remote sensing

MDPI

Article

A Scale-Separating Framework for Fusing Satellite Land Surface Temperature Products

Yichen Yang * and Xuhui Lee

Yale School of the Environment, Yale University, 195 Prospect Street, New Haven, CT 06511, USA; xuhui.lee@yale.edu

* Correspondence: yichen.yang@yale.edu; Tel.: +1-203-392-4007

Abstract: The trade-off between spatial and temporal resolutions of satellite imagery is a long-standing problem in satellite remote sensing applications. The lack of daily land surface temperature (LST) data with fine spatial resolution has hampered the understanding of surface climatic phenomena, such as the urban heat island (UHI). Here, we developed a fusion framework, characterized by a scale-separating process, to generate LST data with high spatiotemporal resolution. The scale-separating framework breaks the fusion task into three steps to address errors at multiple spatial scales, with a specific focus on intra-scene variations of LST. The framework was experimented with MODIS and Landsat LST data. It first removed inter-sensor biases, which depend on season and on land use type (urban versus rural), and then produced a Landsat-like sharpened LST map for days when MOIDS observations are available. The sharpened images achieved a high accuracy, with a RMSE of 0.91 K for a challenging heterogeneous landscape (urban area). A comparison between the sharpened LST and the air temperature measured with bicycle-mounted mobile sensors revealed the roles of impervious surface fraction and wind speed in controlling the surface-to-air temperature gradient in an urban landscape.

Keywords: spatiotemporal data fusion; land surface temperature; scale; deep-learning; urban heat island; New Haven

check for updates

Citation: Yang, Y.; Lee, X. A Scale-Separating Framework for Fusing Satellite Land Surface Temperature Products. *Remote Sens.* **2022**, *14*, 983. https://doi.org/10.3390/rs14040983

Academic Editors: Wenliang Li and Changshan Wu

Received: 28 January 2022
Accepted: 15 February 2022
Published: 17 February 2022

1. Introduction

Land surface temperature (LST) is a key driver of surface-air energy exchanges [1]. Satellite LST data are used in a large array of studies, ranging from climate change [2–5], to agriculture [6,7], forestry [8,9], hydrology [10,11], and ecology [12,13]. Being continuous in space and repetitive in time, satellite-based LST data are especially useful for studies of the urban heat island (UHI) [14–16].However, satellite thermal imagery always involves a tradeoff between spatial resolution and temporal coverage [17]. For example, Landsat satellites acquire thermal imageries in fine spatial resolution (better than 120 m) every 16 days. On the other hand, MODIS LST data are provided daily but the spatial resolution is coarser (1 km). This trade-off is a technological bottleneck [18] that limits the utility of satellite LST data in characterizing intra-city variations of the UHI intensity. One solution to overcome the bottleneck is to develop scaling methods and produce LST data at a finer spatial resolution (e.g., the resolution of Landsat) from daily images acquired at a coarser spatial resolution (e.g., MODIS LST data products).

A large body of literature has been published in the past decades on methods to downscale satellite thermal imagery. These methods can be broadly characterized as Disaggregation of Remotely Sensed Land Surface Temperature (DLST) [19]. A common DLST method, termed Thermal Sharpening (TSP), incorporates kernel-based algorithms for downscaling [19,20]. Loosely speaking, the kernel relates the LST derived from the thermal band with the physical properties derived from other bands of the same coarse resolution image. For example, DisTrad [21], or disaggregation procedure for radiometric surface temperature, uses the Normalized Difference Vegetation Index (NDVI) as the kernel predictor. The relationship between NDVI and radiometric temperature in the coarse-resolution images is determined using least-square fitting. The acquired statistical relationship is then applied to the NDVI from the images with a finer spatial resolution to produce sharpened thermal images. The key assumption of DisTrad is that the physical impact of NDVI on LST is preserved over scales. Other kernel-based methods use enhanced vegetation index [22], impervious percentage [23], and Normalized Difference Sand Index (NDSI) [24] as the predictor. One potential weakness of kernel-based methods is that by restricting the LST dependence to a limited number of kernels, they oversimplify the physical mechanisms underlying spatial variations in LST, especially if a dominant physical driver cannot be obtained from the spectral channels of the fine-resolution image.

A second class of methods, the first of which is the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) model [25], is based on the idea of spatiotemporal image fusion for sharpening satellite LST imagery. These methods do not require fine-resolution kernel images. They assume that the coarse-resolution homogeneous pixels provide identical temporal changes as fine-resolution observations from the same spectral class. The temporal variation is obtained by differencing the coarse-resolution images acquired on two different days. A fine-resolution image acquired on one of the days is used as the reference image, and the fine-resolution data at the other time is predicted by simply adding the temporal variation obtained from the pair of coarse-resolution images to the reference image. The realism of the predicted fine-resolution data, or sharpened coarse-resolution image, can be further improved by filtering out spectrally dissimilar neighboring pixels. A number of variants of STARFM have also achieved satisfying results. For instance, the Enhanced STARFM (ESTARFM) [26] has improved the prediction accuracy for heterogenous landscapes. In an application of STARFM, Huang et al. [27] deployed bilateral filtering in LST data fusion. Wu et al. [28] extended the STARFM-based LST data fusion model from two sensors to an arbitrary number of sensors, using a method called Spatial-temporal Integrated Temperature Fusion Model (STITFM). The BLEnd Spatiotemporal Temperatures (BLEST) [29] method uses both the diurnal and annual temperature cycle [30,31] to bridge the scale gaps among arbitrarily selected satellite sensors.

Generally, STARFM-based fusion models pay a closer attention than kernel-based approaches to temperature-to-temperature relationships between coarse and fine resolution images, allowing more flexibility for LST sharpening at multiple time scales. However, oversimplification of the relationships between independent and dependent variables is still a concern. The coarse and fine resolution LST data are collected by different satellite sensors with different view angels, processing chains, and possibly mismatched sensing times of the day. As a result, biases in geolocation and atmospheric interference are inevitable [25].

A more reasonable hypothesis than the linear framework of STARFM is that the relationship of LST between different sensors is non-linear [17]. A method that accounts for the nonlinearity may improve the sharpened LST maps.

Recent studies have shown that the high non-linearity of deep learning models may provide an appropriate solution to inter-sensor biases. Song et al. [18] used a spatial-temporal fusion model based on Convolutional Neural Network (CNN) to generate Landsat-like surface reflectance from concurrent MODIS images. The model, named STFDCNN, deploys 5-layer CNNs to learn the inter-sensor biases and another 5-layer CNNs to sharpen the coarse-resolution images. In another study, the surface reflectance maps were sharpened with a two-stream CNN (StfNet). The spatiotemporal temperature fusion network (STTFN) proposed by Yin et al. [17] is among the first attempts to apply deep learning models to LST data fusion, with the aim of fusing MODIS LST and Landsat TM/ETM + thermal data.

CNN models appear to be preferable to regular neural networks in terms of feature extraction from imagery data. However, the convolutional operations posed additional challenges for satellite LST fusion.

LST fusion via CNN models is more challenging than surface reflectance fusion for two reasons. First, the surface reflectance is derived from multiple co-registered bands, but the LST data is obtained from a single band. Second, LST images have lower spatial resolutions than surface reflectance images. These characteristics limit the number of learnable features that can be extracted by CNN models for LST fusion.

A third challenge is that CNN models can only be fed with images without missing values. If a portion of the image is contaminated by cloud, the entire image must be discarded from the training set. A possible solution to cloud contamination is training the CNN models using image patches, which are small subsets sampled from the entire scene [17]. The patches with cloud are excluded from training so that the training still makes full use of the clear-sky portion of the scene. The size of patches is a critical parameter that needs careful determination. A large patch size leads to too much data loss when cloudy pixels occur within the patches, while a small patch size limits the within-patch variations of LST that can be learned by the CNN. For this reason, the CNN models in previous studies are based on clear-sky images.

In comparison to a CNN, a regular neural network is trained using single pixels of the LST data. The pixel-level training minimizes the data loss when discarding the cloudy pixels. Although the spatial variation of LST is not learnable for the regular neural network, its non-linearity is still a useful feature to reduce the inter-sensor biases.

In this study, we propose a novel framework for spatiotemporal fusion of LST to achieve the following goals:

1. To address the non-linearity of inter-sensor LST relationships with incorporation of a neural network,
2. To capture temporal change in LST in multiple scales, and
3. To generate high-quality fine resolution LST images in urban areas to support studies of intra-city temperature variations.

2. Materials and Methods

2.1. Study Area

Our study area is New Haven County, Connecticut, United States (Figure 1). It encompasses the City of New Haven, West Haven, and their outskirts. New Haven County is located in the temperate climate zone. The mean diurnal change of temperature is about 9.2 °C. The monthly mean air temperature ranges from 2.8 °C to 25.7 °C, with an annual mean of about 12.3 °C. Snowfall occurs from November to April. The total population was 847,000 in 2020 census. New Haven is mostly an industrial city supported by a dense road network (black lines in Figure 1). The compact man-made structures in urban land are surrounded by densely vegetated lands. This highly heterogeneous landscape provides a rigorous test of our model performance.

2.2. LST Data

The coarse-resolution LST data is derived from the MODIS satellite. The MOD11A1 Version 6 product provides daily per-pixel LST at 0.928-km spatial resolution. It is available on NASA's Land Processes Distributed Active Archive Center (LP DAAC; https://lpdaac.usgs.gov/products/mod11a1v006/; accessed date: 20 August 2021). The original sinusoidal pixels were resampled to 100 m then aggregated to 1 km. The local time of observation is 11:48 AM Eastern standard Time.

For fine-resolution LST data, we used Landsat 8 Provisional Surface Temperature (Collection 1), hosted by United States Geological Survey (USGS). The data originated from the Landsat Collection 1 Level-1 thermal infrared (TIR) bands. The TIR band 10 has been converted to surface temperature using an operational calibration methodology [32,33]. In this collection, the atmospheric effects have been corrected using atmospheric profiles from North American Regional Reanalysis (NARR) data. The surface emissivity is obtained from ASTER Global Emissivity Database (GED). Specifically, the ASTER GED emissivity is first corrected to Landsat-8 band 10 by a spectral adjustment. An NDVI-based vegetation adjustment is then performed to account for inconsistent phenology between ASTER and Landsat scenes [34]. The time of observation is approximately 10:30 AM Eastern Standard Time. The Landsat LST data are archived at 30-m spatial resolution and 16-day temporal resolution. In this study, we processed the data to 100-m spatial resolution identical to the original resolving quality of the TIR thermal bands. Cloud, snow and ocean pixels were excluded from model training, but inland water pixels were retained.

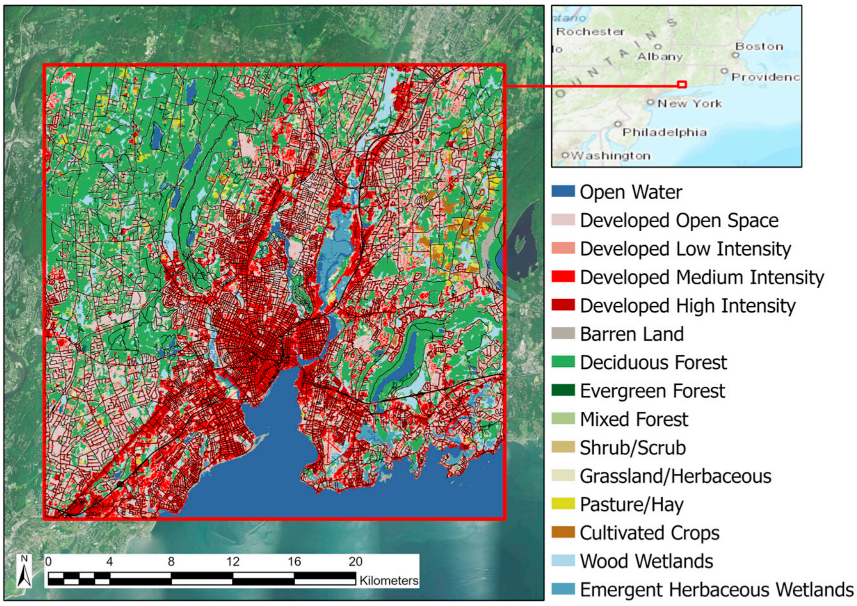


Figure 1. Land use types of New Haven County, Connecticut, United States. The red box indicates the study area. Black lines display the road network. The land use data is provided by National Land Cover Database (NLCD) of 2016 (https://www.mrlc.gov; accessed date: 1 October 2021). The background is a true-color aerial image in 2018 provided by National Agriculture Imagery Program (NAIP).

In this study, we selected a total of 26 image pairs from 1 January 2013 to 1 May 2020 (Figure 2). Each pair consists of one Terra MODIS and one Landsat LST image acquired on the same day, the latter of which is referred to as the target Landsat LST. The target Landsat image in each pair has less than 1% missing pixels (due to cloud) in the study domain. All the images were projected to the World Geodetic System 1984 (WGS84) for co-registration.

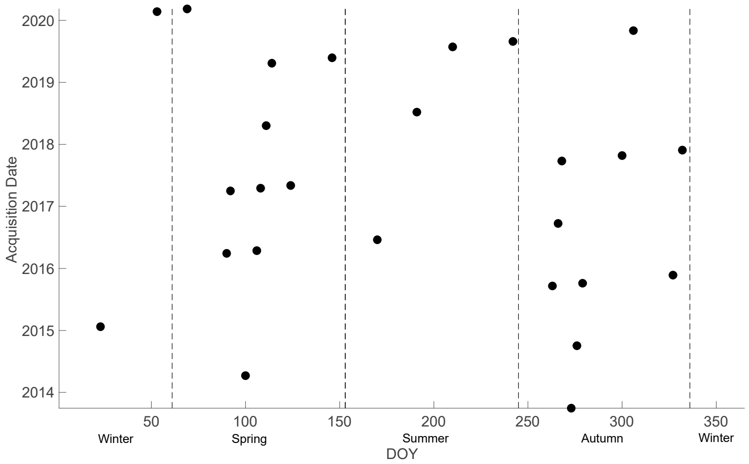
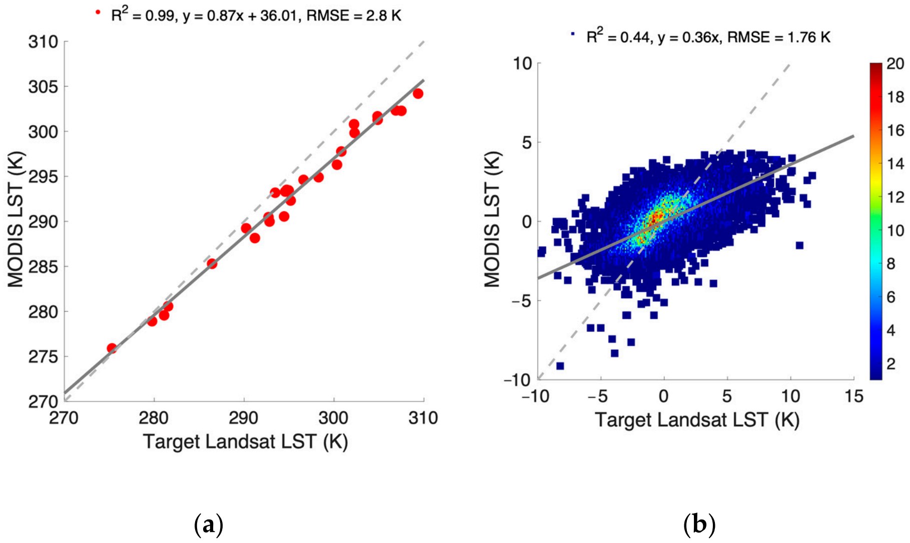


Figure 2. A scatter plot illustrating the acquisition Day of Year (DOY) of 26 MODIS-Landsat image pairs. The gray dash lines show season divisions.

2.3. Sensor-to-Sensor Biases

We found inter-sensor biases at the scene and the patch scale. For the 26 pairs of MODIS and target Landsat LST images, the scene-average LST shows significant correlation (Figure 3a; $R^2 = 0.99$). The mean bias (MODIS minus Landsat) is −2.39 K, and the RMSE is 2.8 K. The slope of the regression of MODIS LST versus Landsat LST is less than 1, indicating that MODIS LST is more negatively biased in warmer conditions. The most negative bias (−5.21 K) occurred on 29 July 2019.



(**a**) (**b**)

Figure 3. Comparisons of LST at difference scales. (**a**) Comparison of scene-scale mean between MODIS and Landsat LST. (**b**) Comparison of patch-scale LST deviation from the scene mean between MODIS and Landsat. Color in (**b**) indicates data density, presented by number of pixels within each 0.1-K bin. The solid grey lines show the linear fitting functions with statistics noted and the dashed grey lines indicate the 1:1 relationship.

Biases also occurred at the patch-scale, or scale of the MODIS pixels. For patch-scale comparison, we averaged the 10 by 10 Landsat pixies in each MODIS pixel grid to produce the patch-scale, or resampled Landsat LST. We then compared the deviation of the patch-scale LST from the scene mean between the resampled Landsat and the MODIS LST in the same image pair. The patch-scale LST is highly correlated (Figure 3b; $R^2 = 0.44$). However, the slope of the regression (MODIS versus Landsat) is only 0.36. Because the scene-level mean LST values were removed, the scatter seen in Figure 3b was caused by spatial variations across the scene. In other words, spatial variations in the MODIS LST were only 36% of those in the resampled Landsat LST. Furthermore, the linear fit can only explain 43% of the variability, suggesting that a model with higher complexity than linear regression is needed to reduce the inter-sensor error.

2.4. Framework Description

2.4.1. Workflow

Our scale-separating framework produces Landsat-like fine-resolution LST with a 3-step workflow. The three steps are: Linear Stretching across Time (LSAT), Neural Network (NN) processing, and Enrichment of Fine-resolution Variation (EOFRV).

Figure 4 illustrates the whole workflow. In this example, the goal is to sharpen the MODIS LST observed on 15 April 2016 (panel a). The MODIS LST of size 22 × 29 passes through the LSAT model to correct the patch-level biases (panel c; called adjusted MODIS). Next, the NN model is applied to the image after the scene-scale mean temperature has been removed (panel e). The result is an improved LST image, called predicted patch-level LST, at the original MODIS resolution (panel g). On this particulate date, we have a Landsat image available for validation. This image (panel b) is resampled to the patch-scale (resampled Landsat LST; panel d) and is subtracted by its scene-scale mean value (panel f). The patch-scale validation is made using images (g) and (f). The predicted patch-level LST is up-sampled to the Landsat resolution (panel h).

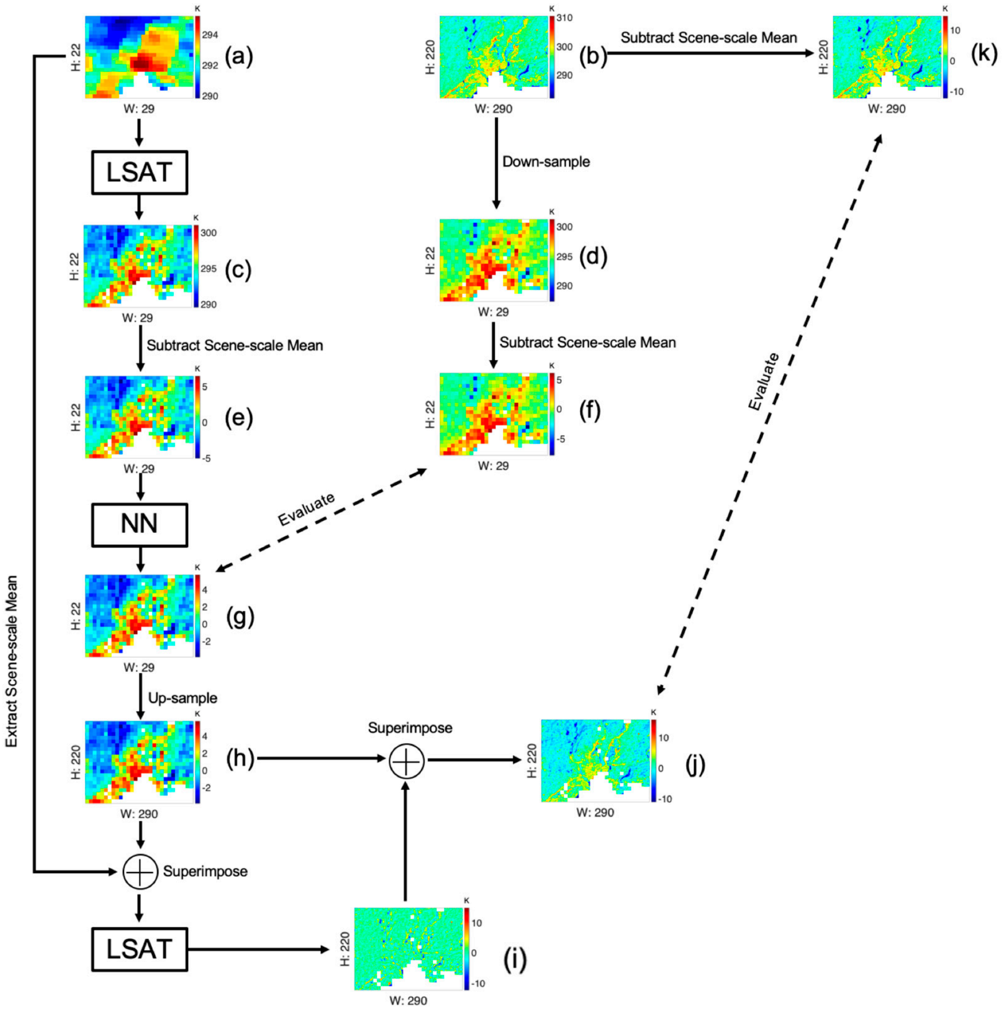


Figure 4. Processing workflow of the scale-separating framework applied to the MODIS scene from 15 April 2016. The width (H) and height (H) of each image are attached to the axes to reflect the dimension changes caused by up-sampling and down-sampling operations. Panel a: MODIS LST; panel b: target Landsat LST; panel c: adjusted MODIS LST; panel d: resampled target Landsat LST; panel e: adjusted MODIS LST variations; panel f: resampled Landsat LST variations for validation; panel g: predicted patch-level LST; panel h: up-sampled predicted patch-level LST; panel i: predicted within-patch variations; panel j: sharpened MODIS LST (intra-scene variations); panel k: target Landsat LST variations for validation.

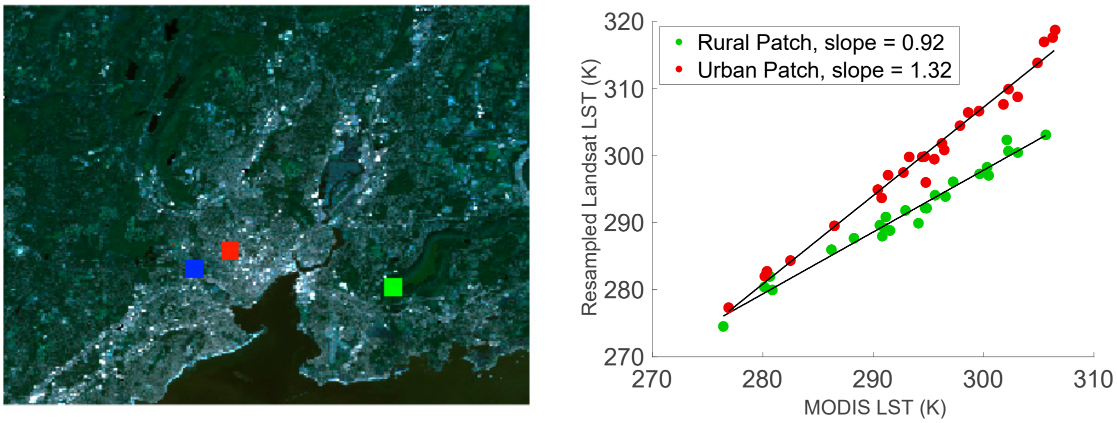
The next step is to enrich the improved patch-level image (panel g) with fine-resolution details. We built EOFRV, a sharpening model, to learn the relationships between the LST of each patch and the Landsat pixel-scale variations within the patch. The across-scale relationships are based on the LST data on all other dates. The model ingests up-sampled predicted patch-scale LST (panel h), superimposed on the scene-scale mean of adjusted MODIS (panel c), and predicts the pixel-scale variations (panel i) on the target date (15 April 2016).

The final step is simply superimposing the pixel-scale variations (panel i) on the up-sampled predicted patch-level LST (panel h). The superimposition produces the sharpened MODIS LST (panel j), which can be compared to the target Landsat LST (panel b) in terms of intra-scene variations (panel k).

In this example, the results can be evaluated (black dashed arrows in Figure 4) at the patch-level (image g versus image f) and at the Landsat pixel resolution (image j versus image k, the target Landsat without scene-scale mean). To sharpen an arbitrary MODIS image not in our archive of image pairs, the workflow ends with image j.

2.4.2. Linear Stretching across Time (LSAT)

The seasonality of LST of urban land is generally larger than that of rural land. This contrast is illustrated by the data shown in Figure 5b for an urban (red square, Figure 5a) and a rural patch (green square, Figure 5a). According to the resampled Landsat data, the urban patch experiences a larger seasonal variation of temperature (from 277.32 K on DOY 23, 2015 to 318.73 K on DOY 191, 2018, a range of 41.41 K) than the rural pixel (from 274.54 K to 303.10 K, a range of 28.55 K). Composed of man-made materials, urban patches absorb a large amount of solar radiation, leading to high LST in the summer. For rural patches, evaporative cooling provided by vegetation prevents LST from rising to high values. These urban-rural contrasts in seasonality are much reduced in the MODIS data (range of variation 29.58 K for the urban patch and 29.22 K for the rural patch). The MODIS and Landsat observations of the urban patch produce a linear fit with the slope greater than 1 (slope = 1.32), indicating that the Landsat LST has a larger seasonality than the MODIS LST. In comparison, for the rural patch, the regression slope is slightly less than 1, indicating a relatively slower response of the Landsat LST to seasonal warming than the MODIS LST at the patch scale. The minimum temperature in both MODIS and Landsat LST is slightly higher for the urban patch than for the rural patch, possibly indicating additional heat source due to anthropogenic activities in urban lands.



(**a**) (**b**)

Figure 5. Temporal LST variations of selected patches in the scene: (**a**) locations of an urban dominated patch (red square) and a rural dominated patch (green square), and (**b**) linear relationships across time between MODIS LST and resampled Landsat LST. The urban patch is dominated by medium-intensity developed area (59%), with the rest area to be high-intensity (31%) and low-intensity development (10%). The rural path is mainly forested (72%), with 25% inland water. The blue square marks a patch of urban land mixed with rural land.

We use a linear stretching model to adjust the MODIS LST pixel-by-pixel (or patch-by-patch) according to the LST relationship across time between MODIS and resampled Landsat. From the 26 image pairs, up to 26 LST observations can be found for a given image patch. For a target date, a linear relationship is developed for each patch using the observations on all other dates. In total, there are 22 × 29 linear regression equations (see the example for 29 July 2019 in Figure S1) for the study area. These equations are used to make patch-by-patch correction to produce the adjusted MODIS LST (Figure 4c).

2.4.3. Neural Network

A neural network is used to further reduce the inter-sensor biases. The adjusted MODIS image after the removal of scene average (Figure 4e) is concatenated with three data layers (DOY, IMP, and MODIS NDVI) to produce a 4-band image stack. The DOY band is the DOY value of the target date and is assigned to all the patches. IMP is an index on imperviousness of the surface provided by the National Land Cover Database (NLCD). MODIS NDVI is the concurrent NDVI layer derived from MOD09GQ. IMP and NDVI are resampled from their original 100-m and of 250-m resolution, respectively, to the 1-km patch scale. DOY is expected to help the model learn seasonal variations of LST. IMP and NDVI are expected to teach the model about morphological relationships between LST spatial variations. The four bands in the image stack are then normalized to the range of 0–1. For applications outside the US, the imperviousness data from Copernicus Global Land Service (CGLS) can be used to replace IMP without degrading the neural network performance.

The image stack and the corresponding resampled Landsat image free of scene average (Figure 4f) are used for network training at the patch level. The network ingests four band values of a patch (Figure 6a) from the image stack, extracting features using three high-dimensional hidden layers (Figure 6b–d), and predicts a new patch-level LST (Figure 6e). Sigmoid activation and batch normalization are used on all the hidden layers. The neural network is trained using mini batches sized at 500 patches. At the batch level, the difference between the prediction and the corresponding LST in the resampled Landsat data (Figure 6f) is measured by the L1 norm loss function (Mean Absolute Error). The loss function is minimized by Stochastic Gradient Descent (SGD) through an iterative training process using a relatively small learning rate of 0.0001. A backpropagation algorithm distributes the error to the model parameters and updates their values to improve the prediction. The prediction error declines as the model parameters are updated iteratively. One epoch of training is completed after the model has passed through all the training batches. To evaluate the improvement at each iterative step, we use the average loss value of all the batches in that epoch.

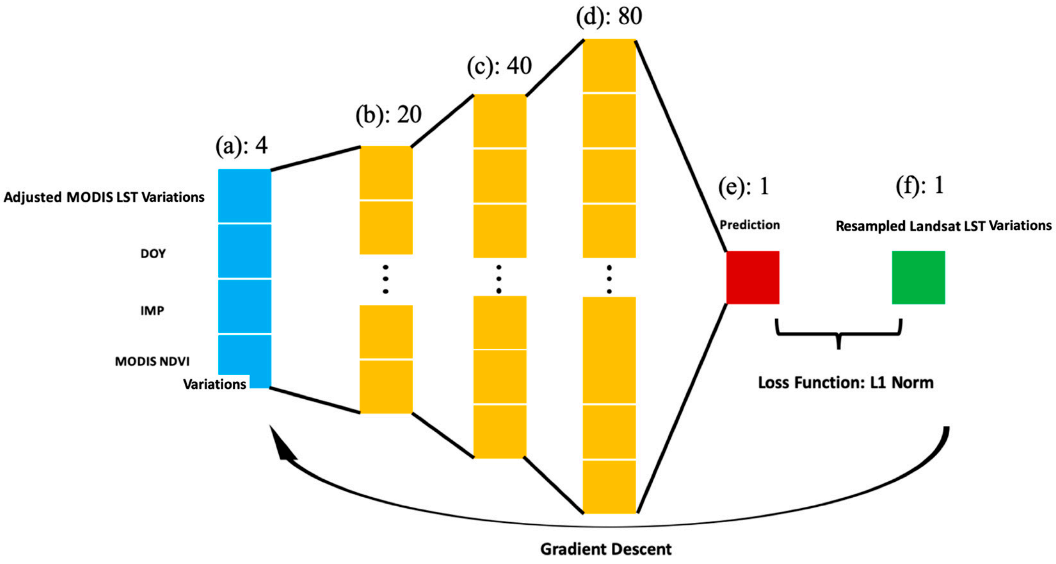


Figure 6. Architecture of the neural network to improve the MODIS LST intra-scene variations. Layer a: 4-dimensional input layer; layer b: 20-dimensional hidden layer; layer c: 40-dimensional hidden layer; layer d: 80-dimensional hidden layer; layer e: one-dimensional output layer; panel f: resampled Landsat LST Variations.

To evaluate model performance, we selected one image pair at a time for validation and train the model using the other 25 pairs. At the end of each epoch, the validation loss or error is the L1 norm between the predicted LST and the corresponding resampled Landsat

LST (Figure 4f). The process is repeated 26 times. The optimal number of epochs (NOE) with the lowest validation error is recorded for each validation.

2.4.4. Enrichment of Fine-Resolution Variations

The within-patch variations at the Landsat pixel scale change with time. The assumption behind the EOFRV is that the temporal change of within-patch variation is caused by the temporal change of local climate if the land cover is unchanged. To validate this, we selected the patch of urban-rural mixture in Figure 5a (blue square) and examined the temporal change of within-patch variations of Landsat LST. For an urban pixel in this patch, the within-patch variation (pixel LST minus patch mean LST) first increases and then decreases with DOY after peaking at DOY 170 (Figure 7, red dots). An opposite trend is observed for a rural pixel, who's within-patch variation first decreases and then rises after DOY 191 (Figure 7, green dots). The variation is positive for the urban pixel and negative for the rural pixel. In other words, the urban pixel is warmer than the patch mean and this deviation is stronger in warm seasons than in cold seasons. On the contrary, the rural pixel is cooler than the patch mean and the cooling tendency is stronger when the local climate becomes warmer.

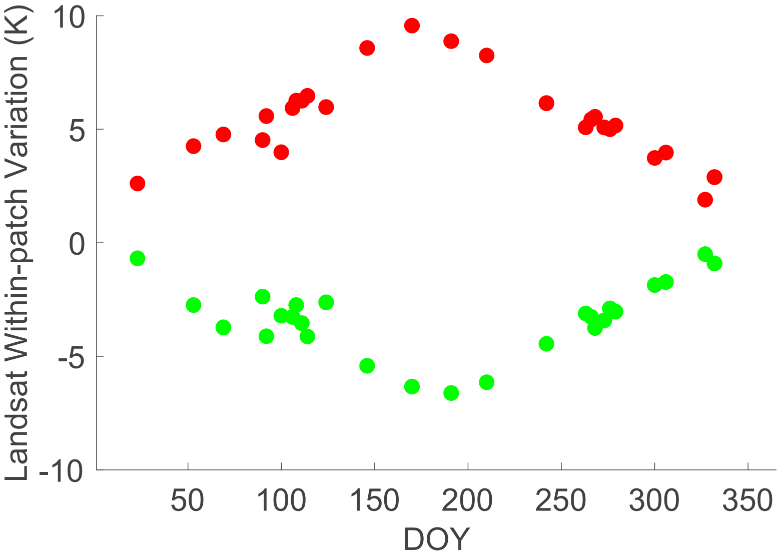


Figure 7. Seasonal change of Landsat within-patch variations (pixel LST minus patch mean LST). Red dots: urban pixel; green dots: rural pixel.

The EOFRV predicts within-patch variations using DOY and the patch-level LST, an additional predictor, on the basis of a quadratic equation (Equation (1)),

$$T_{var} = aDOY + bDOY^2 + cT_{patch} + dT_{patch}^2 + e \tag{1}$$

where T_{var} is the Landsat with-in patch variation for a single pixel and T_{patch} is the mean LST of the patch containing that pixel. In this equation, DOY accounts for seasonal and the day-to-day changes of T_{var}, and T_{patch} is a dynamic variable accounting for inter-annual changes. The coefficients a, b, c, d, and e were found by a least-square method using the patch and pixel-scale Landsat images on all dates excluding the target date (Figure 4d,k). In all, there are a total of 63,800 sets of coefficients, each corresponding to one Landsat pixel.

The modeled T_{var} for the urban and the rural pixels in Figure 7 is shown as a function of DOY and patch mean (Figure 8). In this plot, T_{var} is a convex surface for the urban pixel and a concave surface for the rural pixel. As the patch-level LST increases, the surface skews upward for the urban pixel (Figure 8a), and downward for the rural pixel (Figure 8b). The model surfaces fit well with the observations (black point in Figure 8). The RMSE values are 0.55 K and 0.56 K for the urban and the rural pixels, respectively. Compared with the ranges (~7 K) of within-patch variations, such error is small.

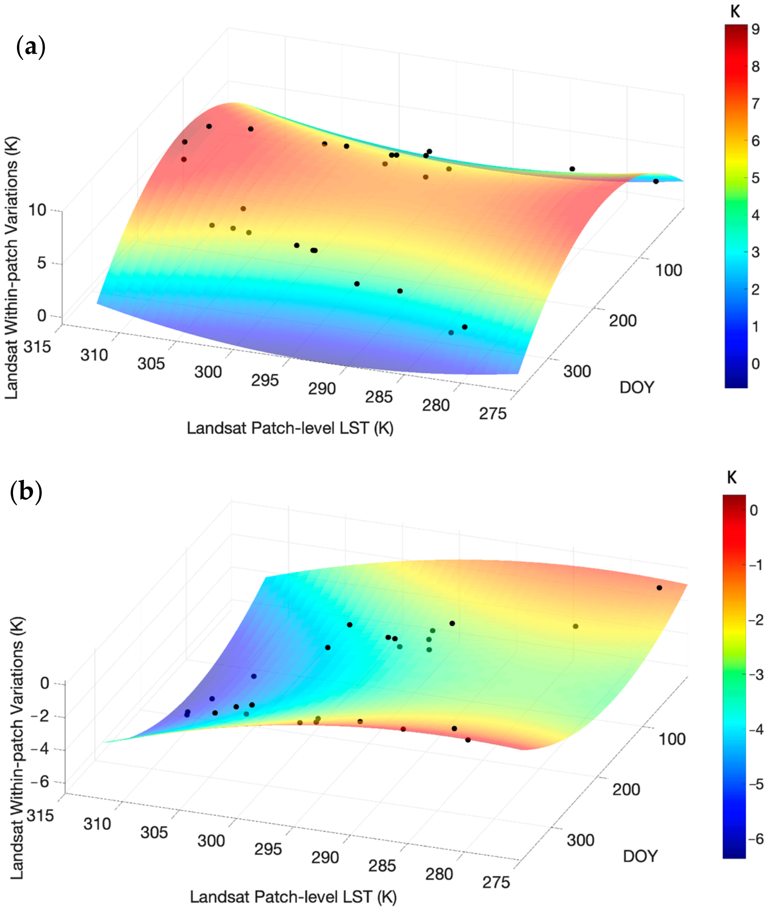


Figure 8. Comparison of modeled within patch temperature variation (pixel LST—patch mean LST) for an urban pixel (**a**) and a rural pixel (**b**). Curved surfaces: EOFRV predictions; black dots: Landsat observations.

To apply this model to a sharpening task, T_{patch} is replaced with the predicated patch-level LST (Figure 4h) added the scene-scale mean of adjusted MODIS (Figure 4c). In Figure S2, we show that the model performs similarly well to the results produced by Landsat T_{patch} as the predictor. Unlike the Landsat within-patch variations, the predicted variations of a patch do not strictly center at zero due to statistical noises.

2.4.5. Sharpening an Arbitrary MODIS Image

The workflow for sharpening the MODIS image on an arbitrary date (that is, image not in our archive of 26 image pairs) is similar to that shown in Figure 4, with two differences. First, the LSAT, NN, and the EOFRV models are trained on all the 26 image pairs. Second, steps shown as panels b, d and f are not implemented.

3. Results

3.1. Training and Validation Loss

The training and validation results are summarized Figure S3, and the ensemble mean training and validation losses are shown in Figure S4. Both the training loss and the validation loss drop drastically in the beginning (number of epoch NOE < 10) and much more gradually when the number of epochs exceeds 30. These losses are stabilized at ~200 epochs, indicating that the global minimum has been reached. For training, the highest final Mean Absolute Error (MAE) is 0.56 K for 15 April 2016, and the lowest is 0.51 K for

4 May 2017. For validation, the final highest MAE is 0.98 K for 9 March 2020 and the lowest is 0.34 K for 6 October 2015. The validation loss is apparently higher than the training loss for 4 May 2017 and 9 March 2020. For all other dates, the training and validation loss are similar, showing no evidence of overfitting. In other words, the trained NN model is able to interpret reasonably well the relationship between the intra-scene variations of MODIS and Landsat LST in the unseen (validation) dataset. In the following, the model uses a NOE of 200 for LST sharpening. The corresponding MAE is 0.53 K (Figure 9).

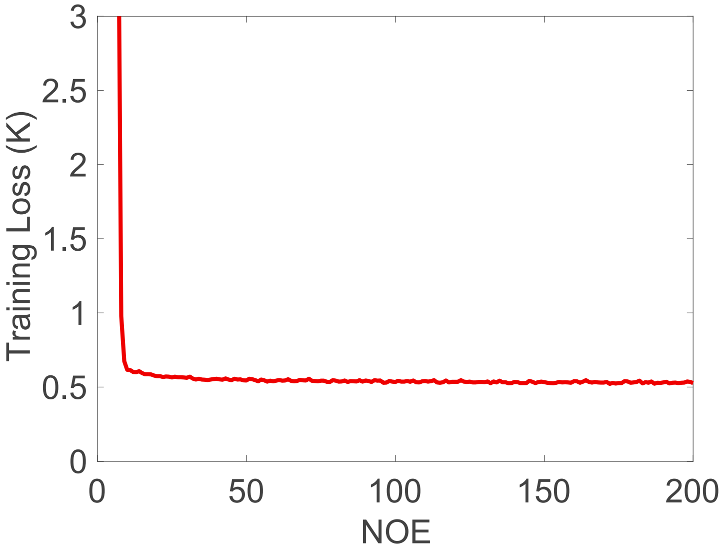


Figure 9. Training loss as a function of NOE (number of epoch). Here the loss function is based on all 26 pairs of MODIS LST and resampled Landsat LST scenes.

3.2. Accuracy Assessment

Accuracy was evaluated using RMSE against the resampled Landsat after the scene-level mean value has been subtracted from both. The RMSE was calculated for the whole scene as well as for urban and rural areas. Of note, the scene-level mean was not corrected. In other words, these RMSE values provide accuracy assessment on the spatial variations in LST, not on the LST absolute accuracy. Determination of whether a patch belongs to the urban or rural class was made with the land cover map for 2017 provided by Copernicus Global Land Service (CGLS) (https://lcviewer.vito.be/about; accessed date: 15 July 2021). The map of 100-m resolution was down-sampled to the 1-km patch scale. A patch is classified as urban if half or more of the original pixels were in the urban class.

Figure 10 is a summary of the accuracy statistics before (Figure 10a) and after the processing steps (Figure 10b,c). Prior to any processing step, the scene-scale RMSE ranges from 0.77 K to 2.94 K, with a mean of 1.68 K. The urban RMSE ranges from 1.15 K to 4.71 K, with a mean of 2.54 K. The rural RMSE is much lower, ranging from 0.66 to 2.45 K with a mean of 1.41 K. LSAT has reduced these errors, especially in the warm season (Figure 10b). After LSAT, errors are similar among urban and rural patches. The NN processing has reduced the RMSE further, by an average of 0.12 K (Figure 10c). Although the RMSE change brought by NN is small, the rural and urban errors are nearly identical for most target dates. Furthermore, the NN-predicted LST shows better spatial correlation with the resampled Landsat than the LST with only the LSAT application (see below).

The improvements brought by LSAT and NN are evident in Figure 11, which compares the original MODIS (panel a), the MODIS after LSAT (panel c) and that after NN (panel d) with the resampled Landsat (panel b) for 23 January 2015. The MODIS LST map without correction appears much smoother than that of the resampled Landsat LST. Their temperature ranges are 3.44 K (Figure 11a) and 6.59 K (Figure 11b), respectively. After LSAT, the temperature range is stretched to 4.56 K (Figure 11c). The resemblance to the resampled Landsat LST is improved with each processing step: the spatial R^2 of the resampled Landsat is 0.45 with the original MODIS, 0.55 after LSAT and finally 0.86 after both LSAT and NN.

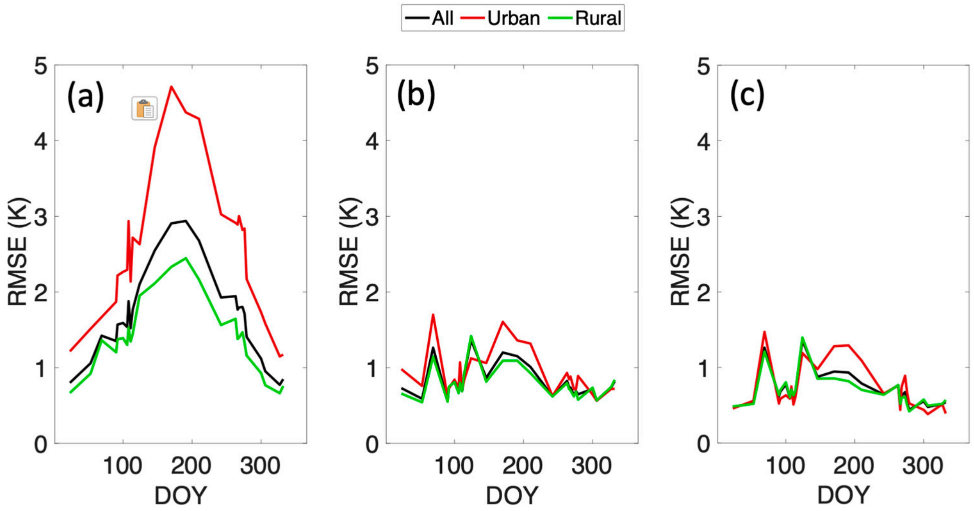
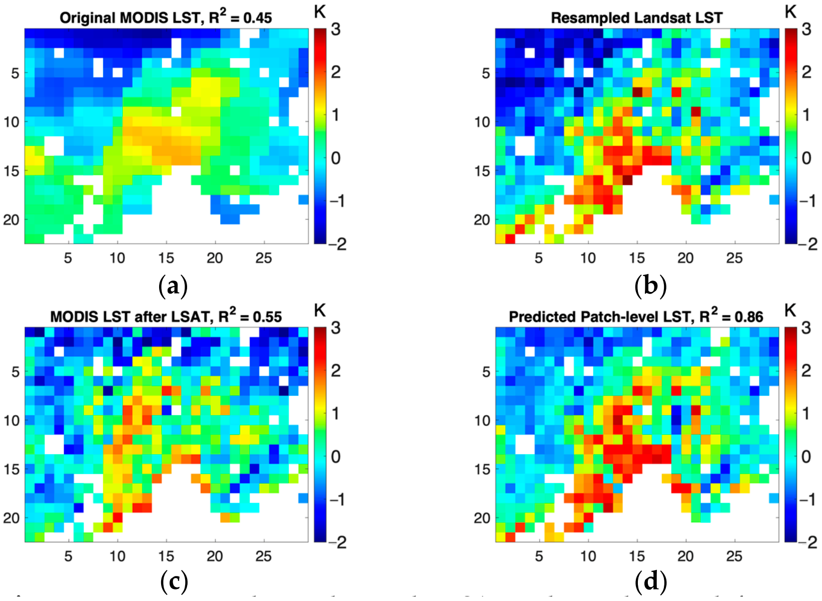


Figure 10. Inter-sensor biases between MODIS and Landsat patch-level LST. (**a**–**c**) display the RMSE of original MODIS LST, adjusted MODIS LST and predicted LST, respectively, against resampled Landsat LST. The black, red and green lines denote the entire scene, urban areas and rural areas, respectively.


Original MODIS LST, R^2 = 0.45
Resampled Landsat LST
MODIS LST after LSAT, R^2 = 0.55
Predicted Patch-level LST, R^2 = 0.86
K
3
2
1
0
−1
−2
5
10
15
20
25

(**a**) (**b**)

(**c**) (**d**)

Figure 11. Inter-sensor bias reduction by LSAT and neural network for 23 January 2015. (**a**) is the original MODIS LST, (**b**) is the resampled Landsat LST on the same date, (**c**) is the MODIS LST adjusted by LSAT, and (**d**) is the LST predicted by the neural network. All the LST images are free of scene-scale mean, showing intra-scene variations only. The spatial R^2 of (**b**) with (**a**,**c**,**d**) is marked.

3.3. Evaluation of Sharpened LST

Figure 12 shows the RMSE for the sharpened LST at the 100 m resolution (e.g., Figure 4j). The whole-scene RMSE ranges from 0.58 K to 1.51 K, with a mean of 0.91 K. The RMSE for rural areas shows a similar seasonal variation (0.55 to 1.54 K, mean of 0.91 K). The urban RMSE is also similar, except for the three marked dates.

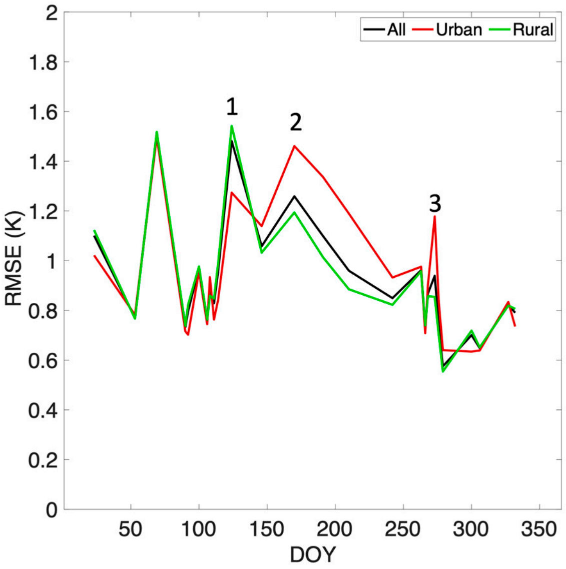


Figure 12. RMSE of the sharpened LST at the 100-m resolution for the entire scene (black line), urban areas (red line), and rural areas (green line). Three error peaks are marked by numbers.

Figure 13 compares the sharpened LST map from the MODIS data with the original Landsat results for a fall date (22 September 2016). Examples for dates in the spring (4 May 2017), summer (29 July 2019), and winter (28 November 2017) are given in supplementary Figures S5–S10. The R^2 of pixel-by-pixel correlation is 0.95 for the entire scene (Figure 14a) and 0.93 for the urban area (Figure 14b). The accuracy is the highest in the spring (Figure S6), and lowest in the fall (Figure 14a,b).

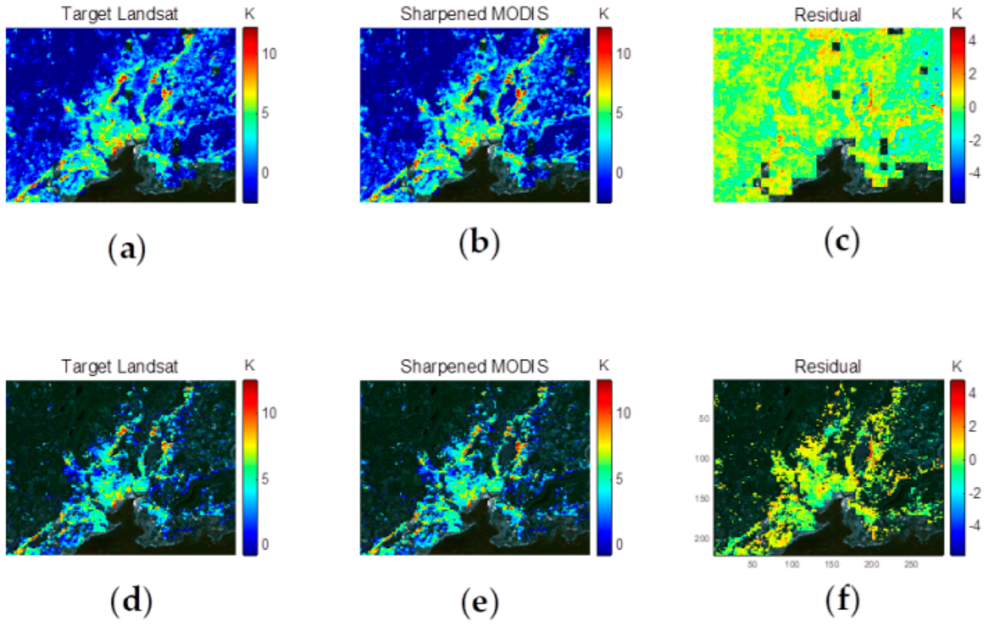


Figure 13. Sharpening LST map for 22 September 2016. (**a**,**b**) are the target Landsat LST and the sharpened MODIS LST for the entire scene, and (**d**,**e**) are for urban areas only. (**c**,**f**) are residual maps (sharpened MODIS minus target Landsat) for the entire scene and for urban areas, respectively.

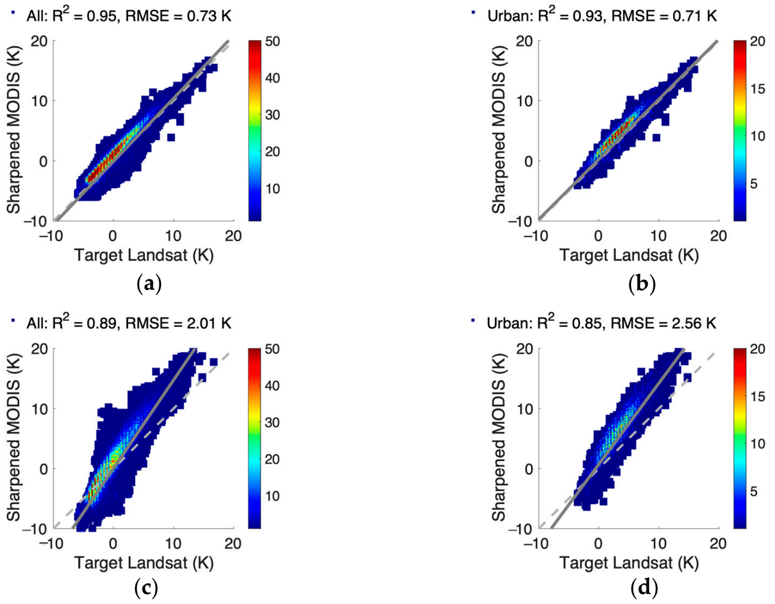


Figure 14. Pixel-by-pixel correlation between the target Landsat LST and the sharpened MODIS LST on 22 September 2016 for (**a**) the entire scene and (**b**) urban areas using our SS method. For comparison, the sharpening results with the Bilateral Filtering (BF) method are shown in (**c**,**d**). Color indicates data density, presented by number of pixels within each 0.1-K bin. The solid gray lines show the linear fitting functions and the dashed gray lines indicate the 1:1 relationship.

3.4. Comparison with Bilateral Filtering

Our scale-separating (SS) framework has achieved better results than bilateral filtering (BF; Figure 15). We chose to bench-mark our framework against the BF method [27] because the latter provides high-quality LST data for UHI studies. We used a moving window of size 10 × 10 in the bilateral filtering process. The spatial weight and similarity weight were determined with a geometric spread of 0.4 and a temperature spread of 2.5. The weights were found with minimized sharpening error to within a small range (0.1–2.5). The average RMSE for BF is 1.63 K for the entire scene, 1.93 K for urban areas, and 1.53 K for rural areas (Figure 15a). The corresponding RMSE values for our SS framework are 0.91 K, 0.94 K and 0.91 K.

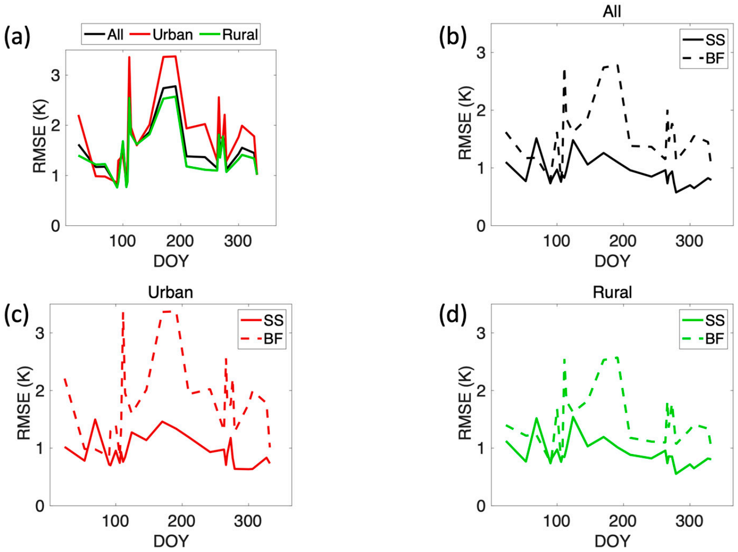


Figure 15. Comparison of accuracy between bilateral filtering (BF) and scale-separating (SS) framework. (**a**) RMSE for BF. (**b**–**d**): comparison between BF and SS for the entire scene, urban areas, and rural areas.

4. Discussion

4.1. Error Analysis

Sharpened LST data are affected by errors at three separate scales (scene, patch and pixel). Inter-sensor biases affect not only the scene–scene mean but also the spatial distribution of LST. In this study, we subtracted the scene-scale mean from the patch-scale temperature to remove the mean bias, but errors in the spatial distribution still remain (Figure 10a).

One established reason for inter-sensor biases is related to the surface emissivity used for MODIS and Landsat LST retrieval. MODIS emissivity is obtained with an NDVI classification method [35]. Landsat emissivity is obtained from the ASTER satellite. The emissivity discrepancy can lead to differences in the conversion from brightness temperature to the LST. However, emissivity parameterization alone cannot fully explain the patch-scale error shown in Figure 3b. A recent study [36] shows that the surface UHI intensity, which is a measure of spatial variations of LST, only changes by about 0.1 K when the Landsat emissivity is replaced with an NDVI-based parameterization used by MODIS.

We postulate that the difference in view angle between MODIS and Landsat sensors may have contributed to these errors. New Haven County is located at the fringe of MODIS swaths, with a view angle of 65° from zenith and from the west of ground targets [37]. In comparison, the Landsat 8 sensor scans the surface from the nadir point (zenith angle about 0°). In other words, the MODIS observations are influenced more by cold shadows than the Landsat observations. As a result of this anisotropic heating, the off-nadir observation will detect weaker thermal radiation than the nadir scan [38,39]. That solar heating difference among ground objects is smaller in the winter and larger in the spring and summer may potentially explain the seasonal variation in the RMSE (Figure 10a).

On several summer dates (DOY 170, 191, and 210), urban RMSE is greater than rural RMSE (Figure 10c). We infer from this that the degree of vertical heterogeneity has impacted the quality of the results from LSAT and NN. Natural surfaces are spatially more uniform and vertically more homogeneous than areas occupied buildings extending tens of meters in the vertical direction. These vertical structures will obstruct each other from the satellite view. This inter-occlusion effect in the urban area increases the LST difference between the two satellites, especially in warm seasons when vegetation in rural areas is most homogeneous. As a result, urban areas suffer from larger sharpening errors in the summer than rural areas.

A third type of sharpening error occurs at the pixel scale within the patches. It originates from the prediction of within-patch variations (Figure 4i) made by EOFRV, which are similar but not identical to the within-patch details of target Landsat (Figure 4b). The overall error has increased by 0.22 K after the EOFRV application, indicating that the pixel-scale error is secondary compared with the patch-scale error (0.70 K).

To analyze the source of sharpening error in detail, we now focus on three example dates marked in Figure 12. They are 4 May 2017 (DOY 124, mark 1), and 18 June 2016 (DOY 170, mark 2), and 30 September 2013 (DOY 273, mark 3). The sharpening error is larger for those dates than average, indicating relatively poor sharpening quality. Figures S11–S13 compare the patch-scale error and the final sharpening error for these dates. On 4 May 2017, the patch-scale error is 1.36 K (Figure S11c). This error has propagated to the sharpened MODIS (Figure S11e), contributing 92% of the final sharpening error (1.48 K), with the rest 9% (0.12 K) explained by the pixel-scale error. The patch-scale error is also the main error source for the other two dates, accounting for 76% and 71% of the final sharpening error for 18 June 2016 and 30 September 2013, respectively. The patch-scale error is consistent because the predicted patch-level LST is biased high by ~1 K in rural areas (Figure S11–S13c).

In generally, the pixel-scale error is small and is usually spread homogeneously over the scene. Among all the 1,354,300 pixels, most (93%) are predicted with pixel-scale error below 1 K, and only 130 pixels have significant error over 7 K. The significant errors occurred at fixed locations. For example, the black boxes in Figures S12f and S13f highlight a strip-shape area in rural land with errors up to −12.74 K. The biases in this area are found for nearly all the target dates, sometimes low for the entire shape (Figure S14a) and sometimes high for parts but low for the others (Figure S14b,c). A satellite map indicates that this area (red dashed line in Figure S15) is a quarry near North Branford, CT. It is mostly bare land in low-lying topography. The quarry was waterlogged from time to time, leading to drastic land cover change in the daily scale. The EOFRV model failed to produce accurate within-patch variations for this area because EOFRV assumes that land cover is constant within the time span of the study.

In addition to quarrying-related land cover changes (labelled as Type A, black points in Figure 16), four other types led to large pixel-scale errors (Figure S16). Type B is cultivated cropland under the influence of irrigation (yellow points in Figure 16). Type C is herbaceous wetland subject to varying levels of tidal water (blue points in Figure 16). Type D is buildings with white-roof whose reflectivity may have degraded over time (red points in Figure 16). Type E is ponds with occasional green algal blooms (green points in Figure 16). The land cover types of these pixels were variable over different years. Because EOFRV assumes constant land cover types, its predictions were inaccurate in those areas. The influences of such land cover changes are secondary in reference to the overall sharpening errors. According to the error histogram, 99% of the data are within the error bounds of ±1.97 K and 95% within ±1.16 K (Figure S17).

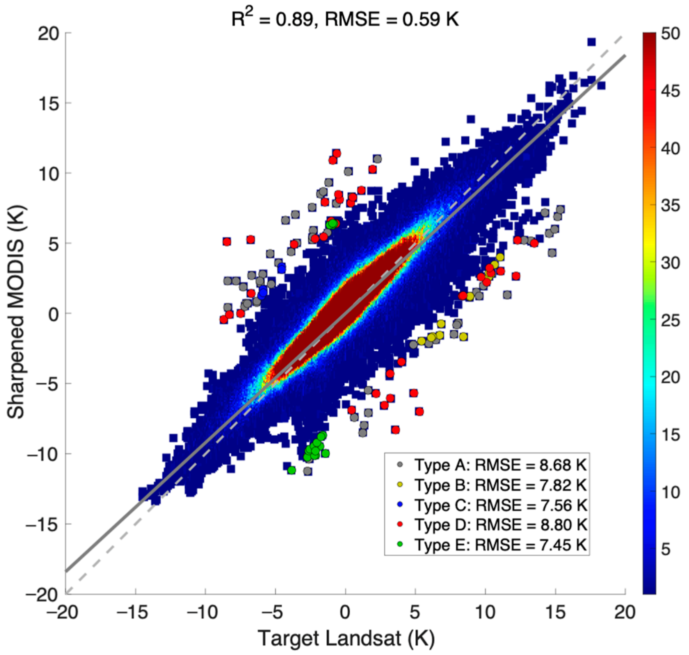
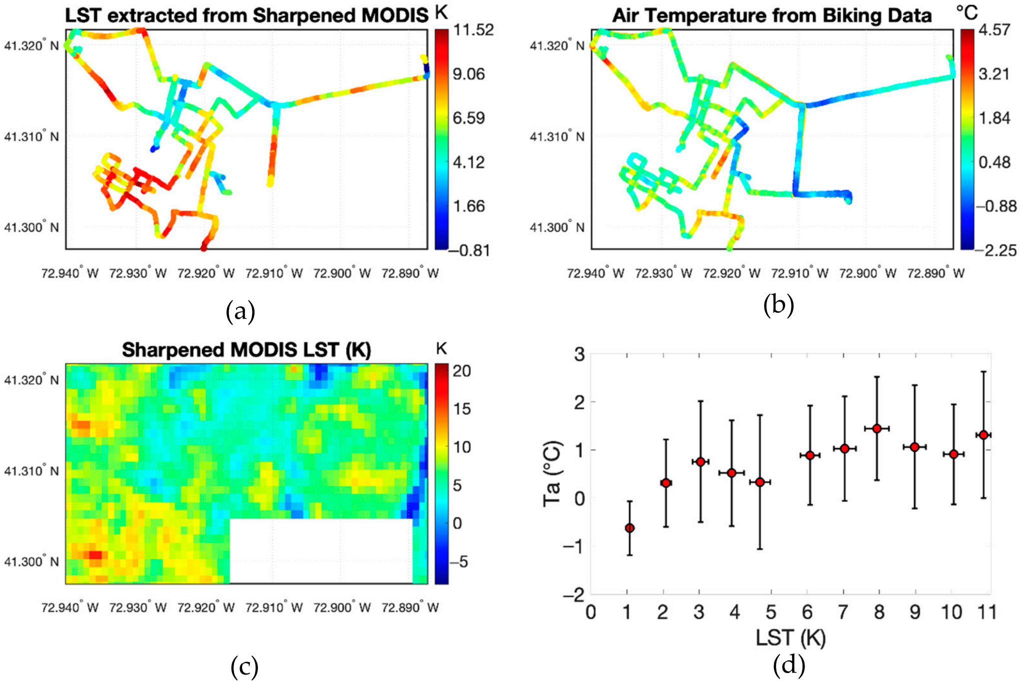


Figure 16. Pixel-by-pixel correlation between sharpened MODIS LST and target Landsat LST for within-patch variations and for all 26 dates. The overall spatial correlation (R^2) and RMSE are noted. The color bar indicates data density, presented by number of pixels within each 0.1-K bin. The solid gray line shows the linear fitting function, and the dashed gray line indicates the 1:1 relationship. Pixels of absolute biases of 7 K or more are marked in different colors depending on the types of land use.

4.2. Comparison with Air Temperature

The scale-separating framework opens new opportunities for UHI studies. To demonstrate this potential, we compared a sharpened MOIDS LST map (Figure 17c) for 12 August 2019 with air temperature (Ta) measured with smart sensors mounted on bicycles [40]. The bicycle measurement took place between 10: 37 and 13:14 local time, along transects that passed through urban core pixels with impervious surface fraction up to 100% as well as rural pixels with nearly zero impervious fraction. The air temperature from mobile sensors was relative to that measured by a stationary sensor. The LST spatial pattern along the bike routes (Figure 17a) bears strong similarity to the spatial variation of Ta (Figure 17b). The spatial correlation between LST and Ta is relatively strong (Figure 17d, $p < 0.001$). Both Ta and LST increase with increasing impervious surface fraction (Figure 18), conforming the role of impervious surface material in warming the surface and the near-surface atmosphere [41–44]. Impervious material limits the evaporative cooling and leads to higher surface temperature in urban lands than in rural lands [45]. Similar urban-rural differences are formed for air temperature as land surface transports sensible heat to the near-surface atmosphere [46]. Such warming effect is stronger on the surface than on the air, again in agreement with previous modeling studies [41,42]. Compared with rural vegetations, impervious material in urban lands tends to redistribute more solar energy to sensible heat [47], leading to greater vertical temperature gradient (LST minus Ta) [41]. As a result, air temperature is less variable in space than surface temperature and the slope of the linear relationship between LST and Ta is less than unity. Such fine-scale intra-city variability in LST would be impossible to discern using the original MODIS LST because the bicycle transect only covered a small number of MOIDS pixels.



(a)

(b)

(c)

(d)

Figure 17. A comparison between air temperature (Ta) acquired with mobile sensors and LST from sharpened MODIS LST for 12 August 2019. (**a**) shows the LST values along bicycle transects, (**b**) shows Ta spatial variations, (**c**) is the sharpened MODIS LST image, and (**d**) displays the relationship between Ta and LST. The red points in (**d**) are bin averages (LST bin size 1 K) and error bars are one standard deviation.

Table S1 provides a summary of Ta and LST analysis for the 12 days when both clear sky MODIS and biking data were available. The mean slope of LST against impervious fraction was 0.11. In other words, the range of LST resulting from 0 to 100% imperious fraction is 11.30 K, implying a surface UHI intensity of similar magnitude [48]. The mean slope of Ta against impervious fraction is 0.015, suggesting an air UHI intensity of 1.54 K. Another notable feature is that the regression slope between Ta and LST becomes smaller with increasing wind speed (Figure S18a), indicating the de-coupling of surface and air temperature in strong-wind conditions. Strong wind also weakens the surface and the air UHI intensity because the slope of LST or Ta versus impervious fraction decreases with increasing wind speed (Figure S18b,c).

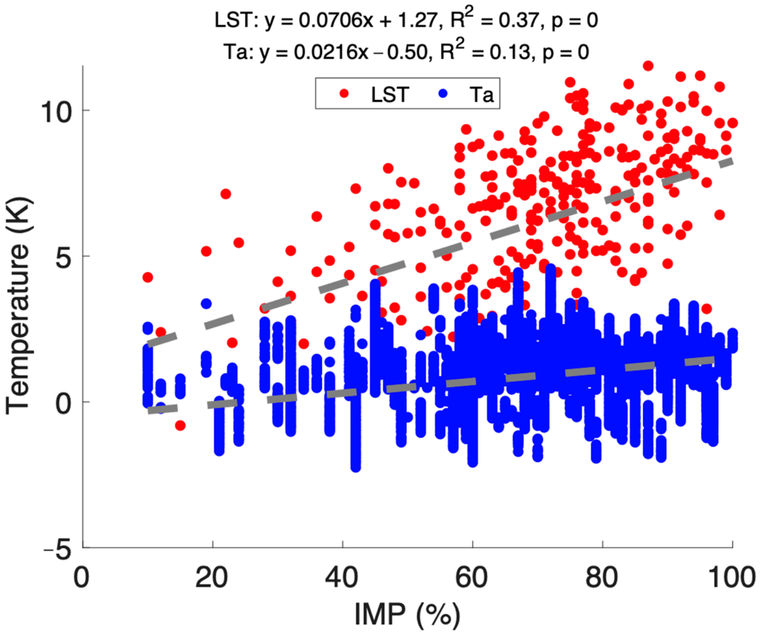


Figure 18. Dependence of air temperature (Ta) and LST on impervious surface fraction (IMP) for 12 August 2019.

5. Conclusions

In this article, we developed a fusion framework to generate new LST data with fine spatiotemporal resolution. The framework is architected on an in-depth understanding of the discrepancy between coarse and fine resolution LST at three separate scales (the scene scale, the patch scale, and the pixel scale). The inter-sensor biases at the patch scale are reduced by a model named Linear Stretching across Time (LSAT) and a neural network (NN) model. The patch-scale outputs of NN are further sharpened to pixel scale (100 m) by a model named Enrichment of Fine-resolution Variations (EOFRV). The framework is designed to improve the intra-scene variation of LST instead of the absolute temperature for advancing UHI studies.

We showed that at the patch (or MODIS pixel) scale, the spatial variations in the MODIS LST are only 36% of those in the resampled Landsat LST. The LSAT model is able to reduce these patch-level biases. The improvement is particularly noteworthy for urban areas, where the error has decreased by 1.63 K. The NN model has improved the accuracy slightly and has homogenized the error distribution over space. We showed that the LST variations within the patches are strongly controlled by the local climate. Based on this understanding, the EOFRV model sharpened the patch-level LST successfully. The overall accuracy of the sharpened MODIS LST is 0.91 K, which outperforms a bilateral filtering method especially in warm seasons. The robustness of the framework is supported by a comparison between the air temperature collected with mobile sensors and the sharpened MODIS LST along an urban street network.

Our scale-separating framework can be improved on several aspects. The remaining error mainly occurs at the patch scale. The error is dependent on season and land cover type. A model that accounts for viewing and illumination effects of ground objects may be able to further quantify and reduce the error. The EOFRV model assumes that no significant change in the landscape has taken place. However, drastic changes in landscape are possible due to human activities (e.g., quarrying, irrigation, and roof coating) and the morphological changes of aquatic organisms. Such land cover changes are found in a few fixed locations in our study domain which can be excluded from the sharpened LST product. Moreover, this assumption may not be accurate in places that have experienced rapid urbanization. Further improvements may be possible with searching for a break point in time when the landscape change took place. Finally, it is possible to refine the cloud-tolerance threshold by incorporating meteorological data so that more partly cloudy image pairs can be added to the image archive for model training.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/rs14040983/s1, Figure S1: Regression statistics for LSAT, Figure S2: EOFRV results using predicted patch-level LST, Figure S3: Training and validation losses for each target date, Figure S4: Ensemble mean training and validation losses, S5: Sharpening results for 4 May 2017, Figure S6: Correlation plot of sharpening results for 4 May 2017, Figure S7: Sharpening results for 29 July 2019, Figure S8: Correlation plot of sharpening results for 29 July 2019, Figure S9: Sharpening results for 28 November 2017, Figure S10: Correlation plot of sharpening results for 28 November 2017, Figure S11: Patch-scale and pixel-scale error for 4 May 2017, Figure S12: Patch-scale and pixel-scale error for 18 June 2016, Figure S13: Patch-scale and pixel-scale error for 30 September 2013, Figure S14: Sharpening residual of the area with significant pixel-scale error, Figure S15: Satellite map of the area with significant pixel-scale error, Figure S16: Locations where significant biases result from land cover changes, Figure S17: Residuals of pixel-scale variations in the form of histograms, Figure S18: Validation using mobile transect data on multiple dates. Table S1: Statistics of validation for each mobile transect.

Author Contributions: Conceptualization, X.L. and Y.Y.; methodology, X.L. and Y.Y.; software, Y.Y.; validation, X.L. and Y.Y.; formal analysis, Y.Y.; investigation, X.L. and Y.Y.; resources, X.L.; data curation, Y.Y.; writing—original draft preparation, Y.Y.; writing—review and editing, X.L.; visualization, Y.Y.; supervision, X.L.; project administration, X.L.; funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Robert Wood Johnson Foundation (grant 77476) and a Leitner Award for Uncommon Environmental Collaboration.

Data Availability Statement: Publicly available satellite datasets were analyzed in this study. This data can be found here: https://earthexplorer.usgs.gov (accessed date: 20 August 2021). The mobile sensor data are available on request from the corresponding author. The data are not publicly available for safety of data collectors.

Acknowledgments: The authors thank Gaboury Benoit, Eric Fine, Colleen Murphy-Dunning, and Steven Zhang for their help with bicycle air temperature measurement.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Oke, T.R. *Boundary Layer Climates*; Routledge: London, UK, 2002.
2. Lee, X.; Goulden, M.L.; Hollinger, D.Y.; Barr, A.; Black, T.A.; Bohrer, G.; Bracho, R.; Drake, B.; Goldstein, A.; Gu, L. Observed increase in local cooling effect of deforestation at higher latitudes. *Nature* **2011**, *479*, 384–387. [CrossRef] [PubMed]
3. Liu, Z.; Ballantyne, A.P.; Cooper, L.A. Biophysical feedback of global forest fires on surface temperature. *Nat. Commun.* **2019**, *10*, 214. [CrossRef] [PubMed]
4. Maimaitiyiming, M.; Ghulam, A.; Tiyip, T.; Pla, F.; Latorre-Carmona, P.; Halik, Ü.; Sawut, M.; Caetano, M. Effects of green space spatial pattern on land surface temperature: Implications for sustainable urban planning and climate change adaptation. *ISPRS J. Photogramm. Remote Sens.* **2014**, *89*, 59–66. [CrossRef]
5. Meehl, G.A. Influence of the land surface in the asian summer monsoon: External conditions versus internal feedbacks. *J. Clim.* **1994**, *7*, 1033–1049. [CrossRef]
6. Bastiaanssen, W.G.; Menenti, M.; Feddes, R.; Holtslag, A. A remote sensing surface energy balance algorithm for land (sebal). 1. Formulation. *J. Hydrol.* **1998**, *212*, 198–212. [CrossRef]
7. Jackson, R.; Reginato, R.; Idso, S. Wheat canopy temperature: A practical tool for evaluating water requirements. *Water Resour. Res.* **1977**, *13*, 651–656. [CrossRef]
8. Chuvieco, E.; Cocero, D.; Riano, D.; Martin, P.; Martınez-Vega, J.; de la Riva, J.; Pérez, F. Combining ndvi and surface temperature for the estimation of live fuel moisture content in forest fire danger rating. *Remote Sens. Environ.* **2004**, *92*, 322–331. [CrossRef]
9. Guangmeng, G.; Mei, Z. Using modis land surface temperature to evaluate forest fire risk of northeast china. *IEEE Geosci. Remote Sens. Lett.* **2004**, *1*, 98–100. [CrossRef]
10. Schmugge, T.J.; André, J.-C. *Land Surface Evaporation: Measurement and Parameterization*; Springer Science & Business Media: Berlin, Germany, 2012.
11. Zink, M.; Mai, J.; Cuntz, M.; Samaniego, L. Conditioning a hydrologic model using patterns of remotely sensed land surface temperature. *Water Resour. Res.* **2018**, *54*, 2976–2998. [CrossRef]

12. Raynolds, M.K.; Comiso, J.C.; Walker, D.A.; Verbyla, D. Relationship between satellite-derived land surface temperatures, arctic vegetation types, and ndvi. *Remote Sens. Environ.* **2008**, *112*, 1884–1894. [CrossRef]
13. Sims, D.A.; Rahman, A.F.; Cordova, V.D.; El-Masri, B.Z.; Baldocchi, D.D.; Bolstad, P.V.; Flanagan, L.B.; Goldstein, A.H.; Hollinger, D.Y.; Misson, L. A new model of gross primary productivity for north american ecosystems based solely on the enhanced vegetation index and land surface temperature from modis. *Remote Sens. Environ.* **2008**, *112*, 1633–1646. [CrossRef]
14. Connors, J.P.; Galletti, C.S.; Chow, W.T. Landscape configuration and urban heat island effects: Assessing the relationship between landscape characteristics and land surface temperature in phoenix, arizona. *Landsc. Ecol.* **2013**, *28*, 271–283. [CrossRef]
15. Kumar, K.S.; Bhaskar, P.U.; Padmakumari, K. Estimation of land surface temperature to study urban heat island effect using landsat etm+ image. *Int. J. Eng. Sci. Technol.* **2012**, *4*, 771–778.
16. Zhao, L.; Lee, X.; Smith, R.B.; Oleson, K. Strong contributions of local background climate to urban heat islands. *Nature* **2014**, *511*, 216–219. [CrossRef]
17. Yin, Z.; Wu, P.; Foody, G.M.; Wu, Y.; Liu, Z.; Du, Y.; Ling, F. Spatiotemporal fusion of land surface temperature based on a convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 1808–1822. [CrossRef]
18. Song, H.; Liu, Q.; Wang, G.; Hang, R.; Huang, B. Spatiotemporal satellite image fusion using deep convolutional neural networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 821–829. [CrossRef]
19. Zhan, W.; Chen, Y.; Zhou, J.; Wang, J.; Liu, W.; Voogt, J.; Zhu, X.; Quan, J.; Li, J. Disaggregation of remotely sensed land surface temperature: Literature survey, taxonomy, issues, and caveats. *Remote Sens. Environ.* **2013**, *131*, 119–139. [CrossRef]
20. Zhan, W.; Chen, Y.; Zhou, J.; Li, J.; Liu, W. Sharpening thermal imageries: A generalized theoretical framework from an assimilation perspective. *IEEE Trans. Geosci. Remote Sens.* **2010**, *49*, 773–789. [CrossRef]
21. Kustas, W.P.; Norman, J.M.; Anderson, M.C.; French, A.N. Estimating subpixel surface temperatures and energy fluxes from the vegetation index–radiometric temperature relationship. *Remote Sens. Environ.* **2003**, *85*, 429–440. [CrossRef]
22. Agam, N.; Kustas, W.P.; Anderson, M.C.; Li, F.; Neale, C.M. A vegetation index based technique for spatial sharpening of thermal imagery. *Remote Sens. Environ.* **2007**, *107*, 545–558. [CrossRef]
23. Essa, W.; Verbeiren, B.; van der Kwast, J.; Van de Voorde, T.; Batelaan, O. Evaluation of the distrad thermal sharpening methodology for urban areas. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *19*, 163–172. [CrossRef]
24. Pan, X.; Zhu, X.; Yang, Y.; Cao, C.; Zhang, X.; Shan, L. Applicability of downscaling land surface temperature by using normalized difference sand index. *Sci. Rep.* **2018**, *8*, 9530. [CrossRef] [PubMed]
25. Gao, F.; Masek, J.; Schwaller, M.; Hall, F. On the blending of the landsat and modis surface reflectance: Predicting daily landsat surface reflectance. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2207–2218.
26. Zhu, X.; Chen, J.; Gao, F.; Chen, X.; Masek, J.G. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sens. Environ.* **2010**, *114*, 2610–2623. [CrossRef]
27. Huang, B.; Wang, J.; Song, H.; Fu, D.; Wong, K. Generating high spatiotemporal resolution land surface temperature for urban heat island monitoring. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1011–1015. [CrossRef]
28. Wu, P.; Shen, H.; Zhang, L.; Göttsche, F.-M. Integrated fusion of multi-scale polar-orbiting and geostationary satellite observations for the mapping of high spatial and temporal resolution land surface temperature. *Remote Sens. Environ.* **2015**, *156*, 169–181. [CrossRef]
29. Quan, J.; Zhan, W.; Ma, T.; Du, Y.; Guo, Z.; Qin, B. An integrated model for generating hourly landsat-like land surface temperatures over heterogeneous landscapes. *Remote Sens. Environ.* **2018**, *206*, 403–423. [CrossRef]
30. Quan, J.; Chen, Y.; Zhan, W.; Wang, J.; Voogt, J.; Li, J. A hybrid method combining neighborhood information from satellite data with modeled diurnal temperature cycles over consecutive days. *Remote Sens. Environ.* **2014**, *155*, 257–274. [CrossRef]
31. Quan, J.; Zhan, W.; Chen, Y.; Wang, M.; Wang, J. Time series decomposition of remotely sensed land surface temperature and investigation of trends and seasonal variations in surface urban heat islands. *J. Geophys. Res. Atmos.* **2016**, *121*, 2638–2657. [CrossRef]
32. Cook, M.; Schott, J.R.; Mandel, J.; Raqueno, N. Development of an operational calibration methodology for the landsat thermal data archive and initial testing of the atmospheric compensation component of a land surface temperature (lst) product from the archive. *Remote Sens.* **2014**, *6*, 11244–11266. [CrossRef]
33. Cook, M.J. *Atmospheric Compensation for a Landsat Land Surface Temperature Product*; Rochester Institute of Technology: New York, NY, USA, 2014.
34. Malakar, N.K.; Hulley, G.C.; Hook, S.J.; Laraby, K.; Cook, M.; Schott, J.R. An operational land surface temperature product for landsat thermal data: Methodology and validation. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5717–5735. [CrossRef]
35. Snyder, W.C.; Wan, Z.; Zhang, Y.; Feng, Y.Z. Classification-based emissivity for land surface temperature measurement from space. *Int. J. Remote Sens.* **1998**, *19*, 2753–2774. [CrossRef]
36. Chakraborty, T.C.; Lee, X.; Ermida, S.; Zhan, W. On the land emissivity assumption and landsat-derived surface urban heat islands: A global analysis. *Remote Sens. Environ.* **2021**, *265*, 112682. [CrossRef]
37. Wan, Z. *Modis Land Surface Temperature Products Users' Guide*; Institute for Computational Earth System Science, University of California: Santa Barbara, CA, USA, 2006.
38. Hu, L.; Monaghan, A.; Voogt, J.A.; Barlage, M. A first satellite-based observational assessment of urban thermal anisotropy. *Remote Sens. Environ.* **2016**, *181*, 111–121. [CrossRef]

39. Krayenhoff, E.S.; Voogt, J.A. Daytime thermal anisotropy of urban neighbourhoods: Morphological causation. *Remote Sens.* **2016**, *8*, 108. [CrossRef]
40. Cao, C.; Yang, Y.; Lu, Y.; Schultze, N.; Gu, P.; Zhou, Q.; Xu, J.; Lee, X. Performance evaluation of a smart mobile air temperature and humidity sensor for characterizing intracity thermal environment. *J. Atmos. Ocean. Technol.* **2020**, *37*, 1891–1905. [CrossRef]
41. Li, H.; Wolter, M.; Wang, X.; Sodoudi, S. Impact of land cover data on the simulation of urban heat island for berlin using wrf coupled with bulk approach of noah-lsm. *Theor. Appl. Climatol.* **2018**, *134*, 67–81. [CrossRef]
42. Li, H.; Zhou, Y.; Wang, X.; Zhou, X.; Zhang, H.; Sodoudi, S. Quantifying urban heat island intensity and its physical mechanism using wrf/ucm. *Sci. Total Environ.* **2019**, *650*, 3110–3119. [CrossRef]
43. Yuan, F.; Bauer, M.E. Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in landsat imagery. *Remote Sens. Environ.* **2007**, *106*, 375–386. [CrossRef]
44. Ziter, C.D.; Pedersen, E.J.; Kucharik, C.J.; Turner, M.G. Scale-dependent interactions between tree canopy cover and impervious surfaces reduce daytime urban heat during summer. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 7575–7580. [CrossRef]
45. Li, D.; Liao, W.; Rigden, A.J.; Liu, X.; Wang, D.; Malyshev, S.; Shevliakiva, E. Urban heat island: Aerodynamics or imperviousness? *Sci. Adv.* **2019**, *5*, eaau4299. [CrossRef]
46. Yap, D.; Oke, T.R. Sensible heat fluxes over an urban area—Vancouver, BC. *J. Appl. Meteorol. Climatol.* **1974**, *13*, 880–890. [CrossRef]
47. Kato, S.; Yamaguchi, Y. Analysis of urban heat-island effect using ASTER and ETM+ Data: Separation of anthropogenic heat discharge and natural heat radiation from sensible heat flux. *Remote Sens. Environ.* **2005**, *99*, 44–54. [CrossRef]
48. Li, H.; Zhou, Y.; Li, X.; Meng, L.; Wang, X.; Wu, S.; Sodoudi, S. A new method to quantify surface urban heat island intensity. *Sci. Total Environ.* **2018**, *624*, 262–272. [CrossRef] [PubMed]