第 54 卷 第 2 期
2017 年 3 月

土 壤 学 报
ACTA PEDOLOGICA SINICA

Vol. 54，No. 2
Mar.，2017

DOI：10.11766/trxb201608150111

基于地形因子的土壤有机碳最优估算模型*

郭治兴[1] 袁宇志[1] 郭 颖[1, 2] 孙 慧[1, 2] 柴 敏[1]
陈泽鹏[3] Mogens H. Greve[4]

（1 广东省生态环境技术研究所广东省农业环境综合治理重点实验室，广州 510650）
（2 山西农业大学资源环境学院，山西晋中 030800）
（3 广东省烟草公司，广州 510610）
（4 Department of Agro-ecology，Faculty of Science and Technology，Aarhus University，8830 Tjele，Denmark）

摘 要 基于数字地面模型（Digital Terrain Model，DTM），同时考虑因子组合和分辨率构建土壤有机碳（SOC）最优估算模型。在7 100 km^2范围内，选取了71个分辨率和22个地形因子中不多于5个因子的所有可能组合，构造了2 514 820个模型。采样点随机分为两组，6 362个训练样点构造数据挖掘模型，其他2 208个为验证样点。根据模型相关系数r值大小从中选取了不同个数因子组合以及相应分辨率的最优模型，并根据这些模型生成对应的土壤有机碳图。结果表明：单个地形因子模型和栅格大小之间的关系表现出多样化，并不是分辨率越高模型结果越好。单因子模型r值的大小并不能决定其在因子组合模型中的重要性。不同的因子及其组合有其特定的最适分辨率，最佳分辨率范围约为60 ~ 150 m。综合数据的存储空间和计算量、模型复杂度、预测精度以及空间表达能力，该地区最优模型由相对坡位、高程、归一化高程及多尺度山谷平坦指数等4个变量组成，对应分辨率为121.6 m。同时与多种克里格空间插值方法生成的土壤有机碳空间分布图进行了对比分析，发现无论几个变量的组合，其空间预测能力均较克里格空间插值方法更能表达SOC的空间变化，预测精度也较高。

关键词 土壤有机碳；数字土壤制图；数据挖掘；最优估算模型；数字地面模型；地形参数
中图分类号 S158 **文献标识码** A

土壤有机碳（SOC）是全球碳库的重要组成部分，是陆地生态系统中最大的有机碳库[1]，在全球碳循环和气候变化过程中起着极其重要的作用[2-3]。同时也是人类可持续利用土地资源进行生产，保证全球粮食安全必须要掌握的基础信息[4]。

土壤有机碳储量受自然和人为两个方面因素的影响[5]。地球环境的进化决定了一个特定地区的气候、植物、地形、地貌以及土壤等自然属性，这些属性之间必定有较高的相关性。一个地区的土壤有机碳和其所处的地质、地貌以及地形等因子是自然形成和演化的共同结果[6]。在此过程中，特定的地质、地貌以及地形等因子均会影响土壤有机碳的形成与演化。因此地形因子的组合与土壤有机碳必然有其内在的联系，能够在一定程度上反映所在地土壤有机碳的状态[7-8]。

土壤有机碳还受人类活动的影响[9-10]。人类利用土壤的方式除受气候、传统等因素影响[11]，特定的地形因子也会直接影响人类利用土地的方式，例如坡度、坡向等。因此，从人类活动的角度看，地形因子与土壤有机碳储量之间也有其必然的

* 广东省科技计划项目（2015B070701017，2014A040401059，2015A030401068）、国家自然科学基金项目（41601558）、广东省科学院创新平台建设专项资助 Supported by the Project of the Science and Technology of Guangdong Province（Nos. 2015B070701017，2014A040401059，2015A030401068），National Natural Science Fund（No. 41601558）and SPICC Program（The Scientific Platform and Innovation Capability Construction Program of GDAS）
作者简介：郭治兴（1966—），男，山西太谷人，博士，研究员，主要从事GIS应用研究。E-mail：zxguo@soil.cn
收稿日期：2016-08-15；收到修改稿日期：2016-11-09；优先数字出版日期（www.cnki.net）：2017-01-03

联系。

土壤有机碳的计算模型较多，大体分为经验模型[12-13]、统计模型[14-15]和机理模型[16-17]。基本均是间接的方法，目前还没有一个方法可以快速测量或估算一个区域内的土壤有机碳储量。采用地形因子等相对较易获取的参数建模就是一个经常被采用的方法。在数字土壤制图领域，土壤属性的预测和计算，包括土壤有机碳的估算，地形因子经常会被纳入模型[18-20]。遥感技术为快速获取数字地面模型（DTM），提供了一个高效的技术手段[21]。

一个能够精确预测土壤属性的模型，还依赖于找到一个最适栅格大小和由其推导出的地形因子[22-23]。众所周知遥感数据是栅格形式的，而地形因子的取值与栅格的大小息息相关，模型的结果也会随之变化[24-27]。如何确定这些因子的最适栅格大小的问题仍然未解决，只有一些针对特定区域的研究和原则性的指南[28]。

通常DTM数据是最容易获取的数据源，常规的做法是直接采用原始分辨率的数据，一般认为这样会提高模型的准确度和精度[29]。但是不恰当的格栅大小通常会带来一些问题[30]。如果栅格过大，地形因子的属性会丢失太多的细节，进而失去预测能力。反之，如果栅格过小，由此导出的地形因子会带有太多的细节，甚至是“噪声”，这样势必会影响最终模型的精度[31]。同时，高分辨率数据由于拥有较高的空间精度，栅格非常小，数据量非常大，需要更大的存储空间；模型会需要更多的计算时间，有的时候这种增长是指数级别的。一个合适的分辨率，应该是栅格尺寸足够大，但又可以充分表达模型所需要的信息[22]。

预测土壤属性采用哪些地形因子也是一个需要考虑的问题[32]。由DTM可以导出很多地形因子，一个模型采用哪些因子，通常均具有较大的随意性和主观性[33-34]。同一个地形因子对不同的土壤属性有着不同的影响。反之，不同土壤属性有属于其特有的决定性的地形因子[35]。地形因子之间存在复杂的相互关系，单个因子在土壤模型中的重要性在多个地形因子组合模型中也会发生变化[36-37]。

利用DTM及相关因子预测SOC前人做出了大量工作，但大都基于一个固定的栅格大小或人为选取的因子或组合[38-39]。本研究试图通过DTM及其推导的相关地形因子，综合栅格大小、模型精度以及数据的冗余度等因素，通过数据挖掘技术，寻找最优因子组合和相匹配的最佳分辨率的土壤有机碳预测模型。

1 数据与方法

1.1 研究区概况

研究区位于56°9′~56°33′N及8°6′~10°57′E之间。是一条从西到东贯穿整个丹麦日德兰半岛（Jutland）的宽约45 km的条状带，总面积约7 100 km^2（图1）。整个区域地势平缓，平均海拔为34 m，最高峰仅为152 m。该区域冬季和夏季平均温度分别为0° C和16° C，年平均降雨量为500~800 mm。主要种植冬小麦、春大麦、油菜、牧草以及玉米。

由于历史上冰川侵蚀活动，造成了研究区内复杂多样的地貌类型[40]。据丹麦土壤分类，研究区范围内的土壤质地从东部的重黏土和黏土到壤质土到西部的粗砂质土。东部地区的土壤由魏克塞冰期冰碛地貌发育而成，以壤质土为主。而西部主要由

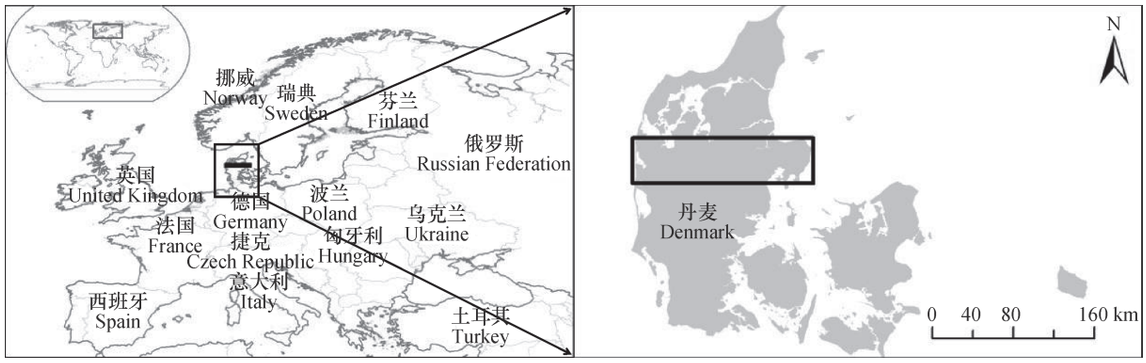


图1 研究区位置

Fig. 1 Location of the study area

严重侵蚀和历史相对较长的撒利冰碛地貌，以及一个相对较大的冰水平原组成，以砂质土为主。

该区域包含有丹麦全部8种土壤类型，全部12种景观类型中10种，全部12种地质类型中的11种，具有典型的代表性。

1.2　土壤采样点和数据库

研究中用到的土壤样点主要来源于全丹麦土壤分类数据库[41]。该数据库建成于20世纪70年代中期，当时建库的主要目的是为了土壤制图和保护城市化进程中日益受到威胁的土地资源，由于该项目要求三年内完成，因此其目标主要集中在采样和分析几个重要和易分析的指标，如土壤质地和耕层（0～20 cm）土壤有机质。全国共采集36 000余个土壤样品。在研究区域内共计有8 570个样点数据，其中约75%，计6 362个样点用来生成和优选模型，其余25%，计2 208个的样点用来验证模型。

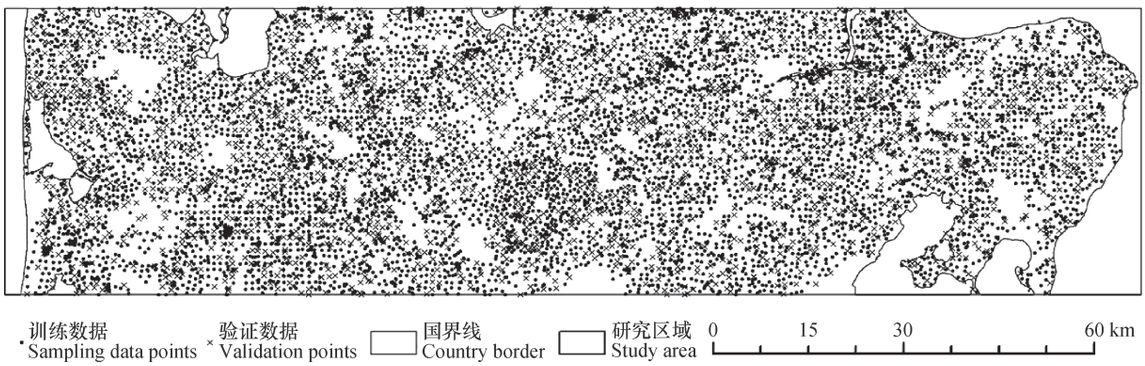


图2　采样点分布图

Fig. 2　Distribution map of sampling sites

1.3　DTM数据

丹麦全国的DTM数据由国家环境部基于光探测与测量（Light Detection and Ranging，LiDAR）数据在2005—2007年间生成。采用不规则三角网方法从LiDAR点云数据中提取全国的地表高程，然后将TIN数据转换为1.6m分辨率的栅格数据。

采用简单求平均值的方法生成其他较低分辨率DTM栅格数据集，共采样计71个，小于150 m的设计了41组，200～1 000 m之间每50 m取一组，1 000 m以上大约每100 m取一组，最小分组的分辨率为2 304 m。基于DTM计算其他地形参数之前，先对每个不同分辨率的栅格DTM数据进行填洼处理，以减少或避免在进行地形运算，特别是推导与水系网有关的地形参数时，由于细小洼和峰导致的错误或出现问题[42]。

1.4　导出变量

基于不同分辨率的DTM数据集，通过SAGA-GIS计算其他地形相关因子。共导出21个地形因子参数，每个因子的说明见表1。

1.5　基于规则的数据挖掘模型

本研究使用澳大利亚Rulequest公司的数据挖掘工具Cubist 2.08生成SOC预测模型，这种模型利用线性的方法模拟非线性的现实。Cubist设计用来处理海量数据，它的数据挖掘模型不仅注重模型的预测精度，同时也重视模型的可解释性。其生成的模型以一组规则集表达，每个规则采用如果满足条件则使用一个线性方程的模式。当自变量满足某一规则给定的条件时，它就使用该规则对应的多变量线性模型进行预测[43]。

Cubist模型利用相对坡位（RSP）、多尺度山谷指数（MRVBF）、高程（Elevation）和归一化高度（Normalh）四个变量预测SOC含量。该模型由29个规则组成，每个规则给出一个条件和一个对应的多元线性模型。例如，第一个规则预测1758个样点的SOC，这些样点的取值范围为0.08～540.1，平均值和估计误差分别为20.315和7.662。该规则列出条件当MRVBF＞8.810 939，并且RSP＞0.227 911时，其对应的用来预测SOC的多变量线性回归方程。从这条规则可以看出，这些样点的SOC随着RSP的增加而减少呈现负相关，随着MRVBF、Elevation和Normalh的增加而增加。该模型的总体相对误差和模型相关系数分别为0.54和

表1 模型地形因子列表

Table 1 List of topographic factors derived by DTM

变量名 Variable name	说明 Brief descriptions	单位 Unit
高程 Elevation	雷达数据生成的地面高程 Elevation of the land surface derived from LiDAR	m
坡向 Aspect	坡向 Direction aslope faces	°
水系网高程 Chnl_alti	相对水系网高程 Base level in elevation of channel network	m
水系网基准面 Chnl_base	水系网基准面 Channel network base level elevations interpolated by channel network	m
汇指数 Convergence	汇指数 Convergence index，the index of convergence/divergence regarding to overland flow	%
地表曲率 Convexity	地表非局地曲率 Terrain surface convexity	–
平面曲率 Hcurv	平面曲率 Plan curvature，curvature of contour drawn through the grid point	m^{-1}
坡长 Ls	USLE计算的坡长 Slope length（LS），factor to calculate slope length as used by USLE	m
中坡位 Midslppst	中坡位 Mid-slope position：a classification of slope position in both valley and crest directions	–
多尺度山脊指数 MRRTF	多尺度山脊平坦指数 Multiresolution index of the ridge top flatness	–
多尺度山谷指数 MRVBF	多尺度山谷平坦指数 Multiresolution index of valley bottom flatness	–
归一化高度Normalh	归一化高度 Normalized height，1 for the highest and 0 for the lowest position within a respective reference area	%
相对坡位 RSP	相对坡位 Relative slope position	%
粗超度 Ruggedness	地形粗糙指数，指中心点高程与周围高程差的平均值 Terrain ruggedness index	–
洼地 Sink	闭合洼地 Closed depressions	
坡度 Slope	坡度 Maximum rate of change between cells and neighbors	°
坡高度 Slph	坡高度：与相邻山脊线的相对高度差 Slope height：The relative height difference to the immediate adjacent crest lines	m
标准化高度Standh	标准化高度：归一化高度与相对高度的乘积 Standardized height is the product of normalized height multiplied with absolute height	–
地表纹理 Texture	地表纹理 Terrain surface texture	–
地形湿度指数 TWI	地形湿度指数 Slope and specific catchment area based topographic wetness index	–
山谷深度Vall_depth	山谷深度：与紧邻水系网的相对高度 Valley depth：Relative height difference between the immediate adjacent channel lines	m
剖面曲率 Vcurv	剖面曲率 Profile curvature，curvature of the surface in the direction of steepest descendent	m^{-1}

0.71。同时，模型还给出了变量在模型中贡献率。贡献率是指相对重要性和相对使用率两个参数，相对重要性指变量在模型全部规则所有条件中出现的频率，相对使用率指变量在多元线性方程中参与计算的频率。在这个模型中RSP的相对重要性和相对使用率分别为89%和81%，较Normalh的9%和52%要高很多。

所有模型均是简单性和精度之间的平衡。22个变量的所有可能的组合数高达4 194 303个，变量过多会引入太多的噪声，增加模型的复杂程度。综合模型复杂度和数据的计算量及存储空间，选取不超过5个变量的组合是比较合适的。这种组合共计有35 442个，每个组合有71个分辨率，共计生成2 514 820个模型并进行计算。

以上所有组合通过Cubist运算获得模型结果，每个模型均有对应的相关系数*r*值。对比所有模型*r*值，最大*r*值所在的因子组合与分辨率即为该因子组合和最佳分辨率。

2　结　果

2.1　数据基本统计特征

整体样本的平均值和标准差分别为36.0 g kg^{-1}和69.7 g kg^{-1}。最小值、中位数和最大值分别为0.08 g kg^{-1}、15.58 g kg^{-1}和540.1 g kg^{-1}。变异系数高达193.6%，说明数据整体变化幅度较大，整体为带偏度的正态分布。尽管两组数据的很多统计量有所不同，但通过柯尔莫诺夫—斯米尔诺夫检验（Kolmogorov-Smirnov test），训练数据组和验证数据组之间并不存在显著性差异（99.9%）[44]，说明两组数据来自同一母体，验证数据可以用来对模型进行验证。

2.2　地形因子与分辨率对模型的影响

表2中列出了单因子在所有分辨率中的模型的*r*平均值、最大值以及最大值对应的分辨率。单个地形变量模型和栅格大小之间的关系表现出多样

表2　单因子模型*r*值及最佳分辨率

Table 2　R and optimal resolution of single factor models

变量 Variables		*r*平均值 Average *r*	*r*最大值 Max.*r*	*r*最大值时对应分辨率（m） Resolutions where the *r* is maximum
相对坡位	RSP	0.272	0.5	92.8
水系网高程	Chnl_alti	0.246	0.48	68.8、88、92.8、97.6、116.8
山谷深度	Vall_depth	0.245	0.38	19.2、64、92.8、97.6、380.8
标准化高度	Standh	0.236	0.43	12.8
汇指数	Convergence	0.228	0.3	384、459.2、604.8、644.8、651.2
高程	Elevation	0.208	0.32	112、182.4、185.6、230.4
归一化高度	Normalh	0.093	0.44	12.8、16
粗超度	Ruggedness	0.079	0.34	20.8
中坡位	Midslppst	0.066	0.4	112、182.4、185.6、230.4
坡度	Slope	0.049	0.27	17.6、25.6
地表纹理	Texture	0.042	0.28	12.8
地表曲率	Convexity	0.038	0.38	44.8、78.4
坡高度	Slph	0.033	0.26	12.8
多尺度山谷指数	MRVBF	0.023	0.28	1 460.8
水系网基准面	Chnl_base	0.022	0.26	25.6
剖面曲率	Vcurv	0.019	0.28	268.8
坡长	Ls	0.018	0.22	1 460.8
地形湿度指数	TWI	0.014	0.3	12.8
平面曲率	Hcurv	0.011	0.28	19.2
洼地	Sink	0.01	0.32	1 670.4、1 737.6
多尺度山脊指数	MRRTF	0.01	0.23	2 188.8
坡向	Aspect	0	0	–

化，并不总是负相关。大部分变量与格栅大小呈偏态的正态分布，如相对坡位（RSP）和水系网高程（Chnl_alti）。个别变量呈正相关，如汇指数（Convergence），随着栅格的增大，模型的r值增大。也有部分变量与栅格的大小无关，呈随机状态，如水系网基准面（Chnl_base），还有部分变量与栅格没有任何关系，其模型的r值在任何分辨率下均为0，如坡向。最好的地形因子为相对坡位（RSP），r平均值为0.272，最大值达到0.5。最差的为坡向，无论是何种分辨率，坡向的r值均为0，即坡向与SOC没有任何关系。

对单个地形因子模型，无论其与栅格大小的关系是正相关还是负相关，r最大值时的分辨率通常并不是最大或最小栅格。不同的变量有一个或多个最佳的栅格大小，例如，相对坡位（RSP）的最佳分辨为92.8m，归一化高度（Normalh）则有2个分辨率，分别为12.8和16m。

任意多个因子的组合与单个因子的表现不同，多个因子的组合表现出了一致的规律，均为呈偏态的正态分布。从最高分辨率开始随着栅格的增大，模型的r值增加至一个最大值，此处的分辨率范围大约为 60～150 m。此后r与分辨率呈负相关，在1 500m附近r值基本达到稳定，之后随着分辨率的改变，r值呈随机变化状态。因此无论使用哪些因子组合建模，分辨率均应在60～150m之间选取。

单因子模型r值的大小并不能决定其在因子组合模型中的重要性。即2个r值最大的单因子模型，它们组合后的模型的r值并不是最大的。单个因子r值最大的6个因子分别为相对坡位（RSP）、水系网高程（Chnl_alti）、山谷深度（Vall_depth）、归一化高度（Standh）、地表曲率（Convexity）和高程（Elevation）（表2）。但是在筛选出的最佳因子组合模型中出现次数最多的6个因子是：水系网基准面（Chnl_base）、高程（Elevation）、水系网高程（Chnl_alti）、多尺度山脊指数（MRRTF）、归一化高度（Normalh）和相对坡位（RSP）。事实上在最佳模型中出现的12个因子，并不是最好的12个单因子。最明显的例子是多尺度山脊指数（MRRTF）因子，该因子与分辨率关系呈随机分布状态，r值总体上也偏低，但是在因子组合模型中该因子共出现了4次。

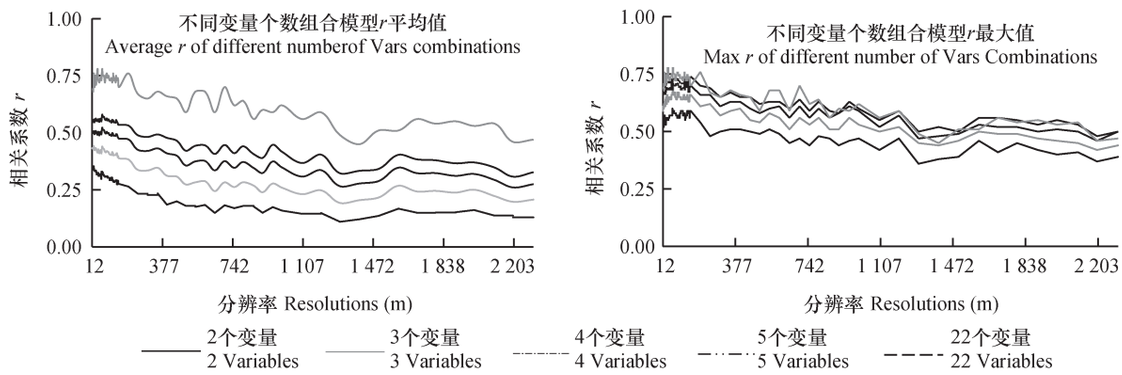


图3 变量组合模型相关系数r值与分辨率的关系

Fig. 3 Relationship between r of combination of variables with resolution

无论何种分辨率，因子的多少决定了r的总体趋势，引入的因子越多，r值越大。当引用全部22个因子时所构建的模型，r值达到最大0.78（图3），但是模型的预测能力并非如此。

2.3 SOC最优估算模型

通过对比r值，筛选出不同数量因子组合的最优模型和对应的分辨率。不同个数因子的最优模型列于表3，这些模型的r值变化见图4。单个因子的模型为相对坡位（RSP）分辨率为 92.8 m。2个因子由相对坡位（RSP）和水系网基准面（Chnl_base）组成，最适分辨率为 60.8 m。3个因子的模型为水系网高程（Chnl_alti），水系网基准面（Chnl_base）和多尺度山谷指数（MRVBF），最佳分辨率为64 m。有6组4个因子的组合模型其r均为0.71，分辨率从64 m至136 m不等。5个因子共有2个组合模型，r值为0.78，分辨率为152 m。每组均包含水系网高程（Chnl_alti）、高

表3　不同数量因子最佳组合及分辨率

Table 3　Optimal combination of different numbers of factors and resolution

因子数量 Number of attributes	变量名 Names of attributes combinations	分辨率 Resolution（m）	模型相关系数r值 r of model	验证集r值 r of validation data	验证集 RMSE Validation RMSE
1	相对坡位RSP	92.8	0.5	0.448	64.256
2	相对坡位及水系网基准面RSP，Chnl_base	60.8	0.6	0.467	63.564
3	水系网高程、多尺度山谷指数及水系网基准面Chnl_alti，MRVBF，Chnl_base	64	0.7	0.526	61.112
4	相对坡位、山谷深度、高程及归一化高度 RSP，Vall_depth，Elevation，Normalh	64	0.71	0.601	57.431
4	相对坡位、高程、归一化高度及多尺度山谷指数 RSP，Elevation，Normalh，MRVBF	121.6	0.71	0.601	57.788
4	水系网高程、标准化高度、归一化高度及多尺度山脊指数 Chnl_alti，Standh，Normalh，MRRTF	136	0.71	0.558	59.640
4	水系网高程、中坡位、地表纹理及坡高度 Chnl_alti，Midslppst，Texture，Slph	59.2	0.71	0.524	61.607
4	山谷深度、高程、地表纹理及水系网基准面 Vall_depth，Elevation，Texture，Chnl_base	88	0.71	0.552	59.916
4	高程、坡高、水系网基准面、多尺度山脊指数 Elevation，Slph，Chnl_base，MRRTF	83.2	0.71	0.564	59.345
5	水系网高程、高程、归一化高度、水系网基准面及多尺度山脊指数 Chnl_alti，Elevation，Normalh，Chnl_base，MRRTF	152	0.74	0.560	59.610
5	水系网高程、高程、中坡位、水系网基准面及多尺度山脊指数Chnl_alti，Elevation，Midslppst，Chnl_base，MRRTF	152	0.74	0.573	58.928
22	全部变量All attributes	40	0.78	0.236	63.229
22	全部变量All attributes	64	0.78	0.329	59.234

程（Elevation）、水系网基准面（Chnl_base）、多尺度山脊指数（MRRTF）四个因子，此外归一化高度（Normalh）和中坡位（Midslppst）2个因子分在不同组中。同时还计算了全部22个因子组合模型的r值为0.78，最佳分辨率为40 m或64 m。

为了确定最终采用的模型，使用预留的25%计2 208个样点对这些预选的模型的预测能力和有效性进行了检验。r值有着明显的差异，22个因子模型的r值最差，甚至小于单个因子的r值（表3）。这可能是由于因子太多，引入的噪声也随之加大，从而无法真正总结内在的规律。验证点的r值也是随着因子数量的增多而增加，但r值最大的是4个因子的组合，共有2组，分辨率分别为64 m和121.6 m。考虑到数据的存储空间前者是后者的4倍，模型运算时间指数级增长，因此本文选择相对坡位（RSP）、高程（Elevation）、归一化高度（Normalh）和多尺度山谷指数（MRVBF）组成，栅格大小为121.6 m的模型为最优模型。

2.4　土壤有机碳含量空间分布特征

为了表达不同模型对原始样点SOC的再现能力和相互间的对比，所有图均采用相同色系和分级值。分级值为11、14、18和28，分别为全部原始8 570个样点SOC值按从小到大排列后，20%、40%、60%和80%的分段值。这些模型生成的SOC

总体空间分布格局是一致的。SOC的高值基本发生在湿地地区。除湿地外，SOC总体上西部较东部高。从空间上对比这些模型生成的SOC空间分布图（图5），1个和2个因子生成的图形明显综合过

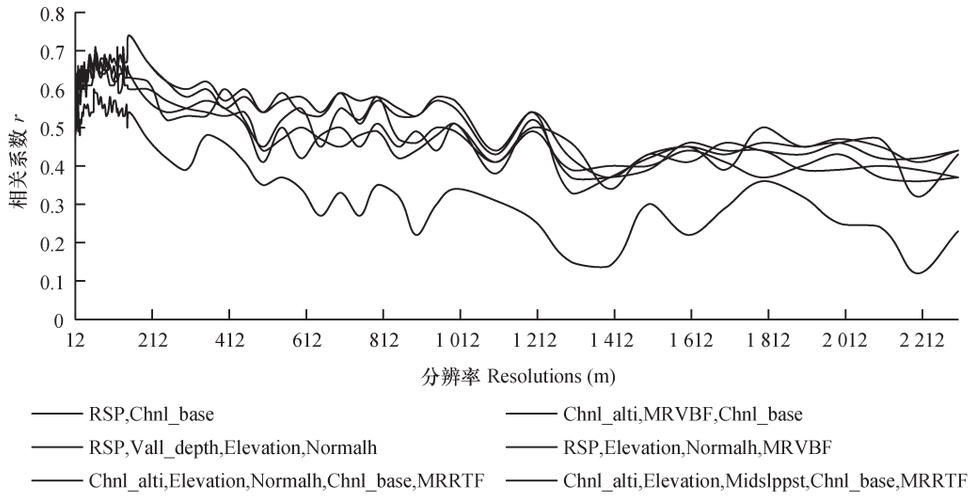
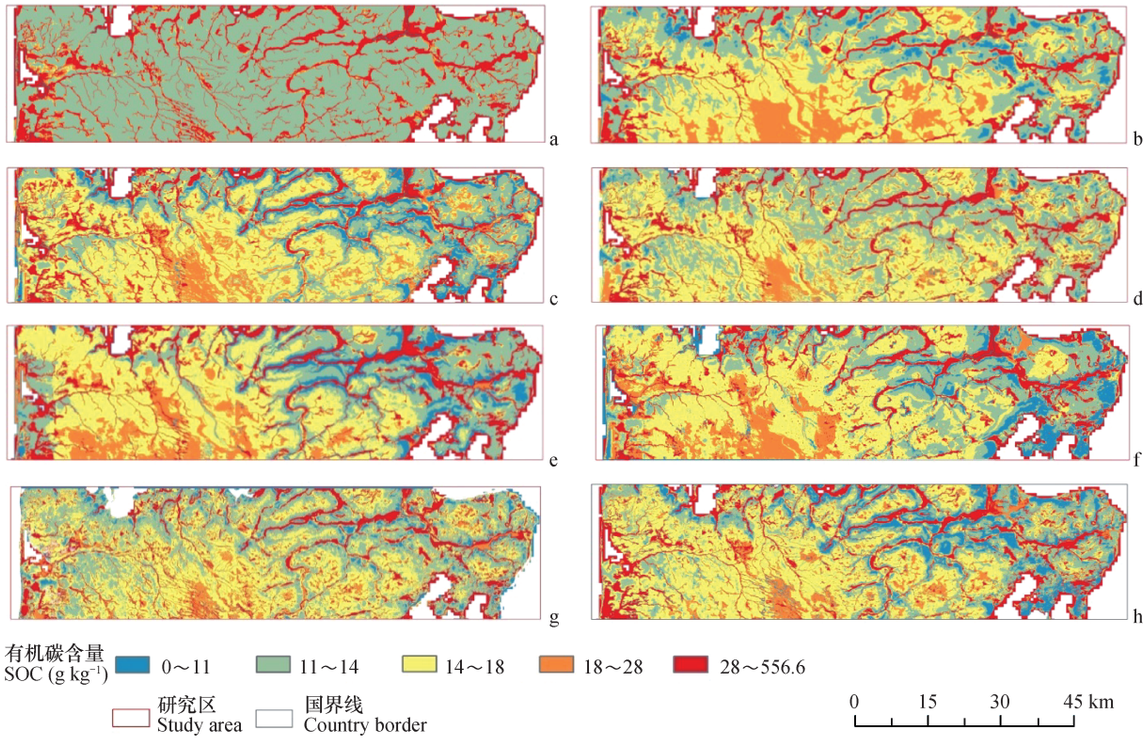


图4 最优模型r曲线

Fig. 4 Curve of r with resolutions of optimae model



注：（a）RSP，92.8 m，（b）RSP及Chnl_base，60.8 m，（c）Chnl_alti，Chnl_base及MRVBF，64 m，（d）RSP，Vall_depth，Elevation，Normalh，4变量64 m，（e）RSP，Elevation，Normalh，MRVBF，4变量，121.6 m，（f）Chnl_alti，Elevation，Normalh，Chnl_base，MRRTF5变量，152 m，（g）全部22变量，40 m，（h）全部22变量，64 m Note:（a）RSP at 92.8 m，（b）RSP and Chnl_base at 60.8 m，（c）Chnl_alti，Chnl_baseand MRVBF AT 64 m，（d）RSP，Vall_depth，Elevation，Normalh，4 attributes at 64 m，（e）RSP，Elevation，Normalh，MRVBF，4 attributes at 121.6 m，（f）Chnl_alti，Elevation，Normalh，Chnl_base，MRRTF 5 attributes at 152 m，（g）all 22 attributes at 40 m，and（h）all 22 attributes at 64 m

图5 不同模型参数生成的土壤有机碳图

Fig.6 Predicted maps of SOC

表4　克里格方法与模型模拟结果对比

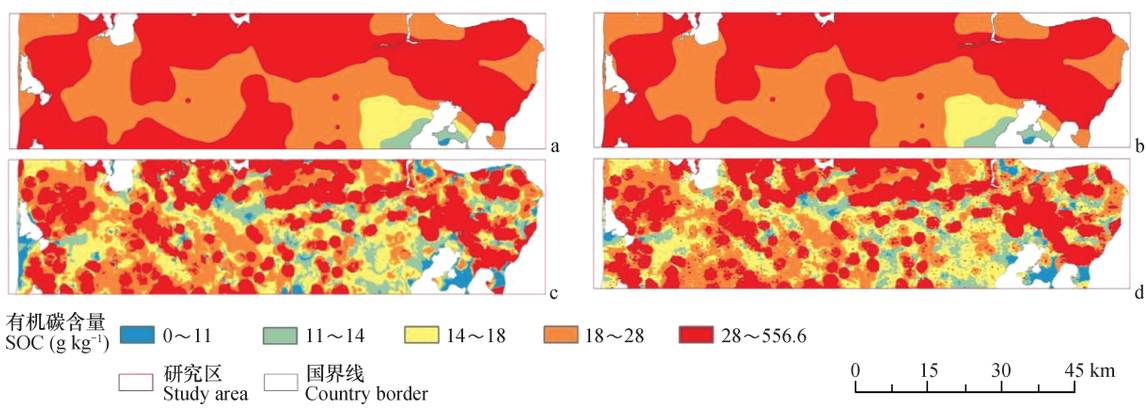
Table 4　Results comparison of fitting of the models with Kriging

不同预测模型 Names of attributes combinations		预测最小值 Predicted Min	预测最大值 Predicted Max	预测平均值 Predicted mean	标准误差 SD
数据挖掘模型 Cubist 模型	相对坡位 RSP	10.66	51.10	20.39	12.87
	相对坡位及水系网基准面 RSP，Chnl_base	0	482.4	24.31	35.08
	水系网高程，多尺度山谷指数及水系网基准面 Chnl_alti，MRVBF，Chnl_base	0	482.4	25.97	43.57
	相对坡位，山谷深度，高程及归一化高度 RSP，Vall_depth，Elevation，Normalh	0	461.6	26.00	46.88
	相对坡位，高程，归一化高度及多尺度山谷指数 RSP，Elevation，Normalh，MRVBF	0	525.9	65.34	123.7
	水系网高程，高程，中坡位，水系网基准面及多尺度山脊指数 Chnl_alti，Elevation，Midslppst，Chnl_base，MRRTF	0	556.6	42.27	64.72
	全部变量（40 m）All attributes（40 m）	0	555.8	26.94	47.22
	全部变量（64 m）All attributes（64 m）	0	556.6	31.3	64.41
克里格模型 Kriging	全局泛克里格Global Universal Kriging	10.10	136.1	34.76	17.63
	全局普通克里格Global Ordinary Kriging	9.47	136.0	33.43	17.00
	全局克里格Universal Kriging	−798.59	749.1	31.93	28.76
	普通克里格Ordinary Kriging	−54.79	531.9	32.89	31.69

度，粒度偏大，22个变量的图形中噪声非常明显，4个和5个变量的综合效果最好。

2.5　与克里格预测模型对比

为了验证模型的预测能力和空间表达能力，与通常采用的克里格空间插值方法生成的空间分布图进行了对比。采用全局泛克里格、全局普通克里格、全局克里格以及普通克里格等多种空间插值方法生成空间分布图（图6）。利用2208个验证点验证其预测精度，r和均方根误差（RMSE）范围分别为0.34～0.47，64.6～67.3，与模型的结果相比差



注：（a）全局泛克里格插值，（b）全局普通克里格插值，（c）全局克里格插值，（d）普通克里格插值Note：（a）Global Universal Kriging interpolation，（b）Global ordinary Kriging interpolation（c）Global Kriging interpolation，（d）Ordinary Krigingin terpolation

图6　不同克里格方法生成的土壤有机碳空间分布图

Fig. 6　Predicted SOC distribution map relative to Kriging method

距较大。从空间分布图也可以明显看出，空间异质性未得到合理的表达，克里格方法生成的空间数据趋向于将数据进行均一化。从表4中的最大、最小以及方差也可以看出克里格生成的数据非常集中。

3 结 论

研究通过数据挖掘的方法，寻找最优的因子组合及最适栅格大小，构建丹麦典型区域土壤有机碳估算模型。结果表明不同的单个地形因子模型与栅格大小有不同的规律。绝大部分变量预测的r值与格栅大小呈负相关，只有个别的呈正相关。r最大值通常并未像预料的那样出现在最大分辨率处。多个变量组合的模型与栅格大小呈偏态的正态分布，不同的变量组合最佳分辨率的位置也不同。最大的r值并不是出现在最高分辨率处，而是出现在60～150 m附近，因此无论何种因子的组合，选择的分辨率不能低于200 m。总体上，变量越多，模型的r值越大，4到5个变量的组合的效果均能较好地预测SOC。但是由于噪声的影响，超过4个变量的组合后，变量越多对未知数据的预测能力越差。综合数据的存贮量和计算量、模型复杂度、预测精度以及空间表达能力，由4个因子相对坡位（RSP）、高程（Elevation）、归一化高度（Normalh）和多尺度山谷指数（MRVBF）组合的，对应的分辨率121.6 m为SOC最优估算模型。

参 考 文 献

[1] Eswaran H, Vandenberg E, Reich P. Organic-carbon in soils of the world. Soil Science Society of America Journal, 1993, 57 (1): 192—194

[2] Lal R. Soil carbon sequestration to mitigate climate change. Geoderma, 2004, 123 (1/2): 1—22

[3] Post W M, Emanuel W R, Zinke P J, et al. Soil carbon pools and world life zones. Nature, 1982, 298 (5870): 156—159

[4] Lal R. Soil carbon sequestration impacts on global climate change and food security. Science, 2004, 304 (5677): 1623—1627

[5] Melillo J, Steudler P, Aber J, et al. Soil warming and carbon-cycle feedbacks to the climate system. Science, 2002, 298 (5601): 2173—2176

[6] Nichols J. Relation of organic carbon to soil properties and climate in the southern great plains. Soil Science Society of America Journal, 1984, 48 (6): 1382—1384

[7] Hontoria C, Saa A, Rodríguez-Murillo J. Relationships between soil organic carbon and site characteristics in peninsular spain. Soil Science Society of America Journal, 1999, 63 (3): 614—621

[8] Percival H J, Parfitt R L, Scott N A. Factors controlling soil carbon levels in new zealand grasslands is clay content important? Soil Science Society of America Journal, 2000, 64 (5): 1623—1630

[9] Stevens A, van Wesemael B. Soil organic carbon stock in the Belgian Ardennes as affected by afforestation and deforestation from 1868 to 2005. Forest Ecology and Management, 2008, 256 (8): 1527—1539

[10] 李凌浩. 土地利用变化对草原生态系统土壤碳储量的影响. 植物生态学报, 1998, 22 (4): 300—302
Li L H. Effects of land-use change on soil carbon storage in grassland ecosystems (In Chinese). Acta Phytoecologica Sinica, 1998, 22 (4):300—302

[11] Fisher M, Rao I, Ayarza M, et al. Carbon storage by introduced deep-rooted grasses in the south American Savannas. Nature, 1994, 371 (6494): 236—238

[12] Burke I C, Yonker C, Parton W, et al. Texture, climate, and cultivation effects on soil organic matter content in US grassland soils. Soil Science Society of America Journal, 1989, 53 (3): 800—805

[13] 于东升，史学正，孙维侠，等. 基于1∶100万土壤数据库的中国土壤有机碳密度及储量研究. 应用生态学报, 2005, 16 (12): 2279—2283
Yu D S, Shi X Z, Sun W X, et al. Estimation of China soil organic carbon storage and density based on 1∶1 000 000 soil database (In Chinese). Chinese Journal of Applied Ecology, 2005, 16 (12): 2279—2283

[14] Sims Z, Nielsen G. Organic carbon in montana soils as related to clay content and climate. Soil Science Society of America Journal, 1986, 50 (5): 1269—1271

[15] 解宪丽，孙波，周慧珍，等. 中国土壤有机碳密度和储量的估算与空间分布分析. 土壤学报, 2004, 41 (1): 35—43
Xie X L, Sun B, Zhou H Z, et al. Organic carbon density and storage in soils of China and spatial analysis (In Chinese). Acta Pedologica Sinica, 2004, 41 (1): 35—43

[16] Franko U, Oelschlägel B, Schenk S. Simulation of temperature, waterand nitrogen dynamics using the model candy. Ecological Modelling, 1995, 81 (1): 213—222

[17] Li C, Frolking S, Crocker G J, et al. Simulating trends in soil organic carbon in long-term experiments using the dndc model. Geoderma, 1997, 81 (1): 45—60

[18] Hao Y, Lal R, Owens L, et al. Effect of cropland management and slope position on soil organic carbon pool at the north appalachian experimental watersheds. Soil and Tillage Research, 2002, 68 (2): 133—142

[19] McBratney A B, Mendonça Santos M L, Minasny B. On digital soil mapping. Geoderma, 2003, 117 (1/2): 3—52

[20] 陆晓宇，张洪江，程金花，等. 地形因子与人工林土壤有机碳密度的关系——以晋西黄土丘陵区为例. 东北林业大学学报，2013，41（1）：48—50，58
Lu X Y, Zhang H J, Cheng J H, et al. Ralationship of topographic factors and soil organic carbon density under plantations: A case in loess hilly region in western Shanxi ProvinceChina (In Chinese). Journal of Northeast Forestry University, 2013, 41 (1): 48—50, 58

[21] Mulder V L, de Bruin S, Schaepman M E, et al. The use of remote sensing in soil and terrain mapping — A review. Geoderma, 2011, 162 (1/2): 1—19

[22] Cavazzi S, Corstanje R, Mayr T, et al. Are fine resolution digital elevation models always the best choice in digital soil mapping? Geoderma, 2013, 195/196: 111—121

[23] Gessler P, Chadwick O, Chamran F, et al. Modeling soil - landscape and ecosystem properties using terrain attributes. Soil Science Society of America Journal, 2000, 64 (6): 2046—2056

[24] Behrens T, Zhu A X, Schmidt K, et al. Multi-scale digital terrain analysis and feature selection for digital soil mapping. Geoderma, 2010, 155 (3/4): 175—185

[25] Greve M H, Greve M B, Bøcher P K, et al. Generating a danish raster-based topsoil property map combining choropleth maps and point information. Geografisk Tidsskrift-Danish Journal of Geography, 2007, 107 (2): 1—12

[26] Kim D, Zheng Y. Scale-dependent predictability of dem-based landform attributes for soil spatial variability in a coastal dune system. Geoderma, 2011, 164 (3/4): 181—194

[27] Smith M P, Zhu A X, Burt J E, et al. The effects of DEM resolution and neighborhood size on digital soil survey. Geoderma, 2006, 137 (1/2): 58—69

[28] Shi X, Girod L, Long R, et al. A comparison of lidar-based dems and usgs-sourced DEMs in terrain analysis for knowledge-based digital soil mapping. Geoderma, 2012, 170: 217—226

[29] Murphy P N C, Ogilvie J, Meng F-R, et al. Modelling and mapping topographic variations in forest soils at high resolution: A case study. Ecological Modelling, 2011, 222 (14): 2314—2332

[30] Thompson J A, Bell J C, Butler C A. Digital elevation model resolution: Effects on terrain attribute calculation and quantitative soil-landscape modeling. Geoderma, 2001, 100 (1/2): 67—89

[31] Sorensen R, Seibert J. Effects of DEM resolution on the calculation of topographical indices: TWI and its components. Journal of Hydrology, 2007, 347 (1/2): 79—89

[32] Grunwald S. Multi-criteria characterization of recent digital soil mapping and modeling approaches. Geoderma, 2009, 152 (3/4): 195—207

[33] Florinsky I V, Eilers R G, Manning G R, et al. Prediction of soil properties by digital terrain modelling. Environmental Modelling & Software, 2002, 17 (3): 295—311

[34] Simbahan G C, Dobermann A, Goovaerts P, et al. Fine-resolution mapping of soil organic carbon based on multivariate secondary data. Geoderma, 2006, 132 (3/4): 471—489

[35] Peng Y, Knadel M, Gislum R, et al. Predicting soil organic carbon at field scale using a national soil spectral library. Journal of Near Infrared Spectroscopy, 2013, 21 (3): 213—222

[36] Adhikari K, Kheir R B, Greve M B, et al. Comparing kriging and regression approaches for mapping soil clay content in a diverse danish landscape. Soil Science, 2013, 178 (9): 505—517

[37] Greve M H, Kheir R B, Greve M B, et al. Quantifying the ability of environmental parameters to predict soil texture fractions using regression-tree model with GIS and lidar data: The case study of Denmark. Ecological Indicators, 2012, 18 (7): 1—10

[38] Kheir R B, Greve M H, Bocher P K, et al. Predictive mapping of soil organic carbon in wet cultivated lands using classification-tree based models: The case study of Denmark. Journal of Environmental Management, 2010, 91 (5): 1150—1160

[39] 刘京，朱阿兴，张淑杰，等. 基于样点个体代表性的大尺度土壤属性制图方法. 土壤学报，2013，50（1）：12—20
Liu J, Zhu A X, Zhang S J, et al. Large-scaled soil attribute mapping method based on individual representativeness of sample sites (In Chinese). Acta Pedologica Sinica, 2013, 50 (1): 12—20

[40] Schou A, Nielsen N. Atlas of Denmark: Text and photographs//Hagerup I H.The landscapes. 1949

[41] Madsen H B, Norr A H, Holst K A. The Danish soil

classification atlas of Denmark. Copenhagen: Rosendahl Press, 1992

[42] Wang L, Liu H. An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. International Journal of Geographical Information Science, 2006, 20 (2): 193—213

[43] Minasny B, McBratney A B. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. Chemometrics and Intelligent Laboratory Systems, 2008, 94 (1): 72—79

[44] Holliday I E. Kolmogorov-Smirnov test (v1.0.3) in free statistics software (v1.1.23-r7). Office for Research Development and Education, 2012

Optimal Estimation Model of Soil Organic Carbon Based on the Terrain Factor

GUO Zhixing[1] YUAN Yuzhi[1] GUO Yin[1, 2] SUN Hui[1, 2] CHAI Min[1] CHEN Zepeng[3] Mogens H. Greve[4]

(1 *Guangdong Key Laboratory of Agricultural EnvironmentPollution Integrated Control*, *Guangdong Institute of Eco-environmental and Soil Sciences*, *Guangzhou* 510650, *China*)

(2 *College of Resources and Environment*, *Shanxi Agriculture University*, *Jinzhong*, *Shanxi* 030800, *China*)

(3 *Guangdong Tobacco Company*, *Guangzhou* 510610, *China*)

(4 *Department of Agro-ecology*, *Faculty of Science and Technology*, *Aarhus University*, 8830 *Tjele*, *Denmark*)

Abstract 【Objective】As an important component of the global carbon pool, soil organic carbon (SOC) is the largest organic carbon pool in the terrestrial ecosystem and plays an extremely important role in the global carbon cycle and global warming. The SOC pool is subject to the impacts of both natural and human activities and sure closely related to terrain attributes or factors. There are a number of methodsfor calculation of SOC, which can roughly be sorted into three types, that is, empirical, statistical and mechanismones. But none of them can be used to predict or calculate reapidly soil organic carbon pool of a region rapidly. Remote sensing is an efficient technical means for fast acquisition of DTM, from which numerous information can be derived with the aid of GIS, thus making it possible to constitute a model for rapid calculation of SOC. 【Method】Based on the Digital Terrain Model (DTM) and the topographic attributesderived thereof, an optimal SOC prediction model was built up, taking into account factor combination and resolution with Cubist, a powerful data mining tool for generating rule-based models. This tool works on condition-specific rules where the output is a set of rules and each rule has a specific multivariate linear model attached. Whenever a situation matches the condition of a rule, the associated model is used to calculate or predictevalues. A total of 8 570 soil samplescollected from the 7 100 km^2 study area were divided into two groups randomly, 6 362 for training and the other 2 208 for model validation, a total of 2 514 820 models were constructed based on 71 selected resolutions and all possible combinations of no more than 5of the 22 terrains attributes. According to the correlation coefficient (R), terrain factors, varying in number, were selected, to form optimal models with their corresponding resolutions, Based on these models, SOC maps were plotted. 【Result】Results show that the relationsships between resolution and single-factor models are diversified, it is not true that the higher the resolution, the better the model. The R value of a single-factor model is not necessarily the factor that determines its importance in a multi-factor model. All the multi-factor modelsexhibit a similar rule of skewed normal distribution. Each factor and its combination has a factor-specific optimal resolution, varying in the range of 60 ~ 150 m. For models composed of whatever factors, the resolution t be selected should not be lower than 200 m. The variable of the optimal single-factor model is RSP, with resolution being 92.8 m, the variables of the optimal two-factor model are RSP and Chnl_base with resolution being 60.8 m, while the variables of the three-factor model are Chnl_alti, Chnl_base and

MRVBF，with resolution being 64 m. There are 6 four-factor models，with R being 0.71 and resolution varying in the range of 64 ~ 136 m，and 2 five-factor models with R being 0.78，and resolution being 152meters. Every model consists at least of Chnl_alti，elevation，Chnl_base and MRRTF，and in the 2 five-factor modelsNormalh or Midslppst is added. The R of all the models consisting of any four ro five of the 22 factors was calculated to be 0.78 with two optimal resolutions，i.e. 40 and 64 m. In general，the more the variables，the higher the R of the models. But owing to impact of the noise，models with more than four factors decine in predictive ability.Four to five is the appropriate number of factors in combination，making the models more capable of predicting SOC. Comparative analysis of the SOC maps plotted with the aid of global fan Kriging，global ordinary Kriging，global kriging and ordinary Kriging shows that regardless of the number of factors in the model，this method is better than all the four Kriging interpolation methods in prediction of spatial variation of SOC and prediction accuracy.【Conclusion】Takinginto comprehensive account storage space，amount of calculationm，sophistication of the model，accuracy of prediction and ability of spatial expression，the optimal model for the study region should be the four-factor model，consisting of relative slope position，elevation，normalheight and MRVBF，with resolution being 121.6 m.

Key words　Soil organic carbon（SOC）；Digital soil mapping；Data mining；Digital Terrain Model（DTM）；Terrain attribute；Optimal model

（责任编辑：檀满枝）