-1.72)	0.012
	sneeze score	6 (3-7)	6 (3-9)	0.85 (0.67-1.07)	0.226
	running nose	5 (3-8)	3 (1-8)	1.09 (0.89-1.34)	0.443
	dry cough	0 (0-10)	6 (0-9)	0.88 (0.75-1.04)	0.265
	Oseltamivir, n/N (%)	9/11 (82)	36/48 (75)	1.50 (0.14-15.75)	0.735
	age	13.2 (8.3-33.3)	12.4 (8.0-22.0)	1.01 (0.96-1.07)	0.585
	Female, n/N (%)	8/11 (73)	25/48 (52)	2.45 (0.43-13.93)	0.311
House	People/house	4 (3-4)	4 (4-5)	0.46 (0.17 - 1.29)	0.14
	Child/house	2 (1-2)	2 (1-3)	0.62 (0.31-1.23)	0.168

Table 7.6: Risk factors for transmission of H1N1-2009 from index case to household contacts during the first pandemic wave.

Variable	Contact status[a]		**OR (CI 95%)**	***p***	**Adjusted OR (CI 95%)**	***p***
	Infected (n = 11)	**Not infected (n = 47)**				
Contact age	16.9 (9.634.6)	31.9 (13.941.9)	0.96 (0.921.01)	0.112	0.94 (0.881.01)	0.115
Index Peak Log 10 Viral load[b]	6.4 (5.87.3)	7.0 (6.77.5)	0.33 (0.120.86)	0.02	0.56 (0.142.23)	0.409
Index wet cough[c]	8 (310)	4 (07)	1.36 (1.071.72)	0.012	1.56 (1.221.99)	<0.001
People/house	4 (34)	4 (45)	0.46 (0.171.29)	0.14	0.20 (0.080.48)	<0.001

[a]Results are presented as median and interquartile range.
[b]Maximum Log 10 cDNA equivalent viral RNA copies/ml detected for each index case.
[c]Summed score for wet cough over the course of illness in the index case ranging from 0 for no cough to 2 for moderate to severe cough.

7.4 Discussion

The current study sought to systematically detect H1N1/2009 index cases within a random household cohort and then intensively investigate viral RNA shedding and symptoms in household members to obtain unbiased estimates of transmission. The vast majority of household members appeared to be susceptible to infection based on pre-pandemic H1N1/2009 HI and MN titres. Eleven household contacts were infected, but 5 (45%) did not develop symptoms. Virus genetic sequencing indicated that 10 (91%) were probably infected within the household rather than from the community, enabling a more precise estimate of SIR. The majority of transmission involved mothers and children with a serial interval of around 2 days. The study was not powered to identify small effects on transmission but wet cough in the index case was found to have a significant effect. Studies such as this are also essential to provide precise estimations of incubation period, duration of virus shedding and relation of shedding to symptoms.

In the current study index and secondary cases were similar in terms of age, virus RNA shedding and symptoms. In contrast, studies using case ascertainment designs report a tendency for more severe symptoms and higher viral shedding for index cases (Cowling et al., 2010; Suess et al., 2010) a bias that could lead to over-inflated SIR estimates. Factors other than severity can also influence health care seeking, leading to bias in case ascertainment studies. Surveys conducted in France and England during the H1N1/2009 pandemic found that the proportion of self-defined ILI cases that sought care was highest for children and males aged below 25 years (Brooks-Pollock et al., 2011; Van Cauteren et al., 2012).

The cohort study design used here facilitated confirmation of susceptibility to infection by serology on pre-pandemic sera. Nevertheless, some index case household members may have had asymptomatic or mild infection before the index case was detected because they seroconverted without ILI or detection of virologically confirmed infection during investigation of the index case episode. This scenario would mean that fewer were susceptible. Virus genetic sequencing enabled discrimination of household from community transmission and we demonstrated that one index case household member was infected in

the community rather than in the household. The within and between household genetic diversity is in agreement with other studies (Gubareva et al., 2002; Pascalis et al., 2012; Poon et al., 2011; Teo et al., 2007) and the magnitude of sequence diversity within individuals, households and between households was consistent with the study of Poon et al Poon et al. (2011). Pascalis et al found evidence of changes in quasi-species dominance within individuals (Pascalis et al., 2012), and we will perform further analysis of deep sequences to describe quasi-species in future. The results demonstrate that intensive investigations involving serology, virology and phylogenetics are required to obtain an accurate estimate of transmission.

A notable feature of the current study was the predominance of females amongst index cases, whereas most other H1N1/2009 transmission studies found that roughly half of index cases were females. In relation, the number and proportion of fathers infected was significantly lower than for mothers and children. Similarly, a study that assessed household contacts of children identified by active case finding during a school camp outbreak found significantly lower infection amongst fathers (France et al., 2010). These findings are also reminiscent of cohort and other studies from the 1950s (Badger et al., 1953; Buck, 1956) suggesting that the pattern of transmission between mothers and children, with sparing of fathers may be a common phenomenon. Fathers in our study did not appear to be less susceptible on the basis of serology implying that they may have less exposure to infection, either via less contact with cases and/or more effective prevention of infection upon exposure. During a survey in 2007, 43% of fathers in the cohort said they cared for children compared to 55% for mothers (Horby et al., 2011). This small difference is unlikely to account for the difference in proportion infected, but may not reflect care patterns for sick children. During the school camp outbreak study described above, 66% of the household contacts that cared for index cases were mothers, 24% were fathers and 3% were siblings (France et al., 2010).

A high proportion of child daughters were index cases. It is generally considered that children are the main influenza transmitters because they have more contacts outside the house, are more susceptible to infection and severity, and shed more virus (Viboud et al., 2006b). We did not detect significant differences in virus RNA shedding or symptom

scores between children and adults, similar to other studies (Loeb et al., 2012; Suess et al., 2012). A systematic review also concluded that shedding duration of influenza H1N1/2009 was no longer among children compared with adults, either between or within studies (Fielding et al., 2014). Perhaps susceptibility to novel virus is more uniform in accordance with the uniform absence of HI antibodies. It should also be noted that viral RNA shedding may not reveal differences in shedding of viable virus, which is relatively shorter in duration (Suess et al., 2012). Contact patterns could influence who is infected as an index or household secondary case. A previous study of contact patterns for this cohort demonstrated that children have the highest numbers of close contacts, both with peers and parents (Horby et al., 2011), but did not differentiate by gender or position in the family. Further verification of contact patterns for different family members, particularly mothers versus fathers, is planned.

Virus RNA shedding dynamics correlated with symptom scores and were generally consistent with reports elsewhere (Cowling et al., 2010; Lau et al., 2010; Suess et al., 2010, 2012). The duration of viral RNA shedding was within the 3–9 day range reported by other studies of cases in the community (Fielding et al., 2014). The serial interval was slightly shorter than in other studies but was based on a small number of secondary cases while tertiary cases were excluded. As noted by Lau et al., serial interval estimates could be shortened by correction for multiple chains of transmission (e.g., tertiary cases), and serial interval estimates are not constant because they reflect a combination of the profile of index cases, contact patterns within households, and incubation period (Lau et al., 2012).

Timely oseltamivir treatment of index cases was not significantly associated with infection of contacts, as reported elsewhere (Carcione et al., 2011). However, cases that took oseltamivir early tended to have higher viral RNA shedding and symptom scores at onset compared to untreated or late-treated cases, whereas levels were similar or lower by day 2. Therefore, timely treatment may have helped to resolve shedding and symptoms.

Forty five percent of virologically confirmed household secondary cases did not develop symptoms, higher than reported by others (Lau et al., 2010; Loeb et al., 2012; Papenburg et al., 2010; Simmerman et al., 2011; Suess et al., 2012). One asymptomatic case did

not seroconvert, which may indicate that viral RNA remained in the respiratory tract without being internalized and eliciting an immune response. Contrary to expectations, the duration of viral RNA shedding was similar for symptomatic cases and asymptomatic cases, perhaps because asymptomatic cases did not take oseltamivir. In contrast Loeb et al. reported a shorter duration of shedding in asymptomatic cases (Loeb et al., 2012).

The extent to which shedding without symptoms contributes to influenza transmission is unclear (Patrozou and Mermel, 2009). A few studies have investigated transmission during pre-symptomatic shedding in humans, but involve only a few index cases, rely on recall, and can't control for exposure (Gu et al., 2011; Hermes et al., 2011). One study has demonstrated transmission before symptoms in ferrets (Roberts et al., 2012). Virus emission is an important component of transmission and is related to both nasopharangeal viral load and the mechanical processes of coughing and sneezing (Bischoff et al., 2013). In the current study viral RNA shedding was lower in asymptomatic compared to symptomatic cases, consistent with Loeb et al. (2012), but in contrast to Suess et al. (2012). Household transmission was also associated with the amount of wet cough in the index case, consistent with several other studies (Carcione et al., 2011; Chang et al., 2011; Looker et al., 2010) and suggesting that transmission from symptomatic cases is more efficient. However, virus emission has been reported to vary substantially between individuals (Bischoff et al., 2013) and this could confound our interpretation of risk factors. Further definition of the contribution of shedding without or before symptoms to transmission is required to estimate the effectiveness of control measures such as case quarantine and timely treatment.

The major limitations of the current study were the small number of index cases, and the selection of households from just one commune. Although nearly 1000 people were included in the cohort, the number of index cases could not be controlled and was not sufficient to robustly assess risk factors for transmission, particularly factors with a lot of variance such as viral load. Households were selected from one commune because we lacked sufficient resources to maintain intensive surveillance in multiple sites, representative of the population. Nevertheless, the commune was representative of a large proportion of the population that reside within the semi-rural deltas. Studies are underway to investigate

urban versus rural differences in transmission and contact patterns.

This cohort study avoided many of the limitations of other studies of A/H1N1/2009 transmission in households including case ascertainment bias, assumptions about immunity/susceptibility and transmission within the household, and failure to detect asymptomatic infection (Klick et al., 2012; Lau et al., 2012). Cohort studies are resource and labour intensive but can provide more reliable estimates of SIR. The intensive assessment of shedding and symptoms demonstrated that a substantial amount of shedding occurs without symptoms but wet cough in the index case was associated with significantly increased transmission.

7.5 Conclusion:

In this cohort of H1N1/2009 susceptible persons, virus sequencing was capable of discriminating household from community transmission. Household transmission involved mothers and children but rarely fathers. Asymptomatic or pre-symptomatic shedding was common.

Chapter 8

General discussion and concluding remarks

8.1 Contribution to knowledge on seasonal and pandemic influenza

In general, the seasonal characteristics of influenza transmission in tropical and sub-tropical areas is not well defined, and the seasonality of influenza in Viet Nam has not previously been characterized. As discussed in Chapter 1, a good understanding of the temporal and geographic patterns of influenza transmission is needed for the planning of influenza immunisation programmes. As discussed in chapter 3, Viet Nam does not yet recommend routine influenza vaccination in the national influenza control guidelines (Ministry of Health, 2009). This is not because Viet Nam does not have access to vaccine, but because the MoH does not yet have an evidence base to set the schedule. In addition, studies of the patterns of influenza transmission can provide new insights into the environmental conditions that are favourable to effective transmission, thereby providing an evidence base for prediction of the timing of epidemics and for prevention and control measures. Whilst data on patterns and determinants of influenza transmission in tropical and sub-tropical areas of south-east Asia are accumulating, they are still limited. The work described in this thesis is therefore a significant contribution in this area.

Firstly, I have taken a large set of routine ILI notification data and through careful filtering of background 'noise', shown patterns that were not initially apparent. Prior to this analysis the routine ILI surveillance data were not well respected. Even people working on the routine system did not have confidence in the value of the data. In chapter 3, I have shown that important patterns do exist in the data, and for the first time it has been

demonstrated that ILI is seasonal in the north of Viet Nam but this seasonality disappears towards the south. In Chapter 5, I have explored the association between ILI notifications and influenza virus activity. Characterising this relationship is critical if ILI notification data are to be used to inform specific programmes to control respiratory pathogens such as influenza, but also other important respiratory pathogens such as respiratory syncytial virus. The work presented in chapter 5 is the first attempt to assess the relationship between ILI notifications and influenza virus activity in Viet Nam, and although the virological surveillance data are of limited extent, completeness and quality, and have been complicated by the arrival of the pandemic in the middle of the surveillance period, I was able to assess the relationship. I found evidence of synchrony between ILI notifications and influenza virus activity, with a lag on about two weeks. However influenza viruses can be detected throughout the year and the relationship between ILI and influenza activity is undoubtedly complicated by other respiratory pathogens. Identifying clear seasonality of ILI in north Viet Nam, and an association between ILI and influenza virus activity throughout the country, is an important step towards utilising these surveillance sources to inform public policy on immunisation. The work has also shown both the value and the limitations of the surveillance data, and I believe that through revealing some weaknesses of the systems, this work will result in strengthening of the surveillance system in Viet Nam. Specifically, the collection of age data in the ILI notifications would aid interpretation and whilst the number of sentinel sites for virological surveillance may be reduced due to budget limitations it is vital that the quality of data is improved. Furthermore, the results from chapter 5 also point out that future research must include the study of other respiratory viruses in Viet Nam, to understand the varying components that make up the ILI time series signal presented in chapter 3 and 4. In addition, the methods developed and described in this thesis will be applied in the future to surveillance data, contributing to quicker and improved analysis and interpretation of the data.

In chapter 4, I was able to assess the association between the seasonality of influenza and the seasonality of a range of climate variables. Absolute humidity and temperature were the variables that explained most of the variance in climate data, and in tree regression analysis, the seasonality of AH was the variable that best explained the seasonality of

ILI notifications. Whilst this finding is in line with other work demonstrating the importance of AH, it is the first time that AH has been identified as an important variable in a non-temperate region. Further work is needed to explain differences in our results compared to others, who found the timing of influenza epidemics to be associated with declines in AH, whereas we found that ILI peaks are associated with AH peaks (Bloom-Feshbach et al., 2013; Tamerius et al., 2010, 2013). The results from chapter 3 and 4 have been presented in several international workshops, including Options for the Control of Influenza VIII (http://optionsviii.controlinfluenza.com/), and the APACI Ha Noi workshop (http://apaci.asia/activities/apaci-meetings/hanoi-vietnam-workshop-2013).

The study presented in chapter 7 is part of a larger program of work arising from the Ha Nam cohort study, for which I have been the field supervisor since it began ((Cauchemez et al., 2012; Horby et al., 2012, 2011; Powell et al., 2012; Thai et al., 2014)). The household transmission study produced several interesting findings. Firstly, the household secondary attack rate was within the range of other studies (Lau et al., 2012) but we found a high proportion (5/11) of cases with virologically confirmed infection but no symptoms. Secondly, genetic analysis suggested that ten of the eleven secondary case acquired infection from within the household, with only one acquiring infection from the community. The high proportion of cases with viral shedding that were asymptomatic, and the high proportion of second cases in the household that are acquired within the home suggests that the isolation of clinical cases, school closure and social distancing measures may not be very effective at preventing onward transmission. However, wet cough in the index case was found to be associated with secondary transmission within the household, suggesting that symptom severity is associated with transmission risk, although this finding does not really improve understanding of the relative contribution of aerosol versus large respiratory droplets transmission since a cough can produce both small and large respiratory droplets Cowling et al. (2013); Killingley and Nguyen-Van-Tam (2013). The data from the cohort have shown that influenza infection rates in Viet Nam are comparable to temperate regions, and perhaps even higher; that mild and asymptomatic influenza infection is common; and that households are an important site for transmission.

In summary, although influenza epidemiology in Viet Nam is complex (and studying

it is challenging due to limitations of the surveillance systems), the work presented here represents significant progress in describing and explaining the epidemiology both at a national level and at the household level.

8.2 Further research directions

Further work is required to reconcile the results of the climate analysis presented in this thesis with the findings of other authors. We are planning to contact Cecile Viboud and other relevant experts to see if a joint analysis of data will be possible, and also to access the global data set published by Tamerius et al. (2013) to attempt to validate our global extrapolation of the Viet Nam results. I also plan to work with others to develop a meta-population mathematical model of influenza transmission in Viet Nam, which will include sub-populations that are subject to different seasonal forcing parameters and a transfer rate of infected individuals between sub-populations, representing travel within Viet Nam. This model will be used to explore the potential impact of various immunisation options whilst accounting for spatial differences in the dynamics of influenza activity in Viet Nam. The Ha Nam cohort runs well, and will be continued in order to provide data on population infection rates and immunological and clinical responses to exposure to new influenza strains. The Ha Nam cohort has provided a bio-archive of respiratory and serum samples which will be used to assess the role of other pathogens in respiratory illnesses in this community. We have also collected a time and age stratified sample of residual serum from four different hospital representing four different regions in Viet Nam. This will be used to further explore the age and season specific incidence of influenza infection in Viet Nam. Activities are also ongoing to improve the quality of the surveillance data. We have developed a web-based system to collect timely ILI data and to automate the analyse of the ILI time-series. All the regional public health institutes have agreed to participate and re-enter all historic data, and all provinces have been trained in data entry. Agreement has also been reached between my institute and Ha Noi medical university to work together on a climate change project, which will provide access to detailed climate data from the Viet Nam Institute of Meteorology, Hydrology and Environment.

References

Adams, B. and McHardy, A. C. (2011). The impact of seasonal and year-round transmission regimes on the evolution of influenza A virus. *Proceedings. Biological sciences / The Royal Society*, 278(1716):2249–56.

Alexander, D. J. (2007). An overview of the epidemiology of avian influenza. *Vaccine*, 25(30):5637–44.

Alonso, W. J., Viboud, C., Simonsen, L., Hirano, E. W., Daufenbach, L. Z., and Miller, M. A. (2007). Seasonality of influenza in Brazil: a traveling wave from the Amazon to the subtropics. *American journal of epidemiology*, 165(12):1434–42.

Altizer, S., Dobson, A., Hosseini, P., Hudson, P., Pascual, M., and Rohani, P. (2006). Seasonality and the dynamics of infectious diseases. *Ecology letters*, 9(4):467–84.

Azziz Baumgartner, E., Dao, C. N., Nasreen, S., Bhuiyan, M. U., Mah-E-Muneer, S., Al Mamun, A., Sharker, M. a. Y., Zaman, R. U., Cheng, P.-Y., Klimov, A. I., Widdowson, M.-A., Uyeki, T. M., Luby, S. P., Mounts, A., and Bresee, J. (2012). Seasonality, timing, and climate drivers of influenza activity worldwide. *The Journal of infectious diseases*, 206(6):838–46.

Badger, G. F., Dingle, J. H., Feller, A. E., Hodges, R. G., Jordan W. S., J., Rammelkamp C. H., J., Jordan, W. S., and Rammelkamp, C. H. (1953). A study of illness in a group of Cleveland families. II. Incidence of the common respiratory diseases. *Am J Hyg*, 58(2):174–178.

Bahl, J., Nelson, M. I., Chan, K. H., Chen, R., Vijaykrishna, D., Halpin, R. a., Stockwell, T. B., Lin, X., Wentworth, D. E., Ghedin, E., Guan, Y., Peiris, J. S. M., Riley, S., Ram-

baut, A., Holmes, E. C., and Smith, G. J. D. (2011). Temporally structured metapopulation dynamics and persistence of influenza A H3N2 virus in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 108(48):19359–64.

Barreca, A. I. and Shimshack, J. P. (2012). Absolute humidity, temperature, and influenza mortality: 30 years of county-level evidence from the United States. *American journal of epidemiology*, 176 Suppl(7):S114–22.

Beckett, C. G., Kosasih, H., Ma'roef, C., Listiyaningsih, E., Elyazar, I. R., Wuryadi, S., Yuwono, D., McArdle, J. L., Corwin, A. L., and Porter, K. R. (2004). Influenza surveillance in Indonesia: 1999-2003. *Clin Infect Dis*, 39(4):443–449.

Bedford, T., Cobey, S., Beerli, P., and Pascual, M. (2010). Global migration dynamics underlie evolution and persistence of human influenza A (H3N2). *PLoS pathogens*, 6(5):e1000918.

Berg JM, Tymoczko JL, S. L. (2002). Section 11.4, Lectins Are Specific Carbohydrate-Binding Proteins. In *Biochemistry. 5th edition*. W H Freeman, New York.

Bischoff, W. E., Swett, K., Leng, I., and Peters, T. R. (2013). Exposure to influenza virus aerosols during routine patient care. *The Journal of infectious diseases*, 207(7):1037–46.

Bloom-Feshbach, K., Alonso, W. J., Charu, V., Tamerius, J., Simonsen, L., Miller, M. a., and Viboud, C. (2013). Latitudinal Variations in Seasonal Activity of Influenza and Respiratory Syncytial Virus (RSV): A Global Comparative Review. *PloS one*, 8(2):e54445.

Boëlle, P.-Y., Ansart, S., Cori, A., and Valleron, A.-J. (2011). Transmission parameters of the A/H1N1 (2009) influenza virus pandemic: a review. *Influenza and other respiratory viruses*, 5(5):306–16.

Bolton, D. (1980). The Computation of Equivalent Potential Temperature. *Monthly Weather Review*, 108(7):1046–1053.

Boni, M. F., Gog, J. R., Andreasen, V., and Feldman, M. W. (2006). Epidemic dynamics and antigenic evolution in a single season of influenza A. *Proc Biol Sci*, 273(1592):1307–1316.

Both, G. W., Sleigh, M. J., Cox, N. J., and Kendal, A. P. (1983). Antigenic drift in

influenza virus H3 hemagglutinin from 1968 to 1980: multiple evolutionary pathways and sequential amino acid changes at key antigenic sites. *J Virol*, 48(1):52–60.

Brankston, G., Gitterman, L., Hirji, Z., Lemieux, C., and Gardam, M. (2007). Transmission of influenza A in human beings. *The Lancet infectious diseases*, 7(4):257–65.

Britten, R. H. (1932). The incidence of epidemic influenza, 1918-19. *Public health reports*, 47(6):303–39.

Brooks-Pollock, E., Tilston, N., Edmunds, W. J., and Eames, K. T. D. (2011). Using an online survey of healthcare-seeking behaviour to estimate the magnitude and severity of the 2009 H1N1v influenza epidemic in England. *BMC infectious diseases*, 11:68.

Broutin, H., Guégan, J.-F., Elguero, E., Simondon, F., and Cazelles, B. (2005). Large-scale comparative analysis of pertussis population dynamics: periodicity, synchrony, and impact of vaccination. *American journal of epidemiology*, 161(12):1159–67.

Buck, C. (1956). Acute upper respiratory infections in families. *Am J Hyg*, 63(1):1–12.

Calatayud, L., Kurkela, S., Neave, P. E., Brock, A., Perkins, S., Zuckerman, M., Sudhanva, M., Bermingham, A., Ellis, J., Pebody, R., Catchpole, M., Heathcock, R., and Maguire, H. (2010). Pandemic (H1N1) 2009 virus outbreak in a school in London, April-May 2009: an observational study. *Epidemiology and infection*, 138(2):183–91.

Call, S. A., Vollenweider, M. A., Hornung, C. A., Simel, D. L., and McKinney, W. P. (2005). Does this patient have influenza? *JAMA : the journal of the American Medical Association*, 293(8):987–97.

Cannell, J. J., Vieth, R., Umhau, J. C., Holick, M. F., Grant, W. B., Madronich, S., Garland, C. F., and Giovannucci, E. (2006). Epidemic influenza and vitamin D. *Epidemiology and infection*, 134(06):1129–40.

Cannell, J. J., Zasloff, M., Garland, C. F., Scragg, R., and Giovannucci, E. (2008). On the epidemiology of influenza. *Virology journal*, 5:29.

Carcione, D., Giele, C. M., Goggin, L. S., Kwan, K. S., Smith, D. W., Dowse, G. K., Mak, D. B., and Effler, P. (2011). Secondary attack rate of pandemic influenza A(H1N1) 2009 in Western Australian households, 29 May-7 August 2009. *Euro surveillance : bulletin*

Européen sur les maladies transmissibles = European communicable disease bulletin, 16(3).

Cauchemez, S., Donnelly, C. A., Reed, C., Ghani, A. C., Fraser, C., Kent, C. K., Finelli, L., and Ferguson, N. M. (2009). Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. *The New England journal of medicine*, 361(27):2619–27.

Cauchemez, S., Horby, P., Fox, A., Mai, L. Q., Thanh, L. T., Thai, P. Q., Hoa, L. N. M., Hien, N. T., and Ferguson, N. M. (2012). Influenza infection rates, measurement errors and the interpretation of paired serology. *PLoS Pathogens*, 8(12):e1003061.

Cauchemez, S., Valleron, A.-J., Boëlle, P.-Y., Flahault, A., and Ferguson, N. M. (2008). Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature*, 452(7188):750–4.

Cazelles, B., Chavez, M., Berteaux, D., Ménard, F., Vik, J. O., Jenouvrier, S., and Stenseth, N. C. (2008). Wavelet analysis of ecological time series. *Oecologia*, 156(2):287–304.

Cazelles, B., Chavez, M., Magny, G. C. D., Guégan, J.-F., and Hales, S. (2007). Time-dependent spectral analysis of epidemiological time-series with wavelets. *Journal of the Royal Society, Interface / the Royal Society*, 4(15):625–36.

Cazelles, B., Chavez, M., McMichael, A. J., and Hales, S. (2005). Nonstationary influence of El Niño on the synchronous dengue epidemics in Thailand. *PLoS medicine*, 2(4):e106.

CDC (2005). Prevention and Control of Influenza: Recommendations of the Advisory Committee on Immunization Practices (ACIP).

CDC, U. (2013). Overview of Influenza Surveillance in the United States. Technical report.

central population, C. and housing steering committee (2010). The 2009 vietnam population and housing census major finding. Technical report.

Chan, P., Mok, H., and Lee, T. (2009). Seasonal influenza activity in Hong Kong and its association with meteorological variations. *Journal of Medical Virology*, 1806(81):1797–1806.

Chang, L.-Y., Chen, W.-H., Lu, C.-Y., Shao, P.-L., Fan, T.-Y., Cheng, A.-L., and Huang, L.-M. (2011). Household transmission of Pandemic (H1N1) 2009 Virus, Taiwan. *Emerging infectious diseases*, 17(10):1928–31.

Cheng, X., Tan, Y., He, M., Lam, T. T.-Y., Lu, X., Viboud, C., He, J., Zhang, S., Lu, J., Wu, C., Fang, S., Wang, X., Xie, X., Ma, H., Nelson, M. I., Kung, H.-f., Holmes, E. C., and Cheng, J. (2013). Epidemiological dynamics and phylogeography of influenza virus in southern China. *The Journal of infectious diseases*, 207(1):106–14.

Chittaganpitch, M., Supawat, K., Olsen, S. J., Waicharoen, S., Patthamadilok, S., Yingyong, T., Brammer, L., Epperson, S. P., Akrasewi, P., and Sawanpanyalert, P. (2012). Influenza viruses in Thailand: 7 years of sentinel surveillance data, 2004-2010. *Influenza and other respiratory viruses*, 6(4):276–83.

Chiu, S. S., Lau, Y. L., Chan, K. H., Wong, W. H. S., and Peiris, J. S. M. (2002). Influenza-related hospitalizations among children in Hong Kong. *The New England journal of medicine*, 347(26):2097–103.

Chowell, G., Cazelles, B., Broutin, H., and Munayco, C. V. (2011). The influence of geographic and climate factors on the timing of dengue epidemics in Peru, 1994-2008. *BMC Infect Dis*, 11:164.

Clancy, S. (2008). Genetics of the Influenza Virus. *Nature Education*, 1(1):83.

Couch, R. B. and Kasel, J. A. (1983). Immunity to influenza in man. *Annual review of microbiology*, 37:529–49.

Coudeville, L., Bailleux, F., Riche, B., Megas, F., Andre, P., and Ecochard, R. (2010). Relationship between haemagglutination-inhibiting antibody titres and clinical protection against influenza: development and application of a bayesian random-effects model. *BMC medical research methodology*, 10:18.

Cowling, B. J., Chan, K. H., Fang, V. J., Lau, L. L. H., So, H. C., Fung, R. O. P., Ma, E. S. K., Kwong, A. S. K., Chan, C.-W., Tsui, W. W. S., Ngai, H.-Y., Chu, D. W. S., Lee, P. W. Y., Chiu, M.-C., Leung, G. M., and Peiris, J. S. M. (2010). Comparative epidemiology of pandemic and seasonal influenza A in households. *The New England*

journal of medicine, 362(23):2175–84.

Cowling, B. J., Ip, D. K. M., Fang, V. J., Suntarattiwong, P., Olsen, S. J., Levy, J., Uyeki, T. M., Leung, G. M., Malik Peiris, J. S., Chotpitayasunondh, T., Nishiura, H., and Mark Simmerman, J. (2013). Aerosol transmission is an important mode of influenza A virus spread. *Nature communications*, 4:1935.

de Jong, M. D., Bach, V. C., Phan, T. Q., Vo, M. H., Tran, T. T., Nguyen, B. H., Beld, M., Le, T. P., Truong, H. K., Nguyen, V. V. C., Tran, T. H., Do, Q. H., and Farrar, J. (2005). Fatal avian influenza A (H5N1) in a child presenting with diarrhea followed by coma. *The New England journal of medicine*, 352(7):686–91.

Dominici, F., McDermott, A., Zeger, S. L., and Samet, J. M. (2002). On the use of generalized additive models in time-series studies of air pollution and health. *American journal of epidemiology*, 156(3):193–203.

Donnelly, C. A., Finelli, L., Cauchemez, S., Olsen, S. J., Doshi, S., Jackson, M. L., Kennedy, E. D., Kamimoto, L., Marchbanks, T. L., Morgan, O. W., Patel, M., Swerdlow, D. L., and Ferguson, N. M. (2011). Serial intervals and the temporal distribution of secondary infections within households of 2009 pandemic influenza A (H1N1): implications for influenza control recommendations. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 52 Suppl 1(suppl_1):S123–30.

Doraisingham, S., Goh, K. T., Ling, A. E., and Yu, M. (1988). Influenza surveillance in Singapore: 1972-86. *Bulletin of the World Health Organization*, 66(1):57–63.

Dowell, S. F. (2001). Seasonal variation in host susceptibility and cycles of certain infectious diseases. *Emerg Infect Dis*, 7(3):369–374.

Dushoff, J., Plotkin, J. B., Levin, S. a., and Earn, D. J. D. (2004). Dynamical resonance can account for seasonality of influenza epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(48):16915–6.

Dwyer, D., Barr, I., Hurt, A., Kelso, A., Reading, P., Sullivan, S., Buchy, P., Yu, H., Zheng, J., Shu, Y., Wang, D., Lam, Aguon, A., Oliva, R. Q., Odagiri, T., Tashiro, M., Verasahib, K., Yusof, M. A., Nymadawa, P., Alexander, B., Gourinat, A.-C., Grangeon,

J.-P., Jennings, L., Huang, S., Horwood, P., Lucero, M., Roque, V., Lee Suy, L., Cardon, P., Tandoc, A., Olveda, R. M., Kang, C., Young-Joon, P., Cutter, J., Lin, R., Low, C., Mai, L. T. Q., Balish, A., Kile, J., Mei, S., Mcfarland, J., Moen, A., Olsen, S., Samaan, G., Xiyan, X., Chea, N., Diorditsa, S., Feldon, K., Fox, K., Jamsran, M., Konings, F., Lewis, H. C., McPherson, M., Nilles, E., Olowokure, B., and Partridge, J. (2013). Seasonal influenza vaccine policies, recommendations and use in the World Health Organization's Western Pacific Region. *Western Pacific surveillance and response journal : WPSAR*, 4(3):51–9.

Eames, K. T. D., Tilston, N. L., and Edmunds, W. J. (2011). The impact of school holidays on the social mixing patterns of school children. *Epidemics*, 3(2):103–8.

Earn, D. J., Dushoff, J., and Levin, S. a. (2002). Ecology and evolution of the flu. *Trends in Ecology & Evolution*, 17(7):334–340.

Elliot, A. and DM, F. (2006). Surveillance of influenza-like illness in England and Wales during 1966-2006. Technical report.

Ferguson, N., Cummings, D., and Fraser, C. (2006). Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448 – 452.

Fielding, J. E., Kelly, H. a., Mercer, G. N., and Glass, K. (2014). Systematic review of influenza A(H1N1)pdm09 virus shedding: duration is affected by severity, but not age. *Influenza and other respiratory viruses*, 8(2):142–50.

Finkelman, B. S., Viboud, C., Koelle, K., Ferrari, M. J., Bharti, N., and Grenfell, B. T. (2007). Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: viral coexistence and latitudinal gradients. *PLoS ONE*, 2(12):e1296.

France, A. M., Jackson, M., Schrag, S., Lynch, M., Zimmerman, C., Biggerstaff, M., and Hadler, J. (2010). Household transmission of 2009 influenza A (H1N1) virus after a school-based outbreak in New York City, April-May 2009. *The Journal of infectious diseases*, 201(7):984–92.

Fukuyama, S. and Kawaoka, Y. (2011). The pathogenesis of influenza virus infections: the contributions of virus and host factors. *Current opinion in immunology*, 23(4):481–6.

Gao, R., Cao, B., Hu, Y., Feng, Z., Wang, D., Hu, W., Chen, J., Jie, Z., Qiu, H., Xu, K., Xu, X., Lu, H., Zhu, W., Gao, Z., Xiang, N., Shen, Y., He, Z., Gu, Y., Zhang, Z., Yang, Y., Zhao, X., Zhou, L., Li, X., Zou, S., Zhang, Y., Li, X., Yang, L., Guo, J., Dong, J., Li, Q., Dong, L., Zhu, Y., Bai, T., Wang, S., Hao, P., Yang, W., Zhang, Y., Han, J., Yu, H., Li, D., Gao, G. F., Wu, G., Wang, Y., Yuan, Z., and Shu, Y. (2013). Human infection with a novel avian-origin influenza A (H7N9) virus. *The New England journal of medicine*, 368(20):1888–97.

Garske, T., Legrand, J., Donnelly, C. A., Ward, H., Cauchemez, S., Fraser, C., Ferguson, N. M., and Ghani, A. C. (2009). Assessing the severity of the novel influenza A/H1N1 pandemic. *BMJ (Clinical research ed.)*, 339(july):b2840.

Gessner, B. D., Shindo, N., and Briand, S. (2011). Seasonal influenza epidemiology in sub-Saharan Africa: a systematic review. *The Lancet infectious diseases*, 11(3):223–35.

Gleeson, M., Fleming, D. M., and Elliot, A. J. (2005). The impact of influenza on the health and health care utilisation of elderly people. *Vaccine*, 23:S1–S9.

Goldstein, E., Cobey, S., Takahashi, S., Miller, J. C., and Lipsitch, M. (2011). Predicting the epidemic sizes of influenza A/H1N1, A/H3N2, and B: a statistical method. *PLoS Med*, 8(7):e1001051.

Gouhier, T. and Grinsted, A. (2013). biwavelet: Conduct univariate and bivariate wavelet analyses.

Grant, W. B. and Giovannucci, E. (2009). The possible roles of solar ultraviolet-B radiation and vitamin D in reducing case-fatality rates from the 1918-1919 influenza pandemic in the United States. *Dermato-endocrinology*, 1(4):215–9.

Grassly, N. C. and Fraser, C. (2006). Seasonal infectious disease epidemiology. *Proceedings. Biological sciences / The Royal Society*, 273(1600):2541–50.

Grenfell, B. T., Bjø rnstad, O. N., and Kappey, J. (2001). Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414(6865):716–23.

Grinsted, A., Moore, J. C., and Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in*

Geophysics, 11(5/6):561–566.

GSO (2012). General Statistics Office Of Vietnam web page.

Gu, Y., Komiya, N., Kamiya, H., Yasui, Y., Taniguchi, K., and Okabe, N. (2011). Pandemic (H1N1) 2009 transmission during presymptomatic phase, Japan. *Emerging infectious diseases*, 17(9):1737–9.

Gubareva, L. V., Novikov, D. V., and Hayden, F. G. (2002). Assessment of hemagglutinin sequence heterogeneity during influenza virus transmission in families. *The Journal of infectious diseases*, 186(11):1575–81.

Gupta, V., Dawood, F. S., Muangchana, C., Lan, P. T., Xeuatvongsa, A., Sovann, L., Olveda, R., Cutter, J., Oo, K. Y., Ratih, T. S. D., Kheong, C. C., Kapella, B. K., Kitsutani, P., Corwin, A., and Olsen, S. J. (2012). Influenza vaccination guidelines and vaccine sales in southeast Asia: 2008-2011. *PloS one*, 7(12):e52842.

Hale, B. G., Randall, R. E., Ortín, J., and Jackson, D. (2008). The multifunctional NS1 protein of influenza A viruses. *The Journal of general virology*, 89(Pt 10):2359–76.

Harper, G. J. (1961). Airborne micro-organisms: survival tests with four viruses. *The Journal of hygiene*, 59:479–86.

Hayward, A. C., Fragaszy, E. B., Bermingham, A., Wang, L., Copas, A., Edmunds, W. J., Ferguson, N., Goonetilleke, N., Harvey, G., Kovar, J., Lim, M. S. C., Mcmichael, A., Millett, E. R. C., Nguyen-van tam, J. S., and Nazareth, I. (2013). Comparative community burden and severity of seasonal and pandemic infl uenza : results of the Flu Watch cohort study. *The Lancet Respiratory*, 2600(14):16–19.

Held, L. and Paul, M. (2012). Modeling seasonality in space-time infectious disease surveillance data. *Biometrical journal. Biometrische Zeitschrift*, 54(6):824–43.

Hemmes, J. H., Winkler, K. C., and Kool, S. M. (1960). Virus survival as a seasonal factor in influenza and polimyelitis. *Nature*, 188:430–1.

Hens, N., Ayele, G. M., Goeyvaerts, N., Aerts, M., Mossong, J., Edmunds, J. W., and Beutels, P. (2009). Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight European countries. *BMC*

infectious diseases, 9:187.

Hermes, J., Bernard, H., Buchholz, U., Spackova, M., Löw, J., Loytved, G., Suess, T., Hautmann, W., and Werber, D. (2011). Lack of evidence for pre-symptomatic transmission of pandemic influenza virus A(H1N1) 2009 in an outbreak among teenagers; Germany, 2009. *Influenza and other respiratory viruses*, 5(6):e499–503.

Heymann, D. L. and American Public Health Association (2008). *Control of communicable diseases manual.* American Public Health Association, Washington, {DC}.

Hien, T. T., Boni, M. F., Bryant, J. E., Ngan, T. T., Wolbers, M., Nguyen, T. D., Truong, N. T., Dung, N. T., Ha, D. Q., Hien, V. M., Thanh, T. T., Nhu, L. N. T., Uyen, L. T. T., Nhien, P. T., Chinh, N. T., Chau, N. V. V., Farrar, J., and van Doorn, H. R. (2010). Early pandemic influenza (2009 H1N1) in Ho Chi Minh City, Vietnam: a clinical virological and epidemiological analysis. *PLoS medicine*, 7(5):e1000277.

Hippocrates. Of the Epidemics, book II, section III. In *book II*, page paragraph 13.

Hoa, L. K., Hiep, L. V., and Be, L. V. (2011). Development of pandemic influenza vaccine production capacity in Viet Nam. *Vaccine*, 29 Suppl 1:A34–6.

Holland, S. M. (2008). Principal components analysis (PCA).

Hood, a. M. (1963). Infectivity of Influenza Virus Aerosols. *The Journal of hygiene*, 61:331–5.

Horby, P., Mai, L. Q., Fox, A., Thai, P. Q., Thi Thu Yen, N., Thanh, L. T., Le Khanh Hang, N., Duong, T. N., Thoang, D. D., Farrar, J., Wolbers, M., and Hien, N. T. (2012). The Epidemiology of Interpandemic and Pandemic Influenza in Vietnam, 20072010 The Ha Nam Household Cohort Study I. *American Journal of Epidemiology*, 175(10):1062–74.

Horby, P., Pham, Q. T., Hens, N., Nguyen, T. T. Y., Le, Q. M., Dang, D. T., Nguyen, M. L., Nguyen, T. H., Alexander, N., Edmunds, W. J., Tran, N. D., Fox, A., and Nguyen, T. H. (2011). Social contact patterns in Vietnam and implications for the control of infectious diseases. *PloS one*, 6(2):e16965.

Horby, P., Sudoyo, H., Viprakasit, V., Fox, A., Thai, P. Q., Yu, H., Davila, S., Hibberd,

M., Dunstan, S. J., Monteerarat, Y., Farrar, J. J., Marzuki, S., and Hien, N. T. (2010). What is the evidence of a role for host genetics in susceptibility to influenza A/H5N1? *Epidemiology and Infection*, 138(11):1550–1558.

Horby, P. W. (2014). Community studies of influenza: new knowledge , new question. *The Lancet Respiratory*, 2600(14):1–2.

Jackson, C., Mangtani, P., Vynnycky, E., Fielding, K., Kitching, A., Mohamed, H., Roche, A., and Maguire, H. (2011). School closures and student contact patterns. *Emerging infectious diseases*, 17(2):245–7.

Jakeman, K. J. and Sweet, C. (1996). Summer outbreaks of influenza. *The Journal of infectious diseases*, 174(3):674–5.

Jian, J.-W., Chen, G.-W., Lai, C.-T., Hsu, L.-C., Chen, P.-J., Kuo, S. H.-S., Wu, H.-S., and Shih, S.-R. (2008). Genetic and epidemiological analysis of influenza virus epidemics in Taiwan during 2003 to 2006. *Journal of clinical microbiology*, 46(4):1426–34.

Jordan, W. S. (1961). The mechanism of spread of Asian influenza. *The American review of respiratory disease*, 83(2)Pt 2:29–40.

Juzeniene, A., Ma, L.-W., Kwitniewski, M., Polev, G. A., Lagunova, Z., Dahlback, A., and Moan, J. (2010). The seasonality of pandemic and non-pandemic influenzas: the roles of solar radiation and vitamin D. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases*, 14(12):e1099–105.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D. (1996). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, 77(3):437–471.

Kamps, B., Hoffmann, C., and Preiser, W. (2006). Influenza report 2006. *Archives of Internal Medicine*, 162(16):1842–1848.

Katz, M. a., Lebo, E., Emukule, G., Njuguna, H. N., Aura, B., Cosmas, L., Audi, A.,

Junghae, M., Waiboci, L. W., Olack, B., Bigogo, G., Njenga, M. K., Feikin, D. R., and Breiman, R. F. (2012a). Epidemiology, seasonality, and burden of influenza and influenza-like illness in urban and rural Kenya, 2007-2010. *The Journal of infectious diseases*, 206 Suppl(ILI):S53–60.

Katz, M. a., Schoub, B. D., Heraud, J. M., Breiman, R. F., Njenga, M. K., and Widdowson, M.-A. (2012b). Influenza in Africa: uncovering the epidemiology of a long-overlooked disease. *The Journal of infectious diseases*, 206 Suppl(Suppl 1):S1–4.

Kelly, H. A., Grant, K. A., Tay, E. L., Franklin, L., and Hurt, A. C. (2013). The significance of increased influenza notifications during spring and summer of 2010-11 in Australia. *Influenza and other respiratory viruses*, 7(6):1136–41.

Khamphaphongphane, B., Ketmayoon, P., Lewis, H. C., Phonekeo, D., Sisouk, T., Xayadeth, S., Ongkhammy, S., Vongphrachanh, P., Tsuyuoka, R., Moen, A., and Corwin, A. (2013). Epidemiological and virological characteristics of seasonal and pandemic influenza in Lao PDR, 2008-2010. *Influenza and other respiratory viruses*, 7(3):304–11.

Killingley, B. and Nguyen-Van-Tam, J. (2013). Routes of influenza transmission. *Influenza and other respiratory viruses*, 7 Suppl 2:42–51.

Kingdon, K. H. (1960). Relative humidity and air-borne infections. *The American review of respiratory disease*, 81(34):504–12.

Klick, B., Leung, G. M., and Cowling, B. J. (2012). Optimal design of studies of influenza transmission in households. I: case-ascertained studies. *Epidemiology and infection*, 140(1):106–14.

Klick, B., Nishiura, H., Ng, S., Fang, V. J., Leung, G. M., Peiris, J. S. M., and Cowling, B. J. (2011). Transmissibility of seasonal and pandemic influenza in a cohort of households in Hong Kong in 2009. *Epidemiology (Cambridge, Mass.)*, 22(6):793–6.

Kohn, M. a., Farley, T. a., Sundin, D., Tapia, R., McFarland, L. M., and Arden, N. H. (1995). Three summertime outbreaks of influenza type A. *The Journal of infectious diseases*, 172(1):246–9.

Komiya, N., Gu, Y., Kamiya, H., Yahata, Y., Yasui, Y., Taniguchi, K., and Okabe, N.

(2010). Household transmission of pandemic 2009 influenza A (H1N1) virus in Osaka, Japan in May 2009. *The Journal of infection*, 61(4):284–8.

Kosasih, H., Roselinda, Nurhayati, Klimov, A., Xiyan, X., Lindstrom, S., Mahoney, F., Beckett, C., Burgess, T. H., Blair, P. J., Uyeki, T. M., and Sedyaningsih, E. R. (2013). Surveillance of influenza in Indonesia, 20032007. *Influenza and other respiratory viruses*, 7(3):312–20.

Kreijtz, J. H. C. M., Fouchier, R. a. M., and Rimmelzwaan, G. F. (2011). Immune responses to influenza virus infection. *Virus research*, 162(1-2):19–30.

Laguna-Torres, V. A., Gómez, J., Ocaña, V., Aguilar, P., Saldarriaga, T., Chavez, E., Perez, J., Zamalloa, H., Forshey, B., Paz, I., Gomez, E., Ore, R., Chauca, G., Ortiz, E., Villaran, M., Vilcarromero, S., Rocha, C., Chincha, O., Jiménez, G., Villanueva, M., Pozo, E., Aspajo, J., and Kochel, T. (2009). Influenza-like illness sentinel surveillance in Peru. *PloS one*, 4(7):e6118.

Lau, K.-M. and Weng, H. (1995). Climate Signal Detection Using Wavelet Transform: How to Make a Time Series Sing. *Bulletin of the American Meteorological Society*, 76(12):2391–2402.

Lau, L. L. H., Cowling, B. J., Fang, V. J., Chan, K.-H., Lau, E. H. Y., Lipsitch, M., Cheng, C. K. Y., Houck, P. M., Uyeki, T. M., Peiris, J. S. M., and Leung, G. M. (2010). Viral shedding and clinical illness in naturally acquired influenza virus infections. *The Journal of infectious diseases*, 201(10):1509–16.

Lau, L. L. H., Nishiura, H., Kelly, H., Ip, D. K. M., Leung, G. M., and Cowling, B. J. (2012). Household transmission of 2009 pandemic influenza A (H1N1): a systematic review and meta-analysis. *Epidemiology (Cambridge, Mass.)*, 23(4):531–42.

Laurie, K. L., Huston, P., Riley, S., Katz, J. M., Willison, D. J., Tam, J. S., Mounts, A. W., Hoschler, K., Miller, E., Vandemaele, K., Broberg, E., Van Kerkhove, M. D., and Nicoll, A. (2013). Influenza serological studies to inform public health action: best practices to optimise timing, quality and reporting. *Influenza and other respiratory viruses*, 7(2):211–24.

Le, M. Q., Lam, H. M., Cuong, V. D., Lam, T. T.-Y., Halpin, R. A., Wentworth, D. E., Hien, N. T., Thanh, L. T., Phuong, H. V. M., Horby, P., and Boni, M. F. (2013). Migration and persistence of human influenza A viruses, Vietnam, 2001-2008. *Emerging infectious diseases*, 19(11):1756–65.

Lee, V. J., Yap, J., Ong, J. B. S., Chan, K.-P. P., Lin, R. T. P., Chan, S. P., Goh, K. T., Leo, Y.-S. S., and Chen, M. I.-C. (2009). Influenza excess mortality from 1950-2000 in tropical Singapore. *PLoS ONE*, 4(12):e8096.

Leung, Y. H. C., Zhang, L.-J., Chow, C.-K., Tsang, C.-L., Ng, C.-f., Wong, C.-k., Guan, Y., and Peiris, J. S. M. (2007). Poultry drinking water used for avian influenza surveillance. *Emerging infectious diseases*, 13(9):1380–2.

Loeb, M., Singh, P. K., Fox, J., Russell, M. L., Pabbaraju, K., Zarra, D., Wong, S., Neupane, B., Singh, P., Webby, R., and Fonseca, K. (2012). Longitudinal study of influenza molecular viral shedding in Hutterite communities. *The Journal of infectious diseases*, 206(7):1078–84.

Lofgren, E., Fefferman, N. H., Naumov, Y. N., Gorski, J., and Naumova, E. N. (2007). Influenza seasonality: underlying causes and modeling theories. *Journal of virology*, 81(11):5429–36.

Loh, W. (2008). Classification and regression tree methods. *Encyclopedia of statistics in quality and reliability*, pages 315–323.

Looker, C., Carville, K., Grant, K., and Kelly, H. (2010). Influenza A (H1N1) in Victoria, Australia: a community case series and analysis of household transmission. *PloS one*, 5(10):e13702.

Loustalot, F., Silk, B. J., Gaither, A., Shim, T., Lamias, M., Dawood, F., Morgan, O. W., Fishbein, D., Guerra, S., Verani, J. R., Carlson, S. A., Fonseca, V. P., and Olsen, S. J. (2011). Household transmission of 2009 pandemic influenza A (H1N1) and nonpharmaceutical interventions among households of high school students in San Antonio, Texas. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 52 Suppl 1:S146–53.

Lowen, A., Mubareka, S., Steel, J., and Palese, P. (2007). Influenza virus transmission is dependent on relative humidity and temperature. *PLoS pathogens*, 3(10):1470–6.

Lowen, A. and Palese, P. (2009). Transmission of influenza virus in temperate zones is predominantly by aerosol, in the tropics by contact: a hypothesis. *PLoS currents*, 1:RRN1002.

Lowen, A. C., Steel, J., Mubareka, S., and Palese, P. (2008). High temperature (30 degrees C) blocks aerosol but not contact transmission of influenza virus. *Journal of virology*, 82(11):5650–2.

Lutwama, J. J., Bakamutumaho, B., Kayiwa, J. T., Chiiza, R., Namagambo, B., Katz, M. a., and Geissler, A. L. (2012). Clinic- and hospital-based sentinel influenza surveillance, Uganda 2007-2010. *The Journal of infectious diseases*, 206 Suppl(September 2010):S87–93.

Macroepidemiology of Influenza Vaccination (MIV) Study Group (2005). The macroepidemiology of influenza vaccination in 56 countries, 1997–2003. *Vaccine*, 23(44):5133–43.

Mahamat, a., Dussart, P., Bouix, A., Carvalho, L., Eltges, F., Matheus, S., Miller, M. a., Quenel, P., and Viboud, C. (2013). Climatic drivers of seasonal influenza epidemics in French Guiana, 2006-2010. *The Journal of infection*, 67(2):141–7.

Mathews, J. D., McCaw, C. T., McVernon, J., McBryde, E. S., and McCaw, J. M. (2007). A biological model for influenza transmission: pandemic planning implications of asymptomatic infection and immunity. *PloS one*, 2(11):e1220.

Matrosovich, M. N., Matrosovich, T. Y., Gray, T., Roberts, N. A., and Klenk, H. D. (2004). Human and avian influenza viruses target different cell types in cultures of human airway epithelium. *Proc Natl Acad Sci U S A*, 101(13):4620–4624.

McKimm-Breschkin, J. L. (2013). Influenza neuraminidase inhibitors: antiviral action and mechanisms of resistance. *Influenza and other respiratory viruses*, 7 Suppl 1:25–36.

Meerhoff, T. J., Meijer, A., and Paget, W. J. (2004). Methods for sentinel virological surveillance of influenza in Europe - an 18-country survey. *Euro surveillance : bulletin*

Européen sur les maladies transmissibles = European communicable disease bulletin, 9(1):34–8.

Members of the Western Pacific Region Global Influenza Surveillance and Response System (2012). Epidemiological and virological characteristics of influenza in the Western Pacific Region of the World Health Organization, 2006-2010. *PloS one*, 7(5):e37568.

Mertz, D., Kim, T. H., Johnstone, J., Lam, P.-P., Science, M., Kuster, S. P., Fadel, S. A., Tran, D., Fernandez, E., Bhatnagar, N., and Loeb, M. (2013). Populations at risk for severe or complicated influenza illness: systematic review and meta-analysis. *BMJ (Clinical research ed.)*, 347(August):f5061.

Mi, X., Ren, H., Ouyang, Z., Wei, W., and Ma, K. (2005). The use of the Mexican Hat and the Morlet wavelets for detection of ecological patterns. *Plant Ecology*, 179(1):1–19.

Minhaz Ud-Dean, S. M. (2010). Structural explanation for the effect of humidity on persistence of airborne virus: seasonality of influenza. *Journal of theoretical biology*, 264(3):822–9.

Ministry of Health, V. (2010). Circulation No 48/2010/TT-BYT on Guildline for reporting, declaring infectious disease.

Ministry of Health, V. (2013). Decision No 1950/QD- BYT on Approval for the plan of influenza virus vaccine development and use in period 2013-2020, vision 2030.

Ministry of Health, V. N. (2009). Decision No 1846/QD-BYT on Approval for the guildline for pandemic influenza H1N1 prevention and control.

Monto, A. S. (1994). Studies of the community and family: acute respiratory illness and infection. *Epidemiologic reviews*, 16(2):351–73.

Monto, A. S., Gravenstein, S., Elliott, M., Colopy, M., and Schweinle, J. (2000). Clinical signs and symptoms predicting influenza infection. *Archives of internal medicine*, 160(21):3243–7.

Morgan, O. W., Parks, S., Shim, T., Blevins, P. A., Lucas, P. M., Sanchez, R., Walea, N., Loustalot, F., Duffy, M. R., Shim, M. J., Guerra, S., Guerra, F., Mills, G., Verani, J., Alsip, B., Lindstrom, S., Shu, B., Emery, S., Cohen, A. L., Menon, M., Fry, A. M.,

Dawood, F., Fonseca, V. P., and Olsen, S. J. (2010). Household transmission of pandemic (H1N1) 2009, San Antonio, Texas, USA, April-May 2009. *Emerging infectious diseases*, 16(4):631–7.

Moura, F. E. a., Perdigão, A. C. B., and Siqueira, M. M. (2009). Seasonality of influenza in the tropics: a distinct pattern in northeastern Brazil. *The American journal of tropical medicine and hygiene*, 81(1):180–3.

Murray, E. L., Klein, M., Brondi, L., McGowan, J. E., van Mels, C., Brooks, W. A., Kleinbaum, D., Goswami, D., Ryan, P. B., and Bridges, C. B. (2012). Rainfall, household crowding, and acute respiratory infections in the tropics. *Epidemiology and infection*, 140(1):78–86.

Nair, H., Nokes, D. J., Gessner, B. D., Dherani, M., Madhi, S. a., Singleton, R. J., O'Brien, K. L., Roca, A., Wright, P. F., Bruce, N., Chandran, A., Theodoratou, E., Sutanto, A., Sedyaningsih, E. R., Ngama, M., Munywoki, P. K., Kartasasmita, C., Simões, E. a. F., Rudan, I., Weber, M. W., and Campbell, H. (2010). Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. *Lancet*, 375(9725):1545–55.

Navarro-Marí, J. M., Pérez-Ruiz, M., Cantudo-Muñoz, P., Petit-Gancedo, C., Jiménez-Valera, M., and Rosa-Fraile, M. (2005). Influenza-like illness criteria were poorly related to laboratory-confirmed influenza in a sentinel surveillance study. *Journal of Clinical Epidemiology*, 58(3):275–279.

Nelson, M. I., Simonsen, L., Viboud, C., Miller, M. A., and Holmes, E. C. (2007). Phylogenetic analysis reveals the global migration of seasonal influenza A viruses. *PLoS pathogens*, 3(9):1220–8.

Neuzil, K. M., Mellen, B. G., Wright, P. F., Mitchel, E. F., and Griffin, M. R. (2000). The effect of influenza on hospitalizations, outpatient visits, and courses of antibiotics in children. *The New England journal of medicine*, 342(4):225–31.

Nguyen, H. L. K., Saito, R., Ngiem, H. K., Nishikawa, M., Shobugawa, Y., Nguyen, D. C., Hoang, L. T., Huynh, L. P., and Suzuki, H. (2007). Epidemiology of influenza in Hanoi,

Vietnam, from 2001 to 2003. *The Journal of infection*, 55(1):58–63.

Nguyen, H. T., Dharan, N. J., Le, M. T. Q., Nguyen, N. B., Nguyen, C. T., Hoang, D. V., Tran, H. N., Bui, C. T., Dang, D. T., Pham, D. N., Nguyen, H. T., Phan, T. V., Dennis, D. T., Uyeki, T. M., Mott, J., and Nguyen, Y. T. (2009). National influenza surveillance in Vietnam, 2006-2007. *Vaccine*, 28(2):398–402.

Nicholson, K. G., Wood, J. M., and Zambon, M. (2003). Influenza. *Lancet*, 362(9397):1733–1745.

Nyatanyi, T., Nkunda, R., Rukelibuga, J., Palekar, R., Muhimpundu, M. A., Kabeja, A., Kabanda, A., Lowrance, D., Tempia, S., Koama, J. B., McAlister, D., Mukabayire, O., Wane, J., Raghunathan, P., Katz, M., and Karema, C. (2012). Influenza sentinel surveillance in Rwanda, 2008-2010. *The Journal of infectious diseases*, 206 Suppl(Suppl 1):S74–9.

Ohmit, S. E. and Monto, A. S. (2006). Symptomatic predictors of influenza virus positivity in children during the influenza season. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 43(5):564–8.

Paget, J., Marquet, R., Meijer, A., and van der Velden, K. (2007). Influenza activity in Europe during eight seasons (1999-2007): an evaluation of the indicators used to measure activity and an assessment of the timing, length and course of peak activity (spread) across Europe. *BMC infectious diseases*, 7(4):141.

Papenburg, J., Baz, M., Hamelin, M.-E., Rhéaume, C., Carbonneau, J., Ouakki, M., Rouleau, I., Hardy, I., Skowronski, D., Roger, M., Charest, H., De Serres, G., and Boivin, G. (2010). Household transmission of the 2009 pandemic A/H1N1 influenza virus: elevated laboratoryconfirmed secondary attack rates and evidence of asymptomatic infections. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 51(9):1033–41.

Park, A. W. and Glass, K. (2007). Dynamic patterns of avian and human influenza in east and southeast Asia. *The Lancet infectious diseases*, 7(8):543–8.

Pascalis, H., Temmam, S., Wilkinson, D. a., Dsouli, N., Turpin, M., de Lamballerie, X.,

and Dellagi, K. (2012). Molecular evolutionary analysis of pH1N1 2009 influenza virus in Reunion Island, South West Indian Ocean region: a cohort study. *PloS one*, 7(8):e43742.

Patrozou, E. and Mermel, L. A. (2009). Does influenza transmission occur from asymptomatic infection or prior to symptom onset? *Public health reports*, 124(2):193–6.

Paul, J. and Freese, H. (1933). An epidemiological and bacteriological study of the common cold in an isolated arctic community (Spitsbergen). *American Journal of Epidemiology*, 17(3):517–535.

Paul Glezen, W., Schmier, J. K., Kuehn, C. M., Ryan, K. J., and Oxford, J. (2013). The burden of influenza B: a structured literature review. *American journal of public health*, 103(3):e43–51.

Pebody, R. G., Harris, R., Kafatos, G., Chamberland, M., Campbell, C., Nguyen-Van-Tam, J. S., McLean, E., Andrews, N., White, P. J., Wynne-Evans, E., Green, J., Ellis, J., Wreghitt, T., Bracebridge, S., Ihekweazu, C., Oliver, I., Smith, G., Hawkins, C., Salmon, R., Smyth, B., McMenamin, J., Zambon, M., Phin, N., and Watson, J. M. (2011). Use of antiviral drugs to reduce household transmission of pandemic (H1N1) 2009, United Kingdom. *Emerging infectious diseases*, 17(6):990–9.

Poon, L. L. M., Chan, K. H., Chu, D. K. W., Fung, C. C. Y., Cheng, C. K. Y., Ip, D. K. M., Leung, G. M., Peiris, J. S. M., and Cowling, B. J. (2011). Viral genetic sequence variations in pandemic H1N1/2009 and seasonal H3N2 influenza viruses within an individual, a household and a community. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*, 52(2):146–50.

Powell, T. J., Fox, A., Peng, Y., Quynh Mai, L. T., Lien, V. T. K., Hang, N. L. K., Wang, L., Lee, L. Y.-H., Simmons, C. P., McMichael, A. J., Farrar, J. J., Askonas, B. A., Duong, T. N., Thai, P. Q., Thu Yen, N. T., Rowland-Jones, S. L., Hien, N. T., Horby, P., and Dong, T. (2012). Identification of H5N1-specific T-cell responses in a high-risk cohort in vietnam indicates the existence of potential asymptomatic infections. *The Journal of infectious diseases*, 205(1):20–7.

Radin, J. M., Katz, M. a., Tempia, S., Talla Nzussouo, N., Davis, R., Duque, J., Adedeji,

A., Adjabeng, M. J., Ampofo, W. K., Ayele, W., Bakamutumaho, B., Barakat, A., Cohen, A. L., Cohen, C., Dalhatu, I. T., Daouda, C., Dueger, E., Francisco, M., Heraud, J.-M., Jima, D., Kabanda, A., Kadjo, H., Kandeel, A., Bi Shamamba, S. K., Kasolo, F., Kronmann, K. C., Mazaba Liwewe, M. L., Lutwama, J. J., Matonya, M., Mmbaga, V., Mott, J. a., Muhimpundu, M. A., Muthoka, P., Njuguna, H., Randrianasolo, L., Refaey, S., Sanders, C., Talaat, M., Theo, A., Valente, F., Venter, M., Woodfill, C., Bresee, J., Moen, A., and Widdowson, M.-A. (2012). Influenza surveillance in 15 countries in Africa, 2006-2010. *The Journal of infectious diseases*, 206 Suppl(Suppl 1):S14–21.

Rambaut, A., Pybus, O. G., Nelson, M. I., Viboud, C., Taubenberger, J. K., and Holmes, E. C. (2008). The genomic and epidemiological dynamics of human influenza A virus. *Nature*, 453(7195):615 – 619.

Reichert, T. a., Simonsen, L., Sharma, A., Pardo, S. a., Fedson, D. S., and Miller, M. a. (2004). Influenza and the winter increase in mortality in the United States, 1959-1999. *American journal of epidemiology*, 160(5):492–502.

Reina, J., Nicolau, A., Galmes, A., and Arbona, B. (2009). [Diagnostic yield of paediatric respiratory samples in the Balearic Islands Sentinel Influenza Surveillance Network]. *Anales de pediatría (Barcelona, Spain : 2003)*, 70(5):438–42.

Riley, S., Kwok, K. O., Wu, K. M., Ning, D. Y., Cowling, B. J., Wu, J. T., Ho, L.-M., Tsang, T., Lo, S.-V., Chu, D. K. W., Ma, E. S. K., and Peiris, J. S. M. (2011). Epidemiological characteristics of 2009 (H1N1) pandemic influenza based on paired sera from a longitudinal community cohort study. *PLoS medicine*, 8(6):e1000442.

Ringnér, M. (2008). What is principal component analysis? *Nature biotechnology*, 26:303–304.

Ripley, B. (2005). Classification and regression trees. *R package version 1.0-35*, pages 0–1.

Roberts, K. L., Shelton, H., Stilwell, P., and Barclay, W. S. (2012). Transmission of a 2009 H1N1 pandemic influenza virus occurs before fever is detected, in the ferret model. *PloS one*, 7(8):e43303.

Rott, O., Charreire, J., and Cash, E. (1996). Influenza A virus hemagglutinin is a B

cell-superstimulatory lectin. *Medical microbiology and immunology*, 184(4):185–93.

Rubel, F. and Kottek, M. (2010). Observed and projected climate shifts 19012100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorologische Zeitschrift*, 19(2):135–141.

Russell, C. A., Jones, T. C., Barr, I. G., Cox, N. J., Garten, R. J., Gregory, V., Gust, I. D., Hampson, A. W., Hay, A. J., Hurt, A. C., de Jong, J. C., Kelso, A., Klimov, A. I., Kageyama, T., Komadina, N., Lapedes, A. S., Lin, Y. P. Y. P., Mosterin, A., Obuchi, M., Odagiri, T., Osterhaus, A. D. M. E., Rimmelzwaan, G. F., Shaw, M. W., Skepner, E., Stohr, K., Tashiro, M., Fouchier, R. A. M., and Smith, D. J. (2008a). The global circulation of seasonal influenza A (H3N2) viruses. *Science (New York, N.Y.)*, 320(5874):340–6.

Russell, R. J., Kerry, P. S., Stevens, D. J., Steinhauer, D. A., Martin, S. R., Gamblin, S. J., and Skehel, J. J. (2008b). Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion. *Proceedings of the National Academy of Sciences of the United States of America*, 105(46):17736–41.

Samaan, G., McPherson, M., and Partridge, J. (2013). A review of the evidence to support influenza vaccine introduction in countries and areas of WHO's Western Pacific Region. *PloS one*, 8(7):e70003.

Schaffer, F. L., Soergel, M. E., and Straube, D. C. (1976). Survival of airborne influenza virus: effects of propagating host, relative humidity, and composition of spray fluids. *Archives of virology*, 51(4):263–73.

Schnell, J. R. and Chou, J. J. (2008). Structure and mechanism of the M2 proton channel of influenza A virus. *Nature*, 451(7178):591–5.

Schulman, J. L. and Kilbourne, E. D. (1963). Experimental transmission of influenza virus infection in mice. ii. some factors affecting the incidence of transmitted infection. *The Journal of experimental medicine*, 118:267–75.

Shaman, J., Goldstein, E., and Lipsitch, M. (2011a). Absolute humidity and pandemic versus epidemic influenza. *American journal of epidemiology*, 173(2):127–35.

Shaman, J., Jeon, C. Y., Giovannucci, E., and Lipsitch, M. (2011b). Shortcomings of vitamin D-based model simulations of seasonal influenza. *PloS one*, 6(6):e20743.

Shaman, J. and Kohn, M. (2009). Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences of the United States of America*, 106(9):3243–8.

Shaman, J., Pitzer, V. E., Viboud, C., Grenfell, B. T., and Lipsitch, M. (2010). Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS biology*, 8(2):e1000316.

Shek, L. P.-C. and Lee, B.-W. (2003). Epidemiology and seasonality of respiratory tract virus infections in the tropics. *Paediatric respiratory reviews*, 4(2):105–11.

Shoji, M., Katayama, K., and Sano, K. (2011). Absolute Humidity as a Deterministic Factor Affecting Seasonal Influenza Epidemics in Japan. *The Tohoku Journal of Experimental Medicine*, 224(4):251–256.

Shortridge, K. and Stuart-Harris, C. (1982). An influenza epicentre? *The Lancet*, 320(8302):812–813.

Sikora, C., Fan, S., Golonka, R., Sturtevant, D., Gratrix, J., Lee, B. E., Jaipaul, J., and Johnson, M. (2010). Transmission of pandemic influenza A (H1N1) 2009 within households: Edmonton, Canada. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*, 49(2):90–3.

Simmerman, J. and Uyeki, T. (2008). The burden of influenza in East and SouthEast Asia: a review of the English language literature. *Influenza and other respiratory ...*, 2(3):81–92.

Simmerman, J. M., Chittaganpitch, M., Levy, J., Chantra, S., Maloney, S., Uyeki, T., Areerat, P., Thamthitiwat, S., Olsen, S. J., Fry, A., Ungchusak, K., Baggett, H. C., and Chunsuttiwat, S. (2009). Incidence, seasonality and mortality associated with influenza pneumonia in Thailand: 2005-2008. *PloS one*, 4(11):e7776.

Simmerman, J. M., Suntarattiwong, P., Levy, J., Jarman, R. G., Kaewchana, S., Gibbons, R. V., Cowling, B. J., Sanasuttipun, W., Maloney, S. A., Uyeki, T. M., Kamimoto, L.,

and Chotipitayasunondh, T. (2011). Findings from a household randomized controlled trial of hand washing and face masks to reduce influenza transmission in Bangkok, Thailand. *Influenza and other respiratory viruses*, 5(4):256–67.

Simmerman, J. M., Thawatsupha, P., Kingnate, D., Fukuda, K., Chaising, A., and Dowell, S. F. (2004). Influenza in Thailand: a case study for middle income countries. *Vaccine*, 23(2):182–7.

Smith, D. J., Lapedes, A. S., de Jong, J. C., Bestebroer, T. M., Rimmelzwaan, G. F., Osterhaus, A. D., and Fouchier, R. A. (2004). Mapping the antigenic and genetic evolution of influenza virus. *Science*, 305(5682):371–376.

Smith, L. I. (2002). A tutorial on Principal Components Analysis.

Smith, W., Andrewes, C., and Laidlaw, P. (1933). A virus obtained from influenza patients. *The Lancet*, 222(5732):66–68.

Stohr, K. (2003). The Global Agenda on Influenza Surveillance and Control. *Vaccine*, 21(16):1744–1748.

Suess, T., Buchholz, U., Dupke, S., Grunow, R., an der Heiden, M., Heider, A., Biere, B., Schweiger, B., Haas, W., and Krause, G. (2010). Shedding and transmission of novel influenza virus A/H1N1 infection in households–Germany, 2009. *American journal of epidemiology*, 171(11):1157–64.

Suess, T., Remschmidt, C., Schink, S. B., Schweiger, B., Heider, A., Milde, J., Nitsche, A., Schroeder, K., Doellinger, J., Braun, C., Haas, W., Krause, G., and Buchholz, U. (2012). Comparison of shedding characteristics of seasonal influenza virus (sub)types and influenza A(H1N1)pdm09; Germany, 2007-2011. *PloS one*, 7(12):e51653.

Sugimoto, J. D., Borse, N. N., Ta, M. L., Stockman, L. J., Fischer, G. E., Yang, Y., Halloran, M. E., Longini, I. M., and Duchin, J. S. (2011). The effect of age on transmission of 2009 pandemic influenza A (H1N1) in a camp and associated households. *Epidemiology (Cambridge, Mass.)*, 22(2):180–7.

Tam, N. D., Thao, N. Q., and Canh, V. T. (2004). *Geographic Atlas of Vietnam*. Viet Nam Education Publishing House.

Tamerius, J., Nelson, M. I., Zhou, S. Z., Viboud, C., Miller, M. A., and Alonso, W. J. (2010). Global Influenza Seasonality: Reconciling Patterns across Temperate and Tropical Regions. *Environmental Health Perspectives*, 119(4):439–445.

Tamerius, J. D., Shaman, J., Alonso, W. J., Bloom-Feshbach, K., Uejio, C. K., Comrie, A., and Viboud, C. (2013). Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. *PLoS pathogens*, 9(3):e1003194.

Tang, J. W., Loh, T. P., Tambyah, P. A., and Koay, E. S. C. (2012). Influenza outbreaks in Singapore: epidemiology, diagnosis, treatment and prevention. *Expert review of anti-infective therapy*, 10(7):751–60.

Taubenberger, J. K. and Morens, D. M. (2010). Influenza: the once and future pandemic. *Public health reports (Washington, D.C. : 1974)*, 125 Suppl:16–26.

te Beest, D. E., van Boven, M., Hooiveld, M., van den Dool, C., and Wallinga, J. (2013). Driving factors of influenza transmission in the Netherlands. *American journal of epidemiology*, 178(9):1469–77.

Tellier, R. (2009). Aerosol transmission of influenza A virus: a review of new studies. *Journal of the Royal Society, Interface / the Royal Society*, 6 Suppl 6(September):S783–90.

Teo, S. S. S., Ellis, J. S., Aitken, C., and Booy, R. (2007). Transmission of influenza A in families. *The Pediatric infectious disease journal*, 26(7):645–7.

Teunis, P. F. M., Brienen, N., and Kretzschmar, M. E. E. (2010). High infectivity and pathogenicity of influenza A virus via aerosol and droplet transmission. *Epidemics*, 2(4):215–22.

Thai, P. Q., Mai, L. Q., Welkers, M. R., Le Khanh Hang, N., Thanh, L. T., Viet Dung, V. T., Thu Yen, N. T., Duong, T. N., Minh Hoa, L. N., Thoang, D. D., Huyen Trang, H. T., de Jong, M. D., Wertheim, H., Hien, N. T., Horby, P., and Fox, A. (2014). PANDEMIC H1N1 virus transmission and shedding dynamics IN index case households OF A prospective VIETNAMESE cohort. *The Journal of infection*.

Thomas, R. E. (2014). Is influenza-like illness a useful concept and an appropriate test of

influenza vaccine effectiveness? *Vaccine*, (December 2013):2–8.

Thompson, W. W., Comanor, L., and Shay, D. K. (2006). Epidemiology of seasonal influenza: use of surveillance data and statistical models to estimate the burden of disease. *The Journal of infectious diseases*, 194 Suppl:S82–91.

Thursky, K., Cordova, S. P., Smith, D., and Kelly, H. (2003). Working towards a simple case definition for influenza surveillance. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*, 27(2):170–9.

Torner, N., Soldevila, N., Martínez, A., Pumarola, T., and Dominguez, A. (2012). Timely Prediction of Peak Seasonal Influenza Activity Estimation Using Sentinel Surveillance Data. *Public Health Research*, 2(3):53–57.

Torrence, C. and Compo, G. P. (1998). A Practical Guide to Wavelet Analysis.

Torrence, C. and Webster, P. J. (1999). Interdecadal Changes in the ENSOMonsoon System. *Journal of Climate*, 12(8):2679–2690.

Truscott, J., Fraser, C., Cauchemez, S., Meeyai, A., Hinsley, W., Donnelly, C. a., Ghani, A., and Ferguson, N. (2012). Essential epidemiological mechanisms underpinning the transmission dynamics of seasonal influenza. *Journal of the Royal Society, Interface / the Royal Society*, 9(67):304–12.

Van Cauteren, D., Vaux, S., de Valk, H., Le Strat, Y., Vaillant, V., and Lévy-Bruhl, D. (2012). Burden of influenza, healthcare seeking behaviour and hygiene measures during the A(H1N1)2009 pandemic in France: a population based study. *BMC public health*, 12:947.

Van-Tam, J. and Sellwood, C. (2010). *Seasonal Influenza: Epidemiology, Clinical Features and Surveillance*. CABI.

Viboud, C., Alonso, W. J., and Simonsen, L. (2006a). Influenza in tropical regions. *PLoS medicine*, 3(4):e89.

Viboud, C., Bjø rnstad, O. N., Smith, D. L., Simonsen, L., Miller, M. A., and Grenfell, B. T. (2006b). Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science (New York, N.Y.)*, 312(5772):447–51.

Viboud, C., Boelle, P. Y., Cauchemez, S., Lavenu, A., Valleron, A. J., Flahault, A., and Carrat, F. (2004a). Risk factors of influenza transmission in households. *Br J Gen Pract*, 54(506):684–689.

Viboud, C., Boëlle, P.-Y., Pakdaman, K., Carrat, F., Valleron, A.-J., and Flahault, A. (2004b). Influenza epidemics in the United States, France, and Australia, 1972-1997. *Emerging infectious diseases*, 10(1):32–9.

Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M., and Kawaoka, Y. (1992). Evolution and ecology of influenza A viruses. *Microbiological reviews*, 56(1):152–79.

Wenger, J. B. and Naumova, E. N. (2010). Seasonal Synchronization of Influenza in the United States Older Adult Population. *PLoS ONE*, 5(4):e10187.

WHO (1999). WHO recommended surveillance standards, Second edition.

Willem, L., Van Kerckhove, K., Chao, D. L., Hens, N., and Beutels, P. (2012). A nice day for an infection? Weather conditions and social contact patterns relevant to influenza transmission. *PloS one*, 7(11):e48695.

Wilschut, J. C., Wilschut, J., McElhaney, J. E., and Palache, A. M. (2006). *Influenza*. Mosby Elsevier.

Wong, C. M., Chan, K. P., Hedley, A. J., and Peiris, J. S. (2004). Influenza-associated mortality in Hong Kong. *Clin Infect Dis*, 39(11):1611–1617.

Wong, C. M., Yang, L., Chan, K. P., Leung, G. M., Chan, K. H., Guan, Y., Lam, T. H., Hedley, A. J., and Peiris, J. S. M. (2006). Influenza-associated hospitalization in a subtropical city. *PLoS medicine*, 3(4):e121.

Wu, P., Goldstein, E., Ho, L. M., Yang, L., Nishiura, H., Wu, J. T., Ip, D. K. M., Chuang, S.-K., Tsang, T., and Cowling, B. J. (2012). Excess mortality associated with influenza A and B virus in Hong Kong, 1998-2009. *The Journal of infectious diseases*, 206(12):1862–71.

Yaari, R., Katriel, G., Huppert, A., Axelsen, J. B., and Stone, L. (2013). Modelling seasonal influenza: the role of weather and punctuated antigenic drift. *Journal of the Royal Society, Interface / the Royal Society*, 10(84):20130298.

Yang, L., Ma, S., Chen, P. Y., He, J. F., Chan, K. P., Chow, A., Ou, C. Q., Deng, A. P., Hedley, A. J., Wong, C. M., and Peiris, J. S. M. (2011). Influenza associated mortality in the subtropics and tropics: results from three Asian cities. *Vaccine*, 29(48):8909–14.

Yang, L., Wong, C. M., Chan, K. P., Chau, P. Y. K., Ou, C. Q., Chan, K. H., and Peiris, J. S. M. (2009). Seasonal effects of influenza on mortality in a subtropical city. *BMC infectious diseases*, 9:133.

Yang, L., Wong, C. M., Lau, E. H. Y., Chan, K. P., Ou, C. Q., and Peiris, J. S. M. (2008). Synchrony of clinical and laboratory surveillance for influenza in Hong Kong. *PloS one*, 3(1):e1399.

Yang, W., Elankumaran, S., and Marr, L. C. (2012). Relationship between humidity and influenza A viability in droplets and implications for influenza's seasonality. *PloS one*, 7(10):e46789.

Yang, W. and Marr, L. C. (2011). Dynamics of airborne influenza A viruses indoors and dependence on humidity. *PloS one*, 6(6):e21481.

Yu, H., Alonso, W. J., Feng, L., Tan, Y., Shu, Y., Yang, W., and Viboud, C. (2013). Characterization of Regional Influenza Seasonality Patterns in China and Implications for Vaccination Strategies: Spatio-Temporal Modeling of Surveillance Data. *PLoS Medicine*, 10(11):e1001552.

Zaman, R. U., Alamgir, A. S. M., Rahman, M., Azziz-Baumgartner, E., Gurley, E. S., Sharker, M. A. Y., Brooks, W. A., Azim, T., Fry, A. M., Lindstrom, S., Gubareva, L. V., Xu, X., Garten, R. J., Hossain, M. J., Khan, S. U., Faruque, L. I., Ameer, S. S., Klimov, A. I., Rahman, M., and Luby, S. P. (2009). Influenza in outpatient ILI case-patients in national hospital-based surveillance, Bangladesh, 2007-2008. *PloS one*, 4(12):e8452.

Appendix A

The routine surveillance system in Viet Nam

Title: The routine surveillance system in Viet Nam

Source: Viet Nam Atlas of Communicable Diseases

Author(s): Authors Nguyen Tran Hien National Institute of Hygiene and Epidemiology, Ha Noi, Viet Nam

Nguyen Van Binh General Department of Preventive Medicine, Ha Noi, Viet Nam

Pham Quang Thai National Institute of Hygiene and Epidemiology, Ha Noi, Viet Nam

Nguyen Thi Thanh Thuy Wellcome Trust Major Overseas Program, Oxford University Clinical Research Unit, National Hospital of Tropical Diseases, Ha Noi, Viet Nam;

Peter Horby Wellcome Trust Major Overseas Program, Oxford University Clinical Research Unit, National Hospital of Tropical Diseases, Hanoi, Vietnam; Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, Oxford University, Oxford, United Kingdom

Heiman F.L. Wertheim Wellcome Trust Major Overseas Program, Oxford University Clinical Research Unit, National Hospital of Tropical Diseases, Hanoi, Vietnam; Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, Oxford University, Oxford, United Kingdom

COMMUNICABLE DISEASE SURVEILLANCE SYSTEM IN VIETNAM

The Vietnamese communicable disease surveillance system is running nationwide under the responsibility of the General Department of Preventive Medicine - Ministry of Health (GDPM; Circular No 48 /2010/TT-BYT of Ministry of Health; 31st December 2010). In 2013 there are 28 communicable diseases under surveillance (see table 1). The list for disease surveillance is decided by MOH and can be changed depending on new developments or emergence of communicable diseases and the expanded program of immunization (EPI; table 2).

Table 1. List of reportable communicable diseases:

A. List of communicable diseases that need to be reported weekly

No	Name of Disease	Group*	Code by ICD-10
1.	Cholera	A	A00
2.	Typhoid	B	A01
3.	Dengue	B	A90/A91
4.	Viral Encephalitis	B	A83
5.	Malaria	B	B50
6.	Hand, Foot, and Mouth disease	B	B08.4
7.	Meningococcal Meningitis	B	A39
8.	Measles	B	B05
9.	Influenza A(H5N1)	A	J09
10.	Severe respiratory infection caused by virus	A	
11.	Dangerous emerging disease with unknown pathogen	A	

B. List of communicable diseases that need to be reported monthly

No	Name of Disease	Group*	Code by ICD-10
1.	Cholera	A	A00
2.	Typhoid	B	A01
3.	Dysenteria	B	A03
4.	Amebiasis	B	A06
5.	Diarrhea	B	A09
6.	Viral Encephalitis	B	A83
7.	Dengue	B	A90/A91
8.	Malaria	B	B50
9.	Viral hepatitis	B	B15
10.	Rabies	B	A82
11.	Meningitis syndrome	B	A39
12.	Varicella	B	B01
13.	Diphtheria	B	A36
14.	Pertussis	B	A37
15.	Neonatal tetanus	B	A33
16.	Other tetanus (not neonatal tetanus)	B	A35
17.	AFP- polio suspected case	A	A80
18.	Measles	B	B05
19.	Mumps	B	B26
20.	Rubella	B	B06
21.	Influenza (seasonal)	B	J10,11
22.	Influenza A(H5N1)	A	J09
23.	Adenovirus pharyngoconjunctivitis (APC)	B	B30
24.	Plague	A	A20
25.	Anthrax	B	A22
26.	Leptospirosis	B	A27
27.	Hand, foot and mouth disease	B	B08.4
28.	Streptococcosis suis	B	B95

*Communicable diseases are classified into three groups: A, B and C. **Group A**: Very dangerous, can spread rapidly, and has a high mortality rate or is caused by an unknown pathogen. **Group B** includes dangerous pathogens that can transmit quickly and can result in death. **Group C** includes less dangerous pathogens with either low transmission or rarely leads to death.

All administrational levels (from commune to national) are responsible for collecting surveillance data and writing reports. The reporting can be performed by fax, phone,or email.The data must be submitted to the upper levelin weekly or monthly reports, depending on the type of disease (see table 1). The surveillance reports include the following aggregated information per disease: the number of new patients, the number of deaths, the cumulativecase and death and intervention has been performed. In particularepidemics, case investigation reports can be submitted (eg: from district to NIHE provincial PMC).Although the majority of communicable diseases are recorded by the hospital system, the community network still plays an important role in early detection of diseases and outbreaks. When a potential outbreak is detected of a reportable disease, an alert goes out to the commune health centers (CHC) to raise awareness.

As shown in the figure of the reporting system, there are twodirections of reporting: reports and information exchange (including disease alerts). Weekly reports will be used for rapid response, usually for outbreak verification and disease control. Rapid response teams are in charge of data analysis and reporting and to provide feedback to outbreak region. Monthly data is mostly used for annual reporting, and to calculate a threshold for outbreak.

A weakness of the present surveillance system is identifying the different reportable diseases correctly, asmost are confirmed clinicallyusing case definitions(see list of case definitions below). Without laboratoryconfirmation, the case definition is not systematically applied to the whole health care system and is not standardized. Besides non-standard case definitions, also the quality of data provided by communes is often poor due to data entry errors. Sentinel surveillance projects can provide more accurate data of diseases like is done for influenza.Despite all the flaws, the surveillance system remains a crucial data source on the situation of communicable diseases in Vietnam. In the near future, a web-based surveillance system will be launched in order to support and improve surveillance activities.

List of case definitions of notifiable diseases.

1. Cholera

Multiple watery stools

Rice-water stool

Vomiting (frequent)

Signs of rapid dehydration

2. Typhoid and paratyphoid fever

High fever 39-40^{o}C for 3-5 days

Severe headache

Constipation or diarrhea

Abdominal distension and tenderness

3. Dysentery syndrome

Abdominal cramp

Tenesmus
Multiple loose stools with blood and mucus

4. Diarrhoea

Loose stools ≥ 3 times per day
Very loose stools or watery stools

5. Viral meningitis

Sudden onset of high fever 39-40^{o}C
Headache
Disorderly movement
Confusion

6. Dengue fever, haemorrhagic dengue fever

High fever above 38^{o}C for 2-7 days
Headache, muscle and joint pain, periorbital pain
Congestion, skin rash
Signs of bleeding
Signs of shock

7. Viral hepatitis

Sudden onset of fatigue, malaise
Anorexia, nausea, abdominal discomfort, lower abdominal pain (upper right quadrant)
Jaundice, discolored stool, dark urine

8. Rabies

Pain along the nerve near the site of animal bite
Agitated
Afraid of water (hydrophobia), wind, light, noise
Increased salivation, difficulty to swallow, delirium, convulsion
Rapid progression and death

9. Meningococcal Meningitis

Sudden onset of high fever
Severe headache
Nausea and vomiting
Stiff neck
Possiblehaemorrhagiclesions

10. Chickenpox/varicella

Mild fever
Begins with red lesions/rash, after few hours developing to shallow blisters, after 1-2 days becoming yellow pustules.
Scattered lesions, predominantly on the scalp, different stages
Itching

11. Diphtheria

Sore throat/pharyngitis, inflammation of tonsil or larynx

Red throat, dysphagia

Pseudomembrane in pharynx, tonsil, larynx, nose

Greyish white cover attached to mucous membrane, cause bleeding when peeled off

12. Whooping cough

Persistent coughing more than 2 weeks

Paroxysmal cough, with episodes of cyanosis and ceasing breathing after a period of intense coughing

'Whoop' sound with sharp intake of breath after a coughing episode

Vomit after coughing

After each episode of coughing, the child is extremely tired, sweating and breathing rapidly

13. Neonatal tetanus

The newborn infant has normal breastfeeding (cry and suck) in the first 2 days after birth

From the 3rd-28th days, inability to nurse (cannot take suck breastfeeding)

Spasm or convulsion when stimulated with light, noise or touch

Signs of spasm/convulsion: stiff jaws, convulsion in arms and legs, tightened lips, bending back (opisthotonos)

Death occurs after 7-14 days after acquired the disease

14. Other tetanus

Painful muscular contractions in the face, neck, trunk

Abdominal rigidity

Generalized spasm occurs when induced by sensory stimuli

Typical features of the tetanic spasm are the position of opisthotonos and the facial expression known as "risussardonicus"

15. Acute Flaccid Paralysis (AFP)

Flaccid paralysis (flaccid muscles, muscle weakness or loss of movement ability) suddenly appear within 1 week in children of less than 15 years old.

- Confirmed poliomyelitis: is AFP with confirmed isolated wild polio virus

- Suspected poliomyelitis: is AFP but unable to obtain or test stool

16. Suspected measles

Fever, with at least one of the following symptoms: coughing, runny nose, conjunctivitis, rash

Confirmed measles diagnosis:

- ***Confirmed lab diagnosis:*** The suspected case has IgM (+) antibody or isolated measles virus

- ***Confirmed epidemiological diagnosis:*** The suspected case has epidemiological exposure with measles cases with confirmed IgM (+) antibody during the incubation period of 7-14 days

Clinical diagnosis:no laboratory confirmation.

17. Mumps

Fever, swelling and tenderness in one or multiple salivary glands. The skin is not red.

18. Influenza

Sudden onset of fever:39-40^{o}C

Severe headache, body, muscle and joint pain

Runny nose, sore throat, coughing

19. APC (Adenoviruspharyngoconjunctivitis)

Conjunctivitis (red eyes)

Pharyngitis

Lymphadenopathy behind parotis and below jaws

20. Plague

Sudden onset of high fever

Headache, malaise

Bubonic plague: swollen lymph nodes, which are inflamed, red, tender, and often occur in the inguinal or axillary areas, or neck (cervical)

Pneumonic plague: Coughing with pus and blood, chest pain, difficulty breathing

21. Anthrax

Cutaneous anthrax: initial itching of the affected site, followed by a lesion that becomes papular, then vesicular, developing in 2-6 days into a depressed black eschar. Moderate to severe and very extensive edema surrounds the eschar.

Inhalation anthrax: Initial symptoms are similar to acute respiratory inflammation with fever, cough, chest pain, difficulty breathing, shock after 2-3 days leading to death.

Digestive/gastrointestinal anthrax: Nausea, vomit, anorexia, severe abdominal pain, accompanied with fever, followed by signs of septicemia and death.

22. Leptospirosis

Sudden onset of high fever, headache, chills, malaise, myalgia (specially in calves and thighs)

Conjunctivaleffusion

Renal failure

Arrhythmia

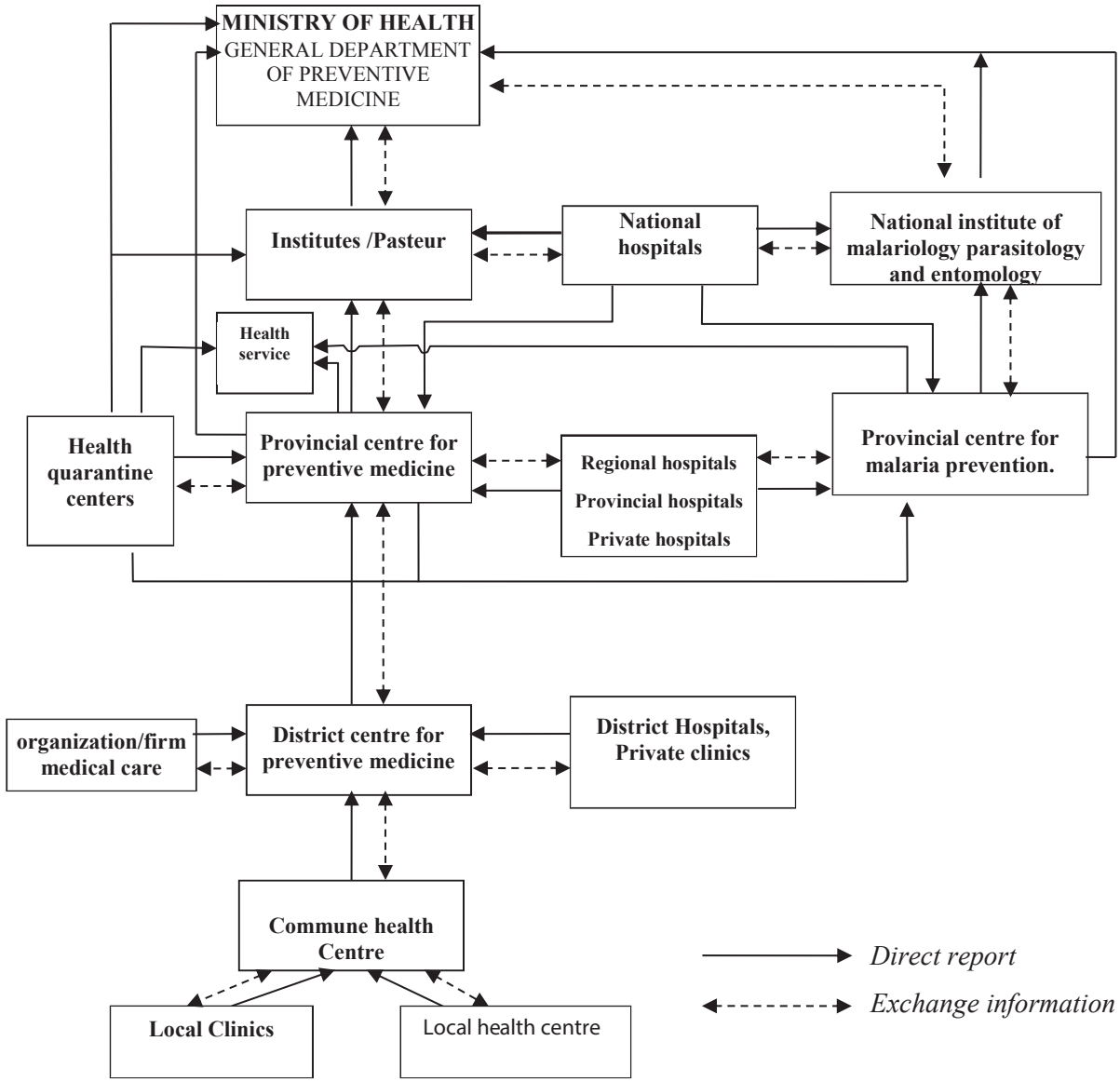
Jaundice

Rash

Table 2. Extended Immunization Program (EPI) in Vietnam (see also Vaccination coverage map)

Vaccine	Target population	Location
Japanese encephalitis (JE)	12 months; + 2 weeks; 2 years	High risk area
Bacille Calmette-Guérin vaccine (BCG)	birth	National
Oral polio vaccine (OPV)	2, 3, 4 months	National
Hepatitis B vaccine	birth	National
Measles vaccine	9,18 months	National
Tetanus toxoid	pregnant women; +1, +6 months; +1 year	National
Cholera	2-5 years	High risk area
Diphtheria, Tetanus and Pertussis (DTP)	2, 3, 4 months	National
Typhoid fever vaccine	3-10 years	High risk area
Haemophilus influenzae b (Hib) [1]	2, 3, 4 months	National

Figure. Data flow of surveillance reporting of communicable diseases to Ministry of Health level.
Based on circular No 48 /2010/TT-BYT of Ministry of Health dated *31 December 2010*



Appendix B

Supplementary for chapter 3 and 4

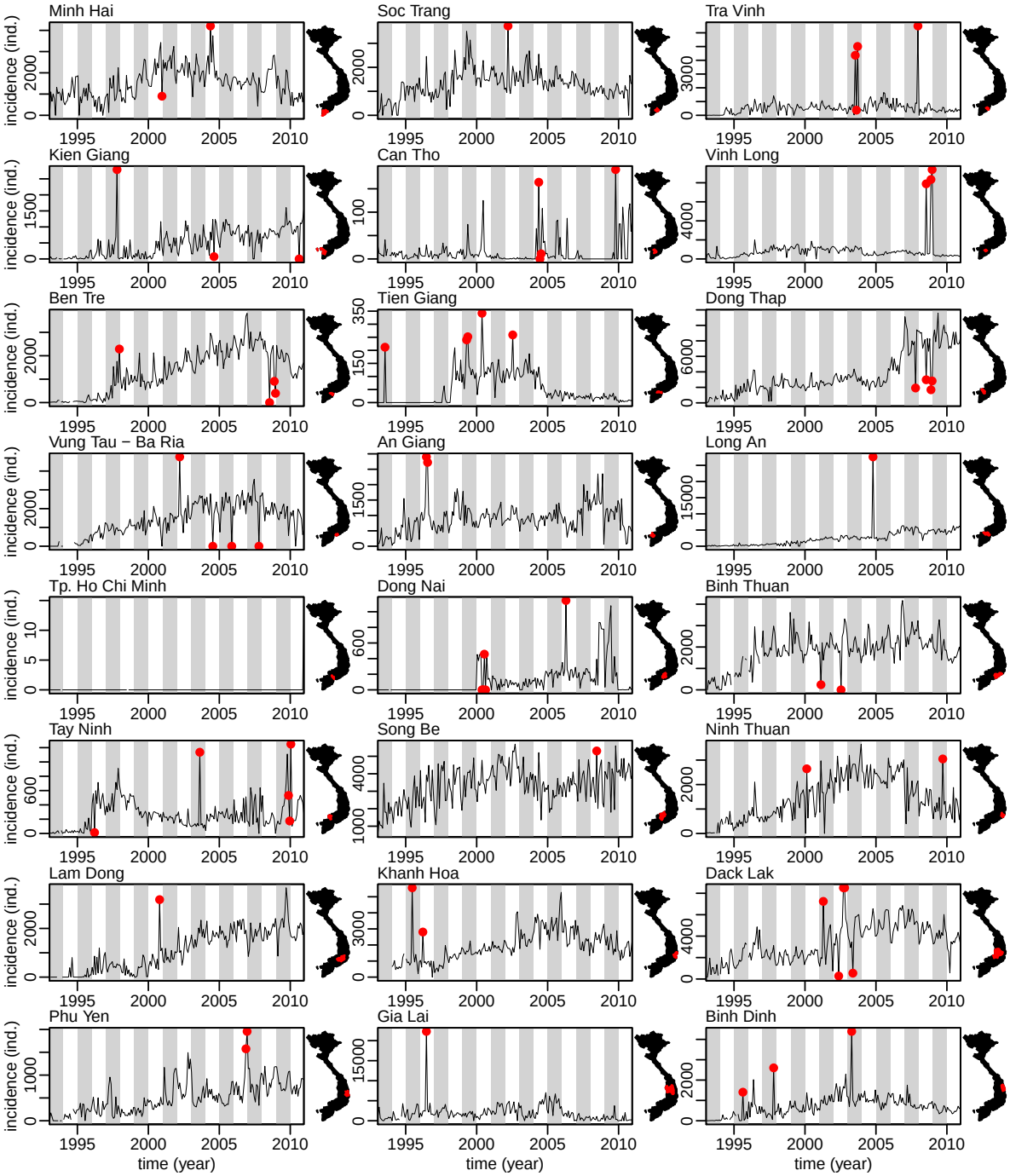


Figure B.1: Raw data. Red dots show identified outliers with criterion $f = 0.01$. Little maps show the populations' centroids of the provinces (red dots).

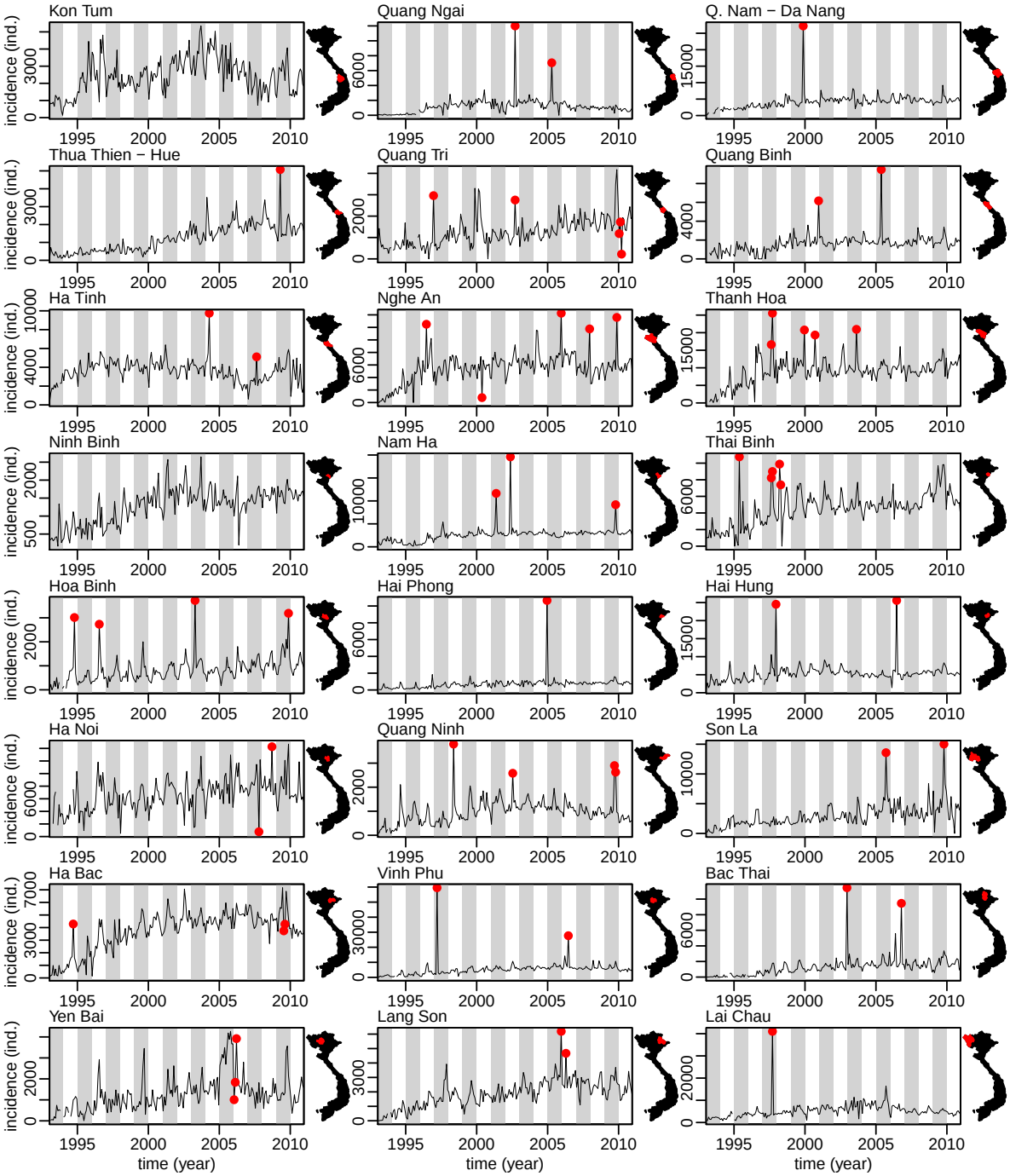


Figure B.2: Raw data, cont'd.

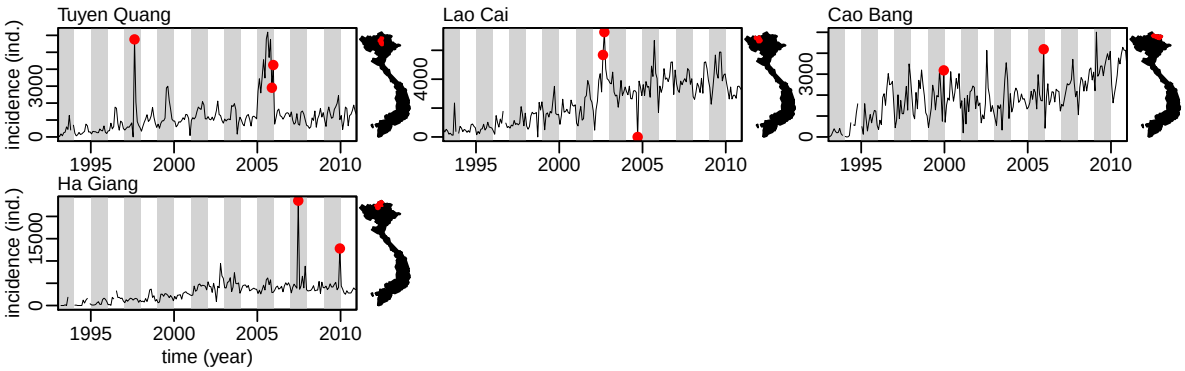


Figure B.3: Raw data, cont'd.

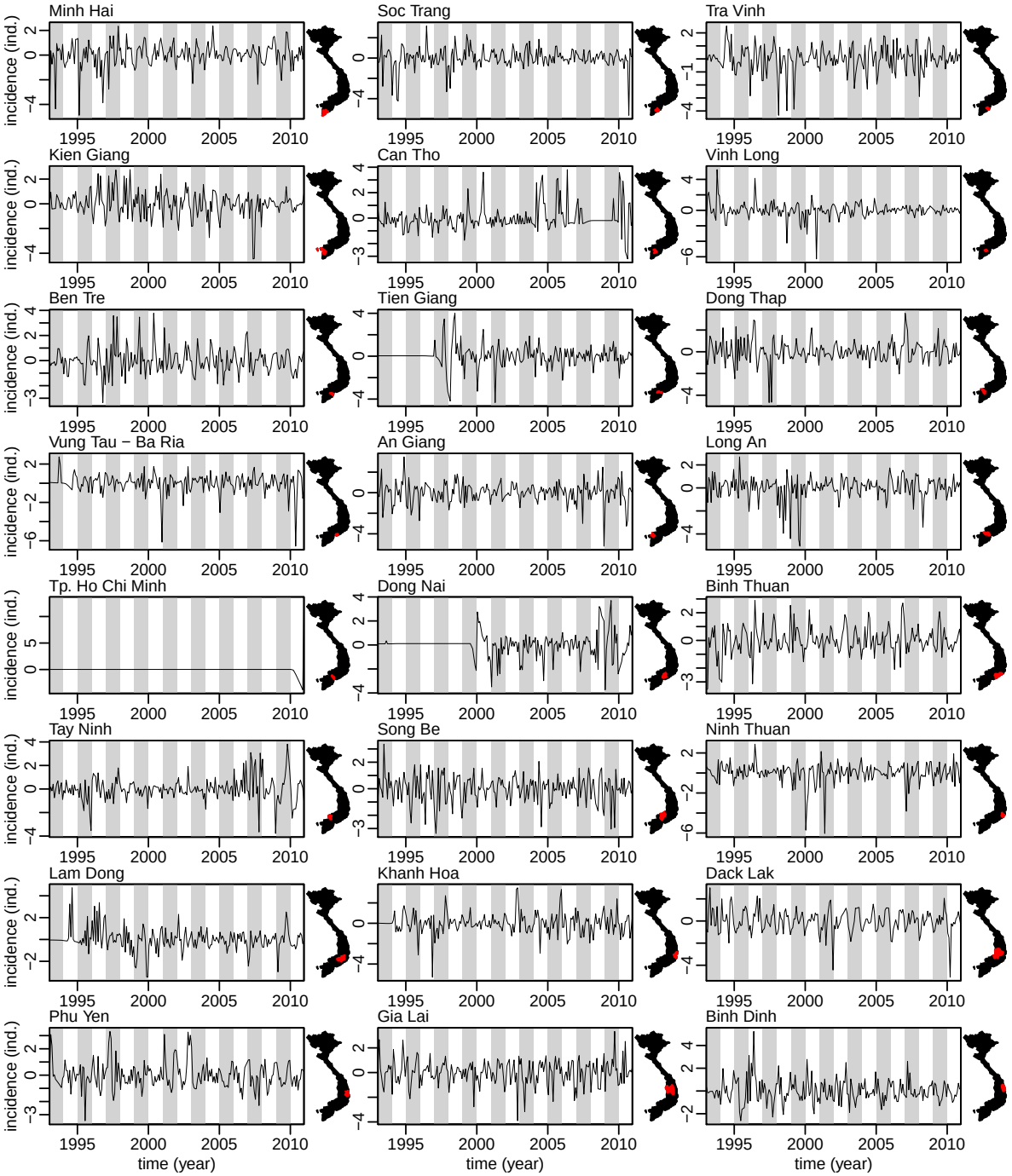


Figure B.4: Transformed data. Outliers ($f = 0.01$) are discarded, missing values linearly interpolated, data square-root transformed before being detrended and scaled.

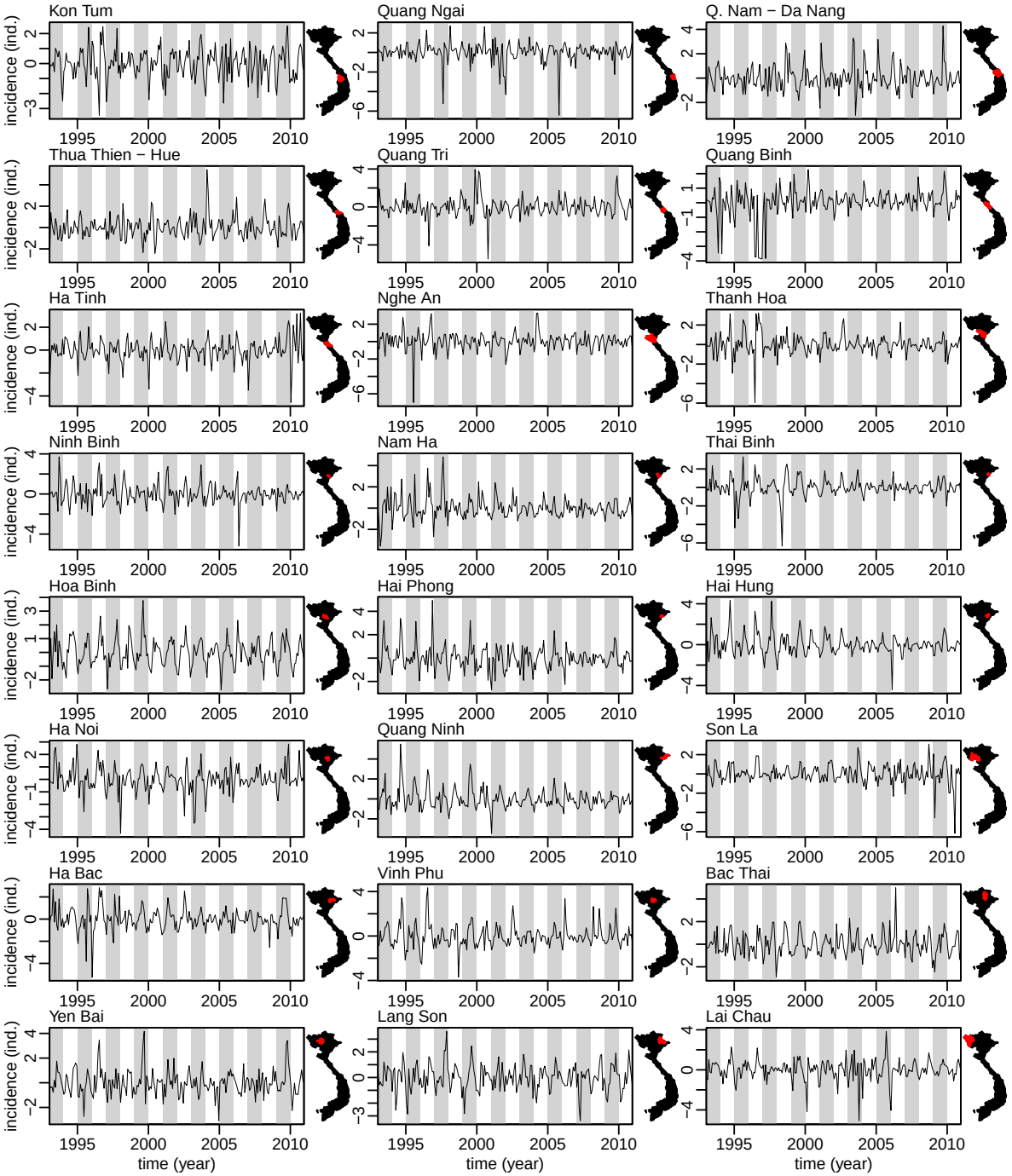


Figure B.5: Transformed data, cont'd.

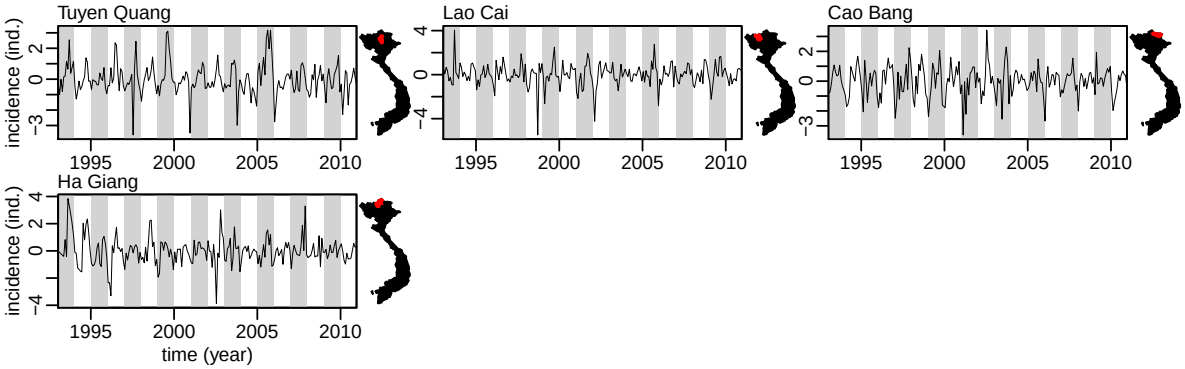


Figure B.6: Transformed data, cont'd.

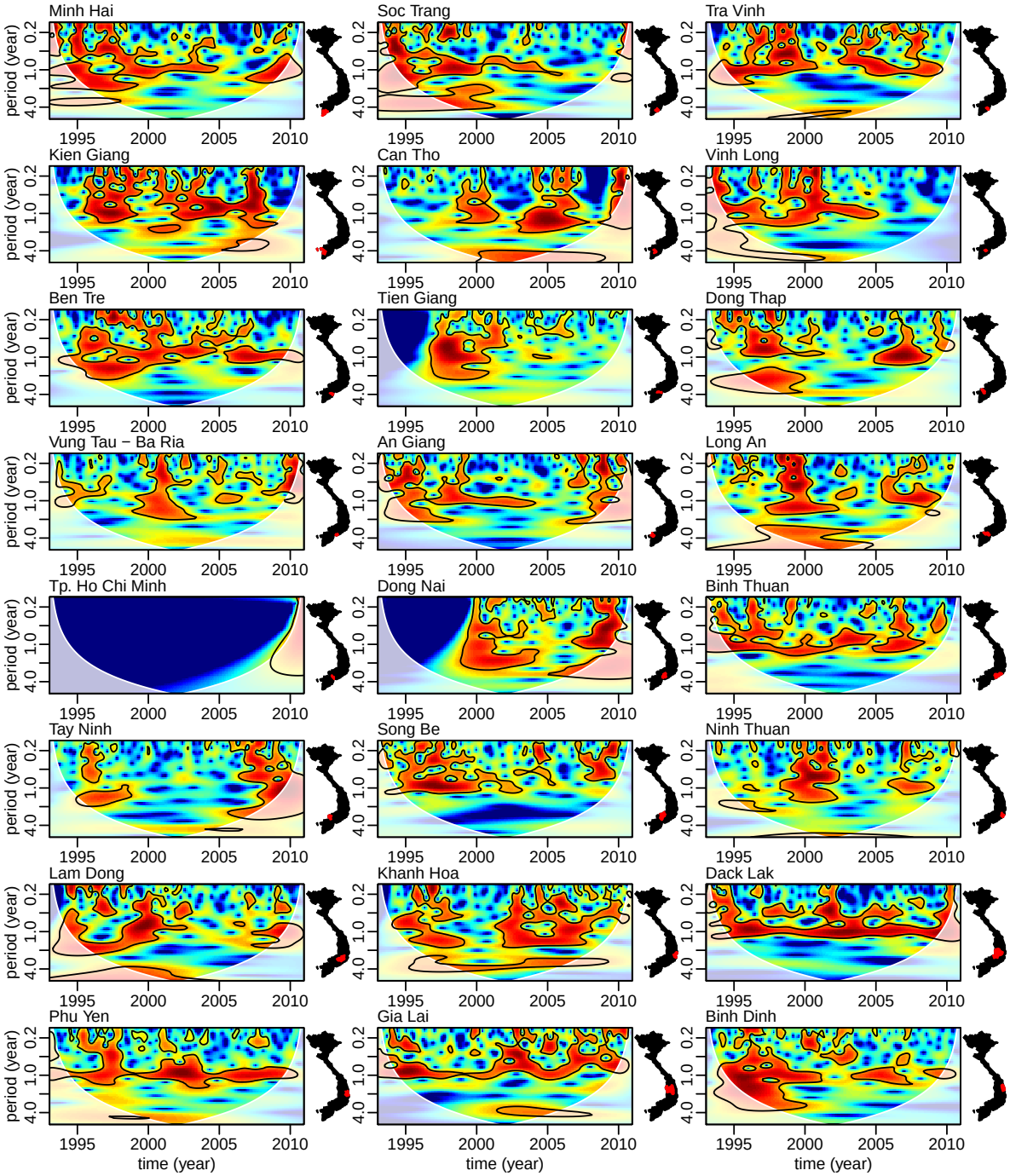


Figure B.7: Wavelet transform. Outliers ($f = 0.01$) are discarded, missing values linearly interpolated, data square-root transformed before being detrended and scaled.

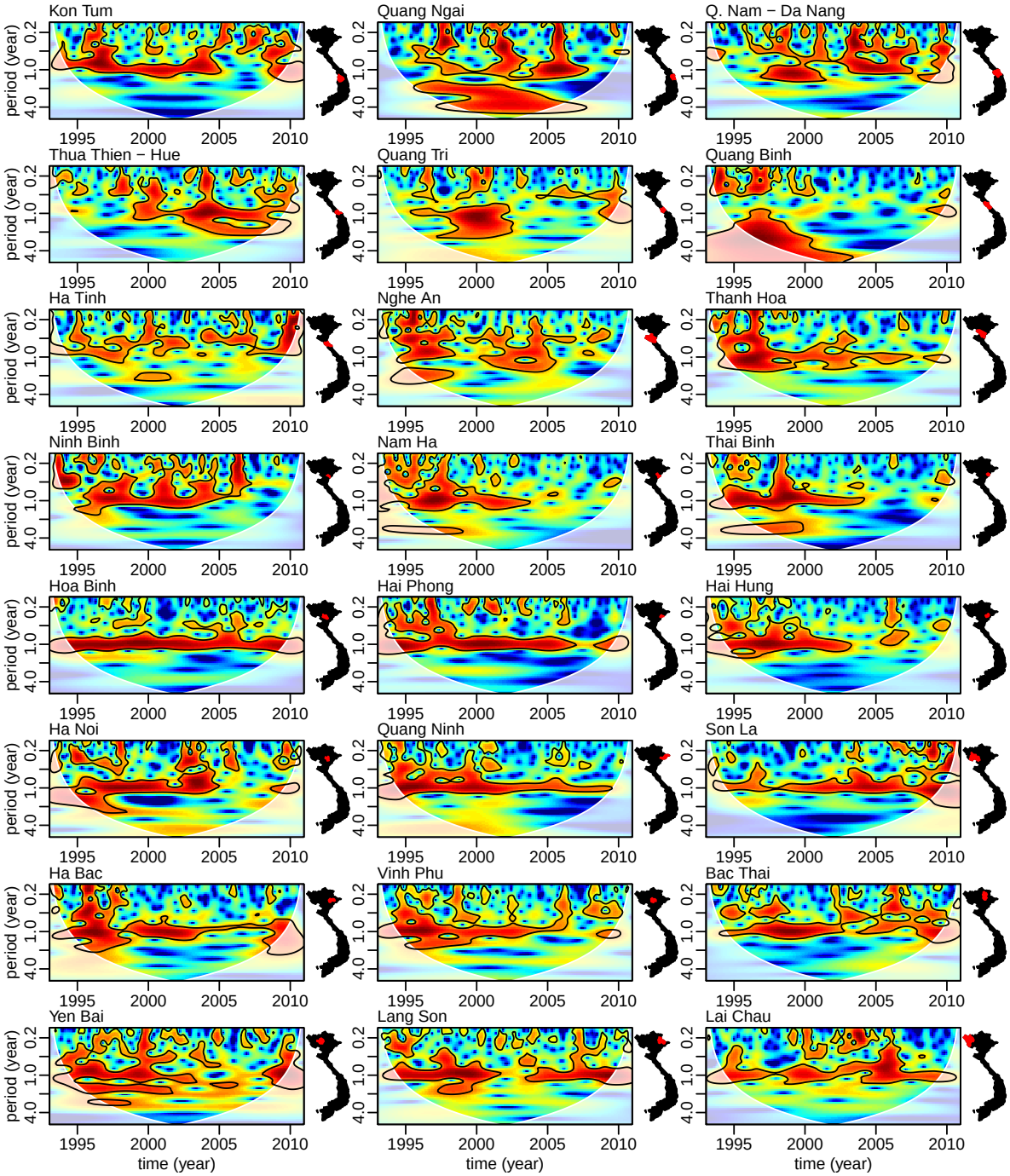


Figure B.8: Wavelet transform, cont'd.

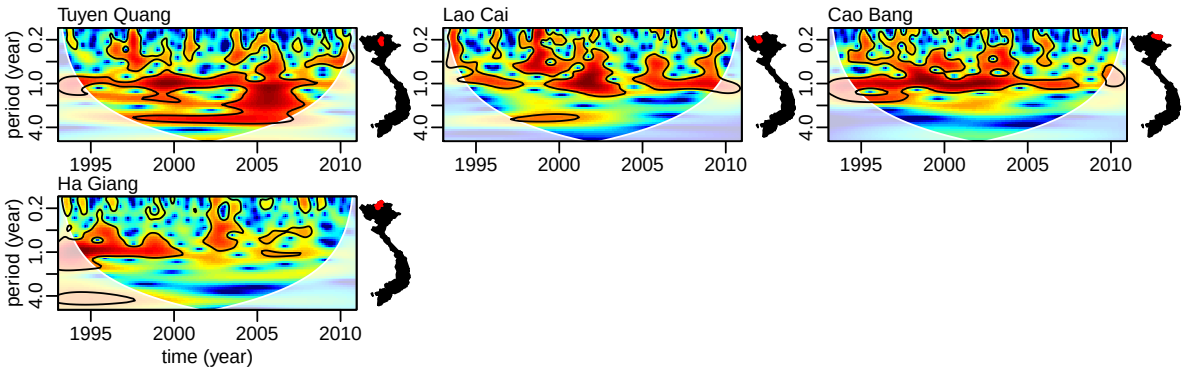


Figure B.9: Wavelet transform, cont'd.

Appendix C

Ha Nam cohort supplementary

Figure C.1: The Thanh Ha commune health worker

Figure C.2: The OUCRU Ha Noi

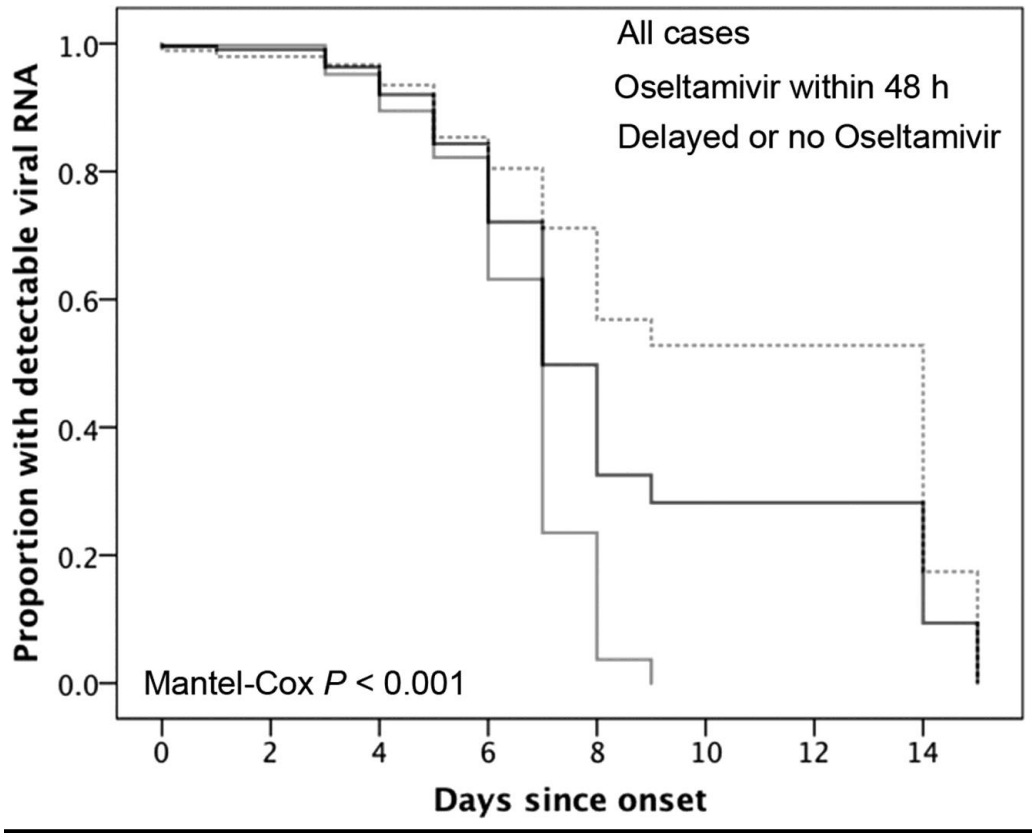


Figure C.3: KaplanMeier curves of time until cessation of viral RNA shedding in virologically confirmed cases.
A P value is shown for the comparison of cases who took timely Oseltamivir (n = 17) versus those who didn't take Oseltamivir or who took Oseltamivir late (n = 16)

Supplementary Table 1. HI and MN antibody titers in pre-pandemic, acute infection and post-pandemic sera for 81 people from index case households

	HI titer [a]		MN titer [a]		
Participant Infection Status	**pre-pandemic**	**post-pandemic**	**pre-pandemic**	**post-pandemic**	**convert**
index	5	160	5	960	yes
index	5	160	5	320	yes
index	5	320	5	640	yes
index	5	80	5	320	yes
index	5	80	5	120	yes
index	5	40	5	160	yes
index	5	40	5	80	yes
index	5	40	5	40	yes
index	5	40	5	80	yes
index	5	40	5	160	yes
index	5	40	5	160	yes
index	5	20	5	80	yes
index	5	*5*	5	320	yes
index	5	*5*	5	40	yes
index	5	*5*	5	320	yes
index	5	*5*	5	40	yes
index	-	-	-	-	
index	-	-	-	-	
index	-	-	-	-	
index	-	-	-	-	
index	5	-	5	-	
index	5	-	5	-	
secondary	5	40	5	240	yes
secondary	5	320	5	640	yes
secondary	5	160	5	640	yes

secondary	5	40	5	160	yes
secondary	5	-	5	-	
secondary	5	-	5	-	
secondary asymptomatic	5	10	5	10	no
secondary asymptomatic	5	320	5	640	yes
secondary asymptomatic	5	100	15	480	yes
secondary asymptomatic	5	10	5	60	yes
secondary asymptomatic	5	-	5	-	
RT-PCR negative	5	80	5	320	yes
RT-PCR negative		40		120	yes
RT-PCR negative	5	20	5	80	yes
RT-PCR negative	5	5	5	80	yes
RT-PCR negative	5	5	5	160	yes
RT-PCR negative	5	5	5	120	yes
RT-PCR negative	20	20	40	40	no
RT-PCR negative	5	5	5	5	no
RT-PCR negative	5	5		-	no
RT-PCR negative	5	5	10	5	no
RT-PCR negative	5	5	15	20	no
RT-PCR negative	5	5	5	5	no
RT-PCR negative	5	5	-	5	no
RT-PCR negative	5	5	5	5	no
RT-PCR negative	5	5	5	5	no
RT-PCR negative	5	5	5	5	no
RT-PCR negative	5	5	-	5	no
RT-PCR negative	5	5	5	5	no
RT-PCR negative	5	5	5	5	no
RT-PCR negative	5	5	5	5	no
RT-PCR negative	5	5	15	10	no
RT-PCR negative	5	5	-	5	no
RT-PCR negative	5	5	-	-	no

RT-PCR negative	5	5	-	-	no
RT-PCR negative	5	5	-	-	no
RT-PCR negative	5	5	20	20	no
RT-PCR negative	5	5	5	5	no
RT-PCR negative	5	5	5	5	no
RT-PCR negative	5	5	8	5	no
RT-PCR negative	5	5		-	no
RT-PCR negative	5	5	5	5	no
RT-PCR negative	5	5	5	5	no
RT-PCR negative	5	5	5	5	no
RT-PCR negative	5	5	5	5	no
RT-PCR negative	5	5	5	5	no
RT-PCR negative	5	5	5	10	no
RT-PCR negative	5	5	-	5	no
RT-PCR negative	-	40	-	-	
RT-PCR negative	--	-	-	-	
RT-PCR negative	-	-	-	-	
RT-PCR negative	5	-	-	-	
RT-PCR negative	-	-	-	-	
RT-PCR negative	-	-	-	-	
RT-PCR negative	-	-	-	-	
RT-PCR negative	-	-	-	-	
RT-PCR negative	5	-	-	-	
RT-PCR negative	5	-	-	-	
RT-PCR negative	5	-	-	-	

a: reciprocal titers are presented as the mean of two measurements

Appendix D

R code

```
source("plot.biwavelet.r")
#tmp <- doXprovince(function(x)scaling(detrend(transform(interpolate(
    discardoutliers(x,perc=.01)),function(x)sqrt(x)),ff=.1)),ili93)


countnbna <- function(data=ili,rules=mergings) {
# This function counts the number of consecutive missing values per
# province, per starting year, as well as the total number of missing
# values.
# data : output of the "readili" function.
# rules: output of the "readmergings" function.
    fct <- function(y) {
        print(y)
        out <- merge_ili(y,data,rules)
        return(sapply(unique(out$province),function(x) {
            tmp <- subset(out,province==x,cases,T)
            return(c(max(c(0,consec.na(tmp))),
                sum(!is.na(tmp))))
        }))
    }
    time <- sort(unique(data$time))
    time <- time[-length(time)] # We need to remove the last time.
    return(list(time,sapply(time,fct)))
}


plot.nb.provinces <- function(data=start_yr_effect,merg=mergings,
    multiple=F) {
# This function plots the number of provinces as a funtion of cutting year
# when the number of consecutive missing values for the exclusion of
```

```
# provinces varies.
# data:     output of the "countnbna" function.
# merg:     output of the "readmergings" function.
# multiple: (boolean) says whether this graph is include in a
#           multiple-panel one.
    time <- data[[1]]
    data <- data[[2]]
    data <- lapply(data,function(x)x[1,])
    plot(time,sapply(data,function(x)sum(x<7)),type="n",
        col="lightgrey",xlab="",
        ylab="number of time series",ylim=c(0,63),axes=F)
    if(!multiple) {axis(1); title(xlab="starting year")}
    axis(2); box()
    for(i in 8:(length(time)+1)) lines(time,sapply(data,function(x)
        sum(x<i)),col="lightgrey")
    for(i in 1:7) lines(time,sapply(data,function(x)sum(x<i)))
    abline(v=c(unique(merg$dates),2008.042),lty=2)
}




plot.nb.data <- function(data=start_yr_effect,merg=mergings) {
# This function plots the number of non-missing data as a funtion of
# cutting year when the number of consecutive missing values for the
# exclusion of provinces varies.
# data:     output of the "countnbna" function.
# merg:     output of the "readmergings" function.
    time <- data[[1]]
    data <- data[[2]]
    timepoints <- length(time)
    out <- sapply(1:timepoints,function(y)sapply(data,function(x)
        sum(x[2,x[1,]<y])))
    plot(time,out[,1],type="n",xlab="starting year",ylim=c(0,12100),
        ylab="number of non-missing data")
    for(i in 8:timepoints) lines(time,out[,i],col="lightgrey")
    for(i in 1:7) lines(time,out[,i])
```

```
67      abline(v=c(unique(merg$dates),2008.042),lty=2)
68  }
69
70
71
72  plot.start.year.effect <- function(data=start_yr_effect,merg=mergings,
73      x1=.12,x2=.99,y1=.1,y3=.99,a=2010,l=-1.5) {
74  #pdf("starting_year.pdf",width=2.75,height=3.5,pointsize=8.5)
75  #plot.start.year.effect(x1=.12,x2=.99,y1=.1,y3=.99,a=1981,l=-1.5); dev.off
        ()
76  # data:     output of the "countnbna" function.
77  # merg:     output of the "readmergings" function.
78      y2 <- y1+(y3-y1)/2
79      opar <- par(plt=c(x1,x2,y2,y3))
80      plot.nb.provinces(data,merg,T)
81      mtext("(A)",at=a,line=l)
82      par(plt=c(x1,x2,y1,y2),new=T)
83      plot.nb.data(data,merg)
84      mtext("(B)",at=a,line=l)
85      par(opar)
86  }
87
88
89  plot.split <- function(color="white") {
90  # This function plots the splitting history of provinces.
91  #pdf("splits.pdf",width=3.5,height=6.33,pointsize=8.5)
92  #par(plt=c(0,1,.055,.84)); plot.split(); dev.off()
93      require("ape") # for read.tree and plot.phylo
94      opar <- par(plt=c(0,1,.055,.84))
95      start <- 1980.042
96      end <- 2010.958
97      b <- 1992.042
98      e <- 2004.042
99      h <- 1997.042
100     k <- 2008.042
101     n <- 1991.042
102     q <- 1990.042
```

```
    a <- end - start
    c <- end - b; d <- b - start;
    f <- end - e; g <- e - start
    i <- end - h; j <- h - start
    l <- k - start; m <- l - c
    o <- end - n; p <- n - start
    r <- end - q; s <- q -start
    t <- c - f; u <- c - i
    v <- c - end + k
# This is the tree in Newick format:
    tree <- paste(
        "(Cao Bang:",a,
        ",(Ha Giang:",c,",Tuyen Quang:",c,")Ha Tuyen:",d,
        ",(Yen Bai:",c,",Lao Cai:",c,")Hoang Lien Son:",d,
        ",(Lai Chau 2:",f,",Dien Bien:",f,")Lai Chau 1:",g,
        ",Lang Son:",a,
        ",(Thai Nguyen:",i,",Bac Kan:",i,")Bac Thai:",j,
        ",(Vinh Phuc:",i,",Phu Tho:",i,")Vinh Phu:",j,
        ",(Bac Giang:",i,",Bac Ninh:",i,")Ha Bac:",j,
        ",Son La:",a,
        ",Quang Ninh:",a,
        ",(Hai Duong:",i,",Hung Yen:",i,")Hai Hung:",j,
        ",(Hoa Binh:",c,",Ha Tay:",v,")Ha Son Binh:",d,
        ",Ha Noi 2:",l,
        ",Hai Phong:",a,
        ",Thai Binh:",a,
        ",((Nam Dinh:",i,",Ha Nam:",i,")Nam Ha:",u,
            ",Ninh Binh:",c,")Ha Nam Ninh:",d,
        ",Thanh Hoa:",a,
        ",(Nghe An:",o,",Ha Tinh:",o,")Nghe Tinh:",p,
        ",(Quang Binh:",r,",Quang Tri:",r,",Thua Thien - Hue:",r,
            ")Binh Tri Thien:",s,
        ",(Quang Nam:",i,",Da Nang:",i,")Quang Nam - Da Nang:",j,
        ",(Quang Ngai:",r,",Binh Dinh:",r,")Nghia Binh:",s,
        ",(Gia Lai:",c,",Kon Tum:",c,")Gia Lai Kon Tum:",d,
        ",(Dak Lak:",f,",Dak Nong:",f,")Dack Lak:",g,
        ",(Phu Yen:",r,",Khanh Hoa:",r,")Phu Khanh:",s,
```

```
        ",Lam Dong:",a,
        ",(Binh Duong:",i,",Binh Phuoc:",i,")Song Be:",j,
        ",Tay Ninh:",a,
        ",(Binh Thuan:",c,",Ninh Thuan:",c,")Thuan Hai:",d,
        ",Dong Nai:",a,
        ",Tp. Ho Chi Minh:",a,
        ",Long An:",a,
        ",An Giang:",a,
        ",Vung Tau - Ba Ria:",a,
        ",Dong Thap:",a,
        ",Tien Giang:",a,
        ",Ben Tre:",a,
        ",(Tra Vinh:",c,",Vinh Long:",c,")Cuu Long:",d,
        ",((Can Tho 2:",f,",Hau Giang 2:",f,")Can Tho 1:",t,
            ",Soc Trang:",c,")Hau Giang 1:",d,
        ",Kien Giang:",a,
        ",(Ca Mau:",i,",Bac Lieu:",i,")Minh Hai:",j,");",sep="")
    tree <- read.tree(text=tree)
    attr(tree,"order") <- NULL
    tree$edge <- tree$edge[nrow(tree$edge):1,]
    tree$edge.length <- rev(tree$edge.length)
    plot(tree,show.tip.label=F,x.lim=c(-10,40),
        y.lim=c(0,65),yaxs="i",lwd=.25,plot=F)
    abline(v=c(10:12,17,24,28)+.042,col="lightgrey")
    par(new=T)
    plot(tree,show.tip.label=F,x.lim=c(-10,40),
        y.lim=c(0,65),yaxs="i",lwd=.25)
    title(xlab="time (year)")
    right <- rev(tree$tip.label)
    right <- gsub("([A-Z])"," \\1",right)
    right <- gsub("^ ","",right)
    right <- gsub("[0-9]","",right)
    right <- gsub("- "," - ",right)
    text(31.3,c(1:43,44.5,46:64),right[-45],cex=.75,adj=0)
    left <- c("Minh Hai","Kien Giang","Hau Giang","Cuu Long","Ben Tre",
        "Tien Giang","Dong Thap","Vung Tau - Ba Ria","An Giang",
        "Long An","Tp. Ho Chi Minh","Dong Nai","Thuan Hai",
```

```
177         "Tay Ninh","Song Be","Lam Dong","Phu Khanh","Dack Lak",
178         "Gia Lai - Kon Tum","Nghia Binh","Q. Nam - Da Nang",
179         "Binh Tri Thien","Nghe Tinh","Thanh Hoa","Ha Nam Ninh",
180         "Thai Binh","Hai Phong","Ha Noi","Ha Son Binh","Hai Hung",
181         "Quang Ninh","Son La","Ha Bac","Vinh Phu","Bac Thai",
182         "Lang Son","Lai Chau","Hoang Lien Son","Ha Tuyen",
183         "Cao Bang")
184     text(-.2,c(1.5,3,4.75,7.5,9:16,17.5,19,20.5,22,23.5,25.5,27.5,29.5,
185         31.5,34,36.5,38,39.75,42:44,45.5,47.5,49,50,51.5,53.5,55.5,
186         57,58.5,60.5,62.5,64),left,cex=.75,adj=1)
187     abline(v=0,col=color,lwd=3)
188     segments(2008.042-1980,44,2008.042-1980,45)
189     segments(2008.042-1980,44.5,end-1980,44.5)
190     axis(1,pretty(0:30),paste(pretty(1980:2010)))
191     par(plt=c(0,1,.84,1),new=T)
192     plot(c(1980,1990:1992,1997,2004,2008,2010),
193         c(40,44,45,53,61,64,63,63),type="n",ylim=c(0,65),
194         xlim=c(1970,2020),axes=F,ann=F)
195     abline(v=c(1990:1992,1997,2004,2008)+.042,col="lightgrey")
196     points(c(1980,1990:1992,1997,2004,2008,2010),
197         c(40,44,45,53,61,64,63,63),type="s")
198     axis(2,pos=1979.5)
199     title(ylab="number of provinces",line=-4)
200     par(opar)
201 }
202
203
204
205
206 plot_maps <- function() {
207 # This function plots the maps with the different provinces' definitions.
208 # provinces80, provinces90, provinces91, provinces92,
209 # provinces97, and provinces04 are output of the "mergeprovinces" function.
210 #pdf("maps.pdf",pointsize=8.5); plot_maps(); dev.off()
211     lwd <- .1
212     col1 <- "lightgrey"
213     col2 <- "red"
```

```
214     x1 <- 0; x2 <- 1
215     y1 <- 0; y2 <- 1
216     x <- 104; y <- 15
217     width <- (x2-x1)/3
218     height <- (y2-y1)/2
219     opar <- par(plt=c(x1,x1+width,y1+height,y1+2*height))
220     plot(provinces80_l,lwd=lwd,col=col1)
221     a <- which(is.element(sapply(1:length(provinces80_l@polygons),
222         function(x)provinces80_l@polygons[[x]]@ID),
223             c("Binh Tri Thien","Nghia Binh","Phu Khanh")))
224     for(i in a) plot(provinces80_l[i],col=col2,add=T,lwd=lwd)
225     text(x,y,"1980-1989")
226     par(plt=c(x1+width,x1+2*width,y1+height,y1+2*height),new=T)
227     plot(provinces90_l,lwd=lwd,col=col1)
228     a <- which(is.element(sapply(1:length(provinces90_l@polygons),
229         function(x)provinces90_l@polygons[[x]]@ID),
230             "Nghe Tinh"))
231     for(i in a) plot(provinces90_l[i],col=col2,add=T,lwd=lwd)
232     text(x,y,"1990")
233     par(plt=c(x1+2*width,x1+3*width,y1+height,y1+2*height),new=T)
234     plot(provinces91_l,lwd=lwd,col=col1)
235     a <- which(is.element(sapply(1:length(provinces91_l@polygons),
236         function(x)provinces91_l@polygons[[x]]@ID),c("Ha Tuyen",
237             "Hoang Lien Son","Ha Son Binh","Ha Nam Ninh",
238             "Gia Lai - Kon Tum","Thuan Hai","Cuu Long",
239             "Hau Giang")))
240     for(i in a) plot(provinces91_l[i],col=col2,add=T,lwd=lwd)
241     text(x,y,"1991")
242     opar <- par(plt=c(x1,x1+width,y1,y1+height),new=T)
243     plot(provinces92_l,lwd=lwd,col=col1)
244     a <- which(is.element(sapply(1:length(provinces92_l@polygons),
245         function(x)provinces92_l@polygons[[x]]@ID),c("Bac Thai",
246             "Vinh Phu","Ha Bac","Hai Hung","Nam Ha",
247             "Q. Nam - Da Nang","Song Be","Minh Hai")))
248     for(i in a) plot(provinces92_l[i],col=col2,add=T,lwd=lwd)
249     text(x,y,"1992-1996")
250     opar <- par(plt=c(x1+width,x1+2*width,y1,y1+height),new=T)
```

```
251     plot(provinces97_l,lwd=lwd,col=col1)
252     a <- which(is.element(sapply(1:length(provinces97_l@polygons),
253         function(x)provinces97_l@polygons[[x]]@ID),
254             c("Lai Chau","Dack Lak","Can Tho")))
255     for(i in a) plot(provinces97_l[i],col=col2,add=T,lwd=lwd)
256     text(x,y,"1997-2003")
257     opar <- par(plt=c(x1+2*width,x1+3*width,y1,y1+height),new=T)
258     plot(provinces04_l,lwd=lwd,col=col1)
259     a <- which(is.element(sapply(1:length(provinces04_l@polygons),
260         function(x)provinces04_l@polygons[[x]]@ID),
261             c("Ha Tay","Ha Noi")))
262     for(i in a) plot(provinces04_l[i],col=col2,add=T,lwd=lwd)
263     text(x,y,"2004-2007")
264     par(opar)
265 }
266
267
268
269
270 transform <- function(data=subset(ili93,province=="An Giang"),
271     transf=sqrt) {
272 # This function transforms the cases of one province.
273 # data  : one province of the output of "readili" or "merge_ili".
274 # transf: the function used for the transformation.
275     data$cases <- with(data,transf(cases))
276     return(data)
277 }
278
279
280 interpolate <- function(data=subset(ili93,province=="An Giang")) {
281 # This function linearly interpolates the missing values of one province.
282 # data: one province of the output of "readili" or "merge_ili".
283     data$cases <- with(data,approx(time,cases,time,rule=2)$y)
284     return(data)
285 }
286
287
```

```
288 detrend <- function(data=interpolate(),ff=.1) {
289 # This function detrends the data with a lowess smoother.
290 # data: one province of the output of "readili" or "merge_ili".
291 #       The cases of these data have to have no missing values.
292 # ff  : the smoother parameter of the "lowess" function.
293     data$cases <- with(data,cases-lowess(time,cases,f=ff)$y)
294     return(data)
295 }
296
297
298
299 scaling <- function(data=interpolate()) {
300 # This functions scales the data (centers and reduces).
301 # data: one province of the output of "readili" or "merge_ili".
302 #       The cases of these data have to have no missing values.
303     data$cases <- with(data,scale(cases))
304     return(data)
305 }
306
307
308
309 outliers <- function(data=scaling(detrend(interpolate())),perc=.01) {
310 # This function identifies outliers and return their index.
311 # data: one province of the output of "readili" or "merge_ili".
312 #       The cases of these data have to have no missing values.
313 #       It's better if it's also detrended and scaled.
314 # perc: the percentile above which to consider a data as an outlier.
315     foo <- diff(as.vector(data$cases))
316     perc <- perc/2
317     perc <- c(perc,1-perc)
318     bar <- qnorm(perc,mean(foo),sd(foo))
319     foo <- which(foo<bar[1] | foo>bar[2])
320     un <- which(diff(foo)==1)+1
321 # In case there are two outliers next to each other
322     deux <- which(diff(foo)==2)
323     return(c(foo[un],foo[deux]+1,foo[deux]+2))
324 }
```

```
325
326
327
328
329 plot1province <- function(data=subset(ili93,province=="An Giang"),
330     xlab=T,ylab=T,lwd=.5,perc=.01,wave=F) {
331 # This function plots the time series of 1 province.
332 # data: one province of the output of "readili" or "merge_ili".
333     a <- .04 # the margins we add to the y axis for the grey areas.
334     with(data,{
335         min <- floor(min(unique(time)))
336         if(wave) {
337             require(biwavelet) # for "wt" function
338             tmp <- wt(with(data,cbind(time,cases)))
339             plot.biwavelet(tmp,xlab="",ylab="")
340             if(xlab) title(xlab="time (year)")
341             if(ylab) title(ylab="period (year)")
342         } else {
343             max <- ceiling(max(unique(time)))
344             plot(time,cases,ann=F,type="n",xaxs="i")
345             if(xlab) title(xlab="time (year)")
346             if(ylab) title(ylab="incidence (ind.)")
347             foo <- range(cases,na.rm=T)
348             foo <- foo + diff(foo)*c(-a,a)
349             for(year in seq(min,max,2))
350             polygon(c(year,year+1,year+1,year),rep(foo,each=2),
351                 col="lightgrey",border=NA)
352             points(time,cases,type="l",lwd=lwd)
353             sel <- outliers(scaling(detrend(
354                 interpolate(data))),perc)
355             with(data,points(time[sel],cases[sel],
356                 col="red",pch=19))
357             box()
358         }
359             mtext(gsub(" [1,2]","",province[1]),
360                 at=min,adj=0,line=.1)
361     })
```

```
362 }
363
364
365
366
367 discardoutliers <- function(data=subset(ili93,province=="An Giang"),
368     perc=.01) {
369     sel <- outliers(scaling(detrend(interpolate(data))),perc)
370     data$cases[sel] <- NA
371     return(data)
372 }
373
374
375 doXprovince <- function(fct1province,data=ili93) {
376 # This function applies a function to the cases of all the provinces.
377 # data        : output of "readili" or "merge_ili".
378 # fct1province: the function used for the transformation.
379 # Examples    :
380 #   doXprovince(function(x)transform(x,function(x)sqrt(x+1)))
381 #   doXprovince(function(x)interpolate(x))
382 #   doXprovince(function(x)detrend(interpolate(x),ff=.1))
383 #   doXprovince(function(x)scaling(x))
384 #   doXprovince(function(x)discardoutliers(x,perc=.01))
385     out <- lapply(unique(data$province),function(x)
386         fct1province(subset(data,province==x)))
387     return(do.call("rbind",out))
388 }
389
390
391
392 plotXprovinces <- function(data=ili93,x1=.06,x2=.99,y1=.045,
393     y2=.98,vspace=.04,hspace=.05,nbrow=8,nbcol=2,shp=provinces92_lh,
394     centroids=centroids93,lwd=.5,perc=0.01,plotcent=T,wave=F) {
395         prov <- unique(data$province)
396     width <- (x2-x1-hspace*(nbcol-1))/nbcol
397     height <- (y2-y1-vspace*(nbrow-1))/nbrow
398     nbprov <- length(prov)
```

```
399     opar <- par()
400     k <- 1
401     for(i in 1:nbrow) {for(j in 1:nbcol) {
402         p1 <- p2 <- pltpar <- c(x1+(j-1)*(width+hspace),
403             x1+j*width+(j-1)*hspace,
404             y2-i*height-(i-1)*vspace,
405             y2-(i-1)*(height+vspace))
406         xint <- p1[2] - par("fin")[2]*
407             (pltpar[4]-pltpar[3])/(2*par("fin")[1])
408         p1[2] <- p2[1] <- xint
409         par(plt=p1,new=!(i<2 & j<2),mgp=c(1.5,.5,0))
410         if(k <= nbprov) {
411             plot1province(data=subset(data,province==prov[k]),
412                 xlab=(i==nbrow)|(k==nbprov),ylab=j<2,
413                 lwd,perc,wave)
414             par(plt=p2,new=T,mgp=c(1.5,.5,0))
415             map_1_province(gsub(" [1,2]","",prov[k]),shp,
416                 centroids,plotcent)
417         }
418         k <- k + 1
419     }}
420     suppressWarnings(par(opar))
421 }
422
423
424
425 map_1_province <- function(prov="An Giang",shp=provinces92_lh,
426     centroids=centroids93,plotcent=T) {
427     require(maptools)
428     plot(vietnam_light,col="black")
429     plot(shp[names(shp)=="Kien Giang"],add=T,col="black")
430     if(plotcent) {
431         centroids$province <- gsub(" [1,2]","",centroids$province)
432         with(subset(centroids,province==prov),
433             points(longitude,latitude,pch=19,col="red"))
434     } else invisible(sapply(prov,function(x)
435         plot(shp[names(shp)==x],add=T,col="red",border="red")))
```

```
436 }
437
438
439
440
441 figure_incidences <- function(data=ili93,
442     x1=.06,x2=.99,y1=.045,y2=.98,vspace=.04,hspace=.025,
443     nbrow=8,nbcol=3,name="raw",thewidth=7,theheight=8,
444     shp=provinces92_lh,centroids=centroids93,lwd=.5,
445     perc=.01,bylat=T,plotcent=F,wave=F) {
446 # Raw data:
447 #figure_incidences()
448 # Transformed data:
449 #figure_incidences(doXprovince(function(x)scaling(detrend(transform(
    interpolate(discardoutliers(x,perc=.01)),function(x)sqrt(x)),ff=.1)),
    ili93),name="transformed",perc=0)
450 # Wavelets on transformed data:
451 #figure_incidences(doXprovince(function(x)scaling(detrend(transform(
    interpolate(discardoutliers(x,perc=.01)),function(x)sqrt(x)),ff=.1)),
    ili93),name="waves",perc=0,wave=T)
452     if(bylat) {
453         data <- merge(data,centroids,"province")
454         data <- data[with(data,order(latitude,time)),]
455         provinces <- centroids[order(centroids$latitude),
456             "province"]
457     } else provinces <- sort(unique(data$province))
458     provperpage <- nbrow*nbcol
459     nbprov <- length(provinces)
460     starts <- seq(1,nbprov,nbrow*nbcol)
461     nbpages <- length(starts)
462     supp <- nbprov%%provperpage
463     if(supp>0) {
464         for(i in 1:(nbpages-1)) {
465             pdf(paste(name,i,".pdf",sep=""),width=thewidth,
466                 height=theheight,pointsize=8.5)
467             plotXprovinces(data,x1,x2,y1,y2,vspace,hspace,
468                 nbrow,nbcol,shp,centroids,lwd,perc,
```

```
469                 plotcent,wave)
470             dev.off()
471             provinces <- provinces[-(1:provperpage)]
472             data <- data[with(data,is.element(province,
473                 provinces)),]
474         }
475         ncol <- min(supp,nbcol)
476         nrow <- (supp-1)%/%nbcol + 1
477         width <- (x2-x1-hspace*(nbcol-1))/nbcol
478         newwidth <- thewidth*(1-(nbcol - ncol)*(width+hspace))
479         height <- (y2-y1-vspace*(nbrow-1))/nbrow
480         newheight <- theheight*(1-(nbrow-nrow)*(height+vspace))
481         x1 <- x1*thewidth/newwidth
482         x2 <- 1-(1-x2)*thewidth/newwidth
483         hspace <- hspace*thewidth/newwidth
484         y1 <- y1*theheight/newheight
485         y2 <- 1-(1-y2)*theheight/newheight
486         vspace <- vspace*theheight/newheight
487         pdf(paste(name,nbpages,".pdf",sep=""),width=newwidth,
488             height=newheight,pointsize=8.5)
489         plotXprovinces(data,x1,x2,y1,y2,vspace,hspace,nrow,ncol,
490             shp,centroids,lwd,perc,plotcent,wave)
491         dev.off()
492     } else for(i in 1:(nbpages)) {
493         pdf(paste(name,i,".pdf",sep=""),width=7,height=8,
494             pointsize=8.5)
495         plotXprovinces(data,x1,x2,y1,y2,vspace,hspace,nbrow,nbcol,
496             shp,centroids,lwd,perc,plotcent,wave)
497         dev.off()
498         provinces <- provinces[-(1:provperpage)]
499         data <- data[with(data,is.element(province,provinces)),]
500     }
501 }
502
503
504
505 heatmap2 <- function(data=ili93,centroids=centroids93,
```

```
506     provinces=provinces92_lh,ff=.1,perc=.01,col1="blue",hl=T,x1=0,x2=.84,
507     x3=.92,x4=.935,y00=.015,y0=.195,y1=.245,y2=.755,y3=.81,y4=.99,
508     ts="Hoa Binh",clr=F,nbcolors=12,expo=.5) {
509 # Compared to "heatmap", this function has 2 example time series: one for
        the North, one
510 # for the south.
511 # pdf("figure_1b.pdf",width=5,height=3.25,pointsize=8.5); heatmap2(); dev.
        off()
512 #   jet.colors <- colorRampPalette(c("#00007F","blue","#007FFF","cyan",
513 #       "#7FFF7F","yellow","#FF7F00","red","#7F0000"))
514 #   colors = jet.colors(nbcolors)
515     colors <- rev(heat.colors(nbcolors))
516     data <- doXprovince(function(x)scaling(detrend(transform(
517         interpolate(discardoutliers(x,perc)),
518             function(x)sqrt(x)),ff)),data)
519 # Ordering the data by latitude:
520     data <- merge(data,centroids,"province")
521     data <- data[with(data,order(latitude,time)),]
522     prov_names <- gsub(" [1,2]","",unique(data$province))
523     ts <- which(prov_names==ts)
524     ts2 <- which(prov_names=="Ninh Thuan")
525 # The time vector:
526     time <- unique(data$time)
527 # Putting the incidence data into a 2x2 matrix:
528     data2 <- with(data,split(cases,province))
529     incidences <- as.matrix(as.data.frame(data2))
530     ind <- sapply(unique(data$province),function(x)
531         which(names(data2)==x))
532     incidences <- incidences[,ind]
533     nbprov <- ncol(incidences)
534 # The heatmap:
535     xint <- x1 + par("fin")[2]*(y2-y1)/(2*par("fin")[1])
536     opar <- par(plt=c(xint,x2,y1,y2))
537 # We may want to remove the big value due to Ho Chi Minh:
538 #   incidences[incidences==max(incidences)] <- 5.493404
539 #incidences[incidences<(-5)] <- -5
540 #incidences[incidences>5] <- 5
```

```
541     incidences2 <- sign(incidences)*abs(incidences)^expo
542     image(time,1:nbprov,incidences2,col=colors,axes=F,
543         ann=F)
544 #       xlab="time (year)",ylab="")
545 # The axes:
546     ats <- seq(floor(min(time)),ceiling(max(time)),2)
547     axis(1,ats,paste(ats))
548     axis(3,ats,paste(ats))
549     axis(4); box()
550     mtext("provinces",4,line=1.5)
551 # The map of Vietnam:
552     par(plt=c(x1,xint,y1,y2),new=T)
553     plot(provinces,yaxs="i",col="lightgrey")
554 #   if(hl) {
555 #       plot(provinces[names(provinces)=="Hoa Binh"],add=T,col=col1,border=
    col1)
556 #       plot(provinces[names(provinces)=="Ninh Thuan"],add=T,col="red",
    border="red")
557 #   }
558 # The lines:
559     x <- c(.5,nbprov+.5)
560     y <- bbox(provinces)[2,]
561     model <- lm(y~x)
562     Y2 <- predict(model,data.frame(x=1:nbprov))
563     centroids <- centroids[order(centroids$latitude),]
564     Y1 <- centroids$latitude
565     X1 <- centroids$longitude
566     X2 <- diff(bbox(provinces)[1,])*.07+bbox(provinces)[1,2]
567     segments(X1,Y1,X2,Y2,lwd=.5)
568     if(hl) {
569         segments(X1[ts],Y1[ts],X2,Y2[ts],col=col1,lwd=1.5)
570         segments(X1[ts2],Y1[ts2],X2,Y2[ts2],col="red",lwd=1.5)
571         plot(provinces[names(provinces)=="Hoa Binh"],
572             add=T,col=col1,border=col1)
573         plot(provinces[names(provinces)=="Ninh Thuan"],
574             add=T,col="red",border="red")
575     }
```

```
576 # The scale:
577     par(plt=c(x3,x4,y1,y2),new=T)
578     thescale <-seq(min(incidences2,na.rm=T),
579         max(incidences2,na.rm=T),le=100)
580     image(1,thescale,matrix(thescale,nrow=1,byrow=T),col=colors,
581         axes=F,ann=F)
582     mtext("incidence z-scores",4,line=1.5); box()
583 #   axis(4)
584     aaa <- pretty(c(floor(incidences),ceiling(incidences)),30)
585     axis(4,sign(aaa)*abs(aaa)^expo,paste(aaa))
586 # Adding the time series of a selected province:
587 if(y3<y4) {
588     par(plt=c(xint,x2,y3,y4),new=T,xaxs="i")
589     if(clr) {
590         hoabinh <- incidences[,ts]
591         l <- length(hoabinh)
592         a <- approx(1:l,hoabinh,seq(1,l,le=1e5))
593         with(a,plot(x,y,col=colors[
594             as.numeric(cut(y,seq(min(incidences,na.rm=T),
595             max(incidences,na.rm=T),le=nbcolors+1)))],pch=".",
596             axes=F,ann=F))
597     } else {
598         plot(incidences[,ts],type="l",axes=F,ann=F,col=col1)
599 #       lines(incidences[,ts2],col="red")
600     }
601     axis(2,line=.5)
602     mtext("z-scores",2,line=2)
603     par(plt=c(xint,x2,y00,y0),new=T,xaxs="i")
604     plot(incidences[,ts2],type="l",axes=F,ann=F,col="red")
605     axis(2,line=.5)
606     mtext("z-scores",2,line=2)
607 }
608 # Back to initial graphic parameter values
609     par(opar)
610 #return(incidences)
611     invisible(prov_names)
612 }
```

```
613
614
615
616 correlations <- function(n=11,col="Spectral",lag=F,below=T) {
617     require(RColorBrewer)
618 #    data <- doXprovince(function(x)scaling(detrend(transform(interpolate(
        discardoutliers(x,.01)),function(x)sqrt(x)),.1)),ili93)
619     meteo <- subset(meteo,year>1992)
620     fit <- find_station()
621     fit[,1] <- as.character(fit[,1])
622     fit[,2] <- as.character(fit[,2])
623     fct <- function(x) {
624         aa <- subset(data,province==fit[x,1],c("year","month","cases"))
625         bb <- subset(meteo,station==fit[x,2],c("year","month","aH"))
626         cc <- merge(aa,bb)
627         aa <- cc[,3]
628         bb <- cc[,4]
629         if(lag) return(cor.test(aa[-1],bb[-length(bb)]))
630         else return(cor.test(aa,bb))
631     }
632     power <- data.frame(province=names(out),power=max_power(out,c(.9,1.1),
            max)$power)
633     out <- lapply(1:nrow(fit),function(x)fct(x))
634     power <- merge(fit,power,sort=F)$power
635
636 #    latitudes <- merge(fit,centroids,sort=F)$latitude
637     colors <- rev(brewer.pal(n,col))
638     colors <- colors[as.numeric(cut(power,n))]
639
640     m1 <- unlist(lapply(out,function(x)return(x$estimate)))
641     m2 <- unlist(lapply(out,function(x)return(x$p.value)))
642
643     out <- data.frame(m1,m2,colors,power)
644     out <- out[order(out$power),]
645     with(out,{
646         if(below) sel <- power <= 4
647         else sel <- power > 4
```

```
        plot(log10(m2),out$m1,type="n",xlim=c(-13,0),ylim=c(-.4,.6))
        points(log10(m2[sel]),out$m1[sel])
    })
}



fig_ILI_aH <- function(data=ili93,ff=.1,perc=.01,aH=meteo,
    centroids=centroids93,stat=stations,low=0,up=30,
    ylim_ili=c(-1,1),ylim_aH=c(15,35),left=T) {
    m <- 1:13
    col1 <- rgb(1,.55,0,.2)
    col2 <- "orange"
    col3 <- rgb(0,0,1,.125)
    col4 <- "deepskyblue"
# ILI:
    data <- doXprovince(function(x)scaling(detrend(transform(
        interpolate(discardoutliers(x,perc)),
            function(x)sqrt(x)),ff)),data)
    data <- subset(data,is.element(province,
        subset(centroids,latitude>low & latitude<=up,province,T)))
    print(length(unique(data$province)))
    foo <- with(data,tapply(cases,month,quantile,c(.25,.5,.75)))
    lower_ili <- sapply(foo,function(x)x[1])
    media_ili <- sapply(foo,function(x)x[2])
    upper_ili <- sapply(foo,function(x)x[3])
    lower_ili <- c(lower_ili,lower_ili[1])
    media_ili <- c(media_ili,media_ili[1])
    upper_ili <- c(upper_ili,upper_ili[1])
# aH:
    bar <- subset(aH,year>1992 & is.element(station,
        subset(stat,latitude>low & latitude<=up,station,T)))
    print(length(unique(bar$station)))
    foo <- with(bar,tapply(aH,month,quantile,c(.25,.5,.75),na.rm=T))
    lower_aH <- sapply(foo,function(x)x[1])
    media_aH <- sapply(foo,function(x)x[2])
    upper_aH <- sapply(foo,function(x)x[3])
```

```
685     lower_aH <- c(lower_aH,lower_aH[1])
686     media_aH <- c(media_aH,media_aH[1])
687     upper_aH <- c(upper_aH,upper_aH[1])
688 # plot ILI:
689     plot(rep(m,3),c(lower_ili,media_ili,upper_ili),type="n",axes=F,
690         ylim=ylim_ili,xlab="months of the year",ylab="")
691     axis(1,seq(2,12,2),c("Feb.","Apr.","Jun.","Aug.","Oct.","Dec."))
692     if(left) {
693         axis(2)
694         mtext("normalized incidence",2,1.25)
695     } else {
696         axis(4)
697         mtext("normalized incidence",4,1.25)
698     }
699     polygon(c(m,rev(m)),c(lower_ili,rev(upper_ili)),col=col1,border=col1)
700     lines(m,media_ili,col=col2)
701 # plot aH:
702     par(new=T)
703     plot(rep(m,3),c(lower_aH,media_aH,upper_aH),type="n",
704         axes=F,ann=F,ylim=ylim_aH)
705     if(left) {
706         axis(4)
707         mtext("absolute humidity (g/L)",4,1.25)
708     } else {
709         axis(2)
710 #       mtext("absolute humidity (g/L)",2,1.25)
711     }
712     polygon(c(m,rev(m)),c(lower_aH,rev(upper_aH)),col=col3,border=col3)
713     lines(m,media_aH,col=col4)
714     box()
715 }
716
717
718
719
720 fig_ILI_aH2 <- function(data=ili93,ff=.1,perc=.01,aH=meteo,
721     centroids=centroids93,stat=stations,thresh=16,
```

```
722     ylim_ili=c(-1.5,1.5),ylim_aH=c(15,35)) {
723     x1 <- .07; x4 <- .93
724     y1 <- .185; y2 <- .98
725     hspace <- .115
726 # For a PNG figure:
727     x1 <- .07; x4 <- .93
728     y1 <- .19; y2 <- .97
729     hspace <- .108
730     width <- (x4-x1-hspace)/2
731     x2 <- x1 + width
732     x3 <- x4 - width
733     opar <- par(plt=c(x1,x2,y1,y2))
734     fig_ILI_aH(data,ff,perc,aH,centroids,stat,low=thresh,up=30,
735         ylim_ili,ylim_aH)
736     par(plt=c(x3,x4,y1,y2),new=T)
737     fig_ILI_aH(data,ff,perc,aH,centroids,stat,low=0,up=thresh,
738         ylim_ili,ylim_aH,left=F)
739     par(opar)
740 }
741
742
743
744
745 heatmap_filtered <- function(data,centroids=centroids93,
746     provinces=provinces92_lh,ff=.1,perc=.01,col1="blue",hl=T,x1=0,x2=.84,
747     x3=.92,x4=.935,y00=.015,y0=.195,y1=.245,y2=.755,y3=.81,y4=.99,
748     ts="Hoa Binh",clr=F,nbcolors=12,expo=.5) {
749     colors <- rev(heat.colors(nbcolors))
750 # The time vector:
751     time <- data$time
752 # Ordering the data by latitude:
753     data <- data$filtered_ts[,with(centroids93,
754         as.character(province[order(latitude)]))]
755     prov_names <- gsub(" [1,2]","",names(data))
756     ts <- which(prov_names==ts)
757     ts2 <- which(prov_names=="Ninh Thuan")
758     nbprov <- ncol(data)
```

```
759 # The heatmap:
760     xint <- x1 + par("fin")[2]*(y2-y1)/(2*par("fin")[1])
761     opar <- par(plt=c(xint,x2,y1,y2))
762     image(time,1:nbprov,as.matrix(data),col=colors,axes=F,ann=F)
763     incidences <- incidences2 <- data
764 # The axes:
765     ats <- seq(floor(min(time)),ceiling(max(time)),2)
766     axis(1,ats,paste(ats))
767     axis(3,ats,paste(ats))
768     axis(4); box()
769     mtext("provinces",4,line=1.5)
770 # The map of Vietnam:
771     par(plt=c(x1,xint,y1,y2),new=T)
772     plot(provinces,yaxs="i",col="lightgrey")
773 # The lines:
774     x <- c(.5,nbprov+.5)
775     y <- bbox(provinces)[2,]
776     model <- lm(y~x)
777     Y2 <- predict(model,data.frame(x=1:nbprov))
778     centroids <- centroids[order(centroids$latitude),]
779     Y1 <- centroids$latitude
780     X1 <- centroids$longitude
781     X2 <- diff(bbox(provinces)[1,])*.07+bbox(provinces)[1,2]
782     segments(X1,Y1,X2,Y2,lwd=.5)
783     if(hl) {
784         segments(X1[ts],Y1[ts],X2,Y2[ts],col=col1,lwd=1.5)
785         segments(X1[ts2],Y1[ts2],X2,Y2[ts2],col="red",lwd=1.5)
786         plot(provinces[names(provinces)=="Hoa Binh"],
787             add=T,col=col1,border=col1)
788         plot(provinces[names(provinces)=="Ninh Thuan"],add=T,
789             col="red",border="red")
790     }
791 # The scale:
792     par(plt=c(x3,x4,y1,y2),new=T)
793     thescale <-seq(min(incidences2,na.rm=T),
794         max(incidences2,na.rm=T),le=100)
795     image(1,thescale,matrix(thescale,nrow=1,byrow=T),col=colors,
```

```
796         axes=F,ann=F)
797     mtext("incidence z-scores",4,line=1.5); box()
798     axis(4)
799 # Adding the time series of a selected province:
800 if(y3<y4) {
801     par(plt=c(xint,x2,y3,y4),new=T,xaxs="i")
802     if(clr) {
803         hoabinh <- incidences[,ts]
804         l <- length(hoabinh)
805         a <- approx(1:l,hoabinh,seq(1,l,le=1e5))
806         with(a,plot(x,y,col=colors[
807             as.numeric(cut(y,seq(min(incidences,na.rm=T),
808             max(incidences,na.rm=T),le=nbcolors+1)))],pch=".",
809             axes=F,ann=F))
810     } else {
811         mx <- max(abs(incidences[,ts]))
812         plot(incidences[,ts],type="l",
813             axes=F,ann=F,col=col1,ylim=c(-mx,mx))
814     }
815     axis(2,line=.5)
816     mtext("z-scores",2,line=2)
817     par(plt=c(xint,x2,y00,y0),new=T,xaxs="i")
818     plot(incidences[,ts2],type="l",axes=F,ann=F,col="red",
819         ylim=c(-mx,mx))
820     axis(2,line=.5)
821     mtext("z-scores",2,line=2)
822 }
823 # Back to initial graphic parameter values
824     par(opar)
825     invisible(prov_names)
826 }
827
828
829
830 wavelet_all <- function(data=doXprovince(function(x)scaling(detrend(
831     transform(interpolate(discardoutliers(x,perc=.01)),
832     function(x)sqrt(x)),ff=.1)),ili93),s0=NULL) {
```

```
# This function calculates the wavelet transform for all the provinces.
    require(biwavelet) # for "wt" function.
    provinces <- unique(data$province)
    nbprov <- length(provinces)
    if(is.null(s0)) w <- wt(with(subset(data,province==province[1]),
        cbind(time,cases)))
    else w <- wt(with(subset(data,province==province[1]),
        cbind(time,cases)),s0=s0)
    wa <- array(NA,dim=c(nbprov,nrow(w$wave),ncol(w$wave)))
    wa[1,,]=w$wave
    if(is.null(s0)) for(i in 2:nbprov) {
        w <- wt(with(subset(data,province==provinces[i]),
            cbind(time,cases)))
        wa[i,,] <- w$wave
    } else for(i in 2:nbprov) {
        w <- wt(with(subset(data,province==provinces[i]),
            cbind(time,cases)),s0=s0)
        wa[i,,] <- w$wave
    }
    return(list(wa=wa,provinces=provinces))
}



distcalc <- function(object) {
# This function calculates the distances between the wavelet transforms of
# all the provinces.
# object: output of "wavelet_all".
    require(biwavelet) # for "wclust" function.
    return(list(distances=wclust(object$wa),
        provinces=object$provinces))
}



clustcalc <- function(object,method="ward") {
# This function does a hierarchical clustering of the provinces based on
```

```
870 # the matrix of distances between their wavelet spectra.
871 # object: output of "distcalc".
872     clusters <- hclust(object$distances$dist.mat,method)
873     clusters$labels <- object$provinces
874     return(clusters)
875 }
876
877
878
879 plotclust <- function(object) {
880 # This function plots the hierarchy of provinces calculated from the
881 # similarities between their wavelet spectra.
882 # object: output of "clustcalc".
883     plot(object,sub="",main="",ylab="dissimilarity",hang=-1)
884 }
885
886
887
888 plotonmap <- function(object,k=3,shp=provinces92_lh) {
889 # object: output of "clustcalc".
890     require(sp)
891     out <- cutree(object,k)
892     provnames <- gsub(" [1,2]","",names(out))
893     plot(shp)
894     for(i in 1:length(out))
895         plot(shp[names(shp)==provnames[i]],add=T,col=out[i]+1)
896 }
897
898
899
900 global_spect_calc <- function(data=doXprovince(function(x)scaling(detrend(
    transform(interpolate(discardoutliers(x,perc=.01)),function(x)sqrt(x)),
    ff=.1)),ili93),dj=1/100,upperPeriod=3,space="province",var="cases") {
901 # This function calculates the global spectrum of a set "data" of
902 # provinces.
903     require(ondelettes) ## for "morlet" and "waveglobal"
904     names(data) <- gsub(space,"province",names(data))
```

```
905     names(data) <- gsub(var,"cases",names(data))
906     provinces <- unique(data$province)
907     wavelet_transformed <- lapply(provinces,function(x)
908         morlet(subset(data,province==x,cases,T),dt=1/12,dj=dj,
909         lowerPeriod=.1,upperPeriod=upperPeriod,pad=2^8))
910     periods <- wavelet_transformed[[1]]$fourier*
911         wavelet_transformed[[1]]$scale
912     globals <- lapply(wavelet_transformed,function(x)
913         return(data.frame(periods=periods,globwave=waveglobal(x))))
914     names(globals) <- provinces
915     return(globals)
916 }
917
918
919
920 filtered_phase_calc <- function(data=doXprovince(function(x)scaling(detrend
        (transform(interpolate(discardoutliers(x,perc=.01)),function(x)sqrt(x))
        ,ff=.1)),ili93),var1="province",var2="cases",dj=1/100,upperPeriod=3,
        space="province",filt_lwr=.9,filt_upp=1.1) {
921 # This function returns the time series and the phases of the filtered
922 # signals.
923     require(ondelettes) ## for "morlet" and "wavefilter"
924     provinces <- unique(data[,var1])
925     tmp <- lapply(provinces,function(x)
926         wavefilter(morlet(data[data[,var1]==x,var2],dt=1/12,dj=dj,
927         lowerPeriod=.1,upperPeriod=upperPeriod,pad=2^8),filt_lwr,
928         filt_upp))
929     thelengths <- sapply(tmp,function(x)length(x$ts))
930     sel <- !(thelengths<max(thelengths))
931     tmp <- tmp[sel]
932     filtered_ts <- as.data.frame(sapply(tmp,function(x)x$ts))
933     filtered_phase <- as.data.frame(sapply(tmp,function(x)x$phase))
934     names(filtered_ts) <- names(filtered_phase) <- provinces[sel]
935     return(list(time=unique(data$time),filtered_ts=filtered_ts,
936         filtered_phase=filtered_phase))
937 }
938
```

```
939
940
941
942 calc_filtered_phase_aH <- function() {
943 # This function filters and calculates the phase of the absolute humidity:
944     meteo <- subset(meteo,year>1992,T)
945     meteo$time <- with(meteo,year+month/12-1/24)
946     out <- filtered_phase_calc(meteo,var1="station",var2="aH")
947     sel <- !sapply(out$filtered_ts,function(x)any(is.na(x)))
948     out$filtered_ts <- out$filtered_ts[sel]
949     out$filtered_phase <- out$filtered_phase[sel]
950     return(out)
951 }
952
953
954
955 angle_transf <- function(data) {
956 # This function transforms a vector "data" of phase angles from
957 # [-pi,+pi] to [-pi,+infty].
958 # data is a vector of phase angles, typically an output of the "angle"
    function
959         zz <- which((diff(data)<0)>0)+1
960         data[zz[which(diff(zz)==1)]] <- NA
961         tmp <- data>0
962         tmp[is.na(tmp)] <- TRUE
963         tmp <- diff(tmp)
964         tmp <- which(tmp<0)
965         end <- length(data)
966         for(i in 1:length(tmp))
967                 data[(tmp[i]+1):end] <- data[(tmp[i]+1):end] + 2*pi
968         return(data)
969 }
970
971
972
973 dff <- function() {
974 # This function calculates the phase angles of the ILI and the
```

```
975 # corresponding absolute humidity for each province.
976     aH <- calc_filtered_phase_aH()$filtered_phase
977     ili <- filtered_phase_calc(data)$filtered_phase
978     interf <- find_station(stat=stations[
979         is.element(stations$station,names(aH)),])
980     sel <- as.character(merge(data.frame(province=names(ili)),
981         interf,sort=F)[,2])
982     aH <- aH[,sel]
983     return(list(aH,ili))
984 }
985
986
987
988
989 plot_phase_diff6 <- function(n=11,col="Spectral",thresh=3) {
990 #pdf("figure_7b.pdf",width=4.5,height=5.4,pointsize=8.5); plot_phase_diff6(
        thresh=4); dev.off()
991 # For a PDF figure of chapter 4.
992     x1 <- .07; x4 <- .93
993     y1 <- .075; y6 <- .98
994     hspace <- .115
995     width <- (x4-x1-hspace)/2
996     x2 <- x1 + width
997     x3 <- x4 - width
998     vspace <- .1
999     height <- (y6-y1-2*vspace)/3
1000    y2 <- y1 + height
1001    y3 <- y2 + vspace
1002    y4 <- y3 + height
1003    y5 <- y4 + vspace
1004
1005    require(RColorBrewer)
1006    colors <- rev(brewer.pal(n,col))
1007    xx <- dff()
1008    provnames <- names(xx[[2]])
1009    phase_abshum <- as.data.frame(lapply(xx[[1]],
1010        function(x)angle_transf(x)))
```

```
1011      phase_ili <- as.data.frame(lapply(xx[[2]],
1012          function(x)angle_transf(x)))
1013      thediff <- phase_ili - phase_abshum
1014      thediff <- thediff%%(2*pi)
1015      fct <- function(x){
1016          sel <- x > pi
1017          sel[is.na(sel)] <- F
1018          x[sel] <- x[sel] - 2*pi
1019          return(x)
1020      }
1021      thediff <- as.data.frame(lapply(thediff,fct))
1022      names(thediff) <- provnames
1023      col <- data.frame(province=names(out),
1024          maxpower=max_power(out,c(.9,1.1),max)$power)
1025      col <- merge(data.frame(province=names(thediff)),col,sort=F)[,2]
1026      thediff <- thediff[,order(col)]
1027      col <- sort(col)
1028      times <- unique(data$time)
1029      index <- as.numeric(cut(col,n))
1030  # Left panel
1031      opar <- par(plt=c(x1,x2,y5,y6))
1032      sel <- col>thresh
1033      print(sum(sel))
1034      index2 <- rep(index[sel],each=length(times))
1035      thediff2 <- thediff[,sel]
1036      plot(rep(times,ncol(thediff2)),unlist(thediff2),type="n",
1037          xlab="time (year)",
1038          ylab="phase difference (radians)",axes=F)
1039  #   abline(h=c(-pi,-2*pi/3,-pi/3,0,pi/3,2*pi/3,pi),col="lightgrey")
1040      abline(h=c(-pi,-5*pi/6,-2*pi/3,-pi/2,-pi/3,-pi/6,0,
1041          pi/6,pi/3,pi/2,2*pi/3,5*pi/6,pi),col="lightgrey")
1042      points(rep(times,ncol(thediff2)),unlist(thediff2),pch=19,
1043          col=colors[index2],cex=.25)
1044      axis(1)
1045      axis(2,c(-pi,-pi/2,0,pi/2,pi),expression(-pi,-pi/2,0,pi/2,pi))
1046      axis(4,c(-pi,-2*pi/3,-pi/3,0,pi/3,2*pi/3,pi),
1047          c("-6","-4","-2","0","2","4","6"))
```

```
1048      box()
1049  # Right panel
1050      par(plt=c(x3,x4,y5,y6),new=T)
1051      sel <- col<=thresh
1052      print(sum(sel))
1053      index2 <- rep(index[sel],each=length(times))
1054      thediff2 <- thediff[,sel]
1055      plot(rep(times,ncol(thediff2)),unlist(thediff2),type="n",
1056          xlab="time (year)",
1057          axes=F,ylab="lag (months)")
1058      abline(h=c(-pi,-5*pi/6,-2*pi/3,-pi/2,-pi/3,-pi/6,0,
1059          pi/6,pi/3,pi/2,2*pi/3,5*pi/6,pi),col="lightgrey")
1060      points(rep(times,ncol(thediff2)),unlist(thediff2),pch=19,
1061          col=colors[index2],cex=.25)
1062      axis(1)
1063      axis(2,c(-pi,-2*pi/3,-pi/3,0,pi/3,2*pi/3,pi),
1064          c("-6","-4","-2","0","2","4","6"))
1065      axis(4,c(-pi,-pi/2,0,pi/2,pi),expression(-pi,-pi/2,0,pi/2,pi))
1066      box()
1067      mtext("phase difference (radians)",4,1.25)
1068  #
1069      par(plt=c(x1,x2,y3,y4),new=T)
1070      correlations(n=11,col="Spectral",lag=F,below=F)
1071      par(plt=c(x3,x4,y3,y4),new=T)
1072      correlations(n=11,col="Spectral",lag=F,below=T)
1073  #
1074      par(plt=c(x1,x2,y1,y2),new=T)
1075      correlations(n=11,col="Spectral",lag=T,below=F)
1076      par(plt=c(x3,x4,y1,y2),new=T)
1077      correlations(n=11,col="Spectral",lag=T,below=T)
1078  # End
1079      par(opar)
1080      invisible(thediff)
1081  }
1082
1083
1084  plot_phases <- function(out) {
```

```
# "out" is an output of the "filtered_phase_calc" function.
    with(out,{
        nbprov <- ncol(out$filtered_ts)
        plot(rep(time,nbprov),unlist(filtered_ts),type="n")
        for(i in 1:nbprov)
            lines(time,filtered_ts[,i],col=rgb(0,0,0,.1))
    })
}



fig_glob_power1 <- function(out,n=11,col="Spectral") {
# This functions draws the global power with a color that reflects the
# latitutde.
# out is an output of "global_spect_calc".
    require(RColorBrewer)
    colors <- brewer.pal(n,col)
    x <- unlist(lapply(out,function(x)x$periods))
    y <- unlist(lapply(out,function(x)x$globwave))
    plot(x,y,type="n",xlab="period (year)",ylab="power",
        axes=F,xlim=c(0,max(x)))
    axis(1); axis(2)
    index <- as.numeric(cut(with(centroids93,
        latitude[match(names(out),province)]),n))
    for(i in 1:length(out))
        with(out[[i]],lines(periods,globwave,col=colors[index[i]]))
}



fig_glob_power2 <- function(out,n=11,col="Spectral",rge=c(.9,1.1),fct=max)
    {
# This function draws the global power with a colors that reflects the
# maximum power at 1 year.
# out is an output of "global_spect_calc".
    require(RColorBrewer)
```

```
1121      colors <- rev(brewer.pal(n,col))
1122      x <- unlist(lapply(out,function(x)x$periods))
1123      y <- unlist(lapply(out,function(x)x$globwave))
1124      plot(x,y,type="n",xlab="period (year)",ylab="power",
1125          axes=F,xlim=c(0,max(x)))
1126      axis(1); axis(2)
1127      mxpwr <- max_power(out,rge,fct)$power
1128      index <- as.numeric(cut(mxpwr,n))
1129      for(i in 1:length(out))
1130          with(out[[i]],lines(periods,globwave,col=colors[index[i]]))
1131  }
1132
1133
1134
1135
1136  map_lat2 <- function(out,prov=provinces92_lh,n=11,col="Spectral",
1137      rge=c(.9,1.1),fct=max) {
1138  # This function draws the map of vietnam with the color of the provinces
1139  # reflecting the seasonality (i.e. maximum power around the period of 1
         year).
1140  # out is an output of "global_spect_calc".
1141      require(RColorBrewer)
1142      colors <- rev(brewer.pal(n,col))
1143      pwr <- max_power(out,rge,fct)
1144      mxpwr <- pwr$power
1145      provpwr <- as.character(pwr$province)
1146      provpwr <- gsub(" 1","",provpwr)
1147      provpwr <- gsub(" 2","",provpwr)
1148      provmap <- sapply(prov@polygons,function(x)x@ID)
1149      index <- as.numeric(cut(mxpwr[match(provmap,provpwr)],n))
1150      colors <- colors[index]
1151      plot(prov,yaxs="i",col=colors)
1152  }
1153
1154
1155
1156
```

```
1157 map_lat1 <- function(prov=provinces92_lh,n=11,col="Spectral") {
1158 # This function draws the map of vietnam the the color of the provinces
1159 # reflecting the latitude.
1160     require(RColorBrewer)
1161     colors <- brewer.pal(n,col)
1162     a <- gsub(" 1","",centroids93$province)
1163     a <- gsub(" 2","",a)
1164     index <- as.numeric(cut(with(centroids93,
1165         latitude[match(sapply(prov@polygons,
1166             function(x)x@ID),a)]),n))
1167     colors <- colors[index]
1168     plot(prov,yaxs="i",col=colors)
1169 }
1170
1171
1172
1173
1174 power_map <- function(out,x1=.1,x2=.9,y1=.1,y2=.9,eps=0) {
1175 #pdf("figure_2.pdf",width=3,height=2,pointsize=8.5); power_map(out,x1=.11,
        x2=1,y1=.17,y2=.99,eps=.07); dev.off()
1176     xint <- x2 - par("fin")[2]*(y2-y1)/(2*par("fin")[1])
1177 # The power
1178     opar <- par(plt=c(x1,xint+eps,y1,y2))
1179     fig_glob_power2(out)
1180 # The map
1181     par(plt=c(xint,x2,y1,y2),new=T)
1182     map_lat2(out)
1183     par(opar)
1184 }
1185
1186
1187
1188 power_map2 <- function(out,x1=.09,x2=1,y1=.17,y2=.99,eps=.04,eps2=.04) {
1189 #pdf("figure_2.pdf",width=3.7,height=2,pointsize=8.5); power_map2(out,x1
        =.09,x2=1,y1=.17,y2=.99,eps=.04,eps2=.04); dev.off()
1190     xxxx <- par("fin")[2]*(y2-y1)/(2*par("fin")[1])
1191     xint <- x2 - 2*xxxx
```

```
1192      ll <- -.75
1193  # The power
1194      opar <- par(plt=c(x1,x2-2*xxxx+eps+eps2,y1,y2))
1195      fig_glob_power2(out)
1196      mtext("(A)",at=-.35,line=ll)
1197  # The map with the power
1198      par(plt=c(x2-2*xxxx+eps2,x2-xxxx+eps2,y1,y2),new=T)
1199      map_lat2(out)
1200      mtext("(B)",at=101.3,line=ll)
1201  # The map with the elevations:
1202      par(plt=c(x2-xxxx,x2,y1,y2),new=T)
1203      x <- 500 ## This is the elevation.
1204      plot(vietnam_light,yaxs="i")
1205      colors = rev(c("#8C510A","#DFC27D"))
1206      colors = c("lightgrey","darkgrey")
1207      image(altitudes,add=T,col = colors,breaks=c(0,x,3500))
1208      plot(vietnam_light,add=T)
1209      contour(altitudes,levels=x,add=T,drawlabels=F,lwd=.5)
1210      mtext("(C)",at=101.3,line=ll)
1211      par(opar)
1212  }
1213
1214
1215
1216
1217  figure_meteo1 <- function(v="Ta",n=9,col="YlOrRd",latmin,latmax) {
1218      require(RColorBrewer)
1219      yylab <- data.frame(Tm="minimal temperature ( C)",
1220          Ta="average temperature ( C)",
1221          Tx="maximal temperature ( C)",aH="absolute humidity (g/L)",
1222          rH="relative humidity (%)",Rf="rainfall (mm)",
1223          Sh="sunshine (hours)")
1224      temperature <- c("Tm","Ta","Tx")
1225      colors <- rev(brewer.pal(n,col))#[-1])
1226      meteo <- subset(meteo,year>=1993)
1227      temprange <- range(meteo[,temperature],na.rm=T)
1228      meteo <- with(meteo,tapply(meteo[,v],list(month,station),mean))
```

```
1229      meteo <- rbind(meteo,meteo[1,])
1230      months <- 1:13
1231      nbcol <- ncol(meteo)
1232      if(is.element(v,temperature))
1233          plot(rep(months,nbcol),as.vector(meteo),type="n",axes=F,
1234              xlab="",ylab=yylab[,v],ylim=temprange)
1235      else plot(rep(months,nbcol),as.vector(meteo),type="n",axes=F,
1236          xlab="",ylab=yylab[,v])
1237  #   axis(1,seq(1,12,2),c("Jan","Mar","May","Jul","Sep","Nov"))
1238  #   axis(1,c(1,4,7,10),c("Jan","April","July","October"))
1239      axis(1,c(1,4,7,10,13),c("Jan.","Apr.","Jul.","Oct.","Jan."),
1240          lwd=0,tick=F)
1241      axis(1,1:13,F)
1242      axis(2)
1243      a <- match(colnames(meteo),stations$station)
1244      meteo <- meteo[,!is.na(a)]
1245      nbcol <- ncol(meteo)
1246      a <- match(colnames(meteo),stations$station)
1247      a <- stations[a,"latitude"]
1248  #   a <- c(latmin,latmax,a)
1249      colors <- colors[as.numeric(cut(a,n))]#-1))]
1250  #   colors <- colors[-(1:2)]
1251      for(i in 1:nbcol) lines(months,meteo[,i],col=colors[i])
1252  }
1253
1254
1255
1256
1257  figure_meteoX <- function(n=9,col="YlOrRd",x1=.1,x2=.99,y1=.1,y2=.99,
1258      nrw=4,ncl=2,hspace=.01,vspace=.01) {
1259  #pdf("figure_3.pdf",width=4,height=6,pointsize=8.5); figure_meteoX(n=11,col
        ="Spectral",x1=.08,x2=.99,y1=.04,y2=1,nrw=4,ncl=2,hspace=.1,vspace=.03)
        ; dev.off()
1260      x <- raster_lat_grad()
1261      latmin <- x@extent@ymin
1262      latmax <- x@extent@ymax
1263      xrange <- x2 - x1
```

```
1264      width  <- (xrange - (ncl-1)*hspace)/ncl
1265      yrange <- y2 - y1
1266      height <- (yrange - (nrw-1)*vspace)/nrw
1267      opar <- par()
1268      variables <- c("Tx","Sh","Ta","Rf","Tm","rH","aH")
1269      j <- 0
1270      plot(1,type="n",ann=F,axes=F)
1271      for(r in 1:(nrw-1)) {
1272          for(c in 1:ncl) {
1273              par(plt=c(x1+(c-1)*(width+hspace),
1274                  x1+c*width+(c-1)*hspace,
1275                  y1+(r-1)*(height+vspace),
1276                  y1+r*height+(r-1)*vspace),
1277                  new=T)
1278              j <- j+1
1279              figure_meteo1(variables[j],n,col,latmin,latmax)
1280  #           mtext("(A)",line=0.2,at=-1.4)
1281          }
1282      }
1283      r <- nrw
1284      c <- 1
1285      eps <- .075
1286      par(plt=c(x1+(c-1)*(width+hspace),
1287          x1+c*width+c(c-1)*hspace,
1288          y1+(r-1)*(height+vspace)-eps,
1289          y1+r*height+(r-1)*vspace),new=T)
1290      x <- raster_lat_grad()
1291      plot(vietnam_light)
1292      image(x,add=T,col=brewer.pal(n,col))
1293      plot(vietnam_light,add=T)
1294      with(stations,points(longitude,latitude))
1295      c <- ncl
1296      par(plt=c(x1+(c-1)*(width+hspace),
1297          x1+c*width+c(c-1)*hspace,
1298          y1+(r-1)*(height+vspace),
1299          y1+r*height+(r-1)*vspace),new=T)
1300      figure_meteo1(variables[7],n,col,latmin,latmax)
```

```
1301      par(opar)
1302  }
1303
1304
1305
1306
1307  raster_lat_grad <- function(n=9,colors="YlOrRd") {
1308      require(raster)
1309  #   require(RColorBrewer)
1310  #   colors <- rev(brewer.pal(n,colors)[-1])
1311      x <- raster(xmn=102,xmx=110,ymn=8,ymx=24)
1312      projection(x) <- projection(altitudes)
1313      values(x) <- matrix(rep(seq(8,24,le=180),360),ncol=360)
1314      x <- crop(x,extent(vietnam_light))
1315      x <- rasterize(vietnam_light,x,mask=T)
1316  #   plot(vietnam_light)
1317  #   image(x,add=T,col=colors)
1318  #   plot(vietnam_light,add=T)
1319      return(x)
1320  }
1321
1322
1323
1324  figure_meteo_pca0 <- function(n=11,col="Spectral") {
1325      require(RColorBrewer)
1326      meteo <- subset(meteo,year>=1993)
1327      meteo <- na.exclude(meteo)
1328      pca <- prcomp(subset(meteo,
1329          sel=c("Ta","Tx","Tm","Rf","rH","Sh","aH")),scale=T)
1330      x <- as.data.frame(pca$x)
1331      limits <- range(with(x,c(PC1,PC2)))
1332      limits <- c(-1,1)*rep(max(abs(limits)),2)
1333      colors <- rev(brewer.pal(n,col))
1334      a <- match(meteo$station,stations$station)
1335      a <- stations[a,"latitude"]
1336      colors <- colors[as.numeric(cut(a,n))]
1337      a <- x[,c("PC1","PC2")]
```

```
1338     shuffle <- sample(1:nrow(a))
1339     a <- a[shuffle,]
1340     colors <- colors[shuffle]
1341     importance <- round(100*summary(pca)$importance[2,1:2])
1342     with(a,plot(PC1,PC2,xlim=limits,ylim=limits,
1343         xlab=paste0("PC1 (",importance[1],
1344         "% of variance)"),
1345         ylab=paste0("PC2 (",importance[2],
1346         "% of variance)"),col=colors))
1347     var <- 10*pca$rotation[,1:2]
1348     for(i in 1:7) arrows(0,0,var[i,1],var[i,2],.1)
1349     return(pca)
1350 }
1351
1352
1353
1354 figure_meteo_pca <- function(n=11,col="Spectral",width=2.75,height=4.1,
1355     x1=0,x2=.98,y1=.005,y2=.98,eps=.03) {
1356 #figure_meteo_pca(width=2.75,height=4.1,x1=0,x2=.98,y1=.005,y2=.98,eps=.03)
1357     require(RColorBrewer)
1358     require(plotrix)
1359     pdf("figure_4.pdf",width,height,pointsize=8.5)
1360     xmid <- x1+(x2-x1)/3
1361     yheight <- width*(x2-xmid)/height
1362 # The map of Vietnam:
1363     par(plt=c(x1,xmid,y1,y1+yheight-eps))
1364     x <- raster_lat_grad()
1365     plot(vietnam_light,xaxs="i",yaxs="i")
1366     image(x,add=T,col=brewer.pal(n,col))
1367     plot(vietnam_light,add=T)
1368     with(stations,points(longitude,latitude))
1369     meteo <- subset(meteo,year>=1993)
1370     meteo <- na.exclude(meteo)
1371     pca <- prcomp(subset(meteo,
1372         sel=c("Ta","Tx","Tm","Rf","rH","Sh","aH")),scale=T)
1373     x <- as.data.frame(pca$x)
1374     limits <- range(with(x,c(PC1,PC2)))
```

```
1375      limits <- c(-1,1)*rep(max(abs(limits)),2)
1376      colors <- rev(brewer.pal(n,col))
1377      a <- match(meteo$station,stations$station)
1378      latitudes <- stations[a,"latitude"]
1379      colors <- colors[as.numeric(cut(latitudes,n))]
1380      a <- x[,c("PC1","PC2")]
1381      shuffle <- sample(1:nrow(a))
1382      a <- a[shuffle,]
1383      colors <- colors[shuffle]
1384      latitudes <- latitudes[shuffle]
1385      importance <- round(100*summary(pca)$importance[2,1:2])
1386  # Plot of the PC1 as a function of latitude:
1387      par(plt=c(xmid,x2,y1,y1+yheight-eps),new=T)
1388      with(a,plot(PC1,latitudes,xlim=limits,yaxs="i",
1389          ylim=bbox(vietnam_light)[2,],ann=F,axes=F,col=colors))
1390      axis(4,line=-3)
1391      mtext("latitude",4,-1.5)
1392      axis(3,at=c(-10,-5,0,5))
1393      mtext("(B)",at=-14,line=1)
1394  # Plot of the PC1 and PC2:
1395      par(plt=c(xmid,x2,y2-yheight,y2),new=T)
1396      with(a,plot(PC1,PC2,xlim=limits,ylim=limits,
1397          xlab=paste0("PC1 (",importance[1],
1398          "% of variance)"),
1399          ylab=paste0("PC2 (",importance[2],
1400          "% of variance)"),col=colors))
1401      f <- 15
1402      var <- f*pca$rotation[,1:2]
1403      for(i in 1:7) arrows(0,0,var[i,1],var[i,2],.1)
1404      draw.circle(0,0,f*.25,lty=2,border="grey")
1405      draw.circle(0,0,f*.5,lty=2,border="grey")
1406      draw.circle(0,0,f*.75,lty=2,border="grey")
1407      draw.circle(0,0,f*1,lty=2,border="grey")
1408      abline(v=0,lty=2,col="grey")
1409      abline(h=0,lty=2,col="grey")
1410      txt <- c("Ta","Tx","Tn","Rf","rH","Sh","aH")
1411      for(i in c(1:4,6,7)) text(var[i,1],var[i,2],txt[i],
```

```
1412            pos=4,offset=.3)
1413        text(var[5,1],var[5,2]+.1,txt[5],pos=2,offset=.4)
1414        mtext("(A)",at=-14,line=-.5)
1415        dev.off()
1416    #   return(list(a,latitudes))
1417    }
1418
1419
1420
1421
1422    max_power <- function(out,rge=c(.9,1.1),fct=max) {
1423    # out: output of function "global_spect_calc".
1424    # rge: a two-value vector giving the range of periods over which to apply
        function "fct".
1425        thenames <- names(out)
1426        if(length(rge)<2) {
1427            out <- sapply(out,function(x)max(x$globwave))
1428        }
1429        else out <- sapply(out,function(x) {
1430            p <- x$periods
1431            sel <- p>=rge[1] & p<=rge[2]
1432            return(fct(x$globwave[sel]))
1433        })
1434        return(data.frame(province=thenames,power=out))
1435    }
1436
1437
1438
1439    fig_lat_power <- function(data=doXprovince(function(x)scaling(detrend(
        transform(interpolate(discardoutliers(x,perc=.01)),function(x)sqrt(x)),
        ff=.1)),ili93),dj=1/100,upperPeriod=3,rge,fct=sum) {
1440    # out: output of function "global_spect_calc".
1441    # rge: a two-value vector giving the range of periods over which to apply
        function "fct".
1442        out <- global_spect_calc(data,dj,upperPeriod)
1443        pwr <- max_power(out,rge,fct=sum)
1444        pwr <- merge(centroids93,pwr)
```

```
1445      with(pwr,plot(latitude,power))
1446  }
1447
1448
1449
1450  meteo_pca <- function(yr=1993) {
1451  # This function returns the PCA component of the climatic stations from
1452  # year yr.
1453      tmp <- na.exclude(subset(meteo,year>=yr))
1454      pca <- prcomp(tmp[,c("Ta","Tx","Tm","Rf","rH","Sh","aH")],scale=T)
1455      pc <- pca$x
1456      out <- as.data.frame(sapply(1:ncol(pc),function(x)
1457          tapply(pc[,x],tmp$station,mean)))
1458      out$station <- rownames(out)
1459      names(out) <- gsub("V","PC",names(out))
1460      out <- na.exclude(out)
1461      rownames(out) <- NULL
1462      return(out)
1463  }
1464
1465
1466
1467  meteo_min_max <- function(yr=1993) {
1468  # This function returns the minimum and maximum of the averages of the
1469  # monthly climatic data from year yr.
1470      tmp <- na.exclude(subset(meteo,year>=yr))
1471      vars <- c("Tm","Ta","Tx","Rf","rH","Sh","aH")
1472  #   mins <- sapply(c("Tm","Rf","rH","Sh","aH"),function(x)
1473      mins <- sapply(vars,function(x)
1474          apply(tapply(tmp[,x],
1475          list(tmp$year,tmp$station),min),2,mean,na.rm=T))
1476      colnames(mins) <- paste0(colnames(mins),"_min")
1477  #   maxs <- sapply(c("Tx","Rf","rH","Sh","aH"),function(x)
1478      maxs <- sapply(vars,function(x)
1479          apply(tapply(tmp[,x],
1480          list(tmp$year,tmp$station),max),2,mean,na.rm=T))
1481      colnames(maxs) <- paste0(colnames(maxs),"_max")
```

```
#    means <- sapply(c("Ta","Rf","rH","Sh","aH"),function(x)
    means <- sapply(vars,function(x)
        apply(tapply(tmp[,x],
        list(tmp$year,tmp$station),mean),2,mean,na.rm=T))
    colnames(means) <- paste0(colnames(means),"_mean")
    out <- as.data.frame(cbind(mins,maxs,means))
    out$station <- rownames(out)
    rownames(out) <- NULL
    out <- na.exclude(out)
    return(out)
}


max_power <- function(out,rge=c(.9,1.1),fct=max) {
# out: output of function "global_spect_calc".
# rge: a two-value vector giving the range of periods over which to apply
# function "fct".
    thenames <- names(out)
    if(length(rge)<2) {
        out <- sapply(out,function(x)max(x$globwave))
    }
    else out <- sapply(out,function(x) {
        p <- x$periods
        sel <- p>=rge[1] & p<=rge[2]
        return(fct(x$globwave[sel]))
    })
    return(data.frame(province=thenames,power=out))
}



meteo_power <- function(data=meteo,yr=1993,rge=c(.9,1.1),fct=max) {
# This functions calculates the power of the annual component of the
# climatic variables.
    variable <- "Ta"
    data <- subset(data,year>=yr)
    data <- data[order(data$year,data$month),]
```

```
# Scaling the climatic variables:
    for(i in unique(data$station)) {
        tmp <- subset(data,station==i)
        tmp <- apply(tmp[,4:10],2,scale)
        data[data$station==i,4:10] <- tmp
    }
    a <- global_spect_calc(data,space="station",var=variable,
        upperPeriod=2)
    a <- max_power(a,rge,fct)
    names(a) <- gsub("power",paste0(variable,"_pwr"),names(a))
    variable <- c("Tx","Tm","Rf","rH","Sh","aH")
    for(i in variable) {
        tmp <- global_spect_calc(data=subset(data,year>=yr),
            space="station",var=i,upperPeriod=2)
        tmp <- max_power(tmp,rge,fct)
        names(tmp) <- gsub("power",paste0(i,"_pwr"),names(tmp))
        a <- merge(a,tmp)
    }
    rownames(a) <- NULL
#   a <- na.exclude(a)
    names(a) <- gsub("province","station",names(a))
    return(a)
}



meteo_thresh <- function(yr=1993,var="Ta",lower=T,le=10) {
# This function finds the number of month below or above a given threshold
# for all the stations, for a given station.
    meteo <- subset(meteo,year>=yr)
    if(lower) {
        fct <- function(x,thresh) sum(x<thresh)
        x <- "l"
    }
    else {
        fct <- function(x,thresh) sum(x>thresh)
        x <- "u"
```

```
    }
    tmp <- meteo[,var]
    values <- round(seq(min(tmp,na.rm=T),max(tmp,na.rm=T),le=le))
    out <- sapply(values,function(y)
        tapply(tmp,meteo$station,function(x)fct(x,y)))
    out <- as.data.frame(out)
    names(out) <- paste0(var,"_",x,"_",values)
    return(out)
}



meteo_thresh_all <- function(yr=1993,le=50) {
    out <- meteo_thresh(yr,"Tm",T,le)
    out <- cbind(out,meteo_thresh(yr,"Tx",F,le))
    out <- cbind(out,meteo_thresh(yr,"Rf",T,le))
    out <- cbind(out,meteo_thresh(yr,"Rf",F,le))
    out <- cbind(out,meteo_thresh(yr,"rH",T,le))
    out <- cbind(out,meteo_thresh(yr,"rH",F,le))
    out <- cbind(out,meteo_thresh(yr,"Sh",T,le))
    out <- cbind(out,meteo_thresh(yr,"Sh",F,le))
    out <- cbind(out,meteo_thresh(yr,"aH",T,le))
    out <- cbind(out,meteo_thresh(yr,"aH",F,le))
    out$station <- rownames(out)
    rownames(out) <- NULL
    return(out)
}



find_station <- function(provinces=centroids93,stat=stations) {
# This function finds the stations that are the closest to each province:
    require(gmt)
    find1 <- function(lat,long,clim_stat) {
        nb <- nrow(clim_stat)
        station_names <- clim_stat$station
        latitudes <- clim_stat$latitude
```

```
1593          longitudes <- clim_stat$longitude
1594          distances <- sapply(1:nb,function(x)
1595              geodist(lat,long,latitudes[x],longitudes[x]))
1596          return(station_names[distances==min(distances)])
1597      }
1598      return(data.frame(province=provinces$province,
1599          station=sapply(1:nrow(provinces),function(x)
1600          find1(provinces[x,"latitude"],
1601          provinces[x,"longitude"],stat))))
1602  }
1603
1604
1605
1606
1607  clim_nbmonths <- function(data=meteo,yr=1993,var="Ta",thresh=20) {
1608  # This function finds the number of months that a given climatic variable
1609  # spends below a given threshold.
1610      data <- subset(data,year>=yr)
1611      thestations <- unique(data$station)
1612      fct <- function(x)
1613          return(sum(data[data$station==x,var]>thresh))
1614      nb <- sapply(thestations,fct)
1615      out <- data.frame(thestations,nb)
1616      names(out) <- c("station",paste0(var,"_nb"))
1617      return(out)
1618  }
1619
1620
1621  clim_nbmonths2 <- function(data1=ili_climate,data=subset(meteo,year>=1993),
         var="Ta",thresh=20) {
1622  # data1 is an output of the function "combined_data".
1623      thestations <- unique(data$station)
1624      fct <- function(x)
1625          return(sum(data[data$station==x,var]>thresh))
1626      nb <- sapply(thestations,fct)
1627      out <- data.frame(station=thestations,nb=nb)
1628      out <- merge(out,data1)
```

```
1629      out <- out[,c("nb","power")]
1630      with(out,plot(nb,power,xlim=c(0,250)))
1631      invisible(out)
1632  }
1633
1634
1635
1636  combined_data <- function() {
1637  # This function combines ILI and climatic data.
1638      out <- global_spect_calc(upperPeriod=2)
1639      power <- max_power(out)
1640      stations <- find_station()
1641  #   meteoPCA <- meteo_pca()
1642      meteominmax <- meteo_min_max()
1643      meteopower <- meteo_power()
1644      meteothresh <- meteo_thresh_all()
1645      out <- merge(stations,power)
1646  #   out <- merge(out,meteoPCA)
1647      out <- merge(out,meteominmax)
1648      out$T_amp <- with(out,Tx_max-Tm_min)
1649      out$Rf_amp <- with(out,Rf_max-Rf_min)
1650      out$rH_amp <- with(out,rH_max-rH_min)
1651      out$Sh_amp <- with(out,Sh_max-Sh_min)
1652      out$aH_amp <- with(out,aH_max-aH_min)
1653      out <- merge(out,meteopower)
1654      out <- merge(out,meteothresh)
1655      return(out)
1656  }
1657
1658
1659
1660  figure5 <- function(eps1=0,eps2=0,eps3=0,eps4=0,eps5=0,eps6=0,eps7=0,eps8
        =0,eps9=0,eps10=0,x1=.1,x2=1,y1=.1,y2=1,hspace=.1) {
1661  #pdf("figure_5.pdf",width=4,height=1.9,pointsize=8.5); figure5(eps1=.085,
        eps2=.035,eps3=.07,eps4=.011,eps5=.02,eps6=.014,eps7=.058,eps8=.015,
        eps9=.045,eps10=.01,x1=.08,x2=.98,y1=.175,y2=.92,hspace=.1); dev.off()
1662      require(tree)
```

```
#    a <- combined_data()
#this function to create tree regression in chapter 4
    width <- (x2-x1-hspace)/2
    opar <- par(plt=c(x1,x1+width,y1,y2))
    plot(1,type="n",
        xlim=range(modelout$yval),
#       xlim=c(0,8),
        ylim=range(modelout$y),axes=F,
        xlab="seasonlity of ILI (power)",ylab="deviance")
    axis(1); axis(2)
    size <- modelout$n/7.5
    marc <- "butt"
    segments(2.664609-eps1,190.03569,5.822562+eps2,190.03569,
        lwd=size[9],lend=marc)
    segments(2.664609,190.03569,2.664609,84.44272,lwd=size[11],
        lend=marc)
    segments(5.822562,190.03569,5.822562,84.44272,lwd=size[10],
        lend=marc)
    segments(2.423463-eps3,84.44272,3.870341+eps4,84.44272,
        lwd=size[11],lend=marc)
    segments(4.909161-eps5,84.44272,7.192664+eps6,84.44272,
        lwd=size[10],lend=marc)
    segments(2.423463,84.44272,2.423463,73.97544,lwd=size[7],lend=marc)
    segments(3.870341,84.44272,3.870341,73.97544,lwd=size[4],lend=marc)
    segments(4.909161,84.44272,4.909161,65.67093,lwd=size[6],lend=marc)
    segments(7.192664,84.44272,7.192664,65.67093,lwd=size[5],lend=marc)
    segments(2.211870-eps7,73.97544,3.481424+eps8,73.97544,
        lwd=size[7],lend=marc)
    segments(2.211870,73.97544,2.211870,67.25974,lwd=size[8],lend=marc)
    segments(3.481424,73.97544,3.481424,67.25974,lwd=size[2],lend=marc)
    segments(2.023522-eps9,67.25974,2.965262+eps10,67.25974,
        lwd=size[8],lend=marc)
    segments(2.023522,67.25974,2.023522,63.71225,lwd=size[1],lend=marc)
    segments(2.965262,67.25974,2.965262,63.71225,lwd=size[3],lend=marc)
    text(2.43,92,"a")
    text(2.2,82,"b")
    text(2.02,74,"c")
```

```
1700      text(6,90.5,"d")
1701      xxx <- (5.822562+2.664609)/2
1702      mtext(expression(paste("power of aH < 17.60")),at=xxx,line=-.1)
1703      bbb <- 5
1704      segments(xxx,190.03569-bbb,xxx,190.03569+bbb,lwd=2)
1705      mtext("(A)",at=1)
1706  # Panel (B):
1707      par(plt=c(x1+width+hspace,x2,y1,y2),new=T)
1708      require(RColorBrewer)
1709      colors <- brewer.pal(11,"Spectral")
1710      yyy <- global_aH_power@data@values
1711      marc <- as.numeric(cut(a$aH_pwr,seq(min(yyy,na.rm=T),max(yyy,na.rm=T),
             le=12)))
1712  print("ok")
1713      colors <- colors[marc]
1714      with(a,plot(aH_pwr,power,axes=F,
1715          xlab="seasonality of abs. humidity (power)",
1716          ylab="seasonality of ILI (power)",col=colors))
1717      axis(1); axis(2)
1718      abline(v=17.6,lty=2)
1719      mtext("(B)",at=6)
1720      par(opar)
1721  }
1722
1723
1724
1725  world_humidity <- function() {
1726      require(maptools) ## for "readShapePoly"
1727      require(raster) ## for "raster", "rotate", and "rasterize"
1728      require(maps) ## for "map"
1729      world <- readShapePoly("/home/choisy/Bureau/Travail/GIS/DIVA/World/
             countries.shp")
1730      first <- raster("ah_monthly_mean_1993_2010.tif",band=1)
1731      first <- rotate(first)
1732      first <- rasterize(world,first,mask=T)
1733      sel <- is.na(first@data@values)
1734      fct <- function(x) {
```

```
1735          out <- raster("ah_monthly_mean_1993_2010.tif",band=x)
1736          out <- rotate(out)
1737  #       out@data@values[which(sel)] <- NA
1738          a <- rasterize(world,out,mask=T)
1739      }
1740      nb <- 216
1741      out <- lapply(1:nb,fct)
1742  #return(out)
1743      b <- sapply(out,function(x)x@data@values)
1744      b <- na.exclude(b)
1745      b <- as.data.frame(t(b))
1746      b <- sapply(b,maxpower)
1747      c <- rep(NA,length(first@data@values))
1748      c[!sel] <- b
1749      out <- out[[1]]
1750      out@data@values <- c
1751      return(out)
1752  #map("world",ylim=c(-55,83))
1753  #image(global_aH_power,add=T,col=rev(terrain.colors(255)))
1754  #map("world",add=T)
1755  }
1756
1757
1758
1759  global_map <- function() {
1760  #pdf("figure_6a.pdf",width=5.41,height=2.25,pointsize=8.5); global_map();
        dev.off()
1761      require(maps) ## for "map"
1762      require(RColorBrewer) ## for "brewer.pal"
1763  #   colors <- rev(terrain.colors(255))
1764      colors <- brewer.pal(11,"Spectral")
1765      world <- map("world",plot=F)
1766      opar <- par(plt=c(0,.9,0,1))
1767      plot(world,ylim=c(-52,80),xlim=c(-167.5,180),type="n",axes=F,ann=F)
1768      image(global_aH_power,add=T,col=colors)
1769      lines(world)
1770      text(-176,81,"(A)")
```

```
1771      par(plt=c(.915,.935,.1,.9),new=T)
1772      x <- global_aH_power@data@values
1773      xval <- seq(min(x,na.rm=T),max(x,na.rm=T),le=1000)
1774      scale <- t(matrix(rep(xval,100),ncol=100))
1775      image(1:100,xval,scale,col=colors,axes=F,ann=F)
1776      axis(4); box()
1777      mtext("seasonality of absolute humidity (power)",4,line=1.5)
1778      par(opar)
1779  }
1780
1781
1782  global_map2 <- function() {
1783  #pdf("figure_6b.pdf",width=5.41,height=2.25,pointsize=8.5); global_map2();
         dev.off()
1784      require(maps) ## for "map"
1785      require(RColorBrewer) ## for "brewer.pal"
1786  #   colors <- rev(terrain.colors(255))
1787      colors <- rev(brewer.pal(11,"Spectral")[c(3,9)])
1788      world <- map("world",plot=F)
1789      opar <- par(plt=c(0,.9,0,1))
1790      plot(world,ylim=c(-52,80),xlim=c(-167.5,180),type="n",axes=F,ann=F)
1791      x <- global_aH_power@data@values
1792      global_aH_power@data@values <- x<16
1793      image(global_aH_power,add=T,col=colors)
1794      lines(world)
1795      text(-176,81,"(B)")
1796  #   par(plt=c(.915,.935,.1,.9),new=T)
1797  #   xval <- seq(min(x,na.rm=T),max(x,na.rm=T),le=1000)
1798  #   scale <- t(matrix(rep(xval,100),ncol=100))
1799  #   image(1:100,xval,scale,col=colors,axes=F,ann=F)
1800  #   axis(4); box()
1801  #   mtext("seasonality of absolute humidity (power)",4,line=1.5)
1802  ###
1803      par(cex=1.5)
1804      wit(subset(tamerius,periodicity=="annual"),points(longitude,latitude,
             pch=17,col="yellow"))
```

```
1805      with(subset(tamerius,periodicity=="annual"),points(longitude,latitude,
              pch=2))
1806      with(subset(tamerius,periodicity=="biannual"),points(longitude,latitude
              ,pch=19,col="blue"))
1807      with(subset(tamerius,periodicity=="biannual"),points(longitude,latitude
              ))
1808      par(opar)
1809  }
1810
1811  maxpower <- function(abs_hum) {
1812  # This function calculates the maximum power around the period of 1 year.
1813  # abs_hum is a time series of monthly averages of absolute humidity.
1814  # This time series is calculated from relative humidity and average
1815  # temperature by function "".
1816  # The wavelet transformations are made thanks to the functions pasted below
          .
1817      require(ondelettes)
1818  # Scaling the data:
1819      data <- scale(abs_hum)
1820  # Calculating the wavelet transform using Morlet wavelet:
1821      data <- morlet(data,dt=1/12,dj=1/100,
1822          lowerPeriod=.1,upperPeriod=3,pad=2^8)
1823      periods <- with(data,fourier*scale)
1824  # Calculating the global wavelet power:
1825      global_wave <- waveglobal(data)
1826  # Filtering between period .9 and 1.1 year:
1827      selection <- periods>.9 & periods<1.1
1828  # Returning the max power of the global wavelet around the period of 1 year
          :
1829      return(max(global_wave[selection]))
1830  }
1831
1832
1833
1834
1835  aha <- function() {
1836      sel <- 11:20
```

```
    foo <- NULL
    for(i in 1:(length(sel))) {
        tmp <- anova(lm(out[,c(3,c(sel[-i],sel[i]))]))
        foo <- rbind(foo,tmp[nrow(tmp)-1,])
    }
    return(foo)
}


## The end ##

################################################################################
```

```
plot.biwavelet <- function (x, ncol = 64, xlab = "Time", ylab = "Period",
    sig.level = 0.95,
    plot.cb = FALSE, plot.phase = FALSE, type = c("power.norm",
        "power", "wavelet", "phase"), plot.coi = TRUE, plot.sig = TRUE,
    bw = FALSE, legend.loc = NULL, legend.horiz = FALSE, arrow.size = 0.08,
    arrow.lwd = 2, arrow.cutoff = 0.9, xlim = NULL, ylim = NULL,
    xtick = TRUE, ytick = TRUE, form = "%Y",lwdcoi=1,lwdsig=1, ...)
{
    if (bw) {
        bw.colors <- colorRampPalette(c("black", "white"))
        fill.colors = bw.colors(ncol)
    }
    else {
        jet.colors <- colorRampPalette(c("#00007F", "blue", "#007FFF",
            "cyan", "#7FFF7F", "yellow", "#FF7F00", "red", "#7F0000"))
        fill.colors = jet.colors(ncol)
    }
    yrange = ylim
    y.ticks = 2^(floor(log2(min(x$period, yrange))):floor(log2(max(x$period
        ,
        yrange))))
    types = c("power.norm", "power", "wavelet", "phase")
    type = match.arg(tolower(type), types)
    if (type == "power.norm") {
```

```
23          if (x$type == "xwt") {
24              zvals = log2(abs(x$wave/(x$d1.sigma * x$d2.sigma)))
25              zlims = range(c(-1, 1) * max(zvals))
26              zvals[zvals < zlims[1]] = zlims[1]
27              locs = pretty(range(zvals), n = 5)
28              leg.lab = 2^locs
29          }
30          else if (x$type == "wtc") {
31              zvals = x$rsq
32              zlims = range(zvals)
33              zvals[zvals < zlims[1]] = zlims[1]
34              locs = pretty(range(zvals), n = 5)
35              leg.lab = locs
36          }
37          else {
38              zvals = log2(abs(x$power/x$sigma2))
39              zlims = range(c(-1, 1) * max(zvals))
40              zvals[zvals < zlims[1]] = zlims[1]
41              locs = pretty(range(zvals), n = 5)
42              leg.lab = 2^locs
43          }
44      }
45      else if (type == "power") {
46          zvals = log2(x$power)
47          zlims = range(c(-1, 1) * max(zvals))
48          zvals[zvals < zlims[1]] = zlims[1]
49          locs = pretty(range(zvals), n = 5)
50          leg.lab = 2^locs
51      }
52      else if (type == "wavelet") {
53          zvals = (Re(x$wave))
54          zlims = range(zvals)
55          locs = pretty(range(zvals), n = 5)
56          leg.lab = locs
57      }
58      else if (type == "phase") {
59          zvals = x$phase
```

```
60          zlims = c(-pi, pi)
61          locs = pretty(range(zvals), n = 5)
62          leg.lab = locs
63      }
64      else {
65          stop("type must be power, power.norm, wavelet or phase")
66      }
67      if (is.null(xlim))
68          xlim = range(x$t)
69      yvals = log2(x$period)
70      if (is.null(ylim))
71          ylim = range(yvals)
72      else ylim = log2(ylim)
73      image(x$t, yvals, t(zvals), zlim = zlims, xlim = xlim, ylim = rev(ylim)
            ,
74          xlab = xlab, ylab = ylab, yaxt = "n", xaxt = "n", col = fill.colors
                ,
75          ...)
76      box()
77      if (class(x$xaxis) == "Date") {
78          xlocs = pretty(x$t) + 1
79          if (xtick)
80              lab = format(x$xaxis[xlocs], form)
81          else lab = NA
82          axis(side = 1, at = xlocs, labels = lab)
83      }
84      else {
85          xlocs = axTicks(1)
86          if (xtick)
87              xticklab = xlocs
88          else xticklab = NA
89          axis(side = 1, at = xlocs, labels = xticklab)
90      }
91      axis.locs = axTicks(2)
92      if (ytick)
93          yticklab = format(2^axis.locs, dig = 1)
94      else yticklab = NA
```

```
95      axis(2, at = axis.locs, labels = yticklab)
96      if (plot.cb) {
97          image.plot(x$t, yvals, t(zvals), zlim = zlims, ylim = rev(range(
                yvals)),
98              xlab = xlab, ylab = ylab, col = fill.colors, smallplot = legend
                    .loc,
99              horizontal = legend.horiz, legend.only = TRUE, axis.args = list
                    (at = locs,
100                 labels = format(leg.lab, dig = 2)), xpd = NA)
101         box()
102     }
103     if (plot.coi) {
104 #       lines(x$t, log2(x$coi), lty = 1, lwd = lwdcoi, col = "white")
105     polygon(c(.75*x$t[1],x$t,1.25*tail(x$t,1)),c(3,log2(x$coi),3),col=rgb
            (1,1,1,.75),border="white")
106     box()
107     }
108     if (plot.sig & length(x$signif) > 1) {
109         if (x$type %in% c("wt", "xwt")) {
110             contour(x$t, yvals, t(x$signif), level = sig.level,
111                 col = "black", lwd = lwdsig, add = TRUE, drawlabels = FALSE
                        )
112         }
113         else {
114             contour(x$t, yvals, t(x$signif), nlevel = 1, col = "black",
115                 lwd = lwdsig, add = TRUE, drawlabels = FALSE)
116         }
117     }
118     if (plot.phase) {
119         a = x$phase
120         locs = which(zvals < quantile(zvals, arrow.cutoff))
121         a[locs] = NA
122         x.ind = seq(max(floor(x$dt/2), 1), length(x$t), length.out = 40)
123         y.ind = seq(max(floor(1/2), 1), length(x$period), length.out = 50)
124         phase.plot(x$t[x.ind], log2(x$period[y.ind]), a[y.ind,
125             x.ind], arrow.size = arrow.size, arrow.lwd = arrow.lwd)
126     }
```

```
}
```

```
analysis <- function(T,RH) {
# This function calculates the maximum power of the absolute humidity in
    one
# locality from time series of average temperature and relative humidity.
# T  : is a vector of monthly averages of daily average temperature in
#   Celcius degree from Jan. 1993 to Dec. 2010.
# RH : is a vector of monthly averages of daily average values of relative
#   humidity in percentage from Jan. 1993 to Dec. 2010.
# Calculating absolute humidity:
    abs_hum <- VP_calc(T,RH)
# Returning the maximum power around the period of one year:
    return(maxpower(abs_hum))
}


VP_calc <- function(T,RH) {
# This function calculates the absolute humidity from the relative humidity
# and the temperature, using the Clausius Clapeyron relation cited in
# Shaman & Kohn (2009).
# T  : temperature in Celcius degree.
# RH : relative humidity in percentage.
        esT0 <- 6.11 #(mb)
        T0 <- 273.15 #(K)
# Latent heat of evaporation for water:
        L <- 2257000 #(J/kg)
# Gaz constant for water vapor:
        Rv <- 461.5 #(J/(kg K))
        T <- T + T0 # Converting temperatures from Celcius to Kelvin.
# Shaman et al (2009)'s formula:
        esT <- esT0*exp((L/Rv-T0)*(1/T0-1/T))
        e <- esT*RH/100
# Returning the output:
        return(e)
}

```

```
35
36
37 maxpower <- function(abs_hum) {
38 # This function calculates the maximum power around the period of 1 year.
39 # abs_hum is a time series of monthly averages of absolute humidity.
40 # This time series is calculated from relative humidity and average
41 # temperature by function "".
42 # The wavelet transformations are made thanks to the functions pasted below
       .
43 # Scaling the data:
44     data <- scale(abs_hum)
45 # Calculating the wavelet transform using Morlet wavelet:
46     data <- morlet(data,dt=1/12,dj=1/100,
47         lowerPeriod=.1,upperPeriod=3,pad=2^8)
48     periods <- with(data,fourier*scale)
49 # Calculating the global wavelet power:
50     global_wave <- waveglobal(data)
51 # Filtering between period .9 and 1.1 year:
52     selection <- periods>.9 & periods<1.1
53 # Returning the max power of the global wavelet around the period of 1 year
       :
54     return(max(global_wave[selection]))
55 }
56
57
58 #
       ##############################################################################
59 # Below is a number of functions used from wavelet analysis:
60
61
62 wavepower <- function(object,time,from,to) {
63 # This function calculates the power of a wavelet decomposition.
64     if(is.list(object)) object <- object$wave
65     else if(!is.matrix(object)) stop(paste(
66         "'object should be either a matrix or a list as",
67         "outputed by a wavelet decomposition function'"))
```

```
68      if(!(missing(time) | missing(from) | missing(to)))
69          object <- object[,time>=from & time<to]
70      return(abs(object)^2)
71  }
72
73
74
75  waveglobal <- function(object,time,from,to) {
76  # This function calculates the global power spectrum of a wavelet.
77      if(missing(time) | missing(from) | missing(to))
78          return(var(object$y)*apply(wavepower(object),1,mean))
79      else {
80          time <- object$time
81          variance <- var(object$ts[time>=from & time<to])
82          power <- wavepower(object,time,from,to)
83          return(variance*apply(power,1,mean))
84      }
85  }
86
87
88
89  morlet <- function(y,dt,dj=.25,lowerPeriod,upperPeriod,pad,ko=6,linear=T) {
90  # This function performs a Morlet wavelet transform of the time series y
91  # over the time period defined by lowerPeriod and upperPeriod.
92
93  # Arguments:
94  # y      : input time series signal.
95  # dt         : sampling rate (e.g. 1/12 for monthy data and time unit
96  #          expressed in year.
97  # dj         : frequency resolution (i.e. inverse of the number of
98  #          sub-octaves in case of base 2 or the inverse of the number
99  #          of sub-scales within 1 Fourier factor year).
100 # lowerPeriod    : lower period of the decomposition.
101 # upperPeriod    : upper period of the decomposition.
102 # pad        : in case of zero padding (it must be a power of two).
103 # ko         : non-dimensional frequency of the Morlet mother wavelet.
104
```

```
# Value:
# wave      : wavelet transform matrix.
# period    : the vector of "Fourier" periods (in time units)
#         that corresponds to the scales.
# scale     : the vector of scale indices.
# coi       : the "cone-of-influence", which is a vector of n_y points
#         that contains the limit of the region where the wavelet
#         transform is influenced by edge effects.
# fourier   : the Fourier factor corresponding to the ko argument.

# General parameters:
    eps1 <- .49999 # for the base 2.
    eps2 <- 1e-5   # for the cone of influence.
    fourier_factor <- (4*pi)/(ko+sqrt(2+ko^2))
    if(missing(lowerPeriod)) so <- 2*dt
    else so <- lowerPeriod/fourier_factor
# Length of the time series before padding:
    n1 <- length(y)
# Zero padding:
    if(pad==0) {
# Pad with zeros to the nearest power of 2 to N:
        base2 <- trunc(log2(n1)+eps1)
        x <- c(y,rep(0,2^(base2+1)-n1))
        pad <- length(x)
    }
    else if(pad>0) {
# Pad with zeros with a specified length of the new time series:
        base2 <- log2(pad)
        if (base2%%1) stop("pad must be a power of two")
        else {
            if(pad<n1) warning("pad is too low: no padding")
            x <- c(y,rep(0,max(0,2^(base2)-n1)))
        }
    }
# Length of the time series after padding:
    n <- length(x)
# Creating the vector of scales:
```

```
142     if(linear) {
143 # Linear repartition of scales:
144         fo <- min(upperPeriod/fourier_factor,n*dt)
145         scale <- seq(so,fo,dj)
146     }
147     else {
148 # Power of 2 repartition of scales:
149 # The largest possible number of scales:
150         largestNumberofScales <- trunc(log2(n*dt/so)/dj)
151 # The current number of scales:
152         currentNumberofScales <- trunc(log2(upperPeriod/so)/dj)
153 # If upperPeriod is too long
154         j1 <- min(currentNumberofScales,largestNumberofScales)
155         scale <- so*2^((0:j1)*dj)
156     }
157 # Creating the vector of angular frequencies (phases) (equation 5):
158     k <- 1:trunc(n/2)
159     k <- k*((2*pi)/(n*dt))
160     k <- c(0,k,-k[floor((n-1)/2):1])
161     ventana <- length(k)
162 # Fourier transform of the time series (equation 3):
163     f <- fft(x)
164 # Calculating the matrix of wavelet transform:
165     wave <- lapply(scale,function(x) {
166 # The daughter Morlet wavelet for the specified scale (table 1):
167         daughter <- pi^(-0.25)*(k>0)*exp(-(x*k-ko)^2/2)
168 # Normalisation (equation 6):
169         daughter <- sqrt(x*k[2]*ventana)*daughter
170 # Fourier inverse transform (equation 4):
171 # Note that the FFT of a (complex) Morlet wavelet is real, thus its
172 # conjugate is equal to itself. Note also that because of the
173 # normalization of the FFT in the R fft function, we need to divide
174 # the result by the length of the fft (see help(fft)).
175         return(fft(f*daughter,inverse=T)/length(f*daughter))
176     })
177     wave <- matrix(unlist(wave),byrow=T,nrow=length(wave))
178 # Caculating the cone of influence:
```

```
179     coi <- fourier_factor*dt*c(eps2,1:((n1+1)/2-1),
180         rev((1:(n1/2-1))),eps2)/sqrt(2)
181 # Give the output after getting rid of the padded zeros:
182     return(list(wave=wave[,1:n1],scale=scale,coi=coi,
183         fourier=fourier_factor,y=y,dt=dt,dj=dj,pad=pad,
184         lowP=lowerPeriod,upP=upperPeriod,ko=ko,linear=linear))
185 }
```

Appendix E

Supplementary research paper

Title: Influenza Infection Rates, Measurement Errors and the Interpretation of Paired Serology

Author(s): Simon Cauchemez*, Peter Horby, Annette Fox, Le Quynh Mai, Le Thi Thanh, Pham Quang Thai, Le Nguyen Minh Hoa, Nguyen Tran Hien, Neil M. Ferguson.

Journal/Publisher: PLOS Pathogens

Type of publication: Major article

Stage of publication: Published

Academic peer-reviewed: Yes

Copyright: Permission obtained from the publisher.

Candidate's role: I conceived of the study and supervised the collection and collation of all the data. I prepared the data for analysis. I am taken part in wrote the first and all subsequent drafts of the manuscript and responded to all reviewers comments.

Candidate's signature:

Date: 25 April 2014

Full Name: **Pham Quang Thai**

Supervisor or senior author's signature to confirm Candidates role:

OPEN ACCESS Freely available online

PLOS | PATHOGENS

Influenza Infection Rates, Measurement Errors and the Interpretation of Paired Serology

Simon Cauchemez[1]*, Peter Horby[2], Annette Fox[2], Le Quynh Mai[3], Le Thi Thanh[3], Pham Quang Thai[3], Le Nguyen Minh Hoa[2], Nguyen Tran Hien[3], Neil M. Ferguson[1]

1 MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Imperial College London, London, United Kingdom, **2** Oxford University Clinical Research Unit - Wellcome Trust Major Overseas Programme, Hanoi, Vietnam, **3** National Institute of Hygiene and Epidemiology, Hanoi, Vietnam

Abstract

Serological studies are the gold standard method to estimate influenza infection attack rates (ARs) in human populations. In a common protocol, blood samples are collected before and after the epidemic in a cohort of individuals; and a rise in haemagglutination-inhibition (HI) antibody titers during the epidemic is considered as a marker of infection. Because of inherent measurement errors, a 2-fold rise is usually considered as insufficient evidence for infection and seroconversion is therefore typically defined as a 4-fold rise or more. Here, we revisit this widely accepted 70-year old criterion. We develop a Markov chain Monte Carlo data augmentation model to quantify measurement errors and reconstruct the distribution of latent *true* serological status in a Vietnamese 3-year serological cohort, in which replicate measurements were available. We estimate that the 1-sided probability of a 2-fold error is 9.3% (95% Credible Interval, CI: 3.3%, 17.6%) when antibody titer is below 10 but is 20.2% (95% CI: 15.9%, 24.0%) otherwise. After correction for measurement errors, we find that the proportion of individuals with 2-fold rises in antibody titers was too large to be explained by measurement errors alone. Estimates of ARs vary greatly depending on whether those individuals are included in the definition of the infected population. A simulation study shows that our method is unbiased. The 4-fold rise case definition is relevant when aiming at a specific diagnostic for individual cases, but the justification is less obvious when the objective is to estimate ARs. In particular, it may lead to large underestimates of ARs. Determining which biological phenomenon contributes most to 2-fold rises in antibody titers is essential to assess bias with the traditional case definition and offer improved estimates of influenza ARs.

Citation: Cauchemez S, Horby P, Fox A, Mai LQ, Thanh LT, et al. (2012) Influenza Infection Rates, Measurement Errors and the Interpretation of Paired Serology. PLoS Pathog 8(12): e1003061. doi:10.1371/journal.ppat.1003061

Editor: Ron A. M. Fouchier, Erasmus Medical Center, Netherlands

Received August 9, 2012; **Accepted** October 14, 2012; **Published** December 13, 2012

Funding: This work was supported by research grants from the Wellcome Trust (grants 081613/Z/06/Z and 077078/Z/05/Z), the NIH MIDAS program, EU FP7 EMPERIE and PREDEMICS projects and the MRC. SC also thanks Research Council UK. SC received consulting fees from Sanofi Pasteur MSD for a project on the modelling of varicella zoster virus transmission. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: SC received consulting fees from Sanofi Pasteur MSD for a project on the modelling of the transmission of varicella zoster virus (i.e. different subject than submission). This does not alter our adherence to all PLOS Pathogens policies on sharing data and materials.

* E-mail: s.cauchemez@imperial.ac.uk

Introduction

Each year, seasonal influenza is responsible for about three to five millions severe illnesses and about 250,000 to 500,000 deaths worldwide [1]. These epidemics can generate important economic losses due to high levels of worker absenteeism as well as a saturation of emergency services at the peak of the epidemic [1]. In addition, avian or swine influenza viruses occasionally adapt to humans and generate influenza pandemics like in 1918, 1957, 1968 and 2009, sometimes with catastrophic consequences like in 1918, when 20 to 50 million people died worldwide.

Appropriate assessment of the epidemiological characteristics of the influenza virus is important to guide control policies. In particular, this requires being able to track the number of influenza cases with severe clinical outcomes (*i.e.* the tip of the severity pyramid) as well as the total number of people infected by an influenza virus (*i.e.* the base of the severity pyramid). For example, the case fatality ratio (proportion of influenza cases who die) is a key measure of severity that informs decision making during influenza pandemics, and takes the number of influenza related death as numerator and the number of influenza cases as denominator. Estimates of infection attack rates are also essential for characterizing the spread of the virus in human populations in order to predict epidemic trajectory, the potential impact of control measures such as social distancing measures, and the likelihood and magnitude of subsequent epidemics arising from continued circulation of the same virus [2,3].

Although it is usually possible to estimate the number of severe influenza cases from sentinel surveillance (e.g. based on data collected at medical practices, clinics or hospitals), it is much harder to estimate the total number of people infected by an influenza virus. First, a substantial proportion of influenza infections are asymptomatic [4,5]. Second, among those with symptoms, only a proportion seek healthcare; and this proportion may vary from season to season or even during the course of an epidemic. Last, Influenza-Like-Illness (ILI) symptoms are not specific to influenza. So, a substantial proportion of patients consulting for ILI may not have been infected by an influenza virus.

Serological studies have become the gold standard approach for estimating influenza infection attack rates due to the difficulty of

Author Summary

Each year, seasonal influenza is responsible for about three to five million severe illnesses and about 250,000 to 500,000 deaths worldwide. In order to assess the burden of disease and guide control policies, it is important to quantify the proportion of people infected by an influenza virus each year. Since infection usually leaves a ''signature'' in the blood of infected individuals (namely a rise in antibodies), a standard protocol consists in collecting blood samples in a cohort of subjects and determining the proportion of those who experienced such rise. However, because of inherent measurement errors, only large rises are accounted for in the standard 4-fold rise case definition. Here, we revisit this 70 year old and widely accepted and applied criterion. We present innovative statistical techniques to better capture the impact of measurement errors and improve our interpretation of the data. Our analysis suggests that the number of people infected by an influenza virus each year might be substantially larger than previously thought, with important implications for our understanding of the transmission and evolution of influenza – and the nature of infection.

estimating infection rates by other means. Although cross-sectional serological surveys can provide valuable and timely information, paired blood samples collected before and after an epidemic in a cohort of individuals is the optimal approach for precisely assessing infection rates. The haemagglutination-inhibition (HI) assay remains the most commonly used approach for detecting serological evidence of recent influenza infection [6–12]. The assay detects the presence of antibodies that prevent the haemagglutinin protein of the influenza virus from agglutinating red blood cells [13,14]. For each serum sample, antibody titers are expressed as the reciprocal of the highest serum dilution that can still prevent a fixed concentration of virus from agglutinating red blood cells. A rise in antibody titers between the first and second blood is taken as a marker of infection. However, because the procedure is susceptible to measurement errors, a 2 fold rise (that is a 1-dilution increase) is usually considered as insufficient evidence for infection. Seroconversion is therefore typically defined as a 4-fold rise (*i.e.* a 2-dilutions increase) or more in antibody titers. This ad-hoc rule became established when these methods were first developed and is now widely adopted [15,16]. In the meantime, however, statistical methods for addressing measurement errors have made substantial progress. In particular, there is now an extensive body of literature on methods to ensure that the presence of measurement errors does not bias estimates of key parameters of interest. Given these developments, it is timely to revisit the way serological data are interpreted.

Central to the traditional approach to analyzing serological data is the belief that data about 2-fold rises provide no information since such increases can be caused by frequent measurement errors. This concern about measurement errors is certainly relevant when trying to make specific diagnoses for individual cases. For example, one may be averse to the risk of false positives; but less so to the risk of false negatives. However, estimating infection attack rates at the population level is a very different aim from setting up a specific diagnostic tool, and may benefit from a different use of the data.

First, it is important to note that estimating infection attack rates is not just a matter of specificity (*i.e.* ensuring that subjects satisfying the diagnostic definition of infection were indeed infected by an influenza virus) but also a matter of sensitivity (*i.e.* ensuring that all subjects infected are diagnosed as such). An approach that favours specificity over sensitivity may lead to underestimating infection attack rates.

A second important observation is that, even in a context of frequent 2-fold errors, data about 2-fold rises may still be informative. Consider for example a situation where all individuals exhibit a 2-fold rise during the season: such a pattern cannot be explained by measurement error alone since measurement errors are made *both* at baseline and post-epidemic and should be about equally distributed provided the sample size is sufficiently large.

Here, we explore how modern statistics for the analysis of data with measurement errors can change and improve our interpretation of serology. We present a new method to quantify errors in the measurement of antibody titers and to estimate the true distribution of paired serological measurements corrected for measurement errors. The methodology is applied to data collected in a cohort study conducted in Vietnam between 2007 and 2009.

Results

Measurement errors

We estimate that the 1-sided probability of a 2-fold error was 9.3% (95% CI: 3.3%, 17.6%) when the true antibody titer was below detection levels, rising to 20.2% (95% CI: 15.9%, 24.0%) otherwise (posterior probability that latter larger than former: 98.7%). There was a satisfying fit of the model to replicate measurement data (Figure 1). The model where measurement errors were independent of true antibody titers failed to fit the data (Figure S2 and Supplementary Material).

Distribution of true paired serology

Figure 2 summarizes the distribution of paired serology, corrected for measurement errors for the different seasons (2008, Spring 2009, Autumn 2009) and subtypes (H1N1, H3N2 and B). A range of observations can be made.

The first observation concerns 2-fold rises in antibody titers between baseline and post serology (yellow bars). Such increases are usually ignored in analyses because 2-fold errors are common. In some instances, like for example subtypes H3N2 and B in 2008 and H1N1pdm09 in Autumn 2009, 2-fold rises appeared negligible and at levels that could be generated by measurement errors alone, since 0 was within the 95% CI of the estimated proportion of subjects having a 2-fold rise (Figures 2B, 2C, 2G). In other instances, however, the proportion of individuals experiencing a 2-fold rise ranged from 20% to 33% with lower bounds of the 95% CIs above 0 (range: 7%–23%), indicating that these rises cannot be solely explained by measurement errors. Assuming that most of these 2-fold rises were due to infection, our estimate of infection attack rates $AR_{\geq 2f.r.}$ for H1N1 in 2008 and H1N1, H3N2 and B in Spring 2009 would be dramatically higher than traditional estimate $AR_{\geq 4f.r.}$ based on 4-fold rises or more (Figure 3A). So, even if only a proportion of the 2-fold rises were due to influenza infections, the traditional estimate $AR_{\geq 4f.r.}$ might still represent a substantial underestimate of the true infection attack rates

The fact that $AR_{\geq 2f.r.}$ and $AR_{\geq 4f.r.}$ were very similar for H3N2 and B in 2008 and virtually identical for H1N1pdm09 in Autumn 2009 (Figure 3A) highlights important heterogeneities in the way antibody titers increase by season/subtype (Figure 3B). For example, for H1N1pdm09 in Autumn 2009, almost all those experiencing a rise in antibody titers exhibited a 4-fold rise or more; but for H1N1 in 2008, most of those experiencing a rise only had a 2-fold increase. The absence of a simple linear relationship between $AR_{\geq 4f.r.}$ and the proportion of 2-fold rises

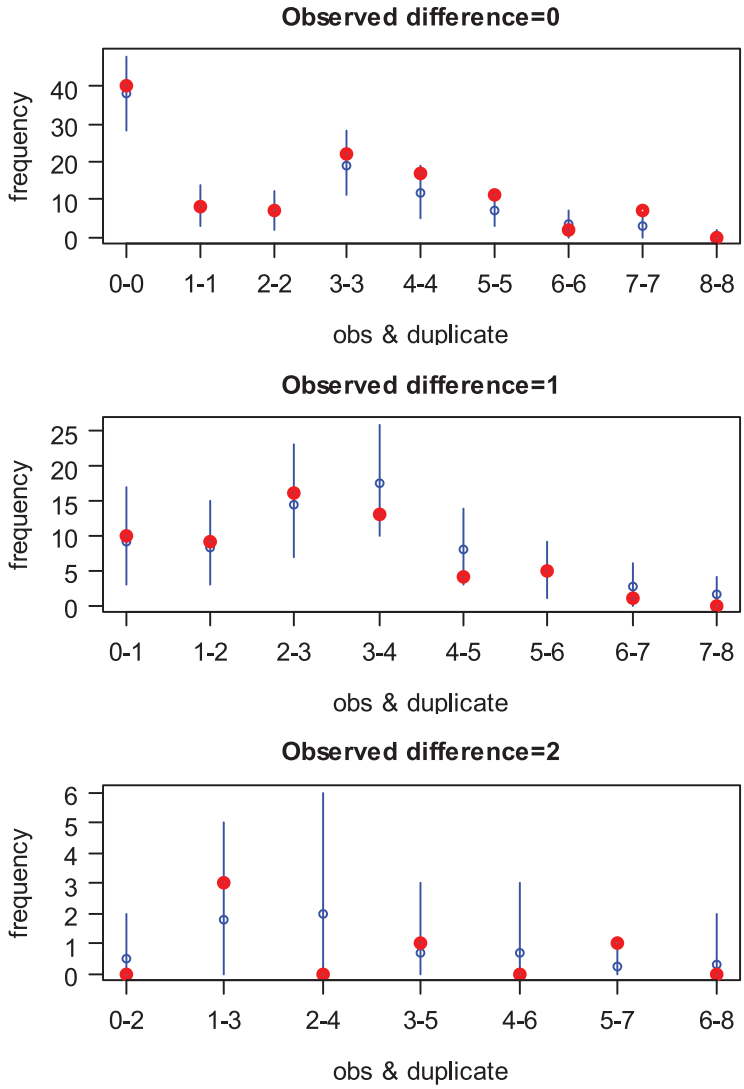


Figure 1. Fit of the model to data on replicate measurements. Observed (red point) and expected (mean: blue point/95% CI: blue bar) number of pairs (observed AT level, replicate AT level). Pairs are sorted by panel according to the number of dilution difference between the observed and the replicate measurement.
doi:10.1371/journal.ppat.1003061.g001

suggests that the standard approach of inflating $AR_{\geq 4f.r.}$ by a fixed proportion (generally equal to the proportion of PCR positive cases who do not seroconvert; around 10–20%) to get corrected estimates of infection attack rates may be inappropriate. Rather, corrections might have to be applied on a season-to-season and subtype-to-subtype basis.

The last notable observation is that decay in antibody titers is observed. For example, 30% (95% CI: 22, 36) of individuals exhibited a decay for subtype H3N2 in 2008.

PCR positive cases

Figure 4 shows the observed rise in antibody titers for PCR positive cases. Twenty seven percent of these cases experienced no rise or only a 2-fold rise in titer during the season. This again suggests that the case definition of a 4-fold rise or more may underestimate attack rates by at least 27%. PCR positive cases with low baseline titers experienced an average increase significantly larger than those with higher baseline titers ($p = 0.026$) (Figure 4) [17,18].

Cross-reactivity between subtypes

Simulations were run to test the hypothesis of an absence of cross-reactivity between subtypes H1N1, H3N2 and B in 2008 and Spring 2009 (see Supplementary Material). We found that there was good adequacy between the data and patterns that would be obtained in the absence of cross-reactivity. The hypothesis of an absence of cross-reactivity could therefore not be rejected (Figure S3).

Model fitting

Figure 5 compares the distribution of observed paired serology as observed in the data (black point) and as predicted by the model. Model fit was satisfactory.

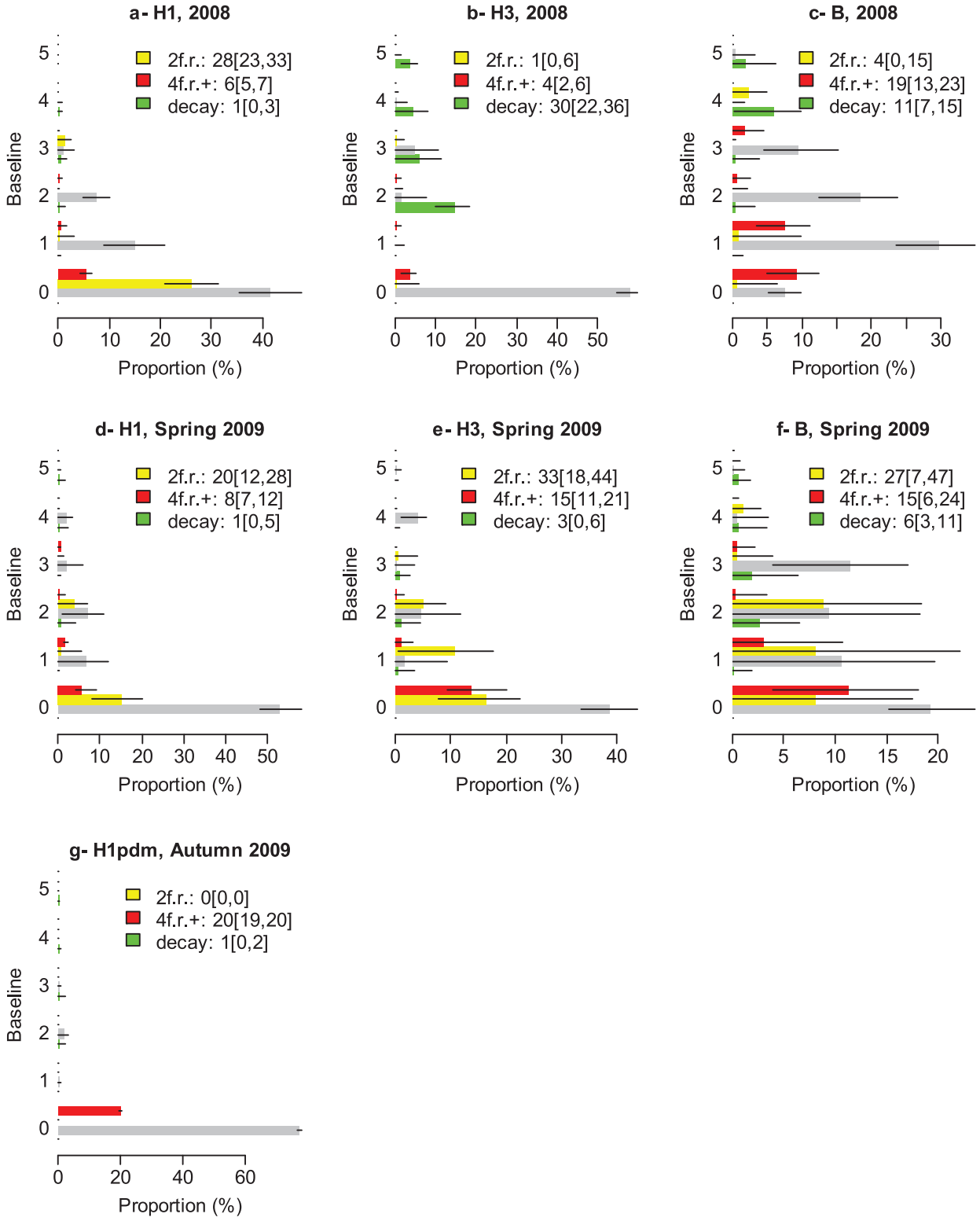


Figure 2. Distribution of paired serology, corrected for measurement errors as a function of season (2008, Spring 2009, Autumn 2009) and subtype (H1N1, H3N2 and B) (in Autumn 2009, subtyping was only conducted for H1N1pdm09). In each panel, individuals are sorted by baseline AT levels on the y-axis. For a given baseline, the grey bar indicates the expected proportion of individuals with post AT level equal to baseline AT level; the yellow bar indicates the proportion with a 2 fold rise (2f.r.); the red bar indicates the proportion with a 4 fold rise or more (4f.r.+); the green bar indicates the proportion with a decay. The black thin lines give the 95% CI. The legend gives the mean [95% CI]. **A**: H1N1, 2008. **B**: H3N2, 2008. **C**: B, 2008. **D**: H1N1, Spring 2009. **E**: H3N2, Spring 2009. **F**: B, Spring 2009. **G**: H1N1pdm09, Autumn 2009.
doi:10.1371/journal.ppat.1003061.g002

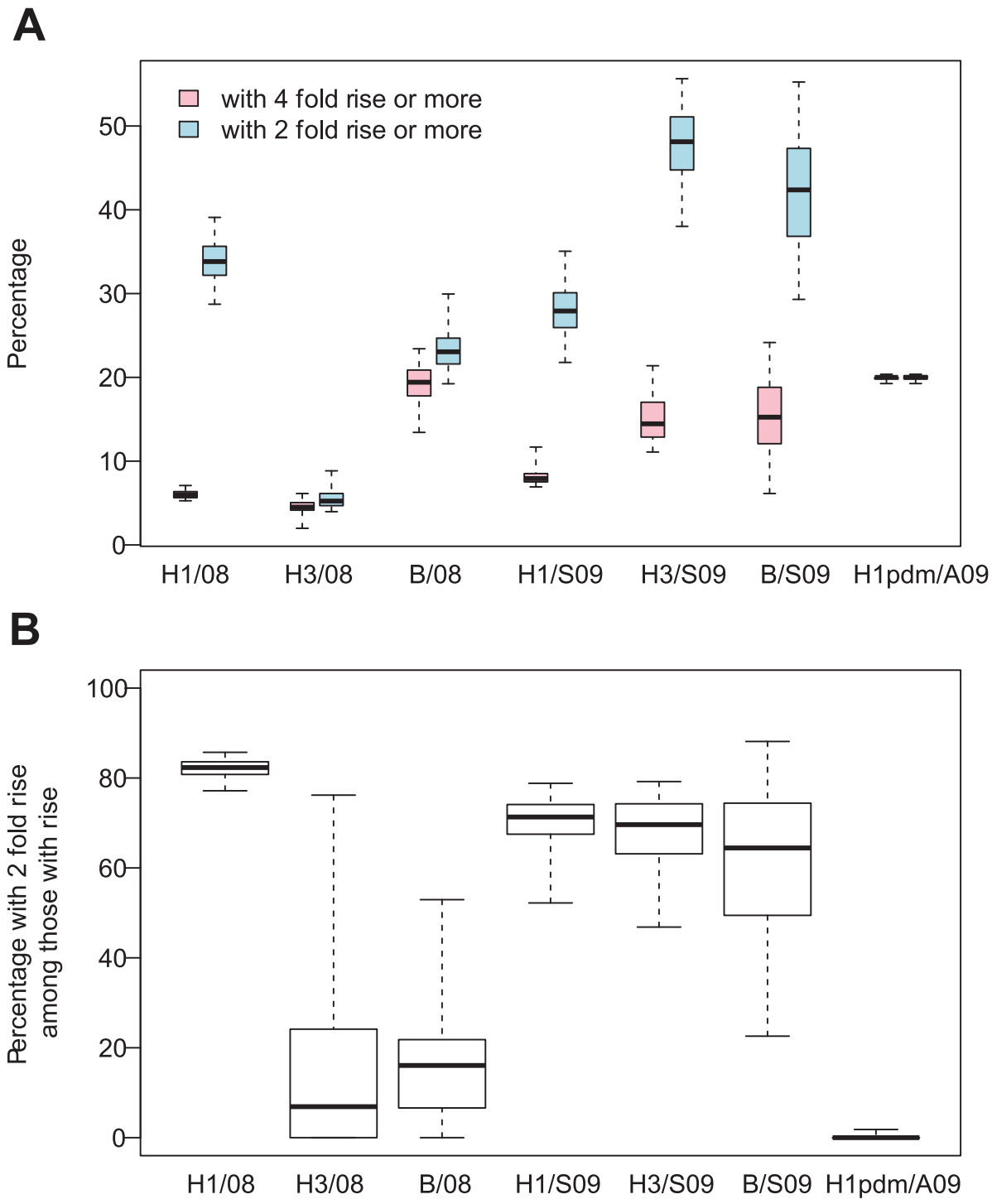


Figure 3. Increases in antibody titers. A: Posterior distribution of the percentage of subjects with a 4 fold rise or more in AT (pink) and with a 2 fold rise or more in AT (blue) for the different subtypes and the different seasons (2008 (08), Spring 2009 (S09), Autumn 2009 (A09)). **B**: Posterior distribution of the percentage of subjects with a 2 fold rise in AT among those with a rise in AT. Boxplots give percentiles 2.5%, 25%, 50%, 75%, 97.5% of the distribution.
doi:10.1371/journal.ppat.1003061.g003

Simulation study

In a simulation study, we found that estimates of parameters characterizing measurement errors were unbiased (Table 1), as well as those characterizing the selection process (Table S2). We also found that estimates of the proportion of subjects with an antibody titer increase (empirical absolute bias: 0.1%), of the proportion of subjects with an antibody titer decay (empirical absolute bias: 0.0%) and of the probabilities characterizing

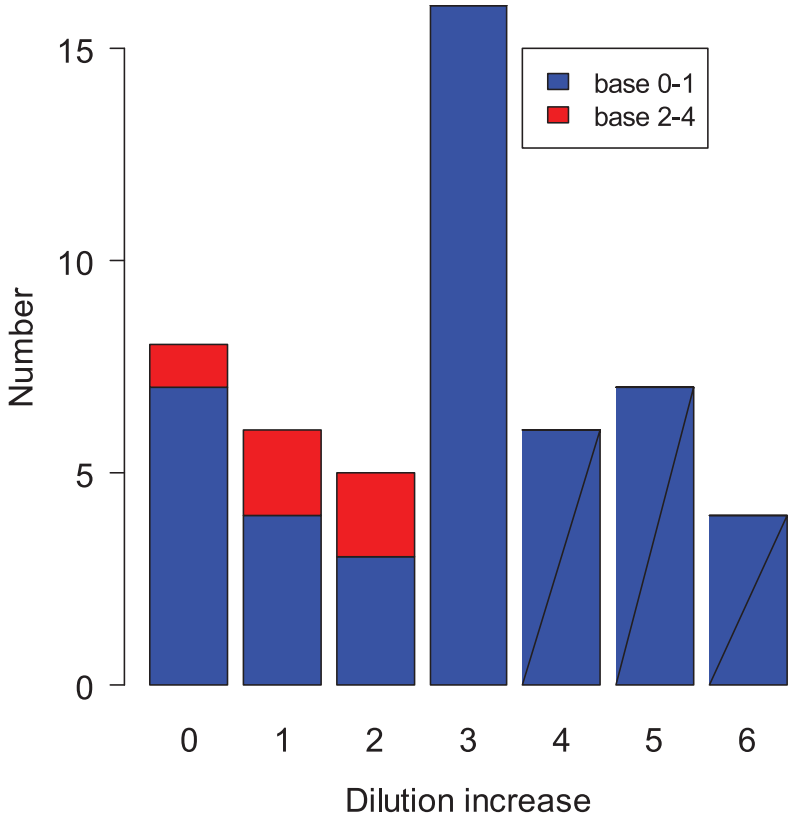


Figure 4. Distribution of observed increase in PCR positive cases as a function of baseline. Individuals with a low antibody titer baseline (0–1) are in blue; those with a higher baseline (2–4) are in red.
doi:10.1371/journal.ppat.1003061.g004

jointly baseline antibody titers and the change in antibody titers during a season (empirical absolute bias: 0.0%) were unbiased (Figure 6).

Age-specific patterns

Our statistical model describes the distribution of paired serology across all subjects. However, since we infer true paired serology for each individual, it is possible to reconstruct a posteriori the distribution of true paired serology for the different age groups. The age-specific distributions for true paired serology are presented in Figure S4. Interesting differences can be noticed between age groups. For example and consistent with the literature, for H1N1pdm09 in Autumn 2009, the proportion of 4-fold rises falls from 39% (95% CI: 37%, 39%) in <18 y.o. to 15% (95% CI: 15%,16%) in 18–48 y.o. and 8% (95% CI: 7,9) in >48 y.o. For H3N2 in 2009, the decay in antibody titers was more important among <18 y.o. (53%; 95% CI: 38%, 65%) than among older age groups (25%, 95% CI 19%, 30% for 18–48 y.o. and 18%, 95% CI 12, 22 for >48 y.o.). For H3N2 in Spring 2009, although the proportions of 4-fold rises were similar across age groups, our analysis suggests that the proportion of 2-fold rises may have been higher among <18 y.o (43%, 95% CI: 23, 58) than in other age groups (30%, 95% CI 17%, 41% for 18–48 y.o. and 27%, 95% CI 13, 38 for >48 y.o.). We find that, for each age group, there is a satisfying adequacy between the observed distribution of paired serology and that predicted by the model (Figure S5).

Discussion

In this paper, we have revisited the traditional interpretation of paired serological measurements of influenza antibody titers. Until now, data on 2-fold rises have been largely ignored because of the belief that measurement errors made them unreliable. Although this may be a valid concern if the aim is to get a specific diagnosis for individual cases, we argue that this is less so when the objective is to interpret antibody titer variations at the population level. We have shown that it is possible to quantify measurement errors, and to reconstruct the distribution of paired serology corrected for measurement errors. Our method gave unbiased estimates in a simulation study.

After correction for measurement errors for the Vietnamese data examined here, we found that for some seasons and subtypes the proportions of individuals with 2-fold rises in antibody titers was too large to be explained by measurement errors alone. Estimates of infection attack rates varied greatly depending on whether or not 2-fold rises were included. It is therefore important to determine the biological phenomenon that could cause such increases, in particular whether they are caused by exposure to influenza viruses.

A first hypothesis is that 2-fold titer increases are caused by infection by an influenza virus. In support of this hypothesis, it is clear that a proportion of virologically- or RT-PCR- confirmed influenza cases do not achieve a 4-fold rise in HI titer. This proportion was 27% in our dataset, similar to a large cohort of confirmed pandemic cases in the US [19]. However, past work has shown this proportion to be as high as 77% in people who have

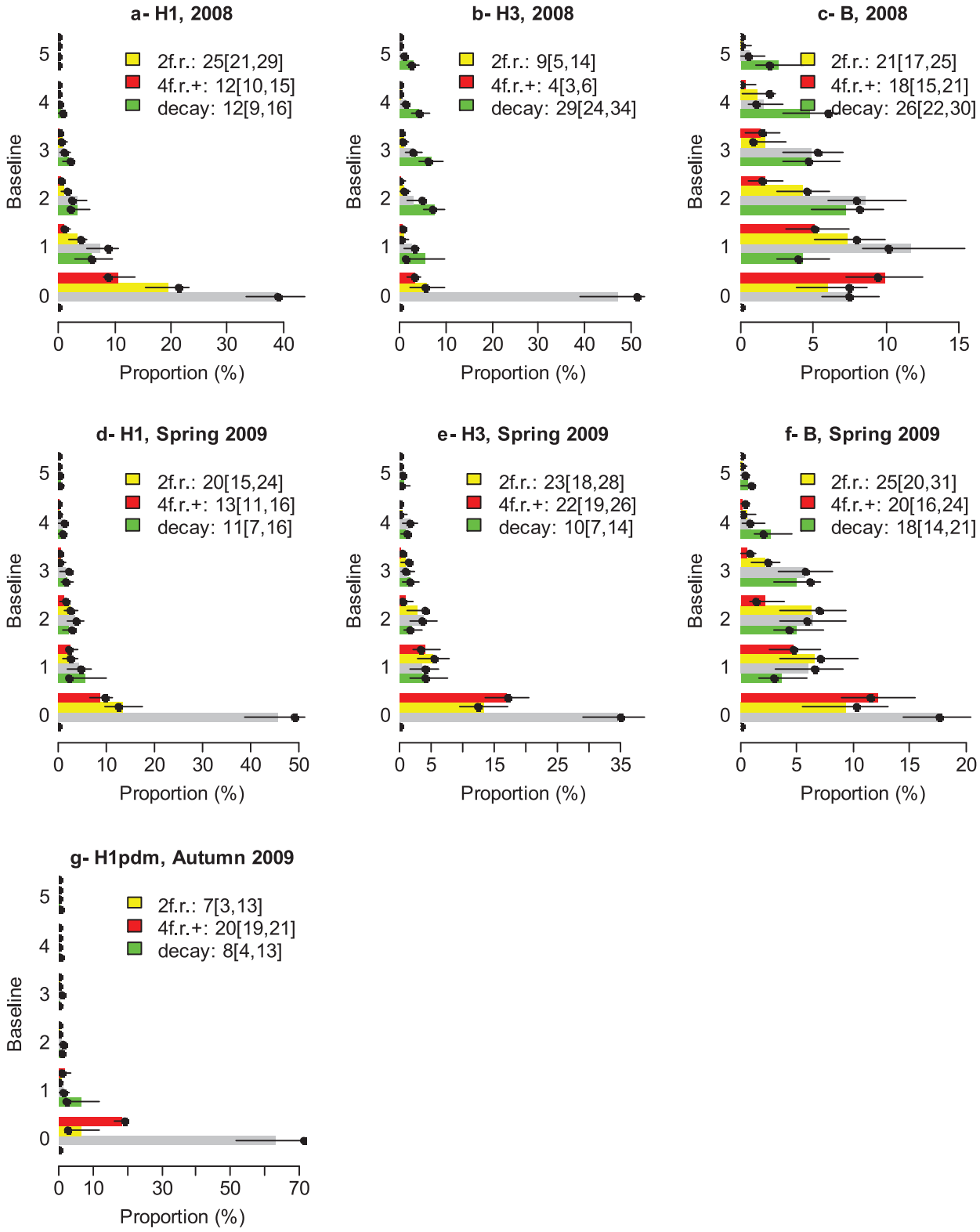


Figure 5. Model adequacy to the data. Distribution of "observed" paired serology as predicted by the model (color bars) and as observed in the data (black point) as a function of season (2008, Spring 2009, Autumn 2009) and subtype (H1N1, H3N2 and B). In each panel, individuals are sorted by baseline AT levels on the y-axis. For a given baseline, the grey bar indicates the expected proportion of individuals with post AT level equal to baseline AT level; the yellow bar indicates the proportion with a 2 fold rise (2f.r.); the red bar indicates the proportion with a 4 fold rise or more (4f.r.+); the green bar indicates the proportion with a decay. The black thin lines give the 95% CI. The legend gives the mean [95% CI]. **A**: H1N1, 2008. **B**: H3N2, 2008. **C**: B, 2008. **D**: H1N1, Spring 2009. **E**: H3N2, Spring 2009. **F**: B, Spring 2009. **G**: H1N1pdm09, Autumn 2009.
doi:10.1371/journal.ppat.1003061.g005

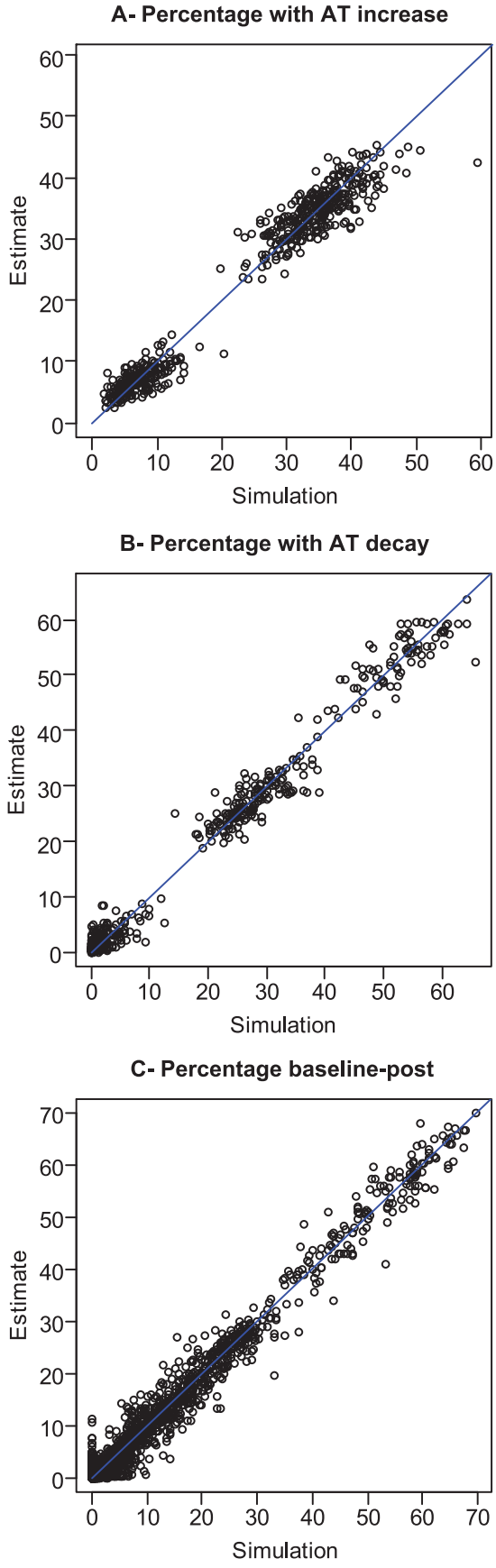


Figure 6. Performance of the method to reconstruct the true distribution of paired serology. Eighty datasets are simulated with known parameters (see Methods). **A**: Estimated percentage of subjects with an increase in antibody titers as a function of the true percentage in the simulated dataset. **B**: Estimated percentage of subjects with a decay in antibody titers as a function of the true percentage in the simulated dataset. **C**: Estimated probabilities characterizing jointly baseline AT level and the change in AT level during the epidemic _ similar to those presented in Figure 1 _ as a function of the true probability in the simulated dataset.
doi:10.1371/journal.ppat.1003061.g006

high pre-existing antibody titers [17], or as low as 10% in patients seeking medical care for pandemic H1N1 infection in 2009 [20]. It is clear that antibody titer changes following infection vary between individuals and are affected by factors including pre-existing titer and timing of serum collection. In particular, since there is an upper limit to antibody concentrations, individuals with high pre-existing titers are limited in their ability to generate 4-fold rises and may produce only a 2-fold titer increase in response to infection [15]. However, the analysis performed here shows that 2 fold titer changes are common even among individuals with low pre-existing titers. Antibody concentrations reach a peak 4–7 weeks after infection and then decay over a period of around six months to a plateau that is maintained for several years [21]. Although the profile of HA antibody decay is not well characterised, the probability of detecting 2- or 4- fold rises will vary with the interval following infection. However, in our data the longest interval between the peak transmission period and blood sampling was in season 3, when the proportion of 2-fold titer rises was lowest.

A second hypothesis is that 2-fold rises correspond to infection which is attenuated by mucosal or serological antibodies to homologous or heterologous strains, or by innate or cell mediated immunity. Antibody responses to inactivated influenza vaccines clearly demonstrate the potential for antigenic stimulation without active infection and the phenomenon of boosting of immunity in exposed yet uninfected individuals is well documented for other viruses (*e.g.* varicella zoster [22]).

A third hypothesis is that 2-fold rises are an artefact unrelated to influenza infection or exposure. Seasonal variation in titres independent of infection might result from the presence of non-specific inhibitors of agglutinination. For example, this could happen if the circulation of other viruses boosted the immune system, leading to small increases in all antibody titers. In such a scenario, one might expect the effect to be similar on the different subtypes. However, in 2007, a large proportion of individuals exhibited 2-fold increases for H1N1 but not for H3N2 or B, suggesting that this hypothesis is not strongly supported by the data.

Table 1. Performance of the method to estimate parameters characterizing measurement errors.

	p_0	p_1	ε
Simulation value	9.0%	20.0%	0.50%
Mean estimate (SD)	9.5% (4.1%)	19.8% (2.3%)	0.065% (0.21%)

p0: probability of a 1-sided 1-dilution error if true AT level is $=0$.
p1: probability of a 1-sided 1-dilution error if true AT level is >0.
ε: probability that measurement goes wrong and that observed AT level is Uniformly drawn in (0,…,K).
Eighty datasets are simulated with known parameters (see Methods). The table gives the simulation value of parameters and the mean (standard deviation) of estimates.
doi:10.1371/journal.ppat.1003061.t001

It is also important to understand why 2-fold titers changes were prominent during some seasonal influenza epidemics but not during the pandemic. One possibility may be that there was greater antigenic mismatch for some seasonal strains because of unrecognised co-circulation of different influenza strains from those used as antigens in the HI assay. In this situation, anti-HA antibodies generated by infection have lower avidity for the HA of the assay virus. Conversely, original antigenic sin, where an infection results in an anamnestic response and the generation of antibodies directed towards an earlier infecting strain, might also explain 2-fold titer rises in response to infection [17]. In all these scenarios however, 2-fold increases would still represent infection by an influenza virus.

It is unlikely that 2-fold increases represent cross-reactivity of HI antibodies to strains of one subtype with strains of other subtypes. This is confirmed by our analysis that did not reject the hypothesis of an absence of cross-reactivity between subtypes.

It is therefore important for future work to determine if 2-fold titer increases represent infection, antigenic stimulation (attenuated infection), or artefact. If influenza infection rates are higher than currently recognised this might change our understanding of influenza transmission and of intra-host and inter-host immune mediated evolutionary pressures, and may have implications for the feasibility of control measures.

In the dataset examined here, 2-fold increases exceeded 4-fold increases for H1N1 in 2008 and H1N1, H3N2 and B in Spring 2009. There was no clear pattern with respect to subtype or strain. The seasonal H1N1 strain circulating in 2008 (A/Brisbane/59/2007) was antigenically distinct from those circulating previously (A/Solomon Islands/03/2006 and A/New Caledonia/20/1999-like), but this strain continued to circulate in Spring 2009. The seasonal H3N2 strain circulating in Spring 2009 (A/Perth/16/2009) was antigenically distinct from the 2007/8 strain (A/Brisbane/10/2007). H3N2 A/Perth/16/2009-like viruses have been difficult to propagate and we had difficulty propagating sufficient virus for the HI assays using A/Perth/16/2009-like viruses isolated from the cohort during the Spring 2009 season. We therefore used a virus isolated from a patient in Hanoi by the National Influenza Center, and propagated in eggs followed by MDCK cells (TX265M2E1) for undertaking HI testing of sera collected in Spring 2009. It is possible that the propagation in eggs this virus underwent might have resulted in some antigenic change, resulting in lower titers in the HI assay. National influenza surveillance data indicates that both influenza B lineages - Yamagata and Victoria- co-circulated during the study period, with the Yamagata lineage dominating in 2007 and 2008 and the Victoria lineage in 2009. For all HI assays, we used the same influenza B virus, which was isolated in 2008 and was characterized antigenically as Yamagata lineage-like, as with all influenza B viruses isolated from the cohort in 2008. While Yamagata viruses dominated the influenza B samples we collected in 2007 and 2008, the Victoria lineage was predominant in 2009. This may be a factor explaining the lower influenza B titer increases seen in that year. If heterogeneities in the proportion of 2-fold titer rises are largely attributable to a poor match between assay antigen and infecting virus, future seroprevalence and seroincidence surveys will need to use a greater diversity of antigens than typically used currently.

There are often strong age-related patterns in influenza serology. Ideally, we would therefore like to fit our statistical model independently for each age group. However, simulation studies indicate that the relatively small number of observations per age group would lead to relatively inaccurate estimates. We have therefore opted for an intermediate estimation strategy. Our statistical model fits a single distribution of true paired serology to all subjects; but since we infer true paired serology for each individual, we can reconstruct a posteriori the distribution of true paired serology for the different age groups. Even with such a conservative approach (*i.e.* it favours scenarios where the different age groups exhibit similar distributions), we were able to detect clear age-related patterns. In particular, it indicated that age may be another factor that influences the occurrence of a 2-fold rise. Larger sample sizes will be needed to investigate this possibility further.

The presence of relatively large proportions of individuals experiencing a 2-fold increase in antibody titers is not a peculiarity of the Vietnamese data examined here. Similar shifts were observed on data gathered by Cowling et al, with micro-neutralization assays for 2009 H1N1pdm09 influenza and on HI assays for seasonal influenza [23] (Figure S6).

It is well known that there may be substantial within- and between- laboratory variability in HI assays as well as in other serological assays such as virus neutralisation (VN) [24]. The level of intra-laboratory variations may depend on both the laboratory and the type of assay used [24]. Here, we have introduced an approach that allows controlling for within-laboratory variations. The only additional data needed compared with standard serological surveys is that replicate measurements are performed for a subset of subjects. These replicate measurements allow within-laboratory quantification of variation in assay performance. With this information, it is then possible to reconstruct the distribution of paired serology that is corrected for the estimated level of within-laboratory variations. Although our approach gives a better control on within-laboratory variation, it does not address the problem of between-laboratory variation. The use of standards in bioassays is critical for minimising the impact of the latter problem [24].

To conclude, while a 4-fold titer increase may be a highly specific diagnostic of infection by an influenza virus for individual cases, this criterion is less justifiable when the objective is to estimate community ARs. Our work shows that requiring a 4-fold titer increase may lead to ARs being substantially underestimated. More research is needed to determine what proportion of 2-fold rises are causally linked to exposure to influenza, and what proportion may be caused by other mechanisms. It will be important to determine whether the high proportion of 2-fold titer increases seen in the settings of Vietnam and Hong Kong [23] are also observed in other (e.g. temperate climate) settings.

Materials and Methods

Data

Samples were collected from a household-based cohort of 940 participants in 270 households in a single community in semi-rural northern Vietnam as previously described [5]. None of the participants had ever received influenza immunisation. Participants were under weekly active surveillance by village health workers for influenza-like-illness (ILI) and in the event of an ILI were asked to provide a nose and throat swab for detection of influenza RNA by reverse-transcription polymerase chain reaction. Participants were also asked to provide serial blood samples at times when national influenza surveillance data indicated that influenza circulation was minimal. The samples described here were collected over a period of three consecutive influenza seasons, from December 2007 through April 2010. The bleeding times were 1st–7th December 2007 (bleed 1), 9th–15th December 2008 (bleed 2), 2nd–4th June 2009 (bleed 3), and on the 3rd April 2010 (bleed 4). This provided three sets of paired samples either

side of an influenza transmission season: 548 paired samples for season 1 (2008), 501 paired samples for season 2 (Spring 2009), and 540 paired samples for season 3 (Autumn 2009). In season 1, the influenza A virus strains detected in the cohort through ILI surveillance were A/H1N1/Brisbane/59/2007-like and A/H3N2/Brisbane/10/2007-like; in season 2, they were A/H1N1/Brisbane/59/2007-like and A/H3N2/Perth/16/2009-like; and in season 3, it was A/H1N1/California/7/2009-like. There was co-circulation of influenza B Yamagata lineage and Victoria lineage in both season 1 and season 2, with a predominance of Yamagata lineage in season 1 and Victoria lineage in season 2.

Laboratory methods

Nasal and oropharangeal swabs were assessed by real-time reverse-transcriptase polymerase chain reaction (RT-PCR), according to WHO/USCDC protocols [25]. Influenza hemagglutination inhibition (HI) assays were performed according to standard protocols [WHO 2011 manual]. The seasonal influenza A viruses used were isolated from participants' swabs or from swabs taken from patients presenting in Ha Noi in the same season and propagated in embryonated hen's eggs or in MDCK cells. A reference antigen supplied by WHO (A/H1N1/California/7/2009-like) was used to assess season 3/pandemic sera. A single influenza B virus isolated from a participant during 2008 was used to assess serum for both the first and second seasons. The virus had a titer of 320 with B/Wisconsin/1/2010 (Yamagata) reference antisera and of <10 with B/Brisbane/60/2008 (Victoria) antisera. Each virus was first assessed for haemagglutination of erythrocytes from chickens, guinea pigs and turkeys then titrated with optimal erythrocytes. Serum was treated with receptor destroying enzyme (Denka Seiken, Japan) then heat inactivated and adsorbed against packed erythrocytes. Eight 2-fold dilutions of serum were made starting from 1:10 and incubated with 4 HA units/25 µl of virus. Appropriate erythrocytes were added and plates read when control cells had settled. Virus, serum and positive controls were included in each assay. Pre- and post-season sera were tested in pairs. Each serum was tested in a single dilution series. The HI titre was read as the reciprocal of the highest serum dilution causing complete inhibition of RBC agglutination, partial agglutination was not scored as inhibition of agglutination. If there was no inhibition of HI at the highest serum concentration (1:10 dilution) the titer was designated as 5. Only one sample had a titer >1280 and this was not adjusted. Replicate HI assay measurements were performed on a subset of samples from patients that seroconverted (i.e. 4-fold rise in titer) as well as some others that had titers ≥ 20 in both pre and post-season sera.

Statistical analysis

A less technical description of statistical methods is given for non-specialists in Box 1 and Figure 7.

Notation. Antibody titers (AT) are discrete measurements that can take a finite number of values. In our dataset, they can take 9 values: $a_0 = 10$, $a_1 = 20$, $a_2 = 40, \ldots$, $a_8 = 2560$, with the general form being $a_t = 10 \times 2^t$ for $t = 0, \ldots, K (K = 8)$. For simplicity, in the rest of the paper, antibody titers are labelled by integer t. For example, AT level $t = 0$ corresponds to antibody titers $a_0 = 10$.

We denote $\left\{O^b_{i,y,s}, O^p_{i,y,s}\right\}$ the "observed" AT levels measured at baseline (b) and post epidemic (p) in individual i, during season y ($= 2008$, Spring 2009, Autumn 2009) and for subtype s ($=$ H1N1, H3N2, B). In addition, for a subset of the blood samples, a replicate measurement of antibody titers was performed. We denote the replicate measurement for individual i, during season y and for subtype s (with $j = b$ for baseline and $j = p$ for post epidemic serology) by $R^j_{i,y,s}$ $R^j_{i,y,s} = NA$ if no replicate measurement was performed.

Measurement errors mean that observed and replicate AT levels may be different from the true (but unobserved) AT levels that we denote by $T^j_{i,y,s}$.

Hierarchical structure of the statistical model. We build a 3-level Bayesian hierarchical model to characterize measurement errors together with the underlying true distribution of baseline and post-epidemic serology. The model is defined by the following equation:

$$P\left(\left\{O^j_{i,y,s}, R^j_{i,y,s}, T^j_{i,y,s}\right\}_{i,y,s,j}, \theta\right) = \prod_{\{i,y,s\}} \left(p_{y,s}\left(\left\{T^b_{i,y,s}, T^p_{i,y,s}\right\}|\theta\right) M\left(O_{i,y,s}, R_{i,y,s}|T_{i,y,s}, \theta\right)\right) P(\theta) \quad (1)$$

where θ is the parameter vector of the model.

The first level $p_{y,s}\left(\left\{T^b_{i,y,s}, T^p_{i,y,s}\right\}|\theta\right)$ of the model characterises the underlying true distribution of baseline and post-epidemic serology for each season and subtype. The second level $M\left(O_{i,y,s}, R_{i,y,s}|T_{i,y,s}, \theta\right)$characterises measurement errors: given true AT levels $T_{i,y,s}$, it gives the probability to measure $O_{i,y,s}, R_{i,y,s}$ for the observed and replicate serology. The third level specifies our priors on model parameters. Each of those levels is described below, with more technical details given in the Supplementary Material.

Model for the underlying true serology. We consider the most general model for the joint distribution of true paired serology. For an individual i, during season y and for subtype s, each pair of serology measurements $\left\{T^b_{i,y,s}, T^p_{i,y,s}\right\}$is drawn from a Multinomial distribution

$$Multinomial\left(1, \left\{p_{y,s}\left(t_b, t_p\right)\right\}_{\{t_b = 0 \ldots K; t_p = 0 \ldots K\}}\right)$$

where $p_{y,s}\left(t_b, t_p\right)$is the probability that $\left\{T^b_{i,y,s} = t_b, T^p_{i,y,s} = t_p\right\}$. We estimate these probabilities from the data.

Model for measurement errors. The quantity of antibodies in the blood of a subject can be thought of as a continuous variable. However, observations (*i.e.* AT titers) are discrete. We build a model of measurement errors that accounts for the continuous nature of the underlying biological variable. As mentioned earlier, AT measurements can take K values $T = 0, \ldots, K$, corresponding to dilution levels of the HI assay. If the true (discrete) AT level is T, we assume that the continuous (unobserved) true quantity of antibodies in the blood, C_T, is uniformly distributed in the interval $[T; T+1[$. Conditional on the true quantity of antibodies C_T, we introduce a function $f(.)$ that indicates how far off from C_T the observation can be:

$$f(C_O|C_T) = \begin{cases} \frac{\gamma_T + 1}{2}(1 - |C_O - C_T|)^{\gamma_T} & \text{if} |C_O - C_T| \leq 1 \\ 0 & \text{if} |C_O - C_T| > 1 \end{cases}$$

Conditional on true AT level T and on the titration not going wrong, the probability that the observed AT level is O is given by:

$$g_1(O|T) = \int_{C_T = T}^{T+1} \int_{C_O = h_{\min}(O)}^{h_{\max}(O)} f(C_O|C_T) dC_T dC_O$$

where $h_{\min}(O) = O$ for $O > 0$ and $h_{\min}(0) = -\infty$; $h_{\max}(O) = O + 1$ for $O < K$ and $h_{\max}(K) = \infty$ (NB: boundaries 0 and K are treated as special cases since data are truncated at those levels).

Box 1. Less-technical description of the statistical method

In this box, we provide a less-technical description of the statistical method to give non-specialists an intuition of how it works. Readers should refer to the methods section for a technically rigorous description. From observed and replicate measurements of baseline and post epidemic ATs, our aim is to i) quantify measurement errors and ii) derive the *true* distribution of paired serology, that is, for example, to be able to estimate the *true* (*i.e.* after correction for measurement errors) proportion of subjects with ATs 10 at baseline and 40 post epidemic. For the sake of clarity, in this box, we restrict to the study of baseline ATs; but extending the approach to the joint analysis of baseline and post epidemic ATs is straightforward. We consider a toy dataset with 5 subjects with observed and replicate measurements for baseline ATs (Figure 7, panel A). Because of measurement errors, true baseline ATs are unknown (Figure 7, panel A). The statistical procedure is iterative. At iteration 1 (Figure 7, panel B), we start by initiating the model parameters and true ATs with arbitrary values (steps a and b). We can then derive the distribution of true ATs (step c) and calculate the probability ('likelihood') of the observed and replicate ATs given this initial set of parameters and characterisation of true ATs (step d). We are then running an iterative procedure called Markov chain Monte Carlo (MCMC) sampling. At each iteration (Figure 7, panel C) we are proposing new values for model parameters (step a) and for the true ATs of subjects (step b) in an attempt to improve the likelihood. After a certain number of iterations, parameters converge to the *posterior* distribution. This distribution gives likely values of parameters and also informs on uncertainty about those parameters. From the large sample of parameter values generated through 150,000 iterations of the MCMC procedure, we can calculate the posterior mean and 95% Credible Intervals (CI) of the parameters.

The probability of a 1-dilution (2-fold) error on one side (e.g. on the left) is $1/(2(\gamma_T+2))$.When the true AT level is not on the boundary 0 or K, the 2-sided probability of a 1-dilution error is $1/(\gamma_T+2)$.

The joint probability for the pair $\{\text{observed } O, \text{replicate } R\}$ is:

$$g_2(O,R|T)=\int_{C_T=T}^{T+1}\int_{C_O=h_{\min}(O)}^{h_{\max}(O)}\int_{C_R=h_{\min}(R)}^{h_{\max}(R)} f(C_O|C_T)f(C_R|C_T)dC_T dC_O dC_R$$

We also assume that there is a probability ε that the titration goes wrong and the resulting titre measurement is an integer uniformly drawn from 0 to K. Conditional on true AT levels T, the probability distribution for O is therefore:

$$g_O(O|T)=(1-\varepsilon)g_1(O|T)+\frac{\varepsilon}{K+1}$$

and the joint probability for the pair $\{\text{observed } O, \text{replicate } R\}$ is:

$$g_R(O,R|T)=(1-\varepsilon)^2 g_2(O,R|T)+\frac{\varepsilon}{K+1}(g_1(O|T)+g_1(R|T))+\left(\frac{\varepsilon}{K+1}\right)^2$$

Prior model. For each season y and subtype s, we assume that the set of probabilities $\{p_{y,s}(t_b,t_p)\}$ characterizing true paired serology has a Dirichlet prior distribution $Dirichlet(\{\alpha_{y,s}\})$, where hyperparameter $\alpha_{y,s}$ has a uniform hyperprior distribution on [0, 1000] (see Supplementary Material). The Dirichlet distribution is the conjugate prior of the multinomial distribution. Other parameters of the model have uniform priors.

Data augmentation and inference. True AT levels $\left\{T^b_{i,y,s},T^p_{i,y,s}\right\}_{i,y,s}$ are considered as augmented data and a Markov chain Monte Carlo (MCMC) sampling algorithm is used to explore the joint distribution of augmented data and parameters [26]. At each iteration of the MCMC, the following updates, which are detailed in the Supplementary Material, are implemented:

- Update 1: For each subject i, season y, subtype s, independence sampler for true AT levels $\left\{T^b_{i,y,s},T^p_{i,y,s}\right\}$;
- Update 2: For each season y and subtype s, Gibbs sampler for the probability distribution of paired serology $\{p_{y,s}(t_b,t_p)\}_{\{t_b=0...K;t_p=0...K\}}$;
- Update 3: For each season y and subtype s, Metropolis-Hastings update of hyperparameter $\alpha_{y,s}$;
- Update 4: Metropolis-Hastings update of parameters characterizing measurement errors.

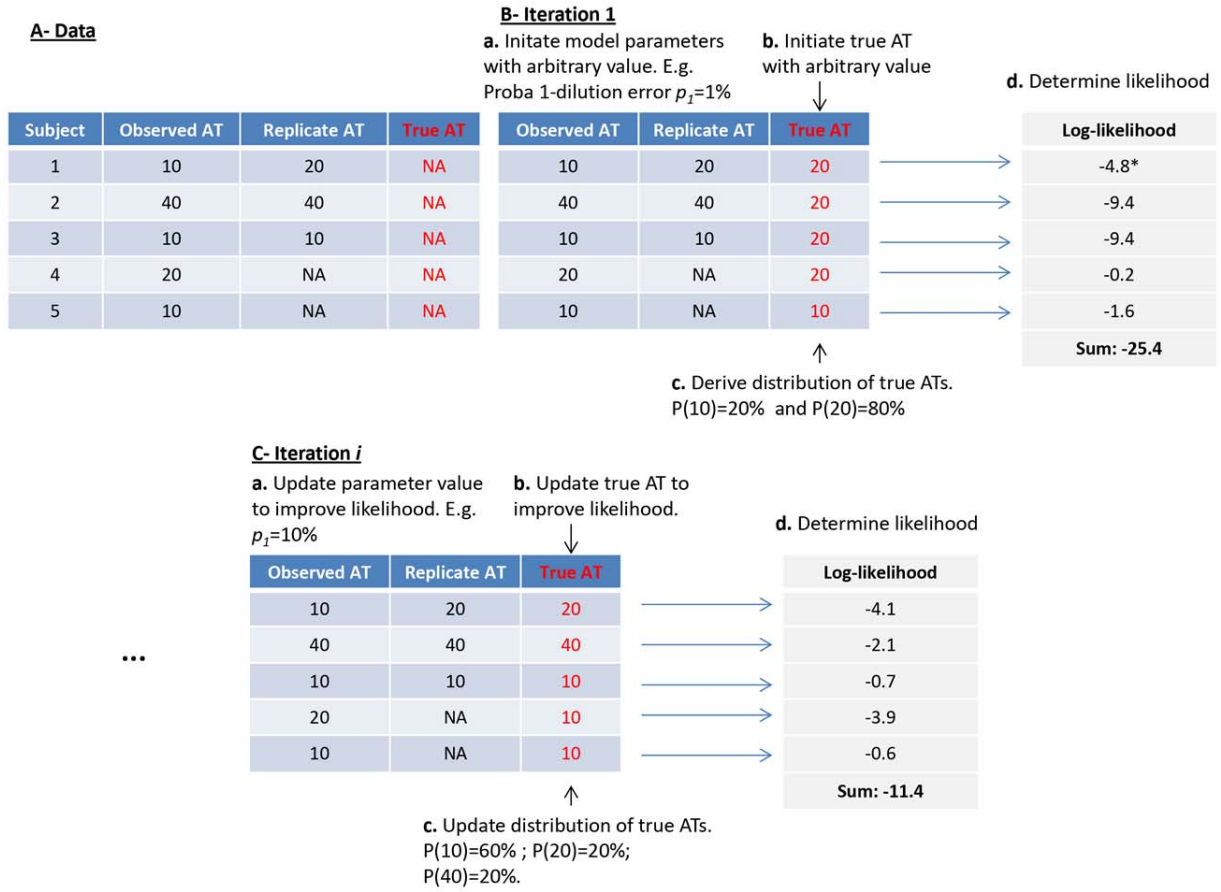
Information on measurement errors is contained in the data from the subset of individuals for whom a replicate measurement was performed. If update 4 (on measurement error parameters) was run on the full likelihood, the inference would suffer a "feedback" problem, with estimates of measurement errors being potentially largely driven by the larger (yet poorly informative) subset of individuals for whom no replicate measurements are available. We therefore use a standard strategy to circumvent this problem that consists in only using the contribution of individuals with replicate measurements in update 4 (see for example, function "cut" in WinBugs) [27–29]. Technical details are given in the Supplementary Material.

Selection of subjects for whom replicate measurements were performed. The subjects for whom replicate measurements were performed were not selected at random (Table S1). For example, those that had low antibody titers at baseline and post epidemic were never selected. To correct for this selection bias we model the selection process and make estimation of parameters characterizing measurement errors conditional on those individuals being selected. Technical details are given in the Supplementary Material.

Simulation study. In order to assess the performance of the method to quantify measurement errors and reconstruct the true distribution of paired serology, a simulation study is implemented. Eighty datasets with a structure similar to ours (i.e. same number of subtype/season, same number of observed paired serology per subtype/season) are simulated from the posterior mean of the parameters and the distribution of the true paired serology. The selection of subjects for whom replicate measurements are performed is simulated as in our model. We then applied our statistical model to each of the simulated datasets and assessed the bias on parameters quantifying measurement errors and on the true distribution of paired serology.

Ethics statement

The research was approved by the institutional review board of the National Institute of Hygiene and Epidemiology, Vietnam; the Oxford Tropical Research Ethics Committee, University of Oxford, UK; and the Ethics Committee of the London School



*: Calculation for subject 1. Likelihood=P(True=20) x P(Obs=10 | True=20) x P(Rep=20 | True=20) = 0.8 x 0.01 x 0.98

Figure 7. Less technical description of the statistical method. This figure illustrates the description of the method that is made in Box 1.
doi:10.1371/journal.ppat.1003061.g007

of Hygiene and Tropical Medicine, UK. All participants provided written informed consent.

Supporting Information

Figure S1 Fit of the model where measurement errors are independent of true antibody titers to data on replicate measurements. Observed (red point) and expected (mean: blue point/95% CI: blue bar) number of pairs {observed AT level, replicate AT level}. Pairs are sorted by panel according to the number of dilution difference between the observed and the replicate measurement.
(EPS)

Figure S2 Adequacy of model where measurement errors are independent of true antibody titers to the data. Distribution of "observed" paired serology as predicted by the model (color bars) and as observed in the data (black point) as a function of season (2008, Spring 2009, Autumn 2009) and subtype (H1N1, H3N2 and B). In each panel, individuals are sorted by baseline AT levels on the y-axis. For a given baseline, the grey bar indicates the expected proportion of individuals with post AT level equal to baseline AT level; the yellow bar indicates the proportion with a 2 fold rise (2f.r.); the red bar indicates the proportion with a 4 fold rise or more (4f.r.+); the green bar indicates the proportion with a decay. The black thin lines give the 95% CI. The legend gives the mean [95% CI]. **A**: H1N1, 2008. **B**: H3N2, 2008. **C**: B, 2008. **D**: H1N1, Spring 2009. **E**: H3N2, Spring 2009. **F**: B, Spring 2009. **G**: H1N1pdm09, Autumn 2009.
(EPS)

Figure S3 Testing the absence of cross-reactivity between subtypes. For each year and each subtype, individuals were partitioned between those with no increase in titers (coded 0), those with a 1-dilution increase (coded 1) and those with a 2 dilution or more increase (coded 2). The population was then partitioned in 27 groups according to outcome for triplet H1N1-H3N2-B. For example triplet 1-0-0 consists of individuals with a 1-dilution increase for H1N1 but no increase for H3N2 and B; 1-2-0 are individuals with a 1-dilution increase for H1N1, 2-dilution increase for H3 but no increase for B etc. Red points show the mean posterior distribution for triplet H1N1-H3N2-B, corrected for measurement errors. The boxplots in the figure show the distribution that would be obtained if there was no cross-reactivity between subtypes.
(EPS)

Figure S4 Age-specific distribution of paired serology, corrected for measurement errors as a function of season (2008, Spring 2009, Autumn 2009) and subtype (H1N1, H3N2 and B) (in Autumn 2009, subtyping was only conducted for H1N1pdm09). In each panel, individuals are sorted by baseline AT levels on the y-axis. For a given baseline, the grey bar indicates the expected proportion of individuals with

post AT level equal to baseline AT level; the yellow bar indicates the proportion with a 2 fold rise (2f.r.); the red bar indicates the proportion with a 4 fold rise or more (4f.r.+); the green bar indicates the proportion with a decay. The black thin lines give the 95% CI. The legend gives the mean [95% CI]. **A**: H1, 2008, <18 y.o. **B**: H3, 2008, <18 y.o. **C**: B, Spring 2009, <18 y.o. **D**: H1, 2008, <18 y.o. **E**: H3, Spring 2009, <18 y.o. **F**: B, Spring 2009, <18 y.o. **G**: H1pdm, Autumn 2009, <18 y.o. **H**: H1, 2008, 18–48 y.o. **I**: H3, 2008, 18–48 y.o. **J**: B, Spring 2009, 18–48 y.o. **K**: H1, 2008, 18–48 y.o. **L**: H3, Spring 2009, 18–48 y.o. **M**: B, Spring 2009, 18–48 y.o. **N**: H1pdm, Autumn 2009, 18–48 y.o. **O**: H1, 2008, >48 y.o. **P**: H3, 2008, >48 y.o. **Q**: B, Spring 2009, >48 y.o. **R**: H1, 2008, >48 y.o. **S**: H3, Spring 2009, >48 y.o. **T**: B, Spring 2009, >48 y.o. **U**: H1pdm, Autumn 2009, >48 y.o.
(EPS)

Figure S5 Model adequacy to age-specific data. Distribution of "observed" paired serology as predicted by the model (color bars) and as observed in the data (black point) as a function of season (2008, Spring 2009, Autumn 2009), subtype (H1N1, H3N2 and B) and age group (<18 y.o., 18–48 y.o., >48 y.o.). In each panel, individuals are sorted by baseline AT levels on the y-axis. For a given baseline, the grey bar indicates the expected proportion of individuals with post AT level equal to baseline AT level; the yellow bar indicates the proportion with a 2 fold rise (2f.r.); the red bar indicates the proportion with a 4 fold rise or more (4f.r.+); the green bar indicates the proportion with a decay. The black thin lines give the 95% CI. The legend gives the mean [95% CI]. **A**: H1, 2008, <18 y.o. **B**: H3, 2008, <18 y.o. **C**: B, Spring 2009, <18 y.o. **D**: H1, 2008, <18 y.o. **E**: H3, Spring 2009, <18 y.o. **F**: B, Spring 2009, <18 y.o. **G**: H1pdm, Autumn 2009, <18 y.o. **H**: H1, 2008, 18–48 y.o. **I**: H3, 2008, 18–48 y.o. **J**: B, Spring 2009, 18–48 y.o. **K**: H1, 2008, 18–48 y.o. **L**: H3, Spring 2009, 18–48 y.o. **M**: B, Spring 2009, 18–48 y.o. **N**: H1pdm, Autumn 2009, 18–48 y.o. **O**: H1, 2008, >48 y.o. **P**: H3, 2008, >48 y.o. **Q**: B, Spring 2009, >48 y.o. **R**: H1, 2008, >48 y.o. **S**: H3, Spring 2009, >48 y.o. **T**: B, Spring 2009, >48 y.o. **U**: H1pdm, Autumn 2009, >48 y.o.
(EPS)

Figure S6 Distribution of observed paired serology in [23]. A: HI assay for seasonal H1N1 influenza (2009). **B**: Micro-neutralization assay for pandemic H1N1 influenza (2009). **C**: HI assay for pandemic A(H1N1)pdm09 influenza (2009).
(EPS)

Table S1 Probability (numerator/denominator) that replicate measurements are performed during 2008 and Spring 2009 seasons, for subtype H1N1, as a function of observed serology at baseline and post epidemic. The colors indicate how we model the probability of selection. Yellow cells correspond to cells for which we assume that the probability of selection is null. The probabilities associated to the 4 other colors are estimated from the data (η_{S1}: orange ; η_{S2}: red; η_{S3}: light green; η_{S4}: green). See Supplementary Material for details.
(DOCX)

Table S2 Performance of the method to estimate parameters characterizing how subjects with duplicate measurements were selected. Those parameters are defined in Table S1 (see also section 1 of Supplementary Material). Eighty datasets are simulated with known parameters (see Methods). The table gives the simulation value of parameters and the mean (standard deviation) of estimates.
(DOCX)

Text S1 Technical details on the model, the estimation procedure, sensitivity analyses, and the test for the hypothesis of cross-reactivity between subtypes.
(DOCX)

Acknowledgments

SC thanks the RAPIDD program from Fogarty International Centre with the Science & Technology Directorate, Department of Homeland Security for the 2011 serology meeting. The authors thank the staff of the Ha Nam Provincial Preventive Medicine Centre, the Hamlet health workers and the National Institute for Hygiene and Epidemiology, Vietnam for their support in conducting the field work.

Author Contributions

Conceived and designed the experiments: SC PH AF NMF. Performed the experiments: SC PH AF LQM LTT PQT LNMH NTH. Analyzed the data: SC PH AF. Wrote the paper: SC PH AF LQM LTT PQT LNMH NTH NMF.

References

1. WHO Influenza (Seasonal). Available at www.who.int/mediacentre/factsheets/fs211/en/index.html. Accessed on 30 May 2012.
2. Van Kerkhove MD, Asikainen T, Becker NG, Bjorge S, Desenclos JC, et al. (2010) Studies needed to address public health challenges of the 2009 H1N1 influenza pandemic: insights from modeling. PLoS Medicine 7: e1000275.
3. Van Kerkhove MD, Ferguson NM (2012) Epidemic and intervention modelling–a scientific rationale for policy decisions? Lessons from the 2009 influenza pandemic. Bull World Health Organ 90: 306–310.
4. Carrat F, Vergu E, Ferguson NM, Lemaitre M, Cauchemez S, et al. (2008) Time lines of infection and disease in human influenza: A review of volunteer challenge studies. American Journal of Epidemiology 167: 775–785.
5. Horby P, Mai le Q, Fox A, Thai PQ, Thi Thu Yen N, et al. (2012) The epidemiology of interpandemic and pandemic influenza in Vietnam, 2007–2010: the ha nam household cohort study I. American Journal of Epidemiology 175: 1062–1074.
6. Edmondson WP, Jr., Rothenberg R, White PW, Gwaltney JM, Jr. (1971) A comparison of subcutaneous, nasal, and combined influenza vaccination. II. Protection against natural challenge. American Journal of Epidemiology 93: 480–486.
7. Evans AS (1975) Serologic studies of acute respiratory infections in military personnel. Yale J Biol Med 48: 201–209.
8. Foy HM, Cooney MK, McMahan R, Bor E, Grayston JT (1971) Single-dose monovalent A 2 -Hong Kong influenza vaccine. Efficacy 14 months after immunization. Journal of the American Medical Association 217: 1067–1071.
9. Fox JP, Cooney MK, Hall CE, Foy HM (1982) Influenzavirus infections in Seattle families, 1975–1979. II. Pattern of infection in invaded households and relation of age and prior antibody to occurrence of infection and related illness. American Journal of Epidemiology 116: 228–242.
10. Clements ML, Betts RF, Tierney EL, Murphy BR (1986) Serum and nasal wash antibodies associated with resistance to experimental challenge with influenza A wild-type virus. Journal of Clinical Microbiology 24: 157–160.
11. Potter CW, Oxford JS (1979) Determinants of immunity to influenza infection in man. Br Med Bull 35: 69–75.
12. Hobson D, Curry RL, Beare AS, Ward-Gardner A (1972) The role of serum haemagglutination-inhibiting antibody in protection against challenge infection with influenza A2 and B viruses. J Hyg (Lond) 70: 767–777.
13. WHO Global Influenza Surveillance Network (2011) Manual for the laboratory diagnosis and virological surveillance of influenza. WHO.
14. Zambon M (1997) Laboratory diagnosis of influenza. In: Nicholson KG, Webster RG, Hay AJ, editors. Influenza. Oxford: Blackwell Publishing. pp. 123–156.
15. Katz JM, Hancock K, Xu X (2011) Serologic assays for influenza surveillance, diagnosis and vaccine evaluation. Expert Rev Anti Infect Ther 9: 669–683.
16. Wood JM, Gaines-Das RE, Taylor J, Chakraverty P (1994) Comparison of influenza serological techniques by international collaborative study. Vaccine 12: 167–174.
17. Petrie JG, Ohmit SE, Johnson E, Cross RT, Monto AS (2011) Efficacy studies of influenza vaccines: effect of end points used and characteristics of vaccine failures. Journal of Infectious Diseases 203: 1309–1315.
18. Cate TR, Couch RB, Parker D, Baxter B (1983) Reactogenicity, immunogenicity, and antibody persistence in adults given inactivated influenza virus vaccines - 1978. Rev Infect Dis 5: 737–747.

19. Veguilla V, Hancock K, Schiffer J, Gargiullo P, Lu X, et al. (2011) Sensitivity and specificity of serologic assays for detection of human infection with 2009 pandemic H1N1 virus in U.S. populations. Journal of Clinical Microbiology 49: 2210–2215.
20. Chen MI, Barr IG, Koh GC, Lee VJ, Lee CP, et al. (2010) Serological response in RT-PCR confirmed H1N1-2009 influenza a by hemagglutination inhibition and virus neutralization assays: an observational study. PLoS ONE 5: e12474.
21. Couch RB, Kasel JA (1983) Immunity to influenza in man. Annu Rev Microbiol 37: 529–549.
22. Thomas SL, Wheeler JG, Hall AJ (2002) Contacts with varicella or with children and protection against herpes zoster in adults: a case-control study. Lancet 360: 678–682.
23. Cowling BJ, Chan KH, Fang VJ, Lau LL, So HC, et al. (2010) Comparative epidemiology of pandemic and seasonal influenza A in households. New England Journal of Medicine 362: 2175–2184.
24. Stephenson I, Das RG, Wood JM, Katz JM (2007) Comparison of neutralising antibody assays for detection of antibody to influenza A/H3N2 viruses: an international collaborative study. Vaccine 25: 4056–4063.
25. WHO/USCDC (2009) CDC protocol of realtime RTPCR for influenza A(H1N1). Available at http://www.who.int/csr/resources/publications/swineflu/CDCRealtimeRTPCR_SwineH1Assay-2009_20090430.pdf. Accessed on 30 May 2012.
26. Gilks WR, Richardson S, Spiegelhalter DJ (1996) Markov Chain Monte Carlo in Practice. London: Chapman and Hall.
27. Best N, Spiegelhalter DJ, Lunn D, Graham G, Neuenschwander B (2006) Cutting "feedback" in Bayesian full probability models. A workshop about the development and use of the BUGS programme. Hanko, Finland. Available at http://mathstat.helsinki.fi/openbugs/IceBUGS/Presentations/BestIceBUGS.pdf. Accessed on 22 June 2012.
28. Spiegelhalter DJ, Thomas A, Best N, Lunn D (2003) WinBUGS User Manual Version 1.4, 2003. Available at http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf. Accessed on 22 June 2012.
29. Erasto P, Hoti F, Auranen K (2012) Modeling transmission of multitype infectious agents: application to carriage of Streptococcus pneumoniae. Statistics in Medicine 31: 1450–1463.