

文章编号: 1008-2786-(2020)6-873-08

DOI:10.16089/j.cnki.1008-2786.000563

基于遥感数据和随机森林算法的黄土高原地区 气温模拟及时空变化

邢立亭, 李 净*

(西北师范大学 地理与环境科学学院, 甘肃 兰州 730070)

摘 要: 气温是影响区域热环境的重要表征和引起气候变化的关键因素, 依靠少量气象站点的内插或者外推获得的空间连续分布的气温无法准确表示大区域气温的空间变化特征。本文利用遥感数据结合随机森林算法模拟了 2003—2016 年黄土高原地区长时空序列的气温数据, 并分析了 11 个输入变量的重要性, 验证了随机森林模型的性能。结果表明: 使用随机森林算法结合遥感数据模拟得到的近地表气温的平均绝对误差为 0.91°C , 均方根误差为 1.06°C , 结果精度较高。根据模拟出来的近地表气温分析了黄土高原地区 2003—2016 年气温的时空变化特征及趋势, 发现 2003—2016 年黄土高原大部分区域气温整体上呈现缓慢上升的趋势。结果表明使用遥感数据结合随机森林算法的气温模拟在黄土高原地区具有很好的适用性, 对研究黄土高原地区的气温演变规律及水热变化具有重要意义。

关键词: MODIS; 随机森林算法; 近地表气温; 黄土高原

中图分类号: P407.8

文献标志码: A

近地表气温是地面气象站测定的常规要素之一, 是反映一个地区气候特征的重要因子, 也是各种气象、水文模型的重要输入数据。由于不同区域气象站点的疏密程度不同, 插值的方法无法准确得到大区域气温的空间分布实情。遥感数据能够提供大范围的空间分布信息, 例如地表温度、植被指数等, 这些数据也被广泛应用于模拟近地表气温的研究当中^[1-4]。

近年来国内外学者开始利用遥感数据模拟近地表气温, 比较成熟的气温模拟方法可以总结为 4 种^[5-7]: (1) 简单统计法, 例如, Davis^[8] 等使用一元线性回归模型, 以北美地区为研究区, 计算得出的误

差为 $1.6^{\circ}\text{C} \sim 2.6^{\circ}\text{C}$ 。(2) 高级统计法, 在简单统计法的基础上, 使用多个影响因子与气温建立模型。例如, Cresswell^[9] 等通过建立与气温的多元回归模型, 计算得出的气温值与观测值误差在 $0.09^{\circ}\text{C} \sim 1.69^{\circ}\text{C}$ 之间。(3) TVX 法, 例如, Stisen^[10-12] 等使用正弦函数插值结合 TVX 法对近地表气温进行模拟, 模拟得出的误差在 $2.55^{\circ}\text{C} \sim 2.99^{\circ}\text{C}$ 之间。(4) 能量平衡方法, 例如, Pape^[13] 等通过建立地表能量平衡方程, 计算得出的气温值与观测值误差 $0.37^{\circ}\text{C} \sim 1.02^{\circ}\text{C}$ 之间。上述 4 种方法虽可以获得空间连续分布的气温数据, 但在应用上具有一定的局限性, 气温模拟结果精度较差, 无法获取精细化的气温空间

收稿日期 (Received date): 2020-05-15; 改回日期 (Accepted date): 2020-11-09

基金项目 (Foundation item): 国家自然科学基金项目 (41861013; 41761083; 41561016)。[National Natural Science Foundation of China (41861013; 41761083; 41561016)]

作者简介 (Biography): 邢立亭 (1996-), 男, 山东济南人, 硕士研究生, 主要研究方向: 定量遥感与气温模拟。[XING Liting (1996-), male, born in Jinan, Shandong Province, M.Sc. candidate, research on quantitative remote sensing and temperature simulation] E-mail: 15117290383@163.com

* 通讯作者 (Corresponding author): 李净 (1978-), 女, 甘肃会宁人, 博士, 主要研究方向: 定量遥感与辐射模拟。[LI Jing (1978-), female, born in Huining, Gansu Province, Ph.D., engaged in quantitative remote sensing and radiation simulation] E-mail: li_jinger@nwnu.edu.cn

分布数据。

近年来很多国内外学者基于随机森林算法结合不同的遥感数据进行了近地表气温的模拟研究,并取得了不错的效果^[14-17]。在这些研究中,大多只用了卫星白天所获取的地表温度数据,无法体现气温在日内的变化。MODIS 传感器一天内可以获取 4 次地表温度数据(2 次白天和 2 次夜晚),由于夜晚地面辐射和大气的保温作用,卫星所获取的夜晚地表温度数据对气温的影响同样不可忽视,相较于使用卫星白天所获取的数据,选取 4 个时间点的地表温度来进行近地表气温的研究可以确保模拟结果的准确性,避免偶然性。由于黄土高原地区气象站点稀少,故本文以黄土高原为研究区,使用随机森林算法结合 MODIS 白天和晚上四个时间点的地表温度数据及其他遥感数据模拟了 2003—2016 年的近地表气温,分析气温的时空变化特征,本研究将为模拟精细化空间分布的近地表气温提供了思路,为黄土高原地区的气候变化情况提供依据。

1 数据来源

1.1 研究区概况

黄土高原处于我国半湿润区,半干旱区以及干旱区的交界位置,复杂的气候造就了区域内复杂的地形条件。由于气象台站在地形复杂地区较难布设,故研究区内气象站点稀少且分布不均,黄土高原涵盖了整个山西省、宁夏回族自治区以及青海省、甘肃省、陕西省、河南省、内蒙古自治区的部分区域,总面积为 64 万 km²,海拔为 800~4874 m(图 1)。

1.2 数据来源

本文使用的 MODIS 地表温度数据有白天 10:30 和晚上 10:30 获取的 MOD11A2 以及在凌晨 1:30 和下午 1:30 获取的 MYD11A2。所获取的 MODIS 地表温度数据为空间分辨率为 1000 m 的 8 d 合成产品,研究区包含 4 个瓦片,收集了 2003—2016 年每年的数据,共计 5152 景;使用的归一化植被指数数据分辨率为 1000 m 的月合成产品 MOD13A3,共计 672 景。所有 MODIS 数据均来源于 NASA 官网。由于研究需要使用逐月的地表温度数据,使用 ArcGIS 像元统计中的最大值合成法将时间分辨率为 8 d 的地表温度数据合成为月值数据。DEM 数据空间分辨率为 30 m,来源于地理空间数据云平台,并由

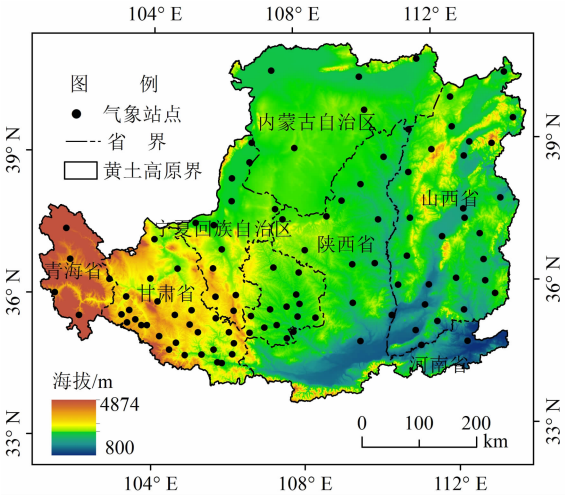


图 1 研究区位置

Fig. 1 Location of study area

DEM 数据计算得到坡度、坡向数据;太阳辐射数据来自 AVHRR 卫星获取的月数据^[2]。气象数据来源于国家气象科学数据共享服务平台下载的研究区内共计 104 个气象站点的月均温数据。文中所用到的自变量和辅助变量如表 1 所示。

表 1 自变量和辅助变量描述

Tab. 1 Description of independent variables and auxiliary variables

输入变量	简称	描述
自变量 a	LSTAN	Aqua 星凌晨 1:30 获取的地表温度
	LSTTD	Terra 星上午 10:30 获取的地表温度
	LSTAD	Aqua 星下午 1:30 获取的地表温度
	LSTTN	Terra 星晚上 10:30 获取地表温度
辅助变量 b	Lon	经度
	Lat	纬度
	DEM	海拔
	Asp	坡向
	Slope	坡度
	NDVI	归一化植被指数
	Rad	太阳辐射

2 研究方法

随机森林算法是一种近来比较流行的机器学习模型,2001 年 Culter Adele 和 LeoBreiman 开发了这种数据算法,这种算法对非线性数据有着很好的拟

合效果,能很好的减小误差、提高模型的模拟精度^[18-20]。

本文使用 Python2.7 中的 Pandas、Sklearn、numpy 等工具包通过一系列运算代码构建随机森林模型,将使用的自变量和辅助变量输入模型进行模拟,最终得到目标变量——近地表气温。具体算法步骤如下^[21-22]:

(1)从 N 个训练数据中用自助抽样的方式有放回地对样本随机抽取 n 次(可重复抽样),组成训练集。使用未被抽取的数据进行类别预测,评估其误差。

(2)随机森林的每一棵树都由一个训练集构成,根据自变量的数量选择节点,再根据节点的个数计算最佳分割方式。

(3)将第二个步骤重复 n 次之后可以生成 n 棵决策树,每棵决策树,都可以模拟出预测值,对预测值进行回归分析,最后以每棵树的输出结果平均值作为模型估计值,具体计算过程如下式所示:

$$Y(i) = \frac{1}{n} \sum_{i=1}^n T_n(x_i) \quad i = 1, 2, 3, \dots, n \quad (1)$$

其中, T_n 为第 n 棵决策树; n 为决策树数量。

3 结果分析

3.1 模型输入因子重要性分析

地表温度数据是模拟空间近地表气温的关键,为了避免使用一种地表温度数据给模拟结果带来的偶然性,使用 4 个时间点的地表温度数据作为输入模型的自变量 a ,其他 7 个变量为与近地表气温相关性较高的辅助变量 b 。分析输入因子的重要性可以去除多余无效的变量,减少数据冗余,使模型的模

拟结果更加准确。训练好随机森林模型后,使用 python 直接调用 sklearn 库里的 feature_importances 属性就能得到每个特征的重要性。模型输入因子的重要性指数不同于相关系数,重要性指数是相对的,它是指一种输入因子对模型的影响系数占总输入因子对模型的影响系数的比例,总输入因子对模型的影响系数为 1。这方便了解并分析自变量 a 和辅助变量 b 对目标变量气温的影响程度,值越大,则说明该变量越重要。图 2 为自变量 a 和辅助变量 b 对目标变量的影响程度图。

由图 2a 可见,自变量对目标变量的影响程度都比较高,其中晚上获取的地表温度数据更为重要。所以在利用地表温度模拟气温的研究中要考虑使用卫星夜晚所获取的地表温度数据。从图 2b 中可以看出,所有辅助变量的重要性指数都大于 0.1,其中经度对目标变量的影响程度高于其他的变量,其次是 NDVI 和高程,其他几个变量的重要性相差不大,但都是影响近地表气温估算的重要因子。故在构建模型时需要选取以上 11 个变量来模拟气温。

3.2 结果验证

以 MODIS 两个白天和夜晚的地表温度数据为自变量,随机抽取 90 个站点上的辅助变量(经纬度、NDVI、海拔、坡度、坡向、太阳辐射)与气温数据建立随机森林模型,模拟了 2003—2016 年黄土高原地区的月平均气温,剩余站点用于模型的验证,使用平均绝对误差(Mean Absolute Error, MAE)和均方根误差(Root Mean Square Error, RMSE)来量化模拟误差,由于研究的时间序列较长,选取了 2003—2016 年 7 月和 8 月的模拟结果进行验证(表 2)。

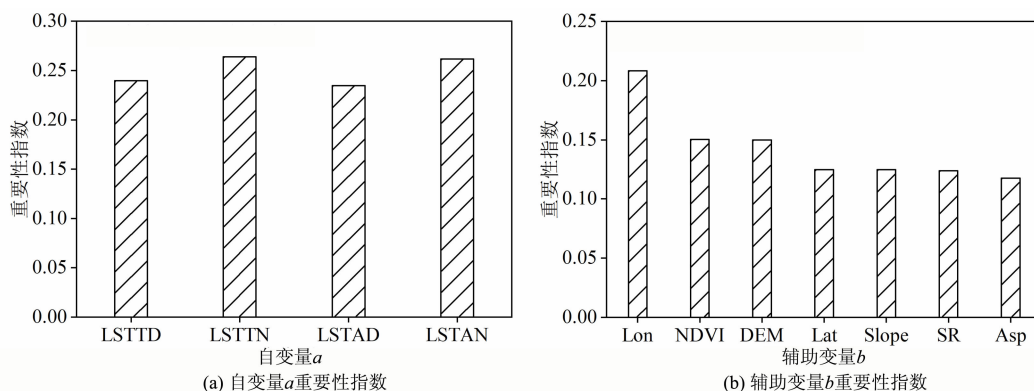


图 2 自变量 a 与辅助变量 b 的重要性指数

Fig. 2 Importance index of independent variable a and auxiliary variable b

表 2 模型验证结果

Tab.2 The validation results of the model

时间	MAE/℃	RMSE/℃	R ² 系数	时间	MAE/℃	RMSE/℃	R ² 系数
2003 年 7 月	1.48	1.77	0.83	2010 年 7 月	1.46	1.73	0.83
2003 年 8 月	1.93	2.35	0.88	2010 年 8 月	1.10	1.21	0.83
2004 年 7 月	1.00	1.24	0.86	2011 年 7 月	1.10	1.29	0.85
2004 年 8 月	1.05	1.43	0.83	2011 年 8 月	1.50	1.90	0.85
2005 年 7 月	0.94	1.17	0.85	2012 年 7 月	0.94	1.19	0.84
2005 年 8 月	1.25	1.58	0.86	2012 年 8 月	1.12	1.25	0.84
2006 年 7 月	0.96	1.18	0.82	2013 年 7 月	0.97	1.20	0.86
2006 年 8 月	1.06	1.16	0.87	2013 年 8 月	1.04	1.28	0.84
2007 年 7 月	0.91	1.06	0.87	2014 年 7 月	1.11	1.4	0.83
2007 年 8 月	1.06	1.24	0.85	2014 年 8 月	1.10	1.23	0.83
2008 年 7 月	1.14	1.37	0.86	2015 年 7 月	1.17	1.55	0.85
2008 年 8 月	1.07	1.27	0.84	2015 年 8 月	0.96	1.28	0.85
2009 年 7 月	1.06	1.36	0.86	2016 年 7 月	1.14	1.22	0.84
2009 年 8 月	1.07	1.23	0.85	2016 年 8 月	0.98	1.28	0.83

注:表中粗体为最优模型方案。

从验证结果来看,模型显示了其良好的优越性及模拟精度,误差保持在一个较低的水平,气温的平均绝对误差在 1℃ 附近波动。均方根误差大多在 1.2℃ 附近。通过验证结果可以得出,随机森林对近地表气温的模拟有着很好的适用性且模拟效果较好。

以上验证结果月份均为夏季,为了保证模拟结果的可靠性,选取了 2003—2016 年各季气温模拟结果进行整体结果验证(图 3)。结果表明所有样本大都聚集在 1:1 线的周围,故模型模拟的气温值与站点气温值非常接近,随机森林在近地表气温模拟上有良好的优越性。从图中还可以看出样本值两极分化较为严重,以 15℃ 和 0℃ 为界限,大于 15℃ 的样本更接近 1:1 线,拟合度较高,说明模型在模拟大于 15℃ 的近地表气温时效果更好。通过后期的计算得出整个夏季的平均 MAE 为 1.15℃,平均 RMSE 为 1.44℃。春季平均 MAE 为 1.53℃,平均 RMSE 为 1.92℃。秋季平均 MAE 为 2.06℃,平均 RMSE 为 2.61℃。冬季平均 MAE 为 3.03℃,平均 RMSE 为 3.74℃。综上,模型在夏季的模拟效果最好,春秋次之,冬季最差。这符合前人所提出的气温较高时模型模拟效果较好的研究结论^[23]。

3.3 黄土高原气温时空分布特征

3.3.1 气温的时空分布特征

通过计算得出了 2003—2016 年逐月近地表气

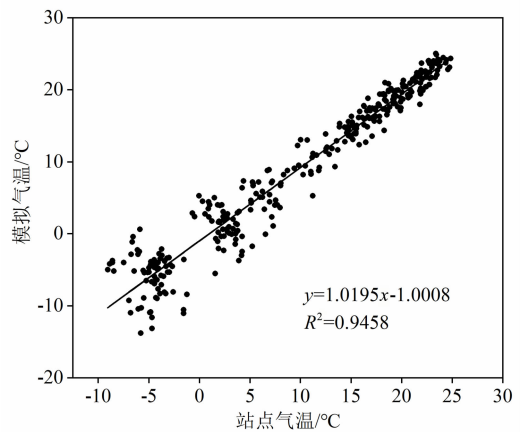


图 3 2003—2016 模拟气温与站点气温散点图

Fig.3 2003 - 2016 scatter plot of simulated temperature and site temperature

温空间分布结果,为了更好的研究气温的时空分布特征,将模拟结果按照春、夏、秋、冬进行划分,由于春秋冬季,多云阴雨天气较多,卫星获取的数据质量较差,个别月份模拟的效果不好,而夏季各月份的模拟结果整体较好,故使用 2003—2016 年夏季的模拟结果来研究气温的时空分布,模拟结果如图 4 所示。

从图 4 可以得出,2003—2016 年夏季气温的高值与低值之间差异明显,总体来说高温一般分布在黄河主要干支流流经区域,低温主要分布在西北部和高南部的高原山地。除了高值与低值,气温中间值

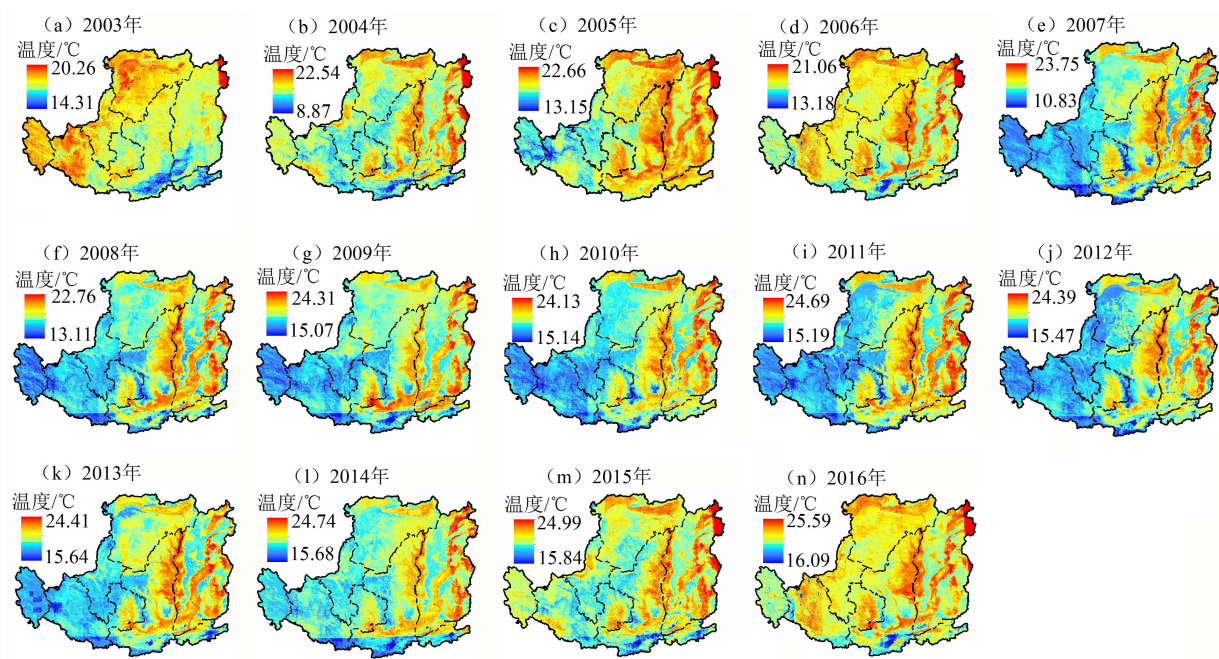


图4 2003—2016 黄土高原夏季气温模拟结果

Fig. 4 2003–2016 summer temperature simulation results on the Loess Plateau

的差异也有较好的体现,以黄土高原北部的河套平原为例,平原内部的气温要高于周边地区,而又由于经纬度等影响因子的影响,气温值普遍低于中部平原地区。鄂尔多斯高原西部由于地形和海拔高度的影响又能区别于纯粹的高原山地地形,使得气温值高于西南部的高原山地地区。汉中平原以西呈现出海拔越高气温越低的趋势。从时间序列上来说 2003—2016 年夏季气温呈现波动上升的趋势,黄土高原的中西部变暖趋势更为明显。为了更好地研究黄土高原各季节的气温年际变化情况,计算得出各季节的平均气温如图 5 所示。

相较于气象站点的气温,模拟得到的黄土高原空间气温数据更加准确,能更好地反映黄土高原地区各季节气温变化情况,从图中可以看出春夏气温总体呈现出波动上升的趋势,冬季气温上升幅度不大。秋季气温变化的波动最大,整体上呈现出下降趋势,从 2007 年气温呈现出逐年上升趋势,其中 2006—2008 年间气温波动最为明显。综上所述黄土高原地区各季气温整体上呈现出逐年上升的趋势。

3.3.2 气温的变化趋势

采用非参数化趋势度 (Sen) 方法来计算黄土高原地区 2003—2016 年年均气温的变化趋势,当系数

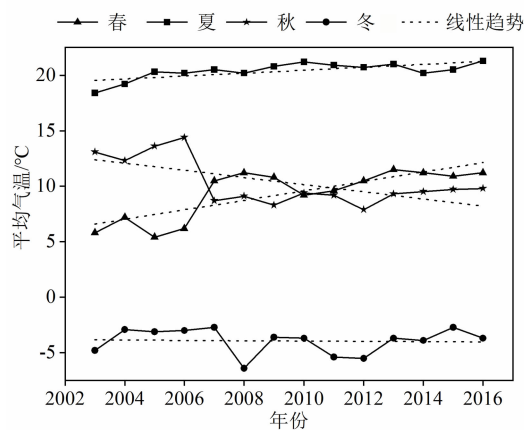


图5 2003—2016 黄土高原各季平均气温趋势图

Fig. 5 2003–2016 seasonal average temperature trend map of the Loess Plateau

为正值时,代表气温随时间的变化呈上升趋势,反之,则呈下降趋势,值的大小为变化的快慢程度。从图 6a 中可以看出黄土高原大多数地区近地表气温总体上呈现出上升趋势,中部最为明显,西部和南部山区呈下降趋势。Mann-Kendall 统计检验方法,用来判断趋势显著性,使用 Mann-Kendall 统计检验法对黄土高原地区 2003—2016 年年均气温变化趋势的显著性进行了检验,检验结果分别通过了置信度为 90%、95% 和 99% 的显著性检验。从图 6b 中可

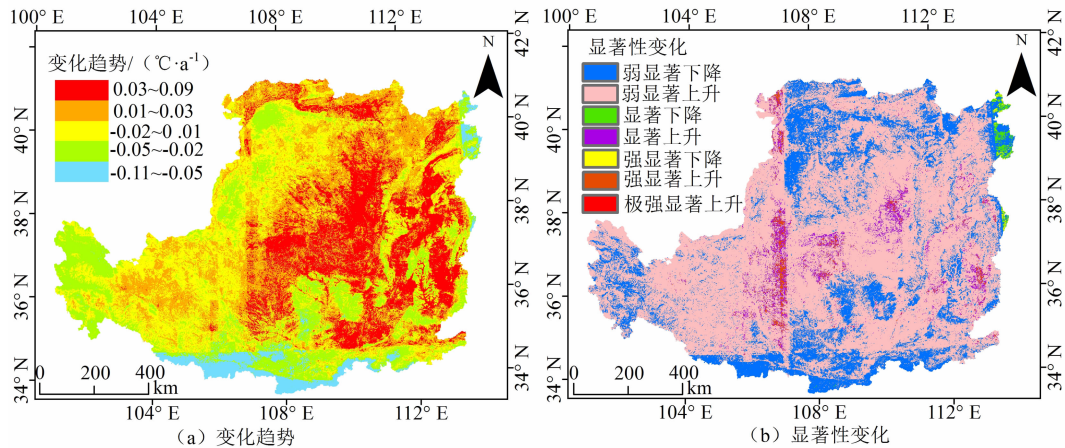


图 6 2003—2016 年均价空间变化趋势及显著性变化

Fig. 6 The spatial change trend and significant change of average temperature from 2003 to 2016

以看出黄土高原地区气温整体上为弱显著上升趋势,只有北部、西部和南部的山地呈弱显著下降趋势,从计算得出的数值来看,气温上升和下降的幅度较小。用所得数值乘 10 为气温倾向率,根据平均倾向系数所得出的气温倾向率为 $0.11\text{ }^{\circ}\text{C}/10\text{ a}$ 。

4 讨论与结论

4.1 讨论

如何获取空间分布的近地表气温数据一直是国内外研究的热点问题,相较于统计方法、TVX 法和能量平衡法,使用随机森林算法结合遥感数据进行气温模拟效果更好,误差更低。本文在前人的基础上探索了近地表气温新思路,加入了卫星夜晚所获取的地表温度数据作为变量,模拟结果相较于 Li、Yoo 等人更好^[1,15]。模拟空间分辨率较高的近地表气温数据能更好的研究区域内气候演变规律及水热变化。通过分析模拟气温的空间分布特征及变化趋势,得出黄土高原地区春季和夏季平均气温总体呈现出波动上升的趋势,这与张定全等人的研究结果一致^[18-20],从年均温的变化趋势与时空分布特征来看,黄土高原气温总体呈上升趋势,中部区域尤为明显。顾朝军、晏利斌等^[25-26]采用线性趋势法、Mann-Kendall 法和 Morlet 小波分析等方法对黄土高原近地表气温变化趋势进行分析,得出黄土高原地区年均温上升趋势显著,气候倾向率为 $0.31\text{ }^{\circ}\text{C}/10\text{ a}$ 。前人的研究结果与本文计算结果相差不大,且黄土高原年均气温在变化趋势上与本文的研究较为一

致。故使用遥感数据结合随机森林算法进行近地表气温的模拟是可行的。由于研究区内气象站点的疏密程度不同,模型在站点较少区域存在一定误差,遥感数据的质量也对模拟结果有着重要影响^[2,14]。因此在以后的研究中,要注重讨论遥感数据质量差及站点较少区域模拟结果较差的问题,探索解决方案,模拟出高精度空间化分布的近地表气温数据。

4.2 结论

采用随机森林算法并结合遥感数据,使用 11 个变量模拟了 2003—2016 年黄土高原地区的月平均气温,分析了不同影响因子对气温的重要性,分析了黄土高原地区气温的时空变化特征,使用气象站点的实测气温值对模拟结果进行了验证。通过本文的研究可以得出以下主要结论:

(1) 随机森林算法相较于其他气温模拟方法,更适合近地表气温的模拟,且模拟的精度较高。本文中最优模拟结果的平均绝对误差达到了 $0.91\text{ }^{\circ}\text{C}$,均方根误差为 $1.06\text{ }^{\circ}\text{C}$ 。相较于其他气温模拟方法,模型的输入因子容易获取,具有更好的普适性。对于模拟不同区域时空分布的气温数据,分析气温的时空变化特征起到一定的参考作用。

(2) 在进行变量重要性分析时发现,卫星夜晚获取的地表温度数据对近地表气温模拟结果更为重要。绝大部分学者在使用地表温度模拟近地表气温的研究中往往会忽视卫星夜晚获取的地表温度数据。事实证明卫星夜晚获取的地表温度数据是模拟近地表气温的重要参数,为使用夜晚地表温度数据进行近地表气温的研究提供了依据。

(3)黄土高原地区2003—2016年年均气温大部分区域呈现出逐年缓慢上升的趋势;在黄土高原的东北部气温呈现出显著下降趋势;西部、北部和南部山区气温呈现出缓慢下降的趋势。综合来说黄土高原气温上升趋势大于下降趋势。

参考文献(References)

- [1] LI Long, ZHA Yong. Estimating monthly average temperature by remote sensing in China[J]. *Advances in Space Research*, 2019, **63**(8): 2345–2357.
- [2] 邢立亭,李净,焦文慧. 基于MODIS和随机森林的兰州市日最高气温和最低气温估算[J]. *干旱区研究*, 2020, **37**(3): 689–695. [XING Liting, LI Jing, JIAO Wenhui. Estimation of daily maximum and minimum temperature of Lanzhou City based on MODIS and random forest[J]. *Arid Zone Research*, 2020, **37**(3): 689–695]
- [3] MAO K B, TANG H J, WANG X F, et al. Near-surface air temperature estimation from ASTER data based on neural network algorithm[J]. *International Journal of Remote Sensing*, 2008, **29**(20): 6021–6028.
- [4] PEDE T, MOUNTRAKIS G, SHAW S B. Improving corn yield prediction across the US Corn Belt by replacing air temperature with daily MODIS land surface temperature[J]. *Agricultural and Forest Meteorology*, 2019, 276–277:107615.
- [5] 韩秀珍,李三妹,窦芳丽. 气象卫星遥感地表温度推算近地表气温方法研究[J]. *气象学报*, 2012, **70**(5): 1107–1118. [HAN Xiuzhen, LI Sanmei, DOU Fangli. Study of obtaining high resolution near-surface atmosphere temperature by using the land surface temperature from meteorological satellite data[J]. *Acta Meteorologica Sinica*, 2012, **70**(5): 1107–1118]
- [6] 祝善友,张桂欣. 近地表气温遥感反演研究进展[J]. *地球科学进展*, 2011, **26**(7): 724–730. [ZHU Shanyou, ZHANG Guixin. Progress in near surface air temperature retrieved by remote sensing technology[J]. *Advances in Earth Science*, 2011, **26**(7): 724–730]
- [7] 张丽文,黄敬峰,王秀珍. 气温遥感估算方法研究综述[J]. *自然资源学报*, 2014, **29**(3): 540–552. [ZHANG Liwen, HUANG Jingfeng, WANG Xiuzhen. A review on air temperature estimation by satellite thermal infrared remote sensing[J]. *Journal of Natural Resources*, 2014, **29**(3): 540–552]
- [8] DAVIS F A, TARPLEY J D. Estimation of Shelter temperatures from operational satellite sounder data[J]. *Journal of Applied Meteorology*, 2010, **22**(3): 369–376.
- [9] CRESSWELL M P. Estimating surface air temperatures, from Meteosat land surface temperatures, using an empirical solar zenith angle model[J]. *International Journal of Remote Sensing*, 1999, **20**(6): 1125–1132.
- [10] 李斌,王慧敏,秦明周,等. NDVI、NDMI与地表温度关系的对比研究[J]. *地理科学进展*, 2017, **36**(5): 585–596. [LI Bin, WANG Huimin, QIN Mingzhou, et al. Comparative study on the correlations between NDVI, NDMI and LST[J]. *Progress in Geography*, 2017, **36**(5): 585–596]
- [11] 丁海勇,李往华. 基于TVX方法的南京市城区时空格局与地表温度的研究[J]. *长江流域资源与环境*, 2018, **27**(4): 735–744. [DING Haiyong, LI Wanghua. Analysis of land use land cover temporal-spatial distribution and land surface temperature in Nanjing City using TVX method[J]. *Resources and Environment in the Yangtze Basin*, 2018, **27**(4): 735–744]
- [12] STISEN S, SANDHOLT I, NØRGAARD A, et al. Estimation of diurnal air temperature using MSG SEVIRI data in West Africa[J]. *Remote Sensing of Environment*, 2007, **110**(2): 262–274.
- [13] PAPE R, LÖFFLER J. Modelling spatio-temporal near-surface temperature variation in high mountain landscapes[J]. *Ecological Modelling*, 2004, **178**(3): 483–501.
- [14] 白琳,徐永明,何苗,等. 基于随机森林算法的近地表气温遥感反演研究[J]. *地球信息科学学报*, 2017, **19**(3): 390–397. [BAI Lin, XU Yongming, HE Miao, et al. Remote sensing inversion of near surface air temperature based on random forest[J]. *Journal of Geo-Information Science*, 2017, **19**(3): 390–397]
- [15] YOO C, IM J, PARK S, et al. Estimation of daily maximum and minimum air temperatures in urban landscapes using MODIS time series satellite data[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, **137**:149–162.
- [16] FU Peng, WENG Qihao. Variability in annual temperature cycle in the urban areas of the United States as revealed by MODIS imagery[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, **146**:65–73.
- [17] 翟丹平,白红英,冯海鹏,等. 基于气象数据和遥感影像的太白山气温直减率[J]. *山地学报*, 2016, **34**(4): 496–503. [ZHAI Danping, BAI Hongying, FENG Haipeng, et al. Temperature lapse rates in the taibai mountain based on meteorological data and remote sensing image[J]. *Journal of Mountain Research Science*, 2016, **34**(4): 496–503]
- [18] BREIMAN L. Random forests[J]. *Machine Learning*, 2001, **45**(1): 5–32.
- [19] 周屹,冯兆祥,白熙卓,等. 基于随机森林算法的数据分析软件设计[J]. *黑龙江工程学院学报*, 2017, **31**(3): 38–41. [ZHOU Yi, FENG Zhaoxiang, BAI Xizhuo, et al. Design of data analysis software based on random forest algorithm[J]. *Journal of Heilongjiang Institute of Technology*, 2017, **31**(3): 38–41]
- [20] 刘剑,曹美燕,高治军,等. 一种基于随机森林的太阳能辐射预测模型[J]. *控制工程*, 2017, **24**(12): 2472–2477. [LIU Jian, CAO Meiyang, GAO Zhijun, et al. A solar radiation prediction model based on random forest[J]. *Control Engineering of China*, 2017, **24**(12): 2472–2477]

- [21] BREIMAN L. Bagging predictors[J]. Machine Learning, 1996, **24** (2): 123–140.
- [22] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, **26** (3): 32–38. [FANG Kuangnan, WU Jianbin, ZHU Jianping, et al. A review of technologies on random forests[J]. Statistics & Information Forum, 2011, **26** (3): 32–38]
- [23] 梅静, 王建, 何亮, 等. 川西甘孜州 1961—2015 年气温和降水时空变化特征研究[J]. 山地学报, 2019, **37** (2): 161–172. [MEI Jing, WANG Jian, HE Liang, et al. Spatio-temporal variations of temperature and precipitation in Ganzi of Western Sichuan in China during 1961–2015 [J]. Mountain Research, 2019, **37** (2): 161–172]
- [24] 张定全, 王毅荣. 中国黄土高原地区春季气温时空特征分析[J]. 高原气象, 2005 (6): 898–904. [ZHANG Dingquan, WANG Yirong. Spatial and temporal characteristics of air temperature in China Loess Plateau in spring [J]. Plateau Meteorology, 2005 (6): 898–904.]
- [25] 顾朝军, 穆兴民, 高鹏, 等. 1961—2014 年黄土高原地区降水和气温时间变化特征研究[J]. 干旱区资源与环境, 2017, **31** (3): 136–143. [GU Chaojun, MU Xingmin, GAO Peng, et al. Characteristics of temporal variation in precipitation and temperature in the Loess Plateau from 1961 to 2014 [J]. Journal of Arid Land Resources and Environment, 2017, **31** (3): 136–143]
- [26] 晏利斌. 1961—2014 年黄土高原气温和降水变化趋势[J]. 地球环境学报, 2015, **6** (5): 276–282. [YAN Libin. Characteristics of temperature and precipitation on the Loess Plateau from 1961 to 2014 [J]. Journal of Earth Environment, 2015, **6** (5): 276–282]

Temperature Simulation and Temporal Variation Based on Remote Sensing Data and Random Forest Algorithm: A Case Study in the Loess Plateau Region, China

XING Liting, LI Jing*

(College of Geographical and Environmental Science, Northwest Normal University, Lanzhou 730070, China)

Abstract: Air temperature is an integral representation of regional thermal environment and one of key factors that cause climate change. A continuous distribution of air temperature derived by spatial interpolation or extrapolation based on observations merely at several scattered meteorological stations cannot accurately characterize the spatial variation of air temperature over a large area. In this paper, an improved approach to model the variation of air temperature was exemplified by a case study in the Loess Plateau region, which used remote sensing data combined with random forest machine learning to analyze the monthly mean air temperature from 2003 to 2016 to obtain its continuous spatio-temporal distribution. The significances of 11 variables related to air temperature were analyzed and the performance of the random forest model was examined. Then these 11 verified variables and random forest model were used to simulate air temperature in the Loess Plateau from 2003 to 2016. The research found that the average absolute error of near-surface air temperature obtained by our proposed approach was $0.91\text{ }^{\circ}\text{C}$, with a mean square error $1.06\text{ }^{\circ}\text{C}$, exhibiting a higher accuracy; According to the simulation, in most areas of the Loess Plateau from 2003 to 2016, the near-surface temperature had presented a slow rising trend, which is consistent with past research. Conclusively, the air temperature simulation using remote sensing data combined with the random forest algorithm has good applicability in the Loess Plateau region, and is of significance to the study of the law of temperature evolution and water and heat changes in the Loess Plateau.

Key words: MODIS; Random forest algorithm; Near surface temperature; Loess Plateau