

Spatiotemporal Fusion of High Resolution Land Surface Temperature Using Thermal Sharpened Images from Regression-based Urban Indices

M. Kim¹, K. Cho¹, H. Kim¹, Y. Kim^{1*}

¹ Seoul National University, Dept. of Civil & Env. Engineering, South Korea - (mhh93, chris091, hanswfg, yik)@snu.ac.kr

Commission VI, WG VI/4

KEY WORDS: Spatiotemporal Fusion, Land Surface Temperature, Urban Index, Thermal Sharpening

ABSTRACT:

Obtaining spatially continuous, high resolution thermal images is crucial in order to effectively analyze heat-related phenomena in urban areas and the inherent high spatial and temporal variations. Spatiotemporal Fusion (STF) methods can be applied to enhance spatial and temporal resolutions simultaneously, but most STF approaches for the generation of Land Surface Temperature (LST) have not focused specifically on urban regions. This study therefore proposes a two-phase approach using Landsat 8 and MODIS images acquired on a study area in Beijing to first, investigate the sharpening of the fine resolution image input with urban-related spectral indices and second, to explore the potential of implementing the sharpened results into the Spatiotemporal Adaptive Data Fusion Algorithm for Temperature Mapping (SADFAT) to generate high spatiotemporal resolution LST images in urban areas. For this test, five urban indices were selected based on their correlation with brightness temperature. In the thermal sharpening phase, the Fractional Urban Cover (FUC) index was able to delineate spatial details in urban regions whilst maintaining its correlation with the original brightness temperature image. In the STF phase however, FUC sharpened results returned relatively high levels of correlation coefficient values up to 0.689, but suffered from the highest Root Mean Squared Error (RMSE) and Average Absolute Difference (AAD) values of 4.260 K and 2.928 K, respectively. In contrast, Normalized Difference Building Index (NDBI) sharpened results recorded the lowest RMSE and AAD values of 3.126 K and 2.325 K, but also the lowest CC values. However, STF results were effective in delineating fine spatial details, ultimately demonstrating the potential of using sharpened urban or built-up indices as a means to generate sharpened thermal images for urban areas, as well as for input images in the SADFAT algorithm. The results from this study can be used to further improve STF approaches for daily and spatially continuous mapping of LST in urban areas.

1. INTRODUCTION

With the rapid growth in data volumes, satellite image accessibility, and increasingly powerful computing efficiency, remote sensing applications are experiencing a paradigm shift towards monitoring land surface dynamics seamlessly over time using densely amassed time series data (Woodcock et al., 2019). In light of this trend, Land Surface Temperature (LST) is a crucial parameter to examine and interpret the processes occurring in the atmosphere and on the Earth's surface at regional and global scales. On a finer scale, LST can be used to study thermal variations across urban areas, such as urban climate conditions, heat wave and drought predictions, and urban heat island effects (Weng et al., 2014). LST can be derived by using thermal images from multiple satellite platforms, such as Moderate Resolution Imaging Spectroradiometer (MODIS), Advanced Spaceborne Thermal Emission and Reflectance (ASTER), Landsat 5, Landsat 7, and Landsat 8. On the one hand, coarse resolution sensors provide frequent data needed for monitoring purposes, but at a spatial resolution which is too low to resolve details of urban LST measurements. On the other hand, fine resolution sensors offer remotely-sensed images which are more suitable for detailed urban analysis, but are limited by long revisit times.

In general, optical satellites are restricted by a trade-off between spatial and temporal resolutions due to technical constraints. However, obtaining spatially continuous high resolution images is crucial to effectively analyze heat-related phenomena in urban areas and the inherent high spatial and temporal variations. Cloud

coverage and sparse data distribution due to missing information or data gaps also compound to this problem. Essentially, this conundrum can be narrowed to two possible solutions: enhancing the spatial resolution or increasing the temporal resolution. Conventional thermal sharpening methods utilize the relationship between LST and other land surface characteristics to downscale the image using statistical regression such as Normalized Difference Vegetation Index (NDVI) and Fractional Vegetation Cover (FVC). More recent studies incorporated machine learning techniques to generate more reliable results (Li et al., 2019). The drawback is that these downscaling methods only improve the image's spatial resolution without accounting for the temporal resolution. Ideally, combining sharpening methods in an image fusion framework can provide a potential solution to combine the resolution advantages of both coarse and fine resolution sensors.

In order to consider spatial and temporal resolutions simultaneously, spatiotemporal fusion (STF) methods can be applied to produce high spatial and temporal resolution products for daily, continuous mapping. Gao et al. (2006) developed the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) by blending a pair of Landsat and MODIS images to generate synthetic "daily" surface reflectance products. Spectrally similar neighbourhood information and the temporal variation between the input images are estimated to generate the synthetic image at a given prediction date. However, the model's performance degrades for highly heterogeneous, fine-grained landscapes (Gao et al., 2006). To overcome the landscape heterogeneity, Zhu et al. (2010) developed an enhanced version

* Corresponding author

of STARFM (ESTARFM) by inputting two pairs of Landsat and MODIS images, and introducing a conversion coefficient based on the ratio of change between the coarse image pixels and the fine image pixels. ESTARFM can interpret mixed pixels comprised of multiple classes, but cannot predict short-term changes unavailable in the input images and suffers in highly complex and fine-grained areas (Zhu et al., 2010).

STF methods were originally designed for reflectance bands, but has also been applied on TIR bands. Weng et al. (2014) proposed the Spatiotemporal Adaptive Data Fusion Algorithm for Temperature Mapping (SADFAT) which improved upon STARFM by incorporating the annual temperature cycle and modified the conversion coefficient from ESTARFM to interpret urban thermal landscape heterogeneity. SADFAT combined Landsat 7+ ETM and MODIS images, but the LST products were limited to a spatial resolution of 120 m, which is still too coarse for an effective analysis of LST in urban areas (Weng et al., 2014). To improve the output's spatial resolution, previous studies varied the input images or modified the original STF model. For instance, ASTER and MODIS images were used to evaluate three different STF models for LST generation, where ESTARFM was determined to produce better results over STARFM and a linear mixture based STF approach (Wu et al., 2015). Also, Landsat 8 and Visible Infrared Imaging Radiometer Suite (VIIRS) images were combined using SADFAT to generate daily LST images at a spatial resolution of 100 m (Jia et al., 2017). Another study integrated an extreme learning machine using a neural network regression model into SADFAT to produce LST at 30 m spatial resolution, but the computation was highly time-consuming (Bai et al., 2015). However, not many studies have explored the application of STF to generate high resolution LST in heterogeneous urban environments.

This study thus consists of two main objectives: first, to use sharpened regression-based urban indices for the enhancement of spatial features in urban areas and second, to investigate the feasibility of using the sharpened images into the SADFAT algorithm to generate LST images at 30 m spatial resolution for continuous, high resolution mapping and for finer-scale urban analysis. In more detail, urban indices were experimented for thermal sharpening in place of the traditional NDVI input, in order to represent the urban areas more accurately. The correlation of each index was compared with the original Landsat 8 thermal image to validate its relationship with LST. The generated LST results were tested using five images acquired from May to September of 2019 to minimize seasonal variation. The first two images were processed to determine the conversion coefficient while the remaining three images were used to test the influence of temporal variation on the STF algorithm. The following content of this paper is structured in the following order: Section 2 introduces key theoretical background information on thermal sharpening and STF, and also displays the study area and satellite image datasets used this research. Section 3 shows an overview of the study and describes the processing steps in more detail. Section 4 presents the generated LST images and quality assessment results with associated discussion points. Lastly, Section 5 provides a conclusion, summarizing the study as well as outlining the limitations and potential of the study.

2. STUDY AREA AND DATASETS

2.1 Study Area

The study area was selected to be in Beijing between $116^{\circ} 17' 26''$ E to $116^{\circ} 38' 33''$ E and $39^{\circ} 50' 34''$ N to $40^{\circ} 6' 50''$ N,

covering an area of approximately 30 km^2 . The location of the area is shown in Figure 1. The majority of the study area is urban land cover, but also includes other heterogeneous land cover types such as vegetation, water, bare soil, and impervious surfaces. Notable features in the study area include the Beijing Capital airport in the northeast and the Forbidden City palace complex near the center. Beijing has a humid, continental climate with a hot summer season influenced by monsoon seasons, followed by a dry autumn with minimal precipitation. According to the National Oceanic and Atmospheric Administration, the monthly average temperature in Beijing reaches an average maximum temperature of 31°C and an average maximum precipitation of 177.9 mm.



Figure 1. Location of the study area

2.2 Satellite Image Datasets

Landsat 8 and MODIS images were used as the fine and coarse resolution images, respectively. Landsat and MODIS data are ideal for STF because of their similar orbit parameters and sensor specifications. For this study, Landsat 8 multispectral and thermal images were obtained with minimal cloud coverage. For MODIS images, MOD09GA surface reflectance and MOD11A1 LST products were acquired. A total of five pairs of Landsat and MODIS images were used. The minimum, maximum and median temperatures of the Landsat and MODIS datasets are shown in Figure 2. VIIRS Nighttime Light (NTL) images were required for the computation of urban indices, discussed in Sub-section 3.1. VIIRS NTL images were acquired by the Day-Night band at a spatial resolution of 750m. The monthly composite NTL images were selected to avoid cloud coverage and stray light effects. In comparison to daily NTL images, the monthly composite images were considered suitable for this study based on the assumption that there would be minimal NTL variations in urban environments within the span of a month.

2.3 Image Pre-processing

Landsat 8 multispectral images require radiometric calibration and atmospheric correction using the FLAASH module in ENVI. Landsat 8 TIR images can be radiometrically calibrated to produce at-sensor radiance and brightness temperature images. Both images are reprojected to UTM zone 50N and are subset to the study area's extents. Likewise, MODIS surface reflectance and LST images are reprojected using the MODIS reprojection toolkit, and are then resampled to 30 m to match Landsat 8's spatial resolution. Lastly, VIIRS NTL images are normalized to compensate for atmospheric effects and reprojected.

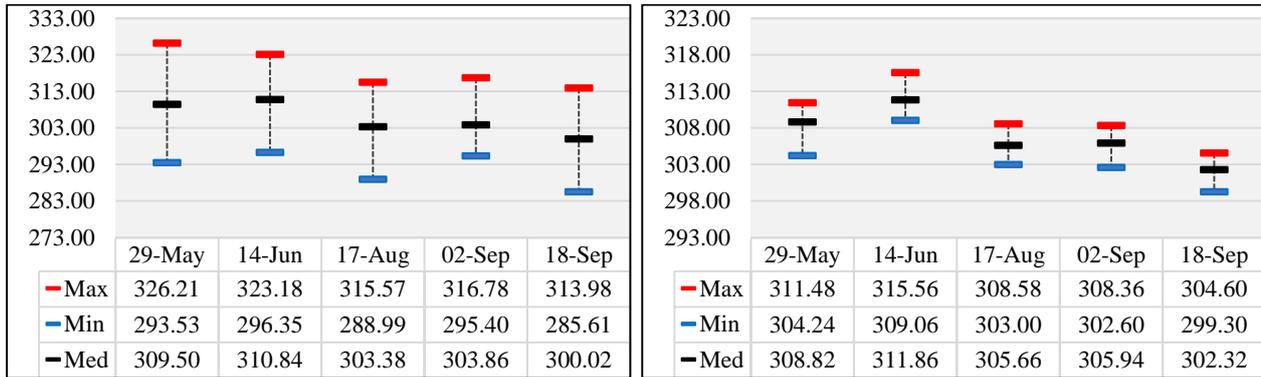


Figure 2. Maximum, minimum and median temperatures for Landsat 8 (left) and MODIS (right) LST in Kelvin (K)

3. METHODOLOGY

This study proposes a “regression-then-fusion” approach to STF as coined by Xia et al. (2018). As presented in Figure 3, the study can be divided into two phases: (1) Data preparation of input images and (2) STF. Sub-sections 3.1 and 3.2 explain the first phase while Sub-section 3.2 describes the second phase in detail.

3.1 Urban Spectral Indices

For this study, urban-related indices were selected based on their correlation with the corresponding Landsat 8 thermal image. First, the Normalized Difference Built-up Index (NDBI) has been extensively used to classify built-up, urban areas (Zha et al., 2013). NDBI is designed to capture urban regions since built-up features tend to have a higher reflectance in the shortwave-infrared (SWIR) region, as shown by the following equation:

$$NDBI = \frac{b_{SWIR} - b_{NIR}}{b_{SWIR} + b_{NIR}} \quad (1)$$

Second, Jieli et al. (2010) developed the New Building Index (NBI) to emphasize the spectral response of barren land from other land covers.

$$NBI = \frac{b_{Red} \cdot b_{NIR}}{b_{SWIR1}} \quad (2)$$

Third, Liu et al. (2014) proposed the Modified Built-up Index (MBI) to accentuate built-up areas for a study on investigating spatiotemporal dynamics of urban sprawl.

$$MBI = \frac{b_{SWIR2} \cdot b_{Red} - b_{NIR}^2}{b_{SWIR2} + b_{NIR} + b_{Red}} \quad (3)$$

Fourth, Lu et al. (2003) proposed the Human Settlement Index (HSI) to extract urban areas by integrating NTL and NDVI data. The maximum value of the calculated NDVI and the normalized NTL data were used to calculate HSI.

$$HSI = \frac{(1 - NDVI_{Max} + NTL_{Norm})}{(1 - NTL_{Norm}) + NDVI_{Max} + (NDVI_{Max} \cdot NTL_{Norm})} \quad (4)$$

Last, Cho (2019) proposed a novel index designed for urban areas called Fractional Urban Cover (FUC), which was inspired by the structure of fractional vegetation cover (FVC) and HSI. In comparison to other thermal sharpening methods, FUC demonstrated superior results for built-up areas, but recorded lower performance for water bodies (Cho, 2019). FUC can be expressed using minimum and maximum values of HSI:

$$FUC = \left(\frac{HSI_{Max} - HSI}{HSI_{Max} - HSI_{Min}} \right)^n \quad (5)$$

Here, the exponent n is determined based on the maximum Correlation Coefficient (CC) derived from linear regression between FUC and brightness temperature.

3.2 Thermal Sharpening Methods

Disaggregating T_r to the NDVI pixel resolution (DisTrad) (Kustas et al., 2003) and Thermal Sharpening (TsHARP) (Agam et al., 2007) are two of the most widely adopted methods in thermal sharpening. The TsHARP method conducts a polynomial regression using FVC to produce a predicted thermal image as shown in Equation 6. FVC is computed based on Equation 7 and uses minimum and maximum NDVI values. A residual term, ΔT_{LR} , is computed in Equation 8 by using the difference between the predicted thermal image and the original low-resolution thermal image $T_{LR,original}$. Lastly, linear least squares regression is applied using the high-resolution FVC_{HR} with the addition of the residual term from the previous equation to estimate the sharpened image \widehat{T}_{HR} in Equation 9.

$$\widehat{T}_{LR} = F(FVC_{LR}) = a_0 + a_1 FVC_{LR} \quad (6)$$

$$FVC_{LR} = \left(\frac{NDVI_{Max} - NDVI_{LR}}{NDVI_{Max} - NDVI_{Min}} \right)^{0.625} \quad (7)$$

$$\Delta T_{LR} = T_{LR,original} - \widehat{T}_{LR} \quad (8)$$

$$\widehat{T}_{HR} = F(FVC_{HR}) + \Delta \widehat{T}_{LR} \quad (9)$$

However, since NDVI or FVC may be inadequate to interpret fine-grained, heterogeneous urban areas, the urban indices introduced in the previous sub-section can be integrated into the TsHARP model to produce sharpened thermal images in a similar manner. For clarity, the urban index input images can be implemented using the following expression:

$$\widehat{T}_{LR} = F(Urban\ Index_{LR}) = a_0 + a_1 Urban\ Index_{LR} \quad (10)$$

where the urban indices are first aggregated to match the thermal image’s resolution prior to calculating the regression.

3.3 Implementing the STF Model (SADFAT)

The SADFAT model proposed by Weng et al. (2014) was adopted to generate high resolution LST images. To differentiate from past studies, the principal contribution of this study lies in the investigation of urban index-based thermal images as input datasets, as shown in Figure 3 by the “Index-based LST (30m) input” highlighted in red text.

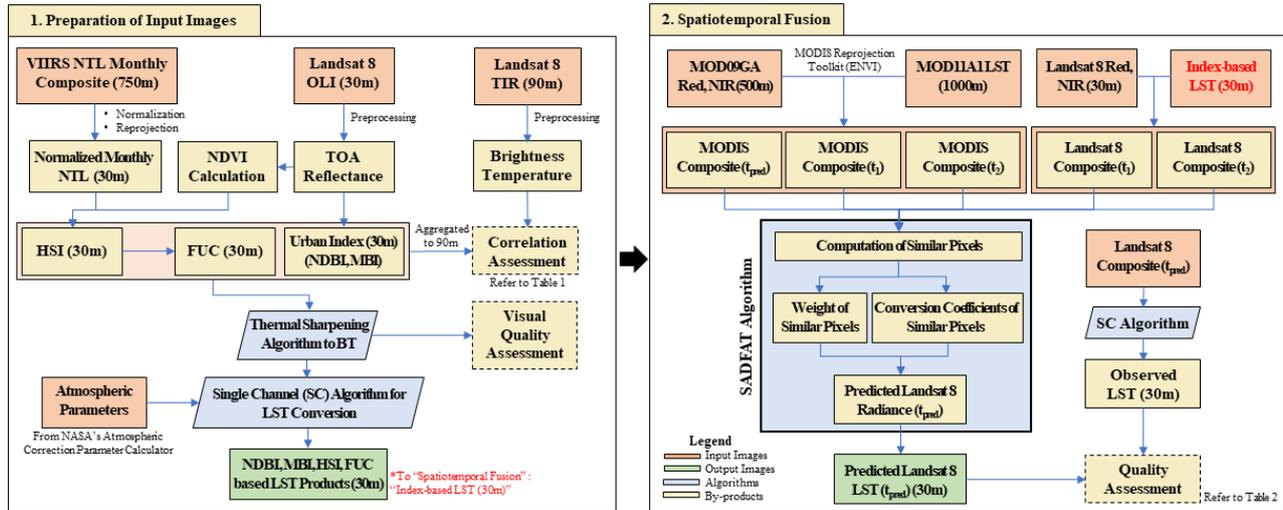


Figure 3. Overview of the two-phase methodology showing the input images and processes involved in this study

Linear spectral mixture analysis was used since the majority of urban areas are filled with mixed pixels. Hence, each of the finer resolution Landsat pixels can be considered as a pure endmember within the resampled MODIS pixel. For this study, Landsat and MODIS images are assigned as the fine resolution and coarse resolution image inputs in the SADFAT algorithm, respectively. Using a pair of Landsat 8 and MODIS radiance images at t_0 and another MODIS image at t_p , the corresponding Landsat image can be generated at t_p using the following formula:

$$R_L(x, y, t_p) = R_L(x, y, t_0) + h(x, y) \times [R_M(x, y, t_p) - R_M(x, y, t_0)] \quad (11)$$

where R_L is the Landsat radiance derived from TIR Band 10, R_M is the MODIS radiance, W_i is the weight of a neighboring similar pixel which considers the spectral similarity of the i^{th} pixel between the Landsat and MODIS images as well as the spatial distance between the selected similar pixel and the center pixel. In more detail, the combined weight term CW_i can be computed by combining the Euclidean distance of the neighboring similar pixel and the central pixel, d_i . The weight of the pixel is calculated by finding the normalized result of the inverse of the combined weight using the expression:

$$CW_i = (1 - R_i) \times d_i \quad (12)$$

$$W_i = \frac{\left(\frac{1}{CW_i}\right)}{\sum_{i=1}^N \left(\frac{1}{CW_i}\right)} \quad (13)$$

Also, h_k is the conversion coefficient originally proposed in ESTARFM which computes the ratio of the difference between the radiance value of Landsat and MODIS at the inputted times, t_1 and t_2 , and can be explained by the following expression:

$$h_k = \frac{R_{kL}(t_2) - R_{kL}(t_1)}{R_{kM}(t_2) - R_{kM}(t_1)} \quad (14)$$

where R_{kL} denotes the radiance pixel of the k^{th} Landsat pixel at a given time, while R_{kM} refers to the corresponding radiance pixel of the k^{th} MODIS pixel at a given time. A moving window can then be used to incorporate additional information from neighboring spectrally similar pixels, and to determine a unique solution of Equation 11 to find the value of the center pixel. Assuming the moving window has a size of w with a center pixel of $x_{w/2}, y_{w/2}$, the Landsat radiance can be predicted at t_p as:

$$R_L(x_{w/2}, y_{w/2}, t_p) = R_L(x_{w/2}, y_{w/2}, t_0) + \sum_{k=1}^N W_i \times h_k \times [R_M(x_{w/2}, y_{w/2}, t_p) - R_M(x_{w/2}, y_{w/2}, t_0)] \quad (15)$$

3.4 Generalized Single Channel (SC) Algorithm

LST can be retrieved from radiometrically calibrated thermal images using single channel, split-window, and mono-window methods (Lu et al., 2004). The Generalized SC algorithm was used since in-situ radio-soundings or effective mean atmospheric temperature values were not required (Jiménez-Muñoz and Sobrino, 2003). The predicted Landsat radiance image at t_p can be converted to LST and corrected for atmospheric effects using the Generalized SC algorithm outlined in Sobrino et al. (2004):

$$LST = \gamma \left[\frac{1}{\varepsilon} (\psi_1 L_s + \psi_2) + \psi_3 \right] + \delta \quad (16)$$

where LST is the calculated Landsat 8 thermal image (in Kelvin), L_s is the at-sensor radiance, ε is the surface emissivity, ψ_1, ψ_2, ψ_3 are atmospheric functions, and γ and δ are parameters dependent on Planck's function which are calculated by:

$$\gamma = \left(\frac{c_2 L_s}{T_s^2} \left[\frac{\lambda^4}{c_1} L_s + 1 \right] \right)^{-1} \quad (17)$$

$$\delta = -\gamma L_s + T_s \quad (18)$$

where λ is the effective wavelength of Landsat 8 TIR band 10 (μm), T_s is the sharpened index radiance image, and c_1 and c_2 are Planck's radiation constants which correspond to 774.8853 and 1321.0789, respectively. Land surface emissivity can vary based on the feature's characteristics, and is denoted as ε_λ . Emissivity products from satellite images are available, but are either outdated (ASTER) or have a very coarse resolution (MODIS). Alternatively, emissivity can be determined conditionally based on NDVI proportions, calculated separately for vegetation, soil, water, and mixed areas using:

$$\varepsilon_\lambda = \varepsilon_{veg, \lambda} P_{veg} + \varepsilon_{soil, \lambda} (1 - P_{veg}) + C_\lambda \quad (19)$$

$$P_{veg} = \left(\frac{NDVI - NDVI_{soil}}{NDVI_{veg} - NDVI_{soil}} \right)^2 \quad (20)$$

where $\varepsilon_{veg,\lambda}$ and $\varepsilon_{soil,\lambda}$ correspond to vegetation and soil emissivities, respectively, C_λ is the surface roughness, P_{veg} is the proportion of vegetation calculated by using NDVI values for vegetation, $NDVI_{veg}$, and for soil, $NDVI_{soil}$. For this study, a surface roughness was used as a constant value of 0.005 (where $C_\lambda = 0$ for homogeneous, flat surfaces), while $NDVI_{veg}$ and $NDVI_{soil}$ were set to 0.5 and 0.2, respectively, based on an empirical validation of the NDVI results for each date. Negative NDVI values were deemed as water pixels and were assigned an emissivity value of 0.991, while other emissivity terms were calculated based on Equation 19 and 20 using the aforementioned four land cover classes from Sub-section 2.1 (vegetation, water, bare soil, and mixed area). The emissivity of mixed pixels, $\varepsilon_{\lambda,mix}$, can be calculated using the following conditioned formula:

$$\varepsilon_{\lambda,mix} = \begin{cases} [1] \varepsilon_{soil,\lambda} \\ [2] \varepsilon_{veg,\lambda} P_{veg} + \varepsilon_{soil,\lambda} (1 - P_{veg}) + C_\lambda \\ [3] \varepsilon_{soil,\lambda} + C_\lambda \end{cases} \quad (21)$$

$[1] NDVI < NDVI_{soil}$
 $[2] NDVI_{soil} \leq NDVI \leq NDVI_{veg}$
 $[3] NDVI > NDVI_{veg}$

Condition [1] is represented by bare soil and was given an emissivity value of 0.996, condition [2] contains both vegetation and soil pixels and was assigned an emissivity value based on Equation 21, and condition [3] is represented by vegetation and was allocated an emissivity of 0.973 (Avdan and Jovanovska, 2016). Lastly, atmospheric functions are required to eliminate the atmospheric effects on the LST product. This study used the Atmospheric Correction Parameter Calculator (ACPC) offered by NASA to estimate transmissivity (τ), upwelling radiance (L^\uparrow), and downwelling radiance (L^\downarrow). These variables are related to the atmospheric functions as shown in Equation 22.

$$\psi_1 = \frac{1}{\tau}, \psi_2 = -L^\downarrow - \frac{L^\uparrow}{\tau}, \psi_3 = L^\downarrow \quad (22)$$

4. RESULTS AND DISCUSSION

4.1 Calculation of Urban Indices

While traditional thermal sharpening methods utilized NDVI or variants of the index, urban indices were hypothesized to be more prevalent in urban areas. In light of this shortcoming, five urban indices were selected on the basis of showing a reasonable correlation with the thermal image. The urban indices were processed using Landsat 8 bands and aggregated to 90 m for fair comparison with the Landsat 8 thermal image. The results of the correlation assessment are provided in Table 1. NBI was removed from the results in Table 1 and the STF phase due to poor performance to narrow the number of input images to four urban indices. MBI showed relatively higher results over NDBI, thereby indicating the effectiveness of combining both the difference between NIR and SWIR in conjunction with the red band. HSI demonstrated the lowest Root Mean Squared Error (RMSE) values while FUC displayed the highest CC values. In comparison to HSI, the relatively higher RMSE values in the FUC results can be attributed to the index's low performance in delineating water body parameters.

4.2 Thermal Sharpened BT Results

The 30 m urban indices were used to sharpen the original 90 m Landsat 8 brightness temperature images, as shown in Figures 4 and 5 using a range of 293 K to 328 K. All of the sharpened results enhanced the urban features in more detail, but varied in terms of sharpness and the noisy pixels in the image. On closer

Index	Metric	Image Acquisition Date				
		5/29	6/14	8/17	9/02	9/18
NDBI	CC	0.578	0.620	0.607	0.631	0.614
	RMSE	0.184	0.237	0.220	0.195	0.199
MBI	CC	0.566	0.636	0.570	0.608	0.621
	RMSE	0.179	0.227	0.198	0.167	0.186
HSI	CC	0.601	0.674	0.609	0.698	0.702
	RMSE	0.157	0.193	0.160	0.141	0.159
FUC	CC	0.683	0.745	0.715	0.766	0.752
	RMSE	0.171	0.209	0.195	0.163	0.180

Table 1. Correlation assessment prior to thermal sharpening

inspection, the NDBI and HSI results reflected the thermal distribution in the palace complex and in the nearby urban regions as shown in Figure 4, but were degraded by noisy pixels due to high heterogeneity. The MBI, TsHARP and FUC results displayed smoother variations, particularly evident near water bodies. All of the indices apart from FUC were able to distinguish the perimeter of water bodies effectively. Although the FUC result exhibited rougher extents, the overall result was smoother and more representative of the original image. Moreover, the spatial details in the Beijing Capital International Airport in Figure 5 were more pronounced in the NDBI, MBI and HSI results, but were nonetheless affected by noisy pixels. In contrast, the smoother TsHARP and FUC images presented similar thermal variations with respect to the original thermal image.

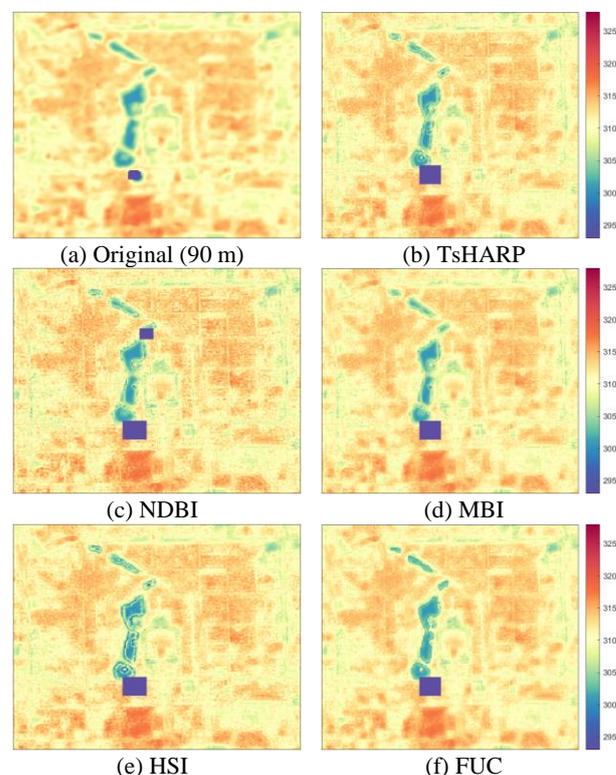


Figure 4. Sharpened images of the Forbidden City palace complex and nearby urban area from 06/14/2019 images using the urban indices.

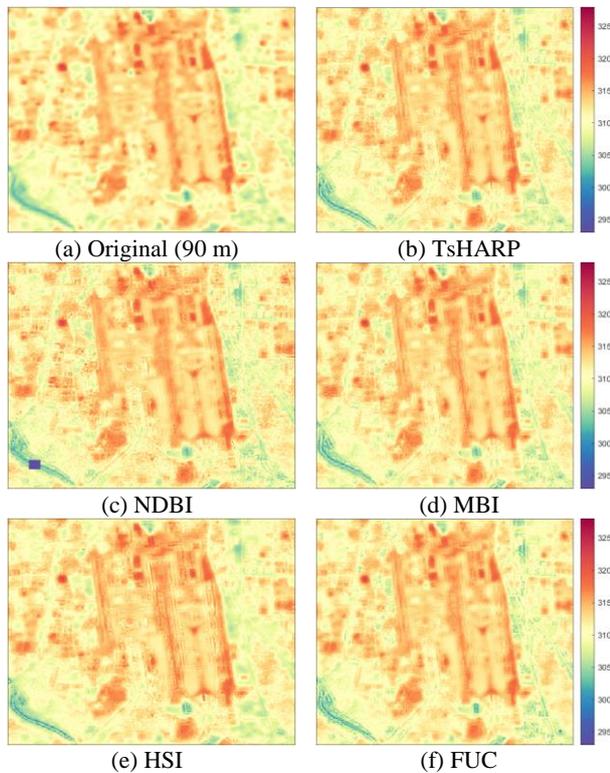


Figure 5. Sharpened images of the Beijing Capital International airport from 06/14/2019 images using urban indices.

4.3 Results from the STF Algorithm

Two pairs of Landsat and MODIS images from 05/29/19 and 06/14/19 were used to compute the conversion coefficients, while the remaining acquisition dates of 08/17/19, 09/02/19, and 09/18/19 were used as validation sets. The generated LST images using the original Landsat 8 thermal band are displayed in Figures 6 to 8 on a scale of 293 K to 328 K.

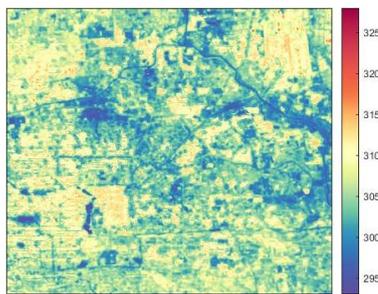


Figure 6. LST Prediction for 08/17/19

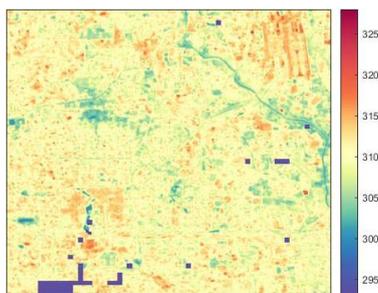


Figure 7. LST Prediction for 09/02/19

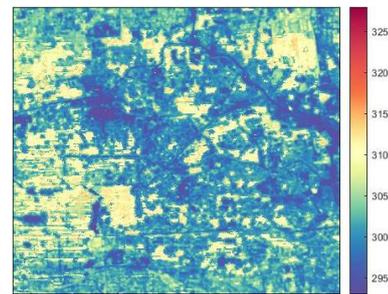


Figure 8. LST Prediction for 09/18/19

4.4 Quality Assessment Results

The generated LST images at t_{pred} were compared with the observed LST on the corresponding acquisition date. RMSE, Absolute Average Difference (AAD), and CC were computed to evaluate the algorithm's prediction accuracy as displayed in Table 2. y_i is the observed LST value and \hat{y}_i is the predicted LST value using the STF algorithm, while μ_{y_i} and $\mu_{\hat{y}_i}$ correspond to the mean values of y_i and \hat{y}_i . These quality assessment metrics were selected based on their common usage in literature.

Metric	Ideal	Formula
RMSE	0	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - \mu_{\hat{y}_i})^2}$
AAD	0	$AAD = \frac{1}{N} \sum_{i=1}^N \hat{y}_i - \mu_{\hat{y}_i} $
CC	1	$CC = \frac{\sum_{i=1}^N (y_i - \mu_{y_i})(\hat{y}_i - \mu_{\hat{y}_i})}{\sqrt{\sum_{i=1}^N (y_i - \mu_{y_i})^2} \sqrt{\sum_{i=1}^N (\hat{y}_i - \mu_{\hat{y}_i})^2}}$

Table 2. Quality assessment results for each index

The spatiotemporally sharpened results are presented in Table 3, and the best results are highlighted in bold text. TsHARP, MBI, and FUC sharpened results scored relatively higher CC values, but MBI and FUC suffered from higher RMSE and AAD values. CC values for TsHARP, NDBI, and FUC are included in Figure 9. Overall, the sharpened inputs were able to distinguish the LST patterns on a finer spatial resolution, especially for temporally closer prediction times. LST variations around larger built-up features such as buildings and airport facilities were more distinguishable, but the central Beijing area and surrounding urban regions were more difficult to predict due to the high landscape heterogeneity and complexity.

08/17/19	TsHARP	NDBI	HSI	MBI	FUC
RMSE (K)	3.232	3.126	3.252	3.367	3.462
AAD (K)	2.404	2.325	2.401	2.515	2.582
CC	0.669	0.580	0.662	0.668	0.671
09/02/19	TsHARP	NDBI	HSI	MBI	FUC
RMSE (K)	3.306	3.223	3.317	3.417	3.487
AAD (K)	2.560	2.501	2.547	2.653	2.696
CC	0.690	0.615	0.684	0.687	0.689
09/18/19	TsHARP	NDBI	HSI	MBI	FUC
RMSE (K)	3.807	3.510	3.849	4.062	4.260
AAD (K)	2.610	2.435	2.613	2.794	2.928
CC	0.658	0.576	0.646	0.651	0.650

Table 3. Quality assessment results for each index

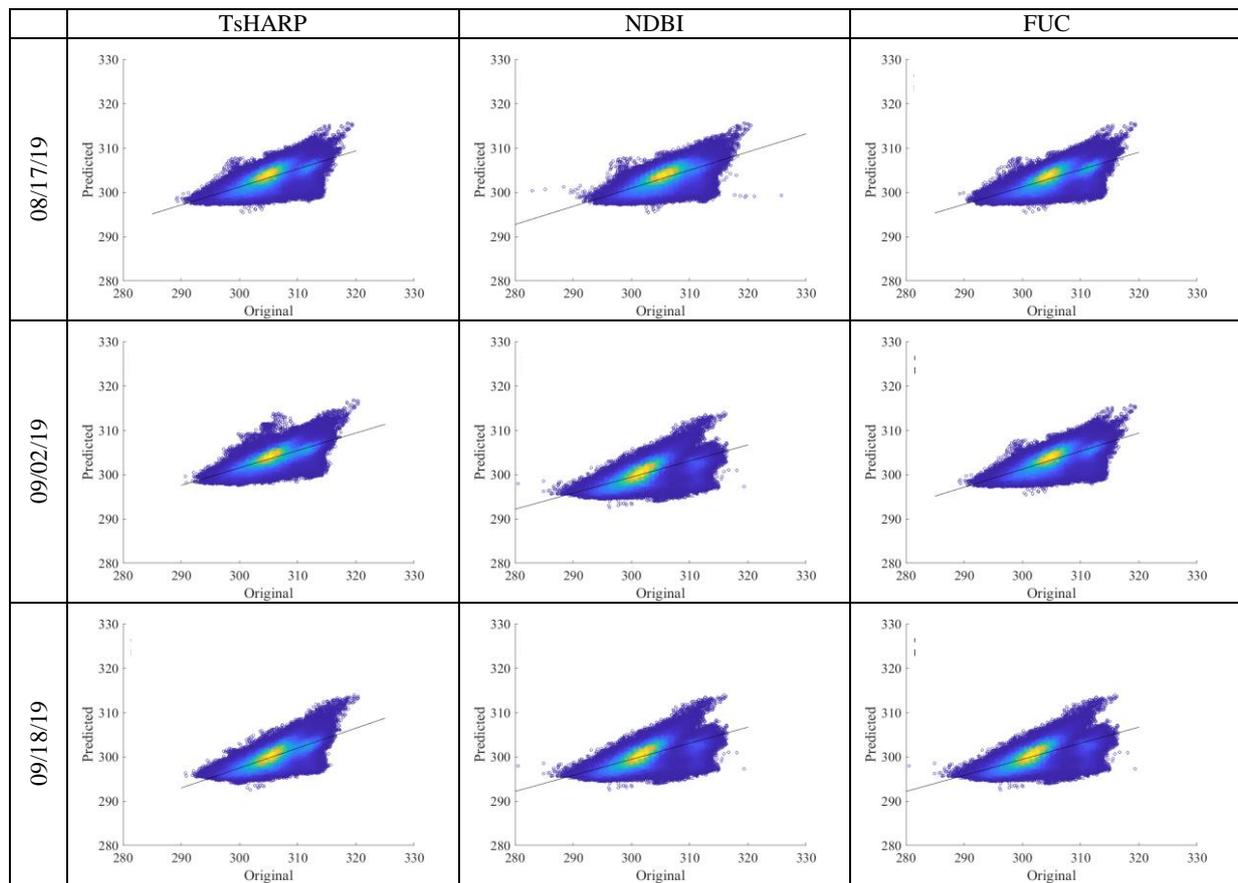


Figure 9. Visual representation of CC results found in Table 3

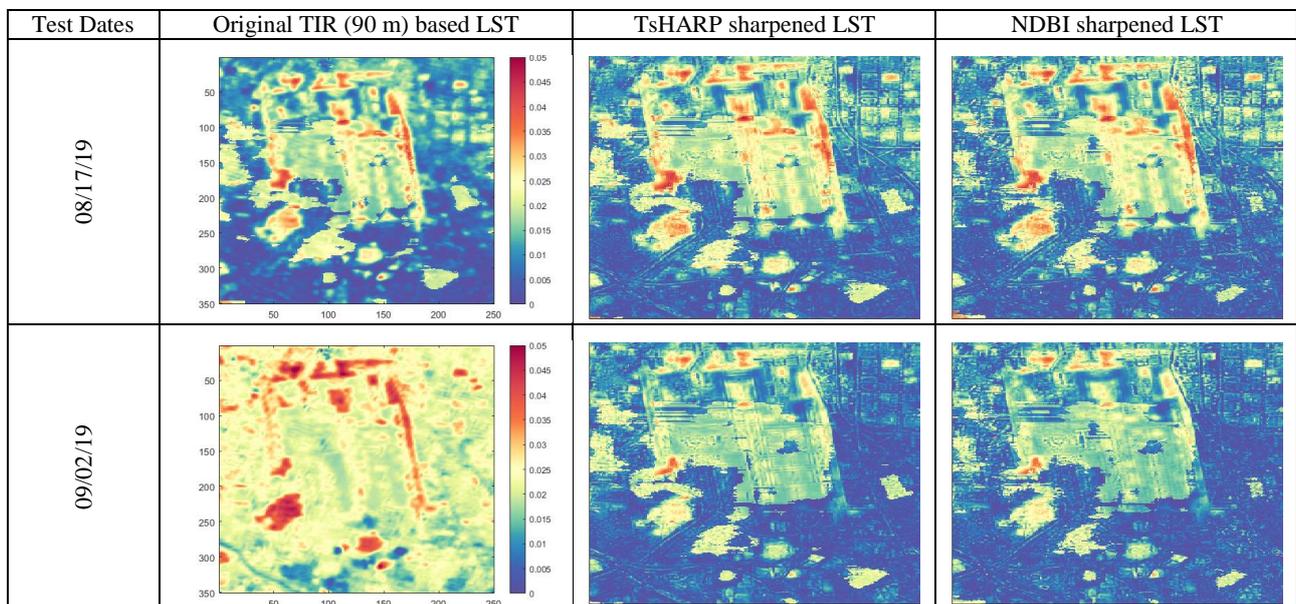


Figure 10. Average difference percentage error of the STF sharpened images using TsHARP, NDBI, and FUC of the Beijing Capital International airport shown in Figure 5. The range of all of the images is 0 and 5%.

Furthermore, the increase in RMSE and AAD over time can be correlated with the model's lack of sensitivity to larger temporal variations. The NDBI sharpened result recorded the lowest RMSE and AAD values for all of the scenarios, but also returned the lowest CC values, indicating that the prediction may have been more accurate, but was unstable. In order to investigate the

prediction errors in more depth, Figure 10 displays the average difference percentage error for each site using TsHARP and NDBI. FUC was omitted since the error map was similar to TsHARP, given that both results were NDVI-based. The average difference percentage error is calculated by dividing the average distance between pixels of the original and predicted LST images

by the original LST image, and is measured on a scale of 0 to 1, with 0 being ideal. The results showed higher levels of error at higher thermal variations, especially near building facilities. Comparing TsHARP and NDBI results, the latter result demonstrated smaller proportions (regions) of error in comparison to the TsHARP result for both test dates, indicating the relatively lower AAD and RMSE values in Table 3. On the contrary, the TsHARP results tended to smoothen the fine spatial details. Also, FUC was able to preserve the details to a reasonable extent, the MBI results tended to pixelate predictions, while the HSI results were found to overestimate predictions. Other sources of error include the uncertainty behind emissivity values. NDVI may not suffice to classify the various land cover classes in urban areas, especially considering the heterogeneity of a megacity like Beijing. In addition, the presence of water bodies tended to deteriorate the sharpening methods (such as for FUC) as well as for the SADFAT algorithm. The SADFAT parameters were also difficult to optimize, given the high computation time.

5. CONCLUSION

STF of LST images provides a method to generate spatially continuous, high resolution thermal imagery. This study demonstrated the feasibility of applying urban indices to sharpen Landsat 8 thermal images in urban areas and of implementing these sharpened images in the SADFAT algorithm to generate high resolution LST images. Although the NDBI sharpened results were found to be unstable in highly heterogeneous urban regions, the low RMSE and AAD statistical parameters signified the potential to modify the index by optimizing its correlation with thermal images. Future works stemming from this study include enhancing the resolution of the coarse image input for SADFAT or other STF methods, utilizing classification methods to delineate land cover classes more accurately, and integrating dictionary-based or deep learning methods in the STF pipeline.

ACKNOWLEDGEMENTS

The authors are grateful to Fu, P. for the open-source SADFAT algorithm. Cho, K. provided the FUC index and thermal heat maps and Kim, H. provided helpful guidance to optimize the algorithm. Also, the authors thank the two anonymous reviewers who helped to improve this research.

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.
- Avdan, U., Jovanovska, G., 2016: Algorithm for automated mapping of land surface temperature using LANDSAT 8 satellite data. *Journal of Sensors*, 2016.
- Bai, Y., Wong, M. S., Shi, W. Z., Wu, L. X., Qin, K., 2015: Advancing of land surface temperature retrieval using extreme learning machine and spatio-temporal adaptive data fusion algorithm. *Remote Sensing*, 7(4), 4424-4441.
- Cho, K., 2019: Modulation and Regression based Hybrid Thermal Sharpening of Landsat-8 TIRS Imagery Using Fractional Urban Cover (*Master's Thesis*, Seoul National University).
- 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.
- Jia, Y., Huang, Y., Yu, B., Wu, Q., Yu, S., Wu, J., Wu, J., 2017: Downscaling land surface temperature data by fusing Suomi NPP-VIIRS and landsat-8 TIR data. *Remote Sensing Letters*, 8(12), 1132-1141.
- Jieli, C., Manchun, L. I., Yongxue, L. I. U., Chenglei, S., Wei, H. U., 2010: Extract residential areas automatically by new built-up index. In *2010 18th International Conference on Geoinformatics* (pp. 1-5). IEEE.
- Jiménez-Muñoz, J. C., Sobrino, J. A., 2003: A generalized single-channel method for retrieving land surface temperature from remote sensing data. *Journal of Geophysical Research: Atmospheres*, 108(D22).
- 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.
- Li, W., Ni, L., Li, Z. L., Duan, S. B., Wu, H., 2019: Evaluation of machine learning algorithms in spatial downscaling of MODIS land surface temperature. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7), 2299-2307.
- Liu, Y., Chen, J., Cheng, W., Sun, C., Zhao, S., Pu, Y., 2014: Spatiotemporal dynamics of the urban sprawl in a typical urban agglomeration: a case study on Southern Jiangsu, China (1983–2007). *Frontiers of Earth Science*, 8(4), 490-504.
- Lu, D., Tian, H., Zhou, G., Ge, H., 2008: Regional mapping of human settlements in southeastern China with multisensor remotely sensed data. *Remote Sensing of Environment*, 112(9), 3668-3679.
- Sobrino, J. A., Jiménez-Muñoz, J. C., Paolini, L., 2004: Land surface temperature retrieval from LANDSAT TM 5. *Remote Sensing of Environment*, 90(4), 434-440.
- Weng, Q., Fu, P., Gao, F., 2014: Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data. *Remote Sensing of Environment*, 145, 55-67.
- Woodcock, C. E., Loveland, T. R., Herold, M., Bauer, M. E. 2019: Transitioning from change detection to monitoring with remote sensing: A paradigm shift. *Remote Sensing of Environment*, 111558.
- Wu, M., Li, H., Huang, W., Niu, Z., & Wang, C., 2015: Generating daily high spatial land surface temperatures by combining ASTER and MODIS land surface temperature products for environmental process monitoring. *Environmental Science: Processes & Impacts*, 17(8), 1396-1404.
- Xia, H., Chen, Y., Zhao, Y., & Chen, Z., 2018: "Regression-then-Fusion" or "Fusion-then-Regression"? A Theoretical Analysis for Generating High Spatiotemporal Resolution Land Surface Temperatures. *Remote Sensing*, 10(9), 1382.
- Zhu, X., Chen, J., Gao, F., Chen, X., Masek, J. G., 2010: An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sensing of Environment*, 114(11), 2610-2623.