APPLICATIONS OF SPATIO-TEMPORAL DATA MINING AND KNOWLEDGE DISCOVERY (STDMKD) FOR FOREST FIRE PREVENTION

T. Cheng J. Wang
School of Geography and Urban Planning
Sun Yat-sun University, Guangzhou, P R China
tao_tao815@hotmail.com cafes123@163.com

KEYWORDS: Data Mining and Knowledge discovery, Spatio-Temporal Data Mining, Forest Fire Prevention, Artificial Neural Network

ABSTRACT:

Forests play an important role for sustaining the natural environment of human living. Forest fires not only destroy natural environment and ecological equivalence, but also threaten security of life and wealth to people. This paper presents applications of Spatio-temporal Data Mining and Knowledge Discovering (STDMKD) for forest fire prevention. The special attention of the research is paid to the spatio-temporal forecasting of forest fires because of the importance of prediction for the fire prevention. It is also due to the fact that most existing spatio-temporal forecasting methods cannot handle the dynamic development of forest fires over space. An improved spatio-temporal integrated forecasting framework – ISTIFF is proposed. The method and algorithm of ISTIFF are presented, which are illustrated by a case study of forest fire area predication in Canada. Comparative analysis of ISTIFF with other methods is implemented, which shows its high accuracy in short-term prediction. Based upon the forecasting result, more intelligent strategies of fire prevention and extinguishments can be delivered to decision makers in fireproofing.

1. INTRODCUTION

Forests play an important role for sustaining the natural environment people live in. Because forest fires are among the best dangers for forest prevention, it is not a surprise to see increasing expenditures for forest fire control. Even so, millions of hectares of forests are still destroyed by fires every year around the world. That this number has not declined implies that controlling fires is a complex task, and indeed forest fires can become as large as 600 km^2 within 9 days and cost millions of dollars to extinguish (Martinus and Junk, 1982). Although most fires are extinguished quickly, a few forest fires become uncontrollable for human intervention after which they cause huge damages to the environment and endanger human lives. For example, Victoria (Australia) fire disaster in 1983, burned 392 000 ha of (grass) land and killed 75 people (Moore and Trevitt, 1991). Therefore, discovering and forecasting forest fire as early as possible is an urgent requirement for forest fire prevention.

In China, forests are a very rare natural resource. Forest fires happen frequently and the loss is very serious each year. It not only destroys natural environment and ecological equivalence, but also directly influences the production of industry and agriculture, and seriously threatens security of life and wealth to people (Zhang, 2004). In order to predict, detect and control forest fire, forest fire prevention information system is urgently needed. In 2002, the National Bureau of Forestry of China has invested twenty millions of RMB in building such forest fire prevention information system. The system plays a very important role in forest fire detection and fire extinguishing.

To build the forest fire prevention information system, not only china, but also other countries such as Canada have built data bases to record forest fires and relevant weather information, which accumulates huge spatio-temporal data. Due to the lack of platform and tools to mine spatio-temporal data, it is difficult to adequately make use of the data for forest fire prevention. Therefore, advanced techniques of spatio-temporal data analysis and data mining should be applied to extract implicit knowledge, spatial and temporal relationships or other patterns not explicitly stored in the system, in order to enhance the intelligence of systems and to facilitate decision-making.

In the following sections, we first introduce spatio-temporal data mining techniques for forest fire prevention. Then we present a spatio-temporal forecasting approach, an improved spatio-temporal forecasting framework (ISTIFF) for dynamic process changing over space (such as forest fire). Finally, we carry out forest fire area forecasting in Alberta of Canada by using the proposed approach. It is shown in our case study that ISTIFF has high accuracy for dynamic forecasting, which provides a very useful tool for forest fire forecasting.

2. SPATIO-TEMPORAL DATA MINING TECHNIQUES FOR FOREST FIRE PREVENTION

Spatio-temporal data mining is the extraction of unknown and implicit knowledge, structures, spatio-temporal relationships, or patterns not explicitly stored in spatio-temporal databases (Yao, 2003). Spatio-temporal data mining techniques and tasks include spatio-temporal forecast and trend analysis, spatio-temporal association rule mining, spatio-temporal sequential patterns mining, spatio-temporal clustering and classification and so on. Difficulty of spatio-temporal data mining relies on how to integrate space and time seamlessly and simultaneously.

The techniques of spatio-temporal mining can be applied for forest fire prevention as follows.

1) Spatio-Temporal Forecasting and Trend Analysis

Spatio-temporal forecasting and trend analysis technique is an effective means of forecasting spatial attribute. Spatio-temporal forecasting has been developed from individual spatial or temporal forecasting and gained heavy attention for its promising performance in handling complex data in which not only spatial but also temporal characteristics must be taken into account. The key issue is to integrate space and time. Some mature analysis tools, e.g., time series or spatial statistics are extended to spatio-temporal problem.

Based upon the information of the forest fires in the past, we are able to predict the forest fires in future, so as to facilitate the fire prevention. Spatio-temporal forecasting and trend analysis technique can predicate the speed of fire spreading, the trend of fire spreading, the area and length of fire field, so as to provide real time optimised fighting plan to minimize the total cost due to the fire.

2) Spatio-temporal Association Rule Mining

Association Rule Mining has been one of the most extensively studied data mining techniques. A spatio-temporal association rule is an implication of strong association between *A* and *B* with the form A $\rightarrow$ B, where *A* and *B* are sets of spatio-temporal and non-spatio-temporal attributes. The implication carries the meaning that if the attributes at *A* take some specific value at a point in time, then with a certain probability, at the same point in time, the attributes at *B* will take some specific value (Gidofalvi, 2004). A spatio-temporal association rule might find that "If forest fire has been found at *A* at a specific time, at the meanwhile, it might be quite possible that fire will occur at *B*."

Despite the abundance of spatio-temporal data, the number of algorithms that mine such data is few. The main reason for the lack of efficient algorithms is due to the exponential explosion in the search space for knowledge caused by the added spatial and temporal attributes. Existing attempts is to modify classical rule mining method to spatio-temporal association rule, thus losing the spatio-temporal characteristics (Verhein and Chawla, 2005).

Spatio-temporal association rule technique might be able to identify the relationship of locations *A* with *B* over time change between weather conditions (wind speed, wind direction, temperature, humidity), forest fuel type, and geographical conditions (degree of slope, aspect of slope, position of slope).

3) Spatio-temporal Sequential Patterns Mining

The task of mining spatio-temporal patterns is to find out sequences of events (an ordered list of item sets) that occur frequently in spatio-temporal datasets. A spatio-temporal sequential pattern has the form A $\rightarrow$ B, where A and B are sets of spatio-temporal and non-spatio-temporal attributes, meaning that if at some point in time and space, the attributes in *A* take some specific value, then with a certain probability at some later point in time, attributes in *B* will take some specific value (Gidofalvi, 2004). A spatio-temporal sequential pattern might be that "If forest fire has been found at *A*, it might be quite possible that fire will occur at *B* in two hours at certain weather condition".

The sequential pattern mining algorithms were first introduced by Agrawal and Srikant in 1995 (Agrawal and Srikant, 1995). Six years later, Tsoukatos and Gunopulos extended these methods by adding spatial dimension. An efficient DFS (Depth First Searching) algorithm was proposed to discover spatio-temporal sequential patterns for weather prediction.

By spatio-temporal sequential patterns mining technique we may discovery a spatio-temporal sequential pattern that tells, "Forest fire always occurs at region R1 prior to the occurrence of haze in nearby region R2."

4) Spatio-Temporal Cluster Characteristic and Discriminate Rule Mining

Cluster characteristic or discriminate rules associate objects belonging to a cluster of some attributes with some probability. A spatio-temporal clustering might discover that "The grid cells with similar values in meteorological satellite image at noon, can be clustered as a spot at high temperature, tending to be a fire spot". Widely used spatial clustering techniques e.g., *K*-means and *K*-medoids and CLARANS (Han and Kamber, 2001), may be extended for spatio-temporal clustering.

In addition to discriminate the fire spot, spatio-temporal cluster characteristic and discriminate mining technique might be able to classify the fire risk ranking and predict the probability of forest fire.

3. SPATIO-TEMPORAL FORECASTING OF FOREST FIRE AREA

3.1 Principle

As one of data mining techniques, forecasting is widely used to predict the unknown future based upon the patterns hidden in the current and past data. Due to the increasing demand for spatio-temporal data mining in many application fields, many spatio-temporal forecasting models are ppproposed. These forecasting models are based on mature analysis tools, e.g., time series or spatial statistics, which are extend to spatial or temporal aspect, respectively. For example, spatial statistics concepts were extended to take the time dimension into account in (Cressie, and Majure, 1997; Pokrajac and Obradovic, 2001). On the contrary, Deutsch et al. incorporated spatial correlation into the multivariate time series analysis with the help of a neighboring distance matrix (Deutsch and Ramos, 1986; Pfeifer, and Deutsch, 1990).

Recently, Li and Dunham proposed a spatio-temporal integrated framework - STIFF, which is applied to forecast the water flow rate at gauging station in the catchment (Li and Dunham, 2002). In STIFF, time series analysis strategy is incorporated to capture the temporal correlations and the artificial neural network technique is employed to discover the hidden and deeply entangled spatial relationships, then the two mechanisms are combined via regression to generate the overall forecasting. It overcomes deficiency of previous works by loosening their stringent assumptions and excessive simplification.

However, STIFF approach is insufficient in forecasting forest fire because forest fire is a dynamic process developing over space, which cannot be handled by a static forward neural network based on BP algorithm that STIFF employed. Elman is a kind of dynamic recurrent neural network (RNN). This recurrent connection allows the network to both detect and generate time-varying patterns as well as spatial-varying patterns. Because the network can store information for future reference, it is able to learn temporal patterns as well as spatial patterns. Therefore, we use Elman network to produce spatial forecasting. To differentiate from STIFF, we call our approach as ISTIFF, i.e. improved STIFF, due to its improved ability of detecting spatial-varying patterns and the improved accuracy of forecasting.

The key idea of spatio-temporal forecasting is as follows: constructing a stochastic time series models to capture the temporal characteristics of each spatially independent subcomponent, then building an dynamic recurrent neural network (RNN) to discover the hidden spatial correlation, finally combining the previous individual temporal and spatial forecasts based upon statistical regression to procure the final forecasting result.

3.2 Problem Definition

The spatio-temporal forecasting can be formally defined as follows:

1. The research area Δ is composed of $n+1$ subcomponents denoted by $\alpha_{0,}\alpha_{1,}\cdots_{,}\alpha_n$, which can be spatially separated from each other. Without loss of generality, α_0 is assumed to be the only target subcomponent where the spatio-temporal forecasting will be conducted.

2. For each $\alpha_i \in \Delta$, there are j time series observations $\alpha_{i1,}\alpha_{i2,}\cdots_{,}\alpha_{ij,}$ that are recorded as $\prod_i$ for convenience.

3. Given the collection of subcomponents $\Delta = \{\alpha_{0,}\alpha_{1,}\cdots_{,}\alpha_n\}$, the whole available dataset $\prod = \{\prod_{0,}\prod_{1,}\cdots_{,}\prod_n\}$ and the look-ahead steps of s, the problem asks to find a mapping relationship f, defined as

$$f:\{\Delta,\Pi,l,s\} \;=\; f\;\{\sigma_{0(l+1),}\sigma_{0(l+2),\cdots,}\sigma_{0(l+s),}\} \qquad (1)$$

which should be as precise as possible, where l （$=|a_i|$）, σ_{ij} (i= 0, 1,…, *n*; *j* = 0, 1,…, *l*) is the j^{th} observation in time series data σ_i.

3.3 Algorithm

The problem defined above can be solved by the algorithm with the following steps:

- *Setp 1*: Define the forecasting problem in terms of the specification by determining the target subcomponent α_0 and its spatially-correlated siblings $\alpha_{1,}\alpha_{2,}\cdots_{,}\alpha_n$.
- *Setp2*: For each subcomponent $\alpha_i \in \Delta$ build a time series model TS_i that will implement the needed temporal forecasting for each subcomponent. Specifically, temporal forecasting for the target subcomponent α_0 is denoted as f_T instead of TS_0 to differentiate from other subcomponents.
- *Setp3*: Based upon the spatial correlation of all non-target subcomponents $\alpha_i\ (i=1,...,n)$, an artificial neural network is built to capture the spatial influence of all non-target subcomponents over the target subcomponent. The network first gets trained and adjusted accordingly. Then forecasts from each time series model $TS_i(i \neq 0)$ are fed into the network. The spatial forecasts at α_0, identified as f_S, can be finally obtained from the network output.
- *Setp4:* The individual spatial forecast, f_T, and the temporal forecast, $f_{S,}$, will be merged together, mostly via a statistical regression mechanism, to generate the final spatio-temporal forecast.

The detailed implementation of the algorithm is illustrated in the next section. The novelty in our approach lies at Step 3, i.e. a dynamic recurrent neural network is applied (please refer to Section 4.2 for detailed implementation), which overcomes the shortcomings in STIFF.

4. CASE STUDY

A practical spatio-temporal forecasting is presented to explain the novel approach discussed in the previous section and to show how accurate the overall forecasting could be.

The case study is based upon the data kindly provided by CFS (Canada Forest Service) (CFS). LFDB (Canada Large Fire Database) records large forest fires of which areas exceed two hundred hektare from 1959 to 1999 and covers every province, region and forest park in Canada. Spatial relationship is given in Figure 1 for the study area. Alberta (AB) province is where the forecasting will be carried out. In other words, it is the target location a_0 in term of the problem definition. We are going to forecast the area (ha^3/month/year) at Alberta (AB), the target subcomponent. The neighbouring (spatially correlated) provinces, the non-target subcomponents, are British Columbia (BC), Saskatchewan (SK), Manitoba (MB), Ontario (ON), Northwest Territories (NWT) and Quebec (QC).

Figure 1 Spatial Distribution of Provinces in Canada

4.1 Time Series Model for Temporal Forecasting

Fist of all, original data is analysed and we find some missing value in original data. After determining the integrality of data, we choose data between 01/1959 and 12/1988 as the training to forecast the data between 01/1989 and 12/1999. A fraction plot of time series data between 01/1959 and 12/1988 for Alberta (AB) province is given in Figure 2.

Because the time series is not steady, we transform the data to steady sequence according to difference method. After the data has been appropriately transformed according to the logarithmic method, autocorrelations and spectral density plot is given in Figure 3.

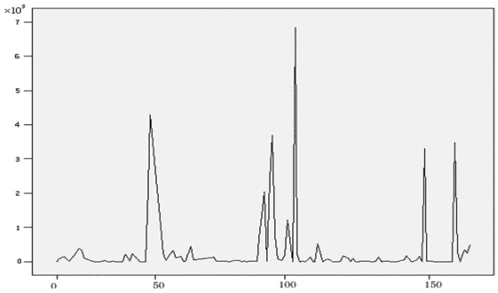


Figure 2 Time series plot

Then, we use the ARIMA (Autoregressive Integrated Moving Average Model) model to forecast the data between 01/1989 and 12/1999 at each province, i.e. time series forecasting values for f_T for the target subcomponent and TS_i (i=1, …, n) for the non-target subcomponents .

4.2 Artificial Neural Network for Spatial Forecasting

First of all, input and output of neural network should been confirmed. Because Alberta province is target location and there are 6 non-target locations British Columbia (BC), Saskatchewan (SK), Manitoba (MB), Ontario (ON), Northwest Territories (NWT) and Quebec (QC), the network would be in a $6-X-1$ structure as shown in Figure 4. That is, there are 6 input neurons, and unknown number (X) of neurons in the hidden layer, and one neuron in the output layer. In order to find an optimal number of the hidden layer, we vary the number of neurons in the hidden layer from 6 to 13 to train network using the data of 6 neighbouring provinces from 01/1959 to 12/1988. It turns out 6 neurons in the hidden layer has the best performance during the training stage. As a result 6 is picked up for its most simple structure. Thus the condensed network with 6 neurons in the hidden layer is finally chosen as the one used to find the spatial forecasting f_S .

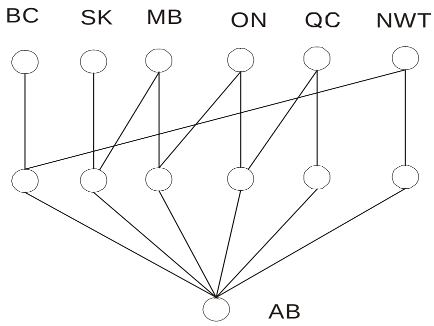


Figure 4 The structure of the recurrent neural network

Because there are six non-target locations, which is closely related to the target spatially, we use Elman with 6 input, 6 hidden layer nodes and 1 output node, as learning model. We will construct Elman network, where the input of two networks is the data of 6 neighbouring in past 30 years with 6 groups. Besides, it is important to select a proper stimulation function. We choose transig function for the hidden layer as stimulation function, whose output range is larger than that of the logsig function. We choose linear purelin function as the stimulation function for the output layer whose output may be of any value. The learning rate is 0.01. The training goal is reached after 20 times of learning. The spatial forecasts reached are f_S . To compare with STIFF, we also used BP (back propagation) network (which was employed in STIFF) for the same predication, denoted as $f_s^{'}$.

4.3 Overall Spatio-temporal Forecasting

So far, we obtained temporal forecasting f_T and spatial forecasting f_S for the target location respectively. Our goal is to produce an optimal overall spatio-temporal forecasting $f_{overall}$. Therefore, we use linear regression to fuse f_T and f_S :

$$f_{overall} = x_1 \times f_T + x_2 \times f_S + \mathrm{Re}\, gression\_Cons\tan t \qquad (2)$$

where both the regression coefficients, include x_1 and x_2 , and regression constant, *t*, have to be estimated beforehand. Before we carry out regression analysis, trend between observation variables is estimated through scatter plotting. Two scatter plots (f_T , $f_{overall}$) and (f_S , $f_{overall}$) are shown in Figure 7.

From Figure 7 we can see the obvious linear trend for temporal forecasting f_T and infirm linear trend for spatial forecasting f_S . It means that time forecasting f_T will occupy more specific gravity than spatial forecasting f_S in Equation 2. The regression coefficient is acquired after analysing the regression, which is 2.538 and 0.976 respectively for x_1 and x_2 .

The spatio-temporal forecasting result is basically identical with the real value. At the same time, we compute $f_{overall}$ ' using f_S ' based on STIFF. The forecasting results of our approach – ISTIFF and STIFF are compared with the real data (Figure 8). In order to see the advantage of spatio-temporal forecasting, we also report the results of temporal forecasting by ARIMA. The errors occurred in the three methods are reported in Table 1.

Table 1 Forecasting Errors of Difference Methods

Model	Average absolute error	Average relative error
ISTIFF	1.34	0.65
STIFF	1.97	0.89
ARIMA	3.78	1.87

The comparison for ARIMA, STIFF and ISTIFF shows that ISTIFF method can achieve better forecasting accuracy than STIFF, which is better than ARIMA. It implies that Elman network in ISTIFF obtains better spatial forecasting than BP network in STIFF. It also indicates that spatio-temporal forecasting is better than pure time series analysis for spatio-temporal data.

5. DISCUSSION AND CONCLUSIONS

This paper presents application of Spatio-temporal Data Mining and Knowledge Discovering (STDMKD) for forest fire prevention. The special attention of the research is paid to the spatio-temporal forecasting of forest fire. An improved spatio-temproal integrated forecasting framework – ISTIFF is proposed, which has been illustrated by a case study of forest fire area predication in Canada. Comparative analysis of ISTIFF with ARIMA and STIFF shows the high predication accuracy of ISTIFF. Based upon the forecasting result, more intelligent strategies of putting out the fires can be delivered to decision maker in fireproofing.

REFERENCE

CFS: http://fire.cfs.nrcan.gc.ca/research/climate_change/lfdb/lfdb_download_e.htm.

Cressie, N. and Majure, J. J. 1997, Spatio-temporal statistical modeling of livestock waste in streams. *Journal of Agricultural, Biological and Environmental Statistics*, 2, pp.24-47.

Deutsch, S. J. and Ramos, J. A., 1986, Space-time modeling of vector hydrologic sequences, *Water Resource Bulletin*, Vol. 22, pp. 967-981.

Gidofalvi, G., 2004, Spatio-temporal Data Mining for Location-based Services, Industrial Ph.D. study proposal at Aalborg University, accessed at http://www.cs.aau.dk/~gyg/docs/STDM.PDF.

Han, J. and Kamber, M., 2001, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers.

Li, Z., Dunham, M. H., 2002, STIFF: A forecasting framework for spatio-temporal data, *International Workshop on Knowledge Discovery in Multimedia and Complex Data (KDMCD 2002)*, pp. 183-188.

Moore, P.F., Trevitt, A.C.F, 1991, Computers in fire management: Limitations of the mechanistic approach. In Andrews, P.L. and Potts, D.F., editors, *Proceedings of the 11th Conference on Fire and Forest Meteorology,* pp.98-108.

Pfeifer, P. E. and Deutsch, S. J., 1990, A statima model-building procedure with application to description and regional forecasting. *Journal of Forecasting,* 9, pp.50～59.

Pokrajac, D. and Obradovic, Z., 2001. Improved spatial-temporal forecasting through modeling of spatial residuals in recent history. *In First SIAM International Conference on Data Mining (SDM'2001), Chicago, April 5-7, paper No. 9*

Agrawal, R. and Srikant, R., 1995, Mining sequential patterns. *Proceedings of the International Conference on Data Engineering (ICDE'95),* pp. 3-14.

Martinus, N. and Junk, W., 1982, Forest fires in North America. In Forest fire prevention and control: proceedings of an international seminar organized by the Timber Committee of the United Nations Economic Commission for Europe, Warsaw, Poland, May 20 – 22, pp.101-108.

Tsoukatos, I. and Gunopulos, D., 2001, Efficient mining of spatio-temporal patterns, *7th International Symposium on Spatial and Temporal Databases (SSTD),* California, July 12-15, pp. 425-442.

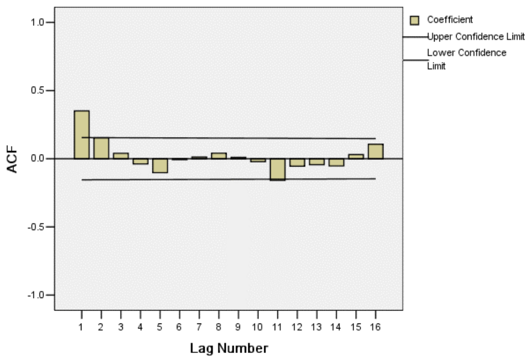
Verhein, F and Chawla, S., 2005, Mining Spatio-Temporal Association Rules, Sources, Sinks, Stationary Regions and Thoroughfares in Object Mobility Databases, *Technical Report, University of Sydney, Number 574.*

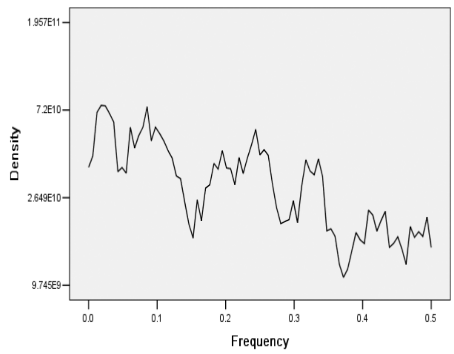
Yao, X., 2003, Research issues in spatio-temporal data mining, A white paper submitted to the University Consortium for Geographic Information Science (UCGIS), workshop on Geospatial Visualization and Knowledge Discovery, Lansdowne, Virginia, Nov. 18-20.

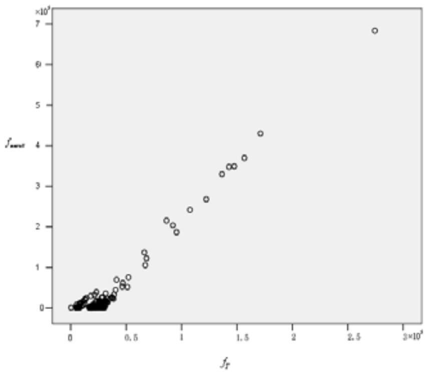
Zhang, G, 2004, Research on forest fire dynamic monitoring in Guangzhou City, *PhD Dissertation, Central South Forestry University, Changsha, P.R China.*

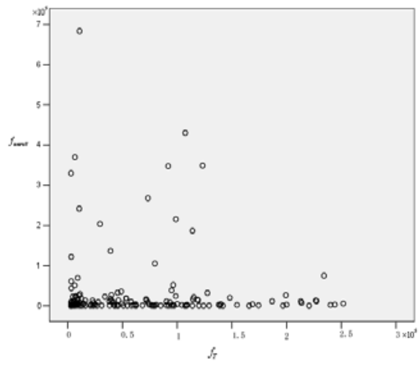
ACKNOWLEDGEMENTS

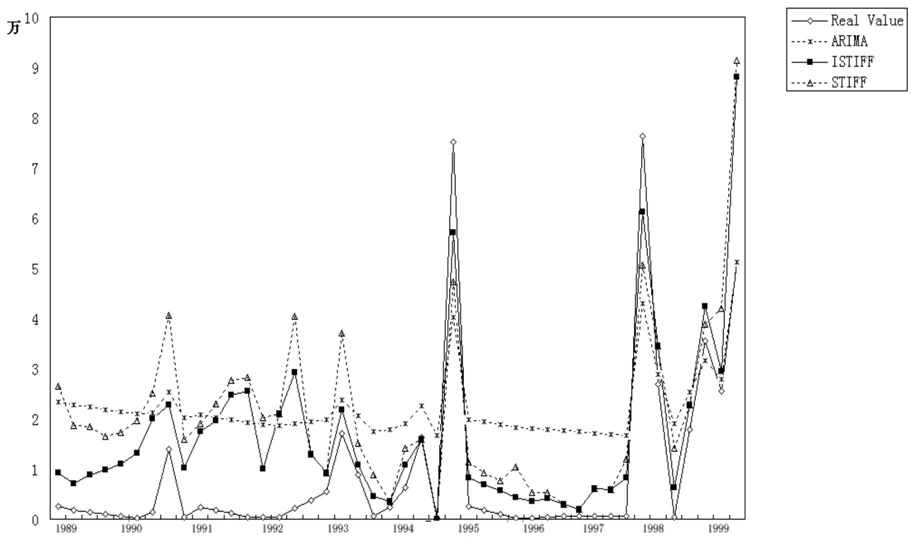
The research is supported by the Major State Basic Research Development Program of China (973 Program, no. 2006CB701306) and the Ministry of Education of China (985 Project, No. 105203200400006).


Coefficient
Upper Confidence Limit
Lower Confidence Limit
ACF
Lag Number


1.957E11
7.2E10
2.649E10
9.745E9
Density
Frequency






万
Real Value
ARIMA
ISTIFF
STIFF