13. Brendel, V., Bucher, P., Nourbakhsh, I. et al., Methods and algorithms for statistical analysis of protein sequences, Proc. Natl. Acad. Sci. USA., 1992, 89: 2002—2006.
14. Subramaniam, S., The biology workbench—A seamless database and analysis environment for the biologist (editorial), Proteins, 1998, 32: 1—2.
15. Persson, B., Argos, P., Prediction of transmembrane segments in proteins utilising multiple sequence alignments, J. Mol. Biol., 1994, 237: 182—192.
16. Sonnhammer, E. L., Heijne, G. Von., Krogh, A., A hidden Markov model for predicting transmembrane helices in protein sequences, Proc. Int. Conf. Intell. Syst. Mol. Biol., 1998, 6: 175—182.
17. Bateman, A., Birney, E., Cerruti, L. et al., The Pfam protein families database, Nucleic Acids Res., 2002, 30: 276—280.
18. Wallace, J. C., Henikoff, S., PATMAT: a searching and extraction program for sequence, pattern and block queries and databases, Comput. Appl. Biosci., 1992, 8: 249—254.
19. Altschul, S. F., Madden, T. L., Schaffer, A. A., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic. Acids. Res., 1997, 25: 3389—3402.
20. Thompson, J. D., Higgins, D. G., Gibson T. J., CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic. Acids. Res., 1994, 22: 4673—4680.
21. Felsenstein, J., PHYLIP——Phylogeny Inference Package (Version 3.2), Cladistics, 1989, 5: 164—166.
22. Saraste, M., Sibbald, P. R., Wittinghofer, A., The P-loop——a common motif in ATP- and GTP-binding proteins, Trends. Biochem. Sci., 1990, 15: 430—434.
23. Kyte, J., Doolittle, R. F., A simple method for displaying the hydropathic character of a protein, J. Mol. Biol., 1982, 157: 105—132.
24. Bonavia, A., Zelus, B. D., Wentworth, D. E. et al., Identification of a receptor-binding domain of the spike glycoprotein of human coronavirus HCoV-229E, J. Virol., 2003, 77: 2530—2538.
25. Ortego, J., Escors, D., Laude, H. et al., Generation of a replication-competent, propagation-deficient virus vector based on the transmissible gastroenteritis coronavirus genome, J. Virol., 2002, 76: 11518—11529.
26. Kuo, L., Masters, P. S., The small envelope protein E is not essential for murine coronavirus replication, J Virol., 2003, 77: 4597—4608.
27. Cyranoskiand, D., Abbott, A., Virus detectives seek source of SARS in China' s wild animals, Nature, 2003, 423: 467

(Received June 3, 2003; accepted June 26, 2003)

Chinese Science Bulletin 2003 Vol. 48 No. 13 1287—1292

Stochastic dynamic model of SARS spreading

SHI Yaolin

Computational Geodynamics Laboratory of Graduate School of Chinese Academy of Sciences, Beijing 100039, China (e-mail: shiyl@gscas.ac.cn)

Abstract **Based upon the simulation of the stochastic process of infection, onset and spreading of each SARS patient, a system dynamic model of SRAS spreading is constructed. Data from Vietnam is taken as an example for Monte Carlo test. The preliminary results indicate that the time-dependent infection rate is the most important control factor for SARS spreading. The model can be applied to prediction of the course with fluctuations of the epidemics, if the previous history of the epidemics and the future infection rate under control measures are known.**

Keywords: SARS, epidemic model, system dynamics, point process.

DOI: 10.1360/03wc0292

It is an important topic for the research to establish the dynamic model of SARS spreading, to understand the characteristics and control factors, and to make predictions of the course of the epidemics. Most dynamic epidemic models in Chinese study are following the classic work of Anderson and May[1], by construction and analysis of the epidemic differential equations[2—11]. Internationally, in addition to this kind of deterministic models[12—14], stochastic models[15] consist of another kind of work, either by consideration of random processes in the corresponding differential equations[16—18], or by performing Markov chain and Monte Carlo simulation[19]. For epidemics of SARS, which has only a small number of tens to thousands patients in a region, and consists of a very small percentage of the entire susceptible population, a stochastic model is probably more appropriate[15,18]. So far, two epidemic models on SARS have been published[20,21], which all adopted the stochastic approach with a benefit recognized that the modeling parameters can be easily adjusted during the evolution of epidemics[22].

On one hand, epidemic equations can be applied not only to epidemics study itself, but also to other kind of social and natural scientific problems, such as biological group distribution, spreading of new technology, spreading of rumors in society, etc.[23,24]. On the other hand, scientific methods from other research areas can be made use of for epidemiology. A suggestive case comes from fluid dynamics: the fluid motion can be studied macroscopically by the partial differential equations of conservation of mass, moment and energy; it can also be studied by tracing the motion, collision, and interaction of all the molecules, and approach the overall behavior by averaging their effects. The latter method as molecular dynamics

became tractable only after the development of advanced computers, which provide the technique possibility of calculating and tracing the motion of a large amount of molecules. In this study, we propose a stochastic model for SARS spreading by tracing the chain of infection of each patient and analyze the Vietnam case as an example. The methodology is similar to the molecular dynamics tracing every molecule to understand the macroscopic system, and the mathematic tools used is the point process similar to the stochastic model we used in the earthquake sequence analysis[25].

1 Model

In a preliminary form of SARS model, the spatial spreading and distribution can be neglected. We firstly focused on the temporal epidemic evolution within a specific area. Assuming that a SARS patient is imported, he will become infectious after the latent period. The latent period is believed to range within 1—12 d. In the model, a random parameter is assumed following the Poisson distribution. After the latent period, the patient becomes infectious, the number of people infected is determined in the computer by chance. Some infections may be traced as a result of close one-to-one contact, some may be infected in an indirect way (such as at the clinic of hospital, or flaws in the sewage system like the Amoy Garden case in Hong Kong). In any case, an averaged rate of direct or indirect infection per patient per day can be obtained, and their summation is the average infection rate for each patient. Because the distribution of infection rate is not yet well known for SARS as a new disease, we assume it to obey the Poisson distribution. The average infection rate may change at different stages of sickness for patient, and may change with social control measures taken. The length of the infectious period and prognosis of patients are all random variables. The infectious period may vary, and recover or die according to the mortality rate. The infection rate can be reduced by isolation measures. Some patients may become infectious first and then be isolated, some others may be quarantined first before they develop the symptoms, both can be produced stochastically in the model by computer. In summary, the latent period, date of infectious unset, date of isolation, and date of recovery or death, how many people newly infected each day for each patient, all these can be produced by the computer stochastically, stored in files, and updated every day of the epidemics. Therefore, with the four parameters (infection rate, latent period, infectious period, and morality rate) known, it is possible to construct a simple dynamic model of SARS spreading. If spatial spreading of SARS is to be studied, it is necessary to add information on the spatial structure of the model, the probability of flow of people among different regions, and situation related infection rate at different regions. The flow chart for computation is shown in Fig. 1.

Among the four parameters, the latent period is not affected by mankind, while the infection rate, length of infectious period, and morality rate are all subjected to changes produced by human measures. The morality rate does not affect the spreading of SARS in the model. Latent period, as a parameter not affected by human beings, does not vary significantly, and small changes of latent period do not make significant effects on the results. In Vietnam case, a one-day increase in the averaged latent period can produce a 2% decrease of duration of epidemics and 5% decrease in total number of patients. The duration of infectious period is reported to be about 10—14 d; change of one day can produce less than 5% changes in the results. Infection rate is the most important factor to

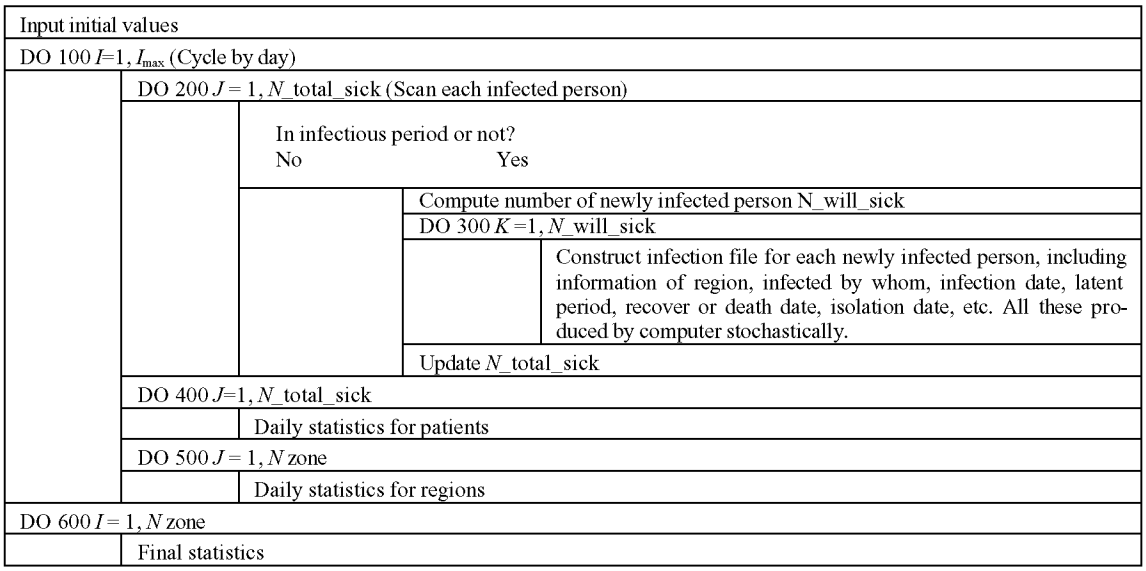

Fig. 1. Computational flow chart of the stochastic model of SARS spreading.

affect SARS spreading. Infection rate is influenced by many factors. It is related to the course of the disease, for example, it is reported that the patient is more infectious from the 3rd to the 5th day of onset, and extremely infectious to medical workers during the operation of tracheotomia, but non-infectious in the latent period. The infection rate is also related to the population characteristics, such as population of the susceptible group, type of confluence as residential area, school and university, building set, village, etc. Each type may have different infection rates and control measures should be adjusted to fit the local conditions. There are some factors which people can control effectively to reduce the infection rate, such as early detection of SARS patients, isolation of patients and quarantine of close contactors to SARS patients, improvement of clinics to reduce cross infection, protection of medical workers, etc. All these will be embodied in one parameter: reduction of the infection rate.

Of course, in more realistic models, many other factors should be included, such as the susceptible people may be divided into different groups, and each group may have different infection rates[18], the recovered people may get immunity for life or just temporarily. There may be cases that the infected people do not become sick (or only have very light symptoms), and they may or may not be infectious. To consider the spatial spreading of SARS, it is necessary to construct a spatial structure of the model, to include the rate of flow or mixing of population between different regions, and the flow probability may vary between different types of regions (such as city—city, city—village), distances, means of transportation, time of seasons, holidays, or panic run away people after announcement of epidemics, etc. Although these factors are not included in the present modeling, it is not difficult to consider them in the future modeling.

2 Results

In the real world, mother nature produces a consequence randomly by her own rules: in the model, computer produces a result based on our understanding of the natural rules.

Two points should be noted: first, a good model or bad model is determined by the understanding of natural rules, we can trust the model more if we know the rules better. Second, even if the model is perfect in replication of the natural rules, it does not mean that we can exactly repeat the natural results. In the stochastic modeling, each run only produces one of the possible results randomly. The evaluation and use of the model should be based on the average results of a large amount runs, i.e. the Monte Carlo experiment. Only the average results can be compared with observation quantitatively.

Taking the example of Hanoi, Vietnam (data from WHO) as an example, Fig. 2 shows the actual daily reports of new patients evolves with time by day. The duration of the epidemics lasted 45 d, and the total number of patients infected is 62. Based on the present knowledge on SARS, the parameters chosen in our stochastic model are: the average latent period is 6.5 d, the average infectious period is 12 d. The infection rate r is defined as the number of people being infected per patient per day (unit: d^{-1}). It is found that the infection rate and its variation with time are the major factor to affect the epidemics. In this work, we made a great number of numerical tests, and by try and error, we found out the best model parameters to fit the constraints of duration of 45 d and accumulated number of 62 patients. The infection rate chosen in the final model is: before any control measures are taken, r_0 = 1.8 d^{-1}, and among the 3rd to 5th days of sickness of the patient, the rate is 5 times higher than the base value; after control measures are taken, the average infection rate r is chosen as 0.01 d^{-1}, and between the 3rd and 5th days, the rate is also 5 times higher as 0.05 d^{-1}. The isolation measures are taken 7 d after the first patient is discovered. Small changes in the infection rate may produce significant changes in the results of modeling. In Vietnam case, an increase of r from 0.01 to 0.015 after control measures

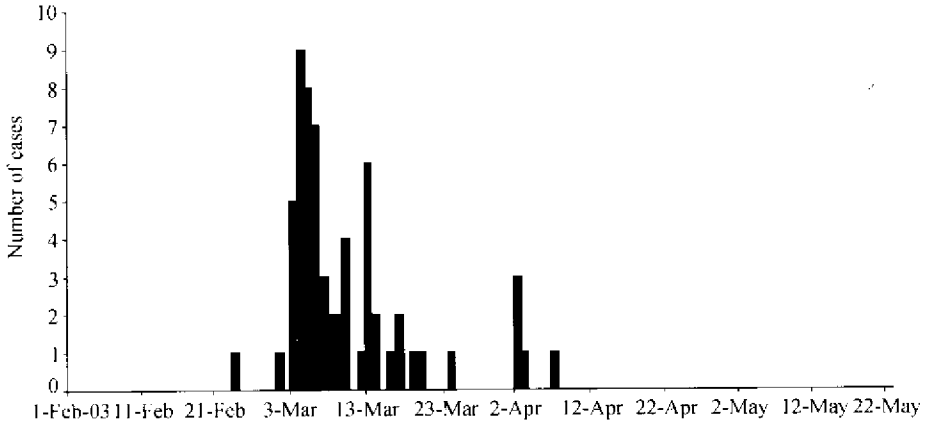

Fig. 2. Daily variation of reported SARS cases in Hanoi, Vietnam (from WHO). n = 62.

were taken can result in an increase of 25% or more in the duration of epidemics and total number of patients.

Fig. 3 shows some typical results as examples of the model. The abscissa is time in day, the ordinate is the number of cases reported each day. Fig. 3(a)—(c) are some typical results similar to the actual case. Fig. 3(a) lasts 44 d of epidemics, infected 59 patients in total; Fig. 3(b) and (c) both last 44 d, and totally infected 67 patients. However, both the duration and the total number of patients may fluctuate significantly. In Fig. 3(d), the duration is 58 d and total number of patients reaches 105. In Fig. 3(e), total number of patients is only 34 within a duration of 31 d. Fig. 3(f) has a duration as long as 102 d, although the total infection is only 42 patients.

All these cases indicate some common features: i) Infected patients increase dramatically without control measures; ii) control measures show their effect not immediately, but with delays, the cases of SARS patients still increase and reach a peak a few days later because of onset of previously infected patients after the latent period; iii) effect of control measures will show up a few days later, and reported cases of SARS can reduce rapidly; iv) random fluctuations are likely to appear in the final stage, and keep the duration of epidemics longer.

These figures also show that fluctuations among different runs of the model can be significant, therefore, in order to understand the overall feature, a lot of test runs have to be carried on, and the averaged results should be used for comparison with observations. We made Monte Carlo tests, and based on 1000 times experiments, it is obtained from the model that duration of the epidemics is (49±15) d, and total number of patients is (61±22), the peak of daily reported cases appears on the (11th ± 3) day, and with a peak value of (8±2) cases, accumulated patients reach (28±13) at this peak day. Fig. 4 shows the frequency distribution of duration of epidemics and total cases reported.

It is a problem of special interest that if such kind of model can provide predictions on the epidemics while it is going on. Assuming that on the 13th d, it is already known that the peak of daily reported cases occurred on the 10th day of 9 cases that day and accumulated cases of 16 up to that day, considering the random fluctuations, we take parts of the results of Monte Carlo tests, which satisfies the constraints that the peak appears on the 9th to 12th day, number of cases reported on the peak day is from 7 to 10 cases, and accumulated cases to the peak day ranges from 15 to 31 cases. Then we observe the subsequent evolution of these series by Monte Carlo modeling. It is found that in the circumstance that the first 12th days' data are known, the conditional probability of subsequent devel- opment of epidemics can be obtained. It is expected that the duration of the epidemics would be (48 ± 13) d, and total number of patients would reach (58±13). Comparing corresponding predictions without knowing the 12th days' data: duration of (49 ± 15) d and total patients of (61± 22). It is found that there are only slight changes in the predicted value, but the estimation of fluctuations has been improved, especially the mean square deviation in the number of total patients reduces significantly. Frequency distribution is shown in Fig. 5, for comparison, the scale of axes is the same as in Fig. 4. The results suggest that as

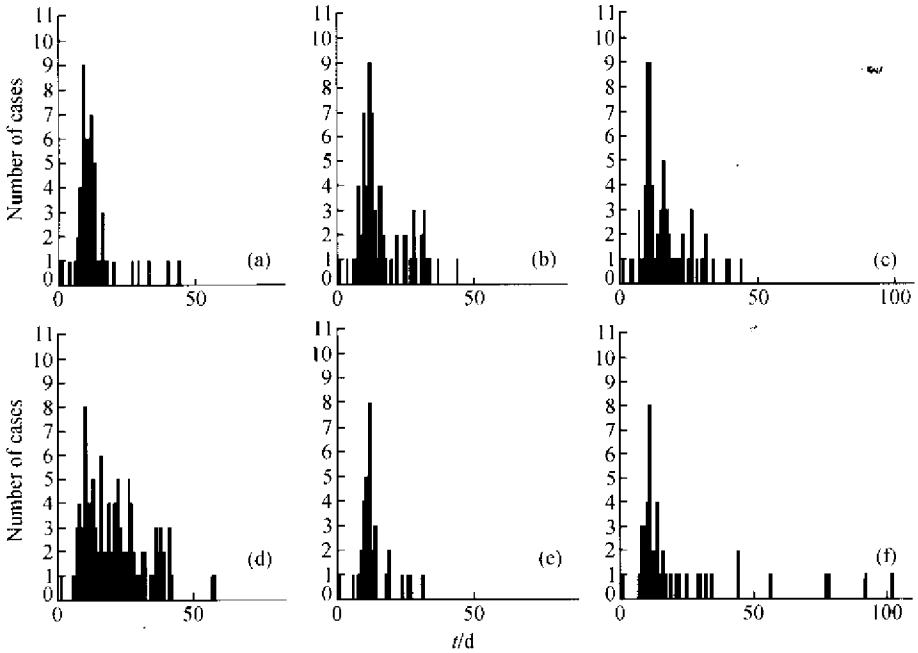

Fig. 3. Some typical examples of the stochastic SARS model.

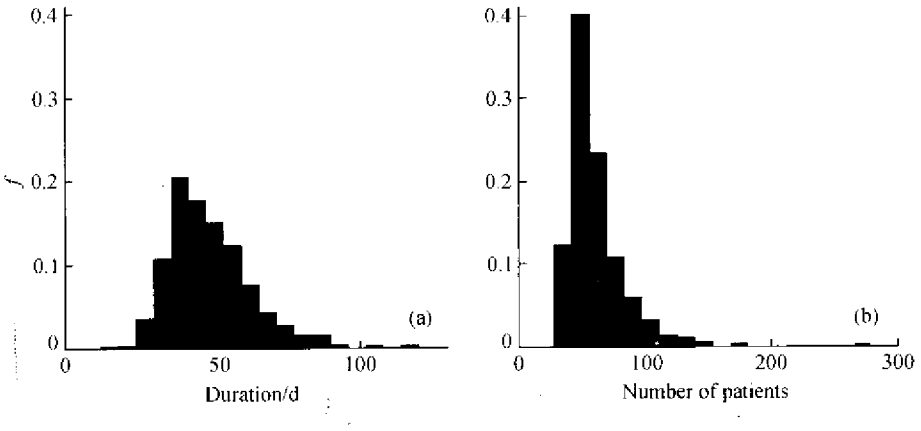

Fig. 4. Results of Monte Carlo tests. (a) Frequency distribution of duration of the epidemics; (b) frequency distribution of number of total SARS patients.

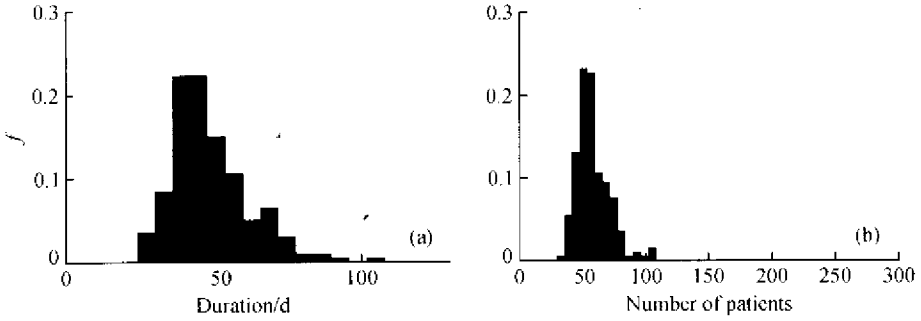

Fig. 5. Monte Carlo tests in condition of the first 12 days' data is similar to the actual case of Vietnam. (a) Frequency distribution of duration of the epidemics; (b) frequency distribution of the number of total SARS patients.

long as the infection rate at different stages can be estimated reasonably, it is possible to make predictions on the duration and total number of patients of the epidemics, providing not only the mean values and most probable values, but also the worst or best situations and the probability of appearance of these extremes. These can play a significant role in fighting against SARS and taking social and economical measures.

3 Discussions and conclusions

In epidemic models, a parameter called reproduction index R, the average number of people being infected by a patient, is considered to be critical. If it is greater than 1 at the beginning of epidemics without precaution measures, denoted as $R_0 > 1$, the epidemics will develop in deterministic models, and develop at a non-zero probability in stochastic models. If control measures are taken to reduce R, and make $R < 1$, the epidemics then can be controlled to decay[18]. In this study, the infection rate r, the number of people infected per patient per day, provides detailed information of daily infection, and the summation of r for all infectious days produces index R. In the discussion of Riley et al.[20], the infected people were divided into two categories: the normal part of infection, and the super-spreading event (SSE). A typical example of SSE was that a SARS patient from Hong Kong infected 25 staff members among the entire 26 staff members in the private French Hospital in Hanoi, Vietnam[26]. Dye and Gay[22] suggested that it is immature to conclude either infections can be divided into two categories as normal events and SSEs, or the so called SSE is just the long tail of a skewed distribution of infection. In this study, it is shown that a unified distribution including both normal events and the so-called SSEs can work well. The so-called super-spreaders just make the average infection rate higher. However, it is worthwhile to try more complex models in future study.

Reducing the infection rate is most important for stopping the spreading of SARS. Early control to reduce infection rate is of top priority. Numerical experiments indicate that the total case in Vietnam is only 62 because they took strict control measures to reduce the infection rate 7 d after the first case of SARS. If the measures are taken one month after the first case of SARS, at reasonable similar parameters, the total cases would be increased to about 1000; if measures are taken 45 d after the first case,

the total cases would reach nearly 3000; and if measures are taken 60 d after the first case, the total number of patients would jump to nearly 9000.

This report is just to provide a test of the suggested methodology of our SARS model, which emphasizes the use of the state of art computation ability to trace infection of each patient and construct a stochastic model for SARS spreading. This model is somehow similar to the Markov chain and Monte Carlo method in the previous studies[19], but it is more flexible to model various complexities. The preliminary test of the example of Vietnam shows the feasibility of this attempt. The model can simulate not only the different stages of SARS spreading (the accelerating stage at first, the delayed appearance of peak infection, the rapid but fluctuated decay under effective control), but also provide quantitative estimation of the duration of epidemics and total number of patients under given model parameters. During the epidemics, if the leading part has become known, it is possible to apply the model for prediction of subsequent evolution of the epidemics. In practical forecast, it is noted that an oversimplified model has no practical values, while an over complex model cannot obtain enough epidemic data to constrain the parameters. Therefore, a trade-off has to be made[19]. This work is preliminary, it needs to collect the detailed data of SARS epidemics and make use of the special capability of tracing every patient in this model, with in-depth comparison of the observation and stochastic modeling, it is possible to develop the present model to contain new findings and details on SARS research as well as the spatial spreading process.

Acknowledgements The author thanks Daryl Daley and David Vere-Jones for helpful discussions, and comments from two anonymous reviewers. This work was supported by the National Natural Science Foundation and the President Fund from the Graduate School of the Chinese Academy of Sciences.

References

1. Anderson, R. M., May, R. M., Infectious Diseases of Humans: Dynamics and Control, Oxford: Oxford Univ. Press, 1991.
2. Wang, W. D., Global behavior of an SEIRS epidemic model with time delays, Applied Mathematics Letters, 2002, 15: 423—428.
3. Wang, W. D., Ma, Z. E., Global dynamics of an epidemic model with time delay, Nonlinear Analysis: Real World Applications, 2002, 3: 365—373.
4. Ma, Z. E., Liu, J. P., Li, J., Stability analysis for differential infectivity epidemic models, Nonlinear Analysis: Real World Applications, 2003, 4: 841—856.
5. Li X. Z., Guper, G., Zhu, G. T., Threshold and stability results for an age-structured SEIR epidemic model, Computers and Mathematics with Applications, 2001, 42: 883—907.
6. Li, J. Q., Ma, Z. E., Qualitative analyses of SIS epidemic model with vaccination varying total population and size, Mathematical and Computer Modelling, 2002, 35: 1235—1243.
7. Fang, B. X., The global existence and uniqueness of the solution of a kind of ODE SIRS model, Journal of Fudan University (Natural Science), 2001, 40(6): 640—687.
8. Zhang, X. Y., Weng, S. Y., Gao, H. Y., An epidemiological model with nonlinear incidence rate and time-lag, Journal of Changchun University, 2000, 10(1): 40—42.
9. Zhang, S. D., Hao, H., Zhang, X. H., Dynamic model of epidemics with latent period, Journal of Mathematical Medicine, 2002, 15(5): 385—386.
10. Fan, A. J., Wang, K. F., Analysis on a species mechanics epidemical model within nonlinear incidence rate, Journal of Sichuan Normal University (Natural Science), 2002, 25(3): 261—263。
11. Yuan, B. Q., The global solutions of nonlinear integral equations on some infection diseases' mathematics model in banach spaces, Journal of Shangdong Normal University (Natural Science), 1997, 12(1): 8—11.
12. Lloyd, A. L., Realistic distributions of infectious periods in epidemic models: Changing patterns of persistence and dynamics, Theoretical Population Biology, 2001, 60: 59—71.
13. Medlock, J., Kot, M., Spreading disease: integro-differential equations old and new, Mathematical Biosciences, 2003, 184(2): 201—222.
14. D' Onofrio, A., Mixed pulse vaccination strategy in epidemic model with realistically distributed infectious and latent times, Applied Mathematics and Computation, 2003, in press.
15. Daley, D. J., Gani, J. M., Epidemic Modelling: An Introduction, Cambridge: Cambridge University Press, 1999.
16. Roberts, M. G., Saha, A. K., The asymptotic behaviour of a logistic epidemic model with stochastic disease transmission, Applied Mathematics Letters, 1999, 12: 37—41.
17. Grasman, J., Stochastic epidemics: the expected duration of the endemic period in higher dimensional models, Mathematical Biosciences, 1998, 152: 13—27.
18. Ball, F., Neal, P., A general model for stochastic SIR epidemics with two levels of mixing, Mathematical Biosciences, 2002, 180: 73—102.
19. O' Neill, P. D., A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods, Mathematical Biosciences, 2002, 180: 103—114.
20. Riley, S., Fraser, C., Donrelly, C. A. et al., Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of Public Health Interventions, Science, 2003, 300: 1961—1966.
21. Lipsitch, M., Cohen, T., Cooper, B. et al., Transmission dynamics and control of severe acute respiratory syndrome, Science, 2003, 300: 1966—1970.
22. Dye, C., Gay, N., Modeling the SARS epidemic, Science, 2003, 300: 1884—1885.
23. Hu, Z. G., Ye, C. S., "Infectious disease model" in new technology dissemination and its empirical evidence, Journal of Wuhan University of Technology, 1998, 20(2): 76—78.
24. Leung, B., Grenfell, B. T., A spatial stochastic model simulating a scabies epidemic and coyote population dynamics, Ecological Modelling, 2003, in press.
25. Shi, Y. L., Liu, J., Vere-Jones, D. et al., Application of mechanical and statistical models to the study of seismicity of synthetic earthquakes and prediction of natural ones, Acta Seismologica Sinica, 1998, 11(4): 421—430.
26. World Health Organization, Communicable disease surveillance and response, severe acute respiratory syndrome (SARS): Status of the outbreak and lessons for the immediate future, Geneva: World Health Organization, 2003.

(Received June 9, 2003; accepted July 4, 2003)