

水库淹没损失遥感估算方法探讨

——以三峡水库石宝寨试验区为例*

周福群

(中国科学院成都地理研究所)

1985年12月12日收到

我国水能资源十分丰富,但目前仅开发了可供开发利用的5%左右。在发展水电时,淹没损失的估算是一个很重要的问题。而采用常规的野外作业方式进行淹没损失的调查,不仅野外工作量大,要花很大的人力物力,而且周期长,已不能满足水电发展的需要。为此,在二滩电站的淹没损失估算中,中国科学院遥感应应用研究所曾采用过遥感分析的方法,并取得了开发性的成果。本文试图用航天遥感方法,对水库淹没损失进行估算,旨在为这类工作探讨一种新的方法。整个工作流程见图1。

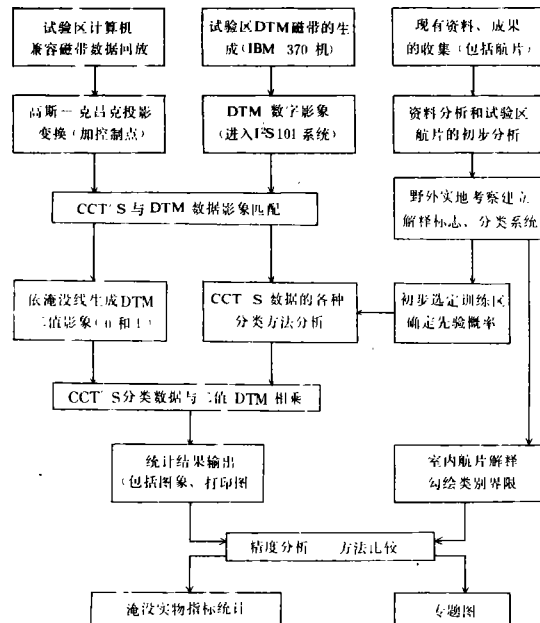


图1 水库淹没损失卫星遥感估算工作框图

Fig. 1 Flowchart of Estimating Reservoir Submergence Loss from Landsat MSS

* 本文是在陈昱导师指导下完成的,试验过程中还得到了四川大学陈祯培副教授和本所高世忠等老师的热情指导,在此一并表示感谢。

本文作者现在西南交通大学航地系遥感教研室工作。

一、淹没损失遥感估算的方法论探讨

从地理信息系统的角度看,遥感数据可作为现有地面地物以及相关地物的信息源之一。考虑到地理信息系统是地理学的一个发展趋势,本文试图借助于地理信息系统中地形因子信息和地表遥感信息的综合分析,来进行三峡水库石宝寨试验区淹没损失遥感估算方法探讨。

在淹没损失估算中,其关键是两个精度问题:即地表实物的分类精度和淹没边界线的精度。前者决定淹没区土地利用、土地覆盖状况,后者决定淹没区的范围以及淹没实物指标的统计。因此,这两个精度便是淹没损失遥感估算的焦点。很显然,利用卫星遥感数据作为实物本底调查的信息源,然后利用地形因子确定淹没边界线,将两者结合起来,是能完成淹没损失估算的。

二、试验区的选择及分类系统

1. 试验区的选择

长江三峡水利工程,拟以三斗坪为坝址。若以水库正常蓄水位 150 米的低坝方案而

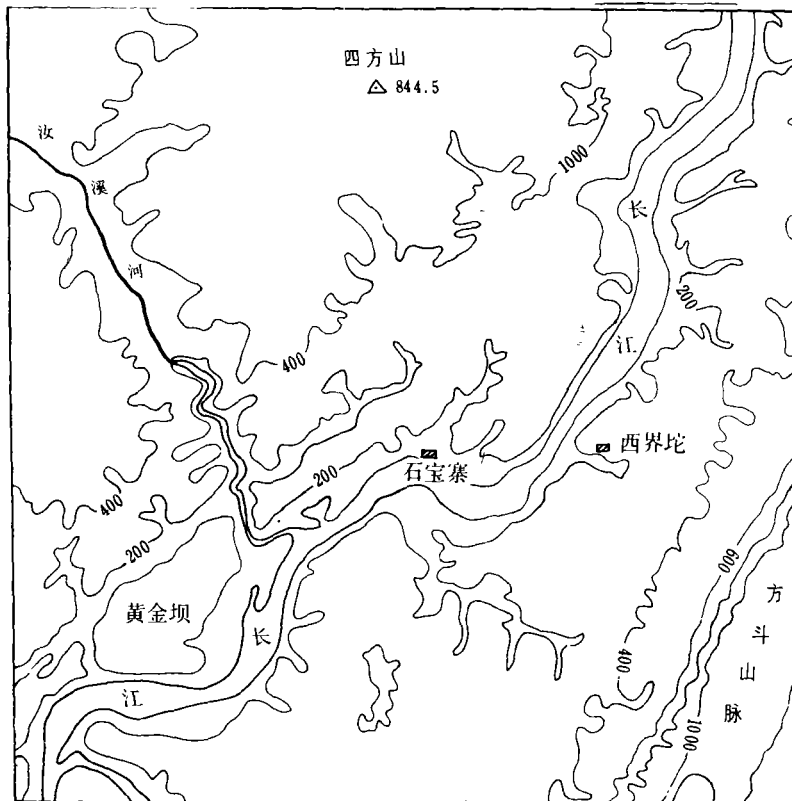


图 2 试验区地形图

Fig. 2 The Topographic Map of Experiment Area

论,水库的迴水区在四川省长寿县附近。从坝址到迴水区,沿江淹没长度达500多公里。限于人力物力及时间的关系,本文只在淹没区内选择了一个边长大约为21公里的区域,即石宝寨试验区进行试验,该区主要部分位于四川省忠县境内。

2. 试验区的地形地貌特征

忠县属川东平行峪谷的一部分,东南边境有方斗山,西北边境有精华山,长江北面的挖断山,成为县内前后乡的自然分界线,各山之间几乎全为丘陵。后乡多为台状或圆形包状深丘,其间有零散的冲积或侵蚀的槽状谷地及平坝;前乡因溪流切割和雨水冲蚀,多呈“V”形谷地及陵峭深丘,沿长江两岸偶有平缓阶地、滩地及江中沙坝。试验区就位于前乡的沿江两岸(图2)。

3. 分类系统

土地利用和土地覆盖分类系统,是以《土地利用现状调查技术规程》为原则,结合试验区的区域特点、卫星影像的地面分辨率等状况而确定的,相当于上述规程中的二级类型。

三、试验区卫星影像数据的处理

1. 高斯-克吕克投影变换

陆地卫星扫描是多中心投影,加上卫星飞行姿态的随机变动、扫描状态的改变等诸因素的影响,使得卫星影像没有一个统一的地理座标。为使它能与同地区的以高斯-克吕克投影为基础的数字地形模型相匹配,也为了处理结果的准确性,必须对原始数据进行几何投影变换,该工作是在 I²S101 系统上使用高斯变换功能(>KRUEGER)完成的。为保证精度,加了地面控制点。具体步骤如下:

(1) 确定投影方式 高斯-克吕克投影。

(2) 选择控制点 选择出理想、准确的地物控制点是成功的关键,控制点还应保证是在进行几何投影变换区域所处的六度带内。

(3) 几何投影变换的数学关系式 本文采用常用的多项式拟合,

$$\left. \begin{aligned} u &= f_1(X, Y) \\ v &= f_2(X, Y) \end{aligned} \right\} \quad (1)$$

式中: u, v 表示控制点的座标; X, Y 表示控制点在卫星影像上的行、列数。

本文选取了七个控制点,用(1)式多次拟合,最后剔出三个拟合误差较大的点,选定四个分布大致均匀、拟合误差较小的点作为确定变换系数的原始数据。四对点的拟合情况见表1。为了使以后卫星影像数据与数字地形模型相匹配,还将图像旋转到正北,经二次采样,像元大小为 80×80 米²。

2. 试验区土地利用、土地覆盖在 I²S101 上的自动分类

(1) 集群分类 规定土地覆盖类别数为8,迭代次数为8。

(2) 最大似然比分类 高斯正态模型中的均值和方差参数,通过训练区统计给出,其余参数由该功能的隐含值给出。

3. 模糊分类

所谓模糊性,主要是指客观事物的差异在中介过渡时所呈现的“亦此亦彼”性。在遥

表 1 高斯变换控制点拟合误差表

Table 1 Fitting Errors of Gauss Transformation Control Points

控制点号	在卫星影像中的行、列数		高斯座标(公里)		拟合误差 (像元)	平均误差 (像元)
	x(抽样)	y(行)	u	v		
1	276	8	3395.4	237.9	0.68	1.75
2	360	71	3389.6	241.3	1.47	
3	182	378	3367.8	224.8	2.88	
4	558	313	3368.5	247.2	1.99	

感图像的分类识别中,就存在着许多模糊性。这主要表现在代表地物的光谱模糊性上。一幅图像可以认为是一个模糊集合。由于人脑具有非凡的处理模糊现象的能力,所以在目视判读中,人能借助于推理、判断、分析以及相关知识的利用等逻辑思维活动,从模糊的现象中抽取出具体的内容来。与此相仿,自动识别利用模糊数学提供的模式,也能解决一定的模糊问题。

(1) 模糊模式识别

设论域为图像场

$$u = \{X | X^T = (x_1, x_2, x_3, x_4)\}$$

又设第 i 类地物在论域 u 上的模糊集为 A_i , 隶属函数为

$$A_i(X) = \frac{1}{4} \sum_{j=1}^4 A_{ij}(x_j) = \frac{1}{4} \sum_{j=1}^4 \exp\left[-\left(\frac{x_j - a_{ij}}{b_{ij}}\right)^2\right] \quad (2)$$

式中: a_{ij} ——第 i 类第 j 分量的均值; b_{ij} ——第 i 类第 j 分量的均方差; $i = 1, 2, \dots, n$ 类别序号; $j = 1, 2, 3, 4$, 波段序号。

本判别用最优从属原则。

设, $\forall X \in u$

$$\text{若 } \tilde{A}_{i_0}(X) = \bigvee_{1 \leq i \leq n} \tilde{A}_i(X)$$

则将 X 判别为 i_0 类,即将 X 归属于 \tilde{A}_{i_0} 模糊集。

计算时还可预先规定一阈值 T ,

若 $\tilde{A}_{i_0}(X) \geq T$, 便将 X 归入 i_0 类,

若 $\tilde{A}_{i_0}(X) < T$, 拒判。

(2) 混合分类法

本文选用了两个子区($30 \times 20, 20 \times 20$),共 1000 个像元进行模糊聚类,得出 8 个聚类中心及方差。然后用模糊模式识别,求出分类结果。该方法称为混合分类法。

模糊聚类所用的隶属度和聚类中心,由下两式表达^[1]:

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_j - v_i\|}{\|x_j - v_k\|} \right)^2} \quad (3)$$

$$v_i = \frac{\sum_{L=1}^n (u_{iL})^2 x_L}{\sum_{L=1}^n (u_{iL})^2} \quad (4)$$

式中: c 为分类数, n 为像元数。

(3)和(4)式是一个反复迭代的过程。

首先用典型区域作聚类分析,尔后用其产生的聚类统计文件用较好的模式将整个区域进行识别。已有文章披露,其结果会比其它方法为好。本文认为这种方法有一定的优点,尤其是对地表复杂的地区。在我国许多地区作土地覆盖分类时,若用有监督模式,一般不易找到理想的训练区,因而影响分类精度,采用此法便能克服以上缺点,尤其在山区、丘陵等破碎的区域,更为优越。

四、试验区数字地形模型的建立及与卫星数据的匹配

1. 数字地形模型 (DTM) 的建立

为了在淹没损失估算中,量化、数字化地利用地形因子,便将自然界连续变化的地形,用离散的数字地形模型模拟出来。本文所建数字地形模型以 80×80 米的矩形网格为结构,以 1:5 万的地形图(高斯-克吕克投影)为底图,人工采集高程数据,该数据在 IBM/370 通用计算机上生成数字高程模型。该模型进入 I'S101 系统后,成为数字高程影像。有了数字高程矩阵,便能进行地形要素的分析。本文用的地形要素主要有两个:

(1) 绝对高程 绝对高程定义为网格中心点的高程 H 。

(2) 坡度 网格中心点流水线方向的坡度为该网格的坡度。

设 $F_K = \arctg[(H_{ij} - H_K)/D_K]$

于是 坡度 = $\max F_K \quad K = 1, 2, \dots, 8$

式中: H_{ij} 为点 (i, j) 的高程; H_K 为点 (i, j) 的相邻点的高程; D_K 为点 (i, j) 至相邻点的水平距离。

2. 数字地形影像与 CCT'S 数据影像的匹配

从理论上讲,以高斯投影为基础的数字地形影像与经过高斯投影变换的卫星数据是完全能匹配起来的,因为两者都具有统一的地理座标。本文所采用的匹配方法可称为控制点法。方法是分别在卫星影像中和数字地形影像中找到同名地物点(控制点),经坐标平移便完成两个影像的匹配(本试验区域不大)。匹配精度分析见表 2。

据分析,匹配误差达两个多像元的主要原因,是研究地区太破碎,控制点难找准,引起一系列的系統误差,如导致高斯变换的误差,在平移时又造成平移计算的误差,最后导致匹配误差。

3. 数字地形模型的应用

(1) 确定水库边界

卫星影像与数字地形影像匹配后,每一像元既具有地表信息又具有地形信息。可以说,卫星影像的主题模型已经建立起来了,因此,根据高程,便能在卫星影像中确定出淹没

边界来。

表 2 地形影像与卫星影像匹配精度分析表
Table 2 Registration Accuracy of DTM and Landsat MSS

控制点号	在地形影像中的行列值		在卫星影像中的行、列值		匹配误差 (像元)	平均误差 (像元)
	S(抽样)	L(行)	S(抽样)	L(行)		
1	-169*	250	-166	252	3.606	2.52
2	-96	292	-98	294	2.828	
3	170	366	173	364	1.414	
4	277	39	275	38	2.236	

* 负号表示朝反方向超出数字地形影像范围。

由于数字地形模型是自然界连续地形的离散表示,采用格网化的数字地形,其抽样点阵是跳跃起伏的,要在这样的模型中求出精确的淹没边界将是困难的,尤其是边长较大的网格,困难更大,甚至不可能。

本文采用如下方法来确定边界线,尔后,经过面积修正,求得相对准确的结果。

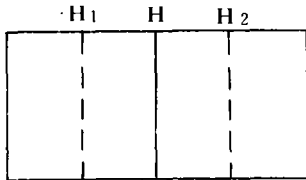


图 3 边界线左右的网格
Fig. 3 Left and Right
Grids of Boundary

如图 3 所示,设格网边长为 Δx ,假设实际淹没高程为 H ,计算出在此高程附近的平均坡度 a ,于是

$$H_2 = H + \frac{1}{2} \Delta x t g a$$

以 H_2 为高程的网格刚好超过实际淹没高程 H 一个格网,而由高程

$$H_1 = H - \frac{1}{2} \Delta x t g a$$

确定的网格刚好在淹没边界之内(图 3)。

由于在数字地形模型中,相邻网格的高程存在于 H_1 至 H_2 高程范围内,那么,凡高程小于 H_2 的格网都作为淹没区域处理。这样可以看出,从概率的意义上说,边界像元的误差为半个像元(这半个像元的误差,在以后统计面积时,作了专门修正)。

(2) 统计淹没实物指标

以 H_2 为临界值,使地形影像成为二值图像,0 和 1。然后去乘分类结果,将各类非零像元分别相加即可。混合法分类结果见表 3 第一行。

表 3 混合法淹没损失估算实物指标(150 米) 单位: 像元
Table 3 Submergence loss index estimated with the hybrid (150m.) unit: pixel

项 目 \ 类 别	水 田	旱 地	有 林 地	疏 林	水 体	滩 深
原始分类结果	2072	1967	48	74	2560	705
坡度因子修正	2100	2049	51	105	2571	735
网格面积修正	1949	1902	47	97	2571	682

五、面积修正

1. 由坡度因子引起的面积修正

根据投影原理可知,水平投影面对倾斜面的校正系数为 $1/\cos a$ 。 a 为该像元对应的地形坡度,修正后的结果见表3第二行。

2. 由网格计算面积产生误差的统计修正

利用一个地区的地理信息系统确定其中的一个区域面积,通常的做法是,输出该面积区域边界内的网格数,其总面积等于网格数乘以每一网格所代表的面积(本文便是如此),但是,要计算的面积边界线,并非刚好全部通过网格的边界,其中必然有很大的部分要穿越网格,因此在计算时,凡包括有边界线的网格都要计算上,那么面积必然增大,反之减小。考虑到上述两种情况,这里提出一种面积修正方法。

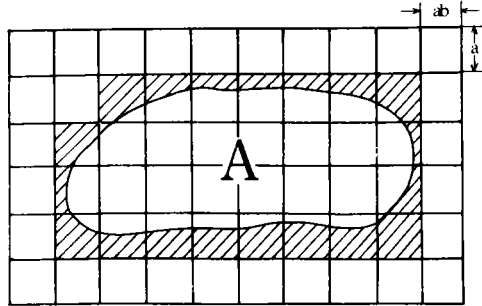


图4 计算面积 A 引起的误差(阴影区)

Fig. 4 Error Incurred in Calculating an Area A (Shaded)

如图4所示,欲求不规则图形 A 的面积,它的边界线横跨过许多网格,阴影区便为计算面积 A 所带来的误差。设网格面积为 $a \times ab$, 借助 Goodchild 等^[2]推论的一些公式,可进行如下推导。

从统计学角度看,一个区域的边界线越过任一网格的方向,长度是随机的。所以可以这样认为, N_b 个边界线像元中,边界将网格分为两部分,凡在边界线外的部分即为误差。由于随机性,这个误差有大有小,设其服从正态分布,其平均误差是半个像元。设网格面积 $S = a \times ab$, 总误差为 ΔA , 于是可得:

$$\Delta A = \frac{S}{2} N_b = \frac{S}{2} \cdot \frac{a K_1 \sqrt{\pi A}}{\bar{L}} = \frac{S K_1 \sqrt{\pi A}}{\bar{L}}$$

式中: K_1 为所求面积的形状因子,区域为圆形时, $K_1 = 1$; \bar{L} 为一网格内边界线的平均随机长度,区域为方形、边长为80米时, $\bar{L} = 63.48$ 。

若 N_1 是区域内包括边界线网格的总网格数,于是,

$$A = N_1 S - \Delta A = N_1 S - \frac{S K_1 \sqrt{\pi A}}{\bar{L}} \quad (5)$$

若 N_2 是区域内不包括边界线网格的总网格数,那么,

$$A = N_2 S + \Delta A = N_2 S + \frac{SK_1 \sqrt{\pi A}}{L} \quad (6)$$

(5)和(6)式均可使用迭代法计算出 A 值,于是面积修正值为

$$\Delta A = N_1 S - A \text{ 或 } \Delta A = A - N_2 S \quad (7)$$

3. 统计面积修正在本课题中的应用

本文对混合法分类结果作了统计面积修正。由于淹没损失是处在沿江两岸狭窄的地带,故可将面积形状视为长方形。经(7)式计算后,在 150 米方案下,总修正值是 362 个像元。再经比例分配到各类中去(除水体外),其结果见表 3 第三行。

4. 面积计算精度

试验区的面积量测用日本产 KP-92 型数字求积仪在 1:5 万的地形图上沿等高线进行。200 米等高线之下的面积 $S_{\text{平}}$ 为

$$S_{\text{平}} = 6545 \times 10^4 \text{ 米}^2$$

由计算机统计经过面积修正的淹没损失(200 米)总像元数为 10564 个,其相应面积 S_{H} 为

$$S_{\text{H}} = 6760.96 \times 10^4 \text{ 米}^2$$

S_{H} 对于 $S_{\text{平}}$ 的相对误差为

$$\Delta S_{\text{H}} = \frac{S_{\text{H}} - S_{\text{平}}}{S_{\text{平}}} = \frac{(6760.96 - 6545) \times 10^4}{6545 \times 10^4} = 3.3\%$$

因此, S_{H} 之值是可靠的。 S_{H} 大于 $S_{\text{平}}$, 是因为对 S_{H} 作了倾斜面积改正, 而 $S_{\text{平}}$ 只是水平投影面积。

5. 淹没实物与高程的相关分析

由于本文涉及的区域不大, 试验区域内建筑物不多, 而且分散, 在卫星影像上无十分明显的信息特征, 所以本文未作单独处理, 而归入了其它类别或拒判。关于人口估计和社会经济方面的情况, 牵涉其它问题, 如地面调查等, 文内也未作考虑。换句话说, 为使问题简化, 本文只对土地资源的淹没损失作了估算。

本文对 150 米, 170 米和 200 米三个正常蓄水位作了淹没损失估算。据此, 我们建立了高程-淹没损失曲线, 由该曲线可以查得任一高程相对应的淹没损失实物指标。这对工程决策时进行方案的比较选择是极其方便和有利的。也有利于数学上的优化处理。

一般设淹没实物与高程的相关分析式为:

$$y = f(x)$$

式中, x 为高程; y 为淹没损失实物指标。

$f(x)$ 的具体函数形式应根据研究地区的具体情况而定, 根据所得的几组数据 (x_i, y_i) , $i = 1, 2, \dots, n$, 在直角坐标系中标出, 目视 y 随 x 的大致变化情况, 确定其函数形式后, 根据最小二乘法原理, 解算系数, 最后求出 $y = f(x)$ 的统计表达式。

本文通过高程与相对应的淹没损失, 计算得出淹没旱地与高程的关系式为:

$$y = 36.21x - 3548.4$$

淹没水田与高程的关系式为:

$$y = 1833.8 \ln x - 7215.4$$

前者为线性关系,后者为对数关系。

图5为根据上述两式绘出的高程与淹没实物总和的相关曲线图。其它各种类型的淹没指标,都可与高程建立相关函数式。图4的回归方程为:

$$\begin{aligned} y &= a + bx \\ &= 66.195x - 2662.42 \end{aligned}$$

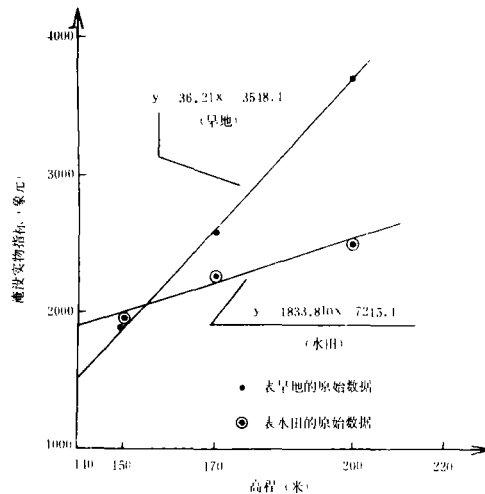


图5 高程—淹没损失相关线性图

Fig. 5 Linear Relation Between elevation and Submergence Loss

六、精度概况与分析

自动图像识别分类精度验证,是以同地区的1:3万的航空像片解译图和野外验证为基础,然后进行统计分析的。统计分析结果受到很多因素的影响,它还取决于所采用的抽样方式、抽样的样本数等。在统计分析中,首先应从整体中抽出一定量的样本,然后对该样本进行检验,从而可以看出样本质量的好坏。它是否能代表总体的特征,直接关系到统计分析的结果。所以抽样数量、抽样方式是比较重要的问题。抽样方式选择目的之一,是以最大限度地减小因抽样而带来的附加随机误差。

本文为保证分析结果具有代表性,每类的抽样数均在50个左右。抽样方式拟用分层随机取样方式。具体方法是先总体取样,计算落入每层的总量,若有一层满足取样规模,那么总体取样便告结束。以后的取样凡落入该层的都不计算,直至每类都达到所规定的取样规模。

取样结束后,经检验得出混淆矩阵表,在此基础上便能进行统计分析。

设 \bar{x} 为样本的抽样精度, M 为总体精度, n 为抽样数^[3],由于每一样本都服从(0—1)分布,故样本方差为

$$\sigma^2 = M(1 - M)$$

根据大数定理, M 的置信区间估计为

$$P\left\{-b < \left(\frac{\bar{x} - M}{\sigma/\sqrt{n}}\right) < b\right\} = 1 - \alpha \quad (8)$$

利用(8)式确定置信水平 α 之后, 便能计算出在 α 之下的总体精度的置信区间。

经统计分析后, 得土地利用二级类型的分类总精度, 混合法是 67.7%, 最大似然比法是 57.7%, 模糊模式识别是 56.9%。一级类型的分类总精度, 混合法为 79.2%, 最大似然比法为 72.3%, 模糊模式识别是 70%。

影响分类精度的原因有: 卫星像片与验证的航空像片时相不一致, 年份相差 9 年, 月份相差半年; 其次是由于卫星图像的分辨率太低, 只用光谱分类, 在山区、在地形破碎的地区其精度受到限制; 第三是未对阴影区的像元进行预处理。

七、结 论

1. 用航天遥感方法估算水库淹没损失, 基本上能达到水电部门规划阶段的要求。本文利用卫星影像与数字地形模型相结合来完成淹没损失估算的方法, 基本上是可行的, 具有一定的实际生产意义。

2. 文中建立的土地覆盖分类的混合法等模式, 有一定的长处和特点, 是有一定实用价值的。

3. 单一的航天遥感信息不能满足水库淹没损失清查的全部要求, 它还必须与其它相关信息, 尤其是包含有大量自然环境信息和社会环境信息的地理信息系统相匹配, 才能使水库淹没损失遥感估算的水平大为提高。

参 考 文 献

- [1] 楼博世、孙章、陈化成, 模糊数学, 科学出版社, 1983 年。
- [2] P. F. Crapper, Errors Incurred in Estimating an Aear of Uniform Land Cover Using Landsat, Photo. Engin. and Remote Sens., Oct., 1980.
- [3] R. Michael Hord, William Brooner, Land-use map Accuracy Criteria, Photo. Engin. and Remote Sens., May, 1976.

A Method Research on Estimating Reservoir Submergence Loss from Landsat MSS

Zhou Fuqun

(South-west China Traffic University)

Abstract

The article primarily experiments a new method for estimating reservoir submergence loss from landsat. First, satellite image of experimental region, transformed to Gaussian coordinate, is registered to digital terrain image of the same region. Secondly, boundaries of reservoir submergence are determined with elevation and slope of digital terrain model (DTM). It develops a statistical method for modifying the error caused by calculating an area with grids. Land use of this region is classified by fuzzy pattern recognition, hybrid pattern (fuzzy cluster and fuzzy pattern recognition together), cluster and maximum-likelihood pattern. The classification accuracy of hybrid pattern is about 79.2 percent.

The experiment shows that above method is able to satisfy the request of submergence loss survey for reservoir planning.

The article pays special attention to the importance of incorporation of geography information with satellite data. Specially, the region, generally speaking, is mountainous and study contents include many social and environmental problems which can't reflected by remote sensing image. Therefore, this incorporation may greatly improve the level of estimating reservoir submergence loss by space remote sensing.