

基于多窗口线性回归模型的浙北地区冬季气温估算

王越,刘小平*,黎夏,姚尧,盛艳玲

(中山大学地理科学与规划学院,广东省城市化与地理环境空间模拟重点实验室,广东 广州 510275)

摘要:气温(T_a)是描述陆地气候环境的一个重要参数,其异常变化直接影响人类的生存环境,因此如何高精度地估算气温成为当前研究的热点。MODIS数据因其分辨率较低不能提供精细的地表信息,为此,本文以更高分辨率的Landsat8影像为数据源,结合自动气象站的气温数据,耦合经纬度、归一化植被指数、归一化建筑指数和改进的归一化水体指数等多种因子,建立了多窗口线性回归模型(Multi-Window Linear Regression Model, MWLR)。最后以浙江北部为研究区,使用MWLR模型对该地区冬季气温进行了估算,模型预测的RMSE在 $1.458\sim 1.551\text{ }^{\circ}\text{C}$ 之间, R^2 在 $0.835\sim 0.842$ 之间,当窗口大小为 3×3 时取最优精度(RMSE= $1.458\text{ }^{\circ}\text{C}$, $R^2=0.835$),优于一般的空间内插方法。研究结果验证了利用MWLR模型和Landsat8影像进行气温估算的有效性,并提供了一种基于遥感数据在局部地区开展高精度和高分辨率气温估算的模型。

关键词:多窗口线性回归模型;气温估算;Landsat8

中图分类号:P407 **文献标识码:**A **文章编号:**1672-0504(2017)02-0045-07

0 引言

气温(Air Temperature, T_a)通常是指离地表 1.5 m 或 2 m 高的近地面空气的温度^[1],是表征陆地环境状况的一项重要指标^[2]。气温异常易引发森林火灾^[3]、干旱^[4]、病虫害^[5]、传染性病菌传播等^[6],从而威胁人类的生存与发展,因此对气温的观测非常重要。目前,在遥感观测技术日趋成熟的形势下,利用卫星遥感大范围空间连续成像的特点对气温进行研究,能够有效弥补气象站点覆盖不均、气温数据空间不连续的缺陷^[1,7,8]。因此,利用遥感影像构建高精度、高实用性的气温估算模型已成为该领域的研究热点。

考虑到大气传输的复杂性和空气信号的薄弱,通过热红外遥感数据直接得到气温数据的难度太大,一般都是基于地表能量平衡的原理通过地表温度(Land Surface Temperature, LST)估算气温^[7,9-11]。目前,常用的气温估算方法主要有温度-植被指数(Temperature Vegetation Index, TVX)法^[2,12-16]、能量平衡法^[8,17]和统计分析法^[18-21]。TVX方法是一种利用地表温度和光谱植被指数间的负相关性从遥感数据中提取气温的空间邻域运算方法,估算精度在 $2\sim 5\text{ }^{\circ}\text{C}$ 之间^[2,12-15],但该方法并不适用于植被覆盖度较低或无植被区域^[16]。能量平衡法以地表能量平衡理论为基础,将显热通量和潜

热通量表达为地气温差(LST- T_a)的函数,以此建立地表温度与气温之间的关系,该方法估算精度在 $3\text{ }^{\circ}\text{C}$ 左右^[8,17]。由于部分参数(如空气动力学阻抗、风速等)无法通过遥感手段获取,只能依赖实地观测,导致此类研究较少^[1,22]。统计分析方法由于能弥补前述两种方法的不足而被广泛使用,主要包括:1)单因子统计法。即建立气温与地表温度之间的线性模型,模型估计误差在 $2\text{ }^{\circ}\text{C}$ 左右^[18],精度较高,但过于简单;2)多元线性回归方法。主要是在模型中加入地理因子(如经纬度、高程等)、反照率、太阳天顶角等多元变量,如Cristóbal等^[19]利用Landsat数据,以经纬度、距海岸距离和太阳入射辐射等地理因子和反照率、归一化植被指数、LST作为多元变量建立模型,将月平均气温估算模型的预测精度提高至 $1.0\text{ }^{\circ}\text{C}$;徐伟燕等^[20]基于MODIS数据,通过比较单因子统计法、多因子统计法和地表能量平衡法,发现多因子统计法精度最高;同时,其他学者也在算法上进行了改进,如姚永慧等^[21]利用MODIS数据针对青藏高原地区采用地理加权回归(GWR)方法建立了高原气温与地表温度、海拔之间的模型。

目前,多数相关研究以MODIS(Moderate-Resolution Imaging Spectroradiometer)数据为主要的遥感数据源^[3,5,9,13,15,17],而基于最新的Landsat8数据建立近地表气温估算模型的研究较少。作为最新一代的Landsat系列卫星,Landsat8数据的应用具

收稿日期:2016-09-28; 修回日期:2016-12-28

基金项目:国家自然科学基金项目(41531176,41371376)

作者简介:王越(1992-),女,硕士研究生,主要从事环境遥感地气温度研究。E-mail:wangy458@mail2.sysu.edu.cn

有很高的研究价值^[23-25]。本文以浙江北部为研究区,采用 2014 年冬季的 Landsat8 影像,基于辐射传输方程反演出地表温度^[26],耦合经纬度、归一化植被指数 (Normalized Difference Vegetation Index, NDVI)、归一化建筑指数 (Normalized Difference Building Index, NDBI)、改进的归一化水体指数 (Modified Normalized Difference Water Index, MNDWI) 等因子,结合高密度的自动气象站点数据建立多窗口线性回归模型 (Multi-Window Linear Regression Model, MWLR), 并采用 10 折交叉验证方法对该模型进行精度评价。

1 研究区与数据

浙江省位于我国东部沿海,西南以山地为主,中部以丘陵为主,东北部是低平的冲积平原;全省属典型的亚热带季风气候,据浙江省统计信息网 (<http://www.zj.stats.gov.cn>) 显示,该省年均气温 15~18℃,极端最高气温 33~43℃,极端最低气温 -8~-2.2℃。为较大程度减少高程对气温模拟的影响,本研究选择水网密集、地形平坦的浙江北部冲积平原地区 (118.38~122.13°E, 29.06°~31.16°N) (包括杭州市、宁波市、湖州市、嘉兴市) 为实验区。

本研究实验数据包括: 1) 气象数据。来源于高密度均匀覆盖于研究区内的 601 个自动气象站 (图 1) 和中国地面气候资料日值数据集 (V3.0)。气象站数据更新频率为每小时,剔除缺失站点,提取余下有效站点的经纬度作为自变量之一,取临近卫星过境时间 (11:00) 的气象温度作为实验的因变量 T_a ; 中国地面气候资料日值数据集 (V3.0) 更新频率为每天,可提取站点高度、气压、平均气温和相对湿度等数据计算大气剖面参数,用于地表温度的反演。2) 遥感数据。通过遥感估算气温需要选择晴朗无云的影像^[7],鉴于浙江省雨热同期的气候特点,本文采用的遥感数据为浙江北部 2014 年冬季 1 月、11 月和 12 月的 Landsat8 影像,卫星过境的拍摄时间约为 10:30,影像清晰无云,质量较好,轨道号为 118/39、118/40、119/39、119/40、120/40,能够完整覆盖研究区范围。Landsat8 卫星主要载荷为陆地成像仪 (Operational Land Imager, OLI) 和热红外传感器 (Thermal Infrared Sensor, TIRS)。其中 TIRS 分辨率为 100 m,考虑到其定标参数发生过多次变化,目前第 11 波段的定标不确定性比第 10 波段大近 1 倍,且 USGS 建议不采用双波段的劈窗算法等^[23,27],因此本文最终只选用 TIRS 第 10 波段反演地表温度。

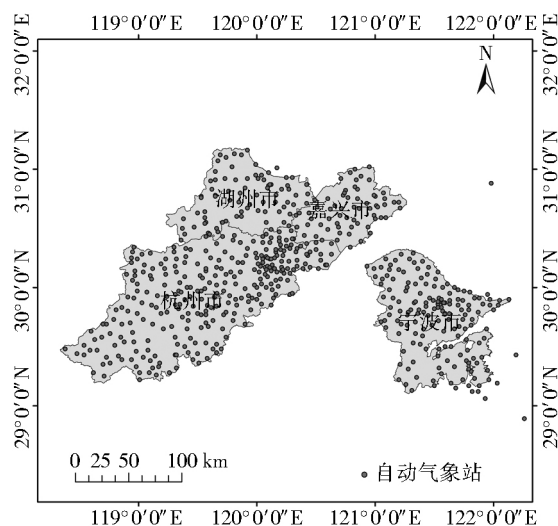


图 1 研究区自动气象站分布

Fig. 1 Distribution of automatic meteorological stations in study area

2 方法

本文首先基于辐射传输方程反演研究区地表温度^[26],引入多窗口处理方法对 LST、NDVI 等 4 种因子进行处理,再对所有因子进行共线性诊断,得到可纳入模型的自变量;其次,结合自动气象站的气温数据进行多元线性回归建模;为增强模型的鲁棒性,使用 10 折交叉验证方法 (10-fold Cross-Validation)^[28]对模型进行检验,最终确定 MWLR 模型并得到气温估算结果 (图 2)。

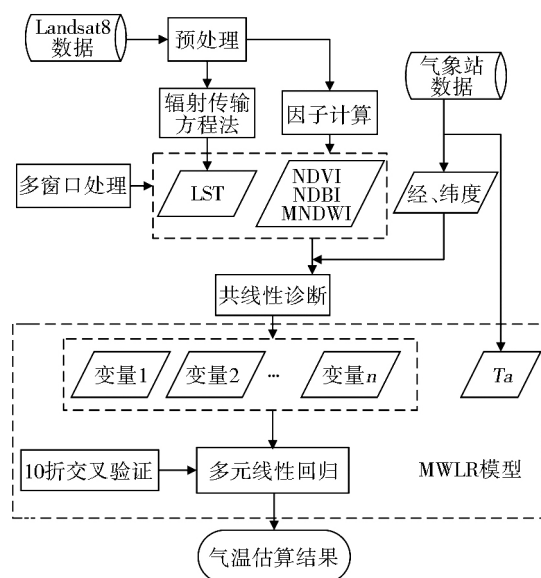


图 2 基于 MWLR 模型估算气温流程

Fig. 2 Flow chart of air temperature estimation by MWLR model

2.1 地表温度反演

辐射传输方程法的基本原理为:首先利用与卫星过境时间同步的大气数据估算大气对地表热辐射的影响;其次将这部分大气影响从卫星传感器所观测到的热辐射总量中剔除,从而得到真实的地表热

辐射强度;最后将其转化为相应的地表温度^[26]。该方法假设大气和地表对热辐射具有朗伯体性质,卫星传感器接收到的热红外辐射亮度值 L_λ 由 3 个部分组成:大气上行辐射、地面的真实辐射亮度经过大气层后到达传感器的能量以及大气向下辐射到达地面后反射穿透大气的能量,其计算公式为:

$$L_\lambda = [\epsilon B(T_s) + (1 - \epsilon)L_\downarrow] \tau + L_\uparrow \quad (1)$$

式中: ϵ 为地表比辐射率; T_s 为地表真实温度(K); $B(T_s)$ 为黑体温度为 T_s 时的普朗克函数; τ 为大气在热红外波段的透过率; $L_\uparrow$ 、 $L_\downarrow$ 分别为大气上行辐射和下行辐射,单位为 $W/(m^2 \cdot sr \cdot \mu m)$, 其与 τ 统称为大气剖面参数,可通过登录 NASA 官网(<http://atmcorr.gsfc.nasa.gov>)输入卫星成像时间、中心经纬度及站点高程、气压、相对湿度、平均气温等生成。

地表比辐射率 ϵ 是地表温度反演过程中非常重要的参数^[29],其不仅依赖于地表物体的组成,而且与物体的表面状态(如粗糙度等)及物理性质(如介电常数、含水量等)有关,并随着所测定的波长和观测角度等因素有所变化,因此对比辐射率进行精确测量的难度很大。目前估计地表比辐射率的方法可分为经验公式法^[30]和混合像元法^[31-33]两种,本文采用简化后的 Sobrino 模型^[32]进行地表比辐射率的计算,公式如下:

$$\epsilon = 0.004P_v + 0.986 \quad (2)$$

式中: P_v 为植被覆盖度,其计算公式如下:

$$P_v = \frac{NDVI - NDVI_{soil}}{NDVI_{veg} - NDVI_{soil}} \quad (3)$$

式中: $NDVI_{soil}$ 为完全是裸土或无植被覆盖区域的 NDVI 值; $NDVI_{veg}$ 则代表纯植被像元的 NDVI 值。本研究取经验值 $NDVI_{veg} = 0.70$, $NDVI_{soil} = 0.05$, 即当某个像元的 NDVI 大于 0.70 时, P_v 取值为 1; 当 NDVI 小于 0.05, P_v 取值为 0。

最后根据普朗克函数将黑体的辐射亮度 $B(T_s)$ 转换为地表真实温度 T_s , 即本实验中的地表温度 (LST), 转换公式为:

$$T_s = K2 / \ln\left(\frac{K1}{B(T_s)} + 1\right) \quad (4)$$

式中: $K1$ 和 $K2$ 为卫星发射前预设的常量,对于 Landsat 8 TIRS 的第 10 波段,常量 $K1 = 774.89 W/(m^2 \cdot \mu m \cdot sr)$, $K2 = 1321.08 K$ ^[25]。

2.2 基于 MWLR 模型的气温估算

本研究基于多窗口处理和多元线性回归方法建立 MWLR 模型,模型中的因子综合考虑了地表温度、站点经纬度、土地覆盖类型等信息。本文的实验

步骤为:首先基于研究区数据求出各个因子,并对各因子进行多窗口处理,通过共线性诊断除去干扰因子;其次,经过多元线性回归得到 MWLR 模型及气温预测值;最后对其预测精度进行评价验证。

2.2.1 多元变量选择 诸多研究表明气温与地表温度之间存在一定的线性关系^[7,9,17],但实际上地、气温度之间的关系非常复杂^[1]。气象站点周围植被覆盖的多少、土地类型的不同、站点所在的地理位置等均会对气温产生较大影响^[34],因此本文综合前人的研究和实际的数据情况,加入 NDVI、NDBI^[35]、MNDWI^[36]以及气象站点所在的经纬度数据,与地表温度一同作为 MWLR 模型中的自变量。

NDVI 能够在一定程度上消除与太阳高度角、卫星观测角、地形、云影等与大气条件有关的辐射变化的影响,是热红外遥感数据建立气温模型时最常用的变量。其公式如下:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (5)$$

式中: NIR 表示近红外波段处的表观反射率值; RED 表示红波段处的表观反射率值。

NDBI 利用只有城镇在近红外和短波红外的反射率持续上升的特点提取城镇用地:

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \quad (6)$$

式中: $SWIR$ 表示短波红外处的表观反射率值。

MNDWI 弥补了 NDWI 在城镇建成区难以提取水体信息的缺陷,同时可以消除地形差异引起的阴影问题。由于水体的比热容大,对气温的影响较高,因此实验中把 MNDWI 也作为一个重要自变量,计算公式为:

$$MNDWI = \frac{GREEN - SWIR}{GREEN + SWIR} \quad (7)$$

式中: $GREEN$ 表示绿波段处的表观反射率值。

气温与地表温度会随着经纬度的变化而变化,但经纬度与温度变化的相关性和贡献率仍需要探讨,因此从自动气象站数据集中提取出每个气象站点对应的经纬度信息,将二者作为模型的自变量,以研究其对气温的影响。

2.2.2 多窗口处理 传统方法获得的某点气温值是该点周围一定范围内的所有因素共同作用的结果,因此可认为气象站点获得的气温值是比一个像元更广泛的范围内热交换的结果,即气象站点测得的气温值与气象站周围的环境要素之间有着密切关系^[34]。为研究气象站周围不同范围像元对中心像元的影响,同时考虑到窗口过大会造成较大误差,本文

最终选取了 1×1 、 3×3 、 5×5 和 9×9 共 4 种窗口, 对除经纬度信息外的其他因子, 分别计算窗口内像元的平均值。

2.2.3 多元线性回归 经过地表温度反演和多窗口处理后, 可得到 4 组多元变量, 在进行建模之前必须进行多元变量的共线性诊断。共线性可能会使回归的结果造成混乱, 甚至会把分析引入歧途, 也可能对参数估计值的正负号产生影响, 特别是各回归系数的正负号有可能同预期的正负号相反。统计学上一般用方差膨胀因子 (Variance Inflation Factor, VIF) 或容忍度 (Tolerance) 作为判定指标, 二者互为倒数关系。通常认为, 满足 $\text{Tolerance} < 0.1$ 或 $\text{VIF} > 10$, 就存在严重的共线性^[37]。因此, 将自动气象站所在的经纬度、多窗口处理得到的地表温度、各类指数因子与自动气象站获得的气温数据进行逐步回归 (Stepwise Regression), 考查自变量之间的共线性以及排除干扰变量和无关变量。

然后, 将符合要求的自变量与气温数据进行多元线性回归, 公式如下:

$$Ta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_n X_n + e \quad (8)$$

式中: β_0 为常数, $\beta_j (j=1, 2, \cdots, n)$ 为回归系数; n 为参与建立模型的自变量数目; $X_j (j=1, 2, \cdots, n)$ 表示 n 个自变量, 包括 LST、经纬度、NDVI、NDBI 和 MNDWI 等因子; e 表示随机误差。

为增强模型稳定性, 同时防止出现过拟合现象, 本文通过 10 折交叉验证选择最优结果: 首先将全部样本随机平均分割成 10 份, 每 1 份子集轮流做测试集 (test set), 其余 9 份作为训练集 (train set)。训练集用来建立模型, 测试集作为检验^[38], 最后比较预测值与测试集实际观测值的均方根误差, 选择 10 次训练中 RMSE 最小的作为最优结果。

3 结果分析

3.1 地气温度相关性分析

在进行建模之前, 首先检验 LST 反演结果与气温 (T_a) 之间的相关性。通过线性回归 (图 3) 发现, 本实验针对 Landsat8 影像反演得到的 LST 结果与气温实际值之间具有良好的相关性, 决定系数 R^2 分别为 0.810、0.813、0.816、0.807。其中 5×5 窗口下的 LST 与 T_a 相关性最高。

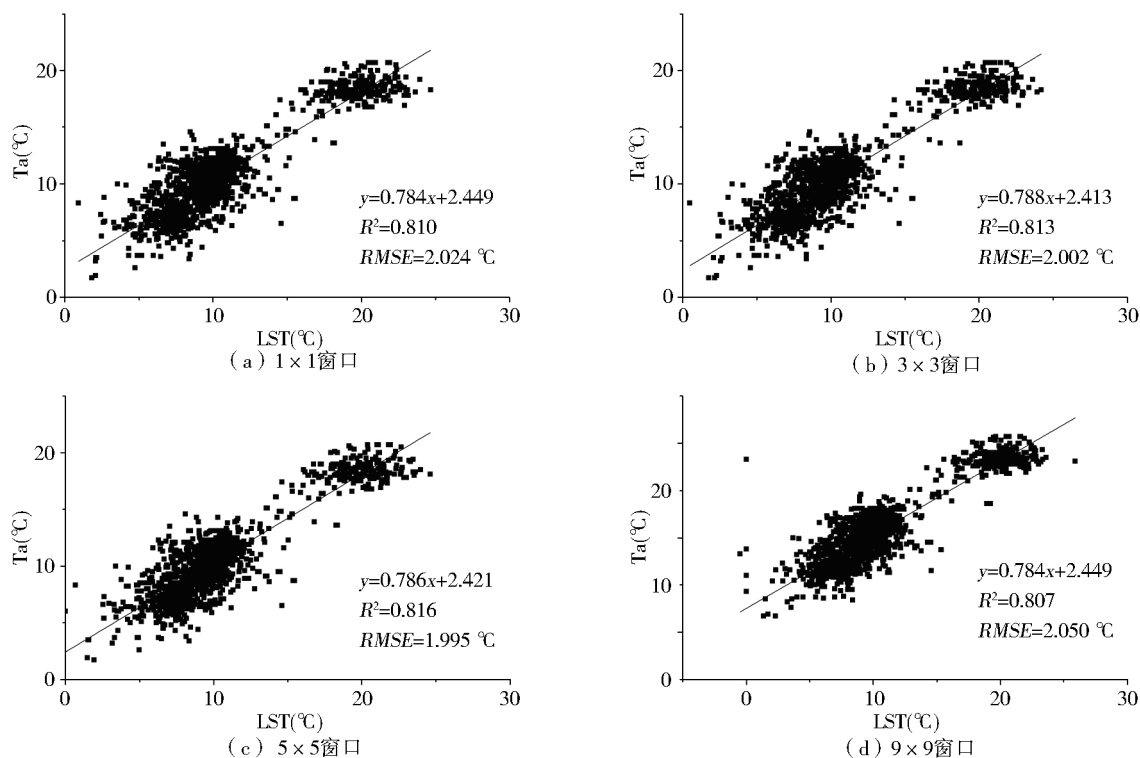


图 3 4 种大小窗口下的 LST 与 T_a 线性关系
Fig. 3 Correlations between LST and T_a of multiple windows

由图 3 可以看出, 随着窗口的增大, 地表温度与气温之间的相关性逐渐增加, R^2 呈现略微增大的趋势, 均方根误差 (Root-Mean-Square Error, RMSE) 略微减小, 然而当窗口增至 9×9 时, 相关性反而减

小, R^2 达到最低值, RMSE 达到最大。该现象表明窗口不能无限制增加, 最佳的窗口大小表示周围像元对站点所在的中心像元影响最准确的范围, 该范围的确定仍需进一步研究。总体而言, 本文反演的

地表温度 LST 与气温实际值均具有很好的相关性,可用于建立气温估算模型。

3.2 多元变量检验结果

在进行多元线性回归之前,本文通过逐步回归进行变量共线性诊断,找到并剔除引起多重共线性的干扰变量。以 1×1 窗口为基准,检验结果如表 1 所示,其中 Lon、Lat 分别代表经度和纬度。从表 1 可以看出,在逐步回归的第 5 个模型中,NDBI、NDVI 和 MNDWI 的容忍度均远远小于 0.1,且方差膨胀因子远大于 10;而 NDBI 进入之前,变量间并没有产生超过标准的共线性现象,因此可以判定 NDBI 的进入引起了与 NDVI、MNDWI 的严重共线性,最终建模时需要剔除这一变量。

3.3 MWLR 模型不确定性分析

为得到稳定可靠的模型,本文引入 10 折交叉验证的方法评价模型精度,并以测试集的预测值与实际观测值的 R^2 和 RMSE 为评价指标,表 2 为 4 种

窗口 10 次模型对应的 R^2 和 RMSE。

表 1 1×1 窗口共线性检验结果
Table 1 Result of collinearity diagnosis from 1×1 window

模型	引入因子	容忍度	VIF
1	LST	1	1
2	LST	0.996	1.004
	Lon	0.996	1.004
3	LST	0.992	1.008
	Lon	0.953	1.049
	MNDWI	0.955	1.047
4	LST	0.9	1.111
	Lon	0.92	1.087
	MNDWI	0.277	3.607
	NDVI	0.259	3.867
5	LST	0.881	1.135
	Lon	0.876	1.141
	MNDWI	0.018	55.131
	NDVI	0.028	35.926
	NDBI	0.062	16.039
6	LST	0.874	1.144
	Lon	0.838	1.193
	MNDWI	0.018	55.167
	NDVI	0.028	36.034
	NDBI	0.062	16.071
	Lat	0.859	1.165

表 2 4 种窗口下 10 次模型的 R^2 和 RMSE 比较
Table 2 Comparison of R^2 and RMSE for 10 models between multiple windows

窗口	评价指标	模型 1	模型 2	模型 3	模型 4	模型 5	模型 6	模型 7	模型 8	模型 9	模型 10
1×1	R^2	0.829	0.826	0.830	0.829	0.830	0.828	0.828	0.829	0.830	0.827
	RMSE	1.599	1.611	1.551	1.682	1.608	1.758	1.717	1.738	1.650	1.834
3×3	R^2	0.833	0.832	0.835	0.830	0.832	0.830	0.835	0.833	0.836	0.833
	RMSE	1.600	1.717	1.509	1.614	1.687	1.789	1.458	1.767	1.599	1.728
5×5	R^2	0.830	0.836	0.841	0.835	0.832	0.831	0.832	0.840	0.839	0.835
	RMSE	1.658	1.617	1.640	1.783	1.546	1.560	1.626	1.477	1.734	1.741
9×9	R^2	0.826	0.822	0.822	0.842	0.823	0.822	0.822	0.824	0.826	0.819
	RMSE	1.716	1.665	1.643	1.510	2.328	1.615	1.650	1.703	1.568	1.705

从表 2 可以看出: 1×1 窗口时, R^2 的范围是 0.827~0.830,最小值与最大值分别对应第 10 和第 3、5、9 次模型,均方根误差的范围是 1.551~1.834 $^{\circ}\text{C}$,对应第 3 次和第 10 次模型; 3×3 窗口时, R^2 的范围是 0.830~0.836,分别对应第 4、6 次和第 9 次模型,均方根误差的范围是 1.458~1.789 $^{\circ}\text{C}$,对应第 7 次和第 6 次模型; 5×5 窗口时, R^2 的范围是 0.830~0.841,分别对应第 1 和第 3 次模型,均方根误差的范围是 1.477~1.783 $^{\circ}\text{C}$,对应第 8 次和第 4 次模型; 9×9 窗口时, R^2 的范围是 0.819~0.842,分别对应第 10 和第 4 次模型,均方根误差的范围是

1.510~2.328 $^{\circ}\text{C}$,对应第 4 次和第 5 次模型。除了 9×9 窗口下第 5 次模型的均方根误差高达 2.328 $^{\circ}\text{C}$,其他窗口均在 1.4~1.8 $^{\circ}\text{C}$ 之间波动,说明 MWLR 模型得到的气温估算结果十分理想。

综合看, R^2 的最大值随着窗口尺寸的增加而增大,在 9×9 窗口时达到最大,为 0.842;然而模型预测的均方根误差的最小值却没有明显的规律,在 3×3 窗口时达到最小,然后又呈现增大的趋势。而且 R^2 最大值对应的模型与均方根误差最小值对应的模型并不完全一致,最终选择均方根误差最小的一次模型作为各窗口下的最优模型,如表 3 所示。

表 3 4 种窗口下的最优模型参数
Table 3 Parameters of the optimal models for multiple windows

窗口	RMSE _{min} ($^{\circ}\text{C}$)	R^2	模型
1×1	1.551	0.830	$Ta = 24.76 + 0.33Lat - 0.27Lon + 0.80LST + 3.27NDVI + 3.66MNDWI$
3×3	1.458	0.835	$Ta = 25 + 0.32Lat - 0.27Lon + 0.81LST + 3.62NDVI + 3.06MNDWI$
5×5	1.477	0.840	$Ta = 21.93 + 0.33Lat - 0.25Lon + 0.80LST + 3.71NDVI + 2.86MNDWI$
9×9	1.510	0.842	$Ta = 23.9 + 0.33Lat - 0.27Lon + 0.80LST + 4.04NDVI + 3.21MNDWI$

由表 3 可以看出,在建立的 MWLR 模型中,只有站点所在的经度与 Ta 呈负相关,说明浙江北部冬季时节的气温随着经度的增加而降低,即由内陆向

沿海地区温度递减。最后,将 3×3 窗口下的最优模型与空间内插方法进行了精度比对:选取 70% 的自动气象站数据采用反距离权重法^[39]进行内插,以剩

余的 30% 自动气象站的气温数据为测试集, 分别计算内插气温和 MWLR 模型计算结果与测试集实际气温之间的误差, 发现内插结果的 $RMSE$ 为 2.24°C , 而 MWLR 模型结果对应的 $RMSE$ 仅为 1.31°C , 说明 MWLR 模型精度明显优于一般的空间内插方法。这是因为传统空间内插方法考虑的因素较少, 加上过度依赖已有的气象站点数据, 在气象站点稀少的地区, 难以保证远离气象站区域的内插精度, 其结果并不能很好地体现出气温的空间异质性, 相对于本文方法精度较低^[40-42]。

4 结论与讨论

本文基于 Landsat8 数据耦合了地表温度、经纬度、归一化植被指数、归一化建筑指数和改进的归一化水体指数等因子, 结合自动气象站的气温数据建立了多窗口线性回归模型 (MWLR), 并有效利用该模型估算了浙江北部冬季的气温, 最后利用 10 折交叉验证的方法对估算精度进行了评价。结果显示: 1) 利用辐射传输方程法反演得到的 LST 与 T_a 呈现出良好的相关性, 其中 5×5 窗口下的相关性最高。2) 共线性检验发现 NDBI 可引起严重的共线性, 因此在多元回归分析中将其剔除。3) 通过对比不同窗口下的最优模型, 发现预测模型的均方根误差在 $1.458 \sim 1.551^{\circ}\text{C}$ 之间, 3×3 窗口下的误差最小; 然而模型的 R^2 却随着窗口尺寸的增大略有增加, 最大值达到 0.842, 说明窗口大小对于气温估算模型具有一定的影响, 在一定范围内调整窗口大小能够提高模型的精度。4) 在建立的 MWLR 模型中, 只有经度与气温呈负相关, 说明浙江北部冬季气温随经度增加而降低, 符合我国冬季气温由内陆向沿海递减的趋势。5) 通过与反距离加权法内插气温结果进行对比, 发现 MWLR 模型计算的气温误差更小, 说明 MWLR 模型估算精度优于普通的空间内插方法。

本文在研究过程中遇到两大难题, 一是 Landsat8 数据反演地表温度方法选择的问题。目前基于 Landsat8 热红外波段的温度反演还没有研究出最佳的方法^[24,25,27], 考虑到 TIRS Band11 定标参数的不确定性太大, 最终选择辐射传输方程法, 反演得到的地表温度与气温实际值的回归决定系数高达 0.816, 效果很好。二是气温预测模型的因子选择问题。近地表气温值受到很多因素的影响, 前人研究中也加入太阳天顶角、高程、风速等因子^[1,7], 然而本文的研究区域是浙北平原地区, 因此没有加入高程信息; 受研究资料的限制, 最终选择了站点经纬度、归一化

植被指数 (NDVI)、归一化建筑指数 (NDBI) 和改进的归一化水体指数 (MNDWI), 与地表温度作为多元变量进行建模, 最终模型的预测精度较高。

本研究证实了基于 MWLR 模型和 Landsat8 数据进行气温估算的可行性, 相比普通的空间内插方法估算更为准确, 为局部地区气温估算提供了一种方法。后续研究可以考虑加入水汽压、地表高程及精确的土地利用分类等数据, 增加气温估算模型的因子并提高地表温度反演的精度, 进而提高气温估算的精度和实用性。

参考文献:

- [1] 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(3): 540—552.
- [2] PRIHODKO L, GOWARD S N. Estimation of air temperature from remotely sensed surface observations[J]. Remote Sensing of Environment, 1997, 60(3): 335—346.
- [3] GUO G M, ZHOU M. Using MODIS land surface temperature to evaluate forest fire risk of northeast China[J]. IEEE Geoscience & Remote Sensing Letters, 2004, 1(2): 98—100.
- [4] KARNIELI A, PANOVA N, GOLDBERG A, et al. Use of NDVI and land surface temperature for drought assessment: Merits and limitations[J]. Journal of Climate, 2010, 23(3): 618—633.
- [5] VANCUTSEM C, CECCATO P, DINKU T, et al. Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa[J]. Remote Sensing of Environment, 2010, 114(2): 449—465.
- [6] KLOOG I, CHUDNOVSKY A, KOUTRAKIS P, et al. Temporal and spatial assessments of minimum air temperature using satellite surface temperature measurements in Massachusetts, USA[J]. Science of the Total Environment, 2012, 432(5): 85—92.
- [7] 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.
- [8] SUN Y J, WANG J F, ZHANG R H, et al. Air temperature retrieval from remote sensing data based on thermodynamics[J]. Theoretical & Applied Climatology, 2005, 80(1): 37—48.
- [9] ZAKŠEK K, SCHROEDTER-HOMSCHEIDT M. Parameterization of air temperature in high temporal and spatial resolution from a combination of the SEVIRI and MODIS instruments[J]. Isprs Journal of Photogrammetry & Remote Sensing, 2009, 64(4): 414—421.
- [10] CZAJKOWSKI K P, GOWARD S N, STADLER S J, et al. Thermal remote sensing of near surface environmental variables: Application over the Oklahoma Mesonet[J]. Professional Geographer, 2000, 52(2): 345—357.
- [11] CRESSWELL M P, MORSE A P, THOMSON M C, et al. 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.
- [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] ZHU W, LÜA, JIA S. Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products

- [J]. Remote Sensing of Environment, 2013, 130: 62—73.
- [14] 徐剑波, 赵凯, 赵之重, 等. 利用 HJ-1B 遥感数据反演西北地区近地表气温[J]. 农业工程学报, 2013(22): 145—153.
- [15] 徐永明, 覃志豪, 沈艳. 基于 MODIS 数据的长江三角洲地区近地表气温遥感反演[J]. 农业工程学报, 2011, 27(9): 63—68.
- [16] 徐永明, 覃志豪, 万洪秀. 热红外遥感反演近地层气温的研究进展[J]. 国土资源遥感, 2011(1): 9—14.
- [17] PAPE R, LÖFFLER J. Modelling spatio-temporal near-surface temperature variation in high mountain landscapes[J]. Ecological Modelling, 2004, 178(3—4): 483—501.
- [18] KAWASHIMA S, ISHIDA T, MINOMURA M, et al. Relations between surface temperature and air temperature on a local scale during winter nights[J]. Journal of Applied Meteorology, 2000, 39(9): 1570—1579.
- [19] CRISTÓBAL J, JIMÉNEZ-MUÑOZ J C, SOBRINO J A, et al. Improvements in land surface temperature retrieval from the Landsat series thermal band using water vapor and air temperature[J]. Journal of Geophysical Research Atmospheres, 2009, 114(D08103).
- [20] 徐伟燕, 孙睿, 金志凤, 等. 基于 MODIS 数据的近地表气温估算[J]. 气象与环境科学, 2015, 38(1): 1—6.
- [21] 姚永慧, 张百平. 基于 MODIS 数据的青藏高原原气温与增温效应估算[J]. 地理学报, 2013, 68(1): 95—107.
- [22] HOU P, CHEN Y, QIAO W, et al. Near-surface air temperature retrieval from satellite images and influence by wetlands in urban region[J]. Theoretical & Applied Climatology, 2012, 111(1—2): 109—118.
- [23] 徐涵秋, 唐菲. 新一代 Landsat 系列卫星: Landsat 8 遥感影像新增特征及其生态环境意义[J]. 生态学报, 2013, 33(11): 3249—3257.
- [24] 宋挺, 段峥, 刘军志, 等. Landsat 8 数据地表温度反演算法对比[J]. 遥感学报, 2015, 19(3): 451—464.
- [25] 蒋大林, 匡鸿海, 曹晓峰, 等. 基于 Landsat8 的地表温度反演算法研究——以滇池流域为例[J]. 遥感技术与应用, 2015, 30(3): 448—454.
- [26] 罗红霞, 邵景安, 张雪清. 基于辐射传导方程的三峡库区腹地地表温度的遥感反演[J]. 资源科学, 2012, 34(2): 256—264.
- [27] 徐涵秋. Landsat 8 热红外数据定标参数的变化及其对地表温度反演的影响[J]. 遥感学报, 2016, 20(2): 229—235.
- [28] 家会臣, 靳竹萱, 李济洪. Logistic 模型选择中三种交叉验证策略的比较[J]. 太原师范学院学报(自然科学版), 2012(1): 87—90.
- [29] 丁凤, 徐涵秋. TM 热波段图像的地表温度反演算法与实验分析[J]. 地球信息科学学报, 2006, 8(3): 125—130.
- [30] VAN DE GRIEND A A, OWE M. On the relationship between thermal emissivity and the normalized difference vegetation index for natural surfaces[J]. International Journal of Remote Sensing, 1993, 14(6): 1119—1131.
- [31] 吴志刚, 江滔, 樊艳磊, 等. 基于 Landsat8 数据的地表温度反演及分析研究——以武汉市为例[J]. 工程地球物理学报, 2016, 13(1): 135—142.
- [32] SOBRINO J A, RAISSOUNI N, LI Z L. A comparative study of land surface emissivity retrieval from NOAA data[J]. Remote Sensing of Environment, 2001, 75(2): 256—266.
- [33] 覃志豪, 李文娟, 徐斌, 等. 陆地卫星 TM6 波段范围内地表比辐射率的估计[J]. 国土资源遥感, 2004, 16(3): 28—32.
- [34] SHENG Y L, LIU X P, YANG X C, XIN Q C, DENG C B, Li X. Quantifying the spatial and temporal relationship between air and land surface temperatures of different land-cover types in Southeastern China[J]. International Journal of Remote Sensing, 2017, 38(4): 1114—1136.
- [35] 查勇, 倪绍祥, 杨山. 一种利用 TM 图像自动提取城镇用地信息的有效方法[J]. 遥感学报, 2003, 7(1): 37—40.
- [36] 徐涵秋. 利用改进的归一化差异水体指数(MNDWI)提取水体信息的研究[J]. 遥感学报, 2005, 9(5): 589—595.
- [37] 鲁茂, 贺昌政. 对多重共线性问题的探讨[J]. 统计与决策, 2007(8): 6—9.
- [38] ARLOT S. V-fold cross-validation improved; V-fold penalization[J]. Mathematics, 2008.
- [39] 庄立伟, 王石立. 东北地区逐日气象要素的空间插值方法应用研究[J]. 应用气象学报, 2003, 14(5): 605—615.
- [40] ISHIDA T, KAWASHIMA S. Use of Cokriging to estimate surface air temperature from elevation[J]. Theoretical and Applied Climatology, 1993, 47(3): 147—157.
- [41] STAHL K, MOORE R D, FLOYER J A, et al. Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density[J]. Agricultural & Forest Meteorology, 2006, 139(3): 224—236.
- [42] 姚永慧, 张百平, 韩芳. 基于 Modis 地表温度的横断山区气温估算及其时空规律分析[J]. 地理学报, 2011(7): 917—927.

Estimation of Winter Temperature in Northern Zhejiang Province Based on Multi-Window Linear Regression Model

WANG Yue, LIU Xiao-ping, LI Xia, YAO Yao, SHENG Yan-ling

(Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, School of Geography and Planning of Sun Yat-Sen University, Guangzhou 510275, China)

Abstract: Air temperature(T_a) is a key descriptor of the terrestrial environment, and how to estimate T_a of high accuracy is a hotspot since the abnormal change of T_a has influence on the living environment of human beings. Considering MODIS data cannot offer much precise surface information, in this paper, a Multi-Window Linear Regression Model(MWLR) is established using Landsat8 images of higher resolution by coupling latitude and longitude, normalized difference vegetation index, normalized difference building index and modified normalized difference water index with T_a obtained from automatic meteorological stations. Finally, the MWLR model is used to estimate T_a of northern Zhejiang Province in winter, and RMSE of prediction is between 1.458~1.551 °C, meanwhile, R^2 reaches 0.835~0.842. The highest precision is achieved when the window size is 3×3 (RMSE=1.458 °C, $R^2=0.835$). In a word, the result shows the accuracy of estimating T_a based on MWLR model and Landsat8 data, which means our model is appropriate to develop high precision and resolution T_a estimation over local area using remote sensing data.

Key words: multi-window linear regression model; air temperature estimation; Landsat8