Spatial Analysis of the 2008 Influenza Outbreak of Hong Kong

Poh-chin Lai and Kim-hung Kwong

Department of Geography, The University of Hong Kong
Pokfulam Road, Hong Kong SAR
pclai@hkucc.hku.hk

Abstract. The deaths of three children amid a series of recent influenza outbreaks in early March 2008 resulted in the immediate shut down of all kindergartens and primary schools in Hong Kong. While many parents welcome the decision, others queried the judgment given that citizens lack sufficient information to evaluate whether there is an outbreak and must follow actions prescribed by the government. We demonstrated in this paper various techniques to visualize disease distribution and present outbreak data for public consumption. Our analyses made use of affected (case) and non-affected (control) schools with influenza cases in March 2008. A series of maps were created to show disease spread and concentration by means of standard deviational ellipses, grid-based spatial autocorrelation, and kernel density. The generalized data did not permit statistical analysis other than the nearest neighbor distance. We also made suggestions about requirements of additional data and possible directions of disease analysis.

Keywords: GIS, spatial autocorrelation, kernel density mapping, standard deviation ellipse, nearest neighbor distance analysis.

1 Introduction

Following a series of flu outbreaks\(^1\) at schools, a hospital and a nursing home for the elderly since 6 March 2008, the government of the Hong Kong Special Administrative Region (HKSAR) announced on 13 March 2008 to suspend classes of all kindergartens, child care centers, and primary schools for two weeks. The announcement came about after three children were allegedly suspected to die from complications arising from influenza A (H1N1 and H3) [1] [2]. While medical experts reported that the flu strain was not more virulent, the measure to help reduce infections and calm

\(^1\) According to the “Guidelines on Prevention of Communicable Diseases in Child Care Centers, Kindergartens, and Schools” by the Center for Health Protection, there is an outbreak from the epidemiological point of view if children or staff members develop similar symptoms one after another and the incidence is higher than usual. More detail is available from http://www.chp.gov.hk/files/pdf/Guild-Booklet-eng.pdf.
public fears has nonetheless received worldwide attention given constant reports on
bird flu cases in the region and vivid reminders of the 2003 outbreak of severe acute
respiratory syndrome (SARS) [3] [4] [5].

The measure to shut down all schools did lower the disease incidence but received
mixed comments from the public [6]. Some argued on hindsight that the move under-
scored Hong Kong’s relative vulnerability to global infectious disease pandemics and
perhaps an over-reaction to the influenza threat.

Geographical or spatiotemporal methods may offer insights and suggest solutions
to this universal measure of school closures in the territory. Tuckel et al. [7] em-
ployed the geographic information systems (GIS) technique to revisit the 1918 epi-
demic pattern of influenza in Hartford. Their study suggested that the use of GIS lends a better understanding of local outbreaks as opposed to viewing the epidemic as
a single incident. Venkatachalam and Mikler [8] used a global stochastic field simu-
lation paradigm to model infectious diseases. Other studies [9] [10] also showed that
the spatial autocorrelation technique helps to reveal local hot spots of influenza cases
and allows geographically focused precautionary measures to take place in due time.

In this paper, we used a number of different methods to investigate outbreak data
on affected school locations in March 2008 released by the Hong Kong Center for
Health Protection (CHP). We had also some background data for Hong Kong – dis-
trict boundaries and the extents of populated and non-populated areas. Our analyses
included two sets of results for case (affected schools) and control (non-affected
schools): (i) maps of standard deviational ellipses, (ii) nearest neighbor distance sta-
tistics, (iii) grid-based spatial autocorrelation, and (iv) kernel density maps. We of-
fered our comments on the results and explained choices of the analytical methods
and parameters used. We also hope to draw on the complementary roles of various
methods which seem to be deficient as the sole measure in decision and policy
matters.

2 Methodology

2.1 Data

Three sets of data were compiled: affected schools, non-affected schools, and back-
ground (Figure 1). The CHP provided over the Internet daily updates of institutions
(including elderly homes, schools, caring centers and hospitals) with flu outbreaks
beginning 6 March 2008. The suspected strains of virus for the outbreaks included
H1N1 (Brisbane) and H3N2 (Brisbane). The outbreak data for 6-13 March 2008 were
assembled to yield a total of 117 cases of affected schools. Data for all schools (kin-
dergarten, primary, and secondary) were compiled from information published on the
website of the Hong Kong Education Bureau. The locations of 2045 schools (includ-
ing the affected) were geocoded to obtain their coordinates for plotting. Background
data were obtained from the Survey and Mapping Office of the Lands Department of
Hong Kong. They were generalized and reconstituted for this study.
2.2 Analytical Methods

Our analysis involves statistically testing spatial patterns exhibited by case and control data produced by a variety of different methods. We employed ArcGIS, an integrated collection of software for geographical or spatial analyses by ESRI [11], to undertake our investigation. We also utilized GeoDa, which offers various functionalities for spatial analysis and visualization of the analysis results, to conduct spatial autocorrelation analysis [12]. But first, a descriptive analysis of the influenza data was conducted to offer background information for the study.

The method of standard deviational ellipses is an attempt to measure the directional trend of a set of points. A distance of one or two standard deviation will respectively cover approximately 68 or 86 percent of the points under study. The ellipse is based on the mean center of the points and its shape helps project the spread and directionality of the points. The weighted center of the points, adjusted by the size of the
schools, is also plotted for comparison. In the weighted case, the center of mass will be pulled towards points representing schools with more student population.

The nearest neighbor distance statistics measures the average distance between points and compares the measurement to the expected measurement of a hypothetical random distribution. The index ranges from 0 to 2.1491 with values less than 1 indicating a clustered pattern, values close to 1 indicating a random pattern, and exceeding 1 indicating a dispersed pattern.

The grid method is a measure of dispersion based primarily on the density of points. Here, we partitioned the study areas into two grid surfaces of cell sizes 1km x 1km and 500m x 500m. Only cells representing populated areas (i.e., excluding country parks and conservation areas) were included in the study. For each cell, the proportion of infection was computed by taking the number of infected schools divided by total schools within a cell. Both grid partitions were within guidelines suggested by researchers, viz. an average of 2 points per cell according to Curtis and McIntosh [13] or 1.6 points per cell by Bailey and Gatrell [14]. Spatial autocorrelation of the grid surfaces were examined using Moran’s I and local indicators of spatial autocorrelation (LISA) originally developed by Getis and Ord [15] (see also [16]).

Moran’s I values range from -1 to 1 much like the Pearson’s correlation coefficient. A value of 1 indicates spatial clustering of like values. A value of -1 signifies spatial dispersion while a zero value typifies spatial randomness. Spatial autocorrelation maps for each grid surface come in pairs – LISA and LISA significance maps. The former categorizes cells into 5 types: High-High (which shows a cell of a high value with adjoining neighboring cells also of high values); Low-Low (which shows a cell of a low value with adjoining neighboring cells also of low values); Low-High (which shows a cell of a low value with adjoining neighboring cells of high values); High-Low (which shows a cell of a high value with adjoining neighboring cells of low values); and not significant (which shows a cell not of the above four types). The latter map shows statistical significance of each cell type.

Kernel density mapping is a partitioning technique where local incidents within a moving 3D kernel of a defined radius or bandwidth are included to compute a density value for each cell in a grid overlaid on the study area [17] [18]. This technique effectively transforms a surface of raw counts into a density or probability surface. The density values are classed and shaded (darker shades to indicate higher values) to highlight hot spots.

3 Results and Discussions

3.1 Background

Our data revealed that the affected schools amounted to about 5.7 percent of the total and more primary schools were affected (Table 1). The Chi-square test was significant and we can be about 99.99 percent confident that the difference between the observed and expected patterns of affected frequencies did not result from mere random variability.
Table 1. Number of affected and non-affected schools by categories of schools

| Schools*          | Secondary | Primary | Kindergarten | Row Total |
|-------------------|-----------|---------|--------------|-----------|
| Affected (diseased) | 16        | 67      | 34           | 117       |
| Not Affected (control) | 447       | 522     | 959          | 1928      |
| Column Total      | 463       | 589     | 993          | 2045      |

Chi-square = 49.03; df =2; \( \rho < 0.0001 \)

* School type is based on profile established by the Education Bureau of Hong Kong (http://chsc.edb.hkedcity.net/en/index.php)

Looking at the 2006-2008 statistics released by the CHP [1], the incidence levels in February and March 2008 showed periodicity but the number of occurrences was not as high as those of 2006 and 2007. We must also bear in mind that the strains of viruses were different in these periods. However, the number of outbreaks by institutions did register a noticeable increase.

3.2 Maps of Standard Deviational Ellipses

Figure 2 shows two maps of standard deviation ellipses – one based on locations of controls (non-affected schools) and another on disease incidence (affected schools). The locations of mean and weighted mean centers (the latter adjusted by student population of each school) for the controls were indifferent; however, those of the affected schools were further apart. Because the exact number of infected cases per school was not available, displacement between the mean centers would indicate student population in the direction of the weighted mean center were potentially more susceptible to infection. The skewed nature of the standard deviation ellipse is a general measure of anisotropy or the property of being directionally dependent. By looking at the orientation of the standard deviational ellipses, we can try to predict which areas should prepare for a rise in incidence of influenza.

Point patterns of the infected cases were further analyzed to detect local clustering or hot spots using 1 km as the search radius. The results revealed four hot spots of influenza outbreak. These highlighted locations would be targets for close monitoring of further outbreaks in the neighborhoods.

3.3 Nearest Neighbor Distance Statistics

Figure 3 shows that the nearest neighbor distance for control cases was more compact than infected cases (nearest neighbor observed mean distance of 112 meters compared to 648 meters). The nearest neighbor statistics of 0.24 for control and 0.44 for infected cases indicated point patterns of significant clustering\(^2\). The results were in

\(^2\) The nearest neighbor statistics (observed mean distance / expected mean distance) range from 0 to 2.149. In general: 0 \(\rightarrow\) perfect clustering, 1 \(\rightarrow\) perfect randomness, 2 \(\rightarrow\) perfect even spacing of a grid formation, and 2.149 \(\rightarrow\) perfect triangular lattice [19]. Values of 0 to 0.5 indicate high degrees of clustering.
part due to the sheer difference in number (1928 control versus 117 infected cases) and partly to the locational distribution of schools that were often situated in regions of population clusters.

Comparing Figures 2 and 3, were the number of infected cases to increase in the next few days but the nearest neighbor observed mean distance had remained relatively stable, we would be sure that the newly infected cases would not be too far from existing infected schools in Figure 2. When the observed mean distance of
infected cases becomes noticeably shorter, we would expect additional outbreaks in localities other than the existing clusters. In theory, there should be a threshold distance to suggest the beginning of widespread infection but current research fall short of a means of determining this critical value.

3.4 Grid-Based Spatial Autocorrelation

Point data about schools were aggregated into areal units using grid cells of two different sizes: 1 km x 1 km and 500m x 500m. The 500m x 500m cell contains an average of 120 buildings per cell, thus ‘ignoring’ the detailed locational information in the observed point distribution (Figure 4). This level of aggregation is a form of data masking [20] to protect against the disclosure of individual identity as revealed by point locations while, at the same time, keeping the number of cell size manageable for desktop computer operations.
A few patches of ‘high-high’ occurrences or hot spots were evident from the surface of coarser cells of 1km x 1km (Figure 5). These hot spots were not extensive in their local coverage and they were buffered by cells of ‘low-high’ values. The surface of finer cells manifested similar patterns but the hot spots appeared more disjoint. These illustrations highlighted the difference of cell sizes on visual impact and analysis. They also brought out the fact that areas of high infection rate were isolated cases and perhaps not a cause for alarm at this stage. Given that the infection rates were computed based on institutions and not by residential locations of individuals, the aggregation might have under-estimated the spatial extent of disease spread.

3.5 Kernel Density Maps

Another method of revealing disease hot spots is by means of kernel density. Figure 6 shows four density surfaces created using different bandwidths (1 km and 500m) and cell sizes (1km x 1km, 500m x 500m, and 250m x 250m). The choices of bandwidth and cell size determine the degree of smoothing applied on a point pattern. A larger bandwidth yields a smoother surface with low intensity levels while a smaller bandwidth a thorny surface with more obvious local variations. Similarly a smaller cell...
(a) Spatial autocorrelation of quadrat counts of 1km x 1km cells

**Fig. 5.** Spatial autocorrelation of quadrat counts using two different cell sizes
(b) Spatial autocorrelation of quadrat counts of 500m x 500m cells

**Fig. 5. (continued)**
Fig. 6. Kernel density surfaces of influenza outbreak using different cell sizes and bandwidths.

partition produces a pattern that resembles more closely that revealed in a point map but too small a cell defeats the original intent of areal generalization. Bandwidth and cell size need not be the same but the former should be at least as large as the latter.

Although the cell sizes used for LISA and kernel density maps in Figures 5 and 6 were the same, the visual impression of hot spots projected by these maps were quite different. The kernel density surfaces appeared smoother and the patterns more contoured. Indeed, they resemble a probability surface of disease occurrence and the patterns are easy to interpret. LISA maps however showed hot spots as a discrete category along with other categories not identifiable on a kernel density surface. Pockets of hot spots buffered by spatial outliers implied that the disease had remained localized. In both cases, the patterns and Moran’s I values can inform about areas of concentration or hot spots are but not the severity of the matter. From the observed pattern, we cannot say for certain that the outbreak needs drastic measures of intervention (such as school closures or designated isolation). Even with daily tracking and reporting of a disease development, the map analysis may detect the occurrence of disease concentration or clustering patterns but still fall short of giving early warning or signal of an outbreak.
4 Observations and Implications

This paper demonstrates that graphic, statistical, and spatial analyses work together to provide clues on clustering tendency and cluster areas. The null hypothesis in point pattern analysis (either a random distribution of points or a homogeneous Poisson distribution) is not appropriate for analyzing cases of a disease which are usually clustered in regions of high population density. The degree of clustering as demonstrated in this paper should be evaluated with respect to the usually non-uniform population distribution.

As one of megacities in the world, Hong Kong’s high population concentration and density form a major source of disease burden. Geo-epidemiological models to enable the identification of disease variance in space can help guide interventions for improving the overall conditions of areas with a higher disease burden. A better understanding of spatial distribution of hot and cold spots would help formulate policies to target specific community groups. Here, the spatial process of influenza was examined in terms of variation from Poisson processes. Certain analyses have become more meaningful because of local policies and jurisdiction. For example, movements of primary school students (allocated based on residential locations) are controlled to school districts thereby reducing cross district interaction. Designated isolation of infected primary schools and schools around the hot spots will likely be an effective intervention measure. Other settings such as secondary schools and hospitals with less movement restrains may be modeled in similar fashion but more radical intervention approach may be warranted.

There is no clear cut definition for an outbreak. From the epidemiological point of view, an outbreak occurs if individuals develop similar symptoms one after another and the disease incidence is higher than usual. However, a single case posing major impact to the population at large (such as SARS in 2003) may sometimes warrant intervention treatment for an outbreak. Studies have also indicated that the burden of disease is significantly higher in slums as compared to affluent areas [21] [22] and people of similar socio-economic background and demographic characteristics tend to share similar activity pattern and action space [23] [24] [25]. As such, disease occurrences will likely spread within the community groups. The current intervention practice of a unified policy for an entire city (such as total school closures) may be disruptive to communities not under immediate threat although variable closures may cause confusion and anxiety in practice.

The effectiveness of policies in establishing a functioning health care system depends critically on the capacity of local governments to implement and enforce the policies. Hong Kong’s urban health administration, supervision and monitoring are segmented. Time lags between notification of suspected cases and confirmation of statutory notifiable diseases may distort counts. If the different administrations can coordinate their policies, more effective means of communication and intervention strategy can be devised to decrease the likelihood of disease transmission and possibly contain a potential flu pandemic.

Our findings, however, were constrained by the data from the Hong Kong SAR Government. First, we did not have data about the intensity of the infection (e.g. number of individuals reported ill for each institution). Therefore, we were unable to weight severity by institution. Second, our data were at institutional as opposed
to individual level. We had no specific data about the infected individuals (e.g. residential location of those infected) and thus unable to delimit their zones of active activity even though residential locations and transport preference have been shown to affect transmission patterns [26]. Third, district boundary is artificial. The presence of such a boundary separating infected and non-infected schools will, in no way, reduce the chance of getting infected. Therefore, district-level aggregation or the modifiable areal unit problem [27] [28] [29] in which health policies are based may be debatable.

The strengths of the study include careful assessment of the aggregation level and comparison of different visualization and presentation techniques. We did demonstrate in this paper that “seeing is believing.” Many previous studies on epidemics of respiratory infectious diseases often focused on using deterministic models, charts and tables to analyze the spread of the diseases [30] [31] [32] [33] [34]. They were, however, not able to highlight the spatial characteristics of the diseases. Maps, unlike a printed list of schools, offer a viewable version of the locations or concentrations of a disease which may render a decision, such as school closures, more justifiable. Furthermore, the grid method offers a suitable means of seeing the distribution without disclosing too much detail. For the case of Hong Kong, the opportunity to examine disease by more refined census enumeration units (e.g. tertiary planning units or street blocks) exists to provide fresh insights into the veracity and complexity of the relationship between public health events and neighborhood characteristics. There is further opportunity to apply complex statistical modeling methodology and investigation of cross-level interactions if addresses of the infected subjects were available for geocoding. Such data will allow epidemiologists to see how social mixing patterns might affect disease spread and what measures might protect the public's health.

References

1. CHP: Daily Update of Influenza Situation, http://www.info.gov.hk/healthzone/chp/Flu_dailyupdate_14_3_08_final_en.pdf (Accessed on March 14, 2008)
2. Parry, J.: Hong Kong closes all primary schools in flu outbreak. British Medical Journal 336, 632 (2008)
3. BBC: HK schools close amid flu fears, http://news.bbc.co.uk/2/hi/asia-pacific/7291169.stm (Accessed on March 14, 2008)
4. FOX News: Hong Kong Orders 560,000 Kids to Stay Home for Two Weeks Amid Flu Outbreak, http://www.foxnews.com/story/0,2933,337169,00.html (Accessed on May 5, 2008)
5. TIME: The Hong Kong Flu Scare of 2008, http://www.time.com/time/health/article/0,8599,1722633,00.html (Accessed on May 5, 2008)
6. Cummings, L.M.: Over-securitizing Public Health? The Recent Hong Kong Flu Case, March 17 (2008), http://ponderingir.wordpress.com/2008/03/17/over-securitizing-public-health-the-recent-hong-kong-flu-case/ (Accessed on April 5, 2008)
7. Tuckel, P., Sassler, S., Maisel, R., Leykam, A.: The diffusion of the influenza pandemic of 1918 in Hartford, Connecticut. Social Science History 30(2), 167–196 (2006)
8. Venkatachalam, Mikler: Modeling infectious diseases using global stochastic field simulation (2006), http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1635909&isnumber=34297 (Accessed on November 15, 2008)

9. Crighton, E.J., Elliott, S.J., Moineddin, R., Kanaroglou, P., Upshur, R.E.G.: An exploratory spatial analysis of pneumonia and influenza hospitalizations in Ontario by age and gender. Epidemiology and Infection 135(2), 253–261 (2007)

10. Greene, S.K., Ionides, E.L., Wilson, M.L.: Patterns of influenza-associated mortality among US elderly by geographic region and virus subtype, 1968–1998. American Journal of Epidemiology 163(4), 316–326 (2006)

11. ESRI: ArcGIS 9.2 Desktop Help (2008), http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm (Accessed on March 3, 2008)

12. Anselin, L., Syabri, I., Kho, Y.: GeoDa: an introduction to spatial data analysis. Geographical Analysis 38(1), 5–22 (2006)

13. Curtis, J.T., McIntosh, R.P.: The interrelations of certain analytic and synthetic phytosociological characters. Ecology 31, 434–455 (1950)

14. Bailey, T.C., Gatrell, A.C.: Interactive Spatial Data Analysis. Longman Group Limited, Essex (1995)

15. Getis, A., Ord, J.K.: The analysis of spatial association by use of distance statistics. Geographical Analysis 24, 189–206 (1992)

16. Anselin, L.: Local indicators of spatial association – LISA. Geographical Analysis 27, 93–115 (1995)

17. Levine, N.: CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations. Ver. 3.1, Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC (2007)

18. Ratcliffe, J.: Spatial Pattern Analysis Machine, http://jratcliffe.net/ware/spam1.htm (Accessed on October 15, 2007)

19. Taylor, P.J.: Quantitative Methods in Geography – An Introduction to Spatial Analysis, pp. 156–162. Houghton Mifflin Company, New Jersey (1977)

20. Rushton, G., Armstrong, M.P., Gittler, J., Greene, B.R., Pavlik, C.E., West, M.M., Zimmerman, D.L.: Geocoding in cancer research: a review. American Journal of Preventive Medicine 30(2), S16–S24 (2006)

21. Reddy, S., Shah, B., Varghese, C., Ramadoss, A.: Responding to the threat of chronic diseases in India. The Lancet 366(9498), 1744–1749 (2003)

22. Agtini, M.D., Soeharno, R., Lesmana, M., Punjabi, N.H., Simanjuntak, C., Wangsasaputra, F., Nurdin, D., Pulungsih, S.P., Rofiq, A., Santos, H., Pujarwoto, H., Sjahrrachman, A., Sudarmono, P., von Seidlein, L., Deen, J.L., Ali, M., Lee, H., Kim, D.R., Han, O., Park, J.K., Suwandono, A., Ingerani, O.B.A., Campbell, J.R., Beecham, H.J., Corwin, A.L., Clemens, J.D.: The burden of diarrhoea, shigellosis, and cholera in North Jakarta, Indonesia: findings from 24 months surveillance. BMC Infectious Diseases 5, 89 (2005)

23. White, S.E.: Action space, human needs and interurban migration. The Professional Geographer 29(1), 47–52 (1977)

24. Joseph, A.E., Poyner, A.: Interpreting Patterns of Public Service Utilization in Rural Areas. Economic Geography 58(3), 262–273 (1982)

25. Lee, R.E., Cubbin, C., Winkleby, M.: Contribution of neighbourhood socioeconomic status and physical activity resources to physical activity among women. Journal of Epidemiology and Community Health 61, 882–890 (2007)

26. Lai, P.C., So, F.M., Chan, K.W.: Spatial Epidemiological Approaches in Disease Mapping and Analysis: A handbook on operational procedures. CRC Press, Boca Raton (2009)

27. Openshaw, S.: The Modifiable Areal Unit Problem. Geobooks, Norwich (1984)

28. Holt, D., Steel, D.G., Tranmer, M.: Area homogeneity and the modifiable areal unit problem. Geographical Systems 3, 181 (1996)
29. Rushton, G.: Improving the geographical basis of health surveillance using GIS. In: Gatrell, A.G., Loytonen, M. (eds.) GIS and Health, ch. 5. Taylor & Francis, London (1998)
30. Dye, C., Gay, N.: Modeling the SARS epidemic. Science 300, 1884–1885 (2003)
31. Lipsitch, M., Cohen, T., Cooper, B., Robins, J.M., Ma, S., Jams, L., Gopalakrishna, G., Chew, S.K., Tan, C.C., Samore, M.H., Fisman, D., Murray, M.: Transmission dynamics and control of severe acute respiratory syndrome. Science 300, 1966–1970 (2003)
32. Ng, T.W., Turinici, G., Danchin, A.: A double epidemic model for the SARS propagation. BMC Infectious Diseases 3, 19 (2003)
33. Riley, S., Fraser, C., Donnelly, C.A., Ghani, A.C., Abu-Raddad, L.J., Hedley, A.J., Leung, G.M., Ho, L.M., Lam, T.H., Thach, T.Q., Chau, P., Chan, K.P., Lo, S.V., Leung, P.Y., Tsang, T., Ho, W., Lee, K.H., Lau, E.M.C., Ferguson, N.M., Anderson, R.M.: Transmission Dynamics of the Etiological Agents of SARS in Hong Kong: Impact of Public Health Interventions. Science 300, 1961–1966 (2003)
34. Shannon, G.W., Willoughby, J.: Severe acute respiratory syndrome (SARS) in Asia: a medical geographic perspective. Eurasian Geography and Economics 45(5), 359–381 (2004)