Journal of Public Health | Vol. 30, No. 3, pp. 234–244 | doi:10.1093/pubmed/fdn023 | Advance Access Publication 26 April 2008

Data-driven exploration of 'spatial pattern-time process-driving forces' associations of SARS epidemic in Beijing, China

Jin-Feng Wang[1], George Christakos[2], Wei-Guo Han[3], Bin Meng[4]

[1]Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, A11, Datun Rd, Anwai, Beijing 100101, China
[2]Department of Geography, San Diego State University, San Diego, CA 92182-4493, USA
[3]Center for Spatial Information Science and System, George Mason University, 6301 Ivy Lane, Greenbelt, MD 20770, USA
[4]College of Applied Sciences and Humanities, Beijing Union University, Beijing 100083, China
Address correspondence to Jin-Feng Wang, E-mail: wangjf@igsnrr.ac.cn

ABSTRACT

Background Severe Acute Respiratory Syndrome (SARS) was first reported in November 2002 in China, and spreads to about 30 countries over the next few months. While the characteristics of epidemic transmission are individually assessed, there are also important implicit associations between them.

Methods A novel methodological framework was developed to overcome barriers among separate epidemic statistics and identify distinctive SARS features. Individual statistics were pair-wise linked in terms of their common features, and an integrative epidemic network was formulated.

Results The study of associations between important SARS characteristics considerably enhanced the mainstream epidemic analysis and improved the understanding of the relationships between the observed epidemic determinants. The response of SARS transmission to various epidemic control factors was simulated, target areas were detected, critical time and relevant factors were determined.

Conclusion It was shown that by properly accounting for links between different SARS statistics, a data-based analysis can efficiently reveal systematic associations between epidemic determinants. The analysis can predict the temporal trend of the epidemic given its spatial pattern, to estimate spatial exposure given temporal evolution, and to infer the driving forces of SARS transmission given the spatial exposure distribution.

Keywords associations, determinants, epidemic, SARS, spatial pattern, statistics, time evolution

Introduction

Severe Acute respiratory Syndrome (SARS) is a viral respiratory illness caused by a coronavirus known as SARS-associated coronavirus, SARS-CoV.[1,2] According to the World Health Organization (WHO), a total of 8098 people worldwide became sick with SARS during the 2003 outbreak; of these, 774 people died. SARS quickly obtained the status of the first pandemic of the twenty-first century.[3,4] In this work we focus on the SARS epidemic in the city of Beijing (China) during the year 2003.

SARS usually spreads by person-to-person transmission. The virus transmits most readily by respiratory droplets produced when an infected person coughs or sneezes. The virus can also spread when a person touches a surface or object contaminated with infectious droplets and then touches his/her mouth, nose or eyes. In addition, it is possible that the SARS virus might spread more broadly through the air (airborne spread) or by other ways that are not now known.[5] Naturally, composite space-time variations, disease parameters and determinants (proximity to infected persons, population density, number of doctors, etc.) are essential features of the SARS epidemic.[6] Understanding these features is

Jin-Feng Wang, Professor of Geoinformatics
George Christakos, Professor of Geography
Wei-Guo Han, Scientist of Computation
Bin Meng, Lecturer of Human Geography

critical for the scientific representation of the epidemic and the efficient control of its transmission.

In general, the evolution of an epidemic is represented in terms of some kind of a mechanistic model.[7–14] However, in the case of SARS (as well as in many other epidemic situations) only datasets of various kinds are available, whereas little is known about the underlying disease mechanisms. Mainstream statistics, time series, meta-analysis, randomized control trials and spatial techniques have been used to detect disease characteristics in a given dataset,[15–18] but the substantive relationships between the disease characteristics across space-time often remain unknown. Similarly, the procedure generating epidemiologic curves[19] (i.e. curves showing the day-to-day rate of SARS growth) does not account for important associations between epidemic characteristics and dominant space-time disease patterns and dependencies. Despite the aforementioned limitations of the mainstream approaches, it is highly desirable to develop methods that can incorporate the scattered and incomplete datasets generated by various surveillance systems[20–22] and by different programmatic and sectoral actions[23] in order to evaluate salient associations between epidemic determinants and quantify consequential space-time variations and patterns.

In view of similar considerations, the aim of the present work is to develop a mathematically rigorous and scientifically meaningful SARS modelling framework that accounts for the crucial epidemic associations mentioned above by combining information about space, time, disease parameters and epidemic determinants from diverse datasets. The notation used in this paper is described in Appendix 1; the reader should peruse it before proceeding.

Data and methods

SARS is a typical case of a communicable disease about which a set of observed data has been obtained across space and time, and its potential determinants are available in a geographical information system (GIS) format. Below, several tools are considered for exploring epidemic determinants, spatial patterns and temporal processes on the basis of the available datasets. Certain statistics pairs share common elements that can be used to link these determinants, patterns and processes, and to formulate informative associations between them.

In this study, it is assumed that the pair-wise linking of the various SARS determinants is mathematically represented by a system of the general form $f(R, C, F, \Theta, s, t) = 0$, which is constructed mainly on the basis of the available datasets. The underlying epidemic mechanisms (usually represented in terms of differential equations) are not considered in this formulation, but they will be the topic of forthcoming publications.

Datasets

The first SARS case in Beijing was recorded on 1 March 2003. The city consists of 16 districts and 2 counties, with a total population of 12.5 million people occupying an area of 17 800 km^2. Information about SARS cases was available daily from authorized reports, beginning on 20 April 2003 and continuing till the end of the epidemic on 24 June 2003. Since April 27, the data in each of the 18 Beijing districts were obtained in GIS format. GIS data also provided the residential locations of all 11 108 people who had been in close contact with infected persons. This information was collected by exhaustively tracing SARS cases from 20 April 2003 onwards and backwards. Other data that were considered relevant to the epidemic were also obtained, including population counts in 246 census units, hospital locations, number of doctors in each hospital and the topography of the main traffic lines.

By way of a summary, the following datasets were systematically used in the present SARS study:

(i) New SARS cases, $R(s, t)$, reported daily in all 18 districts; s refers to the geographical district and t to the day considered.
(ii) Epidemic-relevant determinants, $F(s)$, i.e. population (P), number of doctors (D), number of hospitals (Hosp) and proximity to traffic (T) for each of the 18 districts.
(iii) Residential locations, H(s), of each one of the 11 108 individual close contacts to the infected persons.

Temporal evolution of infected cases

A SEIR (susceptible-exposed-infectious-recovered) model was developed to represent SARS epidemic transmission in a dynamic context. The model basically involves the following successive stages of the individuals involved:

Susceptible → Exposed → Infectious → Recovered,

see Appendix 2. The SEIR model for the SARS epidemic is based on the theoretical considerations discussed in Hamer[24] and Anderson.[25] Similar dynamic models have been used in early SARS studies by Lipsitch *et al.*[3] and Riley *et al.*,[5] which concluded that the SARS coronavirus, if uncontrolled, would infect the majority of people wherever it was introduced. Epidemic parameters (as described in Chowell *et al.*[26]), temporal process and total size can be easily simulated using a time-dependent SEIR model and a small amount of data.

Spatial distribution of identified contacts

In general, the determination of spatial dependence contributes significantly to adequate risk exposure assessment.[27–31] In the present study, the nearest-neighbour hierarchical clustering (NHC) model (Appendix 3) was used to identify spatial patterns of SARS distribution. Points were randomly distributed in space and circled in a spatial cluster, if their inter-distance was significantly smaller than the mean distance. Note that the primary spatial clusters may be further clustered using the same algorithm.

Table 1 Common items of different statistics

Common item	R(t) *and* Θ(t)	C(s) ~ T(s)	C(t)	F(t): T, P, D, ...
R(t) and $\Theta(t)$			*T*	*t*
C(s) ~ *T(s)*			*C*	*T*
C(t)	*t*	*C*		*t*
F(t): T, P, D, ...	*t*	*T*	*T*	

Note: '*C* ~ *T*' denotes that *C* and *T* are associated (connected).

Temporal changes in spatial clustering of cases

We measured the degree of global spatial data clustering using Moran's I_M coefficient (Appendix 4). A positive value of the coefficient meant that adjacent districts had similar values, whereas a negative value indicated that adjacent districts were dissimilar. A temporal measure of spatial clustering was obtained by calculating this coefficient on a daily basis. This time series was then filtered using a discrete wavelet transform implemented by the MatLab computer library (http://www.mathworks.com/). In this way, the time series was decomposed into low-frequency components (reflecting the fundamental trend of a time series) and high-frequency components (reflecting noise caused by random factors).

Factors associated with dispersion of infection

The Black-White (BW) join-count test (Appendix 5) was used to measure the extent to which districts of a disease network shared the same infection pattern. We started with an initial disease dispersal network, which consisted of connections (or joints) between districts. Each day a district was coded black (B) if a SARS case was reported on that day; otherwise it was coded white (W). Every network joint connected two B districts (BB), two W districts (WW), or a B and a W district (BW). The observed number of BW joints was compared with the expected number, and a standard normal deviation (z-score) was used to test the significance. High negative values of the statistics indicated clustering of infected cases on the network, whereas high positive values provided evidence of spacing. Seven networks were considered as follows:

N1. Local transmission: Two districts were considered connected if they shared a common geographical boundary.

N2. Nearest district: Each district was connected to its nearest neighbour as measured by the distance between the centroids of the districts.

N3. Population size: Districts were ranked according to population size, and consecutive districts in the corresponding hierarchy were appropriately connected.

N4. Population density: The same as in N3, but ranked by population density, instead.

N5. Number of doctors: The same as in N3, but ranked by number of doctors in the district.

N6. Number of hospitals: The same as for N3, but ranked by number of hospitals in the district.

N7. Urban-rural: Eight districts were designated urban and the rest rural; a rural-urban pair was considered connected if (i) the districts shared a boundary, or (ii) the urban district could be reached from the rural district by passing through just one other rural district, or (iii) the rural district could be reached from the urban district by passing through just one other urban district.

For each network, the BW join-count statistics was calculated for each day, and statistics changes were plotted over time.

Combining space, time, parameters and factors of epidemic transmission

The above separate SARS statistics were combined by means of pair-wise linking of common items to form a network connecting (Table 1):

- –The time series of infectives $R(t)$ and epidemic parameters $\Theta(t)$; time t is here the common item.
- –The spatial pattern of risk exposure $C(s)$ and the city traffic loop $T(s)$; space s is the common item.
- –The time process of spatial clusters $C(t)$ and the transmission determinants $F(t)$, including spatial proximity $T(t)$; a link is established by the consistency between the spatial patterns of C and T.

Careful considerations of the combinations described in Table 1 play a central role in the evaluation of determinant associations in a space-time context.

Results

The statistics described above were applied to the datasets, in which case a systematic connection between the separate statistics was established and described below.

Separate statistics

The time process of SARS transmission, $R(t)$, and the transmission rate in $\Theta(t)$ were simulated using the SEIR model. The risk of spatial exposure to the epidemic, $C(s)$, was disclosed by mapping data on the residences of the 11 108 identified contacts of the Beijing SARS cases using the NHC model.[32] The first-order clusters indicated spatial clustering of high-risk susceptibles, whereas second-order clusters indicated regions with a high concentration of primary clusters. The $C(s)$ exhibited a strong visual association with traffic infrastructure, $T(s)$ (Fig. 1).

Temporal changes in the spatial clustering of cases, $C(t)$, were determined by the time-variation of Moran's I_M coefficient. The I_M-curve was then filtered using a discrete wavelet transform implemented by MatLab (with Daubechies db3 as mother wave and four decomposition levels). In this way, the curve was decomposed into a low-frequency component (a_4) and high-frequency components (d_1, d_2, d_3, d_4) (Fig. 2). Daily data on new SARS cases were grouped according to the district of residence. The time-variation of a_4 indicated that SARS cases in Beijing became increasingly clustered

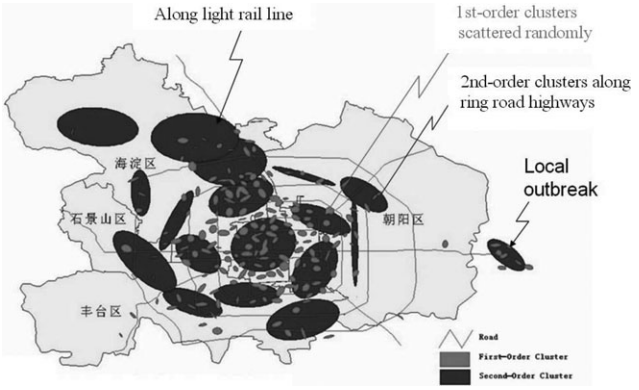

Fig. 1 Identified contacts (11 108) of Beijing SARS cases and their spatial patterns (modified from Wang *et al.*, 2006).

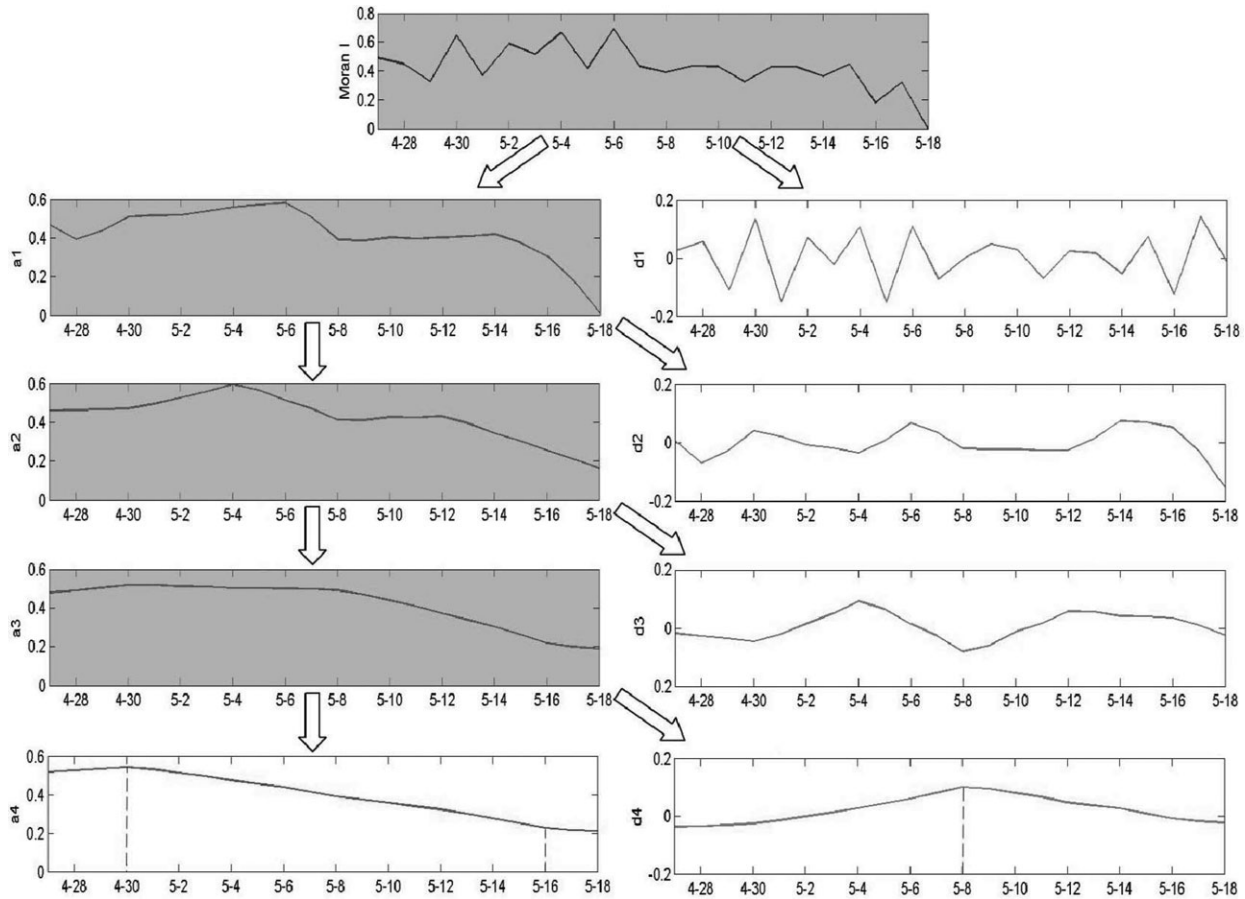

Fig. 2 Wavelet decomposition of the time series of Moran's coefficient I_M of spatial clustering (a_i and d_i, $i = 1, \ldots, 4$, correspond to low and high frequency signals respectively; grey figures were further decomposed in the arrows' directions).

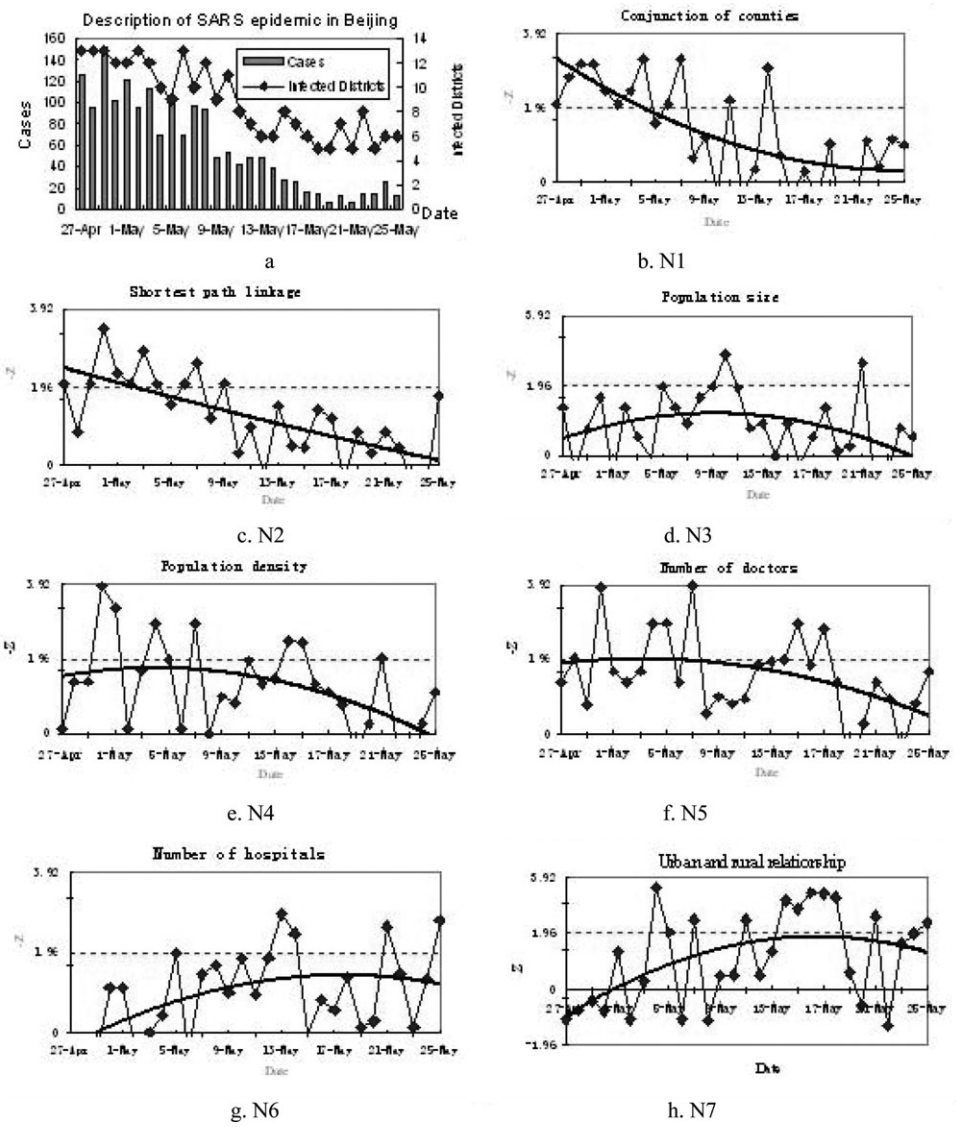

Fig. 3 Relationships between incidence of new SARS cases and spatial spread factors. Each panel tests one suspect determinant (N1, …, N7); dots are BW scores in each date; solid line is simple regression of the BW zigzag; horizontal dashed line denotes a significant threshold (modified from Meng *et al.*, 2005).

towards the end of April, 2003, followed by a decrease in clustering during 1–8 May 2003. After 9 May 2003, the time-variation of a_4 and d_4 indicated that spatial clustering was sharply declining.

Seven suspect determinants of the space-time SARS transmission, jointly denoted by $F(t)$, were investigated by means of the BW join-count test.[33] The spatial proximity factor is denoted by $T(t)$. Accordingly, Fig. 3 displays the determinants associated with SARS transmission dynamics.

Systematic analysis

The findings obtained on the basis of the different statistics of $R(t)$, $C(t)$, $C(s)$ and $F(t)$ above were pair-wise combined by means of their common elements and the corresponding network was formed as shown in Fig. 4. Logical inferences concerning the SARS epidemic evolution based on the connected network of Fig. 4 are discussed.

In the following inference analysis, the symbol 'A&B' means A and B; '$A \sim B$' denotes that the entities A and B are connected; the '$A \sim m \sim B$' means that A is linked to B via m; and the '$A \rightarrow B$' denotes that B is inferred from A.

Evolution of spatial clusters

Inference concerning the evolution of spatial SARS clusters was based on the rule

$$C(s)\&T(s)_{\text{Fig. 1}} \sim T \sim T(t)_{\text{Fig. 3}} \rightarrow C(s, T(t)). \qquad (1)$$

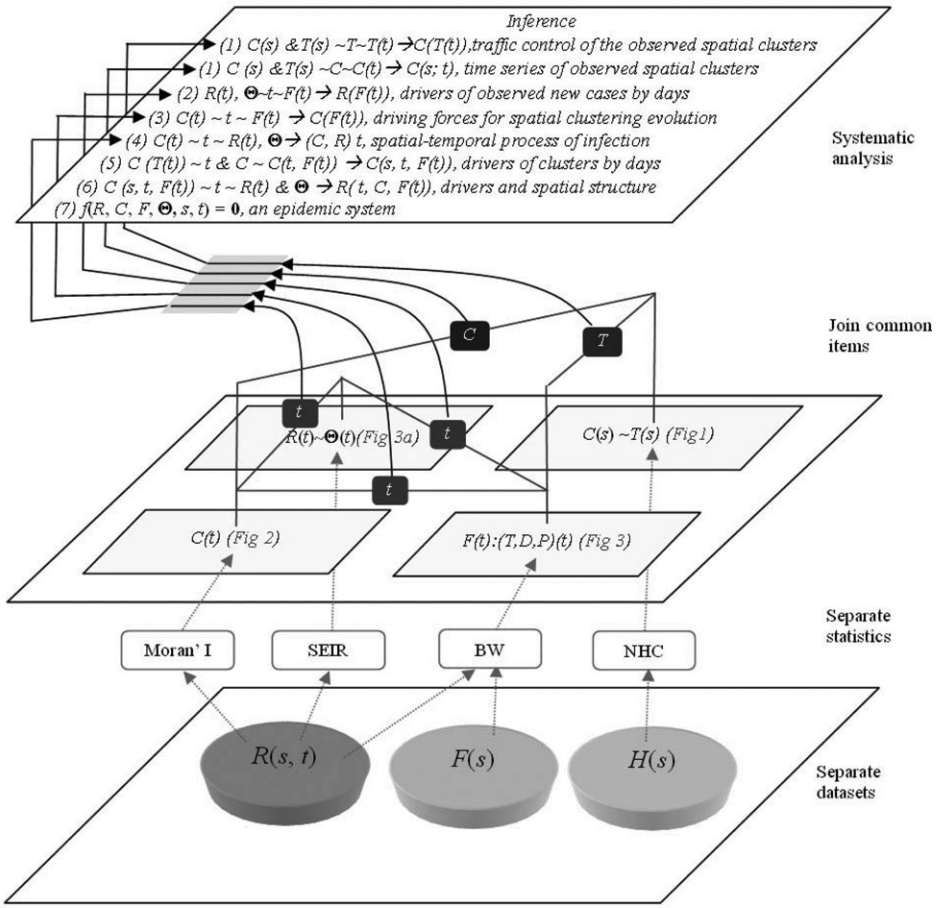

Fig. 4 Network of systematic connections between different statistics.

Figure 1 illustrates that the larger spatial clusters $C(s)$ were controlled by the urban traffic loops $T(s)$, i.e. a proxy of the macro spatial proximity, $T(t)$, the influence of which to the epidemic is detectable and changes with time (Fig. 3b and c). The smaller spatial clusters in Fig. 1 are randomly distributed over space, which is consistent with the spatial distributions of the population P and the number of doctors D, the two factors playing a significant role in the epidemic decline period (Fig. 3d–f). We concluded that the larger spatial clusters were dominant during the peak epidemic period and the smaller ones became more active during the decline period.

A similar conclusion was drawn by means of the inference rule

$$C(s)\&T(s)_{\text{Fig. 1}} \sim C \sim C(t)_{\text{Fig. 2}} \rightarrow C(s,t). \qquad (2)$$

In other words, Fig. 1 clearly shows that there are spatial clusters at two $C(s)$ scales, whereas Fig. 2 depicts the time series of the spatial clustering of the two scales. The large cluster (lower frequency part, left curve in Fig. 2) grew until 30 April 2003, and diffused into smaller clusters thereafter (as shown in the right part of Fig. 2), when there was a slight increase in the building up of higher-frequency components (or, equivalently, of smaller spatial clusters). Both the larger and the smaller clusters lost their strength after 8 May 2003 (Fig. 2).

Driving forces of the temporal process

Inference concerning the driving forces of the temporal process was based on the rule

$$R(t)\&\Theta(t) \sim t \sim F(t)_{\text{Fig. 3}} \rightarrow R(F(t)). \qquad (3)$$

The rapid decline of the daily new cases $R(t)$ and the infective rate in $\Theta(t)$ after 30 April 2003 (Fig. 3a) were predominantly driven by the spatial proximity $T(t)$ (Fig. 3b and c); and intermittently accompanied by the interaction between (infected) doctors D and the urban population density P (Fig. 3e and f). This trend was interrupted (see the diamond line in Fig. 3a) by a significant transmission between rural and urban areas on 15–18 May 2003 (Fig. 3h).

Driving forces of spatial clustering

Inference about the driving forces of spatial clustering was based on the rule

$$C(t)_{\text{Fig. 2}} \sim t \sim F(t)_{\text{Fig. 3}} \rightarrow C(t, F(t)). \quad (4)$$

In other words, the larger spatial clusters $C(t)$ (associated with the a_4 component in Fig. 2) had a consistent trend and were controlled by spatial proximity $T(t)$ (Fig. 3b and c), whereas the smaller clusters $C(t)$ (d_4 component in Fig. 2) were still building up, driven by significant P (population density) and D (number of doctors) during the specified time period (Fig. 3e and f).

Spatial clustering and intervention

Inference concerning spatial clustering and intervention was based on the rule

$$C(t)_{\text{Fig. 2}} \sim t \sim R(t), \Theta(t) \rightarrow (C, R, \Theta)(t). \quad (5)$$

That is, the larger spatial clusters represented by the a_4 component of $C(t)$ in Fig. 2 dominated the trend throughout the entire epidemic period, whereas the smaller clusters reflected in the d_4 component of $C(t)$ became more active during the decline period. The spatial and temporal patterns of transmission were controlled dynamically by intervention measures, as reflected in the epidemic parameters $\Theta(t)$ (Fig. 3a).[32] E.g. the '$(C, R, \Theta)(t)$' denotes that the triple within the bracket change with time.

In addition to the above direct pair-wise relationship between the variables of the SARS epidemic system, some composite inferences can be also drawn. A few examples follow. The

$$C[s, T(t)]_{\text{Eq.(1)}} \sim t \& z C \sim C[t, F(t)]_{\text{Eq.(3)}} \rightarrow C[s, t, F(t)] \quad (6)$$

demonstrates the spatiotemporal (s, t) dynamics of the observed spatial clusters (C) and the corresponding determinants $F(t)$ on a daily basis. The

$$C(s, t, F(t))_{\text{Eq.(5)}} \sim t \sim R(t) \ \& \ \Theta(t) \rightarrow R(t, C, F) \quad (7)$$

represents the fact that the observed time series of the infected cases $R(t)$ was driven by the spatial clustering variable $C(t)$ and the epidemic determinants $F(t)$. The role of the traffic T in F was to control the larger $C_0(s)$ [see Eq. (2)] and $C(t)$ [Eq. (4)]. The P and D in F control the smaller $C(s)$ [Eq. (2)] and $C(t)$ [Eq. (4)]. The role of $C(s, t)$ is to control the $R(t)$ [Eq. (5)]. The larger $C(s)$ occurred during the peak $R(t)$-values and the smaller $C(s)$ occurred during the decline phase of the epidemic, see Eqs. (2) and (5) or Eqs. (2) and (3). Accordingly, the peak and decline times of the epidemic spread were the two best times to efficiently conduct travel-related control at the ring roads T and the doctor (D)- and population (P)- related measures in the community, given a limited budget for epidemic control.

Discussion

Main finding of this study

The present work proposed a methodological framework to determine the main features of SARS transmission during the Beijing epidemic of 2003 using various datasets and space-time statistics techniques. Epidemic determinants and the relevant statistics are pair-wise linked in a space-time context, which means that the study of determinant associations can considerably enhance the mainstream epidemic analysis and improve the understanding of previously undetected relationships within the epidemic system.

Several interesting findings were obtained. Spatial neighbourhood is a major component of epidemic spread modelling, whereas coupling spatial transmission with population density and healthcare workers was sporadically significant. Changes in the spatial spread of SARS over time indicated which intervention measures are likely to be most effective at different phases of the outbreak. When the epidemic is growing via diffused infection, efficient intervention measures include isolating cases and reducing inter-regional movement. When cases of the disease appear in clusters, resources should be directed towards curbing transmission within localities that have a high incidence of infection. Our results suggested that improving control measures at the local level would have been the most efficient approach, before the end of April, 2003. Local intervention strategies include quarantine of family members. Once these measures take effect, transmission is reduced substantially and becomes dominated by more remote contacts. The evidence that diffusion of infections drove the declining stage of the outbreak may indicate that the measures were successfully controlling local transmission.

Temporal variables (e.g. number of infectives, season and weather), spatial variables (e.g. risk exposure, surveillance network, travel warnings and isolation) and other factors (e.g. immunity, population, doctors, hospitals and transportation) are mapped on the SARS epidemic system. This allows transferring operations from a less detectable or operable domain to a domain with features that improve determinant identification, generate accurate predictions and sound inferences, and contribute to strategies aiming at controlling disease spread. E.g. the free travel or isolation over regions (space) lead to an increase or decrease of the total number of

infectives (time); and the seasonal temperature change (time) alters the spatial pattern of risk exposure (space).

More than one statistics have to be calculated in epidemic data analysis, so more resources are needed to improve the accuracy of such calculations. Fortunately, there is a plethora of computer software and open source codes that are easily accessible on the Internet and greatly facilitate data analysis computations. In this study, the datasets are managed by the GIS software ARCGIS (http://www.esri.com/software/arcgis/). The hierarchical clustering is implemented by the Crimestat software (http://www.icpsr.umich.edu/CRIMESTAT/). Moran's coefficient I_M is calculated using the GeoDA software (http://www.geoda.uiuc.edu/). Wavelet decomposition and SSEIR are computed using Matlab (http://www.icpsr.umich.edu/CRIMESTAT/), and the BW test is performed using software compiled by the authors. Primary training on GIS and spatial statistics are needed in order to operate the above software.

What is already known on this topic

Several epidemiologic assessments of the SARS-CoV have concluded that this coronavirus is sufficiently transmissible to cause a very large epidemic, but not so contagious as to be uncontrollable with good, basic public health measures.[3,5] It is a common modelling practice to use a single or several separate data statistics to investigate individual epidemic properties and determinants,[6] and to quantify policies addressing problems caused by the disease.[34] Among other things, the spatial pattern, time process and driving forces of an epidemic can be explored by hierarchical clustering, SSEIR, and BW, respectively. Also, epidemiologic curves have been generated that display the day-to-day rate of SARS growth[19] but without accounting for the 'spatial pattern-time process-driving forces' associations of the epidemic. Only a limited number of studies have attempted to partially represent and evaluate determinant associations and links between separate statistics under conditions of uncertainty.

What this study adds

The present work is the first systematic effort to represent the aforementioned associations and links in a rigorous manner and to assess their considerable role in understanding the various aspects of epidemic distribution across space-time. Advanced mathematical modelling of an epidemic could potentially unveil the relationships between various determinants, but this modelling is often impossible due to the unavailability of the necessary scientific knowledge, including a lack of understanding of the fundamental disease mechanisms.[35,36] Nevertheless, the present study shows that using the available datasets an informative modelling approach can be developed that is based on the pairwise linking of epidemic determinants and the formulation of a network that systematically connects the relevant statistics. In addition, unlike earlier SARS modelling studies,[5] the present approach accounts in an efficient manner for the spatial variation of the data available. This approach is particularly useful when the disease mechanisms are not clear and the corresponding mathematical model is difficult to develop for reasons mentioned above. Moreover, the approach can be successfully implemented to simulate the response of SARS transmission to various epidemic control factors, identify target areas and determine the critical time and relevant factors. Given the response by individuals, institutions and governments to a new and dangerous disease, the findings of the present study can be useful in the implementation of efficient epidemic intervention strategies and effective population protection policies.

Limitations of this study

The present modelling approach is concerned mainly with statistical associations rather than causal laws. The latter are available primarily in cases in which the fundamental disease mechanisms (molecular biology, etc.) are well understood. In this sense, SARS modelling could benefit from future developments in basic scientific research in disease etiology and mathematical theories of epidemic propagation, as well as from sophisticated surveillance capabilities and meta-analysis. Furthermore, the present SARS modelling could be improved by considering epidemic variation and dependence in a composite space-time continuum under conditions of multi-sourced uncertainty. This may involve a combination of random field theories with temporal GIS techniques.

Future applications

By means of the novel methodological framework considered in the present work, future epidemic studies may not only benefit from 'meta data analysis' but also from 'meta model analysis'. A wide variety of inter-disciplinary datasets that are relevant to the epidemic spread can be integrated by means of this framework. The statistics functions and spatial analysis techniques described above can be readily applied to the study of epidemics other than SARS. Furthermore, the methodological framework is not limited to the statistical tools and datasets considered in this paper, but is more general and flexible, in the sense that it may include additional information sources, statistics functions and models; and it can allow the performance of other types of

logical inferences, thus generating new findings and conclusions concerning the epidemic under investigation. By generalizing the current approach within the context of random field theories with temporal GIS techniques, one may consider additional issues related to the next generation of (stochastic) SARS models allowing for space-time variations, multi-sourced uncertainty, seasonality and varying transmission modes, the case of SARS virus long-time persistence, and examining the case of global eradication rather than local control in terms of early epidemic detection measures.

Acknowledgements

We would like to thank A. J. McMichael, N. Becker and K. Glass for their help in this research, and the Beijing Center for Disease Control and E. Zhong and D. Zhuang for preparing the SARS information and relevant data.

Funding

The work was funded by the NSFC (40471111, 70571076), the MOST (2006AA12Z215, 2007AA12Z241) and CAS (KZCX2-YW-308).

Appendix 1: notation

s	spatial location.
t	time period.
$R(s, t)$	daily (t) reported new cases of SARS in each of the 18 districts (s) of the city of Beijing (China).
$\Theta(t)$	epidemic parameters.
$H(s)$	residential locations of all 11 108 people who had been in close contact with infected persons.
$F(s)$	determinants of SARS.
P	population counts in 246 census units.
Hosp	location of hospitals.
D	number of doctors in a hospital.
$C(s)$, $C(t)$ and C	spatial pattern of risk exposure, temporal change in spatial clustering of cases, and spatial pattern in general, respectively.
$T(s)$, $T(t)$ and T	main traffic lines with spatial location, traffic factor of SARS transmission in time, and traffic in general, respectively.
$(C, R, \Theta)(t)$	triplet of entities within the bracket change with time.
A&B	entities A and B.
$A \sim B$	entities A and B are connected.
$A \sim m \sim B$	A is linked to B via m.
$A \rightarrow B$	B is inferred from A.

Appendix 2: the SEIR model

On the basis of the discussion in Kermack and McKendrick[7] and Anderson,[25] the following SEIR model has been proposed.

$$\left.\begin{array}{l} \mathrm{S} \xrightarrow{\ell(\mathrm{t})} \mathrm{E} \xrightarrow{g} \mathrm{I} \xrightarrow{a} \mathrm{R} \\ \frac{\mathrm{d}}{\mathrm{d}t}E(t) = \ell(t)S(t) - gE(t) \\ \frac{\mathrm{d}}{\mathrm{d}t}I(t) = gE(t) - aI(t) \\ \ell(t) = b + c(1 + \mathrm{e}^{d(t-e)})^{-1} \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{R}(t) = aI(t) \end{array}\right\} \quad (8)$$

where $E(t)$, $I(t)$ and $R(t)$ are the number of exposed, infectious and removed individuals at time t, $\ell(t)$ average number of contacts per infectious person (depends on time t because the control effort changes with t), G the rate at which exposed (latent) individuals become infectious. A the rate at which infectious individuals are removed (recover or are isolated), $R_0 = \ell(0)/a \approx (b + c)/a$ the basic reproduction number for this model, $R_0 \approx b/a$ approximated eventual reproduction number, $1/g$ average incubation period, $1/g$ average infection period and $a = 0.252$, $b = 0.008$, $c = 0.588$, $d = 0.368$, $e = 54$ and $g = 0.200$.

Appendix 3: nearest-neighbour hierarchical clustering

Our datasets included the geographical locations of the residences of the 11 108 identified contacts. In the absence of clustering, the mean distance of an identified contact from its nearest neighbour and the standard error of this mean are given by, respectively,

$$\left.\begin{array}{l} \mu(d) = \frac{1}{2}\sqrt{A/N} \\ s(d) = 0.26136\sqrt{A}/N \end{array}\right\}, \quad (9)$$

where A is the area of the region and N is the number of identified contacts in that region. We used the CrimeStat software package,[37] which defines a threshold distance

$$L = \mu - 1.645s, \quad (10)$$

and identifies all points having neighbours within this threshold. The software selects first-order clusters

sequentially, beginning with the point having the largest number of neighbours. The second-order clusters were obtained by calculating the centre of each first-order cluster, and then repeating the above procedure on these centre points. The first-order clusters included at least eight points and the second-order clusters included at least four points.

Appendix 4: Moran's *I* coefficient

To avoid confusion with Panel 1, we denoted Moran's *I* coefficient[35] as

$$I_M(t) = \frac{N\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}[x_i(t)-\bar{x}(t)][x_j(t)-\bar{x}(t)]}{(\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij})\sum_{i=1}^{N}[x_i(t)-\bar{x}(t)]^2}, \tag{11}$$

where

$x_i(t)$ = number of new SARS cases observed in district i on day t,
$\bar{x}(t)$ = mean number of new SARS cases per district on day t,

$$W_{ij} = \begin{cases} 1 & \text{if } ij \text{ are adjacent districts} \\ 0 & \text{if } ij \text{ are not adjacent districts or } i = j, \end{cases}$$

N = number of districts.

Appendix 5: BW join-count statistic

We started with the proposed disease dispersal network, which consisted of connections (or 'joins') between districts. Each day a district was coded black (B) if it reported a SARS case on that day; otherwise it was coded white (W). Every joint in the network linked two B districts (BB), two W districts (WW) or a B and a W district (BW). The observed number of BW joins was compared with the expected number, and a standard normal deviation (z-scores) was used to test the significance.[39,40] High negative values of the statistics indicated evidence of clustering of cases on the network and high positive values indicated evidence of spacing. The statistics are as follows:

$$z(\mathrm{BW}) = \frac{N_{\mathrm{BW}} - E(N_{\mathrm{BW}})}{\sqrt{\mathrm{Var}(N_{\mathrm{BW}})}}, \tag{12}$$

$$N_{\mathrm{BW}} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(x_i - x_j)^2, \tag{13}$$

$$E(N_{\mathrm{BW}}) = \frac{1}{2}[S_0 N_B N_W/N(N-1)], \tag{14}$$

$$\mathrm{Var}(N_{\mathrm{BW}}) = \frac{1}{4}\left[\frac{2S_1N_BN_W}{N(N-1)} + \frac{(S_2-2S_1)N_BN_W(N-2)}{N(N-1)(N-2)} + \frac{4(S_0^2+S_1-S_2)N_B\times(N_B-1)N_W(N_W-1)}{N(N-1)(N-2)(N-3)}\right] - E(N_{\mathrm{BW}})^2 \tag{15}$$

$$\left.\begin{aligned} S_0 &= \textstyle\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} \\ S_1 &= \tfrac{1}{2}\textstyle\sum_{i=1}^{N}\sum_{j=1}^{N}(w_{ij}+w_{ji})^2 \\ S_2 &= \textstyle\sum_{i=1}^{N}(w_{i.}+w_{.i})^2 \end{aligned}\right\}, \tag{16}$$

N_{BW} number of BW joins,

x_i = $\begin{cases} 1 & \text{if district } i \text{ is black} \\ 0 & \text{if district } i \text{ is white,} \end{cases}$

w_{ij} = $\begin{cases} 1 & \text{if district } i \text{ are connected} \\ 0 & \text{otherwise,} \end{cases}$

$w_{i.}$ average of w_{ij} over the second subscript,
N total number of districts,
N_B, N_W numbers of BB and WW joins, respectively.

References

1 Satija N, Lal SK. The molecular biology of SARS coronavirus. *Ann NY Acad Sci* 2007;**1102**:26–38.

2 Huang Y, Lau SKP, Woo PCY *et al*. CoVDB: a comprehensive database for comparative analysis of coronavirus genes and genomes. *Nucleic Acids Res* 2008;**36**:D504–D511.

3 Lipsitch M, Cohen T, Cooper B *et al*. Transmission dynamics and control of Severe Acute Respiratory Syndrome. *Science* 2003;**300**:1966–70.

4 Rota PA, Oberste MS, Monroe SS *et al*. Characterization of a novel coronavirus associated with Severe Acute Respiratory Syndrome. *Science* 2003;**300**:1394–9.

5 Riley S, Fraser C, Donnelly CA *et al*. Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science* 2003;**300**:1961–6.

6 Donnelly CA, Ghani AC, Leung GM *et al*. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet* 2003;**361**:1761–6.

7 Kermack W, McKendrick A. A contribution to the mathematical theory of epidemics. *Proc R Soc Lond A* 1927;**115**:700–21.

8 Christakos G, Hristopulos DT. *Spatiotemporal Environmental Health Modelling*. Boston, MA: Kluwer Academic Publishers, 1998.

9 Christakos G, Olea RA, Serre ML *et al*. *Interdisciplinary Public Health Reasoning and Epidemic Modelling: The Case of Black Death*. New York, NY: Springer, 2005.

10 Xu B, Gong P, Seto E *et al*. A spatial-temporal model for assessing the effects of intervillage connectivity in Schistosomiasis transmission. *Ann Assoc Am Geograph* 2006;**96**:31–46.

11 Nakaya T, Nakase K, Osaka K. Spatio-temporal modelling of the HIV epidemic in Japan based on the national HIV/AIDS surveillance. *J Geograph Syst* 2005;**7**:313–36.

12 Becker N, Glass K. Controlling emerging infectious diseases like SARS. *Math Biosci* 2005;**193**:205–21.

13 Glass K. Ecological mechanisms that promote arbovirus survival: a mathematical model of Ross River virus transmission. *Trans R Soc Trop Med Hyg* 2005;**99**:252–60.

14 Yu HL, Kolovos A, Christakos G *et al*. Interactive spatiotemporal modelling of health systems: The SEKS-GUI framework. Jour. *Stochast Environ Res Risk Assess* 2007;**21(5)**:555–72.

15 Stroup D, Berlin J, Morton S *et al*. Meta-analysis of observational studies in epidemiology. A proposal for reporting. *JAMA* 2000;**283**:2008–12.

16 Tania B, Sánchez-Meca J, Marın-Martınez F *et al*. Assessing Heterogeneity in Meta-Analysis: Q Statistic or I2 Index? *Psychol Methods* 2006;**11**:193–206.

17 Viechtbauer W. Confidence intervals for the amount of heterogeneity in meta-analysis. *Stat Med* 2007;**26**:37–52.

18 Aldstadt J. An incremental Knox test for the determination of the serial interval between successive cases of an infectious disease. *J Stochast Environ Res Risk Assess* 2007;**21(5)**:487–500.

19 Banos A, Lacasa J. Spatio-temporal exploration of SARS epidemic. *Cybergeo*, Systèmes, Modélisation, Géostatistiques 2007. Art. no. 408 (mis en ligne le 27 Novembre 2007; modifié le 28 Novembre 2007): http://www.cybergeo.eu/index12803.html.

20 McMichael T. *Human Frontiers, Environments and Disease: Past Patterns, Uncertain Futures*. Cambridge, New York, NY: Cambridge University Press, 2001.

21 Murray C, Lopez A, Wibulpolprasert S. Monitoring global health: time for new solutions. *BMJ* 2004;**329**:1096–100.

22 Marris E. Inadequate warning system left Asia at the mercy of a tsunami. *Nature* 2005;**433**:3.

23 Holveck J, Ehrenberg J, Ault S *et al*. Prevention, control, and elimination of neglected diseases in the Americas: Pathways to integrated, inter-programmatic, inter-sectoral action for health and development. *BMC Public Health* 2007;**7**:6 (doi:10.1186/1471-2458-7-6).

24 Hamer W. Epidemic disease in England. *Lancet* 1906;**1906**:733–9.

25 Anderson R. *Infectious Diseases of Humans*. Oxford: Oxford University Press, 1991.

26 Chowell G, Nishiura H, Bettenciourt L. Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *J R Soc Interface* 2006. doi:10.1098/rsif.2006.0161.

27 Christakos G, Lai J. A study of the breast cancer dynamics in North Carolina. *Soc Sci Med* 1997;**45(10)**:1503–17.

28 Christakos G, Vyas V. A novel method for studying population health impacts of spatiotemporal ozone distribution. *Soc Sci Med* 1998;**47(8)**:1051–66.

29 Haining R. *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press, 2003.

30 Keeling M, Woolhouse M, May R *et al*. Modeling vaccination strategies against foot-and-mouth disease. *Nature* 2003;**421**:136–42.

31 Koch T, Denike K. Certainty, uncertainty, and the spatiality of disease: a West Nile virus example. Jour. *Stochast Environ Res Risk Assess* 2007;**21(5)**:523–31.

32 Wang JF, McMichael AJ, Meng B *et al*. Spatial dynamics of a SARS epidemic in an urban area. *Bull World Health Organ* 2006;**84**:965–8.

33 Meng B, Wang JF, Liu J *et al*. Understanding the spatial diffusion process of severe acute respiratory syndrome in Beijing. *J Public Health* 2005;**119**:1080–7.

34 Mikler AR, Venkatachalam S, Ramisetty-Mikler S. Decisions under uncertainty: a computational framework for quantification of policies addressing infectious disease epidemics. *J Stochast Environ Res Risk Assess* 2007;**21(5)**:533–43.

35 Dye C, Gay N. Modelling the SRAS epidemic. *Science* 2003; **300**:1884–5.

36 Sorensen MD, Sorenses B, Gonzalez-Dosal R *et al*. Severe Acute Respiratory Syndrome (SARS): development of diagnostics and antivirals. *Ann N Y Acad Sci* 2006;**1067**:500–5.

37 Levine N. CrimeStat III: A spatial statistics program for the analysis of crime incident locations (version 3.0). *Ned Levine & Associates: Houston, TX/ National Institute of Justice: Washington, DC*, 2004.

38 Moran P. The interpretation of statistical maps. *J R Statist Soc Ser B* 1948;**10**:243–51.

39 Krishna-Iyer PV. The theory of probability distributions of points on a lattice. *Ann Math Stat* 1950;**21**:198–217.

40 Haggett P. Hybridizing alternative models of an epidemic diffusion process. *Economic Geography* 1976;**52**:136–46.