

**Spatial and Temporal  
Analysis of Avian Influenza  
H5N1**

**GE, Erjia**

A Thesis Submitted in Partial Fulfilment  
of the Requirements for the Degree of  
Doctor of Philosophy  
**in**  
Geography and Resource Management

The Chinese University of Hong Kong  
July 2011

UMI Number: 3497743

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent on the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3497743

Copyright 2012 by ProQuest LLC.

All rights reserved. This edition of the work is protected against unauthorized copying under Title 17, United States Code.



ProQuest LLC,  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 - 1346

**Thesis/Assessment Committee**

Professor FUNG Tung(Chair)  
Professor LEUNG Yee (Thesis Supervisor)  
Professor LIN Hui (Committee)  
Professor Gerard Rushton (External Examiner)  
Professor LAI Poh Chin (External Examiner)

Abstract of thesis entitled:

Spatial and Temporal Analysis of Avian Influenza H5N1

Submitted by GE, Erjia

for the degree of Doctor of Philosophy

at The Chinese University of Hong Kong in July 2011

Avian influenza H5N1 is one kind of important bird flu. Unfortunately, this virus has swiftly evolved and become highly pathogenic to humans and poultry, resulting in 100% of death in infected poultry and over 60% of mortality among infected human population. Moreover, the virus tends to reassort with other influenza viruses, such as the current swine flu H1N1, to establish themselves in environments and further this epidemic all over the world. The World Health Organization (WHO) has in fact warned that highly pathogenic avian influenza H5N1 poses a graver risk of a global human pandemic than at any time since the Hong Kong outbreak (H3N2) in the 1960s.

One key to preventing such a calamity is to obtain a thorough understanding of the mechanisms of avian influenza transmission and its spatio-temporal patterns of dispersal. The issues at stake are outbreaks' spatial and temporal patterns, the interrelationship of these with the evolution of influenza viruses in such a way that geography is understood as a dimension of the disease's virology, and the human and avian behaviors and socio-ecological environments associated with H5N1 spread. This thesis sets out to study these problems in detail

and propose solutions.

First, we apply multifractal detrended fluctuation analysis to determine the temporal scaling behavior of outbreaks in Asia, Europe, Africa, and the whole of the world between December 2003 to March 2009. Long-range correlation and multifractality, two important properties characterizing the scaling behavior of complex dynamics, are first detected in the outbreak time series. In addition, this study identifies different temporal scaling behaviors of outbreaks of these continents and specific seasonal patterns in Asia. These findings confirm our perspective that avian-influenza outbreak behaviors are self-similar over time and are spatially heterogeneous.

Second, we conduct a spatial analysis for global trends and local clusters of H5N1 outbreaks at multiple geographical scales. Currently, the local  $K$  function used in a point pattern analysis searches outbreak clusters, assuming the disease is spatially homogeneous. The thesis proposes a much more efficient method to measure the degree of clusters accurately. The modified function works by weighting outbreaks through distances, counting the number of the weighted outbreaks for each lattice point no matter whether the disease emerges in a grid. This weighted local  $K$  function extends cluster analysis from a point pattern to lattice data. Spatial representation in these terms then seeks to explore local patterns of H5N1 over a continuous space.

Third, we study a set of socio-environmental factors, which are plausibly associated with the occurrence of H5N1. Spatial epidemiological models are built for predicting the disease at both continental and national levels, covering Indonesia, China, and the whole of East-Southeast Asia. We evaluate the statistical models using 1,000 bootstrap replicates, showing a consistently high rate of prediction, assessed

by statistics: AUC, Kappa Index, and pseudo  $R$  square.

Finally, avian influenza is an inter-disciplinary issue across virology, medical geography, and spatial epidemiology. How to quantify and integrate knowledge from different disciplines remains a challenge in fully understanding the disease. We propose a method to formally integrate genetic analysis that identifies the evolution of the H5N1 virus in space and time, epidemiological analysis that determines socio-environmental factors associated with H5N1 occurrence and statistical analysis that identifies outbreak clusters. Our integrated results show a significant advance in findings over reports in, for instance, Gilbert et al. (2008) and we believe our findings are more precise and informative in representing the occurrence and the space-time dynamics of H5N1 spread. Overall, unlike traditional influenza studies, our work sets up a solid foundation for the inter-disciplinary study of this and other spatial infectious diseases.

## 摘 要

H5N1 是其中一种严重的鸟类禽流感病毒。这种病毒已经迅速地演变成为高致病性的病毒，导致感染病毒的家禽的死亡率达到 100%。然而，这种病毒同样对人类产生高致病性，在感染病毒的病人中，有超过 60% 的人死亡。更糟糕的是这种病毒已经出现与其他流感病毒结合产生新的病毒类型，如近期的 H1N1 猪流感病毒。而这些新的病毒将在人类生存的环境中存活下来加剧全球性的流感传播。实际上，世界卫生组织已经发出警告：高致病性 H5N1 禽流感是对人类全球性流感疾病的重大威胁。这将比六十年代香港爆发 H3N2 流感以来的任何时刻更为严重。

掌握禽流感的传播机制及其时空分布模式为实现有效阻止疾病的爆发和传播提供重要手段。目前，我们所面临的问题包括：禽流感爆发的时间和空间分布，以其与流感病毒演化的关系，人类活动，家禽饲养，还有各种社会生态环境因素与 H5N1 病毒传播的关系。该论文对上述问题做了详尽的分析研究并提出相应的解决方案。

首先，我们通过重分形去势波动分析方法在亚洲，欧洲，非洲，以及全球不同尺度下对从 2003 年 12 月份到 2009 年 3 月之间爆发的禽流感做了时间序列方面的分析。在分析中，我们发现禽流感的爆发在时间尺度上是具有长相关性和重分形性质的。这些特性在禽流感的研究中是第一次被发现，反映了这种疾病复杂的时间尺度变化行为。此外，我们的研究还发现禽流感爆发的时间尺度变化在不同的空间上也是不一样的。尤其在亚洲，具有明显的季节性。这些发现证实了我们的观点：禽流感的爆发在时间尺度上是具有自相似性的，而在空间又不尽相同。

其次，我们基于多个不同的地理空间尺度对禽流感 H5N1 的爆发做了全局和局部模式的分析。局部  $K$  函数是一种常用于点模式分析的空间统计量。在这里，我们将用这类方法确定疾病集中爆发的空间位置。

然而，空间无差异性的假设条件，限制了局部  $K$  函数的精确性，也很难符合现实环境中的情况。因此，论文提出了一种更加有效精确地测量空间聚合度的方法。改进后的  $K$  函数通过对每个栅格位置周围的爆发事件求取距离加权和的方式来估计疾病在空间中的聚合分布。这种加权局部  $K$  函数是对原来基于点模式的分析进行拓展，使得对禽流感 H5N1 的空间聚合分析可以在连续的栅格数据中实现。

第三，为了掌握禽流感爆发与环境的关系，我们对一些有可能导致疾病的社会生态环境因素进行了分析。基于这些因素的分析，我们为印尼，中国，以及整个东南亚地区建立空间流行病的预测模型，并且通过一系列的统计分析，如 AUC，Kappa 指数，和方差  $R^2$  对该模型进行了评价，证实了模型良好的预测效果。

最后，我们意识到禽流感的研究是一个跨越病毒学，医学地理，以及流行病学的交叉学科。如何将这些不同学科的知识量化并结合在一起实现对该疾病全面认识是跨领域研究中的一大挑战。一方面，论文通过对 H5N1 病毒的基于分析，加入空间和时间因素，分析了病毒在时空中的演化。另一方面，论文提出了一种方面，将上述的空间流行病模型分析，疾病爆发的聚合分析，以及基因分析的结果进行量化并结合起来用于对 H5N1 病毒演化的时空建模。我们的工作比起一些已经发表的结果，如 Gilbert et al. (*PNAS*, 2008)，无论在对疾病爆发的预测和病毒时空传播的模型的精确度和信息量的表达上都具有显著的优势。总的来说，我们的工作并非一种传统的流感研究，而是为类似于禽流感和其他的流感疾病的交叉学科研究建立了一个良好的基础。

# Acknowledgement

My supervisor, Prof. Yee Leung, guided me through this research project. I sincerely appreciate his teaching and encouragement along the way. It was also fortunate for me to be in the Influenza Epidemic Research Group (with Prof. Mary Waye, Prof. Zuguo Yu, and Prof. Kahou Chu) led by my supervisor. It provides a rich learning environment throughout my study. My heartfelt “thanks” goes to Prof. Robert Haining, my host supervisor at Cambridge University during my visit (from July to December, 2010), for his inspiring thoughts and stimulating discussions.

During this research project, it was my pleasure to work with several people who also contributed to this thesis in various ways. In Chapter 3, Yee Leung provided insights on applying multifractal detrended fluctuation analysis (MF-DFA) to the time series of avian-influenza outbreaks; He also guided me to devise a statistical procedure that makes for the first time the detection of crossover time scales in MF-DFA rigorous, robust, and with statistical sense. Zuguo Yu provided a mathematical explanation for the long-range correlation and multifractality of the outbreaks. In Chapter 5, Yee Leung, again, guided me to construct a research framework that formally integrates multidisciplinary H5N1 studies into a unified framework within which constituent components can be individually analyzed and their interplays

can be effectively scrutinized. Robert Haining improved our assessment method by suggesting a rigorous measure for the spatial correspondence between our H5N1 prediction and the outbreak observation. Albert Lee explained the H5N1 evolution, including genetic sequence alignments, bio-evolutionary model selections, and practical implementation of phylogenetic analysis. Mary Waye and Kahou Chu gave useful perspective on DNA sequence analysis.

I would also like to thank the external examiners Prof. Gerard Rushton and Poh Chin Lai, and my internal committee members Prof. Tung Fung and Hui Lin for their comments and suggestions.

Finally, all of my work would not be possible without the support and encouragement from my family. I thank my father for his love and encouragement, although he passed away many years ago, and also my mother for her patience, encouragement, and always being with me. This work is dedicated to my parents.

# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgement</b>	<b>vi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Contribution . . . . .	5
1.3 Thesis Outline . . . . .	7
<b>2 Literature Review</b>	<b>11</b>
2.1 Importance of H5N1 Studies . . . . .	11
2.1.1 History of H5N1 subtype . . . . .	12
2.1.2 The Viruses . . . . .	15
2.1.3 Transmission . . . . .	17
2.2 State of Current H5N1 Investigations . . . . .	20
2.2.1 Studies of H5N1 Evolution . . . . .	21
2.2.2 Spatial Epidemiological Studies of H5N1 . . . . .	24
2.2.3 Geographic Studies of H5N1 . . . . .	25
2.3 Research Issues and Trends . . . . .	28
2.4 Summary . . . . .	31
<b>3 Research Framework</b>	<b>32</b>

3.1	Conceptual Framework . . . . .	32
3.2	Research Issues . . . . .	34
3.2.1	Spatial and Temporal Patterns . . . . .	34
3.2.2	H5N1 Evolution in Space and Time . . . . .	36
3.2.3	Factors Associated with H5N1 . . . . .	37
3.2.4	Integrated Prediction of the Epidemic . . . . .	38
3.3	Summary . . . . .	39
<b>4</b>	<b>Temporal Analysis of H5N1 Outbreaks</b>	<b>40</b>
4.1	Introduction . . . . .	41
4.2	Experimental Data . . . . .	42
4.3	Fractal Methods . . . . .	48
4.3.1	Fractal . . . . .	50
4.3.2	Correlation . . . . .	51
4.3.3	R/S Analysis . . . . .	52
4.3.4	Multifractal Analysis . . . . .	53
4.3.5	Multifractal Detrended Fluctuation Analysis . . . . .	55
4.4	Analysis Results and Interpretations . . . . .	58
4.4.1	Long-range correlation . . . . .	58
4.4.2	Multifractality . . . . .	75
4.5	Summary . . . . .	80
<b>5</b>	<b>Spatial Analysis of H5N1 Outbreaks</b>	<b>82</b>
5.1	Introduction . . . . .	83
5.2	Experimental Data . . . . .	84
5.3	Spatial Statistics Methods . . . . .	87
5.3.1	$K$ Function . . . . .	87
5.3.2	Local $K$ Function . . . . .	91
5.3.3	Modified Local $K$ Function . . . . .	92

5.4	Analysis Results and Interpretation . . . . .	94
5.4.1	Global Trends . . . . .	95
5.4.2	Local Clusters . . . . .	99
5.5	Summary . . . . .	109
<b>6</b>	<b>The Integrated Approach to Map H5N1 in Space and Time</b>	<b>111</b>
6.1	Introduction . . . . .	112
6.2	Data Description . . . . .	114
6.2.1	DNA sequences . . . . .	114
6.2.2	Outbreak Data . . . . .	117
6.2.3	Socio-environmental Data . . . . .	119
6.3	The Quantitative Analysis . . . . .	120
6.3.1	Quantifying the Phylogenetic Tree . . . . .	121
6.3.2	Risk Factor Analysis . . . . .	127
6.3.3	Knowledge Fusion and Dempster-Shafer Theory of Evidence . . . . .	133
6.3.4	Spatial Correspondence Analysis . . . . .	141
6.4	Results and Interpretations . . . . .	142
6.4.1	Thailand and Vietnam . . . . .	142
6.4.2	Indonesia, China, and East-Southeast Asia . . . . .	145
6.5	Summary . . . . .	150
<b>7</b>	<b>Conclusion</b>	<b>152</b>
7.1	Summary . . . . .	152
7.2	Directions for Further Research . . . . .	156
<b>A</b>	<b>Detection of Crossover Time Scales in MF-DFA</b>	<b>158</b>
A.1	Scaling-identification Regression and Detection of Crossover Points . . . . .	161

A.2 Determination of Crossover Time Scales by Statistical Inference . . . . .	163
A.3 Model Fitting and Detection of Crossover Points . . . . .	165
<b>Bibliography</b>	<b>169</b>

## List of Figures

2.1	Genealogical tree provided by Haeckel (1866) . . . . .	23
3.1	Conceptual framework for the study of avian influenza.	33
4.1	Time series of H5N1 outbreaks . . . . .	47
4.2	Cross correlation of H5N1 outbreaks . . . . .	49
4.3	Time series of Fractional Brownian motion and Fractal Gaussian noise and their MF-DFA results . . . . .	59
4.4	The MF-DFA results of H5N1 outbreaks . . . . .	62
4.5	Time series of Asian outbreaks at 90-day time scale . .	66
4.6	Time series of Asian outbreaks at 20-day time scale . .	71
4.7	The multifractality of H5N1 outbreaks . . . . .	79
5.1	The spatial distribution of H5N1 outbreaks . . . . .	86
5.2	Spatial point patterns . . . . .	88
5.3	Illustration of the traditional and modified local $K$ func- tions. . . . .	94
5.4	The results of $K$ and $L$ functions of humans and avian outbreaks . . . . .	97
5.5	Clusters of H5N1 outbreaks at a scale of (a) 30 km, (b) 120 km, (c) 260 km, and (d) 330 km. . . . .	102
5.6	Result of modified local $K$ function analysis of H5N1 outbreaks in East-Southeast Asia. . . . .	103

5.7	Observed H5N1 outbreaks and the results of the modified local $K$ function . . . . .	107
6.1	The spatial distribution of H5N1 outbreaks and Socio-Environmental factors . . . . .	115
6.2	The combination of HA and NA sequences . . . . .	116
6.3	The neighbor-joining (NJ) tree of the H5N1 virus in East-Southeast Asia . . . . .	118
6.4	The NJ trees of Thailand, Vietnam, Indonesia, and China	125
6.5	The Quantification of the phylogenetic tree of the H5N1 virus from Thailand and Vietnam . . . . .	128
6.6	ROC curves for the logistic regression model . . . . .	132
6.7	The results of geo-evolutionary and logistic regression analysis of the H5N1 virus from Indonesia, China, and East-Southeast Asia . . . . .	137
6.8	Integrated H5N1 patterns of Indonesia, China, and East-Southeast Asia. . . . .	148
A.1	Optional caption for list of figures . . . . .	160
A.2	Inference procedure for the determination of Scaling-identification Regression Model . . . . .	164

## List of Tables

4.1	Attributes of H5N1 outbreak records . . . . .	44
4.2	A sample of outbreak records by continents . . . . .	44
4.3	Pearson's Correlation . . . . .	48
6.1	Logistic regression model assessment for the H5N1 occurrences in East-Southeast Asia, Indonesia, and China, 1996-2009 and the two epidemic waves between 1996-2004 and 2005-2009. . . . .	133
6.2	Summary results of the logistic regression model for the avian influenza H5N1 epidemics in East-Southeast Asia, Indonesia, and China, 1996-2009. . . . .	134
6.3	Dempster's combination for the three sources of evidence on H5N1 in Thailand. . . . .	140
6.4	Test for spatial correspondence of H5N1 outbreaks and the empirical patterns . . . . .	144

# Chapter 1

## Introduction

This thesis sets out to study spatial and temporal processes in the incidence of avian influenza H5N1. It surveys different existing H5N1 studies in virology, medical geography, and spatial epidemiology before pointing to problems in existing types of analysis in these disciplines that might limit our being able to reconstruct the scaling behaviors of the disease in space and time. As an alternative to current approaches, the study then proposes a range of advanced techniques, including multifractal detrended fluctuation analysis, phylogenetic analysis, spatial epidemiological models, and theory of evidence, which it applies to modeling the spatial and temporal spread of H5N1. The study finally suggests further research extending the methods and topics it has introduced.

### 1.1 Motivation

The highly pathogenetic avian influenza (HPAI) A H5N1 has become a serious public health problem fatal to poultry and humans all over the world (Enserink, 2006; Adams and Sandrock, 2010). This form of influenza can cause fatalities in humans and birds alike. The World

Health Organization has cautioned that H5N1 has the potential to cause a global human pandemic (WHO, 2005), suggesting the urgent need to put in place effective control measures to prevent fresh outbreaks, or (failing that) the spread of the disease. The study of the spatial and temporal patterns of avian influenza is thus an important area of research in so far as it can help science understand the mechanisms that drive the establishment and transmission of the H5N1 virus (Wallace et al., 2007; Smallman-Raynor and Cliff, 2008; Si et al., 2009).

Traditional genetic analysis, which infers the phylogenetic relationships associated with the H5N1 virus from its genetic sequences, has already made a fair amount of progress on understanding the evolution of avian influenza viruses (Li et al., 2004; Duan et al., 2008). These genetic means allow the inference of possible sources of, and pathways taken by, the disease (Smith et al., 2006b). However, this form of analysis, which provides a micro-scale insight into the process of viral evolution, is insufficient to represent the macro-scale spread of avian influenza.

The spread of the *evolution* of influenza in large-scale, that is, geographic terms, has only recently become a subject of research. Geographers' main concern has been, up to now, not the virology of H5N1 but more simply its spatial distribution and mapping (Smallman-Raynor and Cliff, 2008; Carrel et al., 2010). Thus, although geography can properly be accounted an important dimension in the virus's evolution, as yet no rigorous and systematic analysis has been carried out on the time series of the epidemic. While virological attempts to place phylogenetic influenza variants in some sort of sequence have necessarily relied on assumptions of variants occurring at different times and in different places, 'tree' models of the disease, as noted above, have been

put together purely out of the analysis of variants' genetic data. Further, despite the suggestion being made that avian influenza outbreaks are seasonal and follow a random distribution (Chaichoune et al., 2009), very few studies have been offered formally analyzing the influenza's temporal behaviors. There is consequently a need to unpick the temporal processes of H5N1 in order to shed light on the spatial spread of the disease. As Chapter 3 discusses in more detail, an accurate measure of the temporal scaling behavior of avian influenza outbreaks will enable us to understand the physical mechanisms driving this complex system.

Identifying the spatial patterns of avian influenza is also essential if we want to prevent H5N1. Si et al. (2008) apply a range of spatial statistics to explore the virus's spatio-temporal pattern. The usefulness of this study, though, is restricted by its assumption of an equal probability of an occurrence of the disease over an entire and homogeneous space (and/or time) (Diggle, 1983; Cressie, 1991). The local  $K$  function, a measure of the clustering of events, also suffers from this assumption, meaning that it is unlikely to yield a satisfactorily accurate (or detailed) picture of local patterns of the disease as shown through a point process. There is a particular discrepancy between the reality of the disease and these basic modeling methods in that the risk posed by avian influenza extends continuously in space (and time). The implication from modeling or predication is that it is not sufficient to concentrate on points where outbreaks occurred and clustered intensively in the past. Modeling has to be able to take account of the risk of new outbreaks and to understand the likelihood of past incidences of H5N1 incubating fresh occurrences.

Given that avian influenza has spread across the globe (Enserink,

2006), it is particularly urgent and important to understand the factors associated with the disease. Gilbert et al.'s analysis focuses on the association between the spread of H5N1 and domestic poultry in Thailand (Gilbert et al., 2006a, 2007). The increasing availability of environmental data means that the study can quantify this research by statistical modeling (Gilbert et al., 2008). Gilbert et al.'s model attempts in this way to predict the probability of occurrence of the disease in both Thailand and Vietnam. However, the value of these statistical analysis is limited by persistent forms of uncertainty arising from incomplete data, limited domain knowledge, and the application of an insufficiently sophisticated methodology. Specifically, by not taking into account effects caused by geographical scale (the so-called "modifiable areal unit problem" (Openshaw, 1984)), studies of this kind have the potential to be misinterpreted at the stage where their results offer themselves up for modeling in terms of area-based epidemiology.

Finally, current research recognizes that the study of avian influenza is inter-disciplinary (Yee et al., 2009). It is acceptable that integrating multi-disciplinary studies from virology, medical geography, and spatial epidemiology may in theory be able to get around the problems of needing to factor geographic scale into predictive models of the spread of an evolving disease. However, as the study describes in Chapter 5, the current work can go no further than the early stage of analytically integrating data from different sources (Kilpatrick et al., 2006; Liang et al., 2010) and performing basic statistical analysis (Carrel et al., 2010). Even though knowledge can be derived from different disciplinary studies, how best to quantify and integrate this knowledge remains a challenge in the ongoing study of H5N1.

Based on the above issues, the thesis, unlike traditional influenza studies, sets out to investigate the mechanisms of the H5N1 virus's evolution and its spread in space and time. The primary objective of the thesis is to model the spatial and temporal spread of H5N1, aiming at the prediction and prevention of the disease. Specifically, this study consists of four sub-objectives: (1) to explore the temporal scaling behaviors of H5N1 outbreaks in Asia, Europe, Africa, and the world; (2) to identify the spatial patterns of the disease over a range of geographical scales, covering Thailand, Vietnam, Indonesia, China, and the whole of East and Southeast Asia; (3) to determine the socio-environmental factors that might have affected the spread and outbreaks of the disease; (4) to integrate phylogenetic analysis with the above studies for modeling the occurrence of the H5N1 virus in East-Southeast Asia.

## 1.2 Contribution

This thesis examines problems of geography, geographic scale and virus's continuous evolution in more detail, making the following specific contributions.

1. The thesis represents first full-dress study of the long-range correlation and multifractality of the time series of H5N1 outbreaks in Asia, Europe, Africa and the rest of the world. The study reveals a seasonal pattern to Asian outbreaks, further identifying well-formed and distinct temporal scaling behaviors in outbreaks on the three key continents. These results indicate that avian influenza outbreaks are heterogeneous in space and time (Leung et al., 2011).

2. The thesis lays out a statistical procedure for the modeling of complex fractal scaling behaviors capable of reliably identifying crossover time scales under multifractal detrended fluctuation analysis (MF-DFA). Further, it proposes a statistical procedure for automatically detecting crossover time scales in fractal analysis. This technique, it is suggested, can replace the traditional identification of crossover time scales by eyeballing, improving the efficiency, precision and robustness of this process (Ge and Leung, 2011).
3. The thesis unravels the spatial patterns of avian-influenza (H5N1) outbreaks in humans and birds across the world. This study explicitly indicates that the global trends of the outbreaks for these two host populations differ significantly over a wide range of spatial scales. This finding implies that mechanisms driving the disease are different in humans and birds.
4. The thesis investigates clusters of avian influenza (H5N1) outbreaks using a weighted local  $K$  function. This modified function takes into account the spatial association of outbreaks and shows a substantial improvement over the standard  $K$  function in detecting outbreak clusters. Moreover, the thesis's technique developing a modified  $K$  function further refines cluster analysis by shifting from a discrete point pattern to the representation of a continuous space set out on a lattice.
5. The thesis identifies socio-environmental factors associated with outbreaks in East and Southeast Asia, notably Indonesia and China, and quantified factors' relative importance in triggering the disease using logistic regression models. By formulating the

relationship, we can predict the likelihood of disease outbreaks for the three areas at various space scales.

6. The thesis proposes a novel method formally integrating H5N1 studies in virology, medical geography, and spatial epidemiology. This method, as underpinned by the thesis' geographic perspective, facilitates the framing of a new research paradigm for the inter-disciplinary investigation of this and other influenza epidemics (Ge et al., 2011).

### 1.3 Thesis Outline

Chapter 2 seeks to underline the importance of studying H5N1 by coming at the topic from three distinct directions: from that of the history of avian influenza, from a virology account of H5N1's characteristics, and from a consideration of its transmission. The chapter proceeds to offer a comprehensive review of previous work on H5N1 in the fields of virology, medical geography, and spatial epidemiology. It explains this thesis's basic concepts, points to trends in current researches, and ends by proposing the set of research problems that the current study goes on to tackle.

Chapter 3 aims to provide a conceptual research framework within which the spread of avian influenza can be examined from a unified view that takes into account the evolution of H5N1, the spatial and temporal spread of avian influenza, and the environmental and socio-economic settings conducive to the outbreaks of the disease. This chapter elaborates the interplays of these different analysis components for the construction of the overall patterns of avian-influenza spread and its interpretation.

Chapter 4 engages with these problems by studying the temporal scaling behaviors of avian influenza (H5N1) over multiple time scales. It applies multifractal detrended fluctuation analysis (MF-DFA) to detect the long-range correlation and multifractality of the time series of global and continental outbreaks over a period stretching from December 2003 to March 2009. Experimental results show that H5N1 outbreaks differ in long-range correlation and multifractal properties among the different spatial areas of Asia, Europe, Africa, and the world. More particularly, it is possible to detect a crossover timescale in Asian outbreaks. This study demonstrates how H5N1 outbreaks follow different time series depending on the part of the world where they occur. The implication is that H5N1 outbreaks behave differently under different ecosystems, different poultry farm practices, and different public health measures.

In applying MF-DFA, this study notes that the detection of crossover time scale(s) in previous work has been relatively subjective, since it has not been made on the back of rigorous statistical procedures, rather having been generally determined by “eye balling” or subjective observation. The crossover time scales identified in this way may, then, turn out to be problematic even spurious. As an advance on subjective eye-balling procedures, this work proposes a statistical method modeling complex fractal scaling behaviors, which is capable of reliably identifying unobservable crossover time scales: it can determine the number and locations of the genuine crossover time scales, providing confidence intervals for the crossover time scales detected. In order not to break the flow of our discussion of avian influenza, the statistical test for the crossover time scales is detailed in the Appendix. This work has been organized for publication (Ge and Leung, 2011).

Chapter 5 studies spatial patterns of avian influenza (H5N1) over a range of space scales. The chapter employs  $K$  and local  $K$  functions to identify global trends and local outbreak clusters across the world. A modified local  $K$  function, as defined by the thesis, tackles the problem of H5N1's spatial heterogeneity (Diggle, 1983; Cressie, 1991) by allowing for (the possibility of) outbreaks' spatial association. By weighting the number of outbreaks by the distances between them, such a modified function can dramatically increase the precision of cluster estimates. This method is extended from point to lattice data to model outbreak clusters arranged along a theoretically continuous area.

Chapter 6 studies factors associated with the occurrence of H5N1. It selects a set of socio-environment variables and works out the relationship between these using logistic regression models. This socio-environmental model is then used to predict a possible pattern of incidence for H5N1 in Indonesia, China, and across the whole of East and Southeast Asia. The study assesses this model's predictive power by three statistical measures: Cohen's kappa index, Nagelkerke/Cragg & Uhler's Pseudo  $R^2$ , and plots of receiver-operating characteristic (ROC).

This chapter also says a few words on the inter-disciplinary nature of H5N1 research. The study of avian influenza has become interdisciplinary across virology, medical geography, and spatial epidemiology and any study that relies on only one kind of knowledge may miss important connections. Chapter 5 proposes a novel approach to uncovering the space-time dynamics of the disease by integrating the findings, first, of phylogenetic analysis, which unravels H5N1 evolution in space and time, second, of an analysis based on a modified local  $K$

function, which identifies outbreak cluster in space, with lastly, with the methods of spatial epidemiology, which seek to determine socio-environmental factors associated with H5N1 occurrence. These three strands of theory are integrated in formal terms using Dempster-Shafer theory of evidence (Dempster, 1967; Shafer, 1976). “Fusing” knowledge from multiple disciplinary domains can mark a significant advance on previous methods in the following ways. First, it can provide more precise predictions of future outbreaks and outbreak clusters by more fully mining information from different sources. Second, the integrative method does not just yield a description of the spatial distribution of H5N1, but is more informative in depicting the space-time dynamics of the disease and so more realistically represents its spread.

Chapter 7 concludes the thesis and suggests future work that can build on the foundation of concerns and techniques it puts in place.

---

□ End of chapter.

## Chapter 2

# Literature Review

This chapter explains why the study of H5N1 is important, looking first at its history and then considering the specific characteristics of the H5N1 virus and its vector(s) of transmission. The chapter next reviews a range of studies of the H5N1 virus as they have been offered in the disciplines of virology, medical geography, and spatial epidemiology. This section is intended to help orient readers of this thesis by explaining the work's basic concepts and identifying its research problems.

### 2.1 Importance of H5N1 Studies

The highly pathogenic avian influenza (HPAI) designated H5N1 has been responsible for large outbreaks in poultry, leading to high rate of human mortality as the disease has leaped the species barrier. H5N1 thus poses a serious threat to the health of both humans and animals. Since its first detection in 1996, the H5N1 subtype has exhibited a great ability to spread across southeastern Asia, Europe and Africa. The World Health Organization (WHO) has warned that H5N1 has the potential to cause a global human pandemic that could well result

in a total of over 5 million deaths and huge treatment and other costs (WHO, 2004).

By March 2010, a total of 489 patients from 15 countries were known to have been infected by H5N1, of which 289 died (WHO, 2010a). These deaths, along with its many outbreaks, have led to HPAI H5N1's being recognized by researchers in virology, molecular biology, epidemiology, and medical geography as an important research issue of significant practical value. By way of illustration, this section will give a general description of H5N1 from perspective of its history, viral characteristics, and transmission.

### 2.1.1 History of H5N1 subtype

Early in 1878, highly pathogenic avian influenza was initially detected in chickens in Italy (Preiser, 2006). These viruses are usually able to circulate for hundreds of years. Such a degree of persistence makes them more likely to set up viral pools for genetic reassortments and mutations (Duan et al., 2008). This may facilitate the evolution of viruses of an increased degree of virulence or extended host range. Importantly, virus persistence and pool depth also increase the risk of transmitting the virus to humans.

In the mid-1990s, H5N1 via a host of adaption phrase saltatorially evolved into a highly pathogenic form (Xu et al., 1999). This rapid evolution became one of the causes of large outbreaks in Thailand, Vietnam, Indonesia, and southern China between 2002 and 2004 (Chen et al., 2006). Since then, the virus has become predominant in East and Southeast Asia (Chen et al., 2006). Repeated outbreaks of the H5N1 strain of disease pose a challenge for current disease control measures. Avian influenza A H5N1 outbreaks have thus far gone

through three epidemic waves (Yee et al., 2009).

- *Wave I: December 2003–March 2004*

Before 2003, a Gs/GD-like influenza lineage was predominant in poultry in Hong Kong and Guangdong, southern China (Xu et al., 1999). In late 2003, this precursor lineage was replaced by a novel H5N1 genotype Z, which resulted in the first epidemic wave across southeast Asia (Li et al., 2004). This genotype is widespread across Thailand, Vietnam, Japan, South Korea, Philippine, Laos, Cambodian, Indonesia, and China (Alexander, 2007b). Compared to other viruses, H5N1 proved especially virulent in poultry, leading to over 120 million birds dying in the affected countries within a short period (Ligon, 2005). Unlike previous outbreaks, this lethal virus had a much higher mortality, nearly 100% in infected poultry in Thailand and Vietnam (WHO, 2004). Bird losses in numerical terms came in at a much higher figure than for all other outbreaks cumulatively worldwide for previous decades (Ligon, 2005).

Furthermore, this genotype extended its range of host species to mammals, including tigers, leopards, swine, cats, and dogs (OIE, 2010). The virus caused severe respiratory disease in infected humans, who suffered a high mortality rate accordingly (Chotpitayasunondh et al., 2005). Subtype H5N1 also significantly deepened the influenza risk of influenza in humans and other animals. To prevent the disease from spreading further, strict control measures were implemented in the infected countries. Vaccination for H5N1 in commercial domestics was extensively implemented in China, Vietnam, and Indonesia (FAO, 2010). However, these efforts failed to stop the recurrence of the epidemic.

- *Wave II: June 2004–November 2004*

In contrast with other strains, subtype H5N1 tends to prove more tenacious, even though it has to survive inter-outbreak seasons on the basis of a few remaining sublineages (Chaichoune et al., 2009). In the second wave, the persistent virus, again, moved the disease to several Asian countries, including Malaysia (WHO, 2005). Outbreaks appeared to be contained in Hong Kong, Japan, and Korea (FAO, 2010), but it is widely believed that during this period the H5N1 virus was going through a silent genetic evolution with the potential to initiate a larger pandemic in future outbreak seasons (Li et al., 2004; Chen et al., 2006).

- *Wave III: December 2004–May 2006*

Compared to the previous waves, the third epidemic wave was the largest and most devastating in terms of the magnitude of its outbreaks in poultry, the spread of the disease and the increasing number of species falling susceptible to it. Human H5N1 infections in this wave were reported every month in Asia, Eastern Europe, and Africa (WHO, 2010b). Human-to-human H5N1 transmission is rare, with poultry-to-human infections being the dominant vector by which the disease pass over to human in southeastern Asia. As a result of this third wave, subtype H5N1 has become highly epidemic in human populations. The rate of human-case fatality stands at around 60% in Asia (WHO, 2006b). In contrast, Indonesia shows a worse rate of about 77% considering the repeated infections of humans with the influenza over the whole, three-wave period.

After a seemingly silent period, a new variant of H5N1, character-

ized by what has been called “Fujian-like” lineage, was found in place of the virus’s previous multiple sublineages (Smith et al., 2006a). This dominant virus was more panzootic in poultry and birds and prove more virulent (FAO, 2010). Over 6000 infected migratory birds died at Qinghai Lake, western China in April 2005 (Chen et al., 2005; Liu et al., 2005). This large outbreak caused an estimated 10% decrease in the global population of bar-headed geese, highlighting the virus’s potentially devastating effects on vulnerable wildlife (Olsen et al., 2006). Subsequently, the virus has appeared across western Siberia, spreading to Europe and several countries of Africa (Olsen et al., 2006). Large outbreaks in poultry and wild birds were reported in these regions. At the time of these outbreaks in birds, the H5N1 virus was also isolated in pigs, tigers, leopards, cats, and dogs (OIE, 2010).

The persistence of avian influenza seriously threatens the health of humans and animals. The disease’s spread has caused large economic losses in the poultry trade both through the loss of stock and the high costs of vaccination. H5N1 spreads so quickly, and had proved so difficult to eradicate, that the efficacy of current disease control measures in poultry must be placed in question. Therefore, understanding the spatial and temporal patterns of H5N1 is crucial to forming an effective strategy of containing the disease.

### 2.1.2 The Viruses

Virological studies have shown a high degree of similarity between the H5N1 strain of avian influenza and the 1918 “Spanish” influenza virus sequence, with no more than ten amino acid changes being detected between the two strains’ polymerase proteins (Preiser, 2006). Moreover, a number of the same changes have been found in recently circulating

highly pathogenic H5N1 viruses (Taubenberger et al., 2005). This similarity between virus subtypes implicates that H5N1 could potentially cause a major pandemic in humans in coming decades (Preiser, 2006).

Influenza viruses are enveloped with a segmented genome made up of eight single-stranded negative RNA segments. These viruses belong to a family of Orthomyxoviridae (Webster et al., 1992). Based on antigenic differences between their nucleo- and matrix proteins, influenza viruses are classified into types A, B or C. Avian influenza is caused by type A viruses. Two main antigenic determinants of influenza A and B viruses are haemagglutinin (HA) and neuraminidase (NA) transmembrane glycoproteins. HA attaches to cell surface sialic acid receptors in association with entry of the virus into host cells (Takeda et al., 2003). NA cleaves terminal sialic acid to release of the virus from infected cells; infection then spreads through the respiratory tract (Lamb and Krug, 2001). The interest of this thesis is in looking at both HA and NA in order to infer phylogenetic relationships among H5N1 viruses. The other segments of the virus genome are a polymerase basic (PB2 and PB1), polymerase acidic (PA), nucleoprotein (NP), matrix (M1 and M2), and nonstructural proteins (NS1 and NS2). Based on an analysis of glycoproteins, influenza viruses are clustered into sixteen H (H1-H16) and nine N (N1-N9) subtypes. Three subtypes, H1N1, H1N2, and H3N2 are the only known influenza A viruses currently circulating among humans (Lipatov et al., 2004). Few avian influenza subtypes can cross species barriers, but H5N1 has caused responsible of severe disease and a high rate of mortality in humans.

The replication of influenza viruses as a cause of antigenic change has, in effect, made the disease continuously circulate in human populations. Selective pressures in viruses, for instance, neutralizing antibod-

ies, suboptimal receptor binding, or chemical antivirals, are likely to raise replications in a host or population scale (Preiser, 2006). In this case of H5N1, mutants with corresponding selective advantages have gradually become the dominant variants in that host and population (Ferguson et al., 2003). Such changes, named “antigenic drift”, have been found mainly to occur in HA and NA, generating new influenza strains that have the potential to spread annually (Proenca-Modena et al., 2007). “Antigenic shift”, by contrast, names a sudden and profound change, caused by the acquisition of two or more influenza A viruses of different subtypes (so-called reassorts). However, both genetic mutation and reassortment are likely to raise novel subtypes, such, for instance, as strains of H5N1 associated with genetic reassortment between human and avian viruses (Guan et al., 2002; Taubenberger et al., 2005). Such influenza subtypes with their increased virulence and host’s limited levels of prior protective immunity, may conceivably cause epidemic or even pandemic in humans. More importantly, the establishment of multiple H5N1 sublineages has the potential to set up viral pools for subsequent genetic reassortment and mutations. This in turn increase the possibility of a pandemic flowing from future influenza waves (Chen et al., 2006).

### 2.1.3 Transmission

Understanding the transmission of H5N1 viruses is important if we want to learn which areas are seeding recurrent outbreaks and how these are epidemiologically connected. The mechanisms by which avian influenza viruses pass through one species to another and bring about infection are extremely complex, depending on the strain of virus, the species of host, and environmental factors. To allow readers to better

understand this study, this section presents some background in terms of natural hosts, H5N1 pathogeny, and the modes of viruses transmission.

Wild aquatic birds, including *Anseriformes* (ducks and geese) and *Charadriiformes* (gulls and shorebirds), are natural hosts, indeed carrying all subtypes of influenza A viruses (Webster et al., 1992; Krauss et al., 2004). These viruses are often stable in their natural hosts in a low pathogenic form. However, birds without apparent clinical symptoms may shed H5N1 viruses over a long time period (Hulse-Post et al., 2005). Viruses, which cause birds only a slight reduction in weight or a decline in egg production, are referred to as low pathogenic avian influenza (LPAI). However, LPAI H5N1 has the potential to become highly pathogenic avian influenza (HPAI) after transmission or adaptation to new hosts (Preiser, 2006). Domestic poultry, such as chickens, turkeys, guinea fowls, quail and pheasants, are vulnerable to infection (Perkins and Swayne, 2002). Once LPAI phenotypes are transmitted to domestic poultry, they will quickly mutate to a highly pathogenic form and cause severe respiratory diseases within 24 hours. Death tends to follow in 48 hours (Swayne and Suarez, 2000). So far, the HPAI H5N1 strain that originated in poultry in East and Southeast Asia has resulted in mortalities in more than 60 species of wild birds (Olsen et al., 2006). The actual number of fully susceptible species is likely to be greater than that suggested by laboratory experiments or investigation of captive birds (Alexander, 2007a).

The main mode of transmission among birds is through faecal-oral chains (Preiser, 2006). Transmission enables both direct infection from host to host and the indirect spread of the virus via contaminated water and fomites. Birds with large populations that co-mingle with other

species are at high risk for transmission of the disease. 'Wet' markets, where live birds are sold under crowded conditions, have been deemed as one of primary sources for the dissemination of avian influenza diseases (Mounts et al., 1999). Compared to poultry-to-poultry transmission, transmission from poultry to wild birds were not so apparent at the onset of the disease in 1997.

The virus's mode of transmission becomes more intricate after the re-emergence of HPAI H5N1 in poultry 2003 (Ligon, 2005). Since that time, poultry-to-human transmission has been found in relation to human's direct contact with infected poultry or with surfaces and objects heavily contaminated with their droppings (Yee et al., 2009). Even though the suspicion exists that the virus may be disseminated over long distances by migratory wild birds (Normile, 2006a; Brown and Stallknecht, 2008), wild birds' role in the geographic spreading of avian influenza diseases remains unclear (Normile, 2005a, 2006b; Olsen et al., 2006). So far cases of human-to-human transmission have been limited only among family members, at the same time, the highly lethal H5N1 viruses seem capable of naturally infecting other mammals including tigers, swine, cats, and dogs. These infections impose significant threats on public health, as an avian strain of influenza A may jump directly from aquatic birds to an intermediate animal host and then to human populations (Preiser, 2006). Furthermore, H5N1 subtypes pose the risk of raising novel influenza viruses that might exhibit a higher degree of virulence exacerbating the current influenza pandemic in the case of HPAI H5N1 viruses potentially reassorting with the current swine flu H1N1 (Dawood et al., 2009). In addition, the illegal trades of poultry and exotic birds provide another route for possible spread of H5N1 over large geographic scales (Normile, 2005a; Kilpatrick

et al., 2006).

Previous epidemiological studies provide a general description of relationships, for example, between avian influenza and water (Ito et al., 1995; Okazaki et al., 2000; Rogers et al., 2004), climate (Fang et al., 2008), and poultry and farming (Gilbert et al., 2006a, 2007, 2008). However, these studies need to be further developed for different spatial and temporal scales (Kapan et al., 2006). Scientists may apply different spatio-temporal scales to explore the spatial and temporal patterns of socio-ecological changes. The patterns identified will elaborate trends in how the relationships vary with the changes of scales. Scale-based and -variant analysis of this kind may provide substantive information towards the prediction of avian-influenza outbreaks, which can be used to prevent the disease at global, continental, and regional levels.

## 2.2 State of Current H5N1 Investigations

Avian influenza is the subject of research across a range of disciplines in virology and molecular biology, spatial epidemiology, and medical geography. In the last decades, intensive studies have been devoted to understand the evolution of H5N1 subtypes, relationships between disease presence and socio-ecological factors, and geographical patterns of outbreaks. Drawing on different disciplinary knowledge, data sources, and methodologies, these studies share the objective of identifying the sources and pathways of H5N1. In order to frame the current research topic more clearly, this thesis will give a comprehensive review of previous studies, with focus on H5N1 evolution, disease epidemiological modeling, and the spatial distribution of outbreaks.

### 2.2.1 Studies of H5N1 Evolution

Virologists around the world have made significant progress in understanding the evolution of H5N1 viruses (Guan et al., 2002; Li et al., 2004; Smith et al., 2006b; Suwannakarn et al., 2009). Traditional genetic studies primarily include antigenic analysis and phylogenetic analysis. Antigenic analysis using antisera raised from immunized ferrets provides a “gold standard” for evaluating antigenic change to influenza viruses (Smith et al., 2004; Ducatez et al., 2006; Chen et al., 2006). Analysis can be used to predict antigenic drift, a factor aiding in the selection of vaccine strains (Bush et al., 1999; Smith et al., 2004; Wu et al., 2008).

Phylogenetic analysis was first proposed in biology for the study of evolutionary relatedness among various groups of organisms such as species and populations (see Fig. 2.1) (Edwards and Cavalli-Sforza, 1964). In using RNA sequences to construct phylogenetic trees this method has been widely applied to infer the evolutionary relationships among influenza viruses (Guan et al., 2002; Li et al., 2004; Chen et al., 2006; Duan et al., 2008). Phylogenetic analysis inferentially builds up a gene tree out of sampled sequences using (one of) the computational methods of maximum parsimony (that is, inferring the simplest or most economical genetic connections), maximum likelihood (inferring the most plausible genetic connections), or distance (inferring the shortest genetic connections) (Felsenstein, 1981; Saitou and Nei, 1987; Felsenstein, 1996; Swofford, 2002). Analysis then graphically represents inferred evolutionary relationships among viruses based on similarities and differences in their genetic characteristics (Cavalli-Sforza and Edwards, 1967). For a rooted phylogenetic tree, each node with descendants represents an inferred common ancestor while the lengths

of edges can be interpreted as virus's evolution processes. Rooted tree may actually be understood as direct in the sense of showing a unique node corresponding to the most common ancestor of all the tree's leaves (*taxa*) in the tree (Benton and Ayala, 2003). The most common ancestor for specific phylogenetic trees is usually designated as an outgroup close enough to allow inference from the sequence but far enough to sustain the assertion of distinct identity (Nixon and Carpenter, 1993). A/Goose/Guangdong/1/96, which has been commonly viewed as a source of H5N1 viruses (Xu et al., 1999; Li et al., 2004; Preiser, 2006), is chosen to be an outgroup of the phylogenetic analyses considered in this study.

Virologists attempt to infer about the spread of avian influenza disease from observing the topology of phylogenetic trees (Smith et al., 2006b; Ducatez et al., 2007). However, phylogenetic analysis has important limitations both in terms of the nature of the data in analysis and the bio-evolutionary models it supposes (Penny et al., 1992). This mode of analysis may not accurately represent species' evolutionary history when data for antigen movements is noisy (Woese, 2002). Although phylogenetic analysis provides a micro-scale insight into the evolution of avian influenza viruses, inferring evolution of H5N1 solely from a virological perspective will not be enough to give a complete and direct picture of the spread of avian influenza in space and time. There is a need to call on epidemiological studies and geographical analysis for supplementary information helping us better to understand the dynamic of this epidemic.

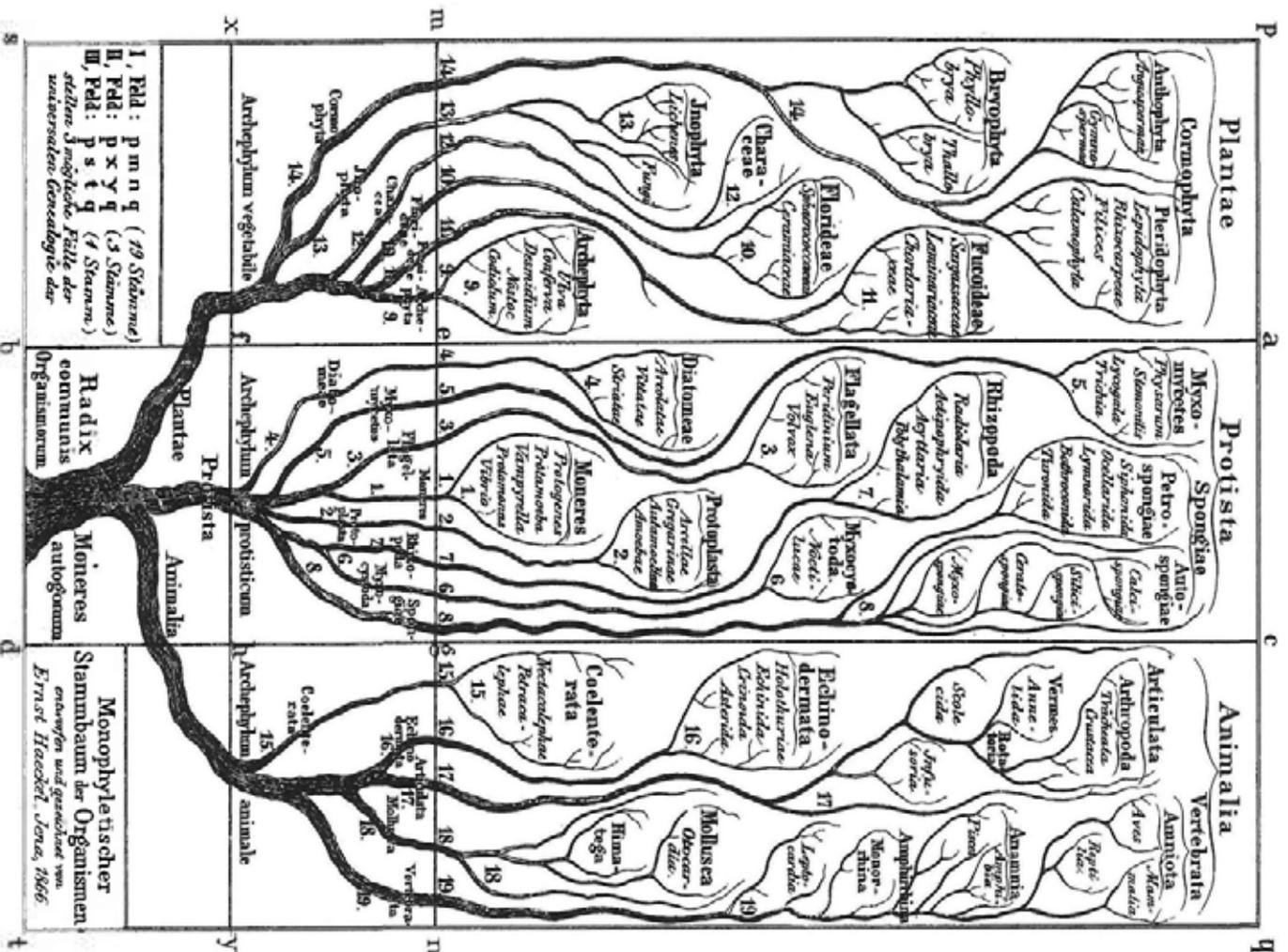


Figure 2.1: Genealogical tree provided by Haeckel (1866)

### 2.2.2 Spatial Epidemiological Studies of H5N1

In taking both space and time into account, spatial epidemiology is centrally concerned to identify factors relating to the causation of disease in population (Thomas, 1990). The purposes of spatial epidemiology are to describe spatial patterns of disease, to identify disease clusters to interpret or predict epidemic risk (Elliott et al., 2001; Pfeiffer et al., 2008). The discipline involves a range of quantitative analysis ranging from statistical description to sophisticated mathematical modeling (Lawson, 2001).

Avian influenza H5N1 is an important infectious disease spreading across large areas in Asia, Europe, and Africa (Enserink, 2006). This widespread disease has attracted much interest in spatial epidemiology, with intensive studies being devoted identifying risk factors in relation to the introduction and persistence of infection (Kapan et al., 2006; Pfeiffer et al., 2007; Uchida et al., 2008). Gilbert et al. (2006b) analyze the association between wild bird migration and patterns of HPAI H5N1 spread in western palearctic. Gilbert's analysis supports the inference that *Anatidae* species, including ducks, geese, and swans, have seeded influenza viruses along their seasonal migration routes. Li et al. (2006) assess the risk of avian influenza disease by conducting an analysis of the association between outbreaks and temperature in China. Spatial epidemiological studies have sorted through a large amount of socio-ecological data to model and predict the occurrence of HPAI H5N1 risk in southeastern Asia (Fang et al., 2008; Gilbert et al., 2008). Free-grazing duck, rice cropping, and population density have all been proved as central factors in virus transmission in Thailand (Gilbert et al., 2006a, 2007) and Vietnam (Pfeiffer et al., 2007).

These epidemiological studies have made a significant contribution

to understanding the relationships between avian influenza disease and ecological environments. However, spatial epidemiology is limited in the support that it can give inferences of the causal nature of those relationships it establishes (Green et al., 2008). In other words, while it can show correlations between outbreaks and model the spread of the disease, it lacks the virological or other foundations to assert these correlations as causes. Epidemiological analysis may further give a misleading impression of the causes of disease. This ambiguity is a consequence of a similar “modifiable areal unit problem” (Openshaw, 1977) in the sense that estimates of causation will vary with change in the scale at which analyses are implemented (Thomas, 1990). In addition, spatial epidemiological studies that explore the interrelationships between diseases and risk factors may suffer from uncertainty arising out of incomplete data or sometimes imprecise methodologies (Kapan et al., 2006). We will better understand the spatial pattern of H5N1, then, if we carry out geographical analysis aiming to identify disease dynamics over multiple spatial and temporal scales.

### 2.2.3 Geographic Studies of H5N1

Medical geography uses the concepts and technologies of the geography discipline to investigate health-related topics (Meade and Earickson, 2000). Getting underway in the 1950s, medical geography has been defined as a study of the relationships of factors connecting pathogens and so-called “geogens” (May, 1977). This study, as a subdiscipline and specialty of geography, brings to bear geographical perspectives and methodologies in various studies of health, disease, and care (May, 1954, 1977). Generally unlike spatial epidemiology, medical geography relates observed difference in disease incidence to differences in local

environments by mapping the incidence of diseases (Cliff and Haggett, 1986; Meade and Earickson, 2000).

One of the fundamental concepts in geography is *scale*, which may be broken down further into concepts of cartographic scale, analysis scale, and phenomenon scale. Cartographic scale refers to the depicted size of a feature on a map relative to its actual size in the world. Phenomenon scale indicates the size of physical processes that actually exist (Smelser et al., 2001). Unlike these two scales, analysis scale refers to the size of a unit within which quantitative analysis will be implemented (Openshaw, 1977; Meade and Earickson, 2000). Change of unit size often alters the level of analysis, generating different spatial patterns and affecting the interpretation of analysis results. This well-known phenomenon has been labeled as the modifiable area unit problem (MAUP or MTUP in the case of temporal scale) (Gehlke and Biehl, 1934; Openshaw, 1984).

The MAUP has, in fact, disrupted, or had considerable effects on the analysis of spatial and temporal patterns, particularly in association with epidemic diffusion (Meade and Earickson, 2000). Geography offers a straightforward and effective way to deal with such problems by using multiple scales or zooms (Lam and Quattrochi, 1992; Montello and Golledge, 1998).

Geographers have taken interest in the study of epidemics and pandemics for some time (Cliff and Haggett, 1986). Mapping the distribution of diseases (Cliff and Haggett, 1986; Gilbert et al., 2008) and modeling the processes of outbreaks (Cliff and Ord, 1980; Gibson, 1997a,b; Williams and Peterson, 2009) are typical examples of the application of geography to epidemiological themes. With more data available, geographical studies, however preliminary, become possible, in the cases

of avian influenza, geographic research is still thin on the ground and remains at an initial stage.

Geographic factors have not been explicitly considered as a dimension in the study of avian influenza until recently. In this connection, Wallace et al. (2007) apply the concept of spatial association in phylogenetic analysis to identify the epicenter of the disease and to infer the possible direction of spread of H5N1. Gilbert et al. (2008) also introduced spatial autocorrelation into their epidemic model for analyzing the factors associated with H5N1 outbreaks in southeast Asia, particularly Thailand and Vietnam. Most geographers' main concern is to establish H5N1's spatial and temporal patterns and to map the disease's spread (Si et al., 2008; Carrel et al., 2010). Smallman-Raynor and Cliff, for instance, propose a number of problems associated with the spread of avian influenza, coming up with a statement of the potential significance of geographical research on the disease (Smallman-Raynor and Cliff, 2008).

With respect to space and time, the spatial and temporal patterns of global spread of avian-influenza disease have been studied in relation to the movements of migrating birds (Olsen et al., 2006; Kilpatrick et al., 2006; Si et al., 2009). It is believed on the basis of such studies that migratory birds are responsible for the distribution of avian influenza (Normile, 2006a,b; USGS, 2010). However, such claims have not been rigorously confirmed in the statistical sense; and in fact, the role of migratory birds in the long-distance spread of influenza A H5N1 remains to be definitively determined (Normile, 2005a; Ducatez et al., 2006; Krauss et al., 2007; Wang et al., 2008a). Actually measuring the geographic scaling behaviors of influenza A H5N1 will be key to understanding the physical mechanisms of the disease's complex dynamics.

If medical geography is then to illuminate the process of avian influenza H5N1, it must integrate multiple knowledge bases, in a rigorous way, bringing together studies originating in different disciplines.

### 2.3 Research Issues and Trends

In virology, research on the avian influenza A viruses have centered around the detection of new subtypes and sublineages by phylogenetic analysis on nucleotides (Li et al., 2004; Chen et al., 2005). These studies attempt to trace how viruses jump from wild birds to poultry and possibly on to humans. Some of the motivation behind these studies is the fear that a new subtype, not previously circulating among humans, might be introduced into the human population and cause a pandemic. Virologists using the topology of phylogenetic trees attempt to infer the spread of avian influenza viruses over space and time (Chen et al., 2006; Smith et al., 2006b; Ducatez et al., 2007; Wang et al., 2008a). However, referring solely from the evolution of H5N1 in virological terms can only yield an incomplete picture of the spread of avian influenza, which has distinct spatial pattern (Kilpatrick et al., 2006). On this theme, Wallace et al. (2007) find that their results of phylogeographical analysis are statistically non-significant and conflict with previous studies such as those of Kilpatrick et al. (2006), although this may be due to the shortage of data.

Medical geography investigates patterns of mapping diseases and their association with attributes of the human and physical environments (Meade and Earickson, 2000). Mapping disease may raise valuable questions and hypotheses about its causation (May, 1954). This method, though, is unlikely by itself to arrive at correct understanding of a disease's etiology. Howe (1977), for example, attempts to identify

the origin of stomach cancer in Japan by mapping its rates of incidence both at international and national scales. However, his study cannot derive any single interpretation of the two maps. In this case, spatial variation fails to be consistent with the etiology. A similar situation currently characterizes H5N1 studies. Si et al. (2008) apply a set of spatial statistics to detect spatial and temporal clusters of H5N1 disease. The analysis appears to identify clusters by pointing out areas of a high density of outbreaks. However, it cannot fully interpret mechanisms of avian-influenza outbreaks and H5N1's mode of transmission. In addition, the MAUP (Openshaw, 1977) can lead to interpretation errors or misunderstanding of the spread of diseases in geographic studies. Even though sophisticated statistics features in current research work (Cliff and Ord, 1980; Cliff and Haggett, 1986; Getis, 1991; Getis and Ord, 1996; Getis and Aldstadt, 2004), studies cannot rely on the information essential for making detailed forecasts of disease's outbreak and spread.

The focus of spatial epidemiology, unlike that of medical geography, is measures of the prevalence of disease (Thomas, 1990). One of important factors affecting statistical analyses in epidemiology is uncertainty, which may arise from the limitations in data, methodologies, or even domain knowledge. Not being able to access relevant data in terms of numbers of chickens and ducks, Gilbert et al. (2008) failed to give a high rate of HPAI H5N1 prediction in Vietnam and Indonesia. On the other hand, spatial epidemiology is also affected by the MAUP (Openshaw, 1977) when statistical analyses or models are implemented on the basis of lattices or grid data. Without rigorous analysis procedure, Fang et al. (2008) convert point and line-type information for H5N1 outbreaks, wild bird migration, and environmen-

tal variables (including water bodies, wetlands, transportation routes, main cities, precipitation and elevation) into a raster-type layer for modeling H5N1 risk. This study, though, suffer from its lack of a rigorously defined analytical procedure, Fang's logistic regression model, then represents the distribution of the disease misleadingly, showing the highest risk in northern China and a weaker risk in southeastern coastal and near-coastal areas. This risk map obviously conflicts with real-world observations (FAO, 2010; WHO, 2010b) and with previous studies (Smith et al., 2006a; Wallace et al., 2007; Smallman-Raynor and Cliff, 2008).

Existing studies of H5N1 have generated insights into the evolution and space-time spread of avian influenza. The study of H5N1 is a multi-disciplinary investigation across virology, medical geography, and spatial epidemiology. Integrating different types of domain knowledge will be key to framing a direct, comprehensive understanding of this spatial infectious disease.

Although current research tries to integrate multi-disciplinary studies of avian influenza, it tends to stop at the initial stages of analytically integrating data or implementing only basic statistical analysis between phylogenetic distance and geographic distance. The research group at the University of North Carolina (UNC) and the University of Iowa have begun to undertake work on the geography of avian influenza with a view to depicting the spatial and temporal patterns of influenza viral genotypes and thereby enhancing our understanding of ecosystem diversity in avian-influenza evolutions (Carrel et al., 2010). Kilpatrick et al. (2006) integrate data for phylogenetic relationships for virus isolates, migratory bird movements, and trade in poultry and wild birds to determine the pathways of H5N1 transmission into coun-

tries and to predict the disease's spread in future. The combination of data on human population density, natural host movements, and viral evolution also forms a basis for the identification of hot spots (Kapan et al., 2006). However, even though all these studies are informative in different ways, the questions remain of how current H5N1 studies can best quantify and integrate these different forms of knowledge.

## 2.4 Summary

In providing a comprehensive literature review, this chapter first illustrated the importance and necessity of this study by offering a brief perspective on avian influenza history, viruses' genetic characteristics and viruses' transmission. It took stock of different studies of H5N1 as offered in the disciplines of virology, medical geography, and spatial epidemiology. These investigations all suffer from certain problems, even if many point in the same direction and suggest research trends. The final part of the chapter began to expound the research issues to be discussed in this thesis. In the next chapter, we will present a conceptual framework of this research and then detail each component and their interplays within the proposed framework.

---

□ End of chapter.

## Chapter 3

# Research Framework

This chapter presents the conceptual framework of the thesis in general and elaborates research problems to be investigated under the framework in particular. It aims to give a complete outline of this study and to identify the interplays among the constituent components.

### 3.1 Conceptual Framework

Avian influenza is a spatial epidemic problem, which calls for the analysis of its patterns of spread at multiple spatial and temporal scales; the identification of hot spots and clusters of outbreaks; the relationship between H5N1 evolution and the outbreaks in humans and birds; and the socio-environmental factors affecting the spread and outbreak of the disease. The issues to be investigated are: (1) How do the patterns of spatial and temporal spread of avian influenza vary with the change of spatial and temporal scales; (2) How does the H5N1 virus evolve and how is it related to the outbreak of avian influenza in space and time; (3) What are the socio-economic and ecological-environmental factors affecting the outbreak and transmission of the disease; (4) Can the future occurrence of avian influenza be predicted?

The section constructs an integrated conceptual research framework for these studies. Within the framework, the spread of avian influenza can be examined under a unified view that takes into account the evolution of H5N1, the spatial and temporal spread of avian influenza, and the environmental and socio-economic factors conducive to the outbreak of the disease. Figure 3.1 shows the conceptual framework of this study.

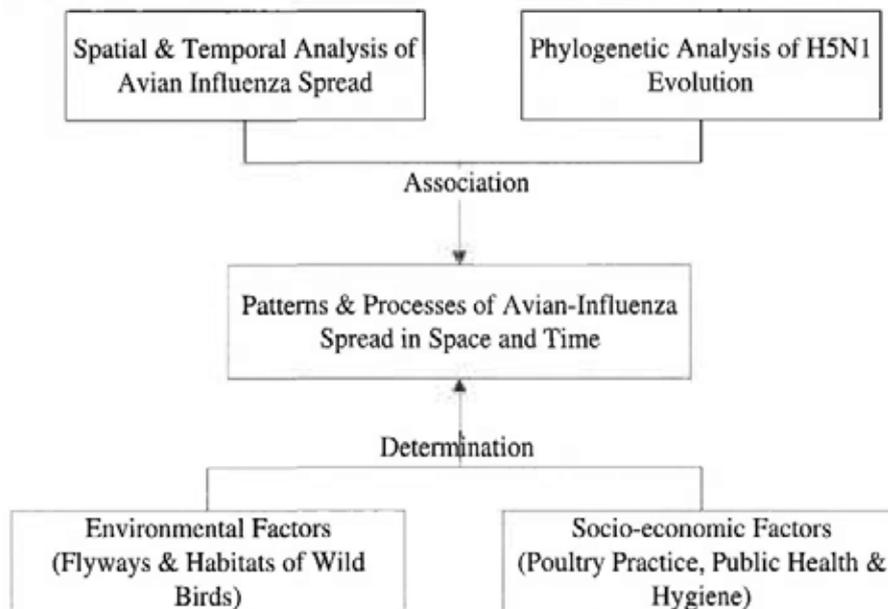


Figure 3.1: Conceptual framework for the study of avian influenza.

The framework pays particular attention to the interactions among these processes. The spatial and temporal spreads of avian influenza is first scrutinized to explore the patterns and processes of the disease over multiple spatial and temporal scales. Phylogenetic analysis is applied to explore the evolution of the H5N1 virus and its transmission among humans and birds so that a link to the spread of the disease in space and time can be established. Environmental factors and human socio-

economic behaviors, such as wild birds migration and poultry practice, are then investigated with respect to their effects on the occurrence and spread of the disease. Through these interactions, we can come up with a more comprehensive spatial and temporal analysis of H5N1.

## 3.2 Research Issues

Based on the proposed framework, this section sets to give an elaboration on the issues, including the spatial and temporal patterns of H5N1 outbreaks, the evolution of the H5N1 virus and its transmission in space and time, factors associated with the occurrence of H5N1, and the prediction of the disease. To understand the process of avian influenza, the thesis seeks to answer these questions through a systematic and rigorous analysis.

### 3.2.1 Spatial and Temporal Patterns

Geographical studies are essential of H5N1 outbreaks providing clues to the aetiology of a disease (Elliott et al., 1996). The purpose of spatial and temporal analysis is to explore the behavior and tendency of H5N1 outbreaks over multiple spatial and temporal scales. Spatial distribution and temporal scaling behaviors of avian influenza obtained from such study can be used to analyze the patterns and processes of the disease (see Figure 3.1).

First, the thesis sets to examine the process of avian influenza over time. Outbreaks of avian influenza appear to be complex, irregular, and self-similar at different temporal scales, evolving intermittently with interwoven periods of tranquility and spiky bursts of devastating occurrences. Understanding the scaling behavior of H5N1 outbreaks is essential to formulating the physical mechanisms of this complex

dynamics. The temporal analysis for outbreak time series seeks to answer: (1) Are avian-influenza outbreaks temporally correlated? (2) Do the outbreaks have any specific periodicity? and (3) Are the temporal behaviors different in different parts of the world?

Long-range correlation and multifractality are two important properties which characterize the scaling behaviors of time series. Multifractal detrended fluctuation analysis (MF-DFA) (Kantelhardt et al., 2002) provides an approach to detect the long-range correlation and multifractal property of non-linear time series. This method is used to analyze the time series of global and continental outbreaks of H5N1 from December 2003 to March 2009. The general Hurst exponents derived from the analysis can be employed to explicate the degree of long-range correlation in the outbreak of avian influenza over time. Specific crossover time scales separating regimes that have different long-range correlations can be uncovered to indicate changes in the behavior of the outbreak time series. Seasonal patterns and specific periodicity are thus determined. In addition, to examine if there are heterogeneities among the time series in different parts of the world, the outbreak series of Asia, Europe, Africa, and the world are examined and compared in a quantitative manner. Detailed analysis and interpretation are made in Chapter 4.

Second, this thesis seeks to identify the spatial distribution of avian influenza, particularly the global trends and local clustering of H5N1 outbreaks. It collects and collates outbreak data from the World Organisation for Animal Health (OIE) and the World Health Organization (WHO). The  $K$  and local  $K$  functions (Ripley, 1976; Getis, 1984) are applied to these data to determine the trend of avian-influenza outbreaks of the world showing a pattern of clustering over a range

of spatial scales. Through this method, hot spots of avian-influenza outbreaks can be identified at different scales.

Allowing for the spatial association of outbreaks, a modified local  $K$  function is proposed in this study to the outbreaks of H5N1 in Thailand, Vietnam, Indonesia, China, and the whole of East-Southeast Asia. Weighting the number of outbreaks by the distance between the outbreaks can significantly increase the precision of cluster estimation. Moreover, it extends the analysis from point to lattice data to model outbreak clusters under a theoretically continuous area. This study provides a detailed description of the distribution of H5N1 clusters in space (as detailed in Chapter 5).

Finally, these analyses explore the spatial and temporal patterns of H5N1 outbreaks and sheds more lights on the scaling behaviors of avian influenza from a macro view in space and time. Findings of this study will be integrated in Chapter 6, with the analysis of H5N1 evolution to model the processes of avian-influenza spread.

### 3.2.2 H5N1 Evolution in Space and Time

To have a complete understanding of avian influenza, phylogenetic analysis provides an approach to study the processes of H5N1 evolution. The phylogenetic tree obtained from this analysis can help us infer sources and pathways that the H5N1 virus evolves. How to extract this kind of knowledge from viral evolution remains a challenge to tracing the spread of the disease.

In Chapter 6, we first use DNA sequences and space-time data to create a phylogenetic tree to estimate the H5N1 virus' capability of spreading. Unlike traditional studies in influenza disease, it is the first attempt to quantify the evolution process of a virus in space and

time and provide a mapping of H5N1 viruses directly derived from the phylogenetic tree. This study is significant in shedding light on the spread of avian influenza through the microscope of the evolution of H5N1 in space and time.

### 3.2.3 Factors Associated with H5N1

Within the research framework (Figure 3.1), factors associated with avian-influenza outbreaks and spreads are to be determined for the prediction of avian influenza. To explain the relationship between the environment and the disease, factors, such as altitude, the shortest path to wild birds flyways, inland water, coast lines, are collected and correlated for the analysis. On the other hand, socio-economic factors, such as population density, poultry density, the shortest path, to railways and roads, are also employed to explain the occurrence of the disease.

To provide a rigorous analysis procedure, logistic regression analysis is used to model the relationship between socio-environmental factors and the occurrence of avian influenza H5N1 over different geographical scales ranging from Indonesia, China, and East-Southeast Asia, respectively (see the discussion in Chapter 6). In addition, a series of statistics, including the receiver-operating characteristic ROC, Cohen's Kappa index, and Nagelkerke / Cragg & Uhler's pseudo- $R^2$ , are calculated to assess the predictive power of the model. The findings obtained from this analysis give a quantitative description of the relationship between avian influenza and the environments where the H5N1 virus evolves, as well as human social behaviors that might have triggered the wide spread of the disease all over the world. The analysis result can be used as a piece of valuable evidence for the prediction of avian influenza. It also provides governments or public health organizations

substantial information for containing the disease.

#### 3.2.4 Integrated Prediction of the Epidemic

The above analyses provide evidence from three different angles for the processes of avian influenza and its viral evolution in space and time. Relying on knowledge or analysis results from only one perspective would only give us a partial picture of the spread of H5N1 in space and time. These pieces of knowledge, however, can complement one another in the construction of the overall situation of the disease. Therefore, effective integration of these domain specific knowledge is essential to understanding the process of avian-influenza outbreaks and spreads, as well as to predicting the occurrence of the disease.

In this part of the framework, Dempster-Shafer theory of inference is employed to integrate the findings of phylogenetic analysis, which unravels H5N1 evolution in space and time, with modified local  $K$  function analysis, which identifies outbreak clusters in space, and also with spatial epidemiology, which determines socio-ecological factors associated with the occurrence of H5N1. In Chapter 6, we apply this approach to study and map the risk of H5N1 across multiple geographical scales, including the whole of East-Southeast Asia and individual countries: Thailand, Vietnam, Indonesia, and China, respectively. As indicated in Chapter 6, bivariate spatial association analysis shows that our integrated study outperforms existing studies in giving a closer correspondence to the actual observation of H5N1 outbreaks in these places. This finding holds over a range of spatial scales. By this means, we may confirm that our results are more precise and informative in analyzing the space-time dynamic of H5N1. This study lays a solid foundation for the inter-disciplinary approach to the

study of influenza epidemics in general and H5N1 in particular.

### 3.3 Summary

This chapter provides the conceptual framework of the thesis by depicting the overall approach of the present study and the interrelationships of the constituent components. The central idea of the framework is to first identify the importance in understanding the spatial and temporal aspects of avian influenza, the phylogenetic analysis of H5N1, and the spatial epidemiological modeling and prediction of the disease. It then stipulates the way these individual studies can be integrated to render a comprehensive analysis of the patterns and processes of the spread of avian influenza in space and time. The methodologies and the analysis results of the above analysis are detailed in the chapters to follow.

---

□ End of chapter.

## Chapter 4

# Temporal Analysis of H5N1 Outbreaks

*Why is geometry often described as 'cold' and 'dry'? One reason lies in its inability to describe the shape of a cloud, a mountain, a coastline, or a tree. Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line ... Nature exhibits not simply a higher degree but an altogether different level of complexity.*

— Benoit Mandelbrot, 1977

This chapter studies the temporal behaviors of avian influenza A (H5N1) over multiple time scales by analyzing the global and continental outbreak time series from December 2003 to March 2009. Experimental results show that H5N1 outbreaks are long-range correlated and multifractal. The temporal patterns are heterogenous over space.

## 4.1 Introduction

In terms of time, avian-influenza outbreaks have been suggested to be seasonal with random distribution (Chaichoune et al., 2009). However, there are very few studies on the formal analysis of its temporal behaviors. To shed some more light on the spatial spread of the disease, this chapter seeks to unravel the temporal processes of H5N1 outbreaks. In particular, one needs to study its complex dynamics with reference to long-range correlation and self-similarity over time. In other words, the scaling behavior is a key to model the temporal process and to formulate the physical mechanisms of the complex dynamics. Studying the temporal behavior of avian influenza H5N1 outbreaks over multiple time scales will enable us to detect long-range correlation and underlying multifractal properties characterizing the complex dynamics (Feder, 1988). In addition, the temporal patterns of the H5N1 outbreak series can be compared to show whether avian influenza outbreaks are heterogeneous in different parts of the world.

The purpose of this chapter is to determine the scaling behavior of H5N1 outbreaks in wild birds and poultry over time. Specifically, answers are sought in quantitative terms for the following issues: (1) Are previous H5N1 outbreaks responsible for the current spells of infection? (i.e., Is it long-range correlated?); (2) Do H5N1 outbreaks exhibit multifractality properties; and (3) Are the temporal behaviors different among the continents?

To facilitate our discussion, we first give a description of the data in section 4.2. Relevant methodologies related to time series analysis, such as fractal, rescale range analysis, multifractal analysis and multifractal detrended fluctuation analysis are then briefly discussed in section 4.3. In section 4.4, the analysis results are examined and interpreted. We

then conclude the chapter with a summary of the analysis.

## 4.2 Experimental Data

To facilitate our analysis, an integrated avian-influenza database has been built by pulling together information provided by the World Organisation for Animal Health (OIE) ([www.oie.int](http://www.oie.int)), which are responsible for the global surveillance of the outbreaks of avian influenza A H5N1 in animals and humans. Specifically, the OIE collects and collates confirmed information on animal outbreaks from the sub-reference laboratories across the world. On the other hand, WHO collects data on human infections from its member countries distributed in different parts of the world. Since 2003, the organization has provided statistical data on H5N1 infections confirmed in humans. In this study, each record of H5N1 confirmed infections provided by the OIE is called, as officially addressed by OIE, an “outbreak” of H5N1. The Genbank is a sequence database of the National Institutes of Health (NIH) that collects all publicly available nucleotide sequences. This official database offers genetic sequences of H5N1 viruses in different parts of the world, making it possible to study the viral evolution via molecular analysis (Carrel et al., 2010).

This chapter is mainly concerned with the temporal behaviors of avian-influenza outbreaks in poultry and birds. The outbreak data are extracted from the official reports provided by the OIE. Although surveillance programs vary among countries, data provided by OIE are considered generally reliable and official records of the overall situation across the world. The OIE database has been commonly used in the study of avian influenza in medical geography (see for example Smallman-Raynor and Cliff (2008); Si et al. (2009)).

More than 5000 outbreak data have been involved in the present study, covering the period from 10<sup>th</sup> December 2003 to 21<sup>st</sup> March 2009. All these outbreak records were reported by the OIE. Most records contain attributes such as location, latitude, longitude, start time, end time, report time, affected species, and number of deaths (as explained in Table 4.1). Based on these reported data, Asia appears to be an epicenter with more than 57% outbreaks in general, compared to 25% in Europe and 18% in Africa. Nevertheless, the situation is quite different from the outbreaks in wild birds and poultry. For wild birds, less than 1% outbreaks happened in Africa, while beyond 40% were in Europe and 59% were in Asia. But for poultry, intensive outbreaks emerged in Asia and Africa including more than 80% in total. This preliminary examination suggests a spatial heterogeneity of the temporal outbreaks, which may imply distinct behaviors of the disease in different parts of the world over time.

To construct the time series for analysis, the OIE data are grouped into continents with respect to time (in year, month, and date). Table 4.2 shows a small part of the time series that indicates the frequencies of H5N1 outbreaks in each continent at each point in time. The analysis at the continental scale is to have time series long enough for a reliable temporal investigation of daily outbreaks of H5N1 across the world. However, our analysis is also applicable to finer spatial scale in the future if the outbreaks in places at such scale are long enough with respect to time.

This chapter aims at unraveling the temporal behaviors of avian-influenza outbreaks of Asia, Europe, Africa, and the world. Four time series of the outbreaks have been constructed, as depicted in Figure 4.1. Advanced statistical method, the MF-DFA, is applied to examine

Table 4.1: Attributes of H5N1 outbreak records

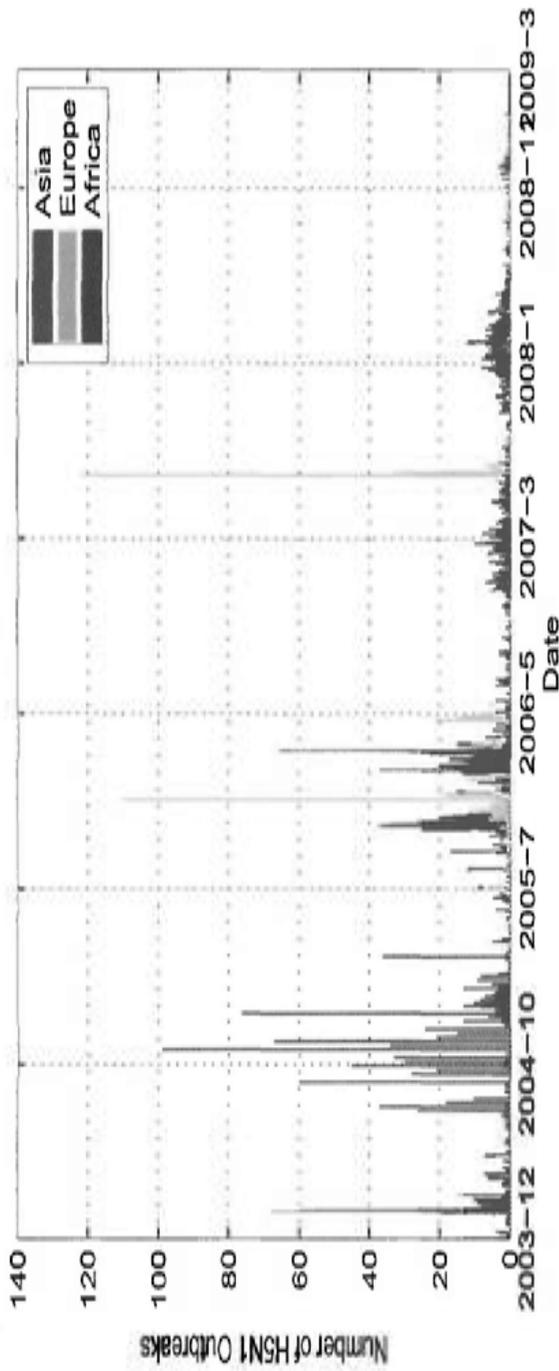
Field	Definition	Examples
FID	Shape ID by location	221
Shape	Type of shape	Point
Case No	Unique case number by time	4206
Continent	Part of the continent where the outbreak occurred	East Asia
Country	Country where the outbreak occurred	China
Province	Province of country (if any) where the outbreak occur	Guangdong
District	Local village where the outbreak occurred	Panyu
Location	Local village where the outbreak occurred	Sixian
Unit	Unit of outbreak size (e.g. village, farm)	village
Latitude	Latitude of the outbreak occurred	23.05
Longitude	Longitude of the outbreak occurred	113.42
Report Date	Date of report by WHO	9/15/2007
Date Start	Starting date of outbreak	9/5/2007
End Date	End date of outbreak	N/A
Virus	Virus of avian influenza	H5N1
Species	Species of the infected animal	ducks
Deaths	Number of Deaths in the outbreak	9830
Source	WHO OIE Reference	OIE Ref:6213

Table 4.2: A sample of outbreak records by continents

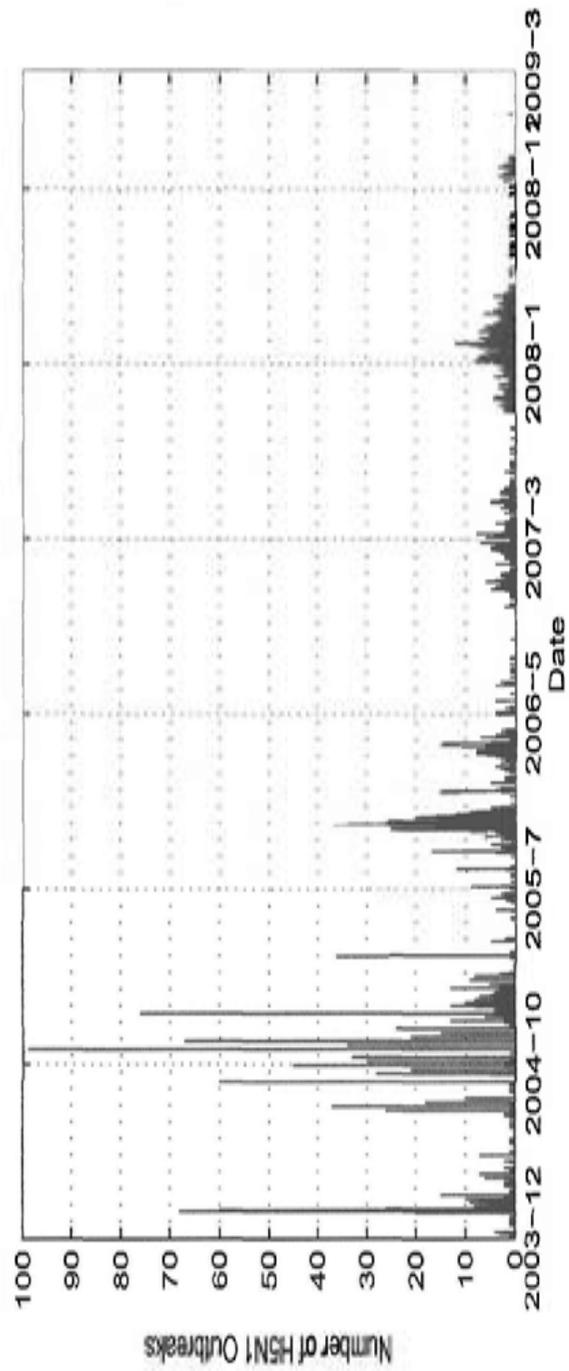
Year	Month	Date	World	Asia	Africa	Europe
2007	2	12	8	2	5	1
2007	2	13	6	2	3	1
2007	2	14	9	2	4	3
2007	2	15	3	0	1	2
2007	2	16	2	1	1	0
2007	2	17	4	2	2	0

the behavior of avian influenza over a wide range of time scales from 5 to 90 days. This renders a quantitative exploration of the behavior of the epidemic. The time series of Asia, similar to the world, contains the daily outbreaks from 10<sup>th</sup> December 2003 to 21<sup>st</sup> 2009. It should be noted that the time series of Europe and Africa are shorter, which contain 1082 and 901 records starting from the first detection on 18<sup>th</sup> July 2005 and 10<sup>th</sup> 2006, respectively. From Figure 4.1(b), five distinct peaks of avian-influenza outbreaks can be observed, with the first three occurring around the winter of 2003, 2004, and 2005, mainly in Asia, and the last two occurring towards the early spring in Asia, Europe, and Africa. The diffusion process of H5N1 is thus reflected by the distinct phases of outbreaks, originating seemingly from eastern Asia and spreading across Eurasia and finally into Africa.

To initially explore the spread of avian influenza A H5N1 across the world, we first examine the correlation and cross-correlation of the outbreaks emerging in different continents. The Pearson's correlation analysis is employed to examine whether the outbreaks of Asia, Europe, and Africa are temporally associated. From the results in Table 4.3, outbreaks in Europe are slightly correlated with those of Asia and Africa, but no significant correlations are found between the latter two continents. This result indicates that outbreaks in Europe have distinct associations (overlappings) in time with these of Asia and Africa. This implies that an indirect spread of the disease from Asia to Africa goes through Europe, an intermediate hub. The correlation analysis of the outbreak time series is insufficient to determine the degree of correspondence of the three continents. Hence the cross-correlation analysis (Haggett and A.E.Frey, 1977) has been applied to the time series of the outbreaks. Figure 4.2 indicates that the outbreaks in Asia are one and



(a) The world



(b) Asia

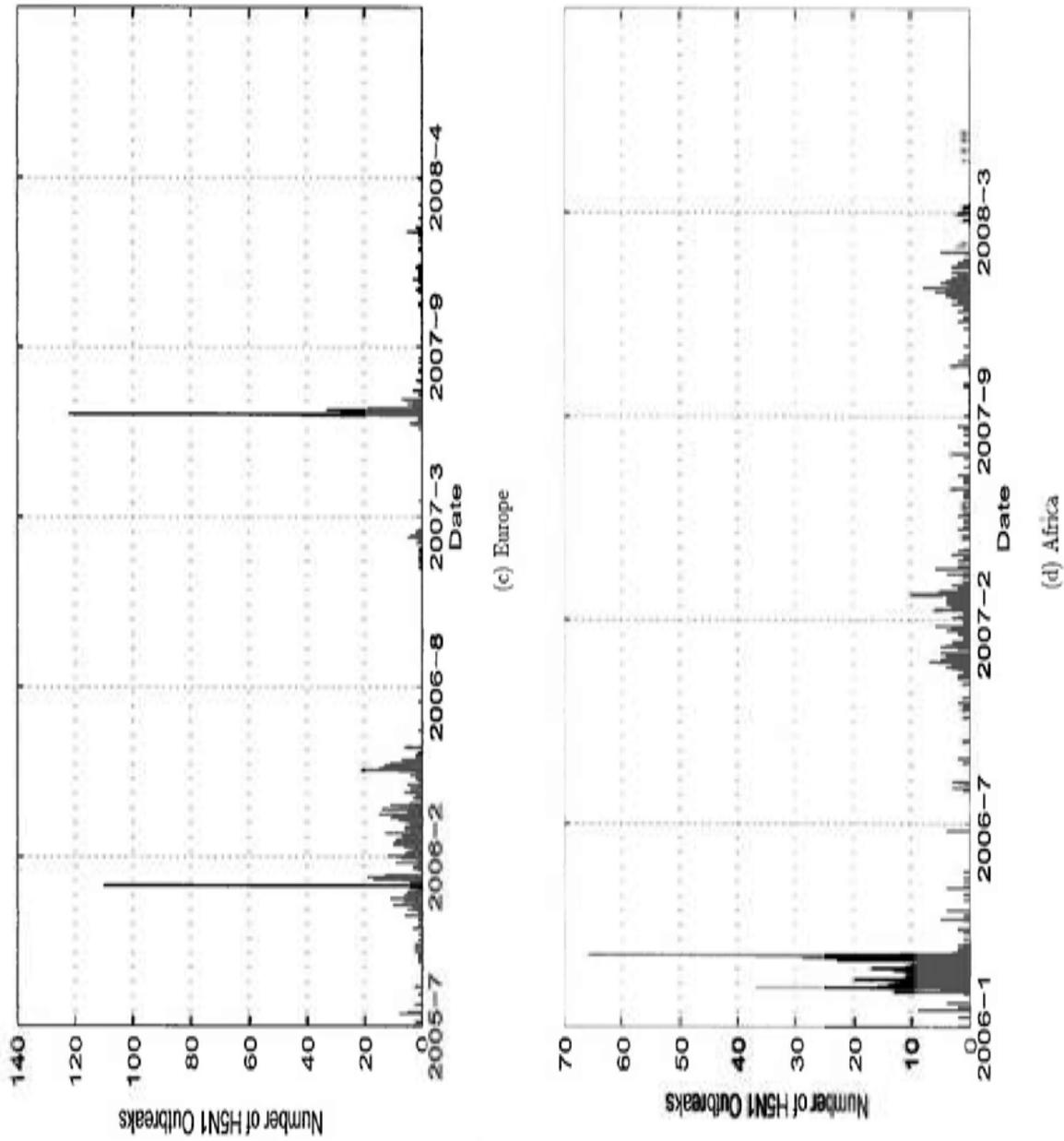


Figure 4.1: Time series of H5N1 outbreaks of (a) the world; (b) Asia, 10<sup>th</sup> December 2003 – 21<sup>st</sup> March 2009; (c) Europe, starting from 18<sup>th</sup> July 2005; and (d) Africa, starting from 10<sup>th</sup> January 2006.

Table 4.3: Pearson's Correlation

		Africa	Europe	Asia
Africa	Pearson Correlation	1	.151(**)	.022
	Sig.(2-tailed)		.000	.362
	N	1652	1652	1652
Europe	Pearson Correlation	.151(**)	1	.134(**)
	Sig.(2-tailed)	.000		.000
	N	1652	1652	1652
Asia	Pearson Correlation	.022	.134(**)	1
	Sig.(2-tailed)	.362	.000	
	N	1652	1652	1652

\*\* Correlation is significant at the 0.01 level (2-tailed).

a half years prior to those of Europe and Africa. This seems to imply that avian influenza tends to peak first in Asia and subsequently spread into the surrounding continents. Although the phenomenon has been observed and reported by Enserink (2006) and Alexander (2007a), this chapter provides a statistical description of such associations.

Correlation and cross-correlation analysis only give us a preliminary examination of avian influenza outbreaks across the continents. They, however, are insufficient for our understanding of the situation that is marked by the complex temporal dynamics across the continents. To shed some more light on the spatial processes, rigorous analysis of the temporal processes and their comparisons are essential. This is the focus of this chapter and is detailed in the discussion to follow.

### 4.3 Fractal Methods

Avian influenza outbreaks appear to be a complex process that evolves intermittently with interwoven periods of tranquility and spiky bursts

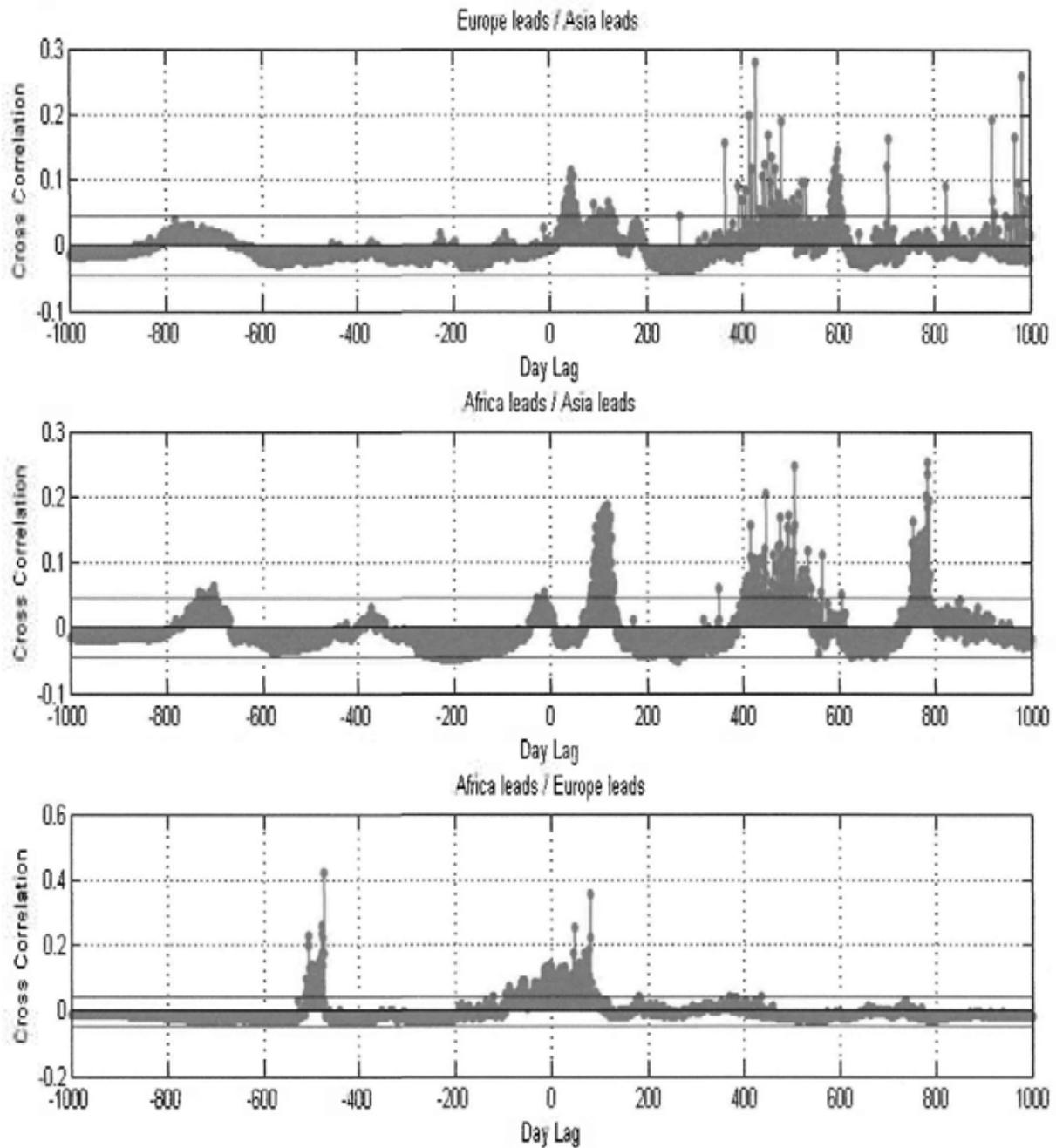


Figure 4.2: Cross-correlation analysis of H5N1 outbreaks of Asia, Europe, and Africa

of devastating occurrences. These intricate processes may be long-range correlated and multifractal. To unravel the temporal processes of avian influenza, fractal and multifractal analysis are useful for studying its dynamics, particularly its long-range correlation and self-similarity. To facilitate our discussion, we first give a brief description of fractal and multifractal analyses. Then we will explain why detrended fluctuation analysis in general and multifractal detrended fluctuation analysis in particular are needed to study the complex outbreaks in time of avian influenza.

#### 4.3.1 Fractal

Complex systems usually contain many components with multiple interwoven fractal subsets of time series having multitudes of scaling exponents. Accurate measure of the scaling exponents play an essential role in the classification and modeling of empirical data, as well as in the analysis of the physical mechanisms producing the scaling phenomena.

Fractal provides a mathematical formalism describing complex spatial and temporal structures (Feder, 1988; Mandelbrot, 1983). It intends to study extremely irregular objects, with fractional dimension, that cannot be easily described by the language of Euclidean geometry. Self-similarity is an underlying concept of all fractals. Self-similar objects repeat themselves on finer and finer stages/levels ad infinitum. Fractals are actually quite common in geographical structures, such as urban patterns and river networks, which have been intensively studied over the years (Lam and Lee, 1993).

Time series may also be recursively divided into self-similar components. In terms of avian-influenza-outbreak over time, one may wonder

whether it is self-similar in multiple time scales. That is, whether it is a self-similar process that consists of miniatures of itself.

The size of a fractal set can be measured by the fractal dimension, which is defined as follows:

$$M_{\delta \rightarrow 0}(F) \propto c\delta^{-s}, \quad s \in \mathbf{R}, \quad (4.1)$$

where  $F$  is a set,  $M_\delta(F)$  is the measurement of  $F$  with dimension  $s$ ,  $\delta$  is the scale, and  $c$  is the  $s$ -dimensional measure of  $F$ . It can be interpreted as the number of covers required to cover the set  $F$ . In short, the required number of covers scales with the dimension.

### 4.3.2 Correlation

For time series, the properties of fractal can be measured from the correlation function. For a time series,  $x_i, i = 1, 2, \dots, N$ , the autocorrelation function can be defined as:

$$C(s) = \langle \bar{x}_i \bar{x}_{i+s} \rangle = \frac{1}{N-s} \sum_{i=1}^{N-s} \bar{x}_i \bar{x}_{i+s}, \quad s > 0, \quad (4.2)$$

$$\bar{x}_i \equiv x_i - \frac{1}{N} \sum_{i=1}^N x_i, \quad (4.3)$$

where  $s$  is the time interval.

If the time series  $x_i$  are uncorrelated, then  $C(s)$  is zero. When  $C(s)$  has the exponential function like  $C(s) \propto \exp(-s/s_c)$  ( $s_c$  is the time decay), the time series has short-range correlation, which means the correlation of  $x_i$  declines exponentially over time. However, when  $C(s)$  is distributed as the power law,  $C(s) \propto s^{-\gamma}$  with the exponent  $0 < \gamma < 1$ , the time series  $x_i$  is supposed to have long-range dependency, which means the autocorrelation function decays gradually to zero with heavy

tails over time.

This chapter is to detect whether there are long range dependencies in the outbreak processes of avian influenza, i.e. whether previous outbreaks have long-term effects on the current occurrences. However, when noise is superimposed on the collected data  $x_t$  and the underlying trends, the direct calculation of  $C(s)$  is inappropriate (Chen et al., 2002).

### 4.3.3 R/S Analysis

Hurst (1951) proposed the rescaled range analysis (*R/S* analysis) to study the scaling behaviors of complex systems. For time series, the *R/S* analysis can be used to detect the underlying temporal dependency, indicated by the Hurst exponent.

Given a time series of length  $n$ ,  $X \equiv \{X_t : t = 1, 2, \dots, n\}$  with mean  $\langle x \rangle_n$  and variance  $S^2(n)$ , the ratio  $R(n)/S(n)$  is defined as:

$$R(n)/S(n) \propto (n/2)^H, \quad (4.4)$$

where

$$R(n) = \max_{1 \leq i \leq n} X(i, n) - \min_{1 \leq i \leq n} X(i, n), \quad (4.5)$$

$$S(n) = \left[ \frac{1}{n} \sum_{i=1}^n (x_i - \langle x \rangle_n)^2 \right]^{1/2}. \quad (4.6)$$

$H$  is the Hurst exponent of time series  $X$ , which can be estimated by the least-squares linear fit of the log-log transform of Equation 4.4.

Theoretically, when the Hurst exponent  $H$  is larger than 0.5, the time series is said to have long-range correlation. When  $H$  is less than 0.5, the time series is said to be long-range anti-correlated. When  $H$

is equal to 0.5, then it means that no correlations exist and the series is considered as random.

Extending on this statistical method, some advanced fractal methods have been developed to examine the long-range correlation and multifractal properties of complex time series in varying time scales.

#### 4.3.4 Multifractal Analysis

In many cases, a single exponent (fractal dimension) is not sufficient to describe the complex dynamics of a multifractal system, such as the temporal behavior of avian influenza as we will show in this chapter. Different regions of an object or series may have different fractal properties. An appropriate description may require a continuous spectrum of exponents instead. In some cases, there might exist crossover time scales separating distinct regimes that may indicate different patterns of fractal structures. Hence, multiple scaling exponents might be required for the full description of the multiscaling behavior.

Multifractal analysis, which enables us to obtain more insight into the scaling behaviors of complex time series, has been used to characterize the spatial heterogeneity of the theoretical and experimental fractal pattern in general (Grassberger and Procaccia, 1983). The most common numerical implementation of multifractal analysis is the fixed-size box counting algorithm (Halsey et al., 1986). For the one-dimension case, the simplest type of multifractal analysis based on the standard partition function is as follows:

Given a measure  $\mu$  with support  $E \subset \mathbf{R}$ , the partition sum can be defined as:

$$Z_\epsilon(q) = \sum_{\mu(B) \neq 0} \mu(B)^q, \quad (4.7)$$

where  $Z_\epsilon(q)$  is the sum of different non-empty half-open-to-the-right

boxes  $B$  with a given scale (width)  $\epsilon$  in a lattice covering the whole support  $E$ , where  $B = [k\epsilon, (k+1)\epsilon[$ . Thus, summing over all possible outcomes is equivalent to summing over all lattice boxes. The dimensions are then defined by scaling via the box width. Then the scaling exponent  $\tau(q)$  is defined as:

$$\tau(q) = \lim_{\epsilon \rightarrow 0} \frac{Z_\epsilon(q)}{\log \epsilon}. \quad (4.8)$$

The scaling exponent (mass function)  $\tau(q)$  is numerically estimated by the linear fit of  $\log Z_\epsilon(q)$  against  $\log \epsilon$  for any value of  $q$ ,  $q \in \mathbf{R}$ .

A relationship between the scaling exponent (mass function)  $\tau(q)$  and the generalized fractal dimension  $D(q)$  is obtained as:

$$D_q = \frac{\tau(q)}{q-1}, \quad q \in \mathbf{R} \quad \text{and} \quad q \neq 1. \quad (4.9)$$

Both  $\tau(q)$  and  $D(q)$  have been employed to examine the multifractal properties of time series (V.V. Anh, 2000; Yu et al., 2009).

**Remark.** Methods such as the  $R/S$  analysis and multifractal analysis work well when the time series are long and trends are not involved. They are, however, not suitable for the analysis of nonstationary time series affected by trends. The existence of underlying trends or exogenous trends might affect the scaling behavior of long-range dependent processes. The detrended fluctuation analysis (DFA) proposed by Peng et al. (1994) provides an approach to identify the monofractality of noisy and non-stationary processes. Extended on this, Kantelhardt et al. (2002) proposed the multifractal detrended fluctuation analysis (MF-DFA) for the study of long-range correlation and multifractal property of time series. It is an appropriate method for the study of scaling behaviors of H5N1.

### 4.3.5 Multifractal Detrended Fluctuation Analysis

The multifractal detrended fluctuation analysis (Kantelhardt et al., 2002) is an extension of the DFA for the detection of correlation and multifractal properties of noisy and non-stationary time series. Since its inception, the MF-DFA has been widely used to identify long range dependency and multifractal properties in economic time series (Lim et al., 2007; Du and Ning, 2008), chemical engineering (Niu et al., 2008), traffic system (Shang et al., 2008), heart rate (Movahed et al., 2006a). It has also been applied to the study of geomagnetic field (Anh et al., 2007; Yu et al., 2010), earthquake, hydrology, and sunspot activities. In this chapter, we attempt to explore the existence of multifractal scaling behavior in the H5N1 outbreak process. The generalized multifractal DFA procedure contains five steps. Supposed that  $x_k$  is a series of length  $N$  with compact support.

Step 1. Determine the 'profile'

$$Y(i) \equiv \sum_{k=1}^i [x_k - \langle x \rangle], \quad i = 1, 2, \dots, N. \quad (4.10)$$

According to Kantelhardt et al. (2002), the subtraction of the mean  $\langle x \rangle$  is not compulsory, since it will be eliminated by the detrending in the third step.

Step 2. Divide the profile  $Y(i)$  into  $N_s \equiv \text{int}(N/s)$  non-overlapping segments of equal length  $s$ . Since the length  $N$  of the series is often not a multiple of the time scales  $s$ , a short part at the end of the profile may remain. In order not to ignore this part of the series, the same procedure is repeated starting from the opposite end. Thereby,  $2N_s$  segments are obtained altogether.

Step 3. Calculate the local trends for each of the  $2N_s$  segments by

a least-squares fit of the series. Then determine the variance

$$F^2(s, v) = \frac{1}{s} \sum_{i=1}^s \{Y[(v-1)s+i] - y_v(i)\}^2, \quad (4.11)$$

for each segment  $v$ ,  $v = 1, \dots, N_s$ , and

$$F^2(s, v) = \frac{1}{s} \sum_{i=1}^s \{Y[N - (v - N_s)s + i] - y_v(i)\}^2, \quad (4.12)$$

for  $v = N_s + 1, \dots, 2N_s$ . Here,  $y_v(i)$  is the fitting polynomial in segment  $v$ . Linear, quadratic, cubic or higher order polynomials can be used in the fitting procedure (conventionally called DFA1, DFA2, DFA3, ...). Since detrending of the time series is accomplished by the subtraction of the polynomial fits from the profile, different orders employed in the DFA differ in their capabilities of eliminating trends in the series. In (MF)-DFA $m$  ( $m$ th-order (MF)-DFA), trends of order  $m$  in the profile (or equivalently, of order  $m-1$  in the original series) are eliminated. Thus, a comparison of the results obtained from different orders of the DFA enables us to estimate the types of the polynomial trends existed in the time series (Hu et al., 2001a; Chen et al., 2002). For simplicity, the linear trends are considered in the present study.

Step 4. Average over all segments to obtain the  $q$ th-order fluctuation function, defined as

$$F_q(s) \equiv \left\{ \frac{1}{2N_s} \sum_{v=1}^{2N_s} [F^2(s, v)]^{q/2} \right\}^{1/q}. \quad (4.13)$$

In general, the index variable  $q$  can take on any real value except zero. For  $q = 2$ , the standard DFA procedure is retrieved.  $F_q(s)$  is only defined for  $s \leq m + 2$ .

Step 5. Determine the scaling behavior of the fluctuation functions

by analyzing the log-log plots of  $F_q(s)$  versus  $s$  for each value of  $q$ . If the series is long-range correlated, it then follows the power law below:

$$F_q(s) \propto s^{h(q)}, \quad (4.14)$$

where  $h(q)$  is the generalized Hurst exponent. When the scaling exponent  $h(q)$  depends on the order of  $q$ , the time series are said to be multifractal. On the contrary, when a time series is monofractal, the scaling exponent  $h(q)$  does not depend on  $q$ . In this case, the scaling behavior of the variance  $F_q(s)$  is identical for all segments  $v$ , and the averaging procedure in Equation 4.13 will obtain the same scaling behavior for all values of  $q$ .

For positive values of  $q$ , the average  $F_q(s)$  will be enlarged by high value of  $F^2(s, v)$  in each segment  $v$ . So,  $h(q)$  describes the scaling behavior of the segment with large fluctuation. For negative values of  $q$ , the average  $F_q(s)$  will be calculated by the small value of  $F^2(s, v)$  in each segment  $v$ . Thus,  $h(q)$  describes the scaling behavior of the segments with small fluctuation.

Figure 4.3, for example, shows two distinct time series of Brownian motion and Gaussian noise, as well as their MF-DFA results. For non-stationary time series like fractional Brownian motion, the scaling exponent  $h(2)$  is larger than 1 and it satisfies the formula:  $H = h(2) - 1$ . For stationary time series such as fractional Gaussian noise, the corresponding scaling exponent is  $0 < h(2) < 1$  and it is equal to the well-known Hurst exponent  $H$  as mentioned above (Feder, 1988; Peng et al., 1994).

The relationships discussed above can shed light on our analysis of the long range dependency and multifractal properties of the avian-influenza outbreak series. The experiments and analysis results are

elaborated in the section to follow.

## 4.4 Analysis Results and Interpretations

The MF-DFA results indicate that avian-influenza outbreaks are long-range correlated and multifractal. Long-range correlation implies that H5N1 can transmit in time and space. Specifically, previous outbreaks of H5N1 infections are responsible for and have long-term effects on the current outbreaks. The property of multifractality indicates that the dynamics of avian-influenza outbreaks are self-similar at all moments  $q$  examined in this chapter, which may be caused by the long-range correlation of the epidemic.

### 4.4.1 Long-range correlation

Figure 4.4 depicts the MF-DFA result of avian-influenza outbreaks of the world. It suggests a strong long-range correlation of the disease because the value of the scaling exponent  $h(2)$  is up to 0.8. In other words, the epidemic is long-range correlated. That is, when a large outbreak occurs at the present moment, a large outbreak is more likely to follow at the next moment in time. The situation is similar for small outbreaks. The long-range correlation of avian influenza detected by the present analysis give a quantitative description of the behaviors of this disease. It has been discovered that avian influenza H5N1 viruses have adapted to the ecological environments (Smith et al., 2006b) and have evolved to establish their genetic diversities crossing large areas of the world (Chen et al., 2006; Cattoli et al., 2009). The multiple sublineages of H5N1 coexisting in common environments are more likely to set up the pool for genetic reassortments and mutations that may facilitate the evolution of the viruses with increased virulence or expanded

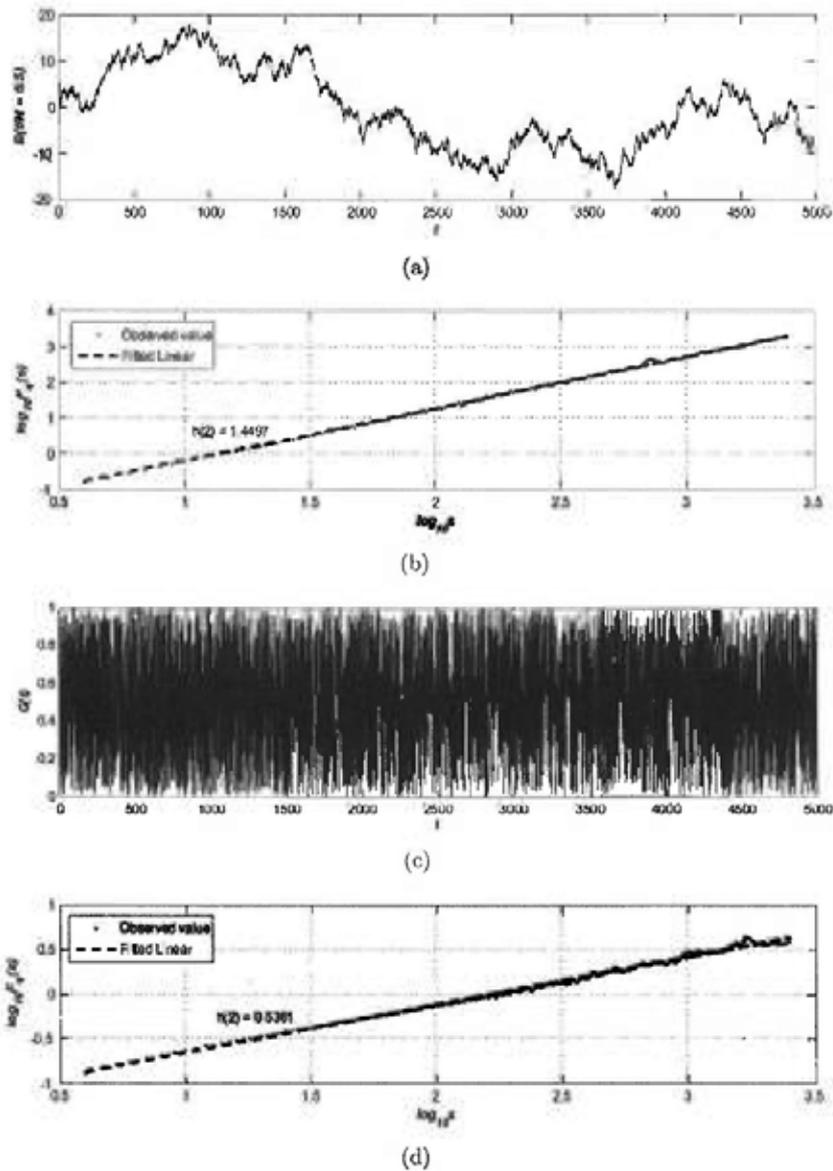
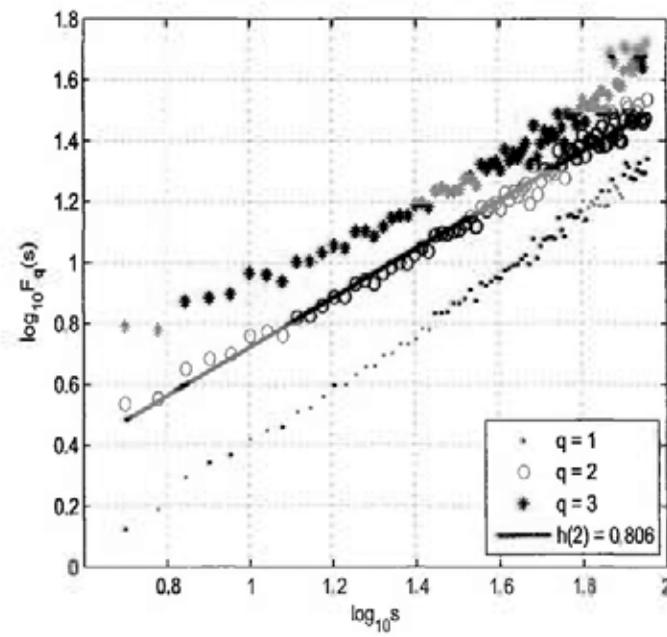


Figure 4.3: Time series of (a) the fractional Brownian motion and (b) the MF-DFA results with  $h(2) = 1.4497$ ; (c) the fractional Gaussian noise and (d) the MF-DFA results with  $h(2) = 0.5361$  for the fractional Gaussian noise

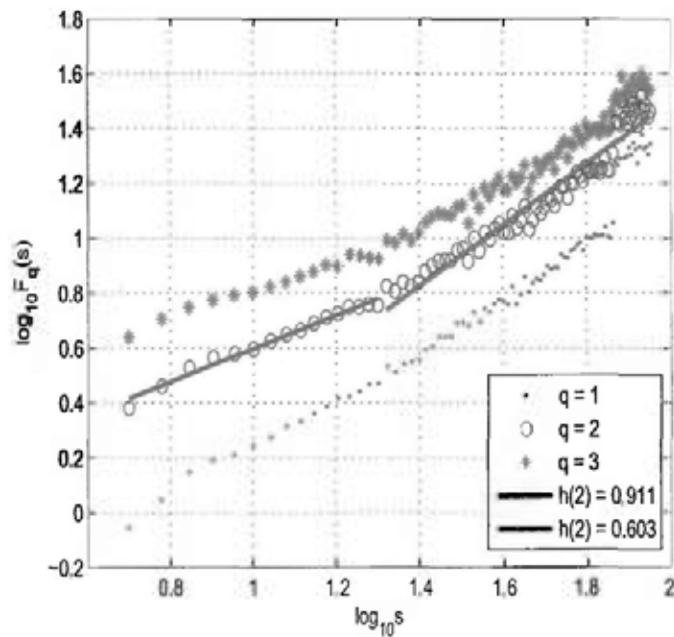
host range (Aubin et al., 2005; Duan et al., 2008). Continued circulation of the avian-influenza viruses in poultry and wild birds have led to repeated outbreaks over the years (Chen et al., 2006; Tiensin et al., 2007). In general, the continuous outbreaks that are temporally associated demonstrate a long-term persistence, derived probably from the genetic behaviors of H5N1. Specifically, the degree of the long-range correlation of the global outbreaks has mainly been intensified by the diseases in southeastern Asia and Africa, since over 80% of the world's outbreaks emerged continuously in the two continents, infecting primarily the poultry. This might imply that current public health practices and disease control measures are inadequate for preventing the outbreaks and the spread of avian influenza at a global level.

Figure 4.4(b) is the MF-DFA plots of H5N1 outbreaks of Asia. The disease appears to be long-range correlated in general. The scaling behavior in Asia exhibits a special crossover time scale at around 20 days, separating two distinct outbreak regimes. For scales of 5 days to around 20 days, the disease outbreaks display a weaker long-range correlation as the value of the scaling exponent  $h(2)$  is close to 0.6, but for scales up to 90 days and beyond, the outbreaks are long-range correlated and the value of the scaling exponent  $h(2)$  is larger than 0.9. This implies that there might be different mechanisms driving the outbreaks in the two temporal regimes.

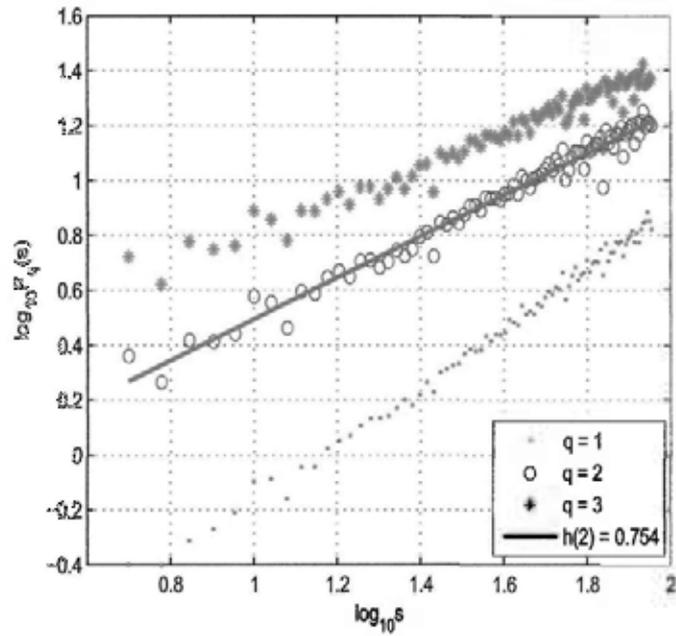
The crossover time scale identified in this study roughly coincides with the incubation period of avian influenza A H5N1, ranging from 2 to 8 days or as long as possibly 17 days (WHO, 2006a). Si et al. (2009) also discern that the time length of incubation for avian influenza is about 21 days, depending on the species exposed. Our result essentially indicates in mathematical manner the crossover time scale of around



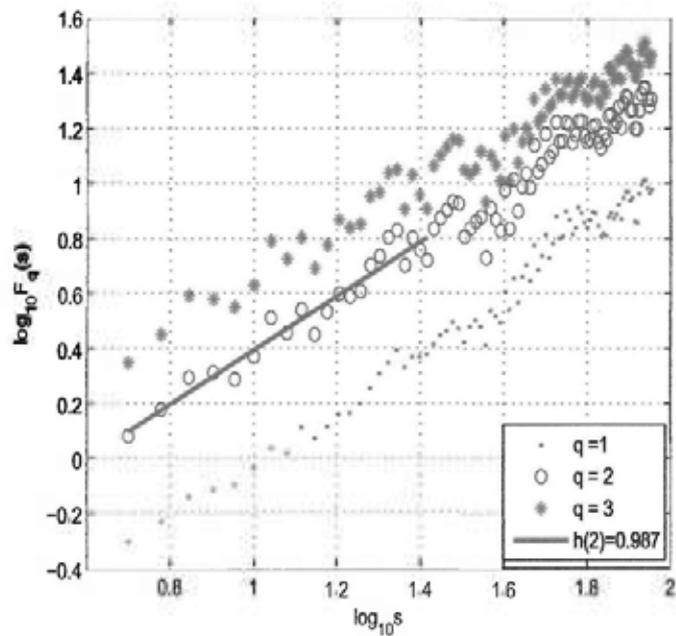
(a) The whole world



(b) Asia



(c) Europe



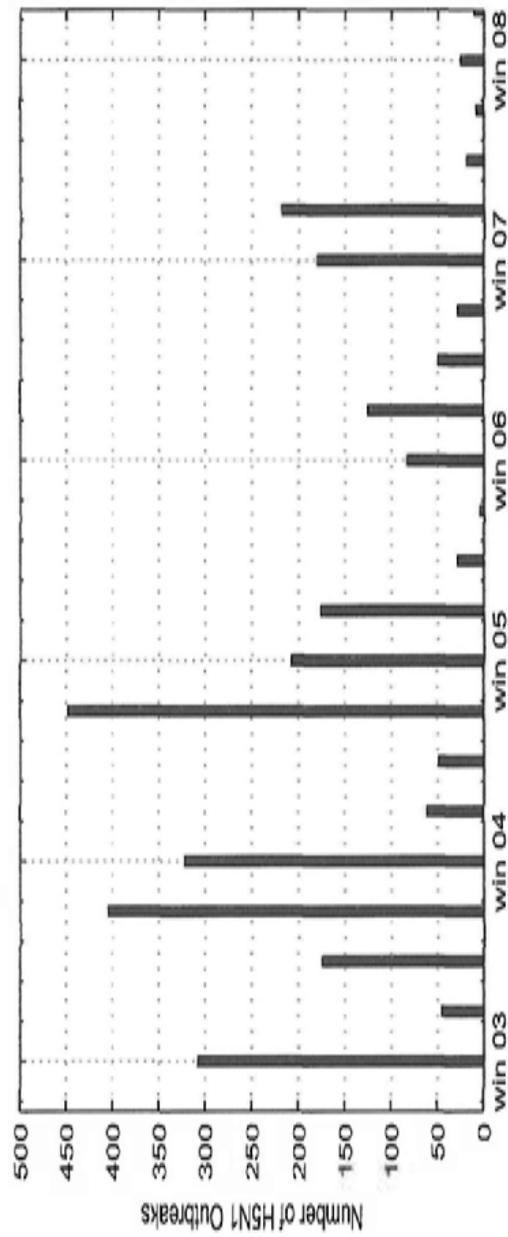
(d) Africa

Figure 4.4: The MF-DFA results of H5N1 outbreaks of (a) the whole world, (b) Asia, (c) Europe, and (d) Africa.

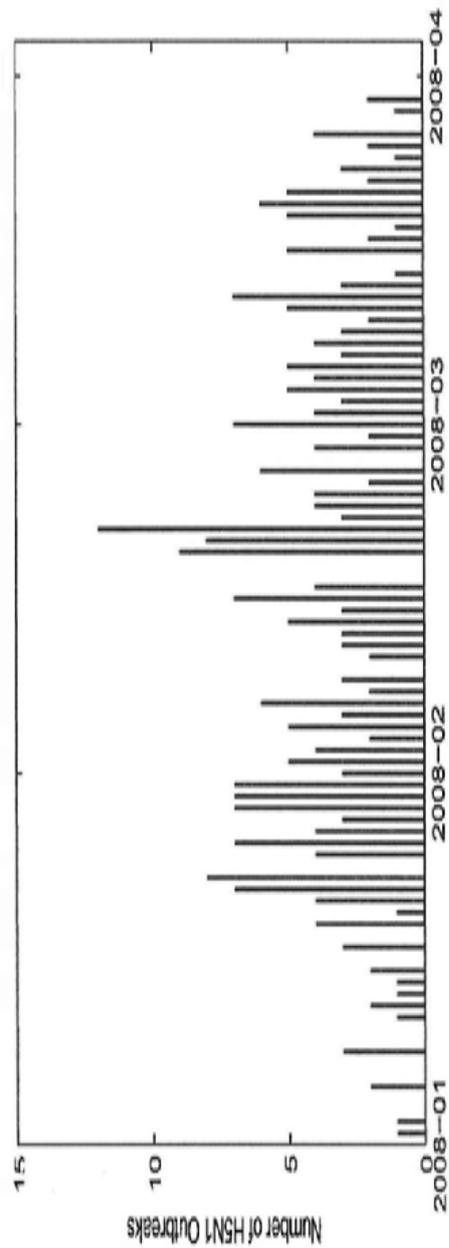
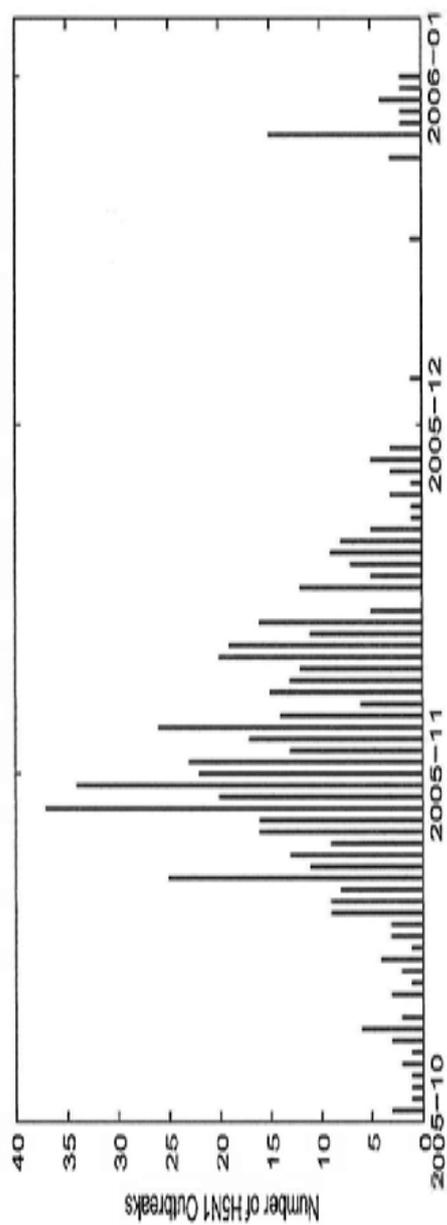
20 days, at which the behavior of avian-influenza outbreaks happens to change.

For larger scales (the 20-day and above scales), the outbreaks in Asia have strong long-range correlation, with the scaling exponent  $h(2)$  having a value larger than 0.9. It generally depicts the process in which large outbreaks are usually followed by another large outbreak at the next moment of time. Avian-influenza outbreaks in Asia, depicted in Figure 4.5(a) shows a strong seasonal pattern with a high frequency of outbreaks that start from late autumn and last through the entire winter, and then a reflux with the coming of the following spring. The seasonal pattern has been observed and reported by Guan et al. (2004) and Smith et al. (2006a).

Indeed, the strong long-range correlation obtained in our analysis further characterizes quantitatively the behaviors of the disease in a formal manner. For the period from late autumn in October to the early spring of next year in around April, repeated H5N1 outbreaks have largely taken place in most areas of Asia, including Thailand, Vietnam, Indonesia, and southern China (Yee et al., 2009). Figure 4.5(b) and 4.5(c) give instances of the intensive H5N1 outbreaks in the winter of 2005 and 2008, respectively. Previous studies have shown that lower temperatures make the viruses more viable with persistence, increasing the likelihood of the wide spreading of the epidemic (Li et al., 2006; Preiser, 2006). Moreover, cold air may inhibit the immunity of poultry against avian-influenza disease, leading to high frequency of H5N1 outbreaks in winter. In addition, Si et al. (2009) indicate that during the autumn migration from August to around November, wild birds migrated from north Siberia to Australia via East Asia. This also leads to the seasonality of the outbreaks in southeastern Asia. Besides these



(a) H5N1 outbreaks of four seasons from 2003 to 2008 in Asia (note: win is for winter)



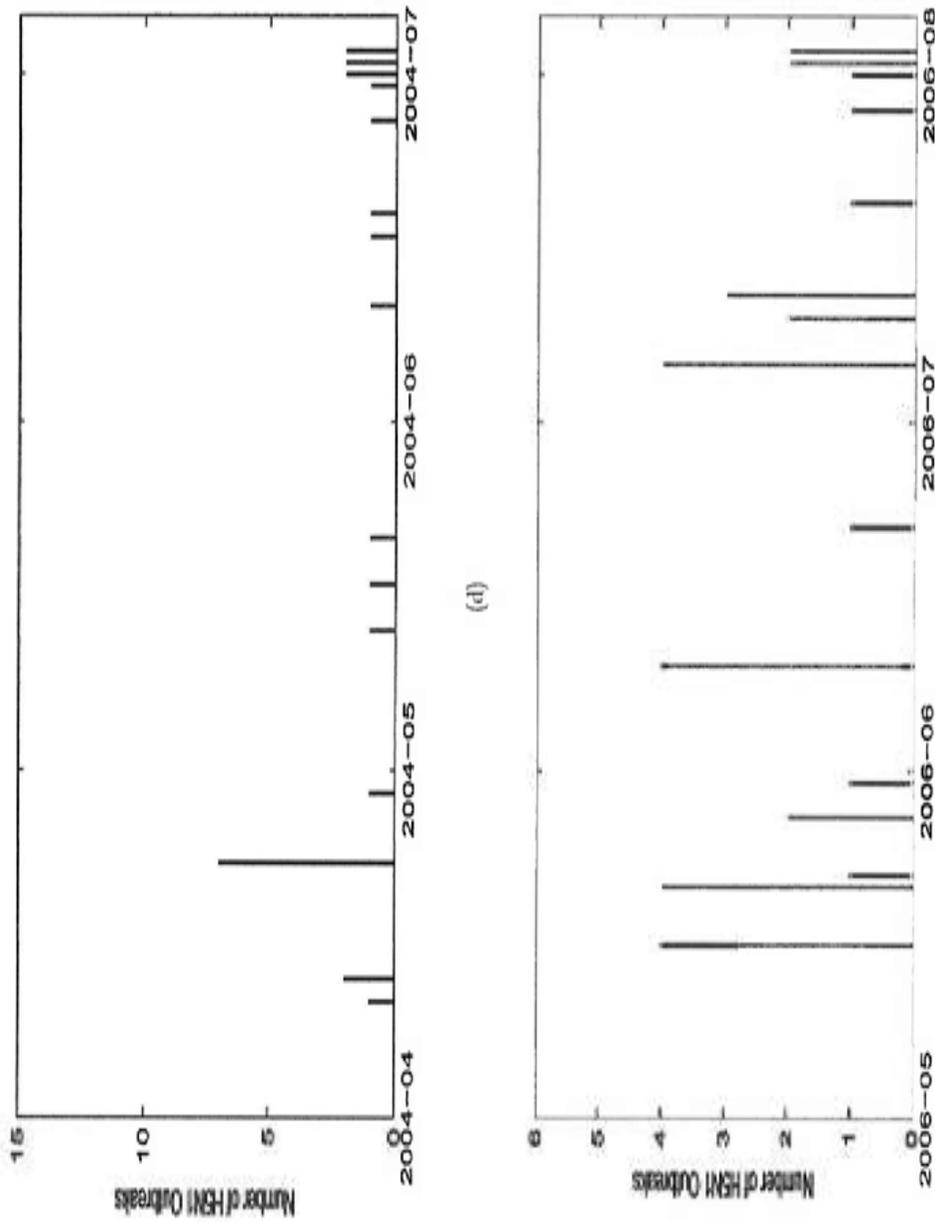


Figure 4.5: Avian influenza outbreaks in Asia (a) seasonal pattern of H5N1 outbreaks from the winter of 2003 to the end of 2008 in Asia; (b) and (c) large outbreaks repeated in late autumn 2005 and in winter 2008; (d) and (e) sporadic outbreaks starting in spring 2004 and in early summer 2005.

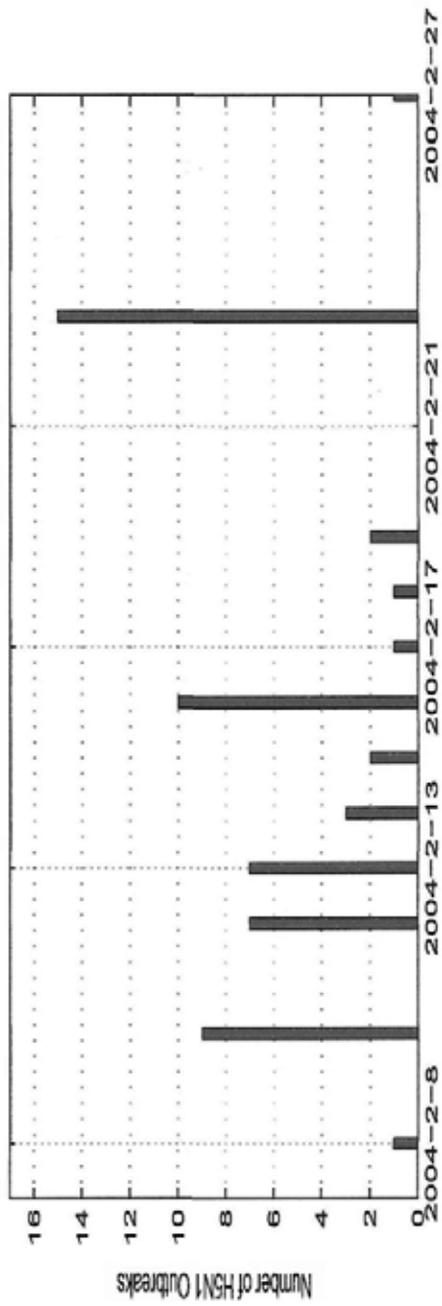
ecological and environmental factors, the long-range correlation of the outbreaks in Asia might also be ascribed to social-economic behaviors. Traditional agricultural modes and free-grazing practices have been attributed to the persistence of the epidemic (Gilbert et al., 2006a, 2007, 2008), as domestic ducks have been affirmed being the Trojan horse of the highly pathogenic H5N1 influenza in Asia (Hulse-Post et al., 2005). Furthermore, we may notice that most festivals, such as Thanksgiving, Christmas, and New year, occur in winter. This may increase the risk of avian-influenza outbreaks owing to an increase in the import of poultry from other countries to meet market demands (Smith et al., 2006a). For the period starting from the early spring in April, sporadic outbreaks lasting through the entire summer also demonstrate a strong long-range correlation. Figure 4.5(d) and 4.5(e) show the sporadic outbreaks in the early spring of 2004 and in the summer of 2006, respectively. The disease is rather long-range correlated: sporadic outbreaks appear at this moment tend to take place at the next moment in time. The near disappearance of the disease seems to suggest that avian-influenza viruses of Asia undertake an hibernation over the whole summer. From the phylogenetic analysis, Chaichoune et al. (2009) believed that a small number of H5N1 viruses silently persist and survive the inter-outbreak period in summer, while most H5N1 viruses were extinct, causing a bottleneck effect in evolution. This may suggest an appropriate time to eradicate the influenza H5N1 viruses. Oftentimes, these small pool of viruses are periodically amplified to facilitate genetic reassortments leading to large epidemics when outbreak season returns. All the factors attribute to the specific scaling behavior of the outbreaks in Asia. Long-range correlation actually signifies the scaling behaviors of the disease because the more long-range correlated the

outbreaks are, the higher is the value of the scaling exponent  $h(2)$ .

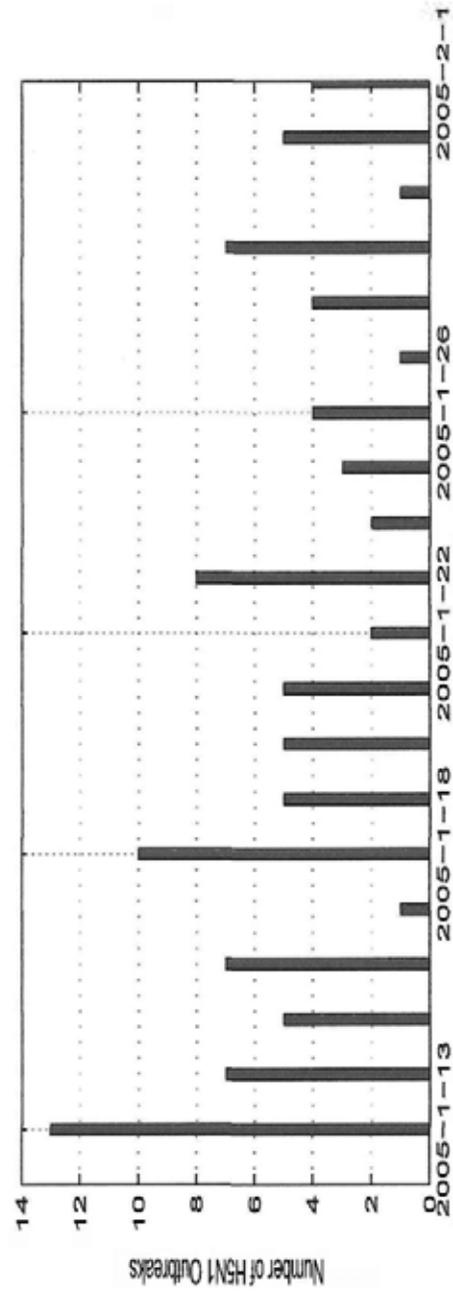
However, such long-range dependence does not hold at time scales smaller than 20 days. The value of the scaling exponent  $h(2)$  is close to 0.6, which suggests a weaker long-range correlation of the disease in Asia (see the Figure 4.4(b)). For smaller scales from 5 to around 20 days, the outbreaks appear to be nearly random as the value of the  $h(2)$  is approaching 0.5, as explained above by the  $R/S$  analysis and the MF-DFA. The distinct scaling behavior may imply a different mechanism of the outbreak at these shorter time scales. To generate an intuitive insight into this epidemic, the time series of the outbreaks from 10<sup>th</sup> December 2003 to 21<sup>st</sup> March 2009 was divided into 96 sections based on the 20-day time scales. Figure 4.6 shows the instances of the outbreaks, extracted from 2004, 2005, 2006, and 2008, respectively. Compared to Figure 4.6, the outbreaks at smaller time scales of 20 days are not so long-term persistent. In these cases, large outbreaks occur at the present moment may not be followed by large outbreaks at the next moment. It may be interpreted that the incubation period for avian influenza H5N1 is longer than that of the conventional seasonal influenza (WHO, 2006b), which usually lasts a few days but rarely up to 21 days depending upon the characteristics of the isolate, the dose of inoculum, and even the host species and its age (Preiser, 2006). Many birds die suddenly and large outbreaks subsequently happen with premonitory signs. This has been speculated that wild aquatic birds play a crucial role in carrying the H5N1 influenza virus over long distances during the incubation periods (Normile, 2005a). Effective control measures and rigorous biosecurity practice have been conducted by the infected countries or regions to prevent the disease from spreading wide and causing larger outbreaks. The Hong Kong authorities, for example,

culled the entire poultry population (around 1.5 million birds) for the immediate eradication of the highly pathogenic avian influenza soon after the early detection of the outbreaks in 1997 (Preiser, 2006). In addition, Japan and Korea quickly took aggressive strategies through culling massive number of birds and implementing movement controls (Ozawa et al., 2006) when the HPAI H5N1 were first detected in January 2004 and December 2003, respectively. The rapid and determined responses have effectively controlled the disease, preventing the large outbreaks that might follow. On the other hand, such measures, aiming at the immediate eradication of the disease at the cost of birds culling, are usually infeasible in those countries, i.e., Cambodia, Vietnam, and Thailand, that pursue traditional forms of poultry holding. Chickens and ducks roam freely and share common water sources together with wild birds (Yee et al., 2009). Even vaccination fails to control the disease in China and Indonesia where repeated outbreaks appear long-term persistent (Webster et al., 2006; Yee et al., 2009). The weaker long-range correlation essentially characterizes the intrinsic dynamics of the outbreaks in smaller scales. From this result, we may affirm that massive control measures have been carried out to prevent the outbreaks and the spreading of avian influenza. However, the current disease surveillance and control practice are inadequate in effectively eradicating the epidemic. The scaling exponent discussed in this study may in a sense serve as an indicator for the evaluation of the effects of disease control and prevention.

With respect to Europe and Africa, avian influenza appears to be long-range correlation also (see Figure 4.4(c) and 4.4(d)). The European outbreaks exhibit weaker long-range correlation with its scaling exponent  $h(2) = 0.754$  for the time scales from 5 to 90 days. In Africa,



(a)



(b)

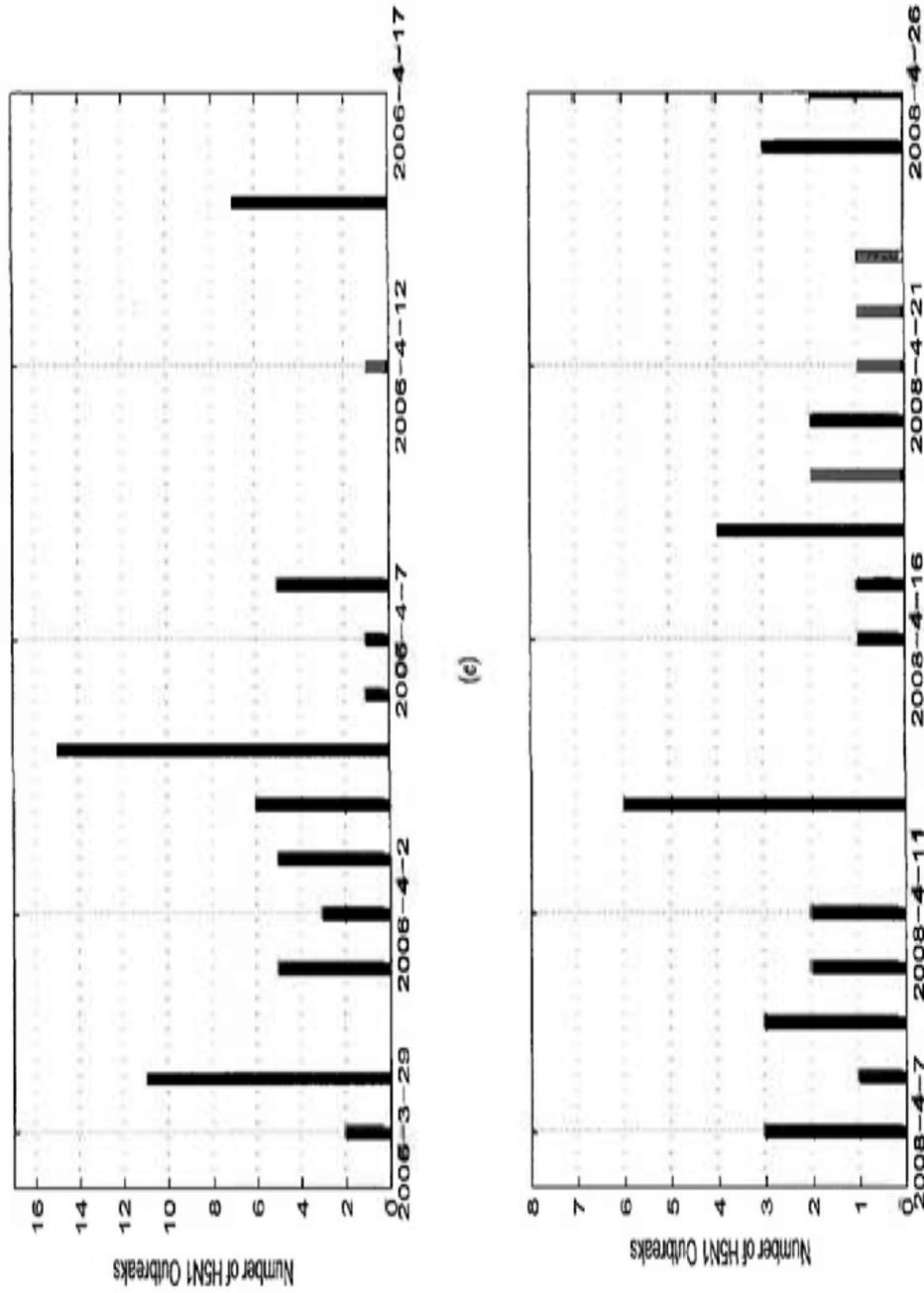


Figure 4.6: The H5N1 Outbreaks of Asia at 20-day time scale, starting from (a) 8<sup>th</sup> February 2004, (b) 13<sup>th</sup> January 2005, (c) 29<sup>th</sup> March 2006 and (d) 7<sup>th</sup> April 2008, respectively

long-range correlation is considerably more apparent as the value of  $h(2)$  is larger than 0.98. The distinct values of the scaling exponent in the three continents may indicate that avian-influenza processes are heterogeneous with plausibly different driving mechanisms.

The majority of the infections take place among wild birds rather than poultry (WHO, 2010b). Since the H5N1 outbreaks in migratory birds at Qinghai Lake in April 2005 (Chen et al., 2005; Liu et al., 2005), the highly pathogenic avian influenza has spread across eastern and central Europe arriving at Croatia, Romania, and the European part of the Russian Federation at the end of the year. Within a short period between January and March 2006, 18 European countries were infected, primarily in wild birds, east of the line from south-eastern Sweden to south-western Italy (OIE, 2010). Such rapid spreads of the disease seem to be associated with the dispersal of wild birds (Lebarbenchon et al., 2009). The outbreaks that emerged in domestic poultry in Europe have been assumed to be primarily linked to the behaviors of wild birds also (Martinez et al., 2009; Szelezky et al., 2009). Quick actions have been carried out by the European Commission on the curbing of the disease (Normile, 2005b; Munasinghe et al., 2008). Many countries in Europe urged their farmers situated along the migratory routes or near wetlands to keep their poultry indoor in order to minimize the chance of having close contacts with wild birds that may carry influenza viruses (Normile, 2005a). This makes the outbreaks of Europe appear to be not as long-term persistent as that of Asia and Africa. The positive surveillance aiming at the early detection of H5N1 in wild birds has been established in many European countries. However, large outbreaks of the disease that emerged at the current moment usually re-emerged at the next moment in time. The persistence of the

epidemic may be due to the behaviors of wild birds (Munster et al., 2007). Specifically, different species have different behaviors in seasonal migration for wintering and breeding, as well as movements for foraging and moulting (Globig et al., 2009). Such behaviors that depend on climate conditions may interfere with disease prevention and control activities, leading to the recurrence of avian influenza in Europe (Gilbert et al., 2006b; Liu et al., 2007). Our results reveal different mechanisms of the outbreaks in Europe and in Asia and Africa. The weaker long-range correlation in Europe might be a result of its rigorous biosecurity and public health practice. As migratory birds share common areas for wintering and breeding, it is important for Europe to implement surveillance of wild birds at distinct geographical sites and to employ comprehensive control measures to prevent the spread of the epidemic.

Although the H5N1 viruses of Africa show a close phylogenetic relationship with that of Europe (Salzberg et al., 2007), the African outbreak mechanism is considerably more complex and distinct. The outbreaks of Africa demonstrate strongest long-range correlation with the value of the scaling exponent  $h(2)$  measured up to 0.98. Since the first H5N1 outbreak was reported in Kaduna State, Nigeria, in mid-January 2006, the disease has spread across Africa within a short time period, affecting Nigeria, Niger, Sudan, Egypt, Burkina Faso, Djibouti, Ivory Coast, Ghana, Togo, Cameroon, and Benin (OIE, 2010). Oftentimes, the mechanism of the outbreak might vary with the evolution of H5N1 (Kilpatrick et al., 2006). Three H5N1 lineages have been identified in Nigeria, suggesting multiple introductions of the epidemic from distinct sources into Africa (Ducatez et al., 2006; Cattoli et al., 2009). At least, three migratory bird flyways, with many bird sanctuaries along

western Africa, link the territories of western Siberia, Europe, Mediterranean, and western Asia. The association of outbreaks and flyways implies the role migratory birds play in the spread of H5N1, although poultry imports and trades cannot be ruled out (Ducatez et al., 2007). The complex ecological systems of wild birds and poultry, moreover, exacerbate the long-term co-circulation of the multiple H5N1 lineages, leading to large repeated outbreaks in Africa. Backyard rearing is very common in Africa. Over 65% of 140 million birds in Nigeria for example are free-range poultry, the majority of which are chickens and guinea-fowls (Adene and Oguntad, 2006). This also facilitate the transmission of avian influenza in poultry. In addition, the wetlands that serve as important habitats for both migratory and resident water birds increase the risk of H5N1 infections in poultry and humans in places where extensive rice farming, fishing, cattle grazing, and poultry rearing make frequent contacts between humans and birds (Cecchi et al., 2008). Recent climate change may also exert influences on birds migration causing extensive disease outbreaks across the continent (Cecchi et al., 2008). For Africa, poverty, inadequate infrastructure including poor hygiene facilities and insufficient medical care, and political hurdles further compromise the control measures (Enserink, 2006). These result in the long-term persistence of the disease and the multiple introductions of H5N1 viruses (Ducatez et al., 2006). The new genetic reassortment of H5N1 viruses, commonly identified in the poultry of Nigeria, reflects the inadequate biosecurity measures and poor disease prevention in Africa (Moone et al., 2008). All of the above are perhaps the reasons for such repeated large outbreaks in this continent. Our result actually indicates the strongest long-range correlation of the outbreaks in Africa, and echoes the call in the literature for strict

biosecurity practices and sustainable control measures in order to keep the disease from further spreading.

To recapitulate, this chapter has identified different long-range correlations of avian-influenza outbreaks among the world as a whole and the three continents, namely Asia, Europe, and Africa. The analysis shows that the mechanisms driving the outbreaks are spatially heterogeneous under different ecological and social-economic environments. The continents need to devise control measures appropriate to their own situations. Furthermore, this result, different degrees of the long-range correlation, may serve as a reflection of different disease surveillance and controls put into practice. The crossover time scales at which outbreak behaviors varied may indicate distinct mechanisms of the disease at different time scales. This finding is also instrumental to the design of different biosecurity strategies and hygiene practices with respect to different time scales.

#### 4.4.2 Multifractality

In the present study, we have detected the multifractality of the time series of H5N1 outbreaks. This property indicates that the multifractal behaviors extend over different mathematical moments  $q$ . In other words, avian-influenza outbreaks exhibit diverse self-similarities at different moments. Oftentimes, the property of a data set can be statistically evaluated in terms of mean, variance, and even higher moments. Similarly, the property of self-similarity of time series can be described by distinct moments  $q$ .

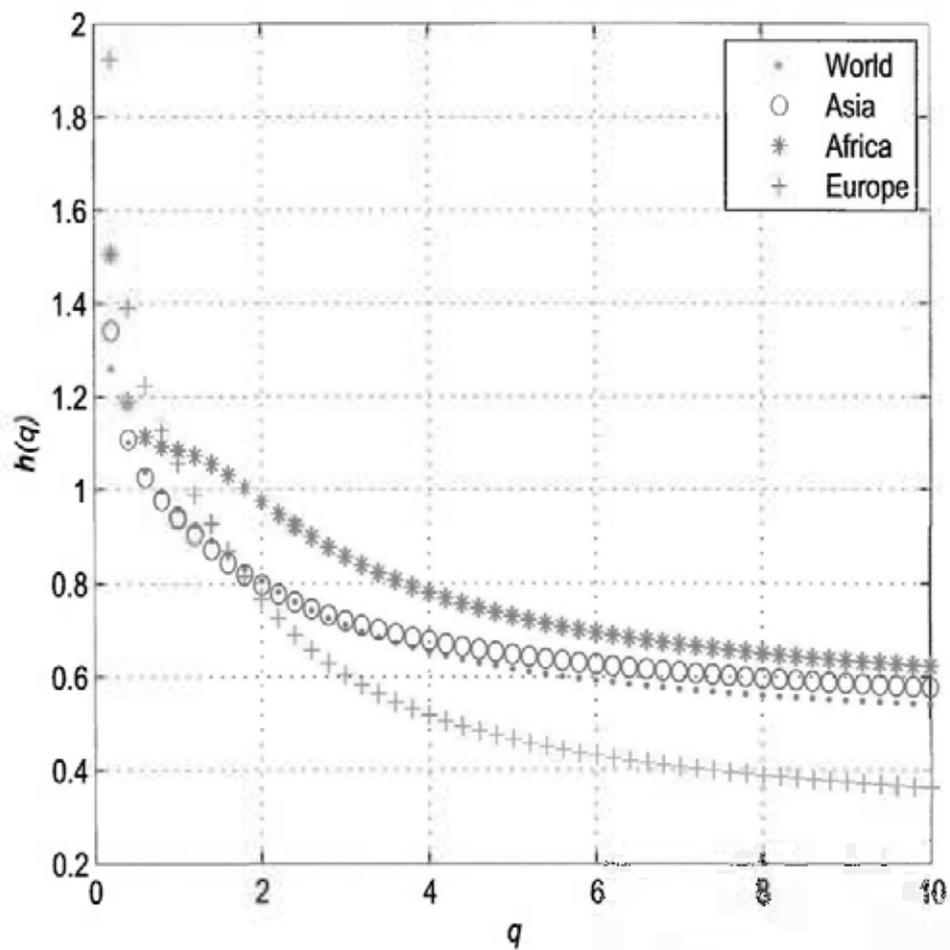
Mathematically, the multifractality of time series can be detected by the generalized Hurst exponent  $h(q)$ , the mass function  $\tau(q)$ , or the generalized fractal dimension  $D(q)$ . Figure 4.7(a),(b) and (c) show the

dependence of  $h(q)$  on  $q$ , and the nonlinear  $\tau(q)$  and  $D(q)$ , respectively. All these indicate that the time series of the H5N1 outbreaks in general and the three continents in particular are multifractal.

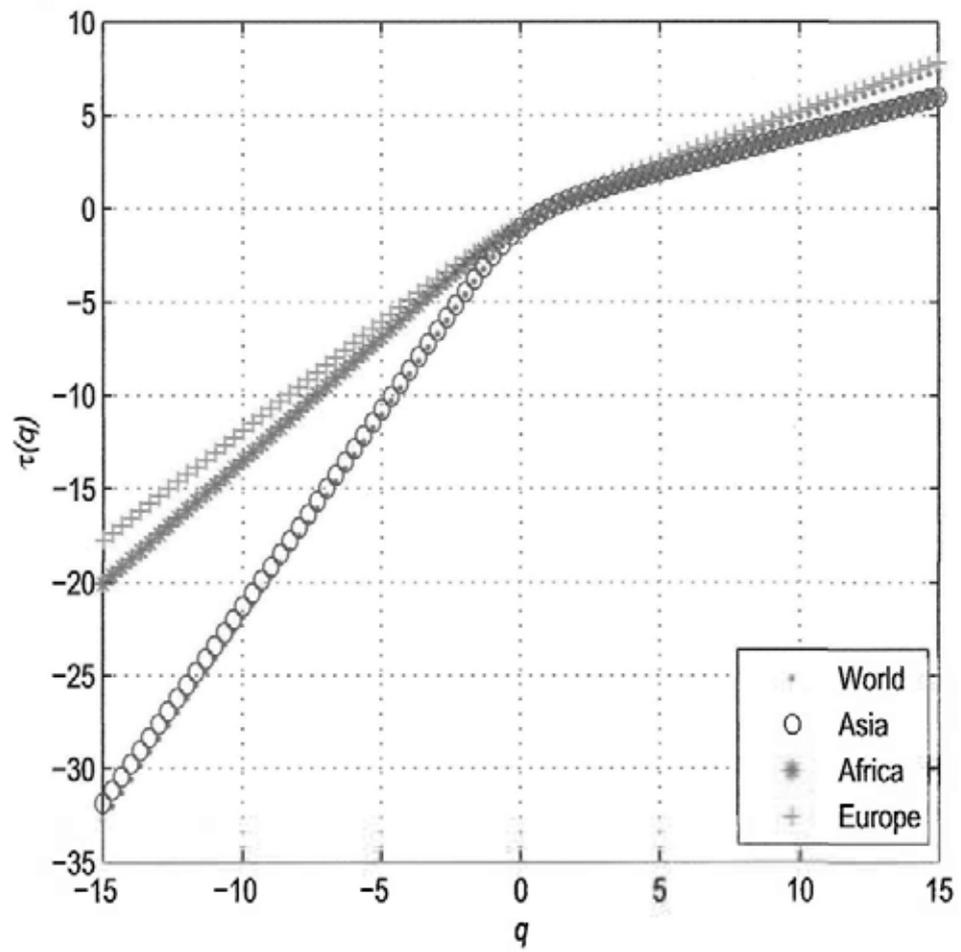
Specifically, the generalized Hurst exponent  $h(q)$  is the dimension of the measure  $F_q(s)$  in Equation 4.14, which has often been used to characterize scaling behavior and multifractality of time series. When the time series is multifractal,  $h(q)$  will have diverse values describing fractal properties from different moments  $q$ . For monofractal time series, the value of  $h(q)$  will not change with different moments, depicting a single form of self-similarity.

Alternatively, the multifractal property can also be reflected by  $\tau(q)$  and  $D(q)$ . The mass exponent function  $\tau(q)$  from the partition function proposed by Halsey (1986), especially the value of  $\tau(q)$  at  $q = 0, 1$ , and  $2$ , is useful for describing the degree of multifractality over multiple moments  $q$ . It measures the variation of the  $q^{\text{th}}$  moment of the fractal measure with respect to scales. If the measure is multifractal, the mass exponent  $\tau(q)$  must be a nonlinear function. Similarly, the general fractal dimension  $D(q)$ , having the relationship with  $\tau(q)$  described in Equation 4.9, shows another identification of multifractality in terms of different moments. For instance,  $D(q)$  can be interpreted as a fractal dimension when  $q$  is equal to 1. If the value of  $q$  are 2 and 3, then  $D(q)$  represents the information dimension of entropy and the correlation dimension of a time series, respectively. For values larger than 3, it is, however, difficult to find appropriate physical meanings of  $D(q)$ . Mathematically, multifractality can be completely determined by the entire moments of  $\tau(q)$  and  $D(q)$ .

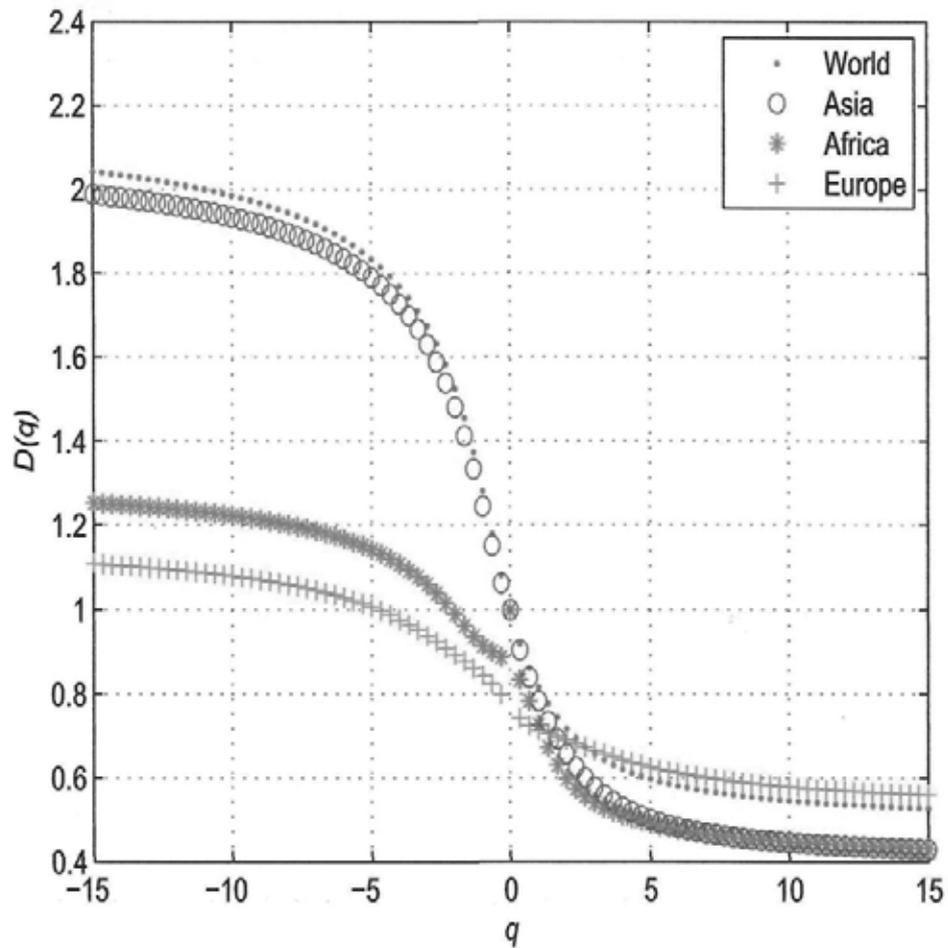
The experimental results indicate the self-similarity of avian influenza for all states, which seem to suggest that multifractality of the



(a)



(b)



(c)

Figure 4.7: The multifractalities of the H5N1 outbreaks of the whole world, Asia, Europe, and Africa: (a)  $h(q)$ , (b)  $\tau(q)$ , and (c)  $D(q)$

disease are probably due to the diversity of ecological environments and the complexity of social behaviors. The long-range correlation, in mathematics, may lead to multifractal scaling behaviors, but more comprehensive studies will be needed to confirm and further explain this interesting property. The multifractal behavior of avian-influenza outbreaks appears to be first discovered by the present study.

## 4.5 Summary

In this chapter, we have studied the scaling behavior of avian influenza A H5N1 in wild birds and poultry over time. Through the application of MF-DFA to the analysis of the H5N1-outbreak time series, we have successfully detected the long-range correlation and multifractal properties of the disease. Our results show that repeated outbreaks have long-range correlation in the sense that outbreaks have great effects on and are responsible for recurrent infections and future spreads. Furthermore, crossover time scales separating distinct outbreak regimes have also been unraveled. In particular, the seasonal pattern has been identified. Comparisons of the scaling behaviors of the outbreak time series by continents, however, reveal that the spread mechanisms are spatially heterogeneous. It is possibly due to different ecological environments, public health policies, and poultry farm practices. The results suggest that tighter surveillance and control with international cooperation is necessary for the prevention of avian influenza outbreaks and spreads.

The present study is limited by the availability of the time series data. With finer and more data made available in the future, we can achieve a deeper understanding of the underlying process generating the temporal scaling behavior of the H5N1 time series. To have a more

complete picture, the next chapter is to study the scaling behavior of H5N1 in space. The spatial and temporal dimensions can then be tied together to construct a space-time process of avian-influenza outbreaks and spreads. Such studies will advance the rigorous analysis of the disease and provide practical guides to policy formulation.

---

□ End of chapter.

## Chapter 5

# Spatial Analysis of H5N1 Outbreaks

*The sound of progress is perhaps the sound of plummeting hypotheses.*

— Peter Haggett, 1965

The last chapter explored the long-range correlation and multifractality of the time series of H5N1 outbreaks. It showed different scaling behaviors of H5N1 outbreaks in Asia, Europe, and Africa. The scaling behavior can be used to interpret the mechanisms by which avian influenza outbreaks move spatially in a heterogeneous fashion, as mediated by the world's different ecological environments.

This chapter deals with the study of the spatial patterns of H5N1 outbreaks over a range of spatial scales. By applying a general and local  $K$  function to the outbreak data, the global trend and local clusters of avian influenza are identified respectively. A modified local  $K$  function is proposed to take into account the spatial effects caused by distance between outbreaks. This method in effect extends the analysis of the local  $K$  function from a point pattern problem to a lattice for points

over continuous space. Spatial representation in these terms then seeks to explore local patterns of H5N1 at various scales, including for the whole of East and Southeastern Asia and for the individual countries: Thailand, Vietnam, Indonesia, and China.

## 5.1 Introduction

It is essential to understand the spatial distribution of avian-influenza outbreaks if the disease is to be better understood. Avian-influenza outbreaks appear in clusters in both the global and local contexts. Parameters and measures taken from spatial statistics, for example, the  $K$  function (Ripley, 1976) and the local  $K$  function (Getis, 1984), have been commonly used to identify spatial patterns represented as a point process. We can further analyze avian influenza by these means.

However, the usefulness of the local  $K$  function is restricted when the events under modeling, such as avian-influenza outbreaks, cannot plausibly be characterized as spatially homogeneous. In most real world cases of epidemics, an assumption of spatial homogeneity is hard to sustain (Diggle, 1983; Cressie, 1991), as in the case of avian influenza. Avian-influenza outbreaks, rather, appear to be spatially associated with each other without being homogeneous. The closer it is to a disease cluster, the greater the number of outbreaks a site can expect (Tobler, 1970). In order to describe this pattern of spatial association accurately, this work proposes a weighted form of the local  $K$  function specifically adapted for examining clusters of H5N1 outbreaks. This modified function, which measures the count of outbreaks by the weight of the distance between them, enables the study to extend the cluster analysis of outbreaks from a depiction of a number of point process to the representation of data arranged in lattices.

The objective of this chapter, then, is to unravel the spatial scaling behavior of H5N1 outbreaks in humans and poultry. The chapter has three more specific aims: (1) to determine global patterns of the H5N1 outbreaks in humans and avians; (2) to explore local clusters of poultry outbreaks over a range of spatial scales; and (3) to estimate the probability of an occurrence of the disease being caused by outbreaks in surrounding areas. Findings for all of these questions promise to help improve public health practice by offering guidelines for the more effective management of infectious disease.

The chapter proceeds by first describing (in section 5.2) the data for avian-influenza outbreaks. In section 5.3, traditional  $K$  and local  $K$  functions are applied to this data in tracing global and local patterns of H5N1 across the world. In response to the problems of applying a traditional local  $K$  function to spatially heterogeneous phenomena, we propose a weighted local  $K$  function for identifying outbreak clusters, and then estimating the probability of the presence of H5N1 across lattices. In section 5.4, we examine and interpret the results of this lattice analysis.

## 5.2 Experimental Data

This chapter takes as its topic the spatial patterns of avian-influenza outbreaks in humans and birds. Outbreak data are taken from the WHO and OIE official Reports (WHO, 2010b; OIE, 2010)

Each report states outbreak attributes, noting outbreaks' latitude, longitude, report data, start date, end date, infected species, and number of deaths caused. The OIE's term for any reported occurrence of the disease is an "outbreak" of H5N1. The chapter subjects to analysis data from more than 5,000 separate outbreaks, covering a period from

late 2003 to early 2009. In the earliest two years of reports, longitude and latitude data are absent from the reports provided by the OIE. These outbreaks in this analysis are geocoded on the basis of their locations. The World Geodetic System (WGS) 84, currently widely used in cartography, geodesy and navigation, is applied by the study as geographic coordinate system. Figure 5.1 shows the locations of H5N1 outbreaks.

In order to display in our analysis, points in a geographic coordinate system need to be expressed as a pair of Cartesian coordinates (Robinson, 1978). The study uses the Mercator projected coordinate system (Raisz, 1938), a standard cylindrical map projection, to make the conversion. This projection has severe distortion with increasing distance from the Equator, but maintains the shape of the continents. Taking the advantage of Mercator in mapping, we used this projection for displaying avian-influenza outbreak data. Actually, the modified local  $K$  function and other spatial quantitative methods in Chapter 6 are developed on the basis of lattice points and distances, not the Mercator projection. That is, the distortion caused by the Mercator projection has essentially no effect on the computation results. Further, the study applies a modified local  $K$  function to extend cluster analysis on lattice data. Outbreak data are collated for lattices based on different spatial resolutions: 8.4 km<sup>2</sup>, 34.22 km<sup>2</sup>, 0.94 km<sup>2</sup>, 0.32 km<sup>2</sup>, and 0.24 km<sup>2</sup> for East-Southeast Asia, China, Indonesia, Thailand, and Vietnam, respectively. Spatial resolutions are determined in terms of the area under study. The larger is the area, the lower is the resolution adopted. Due to the limitation of computation power, East-Southeast Asia and China have to adopt lower resolutions than Indonesia, Thailand, and Vietnam. On the other hand, to provide a finer description

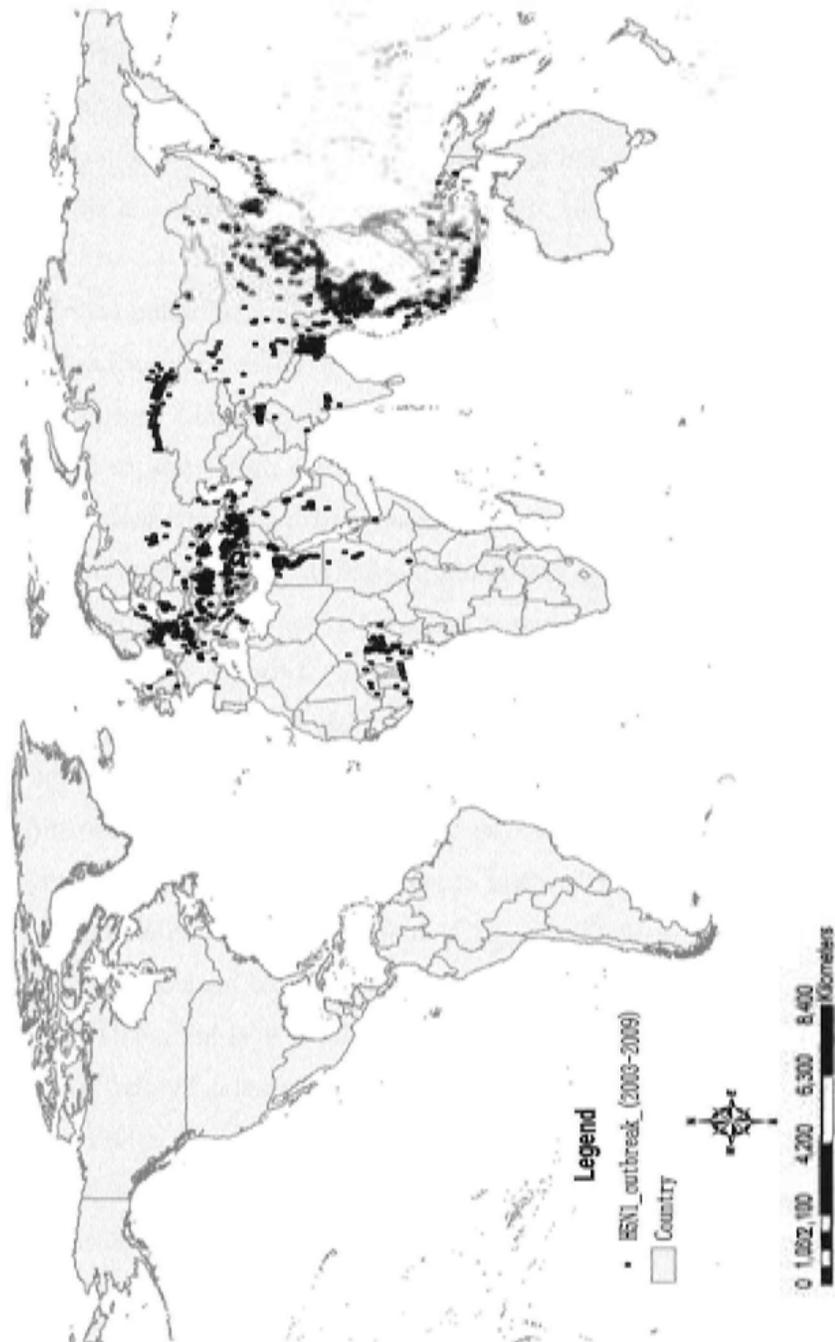


Figure 5.1: The spatial distribution of H5N1 outbreaks from December 2003 to March 2009

of avian influenza in the latter three countries, a relatively larger resolution is employed. By this approach, the spatial pattern of H5N1 can be better identified over multiple scales in these different areas.

### 5.3 Spatial Statistics Methods

The process of influenza outbreaks and spreads demonstrates a close association with properties typically studied in geography (Cliff and Haggett, 1986). The spatial process of avian influenza is complex and the patterns of outbreaks may vary with geographical scales. To determine the spatial patterns of avian influenza, this section first employs a  $K$  function and local  $K$  function to study the spatial dynamic of avian influenza, particularly its global trends and local clusters. In addition, a modified local  $K$  function is then proposed and used to evaluate the possibility of H5N1 cluster occurrence for a succession of map lattices.

#### 5.3.1 $K$ Function

In spatial analysis, a spatial point pattern is defined as a set of locations which are irregularly distributed within a designated region (Diggle, 1983). A wide range of phenomena may be represented through spatial point patterns, including, for example trees in forest, traffic accidents on highways and influenza outbreaks in cities. When all the events of a realization are recorded, a point pattern is mapped. Mapped point patterns can then be separated as random, cluster, and dispersed, although change of patterns are continuous. Figure 5.2 shows a series of spatial point patterns.

Informally defined, randomness, which is usually called spatial complete randomness (CSR), represents an idealized realization where the intensity of events remains invariant over the study plane and shows

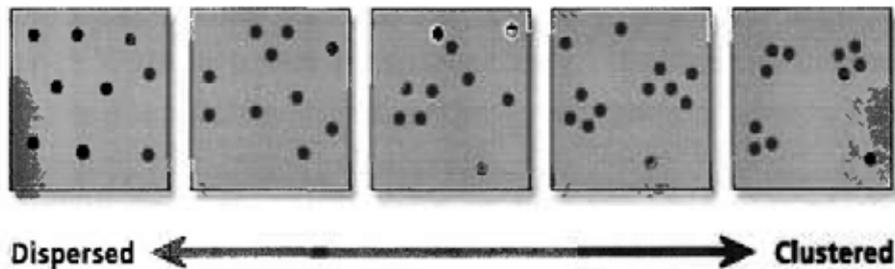


Figure 5.2: Spatial point patterns

absolutely no (spatial) interactions between events. A mathematical manifestation of CSR is described by Diggle (1983). Importantly, CSR provides a baseline against which to distinguish other patterns as they may be broadly categorized as “cluster” and “dispersed”. For a cluster pattern, the average distance between an event and its nearest-neighbor event is smaller than that in CSR. Meanwhile, for a dispersed pattern, the average distance is larger than would be expected under CSR.

However, the analysis of spatial point patterns also falls foul of the modifiable areal unit problem (MAUP) (Openshaw, 1984), a source of statistical bias that can significantly affect the results of statistical analysis. In the study of ecology, for instance, the nature of the pattern generated by a biological process can be affected by the physical scale at which the process is observed (Diggle, 1983). Specifically, at a large scale, most natural environments appear to be spatially heterogeneous and in aggregate terms would seem to be characterized by many different kinds of vegetation. At a small scale, this environmental variation is not so pronounced. A pattern of vegetation may then exhibit a more dispersed pattern.

The properties of spatial point patterns have been intensively studied over many decades. A broad range of statistical methods have also been developed for point pattern analysis, including quadrant analysis,

nearest-neighbor distance, kernel density estimation, and the  $K$  function (Cliff and Ord, 1980; Diggle, 1983; Cressie, 1991; Schabenberger and Gotway, 2005). The  $K$  function offers a way to characterize the patterns as they emerge across multiple spatial scales. This measure was first proposed as a concept by Bartlett (1964) and defined mathematically by Ripley (1976) as follows.

$$K(h) = \lambda^{-1} E(\text{number of events within distance } h \text{ of an arbitrary event}). \quad (5.1)$$

where  $E(\cdot)$  is the expectation of events within distance  $h$ ,  $h \geq 0$  and  $\lambda$  is the intensity of a point process.

In experimental terms, when a point pattern is CSR, the theoretical value of  $K(h)$  should be  $\pi h^2$ . Supposing that  $N$  events fall within a study area  $A$ , the intensity of these events can be estimated by  $\hat{\lambda} = N/A$ . The estimate of  $K$  function is then as follows.

$$\hat{K}(h) = \frac{A}{N^2} \sum_i \sum_{j \neq i} I(d_{ij} < h), \quad (5.2)$$

where  $d_{ij}$  is the distance between events  $i$  and  $j$ .  $I(d_{ij})$  is an indication function as below.

$$I(d_{ij}) = \begin{cases} 1 & \text{if } d_{ij} \leq h \\ 0 & \text{others} \end{cases}$$

When  $\hat{K}(h)$  is smaller than  $\pi h^2$ , it indicates a dispersal point pattern. On the other hand, when  $\hat{K}(h)$  is larger than  $\pi h^2$ , it implies a tendency of cluster. For simplicity, Besag (1977) suggested a linear form of the  $K$  function taking the form of a straight line proceeding upwards from

CSR. i.e., from a dispersed pattern showing no clustering or density of incidence. This linear form has been called the  $L$  function.

Edge effects may arise when events outside the study region  $A$  interact with events that can be observed within  $A$ . Spatial geographers have therefore devoted efforts to developing a more sophisticated  $K$  function taking into account these edge effects and thus seeking to mitigate the arbitrariness of particular of points or delimitation of areas (Cressie, 1991; Yamada and Rogerson, 2003). The value of this  $K$  function can be estimated by

$$\hat{K}(h) = \frac{A \sum_i \sum_{j \neq i} I(d_{ij} < h)}{N^2 c_{ij}}. \quad (5.3)$$

where  $c_{ij}$  is a factor of edge effect for a pair of events  $i$  and  $j$ . It can be calculated in terms of taking a proportion of a circumference of a circle with a concentration at event  $i$  which passes through event  $j$  (Cressie, 1991).

This modified function has been widely applied in the current studies of ecology (Haase, 1995; Barot et al., 1999; He and Duncan, 2000), spatial epidemiology (Pfeiffer et al., 2008; Lai et al., 2009), traffic accidents (Jones et al., 1996; Okabe and Yamada, 2001), wild animal behaviors (Bailey and Gatrell, 1995), and even testing good-of-fit for model selections (Diggle, 1983; Cressie, 1991). H5N1 outbreaks from late 2003 to early 2009 can also be regarded in terms of a spatial point process. In this study, we will apply the  $K$  function to determine the global trend of the disease by making a comparison between outbreaks and CSR. In order to arrive at a more explicit estimation, 1,000 Monte Carlo simulation runs of CSR are carried out to examine the statistical significant of H5N1 outbreak patterns as they have been detected.

### 5.3.2 Local $K$ Function

The overall structure of point patterns reflects a general average trend of all possible spatial relationships. Another important characteristic of a point pattern is its local structure (Anselin, 1995). The analysis of local pattern studies deviations from the mean of overall trends by seeking out portions of an area that vary spatially and offer some kind of exception to randomness. Study of local pattern effects originated from work carried out on the detection of rare disease clustering in Besag and Newell (1991). As an increasing number of georeferenced data sets become available for complex spatial studies, an increasing number of researchers have taken an interest in exploring local patterns based on 'local statistics' (Getis and Ord, 1992; Leung and Mei, 2003), for example,  $G_i$  and  $G_i^*$  statistics (Ord and Getis, 1995) and Anselin's *LISA* (Anselin, 1995) which include a local form of Moran's  $I$  (Moran, 1950) and Geary's  $C$  (Geary, 1954).

These statistics developed for the study of continuous space, however, might not be appropriate to the study of avian-influenza outbreaks. The local  $K$  function, a local statistical measure supposing spatial point patterns, may provide a better approach to understanding the spatial distribution over relatively small scales of H5N1 outbreaks. This statistic was first proposed by Getis (1984), and defined in Getis and Franklin (1987) as follow.

$$K_i(h) = E(\text{the number of events within distance } h \text{ of an arbitrary event } i), \quad (5.4)$$

Mathematically, the function can be estimated by

$$\hat{K}_i(h) = \frac{|A| \sum_j I(d_{ij} < h)}{N c_{ij}}. \quad (5.5)$$

where  $I$  is an indicator of a number of events  $j$ . The distance between  $i$  and  $j$  must be less than  $h$ .  $c_{ij}$  is an edge correlation.

Unlike the *LISA*, the local  $K$  function can be used to evaluate the degree of clustering in a location, regardless of whether events occur or not at a specific point (Yamada and Thill, 2007). The ability to take a measure of this variable i.e. cluster has meant that the metric has been widely applied to the study of wild animal behaviors (Potvin et al., 2003), traffic incidence (Yamada and Thill, 2007), and indeed avian-influenza epidemic (Si et al., 2008).

### 5.3.3 Modified Local $K$ Function

The local  $K$  function, as an indicator of cluster, can also be used to indicate the effects caused by outbreaks at the point of a specific occurrence of disease. This effect may be understood in theory as spatially continuous and gradually decreasing with distance (Cliff and Haggett, 1986). However, as a spatial statistic of point patterns, the local  $K$  function fails to estimate continuous spatial effects. Motivated by this problem, this work proposes a modified local  $K$  function weighing the spatial effects of disease in causative terms across a continuous space by distance.

The modified local  $K$  function sets out to analyze the spatial effect of outbreak clusters as they may be conceived and represented on lattices. Outbreaks, represented by point data (see Figure 5.1), are thus assigned to a lattice form. In each grid, analysis records the number of outbreaks, including in humans and birds. For each lattice, then, the

effect of surrounding outbreaks can be evaluated by

$$K_i(h) = E(\text{the weighted number of outbreaks within distance } h \text{ of a lattice } i), \quad (5.6)$$

Mathematically, the modified function can be estimated by

$$\hat{K}_i(h) = \frac{\sum_{j=1}^n w_{ij} I(d_{ij} < h)}{\lambda c_i} \quad (5.7)$$

where  $\lambda$  is the global density of outbreaks and  $d_{ij}$  is a distance between lattices  $i$  and  $j$ .  $I$  is an indicator function, giving the number of outbreaks surrounding a lattice  $i$  within the distance of  $h$ . Unlike traditional local  $K$  function,  $I$  is weighted by the inverse of Euclidean distance, i.e.  $w_{ij} = 1/d_{ij}$ .

Intuitively, Figure 5.3 gives an example of a distribution of outbreaks in lattices. The total number of outbreaks is 11 for two different circles in the plot. The estimate obtained from the local  $K$  function is  $\hat{K}_i(h) = \frac{|A| \sum_j I(d_{ij} < h)}{c_{ij}}$ , where  $\sum_j I(d_{ij} < h) = 11$ , the same value within both circles. However, such estimates do not fully capture the spatial effect caused by the locations of outbreaks. In this case, the modified function weighting outbreaks by distances away from a point offers a more precise estimation of clusters and their likely effects.

Through applying these spatial statistical methods, we can explore the spatial patterns of avian influenza outbreaks in a global-local context. The modified local  $K$  function can also be used to analyze other spatial point pattern analysis, such as criminology and traffic accident. Similar to this study, an event, represented as a spatial point, can be converted to lattice based on its corresponding latitude and longitude. By this approach, a risk map characterizing the probability

0	0	0	0	0	0
0	0	1	2	0	0
0	0	2	2	1	0
0	0	0	3	0	0
0	1	0	0	0	0
5	5	0	0	0	0

Figure 5.3: *Note:* the original local  $K$  function based on a count of events cannot identify the different local  $K$  function taking into account the distance between events. In this instance, the modified function is able to distinguish the two above patterns.

of the occurrence of events is thus generated over a continuous area (see Figure 5.6 and 5.7). In next section, we will run and analyze a number of statistical experiments studying the spatial distribution of avian influenza.

#### 5.4 Analysis Results and Interpretation

The results of spatial analysis indicate that the global patterns of avian-influenza outbreaks vary with the change of spatial scales. These patterns are different between the outbreaks in humans and birds. The difference suggests that the mechanisms of H5N1 outbreaks and patterns of spread are various in space and among host species. In addition, our result allow us to examine clusters of outbreaks according to a series of scales. Understanding the scaling behaviors of H5N1 in this

way should strengthen prevention at national, continental, and global levels.

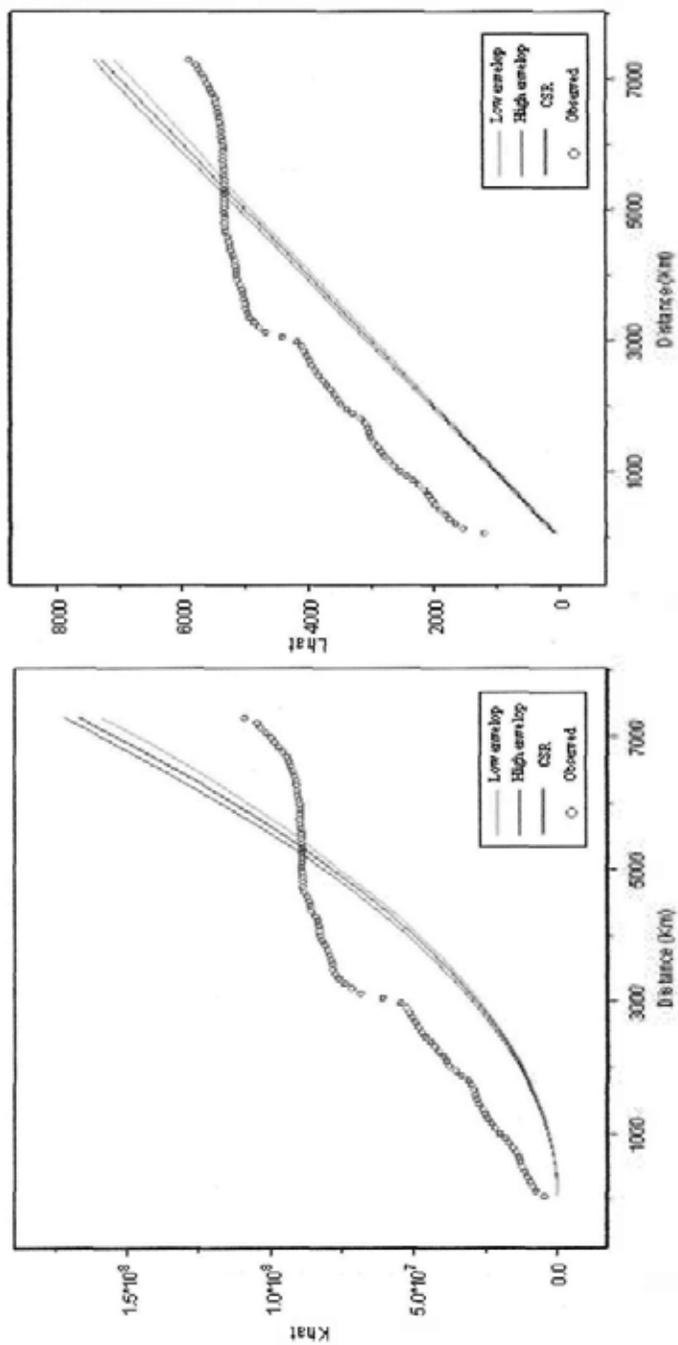
#### 5.4.1 Global Trends

Figure 5.4 shows the  $K$  and  $L$  functions of avian-influenza outbreaks of the world. Figures 5.4(a) and (b) depict the patterns of outbreaks in humans and birds. The plots indicate that these spatial patterns vary with the change of scales, with the pattern of avian influenza appearing to be more complex than in humans. This may suggest that the mechanisms driving the outbreak and spread of this disease are spatially different between the two host species.

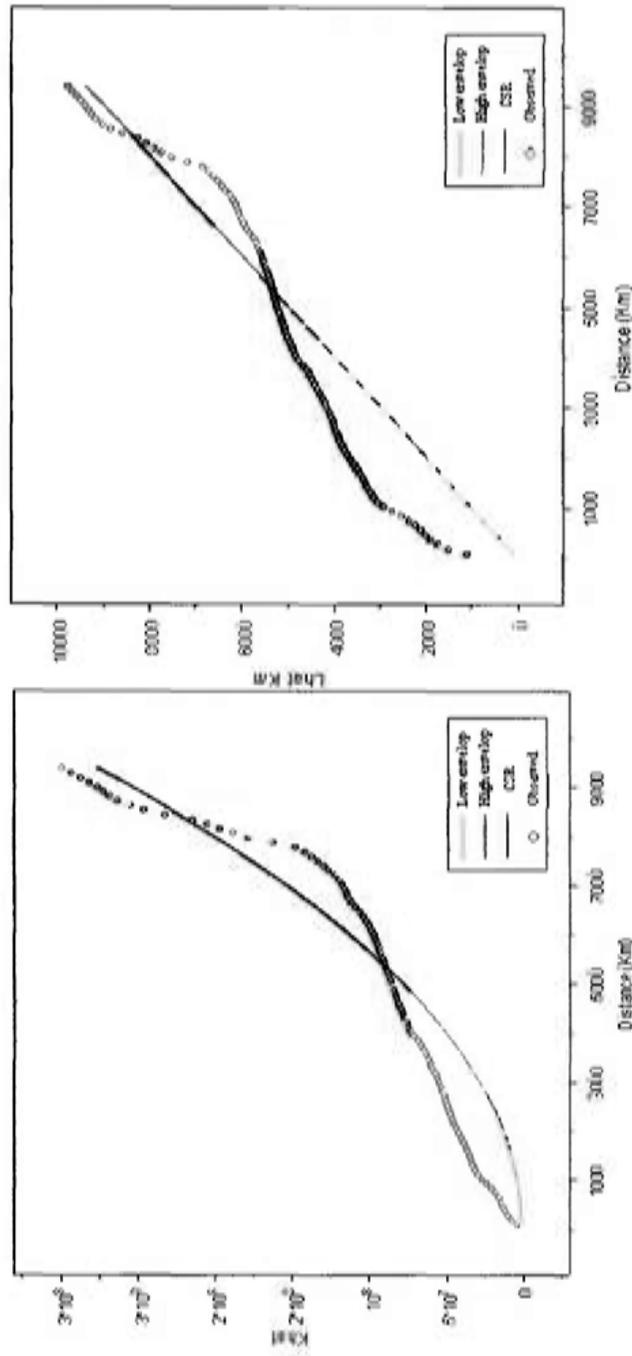
For human outbreaks, results show a significant degree of clustering in a wide range from around 500 to 5000 km, with patterns then showing a much more dispersed form thereafter (see Figure 5.4(a)). The highest degree of cluster appears at about 3500 km.

This pattern suggests that H5N1 may be transmitted across a large space. Geographically, this wide range stands for a series of spatial levels from transmission within a particular country to transmission across a continent. Humans' socio-economic behaviors may plausibly be taken as the cause of such a persistent degree of clustering in the disease. First, intensive agricultural activities and the free grazing of domestic poultry facilitate the establishment and circulation of the H5N1 virus (Chen et al., 2004; Gilbert et al., 2007). Direct contacts with sick or deadly H5N1-infected poultry have caused large outbreaks in most countries of southeast Asia (WHO, 2010b).

Second, long-distance movements are also responsible for clusters of infected family. Having said that, Human-to-human transmission is rare and limited (Ungchusak et al., 2005). However, about 25% have



(a)



(b)

Figure 5.4: The results of  $K$  and  $L$  functions of (a) humans outbreaks and (b) avian outbreaks.

occurred in clusters of two or more epidemiologically linked peoples (Wang et al., 2008b). In addition, intensive trading in poultry and illegal bird smuggling across large geographical areas have caused the spread of H5N1 at a continental level (Smith et al., 2006b).

These outbreaks, however, become dispersed at a scale of 5000 km and upwards. The suggestion here is that avian influenza has yet to become a global pandemic in human beings, despite the fact that outbreaks cluster at a wide range of spatial scales. For avian outbreaks, analysis of clustering by scales shows two clusters, the first starting between around 500 and 5,000 km, and the second at a scale of 8,000 km and upward. Between these two levels, outbreaks are dispersed. This interphase pattern implies a complex mechanism of outbreaks and distribution in bird flu. Specifically, these outbreaks appear to be associated with various factors including the ecological environment and human behaviors (Alexander, 2007a). Human socio-economic behaviors, again, have played an important role in leading to clusters of the disease in birds at both community and country levels (Gilbert et al., 2006a; Wang et al., 2008a).

Migratory birds have been regarded as a vehicle leading to the global spread of H5N1 (Normile, 2006a; Olsen et al., 2006; Si et al., 2009). The clustering behavior of the disease at large scales of 8,000 km and above can be interpreted as a support to the claim that avian influenza has crossed continental barriers and become a global pandemic in poultry. On the other hand, the varying dispersal of outbreaks at the continental level may suggest the variation of inhibiting factors among the continents that have yet to be detected in the study.

### 5.4.2 Local Clusters

In this chapter, we apply the two different local  $K$  functions to identify local pattern of outbreaks in poultry. The traditional local  $K$  function examines the degree of clusters for each location of the outbreaks. The modified local  $K$  function, on the other hand, is used to estimate the effect caused by the outbreaks in a continuous manner as they are taken to get further and further away from a location.

Figure 5.5 shows the results of traditional local  $K$  function analysis. It shows clusters of avian H5N1 outbreaks at a number of scales, including 30 km, 120 km, 260 km, and 330 km. At the scales for 30 and 120 km, the outbreaks showing small values of the local  $K$  estimates indicate a low degree of clustering (see subfigures 5.5(a) and (b)). The patterns seem to suggest an equal possibility of outbreaks at these small scales.

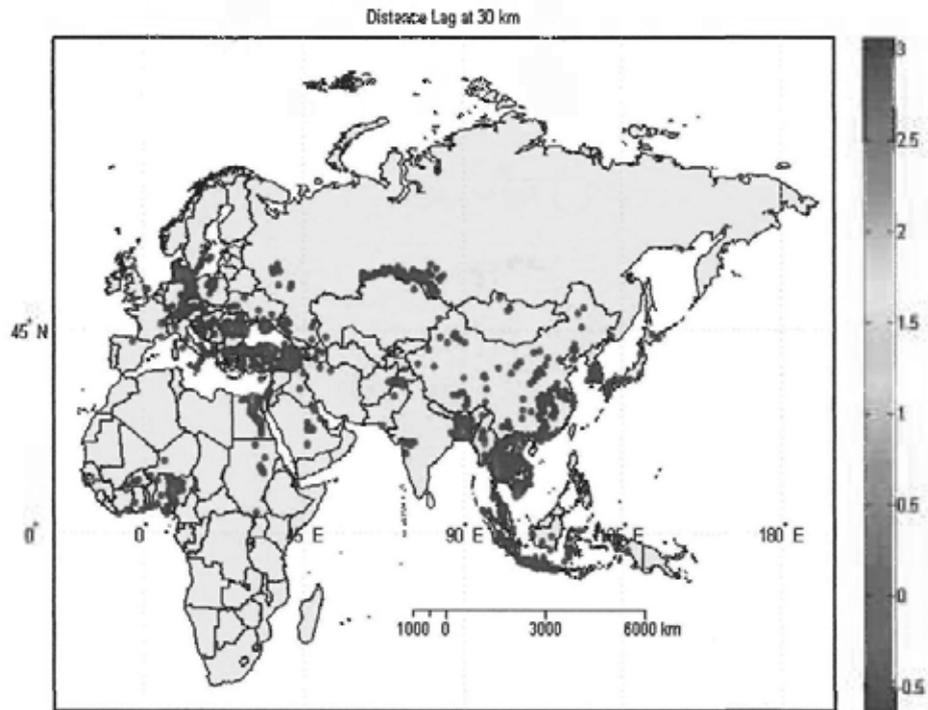
These results are explicable, however, in terms of the similarity of the socio-ecological environments across an area at a small scale. However, at large scales (260 and 330 km), clusters can be observed in Thailand and Vietnam, southeast Asia, and Egypt, Africa (see Figures 5.5(c) and (d)). Since late 2003, the widespread HPAI H5N1 has caused intensive outbreaks among poultry in Thailand and Vietnam (WHO, 2010b). This virus has established itself in Thailand (Lipatov et al., 2005; Tiensin et al., 2007; Suwannakarn et al., 2009) and also evolved into multiple sublineages in Vietnam (Nguyen et al., 2008). The first H5N1 infection was reported in Egypt, February 2006 (WHO, 2010b). Since then, this disease has caused large outbreaks in poultry and backyard flocks, resulting in dramatic economic loss in North African countries (WorldBank, 2009). Although rigorous control measures are in place, the transmission of avian influenza (H5N1) has been continuous

(Kandeel et al., 2010).

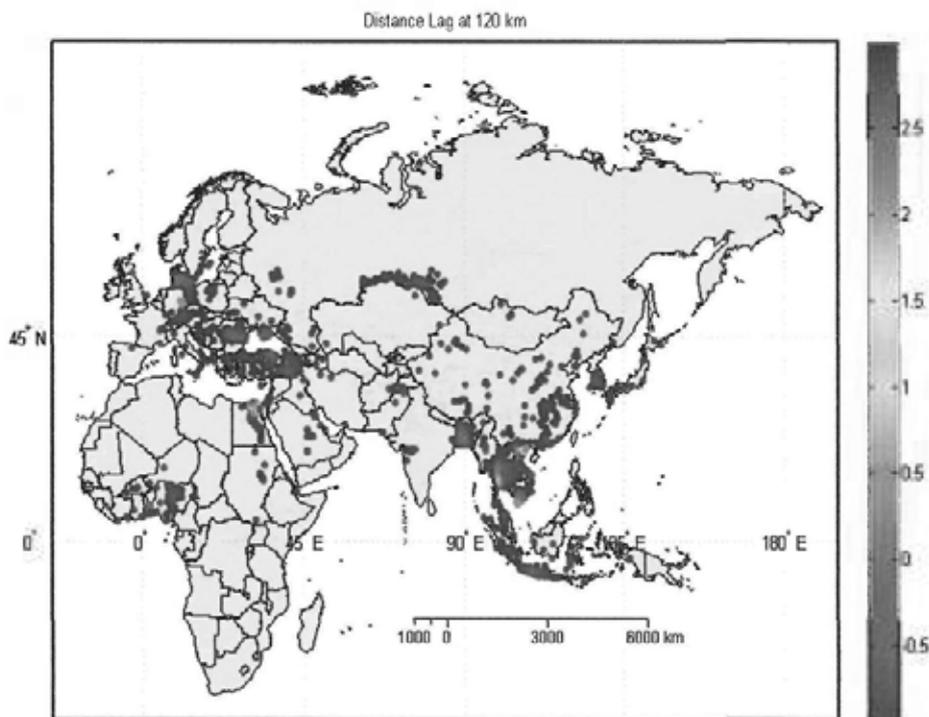
In this chapter, the modified local  $K$  function facilitates analysis of what it is taken to be the H5N1 virus's spatially continuous spread. The scale  $h = 113$  (see the Equation 5.7) is selected as showing a typical local spread of influenza viruses in Viboud et al. (2006). Each estimate of the modified local  $K$  function indicates a degree of a clustering in an outbreak. By this means, we can evaluate the effect caused by outbreaks for lattices strung across East-Southeast Asia, Thailand, Vietnam, Indonesia, and China, respectively. These values via normalization can be used as a probability for predicting the occurrence of H5N1 at a location. That is, a lattice with a high value approaching to 1 implies the large likelihood of a location having the disease.

Figure 5.6 is the normalized result for investigation of a modified local  $K$  function of East-Southeast Asia. It indicates that intensive outbreaks aggregate in Thailand, Vietnam, and Java island in Indonesia. Specifically, two large clusters can be observed in the Red River delta and a low-level plain of the Mekong delta in northern and southern Vietnam, respectively. Further, one cluster lies in the flat Mekong delta of central Thailand. In contrast to the results of traditional analysis, the outbreak effects gradually decreasing with distance away from the three clusters have been detected as operating across continuous space. Interestingly, China, regarded as one of epicenters, shows a low degree of clustering for outbreaks in the wider East and southeastern Asian region.

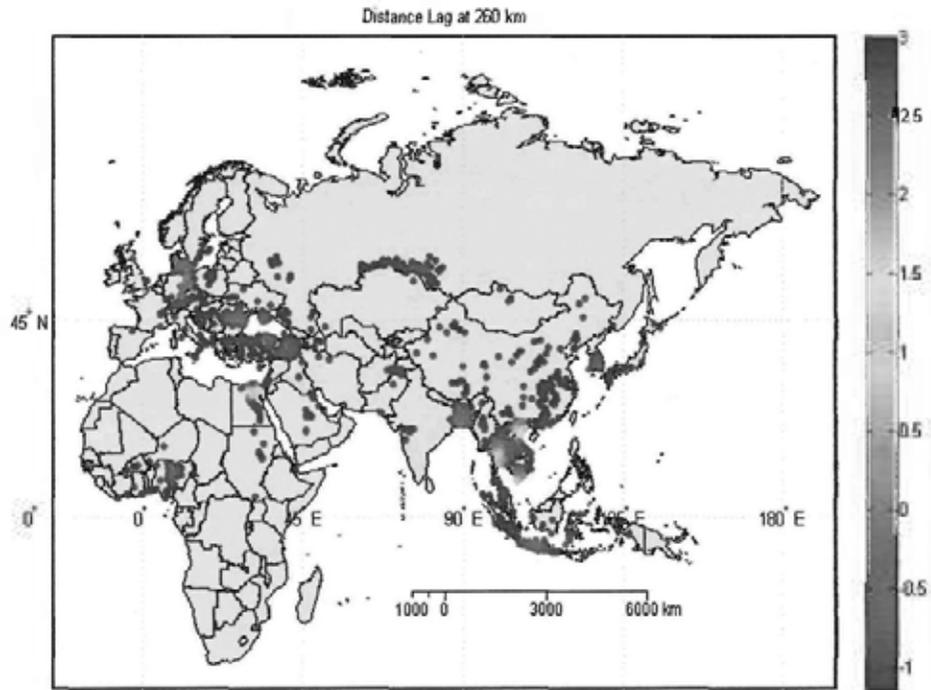
For purposes of elaboration, a modified local  $K$  function is employed to show clustering behavior for the individual countries of Thailand, Vietnam, Indonesia, and China, respectively. These countries are different in their resolutions as discussed in section 5.2. Figure 5.7 shows



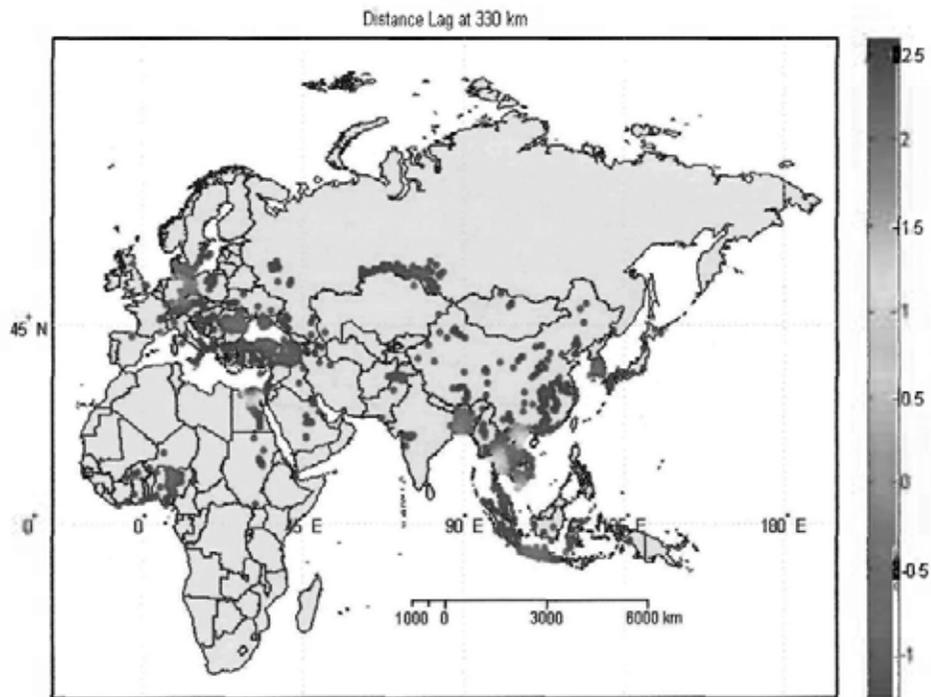
(a)



(b)



(c)



(d)

Figure 5-5: Clusters of H5N1 outbreaks at a scale of (a) 30 km, (b) 120 km, (c) 260 km, and (d) 330 km.

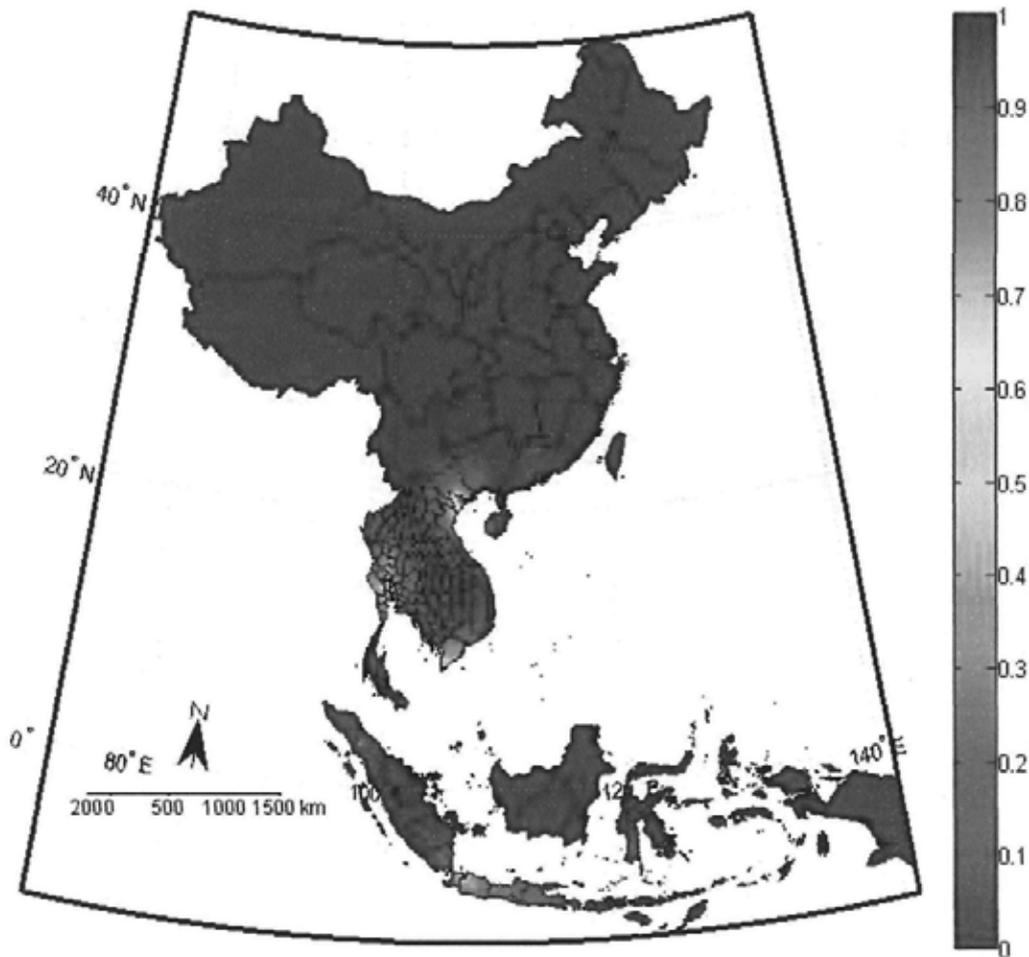
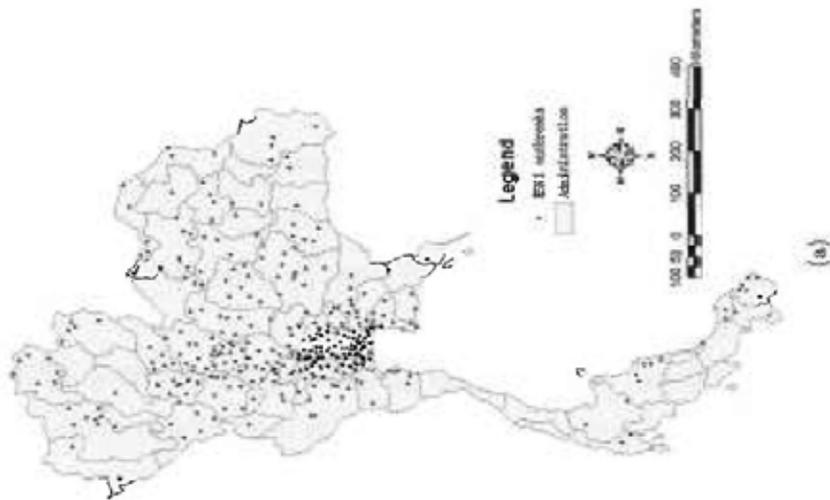
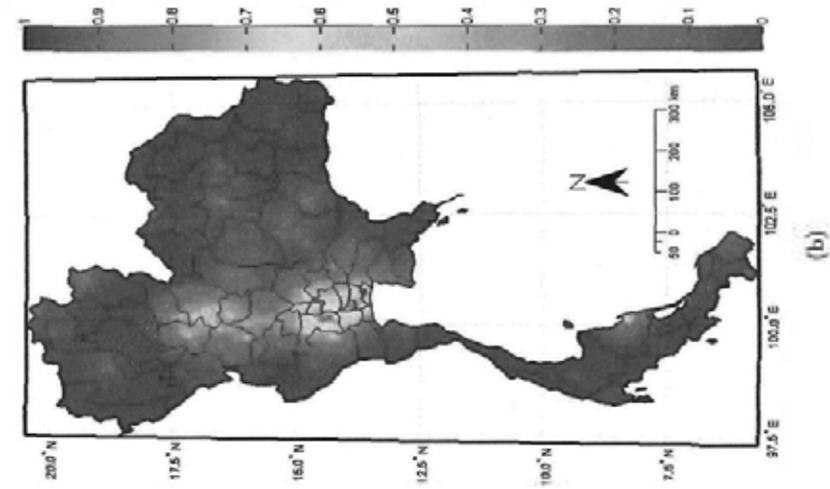
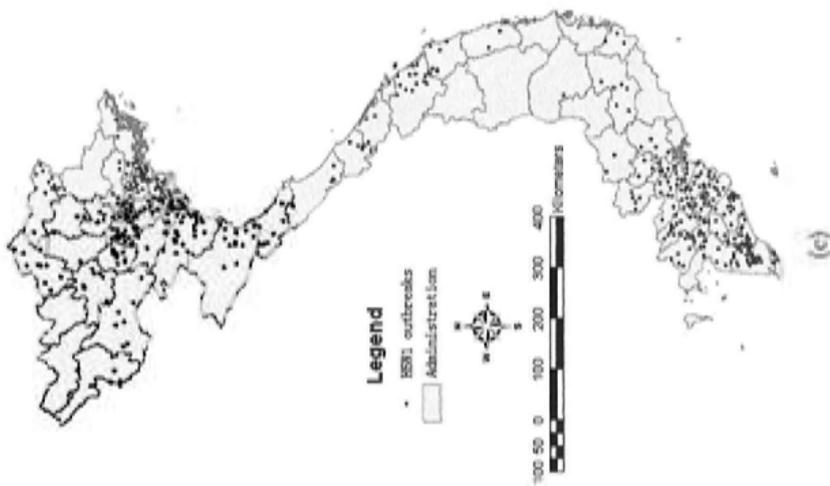
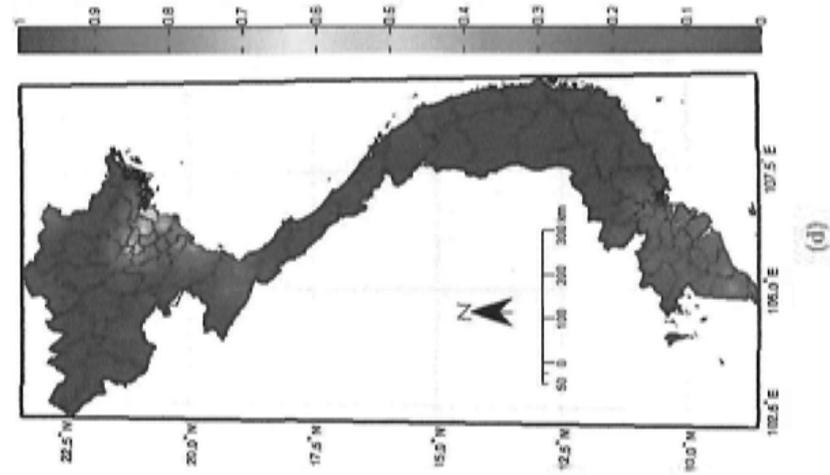


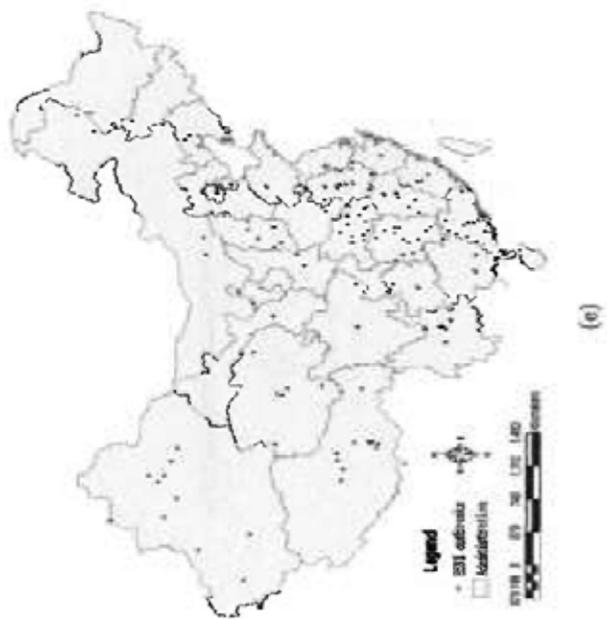
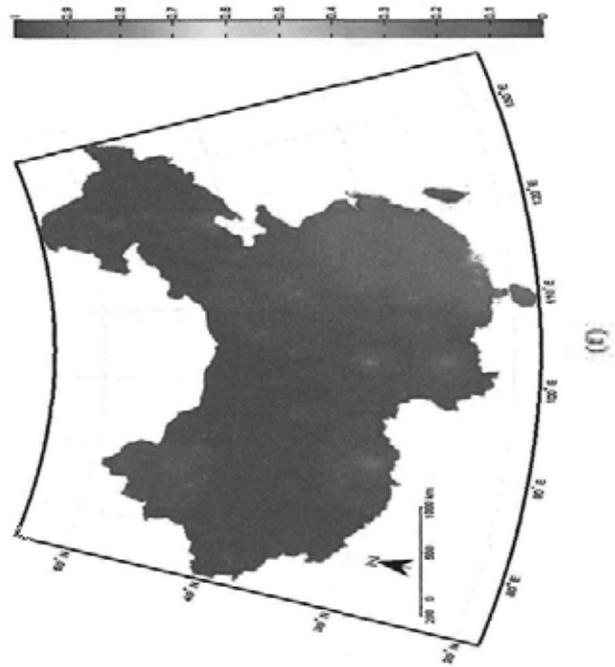
Figure 5.6: Result of modified local  $K$  function analysis of H5N1 outbreaks in East-Southeast Asia.

the normalized results of the modified local  $K$  function for the outbreaks of the four countries.

Specifically, these patterns support interpretations asserting differences between outbreaks in the four countries in terms of spatial ranges, geographical locations and socio-ecological environments, as well as the spread of H5N1 virus. In Thailand, analysis indicates a large cluster close to Bangkok (see Figure 5.7(b)). The risk posed by this clus-







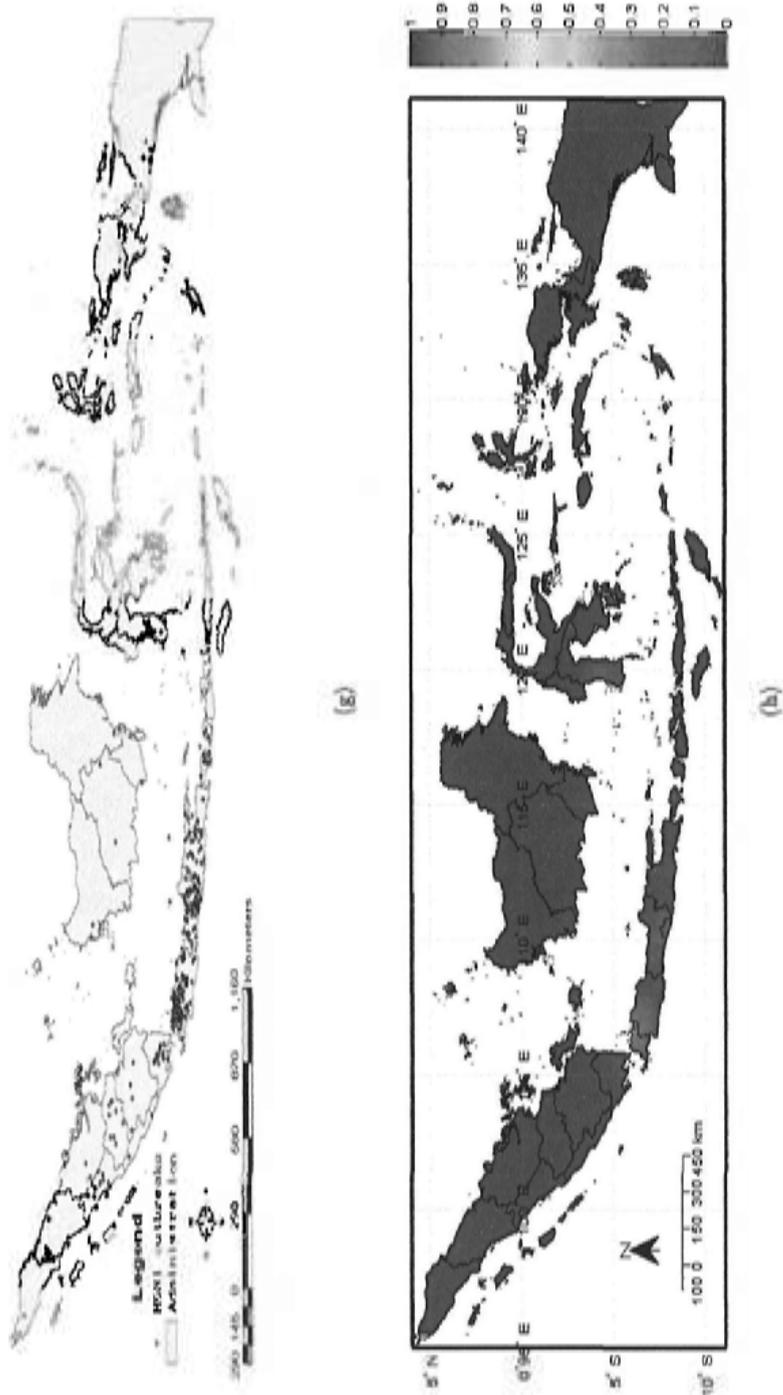


Figure 5.7: The observed H5N1 outbreaks and the results of the modified local  $K$  function: (a) and (b) Thailand, (c) and (d) Vietnam, (g) and (h) Indonesia, and (e) and (f) China

ter spatially disperses over a broad area from Nakhon Ratchasima to Phitsanulok. Such a pattern may suggest that the H5N1 virus locally persists and circulates without further spread outside the central part of Thailand. In Figure 5.7(d), the highest risk can be observed in Hanoi, northern Vietnam. This risk also gradually diminishes with distance. Separated by a low risk in the central area of the country, a high degree of risk, again, characterizes southern Vietnam, especially the area around Ca Mau City. The interphase pattern of clusters implies that multiple H5N1 sublineages have established themselves in Vietnam. Showing a different pattern from the above two countries, China shows an intensive outbreak in Guangdong province (see Figure 5.7(f)). The effect of this cluster extends across a wide area to affect provinces in the southwest and northeast, as well as the whole of southeastern China. Such a continuing pattern reflects how H5N1 is transmitted in part as a consequence of extensive human behaviors in mainland China. In addition, Indonesia, isolated by the Pacific, exhibits a different pattern again, where a high degree of clustering behaviors center on western Java (see Figure 5.7(h)). With over 60% of poultry production of the whole country (Smith et al., 2006b) and the highest density of population (more than 940 people per square kilometer)[<http://www.worldislandinfo.com>], Java has been deemed as one of epicenters of the world.

The modified local  $K$  function has extended spatial point pattern analysis from a vector to lattice data. The modified function allows for the spatial association of outbreaks over a continuous space. The analysis result, in terms of outbreak clusters, is a risk map indicating the probability of the occurrence of H5N1. In contrast to traditional analysis, the results obtained from the modified local  $K$  function are

more precise and comprehensively illustrate the distribution of H5N1 risk.

This study has also described the spatial effect of outbreak clusters. Its representation of virus's clustering behavior in terms of normalized patterns may also be thought more informative than previous work in presenting the possibility of H5N1 occurrence in a location.

## 5.5 Summary

In this chapter, we have studied the spatial patterns of avian influenza A H5N1 in humans and poultry. Through the application of  $K$  and local  $K$  functions, the chapter has identified the global trend and local clusters of the disease over multiple spatial scales. Our results indicate that H5N1 outbreaks tend to follow different patterns in humans and birds, suggesting distinct mechanisms of H5N1 outbreak for these two hosts. Clusters of outbreaks have also been detected in Thailand and Vietnam.

To enhance the power of this statistical analysis, the chapter proposed a modified local  $K$  function to estimate clusters in Thailand, Vietnam, China, Indonesia, and the whole of East and Southeastern Asia. The analysis result can be used to assess the spatial impact of the disease on a location given a certain density and pattern of incidence of outbreaks in surrounding areas at a range of spatial scales. Here clustering analysis points up different patterns for different countries and across environmental features of the whole region. The suggestion is that different transmission mechanisms for H5N1 avian influenza may come into play for different countries and different environments.

This study is also limited by the limited availability of outbreak data. With finer and more accurate location data of outbreaks, we

could further explore the spatial pattern of H5N1. The next chapter moves to deepen our understanding of the disease by considering factors associated with its occurrence. The chapter will consider the evolution of h5N1 virus and the ecological environments in which it spreads, before moving to integrate these with identification of space-time process of outbreaks and their spread in avian influenza. The study as a whole is intended to provide a firm foundation for inter-disciplinary research on relevant research fields; it should further provide valuable knowledge guiding policy decision in practice.

---

□ End of chapter.

## Chapter 6

# The Integrated Approach to Map H5N1 in Space and Time

*Science provides us with very sharp tools, but as any craftsman will tell you it is the sharp tools which can do most damage when misapplied.*

— David Harvey, 1973

Though many disciplines have made important contributions to our understanding of H5N1, it remains a challenge to integrate knowledge from different disciplines. This chapter applies genetic analysis that identifies the evolution of the H5N1 virus in space and time, epidemiological analysis that determines socio-ecological factors associated with H5N1 occurrence and statistical analysis that identifies outbreak clusters, and then applies a methodology to formally integrate the findings of the three sets of methodologies to obtain a more comprehensive examination of avian influenza occurrence in space and time. It reports findings from about the spatial pattern of the highly pathogenic avian

influenza, H5N1, risk in East-Southeast Asia where the disease is both persistent and devastating.

## 6.1 Introduction

Traditional genetic analysis, which examines phylogenetic relationships associated with the H5N1 virus from its DNA sequences, has enabled progress to be made in understanding the evolution of avian influenza viruses (Li et al., 2004; Smith et al., 2006b). By this means, possible sources and pathways that are associated with the spread of the H5N1 virus can be inferred (Smith et al., 2006b; Duan et al., 2008). Phylogeographic analysis (Avice et al., 1987) offers a method of tracking migration of the H5N1 virus. It analyzes the topology of the phylogenetic tree and use evolutionary models to statistically infer the resident localities of the H5N1 virus (Wallace et al., 2007). Medical geography examines the spatial pattern of H5N1 looking for localized hot spots where outbreaks are significantly clustered (Si et al., 2009). In spatial epidemiology, risk factor analysis focuses on the identification of factors associated with H5N1 occurrence and statistical modeling is used to predict the incidence of the disease (Gilbert et al., 2008). All these different forms of analyses have a common objective which is to understand the distribution and the spread of avian influenza. The study of avian influenza H5N1 is multi-disciplinary across virology, molecular biology, evolutionary biology, medical geography, and spatial epidemiology.

However, any study that relies on only one kind of disciplinary knowledge may miss important dimension and connections. For example, ignoring the “modifiable areal unit problem” could lead to interpretation errors when reviewing results from statistical analysis and

modeling in area-based epidemiology. The predictive mapping of H5N1 risk in China, for instance, in Fang et al. (2008), conflicts with empirical observations (WHO, 2010a; FAO, 2010) and previous studies (Smith et al., 2006a; Smallman-Raynor and Cliff, 2008). Also, phylogenetic analysis (Cavalli-Sforza and Edwards, 1967) has its own limitations when genetic data are incomplete, and evolutionary models inappropriate (Penny et al., 1992). This form of analysis, which provides a microscopic insight into the process of viral evolution, is insufficient for understanding macroscopic spread of avian influenza. In addition, uncertainty, perhaps arising from incomplete data, limited domain knowledge, or the application of an insufficiently sophisticated methodology could limit the value of these analyses. Wallace et al. (2007) states that limited sampling may lead to results that are not statistically significant.

Although current research tends to integrate multi-disciplinary studies of avian influenza, it stops at the early stage of analytically integrating data, for example, on phylogenetic relationships between isolated occurrence of the virus, migratory bird movements, and trade in poultry and wild birds (Kilpatrick et al., 2006; Liang et al., 2010) or implementing only basic statistical analysis between genetic distance and geographic distance (Carrel et al., 2010). Even though knowledge can be obtained from different studies, how to quantify and integrate this knowledge remains a challenge in the study of H5N1.

This chapter proposes a novel approach to integrating the findings of phylogenetic analysis, which unravels H5N1 evolution in space and time, with modified local  $K$  function, which, proposed in the last chapter, identifies outbreak clusters in space, and also with spatial epidemiology, which determines socio-ecological factors associated with the oc-

currence of H5N1. In this study, Dempster-Shafer theory of evidence (Dempster, 1967; Shafer, 1976), a mathematical method for making inferences based on multiple forms of evidence and which recognizes the uncertainties associated with the different sources of evidence, is used to formally integrate the three sets of findings. Finally, we apply the methodology across multiple scales; that is to the whole of East-Southeast Asia as well as the individual countries of Thailand, Vietnam, Indonesia, and China, respectively.

## 6.2 Data Description

Three kinds of data are used in this study: DNA sequences, reported H5N1 outbreak records, and socio-environmental factors, including altitude, population density, poultry density, and the shortest path distances to inland water, coastlines, migratory bird pathways, railways, and roads (see Figure 6.1). All these data were collated for lattice based on different spatial resolutions: 8.4 km<sup>2</sup>, 34.22 km<sup>2</sup>, 0.94 km<sup>2</sup>, 0.32 km<sup>2</sup>, and 0.24 km<sup>2</sup> for East-Southeast Asia, China, Indonesia, Thailand, and Vietnam, respectively.

### 6.2.1 DNA sequences

From GenBank (Benson et al., 2006), 888 DNA sequences of influenza A H5N1 hemagglutinin (HA) and neuraminidase (NA) genes were isolated from a variety of hosts between 1996 and 2009 across the areas of East-Southeast Asia covering Thailand, Vietnam, Laos, Cambodia, Indonesia, and China. Figure 6.2 shows the combination of the HA and NA sequences.

With reported geographic localities, the sequences can be assigned to lattices by geocoding. Multiple alignment of all sequences from each

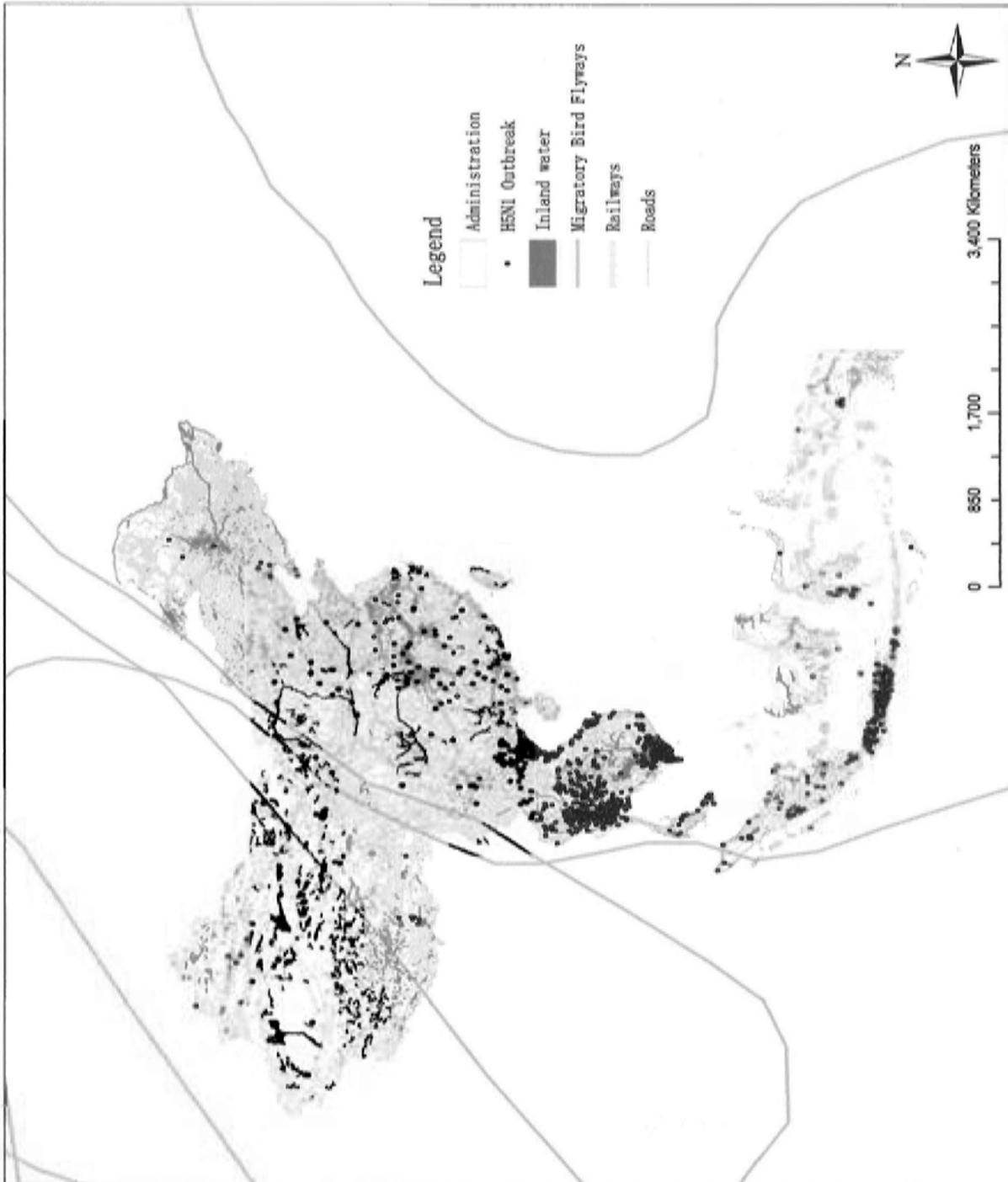


Figure 6.1: Spatial distribution of reported H5N1 outbreaks, inland water bodies, migratory bird pathways, and roads in East-Southeast Asia.



HA and NA gene was carried out using MUSCLE (Edgar, 2004), and the HA and NA sequences were combined using TaxonDNA (Meier et al., 2006). A phylogenetic tree for the combined dataset was constructed using neighbor-joining (NJ) (Saitou and Nei, 1987) in PAUP\* 4.0 (Swofford, 2002). The best-fit model of DNA substitution for the NJ analysis was assessed by Model test version 3.7 (Posada and Crandall, 1998). Base composition and pairwise comparisons were examined using MEGA version 4.02 (Tamura et al., 2007). The NJ algorithm is adopted in this study because this distance-based method of phylogenetic reconstruction gives the genetic distances among the sequences that need to be combined with the results of the other analysis in the knowledge fusion step. Figure 6.3 shows the NJ tree of the H5N1 virus in East-Southeast Asia.

### 6.2.2 Outbreak Data

Outbreaks of the disease are assigned to lattice points. Data on avian influenza H5N1 outbreaks in East-Southeast Asia include 2204 avian and 327 human H5N1-confirmed cases between May 1<sup>st</sup> 1997 to March 16<sup>th</sup> 2009, compiled by the World Organisation of Animal Health (OIE, [www.oie.int](http://www.oie.int)) and the World Health Organization (WHO, [www.who.int](http://www.who.int)), respectively. Each record contains the following attributes: country, province, location (latitude, longitude), start time, affected species, and the number of deaths. Using latitude and longitude, any outbreak can be assigned to the lattice. For each lattice point, the numbers of outbreaks of both human and avian influenza are recorded.

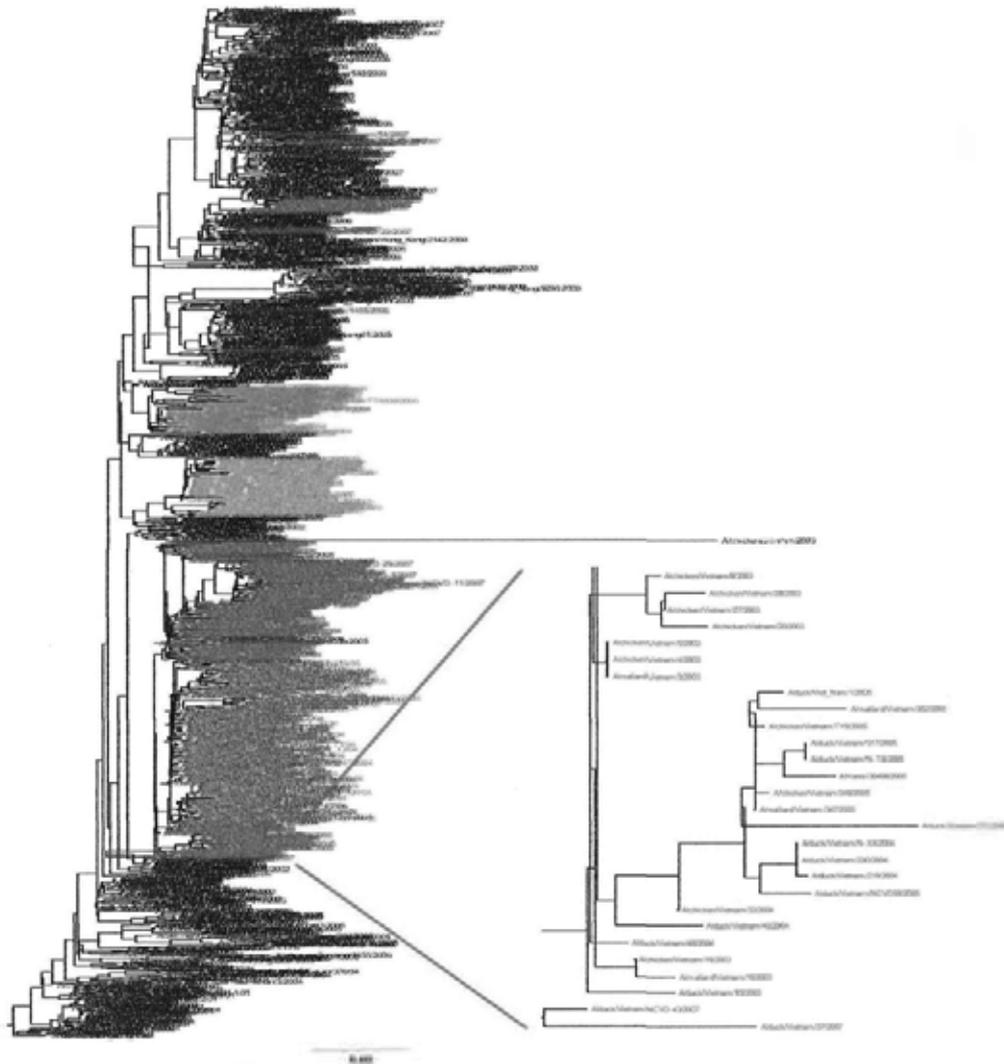


Figure 6.3: NJ tree of the 888 H5N1 concatenation of hemagglutinin (HA) and neuraminidase (NA) DNA sequences from East-Southeast Asia covering Thailand, Vietnam, Cambodia, Laos, Indonesia, and China. The best model is TVM+I+G (transversional model incorporating invariable sites and rate variation among sites). The goose H5N1 DNA sequence from Guangdong in 1996 (A/Goose/Guangdong/1/96) is used to root the tree. The length of a unit is 0.02. The taxa are colored by locality with green for Thailand, blue for Vietnam, orange for Indonesia, and black for all others including China, Cambodia, Laos, and purple, especially, for Qinghai province, western China.

### 6.2.3 Socio-environmental Data

The socio-environmental data were compiled for the raster grids based on the spatial resolutions of East-Southeast Asia, Indonesia and China, respectively. Poultry census data in 2005 are obtained from Food and Agriculture Organization's Animal Production and Health Division (FAO-AGA).

These poultry data, collected from sub-national livestock census data and corresponding to administrative boundaries, have been converted into densities by excluding the areas unsuitable for livestock. The poultry density data were downloaded from GeoNetwork ([www.fao.org/geonetwork/srv/en](http://www.fao.org/geonetwork/srv/en)), and collated to raster grids with each pixel value representing actual density per km<sup>2</sup>. Population density was chosen as an indicator of the volume of viral traffic (Gilbert et al., 2008). The 2010 estimate of population density, produced by the Center for International Earth Science Information Network (CIESIN), Columbia University and the United Nations Food and Agriculture Programme (FAO), were downloaded from the CIESIN ([www.ciesin.org](http://www.ciesin.org)). Railways and roads were chosen because these two variables were used as surrogate indicator of long-range movements of human and poultry. The shape data for the two transport networks were provided by GIVA-GIS ([www.diva-gis.org](http://www.diva-gis.org)), an open source for mapping and geographic data. The area and line-based data, including inland water bodies, coastline, and migratory bird pathways, downloaded from the GIVA-GIS, were used to determine the association between birds and the outbreaks. The migratory bird pathways were specified by 70 km buffers on each side because of uncertainty about the behavior of migrating birds. All shapes were converted to binary lattice data indicating presence or absence. In addition average elevation data [90\_m resolution

Digital Terrain Model from the Shuttle Radar Topography Mission data, STRM V3 (<http://srtm.csi.cgiar.org>)] were used to capture topographic features that might be associated with the establishment of an H5N1 epidemic (Gilbert et al., 2008).

### 6.3 The Quantitative Analysis

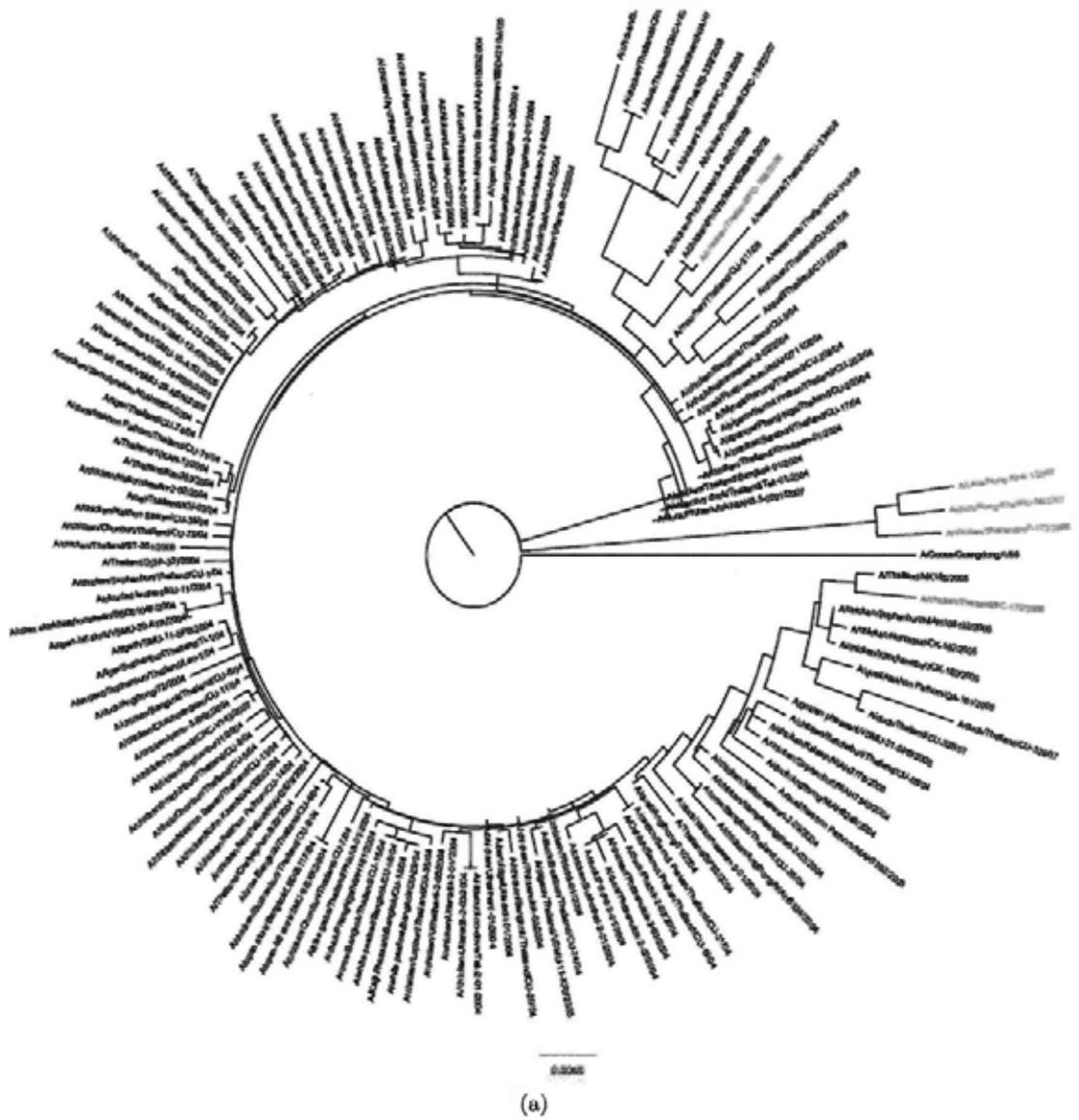
Three analyses are implemented on raster data frames. The probability of the occurrence of HPAI H5N1 is estimated for each lattice point (or small pixel). Specifically, this study consists of four parts. First, phylogenetic trees are built for the evolution of the H5N1 virus. The branches and the topology of a tree describe the processes of viral evolution. The ability of a virus to survive in nature (its “capability”) is a characteristic of a virus which goes through a long evolutionary process with wide spatial dispersion and persistence over time. Strong capability may lead to a high probability of the disease spreading widely. In this study, the years and localities from which the DNA sequences were sampled are collated for each H5N1 virus. Integrating in space and time, the quantification of the phylogenetic tree is a feasible way to measure and map the capability of the H5N1 virus. Second, the local  $K$  function, a spatial point pattern statistic (Botts and Getis, 1988), has been modified in the last chapter for the purpose of identifying the local pattern of outbreaks. The estimates obtained from the analysis are an indicator of outbreak clusters. Third spatial epidemiological analysis involves building a logistic regression model for analyzing the statistical association between the presence/absence of reported H5N1 outbreaks and eight socio-environmental variables. This model can be used to predict the probability of the occurrence of an H5N1 outbreak. The findings obtained from the three analyses provide evidence with

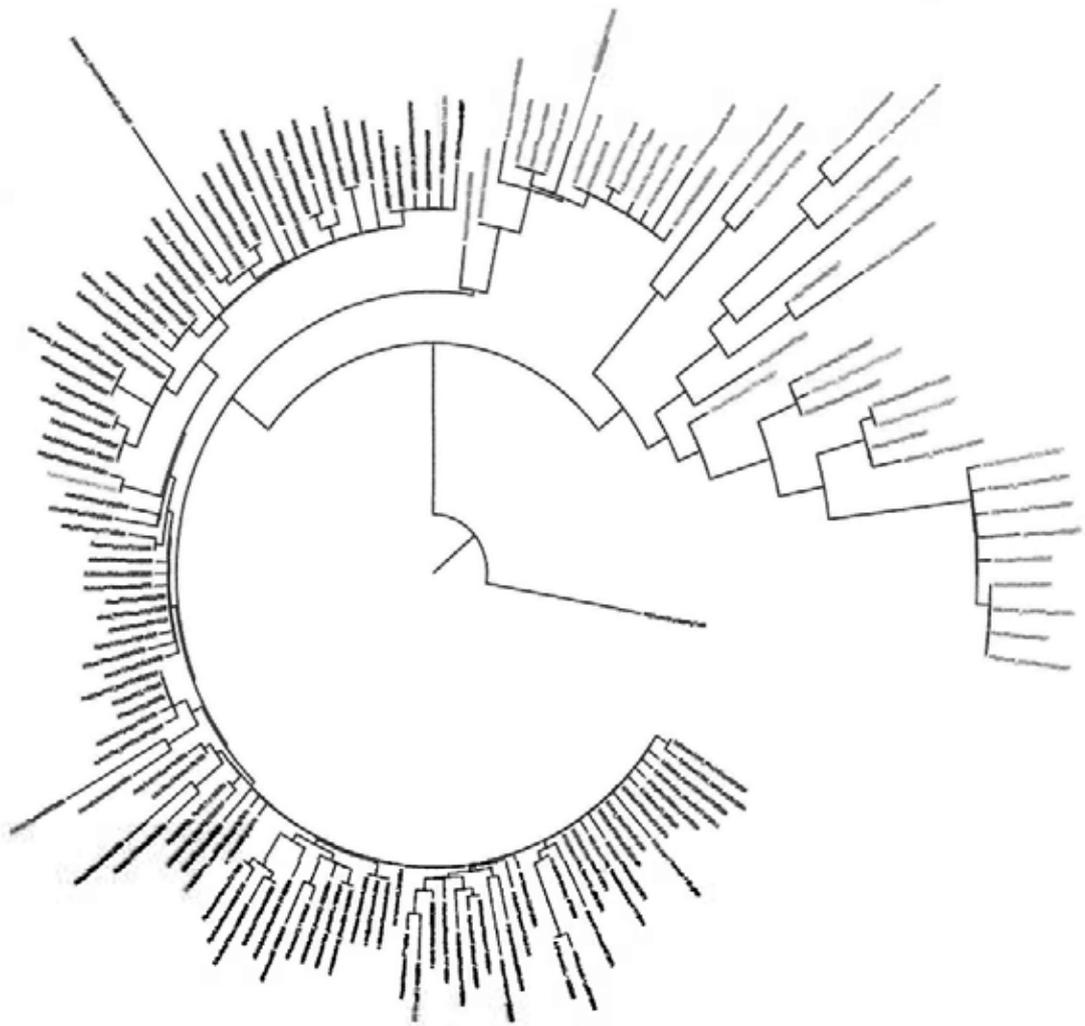
which to explore the spatial distribution of H5N1. To present a powerful, robust, and unified result, Dempster-Shafer theory of evidence is applied to integrate the different forms of evidence.

### 6.3.1 Quantifying the Phylogenetic Tree

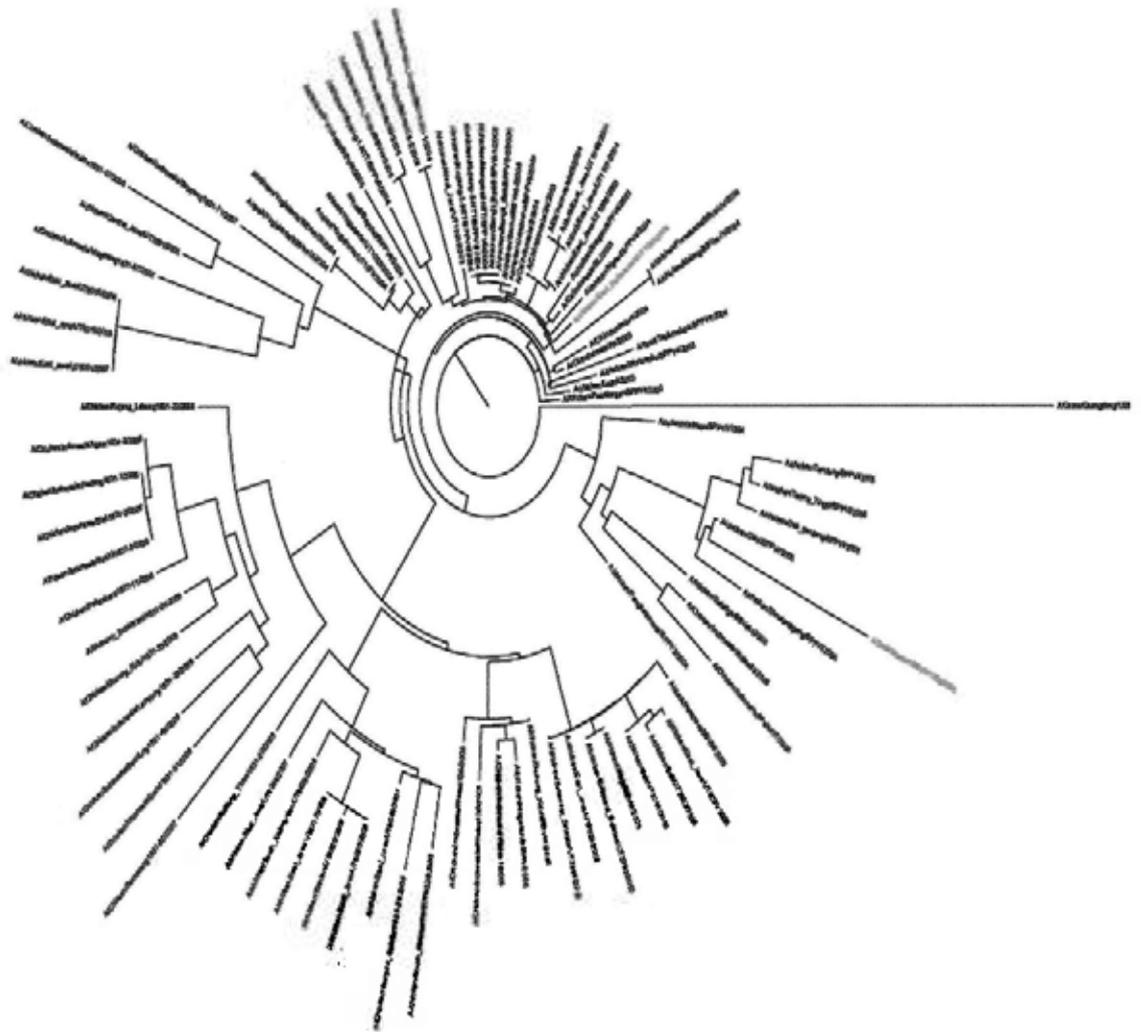
Figure 6.4 shows the NJ trees of the H5N1 virus from (a) Thailand, (b) Vietnam, (c) Indonesia, and (d) China. In the phylogenetic trees of the H5N1 virus, the process of viral evolution is composed of a set of branches, originating from a common ancestor at the root of the tree. The length of a branch describes an evolutionary stage which starts from a previous hypothesized ancestor (or a node). The taxa having a common node can be regarded as a subgroup (or a clade). By this means, a phylogenetic tree is usually divided and the taxa are grouped into different subgroups, with the members of each subgroup are phylogenetically close. In virology, subgroups are usually determined by eye balling (Li et al., 2004; Smith et al., 2006b). A subgroup is believed to have a strong capability of surviving if its members show a pattern of wide spatial dispersion and extensive persistence over time (Chen et al., 2006). This also indicates the capability of a subgroup to spread the disease.

However, for a virus, this capability can be measured by both the process of the viral evolution and the capability of the subgroups which the virus belongs to. The longer is the evolutionary process the greater is the ability of a virus to survive and the higher is the possibility of the spread of the virus. Also, a virus is believed to have a large capability if it is a member of a strong subgroup of the strains showing extensive persistence in space and time. For evaluation, it is necessary to sum up all evolutionary stages that a virus goes through. Each stage contains





(b)



(c)

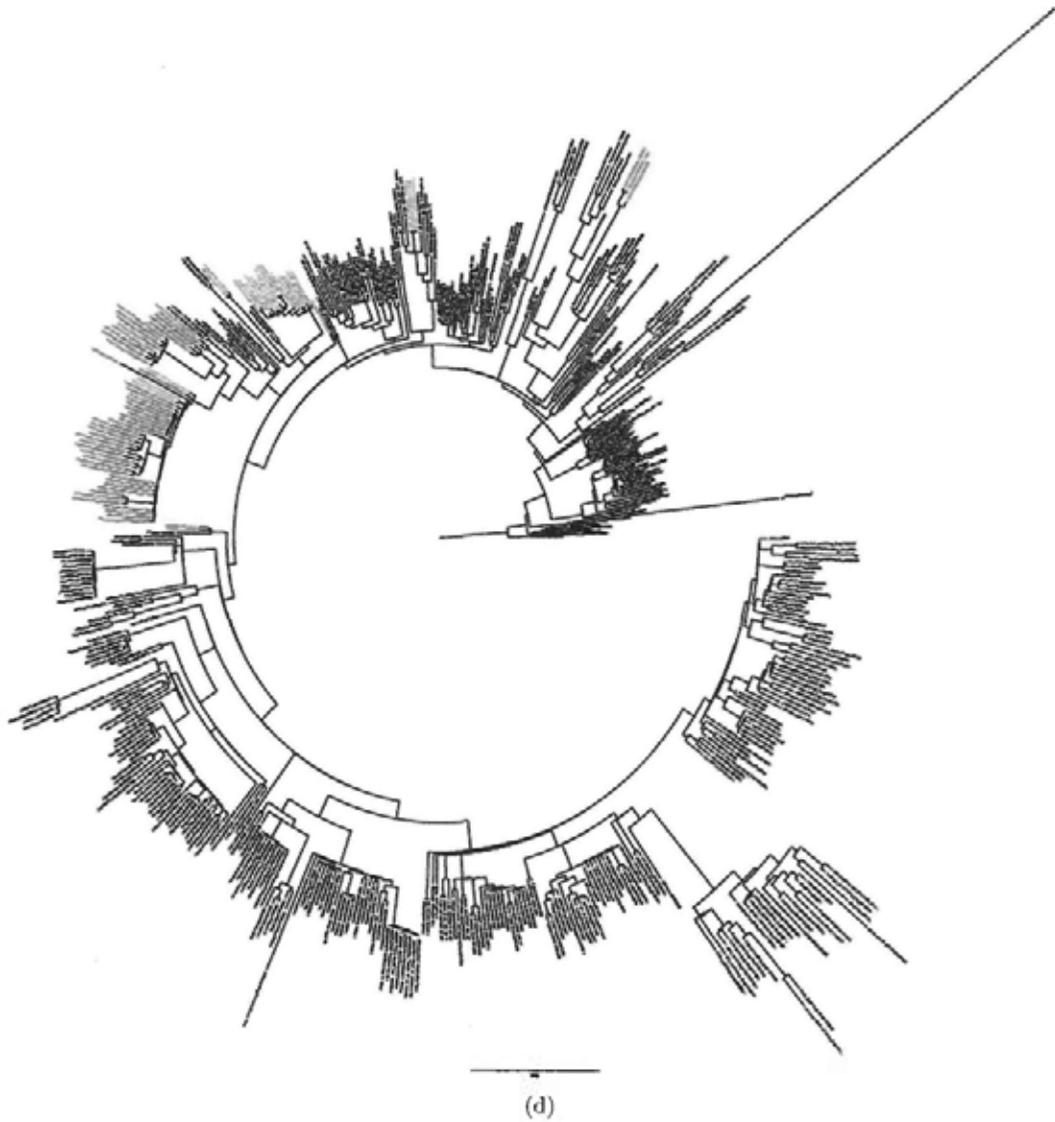


Figure 6.4: The NJ trees for the H5N1 virus from (a) Thailand, (b) Vietnam, (c) Indonesia, and (d) China. The best model is GTR+I+G (General Time Reversible incorporating invariable sites and rate variation among sites.)

a branch and a node which is shared by a subgroup of the taxa. As a first approximation,  $p_i$ , an estimate of the capability of a virus, can be measured by a linear sum,

$$p_i = w_1 e_i + w_2 s_i + w_3 t_i, \quad (6.1)$$

where  $e_i$ ,  $s_i$ , and  $t_i$  are viral evolution, spatial dispersion, and time span at a stage  $i$ .  $w_1$ ,  $w_2$ , and  $w_3$  are weights for the three variables, assigned the values 0.4, 0.4, and 0.2, respectively. The lower weight for the time span is because a temporal scale of a year is coarse relative to the other two variables. The normalized values of  $e_i$ ,  $s_i$ , and  $t_i$  can be estimated from the hierarchical structure of the tree. First, the length of the branch at a stage  $i$  is used to measure viral evolution. Second, we measure how widely dispersed, geographically, members of the same subgroup are. Spatial dispersion is estimated by the total inertia (sum of the variances) of the  $2 \times 2$  covariance matrix calculated using the locations (latitude, longitude) of the members at the same stage. Third, the length of time for a subgroup is measured by the time span from early to late occurrences of virus members. The time span  $t_i$  is a ratio, including the persistence of a subgroup at stage  $i$ . It is calculated as the time length  $i$  divided by the time length of the whole tree. Estimates of  $e$  and  $s$  are normalized to avoid the effect of large values. Mathematically, the capability of a virus can be evaluated by

$$\sigma = \sum_{i=1}^n p_i. \quad (6.2)$$

where  $n$  is the number of evolutionary stages of the virus. Quantifying a phylogenetic tree enables us to estimate the capability of a virus.

To determine the range of a virus, we also need to examine the

density of surrounding outbreaks at a series of scales from 10 to 250 km. Within this range, a scale  $l$  at which outbreak density is at a maximum was selected. The area affected by a virus is estimated using a spatial Gaussian kernel function:

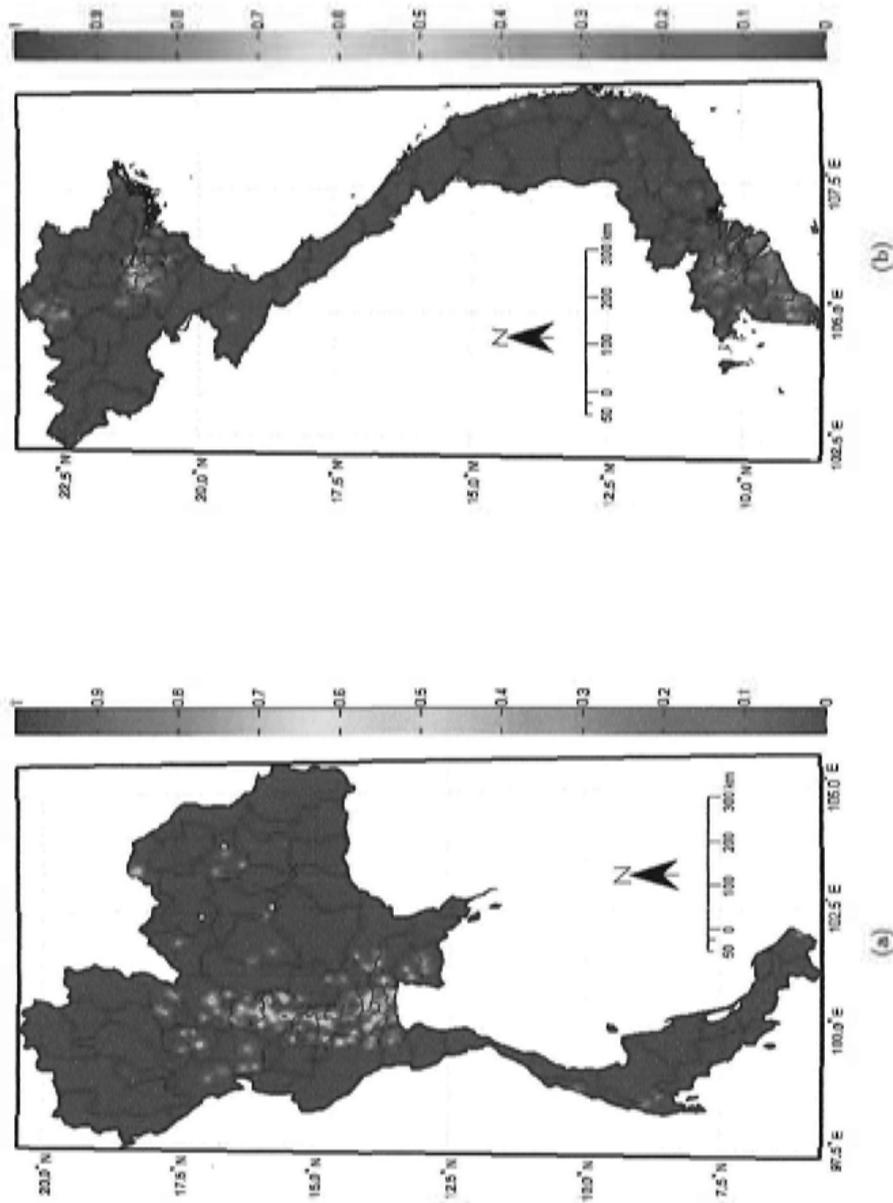
$$G(\mathbf{x}, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{\|\mathbf{x}\|^2}{2\sigma^2}}, \quad (6.3)$$

If  $\|\mathbf{x}\| > l$  and  $G(\mathbf{x}, \sigma) = 0$ , the range of the affected surface is  $l$ ; otherwise, if  $\|\mathbf{x}\| < l$  and  $G(\mathbf{x}, \sigma) = 0$ , the range is  $\mathbf{x}$ . Figures 6.5 show the capability of H5N1 in Thailand and Vietnam based on quantifying the phylogenetic trees of the H5N1 virus (Figures 6.4(a)) and (b)). In Figure 6.5, high value represented by red points indicates a large probability of occurrence of avian influenza in terms of the evolution capability of H5N1 analyzed above.

### 6.3.2 Risk Factor Analysis

A logistic regression was fitted to model the association between the occurrence of H5N1 outbreaks and eight socio-environmental factors in East-Southeast Asia, Indonesia, and China. Data on the predictors were compiled for each lattice point, and the outbreaks converted to 'presence of outbreaks' (1) or 'absence of outbreaks' (0) for each lattice point. Mathematically, the relationship between the occurrence of H5N1 and the predictors can be formulated by a binary logistic regression model:

$$P_i = \frac{e^{z_i}}{1 + e^{z_i}}; \quad (6.4)$$



**Figure 6.5:** Probability maps predicting the occurrence of avian influenza (H5N1) in Thailand and Vietnam. (a) and (b) show the probabilities derived from the phylogenetic trees analyses (see Figures 6.4(a)) and (b). The experimental data covers the H5N1 outbreaks from late 2003 to 2009.

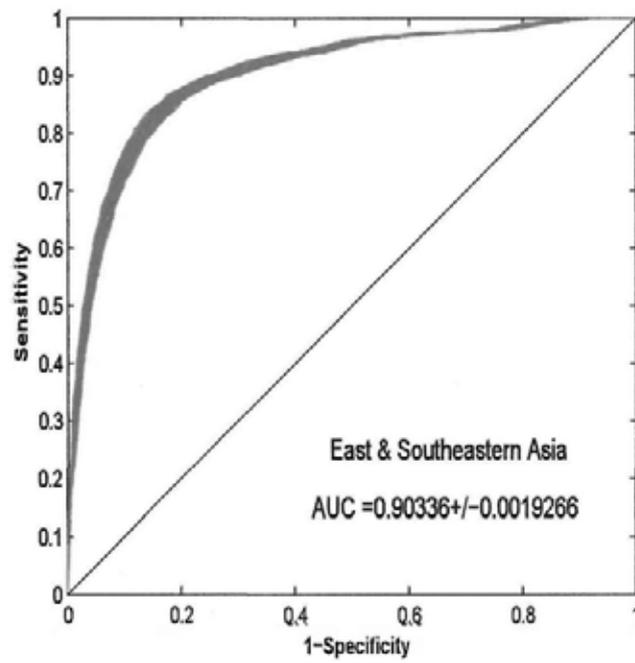
where  $P_i$  is a probability for the occurrence of the disease at location  $i$  and  $z_i$  is usually defined as:

$$z_i = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n . \quad (6.5)$$

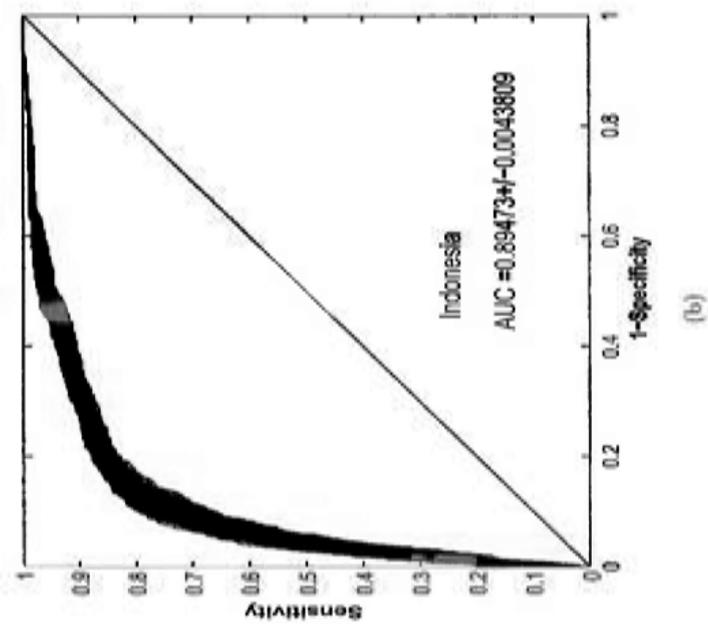
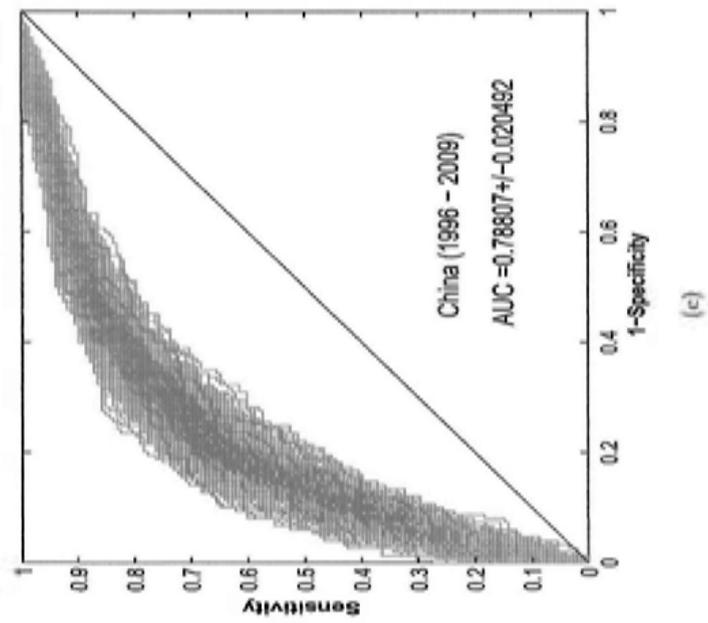
$\beta$  is correlation coefficient and  $x$  is independent variable. That is, the eight predictors in the analysis.

A very large majority of lattice points have no outbreaks which makes fitting the logistic model to all the lattice points unsatisfactory (see also Gilbert et al. (2008)). We build models by selecting all outbreak-present lattice points and randomly selecting double the number of lattices points with no outbreaks. The number of H5N1 outbreaks in East-Southeast Asia including Indonesia and China are much lower than in Thailand and Vietnam. 1000 bootstrap replicates were implemented to ensure a satisfactory sample size for carrying out model inference. In each repeat, the coefficient and  $p$  value of each predictor were estimated. Cohen's kappa index was used to evaluate the observed/model predicted misclassification matrix and Nagelkerke / Cragg & Uhler's psuedo- $R^2$  was used as the goodness-of-fit statistic for the logistic regression models (Long, 1997) . In addition, the receiver-operating characteristic ROC provides a two-dimensional depiction of predictive performance (Fawcett, 2006). The area under the ROC curve (AUC) measures the probability of a correct classification (Hanely and McNeil, 1982). Figure 6.6 shows the ROC curves for the logistic regression models of Indonesia, China, and the whole of East-Southeast Asia.

The ROC, Cohen's kappa index, and Nagelkerke / Cragg & Uhler's psuedo- $R^2$  were calculated for the purpose of assessing the predictive power of the model (see Table 6.1). All these estimates were averaged



(a)



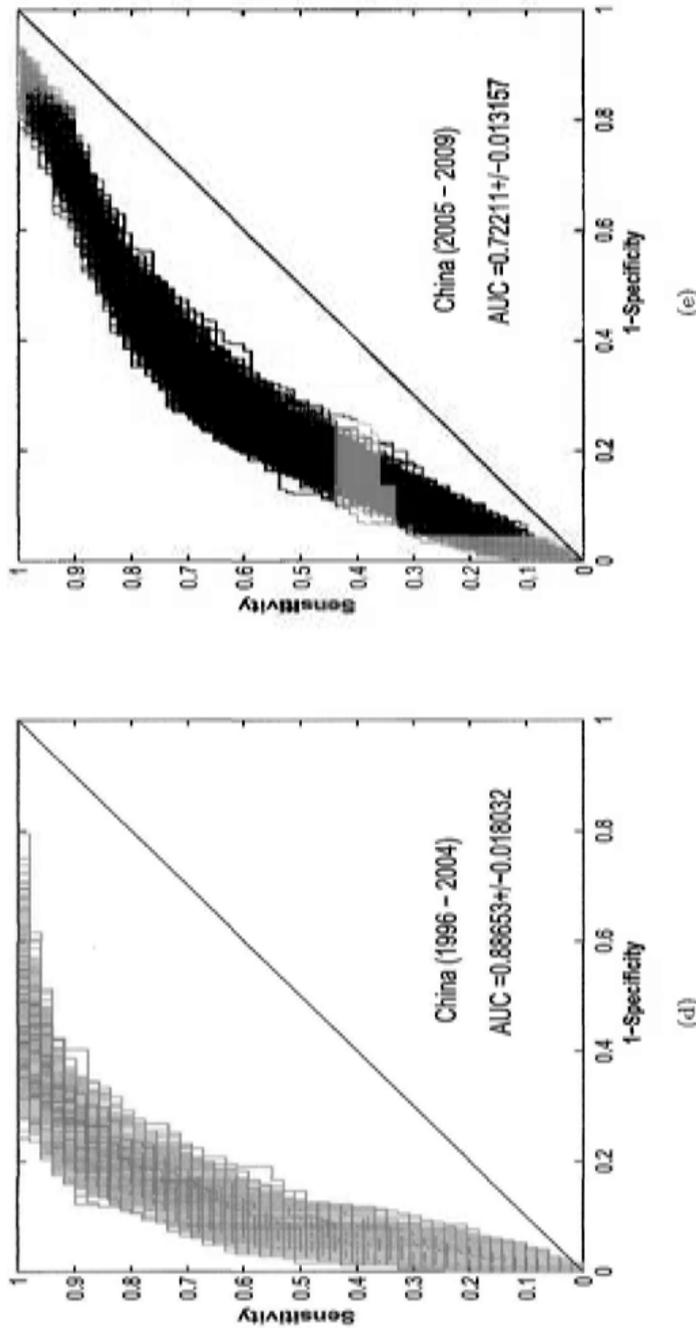


Figure 6.6: ROC curves for the logistic regression model. The area under the ROC curve (AUC) indicates the probability of a correct classification. (a) and (b) show the average AUC of the models for outbreaks of East-Southeast Asia and Indonesia from 1996 to 2009. (c), (d), (e) are the AUC of the models for China for the periods between 1996 and 2009, 1996 and 2004, and 2005 and 2009, respectively. The blue areas are the envelopes of the 1000 bootstrap replicates.

Table 6.1: Logistic regression model assessment for the H5N1 occurrences in East-Southeast Asia, Indonesia, and China, 1996-2009 and the two epidemic waves between 1996-2004 and 2005-2009.

Region or Country	Model Assessment		
	AUC $\pm$ SD	Kappa $\pm$ SD	Pseudo-R <sup>2</sup> $\pm$ SD
East-Southeast Asia	.9034* $\pm$ .0019	.8467 $\pm$ .0130	.6061 $\pm$ .0060
Indonesia	.8947 $\pm$ .0044	.6491 $\pm$ .0042	.4751 $\pm$ .0192
China (1996-2009),	.7806 $\pm$ .0097	.6063 $\pm$ .0408	.2354 $\pm$ .0241
China (1996-2004),	.8865 $\pm$ .0180	.6497 $\pm$ .0181	.4901 $\pm$ .0681
China (2005-2009),	.7221 $\pm$ .0132	.5796 $\pm$ .1021	.1679 $\pm$ .0269

\* Average value of 1000 logistic models

over the 1000 bootstrap replicates. The predictions for the H5N1 in Indonesia, China, and East-Southeast Asia are shown in Table 6.2 and Figures 6.7(b), (d), and (f). In these figures, the risk potential for the occurrence of H5N1, with respect to environment, is highlighted by large values in red. That is, the larger is the value approaching to 1, the higher is the probability of occurrence of the disease.

### 6.3.3 Knowledge Fusion and Dempster-Shafer Theory of Evidence

Knowledge fusion provides a framework for integrating information from different research domains. Uncertainty, perhaps arising from incomplete data, limited domain knowledge, or the application of an insufficiently sophisticated methodology, could limit the value of the above analyses. Dempster-Shafer (D-S) theory of evidence seeks to overcome the limitations associated with conventional probability theory when the researcher seeks to quantify and reason with imprecise, uncertain and/or weak information (Dempster, 1967; Shafer, 1976). It has been widely used in applications involving the use of geographical

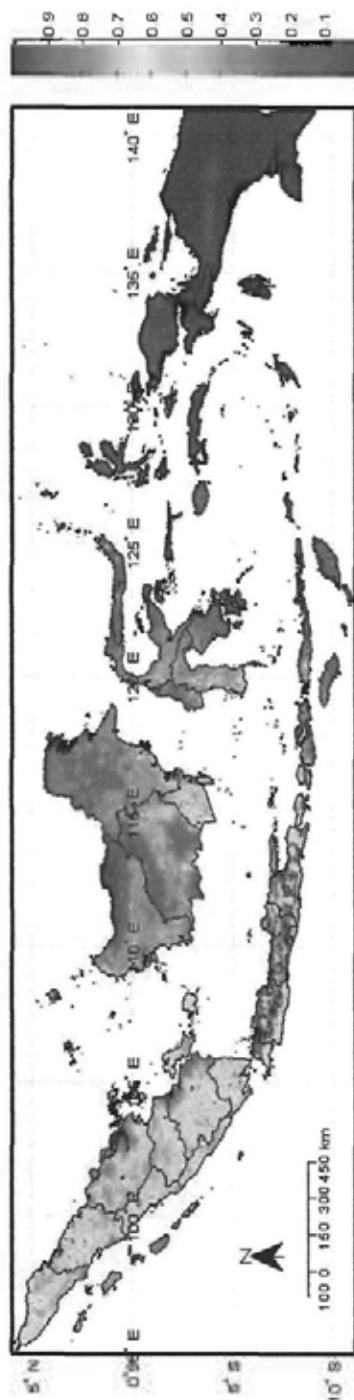
Table 6.2: Summary results of the logistic regression model for the avian influenza H5N1 epidemics in East-Southeast Asia, Indonesia, and China, 1996-2009.

Regent or Country	Con	Alt	PopDen	PolDen	D2Wat	D2Coast	D2Flyway	D2Rail	D2Road
East-Southeast Asia	.5371	-.001	.0007	$-4.01^{-5}$	-.0037	-.0008	-.0007	-.0004	-.249
		$p < .001$	$p < .001$	$p = .142$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
Indonesia	-1.8408	$3.654^{-4}$	$4.298^{-4}$	$4.142^{-4}$	$5.597^{-3}$	$-4.383^{-3}$	$3.904^{-4}$	$-1.853^{-3}$	-.014
		$p = .074$	$p < .001$	$p < .001$	$p = .030$	$p = .012$	$p = .011$	$p < .001$	$p = .082$
China (1996-2009)	-1.637	$3.065^{-4}$	$8.289^{-4}$	$-4.74^{-6}$	-.012	$-2.458^{-4}$	$7.951^{-4}$	$-9.076^{-4}$	-.025
		$p = .036$	$p = .004$	$p = .580$	$p < .001$	$p = .094$	$p = .003$	$p = .166$	$p = .012$
China (1996-2004)	-1.367	$6.515^{-4}$	$1.407^{-3}$	$-4.947^{-5}$	-.034	$-3.718^{-4}$	$1.353^{-3}$	$-4.431^{-3}$	-.041
		$p = .074$	$p = .078$	$p = .459$	$p = .0002$	$p = .288$	$p = .028$	$p = .049$	$p = .175$
China (2005-2009)	-1.422	$2.547^{-4}$	$6.517^{-4}$	$-3.429^{-6}$	-.014	$-4.7^{-6}$	$4.184^{-4}$	$-7.944^{-4}$	-.027
		$p = .145$	$p = .083$	$p = .621$	$p = .001$	$p = .721$	$p = .169$	$p = .289$	$p = .020$

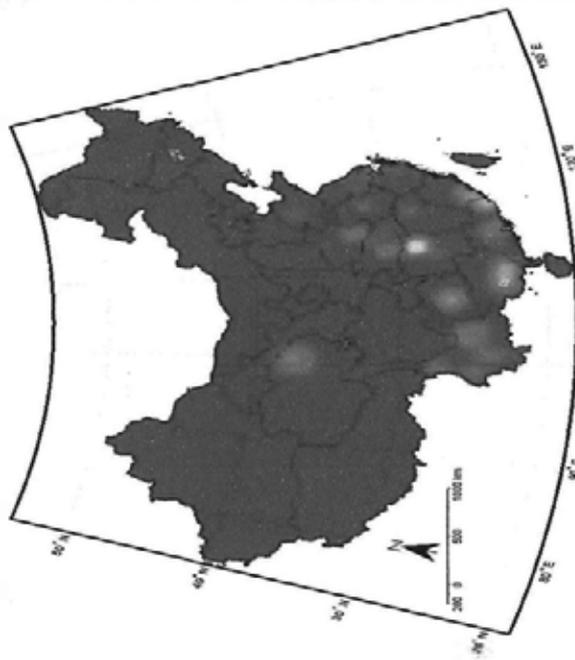
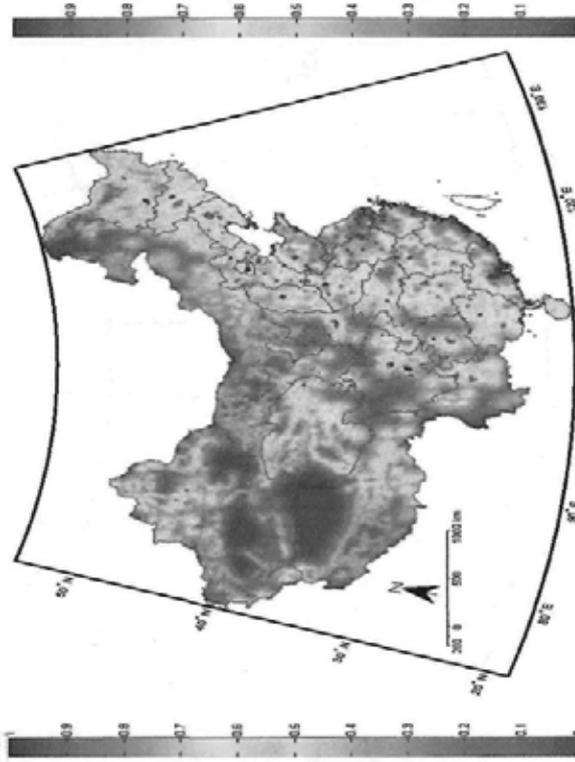
These values are the average of 1000 bootstrap replicates of the logistic regression model. The meaning of the abbreviation shows as follow: Alt = average altitude; PopDen = population density; PolDen = poultry density; D2Water = minimal distance to inland water bodies; D2coast = minimal distance to coastline; D2Flyway = minimal distance to migratory bird pathways; D2Rail = minimal distance to railways; D2Road = minimal distance to roads. Con is the constant of the logistic regression models.



(a)



(b)



(c)

(d)

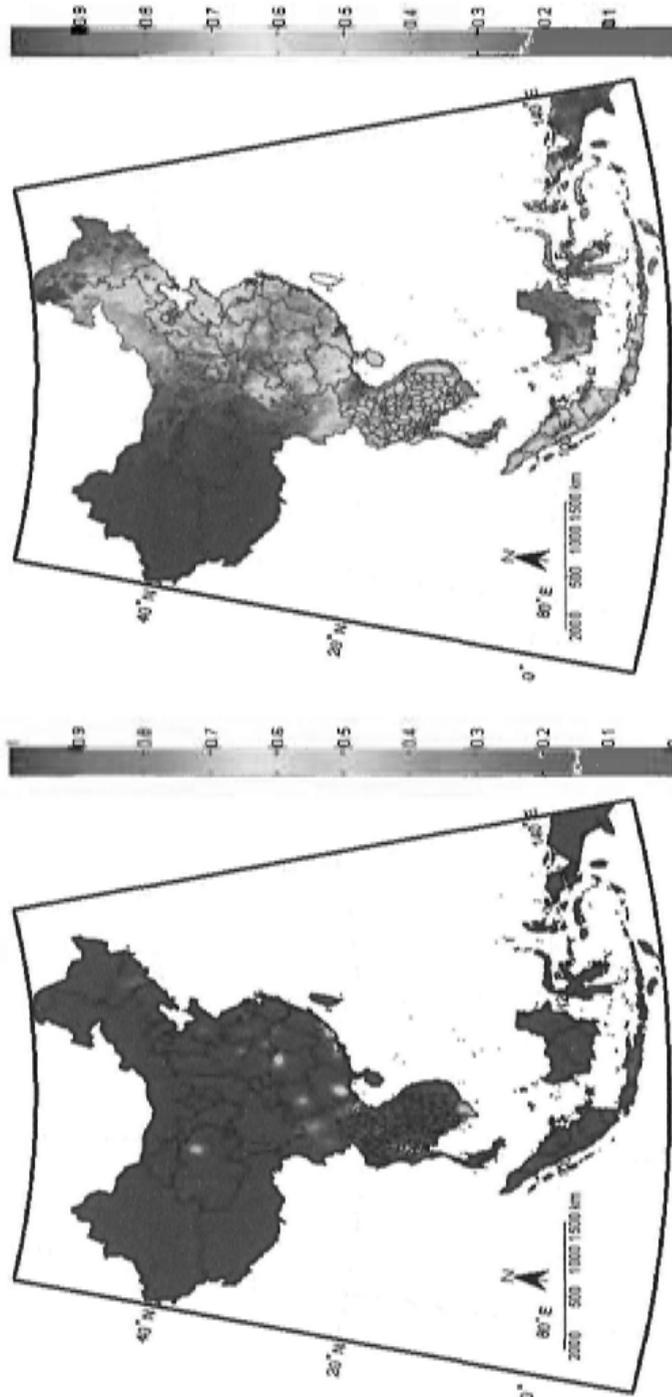


Figure 6.7: Results from phylogenetic tree analysis, and logistic regression analysis for Indonesia, China, East-Southeast Asia. The values shown in the color bar are the probability of outbreaks of H5N1. (a), (c), (e) are the outcomes from the phylogenetic tree analysis (Figs. 6.4 and 6.3), showing the spatial profile of the H5N1 virus; (b), (d), and (f) show the predictive results from the logistic regression models.

information systems (Malpica et al., 2007) and image processing technologies (Kaftandjian et al., 2003; Adamek and O'Connor, 2007); it has also been used in climate change research (Luo and Caselton, 1997), detecting credit card fraud (Panigrahi et al., 2009), and in evaluating outcomes associated with medical intervention (Jones et al., 2006).

Recently, this method has been employed to identify areas at risk from rift valley fever in Africa (Clements et al., 2006). To provide a formal description of D-S theory,  $\Theta$  defines a finite set of mutually exclusive and exhaustive elementary hypothesis  $\{H_1, H_2, \dots, H_n\}$  called the frame of discernment (*FOD*). The number of possible subsets of  $\Theta$  is  $2^{|\Theta|}$ , including the full and null hypotheses. In this study, each lattice point is a binary frame of a discernment containing two elementary hypotheses:  $\{yes\}$  for the presence of H5N1 risk and  $\{no\}$  for absence, i.e.,  $\Theta = \{yes, no\}$ . The subsets of  $\Theta$  are  $\{\phi\}$ ,  $\{yes\}$ ,  $\{no\}$ , and  $\{yes, no\}$ . In particular, the subset  $\{\phi\}$  and  $\{yes, no\}$  stand for empty and unknown (or uncertainty).

In D-S theory, each subset is assigned a belief value by the available evidence, called probability mass function,  $m(\cdot)$ . In particular,  $m(\{\phi\}) = 0$ ,  $\sum_{H \subset \Theta} m(H) = 1$ , where  $H$  represents the subsets  $\Theta$  and  $0 \leq m(\cdot) \leq 1$ . In this study, the results of the three analyses (the capability measure of the H5N1 virus, the clusters of outbreaks, and the results from the logistic regression analysis) provide evidence indicating the belief for the risk of occurrence of H5N1, denoted by  $m_1(\{yes\})$ ,  $m_2(\{yes\})$ , and  $m_3(\{yes\})$ . However, no evidence is directly provided by the first two studies regarding the absence of the disease. Therefore, the values of  $m_1(\{no\})$  and  $m_2(\{no\})$  cannot be determined and so the mass functions of unknown, i.e.  $m_1(\{yes, no\})$  and  $m_2(\{yes, no\})$ , are assigned:  $1 - m_1(\{yes\})$  and  $1 - m_2(\{yes\})$ .

The logistic regression model results are different however because the predictive power of the model allows us to assign values to the mass functions. The average AUC is used to assess model predictions. The multiplication of the predicted value and the average AUC is the evidence for  $m_3(\{yes\})$ . The values of  $m_3(\{no\})$  and  $m_3(\{yes, no\})$  are thus  $(1 - m_3(\{yes\}))(1 - AUC)$  and  $1 - m_3(\{yes\}) - m_3(\{no\})$ , respectively.

The Dempster's rule of combination offers an approach to combining evidence from different sources. The joint probability mass function, for instance  $m(\{yes\})$ , can be obtained from the combination of the two mass functions,  $m_1(\{yes\})$  and  $m_2(\{yes\})$ :

$$m(\{yes\}) = (m_1 \oplus m_2)(\{yes\}) = \sum_{s_1 \cap s_2 = \{yes\}} \frac{m_1(s_1)m_2(s_2)}{k}, \quad (6.6)$$

where

$$k = 1 - \sum_{s_1 \cap s_2 = \emptyset} m_1(s_1)m_2(s_2) > 0, \quad (6.7)$$

where  $k$  is a normalization factor, and  $s_1$  and  $s_2$  are subsets of  $\Theta$ . Dempster's rule is commutative and associative, and thus the joint mass function is independent of the order in which evidence are combined Dempster (1967).

In this study, the three sources of evidence are combined via an iterative procedure in order to identify for each lattice point the 'degree of belief' we have in the likely occurrence of H5N1. To simplify the text we have referred to this as the "risk". Thailand, for example, is represented by a matrix  $3088 \times 1738$  lattice points. Dempster's combination procedure is demonstrated for Thailand in Table 6.3.

Table 6.3: Dempster's combination for the three sources of evidence on H5N1 in Thailand.  
Lattice Point Locations (Row: 1330, Column: 621-630)

Evidence	621	622	623	624	625	626	627	628	629	630
$m_1(\{yes\})$	.5920	.6144	.6357	.6558	.6747	.6922	.7081	.7222	.7345	.7447
$m_1(\{no\})$	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
$m_1(\{yes, no\})$	.4080	.3856	.3643	.3442	.3253	.3078	.2919	.2778	.2655	.2553
$m_2(\{yes\})$	.6623	.6650	.6678	.6709	.6743	.6765	.6810	.6864	.6927	.7002
$m_2(\{no\})$	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
$m_2(\{yes, no\})$	.3377	.3350	.3322	.3291	.3257	.3235	.3190	.3136	.3073	.2998
$m_3(\{yes\})$	.6300	.6304	.6211	.6887	.3687	.3682	.5207	.5953	.6559	.6562
$m_3(\{no\})$	.1110	.1109	.1437	.1894	.1894	.1895	.1438	.1214	.1032	.1031
$m_3(\{yes, no\})$	.2590	.2587	.3352	.4419	.4419	.4423	.3355	.2833	.2409	.2407
Combination										
$m(\{yes\})$	.9436	.9472	.9337	.9141	.9195	.9241	.9487	.9604	.9690	.9709
$m(\{no\})$	.0169	.0159	.0199	.0258	.0242	.0228	.0154	.0119	.0093	.0087
$m(\{yes, no\})$	.0395	.0370	.0464	.0601	.0564	.0531	.0359	.0278	.0217	.0204

### 6.3.4 Spatial Correspondence Analysis

Spatial correspondence analysis using Pearson's correlation coefficient,  $R$ , Haining (1991, 2003) measures the association between observed avian-influenza outbreaks and model predictions. A sample size adjusted t-test was employed to test statistical significance of  $R$ . The reduction in the degrees of freedom is a function of the level of spatial autocorrelation in the two maps that are being correlated. This is because spatial autocorrelation introduces redundancy into a set of data and the adjustment procedure identifies the "equivalent" number of independent observations.

Mathematically, the Pearson product moment correlation coefficient is defined as:

$$\hat{r} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (x_i - \bar{x})^2}} \quad (6.8)$$

where  $\bar{x}$  and  $\bar{y}$  are sample means and  $n$  is the sample size. The adjusted sample size can be calculated from

$$N' = 1 + n^2(\text{trace}(\hat{R}_x \hat{R}_y))^{-1}, \quad (6.9)$$

$\hat{R}_x$  and  $\hat{R}_y$  are the estimated  $n \times n$  spatial correlation matrices for two spatial processes  $X$  and  $Y$ . If they are spatially uncorrelated,  $\text{trace}(\hat{R}_x \hat{R}_y) = n$  and  $N' \cong n$ .

Assuming approximate normality, and null hypothesis  $H_0$  is  $r = 0$ ,  
If

$$t = (N' - 2)^{1/2} |\hat{r}| (1 - \hat{r}^2)^{-1/2} \quad (6.10)$$

exceeds the critical value of t-statistic with the degree of freedom  $N' - 2$ ,  $H_0$  will be rejected.

For the implementation, lattices were aggregated into blocks vary-

ing in size from  $60 \times 60$  to  $30 \times 30$  because the number of cases is small relative to the original spatial resolution of the data which is extremely fine-grained. The observed outbreaks were converted to a rate by calculating for each areal unit the number of outbreak cases divided by the population at risk which is the human plus poultry population,  $(\text{number of outbreak})/(\text{population} + \text{poultry})$ . Model predictions were obtained by averaging results across the lattice points falling within any block. Results from the analysis of spatial correspondence using the sample size adjusted t-test for significance testing are shown in Table 6.4.

## 6.4 Results and Interpretations

### 6.4.1 Thailand and Vietnam

Table 6.4 and Figure 6.5 show the result for Thailand and Vietnam. Table 6.4(c) and (f) demonstrate our estimate of the spatially varying ‘degree of belief’ in the level of risk of H5N1 obtained by integrating the three forms of analysis (see Figs. 6.5(a) and (c), and Figures. 6.5(b) and (d), and Table 6.4(b) and (e)) using the Dempster-Shafer theory of knowledge fusion. Tables 6.4(b) and (e) are the results from the epidemiological analysis of Gilbert. et al. (Gilbert et al., 2008). In both cases the closer any area’s value is to 1 (the redder it is) the greater the likelihood of an H5N1 outbreak in that area. Table 6.4 also reports Pearson correlation coefficients (R) and associated  $p$ -values which show that our experimental patterns ((c) and (f)) have a closer correspondence to the observed pattern of cases ((a) and (d)) than the results of Gilbert et al. (Gilbert et al., 2008) ((b) and (e)). This finding holds over a range of spatial scales from  $60 \times 60$  cell aggregates to  $30 \times 30$

cell aggregates.

Table 6.4(c) shows that the greatest risk of H5N1 is in the upper central region and the lower part of north Thailand. It indicates less risk in central Thailand than predicted by Gilbert et al. (Gilbert et al., 2008) (Table 6.4(b)). This pattern corresponds with fewer observed outbreaks in this area of Thailand.

For Vietnam, the greatest risk occurs in the north and to a lesser extent the south of the country (Table 6.4(f)). Unlike the work of Gilbert et al. (2008), our results not only model the spatial distribution of H5N1 outbreaks, but also the space-time dynamics of viral evolution. Our analysis combines real-world outbreak data with evidence on viral evolution. For instance, the H5N1 virus was first detected and became established around Hanoi, north Vietnam, in 2001 and consequently spread to the south around Ho Chi Minh city (Wan et al., 2008). Phylogenetic analysis shows that the virus, isolated from the north, has multiple sublineages and shares a close phylogenetic relationship with the virus from Thailand, Malaysia, Laos, and provinces in southern China (Chen et al., 2006; Smith et al., 2006b). Furthermore, the northern H5N1 virus is associated with novel genetic subtypes, and these have facilitated the spread of the disease both within and outside the country (Wan et al., 2008). Finally, it was reported that the number of H5N1 outbreaks decreased in the south in late 2005, but the disease still persists in causing outbreaks in northern Vietnam (FAO, 2010; Wan et al., 2008). Table 6.4 again demonstrates that the risk estimates obtained from our integrated analysis correspond more closely than the results from Gilbert et al. to the empirical outbreak pattern at a range of different scales.

Table 6.4: Test for spatial correspondence of H5N1 outbreaks and the empirical patterns  
Assessment of Bivariate Spatial Association

Country	OBK	ALM	KF	Pattern	Group Size	R	n	AdjN'	DF	p-value
Thailand				OBK-ALM	60x60	.2035	1508	85	83	.050
				OBK-KF	60x60	.2871	1508	84	82	.010
				OBK-ALM	50x50	.1870	2170	99	97	.050
				OBK-KF	50x50	.2630	2170	99	97	.010
				OBK-ALM	40x40	.1487	3432	141	139	.075
				OBK-KF	40x40	.2159	3432	142	140	.010
Vietnam				OBK-ALM	30x30	.1178	5974	178	176	.100
				OBK-KF	30x30	.1778	5974	182	180	.010
				OBK-ALM	60x60	.1206	1770	263	261	.050
				OBK-KF	60x60	.1530	1770	265	263	.010
				OBK-ALM	50x50	.1012	2485	305	303	.070
				OBK-KF	50x50	.1320	2485	306	304	.020
			OBK-ALM	40x40	.0903	3872	324	322	.100	
			OBK-KF	40x40	.1179	3872	325	323	.025	
			OBK-ALM	30x30	.0598	6962	537	535	.150	
			OBK-KF	30x30	.0811	6962	539	537	.050	

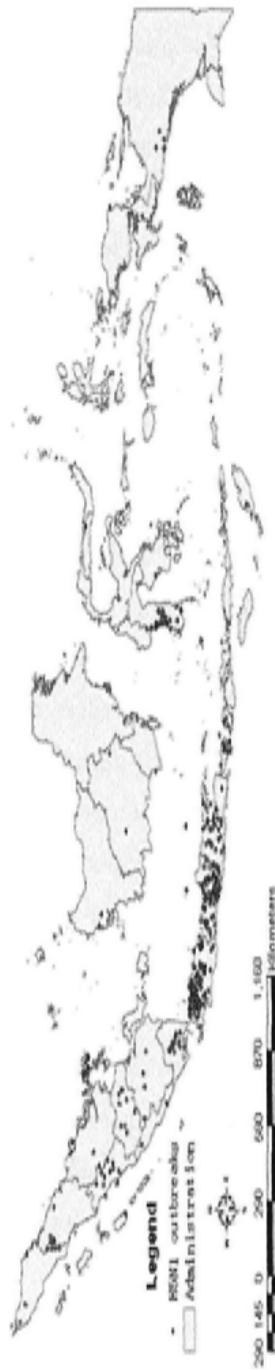
OBK: outbreaks; ALM: Auto-logistic model; KF: Knowledge Fusion; R: Pearson's correlation coefficient; n: sample size; AdjN': adjusted sample size; DF: degree of freedom; Group sizes: the number of lattice points grouped as an experiment unit in assessing the spatial correspondence between outbreaks and experimental patterns: ALM and KF.

#### 6.4.2 Indonesia, China, and East-Southeast Asia

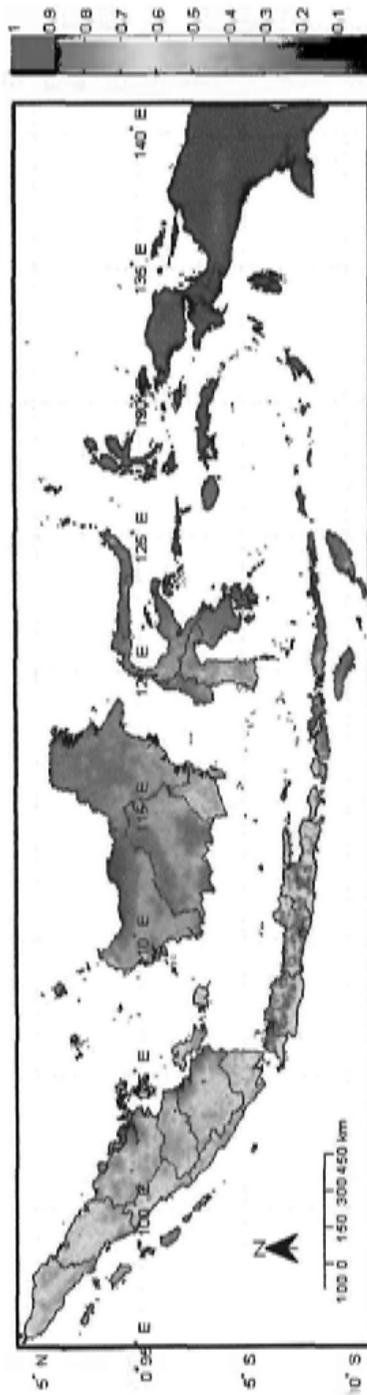
Figure 6.8 shows the observed distribution of H5N1 outbreaks and our estimates of the risk of the disease in Indonesia, China, and the whole of East-Southeast Asia. The integrated patterns shown in Figures 6.8(b), (d), and (f), actually, are risk maps which not only reflect avian-influenza outbreaks, but also characterize the probabilities of occurrence of the disease by integratedly considering the H5N1 evolution, outbreak clusters, and environment. In these Figures, the larger is the value approaching to 1, the higher is the probability of developing the disease. Table 6.2 summarizes the logistic regression analysis in the three areas.

For Indonesia, the highest risk of H5N1 is in central Java (Figure 6.8(b)). It shows that the risk extends through the island to its surrounding archipelagos. The findings from the logistic regression analysis provide an interpretation for this pattern. Table 6.8 illustrates that population density, poultry density, and the shortest path distance to railways are significantly ( $p < 0.001$ ) associated with occurrences of the disease. The results indicate that the very high density of population (more than 940 people per  $\text{km}^2$ ) and the high concentration of poultry production (60% of Indonesia's production takes places in Java) are underlying factors in the establishment of multiple H5N1 subgroups (Smith et al., 2006b). These factors greatly increase the risk of H5N1 occurrence in Java. In addition, the significant negative relationship between H5N1 and the shortest path distance to railways implies that avian influenza is spread to other surrounding archipelagos through the production and trading of poultry (Smith et al., 2006b).

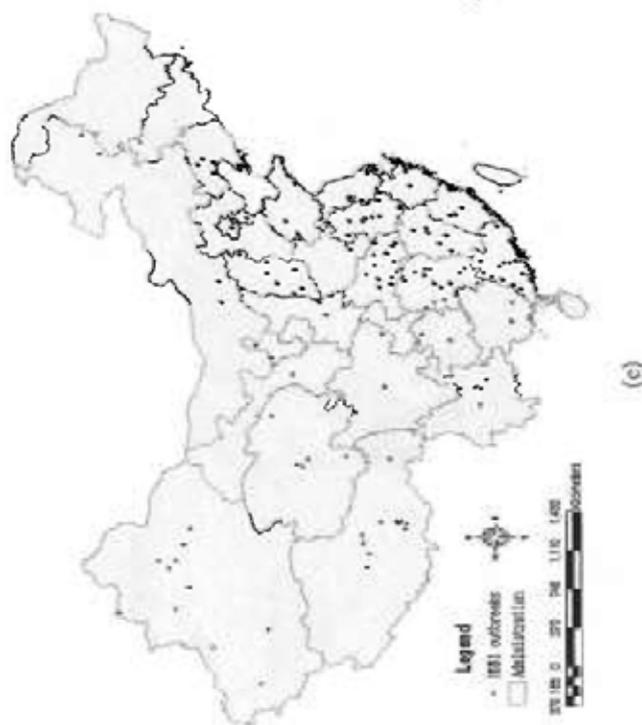
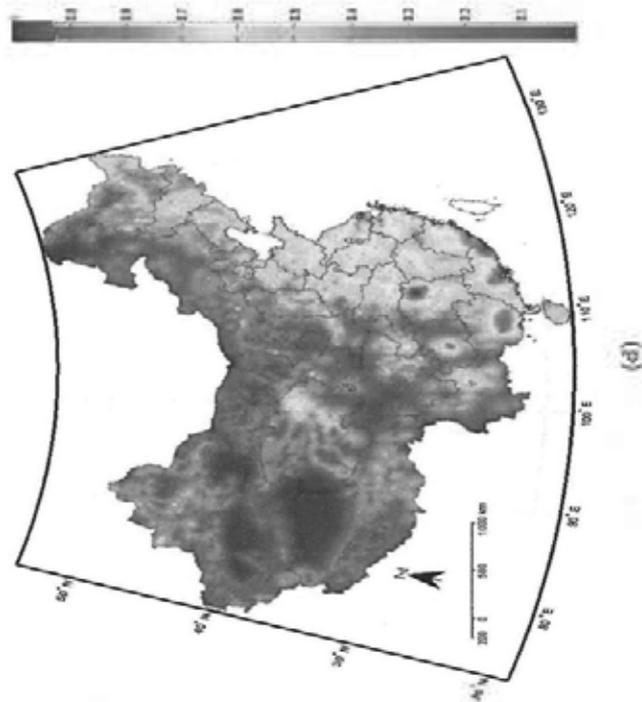
For China, our result shows that the highest risk of H5N1 is in Guangdong, Yunnan, Fujian, and areas close to Dongting Lake in Hu-



(a)



(b)



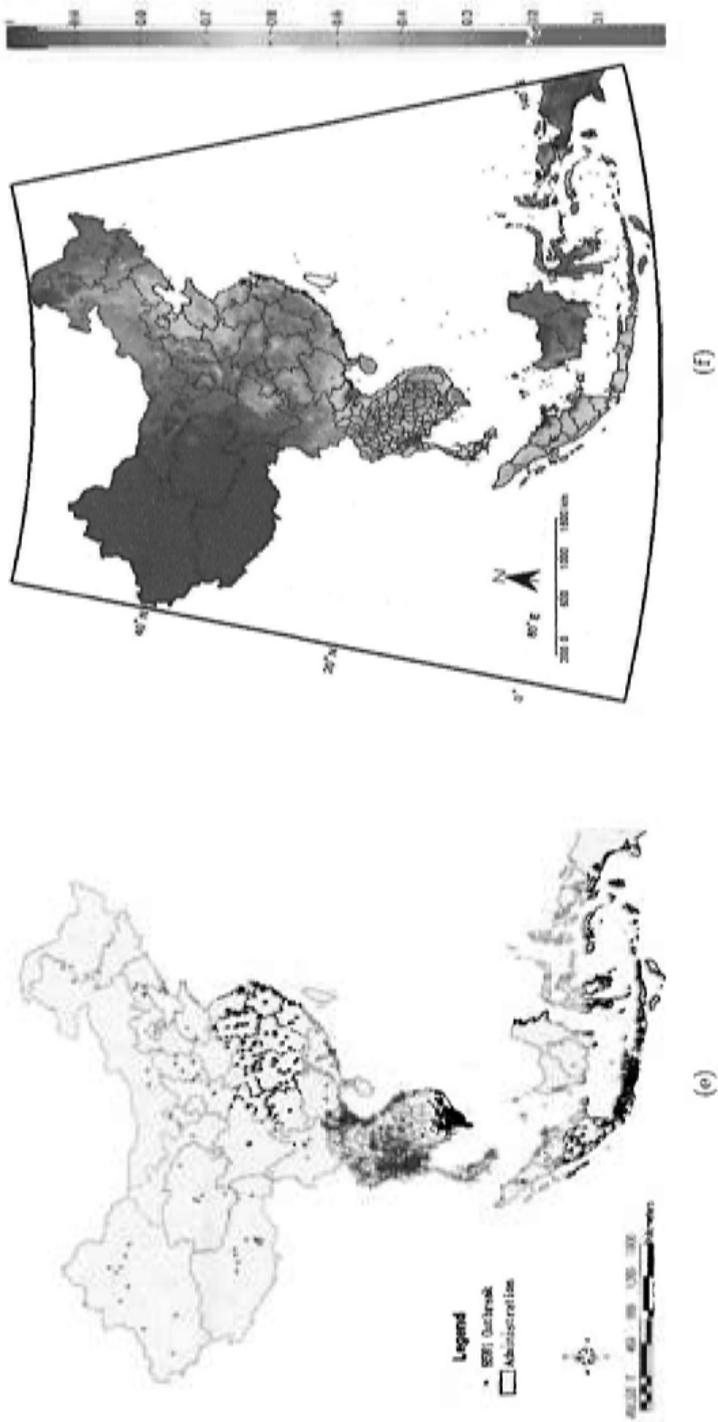


Figure 6.8: The spatial pattern of H5N1 in Indonesia, China, and East-Southeast Asia. (a), (c), (e) show the distribution of observed H5N1 outbreaks; (b), (d), and (f) show the probability maps integrating the findings of the phylogenetic analysis (Figs. 6.7(a), (c), and (e)), the modified local  $K$  function analysis (Figs. 5.7(h), (f), and Fig. 5.6), and the logistic regression analysis. The closer the probability is to 1 the greater is the probability of an H5N1 outbreaks.

nan province, south China. High risk also appears to extend along the coastline of southern China (see Fig. 6.8(d)). The risk increases from the northwest to southeast. Qinghai Lake, where over 6000 migratory birds were infected and killed by HPAI H5N1 in early 2005 (Chen et al., 2005; Liu et al., 2005), was also highlighted as a potential source for the disease. The logistic regression analysis identified population density and the shortest path distance to inland water bodies as significantly ( $p < 0.005$ ) associated with the occurrence of H5N1 (Table 6.2). In eastern China, farming, free-grazing poultry, and economic trade and movement, have played significant roles in the maintenance and the transmission of avian influenza. The significant positive relationship ( $p < .005$ ) between the shortest path distance to migratory bird pathways and outbreaks, however, appears to suggest that bird migration may not be a key factor triggering large outbreaks and viral transmission in China, particularly in eastern China between 1996 and 2009, even though bird migration has been widely thought as a cause of wide spread of the disease globally (Chen et al., 2005; Liu et al., 2005; Olsen et al., 2006). The other socio-environmental variables (as shown in Table 6.4), including altitude, population density, poultry density, and the shortest distances to inland water bodies, coastlines, migratory bird pathways, railways, and roads, fail to show statistical significance within the two periods between 1996 and 2004, and 2005 and 2009. This might be due to data limitations associated with using such a short time span.

Figure 6.8(f) depicts the risk of occurrence of avian influenza H5N1 in East-Southeast Asia. The large-scale mapping shows that the highest risk is in central Thailand and the northern and southern parts of Vietnam. The central part of Indonesia also has high risk. Compared

to these countries, China appears to have lower levels of risk, especially in the northwest including Tibet and Xinjiang autonomous regions, and the northern part of Qinghai province. This pattern can be regarded as a reflection of the relationship between the disease and heterogeneity of the local socio-ecological environment. The logistic regression analysis indicates that outbreaks significantly ( $p < 0.001$ ) associated with altitude, population density, and the shortest path distance to inland water bodies, coastlines, railways, and roads (Table 6.2). In East-Southeast Asia, most cities and countries are usually located in areas where the environment is suitable for human habitation and agricultural production. Rice cropping and poultry rearing are popular in Thailand, Vietnam, Indonesia, and south China. These countries benefit from abundant hydrological resources (e.g., river deltas), but suffer from ecological-environmental problems caused by a rapidly increasing population which in turn facilitates the establishment of multiple H5N1 sublineages (Chen et al., 2005). In addition, convenient transport networks like railways and roads become significant in spreading and triggering the re-occurrence of H5N1 in East-Southeast Asia.

## 6.5 Summary

This chapter employed genetic analysis that identifies the evolution of the H5N1 virus in space and time, epidemiological analysis that determines socio-ecological factors associated with H5N1 occurrence and statistical analysis that identifies outbreak clusters, and then applied a methodology to formally integrate the findings of the three sets of methodologies.

This study is novel and significant in seeking to lay a solid foundation for the inter-disciplinary study of this and other relevant influenza

epidemics. First, it uses DNA sequences and space-time data to create a phylogenetic tree to estimate the virus' capability of spreading. This is the first attempt to provide a mapping of H5N1 viruses derived from the phylogenetic tree. Second, by integrating the results obtained from the three analyses, we offer insights into the occurrence and space-time spread of H5N1 that have a higher level of correlation with empirical evidence than is found when analysis is based on only one methodology.

In addition, we applied the methodology across multiple scales; that is to the whole of East-Southeast Asia as well as the individual countries of Thailand, Vietnam, Indonesia, and China, respectively. Our analysis results in a significant advance in findings over those reported in, for example, Gilbert et al. (2008), and we believe our findings are more precise and informative in representing the occurrence and the space-time dynamics associated with the spread of H5N1.

---

□ End of chapter.

## Chapter 7

# Conclusion

### 7.1 Summary

This research set out to explore the spatial and temporal patterns of avian influenza H5N1 and to identify factors affecting the spread of the disease. On these questions, the work makes three primary contributions: 1) it explores the long-range correlation and multifractality of the time series of H5N1 outbreaks; 2) it devises an efficient method for mapping the risk of H5N1 disease by integrating studies and techniques initially proposed in different knowledge domains, and 3) it models the association between the occurrence of H5N1 and socio-ecological environments in Thailand, Vietnam, Indonesia, and China, as well as the whole of East-Southeast Asia. In addition, we proposed a statistical method for automatically identifying the crossover point of time series. The method was discussed in Ge and Leung (2011).

We started, in Chapter 2, by illustrating the importance of studies of avian influenza as they have been offered from three perspectives: that of H5N1 history, characteristics of the H5N1 virus, and of its mode of transmission. In Chapter 3, we give the conceptual framework

within which the spread of avian influenza can be examined under a unified view that takes into account the evolution of H5N1, the spatial and temporal spread of avian influenza, and the environmental and socio-economic factors conducive to the outbreak of the disease. The framework pays particular attention to the interactions among these processes. Based on the framework, we first investigate in Chapter 4 the scaling behavior of H5N1 outbreaks in wild birds and poultry over time. This chapter sought to answer the following issues: (1) Are previous H5N1 outbreaks responsible for current spells of infection? (i.e., Are outbreaks long-range correlated?); (2) Do H5N1 outbreaks exhibit multifractality properties; and (3) Are the temporal behaviors different among the continents? Through the application of multifractal detrended fluctuation analysis (MF-DFA) to the H5N1-outbreak time series, the study was able to work out the long-range correlation and multifractal properties of the disease. The seasonal pattern was also determined by the crossover time scale. A comparison of the scaling behaviors of the outbreak time series suggested different mechanisms of the transmission of the disease in Asia, Europe, and Africa, respectively. These results underlined a need for tighter surveillance and control of H5N1 through with international cooperation in order to prevent outbreaks and spread of avian influenza.

In applying MF-DFA, we found that the detection of crossover time scale(s) was relatively subjective since did not rest on rigorous statistical procedures and was generally determined by eyeballing or subjective observation. Crossover time scales determined in this way risk being problematic, even spurious. The results of subjective estimates may not thus reflect the genuine underlying scaling behavior of a time series. In place of these approximate method, we proposed a statisti-

cal procedure to terminate the number and location of crossover time scales hidden in the fluctuation of a time series. In order to ensure statistical significance, the method also establishes confidence intervals for the crossover time scales. By this means, genuine crossover time scales can be rigorously determined and crossover time scales which went unnoticed by conventional observation method can be successfully captured. This makes the detection of crossover time scales in MF-DFA more objective, reliable, and with statistical sense. The proposed method was the first attempt to employ a statistical procedure for the determination of crossover time scales in MF-DFA.

In Chapter 5, we went on to study the spatial pattern of H5N1 outbreaks. This study included the identification of both global and local patterns of the outbreaks in humans and bird populations of the world. We applied the  $K$  function and the local  $K$  function to explore trends and clusters of the disease at multiple spatial scales.  $K$  function analysis indicated that the global trends of the disease were different in humans and avian. Our local pattern analysis using a local  $K$  function further identified clusters of outbreaks in Thailand, Vietnam, and Egypt.

However, we recognized that the assumption of spatial homogeneity implicit in the working of traditionally defined  $K$  function had the potential to limit the value of analyses of the disease's spatial scaling behavior. To meet this concern, this work then proposed a modified local  $K$  function to estimate the degree of outbreak clusters on lattices, allowing for the spatial effects caused by distance. This modified function was used to detect clusters of outbreaks in Thailand, Vietnam, Indonesia, China, and East-Southeast Asia. We identified different patterns as characterizing these countries and various ecological envi-

ronments of the wider region. The existence of these different forms of patterning behavior suggests that H5N1 may transmit itself via distinct mechanisms in different countries and different environments. It seems plausible to suggest that these mechanisms will vary in association with the different local socio-ecological environments and the behavioral patterns of hosts, including humans and birds.

Finally, Chapter 6 studied a set of socio-environmental factors possibly associated with the disease, including altitude, population density, poultry density, and the shortest path distances to inland water bodies, coastline, migrating bird pathways, railways, and roads. The relationship between outbreaks and these factors were quantified by the logistic regression model. By this means, we predicted the occurrence of H5N1 risk for Indonesia, China, and East-Southeast Asia.

Through our attempts at modeling, we note that uncertainty, whether arising from incomplete data, limited domain knowledge, or the application of an insufficiently sophisticated methodology could limit the rigor of H5N1 studies. In essence, the study of avian influenza H5N1 is multidisciplinary across virology, molecular biology, medical geography, and spatial epidemiology. Any study that relies on only one kind of disciplinary knowledge may fail to make important connection with various domains. With this perspective, this work proposed a novel approach for formally integrating multiple information from different data sources and the distinct domain knowledge. Overall, our analysis result is significantly more revealing than that of the analyses based on only one kind of methodologies. In addition to having higher predictive power, the integrated patterns shown on our maps are substantively informative in depicting the space-time dynamics of viral evolution.

To sum up, this thesis proposed three methods – (1) applying multi-

fractal detrended fluctuation analysis to determine the temporal scaling behavior of avian-influenza outbreaks (Chapter 4); (2) modifying the local  $K$  function to identify the clusters of outbreaks in space (Chapter 5) and (3) applying spatial epidemiological model and Dempster-Shafer evidence theory to explore the distribution of H5N1 risk (Chapter 6). The study seeks to lay a solid foundation for the future interdisciplinary study of influenza outbreaks in general and avian influenza in particular.

## 7.2 Directions for Further Research

Building on the knowledge gained in this research, it is beneficial in carrying out further research along the following directions.

1. The multifractal detrended fluctuation analysis (MF-DFA) presented in Chapter 4 could be applied to detect the temporal scaling behavior of avian influenza and other infectious diseases (such as HIV and dengue fever) at multiple spatial scales based on city, country, continent, and the world. To make it viable in practice, further work need be done to collect space-time diseases data at different levels. This analysis could also be combined with method of geographical statistics introduced in Chapter 5, depending on whether a substantial database is available for the disease under consideration.
2. In Chapter 5, it is possible to include the variable of time into spatial analysis, e.g. based on the  $K$  and local  $K$  function, instead of just estimating the effect of distance. Hence, by adding variables of space and time, the detection of disease hot spots could be examined along the two dimensions. This will enable

better decision making in the prevention of the disease.

3. It is worthy to extend the techniques presented in Chapter 4 and 5 to other subtypes of avian influenza, thereby seeking to understand the spatial and temporal association between H5N1 and other avian viruses. A conceptual and technical challenge will be on measuring the association between these viruses in terms of space, time, and the viral evolutionary processes.
4. In Chapter 6, it might be profitable to extend the analysis of phylogenetic tree to the analysis of genetic networks. The quantification of a phylogenetic network would enable us to formally measure the relationships among the viruses. By linking up with space and time, such an identification of networks might enhance our understanding of the association of avian influenza based on their space-time evolutionary processes.
5. Finally, the thesis can be made more comprehensive and policy relevant if detailed data about poultry trades and products, human socio-economic activities, wild birds' migration, investment in public for hygiene, health care service, and public health practice could be made available. Applying our approach to the analysis of these data will be a direction for future research capable of providing effective measures and strategies for prevention of the spread of infectious diseases.

---

□ End of chapter.

## Appendix A

# Detection of Crossover Time Scales in MF-DFA

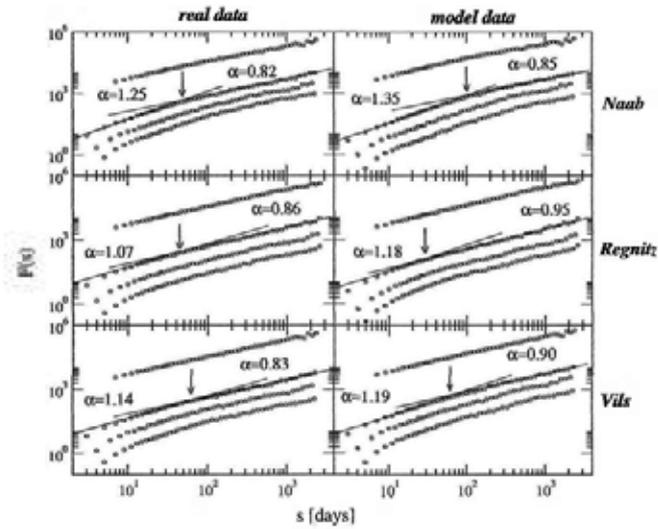
Fractal analysis provides a mathematical formalism to characterize intricate spatial and dynamical structures (Feder, 1988). The method intends to study extremely irregular objects that cannot be described by Euclidean geometry because of their fractional dimension. The study of fractal scaling behaviors examines how time-series patterns vary with the change of temporal scales. In other words, the scaling relationship between pattern and its measurement scale can be described as the size of steps required to cover a fractal series that varies as the scale is raised to a scaling exponent given by the fractal dimension. The study explores the characteristics of time series in terms of long-rang correlation and fractal property, giving rise to diverse behaviors with respect to the change of scale. In practice, to determine fractal scaling behaviors, we measure the series using various step sizes

Oftentimes, complex fractal series does not exhibit mono-scaling behaviors characterized by a single scaling exponent. Multi-scaling behaviors are actually very common in natural phenomena, such as

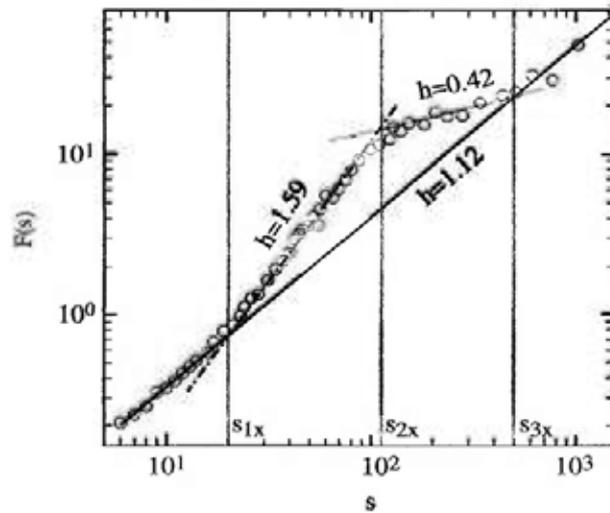
hydrological processes (Kantelhardt et al., 2003), temperature fluctuations (Alvarez-Ramirez et al., 2008), sunspot activities (Movahed et al., 2006b), avian-influenza outbreaks (Leung et al., 2011), etc.. One of the important concepts is the crossover time scale(s) that separates distinct regimes having different scaling behaviors. Actually, crossover time scale(s) can be treated as the crossover point at which the fractal structure changes its behavior. For example, the time series may be long-range correlated in large scales of  $s > s_x$ , while it may not happen in small scales  $s < s_x$ , where  $s_x$  is the crossover time scale. Figure A.1 shows two examples of crossover time scales reported by Livina et al. (2007) and Movahed et al. (2006b).

The detection of crossover time scale(s) is consequently employed to distinguish multi-scaling behaviors. Some investigations intend to find the factors leading to the occurrence of crossover time scales and mixed fractal structures by detrending (Kantelhardt et al., 2001; Hu et al., 2001b). Multifractal detrended fluctuation analysis (MF-DFA) (Kantelhardt et al., 2001) is a commonly used method. By convention, the detection of crossover time scales in MF-DFA is by observing the change of the log-log plot of the fluctuation function. However, exactly how one should identify and locate the genuine crossover points that separate different regimes remains an open issue. Although Butler et al. (2001) analytically identified crossover points by minimizing the residual of the linear fits on fractal scalings, it has no solid theoretical foundations for the quantitative detection of crossover time scales. Up to this moment, the detection of crossover time scales has been done by the visualization of the log-log plot and the presence of seemingly different exponent values. Whether the observed crossover points are significant in the statistical sense and whether there are genuine

APPENDIX A. DETECTION OF CROSSOVER TIME SCALES IN MF-DFA160



(a)



(b)

Figure A.1: Two examples for crossover points provided by Livina et al. (2007) and Movahed et al. (2006b): (a) “Arrows denote average points of crossovers, where the scaling exponents change” and (b) “Crossover behavior of the log-log plot of  $F(s)$  versus  $s$  for the sunspot time series for  $q = 2.0$ . There are three crossover timescales in the plot of  $F(s)$ , at scales  $s_{1x}$ ,  $s_{2x}$  and  $s_{3x}$ ”.

crossover points that went unnoticed have not been investigated.

Actually, crossover time scales can be identified by some rigorous statistical methods that search for the best fit of the log-log plot. For simple fractal scaling, it can be fitted by a linear regression model. The gradient is the index characterizing the fractal scaling behavior. For complex fractal series, multiple scaling behavior exhibits itself as a kinked kind of curve with piecemeal linear fits. And the crossover points should be the joinpoints of the kinked curve. Segmented line regression model (Muggeo, 2003; Kim et al., 2004; Tiwari et al., 2005; Yu et al., 2007), which is composed of a few continuous linear phrases, appears to be a natural model to characterize the multi(mono)-scaling behaviors of time series. Through the procedure of model selection and parameter estimation, crossover points in multi-scaling time series can be statistically identified with rigorous justifications.

To facilitate our discussion, the statistical model for the description of multi-scaling behaviors and the statistical procedure for detecting the crossover points will be detailed in the sections to follow.

### **A.1 Scaling-identification Regression and Detection of Crossover Points**

As discussed above, a single scaling behavior can be modeled by a linear fit of the fluctuation function. However, for multi-scaling behaviors, it is apparent that a single linear fit fails to characterize the fluctuation function. It appears that if there multiple crossover time scales, the linear fit will be segmented into a kinked curve with each linear segment representing a distinct regime of the multi-scaling behavior of the time series under study. Thus, it is essential to have an appropri-

ate model with which we can rigorously characterize such behavior by determining the number of crossover time scales (crossover points) in strict statistical sense. Due to the kinked nature of the linear fit, the segmented linear regression model (also called multi-phase regression, piecewise regression, broken line regression, and joinpoint regression) (Feder, 1975; Ertel and Fowlkes, 1976) with continuous linear phases should be appropriate for the statistical description of multiple scaling behaviors. Here, we apply the segmented linear regression model to detect the crossover points of the fluctuation function, derived by the MF-DFA.

Supposed that  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  are the observations of the fluctuation values at each scale, i.e.,  $x = \log s$  and  $y = \log F_q(s)$ , the responses are  $\hat{y}_i = E(y|x_i) + \varepsilon_i$ ,  $i = 1, \dots, n$ , assuming that  $E(\varepsilon_i) = 0$  and  $V(\varepsilon_i) = \sigma_i^2$  for random errors  $\varepsilon_i$ . For each regime, the scaling behavior is described by a linear fit expressed as:

$$E(y|x) = \beta_{k,0} + \beta_{k,1}x, \quad \tau_{k-1} < x \leq \tau_k, \quad (\text{A.1})$$

where  $\beta_{k,0}$  and  $\beta_{k,1}$  are the intercept and gradient that can be regarded as parameters characterizing the scaling exponents in the MF-DFA.

In general, the scaling model with  $k$  crossover time scales can be expressed as

$$E(y|x) = \beta_{10} + \beta_{11}x + \sum_{k=1}^K \delta_k(x - \tau_k)^+, \quad (\text{A.2})$$

where  $\beta_{10}$  and  $\beta_{11}$  stand for the intercept and gradient of the first regime, respectively;  $\tau_k$ ,  $k = 1, \dots, K + 1$ , is the  $k$ th crossover time scale to be determined;  $\delta_k(x - \tau_k)^+$ , depicts the fluctuation of the  $k$ th subregion, when  $x > \tau_k$ , and  $x - \tau_k = 0$ , if  $x \leq \tau_k$ ;  $\delta_k$  is the coefficient

equals to  $\beta_{k+1,1} - \beta_{k,1}$ .

This is actually the joinpoint regression model that has been used to describe the trends of cancer incidents over time (Muggeo, 2003; Kim et al., 2004). In the context of MF-DFA, we call it the scaling-identification regression model. And, it is used to determine the exact form of the fractal scaling behavior by identifying the significant crossover points of the MF-DFA fluctuation function. Specifically, it is employed to determine the locations and number of crossover points that are statistically significant. That is, if there are only  $N$  statistically significant crossover points, then the exact form of the scaling-identification regression model can be statistically determined. Kim et al. (2000) proposed an iterative hypothesis testing procedure to determine the number of joinpoints in the regression model. Similar approach is employed in our investigation.

## A.2 Determination of Crossover Time Scales by Statistical Inference

The key to the determination of crossover time scales is the identification of the significant crossover points of the log-log plot of the MF-DFA fluctuation function. Supposed that  $k_0$  and  $k_1$ , with  $0 \leq k_0 < k_1$ , are the minimum and the maximum crossover points respectively. The iterative statistical inference starts from the null hypothesis  $H_0$ : there are  $k_0$  crossover points against the alternative hypothesis  $H_1$ : there are  $k_1$  crossover points. If  $H_0$  is rejected at significance level  $\alpha_1$  and  $k_1 - k_0 \geq 2$ , we will test the hypothesis:  $k_0 + 1$  vs  $k_1$ . On the other hand, if  $H_0$  is not rejected and  $k_1 - k_0 \geq 2$ , we will test the hypothesis:  $k_0$  vs  $k_1 - 1$ . The iterative inferential procedure ends when the null

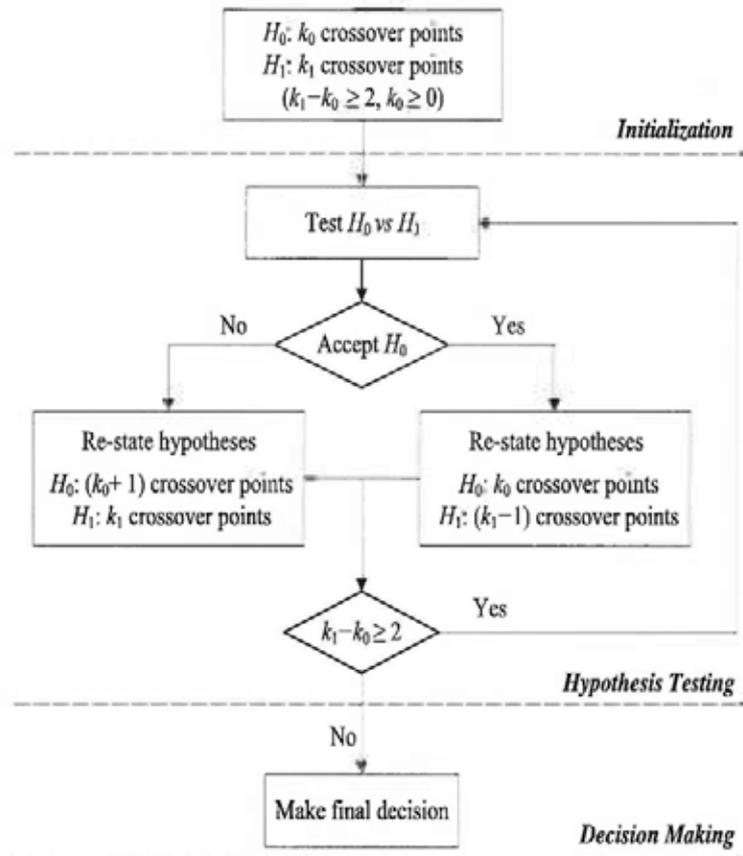


Figure A.2: Inference procedure for the determination of Scaling-identification Regression Model

hypothesis of  $k$  crossover points is tested against the alternative hypothesis of  $k+1$  crossover points, where  $k_0 \leq k < k_1$ . Hence, the number of crossover points is determined as  $k+1$  when the null hypothesis assuming  $k$  crossover points is rejected, and it is  $k$  otherwise. In each step, the significance level is adjusted as  $\alpha_1 = \alpha / (k_1 - k_0)$  by the Bonferroni correction in order to ensure that the entire significance level equal to  $\alpha$ . The iterative procedure of this statistical inference is depicted in figure A.2.

For calibration, we applied the grid-search method proposed by Lerman (1980) to obtain the best-fit curve of the data and give a significance test by permutation (Kim et al., 2000) for each hypothesis. The implementations of the model fitting and the crossover time scales detection procedures are elaborated in the next section.

### A.3 Model Fitting and Detection of Crossover Points

Lerman (1980) developed the grid-search method of fitting segmented regression curves with unknown transition points. This technique is suitable for fitting the scaling-identification regression model and for providing as a by-product a way of making reliable inference on the crossover points. Essentially, it estimates the parameters,  $\beta_0, \beta_1, \delta_1, \dots, \delta_n$ , and  $\tau_1, \dots, \tau_n$ , in the proposed model by minimizing the residual sum of squared error (*SSE*)

$$R^2 = \sum_{i=1}^n (y_i - \mu_i^{(k)})^2, \quad (\text{A.3})$$

where  $\mu_i^{(k)}$  denotes the fitted value of  $y$  at  $x_i$ , and  $\hat{\epsilon}_i^{(k)} = y_i - \hat{\mu}_i^{(k)}$  denotes the residual.

The entire procedure of the grid-search method consists of two main steps. Firstly, assuming  $k$  unknown crossover points fixed at  $\tau_1, \dots, \tau_k$ , we fit a least square regression model (*LS*) on the covariates  $\{x, (x - \tau_1)^+, \dots, (x - \tau_k)^+\}$ , and thus obtain a residual sum of squared error. Secondly, we examine all possible combinations of  $(\tau_1, \dots, \tau_k)$  to find out the genuine crossover points  $\tau_1, \dots, \tau_k$  with the minimum *SSE*. The confidence interval of the crossover points can be calculated as

$$S_m(\mathbf{X}) \leq \text{MinSSE} \times (1 + F_\alpha(k, p)), \quad (\text{A.4})$$

where  $S_m(\mathbf{X})$  is the minimum value of  $SSE$  at  $\mathbf{X}$ . In this expression,  $k$  is the number of crossover points,  $p$  is the dimension of the parameter space of the model, and  $F_\alpha(k, p)$  is the  $(1 - \alpha)$  percentile of an F-distribution with  $k$  and  $p$  degrees of freedom.

The grid-search technique is capable of finding the exact locations of the crossover points through the search of the best fit for the scaling-identification regression model. However, the method does not provide a means to determine whether one should reject the model under the null hypothesis and accept the model under the alternative hypothesis. Furthermore, it does not say whether the selected model is significant enough to represent the observed multi-scaling behavior under MF-DFA. In order to answer these two questions, the construction of an appropriate test statistic is thus crucial.

Obviously, it is reasonable to believe that there is significant difference between the two hypothesized model when the ratio of their residual sum of squares approaches 1. The test statistic is thus determined as (Kim et al., 2000):

$$T(\mathbf{y}) = \frac{[\hat{\varepsilon}^{(k_0)}(\mathbf{y})]'[\hat{\varepsilon}^{(k_0)}(\mathbf{y})]}{[\hat{\varepsilon}^{(k_1)}(\mathbf{y})]'[\hat{\varepsilon}^{(k_1)}(\mathbf{y})]}, \quad (\text{A.5})$$

where  $\hat{\varepsilon}^{(k_0)}(\mathbf{y})$  is the residuals vector of the null model and  $\hat{\varepsilon}^{(k_1)}(\mathbf{y})$  is that of the alternative model. Essentially,  $T(\mathbf{y})$  is a goodness-of-fit measure examining the difference between the two models.

However, how large (or small) should the value of  $T(\mathbf{y})$  be before one can say that there is a significant difference between the models? The distribution of the test statistic  $T(\mathbf{y})$  is the key to the answer of this question. Oftentimes, the exact distribution of  $T(\mathbf{y})$  is unclear. Permutation, a Monte Carlo method, can be used to construct a permutation distribution for the test statistic that approximates the property of its

sampling distribution when permutation resamples are large. We give a brief description of the procedure in Kim et al. (2000) as follows:

Firstly, we apply the grid-search method to fit the null model and obtain the residuals  $[\hat{\varepsilon}_1^{(k_0)}, \hat{\varepsilon}_2^{(k_0)}, \dots, \hat{\varepsilon}_n^{(k_0)}]$ . The procedure of permutation is repeated to ensure large enough resamples. For each process, we permute the residuals and add them back to the null-modeled means. Let  $\pi'_a = [\pi_{a1}, \dots, \pi_{an}]$  be an  $n \times 1$  vector of permutations of the integers from 1 to  $n$ . The permuted data set associated with  $\pi_a$  has the same covariates as the original data, and the permuted responses are of the form

$$y'_{(a)} = \hat{\mu}^{(k_0)'} + [\hat{\varepsilon}_1^{(k_0)}, \dots, \hat{\varepsilon}_n^{(k_0)}], \quad (\text{A.6})$$

where  $\hat{\mu}^{(k_0)'} = [\hat{\mu}_1^{(k_0)'}, \dots, \hat{\mu}_n^{(k_0)'}]$ . The next step is to calculate the test statistic  $T(y_{(a)})$  by fitting the null and alternative hypothesis models for each permuted series. The values of  $T(y_{(a)})$ ,  $a = 1, \dots, N_p - 1$  are therefore obtained by the repeated  $N_p - 1$  permutations.

The purpose of the permutation test in this study is to evaluate the significance of the difference between the hypothesized models. The  $p$ -value is then calculated by the following formula to depict how extreme  $T(y)$  is in the entire permutation values of  $T(y_{(a)})$ ,  $a = 1, \dots, N_p - 1$ ,

$$p = \frac{\# [T(y_{(a)}) \geq T(y)], a \in \{1, 2, \dots, N_p - 1\}}{N_p}, \quad (\text{A.7})$$

where  $N_p - 1$  is the number of permutations. By this statistical procedure, we can decide whether the two hypothesized models are different, and determine simultaneously the number of crossover points underlying the multi-scaling behavior. Kim et al. (2000) called it the approximate permutation test, because of the Monte Carlo nature of the calculation and some other factors. With regard to the issue of whether the selected model is significant enough to represent the multi-scaling

behavior under MF-DFA, conventional methods, such as T-test and normal test, are useful for testing each estimation of the parameters of the kinked curve.

---

□ **End of chapter.**

## Bibliography

- Adamek, T. and O'Connor, N. (2007). Using Dempster-Shafer theory to fuse multiple information sources in region-based segmentation. In *IEEE International Conference on Image Processing, 2007 (ICIP 2007)*, volume 2 of *IEEE International Conference on Image Processing, 2007 (ICIP 2007)*, pages II-269-II272.
- Adams, S. and Sandrock, C. (2010). Avian influenza: update. *Med Princ Pract*, 19:421-432.
- Adene, D. and Oguntad, A. (2006). *The structure and importance of the commercial and village based poultry industry in Nigeria*. FAO, Rome, Italy.
- Alexander, D. J. (2007a). An overview of the epidemiology of avian influenza. *Vaccine*, 25(30):5637 - 5644. 4th International Veterinary Vaccines and Diagnostics Conference, Oslo, 25-29 June 2006, 4th International Veterinary Vaccines and Diagnostics Conference.
- Alexander, D. J. (2007b). Summary of avian influenza activity in europe, asia, africa, and australasia, 2002-2006. *Avian Dis*, 51(1 Suppl):161-166.
- Alvarez-Ramirez, J., Alvarez, J., Dagdug, L., Rodriguez, E., and Echeverria, J. C. (2008). Long-term memory dynamics of continental

- and oceanic monthly temperatures in the recent 125 years. *Physica A: Statistical Mechanics and its Applications*, 387(14):3629 – 3640.
- Anh, V., Yu, Z.-G., and Wanliss, J. A. (2007). Analysis of global geomagnetic variability. *Nonlinear Proc Geoph*, 14(6):701–708.
- Anselin, L. (1995). Local indicators of spatial association. *Geogr Anal*, 27:93–115.
- Aubin, J.-T., Zebi, S., Balish, A., Banks, J., Bhat, N., Bright, R. A., Bronw, L., Buchy, P., Burguiere, A. M., Ian Chen, H., Cheng, P., Cox, N. J., Crosier, A., Curus, A., Deng, F. C. G., Desheva, J., Devaux, S., Diep, N. H., Donis, R. O., Douglas, A., Dowell, S. F., Dung, N. T., Edwards, L., Fukuda, K., Garten, R., Govorkova, E., Gregory, V., Hampson, A., Hanh, N. T. H., Harper, S., Hay, A., Hoffmann, E., Hulse, D., Masaki Inai, Itamura, S., Jadhao, S., Jeanin, P., Kang, C., Katz, J., Kim, J.-H., Klimov, A., Kuk Kwon, Y., Lee, C.-W., Lien, P. S., Li, Y., Lim, W., Lin, Y. P., Lindstrom, S., Loftin, L., Mabry, J., Mai, L. Q., Maines, T., Manuguerra, J.-C., Mase, M., Matsuoka, Y., McCarron, M., Medina, M.-J., Nguyen, D., Ninomiya, A., Obuchi, M., Odagiri, T., Peiris, M., Perdue, M. L., Reynes, J.-M., Roberson, J., Rousseaux, C., Saito, T., Sangkipom, S., Shaw, M., Simmerman, J. M., Slomka, M., Smith, C., Som, S., Spackman, E., Stohr, K., Suarez, D. L., Sung, H. W., Swayne, D. E., Tardy-Panit, M., Tashiro, M., Thawatsupha, P., Tumpey, T., Uyeki, T., Tu, P. V., vander Werf, S., Vong, S., Webby, R., Webster, R., Wood, J., Xu, X., Yi, G., and Zhang, W. (2005). Evolution of H5N1 avian influenza viruses in Asia. *Emerg Infect Dis*, 11(10):1515–1521.
- Avise, J. C., Arnold, J., R. M. Ball, J., Bermingham, E., Lamb, T., Neigel, J. E., Reed, C. A., and Saunders, N. C. (1987). Intraspecific

- phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annu Rev Ecol Syst*, 18:489–522.
- Bailey, T. and Gatrell, A. (1995). *Interactive spatial data analysis*. Longman, Harlow.
- Barot, S., Gignoux, J., and Menaut, J.-C. (1999). Demography of savanna palm tree: predictions from comprehensive spatial pattern analysis. *Ecology*, 80:1987–2005.
- Bartlett, M. S. (1964). The spectral analysis of two-dimensional point process. *Biometrika*, 51:299–311.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2006). GenBank. *Nucleic Acids Res*, 34:D16.
- Benton, M. J. and Ayala, F. J. (2003). Dating the tree of life. *Science*, 300:1698–1700.
- Besag, J. (1977). Comment on “modelling spatial patterns” by B.J. Ripley’s. *J. Roy. Statistical Society*, 39:193–195.
- Besag, J. and Newell, J. (1991). The detection of clusters in rare diseases. *J. Roy. Statistical Society*, 154:143–155.
- Botts, B. N. and Getis, A. (1988). *Point Pattern Analysis*, volume 8 of *Scientific geography series*. Sage publications, Inc.
- Brown, J. D. and Stallknecht, D. E. (2008). Wild bird surveillance for the avian influenza virus. *Methods Mol Biol*, 436:85–97.
- Bush, R. M., Bender, C. A., Subbarao, K., Cox, N. J., and Fitch, W. M. (1999). Predicting the evolution of human influenza A. *Science*, 286(5446):1921–1925.

- Butler, J. B., Lane, S. N., and Chandler, J. H. (2001). Characterization of the structure of river-bed gravels using two-dimensional fractal analysis. *Mathematical Geology*, 33:301-330.
- Carrel, M., Emch, M., Jobe, R., Moody, A., and Wan, X. (2010). Spatiotemporal structure of molecular evolution of H5N1 highly pathogenic avian influenza in Vietnam. *PLoS ONE*, 5:e86311-e86319.
- Cattoli, G., Monne, I., Fusaro, A., Joannis, T. M., Lombin, L. H., Aly, M. M., Arafa, A. S., Sturm-Ramirez, K. M., Couacy-Hymann, E., Awuni, J. A., Batawui, K. B., Awoume, K. A., Aplogan, G. L., Sow, A., Ngangnou, A. C., El Nasri Hamza, I. M., Gamatie, D., Dauphin, G., Domenech, J. M., and Capua, I. (2009). Highly pathogenic avian influenza virus subtype H5N1 in Africa: A comprehensive phylogenetic analysis and molecular characterization of isolates. *PLoS ONE*, 4(3):e4842.
- Cavalli-Sforza, L. L. and Edwards, A. W. F. (1967). Phylogenetic analysis: Models and estimation procedures. *Am J Hum Genet*, 19:233-257.
- Cecchi, G., Ilemobade, A., Brun, Y. L., Hogerwerf, L., and Slingenbergh, J. (2008). Agro-ecological features of the introduction and spread of the highly pathogenic avian influenza (HPAI) H5N1 in northern Nigeria. *Geospat Health*, 3(1):7-16.
- Chaichoune, K., Wiriyarat, W., Thitithanyanont, A., Phonarknguen, R., Sariya, L., Suwanpakdee, S., Noimor, T., Chatsurachai, S., Suriyaphol, P., Ungchusak, K., Ratanakorn, P., Webster, R. G., Thompson, M., Auewarakul, P., and Puthavathana, P. (2009). Indigenous sources of 2007-2008 H5N1 avian influenza outbreaks in Thailand. *J Gen Virol*, 90(1):216-222.

- Chen, H., Deng, G., Li, Z., Tian, G., Li, Y., Jiao, P., Zhang, L., Liu, Z., Webster, R. G., and Yu, K. (2004). The evolution of H5N1 influenza viruses in ducks in southern China. *Proc Natl Acad Sci U S A*, 101(28):10452-10457.
- Chen, H., Smith, G. J. D., Li, K. S., Wang, J., Fan, X. H., Rayner, J. M., Vijaykrishna, D., Zhang, J. X., Zhang, L. J., Guo, C. T., Cheung, C. L., Xu, K. M., Duan, L., Huang, K., Qin, K., Leung, Y. H. C., Wu, W. L., Lu, H. R., Chen, Y., Xia, N. S., Naipospos, T. S. P., Yuen, K. Y., Hassan, S. S., Bahri, S., Nguyen, T. D., Webster, R. G., Peiris, J. S. M., and Guan, Y. (2006). Establishment of multiple sublineages of H5N1 influenza virus in Asia: Implications for pandemic control. *Proc Natl Acad Sci U S A*, 103(8):2845-2850.
- Chen, H., Smith, G. J. D., Zhang, S. Y., Qin, K., Wang, J., Li, K. S., Webster, R. G., Peiris, J. S. M., and Guan, Y. (2005). H5N1 virus outbreak in migratory waterfowl. *Nature*, 436(7048):191-192.
- Chen, Z., Ivanov, P. C., Hu, K., and Stanley, H. E. (2002). Effect of nonstationarities on detrended fluctuation analysis. *Phys. Rev. E*, 65(4):41107-41123.
- Chotpitayasunondh, T., Ungchusak, K., Hanshaoworadul, W., Chunsuthiwat, S., Sawanpanyalert, P., and Kijphati, R. (2005). Human disease from influenza a (H5N1) Thailand. *Emerg Infect Dis*, 11:201-209.
- Clements, A., Pfeiffer, D., and Martin, V. (2006). Application of knowledge-driven spatial modelling approaches and uncertainty management to a study of Rift Valley fever in Africa. *Int J Health Geogr*, 5(1):57.

- Cliff, A. and Haggett, P. (1986). *Spatial Aspects of Influenza Epidemic*. London: Poin Limited.
- Cliff, A. and Ord, J. (1980). *Spatial Processes Models & Application*. Bros Limited, London.
- Cressie, N. A. (1991). *Statistics for spatial data*. John Wiley & Sons. Inc.
- Dawood, F., Jain, S., Finelli, L., Shaw, M., Lindstrom, S., Garten, R., Gubareva, L., Xu, X., Bridges, C., and Uyeki, T. (2009). Emergence of a novel swine-origin influenza A (H1N1) virus in humans. *N Engl J Med.*, 25:2605-2615.
- Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Ann Math Stat*, 38:325-339.
- Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Patterns*. New York: Oxford University Press Inc.
- Du, G. and Ning, X. (2008). Multifractal properties of Chinese stock market in Shanghai. *Physica A*, 387(1):261 - 269.
- Duan, L., Bahl, J., Smith, G. J. D., Wang, J., Vijaykrishna, D., Zhang, L. J., Zhang, J. X., Li, K. S., Fan, X. H., Cheung, C. L., Huang, K., Poon, L. L. M., Shortridge, K. F., Webster, R. G., Peiris, J. S. M., Chen, H., and Guan, Y. (2008). The development and genetic diversity of H5N1 influenza virus in China, 1996-2006. *Virology*, 380(2):243-254.
- Ducatez, M. F., Olinger, C. M., Owoade, A. A., De Landtsheer, S., Ammerlaan, W., Niesters, H. G. M. Osterhaus, A. D. M. E., Fouchier, R. A. M., and Muller, C. P. (2006). Multiple introductions of H5N1 in Nigeria. *Nature*, 442:37.

- Ducatez, M. F., Olinger, C. M., Owoade, A. A., Tarnagda, Z., Tahita, M. C., Sow, A., De Landtsheer, S., Ammerlaan, W., Ouedraogo, J. B., Osterhaus, A. D. M. E., Fouchier, R. A. M., and Muller, C. P. (2007). Molecular and antigenic evolution and geographical spread of h5n1 highly pathogenic avian influenza viruses in western Africa. *J Gen Virol*, 88(8):2297–2306.
- Edgar, R. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, 32:1792–1797.
- Edwards, A. and Cavalli-Sforza, L. (1964). *Phenetic and Phylogenetic Classification*. Systematics Assoc. Publ.
- Elliott, P., Cuzick, J., English, D., and Stern, R., editors (1996). *Geographical and Environmental Epidemiology – Methods for Small-Area Studies*. Oxford University Press.
- Elliott, P., Wakefield, J., Best, N., and Briggs, D. (2001). *Spatial Epidemiology: Methods and Applications*. Oxford University Press.
- Enserink, M. (2006). H5N1 moves into Africa, European Union, deepening global crisis. *Science*, 311(5763):932.
- Ertel, J. E. and Fowlkes, E. B. (1976). Some algorithms for linear spline and piecewise multiple linear regression. *Journal of the American Statistical Association*, 71(355):640–648.
- Fang, L.-Q., de Vlas, S. J., Liang, S., Looman, C. W. N., Gong, P., Xu, B., Yan, L., Yang, H., Richardus, J. H., and Cao, W.-C. (2008). Environmental factors contributing to the spread of H5N1 avian influenza in mainland China. *PLoS ONE*, 3(5):e2268.

- FAO (2010). Technical task force on avian influenza. Update on the avian influenza situation. <http://www.fao.org/avianflu/en/aidenews.html> (latest accessed on 5 August, 2010).
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recogn Lett*, 27:861–874.
- Feder, J. (1988). *Fractals*. Plenum Press.
- Feder, P. I. (1975). The log likelihood ratio in segmented regression. *The Annals of Statistics*, 3(1):84–97.
- Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol*, 17:368–376.
- Felsenstein, J. (1996). Inferring phylogenies from protein sequences by parsimony, distance and likelihood methods. *Meth Enzymol*, 266:418–427.
- Ferguson, N., Galvani, A., and Bush, R. (2003). Ecological and immunological determinants of influenza evolution. *Nature*, 422:428–433.
- Ge, E., Haining, R., Li, C., Yu, Z., Wayne, M., Chu, K., and Leung, Y. (2011). Using knowledge fusion to analyze avian influenza H5N1 in East and Southeast Asia (unpublished paper).
- Ge, E. and Leung, Y. (2011). The detection of crossover time scales in multifractal detrended fluctuation analysis (unpublished paper).
- Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Ineor Stat*, 5:115–145.

- Gehlke, C. and Biehl, H. (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *J Amer Statist Assoc*, 29:169-170.
- Getis, A. (1984). Interaction modeling using second-order analysis. *Environ Plann A*, 16:173-183.
- Getis, A. (1991). Spatial interaction and spatial autocorrelation: a cross product approach. *Environ Plann A*, 23:1269-1277.
- Getis, A. and Aldstadt, J. (2004). Constructing the spatial weights matrix using a local statistic. *Geogr Anal*, 36(2):90 - 104.
- Getis, A. and Franklin, J. (1987). Second-order neighborhood analysis of mapped point patterns. *Ecology*, 68:473-477.
- Getis, A. and Ord, J. (1992). The analysis of spatial association by use of distance statistics. *Geogr Anal*, 24:189-206.
- Getis, A. and Ord, J. K. (1996). *Local Spatial Statistics: An overview*, pages 261-277. Cambridge: Geoinformation International.
- Gibson, G. (1997a). Investigating mechanisms of spatiotemporal epidemic spread using stochastic models. *Am Phytopathol Soc*, 87:139-146.
- Gibson, G. (1997b). Markov chain monte carlo methods for fitting spatiotemporal stochastic models in plant epidemiology. *J. Roy. Statistical Society*, pages 215-233.
- Gilbert, M., Chaitaweesub, P., Parakamawongsa, T., Premashtira, S., Tiensin, T., and Kalpravidh, W. (2006a). Free-grazing ducks and highly pathogenic avian influenza, Thailand. *Emerg Infect Dis.*, 12:227-234.

- Gilbert, M., Xiao, X., Chaitaweesub, P., Kalpravidh, W., Premashthira, S., Boles, S., and Slingenbergh, J. (2007). Avian influenza, domestic ducks and rice agriculture in Thailand. *Agric Ecosyst Environ*, 119:409–415.
- Gilbert, M., Xiao, X., Domenech, J., Lubroth, J., Martin, V., and Slingenbergh, J. (2006b). Anatidae migration in the western Palearctic and spread of highly pathogenic avian influenza H5N1 virus. *Emerg Infect Dis*, 12(11):1650–1656.
- Gilbert, M., Xiao, X., Pfeiffer, D. U., Epprecht, M., Boles, S., Czarnecki, C., Chaitaweesub, P., Kalpravidh, W., Minh, P. Q., Otte, M. J., Martin, V., and Slingenbergh, J. (2008). Mapping H5N1 highly pathogenic avian influenza risk in southeast Asia. *Proc Natl Acad Sci U S A*, 105(12):4769–4774.
- Globig, A., Staubach, C., Beer, M., Kppen, U., Fiedler, W., Nieburg, M., Wilking, H., Starick, E., Teifke, J. P., Werner, O., Unger, F., Grund, C., Wolf, C., Roost, H., Feldhusen, F., Conraths, F. J., Mettenleiter, T. C., and Harder, T. C. (2009). Epidemiological and ornithological aspects of outbreaks of highly pathogenic avian influenza virus H5N1 of Asian lineage in wild birds in Germany, 2006 and 2007. *Transbound Emerg Dis*, 56(3):57–72.
- Grassberger, P. and Procaccia, I. (1983). Characterization of strange attractors. *Phys. Rev. Lett.*, 50(5):346–349.
- Green, M., Freedman, D., and Gordis, L. (2008). Reference guide on epidemiology.
- Guan, Y., Peiris, J. S. M., Lipatov, A. S., Ellis, T. M., Dyrting, K. C., Krauss, S., Zhang, L. J., Webster, R. G., and Shortridge, K. F.

- (2002). Emergence of multiple genotypes of H5N1 avian influenza viruses in Hong Kong SAR. 99(13):8950–8955.
- Guan, Y., Poon, L. L. M., Cheung, C. Y., Ellis, T. M., Lim, W., Lipatov, A. S., Chan, K. H., Sturm-Ramirez, K. M., Cheung, C. L., Leung, Y. H. C., Yuen, K. Y., Webster, R. G., and Peiris, J. S. M. (2004). H5N1 influenza: a protean pandemic threat. *Proc Natl Acad Sci U S A*, 101(21):8156–8161.
- Haase, P. (1995). Spatial pattern analysis in ecology based on Ripley's K-Function: introduction and methods of edge correction. *J VEG SCI*, 6:575–582.
- Haeckel, E. H. P. (1866). *Generelle morphologie der organismen. Allgemeine grundzge der organischen formen-wissenschaft, mechanisch begrndet durch die von Charles Darwin reformirte descendenztheorie.* Berlin G. Reimer.
- Haggett, P., A. C. and A.E.Frey (1977). *Local Analysis in Human Geography.* London:Edward Arnold.
- Haining, R. (1991). Bivariate correlation with spatial data. *Geogr Anal*, 23:210–27.
- Haining, R. (2003). *Spatial Data Analysis: Theory and Practice.* Cambridge University Press.
- Halsey, T. C., Jensen, M. H., Kadanoff, L. P., Procaccia, I., and Shraiman, B. I. (1986). Fractal measures and their singularities: The characterization of strange sets. *Phys. Rev. A*, 33(2):1141–1151.
- Hanely, J. and McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143:29–36.

- He, F. and Duncan, R. P. (2000). Density-dependent effects on tree survival in an old-growth douglars fire forest. *Acta Oecol*, 88:676–688.
- Howe, G. (1977). *A world geography of human disease*. Academic Press, New York.
- Hu, K., Ivanov, P. C., Chen, Z., Carpena, P., and Stanley, H. E. (2001a). Effect of trends on detrended fluctuation analysis. *Phys Rev E Stat Nonlin Soft Matter Phys*, 64(1 Pt 1):1–19.
- Hu, K., Ivanov, P. C., Chen, Z., Carpena, P., and Stanley, H. E. (2001b). Effect of trends on detrended fluctuation analysis. *Phys Rev E Stat Nonlin Soft Matter Phys*, 64(1 Pt 1):11114.1–11114.19.
- Hulse-Post, D. J., Sturm-Ramirez, K. M., Humberd, J., Seiler, P., Govorkova, E. A., Krauss, S., Scholtissek, C., Puthavathana, P., Buranathai, C., Nguyen, T. D., Long, H. T., Naipospos, T. S. P., Chen, H., Ellis, T. M., Guan, Y., Peiris, J. S. M., and Webster, R. G. (2005). Role of domestic ducks in the propagation and biological evolution of highly pathogenic H5N1 influenza viruses in Asia. *Proc Natl Acad Sci U S A*, 102(30):10682–10687.
- Hurst, H. (1951). Long term storage capacity of reservoirs. *T Ame Soc Civ Eng*, 116.
- Ito, T., Okazaki, K., Kawaoka, Y., Takeda, A., Webster, R., and Kida, H. (1995). Prepetuation of influenza A viruses in Alaskan waterfowl reservoirs. *Arch Virol.*, 140:1163–1172.
- Jones, A. P., Langford, I. H., and Bentham, G. (1996). The application of K-function analysis to the geographical distribution of road traffic accident outcomes in Norfolk, England. *Soc Sci Med.*, 42(6):879 – 885.

- Jones, L., Beynon, M. J., Holt, C. A., and Roy, S. (2006). An application of the Dempster-Shafer theory of evidence to the classification of knee function and detection of improvement due to total knee replacement surgery. *J Biomec*, 39(13):2512–2520.
- Kaftandjian, V., Dupuis, O., Babot, D., and Zhu, Y. M. (2003). Uncertainty modelling using Dempster-Shafer theory for improving detection of weld defects. *Pattern Recogn Lett*, 24(1-3):547–564.
- Kandeel, A., Manoncourt, S., Kareem, E., Ahmed, A., EI-Refaie, S., Essmat, H., Tjaden, J., Mattos, C., K.C.Earhart, Marfin, A., and EI-Sayed, N. (2010). Zoonotic transmission of avian influenza virus (H5N1), Egypt, 2006–2009. *Emerg Infect Dis*, 16:1101–1107.
- Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H. H. A., Havlin, S., and Bunde, A. (2001). Detecting long-range correlations with detrended fluctuation analysis. *Physica A: Statistical Mechanics and its Applications*, 295(3-4):441 – 454.
- Kantelhardt, J. W., Rybski, D., Zschiegner, S. A., Braun, P., Koscielny-Bunde, E., Livina, V., Havlin, S., and Bunde, A. (2003). Multifractality of river runoff and precipitation: comparison of fluctuation analysis and wavelet methods. *Physica A: Statistical Mechanics and its Applications*, 330(1-2):240 – 245.
- Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Bunde, A., Havlin, S., and Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A*, 316:87.
- Kapan, D. D., Bennett, S. N., Ellis, B. N., Fox, J., Lewis, N. D., Spencer, J. H., Saksena, S., and Wilcox, B. A. (2006). Avina in-

- fluenza (H5N1) and the evolutionary and social ecology of infectious disease emergence. *EcoHealth*, 3:187–194.
- Kilpatrick, A. M., Chmura, A. A., Gibbons, D. W., Fleischer, R. C., Marra, P. P., and Daszak, P. (2006). Predicting the global spread of H5N1 avian influenza. *Proc Natl Acad Sci U S A*, 103(51):19368–19373.
- Kim, H., Fay, M. P., Yu, B., Barrett, M. J., and Feuer, E. J. (2004). Comparability of segmented line regression models. *Biometrics*, 60(4):1005–1014.
- Kim, H.-J., Fay, M. P., Feuer, E. J., and Midthune, D. N. (2000). Permutation tests for joinpoint regression with applications to cancer rate. *Statistics in Medicine*, 19:335–351.
- Krauss, S., Obert, C. A., Franks, J., Walker, D., Jones, K., Seiler, P., Niles, L., Pryor, S. P., Obenauer, J. C., Naeve, C. W., Widjaja, L., Webby, R. J., and Webster, R. G. (2007). Influenza in migratory birds and evidence of limited intercontinental virus exchange. *PLoS Pathog*, 3(11):e167.
- Krauss, S., Walker, D., Pryor, S., Niles, L., Chenghong, L., Hinshaw, V., and Webster, R. (2004). Influenza A viruses of migrating wild aquatic birds in North America. *Vector Borne Zoonotic Dis*, 4:177–189.
- Lai, P.-C., So, F.-M., and Chan, K.-W. (2009). *Spatial Epidemiological Approaches in Disease Mapping and Analysis*. CRC Press.
- Lam, N. s.-n. and Lee, D. C. (1993). *Fractals in Geography*. PTR Prentice-Hall, Inc.

- Lam, S. and Quattrochib, D. (1992). On the issues of scale, resolution, and fractal analysis in the mapping sciences. *Prof Geogr*, 44:88–98.
- Lamb, R. and Krug, R. (2001). *Orthomyxoviridae: The viruses and their replication*. Lippincott Williams & Wilkins.
- Lawson, A. B. (2001). *Statistical methods in spatial epidemiology*. John Wiley.
- Lebarbenchon, C., Albespy, F., Brochet, A.-L., Grandhomme, V., Renaud, F., Fritz, H., Green, A. J., Thomas, F., van der Werf, S., Aubry, P., Guillemain, M., and Gauthier-Clerc, M. (2009). Spread of avian influenza viruses by common teal (*Anas crecca*) in Europe. *PLoS ONE*, 4(10):e7289.
- Lerman, P. (1980). Fitting segmented regression models by grid search. *Appl. Statist.*, 29:77–84.
- Leung, Y., Ge, E., and Yu, Z. (2011). Temporal scaling behavior of avian influenza A (H5N1) – the multifractal detrended fluctuation analysis. *Annals of the Association of American Geographers (in press)*.
- Leung, Y. and Mei, C.-L. (2003). Statistical test for local patterns of spatial association. *Environ Plann A*, 35:725–755.
- Li, J., Wang, J., Wu, C., Yang, Y., Ji, Z., and Wang, H. (2006). Spatial and temporal analysis of the effects of temperature on the spread of avian influenza. *Chin. J. Prev. Vet. Med.*, 28:408–411.
- Li, K. S., Guan, Y., Wang, J., Smith, G. J. D., Xu, K. M., Duan, L., Rahardjo, A. P., Puthavathana, P., Buranathai, C., Nguyen, T. D., Estoepongstie, A. T. S., Chaisingh, A., Auewarakul, P., Long, H. T.,

- Hanh, N. T. H., Webby, R. J., Poon, L. L. M., Chen, H., Shortridge, K. F., Yuen, K. Y., Webster, R. G., and Peiris, J. S. M. (2004). Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature*, 430(6996):209–213.
- Liang, L., Xu, B., Chen, Y., Liu, Y., Cao, W., Fang, L., Feng, L., Goodchild, M. F., and Gong, P. (2010). Combining spatial-temporal and phylogenetic analysis approaches for improved understanding on global H5N1 transmission. *PLoS ONE*, 5(10):e13575.
- Ligon, B. L. (2005). Avian influenza virus H5N1: A review of its history and information regarding its potential to cause the next pandemic. *Pediatr. Infect. Dis.*, 16:326–335.
- Lim, G., Kim, S., Lee, H., Kim, K., and Lee, D.-I. (2007). Multi-fractal detrended fluctuation analysis of derivative and spot markets. *Physica A: Statistical Mechanics and its Applications* *Physica A*, 386(1):259 – 266.
- Lipatov, A. C., Smirnov, I. A., Kaverin, N. V., and Webster, R. G. (2005). Evolution of avian influenza viruses H5N1 (1997-2004) in southern and south-eastern Asia. *Vopr Virusol*, 50(4):11–17.
- Lipatov, A. S., Govorkova, E. A., Webby, R. J., Ozaki, H., Peiris, M., Guan, Y., Poon, L., and Webster, R. G. (2004). Influenza: Emergence and control. *J. Virol.*, 78(17):8951–8959.
- Liu, C.-M., Lin, S.-H., Chen, Y.-C., Lin, K. C.-M., Wu, T.-S. J., and King, C.-C. (2007). Temperature drops and the onset of severe avian influenza A H5N1 virus outbreaks. *PLoS ONE*, 2(2):e191.
- Liu, J., Xiao, H., Lei, F., Zhu, Q., Qin, K., Zhang, X.-W., Zhang, X.-L., Zhao, D., Wang, G., Feng, Y., Ma, J., Liu, W., Wang, J., and

- Gao, G. F. (2005). Highly pathogenic H5N1 influenza virus infection in migratory birds. *Science*, 309(5738):1206.
- Livina, V., Kizner, Z., Braun, P., Molnar, T., Bunde, A., and Havlin, S. (2007). Temporal scaling comparison of real hydrological data and model runoff records. *Journal of Hydrology*, 336(1-2):186 – 198.
- Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage publications, Inc.
- Luo, W. B. and Caselton, B. (1997). Using Dempster-Shafer theory to represent climate change uncertainties. *J Environ Manage*, 49(1):73–93.
- Malpica, J., Alonso, M., and Sanz, M. (2007). Dempster-Shafer theory in geographic information systems: A survey. *Expert Syst Appl*, 32(1):47–55.
- Mandelbrot, B. (1983). *The fractal geometry of Nature*. New York: W.H. Freeman.
- Martinez, M., Muoz, M. J., Torre, A. D. L., Iglesias, I., Peris, S., Infante, O., and Sanchez-Vizcaino, J. M. (2009). Risk of introduction of H5N1 HPAI from Europe to Spain by wild water birds in autumn. *Transbound Emerg Dis*, 56(3):86–98.
- May, J. (1954). *Medical Geography*. In P.E. James & C.F. Jones.
- May, J. M. (1977). Medical geography: its methods and objectives. *Soc Sci Med.*, 11:715–730.
- Meade, M. S. and Earickson, R. J. (2000). *Medical Geography*. The Guilford Press.

- Meier, R., Kwong, S. Vaidva, G., and Ng, P. (2006). DNA barcoding and taxonomy in Diptera: A tale of high intraspecific and low identification success. *Syst Biol*, 55:715–728.
- Montello, D. R. and Golledge, R. G. (1998). Scale and detail in the cognition of geographic information. Technical report, California University of Santa Barbara.
- Moone, I., Joannis, T. M., Fusaro, A., Benedicis, P. D., Lombin, L. H., Ularanu, H., Egbuji, A., Solomon, P., Obi, T. U., Cattoli, G., and Capua, I. (2008). Reassortant avian influenza virus (H5N1) in poultry, Nigeria, 2007. *Emerg Infect Dis*, 14:637–640.
- Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37:17–33.
- Mounts, A., Kwong, H., Izurieta, H., Ho, Y., Au, T., Lee, M., Bridges, C., Seymour, W., Kowk, H., Katz, J., Thompson, W., Cox, N., and Fukuda, K. (1999). Case-control study of risk factors for avian influenza a (H5N1) disease, Hong Kong 1997. *J. Infect Dis.*, 180:505–508.
- Movahed, M. S., Jafari, G. R., Ghasemi, F., Rahvar, S., and Tabar, M. R. R. (2006a). Multifractal detrended fluctuation analysis of sunspot time series. *J Stat Mech Theor Exp*, 2006(02):1–17.
- Movahed, M. S., Jafari, G. R., Ghasemi, F., Rahvar, S., and Tabar, M. R. R. (2006b). Multifractal detrended fluctuation analysis of sunspot time series. *Journal of Statistical Mechanics: Theory and Experiment*, 2006:1–17.
- Muggeo, V. M. R. (2003). Estimating regression models with unknown break-points. *Stat Med*, 22(19):3055–3071.

- Munasinghe, S., Brown, G., Pereira, A., Keeble, B., Nair, P., and Sundkvis, T. (2008). Public health response to an avian influenza influenza (h5n1) poultry outbreak in Suffolk, United Kingdom, in November 2007. *Eurosurveillance*, 13:1-3.
- Munster, V. J., Baas, C., Lexmond, P., Waldenström, J., Wallensten, A., Fransson, T., Rimmelzwaan, G. F., Beyer, W. E. P., Schutten, M., Olsen, B., Osterhaus, A. D. M. E., and Fouchier, R. A. M. (2007). Spatial, temporal, and species variation in prevalence of influenza A viruses in wild migratory birds. *PLoS Pathog*, 3(5):636-638.
- Nguyen, T. D., Nguyen, T. V., Vijaykrishna, D., Webster, R. G., Guan, Y., Peiris, J. M., and Smith, G. J. (2008). Multiple sublineages of avian influenza A virus (H5N1), Vietnam, 2005-2007. *Emerg Infect Dis*, 14:632-636.
- Niu, M., Wang, F., Liang, Q., Yu, G., and Yu, Z. (2008). Multifractal detrended fluctuation analysis of pressure fluctuation signals in an impinging entrained-flow gasifier. *CHEM ENG J*, 136(2-3):364-372.
- Nixon, K. C. and Carpenter, J. M. (1993). On outgroups. *Cladistics*, 9:413-426.
- Normile, D. (2005a). Are wild birds to blame. *Science*, 310:426-428.
- Normile, D. (2005b). Avian influenza. Europe scrambles to control deadly H5N1 strain. *Science*, 310(5747):417.
- Normile, D. (2006a). Evidence points to migratory birds in H5N1 spread. *Science*, 311:1225.
- Normile, D. (2006b). Wild birds only partly to blame in spreading H5N1. *Science*, 312:1415.

- OIE (2010). Update on highly pathogenic avian influenza in animals ([http://www.oie.int/downld/avian%20influenza/a\\_ai-asia.htm](http://www.oie.int/downld/avian%20influenza/a_ai-asia.htm)).
- Okabe, A. and Yamada (2001). The K-function method on a network and its computational implementation. *Geogr Anal*, 3:271-290.
- Okazaki, K., Takada, A., Ito, T., Imai, M., Takakuwa, H., Hatta, M., Ozaki, H., Tanizaki, T., Nagano, T., Ninomiya, A., Demenev, V., Tyaptirganov, M., Karatayeva, T., Yamnikova, S., Lvov, D., and Kida, H. (2000). Precursor genes of future pandemic influenza viruses are perpetuated in ducks nesting in Siberia. *Arch Virol.*, 145:885-893.
- Olsen, B., Munster, V. J., Wallensten, A., Waldenström, J., Osterhaus, A. D. M. E., and Fouchier, R. A. M. (2006). Global patterns of influenza A virus in wild birds. *Science*, 312(5772):384-388.
- Openshaw, S. (1977). A geographical solution to scale and aggregation problems in region building, partitioning and spatial modeling. *T I BRIT GEOGR*, 2:459-472.
- Openshaw, S. (1984). *The Modifiable Areal Unit Problem*. Norwich: Geo Books. Geo Books. Norwich-Printed by Headley Brothers Ltd. Kent.
- Ord, J. and Getis, A. (1995). Local spatial autocorrelation statistics: distribution issue and an application. *Geogr Anal*, 27:286-306.
- Ozawa, Y., Makino, S., Park, J., Chang, J., Kim, J., and An, S. (2006). A review of recent unexpected animal disease events in Japan and Korea and the follow-up action taken. *Rev. sci. tech. Off. int. Epiz.*, 25:125-135.
- Panigrahi, S., Kundu, A., Sural, S., and Majumdar, A. (2009). Credit card fraud detection: A fusion approach using Dempster-Shafer the-

- ory and Bayesian learning. *Inform Fusion*, 10(4):354–363. Special Issue on Information Fusion in Computer Security.
- Peng, C. K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., and Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics*, 49(2):1685–1689.
- Penny, D., Hendy, M. D., and Steel, M. A. (1992). Progress with methods for constructing evolutionary trees. *Trends Ecol Evol*, 7(3):73–79.
- Perkins, L. and Swayne, D. (2002). Pathogenicity of a Hong Kong-origin H5N1 highly pathogenicity of a highly pathogenic avian influenza virus for emus, geese, ducks, and pigeons. *Avian Dis*, 46:53–63.
- Pfeiffer, D. U., Minh, P. Q., Martin, V., Epprecht, M., and Otte, M. J. (2007). An analysis of the spatial and temporal patterns of highly pathogenic avian influenza occurrence in Vietnam using national surveillance data. *Vet J*, 174(2):302–309.
- Pfeiffer, D. U., Robinson, T. P., Stevenson, M., Stevens, K. B., Rogers, D. J., and Clements, A. C. A. (2008). *Spatial Analysis in Epidemiology*. Oxford University Press.
- Posada, D. and Crandall, K. (1998). MODELTEST: Testing the model of DNA substitution. *Bioinformatics*, 14:817–818.
- Potvin, F., Boots, B., and Dempster, A. (2003). Comparison among three approaches to evaluate winter habitat selection by white-tailed deer on Anticosti island using occurrences from an aerial survey and forest vegetation maps. *Can. J. Zool.*, 81:1662–1670.
- Preiser, K. H. (2006). *Influenza Report 2006*. Flying Publish.

- Proenca-Modena, J. L., Macedo, I. S., and Arruda, E. (2007). H5N1 avian influenza virus: an overview. *Braz J Infect Dis*, 11:125–133.
- Raisz, E. (1938). *General cartography*. New York, London : McGraw-Hill Book Company, inc.
- Ripley, B. (1976). The second-order analysis of stationary point processes. *J Appl Probab*, 13:255–266.
- Robinson, A. H. (1978). *Elements of Cartography*. New York [etc.] : Wiley.
- Rogers, S., Starmer, W., and Castello, J. (2004). Recycling of pathogenic microbes through survival in ice. *Med Hypothesis*, 63:773–777.
- Saitou, N. and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Bio Evol*, 4:406–425.
- Salzberg, S. L., Kingsford, C., Cattoli, G., Spiro, D. J., Janies, D. A., Aly, M. M., Brown, I. H., Couacy-Hymann, E., Mia, G. M. D., Dung, D. H., Guercio, A., Joannis, T., Ali, A. S. M., Osmani, A., Padalino, I., Saad, M. D., Savi, V., Sengamalay, N. A., Yingst, S., Zaborsky, J., Zorman-Rojs, O., Ghedin, E., and Capua, I. (2007). Genome analysis linking recent European and African influenza (H5N1) viruses. *Emerg Infect Dis*, 13(5):713–718.
- Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. Chapman & Hall/CRC.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press.

- Shang, P., Lu, Y., and Kamae, S. (2008). Detecting long-range correlations of traffic time series with multifractal detrended fluctuation analysis. *Chaos soliton fract*, 36(1):82 – 90.
- Si, Y., Debba, P., Skidmore, A., Toxopeus, A., and Li, L. (2008). Spatial and temporal patterns of global H5N1 outbreaks. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.
- Si, Y., Skidmore, A. K., Wang, T., de Boer, W. F., Debba, P., Toxopeus, A. G., Li, L., and Prins, H. H. (2009). Spatio-temporal dynamics of global H5N1 outbreaks match bird migration patterns. *Geospatial Health*, 4:65–78.
- Smallman-Raynor, M. and Cliff, A. D. (2008). The geographical spread of avian influenza A (H5N1): Panzootic transmission (December 2003-May 2006), pandemic potential, and implications. *Ann Assoc Am Geogr*, 98:553–582.
- Smelser, N. J., Wright, J. D., and Baltes, P. B. (2001). *International Encyclopedia of Social & Behavioral Sciences*. Oxford: Pergamon Press.
- Smith, D. J., Lapedes, A. S., de Jong, J. C., Bestebroer, T. M., Rimmelzwaan, G. F., D.M.E.Osterhaus, A., and Fouchier, R. A. M. (2004). Mapping the antigenic and genetic evolution of influenza virus. *Science*, 305:371–376.
- Smith, G. J. D., Fan, X. H., Wang, J., Li, K. S., Qin, K., Zhang, J. X., Vijaykrishna, D., Cheung, C. L., Huang, K., Rayner, J. M., Peiris, J. S. M., Chen, H., Webster, R. G., and Guan, Y. (2006a). Emergence

- and predominance of an H5N1 influenza variant in China. *Proc Natl Acad Sci U S A*, 103(45):16936–16941.
- Smith, G. J. D., Naipospos, T. S. P., Nguyen, T. D., de Jong, M. D., Vijaykrishna, D., Usman, T. B., Hassan, S. S., Nguyen, T. V., Dao, T. V., Bui, N. A., Leung, Y. H. C., Cheung, C. L., Rayner, J. M., Zhang, J. X., Zhang, L. J., Poon, L. L. M., Li, K. S., Nguyen, V. C., Hien, T. T., Farrar, J., Webster, R. G., Chen, H., Peiris, J. S. M., and Guan, Y. (2006b). Evolution and adaptation of H5N1 influenza virus in avian and human hosts in Indonesia and Vietnam. *Virology*, 350(2):258–268.
- Suwannakarn, K., Amonsin, A., Sasipreeyajan, J., Kitikoon, P., Tantilertcharoen, R., Parchariyanon, S., Chaisingh, A., Nuansrichay, B., Songserm, T., Theamboonlers, A., and Poovorawan, Y. (2009). Molecular evolution of H5N1 in Thailand between 2004 and 2008. *Infect Genet Evol*, 9(5):896–902.
- Swayne, D. and Suarez, D. (2000). Highly pathogenic avian influenza. *Rev Sci Tech*, 19:463–482.
- Swofford, D. (2002). *PAUP\*: Phylogenetic Analysis Using Parsimony (\* and other methods)*. Sinauer Associates, Sunderland, MA., 4 edition.
- Szelezcky, Z., D., ., Ursu, K., a Ivanics, Kiss, I., Erdyi, K., Bel, S., Muller, C. P., Brown, I. H., and Bint, . (2009). Four different sub-lineages of highly pathogenic avian influenza H5N1 introduced in Hungary in 2006–2007. *Vet Microbiol*, 139(1–2):24 – 33.
- Takeda, M., Leser, G. P., Russell, C. J., and Lamb, R. A. (2003). Influenza virus hemagglutinin concentrates in lipid raft microdomains

- for efficient viral fusion. *Proc Natl Acad Sci U S A*, 100(25):14610–14617.
- Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Bio Evol*, 24:1596–1599.
- Taubenberger, J., Reid, A., Lourens, R., Wang, R., Jin, G., and Fanning, T. (2005). Characterization of the 1918 influenza virus polymerase genes. *Nature*, 437:889–893.
- Thomas, R. (1990). *Spatial Epidemiology*. London papers in regional science 21 a poin publication.
- Tiensen, T., Nielen, M., Songserm, T., Kalpravidh, W., Chaitaweesub, P., Amonsin, A., Chotiprasatintara, S., Chaisingh, A., Damrongwatanapokin, S., Wongkasemjit, S., Antarasena, C., Songkitti, V., Chanachai, K., Thanapongtham, W., and Stegeman, J. A. (2007). Geographic and temporal distribution of highly pathogenic avian influenza A virus (H5N1) in Thailand, 2004-2005: an overview. *Avian Dis*, 51(1 Suppl):182–188.
- Tiwari, R. C., Cronin, K. A., Davis, W., Feuer, E. J., Yu, B., and Chib, S. (2005). Bayesian model selection for join point regression with application to age-adjusted cancer rates. *Journal Of The Royal Statistical Society Series C*, 54(5):919–939.
- Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Econ Geogr*, 46:234–240.
- Uchida, Y., Chaichoune, K., Wiriyarat, W., Watanabe, C., Hayashi, T., Patchimasiri, T., Nuansrichay, B., Parchariyanon, S., Okamatsu, M., Tsukamoto, K., Takemae, N., Ratanakorn, P., Yamaguchi, S.,

- and Saito, T. (2008). Molecular epidemiological analysis of highly pathogenic avian influenza H5N1 subtype isolated from poultry and wild bird in Thailand. *Virus Res*, 138(1-2):70-80.
- Ungchusak, K., Auewarakul, P., Dowell, S. F., Kitphati, R., Auwanit, W., Puthavathana, P., Uiprasertkul, M., Boonnak, K., Pittayawonganon, C., Cox, N. J., Zaki, S. R., Thawatsupha, P., Chittaganpitch, M., Khontong, R., Simmerman, J. M., and Chunsutthiwat, S. (2005). Probable person-to-person transmission of avian influenza A (H5N1). *N Engl J Med*, 352(4):333-340.
- USGS (2010). Migratory Bird Flyways: Potential Pathways for Distribution of Avian Influenza.  
([http://alaska.usgs.gov/science/biology/avian\\_influenza/flyways.html](http://alaska.usgs.gov/science/biology/avian_influenza/flyways.html))  
(lastest accessed on 22 march, 2010).
- Viboud, C., Bjornstad, O. N., Smith, D. L., Simonsen, L., Miller, M. A., and Grenfell, B. T. (2006). Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science*, 312(5772):447-451.
- V.V. Anh, K.C. Lam, Y. L. Q. T. (2000). Multifractal analysis of Hong Kong air quality data. *Environmetrics*, 11:139-149.
- Wallace, R. G., Hodac, H., Lathrop, R. H., and Fitch, W. M. (2007). A statistical phylogeography of influenza A H5N1. *Proc Natl Acad Sci U S A*, 104(11):4473-4478.
- Wan, X.-F., Nguyen, T., Davis, C. T., Smith, C. B., Zhao, Z.-M., Carrel, M., Inui, K., Do, H. T., Mai, D. T., Jadhao, S., Balish, A., Shu, B., Luo, F., Emch, M., Matsuoka, Y., Lindstrom, S. E., Cox, N. J., Nguyen, C. V., Klimov, A., and Donis, R. O. (2008). Evolution of

- highly pathogenic H5N1 avian influenza viruses in Vietnam between 2001 and 2007. *PLoS ONE*, 3(10):e3462.
- Wang, G., Zhan, D., Li, L., Lei, F., Liu, B., Liu, D., Xiao, H., Feng, Y., Li, J., Yang, B., Yin, Z., Song, X., Zhu, X., Cong, Y., Pu, J., Wang, J., Liu, J., Gao, G. F., and Zhu, Q. (2008a). H5N1 avian influenza re-emergence of Lake Qinghai: phylogenetic and antigenic analyses of the newly isolated viruses and roles of migratory birds in virus circulation. *J Gen Virol*, 89(3):697–702.
- Wang, H., Feng, Z., Shu, Y., Yu, H., Zhou, L., Zu, R., Huai, Y., Dong, J., Bao, C., Wen, L., Wang, H., Yang, P., Zhao, W., Dong, L., Zhou, M., Liao, Q., Yang, H., Wang, M., Lu, X., Shi, Z., Wang, W., Gu, L., Zhu, F., Li, Q., Yin, W., Yang, W., Li, D., Uyeki, T. M., and Wang, Y. (2008b). Probable limited person-to-person transmission of highly pathogenic avian influenza A (H5N1) virus in China. *Lancet*, 371(9622):1427–1434.
- Webster, R., Bean, W., Gorman, O., Chambers, T., and Kawaoka, Y. (1992). Evolution and ecology of influenza A viruses. *Microbiol Rev*, 56:152–179.
- Webster, R. G., Peiris, M., Chen, H., and Guan, Y. (2006). H5N1 outbreaks and enzootic influenza. *Emerg Infect Dis*, 12(1):3–8.
- WHO (2004). Estimating the impact of the next influenza pandemic: enhancing preparedness [<http://www.who.int>] (lastest accessed on 4 august, 2010).
- WHO (2005). Avian influenza: assessing the pandemic threat.
- WHO (2006a). Avian influenza (“bird flu”) ([http://www.who.int/mediacentre/factsheets/avian\\_influenza/en/](http://www.who.int/mediacentre/factsheets/avian_influenza/en/)).

- WHO (2006b). Weekly epidemiological record; epidemiology of WHO-confined human cases of avian influenza A (H5N1) infection.
- WHO (2010a). Cumulative number of confirmed human cases of avian influenza A/(H5N1)  
[http://www.who.int/csr/disease/avian\\_influenza/country/cases\\_table\\_2010\\_08\\_31/en/index.html](http://www.who.int/csr/disease/avian_influenza/country/cases_table_2010_08_31/en/index.html) (latest accessed on 17 September, 2010).
- WHO (2010b). H5N1 avian influenza: Timeline of major events. 04 January 2010. Available from: [http://www.who.int/csr/disease/avian\\_influenza/timeline\\_10\\_01\\_04.pdf](http://www.who.int/csr/disease/avian_influenza/timeline_10_01_04.pdf).
- Williams, R. and Peterson, A. T. (2009). Ecology and geography of avian influenza (HPAI H5N1) transmission in the Middle East and northeastern Africa. *Int J Health Geogr*, 8(1):47.
- Woese, C. R. (2002). On the evolution of cells. *Proc Natl Acad Sci U S A*, 99:8742–8747.
- WorldBank (2009). Economic Impact of Avian Flu [<http://web.worldbank.org>] (last accessed on 9 Apr. 2010) .
- Wu, W. L., Chen, Y., Wang, P., Song, W., Lau, S.-Y., Rayner, J. M., Smith, G. J. D., Webster, R. G., Peiris, J. S. M., Lin, T., Xia, N., Guan, Y., and Chen, H. (2008). Antigenic profile of avian H5N1 viruses in Asia from 2002 to 2007. *J. Virol.*, 82(4):1798–1807.
- Xu, X., Subbarao, K., Cox, N. J., and Guo, Y. (1999). Genetic characterization of the pathogenic influenza A/Goose/Guangdong/1/96 (H5N1) virus: Similarity of its Hemagglutinin gene to those of H5N1 viruses from the 1997 outbreaks in Hong Kong. *Virology*, 261(1):15–19.

- Yamada, I. and Rogerson, P. A. (2003). An empirical comparison of edge effect correction methods applied to K-function analysis. *Geographical Analysis*, 35:97–109.
- Yamada, I. and Thill, J.-C. (2007). Local indicators of network-constrained clusters in spatial point patterns. *Geogr Anal*, 39:268–292.
- Yee, K. S., Carpenter, T. E., and Cardona, C. J. (2009). Epidemiology of H5N1 avian influenza. *Comp Immunol Microbiol Infect Dis*, 32(4):325 – 340. Avian Influenza.
- Yu, B., Barrett, M. J., Kim, H.-J., and Feuer, E. J. (2007). Estimating joinpoints in continuous time scale for multiple change-point models. *Computational Statistics & Data Analysis*, 51(5):2420–2427.
- Yu, Z.-G., Anh, V., and Eastes, R. (2009). Multifractal analysis of geomagnetic storm and solar flare indices and their class dependence. *J. Geophys. Res.*, 114.
- Yu, Z.-G., Leung, Y., Chen, Y. D., Zhang, Q., and Anh, V. (2010). Multifractal analysis of daily rainfall in the Pearl River basin of China. *Water Resour Res (revised version)*.