

第二节 数据的误差来源和分类



知识点



地理空间数据源误差

The diagram features a large, light blue semi-circle on the left side. A thin, light yellow curved line extends from the top of this semi-circle, passing through three circular nodes. Each node is a small circle with a metallic border and a colored center (yellow, green, and orange from top to bottom). To the right of each node is a horizontal, light purple rounded rectangle containing text. The background is a light blue sky with white clouds.

数据转换和处理的误差

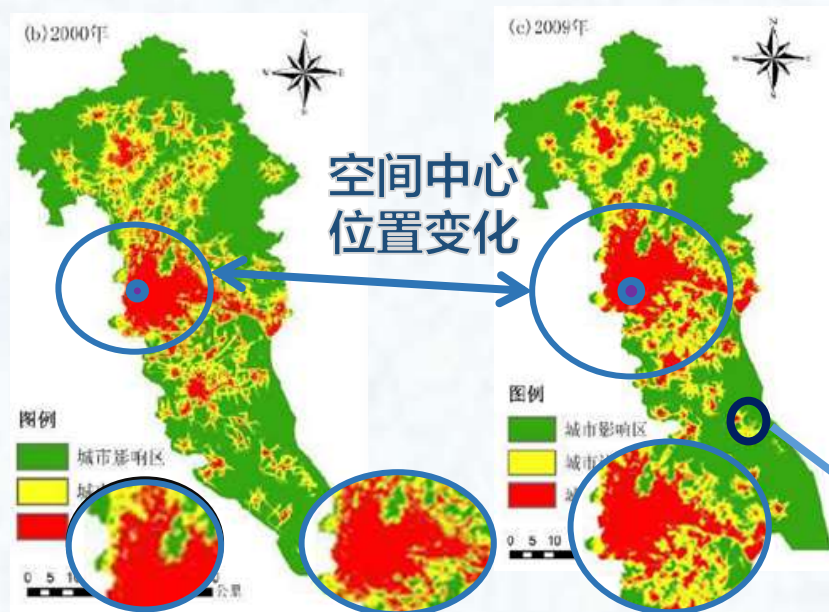
应用分析产生的误差

地理空间数据误差的来源与分类

误差来源---空间现象自身的不稳定性

空间现象自身存在的不稳定性包括空间特性和过程在空间、专题、时间和内容上的不确定性。

数据源误差



2000年

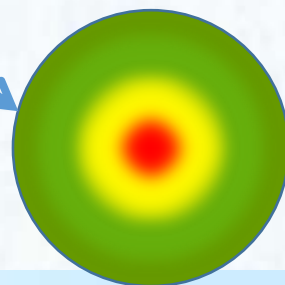
2009年

空间现象的游移性

空间上的不确定性

时间上的不确定性

属性的不确定性



边界的模糊性

地理空间数据误差的来源与分类

误差来源---空间现象的表达

数据源误差

定义

概念理解的不一致性导致数据测量误差。



测量

测量仪器本身有一定的设计精度。



表达方式

图形的表达不同会产生不同的误差。



物理介质的变化



地理空间数据误差的来源与分类

数据源误差

误差的表现

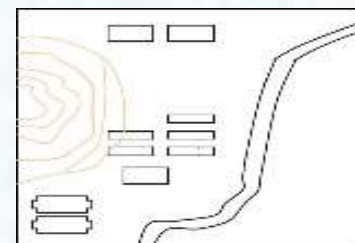
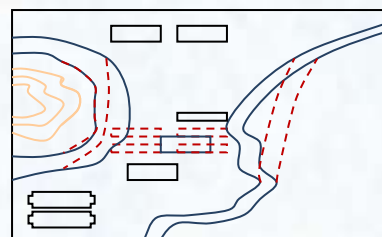
- 时域误差
数据采集周期不一致。
- 逻辑不一致性误差
要素之间空间冲突。
- 数据不完整性误差
同一目标的分割。



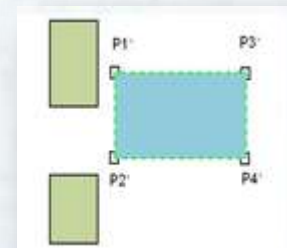
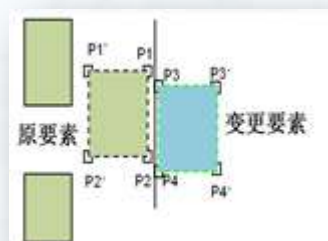
道路采集周期：1年



建筑物采集周期：3年



对更新后变化要素进行修改



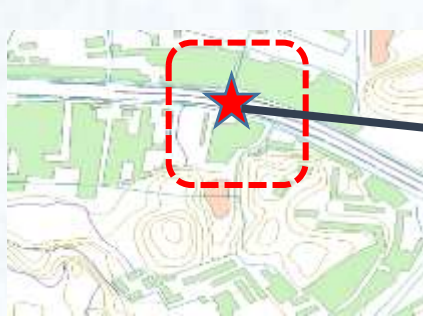
对同一要素进行合并

地理空间数据误差的来源与分类

质量控制---对数据源选择和处理

数据源误差

选择满足系统和应用要求的数据源



1:10000 旅游地图



更大比例尺 宗地测量

数据源的误差至少不能大于系统对数据误差的要求。

逐步减少或取消不必要的数据处理中间环节



让用户根据具体的应用进行相应的数据处理

地理空间数据误差的来源与分类

质量控制---属性误差的检测与表达

确定抽样方法

- 系统分区随机抽样：对整个区域内等面积分区，并在区内随机布点。

确定抽样数

- 使用其他方法确定每个样本的属性做为参考数据。

建立误差矩阵

参考数据

被检测数据

$$\begin{bmatrix} X_{11} & X_{12} & \dots & X_{1N} & R_{1N} \\ X_{21} & X_{22} & \dots & X_{2N} & R_{2N} \\ \dots & \dots & \dots & \dots & \dots \\ X_{N1} & X_{N2} & \dots & X_{NN} & R_{NN} \\ C_{N1} & C_{N2} & \dots & C_{NN} & T_{NN} \end{bmatrix}$$

行累加

列累加

误差矩阵举例

	盐碱地	非盐碱地	水体	综合
盐碱地	28993	1830	375	31198
非盐碱地	609	118220	1131	119960
水体	13	1960	182275	184248
综合	29615	122010	183781	335406

地理空间数据误差的来源与分类

质量控制---属性误差的检测与表达

计算误差的常用指标

数据源误差

总准确度

对角线样点数之和除以总样点数

生产者准确度

每个属性值的正确样点数除以该属性值总的验证样点数

使用者准确度

每个属性值的正确样点数除以该属性值的总数

遗漏误差

生产者准确度 (列合计)

错判误差

使用者准确度 (行合计)

误差矩阵

计算例子：

$$(176 + 127 + 277 + 129) / 756$$

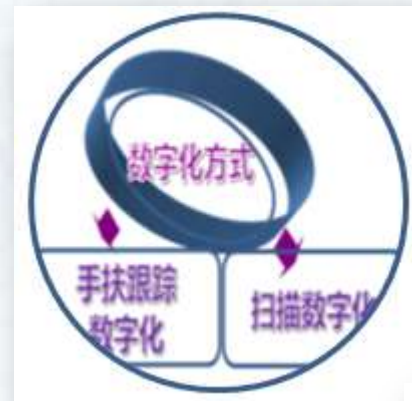
$$176 / 202 \quad 26 / 202$$

$$176 / 197 \quad 21 / 197$$

Data	农田	建筑用地	草地	林地	总计
农田	176	21	0	0	197
建筑用地	26	127	0	0	153
草地	0	0	277	0	277
林地	0	0	0	129	129
总计	202	148	277	129	756

地理空间数据误差的来源与分类

数字化中的误差



图数转换处理中的误差

误差类型

- 数字化仪
 - 数字化方式
 - 数字化操作人员
 - 数字化要素对象
- 分辨率
- 技能和经验
- 纸张变形

误差原因



采样点的方式和密度

定点误差

重复误差



地理空间数据误差的来源与分类

数字化过程中的数据质量控制

图数转换处理中的误差

数据
预处理工作

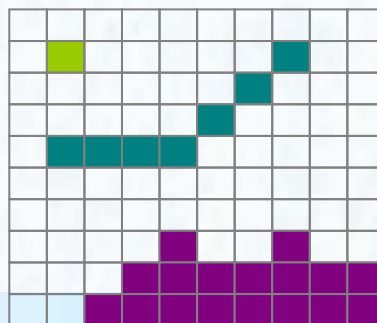
- 原始地图、表格的整理和重新清绘。



Shape #	NAME	Shape_Leng	RAIL_4M_LC	RAIL_4M_1
Polyline	津浦线	229655.720322	47	1729
Polyline	石德线	56629.378996	48	1804
Polyline	津浦线	116262.523626	49	1688
Polyline	津浦线	178849.416924	50	1650
Polyline	胶济线	323842.874597	51	1693
Polyline	胶济线	46422.912626	52	1826
Polyline	津浦线	3374.037801	62	1471
Polyline	沿海线	217194.173296	63	1472
Polyline	津浦线	162904.891196	64	1502
Polyline	津浦线	185174.209275	69	1501

数字化设备的
选用

- 手扶数字化仪、扫描仪的分辨率挑选。



地理空间数据误差的来源与分类

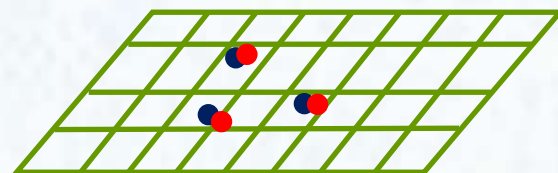
数字化过程中的数据质量控制

图数转换处理中的误差

数字化 对点精度



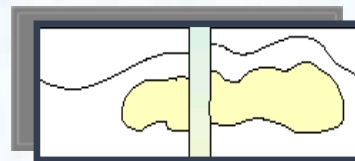
- 数字化时数据采集点与原始点重合的程度。



数字化 限差



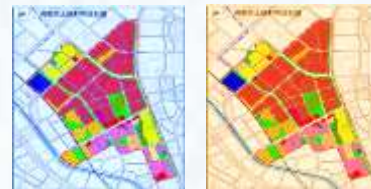
- 采点精度、采点密度、接边误差、接合距离.....



数据的 精度检查



- 输出图与原始图之间的点位误差



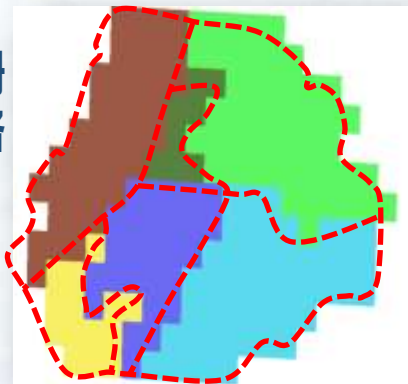
地理空间数据误差的来源与分类

数据转换的误差

● 数据结构转换



矢量 ↔ 栅格



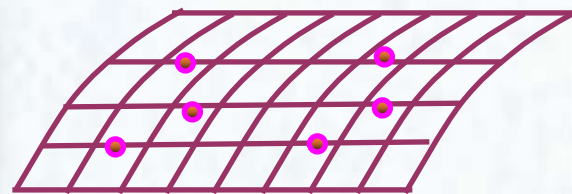
● 数据格式转换



Shp ↔ CAD



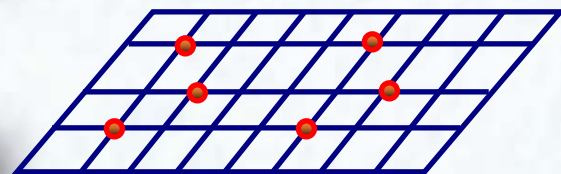
● 数据计算变换



两投影间的变换



$$\begin{aligned} X &= f1(x,y) \\ Y &= f2(x,y) \end{aligned}$$

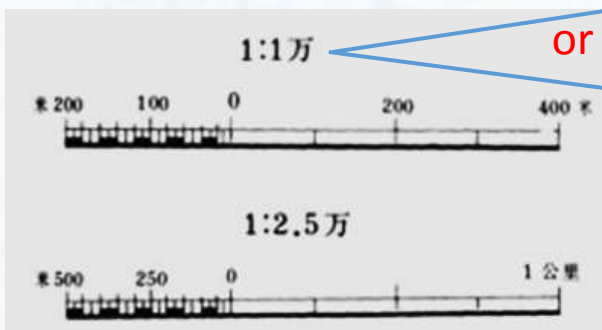


地理空间数据误差的来源与分类

应用分析产生的误差

数据使用中的误差来源

- 对数据的解释过程



1cm:10000cm

1cm:10000m

1:1万 > or < 1:2.5万

比例尺理解不正确

- 缺少文档



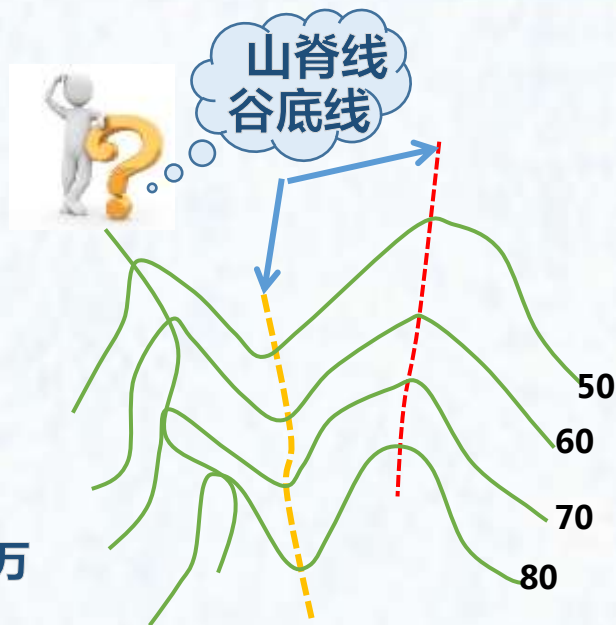
地理坐标系

编码

精度说明

⋮

=



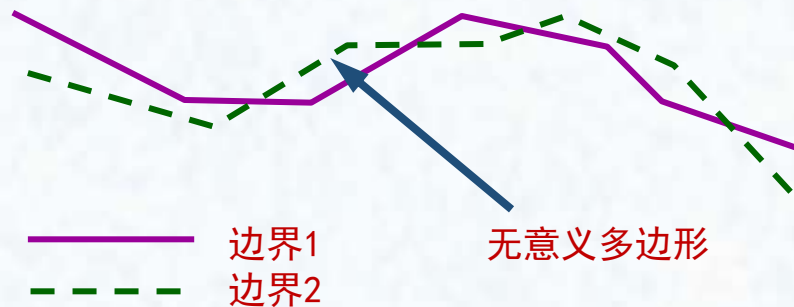
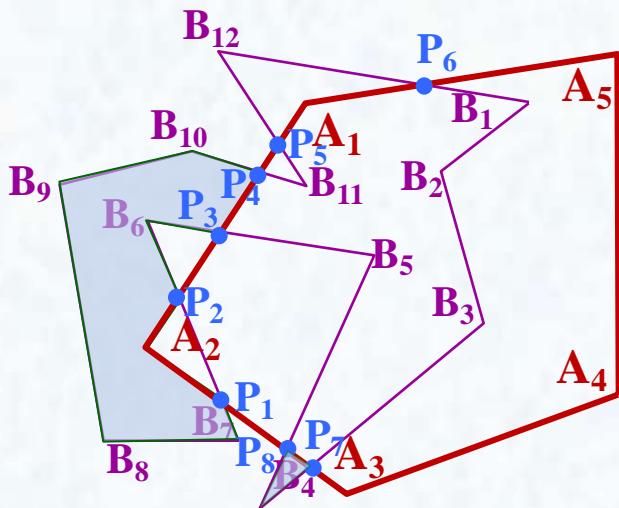
地图符号理解不正确

地理空间数据误差的来源与分类

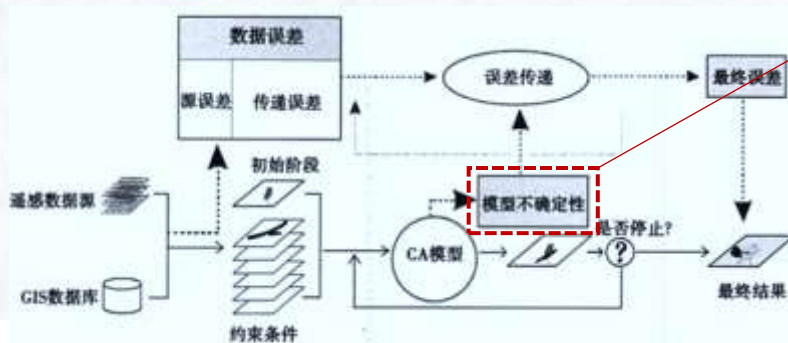
应用分析的误差来源

应用分析产生的误差

● 数据层叠加时的冗余多边形



● 由应用模型引进的误差



CA模型的不确定性
引进的误差

地理空间数据误差的来源与分类

误差控制

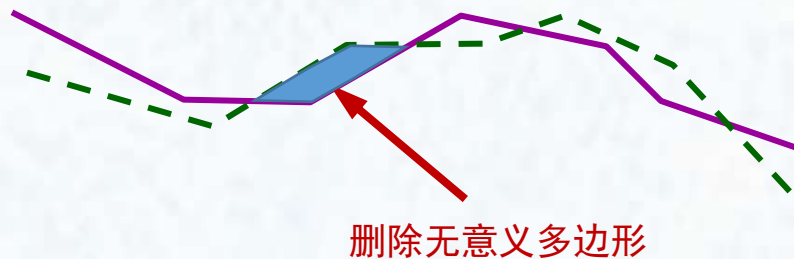
应用分析产生的误差

● 地图符号的解释

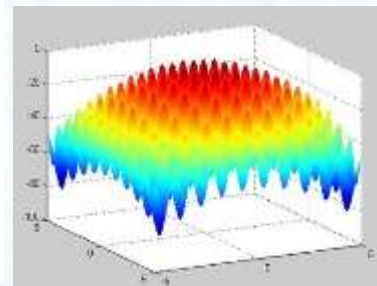
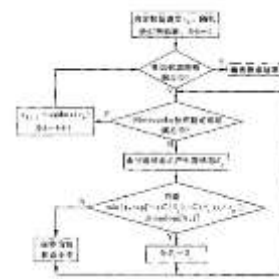


对地图添加图例和说明文档

● 删除叠加后的冗余多边形



● 采用多种模型对比验证



模拟退火和粒子群算法



谢谢大家！

