

文章编号: 1000-0240(2011)05-1088-10

基于多元自适应回归样条的青藏高原 温泉区域的冻土分布制图

张秀敏¹, 南卓铜^{2*}, 吴吉春¹, 杜二计², 王通¹, 游艳辉¹

(1. 中国科学院寒区旱区环境与工程研究所冻土工程国家重点实验室, 甘肃兰州 730000;

2. 中国科学院寒区旱区环境与工程研究所, 甘肃兰州 730000)

摘要: 以探地雷达、电磁测深、钻探等技术方法获得野外数据及数字高程(DEM)遥感数据为基础, 通过聚类分析和相关性分析对高程、坡度、坡向等因素对多年冻土分布的影响进行了量化研究. 利用非线性的多元自适应回归样条(MARS)方法建立了基于高程、太阳辐射的多年冻土分布模型, 通过自身的交叉验证及对对比年平均地温模型和逻辑回归模型的总体分类精度, 说明 MARS 模型具有较好的分类精度. 运用 MARS 模型模拟了整个温泉区域冻土的空间分布特征. 结果表明: MARS 模型分类精度较高, 验证了此模型模拟温泉区域冻土分布的可行性; 此模型除了考虑高程对多年冻土分布的控制作用外, 还体现了太阳辐射这一局地综合因素对多年冻土分布的调整作用, 较好地模拟了高程相对较低的低山区多年冻土的存在.

关键词: 冻土分布模拟; 多元自适应回归样条; 青藏高原; 多年冻土

中图分类号: P642.14 **文献标识码:** A

0 引言

多年冻土是高寒环境赖以生存及维持生态平衡的物质基础, 多年冻土的分布、变化极大地影响着各种寒区生态系统的稳定与演替^[1]. 在工程设施领域, 多年冻土的空间分布状况影响到工程建筑物的稳定性^[2]. 因此, 精确的冻土制图对于冻土区的生态环境和工程建设具有重要的作用. 随着地理信息系统(GIS)和遥感(RS)的发展, 国内外学者在不同区域针对不同的空间尺度开发了多种冻土空间分布模型. 根据其理论基础的不同, 可以将多年冻土分布模型分为两类: 面向过程的物理模型和经验-统计模型^[3], 其中 1) 物理模型是建立在地表能量平衡基础上的, 主要描述冻土与大气系统的水热交换过程, 比如 PERMEBAL 模型^[3-5]、PERMACLIM 模型^[6]等; 2) 经验-统计模型不考虑地气能量交换过程和具体的冻土运作机理, 以多年冻土存在为因变量, 以影响多年冻土分布的因素(纬度、气温、坡

度、坡向、太阳辐射、高程、积雪和岩性等)为自变量, 通过分析二者的关系来构建和描述研究区多年冻土分布的统计函数模型^[7]. 国外应用较多的模型为 PERMAKART 和 PERMAMOD 地形经验模型^[8-10]、等效纬度^[11]、直接辐射^[12]、冻结数^[13-15]、多年冻土顶板温度^[16-17]、雪底温度^[18-22]和逻辑回归模型^[20, 23-24]. 与高纬冻土分布模型相比, 由于我国青藏高原特殊的高寒环境, 野外调查资料相对缺乏, 在模拟和预测青藏高原冻土分布时, 国内学者较少利用物理模型^[25], 相比较而言, 经验-统计模型得到较广泛的研究, 主要有高程模型^[26-27]、等效高程^[28]和地温模型^[29-31]等.

目前, 应用于青藏高原的经验-统计模型主要专注于纬度、高程等宏观因素对冻土的控制作用, 忽略了微观局地因素对冻土分布的影响. 统计方法方面, 多用线性回归(一元、多元)来判断多年冻土的空间分布, 由于多年冻土与环境因子间存在着非线性关系, 造成冻土制图的精度较低. 一些研究者

收稿日期: 2011-01-03; 修订日期: 2011-04-06

基金项目: 科技部基础性工作专项(2008FY110200); 全球变化研究国家重大科学研究计划项目(2010CB951402)资助

作者简介: 张秀敏(1982—), 女, 山东菏泽人, 2007年毕业于鲁东大学, 现为在读博士研究生, 主要从事普通冻土学的研究工作.

E-mail: zhangxm@lzb. ac. cn

* 通讯作者: 南卓铜, E-mail: nztong@lzb. ac. cn

开始尝试非线性方法在青藏高原冻土制图领域的应用. Li 等^[24]利用了非线性的逻辑回归方法, 模拟了祁连山东部的冻土分布, 取得较好的模拟结果. 然而逻辑回归要求数据符合逻辑函数的指数形式, 在一定应用程度上, 普适性较差.

多元自适应回归样条(Multivariate Adaptive Regression Splines, MARS)是一种泛化能力很强的专门针对高维数据的非线性回归方法, 1991年由统计学家 Jerry Friedman 第一次提出^[32]. 自从 20 世纪 90 年代以来, MARS 已应用于土地覆盖分类^[33]、预测物种分布^[34-35]等方面的研究, 并取得较好的预测结果. 但到目前为止, 研究学者利用 MARS 方法用于冻土分布的研究较少, 而多年冻土的存在是气候条件和地理因素等因素综合作用的结果, 它们之间的关系具有运用 MARS 方法的数据特点. 因此, 本文的研究目的是通过分析青藏高原温泉区域多年冻土(野外物探、钻探)和地形因素(主要从 DEM 上提取和计算)的资料, 通过 MARS 方法来建立二者的回归关系, 得到一种新的经验-统计模型, 并运用此模型模拟了温泉区域的冻土分布, 通过与年平均地温模型、逻辑回归模型的对比分析, 说明了 MARS 模型在冻土分布模拟方面具有一定的适用性.

1 多元自适应回归样条方法

MARS 方法通过样条函数来模拟复杂的非线性关系, 它将整个非线性模型划分为若干个区域, 在每个特定的区域由一段线性回归直线来拟合. MARS 模型^[32]的定义为:

$$\begin{aligned} \hat{y} &= \hat{f}_M(x) = a_0 + \sum_{m=1}^M a_m S_m(x) \\ &= a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{k_m} [S_{k_m}(x_{v(k,m)} - t_{k_m})]_+ \quad (1) \end{aligned}$$

式中: \hat{y} 为依赖变量的预测值; a_0 为参数; a_m 为第 m 个样条函数的系数; $S_m(x)$ 为第 m 个样条函数; M 为模型中含有的样条函数的数目; 区域之间的线性回归线的交点称为结点, k_m 为结点数; S_{k_m} 的值取 1 或 -1, 表示右侧或左侧的样条函数; $v(k, m)$ 为独立变量的标识; t_{k_m} 标识结点的位置.

每个基函数代表依赖变量的给定区域, MARS 的基函数是单一样条函数或者是两个(多个)样条函数的交互结果. 右侧(见式(2))和左侧(见式(3))的样条函数分别定义如下:

$$[S_{k_m}(x_{v(k,m)} - t)]_+ = \begin{cases} (x - t_{k_m}) & \text{当 } x \geq t_{k_m} \\ 0 & \text{其他} \end{cases} \quad (2)$$

$$[S_{k_m}(x_{v(k,m)} - t)]_+ = \begin{cases} (t_{k_m} - x) & \text{当 } x \leq t_{k_m} \\ 0 & \text{其他} \end{cases} \quad (3)$$

式中: t 为结点的位置; $x - t_{k_m}$ 和 $t_{k_m} - x$ 为描述给定 t 时右侧和左侧区域的样条函数; “+”为对于负值取 0.

MARS 模型构建算法包括前向逐步选择基函数的过程、剪枝过程和确定最优模型 3 个过程. 在第一过程中需要给定基函数个数的最大值 M 和交互的基函数数目的最大值 N , M 一般是指自变量的个数的 2 倍, N 根据用户的需要而定; 剪枝过程是基于广义交互验证(Generalized Cross Validation, GCV)标准进行的, 当 GCV 的值达到最小时, 对应的预测模型为最佳模型.

$$GCV(\lambda) = \frac{\sum_{i=1}^N (y_i - \hat{f}_\lambda(x_i))^2}{(1 - M(\lambda)/N)^2} \quad (4)$$

式中: $M(\lambda)$ 为模型中有效的参数个数; \hat{f}_λ 为每一步估计的最佳模型; λ 为模型中项的个数; N 为基函数的个数.

MARS 具有以下特点:

(1) MARS 在处理复杂的非线性变量关系时, 不需要假设预测变量和预报因子的线性关系、指数关系及正态假设.

(2) MARS 是一种泛化能力很强的专门针对高维数据的回归方法, 以“前向”和“后向”算法逐步筛选因子, 具有很强的自适应性. 在整个运算过程中, 基函数的确定都是根据数据自动完成, 不需要人工设定. 尽管线性回归和逻辑回归可以根据一定的“前向”算法进行因子的筛选, 但是不能自动“后向”删除相关性不大的因子. 相比较而言, MARS 具有线性回归方法和逻辑回归所无法比拟的特点.

(3) 整个运算过程快捷且得到的模型具有较好的解释能力, 对于说明预测变量的变化往往与某几种环境因素具有重要的联系具有直观性.

2 研究区概况及数据准备

2.1 研究区概况

研究区处于青藏高原东部的温泉区域, 即青海省玛多县和兴海县的交界处(图 1a), 地理范围为 $99^{\circ}6' \sim 99^{\circ}42' E$, $35^{\circ}6' \sim 35^{\circ}42' N$ 之间, 总面积约 $2\,520 \text{ km}^2$, 区内海拔介于 $3\,430 \sim 5\,300 \text{ m}$ 之间. 区内主要有鄂拉山和姜路岭两山脉分布, 呈北西-南东走向, 和青康公路方向垂直, 形成一系列相间排

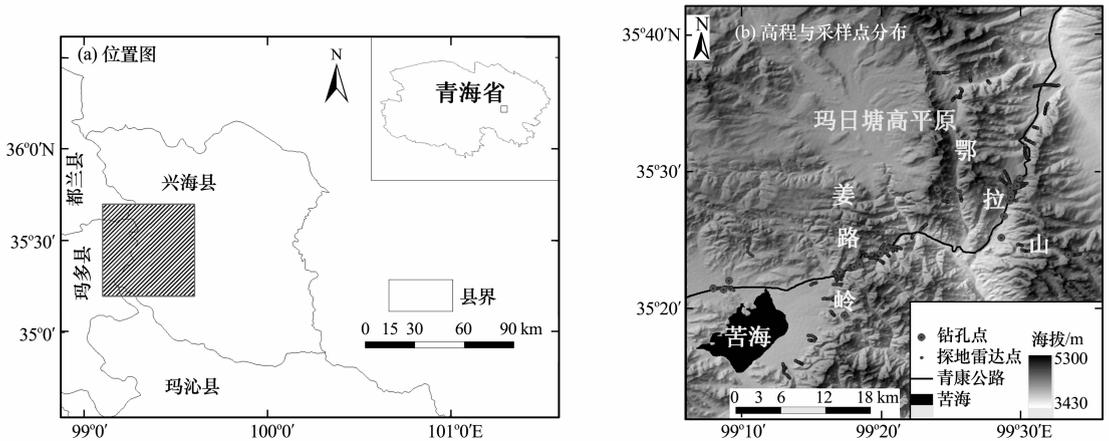


图1 温泉区域的位置图和等高及采样点分布

Fig. 1 Map showing the location of Wenquan area (a) and its topographic map with sampling plots (b)

列的山地和断陷盆地，主要有温泉谷地和苦海滩地(图1b)。根据研究区附近的花石峡气象站的气象资料^[36]，该区年平均气温为 $-3.2\text{ }^{\circ}\text{C}$ ，年极端最高气温为 $18.5\text{ }^{\circ}\text{C}$ ，年极端最低气温为 $-32.5\text{ }^{\circ}\text{C}$ ；年降水量达 $500\sim 600\text{ mm}$ ，属寒温带大陆性气候。研究区内河流纵横，植被发育良好，植被类型主要包括高寒草甸、高寒草原、高寒沼泽草甸和高寒灌丛，其中以高寒草甸和高寒草原为主，二者的面积占整个区域的80%左右。温泉区域作为“青藏高原多年冻土本底调查项目”的典型区，由于它是青藏高原东部由季节冻土区向大片连续多年冻土区的过渡区域，同时20世纪90年代和2004年对青康公路改、扩建期间，在该区域已经开展过多年冻土调查工作，对区内多年冻土分布及特征有一定的认识。

2.2 数据来源及处理方法

在“青藏高原冻土本底调查项目”的支持下，2009年9—10月对温泉区域开展了多年冻土的野外调查，利用探地雷达和电磁测深法对此区域进行了野外勘测。共采集130条剖面，以2m的间隔记录1个测点，共626个点(图1b)，并记录了测点的经纬度坐标及高程。其中，探地雷达在冻土探测中使用的探测方式主要共偏移距测量法和共中心点测量法，电磁测深法主要是瞬变电磁法和频率域电磁法。根据探测结果及结合该区已有知识，在室内我们判读了上述测点的冻土类型及剖面大致的多年冻土下界海拔。为了验证物探方法的准确性，在温泉区域布设了21个钻孔(图1b)，钻探现场详细记录了钻孔点经纬度、高程、孔深、多年冻土上限及地表状况等信息。利用钻孔揭示的实地情况，对探地雷达、电磁测深点判定结果做进一步的校正。校正

后的探地雷达、电磁测深点和钻孔点为本文工作提供了冻土类型的基础信息。

应用的遥感数据为30m ASTER-GDEM数据，通过从GDEM上直接提取和计算得到了经度、纬度、高程、坡度、坡向、曲率、平面曲率和剖面曲率的空间分布值；太阳辐射作为影响冻土分布的重要因子，其中月潜在直接入射辐射相比于年总辐射更能体现太阳辐射与多年冻土存在的关系。因此，基于ArcGIS的Solar Analyst模块计算了2009年1—12月份的潜在直接入射辐射^[37]，得到共20个地形因素的空间分布信息，然后建立626个物探点及21个钻孔点的矢量点图层，用此矢量点图层分别提取上述20个地形因素的数据。

2.3 聚类分析和相关性分析

由于物探方法在进行冻土探测时，两个测点之间的距离仅是2m，间隔距离较小，导致相近样点的环境条件相似，在建立多年冻土存在与环境因子的关系时，过多的样点并不具有代表性和典型性。因此，通过聚类分析来进行样本点个数的筛选。聚类分析就是直接比较各事物的性质，将性质相近的归为一类，将性质差别较大的归入不同的类。本文选用Q型聚类分析方法主要是由于此聚类方法是对研究对象本身进行分类，而我们关注的就是多年冻土是否存在这一研究对象，数据特点符合使用Q型聚类分析方法的条件，其中最短距离法的聚类特点在于首先合并最相似的两项，用两类间最短的距离代表两类间的距离，即距离最近最相似的聚为一类^[38]。因此，利用最短距离聚类方法对626个样本点进行了样本点的筛选分析，根据逐步合并的过程，最终得到代表性强的样本点425个。

多年冻土的存在往往是某一种或几种因素起主

导作用, 通过对多年冻土存在与地形因素进行相关性分析可以来判定影响多年冻土分布的主要影响因素. 利用上述经过聚类分析筛选后得到的 425 个样点的多年冻土存在指标数据和从遥感数据提取的影响冻土分布的 20 个因素进行了相关分析, 为后面冻土制图模型选取主要影响因素提供基础. 结果如表 1 所示: 多年冻土存在与高程呈正相关关系, 相关系数为 0.35, 说明海拔越高, 多年冻土存在的可能性越大, 这一结果体现了海拔是控制和影响多年冻土的主导因素. 同时多年冻土存在和 1—5 月份的潜在直接入射辐射及 7—12 月份的潜在直接入射辐射的关系显著, 并且呈负相关关系, 说明随潜在直接入射辐射的值增大, 多年冻土存在的几率越小. 这和文献[28]中提到的直接入射辐射与冻土地温存在较明显的正相关关系的观点是一致的, 即潜在入射太阳辐射越高, 冻土地温越高. 潜在直接入射辐射作为一个综合坡度、坡向等因素对多年冻土综合作用的因子, 其与多年冻土的存在有相关性. 相关性分析也表明, 多年冻土与坡度、坡向、曲率、平面曲率、剖面曲率等单一地形因子的相关性不明显; 同时多年冻土与经纬度的相关性不明显可能是由于本研究区的经纬度区间较小造成的(表 1).

3 模型的建立和精度验证

3.1 基于 MARS 的模拟结果

经过聚类分析和相关性分析得到 425 个多年冻土存在指标数据与高程、1—5 月份的潜在直接入射辐射、7—12 月份的潜在直接入射辐射数据 12 个影响因子之间的关系较为显著. 在此数据基础上, 利用 MARS 方法来建立温泉区域的冻土预测模型. 其中, 设定交互基函数数目的最大值 N 为 5; 基函数个数的最大值 M 为 24; 当 GCV 的最小值为 0.1834 时得到了温泉区域的最优 MARS 模型, 如公式 5 所示: 式中各个基函数如表 2 所示.

$$y = 0.89821 + 8.255 \times 10^{-3} BF_2 - 1.8234 \times 10^{-7} BF_2 \\ - 1.7234 \times 10^{-6} BF_3 + 1.9671 \times 10^{-6} BF_4 \\ + 1.097 \times 10^{-6} BF_5 - 4.644 \times 10^{-7} BF_6 \\ + 1.0853 \times 10^{-6} BF_7 - 7.8911 \times 10^{-9} BF_8 \\ + 7.5039 \times 10^{-9} BF_9 \quad (5)$$

根据 MARS“前向”和“后向”算法逐步筛选因子, 得到多年冻土与高程、1—3 月、5 月、8 月、10 月、12 月的潜在直接入射辐射等环境因子有联系(式 5). 结果表明, 高程是控制山地冻土分布的主导因素之一; 作为地面热量根本来源的太阳辐射是

影响多年冻土分布的另一主导因素, 尤其是太阳辐射较弱的冷季(9 月至翌年 4 月), 对多年冻土存在具有重要的作用. 潜在直接入射辐射作为一个综合坡度、坡向、地表状况等因素的因子, 在冬季时候, 地表可能有积雪存在, 由于雪盖的反射率大, 地面吸收辐射将小于有效辐射, 有可能导致到达地面的辐射平衡为负值, 此负值存在时间的长短对土的冻结具有重要的意义.

把高程、1—3 月、5 月、8 月、10 月和 12 月的潜在直接入射辐射的空间分布代入式(5), 得到了温泉区域冻土的预测值 y . y 是一个 $[0, 1]$ 之间的连续性数值, 对于预测的结果需要设定一个临界值进行冻土的分类, 以 0.1 为间隔划分了 10 个临界值, 通过对 425 个样点的实测冻土类型和预测冻土类型进行对比分析发现在临界值为 0.5 时, 得到的冻土类型总体正确率最高为 78%, 因此将临界值设置为 0.5, 即预测值 $y \geq 0.5$ 的区域为多年冻土区, 反之为季节冻土区(图 2a).

3.2 模型适用性验证

本文采用 5 重交叉验证, 将筛选后的 425 个样本点随机分解为 5 份, 将其中 340 个点作为训练样本, 85 个点作为验证样本, 反复验证 5 次. 其中应用的变量为高程、1—5 月的潜在直接太阳辐射及 7—12 月份的潜在直接入射辐射, 每次的总体精度分别为 78%、80%、72%、72%、81%, 平均总体精度为 76.6%, 对比文献[24]74%的冻土制图总体精度, 说明模型在研究区有相对较好的模拟精度.

3.3 MARS 与地温模型、逻辑回归模型的对比

目前, 关于我国青藏高原冻土的分布模拟, 利用较多的为线性回归的地温模型^[29-21]; 用非线性回归方法较少, Li 等^[24]利用了非线性的逻辑回归方法成功模拟了祁连山东部的冻土分布, 但是对于青藏高原其它地区的冻土是否具有实用性是未知的. 因此, 利用地温模型和逻辑回归模型模拟了温泉区域的冻土分布, 并与 MARS 的研究结果进行对比.

根据年平均地温模型(Mean Annual Ground Temperature, MAGT)^[31], 以 21 个钻孔的年平均地温值(10 m 深度)为因变量, 纬度、经度、高程影响因素为自变量, 采用逐步回归法进行了自变量的筛选, 删除不显著变量, 最终得到以年均地温为因变量, 高程为自变量的一元线性回归方程:

$$T = -0.0066x + 27.8343, r^2 = 0.9731 \quad (6)$$

表 1 多年冻土存在与局地因素的相关性分析
Table 1 Spearman correlation matrix between topo-climatic factors and binary permafrost factors

	<i>Per</i>	<i>Lon</i>	<i>Lat</i>	<i>Aspect</i>	<i>Slope</i>	<i>Ele</i>	<i>PSR₁</i>	<i>PSR₂</i>	<i>PSR₃</i>	<i>PSR₄</i>	<i>PSR₅</i>	<i>PSR₆</i>	<i>PSR₇</i>	<i>PSR₈</i>	<i>PSR₉</i>	<i>PSR₁₀</i>	<i>PSR₁₁</i>	<i>PSR₁₂</i>	<i>Cur</i>	<i>Plan</i>	<i>Pro</i>	
<i>Per</i>	1																					
<i>Lon</i>	-0.01	1																				
<i>Lat</i>	-0.03	0.72**	1																			
<i>Aspect</i>	-0.04	-0.09*	0.08	1																		
<i>Slope</i>	0.06	0.21**	0.33**	0.14**	1																	
<i>Ele</i>	0.35**	-0.14**	0.09*	0.18**	0.19**	1																
<i>PSR₁</i>	-0.25**	-0.06	-0.02	-0.17**	-0.14**	0.08*	1															
<i>PSR₂</i>	-0.25**	-0.07	-0.02	-0.17**	-0.18**	0.09*	0.99**	1														
<i>PSR₃</i>	-0.24**	-0.11**	-0.05	-0.18**	-0.25**	0.09*	0.99**	0.99**	1													
<i>PSR₄</i>	-0.22**	-0.18**	-0.13**	-0.18**	-0.41**	0.10**	0.95**	0.96**	0.98**	1												
<i>PSR₅</i>	-0.14**	-0.26**	-0.24**	-0.15**	-0.68**	0.13**	0.76**	0.78**	0.83**	0.92**	1											
<i>PSR₆</i>	-0.04	-0.27**	-0.27**	-0.11**	-0.83**	0.15**	0.51**	0.55**	0.61**	0.75**	0.94**	1										
<i>PSR₇</i>	-0.1*	-0.33**	-0.34**	-0.13**	-0.77**	0.12**	0.62**	0.64**	0.71**	0.83**	0.98**	0.97**	1									
<i>PSR₈</i>	-0.2**	-0.24**	-0.21**	-0.17**	-0.50**	0.11**	0.89**	0.91**	0.94**	0.99**	0.96**	0.82**	0.90**	1								
<i>PSR₉</i>	-0.24**	-0.15**	-0.10*	-0.18**	-0.30**	0.09*	0.98**	0.99**	0.99**	0.86**	0.65**	0.75**	0.96**	1								
<i>PSR₁₀</i>	-0.25**	-0.1*	-0.04	-0.17**	-0.20**	0.08*	0.99**	0.99**	0.97**	0.80**	0.57**	0.67**	0.92**	0.99**	1							
<i>PSR₁₁</i>	-0.25**	-0.08*	-0.04	-0.17**	-0.16**	0.08*	0.99**	0.99**	0.95**	0.77**	0.53**	0.63**	0.90**	0.98**	0.99**	1						
<i>PSR₁₂</i>	-0.26**	-0.06	-0.01	-0.17**	-0.14**	0.08*	0.99**	0.99**	0.94**	0.75**	0.51**	0.61**	0.89**	0.98**	0.99**	0.99**	1					
<i>Cur</i>	0.02	0.09*	0.06	-0.03	0.02	-0.02	-0.08	-0.08*	-0.09*	-0.09*	-0.08	-0.06	-0.07	-0.08*	-0.09*	-0.08*	-0.08	-0.07	1			
<i>Plan</i>	0.05	0.09*	0.07	-0.04	0.07	0.003	-0.10	-0.10*	-0.11**	-0.12**	-0.12**	-0.11**	-0.11**	-0.12**	-0.11**	-0.10**	-0.10**	-0.09**	0.91**	1		
<i>Pro</i>	0.02	-0.07	-0.03	0.01	0.04	0.04	0.04	0.04	0.04	0.04	0.01	0.001	0.002	0.03	0.04	0.04	0.04	0.04	-0.90**	-0.63**	1	

注：*：显著水平为 0.05 (2-tailed)；**：显著水平为 0.01 (2-tailed)；*Per*：多年冻土 (Permafrost)；*Lon*：经度 (Longitude)；*Lat*：纬度 (Latitude)；*Ele*：高程 (Elevation)；*PSR_n*：*n* 月份的潜在直接太阳辐射；*Cur*：曲率 (Total curvature)；*Plan*：平面曲率 (Plan curvature)；*Pro*：剖面曲率 (Profile curvature)。

表 2 基函数

Table 2 Primary functions used in the multivariate adaptive regression splines model

基函数编号	基函数	变量	系数
BF_1	$\max(0, 4237 - x_1)$	Ele	8.255×10^{-3}
BF_2	$\max(0, x_3 - 51891) \times \max(0, 4389 - x_1)$	PSR_2, Ele	-1.8234×10^{-7}
BF_3	$BF_1 \times \max(0, 91394 - x_9)$	PSR_{10}, Ele	-1.7234×10^{-6}
BF_4	$BF_1 \times \max(0, x_7 - 178030)$	PSR_8, Ele	1.9671×10^{-6}
BF_5	$BF_1 \times \max(0, 178030 - x_7)$	PSR_8, Ele	1.097×10^{-6}
BF_6	$BF_1 \times \max(0, x_2 - 46864)$	PSR_1, Ele	-4.644×10^{-7}
BF_7	$BF_1 \times \max(0, 46864 - x_2)$	PSR_1, Ele	1.0853×10^{-6}
BF_8	$\max(0, x_6 - 181200) \times \max(0, 43398 - x_{11})$	PSR_5, PSR_{12}	-7.8911×10^{-9}
BF_9	$\max(0, x_6 - 181200) \times \max(0, 121240 - x_4)$	PSR_5, PSR_3	7.5039×10^{-9}

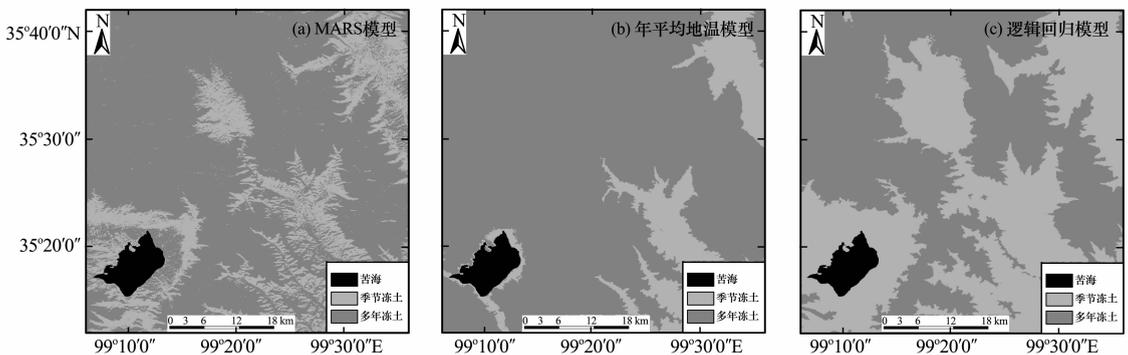


图 2 基于 MARS 模型和年平均地温模型和逻辑回归模型的研究区域多年冻土分布

Fig. 2 Distribution maps of permafrost and seasonal frozen soil in the study area simulated with the MARS model, the MAGT model and the logistic regression model, respectively

式中： T 为年平均地温($^{\circ}\text{C}$)； x 为高程(m). 以 30 m 的 GDEM 数据为基础，根据地温模型计算了温泉区域的年平均地温值，按 0.5°C 作为多年冻土与季节冻土的界线将研究区分为多年冻土区和季节冻土区^[31]，如图 2(b)所示。

同时，利用与 MARS 模型相同的 425 个样点的 12 个属性数据进行二值逻辑回归的运算，其中 172 个季节冻土点，253 个多年冻土点。采用逐步回归法进行了自变量的筛选，结果只有高程进入到模型中，得到公式(7)所示的逻辑回归方程：

$$p = \frac{e^{-30.621+0.007dem}}{1 + e^{-30.621+0.007dem}} \quad (7)$$

结合 30 m 的 GDEM 数据，逻辑回归模型得到温泉区域多年冻土空间分布概率值图，将概率 $\geq 50\%$ 的区域视为多年冻土区，则 $< 50\%$ 的区域为季节冻土区，得到区域的多年冻土空间分布格局，如图 2(c) 所示。

精度评价是遥感数据分类过程中一项不可缺少的工作，通过精度分析，分类者将改进分类模式提高分类精度；使用者能正确、有效地获取分类结果

中的信息^[39]。本文利用经过聚类分析后删除掉的 201 个野外勘测点，其中 86 个季节冻土点，115 个多年冻土点。通过建立误差矩阵，计算总体分类精度，如表 3 所示。MARS 模型的综合精度最高为 85.6%，而年平均地温和逻辑回归模型的综合精度较低，分别为 62.2%、50.3%。说明年平均地温模型和逻辑回归模型出现错分现象较多，年平均地温模型将 73 个季节冻土点错分为多年冻土点而逻辑回归模型将 90 个多年冻土点错分为季节冻土点。造成这些差异的主要原因在于 MARS 模型综合考虑了高程、潜在直接入射辐射等因素对多年冻土分布的影响，而年平均地温模型和逻辑回归模型只考虑了高程因素的作用，进而说明 MARS 模型相比其它两个模型在温泉区域具有较好的适用性。

3.4 多年冻土分布特征

依据 MARS 模型绘制的温泉区域多年冻土分布图(图 2a)，多年冻土的分布面积为 $1\ 881\ \text{km}^2$ ，占区域总面积的 76%。根据平均间隔的方法，将研究区分为 5 个高度带，其中海拔 3 800 m 以下的区

表 3 基于各模型的冻土类型分类精度评价

Table 3 Precision evaluation of the permafrost and seasonal frozen soil classified with the MARS model, the MAGT model and the logistic regression model

类别	MARS 模型		MAGT 模型		逻辑回归模型	
	多年冻土	季节冻土	多年冻土	季节冻土	多年冻土	季节冻土
多年冻土	97	11	112	73	25	10
季节冻土	18	75	3	13	90	76
总体分类精度	85.6%		62.2%		50.3%	

域占整个区域面积的 1.81%，主要位于温泉河谷区；海拔 3 800~4 200 m 高度带内的区域占整个区域面积的 24.34%，主要分布于苦海滩地、东南部的低山区及东北部的冲沟区；海拔 4 200~4 600 m 是整个研究区域的主要海拔高度带，占整个研究区域面积的 61.14%，分布于玛日塘盆地、鄂拉山和姜路岭的山前缘区及两山的低山区；海拔 4 600~5 000 m 高度带的面积占整个研究区域的 12.55%，分布于鄂拉山和姜路岭的中山区；高于海拔 5 000 m 的高山区区域仅占区域总面积的 0.16% (图 3a)。从空间分布特征上来看，多年冻土广泛分布于高、中、低山区，主要分布于鄂拉山和姜路岭山区及山前缘区(图 2a)，在各高度带内多年冻土面积所占比例分别为 71.35%、44.50%、82.86%、98.86% 和 97.58%(图 3b)；季节冻土在高、中和低山区也有分布，主要分布于东南部的低山区、大河坝谷地、温泉谷地、苦海滩地及玛日塘盆地区(图 2)，各高度带的分布比例分别为 28.65%、55.50%、17.14%、1.14% 和 2.42%(图 3b)。图 2(a)和图 3(b)表明，高程是控制多年冻土发育的主要因子；图 2(a)同时表明，对于微观尺度上温泉区域的多年冻土的发育特征，潜在直接入射辐射对其的作用比较显著的。虽然温泉研究区处于相对较小

的地理纬度内，有近似的太阳辐射天文总量，但是由于地面起伏变化(坡度、坡向)造成局部地区接受的太阳辐射仍然存在很大差异；再者由于太阳辐射与下垫面的植被存在一定的关系，夏季时，植被的存在对到达地表的太阳辐射起到郁闭作用；冷季时，高原上的植被也对应的进入衰落期，用于植被的热量将减少，到达下伏土层中的负热通量较大，有利于多年冻土的冻结。对于温泉区域东南部的低山丘陵区来说，即使海拔较低，仍然有多年冻土的存在，这是由于低山区具有相对较陡的坡度，减少了到达地面的太阳辐射；同时阴坡处有沼泽草甸和灌丛的分布，地表水分条件较好，植被覆盖率高，到达地表的太阳辐射部分被植物郁闭，部分因地表水分蒸发而消耗，因此，到达地表的热量被层层消耗掉而减少，有利于多年冻土的发育。而玛日塘盆地区由于其地形比较平缓，接受较多的太阳辐射，同时该地区地表水分状况比较干燥，因水分蒸发消耗掉的热量较少，再者覆盖度相对较低的高寒草原广泛分布在此区域，即使高程相对较高，仍有季节冻土存在。

从空间分布特征上来看，各个模型得到的季节冻土主要集中分布于海拔 3 400~4 200 m，而多年冻土主要集中分布于高于海拔 4 200 m 的地区，这

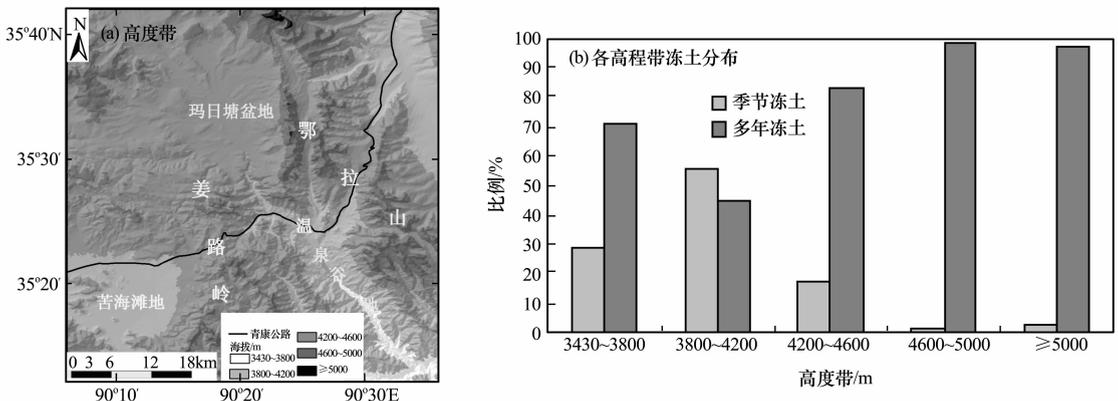


图 3 研究区高度带图(a)和基于 MARS 模型的高程带上冻土分布图(b)

Fig. 3 Altitudinal zones of the study area and percentages of permafrost and seasonal frozen soil in different altitudinal zones calculated with the MARS model

表 4 基于各模型的不同高程带的冻土分布

Table 4 Percentages of permafrost and seasonal frozen soil in different altitudinal zones calculated with different models

高程带 /m	MARS 模型		MAGT 模型		逻辑回归模型	
	多年冻土	季节冻土	多年冻土	季节冻土	多年冻土	季节冻土
3400~3800	71.35%	28.65%	0	1	0	1
3800~4200	44.50%	55.50%	30.97%	69.03%	0	1
4200~4600	82.86%	17.14%	1	0	60.72%	39.28%
4600~5000	98.86%	1.14%	1	0	1	0
≥5000	97.58%	2.42%	1	0	1	0

一结果体现了海拔是控制和影响多年冻土的主导因素。但是，在相同高程带上，各个模型模拟的冻土分布格局又有一些差异性，具体表现在年平均地温模型和逻辑回归模型模拟的结果在低于海拔 3 800 m 没有多年冻土的分布及高于海拔 4 600 m 的区域没有季节冻土存在，MARS 模型模拟的结果在各个高程带季节冻土和多年冻土都有分布(表 4)。造成这些差异的主要原因在于在各个高程带内，由于坡度、坡向及下垫面的不同，造成地表接受的太阳辐射不同，太阳辐射的不同将会影响多年冻土的分布。年平均地温模型和逻辑回归模型没有考虑这些局地因素对多年冻土分布的影响，这些细部冻土分布特征在 MARS 模拟中得到较好的表现。

4 结论

利用青藏高原温泉区域的探地雷达、电磁测深、钻孔资料和数字高程资料，通过聚类分析和相关分析对样本点和地形因素进行筛选，然后基于 MARS 的方法建立了多年冻土与高程，太阳辐射的回归关系，通过交叉验证和已有模型的对比分析，说明 MARS 模型具有较好的模拟结果，并用此模型模拟了整个温泉区域的多年冻土的空间分布，得到以下结论：

(1) 对比 MARS 模型、年平均地温模型及逻辑回归模型模拟冻土分布的精度和结果，表明 MARS 模型的分类精度最高。高程是控制区域多年冻土分布的主导因素，直接决定了区域冻土分布的宏观格局；三者的差异主要表现在高海拔地势比较平缓的玛日塘盆地区及低海拔东南部山区。这是由于 MARS 模型更多得考虑了潜在直接入射辐射因素对多年冻土分布的影响，年平均地温模型和逻辑回归模型仅仅考虑了高程这一因素，使得 MARS 模型的结果好于年平均地温模型和逻辑回归模型的结果。

(2) MARS 模型的结果表明，温泉区域多年冻土的分布面积为 1 881 km²，占区域总面积的 76%，在高于海拔 4 200 m 的高度带上广泛分布；而季节冻土的分布面积为 595.1 km²，主要集中在海拔 3 400~4 200 m 的高度带，说明高程是温泉区域多年冻土分布的主要制约因素；同时 MARS 模型纳入太阳辐射对多年冻土的影响，较好模拟了高程相对较低的低山区多年冻土的存在及高平原区季节冻土的存在。

致谢：本研究是青藏高原多年冻土本底调查研究项目的一部分，谨此对参加本项研究的盛煜老师、赵林老师、陈继老师、史健宗、石伟等表示衷心感谢；同时，感谢李静博士对论文提出的中肯意见和建议；感谢国际科学数据镜像提供 30 m ASTER-GDEM 数据 <http://datamirror.csdb.cn>。

参考文献 (References) :

- [1] Wang Genxu, Li Yuanshou, Wu Qingbai, *et al.* Impacts of permafrost changes on alpine ecosystem in Qinghai-Tibet Plateau[J]. *Science in China (Series D) Earth Sciences*, 2006, **36**(8): 743-754. [王根绪, 李元寿, 吴青柏, 等. 青藏高原冻土区冻土与植被的关系及其对高寒生态系统的影响[J]. *中国科学(D辑)*, 2006, **36**(8): 743-754.]
- [2] Li Jing, Sheng Yu, Chen Ji, *et al.* Modeling the distribution of permafrost along a transportation corridor from Cetar to Muri in Qinghai Province[J]. *Progress in Geography*, 2010, **29**(9): 1100-1106. [李静, 盛煜, 陈继, 等. 青海省柴达尔-木里地区道路沿线多年冻土分布模拟[J]. *地理科学进展*, 2010, **29**(9): 1100-1106.]
- [3] Hoelzle M, Mittaz C, Etzelmüller B, *et al.* Surface energy fluxes and distribution models of permafrost in European mountain areas: an overview of current developments[J]. *Permafrost and Periglacial Processes*, 2001, **12**: 53-68.
- [4] Lunardini V J. Climatic warming and the degradation of warm permafrost[J]. *Permafrost and Periglacial Processes*, 1998, **7**(4): 311-320.
- [5] Smith M W, Riseborough D W. Permafrost monitoring and detection of climate change[J]. *Permafrost and Periglacial Processes*, 1998, **7**(4): 301-309.

- [6] Guglielmin M, Aldighieri B, Testa B. PERMACLIM: a model for the distribution of mountain permafrost, based on climatic observations[J]. *Geomorphology*, 2003, **51**(4): 245–257.
- [7] Julián A, Chueca J. Permafrost distribution from BTS measurements (Sierra de Telera, Central Pyrenees, Spain): assessing the importance of solar radiation in a mid-elevation shaded mountainous area[J]. *Permafrost and Periglacial Processes*, 2007, **18**(2): 137–149.
- [8] Keller F. Automated mapping of mountain permafrost using the program PERMAKART within the geographical information system ARC/INFO [J]. *Permafrost and Periglacial Processes*, 1992, **3**(2): 133–138.
- [9] Imhof M. Modelling and verification of the permafrost distribution in the Bernese Alps (Western Switzerland)[J]. *Permafrost and Periglacial Processes*, 1996, **7**(3): 267–280.
- [10] Frauenfelder R, Allgwer B, Haerberli W. Permafrost investigations with GIS—a case study in the Fletschhorn area, Wallis, Swiss Alps[C]//PERMAFROST—Seventh International Conference, Yellow Knife (Canada) Collection Nordicana, 1998: 291–295.
- [11] Jorgenson M T, Kreig R A. A model for mapping permafrost distribution based on landscape component maps and climatic variables[C]//PERMAFROST—Fifth International Conference, Tapir, 1988: 176–182.
- [12] Funk M, Hoelzle M. A model of potential direct solar radiation for investigating occurrences of mountain permafrost [J]//*Permafrost and Periglacial Processes*, 2006, **3**(2): 139–142.
- [13] Nelson F E, Outcalt S I. A frost index number for spatial prediction of ground-frost zones[C]//*Proceedings of the 4th International Conference on Permafrost*, Fairbanks, 1983: 903–911.
- [14] Nelson F E, Anisimov O A. Permafrost zonation in Russia under anthropogenic climatic change[J]. *Permafrost and Periglacial Processes*, 2006, **4**(2): 137–148.
- [15] Anisimov O A, Nelson F E. Permafrost distribution in the Northern Hemisphere under scenarios of climatic change[J]. *Global and Planetary Change*, 1996, **14**(1–2): 59–72.
- [16] Riseborough D W. The mean annual temperature at the top of permafrost, the TTOP model, and the effect of unfrozen water[J]. *Permafrost and Periglacial Processes*, 2002, **13**(2): 137–143.
- [17] Juliussen H, Humlum O. Towards a TTOP ground temperature model for mountainous terrain in central-eastern Norway [J]. *Permafrost and Periglacial Processes*, 2007, **18**(2): 161–184.
- [18] Riseborough D, Shiklomanov N, Etzelmüller B, *et al.* Recent advances in permafrost modeling[J]. *Permafrost and Periglacial Processes*, 2008, **19**(2): 137–156.
- [19] Hoelzle M. Permafrost occurrence from BTS measurements and climatic parameters in the Eastern Swiss Alps[J]. *Permafrost and Periglacial Processes*, 2006, **3**(2): 143–147.
- [20] Lewkowicz A G, Ednie M. Probability mapping of mountain permafrost using the BTS method, Wolf Creek, Yukon Territory, Canada [J]. *Permafrost and Periglacial Processes*, 2004, **15**(1): 67–80.
- [21] Ishikawa M, Hirakawa K. Mountain permafrost distribution based on BTS measurements and DC resistivity soundings in the Daisetsu Mountains, Hokkaido, Japan[J]. *Permafrost and Periglacial Processes*, 2000, **11**(2): 109–123.
- [22] Isaksen K, Hauck C, Gudevang E, *et al.* Mountain permafrost distribution in Dovrefjell and Jotunheimen, southern Norway, based on BTS and DC resistivity tomography data [J]. *Norsk Geografisk Tidsskrift-Norwegian Journal of Geography*, 2002, **56**(2): 122–136.
- [23] Jason R J. The occurrence of alpine permafrost in the Front Range of Colorado[J]. *Geomorphology*, 2005, **67**(3–4): 375–389.
- [24] Li Jing, Sheng Yu, Wu Jichun, *et al.* Probability distribution of permafrost along a transportation corridor in the northeastern Qinghai province of China[J]. *Cold Regions Science and Technology*, 2009, **59**(1): 12–18.
- [25] Li Shuxun, Cheng Guodong, Guo Dongxin. The future thermal regime of numerical simulating permafrost on Qinghai-Xizang (Tibet) Plateau, China, under climate warming[J]. *Science in China (Series D)*, 1996, **26**(4): 342–347. [李述训, 程国栋, 郭东信. 气候持续变暖条件下青藏高原多年冻土变化趋势数值模拟[J]. *中国科学(D辑)*, 1996, **26**(4): 342–347.]
- [26] Cheng Guodong, Wang Shaoling. On the zonation of high altitude permafrost in China[J]. *Journal of Glaciology and Geocryology*, 1982, **4**(2): 12–16. [程国栋, 王绍令. 试论中国高海拔多年冻土带的划分[J]. *冰川冻土*, 1982, **4**(2): 12–16.]
- [27] Cheng Guodong. Problems on zonation of high altitude permafrost[J]. *Journal of Geographical Science*, 1984, **39**(2): 185–193. [程国栋. 我国高海拔多年冻土地带性规律之探讨[J]. *地理学报*, 1984, **39**(2): 185–193.]
- [28] Sheng Yu, Li Jing, Wu Jichun, *et al.* Distribution patterns of permafrost in the upper area of Shule River with the application of GIS technique[J]. *Journal of China University of Mining and Technology*, 2010, (1): 32–39. [盛煜, 李静, 吴吉春, 等. 基于GIS的疏勒河流域上游多年冻土分布特征[J]. *中国矿业大学学报*, 2010, (1): 32–39.]
- [29] Wu Qingbai, Li Xin, Li Wenjun. The response model of permafrost along the Qinghai-Tibetan Highway under climate change[J]. *Journal of Glaciology and Geocryology*, 2001, **23**(1): 1–6. [吴青柏, 李新, 李文君. 全球气候变化下青藏公路沿线冻土变化响应模型的研究[J]. *冰川冻土*, 2001, **23**(1): 1–6.]
- [30] Wu Qingbai, Li Xin, Li Wenjun. Computer simulation and mapping of the regional distribution of permafrost along the Qinghai-Xizang Highway[J]. *Journal of Glaciology and Geocryology*, 2000, **22**(4): 323–325. [吴青柏, 李新, 李文君. 青藏公路沿线冻土区域分布计算机模拟与制图[J]. *冰川冻土*, 2000, **22**(4): 323–325.]
- [31] Nan Zhuotong, Li Shuxun, Liu Yongzhi. Mean annual ground temperature distribution on the Tibetan Plateau permafrost distribution mapping and further application [J]. *Journal of Glaciology and Geocryology*, 2002, **24**(2): 142–148. [南卓铜, 李述训, 刘永智. 基于年平均地温的青藏高原冻土分布制图及应用[J]. *冰川冻土*, 2002, **24**(2): 142–148.]
- [32] Friedman J. Multivariate adaptive regression splines[J]. *The Annals of Statistics*, 1991, **19**(1): 1–67.

- [33] Quirós E, Felicísimo Á M, Cuartero A. Testing multivariate adaptive regression lines (MARS) as a method of land cover classification of TERRA-ASTER satellite images[J]. *Sensors*, 2009, **9**(11): 9011–9028.
- [34] Leathwick J R, Rowe D, Richardson J, *et al.* Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish[J]. *Freshwater Biology*, 2005, **50**(12): 2034–2052.
- [35] Elith J, Leathwick J. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines[J]. *Diversity and Distributions*, 2007, **13**(3): 265–275.
- [36] Chou Yaling, Sheng Yu, Wei Zhenming. Temperature and deformation differences between southern and northern slopes of highway embankment on permafrost[J]. *Chinese Journal of Rock Mechanics and Engineering*, 2009, **28**(9): 1896–1903. [丑亚玲, 盛煜, 韦振明. 多年冻土区公路路基阴阳坡温度及变形差异分析[J]. *岩石力学与工程学报*, 2009, **28**(9): 1896–1903.]
- [37] Fu P, Rich P M. Design and implementation of the Solar Analyst: an ArcView extension for modeling solar radiation at landscape scales[C]//Proceedings of the Nineteenth Annual ESRI User Conference, San Diego, USA, 1999.
- [38] Wang Lu. Application and practice of SPSS statistical analysis [M]. Beijing: Chemical Industry Press, 2010: 230–250. [王璐. SPSS 统计分析基础、应用与实践[M]. 北京: 化学工业出版社, 2010: 230–250.]
- [39] Zhao Ping, Fu Yunfei. Cart-based land use /cover classification of remote sensing images[J]. *Journal of Remote Sensing*, 2005, **9**(6): 708–716. [赵萍, 傅云飞. 基于分类回归树分析的遥感影像土地利用/覆被分类研究[J]. *遥感学报*, 2005, **9**(6): 708–716.]

Modeling Permafrost Distribution in Wenquan Area over Qinghai-Tibet Plateau by Using Multivariate Adaptive Regression Splines

ZHANG Xiu-min¹, NAN Zhuo-tong², WU Ji-chun¹, DU Er-ji², WANG Tong¹, YOU Yan-hui¹

(1. State Key Laboratory of Frozen Soil Engineering, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou Gansu 730000, China; 2. Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou Gansu 730000, China)

Abstract: In order to understand the distribution patterns of permafrost in the Wenquan area on the Qinghai-Tibet Plateau, the effects of altitude, slope and aspect and other topo-climatic factors on the distribution of permafrost were studied, using the correlation analysis with digital elevation (DEM) data, borehole observations and measures from ground penetrating radar (GPR) and the electromagnetic sounding method. A permafrost distribution model based on the nonlinear multiple adaptive regression splines (MARS) method was developed, taking elevation and direct solar radiation as variables. Five-fold cross validation shows

that this model has a good simulation capability in describing the permafrost distribution spatial pattern in the study area. Applying the model to the study area indicates that in the Wenquan area there is 1 881 km² of permafrost area, accounting for 76% of the total Wenquan area. The MARS model is better than the mean annual ground temperature model and the logistic model, because the MARS model takes into account not only elevation, the predominantly controlling factors in a mountainous area, but also solar radiation, which has adjusting effect to the permafrost distribution.

Key words: permafrost distribution modeling; multivariate adaptive regression splines (MARS); Qinghai-Tibet Plateau; permafrost