DOI:10.13203/j.whugis1993.01.005

第18卷第1期　　武汉测绘科技大学学报　　Vol. 18 No. 1
1993年3月　　Journal of Wuhan Technical University of Surveying and Mapping　　Mar. 1993

GPS水准计算的非参数回归法

薄志鹏　刘国辉

摘　要　非参数回归是一种广义的回归方法，它具有直观，对模型要求不严密，计算简单的优点。本文介绍非参数回归的原理及其在GPS水准计算中的应用，通过对两个测区的试算，得出了一些有益的结论．

关键词　GPS水准；非参数回归；权函数估计法；近邻估计

1　引　言

众所周知，由GPS相对定位测量和几何水准测量可以求定点间的大地高差和正常高差，当给定参考基准后，点的大地高 H，正常高 H_r 和高程异常 ζ 间有如下关系式

$$H = H_r + \zeta \tag{1}$$

当测区中有一部分点已用GPS定位技术和常规高程测量方法求得其大地高和正常高，则按式(1)可计算得点的高程异常，我们称这些点为支撑点或已测点。若测区中支撑点的数量足够多，且分布较为均匀，则可以拟合测区的似大地水准面形状，进而推算测区中其余未进行水准联测的GPS点的高程异常和求定点(一般称为未测点或计算点)的正常高，这种方法可称为“几何法”GPS水准。

目前可用于GPS水准计算的各种方法主要有：

1)解析内插(直线内插，曲线内插，样条函数，……)

2)参数回归将已知数据拟合为某种数学面(平面[2]，高次曲面[1,3,4]，多面函数……)

3)滤波推估

上述这些方法大部分已在GPS水准计算中得到应用，当测区地形变化较大时，计算中还应顾及估计量与高程间的相关性[1,7]。

文[4]对解析内插，非线性回归和样条函数在GPS水准计算中的应用作了分析比较，获得了较好的结果。

在回归分析中，设 X 和 Y 分别是 d 维和一维随机变量。假定 $E|Y|<\infty$，则 $\mu(x)=E(Y|X=x)$ 存在，$\mu(x)$ 称为 Y 对 X 的回归函数。回归分析的基本问题在于通过从 (X,Y) 抽出的独立同分布样本(*iid.* 样本)，即 $(X_i,Y_i)(i=1,2,\cdots,n)$ 去估计回归函数 $\mu(x)$。传统的回归分析假定回归函数有某种特定的数学形式，并假定“误差”$Y-\mu(x)$ 的分布为正态。这时用最小二乘法对回归系数进行估计就可以得到 $\mu(x)$ 的估计。在上述前提下，这种估计有许多优良性质，但经验和理论

收稿日期：1992-07-04

都显示，这种方法并非总能提供良好的结果，这主要是因为在不少情况下，关于参数模型的基本假定与实际情况有较大的差距。例如，采用多项式作为核函数时，为了取得较好的逼近效果，常取较多的项，但此时对未测点的预测效果很差。基于这一事实，促使人们去寻找别的出路，非参数回归就是一个努力方向。

滤波推估的主要困难在于协方差函数的确定。由于半值宽度随地区和网的结构而异，没有足够的资料协方差函数就难以求定。然而，正是因为测区数据不足才需要推估，所以这是一个较难解决的问题。此外，该法的计算工作量也较大。

本文主要介绍用非参数回归法计算 GPS 水准。

2 非参数回归及其在 GPS 水准计算中的应用

非参数回归的特点是对模型的要求很松：回归函数的形式可以任意，随机误差也不必服从正态分布。广义地说，自变量 X 与因变量 Y 之间的回归关系可以理解为 X 与 Y 间虽无确定性关系，但 Y 在给定 $X=x$ 时的条件分布取决于 x，因而更一般地可提出对此条件分布进行估计的问题。这种问题都可以用估计回归函数的形式表达之。例如要估计条件概率 $P(a\leqslant Y\leqslant b\mid X=x)$，则可定义一个新的因变量 $Z=I_{[a,b]}(Y)$，这时所述条件概率转化为 Z 对 X 的回归函数 $E(Z\mid X=x)$。

非参数回归权函数估计法的定义[5]为：

设 $\mathrm{W}_{ni}=\mathrm{W}_{ni}(x)=\mathrm{W}_{ni}(x;X_1,X_2,\cdots,X_n)(i=1,2,\cdots,n)$是选定的依赖于 x 和 $X_1,X_2,\cdots,X_n$ 的函数，则称

$$\mu_n(x)=\sum_{i=1}^{n}\mathrm{W}_{ni}Y_i \tag{2}$$

为回归函数 $\mu(x)$的权函数估计，$\{\mathrm{W}_{ni}\}$称为权函数。

在实用问题中，有时尚要求权函数满足自然的条件

$$\mathrm{W}_{ni}(x;X_1,X_2,\cdots X_n)\geqslant 0,\ \sum_{i=1}^{n}\mathrm{W}_{ni}(x;X_1,X_2,\cdots X_n)=1 \tag{3}$$

满足上述条件的权函数称为概率权函数。

从应用的观点看，最重要的问题是如何选择适当的权函数以用于特定的问题中。目前在文献中考虑较多的定义权函数的方法有两种：

1. 核函数法　选定 R^d(X 是 d 维的)上的非负函数 K，窗宽 $h_n>0$，令

$$\mathrm{W}_{ni}(x;X_1,X_2,\cdots X_n)=K\left(\frac{x-X_i}{h_n}\right)\Big/\sum_{j=1}^{n}K\left(\frac{x-X_j}{h_n}\right) \tag{4}$$

K 称为核函数．当简单地取

$$K(x)=\begin{cases}1, & \|x-X_i\|\leqslant h_n\\ 0, & \|x-X_i\|>h_n\end{cases} \tag{5}$$

称为“均匀核”。这个核函数表示只有离 x 很近的几个点在估计中起作用。重力异常推估中的平均值法[8]就可归属于这一选择。

2. 近邻法　该法先引进一个衡量 R^d 中两点 $U=(u_1,u_2,\cdots u_d)$和 $V=(v_1,v_2,\cdots v_d)$的距离的函数 $\|U-V\|$。若要考虑到 X 的各个分量的重要性，则可以再引进反映这一点的权因子 C_i，例如采用欧氏距离时，

$$\| U - V \|^2 = c_1(u_1 - v_1)^2 + c_2(u_2 - v_2)^2 + \cdots c_d(u_d - v_d)^2 \quad (c_i > 0) \tag{6}$$

给定一个自然数 $k \leqslant n$，以及 k 个正数 $C_{n1}, c_{n2}, \cdots c_{nk}$，满足条件

$$c_{n1} \geqslant c_{n2} \geqslant \cdots \geqslant c_{nk} > 0, \sum_{i=1}^{k} c_{ni} = 1 \tag{7}$$

现设有样本 $(X_i, Y_i)(i=1,2,\cdots n)$，并指定 R^d 中的一个点 x，将 $X_1, X_2, \cdots X_n$ 按其在距离 $\| \cdot \|$ 的意义下与 x 接近的程度排序

$$\| X_1 - x \| < \| X_2 - x \| < \cdots < \| X_n - x \|$$

然后给与 X_i 对应的 Y_i 赋以权 W_{ni}

$$\begin{cases} W_{ni}(x; X_1, X_2, \cdots X_n) = c_{ni} (i = 1, 2, \cdots k) \\ W_{ni} = 0, i > k \end{cases} \tag{8}$$

于是

$$\mu(x) = \sum_{i=1}^{n} W_{ni}(x) Y_{ni} = \sum_{i=1}^{k} c_{ni} Y_{ni} \tag{9}$$

若令 $W_{ni}(xX_1, X_2, \cdots, X_n) = 1/k, (i=1, \cdots, k)$，则又得出前述的平均值法(按均匀核)。

这两种定义权函数的方法出自概率密度估计的核估计(Kernel estimates)和最近邻估计(Nearest Neighbor Estimate)理论，虽然其出发点有所不同，但结果是一致的。

非参数方法具有直观性强，计算简便的优点。"从理论上说可以有把握地断言，它不劣于传统的方法"[5]。近邻权是一种具有优良大样本性质的权。根据"权函数估计的矩相合性"理论，文献[6]从预测风险来估计近邻权的推估效果，证明了：在一定条件下(即使只从 n 个样本中选用离 x 最近的一个 X_1 所对应的 Y_1 预测 Y，称为最邻近预测，其风险也只是最小风险的两倍。

在 GPS 水准计算中，已知支撑点 $P_i(i=1,2,\cdots n_j)$ 的坐标及其高程异常 (x_i, y_i, ζ_i) 和计算点 P_j 的坐标 (x_j, y_j)。由式(8)和(9)，根据重力异常内插和推估的理论，我们同样地以支撑点到计算点的距离 d_{ij} 的倒数为权，则不难写出计算点 P_j 的高程异常估计式为

$$\begin{cases} \zeta_j = \left(\sum_{i=1}^{n_j} D_{ij} \right)^{-1} \sum_{i=1}^{n_j} \left(\zeta_i \prod_{\substack{k=1 \\ k \neq i}}^{n_j} D_{kj} \right) \\ D_{kj} \leqslant D_0 \end{cases} \tag{10}$$

其中 D_0 是一个按经验给定的距离限值，相当于窗宽。在计算中也可以限制计算未测点所需要的最邻近支撑点数。

3 数字研究

图 1 和图 2 表示某两测区(设为 A, B 测区)的似大地水准面形状，图中等高线注记以毫米为单位，大数已删去。两测区的基本情况见表 1。由图可知，A 测区似大地水准面起伏平缓，而 B 测区则变化较大。数字研究的目的在于：

1. 探讨 GPS 水准的精度及其适用范围；
2. 评定非参数回归的效果；
3. 确定支撑点的合理密度。

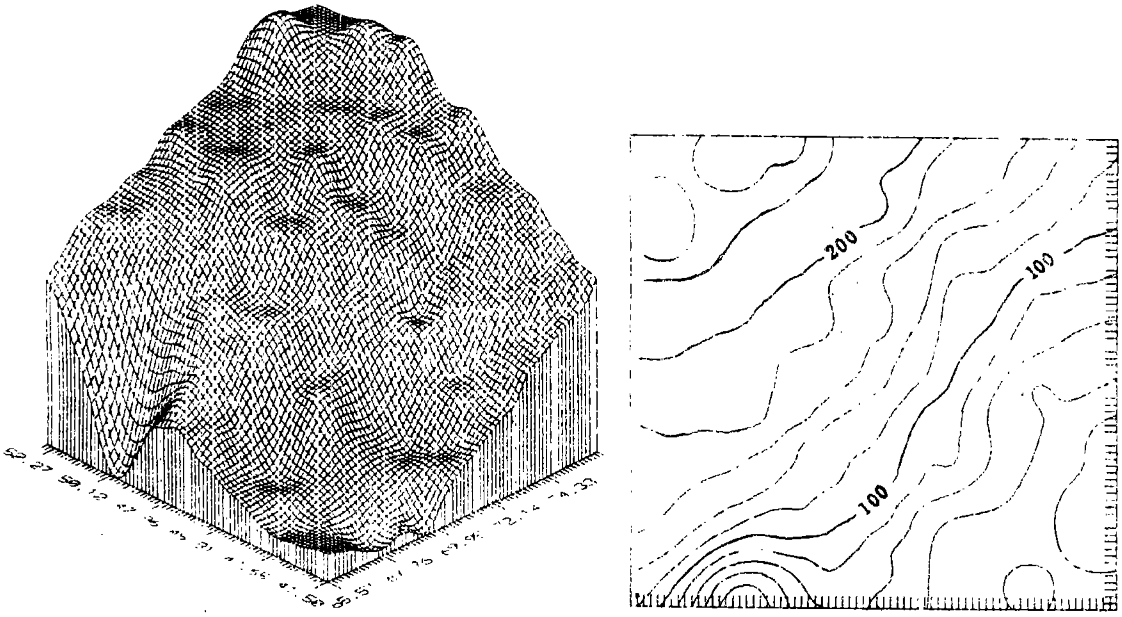

图 1　*A* 测区似大地水准面形状

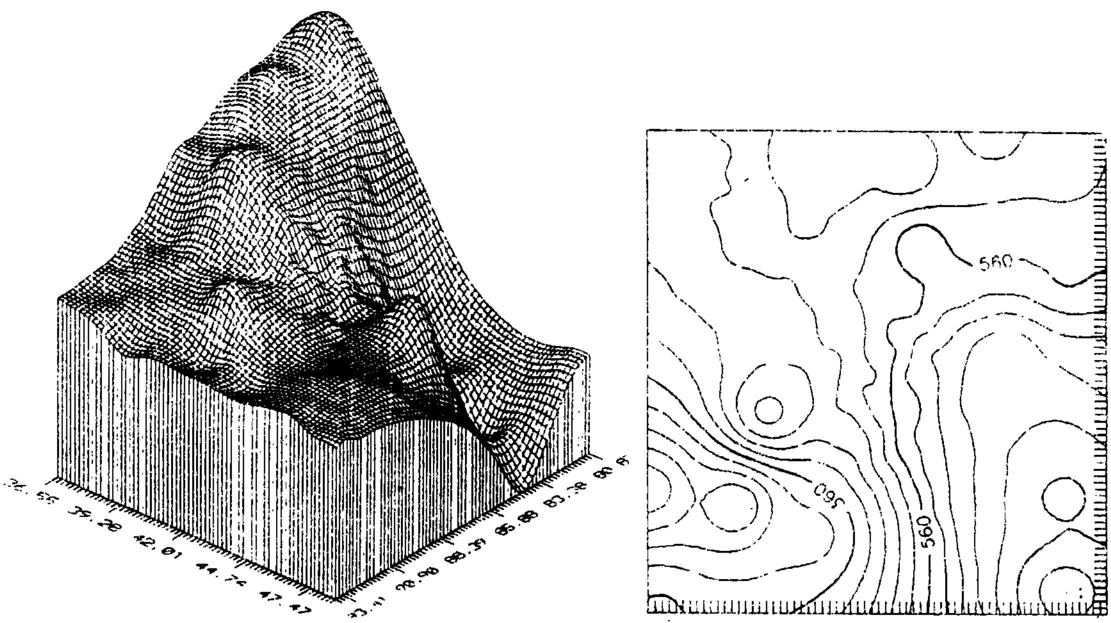

图 2　*B* 测区似大地水准面形状

表　1

测区	已测点数	测区面积	$\zeta_{max}-\zeta_{min}$	高程异常最大变化率
A	29	120km^2	0.25m	5cm/km
B	19	140km^2	0.65m	15cm/km

由已知数据按不同方案将部分已测点作为支撑点，其余点作为“未测点”按式(10)计算未测点的高程异常。然后，由已知的高程异常与其估计值之差计算 ζ 的估计中误差，计算结果已

列于表 2。

表 2

测区名称	A			B				
计算方案	1	2	3	1	2	3	4	5
支撑点个数	8	9	11	5	5	6	8	9
支撑点平均密度(km^2/点)	15	13	11	28	28	23	13	15
支撑点分布状况	均属外围点	外围 8 点内部 1 点	外围 8 点内部 3 点	4 个外围点 1 个边区点	均属外围点	均属外围点	5 个外围点 2 个内部点	均匀布点
计算点数	21	20	18	14	14	13	11	10
估计中误差	±24mm	±13	±12	±64	±56	±31	±26	±26
$\max\{\zeta_{已知}-\zeta_{估计}$	42mm	34	34	124	100	60	42	46

计算结果表明：

1. 未测点高程异常的估计精度不仅与似大水准面起伏变化的幅度，支撑点的平均密度有关，而且还与支撑点的分布有关。当多数支撑点与似大地水准面的“特征点”一致时，可以获得良好的计算结果。

2. 就 A、B 两测区言，当支撑点的平均密度不大于 20km²/点时，未测点高程异常的估算精度约为±3cm；在不利情况下，高程异常的估算精度可达±5cm。

3. 支撑点的分布，测区边缘(外围)应适当密一些(点间距约 5～6km 为好)，以便了解测区似大地水准面总的变化趋势。内部点的密度应视似大地水准面的变化幅度而定，最好使支撑点至计算点的距离 D_k,保持在 3～4km 以内。

4. 试算表明。对外围点而言，当计算点位于两支撑点联线附近时(距不超过 1km)，采用直线内插同样可以获得良好的结果。

3　结论与建议

综上所述可有如下结论：

1. 非参数回归是一种广义的回归方法，它具有直观，对模型要求宽松，计算简便，估计精度不低于传统方法的优点。此外它不像一般 的数值逼近方法(样条函数除外)，在支撑点上会出现裂隙。

2. 当支撑点的密度及其分布能满足前述要求时，GPS 水准可以满足城市一般工程和测图的需要。

3. 为了更好地做好支撑点的高程联测工作，建议联测工作分两步走：第一步按一定密度尽量联测网的外围点，然后计算支撑点的高程异常。根据 ζ 的变化幅度和趋势进一步确定测区内部的联测计划，这样可以使支撑点的分布能较好地接近测区似大地水准面的“特征点”。

4. 非参数回归是近十余年来发展起来的，这一貌似“近似”的方法具有很好的性质[5]。关于权函数的选择，专业知识和经验起很大作用。在正文的计算中，由于测区高差不大，仅简单地按距离的倒数定权，并取得较好的效果，当测区高差较大时，除可按式(6)给高程分量以适当的权因子(需要通过试验确定)外，也可以按其他方法确定权函数，这是今后需要进一步研究的问题。

致谢 在本文撰写过程中，得到了徐绍铨，于正林和晁定波同志的不少帮助，特此致谢。

参 考 文 献

1 丁行斌，许其凤. 关于 GPS 水准及应用. 大地测量综合性学术年会军测论文集，1991.
2 King R W. GPS 测地. 解放军出版社，1988.
3 李征航. GPS 测定高程试验. 测绘通报. 1991(2)
4 徐绍铨，李征航. 拟合法求定 GPS 点的正常高. 武测科技. 1992(2)：24～26
5 陈希孺，方兆本等. 非参数统计. 上海科学技术出版社. 1989，310
6 徐因. 用近邻权法推估重力异常. 武汉测绘科技大学学报，1984(2)
7 兰虎彪. GPS 网正常高求解方法的研究.（硕士论文）. 武汉测绘科技大学，1992.
8 Dan H. Obtanining Centimeter-precision Heights by GPS Observations over Small Areas, GPS World. 1990(1)

Nonparametric Regression Method of GPS-leveling Computation

Bo Zhipeng　　*Liu Guohui*

Abstract Nonparametric regression is a generalized regression which has some advantages of intuition, relaxed restriction on the adopted model and easy computation. The principle of nonparametric regression and it is application to GPS-leveling prediction are introduced. Finally, the method is demonstrated by test computation of two real GPS-leveling networks and it comes to some valuable conclusions.

Key words GPS-leveling; nonparametric regression; weight function estimate; nearest neighbor estimate