doi: 10.7522/j.issn.1000－0240.2013.0009

Huang Peipei, Nan Zhuotong. Estimation of 0-cm soil temperature over the Tibetan Plateau based on the wavelet analysis and adaptive network-fuzzy inference system [J]. Journal of Glaciology and Geocryology, 2013, 35(1): 74－83. [黄培培, 南卓铜. 基于Wavelet-ANFIS和MODIS地表温度产品的青藏高原0 cm土壤温度估算方法[J]. 冰川冻土, 2013, 35(1): 74－83.]

基于Wavelet-ANFIS和MODIS地表温度产品的青藏高原0 cm土壤温度估算方法

黄培培[1], 南卓铜[1,2*]

(1. 中国科学院 寒区旱区环境与工程研究所, 甘肃 兰州 730000; 2. 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)

摘 要: 0 cm土壤温度是冻土模型的上边界条件, 连续的、高质量的青藏高原0 cm土壤温度数据是进行准确冻土模拟的必要条件. 然而受复杂下垫面的影响, 遥感手段无法获取可靠的0 cm土壤温度. 利用自适应网络模糊推理系统(ANFIS)结合青藏高原实测资料建立遥感地表温度产品(LST)与0 cm土壤温度的关系, 以实现通过LST估算青藏高原逐日0 cm土壤温度. 研究了ANFIS的各种参数组合, 发现筛选合适的小波函数、小波窗口、小波层数建立起来的Wavelet-ANFIS模型能较准确实现估算0 cm土壤温度的目的. 验证表明, 估算结果与气象站点实测0 cm土壤温度绝对误差在2 K以下, 相关系数0.98以上. 考虑到原始MODIS LST误差在0～2 K之间, 该方法可以获取较为理想的0 cm土壤温度, 为冻土模型提供准确的上边界输入.

关键词: 小波分析; 自适应网络模糊推理系统; MODIS地表温度产品; 青藏高原; 0 cm土壤温度

中图分类号: P407.8 **文献标识码**: A

0 引言

在冻土空间分布研究中, 冻土学家相继建立了各种基于经验统计或者物理过程的冻土分布模型[1], 借助有限的观测资料模拟多年冻土分布. 例如, 吴青柏等[2]在青藏高原应用了多年冻土顶板温度模型; 王之夏[3]利用多年冻土顶板温度模型和地面冻结数模型[4]分别进行了青藏高原多年冻土模型和评价, 其特点是结合了遥感数据. 南卓铜等[5]对地面冻结数模型进行重新推导, 使地面冻结数模型意义更明确. 这些冻土分布模型都需要0 cm土壤温度作为模型的上边界条件, 且也可以与一些陆面过程的温度输出[6－7]相互验证.

冻土模型中用到的0 cm土壤温度大多基于有限的实测资料, 而对于广袤的青藏高原, 实测站点稀少, 分布不均, 不具备足够的代表性. 为了解决0 cm土温资料不足的问题, 有学者提出n系数[8]以反映不同下垫面的气温和0 cm土温的关系, 从而通过资料相对丰富的气温来估算0 cm土温. 然而, 李述训等[9]的工作表明n系数存在季节和年际变化, 并不是很好的解决方案. 另一种可能的途径是通过气候模式或者陆面过程模型进行动力模拟, 然而, 受限于驱动这些模型的资料和模型本身的不完善, 以及各种下垫面参数对土温模拟的敏感性不同[10], 计算得到的0 cm土温精度往往无法满足冻土模型的要求. 王之夏[3]在冻土模型应用中, 使用了每天2次观测的MODIS LST, 取2次算术平均作为0 cm土壤温度的近似值. 王之夏等[11]工作也表明, MODIS LST与实测0 cm土温有较为一致的变化趋势, 但平均偏差较大, 青藏高原上达5 K以上; 二者在冷暖季有不同的空间变化特征, 这是因为MODIS LST更多反映的是地面覆盖层上的温

收稿日期: 2012-08-17; 修订日期: 2012-12-29

基金项目: 冻土工程国家重点实验室开放基金项目(SKLFSE201009); 科技基础性工作专项(2008FY110200)资助

作者简介: 黄培培(1986—), 女, 河南永城人, 2009年毕业于河南大学, 现为硕士研究生, 主要从事于GIS、RS的应用研究. E-mail: huangyingbing@yeah.net

* 通讯作者: 南卓铜, E-mail: nztong@lzb.ac.cn

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

度，由热红外辐射反演得到，而 0 cm 土壤温度受到近地面气温状况、植被、土壤状况等多层面的综合影响. 欧阳斌等[12]给出了一种结合 MODIS LST 一日 4 次观测拟合日均 0 cm 温度的方法，但一些参数(峰值、Shift 因子、高程等)的确定还需要进一步研究.

遥感提供了较为可靠的长时间序列且有足够空间分辨率的面域地表温度，如果能建立起 LST 与 0 cm 土温的关系，则为估算 0 cm 土温提供了一种新思路. 需要注意的是，0 cm 土壤温度是近地表温度和下垫面状况的综合结果，表现出极大的空间分异性，与 LST 是显著的非线性关系；遥感观测的 LST 反映的是卫星过境时刻的瞬时地表温度，是在一定空间分辨率的平均情况，与观测得到的 0 cm 地温在时间和空间分辨率上并不匹配. 基于这些认识，在本研究中我们采用一种自适应模糊神经网络(ANFIS)[13]的方法，通过实测的逐日 0 cm 土温训练建立 MODIS LST 与日均 0 cm 土温间的模型. 其技术难度在于设置合适的模型变量以体现影响 0 cm 土温的各种因子.

1　数据及方法

1.1　研究区及数据

研究区青藏高原范围在 75. 69°～104. 43° E，26. 00°～39. 82° N，平均海拔在 4 000～5 000 m，面积约 2. 60×10^6 km^2. 大部分在中国西南部，包括了西藏和青海的全部、四川西部、新疆南部，以及甘肃、云南的一部分.

本文使用的遥感数据，包括 2005 年 1 km 逐日 MODIS Terra 地表温度产品(MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1 km SIN Grid V005-MOD11)和 MODIS NDVI 16 d 合成数据. 提取 LST 白天和晚上两个温度数据波段(LST-Day-1 km、LST-Night-1 km)，以及对应的质量控制波段(QC_day、QC_night). 根据质量控制波段移除有云(“clouds”、“cirrus cloud”)、数据缺失(“missing pixel”)、低质量(“poor quality”)、发射误差过大(“Average emissivity error ＞ 0. 04”)、LST 误差大于 2 K(“LST error 2 K～3 K”、“LST error＞3 K”)及其他误差(“other error”)的像元. 选择 2005 年是考虑到此年度数据质量总体较高，经质量控制处理后仍有较多数据，可以满足本文建议方法的研究需求.

实测资料包括青藏高原 71 个气象站点的 2005 年日均 0 cm 土壤温度数据，使用其中 61 个站点逐日数据用于 ANFIS 训练，其余 10 个站点作为结果验证. 研究区及站点(61 个训练站点和 10 个检验站点)的空间位置如图 1 所示.

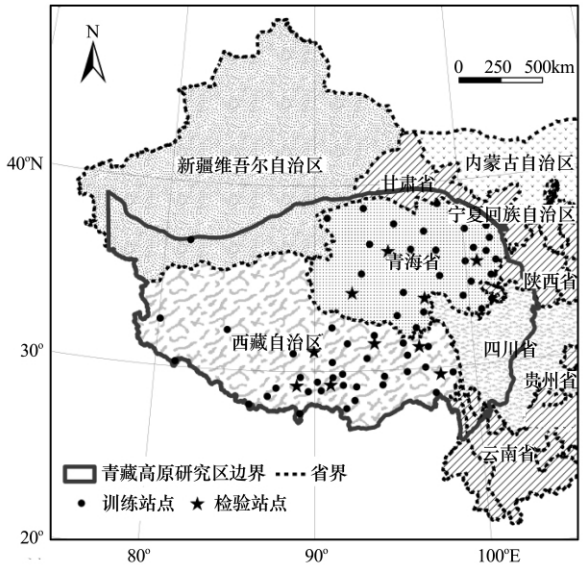


图 1　青藏高原及站点分布

Fig. 1　Map showing the Tibetan Plateau and the meteorological sites

地面气象场地一般选择有代表性的自然下垫面，我们通过 NDVI 检查这些站点下垫面的代表性. 对 71 个站点的暖季 NDVI 值统计分析显示，有 3 个站点暖季 NDVI 均值小于 0. 1，具体在 0～0. 06 之间，可近似为 0，可认为下垫面为裸露岩石或裸土. 有 7 个站点 NDVI 值为 0. 1～0. 2，7 个站点值为 0. 2～0. 3，15 个站点值为 0. 3～0. 4，17 个站点值 0. 4～0. 5，15 个站点值 0. 5～0. 6，7 个站点值为 0. 6 以上. 图 2 中为每个站点的全年和暖季的 NDVI 均值，说明了 71 个气象站点植被覆盖情况迥异，几乎分布在 0～1 的每个区间，由此可大致推

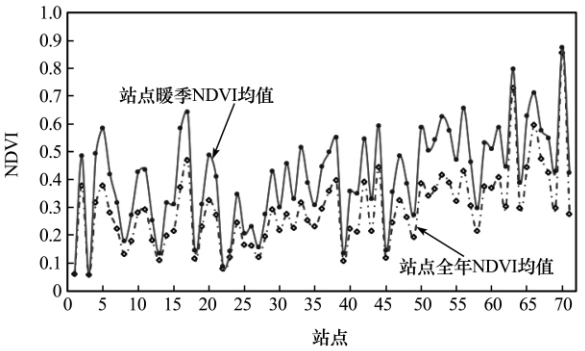


图 2　站点 2005 年平均 NDVI 和暖季平均 NDVI

Fig. 2　Annual mean and warm seasonal mean NDVI in the sites

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

断植被覆盖度处于裸土到茂密的不同状态. 因此，本文估算下垫面对 0 cm 土温影响时具备一定的代表性.

1.2 估算方法

1.2.1 自适应网络模糊推理系统(ANFIS)

自适应网络模糊推理系统[13](ANFIS，Adaptive-Network-Based Fuzzy Inference System)在很多领域都有成功的应用[14-15]. 它综合了神经网络[16]和模糊逻辑推理[17]的过程，通过自身训练和学习过程，不断调整变量隶属度函数图形参数，使模糊推理系统的输入、输出关系达到最佳组合.

ANFIS 采用一阶规则的 Takagi-Sugeno(TS)推理结构[18-19]，即规则的输出是输入变量的线性组合，多输入单输出，输出结果是每个规则输出的加权平均，其权重由隶属函数确定. 图 3 示例了两输入单输出的 ANFIS 5 层结构图.

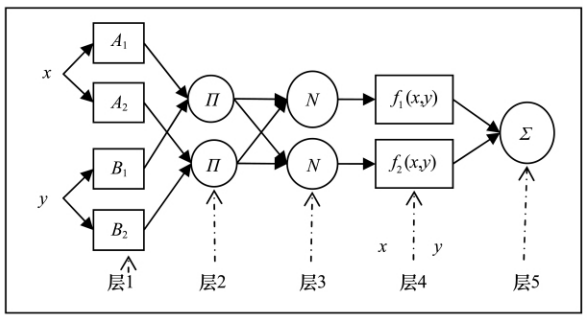


图 3 两输入单输出的 ANFIS 5 层结构图

Fig. 3 A simple ANFIS scheme for two inputs and single output

$$o_i^1 = u_{A_i}(x),\ i = 1,2 \tag{1}$$

$$o_i^1 = u_{B_i}(y),\ i = 1,2 \tag{2}$$

$$o_i^1 = u_{A_i}(x) = \frac{1}{1+\left[\left(\frac{x-c_i}{a_i}\right)^2\right]^{b_i}},\ i = 1,2 \tag{3}$$

$$o_i^2 = w_i = u_{A_i}(x)u_{B_i}(y),\ i = 1,2 \tag{4}$$

$$o_i^3 = \overline{w_i} = \frac{w_i}{\sum_i w_i},\ i = 1,2 \tag{5}$$

$$o_i^4 = \overline{w_i} f_i = \overline{w_i}(p_i x + q_i y + r_i),\ i = 1,2 \tag{6}$$

$$o_i^5 = F = f(x,y) = \sum_i \overline{w_i} \cdot f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{7}$$

层 1，隶属函数节点层或输入节点层，该层每一节点都生成模糊集隶属度，对于给定的输入 x，y，则输出对应变量的隶属度式(1)，(2). A_i、B_i 为模糊集，u_{A_i}、u_{B_i} 为 A_i 和 B_i 模糊集的隶属函数，o_i^1 为对应的隶属度. 本文选用钟型隶属度函数，如式(3)所示.

层 2，模糊规则前向节点层，表示输入信号相乘(式(4))，即规则的可信度.

层 3，归一化节点层，根据式(5)求解每一规则可信度占总规则可信度之和的比例.

层 4，模糊规则输出节点层(式(6))，$\overline{w_i}$ 为前一层的标准化可信度，$[p_i, q_i, r_i]$ 为函数 f_i 的参数集.

层 5，输出节点层，系统的输出为所有输入信号的和(式(7)).

1.2.2 ANFIS 结构优化及数据异常检测

当 ANFIS 输入变量增多，计算量急剧增大. 本文采用聚类分析法，在 ANFIS 模糊规则建立后按聚类法则[20]形成简化的神经网络，提高训练的计算效率. 步骤是：首先通过减法聚类方程通过提取一系列输入输出规则，决定“输入输出规则”和“隶属函数”的个数，再结合最小二乘估计决定每个规则对应的结果方程. 经结构优化后的效率提升可由图 4 演示. 当输入因子数为 9，ANFIS 按照排列组合自动生成神经网络结构如图 4(a)，减法聚类后形成的简化神经网络结构如图 4(b)，显著减少了计算量.

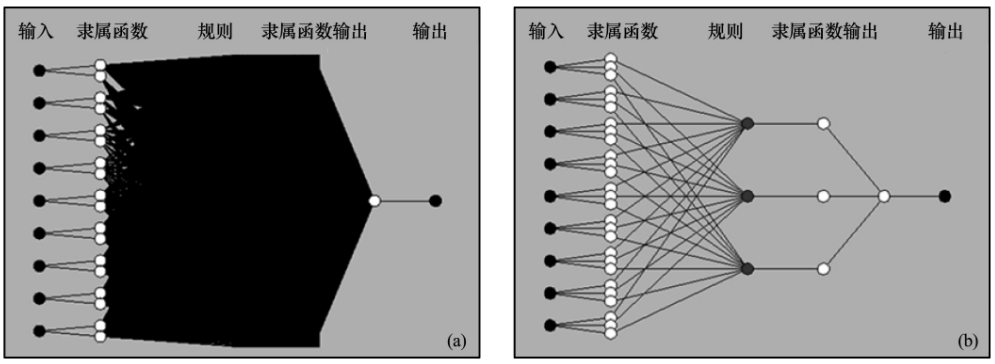


图 4 ANFIS 结构及减法聚类的 ANFIS 结构

Fig. 4 An original 9-input ANFIS structure and its simplified clustered structure

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

数据异常将极大影响 ANFIS 的训练结果，MODIS LST 尽管经过了初步的质量控制筛选，仍然需要进行进一步的异常检测. 本研究采用的方法是，首先将训练数据进行快速的 ANFIS 训练，确定出误差异常大的一些点，从而移除对应的可能异常的 MODIS LST 数据. 图 5 是通过快速 ANFIS 训练得到的绝对误差升序列(8 305个点集)，误差绝对值处于陡然上升的这些点和对应的 MODIS LST 被排除. 本文绝对误差大于 6 的值被去除，这些值处于误差陡然上升的部分且数量所占比例较小，约占总数的 5%.

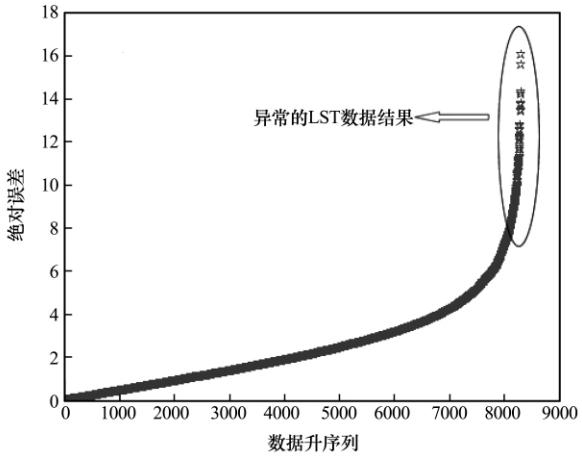


图 5　LST 异常数据检测

Fig. 5　Detection of LST outliers

1.3　ANFIS 参数组合与模型优化

0 cm 土壤温度是近地表温度、下垫面综合影响的结果. 结合 ANFIS 通过 LST 估算 0 cm 土温的技术关键是设置合适的模型变量. 本文从两个角度来考虑模型变量问题：一方面从可获取的自然因子中寻找合适的模型参数组合，具体考察植被、位置、地形等各因子组合，确定最优的自然因子组合；另一方面，由于有限的自然因子可能无法代表 0 cm 土温的最大变化组分，考虑通过小波分析(Wavelet)的方法进行主成分分析，筛选出若干最大变化组分以优化 ANFIS 模型. 此外，由于要估算的是日均 0 cm 地温，而 LST 是卫星过境瞬时地表温度，时间尺度的不匹配是影响估算精度的重要原因之一. 考察昼夜长、卫星过境时间和当地日出时刻 3 个时间因子对 ANFIS 估算精度的影响，确定出最显著影响的时间因子. 在上述两方面优化模型基础上增加时间因子再次优化，从而筛选出最优参数组合用于实际预测.

1.3.1　自然因子组合

考虑到 0 cm 土壤温度受多个层面的综合影响(地表气温、地理位置(经纬度)、地形(高程、坡度和坡向等)、下垫面状况(如冰雪覆盖，土地利用，植被覆盖、植被物候等)、表面土壤状况(如质地、土壤湿度等))，以及这些因素的可获取性，我们选取植被指数(NDVI)、经纬度位置和表示地形的高程、坡度和坡向进行优化组合.

自然因子分别和表示 LST 白天、夜晚两次观测的 LST_Day、LST_Night 一起形成 8 个候选模型. 如表 1 所列，M2～4 分别表示只考虑植被、位置、地形影响的模型，M5～7 分别表示此三类自然要素两两组合形成的模型，M8 是全部三类自然要素参与的模型. 作为对照的 M1 只有两次的 LST 观测变量，自然因子不参与其中. 通过 M1～M8 的对比，我们筛选最优的自然因子组合.

1.3.2　Wavelet 分解

0 cm 温度是诸多自然因子综合影响的结果，有限的自然因子组合可能仍不足以反映 0 cm 土温变化的主要成分. 一种可选的方案是使用小波分析[21－22]将 MODIS LST 两次观测分解为若干层，包括 1 层反映主成分的低通滤波和 n 层反映细节的

表 1　自然因子组合形成的候选模型

Table 1　Candidate ANFIS models constructed with natural factors

ANFIS 模型	输入变量	输入变量数
M1	LST_Day，LST_Night	2
M2	LST，植被信息(LST_Day，LST_Night，NDVI)	3
M3	LST，位置(LST_Day，LST_Night，经度，纬度)	4
M4	LST，地形(LST_Day，LST_Night，高程，坡度，坡向)	5
M5	LST，植被信息，位置(LST_Day，LST_Night，NDVI，经度，纬度)	5
M6	LST，植被信息，地形(LST_Day，LST_Night，NDVI，高程，坡度，坡向)	6
M7	LST，位置，地形(LST_Day，LST_Night，经度，纬度，高程，坡度，坡向)	7
M8	LST，植被信息，位置，地形(LST_Day，LST_Night，NDVI，经度，纬度，高程，坡度，坡向)	8

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

高通滤波. 低通、高通滤波作为 ANFIS 的输入因子进行训练. 我们需要优化的是有较好效果的小波函数及分解层数. Wavelet 分解待率定的小波函数包括 Haar 小波、Daubechies 小波(dbN)、SymletsA 小波(symN)、Biorthogonal 小波(biorNr. Nd)、Coiflet 小波组(coifN)、Morlet 小波和 Meyer 小波. 两次 LST 观测(白天、夜晚)被分解为 $2(n+1)$个 ANFIS 参数输入,形成一系列 Wavelet-ANFIS 候选模型,通过对各小波函数及 n 层 Wavelet-ANFIS 模型进行训练,筛选出最优小波函数及最优 n 值.

1.3.3 最优候选模型

在自然因子组合模型和小波分解模型的筛选基础上,得到优化后的自然因子和小波分解模型,再与率定出的时间因子组合形成模型,对该 4 种最优候选模型再次进行对比选取最优化的模型,对其误差进行分析,以用于实际的青藏高原日均 0 cm 土温估算.

1.3.4 评价指标

本文应用绝对误差 ME、均方根误差 RMSE 和相关系数 CORR 对模型训练和检验结果进行评价:

$$ME = \frac{\sum^{N} | T_a - T |}{N} \tag{8}$$

$$RMSE = \left[\sum^{N} \frac{(T_a - T)^2}{N} \right]^{\frac{1}{2}} \tag{9}$$

$$CORR = \frac{\sum^{N} (T_a - \overline{T_a})(T - \overline{T})}{\sqrt{\sum^{N} (T_a - \overline{T_a})^2 (T - \overline{T})^2}} \tag{10}$$

式中:T_a为预测日均 0 cm 土壤温度;T 为站点实测的日均 0 cm 土壤温度;$\overline{T_a}$和 $\overline{T}$ 分别为 T_a和 T 的均值;N 为样本个数. 绝对误差和均方根误差越低越好,相关系数越高则越显著.

2 结果与讨论

2.1 自然因子优化

站点位置上有效的 MODIS LST 和自然因子的数据特征表现见表 2. LST_day 和 LST_night 最大值分别为 51.51 ℃、20.25 ℃,最小值分别为 −21.51 ℃、−30.91 ℃,均值 16.84 ℃、−4.98 ℃,温度范围较大. NDVI 范围 0.049～0.93,说明在不同站点和时序上植被状况有显著差异,均值 0.24 表明植被状况整体上不是很好. 纬度范围遍及高原南北,经度范围则说明站点多集中于高原东部;海拔、坡向和坡度也表明出较为明显的差异性. 这些数据特性表明本文选取的这些自然因子具备了一定的代表性.

表 3 给出了自然因子组合的 ANFIS 候选模型的模拟结果绝对误差、均方根误差和相关系数,可以看出:1)自然因子参与的模型较之无自然因子参与的模型 M1,各统计指标都有改善,表明植被指数、位置、地形均不同程度的影响训练结果. 但单一自然因子(M2～4)整体影响不显著,其中位置和地形因素较植被因子显著,这可能由于青藏高原大多站点附近植被稀疏,基本为裸地或稀疏草地所致,而位置和地形信息的作用相当;2)对比植被信息、位置和地形自然要素两两组合的模型,增加 NDVI 的候选模型并没有显著的模拟提升. 其中,M7 为位置和地形的组合,在训练和检验站点上都有不错表现,而 M8 是植被、位置和地形的组合,在训练站点结果较好,但检验站点上模拟异常,可能由于有些站点的 NDVI 和其位置和地形之间有显著的关系,有些则不显著,使得模型结构不稳定. 因此,M7 被作为各自然因子组合模型的最优化模型.

最优组合 M7 不包括显式表达植被状况的 NDVI 项,这可能与 M7 中已经包括地表温度、地形因子有关. 由于地表温度本身与下垫面有关,地形因子与植被发生发育有关,植被状况等下垫面信息被隐含在这些变量里. ANFIS 包括一个自适应的过程,当一个变量与其余变量相关,且不能很好表达变化(variance)时,该变量甚至可能降低模拟结果,导致考虑 NDVI 的 M8 不如不考虑 NDVI 的 M7. 正是因为MODIS LST与0 cm土温间偏差的复杂

表 2 自然因子的数据特征

Table 2 Statistical characteristics of the natural parameters

变量	LST_day/℃	LST_night/℃	NDVI	经度	纬度	高程/km	坡向	坡度
最大值	51.51	20.25	0.93	102.02°	38.80°	5.06	350.54°	15.58°
最小值	−21.51	−30.91	0.049	80.08°	27.73°	1.42	2.28°	0.01°
均值	16.84	−4.98	0.24	94.33°	32.69°	3.67	164.07°	3.25°
标准方差	10.94	9.36	0.16	5.17°	3.29°	0.73	107.83°	3.92°

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

表 3　自然因子候选模型的整体误差和相关系数

Table 3　Statistical metrics of candidate ANFIS models consisting of natural impact factors

	ME	RMSE	CORR
M1	2.318(2.376)	3.039(3.037)	0.959(0.960)
M2	2.313(2.321)	2.986(2.979)	0.960(0.962)
M3	2.168(2.348)	2.854(3.023)	0.964(0.960)
M4	2.199(2.577)	2.892(3.277)	0.963(0.953)
M5	2.168(2.281)	2.823(2.915)	0.965(0.963)
M6	2.089(2.484)	2.703(3.208)	0.968(0.956)
M7	1.975(2.337)	2.588(2.973)	0.970(0.962)
M8	1.894(27.803)	2.380(46.009)	0.975(0.513)

注：ME 为绝对误差；RMSE 为均方根误差；CORR 为相关系数；括号外数值为训练统计结果，括号内为检验统计结果.

原因及伴随而来的不确定性，本文建议通过引进 Wavelet 来分解引起变化的各种组分，来弥补这种通过确定性因子无法充分表达变化的不足，然后通过 ANFIS 的模糊神经网络方法忽略变量间的物理机制，达到最优的模拟结果.

2.2　小波优化

试验表明，小波函数 rbio3.1 和分解层数 3 可做为小波优化的结果. 小波优化时，首先，固定小波分解层数来率定小波函数. 待率定的小波函数选自常用小波函数系，Haar 小波、Daubechies 小波(dbN)、SymletsA 小波(symN)、Biorthogonal 小波(biorNr.Nd)、Coiflet 小波组(coifN)、Morlet 小波和 Meyer 小波分别对 LST_day 和 LST_night 分解作为 ANFIS 输入，对模型结果进行对比分析，发现小波函数 bior1.1 和 rbio3.1 有不错的模拟结果，rbio3.1 则更胜一筹，所以本文选用小波函数 rbio3.1；然后，在率定得到小波函数 rbio3.1 的基础上，分析分解层数对结果的影响. 表4显示了采用 rbio3.1 小波函数分解不同层数 n 对应的模拟情况. 其中 $n=3$，4 时训练效果相当，而 $n=3$ 时检验效果明显好于 $n=4$，当 $n=5$ 时误差开始增大，相关系数下降.

表 4　不同 Wavelet-ANFIS 模型结果的整体误差和相关系数

Table 4　Statistical metrics of Wavelet-ANFIS models with different wavelet parameters

小波分解层数	ME	RMSE	CORR
$n=1$	1.869(1.917)	2.347(2.369)	0.976(0.976)
$n=2$	1.745(1.821)	2.173(2.236)	0.979(0.979)
$n=3$	1.699(1.783)	2.128(2.182)	0.980(0.980)
$n=4$	1.668(1.949)	2.120(2.413)	0.980(0.976)
$n=5$	1.731(2.314)	2.219(3.145)	0.978(0.959)

注：ME 为绝对误差；RMSE 为均方根误差；CORR 为相关系数；括号外数值为训练统计结果，括号内为检验统计结果.

小波优化结果与自然因子优化相比，小波函数 rbio3.1 和 $n=3$ 时基于小波分解的 ANFIS 模型估算结果明显好于自然因子，其 ME 为 1.699(检验为 1.783)，RMSE 为 2.128(检验为 2.182)，CORR 为 0.98(检验为 0.98)，全面好于 M7 的 1.975(2.337)、2.588(2.973)、0.970(0.962). 并且在训练和检验过程中均未出现异常，说明小波分解可以一定程度上弥补有限自然因子不能全面反映 0 cm 地温变化特征的问题，且有较为稳定的模拟表现.

2.3　最优模型及模拟结果

表 5 对比了时间因子，即卫星白天、晚上 2 次过境的当地时间(Time_Day、Time_Night)、当地昼夜长(PD，PN)和日出时间(T_sunrise)对模型的敏感性分析. 可以看到：1)昼夜长和日出时间较自然因子组合影响显著，从指标上看有较大的提高，其 ME 在 1.8～1.9，RMSE 在 2.3～2.5 之间(表 5). 昼夜长和日出时间对 ANFIS 模型模拟结果影响接近，考虑到昼夜长短和日出时刻有紧密关系，其结果相近则在情理之中；2)卫星过境的当地时刻作用不明显，ME、RMSE 及 CORR 与 M1 相当，说明利用 MODIS LST 在估算日均 0 cm 土温时，昼夜长或当地日出时间因素非常重要，而过境时刻因素可忽略. 原因可归结为气象站点逐日 0 cm 土温是 24 h 平均的结果，两过境时刻 LST 大致能代表全天高低温情况，致使昼夜长短对估算作用显著. 考虑到昼夜长有 2 个输入参数，而日出时刻只有 1 个，在计算复杂性上有优势. 因此，我们结合日出时刻与优化的自然因子组合和小波分解方法进一步形成候选的 4 个最优模型(表 6).

表 7 给出了参照模型 M1(只有 MODIS LST 两次观测作为输入)、优选的自然因子模型(M7)和优

表 5　时间因子对 ANFIS 模型模拟的敏感性指标

Table 5　Statistical metrics of ANFIS models in combination with different temporal factors

时间信息	ME	RMSE	CORR
昼夜长	1.864(1.875)	2.432(2.383)	0.974(0.976)
日出时刻	1.891(1.857)	2.506(2.368)	0.972(0.976)
卫星过境时刻	2.315(2.376)	3.026(3.037)	0.959(0.960)

注：ME 为绝对误差；RMSE 为均方根误差；CORR 为相关系数；括号外数值为训练统计结果，括号内为检验统计结果.

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

表 6 4 个候选最优模型

Table 6 Four most optimized candidate ANFIS models

ANFIS 模型	输入变量	输入变量数
M7	LST，位置，地形	7
O2	优化的小波分解(小波函数 rbio3.1、小波分解层 $n=3$)	8
O3	M7，日出时刻	8
O4	O2，日出时刻	9

表 7 候选最优 ANFIS 模型结果的误差和相关系数

Table 7 Statistical metrics of most optimized candidate ANFIS models

	ME	RMSE	CORR
M1	2.318(2.376)	3.039(3.037)	0.959(0.960)
M7	1.975(2.337)	2.588(2.973)	0.970(0.962)
O2	1.700(1.783)	2.128(2.182)	0.980(0.980)
O3	1.723(1.834)	2.208(2.294)	0.979(0.978)
O4	1.527(1.582)	1.926(1.957)	0.983(0.982)

注：ME 为绝对误差；RMSE 为均方根误差；CORR 为相关系数；括号外数值为训练统计结果，括号内为检验统计结果.

选的小波分解模型(O2)、以及 M7、O2 增加日出时刻的 2 个最优候选模型(O3、O4)的整体训练(括号外)和检验(括号内)结果的 ME、RMSE 和 CORR. 统计结果显示：1)O4 结果表现最好，其误差较之其他模型最小，ME 为 1.527(检验为 1.582)，RMSE 为 1.926(检验为 1.957)，CORR 为 98.3%(检验为 98.2%)，反映了时间信息结合优化的小波分解的 ANFIS 模型可以较高精度地估算 0 cm 土壤温度；2)优选的小波分解模型 O2 明显好于优选的自然因子模型 M7，结合时间信息的结果又优于无时间信息的模型，说明时间信息是重要的不可忽略的输入. 候选的 4 个模型相关系数都在 95%以上，说明全部候选模型都可以较好的反映 0 cm 土温的变化趋势.

图 6 反映了训练站点和检验站点的误差和相关系数情况，模型对站点的误差及相关系数情况和表 7 整体情况分析一致. 图 6 可明显看出模型之间的差异，O4 的 ME 和 RMSE 均低于其他模型结果，处于最低处，CORR 则处于最高处. 模型的优势及模型间的差别，尤其在站点误差波动明显较大、

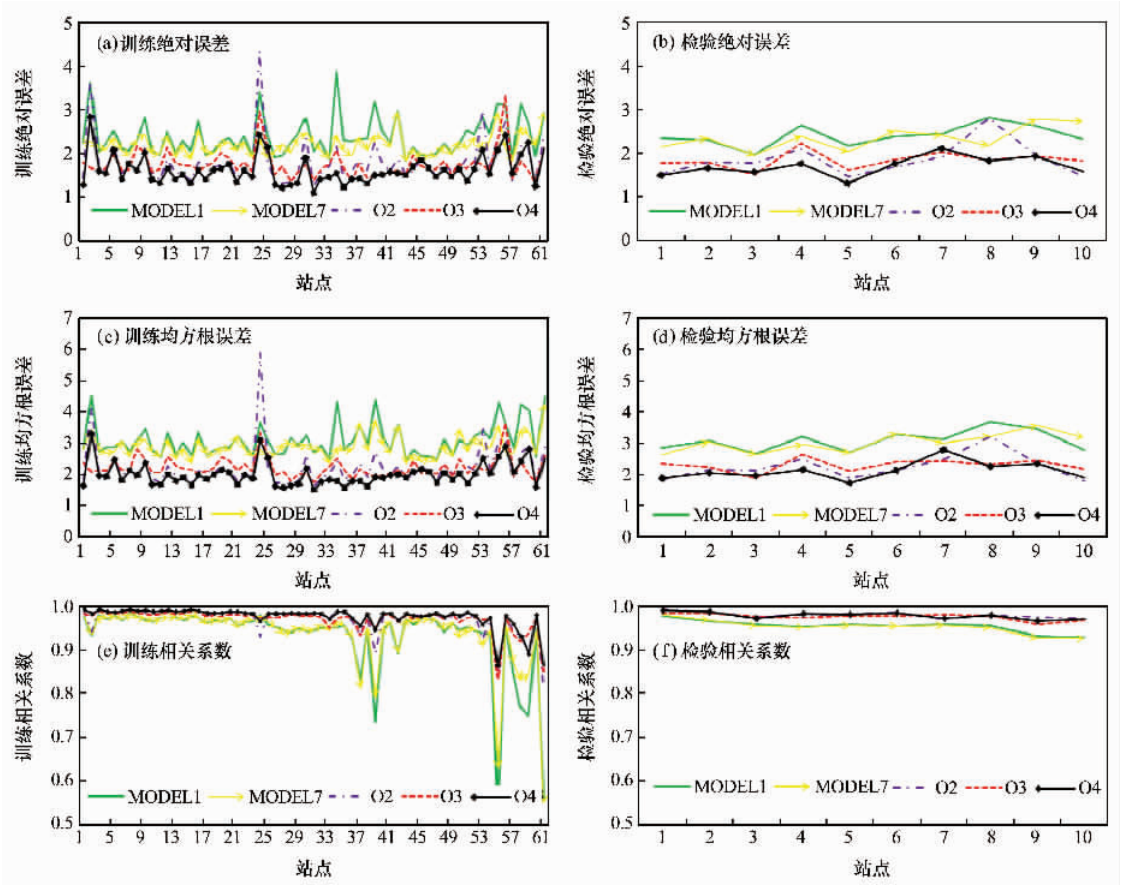


图 6 对照模型 M1 和 4 个候选最优 ANFIS 模型的绝对误差(ME)，均方根误差(RMSE)，相关系数(CORR)

Fig. 6 Evaluation metrics (ME, RMSE and CORR) of a control model (M1) and four most optimized candidate ANFIS models (M7, O2, O3, O4)

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

CORR 明显较低的站点中更容易表现出来．在这些站点上，O4 误差会显著降低，CORR 则相应的会有很大提高．检验站点对 ANFIS 结果检验中，趋势更为明显，模型间的差别容易被区分出，O4 在误差和相关系数方面依然表现最好．说明优化的小波分解和日出时刻作为输入的 ANFIS 模型能做到较精确的通过 LST 估算 0 cm 土温．

图 7 给出了表 6 中最优 ANFIS 候选模型逐日 0 cm 土温估算结果与实测值的散点图．整体上，模型自变量系数都在 0.9～1 之间，说明0 cm土温估算结果普遍稍低于实测值．由于样本数量足够多，线性关系表现的较为显著，复相关系数都在 90% 以上．

最优模型O4结果与实测逐日0 cm土温线性

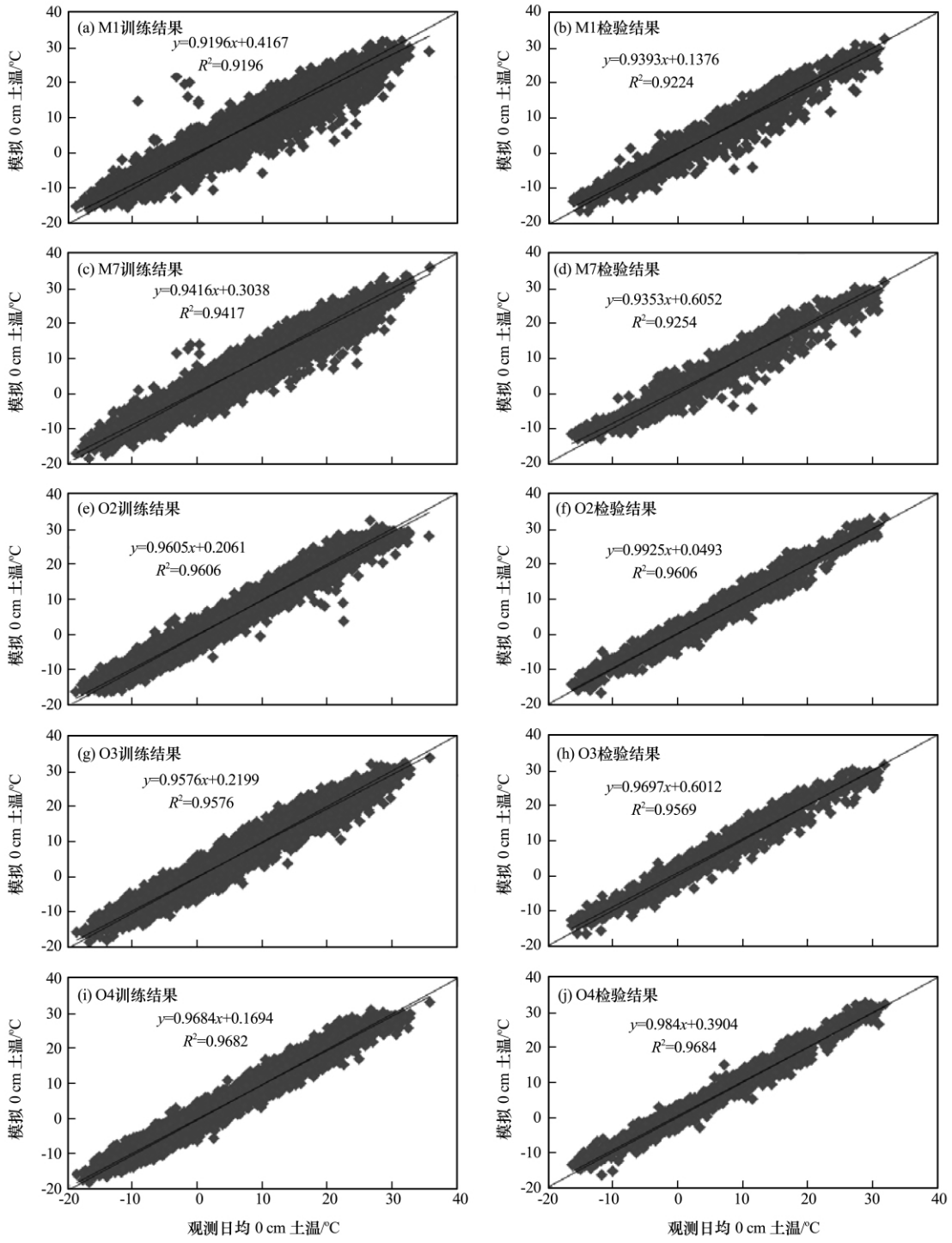


图 7　ANFIS 模型 0 cm 土温估算结果与实测日均 0 cm 土温散点图

Fig. 7　Scatterplots of daily 0-cm soil temperatures predicted with ANFIS models and measured at meteorological stations

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

关系表现的最为显著，训练和检验点集较之其他模型更加集中于对角线两侧，自变量系数达0.968(检验为0.984)，复相关系数为96.8%(检验为96.8%)，且出现的误差异常点相对较少，其趋势线和1∶1线几乎重合.

模型之间的差异与表7和图6中的结果分析类似. 散点图中异常点能在一定程度上说明模型间的优劣，如训练结果散点图中(左)的对角线左上方及右下方存在若干异常点，在M7中出现较多，而其他模型中则出现有所减少，说明自然因子的选取不够完美或者影响较小. 小波分析很好弥补这方面的缺陷，时间信息结合小波分析能够进一步提高精度. O2从一方面也说明了0 cm土壤温度是复杂自然因素综合影响的结果，小波分解进一步涵盖了更多的影响因子信息，从而使得效果比M7更好. 优化的小波分解组合时间信息使得O4表现最好，说明了通过优化的小波分解结合显著性的时间信息可以给ANFIS提供较好的参数，提高结果精度.

3 结论

本文对比建立了多种ANFIS模型利用MODIS LST及自然因子、小波分析来估算青藏高原日均0 cm土壤温度，优选出一种好的算法，即优化的小波分解组合率定的时间信息建立的ANFIS模型. 由于MODIS LST原始数据精度在2 K左右，本文ANFIS参数选择和结构优化后确定一种较好的模型使得0 cm土温估算结果稳定控制在2 K内. ANFIS估算结果结合站点实测0 cm土壤温度，进行了误差、相关系数和散点图分析，得出如下结论：

(1) 优化的小波分解组合率定后的时间信息建立的ANFIS模型估算0 cm土温较之其他模型表现最好. 误差降至1.527(检验为1.582)，相关系数98.3%(98.2%). 由于MODIS 2个波段的LST误差都在2 K内，影响0 cm土壤温度的因素具不确定性，可以认为在现有的条件下估算0 cm土壤温度精度是可以接受的. 在现有MODIS LST精度条件下，符合数据的使用要求.

(2) 自然因子对估算结果贡献不大，小波分析能够改善无法准确描述自然影响因素的局限. 卫星过境当地日出时刻或昼夜长对推求0 cm土温较之自然因子(下垫面植被、位置和地形信息)显得更为重要. 小波分析通过小波函数和分解层的率定，可以有效改善估算结果，可以弥补寻找自然因子中影响0 cm土壤温度环境信息的困难，其高频信息更多地反映了这方面信息的综合影响. 优化的小波分解和率定的时间信息结合可进一步提高模型的精度.

本文得到的最优模型可以扩展到青藏高原上没有观测站点的区域利用MODIS LST估算逐日0 cm土温，从而可以为许多冻土分布模型提供较好的上边界条件. 然而整个青藏高原上MODIS LST受系统误差、云、气溶胶等影响，缺失严重，下一研究阶段，需要根据青藏高原及MODIS LST的特点，结合多种卫星数据源发展更好的算法，或者发展一种好的插值方法以获取更高质量的LST，获得整个青藏高原面上的逐日1 km分辨率的LST.

参考文献(References)：

[1] Heginbottom J A. Permafrost mapping: a review [J]. Progress in Physical Geography, 2002, **26**(4): 623－642.

[2] Wu Qingbai, Zhu Yuanlin, Liu Yongzhi. Application of the permafrost table temperature and thermal offset forecast model in the Tibetan Plateau[J]. Journal of Glaciology and Geocryology, 2002, **24**(5): 614－617. [吴青柏，朱元林，刘永智. 青藏高原多年冻土顶板温度和温度位移预报模型的应用[J]. 冰川冻土，2002，**24**(5)：614－617.]

[3] Wang Zhixia. Applications of Permafrost Distribution Models on the Qinghai-Tibet Plateau [D]. Master Thesis, Lanzhou: Lanzhou University, 2010. [王之夏. 多年冻土分布模型在青藏高原的应用研究[D]. 硕士论文，兰州：兰州大学，2010.]

[4] Nelson F E, Outcalt S I. A computational method for prediction and regionalization of permafrost[J]. Arctic and Alpine Research, 1987, **19**(3): 279－288.

[5] Nan Zhuotong, Li Shuxun, Cheng Guodong, *et al*. Surface frost number model and its application to the Tibetan Plateau [J]. Journal of Glaciology and Geocryology, 2012, **34**(1): 89－95.[南卓铜，李述训，程国栋，等. 地面冻结数模型及其在青藏高原的应用[J]. 冰川冻土，2012，**34**(1)：89－95.]

[6] Zhang Wei, Wang Genxu, Zhou Jian, *et al*. Simulating the water-heat processes in permafrost regions in the Tibetan Plateau based on CoupModel[J]. Journal of Glaciology and Geocryology, 2012, **34**(5): 1099－1109. [张伟，王根绪，周剑，等. 基于CoupModel的青藏高原多年冻土区土壤水热过程模拟[J]. 冰川冻土，2012，**34**(5)：1099－1109.]

[7] Chen Hao, Nan Zhuotong, Wang Shugong, *et al*. Water-heat simulation on typical sites in upstream mountainous area of the Heihe River Basin[J]. Journal of Glaciology and Geocryology, 2013, **35**(1): 126－137. [陈浩，南卓铜，王书功，等. 黑河上游山区典型站点的水热过程模拟研究[J]. 冰川冻土，2013，**35**(1)：126－137.]

[8] Lunardini V J. Heat Transfer in Cold Climates [M]. New York: Van Nostrand Reinhold Co., 1981: 1－731.

[9] Li Shuxun, Wu Tonghua. The relationship between air temperature and ground temperature in the Tibetan Plateau[J]. Journal of Glaciology and Geocryology, 2005, **27**(5): 627－632. [李述训，吴通华. 青藏高原地气温度之间的关系[J]. 冰川冻土，2005，**27**(5)：627－632.]

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

[10] Cai Fu, Zhou Guangsheng, Li Rongping, *et al*. Sensitivity of land surfaces model to dynamic land surface parameters [J]. Advances in Earth Science, 2011, **26**(3): 300－310. [**蔡福，周广胜，李荣平，等. 陆面过程模型对下垫面参数动态变化的敏感性分析**[J]. **地球科学进展**, 2011, **26**(3): 300－310.]

[11] Wang Zhixia, Nan Zhuotong, Zhao Lin. The applicability of MODIS land surface temperature products to simulating the permafrost distribution over the Tibetan Plateau [J]. Journal of Glaciology and Geocryology, 2011, **33**(1): 132－143. [**王之夏，南卓铜，赵林. MODIS 地表温度产品在青藏高原冻土模拟中的适用性评价**[J]. **冰川冻土**, 2011, **33**(1): 132－143.]

[12] Ouyang Bin, Che Tao, Dai Liyun, *et al*. Estimating mean daily surface temperature over the Tibetan Plateau based on MODIS LST products [J]. Journal of Glaciology and Geocryology, 2012, **34**(2): 296－303. [**欧阳斌，车涛，戴礼云，等. 基于** MODIS LST **产品估算青藏高原地区的日平均地表温度**[J]. **冰川冻土**, 2012, **34**(2): 296－303.]

[13] Jang J-S R. ANFIS: Adaptive-network-based fuzzy inference system [J]. IEEE Transactions on Systems, Man and Cybernetics, 1993, **23**(3): 665－685.

[14] Firat M, Güngör M. Hydrological time-series modeling using an adaptive neuro-fuzzy inference system [J]. Hydrological Processes, 2008, **22**(13): 2122－2132.

[15] Lo S-P. The application of an ANFIS and grey system method in turning tool-failure detection [J]. The International Journal of Advanced Manufacturing Technology, 2002, **19**(8): 564－572.

[16] Williams R J, Zipser D. A learning algorithm for continually running fully recurrent neural networks [J]. Neural computation, 1989, **1**(2): 270－280.

[17] Nelles O. Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models [M]. Berlin Heidelberg: Springer, 2001.

[18] Takagi T, Sugeno M. Fuzzy identification of system and its applications to modeling and control [J]. IEEE Transactions on Systems, Man, and Cybernetics, 1985, **15**(1): 116－132.

[19] Sugeno M, Kang G T. Structure identification of fuzzy model [J]. Fuzzy Sets and Systems, 1988, **28**(1): 15－33.

[20] Anderberg M R. Cluster Analysis for Applications [M]. New York: Academic Press, 1973: 1－359.

[21] Mallat S G. A theory for multiresolution signal decomposition: The wavelet representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, **11**(7): 674－693.

[22] Burrus C S, Gopinath R A, Guo H. Introduction to Wavelets and Wavelet Transforms: A Primer [M]. Upper Saddle River, New Jersey: Prentice Hall, 1998: 1－268.

Estimation of 0-cm Soil Temperature over the Tibetan Plateau Based on the Wavelet Analysis and Adaptive Network-fuzzy Inference System

HUANG Pei-pei[1], NAN Zhuo-tong[1, 2]

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

Abstract: 0-cm soil temperature is the upper boundary condition of many permafrost models. Continuous, high-quality 0-cm soil temperature data are necessary inputs to simulate permafrost distribution. However, owing to the influence of complex underlying surface, remote sensing approaches cannot provide reliable 0-cm soil temperature. In this study, in order to estimate 0-cm soil temperature, adaptive network-fuzzy inference system (ANFIS) combining with the data measured in the Tibetan Plateau is used to establish the relations between remote sensing land surface temperature (LST) and 0-cm soil temperature. In this paper, different parameter combinations of ANFIS are examined, and a Wavelet-ANFIS model established by optimized wavelet functions, wavelet windows and wavelet layers is found able to estimate the 0-cm soil temperature more accurately. A comparison analysis of the estimated results and the 0-cm soil temperatures measured at the meteorological sites shows that the approach can achieve desirable estimation with an absolute error less than 2 K and a correlation coefficient greater than 0.98. In view of the original MODIS LST error range from 0 to 2 K, the proposed method may provide more accurate 0-m soil temperature inputs to permafrost models.

Key words: wavelet analysis; adaptive network-fuzzy inference system (ANFIS); MODIS land surface temperature product; Tibetan Plateau; 0-cm soil temperature

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net