

土地利用变化数据的误差特性

朱会义

(中国科学院地理科学与资源研究所, 北京 100101)

摘要: 土地利用变化数据, 主要由不同时点的土地利用数据通过统计、对比、空间叠置分析等途径派生而来; 而不同时点的土地利用数据主要来源于遥感数据、土地调查数据、历史图件数据以及统计数据。由于这些数据来源于本身不可避免地带有不同程度的误差, 由此派生的变化数据也就具有不同的精度。现有的文献中通常会给出时点数据的误差, 但很少关注土地利用变化数据的误差。本文首先建立变化数据的误差分析方法, 然后举例分析了土地利用类型变化幅度、时点数据误差及其不同组合对变化数据误差的影响。最后得出如下结论: 时点数据的高精度并不能完全保证变化数据的高精度, 还要看误差的方向是否一致; 土地利用变化幅度越小, 时点数据的误差对变化数据的精度影响越大, 精度水平越难保证。这些结论提醒我们, 土地利用变化研究中要对数据误差问题给以足够的重视, 并应从数据源质量、数据获取方式、数据分析方法等方面提高数据的精度水平。

关键词: 土地利用变化; 数据; 误差

1 引言

土地利用变化研究, 除了纯理论分析或研究综述, 无论时空尺度如何, 都离不开具体数据的支持。目前相关研究中所用的数据, 基本上有如下一些来源^[1-12]: (1) 遥感数据, 利用 TM、SPOT、MODIS、航空相片等遥感资料, 通过自动分类或人工目视解译提取的不同时点的土地利用数据。这是一种主流来源; (2) 土地调查数据, 来自国家土地部门组织的土地详查与概查以及变更调查, 具体调查中实际也参考了遥感资料; (3) 历史图件数据, 包括地形图、早期的土地利用图、土地资源图等, 用于弥补早期遥感数据与土地调查数据的缺乏; (4) 统计数据, 来自于政府部门的各种统计资料, 见用于部分宏观分析或驱动力分析中。

这些数据, 在不同的精度水平上, 反映了一定时点特定区域的土地利用状况。一般认为前 3 类数据精度较高, 对空间分布的反映更详细, 国内外也多认可。不过, 不同区域土地利用情况不同, 调查人员的素养与工作态度相异, 调查手段与分类方法也不完全一致, 所以数据本身带有不同程度的误差。这些误差的大小有的可以通过实地抽样进行校验, 有的因时过境迁而难以评定, 精度如何依赖于数据采集时对误差水平的说明。

将上述数据应用到区域土地利用变化研究中, 还不免要对上述不同时点、不同来源的数据进行统计、对比、或在地理信息系统环境下进行空间叠置分析, 这样才能揭示土地利用类型间的转移与空间变化特征。具体工作中, 有时也将一个时点的土地利用数据直接叠置在另一时点的遥感影像上, 直接目视解译出土地利用变化情况, 但这种数据获取方式仍然是由两个时点数据的叠置分析来实现的, 对误差分析而言, 与两个以上时点

收稿日期: 2003-04-11; 修订日期: 2004-03-01

基金项目: 国家自然科学基金资助项目 (40271008) [Foundation: National Natural Science Foundation of China, No. 40271008]

作者简介: 朱会义 (1966-), 男, 副研究员, 博士。主要从事土地利用 / 覆被变化、地理信息系统应用研究。

E-mail: zhuhy@igsnr.ac.cn

图形数据的叠置分析并无本质上的差别。这些过程都可归结为由刻画状态的数据派生出刻画状态变化的数据。变化数据的误差主要来源于状态数据的误差,是误差传递的结果。在目前接触到的文献中,还没有看到对变化数据误差的讨论与说明。直觉告诉我们,即使状态数据的误差满足分析的精度要求,由于2个以上状态数据的误差传递与组合影响,变化数据不一定能满足分析精度。数据的可靠性关乎科学研究的立论基础,所以有必要探讨一下变化数据的误差特性。

2 分析方法

对于一幅能满足精度要求的土地利用图,其误差包括定性误差与定量误差两个方面。定性误差是指类型判别方面的偏差,定量误差则指空间界线勾画上的偏差,具体反映在制图图斑的类型识别与多边形范围的描绘上。在遥感解译工作中,往往会对这两种误差提出具体精度要求,作为抽样校验考核的标准。两者集合在一起,实际上反映的是土地利用各种类型的误差,即各类型面积较实际值的偏差。所以不妨从类型误差着手来分析变化数据的误差特性。

首先来确立分析方法。先设某种土地利用类型在状态1时的实际面积为 S_{1a} ,制图或调查所得面积为 S_1 ,在状态2时的实际面积为 S_{2a} ,制图或调查得到的面积为 S_2 ,状态1到状态2相应的变化面积分别为 S_a 和 S ;对应的相对误差分别为 r_1 、 r_2 、 r 。那么就有如下的关系式:

$$S_{1a} = S_1 / (1 - r_1) \quad (1)$$

$$S_{2a} = S_2 / (1 - r_2) \quad (2)$$

$$S_a = S_{2a} - S_{1a} \quad (3)$$

$$S = S_2 - S_1 \quad (4)$$

$$S_a = S_{2a} - S_{1a} = S / (1 - r) = (S_2 - S_1) / (1 - r) \quad (5)$$

根据上述表达式(1)~(5)有:

$$S_2 / (1 - r_2) - S_1 / (1 - r_1) = (S_2 - S_1) / (1 - r) \quad (6)$$

$$r = 1 - \frac{(S_2 - S_1)}{\frac{S_2}{1 - r_2} - \frac{S_1}{1 - r_1}} \quad (7)$$

再设 p 为变化面积与状态1时测得面积的百分比,即 $p = S/S_1$,代入式(7),那么有:

$$S_1 = S/p \quad (8)$$

$$S_2 = S_1 + S = S/p + S = S(p + 1)/p \quad (9)$$

$$r = 1 - \frac{p(1 - r_1)(1 - r_2)}{(p + 1)(1 - r_1) - (1 - r_2)} \quad (10)$$

上述表达式中, p 、 r_1 、 r_2 和 r 为实数,既可以取正值,也可取负值。由此,我们得到了变化数据的误差与变化面积百分比、状态1和状态2数据误差 r_1 与 r_2 之间的数量关系。该关系式表明,变化数据的误差不仅与两个状态误差有关,同时也与变化的幅度有关。考虑到变化误差的性质,这是合理的。

由变化误差的关系式(10),不难发现两个特异点。其一是当 r_1 等于 r_2 时,即两个状态误差相等时(误差值相等,误差方向相同),变化误差与变化面积百分比无关,恒等于状态误差。其二是当 $p \rightarrow (r_1 - r_2)/(1 - r_1)$ 时, $r \rightarrow \infty$,即 p 离该数值越远,变化误差越小,反之,则变化误差越大。下面以实例来探讨这些因素对变化数据误差的影响。

3 实例分析

3.1 变化面积百分比与变化误差

下面以北京市 1985 年与 1995 年的耕地遥感调查结果来分析两期数据误差与该时段耕地变化误差的关系^[1]。这两期遥感调查数据的制图比例尺分别为 1:25 万和 1:10 万, 根据类型图斑抽样统计的结果, 订正前误差大致分别在 10% 与 5% 的水平上, 变化幅度约为 -19.3%。我们先根据公式 (10) 采用枚举法勾画变化曲线, 来考察两个状态误差已知和变化面积百分比由 -100% 到 100% 过程中变化数据误差的特性, 然后再来计算耕地类型变化的误差。

从状态误差分别为 10% 与 5%, 而 p 由 -100% 至 100% 变动时的变化数据误差曲线可看出 (图 1), 当状态误差分别等于 10% 和 5% 时, 如果 p 的绝对值越大, 变化误差越小; p 不管从正向还是负向趋近 5/90 时, 误差都趋向于无穷大。根据误差计算式, 容易推出当 $p \rightarrow (r_1 - r_2)/(1 - r_1)$ 时, S_{1a} 与 S_{2a} 趋于相等, 两者相等时土地利用类型面积实际上没有变化。怎么理解呢? 设想有这样一个区域, 其耕地面积的真实值为 1 000 hm^2 , 实际上耕地面积并未发生变化, 但由于数据获取过程中带来的误差, 提取时点 1 时的误差为 10%, 所得耕地面积的结果为 1 100 hm^2 , 提取时点 2 时的误差为 5%, 所得结果为 1 050 hm^2 。根据两期数据的对比或空间叠置分析, 将得到耕地的面积变化大致为 -50 hm^2 。但实际上耕地面积并未变化, 变化的结果完全是由数据误差造成的。再根据误差关系式 (10) 的计算结果, 该实例中耕地变化数据的误差为 26%, 远高于两期耕地调查数据的误差。

上述现象说明, 即使每一时点的土地利用数据都能满足很高的精度要求, 如误差在 10% 以下, 但用以生成土地利用变化数据时, 所得结果的精度可能根本满足不了分析要求 (如本例中为 26%), 甚至产生误导。所以有必要根据本文提出的误差分析方法, 考察一下所得结果的误差水平, 对于变化幅度小的类型尤其要给予足够的重视。

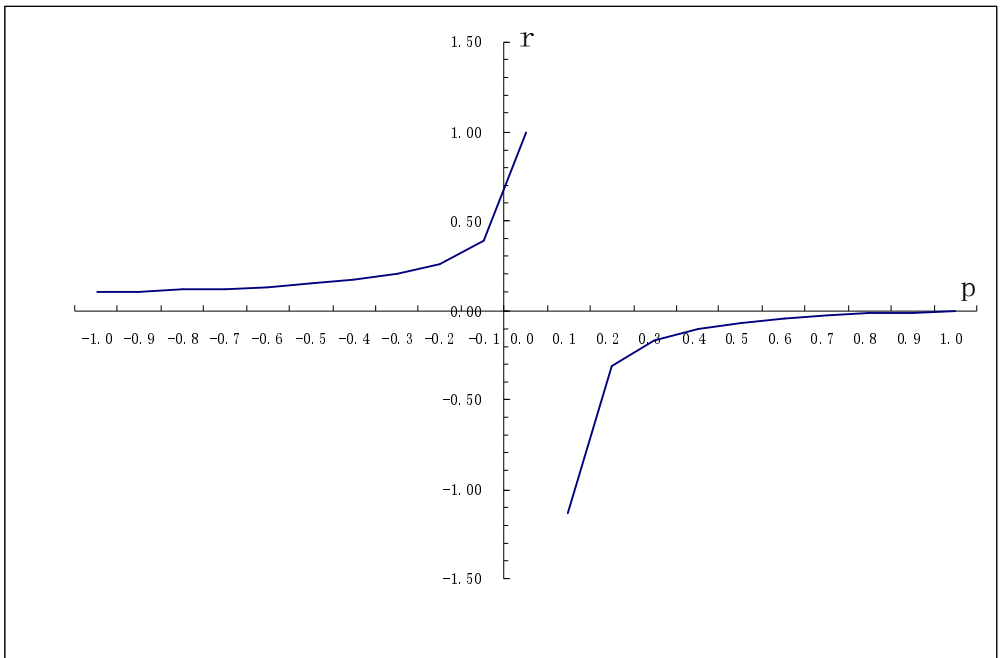


图 1 状态误差为 10% 和 5% 时的变化误差曲线

Fig. 1 Error variation with p at r_1, r_2 equals 10% and 5% respectively

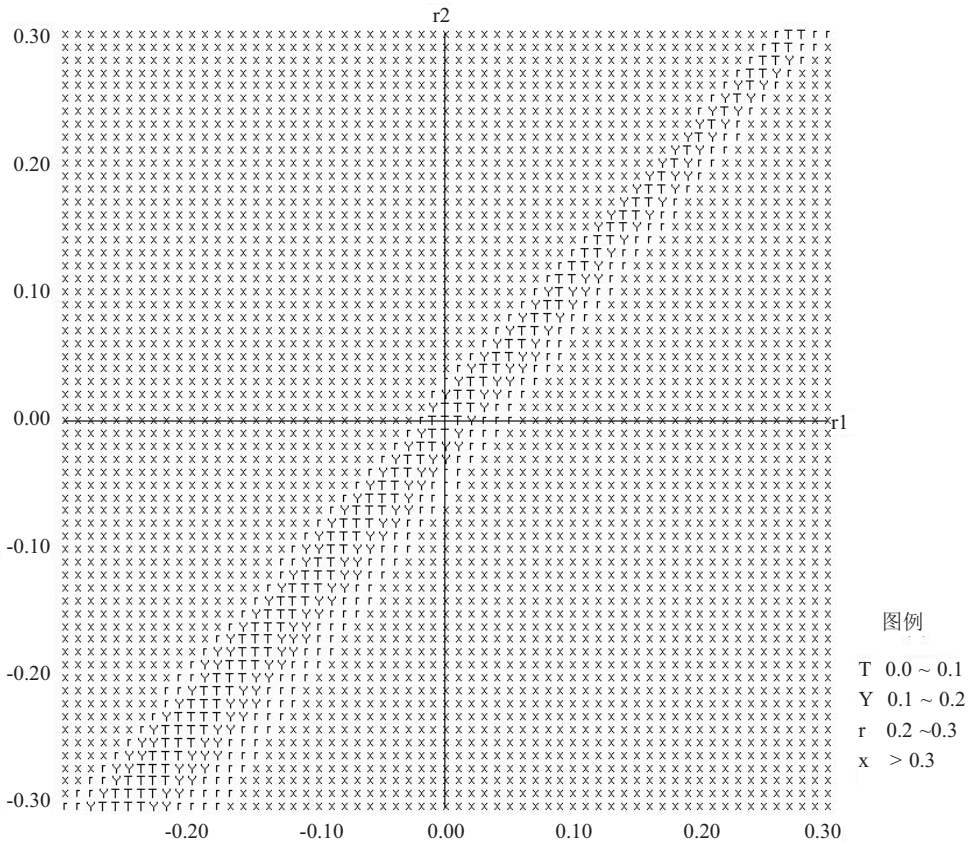


图 2 p 为 -0.12 时不同状态误差组合带来的变化误差特性

Fig. 2 The characteristics of change data error caused by different combinations of period data error as p equaling -0.12

3.2 状态误差与变化误差

以上重点考察了变化幅度 (变化面积占初始状态面积的百分比实际反映了类型变化的幅度) 与变化数据误差的关系。下面再来考察状态误差的不同组合对变化误差的影响。根据环渤海地区研究的经验^[1], 该区在 1985~1995 年之间耕地面积变化幅度为 -12.6%。据此假设变化面积百分比 p 为 -12%, 分析在此条件下状态误差的组合如何影响变化数据的误差。

图 2 为 p 等于 -0.12 时, 以 r_1 为横坐标、 r_2 为纵坐标和 r 为属性值, 在 ArcView 环境下编程计算的图形结果。图例中各种符号指代的数值为变化误差的绝对值。由图 2 可以看出, 当状态误差分别在 -30%与 30%之间变化时, 变化误差会出现非常复杂的变化, 其误差甚至可以达到无穷大。其中一个最为明显的特征是当状态误差的误差方向相同、误差值差别不大时 (其组合分布呈沿 45° 对角线方向的一个条带), 变化数据的误差相对较小; 如果状态误差的方向不同, 不同时为正或为负, 则状态误差即使在正负 10% 的范围内, 也不一定能保证变化数据的精度, 只有在其误差值都很小的时候, 变化误差才能达到较高的精度。如果要对状态误差、变化误差都给以适当的精度限制, 如都要求误差值小于 10%, 那么只有离对角线很近且接近原点的状态误差组合才能满足要求。怀疑该结果可能受土地利用类型变化方向性的影响, 尝试将 p 设为 0.12, 在进行同样的制图分析, 仍然存在类似的现象, 说明这是变化数据误差的一个一般特性。

变化误差的这一特性表明, 在土地利用变化研究中, 为了保证研究数据的精度, 不仅要注意时点数据误差的大小, 还要注意该类误差的方向, 是偏大了还是偏小了。因为误差水平即使都在 10% 以内的时点数据, 如果其误差方向不同, 变化数据的分析结果极有可能在 30% 以上, 而这种精度水平很难满足进一步分析的需要。状态误差的方向性目前显然没有引起足够的重视。

4 结论与讨论

综合以上分析, 可以得出这样一些结论: (1) 土地利用变化数据, 由于来自两个以上时点土地利用数据的分析结果, 会出现状态误差的传递现象, 其误差水平受变化面积百分比的大小以及状态误差组合的复杂影响。(2) 状态误差相等时, 即误差值相等、方向相同时, 变化误差总是等于状态误差。(3) 状态误差不等时, 变化误差存在一个无穷大点。土地利用类型变化幅度越接近于该点, 变化误差越大, 此时所得结果甚至可能完全由状态误差产生, 从而产生误导。(4) 状态数据即使都具有较高的精度, 所得变化数据的结果不一定能满足精度要求, 还与误差的方向有关。

上述结论从土地利用类型变化误差的角度揭示了土地利用变化数据的误差特性。由于土地利用数据更多情况下呈现为空间数据的形式, 而空间数据的误差水平在各个类型斑块上并不会完全相同, 所以土地利用类型时点数据误差的评定将依赖于抽样验证时图斑样本的选取。这不仅涉及到图斑样本的类型、样本容量、样本的形态特征, 还涉及到样本的空间代表性等方面, 具体工作中可引入数学、测量学中有关的误差理论以及分析和处理方法。总之, 土地利用变化研究不能忽视变化数据的误差问题, 应采用相关误差分析方法对数据误差进行评定, 以明了数据的可靠程度。

不过, 核心问题仍然是如何提高土地利用变化数据的精度。该问题可进一步归结为如何提高时点土地利用数据的精度以及协调各时点数据的误差水平, 也就是要在减少数据获取与处理过程中可能产生的系统误差与人为误差的同时, 进一步关注以下几个环节: 土地利用分类系统应尽可能相同; 信息源应具有尽可能相同的空间分辨率; 影像处理、自动分类应尽可能在形同的模式下进行; 目视解译的类型判别标准、图斑界线的勾画精度也尽可能一致; 分类结果应保持相同的比例尺等。

参考文献 (References)

- [1] Shi Peijun, Gong Peng, Li Xiaobing et al. The Method and Practice of Land Use/Land Cover Change. Beijing: Science Press, 2000. [史培军, 宫鹏, 李小兵等. 土地利用 / 土地覆被变化研究的方法与实践. 北京: 科学出版社, 2000.]
- [2] Xu Lan, Zhao Yi. Forecast of land use pattern change in Dongling District of Shenyang: an application of Markov process. Chinese Journal of Applied Ecology, 1993, 4(3): 272-277. [徐岚, 赵羿. 利用马尔柯夫过程预测东陵区土地利用格局的变化. 应用生态学报, 1993, 4(3): 272-277.]
- [3] Tsai Borwen, Chang Changyi, Ding Tsujen. Spatial analysis in GIS: the land use changes in the coastal area of Yunlin county, Taiwan. Journal of Geographical Science (National Taiwan University), 1997, (23): 1-12.
- [4] Liu Shenghe, Wu Chuanjun, Shen Hongquan. A GIS based model of urban land use growth in Beijing. Acta Geographica Sinica, 2000, 55(4): 407-416. [刘盛和, 吴传钧, 沈洪泉. 基于 GIS 的北京城市土地利用扩展模式. 地理学报, 2000, 55(4): 407-416.]
- [5] Xie Gaodi, Cheng Shengkui, Ding Xianzhong. A study on global land use change under the pressure of population growth. Journal of Natural Resources, 1999, 14(3) : 193-199. [谢高地, 成升魁, 丁贤忠. 人口增长胁迫下的全球土地利用变化研究. 自然资源学报, 1999, 14(3): 193-199.]
- [6] Zhang Yili, Li Xiubin, Fu Xiaofeng. Urban land use change in Lhasa. Acta Geographica Sinica, 2000, 55(4): 395-406. [张懿铨, 李秀彬, 傅小锋. 拉萨城市用地变化分析. 地理学报, 2000, 55 (4): 395-406.]
- [7] Brogaard Sara, Prieler Sylvia. Land cover in the Horqin grasslands, north China: detecting changes between 1975 and 1990 by means of remote sensing. IIASA Interim report IR-98-044/July.

- [8] Zhang Yili, Yan Jianzhong, Liu Linshan et al. Impact of Qinghai-Xizang Highway on land use and landscape pattern change. *Acta Geographica Sinica*, 2002, 57(3): 253-266. [张懿铨, 阎建忠, 刘林山 等. 青藏公路对区域土地利用和景观格局的影响. *地理学报*, 2002, 57(3): 253-266.]
- [9] Wang Siyuan, Liu Jiyuan, Zhang Zengxiang et al. Analysis on spatial-temporal features of land use in China. *Acta Geographica Sinica*, 2001, 56(6): 631-639. [王思远, 刘纪远, 张增祥 等. 中国土地利用时空特征分析. *地理学报*, 2001, 56(6): 631-639.]
- [10] Zhu Huiyi, He Shujin, Zhang Ming. Driving forces analysis of land use change in Bohai Rim. *Geographical Research*, 2001, 20(6): 669-678. [朱会义, 何书金, 张明. 环渤海地区土地利用变化的驱动力分析. *地理研究*, 2001, 20(6): 669-678.]
- [11] Zhu Huiyi, Li Xiubin, He Shujin et al. Spatial-temporal change of land use in Bohai Rim. *Acta Geographica Sinica*, 2001, 56(3): 253-260. [朱会义, 李秀彬, 何书金 等. 环渤海地区土地利用的时空变化分析. *地理学报*, 2001, 56(3): 253-260.]
- [12] Li Xiubin et al. The spatial relation analysis of land use change in typical region: in project research report of land use change and sustainable land use models in Bohai Rim, 2002, 15-18. [李秀彬 等. 典型地区土地利用变化的空间关系分析——环渤海地区土地利用变化与土地持续利用模式项目研究报告, 2002, 15-18.]

The Error Characteristics of Land Use Change Data

ZHU Huiyi

(Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China)

Abstract: Land use change data are mainly derived from land use data of different periods through statistical analysis or spatial analysis in GIS environment, and those land use data usually came from remote sensing images, land survey, historical maps or regional statistical information. Because of the errors in those land use data, land use change data are in dissimilar precision, but the available papers on land use change have not paid more attention to the point, but gave some descriptions of the error of land use data. Aimed at the error characteristics of land use change data, error analysis method was firstly discussed in this paper, and then the impacts from influencing factors, including the percentage of land use change and the error of periodical land use data, were analyzed in detail to a certain extent. The result showed that high precision of land use change data cannot always be guaranteed by high precision of land use data, which also depends on the error direction and the percentage of changed area. The results imply that the error of land use change data have been ignored at present and should be improved in the process of land classification, image collection, processing, interpretation and analysis.

Key words: land use change; data; error