

三“S”一体化技术和方法的探讨

刘震 李树楷

(中国科学院遥感应用研究所 北京 100101)

摘要 本文对 GPS、GIS、RS 三 S 一体化的必要性进行了阐述,进而建立参数方程,对一体化处理中的数据离散化、误差模型加以研究证明了其操作的可行性,最后在此基础上得出一体化在实际工作中的程序,以及可行的方法。

关键词 三“S”一体化信息技术,地理信息,误差模型,传递函数

1 引言

全球定位系统 GPS (Global Position System), 遥感 RS (Remote Sensing)、地理信息系统 GIS (Geographical Information System) 这 3 项技术随着计算机技术、无线电技术以及地球科学的发展已从各自独立发展进入了相互融合,相互渗透共同发展的阶段。这 3 种技术的共同发展和相互结合称之为三“S”一体化信息技术。GPS、RS、GIS 就单项技术而言已从研究阶段进入实用阶段,而三“S”一体化信息技术还处于萌芽状态,因此有关三“S”一体化信息技术、理论、方法的探讨是非常重要的。以下从物质运动的形式、地学信息的特点,以 3 个“S”的各自特点出发分别从思想、数学理论和技术方法不同的角度和层次阐述了三“S”一体化信息技术的必要性、可行性和具体结合的技术途径。

2 问题的提出

客观物质世界的各种物质不论其存在的具体表现形式如何,其运动形式可分为 5 种:机械运动、物理运动、化学运动、生命运动和社会运动,并且依照该顺序依次由低级向高级排列。各种运动形式相互渗透,高级形式包含了低级形式,并以此为基础,因此对机械运动的研究是研究其它各种形式的运动的基础,进而也是研究客观物质世界运动的起点。对客观物质世界的研究可以简化为朴素的机械运动的研究,也就是什么东西 (WHAT)、在什么地方 (WHERE)、什么时间 (WHEN)、发生了什么事情 (THING),因此三“W” (WHAT, WHERE, WHEN) 也就成为研究事情 (THING) 3 个重要特征,将这 3 个方面割裂研究问题的方法从辩证的角度看是片面的,很难反映其本质,因此从认识的角度出发,将 3 个方面作为一个有机的整体加以研究是十分必要的。

过去由于认识的不足和科学技术的局限性,只能从某一个方面研究地物的运动,这种研究方式或多或少带有很大的主观性。而地物反映的信息是以一种信息流的方法多元多维地表达出来的,对事物的认识应该从事物本身发展过程出发,按其本身发展规律进行纵向的研究,单纯分为几个层次或几个阶段的研究方法很难达到预期目的。

GPS 是利用 24 颗卫星在两万公里高度构成的卫星群发出的信号,采用三角测量的原理,确定出地球空间中任意位置的精确空间坐标,因此物体的空间位置(WHERE)的问题利用 GPS 就可以迎刃而解了。RS 是指不与物体接触而获得该物体信息的技术与科学,他主要从物体的光谱特性上认识物体,达到了解物体性质的目的,也就是解决是什么的问题(WHAT)。而 GIS 是采集、存储、管理、分析和描述整个或部分地球表面与空间和地理分布有关的空间信息的系统。通过对数据的分析可以解决事物随时间、空间而变化的问题,从而进行动态的管理,实质上解决了三“W”的问题。

在现阶段的大多数的工作中 GPS,RS,GIS 的很多工作是相互分离的,存在很多要进一步讨论的问题。下面从 RS 和 GPS,GIS 和 RS,GIS 和 GPS 3 个方面进行说明。

RS 和 GPS 在遥感处理中几何纠正主要依靠的是地面控制点(GCP),GCP 点的选取有两种方式,一种是从地形图上选取,另一种主要靠野外实地测量。前者有很大局限性:一是地表变化大,地形图上的明显地物在地面有时难以找到,更有甚者找不到。而且其工作量很大,一幅 TM 影像相当于 250 幅 1:10000 地形图^[1]。实地测量可以克服前者的问题,但也存在障碍。在一些地区自然环境恶劣条件艰苦,甚至于人不能正常工作。虽然 GPS 比传统的方法效率高,速度快,精度高,但地面设站的工作仍非常困难,因此全球定位系统的缺陷也显而易见。

GIS 和 RS 这二者的结合主要指的是遥感处理过程中有效的利用 GIS 数据以及遥感处理结果作为数据源更新,修正地理信息系统数据。例如利用遥感手段进行森林资源调查,在进行树种分类时,由生物学知识得知树种 A 只能生长在坡度 B 和 C 之间,因此只须利用 GIS 中的 DEM 便可以划分出适宜 A 生长的地区,在此基础上再结合其它方法进行准确分类。DEM 可由地形图数字化、实地测量和航空摄影测量得到。但无论那一种方法都存在周期长,工作量大的缺陷。如能利用地理信息系统中的基础数据 DEM 则能大大减少工作量,缩短时间。反之遥感处理的结果数据可以进入地理信息系统,作为其数据源。同例中遥感精确处理后得到的 A 的生长区域与地理信息系统中 DEM 所限定的区域不同,可以考虑地球内外因素的影响导致地形变化,使之与 GIS 中的数据不一致,因此可以利用 RS 结果数据修改地理信息系统中的 DEM。但空间数据精度不一致,数据采集的时间不一致是两者结合的障碍。

GPS 和 GIS GPS 一般作为 GIS 的一个空间数据的数据源。实际工作中首先在地形图中或 RS 影像中找到明显地物,再到实际测区进行测量。由于地形图与实际地表的 inconsistency 使得很多时间用于寻找目标。一种有效的方法是在地理信息系统中得到地物的空间坐标,用全球定位系统导航到地物很近的地方,再施测。其结果在地形图上难以准确标识,给地形图修测带来困难,但却很容易用于修改地理信息系统数据。

从以上 GPS 和 GIS,RS 和 GPS,GIS 和 RS 的两两结合可以看出:两两结合比单独操作有更好的效果。但是仍有一些缺陷。主要是缺乏统一的坐标空间,光谱数据和空间数据时间上的不一致,以及具备封装独立的数据和方法能力的技术。因此把三者结合起来形成一体化的信息技术体系是非常迫切的。这主要指数据获取平台的革新和新的信息融和方法的应用。

信息是用文字、符号等来表示事件、实物、现象的内容、数量或特征。信息的来源是数

据。数据是未加工过的原始资料,它是对客观世界的表示,而信息是数据内涵的意义,是

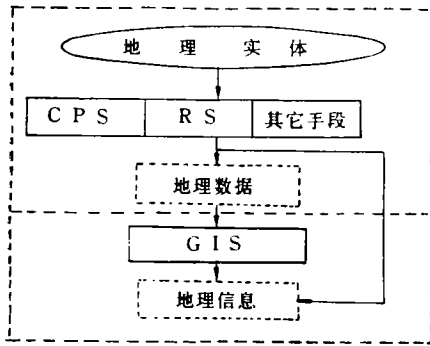


图 1 3“S”一体化模型

Fig. 1 Model of 3“S” integration

数据的内容和解释。地理信息是指与所研究对象的空间地理分布相关的信息,它表示地物及环境固有的数量、质量、分布、联系以及规律^[2]。从地理实体到地理数据,再到地理信息是人类认识地物的必由之路。GPS、RS 主要面对的是地理数据的问题,而 GIS 的对象是地理信息。由于 GPS、RS、GIS 在人类认识地物的过程中所起作用,处理的问题,处理问题的手段不同,因此它们的结合不是一种等结构的结合,而是有层次地有机的结合如图 1, 这种结合的方式是由其各自特点决定的。地理数据和地理信息既相互区别又相互联系,二者既不可截然分开

又不能混为一谈。这种结合方式使人们能够有效地进行控制地学信息的处理,以获得最佳处理效果。

地理数据,地理信息的概念已通用化, GPS,RS,GIS 各自成为成熟技术,这类技术的有机结合在理论上没有障碍,技术上经过努力可以实现,这使三 S 一体化信息技术成为可能^[3]。

因此无论从物质的运动形式,地学信息本身特征,还是从三“S”各自的技术特性出发,GPS、RS、GIS 的结合是科技发展的必然结果。

3 系统模型

以上阐述了三“S”一体化信息技术的可能性和必要性,下面从理论角度给予初步说明。

3.1 简化模型建立参数方程

首先地学信息的时间特征、光谱特征和几何特征等构成了一个线性正交的地理信息空间 β , 这一点是很容易从实践中得到证明的。该空间由时间向量 G 空间向量 W 和光谱向量 E 等组成,即

$$\beta = (W, E, G) \quad (1)$$

不管光谱的变化如何,沿波长变化方向总可以将光谱分解为在一个连续区间 $[u, v]$ 的无限个谐波的叠加:

$$r = \frac{a_0}{2} + \sum_{i=1}^{\infty} [a_i \cos(\omega i x) + b_i \sin(\omega i x)] \quad x \in [u, v] \quad (2)$$

r 是一个时间序列,其变化符合(2)式。

在 β 空间内的连续的地学信息,人们只能用离散的数据来近似表达这种连续的变化。为了用一种方法去适合各种不同的数据,则必须建立一种统一的关系,利用参数方程可以

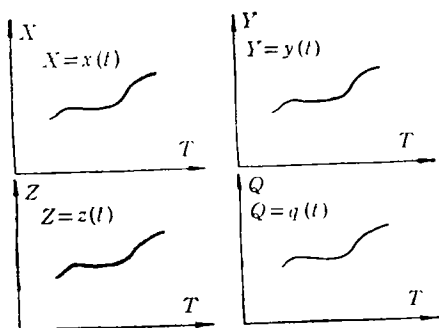


图 2 参数方程组

Fig. 2 Parameter equation group

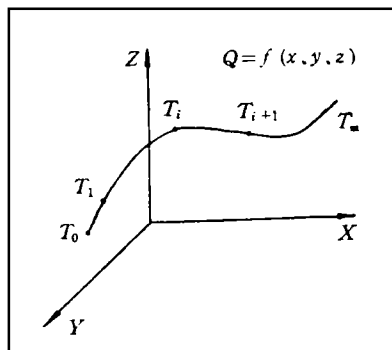


图 3 光谱质量的概念

Fig. 3 Principle of spectral mass

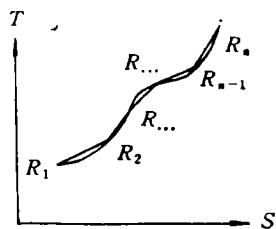


图 4 参数方程

Fig. 4 Parameter equation

达到这个目的^[3],可以借助于空间的光谱质量的一段曲线加以说明,如图 2,即可以用公式:

$$X = x(t) \quad Y = y(t) \quad Z = z(t) \quad Q = q(t) \quad (3)$$

来描述该曲线。 P 是指 (T_i, T_{i+1}) 之间的曲线的光谱质量,对空间任一点 (x, y, z) 当 $G = G_0$ 时,其光谱密度特征 Q 唯一(图 3),即满足:

$$Q = f(x, y, z) \quad (4)$$

因此光谱质量定义为

$$P_i = \iiint_{T_i}^{T_{i+1}} f(x, y, z) dx dy dz \quad (5)$$

以下我们不必要去描述公式(3)的各种关系而可以在一个 S 为横坐标, T 为纵坐标的 TS 平面内进行讨论,这里 S 代替了 X, Y, Z, Q (图 4)。即

$$T = T(S) \quad (6)$$

3.2 简化模型分析

首先多项式是用来内插的最直接的方法

$$T = T(S) = \sum_{i=0}^{n-1} a_i S^i \quad (7)$$

多项式参数 a_i 可由以下线性方程求出:

$$T_i = \sum_{i=0}^{n-1} a_i S_i^i \quad (8)$$

由于多项式展开系数求解比较困难,为了解决多项式求解问题,拉格朗日提出一种巧妙方法对每个数据点 T_i 引起一个基函数 g_i :

$$g_i(S) = \frac{(S - S_1)(S - S_2) \cdots (S - S_{i-1})(S - S_{i+1}) \cdots (S - S_n)}{(S_i - S_1)(S_i - S_2) \cdots (S_i - S_{i-1})(S_i - S_{i+1}) \cdots (S_i - S_n)} \quad (9)$$

多项式 $T = T(S)$ 最终形式由已知值 T_i 组合得到

$$T = T(S) = \sum_{i=1}^n g_i(S) T_i \quad (10)$$

按照上述公式,利用有限数据点计算出参数方程的参数,再利用组合多项式, Akima 法等各种方法进行内插逼近^[4]。下面对数据点推导出曲线的中误差,以及用怎样的点距采样才能使获得的数据达到精度,即达到事先规定的精度指标这些问题进行讨论。

在一个有限区间 $[0, l]$ 上连续的 $T(S)$ 可以用各种频率的正、余弦波组成富立叶级数表示,为了去掉(2)中的 $a_0/2$, 则 S 轴平移至 T 轴的中值上,即 $T = \frac{a_0}{2}$ 处,则公式(2)变为:

$$T(s) = \sum_{i=1}^{\infty} [a_i \cos(\omega_i s) + b_i \sin(\omega_i s)] \quad s \in [0, l] \quad (11)$$

即:

$$T(s) = \sum_{k=1}^{\infty} C_k \cos\left(\frac{2k\pi}{L}s - \varphi_k\right) \quad k = 1, 2, \dots, \infty \quad (12)$$

式中 φ_k 为相位, C_k 为振幅,二者有如下关系:

$$C_k = \sqrt{a_k^2 + b_k^2} \quad (13)$$

$$\varphi_k = \arctg(b_k/a_k) \quad (14)$$

C_k 的总体叫振幅谱, C_k^2 叫功率谱。

相对于各个波长的倒数叫频率:

$$f_k = \frac{1}{L_k} = \frac{k}{L} \quad (15)$$

式中 f_k 为频率, L_k 为波长。

为了确立 $2m$ 个 a_k, b_k , 并由此推导出 C_k , 所以离散点要有 $n \geq 2m$ 个, 由采样定理得知 Δs 与 f_{\max} 即 L_{\min} 之间的关系:

$$\Delta s < \frac{L_{\min}}{2}, \quad \Delta s < \frac{1}{2f_{\max}}$$

若是 $f > f_{\max}$ 不存在, 则(11)式限定在一个有限范围之内:

$$T(s) = \sum_{k=1}^m \left(a_k \cos \frac{2k\pi}{L} s + b_k \sin \frac{2k\pi}{L} s \right) \quad (16)$$

即

$$T(s) = \sum_{k=1}^m c_k \cos \left(\frac{2\pi k}{L} s - \varphi_k \right) \quad (17)$$

对于 $(s_j, t_j), s_j = j\Delta s$ 得到方程

$$T_j = \sum_{k=1}^m \left(a_k \cos \frac{2\pi k j \Delta s}{L} + b_k \sin \frac{2\pi k j \Delta s}{L} \right) \quad (18)$$

$$k = 1, 2, \dots, m, \quad j = 1, 2, \dots, n$$

当 $n = 2m$ 时,有

$$a_k = \frac{2}{n} \sum_{j=1}^n T_j \cos \frac{2\pi k j \Delta s}{L}$$

$$b_k = \frac{2}{n} \sum_{j=1}^n T_j \sin \frac{2\pi k j \Delta s}{L} \quad (19)$$

功率谱

$$C_k^2 = a_k^2 + b_k^2 = 4a_{-k}a_k \quad (20)$$

对于系统的传递函数有

$$H(f_k, S) = \frac{\bar{c}_k}{c_k} \quad (21)$$

式中 c_k 为原始信号, \bar{c}_k 为衰减后信号。

c_k 由非常密的采样点获得并由(13)(14)(20)计算得出, \bar{c}_k 由稀疏点和内插得到。大量文献表明 H 取决于间隔 Δs 的大小及内插方法^[4,5]。

令 $u_k = \Delta s/L_k = f_k \Delta s$, 则(21)变为:

$$H(u_k) = \frac{\bar{c}_k}{c_k} \quad u_k = f_k \Delta s \quad \text{同时} \quad u_k < \frac{1}{2}, \quad k = 1, 2, \dots, m \quad (22)$$

内插获得的数字模型的精度

$$\sigma_T^2 = \frac{1}{L_k} \int_{s=0}^{L_k} [T(s) - \bar{T}(s)]^2 ds = \frac{(c_k - \bar{c}_k)^2}{L_k} \int_{s=0}^{L_k} \cos^2 \frac{2\pi}{L_k} s ds$$

$$= \frac{(c_k - \bar{c}_k)^2}{2} \quad (23)$$

由(22)得

$$\sigma_T^2 = \frac{1}{2} [1 - H(u_k)]^2 c_k^2 \quad (24)$$

进而得到整个波形方差

$$\sigma_T^2 = \frac{1}{2} \sum_{k=1}^m [1 - H(u_k)]^2 c_k^2 \quad (25)$$

由于在数据采样时到 T 存在量测误差 σ_m

则:

$$\sigma_{\text{插}}^2 = \sigma_{\text{插}}^2 + k\sigma_m^2 \quad (26)$$

k 为与内插方法有关的常数。

由以上结果可以看出,整体误差由采样间隔,内插方法,和采样的量测误差 3 个部分构成。

3.3 实际问题分析

现在我们把 GIS、GPS、RS 的处理结果在一起进行考虑,分别从空间数据,光谱数据和观测时间数据 3 个方面进行分析。

对于空间数据有 $W = w(X, Y, Z)$

对于光谱数据有 $E = e(E_1, E_2, \dots, E_n)$ (27)

对于时间数据有 $G = g(G_1, G_2, \dots, G_m)$

E_i 为第 i 波段数据 G_j 为 j 时刻时间向量, $1 \leq i \leq n, 1 \leq j \leq m, n, m$ 为自然数。

考虑时间、空间、光谱和进行分析的结果用一非线性函数表示, U, V, E 分别表示分析结果的时间、空间、光谱数据。

$$\begin{aligned} U &= u(W, E, G) \\ V &= v(W, E, G) \\ R &= r(W, E, G) \end{aligned} \quad (28)$$

则

$$\begin{aligned} \delta_U &= \varepsilon_W \frac{\partial U}{\partial W} + \varepsilon_E \frac{\partial U}{\partial E} + \varepsilon_G \frac{\partial U}{\partial G} \\ \delta_V &= \varepsilon_W \frac{\partial V}{\partial W} + \varepsilon_E \frac{\partial V}{\partial E} + \varepsilon_G \frac{\partial V}{\partial G} \\ \delta_R &= \varepsilon_W \frac{\partial R}{\partial W} + \varepsilon_E \frac{\partial R}{\partial E} + \varepsilon_G \frac{\partial R}{\partial G} \end{aligned} \quad (29)$$

$$Q_{\delta\delta} = \begin{bmatrix} Q_{UU} & Q_{UV} & Q_{UR} \\ Q_{VU} & Q_{VV} & Q_{VR} \\ Q_{RU} & Q_{RV} & Q_{RR} \end{bmatrix} \quad (30)$$

令

$$\begin{aligned} A_U &= \left[\frac{\partial U}{\partial W} \quad \frac{\partial U}{\partial E} \quad \frac{\partial U}{\partial G} \right] & A_V &= \left[\frac{\partial V}{\partial W} \quad \frac{\partial V}{\partial E} \quad \frac{\partial V}{\partial G} \right] \\ A_R &= \left[\frac{\partial R}{\partial W} \quad \frac{\partial R}{\partial E} \quad \frac{\partial R}{\partial G} \right] & B &= \begin{bmatrix} Q_{WW} & Q_{WE} & Q_{WG} \\ Q_{EW} & Q_{EE} & Q_{EG} \\ Q_{GW} & Q_{GE} & Q_{GG} \end{bmatrix} \end{aligned}$$

则:

$$\begin{aligned} Q_{UU} &= A_U B A_U^T & Q_{VV} &= A_V B A_V^T \\ Q_{VR} &= Q_{RV} = A_V B A_R^T & Q_{RR} &= A_R B A_R^T \\ Q_{UV} &= Q_{VU} = A_U B A_V^T & Q_{UR} &= Q_{RU} = A_U B A_R^T \end{aligned} \quad (31)$$

Q_{ij} 与 σ_{ij} 的关系表现为

$$Q_{ij} = \frac{\sigma_{ij}}{m_0^2} \quad m_0^2 \text{ 为权为 1 时的方差} \quad (32)$$

则(30)为:

$$\begin{aligned} \sigma_{UU} &= m_0^2 A_U B A_U^T & \sigma_{VV} &= m_0^2 A_V B A_V^T \\ \sigma_{RR} &= m_0^2 A_R B A_R^T & \sigma_{UV} &= \sigma_{VU} = m_0^2 A_U B A_V^T \\ \sigma_{UR} &= \sigma_{RU} = m_0^2 A_U B A_R^T & \sigma_{VR} &= \sigma_{RV} = m_0^2 A_V B A_R^T \end{aligned} \quad (33)$$

4 应用举例

上面分别定性定量阐述了三“S”一体化信息技术的必要性,可能性,以及误差和采样要求,在理论上为三“S”一体化信息技术奠定了实践的基础,下面从实际工作的角度加以说明。

在应用上述介绍的理论时,作者推荐以下程序,首先利用很小的采样间隔对典型样区进行数字化,利用(13)(14)(20)计算出振幅谱,并且确定振幅谱略去不计的最高频率,再导出整个空间的采样间隔。大量研究表明,经内插获得的结果很大程度上与内插方法无关,而与采样间隔关系很大^[3]。

以作物估产为例,在条件可能的情况下,首先分析被估产作物所允许的最终误差,按(28)导出 U, V, R 的逆函数,依据(13)(14)(20)(29)(31)(33)计算出时间、光谱、空间所允许的误差。求出 $Q_{T_{\text{总体}}}^2$ 和 Q_m^2 , 用很小的采样间隔对作物波谱曲线,地面形状,作物随时间变化的生长情况进行采样,分别利用(13)(14)(20)计算出振幅谱和略去不计的最高频率,再导出整个空间的采样间隔。按照生物量随时间变化曲线导出的时间间隔是对目标进行时间采样的标准,也就是说对目标重复观测的时间间隔不大于 ΔG 。而每次观测必须在光谱空间内使其采样步长小于或等于 ΔE (由光谱曲线导出),空间上小于 ΔW ,这样才能保证最终求得作物的生物量在要求的误差之内。即:

- (1) 设计系统所要达到的目标 Q_{ss} 。
- (2) 分别确定时间、光谱、空间 $\sigma_{T_{\text{总体}}}^2, \sigma_T^2, k, \sigma_m^2$ 。
- (3) 根据作物生长生物量与时间的关系曲线计算出合理的时间采样间隔 ΔG 。
- (4) 依据作物光谱曲线推导出理想的探测器的波段数和通道的宽度 ΔE 。
- (5) 计算出合乎要求的空间采样尺度 ΔW 。
- (6) 计算时间、光谱、空间的整体误差能否达到(1)所要求的 Q_{ss} , 达成标准则继续,否则重复(1)到(6)。
- (7) 按照(4),(5)所确定的采样间隔 $\Delta E, \Delta G$ 对目标进行观测。
- (8) 以(3)推导出的时间间隔 ΔG 重复(6)直至完成任务。
- (9) 将数据输入地理信息系统,建立分析模型进行估产。

5 结束语

由此不难得出这样一个结论：设计合理的采样间隔，具有同步的获取空间和波谱数据的高重复观测能力的平台，支持具有数据封装能力的地理信息系统是三“S”一体化信息技术的核心。一体化的三“S”信息技术其功能远远大于各分技术。

参 考 文 献

- [1] 李树楷. 遥感图像对地定位研究. 北京: 测绘出版社, 1991.
- [2] 李德仁, 龚建雅, 边馥苓. 地理信息系统导论. 北京: 测绘出版社, 1993.
- [3] 卡尔克劳斯. 摄影测量处理系统的理论和实践. 北京: 测绘出版社, 1989.
- [4] Morite. H. & Sunkel. h. Method in Geodetic. Wichman, 10.
- [5] Akima. H. J. Hournal of the Association for Computer Maching. 1970.

The Practice of Integration of GPS,RS and GIS

Liu Zhen Li Shukai

(*Institute of Remote Sensing, Chinese Academy of Sciences, Beijing 100101*)

Abstract From the philosophical viewpoint necessity of integration of GPS,RS and GIS is introduced briefly, on the base of which authors prompt theoretical suport to data adaptation and it's error model. At last a practical method and an example of 3“S” integration are given.

Key word 3“S” Integration, Geographical information, Eerror model, Mode transfer function