

地形起伏度最佳统计单元算法的比较研究

陈学兄¹, 常庆瑞², 毕如田¹, 刘正春¹, 张小军³

(1. 山西农业大学 资源环境学院, 山西 太谷 030801; 2. 西北农林科技大学
资源环境学院, 陕西 杨凌 712100; 3. 山西农业大学 园艺学院, 山西 太谷 030801)

摘要: 基于 90 m 分辨率 SRTM DEM 数据, 利用邻域分析法在不同大小窗口下 ($2 \times 2, 3 \times 3, 4 \times 4, \dots, 30 \times 30$) 对山西省地形起伏度进行了提取, 运用人工工作图法、最大高差法、模糊数学法、均值变点分析法和累积和 (CUSUM) 分析算法分别计算最佳统计单元, 对统计单元计算方法的准确性及适用性进行分析, 提出山西省地形起伏度提取的最佳统计单元。研究分析表明: 均值变点分析法和累积和分析算法是相对比较有效的方法; 运用均值变点分析法与累积和分析算法计算所得的最佳统计单元分别为 $11 \times 11 (0.980\ 1\ \text{km}^2)$ 和 $14 \times 14 (1.587\ 6\ \text{km}^2)$; 均值变点分析法是计算最佳统计单元的一种最为理想的方法; 对于同一地区而言, 不同 DEM 数据类型、不同分辨率 DEM, 所使用的最佳统计单元大小也不同。

关键词: 数字高程模型; 地形起伏度; 最佳统计单元; 算法比较; 山西省

中图分类号: S157; P237

文献标识码: A

文章编号: 1005-3409(2018)01-0052-05

Comparison Study on the Best Statistical Unit Algorithms of Relief Amplitude

CHEN Xuexiong¹, CHANG Qingrui², BI Rutian¹, LIU Zhengchun¹, ZHANG Xiaojun³

(1. College of Resources and Environment, Shanxi Agricultural University, Taigu, Shanxi 030801, China;

2. College of Natural Resources and Environment, Northwest A&F University, Yangling,

Shaanxi 712100, China; 3. College of Horticulture, Shanxi Agricultural University, Taigu, Shanxi 030801, China)

Abstract: Based on the 90-m resolution SRTM DEM data, the relief amplitude of Shanxi Province was extracted with the neighborhood analysis method at different neighborhood windows ($2 \times 2, 3 \times 3, 4 \times 4, \dots, 30 \times 30$), the best statistical unit was calculated by using the artificial drawing method, maximum elevation difference method, fuzzy mathematics method, mean change-point analysis method and cumulative sum analysis method, the accuracy and applicability of statistical unit calculation methods was analyzed, and the best statistical unit of relief amplitude in Shanxi Province was determined. The results indicate that the mean change-point analysis method and cumulative sum analysis method are more effective than the artificial drawing method, maximum elevation difference method and fuzzy mathematics method, the best statistical units of the mean change-point analysis method and cumulative sum analysis method are $11 \times 11 (0.980\ 1\ \text{km}^2)$ and $14 \times 14 (1.587\ 6\ \text{km}^2)$, respectively. The mean change-point analysis method is one of the best ideal methods to calculate the best statistical unit. The best statistical unit is different between different DEM data types and different resolution DEM for the same area.

Keywords: digital elevation model; relief amplitude; best statistical unit; algorithm comparison; Shanxi Province

地形起伏度是定量描述地貌形态, 宏观描述地形变化整体趋势的主要指标^[1-2], 也是导致水土流失的最直接因素, 它不仅具有数学意义且具备土壤侵蚀和

地貌学的意义, 已被广泛应用于各个研究领域, 如水土流失定量评价^[3]、潜在水土流失评价^[4]、地质环境质量定量评价^[5]、区域滑坡灾害评价^[6]、生态敏感性

评价^[7]、水土流失敏感性评价^[8-9]等。在此基础上,国内学者对地形起伏度最佳统计单元的计算方法有了一定的研究^[10-12],最佳统计单元的确定属于典型的变点分析问题,其计算方法有:人工作图法、最大高差法、模糊数学法、均值变点分析法和累积和 CUSUM 分析算法,除人工作图法外,其他方法都以定量的形式寻找突变点,哪种方法更适合山西省地形起伏度最佳统计单元的计算还需研究。鉴于此,本研究使用 SRTM DEM 数据,分析并探讨最佳统计单元各计算方法的准确性及适用性,确定山西省地形起伏度提取的最佳统计单元大小,以期生态敏感性评价、水土流失定量评价等相关研究提供数据资料,也为山西省地形起伏度的进一步应用研究提供参考。

1 研究区概况

山西省,位于我国华北西部的黄土高原地带,东隔太行山与河北为邻,西隔黄河与陕西相望,南与河南省接壤,北与内蒙古自治区相连,疆域轮廓呈由东北斜向西南的平行四边形,全省总面积约 16 万 km²,地貌类型复杂多样,有平原、台地、山地、丘陵、高原等,其中山地丘陵约占全省总面积的 80% 之余。山西属温带暖温带、半湿润半干旱大陆性季风气候,降水偏少,年降水量为 400~600 mm,自西北向东南递增。境内矿产资源丰富多样,且分布广泛,尤以煤炭资源储量大,质量优,品种齐全,易开采而著称于世。

2 数据与研究方法

2.1 基础数据

本研究使用分辨率为 90 m 的 SRTM DEM 数据,数据采样网格大小为 90 m,格式为 ArcGIS 的 Grid 格式,地理坐标系为:WGS_1984_Albers。

2.2 研究方法

2.2.1 地形起伏度的提取 使用邻域分析法提取山西省地形起伏度,所使用矩形窗口大小为 $n \times n$ 像元 ($n=2,3,\dots,30$)。地形起伏度提取的步骤如下:(1) 统计 $n \times n$ ($n=2,3,\dots,30$) 窗口内像元的最大和最小值;(2) 计算各窗口内最大值和最小值的差值;(3) 统计不同窗口的平均起伏度值(表 1)。

2.2.2 最佳统计单元的计算 运用人工作图法^[10-11]、最大高差法^[10]、模糊数学法^[10]、均值变点分析法^[1,11,13]和累积和 CUSUM 分析算法^[12]分别计算最佳统计单元,研究各计算方法在确定最佳统计单元上的区别与适用性。

(1) 人工作图法。利用统计软件对平均起伏度进行回归分析,提取拟合度最好的拟合曲线,从拟合

曲线上人工判断拐点。

(2) 最大高差法。在 (i,j) 样本集中,样点 i 在邻域面积 j 处的高差变化值 (dH) 为: $dH_j = H_j - H_{(j-1)}$,比较 i 点所有的高差变化值 dH_j ,取最大的 dH_j 所对应的 j 点为拐点,则拐点所对应的面积为最佳统计单元的面积。

(3) 模糊数学法。样本集 (i,j) 中,样点 i 在 j 邻域面积处,有:

$$R_j^L = \frac{\{\sum_{k \leq j} a_k [\Delta H_k - \Delta H_{(k-1)}]^2\}}{\sum_{k \leq j} a_k} \quad (1)$$

$$R_j^R = \frac{\{\sum_{k \geq j} a_k [\Delta H_k - \Delta H_{(k-1)}]^2\}}{\sum_{k \geq j} a_k} \quad (2)$$

$$R_j = |R_j^L - R_j^R| \quad (3)$$

式中:模糊系数 (a_k) 均匀分布于 $(0,1]$ 区间,其中 $2 \leq k \leq N$, N 为邻域面积个数 ($N=29$), $a_2=1$;定义拐点为最大 R_j 值对应的那一点。

(4) 均值变点分析法。均值变点分析法对恰有一个变点的检验最有效,该方法已被广泛应用于计算地形起伏度最佳统计单元,根据均值变点分析法的原理构建样本序列,具体过程为^[11]:首先,计算各统计单元下的单位地势度 (T),单位地势度 (T_i) = 平均起伏度 (t_i) / 邻域面积 (s_i), ($i=2,3,\dots,30$);其次,对单位地势度 (T) 取对数,得到序列 X , X 为 $\{x_k, k=1,2,3,\dots,29\}$;最后,根据下面公式对表 1 数据进行计算处理,计算出统计量 S 和 S_i 的值。

$$\bar{X} = \sum_{i=1}^N \frac{x_i}{N} \quad (4)$$

$$S = \sum_{i=1}^N (x_i - \bar{X})^2 \quad (5)$$

$$S_i = \sum_{t=1}^{i-1} (x_t - \bar{X}_{i1})^2 - \sum_{t=i}^N (x_t - \bar{X}_{i2})^2 \quad (6)$$

式中: \bar{X} 表示总体样本的算术平均值; N 表示样本总数; S 和 S_i 分别为总的离差平方和及两段样本的离差平方和之差。

(5) 累积和 CUSUM 分析算法。累积和分析算法可对原始数据序列中是否存在突变以及突变点的位置情况进行分析,具体计算过程为^[12]:① 计算数据序列 $\{x_i, i=1,2,3,\dots,29\}$ 的算术平均值 \bar{X} ;② 计算 \bar{X} 与数据序列第 i 个元素的差值求取累加值 S_i ,计算公式为: $S_i = S_{i-1} + (x_i - \bar{X})$, ($i=1,2,3,\dots,N$),其中 $S_0=0$ 。③ 在步骤 2 的基础上计算 S_{diff} , $S_{\text{diff}} = S_{\text{max}} - S_{\text{min}}$, $S_{\text{max}} = \max_{i=0,1,\dots,N} S_i$, $S_{\text{min}} = \min_{i=0,1,\dots,N} S_i$ 。④ 基于 CUSUM 指标确定突变点的位置,如式: $|S_m| = \max_{i=0,1,2,\dots,N} |S_i|$,式中 m 为原始数据序列中突变发生的最后一个点, $m+1$ 为突变发生后的第一个点。

3 结果与分析

3.1 地形起伏度与邻域面积关系分析

计算不同统计单元的邻域面积(表 1)。由表 1 可知,平均起伏度值随邻域面积的增大逐渐增大,但当邻域面积达到一定的阈值时,其变化趋势逐

渐平缓。

3.2 山西地形起伏度最佳统计单元的探讨

3.2.1 最佳统计单元的确定

(1) 人工工作图法。利用统计软件对表 1 中的平均起伏度值进行回归分析,提取拟合度最好的 2 条拟合曲线,见图 1。

表 1 统计单元与地形起伏度对应关系

单元大小	项目	2×2	3×3	4×4	5×5	6×6	7×7	8×8	9×9	10×10	11×11
邻域面积/10 ⁴ m ²		3.24	7.29	12.96	20.25	29.16	39.69	51.84	65.61	81.00	98.01
平均起伏度/m	原始数据	25.53	46.38	63.75	78.62	91.84	103.71	114.74	124.93	134.43	143.44
	变化值		20.75	17.37	14.87	13.22	11.87	11.03	10.19	9.50	9.01
单元大小		12×12	13×13	14×14	15×15	16×16	17×17	18×18	19×19	20×20	21×21
邻域面积/10 ⁴ m ²		116.64	136.89	158.76	182.25	207.36	234.09	262.44	292.41	324.00	357.21
平均起伏度/m	原始数据	151.93	157.94	167.79	175.44	182.88	189.67	196.51	203.19	209.83	215.78
	变化值	8.49	6.01	9.85	7.65	7.44	6.79	6.84	6.68	6.64	5.95
单元大小		22×22	23×23	24×24	25×25	26×26	27×27	28×28	29×29	30×30	
邻域面积/10 ⁴ m ²		392.04	428.49	466.56	506.25	547.56	590.49	635.04	681.21	729.00	
平均起伏度/m	原始数据	221.90	227.81	233.61	239.39	245.06	250.67	255.92	261.16	266.13	
	变化值	6.12	5.91	5.80	5.78	5.67	5.61	5.25	5.24	4.97	

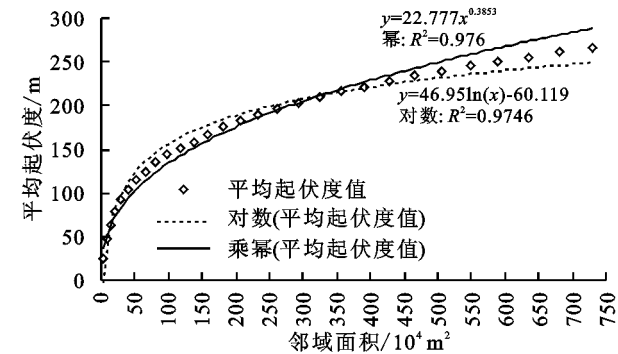


图 1 邻域面积与平均地形起伏度对应关系拟合曲线

由图 1 可知,使用 SRTM DEM 数据分析研究山西省地形起伏度时,采用幂模型和对数模型对平均起伏度

值和邻域面积进行拟合,幂模型的拟合效果最好,而对数模型次之,拟合方程分别为 $y=22.777x^{0.3853}$, $y=46.95\ln(x)-60.119$;平均起伏度值随邻域面积的增大逐渐增大,但其增幅一直在降低并渐渐趋于平缓,由于从拟合曲线上人工判断拐点容易受主观因素的影响,所以此方法不参与比对分析。

(2) 最大高差法。根据最大高差法的含义,由表 1 数据得知平均起伏度的最大变化值在 3×3 处,即 3×3 (0.072 9 km²)为使用此方法计算得出的最佳统计单元。

(3) 模糊数学法。根据模糊数学法的概念,各变量的取值为 $a_k=1, 27/28, \cdots, 1/28 (k=2, 3, 4, \cdots, N), N=29$ 。计算结果见表 2。

表 2 模糊数学法统计结果

j	3	4	5	6	7	8	9	10	11
R_j^{\pm}	434.72	369.43	321.82	287.14	260.14	239.32	222.29	208.14	196.39
R_j^{\mp}	105.33	90.22	79.75	71.83	65.82	60.75	56.64	53.28	50.34
R_j	329.39	279.21	242.07	215.31	194.32	178.57	165.65	154.86	146.05
j	12	13	14	15	16	17	18	19	20
R_j^{\pm}	186.34	175.65	170.70	164.42	158.99	153.97	149.72	146.01	142.82
R_j^{\mp}	47.93	49.32	43.35	41.33	39.33	38.29	36.87	35.46	33.73
R_j	138.41	126.33	127.35	123.09	119.66	115.68	112.85	110.55	109.09
j	21	22	23	24	25	26	27	28	29
R_j^{\pm}	139.84	137.35	135.19	133.34	131.81	130.55	129.56	128.80	128.30
R_j^{\mp}	33.36	32.34	31.60	30.92	29.93	28.82	27.05	26.54	24.70
R_j	106.48	105.01	103.59	102.42	101.88	101.73	102.51	102.26	103.60

由表 2 可知,在 $j=3$ 时, R_j 的值为最大,即使用该方法计算得到的最佳统计单元为 3×3(0.072 9 km²)。

(4) 均值变点分析法。根据均值变点分析法的计算过程,计算得出统计量 $S=22.23, S_i$ 的值见表 3。

由表 3 可以得知,在第 10 个点时 S 与 S_i 的差值达到最大,这一点就是所要求的点,即利用该方法计算得到的最佳统计单元为 $11 \times 11(0.980\ 1\ \text{km}^2)$ 。

(5) 累积和(CUSUM)分析算法。为确定其突变情况,依据累积和算法分析突变点的原理对表 1 中的数据序列进行了分析,分析原始数据序列中累加

和 S_i 与统计单元窗口的关系,见图 2。由图 2 可知,随着邻域面积的增大,CUSUM 值逐渐减小,最小值(-827.34)出现在 14×14 处,然后随着邻域面积的增大,累积和(CUSUM)的值也逐渐增大,因此可确定 $14 \times 14(1.587\ 6\ \text{km}^2)$ 为原始数据序列发生突变的位置。

表 3 均值变点分析法的统计结果

i	2	3	4	5	6	7	8	9	10	11	12	13	14	15
S_i	18.19	14.70	11.81	9.40	7.33	5.52	3.90	2.43	1.06	0.23	1.46	2.65	3.80	4.93
$S-S_i$	4.04	7.53	10.42	12.83	14.90	16.71	18.33	19.80	21.17	22.00	20.77	19.58	18.43	17.30

i	16	17	18	19	20	21	22	23	24	25	26	27	28	29
S_i	6.05	7.15	8.26	9.36	10.47	11.59	12.71	13.85	15.00	16.17	17.35	18.54	19.75	20.98
$S-S_i$	16.18	15.08	13.97	12.87	11.76	10.64	9.52	8.38	7.23	6.06	4.88	3.69	2.48	1.25

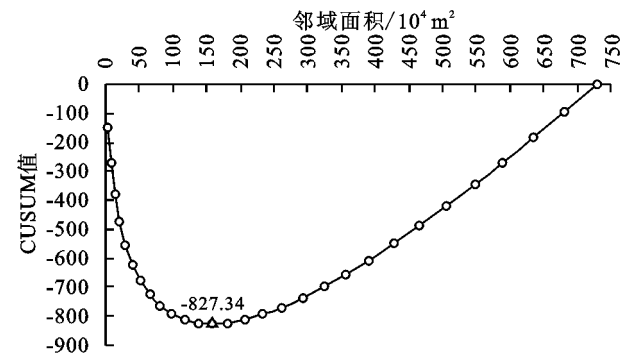


图 2 平均起伏度的累积和 CUSUM 图

3.2.2 计算结果对比分析 不同计算方法计算所得的最佳统计单元面积见表 4。与已有研究成果^[13]进行对比分析,以验证不同计算方法其结果的准确性。在 ArcGIS 支持下,首先将不同计算方法所得的最佳

统计单元下的山西省地形起伏度分为 5 个等级:平坦($0 \sim 30\ \text{m}$)、微起伏($30 \sim 70\ \text{m}$)、小起伏($70 \sim 200\ \text{m}$)、中起伏($200 \sim 500\ \text{m}$)和大起伏($>500\ \text{m}$),统计各级所对应的像元数,并计算各级所占的面积百分比(表 5),然后依据《中国 1:100 万地貌制图规范(试行)》中的地貌基本形态划分体系及其他相关研究成果^[14-15],对基本地貌类型进行了简单的划分,并计算各类地貌所占的面积百分比,结果见表 5。

表 4 不同方法得到的统计单元面积结果对比

计算方法	最大高差法	模糊数学法	均值变点分析法	累积和算法
最佳统计单元面积/ km^2	0.0729	0.0729	0.9801	1.5876

表 5 不同方法计算结果分类

项目	比例	最大高差法/%	模糊数学法/%	均值变点分析法/%	累积和算法/%	文献 ^[13] 数据/%
地形起伏度	平坦($0 \sim 30\ \text{m}$)	35.82	35.82	13.90	11.97	15.69
	微起伏($30 \sim 70\ \text{m}$)	42.94	42.94	9.83	8.61	21.65
	小起伏($70 \sim 200\ \text{m}$)	21.06	21.06	53.60	47.70	56.14
	中起伏($200 \sim 500\ \text{m}$)	0.18	0.18	22.30	30.60	6.51
	大起伏($>500\ \text{m}$)	0.00019	0.00019	0.37	1.12	0.01
基本	平原、盆地($<30\ \text{m}$)	35.82	35.82	13.90	11.97	
地貌	丘陵($30 \sim 200\ \text{m}$)	63.99	63.99	63.43	56.31	
类型	山地($>200\ \text{m}$)	0.19	0.19	22.67	31.72	

为比对不同方法其计算结果的准确性,使用 SPSS 统计软件分别分析最大高差法(或模糊数学法)、均值变点分析法和累积和算法与文献^[13]数据的相关性,分析得出均值变点分析法与文献^[13]中的起伏度最接近,相关系数为 0.89,累积和分析算法次之,相关系数为 0.757,最大高差法与模糊数学法相关性最差,相关系数为 0.39,说明最大高差法与模糊数学法已不适用,这与前人研究结果相一致^[10],均值

变点分析法和累积和分析算法是相对比较有效的方法,而且均值变点分析法更适合山西省地形起伏度最佳统计单元的确定。

通过上述分析可以得出:不同方法计算所得的最佳统计单元面积有所不同(表 4),人工作图法易受主观因素的影响,很难确定最佳统计单元,最大高差法和模糊数学法所得的最佳统计单元均为 $3 \times 3(0.072\ 9\ \text{km}^2)$,均值变点分析法所得的最佳统计单元为 11×11

(0.980 1 km²),而累积和分析算法的最佳统计单元为 14×14(1.587 6 km²);由表 5 得知,均值变点分析法和累积和分析算法所得的最佳统计单元下的起伏度与文献^[13]中的起伏度较接近,以小起伏为主,其面积占研究区总面积的 1/2 左右,且这两种方法划分的基本地貌类型所占面积比例与实际较相符,山地丘陵占 80% 之余;均值变点分析法是计算山西省地形起伏度最佳统计单元的一种最为理想的方法。

4 结论与讨论

(1) 基于 90 m 分辨率 SRTM DEM 数据,利用邻域分析法在不同大小窗口下(2×2,3×3,4×4,...,30×30)对山西省地形起伏度进行了提取(表 1),平均起伏度值起初随邻域面积的增大逐渐增大,但邻域面积达到一定的阈值时,其变化趋势逐渐平缓,采用幂模型对平均起伏度值和邻域面积的拟合效果最好,对数模型次之;

(2) 运用人工工作图法、最大高差法、模糊数学法、均值变点分析法和累积和(CUSUM)分析算法分别计算最佳统计单元。结果表明:人工工作图法易受主观因素的影响,最大高差法和模糊数学法计算所得最佳统计单元下的地形起伏度分级结果与实际很不相符,运用均值变点分析法与累积和分析算法计算所得的最佳统计单元分别为 11×11(0.980 1 km²)和 14×14(1.587 6 km²),最佳统计单元下的地形起伏度分级结果与实际较相符,且使用均值变点分析法和累积和分析算法划分的基本地貌类型与实际较相符,均值变点分析法和累积和分析算法是相对比较有效的方法,其他方法已不适用,这与前人研究结果相一致;

(3) 均值变点分析法的结果与已有研究结果^[13]最接近,研究区地形起伏度以小起伏为主,说明该方法是计算最佳统计单元的一种最为理想的方法;

(4) 已有研究^[13]所用数据为 30 m 分辨率的 ASTERG DEM 数据,与本文所用数据类型及数据分辨率有所不同,而文献^[13]使用均值变点分析法计算所得的地形起伏度最佳统计单元为 17×17(0.260 1 km²),可见对于同一地区而言,不同 DEM 数据类型、不同分辨率 DEM,所使用的最佳统计单元大小是不一样的。

本研究分析研究出了适合山西省地形起伏度最佳统计单元的最为理想的计算方法,对均值变点分析法和累积和分析算法计算所得最佳统计单元大小存在差异还需进一步的研究。

参考文献:

- [1] 刘育成,赵廷宁.基于变点分析法提取废弃采石场地形起伏度的方法[J].水土保持研究,2016,23(3):269-273.
- [2] 涂汉明,刘振东.中国地势起伏度研究[J].测绘学报,1991,20(4):311-319.
- [3] 刘新华,杨勤科,汤国安.中国地形起伏度的提取及在水土流失定量评价中的应用[J].水土保持通报,2001,21(1):57-59,62.
- [4] 马晓微,杨勤科,刘宝元.基于 GIS 的中国潜在水土流失评价研究[J].水土保持学报,2002,16(4):49-53.
- [5] 闫满存,李华梅,王光谦.广东沿海陆地地质环境质量定量评价研究[J].工程地质学报,2000,8(4):416-425.
- [6] 祁元,刘勇,杨正华,等.基于 GIS 的兰州滑坡与泥石流灾害危险性分析[J].冰川冻土,2012,34(1):96-104.
- [7] 刘康,欧阳志云,王效科,等.甘肃省生态环境敏感性评价及其空间分布[J].生态学报,2003,23(12):2711-2718.
- [8] 凡非得,王克林,熊鹰,等.西南喀斯特区域水土流失敏感性评价及其空间分异特征[J].生态学报,2011,31(21):6353-6362.
- [9] 徐天献,王玉宽,傅斌.汶川地震重灾区土壤侵蚀敏感性评价[J].中国水土保持,2011,18(1):39-42.
- [10] 赵斌滨,程永锋,丁士君,等.基于 SRTM-DEM 的我国地势起伏度统计单元研究[J].水利学报,2015,46(S1):284-290.
- [11] 陈学兄,常庆瑞,郭碧云,等.基于 SRTM DEM 数据的中国地形起伏度分析研究[J].应用基础与工程科学学报,2013,21(4):670-678.
- [12] 王志恒,胡卓玮,赵文吉,等.应用累积和分析算法的地形起伏度最佳统计单元确定[J].测绘科学,2014,39(6):59-64.
- [13] 陈学兄,毕如田,刘正春,等.基于 ASTER GDEM 数据的山西地形起伏度分析研究[J].山西农业大学学报:自然科学版,2016,36(6):417-421.
- [14] 中国科学院地理研究所.中国 1:1000000 地貌图制图规范(试行)[M].北京:科学出版社,1987:33-34.
- [15] 李炳元,潘保田,韩嘉福.中国陆地基本地貌类型及其划分指标探讨[J].第四纪研究,2008,28(4):535-543.