



2017年全国自然地理学大会，南京大学，2017年11月20-22日

基于领域知识和数据挖掘的 模糊坡位提取自动化方法

朱良君

资源与环境信息系统国家重点实验室

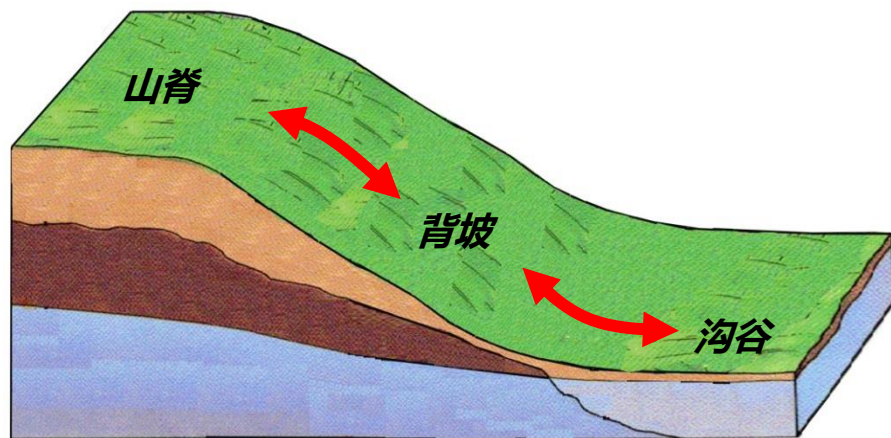
中国科学院地理科学与资源研究所

2017年11月22日

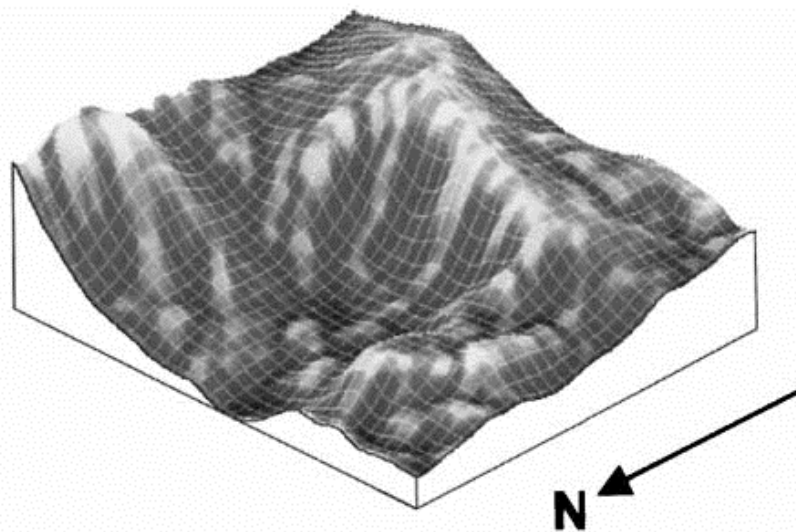
1. 研究背景和科学问题的提出
2. 基本思路
3. 方法设计与实现
4. 案例研究
5. 小结

1 研究背景与科学问题的提出

- 坡位的空间渐变信息可为地学或生态学建模提供重要的信息，如土壤-景观建模 (Deng, 2007; Qin等, 2012; Bui等, 2017)
- 空间任一位置对各坡位的模糊隶属度 (或相似度) 即为模糊坡位

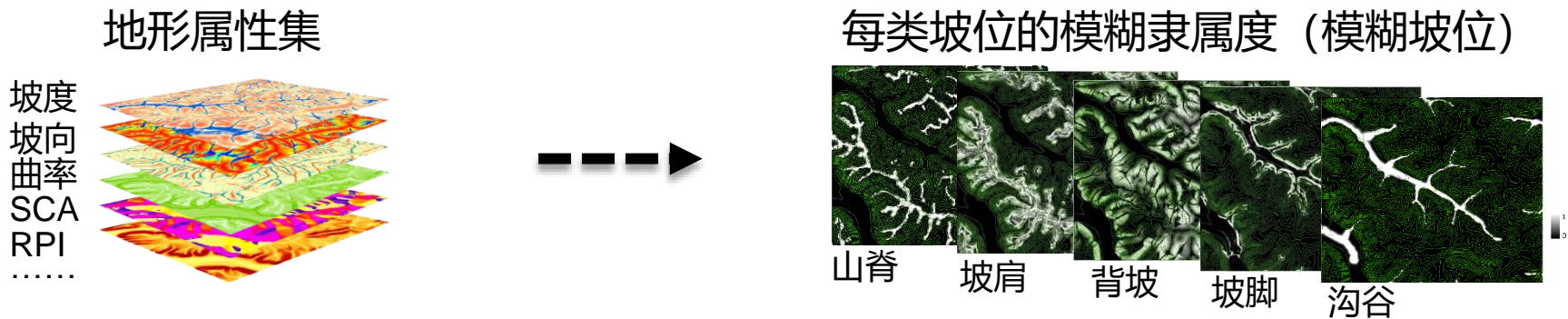


自然界中的坡位渐变



坡肩相似度 (MacMillan等, 2000)

现有的模糊坡位定量化方法

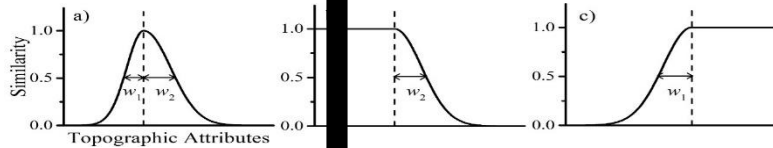
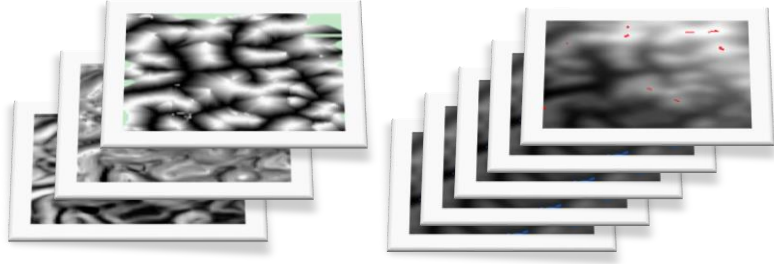


类别	基本思想	缺点
基于模糊聚类	在地形属性空间上进行模糊聚类(如 Irvin等, 1997; Burrough等, 2000)	<ol style="list-style-type: none">1. 忽视空间位置信息2. 结果难以用专家知识解释3. 难以提取低频出现的特殊地形部位
基于模糊分类	以诸多地形属性完善、准确地描述各类坡位的语义输入方法 (如MacMillan等, 2000; Schmidt和 Hewitt, 2004; Gercek等, 2011)	<ol style="list-style-type: none">1. 仅通过反映相对位关系的地形属性间接体现空间位置信息2. 难以形式化表达地形特征

基于原型的模糊推理方法 (Qin等, 2009)

地形属性集

坡位典型位置 (原型)



属性域: 对待推测性计算相对于某一并计算综合相似度

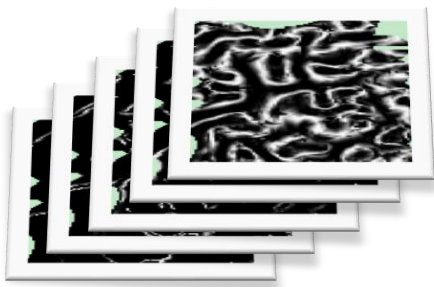
位置, 对每一地形属典型位置的相似度

$$S_{ij}^t = \min(S_{ij}^{1,t}, S_{ij}^{2,t}, \dots, S_{ij}^{n,t}, \dots, S_{ij}^{m,t})$$

空间域: 反距离加权

$$S_{ij} = \frac{\sum_{t=1}^m (d_{ij}^t)^{-8} S_{ij}^t}{\sum_{t=1}^m (d_{ij}^t)^{-8}}$$

模糊坡位信息



✓ 优点

- 充分利用坡位典型位置的地形特征
- 同时考虑属性域和空间域信息

✗ 易用性较差

- ✗ 流程繁琐: 准备地形属性集, 选取典型位置, 确定模糊推理参数
- ✗ 计算耗时: 串行版本实现耗时长

如何自动化实现基于原型的模糊坡位推理方法，
提高易用性



2 基本思路

结合领域知识和数据挖掘实现整个工作流的自动化

准备地形属性集

- 根据领域知识选取的地形属性(如Pennock等, 1987; MacMillan等, 2000; Schmidt和Hewitt, 2004)
- 在不同研究区的选择往往相似 (如Miller和Schaetzl, 2015; Qin等, 2009)
- 可用现有的数字地形分析算法

提取坡位典型位置

- 结合领域知识和数据挖掘确定每个地形属性的典型坡位值域, 通过叠加分析得到坡位典型位置

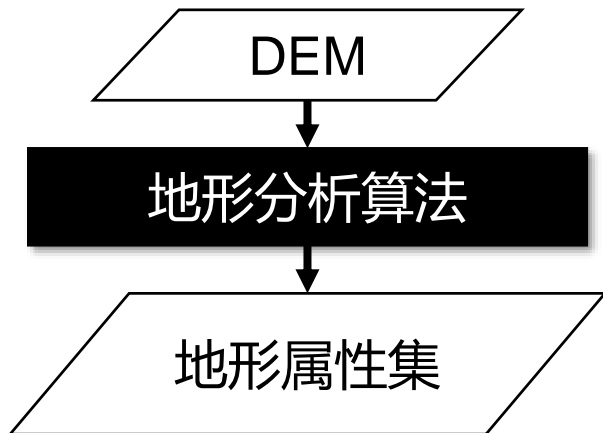
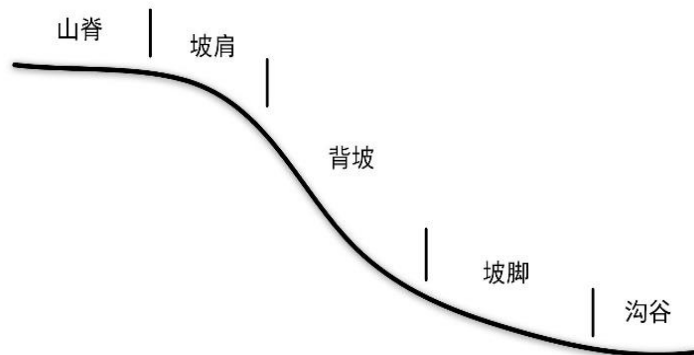
模糊推理

- 确定模糊推理函数参数并进行推理计算

3 方法设计与实现

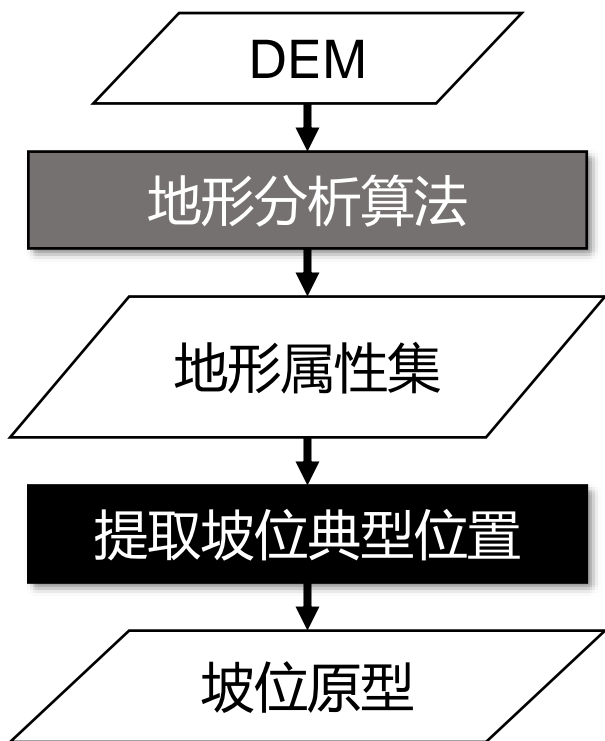
1) 自动准备地形属性集

- 以五类坡位系统为例 (Wysocki等, 2000; Qin等, 2009)



- 根据领域知识选取的地形属性集 (Qin等, 2009)
 - 区域属性: 相对位置 RPI (Skidmore, 1990)
 - 局域属性: 剖面曲率、坡度、高程
- 基于消息传递 (MPI) 的并行化 (Tesfa等, 2011; Qin等, 2014; Tarboton, 2014)

2) 自动提取坡位典型位置

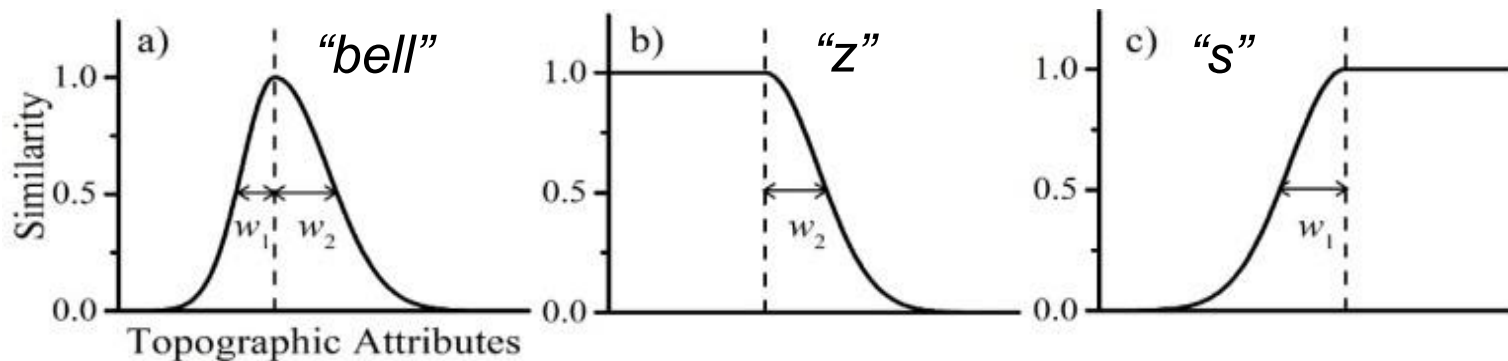


领域知识

坡位	RPI	剖面曲率	坡度	高程
山脊	S	S	Z	S, or N
坡肩	Bell	S	Bell	N
背坡	Bell	Bell	S	N
坡脚	Bell	Z, or Bell	Z, or Bell	N
沟谷	Z	B	Z	N

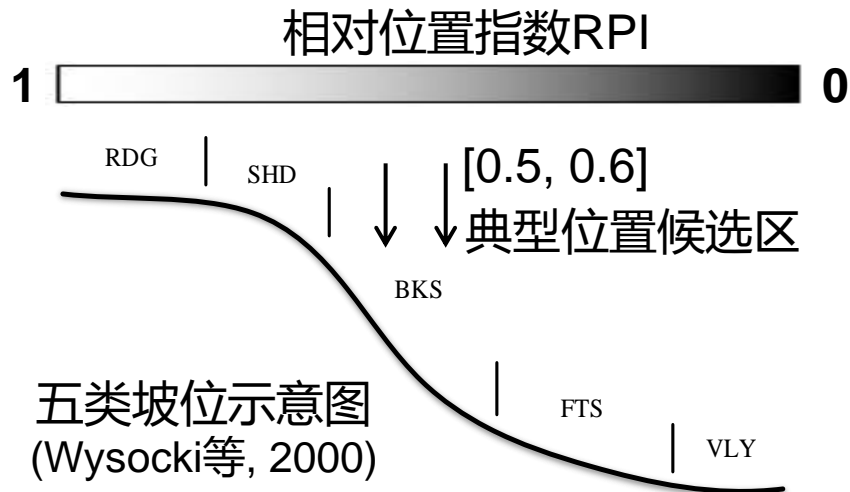
注：N为不采用该地形属性

(据Qin等, 2009, 2012)



2) 自动提取坡位典型位置

- (1) 利用区域地形属性粗略确定典型坡位的候选区域

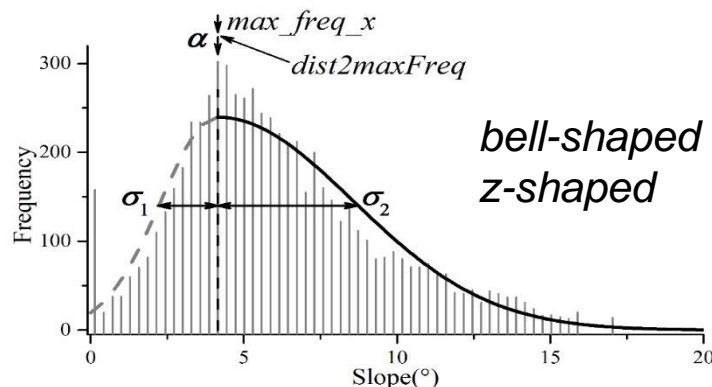


	山脊	坡肩	背坡	坡脚	沟谷
确定典型位置候选区的RPI范围	≥ 0.99	[0.9, 0.95]	[0.5, 0.6]	[0.15, 0.2]	≤ 0.1
RPI的模糊推理函数参数	$w_1 = 0.04$	$w_1 = w_2 = 0.04$	$w_1 = w_2 = 0.3$	$w_1 = w_2 = 0.05$	$w_2 = 0.05$

- (2) 利用bi-Gaussian模型(Yu和Peng, 2010)拟合候选区地形属性频率分布, 并估计模糊推理函数及参数区域

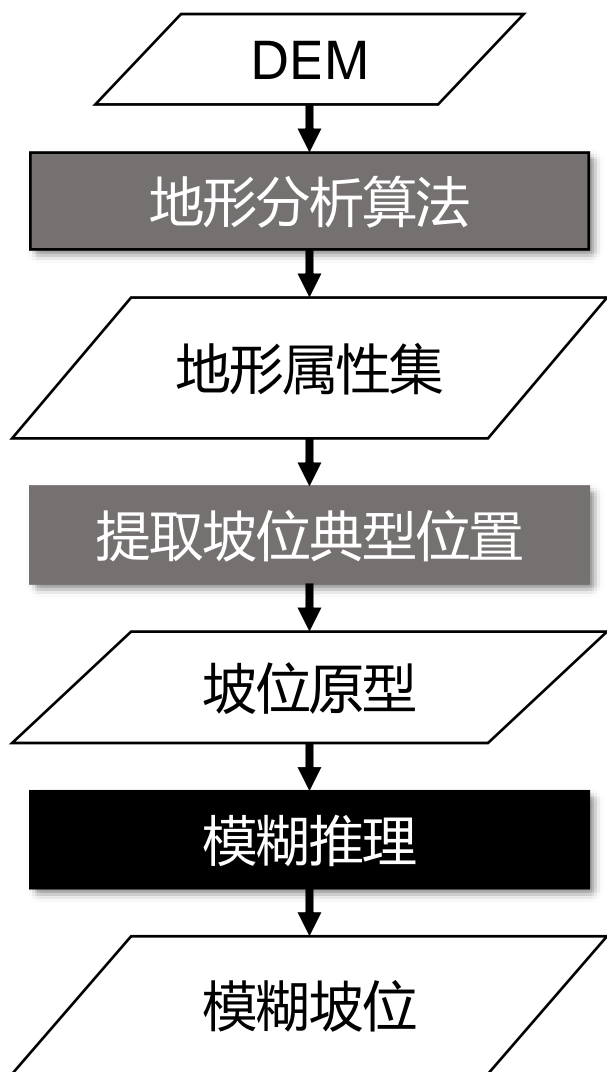
$$y = \frac{\delta}{\sqrt{2\pi}} e^{-\frac{(x-\alpha)^2}{2\sigma^2}}$$

if $x < \alpha$, $\sigma = \sigma_1$; else $\sigma = \sigma_2$



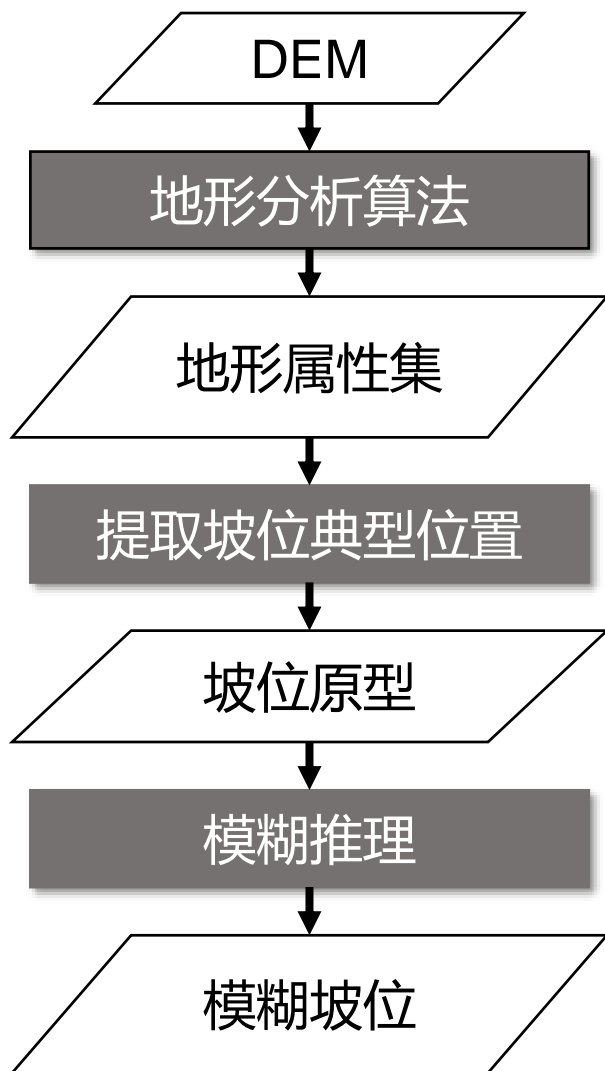
- (3) 兼顾典型位置代表性和计算效率, 通过叠加地形属性集按某比例范围 (如候选个数的10 ~ 30%) 提取坡位典型位置

3) 自动模糊推理



- 确定模糊推理参数并进行推理计算

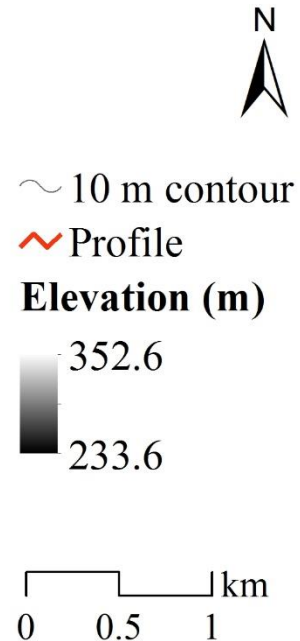
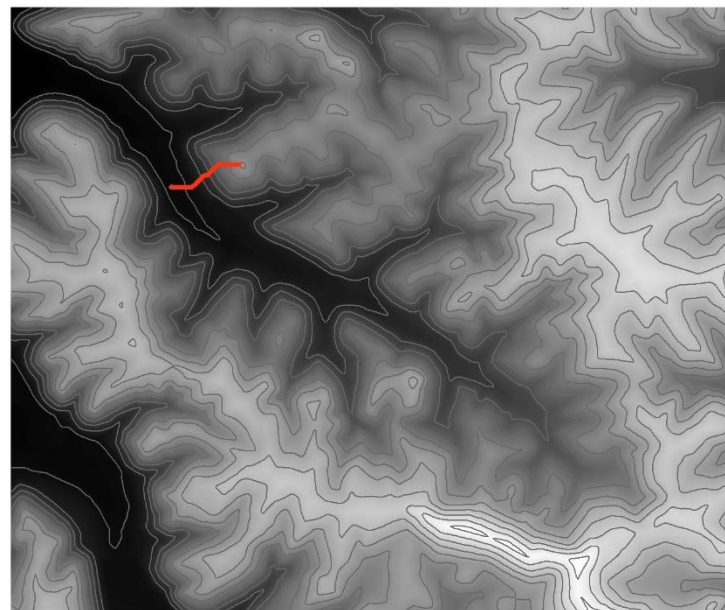
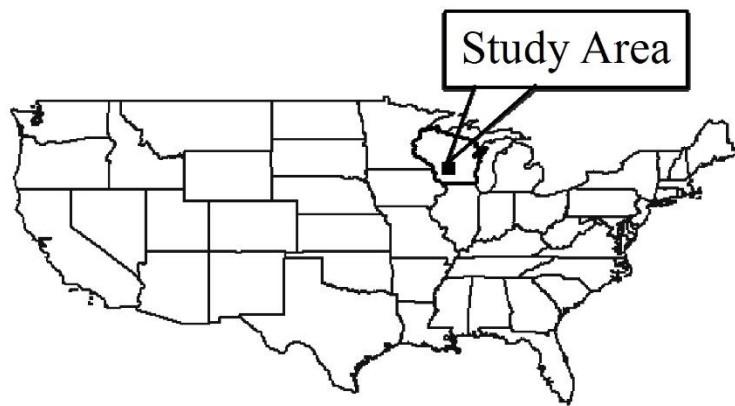
自动化 workflows 的实现



- 实现了模糊坡位提取自动化 workflow
- 默认仅需一个输入数据，即研究区 DEM
- 允许用户自定义配置，如地形属性集等
- 基于消息传递 (MPI) 实现并行化

4 案例研究

- 美国威斯康星州Pleasant Valley小流域，面积约 12.7 km²
- 栅格分辨率 9.14 m
- 最大坡度为 35.5°，平均 9.7°



美国威斯康星州西南部Pleasant Valley位置示意图

实验设计

测试环境

- Linux集群
 - 1个管理节点, 4个计算节点
 - GCC 4.8.4, MPICH 3.1.4, 及Python 2.6.6
- 采用默认参数配置运行程序

结果评价

- 提取模糊坡位信息合理性
 - 坡位典型位置提取参数和模糊推理参数
 - 模糊坡位信息的空间分布
 - 与Qin等 (2009)结果的对比
- 不同进程下 (1 ~ 32) 的计算效率

结果分析

1) 模糊坡位信息提取参数

坡位典型位置提取参数 (自动化方法)

	RPI	剖面曲率($\times 10^{-3} \text{ m}^{-1}$)	坡度 ($^{\circ}$)	高程(m)	个数
山脊	≥ 0.99	≥ 3.36	≤ 6.92	-	868
坡肩	[0.9, 0.95]	≥ 2.7	[4.89, 10.42]	-	1875
背坡	[0.5, 0.6]	[-0.95, 1.81]	≥ 11.63	-	4473
坡脚	[0.15 0.2]	[-2.04, 1.03]	[3.24, 10.3]	-	2581
沟谷	≤ 0.1	[-3.55, 0.4]	≤ 3.26	-	4841

坡位典型位置提取参数 (Qin等, 2009)

	RPI	剖面曲率($\times 10^{-3} \text{ m}^{-1}$)	坡度 ($^{\circ}$)	高程(m)	个数
山脊	≥ 0.99	≥ 0	≤ 1	≥ 285	1305
坡肩	[0.9, 0.95]	≥ 5	-	-	2337
背坡	[0.5, 0.6]	[-0.1, 0.1]	≥ 10	-	1281
坡脚	[0.15 0.2]	≤ -5	-	-	122
沟谷	≤ 0.1	[-0.1, 0.1]	≤ 1	-	1578

1) 模糊坡位信息提取参数

模糊推理参数 (自动化方法)

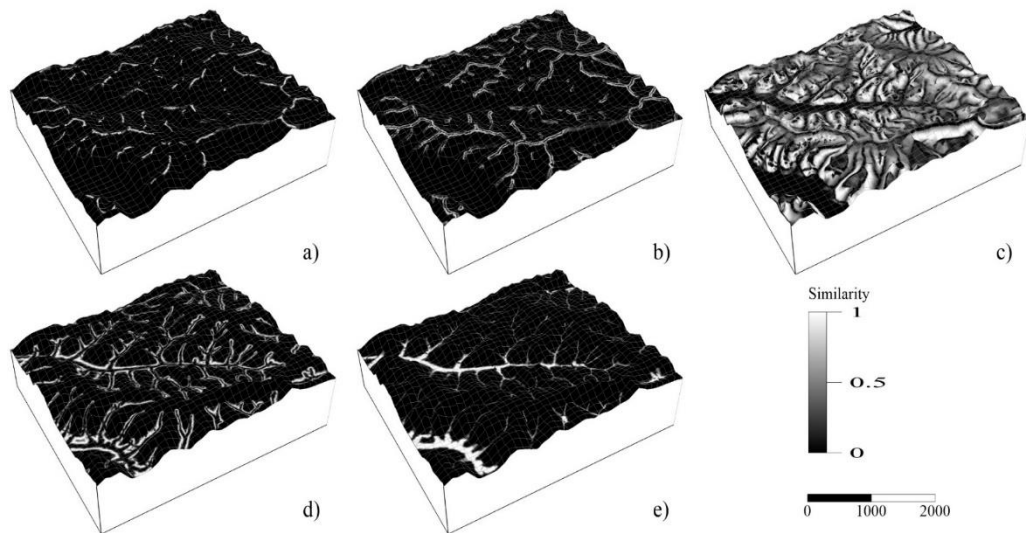
	RPI	剖面曲率($\times 10^{-3} \text{ m}^{-1}$)	坡度 ($^{\circ}$)	高程(m)
山脊	S: $w_1 = 0.05$	S: $w_1 = 7.48$	Z: $w_2 = 5.29$	-
坡肩	B: $w_1 = w_2 = 0.04$	S: $w_1 = 4.87$	B: $w_1 = 3.24, w_2 = 6.43$	-
背坡	B: $w_1 = w_2 = 0.3$	B: $w_1 = 2.71, w_2 = 2.21$	S: $w_1 = 7.23$	-
坡脚	B: $w_1 = w_2 = 0.05$	B: $w_1 = 3.2, w_2 = 2.01$	B: $w_1 = 4.45, w_2 = 5.62$	-
沟谷	Z: $w_2 = 0.1$	B: $w_1 = 5.31, w_2 = 1.77$	Z: $w_2 = 4.88$	-

模糊推理参数 (Qin等, 2009)

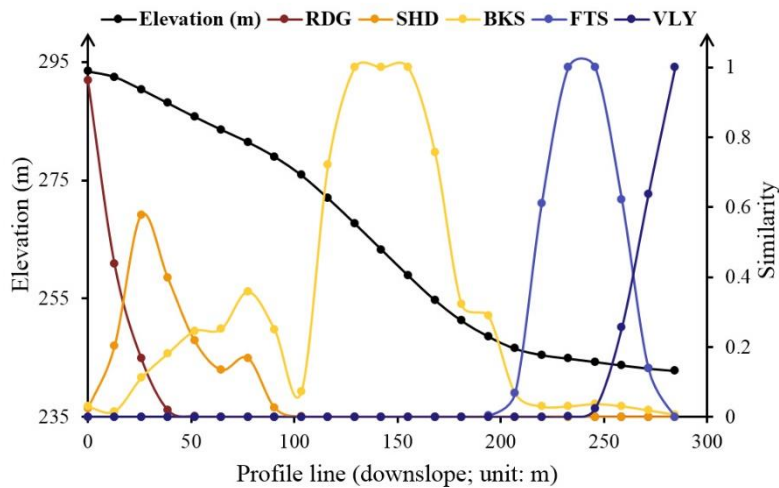
	RPI	剖面曲率($\times 10^{-3} \text{ m}^{-1}$)	坡度 ($^{\circ}$)	高程(m)
山脊	S: $w_1 = 0.1$	S: $w_1 = 5$	Z: $w_2 = 5$	S: $w_1 = 5$
坡肩	B: $w_1 = w_2 = 0.05$	S: $w_1 = 5$	B: $w_1 = w_2 = 5$	-
背坡	B: $w_1 = w_2 = 0.3$	B: $w_1 = w_2 = 5$	S: $w_1 = 5$	-
坡脚	B: $w_1 = w_2 = 0.05$	Z: $w_2 = 5$	B: $w_1 = w_2 = 5$	-
沟谷	Z: $w_2 = 0.1$	B: $w_1 = w_2 = 5$	Z: $w_2 = 5$	-

2) 模糊坡位信息的空间分布

自动化方法



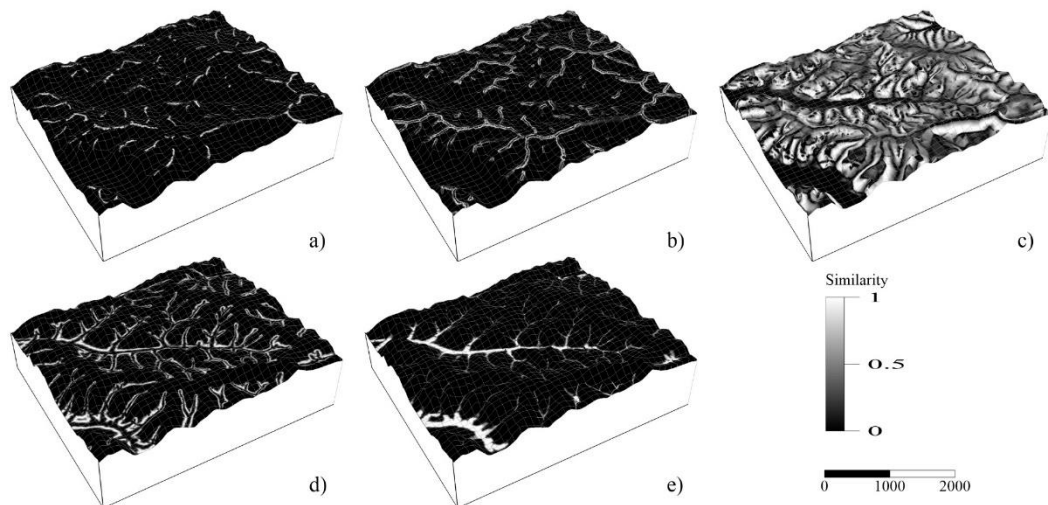
(a)山脊; (b)坡肩; (c)背坡; (d)坡脚; (e)沟谷



某下坡方向剖面线模糊坡位变化曲线

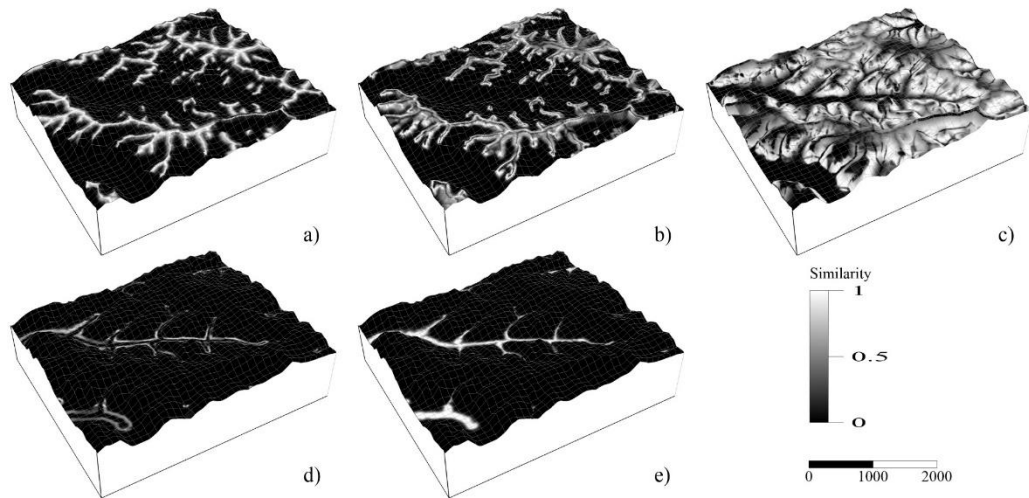
- 模糊坡位信息能够定量刻画从坡顶到坡底的坡位渐变

2) 模糊坡位信息的空间分布



自动化方法

(a)山脊; (b)坡肩; (c)背坡; (d)坡脚; (e)沟谷



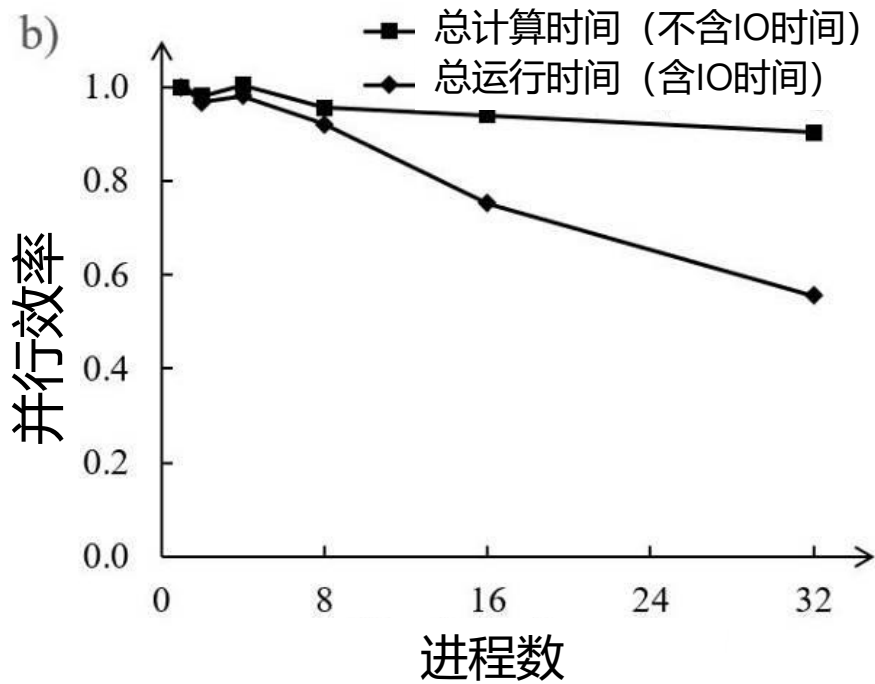
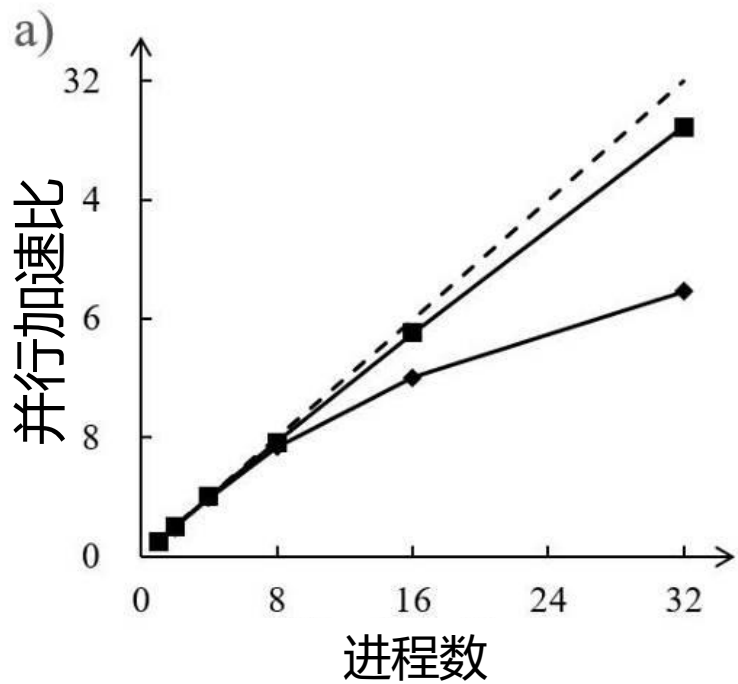
Qin等 (2009)

自动化方法与Qin等 (2009) 对比

- 两者结果整体一致
- 自动化方法具有更详细的空间格局
- 空间格局受区域地形属性 (如 RPI) 影响
- 这种差异也反映出地形元素信息的多尺度特征 (Wang et al., 2010; Drăguț et al., 2011; Gorini and Mota, 2016)

通过综合不同地形骨架尺度 (由山脊和山谷确定) 下的模糊坡位获得**多尺度综合的模糊坡位**

3) 计算效率



5 小结

- 提出了一种基于原型的模糊坡位提取自动化方法
 - 默认配置下仅需一个输入数据，即研究区的栅格DEM
 - 允许用户自定义配置，如地形属性集
 - 并行计算
- 本研究结合领域知识和知识挖掘的基本思想可用于其他类似地学空间分析算法的自动化研究



中国科学院
CHINESE ACADEMY OF SCIENCES

谢谢各位，敬请指正

朱良君, 朱阿兴, 秦承志, 刘军志. 基于领域知识和数据挖掘的模糊坡位提取自动化方法. 2017年全国自然地理学大会, 南京, 2017年11月20-22日.

E-mail : zlj@lreis.ac.cn

Website: zhulj.net

