City Environment pulsation research based on Spatial-temporal Data. Mining-Air quality as an example

Xiang Li[1,2,a], Ling Peng[1,b], Tianhe Chi[1], Jing Shao[1,2], Shaolong Cui[1]

[1]Institute of Remote Sensing and Digital Earth, CAS, 100094 Beijing, China

[2]University of Chinese Academy of Sciences, Beijing 100049, China

[a]lixiang01@radi.ac.cn, [b]pengling@radi.ac.cn

Keywords: Spatial-temporal Data Mining; Air Quality; Pulsation; Remote Sensing.

Abstract. Recent years, with our economy sustained and fast development, environmental problems become increasing serious, especially for air pollution problems. Smog problems are intruding more and more big cities. Improving air quality has become an urgent demand of city development. This paper proposed a spatial-temporal data mining based city air quality pulsation analysis method. By introducing remote sensing data source and integrating several spatial-temporal data mining methods, we conducted a deep research about city air quality pulsing features.

Introduction

Along with our country's economy growing continuously and rapidly, a lot of environmental problems comes up, for example, air pollution, water pollution, ecological degradation and other compound environment pollution. These problems has become a vital factor that limits the economy growing and damages people's health and social stability. Environment problems has drawn great attention of people, and we have taken a lot of measures to protect the environment. However, traditional environment monitoring method can't not satisfy our need for environment information collection, management, analysis and decision making, because of limited data source, not enough time-space continuity, single analytical method, bad timeliness and so on [1, 2].

Recent years, with the rapid development of Digital City, Big Data and Cloud Computing technologies, the basic framework and technical architecture of Smart City is gradually towards perfect [3]. The goal of Smart City is to propel city development in a low-carbon, green, harmonious, and sustainable way. Based on the construction requirement of Smart City, exploring the spatial-temporal distributing features of city environment, finding the leading factors and intelligence solutions for environment problems, conducting pulsating analysis for city environment is of great importance. This paper presents a city environment pulsating analysis approach based on multi-source data and spatial-temporal data mining technology, and gives a profound discussion on the air quality pulsating.

Traditional Methods

Traditional environment monitoring methods are often conducted by using the mathematic model to analyze the data of weather station. Recent years, new methods have come up by using remote sensing technology. Take air quality monitoring as an example, general monitoring methods include air pollutant dispersion models and remote sensing monitoring.

Air pollutant dispersion models use the density data of pollutants and meteorological data to analyze the transmission, diffusion features of air pollutants, they are widely used to simulation and

predict the pollutants distribution, so as to evaluate the air quality [4]. Along with the development of computer science, by using the computer to conducting numerical computation, we can solve inhomogeneous and unsteady pollutant diffusion problems, and the application range extends to middle-scale and large-scale. However, because of the complicity of air pollutants diffusion mechanism, these models still have a lot of problems: insufficient meteorological data and pollutants source data declines the simulation precise, the theoretical simulation result is uncertain, studies of dispersion under stable conditions aren't terribly clear-cut, nesting between different patterns is difficult [5, 6].

Remote sensing technology has the features of observing large area synchronously, timely and economically, and it's suitable for long-term dynamic monitoring. It can find some pollutants and dispersion states which can hardly revealed by traditional methods. It can not only monitoring the air environment change and air pollution in a large area in a fast, real-time, dynamic, time-saving way, but also can tracing and monitoring the occurrence and development of air pollution events, so as to take timely measures to reduce losses caused by air pollution. Remote sensing monitoring technique is playing an irreplaceable role in managing air environment and controlling air pollution [7]. However, we have to realize that using remote sensing technique to retrieve atmospheric parameters is a complicated process, the precise and universality of widely used empirical model is not satisfying. So, the present quantification of air environment remote sensing is far from satisfying the needs for environment monitoring [8].

Process and Key Technologies

City environment pulsation analysis means collecting multi-source time-space data, integrating various data mining method to find the pulsation features of the city, explore the internal factors of environment problems. Generally speaking, it include the following process: data acquisition, data preprocessing, data mining and results presenting. City environment pulsation analysis framework shows in figure 1.

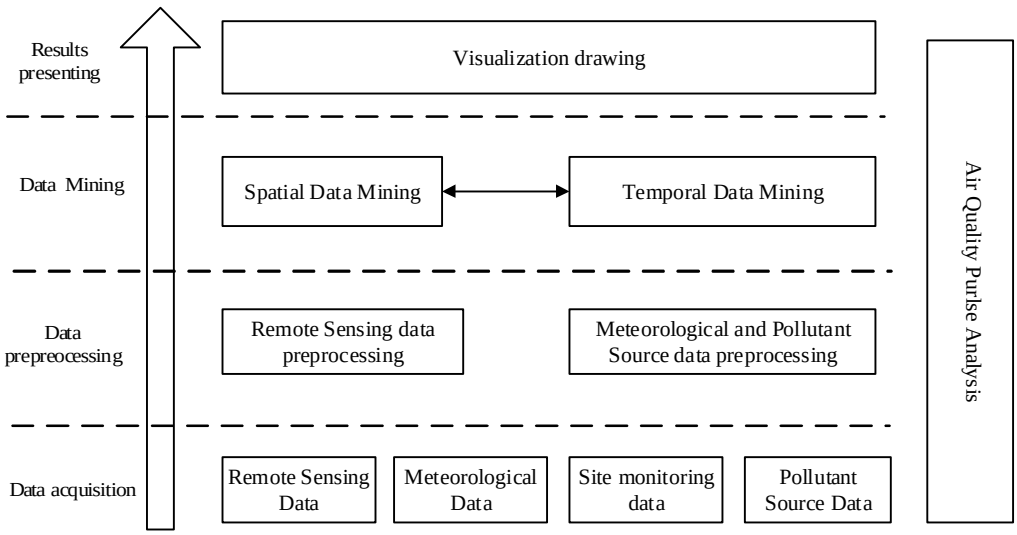

Fig. 1. City environment pulsation analysis framework

Multi-source Data Acquisition

For the single data problem widely exist in air pollutant dispersion models and remote sensing monitoring method, we use multi-source data to conduct air quality pulsation analysis. The data source includes: meteorological data, pollutant data, and remote sensing data.

Meteorological data

We can get various meteorological data by weather station, it includes the density of pollutants, air temperature, humidity, wind speed, wind direction and so on. By means of spatial interpolation and modeling, we can acquire the time-space distribution map of air quality parameters, we use it as the input and verify data.

Remote sensing data

In this paper, we use the EOS/MODIS (Moderate Resolution Imaging Spectroradiometer) as the remote sensing data source. MODIS has some excellent properties, its multi-band data can simultaneously reflect the all kinds of information such as land cover, cloud, moisture, aerosol, land temperature, atmosphere temperature, and ozone, which makes it the most often used data source in atmospheric remote sensing. The MODIS Aerosol Optical Depth (AOD) product is nearly stable after some times of vital version modifying. Systematically verifying has been conducted worldwide, the result shows that the AOD inversion error is about ±(0. 05-0. 15%) compared to ground-based AERONET [9, 10].

Pollutants data

Air pollutants includes point source pollutants, non-point source pollution, mobile source pollutant and biological sources. In the city ecosystem, mainly includes industrial pollution, living pollution, sand and dust, traffic source and so on [4].

As for point source data, apart from location, it should contain the physiochemical parameters relating to the pollutant diffusion. Take a chimney stalk for example, we need the information about location, pollutant type, chimney height, radius, gas temperature, exhaust velocity, total emission of pollutants and so on [4]. As for living pollution data, we apply a spatial grid cut, calculate the total emission inside the grid. For sand and dust data, we build a time-space model according to weather station and wind filed data to get the pollutant distribution. For traffic pollutant, we use the vehicle-mounted GPS tracks data to get the pollutant location, combined with vehicle pollutant emissions data to calculate the pollutant distribution.

Data Preprocessing

To conduct data mining, we should ensure the data quality first. Good data quality can the increase the accuracy and efficiency of data mining. So, data preprocessing is an essential premise and foundation for data mining [11].

MODIS AOD product is HDF format, a kind of hierarchical data format. It storages the longitude-latitude file, but the images have not been geometrical corrected, we should conduct image registration between images and the longitude-latitude file. After that, we should use vector boundary of study area to cut the AOD image data [12].

Weather station data and pollutant data is ambiguous, incomplete, and redundant, and can seldom meet the needs for data mining. Some preprocessing work, includes data cleaning, data integration, data conversion, data reduction shall be done before put them into use [13].

Spatial-temporal Data Mining

Air quality data source has obvious time-space pulsation features, by using spatial-temporal data mining technologies, we can explore their potential rules and patterns and find internal factors which may influence the air quality and the relations between these factors.

Spatial data mining technologies

The theories and methods we choose to conduct spatial data mining will directly affect the quality of discovered knowledge. The main theories and methods of spatial data mining includes probability theory, evidence theory, spatial statistics, rules introduction, clustering analysis, spatial analysis, fuzzy sets, rough sets cloud theory, neural network, genetic algorithms, visualization, decision tree,

spatial online analytical mining. These methods are not isolated applied, to discovery a special kind of knowledge, we often need to integrate several theories and methods according to our special needs. In addition, making good use of machine learning and artificial intelligence to improve automation degree, can decrease the participation degree of human-computer interaction [14].

Time serial data mining technologies

Time serial data mining is an important branches of data mining, it means data mining base on one or more time serial, used to discover potential rules in time serial data [15]. Time serial data mining includes four research field, trend analysis, similitude search, sequential patterns mining and period patterns mining [16].

Results Presenting

Visual mapping can show our data mining knowledge and results in a visual way, make the original abstract results concise, and deepen our understanding about features, relations, patterns and trends about the knowledge.

To directly reflect the pulsation features of air quality, we use statistic graphs and time-space evolutionary charts to show our data mining results. Common statistic tables include tables, histograms, pie charts, and boxplots, and so on. These graph can succinctly reflect statistic information under different scales.

Application Examples

We used the Smart City Cloud Platform of Sino-Singapore Tianjin Eco-city to get various data from Environment Department and remote sensing data from the platform. Then, we integrated several theories and methods of spatial-temporal data mining to analyze the pulsation features of air quality.

Basic Statistics

Figure 2 and figure 3 shows the some basic statistical diagrams using environment data in Tianjin city. Figure 2 shows average pollutants concentration from year 2009 to 2015, and figure 3 shows the API level ratio distribution during this time period.

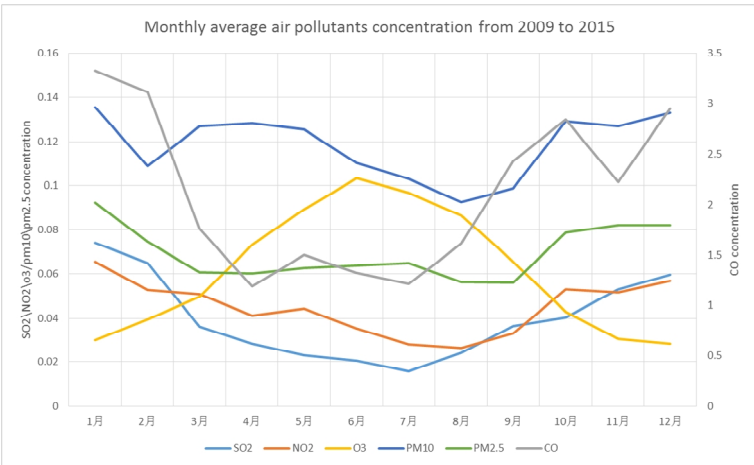

Fig. 2. Average pollutants concentration from 2009 to 2015

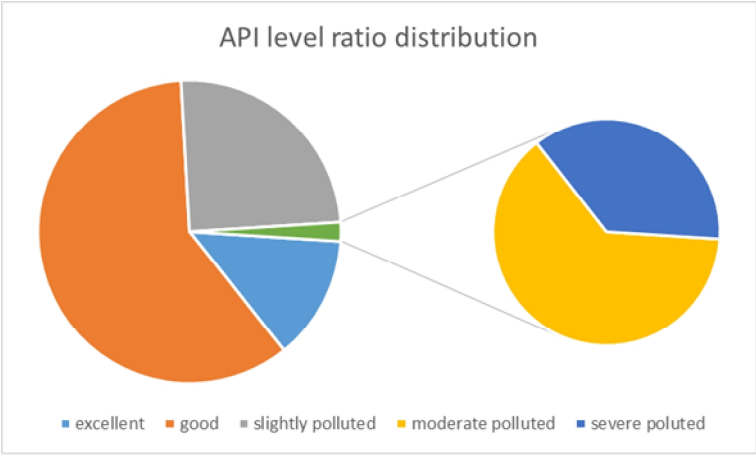

Fig. 3. API Level ratio distribution

Trend Analysis

Using time serial trend analysis to predict the change rules of global and local meteorological parameters. Common time serial trend analysis methods includes neural network, linear and non-linear function, phase space reconstruction theory, support vector machine and so on [17]. Figure 4 shows the prediction result of heat exhaustion at one of the community in Tianjin city using Support Vector Regression (SVR).

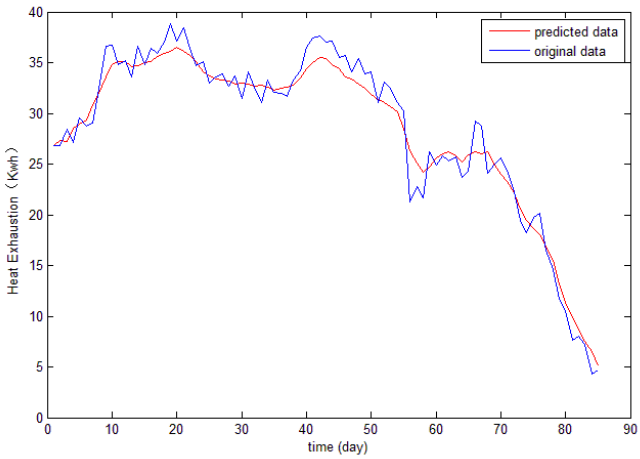

Fig. 4. Heat exhaustion prediction result at one of the community in Tianjin city

Spatial-temporal Interpolation

City weather station are sparse distributed, we should use spatial interpolation to generate spatial distribution diagram. Common spatial-temporal interpolation methods includes BME (Bayesian Maximum Entropy), Spatial-temporal Krging interpolation. Figure 5 shows the result of PM2.5 density map of Tianjin city using Spatial-temporal Kriging interpolation.

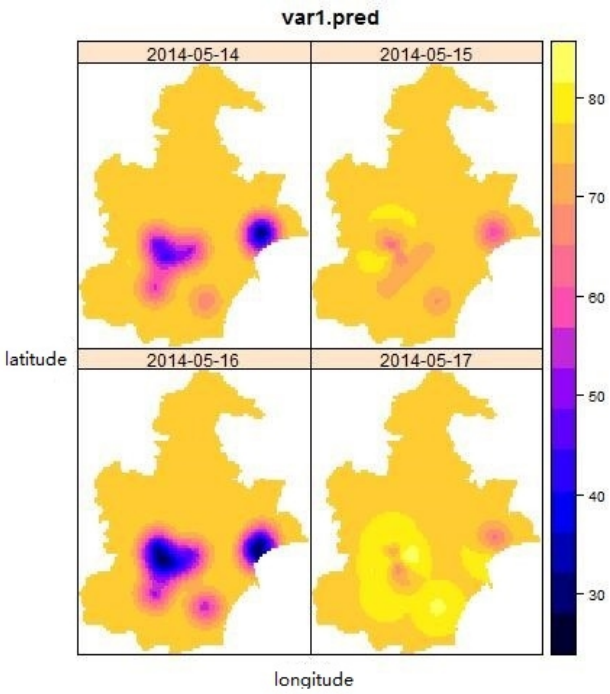

Fig. 5. PM2.5 density map of Tianjin city 2014-5-14 to 2014-5-17

Correlational Analysis

Research shows that MODIS AOD product have a strong relativity with API (Air Pollution Index) [4, 12, 18, 19]. Making the correlation analysis between MODIS AOD data and air quality parameters collected by weather stations, we can build a linear or non-linear regression model, shows in figure 6.

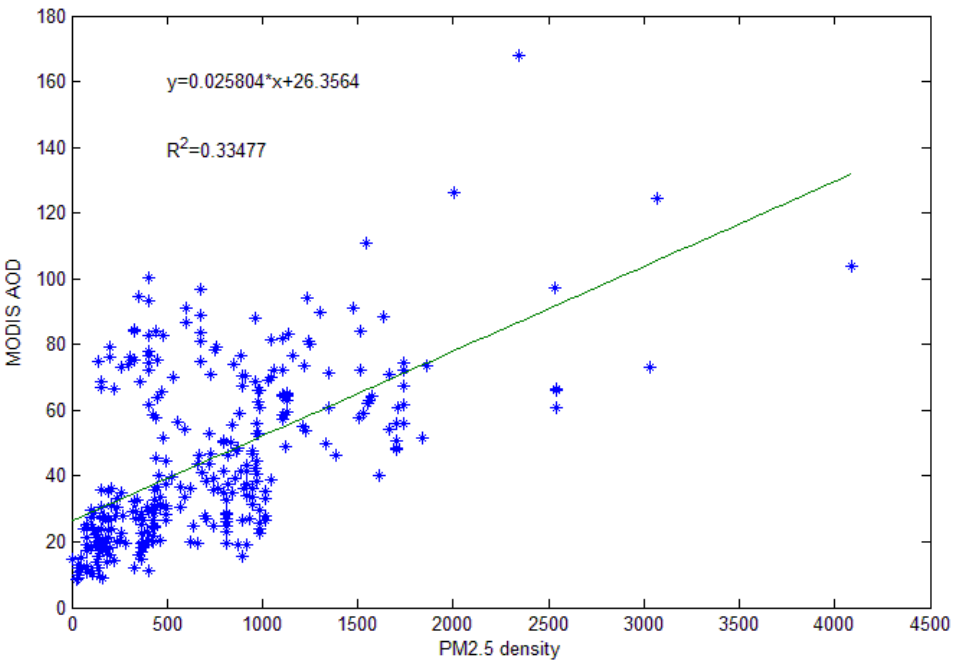

Fig. 6. PM2.5 and MODIS AOD at air monitoring station in Tianjin city

Different air pollutants have some different levels of correlation. Table 1 shows the Pearson correlation coefficient of various air pollutants in 2014.

Table 1. The Pearson correlation coefficient of various air pollutant in 2014

	SO2	NO2	CO	O3	PM10	PM2.5
SO2	1	0.562	0.404	-0.372	0.392	0.442
NO2	0.562	1	0.222	-0.198	0.538	0.604
CO	0.404	0.222	1	-0.313	0.296	0.344
O3	-0.372	-0.198	-0.313	1	-0.09	-0.081
PM10	0.392	0.538	0.296	-0.09	1	0.784
PM2.5	0.442	0.604	0.344	-0.081	0.784	1

Anomaly Detection

Using anomaly detection technologies. Common spatial-temporal anomaly detection methods includes, Dual Deviation-based Spatial-Temporal Outlier Detection (DDSTOD) and SatScan (a Spatial-temporal hot spot scan technology). Figure 7 shows the anomaly detection results of gas consumption at one of the community in Tianjin city, using DDSTOD method.

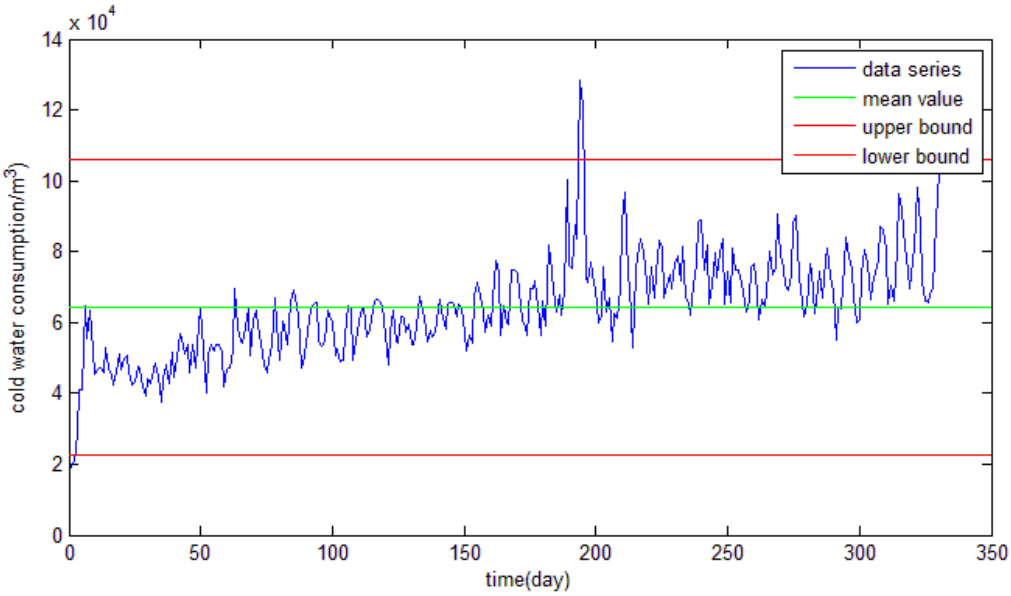

Fig. 7. Anomaly detection results of gas consumption at one of the community in Tianjin city

Conclusion

This paper proposed a spatial-temporal data mining based city air quality pulsation analysis method, and introduced the key technologies and research schemes. Except for traditional data source for air environment research, including weather station data and pollutants data, we introduced remote sensing data. By adopting spatial-temporal data mining techniques, we acquire useful information about city environment pulsation features, and then use visual mapping technique to present our results.

Acknowledgements

This work was financially supported by the National High Tech Research and development program (2012AA12A401).

References

1. Jing, W., Shaanxi Province remote sensing ecological environment monitoring and evaluation based on 3S technology. 2008, Chang'an University. p. 87.(in chinese)
2. Huiying, J. and Z. Biyun, Summarization of Remote Sensing Technology on Environmental Monitoring. Environmental Science Survey, 2008(2): p. 21-24. (in chinese)
3. Derren, L., Y. Yuan and S. Zhenfeng, Big Data in Smart City. Geomatics and Information Science of Wuhan University, 2014. 39(6): p. 7-10. (in chinese)
4. YF, W., Z. HC and M. YJ, Numerical simulation stud of Models3 in air quality of Shenyang. Acta Scientiae Circumstantiae, 2007. 27(3): p. 487-493. (in chinese)
5. Yanyan, C. and Z. Huiyuan, A review of the development and application of air pollutant dispersion models. Environmental Pollution and Control, 2007(5): p. 376-381. (in chinese)
6. Wenbo, X., et al., Domestic and Foreign Research Progress of Air Quality Model. Environment and Sustainable Development, 2013(3): p. 14-20. (in chinese)
7. Ligang, C., W. Yanjiao and W. Yaoting, Summarization of the application of remote sensing in atmospheric environmental monitoring. Environmental Monitoring in China, 2005. 21(5): p. 17-23. (in chinese)
8. QIU, J., H. CHEN and P. WANG, A Prospect on Future Atmospheric Remote Sensing. Chinese Journal of Atmospheric Sciences, 2005. 29(1): p. 131-136. (in chinese)
9. Holben, B.N., et al., AERONET—A federated instrument network and data archive for aerosol characterization. Remote sensing of environment, 1998. 66(1): p. 1-16.
10. Levy, R.C., L.A. Remer and O. Dubovik, Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land. Journal of Geophysical Research: Atmospheres (1984–2012), 2007. 112(D13).
11. GUAN, D., Data Preprocessing in Data Mining. 2006, Jilin University. (in chinese)
12. ZHANG, L., Study on the Correlation between MODIS AOD Products and the Air Pollution in Beijing. 2009, China University of Geosciences (Beijing). (in chinese)
13. Gaohui, P. and W. Zhiliang, Research on Data Preprocessing in Data Mining. 2008, Journal of North China Institute of Water Conservancy and Hydroelectric Power. (in chinese)
14. Derren, L., et al., Theories and Technologies of Spatial Data Mining and Knowledge Discovery. Geomatics and Information Science of Wuhan University, 2007. 27(3): p. 221-233.
15. Yuanyuan, H., Research and Practice of Time Series Data Mining and Visualization. 2010, Donghua University. (in chinese)
16. Han, J. and M. Kamber, Data Mining: Concept and Technology. 2003: Machinery Industry Press.
17. Zhuyan, H., Time Series Data Mining and Its Application in fault diagnosis. 2006, Wuhan University Of Technology. (in chinese)
18. LIU, Y., Study on the Correlation between MODIS Aerosol Optical Thickness and Air Pollution Index in Nanjing. 2007, Nanjing Normal University. (in chinese)
19. LI, C., et al., Using MODIS Aerosol AOD Products to research the air pollution around Beijing City. Chinese Journal of Atmospheric Science, 2003. 27(5): p. 869-880. (in chinese)