

# DESIGN AND IMPLEMENTATION OF CENSUS SPATIO-TEMPORAL DATABASE BASED ON MODELS OF BASE STATE WITH AMENDMENTS

Ma Weijun , Liu Deqin

Chinese Academe of Surveying and Mapping

Beijing 100039, China

[weijunma@casm.ac.cn](mailto:weijunma@casm.ac.cn)

## [Abstract]

The advantages and disadvantages of many kinds of spatio-temporal data models, especially the data models of base state with amendments, were introduced in this paper. A new spatio-temporal data model of base state with amendments was presented. Based on the model, a population census spatio-temporal database of China was established. The experiment shows that the database has the ability to manage the data of different historical time efficiently .

**Key words:** base state with amendments; spatio-temporal data model; census; spatio-temporal database

## 1. Introduction

Census is a main method that a country used to collect population data .Most of countries in the world has established policy of census. In 1953,1964,1982,1990 and 2000,five census have been performed since People Republic of China was established in 1949. Census data was gathered and reported by 7 hierarchical levels, namely state, province, district, county, township, census district and Enumeration district. The top 5 levels are consistent with the administrative district, the last tow level districts are subdivided based on the higher level district.

So far,China has gathered lots of population datum of different time .It is necessary to establish a database to manage these datum ,not only the population attribute information,but also the temporal and spatial information. The study of managing these population datum of different time is now a hot issue in China. A new spatio-temporal data model of base state with amendments was presented in this paper. Based on the model, a population census spatio-temporal database was established.

## **2. Why to establish census spatio-temporal database**

Spatial-Temporal Database, which combines the temporal database and the spatial database, has the capacity to store and manage temporal,spatial and attribute data simultaneously. The significance of building population census spatio-temporal database are:

- a. Serving the population census;
  - subdividing census district more quickly;
  - guarantee the higher quality of the census data
  - managing the census data more efficiently
- b. Providing a method to get the population data of different time quickly  
Because some census district in China is changing annually,it is too difficult to get the history population data of a region . Based on the census spatio-temporal database ,population data of anytime ,anywhere can be recovered expediently.
- c. Providing a method to study the rule of population spatio-temporal distributing and the change of population spatio-temporal distributing.

## **3.The development of studying of spatio-temporal data model**

Spatio-temporal data model is a spatial temporal integrated data model . It has the capability in organizing and managing temporal geo-information efficiently,it is a information data model with more integrated attribute,spatial and temporal information[Wang Yingjie etc 2003].It is the base of constructing spatio-temporal database.Many kinds of spatio-temporal data models have been put forward.

Snapshot model , which combined time information and spatial data model by adding time stamp on map layer , is the simplest spatio-temporal data model.The obvious disadvantage of this model is the redundance of data recorded,and the changed information has not been recorded directly;

The principle of spatial temporal composite (STC) data model is to project every line of spatial temporal on a space plane,the lines intersected each other,formed lots of grid.Each grid has its own attributes.After each spatio-temporal change,lots of grids with independence attributes is formed.This data model hold out many kinds of spatio-temporal queries,but it is difficult to querying spatio-temporal relation and spatio-temporal action[Liu Renyi etc 2000,Zhang Zuxun etc 1996, N.Pelekis].

Trifield model(M.Yuan) describes geo-processing and geo-phenomena in expressing semantic,time and spatial separately,and providing relationship among them[M.Yuan

1994].

Historical graph model divides object into three classes: continuously changing object, abruptly changing object and never changing object. The purpose is to identify time action, manage objects and events [A. Renolen 1996].

Based on entity relation (ER) model, Tryfona put forward Spatio-Temporal ER (STER) model.

All models above do not come down to change treatment, so, C. Claramunt, C. Parent put forward Object-Relationship Model [C. Claramunt etc 1995]. Based on object-oriented technology, many scholars put forward object-oriented spatio-temporal data model [M.-A. Aufauer-Parent 1995, C. Bonfatti 1999, A. Raza 1999]. Aimed at the moving object, M. Erwig, R. H. Gutting, L. Forlizzi have studied the moving object spatio-temporal data model [M. Erwig etc 1999, R. H. Gutting 1999, L. Forlizzi etc 2000].

Zhang Z, Zhang B and Liu R have studied base state with amendments data models, and gave many different base state with amendments extended data models [Liu Renyi etc 2000, Zhang Zuxun 1996, Zhang Baogang 2005].

Zheng put forward a spatio-temporal data model which based on state and change [Zheng Kougen ]; Chen J and Jiang J put forward a spatio-temporal data model based on event [J. Chen 1998].

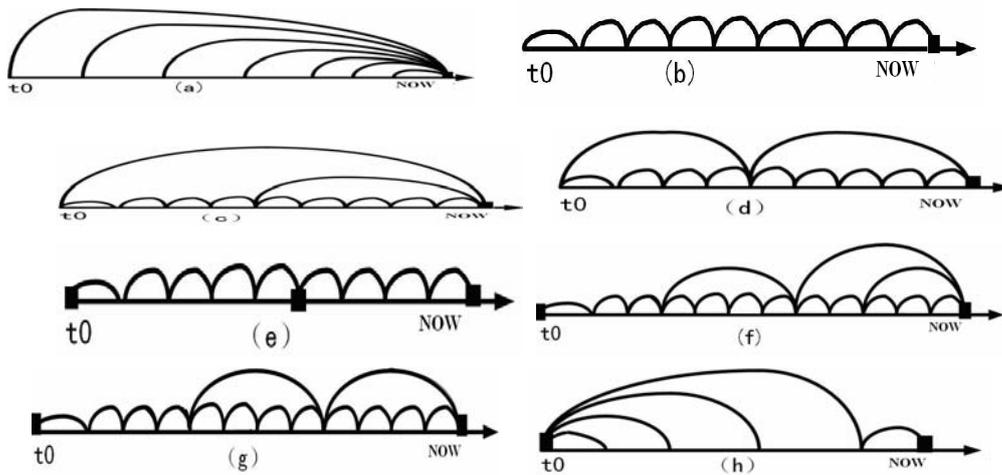
Each model above has advantages and disadvantages. The data models of base state with amendments were introduced particularly in this paper.

#### **4. Base state with amendments data model and extended models**

With the data models of base state with amendments, only the data of one time point (base state), which is called the base file, and the changed data of other times, which are called the delta files, are saved in the database [Liu Renyi 2001].

Because of the size of delta file is far smaller than the base file, the size of database based on the data model of base state with amendments is far smaller than the usual database.

Many kinds of data models of base state with amendments have been put forward.



**Figure 1 Data Models of Base State With Amendments and Extended Models**

With model(a),the current state is looked as base state,Any changed data related to the current time are saved as delta file in database.Every delta files mustbe modified when the base file changed.Along with the numbers of detal files rising,the workload of modifying the delta files would be very heavy.

With model(b),the current state is looked as base state.The changed data between two time points in sequence is saved in database as delta files.The disadvantages of model(a) is avoided.Only one delta file be created and the other delta files without any change when the base state has changed.The disadvantage of this model is When serching for any history data,many dotal files are used . The more time point there, the more heavy the searching workload .

Model(c) and (d) (Zhang Z and Huang M)are extended models[Zhang Zuxun 1996].Multilevel index files are added into these models.The efficiency of searching for history data is improved evidently[Cao Zhiyue 2000].

With model(e) [Cao Zhiyue 2000],many base files are saved in database.The number of delta files between two base state are called Base State Distance. The searching speed is improved sharply. Because many base files are saved in the database ,its size is increased.

Zhang B etc combined model(c),(d)and (e) to construct model(f) and(g).with model (f) and (g),The state of the initial time and the current are saved as base file , besides delta files between any two sequent time point,3 other delta files are added into the

database. Compared with the model (c),(d) and (e), the size of database increased little, the searching speed improved sharply. The same as the model (c),(d) and (e), along with the number of time point increasing, the size of the database increasing.

Based on the disadvantage of above data models, an extended data model of base state with amendments was put forward—model (h). With this model, the data of initial time is looked as base state, and any other time point, except the current time point, are saved as delta files. The delta file are relative to this base state instead of the previous time point. Considering of the frequent search for current time data, another base file (data of current time) is added into the database. The difference between this model and others is that this model uses the start time data as the base file, and the delta file is relative to this base file. As a result, only one delta file needed to add to the database when data has been changed. For the data of start time and the current time, it can be searched directly. In other cases, base file and delta file are needed for searching for historical data. Only two files are used in searching for any history data—one base file and one delta file. This model has the advantage of smaller data size to store and faster speed to search than other models. Obviously the data size will not increase sharply and the searching speed will not decline sharply along with the time. The more time point there, the more obvious the advantage of this model.

Table one list the database size, max file number used in searching and the number of files to be modified on the assumption that the number of time point is 30 and the Base State Distance is 5.

**Table 1 Comparisons of Base State With Amendments Data Models**

Model	Number of files saved	Max File Number Searched	File Number Modified
a	1bs +29df	1bs+1df	30df
b	1bs +29df	1bs+29df	1df
c	6bs +36df	1bs+2df	1 or 3df
d	6bs +36df	1bs+2df	1 or 2df
e	7bs +29df	1bs+2df	1df
f	7bs +35df	1bs+1df	1 or 2df
g	7bs +35df	1bs+1df	1 or 2df
<b>h</b>	<b>2bs +29df</b>	<b>1bs+1df</b>	<b>1df</b>

bs:base file,df:delta file

Concluded from table one: Model (a) and (b) has the advantage of small size to save, but

the searching speed is slow and the workload is heavy;From model(c) to model(g),the searching speed is fast ,but the data size saved in the database is very big.Model (H) have obvious advantage both in the searching speed and data size. The more time point there , the more obvious the advantage of this model. This model is fit for dealing with slight change of census district.when there is huge change,the workload to modify the delta file will be very heavy and the data size will be very big.

## 5.The scheme of population census spatio-temporal database

### 5.1 The changing characteristic of population census data

Population census data includes both population data and the census district data ,so two kinds of data should be taken into consider .It can be found that the changing characteristics of population census data is that the census district data has little change and the population data has lots of change .The Census district of each time of census is bound to be changed ,but the rate of changed district is not high .The type of change include union of two district and split etc,just as the figure 2 has shown:



Figure 2 The Change Type of Census District

The types of change of census district include:a.split :one district is split into many census districts;b. full district union:one census district united with another census district fully ;c.part district union:one census district is split into many parts,each part united with different census district; d. complex change :many kind of change occurred synchronously

### 5.2 The scheme of population census spatio-temporal data base

Taking into account the changing characteristic of population census ,the author consider the best scheme of population census spatio-temporal database is that the population data and the census district data to be saved separately. The population data stored in RDBMS using snapshot data model, the census district data stored using the model(h).Two kinds of data joint by a unique value field.

### 5.3 Experiment

In this experiment,two database were set up to manage population data and census district data separately.Census district is managed by the ESRI spatial data manage database—geodatabase,and population data managed by Microsoft Access database.The census district database is named *spatio.mdb*,and the population database is named *popu.mdb*.The structure of sptio.mdb is(example of county level census district):

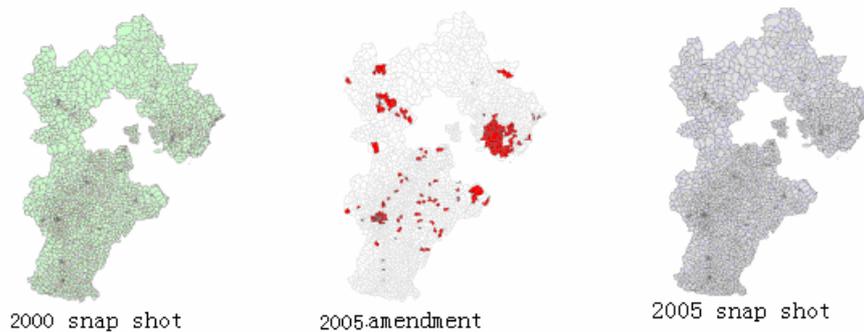
Table 2 Structure of County Census District Table

OBJECT ID	SHAP E	ARE A	PERIMET ER	NAM E	NAME -D	STA R	END	CHANG E
-----------	--------	-------	------------	-------	---------	-------	-----	---------

Table 3 Example of Spatial Data Table

OBJECTID*	Shape*	AREA	PERIMETER	NAME	NAME_D	STAR	END	CHANGE
1	多边形	272815919.785542	96337.889805	九连城镇	张家口市	2005	0	2
2	多边形	102177236.254549	53462.457102	察北虚拟乡	张家口市	2005	0	1
3	多边形	220189018.521255	78341.848270	卧龙镇	承德市	2005	0	1
4	多边形	99008815.012692	56107.483470	沙沟乡	张家口市	2005	0	1
5	多边形	69924452.726546	38072.383002	宇宙营乡	张家口市	2005	0	1
6	多边形	57233194.851196	31255.805890	察北虚拟乡	张家口市	2005	0	1
7	多边形	202814574.786855	68325.803381	杨树岭镇	承德市	2005	0	1
8	多边形	2742198.202063	7091.950570	察北虚拟乡	张家口市	2005	0	1
9	多边形	19596913.627480	19690.641942	察北虚拟乡	张家口市	2005	0	1
10	多边形	139401984.13871	57532.984680	三工地镇	张家口市	2005	0	2
11	多边形	5346175.864367	10815.087910	承德钢铁公司虚拟街道	承德市	2005	0	1
12	多边形	5113283.042240	9973.065454	滦河镇	承德市	2005	0	1
13	多边形	303963843.413387	100642.93120	高家营镇	张家口市	2005	0	1
14	多边形	71134402.937517	58427.923690	东窑子镇	张家口市	2005	0	1
15	多边形	180492249.311722	62608.920486	东望山乡	张家口市	2005	0	1
16	多边形	18030122.671492	21686.308134	五一路街道	张家口市	2005	0	1

Field OBJECTID ,the unique value field ,is used to joint with the population data .Field STAR is used to express the created time of the census district,and the Field END is used to express the changed time of the census district.if the value of the END field is NULL,it express that the district had never been changed form it has been created;The field CHANGED is used to express the change type .



**Figure 3 The Census District of Different Time Based on the Model(h)**

From the fig3, we can find that the data size was reduced sharply by using the model(h). In this example, the data size of 2005 amendment is 495KB, and the data size of 2005 snapshot is 4.59MB. The data size of amendment is just 1/10 to 1/9 of the snapshot's. Meanwhile the searching for the history data is very convenient:

*Select \* from 2000 where change=0*

*Union*

*Select \* from 2005amendment where change<>0*

## 6. Conclusion

Population census spatio-temporal database has the ability to manage population census data of different time efficiently. The data size and the searching speed may be reduced sharply by using the model(h). The more time point there, the more obvious the advantage of this model. It provided an useful example of building a population census spatio-temporal database.

## References

- [1] Liu Renyi, Liu Nan, Su Guozhong. *Analyzing of several typical spatial-temporal data models and expanding of spatial-temporal data models of base state with amendments*[J]. Journal of ZheJiang University (Science Edition) 2000, 27(2).
- [2] Wang Yingjie, Yuan Kansheng, Yu Zhuoyuan. *Visualize of multi-dimension dynamic geo-information*[D]. Beijing: Science Press, 2003.
- [3] Liu Renyi, Liu Nan. *Extension of spatial-temporal data models of base state with amendments and its implementation in land registration*[J]. Acta Geodaetica et cartographica sinica, 2001, 30(2): 168-172.
- [4] Zhang Zuxun, Huang Mingzhi. *Study of data structure of temporal GIS*[J]. Bulletin

- of surveying and mapping,1996,(1):19-22.
- [5] Cao Zhiyue,Liu Yue.*An object-oriented spatio-temporal data model* [J],Acta Geodaetica et cartographica sinica,2000,31(1):88-92.
- [6] Zhang Baogang,Zhu Chongguang,Wang Runsheng.*Improved methods of base state with amendments on spatio-temporal data.*[J]Acta Geodaetica et cartographica sinica,2005,34(3):252-256.
- [7] N.Pelekis,B.Theodoulidis,I.Kopanakis,Y.Thodoridis.Literature Review of *Spatio-Temporal Database Models* (EB/OL) .
- [8] M.Yuan,*Wildfire Conceptual Modeling for Building GIS Space-Time Models*,Proceeding of the GIS/LIS'94,pp.860-869,1994.
- [9] A.Renolen, *History Graphs: Conceptual Modeling of Spatio-Temporal Data*[J],In GIS Frontiers in Business and Science,Vol.2, International Cartographic Association,Brno,Czech Republic,1996.
- [10] N.Tryfona.*Modeling phenomena in Spatiotemporal Database :Desiderata and Solution*,Department of Computer Science,Aalborg University,Denmark,1997.
- [11] C.Claramunt,C.Parent,S.Spaccapietra,M.Theriault."Database Modeling for Environmental and Land Use Changes"[J],Geographical Information and Planning,Chapter 20,Springer-Verlag,Zurich,pp.23-42,1995.
- [12] M.-A.Aufauer-Portier ,C.Bonhomme. *A High Level Visual Language for Spatial Data Management*,Proceedings of Visual'99,3<sup>rd</sup> International Conference on Visual Information System,Amsterdam,1999.
- [13] C.Bonfatti.*A Visual Language for Querying Spatio-Temporal Database*. Proceeding of the 7<sup>th</sup> ACM Symposium on Advances in Geographic Information Systems,Kansas City,Mo,1999.
- [14] A.Raza ,W.Kainz.*Cell Tuple Based Spatio-Temporal Data Model:An Object Oriented Approach*.ACM-GIS'99:Proceedings of the 7<sup>th</sup> International Symposium on Advances in Geographic Information Systems,Kansas City,Mo,1999.
- [15] M.Erwig,R.H.Guting,M.Schneider,M.Vazirgiannis.Spatio-TemporalData Types:*An Approach to Modeling and Querying Moving Object in Database* [J] , Geoinformatica , 3(3): 264-291,1999
- [16] R.H.Guting,M.H.Bohlen,M.Erwig,C.S.Jensen,N.A.Lorentzos,M.Schneider,and M.Vazirgiannis. *A foundation for Representing and Querying Moving Objects*[J].ACM Transactions on Database Systems,25(1):1-42,2000.
- [17] L.forlizzi, R.H.Guting,E.Nardelli,M.Scheider.*A data Model and Data Structures for Moving Objects Databases*.Proceedings ACM SIGMOD Intl.Conf.on Management of Data,pages 319-330,2000.
- [18] Zheng Kougen etc.*A unified spatio-temporal data model based on state and change*[J].Journal of software,12(9),1360-1365.

- [19] J.Chen,J.Jiang.Event-Based Spatio-temporal Database Dasign,IAPRS,Vol 32,Part4  
“GIS-Between Visions and Applications”,Stuttgart,1998.