Guo, F., **Hu, D.**, **Schlink, U.** (2022):
A new nonlinear method for downscaling land surface temperature by integrating guided and Gaussian filtering
Remote Sens. Environ. **271** , art. 112915

A new nonlinear method for downscaling land surface temperature by integrating Guided and Gaussian filtering

Fengxiang Guo (Corresponding author)
Affiliation: Department of Urban and Environmental Sociology, UFZ-Helmholtz Centre for Environmental Research
E-mail address: fengxiang.guo@ufz.de
Postal address: Permoserstraße 15, 04318 Leipzig, Germany
Phone number: 0049 176 6864 3237

Die Hu
Affiliation: Department of Urban and Environmental Sociology, UFZ-Helmholtz Centre for Environmental Research
E-mail address: die.hu@ufz.de

Uwe Schlink
Affiliation: Department of Urban and Environmental Sociology, UFZ-Helmholtz Centre for Environmental Research
E-mail address: uwe.schlink@ufz.de

Abstract

Land surface temperature (LST), retrieved from thermal infrared (TIR) bands of remote sensing satellites, is an important parameter for various climate and environmental models. TIR bands detect a range of low-energy wavelengths, resulting in a coarser spatial resolution than other multispectral bands, and limiting applicability in heterogeneous urban regions. In this study, a new nonlinear method for LST downscaling, called Three Layers Composition (TLC), was proposed. The TLC integrates large-scale temperature variations, re-constructed detailed characteristics of LSTs, and strong boundary information. The performance of TLC is compared with disaggregation of radiometric surface temperature (DisTrad), thermal imagery sharpening (TsHARP), and random forest (RF) for a complex landscape in Beijing city, which has agriculture, forest, and massive impervious surfaces. The scale effects on the downscaled LSTs (DLST) were analyzed from the aspects of spatial resolution and spatial contexts. The experimental results indicate that the nonlinear algorithms (TLC and RF) perform better than linear methods (DisTrad and TsHARP). Indicated by coefficient of determination (R^2), centered root-mean-square error (*CRMSE*), and correlation coefficient (*CC*), TLC (R^2 = 0.901, *CRMSE* = 0.319, *CC* = 0.951) was the most effective and workable technique for predicting LSTs, followed by RF (0.768, 0.502, 0.874), TsHARP (0.544, 0.652, 0.734), and DisTrad (0.518, 0.751, 0.719). Larger experimental regions and larger ratios between initial and target resolution weaken the accuracy of DLST. TLC indicated a stronger ability to resist the influence of such scale effects. Traditional downscaling methods (DisTrad, TsHARP, and RF) are trained with global LST-predictor relationships and predict the DLST point by point, which can result in significantly biased estimates for very high or very low temperatures. Addressing this issue, TLC advantageously preserves the texture similarity between

LST and its predictors, and yields more precise DLST, which showed higher consistency with the reference LST. Considering high accuracy and low computation time, TLC may be a safe technique for LST downscaling in other regions and different remote sensing sensors.

Keywords: Downscaling land surface temperature; Landsat 8; Linear regression; Random Forest; Scale effect; Three Layers Composition method

1. Introduction

Land surface temperature (LST) is the skin temperature of the Earth's surface; it is required to properly study urban moisture and drought (Wan et al., 2004), to monitor spatiotemporal dynamics of urban heat islands (Nichol, 2005; Huang and Wang, 2019), and to describe the growth status of vegetation (Julien and Sobrino, 2009). Remote sensing (RS) satellites can supply up to date, highly covered and spatially explicit LSTs, but hardly get LSTs at both high spatial and high temporal resolution (Hutengs and Vohland, 2016; Pu, 2021). The LST products retrieved from Moderate Resolution Imaging Spectroradiometer (MODIS), for example, are available daily but the spatial resolution is only 1km, while Landsat thermal infrared data have a finer resolution (100 m), but a 16-days revisit cycle. A high spatiotemporal resolution of LST would be desirable for the assessment of thermal performance over multiple landscape configurations at local and block scale (Xu et al., 2020; Zawadzka et al., 2020). One solution would be an improvement of hardware and detector instruments, particularly the capacity of data transfer in orbit, however, this has high production costs and is time consuming (Wang et al., 2021). Another solution is to develop downscaling models based on the correlations between LST and ancillary biophysical parameters (e.g., surface reflectance ratio, land use and land cover types, and vegetation indices), which can be extracted from visible and near infrared bands of high-resolution RS data (Kustas et al., 2003; Hutengs and Vohland, 2016).

LST downscaling has attracted more and more interest during the past two decades, and techniques utilize image fusion, kernel-driven approaches and the combination of both (Gao et al., 2006; Weng, 2009; Wang et al., 2021). Kernel-driven statistical models are frequently used due to their simplicity and effectiveness in multiple natural conditions. Disaggregation of radiometric surface temperature (DisTrad) and thermal imagery sharpening (TsHARP) are two classical linear models (Kustas et al., 2003; Agam et al., 2007; Jeganathan et al., 2011), which apply least square regression between LST and normalized difference vegetation index (NDVI). These linear models are well suited and workable for relatively uniform landscapes, while they may be less appropriate for urban areas with high heterogeneity (Hutengs and Vohland, 2016; Xu et al., 2020). LSTs, representing thermal performance over complex Earth surfaces, are affected by multiple factors (e.g., wind, topography, and surface material), and applying only NDVI as predictor for LST is insufficient. Random forest (RF), as a nonlinear statistical ensemble algorithm, can solve these problems by building sequentially randomised and de-correlated decision trees for multi-factorial regression (Hutengs and Vohland, 2016; Xu et al., 2020). Compared with linear models, RF avoids over-fitting, handles multi-collinearity, and can model complex relationships between LST and multiple influencing factors. Applying the RF algorithm, Hutengs and Vohland (2016) re-constructed the LSTs at high resolution in Jordan for varied geographical environments with improved performance in comparison to TsHARP.

A commonality of RF and linear methods is that all these models are trained with global LST-predictor relationships, and then predict the LST point by point (Wu and Li, 2019; Pu, 2021). It is unquestioned that the underlying surface attributes are highly correlated with LSTs, but this association varies with the locations. Therefore, a global relationship might be not suitable for local LST downscaling, particularly in urban regions (Wang et al., 2021). Moreover, a point-by-point procedure is likely to result in a disruption of the spatial texture characteristics of LST, and generate a significant bias for very high or very low temperatures in the downscaled LST (DLST). The differences of LST between initial (low) resolution and target (high) resolution mainly suggest that lots of detailed information in sub-pixels at initial resolution is missing and the boundaries at which temperatures greatly change are inaccurate (Fig. 1). Therefore, it is important to develop a new downscaling model that can simultaneously consider the temperature value and its spatial neighborhood relationships.

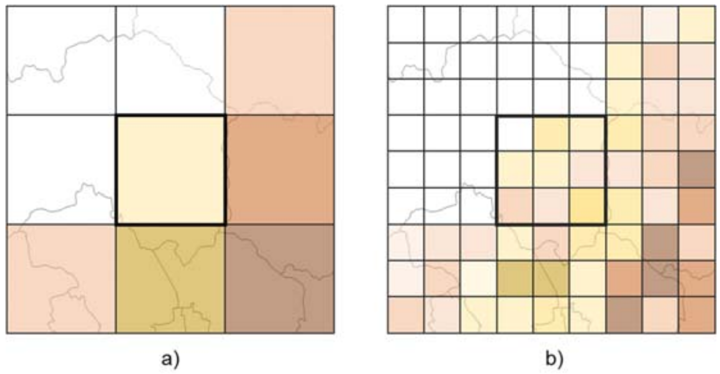

Fig. 1. Visual differences between LST at low (a) and high resolution (b).

Common methods for LST downscaling assume that the relationship between LSTs and predictors is scale-invariant, which has been questioned and needs more in-depth examination (Jeganathan et al., 2011; Chen et al., 2012; Pu, 2021). Previous researchers demonstrated the occurrence of scale effects when downscaling LST, which is usually caused by varied probability distributions of LST and different influencing factors between the initial and target resolution (Zhou et al., 2016; Pu, 2021). The spatial context, i.e. the region covered by the LST map, is another variable affecting the accuracy of DLST. Generally, the accuracy of DLST tends to decrease with an increase of the ratio from initial to target resolution, and an increase of the spatial context (Chen et al., 2012). At low resolution, the thermal performance of a pixel results from the within-pixel mixture of land cover and is constrained by the dominant land cover type. Their influence on the spatial variations of LST are relatively uniform and the scale remains relatively stable. At high resolution, each pixel is relatively pure, and the influence of varied land cover types on LST might be not scale-invariant, but multiscale (Pu, 2021): particularly when the target resolution is in a range of 20-30 m, downscaling processes proved to be not safe and results were not reliable.

This study proposes a new downscaling method called Three Layers Composition (TLC) based on image processing, and aims to demonstrate its suitability and advantages by 1) evaluating the downscaled LST over different land use and land cover types, estimated with a machine learning model (RF) and two linear methods (DisTrad and TsHARP); and 2) discussing the performance of different methods at varied target resolutions and contexts, and assessing the ability of TLC to reduce the influence of scale effects on DLST maps. For applicability in operational LST downscaling, we clarify which method works safely at which scale requirements (resolution and context).

2. Study area and Data

2.1. Study area

Beijing (39˚54′N, 116˚23′E), covering a total area of about 16000 km^2, is the political and cultural center of China. The terrain of Beijing gradually decreases from northwest to southeast, and the main urban region is located in the south plain (Fig. 2d). Beijing has a humid continental monsoon climate with severe, dry winters, hot summers and strong seasonality with an annual mean temperature ranging from 10 ℃ to 12 ℃ and mean precipitation ranging from 450 mm to 550 mm. The study area is located in the center of Beijing, with a spatial extension of 60 × 60 km, and the main land cover types include impervious surface distributed in the south, agriculture land distributed in the center and east, and forest distributed in the northeast mountain region (Fig. 2b). Since the 1980s, Beijing has witnessed a rapid urbanization and the urbanization level has reached 86% in 2010, having a significant influence on the urban thermal environment (Xiao et al., 2008; Peng et al., 2016).

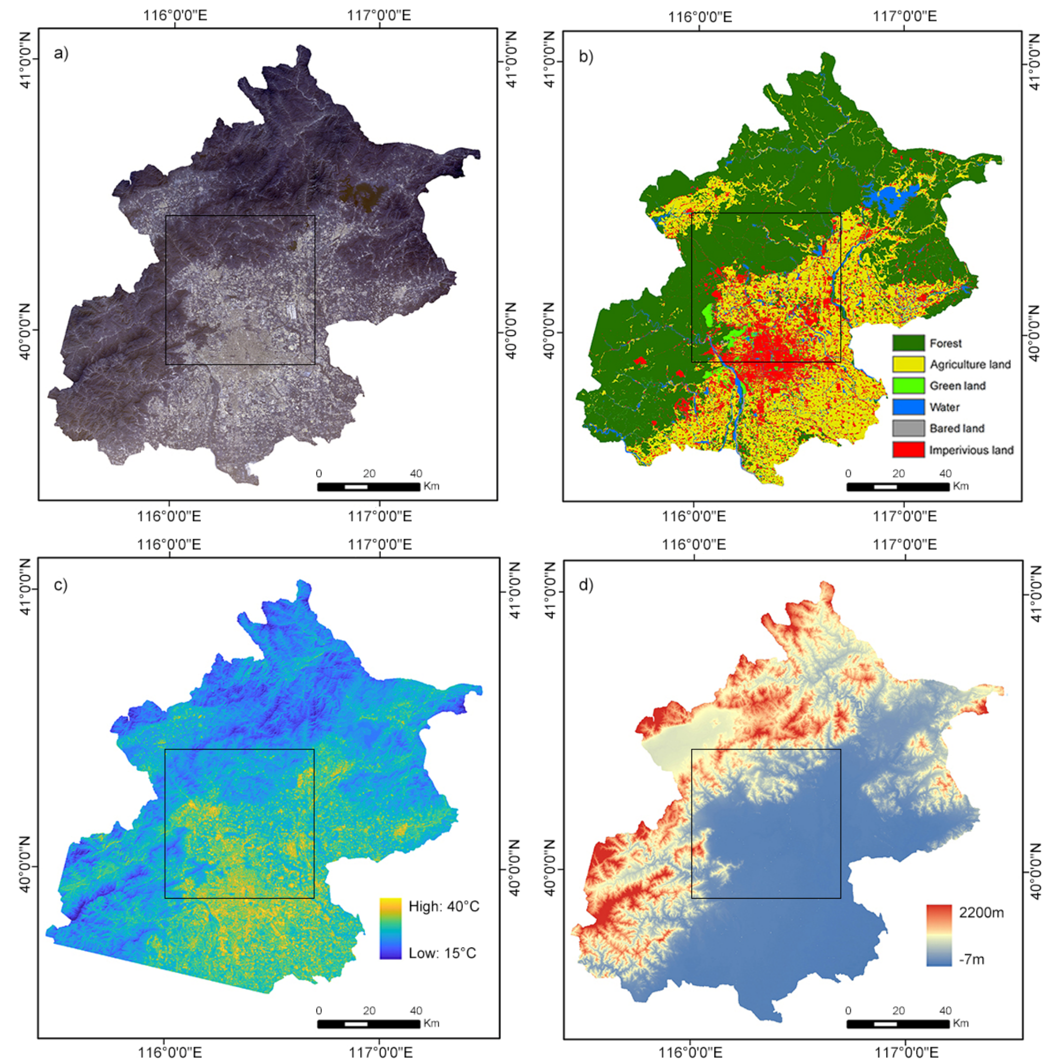

Fig. 2. Beijing region (black square marks the experimental area). (a) Landsat 8 band composite (RGB-band 213); (b) map of land use and land cover; (c) spatial distribution of LST; (d) elevation a.s.l.

2.2. Data

Landsat 8 OLI/TIRS data of September 12, 2017, obtained from the USGS website (https://earthexplorer.usgs.gov/), has been systematically processed with radiometric and geometric correction. Most of Landsat 8 bands have a resolution of 30 m except thermal infrared bands (100 m) and panchromatic band (15 m). Landsat 8 supplies LST products retrieved using the atmospheric correction method with 30 m spatial resolution (Yu et al., 2020), and provides auxiliary parameters for LST downscaling. NASA's

Shuttle Radar Topography Mission (SRTM) data were downloaded to extract the terrain factors (elevation, slope and aspect), which have a spatial resolution of 90 m.

The resampling method of nearest-neighbor interpolation was applied in this study to upscale Landsat 8 LST and auxiliary parameters to coarser resolutions of 60 m, 90 m, 120 m, 150 m and 300 m. SRTM data was resampled to 30 m and 60 m, and aggregated to 120 m, 150 m, and 300 m using spatial averaging. The LST map with 300 m resolution was taken as the initial resolution LST for downscaling, while others were taken as reference LST (RLST). The purpose of the RLST is to evaluate the accuracy of DLST at different target resolutions. All the data were geometrically corrected to WGS84/UTM Zone 50 N.

3. Methods

3.1. Three Layers Composition (TLC) method

Land use and land cover (LULC) types affect the spatial variations of LST significantly (Berger et al., 2017; Yu et al., 2020). Similar land covers at large scale form relatively smooth temperature variations with small gradients. Long stripes of land cover (e.g., rivers) and the junction of different land covers form a boundary texture in the temperature pattern, while tiny land patches shape detailed features. Both boundary and detailed features have large gradients. In an LST map with low resolution, texture boundaries and detailed features are severely missing. Therefore, for downscaling LST, we propose a new nonlinear approach TLC, which takes into account large-scale temperature variations, detailed LST characteristics, and boundary information. Using (1) a cubic convolution model, (2) a Gaussian low-pass filtering, and (3) a guided filtering, the TLC can properly extract the above mentioned three features from the low-resolution temperature image (T_{low}) and the high-resolution predictors (P) (see flow chart in Fig. 3).

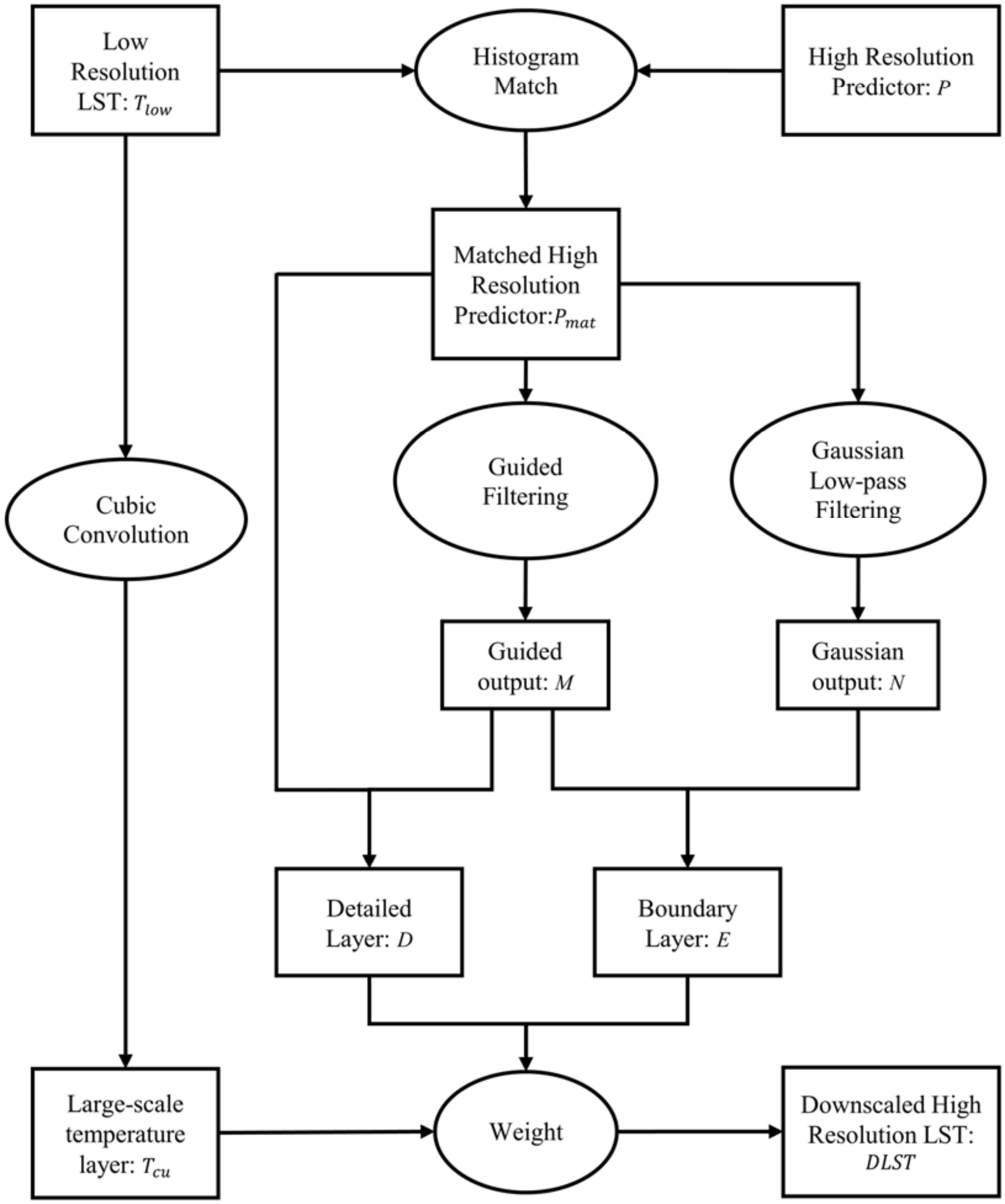

Fig. 3. Flow chart of the TLC method downscaling a low-resolution LST map.

3.1.1. Extraction of a large-scale temperature layer

The large-scale temperature layer (T_{cu}) represents the regional variations of LSTs at high spatial resolution, and is extracted from the low-resolution LST product (T_{low}). Remote sensing images describe LST maps as continuous surfaces with high spatial autocorrelation (Hutengs and Vohland, 2016). Although a low-resolution LST map $T_{low}(x, y)$ misses temperature characteristics at sub-pixel scale, it can well describe the large-scale temperature fluctuation in the study area. The two-dimensional cubic convolution interpolation function (1) was applied to obtain the large-scale temperature characteristics $T_{cu}(i_x, j_y)$ at pixel (i_x, j_y) of high resolution that is located inside pixel (x, y) of low resolution as follows (Keys, 1981):

$$T_{cu}(i_x, j_y) = \sum_{row=-1}^{2} \sum_{col=-1}^{2} T_{low}(x + row, y + col) S(\frac{i_x - i_{x+row}}{l}) S(\frac{j_y - j_{y+col}}{l}) \quad (1)$$

where (x, y) represents the interpolation node location, and l represents the sampling increment. The convolution kernel $S(x_1)$ is composed of piecewise cubic polynomials defined on the subintervals (-2, -1), (-1, 0), (0, 1), and (1, 2); this symmetric kernel vanishes outside the interval (-2, 2) (Keys, 1981):

$$S(x_1) = \begin{cases} (a_1 + 2)|x_1|^3 - (a_1 + 3)|x_1|^2 + 1, & 0 \le |x_1| \le 1 \\ a_1|x_1|^3 - 5a_1|x_1|^2 + 8a_1|x_1| - 4a_1, & 1 < |x_1| \le 2 \end{cases} \quad (2)$$

$S(x_1) = 0$ when $|x_1| > 2$. Parameter a_1 can be used to approximate different spline functions, specifically the interpolation error approaches 0 at a rate proportional to the third power of the sampling interval when $a_1 = -0.5$ (Keys, 1981; Reichenbach and Geng, 2003). Cubic convolution interpolation is theoretically an optimal approximation of the $sinc$ function (Meijering et al., 1999), which is effective for edge enhancement and the preservation of subtle features in comparison to nearest-neighbor and bilinear interpolation.

3.1.2. Extraction of boundary layer and detailed layer

Boundary and detailed layers are the key to reconstruct high-resolution LST using TLC, because they supply detailed information at sub-pixel scale. In contrast to the large-scale temperature layer (T_{cu}) obtained from low-resolution LST (section 3.1.1), the boundary and detailed layers are extracted from high-resolution predictors in two steps: Firstly, predictors (P), such as NDVI, ranging from -1 to 1, are linked to temperatures (T_{low}) by means of histogram matching, which preserves mean value and standard deviation of low-resolution LST (Zhang et al., 2019). This procedure ensures that the spatial locations and patterns of matched predictors (P_{mat}) are highly correlated with LSTs. The second step is to separate the boundary layer from the detailed layer. While both involve large gradients, the features of the latter are isotropic, i.e. in all directions. In contrast, boundary features have large gradients only in the normal direction and smaller gradients in the tangential direction. Based on this difference, this study applied a guided filtering to obtain the detailed layer, and a combination of Gaussian low-pass filtering and guided filtering to obtain the boundary layer.

The guided filtering (eq. 3) acts on a local square window ω_k centered at pixel k, and is a linear function between the guided map (T_{cu}) and the output (M), retaining the information which parts are boundaries and which are regions (He et al., 2012):

$$M(i,j) = c_k\, T_{cu}(i,j) + d_k, \qquad \forall (i,j) \in \omega_k \tag{3}$$

$$n(i,j) = P_{mat}(i,j) - M(i,j)\;; \quad Ex[n] = mean(n) \xrightarrow{c_k, d_k} Min \tag{4}$$

where c_k and d_k represent linear coefficients assumed to be constant in the local window ω_k. To determine these coefficients, we include the constraint (eq. 4) that the output M results from the input image P_{mat} subtracting some unwanted components like noise/textures (n). Ex represents the mathematical expectation, and a linear ridge regression with regular terms was applied to minimizes the difference between input P_{mat} and output M (He et al., 2012). In result of guided filtering, the linear relationship (eq. 3) ensures that the spatial texture of M is as similar as possible to that of T_{cu} (Fig. 4). Finally, the detailed layer D representing features of small land patches in the natural world, is calculated from subtracting the guided output from the predictor map, which provides important detailed information on the DLST (eq. 5).

$$D(i,j) = P_{mat}(i,j) - M(i,j) \tag{5}$$

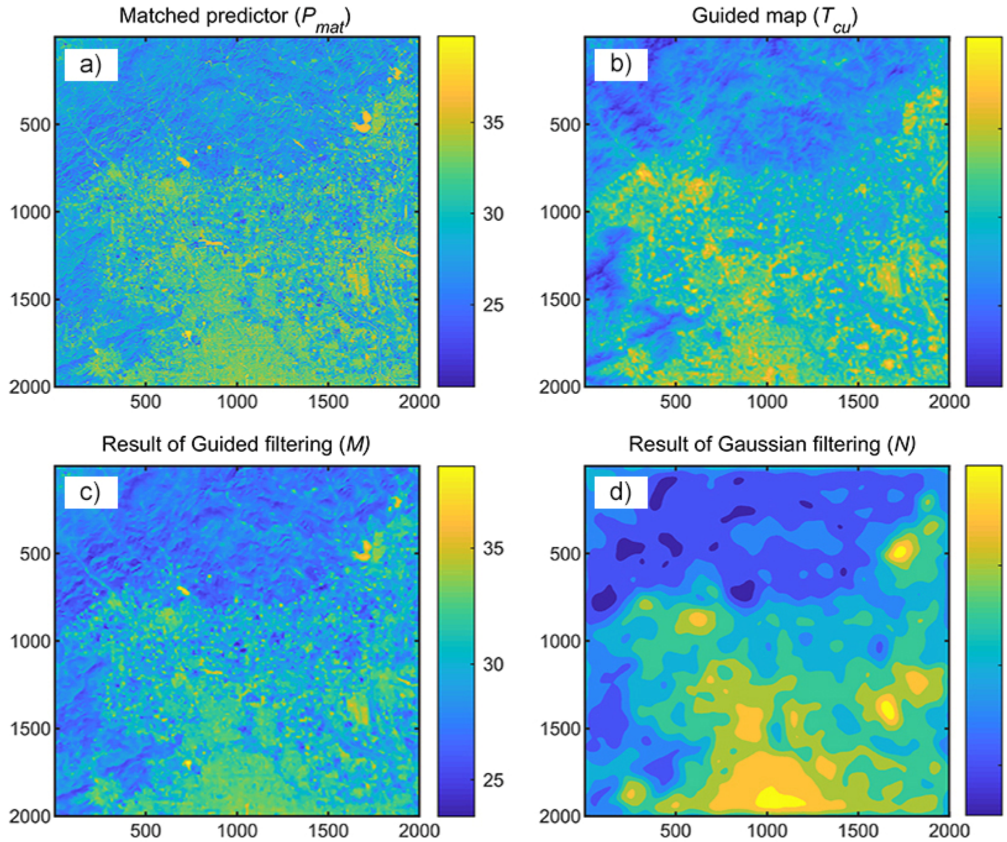

Fig. 4. (a) map of the predictor after matching; (b) guided map (T_{cu}); (c) result of guided filtering with window size 11; (d) result of Gaussian low-pass filtering with cut-off frequency 3.

To extract the boundary layer, a Gaussian low-pass filter (g) was applied to the predictor (eq. 6). It utilizes a Gaussian distribution kernel, removes high-frequency noise and preserves the low-frequency components (Haddad and Akansu, 1991).

$$\boldsymbol{N} = \boldsymbol{g} * \boldsymbol{P_{mat}} \tag{6}$$

Unlike guided filtering, the Gaussian low-pass acts isotropic and both the details and the boundaries in P_{mat} are weakened or even disappeared after filtering (Fig. 4d). Considering that the output M of guided filtering retains the boundary and regional features, the boundary layer can be described as (eq. 7):

$$\boldsymbol{E} = \boldsymbol{M} - \boldsymbol{N} \tag{7}$$

To properly merge all layers, the boundary and detailed layer obtained from high-resolution predictors were then transferred into the Tcu space by weighing ($\boldsymbol{weight} = \boldsymbol{Tcu}/\boldsymbol{P_{mat}}$), and the downscaled LST is (eq. 8):

$$\boldsymbol{DLST} = \boldsymbol{T_{cu}} + \boldsymbol{weight} \cdot (a \cdot \boldsymbol{D} + b \cdot \boldsymbol{E}) \tag{8}$$

a and b are constants adjusting the weight and integrating the detailed and boundary layers; the cut-off frequency of the Gaussian low-pass filtering and the window size of the guided filtering are another two parameters that need to be optimized. In this study, we set the cut-off frequency to 3, $\omega_k = 11,\ a = 0.3,\ b = 0.6$ (for detailed parameter specifications see supplementary material).

To predict urban LST, it is insufficient to use only one single predictor for different land cover types. In this study, NDVI was applied to predict the temperature in vegetated regions, NDBI was applied to predict the temperature in urban impervious regions, and NDWI was applied to predict the temperature in water bodies.

3.2. Additional methods for comparison with TLC

To assess the performance of the suggested TLC approach, three alternative models were applied (see Table 1 for input variables): DisTrad, TsHARP and RF. The first two are classical methods for LST downscaling based on its linear correlation to NDVI, while the nonlinear machine learning algorithm RF can model complex relationships between LST and predictors, accounts for both multicollinearity and nonlinearity, and avoids overfitting by averaging a large number of de-correlated individual trees (Hutengs and Vohland, 2016). The RF model parameters for the regression trees set-up include: 1) the number of regression trees (600); 2) minimum number of observations per tree leaf (5).

Table 1. Variables selected for downscaling methods in this study

Index	Name	Functions and characteristics	Application
NDVI	Normalized difference vegetation index	Well-documented negative relationship with LST, and positive relationship with soil moisture.	DisTrad; TsHARP; RF; TLC
NDBI	Normalized difference building index	High correlation with impervious surface area and less sensitive to seasonal change than NDVI.	RF; TLC
NDWI	Normalized difference water index	High linear correlation with LST, particularly over the water.	RF; TLC
SAVI	Soil-adjusted vegetation index	Interaction of soil properties and vegetation systems.	RF; TLC
BSI	Bare soil index	High correlation with bare soil.	RF
LULC	Land use and land cover types	Influence of underlying surface attributes on LST.	RF
Elevation	Terrain factors	High negative correlation with LST in mountain area.	RF
Slope			RF
Aspect			RF

DisTrad, TsHARP and RF for DLST calculation include following steps (Fig. 5):

(1) Assessment of the association between LST map and predictors at initial (lower) resolution:

$$\widetilde{T_{low}} = f(predictors_{low}) \tag{9}$$

For DisTrad, the prediction variable is NDVI, and a linear regression (eq. 10) is performed:

$$\widetilde{T_{low}} = a_0 + a_1 * NDVI_{low} \tag{10}$$

For TsHARP, the vegetation cover (*FVC*) was calculated and taken as predictor (eq. 11 and 12).

$$FVC = (1 - NDVI)^{0.625} \quad (11)$$

$$\widetilde{T_{low}} = b_0 + b_1 * FVC_{low} \quad (12)$$

For RF, sequential randomised and de-correlated decision trees represent the complex relationship between multiple predictors and low-resolution LST:

$$\widetilde{T_{low}} = f_{RF}(predictors_{low}) \quad (13)$$

(2) Prediction of LSTs at high resolution ($\widetilde{T_{high}}$) utilizing the downscaling models trained in Step (1):

$$\widetilde{T_{high}} = f(predictors_{high}) \quad (14)$$

(3) Improvement of the accuracy of high-resolution LST maps $\widetilde{T_{high}}$ calculated in Step (2) by help of error calibration. The error $\Delta\widetilde{T_{low}}$ was calculated as the difference between the LST product and the estimated LST at low resolution (eq. 15), and then downscaled into high resolution $\Delta\widetilde{T_{high}}$ by Kriging interpolation. Then the final DLST map at high resolution $\overline{T_{high}}$ was calculated (eq. 16):

$$\Delta\widetilde{T_{low}} = T_{low} - \widetilde{T_{low}} \quad (15)$$

$$\overline{T_{high}} = \widetilde{T_{high}} + \Delta\widetilde{T_{high}} \quad (16)$$

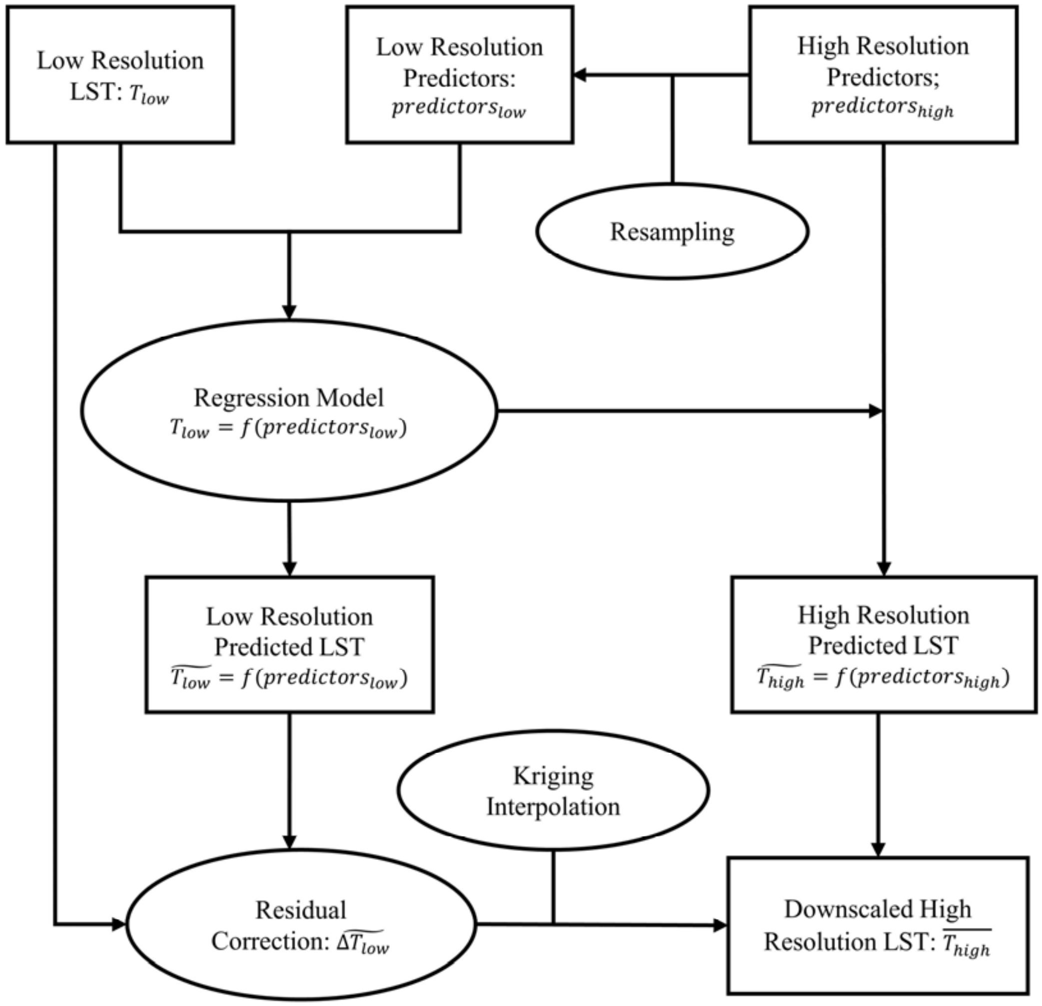

Fig. 5. Flow chart of traditional methods (DisTrad, TsHARP, and RF) for downscaling.

3.3. Descriptive statistics and error analyses

For the DLST and RLST maps, histograms and scatter density plots were created to evaluate the consistency of their spatial distributions, and the coefficient of determination (R^2) for their association. Taylor Diagrams were used to comparatively assess the different downscaling methods. Three statistics: standard deviation (STD), centered root-mean-square error ($CRMSE$), and correlation coefficient (CC) satisfy the cosine theorem utilized for the Taylor Diagram (Taylor, 2001):

$$CRMSE^2 = {STD_{DLST}}^2 + {STD_{RLST}}^2 - 2 * STD_{DLST} * STD_{RLST} * CC \qquad (17)$$

4. Results

4.1. Comparison between TLC and other methods for LST downscaling

TLC, DisTrad, TsHARP and RF were applied to downscale LST from initial resolution (300 m) to target resolution (30 m), respectively, and the LST map inversed from Landsat 8 was taken as reference LST (RLST, see Fig. 6b). All four DLST maps were similar to the RLST, and the best result was achieved by TLC, followed by RF, DisTrad and TsHARP based on visual inspection of spatial distribution locations and patterns.

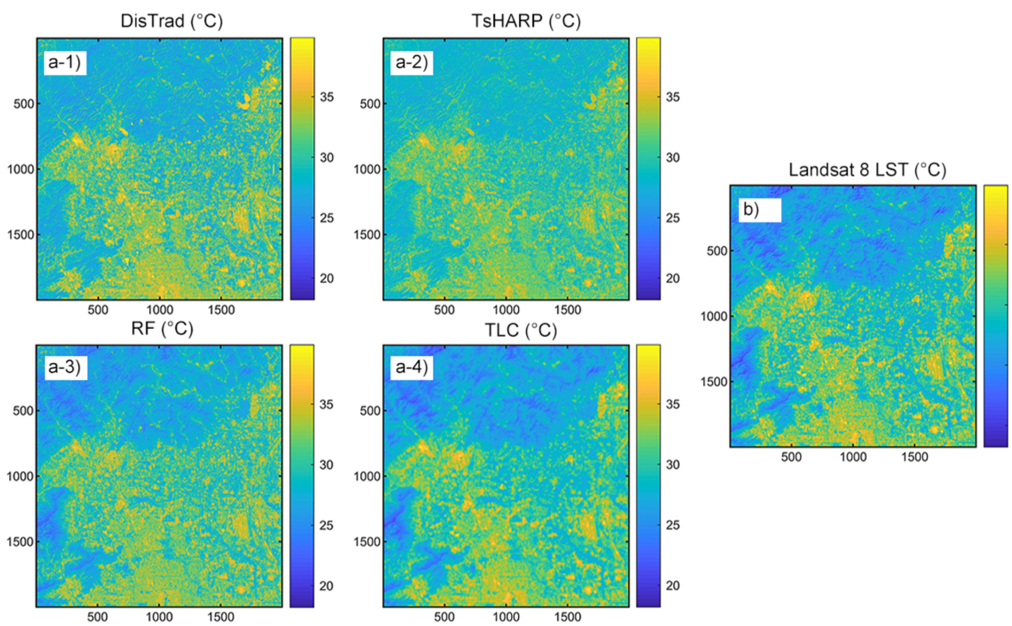

Fig. 6. (a) Downscaled LST maps and (b) Landsat 8 RLST.

Comparing paired histograms between DLST and RLST (Fig. 7a), the DLST downscaled by TLC were in closer agreement with the reference than the other methods. For DisTrad and TsHARP, there was significant bias, especially in the very low temperature range. Overestimates arise from RF around moderate temperatures (near 28 °C). Most residual errors between DLST and RLST were around zero (Fig. 7b). Residual errors of TLC were in the range from -1 °C to 1 °C, whereas the linear methods were most prone to residual errors exceeding 5 °C. At points of large noise the DLST calculated by linear methods was considerably deviating from the corresponding RLST (e.g. in the northern and western areas, where reverse temperature characteristics occurred). The best downscaling results were achieved for TLC (R^2 = 0.901), followed by RF (R^2 = 0.768), TsHARP (R^2 = 0.544) and DisTrad (R^2 = 0.518).

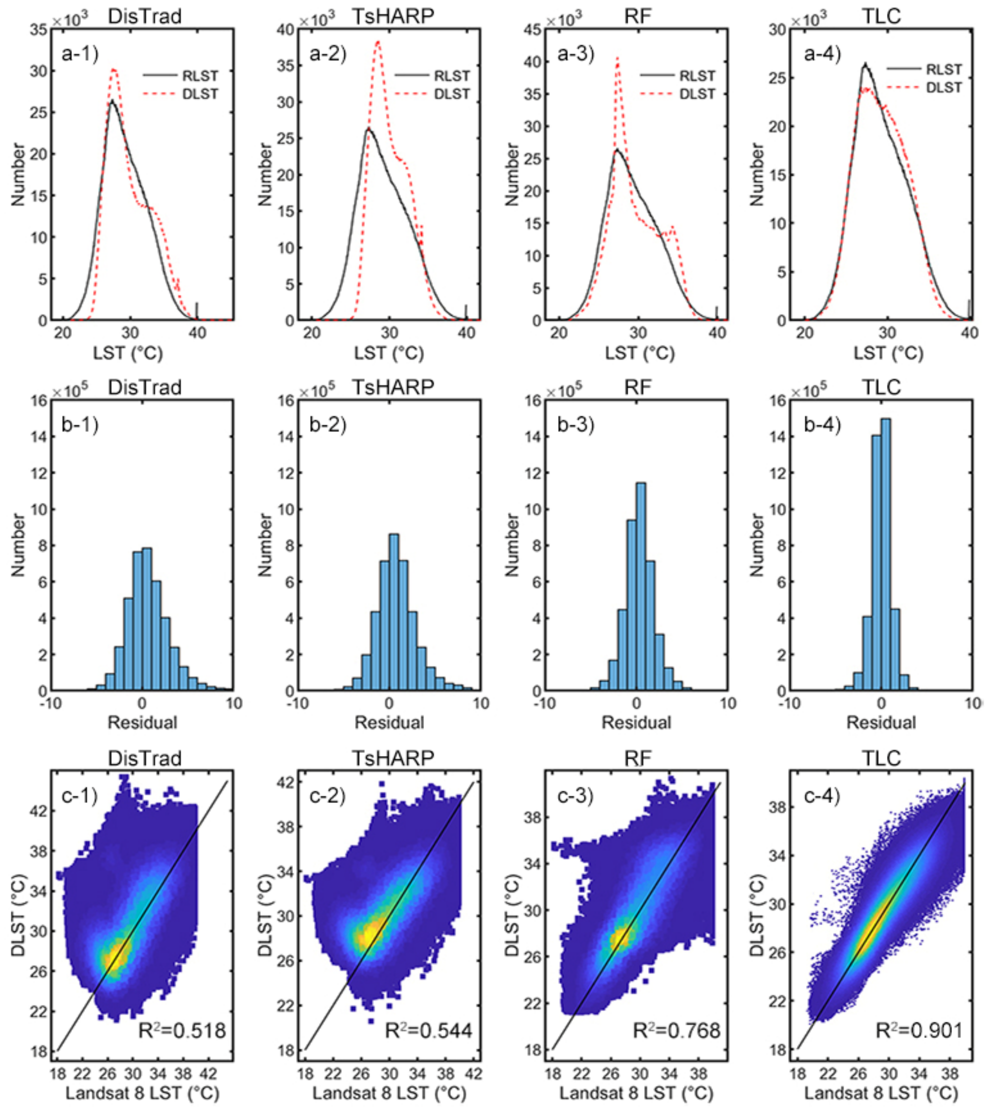

Fig. 7. (a) Histograms of DLST compared to Landsat 8 RLST. (b) Error distribution between Landsat 8 RLST and downscaled LSTs. (c) Scatter density plots of DLST (y-axis) versus RLST (x-axis) downscaling from initial resolution 300 m to target resolution 30 m. From left to right: DisTrad, TsHARP, RF and TLC.

In the Taylor Diagram (Fig. 8), the downscaling models were represented by solid dots of different colors, and the distance between each model and the reference point labelled 'XEF' is a measure of the performance of the downscaling method. A closer distance to the reference point means higher accuracy of the method. The Taylor diagram directly indicated that DLST accuracy was best for TLC (shortest distance to XEF, Fig. 8). TLC had lowest *CRMSE* (0.319) and highest *CC* (0.951), followed by RF (*CRMSE* = 0.502, *CC* = 0.874), TsHARP (*CRMSE* = 0.652, *CC* = 0.734) and DisTrad (*CRMSE* = 0.751, *CC* = 0.719). The *STDs* of DLST using DisTrad, RF and TLC were generally consistent with RLST, while TsHARP had much less variability in downscaled temperatures.

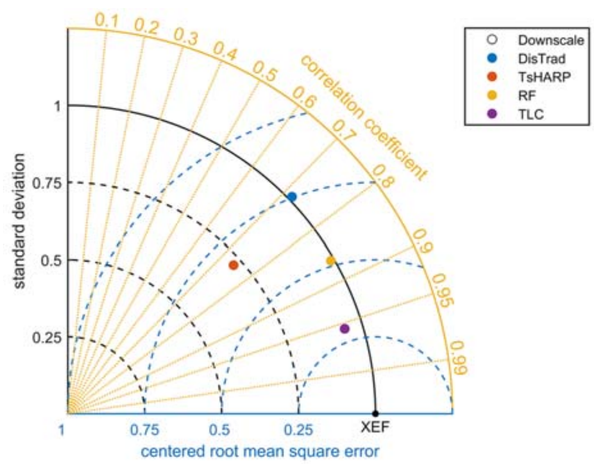

Fig. 8. Taylor diagram evaluating the accuracy of DLST using different downscaling methods. 'XEF' represents the reference data, the *CC* is related to the azimuthal angle (yellow lines), the *CRMSE* of the downscaled results is proportional to the distance to the reference point XEF (blue lines), and the *STD* is proportional to the radial distance to the origin (black lines).

Fig. 9 illustrated the accuracy of the downscaling methods in terms of different land cover types. Compared with TLC, RF was less accurate, but still outperformed TsHARP and DisTrad. For linear methods, the accuracy of DLST maps for forest was much lower than that of imperious surface land and agriculture land, because the terrain of imperious surface and agriculture land is relatively flat, while forest is mainly distributed in the mountain area. The linear methods do not consider any changes of the terrain and involve just the association with NDVI, which leads to a significant misrepresentation of forest temperature. RF and TLC predicted forest temperatures more accurately than the impervious surface and cultivated land temperatures. For the RF, terrain fluctuation is an important variable for LST prediction, and the influence of terrain on the temperature is well considered during the model training process. The higher spatial aggregation level and less heterogeneity of forest landscape cause a relatively higher prediction accuracy than the impervious surface with high spatial heterogeneity. Differently from the other methods, TLC takes the texture similarity between land cover and LSTs into consideration, which can well capture local temperature fluctuations and avoid massive noise in the DLST map.

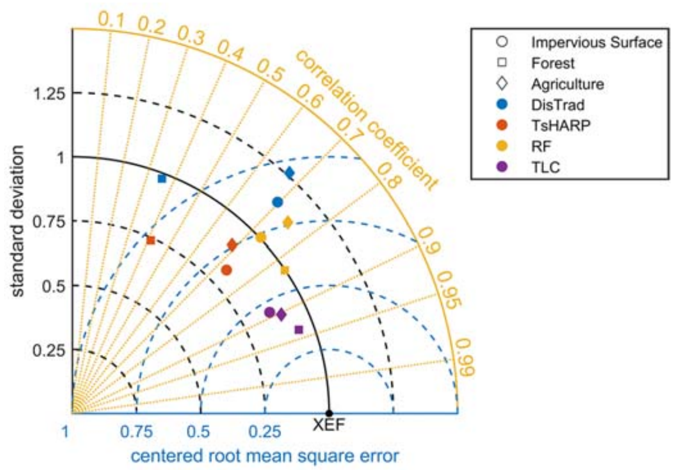

Fig. 9. Taylor diagram evaluating the accuracy of DLST maps for different land cover types (marked by symbol shapes: for example, the blue rectangular box marks the accuracy of forest LST obtained from DisTrad).

4.2. The influence of scale effects on DLST calculated by different methods

4.2.1. Accuracy depending on spatial resolution

To study the influence of spatial resolution on LST downscaling for all considered methods, we firstly downscaled LST from 300 m spatial resolution to 30 m, 60 m, 90 m, 120 m, and 150 m within a spatial context 60 × 60 km (Fig. s1 in supplementary material). Generally, the spatial distribution and patterns of DLST were consistent to RLST, but the DLST maps using nonlinear methods were visually much smoother and more similar to RLST than those using linear methods.

With increasing ratio from initial to target resolution, R^2 tended to decrease (Fig. 10), because a higher ratio means that more detailed information needs to be added to DLST maps and more errors might be introduced. The drop of R^2 using TLC was only 0.038, much lower than that using RF (0.272), TsHARP (0.249) and DisTrad (0.326), which indicated that the TLC method might be more reliable than other methods.

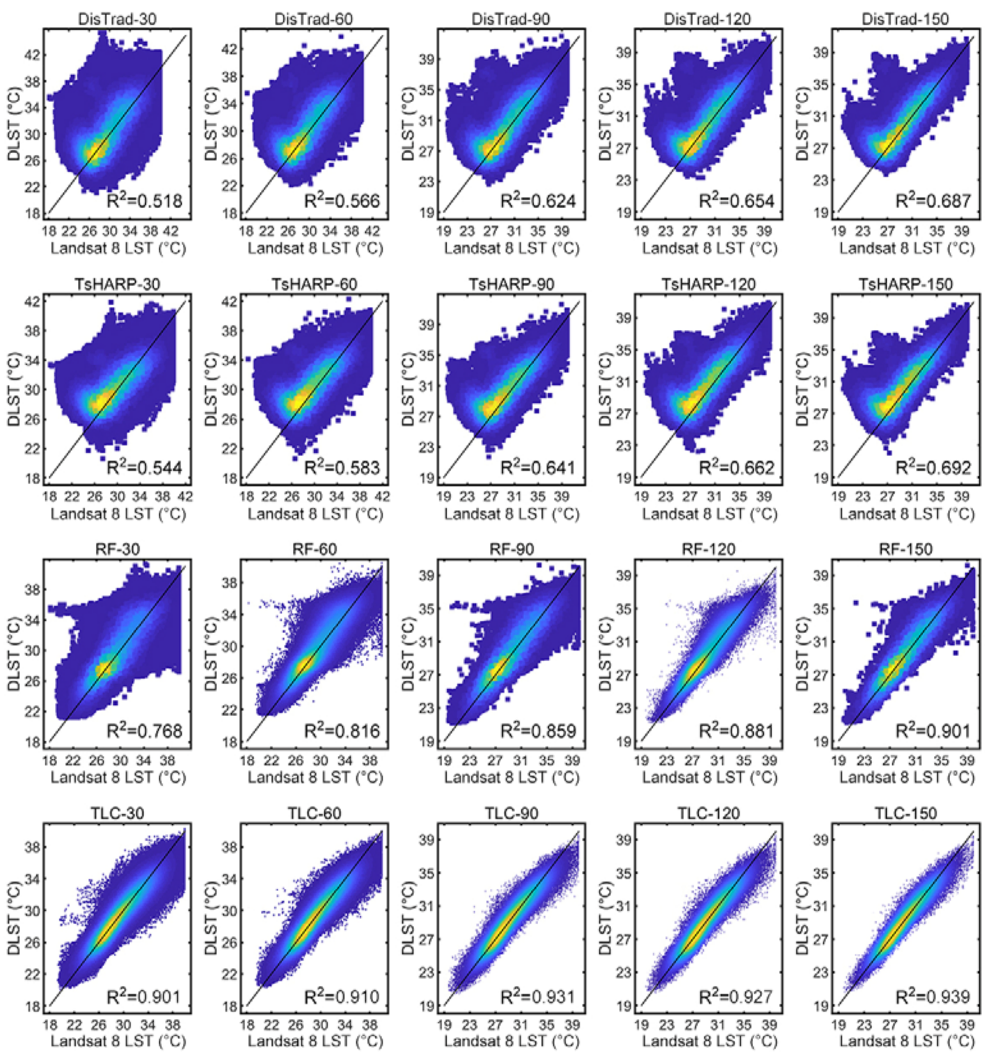

Fig. 10. Scatter density plots of DLST (y-axis) versus RLST (x-axis) extracted from Landsat 8 30 m LST products (from left to right: target resolution from 30 m to 150 m; from top to bottom: DisTrad, TsHARP, RF, and TLC).

An increase of spatial resolution clearly weakened the quality of DLST maps, but in different amounts (Fig. 11): The *CRMSE* of TLC decreased by around 0.07, much lower than for RF (0.224), TsHARP (0.231), and DisTrad (0.178). The *CC* showed similar changes, which indicated that the TLC might better resist the influence of scale effects.

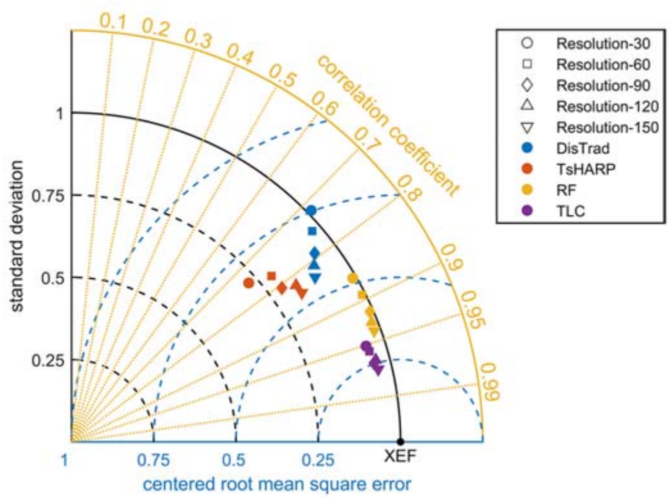

Fig. 11. Taylor diagram evaluating the accuracy of DLST maps for varied target resolutions.

4.2.2. Accuracy depending on spatial context

The LST at 300 m resolution was downscaled into 30 m resolution for different spatial contexts: 15 × 15 km, 30 × 30 km, 45 × 45 km, and 60 × 60 km. For varied regions, the smaller region was included in the larger region, for example, the region 15 × 15 km was located in the center of the region 30 × 30 km. In results (Fig. s2, supplementary material), for extreme temperatures, the smoothness in DLST maps from DisTrad and TsHARP was significant, and there was an obvious overestimation of low temperatures compared with nonlinear methods.

Downscaling performance of TLC and RF was reasonable, while the results of

DisTrad and TsHARP were relatively poor (Fig. 12). With increasing spatial context, the R^2 of TsHARP, RF and TLC showed similar variations that the values increased first from 15 × 15 km to 30 × 30 km, and then decreased gradually. A smaller area has less pixel samples, that is why the model accuracy is relatively lower, particularly for the linear models, suggested by the R^2 of DisTrad (0.359) and TsHARP (0.445). For study regions of very large size, the trained models can hardly capture each detailed characteristics between LST and its predictors, which might lead to a decrease of DLST accuracy. Compared with traditional methods, the R^2 of TLC stayed at a relatively high level, and its detoriation was much lower with increasing spatial context (similar conclusions result from the Taylor diagram (Fig. 13). We find that the main reason for the high performance of TLC is that traditional methods downscale LST with global LST-predictor relationships, which might be not suitable for local temperature prediction. TLC relied on the texture similarity between LST and predictors, which can well consider the autocorrelation of LST, and can preserve the local variations of actual LST as much as possible.

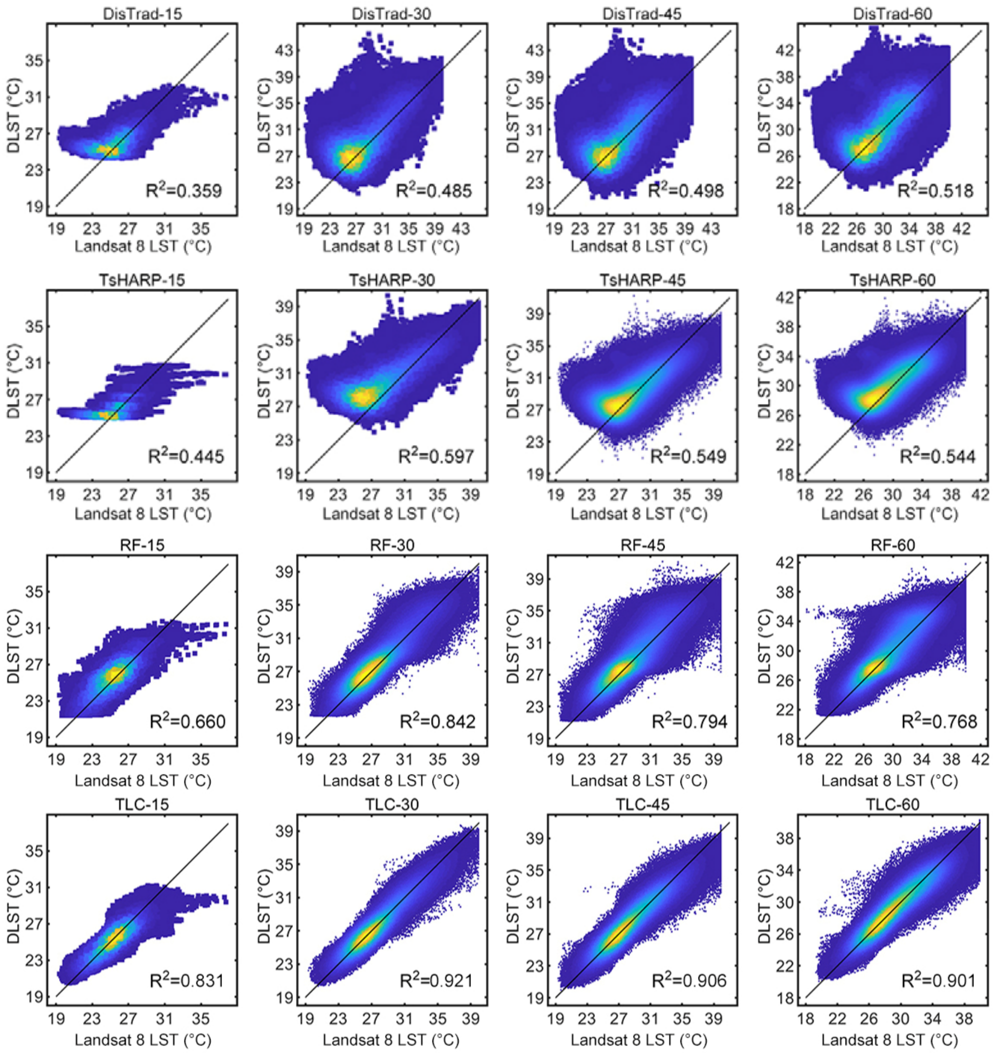

Fig. 12. Scatter density plots of DLST (y-axis) versus RLST (x-axis) extracted from Landsat 8 30 m LST products (from left to right: varied spatial context from 15 × 15 km to 60 × 60 km; from top to bottom: DisTrad, TsHARP, RF, and TLC).

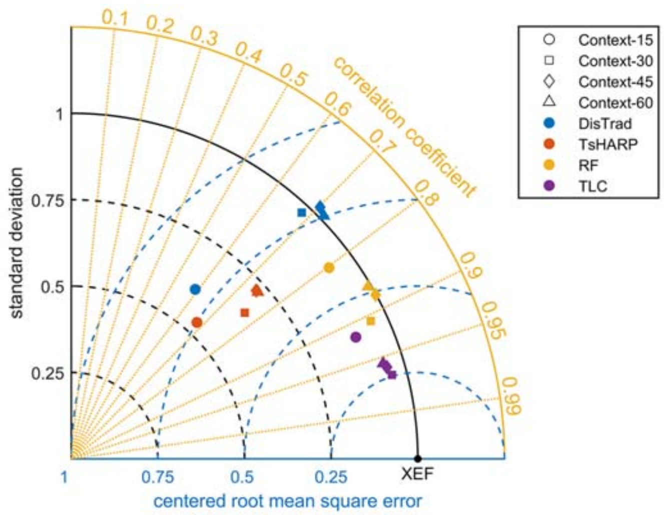

Fig. 13. Taylor diagram evaluating the accuracy of DLST maps over varied spatial context.

5. Discussions and Conclusions

5.1. Advantages of TLC in LST downscaling

One of the main goals in this study was to evaluate the suitability and effectiveness of the new TLC method for downscaling LST in highly heterogeneous regions, and the results, particularly the Taylor Diagram indicated that TLC outperforms traditional methods. The accuracy of DLST obtained from RF was much higher than that using linear methods, similar to the findings of Hutengs and Vohland (2016). In the urban region, the relationship between LST and its influencing factors is not linear (Peng et al., 2016; Wang et al., 2021), and the complexity of landscape composition and configuration is likely to result in significant variations of temperature (Berger et al., 2017; Yu et al., 2020). Among the nonlinear methods, TLC was superior for downscaling LST than RF, because the decision trees used in the RF regression were carried out based on the global LST-predictor relationships, which lack of the consideration of spatial autocorrelation and local correlation. This insufficiency might disrupt the spatial neighborhood relationships, and lead to massive noise compared to RLST. Unlike global LST-predictor relationships, the Gaussian low-pass filtering and guided filtering applied in TLC both act in a local window, ensuring that the local DLST variations are only affected by the local predictors and can effectively avoid interferences from other pixels. TLC has the advantage of being able to reconstruct the missing local information at target resolution based on texture similarity between LST and multiple predictors, and the combination of detailed information and large-scale temperature variations help maintain the continuity and consistency of DLST.

Among different land cover types, the predicted temperature at forest regions using linear methods showed lower accuracy and weak consistency with RLST than that at impervious surface land. However, this relationship was reverse for nonlinear methods, supporting findings of Wu and Li (2019) who also applied TsHARP and RF models to downscale LST in Beijing city. The differences over varied land cover types are mainly caused by terrain factors. In Beijing, forests are mostly distributed in the northern and western mountain region with a stronger variability of the terrain. For linear models, the influence of terrain on the DLST was not included during the model training process. The nonlinear methods (RF and TLC) avoid this insufficiency, particularly the latter suggested better performance in forested regions than for sealed surfaces. In the urban region highly covered with impervious surfaces, the spatial heterogeneity exceeds that of other areas significantly, and temperature patterns change greatly, leading to difficulties and errors in LST downscaling. However, the DLST using TLC showed significant over-smoothed characteristics in comparison to other methods (Fig. 14). The reason for this might be related to the selection of predictors. For each land cover type, TLC applied a single predictor for downscaling, while RF allows multiple indices to complement each other with detailed information. Despite different combinations of predictors in TLC, the correlation between NDBI and NDVI in the study region exceeds 0.93, indicating that NDVI is unlikely to provide more detailed information compared to NDBI. In addition, the quality of predictors, particularly the systematic error brought by the sensors in the remote sensing images also has a great influence on DLST. In future studies, the TLC algorithm should consider combinations of multiple predictors to deal with this insufficiency.

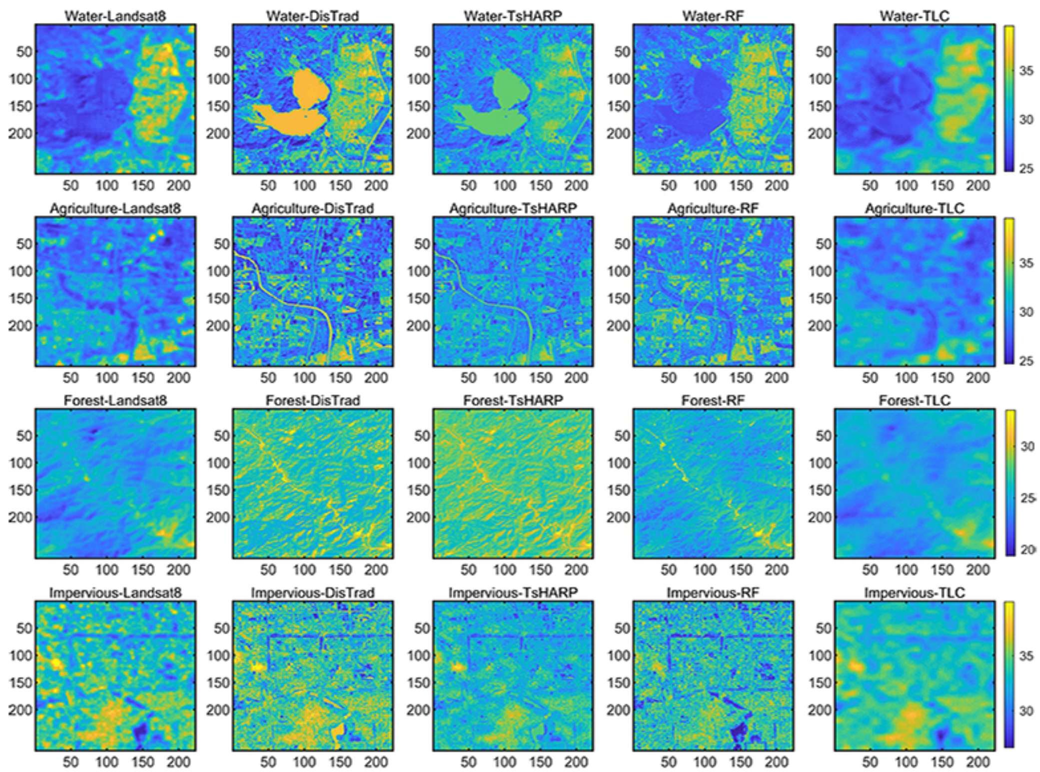

Fig. 14. Comparison of Landsat 8 RLST and downscaled LST maps for different land cover types (from left to right: Landsat 8, DisTrad, TsHARP, RF, and TLC; from top to bottom: water, agriculture land, forest, impervious surface land).

For the extreme temperature range, the DLST tended to be biased with an overestimation of low temperatures and an underestimation of high temperatures. A similar phenomenon was previously noted by Hutengs and Vohland (2016), and Xu et al. (2020). Compared to TLC, this bias was much more pronounced for DisTrad, TsHARP and RF. The inability of traditional downscaling methods to predict very high or very low temperatures might be caused by an insufficient number of training samples. Models trained with global variables must represent the entire LST map, which can make conventional models insensitive to local temperature extremes. TLC can overcome this shortcoming and achieved both a local as well as a global optimal solution for DLST by adding detailed information and strong boundary information into the large-scale temperature layer. The local optimal solution helps to control the dispersion and bias for extreme LST values, while the global optimal solution contributes to preserve the spatial texture and control the deviation of DLST relative to RLST.

5.2. Advantages of TLC for reducing scale effects in LST downscaling

The experimental results of scale effects indicated that the accuracy of downscaling LST was affected by the complexity of surface coverage significantly, and it is recommended that a range of spatial resolutions and contexts should be pre-calculated before conducting an LST downscaling project (Chen et al., 2012; Zhou et al., 2016). In this study, the accuracy of DLST maps was directly related to the ratio from initial to target resolution. With the increase of target resolution, the 'pureness' of pixels tends to be stronger and the thermal conditions over a surface are usually controlled by a specific land cover or land use type, different from 'mixed' pixels at coarser resolution with similar scaling factor pixel values. This effect leads to scale dependence in LST downscaling: the larger the ratio between initial and target resolution is, the more detailed information needs to be determined and added into the downscaling models (Jeganathan et al., 2011; Wu and Li, 2019), which potentially can increase the errors and decrease the accuracy of DLST maps.

To analyze the impact of the spatial context, this study considered regions of different size. As the size increases, the complexity of the surface coverage changes. The R^2 and Taylor Diagrams indicated that a region of 30 × 30 km might be most useful for RF and TLC downscaling. For linear methods, a larger experimental region means more training samples, for which the accuracy tended to increase, but was still much lower than for nonlinear methods. Considering both spatial resolution and spatial context, nonlinear methods, particularly TLC, indicated a better performance to resist scale effects on DLST. However, these implications arose only from applying our approach to the selected urban hotspot. The conclusion may be different for other regions depending on the location's peculiarities of terrain, climate and LULC types, and an increase of the study area size does not always result in an increased surface complexity of the considered area.

In addition to accuracy requirements, computation time is another key parameter for downscaling in a larger region. The ideal method should be highly accurate and less time consuming. RF is a widely used model for LST downscaling due to its high accuracy. However, many prediction parameters have to be pre-calculated and much time is consumed to train the regression model. In this study, RF took 100.4 seconds to downscale LST from 300 m to 30 m resolution within a spatial context of 60 × 60 km, nearly 10 times longer than DisTrad and TsHARP, while TLC took only 6.3 seconds. As the spatial context increases, the time requirement of RF increases sharply due to the larger number of training samples. Guided filtering assumes linearity between guided output and guided image in a local window, thus the computational speed of TLC benefits from this linearity. We conclude that the TLC is more accurate and less time-consuming, compared with traditional downscaling methods, which provides a good basis for obtaining regional or even global LST with high spatial resolution.

5.3. Limitations of this study and potential applications of TLC

TLC performed better than traditional methods for LST downscaling, however, it has a stronger dependence on the quality of predictors. TLC is guided by techniques of image processing, which requires that the predictors for LST downscaling should well reflect the land surface at one single day. Accuracy might be affected if the predictors are poor or multiday composites. RF is more flexible than TLC in selecting predictors and even multiday composite products can be included. In the future, the TLC algorithm should be tested with the inclusion of different types of predictors, as these can provide detailed information and constraints for downscaling, and expand the applicability of TLC.

The experimental design of this study had limitations. Firstly, downscaling accuracy was not analyzed for target resolutions greater than 30 m, due to limited spatial resolution of the thermal infrared sensors. Some literature, for example, Pu (2021) found significant differences between DLST and RLST within a target resolution range of 15-20 m. Secondly, the accuracy of DLST as well as Landsat 8 LST product was not verified using ground measured temperature. The latter is only based on existing literature (Berger et al., 2017). In the future, unmanned aerial vehicles with thermal infrared cameras and ground measured data might be used to demonstrate the superiority of the TCL method as well as the influence of scale effects on the DLST. In addition, this study proposed a new nonlinear method for LST downscaling, but tested only on LST products from Landsat 8. In future studies, TLC could be applied to other remote sensing sensors with coarser spatial resolution, such as MODIS, Advanced Very High Resolution Radiometer (AVHRR), and Infrared Spectrograph (IRS).

Author Contributions

Conceptualization, FG; Data processing, FG and DH; Formal analysis, FG and US; Methodology, FG; Writing—original draft, FG; Manuscript modification, FG, DH and US.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The first author would like to express his gratitude for the research support from China Scholarship Council under Grant No. 202008080124.

References

Agam, N., Kustas, W. P., Anderson, M. C., Li, F., & Neale, C. M. (2007). A vegetation index based technique for spatial sharpening of thermal imagery. *Remote Sensing of Environment,* 107(4), 545-558. DOI: 10.1016/j.rse.2006.10.006

Berger, C., Rosentreter, J., Voltersen, M., Baumgart, C., Schmullius, C., & Hese, S. (2017). Spatio-temporal analysis of the relationship between 2D/3D urban site characteristics and land surface temperature. *Remote sensing of environment*, 193, 225-243. DOI: 10.1016/j.rse.2017.02.020

Chen, X., Yamaguchi, Y., Chen, J., & Shi, Y. (2012). Scale effect of vegetation-index-based spatial sharpening for thermal imagery: A simulation study by ASTER data. *IEEE Geoscience and Remote Sensing Letters*, 9(4), 549-553. DOI: 10.1109/LGRS.2011.2174453.

Gao, F., Masek, J., Schwaller, M., & Hall, F. (2006). On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance. *IEEE Transactions on Geoscience and Remote sensing*, 44(8), 2207-2218. DOI: 10.1109/TGRS.2006.872081

Haddad, R. A., & Akansu, A. N. (1991). A class of fast Gaussian binomial filters for speech and image processing. *IEEE Transactions on Signal Processing*, 39(3), 723-727. DOI: 10.1109/78.80892

He, K., Sun, J., & Tang, X. (2012). Guided image filtering. *IEEE transactions on pattern analysis and machine intelligence*, 35(6), 1397-1409. DOI: 10.1109/TPAMI.2012.213

Huang, X., & Wang, Y. (2019). Investigating the effects of 3D urban morphology on the surface urban heat island effect in urban functional zones by using high-resolution remote sensing data: A case study of Wuhan, Central China. *ISPRS*

Journal of Photogrammetry and Remote Sensing, 152, 119-131. DOI: 10.1016/j.isprsjprs.2019.04.010

Hutengs, C., & Vohland, M. (2016). Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sensing of Environment*, 178, 127-141. DOI: 10.1016/j.rse.2016.03.006

Jeganathan, C., Hamm, N. A., Mukherjee, S., Atkinson, P. M., Raju, P. L. N., & Dadhwal, V. K. (2011). Evaluating a thermal image sharpening model over a mixed agricultural landscape in India. *International Journal of Applied Earth Observation and Geoinformation*, 13(2), 178-191. DOI: 10.1016/j.jag.2010.11.001

Julien, Y., & Sobrino, J. A. (2009). The Yearly Land Cover Dynamics (YLCD) method: An analysis of global vegetation from NDVI and LST parameters. *Remote sensing of environment*, 113(2), 329-334. DOI: 10.1016/j.rse.2008.09.016.

Keys, R. (1981). Cubic convolution interpolation for digital image processing. IEEE transactions on acoustics, speech, and signal processing, 29(6), 1153-1160. DOI: 10.1109/TASSP.1981.1163711.

Kustas, W. P., Norman, J. M., Anderson, M. C., & French, A. N. (2003). Estimating subpixel surface temperatures and energy fluxes from the vegetation index–radiometric temperature relationship. *Remote sensing of environment*, 85(4), 429-440. DOI: 10.1016/S0034-4257(03)00036-1

Meijering, E. H., Zuiderveld, K. J., & Viergever, M. A. (1999). Image reconstruction by convolution with symmetrical piecewise nth-order polynomial kernels. *IEEE transactions on image processing*, *8*(2), 192-201. DOI: 10.1109/83.743854

Nichol, J. (2005). Remote sensing of urban heat islands by day and night. *Photogrammetric Engineering & Remote Sensing*, 71(5), 613-621. DOI:

10.14358/PERS.71.5.613

Peng, J., Xie, P., Liu, Y., & Ma, J. (2016). Urban thermal environment dynamics and associated landscape pattern factors: A case study in the Beijing metropolitan region. *Remote Sensing of Environment*, 173, 145-155. DOI: 10.1016/j.rse.2015.11.027

Pu, R. (2021). Assessing scaling effect in downscaling land surface temperature in a heterogenous urban environment. *International Journal of Applied Earth Observation and Geoinformation*, 96, 102256. DOI: 10.1016/j.jag.2020.102256

Reichenbach, S. E., & Geng, F. (2003). Two-dimensional cubic convolution. *IEEE Transactions on Image Processing*, 12(8), 857-865. DOI: 10.1109/TIP.2003.814248

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7), 7183-7192. DOI: 10.1029/2000JD900719

Wan, Z., Wang, P., & Li, X. (2004). Using MODIS land surface temperature and normalized difference vegetation index products for monitoring drought in the southern Great Plains, USA. *International journal of remote sensing*, 25(1), 61-72. DOI: 10. 1080/0143116031000115328.

Wang, S., Luo, Y., Li, X., Yang, K., Liu, Q., Luo, X., & Li, X. (2021). Downscaling Land Surface Temperature Based on Non-Linear Geographically Weighted Regressive Model over Urban Areas. *Remote Sensing*, 13(8), 1580. DOI: 10.3390/rs13081580

Weng, Q. (2009). Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(4), 335-344. DOI: 10.1016/j.isprsjprs.2009.03.007

Wu, H., & Li, W. (2019). Downscaling land surface temperatures using a random forest regression model with multitype predictor variables. *IEEE Access*, 7, 21904-21916. DOI: 10.1109/ACCESS.2019.2896241

Xiao, R., Weng, Q., Ouyang, Z., Li, W., Schienke, E. W., & Zhang, Z. (2008). Land Surface Temperature Variation and Major Factors in Beijing, China. *Photogrammetric Engineering and Remote Sensing*, 74(4), 451-461. DOI: 10.14358/PERS.74.4.451

Xu, J., Zhang, F., Jiang, H., Hu, H., Zhong, K., Jing, W., ... & Jia, B. (2020). Downscaling ASTER land surface temperature over urban areas with machine learning-based area-to-point regression Kriging. *Remote Sensing*, 12(7), 1082. DOI: 10.3390/rs12071082

Yu, S., Chen, Z., Yu, B., Wang, L., Wu, B., Wu, J., & Zhao, F. (2020). Exploring the relationship between 2D/3D landscape pattern and land surface temperature based on explainable eXtreme Gradient Boosting tree: A case study of Shanghai, China. *Science of The Total Environment*, 138229. DOI: 10.1016/j.scitotenv.2020.138229

Zawadzka, J., Corstanje, R., Harris, J., & Truckell, I. (2020). Downscaling Landsat-8 land surface temperature maps in diverse urban landscapes using multivariate adaptive regression splines and very high resolution auxiliary data. *International Journal of Digital Earth*, 13(8), 899-914. DOI: 10.1080/17538947.2019.1593527

Zhang, W, Gong C, Hu Y, Song W, & Kuang D. (2019). A Research on spatial downscaling of thermal infrared image based on improved three-layer decomposition model. *J. Infrared Millim. Waves*, 38(2), 203-209. DOI: 10.11972/j.issn.1001-9014.2019.02.013

Zhou, J., Liu, S., Li, M., Zhan, W., Xu, Z., & Xu, T. (2016). Quantification of the scale effect in downscaling remotely sensed land surface temperature. *Remote*

Sensing, 8(12), 975. DOI: 10.3390/rs8120975.

List of Figure Captions

Fig. 1. Visual differences between LST at low (a) and high resolution (b).

Fig. 2. Beijing region (black square marks the experimental area). (a) Landsat 8 band composite (RGB-band 213); (b) map of land use and land cover; (c) spatial distribution of LST; (d) elevation a.s.l.

Fig. 3. Flow chart of the TLC method downscaling a low-resolution LST map.

Fig. 4. (a) map of the predictor after matching; (b) guided map (T_{cu}); (c) result of guided filtering with window size 11; (d) result of Gaussian low-pass filtering with cut-off frequency 3.

Fig. 5. Flow chart of traditional methods (DisTrad, TsHARP, and RF) for downscaling.

Fig. 6. (a) Downscaled LST maps and (b) Landsat 8 RLST.

Fig. 7. (a) Histograms of DLST compared to Landsat 8 RLST. (b) Error distribution between Landsat 8 RLST and downscaled LSTs. (c) Scatter density plots of DLST (y-axis) versus RLST (x-axis) downscaling from initial resolution 300 m to target resolution 30 m. From left to right: DisTrad, TsHARP, RF and TLC.

Fig. 8. Taylor diagram evaluating the accuracy of DLST using different downscaling methods. 'XEF' represents the reference data, the *CC* is related to the azimuthal angle (yellow lines), the *CRMSE* of the downscaled results is proportional to the distance to the reference point XEF (blue lines), and the *STD* is proportional to the radial distance to the origin (black lines).

Fig. 9. Taylor diagram evaluating the accuracy of DLST maps for different land cover types (marked by symbol shapes: for example, the blue rectangular box marks the accuracy of forest LST obtained from DisTrad).

Fig. 10. Scatter density plots of DLST (y-axis) versus RLST (x-axis) extracted from Landsat 8 30 m LST products (from left to right: target resolution from 30 m to 150 m; from top to bottom: DisTrad, TsHARP, RF, and TLC).

Fig. 11. Taylor diagram evaluating the accuracy of DLST maps for varied target resolutions.

Fig. 12. Scatter density plots of DLST (y-axis) versus RLST (x-axis) extracted from Landsat 8 30 m LST products (from left to right: varied spatial context from 15 × 15 km to 60 × 60 km; from top to bottom: DisTrad, TsHARP, RF, and TLC).

Fig. 13. Taylor diagram evaluating the accuracy of DLST maps over varied spatial context.

Fig. 14. Comparison of Landsat 8 RLST and downscaled LST maps for different land cover types (from left to right: Landsat 8, DisTrad, TsHARP, RF, and TLC; from top to bottom: water, agriculture land, forest, impervious surface land).