

引用格式: Wang Meiya, Xu Hanqiu. Study on the Random Forest Regression Model of Land Cover and Thermal Environment in Megacities[J]. Remote Sensing Technology and Application, 2022, 37(2): 379-388. [王美雅, 徐涵秋. 超大城市土地覆盖与热环境的随机森林回归模型研究[J]. 遥感技术与应用, 2022, 37(2): 379-388.]
DOI: 10.11873/j.issn.1004-0323.2022.2.0379

超大城市土地覆盖与热环境的随机森林回归模型研究

王美雅¹, 徐涵秋²

(1. 闽南师范大学 历史地理学院, 福建 漳州 363000;
2. 福州大学 环境与资源学院, 福州大学遥感信息工程研究所,
福建省水土流失遥感监测评估与灾害防治重点实验室, 福建 福州 350116)

摘要: 目前对于超大城市土地覆盖和热环境定量模型研究报道不足, 这主要是因为大城市地表温度和地表生物物理组分之间存在复杂的潜在非线性关系, 这使得准确评估城市热环境情况遇到了严峻的技术挑战。研究选取中外6个典型超大城市(北京、上海、广州、伦敦、纽约和东京)为研究对象, 以Landsat遥感影像为主要数据源, 利用单通道算法反演各城市地表温度, 采用随机森林回归模型(RFR)建立土地覆盖类型与城市热环境定量关系模型(LCT), 综合分析城市土地覆盖因子与热环境间的多维定量关系。土地覆盖与地表温度的定量关系显示, 城市地表热场的空间结构在很大程度上被下垫面用地类型所左右, 不透水面会导致高温热场的聚集, 而植被和水体则有降温作用。6个超大城市地表覆盖结构变化产生的升温/降温效应有所差异, 北京、上海、纽约和东京等城市区域的植被和水体降温效应较广州和伦敦显著。基于随机森林回归方法建立了NDVI、MNDWI和NDISI等3种土地覆盖类型与城市热环境的综合定量关系模型(LCT), 模型得到的精度高于基于多元线性回归方法建立的模型。LCT_RF模型的 R^2 值在0.623~0.826之间, 比LCT_MLR模型高0.021~0.074; RMSE比LCT_MLR模型低0.07℃~0.35℃。研究超大城市土地覆盖与城市热环境的互动作用机理, 能为未来生态城市建设提供宝贵建议。

关键词: 超大城市; 土地覆盖; 城市热环境; 遥感; 随机森林回归模型

中图分类号: P237 **文献标志码:** A **文章编号:** 1004-0323(2022)02-0379-10

1 引言

超大城市建筑和道路等不透水面分布规模大且密集, 加上大量人为活动产生了大体量的水和热等城市代谢, 这种“超大性”对城市生态系统产生了强大外力干扰, 改变了城市地表热量平衡, 其产生的城市热岛问题远大于其他中小城市和乡村地区^[1-4]。探究超大城市不同地表覆盖类型与地表温

度的复杂非线性关系, 有助于提高城市地表生物物理组分与地表辐射能量平衡互馈关系内在机理的认知能力, 为城市生态规划与整治等提供科学参考。这其中, 选择合适方法建立地物指数和地表温度间的综合定量模型是研究关键。

遥感技术覆盖面广、分辨率高、处理速度快, 利用遥感技术来监测城市下垫面性质, 可以克服传统地面观测站点资料数据极其有限的问题, 实现海量

收稿日期: 2021-10-21; 修订日期: 2022-02-21

基金项目: 国家重点研发计划专项课题(2016YFA0600302), 福建省创新战略研究项目(2020R0155), 闽南师范大学校长基金项目(KJ19013)。

作者简介: 王美雅(1991—), 女, 福建泉州人, 博士, 副教授, 主要从事环境与资源遥感研究。E-mail: 286097145@qq.com

通讯作者: 徐涵秋(1955—), 男, 江苏盐城人, 博士, 教授, 主要从事环境与资源遥感研究。E-mail: hxu@fzu.edu.cn

“面”数据的获取^[5]。此外,多时相、多光谱、多平台的遥感卫星影像能保证在全球范围内提供准确、一致的数据,被广泛用于城市热环境研究。从原始的目视解译到海量遥感影像分类、遥感因子回归分析等问题引入了众多模式识别方法,如最大似然法、最小距离法、KNN(K-Nearest Neighbor)算法、ANN(Artificial Neural Network)、支持向量机、随机森林、遗传算法和面向对象等算法等^[6],可以看出与日俱增的海量遥感影像推动着遥感图像处理算法的进步。

快速城市化导致的城市热环境变化,产生城市热岛效应^[7]。一方面,城市人口的大幅增加,城市规模不断扩大,能耗需求剧增,人为温室气体排放总量持续增加。另一方面,城市化过程中大量的建筑、水泥路等不透水面取代了植被、水体等生态用地,导致地表蒸腾明显减少,显热表面比例升高而潜热表面比例降低,促使城市热岛效应日益严峻^[8-9]。众多学者围绕城市热环境变化进行研究分析。Zheng等^[10]则在利用Moran'I指数对凤凰城的地表温度和LULC类型的相关性研究中发现,相对于城市建筑物的空间格局,其外立面不同涂料的区别对地表温度的影响更大,且夜间的正相关性更高。Chen等^[11]运用Landsat影像反演广州城市地表温度,进一步计算城市热变异区域,并采用多种景观格局指数分析城市热变异区的特征。结果表明,城市地物组成对城市热变异区的影响最大,地表景观空间格局也会影响城市热变异区。Kotharkar等^[10]分析印度那格浦尔市局地气候分区(Local Climate Zones, LCZ)的热岛情况,结果表明冬季建筑用地气候区内的UHII介于1.76℃~4.09℃之间,城区中紧凑型低层LCZs区比城市内部的其它主要LCZs区的气候暖和。

植被、不透水面和水体等三大土地覆盖类型是城市复合生态系统的重要组成部分,其比例和空间变化很大程度地影响城市热环境状况。植被能够选择性吸收和反射太阳辐射,调节城市地表潜热和显热的交换,从而起到调节城市温度作用^[13]。研究土地覆盖变化导致的城市热岛变化情况,结果表明,增加绿地面积对缓解局部热岛的范围和强度具有重要作用。水体对周边环境具有降温、增湿的效应^[14-15]。在不透水面变化对城市热环境的影响研究中,国内外学者重点关注不透水面对城市热环境的影响及其与地表温度的定量关系^[16-18]。不透水面比

例的增加,使地表热容和导热率增加,地表温度上升。

然而目前对于超大城市土地覆盖和热环境定量模型研究报道不足,这主要是因为大城市地表温度和地表生物物理组分之间存在复杂的潜在非线性关系,这使得准确评估城市热环境情况遇到了严峻的技术挑战。因此,本研究选取中外6个典型超大城市,研究超大城市土地覆盖与城市热环境的互动作用机理,能为未来生态城市建设提供宝贵建议。

2 研究区概况

研究区为中外6个典型超大城市(北京、上海、广州、伦敦、纽约和东京)。北京、上海、广州是中国城市规模排名前3的超大城市,伦敦、纽约和东京是城市规模居世界前列的3个超大城市,分别分布在欧洲、美洲和亚洲。6个超大城市的城市规模相当,人口众多,GDP高,超大城市的“超大性”产生的生态环境问题远大于其他中小城市和乡村地区。由于中、美、英、日等国的行政区划差异较大,以6个城市2015年的建成区外扩5 km缓冲区的范围作为研究区(图1),以此划定具有对比性的城市边界^[19]。

3 研究方法

将植被、水体和不透水面等3大主要地表覆盖类型的遥感指数与城市热环境进行回归统计,定量分析中外6个超大城市的城市地表参数与城市热环境的互动关系。

3.1 地表温度反演

主要研究数据为Landsat 8卫星影像,北京、上海、广州、伦敦、纽约和东京的影像获取日期分别为2015-08-22、2015-08-03、2015-10-18、2015-10-02、2015-08-26和2015-10-09。采用Chander和Chavez的模型和参数以及Landsat 8网站提供的公式和参数对Landsat 8影像进行辐射校正,将影像的DN值转换为传感器处反射率^[20-21]。

地表温度反演采用Jiménez-Muñoz等提出的单通道算法^[22-24],根据Chander和Chavez的模型和参数以及Landsat 8网站提供的公式和参数,将影像的亮度值转换成大气顶部的光谱辐射值。进一步通过下式转换为亮温(T_{sensor}),最后反演出地表温度(Land Surface Temperature, LST):

$$T_{\text{sensor}} = K_2 / \ln(K_1 / L_{\text{sensor}} + 1) \quad (1)$$

$$\text{LST} = \gamma [\varepsilon^{-1} (\psi_1 L_{\text{sensor}} + \psi_2) + \psi_3] + \delta \quad (2)$$

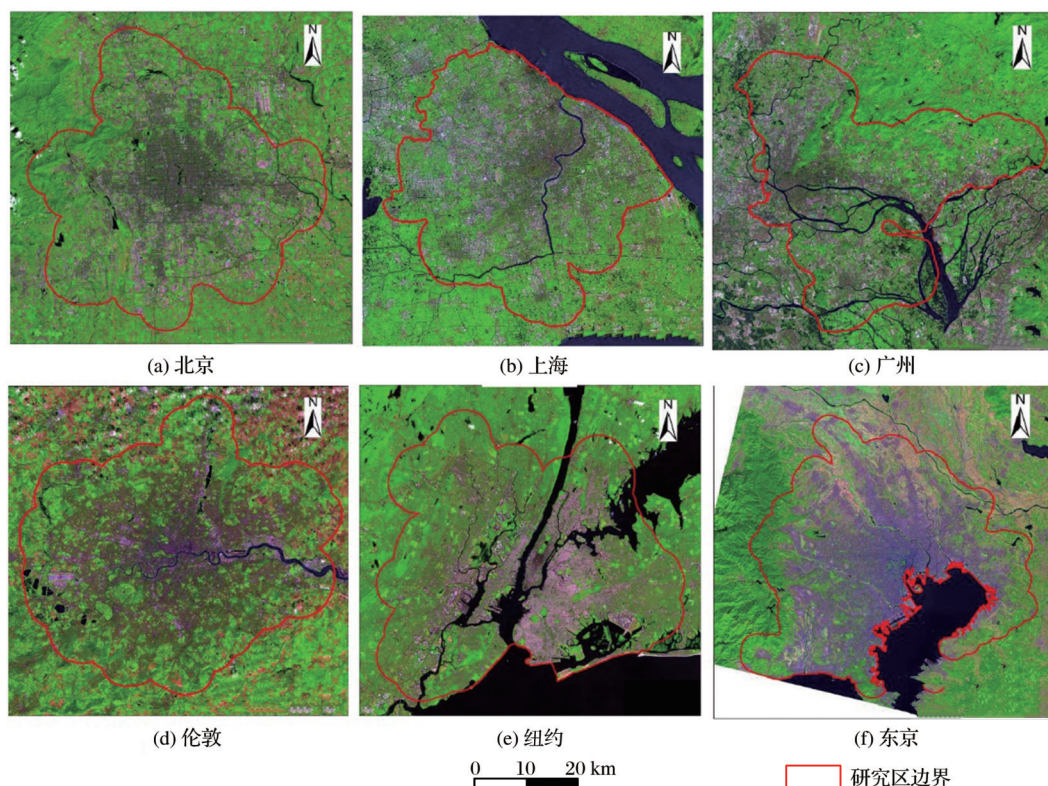


图1 研究区 Landsat 遥感影像

Fig.1 Landsat images of the study area

其中: T_{sensor} 为传感器处亮温值; K_1 和 K_2 为热红外波段的定标常数; L_{sensor} 为热红外波段的辐射值; γ 和 δ 是基于 Planck 函数的 2 个参数; ϵ 为地表比辐射率, ψ_1 、 ψ_2 、 ψ_3 是大气水汽含量的函数。

3.2 主要地物遥感指数

选用归一化植被指数(NDVI)、改进型归一化差值水体指数(MNDWI)、归一化差值不透水面指数(NDISI)来分别表征不透水面、植被、水体等 3 种主要地表覆盖类型,并量化这 3 种因子与地表温度的二维关系。其中,NDVI 是植被研究中应用最为广泛的一种指数,是植物生长状况、植被生物量和植被覆盖度的较好指示因子^[25];MNDWI 指数能很好地提取城市污染和浑浊水体,剔除土壤、阴影、植被等信息^[26];NDISI 指数能有效地增大土壤和不透水面的光谱差异,提取不透水面信息^[27]。这 3 个指数已被广泛运用于表征土地覆盖信息,具有代表性和可靠性^[25-28]。

植被指数: $\text{NDVI} = (\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$ (3)

水体指数: $\text{MNDWI} = (\text{Green} - \text{MIR1}) / (\text{Green} + \text{MIR1})$ (4)

不透水面指数^[28]:

$$\text{NDIST} = \frac{\text{TIR} - (\text{MNDWI} + \text{NIR} + \text{MIR1})/3}{\text{TIR} + (\text{MNDWI} + \text{NIR} + \text{MIR1})/3} \quad (5)$$

其中: Green、Red、NIR、MIR1 分别代表绿光、红光、近红外和中红外 1 波段;MNDWI 为新型归一化水体指数,TIR 为热红外波段。

3.3 土地覆盖与城市热环境的综合定量关系模型构建

城市地表温度和地表生物物理组分之间存在复杂的潜在非线性关系,这使得准确评估城市热环境情况遇到了严峻的技术挑战。研究采用随机森林回归模型来模拟植被、水体和不透面对地表温度的综合作用关系,建立土地覆盖类型与城市热环境定量关系模型(Land Cover-LST Regression, LCT),并与传统的多元线性回归模型进行对比。

3.3.1 模型输入样本选择

用于建模与检验的样本需要充分客观地代表各研究区的土地覆盖类型和对应的地表温度特征,为此,本研究采用网格采样的方法进行采样。选用 NDVI、MNDWI 和 NDISI 指数来表征植被、水体和不透水面比例,作为模型的 3 个输入因子。采用地物比例影像进行回归样本采样能有效地解决混合像元效应问题。首先将各研究区植被的 NDVI 影

像、水体的 MNDWI 影像和不透水面的 NDISI 影像归一化到 0~1 之间,并转换成百分率,使其数值分布在 0~100 之间^[28-29],然后根据研究区大小,采用适当尺寸的网格进行采样,北京、上海、广州、伦敦、纽约和东京所用的网格大小分别为 15×15、15×15、15×15、14×14、13×13 和 18×18,总采样点分别为 16 894 个、11 404 个、12 653 个、14 684 个、12 525 个和 17 080 个,统计出对应的植被比例、水体比例、不透水面比例和地表温度 LST,足够多的样品以确保回归分析的客观性。

3.3.2 随机森林回归模型构建

随机森林回归模型(Random Forest Regression, RFR)是一种基于决策树的集成学习算法。由于随机森林回归能够有效处理高维数据,且模型中需要调整的参数较少,只需设置两个参数:生成的决策树数量(Ntree)和决策树生长分裂时被选择用于测试的特征数目(Mtry),因此,比传统的多元回归分析法更具有优势^[30-31]。本研究引入随机森林回归模型(RFR)来解释土地覆盖与地表温度之间的综合定量关系。本研究将 6 幅影像的 NDVI、MNDWI 和 NDISI 指数作为自变量,LST 作为目标值,建立 6 幅影像的随机森林回归模型(LCT_RF)。根据前人研究成果,本研究将 Ntree 值设为 500^[32],Mtry 值设为 1(输入特征数的三分之一),模型表达式可简写为:

$$\text{LCT_RF} = \text{RFR}(\text{NDVI}, \text{MNDWI}, \text{NDISI}) \quad (6)$$

图 2 给出随机森林回归模型的构建流程图。

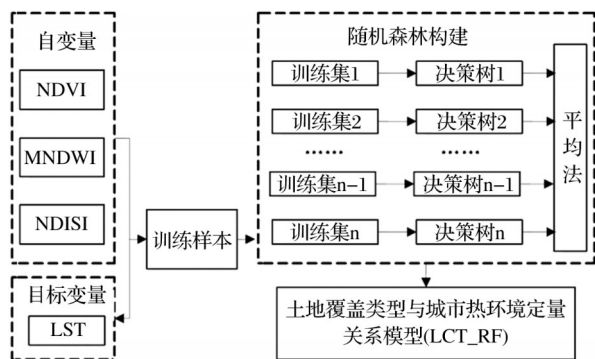


图 2 随机森林回归模型建立过程

Fig.2 Process of random forest regression model

研究对比了随机森林模型与多元线性回归模型对 LST 的解释能力,以均方根误差(Root Mean Square Error, RMSE)来定量评价模型的预测精度。RMSE 反映的是预测值与真实值之间的偏离程度。把卫星影像反演的 LST 作为真实的 LST 来评价模

型的预测效果。此外,研究也绘制了 2 种模型真实值与预测值的 1:1 的散点图来评价模型的精度,如果模型具有较高的预测精度,其 1:1 的散点图会显示 45° 的直线趋势。

4 研究结果

4.1 地表温度反演结果

依据上述 LST 反演算法组合以及相关关键参数的获取和计算方法,得到所研究 6 个超大城市的 LST 反演结果(图 3)。

研究涉及多城市多区域的地表温度数据,通过收集标准气象站点历史温度数据进行验证存在一定困难。因此,研究利用同日的 MODIS 地表温度产品进行精度验证。通过美国航空航天局(NASA)获取与所反演 6 个城市各时期 LST 影像同日的 MODIS LST 逐日数据(MOD11A1),统计对比各研究区相同区域内的地表温度均值,验证结果见表 1。通过取温度均值进行对比的方法可以避免反演的温度和 MODIS LST 数据空间分辨率上不同而产生的差异。结果表明(表 1),所反演的 Landsat LST 与 MODIS LST 的均值比较接近,除了伦敦 2015 年 10 月 2 日和东京 2015 年 10 月 9 日的平均温差绝对值略大于 1.0℃,其余影像的平均温差均小于 1.0℃。这说明,地表温度反演结果总体上符合精度要求。有一定差距的原因可能是由于两种卫星的过空时间不完全相同所致。

4.2 3 种地表覆盖类型与地表温度的二维关系分析

采用线性和非线性多种拟合分析来定量刻画不同地表覆盖物变化的升温/降温效应。对比线性、二次多项式、对数、指数、幂函数等 5 种拟合曲线函数,选择最佳拟合回归函数分析 3 种地物类型与地表温度间的关系。根据拟合结果,选择线性回归方程表示植被和水体与地表温度的定量关系,选择指数回归方程表示不透水面与地表温度的定量关系。

回归分析表明(图 4),植被比例和地表温度呈显著负线性相关关系,所有线性回归方程的 r 绝对值也都在 0.72 以上(图 4(a)~图 4(f));水体比例和地表温度呈线性负相关关系,线性回归方程的 r 绝对值均在 0.5 以上(图 4(g)~图 4(l))。不透水面比例和地表温度呈显著的指数正相关关系,指数回归方程的 r 值均在 0.7 以上(图 4(m)~图 4(r))。这说明,植被和水体比例的增加会使地表温度降低,而不透

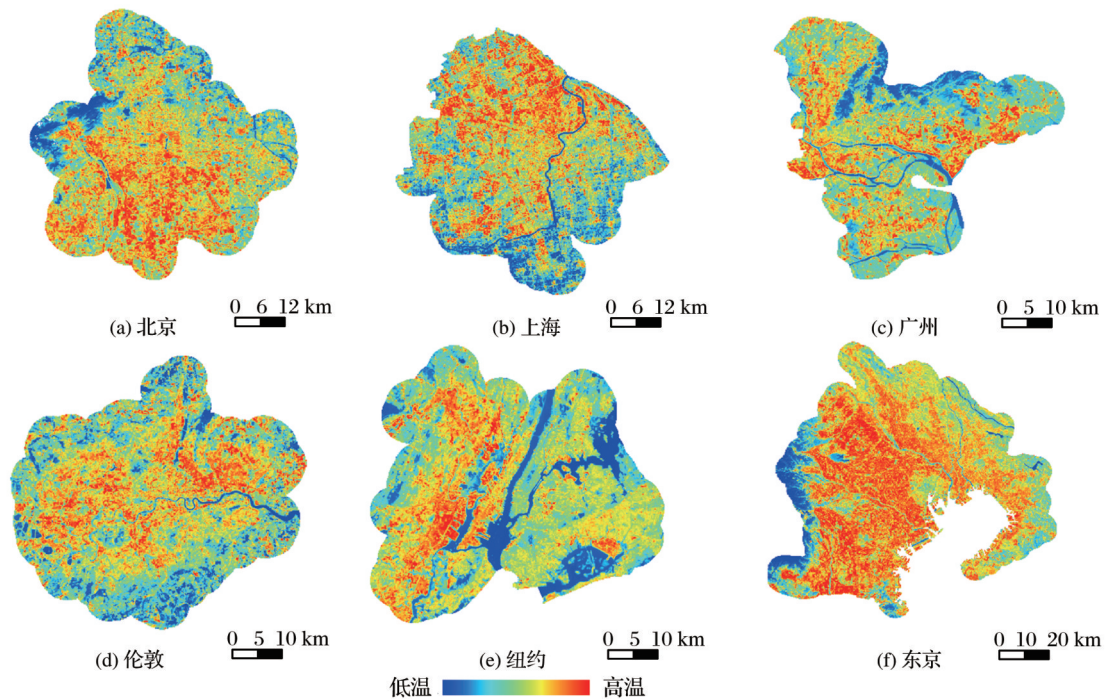


图 3 6个超大城市地表温度反演结果

Fig.3 LST maps of the six megacities

表 1 地表温度反演精度对比

Table 1 Results comparison of retrieve accuracy of LST

城市	日期	过空时间	卫星	LST 均值/ $^{\circ}\text{C}$	平均 LST 差值/ $^{\circ}\text{C}$
北京	2015/8/22	11:48a.m.	MODIS	37.691	0.135
		10:53a.m.	Landsat	37.556	
上海	2015/8/3	11:36a.m.	MODIS	38.557	-0.855
		10:24a.m.	Landsat	37.702	
广州	2015/10/18	10:48a.m.	MODIS	32.767	0.823
		10:52a.m.	Landsat	33.590	
伦敦	2015/10/2	10:24a.m.	MODIS	19.177	-1.312
		10:58a.m.	Landsat	17.865	
纽约	2015/8/26	10:12a.m.	MODIS	31.600	0.869
		10:39a.m.	Landsat	32.469	
东京	2015/10/9	11:42a.m.	MODIS	26.745	1.314
		10:15a.m.	Landsat	28.059	

水面比例的升高会导致地表温度的上升,且随着不透水面比例的增加,地表温度上升的幅度增加。

不同城市地表覆盖升温/降温效应有所差别,在北京、上海、纽约和东京等城市区域,植被和水体增加产生的降温效应较明显,植被比例每提升 10%,地表温度可降低 $1^{\circ}\text{C}\sim 1.6^{\circ}\text{C}$;水体比例每提升 10%,地表温度可降低 0.7°C 以上。广州和伦敦城市区域,植被比例每提升 10%,地表温度降低 $0.5^{\circ}\text{C}\sim 0.9^{\circ}\text{C}$;水体比例每提升 10%,地表温度降低 $0.2^{\circ}\text{C}\sim$

0.4°C 。北京、上海和纽约等城市区域的不透水面增加产生的升温效应也高于广州、伦敦和东京等区域,当 $\text{NDISI} \leq 30\%$ 时,不透水面比例每增加 10%,只导致地表温度增加 1.8°C 左右,但当 $\text{NDISI} \geq 80\%$ 时,不透水面比例每增加 10%,地表温度的增加幅度会达 2.4°C 以上。在广州和东京等城市区域,当 $\text{NDISI} \leq 30\%$ 时,不透水面比例每增加 10%,只导致地表温度增加 1.3°C 左右,但当 $\text{NDISI} \geq 80\%$ 时,不透水面比例每增加 10%,地表温度的增加幅度会达 1.8°C 以上。升温效应相对低的伦敦,地表温度随不透水面比例增加的上升幅度在 $0.7^{\circ}\text{C}\sim 0.9^{\circ}\text{C}$ 左右。此外,从北京、广州、纽约和东京植被比例和地表温度的回归散点图看出,在 NDVI 高值 ($\text{NDVI} > 80\%$) 区域,回归曲线的下降速率比 NDVI 低值区域有所升高。这说明,在高植被比例 ($\text{NDVI} > 80\%$) 情况下,这些城市区域植被的降温效应好于低植被比例的情况。结合这 4 个城市的遥感影像(图 1)可以发现,北京、广州、纽约和东京研究区中的 NDVI 高值区域对应的为城郊地区覆盖林地的山体区域,而上海和伦敦城市周边则无大片山体林地,这在一定程度说明了大面积林地的降温效应明显好于其他植被类型。

4.3 LCT 模型结果分析

基于 3 种土地覆盖比例指数和随机森林回归模

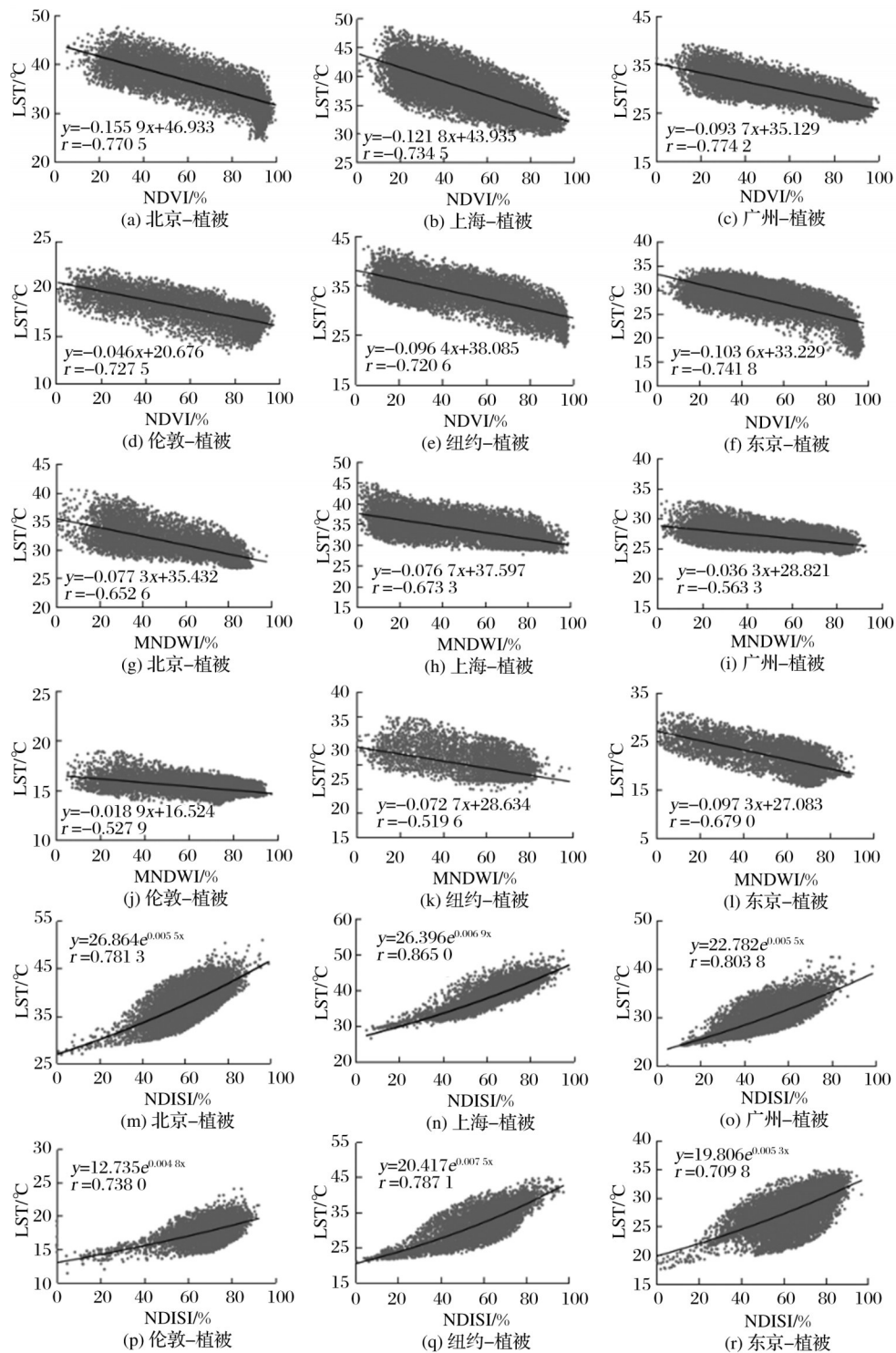


图4 植被、水体和不透水面比例与地表温度回归散点图

Fig.4 Relationship of NDVI, MNDWI, NDISI with LST

型、多元线性回归模型分别对中外6个超大城市的地表温度进行反演,得到基于RF方法的土地覆盖类型与地表温度的回归模型(LCT_RF)和基于MLR方法的回归模型(LCT_MLR)。LCT_RF模型和LCT_MLR模型都使用了中心像元的植被比例(NDVI)、水体比例(MNDWI)和不透水面比例(NDISI)

作为模型输入自变量,且这3个自变量都通过了1%的显著性检验,说明3个变量都是影响地表温度的重要因子。模型获得的预测LST与遥感数据反演的LST进行精度检验,计算了 R^2 和RMSE 2个统计指标来评价模型的精度(表2)。图5~6分别为LCT_RF和LCT_MLR模型的1:1散点图。

表2 LCT_RF和LCT_MLR方法预测LST精度对比
Table 2 Results comparison of LCT_RF and LCT_MLR model based on LST prediction

城市	LCT_RF(3指数)		LCT_MLR(3指数)	
	R^2	RMSE/℃	R^2	RMSE/℃
北京	0.656	1.646	0.635	1.956
上海	0.775	1.347	0.754	1.592
广州	0.726	1.507	0.686	1.613
伦敦	0.623	0.882	0.559	0.954
纽约	0.826	1.814	0.752	2.164
东京	0.761	1.636	0.698	1.897

结果表明,LCT_MLR模型的精度低于相应的LCT_RF模型。LCT_RF模型的 R^2 值比LCT_

MLR模型高0.02~0.07;LCT_RF模型检验的RMSE则比LCT_MLR模型低0.1℃~0.4℃。这说明,LCT_RF回归方法所建立的非线性模型在不同城市土地覆盖对LST的预测上都表现出比线性模型更高的精度,而其算法的效率也较高。综上所述,LCT_RF模型通过对非线性关系的模拟,比传统的多元线性回归模型对复杂城市环境内LST的变异提供更好的解释。

从LCT_RF模型和LCT_MLR模型真实值与预测值的1:1的散点图可以看出(图5~6),LCT_RF模型预测值与实际值的1:1的散点图更接近于45°的直线,LCT_MLR模型的1:1散点图偏离

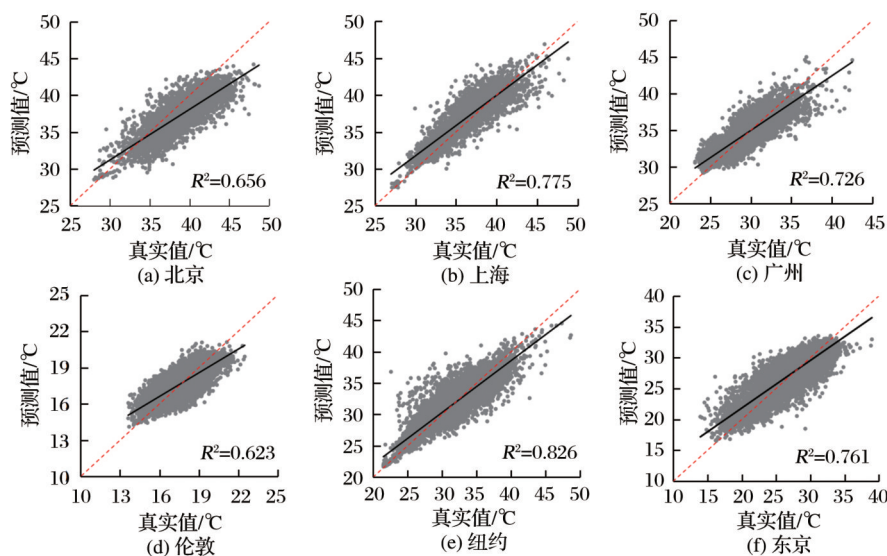


图5 6个城市LCT_RF(3指数)模型结果

Fig.5 Results comparison of LCT_RF(3 indices) of the six megacities

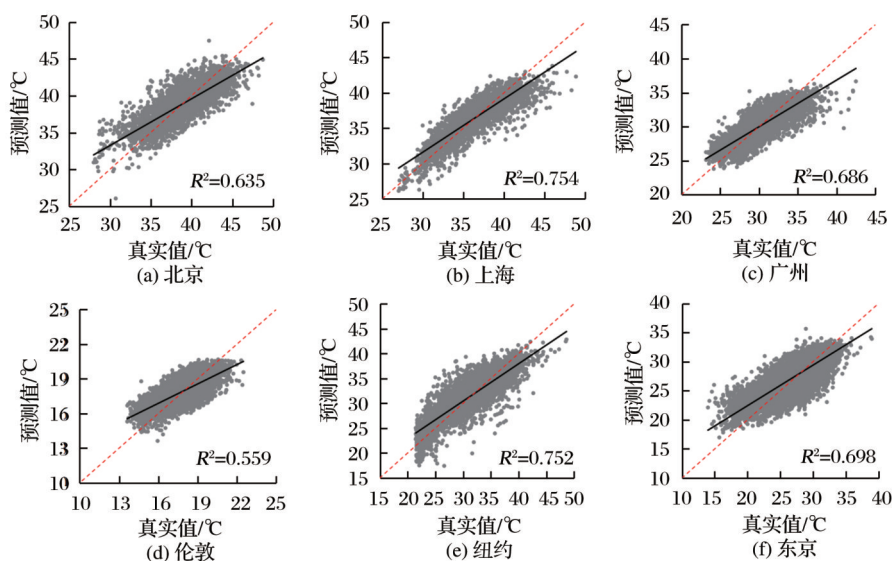


图6 6个城市LCT_MLR(3指数)模型结果

Fig.6 Results comparison of LCT_MLR(3 indices) of the six megacities

45°直线更加明显,且不同城市的1:1散点图存在差异。但同时也可看到当LST值较高时,LCT_RF和LCT_MLR模型都存在一定的低估;而当LST值较低时,LCT_RF模型和LCT_MLR模型都出现一定的高估,且这一现象在不同城市LST的预测过程中都有不同程度地显现。

综上,LST定量模拟结果表明复杂城市景观中的地表覆盖组分与LST存在明显的非线性关系。植被、水体和不透水面等地表覆盖的综合作用会导致城市地表温度的变化。

5 讨 论

本研究采用随机森林回归方法建立了土地覆盖类型与城市热环境的综合定量关系模型(LCT),反映了复杂城市景观中下垫面用地类型和地表温度之间的非线性关系,精度高于基于多元线性回归方法建立的模型,采用的Landsat影像数据易获取,覆盖面广,便于实现大范围应用,但仍有许多不足之处。

(1)以北京、上海、广州、伦敦、纽约和东京等6个超大城市为研究区开展实验,建议的土地覆盖类型与城市热环境的综合定量关系模型(LCT)有较高的精度,然而该方法在其他大城市或中小城市的适用性还有待进一步研究。此外,实验构建的基于Landsat 8影像的回归模型,所选城市区域夏季多云多雨,数据获取受限,未来将进一步探究使用其他多源遥感影像补充获取更多时相云覆盖区域的数据进行研究。

(2)使用的随机森林回归模型是一种黑盒模型,虽然随机森林可以评估每个参数的重要性,但是每个参数具体如何影响分类结果却未可知;本研究构建500棵分类树的随机森林,分类树棵树的增加,一定程度上提高了分类精度,但是分类花费的时间也相应增加,已有的输入参数可以帮助模型实现最优分类精度,但大大增加了算法运行时间。接下来将继续优化输入参数、分类树数量,在回归精度和时间之间谋求平衡。

(3)植被、不透水面和水体等三大土地覆盖类型是城市复合生态系统的重要组成部分,选择这3种土地覆盖类型作为影响因子,可有效地反映地表覆盖与地表温度的综合作用关系,然而大城市地表温度和地表生物物理组分之间作用机理复杂,将来可进一步细分更多地表覆盖类型增加影响因子类型,持续改进评估模型。

6 结 论

研究利用单通道算法反演中外6个超大城市地表温度,并选取NDVI、MNDWI、NDISI指数作为植被、水体和不透水面等3种主要土地覆盖类型的表征因子。在定量分析了每种土地覆盖类型对地表温度的作用基础上,采用随机森林回归方法建立了土地覆盖类型与城市热环境的综合定量关系模型(LCT)。结果表明:

(1)地表温度和下垫面用地类型之间表现出密切的关联,高温热场的空间结构在很大程度上被下垫面用地类型的空间格局所左右,不透水面会导致高温热场的聚集,而植被和水体则有降温作用。当 $NDISI \leq 30\%$ 时,不透水面比例每增加10%,只导致地表温度增加 $0.7^{\circ}\text{C} \sim 1.8^{\circ}\text{C}$ 左右,但当 $NDISI \geq 80\%$ 时,不透水面比例每增加10%,地表温度的增加幅度会达 $0.9^{\circ}\text{C} \sim 2.4^{\circ}\text{C}$ 左右。植被比例每提升10%,地表温度可降低 $0.5^{\circ}\text{C} \sim 1.6^{\circ}\text{C}$;水体比例每提升10%,地表温度可降低 $0.4^{\circ}\text{C} \sim 0.9^{\circ}\text{C}$ 。不同城市地表覆盖升温/降温效应有所差别,在北京、上海、纽约和东京等城市区域,植被和水体增加产生的降温效应较明显。北京、上海和纽约等城市区域的不透水面增加产生的升温效应也高于广州、伦敦和东京等区域。

(2)基于随机森林回归方法建立的NDVI、MNDWI和NDISI等3种土地覆盖指数因子与LST的定量关系模型(LCT_RF)得到的精度高于基于多元线性回归方法建立的模型。LCT_RF模型的 R^2 值在0.623~0.826间,比LCT_MLR模型高0.021~0.074;RMSE比LCT_MLR模型低 $0.07^{\circ}\text{C} \sim 0.35^{\circ}\text{C}$ 。本研究可用于衡量超大城市土地覆盖变化对城市热环境的定量作用,作为生态城市建设重要理论和技术支撑。

参考文献(References):

- [1] Kuang Wenhui, Yang Tianrong, Liu Ailin, *et al.* An Eco-City model for regulating urban land cover structure and thermal environment: Taking Beijing as an example[J]. Science China Earth Sciences, 2017, 47(7): 847-859.[匡文慧, 杨天荣, 刘爱琳, 等. 城市地表覆盖结构组分与热环境调控模型(EcoCity)研究——以北京城市为例[J]. 中国科学: 地球科学, 2017, 47(7): 847-859.]
- [2] Wang Yu, Tang Li, Zhu Haitao, *et al.* The study of urban thermal environment dynamics and attribution analysis based on multiple remote sensing dataset: In the case of Shenzhen[J]. Acta Ecologica Sinica, 2021, 41(22): 8771-8782.[王煜, 唐力,

- 朱海涛,等.基于多源遥感数据的城市热环境响应与归因分析——以深圳市为例[J].生态学报,2021,41(22):8771-8782.]
- [3] Xu H Q, Wang M Y, Shi T T, *et al.* Prediction of ecological effects of potential population and impervious surface increases using a Remote Sensing based Ecological Index (RSEI) [J]. Ecological Indicators, 2018, 93: 730-740.
- [4] Zhang Xiaodong, Zhao Yinxin, Chu Xiaodong, *et al.* Spatial and temporal evolution characteristics of thermal environment and its influencing factors in urban area of Yinchuan City based on remote sensing [J]. Research of Soil and Water Conservation, 2020, 27(6): 180-187.[张晓东,赵银鑫,褚小东,等.基于遥感的银川市城区热环境及其影响因素的时空演变特征[J].水土保持研究,2020,27(6):180-187.]
- [5] Gao Yang, XiongJuhua, Wu Hao, *et al.* Frontier hotspots and development directions of geographical science research: From a perspective of national natural science foundation application keywords in 2021[J/OL]. Scientia Geographica Sinica: 1-16. [高阳,熊巨华,吴浩,等.2021年度自然科学基金申请书关键词透视地理科学研究前沿热点与发展方向[J/OL].地理科学: 1-16. <http://kns.cnki.net/kcms/detail/22.1124.P.20220129.1449.004.html>.]
- [6] Li Deren. Brain cognition and spatial cognition: On integration of geospatial big data and artificial intelligence [J]. Geomatics and Information Science of Wuhan University, 2018, 43(12): 1761-1767.[李德仁.脑认知与空间认知——论空间大数据与人工智能的集成[J].武汉大学学报:信息科学版,2018,43(12):1761-1767.]
- [7] Zhang Fei, Shao Yuan, Huang Hui, *et al.* Review of urban remote sensing research in the last two decades [J]. Acta Ecologica Sinica, 2021, 41(8): 3255-3276.[张飞,邵媛,黄晖,等.近20年城市遥感研究现状及其发展趋势[J].生态学报,2021,41(8):3255-3276.]
- [8] Feyisa G L, Meilby H, Jenerette G D, *et al.* Locally optimized separability enhancement indices for urban land cover mapping: Exploring thermal environmental consequences of rapid urbanization in Addis Ababa, Ethiopia [J]. Remote Sensing of Environment, 2016, 175: 14-31.
- [9] Yang Yuting, Tang Jiafa, BianJinhu, *et al.* Seasonal variations in the relationship between land surface temperature and impervious surface percentage in Kolkata [J]. Remote Sensing Technology and Application, 2021, 36(1): 79-89.[杨玉婷,汤家法,边金虎,等.加尔各答市地表温度与不透水面比例季相关性研究[J].遥感技术与应用,2021,36(01):79-89.]
- [10] Zheng B J, Myint S W, Fan C. Spatial configuration of anthropogenic land cover impacts on urban warming [J]. Landscape and Urban Planning, 2014, 130: 104-111.
- [11] Chen Y J, Yu S X. Impacts of urban landscape patterns on urban thermal variations in Guangzhou, China [J]. International Journal of Applied Earth Observation and Geoinformation, 2017, 54: 65-71.
- [12] Kotharkar R, Bagade A. Evaluating urban heat island in the critical local climate zones of an Indian city [J]. Landscape and Urban Planning, 2018, 169: 92-104.
- [13] Nguyen T, Yu X X, Zhang Z M, *et al.* Relationship between types of urban forest and PM_{2.5} capture at three growth stages of leaves [J]. Journal of Environmental Sciences, 2015, 27: 33-41.
- [14] Balci F B. Determining the impact of urban components on land surface temperature of Istanbul by using remote sensing indices [J]. Environmental Monitoring and Assessment, 2014, 186(2): 859-872.
- [15] Nsubuga F W N, Botai J O, Olwoch J M, *et al.* Detecting changes in surface water area of lake Kyoga sub-basin using remotely sensed imagery in a changing climate [J]. Theoretical and Applied Climatology, 2017, 127(1-2): 327-337.
- [16] Xu Hanqiu. Urban expansion process in the center of the Fuzhou basin, Southeast China in 1976-2006 [J]. Scientia Geographica Sinica, 2011, 31(3): 351-357.[徐涵秋.近30年来福州盆地中心的城市扩展进程[J].地理科学,2011,31(3): 351-357.]
- [17] Shen H F, Huang L W, Zhang L P, *et al.* Long-term and fine-scale satellite monitoring of the urban heat island effect by the fusion of multi-temporal and multi-sensor remote sensed data: A 26-year case study of the city of Wuhan in China [J]. Remote Sensing of Environment, 2016, 172: 109-125.
- [18] Son N T, Thanh B X. Decadal assessment of urban sprawl and its effects on local temperature using Landsat data in Cantho city, Vietnam [J]. Sustainable Cities and Society, 2018, 36: 81-91.
- [19] Van de Voorde T, Jacquet W, Canters F. Mapping form and function in urban areas: An approach based on urban metrics and continuous impervious surface data [J]. Landscape and Urban Planning, 2011, 102(3): 143-155.
- [20] Chander G, Markham B L, Helder D L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors [J]. Remote Sensing of Environment, 2009, 113(5): 893-903.
- [21] Charvz Jr P S. Image-based atmospheric corrections—revisited and revised [J]. Photogrammetric Engineering and Remote Sensing, 1996, 62(9): 1025-1036.
- [22] Jiménez-Muñoz J C, Cristobal J, Sobrino J A, *et al.* Revision of the single-channel algorithm for land surface temperature retrieval from Landsat thermal-infrared data [J]. IEEE Transactions on Geoscience and Remote Sensing, 2009, 47(1): 339-349.
- [23] Jiménez-Muñoz J C, Sobrino J A. A generalized single-channel method for retrieving land surface temperature from remote sensing data [J]. Journal of Geophysical Research Atmospheres, 2003, 108(D22): 4688.
- [24] Jiménez-Muñoz J C, Sobrino J A, Skoković D, *et al.* Land surface temperature retrieval methods from Landsat 8 thermal infrared sensor data [J]. IEEE Geoscience & Remote Sensing Letters, 2014, 11(10): 1840-1843.
- [25] Dematte J A M, Sayao V M, Rizzo R, *et al.* Soil class and attribute dynamics and their relationship with natural vegetation based on satellite remote sensing [J]. Geoderma, 2017, 302: 39-51.
- [26] Tulbure M G, Broich M, Stehman S V, *et al.* Surface water extent dynamics from three decades of seasonally continuous Landsat time series at subcontinental scale in a semiarid region

- [J]. Remote Sensing of Environment, 2016, 178, 142–157.
- [27] Li Deren, Luo Hui, Shao Zhenfeng. Review of impervious surface mapping using remote sensing technology and its application[J]. Geomatics and Information Science of Wuhan University, 2016, 41(5): 569–703.[李德仁, 罗晖, 邵振峰. 遥感技术在不透水层提取中的应用与展望[J]. 武汉大学学报(信息科学版), 2016, 41(5): 569–703.]
- [28] Xu H Q. Analysis of impervious surface and its impact on urban heat environment using the Normalized Difference Impervious Surface Index (NDISI)[J]. Photogrammetric Engineering & Remote Sensing, 2010, 76(5): 557–565.
- [29] Wentz E A. Getting started with geographic information systems, 3rd edition[J]. International Journal of Geographical Information Science, 2002, 16(2): 204–205.
- [30] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5–32.
- [31] JiaJiaqiong, Liu Wanqing, MengQingyan, *et al.* Estimation of maize leaf area index based on GF-1 WFV image and machine learning random algorithm [J]. Journal of Image and Graphics, 2018, 23(5): 719–729.[贾洁琼, 刘万青, 孟庆岩, 等. 基于 GF-1 WFV 影像和机器学习算法的玉米叶面积指数估算[J]. 中国图象图形学报, 2018, 23(5): 719–729.]
- [32] Du P J, Samat A, Waske B, *et al.* Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2015, 105: 38–53.

Study on the Random Forest Regression Model of Land Cover and Thermal Environment in Megacities

Wang Meiya¹, Xu Hanqiu²

(1.School of History and Geography, Minnan Normal University, Zhangzhou 363000, China;

2.College of Environment and Resources; Institute of Remote Sensing Information Engineering;

Fujian Provincial Key Laboratory of Remote Sensing of Soil Erosion and Disaster Prevention,

Fuzhou University, Fuzhou 350116, China)

Abstract: Rapid urbanization has led to rapid change in land cover and the landsurface heat balance in megacities. Due to the complex potential nonlinear relationship between land surface temperature and surface biophysical components in megacities, the quantitative models and the response mechanism between land cover and thermal environment in megacities is not yet clear. Takesix Chinese and foreign megacities (Beijing, Shanghai, Guangzhou, London, New York and Tokyo) as the study area, Landsat images were used to comprehensively analyze the quantitative relationship between urban land cover factors and thermal environment. The single-channel algorithm was used to retrieve the land surface temperature of thesix megacities. The random forest regression model was used to establish the quantitative relationship (LCT) model between land cover types and urban thermal environment (LST). The quantitative relationship between land cover type and LST showed that the LST was closely related to urban land surface types. The spatial pattern of the urban thermal field depends to a great extent on the spatial distribution pattern of the urban land surface types. The impervious surface will lead to the accumulation of high LST fields, while vegetation and water had a significant cooling effect. The land cover compositionin six megacities had different heating/cooling effects. In urban areas, such as Beijing, Shanghai, New York, and Tokyo, the cooling effects of vegetation and water were more pronounced than those in Guangzhou and London. The established LCTmodel between the three land cover types, NDVI, MNDWI, and NDISI, and the urban thermal environment showed that the LCT model had higher precisionthan that was based on the multiple linear regression method. The R^2 value of the LCT_RF model is 0.021~0.074, which is higher than that of the LCT_MLR model. The RMSE is 0.07°C~0.35°C, which is lower than that of the LCT_MLR model. It will be helpful for future construction of eco-cities by studying the interaction mechanism between the land cover and the urban thermal environment in megacities.

Key words: Megacities; Land cover; Urban heat environment; Remote sensing; Random forest regression model