

# 第一章 导 论

## 1.1 概 述

本书将讨论空间数据建模与处理中的位置和属性以及两者结合时的不确定性问题。该问题是地理信息系统（以下简称 GIS）与遥感集成的关键之一。

本章的 1.2 节首先介绍 GIS、RS 及其集成技术。为了探讨集成技术中的不确定性特征，在 1.3 节中分析与总结了其中的误差源。1.4 节回顾了在这方面的有关研究进展。根据对现有问题及研究现状的分析，我们提出了本研究的目标及研究问题。这些将在 1.5 节中介绍。1.6 节给出了本书的结构，使读者在阅读本书时有一个比较清晰的指引。

## 1.2 地理信息系统、遥感及其集成技术

本节将简要介绍 GIS、遥感及其集成技术 其中包括其定义及有关的应用。

### 1.2.1 地理信息系统

GIS 是在信息系统中的一个特例（de Man, 1988; Carter,

1989)。信息是关于特征的符号表示数据的解译 ( Begon, 1990)。信息系统有若干种属性 ( de Man, 1988; Carter, 1989) 系统中的信息必须有效地组织, 使其在使用时可被方便地检索到; 系统必须支持对信息的查询; 系统必须被不断地更新与维护。根据所承担的任务不同, 信息系统可以分为两大类: 信息处理系统 (例如航空公司售票系统) 以及决策支持系统 (例如市场决策分析系统)。信息处理系统主要是实现对数据、信息的有关操作, 而决策支持系统主要是进行操作, 尤其是为决策支持者提供分析与建模有关操作 (Maguire, 1991)。

GIS与其他信息系统有许多共同之处, 而两者之间的主要差别在于对空间实体及其关系的描述与处理, 以及更进一步地进行空间分析与建模操作。空间查询与叠加分析是 GIS 主要功能特征之一。

对于 GIS 的定义, 有三种观点, 即地图、数据库与空间分析 (Maguire, 1991) 持地图观点的定义侧重于制图有关的内容, 因此 GIS 被视为一个地图分析与处理系统 (McHarg, 1969; Berry, 1987; Tomlin, 1990, 1991)。持数据库观点的 GIS 定义侧重于数据库的设计与实现的完美性 (Frank, 1988)。一个复杂的数据库管理系统被视为 GIS 不可分割的一部分。持空间分析观点的定义侧重于分析与建模, GIS 被视为是一门空间信息科学而不是一门技术 (Goodchild, 1990; Openshaw, 1991)。

笔者认为, GIS 是一个有关空间信息的决策支持系统, 它可以为决策者提供有关的空间信息及其质量指标。该系统应具有以下功能: 即数据获取、空间数据分析、空间信息推导、结果表示、导出结果不确定性的描述指标及其对最终决策的影响。物理上, 一个 GIS 是由空间及非空间数据、分析模型, 计算机软件、硬件及有关人员组成。

GIS的发展源于对土地属性信息与几何表达的集成以实现空

间分析。GIS 的应用可以被划分为三个阶段：①调查；②分析；管理。大多数 GIS 最初建立的原因是对某一感兴趣的特征的数据采集、组织以及实施调查，例如实现对森林类型、土壤、管线网络、学校等的调查。

在 GIS 应用的第二个阶段，根据用户不同的需求，需要实施更加复杂的分析操作。在这些操作中，对多层数据的检索、统计及空间分析技术将会被用到。GIS 的应用，如分析适用某一类型的用地地址，又如监测土地覆盖类型等，均属于 GIS 的第二阶段应用。

第三阶段的 GIS 应用，即最复杂、最深入的应用，是从一个处理系统发展到一个空间决策支持信息系统。在这个管理阶段，GIS 用于支持决策者的决策行为。该应用主要侧重于复杂的空间分析与建模。例如，决策在若干所医院中哪一所应当关闭；怎样安排某一地区的用地类型等，都属于 GIS 的第三阶段应用。在该阶段应用中，可能涉及对于现实世界的某种建模。这些可以是解析的、模拟的、逻辑的以及数学的（随机或确定的）模型。模型中可能含有近似与条件限制，而所使用的数据亦含有误差。基于 GIS 的决策支持，通常是通过对一组数据在相应模型中的运行来实现的。因此，提供给决策的选择通常是含有不确定性的。如果忽略这些不确定性，往往会导致错误的决策，有时是后果严重的。因此，对于 GIS 用户来讲，理解数据、模型及其最终结果中的不确定性是极为重要的。因此，对该不确定性的研究是当前 GIS 理论工作者的主要研究方向之一。

### 1.2.2 遥 感

遥感是一种不通过直接接触目标物而获取其信息的一种新型探测技术。“遥感”一词通常是指获取和处理地球表面的信息，尤其是自然资源与人文环境方面的信息，其最后反映在主要通过使用飞机或卫星获取的像片或数字影像（Colwell, 1983）上。遥感影

像通常需要进一步处理方可使用，用于该目的技术称之为图像处理。图像处理包括各种可以对像片或数字影像进行处理的操作，这些包括影像压缩、影像存贮、影像增强、处理、量化、空间滤波以及影像模式识别等 (Colwell, 1983)。当然图像处理还包括其他更加丰富的内容。

以飞机为载体获得的遥感数据的应用日趋广泛，因为不同学科的专业人员不断地发现航空遥感不同数据在各领域内的潜在应用可能性。由于遥感的多学科性，使得其成为环境研究中极有价值的工具。遥感数据已经被广泛应用于许多地学有关的学科领域和社会工程建设中。

### 1.2.3 遥感、图像处理及地理信息系统的集成技术

遥感和图像处理技术被用于获取和处理地球表面有关的信息。GIS 的发展则源于对土地属性信息与相应的几何表达的集成及空间分析的需求。这两项技术在过去是相互独立发展的，尽管实际上它们是互补的。对 GIS 而言，遥感是一种数据获取的手段，另一方面 GIS 中的数据可以作为遥感影像分析的一种辅助数据。因此，对这种数据获取与分析技术的集成变得愈来愈重要。在过去的几年中，许多研究人员在这方面进行了有益的探索。

在该集成过程中，GIS 主要用于数据处理、操作和分析；而遥感则作为一种数据获取、维护与更新 GIS 中的数据的手段。另一方面，GIS 可以用于基于知识的遥感影像分析。

Ehlers 等人 (1989) 分析了遥感与 GIS 集成上的主要困难。它涉及到技术方面和社会方面的考虑。在技术上，主要困难在于数据结构上的差别。遥感数据是栅格格式的，而 GIS 的几何数据通常是矢量格式的。尽管人们已经发展了许多从一种格式向另一种格式转换的算法，但是格式上的差别仍是有效集成及应用上的主要局限所在。这一问题仍需在技术上进一步发展。GIS 与遥感的

集成可以分为三个阶段，它们是：(1) 分离但相等；(2) 无缝集成；(3) 完全集成。集成的程度取决于对 GIS 和遥感影像数据融合的能力。

#### 1.2.3.1 地理信息系统与遥感集成中的主要研究问题

地理信息与分析（美国）国家中心（NCGIA），受美国国家科学基金会（NSF）资助，于 1988 年成立。该中心成立的目的在于解决 GIS 进一步应用和地理分析中的问题。12 个研究主题被提出来进一步研究，其中第十二专题是关于遥感与 GIS 集成中的有关问题。该研究小组确定了对若干问题研究之先后次序，其目标在于提高遥感与 GIS 的集成技术（Star et al., 1991）。第十二专题研究小组提出了集成中 6 个主要研究问题：

- 误差分析；
- 数据结构与检索；
- 数据处理流程与方法；
- 人-机交互界面；
- 硬件环境 以及
- 社会因素。

误差是遥感应用中存在的主要问题之一，因此对可能产生的误差源，及其相应的误差影响的大小之分析即成为一个重要的问题。由于遥感可以并且已经为 GIS 提供了大量数据，因此 GIS 用户十分关心在使用这些遥感数据时可能产生的误差，包括其类型及其大小（Star et al., 1991）。NCGIA 的第一研究专题就把 GIS 的精度问题定为最优先研究的主题；而第十二研究专题小组又把误差问题列为 6 个研究问题中的第一位，这充分显示现阶段 GIS 与遥感研究领域，不确定性问题是最为重要的研究主题之一。

在决策支持中不确定性起着两个方面的作用。首先，为避免错误地使用信息而带来决策上的失误，需要理解 GIS 所提供信息

中不确定性的量值大小及其特征；其次，从信息论的观点来看，不确定性是信息的一种量度。因此，本书的研究主题定为遥感与 GIS 及其集成中的不确定性问题。

### 1.2.3.2 GIS与遥感集成中的主要研究问题

NCGIA 的第十二研究专题小组在 1991 年提出了遥感与 GIS 集成中的 6 个主要研究问题 (Lunetta et al. 1991)。

#### (1) 评估与提出误差报告及历程 (Lineage 文件)

在 GIS 和遥感领域中，常常提到以下需求：①评估分类的误差 (Veregin, 1989)，并且提出空间数据转换标准 SDTS (NCDGNS, 1988)；进一步细化与扩展所提出的 SDTS 以满足误差评估报告的目标；③提出新的误差评估过程以满足新的需求；④对已有的 SDTS 提出新的改进与增补。误差报告标准化的最终目的在于对用遥感数据产生的 GIS 产品的应用提出适当的评估方法，进一步方便各种结果之比较。

#### (2) 改进现有遥感误差评估程序

现时所使用遥感中误差评估的方法是基于统计学中有关的方法，而这些方法并非对空间数据而设计的。这些方法在对于区域或全局范围的使用上并不经济，因为现有方法仅仅可以提供总体分类的误差，并未指出误差的空间分布，因此，对综合分类中的误差结构需要进一步的研究，比如说分类误差与多边形边界的关系等等。

#### (3) 实地检查数据获取程序

如何获取高质量的外业数据用于地图精度检查，这方面的原则涉及一般性指引并未详尽地提出过。有关各种形式的外业检测数据精度需要进行进一步的研究。

#### (4) 栅格到矢量以及矢量到栅格的数据转换

栅格到矢量转换中的一个主要问题是分类后的影像中有大量

的多边形，若在此时将栅格数据直接转换到矢量数据将会产生许多问题。只有当这种转换所产生的问题得以充分理解并可以定量其误差大小时，这些技术方可被适当地使用。

#### (5) 位置误差的特点

遥感数据的位置误差特性以及位置与分类误差之间的关系都需要有更多的信息加以描述，需要有更多的关于各种不同遥感平台特性的有关数据 以及应用 GPS 来提高遥感数据位置精度。此外，需要进一步地探讨数字高程的比例尺精度地理配准处理与地形变形的关系。另外，大地控制对地理配准精度亦有较大的影响。

#### (6) 发展标准化的位置及属性可靠性图示

最终的地图或统计结果应附有对数据精度及可靠性描述的信息，这就需要开展关于几何和属性方面的可靠性之可视化的研究，以此来描述遥感和 GIS 最终产品的精度。

在遥感与 GIS 集成的 6 个研究问题中 (Lunetta et al., 1991) 其中有 3 个直接与误差的空间分布有关。而这些不确定性则源于位置与属性不确定性的结合：①“对评估集成遥感分类结果误差的空间结构技术需进一步研究，如误差与多边形边缘的关系”；②对于遥感数据的位置误差特性及其与分类误差的关系需提供进一步的信息”；③需对遥感及 GIS 最终产品提供几何与属性可靠性显示” (Lunetta et al., 1991)。因此，在遥感与 GIS 集成中误差的研究范畴内，位置与属性不确定性的空间分布及其集成是关键问题。本研究范围进一步集中于这 2 个问题内。以上 与 研究方向可以进一步展开为以下 4 个研究问题：

#### (1) 位置数据不确定性的空间分布

点的误差分布在测绘界已进行了长期的研究，然而对于线段、线及面状元素的不确定性尚未严格地从理论上系统地研究过。

#### (2) 遥感分类数据属性不确定的空间分布

通常用于描述遥感分类结果的不确定方法是误差矩阵。但是

这种方法不能描述出不确定性的空间分布，而概率矢量方法则具有描述不确定的空间分布的潜在可能。

### (3) 位置与属性数据结合过程中不确定性的传递

现有的技术中，并没有直接可以用于处理集成位置与属性不确定的方法。但是，若两种不同的不确定性可以转化为用同一种指标描述，则它们可以被结合。这意味着，两种不确定性应在同一理论基础下用统一的方法描述。

### (4) 位置与属性不确定性的关系

Goodchild 等人(1992)曾经探讨过这个问题。他们使用两个描述指标，称之为包含 (inclusion) 和自相关 (auto-correlation) 来描述误差分布模式。因此，本书将不再进一步探索这个问题。

从对现实世界进行抽象表达的角度来看，GIS 和遥感数据是对自然现象不同水平的抽象。因此，集成这两类数据对于空间决策支持将十分有益。这也是 GIS 发展的一个趋势。在决策支持的过程中，需要考虑以下几个问题：首先是地图数据的不确定性，这是由于量测或进一步处理过程中引入的位置误差，这些几何数据当中也有可能含有属性误差；其次，分类遥感数据的误差，主要是由于空间、光谱和时域特性的不确定性而造成的属性误差，地理配准时也会造成位置上的不确定性；第三种不确定性是在结合以上两种类型的数据时引入的不确定性。由于不可能同时处理几个方面的不确定性，本书侧重讨论以下问题：①矢量数据的位置不确定性；②栅格数据的属性不确定性；③结合这两者数据引入的不确定性。

地理信息在 GIS 中可以用以下两种方法之一来表达：目标 (object) 方法或者是域 (field) 方法 (Ehlers et al., 1989; Goodchild, 1989; 1992)。在目标方法中，GIS 用目标来表示自然界，这可能包含有点、线以及面状目标，其可以被位置、属性或拓扑关系特性描述。使用这种方法时，异质专题数据不确定性的空间分布世界区

域内的渐变特性不容易表达出来。但是，这种方法比较易于表达目标的位置不确定性。例如，GIS 中的空间目标可以是电线杆、道路及地籍宗图，其相应的属性可以是电线杆使用了的时间、道路的宽度及宗地的归属、拓扑关系可以是相邻宗地的关系等等。

在使用基于域的方法中，GIS 在描述现实世界时是通过每一个像素赋予属性来实现的，而不是通过对目标的抽象或描述它们之间的拓扑关系来实现的，例如来自于遥感影像的土地覆盖分类数据。利用这种方法有可能描述异质数据不确定性的空间分布以及渐变区域的不确定性。与这两种方法相对应的数据通常称之为矢量或栅格数据。在表达不确定的空间分布方面时，这两种方法是互补的。这是 GIS 矢量数据与遥感栅格数据集成的另一个优点。

大多数现有的 GIS 设计是基于地理目标没有误差的假设，因此只可以处理确定的目标。用一个处理“确定”数据的 GIS 来处理带有不确定性的数据将会出现问题。因此，发展具有处理不确定性数据的功能的 GIS 也成为 GIS 发展的一个趋势。设计这种 GIS 的主要问题有：①描述一个带有不确定性的物体；②描述物体之间的不确定性关系。在现阶段，GIS 与遥感集成中的误差源以及 GIS 操作过程中的不确定性是两个主要研究问题。

### 1.3 地理信息系统与遥感中的误差源

“尽管利用遥感方法在为空间数据库获取数据方面发展得十分迅速，但我们对于数据处理过程中误差的理解，尤其是多种空间数据的集成中的误差了解甚少” (Lunetta et al., 1991)。以下，首先讨论遥感与 GIS 中的误差源。

### 1.3.1 地理信息系统中的误差源

GIS 中的误差是指 GIS 中描述的物体与其在现实世界中的真实情况的差别。这些误差的类别包括随机误差、系统误差以及粗差。数据是通过现实世界中的实体进行解译、量测、数据输入、空间数据处理以及数据表示而完成的。其中每一个过程均有可能产生误差，从而导致相当数量的误差积累。图 1-1 表示出 GIS 中数据的误差源。

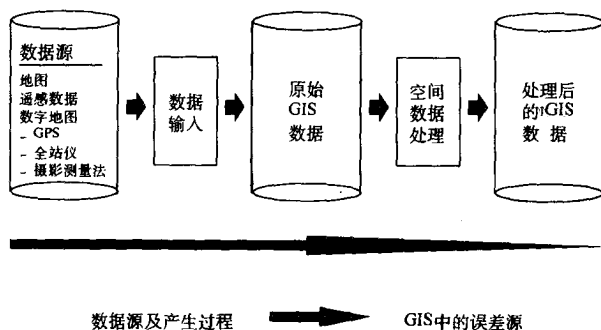


图 1-1 GIS 中数据的误差源

GIS 的数据源可以是一张纸质地图，一张分类的遥感影像来自航测的数字数据以及 GPS、全站测距仪数据。数据可以通过解译或者量测而获得。GIS 中的数据通常是这些数据及其相应的误差的传递结果。进一步的处理将会引入新的误差。与直接从地面采集的数据（如从 GPS 获取的数据）相比，从地图或影像中获取的数据将含有更多的误差。纸质地图的误差包括：目标解译、目标量测、地理配准、地图编辑（综合）、印刷以及图纸变形。在这些误差源中，目标解译是一个主要来源之一。例如，最常用的地质图、土地利用图或土壤图是利用不同的专题模型表示其属性，即具有相

同属性的区域用边缘线包围起来。然而，在许多情况下，这种模型不能准确地对现实世界给予准确的描述。在现实世界中，许多物体不具有明确的边界，而是一个渐变的过程。例如，土壤边界就是这种情形；而且在边界的内部亦不是同质的，而是异质的（Burrough, 1986, 1993; Goodchild, 1989）。因此，用确切边界及均质多边形模型的表示在许多情况下会导致误差，这种误差主要是属性误差，同时与边界的位置误差相关。

在一个指定的数据采集过程中，量测误差应局限于某一限差内，例如对某一地图产品，其限差在地图比例尺的 0.2 mm 内。为确定点的位置需要  $x$ 、 $y$ 、 $z$  坐标的量测值，而且需要相应的属性（例如道路宽度）。对于量测点而言，其误差本质的有关研究在测量学中已经发展得比较成熟（Mikhail and Ackermann, 1976）。大多数的量测误差具有随机特性。

源于数据内插的误差有时是十分显著的，尤其是那些初始值是通过手工量测系统获得最少量的观测值。Tempfli (1980) 指出，在建立数字地面模型时，内插误差较量测误差要大得多。

Burrough (1986) 曾较系统地分析了 GIS 中的误差。他将误差分为：①明显误差 ②源于自然或原始量测值的误差；③源于数据处理的误差。在这三组误差中，第一组误差是十分明显而易于查出的，第二组误差相对比较复杂而在与数据接触时可以检查出；第三组误差是在执行某一处理时才会出现。这种误差最难于量测，因为误差不仅与所使用的数据有关，而且与所使用的算法、数据结构有关（Burrough, 1986）。

#### 1.3.1.1 明显误差

##### (1) 数据年龄

大多数 GIS 使用已有的数据，而这种数据往往是过时的。遥感技术可以提供最新的数据，但它不能取代决策支持时所需的全

部 GIS 数据。Mead(1982)指出：“除地质数据外，数据的可靠性随其年龄递增而递减”。因此，若使用的数据愈旧，出现误差的可能性就愈大。

## (2) 地图比例尺

土地覆盖可以表示在不同比例尺的地图上或以数字的形式表示于地图上。地图比例尺越大，表示的内容越详尽，一般来讲，精度也越高。但是在 GIS 存贮时，比例尺越大，占据的存贮空间也越大。因此，地图比例尺的选择取决于 GIS 的应用。

## (3) 观测值的密度及其分布模式

一幅专题图(如土壤图)是对一系列观测值的内插得来的。有两个因素可影响其精度，即观测值的密度及其分布模式。

### 1.3.1.2 源于自然变量或原始量测值

#### (1) 位置精度

地理数据的位置精度大多数与相应的数据类型有关。地形特征，如房角、街道中心线等，通常是以高精度量测所得的。另一种类型，如森林或土壤类型的边界，则可能具有相当低的精度。这里涉及到两类误差，即线的识别误差和量测误差。其中，前者的误差更为显著。在这种情况下，线的位置误差可能会相当大。

#### (2) 属性误差

属性数据有两类 即定量数据 如 pH 值 以及定性数据 如某一宗地的用地类别为工业用地)。相应地，属性数据的精度也有两类，即定量数据精度和定性数据精度。

#### (3) 数据偏差

数据输入时的错误可能会导致损坏原来高精度的数据。另一类偏差是观测偏差。除随机误差之外，每一个观测员都会用一定的不同的方式进行观测。例如，在数字化一条线时，一个操作员可能会习惯地偏向被数字化线的左侧，而另一个操作员可能会偏右。

### 1.3.1.3 处理过程中产生的误差

#### (1) 计算机中的数字误差

数字在计算机内是用有限的位来表示的。在计算时可能会产生严重的错误，尤其是使用两组大值数据和数组运算时，而使得精度不够。舍位误差是一种明显误差。

#### (2) 拓扑分析误差

在地形图叠加并且使用特定的算法时会产生误差。这可能会出现在 GIS 处理中，如地图叠加、边界交叉、矢量地图的栅格化、两层或多层网络系统叠加等等。这些可导致拓扑关系上的误差。

#### (3) 地图数字化

地图数字化是数据获取的一种常用方法。在手工数字化中有两种方式可供选择：点式或串式数字化。数字化的误差特性取决于所选择的数字化方式、操作员及数字化设备。数字化误差可以被定义为数字化的线对原地图上的线的偏差。Bolstad 等 (1990) 研究了手工数字化的误差，他们发现手工数字化所产生的误差一般比较小，一般偏差为 0.054 mm，最大观测距离偏差为 0.261 mm。数字化误差也曾被其他研究人员探讨过 (Maffini et al. 1989; Keefer et al., 1991)。

### 1.3.2 遥感中的误差源

在本节中，我们将讨论遥感作为 GIS 的数据输入手段中的误差源。其误差可以被分为以下几个部分：(1) 数据获取误差；(2) 数据处理误差；(3) 数据分析误差；(4) 数据转换误差；(5) 人工解译误差。Lunetta 等 (1991) 曾对误差源问题进行了探讨。

#### 1.3.2.1 数据获取误差

数据可以在实地或用遥感的方法获得。数据获取过程中，有

些误差源是难以控制的，诸如大气条件和地面景观的自然变化。其它数据采集误差，如几何或辐射误差，则可以控制或改正。关于数据获取误差源的进一步阐述可见文献 Salsig(1990)，Desachy 等人(1985)或 Duggin 等人(1985)。数据获取中可控制的误差讨论如下。

### (1)几何误差

保持一个定常的影像比例尺会方便直接将其输入 GIS 中。许多因素可能造成比例尺的局部变化，例如非底点部分的影像、地形起伏变形等等。

影像几何与所使用的传感器有关。如何获得稳定的影像，在航空摄影测量中已进行了许多研究。其主要是通过竖直摄影测量实现。许多其他遥感系统由于包含了连续影像生成装置，因此更易产生几何变形而影响其与 GIS 的集成 Lunetta et al.,1991)。

多光谱扫描(MSS)系统在工作时连续地获取影像，所以传感平台在获取数据时的所有运动都会影响构像几何。在一条扫描带上，每一个像素对应于一个不同的视角。因此，像素的大小在一条成像带中是视角的余弦函数。带有 CCD 传感器的构像系统消除了 MSS 中的多数几何误差 Slama,1980)。

### (2)平台

传感器平台的飞行或轨道高度、传感器的视场及方向影响着成像几何。因此，如果需对遥感数据进行精确地纠正而输入 GIS 中，就必须已知飞行器的瞬时高度与方位，以及传感器的方向。全球定位系统 GPS 技术可以为  $x$ 、 $y$ 、 $z$  位置量测提供良好的基础。对于高频运动的传感器平台的几何改正可能会要求一个基于每一个像素的复杂的点模型(Gibson, 1984; Gibson et al.,1987;Reimer et al.,1987,Rickman et al.,1989;Till,1987)。

### 1.3.2.2 数据预处理误差

#### (1) 地面控制

纠正过的遥感数据或最终地图产品的精度不会高于用于纠正的地面控制点的精度。基于地面控制点的摄影测量纠正技术已经发展得十分成熟。

#### (2) 几何改正

几何改正将会产生位置或属性上的误差。属性误差可源于重采样。位置误差来源于纠正模型的近似性以及用于解求纠正参数的控制点的误差。递归或离散技术，如有限元规划，经常被用来解算飞行器不稳定性而造成的复杂变形。一个微分纠正，如使用 DEM，较之基于少量地面控制点的多项式纠正会产生较小的误差。Welch 等(1985)用一级多项式纠正法对 Landsat-4 数字影像纠正，其结果的几何精度非常好。

#### (3) 影像增强

空间数据的影像处理经常涉及影像变换。常常有可能出现这种情况，即重采样后的几何与辐射特性与原始数据具有相当差的关系。例如，滤波技术常常用于消除分类中的异质“噪声”。这里必须考虑由影像增强而产生的误差，以及在传入 GIS 之前量化这些影响的程度。

#### (4) 分类

分类本身可能产生相当显著的误差。潜在的误差源包含：遥感数据中的随机误差；类别定义时的不明确性；对混合类及变化边界处分类系统的能力较差；不同分类系统的一致性较差。

这里“分类”一词有两方面的涵义 第一个是遥感分类 即将某一像素赋予某一类别值；其二是分类系统，即将一组具有相同或相似性质的物体归为一个类别。

遥感数据表示对具有一定空间、时间及属性特征物体的抽象。

信息从地面到传感器的传递可能导致新的随机误差。分类之前的预处理过程也会引入误差。因此，用于分类的遥感数据含有随机或其他类型的误差。

类别定义的模糊是引起分类误差的另一个常见的误差源。经常有这样的可能，即某一给定的特征不能完全地落入某一定义类别内，由此而导致不一致性误差。因此，应寻求更好的类别定义以及分类体系。

一个遥感影像上的像素可能被分到一个错误的类别中。这是由于在分类决策时缺少有关的证据。在分类中引入其他类型的辅助数据如 GIS 数据，是提高分类精度的一个途径。另一引入误差的可能是分类时综合不同种证据的方法。常用的分类方法，如最大似然分类法，是基于概率论的。另一个可用的理论是 Evidence 理论 (Shafer, 1976)，该理论可视为概率论的一个扩展。

在遥感数据分类时总是要参照一定的分类系统，例如根据国家标准把土地覆盖分成若干大类，更进一步分为若干子类、亚类等以便 GIS 中的分析。在对混合像片、变换区域或动态系统进行标识类别时，由于所依据的分类系统的不同可能会导致对某一像素给予不同的类别相识。这种误差常常难以量化。在 GIS 空间分析中，常常需要两层或以上的数据。各层数据若所依赖的分类系统可能不同，常常会导致严重的后果，甚至致使某些类别不能进行叠加分析。

#### 1.3.2.3 数据转换误差

在数据转换的过程中，不论是矢量到栅格或是从栅格到矢量的转换，均可能产生严重的误差。在许多情况下，一张矢量地图需要转换成栅格格式，这是因为有许多 GIS 的空间叠加分析在栅格环境中更易实现。在栅格化的过程中有两类主要的误差源：一类较为明显，在栅格格式中，每一像素只含有一种类别。而当在

相应的矢量格式中一条线从一个栅格的中间穿过，理论上讲在栅格格式中该栅格内含有两种不同的属性。由于栅格格式的局限性，该栅格内只能近似地用一种属性表达出来。因此造成转换以后的误差。另一误差源是拓扑上的不匹配。栅格化的误差大小与诸多因素有关，例如栅格的大小、栅格表示方法、转换算法等。关于转换误差的讨论见 Burrough(1986)；Dorst 和 Smeulders(1987)；Bregt 等(1991)以及 van der Knaap(1992 等人的讨论。

#### 1.3.2.4 人工解译误差

在数据分析与解译过程中，人为造成的误差是最难量化的。人为误差可能出现在遥感及 GIS 数据的使用过程中，因为该误差是主观的，而难以量化。这种误差与人对遥感影像的理解能力以及所需的从遥感影像中抽象水平有关。人工解译误差可以归属为分析误差；由于主要因素是人，因此这种误差又与前述的分析误差有较大的差别。

到此为止 我们讨论了 GIS 和遥感中可能的误差。篇幅所限，本书不可能详细地讨论所有这些误差的特性。在 1.5.1 节中将阐述本书所涉及的误差范围。

## 1.4 有关研究的简要回顾

在过去几年里，有许多关于空间数据处理研究进展的报道。其中最多的是与误差传播有关的主题。Newcomer、Szajgin(1984)和 Veregin(1989)讨论了在 GIS 叠加时布尔运算的误差传播；Lodwick(1989)，Lodwick 和 Monson(1988)和 Heuvelink 等(1989)分析了加权叠加与输入层及权重不确定性的灵敏性关系。Arbia 和 Haining(1990)提出了一个一般性的栅格模型。而叠加中的不确定性及从一个数据模型到另一个模型转换时产生的不确定性需进一步地研

究。

位置不确定性的有关研究已经进行了许多。Perkal (1956, 1966)提出了 Epsilon 带模型。Chrisman(1982 )对该模型又进一步发展并用于量测地图的不确定性。Epsilon 带是沿着量测线的两侧平行等距的一个条带。Honeycutt(1986)讨论了基于制图综合的 Epsilon 带 以及 Epsilon 带内的概率分布。Dutton(1992 )用 Monte Carlo模拟法模拟了线段不确定性的分布。点位置误差已经经过了长期的讨论,而面状元素之不确定性的空间分布尚未进行讨论。其中的关键问题是线段的不确定性模型(它是面状目标的基本要素)该内容将在第三章中进一步讨论。

人们对若干种关于空间定性数据的误差传播方法进行过探讨(Newcomer and Szajgin, 1984; Veregin, 1989; Lodwick et al., 1990; Van der Wel, 1991; Lanter and Veregin, 1992)。Goodchild(1992)讨论了关于相关类别空间数据的误差理论。Bierkens 和 Burrough(1993)从非参数地理统计的观点讨论了类别空间数据误差模型。

尽管在属性及位置数据的不确定性方面已经进行了较多的探讨,然而这两者的结合的误差模型还讨论甚少。因此,本书将进一步讨论结合两种不确定性的误差模型。

## 1.5 本研究的范畴与目标

### 1.5.1 本研究涉及的不确定性

#### 1.5.1.1 GIS 中的误差

GIS 中点的误差来源于其原始数据源中的误差(如纸质地图)以及进一步积累的数据输入、表示、空间分析中的误差。在本研究中,将不进一步讨论构成这些误差的原因,而只讨论它们之间的传递。在 GIS 中,只讨论位置误差,因为它被认为是最主要的误差。

假设所有输入 GIS 中的数据都经过认真的检查，因而在数据中没有粗差。GIS 中数据可能会受到若干种系统误差的影响。这里假设较大的系统误差已经经过改正。因此，本研究主要针对 GIS 中点位置随机误差。根据点的误差，进一步讨论线段、线特征及面状物体的不确定性。

#### 1.5.1.2 遥感中的不确定性

这里只考虑遥感数据中的属性不确定性并且侧重于分类误差的讨论，而该误差是由于分类模型本身造成的。

### 1.5.2 本研究的目标与范畴

NCDCDS(1988)曾提出过数字制图数据中的五种数据质量指标。它们是：历程 Lineage)、位置精度、属性精度、逻辑一致性以及完整性。而时域不确定性应是要考虑的一个方面。我们认为，有三种不确定性是最基本的：位置、属性及逻辑上的不一致性。另一方面，从某种程度上分析，逻辑不一致性可以看成是位置、属性或两者之不确定性所造成的。例如，将房屋放在街道之中心是一个逻辑不一致性错误。该错误可能是由于数字化时所造成的位置误差所导致。逻辑不一致误差也可出现于分类当中，如一球场的内部的某一处被误分为建筑用地。本研究将讨论 GIS 和遥感中的两类最基本误差：位置和属性误差。

根据 1.2.3.2 节中对 NCGIA 所研究专题的理解，1.4 节中的分析以及以上讨论，本研究的主题确定为：研究 GIS、遥感及其集成中的位置、属性之不确定性。在该主题下，进一步确定本研究的子目标为：

- 研究 GIS 中面状物体位置不确定性的空间分布模型；
- 研究分类的遥感数据属性不确定性的空间分布模型；
- 研究遥感与 GIS 数据集成时属性与位置不确定性的结合模

型。

(1) GIS 中的面状物体是空间分析的基本单位。该物体位置不确定性的空间分布可以用某一点属于该面状物体的概率来表达。

Blakmore(1984)曾对该问题探讨过。根据他的结论,一个点与一个多边形的关系可以是以下五种之一:一定在其内,一定在其外,可能在其内,可能在其外,以及不确定。该描述给出点与多边形关系的不确定的五种定性描述。但是没有定量指标描述。在空间分析与其他指标相结合时,往往需要有定量的指标描述。

在我们的问题中,面状物体的不确定性是由于其所构成的节点的位置不确定性所造成的。因此,需探讨由于构成节点的随机误差所形成的面状物体的不确定性模型。

(2)通常地,分类的遥感影像通过用正确分类像素占总像元数的百分比来描述某一类或整个影像的分类精度。但是,这种精度指标不能描述出分类不确定性的空间分布。利用这一结果,我们不可能知道某一特定象元的分类精度。例如一个这样的问题就无法回答:对于某一分为森林的区域,其中心区以及边界区域的不确定性各是多少?因此,不确定性的空间分布需要进一步的研究。根据一个具有空间不确定性分布的分类结果,可以知道哪一区域应给予较高的可靠性,哪一区域给予较低的可靠性。从而减少在使用数据时由于对其数据质量缺乏认识而造成的决策失误。这种具有不确定空间分布的遥感分类结果可以作为与具有位置不确定性的 GIS 数据结合的数据基础。

(3) GIS 与遥感集成的一个最基本的问题是集成 GIS 中(通常是矢量格式)的数据与分类的遥感(通常是栅格格式)数据。因此, GIS 与遥感集成中的误差分析的核心问题是位置误差与属性误差的结合的问题。这一核心问题的重要性,可以从 NCGIA 第十二专题研究小组提出的有关 GIS 与遥感集成中的不确定性分析问题看

出。那里，所提出的六个问题中有三个涉及位置与属性不确定性的结合及其空间分布问题。

### 1.5.3 本研究拟解决的问题

为达到以上所提出的目标，必须解决以下几个主要问题。

(1) 系统地分析不确定性的类型以及相应的理论基础，如（空间统计学、概率论、模糊数学、Evidence 数学理论。选择适当的理论作为解决本研究所提出问题的理论基础。

(2) 用于处理位置及属性不确定性模型与方法的研究现状。这些模型与方法可以应用的领域。哪些模型可以选择并进一步发展来处理本研究所提出的问题。这些问题可以通过系统地分析已有的模型与方法进行探索。

(3) 推导矢量 GIS 中一条线段的位置不确定性模型。一个面状物体几何上是由其边界线来描述的，而边界线是由若干条线段构成的。一条线段是由连接其两端的端点所构成的。因此，为了描述面状物体的不确定性，需要从点的不确定性导出线段、线性特征的不确定性。由于点的不确定性在测量界已研究得比较成熟，因此可以从线段的不确定性开始讨论面状物体的不确定性问题。

(4) 导出线状边界和面状物体的不确定性模型。根据线段的位置不确定性模型，边界线性特征及面状物体的不确定性模型可以被导出。

(5) 描述分类影像的属性不确定性并定义描述属性不确定性的误差指标。

(6) 研究结合位置与属性不确定性模型。根据导出的 GIS 中的面状物体不确定性的描述和分类遥感影像的属性不确定性，研究结合这两种不确定性的模型。

(7) 研究位置不确定性，属性不确定性及其两者结合的不确定性可视化模型。

(8) 开发一个软件原型用于实现提出的理论模型，以显示其应用的潜在可能性。

## 1.6 本书结构

本书共有六章。在本章之后，第二章主要分析与总结现有不确定模型所依据的理论，讨论每一理论的局限性及应用范围，之后将指出本研究问题所应用的理论。

第三章将分析与回顾一般误差分析所使用的模型与方法，并且进一步分析这些模型的可应用领域。基于这些分析，指出在模型方面还需就 GIS 中位置不确定性描述的进一步发展方向。用于处理属性不确定性的模型也将在此章中介绍。

第四章介绍“S-带”模型，它是用来结合位置不确定性及属性不确定性的一个模型。该章将主要解决两个主要问题：首先，需要导出线段的不确定性模型，这是描述面状物体不确定性的基础；另一个问题是发展一个用于位置及属性不确定性结合的模式。

第五章描述一个不确定性可视化的软件原型，以及基于第四章中的模式的可视化结果。

第六章总结出本书的成果以及对进一步研究的建议。

## 第二章 地理信息系统与遥感中 不确定性处理的理论概述

### 2.1 概 述

在 GIS 与遥感领域中人们已经研究了许多处理不确定性的模型与方法。然而，到目前为止尚未有完整的理论体系用于处理地学中的不确定性。若干现有的理论如概率论、证据理论、模糊数学以及空间统计可以被用来处理地学中的不确定性问题。以上的每一个理论都有其适用范围，且仅能处理 GIS 和遥感中不确定性的一部分。

在本章中，GIS 和遥感中的不确定性首先被分成若干类。然后指出相应的、可以处理每一类不确定性的理论。对每一种理论中不确定性的度量指标作了进一步阐述。最后指出本书中所涉及的不确定性类型及相应的理论基础。

### 2.2 地理信息系统与遥感中的 不确定性及有关理论

本章旨在讨论 GIS 和遥感中的不确定性类型，以及相应的可用于处理这些不确定性的理论。这里通过对一个经典问题的讨论

来展开论述。

GIS 和遥感中涉及的主要问题包含：

(1) 确定一个物体的空间位置  $O(x, y, z)$ ；

(2) 确定两空间物体  $O_1$  和  $O_2$  的关系  $R(O_1, O_2)$ ；

(3) 决定一个空间物体对一个集合的归属关系  $O \in A$  (物体  $O$  属于集合  $A$ )。

问题 1) 中的  $O(x, y, z)$  表示了一个物体的几何描述方面 需要讨论其空间位置。一个点的坐标可以通过直接量测或是通过对一系列量测数据的处理而获得。所涉及的误差有随机误差、系统误差、粗差或是其两者以上之综合影响。在一个可靠的系统内 粗差出现的机会很小。在系统误差被改正之后，一个系统内仅含有随机误差。在测量中广泛使用概率论与数理统计来处理这种误差。这类不确定性的统计描述量的本质是定量的。

在问题(2)中 需要讨论两物体  $O_1$  和  $O_2$  的空间关系。模糊数学理论的空间拓展有可能是解决该问题的一种途径。所以，模糊数学是解决空间关系不确定的一个潜在理论。

问题 3) 即  $O \in A$  相对比较复杂。在该问题中 需要讨论以下三个方面的问题：

(1) 子集  $A$  的定义可能模糊。在 GIS 中这可能意味着某一物体的自然边界的定义不清楚。这个模糊性可能在空间域、属性域、时间域、或 遥感影像的 频谱域。例如 在空间域中的“城市区域”属性域的“湿地”。

对一个模糊集合 对于全集  $O$  中的每一个元素  $O$  需给出一个隶属函数值  $\mu_A(O) \in [0, 1]$ ，以表示该元素对于子集  $A$  的隶属度。对一个一般集合而言，一个元素可以属于或者不属于这个集合。因此 隶属度函数值等于 0 或 1 即  $\mu_A \in \{0, 1\}$ 。模糊数学中的规则 主要是定性的，因为对于某一模糊集合而言其隶属度函数值  $\mu_A(O)$  是难以确定的。隶属度函数  $\mu_A(O)$  也可以用一个概率值

$P(0 \in A) \in [0, 1]$  来表示。为了确定其概率值,就需要一组观测值。在许多情况下,这些观测值是不可获得的。有时我们是取一个经验值为隶属度值,而不是用概率值。

(2) 对物体的描述可能含有误差。在一个 GIS 中,  $O$  可以是一个具有空间、属性、时域值的一个物体。例如,  $O$  可以是由边界坐标描述其位置的一宗地,其属性为:其中森林覆盖率为 75%。描述边界的坐标及其密度值均可能含有误差。在该问题中,概率论及统计可用于描述其中的随机误差及其传播。

(3) 我们有可能没有足够的证据将物体  $O$  分入子集  $A$  中。注意其中子集  $A$  可能是一个普通集合或是一个模糊集合。这个问题的本质是不能十分肯定物体  $O$  是否属于子集  $A$ 。例如,在遥感影像分类中,光谱信息可能不能提供足够的信息来将物体  $O$  分类给某一类别。在这种情况下,物体  $O$  被精确地确定。而子集  $A$  也是确定的,是一个普通集合。只是光谱信息太差而不能肯定  $O \in A$ 。这种不确定性可以用证据理论来处理。这种不确定性有时是可以估计的。例如,可以表示为似然函数或 Bayes 概率。然而,在许多情况下不是如此。因此,对应的规则本质上是定性的 (Molenaar, 1991)。

基于以上的分析可以看出,用于处理 GIS 和遥感中不确定性的主要理论有:空间统计学、概率论、模糊数学以及证据理论。

在本章的其余部分,将介绍这些理论的本质及其可以处理的不确定性。每一理论所描述不确定性的侧重面有所不同。有关这些量度的详细阐述见 Klir 和 Folger (1988) 的文章。本章最后将简单说明这些理论之间的关系。

## 2.3 概率论

概率论可用于处理由于随机误差而产生的不确定性。在概率

论中，不确定性被描述成给定某些观测值的条件下某一假设为真的条件概率。该假设的条件概率表示了一个概率从 0 到 1 区间的定量描述。Bayes 定律 Rao, 1965 如下：

$$P(H_i/X) = P(X/H_i)P(H_i)/P(X)$$

其中

$$P(X) = \sum_{i=1}^n P(X/H_i)P(H_i)$$

这里

$n$ ——假设的总数；

$P(H_i/X)$ ——在给定观测值  $X$  的条件下假设  $H_i$  为真的条件概率；

$P(H_i)$ ——假设  $H_i$  为真的条件概率；

$P(X)$ ——观测值  $X$  出现的概率。

例如在遥感分类中， $H_i$  可以是一个给定像素属于类别  $i$  的假设。 $X$  为频谱空间中一个密度矢量。不确定性是由条件概率  $P(H_i/X)$  表示出来的，它指出在划分某一个像素为某一类别时的概率。条件概率  $P(H_i/X)$  是由 Bayes 公式确定的。在第四章将详细描述该公式在遥感分类中的应用。在一个最大似然分类中， $\{P(H_i/X), i=1, 2, \dots, n\}$  中的最大值将被选定，即该像素被分为类别  $i$  如果  $P(H_i/X) > P(H_j/X)$  对于所有的  $j \neq i$ 。在最大似然分类中，每一像素的所有概率值  $\{P(H_i/X)\}$  称之为概率矢量。该矢量将被用作描述分类遥感影像的属性不确定性，有关内容在第四章将进一步阐述。

在 Bayes 公式中，假设先验概率  $[P(H_i)]$  可以正确地确定，这在许多应用中是难以达到的。该理论可以使用相当复杂的联合概率，但需要大量的概率观测值。这里我们没有一个概率值质量的描述指标，即本模型不能提供其可靠性的度量指标。

## 2.4 证据数学理论

证据数学理论 也称为 Dempster-Shafer 理论 是对传统概率论的一个扩展 (Shafer, 1976)。在证据数学理论中 假设  $H_i$  是真的概率量测值由支持某一假设的可获得的证据的概率来代替。证据理论是基于可信度 belief 函数和可能 possibility 函数所确定的一个区间, 可信函数度量已有的证据对假设支持的程度。而另一方面, 可能函数量测根据已有的证据不能否定假设的程度。可信函数表示了证据支持假设的最低程度, 而可能函数则表达了证据支持假设的最高程度。

例如 假设  $Y$  是一系列的假设并且知道每一个可能的假设为真的概率  $[P(H_i)]$ 。

$$Y = \{P(H_1), P(H_2), P(H_3), \dots, P(H_n), \textcircled{H}\}$$

这里

$P(H_i)$ ——假设  $H_i$  为真的概率;

$n$ ——总的假设的个数;

$\textcircled{H}$ ——没有讨论到的部分 见以下说明。

我们定义

$$\sum_{i=1}^n P(H_i) + \textcircled{H} = 1 \quad (\textcircled{H} \geq 0)$$

例如 在一个最大似然分类中 在某一区域内有五种可能的类别。对于某一像素, 有以下概率值  $P(H_1) = 0.4, P(H_2) = 0.2$  和  $P(H_3) = 0.1$  则没有支持的部分  $\textcircled{H}$  是 0.3。

可信函数定义如下:

$$\text{Bel}(H_i) = P(H_i)$$

可能函数定义如下:

$$\begin{aligned} \text{Pl}(H_i) &= 1 - \sum_{j=1}^n P(H_j) \quad (i \neq j) \\ &= \text{Bel}(H_i) + \textcircled{B} \end{aligned}$$

区间  $[\text{Bel}(H_i), \text{Pl}(H_i)]$  表示了由于证据不完整性对于假设  $H_i$  支持的不确定性，这是由于未支持部分  $\textcircled{B}$  所造成的。

在  $\textcircled{B} = 0$  的情况下， $\text{Bel}(H_i) = \text{Pl}(H_i) = P(H_i)$  此时证据理论等同于概率论。因此，证据理论是概率论的一个扩展。

Dempster 的结合规则是证据理论中的重要部分。该理论的一个基本策略是把证据集合分解成一系列不相关的证据集合，在这些证据集合中分别作判断 最后利用 Dempster 结合规则将这些判断结合起来。

证据理论在专家系统设计中变得越来越重要。该理论也被应用到 GIS 和遥感领域中。例如，该理论被应用到一专家系统当中，利用遥感影像更新森林图 (Goldberg et al., 1985)。Lee 等人 (1987) 将这种方法与基于 Bayes 方法的对 MSS 影像分类进行了比较。该理论还被用于在 GIS 中应用多源数据为直升机选择最佳路线 (Carvey, 1987) 证据数学理论的优点在于其为概率论的更一般性表达，Bayes 的可信函数是 Dempster 可信函数的一个子集；Bayes 的条件规则是 Dempster 结合规则的特例，Bayes 对统计估计的求解可以通过 Dempster 的结合规则实现 (Shafer, 1976)。

然而证据理论也有其弱点。该理论不能就矛盾证据或不同假设之间具有微弱的支持问题提出解决办法。对该理论值得信服的解释尚未给出 (Veregin, 1989)。对于长链的推证问题，现有理论给出的表达方式还过于繁琐。

由于证据理论是对概率论的一个扩展，因此以往概率论应用的领域，本理论应也可适用。本理论有可能解决决策问题： $0 \in A$  (物体  $0$  属于子集  $A$ ) 该问题定义见 2.2 节。

## 2.5 模糊数学

模糊数学是由 Zadeh(1965) 提出的由计算机处理不精确概率的一种理论。

模糊集合  $A$  是一个具有连续隶属度的集合的物体所组成。该集合的特性由隶属度函数  $\mu_A(0)$  确定 其对每一个物体  $O$  赋予一个属于集合  $A$  的隶属度值 其值范围从 0 到 1 详见 2.2 节。

在模糊运算时, 不精确输入集合首先被精确地确定其隶属度值。模糊运算才可在隶属度函数基础上运算。在模糊数学中的标准运算操作包括余、并、交运算。其分别定义如下:

$$\mu_{\bar{A}}(0) = 1 - \mu_A(0)$$

$$\mu_{A \cup B}(0) = \max[\mu_A(0), \mu_B(0)]$$

$$\mu_{A \cap B}(0) = \min[\mu_A(0), \mu_B(0)]$$

这里  $\max$  和  $\min$  分别是最大和最小操作。 $\bar{A}$  是  $A$  的补集。基于以上三式的模糊数学理论有时亦称之为可能性理论。模糊集合与普通集合运算上的不同在于前者是对模糊隶属度运算, 而后者是对于集合中的元素运算。

模糊数学理论可用于空间数据的不确定性。例如, 利用普通集合理论, 我们难以描述土壤类型的空间扩展、分布。其中的一个困难在于描述两种土壤间的渐变区域。应用模糊数学理论, 可以解决该问题。土壤体可以定义为一个具有空间扩展的集合体。其边界区域可以用隶属度函数定义为一个传递区域。该隶属度函数可用于描述空间内任何一点属于该土壤的程度。

模糊数学的优点在于其处理不确定性的能力。因此, 该理论在进行自然语言有关空间查询方面具有较强的潜在应用可能。例如这样一个查询, 靠近某一建筑 ( $ID = 88$ ) 的范围在哪里? 靠近某一建筑'区域可以用空间域中的模糊子集来描述。

模糊数学的一个弱点是它没有一个像概率论经过严格证明的过程。模糊数学与概率论的区别在于前者的处理对象是可能性而不是概率。因此,模糊数学用于处理不确定性中的不准确性 inexactness 而非随机性 randomness)。

在本研究中,模糊数学将用于讨论连接两线段连接处的不确定性问题。

## 2.6 空间统计理论

统计学是用有序的模型描述无序事件的一门不确定性理论。空间数据是许多地学科学所研究的主题之一。Cressie(1991)将空间数据分为三个大类:地理统计数据、栅格(Lattice)数据以及点数据。地理统计数据描述连接空间内的统计过程[也称为地理统计(geostatistics)]。栅格数据是通过空间中的栅格内的空间过程(时间序列的空间模拟)点模式描述空间点过程,包括标识的点和随机点集过程,对现实世界进行描述。

Cressie(1991)提出了一个通用模型,该模型是适用于地理统计数据、栅格数据、点模式及物体。

设  $S \in R^n$  是一个  $n$  维欧几里德空间的一个一般数据。设一潜在随机数据  $z(s)$  在空间位置  $s$ 。现在假设  $s$  在一个索引集合  $D \in R^n$  中,因此而产生多维随机矢量场(或随机过程)

$$\{z(s): s \in D\}$$

上式的一个实现记为  $\{z(s): s \in D\}$ 。根据  $D$  的定义不同可以有不同的空间数据。

(1) 地理统计数据:  $D$  是  $R$  空间内的一个固定子集,它包含一个正值体积的  $n$  维矩形;  $z(s)$  是在  $s \in D$  内的一个  $n$  维随机矢量。地理统计数据是空间索引  $s$  在  $R^n$  子集中连续变化的数据。

地理统计数据的一个重要应用是根据一系列样本预测矿产的

含量。它既模拟空间趋势又描述其相关关系。空间关系将影响空间模型的估计、预测及设计。

(2) 栅格数据： $D$  是  $R^n$  空间内的一个固定（规则或不规则）的可数点集； $z(s)$  是位于  $s \in D$  的一个随机矢量。栅格引起在  $R^n$  空间内规则的点的设想，而连接到最邻近、次邻近的概念等。

(3) 点模式： $D$  是  $R^n$  内点的过程或是  $R^n$  的一个子集； $z(s)$  是位于  $s \in D$  原一个随机矢量。在最一般的形式中，该产生一个标识的空间点过程。当  $z$  没有定义时，可获得一般点过程。

(4) 物体： $D$  是  $R^n$  内的一个点过程； $z(s)$  其自身是一个随机集合。这就产生了如布尔模型及一个相关的格网增长过程。

传统的非空间模型是空间模型的一个特例。空间统计较之非空间统计是一个更一般性的理论。它可以被认为是对非空间变量的一个空间扩展。而时空统计则是更一般性的扩展。统计理论的发展是从非空间、空间最后到时空域内。

在本研究中，我们将涉及经最大似然分类的遥感影像中的属性数据。该影像含有像素。每一像素被赋予一个类别值。这符合以上栅格数据的定义。因此，空间统计理论可应用于本研究用于描述栅格属性数据。此外，统计学将用于描述线段的不确定性。

## 2.7 本研究中应用的不确定性理论

本研究主要涉及两个问题：(1) 描述 GIS 中物体的位置不确定性；(2) 研究结合位置和属性不确定性的模型。第一个问题的本质是处理空间域的随机误差问题。第二个问题涉及到基于可获得的证据而作一个决策，概率论将用于该问题中。在专家系统中广泛应用的确定性因素模型（certainty factor model）将进一步地被发展来解决本问题。该模型基于证据理论。

基于对不确定性有关理论及需处理的不确定本质的讨论，将

用三个理论构成本研究处理不确定性的理论依据(空间统计、概率论及模糊数学。

## 2.8 本章小结

本章分析了 GIS 和遥感中不确定性的本质。在地学中，还没有提出一个处理不确定性的一般性理论。但是，现有的若干理论可用于处理这些不确定性的若干侧面的问题 这些理论包括 概率论、模糊数学、证据数学理论以及空间统计。这些理论在处理不确定性问题时具有互补性，而且每一理论只可以处理某几方面的不确定性(空间统计、概率论和模糊数学构成了本研究中不确定性处理的理论基础。

## 第三章 地理信息系统遥感及其集成中不确定性的处理方法

### 3.1 概 述

基于前一章所介绍的涉及不确定性有关理论，有许多处理不确定性的方法已经在 GIS 和遥感领域中提出并研究过。本章介绍位置和属性不确定性建模的研究状况。通过分析，确定对这些模型、误差指标的进一步发展领域。通过对 GIS 和遥感领域中广泛使用的不确定性处理模型与方法的分析与比较总结出其功能及局限性。本章首先讨论位置不确定有关处理方法，然后是属性不确定性有关处理方法。一个面状物体是由其边界（闭合的线状特征）构成的，一条线状特征是由一系列线段组成的而一线段则是由连接其两端点的线构成。因此，为了讨论一面状物体的不确定性需要首先研究点、线段、线特征的不确定性。

### 3.2 地理信息系统中位置不确定性的处理方法

本节将分析 GIS 中位置不确定性的各种处理方法，这包括对点、线段和线特征的处理方法。

### 3.2.1 点的位置不确定性的处理方法

点是 GIS 中描述几何特性的最基本元素。一个点通常是由描述其位置成其的坐标  $x_1, x_2, (x_3)$  所构成的。为理解 GIS 中的位置不确定性, 必须研究点坐标的误差。坐标是通过量测值或者对量测值计算而得到的。因此 需要确定点的量测值中的误差 以及由于 GIS 有关操作而导致的误差传播。最后, 需要选择适当的不确定性指标来描述其不确定性的量值大小。

#### 3.2.1.1 GIS 中生成一个点的过程

根据概率论与数理统计, 一幅地图的精度及由此而导出的数据的精度可以描述出来。许多研究人员讨论了数字化的误差, 如 Keefe 等 (1988) 然而, 从统计的观点上看, 需考虑数据获取的全过程, 这包括对自然界的一个点一直到将其表达于 GIS 中的一个点的全过程, 而不是只看其中间的某一个过程。我们需根据一定的假设确定原始观测量的统计特性, 还需要导出 GIS 中一个数字点的统计特性, 从而导出该点的不确定性本质。

产生 GIS 中一个数字点的过程包括: 在现实世界对该点的量测 根据此观测值导出该点的统计参数 例如对于平面坐标导出其期望值  $(\mu, \nu)$  方差、协方差  $(\sigma_{11}, \sigma_{22}, \sigma_{12})$ ; 根据坐标的转换函数及假设传递统计参数 在 GIS 描述点的坐标及以不确定性指标描述其不确定性。产生 GIS 中一点的过程见图 3-1。

大多数现有 GIS 中的点不是通过对地面点进行实地、多次观测而获得的 而是通过对现有地图或影像数字化得来的。因此 它们不是对现实世界中点的多次观测量。对于那些没有多余观测量而获取的点, 这些点的坐标需要通过其它实验而确定其统计参数。因此 图 3-1 是 GIS 中确定一点的理想过程, 有些实际情况并不一定完全满足上述全过程。

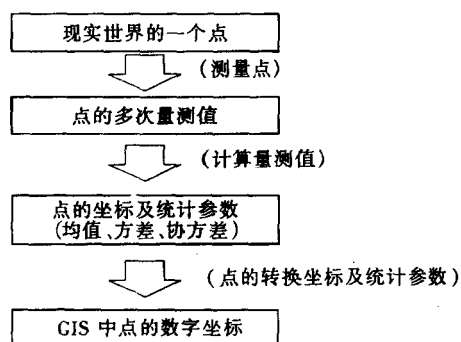


图 3-1 产生 GIS 中一个数字点的过程

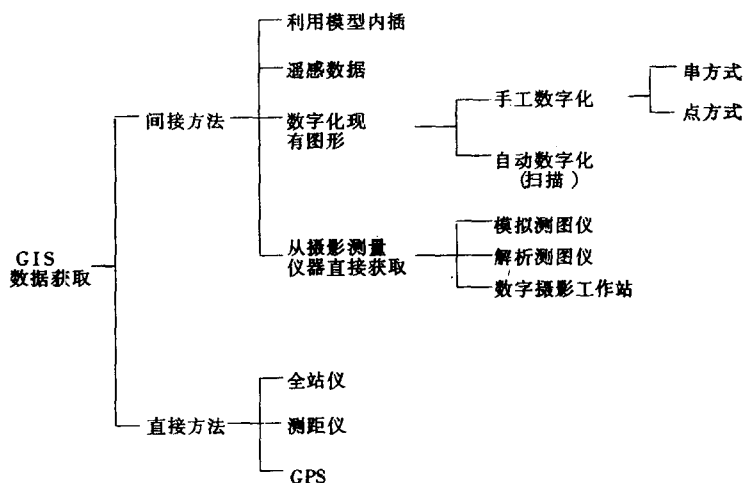


图 3-2 各种数据获取的方法

在上述过程中,有几个问题需要讨论。测量点的统计假设是什么?观测量传递的统计特征是什么?从统计的观点看, GIS 中点

的不确定有什么特点？

根据以上所述 GIS 中点生成的过程 其误差可以被分为以下两类：① 原始观测量的误差；② 数据处理中产生的误差。

GIS 中的数据可以通过直接或间接方法获取。在直接方法的数据获取中（例如用 GPS），GIS 中点坐标的误差是原始测值的传播。通常没有进一步的误差引入。但是，如果该点是通过非直接方法获得的（如通过手工数字化用航空摄影测量获得的纸质地图）其结果的统计不确定性将与量测值的误差及进一步处理过程引入的误差有关。

### 3.2.1.2 估计 GIS 中点不确定性的方法

有三种确定 GIS 中点不确定的方法：一是解析法 即基于统计学中的误差传播定律 其分布、方差协方差的传播 第二种方法是基于试验法；第三种方法是基于模拟的方法。以下分别描述这三种方法。

#### （1）解析法

传播包含有获取相关变量的随机特性，给出独立变量的特性以及两组变量的函数关系。Mikhail 和 Ackermann (1976) 详细阐述了传播原理与技术。

设  $\{x_i\}$  是一组服从  $n$  维密度函数  $f(x_1, x_2, \dots, x_n)$  的一维随机变量。设  $y_k$  是与  $x_i$  函数相关的另一组随机变量

$$y_k = g_k(x_1, x_2, \dots, x_n) \quad (3-1)$$

我们的任务是以  $\{x_i\}$  决定  $y_k$  的随机特性，即确定密度函数  $f(y_1, y_2, \dots, y_n)$ 。

分布的传递

设有一组连续和可导的函数

$$y_i = g_i(x_1, x_2, \dots, x_n) = g_i(\mathbf{x}) \quad (3-2)$$

并且有其传递函数，它们是唯一、连续和可导的。

$$x_j = h_j(y_1, y_2, \dots, y_n) = h_j(\mathbf{y}) \quad (3-3)$$

如果  $f(x_1, x_2, \dots, x_n)$  是随机矢量  $\mathbf{x}$  的联合密度函数 则

$$f(y_1, y_2, \dots, y_n) = f[h_1(y_1, y_2, \dots, y_n), \\ h_2(y_1, y_2, \dots, y_n), \dots, h_n(y_1, y_2, \dots, y_n)] |J_{xy}| \quad (3-4)$$

是导出的随机变量  $y_1, y_2, \dots, y_n$  的联合密度函数。这里  $|J_{xy}|$  是逆 Jacobi 矩阵变化的行列式值。

$$x_j = h_j(\mathbf{y}) \text{ 或 } J_{xy} = \partial x / \partial y \quad (3-5)$$

确定分布传播的困难在于我们需要假定逆函数是唯一、连续和可导的。当  $\mathbf{g}$  是正态随机分布矢量  $\mathbf{x}$  的线性变换时 函数  $\mathbf{y}$  也表示正态分布的随机变量。因此，传播定律的应用通常局限于线性或者可线性化的函数。这时均值和方差、协方差的传播与随机矢量  $\mathbf{x}$  概率分布无关 (Mikhail and Ackermann, 1976)。因此 分布的传播可以简化为均值与方差、协方差的传播。

由于从外业量测值到 GIS 中的数字坐标的转化过程涉及许多复杂的过程，而该进程难以用一个函数或一系列的函数予以定义。我们不能确切地描述出坐标的密度函数。因此，GIS 中数字点坐标的随机特性需进一步研究。

#### · 误差(方差、协方差)的传播

为了应用误差传播定律，首先需对一非线性函数进行线性化，比如利用 Taylor 展开式。这个线性化的过程将会引入近似误差。因此，该线性化只能用于这种误差是可以忽略的情况。线性化以后的函数可以表示为

$$\mathbf{y} = J_{xy}\mathbf{x} + \mathbf{c} \quad (3-6)$$

这里

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Jacobi 矩阵  $J_{yx}$  被定义为

$$J_{xy} = \partial y / \partial x \quad (3-7)$$

或者

$$J_{yx} = \begin{bmatrix} \partial y_1 / \partial x_1 & \partial y_1 / \partial x_2 & \cdots & \partial y_1 / \partial x_n \\ \vdots & & & \vdots \\ \partial y_n / \partial x_1 & \partial y_n / \partial x_2 & \cdots & \partial y_n / \partial x_n \end{bmatrix}$$

并且  $c$  为常系数矢量

$$c = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

矢量  $x$  的协方差阵为

$$\Sigma_{xx} = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1 x_2} & \cdots & \sigma_{x_1 x_n} \\ \vdots & & & \vdots \\ \sigma_{x_n x_1} & & \cdots & \sigma_{x_n}^2 \end{bmatrix}$$

通过使用误差传播定律, 可以获得矢量  $y$  的方差、协方差矩阵

$$\Sigma_{yy} = (J_{yx}) \Sigma_{xx} (J_{yx})^T \quad (3-8)$$

这里

$$\Sigma_{yy} = \begin{bmatrix} \sigma_{y_1}^2 & \sigma_{y_1 y_2} & \cdots & \sigma_{y_1 y_n} \\ \vdots & & & \vdots \\ \sigma_{y_n y_1} & & \cdots & \sigma_{y_n}^2 \end{bmatrix}$$

应用公式 (3-8) 的条件是当线性化近似处理与函数曲线在  $x_i^0 (i = 1, 2, \cdots, n)$  的邻域内对  $x_1, x_2, \cdots, x_n$  的偏离很小。换言之, 在感兴趣的范围内 函数可用其正切值近似地代替。线性函数的方差、协方差传播与非线性函数类似, 只是不需要进行线性化处理。因其已经是式 (3-6) 的形式。

考虑公式 (3-6) 若随机变量  $x_1, x_2, x_3, \dots, x_n$  是两两独立的, 则 (3-8) 中仅对角线上的元素是非零的而非对角线元素均为零。从而随机矢量  $y$  的方差为

$$(\sigma_{yk})^2 = \sum_{i=1}^n (\partial y_k / \partial x_i)^2 (\sigma_{xi})^2 \quad (k = 1, \dots, n) \quad (3-9)$$

进一步地 若变量  $x_1, x_2, \dots, x_n$  不仅相互独立 而且对  $y$  的贡献相等 即等权。  $\partial y / \partial x_i = 1$  对于  $i = 1, \dots, n$ 。则有

$$(\sigma_y)^2 = \sum_{i=1}^n (\sigma_{xi})^2 \quad (3-10)$$

公式 (3-10) 可用于估计 GIS 中点的误差。例如 设想 GIS 中的一个点是通过数字化一幅由摄影测量而获得的纸质地图上的点而得到的。若 GIS 中点的误差由以下独立误差所致：控制测量、摄影、解析空中三角测量、定向、编辑、清绘、印刷以及数字化。则有：

$$\sigma_{GIS} = (\sigma_{控制}^2 + \sigma_{摄影}^2 + \sigma_{空三}^2 + \sigma_{定向}^2 + \sigma_{编辑}^2 + \sigma_{清绘}^2 + \sigma_{印刷}^2 + \sigma_{数字化}^2)^{1/2} \quad (3-11)$$

该“经验”公式也在 Moellering(1987) 中讨论过。可以假设以上各过程是相互独立的，但各过程产生的误差对总误差的贡献并不一定是等权的。但由于权重系数难以确定而假设其为等权的。

#### • 均值的传播

在(3-6)中 若  $x$  的均值矢量为

$$M_x = \begin{bmatrix} \mu_{x_1} \\ \vdots \\ \mu_{x_n} \end{bmatrix}$$

这里  $\mu_{x_i}$  是  $x_i$  的均值。利用传播定律 可得出  $y$  的均值矢量为

$$M_y = J_{yx} M_x + c \quad (3-12)$$

这里

$$M_y = \begin{bmatrix} \mu_{y_1} \\ \vdots \\ \mu_{y_n} \end{bmatrix}$$

如果从实地量测值到 GIS 中的数字坐标的全部转换过程可用一个函数来表达, 则在函数为连续、可导的条件下, 可以用传播定律导出均值、方差和协方差。进一步地, 若这些变换函数是线性的或者可线性化的, 且逆变换函数存在、连续、可导, 则可以导出 GIS 中点的分布。

#### • 关于传播定律的讨论

基于传播定律的解析方法的计算量是适中的。其优点在于可以提供一个输出误差的解析方差形式, 尽管这是一个近似的解。其缺点在于当  $g_k$  [见式(3-1)] 是非线性时, 它仅是一个近似解。若  $g_k$  在  $x_i (i=1, \dots, n)$  处是强烈非线性时, 则这种近似化是不可接受的 (Kuczera, 1988)。另一个局限性在于  $g_k$  必须是连续、可导的。

应用理论的评估方法, 有时不可能知道 GIS 中点的真正统计特性。例如, 系统误差在实际中是有可能遇到的, 由于难以模拟而没有在此处引入。但是我们可以将理论方法获得的结果与实验方法所得结果相比较, 若两者之间有显著差别, 则需要找出其原因。它可能是源于对误差估计的偏差或模型设计上的误差, 例如使用了不合理的近似。因此, 需要对误差模型进行修改。在下一节中我们将讨论试验方法。

### (2) 试验法

#### • 试验场法

为了实施试验场法, 对某一试验目标, 首先需确定试验场中检验点的个数及其分布。这些点的已知值是用比被测试方法更高精度、稳定性的仪器测得的值。之后, 这些检验点再用将被检测的方法测定。后一种方法的精度用最小二乘估计计算出。坐标点的误

差为：

$$RMS_{x_1} = [\Sigma(x'_1 - x_1)^2/n]^{1/2} \quad (3-13)$$

$$RMS_{x_2} = [\Sigma(x'_2 - x_2)^2/n]^{1/2} \quad (3-14)$$

$$RMS_{x_3} = [\Sigma(x'_3 - x_3)^2/n]^{1/2} \quad (3-15)$$

在  $x_1$  与  $x_2$  方向的误差是独立的条件下，点的误差为

$$RMS_{x_1x_2} = [(RMS_{x_1}^2 + RMS_{x_2}^2)/2]^{1/2} \quad (3-16)$$

这里  $x'_1, x'_2$  和  $x'_3$  是平面和高程坐标的“已知”值，即由较高精度的方法测得。而  $x_1, x_2, x_3$  是用被检验的方法测得的观测值， $n$  是检查点的个数。

试验场法适用于那些试验场检查点易于维护且外业试验易于实施的地方。

#### • 实地检验法

当一个被检查区建立之后，我们可以从中选择出若干点进行检验。这些点在实地以更高的精度再重新量测。根据 (3-13) 至 (3-16) 公式，可以计算出这些点的误差。根据美国国家制图精度标准 (USNMAS)，当被选择的点有 90% 高于标准规定的精度，即该结果方可接受。

试验场法与实地检验法有所不同，前者是用预先选定场地中预先选定的固定个数的点进行检验。因此，前一种方法适于对某种方法或仪器进行检验，即对于某一单一因素进行检验。而后一种方法适用于对某一产品结果进行检验，例如一幅地图的精度检验。其中的最小二乘误差包含所有的误差，它是整体误差的量测值。例如，使用该方法量测 GIS 中的图形数据的精度，则最小二乘误差包含了实地量测、绘图、图纸变形、数字化以及数据表示等几方面误差的内容。

#### • Monte Carlo 模拟法

如果式 (3-1) 中的  $g_k$  函数不是连续、可导的，或者由误差传播

定律引起的近似误差是不可接受的，这时解析方法则不能使用。在这种情况下，可以用 Monte Carlo 模拟法 (Hammersley and Handscomb, 1979) 描述其误差传播。Monte Carlo 的基本原理是重复地输入  $x_i (i = 1, 2, \dots, n)$  计算式 3-1 的输出值。其中  $x_i (i = 1, \dots, n)$  是依其分布随机抽样的结果。这样，就可以获取  $y_k$  的样本 并依此估计其统计参数 如均值  $\mu_y$  和方差  $\sigma_y^2$ 。

Monte Carlo 法的最重要的优点是它可以在任意精度水平下产生  $y_k$  的分布，另外它对于简单过程很容易实施。复杂过程（如用摄影测量确定点坐标）则比较难以模拟。此时 实施 Monte Carlo 模拟则较为复杂。该方法的另一缺点是它比其它方法通常需要更长的计算时间 此外 该方法不可能生成一个  $y_k$  分布的解析表达式。因此 对于不同的情况 必须进行不同的、新的模拟。这使得分析十分费时。

在讨论了 GIS 中不确定性之后，现在需要找出表达这些不确定性的方法。这是引出不确定性指标的原因。

### 3.2.1.3 位置不确定性指标

#### (1) 点坐标的准确度

在描述不确定性时，首先需要区分两个概念：精度与准确度。精度是指量测值与其真值的差别程度。在统计学词汇里，准确度是对于一均值离散程度（通常用标准方差表示），通常如果不是可能的话 获得一真值是比较困难的。因此在描述不确定性常常使用准确度而不是精度的概念。理论上，准确度不同于精度。然而，在许多实际情况中，准确度度量值被用作精度的量测指标。

随机变量的准确度是通过其方差、协方差矩阵表示的。点坐标的随机矢量  $y$  的准确度依赖于量测矢量  $x$  的准确度 以及量测量与坐标之间的关系。如果坐标与量测值之间有函数关系  $g$

$$y = gx \quad (3-17)$$

在线性变换之后 导出坐标的方差、协方差可以依据下式用误差传播定律可得

$$\Sigma_{yy} = J_{yx} \Sigma_{xx} J_{yx}^T$$

这里  $J_{yx}$  是式 3-17 的 Jacobi 矩阵。

在对于原始观测量后续处理的每一步，例如一个坐标变换  $y' = h(y)$  都对坐标准确度有一个影响。

$$\Sigma'_{yy} = K \Sigma_{yy} K^T \quad (3-18)$$

其中  $K$  是某一 GIS 操作的误差影响。

在获得点的方差、协方差阵之后，一个点的误差可以用误差椭圆或圆形正态模型表示出。

## (2) 标准椭圆

若一个点的  $x_1, x_2$  坐标服从正态分布 则其二维正态密度函数有一个类似于钟形的形状，可以用标准椭圆来表示点座标的准确度 (Mikhail and Ackermann, 1976)。见图 3-3 所示。

两坐标可以写为如下随机矢量：

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

设二维随机矢量的概率密度函数可以写为如下形式：

$$f(x_1, x_2) = (|\Sigma_{xx}|^{-1/2} / 2\pi) \exp \left[ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma_{xx} (\mathbf{x} - \boldsymbol{\mu}) \right]$$

这里  $\boldsymbol{\mu}$  是均值矢量 且  $\Sigma_{xx}$  是  $x_1, x_2$  的方差、协方差阵。

为了找出具有相同概率点的轨迹，以下条件必须满足：

$$f(x_1, x_2) = \text{常量} \quad (3-19)$$

有变量椭圆的等式

$$\mathbf{Q} = [q_1, q_2]^T$$

其为  $[\Delta x_1, \Delta x_2]^T$  的变换。它是量测值与均值的一个微分矢量 (Bouloucos, 1984)。

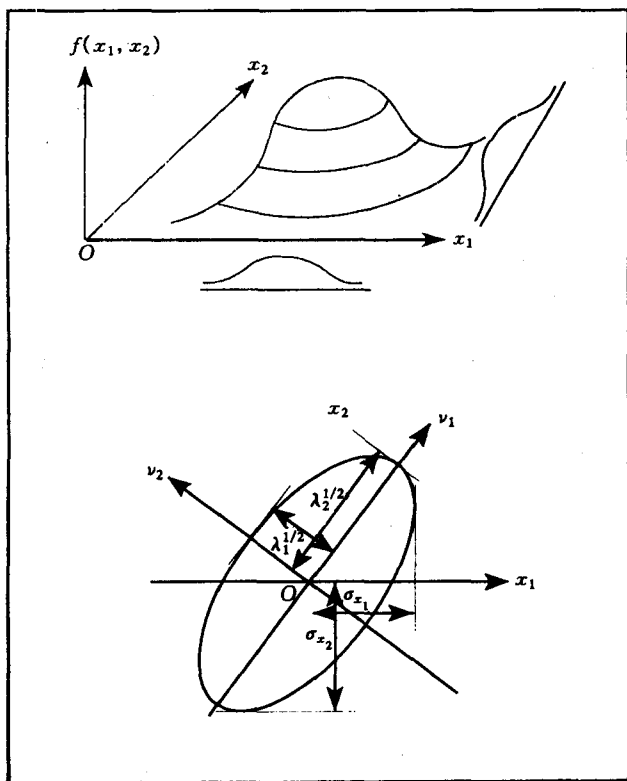


图 3-3 二维正态分布与标准椭圆

$$q_1^2/(x_0 \cdot \text{sqrt}(\lambda_1))^2 + q_2^2/(x_0 \cdot \text{sqrt}(\lambda_2))^2 = 1 \quad (3-20)$$

这里  $\lambda_1, \lambda_2$  是矩阵  $\Sigma_{xx}$  的特征值 而  $V = [v_1, v_2]^T$  是对应的特征矢量。

$$Q = \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = v^T \Delta x \quad (3-21)$$

这里

$$\Delta x = x - \mu_x$$

当常量  $x_0^2 = 1$  时 式 3-20 变为

$$q_1^2/(\text{sqrt}(\lambda_1))^2 + q_2^2/(\text{sqrt}(\lambda_2))^2 = 1 \quad (3-22)$$

这就是大地测量中所用的误差椭圆。标准椭圆的中心位于  $(\mu_{x_1}, \mu_{x_2})$ ; 特征矢量  $\nu_1$  和  $\nu_2$  确定椭圆的轴方向 半轴的长度等于特征值  $\lambda_1$  和  $\lambda_2$  的平方根。轴对于坐标系  $Ox_1x_2$  的斜度指出了  $x_1$  与  $x_2$  的相关程度 见图 3-3)。

### (3) 点误差的圆形正态模型

如果点在  $x_1$  和  $x_2$  方向上的误差相同并且相互独立, 误差椭圆则成为一个圆, 同样其中心也在  $(\mu_{x_1}, \mu_{x_2})$  上。当点的准确度在  $x_1$  和  $x_2$  方向上相差不大时, 亦可近似地使用该模型。其结果是一个圆形正态模型 (见图 3-4)。点的位置误差表示成为一个以  $(\mu_{x_1}, \mu_{x_2})$  为中心的钟形曲面, 某一点为真的概率用这一点相应于钟形模型中的高度表示出来 (Goodchild, 1991)。根据正态分布模型及圆形正态模型, 我们可以定义若干用于误差指标的统计量的名称。

由于是随机误差, 误差的频率分布可以从数学上用正态分布表示。一维坐标的误差, 例如坐标的高程值, 用一维正态分布来估计。二维位置数据, 如一个点的  $(x_1, x_2)$  坐标用二维正态分布表示。这个模型被简化、标准化地表达为相等的圆形误差概率。因此, 水平位置精度可以表示为一个圆形的分布。

二维圆形误差是一维线性误差的扩展。而圆形误差分布是对椭圆误差分布的一个简化以便于概率计算。我们从线性误差指标讨论开始。

对于一组一维量测值  $h_i (i = 1, 2, \dots, n)$  使用一维正态分布, 标准差  $(\sigma')$  即方差  $\sigma^2$  的估计量 通常可以用以下方法计算:

$$\sigma' = [(h_i - \mu')^2 / (n - 1)]^{1/2} \quad (3-23)$$

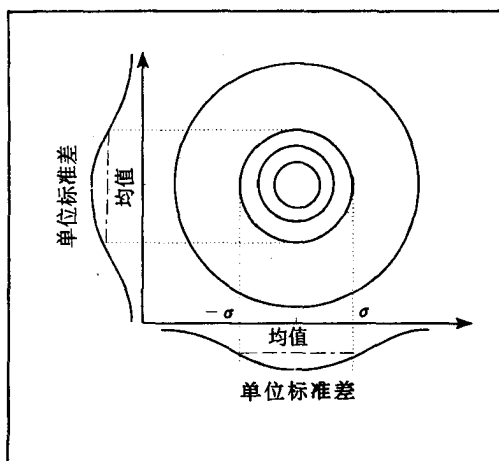


图 3-4 圆形正态模型

这里  $\mu'$  是量测值的平均值。

在一维正态分布曲线下，在  $[-\sigma, \sigma]$  区间的面积为总面积的 68.27%。

标准差  $\sigma'$  是用于量测数据的基本统计参数。但是地图的精度通常表示为 90% 的置信水平，称之为线性地图精度 (LMAS) 详见 CCOSAM(1982)。

其他两个误差指标：可能误差 probable error 和接近确定误差 (near certainty error,  $3\sigma$ ) 被用于表示 50% 和 99.73% 的概率。高度的均方差误差 (MSEH) 用于指示地图中的高度误差。这些指标的比较见表 3-1。对于表 3-1 和表 3-2，“P” 表示“百分比概率”。

圆形分布是从二维正态分布中导出的。它的基本统计参数是标准差  $\sigma_c$  详细定义见 CCOSAM(1982)。

$$\sigma_c = 0.707(\sigma_{x_1}^2 + \sigma_{x_2}^2)^{\frac{1}{2}}$$

表 3-1 线性误差指标

线性误差			
名 称	符 号	$P(\%)$	公 式
标准误差	$\sigma$	68.27	$1.0\sigma$
可能误差	PE	50	$0.6745\sigma$
线性地图精度	LMAS	90	$1.6449\sigma$
接近确定误差	$3\sigma$	99.73	$3.0\sigma$
均方根误差*	RMSEH	68.27	$1.0\sigma$
数据	高程点、等高线		

\* 在文献中定义为均方差。

在一个圆形正态分布中，半径为  $\sigma_c$  的圆表示了 39.35% 的概率。除该误差指标外，其他指标包括：圆形可能误差 (circular error probable, CEP) 它是一个半径为  $1.1774\sigma_c$  的圆表示了 50% 的概率 圆形地图标准 (circular map accuracy standard, CMAS) 它是一个半径为  $2.146\sigma_c$  的圆表达了 90% 的概率；圆形接近确定误差 (circular near certainty error) 半径为  $3.5\sigma_c$ ，以及平面均方根误差 (root mean square error of planimetry, RMSEP)。表 3-2 列出了对这些指标的比较。

表 3-2 圆形误差指标

圆形误差			
名 称	符 号	$P(\%)$	公 式
圆形标准误差	$\sigma_c$	39.35	$\sigma_c$
圆形可能误差	PE	50	$1.1774\sigma_c$
圆形地图精度标准	UMAS	90	$2.146\sigma_c =$ $1.5117RMSEP$
圆形接近确定误差	$3.5\sigma_c$	99.78	$3.5\sigma_c$
均方根误差*	RMSEP	63.21	$1.412\sigma_c$
数据	平面		

\* 在文献中定义为均方差。

从以上分析可以看出，我们已发展了描述一维变量误差分布的误差指标 (线性误差指标) 以及描述二维变量的指标 (圆形误差

指标)但是 我们还没有描述二维以上变量的误差指标 比如一条线段,他是由四维随机矢量构成的。由于线段是构成 GIS 中其它几何特征的基本元素,因此对于线段误差指标的定义就十分重要。在 4.4.1.4 节中将介绍对线段描述的误差指标。

### 3.2.2 线段及线特征不确定性的处理方法

#### 3.2.2.1 GIS 中线的分类

在 GIS 中,若干种不同的地理特征被表达成为线 ( Mark, et al.,1989)。表 3-3 表示出对若干种在 GIS 中表示为线的特征。

表 3-3 地理线的种类 选自 Mark and Csillag, 1989)

现实世界特征	示 例
1 数学线	经、纬度
2. 法律线	政治边界线
3. 线特征	
a) 线特征	铁路、公路
b) 似线特征(变宽)	河流
4. 0集表面	
a) 定义好的表面	等高线
b) 复杂、动态面	海岸线
5. 面状边界	气候区域边界
	植被边界
	土壤边界

根据不确定性引入 GIS 中线过程不同,可以把以上线特征分为两类 即类型-I 和类型-II 线。这两类线的最基本的差别在于 类型-I 线在现实世界中有真正的点确定其位置;而类型-II 线特征则在现实世界中没有明确点描述其位置,因而需要我们确定其位置。比较这两类型的线我们知道,类型-II 线特征有额外的误差源 即确定现实世界中描述线的点位置误差 有时称之为解译误

差。这两类型的线列于表 3-4 中。

表 3-4 类型-I 线与类型-II 线特征的比较

类型-I 线	类型-II 线
现实世界中有确切的点定义线特征	现实世界中没有确切的点定义线特征， 人们需要解译而确定这些点的位置
GIS 中线不确定性误差源	GIS 中线不确定性误差源
——量测误差	——线解译误差
——操作误差	——量测误差
	——操作误差
示例	示例
——地籍边界	——土壤边界
——政治边境线	——森林边界

第 I 种线的误差源主要来源于量测和操作误差。因此，第 I 种类型的线不确定性可以量化。例如，我们可以假设坐标的误差是独立的并且服从正态分布。第 II 类型线的除具有像第 I 种线所具有的量测和操作不确定性误差源外，还具有最为显著的解译误差。由于这种误差涉及人为解译误差，因此它相对地难以量化。

这里侧重考虑由具有随机误差的点所构成的线的不确定性。由于类型-II 线存在人为解译的误差，其特性不易量化而不是十分清晰。因此，在本节中把讨论的范围局限于类型-I 的线特征上。而线段的不确定性是由于所构成点的不确定性所产生的。

### 3.2.2.2 线的位置不确定性模型

确定类型-I 线不确定性的最基本设想就是通过所构成点的不确定性导出线的不确定性。在点之间的误差相互独立并已知其方差、协方差的假设下，有两种基本的模型可用于描述线的不确定性： $\epsilon$ 带和误差带。

在许多 GIS 中，一条矢量的线是由连接所构成的点的线段而

构成的。因此，最直观的方法就是从所构成点的误差导出相应的线的误差 如图 3-5 所示，这里实线为真实的线，而虚线为量测线。

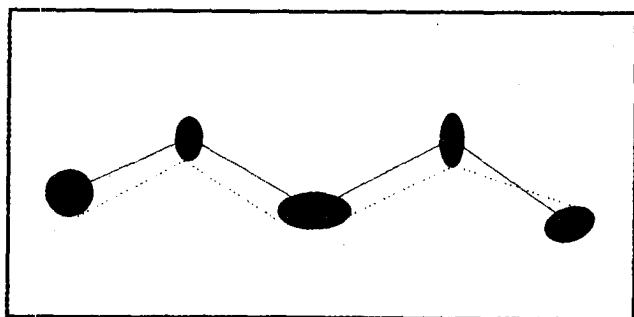


图 3-5 线位置误差的建模是基于各数字化点的随机误差。误差椭圆用于表示点的误差。其中实线为真实线 而虚线表示量测线

本模型不适用于串方式数字化的线 因为在那种情况下 不可以假设点之间的误差是相互独立的。即使对于那些以点方式数字化的线 若其中含有大量的未改正的系统误差 也不能作点之间误差相互独立的假设。例如纸质系统变形即属这种情况。

通常地 在数字化前的几何配准有一个移位 这将对地图上的每一个点产生一个位移。真实线与数字化的线之间不能用一系列相互独立的点误差的模型予以描述 ( Goodchild, 1991)。但是若有一个模型 它可以结合系统误差 如由于地图配准误差所致 和随机误差 (如数字化所引起的误差)，这样一个模型可以作为线位置误差模型。

### (1) Epsilon 带模型

#### ·定义

Epsilon带是这样—个定宽带，它是沿着—条线或多边形边界线两侧具有定宽 Epsilon 的带所构成。该模型是由 Perkal (1956,

1966 提出的 而 Chrisman(1982)和 Blakemore(1984) 对之作了进一步讨论。假设存在抽象的或真实的线，该模型提出真实线将落在 Epsilon 带内。

Epsilon 带模型是基于若干假设的：① 对于 GIS 中每一特定线的每一个误差影响可以被视为一个随机变量 ② 在 GIS 中产生一条数字线的过程可以被看成是独立的。

设一条制图线是一条直线的近似，可以假设真实线是在量测线两侧的 Epsilon 宽度之内。对一条直线段，其轨迹较为简单，是一个平行于线段的矩形。其宽度为 Epsilon 的两倍 并且在两端点处几何地与两个以 Epsilon 为半径的圆相并。基于线段的 Epsilon 带模型，我们可以模拟更复杂的图形(Chrisman, 1982)。Epsilon 带模型见图 3-6 其中实线是量测线，而虚线为线的真实位置。真实

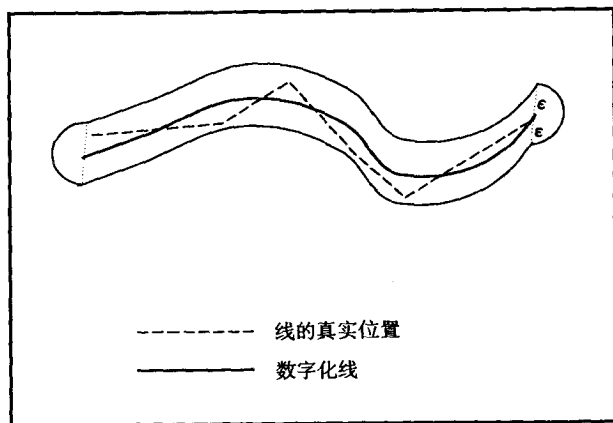


图 3-6 Epsilon 带模型

线以某一概率落于 Epsilon 带内。Epsilon 带也可以用通过沿量测线滑动一球所生成的轨迹来生成。

#### •应用

Epsilon 带模型可用于指示出面积计算时的误差。Chrisman (1982) 应用 Epsilon 带模型描述土地利用覆盖专题图的误差。首先他对于该问题计算 Epsilon 的值, 有三个可度量误差因素影响着 Epsilon 的宽度: 清绘、数字化和舍入误差。这三项误差的偏差值分别为 12.5 m、8.3 m 和 2.9 m。根据这三种误差相互独立的假设, 可以应用误差传播定律计算出总的误差为 15.2 m。出于保守考虑在这里 Epsilon 的宽度定为 20 m。其次对于所有的线生成其 Epsilon 带。Epsilon 带共占有的面积为 7 191 hm<sup>2</sup> 而该图的总面积为 100 000 hm<sup>2</sup> 即约 7% 的面积被 Epsilon 带占有。这个面积意味着可能落入其他土地利用覆盖类型的面积数, 即误差。

#### •分析

Epsilon 带模型在应用于描述 GIS 中线的不确定性时是一个有价值的模型。但是, 在应用该模型时仍有若干局限性值得注意。Goodchild 和 Dubuc (1987) 曾对此作过讨论。

Epsilon 带模型曾用于确定和概率两种描述的形式。在确定的形式中假定该 Epsilon 带以 1.0 的概率包含了真实线。另一方面以概率的形式的描述人们可以假定该线以某一概率包含了真实的线。

Epsilon 带模型有若干方面的局限性。确定形式的 Epsilon 带模型指出所有的线全部落入该 Epsilon 带内。但是直观感觉告诉我们这种上界是不存在的事实上只能期望较大的误差出现的机会较小。量测值的误差模型如高斯分布模型没有定义误差的上限。

其次该模型没有提供在 Epsilon 带内分布的模型。Honeycutt (1986) 发现制图综合产生的误差趋向一个双峰分布。

从以上两个方面的局限性来看在 Epsilon 带内的分布问题值得进一步讨论。在第四章中将进一步详细阐述。Honeycutt 提出的

双峰分布仅是描述制图综合中的误差所产生的分布。而制图综合仅可能是造成 GIS 中误差诸多因素之一。我们更需要理解诸多因素所构成的总体误差的分布。

Goodchild 和 Dubuc(1987)指出,Epsilon 带以及其改进版本指出了线上的一个点对其偏差的模型,而没有提出关于线本身的模型。为进一步研究此问题,Dutton(1992),Caspary 和 Scheuring(1982)用 Monte Carlo 方法模拟了线段的概率分布。然而,如果能为该问题可以给出理论上的证明,就更加完美。

## (2) 误差带模型

在 Epsilon 带模型的基础上,人们又进一步地研究了误差带模型(Zhang and Tulip,1990;Dutton,1992;Caspary and Scheuring,1992)。Epsilon 带与误差带的差别在于,Epsilon 带认为不论在何处带的宽度都是等宽的;而误差带模型指出,该带在中间最窄而在两端较宽。

误差带是基于以下两个假设的 ① 两端点的误差是相互独立的 ② 两端点的误差有相同的方差、协方差。

在这两个假设下,可以根据误差传播定律导出线段上任意一点的误差。在线段  $z_1 z_2$  上的任意一点  $z_i$  在  $x_1$  和  $x_2$  方向上的方差  $\sigma_{x_{1i}}^2$  和  $\sigma_{x_{2i}}^2$  与  $z_1$  和  $z_2$  的方差。在线段上的任意一点  $z_i$  可以表示为

$$x_{1i} = (1 - r)x_{11} + rx_{12}$$

$$x_{2i} = (1 - r)x_{21} + rx_{22}$$

应用误差传播定律 有

$$\sigma_{x_{1i}} = \sigma_{x_{2i}} = (1 - 2r + r^2)^{\frac{1}{2}} \sigma \quad (0 \leq r \leq 1) \quad (3-25)$$

这里  $r = l_i/l$ 。  $l_i$  是  $z_1$  到  $z_i$  的距离  $l$  是  $z_1$  与  $z_2$  之间的距离  $\sigma^2$  是两端点在  $x_1$  和  $x_2$  方向上的方差,即对于端点  $z_1, z_2$  有  $\sigma^2 = \sigma_{x_1}^2 = \sigma_{x_2}^2$ 。  $z_i$  点的 RMS 误差为

$$RMSP_i = \sqrt{2}\sigma_{x_i} \quad (3-26)$$

基于以上公式 误差带的形状是由  $RMSP_i$  决定的(见图 3-7) 而不是 Epsilon 带所描述的一个矩形。

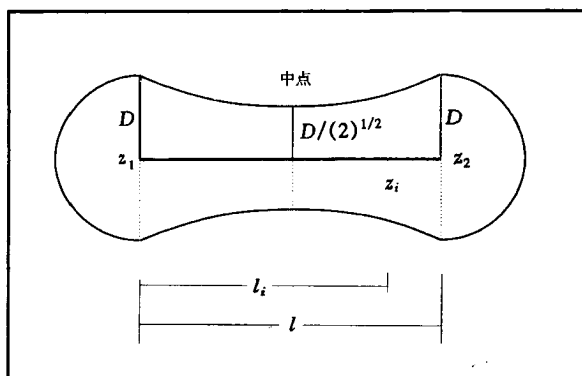


图 3-7 线段  $z_1z_2$  所构成的误差带例子

我们可以根据以上结果计算出该误差带的面积  $A_2$

$$\begin{aligned} A_2 &= \pi\sigma^2 + 2l\sigma \int_1^0 (1 - 2r + 2r^2)^{\frac{1}{2}} dr \\ &= 0.8A_1 \end{aligned}$$

其中  $A_1$  是在相同条件下算得的 Epsilon 面积。由此可以看出 Epsilon 带的面积较之误差带的面积略大一点。

误差带是由四维随机矢量所构成的二维影像。Caspary 和 Scheuring(1992)认为解析地描述其形状是比较困难的 因此他们及 Dutton(1992)使用了 Monte Carlo 模拟的方法模拟出误差带及分布的形状。误差带的解析描述尚未有深入探讨过。这是应用误差模型于 GIS 中的一个主要障碍。因此,该问题被确定为本研究的主要研究问题之一。在第四章中,将详细阐述本研究在该问题上的

统计证明过程。

### 3.2.3 多边形位置不确定性的处理方法

在我们讨论的范围内，线性特征是由线段所构成的线或多边形。一条线段是由其两端点所构成的。因此，多边形的精度是由其所构成的结点精度所决定的。有三种方法可用于决定多边形的不确定性 这些方法分别是基于：(a) 随机矢量分析；(b) 误差传播定律；(c) Monte Carlo 模拟方法。

#### 3.2.3.1 基于误差传播定律的方法

为了简化问题，多边形的不确定性由其若干特征的不确定性所描述。这些特征包括多边形的面积、周长长度或其中心的位置。这里假设多边形节点的具有随机误差、相互独立且服从正态分布。首先，我们建立多边形的特征与其节点的函数关系。应用误差传播定律于线性化后的这些函数，就可得出这些特征的不确定性。Caspary 和 Scheuring(1992) 曾导出这些特征的误差计算公式。

#### 3.2.3.2 基于 Monte Carlo 的模拟法

Monte Carlo 模拟法是最直接的方法。在这种方法中，根据每节点的误差特性产生其一次实现，将每节点的这一次生成相连而构成多边形的一次实现，重复此过程到  $n$  次(例如  $n = 1000$ ) 我们可以得到多边形的分布。Dutton(1992) 完成了用 Monte Carlo 模拟多边形分布的实验 但基于这种模拟法的实验 不可能导出密度函数的解析表达式。

### 3.2.4 面状物体位置不确定性的处理方法

面状物体的不确定性是由一点属于该面状物体的描述实现的。Blakemore(1984) 应用 Epsilon 带模型于“点属于多边形”的问题

之中，即描述由一个多边形所包括的面状物体的不确定性。他指出，一个多边形与一个面状目标有以下五种关系：

- 一定在其内 (点 5)
- 一定在其外 (点 1)
- 可能在内部 (点 4)
- 可能在外部 (点 2)
- 不确定 (点 3)

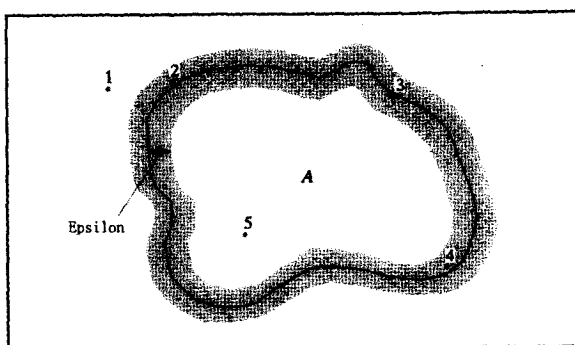


图 3-8 面状物体的不确定性

若一点落在面状物体的内部核心部分，则该点与面状物体之关系为“一定在其内”；如果一点落在 Epsilon 带的内半部，则其为“可能在内部”；若一点落在 Epsilon 带的外半部，则该点“可能在外部”；若一点落在 Epsilon 带及该状物体的外部，则其为“一定在外部”；若一点恰好落在数字化的线上，则其为“不确定”。面状物体的不确定性是由 Epsilon 所占有的面积表达的。若这部分面积越小，则越少点有机会落在“可能在其内”或“可能在其外”的区域内，即其不确定性就较小。Blakemore(1984) 曾用此方法描述一多边形网的不确定性。

### 3.2.5 位置不确定性模型中需进一步探讨的若干问题

#### (1) 点模型

点的位置不确定性处理模型已经发展得比较完善,这包括利用误差传播定律来决定点的误差 描述点分布的模型 二维正态分布或简化的圆形正态分布模型;误差椭圆以及一组误差指标。理论上讲,这些方法可直接用于描述 GIS 中点的不确定性。但是由于 GIS 中的点仅有一次观测值,而且不附有方差、协方差阵。因此 实际上以上方法直接用于描述 GIS 中的点的不确定性仍有困难。为了应用这些模型 我们或者要获得更多的观测量 或设立某些统计上的假设。例如我们可以利用 3.2.1.2(2) 节中的实验法获得点的方差值。

#### (2) 线段模型

线段的不确定模型已经过相当的讨论,但是到目前为止有若干问题有待解决。首先,如何确定一条围绕量测线段的区域线段的真值以一定的概率被包含在此区域内?该区域的形状如何?所确定的概率与区域大小的关系如何?尽管 Epsilon 带模型提供了一个包含线的真实位置的区域,但它没有定义区域大小与预先定义的概率、区域与线段两端方差之关系。

其次,如何确定量测线段围绕其真值位置的分布。Dutton (1992) 用模拟的方法讨论了其分布问题。但是问题在于这种方法未能给出分布的解析表达形式,例如概率密度函数。

第三,在讨论中有两个概念经常被混淆。一个是围绕着量测线所构成的区域 线的真实位置以某一概率被包含在该区域内 另一个是概率分布范围 它是围绕着线的“真实”或期望值的一个范围。

第四 到目前为止尚未有对于线段而定义的误差指标 尽管在测量中已定义了许多关于点的误差指标。

### (3) 线性特征模型

在现有模型中，对于线段及线的处理是用相同的方法。但是，在线段交叉处其不确定性比较复杂需进一步的讨论。为了描述面状物体的不确定性，还必须描述线状边界之不确定性。同样出于此原因，两线段连接处的不确定性需进一步考虑。

### (4) 面状物体模型

Blakemore(1984) 提出一个描述面状物体不确定性的模型，他建立了五种点与面状物体的定性关系。应用这种模型与其它不确定性相结合时会有困难 因为该模型不能提供一个连续变化的、定量的不确定性指标。这是应用该模型于定量分析时的一个局限性。因此 需要发展一个定量的、可连续描述点与面状物体概率关系的模型。

## 3.3 属性不确定性的处理方法

在本节中 我们将讨论两种属性数据 类别属性数据与连续属性数据。由于对于连续属性不确定的处理方法可以与用于处理位置不确定方法类似。因此，侧重讨论类别属性不确定性的模型与方法。我们从属性数据的定义开始 之后讨论其数据不确定性 最后阐述用于处理这些不确定性的方法。

### 3.3.1 属性数据

属性数据是指对于某一 GIS 中点、线、多边形或遥感影像的属性值或属性类别描述。一般地，属性数据被认为是对物质、特性、变量或一地理目标的质、量的描述指标 ( Stutheit, 1990; Rodcay, 1990)。

属性数据可以具有 ( 离散 ) 类别值或连续值。他们也可区分为定性或定量属性值。一个类别变量可以仅仅是一个有限集合内的

有限个元素，例如包含所有土地覆盖的一个分类系统中的有限土地类型。另一方面，一个连续的变量，可以取某一个区间内的任何值。对于类别属性而言，数值本身并不一定具有先后、大小的涵义。例如，环境质量指标从 1 到 4，依次表示最好到最差。这时，类别值有先后次序的涵义。又如类别 1 至 4 分别表示水、森林、城市用地、植被四种不同类别用地。这时，类别值并没有任何大小先后的涵义。一个连续变化的属性值，某地区温度从  $-40^{\circ}\text{C}$  到  $+50^{\circ}\text{C}$  这时属性值可以是区间  $[-40, +50]$  之间的任意值 取值的个数是无限的。

### 3.3.2 属性数据的不确定性

属性的不确定性是由属性的取值与其真值的相差程度决定的。由于有两种不同的属性数据，因此我们也定义两种不同的属性不确定性。连续属性数据的不确定性可以用量测误差决定。对于这种不确定性的建模，可使用与位置不确定性相同的方法，例如误差传播定律。

类别属性数据不确定的评估是分类不确定性评估的范畴。分类精度评估是一个复杂的过程，在遥感领域中已作过长期的研究。分类精度评估的难度在于其受到若干因素的影响，如类别，具体区域的形状、大小，试验点的选择方法，以及相互混合类别情况 (Aronoff, 1989)。

GIS 中一个物体的属性不确定性与数据获取的过程有关。从现实世界到 GIS 中一物体的理解过程被示于图 3-9。若 GIS 中一物体的获取是通过一分类影像，则主要的误差来源于遥感分类。关于确定其不确定性的方法将在以下部分阐述。有时，人工译读误差也会介入。这时 不确定性与操作员的经验、技能有关 也会因人而异。下面侧重于讨论由于遥感分类技术本身而造成的属性不确定性。

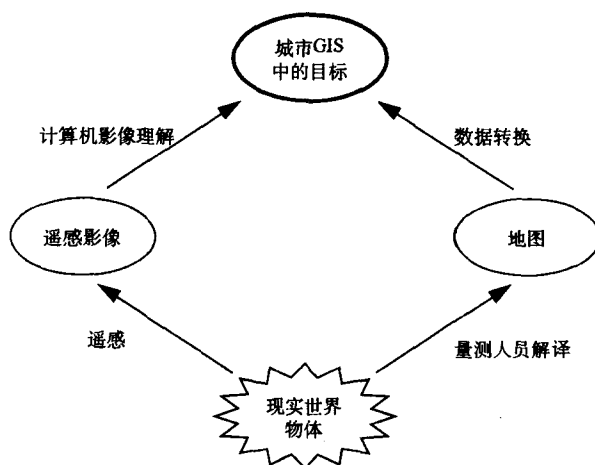


图 3-9 从现实世界到 GIS 理解一个物体的过程

在这些过程中 以下因素会导致不确定性 自然界本身、人工解译、自动分类或后分类过程。

由 2.1 节所述，分类是决定一物体  $O$  属于子集  $A$  的一个过程。因此其不确定应含有以下三方面：类别的定义；物体的量测；决策物体  $O$  属于  $A$  的过程。

不确定性可能在于类别的定义中。类别  $A$  的定义可能模糊，使得在区别  $A$  与另一类  $B$  时有困难。土壤分类是一个比较典型的例子。物体量测不确定性是第二个因素 有关内容见 1.3.1 节。

在作出决定  $O \in A$  时，对于不同的地区、像素可能会有所不同。例如 在一个多边形的中心 我们对于  $O \in A$  有比较高的信心。而在多边形的边界附近，我们的信心会较低。Burrough 和 Heuvelink(1992) 定义了变换区域模型用于描述此类不确定性。一个多边形划分的区域内可能会是异质的，例如一个区域内含有 80% 类别  $A$ , 20% 的类别  $B$ 。但是这一描述中没有说明不确定性

的空间分布 即哪些地方  $A$  多  $B$  少 而哪些地方  $B$  多  $A$  少。概率矢量可对此类问题提供一个方面的解决方法, 有关内容将在下一章内详细介绍。

### 3.3.3 基于采样的检验方法

在现有的文献中总结有三种基于实验的检验方法: ① 对于某一类或全集正确量测的百分比 (Rosenfield, 1986); 在某一置信水平下, 某一类或全集正确量测的百分比 (Aronoff, 1985; Hord and Brooner, 1976); 基于某些参数的某一类或全集正确量测的百分比 (Greenland et al., 1985; Rosenfield and Fitzpatrick-Lins, 1986)。

以上三种方法适用于各种非连续属性值的精度评估。本节将集中讨论遥感分类影像的精度评估问题。非连续属性数据的评估可以通过对一组分类结果的评价得以实现。地面真实数据有时也称之为参考数据。通过将量测数据与参考数据的比较, 我们可以建立一个误差矩阵。该矩阵可以描述某一类别的分类精度、或者是整体分类精度。基于这一矩阵, 可以对分类精度进行进一步地讨论。

#### 3.3.3.1 参考数据的采样

在基于采样数据的属性不确定性评估方法中, 采样数据作为误差矩阵或其他统计分析的输入部分。在这个过程中, 选择适当的采样数据是非常重要的。有两个因素影响采样数据的选择: 采样样本的大小及采样数据分布模式。

在评估分类遥感影像的分类精度时, 采样点的数量是十分重要的。获取地面采样数据是昂贵的, 因此采样点的个数尽可能减少。另一方面, 为了在统计上有意义, 采样点的数目应尽可能大, 至少大于某一给定数目 例如 30 个。人们在采样点个数方面进行了许多探讨 (van Genderen and Lock, 1977; Rosenfield, 1982; Congal-

ton, 1988; Fukunaga and Hayes, 1989)。

另一个重要因素是采样模型。选择适当的采样点分布方式，使所选择的样本可以代表全部分类的影像是十分重要的。一个较差的采样可能导致严重精度评估的偏重，可能对精度估计得过高或过低。通常使用的采样模式有五种：简单随机采样 (Cochran, 1977); 集群采样 (Kish, 1965)；分类随机采样 (Barrett and Nutt, 1979); 系统采样 (Barreft and Nutt, 1979) 以及分类系统非一致采样 (Berry and Baker, 1968)。

Congalton(1988)对各种不同的地区进行了采样模拟，并总结出在各种情况下简单随机与分类随机采样提供最佳的采样结果。根据面域上的空间自相关属性，其他采样方法也适用。

当采样个数与模式确定之后，即可实施采样，从而可以生成误差矩阵，依此进一步进行有关的属性精度评估。

### 3.3.3.2 误差矩阵

误差矩阵，有时也称为混淆矩阵，是一个用于表示分为某一类别的像素个数与地面检验为该类别数的比较阵列。一个误差矩阵的实例见表 3-5。表中的列通常表示参考数据 而行表示遥感分类的结果。误差矩阵通常用于表示分类的精度，因为它可用于指出某一类的或整体分类的精度。此外，用误差矩阵还可以表示出包含与丢失两种误差。在表 3-5 中，A、B、C 是三类待分类的类别名

表 3-5 误差矩阵实例

分 类 数 据		地面真实数据			行总和
		A	B	C	
	A	45	2	5	52
	B	3	63	5	71
	C	2	5	70	77
	列总和	50	70	80	200

称。第一行的数目，如总数‘52’表示为 A 的像素中根据实地检查有 45 个被分为类别‘A’；2 个像素被分为类别‘B’；5 个像素被分为‘C’。

根据误差矩阵，可以导出若干关于总体分类，或对于某一类别分类的精度描述指标。

分类的总体精度是用误差矩阵内对角线元素之和除以总的采样个数。例如，在表 3-5 的例子中，该值为  $(45 + 63 + 70)/200 = 89\%$ 。即总体分类精度为 89%。为描述对某一类别的分类精度，我们定义了用户精度和生产者精度 (Story and Congalton, 1986)。

对于类别 A 的生产者精度是用下列公式计算的，即类别 A 的正确分类个数除以对于类别 A 的总采样个数即 A 的列总和。例如在以上的例子中，生产者精度为： $45/50 = 90\%$ 。该指标指出，一个地面采样点被正确地分类的概率。事实上，它是对丢失误差的一个量度，该误差指出了该采样数据中没有被正确分类的百分比。丢失误差是由该类所在列中非对角元素之和除以该列总和而得。例如在以上的例子中，丢失误差为： $(3 + 2)/50 = 10\%$ 。因此，有

$$\text{生产者精度} + \text{丢失误差} = 100\%$$

另一方面，类别 A 的用户精度定义为正确分类为 A 的个数除以分为类别 A 的总和（即 A 所在行的总和）。在以上例子中，该值为  $45/52 = 86.5\%$ 。事实上，该指标指出一个采样分类点表示实际地面真实情况的百分比。用户精度表示了包含误差。A 类的包含误差用 A 所在行的非对角线元素之和除以该行的总和。在以上的实例中，该值为： $(2 + 5)/52 = 13.5\%$ 。用户精度与包含误差有以下关系：

$$\text{用户精度} + \text{包含误差} = 100\%$$

Chrisman(1986)指出，输入 GIS 中数据应附有一个误差矩阵。这应以原始的误差矩阵的形式，而非由该矩阵导出的一系列参数。只有这个原始矩阵才能表示出每一类别的各种精度与误差，用户

可根据其自己的要求从中导出新的参数。

### 3.3.3.3 Kappa 系数

两个最常用的属性精度测量是二维正态概率和 Kappa 系数。由于二维正态概率是基于“正确百分率”因而不能统计出包含与丢失误差。另一方面，Kappa 系数提供两幅图观测协议的不同量度，而协议是由机会形成的 (Congalton and Mead, 1983; Congalton et al., 1983)。Kappa 系数定义为

$$K = (P_o - P_e) / (1 - P_e)$$

这里  $P_o$  是观测精度估计而  $P_e$  是期望精度估计。

一个 0.80 的 Kappa 系数可以解释为该分类以 80% 的程度优于随机地给像素赋类别值。Kappa 系数的优点在于它已经包含了丢失误差和包含误差。一个条件 Kappa 系数可以表示对于某一类别的分类精度 (Campbell, 1987)。Rosenfield (1986) 和 Chrisman (1984) 建议 Kappa 参数应成为表示总体属性不确定性的一个标准指数而条件 Kappa 参数则应成为某一类别精度描述的指标。

误差矩阵是一个常用的分类遥感影像不确定性描述模型。为使对分类精度的描述具有代表性，样本大小及采样模式是两个重要的考虑因素。基于误差矩阵，一系列的误差指标可以被导出如用户精度、生产者精度、包含误差、丢失误差等。这些参数可用于描述某一类或整体分类的精度。然而，在某些情况下需要知道每一个像素的不确定性。以上的误差描述指标不能满足此要求。第四章将详细讨论这方面内容。

### 3.3.3.4 内部与外部检验

现有三种方法可用于确定属性数据的统计质量：演绎法推论、内部检验和外部检验 (Kennedy-Smith, 1986)。通常用演绎法推断属性数据质量是利用具有由内部或外部检验导出的属性质量的量

测值。内部检验方法是通过比较若干相互独立的重复观测量，其平均值被视为“真值”。在质量控制中，内部检验的结果是准确性。

另一方面，外部检验是通过将量测值与“真值”或可以写作“真值”的量进行比较。外部检验的结果得出量测值的精度，外部检验的结果可以满足用户的需求（Kennedy-Smith, 1986），但是这种检验不能区分开各种误差源或过程的误差影响，其结果是包含了各种误差的影响。

在利用外部检验确定属性的数据质量时，我们首先要选定一定检查点。我们可以选择随机抽样点。为使每一类别内均有一定的点被选中为检查点，人们通常建议使用分类随机采样方式。

Hay(1979)建议对总体至少应选择 50 至 100 个采样点，而对于每一类至少应选择 30 个采样点。使用外部检验法确定属性数据质量的过程描述如下。详细描述见 Hord 与 Brooner(1976) 的文章。该过程的目的在于确定真实地图的精度。

① 定义一置信水平（例如 99.7%）从正态分布表中查出此表所对应的值，即  $z_\alpha = 3.0$ 。

② 确定采样个数（ $N$ ）例如  $N = 200$ 。

③ 计算检查点正确分类的百分比（ $P$ ）例如  $P = 89\%$ 。

④ 利用以下不等式确定检验精度（Drummond, 1991）

$$(-z_\alpha^2 - N)x^2 + (z_\alpha^2 + 2NP)x - NP^2 > 0$$

对于  $N = 200$  以及  $z_\alpha = 3.0$  有

$$0.81 < x < 0.95$$

因此我们可以说，在 99.7% 的置信水平，对于 200 个点采样精度为 89% 的情况下，分类精度为 81% 到 95%。很明显，如果减少检查点的个数或升高置信水平，确定的分类区间宽度将加大。

该方法的一个缺点是整体分类精度有可能被拒绝，尽管对某一类的分类精度是可以接受的。

对于某些应用，遥感分类结果应具有一个最小的正确分类的

百分比。这种情况下，假设检验最为适合。预先确定精度的假设检验是一个经常使用的质量控制方法。接受性采样是质量控制的一个重要分支。关于统计质量控制的详细讨论见 Grant 和 Leavenworth(1988)。

除类别属性数据外，另一种是连续属性数据。以下将讨论连续属性数据不确定性的处理方法。

### 3.3.4 连续属性数据的误差传播

由于位置数据与连续属性数据均属连续随机变量。因此，用于连续属性数据不确定性的处理方法与位置不确定性方法类似。因此 前述的误差传播定律 及 Monte Carlo模拟方法也可用于连续属性数据不确定性的处理。但是，我们需在这里重新用公式对现有问题描述，即将坐标的量测值变为某一位置的连续属性值。对于 (3-1) 式中的位置变量  $y_k, x_1, x_2, \dots, x_n$  我们将其变换成在这些位置  $\{s \in R^2\}$  上的属性值 即

$$y_k(s) = g_k(x_1(s), x_2(s), \dots, x_n(s))$$

这里  $x_1(s), x_2(s), \dots, x_n(s)$  是在位置  $s$  上的属性随机变量。问题是根据  $x_1(s), x_2(s), \dots, x_n(s)$  的统计特性确定  $y_k(s)$  的统计特性。

根据以上变换 可以用 3.2 节中讨论位置数据误差传播定律，来处理连续属性误差传播的问题。

到此，讨论了位置数据不确定性与属性数据不确定性的处理方法。但是，可以结合这两种不确定性的方法讨论得甚少。这是本书将要研究的方面之一，这在第四章将对其进行详细的讨论。

## 3.4 本章小结

根据对已有的位置、属性不确定性问题处理的方法、模型的分

析讨论 可以看出 为在 GIS 与遥感环境中结合这两种不确定性需进一步地研究与探讨。

### 3.4.1 地理信息系统中位置不确定性的建模

根据 3.2 节中的讨论,发现为了建立一个面状物体不确定性模型需要建立线段的置信域模型及其概率分布模型 其次 根据所组成线段的不确定性描述 边界 线特征的不确定性 第三 根据上述两个方面的模型,描述面状物体的不确定性模型。

### 3.4.2 遥感数据中属性不确定性的建模

基于采样的检查方法可用于确定遥感影像的整体或某一类分类精度。如果更进一步区分采样点在区域上的差别,面状目标的中心区域或边界区域,可以在一定程度上表示出不确定性的空间分布。但是它不能指出某一像素的不确定性,而这正是在进一步的空间分析中所要求的。对此,将在第四章中提出概率矢量的概念用以描述每一像素的不确定性。连续属性数据不确定性的处理方法与位置不确定性方法基本类似。

### 3.4.3 结合位置与属性不确定性的方法

基于采样的检验方法可指出分类的结果精度。将分类的结果与“真值”相比较 实际上检出的分类误差中包含属性及位置误差。因此,基于采样的方法可检出位置与属性 (PAT) 误差。但这种方法未能表示出每一个像素的 PAT 不确定性。

误差传播定律适用于定量数据,包括对于位置数据或定量属性数据。因此,该方法可用于确定结合位置不确定性和连续属性不确定性。因此,用于确定位置不确定性的误差传播定律也可处理部分的 PAT 不确定性。

然而,在本研究中我们希望结合位置不确定性与类别属性不

确定性，因此，需要发展结合定性与定量数据的方法。这是一个基于多源数据（GIS、遥感和其他类型数据）的决策支持问题。而误差传播定律不能直接用于解决该类问题，需要发展新的模型，或借用已用于其他领域解决类似问题的方法。

## 第四章 “S-带”位置与属性 不确定性的处理模型

### 4.1 概 述

在 GIS 中，一个物体是由其两方面确定的，即位置与属性值。如前面章节所描述，现已发展出若干种描述位置及属性不确定性的模型，但是尚没有模型可以结合这两种不确定性。为了发展这样一个模型，需要进一步扩展现有的位置、属性不确定性模型，进而发展出结合两种不确定性的模型。

本章首先定义位置、属性不确定性相结合的描述指标。接着讨论在两个层次上建立“S-带”模型：一般理论模型和对于特定应用的实际操作模型。为了设计一个实际的“S-带”模型，首先根据概率统计理论导出一条线段的位置不确定性模型，接着讨论基于线段不确定性模型而导出的线性特征和面状物体不确定性模型。为了评估分类遥感影像的属性不确定性，使用了概率矢量的概念，以及由此而导出的参数作为属性不确定性的描述指标。属性与位置不确定性指标被转换成概率值或确定因子。最后将两者结合以达到描述位置与属性不确定性合成之目的。

为此，必须解决以下三个方面的主要问题。

(1) 导出一物体位置误差分布模型。这意味着导出线段上任

意一点的分布，以及在垂直线段方向上的分布；线段的概率分布；线段的置信域；边界线的不确定性；GIS 中面状物体的不确定性；以及定义一组线段的误差指标。

(2) 分类影像的属性不确定性。这包括定义“概率矢量”基于该矢量导出属性误差描述参数，运用这些参数描述属性不确定性。

(3) 位置、属性不确定性的结合模型。该模型的实现涉及定性、定量不确定性指标的结合策略。

## 4.2 位置与属性不确定性

位置与属性 Position and Thematic, PAT指标定义为这样一个不确定性指标，它描述了集成的位置与属性不确定性，即 PAT 不确定性。在描述一个二维的矢量 GIS 中的物体时，我们可以区分两个区域：模糊边界区域；②内部区域。图 4-1 描述了这两个区域。

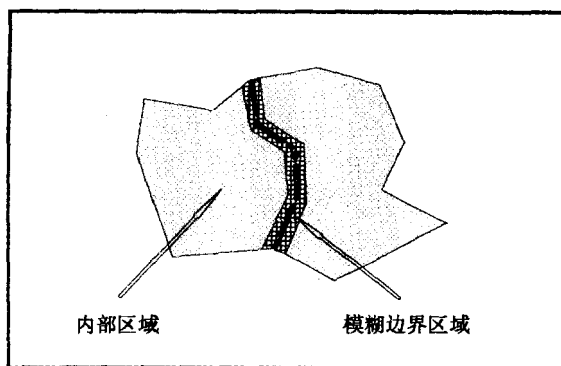


图 4-1 模糊边界与内部区域

一般地来讲，模糊边界内的不确定性比内部区域的不确定性更加复杂。

内部区域与模糊边界的差别源于位置不确定性。模糊边界区域定义为这样的一个区域，在其内，物体的不确定性受到构成物体节点坐标的误差影响，而内部区域则不受到这些节点误差的影响。在内部区域，其不确定性主要来源于属性误差。在模糊边界内，物体受到属性以及位置误差的影响。例如，一个物体是由一层行政管理边界与一层分类遥感影像重叠所生成的，则该物体边界处的不确定性来自以上两个层。当考虑 3.2.2.1 节中所定义的类型-II 线时，模糊边界会更加明显。这里集中讨论模糊边界内的不确定性。在以下的章节中，将导出“S-带”模型用于描述模糊边界内的 PAT 不确定性。

### 4.3 “S-带”的一般理论模型

与通常所用的属性、位置不确定性相分离的模型相比，“S-带”模型的主要思想在模糊边界内将位置、属性不确定性及其关系结合起来表达物体的不确定性。“S-带”模型可用以下公式表达：

$$U_{PAT}(x) = F(U_p(x), U_T(x), \rho(U_p(x), U_T(x))) \quad (4-1)$$

这里  $U_{PAT}(x)$  是  $x$  处的 PAT 不确定性， $x \in \mathbf{R}^2$  是一个二维欧几里德空间的格网或某一模式。 $x$  在索引集合  $D \in \mathbf{R}^2$  内变化 其为模糊边界区域； $U_p(x)$  是  $x$  处的位置不确定性； $U_T(x)$  是  $x$  处的属性不确定性； $\rho(U_p(x), U_T(x))$  是  $x$  处  $U_p(x)$  与  $U_T(x)$  的相互关系。

“S-带”的一般理论模型表达出这样一个概念，即  $x$  处的不确定性与其位置不确定性、属性不确定性及其相互关系有关。其中， $U_p(x)$  是定量数据， $U_T(x)$  可能是定量或定性数据，而  $\rho(U_p(x), U_T(x))$  可能是两者之一。因此，我们不可能定义  $F$  的一般函数形式来包括所有情况， $F$  的具体形式仅当具体问题确定之后才可以定义。

## 4.4 实用‘S-带’模型

根据“S-带”的一般理论模型的指导，我们提出一个实用“S-带”模型发展的过程（见图 4-2）。该发展过程可以是针对以下不确定有关的问题。在一个层内具有行政边界，该层的不确定性来源

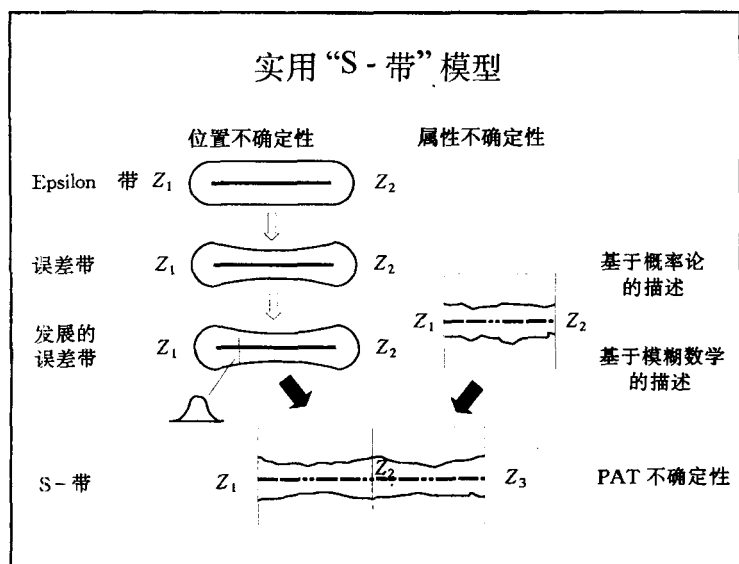


图 4-2 一个实用‘S-带’模型

这里  $Z_1$ 、 $Z_2$ 、 $Z_3$  是确定线段的端点

于位置误差，例如来自数字化的误差。第二层是分类的遥感影像，该类的不确定性来源于分类误差，主要是属性误差。当这两层叠加时，新生成层的不确定性如何描述？实用“S-带”模型的发展将用于解决这个实际问题。

在 4.4.4 节中将介绍一个具体问题。这个问题的本质是属

性、位置不确定性的传播与结合。为了解决这个通用问题，三个方面的问题要予以解决（见图 4-2）。

在 4.4.1 和 4.4.2 节中我们需探讨物体的位置不确定性  $U_p$ 。这里提出一个“进一步发展的误差带”模型用于处理一线段的位置不确定性。该模型有两个方面的新特点：① 以统计的形式定义一个置信域；② 以一个垂直于线段方向的分布模型。基于面状物体所构成线段的不确定模型将导致对两连接线段处不确定性的研究。

分类影像的属性不确定性  $U_T$  将在 4.4.3 节中讨论。由最大似然分类所生成的概率矢量以及由此而导出的一组参数将被用于描述属性不确定性。

在 4.4.4 节中描述的“S-带”模型，将用于结合位置与属性不确定性。对此 我们研究了两个模型，一个是基于概率论中的乘积定理，而另一个是基于具有概率解释的确定因子模型。一个实用“S-带”模型是由实际所需处理的数据所决定的。这些因素包含：各层数据之间的关系，每一层的不确定性本质，以及数据的层次。这里提出的实用“S-带”模型将考虑其中一层含有位置误差（如矢量 GIS 中的线误差），一层含有属性误差（如分类的遥感影像）以及以上两层是相互独立的假设。在 GIS 与遥感集成中，第一步是实现 GIS 中位置不确定性与遥感中属性不确定性的结合。这里假设两种数据误差相互独立的。在解决这一问题的基础上，人们可以进一步地考虑两种数据相关的情况，以及每种数据同时含有位置及属性-误差的情况。

在结合具有位置与属性误差的两层数据之后，在新产生层上误差分布的形态既不是 Epsilon 带所描述的“矩形”也不是误差带所描述的“哑铃形”而是由若干个小“S 形”围绕着力特征所构成。因此 我们在这里称之为“S-带”模型（见图 4-3）。

在位置、属性不确定性可以结合之前，首先需要导出面状物体

的位置不确定性描述。GIS 中的最基本元素是点，两点连接构成一条线段。线特征是由线段构成的。一个面状物体是由其边界线包围所构成的。因此 我们有以下构成面状物体的过程：点、线段、线特征、边界线、面状物体。这个过程及所对应的误差指标表示于图 4-3 中。

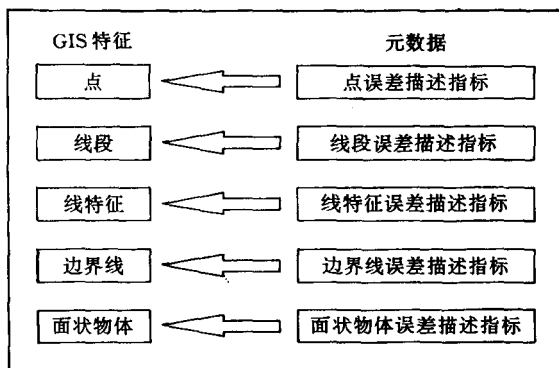


图 4-3 构成面状目标的过程及相应的误差描述指标

点误差已在第三章中讨论过。下面将研究线段、线特征、边界线及面状物体的不确定性模型。

#### 4.4.1 线段的位置不确定性

在测量中，我们已经发展了许多模型用于描述点的不确定性，如 3.2.1 节中所述。描述点的位置不确定性包括各种精度指标（如 RMSE）及误差椭圆。但是线段误差的概率分布仍是一个未解决的问题。该问题比较复杂且根据具体情况而变化。我们只在一定假设下讨论该问题。

#### 4.4.1.1 线段的定义

一条线段是由其两端点  $Z_1$  和  $Z_2$  定义的。 $Z_r$  是线段上  $Z_1$  与  $Z_2$  间的任意一点 (见图 4-4)。其关系由公式 4-4 表达出来。

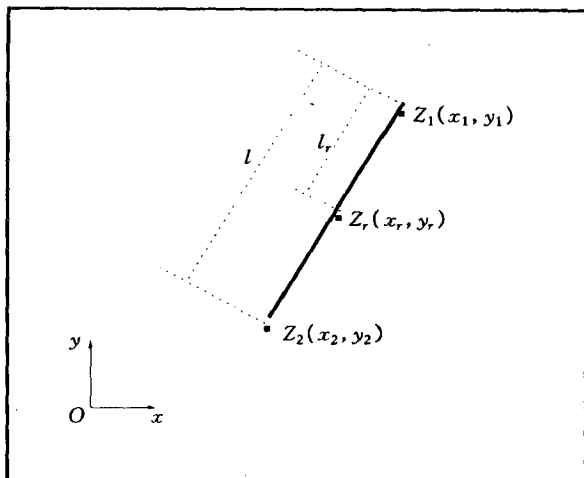


图 4-4 构成一条线段:  $Z_1$  与  $Z_2$  是两端点,  $Z_r$  是线段上的任意一点。  $r = l_r/l$  且  $r \in [0, 1]$

我们假设两量测随机矢量:  $Z_1 = (x_1, y_1)^T$ ,  $Z_2 = (x_2, y_2)^T$  相互独立的且服从以下二维正态分布 ( $N_2$ ):

$$Z_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \sim N_2 \left[ \begin{bmatrix} u_1 \\ v_1 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \right] \quad (4-2)$$

$$Z_2 = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \sim N_2 \left[ \begin{bmatrix} u_2 \\ v_2 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \right] \quad (4-3)$$

这里  $Z_1$  和  $Z_2$  是真实线段两端点  $\zeta_1 = (\mu_1, \nu_1)$  与  $\zeta_2 = (\mu_2, \nu_2)$  的估计量。一条线段可由其上任意一点连续地从一端移动到

另一端来表示。任意一点可用下式来表示：

$$\begin{aligned}\zeta_r &= (1-r)\zeta_1 + r\zeta_2 = \begin{bmatrix} (1-r)\mu_1 + r\mu_2 \\ (1-r)\nu_1 + r\nu_2 \end{bmatrix} \\ r &\in [0,1]\end{aligned}\quad (4-4)$$

#### 4.4.1.2 量测线段上任意点的分布

在这里研究线段的不确定性时，首先排除由于用一系列的线段去近似地表达一条线特征（通常是曲线）而引起的误差。我们仅考虑由于两端点的随机误差而造成的线段的误差，由于  $Z_r$  是两端点  $Z_1$  和  $Z_2$  的线性函数并且假设  $Z_1$  和  $Z_2$  独立且服从二维正态分布。因此，可以得出： $Z_r$  也服从二维正态分布。该分布的描述统计参数可以从  $Z_1$  和  $Z_2$  的分布参数中导出。进一步地，我们还感兴趣于该点 ( $Z_r$  在垂直于线段  $Z_1 Z_2$  方向上的分布。在坐标旋转之后，它可以描述成一个边缘分布。以下给出证明的详细过程。

##### (1) 期望值、方差与协方差

假设参数  $\sigma_{11}, \sigma_{22}$  以及  $\sigma_{12} = \sigma_{21}$  是可以估计的，根据 4-4 式，“量测”线段上的任意一点为

$$Z_r = (1-r)Z_1 + rZ_2 = \begin{bmatrix} (1-r)X_1 + rX_2 \\ (1-r)Y_1 + rY_2 \end{bmatrix} \quad (4-5)$$

$X_r$  的期望值为

$$\begin{aligned}E(X_r) &= E[(1-r)X_1 + rX_2] = (1-r)E(X_1) + rE(X_2) \\ &= (1-r)\mu_1 + r\mu_2 = \mu_r\end{aligned}\quad (4-6)$$

这里  $E(X_r)$  是  $X_r$  的期望值， $\mu_1$  是  $X_1$  的期望值， $\mu_2$  是  $X_2$  的期望值。类似地，有

$$E(Y_r) = (1-r)\nu_1 + r\nu_2 = \nu_r \quad (4-7)$$

这里  $E(Y_r)$  是  $Y_r$  的期望值， $\nu_1$  是  $Y_1$  的期望值， $\nu_2$  是  $Y_2$  的期望值。

方差与协方差的推导是根据误差传播定律进行的 [见 3.2.1. 2 节中 1) 部分]。  $X_r$  的方差  $\sigma_{xx}$  源于式 (4-5) 和式 (4-2)。为

$$\sigma_{xx} = (1-r)^2 \sigma_{11} + r^2 \sigma_{11} = ((1-r)^2 + r^2) \sigma_{11}$$

类似地，可以导出  $Y_r$  的方差  $\sigma_{yy}$  为

$$\sigma_{yy} = [(1-r)^2 + r^2] \sigma_{22}$$

$X_r$  与  $Y_r$  的协方差导出如下：

$$\begin{aligned} \sigma_{xy} &= E\{(x_r - \mu_r)(y_r - \nu_r)\} \\ &= E\{[(1-r)X_1 + rX_2] - [(1-r)\mu_1 + r\mu_2]\} \\ &\quad \{[(1-r)Y_1 + rY_2] - [(1-r)\nu_1 + r\nu_2]\} \\ &= E\{[(1-r)(X_1 - \mu_1) + r(X_2 - \mu_2)] \cdot \\ &\quad [(1-r)(Y_1 - \nu_1) + r(Y_2 - \nu_2)]\} \\ &= E[(1-r)^2(X_1 - \mu_1)(Y_1 - \nu_1)] \\ &\quad + E[r^2(X_2 - \mu_2)(Y_2 - \nu_2)] \\ &\quad + E[(1-r)r(X_1 - \mu_1)(Y_2 - \nu_2)] \\ &\quad + E[(1-r)r(X_2 - \mu_2)(Y_1 - \nu_1)] \end{aligned}$$

因为  $Z_1$  和  $Z_2$  是相互独立的，有

$$(1-r)rE[(X_1 - \mu_1)(Y_2 - \nu_2)] = 0$$

$$(1-r)rE[(X_2 - \mu_2)(Y_1 - \nu_1)] = 0$$

根据 4-2 式和 4-3) 式有

$$E[(X_1 - \mu_1)(Y_1 - \nu_1)] = E[(X_2 - \mu_2)(Y_2 - \nu_2)] = \sigma_{12}$$

因此， $X_r$  与  $Y_r$  的协方差  $\sigma_{xy}$  为：

$$\sigma_{xy} = ((1-r)^2 + r^2) \sigma_{12} = \sigma_{yx} \quad (4-8)$$

根据 4-6) 式和 4-8) 式有

$$Z_r \sim N_2 \left[ \begin{bmatrix} (1-r)\mu_1 + r\mu_2 \\ (1-r)\nu_1 + r\nu_2 \end{bmatrix}, (1-r)^2 + r^2 \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \right] \quad (4-9)$$

根据前面的假设，我们可以说量测线段  $Z_1 Z_2$  上的任意一点  $Z_r$  服从式 (4-9) 所定义的二维正态分布。

(2) 任意点在线段垂直方向上的分布

线段上任意一点在  $y'$  方向上的密度函数可以很容易地用  $y'$  方向上的边缘分布密度函数表示出来：

$$f'_{y'}(y') = \int_{-\infty}^{+\infty} f'(x', y') dx'$$

这里  $x'$  和  $y'$  为任意点  $Z_r$  在旋转坐标系  $OX'Y'$  中的坐标。这里，我们引入旋转坐标系，使  $y'$  与线段自身方向垂直（见图 4-5）。 $f'_{y'}(y')$  是  $z'_r$  在旋转坐标系内密度函数。

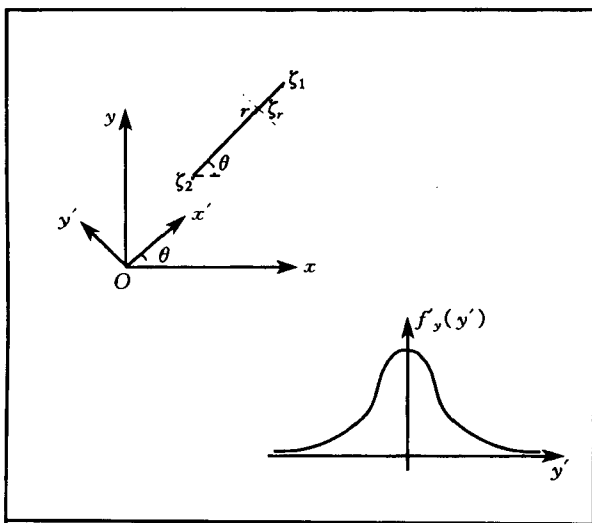


图 4-5 坐标系从  $Oxy$  旋转  $Ox'y'$  沿  $y'$  方向的误差服从一维正态分布

由于  $Oxy$  到  $Ox'y'$  的变换

$$Z'_r = DZ_r$$

$$D = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

是线性的,  $Z'_r$  也服从二维正态分布 (与  $Z_r$  类似) 其在  $y'$  方向上的边缘分布是一个一维正态分布。根据误差传播定律和 (4.9) 式,  $Z'_r$  的期望值为

$$\begin{aligned} E(Z'_r) &= \begin{bmatrix} \mu'_{r'} \\ \nu'_{r'} \end{bmatrix} = DE(Z_r) \\ &= \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} (1-r)\mu_1 + r\mu_2 \\ (1-r)\nu_1 + r\nu_2 \end{bmatrix} \end{aligned}$$

且  $Z'_r$  的协方差阵为

$$\begin{aligned} \Sigma' &= D\Sigma D^T = \begin{bmatrix} \sigma_{x'} & \sigma_{x'y'} \\ \sigma_{y'x'} & \sigma_{y'} \end{bmatrix} \\ &= \begin{bmatrix} \sigma'_{11} & \sigma'_{12} \\ \sigma'_{21} & \sigma'_{22} \end{bmatrix} ((1-r)^2 + r^2) \\ &= \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \\ &\quad \cdot ((1-r)^2 + r^2) \end{aligned}$$

由于  $Z'_r$  是  $Z_r$  的线性变换 而  $Z_r$  服从一个二维正态分布, 因此  $Z'_r$  也服从一个二维正态分布。其密度函数为

$$\begin{aligned} f(Z') &= \frac{1}{2\pi |\Sigma'|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (Z' \right. \\ &\quad \left. - E(Z'_r))^T (\Sigma')^{-1} (Z' - E(Z'_r)) \right) \end{aligned} \quad (4-10)$$

这里  $Z' = [x' \ y']^T$  在  $y'$  方向上的边缘概率分布的密度函数为

$$f'_{y'}(y') = \int_{-\infty}^{+\infty} f(x', y') dx'$$

$$= \frac{1}{(2\pi\sigma_{y'})^{\frac{1}{2}}} \exp[-(y' - \nu_{y'})^2 / 2\sigma_{y'}] \quad (4-11)$$

这里

$$y' = -\sin(\theta)[(1-r)X_1 + rX_2] + \cos(\theta)[(1-r)Y_1 + rY_2]$$

$$\nu_{y'} = -\sin(\theta)[(1-r)\mu_1 + r\mu_2] + \cos(\theta)[(1-r)\nu_1 + r\nu_2]$$

$$A = \cos(\theta)\sigma_{21} - \sin(\theta)\sigma_{11}$$

$$B = \cos(\theta)\sigma_{22} - \sin(\theta)\sigma_{12}$$

$$\sigma_{y'} = [A(-\sin(\theta)) + B(\cos(\theta))][(1-r)^2 + r^2]$$

由于线段垂直于  $y'$  轴方向,  $f'_{y'}$  ( $y'$  即为任意点  $Z$ , 在垂直于线段  $Z_1Z_2$  方向上分布的密度函数。

#### 4.4.1.3 线段 $Z_1Z_2$ 的概率分布

线段  $Z_1Z_2$  是由四个随机变量, 即  $Z_1$  和  $Z_2$  的坐标所定义的。因此,  $Z_1Z_2$  的随机特性即由这四个随机变量所决定。由于随机的特性, 量测线段是围绕其真值  $\zeta_1\zeta_2$  分布的。线段的分布可以分为三个区域:  $A$ 、 $B$  和  $C$  (见图 4-6 所示)。区域  $A$  内的分布是由  $Z_1$  决定的;  $C$  内的分布由  $Z_2$  决定, 而  $B$  内的分布是由  $Z_1$  和  $Z_2$  所决定的。因此, 线段  $Z_1Z_2$  的分布可由这三个区域内的分布所描述。

##### (1) 线段 $Z_1Z_2$ 在 $B$ 子区域内的分布

线段  $Z_1Z_2$  可由其上的任意点  $Z_r$  ( $r \in [0, 1]$ ) 的集合来表示。因此, 线段的分布可由其上的任意点的分布 (对于所有  $r \in [0, 1]$ ) 来表示。而我们特别感兴趣的垂直于线段方向上的分布。

对于  $r \in [0, 1]$ , 已经获得其线段垂直分布, 由 (4-11) 式表示。基于该式, 我们知道, 线段  $Z_1Z_2$  在垂直于  $\zeta_1\zeta_2$  方向上的密度函数可以表达出来。

事实上, 这里定义了一组一维正态分布, 当旋转角  $\theta$  一定时, 这些分布是随  $r$  ( $r \in [0, 1]$ ) 而变化的。为了研究在垂直方向方差

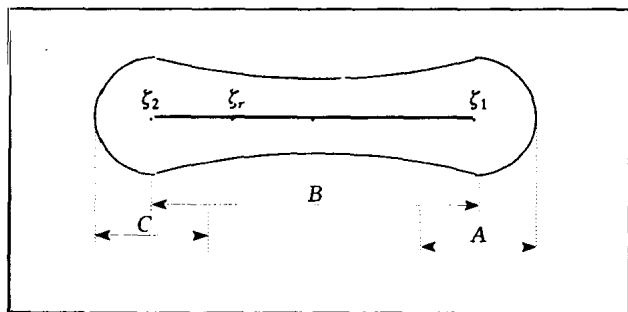


图 4-6 将线段  $\zeta_1 \zeta_2$  的分布分成 A、B、C 三个区来描述

的变化,现将  $\sigma_Y(r)$  改写为以下形式

$$\sigma_Y(r) = \sigma_0[(1-r)^2 + r^2] \quad (4-12)$$

这里

$$\begin{aligned} \sigma_0 = & [(\cos(\theta)\sigma_{21} - \sin(\theta)\sigma_{11})[-\sin(\theta)] \\ & + [\cos(\theta)\sigma_{22} - \sin(\theta)\sigma_{12}][\cos(\theta)] \end{aligned}$$

现在讨论方差  $\sigma_Y(r)$  的变化,设

$$(a) \quad R = (1-r)^2 + r^2$$

由于

$$(b) \quad 0 \leq r \leq 1$$

因此容易看出

$$(c) \quad R > 0 \text{ 且}$$

$$(d) \quad 0 \leq (1-r) \leq 1$$

将 (b) 和 (d) 相乘,并且 2 倍不等式两侧,有

$$\begin{aligned} 0 & \leq 2r(1-r) \leq 2 \\ -1 & \leq 1 - 2(r(1-r)) \leq 1 \end{aligned}$$

$$(e) \quad R = (1-r)^2 + r^2 \leq 1$$

根据 (c) 和 (e) 有

$$(f) \quad 0 \leq R \leq 1$$

容易看出 当  $r=1$  或  $0$  时  $R$  取其最大值 即  $R=1$ 。设

$$(g) \quad \partial R / \partial r = 0$$

则

$$(h) \quad r = 1/2$$

也即 当  $r=1/2$  时  $R$  取最小值。因此可以概括如下： $\sigma_y(r)$  在其两端点处 即  $r=0$  或  $1$  时 取最大值 而在中间处 ( $r=1/2$  时) 取最小值。垂直分布曲线见图 4-7。

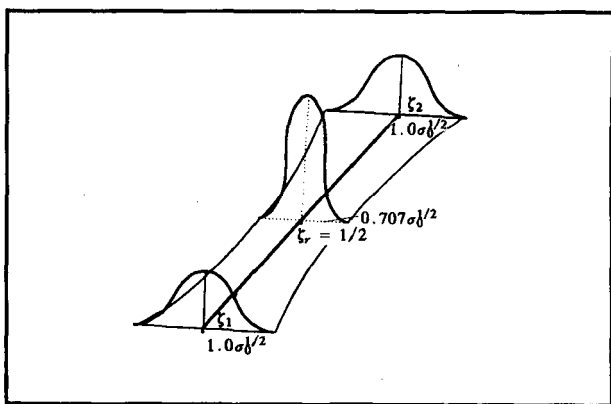


图 4-7 线段  $Z_1 Z_2$  在子区  $B$  内的分布曲线

## (2) $Z_1 Z_2$ 在子区 $A$ 内的分布

在段  $Z_1 Z_2$  在子区  $A$  内的分布是由点  $Z_1$  确定的。为了与  $B$  区内的分布保持一致，使用  $Z'_1$  分布，它是  $Z_1$  通过转换后在  $OX'Y'$  坐标系内相应的值。在此情况下，不仅需要考虑垂直于该线段方向上的分布，而且也需要考虑平行于该线段上的分布。因此，在这里将使用二维分布。其概率密度函数由 (4-10) 式表示 此时  $r=0$  即

$$f_1(Z') = \frac{1}{2\pi |\Sigma'_1|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (Z' - E(Z'_1))^T \cdot (\Sigma'_1)^{-1} \cdot (Z' - E(Z'_1)) \right) \quad (4-13)$$

这里

$$Z_1 = \begin{bmatrix} x'_1 \\ y'_1 \end{bmatrix}$$

$$\Sigma'_1 = \begin{bmatrix} \sigma'_{11} & \sigma'_{12} \\ \sigma'_{21} & \sigma'_{22} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

$$\cdot \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

$$E(Z'_1) = \begin{bmatrix} \mu'_1 \\ \nu'_1 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \mu_1 \\ \nu_1 \end{bmatrix}$$

(3) 线段  $Z_1 Z_2$  在子区  $C$  内的分布

用相似的方法，可以确定线段  $Z_1 Z_2$  在子区  $C$  内的分布。其主要是由  $Z_2$  的分布决定的。概率密度函数为

$$f_2(Z') = \frac{1}{(2\pi) |\Sigma'_1|^{\frac{1}{2}}} \exp \left[ -\frac{1}{2} (Z' - E(Z'_2))^T \cdot (\Sigma'_2)^{-1} \cdot (Z' - E(Z'_2)) \right] \quad (4-14)$$

这里

$$Z_2 = \begin{bmatrix} X'_2 \\ Y'_2 \end{bmatrix}$$

$$\Sigma'_1 = \begin{bmatrix} \sigma'_{11} & \sigma'_{12} \\ \sigma'_{21} & \sigma'_{22} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

$$\cdot \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

$$E(Z'_2) = \begin{bmatrix} \mu'_2 \\ \nu'_2 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \mu_2 \\ \nu_2 \end{bmatrix}$$

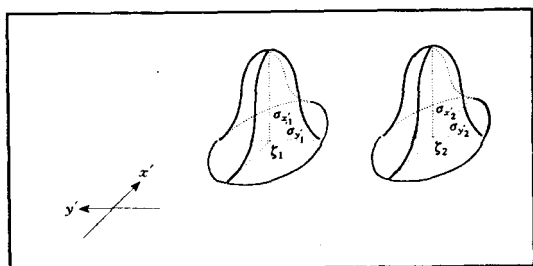


图 4.8 线段  $Z_1, Z_2$  在子区  $A$  和  $C$  内的概率分布

线段  $Z_1, Z_2$  的分布是由其在  $A, B$  和  $C$  区域内的分布所构成的。概率“分布曲面”是由两个二维正态分布 [由式 (4-13) 和 (4-14) 表示] 及一组一维正态分布 [由 (4-11)] 表示, 对所有的  $r \in$

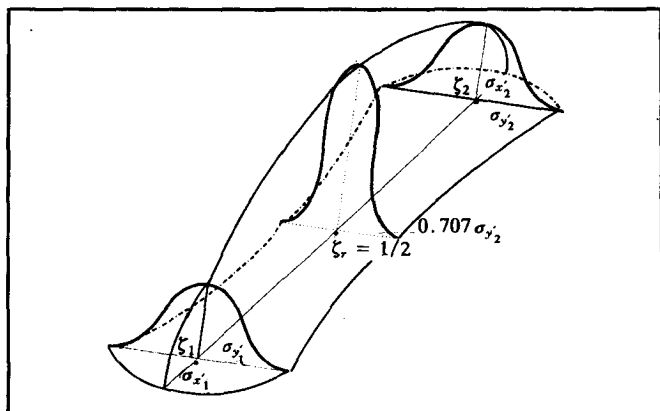


图 4.9 线段  $Z_1, Z_2$  的概率分布

$[0,1]$ 。该图形曲面由图 4-9 表示出来。

线段的概率分布描述了由四个随机变量所构成的随机矢量围绕其“真值” $\zeta_1 \zeta_2$  的分布状况。该模型给出了线段概率分布的解析表达式而 Dutton(1992)用 Monte Carlo 模拟技术对此作了模拟实验,两者在结论方面是相似的。利用所导出的解析表达式,可以定义若干描述线段不确定性的误差指标。理论上,我们也可以用该概率分布解析表达式解决定量地表达一个点属于某一面状目标的概率。如果对线段的真值有一个良好的估计,可以较好地描述出围绕量测线段  $Z_1 Z_2$  的一个不确定区域,并且对区域中任何一点赋予一个概率值以描述其不确定性。然而在一个 GIS 中,对一条线段我们仅有一次观测量,而不是多个观测量。因此,对于点属于一个面状物体仅可给出一个近似描述值。另一方面,若可以导出一个围绕量测线段的误差区域,线段的真值以某一概率落在该区域内。该结果将可以直接在现有的 GIS 所使用。我们称这个区域为线段的“置信区域”。有关公式的推导将在 4.4.1.5 节与 4.4.1.6 节中讨论。

#### 4.4.1.4 基于概率分布的线段误差指标

在第三章中,分析与回顾了线性与圆形误差的若干指标。根据线段的概率分布,现在可以定义对应的线段不确定性的误差指标。我们称之为线段的精度表达式有:线段标准差、线段均方根差、线段接近确定误差、线段地图精度标准以及可能线段误差。在定义线段的误差指标时,主要考虑线段  $Z_1 Z_2$  在子区域  $B$  中的概率分布。引入这个简化是因为:与子区  $B$  相比,区域  $A$  和  $C$  相对较小,尤其是线段  $Z_1 Z_2$  较长时更为显著;其次,若考虑子区  $A$  和  $C$  的分布,则误差指标的定义将过于复杂。因此,对线段误差指标的定义主要依赖于公式(4-11)和(4-12)。

##### (1) 线段标准差 $[\sigma_r, r]$

对于所有的  $r \in [0, 1]$  函数  $\sigma_s(r)$  定义为线段标准差。它是描述线段分布的基本统计参数。根据式 (4-12) 线段标准差描述如下

$$\sigma_s(r) = [\sigma_0((1-r)^2 + r^2)]^{\frac{1}{2}} \quad (4-15)$$

这里

$$\begin{aligned} \sigma_0 = & (\cos(\theta)\sigma_{21} - \sin(\theta)\sigma_{11})(-\sin(\theta)) \\ & + (\cos(\theta)\sigma_{22} - \sin(\theta)\sigma_{12})(\cos(\theta)) \end{aligned}$$

对于所有的  $r \in [0, 1]$ 。当两端点的统计参数及旋转角  $\theta$  固定时,  $\sigma_0$  为一个常量值。 $\sigma_s(r)$  随  $r$  而变化, 当在两端点处  $r=0$  或  $1$  时,  $\sigma_s(r)$  取其最大值 ( $\sigma_0^{\frac{1}{2}}$ ), 而在中间处 ( $r = \frac{1}{2}$ ) 取最小值 ( $(\sigma/2)^{\frac{1}{2}}$ )。沿着线段  $\zeta_1, \zeta_2$  的两侧以  $\sigma_s(r)$  的距离连续地对于  $r \in [0, 1]$  画出一个区域, 其所占的面积为总分布曲面所占面积的

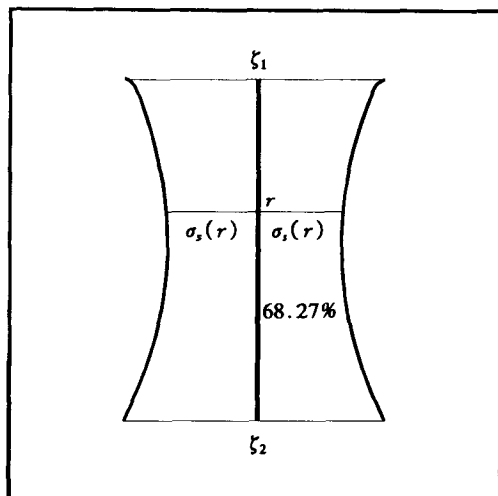


图 4-10 线段标准差的区域及边界

68.26% 见一维正态分布表 (Snedecor and Cochran, 1989)。图 4-10 表示出该结果。线段标准差的另一个名称为线段均方根差 (RM-SES( $r$ ))。

## (2) 线段可能误差 [SPE( $r$ )]

线段标准差描述了一个占总面积的 68% 的区域。与线段标准差相类似, 线段可能误差描述了一个占总区域 50% 面积的区域。根据一维正态分布表 (Snedecor and Cochran, 1989) 它可以表示为

$$SEP(r) = 0.675\sigma_s(r) \quad (4-16)$$

线段可能误差区域平衡“内/外概率”即一量测线段有一半机会落在 SEP( $r$ ) 区域外部, 一半机会落在其内部。

## (3) 线段地图精度标准 [SMAS( $r$ )]

与圆形地图精度标准类似, 线段地图精度标准描述了一个 90% 的分布区域。根据一维正态分布表, 我们获得

$$SMAS(r) = 1.645\sigma_s(r) \quad (4-17)$$

## (4) 线段接近确定误差

线段接近确定误差利用“分布曲面函数”描述了一个占总面积 99.74% 的区域。根据一维正态分布表, 我们有

$$SNCE(r) = 3.0\sigma_s(r) \quad (4-18)$$

以上定义的线段误差指标总结于表 4-1 中。通过将表 4-1 与表 3-1 和表 3-2 相比较, 可以看出这里新定义的线段误差指标是对线性和圆形误差指标的一般化。从线性误差、圆形误差到线段误差指标的发展过程见图 4-11。本研究从一维线性误差指标、二维圆形误差指标发展到四维线段误差指标。此发展过程是由于 GIS 的实用发展中, 对线段不确定性描述的要求而引起的。

线段误差指标提供了一系列描述线段精度的度量值。例如, 为了表示 GIS 中边界图的精度, 可以由 (4-15) 式计算  $\sigma_0$  之后可以选择 5 个误差指标中的一个, 例如 SPE( $r$ ) 对其描述。对于每一条

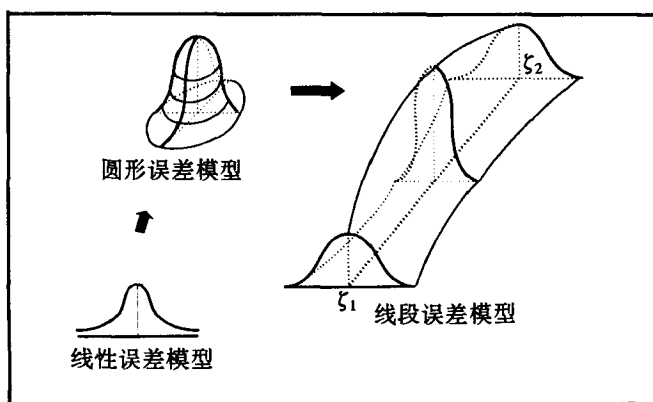


图 4-11

表 4-1 线段误差指标

线段误差			
名称	符号	$P/(\%)$	公式
线段标准误差	$\sigma_s(r)$	68.25	(4-15)
线段可能误差	$SPE(r)$	50	(4-16)
线段地图精度标准	$SMAS(r)$	90	(4-17)
线段接近确定误差	$SNCE(r)$	99.74	(4-18)
线段均方根误差	$RMSES(r)$	68.26	(4-15)
数据	线段		

线段我们计算  $SPE(r)$ ，并将其附于该线段上作为元数据。例如一条线段的单一精度指标 可以计算出来 即  $SPE(r)$  所占面积与线段长度的比例（见图 4-10）。

#### 4.4.1.5 一条线段的置信域

##### (1) 一条线段的置信区域（推导之一）

围绕着一量测线段，我们需要建立这样一个区域使之包含

真实线段  $\zeta_1, \zeta_2$ 。即所有的点  $\zeta_r$  (见 4.4.1.3 节) 同时都以大于某一给定置信水平  $\gamma$  的概率包含于该区域内 即

$$P(\zeta_r \in J, 0 \leq r \leq 1) \geq \gamma \quad (4-19)$$

这里 选择  $\nu = 90\%$  即有 10% 的风险  $\zeta_r$  落在区域之外。

现在对于期望值  $\{(1-r)\mu_1 + r\mu_2\}$  和  $\{(1-r)\nu_1 + r\nu_2\}$  建立其置信区间。对于一个确定的  $r$ ,  $\{(1-r)\mu_1 + r\mu_2\}$  的置信区间可以通过对标准化的随机变量  $\{(1-r)x_1 + rx_2\}$  的处理开始 因此有

$$\frac{(1-r)x_1 + rx_2 - [(1-r)\mu_1 + r\mu_2]}{[\{(1-r)^2 + r^2\}\sigma_{11}]^{\frac{1}{2}}} \sim N(0,1)$$

假设  $X_1$  和  $X_2$  的方差  $\sigma_{11}$  是已知的。我们所感兴趣的是同时获得所有  $\mu_r (r \in [0,1])$  的置信域 而不是线段上某一特定点的置信区域。如果线段上的每一点同时落在置信区域  $J$  内 则整条线段则落在置信域内。因此, 将构造这样一个区域, 所有点 (对于所有的  $r \in [0,1]$ ) 以大于某一预先给定的置信水平的概率同时落在该区域内。

为了获得对于所有  $r \in [0,1]$  同时满足的表达式 引入 Cauchy-Schwarz 不等式。对于矢量  $\mathbf{a} = (a_1, \dots, a_2)^T$  和  $\mathbf{b} = (b_1, \dots, b_2)^T$ , Cauchy-Schwarz 不等式 (EGHM, 1977) 可以写成以下形式

$$\left| \sum_{i=1}^P a_i b_i \right| \leq \left( \sum_{i=1}^P (a_i)^2 \right)^{\frac{1}{2}} \left( \sum_{i=1}^P (b_i)^2 \right)^{\frac{1}{2}}$$

基于 Cauchy-Schwarz 不等式 有

$$\begin{aligned} & |(1-r)x_1 + rx_2 - (1-r)\mu_1 - r\mu_2| \\ &= |(1-r)(x_1 - \mu_1) + r(x_2 - \mu_2)| \\ &\leq [(x_1 - \mu_1)^2 + (x_2 - \mu_2)^2]^{\frac{1}{2}} [(1-r)^2 + r^2]^{\frac{1}{2}} \end{aligned}$$

在不等式的两边同时除以  $[\{(1-r)^2 + r^2\}\sigma_{11}]^{\frac{1}{2}}$  有

$$\frac{|(1-r)x_1 + rx_2 - (1-r)\mu_1 - r\mu_2|}{[\{(1-r)^2 + r^2\}\sigma_{11}]^{\frac{1}{2}}}$$

$$\leq \left[ \frac{(x_1 - \mu_1)^2}{\sigma_{11}} + \frac{(x_2 - \mu_2)^2}{\sigma_{11}} \right]^{\frac{1}{2}} \quad (4-20)$$

该式的右边不依赖于  $r$ ，因此它对于整条线段适用。由于  $x_1$  和  $x_2$  是相互独立的，并且服从正态分布有

$$\frac{(x_1 - \mu_1)^2}{\sigma_{11}} + \frac{(x_2 - \mu_2)^2}{\sigma_{11}} \sim \chi^2_2 \quad (4-21)$$

这就是说  $(x_1 - \mu_1)^2/\sigma_{11} + (x_2 - \mu_2)^2/\sigma_{11}$  满足两个自由度的 Chi-平方分布。现在来建立一个区域，使式 (4-19) 成立，例如对于  $\gamma = 0.90$  的情况。首先对  $(1-r)\mu_1 + r\mu_2$  和  $(1-r)\nu_1 + \nu_2$  分别建立置信区间，并使之同时成立。先考虑 95% 的概率，取该值的原因稍后就会清楚。从 (4-21) 式可以得出

$$P \left[ \frac{(x_1 - \mu_1)^2}{\sigma_{11}} + \frac{(x_2 - \mu_2)^2}{\sigma_{11}} \leq 5.99 \right] = 0.95 \quad (4-22)$$

这里 5.99 是从一个 Chi-平方分布表中查得的 (Li and Tao, 1982)，在两个自由度和 95% 的置信水平下获得的。应用不等式 (4-20) 有

$$P \left[ \frac{|(1-r)x_1 + rx_2 - (1-r)\mu_1 - r\mu_2|}{[(1-r)^2 + r^2]\sigma_{11}} \leq (5.99)^{\frac{1}{2}}, \text{ 对所有 } r \right] \geq 0.95$$

该式等价于

$$P[x_r - c \leq \mu_r \leq x_r + c, \text{ 对所有 } r] \geq 0.95$$

这里

$$c = (5.99)^{\frac{1}{2}} [((1-r)^2 + r^2)\sigma_{11}]^{\frac{1}{2}}$$

至此，建立了对于所有  $\mu_r = (1-r)\mu_1 + r\mu_2$  ( $r \in [0, 1]$ ) 同时成立置信区间。类似地，亦可对所有参数  $\nu_r = (1-r)\nu_1 + r\nu_2$  ( $r \in [0, 1]$ ) 同时成立的置信区间，因而有

$$P[y_r - d \leq \nu_r \leq y_r + d \text{ 对所有 } r] \geq 0.95$$

其中

$$d = (5.99)^{\frac{1}{2}} [((1-r)^2 + r^2)\sigma_{22}]^{\frac{1}{2}}$$

现在引入 Kolmogorov 定理 (Rao, 1965) 来构成联合置信区间。  
如是对于随机事件  $A_1, \dots, A_k$ , 下列不等式成立 即

$$P(A_1 \cup \dots \cup A_k) \leq P(A_1) + \dots + P(A_k) \quad (4-23)$$

即任意随机事件  $A_1$  或...或  $A_k$  发生的概率, 将小于所有单独事件发生概率之和。对于  $k=2$  的情况,  $A_1$  为事件

$$\frac{(x_1 - \mu_1)^2}{\sigma_{11}} + \frac{(x_2 - \mu_2)^2}{\sigma_{11}} \geq 5.99$$

并且  $A_2$  为事件

$$\frac{(y_1 - \nu_1)^2}{\sigma_{22}} + \frac{(y_2 - \nu_2)^2}{\sigma_{22}} \geq 5.99$$

$A_1$  事件的余事件为  $\neg A_1$ 。它可以定义为

$$\frac{(x_1 - \mu_1)^2}{\sigma_{11}} + \frac{(x_2 - \mu_2)^2}{\sigma_{11}} \leq 5.99$$

$A_2$  事件的余事件为  $\neg A_2$ 。其可以定义为

$$\frac{(y_1 - \nu_1)^2}{\sigma_{22}} + \frac{(y_2 - \nu_2)^2}{\sigma_{22}} \leq 5.99$$

根据  $A_1, \neg A_1, A_2, \neg A_2$  事件的表达式, (4-22) 式可以写为

$$P(\neg A_1) = 0.95$$

因此

$$P(A_1) = 0.05$$

类似地

$$P(\neg A_2) = 0.95$$

因此

$$P(A_2) = 0.05$$

在两个事件的情况下, (4-23) 式可以写成

$$P(A_1 \cup A_2) \leq P(A_1) + P(A_2) \quad (4-24)$$

因此有

$$P(A_1 \cup A_2) \leq P(A_1) + P(A_2) = 0.05 + 0.05 = 0.10$$

根据余集的定义有

$$P(\neg(A_1 \cup A_2)) > 1 - 0.10 = 0.90$$

根据 DeMorgan 定律 (Klir and Folger, 1988) 有

$$P(\neg(A_1 \cup A_2)) = P((\neg A_1) \cap (\neg A_2))$$

因此

$$P((\neg A_1) \cap (\neg A_2)) \geq 0.90$$

现在、我们可以得出不等式

$$\frac{(x_1 - \mu_1)^2}{\sigma_{11}} + \frac{(x_2 - \mu_2)^2}{\sigma_{11}} \leq 5.99 \quad (4-25)$$

和

$$\frac{(y_1 - \nu_1)^2}{\sigma_{22}} + \frac{(y_2 - \nu_2)^2}{\sigma_{22}} \leq 5.99 \quad (4-26)$$

以大于 90% 的概率同时成立 即

$$P\left[\frac{(x_1 - \mu_1)^2}{\sigma_{11}} + \frac{(x_2 - \mu_2)^2}{\sigma_{11}} \leq 5.99\right]$$

$$\text{且} \quad \left[\frac{(y_1 - \nu_1)^2}{\sigma_{22}} + \frac{(y_2 - \nu_2)^2}{\sigma_{22}} \leq 5.99\right] > 0.90$$

这时我们比较清楚为什么在证明开始阶段假设 0.95 的概率是为了导出 0.90 的总体置信区间。因为 (4-14) 和 (4-16) 不等式 (4-25) 意味着

$$x_r - c \leq \mu_r \leq x_r + c \quad (4-27)$$

对于所有的  $r$ 。

同样地，不等式 4-26 意味着

$$y_r - d \leq \nu_r \leq y_r + d \quad (4-28)$$

对于所有的  $r$ 。

因此有

$$P \{ x_r - c \leq \mu_r \leq x_r + c \quad \text{且} \\ y_r - d \leq \nu_r \leq y_r + d \} > 0.90$$

现在定义区域  $J_r$  为一个由点  $(x, y)^T$  所构成的集合, 而且  $x$  满足

$$x_r - c \leq x \leq x_r + c$$

$y$  满足

$$y_r - d \leq y \leq y_r + d$$

可以得出

$$P(\zeta_r \in J_r \text{ 对于所有 } r \in [0, 1]) > 0.90 \quad (4-29)$$

这里  $\zeta_r = (\mu_r, \nu_r)$ 。因此, 我们获得了一个区域  $J$  满足 (4-19)。  $J$  为  $J_r$  的并集,  $r \in [0, 1]$ 。

以上的推导给出一条线段的置信区域。在推导的过程中, 是以 0.90 的置信水平为例。现在总结出置信域的一般结果如下。

置信区域  $J$  是这样一个区域, 线段上所有点  $\zeta_r$  ( $r \in [0, 1]$ ) 均以大于预先指定的置信水平 ( $\gamma$ ) 的概率被包含在区域内, 即

$$P(\zeta_r \in J_r, \text{ 对所有 } r \in [0, 1]) > \gamma \quad (4-30)$$

$J$  是所有  $J_r$  的并集,  $r \in [0, 1]$ 。  $J_r$  是这样一个  $(x, y)^T$  的点集 其中  $x$  满足

$$x_r - c \leq x \leq x_r + c \quad (4-31)$$

而且  $y$  满足

$$y_r - d \leq y \leq y_r + d \quad (4-32)$$

其中

$$c = K^{\frac{1}{2}} [((1-r)^2 + r^2) \sigma_{11}]^{\frac{1}{2}}$$

$$d = K^{\frac{1}{2}} [((1-r)^2 + r^2) \sigma_{22}]^{\frac{1}{2}}$$

式中参数  $k$  与所选定的置信水平  $\gamma$  有关 可以从一个 Chi-平方分布表中查到,  $K = k_{2i(1-r)/2}^2$  例如,  $\gamma = 0.90, (1-r)/2 = 0.95, K = 5.99$ 。

为了分析置信域的形状, 对以下两式求偏导

$$c = K^{\frac{1}{2}} [((1-r)^2 + r^2) \sigma_{11}]^{\frac{1}{2}}$$

和

$$d = K^{\frac{1}{2}} [((1-r)^2 + r^2) \sigma_{22}]^{\frac{1}{2}}$$

当  $\partial c / \partial r = 0$  和  $\partial d / \partial r = 0$  的条件下 可以求得  $r = \frac{1}{2}$ 。另外两个极点是  $r = 0$  和  $1$ 。这意味着置信区域在中间部分取得最小值 ( $r = \frac{1}{2}$ ) 而在两端点处 ( $r = 0$  或  $1$ ) 取得最大值。置信区域的形状示于图 4-12 中。

## (2) 线段的置信区域 推导之二)

现在用另一种方法来推导线段的置信区域, 其基础是根据 4.1.1.2(2) 节中所导出的线段分布的公式。为了导出置信区域, 把线段  $Z_1 Z_2$  划分为三个区域  $A$ 、 $B$  和  $C$  见图 4-13。

### 1) 在子区 $B$ 中的置信域

根据式 4-11 对于所有  $r \in [0, 1]$   $Z_1 Z_2$  在  $y'$  方向上的分布为一个一维正态分布。因此, 对于一给定的  $r$  有

$$\frac{y'(r) - \nu_y(r)}{(\sigma_y(r))^{1/2}} \sim N(0, 1) \quad (4-33)$$

这里

$$y'(r) = (1-r)y'_1 + rY'_2$$

$$\nu_y(r) = (1-r)\nu'_1 + r\nu'_2$$

$$\sigma_y(r) = ((1-r)^2 + r^2)\sigma'_{22}$$

因此

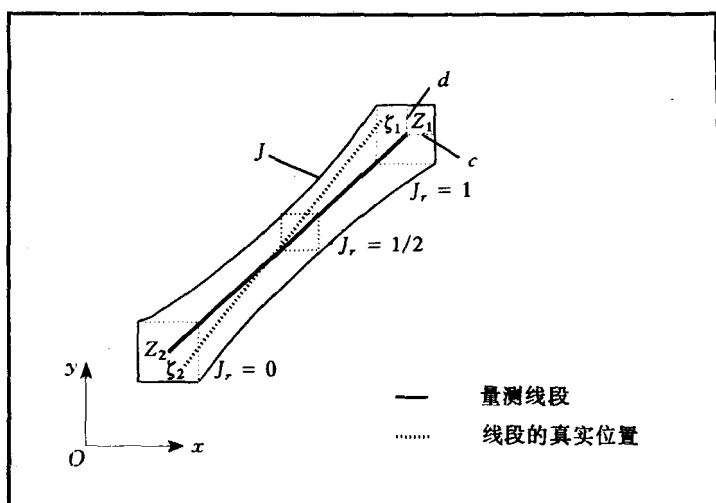


图 4-12 一线段的置信区域是由其所组成点  
置信区域的并集所构成的

$$\frac{(1-r)y'_1 + ry'_2 - (1-r)v'_1 - rv'_2}{[(1-r)^2 + r^2]\sigma'_{22}} \sim N(0,1)$$

如果线段上的任意一点同时落在某一区域内，则整条线段落在该区域内。因此，要建立这样一个区域，所有的点以某一置信水平落在该区域内。根据 Cauchy-Schwarz 不等式，有

$$\begin{aligned} & |(1-r)y'_1 + ry'_2 - (1-r)v'_1 - rv'_2| \\ &= |(1-r)(y'_1 - v'_1) + r(y'_2 - v'_2)| \\ &\leq [(y'_1 - v'_1)^2 + (y'_2 - v'_2)^2]^{\frac{1}{2}} [(1-r)^2 + r^2]^{\frac{1}{2}} \end{aligned}$$

不等式的两边同时除以  $[(1-r)^2 + r^2]\sigma'_{22}$  可得

$$\frac{|(1-r)y'_1 + ry'_2 - (1-r)v'_1 - rv'_2|}{[(1-r)^2 + r^2]\sigma'_{22}} \leq 1$$

$$\leq \left[ \frac{(y'_1 - \nu'_1)^2}{\sigma'_{22}} + \frac{(y'_2 - \nu'_2)^2}{\sigma'_{22}} \right]^{\frac{1}{2}} \quad (4-34)$$

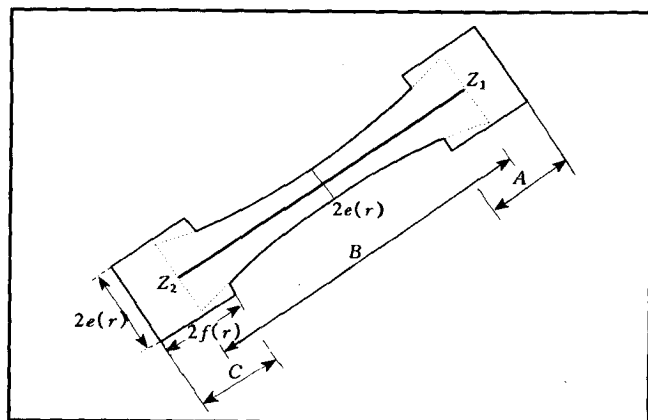


图 4-13 线段的三个区域 A、B、C。其置信区域是由这三个不同的区域内的置信区域构成

由于  $y'_1$  和  $y'_2$  相互独立且服从正态分布，有

$$\frac{(y'_1 - \nu'_1)^2}{\sigma'_{22}} + \frac{(y'_2 - \nu'_2)^2}{\sigma'_{22}} \sim \chi^2_2$$

这就是说， $(y'_1 - \nu'_1)^2/\sigma'_{22} + (y'_2 - \nu'_2)^2/\sigma'_{22}$  服从一个自由度为 2 的 Chi-平方分布。因此，有

$$P \left[ \frac{(y'_1 - \nu'_1)^2}{\sigma_{11}} + \frac{(y'_2 - \nu'_2)^2}{\sigma_{11}} \leq 5.99 \right] = 0.95$$

这里 5.99 是从一个 Chi-平方分布表 Li and Tao, 1982 中查得的。应用不等式 (4-34) 得

$$\left[ \frac{|(1-r)y'_1 + ry'_2 - (1-r)\nu'_1 - r\nu'_2|}{[(1-r)^2 + r^2]\sigma'_{22}} \right]^{\frac{1}{2}} \leq (5.99)^{\frac{1}{2}}, \text{ 对所有 } r$$

$$\geq 0.95$$

该式等价于

$P[y'(r) - e(r) \leq \nu_r(r) \leq y'(r) + e(r), \text{对所有 } r] \geq 0.95$   
这里

$$e(r) = 1 (5.99)^{\frac{1}{2}} [((1-r)^2 + r^2) \sigma'_{22}]^{\frac{1}{2}} |$$

至此 对于所有的  $r \in [0, 1]$  我们对参数  $\nu_r(r)$  建立了联合置信区间。 $y'(r) - e(r)$  和  $y'(r) + e(r)$  对所有  $r \in [0, 1]$  的轨迹是沿着  $Z'_1 Z'_2$  的两侧, 它们构成了  $Z_1 Z_2$  在  $B$  区域内的置信区域。

## 2) 子区 $A$ 和 $C$ 的置信域

子区  $A$  和  $C$  内的置信域分别是针对  $\zeta_1$  和  $\zeta_2$  而言的。在第一种推导方法中, 已经导出对于任意一点  $s_r$  的置信域。该结果可直接用于  $\zeta_1$  和  $\zeta_2$  因为它们都是  $s_r$  中的两个极端的情况。仅有的差别是 前者是在  $Oxy$  坐标系内导出的, 而现在旋转的坐系  $Ox'y'$  内讨论问题。

置信域  $J_r (r \in [0, 1])$  是以大于某一置信水平 如  $\gamma = 0.95$  的概率包含着真值  $\zeta_r$  的一个区域。

$$P(\zeta_r \in J_r) \geq 0.95 \quad (r \in [0, 1])$$

区域  $J_r$  是由一个  $[x', y']^T$  的点集合 其中  $x'$  满足

$$x'(r) - f(r) \leq x' \leq x'(r) + f(r)$$

并且  $y'$  满足

$$y'(r) - e(r) \leq y' \leq y'(r) + e(r)$$

这里

$$f(r) = 1 (7.38)^{\frac{1}{2}} [((1-r)^2 + r^2) \sigma'_{11}]^{\frac{1}{2}} |$$

$$e(r) = 1 (7.38)^{\frac{1}{2}} [((1-r)^2 + r^2) \sigma'_{22}]^{\frac{1}{2}} |$$

这里 7.38 是从 chi-平方分布表中查得的。当  $r$  为 0 时,  $J_r$  是  $A$  子

区内  $\zeta_1$  的置信域。而当  $r$  为 1 时,  $J_r$  是  $C$  子区内  $\zeta_1$  的置信区域。

$\zeta_1 \zeta_2$  的置信区域是其分别在子区  $A$ 、 $B$  和  $C$  内置信区域内的并集。图 4-13 显示出该置信域。

在用两种不同方法导出同一线段置信域之后, 我们希望知道两者的差别。一条线段的置信域是围绕量测线段的某一个区域, 该区域以大于某一置信水平的概率包含了线段的真实位置。通常地, 一个参数的置信区间定义为参数落入该区间的概率等于置信水平。在这里, 使用的是大于置信水平的概率。因此, 对于同一参数。相同的假设及正确的推导过程下, 得出了两个结果。事实上, 置信区间越窄 (即置信域越窄), 表示其结果越精确。现在来比较两个置信区域的宽度。为方便起见, 假设  $\sigma_{11} = \sigma_{22}$ ,  $\sigma_{12} = \sigma_{21} = 0$ , 并且  $\theta = \pi/2$ 。在这组假设下, 基于第一种推导的置信域宽度为

$$D_1 = 2(2)^{\frac{1}{2}} d = 2(2)^{\frac{1}{2}} [5.99((1-r)^2 + r^2)\sigma_{11}]^{\frac{1}{2}}$$

而基于第二种推导的置信域在子区  $B$  内的宽度为

$$D_2 = 2e(r) = 2[7.38((1-r)^2 + r^2)\sigma_{11}]^{\frac{1}{2}}$$

两种置信区域的宽度之比为

$$R = D_1/D_2 = 1.27$$

这意味着第二种推导方法提供了置信区域的更精确描述。而另一方面, 第一种推导在解析表达方面更加完美。另一个不同之处是置信域的形状, 第一种推导出的置信域的形状在两端点处随线段的不同方向而略有不同; 而第二种方法导出的结果则不随线段的方向而变化, 因为它根据线段本身所确定的坐标而导出的。

根据导出的线段置信区域, 可以说线段的真实位置以大于某一置信水平的概率被包含于置信域内。该置信域取决于线段两端点的协方差阵及所选定的置信水平。这样, 就可以用置信域模型来描述线段的不确定性, 即矢量 GIS 中的最基本要求之一。例如可以用 GIS 中一个矢量线划层的所有线段的置信区域之面积

总和与该层所占区域面积之比来指出该矢量层的位置精度。该比值越大，则说明其精度越低。

GIS 中一个面状物体的模糊边界区域及内部区域可由置信区域来描述。若一面状目标是由  $N$  条线段所构成的。对于每一条线段，在某一置信水平  $\gamma$  下导出其置信域。面状物体边线的置信区域是由所构成线段置信区域的并集所构成的。面状物体的模糊边界区域定义为其边界线的置信区域。而其内部区域则是总区域减去模糊边界所占的区域。模糊边界区域可解释成这样一个误差域，它在置信水平  $\gamma$  下，受到边界的顶点的误差的影响而构成的区域；而内部区域在置信水平  $\gamma$  下，则不受边界线顶点的误差的影响。模糊边界的宽度与所给出的置信水平及边界线顶点的误差有关。置信水平越高、边界点误差越大，则模糊边界越宽。反之就越窄。

#### 4.4.1.6 “进一步发展的误差带”模型

根据前面章节中提出的两个描述线段不确定性的指标，现在可以定义“进一步发展的误差带”模型。该模型是由两个指标所描述的。第一个描述指标是线段的置信域，第二个指标是线段的分布。这两个指标描述了线段不确定性的两个方面。置信区域提供了一个围绕量测线段的范围，线段的真实位置以大于某一置信水平的概率被包含在该区域内，而概率分布则描述了量测线段  $Z_1, Z_2$  对于其真实位置  $\zeta_1, \zeta_2$  的分布。这些被表示于图 4-14 中。

尽管线段的置信域与其概率分布有类似的形状，但是从统计学的角度看，它们具有不同的涵义。由于线段的置信域是围绕着“量测”线段的，在 GIS 中是可以获得的。因此，可以被直接用于描述矢量 GIS 中的位置不确定性。例如，在 GIS 中一层数据中有一组“量测”线段，对每一线段产生其置信域，并求置信域的面积总和，如为  $80 \text{ cm}^2$ 。将该值与所在层的面积总和（例如为  $1200 \text{ cm}^2$ ）

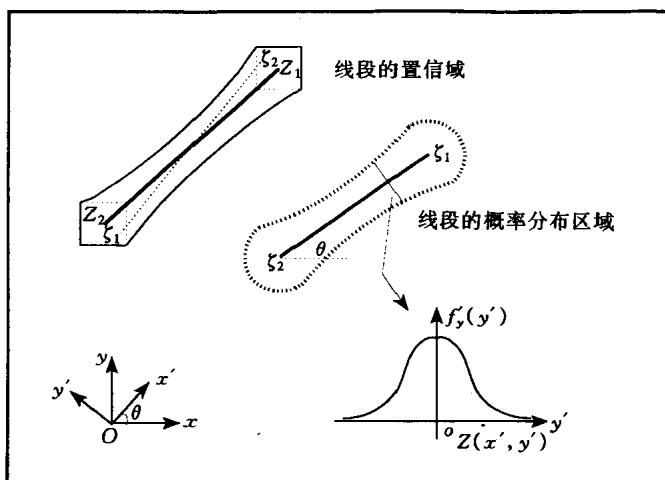


图 4-14 “进一步发展的误差带”模型

相比 其结果为 6.7% 即为描述该 GIS 层位置不确定性的一个指标。

另一方面，线段的概率分布在位置不确定性理论研究上更有价值，因为它不仅描述了一个不确定性区域，而且提供了更多的信息。在下一节中，还将进一步用它研究边界线段的不确定性。

我们称该研究中所提出的两个不确定性特征误差带所描述的模型为‘进一步

$Z_1, Z_2$  存在随机误差, 不能肯定  $Z$  一定属于  $O_1$ 。利用线段的概率分布来估计  $Z$  属于  $O_1$  的概率  $[P(Z \in O_1)]$  以及  $Z$  属于物体  $O_2$  的概率  $[P(Z \in O_2)]$ 。用这些概率值作为点  $Z$  不确定性的描述指标。

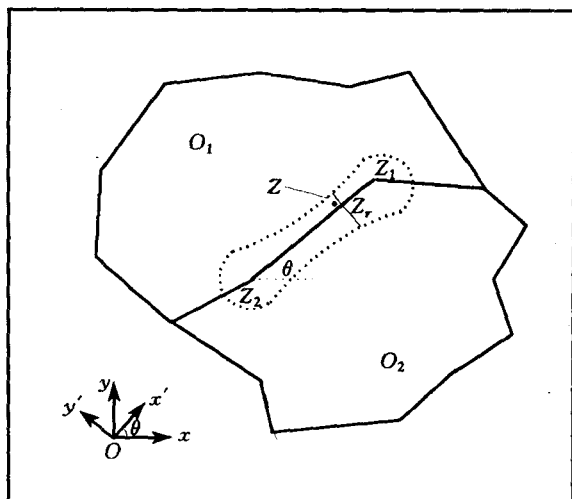


图 4-15 边界线段的不确定性  
( $Z$  可能是  $O_1$  或  $O_2$  内的一个元素)

假设点  $Z$  自己不具有随机误差。点  $Z_1$  和  $Z_2$  的随机误差是造成不确定性的误差源。点  $Z$  是由其在  $OXY$  坐标系内的坐标  $(x, y)^T$  确定的。我们可以把它转换到  $Ox'y'$  坐标系内, 记为  $(x', y')^T$ 。

出于简化的考虑, 用  $Z_1, Z_2$  来取代  $\zeta_1, \zeta_2$ , 并且基于垂直分布  $f_z(y')$  (见图 4-16), 现在可以定义  $Z$  属于  $O_1$  的概率  $[P(Z \in O_1)]$  如下:

$$P(Z \in O_1) = 1 - S(y'_z)$$

这里

$$S(y'_z) = \int_{-\infty}^{y'_z} f'_y(y') dy$$

而  $Z$  属于  $O_2$  的概率  $P(Z \in O_2)$  定义为

$$P(Z \in O_2) = 1 - P(Z \in O_1) = P(\neg(Z \in O_1))$$

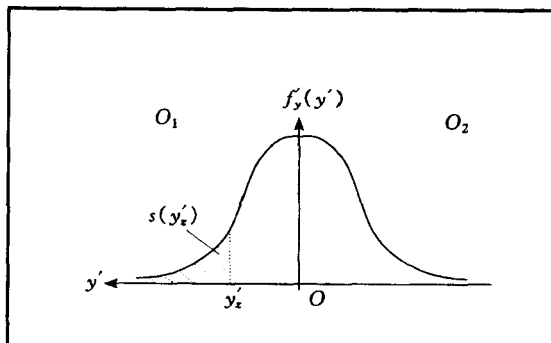


图 4-16 根据一正态分布及点的位置  $(x', y')$  计算概率矢量

现在构造一个矢量  $\{P(Z \in O_1), P(Z \in O_2)\}$  或  $\{P(Z \in O_1), P(\neg(Z \in O_1))\}$  来确定  $Z$  属于物体  $O_1$  或  $O_2$  的不确定性。矢量中两元素的差别越小，其不确定性越大。该矢量在以后的章节中将是被用做描述 PAT 不确定性的主要描述指标。

#### 4.4.2 边界线特征及面状物体的位置不确定性

##### 4.4.2.1 边界线特征的位置不确定性

一条线特征是由一条或多于一条线段所组成的。如果该线段特征为一面状物体的边界，则称之为边界线特征。本节将主要讨论边界线特征的不确定性问题。这里使用线段的分布，以及由此而导出的误差描述指标。

描述边界特征位置不确定性中的一个主要问题是处理两线段连接处的不确定性（见图 4-17）。在这区域内对于一给定点  $Z$  有两个属于面状物体  $O$  的概率。其一，根据线段  $L_1$  的分布有  $Z$  属于  $O_1$  的概率记为  $P(Z \in O_1; L_1)$  其二，根据线段  $L_2$  的概率分布有  $P(Z \in O_1; L_2)$  即点  $Z$  属于物体  $O_1$  的概率。现在需要根据线性特征  $L_{1 \cap 2}$  它是由线段  $L_1$  和  $L_2$  构成的 即解求  $P(Z \in O_1; L_{1 \cap 2})$  确定点  $Z$  属于物体  $O_1$  的概率 见图 4-17 所示。

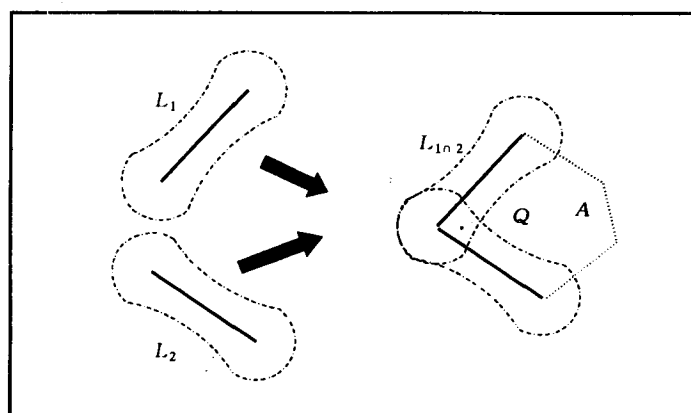


图 4-17 基于所构成线段的不确定性确定线状特征的不确定性

如果  $P(Z \in O_1; L_1)$  和  $P(Z \in O_1; L_2)$  是相互独立的，可以直接运用概率论中的乘积定律来解决这个问题。即

$$P_{1 \cap 2}(Z \in O_1; L_{1 \cap 2}) = [P(Z \in O_1; L_1)][P(Z \in O_1; L_2)]$$

当独立的假设不成立时，可以近似地应用模糊数学的方法解决这个问题。为了应用模糊数学的方法，用概率值当做隶属度值。例如  $P(Z \in O_1)$  被当作  $Z$  属于物体  $O_1$  的隶属度值。可以这样做

的原因是由于概率的主观解释认为概率是可信度 (belief) 的一种度量值。

在概率论中, 一个元素属于两个集合  $A$ 、 $B$  交集的隶属度值, 可以用下式计算

$$\mu_{A \wedge B}(x) = \min[\mu_A(x), \mu_B(x)] \quad (4-36)$$

这里  $\mu_{A \wedge B}(x)$  是  $x$  属于两模糊子集  $A$ 、 $B$  的交集的隶属度值, 而  $\mu_A(x)$  是元素  $x$  属于模糊子集  $A$  的隶属度,  $\mu_B(x)$  是  $x$  属于模糊子集  $B$  的隶属度值。在模糊数学中的隶属度值对应于概率值。

根据 4-36) 式 近似有

$$P_{1 \wedge 2}(Z \in O_1 : L_{1 \wedge 2}) = \min[P_1(Z \in O_1 : L_1), P_2(Z \in O_1 : L_2)] \quad (4-37)$$

这意味着, 可以近似地用“最小”运算来解求两线段  $L_1$  和  $L_2$  相交处的不确定性描述。相应地, 可以解求出所构成线特征不确定性的描述。

公式 4-36) 是原始的模糊数学计算方法。Klir 和 Folger(1988) 总结了计算交集的其他算法。对于某一应用, 可以根据问题的实际情况选用 4-36 式或者其他算法。

这里有两种方法求解  $P_{1 \wedge 2}(Z \in O_1 : L_{1 \wedge 2})$ : 其一是基于概率论; 其二是基于模糊数学。基于前者的解算值要小于后者。

基于概率论的解法逻辑上更为一致, 因为在前面的解求不确定性时一直是基于概率值的。

现在基于所构成线段的不确定性来描述线特征的不确定性。当计算  $P(Z \in O)$  时, 必须检查这种不确定性是否源于多于一条线段。由此, 可以确定一线特征的不确定性。线段与线特征的不确定性在于后者还需进一步考虑线段连接处的不确定性。位置不确定性通常只影响面状物体的边界区域, 内部区域不受影响或影响甚少。因此, 面状物体的不确定性是由其整体边界不确定性 (即构

成其的边界线段)所决定的。

#### 4.4.2.2 面状物体的位置不确定性

一面状物体定义为由一封闭的线特征所构成的面域。面状物体的位置不确定性是由其边界所确定的。面状物体的不确定性可以用属于该面状物体的概率,即  $P(x, y) \in [0, 1]$  来表示。当一点从物体的外部向中心移动时,该概率值由 0 逐渐地变到 1。

与基于 Epsilon 带的面状物体不确定性描述(见 3.2.4 节)相比这里所提出的不确定描述是一个连续的变化过程。而基于 Epsilon 带的描述则是用五个离散的定性描述来表达面状物体的不确定性即一点一定在内部、一定在外部、可能在外部、可能在内部以及不确定。而本研究提出的基于线段概率分布的描述,则给出了一个连续的变化区间  $[0, 1]$ 。这是一种定量的、连续的不确定性描述方法,这有利于进一步地与属性不确定性的结合以进行空间分析。

通过所构成的边界线段的不确定性对面状物体的不确定性进行了描述,而线段的不确定性则来源于其端点的随机误差。因此,面状物体的不确定性是由于所构成边界结点的误差所决定的。

以上讨论了位置不确定性的处理方法。在以下的章节中,将讨论属性不确定性问题。

#### 4.4.3 属性不确定性的空间分布

本研究所讨论的属性不确定性是指遥感影像分类中的不确定性。该不确定性产生的原因是在分类时没有足够的证据用于分类决策。如 3.3 节中所讨论的,现有的方法不能够描述出属性不确定性的空间分布。在以下章节中先简要讨论分类的原理,接着引入概率矢量的概念。最后根据概率矢量定义若干个属性不确定性指标。

#### 4.4.3.1 最大似然分类

在影像分类中，有几种方法是根据所感兴趣类别的概率分布进行的。这些方法包括最大似然分类、最小距离分类、平行四面体分类等。在这些方法中，最大似然分类方法是最基本的方法，而最小距离分类法是对其在某些假设下的简化。因此，在本研究主要讨论最大似然分类方法。现简要介绍该分类方法的原理。

设一影像中的光谱类别表示为

$$C_i, \quad i = 1, 2, \dots, m$$

这里  $m$  是总的类别的个数。当判别位于  $Z_T$  的某一像素的类别时 使用条件概率

$$P(C_i/Z_T), \quad i = 1, 2, \dots, m$$

矢量  $Z_T = (b_1, b_2, \dots, b_n)^T$  是一个光谱反射值的矢量，其中  $n$  为用于分类的光谱波段个数。 $Z_T$  是亮度为  $(b_1, b_2, \dots, b_n)^T$  在特征空间的一点。图 4-18 是特征空间为二维时的一个实例。

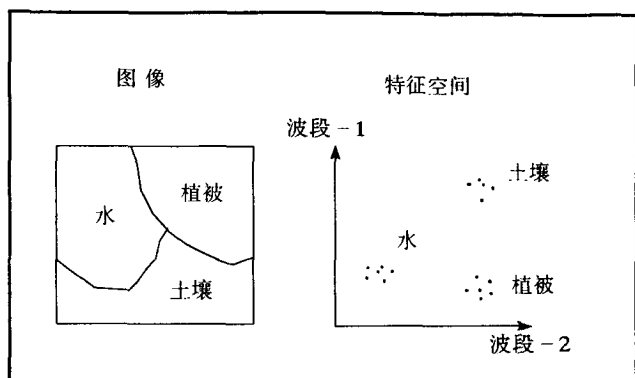


图 4-18 遥感分类中二维特征空间实例

条件概率  $P(C_i/Z_T)$  给出一个给定特征空间一点为类别  $C_i$  的概率。分类是根据以下规则进行的：

$$Z_T \in C_i \quad \text{if } P(C_i/Z_T) > P(C_j/Z_T) \quad \text{对所有 } j \neq i \quad (4-38)$$

即如果  $P(C_i/Z_T)$  是最大的 则该像素就分为类别  $C_i$ 。这是 Bayes 分类的一个特例。

在式 (4-38) 中,  $P(C_i/Z_T)$  通常是未知的。但是 在有足够的训练数据以及地面覆盖的先验知识 (即每一种地面覆盖所占的百分比) 的条件下, 该值可以估计出。条件概率  $P(C_i/Z_T)$  及  $P(Z_T/C_i)$  可以用 Bayes 定律表达如下：

$$P(C_i/Z_T) = P(Z_T/C_i) \cdot P(C_i)/P(Z_T) \quad (4-39)$$

这里  $P(Z_T/C_i)$  指值在  $C_i$  类中出现的概率。 $P(C_i)$  指  $C_i$  类在整幅影像内出现的概率。该值可以用  $C_i$  类所占的面积与总面积之比来估计。

对 (4-39) 式的左、右两侧取自然对数, 以解算决策函数。由于  $P(Z_T)$  与类别  $C_i$  无关, 它可以作为公共项而删去。

$$g_i(Z_T) = \ln(P(Z_T/C_i) + \ln(P(C_i))) \quad (4-40)$$

因此 (4-38) 式可以改写为

$$Z_T \in C_i, \text{ 若 } g_i(Z_T) > g_j(Z_T), \text{ 对所有 } j \neq i \quad (4-41)$$

这就是最大似然分类中使用的决策规则。而  $g_i(Z_T)$  被称为判别函数。在  $P(Z_T/C_i)$  服从多元正态分布的假设下, (4-40) 式可以改写为

$$g_i(Z_T) = -\ln |\Sigma_i| - (Z_T - M_i)^T \Sigma_i^{-1} (Z_T - M_i) \quad (4-42)$$

这里  $M_i$  和  $\Sigma_i$  是  $C_i$  的数据矢量的均值矢量及协方差阵。在最大似然分类中, 这些数据可以从训练数据中获得。(4-42) 为标准最大似然分类方法中使用的判别函数。详细的讨论见 Richards (1986)。

#### 4.4.3.2 概率矢量

从 (4-38) 公式中可以看出, 将像素  $Z_T$  分为类别  $C_i$  时, 并非百分之百地肯定  $Z_T$  应当属于  $C_i$ 。因此, 这是一个带有不确定性的决策过程。为了描述在最大似然分类中引入的不确定性, 这里使用分类过程中本身所具有的概率矢量。

根据 (4-39) 式, 有以下矢量:

$$[P(C_1/Z_T) \ P(C_2/Z_T) \ \cdots \ P(C_i/Z_T) \ \cdots \ P(C_n/Z_T)]^T \quad (4-43)$$

将该矢量中的零元素删除, 并将其从大到小的顺序排列出, 即有如下形式的一个矢量:

$$PV(Z_T) = [P(C_{i1}/Z_T) \ P(C_{i2}/Z_T) \ \cdots \ P(C_{ik}/Z_T)]^T \quad (4-44)$$

这里  $k$  是 (4-43) 式中非零元素的个数, 并且有  $P(C_{i1}/Z_T) \geq P(C_{ij}/Z_T)$  对于  $i < j$ 。由 (4-44) 所构成的矢量就定义为最大似然分类所产生的像素  $Z_T$  的概率矢量。容易看出, 在一个最大似然分类中, 一个像素被分为该矢量中第一个元素所属的类别。而概率矢量中的其他元素可以用来描述分类的不确定性。

#### 4.4.3.3 由概率矢量导出的参数

根据 (4-44) 式所定义的概率矢量, 可以进一步定义若干个参数来描述分类像素之不确定性。首先定义绝对不确定性参数 ( $U_A$ ) 如下:

$$U_A(Z_T) = P(C_{i1}/Z_T) / [1 - P(C_{i1}/Z_T)] \quad (4-45)$$

这里  $U_A$  的取值范围为  $(0, +\infty)$  该值越大, 说明将  $Z_T$  分为类别  $C_{i1}$  的不确定性越小。例如,  $P(C_{i1}/Z_T)$  为 90% 时, 根据 (4-45) 式,  $U_A(Z_T)$  为 9.0。又如, 如果  $P(C_{i1}/Z_T)$  为 20% 则  $U_A(Z_T)$  为 0.25。

由此, 可以根据  $U_A(Z_T)$  的取值看出在第一个分类中有较小的不确定性, 而在第二个分类中有较大的不确定性。而不确定性的太小可以从  $U_A(Z_T)$  参数值的大小表示出来。对于每一个像素都用该值描述之后, 就可以描述出分类的属性不确定性在整幅影像上的分布。

第二个参数是相对不确定性参数 ( $U_R$ ) 它是用于描述一个像素可能错误地在  $C_k$  和  $C_j$  类之间误分类。该参数定义为

$$U_R(Z_T, i, j) = |P(C_k/Z_T) - P(C_j/Z_T)| \quad (4-46)$$

参数  $U_R(Z_T, i, j)$  的取值范围为  $[0, 1]$ 。该参数取值越大, 像素  $Z_T$  就越易于在  $C_k$  和  $C_j$  类之间区别开, 即不易混分为  $C_k$  或  $C_j$ 。例如, 根据 4-46 式, 如果  $P(C_{11}/Z_T)$  为 80% 而  $P(C_{12}/Z_T)$  为 10%。则  $U_R(Z_T, 1, 2)$  为 70% 即 0.70。即该像素很容易被分为类别  $C_{11}$  而不会混为  $C_{12}$  类。而在另一个例子中,  $P(C_{11}/Z_T) = 33\%$  而  $P(C_{12}/Z_T)$  为 30%, 从而  $U_R(Z_T, 1, 2) = 0.03$ 。这种情况意味着, 很难区分  $Z_T$  是类别  $C_{11}$  或是  $C_{12}$  类。在第一个例子中, 很容易区别  $Z_T$  究竟是  $C_{11}$  类还是  $C_{12}$  类, 而在第二个例子中, 不容易确定  $Z_T$  属于  $C_{11}$  类还是  $C_{12}$  类。

第三个参数是用于度量一个像素为混合像素的程度, 即一个像素含有多种不同的类别。该参数定义为

$$M = \frac{\sum_{j=1}^n |P(x \in C_i) - P(x \in C_j)|}{P(x \in C_i)}, \text{ 对所有 } i \neq j$$

这里  $P(x \in C_i)$  是像素属于类别  $C_i$  的概率。

如果概率矢量中各元素的量值大小类似, 则  $m$  之值较小, 即说明该像素有较大的可能成为一个混合像素。

现在定义第四个参数 ( $\Theta$ ), 用于描述证据的不完整性。对于像素  $Z_T$  定义如下:

$$\sum_{i=1}^k P(C_i/Z_T) + \textcircled{H} = 1$$

$\textcircled{H}$ 的存在指出这样一个事实，即有时对一个像素进行分类时我们不能肯定它属于哪一类别，而定义为“未分类类别”。 $\textcircled{H}$ 值越大，意味着该像素被分为“未分类类别”的可能性越大。 $\textcircled{H}$ 的取值范围在 $[0,1]$ 内。

以上讨论了几种描述属性不确定性的方法：概率矢量及四个参数 $(U_A, U_R, M \text{ 和 } \textcircled{H})$ 。以下的讨论中，将直接使用概率矢量描述“S-带”中的属性不确定性。

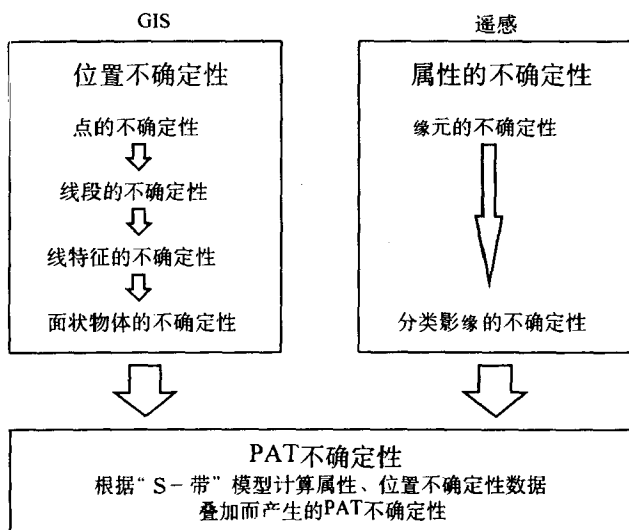
在最大似然分类中有两个主要的误差源。首先，是训练数据。概率矢量的取值取决于训练数据的质量。分类中计算的统计特征依赖于训练数据，是带有主观性的不确定性。为了克服这种局限性，用于评估总体统计量评价方法可以使用。其次，这里假设每一类别的概率分布是服从多元正态分布的。在现实世界中，这一假设未必总是成立。这是造成分类不确定性的第二个因素。

存贮问题是使用概率矢量描述分类不确定性的一个障碍。例如 概率矢量的维数为 8，则用于存贮概率矢量的存贮量至少要八倍于存贮一个单波段影像所用的存贮量。解决该问题的方法是将概率矢量转化为几个参数。本节所提出的四个参数可用于此目的。

在本节中，讨论了分类遥感影像的属性不确定性。所用的方法是概率矢量及所导出的参数，这些可做为描述位置与属性不确定性模型的基本因素。

#### 4.4.4 结合位置与属性不确定性的模型

在 4.4.1 节和 4.4.2 节中 讨论了面物体的位置不确定性 而在 4.4.3 节中讨论了属性不确定性。现在来讨论结合位置与属性不确定性 PAT 的模型。该过程见图 4-19 所示。



#### GIS与遥感

图 4-19 用“S-带”模型处理物体的 PAT不确定性

首先通过一个例子，介绍对位置与属性不确定性的客观要求。例如在一个土地调查项目中，经常会要求调查出某一行政边界内（如一个县界内）各种土地覆盖类型的面积。利用遥感技术，可以对影像分类而获得各种用地类型的分布。从一张行政边界图上数字化行政管理边界，而输入该数据于 GIS 中。最后，可以叠加这两层数据，即分类影像和数字化边界而求出各行政边界的各种用地类型的面积。这是利用遥感和 GIS 技术进行土地利用调查的一种常用的方法（见图 4-20 所示）。

在以上的实例中，假定分类的影像及数字化的地图均没有误差。然而，事实上以上两层数据均含有误差。例如，影像是用最大似然法分类生成的。例如一个像素被分为森林，有可能只有 70%

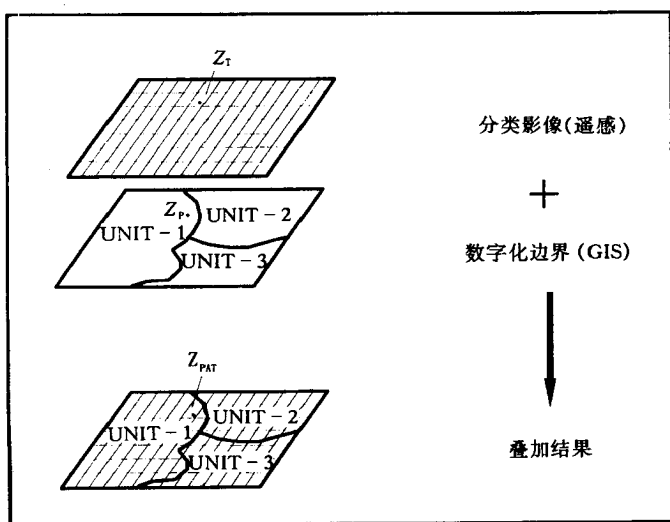


图 4-20 应用遥感与 GIS 技术进行土地调查

的把握说该像素应该为森林，而 30% 不为森林的可能性。因此，在分类的影像中是隐含着不确定性。

另一方面，GIS 中的数字化边界具有位置误差。若某一点靠近某一边界线段，则不是十分确定该点一定在边界内部。该不确定性源于由诸多因素所造成的点的随机误差。有关内容见第一章的讨论。这属于位置不确定性。可以用 4.4.2 节中所描述的方法来确定这种不确定性。

因此，迭叠操作并不是没有误差的。这将涉及到不确定性的结合及误差传递的问题。因此，需要发展描述结合属性与位置不确定的方法。下面将首先定义和公式化所提出的问题，之后导出其结合模型。

#### 4.4.4.1 位置与属性不确定性：问题的定义

对一遥感影像用最大似然法分类。对于一给定的像素  $Z_T$  ,可以解算该像素属于类别  $C_i$  的概率 即

$$P[Z_T \in C_i / Z_T(x)]$$

其中  $Z_T(x) = (b_1, b_2, \dots, b_n)^T$  是  $n$  维特征空间中的一个矢量, 其定义见 4.4.2 节;  $C_i$  是总的类别系统中的一类。

在 GIS 中的一个数字化层里, 有靠近边界的一点  $Z_p$  并且有

$$P[Z_p \in O_i / Z_p(x)]$$

即对于给定点  $Z_p$  可以解算出该点属于面物体  $O_i$  的概率。该概率是一个描述位置不确定性指标。  $Z_p(x) \in R^2$  , 是一个二维 Euclidean 空间内的点。通常地  $Z_p(x) = (x, y)^T$ 。而  $O_i$  是某一面状目标名, 例如可以是一个县名。

现在将问题定义如下 在 GIS 中的两层叠加之后, 一点  $Z_{PAT}$  属于类别  $C_i$  同时属于物体  $O_i$  的概率是多大? 也即

$$P\{[(Z_T \in C_i / Z_T(x)) \wedge Z_p \in O_i / Z_p(x)]\} = ?$$

这里  $Z_{PAT} \in R^{n+2}$  , 即  $Z_{PAT} = (b_1, \dots, b_n, x, y)^T$ 。这意味着需要研究一个可以结合属性及位置不确定性的模型。

为方便后续的讨论, 用以下符号代表所讨论的事件:

$$H_T: Z_T \in C_i$$

$$H_p: Z_p \in O_i$$

$$H_{PAT}: (Z_T \in C_i / Z_T(X) \wedge (Z_p \in O_i / Z_p(x)))$$

因此, 我们的研究问题是, 给定  $P(H_p / Z_p(X))$  和  $P(H_T / Z_T(X))$  的条件下,  $P(H_{PAT})$  是多大?

本节主要研究可用于结合位置与属性不确定性的模型。为此, 引入两种方法: 其一是基于概率论的方法; 其二是基于具有概

率解释的确定因子模型。

#### 4.4.4.2 基于概率论的结合模型

在概率论中，有如下乘积定理：

$$P(A \wedge B) = P(A)P(B/A) = P(B)P(A/B) \quad (4-47)$$

这里  $P(A \wedge B)$  是事件  $A$  与  $B$  同时成立的概率。 $P(B/A)$  是  $A$  成立的条件下  $B$  发生的概率。 $P(A)$  是事件  $A$  发生的概率。 $P(A/B)$  是事件  $B$  发生的条件下事件  $A$  发生的概率。 $P(B)$  是事件  $B$  发生的概率。

在事件  $A$  与  $B$  相互独立的条件下，(4-47) 式所表达的乘积定理为

$$P(A \wedge B) = P(A)P(B) \quad (4-48)$$

现在可以用概率的乘积定理来解决我们所定义的问题。由于  $H_T$  与  $H_P$  是相互独立事件，直接可以使用 4-48 式 即有

$$P(H_{P \wedge T}) = P(H_P/Z_P(x))P(H_T/Z_T(x)) \quad (4-49)$$

以上基于概率论中的乘积定理的方法可以解决 4.4.1 节中所提出的问题。然而，我们还希望研究出具有更一般性意义的模型。乘积定律的方法是一般的概率推理方法。为此，需要满足具有大量采样数据的要求，并且通常的计算量较大。而在实际应用中，获取大量的采样数据通常是十分昂贵的。因此希望有其他方法来解不确定结合问题。在一些推理问题中，可能会遇到如下情形。对于某一假设，某些证据对它起正面的作用，而有些则起负面的作用。例如，用多源数据确定某一地面覆盖的类型。从分类遥感影像上得到的结果是该地块为“森林”而从其他来源数据（如 GIS）得到的信息是该地块不是“森林”而是“草地”。这里需要一不确定模型，它可以描述正面的可信度（如地块是森林）以及负面的可信度（如地块不是森林）。它的可信度区间可以为  $[-1, 1]$ 。而用概率描述的可信度范围为  $[0, 1]$ ，即只可以描述正面的可信度。利用

专家系统中常用的确定因子模型，可以克服使用概率论中的乘积定律模型的局限性。

#### 4.4.4.3 基于确定因子的结合模型

这里同时考虑确定性与不确定性的传播问题。用于 MYCIN 系统 (Shortliffe and Buchanan, 1975) 中的确定因子模型可用于基于不确定性的推理。MYCIN 是一个用于医疗诊断的专家系统。由于确定性因子模型可以克服概率论中的乘积定理中的某些限制，我们把确定因子模型用做为处理结合不确定性的另一种方法。

将使用具有概率解释的确定因子模型，该模型是 Heckerman (1986) 根据确定因子的原始模型而进一步发展而得出的。以下，首先在 1) 和 2) 节中，分别介绍确定因子模型及其概率解释。之后，将讨论应用这一模型用于所讨论的问题中。

##### (1) MYCIN 的确定因子模型

MYCIN 确定因子模型是用于处理推理过程中具有不确定性的方法。发展 MYCIN 确定因子模型的基本目标是提供处理不确定性的一个方法，以避免常规概率统计推理过程中对大量数据以及交互式计算的要求。

MYCIN 的知识存以规则的形式：如果 { 证据 } 那么 { 假设 }。在医学上，证据与假设之间的关系通常是不确定的。确定因子模型就是用来处理这种不确定关系的。一条规则都附有一个确定因子，用来表示给出某一证据对假设的可信度的改变。确定因子的取值范围为  $[-1, 1]$ 。正值表示对假设可信度的增加，而负值则表示对可信度的减低。确定因子并不对应于绝对可信度的度量。

以下表达了一个规则：

$$E \xrightarrow{CF(H, E)} H$$

这里  $H$  是一个假设， $E$  是与假设相关的一个证据而  $CF(H, E)$  则

是附于该规则的确定的因子。

对于概率的解释主要有两个基本的观点。第一种观点是概率的主观解释，在这种解释中，概率被视为多次试验的平均值，即客观解释。另一种观点则认为概率是可信度的一种度量，即主观解释。确定因子是根据概率的主观解释 (Shortliffe and Buchanan, 1975) 而定义的。确定因子  $CF(H, E)$  表达了证据  $E$  对假设  $H$  的信度的改变。 $CF(H, E)$  定义为：

$$CF(H, E) = \begin{cases} \frac{P(H/E) - P(H)}{1 - P(H)} & \text{对于 } P(H/E) > P(H) \\ \frac{P(H/E) - P(H)}{P(H)} & \text{对于 } P(H/E) < P(H) \end{cases} \quad (4-50)$$

这里  $P(H)$  是假设  $H$  发生的先验概率而  $P(H/E)$  则是给定  $E$  下  $H$  的后验概率。这个定义指出，确定因子是可信度变化的度量指标。

在 MYCIN 中，有可能会出现几个有关项共用一个假设。另一种可能是一个假设是另一个假设的证据。其结果即为图 4-21 所示的规则网络。这个网络结构经常被称为推理网络 (Duda et al., 1976) 确定因子的主要用途是通过推理网络描述不确定性的传递。

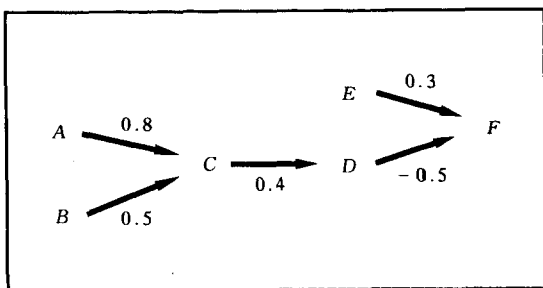


图 4-21 一个推理网络的实例

其中的数值表示确定因子的值，如  $CF(C, A) = 0.8$

推理网络的不确定性传递，可以通过重复地结合网络基本单元来实现。图 4-22 的左侧是基本网络的一种。这里有两项证据  $E_1$  和  $E_2$  同时支持同一个假设  $H$ 。两个确定因子  $CF(H, E_1)$  和  $CF(H, E_2)$  结合产生一个确定因子  $CF(H, E_1 E_2)$ 。而该结合的确因子是对于假设规则：如果  $\{E_1$  和  $E_2\}$  则有  $\{H\}$ 。该过程见图 4-22 的右半侧。这个过程被称为平行结合。确定因子平行结合

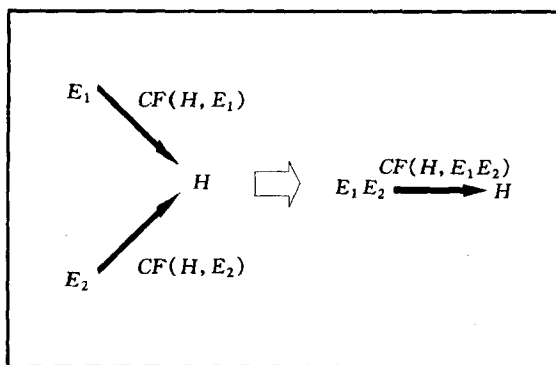


图 4-22 确定因子的平行结合

的函数定义如下 (Heckerman, 1986):

$$Z = \begin{cases} x + y - xy & x, y \geq 0 \\ \frac{x + y}{1 - \min(|x|, |y|)} & x \cdot y < 0 \\ x + y + xy & x, y < 0 \end{cases} \quad (4-51)$$

这里  $x = CF(H, E_1)$ ,  $y = CF(H, E_2)$ ,  $z = CF(H, E_1 E_2)$ 。

第二个网络基本单元是顺序结合。在这种情况下，一个假设是另外一规则的证据。例如图 4-23 的上半部分。两个规则的确因子  $CF(E, E')$  与  $CF(H, E)$  结合而产生新的假设的确因子  $CF(H, E')$ 。见图 4-23 的下半部分。

· 确定因子的序列结合函数定义如下：

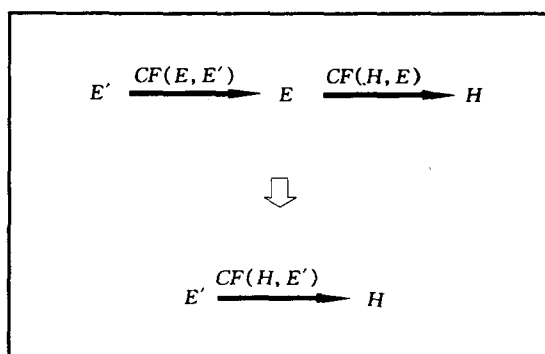


图 4-23 确定因子的顺序结合

$$z = \begin{cases} wx & w \geq 0 \\ -wy & w < 0 \end{cases} \quad (4-52)$$

这里

$$w = CF(E, E')$$

$$x = CF(H, E)$$

$$y = CF(H, \neg E)$$

$$z = CF(H, E')$$

基于(4-51)式和(4-52)式两个确定因子的结合模型，可以重复地计算推理网络中的平行与顺序结合，从而得出整个网络的推理结果。图 4-24 表示出对图 4-21 所提出的推理网络运用确定因子的平行、顺序结合模型进行推理的过程。

图 4-24 中的推理过程如下。图 4-24(a)示出原始的推理网络。假使  $A$ 、 $B$  和  $E$  真，我们想知道，这将如何影响  $F$  的可信度。证据  $A$ 、 $B$  的确定因子用平行结合模型结合，等价的推理网络示于图 4-24(b)中。 $AB$  和  $C$  的确定因子用顺序结合模型结合后，等价网络图示于图 4-24(c)中。 $AB$  和  $D$  序列结合后产生 4-24(d)

图所示的结果。最后  $E$  与  $AB$  平行结合而导致图 4-24(e) 所示的网络。所得出的结果为证据  $A$ 、 $B$  和  $E$  成立的条件下  $F$  的确定因子为 0.15。有关确定因子详细的过程描述见 Heckerman(1986)。

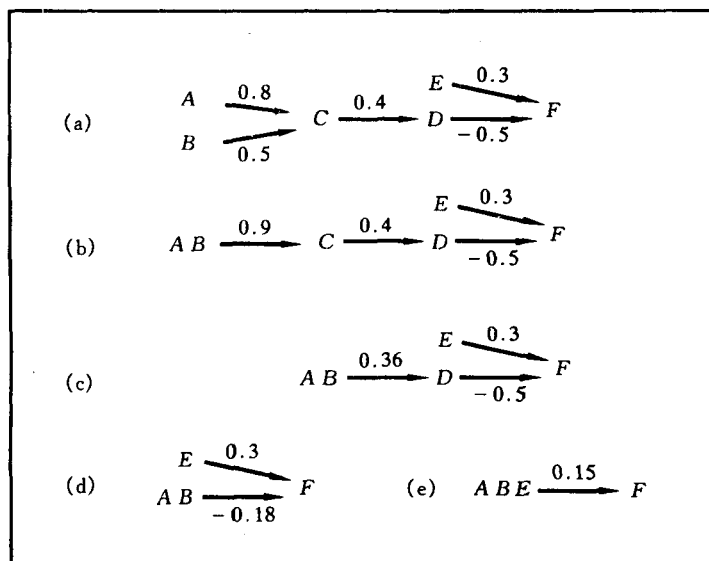


图 4-24 应用推理网络进行推理过程及其中不确定性传递

## (2) 确定因子的概率解释

由 (4-50) 式所定义的原始确定因子定义有两个局限性。首先它与 MYCIN 中所使用的用于结合量的函数不一致。其次原始确定因子模型与概率论的关系不密切 (Heckerman, 1986)。因此, Heckerman(1986) 将其重新定义如下:

$$CF_1(H, E) = \begin{cases} (\lambda(H, E) - 1) / \lambda(H, E) & \lambda \geq 1 \\ \lambda(H, E) - 1 & \lambda < 1 \end{cases} \quad (4-53)$$

其中的似然比定义如下:

$$\lambda(H, E) = P(E/H) / P(E/\neg H)$$

这里,  $P(E/H)$  是给定假设  $H$  条件下  $E$  成立的概率, 而  $P(E/\neg H)$  是给定  $\neg H$  成立的条件下  $E$  发生的条件概率。似然比  $\lambda$  是可信度变化的一个度量指标, 其取值范围为  $[0, +\infty)$ 。这就是具有概率解释的确定因子模型。

根据以上新定义的确定因子, 平行结合函数重新定义为

$$\lambda(H, E_1 E_2) = \lambda(H, E_1) \lambda(H, E_2)$$

因此, 平行结合是似然比的乘积。顺序结合定义为

$$\lambda(H, E')$$

$$= \frac{\lambda(E, E') \lambda(H, E) (1 - \lambda(H, \neg E)) + \lambda(H, \neg E) (\lambda(H, E) - 1)}{\lambda(E, E') (1 - \lambda(H, \neg E)) + (\lambda(H, E) - 1)}$$

(4-55)

关于这两个结合模型的详细描述见 Heckerman(1986)。在平行、顺序结合之后, 可利用 (4-53) 式将似然比转换成确定因子。因此, 可以得到结合后的确定因子值。

至此, 已经介绍了两组确定因子模型及相应的平行与顺序结合方法: 基于原始定义的确定因子模型和具有概率解释的确定因子模型。在本研究中, 采用了具有概率解释的确定因子模型。这是由于对位置、属性两方面的不确定性描述都是基于概率论的考虑。为了与之一致, 在发展结合模型时也首先考虑了具有概率解释的模型, 以保持模型在理论上的一致性。

### (3) 基于确定因子模型的 'S-带' 结合模型

下面我们应用具有概率解释的确定因子模型来解决位置与属性不确定性的结合问题。即已知  $P[H_P/Z_P(X)]$  和  $P[H_T/Z_T(X)]$ ,  $P(H_{P \wedge T}) = ?$

这里  $H_T: Z_T \in C_T$

$$H_P: Z_P \in O_P$$

$$H_{P \wedge T}: (Z_T \in C_i / Z_T(\mathbf{x})) \wedge (Z_P \in O_i / Z_P(\mathbf{x}))$$

其中 证据  $E$  是特征空间或 Euclidean 空间内点的坐标，它们分别表示为  $Z_T(\mathbf{x})$  和  $Z_P(\mathbf{x})$ 。

· 公式推导

$$[(Z_T \in C_i / Z_T(\mathbf{x})) \wedge (Z_P \in O_i / Z_P(\mathbf{x}))] \subseteq [Z_P \in O_i / Z_P(\mathbf{x})]$$

$$[(Z_T \in C_i / Z_T(\mathbf{x})) \wedge (Z_P \in O_i / Z_P(\mathbf{x}))] \subseteq [Z_P \in C_i / Z_T(\mathbf{x})]$$

由以上条件 并有似然比  $\lambda$  定义 可得到

$$\lambda(H_{P \wedge T}, Z_P(\mathbf{x}) Z_T(\mathbf{x})) \leq \lambda(H_P, Z_P(\mathbf{x})) \quad (4-56)$$

$$\text{且 } \lambda(H_{P \wedge T}, Z_P(\mathbf{x}) Z_T(\mathbf{x})) \leq \lambda(H_T, Z_T(\mathbf{x})) \quad (4-57)$$

在 4.4.4.1 节中所定义的问题，其本质属于平行结合问题。

因此应用 4-54) 式有

$$\lambda(H_{P \wedge T}, Z_P(\mathbf{x}) Z_T(\mathbf{x})) \leq \lambda(H_{P \wedge T}, Z_R(\mathbf{x})) \lambda(H_{P \wedge T}, Z_T(\mathbf{x}))$$

根据(4-56)和(4-57)式,有

$$\lambda(H_{P \wedge T}, Z_P(\mathbf{x}) Z_T(\mathbf{x})) \leq \lambda(H_P, Z_P(\mathbf{x})) \lambda(H_T, Z_T(\mathbf{x})) \quad (4-58)$$

这里

$$T = \lambda(H_P, Z_P(\mathbf{x})) \lambda(H_T, Z_T(\mathbf{x})) \quad (4-59)$$

其中  $T$  是  $\lambda(H_{P \wedge T}, Z_P(\mathbf{x}) Z_T(\mathbf{x}))$  上限值的一个域值 即结合后的似然值小于或等于该域值。也即

$$\lambda(H_{P \wedge T}, Z_P(\mathbf{x}) Z_T(\mathbf{x})) \leq T \quad (4-60)$$

应用 Bayes 公式和 4.4.4.8) 式，可以用下式计算出似然比值：

$$\lambda[H_T, Z_T(\mathbf{x})] = \frac{P[H_T / Z_T(\mathbf{x})]}{P[H_T / Z_T(\mathbf{x})]} \cdot \frac{P(H_T)}{P(H_T)} \quad (4-61)$$

在一个分类影像中，在用最大似然分类之后，可以知道每一个像素的属性条件概率  $P(H_T / Z_T(\mathbf{x}))$ 。而  $P(H_T / Z_T(\mathbf{x}))$  可以用  $1 - P(H_T / Z_T(\mathbf{x}))$  计算出来。先验概率  $P(H_T)$  和  $P(H_T)$  可以用分类中的标准技术解算出来。利用 (4.4.4.16) 以及上值，可以解算

出  $\lambda(H_T, Z_T(x))$ 。

位置不确定性的似然比可以计算如下

$$\lambda(H_p, Z_p(x)) = \frac{P(H_p/Z_p(x))}{P(H_p)} \frac{P(H_p)}{P(H_p)} \quad (4-62)$$

其中  $P(H_p/Z_p(x))$  和  $P(H_T/Z_T(x))$  可用 4.4.2 节所讨论的面状物体位置不确定性方法解算出来。而  $P(H_p)$  可以用面状物体所占有的面积与总面积之比求得。而  $P(H_T)$  等于  $1 - P(H_p)$ 。将上述值代入式 (4-62) 可以解算出  $\lambda(H_p, Z_p(x))$ 。

· 计算 PAT 不确定性指标的过程

位置与属性不确定性的结合，在使用以上导出的方法解算时可以采用以下步骤。

第一步 根据  $P(H_p/Z_p(x))$  和  $P(H_T/Z_T(x))$  式计算  $P(H_p/Z_p(x))$  和  $P(H_T/Z_T(x))$

$$P(H_p/Z_p(x)) = 1 - P(H_T/Z_T(x))$$

$$P(H_T/Z_T(x)) = 1 - P(H_p/Z_p(x))$$

而  $P(H_p)$  和  $P(H_T)$  可以从对于地图或影像的先验知识中获得。

第二步 根据 (4-61) 和 (4-62) 式计算  $\lambda(H_T, Z_T(x))$  以及  $\lambda(H_p, Z_p(x))$ 。

第三步 根据 (4-59) 式计算  $\lambda(H_{p \wedge T}, Z_p(x), Z_T(x))$  的域值 ( $T$ )。

第四步 根据 (4-53) 式计算  $T$  所对应的确定因子值。确定因子值是 PAT 不确定性的上限值指标。也就是说，结合的不确定性值 (确定因子值) 小于或等于  $T$  所对应的确定因子值  $CF(T)$ 。

现在得到了确定因子的一个域值，它是给定证据  $Z_p(x) \cdot Z_T(x)$  的条件下假设  $H_{p \wedge T}$  成立的最大确定因子值。这是基于具有概率解释的确定因子模型的位置与属性不确定性 (PAT) 指标。

在实际应用中，可以首先将 GIS 与遥感集成中的有关推理表

达为一个推理网络。根据问题中结合的本质，可选用平行或顺序结合的方法对网络进行推理。应用以上推导的算法，对每个像素最终可以得到确定因子指标，即为 PAT 不确定性的描述指标。该指标对每一像素给出一个上界值域值，该像素的不确定性小于该域值。

## 4.5 本章小结

在本章中，我们对三个方面的问题进行了理论推导。首先，研究了 GIS 中面状物体的位置不确定性空间分布的模型；其次，对分类遥感影像中的属性不确定性进行了讨论；第三，发展了“S-带”模型用于结合位置和属性不确定性。

### 4.5.1 地理信息系统中面状物体的不确定性

GIS 中面状物体的不确定性是由线段之不确定性、边界线段之不确定性导出的。这些特征的不确定性是源于所构成的节点之不确定性。在以下两个假设下，即这些节点相互独立以及服从二维正态分布，我们导出了以下结果。

(1) 首先，导出了线段的置信区间的模型。这是一个围绕线段量测值的误差区域，线段的真实位置以大于某一置信水平的概率被包含于该区域内。置信区域的大小取决于所选定的置信水平及两端点的误差。置信区域的概念可用于定义由于位置误差所引起的模糊边界区域及内部区域。

(2) 导出了线段的概率分布模型。它是由一组一维正态分布和两个二维正态分布所描述的。它们的导出是基于线段上的任意一点在垂直于线段方向上的概率分布。

(3) 根据线段的分布，定义了一组线段的误差指标。在描述几何特征之不确定性的领域中，这是对传统的一维、二维误差指标

的进一步发展，即四维随机变量线段的误差指标。

(4) 本文在以下两个方面对误差模型进行了进一步的发展：

导出了线段的置信区域的解析表达式；② 线段的概率分布。这是对传统的误差带和 Epsilon 带的进一步发展与完善。

(5) 当某一线段为某一物体边界的一部分时，边界线段的不确定性是由某一点属于该线段所表达的物体的概率。该概率值是根据线段上任意一点在垂直线段方向上的分布导出的。

(6) 线状特征和面状物体的不确定性是根据所构成线段的不确定性以及对线段连接处的不确定性的进一步讨论而完成的。一面状物体的不确定性是由一点属于面状物体的概率来描述的。当一点从面状物体的外部连续地移入物体的内部，则该概率值从 0 变到 1。该描述方便于不确定性的定量描述，并且有助于与属性不确定性的结合。图 4-25 显出 GIS 中位置不确定性描述指标发展的内在关系。

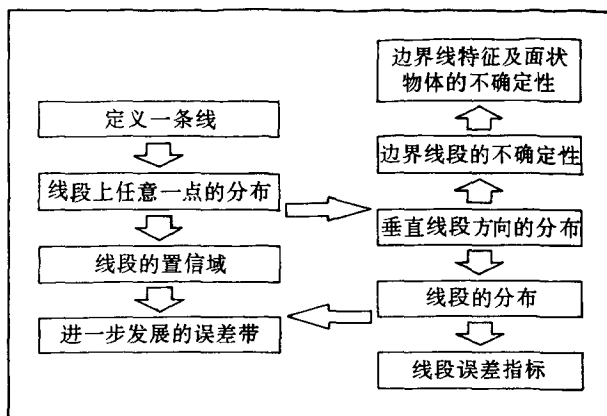


图 4-25

### 4.5.2 分类影像的属性不确定性

定义了一个概率矢量的概念用于描述基于最大似然法分类的影像的不确定性。根据概率矢量模型，进一步地定义了一组属性不确定性指标。在本研究中，概率矢量用于描述最大似然分类影像中的属性不确定性。

### 4.5.3 “S-带”模型与 PAT 不确定性

我们首先提出了 PAT 不确定性及其度量指标，即一个既含有位置方面的，又含有属性方面的不确定性。“S-带”的发展是用来处理 PAT 不确定性，尤其是模糊边界内的不确定性。

“S-带”模型有两个层次的概念：其一是理论上的“S-带”模型。根据这一模型，在描述 PAT 不确定性时，需要讨论位置不确定性、属性不确定性以及两者之间的互相关系。在理论模型的概念指导下，可以针对具体的应用提出实用“S-带”模型。

在本研究中，实用“S-带”模型是在概率论以及具有概率论解释的确定因子模型的范畴内讨论的。（“S-带”是在模糊边界区域内对位置及属性不确定性的结合而形成的。由具有相同或相似不确定性值的点连成的等值线是由许多“S-带”图构成的。因此，我们称之“S-带”。）在属性与位置误差不相关的假设下，可以直接应用乘积定理来结合位置与属性不确定性。第二种结合属性与位置不确定性的模型是具有概率解释的确定因子模型。在这种模型中，不确定指标的取值范围由第一个模型的  $[0,1]$  扩展到  $[-1,1]$ 。这对于确定指标值是可能出现负值的推理尤其重要。

本章中给出了“S-带”的理论推导过程，下一步是在计算机原型中实现该模型以表示出该模型在处理 GIS 与遥感中的不确定性时是怎样工作的。第五章将描述该实现过程并且给出一些初步的结果。

## 第五章 不确定性模型的实现

在研究出位置与属性不确定性的理论模型之后，下一个问题是用计算机软件实现这些模型并应用于解决实际问题。GIS 中位置不确定性首先要定量化以使其能够用于数据分析与可视化当中。在某些模型的实现时，首先要对之简化从而能够实现线段、线特征和面状物体不确定性的可视化。为了显示由于遥感分类而引入的属性不确定性，将使用最大似然分类以及概率矢量。并将引入一个具有位置与属性数据及不确定性的叠加例子。叠加后结合的位置与属性不确定性将被定量化地描述为概率值，并用灰度及彩色以及三维显示技术可视化地表达其不确定性。

为了显示不确定性的空间分布，本模型的实现将在栅格环境下完成。假定所用的矢量 GIS 数据可以转化为栅格形式，而栅格 GIS 中的栅格的大小与遥感影像中的一样。假设矢量 GIS 中的一点可以落入相应的栅格 GIS 中的一栅格内，且该栅格的位置即表示原始点的位置。另外，进一步假定由上述假设而造成的误差可以略而不计。

### 5.1 点的不确定性

在本研究中，假定已知 GIS 中点的各种误差且用一个方差、协方差阵表示为

$$\Sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix}$$

对同一类数据，所有的点具有相同的精度。在第四章中设定所有点是相互独立的。例如一类点是行政边界，具有同一类精度，而另一类是建筑物的房角坐标，则具有另一不同的精度。

## 5.2 线段的不确定性

一线段的位置不确定性可以用两种方法表达，即线段置信域以及线段的概率分布。

### 5.2.1 线段的置信域

在 4.4.1.5 节中导出了线段置信区间的理论模型，该模型定义了一个围绕着“量测”线段的区域，线段的真实位置以大于某一置信水平的概率被该区域所包含。该模型不需作任何简化即可以直接实施。置信域是由一系列的、沿线段的矩形所构成的，详见 (4-29) 式。置信域所覆盖的面积与线段两端点的误差，以及定义的置信水平有关。这些被示于图 5-1 和图 5-2 中。

图 5-1 表示了在不同的端点误差情况下，置信域随置信水平变化的情形。端点的方差相等并设为  $\sigma_{xx} = \sigma_{yy} = 1$ 。为了能够看出置信域的大小，在图示上我们夸大了置信域的宽度。每一置信域用同一种灰度表示出来。其中线段两端处的白点表示出量测线段的端点位置。图 5-1(a) 表示出该区域以大于 50% 的概率包含了线段的真实位置。也就是说真实位置落在该域外面的风险小于 50%。类似地 图 5-1 中的 (b)、(c)、(d) 分别表示了置信水平为 90%、95% 和 99.8% 的情况。(b) 的面积是 (a) 的两倍；(c) 的面积是 (a) 的三倍；而 (d) 是 (a) 面积的五倍。在  $\sigma_{xy} = \sigma_{yx} = 0$  的情况下 置信区域的宽度

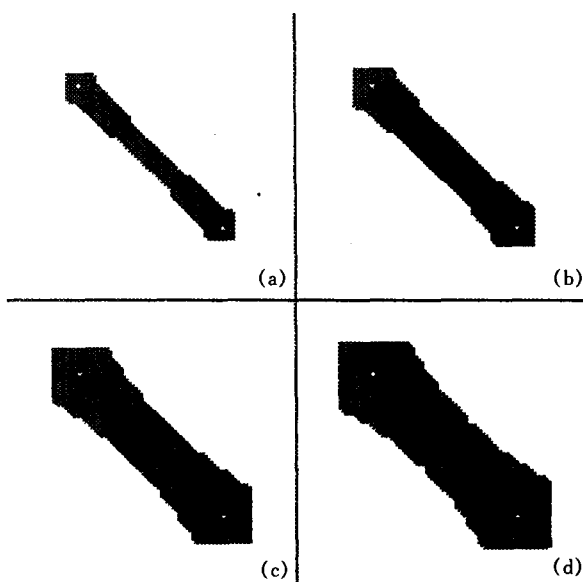


图 5-1 线段的置信域随置信水平而变化

与  $[K_{2,(1+\gamma)/2}]^{\frac{1}{2}}$  成正比。对于置信水平分别为 50%, 90%, 95% 以及 99.8% 的情况下,  $k$  的值分别为 2.77, 5.99, 7.38 和 13.82。

除置信水平之外, 置信域的宽度还取决于两端点的误差。在  $\sigma_{xy} = \sigma_{yx} = 0$  的情况下, 置信域的宽度与  $(\sigma_{xx} + \sigma_{yy})^{\frac{1}{2}}$  成正比。图 5-2 表示出当置信水平为 90% 时, 置信域随端点方差变化的情况。假定  $\sigma_{xx} = \sigma_{yy}$  图 5.2 中的 (a)、(b)、(c) 和 (d) 分别表示了  $\sigma_{xx}^{\frac{1}{2}} = \sigma_{yy}^{\frac{1}{2}}$  为 0.5, 1.0, 2.0 和 3.0 时置信域的形状与大小。

### 5.2.2 线段的概率分布

4.4.1.3 节中讨论了线段的概率分布, 该分布描述了“量测”线

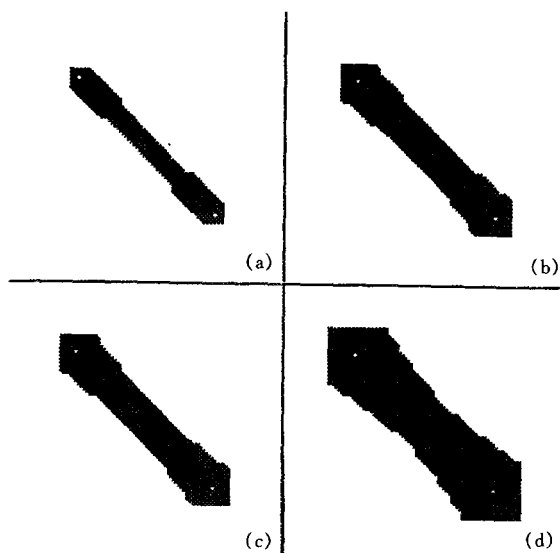


图 5-2 线段的置信域随端点的方差而变化

段围绕着其“真实”位置的分布状况。在 GIS 中，由于通常不知道真实位置，因此只能用真值的估值来代替。

直接实现线段概率分布的理论模型是十分困难的，为了便于实现，用三角分布来代替正态分布。首先将正态分布的范围限制在  $[-3\sigma, 3\sigma]$  之内 因此 99.7% 的误差将发生于此范围之内。

$$\text{设} \quad f(x=0) = \frac{1}{3}\sigma$$

所以三角形内的面积为 1。该近似过程如图 5-3 所示。

第二个引入的近似是分布区域，即一个三角分布所占的区域。原“骨形”的分布区域，经过近似化之后成为由以下几个规则图形所构成的区域 两个三角形和两个梯形 见图 5-4 所示。

在条件  $\sigma_{xy} = \sigma_{yx} = 0$  的条件下，以下两个值需要计算出来以确定

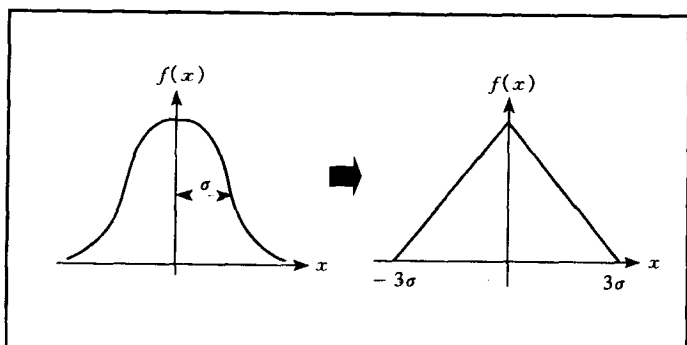


图 5-3 将正态分布近似地表示为一个三角分布

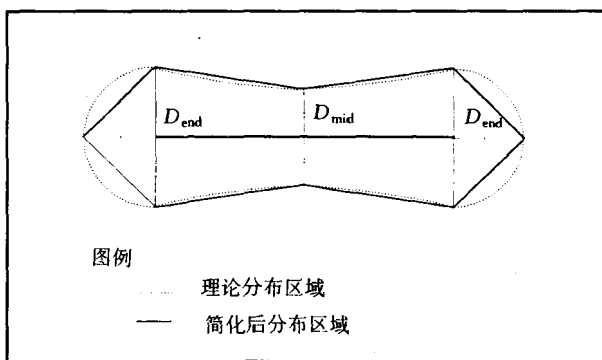


图 5-4 简化后的线段分布区域

近似化以后的分布区域。

$$D_{\text{end}} = [\sigma_{xx} + \sigma_{yy}]^{\frac{1}{2}} \quad (5-1)$$

$$D_{\text{mid}} = 0.707 D_{\text{end}} \quad (5-2)$$

其中  $D_{\text{end}}$  为两结点处的均方差根值，而  $D_{\text{mid}}$  为 midpoint 处的均方差根差。

若一条线段为两相邻物体  $O_1$  与  $O_2$  相邻边界的一部分，而需

要计算一个点属于其中一物体（如  $O_1$ ）的概率时，需进一步地假设。我们近似地认为可以用 GIS 中的量测值代替线段的真实位置。

$$P((x, y) \in O_1) = S_1 \quad (5-3)$$

这里  $S_1$  是图 5-5 中阴影部分所占的面积。如果一个落在线段的尾部且在三角形分布区域内 则概率定义为 (见图 5-5)

$$P((x, y) \in O_1) = S_2 \cdot K^2 \quad (5-4)$$

这里

$$K = d/D_{\text{end}} \quad (k \in [0, 1])$$

以下章节的实验结果均是基于这一假设所得出的。

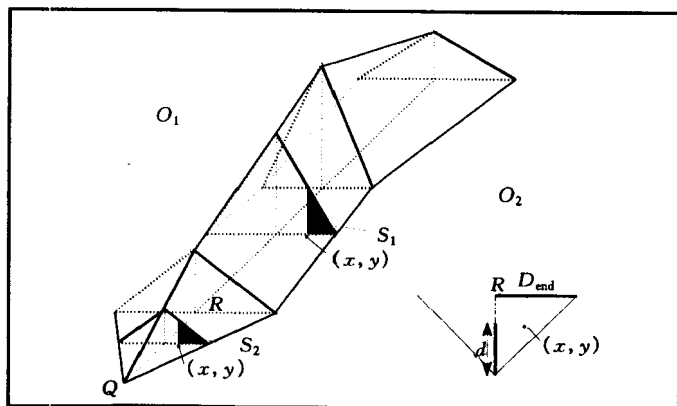


图 5-5 线段的简化概率分布模型

在 4.4.1.4 节中，根据线段的分布模型定义了五个线段误差指标。这些误差指标所对应的分布区域被描述于图 5-6 中。这里  $\sigma_{xx} = \sigma_{yy} = 1.0$ ，而分布的范围被放大了四倍，以增强可视化效果。图 5-6 中的 (a) 图显示了线段可能误差的分布范围，即该区域包含了整体分布区域的 50%，其中两个白点表示线段两端点的“真值”

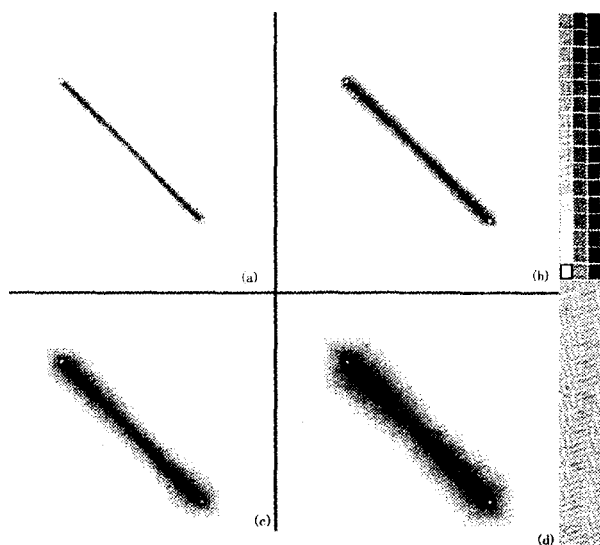


图 5-6 各线段误差指标所对应的概率分布区域

位置。(b) 图显示了线段标准差或均方根差指标所对应的分布范围，该区占据了全体分布区域的 68.25%。(c) 图显示了线段地图标准误差指标所对应的分布范围，它占有整个分布区域的 90%。(d) 图表示了接近确定误差指标所对应的分布范围，即占总分布范围的99.74%。

线段的分布区域随线段两端点的方差而改变，其解析上的关系由 (5-1)式和 (5-2)式表示。图 5-7表示了线段的概率分布范围如何随端点的误差而改变。这里，选用的误差指标是线段地图标准，即 90% 的概率分布区域。其中假设  $\sigma_{xy} = \sigma_{yx} = 0$ ，(a)、(b)、(c)和 (d)图分别表示了  $(\sigma_{xx})^{\frac{1}{2}} = (\sigma_{yy})^{\frac{1}{2}} = 0.5, 1.0, 2.0, 3.0$  时的线段分布状况。为增强可视化的效果，分布图的宽度比实际情况扩大了四倍。

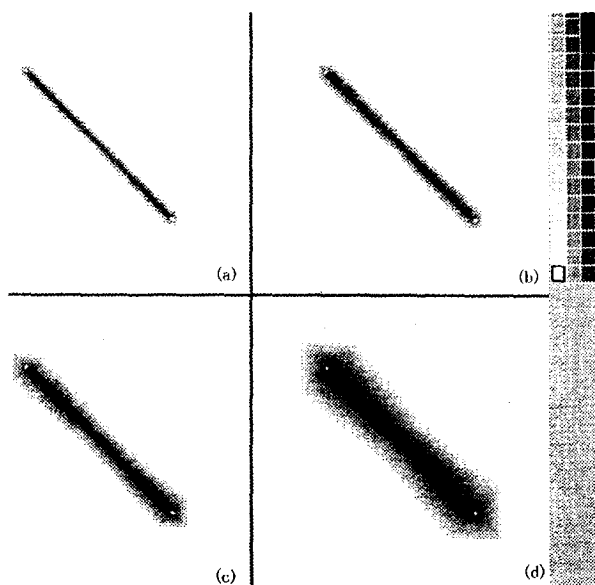


图 5-7 线段的概率分布区域随两端点误差而变化的情况

## 5.3 线特征的位置不确定性

这里，一条线特征定义为由一条以上的线段所构成的。例如，一条线或一多边形的边界就是线特征。在讨论不确定性时，所感兴趣的是：(a)线特征的置信区域；(b)一点属于一物体的概率。

### 5.3.1 线特征的置信区域

线特征的置信区域是所构成的线段的置信区域的并集。我们首先给出一公共的置信水平 比如 90% 对每一线段生成其置信域；之后，用集合论中的并集运算来计算所有线段的置信域的并集，从

而获得线特征的置信域。线特征的置信区域解释成这样一个区域：线特征的真实位置以大于某一置信水平的概率被包含于该区域内。图 5-8 表示了两条线特征的置信区域，分别为三边及五边的多边形。这两个线特征的置信域的置信水平均为 90%， $\sigma_{xy} = \sigma_{yx} = 0.0$ ； $\sigma_{xx} = \sigma_{yy} = (3.0)^{\frac{1}{2}}$ 。图 5-8 中的置信域的宽度放大了三倍以方便观察。

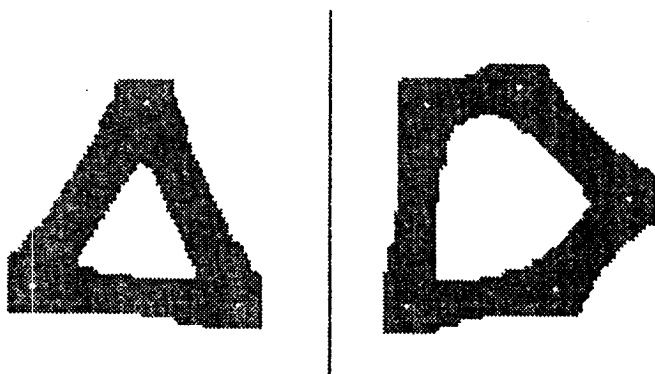


图 5-8 线特征的置信区域，是由所组成线段的置信区域的并集所构成的

### 5.3.2 边界线特征的概率分布

在描述边界线特征的位置不确定性时，重要的是描述所构成线段连接处的位置不确定性。其理论推导过程见 4.4.2 节。基于所得的结果，我们用最小运算计算连接处的概率值。

图 5-9 表示出由两条线段所构成线的线特征的不确定性。概率的分布区域为 95% 的误差指标所对应的区域，并且设对于线段端点  $\sigma_{xy} = \sigma_{yx} = 0, \sigma_{xx} = \sigma_{yy} = (3.0)^{\frac{1}{2}}$ 。对于图 (a), (b), (c) 和 (d) 连接

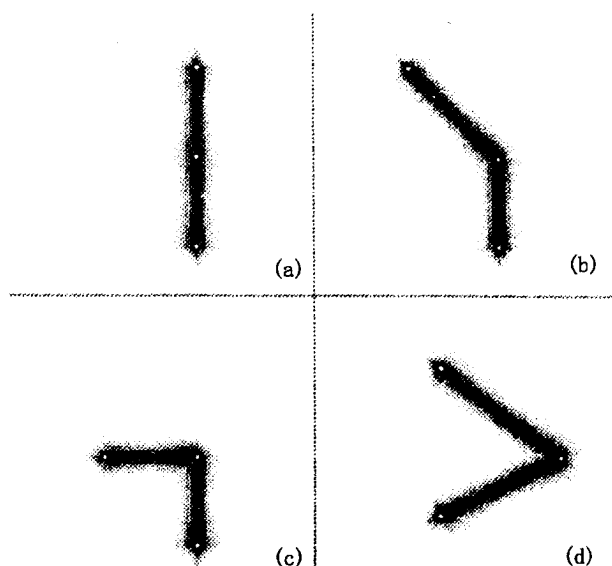


图 5-9 线的概率分布

角度  $\theta$  分别取  $\theta = \pi, 0 > \theta > \pi, \theta = \pi/2$  和  $0 < \theta < \pi/2$ 。图中的白色点表示线段的端点。不同的灰度表示线落在该位置的不同概率。灰度越深，表示概率越大。

图 5-10 中表示出多边形的概率分布。与上面假设一样，即  $\sigma_{xy} = \sigma_{yx} = 0, \sigma_{xx} = \sigma_{yy} = (3.0)^{\frac{1}{2}}$ ，概率分布区域为 95%。其中的白点表示多边形节点的位置。灰度表示线段的真实位置出现的概率。灰度越深，则表示多边形在此位置出现的概率越大。

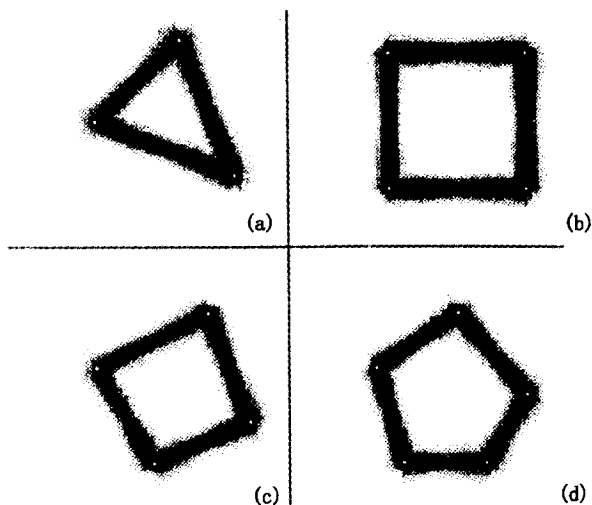


图 5-10 多边形的概率分布

## 5.4 面状物体的位置不确定性

由 4.4.2.2 节所描述，面状物体的不确定性是由其边界不确定性导致的。在描述面状物体的不确定性时，需要确定点属于物体的概率。根据 4.4.2.2 节中的结果，当一个点从物体的外部到内部时，概率值从 0 变到 1。在边界区域内依赖于边界线特征的概率分布，而它是由垂直于所对应线段的方向上的累积分布函数所决定的。图 5-11 给出了四个面状物体不确定性的例子。图中每一像素的灰度表示了所对应点属于面状物体的概率，灰度越深，属于面状物体的概率越大，其中白色的点表示多边形的顶点真实位置的估计值。从该面状物体的灰度分布可以看出，其内部区域是全黑的，即点属于物体的概率为 1，也即不受位置不确定性的影

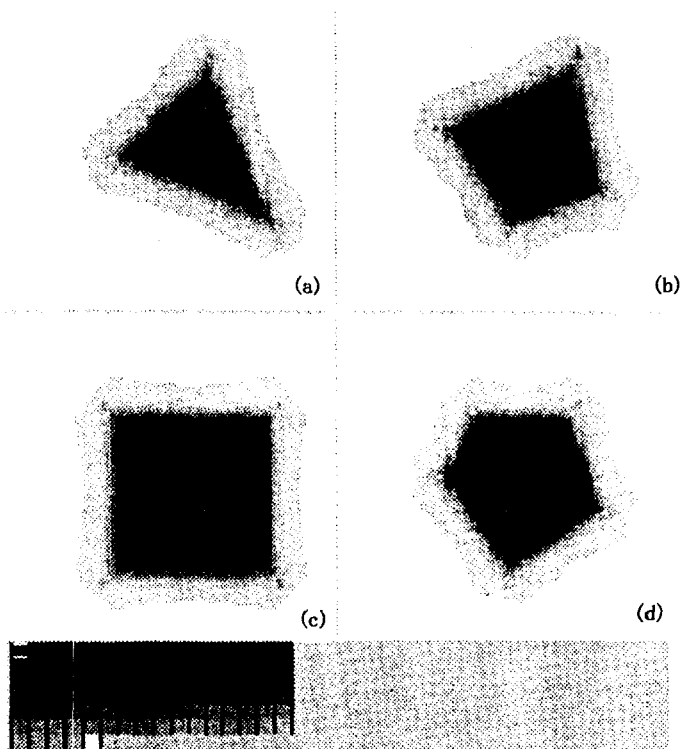


图 5-11 面状物体的不确定性描述

响。当点连续地从内部向外部移动时，在边界区域的灰度连续地从黑变到白，这表明点属于面状物体的概率的不确定性连续地增加。

与 3.2.4 节中基于 Epsilon 带的描述相比 基于概率分布的面状物体不确定性的描述是一种更合理的描述。使用这种方法，点属于面状物体的概率连续地由 0 变到 1 这将适合于定量的空间分析以及属性不确定性的结合的要求。而另一方面，基于 Epsilon 带的描

述只能提供五种定性的描述。

## 5.5 分类影像的属性不确定性

由前面章节所描述，我们可以用概率矢量来描述遥感影像的属性不确定性。ERDAS 系统具有图像处理以及栅格 GIS 的功能。在本研究中将使用 ERDAS 系统的分析及可视化功能。由于 ERDAS 系统不能提供概率矢量，这部分功能是利用新开发的外部程序实现的。需要输入的数据包括遥感影像及训练数据，其中包括方差、协方差阵。而输出结果则为分类影像以及每一个像素的概率值。

首先选用某一地区的 TM 影像。在该地区内的类别包括：城市用地、森林、草地和水。分类之后，对于每一个像素有四个概率值，分别指出该像素属于每一类别的概率。这些值构成概率矢量。例如，一个像素的概率矢量如下：

$$\begin{aligned} [P(x \in \text{城市用地}) = 0.33, P(x \in \text{森林}) = 0.31, \\ P(x \in \text{草地}) = 0.31, P(x \in \text{水}) = 0.05] \end{aligned}$$

根据最大似然分类原理，该像素被分为“城市用地”。这里我们不仅用其分类结果，而且保留分类过程中所产生的概率矢量，则可以表达出分类时所产生的不确定性。根据该概率矢量，可以知道该像素属于每一个类别的概率。在图 5-12 至图 5-19 中，概率矢量被表达成为影像的形式。图 5-12 用灰度表示出每一个像素的最大概率值。灰度值越深，表示概率越大。图 5-13、图 5-14 及图 5-15 分别表示出次大、第三及最小的概率值。

从这组影像结果中，可以看到那些在最大概率的概率影像层中非常黑的地方，相应地在其他概率影像层中的灰度很浅。这些像素所造成的分类不确定性较小。在最大概率影像中灰度很浅的地方，则在次大、第三及最小概率影像中的灰度则较大。这些地方的像素由分类所造成的不确定性则较大。

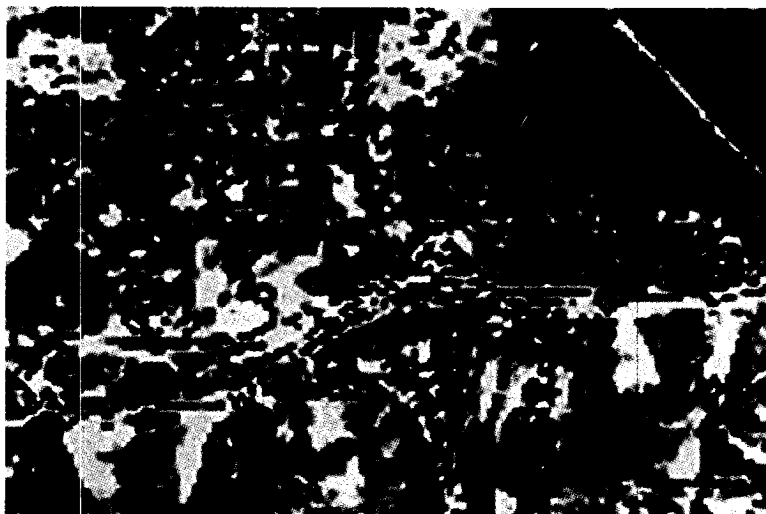


图 5-12 最大概率值影像

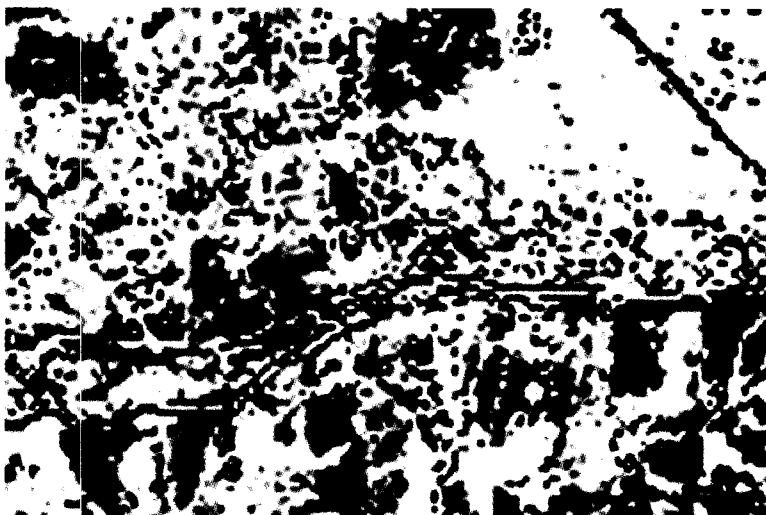


图 5-13 次大率值影像



图 5-14 第三大概率值影像

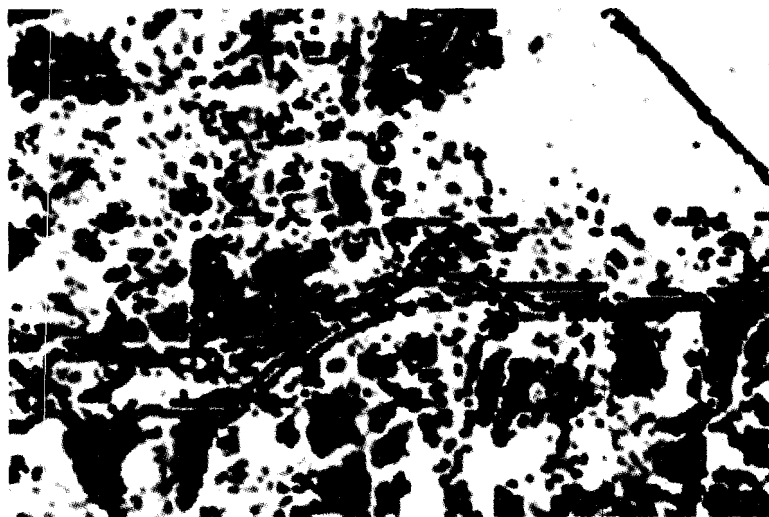


图 5-15 最小概率值影像

## 5.6 位置与属性不确定性的结合

### 5.6.1 问题定义

现在考虑以下不确定性的结合的一个实际问题。我们希望调查某一区域（例如一个县内）土地利用的分布。县的边界数据是从一地图数字化而获得的。土地覆盖分布数据是从一分类遥感影像图中获得。我们希望知道这一县界内各种用地的面积，不仅希望知道面积的大小，而且也感兴趣这些数据中的不确定性。

如果不需要处理不确定性的问题，该问题可用普通的 GIS 叠加分析操作解决。若需要处理其中的不确定性的问题，则需要运用第四章所提出的“S-带”模型来解决该问题。基于面状物体不确定模型可以确定  $P(Z \in O)$  即一点  $Z$  属于物体  $O$  的概率。该不确定性是由数字化地图边界及地图本身的误差所致。而另一方面，分类的误差可由  $P(Z \in C_i) (i = 1, 2, 3, 4)$ ，即该像素属于各类别的概率在最大分类中所生成的概率矢量所表示。

### 5.6.2 位置不确定性的描述

为了表示出结合位置与属性不确定性的影响，现在使用 4.4.4.2 节中所述的乘积定理。在分类遥感影像没有位置误差的前提下模型成立，在实际应用中这个假设不一定总是成立。假定的县界的位置误差示于图 5-16(a) 中。图 5-16(b) 表示出县界内部的不确定性概率，即灰度值越大表示概率值越大。表 5-1 给出了概率值分区间的统计结果。例如该表指出，在总共 1 638 个像素中有 989 个像素的概率值在区间 91% ~ 100% 之内，即该区间内有 60% 的像素以 0.91 至 1.00 的概率属于该面物体。

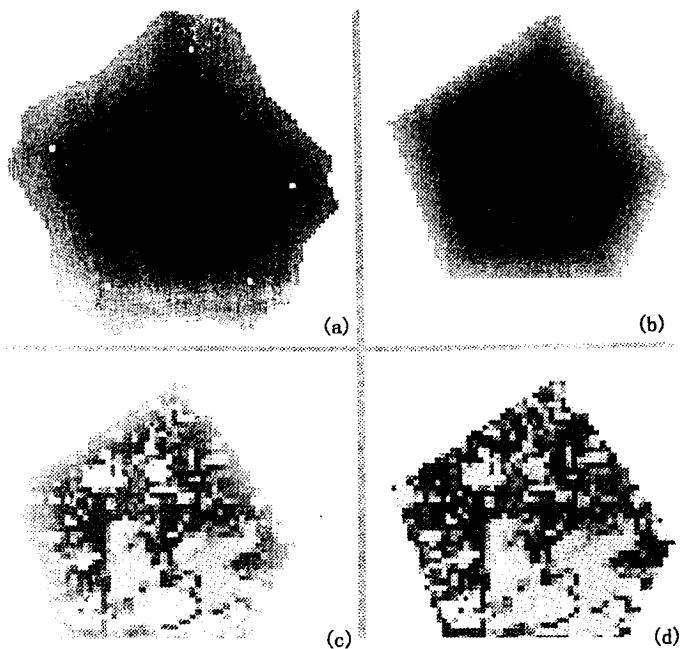


图 5-16 位置不确定性，属于不确定性以及两者的结合

表 5-1 县域边界内位置不确定性的统计结果

概率区间值	0 ~ 10	11 ~ 20	21 ~ 30	31 ~ 40	41 ~ 50	51 ~ 60	61 ~ 70	71 ~ 80	81 ~ 90	91 ~ 100	总数
像素个数	0	0	0	0	20	107	154	170	198	989	1 638

### 5.6.3 属性不确定性的描述

一个分类的遥感影像可用于描述各种类别的空间分布。表 5-2 给出分类后各类别所占有的像素（面积）数。然而这一结果并没有



#### 5.6.4 叠加分析中的不确定性传播

当准备的边界数据具有位置不确定性，而分类的遥感数据具有属性不确定性时，可用(4-49)式来计算其两者叠加时不确定性的传递。结合的结果示于图 5-16(c) 中。比较图 5-16 中的 c) 与 d) 图，在边界区域内叠加之后灰度变浅，即不确定性增加。而在内部区域的灰度没有改变，即位置不确定性并不影响到内部区域的不确定性。叠加后，属性与位置不确定性结合的结果见表 5-4。

表 5-4 叠加后位置与属性不确定性统计结果

概 率 区间值	0~ 10	11~ 20	21~ 30	31~ 40	41~ 50	51~ 60	61~ 70	71~ 80	81~ 90	91~ 100	总数
城市	0	24	42	7	7	15	12	4	2	4	117
草地	0	9	19	13	18	26	25	29	29	57	222
水体	0	4	8	0	2	4	1	0	0	0	19
森林	0	80	344	62	95	93	109	133	124	240	1 280
总数:1 638											

与表 5-3 相比较，可以发现结合后具有较低的概率值的数目增加。例如，“城市用地”类别中在概率区间 91%~100% 之间的像素由 13 减少到 4；“草类”由 138 减少到 57；“水”类由 7 减少到 0；“森林”由 452 减少到 240。另一方面，具有较低概率的像素增加。在概率区间 11%~20%，“城市用地”类别由 0 增加到 24；“草地”的类别由 0 增加到 9；“水”类别像素由 0 增加到 4；“森林”类像素则由 0 增加到 80。这些数据给出在位置与属性不确定性结合之后，不确定性增加的定量描述。

#### 5.6.5 三维与彩色的不确定性可视化技术

在图 5-16 中，不确定性的是用灰度表示的，这是不确定性可视化的常用方法之一。由于人眼对灰度差识别的局限性（如通常少于

64个灰阶)灰度表示法通常不是最好的一种方法。例如在图 5-16 中只能看出少量的像素在 c)和 d)中灰度上有所不同,但是两者在灰度上的差别很难看得十分清楚。因此,本研究提出一种彩色与三维相结合的不确定性可视化方法。

如果只希望可视化一个专题,例如不确定本身,则要直接用三维显示的方法。例如,图 5-17 表示出一个面状物体的位置不确定性。平面坐标表示一个像素的位置,而第三维则表示出该像素的不确定性。高度越高表示其不确定性越小。从本图可以清楚地看到,当一个点从外部移至物体的内部时,其高度增加,即不确定性减少。在内部区域内其高度一致,即确定性值在该区域内为一常量。

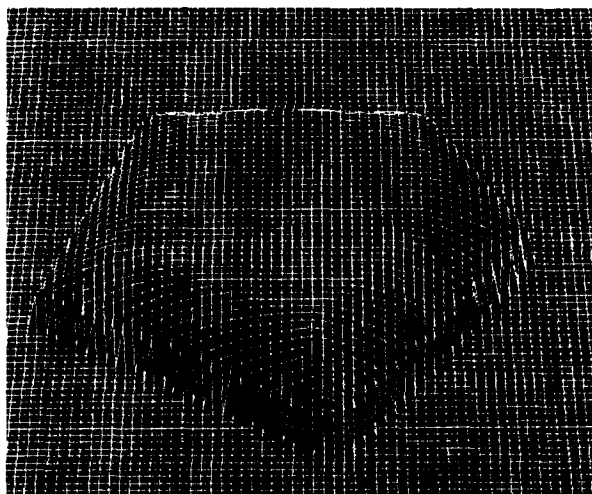


图 5-17 面状物体的位置不确定性

如果不仅希望表达不确定性,而且也要表达相应的专题内容,我们则提出用三维加彩色的方法表示。例如,图 5-18 表示出一分类的结果以及分类时所用的每一像素的概率值。彩色表示类别:红色代表城市用地 绿色代表森林 黄色代表草 蓝色代表水。而其中的

高度代表概率值——不确定性指标。高度越高，概率越大，不确定性越小。

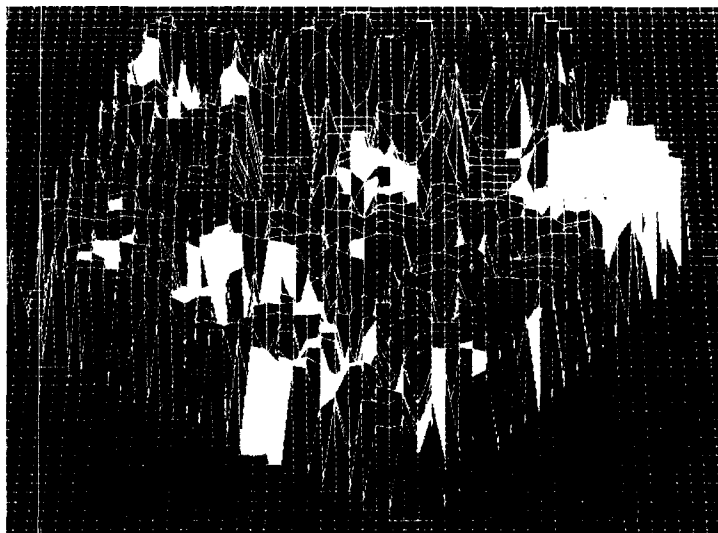


图 5-18 分类影像的专题不确定性

类似地，图 5-19 表示出位置与属性数据不确定性结合后的可视化结果。这里的高度值表示出某一像素属于某一类别（如“水”）并且属于某一面状物体（如某一县界内）的概率。比较图 5-18 和图 5-19 我们可以清楚地看到在叠加分析前后不确定性变化的情况。与图 5-18 相比，不确定性及其变化更明确地表达了出来。从图 5-18 不仅可以看出类别，而且可以该分类的不确定性。例如，在中心附近的两个蓝色的像素被分类为水的概率值很小，即该分类是正确的把握较小。

不确定性及专题可视化的结果对于数据的提供者及使用者均是十分有用的。对于数据提供者，他应该重点检查那些概率值较低的区域，因那里发生误分类的可能性较大。另一方面，对于数据使

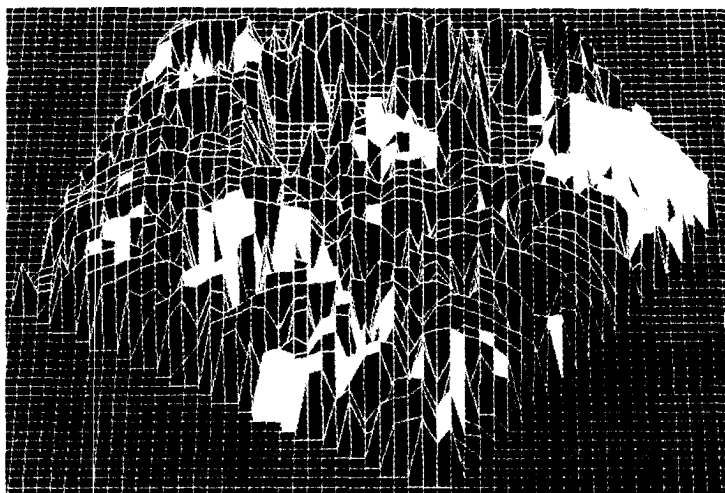


图 5-19 面状物体的位置与属性不确定性  
(彩图见封面)

用者，为了降低由于使用错误数据的风险，对概率降低的数据应赋予较小的权重。

## 5.7 本章小结

本章介绍了对前一章所提出的理论模型进行实现的方法与过程，其目的在于显示不确定模型在 GIS 和遥感环境中如何运用。实现有两个方面的结果：其一是定量的结果，如参数与统计量；另一方面是可视化结果。

线段的位置不确定性模型，包括其置信域及概率分布，可用计算程序来实现。计算结果显示了置信区域的形状以及它与置信水平、线段端点误差在量值上的关系。线段的密度函数由于线段误差指标已指示出，线特征（线及多边形）及面状目标的位置不确定性也

被实现并通过图形表示出来。分类影像的属性不确定性是由概率矢量及四个概率灰度影像表示。位置与属性不确定性的结合是通过用叠加分析方法来实现概率乘积定律。不确定性分析结果表达在统计表格以及概率影像中。

不确定性的可视化是模型实现中的一个重要问题。我们不仅使用了可视化中常用的灰度方法，此外还提出了彩色加三维的方法。本章对灰度可视化方法与三维加彩色的方法进行了比较，后者在可视化方面的效果更好一些。因为它不仅可以清楚地提供专题信息，而且还可以同时提供该专题所对应的不确定性，因此比单一地用灰度的方法可以提供更多的信息。

## 第六章 总结与展望

GIS 与遥感集成中的不确定性问题被 NCGIA 第十二专题研究小组确定为最重要的研究专题，其中一个基本的问题是结合位置与属性不确定性。这是基于 GIS 是一个空间决策支持系统并且具有以多源数据产生信息的功能。本研究的目标是探讨以下问题的理论模型并加以实现机制 ① 位置不确定性的空间分布 ② 属性不确定性 ③ 以上两种不确定性的结合。在一定的理论假设下，以上目标得到了实现。

### 6.1 研究结果总结

#### 6.1.1 地理信息系统中面状物体位置不确定性的建模

为了与属性不确定性相结合，本研究首先对面状物体的位置不确定性进行了建模。该问题进一步分为对以下目标的建模：线段、边界线段、边界线特征以及面状物体，其中线段不确定性的建模是最基本的问题。

本研究首先导出了线段的置信域理论模型。这是一个围绕着“量测”线段所构成的一个区域，该线段的真实位置以大于某一预先给定的置信水平的概率被包含于该区域内。应用线段置信域模型，可以确定模型边界及内部区域。位置不确定性对模糊区域造成影响，在给定置信水平下内部区域不受位置不确定性之影响。与 Epsi-

lon带相比，本研究导出的线段置信域模型给出了置信域的大小、形状与置信水平及端点的端误差的定量关系。导出的置信区域的形状与 Epsilon 带模型有所不同，它是围绕线段的一个两端较大、中间较小的区域。线段置信区域的导出考虑了由于端点的误差所成的影响，且两端点的误差相互独立并且服从二维正态分布。

其次，导出了线段的概率分布。它是由一系列的一维正态分布与两个二维正态分布构成。根据这一分布，边界线段的不确定性被导出。而这是描述边界线特征及面状物体不确定性的基础。

在现有的几何数据精度标准中，人们已经定义了坐标、点的不确定描述指标。在本书中，这些指标被进一步扩展到线段的误差指标。这些指标包括 线段标准差、线段可能误差、线段地图精度标准差、线段接近确定误差以及线段均方根差。这组线段误差指标是基于确定线段的四维的随机矢量导出的。

面状物体的边界的不确定性描述是根据所构成的边界线段不确定性而导出的，其中一个主要问题是解决两线段相连接处的不确定性。面状物体的不确定性是用一个点属于该面状物体的概率所表示的。当一个点连续地从外部移至内部时，该概率值从 0 变到 1。与基于 Epsilon 的面状物体的不确定性相比，该描述提供了一种定量的描述。这为定量分析以及与属性不确定性的结合提供了条件。

### 6.1.2 分类影像的属性不确定性

概率矢量模型用于描述最大似然分类结果所含有的不确定性。基于概率矢量模型，本研究又定义了若干描述属性不确定性的参数：绝对不确定性、相对不确定性、混合程度、以及不完整性参数。概率矢量以及导出的参数被用于描述属性的不确定性及其空间分布。

### 6.1.3 “S-带”模型与 PAT 不确定性

本研究提出的“S-带”模型是用于描述位置不确定性与属性不确定性结合。其应用范围有：(1) 基于多源数据的决策支持及 (2) 空间分析与处理中不确定性的描述。

“S-带”模型在两个层次被定义，其一是理论模型，其二是实用模型。“S-带”的理论模型阐述了处理 PAT 不确定性的一般方法，即分别处理位置不确定性、属性不确定性及其相互关系。根据这一一般性的方法，本研究针对具体的问题，定义了实用“S-带”模型。本研究中讨论了一个比较常见的问题，即在 GIS 中叠加一层仅有位置误差的矢量数据与另一层遥感分类的、具有属性不确定性的影像。在结合这两层数据时则产生 PAT 不确定性。

“S-带”模型主要是在概率论及具有概率解释的确定因子模型的范畴内讨论的。在位置不确定性与属性不确定性相互独立的条件下，这里使用了乘积定理对两种不确定性进行了结合。更进一步地，使用确定因子模型可使得不确定因子的描述指标以概率值从  $[0,1]$  扩展到  $[-1,1]$ ，因此适用于同时具有正、负影响因子的推理。

“S-带”模型提供了一个结合两类不同数据的方法：定量（位置）及定性（属性）数据。通常的算法仅适用定性或定量其中一种数据。而“S-带”模型则是结合这两种数据的一个方法。因此，该方法将基于多源数据的推理和分析，有广泛的应用前景。

### 6.1.4 不确定性模型的实现

模型的实现是在本研究所设计的软件以及现有系统上完成的。为了减少计算量，实现的模型是对原理论模型进行了部分简化与近似处理。在研究中发现常用基于灰度可视化方法有一定局限性。本研究提出了三维加彩色的不确定性可视化方法。使用该方法，用户可以同时清楚地看到两方面的信息：一是专题信息，如土地覆盖

类型；二是不确定性信息，如概率值。这可使用户更加清楚地同时了解到专题内容及相应的不确定性。

## 6.2 研究展望

本研究的范畴为位置与属性不确定性，而逻辑以及时域不确定性尚未讨论。在为 GIS 与遥感中的数据进行统计建模时，为了问题讨论的方便，我们提出了几个假设。例如 GIS 中的边界数据仅有位置不确定性；分类遥感影像仅有属性不确定性；以上两种不确定性是相互独立的。对于 GIS 中的线，我们仅讨论了第三章所定义的第 I 类线。这里还假设 GIS 中仅有随机误差，而不含粗差与系统误差。线段两端点的误差相互独立并且服从正态分布。为了对该领域中的问题进行更全面的探讨，以下几个方面的问题值得进一步研究。

### 6.2.1 理论上的进一步发展

#### (1) 时域与逻辑关系上的不确定性

“S-带”模型提供的是位置与属性不确定性范畴内的问题。但是，这里没有涉及由逻辑上的不一致性所造成的不确定性。更进一步地在时域 GIS 中，我们必须考虑变化所带来的不确定性，即时域不确定性。

#### (2) 位置与属性不确定性的进一步探讨

在本研究中我们假定 GIS 中的面状物体仅有位置不确定性，而分类的遥感影像仅有属性不确定性。然而，在现实世界中，GIS 中的面状物体也可能含有属性不确定性。例如，对某一多边形赋予了一个错的类别。另一方面，分类的遥感影像也可能含有位置误差。例如，由地理配准所造成的位置误差。位置误差与属性误差在某些情况下不一定独立。如何在“S-带”模型中处理这些因素值得进一步研究。

### (3) 进一步研究 GIS 中几何特征的位置不确定性

在本研究中 第Ⅱ类型线（见第三章定义）的不确定性尚未讨论。这里涉及到对人为造成的误差的量化问题。例如，由人工实地确定边界的各点位置量所造成误差即为人为误差。本研究主要讨论了 GIS 中的随机误差，而系统误差和粗差模型尚未讨论。在证明线段的置信区域及其概率分布时，我们假定线段的两端点误差相互独立且服从正态分布。在有些情况下它们不一定独立并且服从正态分布。此时的置信区域及概率分布应是怎样的？该问题值得进一步探讨。

### (4) 发展空间数据的不确定理论

对自然界的理解包含着一定的不确定性。基于 GIS 和遥感的数据分析与决策支持是一个具有不确定性的过程。由于在 GIS 和遥感领域缺乏系统的不确定性理论，这里使用的许多有关不确定性的模型与方法是源于其它领域的模型与方法。几个有关于不确定性的理论可以用于构成空间数据不确定性理论的基础。其中空间数据统计方法是最接近的理论。其它理论包括，例如概率论、证据数学理论、模糊数学、可靠性分析、专家系统中的推理技术等。空间数据的不确定性理论应可以描述物体的不确定性、物体之间的不确定性、空间操作的不确定性、对空间决策支持、空间分析的支持等。

## 6.2.2 “S-带”模型在商用地理信息系统和遥感系统中的实现

通过对“S-带”模型的深入研究 其中包括模型的实现 可以考虑将模型在现有的商用 GIS 软件上实现。有两种方法可用于处理 GIS 中的不确定性 其一是在现有的 GIS 上增加元数据管理模块 另一种方法是用不确定 GIS 来代替现有的 GIS。在目前阶段，第一种方法更为现实可行。我们可以从增加一些已经成熟的误差处理模块于商用 GIS 中着手。

在实现中有两个主要问题要予以考虑：一是数据格式；二是数据量。为了增加元数据，这里指 GIS 中数据的不确定性指标，我们在某些情况下需要对现有的 GIS 数据结构进行修改。一个比较简单的方案就是增加元数据模块，并且建立这些元数据与现有 GIS 数据的链接关系。较为完美的解决方案是在 GIS 数据结构中集成以元数据。元数据包括对物体不确定的描述以及物体之间的不确定关系。物体之间的不确定关系是一个值得进一步探讨的问题。另一个问题是数据量。选用由不确定值导出的不确定指标是减少数据量的一个方法。数据压缩技术、数据结构也是影响到数据量的因素。

### 6.2.3 本研究之应用领域

“S-带”理论模型提出一个结合两种不确定性的原则，不仅仅是属性与位置不确定性。应用领域可以包括多源影像分析、空间分析、基于多源数据的决策支持。“S-带”模型可以提供两个方面的功能支持基于不确定性数据的决策问题以及描述多种数据（如定量与定性数据）结合时的不确定性描述。

## 参 考 文 献

- Amarsaikhan D. 1992. Knowledge-Based Approach to Produce Land Use Map Using Multisource Data (Unpublished manuscript)
- Amrhein C G, Griffith D A. 1987. GIS, Spatial Statistics and Statistical Quality Control. Washington DC: Paper presented at IGIS'87
- Amrhein C G, Griffith D A. 1991. A Statistical Model for Analyzing Error in Geographical Data in an Information System. University of Toronto: Discussion paper, (38)
- Aronoff S. 1985. The Minimum Accuracy Value as an Index of Classification Accuracy. *Photogrammetric Engineering & Remote Sensing*, 51(1): 1687 ~ 1694
- Aronoff S. 1989. *Geographic Information Systems: A Management Perspective*. Ottawa: WDL Publications, 294
- ASP, The Committee for Specification and Standards. 1985. Accuracy Specification for Large-Scale Line Map. *Photogrammetric Engineering & Remote Sensing*, 51(2): 195 ~ 199
- Barrett J F and Nutt M E. 1979. Survey Sampling in the Environmental Sciences: A Computer Approach. Compress, Wentworth: Inc., P. O. Box 102, (N. H) 319
- Berlyon D. 1990. *Information and Data Modelling*. Oxford: Blackwell Scientific Publications
- Berry B J L, Baker A M. 1968. Geographic Sampling. Spatial Analysis: A Reader. *Statistical Geography* (B. J. L. Berry and D. F. Marble eds) Prentice Hall, Englewood Cliffs, (N. J.). 91 ~ 100
- Bierkens M F P, Burrough P A. 1993. Modelling of Map Impurities Using Sequential Indicator Simulation. In: Soares A, ed, *Geostatistics Tróia '92*. Dordrecht: Kluwer, 637 ~ 648
- Blakemore M. 1984. Generalization and Error in Spatial Data Bases. *Cartographica*, 21: 131 ~ 139
- Bolstad P V, Gessler P, Lillesand T M. 1990. Positional Uncertainty in Manually Digitized Map Data. *International Journal of Geographical Information Systems*, 4(4): 399 ~ 412
- Bouloucos T. 1984. Quality Control of Point Determination Systems. ITC lecture notes: ITC
- Bregt A K, Denneboom J, Gesink H J and van Randen Y. 1991. Determination of Rasterizing Error: A Case Study with the Soil Map of The Netherlands. *International Journal of Geographical Information Systems*, 5(3): 361 ~ 367
- Burrough P A. 1993. Soil Variability: A Late 20th Century View. *Soils and Fertilizers* (in press). Next line.
- Burrough P A and Heuvelink G. 1992. The Sensitivity of Boolean and Continuous (fuzzy) Logical

- Modelling to Uncertainty Data. In: Proceedings of EGIS'92, 1032 ~ 1039
- Burrough P A. 1986. Principles of Geographical Information Systems for Land Resources Assessment. Oxford: Clarendon Press
- Campbell J B. 1987. Introduction to Remote Sensing. New York-London: The Guilford Press, 551
- Carter J R. 1989. On Defining the Geographic Information System. In: Ripple, W.J., ed. Fundamentals of Geographic Information Systems: a compendium. ASPRS/ACSM, Falls Church Virginia, 3 ~ 7
- Carver S. 1991. Adding Error Handling Functionality to GIS Toolkit. In: Proceedings of EGIS'91, 187 ~ 196
- Caspary W and Scheuring R. 1992. Error-Band as Measurers of Geometrical Accuracy. In: Proceedings of EGIS'92, 226 ~ 233
- CCOSAM(Canadian Council on Surveying and Mapping). 1982. National Standards for the Exchange of Digital Topographic Data, II-Standards for the Quality Evaluation of Digital Topographic Data
- Chrisman N R. 1989. Modelling Error in Overlaid Categorical Maps. In: Goodchild M F and Gopal S, ed. The Accuracy of Spatial Databases. New York: Taylor and Francis, 21 ~ 34
- Chrisman N R. 1986. Alternatives for Specifying Quality Standards for Digital Cartographic Data. In: Issues In Digital Cartographic Data Standards Report 4. Moellering H, ed. National Committee for Digital Cartographic Data Standards. Columbus, (Ohio): ACSM
- Chrisman N R. 1984. The Role of Quality Information in the Long-term Functioning of a Geographic Information System. Cartographica, 21(2&3)
- Chrisman N R. 1982. A Theory of Cartographic Error and Its Measurement in Digital Data Base. In: Proceedings of Auto-Carto 5
- Cochran W G. 1977. Sampling Techniques. New York: John Wiley and Sons, 428
- Colwell R. (ed) 1983. Manual of Remote Sensing, 2nd ed. Falls Church(Virg.): American Society of Photogrammetry
- Congalton R G, Mead R A. 1983. A Quantitative Method to Test for Co sistency and Correctness in Photointerpretation. Photogrammetric Engineering & Remote Sensing, 49(1): 69 ~ 74
- Congalton R G, Oderwald R G, Mead R A. 1983. Assessing Landsat Classification Accuracy Using Discrete Multivariate Analysis Statistical Techniques. Photogrammetric Engineering & Remote Sensing, 49(12): 1671 ~ 1678
- Congalton R G. 1988. A Comparison of Sampling Schemes Used in Generating Error Matrices for Assessing the Accuracy of Maps Generated from Remotely Sensed Data. Photogrammetric Engineering & Remote Sensing, 54(5): 593 ~ 600

- Cressie N. 1991. *Statistics for Spatial Data*. John Wiley and Sons, Inc. 900 P.
- Mark D M, Csillag F. 1989. The Nature of Boundaries on area-class Maps. *Cartographica*, 26(1): 65 ~ 78
- de Man E. 1988. Establishing A Geographical Information System in Relation to Its Use: A Process of Strategic Choice. *International Journal of Geographical Information Systems*, 2:245 ~ 61
- Desachy J G, Bossi B, Perbos J. 1985. Investigation of Landsat-4 Thematic Mapper Line-to-Line and Band-to-Band Registration and Relative Detector Calibration. *Photogrammetric Engineering & Remote Sensing*, 51(9): 1291 ~ 1298
- Dettinger M D and Wilson J L. 1981. First Order Analysis of Uncertainty in Numerical Models of Ground Water Flow. Part I. Mathematical Development. *Water Resources Research* 17, 149 ~ 161
- Dorst L and Smeulders A W M. 1987. Length Estimators for Digitised Contours. *Computer Vision. Graphics and Image Processing*, 40: 311 ~ 333
- Drummond J E. 1991. *Determining and Processing Quality Parameters in Geographic Information Systems*. Ph D Thesis
- Duda R, Hart P and Nilsson N. 1976. Subjective Bayesian Method for Rule-Based Inference System. In: *Proceedings of National Computer Conference*. AFIPS, 45: 1075 ~ 1082
- Duggin M J, et al. 1985. Systematic and Random Variations in Thematic Mapper Digital Radiance Data. *Photogrammetric Engineering & Remote Sensing*, 51(9): 1427 ~ 1434
- Dutton G. 1984. Geodetic Modelling of Planetary Relief. *Cartographica*, 21: 188 ~ 207
- Dutton G. 1989. Modelling Locational Uncertainty via Hierarchical Tessellation. Goodchild M F, Gopal S, ed. 1989. *The Accuracy of Spatial Databases*. New York: Taylor and Francis.
- Dutton G. 1992. Handling Positional Uncertainty in Spatial Databases. In: *Proceedings of 5th International Symposium on Spatial Data Handling*, 460 ~ 469
- Ehlers M, Edwards G and Bedard Y. 1989. Integration of Remote Sensing with Geographic Information Systems: A Necessary Evolution. *Photogrammetric Engineering & Remote Sensing*, 55(11): 1619 ~ 1627
- EGHM (Editing Group for The Handbook of Mathematics). 1977. *The Handbook of Mathematics*. The People's Publishing House (in Chinese)
- Ehlers M, Greenlee, D Smith T and Star J. 1991. Integration of Remote Sensing and GIS: Data and Data Access. *Photogrammetric Engineering & Remote Sensing*, 57(6): 669 ~ 675
- Footy G M, Campbell N A, Trodd N M and Wood T F. 1992. Derivation and Applications of Probabilistic Measures of Class Membership from the Maximum-likelihood Classification. *Photogram-*

- metric Engineering & Remote Sensing, 58(9): 1335 ~ 1341
- Fukunaga K and Hayes R R. 1989. Effects of Sample Size in Classifier Design. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-11(8)
- Garvey T D. 1987. Evidential Reasoning for Geographic Evaluation for Helicopter Route Planning. IEEE Transactions on Geoscience and Remote Sensing, GE-25(3): 294 ~ 304
- Gibson J R. 1984. Processing Stereo Imagery From Line Images. In: Proceedings 9th Canadian Symposium on Remote Sensing, 471 ~ 488
- Gibson J R, Park W M, Hollinger A B, Dunlop J D and O'Neil N T. 1987. Geometric Correction of Airborne Line Scanner Data. In: Proceedings 9th Canadian Symposium on Remote Sensing
- Ginevan M E. 1979. Testing Land-Use Map Accuracy: Another Look. Photogrammetric Engineering & Remote Sensing, 45(10): 1371 ~ 1377
- Goldberg M, Goodenough D G, Alvo M and Karam G M. 1985. A Hierarchical Expert System for Updating Forestry Maps with Landsat Data. In: Proceedings of the IEEE, 73(6): 1054 ~ 1062
- Goodchild M F and Dubuc O. 1987. A Model of Error for Choropleth Maps, with Applications to Geographic Information Systems. In: Proceedings of Auto-Carto 8, 165 ~ 174
- Goodchild M F and Wang F. 1989a. Modelling Errors for Remotely Sensed Data into GIS, Auto-Carto, 9: 530 ~ 537
- Goodchild M F and Gopal S, ed, 1989b. The Accuracy of Spatial Databases. New York: Taylor and Francis
- Goodchild M F. 1990. Spatial Information Science. In: Proceedings of the Spatial Data Handling Symposium. International Geographical Union. 3 ~ 14
- Goodchild M F. 1991. Issues of Quality and Uncertainty. In Advances In Cartography by Muller J. C
- Goodchild M F, Sun G and Yang S. 1992. Development and Test of An Error Model for Categorical Data. International Journal of Geographic Information Systems, 6(2): 87 ~ 104
- Goodchild M F and Gopal S, ed, 1989b. Accuracy of Spatial Databases. New York: Taylor and Francis, 81 ~ 90
- Gordon J and Shortliffe E H. 1984. The Dempster-Shafer Theory of Evidence: Rule-Based Expert System - The MYCIN Experience of the Stanford Heuristic Programming Project. Buchanan B G and Shortliffe E H, ed. Reading(MA): Addison-Wesley, 272 ~ 292
- Grant M E and Leavenworth R S. 1988. Statistical Quality Control. New York: McGraw-Hill International Editions, 714
- Greenland A, Socher R M, and Thompson M R. 1985. Statistical Evaluation of Accuracy for Digital

- Cartographic Database. In: Proceedings of Auto-Carto 7. Washington DC: ASP-ACSM
- Griffith D A. 1989. Distance Calculations and Errors in Geographic Databases. In: Goodchild M F and Gopal S, ed. The Accuracy of Spatial Databases. New York: Taylor and Francis
- Haemers P B M. 1990. Integration of GIS and Remote Sensing Data: Accuracy Assessment. In: Proceedings of EGIS'90, 428 ~ 436
- Hammersley J M and Handscomb D C. 1979. Monte Carlo Methods. London: Chapman and Hall
- Heckerman D. 1986. Probabilistic Interpretations for MYCIN's Certainty Factors, in Uncertainty in Artificial Intelligence. Kanal L N and Lemmer J F, ed. North-Holland: Elsevier Science Publishers B.V
- Heuvelink G B M. 1993. Error Propagation in Quantitative Spatial Modelling, Application in Geographic Information Systems. PhD Thesis
- Heuvelink G B M, Burrough P A and Stein A. 1989. Propagation of Errors in Spatial Modelling with GIS. International Journal of Geographic Information Systems. 3(4): 303 ~ 322
- Honeycutt D M. 1985. Epsilon Bands Based on Probability. Unpublished Manuscript
- Hord R M, Brooner W. 1976. Land Use Map Accuracy Criteria. Photogrammetric Engineering & Remote Sensing 42(5)
- Jassen L L F and van der Wel FL J M. Accuracy Assessment of Satellite Derived Land Cover Data (Unpublished manuscript)
- Kandel A. 1986. Fuzzy Mathematical Techniques with Applications. Massachusetts: Addison-Wesley
- Kaufmann A. 1975. Introduction to the Theory of Fuzzy Subsets. Academic Press, Inc
- Keefer B J, Smith J L and Gregoire T G. 1988. Simulating Manual Digitizing Error with Statistical Models. In: Proceedings of GIS/LIS'88, 475 ~ 483
- Kennedy-Smith G M. 1986. Data Quality—A Management Philosophy. In: Proceedings of Auto-Carto
- Kish L. 1965. Survey Sampling. New York: John Wiley and Sons. 643
- Klir G J and Folger T A. 1988. Fuzzy Sets, Uncertainty, and Information. New Jersey: Prentice Hall, Englewood Cliffs
- Kuczera G. 1988. On the Validity of First Order Prediction Limits for Conceptual Hydrologic Models. Journal of Hydrology, 103: 229 ~ 247
- Lanter D P and Veregin H. 1992. A Research Paradigm for Propagating Error in Layer-Based GIS. Photogrammetric Engineering & Remote Sensing, 58: 825 ~ 833
- Lee N S, Grize Y L and Dehmad K. 1987. Quantitative Model for Reasoning Under Uncertainty in Knowledge-based Expert Systems: International Journal of Intelligent System, 11(1): 15 ~ 38

- Lee T, Richards J A and Swain P H. 1987. Probabilistic and Evidential Approaches for Multi-Sources Data Analysis; IEEE Transactions on Geoscience and Remote Sensing, GE-25, (3): 283 ~ 293
- Lemmens M J P M. 1991. GIS: The Data Problem. In: Proceedings of EGIS'91. 626 ~ 636
- Li Q H and Tao B Z. 1982. Principle of Probability and Statistics and Their Application in Surveying. Beijing: Chinese Publishing House of Surveying and Mapping (in Chinese), 321
- Lo C P. 1986. Applied Remote Sensing. Longman Scientific & Technical, 293
- Lo K C. 1989. The Evaluation of the Quality of Digital Graphic Data. The proceedings of the Symposium of Geographic Information System in Gaoxiong (in Chinese)
- Lodwick W A, Monson W and Svoboda L. 1990. Attribute Error and Sensitivity Analysis of Map Operation in Geographical Information Systems: Suitability Analysis. International Journal of Geographic Information Systems, 4: 413 ~ 428
- Lodwick W A. 1989. Developing Confidence Limits on Errors of Suitability Analyses in Geographic Information Systems. In: Goodchild M F and Gopal S, ed. The Accuracy of Spatial Databases. New York: Taylor and Francis; 69 ~ 78
- Lodwick W A and Monson W. 1988. Sensitivity Analysis in Geographic Information System. University of Colorado at Denver Paper, 886
- Lunetta R S, Congalton R G, Fenstermaker L K, Jensen J R, McGwire K C and Tinney L R. 1991. Remote Sensing and Geographic Information Data Integration: Error Sources and Research Issues. Photogrammetric Engineering & Remote Sensing, 57(6): 677 ~ 687
- Maffini G, Arno M and Bitterlich W. 1989. Observations and Comments on the Generation of Error in Digital GIS Data. In: Goodchild M F and Gopal S, ed. The Accuracy of Spatial Databases. New York: Taylor and Francis, 55 ~ 67
- Maguire D J, Goodchild M F and Rhind D W. 1991. Geographical Information Systems: Principles and Applications. Longman Science and Technical
- Maguire D J. 1991. An Overview and Definition of GIS. In: Maguire D J, Goodchild M F and Rhind D W, ed. Geographical Information Systems: Principles and Applications. Longman Science and Technical
- Mark D and Caillag F. 1989. The Nature of Boundaries on Area-Class Map. Cartographica, 26(1): 65 ~ 77
- McHarg I L 1969. Design with Nature. New York: Doubleday
- Mead D A. 1982. Assessing Data Quality in Geographic Information Systems. In: Remote Sensing for Resource Management Johannsen C J and Sanders J L ed. Chapter 5, Ankeny(IA): Soil Conservation Society of America, 51 ~ 62

- Mikhail E M and Ackerman F. 1976. *Observations and Least Squares*. New York: IEP-A Dun-Donnelley Publisher, 497
- Moellering H ed. 1987. *A Draft Proposed Standard for Digital Cartographic Data*. Report, 8 Columbus: National Committee for Digital Cartographic Data Standards
- Molenaar M. 1991. *Object Hierarchies and Uncertainty in GIS or Why is Standardisation So Difficult?* (Unpublished manuscript)
- Montgomery D and Peck E. 1982. *Introduction to Linear Regression Analysis*, New York: John Wiley and Sons
- NCDCLS. 1988. *The Proposed Standard for Digital Cartographic Data*. *The American Cartographer*, 15(1): 142
- Newcomer J A, Szaigin J. 1984. *Accumulation of Thematic Map Errors in Digital Overlay Analysis*. *The American Cartographer* 11(1): 58 ~ 62
- Openshaw S. 1991. *Developing Appreciate Spatial Analysis Methods for GIS*. In: Maguire D J, Goodchild M F and Rhind D W, ed. *Geographical Information Systems: Principles and Applications*. Longman Science and Technical
- Perkal J. 1956. *On Epsilon Length*. *Bulletin de l'Academie Polonaise des Sciences*, 4:399-403
- Perkal J. 1966. *On the Length of Empirical Curves; Discussion Paper 10*, Ann Arbor MI, Michigan Inter-University Community of Mathematical Cartographers
- Poorteman K. 1993. *Personal Communication*
- Rao C R. 1965. *Linear Statistical Inference and Its Applications*. John Wiley and Sons, 625
- Reimer J H, Heyada J R, Carpenter S C, Deich W T S and Lee M. 1987. *Avisis Ground Data-Processing System*. *SPIE*, 834(Imaging Spectroscopy II): 79 ~ 90
- Rhode W G. 1978. *Digital Image Analysis Techniques for Natural Resources Inventory*. In: *National Computer Conference Proceedings*, 43 ~ 106
- Richards J. 1986. *Remote Sensing Digital Image Analysis: An Introduction*. New York, London, Paris, Tokyo: Springer-Verlag, Berlin Heidelberg
- Rickman D, Ochoa M C, Holladay K W and Huh O K. 1989. *Georeferencing Airborne Imagery Over New Deltas in Louisiana*. *Photogrammetric Engineering & Remote Sensing* 55(8): 1161 ~ 1165
- Robinson V and Thongs D. 1986. *Fuzzy Set Theory Applied to the Mixed Pixel Problem of Multi-spectral Landcover Databases*. In: Opitz B, ed. *Geographic Information System in Goveret 2*, Deepuk, A Publishing, Hampton Va USA
- Robinson V B and Strahler A H. 1984. *Issues in Designing Geographic Information Systems Under*

- Conditions of Inexactness: Machie Processing of Remotely Sensed Data Symposium, 198 ~ 204
- Rodday G. 1990. Tucson Plans with GIS. *GIS World*, 3(4): 51 ~ 53
- Rosenblueth E. 1975. Points Estimates for Probability Moments. In: *Proceedings of the National Academy of Sciences of the United States of America* 72, 3812 ~ 3814
- Rosenfield G H, 1982. Sample Desing for Estimating Change in Land Use and Land Cover. *Photogrammetric Engineering & Remote Snesing*, 48(5): 793 ~ 801
- Rosefiend G H, and FitzPatrick-Lins K. 1986. A Coefficient of Agreement as A Measure of Thematic Classification Accuracy. *Photogrammetric Engineering & Remote Sensing*, 52(2): 223 ~ 227
- Rosenfield G H. 1986. Analysis of Thematic Map Classification Error Matrices. *Photogrammetric Engineering & Remote Sensing*, 52(5): 681 ~ 686
- Salsig C. 1990. Counting the Errors: Presentation of Aerotriangulation Residuals for Easy Evaluation. *Photogrammetric Engineering & Remote Sensing*, 56(5): 627 ~ 629
- Scarborough J B. 1958. *Numerical Mathematical Anaysis*. 4th ed. The Johns Hopkins Press
- Schmidt M and Wong K. 1985. *Fundamentals of Surveying*. Boston: PWS Publishers.
- Serfling R J. 1980. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, Inc
- Shafer C. 1976. *A Mathematical Theory of Evidence*. Princeton University Press
- Shortliffe E H and Buchanan B G. 1975. A Model of Inexact Reasoning in Medicine. *Mathematical Biosciences*, 23, 351 ~ 379
- Slama C C. 1980. *Manual of Photogrammetry*. 4th ed. American Society of Photogrammetry, Falls Church(Virginia), 1056
- Snedecor G W and Cochran W G. 1989. *Statistical Methods*. 8th ed. Iowa State University Press/ AMES, 503
- Srivastava M S and Carter E M. 1983. *An Introduction To Applied Multivariate Statistics*. Elsevier Science Publishig Co., Inc
- Star J L, Estes J E and Davis F. 1991. Improved Integration of Remote Sesting and Geographic Information Systems: A Background on NCGIA Initiative 12. *Photogrammetric Engineering and Remote Sensing*, 57(6): 643 ~ 645
- Stoms D. 1987. Reasoning With Uncertainty In Intelligent Geographic Information Systems, *GIS'87*, 693 ~ 700
- Story M and Congalton R G. 1986. Accuracy Assessment: A User's Perspective. *Photogrammetric Engineering & Remote Sensing*, 52(3): 397 ~ 399
- Stutheit J ed. 1990. *The 1990 GIS Sourcebook*. GIS WORLD Inc. Fort Collins(Colorado)

- Tempfli K. 1980. Spectral Analysis of Terrain Relief for the Accuracy Estimation of Digital Terrain Models. *ITC Journal*, 3: 478 ~ 510
- Thapa K and Bossler J. 1992. Accuracy of Spatial Data Used in Geographic Information Systems. *Photogrammetric Engineering & Remote Sensing*, 58: 835 ~ 841
- Till S M. 1987. Airborne Electro-optical Sensors for Resource Management. *Geocarto International*, 3: 13 ~ 23
- Tomlin C D. 1990. *Geographic Information Systems and Cartographic Modelling*. Prentice Hall, Englewood Cliffs, N.J. 07632
- Tomlin C D. 1991. Cartographic Modelling. In: Maguire, D J, Goodchild, M F and Rhind, D W, 1991, *Geographical Information Systems: Principles and Applications*. Longman Science and Technical
- Van der Knaap W G M. 1992. The Vector to Raster Conversion: (mis)Use in Geographical Information Systems. *International Journal of Geographical Information Systems*, 6(2): 159 ~ 170
- Van der Wel F. 1991. The Integration of Remotely Sensed Data into a GIS: Estimation of the Accuracy. In: *Proceedings of EGIS'91*, 1219 ~ 1227
- Van Genderen J L and Lock B F. 1977. Testing Land Use Map Accuracy. *Photogrammetric Engineering & Remote Sensing*, 43(9): 1135 ~ 1137
- Veregin H. 1989. Error Modelling for the Map Overlay Operation. In: Goodchild M F. and Gopal, ed. 1989. *The Accuracy of Spatial Data Base*, New York: Taylor and Francis
- Veregin H. 1989. A Taxonomy of Error in Spatial Databases. National Center for Geographic Information and Analysis. Technical Paper 89 ~ 12: 11
- Webster R and Oliver M A. 1990. *Statistical Methods in Soil and Land Resource Survey*. Oxford University Press
- Welch R, Jordan T R and Ehlers M. 1985. Comparative Evaluations of the Geodetic Accuracy and Cartographic Potential of Landsat-4 and Landsat-5 Thematic Mapper Image Data. *Photogrammetric Engineering & Remote Sensing*, 51(9): 1249 ~ 1262
- Zadeh L A. 1965. Fuzzy Sets, *Information and Control*, 8
- Zhang G Y and Tulip J. 1990. An Algorithm for the Avoidance of Sliver Polygons and Clusters of Points in Spatial Overlay. *Proceedings of 4th Spatial Data Handling*, 141 ~ 150

# 前 言

本书主要介绍了地理信息系统（GIS）与遥感（RS）中不确定性分析的理论与方法。该领域是最近一个时期 GIS 界的主要研究热点之一，主要原因是到目前为止尚未有成熟的、系统的理论与方法，可用于处理 GIS 中的不确定性。

近年来，中国在 GIS 的理论研究及产业化发展方面都进入了一个活跃的阶段。与国外许多国家一样，有许多成功发展 GIS 的范例，同时也存在着发展中的问题。GIS 中的数据质量，即不确定性问题，是许多不成功系统的主要原因之一。

作者希望通过本书向国内同行提供 GIS 中不确定性分析方法方面一本新的参考文献。并希望更多的人了解到 GIS 中不确定的一面，从而在 GIS 的发展方面，可以更加全面、系统地考虑 GIS 的系统设计与开发。

本书的出版得到了佳能测量师行及香港空间信息顾问公司的资助，作者深表谢意！

由于作者水平有限，望广大读者在阅读本书时能提出宝贵意见，不胜感谢！