

均值变点分析法在最佳集水面积阈值确定中的应用

常直杨¹, 王建^{1,2}, 白世彪^{1,2}, 张志刚¹

(1. 南京师范大学地理科学学院, 江苏 南京 210023)

(2. 南京师范大学虚拟地理环境教育部重点实验室, 江苏 南京 210023)

[摘要] 利用 DEM 数据提取河流地貌参数在地貌学和水文学研究中具有重要意义, 其中集水面积阈值的准确确定是关键环节, 然而, 目前对集水面积阈值的确定存在随意性和主观性. 为了探讨如何确定提取河网的最佳集水面积阈值, 本文以青藏高原东缘杂谷脑河流域为例, 基于 SRTM-DEM 数据, 在利用 ArcGIS 水文分析模块计算不同集水面积阈值条件下河网密度参数的基础上, 通过均值变点分析法获取河网密度值变化的拐点, 并计算最佳集水面积阈值. 研究表明: 随着集水面积阈值的改变, 河网密度逐渐降低, 利用均值变点分析法所确定的最佳集水面积阈值为 8.1 km², 该结果与 1:50 万水系图相比, 主干上吻合, 支流上更具真实性. 该方法可以较好地确定最佳集水面积阈值, 可为其它地区水系提取提供参考.

[关键词] 杂谷脑河流域, 河网密度, 集水面积阈值, 均值变点分析法

[中图分类号] P542 **[文献标志码]** A **[文章编号]** 1001-4616(2014)01-0147-04

Determination of Accumulation Area Based on the Method of Applying Mean of Change-Point Analysis

Chang Zhiyang¹, Wang Jian^{1,2}, Bai Shibiao^{1,2}, Zhang Zhigang¹

(1. School of Geography Science, Nanjing Normal University, Nanjing 210023, China)

(2. MOE Key Laboratory of Virtual Geographic Environment, Nanjing Normal University, Nanjing 210023, China)

Abstract: To gain the threshold of catchment area is the critical step of extracting drainage network from DEM. But this index determination were still subjective and random with no theoretical support in most studies. Taking Zagunao Drainage Basin as an example, we calculated the drainage densities of different accumulation areas with SRTM-DEM data by using the hydrology model in Arcgis followed by gaining the optimum accumulation area with the method of mean of change-point. The result indicates that the drainage density correlated with accumulation area negatively and the 8.1 km² optimum accumulation shows more authenticity compared with the geology map of 1:500 000. It means that this kind of way can be applied to determine the accumulation area.

Key words: Zagunao River, channel density, threshold of accumulation area, mean of change-point method

利用河网求取的河流地貌参数在地貌学、水文学的研究中具有十分重要的意义^[1-3], 能否基于 DEM 数据提取出符合实际状况的河网则是提取这些地貌参数的前提, 其中最佳集水面积阈值的确定则是河网提取过程中的关键环节. 目前, 基于 DEM 确定最佳集水面积阈值的方法主要包括分形法、河网密度法、流域宽度法、二阶导数法等^[4-8], 虽然这些方法可以较准确地提取最佳河网阈值, 但是仍然存在一定的主观性及不便捷性. 例如, 孔凡哲^[4]提出利用河源密度与集水面积阈值的关系可以确定理想的集水面积阈值, 即当河源密度趋于稳定时的面积阈值为最佳阈值, 但这是通过目视判断河源密度趋于稳定的点, 与真实值相比, 仍存在一定的误差. 关颖慧等^[5]通过对河网密度及河源密度与集水面积阈值的幂函数二阶求导确定了集水面积阈值出现拐点的范围, 未能求出最佳值. 杨锦玲等^[6]虽然探讨了水系分维数与集水面积阈值的拟合关系, 但并未计算出最佳集水面积阈值.

收稿日期: 2013-09-16.

基金项目: 国家自然科学基金(40871010).

通讯联系人: 王建, 教授, 博士生导师, 研究方向: 自然地理学、地貌学、第四纪地质学. E-mail: jwang@vip.sina.com

基于目前对最佳集水面积阈值研究中存在着主观性以及不便捷性的问题,本文利用 GIS 软件及 SRTM-DEM 数据,以青藏高原东缘杂谷脑河流域为例,利用统计学上的均值变点分析法,快捷地求取了杂谷脑河的最佳集水面积阈值。

1 研究基础

1.1 DEM 数据

所使用的 SRTM-DEM 数据下载自中国科学院计算机网络信息中心国际科学数据服务平台(<http://datamirror.csdn.cn>),经过投影转换、拼接、裁切而成,格网分辨率为 30 m,绝对高程精度为 ± 16 m,绝对平面精度为 ± 20 m,经检查,研究区内并未存在数据空洞。

1.2 研究区概况

杂谷脑河流域位于青藏高原东缘($102^{\circ}35'E \sim 103^{\circ}39'E$, $31^{\circ}11'N \sim 31^{\circ}56'N$),面积约 4 600 km²,流域内部高程在 1 331 ~ 5 725 m 之间变化,平均海拔 3 615 m,属于典型的高山峡谷区(图 1)。干流杂谷脑河为岷江上游的一级支流,发源于鹧鸪山脉南麓,流经米亚罗、沙坝、理县、薛城等地,于汶川汇入岷江,河流溯源侵蚀,进入青藏高原内部,河道狭窄,水流湍急^[9]。

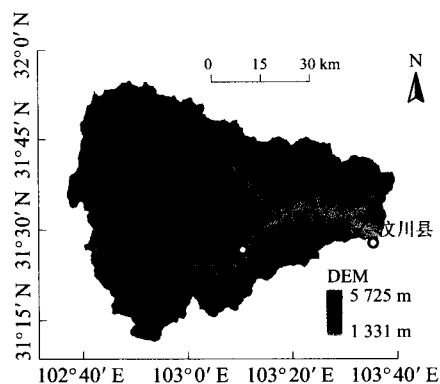


图 1 研究区位置

Fig. 1 The location of study area

2 研究方法

2.1 流域水系提取

目前提取流域水系的工具及方法有多种^[10],其中 ArcGIS 水文分析模块是常用的工具,主要步骤为:①洼地填平:DEM 表面存在着一些凹陷区,这些区域的存在在计算水流方向时会产生不合理甚至错误的结果,因此,首先需要对原始 DEM 数据进行洼地填充,得到无洼地 DEM。②水流方向确定:对于每一格网,水流方向是指水流离开格网时的流向,ArcGIS 中计算水流方向采用的 D8 算法。③汇流累积量生成:根据区域地形的水流方向数据计算每点处所流过的水量数值,便得到了该区域的汇流累积量。④水系提取:基于汇流累积量设置不同的集水面积阈值,经过栅格运算及栅矢转化生成最终的水系。

2.2 均值变点分析法

均值变点分析法是一种对非线性数据进行处理的数据统计方法,该方法对恰有一个变点的检验最为有效^[11],其中的变点是指:设 $\{x_i, i=1, \dots, N\}$ 为非线性系统的输出,其系统模型或输出序列在某未知时刻突然变化,该时刻即为变点,该方法在地势起伏度的求取中应用较多^[12-14]。变点的计算主要如下:对于有序数列 $\{x_i\}, i=1, \dots, N, N$ 为样本数,样本以 x_i 点为界分为两段,分别计算每段样本的算术平均值 $\bar{x}_{i_1}, \bar{x}_{i_2}$ 及样本均值 \bar{x} ,并计算统计量:

$$S_i = \sum_{t_1=1}^{i-1} (x_{t_1} - \bar{x}_{i_1})^2 + \sum_{t_2=i}^N (x_{t_2} - \bar{x}_{i_2})^2, \quad (1)$$

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

对于(1)式和(2)式中,下脚标 $t_1=1, 2, \dots, i-1; t_2=i, i+1, \dots, N$ 。变点的存在会使 S 和 S_i 的差距增大,两者最大差值对应的点称为变点。本文所用样本是指不同集水面积阈值下对应的平均河网密度。

3 结果与分析

3.1 集水面积阈值变化与河网的关系

本文采用的集水面积阈值为栅格的数目,试验设置的栅格数目分别为 300、400、500、600、……、3 500、3 750、4 000。以栅格数 300、1 000、1 500、2 000、4 000 的集水面积阈值为例,提取的河网如图 2 所示。并统计了不同集水面积阈值与河网密度的变化关系,绘制集水面积阈值与河网密度的关系图,如图 3 所示。

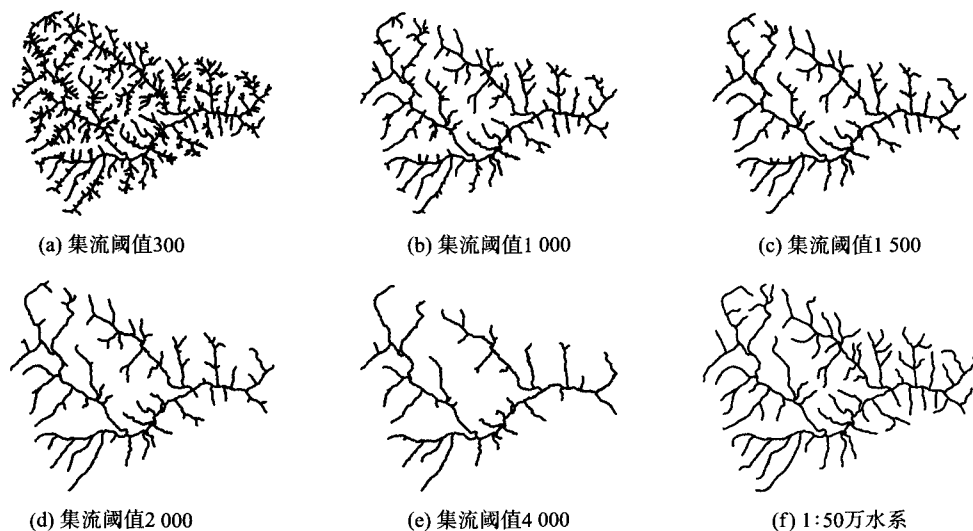


图2 不同集水面积阈值下提取的水系对比

Fig. 2 The comparison of drainage in different accumulation area

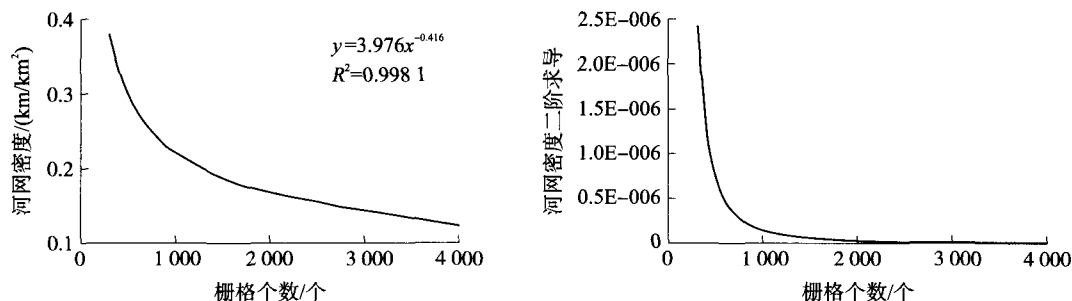


图3 集水面积阈值与河网密度的关系

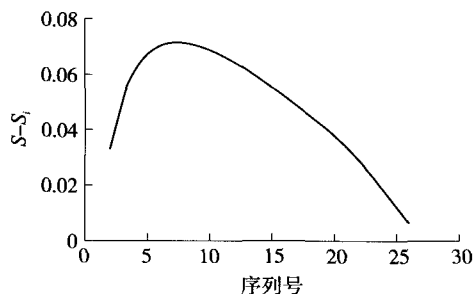
Fig. 3 The correlation between accumulation and channel density

从图2可以看出,随着集水面积阈值的由小到大变化,所提取的河网结构由复杂趋于稀疏,河网的分级程度也由多逐渐减少,并且河网密度逐渐变小。这主要是由于随着集水面积阈值的增大,汇流累积栅格中超过集水面积阈值的栅格变少,河道起点位置向地势平坦处“退缩”。另外,集水面积阈值与河网密度的函数拟合表明两者之间呈幂函数关系,拟合系数 R^2 为0.9981(图3)。在集水面积阈值小于1000的范围内,河网密度减小的幅度最大,而在大于1000之后,河网密度的减小趋于平缓,也就是说在该曲线上,存在一个平均河网密度变化由陡变缓的拐点^[5],该拐点所对应的集水面积阈值就是提取河网水系的最佳集水面积阈值^[4,5]。

3.2 均值变点分析法计算最佳集水面积阈值

根据公式(1)计算分段样本的算术平均值与方差,根据公式(2)计算样本的总体平均值和方差,可以得出 S 与 S_i 的差值变化曲线如图4所示。由图4可知, S 与 S_i 之差在序列号为7时最大,达到0.072,该序列点所对应的栅格数为1000,即提取杂谷脑河流域水系的最佳集水面积阈值为 8.1 km^2 。另外,利用河网密度与积水面积阈值函数的二阶导数作为临界阈值的对比判断^[5],通过二阶求导发现集水面积阈值在栅格数为1000左右时出现拐点(图3),这也印证了利用均值变点分析法计算最佳集水面积阈值的可靠性。

利用 8.1 km^2 作为集水面积阈值提取的河网与中华人民共和国1:50万数字地质图数据库中河网图进行对比(图2b, f),发现主干河流拟合最好,局部支流有所差异,可能是由于DEM数据自身的误差所致,或者是由于1:50万比例尺较小,

图4 S 和 S_i 差值变化曲线Fig. 4 S and S_i different value change curve

忽略了一些河网的存在. 另外, 前人对河网水系的提取研究多集中在中山、低山-丘陵地区^[4-8,10], 杂谷脑河流域位于高山峡谷地区, 均值变点分析方法在该地区确定最佳集水面积阈值中的应用, 也说明了利用 SRTM-DEM 数据提取高山峡谷地区的河网具有可信度, 可以降低人们对高山峡谷地区水系测量的难度.

4 结语

本文以高山峡谷地区杂谷脑河流域为例, 分析了在不同集水面积阈值下的水系形态结构及河网密度的变化规律, 可知集水面积阈值对主干河道空间位置影响不大, 但随着集水面积阈值的增加, 提取的水系变得稀疏, 河网密度逐渐降低, 且河网密度与集水面积阈值之间呈幂函数关系. 采用均值变点分析法的计算结果表明在集水面积阈值为 8.1 km^2 时, 提取的杂谷脑河流域水系接近真实值, 该方法的应用可以较为准确地确定河网的提取阈值, 基于 DEM 提取的河网与实际河网相比误差较小, 对于生产实践及水文模拟具有重要意义.

[参考文献]

- [1] Tejaswini N, Amba S, Hegde V S. Spatial variation in drainage characteristics and geomorphic instantaneous unit hydrograph (GIUH); implications for watershed management—A case study of the Varada River basin, Northern Karnataka[J]. Catena, 2011, 87: 52–59.
- [2] Rachna R. Active tectonics of NE Gujarat (India) by morphometric and morphostructural studies of Vatrak River basin[J]. Journal of Asian Earth Sciences, 2012, 50: 66–78.
- [3] 陈远川, 陈洪凯, 唐红梅. 沟口公路小流域山洪危险性评价——以四川省凉山州美姑河省道公路为例[J]. 中国地质灾害与防治学报, 2012, 23(4): 31–39.
- [4] 孔凡哲, 李莉莉. 利用 DEM 提取河网时集水面积阈值的确定[J]. 水电能源科学, 2005, 23(4): 65–68.
- [5] 关颖慧, 郑粉莉, 王彬. 基于 DEM 的黑龙江宾州河流域水系提取试验研究[J]. 水土保持通报, 2012, 32(1): 127–131.
- [6] 杨锦玲. 基于分形的数字水系集水面积阈值确定研究[J]. 测绘科学, 2011, 36(4): 659–662.
- [7] 梁启君, 梁军, 王少华, 等. 分形在数字水系累积汇水量阈值确定中的应用研究[J]. 地理与地理信息科学, 2013, 29(1): 111–112.
- [8] 杨邦, 任立良. 集水面积阈值确定方法的比较研究[J]. 水电能源科学, 2009, 27(5): 11–14.
- [9] 刘勇, 赵志军, 李才林, 等. 川西高原杂谷脑河阶地的形成[J]. 地理学报, 2006, 61(3): 249–254.
- [10] 曾红伟, 李丽娟, 柳玉梅, 等. Arc Hydro Tools 及多源 DEM 提取河网与精度分析——以洮儿河流域为例[J]. 地球信息科学学报, 2011, 13(1): 22–31.
- [11] 项静恬, 史久恩, 李笑吟. 非线性系统中数据处理的统计方法[M]. 北京: 科学出版社, 2000: 3–6.
- [12] 王玲, 同小娟. 基于变点分析的地形起伏度研究[J]. 地理与地理信息科学, 2007, 23(6): 113–116.
- [13] 韩海辉, 高婷, 易欢, 等. 基于变点分析法提取地势起伏度——以青藏高原为例[J]. 地理科学, 2012, 32(1): 101–104.
- [14] 张学儒, 官冬杰, 牟凤云, 等. 基于 ASTER GDEM 数据的青藏高原东部山区地形起伏度分析[J]. 地理与地理信息科学, 2012, 28(3): 11–14.

[责任编辑: 丁 蓉]