International Journal of Environmental Science and Technology
https://doi.org/10.1007/s13762-022-04510-3

ORIGINAL PAPER

Generating synthetic daily remote sensing products suitable for surface heat island and heatwaves assessments at urban scale

I. Buo[1] · V. Sagris[1] · J. Jaagus[1]

Received: 9 March 2022 / Revised: 12 July 2022 / Accepted: 26 August 2022

Abstract

Presently, no single satellite provides cloud-free daily land surface temperature and normalized difference vegetation index at high spatial resolutions to assess heatwaves in cities. Heatwaves are ephemeral, and it is important to have data of adequate temporal and spatial resolution to study them. This research presents a two-part machine learning approach that yields seamless synthetic daily land surface temperature and normalized difference vegetation index at resolutions (500 and 100 m) suitable for city-scale assessments. The first part involves creating a model to predict normalized difference vegetation index. The second predicts land surface temperature using a model created with the output of the first step and other features as predictors. The outputs were validated with Landsat products from two days. The predicted normalized difference vegetation index had a RMSE of 0.09 and 0.14 for the 500 m and 100 m products, respectively, when compared to the validation data. For the land surface temperature products, the RMSE was 4.28 and 4.33 for the 500 and 100 m products, respectively. Furthermore, a trend analysis reveals a good temporal correlation (500 m:0.66, and 100 m: 0.70) between the land surface temperature and surface-air temperature. The validations reveal that the generated cloud-free synthetic products are more suitable for intra-urban analysis because the models performed with a higher accuracy in built-up areas. Further investigations on how the models respond to changes in the predictors will provide useful for planners in making city-scale heat mitigation decisions.

Keywords Downscaling · Landsat 8 · Land surface temperature · MODIS · Normalized difference vegetation index · Random forest

Introduction

Remotely sensed land surface temperature (LST) and the normalized difference vegetation index (NDVI) are useful parameters in studying the surface urban heat island (SUHI) phenomenon (Bechtel et al. 2012). The NDVI of a pixel informs on the quality of vegetation and its abundance in the pixel. The vegetation abundance determines the LST of a pixel because it controls the process of evapotranspiration(Yuan and Bauer 2007). LST, which is the radiative "skin" temperature of the Earth's surface helps in understanding the relationship between surface materials and the surface thermal environment. Existing LST and NDVI data with high temporal resolution from the moderate resolution imaging spectroradiometer (MODIS) is available twice per day but have a poor spatial resolution of 1 km. This is inadequate for studying surface urban heat at neighborhood scales. For SUHI and urban climate, it is important to have such data within the range of the block and neighborhood scales (Bechtel et al. 2012; Oke et al. 2017). Furthermore, in studying the dynamics of surface urban heat during a heatwave period, it is essential to have seamless LST data for each day of the heatwave period, this can be challenging due to cloud contamination.

To address the spatial resolution limitations, thermal sharpening, statistical, and machine learning (ML) methods have been used to generate synthetic LST and NDVI at higher resolutions. Bechtel et al. (2012) and Yu et al. (2014) used linear regression to downscale coarse resolution LST to fine resolutions suitable for urban scale assessment. Recently, several researchers employed various ML

Editorial responsibility: Samareh Mirkia.

✉ I. Buo
isaac.buo@ut.ee

[1] Department of Geography, University of Tartu, Vanemuise 46, 50410 Tartu, Estonia

Published online: 20 September 2022

algorithms to generate LST at high resolutions. ML algorithms allow for modeling complex relationships between target variables and predictors—an advantage over statistical methods (Osborne and Alvares-Sanches 2019). Li et al. (2019) indicated that a machine learning approach provides an insight into the significance of predictors in estimating the LST of a pixel. This is a major advantage machine learning has over the traditional thermal sharpening algorithms (Li et al. 2019). The random forest (RF) algorithm was used to downscale MODIS LST using surface reflectance bands, elevation, solar incidence, and sky-view factor as predictors (Hutengs and Vohland 2016). RF and area-to-point regression kriging was combined to downscale a 90 m resolution Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) LST to 10 m using spectral indices from Sentinel-2 (Xu et al. 2020). Although these ML attempts were successful, they focused on downscaling products from a day or a few days to assess the performance of their approach. It is important to extend the downscaling approach to generate daily products with which researchers and interested bodies can assess the spatial and temporal variability of the surface temperature during a heatwave. A downscaling approach of this nature addresses the temporal and spatial resolution limitations. Weng et al. (2014) addressed this by fusing MODIS and Landsat LST using a modification of the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM). In addition, most downscaling research limit the model assessments to performance metrics like root mean squared error (RMSE) and mean absolute error (MAE). Furthermore, in most existing research, the response in ML model predictions due to a change in predictor values gets little to no attention. This shortcoming presents the models as black boxes.

To produce daily LST products, it is important to have predictors at the desired resolution and they must be available for each day. Whereas predictors like elevation and land use and land cover may not change daily within a limited period, other predictors like spectral indices and meteorological parameters tend to change. The NDVI is a good predictor of LST and this is corroborated by its utilization as a predictor in most studies (Pan et al. 2018; Bala et al. 2020; Xu et al. 2020; Abdollahipour et al. 2021). Daily NDVI products are mostly at coarse spatial resolutions and are not suitable for generating daily LST at resolutions fit for urban scale analysis. Therefore, to use NDVI as a predictor of LST at such resolutions, it is imperative to first generate daily NDVI at those resolutions. Zhao et al. (2017) used daily NDVI products derived from remote sensing images to classify land use land cover (LULC) using several machine learning algorithms. The results of their work suggest there is a strong relationship between these variables. Therefore, this research assumes that with a ML algorithm, the NDVI of a grid cell is predictable given its LULC information. In this research, the fractions of the various LULCs present in a grid cell derived from high-resolution mapping databases are used instead of using one value of LULC per cell to predict a cell's NDVI.

This research establishes an approach to generate synthetic daily NDVI and LST at resolutions appropriate for urban scale studies using ML. The approach addresses the spatial and temporal limitations associated with these products in assessing SUHI and the spatial variability of surface temperatures in heatwaves. In addition, the synthetic products generated are free of gaps that are present in actual products obtained from remote sensors because of cloud contamination. The possibility of using fractions of LULCs in a grid cell and time as predictors of daily NDVI is explored. Furthermore, the usefulness of the resulting products as predictors in combination with other features to generate synthetic daily LST at the same resolutions was investigated. This research reveals how the models operate by identifying the importance of the predictors and how the models respond given specific values of the predictors. Finally, the performance of the models based on LULC was assessed in addition to commonly used metrics such as RMSE.

Study area

The research focuses on the city of Tallinn, the Estonian capital, and its neighboring settlements. Tallinn is located at 59.4370° N, 24.7536° E. The city of Tallinn is considered as the Central Urban District (CUD) and the entire area as the Tallinn Urban Agglomeration (TUA). To generate the synthetic products, the multi-level grids Statistics Estonia (SE) used to map the country's population was adopted. This informed the choice of the resolutions because it allows for the comparison of population distribution and LST to assess the impact of extreme heat. The 1 km grid is the basic mapping unit, additionally, the SE maps the CUD at 100 m grid cells and the TUA with a 500 m grid. The limits set for this research included settlements within the TUA and those that are within a 5 km buffer. Half of Estonia's population live in this area according to SE and therefore it is an important study area. Figure 1 shows the study area for this research and the land use and land cover distribution within the area.

According to the Estonian Weather Service (EWS), a period of three or more days with a maximum temperature of ≥ 27 °C or an average ≥ 20 °C is a dangerous Level 1 heatwave. The EWS classifies a period of three or more days with a maximum temperature of ≥ 30 °C or an average ≥ 25° C as a dangerous Level 2 heatwave. The summer period of 2018 was the focus of this study because several heatwaves occurred in the study area (Buo et al. 2021).

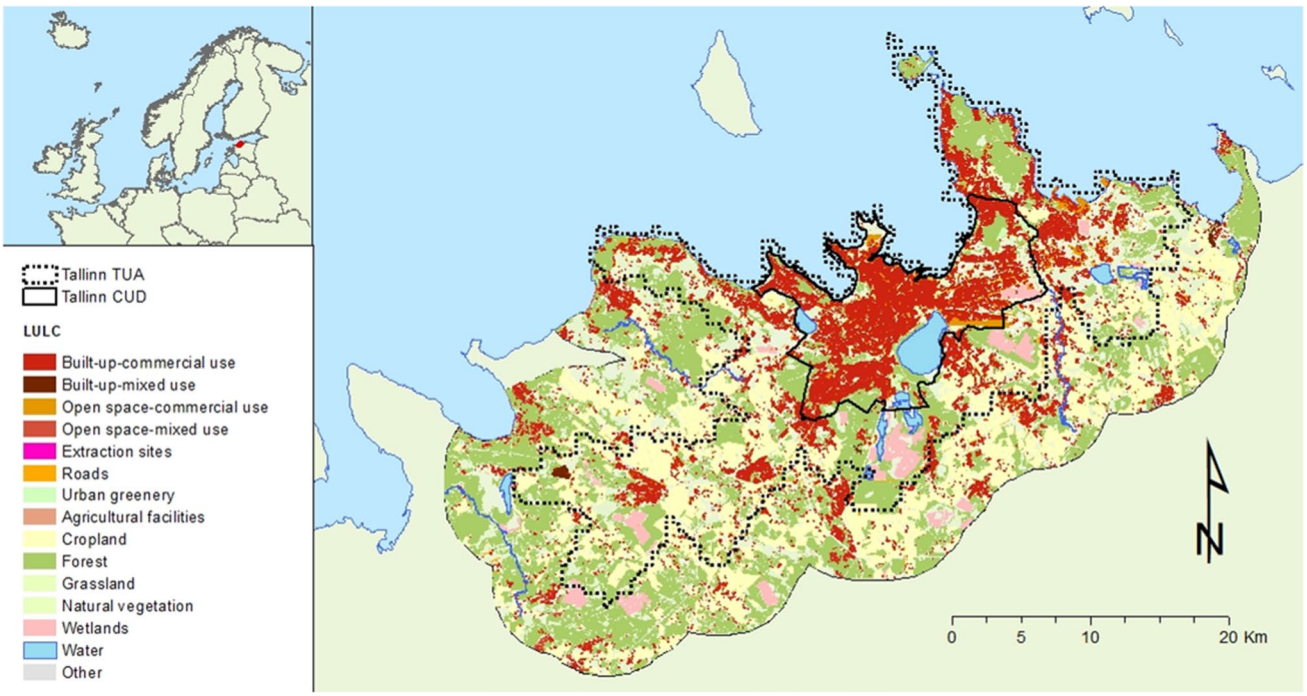

Fig. 1 Map of the study area showing LULC distribution. The insert map shows the location (red mark) of the study area relative to other parts of Europe

Materials and methods

Data and pre-processing

The 1 km grid provided by Statistics Estonia was used as the framework for modeling the NDVI and LST. Additionally, 500 m and 100 m grids were created for the study area which were used in predicting the LST and NDVI at the desired resolutions. The 1 km, 500 m, and 100 m resolution grids have 1374, 5262, and 125,301 cells, respectively. The Land Use and Land Cover (LULC) database of the Estonian Land Board was employed to compute the LULC fractions. The LULC database is highly detailed with 55 classes. For this research, the objects were reclassified into 15 broad classes (Fig. 1). In addition, the LULC database was categorized into surface types which comprised of pervious, impervious, building surfaces, and water. The LULC and surface type fractions were then determined for the three sets of grid cells by calculating the percentage of area that each LULC class or surface type covered in each of the cells following Geletič and Lehnert (2016) and Buo et al. (2021). A Digital Elevation Model (DEM) from the Estonian Land Board was used to sample the mean elevation per grid cell. In addition, the distance from each grid cell to the seacoast was calculated for all three grids. The hourly surface-air temperature (SAT) measured at the local weather station within the study period was obtained to validate the temporal trends of the predicted LST products.

The NDVI and LST products from Landsat 8 was utilized in this research. The Landsat 8 images used in this study covered 19 days within the summer period of 2018 after cloud-contaminated images were discarded. The days with images for each month and the total number of images per month are shown in Table 1. Landsat 8 data for 2 days (17-May and 2-June) were reserved to validate the models. The NDVI for the Landsat 8 images were calculated using Eq. 1 and the LST was calculated using the single-channel algorithm in Eq. 2 (Cristóbal et al. 2018). Landsat 8 has a resolution of 30 m; therefore, the LST and NDVI values were resampled to 1 km using spatial averaging. In addition, the Landsat 8 NDVI and LST images used for the validations were resampled to 500 m and 100 m using spatial averaging.

Table 1 Summary of Landsat 8 images used for summer 2018

Month	Dates with images	Number of days with images
May	8, 10,17, 24, 26	5
June	2, 9, 11, 18, 27	5
July	4, 11, 13, 20, 27, 29	6
August	5, 14, 21	3

The images covered rows: 018 and 019; paths: 187 and 188

$$\mathrm{NDVI} = \frac{(\rho_{\mathrm{nir}} - \rho_r)}{(\rho_{\mathrm{nir}} + \rho_r)} \quad (1)$$

where ρ_{nir} is the surface reflectance in the near-infrared band, and ρ_r is the surface reflectance in the red band.

$$\mathrm{LST} = \gamma\left[\frac{1}{\varepsilon}\left(\psi_1 L_{\mathrm{sen}} + \psi_2\right) + \psi_3\right] + \delta \quad (2)$$

where L_{sen} is at sensor radiance, γ and δ are parameters based on Planck's functions. ψ_1, ψ_2, and ψ_3 are atmospheric functions which were calculated following Barsi et al. (2003) and ε is the surface emissivity which was estimated using the NDVI threshold approach (Sobrino et al. 2008).

The MODIS NDVI (MODIS/006/MCD43A4_006_NDVI) products were downloaded from Google Earth Engine (GEE). The NDVI images from MODIS are daily products covering the entire summer period (May–August) for the year 2018. Also, the mean daily composite MODIS LST (TERA-MOD11A1.006; AQUA-MYD11A1.006) were obtained from GEE. The MODIS LST for this research were limited to only daytime products. The computational capabilities of the GEE platform were harnessed to compute the mean daily MODIS LST composite representing the mean surface temperature. Furthermore, combining both products allowed the reduction of missing data. A summary of all data used and processing in this research is provided in Table 2. The native resolution of the MODIS products is 1 km.

Modeling

The approach adopted for this research is in two parts (Fig. 2a), in each, the data is fitted to an algorithm to create a model. The first part involved creating a model to predict NDVI at the desired resolutions. The second dealt with creating a model that takes the predicted NDVI and other features as predictors to determine the LST at the desired resolutions. The NDVI and LST modeling were executed using the schematic in Fig. 2b. The process begins with using an algorithm to model the relationship between the target feature and the predictors at a coarse resolution (Fig. 2b). With an established model in place, the target feature was predicted at a fine resolution using the predictors at the same resolution. The underpinning assumption is that relationships between the target and the predictors are similar both at coarse and high resolution (Bechtel et al. 2012). The main algorithm used at both stages was Random Forest (RF), which uses an ensemble of decision trees to make predictions (Breiman 2001). In RF, the final prediction is made either by a majority vote for classification problems or averaging all predictions in regression problems (Moisen 2008; Shalev-Shwartz and

Table 2 Summary of datasets used

Dataset	Resolution	Processing
Landsat 8 NDVI	30 m	Resampled to 1000 m, 500 m, and 100 m
MODIS NDVI	1000 m	Used as a target in MODIS-NDVI and a predictor in the LST model
Landsat 8 LST	30 m	Resampled to 500 m and 100 m
MODIS LST	1000 m	Target feature in LST models
DEM	25 m	Resampled to 1000 m, 500 m, and 100 m
Surface-air temperature (Tallinn weather station)	Hourly	Mean daily temperature for daytime MODIS overpass window (10:00 am−2:00 pm)

The surface-air temperature station is located on 59.3981 N and 24.6029 E

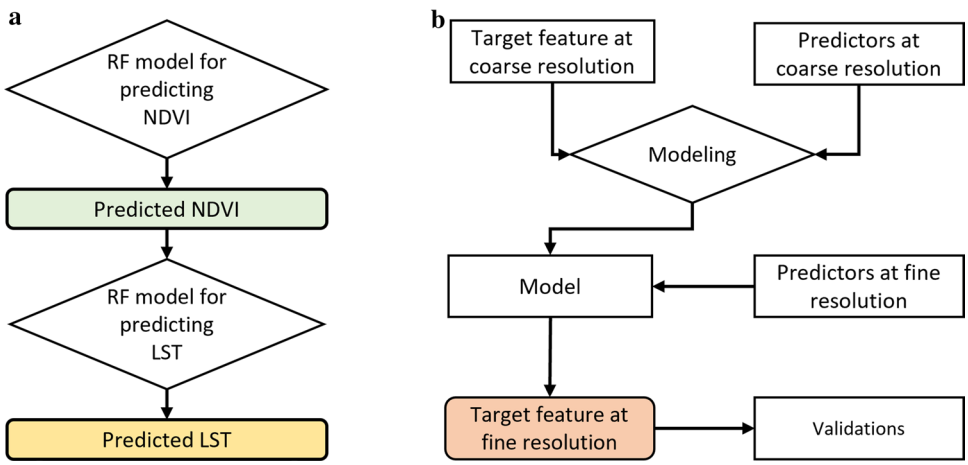

Fig. 2 **a** Schematic of the adopted two-step approach to generate synthetic products. **b** General workflow adopted for modeling at each step of the approach in (**a**)

Ben-David 2013). The partial dependence plots were utilized to identify what the models would predict given specific values of the predictors. The partial dependence plot is a model agnostic method that shows a predictor and the average predicted outcome in the form of curves (Molnar 2019).

Predicting NDVI

The detailed approach adopted in creating the NDVI models and predicting the NDVI at the desired resolutions is shown in Fig. 3a. The NDVI of a cell is expressed as a function of the LULC fractions and the Day of the Year (DOY) (Eq. 3).

$$\mathrm{NDVI}_i = f\left(\propto \mathrm{LULC}_i, \mathrm{DOY}\right) \tag{3}$$

where NDVI_i is the NDVI of cell i, $\propto \mathrm{LULC}_i$ represents the fractions (shares in percentages) of LULCs present in cell i and DOY is the Day-of-the-year.

A data matrix was created from the 1 km grid in which each grid cell (instance) had 15 columns for the fractions of the LULC, DOY, and MODIS NDVI. This data matrix was replicated by substituting the MODIS NDVI and DOY for each grid cell with the resampled Landsat 8 NDVI for each grid cell and the corresponding DOY of their imaging. The preliminary analysis focused on how NDVI from both satellite missions correlated by performing a correlation analysis at 1 km resolution.

Following the outcome of the preliminary analysis, two models were created. One of the models predicted a MODIS-like NDVI and the other Landsat-8 NDVI. In each case, the datasets were split into two, 60% for training and 40% for testing. Three baseline models were run using linear regression, ridge regression, and RF without tuning hyperparameters. From the assessment of all the baseline models, the RF model was found to outperform the rest; therefore, the main models were created using RF.

In creating the models in each case, a grid search cross-validation which is an iterative process on the training set to find the best combination of hyperparameters was performed (Siji George and Sumathi 2020). During this process, the training data set is divided into k-folds, for this research k was 5.

With the two well-performing models in place, the daily NDVI at 500 m and 100 m resolutions from the MODIS-like and Landsat 8-like models were generated. The mean predicted NDVI from both models was calculated next. For the validations, the R^2 and RMSE between predictions and the validation Landsat 8 data were calculated. In addition, grid cells with underestimated and overestimated predictions and their predominant LULC were examined to assess the performance of the models for different LULC. If the difference between the predicted NDVI and the validation at a cell is $\geq +0.3$, it was classified as overestimated and underestimated if the difference is ≤ -0.3.

Predicting LST

The approach used in modeling and predicting the LST at the desired resolutions in detail is shown in Fig. 3b. To model the LST, the MODIS LST was used as the target variable. The instances for training the model were limited to MODIS LST pixels that had errors less than 3 °C based on the quality assurance data from the provider. This filtering process was done in GEE. The fractions of surface types for each grid cell (building; impervious; pervious; and water), NDVI, Day of the Year, and spatial predictors (X coordinates, Y coordinates, elevation, and distance to coast) were used as the predictors. The LST of a grid cell is expressed as a function of the predictors in Eq. 4. The NDVI used here was the MODIS NDVI also at 1 km resolution like the LST.

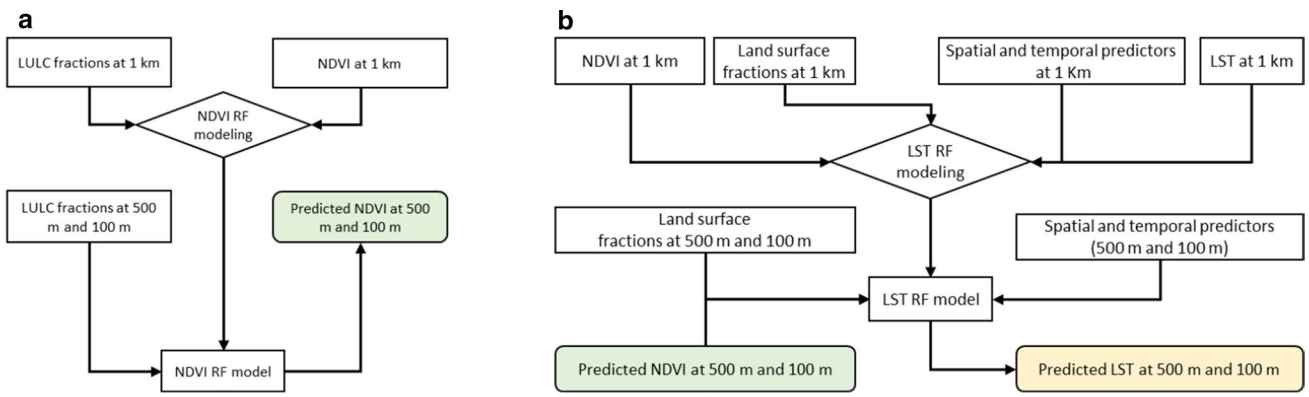

Fig. 3 **a** Detailed approach to generate synthetic NDVI at desired resolutions. **b** Detailed approach to generate synthetic LST at desired resolutions. The NDVI used at the modeling stage is the MODIS NDVI also used in 3a. The predicted NDVIs at 500 m and 100 m are the mean predictions from 3a

$$\mathrm{LST}_i = f(\mathrm{NDVI}_i, \propto \mathrm{LS}_i, \mathrm{CC}_i, \mathrm{DC}_i, \mathrm{ME}_{i,}\mathrm{DOY}) \quad (4)$$

where, LST_i is the LST of cell i, NDVI_i is the NDVI of cell i, $\propto \mathrm{LS}_i$ represents the fractions (shares in percentages) of land surface types in cell$i$, CC_i are the centroid coordinates of cell$i$, DC_i is the distance to coast from cell$i$,$\mathrm{ME}_{i,}$ is the mean elevation in cell i, and DOY is the Day of the Year.

In creating the LST model, an approach like that employed in modeling the NDVI was adopted. The data was split into training (60%), and testing (40%). Baseline models were created using linear regression and RF with default hyperparameters. In this phase, RF stood out as well. Following this, a final model was created after hyperparameter tuning. With the LST model in place, the LST at the desired resolutions were generated using the predictors at those resolutions. The mean predicted NDVI from the previous section was used here. The predictions were validated by comparing the predicted LST with the Landsat 8 LST reserved for validation. Furthermore, the predominant LULC types in the grid cells in which the LST was over-or underestimated were identified using differences of +3 °C and—3 °C as the thresholds respectively, following Malamiri et al. (2018). Furthermore, the relationships between the temporal trends of the predicted products and the trend from SAT measurements made at the local weather station were evaluated.

Results and discussion

Results

NDVI models

The assessment of the relationship between MODIS NDVI and Landsat 8 NDVI shows the existence of a higher correlation ($R^2 = 0.78$; RMSE = 0.07). In the baseline modeling, it was noticed that the same features (built-up commercial, agricultural facilities, grasslands) were not important (had a feature importance of 0) in determining the NDVI of a cell by both models and hence they were excluded in the final models. The importance of the predictors for both models is shown in Table 3. The importance of each predictor can be any value between 0 and 1. The sum of all the feature importance values is 1. The fraction of forest per grid cell is the most important predictor in determining the NDVI of a grid cell followed by water for the MODIS NDVI model, while for the Landsat 8 model, it is water followed by forest (Table 3). Also, the DOY is significantly important in both models (Table 3). The other features are less than 0.1 for both models and in some cases have close values. However, in the MODIS model, croplands, reeds, and bushes were of higher importance compared to what they were in the Landsat 8 model. The feature importance reveals which predictors the models mostly rely on to make a split at the nodes of the decision trees. However, it does not provide information on what the models would predict given a specific predictor value. The partial dependence plots for the MODIS NDVI and Landsat 8 NDVI models are shown in Fig. 4a and b, respectively. From the figures, it is evident that both models seem to think along the same lines. The models predict lower NDVI as the fraction of water in a grid cell increases (Fig. 4a and b). Furthermore, as the fraction of reeds and bushes increases, both models predict higher NDVI. However, there are some notable differences. The MODIS model predictions are in the range of 0.4 and 0.8 irrespective of the predictor value. On the other hand, the Landsat 8 model seems to keep its predictions between about −0.25 and 0.75.

The models performed well both on their training and test data. From Table 3, the closeness of the metrics suggests neither of the two models overfits—a situation where the model performs far better on the training data than the test (unseen) data.

Table 3 Feature importance and performance summary for NDVI models

Feature	MODIS NDVI model	Landsat 8 NDVI model
Water	0.156	0.413
Roads	0.081	0.058
Urban greenery	0.032	0.010
Other	0.036	0.021
Extraction sites	0.027	0.019
Open space-commercial use	0.028	0.035
Cropland	0.081	0.041
Reeds and bushes	0.075	0.019
Open space-mixed used	0.034	0.017
Forest	0.226	0.194
Wetlands	0.015	0.005
Built-up-mixed use	0.024	0.015
DOY	0.185	0.151
Performance metrics		
Training size	80,795	4321
Test size	53,864	2882
RMSE on training data	0.01	0.03
RMSE on test data	0.02	0.05
R^2 training data	0.99	0.97
R^2 test data	0.99	0.89

Features excluded from the final model were built-up commercial, agricultural facilities, and grasslands

LST model

The importance of the predictors and the performance of the LST model are highlighted in Table 4. The DOY stands out as the most important feature for the LST model,

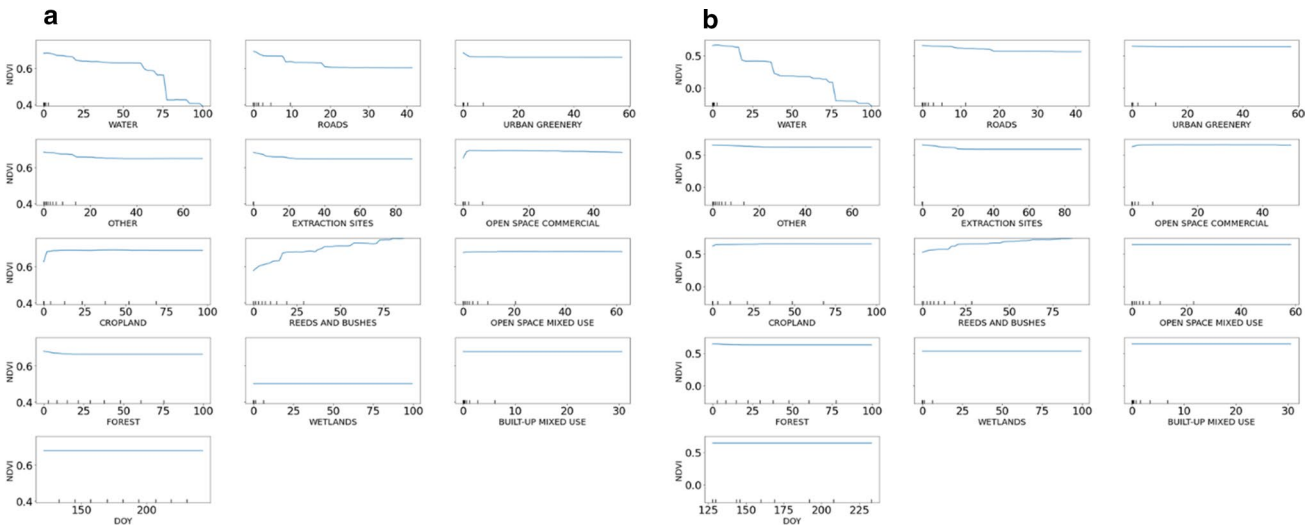

Fig. 4 **a** Partial dependence plots for MODIS NDVI model; **b** Partial dependence plots for Landsat 8 NDVI model. The horizontal axis for the LULC classes represents the possible percentages per grid cell based on the data

Table 4 Feature importance and performance summary for LST models

Feature	LST model
Water	0.016
Building Surface	0.007
Impervious Surface	0.005
Pervious Surface	0.005
NDVI	0.064
X	0.061
Y	0.050
Elevation	0.022
Distance to coast	0.042
DOY	0.728
Performance metrics	
Training size	2,006,836
Test size	1,337,892
RMSE on training data	0.90
RMSE on test data	1.32
R^2 training data	0.96
R^2 test data	0.91

indicating the influence of time (Table 4). The NDVI of the cell came out as the next important feature for the model in estimating the LST of a cell. The spatial features, X coordinates, Y coordinates, elevation, and the distance to coast followed in that order. The shares of surface types were the least important features in predicting the LST of a cell according to the model. From the partial dependence plot in Fig. 5, it is noticeable that the model predicts lower temperatures as the fraction of water in the cell increases from zero percent until about 60%. The predicted LST does not go any lower even though the fraction of water increases. Looking at the behavior of the model concerning the NDVI of a cell, it is evident that the model predicts the same LST for all negative NDVI values and positive values from 0 to + 0.5. The model predicts lower LST as the NDVI of a cell appreciates from + 0.5. The plots from spatial predictors give an idea of the variation in space while the DOY plot shows the temporal variations in LST.

Synthetic NDVI at 500 m and 100 m

The predictions from the MODIS and Landsat 8 NDVI models were similar. However, to get a single NDVI product, the mean of both predictions was used. The resampled Landsat 8 NDVI at 500 m and 100 m for the validation days, the mean predicted NDVI from both models, and the difference between the (resampled) validation and predicted values are shown in Fig. 6a and b. Similar figures for the predicted NDVI from the MODIS model and Landsat 8 model are provided in the Supplementary materials (SI1 and SI2). The predicted NDVIs for distinctive water bodies (Fig. 1) are generally higher than the validation NDVI values (Fig. 6a and b). Therefore, for water bodies, the models overestimate the NDVI. From Fig. 6a and b, it can be noted that the differences are smaller at 500 m compared to 100 m. More than 95% of the NDVI predictions (from both models) for the validation days had differences (Δ) that were greater than −0.3 or less than + 0.3 when compared to the NDVI values for the validation sets, as indicated in Table 5. The mean predicted products have more estimations within the range compared to the MODIS and Landsat 8 versions. The correlations between the predicted values and validation values

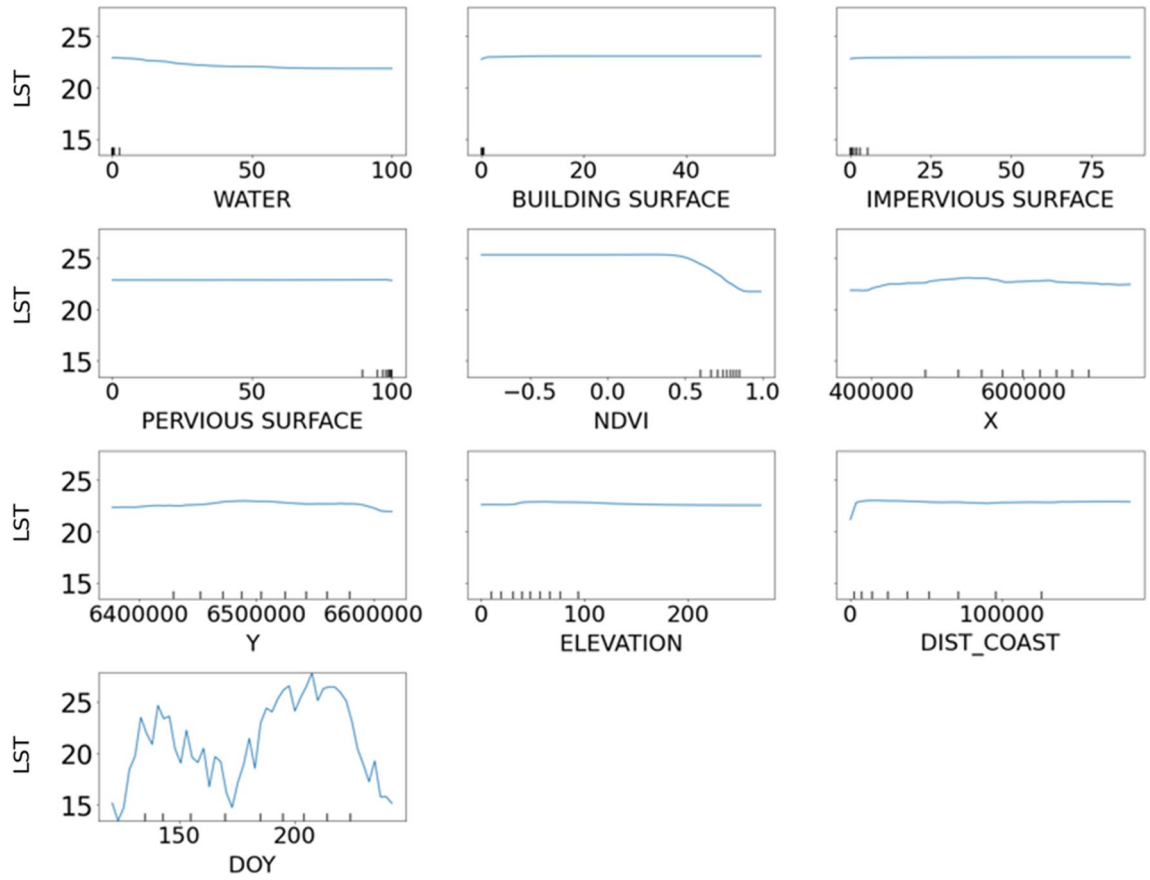

Fig. 5 Partial dependence plots for LST model

are shown in Table 6. The correlation is higher in the mean predicted product compared to the rest at 500 m. There was no improvement at the 100 m resolution.

Synthetic LST at 500 m and 100 m

The resampled Landsat 8 LST at 500 m and 100 m for the validation days, the predicted LST, and the difference between the (resampled) validation and prediction values are shown in Fig. 7a and 7b. From Fig. 7a and b, the predictions are mostly higher than the validation values. This can be attributed to the difference in sensing times for MODIS and Landsat 8. However, there are evident similarities from Fig. 7a and b, especially in the CUD area when comparing the predictions to the validations. About 26% (500 m) and 29% (100 m) of the predicted LST had a difference within the range of -3 °C and + 3 °C. The correlation and RMSE between the validation data and the predicted data were 0.61(RMSE = 4.28) and 0.52 (RMSE = 4.33) at 500 m and 100 m, respectively.

Assessment of synthetic NDVI and LST based on LULC

The mean predicted NDVI and LST product assessments considering predominant LULC types in a grid cell are shown in Fig. 8a and b. For NDVI, the overestimated or underestimated grid cells are dominated by water, cropland, and extraction sites. In the case of LST, it is the grid cells that have forest, cropland, reeds, and bushes as their predominant LULC. Interestingly, most of the predictions for built-up dominated cells are within range, therefore this approach produces promising results for urban areas.

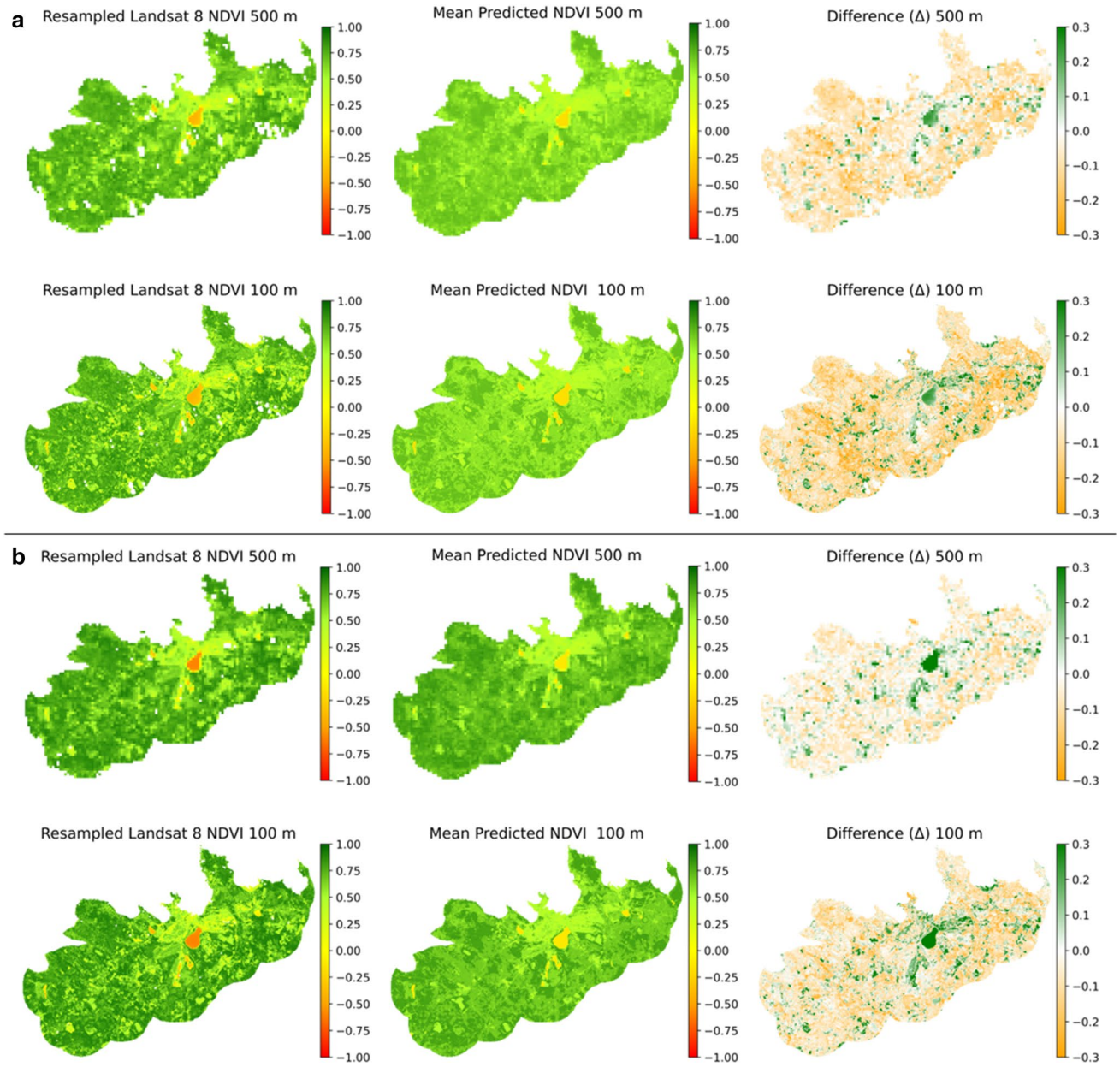

Fig. 6 **a** Resampled Landsat 8 NDVI, mean predicted NDVI, and difference map for 17–05-2018. **b** Resampled Landsat 8 NDVI, mean predicted NDVI, and difference map for 02–06-2018. The difference maps are the results of subtracting the resampled NDVI from the mean predicted Landsat NDVI

Table 5 NDVI estimations within range

	Number of validation grid cells		Estimations within range ($-0.3 < \Delta < 0.3$)	
	500 m	100 m	500 m	100 m
MODIS model products	9772	245,402	9623 (98.4%)	235,779 (97.0%)
Landsat 8 model products	9772	245,402	9684 (99.1%)	236,745 (96.4%)
Mean predictions	9772	245,402	9706 (99.3%)	239,168 (97.4%)

Table 6 Performance metrics of predicted NDVI against validations

	R^2		RMSE	
	500 m	100 m	500 m	100 m
MODIS model products	0.73	0.63	0.09	0.13
Landsat 8 model products	0.72	0.59	0.11	0.15
Mean predictions	0.76	0.63	0.09	0.14

Trend analysis on predicted LST

The predicted LSTs at both resolutions have temporal trends identical to the one observed from the local weather station (Fig. 9). At the beginning of the study period, there is a significant difference between the LST and the measured surface-air temperature because of the high amount of sunshine,

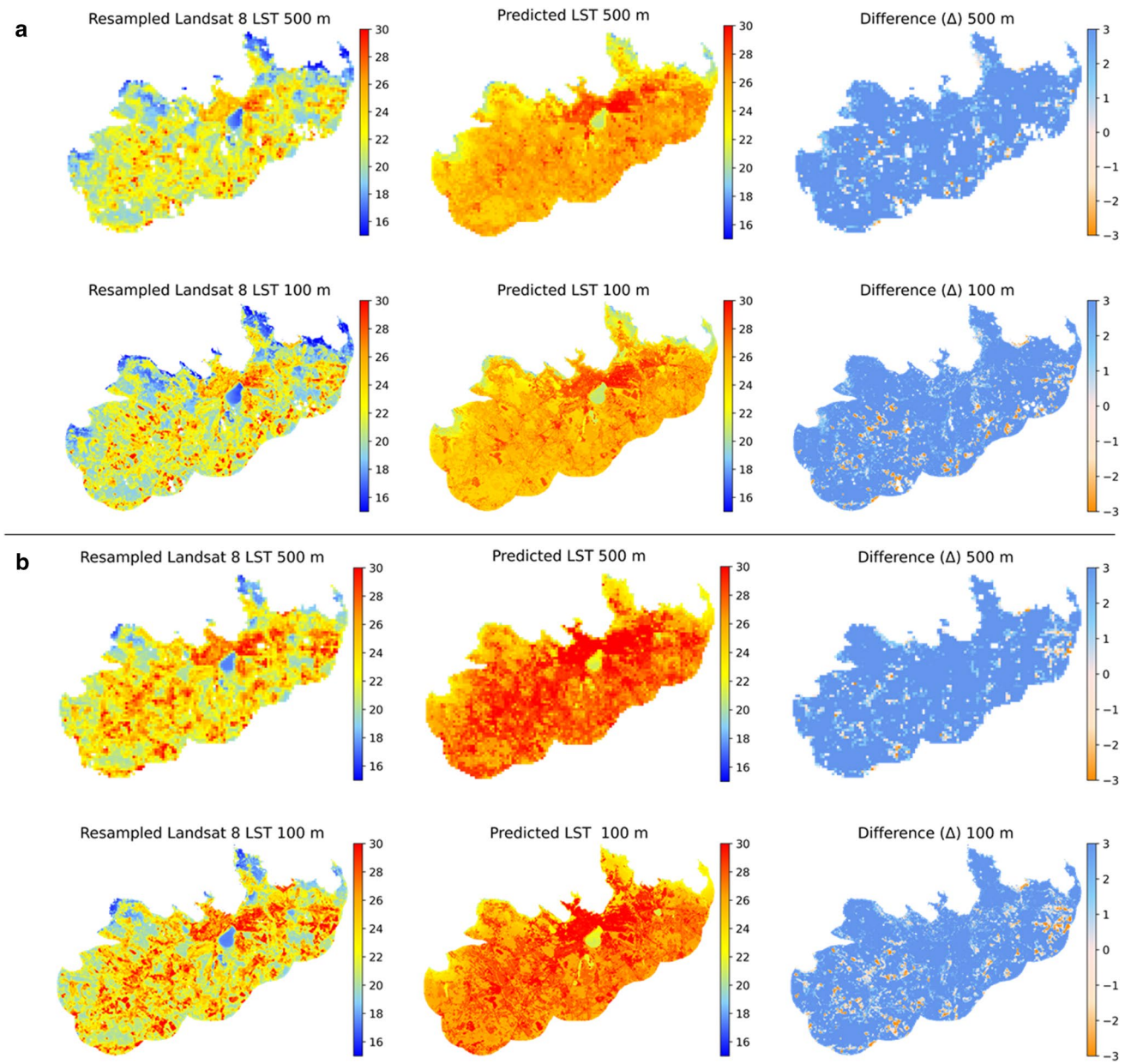

Fig. 7 **a** Resampled Landsat 8 LST, mean predicted LST, and difference map for 17–05-2018. **b** Resampled Landsat 8 LST, mean predicted LST, and difference maps for 02–06-2018. The difference maps are the results of subtracting the resampled LST from the predicted Landsat LST

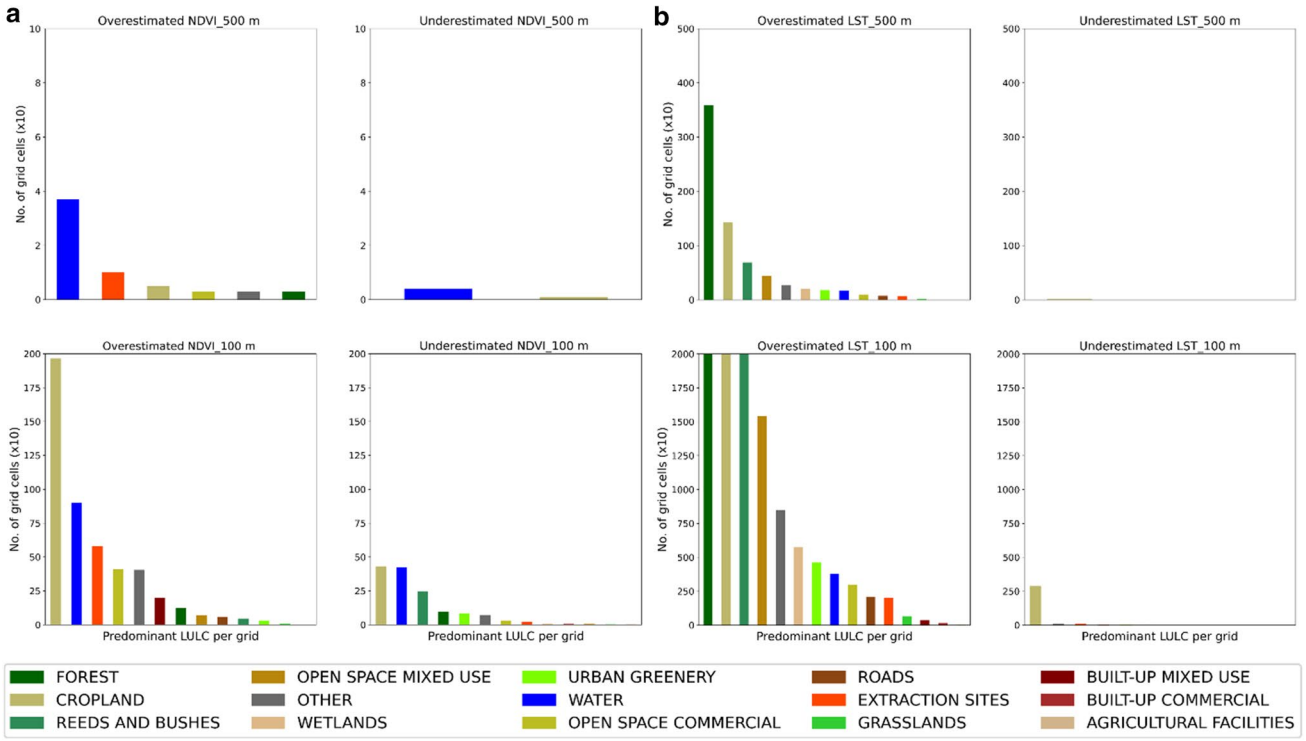

Fig. 8 **a** Predominant LULC in overestimated and underestimated grid cells for mean NDVI products for all validation days. **b** Predominant LULC in overestimated and underestimated grid cells for LST products for all validation days. Overestimated NDVI ($\Delta \geq +0.3$); underestimated NDVI ($\Delta \leq -0.3$) Overestimated LST ($\Delta \geq +3$ °C); underestimated LST ($\Delta \leq -3$ °C)

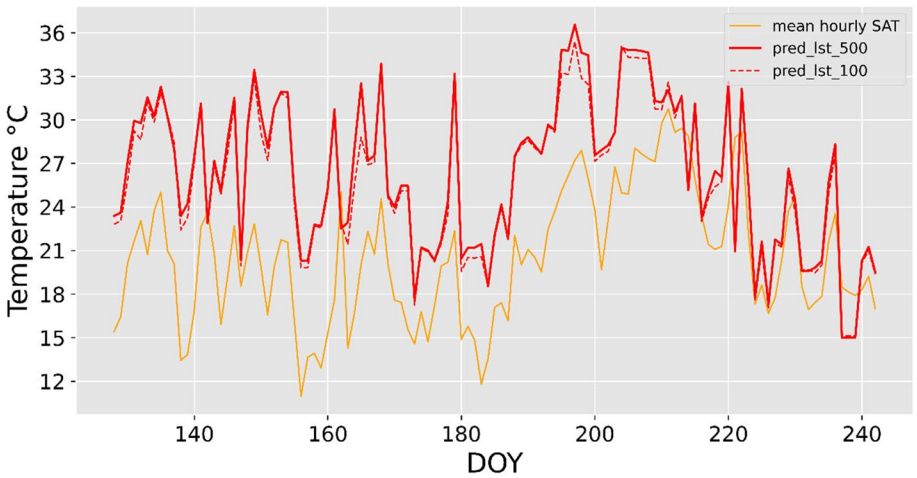

Fig. 9 SAT and predicted LST trends

which is the case in the beginning and middle of summer. However, at the end, both parameters become close as the summer period comes to an end the number of cloudy days increase. Having close values for both parameters were not expected because the SAT measurements represent a point (shaded) in space whereas the predicted LST represents an area which contains the point. Furthermore, both parameters are well-known to be correlated despite the potential disparities in absolute values (Good et al. 2017). The Pearson's coefficients of correlation between the SAT trend and the predicted LST trends are 0.68 (500 m) and 0.70 (100 m). Both trends from the predicted LST had p values < 0.005 when compared with the SAT trend indicating significance.

Discussion

NDVI

The results show that the NDVI from Landsat 8 and MODIS are highly correlated. This suggests that for vegetation quality assessment at an urban scale, products from both sensors are useful. The findings in this research are further corroborated by the results in Ke et al.(2015) where NDVIs from the two sensors were compared for four days at different sites. However, for certain land cover types, the NDVI from both sensors may be the same or different depending on the wavelength ranges in their red and near-infrared bands. In addition, differences in the processing algorithms for atmospheric corrections and computation of surface reflectance can introduce some differences in the NDVI (Ju et al. 2012; Ke et al. 2015). Other instrument-specific attributes like sensing time and viewing angle can introduce differences in the measurements (Gastellu-Etchegorry et al. 2012).

The validations (Table 5) indicate the correlation between the synthetic NDVI, and the validation data is like what was found at 1 km resolution. This demonstrates the ability of the models to downscale and the robustness of the RF algorithm (Li et al. 2019). Hence, the products from the models are satisfactory on their own. However, the mean of predictions from both models is considered a better product. Generating the mean of predictions reduces some biases in the NDVI from both sensors, which are inherent in the model predictions. This is evident from the partial dependence plots. Furthermore, although the NDVI from different sensors are highly correlated, they have notable differences, as intercalibration studies reveal (Steven et al. 2003). The mean predictions have more estimations within range (Table 4), and the performance metrics in Table 5 suggest that they are as reliable as the outputs from both models.

In comparing the output of the NDVI models and the mean predicted NDVI products with the validation sets based on predominant LULC per grid cell, the MODIS model overestimated the NDVI in grid cells predominantly covered by water (SI3). The reason for this overestimation is noticeable in Fig. 4a where it is evident that the MODIS model does not predict below + 0.4, however, water bodies typically have lower NDVI values. The Landsat-8 model overestimated and underestimated in grid cells predominantly covered by water, cropland, extraction sites, and reeds and bushes (SI3). However, in the mean predictions, there is a significant reduction in the count of over and underestimations for these LULC types, signifying the advantage of the mean product (Fig. 6a and b). The predominant LULC in the over and underestimated grid cells at 500 m were the same at the 100 m resolution. Particularly, the count of predominantly cropland grid cells that were over or underestimated exceeded the count of water at 100 m. This suggests that at 100 m both models do not perform well in estimating the NDVI of croplands and this is also evident in the mean predictions. One reason for this is the differences in crop types and growing seasons. In the case of water bodies, flora and fauna activities in the water bodies can influence the surface properties which can affect the NDVI. Therefore, the proportion of water present in the grid cell may not be enough information to predict the NDVI of a cell in all cases. Hence, incorporating other features like water indices into a ML model in addition to the LULC fractions per grid cell can improve the NDVI predictions for water. Despite these limitations, the counts of predominantly built-up cells that were over and underestimated are significantly low in the MODIS-like, Landsat 8-like, and the mean products. This makes them useful for further analysis of vegetation quality in urban areas and particularly the Tallinn CUD area.

LST

In the evaluation of the LST from both sensors at 1 km, it was noted that the MODIS LST is higher than that of Landsat 8, however, they are highly correlated.

While there is a good correlation between the predicted values and validation data, there is a high number of overestimations (Fig. 7a and b). This is primarily because of the difference in sensing time. The model was built with a MODIS product which is a composite of two observations made at two separate times of the day, while the validations were observed at a different time which does not coincide with any of the two MODIS observation times. The percentages would have been higher if the predictions were validated with MODIS products upscaled from a higher resolution to the desired resolution, however, no such product exists. Furthermore, LST is a parameter that varies diurnally (Weng and Fu 2014; Hu et al. 2020) and therefore, it was not expected that the predicted LST would be as close to the validations as they were for the NDVI. Despite these, it cannot be conclusively claimed that this is true for all cases as other researchers have achieved better RMSEs when they used products from different sensors (Hutengs et al. 2018; Xu et al. 2020). The synthetic products are considered suitable for any analysis in the CUD and TUA that would seek mean daytime LST at the desired resolutions. The synthetic LST products reflect the day-to-day variation of temperature in the study area for the period of interest (Fig. 9).

The validation of the predicted LST products based on LULC reveals that the model is more likely to overestimate. Predominantly forested grid cells seem to be the hardest to estimate as the model consistently overestimates the LST at both scales. In addition, cropland-dominated grid cells are

both overestimated and underestimated by the model. However, it mostly overestimates rather than underestimates for these kinds of grid cells. Furthermore, the model overestimated LST for grid cells which are predominantly reeds and bushes. These limitations are attributed to differences in forest and crop types. Adding extra predictors that distinguish forest and cropland types in addition to NDVI-which also acts as a proxy distinguisher in terms of vegetation quality, can improve the performance. The model does not perform well for vegetated surfaces, because the vegetated surfaces were classified as impervious surfaces—a broad class that includes non-vegetated surfaces which are also impervious. This is the trade-off dealt with when a 'simple' ML model with fewer features (other features and surface type fractions) is chosen over one with more features (other features and LULC fractions) with the same performance during baseline modeling. In a previous study validating sharpened products, different vegetated surfaces had varying RMSEs, indicating the variations of LST that different vegetated surfaces (Bartkowiak et al. 2019). In addition, the Estonian Topographic Database does not have records of clear-cut areas, this can affect the performance of the model. Areas like these remain vegetated in the database even though they are not in reality. In instances like this, the LST will not match the information fed into the model.

Despite these limitations, the model performs effectively for predominantly built-up grid cells and other LULC types like roads. This is of prime importance because generating LST useful for assessing heat islands and heatwaves in the CUD and TUA, where most people reside was the main aim. The predicted products from the validations are suitable for intra-urban SUHI assessment and spatial variability of heatwave within the CUD and the TUA. Furthermore, the results of the correlation of the surface-air temperature and predicted LST trends give credence to the usefulness of the synthetic products.

Assessment of the approach

Overall, the approach yields products useful in assessing urban thermal environments at suitable scales. In addition, the resulting products are free of gaps that are often present in optically sensed products due to clouds and shadows. More importantly, the models turn out products that are accurate within considerable ranges for the NDVI and to some extent, LST in the built-up areas. Although the metrics from the validations were like those obtained when NDVI and LST from both missions were compared at the coarse resolution, it cannot be conclusively claimed that the approach retains original relationships after downscaling. This will require further statistical analysis as suggested in Dong et al. (2020). In this research, the models were assessed on two levels, the first being feature importance, and the second, partial dependencies. Many researchers have used the RF algorithm and other similar ones to successfully downscale or generate synthetic NDVI and LST. Most works do not further investigate the models beyond the feature importance, making the models seem like black boxes. By employing the partial dependence plots, how the model works based on changes in each feature is revealed in this research. This information is useful in explaining the model to non-technical audiences and in simulation studies.

Comparing the results of the NDVI validations to the work of Filgueiras et al. (2020), it was noted that the RMSEs although low, were not as low as theirs (< 0.09). This is attributed to differences in the features used. Surface reflectance were employed in their work while LULC was used in this research. To get NDVI for each day like this research aimed at, it is important to use features that are available at the same temporal and spatial resolutions. Therefore, it may be challenging to use surface reflectance, especially when satellite missions are the main sources. With the proliferation of drones, this challenge may be overcome as seen in the case of Bonafoni et al. (2016) but it might require a good deal of fieldwork and planning. Another alternative will be using radiative models to obtain the spectral reflectance as the sensors would (Gastellu-Etchegorry et al. 2012). This approach uses the proportions of land use types in a cell as well as time to predict the NDVI with considerable accuracy and is replicable using any land use database given its simplicity.

This approach used one spectral index to predict LST, but other spectral indices like normalized difference sand index (Pan et al. 2018) and normalized difference built-up index (Xu et al. 2020) are also useful. However, some surface materials like roofs are not properly represented by spectral indices as indicated in Bonafoni et al.(2016). Hence this approach of using a good spectral index in addition to information on fractions of land surface types and spatiotemporal features improves prediction quality. This is evident in the works of Hutengs and Vohland (2016) and Li et al. (2019) where spectral indices and other spatial features were combined. Furthermore, some spectral indices may be correlated and would be redundant features in the model. Other parameters that are available daily, like solar radiation and wind speed could be considered as potential features to improve the LST predictions. The sky-view factor is a good parameter and could also be used to estimate solar radiation and model LST (Hutengs and Vohland 2016; Scarano and Mancini 2017). Even though the desired LST was achieved in two steps, the intermediary output-NDVI is a useful parameter for assessing UHI and other phenomena making it worthwhile. A limitation of this approach lies in the number of over and underestimations. Unfortunately, forest and cropland patches massively cover the study area;

patches where the model fails to perform well accounting for the high number of overestimations. Therefore, the synthetic LST will not be suitable for studies that will focus on forested areas. Regardless, this does not affect the usability of the data for assessing UHI and the spatio temporal variability of surface temperature within the built-up areas.

Furthermore, while using a grid cell as the spatial unit for analysis in this approach is more convenient, spatial objects do not exist in space as regular grids, but rather patches (Blaschke 2010). Hence adopting an object-based approach at the desired resolutions could yield products that better capture the NDVI and LST of certain LULC types. The Estonian LULC database was used because of its level of detail and extent of the study area. However, to scale this approach to cover urban areas from different geographic regions, it will be convenient to explore LULC data from Open Street Map or Local Climate Zone (LCZ) data from the World Urban Database and Access Portal Tools (WUDAPT). The LCZ from WUDAPT is based on a land use classification scheme that incorporates the physical and thermal properties of surface features (Stewart and Oke 2012; Bechtel et al. 2019; Dutta et al. 2021). Furthermore, this approach focuses on the 2D urban space; however, Zhang et al. (2019) point the relevance of vertical urban forms for explaining daytime LST. Hence replacing 2D information with 3D information will be appropriate if the influence of vertical urban form is of interest (Middel et al. 2017; Zhang et al. 2019).

Conclusion

This study presents a two-part approach to generate daily synthetic NDVI and LST at suitable resolutions to assess SUHI and heatwaves at an urban scale. The target resolutions were selected based on how the SE maps the population in the study area.

The first part focused on generating Landsat 8-like NDVI and MODIS-like NDVI at the desired resolutions using the RF algorithm with LULC fractions and DOY as the predictors. The modeling of the NDVI reveals that while the NDVIs from both sensors are highly correlated, they have differences for certain land cover types, especially water. To proceed, the mean NDVI predictions of the Landsat 8-like NDVI and MODIS-like NDVI were used. The mean was considered a better product because it compensates for biases from both models. The second part deals with modeling MODIS LST with RF using NDVI and other spatial and temporal features. The final outputs were daily LST similar to daily mean LST from MODIS Terra and Aqua satellites. The synthetic LSTs are useful for assessing the spatial and temporal variability of surface temperatures during the heatwave. Furthermore, the vulnerability of citizens to heatwaves in the study area can be studied using the synthetic products. This is because they were generated to match the spatial units in which the authorities map the population distribution. Despite these advantages, the synthetic LST products are most suitable for built-up areas and not for forest or cropland areas per the assessments. To make better predictions of LST for forests and croplands, other inputs like solar radiation and longwave radiations could be included.

RF is further established as a robust algorithm for modeling complex spatial relationships and for generating synthetic products of this nature by this approach (Li et al. 2019; Filgueiras et al. 2020). Furthermore, the partial dependency assessments performed on the models reveal how the models operate and could be informative for policymakers and researchers. Overall, the approach is simple to implement given the availability of open-source data. While the approach is considered a success, incorporating other useful features mentioned in the discussions and using spatial objects instead of regular grid cells could improve the accuracy of the estimations. Future works would consider other useful features to improve the model in areas where it performs poorly.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13762-022-04510-3.

Acknowledgements The Ministry of Education and Research of Estonia supported this work was under Grant PRG-352 and Distribution and Trend of the Extent of Heat Islands in Tallinn in 2014–2018 Project under Grant LLTOM19649

Declarations

Conflict of interest The authors declare that they have no known competing interests concerning this work.

References

Abdollahipour A, Ahmadi H, Aminnejad B (2021) A review of downscaling methods of satellite-based precipitation estimates. Earth Sci Inf. https://doi.org/10.1007/S12145-021-00669-4

Bala R, Prasad R, Yadav VP (2020) Thermal sharpening of MODIS land surface temperature using statistical downscaling technique in urban areas. Theor Appl Climatol 141:935–946. https://doi.org/10.1007/s00704-020-03253-w

Barsi JA, Barker JL, Schott JR (2003) An atmospheric correction parameter calculator for a single thermal band earth-sensing instrument. Int Geosci Remote Sens Symp 5:3014–3016. https://doi.org/10.1109/igarss.2003.1294665

Bechtel B, Alexander PJ, Beck C et al (2019) Generating WUDAPT Level 0 data—current status of production and evaluation. Urban Clim 27:24–45. https://doi.org/10.1016/j.uclim.2018.10.001

Bechtel B, Zakšek K, Hoshyaripour G (2012) Downscaling land surface temperature in an urban area: a case study for Hamburg,

Germany. Remote Sens 4:3184–3200. https://doi.org/10.3390/rs4103184
Blaschke T (2010) Object based image analysis for remote sensing. ISPRS J Photogramm Remote Sens 65:2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004
Bonafoni S, Anniballe R, Gioli B, Toscano P (2016) Downscaling landsat land surface temperature over the urban area of Florence. Eur J Remote Sens 49:553–569. https://doi.org/10.5721/EuJRS20164929
Breiman L (2001) Random forests. In: Schapire RE (ed) Machine Learning, 45th edn. Kluwer Academic Publishers, pp 5–32
Buo I, Sagris V, Jaagus J (2021) Gap-filling satellite land surface temperature over heatwave periods with machine learning. IEEE Geosci Remote Sens Lett 1–5. https://doi.org/10.1109/LGRS.2021.3068069
Cristóbal J, Jiménez-Muñoz JC, Prakash A et al (2018) An improved single-channel method to retrieve land surface temperature from the landsat-8 thermal band. Remote Sens 10. https://doi.org/10.3390/rs10030431
Dong P, Gao L, Zhan W et al (2020) Global comparison of diverse scaling factors and regression models for downscaling Landsat-8 thermal data. ISPRS J Photogramm Remote Sens 169:44–56. https://doi.org/10.1016/j.isprsjprs.2020.08.018
Dutta K, Basu D, Agrawal S (2021) Evaluation of seasonal variability in magnitude of urban heat islands using local climate zone classification and surface albedo. Int J Environ Sci Technol. https://doi.org/10.1007/S13762-021-03602-W
Filgueiras R, Mantovani EC, Fernandes-Filho EI et al (2020) Fusion of MODIS and landsat-like images for daily high spatial resolution NDVI. Remote Sens 12:11–13. https://doi.org/10.3390/RS12081297
Gastellu-Etchegorry JP, Grau E, Lauret N (2012) DART: A 3D Model for Remote Sensing Images and Radiative Budget of Earth Surfaces. In: Modeling and Simulation in Engineering. InTech
Geletič J, Lehnert M (2016) GIS-based delineation of local climate zones: the case of medium-sized Central European cities. Morav Geogr Reports 24:2–12. https://doi.org/10.1515/mgr-2016-0012
Good EJ, Ghent DJ, Bulgin CE, Remedios JJ (2017) A spatiotemporal analysis of the relationship between near-surface air temperature and satellite land surface temperatures using 17 years of data from the ATSR series. J Geophys Res Atmos 122:9185–9210. https://doi.org/10.1002/2017JD026880
Hu L, Sun Y, Collins G, Fu P (2020) Improved estimates of monthly land surface temperature from MODIS using a diurnal temperature cycle (DTC) model. ISPRS J Photogramm Remote Sens 168:131–140
Hutengs C, Vohland M (2016) Downscaling land surface temperatures at regional scales with random forest regression. Remote Sens Environ 178:127–141. https://doi.org/10.1016/j.rse.2016.03.006
Hutengs C, Vohland M, Hazaymeh K et al (2018) Downscaling Land Surface Temperature in an Urban Area: A Case Study for Hamburg, Germany. Remote Sens 10:3184–3200. https://doi.org/10.1038/s41598-018-27905-0
Ju J, Roy DP, Vermote E et al (2012) Continental-scale validation of MODIS-based and LEDAPS Landsat ETM+ atmospheric correction methods. Remote Sens Environ 122:175–184. https://doi.org/10.1016/j.rse.2011.12.025
Ke Y, Im J, Lee J et al (2015) Characteristics of Landsat 8 OLI-derived NDVI by comparison with multiple satellite sensors and in-situ observations. Remote Sens Environ 164:298–313. https://doi.org/10.1016/j.rse.2015.04.004
Li W, Ni L, Li ZL et al (2019) Evaluation of machine learning algorithms in spatial downscaling of modis land surface temperature. IEEE J Sel Top Appl Earth Obs Remote Sens 12:2299–2307. https://doi.org/10.1109/JSTARS.2019.2896923
Malamiri HRG, Rousta I, Olafsson H et al (2018) Gap-filling of MODIS time series land surface temperature (LST) products using singular spectrum analysis (SSA). Atmosphere (basel) 9. https://doi.org/10.3390/atmos9090334
Middel A, Lukasczyk J, Maciejewski R (2017) Sky view factors from synthetic fisheye photos for thermal comfort routing—A case study in Phoenix, Arizona. Urban Plan 2:19–30. Doi: https://doi.org/10.17645/up.v2i1.855
Moisen GG (2008) Classification and regression trees. In: Orgensen SE, Fath BD (eds) Encyclopedia of Ecology, 1st edn. Routledge, Oxford, U.K., pp 582–588
Molnar C (2019) "Interpretable machine learning. A guide for making black box models explainable. Lulu, Morrisville, North Carolina, USA
Oke TR, Mills G, Christen A, Voogt JA (2017) Urban Climates. Cambridge University Press, Cambridge
Osborne PE, Alvares-Sanches T (2019) Quantifying how landscape composition and configuration affect urban land surface temperatures using machine learning and neutral landscapes. Comput Environ Urban Syst 76:80–90. https://doi.org/10.1016/j.compenvurbsys.2019.04.003
Pan X, Zhu X, Yang Y et al (2018) Applicability of Downscaling Land Surface Temperature by Using Normalized Difference Sand Index. Sci Rep 8:1–14. https://doi.org/10.1038/s41598-018-27905-0
Scarano M, Mancini F (2017) Assessing the relationship between sky view factor and land surface temperature to the spatial resolution. Int J Remote Sens 38:6910–6929. https://doi.org/10.1080/01431161.2017.1368099
Shalev-Shwartz S, Ben-David S (2013) Understanding machine learning: From theory to algorithms
Siji George CG, Sumathi B (2020) Grid search tuning of hyperparameters in random forest classifier for customer feedback sentiment prediction. Int J Adv Comput Sci Appl 11:173–178. Doi: https://doi.org/10.14569/IJACSA.2020.0110920
Sobrino JA, Jiménez-Muñoz JC, Sòria G et al (2008) Land surface emissivity retrieval from different VNIR and TIR sensors. IEEE Trans Geosci Remote Sens 46:316–327. https://doi.org/10.1109/TGRS.2007.904834
Steven MD, Malthus TJ, Baret F et al (2003) Intercalibration of vegetation indices from different sensor systems. Remote Sens Environ 88:412–422. https://doi.org/10.1016/J.RSE.2003.08.010
Stewart ID, Oke TR (2012) Local Climate Zones for Urban Temperature Studies. Bull Am Meteorol Soc 93:1879–1900. https://doi.org/10.1175/BAMS-D-11-00019.1
Weng Q, Fu P (2014) Modeling diurnal land temperature cycles over Los Angeles using downscaled GOES imagery. ISPRS J Photogramm Remote Sens 97:78–88
Weng Q, Fu P, Gao F (2014) Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data. Remote Sens Environ 145:55–67. https://doi.org/10.1016/j.rse.2014.02.003
Xu J, Zhang F, Jiang H et al (2020) Downscaling ASTER land surface temperature over urban areas with machine learning-based area-to-point regression kriging. Remote Sens 12. https://doi.org/10.3390/rs12071082
Yu X, Guo X, Wu Z (2014) Land surface temperature retrieval from landsat 8 TIRS-comparison between radiative transfer equation-based method, split window algorithm and single channel method. Remote Sens 6:9829–9852. https://doi.org/10.3390/rs6109829

Yuan F, Bauer ME (2007) Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in Landsat imagery. Remote Sens Environ 106:375–386. https://doi.org/10.1016/j.rse.2006.09.003

Zhang Y, Middel A, Turner BL (2019) Evaluating the effect of 3D urban form on neighborhood land surface temperature using Google Street View and geographically weighted regression. Landsc Ecol 34:681–697. https://doi.org/10.1007/s10980-019-00794-y

Zhao L, Zhang P, Ma X, Pan Z (2017) Land Cover Information Extraction Based on Daily NDVI Time Series and Multiclassifier Combination. Math Probl Eng 2017. https://doi.org/10.1155/2017/6824051