

国家尺度社会经济数据格网化原理和方法

胡云锋¹, 王倩倩^{1,2}, 刘越¹, 李军³, 任旺兵³

(1. 中国科学院地理科学与资源研究所, 北京 100101; 2. 辽宁工程技术大学 测绘与地理科学学院, 阜新 123000;
3. 国家发展和改革委员会宏观经济研究院, 北京 100038)

摘要: 空间信息获取和处理技术, 可将传统关系型社会经济数据转变为空间化的矢量或者格网数据。首先, 分析了传统关系型社会经济数据库的不足, 并对社会经济数据格网化理论研究及格网数据库建设的历史做了回顾; 在明确社会经济数据格网化基本概念的基础上, 提出了社会经济数据格网化的3个基本要求, 即时间可比、空间一致和逻辑自治; 同时提出了一个包括24个关键指标的国家尺度社会经济格网数据库的指标体系, 认为社会经济格网数据库生产过程的主要步骤为逻辑检查、空间匹配、代码匹配、空间离散和检查校验; 研究对国家尺度社会经济指标的空间离散过程和离散模型、不同层级社会经济数据的整合和离散策略进行了重点分析。研究最后就社会经济数据格网化过程中存在的主要问题进行了总结。

关键词: 国家尺度; 关系型数据; 空间离散; 格网数据

DOI: 10.3724/SP.J.1047.2011.00573

1 引言

全国总人口、国内生产总值(GDP; Gross Domestic Product)等社会经济数据一般是以基层行政区为基本单元, 通过普查、抽样统计等方式, 经过逐级汇总后最终形成的二维表格; 这些数据通常使用关系型数据库(RDBMS; Relationship Data Base Management System)以实现存储、检索、更新、数据分析和信息重现等功能。传统关系型社会经济数据具有强烈的法律属性, 并以严谨的统计学理论和方法作为支撑, 具有权威、系统、规范等特点。

上述传统的关系型、表格化的社会经济数据在实际应用中也存在许多不足。具体有: (1) 时间分辨率低、更新周期长。国家级普查通常要间隔5年或10年以上, 抽样调查一般以一年期为间隔; (2) 空间分辨率低, 调查单元不稳定、不规则。公开发布的国家级普查数据多以县(县级市、旗、市辖区)为统计单元, 而抽样调查数据通常多以省区、地级市为统计单元; (3) 直观性差, 无法以可视化形式展现社会经济要素的空间分布形态及其发展规律和发展态势; (4) 不支持空间运算和分析。不利于挖

掘社会经济运行中的深层次规律和问题, 也不利于结合区域资源环境禀赋对社会经济运行开展模拟和预测研究。

随着空间信息技术的快速发展, 特别是在地理信息系统技术(GIS; Geographic Information System)和遥感技术(RS; Remote Sensing)的支持下, 研究人员逐渐提出了社会经济数据空间化和格网化技术。在现代空间信息技术, 尤其是海量和快速更新的对地观测信息和愈来愈强大的地理信息空间分析和空间表达功能支持下, 通过社会经济数据的空间化和格网化技术, 可以有效克服传统社会经济数据的不足。

社会经济数据空间化和格网化最早可上溯至20世纪20年代经济地理学家应用的计量地理方法, 但就国家和全球大尺度区域社会经济数据格网化理论、方法论研究以及数据库建设实践而言, 这一过程则是从20世纪90年代在GIS和RS得到较为广泛和深入应用后才真正开始的。1993年, Goodchild提出以GIS为支撑, 采用面域插值方法分配社会经济数据^[1]; 陈述彭于2000年前后即呼吁建立融合气象、遥感、人口和经济、自然资源和环

收稿日期: 2011-08-09; **修回日期:** 2011-10-01.

基金项目: 科技部科技支撑项目(2008BAH31B04); 科技部“973”计划项目(2010CB950904); 国家自然科学基金项目(40971223); 中国科学院知识创新方向性项目(KZCX2-EW-306)。

作者简介: 胡云锋(1974-), 男, 博士, 副研究员, 主要从事区域资源环境监测和可持续发展评价研究。

E-mail: huyf@lreis.ac.cn

境等要素的新一代网格数据库^[2]。在数据库建设方面,目前已建成一批全球尺度和国家尺度的社会经济格网数据库,如:世界栅格人口数据库(Gridded Population of the World: GPW)^[3-4]、LandScan 人口分布数据库(LandScan Population Distribution Database: LandScan)^[5-6]、世界地理栅格化经济数据集(Geographically based ECONomic data: G-ECON)^[7-8]、中国 1km 栅格人口和经济数据集^[9-11]等。进入 21 世纪,随着经济地理量化研究的深入发展,尤其是新时期区域规划及其动态监测评估需求,国内外对于社会经济数据格网化理论、方法论研究以及空间数据库建设需求更加强烈^[12-15]。

回顾社会经济数据格网化研究和数据库建设历史,可以发现:对人口数据空间化研究较早,人口要素格网化方法最为成熟,全球和国家尺度的人口格网化数据库数量也较多。从方法上看,人口数据空间化经历了一个从最早将人口普查数据与行政区划或者普查小区数据简单作空间连接,到目前利用土地覆被和土地利用数据、交通数据、夜景灯光数据等资源环境数据,根据人口空间分布的自然规律逐格点分配的过程。在经济要素格网化研究方面,由于经济活动自身的极端复杂性,研究人员在理解经济活动形成机制、关联地理和资源背景以及经济要素表达形式等方面存在较大分歧。因此,目前有关经济要素格网化的理论、方法论研究以及数据库成熟度尚不能令人满意。

2 网络化概念、要求和指标

2.1 社会经济数据格网化概念和基本要求

社会经济数据格网化是指:在地理信息系统和遥感等现代空间信息技术支持下,将传统的、以基层行政区为统计单元的社会经济数据按照一定规则分配到地理格网上。社会经济数据格网化是自然科学研究和社会科学研究交叉、融合的关键领域之一。格网化的社会经济数据不仅能够更加直观、更加接近真实地反映现实,同时也为自然、经济和社会数据的融合提供了统一的空间基准,为分析、模拟和预测各类社会经济要素的发展和演化提供了基本的技术支撑^[2, 13, 16]。

社会经济数据格网化应满足三个基本要求:时间可比、空间一致以及逻辑自洽。

所谓时间可比,首先是指需要将不同时期的社会经济数据“归一化”到某一个基准点,最典型的案例就是将经济要素中的各种现价(当年价)转换为不变价以消除通货膨胀的影响。其次是指用于辅助生成社会经济格网数据的相关自然资源和地理环境背景数据所表征的时间节点应当和社会经济数据所表征的时间节点相同或相近,典型案例就是将 2000 年的土地利用/土地覆被数据、基础地理数据等用于配置同一年的人口数据。

所谓空间一致,是指综合考虑社会经济要素特点、研究区域位置和面积等特点,将各类空间数据处理成坐标系相同、投影一致、空间起算位置相同、空间格网大小相同的空间数据集。对于中国国家尺度的社会经济数据空间离散,可以采用中国国家 2000 坐标系统或西安 1980 坐标系统,Albers 等积圆锥投影(双标准纬线分别为 25°和 47°,中央经线为 105°),空间格网一般采用 1km 的 ArcInfo GRID 栅格。此外,还要将不同来源的空间数据、属性数据处理成具有相同空间覆盖区域和相同处理单元。

所谓逻辑自洽,是指社会经济空间格网数据在任意层级和任意位置上的数量关系应当符合社会经济内在的分布特点和运行规律。最典型的例子有:在任意格点上应保证总人口等于城镇人口与农村人口之和,城镇总人口等于城镇男性人口与城镇女性人口之和,GDP 等于一产增加值、二产增加值、三产增加值之和,二产增加值等于工业产值与建筑业产值之和;上一级格点(或行政区)上的总人口数值应当与下一级同样位置的全部格点上的总人口数值相等。

2.2 国家社会和经济格网数据库指标体系

在国家统计部门开展的年度抽样调查或普查过程中,工作人员通常要调查十数个、甚至上百个的问题,最终形成数千、甚至数万个具体指标项。例如:2010 年全国第六次人口普查中,普查表短表共有 18 个项目,普查长表共有 45 个项目,经汇集整理后大致可形成 10 大类、5000 余个具体指标项;2008 年全国第二次经济普查项目包括单位基本属性、从业人员、财务状况、生产经营情况、生产能力、能源消耗、科技活动情况等 7 大类,经汇集整理后将形成 40 余类、3 万余个具体指标项。

为构建国家尺度社会经济格网数据库,必须从上述纷繁复杂的数据项中进行指标遴选。具体要

求是需考虑指标的系统性、简洁性、通用性和动态性,同时注意指标数据的可获得性与衍生性,并体现科学发展观的要求和现代政府职能转变要求^[17],最后形成一个系统、简练、可扩展的空间数据库。根据上述要求,我们建立了包括以下 3 大类、24 个指标项在内的中国国家尺度社会经济格网数据库指标体系(表 1)。

表 1 国家尺度社会经济格网数据库指标体系

Tab. 1 Index system of national society and
economy grid database

大类	中类	指标和代码
人口格网	人口总量	总人口(TPOP)
	城乡构成	城镇人口(CTPOP)
		农村人口(COPOP)
	性别构成	男性人口(MAL_POP)
		女性人口(FEM_POP)
	年龄构成	老年人口(OLD_POP)
儿童人口(CHD_POP)		
综合经济	GDP 总量	国内生产总值(GDP)
	产业构成	第一产业增加值(GDP1)
	GDP 密度	第二产业增加值(GDP2)
		第三产业增加值(GDP3)
		人均 GDP(GDP_POP)
行业经济	工业和 建筑业	工业产值(IND)
		建筑业产值(CONST)
		制造业产值(PROUD)
		石化产业产值(PET_CHEM)
		冶金产业产值(METAL)
		其他工业行业产值(OTH2)
	服务业	批发和零售业产值(SALE)
		交通运输、仓储和邮政业产值(TRAF)
		住宿和餐饮业产值(ACOM)
		金融业产值(FINAN)
		房地产业产值(ESTAT)
		其他服务行业产值(OTH3)

3 格网化流程和方法

3.1 社会经济数据格网化技术流程

根据社会经济数据格网化的概念和基本要求,社会经济数据格网化过程可以分为 5 个步骤:逻辑检查、空间匹配、代码匹配、空间离散和检查校验。

(1)逻辑检查:首先是要检查同一层级、不同要素之间的平衡关系,如在任意行政单元内,GDP 应等于一产增加值、二产增加值、三产增加值之和。其次是要检查不同层级、同一要素之间的关系,如国家级年鉴中发布的省区 GDP 与省级年鉴中发布的该省区内部全部县、市的 GDP 总和之间的关系;如不相等,则应该采取一定的技术手段进行平差。

(2)空间匹配:将空间数据转换到相同的坐标系系统上(如西安坐标系 1980,中国国家坐标系 2000),转换到适合经济地理分析投影类型上(如 Albers 等积圆锥投影,双标准纬线分别为 25°和 47°,中央经线为 105°),统一到适合中国国家尺度社会经济分析的 1km 格网上,统一到相同的空间起算位置(即栅格数据的左下角坐标相同)。

(3)代码匹配:使用民政部、国家标准委员会等单位公布的标准行政区代码将空间数据以及业务数据(社会经济数据)中的相关代码统一起来,并在 Arc/Info 支持下进行空间连接,由此实现空间数据和业务数据的逻辑连接。

(4)空间离散:即在 GIS 软件的空间分析模块支持下,通过批处理或高级编程方式(如在 ArcInfo Workstation 下通过 AML 语言编程、或者是在 Windows 图形界面下基于 ArcEngine 组件编程),将以行政区域为单元的社会经济类数据,按照一定的规则分配到规则的地理格网。在空间离散过程中,构建科学合理和现实可行的空间离散模型是关键。

(5)检查校验:针对空间离散所得到的社会经济格网数据,在 GIS 软件空间统计功能和显示功能支持下进行分析和比对。即通过浏览社会经济格网数据的空间分布格局、比对成果数据与资源环境背景要素的空间关系,检查和评价离散模型是否科学合理;通过空间统计将离散成果回算至行政区单元,检查空间离散模型及离散过程是否正确无误。

3.2 人口数据的空间离散过程和模型

人口数据空间离散按如下顺序进行:城镇人口(CTPOP)、农村人口(COPOP)、总人口(TPOP)、男性人口(MAL_POP)、女性人口(FEM_POP)、老年人口(OLD_POP)、儿童人口(CHD_POP),具体方法和模型如下:

在全国尺度上可以将城镇人口在城镇斑块内部的空间分布视作均匀,也就是说可以在县(县级

市)斑块上直接将城镇人口按城镇斑块面积平摊。具体可表示为:

$$CTPOP_i = CTPOP \cdot \frac{S_i}{\sum_{i=1}^n S_i} \quad (1)$$

式中, $CTPOP_i$ 是该县(县级市)城镇用地地块栅格 i 上的人口, $CTPOP$ 为该县(县级市)的统计城镇人口总量, S_i 为栅格 i 的面积, 该县(县级市)共有 n 个城镇用地地块栅格, $\sum_{i=1}^n S_i$ 即为该县城镇用地总面积。

农村人口的空间分布具有较大的不均匀性, 且依赖于土地利用类型、居民点分布、距交通线远近等多种自然地理和人文地理要素。为此, 可以层次分析方法, 逐一考虑各种影响因子对人口的影响(主要是距离衰减作用、斑块限定作用)、影响力衰减梯度以及影响力的权重, 建立加权多因子影响力距离衰减模型。在得到各栅格点的影响力总和 W_i 之后, 将 W_i 视作面积调整系数, 与城镇人口空间离散模型相仿, 建立农村人口的空间离散模型。具体可表示为:

$$W_i = \sum_{j=1}^m \sum_{i=1}^n K_j \cdot D_{ij} \quad (2)$$

$$COPOP_i = COPOP \cdot \frac{W_i \cdot S_i}{\sum_{i=1}^n W_i \cdot S_i} \quad (3)$$

式(2)中, W_i 为栅格 i 上加权多因子综合影响力, 也即栅格 i 上的人口(农村人口)调整系数; i 为栅格序号, 该县(县级市)内共有 n 个农村用地地块栅格; j 为因子编号; 本模型中共考虑 m 个因子; K_j 为第 j 个因子在整个模型影响力的权重值; D_{ij} 为 j 因子在栅格 i 上的影响力。

式(3)中, $COPOP_i$ 是某县(县级市)农村用地地块栅格 i 上的人口, $COPOP$ 是该县(县级市)的统计农村人口总量, S_i 为栅格面积, W_i 为加权多因子综合影响力。

总人口与城镇人口、农村人口之间存在内在的加和平衡关系, 因此, 总人口格网数据是由城镇人口格网与农村人口格网叠加生成。具体可示表为:

$$TPOP_i = CTPOP_i + COPOP_i$$

式中, $TPOP_i$ 为任意栅格 i 处的人口, $CTPOP_i$ 为栅格 i 处的城镇人口, $COPOP_i$ 为栅格 i 处的农村人口数, i 为栅格序号。

在得到总人口($TPOP$)格网数据后, 可以依据

行政区统计单元上后续各指标项(即男性人口(MALE_POP), 女性人口(FEM_POP)、老年人口(OLD_POP)、儿童人口(CHD_POP))等占该行政区总人口($TPOP$)的比例数据, 得到其他各项指标。具体可表示为:

$$X_POP_i = TPOP_i \cdot X_i \quad (4)$$

式中, X_POP_i 为任意栅格 i 处的特征人口(如: 男性人口(MALE_POP), 女性人口(FEM_POP)、老年人口(OLD_POP)、儿童人口(CHD_POP)等), $TPOP_i$ 为栅格 i 处的总人口; X_i 为栅格 i 处的特征人口占总人口的比例, 它是逐县域变化的; i 为栅格序号。

3.3 经济数据的空间离散过程和模型

经济数据的空间离散将按照“先行业, 后综合”的顺序进行。具体将按照如下顺序开展: 服务业各部门产值(批发和零售业产值(SALE), 交通运输仓储和邮政业产值(TRAN)、住宿和餐饮业产值(ACOM)、金融业产值(FINAN)、房地产业产值(ESTAT), 其他服务业产值(OTH3); 工业各部门产值(制造业产值(PROD), 石化产业产值(PET_CHEM)、冶金产业产值(METAL)), 其他工业行业产值(OTH2), 工业总产值(IND), 建筑业产值(CONST); 一产增加值(GDP1), 二产增加值(GDP2), 三产增加值(GDP3), 国内生产总值(GDP), 人均GDP(GPD_POP)等顺序进行。

第二和第三产业下的各部门产业活动主要发生在城镇地区, 因此在全国尺度上可以将上述部门产值在城镇斑块内部的空间分布视为均匀的, 也可在县(县级市)斑块上直接将部门产值按城镇斑块面积平摊。具体可参照城镇人口的离散模型, 此处不再赘述。

第一产业主要存在于农村地区, 且由农业、林业、牧业、渔业等 4 个产业部门组成; 考虑到在国家尺度上, 农、林、牧、渔等 4 个行业的产值可以视为均匀分布于耕地、林地、草地、水体等 4 类土地覆盖/土地利用类型上。因此, 可以首先建立分土地利用类型影响第一产业增加值空间分布权重层; 在得到权重层之后, 再对第一产业增加值进行离散。其具体方法可以参照乡村人口的离散模型, 此处不再赘述。

在得到第二产业、第三产业下各部门的产值以及第一产业增加值后, 可以利用上述各项指标间的

内在的平衡关系,采用格网数据叠加的方法,得到第二产业增加值(GDP2)、第三产业增加值(GDP3)以及国内生产总值(GDP)总量的空间分布数据。最后,将 GDP 总量格网数据(GDP)与总人口(TPOP)数据叠加,采用格网数据库相除方法,得到人均国内生产总值数据(GDP_POP)。

3.4 不同层级数据的整合与离散策略

在社会经济数据格网化过程中,通常会遇到不同层级的数据源。例如,对于某一地区的 GDP 数据,同时可从国家级年鉴以及省区一级的年鉴中获得,但省级年鉴和国家级年鉴数据之间往往会存在一定的差异。对于这种不同层级来源数据的空间化,应当采取一定的策略和模型进行预处理。

一般地,考虑到省级年鉴可以提供到较高精度的统计数据(到县、县级市一级),但数据的权威性略差,数据指标项在时间序列上通常不连续;而国家级年鉴通常只能提供较粗分辨率的统计数据(到省、地区或地级市一级),但数据权威性较高,并且数据指标项在时间序列上连续且完整。因此一般以县级统计数据获取的高精度空间数据作为空间分布格局控制,用相同或者相近年份的省区(地区)的数据作为总量控制进行空间离散。

具体离散模型可表示为:

$$X_{Mi} = X_M \cdot \frac{X_{0i}}{\sum_{i=1}^k X_{0i}} \quad (6)$$

式中, X_{Mi} 是 M 年某省区(地区)内部任意栅格上指标 X 的空间格点数值, X_M 为该省区(地区)指标 X 的统计数值, k 为该省区(地区)内全部的格点数,为已知的某一年指标 X 的空间数据格点值。

4 结论和问题

本文在分析传统关系型社会经济数据库不足的基础上,介绍了国家尺度社会经济数据格网化的概念、要求、指标体系、格网化步骤、离散模型以及相关策略。鉴此,作者构建了 1990 - 2010 年、每 5 年一次的中国 1km 分辨率人口和社会经济空间格网数据库。并将其纳入国家发改委宏观经济研究院的基本数据库系列,且应用于国家主体功能区规划、京津冀都市圈区域规划等项目中。其中也存在一些问题,主要有:

(1) 研究人员对于社会经济的运行方式、关联及决定要素、表达形式和表达内容等存在不同看法,对社会经济要素空间分布及其运行规律的数学模型研究存在诸多假设之处。针对社会经济要素内涵以及空间化模型认识的不确定和不统一,不仅会导致不同研究人员所得空间化成果之间存在明显差异,同时也对空间离散成果的精度检验和可靠性评价带来疑问。

(2) 社会经济数据格网化过程中所涉及的统计数据及相应的环境背景数据存在时间匹配困难以及处理工作量巨大的特点。社会经济空间数据库建设过程中一般都会使用土地利用与土地覆被数据(LUC: Land Use / Land Cover)作为地理环境控制因素之一;但是由于以计算机辅助、人工目视解译技术得到的国家级 LUC 数据集更新速率相对较慢,这会影响到社会经济数据空间离散化的制作。

参考文献:

- [1] Goodchild M F, Anselin L, Deichmann U A. A Framework for the Areal Interpolation of Socioeconomic Data [J]. Environment and Planning A, 1993, 25: 383 - 397.
- [2] 陈述彭,陈秋晓,周成虎. 网格地图与网格计算[J]. 测绘科学, 2002, 27(4): 1 - 6.
- [3] <http://sedac.ciesin.columbia.edu/gpw/>. [DB/OB].
- [4] Wilson S J, Steenhuisen F, Pacyna J M, et al. Mapping the Spatial Distribution of Global Anthropogenic Mercury Atmospheric Emission Inventories[J]. Atmospheric Environment, 2006, 40(24): 4621 - 4632.
- [5] <http://www.ornl.gov/sci/landscan/>. [DB/OB].
- [6] Dobson J E, Bright E A, Coleman P R, et al. LandScan: A Global Population Database for Estimating Populations at Risk[J]. Photogrammetric Engineering and Remote Sensing, 2000, 66(7): 849 - 857.
- [7] Nordhaus W, Azam Q, Corderi D, et al. The G-Econ Database on Gridded Output: Methods and Data[M]. Yale University, New Haven, 2006.
- [8] Nordhaus W D. Geography and Macroeconomics: New Data and New Findings[C]. Proceedings of the National Academy of Sciences of the United States of America, 2006, 103(10): 3510 - 3517.
- [9] 江东,杨小唤,王乃斌,等. 基于 RS、GIS 的人口空间分布研究[J]. 地球科学进展, 2002, 17(10): 734 - 738.
- [10] 刘红辉,江东,杨小唤,等. 基于遥感的全国 GDP 1km 格网的空间化表达[J]. 地球信息科学, 2005, 7(2):

- 120-123.
- [11] 廖顺宝, 孙九林. 基于 GIS 的青藏高原人口统计数据空间化[J]. 地理学报, 2003, 58(1): 25-33.
- [12] 陈宣庆. 关于我国区域规划问题的探讨[J]. 宏观经济管理, 2005, (7): 17-20.
- [13] 胡云锋, 曾澜, 李军, 等. 新时期区域规划的基本任务与工作框架[J]. 地域研究与开发, 2010, 29(4): 6-9.
- [14] 刘卫东, 甄峰. 信息化对社会经济空间组织的影响研究[J]. 地理学报, 2004, 59(s1): 67-76.
- [15] 杨瑞霞, 张莉, 闫丽洁, 等. 省级主体功能区规划支持系统研究[J]. 地域研究与开发, 2009, 28(1): 22-26.
- [16] 王雪梅, 李新, 马明国. 基于遥感和 GIS 的人口数据空间化研究进展及案例分析[J]. 遥感技术与应用, 2004, 19(5): 320-327.
- [17] 史育龙, 申兵, 曾澜, 等. 区域规划信息空间整合的指标体系研究[J]. 地理信息世界, 2006(3): 49-56.

Index System and Transferring Methods to Build the National Society and Economy Grid Database

HU Yunfeng¹, WANG Qianqian^{1,2}, LIU Yue¹, LI Jun³, REN Wangbing³

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

2. *Department of Geodesy and Geographic Sciences, Liaoning Technical University, Fuxin 123000, China;*

3. *Academy of Macro-economy Research, Beijing 100038, China)*

Abstract: With the developments of modern space information sciences, specifically relying on the quickly updated and large scale remote sensing observation technology, the derivative LUC (land use and land cover) data generations, the strong spatial analysis functions and the visual information representation functions supported by geographic information system (GIS) technology, the traditional social and economic statistics data with Relational format can be transferred to the spatial data (Vector or Grid format). This paper firstly analyzed the shortcomings of traditional statistical data in the new-era applications, which characterized by long update-period, low spatial-resolution, poor visualization effects and non-support for the spatial computing and spatial analysis. We briefly reviewed the history of theoretical studies and database constructions of society and economy grid database. Based on the clarification about the concepts of grid process of society and economy statistics data, three basic requirements, i. e. temporal compatibility, space reference system compatibility and self-consistent logic, were put forward as the main principles in the transferring process, and a series of key index including 24 items to build a national scale society and economy grid database were posted. Then, based on the above data transferring principles and index system, we discussed the 5-step transferring processes which include (1) logical checking, (2) spatial matching, (3) code matching, (4) spatial discretization, and (5) checking and calibration. We also analyzed the general models to transfer the rational datasets into grid dataset in detail, and discussed the corresponding strategies when people integrate different datasets. Finally, we summarized the common questions and suggested that the spatial discrete model construction and the quick update routine works to prepare the background resources / environment datasets are the main focuses in the future studies.

Key words: national scale; relationship data; spatial discretization; grid data