

# 基于 Shapefile 的基态修正模型设计与实现

胡 辉<sup>1</sup>, 王昌佐<sup>2</sup>, 梁万杰<sup>1,3</sup>, 江 东<sup>1\*</sup>, 鞠 斌<sup>1</sup>

(1. 中国科学院地理科学与资源研究所, 北京 100101; 2. 环境保护部卫星环境应用中心, 北京 100094;  
3. 南京农业大学资源与环境科学学院, 南京 210095)

**摘要:** 本文通过对 Shapefile 文件进行扩展, 使用改进的基态修正模型, 构建了一种 ArcGIS 的时空数据模型, 支持对时空变化过程的存储、分析和模拟, 同时支持传统的 GIS 空间分析功能。本研究还设计和实现了一种基于多时态矢量数据集的时空数据变化过程发现和时空数据转换算法, 此算法可以自动提取多个时态之间的变化过程, 将多时态数据转换成基于 Shapefile 的时空数据模型存储。鉴于该数据模型, 设计和实现了时空数据的快照恢复、变化查询、历史演化分析和数据更新方法, 验证了模型的可应用性。

**关键词:** 时空数据模型; 时态 GIS; 时空变化; 基态修正模型

**DOI:** 10.3724/SP.J.1047.2012.00313

## 1 引言

空间对象有 3 个基本的属性: 空间属性, 专题属性和时间属性。传统的 GIS 在处理空间对象时, 主要突出空间属性和专题属性, 对时间属性则予以固定<sup>[1]</sup>。这种取舍造成传统的 GIS 只能处理世界的当前状态, 或者某一历史状态的快照, 而无法记录数据随时间的动态变化, 无法进行时序分析和处理时态现象<sup>[2]</sup>。地理对象的变化与时间紧密相关, 为了跟踪和记录变化, 需要在 GIS 中引入时间信息, 这就是时态 GIS(TGIS)所要解决的主要问题。作为 TGIS 的核心, 时空数据模型的研究成为 TGIS 实现的前提条件。

在过去的 20 多年里, 研究者们提出了大量的时空数据模型, 如时空立方体模型、基态修正模型、时空复合模型、快照模型<sup>[3]</sup>, 面向对象时空数据模型<sup>[4]</sup>, 将时间语义理解为事件序列的 ESTDM 模型以及改进模型<sup>[2,5]</sup>, 以过程为核心的时空数据模型<sup>[6]</sup>等。这些模型多以概念模型为主, 有的还没有实用的系统平台, 或仅停留在原型验证阶段。ArcGIS 平台作为当今主流的 GIS 软件平台之一, 在地学处理和空间分析等方面功能强大, 积淀丰富, 但

在时态 GIS 领域, 仍少有涉及。

因此, 本研究以支持时空变化表达为主要目的, 设计了支持时空变化过程的时空数据模型, 并通过多时态矢量空间数据转换算法自动发现和提取时空变化结构信息, 将多时态的矢量空间数据转换成系统支持的时空数据结构存储。

## 2 基于 Shapefile 的基态修正模型设计

### 2.1 基态修正模型

对地理空间实体变化的认知是 TGIS 的基础<sup>[7]</sup>。Yuan 将地理对象的变化划分为 6 种类型<sup>[8]</sup>, Tamas Abraham 在 1999 年使用 W 三角形也总结了 8 种空间变化类型<sup>[9]</sup>, 这些变化描述能够对特定的应用领域和应用需求进行较好的模拟, 但通用性比较差<sup>[10]</sup>。在前人研究的基础上, 本研究将地理对象的变化抽象为以下 8 种类型: ①产生; ②消失; ③属性变化; ④空间变化; ⑤属性和空间变化; ⑥1 对多变更; ⑦多对 1 变更; ⑧不变。

基态修正模型最早由 Gail Langran 提出, 其主要思想是只保存某个时间的数据状态(称为基态), 和变化量<sup>[1, 3, 11]</sup>。基态修正模型的缺点是: 对较远

收稿日期: 2011-02-16; 修回日期: 2012-02-10.

基金项目: 地震行业科研专项经费项目(201208018); 国家科技支撑课题“中国重大自然灾害孕险环境分析技术”(2008BAK50B01); 南水北调水资源综合配置技术研究(2006BAB04A16)。

作者简介: 胡 辉(1980-), 男, 湖南株洲人, 硕士, 研究方向为 GIS 应用研究。E-mail: huhui1007@126.com

\* 通讯作者: 江 东(1972-), 男, 博士, 副研究员, 研究方向为资源环境遥感应用。E-mail: jiangd@igsnrr.ac.cn

的历史数据进行查询时,效率较低。为提高查询效率,刘睿等提出了基于动态基态的时空数据基态修正方法,引入了动态基态数据与动态数据存储区域的概念<sup>[12]</sup>。本研究综合前人的研究结果,采用多基态和多级差文件<sup>[13-14]</sup>对模型进行改进。主要的改进包括:(1)以研究对象的最新状态作为当前的工作状态;(2)以整个数据系统较早的状态作为最原始的基态,最早的变化基于此基态,当差文件累积到一定程度时,重新建立新的基态,之前的基态和差文件不再变更,新的变化发生时,在新基态的基础上更新差文件和当前状态;(3)对于基态距(相邻两个基态之间的时间距离)较大的区域,对其间的差文件进行合并,建立多级差文件,当查询的历史时刻距离最近的基态较远时,优先使用时间跨度较大的二级差文件进行叠加操作。改进后的基态修正模型数据组织结构如图 1 所示。

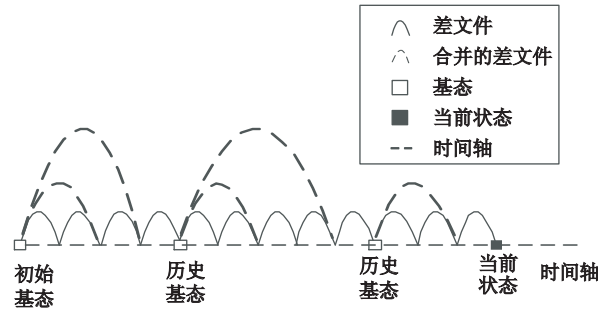


图 1 本文的基态修正模型组织策略

Fig. 1 The organization strategy of the base state with amendments model in this study

2.2 基于 Shapefile 的基态修正模型的设计

2.2.1 时空数据结构设计

数据模型要在实际中应用,需要合适的数据结构实现。本研究以矢量的时空模型为例,采用 Shapefile 文件保存数据,并对 Shapefile 文件做如下扩展:首先,增加 GlobalID 字段,作为地理对象的唯一标识,为了维护数据模型的完整性,对 GlobalID 进行如下约定:①唯一性;②不可重用性;③永久性,即一个对象对应的 GlobalID 一经产生,便不能再更改。然后,增加地理对象的开始时间(StartTime),结束时间(EndTime),入库时间(DBTime)字段,其中,入库时间应不早于开始时间,且时空变化过程的模拟时间粒度要达到秒级的精度<sup>[15-16]</sup>。最后,为了完整的存储和重现时间拓扑信息,增加变化类型(ChangeType),前一状态时空对象 ID

(PriorIDs)和后一状态时空对象 ID(NextIDs)字段。时刻  $t$  新产生的时空对象,前一状态的时空对象应为空;时刻  $t$  消失的时空对象,其后一状态的时空对象也为空;一对多变化,该时空对象的 NextIDs 应存储多个后续对象;多对一变化,该时空对象的 PriorIDs 完整存储所有前驱对象信息。基态文件和差文件的存储数据结构相同。增加必要的字段后,整个 Shapefile 数据的数据结构如图 2 所示。

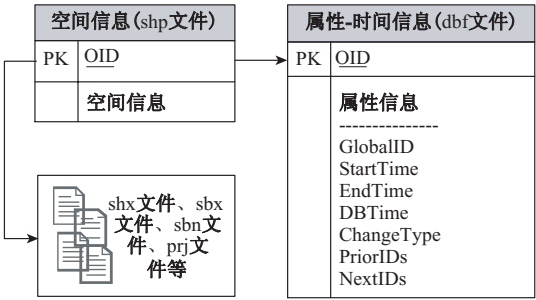


图 2 时空数据存储结构

Fig. 2 The storage structure of spatio-temporal data

2.2.2 时空数据索引

随着时空数据的时间跨度的增加,对应的基态文件和差文件会越来越多。为快速地找到指定时间点对应的基态文件和差文件,周辉等 2010 年提出了基于时间的面向对象的差文件索引设计方法<sup>[17]</sup>,刘仁义等提出了区段快速索引和变粒度存储因子方法<sup>[18]</sup>。本研究在前人研究的基础上,分别建立基于时间的基态文件索引和差文件索引。以基态文件对应的时间主键,按时间从早至晚的顺序排列,建立基于时间的基态文件索引。同时以差文件所记录的变化事件中最早的发生时间为主键,按照从早至晚的顺序,建立基于时间的差文件索引。基于时间的基态文件索引和差文件索引的结构及其表结构如图 3 所示。

3 模型的实现与应用

3.1 多时态矢量空间数据的转换

虽然时态 GIS 的概念已经产生 20 多年,但至今仍没有一个可以广泛应用的时态数据模型,因此,当前的多时态数据大部分仍然是以多快照模式存储的,即对应不同的时刻,存储一个对应的快照状态,不同的时刻对应着多个快照状态。而时空数据模型发展中的一个重要问题就是应该顾及到新的时空数据模型与历史积累的大量实验数据和生



图 3 时空数据索引

Fig. 3 Spatio-temporal data index

产数据之间的移植性,至少应提供一种转换方法和转换接口<sup>[19]</sup>。为了能够应用本文研究中设计的数据模型,设计了将多时相快照数据自动转换到基于基态修正的差文件存储的算法。考虑到矢量数据在 ArcGIS 中主要以点、线、面的形式存储,多时态矢量空间数据自动转换算法也相应的分别针对点、线和面的转换算法。针对 3 种不同类型的转换算法的输入是不同时刻的两个矢量数据集,设为数据集  $DB_0$  和数据集  $DB_1$ ,其中,数据集  $DB_0$  为旧时态数据集, $DB_1$  为新时态数据集;同时设  $obj_0$  和  $obj_1$  分别为数据集  $DB_0$  和  $DB_1$  上的对象;首先设置若干个属性字段作为属性查找比较的依据字段;其次设置一个容差范围  $\sigma$ ,作为空间比较的容差距离,在算法技术流程方面,参考汪汇兵的研究工作后设计的算

法,主要思路如下<sup>[10]</sup>:

(1) 点状数据转换算法

首先,遍历输入数据集  $DB_0$  和  $DB_1$ ,通过比较点之间的距离,建立距离  $\Delta d < \sigma$  的点之间的关联关系(若有多个点对象符合要求,则取距离最近的一个作为关联对象);然后,比较有关联关系的点的空间信息和属性信息,提取出变化信息(属性变化、空间变化、空间—属性均变化等);再是,对所有未找到关联对象的点  $obj_0$ ,通过设定的属性字段查找关联对象,若找到关联对象,则比较其信息,提取变化信息,否则,将未找到关联对象的点  $obj_0$  设置为消失;最后,将所有未找到关联对象的  $obj_1$  设置为新产生的对象。点状数据转换算法的流程如图 4 所示。

(2) 线状数据转换算法

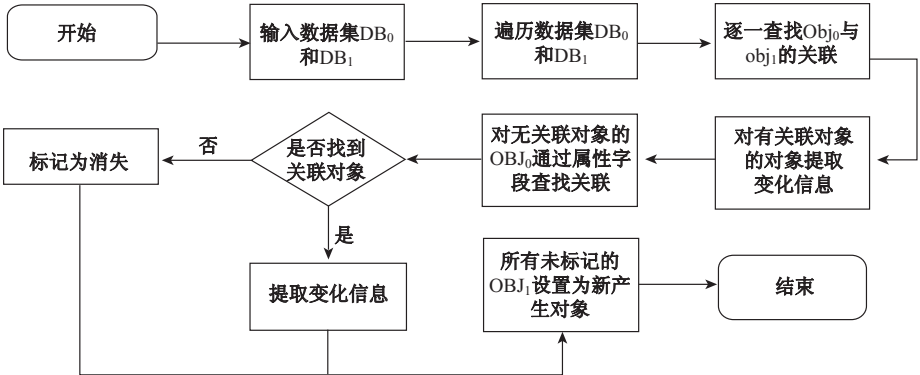


图 4 点数据转换算法流程

Fig. 4 Flowchart of point data conversion algorithm

对于线状数据,上文中论述的 8 种变化类型都可能涉及。为了快速查找到在空间上可能相关的对象,先对两个输入数据集  $DB_0$  和  $DB_1$  中的每一个线对象,建立其最小外接矩形(MBR)索引。线与线之间的关联性是根据两者节点匹配率来计算的;两个节点之间的距离小于指定的容差  $\sigma$ ,则认为两个节点可以匹配;设定一个阈值  $TL$ ,若两个线对象的节点的匹配数量超过占总节点数的百分比大于设定的阈值  $TL$ ,则认为  $obj_0$  与  $obj_1$  能够关联,其中,阈值  $T$  可以根据实际情况进行设定。为了提高计算效率,对于任意  $obj_0$  和  $obj_1$ ,首先比较其 MBR,只

有两者 MBR 相交时其才可能具有关联性,若两者 MBR 不相交,可直接认定两者空间上不关联。

首先,遍历输入数据集  $DB_0$  和  $DB_1$ ,对每一个  $obj_0$  和  $obj_1$  都尝试建立对象关联关系,此时可能出现一对多或者多对一的关联关系,根据关联关系,提取变化信息(空间变化、属性变化、空间属性变化、多对一、一对多等);其次,对未找到关联的  $obj_0$ ,通过属性查找关联对象,若找到关联对象,则比较其属性和空间信息,提取变化信息,否则标记为消失;最后,对所有未标记的  $obj_1$ ,设置为新产生对象,线状数据转换的流程如图 5 所示。

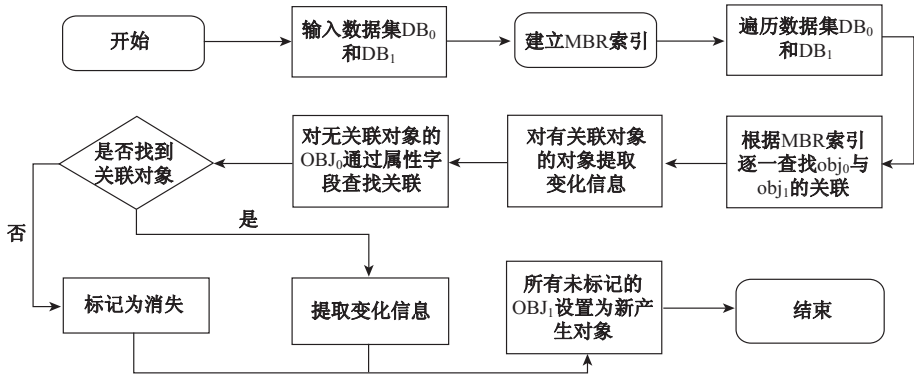


图 5 线、面状数据转换算法流程

Fig. 5 Flowchart of polyline and polygon data conversion algorithm

(3) 面状数据转换算法

为快速地找到可能与一个旧时态面对象相交的新时态面对象,此处仍然使用 MBR 作为快速判断的依据。面状数据的转换算法流程与线状数据转换一样,不同之处在于面状数据的关联是通过其相交部分的面积占其总面积的比例来确定的。同样设定阈值  $TA$ ,对任意两个  $obj_0$  和  $obj_1$ ,若两者重叠的面积与任意一个对象的面积的比例达到指定的阈值  $TA$ ,则认为两者可以关联,其中,阈值  $TA$  可以根据实际情况进行设定。在计算重叠面积时,一般而言使用栅格数据比较合适,矢量数据叠置比较费时。但是矢量-栅格数据转换存在着无法确定合适分辨率的问题,若分辨率过高,转换效率太低;若分辨率过低,转换后一些小多边形容易遗漏,转换误差也较大,因此,本文中仍然采用矢量数据直接叠置的方法<sup>[15-16]</sup>。面状数据转换算法流程与线状数据转换流程一致,如图 5 所示。

3.2 模型的应用

前文中从实用性和应用性的角度对基态修正

模型进行改进,并设计了其基态文件和差文件的存储结构,考虑了其数据组织和索引策略,但其具体可应用性究竟如何,还需要在实际中来验证。因此,本文设计和实现了一个原型系统对本文时空数据模型的实用性进行了验证,系统的整体架构如图 6 所示。

(1) 时空快照的恢复

时空快照的恢复是指建立用户指定任意时刻的时空快照状态,这是时态 GIS 的一个核心业务,也是时态 GIS 区别于普通 GIS 的关键所在。时空快照的恢复包括 3 个步骤。首先,根据用户指定的时刻,遍历基态文件索引,从索引中找到时间版本与用户指定时刻最临近的基态文件。第二步,从差文件索引中查找到所有涉及到的差文件。最后,将基态文件和差文件按时间顺序逐一合并,生成时空快照。根据变化的时空数据进行时空快照恢复结果如图 7 所示,其中绿色框标和蓝色框标记部分是历史状态相比于当前状态的变化部分。

(2) 时空数据的分析

时空数据的分析主要包括对变化的分析、对象



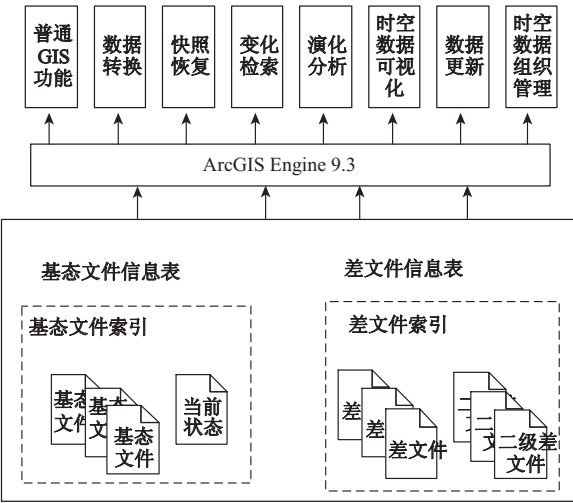


图6 系统架构

Fig. 6 Architecture of the system

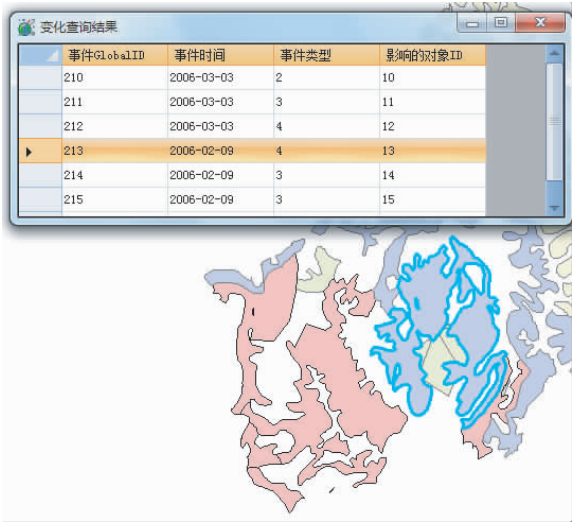


图8 变化的查询结果

Fig. 8 Result of the query

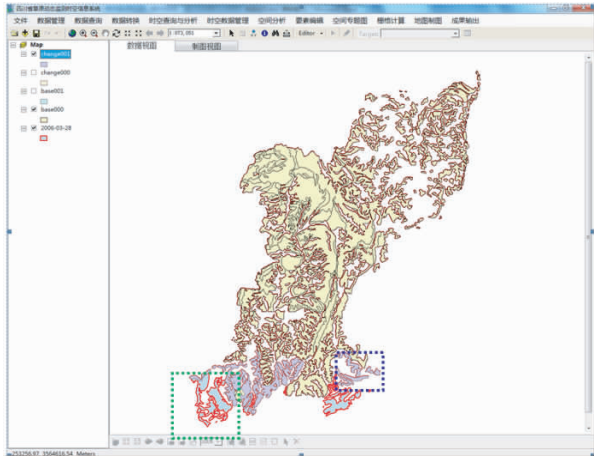


图7 时空快照恢复结果

Fig. 7 The result of spatio-temporal snap recover

历史状态的分析等。

对变化的查询主要查询符合条件的变化事件的信息。要对整个时空数据查询变化信息,首先,根据查询的时间范围在文件索引中找到符合该时间范围的基态文件和差文件,然后,在符合条件的差文件中根据查询条件逐一查询并记录查询结果。由于时间域信息和属性域信息都是使用 dbase 表进行存储的,因此,可以使用标准的 SQL 语句进行查询。如图 8 所示是查询到的 2004-03-28 至 2006-03-28 的事件信息。

在对象历史状态进行查询时,用户主要需要获取感兴趣对象在整个时间过程中的历史演化。在设计时空数据模型时,本文已经考虑到这种需求,

将每个时空对象的前一状态时空对象信息和后一状态时空对象信息都进行了存储,因此,只需要找到感兴趣的对象的最新状态,然后根据其 PriorIDs 和 NextIDs 进行回溯即可给出该时空对象的完整演化过程。图 9 是对 GlobalID 为 209 的对象进行演化分析,查询其历史变化过程的结果。



图9 对象演化分析结果

Fig. 9 Result of evolution analysis on the objects

(3) 时空数据的更新

时空数据的更新主要涉及到对时空数据的编辑和更新,与传统的 GIS 中数据编辑和更新只需要保存一个最新状态不同的是,在时态 GIS 的数据编辑过程中,不仅要更新数据的最新状态,其历史状态也必须被妥善的保存下来以供未来的时空查询,而不能被简单的替代或者删除。同时,数据的编辑过程也不再是简单的更新属性和空间信息,还需要对数据的时间拓扑关系进行维护,以保证未来时空分析的正确性。时空数据更新操作的对象可能是当前状态,也可能是生成的历史状态快照。在本文讨论的基态修正时空数据模型中,对一条记录的编辑主要涉及到当前状态的更新、差文件的更新、基态文件索引的更新、差文件索引的更新、基态文件

信息表的更新和差文件信息表的更新。图 10 所示为删除一个对象的过程,可以看出,在编辑模式下

删除该对象后,在差文件中记录了一个删除操作(选中的记录),同时,该对象被标记为已删除。

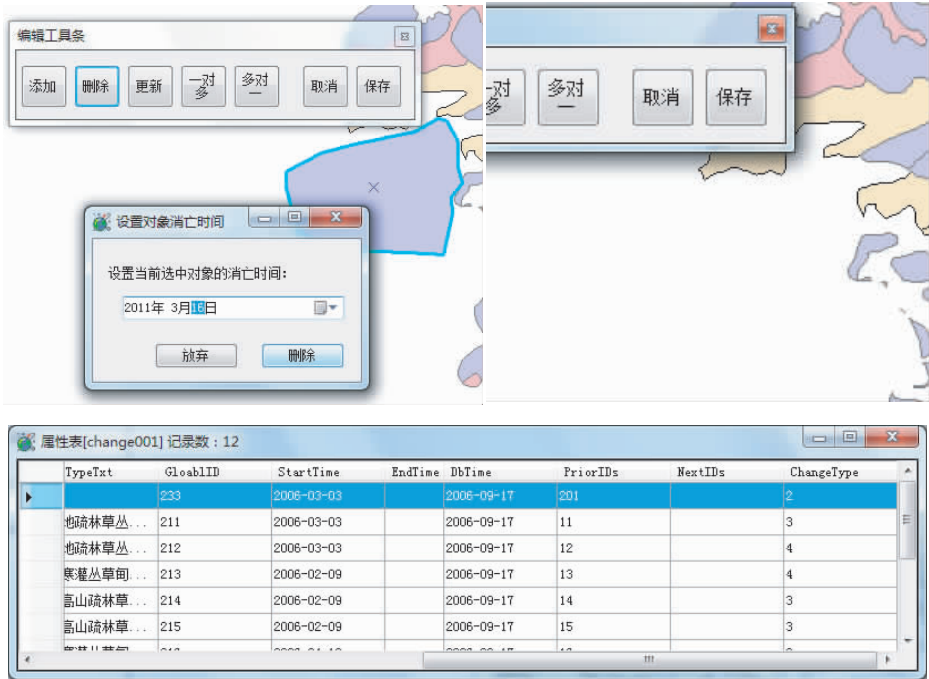


图 10 删除一个对象的过程  
Fig. 10 Deleting one spatio-temporal object

4 结 语

本文设计和实现了一种基于 Shapefile 文件格式的时空数据模型,使用改进的基态修正模型进行数据存储和组织,节省了数据的存储空间,同时也完整地保存了数据的历史变化信息,突破了传统 GIS 只能存储和表达单一时刻地理对象信息的限制,并支持对历史变化信息的查询和时序分析,有利于分析和发现地理对象随时间的变化规律和变化趋势等。鉴此,对时空数据模型的设计开发出一套原型系统。这说明该模型能够完整地存储和表达地理对象随着时间的变化,能够支持基本的时空快照恢复、基于时间的分析功能、数据的更新管理,具有较好的可行性和实用性。

下一步对模型的改进和完善包括:(1)进一步研究时空分析的需求,完善时间信息的组织,提高存储和分析效率;(2)完善时空数据管理和分析系统,提供一个通用的框架,将时空数据的编辑、管理、分析、查询等操作进行统一的管理,提供模型的可扩展性和应用性;(3)研究时空数据动态可视化的方法。

参考文献:

[1] Langran G. A framework for temporal geographic information[J]. Cartographica, 1988, 25(3): 1-14.  
[2] 牛方曲,朱德海,程昌秀. 改进基于事件的时空数据模型[J]. 地球信息科学, 2006, 8(3): 104-108.  
[3] Langran G. Temporal GIS design tradeoffs[J]. Temporal GIS Design Tradeoffs, 1989, 2(2): 16-25.  
[4] Worboys M F. A unified model for spatial and temporal information[J]. The Computer Journal, 1994, 37(1): 26-34.  
[5] Peuquet D J, Duan N. An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data[J]. International Journal of Geographical Information Science, 1995, 9(1): 7-24.  
[6] 薛存金,周成虎,苏奋振,等. 面向过程的时空数据模型研究[J]. 测绘学报, 2010, 39(1): 95-101.  
[7] 周晓光,陈军. 基于变化映射的时空数据动态操作[J]. 遥感学报, 2009, 13(4): 653-658.  
[8] Yuan M. Temporal GIS and spatio-temporal modeling [C]. Santa Fe, New Mexico, USA: 1996.  
[9] Abraham T, Roddick J F. Survey of spatio-temporal databases[J]. GeoInformatica, 1999, 3(1): 61-99.  
[10] 汪汇兵,唐新明,史绍雨,等. 一种基于时态树结构的

- 时空数据的动态可视化方法[J]. 测绘科学, 2009, 34(1): 110 - 112.
- [11] 陆琪,周晓光,周辉. 基于基态修正模型的时空数据动态操作[J]. 计算机应用研究, 2010, 11(27): 4155 - 4158.
- [12] 刘睿,周晓光,李晓蕾. 基于动态基态方法的基态修正时空数据模型[J]. 测绘科学, 2009, 34(1): 130 - 224.
- [13] 张保钢,朱重光,王润生. 改进的时空数据基态修正方法[J]. 测绘学报, 2005(3): 252 - 256.
- [14] 张祖勋,黄明智. 时态 GIS 数据结构的研讨[J]. 测绘通报, 1996(1): 19 - 22.
- [15] 陈珺. 对时空数据管理中时间精度问题的讨论[J]. 现代测绘, 2009, 32(3): 12 - 14.
- [16] 魏海平. 时空 GIS 建模研究与实践[D]. 解放军信息工程大学, 2007.
- [17] 周辉,周晓光,何凭宗,等. 基态修正模型的时空数据组织和快照查询方法研究[J]. 地理信息世界, 2010, 2: 49 - 53.
- [18] 刘仁义,刘南. 动态土地信息系统时空过程及时空数据存储[J]. 中国图象图形学报, 2002(4): 388 - 393.
- [19] 姜晓轶,周云轩. 从空间到时间——时空数据模型研究[J]. 吉林大学学报·地球科学版, 2006, 36(3): 480 - 485.

## Design and Implementation of a Spatio-temporal Data Model on Account of Shapefile and Base State with Amendments Model

HU Hui<sup>1</sup>, WANG Changzuo<sup>2</sup>, LIANG Wanjie<sup>1,3</sup>, JIANG Dong<sup>1</sup> and JU Bin<sup>1</sup>

(1. *Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China;*

2. *Satellite Environment Center, Ministry of Environmental Protection, Beijing 100049, China;*

3. *College of Resources and Environmental Sciences, Nanjing Agricultural University, Nanjing 210095, China*)

**Abstract:** The research of spatio-temporal data model is the core of temporal GIS. Though the concept of temporal GIS has been proposed for years, there isn't any practical system till now. In order to simulate and save the temporal and spatial variation as well as apply the traditional GIS functions, in this paper we chose to extend the existing ArcGIS vector data model to support temporal and spatial analysis functions. Firstly, we analysed the type of temporal and spatial variation and divided the common temporal and spatial variation into 8 different kinds. Then based on the analysis of variation types, we constructed a kind of spatio-temporal data model that support to save, analyse and simulate spatio-temporal changes in view of the vector data model of ArcGIS by using the base state with amendments model. Because that most multi-period spatial data are saved into multi-snapshots format, in other words, they are saved separately, it's unrealistic to convert them into base state with amendments format manually. So in this paper, we also design and implement an algorithm which can extract and save temporal and spatial changes into base state with amendments format automatically from multi-period data. In order to verify the feasibility and efficiency of the data model, we designed and implemented a demo system to support the query, analysis, management and dynamic visualization of spatio-temporal data. The verify result showed that this data model could be used to support spatio-temporal change recording, analysis and management.

**Key words:** spatio-temporal data model; temporal GIS; spatio-temporal change; base state with amendments model