Modifying an Image Fusion Approach for High Spatiotemporal LST Retrieval in Surface Dryness and Evapotranspiration Estimations

Tri Wandi Januar 1, Tang-Huang Lin 1,*, Chih-Yuan Huang 1,2 and Kuo-En Chang 3

1 Center for Space and Remote Sensing Research, National Central University, Taoyuan City 32001, Taiwan; januar.triwandi@g.ncu.edu.tw (T.W.J.); cyhuang@csrsr.ncu.edu.tw (C.-Y.H.)
2 Department of Civil Engineering, National Central University, Taoyuan City 32001, Taiwan
3 Graduate Institute of Space Science, National Central University, Taoyuan City 32001, Taiwan; 101683007@cc.ncu.edu.tw
* Correspondence: thlin@csrsr.ncu.edu.tw; Tel.: +886-03-425-7232

Received: 22 December 2019; Accepted: 3 February 2020; Published: 4 February 2020

Abstract: Thermal infrared (TIR) satellite images are generally employed to retrieve land surface temperature (LST) data in remote sensing. LST data have been widely used in evapotranspiration (ET) estimation based on satellite observations over broad regions, as well as the surface dryness associated with vegetation index. Landsat-8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) can provide LST data with a 30-m spatial resolution. However, rapid changes in environmental factors, such as temperature, humidity, wind speed, and soil moisture, will affect the dynamics of ET. Therefore, ET estimation needs a high temporal resolution as well as a high spatial resolution for daily, diurnal, or even hourly analysis. A challenge with satellite observations is that higher-spatial-resolution sensors have a lower temporal resolution, and vice versa. Previous studies solved this limitation by developing a spatial and temporal adaptive reflectance fusion model (STARFM) for visible images. In this study, with the primary mechanism (thermal emission) of TIRS, surface emissivity is used in the proposed spatial and temporal adaptive emissivity fusion model (STAEFM) as a modification of the original STARFM for fusing TIR images instead of reflectance. For high a temporal resolution, the advanced Himawari imager (AHI) onboard the Himawari-8 satellite is explored. Thus, Landsat-like TIR images with a 10-minute temporal resolution can be synthesized by fusing TIR images of Himawari-8 AHI and Landsat-8 TIRS. The performance of the STAEFM to retrieve LST was compared with the STARFM and enhanced STARFM (ESTARFM) based on the similarity to the observed Landsat image and differences with air temperature. The peak signal-to-noise ratio (PSNR) value of the STAEFM image is more than 42 dB, while the values for STARFM and ESTARFM images are around 31 and 38 dB, respectively. The differences of LST and air temperature data collected from five meteorological stations are 1.53 °C to 4.93 °C, which are smaller compared with STARFM’s and ESATRFM’s. The examination of the case study showed reasonable results of hourly LST, dryness index, and ET retrieval, indicating significant potential for the proposed STAEFM to provide very-high-spatiotemporal-resolution (30 m every 10 min) TIR images for surface dryness and ET monitoring.

Keywords: Landsat-8 OLI and Himawari-8 AHI; TIR image fusion; thermal emission; high spatiotemporal retrieval; LST; surface dryness and evapotranspiration

1. Introduction

Soil moisture and vegetation are key parameters in land–atmosphere interactions [1,2]. As the sum of soil evaporation and canopy transpiration, evapotranspiration (ET) plays a significant role in

determining the exchanges of energy and mass between the hydrosphere, atmosphere, and biosphere [3]. Therefore, ET monitoring is important in various disciplines, such as land surface modelling, global circulation models, irrigation systems, hydrology, and water resources management [4].

Actual ET can be measured directly using lysimeters [5] and eddy covariance flux towers [6] at the individual plant and local scale, respectively. However, since an in situ measurement station is limited in spatial coverage and considered relatively expensive, remote sensing data have been used for ET estimation on the local, regional, or even global scale. Hence, ET estimation using remote sensing data is important and has been increasingly studied in order to expand the information that an in situ measurement station cannot provide.

Land surface temperature (LST) is an important indicator to assess actual evaporation and surface moisture status related to ET and water availability [7,8]. Previous studies used various remote sensing algorithms of ET estimation, such as the surface energy balance algorithm for land (SEBAL) [9,10], mapping evapotranspiration at high resolution and with internalized calibration (METRIC) [11,12], and surface energy balance system (SEBS) [13,14]. However, satellite-based LST data are generally limited in temporal resolution, while in situ meteorological data can be provided at up to an hourly scale. On the other hand, meteorological factors such as net radiation, temperature, humidity, and wind speed have significant impacts on ET variability, in that all factors influence transpiration and evaporation [15]. Therefore, rapid changes in meteorological factors affect the dynamics of ET, and a high temporal resolution is needed in ET estimation using remote sensing data for daily, diurnal, or even hourly analysis.

LST retrieval methods based on satellite observations are continuously being improved in terms of both accuracy and resolution [16,17]. Landsat-8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) can provide LST data at a 30-m spatial resolution that have been widely used. On the other hand, Landsat-8 only provides data every 16 days, which is a limitation in temporal resolution, as mentioned above. Previous studies employed Himawari-8 data to develop LST retrieval algorithm of non-linear three bands (NTB) [18] and LST retrieval algorithm based on water vapor content and the time of day [19]. The studies achieved LST retrieval with an ultra-high temporal resolution that allows observation every 10 min and even 2.5 min in the area of Japan by employing a Himawari-8 advanced Himawari imager (AHI). However, the spatial resolution of LST obtained from Himawari-8 data is 2 km. Thus, it has been a challenge that high-spatial-resolution satellite sensors have a limitation in temporal resolution, and vice versa.

The spatial and temporal adaptive reflectance fusion model (STARFM) [20] was developed to solve the limitations of remote sensing by fusing Moderate Resolution Imaging Spectroradiometer (MODIS) and Landsat images. The algorithm was proposed to synthesize daily surface reflectance with a 30-m spatial resolution. The main idea is that STARFM predicts fine-resolution images at prediction times based on the weightings of fine-resolution and coarse-resolution images at reference times. Since the performance of the original STARFM in heterogeneous landscapes was considered unsatisfactory, the algorithm was modified and extended as the enhanced spatial and temporal adaptive reflectance fusion model (ESTARFM) [21]. As an enhancement of the original algorithm, ESTARFM employs two pairs of images of two reference times and adds a conversion coefficient to better predict fine-resolution images.

The purpose of this study is to hourly estimate actual evapotranspiration based on a combination of remote sensing and meteorological data. However, remote sensing data in terms of satellite-based LST are not provided at a high spatial and temporal resolution, while in situ meteorological data can be provided at up to an hourly scale. To solve the limitation, STARFM can be applied in LST retrieval with a high spatial and temporal resolution. However, the original STARFM was commonly used to predict surface reflectance of visible spectral bands. TIR images from the same satellite are generally coarser in spatial resolution compared with visible images; for instance, the spatial resolution of Landsat TIR bands is originally 100 m before being sharpened to 30 m as the spatial resolution of visible spectral bands, whereas MODIS LST products are provided in a 1-km spatial resolution while visible surface
reflectance images are provided in 250–500 m, and the Himawari-8 advanced Himawari imager (AHI) has a 2-km and 0.5–1-km spatial resolution for TIR and visible spectral bands, respectively.

Finding spectrally similar neighboring pixels based on a classification map generated from fine-resolution images at a reference time is crucial for the spatiotemporal image fusion method, because coarse-resolution pixels might contain multiple land-cover types. Due to the dynamics of Earth’s radiation, bias could happen if similar pixels are found based on a classification map generated from TIR images. According to the solar energy balance description provided by Schneider [22], 45 units of solar energy are absorbed by the Earth’s surface for every 100 units of incoming solar radiation falling on the Earth and its atmosphere. To remain stable, some energy is radiated to space by the Earth as a part of the longwave portion, which is the outgoing infrared radiation. This condition causes the LSTs to vary during day and night hours. Therefore, the modification of the image fusion approach is significant for better predicting TIR bands.

Himawari-8 is a geostationary satellite that was launched in 2014 equipped with a 10-min temporal resolution. As an advantage among other satellites, in terms of ultra-high temporal resolutions, Himawari-8 TIR images can be fused with Landsat-8 TIR images to synthetize Landsat-like TIR images every 10 min. A previous study modified the original STARFM as the spatial and temporal adaptive reflectance fusion model—aerosol property (STARFM-AP) to predict the top-of-atmosphere (TOA) reflectance of visible bands with a 30-m spatial resolution and 10-min temporal resolution by utilizing Landsat-8 and Himawari-8 [23]. In this study, a modification of the original STARFM to fuse Landsat-8 and Himawari-8 TIR bands includes three stages. First, bias coming from the large difference in the spatial resolution of Himawari-8 and Landsat-8 TIR images is reduced by sharpening the coarse images, which are Himawari-8. Second, land surface emissivity (LSE) is used to find spectrally similar neighboring pixel. Third, the shortwave infrared (SWIR) spectral band is used to preserve surface reflectivity and reduce the atmospheric water vapor effect adopted from STARFM-AP algorithm [23]. Finally, we can produce LST data with a 30-m spatial resolution and temporal resolution up to 10 min and have an opportunity to combine hourly in situ meteorological data and hourly fused LST images to estimate hourly ET.

2. Literature Review

Previous studies enhanced LST retrieval technique by employing visible and NIR spectral bands to sharpen the spatial resolution of LST data [24–26]. The relationship between vegetation indices and LST was explored to define the disaggregation procedures of LST estimation. Sequential Bayesian methods were used to enhance TIR images by fusing two image sequences that were evaluated by application of different datasets from SEVIRI (Spinning Enhanced Visible and Infrared Imager) and MODIS [27]. Another study combined the fusion method and regression method to generate high spatiotemporal LST with a comparison of “regression-the-fusion” and “fusion-then regression” algorithms [28]. The results show that the performing regression method followed by the fusion method can generate high spatiotemporal LST with less errors.

The estimation of actual ET with a high spatial and temporal resolution generally uses two or more datasets in order to get the advantages in terms of resolution for each dataset. For example, geostationary and polar orbiting satellite images were used for ET mapping on the field to continental scale [8]. A surface energy balance model driven by TIR remote sensing was used to obtain subsurface moisture conditions based on the information of LST and normalized difference vegetation index (NDVI). A multisensor TIR approach, the atmosphere–land exchange inverse (ALEXI) model, was used to estimate daily ET at the continental scale using geostationary satellites with a 5–10-km spatial resolution. The ALEXI model was spatially disaggregated by the DisALEXI algorithm using moderate-resolution TIR images from polar orbiting satellites.

Another high-spatiotemporal ET mapping study introduced a novel approach to fusing characteristics of daily MODIS and biweekly Landsat LST images to provide optimal spatiotemporal coverage [29]. Daily ET maps were generated with a surface energy balance model using geostationary
satellite data that were spatially disaggregated using daily MODIS with a 1-km spatial resolution and biweekly Landsat LST images sharpened to 30 m.

The region-based and pixel-based disaggregation methods were compared to actual ET estimation [30]. Actual ET data with a 1-km resolution images were disaggregated to a 250 and 30-m resolution in two steps using MODIS and Landsat-5. The results show that the region-based disaggregation method using NDVI was in good agreement with the Landsat-5 actual ET.

Daily ET at a 30-m spatial resolution was computed by employing STARFM combined with a multiscale ET retrieval algorithm based on the two-source energy balance (TSEB) land–surface representation [31]. The TSEB was run using Geostationary Environmental Operational Satellites (GEOS) TIR images with a 4-km spatial resolution and hourly temporal sampling, daily MODIS LST product with a 1-km spatial resolution, and Landsat-8 TIR images with a 30-m spatial resolution and a 16-day temporal resolution.

The MODIS 8-day 1 km (ET) downscaling method based on Landsat-8 data was adopted, using a machine learning approach with 11 indicators, including albedo, LST, and vegetation indices (Vis) derived from Landsat-8 data, upscaled to a 1-km resolution [32]. The relationship between Landsat indicators and MODIS 8-day 1 km ET was developed by using support vector regression (SVR), cubist, and random forest (RF) algorithms. Finally, the models were used to predict ET with 30-m spatial resolution based on Landsat-8 indicators.

A multisensor ET data fusion system was applied to a mixed forest/agriculture landscape in North Carolina, USA [33]. LST data from multiple satellite platforms (hourly geostationary satellite data at a 4-km resolution, daily 1-km MODIS, and biweekly Landsat TIR sharpened to 30 m) were retrieved by applying TSEB to estimate ET. The STARFM was applied to the multiple ET data streams to estimate daily ET at a 30-m resolution.

3. Study Area and Materials

Chiayi County, located in the central part of Taiwan, was chosen as the study area. The study area is located between 120°09′05″E–120°28′30″E and 23°17′45″N–23°35′45″N. The area is mostly covered by agricultural land, along with built-up areas, bare land, vegetation in mountainous areas, and water bodies, as shown in Figure 1.

Figure 1. The study area is a 1063 km² (120°09′05″E–120°28′30″E, 23°17′45″N–23°35′45″N) plot in the Chiayi County, Taiwan, mostly covered by agricultural land. The black circles indicate locations of weather stations A (23°34′9.13″N, 120°17′13.2″E), B (23°33′12.6″N, 120°25′13.08″E), C (23°24′47.16″N, 120°18′0.72″E), D (23°20′57.12″N, 120°24′22.68″E), and D (23°18′44.64″N, 120°18′30.96″E).
Clear-sky images from Landsat-8 and Himawari-8 at 3 acquisition dates were collected for the image fusion model and the assessment of results. Reference dates for fusing TIR images in this study are 30 October and 15 November 2018, and the prediction date is 18 January 2019, as shown in Table 1. Since Himawari-8 images are provided every 10 min, their recording time must be the same as the Landsat-8 recording time in the study area. Besides remote sensing data, we collected hourly meteorological data on 18 January 2019 from 17 weather stations for hourly actual ET estimation.

Table 1. Datasets collected for thermal infrared (TIR) image fusion methods. H8, Himawari-8; L8, Landsat-8; STARFM, spatial and temporal adaptive reflectance fusion model; ESTARFM, enhanced STARFM; STAEFM, spatial and temporal adaptive emissivity fusion model; images used for image fusion method are marked as the symbol “✓”.

<table>
<thead>
<tr>
<th>Specific Bands (Central Wavelength)</th>
<th>Acquisition Dates</th>
<th>Image Fusion Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>STARFM</td>
</tr>
<tr>
<td>H8-B13 (10.41 µm)</td>
<td>30 October 2018</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>15 November 2018</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>18 January 2019</td>
<td>✓</td>
</tr>
<tr>
<td>H8-B05 (1.61 µm)</td>
<td>30 October 2018</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>15 November 2018</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>18 January 2019</td>
<td>✓</td>
</tr>
<tr>
<td>L8-B10 (10.9 µm)</td>
<td>30 October 2018</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>15 November 2018</td>
<td>✓</td>
</tr>
<tr>
<td>L8-B06 (1.61 µm)</td>
<td>30 October 2018</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>15 November 2018</td>
<td>✓</td>
</tr>
</tbody>
</table>

As an important idea in spatiotemporal image fusion methods, every coarse-resolution image must be resampled to the pixel size of the fine-resolution image and georeferenced to its bounds so that they are assumed to be the same in image size, pixel size, and coordinate system. In this study, the bilinear interpolation method, which is available to the continuous data type such as satellite image, was used to resample Himawari-8 images with a 2-km spatial resolution to 30 m, corresponding to the Landsat-8 spatial resolution. Thereafter, image registration was applied to Himawari-8 and Landsat-8 images before the image fusion procedure. In addition, since meteorological data are provided as points, a spatial interpolation technique was used to generate air temperature, air humidity, and wind speed maps.

4. Methodology

4.1. Spatiotemporal Image Fusion Methods

In this study, STAEFM, as an improvement of the original STARFM, is proposed to fuse TIR images. The performance of STAEFM in fusing Landsat-8 and Himawari-8 TIR images was compared to the original STARFM and ESTARFM algorithms. We also employed the ESTARFM algorithm to fuse Landsat-8 OLI and Himawari-8 AHI images to estimate net radiation and soil heat flux density as input data in ET estimation, which will be explained in the next section. Prediction time in these spatiotemporal image fusion methods is denoted as \( t_1 \), and the first and second reference times are denoted as \( t_R \) and \( t_{RA} \), respectively.

4.1.1. STARFM

The STARFM algorithm has been widely used to predict surface reflectance of visible spectral bands. The algorithm employs one pair of images, consisting of a fine-resolution and a coarse-resolution image, at reference time \( t_R \) and another coarse-resolution image at prediction time \( t_1 \) to predict the fine-resolution image at \( t_1 \). Finding spectrally similar neighboring pixels is crucial in the spatiotemporal...
image fusion method, because coarse-resolution pixels might contain multiple land-cover types. In order to efficiently select neighboring pixels of the same land-cover type, unsupervised classification must be applied to the fine-resolution image at $t_R$ before data blending, and pixels of the same class are considered spectrally similar [20].

Under the assumption that differences in fine resolution and coarse resolution for each pixel at $t_R$ and $t_1$ are equal, a pixel of predicted fine-resolution image at $t_1$ can be defined as shown in Equation (1):

$$F(x_i, y_j, t_1) = C(x_i, y_j, t_1) + F(x_i, y_j, t_R) - C(x_i, y_j, t_R)$$

where $x_i$ and $y_i$ are the given pixel location and $F$ and $C$ stand for fine-resolution and coarse-resolution images, respectively.

Since a pixel in a coarse-resolution image might contain multiple land-cover types and Equation (1) is basically used for the ideal situation, which is homogeneous regions as assumed above, a search window size $w$ is applied to find spectrally similar neighboring pixels, as presented in Equation (2):

$$F(x_{w/2}, y_{w/2}, t_1) = \sum_{i}^{w} \sum_{j}^{w} W_{ij} \times [C(x_i, y_j, t_1) + F(x_i, y_j, t_R) - C(x_i, y_j, t_R)]$$

where $(x_{w/2}, y_{w/2})$ is the central pixel within the moving window and $W_{ij}$ is a combined weighting based on spectral, temporal, and distance information. Combined weightings in this algorithm can be calculated by using Equation (3):

$$W_{ij} = \left(\frac{1}{C_{ij}}\right) / \sum_{i}^{w} \sum_{j}^{w} \left(\frac{1}{C_{ij}}\right)$$

whereas $C_{ij}$ is defined by Equation (4):

$$C_{ij} = S_{ij} \times T_{ij} \times d_{ij}$$

where $S_{ij}$, $T_{ij}$, and $d_{ij}$ are weightings of spectral, temporal, and distance, which are explained in the following section.

Spectral Weighting

Neighboring pixels of the same land-cover type are considered spectrally similar and selected for spectral weighting calculation. For each central pixel within a search window, a spectral weighting between coarse-resolution and fine-resolution images at reference time $t_R$ can be defined by Equation (5):

$$S_{ij} = \left|F(x_i, y_j, t_R) - C(x_i, y_j, t_R)\right|$$

A larger difference is defined as a smaller value of $S_{ij}$, which means the features of fine-resolution pixels are spectrally closer to the surrounding pixels on average. Under a further assumption that coarse-resolution and fine-resolution images are consistently comparable spatially and temporally, the $S_{ij}$ values for reference time $t_R$ are equal for prediction time $t_1$.

Temporal Weighting

Land-cover change and noise, including cloud cover, are identified by smaller temporal weight. The changes between coarse-resolution images at $t_R$ and $t_1$ are defined by Equation (6):

$$T_{ij} = \left|C(x_i, y_j, t_R) - C(x_i, y_j, t_1)\right|$$
A small value of \( T_{ij} \) defines less change between \( t_R \) and \( t_1 \), which means the temporal weighting of the given pixel location is large.

Distance Weighting

The spatial distance between central and neighboring pixels can be measured by using Equation (7):

\[
d_{ij} = \sqrt{(x_{w/2} - x_i)^2 + (y_{w/2} - y_j)^2}
\]  

(7)

Closer neighboring pixels normally have better similarity and contribute more to the central pixel, so the weight is higher.

4.1.2. ESTARFM

The ESTARFM algorithm predicts fine-resolution images based on the differences of fine- and coarse-resolution images at two references times, \( t_R \) and \( t_{RA} \). The algorithm minimizes the system biases coming from differences in sensor systems by using correlation to blend multisource data that differ in orbit parameters and spectral response function. In addition, the conversion coefficient and some adjustments of weighting calculation were introduced to enhance the accuracy of prediction for heterogeneous landscapes [21].

Under the condition that a coarse-resolution pixel is only covered by one land-cover type, the relationship between resampled coarse-resolution and fine-resolution pixels at \( t_1 \) can be linearly modeled as Equation (8):

\[
F(x_i, y_j, t_1) = a \times C(x_i, y_j, t_1) + b
\]  

(8)

where \( a \) and \( b \) are coefficients of the linear regression. Accordingly, we can apply Equation (8) to another pair of images at \( t_R \) under the same condition, which means coefficients \( a \) and \( b \) are stable at a given location. Hence, the relationship between the two pairs of images at \( t_1 \) and \( t_R \) can be expressed as Equation (9):

\[
F(x_i, y_j, t_1) = F(x_i, y_j, t_R) + a \times (C(x_i, y_j, t_1) - C(x_i, y_j, t_R))
\]  

(9)

However, coefficient \( a \) in Equation (9) is only stable over nonchanging surfaces such as deserts and water bodies without system biases and after performing atmospheric correction. Considering that the pixels in coarse-resolution images are mostly mixed, a conversion coefficient is used to describe the relationship between two pairs of images at \( t_1 \) and \( t_R \) as Equation (10):

\[
F(x_i, y_j, t_1) = F(x_i, y_j, t_R) + v_i(x_i, y_j) \times (C(x_i, y_j, t_1) - C(x_i, y_j, t_R))
\]  

(10)

The conversion coefficient \( v_i \) indicates the ratio of the change of reflectance for the \( i \)th end member to the change of reflectance for a mixed coarse-resolution pixel (refer to [21] for more details).

The ESTARFM considers spectrally similar neighboring pixels for weighting calculation, similar to the STARFM algorithm. Hence, Equation (10) is extended as Equation (11), which applies a search window size and weighting calculation:

\[
F_R(x_{w/2}, y_{w/2}, t_1) = F(x_{w/2}, y_{w/2}, t_R) + \sum_{i=1}^{N} W_i \times v_i \times \left[C(x_i, y_j, t_1) - C(x_i, y_j, t_R)\right]
\]  

(11)

where \( F_R(x_{w/2}, y_{w/2}, t_1) \) refers to the fine-resolution central pixel within a moving window at \( t_1 \), which is predicted based on information from \( t_R \) images, \( W_i \) is the weighting for each pixel, and \( v_i \) is the conversion coefficient of similar pixels.

As mentioned above, this algorithm uses the information of two reference times to predict a more accurate fine-resolution image at \( t_1 \). Thus, Equation (11) is applied to another reference time \( t_{RA} \) and

Remote Sens. 2020, 12, 498
the final prediction pixel can be estimated by combining two prediction pixels based on \( t_R \) and \( t_{RA} \) as Equation (12):

\[
F(x_{w/2}, y_{w/2}, t_1) = T_R 	imes F_R(x_{w/2}, y_{w/2}, t_1) + T_{RA} 	imes F_{RA}(x_{w/2}, y_{w/2}, t_1)
\]  
(12)

\( T_R \) and \( T_{RA} \) in Equation (12) are temporal weightings corresponding to reference times \( t_R \) and \( t_{RA} \). The calculation of the temporal weightings is described in Equation (13):

\[
T_k = \frac{1 / \sum_{j}^{n} \sum_{l}^{w} C(x_j, y_l, t_k) - \sum_{j}^{n} \sum_{l}^{w} C(x_j, y_l, t_1)}{\sum_k (1 / \sum_{j}^{n} \sum_{l}^{w} C(x_j, y_l, t_k) - \sum_{j}^{n} \sum_{l}^{w} C(x_j, y_l, t_1))}, \quad (k = R, RA)
\]  
(13)

Finally, the combined weightings based on the two reference times \( t_R \) and \( t_{RA} \) can obtain a more accurate prediction at \( t_1 \).

### 4.1.3. STAEFM

The STARFM and ESTARFM algorithms were proposed for surface reflectance of visible bands, as mentioned earlier. Since the purpose of this study is to fuse Landsat-8 and Himawari-8 TIR bands, there might be uncertainty when using the original STARFM to fuse the TIR bands, caused by large differences in spatial resolution, spectral response, and bias coming from atmospheric water vapor. Thus, an improvement of the original STARFM is undertaken in this study to solve the problem.

STAEFM for fusing Landsat-8 and Himawari-8 TIR bands in this study consists of three stages. Figure 2 shows the workflow of the proposed STAEFM. First, a sharpening method is applied to Himawari-8 TIR images before data blending to minimize the bias coming from large differences in spatial resolution between the Landsat-8 and Himawari-8 TIR bands. Second, since the TIR band is based on longwave radiation, land surface emissivity (LSE) is used to find spectrally similar neighboring pixels. Third, the atmospheric water vapor effect is reduced by using shortwave infrared (SWIR) bands in the weighting calculation.

![Figure 2. Flowchart of STAEFM approach. NIR, near infrared; SWIR, shortwave infrared; TIR, thermal infrared; mNDVI, modified normalized difference vegetation index.](image-url)
Sharpening Himawari-8 TIR Images

Himawari-8 TIR images have a 2-km spatial resolution, which means the difference in spatial resolution is quite large compared to Landsat-8 TIR images, which have a 30-m spatial resolution after being sharpened from 100 m. However, the spatial resolution of Himawari-8 TIR images can be enhanced by employing a sharpening method to minimize the gap between the two images. A previous study employed MODIS visible and near infrared (VNIR) band data to increase the spatial resolution of LST images from MODIS TIR band data [34]. Landsat-8 TIR images were successfully sharpened using the red band [35]. Thus, we have an opportunity to enhance the spatial resolution of Himawari-8 TIR images.

In this study, Himawari-8 red images with a 500-m spatial resolution were used for TIR sharpening. Figure 3 shows the Himawari-8 TIR images with and without sharpening compared with Landsat-8 TIR images at the same acquisition time. The Himawari-8 TIR images were previously resampled to 30 m, which is the pixel size of Landsat-8 images. The spatial structure of sharpened Himawari-8 TIR images is much closer to the Landsat-8 TIR images, because the spatial resolution is already disaggregated from 2 km to 500 m, which is the spatial resolution of Himawari-8 red images.

![Figure 3. Digital number of Himawari-8 TIR images (a) without sharpening and (b) with sharpening, compared to (c) the digital number of the Landsat-8 TIR image at the same acquisition time (15 November 2018).](image)

Generating a Classification Map Based on Land Surface Emissivity

Unsupervised classification must be applied to the fine-resolution image at $t_R$ to find spectrally similar neighboring pixels in the spatiotemporal image fusion method, as mentioned earlier. If we follow the original STARFM to generate a classification map, Landsat-8 TIR at $t_R$ should be used for classification. However, there may be bias if the TIR image is used for classification due to the dynamics of Earth’s radiation. The visible spectral bands depend on reflections of objects illuminated by the sun as the Earth’s albedo, which is the shortwave portion of the energy budget. In contrast to visible bands, TIR bands depend on the longwave portion of the energy budget. Therefore, the classification map in the spatiotemporal fusion method for TIR images is not based on a fine-resolution TIR image at $t_R$, instead, it is generated from LSE.

LSE, denoted as $\varepsilon$, can be calculated using Equation (14), provided by Sobrino et al. [36]:

$$\varepsilon = m \times P_v + n$$  \hspace{1cm} (14)

with

$$m = \varepsilon_v - \varepsilon_s - (1 - \varepsilon_s) P_s$$  \hspace{1cm} (15a)
\[ n = \varepsilon_s + (1 - \varepsilon_s) F_{\varepsilon_v} \] (15b)

where \( P_v \) is the proportion of vegetation, \( \varepsilon_v \) is emissivity for the vegetated area, \( \varepsilon_s \) is soil emissivity, and \( F \) is a shape factor considered to be 0.55. In order to apply this method, Sobrino et al. [36] suggested using the values of 0.99 and 0.97 for vegetation and soil emissivity, respectively. Finally, Equation (14) can be rewritten as Equation (16) (refer to the reference for more details).

\[ \varepsilon = 0.004 \times P_v + 0.986 \] (16)

\( P_v \) in Equations (14) and (16) is originally obtained from the NDVI, as shown in Equation (17):

\[ P_v = \left[ \frac{\text{NDVI} - \text{NDVI}_{\text{min}}}{\text{NDVI}_{\text{max}} - \text{NDVI}_{\text{min}}} \right]^2 \] (17)

The NDVI, which is based on the spectral difference between red and near infrared (NIR), is defined as Equation (18), proposed by Rouse et al. [37]:

\[ \text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}} \] (18)

Finally, an LSE map derived from Equation (16) is used to generate a classification image using the unsupervised classification algorithm.

Weightings of SWIR Bands

The TIR band has is sensitive to atmospheric profiles. Rosas et al. [38] studied the sensitivity of Landsat-8 surface temperature to atmospheric profile data. Atmospheric conditions, especially water vapor content, can cause bias in the spatiotemporal image fusion method for the TIR band.

Atmospheric correction is generally performed to minimize the bias coming from the atmosphere. However, the atmospheric correction procedure is considered time consuming and not efficient for Himawari-8 TIR images with a 10-min temporal resolution. In this study, the STARFM-AP proposed by Huang et al. [23] is adopted for weighting calculation. The main idea of STARFM-AP is that certain wavelength bands with a low sensitivity to atmospheric effects can be used to efficiently fuse spectral bands that are sensitive to the atmosphere. The SWIR spectral band, which is not significantly affected by the atmospheric profile, is utilized to preserve surface reflectivity and reduce the atmospheric water vapor effect. SWIR spectral bands of fine- and coarse-resolution images at reference time \( t_R \) are applied to generate the weightings. Finally, the weightings are applied to Equation (2) to predict the fine-resolution TIR image.

4.2. Temperature Vegetation Dryness Index

Fused TIR images in this study are then applied to LST retrieval. The conversion of the digital number (DN) to physical units is performed to derive brightness temperature data corresponding to Landsat-8 data user handbook [39]. The LST data is then estimated based on brightness temperature and LSE proposed by Artis and Carnahan (1982) [40].

The temperature vegetation dryness index (TVDI) proposed by Sandholt et al. [41], has been widely used in various studies related to soil moisture estimation and drought assessment [42–45]. The mechanism is based on the relationship between LST and the vegetation index (VI) as a simplified land surface dryness index. The scatterplot in which VI and LST are plotted on the \( x \)-axis and \( y \)-axis, respectively, takes the shape of a triangular trapezoid. The locations of pixels in the scatterplot determine the dryness of the land surface. Pixels located on the upper sloping edge of the triangular trapezoid are defined as the dry edge (\( T_{s_{\text{max}}} \)), representing the minimum soil moisture and evapotranspiration, whereas pixels located on the lower sloping edge are defined as the wet edge (\( T_{s_{\text{min}}} \)), representing the maximum soil moisture and potential evapotranspiration.
The VI calculation in remote sensing is generally based on the NIR spectral band because it has high reflectivity to vegetation. In order to reduce the atmospheric effect, the red spectral band is employed because it has lower reflectivity to vegetation. In this study, we used the enhanced vegetation index (EVI) as the VI. Besides the red spectral band, the EVI also uses the blue spectral band, which also has low reflectivity to vegetation, in order to further reduce the atmospheric effect.

The TVDI is derived from Equation (19) as follows:

$$\text{TVDI} = \frac{LST - T_{\text{min}}}{T_{\text{max}} + T_{\text{min}}}$$

with

$$T_{\text{max}} = a + b \times \text{EVI}$$

$$T_{\text{min}} = a' + b' \times \text{EVI}$$

where $T_{\text{max}}$ is the slope of the dry edge and $T_{\text{min}}$ is the slope of the wet edge. The EVI is derived from Equation (21), corresponding to the Landsat-8 data user handbook provided by the US Geological Survey [39]:

$$\text{EVI} = 2.5 \times \frac{\text{NIR} - \text{Red}}{\text{NIR} + 6 \times \text{Red} - 7.5 \times \text{Blue} + 1}$$

### 4.3. Solar Radiation

The parameters of solar radiation, such as net radiation at the surface ($R_n$) and soil heat flux density ($G$), in this study are derived from remote sensing data. Equation (22), proposed by Allen et al. [46], was used to obtain $R_n$ (W m$^{-2}$):

$$R_n = (1 - \alpha) \times R_{\text{down}} + R_{\text{L}}^{\downarrow} - R_{\text{L}}^{\uparrow} - (1 - \varepsilon) \times R_{\text{L}}^{\downarrow}$$

where $\alpha$ is surface albedo, $R_{\text{down}}$ is the incoming shortwave radiation (W m$^{-2}$), $R_{\text{L}}^{\downarrow}$ is the incoming longwave radiation (W m$^{-2}$), $R_{\text{L}}^{\uparrow}$ is the outgoing longwave radiation (W m$^{-2}$), and $\varepsilon$ is LSE. The $\alpha$ is calculated in hours by Equation (23):

$$\alpha = \frac{(\alpha_{\text{tot}} - \alpha_{\text{atm}})}{\tau_{\text{sw}}^2}$$

where $\alpha_{\text{tot}}$ is planetary albedo without atmospheric correction; $\alpha_{\text{atm}}$ is atmospheric albedo, the portion of solar radiation reflected by the atmosphere (the value of 0.03 is adopted in this study); and $\tau_{\text{sw}}$ is atmospheric transmittance in the solar radiation domain obtained from Equation (24) [46]:

$$\tau_{\text{sw}} = 0.75 + 2 \times 10^{-5} \times \text{Alt}$$

where Alt stands for altitude of the station, which is considered to be 859 m. The $\alpha_{\text{tot}}$ is obtained from remote sensing data using Equation (25), and in this study we used ESTARFM to synthetize hourly images for hourly $\alpha_{\text{tot}}$ calculation:

$$\alpha_{\text{tot}} = (\omega_2 \times \rho_{\lambda2}) + (\omega_3 \times \rho_{\lambda3}) + (\omega_4 \times \rho_{\lambda4}) + (\omega_5 \times \rho_{\lambda5}) + (\omega_6 \times \rho_{\lambda6}) + (\omega_7 \times \rho_{\lambda7})$$

where $\rho_{\lambda b}$ is the TOA planetary reflectance of specific Landsat-8 bands (bands 2–7) and $\omega_b$ is a weighting calculated as the ESUN (extra-terrestrial solar irradiation) of specific bands divided by the sum of all constant ESUN, as shown in Equation (26):

$$\omega_b = \frac{\text{ESUN}_b}{\sum \text{ESUN}_b}$$
The ESUN values of Landsat-8 bands 2 to 7 are presented in Table 2. In order to convert to TOA planetary reflectance, the DN (digital number) of each band must be converted to TOA planetary spectral reflectance without correction for solar angle, as shown in Equation (27):

$$\rho_\lambda' = M_\rho \times Q_{cal} + A_\rho$$ \hspace{1cm} (27)

where $\rho_\lambda'$ is the TOA planetary spectral reflectance without correction for solar angle, $M_\rho$ is the reflectance multiplicative scaling factor for each band (REFLECTANCE_MULT_BAND_n from metadata), $A_\rho$ is the reflectance additive scaling factor for the band (REFLECTANCE_ADD_BAND_n from the metadata), and $Q_{cal}$ is the DN of the band. The $\rho_\lambda$ for each band is obtained from Equation (28):

$$\rho_\lambda = \frac{\rho_\lambda'}{\cos(\theta_{SZ})} = \frac{\rho_\lambda'}{\sin(\theta_{SE})}$$ \hspace{1cm} (28)

where $\theta_{SE}$ is the local sun elevation angle provided in metadata and $\theta_{SZ}$ is the local solar zenith angle; $\theta_{SZ} = 90^\circ - \theta_{SE}$. However, since in this case we calculated hourly $a_{toa}$ by employing fused images, the sun elevation angles of fused images were obtained from an online calculator of the sun’s position (sunearthtools.com) based on location and prediction times of fused images.

### Table 2. Landsat-8 extra-terrestrial solar irradiation (ESUN) (extra-terrestrial solar irradiation) values.

<table>
<thead>
<tr>
<th>Landsat-8 Band</th>
<th>Wavelength (µm)</th>
<th>ESUN</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 (blue)</td>
<td>0.45–0.51</td>
<td>2067</td>
</tr>
<tr>
<td>3 (green)</td>
<td>0.53–0.59</td>
<td>1893</td>
</tr>
<tr>
<td>4 (red)</td>
<td>0.64–0.67</td>
<td>1603</td>
</tr>
<tr>
<td>5 (NIR)</td>
<td>0.85–0.88</td>
<td>972.6</td>
</tr>
<tr>
<td>6 (SWIR 1)</td>
<td>1.57–1.65</td>
<td>245</td>
</tr>
<tr>
<td>7 (SWIR 2)</td>
<td>2.11–2.29</td>
<td>79.72</td>
</tr>
</tbody>
</table>

$R_{\downarrow}$ is calculated using Equation (29):

$$R_{\downarrow} = Q \times \cos(\theta_{SZ}) + d_r \times \tau_{sw}$$ \hspace{1cm} (29)

where $Q$ is the solar constant (1367 W m$^{-2}$) and $d_r$ is the Earth–sun distance. $R_{\downarrow}$ is calculated using Equation (30):

$$R_{\downarrow} = \varepsilon_a \times \sigma \times T_s^4$$ \hspace{1cm} (30)

with

$$\varepsilon_a = 0.85 \times (-\ln(\tau_{sw}))^{0.09}$$ \hspace{1cm} (31)

where $\varepsilon_a$ is the air emissivity function, $\sigma$ is the Stefan–Boltzmann constant ($5.67 \times 10^{-8}$ W m$^{-2}$ K$^{-4}$), and $T_s$ is the hourly air temperature obtained from meteorological stations. Then, $R_{\downarrow}$ is calculated using Equation (32):

$$R_{\downarrow} = \varepsilon \times \sigma \times T_s^4$$ \hspace{1cm} (32)

where $\varepsilon$ is LSE and $T_s$ is hourly LST.

Since the $R_{\downarrow}$ in Equation (22) is calculated in W m$^{-2}$ as an average amount, it is necessary to convert it to MJ m$^{-2}$ to estimate net radiation during an hour using Equation (33):

$$R_{\downarrow}(\text{MJ m}^{-2}) = \frac{R_{\downarrow}(\text{W m}^{-2})}{1 \times 10^6 \times n_s}$$ \hspace{1cm} (33)
where \( n_s \) stands for number of seconds (3600 for hourly scale). Then, hourly soil heat flux density (MJ m\(^{-2}\)) is calculated using Equation (34), suggested by Su [47]:

\[
G = R_n \times \left[ \Gamma_c + (1 - P_v) \times (\Gamma_s - \Gamma_c) \right]
\]  (34)

where \( R_n \) is calculated in MJ m\(^{-2}\), \( \Gamma_c \) and \( \Gamma_s \) are the ratio of soil heat flux to net radiation for full vegetation canopy (\( \Gamma_c = 0.05 \)) and bare soil (\( \Gamma_s = 0.315 \)), respectively, and \( P_v \) is the proportion of vegetation.

### 4.4. Operational Simplified Surface Energy Balance

Actual ET estimation in this study generally used the operational simplified surface energy balance (SSEBop) modeling approach proposed by Senay et al. [48]. The model was originally used to compute daily actual ET. However, since we can provide LST data at an hourly scale by using the spatiotemporal image fusion method, it is possible to calculate actual ET in hours. Actual ET can be estimated from ET fraction generated from LST and reference ET generated from meteorological data as presented in Equation (35):

\[
ET_a = ET_f \times k \times ET_o
\]  (35)

where \( ET_a \) is actual ET, \( ET_f \) is ET fraction, \( ET_o \) is reference ET, and \( k \) is a scaling factor (in this study, \( k = 1 \)). Since this study aimed to estimate hourly actual ET, we generated ET fraction and reference ET in hours based on hourly fused LST and meteorological data.

#### 4.4.1. ET Fraction

ET fraction in the SSEBop model is developed from LST data using two key parameters, maximum air temperature and constant difference between bare dry surface and canopy-level air temperature. The ET fraction is calculated using Equation (36):

\[
ET_f = \frac{Th - LST}{Th - Tc}
\]  (36)

with

\[
Th = Tc + dT
\]  (37a)

\[
Tc = c \times Ta
\]  (37b)

where LST (K) is obtained from remote sensing data, \( Th \) and \( Tc \) are hot and cold pixels, respectively, \( Ta \) is maximum air temperature (K), \( c \) is a correction factor considered to be 0.993, which relates to \( Ta \) and LST, and \( dT \) is a constant difference between bare dry surface and canopy-level air temperature. \( dT \) is calculated using Equation (38):

\[
dT = \frac{R_n \times r_{ah}}{\rho_a \times C_p}
\]  (38)

where \( R_n \) is calculated in W m\(^{-2}\), \( r_{ah} \) is aerodynamic resistance to heat transfer, which was suggested to be 110 s/m by Senay et al. [48], \( C_p \) is the specific heat of air at constant pressure (1.013 kJ kg\(^{-1}\) °C\(^{-1}\)), and \( \rho_a \) is the density of air (kg m\(^{-3}\)) calculated using Equation (39):

\[
\rho_a = \frac{1000 \times P}{T_{kv} \times R} = \frac{3.486 \times P}{T_{kv}}
\]  (39)

with

\[
P = 101.3 \times \left( \frac{293 - 0.0065 \times z}{293} \right)^{5.26}
\]  (40)

where \( R \) is the specific gas constant (287 J kg\(^{-1}\) K\(^{-1}\)), \( T_{kv} \) is the temperature at which dry air must be heated to equal the density of moist air at the same pressure (so-called virtual temperature), which was suggested by Allen et al. [49] to be \( T_{kv} = 1.01 \times (T + 273) \), where \( T \) is mean temperature (°C) (in
this study we used hourly air temperature data), and \( z \) is elevation (m) obtained from 30 m digital elevation model (DEM).

### 4.4.2. Reference ET

The reference ET in this study is computed from meteorological data using the Food and Agriculture Organization (FAO) Penman–Monteith equation [49]. The method requires solar radiation, air temperature, air humidity, and wind speed data. In this study, hourly meteorological data are used to compute hourly reference ET, as shown in Equation (44):

\[
ET_o = 0.408 \times \Delta \times (R_n - G) + \gamma \times \frac{37.5}{(T_{\text{mean}} + 273)} \times u_2 \times (e_s - e_a) \]

\[\Delta + \gamma \times (1 + 0.34 \times u_2) \]

(41)

where \( ET_o \) is the reference evapotranspiration (mm/hour), \( R_n \) is the net radiation during the hour (MJ \ m^{-2} \ \text{hour}^{-1}), \( G \) is the soil heat flux density during the hour (MJ \ m^{-2} \ \text{hour}^{-1}), \( T_{\text{mean}} \) is the mean air temperature at 2 m of height (°C), \( u_2 \) is the wind speed at 2 m of height (m s^{-1}), \( e_s \) is the saturation vapor pressure (kPa), \( e_a \) is the actual vapor pressure (kPa), \( (e_s - e_a) \) is the vapor pressure deficit, \( \Delta \) is the slope vapor pressure (kPa/°C), and \( \gamma \) is the psychrometric constant (kPa/°C). The constant value 37.5 in Equation (42) is multiplied by 24 for daily data, and \( R_n \) and \( G \) are expressed as MJ m^{-2} \ \text{day}^{-1} (refer to the reference for more details).

### 5. Results and Discussion

#### 5.1. Comparison of Fused Images

In this study, a pair of images from 15 November 2018 were used as a reference time \( t_R \) to predict hourly Landsat-like TIR images using STARFM and STAEFM. For ESTARFM, another pair of images from 30 October 2018 were used as reference time \( t_{RA} \).

Figure 4 shows the results of the spatiotemporal image fusion of TIR Himawari-8 and Landsat-8 images of Chiayi County on 18 January 2019 at 10:20, 11:20, 12:00, and 13:20 h (local time). It seems difficult to distinguish the differences in hourly Landsat-like TIR images from this figure. The performance of the three methods is compared in the next section. Figure 5 shows a comparison of the Landsat LST images and fused LST images. Compared to STARFM, the estimated LST of STAEFM is closer to the LST estimated from Landsat TIR images, indicating that the STAEFM can better predict TIR images. However, the fused images of ESTARFM show differences in the spatial structure compared to Landsat TIR, which might be caused by significant differences in surface radiance between \( t_R \) and \( t_{RA} \). Figure 6 shows a comparison of LST image histograms of Landsat TIR and fused images. The histogram of LST image estimated from STAEFM is closer to the Landsat TIR LST image.

#### 5.2. Error Assessment of Fused Images

In this study, we did not use ground-based measurement LST data for validation due to a lack of data availability. Instead, the fused images in this study were assessed based on similarity to Landsat images and a comparison with air temperature collected from meteorological stations. The peak signal-to-noise ratio (PSNR) was used to demonstrate the accuracy of performance of STARFM, ESTARFM, and STAEFM.

Figure 7 shows the scatter plots of fused TIR and Landsat TIR images. STAEFM has more than 42 dB PSNR, which illustrates that the STAEFM TIR image is the most like the Landsat TIR image, whereas STARFM and ESTARFM have PSNR values of 31 and 38, respectively. STARFM gave overestimated results, whereas ESTARFM and STAEFM predicted better results. However, STAEFM has the best agreement among the methods, indicated by the similarity in spatial structure to the Landsat TIR image.
Figure 4. Results of fused TIR images from STARFM, ESTARFM, and STAEFM approaches.

Figure 5. LST (°C) estimated from (a) Landsat TIR, (b) STARFM, (c) ESTARFM, and (d) STAEFM on 18 January 2019.
5.2. Error Assessment of Fused Images

In this study, we did not use ground-based measurement LST data for validation due to a lack of data availability. Instead, the fused images in this study were assessed based on similarity to Landsat images and a comparison with air temperature collected from meteorological stations. The peak signal-to-noise ratio (PSNR) was used to demonstrate the accuracy of performance of STARFM, ESTARFM, and STAEFM.

Figure 7 shows the scatter plots of fused TIR and Landsat TIR images. STAEFM has more than 42 dB PSNR, which illustrates that the STAEFM TIR image is the most like the Landsat TIR image, whereas STARFM and ESTARFM have PSNR values of 31 and 38, respectively. STARFM gave overestimated results, whereas ESTARFM and STAEFM predicted better results. However, STAEFM has the best agreement among the methods, indicated by the similarity in spatial structure to the Landsat TIR image.

Table 3 shows a comparison of air temperature collected from meteorological stations with LST derived from Landsat TIR and fused images. The smallest difference in air temperature is for the LST derived from Landsat TIR image. The range of differences between air temperature and LST derived from Landsat TIR is 1.52 °C to 4.93 °C. LST derived from STAEFM TIR images has smaller differences in air temperature compared with STARFM and ESTARFM TIR images, with a range of 3.05 °C to 6.16 °C.

<table>
<thead>
<tr>
<th>Station</th>
<th>Ta (°C)</th>
<th>LST (°C)</th>
<th>LST-Ta</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Landsat</td>
<td>STARFM</td>
<td>ESATRFM</td>
</tr>
<tr>
<td>B</td>
<td>19.80</td>
<td>21.95</td>
<td>28.17</td>
</tr>
<tr>
<td>C</td>
<td>19.30</td>
<td>25.97</td>
<td>29.65</td>
</tr>
<tr>
<td>D</td>
<td>19.90</td>
<td>22.19</td>
<td>28.65</td>
</tr>
<tr>
<td>E</td>
<td>20.40</td>
<td>21.92</td>
<td>29.00</td>
</tr>
</tbody>
</table>
5.3. Application of STAEFM

Figure 8 shows the hourly LST data, including histograms around 09:00–10:00, 11:00–12:00, and 13:00–14:00 local time on 18 January 2019, which was in winter. It seems the LST variations in the morning hours did not change a lot, then started changing in the afternoon, affected by changes in meteorological parameters and surface characteristics. The maximum LST for the three periods was around 27 °C belonging to dry soil located in the south of study area. The hourly LST data were then applied to TVDI calculation and hourly ET estimation.

The dry and wet edges defined by the relationships of LST and EVI from fused images at different times, as shown in Figure 9, were then used to calculate TVDI around 09:00–10:00, 11:00–12:00, and 13:00–14:00 local time on 18 January 2019. Each slope of dry and wet edges defines coefficients that are used to calculate TVDI, as shown in Table 4. Thus, Equation (19) was applied to calculate TVDI for each fused image.

Table 4. Slope and offset of dry and wet edges estimated from STAEFM TIR on 18 January 2019. EVI, enhanced vegetation index.

<table>
<thead>
<tr>
<th>Local Time</th>
<th>Dry Edge</th>
<th>Wet Edge</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00–10:00</td>
<td>(T_{\text{max}} = 31.67 - 10.72 \times \text{EVI}) (R^2 = 0.980)</td>
<td>(T_{\text{min}} = 14.70 + 2.79 \times \text{EVI}) (R^2 = 0.883)</td>
</tr>
<tr>
<td>11:00–12:00</td>
<td>(T_{\text{max}} = 30.10 - 11.52 \times \text{EVI}) (R^2 = 0.969)</td>
<td>(T_{\text{min}} = 14.86 + 3.49 \times \text{EVI}) (R^2 = 0.904)</td>
</tr>
<tr>
<td>13:00–14:00</td>
<td>(T_{\text{max}} = 35.17 - 16.53 \times \text{EVI}) (R^2 = 0.991)</td>
<td>(T_{\text{min}} = 8.46 + 8.14 \times \text{EVI}) (R^2 = 0.941)</td>
</tr>
</tbody>
</table>
The dry and wet edges defined by the relationships of LST and EVI from fused images at different times, as shown in Figure 9, were then used to calculate TVDI around 09:00–10:00, 11:00–12:00, and 13:00–14:00 local time on 18 January 2019. Each slope of dry and wet edges defines coefficients that are used to calculate TVDI, as shown in Table 4. Thus, Equation (19) was applied to calculate TVDI for each fused image.

![Figure 9](image_url) Regression of dry and wet edges estimated from STAEFM TIR on 18 January 2019 at (a) 09:00–10:00, (b) 11:00–12:00, and (c) 13:00–14:00.

Table 4. Slope and offset of dry and wet edges estimated from STAEFM TIR on 18 January 2019. EVI, enhanced vegetation index.

<table>
<thead>
<tr>
<th>Local Time</th>
<th>Dry Edge</th>
<th>Wet Edge</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00–10:00</td>
<td>Tsmax = 31.67 – 10.72 × EVI</td>
<td>Tsmin = 14.70 + 2.79 × EVI</td>
<td>0.980</td>
</tr>
<tr>
<td>11:00–12:00</td>
<td>Tsmax = 30.10 – 11.52 × EVI</td>
<td>Tsmin = 14.86 + 3.49 × EVI</td>
<td>0.969</td>
</tr>
<tr>
<td>13:00–14:00</td>
<td>Tsmax = 35.17 – 16.53 × EVI</td>
<td>Tsmin = 8.46 + 8.14 × EVI</td>
<td>0.991</td>
</tr>
</tbody>
</table>

Figure 10 shows the results of TVDI calculation, including histograms. The TVDI distributions around 09:00–10:00, 11:00–12:00, and 13:00–14:00 local time on 18 January 2019 have a similar structure. According to the histograms, TVDI at 13:00–14:00 shows that the surface was slightly drier compared with 09:00–10:00 and 11:00–12:00, which was potentially due to evapotranspiration from the land surface.

![Figure 10](image_url) Same as Figure 8, but with the results of temperature vegetation dryness index (TVDI) on 18 January 2019.

As mentioned above, actual ET is obtained from ET fraction and reference ET. Hourly LST data were used to generate ET fraction, whereas hourly meteorological data were used to generate hourly reference ET. Figure 11 shows hourly actual ET images during 09:00–10:00, 11:00–12:00, and 13:00–14:00 local time on 18 January 2019. Actual ET distribution was lower during 09:00–10:00, when the maximum amount of water transferred to the atmosphere was 0.377 mm/hour, whereas actual
ET distribution was higher during 11:00–12:00 and 13:00–14:00, with the maximum amount of water transferred to the atmosphere up to 0.509 mm/hour because solar radiation was more intense. Beside meteorological factors, land use distribution contributed the variations of TVDI and ET in this study. When the TVDI was higher over bare soil indicated high surface dryness, the ET was lower. Compared with TVDI, actual ET was obviously intensified at 13:00–14:00 (Figure 11c) and could have resulted in a loss of soil moisture, which is consistent with the state of a drier surface (higher TVDI in Figure 10c).

5.4. Correlation between Evapotranspiration and Meteorological Parameters

Actual ET is lower during morning hours and higher during afternoon hours on clear-sky days. The amount of solar energy available to vaporize water significantly determines the ET process. Consequently, actual ET is at its maximum in midafternoon when solar radiation intensity is the highest during the day. Transpiration, as a part of ET, increases due to warmer air temperatures, which trigger plants to open stomata so that the water they contain is transferred to the atmosphere. Under conditions of high humidity, the amount of water transferred to the atmosphere decreases, which means the ET rates drop. In addition, wind speed tends to significantly increase the ET process because wind can transport water vapor contained in the air.

Figure 12 shows correlations of actual ET estimated in this study with meteorological parameters albedo, air temperature, relative humidity, and wind speed. Hourly ET and albedo have a negative correlation with the coefficient of determination ($R^2$), which is 0.6303. On the other hand, there is a significant positive correlation between hourly ET and air temperature, with $R^2 = 0.8808$. In addition, hourly ET and relative humidity have a strong negative correlation, with $R^2 = 0.7878$, and hourly ET and wind speed have a positive correlation, with $R^2 = 0.6835$. Hence, we can conclude that the hourly actual ET in this study has the strongest correlation with air temperature.
Figure 12. Correlation of hourly actual ET estimated from STAEFM TIR on 18 January 2019 with (a) albedo, (b) air temperature, (c) relative humidity, and (d) wind speed.

6. Conclusions

To retrieve high spatiotemporal LST from satellite observations, the STARFM is modified to fuse Landsat-8 OLI and Himawari-8 AHI TIR images. Theoretically, TIR spectral bands depend on the longwave portion of Earth’s energy budget (surface emission dominant), while visible bands depend on the shortwave portion (surface reflectivity dominant). Therefore, surface emissivity (LSE) plays a key role in the emission of energy as thermal radiation, which is essential to the procedure of TIR image fusion. By considering the main mechanism of TIRS sensors, the proposed STAEFM is based on LSE to find spectrally similar neighboring TIR pixels when generating a classification map, which could be the most important contribution of this study.

For the improvement of TIR image fusion, the STAEFM approach consists of three stages: sharpening Himawari-8 TIR images, generating a classification map based on LSE, and using SWIR bands for a weighting calculation to reduce the water vapor absorption effect. The LST image estimated from the STAEFM fused image in the case study has the closest similarity to the LST image estimated from the Landsat TIR image, and the performance of STAEFM in fusing Landsat-8 and Himawari-8 TIR images is evident. The PSNR value of STAEFM is more than 42 dB, while the values for STARFM and ESTARFM are around 31 and 38 dB, respectively. The classification maps developed from LSE are effective for finding spectrally similar neighboring pixels. Hence, the STAEFM demonstrated promising results of hourly LST data with a 30-m spatial resolution. The fused LST images were then applied to the TVDI calculation and the estimation of hourly ET using the operational simplified surface energy balance (SSEBop) model. During the three periods, the maximum LST, which was around 27 °C, belonged to bare soil, and the maximum rate of ET was around 0.509 mm/hour over vegetation in mountainous areas. The estimated hourly ETs have a strong correlation with meteorological parameters.

In summary, the significant potential to provide very-high-spatiotemporal-resolution (30 m every 10 min) TIR fused images using geostationary and polar satellites for surface dryness and ET monitoring is clearly presented by the proposed STAEFM. The proposed method can facilitate the construction of the near-real-time monitoring of surface thermal energy associated with geostationary satellite
receiving systems. As for future work, hourly LST and actual ET need to be validated with in situ measurements. Furthermore, an examination of better methods for LSE calculation with more datasets covering wider regions is required, since classification mapping based on LSE is the most important contribution of this study.

Author Contributions: T.-H.L., T.W.J. and C.-Y.H. conceived and designed the experiments; T.W.J. and K.-E.C. performed the experiments and analyzed the data; C.H.Y. and K.-E.C. contributed satellite data processing tools; and T.-H.L. and T.W.J. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: The authors gratefully acknowledge the US Geological Survey (USGS) and the Meteorological Satellite Center (MSC) of the Japan Meteorological Agency (JMA) for providing Landsat-8 OLI and Himawari-8 AHI images. This research was funded by the Ministry of Science and Technology, Taiwan (grant numbers MOST 107-2111-M-008-024 and MOST 108-2111-M-008-024). We would like to extend our thanks to the editor and reviewers for their careful and constructive comments, which substantially improved the quality of the manuscript.

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

References
13. Shoko, C.; Dube, T.; Sibanda, M.; Adelabu, S. Applying the surface energy balance system (SEBS) remote sensing model to estimate spatial variations in evapotranspiration in Southern Zimbabwe. Trans. R. Soc. S. Afr. 2015, 70, 47–55. [CrossRef]


16. Li, S.; Jiang, G.M. Land surface temperature retrieval from Landsat-8 data with the generalized split-window algorithm. IEEE 2018, 6, 18149–18162. [CrossRef]

17. Sattari, F.; Hashim, M.; Pour, A.B. Thermal sharpening of land surface temperature maps based on the impervious surface index with the TS/HARP method to ASTER satellite data: A case study from the metropolitan Kuala Lumpur, Malaysia. Measurement 2018, 125, 262–278. [CrossRef]


34. Wang, F.; Qin, Z.; Li, W.; Song, C.; Karnieli, A.; Zhao, S. An efficient approach for pixel decomposition to increase the spatial resolution of land surface temperature images from MODIS thermal infrared band data. Sensors 2014, 15, 304–330. [CrossRef]


45. Gao, Z.; Gao, W.; Chang, N.B. Integrating temperature vegetation dryness index (TVDI) and regional water stress index (RWSI) for drought assessment with the aid of Landsat TM/ETM+ images. *Int. J. Appl. Earth Obs. Geoinf.* 2010, 13, 495–503. [CrossRef]


© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).