

用 MODIS 数据反演近地表空气温度的 RM-NN 算法^①

毛克彪^{②***} 马莹* 夏浪* 沈心一***

(* 中国农业科学院农业资源与农业区划研究所/呼伦贝尔草原生态系统国家野外科学观测研究站/
农业部农业信息技术重点实验室,北京 100081)

(** 陕西师范大学国土资源研究中心 西安 710062)

(*** 俄克拉荷马大学俄克拉荷马水文气象学与遥感实验室 诺曼 73072,美国)

摘要 基于辐射传输模型(RM)和动态学习神经网络(NN),成功进行了用 MODIS 数据反演近地表空气温度的研究,并给出了完成这种反演的 RM-NN 算法。该算法用 RM 来模拟不同地面辐射状况下(包括不同的地表温度、近地表空气温度、发射率和大气水汽含量)卫星高度获得的辐射强度数据集,用动态学习神经网络来进行反演计算。反演分析结果表明,近地表空气温度不能直接精确地用 MODIS 数据反演计算得到,如果能把地表温度和发射率以及大气水汽含量作为先验知识,则能够比较精确地反演近地表空气温度。模拟分析表明,平均误差和标准偏差分别大约是 0.8K 和 0.9K,如果考虑地表温度和发射率的误差,平均误差和标准偏差分别为 1.5K 和 1.8K。反演结果和地面气象站点数据比较表明,合理地利用先验知识使得 RM-NN 算法能够用 MODIS 数据比较精确地反演近地表空气温度。

关键词 热红外遥感,神经网络(NN),MODTRAN,中分辨率成像光谱仪(MODIS),近地表空气温度

0 引言

近地表空气温度对地表和空气之间的水分和能量交换有重要影响,因此近地表空气温度是气候变化研究的一个非常重要的参数^[1]。由于近地表空气温度受时间、空间及地表情况的影响,因而至今还没有一种能够很好地估计近地表空气温度的空间分布的方法^[2]。目前气候变化研究通常采用两种方法获得近地表空气温度。一是基于能量平衡的物理方法,该方法需要空气动力学阻抗及地表状态参数,包括水、土壤和植被状态等,这几个参数非常难以获取^[1];另一个是经验方法^[3-7],该方法则是利用地理信息系统(GIS)气象站点获得的近地表空气温度进行插值得到近地表空气温度的分布图。当气象站点不很多且不均匀分布(特别是在山区)时,插值得到的结果不很理想。但由于各种条件的限制,目前普遍采用的是经验方法^[8]。

1999 年、2002 年搭载中分辨率成像光谱仪(MODIS)的对地观测卫星发射成功,为全球和区域资源环境动态监测开辟了又一新的途径。MODIS 是一个拥有 36 个波段的中分辨率遥感系统,每一到两天可获得一次全球观测数据,其飞行与太阳同步,每天同一区域至少可获得昼夜两景图像,并且是免费接收,因此非常适合于中大尺度的区域资源环境动态监测。在 MODIS 的 36 个波段中有 8 个是热红外波段,因而非常合适于区域尺度的地表热量空间差异分析。目前利用 MODIS 遥感数据反演地表温度的算法已有不少^[9-14],但利用 MODIS 数据反演近地表空气温度的算法还没有。为了更准确地分析区域热量空间差异和进行蒸散发研究,本文提出了一种基于辐射传输模型(radiance transfer model, RM) MODTRAN4 和动态学习神经网络(neural network, NN),利用 MODIS 数据反演近地表空气温度的算法,以下简称为 RM-NN 算法。

① 973 计划(2010CB951503),国家科技支撑计划(2013BAC03B00)和 863 计划(2012AA120905)资助项目。

② 男,1977 年生,博士,副研究员;研究方向:农业遥感和气候变化;联系人,E-mail: maokebiao@126.com
(收稿日期:2009-11-30)

1 RM-NN 算法反演近地表空气温度分析

陆地表面温度和发射率的反演是基于通用的热辐射传输方程^[10]

$$B_i(T_i) = \varepsilon_i(\theta)\tau_i(\theta)B_i(T_s) + [1 - \tau_i(\theta)][1 - \varepsilon_i(\theta)]\tau_i(\theta)B_i(T_{ia}) + [1 - \tau_i(\theta)]B_i(T_{ia}) \quad (1)$$

进行的。式中 T_s 表示地表温度, T_i 表示通道 i 在传感器高度获得的星上亮度温度, $\tau_i(\theta)$ 表示通道 i 在观测方向 θ 的大气透过率, θ' 表示大气向下亮温辐射的方向, $\varepsilon_i(\theta)$ 表示通道 i 在观测方向 θ 处的地表发射率。 $B_i(T_i)$ 是传感器接收到的辐射强度, $B_i(T_s)$ 是地表的辐射强度, T_{ia} 表示通道 i 的有效大气平均作用温度。有效大气平均作用温度 (T_{ia}) 是随波长变化而变化的, 它主要由大气水汽含量和近地表空气温度

$$T_{ia} = A_i + B_i T_0 \quad (2)$$

决定^[2,11]。式中 T_0 为 2m 高左右的近地表空气温度, A_i 是常数, B_i 是通道 i 系数。近地表空气温度还受地表温度的影响。在一个给定的地点, 近地表空气温度和地表温度也存在像式(2)一样的关系, 但这个关系不是很稳定, 它随时间和地点变化而变化。在式(1)中, 有 3 个未知数(发射率、地表温度和近地表空气温度), 这是一个典型的病态问题。如果不构造其它条件, 方程组无解。另外, 每个热红外波段的透过率 ($\tau(\theta)$) 也是一个未知数, 它是大气水汽含量和其它气体的函数, 如式

$$\tau_i(\theta) = f(WVC, O) \quad (3)$$

所示。式中 WVC 是大气水汽含量, O 表示其它气体(二氧化碳、一氧化氮、臭氧、甲烷、一氧化碳等), 这些气体相对大气水汽含量是稳定的, 其影响可以通过标准大气剖面模拟得到。热红外波段透过率对水汽非常敏感, 劈窗算法就是利用两个热红外波段对水汽的不同敏感性来消除水汽的影响的, 从而通过反演计算可得到地表温度。对于不同的地物在不同的波段, 发射率几乎是一个常数, Mao 等^[11,15] 提出利用邻近波段发射率之间局部线性关系减少未知数来克服病态问题, 可描述成下式:

$$\varepsilon_i(\theta) = C_i + D_i \varepsilon_j(\theta) \quad (4)$$

式中 $\varepsilon_i(\theta)$ 和 $\varepsilon_j(\theta)$ 是不同波段 (i, j) 在观测角度 θ 时的发射率。 C_i 是常数, D_i 是通道 i 的系数。对于

同一种地物类型, 不同波段的发射率可以用一个波段的发射率表示, 从而将不同波段的发射率减少为 1 个。由于很难用几个函数精确地描述所有的波段, 这个潜在的信息没有被充分利用^[11,15]。另外非线性函数(比如普朗克函数)的线性简化也会产生误差。

传统的反演算法需要花费很多时间进行推算^[9,16]。相对于传统算法, 神经网络(NN)不需要知道输入和输出参数之间的关系, 可以通过训练数据集直接确定输入和输入数据之间的关系^[17], 所以神经网络常被用来解病态反演问题^[11,15]。

2 RM-NN 模拟反演分析

为了训练和测试神经网络, MODTRAN^[18] 被用来产生训练和测试数据集。由于神经网络的精度取决于训练和数据集, 因此建立模拟数据集的过程中, 必须严格按照每种地物的波谱特征(URL: <http://speclib.jpl.nasa.gov>)。在不同的地区(不同的经纬度和高度)的大气剖面是不同的, 为了提高算法精度和实用性, 需要在不同的地区建立训练和测试数据集。RM-NN 算法的优点是能克服测量误差和精确地保持各物理参数之间的关系。MODTRAN 模拟数据集可以当作真实地表的参考数据, 当然可靠地面测试数据集也可以作为补充数据源。

根据光谱光线, 详细的介绍可以参见光谱数据库(URL: <http://speclib.jpl.nasa.gov>)。每种地物的每个波段发射率被当作输入参数输入 MODTRAN4。地表温度的设定范围为 280 ~ 320K, 相应的近地表空气温度(2m 高)被设置成 280 ~ 320 K。大气水汽含量分布范围是 0.2 ~ 4g/cm²。17km 以下水汽剖面和大气温度的设置按照标准大气压变化规则变化。热红外波段主要受大气水汽影响, 所以其它参数按照标准大气压设置为常数。

在本研究中, 动态学习神经网络^[19] 被用来反演近地表空气温度。动态学习神经网络使用了卡曼滤波算法来增加算法学习的收敛速度和提高解非线性边界问题的能力。对于动态学习神经网络的介绍请参见文献[19]。操作步骤如下:

(1) 将模拟数据随机地分成两组: 训练数据集 2215 和测试数据集 1154。

(2) 训练和测试神经网络。

输入节点是 MODIS 的第 29, 31, 32 波段和大气水汽含量, 输出节点是近地表空气温度。在反复测

试和训练后,隐含层两层,并且每层节点数为 30 时, 结果最好。详细的测试数据集信息见表 1。

表 1 反演误差表

| 隐含层 | T_0 | | 隐含节点 | T_0' | | 隐含节点 | T_0'' | |
|-------|-------|-----|-------|--------|-----|-------|---------|-----|
| | r | SD | | r | SD | | r | SD |
| 10-10 | 0.946 | 3.1 | 5-5 | 0.981 | 1.1 | 10-10 | 0.964 | 2.5 |
| 20-20 | 0.946 | 3.1 | 10-10 | 0.981 | 1.1 | 20-20 | 0.968 | 2.4 |
| 30-30 | 0.947 | 3.1 | 15-15 | 0.982 | 1.1 | 30-30 | 0.969 | 2.2 |
| 40-40 | 0.945 | 3.3 | 20-20 | 0.986 | 1 | 40-40 | 0.978 | 2 |
| 50-50 | 0.946 | 3.1 | 25-25 | 0.989 | 0.9 | 50-50 | 0.982 | 1.8 |
| 60-60 | 0.947 | 3.1 | 30-30 | 0.988 | 1 | 60-60 | 0.975 | 2.1 |
| 70-70 | 0.947 | 3.1 | 35-35 | 0.982 | 1.1 | 70-70 | 0.975 | 2.1 |
| 80-80 | 0.948 | 3.0 | 40-40 | 0.981 | 1.2 | 80-80 | 0.974 | 2.2 |

R : 相关系数; SD : 标准偏差. T_0, T_0', T_0'' 为近地表空气温度

从表 1 中第一栏分析可知,近地表空气温度估算精度不是非常高,平均值和标准偏差分别大约是 2.9K 和 3.1K。星上的亮度温度对近地表空气温度不是非常敏感,为了提高反演精度,我们将地表温度和地表发射率作为先验知识,从而可以消除更多的不确定性。详细过程如下:

(1) 从 MODIS 地表温度和发射率产品、大气水汽含量产品读取每一个像元的地表温度、发射率和大气水汽含量,将它们作为 MODTRAN4 的输入参数。举个例子,比如一个像元的地表温度是 300K,波段 29,31 和 32 的发射率分别是 0.95、0.96、0.98,大气水汽含量是 $1.2\text{g}/\text{cm}^2$,将这些产品值作为 MODTRAN 的已知输入参数,近地表空气温度变化范围是 300 K 到 315 K,采用是中纬度大气剖面,模拟过程如图 1 所示。

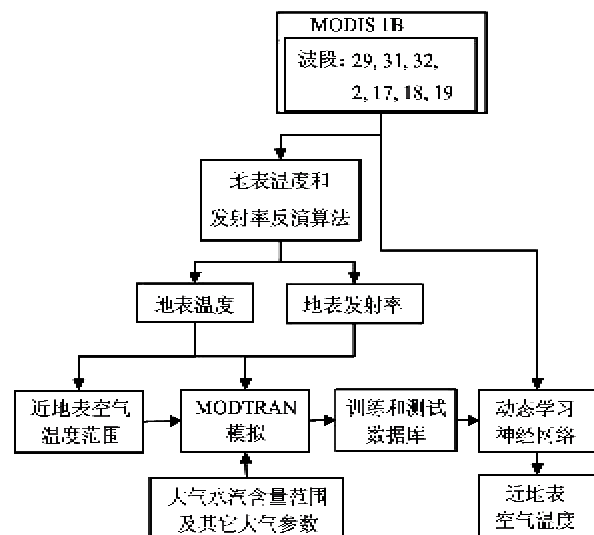


图 1 RM-NN 算法反演近地表空气温度流程

(2) 模拟数据分成训练和测试两部分:训练数据集 89 和测试数据集 51。在反复训练和测试后,近地表空气温度反演结果如表 1 中 T_0' 所示。从表 1 中可以看出,反演结果非常好,平均误差和标准误差分别大约为 0.8K 和 0.9K。精度提高的主要原因是地表温度和发射率以及大气水汽含量作为先验知识。

(3) 地表温度和发射率的反演误差分别大约为 $\pm 1\text{K}$ 和 $\pm 0.015^{[20]}$,大气水汽含量的误差为 $\pm 13\%$ 。这里我们考虑地表温度的反演误差在 $\pm 2\text{K}$ 和发射率误差为 ± 0.03 。对于每个像元, $LST - 2\text{K} \leq LST \leq LST + 2\text{K}$, $\varepsilon_i - 0.03 \leq \varepsilon_i \leq \varepsilon_i + 0.03$, $w - 0.13w \leq w \leq w + 0.13w$,作为每个像元的先验知识输入到 MODTRAN4 中进行模拟并建立训练和测试数据。比如一个像元的地表温度为 300K,波段 29,31 和 32 的发射率分别为 0.95、0.96、0.98,大气水汽含量是 $1.2\text{g}/\text{cm}^2$ 。地表温度(297 ~ 303K),波段 29,31,32 的发射率 0.92 ~ 0.97, 0.93 ~ 0.99, 0.96 ~ 1,大气水汽含量 $1 \sim 1.5\text{g}/\text{cm}^2$ 作为先验知识输入到 MODTRAN4 中。训练数据集为 836,测试数据集为 392。在反复训练和测试后,近地表空气温度如表 1 T_0'' 所示。平均误差和标准偏差分别为 1.5K 和 1.8K,这个在目前的应用中还是能满足要求的。

3 验证比较分析

相对于地表温度,在卫星过境时测量近地表空气温度要容易一些。近地表空气温度受地表温度、土壤湿度、风和地形的影响,所以邻近不同的地方近地表空气温度是不一样的。因此在选择验证区域时

候需要选择地形比较平坦,而且地物比较均一的地表。我们选择了两个气象站点不同时期的气象测量数据作为验证数据^[11]。RM-NN 算法反演结果和地表测量结果比较如图 2 所示,平均误差

$$\left(\frac{\sum_{i=1}^n |T_{0r} - T_0|}{n} \right) \text{ 大约是 } 1.6\text{K}。 \text{ 影响地面近地表}$$

空气温度测量的因素很多,特别是对于混合像元,更多野外测量和比较将会进一步报道。RM-NN 算法的另外一个优点是可以通过补充可靠的训练数据集,从而提高反演精度和实用性。本研究的目的是证明 RM-NN 算法用 MODIS 1B 数据反演近地表空气温度的可行性,将来需要做更多的分析和应用。

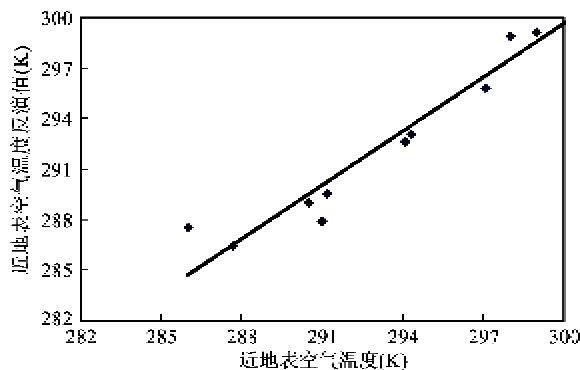


图 2 验证结果

4 结论

本研究对用遥感数据 MODIS 1B 直接反演近地表空气温度的病态问题做了一些分析,提出充分利用地表温度和发射率以及大气水汽含量作为先验知识是可以直接用遥感数据反演近地表空气温度的。大气辐射传输模型 MODTRAN4 被用来产生模拟数据,自动学习神经网络被用来反演近地表空气温度。该方法的优点是能精确地保持地球物理参数(比如不同波段之间的光谱曲线等)之间的关系。反演结果分析表明不能直接用热红外遥感数据反演近地表空气温度。但如果利用地表温度和发射率以及大气水汽含量作为先验知识,则可以精确地反演近地表空气温度,反演的平均误差大约为 1 K。如果考虑地表温度反演误差($\pm 2\text{K}$)和发射率误差(± 0.03)以及大气水汽含量($\pm 13\%$),近地表空气温度的反演平均误差大约是 1.5K。为了进一步提高 RM-NN 算法的反演精度和实用性,可以进一步根据当地气象站点补充可靠的训练数据集。

参考文献

- [1] Sun Y J, Wang J F, Zhang R H, et al. Air temperature retrieval from remote sensing data based on thermodynamics. *Theoretical and Applied Climatology*, 2005, 80: 37-48
- [2] Mao K B, Tang H J, Wang X F, et al. Near-surface air temperature estimation from ASTER data using neural network. *International Journal of Remote Sensing*, 2008, 29 (20): 6021-6028
- [3] Boyer D G. Estimation of daily temperature means using elevation and latitude in mountainous terrain. *Water Resource Bulletin*, 1984, 4: 583-588
- [4] Hutchinson M F. A new objective method for spatial interpolation of meteorological variables from irregular networks applied to the estimation of monthly mean solar radiation, temperature, precipitation and windrun. Technical Report 89-5, Division of Water Resources, the Commonwealth Scientific and Industrial Research Organization (CSIRO), Canberra, Australia, 1989. 95-104
- [5] Hutchinson M F, Corbett J D. Spatial interpolation of climate data using thin plate smoothing splines. In: Proceedings of the Coordination and Harmonization of Databases and Software for Agroclimatic Applications, Proceedings of an Expert Consultation, Rome, Italy, 1995. 211-224
- [6] Ishida T, Kawashima S. Use of cokriging to estimate surface air temperature from elevation. *Theoretical and Applied Climatology*, 1993, 47:147-157
- [7] Russo J M, Liebhold A M, Kelly J G W. Mesoscale weather data as input to a gypsy moth (*Lepidoptera: Lymantriidae*) phenology model. *Journal of Economic Entomology*, 1993, 86: 838-844
- [8] Burrough P A, Mc Donnell R A. Principles of Geographical Information Systems. New York: Oxford University Press, 1998. 30-50
- [9] Wan Z M, Li Z L. A physics-based algorithm for retrieving land-surface emissivity and temperature from EOS/MODIS data. *IEEE Transactions on Geoscience and Remote Sensing*, 1997, 35(4): 980-996
- [10] Mao K, Qin Z, Shi J, et al. A practical split-window algorithm for retrieving land surface temperature from MODIS data. *International Journal of Remote Sensing*, 2005, 26 (15): 3181-3204
- [11] Mao K, Shi J, Li Z L, et al. An RM-NN algorithm for retrieving land surface temperature and emissivity from EOS/MODIS data. *Journal of Geophysical Research-atmosphere*, 2007, D112: 1-17
- [12] 毛克彪,覃志豪,施建成等. 针对 MODIS 数据的劈窗

- 算法研究. 武汉大学学报(信息科学版),2005(8): 703-708
- [13] 毛克彪,唐华俊,周清波等. 用辐射传输方程从 MODIS 数据中反演地表温度的方法. 兰州大学学报(自然科学版),2007,43(4):12-17
- [14] 毛克彪,唐华俊,李丽英等. 一个从 MODIS 数据同时反演地表温度和发射率的神经网络算法. 遥感信息, 2007,92(4): 9-15
- [15] Mao K, Shi J, Tang H, et al. A neural network technique for separating land surface emissivity and temperature from ASTER imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 2008, 46(1): 200-208
- [16] Li Z L, Becker F. Feasibility of land surface temperature and emissivity determination from AVHRR data. *Remote Sensing of Environment*, 1993, 43: 67-85
- [17] Fung A K. Microwave Scattering and Emission Models and Their Application. B. L: *Artech House*, 1994. 527-555
- [18] Berk A, Bemstein L S, Roberttson D C. MODTRAN: A modetate resolution model for LOWTRAN. Burlington, MA, Spectral Science, Inc., Rep. AFGL-TR-87-0220. 1987. 1-40
- [19] Tzeng Y C, Chen K S, Kao W L, et al. A Dynamic learning neural network for remote sensing applications. *IEEE Transactions on Geoscience and Remote Sensing*, 1994, 32(5): 1096-1102
- [20] Wan Z, Zhang Y, Zhang Y Q, et al. Validation of the land surface temperature products retrieved from moderate resolution imaging spectroradiometer data. *Remote Sensing of Environment*, 2002, 83:163-180

An RM-NN algorithm for retrieval of near-surface air temperature retrieval from MODIS data

Mao Kebiao^{***}, Ma Ying^{*}, Xia Lang^{*}, Shen Xinyi^{***}

(^{*} National Hulunber Grassland Ecosystem Observation and Research Station / Key Laboratory of Agri-informatics, Ministry of Agriculture / Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081)

(^{**} Northwest Land and Resources Research Center, Shanxi Normal University, Xi'an 710062)

(^{***} Hydrometeorology and Remote Sensing Laboratory, The University of Oklahoma, Norman 73072, USA)

Abstract

An algorithm for retrieving near-surface air temperature from MODIS 1B data based on radiance transfer model (RM) and a dynamic learning neural network (NN) was developed. The algorithm, called the RM-NN algorithm for short, uses the RM to simulate the radiance transfer from ground with different combinations of land surface temperature, near surface air temperature, emissivity and water vapor content. The dynamic learning neural network is used to estimate near surface air temperature. The analysis indicates that near surface air temperature can not be directly and accurately estimated from thermal remote sensing data. If the land surface temperature and emissivity and water vapor content are made as prior knowledge, the mean and the standard deviation of estimation error are about 0.8K and 0.9K. The mean and the standard deviation of estimation error are about 1.5K and 1.8K under considering the estimation error of land surface temperature and emissivity. Finally, the comparison of estimation results with ground measurement data at meteorological stations indicates that the RM-NN algorithm can be used to estimate near surface air temperature from MODIS 1B data.

Key words: Thermal remote sensing, neural network, MODTRAN, MODIS, near surface air temperature