The assessment of different vegetation indices for spatial disaggregating of thermal imagery over the humid agricultural region

Kai Liu, Shudong Wang, Xueke Li, Yao Li, Bo Zhang & Ruiting Zhai

To cite this article: Kai Liu, Shudong Wang, Xueke Li, Yao Li, Bo Zhang & Ruiting Zhai (2019): The assessment of different vegetation indices for spatial disaggregating of thermal imagery over the humid agricultural region, International Journal of Remote Sensing, DOI: 10.1080/01431161.2019.1677969

To link to this article: https://doi.org/10.1080/01431161.2019.1677969
The assessment of different vegetation indices for spatial disaggregating of thermal imagery over the humid agricultural region

Kai Liu, Shudong Wang, Xueke Li, Yao Li, Bo Zhang and Ruiting Zhai

ABSTRACT

Land surface temperature (LST) plays a significant role in surface water circulation and energy balance at both global and regional scales. Thermal disaggregation technique, which relies on vegetation indices, has been widely used due to its advantage in producing relatively high resolution LST data. However, the spatial enhancement of satellite LST using soil moisture delineated vegetation indices has not gained enough attention. Here we compared the performances of temperature vegetation dryness index (TVDI), normalized difference vegetation index (NDVI), and fractional vegetation coverage (FVC), in disaggregating LST over the humid agriculture region. The random forest (RF) regression was used to depict the relationship between LST and vegetation indices in implementing thermal disaggregating. To improve the model performance, we used the thin plate spline (TPS) approach to calibrate the RF residual estimation. Results suggested that the models based on TVDI performed better than those based on NDVI and FVC, with a reduced average root mean square error and mean absolute error of 0.20 K and 0.16 K, respectively. Moreover, based on the surface energy balance model, we found the surface evapotranspiration (ET) derived with the TVDI disaggregated LST as inputs achieved higher accuracy than those derived with NDVI and FVC disaggregated LST. It is indicated that TVDI, a soil moisture delineated vegetation indices, can improve the performance of LST enhancement and ET estimation over the humid agriculture region, when combining random forest regression and TPS calibration. This work is valuable for terrestrial hydrology related research.

1. Introduction

Satellite observations using thermal frequency can provide available and reliable land surface temperature (LST), which have been widely used in the fields of climate change and surface hydrology (Li et al. 2013; Tran et al. 2006; Zhou et al. 2014). In recent years,
several LST products have been released by different satellite missions, such as Moderate Resolution Imaging Spectroradiometer (MODIS) and Visible Infrared Imager Radiometer Suite (VIIRS). These datasets have proven to be substantially useful in a variety of applications (Huang, Li, and Lu 2008; Gomis-Cebolla, Jimenez, and Sobrino 2018; Li et al. 2019). Yet thermal observations are generally collected at a coarser pixel resolution, partly due to the trade-off between spatial and temporal resolutions for the satellite sensor. Another reason for this issue is that the energy of the thermal band collected is mainly from the emitted energy from the earth other than the reflected energy from the solar. The relatively low spatial resolution of current LST products makes them difficult to capture detailed temporal and spatial variability of land surface. For example, for the heterogeneous agricultural regions, remotely sensed thermal dataset is required at a finer resolution (Wood et al. 2011; Wu et al. 2012; Semmens et al. 2016).

To date, various techniques based on statistical or physical mechanism have been proposed to enhance the spatial resolution of thermal products. In general, these methods can be divided into two categories. The first one is the fusion-based method, which combines information from different sensors to obtain the fine spatial resolution LST (Liu et al. 2018; Weng, Fu, and Gao 2014; Wu et al. 2015). For instance, Moosavi et al. (2015) and Shen et al. (2016) used the integration of multi-temporal information of MODIS and LANDSAT to predict LST images with the high resolution the same as LANDSAT. The other category is the statistical regression-based method, which disaggregates LST resolution using a variety of auxiliary data. Statistical disaggregation methods are highly recognized due to their simplicity and satisfactory accuracy (Zhan et al. 2013). This type of methods generally depends on the regression correlations between LST and ancillary datasets. To obtain fine-resolution LST, the relationship between LST and environment factors was first established at coarse spatial scales. Then, this relationship was applied to an auxiliary dataset with high spatial resolutions. Kustas et al. (2003) developed the DisTrad using a quantitative function of Normalized difference vegetation index (NDVI) and LST, which has been widely used in LST disaggregating. A number of models building on DisTrad have been established to improve the LST disaggregating performance over different regions (Dominguez et al. 2011; Gao, Kustas, and Anderson 2012; Bonafoni 2016). In these models, various predicting variables such as digital elevation parameters, surface albedo, and soil indices were used to enhance the resolution of LST products (Duan and Li 2016; Hutengs and Vohland 2016; Pan et al. 2018). It should be noted that among these ancillary variables, vegetation indices (especially NDVI) have been generally adopted because of their availability.

Although some methods have achieved satisfactory LST disaggregation accuracy over some types of vegetation covers, the errors caused by the variabilities of soil moisture were not well addressed for the relatively complex agricultural regions, especially those sensitive to rainfall and irrigation (Gao, Kustas, and Anderson 2012; Liu et al. 2018b). Accordingly, the models that are closely related to soil moisture have been developed to provide an alternative to downscaling LST over wet regions (Yang et al. 2015; Zhang et al. 2015; Djamai et al. 2015). Additionally, Normalized difference water index (NDWI) and Temperature vegetation dryness index (TVDI) have been used to disaggregate LST (Bayala and Rivas 2014; Merlin et al. 2010; Zhang, Zhao, and Yang 2019). However, the comparison between soil moisture delineated vegetation indices and NDVI (or other related parameters) over humid agriculture regions has not been thoroughly investigated. Moreover,
the downscaling performance for satellite thermal product has been under debate. To the best of our knowledge, at least two issues should be carefully studied. On one hand, some detailed information still needs to be comprehensively considered especially regarding the feasibility of applying the disaggregation model in complex agricultural regions, and the selection of optimal ancillary variables (Merlin et al. 2012; Anderson et al. 2012; Ebrahimy and Azadbakht 2019). On the other hand, residual calibration is important in eliminating the uncertainty of model outputs. The reliable approach allowing for efficient compensation model results is highly required, and this will make the nonlinear and complex issues more tractable in the process of thermal disaggregation (Chen et al. 2014; Xia et al. 2019).

The aim of this study is to comprehensively assess an LST disaggregation model focusing on soil moisture condition based on TVDI. Other two conventional VIs, NDVI and Fractional vegetation coverage (FVC), were used to provide a detailed comparative analysis of the disaggregation methods. To improve the performance of disaggregated models, an error calibration approach based on thin plate spline (TPS) interpolation approach is applied. The models were evaluated on watershed landscapes in central Iowa, USA, which is one rainy and humid agricultural region.

In addition, a couple of points deserve some elaboration in our work. On one hand, linear disaggregation approaches are widely used because of its simplicity and feasibility. Recently, several machine learning strategies have been developed in disaggregating studies, and their performance can be better than linear models (Hutengs and Vohland 2016; Ebrahimy and Azadbakht 2019; Wu and Li 2019). To accurately depict the relationship between LST and auxiliary variables in the process of thermal disaggregation, we adopted the random forest (RF) algorithm that could be an effective solution for fitting the non-linear relationship. On the other hand, LST images generated from thermal disaggregation models can facilitate surface energy fluxes studies. Accordingly, on the watershed scale, the performance of disaggregated LST models was further evaluated by comparing the surface evapotranspiration (ET) estimated from different disaggregated LST, referring to the work of Agam et al. (2008) and Bisquert et al. (2016).

The rest of this manuscript is organized as follows: Section 2 presents the methodology of thermal disaggregation. Section 3 describes the study region, dataset and assessment measures. Section 4 first gives a detailed analysis of the downscaling LST using the simulated dataset, and the following three sub-sections in turns give the sensitive analysis for TPS calibration, and the evaluation related to the simulated dataset and actual dataset in terms of ET. Finally, we make a conclusion in section 5.

2. Methodology

2.1. LST disaggregation

Statistical-based thermal disaggregation method generally uses the relationship between LST and the selected predictor (V). At a coarse scale, low resolution LST (LST\textsubscript{Low}) can be described with Equation (1).

\[
\text{LST}_{\text{Low}} = F(\text{V}_{\text{Low}}) + \epsilon_{\text{Low}}
\]  

\[1\]
where $F$ is the linear or non-linear regression. The predictive residual $\varepsilon_{\text{Low}}$ at this scale can be described as:

$$\varepsilon_{\text{Low}} = \text{LST}_{\text{Low}} - F(V_{\text{Low}})$$ (2)

When adding the residual produced at the coarse spatial resolution, high resolution LST ($\text{LST}_{\text{High}}$) can be described with Equation (3):

$$\text{LST}_{\text{High}} = F(V_{\text{High}}) + \varepsilon_{\text{High}}$$ (3)

If we assume the residual estimations are consistent at low resolution and high resolution scales, Equation (3) can be deduced to Equation (4).

$$\text{LST}_{\text{High}} = F(V_{\text{High}}) + \frac{\text{LST}_{\text{Low}}}{C_0} F(V_{\text{Low}})$$ (4)

However, the uncertainties of the residuals are regionally sensitive. An effective error interpolation method is required to calibrate these residuals via investigating their spatial correlations. Detailed descriptions of the selection of interpolation techniques are shown in section 2.3.

This study focused on one of the soil moisture delineated vegetation indices – TVDI. In addition, NDVI and FVC were used to compare with TVDI in thermal disaggregation.

### 2.2. Introduction of the TVDI

The TVDI has been widely used in assessing and monitoring soil moisture conditions (Zhu, Jia, and Lv 2017; Patel et al. 2009; Rahimzadeh-Bajgiran, Omasa, and Shimizu 2012). As shown in Figure 1, for a given agricultural region, the combination of LST and NDVI could constitute a trapezoid (or triangle) space, in which the LST variations mainly depend on the surface water content. LST increases when the evaporative effects decrease in the initial phase of water stress. In contrast, LST decreases as the vegetation covers the

![Figure 1. Schematic of TVDI in the NDVI-LST space.](image-url)
increase in the later stage. As a result, variations of soil moisture and LST can be assessed using TVDI under this frame. The TVDI of each pixel is derived using Equation (5):

\[
\text{TVDI} \approx \frac{\text{LST} - \text{LST}_{\text{min}}}{\text{LST}_{\text{max}} - \text{NDVI}}
\]

where \( \text{LST}_{\text{min}} \) and \( \text{LST}_{\text{max}} \) is the minimum and maximum surface temperature corresponding to the NDVI at the condition of extreme wet and dry, respectively.

Several algorithms have been developed to determine the extreme edge (Zhu, Jia, and Lv 2017; Long and Singh 2012). Yet no practical method is quite efficient due to the complex hydrological environments caused by meteorological conditions and surface covers. Considering the practicality and operability, this study used one space regression method (Tang, Li, and Tang 2010; Su et al. 2011) to automatically determine the dry edge. The wet edge was identified as the minimum LST corresponding to maximum NDVI. The detailed description of this algorithm is provided in the supplementary materials.

### 2.3. Regression model selection and TSP calibration

In this study, random forest (RF) algorithm was selected as the regression model \( F \). The RF is an advanced machine learning strategy that has been successfully applied in various studies of regression analysis or land-cover classification (Lopatin et al. 2016; Pal 2005). RF algorithm generally uses a randomly training sample subset to produce a tree, multiple of which can be further aggregated to make a final decision. In practice RF uses out-of-bag (OOB) samples, one-third of samples that are excluded for fitting the regression, to calculate mean square error. Through minimizing the error of OOB data, the number of available variables for each regression tree can be optimized. In this study, RF model was carried out with python package. Here, we also examined the performance of ordinary linear approaches and then compared it with the RF approaches in order to comprehensively assess the LST downscaling models. The models based on TVDI, NDVI, and FVC are hereafter referred to as TVDI_RF, NDVI_RF, and FVC_RF, respectively, which are implemented on the RF frame.

The thin plate spline (TPS) (Wood 2003), a spatial interpolation technique, was used to compensate for the model residual. Residuals produced by the RF disaggregation were interpolated using the TPS approach which were then restored to the fine resolution outputs. Generally, the low resolution image is the only available dataset used to distribute the residuals. First, a spatial dependent function was retrieved through interpolating the low resolution dataset. Then, this spatial dependent function was used to obtain high resolution dataset. Considered its availability in interpolating geo-data, the TPS can calibrate the RF outputs. For a given pixel \((x, y)\), it can be described as follows:

\[
\text{f}_{\text{TPS}}(x, y) = a_0 + a_0 \times x + a_2 \times y + \frac{1}{2} \sum_{i=1}^{n} b_i \times r_i^2 \times \log r_i^2
\]

with the following constraints:

\[
\sum_{i=1}^{n} b_i = \sum_{i=1}^{n} b_i \times x_i = \sum_{i=1}^{n} b_i \times y_i = 0
\]
Where \( i \) is the index of pixels in each coarse pixel and \( n \) is the number pixels, and 
\[
    \sum_{i} r_{i}^{2} = (x - x_{i})^{2} + (y - y_{i})^{2}.
\]
\( a \) and \( b \) are the regression coefficients.

3. Dataset and accuracy assessment

3.1. Dataset

The Walnut Creek (WC) watershed in the southern Ames, IA, USA, was selected as the study area. The area of WC watershed (Figure 2) is approximately 100 km\(^2\) and the climate is humid and rainy. The WC region was dominated by corn and soybean during the growing season. In our previous work (Liu et al. 2018a, 2018), this region was successfully used for testing LST disaggregation models.

The micrometeorological data and ET dataset were collected from the eddy covariance towers that were equipped by the Soil Moisture Experiment (SMEX02). The SMEX02 project was conducted over the WC region during May-September 2002 to assess the soil moisture and water balance (Jacobs et al. 2004). In this study, five towers within the WC watershed area were used to assess the performance of the modelled ET.

The satellite imagery of Landsat ETM collected on 8 July 2002 was used to build the model. There was a rainfall occurred around this period, resulting in a substantial alteration for soil moisture condition. Another three scenes of Landsat TM/ETM images collected on 16 July 2002, 17 July 2002 and 2 August 2002 were used to evaluate the model performance. The surface conditions within these scenes were also affected by obvious rainfalls which could have changed the water availability.

The Landsat-derived LST was obtained using the thermal channel, as described by Li et al. (2004). To evaluate the performance of different models, Landsat image collected on 8 July 2002 were first resampled to 960 m, which were then upscaled to 240, 120 and 60 m, respectively. These three scales were considered as the reference resolutions. The other three Landsat imageries were first resampled to 960 m and then were disaggregated to the resolution of 240 m, which was the commonly used scale in practical application considering the data availability of current satellite observations. Despite several resampling approaches were used in downscaling LST studies, one universal method was rarely found causing of the complex cover condition and climatic background. However, as indicated by (Agam et al. 2007a, 2007b), different resampling algorithms could not significantly impact the performance of thermal downscaling model if appropriate dataset and model structure is available.

![Figure 2. Location and classification image of the study area.](image-url)
Accordingly, a simple areal-average resampling approach was used here to upscale surface temperature, following the study of Liu et al. (2018a) and Liu, Hongbo, and Xueke (2016).

In addition, a series of MODIS products were collected during the clear days between 1 July 2002 and 31 July 2002. The LST with the resolution of 1 km was from the daily MODIS products (MOD11A1). The daily NDVI and FVC with the 250 m resolution were estimated from the daily surface reflectance product (MOD09GQ). This dataset was used to evaluate the spatial-temporal ET estimation performance of the downscaled inputs.

3.2. Accuracy assessment

The accuracy of LST disaggregation models was evaluated using the root mean square error (RMSE) and the mean absolute error (MAE). Meanwhile, the Nash-Sutcliffe coefficient (NSE) was adopted to assess the overall accuracy.

\[
\text{NSE} = 1.0 - \frac{\sum_{i=1}^{n} (\text{LST}_i - \text{LST}_i^d)^2}{\sum_{i=1}^{n} (\text{LST} - \text{LST}_i^r)^2} \tag{8}
\]

where LST\(_i^r\) represents the reference surface temperatures, LST\(_i^d\) is the disaggregated temperatures and \(\text{LST}_i^r\) represents the mean value of the reference surface temperatures, and \(i\) is the index of pixels in the image and \(n\) is the number of pixels. Here, a higher NSE indicates a better disaggregation model performance.

The surface evapotranspiration is an effective measure for assessing thermal disaggregating models at regional scales. The two-source energy balance (TSEB) model was used to estimate ET in this study. The TSEB is a two-source energy balance model that can estimate soil evaporation and canopy transpiration respectively, mainly using the remotely sensed surface temperatures (Norman, Kustas, and Humes 1995; Kustas and Norman 2000). The advantage of TSEB is that it does not require prior knowledge. This model has been proven to be robust for a large range of landscapes. The detailed descriptions of the TSEB model can be found in the supplementary materials.

4. Results and analysis

4.1. Evaluation of the results

Figure 3(a–c) shows the statistical accuracy of different models using the RF algorithm. It is clear that larger errors were found with finer resolutions. RMSE ranges from 1.92 K to 2.15 K and MAE ranges from 1.37 K to 1.67 K with the 60 m resolution. While for the 240 m resolution, RMSE ranges from 1.25 K to 1.47 K and MAE ranges from 0.92 K to 1.08 K.

TVDI\(_{RF}\) performed better than NDVI\(_{RF}\) and FVC\(_{RF}\) in downscaling LST. TVDI\(_{RF}\) has the smallest RMSE (1.92 K) and MAE (K). In particular, downscaled LST generated from TVDI\(_{RF}\) has the smallest average RMSE and MAE compared with that from the NDVI\(_{RF}\) (with an RMSE of 0.20 K and an MAE of 0.18 K) and FVC\(_{RF}\) (with an RMSE of 0.17 K and an MAE of 0.12 K). The average NSE values of TVDI\(_{RF}\), NDVI\(_{RF}\) and FVC\(_{RF}\) are 0.71, 0.60, and 0.61, respectively. It indicates that the disaggregated LST derived from TVDI\(_{RF}\) is with higher consistency with the reference imagery in the overall structure.
The comparison between the RF models and the ordinary linear models was conducted to further assess the three vegetation indices in disaggregating LST. Figure 3(d-f) shows the error statistics of the models. Overall, the RF models outperformed the ordinary linear models. What’s more, TVDI produced more accurate downscaling results compared with the NDVI and FVC when using the ordinary linear model, reducing the average RMSE and MAE to 0.14 K and 0.13 K, respectively.

Furthermore, we found a consistent pattern of the underestimation for all the three models, primarily due to the contributions of soil moisture. Figures 4 and 5 shows the input and output imagery results using RF and the ordinary linear model with different indices. Through visual inspection, it can be observed that TVDI_RF provides better results across the watershed area. This is attributed to the compensating effects of soil moisture.

For a detailed analysis of the thermal disaggregation models, we divided the absolute errors into four ranges (0–0.5, 0.5–1, 1–2 and above 2 K). Figure 6 displayed the percentage of pixels within each error range. The most of the errors from TVDI_RF model are located in the 0–0.5 range, while for the other three ranges, there is a relatively smaller portion of errors.

Disaggregated LST was also evaluated with three additional scenes – collected on 16 July 2002, 17 July 2002 and 2 August 2002 – with the most practical resolution (240 m). Table 1 shows the statistical results using the TVDI_RF, NDVI_RF, and FVC_RF disaggregate the LST from the resolution of 960 m to 240 m. TVDI_RF performed the best in
downscaling LST. The highest accuracy of TVDI_RF compared with NDVI_RF and FVC_RF were observed across all the imageries. A similar trend was also found for NSE, which can further demonstrate the superiority of TVDI in downscaling LST when combined with the RF regression.

4.2. Sensitive analysis of the TSP calibration

The residual calibration has a significant influence on the performance of thermal disaggregating. Therefore, the sensitivity analysis was also carried out using residual calibration methods. The performance of TPS was compared with two commonly used methods, IDW (Inverse Distance Weighting) and Kriging.

Figure 7 shows the residual distribution using different calibration methods and the corresponding images at the 240 m scale. By visual inspection, TPS kept more spatial details than the other three methods did, demonstrating its advantage of mitigating the model uncertainties. The accuracies for disaggregated LST with different residual
calibration methods are shown in Figure 8. TPS achieved better results comparing to IDW and Kriging. TPS calibration reduced errors of downscaled LST datasets by considering spatial autocorrelation of the LST residuals, the results presented in this study are consistent with the previous studies (Immerzeel, Rutten, and Droogers 2009; Chen et al. 2014).

4.3. Assess the disaggregated LST using ET estimation

To further assess the accuracy of disaggregated LST, we compared its performance in improving ET estimation. Specifically, the ETs estimated using the disaggregated LSTs (at 240 m, 120 m and 60 m scale, respectively) as driving parameters were compared with those using the reference LSTs as driving parameters. Figure 9 shows the statistical accuracy of the estimated ET. The ET estimates coincided with the results of disaggregated LST. It’s observed the ET estimated using the TVDI_RF downscaled LST produced lower RMSE and MAE. On average, RMSE and MAE of the TVDI_RF derived ET are approximately 5.3 Wm\(^{-2}\) and 4.6 Wm\(^{-2}\) lower than those of NDVI_RF and FVC_RF derived ET, respectively.

ET results were also reflected by NSE values. TVDI_RF achieved an additional average improvement in model ET accuracy of 5.1% and 7.8% compared to NDVI_RF and FVC_RF
did, respectively. This demonstrated the better performance of TVDI_RF in LST enhancement, consequently facilitating the estimation of ET.

Figure 10 shows the visual images of TSEB modelled ET using the 240 m LST maps as inputs. In addition, it should be noted that while the soil moisture obviously affects the LST downscaling, its effects on the ET estimation is relatively small. This decrease is caused by the neutralization of surface ET estimation model.

**Figure 7.** Residual maps with calibration methods of (a) no calibration, (b) TPS, (c) IDW, and (d) Kriging, and the corresponding 240 m resolution disaggregated LST that is calibrated with methods of (a) no calibration, (b) TPS, (c) IDW, and (d) Kriging.
4.4. ET estimation over watershed

For practical purposes, 1 km MODIS LST and high resolution LST was respectively used to estimate the daily ET over the watershed area. High resolution LST with a resolution of 250 m was retrieved with the TVDI-RF model during the clear days between 1 July 2002

Figure 8. Comparison of the accuracy for disaggregated LST with residual calibration methods of (a) no calibration, (b) TPS, (c) IDW, and (d) Kriging.

Figure 9. Statistics of the ET estimation using the disaggregated LST in comparison with the reference LST with the resolution of (a) 240 m, (b) 120 m and (c) 60 m.

4.4. ET estimation over watershed

For practical purposes, 1 km MODIS LST and high resolution LST was respectively used to estimate the daily ET over the watershed area. High resolution LST with a resolution of 250 m was retrieved with the TVDI-RF model during the clear days between 1 July 2002
(DOY182) and 31 July 2002 (DOY212). Two sets of LST dataset were used to drive the TSEB model. The daily ET was evaluated with the EC tower values. Both of the modelled daily ETs agreed well with the field measurements, as shown in Figure 11. The estimated ET using high resolution disaggregated LST were more accurate compared with that using the original MODIS.

Furthermore, we examined the spatial distributions of TSEB-produced ET. As shown in Figures 12 and 13, the daily ET of vegetation is higher than those of non-vegetation covers, indicating the spatial and temporal variations of ET were reasonable. Moreover, the spatial-temporal distributions of ET can be better identified using the high resolution disaggregated ET images. The magnitudes and distributions of ET are easier to be distinguished using the downscaled images. This study is consistent with the studies which suggested that the higher resolution of MODIS was much better in describing spatial patterns of ET, partly because the moisture condition and land-cover types at humid areas are more likely to be discriminated with high resolutions (Agam et al. 2007a; Liu et al. 2018).

Figure 10. ET estimates from TSEB model using inputs of (a) the 960 m LST image, (b) the 240 m reference image, and (c–e) is the thermal disaggregated LST produced by the random forest model using the indices of TVDI, NDVI and FVC, respectively.

Figure 11. Comparison of the TSEB modelled daily ET accuracy for: (a) 1 km MDOSI LST and (b) 250 m disaggregation LST.
5. Summary and conclusion

Current studies have indicated that the accuracy of thermal disaggregation models is susceptible to soil moisture conditions, but quantitative investigation for these effects regarding its magnitude and availability for surface ET estimation is still lacking. To investigate satellite LST disaggregation methods considering the influences of soil moisture, we assessed the TVDI thermal disaggregation models. The TVDI-based models were compared with two commonly used VI-based models across the watershed landscapes in central Iowa, USA. The evaluation was conducted with both the simulated and observational data at different scales.

For the downscaling performance with the simulated dataset, the average RMSE and MAE was decreased from 1.81 K (NDVI_RF and FVC_RF) to 1.61 K (TVDI-RF) and from 1.29 K (NDVI_RF and FVC_RF) to 1.13 K (TVDI-RF), respectively. Compared with NDVI and FVC,
TVDI can delineate soil moisture, which makes it capable to enhance LST retrieval over the wet agriculture regions. This is primarily due to the fact that the variations in water availability and soil moisture condition could be better captured by TVDI, especially over the humid and rainy agricultural regions (Eswar, Sekhar, and Bhattacharya 2016; Gao, Kustas, and Anderson 2012).

Random forest, a non-linear machine learning algorithm, was adopted to establish the relationship between LST and three vegetation indices. RF regression was found competent in downscaling LST especially integrating with TVDI. Compared with ordinary linear regression, RF model has been proved of robust in coping with non-linear problems (Reichstein et al. 2019). Being insensitive to the multicollinearity among explanatory variables makes RF easily address high-dimension parameterization while not considering the overfitting issue, and this facilitates the regression model in training variables and achieving desirable outcome. Moreover, RF is relatively readily implemented due to the combined merits of automatic model fitting and small tuning parameters, and therefore could provide an accessible framework for complex landscapes at different scales. However, it should be noted that RF-based model may be susceptible to training data. For this study over wet agricultural area, enough pixels must be included to ensure the representation of different surface properties. Therefore, properly selecting training pixels should be carefully studied in order to achieve reasonable rules for learning to describe LST changes and the connections with explanatory variables.

On the other hand, a calibration approach based on residual interpolation model is proved to be an essential step for downscaling LST product. Considering the spatial variability caused by the change of land cover types, residuals were interpolated to fine resolutions which were then restored to the RF model outputs. The downscaled LST using TPS calibration can further improve the model accuracy compared with the IDW and Kriging. The integration of TPS calibration into the framework of random forests shows a promising potential to improve LST downscaling because of the decrease in error propagation (Quan et al. 2018).

Accurate ET product is crucial in agriculture monitoring and hydrology studies. This study demonstrated the feasibility of the available downscaled LST in promoting ET estimation over a wet agricultural land, which was supported by recent studies (Bisquert et al. 2016; Semmens et al. 2016). It was noted that the TSEB estimated ET using TVDI disaggregated LST were more accurate than those using NDVI and FVC-derived LST. For the simulated dataset, the TVDI-RF downscaled LST achieved an average 5.1–7.8% higher accuracy in estimating ET compared to NDVI-RF and FVC-RF models. Meanwhile, using the MODIS product, TVDI-RF could improve the estimation of daily ET across the watershed region by enhancing the disaggregation of available LST. It should be noted that even though the effects of soil moisture condition on surface ET estimation is relatively small compared to LST downscaling due to the neutralization of ET estimation model, a compensatory role of TVDI in reducing soil moisture effects on ET estimation cannot be ignored.

The findings presented here would be useful for regions where soil moisture is vital for agriculture management. The better performance of TVDI for the agriculture region indicates that soil moisture plays a possible role in controlling the thermal variations and consequently affects LST. Our study was consistent with and further demonstrated previous work, which reported that LST relies heavily on soil moisture (Sandholt, Rasmussen,
and Andersen 2002; Amazirh, Merlin, and Er-Raki 2019). On the other hand, to eliminate the soil moisture-related bias in LST disaggregation, an attempt was made to delineate these effects using multiple remote sensing data. Yet directly using soil moisture for thermal disaggregation models has been constrained by data availability within the large scale. Alternative indices accounting for soil moisture effects may provide a reference for solving this issue. In future research, developing models that are more applicable to complex surface hydrology environments should be given enough attention.

**Acknowledgements**

We greatly thank to the researchers in SMEX02 for data acquisition and sharing.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

This work was supported by the Natural Science Foundation of China (41671362).

**ORCID**

Ruiting Zhai [http://orcid.org/0000-0002-4357-8393](http://orcid.org/0000-0002-4357-8393)

**References**


