

基于不同大小窗口的移动曲面拟合法探测不规则 DEM 粗差的一种方法

杨晓云¹, 顾利亚¹, 岑敏仪^{1,2}, 李志林²

(1. 西南交通大学土木工程学院地理信息工程中心, 四川成都 610031; 2. 香港理工大学土地测量与地理资讯学系, 香港九龙)

A Method Based Fitting of Moving Quadric Surface with Variable Windows for Detecting Gross Errors in Irregular DEM

YANG Xiao-yun¹, GU Li-ya¹, CEN Min-yi^{1,2}, LI Zhi-lin²

(1. School of Civil Engineering, Southwest Jiaotong University, Chengdu 610031, China; 2. Dept. of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China)

Abstract: The presence of gross errors will distort the spatial variation presentation in digital elevation model (DEM). In some cases, totally undesirable and unacceptable results may be produced in DEM and its derived products due to the existence of such gross errors. Therefore, the detection of gross error in DEM has been becoming the focus that people greatly concern for. Depend on analyzing the existing algorithms of detecting gross errors in DEM based on the point wise method, a new algorithm (YXY algorithm) based on fitting of moving quadric surface with variable windows is established in this paper. Some simulated examples are given by Monte Carlo Method to verify that the method proposed is more reliable and correct than existing method.

Key words: digital elevation model; gross error; fitting of moving quadric surface

摘 要: 粗差的存在会造成数字高程模型 (DEM) 空间上的严重扭曲, 有时能导致 DEM 及其产品严重失真, 甚至完全不能使用。因而有关 DEM 的粗差诊断问题已愈来愈引起人们的关注。本文在对现有的基于点方式的 DEM 粗差探测算法进行分析的基础上, 提出一种基于不同大小窗口的移动曲面拟合法探测 DEM 粗差的一种方法 (简称 XYX 算法), 并通过蒙特卡洛仿真试验来验证该算法的有效性和可行性。

关键词: 数字高程模型; 粗差; 移动曲面拟合法

1 引 言

质量控制是数字高程模型 (Digital Elevation Model, DEM) 生产的重要环节之一。Östman^[1] (1987) 指出并不存在惟一的标准和量测方法来提

高 DEM 的质量, 他认为至少要考虑高程、坡度和曲率的正确性。而在 DEM 生产中则需从误差产生和扩散的每个过程入手, 采用一定的方法来削弱误差。误差按其特性, 可分为系统误差、随机误差和粗差, 其中粗差的产生主要出现在 DEMs 的

收稿日期: 2004-05-10; 修回日期: 2005-04-05

基金项目: 国家自然科学基金资助项目 (40271092); 香港特别行政区研究基金委员会部分资助项目 (香港理工大学 5068/99E)

作者简介: 杨晓云 (1979), 女, 天津塘沽人, 研究生, 专业方向为大地测量学与测量工程。

生产过程中,并最终表现在高程 Z 上。由于 DEM 数据中的粗差难以探测,在实际工作中,研究 DEM 的数据精度一般仅讨论系统误差和随机误差的影响。然而粗差对 DEM 数据所造成的空间扭曲往往最为严重,有时能导致 DEM 及其产品严重失真,甚至完全不能使用,因此很有必要研究一些方法对 DEM 的粗差进行探测和修正。

要有效地探测 DEM 数据中存在的粗差,必须在数据采集后形成的原始 DEM 数据中进行。原始的 DEM 数据结构有规则和不规则两类。从实用的角度来看,在规则格网的 DEM 数据中探测粗差相对简单一些,因而研究成果也相对丰富,如 Hannah^[2] (1981) 的基于坡度信息算法, Felicimo^[3] (1994) 的统计参数法,以及 Lopez^[4] (1997) 的主成分分析法等等。如果原始 DEM 为不规则数据,要转成规则格网点,则需经过数学内插的方法处理。在此过程中,原始数据中粗差点会影响到转换后的多个格网点,从而增加格网点粗差检测的难度。由此,李志林博士在其博士论文中提出点方式^[5]的不规则 DEM 粗差检测算法。本文在此基础上进一步完善该算法,以探讨效果更好地适用于不规则 DEM 数据的粗差探测方法。

2 基于点方式的粗差检测算法

所谓基于点方式的粗差检测算法^[5],是在不规则 DEM 数据中,以检测点 P 为中心,首先确定一个选取数据点范围的窗口(设半径为 R),然后计算窗口范围内所有点高程的算术平均值(或加权平均值)作为 P 点的估值 Z_i (或“真值”),最后计算 P 点高程值 Z_i 与估值 Z_i 的高程较差 δ_i 。对 DEM 中所有的点,可以得到 δ_i 的均值 $\bar{\delta} = \frac{1}{n} \sum_{i=1}^n \delta_i$
 $= \frac{1}{n} \sum_{i=1}^n (Z_i - Z_i)$ 和标准偏差 $S_\delta = \sqrt{\frac{1}{n} \sum_{i=1}^n (\delta_i - \bar{\delta})^2}$ 。以 k 倍 S_δ 作为粗差检测的阈值,对 DEM 任一数据点 i ,如果 $|\delta_i - \bar{\delta}| \geq kS_\delta$,则被认为含有粗差。一旦数据点被检测超限,用其估值 Z 来代替可疑的高程值 Z 。算法需要迭代进行,直到没有超过阈值的高程较差 δ 出现才结束运算。

3 新算法的推导

基于点方式的粗差检测算法,以邻域数据点

的高程均值为基础,建立检验统计量,方法简单易行。众所周知,DEM 内插点的精度,与参与内插的数据点范围(或多少)和内插方式有关。就内插函数而言,采用曲面函数将会优于双线性函数。因此,用邻域数据点的曲面函数拟合高程值取代简单的算术平均值来建立检验统计量,将会有更好的检测效果。此外,局部拟合区域的选点个数 N 也是影响算法对粗差敏感程度的重要因素。在点算法中,作者仅给出了一个普遍意义上的参考值,事实上, N 与检测区域的地形起伏大小以及数据分布密切相关。通常在地形起伏较大的区域,为了更好地逼近地面,应适当地减小参与局部拟合的数据点范围;而对于地形起伏不大的区域,由于地形特征较为单一,可以选定较大的拟合窗口。

新算法在计算中心点的高程估值时,首先以确定范围的数据窗口为计算单位,进行曲面的最小二乘拟合,得到一组高程残差列向量 V_A 。如果在 DEM 范围内移动数据窗口,将会得到一系列的高程残差向量 $V^T = [V_A \ V_B \ \dots \ V_W]^T$,其中 W 为窗口个数。由分析可知,每一个 DEM 高程点会在不同的数据窗口内参与拟合计算。换句话说,同一个 DEM 高程点在不同的拟合区域中将会计算出不同的残差。当以这些残差判断粗差时,不同拟合面内的残差值对同一 DEM 高程点是否为粗差的判断并不一致,即可能会出现一个拟合面内认为该点含有粗差,而在另一个拟合面内则认为该点为正确点。即,拟合残差值在统计意义上是相关的。

鉴于这种情况,必须考虑究竟在多少个拟合面内,DEM 高程点的残差值被检验超限时才应判断为粗差,这是本文的主要研究内容。为叙述方便,这里把有关算法简称 YXY 算法。

3.1 数学模型

设 DEM 全局区域总共有 M 个数据点,对中心点 P ,可选取其周围最近的 N 个点构成局部窗口进行拟合。拟合函数模型为

$$f(x, y) = a_0 + a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2 \quad (1)$$

式中 a_i 为待定参数, x, y 为数据点坐标。局部区域中心点对拟合参数影响很大,为了避免含有粗差的中心点对估计参数的影响,中心点不参与局部区域拟合。

3.2 粗差检测及修正

对于数据窗口 A , 利用最小二乘法进行拟合, 可以得到该窗口内高程残差值列向量 V_A 。当应用于全区域 DEM 时, 所有的数据窗口将可得到一系列的高程残差值列向量, 记作

$$V = [V_A^T \quad V_B^T \quad \dots \quad V_W^T]^T \quad (2)$$

以此为基础计算验后残差值方差 σ_0^2 , 并以 $k\sigma_0$ 作为检测粗差的阈值。

$$\sigma_0^2 = \frac{1}{W \times (N - 1)} (V_A^T V_A + V_B^T V_B + \dots + V_W^T V_W) \quad (3)$$

式中, W 为窗口数, N 为窗口内的数据点数。

当 $|V_{i,j}| > k\sigma_0$ 时, 怀疑第 i 个拟合面内的第 j 个点含有粗差。由于同一个 DEM 高程点在不同的拟合区域中将会计算出不同的残差值, 而这些残差值对同一 DEM 高程点是否为粗差的判断并不一致, 即在一个拟合面内认为该点含有粗差, 而另一个拟合面内认为是正常点。一般地, 含粗差的数据点被判为粗差的可能性会比其他的点更大一些, 因此, 可以统计每个点在相关的拟合面内被怀疑为粗差的百分比(*percent*)

$$percent_i = \frac{\text{点 } i \text{ 被怀疑为粗差的次数}}{\text{点 } i \text{ 出现在局部区域中的总次数}} \times 100\% \quad (4)$$

引入一个限值 a ($0.1 \sim 0.5$), 当 $percent_i \geq a$ 时, 则认为点 i 含有粗差。判断过程迭代进行, 迭代终止的标志是两次迭代的残差值中误差之差的绝对值小于一定的限值, 即 $|\sigma_{0_{r+1}} - \sigma_{0_r}| < \varepsilon$, 其中 ε 取某一极小值(如 $\varepsilon = 0.0001$)。

3.3 消除地性线对 DEM 粗差探测的影响

上述过程是把 DEM 数据作为表面光滑的连续曲面来处理, 当 DEM 为附有地貌特征点线或地形断裂线的非规则数据点时, 新算法可以采用两种不同的策略。第一种, 在拟合运算之前首先判断拟合面内是否有地性线通过, 对含有地性线的曲面, 按地性线将拟合面进行分割, 直到拟合面内不含地性线为止^[5]。虽然这样的策略会增加算法的复杂性, 但是可以保证算法能够克服拟合面跨越地性线时对粗差判断造成的影响。第二种, 通过大量试验数据统计得出最佳怀疑为粗差的百分比限值 a , 利用预先设定的最佳限值 a , 使算法适应各种复杂地形的情况。另外, 在算法中增加一个判断功能: 将探测到的粗差点分类, 不属于断裂线的数据点, 判断为粗差; 属于断裂线的数据点, 根据断裂线分别进行坎上坎下数据点的局部拟

合, 用拟合残差计算其方差和检测阈值, 只有当这些检查的数据点超过阈值时, 才最后判断它们为粗差点。下面的试验均由采用第二种策略的 YXY 算法完成。

4 算例

为使实验更具有代表性, 试验选取 3 组 DEM 数据, 按地形属性可分为平原、丘陵和山地, 每组约 1 000 个非规则数据点, 均取自某铁路新线的实测数据, 其地形特征值见表 1。由于局部区域选点数目 N 直接影响算法对粗差的敏感程度, 本文采用实验方法确定最佳 N 值。先设置初始最小点数, 观察实验效果是否理想, 如果不好, 则不断将点数递增, 以确定较好的点数选取范围, 结果见表 2。

表 1 地形属性参数

	平原	丘陵地	山地
平均坡度	2.558°	5.117°	24.365°
$\bar{\sigma}_0$	0.371 3	0.503 4	2.067 9

表 2 不同地形的试验常数

	平原	丘陵地	山地
N	20	16	12
k	3	3	3

YXY 算法的试验过程如下:

1. 用蒙特卡洛法(Monte Carlo Method)模拟一组含有 n 维粗差的数据($4.5\bar{\sigma}_0 \sim 7\bar{\sigma}_0$);
 2. 将粗差随机加到 DEM 数据点上, 设定限值 a , 用式(4)计算结果进行粗差的判释, 记录试验的正确判断粗差数目, 误判粗差数目;
 3. 重复 1~ 2, 共做 200 次试验。用下式计算粗差探测的正确检测率(pd)与误判率(pc)
- $$pd = \frac{\sum(\text{正确判断粗差数目}/n)}{200} \quad (5)$$
- $$pc = \frac{\sum(\text{误判粗差数目}/M)}{200} \quad (6)$$
4. 改变限值 a 的大小, 重复 1~ 3, 得到与 a 对应的 pd 与 pc ;
 5. 将一系列与 a 对应的 pd 或 pc 曲线拟合, 得 n 维粗差 a 与 pd (或 pc) 的拟合曲线;
 6. 选取不同的粗差维数 n , 重复 1~ 5 的试

验, 获得不同粗差率(模拟的粗差维数 n 除以总的 DEM 数据点数 M) 的 a 与 pd (或 a 与 pc) 拟合曲线。

3 种地形类型的 DEM (平原、丘陵和山地) 的试验方案相同。因论文篇幅所限, 下面仅给出粗差率分别为 1%, 3%, 5% 的 a 与 pd , a 与 pc 拟合

曲线(见图 1(a), 图 1(b), 图中实线代表粗差率为 1%, 虚线代表粗差率为 3%, 点线代表粗差率为 5%)。为更好地比较新算法的优点, 采用基于点方式算法对相同的实验数据经 200 次试验, 统计结果列于表 3。

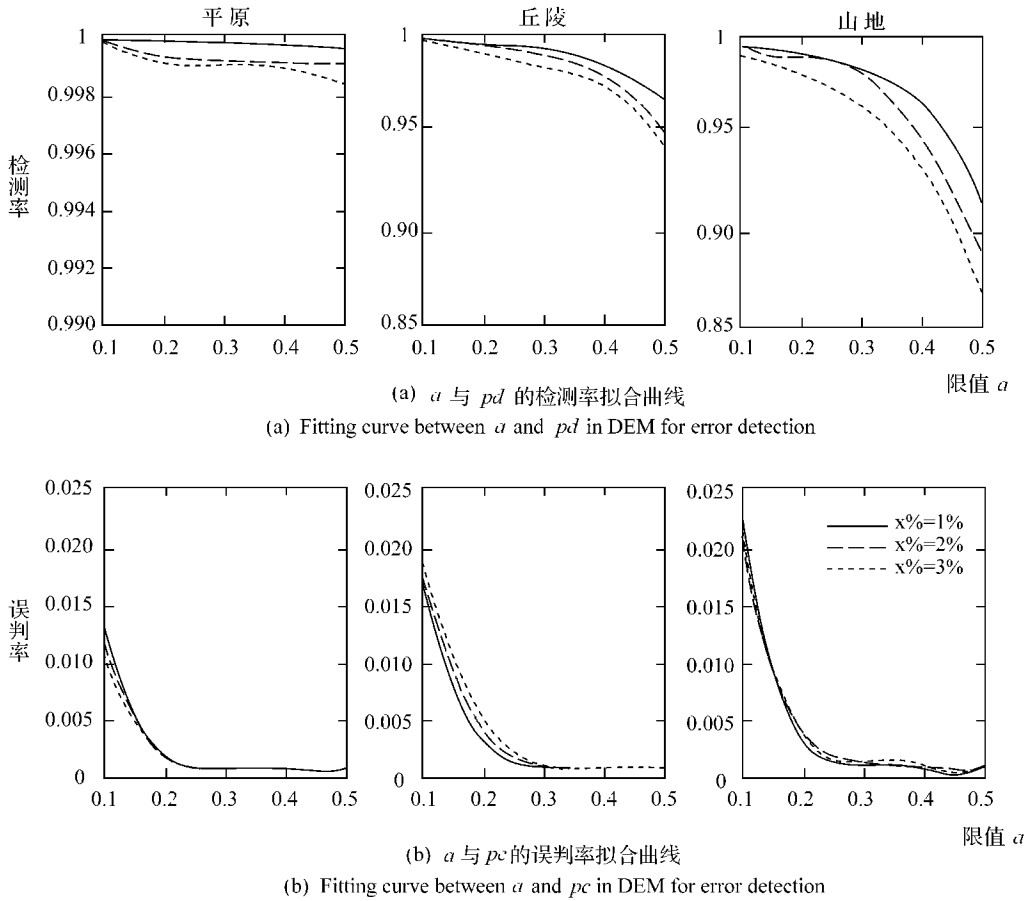


图 1

Fig. 1

表 3 点方式算法与 YXY 算法相差检测比较

Tab. 3 The comparison between the point wise methods and YXY algorithm on detection of gross errors

地形分类	粗差量	点方式算法			YXY 算法		
		检测率	误判率	检测率	限值 a	误判率	限值 a
平地	1%	0.917 5	0	1		0	
	3%	0.910 8	0	1	$a = 0.20$	0	$a = 0.20$
	5%	0.882 5	0	1		0.000 14	
丘陵	1%	0.535 0	0	0.994		0.000 26	
	3%	0.546 7	0	0.992	$a = 0.25$	0.000 38	$a = 0.25$
	5%	0.451 5	0	0.990		0.000 47	
山地	1%	0.392 5	0	0.982		0.000 37	
	3%	0.356 7	0	0.982	$a = 0.3$	0.000 65	$a = 0.3$
	5%	0.304 0	0	0.962		0.000 68	

地形的起伏程度是影响粗差判释的一个重要因素。试验结果显示,无论是点方式算法还是YXY算法,粗差检测率都会随着地形起伏的增大而呈递减趋势。另外,粗差率也会影响检测效果,通常表现为粗差率越高,粗差检测率 pd 越低。这是因为当粗差点增多时,局部数据窗口含有粗差的可能性增大,因而对曲面最小二乘拟合参数的影响增强,使准确探测粗差的可靠性大大降低,从而导致真正的粗差数据未能判释,某些正确的数据却被怀疑,即误判率 pc 增加。

在YXY算法中,局部区域的拟合残差计算的验后中误差 $\bar{\sigma}_0$ 小于点方式算法计算的 $\bar{\sigma}'_0$,因而前者较后者有更强的粗差敏感度。YXY算法在探测粗差中使用百分比($peri$)来判断粗差,当粗差率固定时,设定不同的限值 a 会有不同的 pd 和 pc ,随着 a 增大, pd 和 pc 都会随之降低。粗差率愈高, a 愈大, pd 和 pc 下降得愈快。因此,选取合适的 a 值,既要能有效地抑制 pc ,又不过分降低 pd 。 a 值的确定与原始DEM数据的属性有关。实验数据结果显示,YXY算法在平原DEM中采用 $a=0.2$ 效果最佳;在丘陵DEM中采用 $a=0.25$ 效果最佳;在山地DEM中,采用 $a=0.3$ 效果最佳。表3列有最佳 a 值的检测率,对比点方式的结果可以发现,无论是在平原、丘陵还是在山地,采用YXY算法正确探测的粗差率比点方式算法有明显改进。此外,对于大于 $7\bar{\sigma}_0$ 的粗差数据,YXY算法的检测率几乎为100%。

本文的试验虽然建立在DEM模型为表面光滑连续的模拟数据基础上,但事实上,YXY算法在附有地貌特征点线的实际DEM数据应用中,使用推荐的最佳限值 a ,仍然能获得很好的检测效果。如图2所示为某工程的不规则DEM数据的算例图。其中图2(a)为原始的DEM数据生成的等高线图;图2(b)为 $a=0.001$ 时,被判断为粗差的原始数据点;图2(c)沿地性线加的粗差点;图2(d)在山头加的粗差点。当YXY算法选取推荐的最佳限值 $a=0.2$ 时,原始的DEM数据点没有检测到粗差;但对加到DEM数据上的所有模拟粗差,均能准确判断。

地形特征点线(如山脊线、谷底线)是地形变化的骨架,在DEM采样密度适宜的情况下,与整个地形趋于连续。YXY算法以数据点在不同拟合面内的拟合残差为粗差探测对象,每一个DEM点都会参与多个不同拟合面的平差运算,这些拟

合面中有一部分会跨越地形特征点、线,也有另一部分处于平滑连续的区域。由所有拟合面的残差统计验后方差 $\bar{\sigma}_0$,并根据验后方差 $\bar{\sigma}_0$ 计算检测阈值,再加上经过各种复杂地形试验验证的最佳限值 a ,就完全可以避免因局部窗口含有地形点线给粗差判断造成的影响,提高算法判断粗差的正确率。

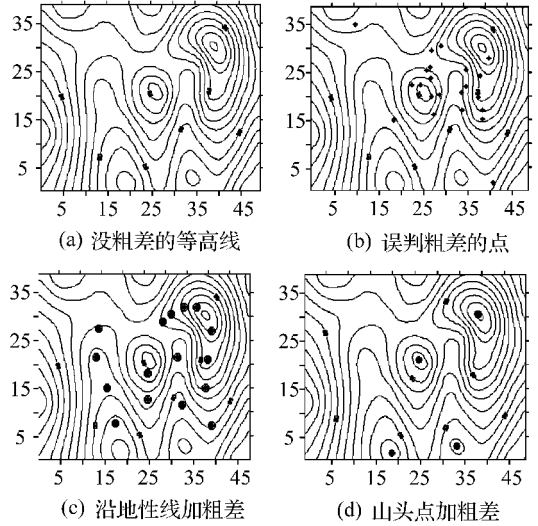


图2 某工程不规则DEM数据的粗差探测算例

Fig. 2 The cases of detecting gross errors in a engineering irregular DEM data

地形断裂线,如陡坎、峭壁等,通常表现为地形的突然变化(陡起或陡落),在非规则数据中表现为坎上和坎下两个点的平面坐标相等(或近似相等),但高程值相差很大。这种情况对算法的影响主要是正常的点误判为粗差点。YXY算法遇到有地形断裂线的DEM,采用第二种处理方式也能表现出稳健的特性。受篇幅所限,试验结果省略。

5 结束语

通过大量的仿真模拟试验分析和一些实际的DEM数据验证,可以得出如下结论:

1. 使用移动窗口曲面拟合的YXY算法,不仅较基于点方式算法更有效地探测DEM不规则数据中的粗差(特别是小粗差),而且由于采用百分比 $percent$ 判断粗差,使其探测粗差的可靠性增强,误判率降低。

2. 地形起伏是影响粗差判释的重要因素,无论是YXY算法还是基于点方式算法,粗差检测率会随着地形起伏的增大而递减。

3. 粗差率也会影响粗差检测率, 在 YXY 算法和基于点方式算法中, 粗差检测率都会随着粗差率的增大而降低。

4. 当粗差率一定时, YXY 算法的粗差检测率和误判率均会随着限值 a 的增大而降低。恰当地设定限值 a , 可以有效地抑制误判率, 同时又不过分降低检测率。本文实验是建立在大量的模拟仿真试验基础上的, 因而限值 a 在不同地形类型中的试验结果将对实际生产工作有参考价值。即使是数据表面并非光滑连续、包含有地形特征点或地形断裂线的不规则 DEM, YXY 算法仍能得到满意的检测结果。

参考文献:

[1] ÖSTMAN A. Quality Control of Photogrammetrically Sam-

pled Digital Elevation Models [J]. Photogrammetric Record, 1987, (12): 333-341.

[2] HANNAH M J. Error Detection and Correction in Digital Terrain Models[J]. Photogrammetric Engineering and Remote Sensing, 1981, 47: 63-69.

[3] FELICÍSIMO A. Parametric Statistical Method for Error Detection in Digital Elevation Models[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 1994, 49: 29-33.

[4] LÓPEZ C. Locating Some Types of Random Errors in Digital Terrain Models[J]. International Journal of Geographic Information Science, 1997, (11): 677-698.

[5] LI Zhi-lin, ZHU Qing. Digital Elevation Model[M]. Wuhan: Press of Wuhan Technical University of Surveying and Mapping, 2000. (in Chinese)

(责任编辑: 雷秀丽)

测绘科技专著出版基金申请办法

国家测绘局测绘科技专著出版基金向全社会开放, 常年受理有关申请。

国家测绘局测绘科技专著出版基金设立于 1999 年, 主要用于资助对推动测绘科技进步具有重要意义, 但因预期经济效益不好而难以出版的测绘学术理论与应用技术专著的出版, 同时适当资助测绘教材和测绘标准的出版。2002 年, 出版基金增设对优秀博士论文出版的资助。

国家测绘局测绘科技专著出版基金设管理委员会。基金管理委员会主任委员: 陈俊勇, 副主任委员: 喻永昌、白泊。

申请测绘科技专著出版基金的注意事项如下:

一、基金委办公室设在中国地图出版社总编办公室, 地址: 北京市宣武区白纸坊西街 3 号, 邮编: 100054, 联系电话: (010) 63531961(传真)、83543902。

二、申请者可随时向基金办索取或从中国地图出版社网站(www.sinomaps.com) 下载《测绘科技专著出版基金申请书》, 也可直接使用国家科学技术学术著作出版基金申请书。

三、申请材料必须一式五份一次性提供齐全, 其中一份必须是签名和加盖公章的原件, 否则不予受理。申请材料包括:

1. 申请书(填写基金办提供的统一表格);
2. 全部书稿(包括内容简介、序、前言、目录、正文、附录、参考文献等);
3. 能反映作者水平的材料(如奖状、成果鉴定书等)。

四、基金办在受理申请和将材料提交基金委评审后, 会将是否资助的结果及时通知申请者, 但申请材料一律不退。请申请者自留申请材料底稿。

五、对于作者通过所在单位或其他渠道能争取到部分出版补贴的项目, 基金委在评审时将优先考虑资助。请作者另行提供一份加盖公章和有关负责人签字的《出版补贴证明》。

六、对于获得“测绘科技专著出版基金”资助的项目, 基金办将以同一申报材料协助作者申报“国家科学技术学术著作出版基金”。如作者不同意申请, 则应附函说明; 如同意, 则请填写《科学技术著作出版意向书》。