Ensemble-Based Spatiotemporal Fusion Methods for Simulating Land Surface Temperature Using Landsat 8/9 and MODIS

Nahid Haghshenas [1], Ali Shamsoddini [2]*

[1] Dept. of RS and GIS, Tarbiat Modares University, Tehran, Iran -n.haghshenas@modares.ac.ir
[2] Dept. of RS and GIS, Tarbiat Modares University, Tehran, Iran - ali.shamsoddini@modares.ac.ir

KEY WORDS: Fusion Methods, MODIS, Landsat 9, Machine Learning, Ensemble Methods.

ABSTRACT:

Land Surface Temperature (LST) is one of the most important parameters for monitoring surface energy and water balance, playing a crucial role in analyzing environmental changes at local and global scales. Despite significant advancements in remote sensing technologies and the availability of time-series data, limitations in sensor design and the inherent trade-off between spatial and temporal resolution still pose challenges for generating accurate LST time series. To overcome these challenges, various Spatiotemporal Fusion (STF) algorithms have been proposed. In this study, ensemble learning algorithms, including Extreme Gradient Boosting (XGBoost), Gradient Boosting Machine (GBM), Random Forest (RF), and Deep Forest, were employed as the core of an STF framework to simulate daily LST data from Landsat 8/9—an approach not widely explored in spatiotemporal fusion until now. Additionally, the classical Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (ESTARFM) was included for comparison. The results indicated that XGBoost achieved the highest accuracy in predicting LST, with Root Mean Square Error (RMSE) values of 2.76, 1.68, and 1.54 ˚K for 2016, 2022, and 2023, respectively, outperforming ESTARFM and other ensemble learning methods. Conversely, GBM showed the weakest performance among the ensemble models, with RMSE values of 2.83, 1.74, and 1.69 ˚K. The ESTARFM algorithm produced RMSE values of 2.78, 1.69, and 2.64 ˚K, reflecting a noticeably different performance compared to ensemble-based approaches. Overall, this research highlights the potential and advantages of ensemble learning techniques, particularly XGBoost, in spatiotemporal fusion for enhancing LST prediction accuracy.

1. INTRODUCTION

LST, as a key physical parameter, provides essential information about spatial-temporal variations in surface energy and water balance at both local and global scales (Ebrahimy, Azadbakht, 2019). It controls many biophysical and biogeochemical processes and regulates the Earth's surface energy balance (Hu et al., 2018). With the rapid growth in data volume, increasing access to satellite imagery, and advancing computational efficiency, remote sensing applications are moving towards monitoring the dynamics of the Earth's surface using massive time-series data (Woodcock et al., 2020). LST is a vital parameter for examining and interpreting processes occurring in the atmosphere and on the Earth's surface at both regional and global scales (Li et al., 2021). Due to limitations imposed by a single satellite platform or sensor, acquiring high-resolution LST data in both spatial and temporal dimensions has become challenging. For instance, daily sensors like MODIS provide LST with lower spatial accuracy (1 km), while high-resolution sensors (100 meters or less), due to coverage and orbital altitude limitations, only acquire data at intervals of several days. To be more specific, the Landsat series platform within eight days (combining Landsat 7, 8, and 9) (Guo et al., 2024). Moreover, frequent cloud cover and fog make it extremely challenging to have images with close temporal intervals and complete spatial coverage (Zhu et al., 2021). Therefore, despite the increasing number of remote sensing satellites launched during the last decades, the process of acquiring high-resolution LST in both spatial and temporal dimensions remain challenging. As an effective solution, Spatio-temporal fusion could facilitate the generation of high spatial-temporal resolution LST data. Currently, several STF methods have been developed, which can be broadly categorized into four main types: Weight-function-based methods, unmixing-based methods, hybrid methods, and learning-based methods (Wang et al., 2023). Weight-function-based methods generally assume that pixels in low-resolution images are pure and create linear models between multi-temporal and multi-scale observations, utilizing a Spatio-temporal weighting strategy to improve downscaling accuracy. Methods in this group include the STARFM model (Gao et al., 2006) and its improved version, ESTARFM (Zhu et al., 2010; Cheng et al., 2017). Unmixing-based methods assume that each low-resolution pixel is a linear combination of the spectral reflectance based on different classes in a region with higher spatial resolution. For instance, some techniques, such as multi-sensor multi-scale methods and their improvements, fall under this category (Zhu et al., 2016; Zhukov et al., 1999; Zurita-Milla et al., 2009). Learning-based methods establish non-linear relationships between images with different resolutions (Chen et al., 2022). Examples of such methods include the dictionary-pair learning-based methods (Huang, Song, 2012) and Artificial Neural Networks (Moosavi et al., 2015). In previous studies,

* Corresponding author

spatiotemporal fusion was mainly conducted based on classical Weight-function-based methods (Guo et al., 2025), hybrid methods (Liang et al., 2024), and deep learning models (e.g., CNN) (Liu et al., 2019; Song et al., 2018; Tan et al., 2019). In most cases, deep learning methods based on CNNs require a large amount of training data. On the other hand, traditional techniques frequently struggle to produce accurate results in areas with intricate spatial patterns or heterogeneous land cover.

Among machine learning approaches, ensemble learning algorithms are recognized as one of the most powerful and flexible methods. By combining a set of base learners and leveraging the diversity among them, ensemble methods offer strong capabilities in improving the accuracy and stability of predictions. This advantage becomes particularly important when dealing with data characterized by spatial heterogeneity, temporal noise, and high structural complexity. In contrast to many conventional machine learning methods that rely on a single model trained from the data (Mendes-Moreira et al., 2012; Zhou, 2009), ensemble learning—through the integration of a set of base learners—provides superior generalization ability and uncertainty management (Galar et al., 2011).

Despite the remarkable success of these approaches across diverse scientific and engineering fields, their application in STF has received relatively limited attention. Previous studies have predominantly focused on deep learning algorithms based on convolutional neural networks (CNNs), whereas the potential of ensemble learning to enhance the accuracy and robustness of STF models remains largely underexplored.

In this context, the present study is designed to investigate and elucidate the role of ensemble learning algorithms in enhancing the spatiotemporal fusion of Land Surface Temperature (LST) data. Considering the inherent challenges of LST data including substantial spatial and temporal variability, limited training samples due to cloud contamination, and the nonlinear nature of spatiotemporal relationships, this study employs ensemble learning algorithms comprising of Gradient Boosting Machine (GBM), Random Forest (RF), and XGBoost as the core of the proposed spatiotemporal fusion framework.

Notably, the use of the Deep Forest algorithm as a deep ensemble learning model represents an innovative aspect of this research. This algorithm facilitates the exploration of synergies between deep and ensemble learning mechanisms and is specifically employed to capture complex spatiotemporal dynamics.

Moreover, to more accurately assess the performance of ensemble models and analyse their relative advantages, the classical ESTARFM algorithm, a well-established, weight-based reference method, was also applied. The results of the comparative analysis of these models can provide a deeper understanding of the role and potential of ensemble learning in improving the accuracy and stability of LST prediction, and serve as a novel step in the development of spatiotemporal fusion methods based on ensemble and deep learning, particularly in the context of monitoring land surface temperature dynamics.

In the following, the article is divided into four sections: the second section introduces the data and study area, the third section describes the algorithms and evaluation indices, the fourth section presents the results and discussion, and finally, the fifth section concludes this paper.

2. STUDY AREA AND DATA

2.1 Study area

As shown in Figure 1, the study area is located in Yanco, within the Murrumbidgee River catchment in southeastern Australia, covering an area of approximately 40 × 80 km^2. The land use in this region is primarily irrigated cropland and pasture. Geographically, the area extends from -34.185° to -34.980° South latitude and from 145.834° to 146.769° East longitude. The topography is predominantly flat, with elevations ranging from 117 to 150 meters above sea level and minimal slope (Senanayake et al., 2021; Ye et al., 2020; Young et al., 2008).

2.2 Data

In this study, nine pairs of LST images from the MODIS and Landsat 8 and 9 were collected. These images correspond to the summer and autumn seasons across three different years. Due to cloud cover and observation constraints, it was not possible to obtain image pairs for all sensor acquisition dates. Full details of this dataset are provided in Table 1.

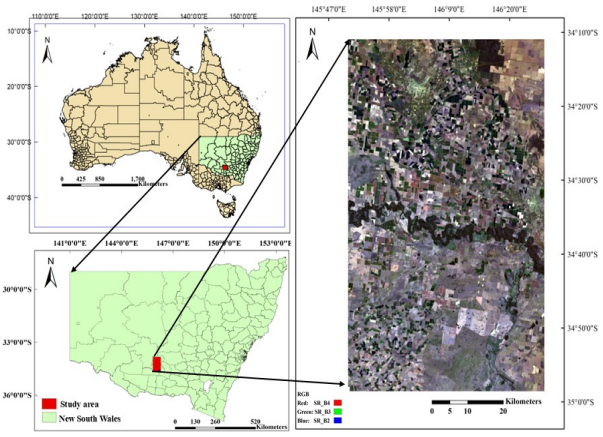

Figure 1. The study area is located in the Yanko region, New South Wales, Australia.

Date of acquisition	Data Collection	LST Retrieval Algorithm
Feb 5, 2016	Landsat 8 LST product MOD11A	Single-channel Split-Window
Mar 8, 2016	Landsat 8 LST product MOD11A	Single-channel Split-Window
Apr 25, 2016	Landsat8 LST product MOD11A	Single-channel Split-Window
Jan 12, 2022	Landsat 8 LST product MOD11A	Single-channel Split-Window
Jan 20, 2022	Landsat 8 LST product MOD11A	Single-channel Split-Window
Feb 13, 2022	Landsat 8 LST product MOD11A	Single-channel Split-Window
Jan 15, 2023	Landsat 9 LST product MOD11A	Single-channel Split-Window
Jan 31, 2023	Landsat 9 LST product MOD11A	Single-channel Split-Window
Apr 21, 2023	Landsat 9 LST product MOD11A	Single-channel Split-Window

Table 1. The used datasets and their main information

2.2.1 Landsat LST data

The Landsat 8 and Landsat 9 satellites, launched on February 11, 2013, and September 27, 2021, respectively, carry Thermal Infrared (TIR) sensors capable of measuring land surface temperature (LST) with a spatial resolution of 100 meters. Both satellites cross the equator around 10:00 AM local time. For this study, LST data were obtained from Level-2 Collection 2 products and processed using the single-channel algorithm. The corresponding Surface Temperature (ST) datasets are generated using version 1.3.0 of the Surface Temperature algorithm developed by the Rochester Institute of Technology, which integrates information such as TOA reflectance, brightness temperature, the Global Emission Database (GED), and atmospheric profiles. All data were accessed through the USGS Earth Explorer platform (https://earthexplorer.usgs.gov).

2.2.2 MODIS LST data

The MOD11A1.006 product, with a spatial resolution of 1 km, is derived from Bands 31 and 32 using the split-window algorithm. The MODIS sensor, onboard the Terra and Aqua satellites, has been providing LST data for over two decades. The MOD11A1 product was selected for analysis due to its close overpass time with Landsat. The accuracy of LST estimates has been reported to be within 1.3 ˚K for homogeneous surfaces. The data were obtained from the USGS Earth Explorer (https://earthexplorer.usgs.gov).

3. METHODOLOGY

The proposed framework in this study (Figure 2) consists of four main stages: **1)** Data preprocessing, including geometric and radiometric corrections. **2)** Dataset splitting into training and testing sets. **3)** Simulation of Land Surface Temperature LST at time T2 using an STF model based on XGBoost, RF, GBM, Deep Forest, and the weighted ESTARFM algorithm. For each prediction date, two pairs of data from the dates before and after the target date, along with the MODIS image of the target date with low spatial resolution, are provided as inputs to the model. The model then leverages different learners to predict the high-resolution Landsat image for the date where a MODIS image is available, but the corresponding Landsat image is missing. **4)** Algorithm performance evaluation using reference Landsat data for the prediction date.

3.1 ESTARFM

The ESTARFM model employs two pairs of high- and low-resolution reference images and assumes that land cover changes between the two acquisition dates follow a linear pattern. To quantify this, a linear conversion coefficient (V) is computed to represent the rate of spectral change (Zhu et al., 2010). The model then integrates spectral, temporal, and spatial differences of spectrally similar pixels through a weighting function (Knauer et al., 2016). In this study, a 3×3 window size was used to define the neighbourhood for identifying spectrally similar pixels.

3.2 RF

RF is an improved version of the Bagging algorithm, in which classification and regression decision trees (CART) are used as base learners. The algorithm enhances accuracy by introducing model diversity through the random selection of data subsets and random sampling of variables at each split. Additionally, approximately 30% of the training samples not used during the bootstrapping process are employed to estimate the out-of-bag (OOB) prediction error (Breiman, 2001).

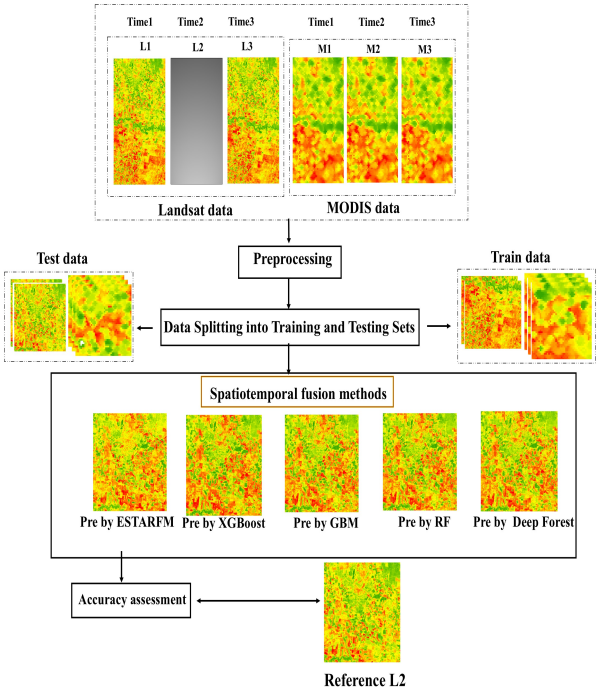

Figure 2. Research Flowchart

3.3 ESTARFM

The ESTARFM model employs two pairs of high- and low-resolution reference images and assumes that land cover changes between the two acquisition dates follow a linear pattern. To quantify this, a linear conversion coefficient (V) is computed to represent the rate of spectral change (Zhu et al., 2010). The model then integrates spectral, temporal, and spatial differences of spectrally similar pixels through a weighting function (Knauer et al., 2016). In this study, a 3×3 window size was used to define the neighbourhood for identifying spectrally similar pixels.

3.4 RF

RF is an improved version of the Bagging algorithm, in which classification and regression decision trees (CART) are used as base learners. The algorithm enhances accuracy by introducing model diversity through the random selection of data subsets and random sampling of variables at each split. Additionally, approximately 30% of the training samples not used during the bootstrapping process are employed to estimate the out-of-bag (OOB) prediction error (Breiman, 2001).

3.5 GBM

GBM is a powerful method for solving regression problems by cumulatively combining weak learners, typically regression trees, to improve prediction accuracy. At each iteration, a new model is added to correct the errors of the previous one, using gradient descent to minimize the loss function. Key characteristics of GBM include iterative learning from residuals, continuous model optimization, and the significant influence of parameters such as tree depth, learning rate, and the number of iterations on overall performance. Because of its ability to model complex and nonlinear relationships, high flexibility, and strong generalization capacity, GBM is widely used across various applications (Konstantinov, Utkin, 2021

3.6 XGBoost

XGBoost, an advanced implementation of GBM, offers faster learning capabilities compared to traditional GBM by employing parallel processing at the node level. Moreover, it incorporates regularization techniques to control model complexity and reduce the risk of overfitting, one of its key advantages over classical GBM approaches (Chen, Guestrin, 2016).

3.7 Deep Forest

Deep Forest, or gcForest, is a deep learning method based on decision forests that uses an ensemble of decision trees instead of neural networks for learning. With its cascade structure, the model processes data layer by layer, enhancing features at each level. Each layer consists of multiple decision forests, increasing the diversity of the model. A key feature of gcForest is multi-grained scanning, which extracts local features using sliding windows and feeds them into decision forests. The depth of the model is automatically adjusted based on performance on a validation set, and k-fold cross-validation is used to reduce the risk of overfitting (Zhou, Feng, 2019).

3.8 Accuracy Assessment

To evaluate the performance of the results, high-resolution LST images acquired on January 31, 2023, that is, the prediction time, were obtained from the Landsat 9 satellite. For the quantitative evaluation of the results, the Peak Signal-to-Noise Ratio (PSNR), Correlation Coefficient (CC), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) were applied. The equations (1) to (4) represent the formulas for these indicators. The effective range of the CC value is between the interval [−1, 1]; values closer to 1 indicate better simulation results and greater similarity between the textural details of the predicted image and the actual image (Zhu et al., 2021). Higher PSNR indicates better prediction, while lower RMSE and MAE values signify higher quality of the simulated image (Guo et al., 2024).

$$\mathrm{RMSE} = \sqrt{\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[y(i,j)-x(i,j)]^2} \quad (1)$$

$$\mathrm{MAE} = \frac{\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}|y(i,j)-x(i,j)|}{m \times n} \quad (2)$$

$$CC = \frac{\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left(x_{(i,j)}-\mu_x\right)\left(y_{(i,j)}-\mu_y\right)}{\sqrt{\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(x_{(i,j)}-\mu_x)^2(y_{(i,j)}-\mu_y)^2}} \quad (3)$$

$$\mathrm{PSNR} = 20 \times \log_{10}\left(\frac{MAX_y}{\sqrt{MSE}}\right) \quad (4)$$

where $X(i, j)$ = actual pixel value at location (i, j)
Y (i, j) = predicted pixel value at location (i, j)
m, n = number of rows and columns in the image
MAXy maximum pixel value in the image
μx, μy = mean values of actual and predicted images

4. RESULT AND DISCUSSION

As shown in Table 2, the results of the Spatiotemporal fusion of LST for the three examined dates are presented. The results indicate that the XGBoost algorithm achieved the highest accuracy among the evaluated models for all three dates, with RMSE values of 2.78, 1.68, and 1.54 ˚K for 2016, 2022, and 2023, respectively. In contrast, the GBM and Deep Forest algorithms exhibited the lowest performance among the ensemble techniques. In addition, comparison with the ESTARFM algorithm indicates that its accuracy was significantly lower, with an approximate difference of 1.09 ˚K compared to the best-performing model (XGBoost) for the year 2023.

Year	Method	RMSE (˚K)	MAE (˚K)	CC	PNSR
2016	ESTARFM	2.7834	1.9707	0.9023	20.83
	RF	2.8512	2.3341	0.9101	20.62
	GBM	2.8354	2.3240	0.9138	20.67
	XG Boost	**2.7612**	**2.2606**	**0.9151**	**20.90**
	Deep Forest	2.8928	2.3663	0.9058	20.49
2022	ESTARFM	1.6939	1.3555	**0.8602**	23.59
	RF	1.7704	1.3522	0.8339	23.20
	GBM	1.7419	1.3979	0.8408	23.34
	XG Boost	**1.6813**	**1.2917**	0.8544	**23.65**

	Deep Forest	1.8259	1.3882	0.8262	22.94
2023	**ESTARFM**	2.6359	2.2485	0.8309	17.06
	RF	1.6130	1.2274	0.8399	21.32
	GBM	1.6989	1.3230	0.8449	20.87
	XG Boost	**1.5487**	**1.1845**	**0.8513**	**21.68**
	Deep Forest	1.6415	1.2429	0.8344	21.17

Table 2. Landsat LST Prediction Results.

As shown in Table 2 and the scatter plots in Figure 3, a clear difference is observed between ESTARFM and the ensemble learning models, with XGBoost outperforming the others. The better performance of XGBoost can be attributed to its upgraded architecture, which incorporates efficient boosting algorithms, node-level parallel processing, and advanced regularization mechanisms. These features allow it to effectively minimize overfitting while maintaining accuracy and robustness (Chen, Guestrin, 2016). Specifically, XGBoost builds predictive models by sequentially combining decision trees. Each new tree focuses on correcting the errors (residuals) made by the previous trees. The model uses the gradient of the loss function to optimize performance, minimizing prediction errors iteratively(Liu et al., 2024).

These findings are consistent with the study of Filgueiras et al. (2020), who demonstrated that ensemble methods such as GBM outperformed linear models (e.g., Principal Component Regression, Ridge Regression, Partial Least Squares Regression, and Linear Support Vector Machine) in generating 30-meter resolution NDVI products with MODIS-like temporal frequency, particularly in regions exhibiting significant vegetation dynamics (Filgueiras et al., 2020). Similarly, Sahin (2020) discovered that XGBoost outperformed Random Forest and GBM in landslide susceptibility mapping (LSM), particularly when selecting optimal subsets of conditioning factors, which supports the current study's findings (Sahin, 2020).

The poor performance of the ESTARFM algorithm in this study is likely due to inherent limitations in its underlying assumptions, particularly under conditions of significant spatiotemporal variation. The temporal gaps of 80, 48, and 96 days in 2016, 2022, and 2023 between the image pairs used for spatiotemporal fusion resulted in substantial changes in land surface characteristics, which markedly affected ESTARFM's predictive accuracy. This performance gap was especially pronounced in 2023, when sudden changes between the base images were more significant.
In contrast, ensemble learning models, by combining multiple learners and leveraging a wide range of input data, offer greater flexibility and adaptability in addressing complex spatiotemporal prediction challenges (Galar et al., 2011). This adaptability is particularly crucial for applications requiring accurate LST estimation under rapidly changing conditions, demonstrating that ensemble-based algorithms can effectively overcome the limitations of traditional methods like ESTARFM.

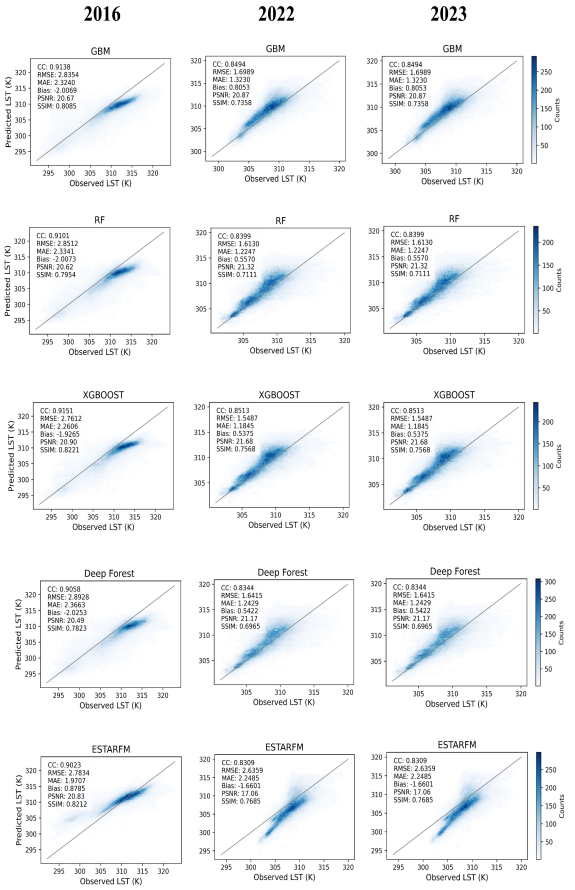

Figure 3. Scatter plots of the reference images and the result.

Figure 3 illustrates the spatial distribution of RMSE produced from LST simulations using various techniques. In this figure, red areas indicate errors higher than 5 ˚K, while blue areas show errors less than 1 ˚K. Because of the varied character of agricultural landscapes, the error distribution is non-uniform and spread throughout the image, as shown in the true-color images of the study area, some agricultural plots exhibited errors exceeding 5 ˚K, reflecting sudden changes such as crop harvesting during the interval between the pairs of base images used for spatiotemporal fusion. As clearly shown in the figure, errors in 2016 and 2023 were higher than those in 2022, which may be attributed to greater changes in crop fields during these years. Previous research has also found limits in spatiotemporal fusion models in locations with considerable land cover variability, particularly in agricultural areas (Tao et al., 2021). In the current study, variability in crop types and differences in harvest timing led to non-uniform land surface changes, resulting in localized high errors. In 2023, due to more abrupt changes, the extent of these areas in the ESTARFM output was noticeably larger compared to the ensemble models.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W8-2025-315-2026 | © Author(s) 2026. CC BY 4.0 License.

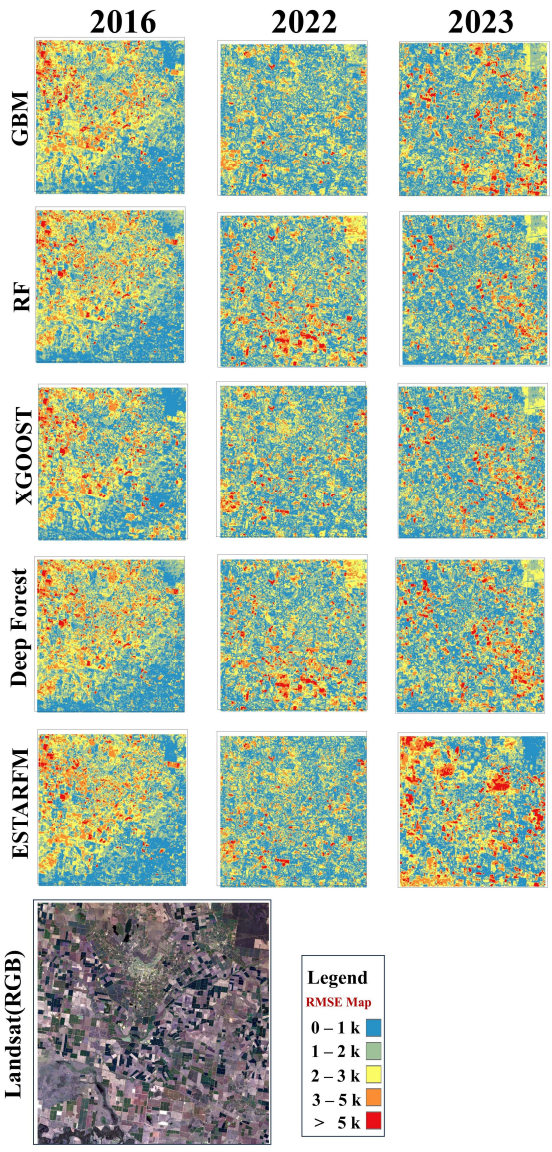

Figure 4. RMSE Map of the LST images and RGB image.

Another noteworthy aspect is the performance of the Deep Forest algorithm in this study. Although this technique is a hybrid model that integrates deep learning and ensemble methods, it did not demonstrate higher accuracy compared to XGBoost. The results of Deep Forest were comparable to those of other ensemble algorithms, showing no clear advantage. This finding suggests that, under certain conditions, particularly in the spatiotemporal fusion of land surface temperature, ensemble learning approaches may outperform the deep learning components within ensemble frameworks. For future studies aiming to improve and develop spatiotemporal fusion models based on ensemble learning, it is recommended to place greater emphasis on optimizing ensemble learning techniques rather than deep learning components in these models.

5. CONCLUSION

In this study, the performance of ensemble learning algorithms and the classical ESTARFM method for spatiotemporal fusion of LST was evaluated using MODIS and Landsat 8/9 data across different temporal intervals for the years 2016, 2022, and 2023.The results showed that the XGBoost algorithm consistently achieved the highest accuracy and stability across all examined dates, outperforming other ensemble algorithms and the ESTARFM method with lower RMSE values. Its superior performance can be attributed to its advanced structure, including efficient boosting mechanisms, parallel processing, and effective regularization, which enable better modeling of complex and nonlinear land surface changes.

Overall, ensemble learning algorithms demonstrated greater robustness than the classical ESTARFM method in dynamic agricultural areas, showing lower sensitivity to gradual and abrupt temporal changes. Conversely, ESTARFM performed reasonably well over short time intervals and under conditions with limited spatiotemporal variation, highlighting the importance of selecting appropriate base image pairs, temporal intervals, and algorithms for specific applications.

Although the Deep Forest algorithm integrates deep learning and ensemble techniques, it did not outperform XGBoost in this study and yielded results comparable to other ensemble algorithms. This suggests that, in tree-based models, ensemble components may be more effective than deep learning elements unless more advanced and optimized hybrid models are developed in future research.

In general, among spatiotemporal fusion models based on ensemble learning algorithms, XGBoost can serve as a highly efficient tool for LST fusion, offering higher accuracy and greater temporal and seasonal stability. However, the findings of this study are limited to datasets with similar characteristics, and broader generalization requires testing across different regions, time periods, and additional algorithms.

REFERENCES

Breiman, L., 2001: *Random forests*. Machine learning, 45, 5-32.

Chen, G., Lu, H., Di, D., Li, L., Emam, M., Jing, W., 2022: StfMLP: Spatiotemporal fusion multilayer perceptron for remote-sensing images, IEEE Geoscience and Remote Sensing Letters, 20, 1-5.https://doi.org/doi.org/10.1109/LGRS.2022.3230720

Chen, T., Guestrin, C., 2016: Xgboost: A scalable tree boosting system, Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 785-794. https://doi.org/10.1145/2939672.2939785

Cheng, Q., Liu, H., Shen, H., Wu, P., Zhang, L., 2017: A spatial and temporal nonlocal filter-based data fusion method, IEEE Transactions on Geoscience and Remote Sensing, 55, 4476-4488. https://doi.org/ doi.org/10.1109/TGRS.2017.2692802

Ebrahimy, H., and Azadbakht, M., 2019: Downscaling MODIS land surface temperature over a heterogeneous area: An investigation of machine learning techniques, feature selection, and impacts of mixed pixels, Computers & Geosciences, 124, 93-102. https://doi.org/ doi.org/10.1016/j.cageo.2019.01.004

Filgueiras, R., Mantovani, E. C., Fernandes-Filho, E. I., Cunha, F. F. d., Althoff, D., and Dias, S. H. B., 2020: Fusion of MODIS

and Landsat-Like images for daily high spatial resolution NDVI, RemoteSensing, 12,1297.
https://doi.org/doi.org/10.3390/rs12081297

Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., Herrera, F., 2011: A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42, 463-484, 2011. https://doi.org/ doi.org/10.1109/TSMCC.2011.2161285

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

Guo, H., Wang, X., Ouyang, Z., Chen, S., Che, T., Zheng, Z., 2025: Application of the ESTARFM algorithm for fusing Sentinel-2 and MODIS NDSI series in the eastern Qilian Mountains, Journal of Hydrology: Regional Studies, 57, 102103. https://doi.org/ doi.org/10.1016/j.ejrh.2024.102103

Guo, S., Li, M., Li, Y., Chen, J., Zhang, H. K., Sun, L., Wang, J., Wang, R., Yang, Y., 2024: The Improved U-STFM: A Deep Learning-Based Nonlinear Spatial-Temporal Fusion Model for Land Surface Temperature Downscaling, Remote Sensing, 16, 322. https://doi.org/ doi.org/10.3390/rs16020322

Hodson, T. O., 2022: Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not, Geoscientific Model Development Discussions, 2022,1-10.
https://doi.org/doi.org/10.5194/gmd-15-5481-2022

Hu, Y., Zhong, L., Ma, Y., Zou, M., Xu, K., Huang, Z., Feng, L., 2018. Estimation of the land surface temperature over the Tibetan Plateau by using Chinese FY-2C geostationary satellite data, Sensors, 18, 376.
https://doi.org/doi.org/10.3390/s18020376

Huang, B., Song, H., 2012: Spatiotemporal reflectance fusion via sparse representation, IEEE Transactions on Geoscience and Remote Sensing, 50, 3707-3716.
https://doi.org/doi.org/10.1109/TGRS.2012.2186638

Ke, Y., Im, J., Park, S., Gong, H., 2016: Downscaling of MODIS one kilometer evapotranspiration using Landsat-8 data and machine learning approaches, Remote Sensing, 8, 215, 2016. https://doi.org/ doi.org/10.3390/rs8030215

Knauer, K., Gessner, U., Fensholt, R., Kuenzer, C., 2016: An ESTARFM fusion framework for the generation of large-scale time series in cloud-prone and heterogeneous landscapes, Remote Sensing, 8, 425.
https://doi.org/doi.org/10.3390/rs8050425

Konstantinov, A. V. and Utkin, L. V., 2021: Interpretable machine learning with an ensemble of gradient boosting machines, Knowledge-BasedSystems, 222,106993.
https://doi.org/doi.org/10.1016/j.knosys.2021.106993

Li, K., Guan, K., Jiang, C., Wang, S., Peng, B., Cai, Y., 2021: Evaluation of four new land surface temperature (LST) products in the US corn belt: ECOSTRESS, GOES-R, landsat, and sentinel-3, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 9931-9945.
https://doi.org/doi.org/10.1109/JSTARS.2021.3114613

Liang, M., Gu, X., Liu, Y., Cheng, T., Cao, H., Zhang, H., Zhang, Q., Ding, Y., Gao, M., Wei, X., 2024: Preliminary Evaluation of Angular Reflectance Downscaling Using FSDAF Spatiotemporal Fusion Model and MODIS BRDF Data, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, 5042-5058.
https://doi.org/doi.org/10.1109/JSTARS.2024.3365826

Liu, Y., Wu, S., Wu, Z., Zhou, S., 2024: Application of gradient boosting machine in satellite-derived bathymetry using Sentinel-2 data for accurate water depth estimation in coastal environments, Journal of Sea Research, 201, 102538. https://doi.org/doi.org/10.1016/j.seares.2024.102538

Liu, X., Deng, C., Chanussot, J., Hong, D., Zhao, B., 2019: StfNet: A two-stream convolutional neural network for spatiotemporal image fusion, IEEE Transactions on Geoscience and Remote Sensing, 57, 6552-6564. https://doi.org/doi.org/10.1109/TGRS.2019.2907310

Mendes-Moreira, J., Soares, C., Jorge, A. M., Sousa, J. F. D., 2012: Ensemble approaches for regression: A survey, Acm computing surveys(csur), 45, 1-40.
https://doi.org/doi.org/10.1145/2379776.2379786

Moosavi, V., Talebi, A., Mokhtari, M. H., Shamsi, S. R. F., Niazi, Y., 2015: A wavelet-artificial intelligence fusion approach (WAIFA) for blending Landsat and MODIS surface temperature, Remote Sensing of Environment, 169, 243-254. https://doi.org/doi.org/10.1016/j.rse.2015.08.015

Ponomarenko, N., Ieremeiev, O., Lukin, V., Egiazarian, K., Carli, M., 2011: Modified image visual quality metrics for contrast change and mean shift accounting, 2011 11th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), 305-311.

Sahin, E. K., 2020: Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest, SN Applied Sciences, *2*(7), 1308.

Senanayake, I., Yeo, I.-Y., Walker, J., Willgoose, G., 2021: Estimating catchment scale soil moisture at a high spatial resolution: Integrating remote sensing and machine learning, Science of The Total Environment, 776, 145924. https://doi.org/doi.org/10.1016/j.scitotenv.2021.145924

Smith, A. B., Walker, J. P., Western, A. W., Young, R., Ellett, K., Pipunic, R., Grayson, R., Siriwardena, L., Chiew, F. H., Richter, H., 2012: The Murrumbidgee soil moisture monitoring network data set, Water Resources Research, 48.
https://doi.org/doi.org/10.1029/2012WR011976

Song, H., Liu, Q., Wang, G., Hang, R., and Huang, B., 2018: Spatiotemporal satellite image fusion using deep convolutional neural networks, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11, 821-829. https://doi.org/doi.org/10.1109/JSTARS.2018.2797894

Tan, Z., Di, L., Zhang, M., Guo, L., Gao, M., 2019: An enhanced deep convolutional model for spatiotemporal image fusion, Remote Sensing, 11, 2898. https://doi.org/doi.org/10.3390/rs11242898

Tao, G., Jia, K., Wei, X., Xia, M., Wang, B., Xie, X., Jiang, B., Yao, Y., Zhang, X., 2021. Improving the spatiotemporal fusion accuracy of fractional vegetation cover in agricultural regions by combining vegetation growth models, International journal of applied earth observation and geoinformation, 101, 102362. https://doi.org/doi.org/10.1016/j.jag.2021.102362

Wang, Q., Tang, Y., Ge, Y., Xie, H., Tong, X., Atkinson, P. M., 2023: A comprehensive review of spatial-temporal-spectral information reconstruction techniques, Science of Remote Sensing, 100102. https://doi.org/doi.org/10.1016/j.srs.2023.100102

Woodcock, C. E., Loveland, T. R., Herold, M., Bauer, M. E., 2019: Transitioning from change detection to monitoring with remote sensing: A paradigm shift, Remote Sensing of Environment, 238, 111558. https://doi.org/doi.org/10.1016/j.rse.2019.111558

Ye, N., Walker, J. P., Wu, X., De Jeu, R., Gao, Y., Jackson, T. J., Jonard, F., Kim, E., Merlin, O., Pauwels, V. R., 2020: The soil moisture active passive experiments: Validation of the SMAP products in Australia, IEEE Transactions on Geoscience and Remote Sensing, 59, 2922-2939. https://doi.org/doi.org/10.1109/TGRS.2020.3007371

Yee, M. S., Walker, J. P., Monerris, A., Rüdiger, C., Jackson, T. J., 2016: On the identification of representative in situ soil moisture monitoring stations for the validation of SMAP soil moisture products in Australia, Journal of hydrology, 537, 367-381. https://doi.org/ doi.org/10.1016/j.jhydrol.2016.03.060

Young, R., Walker, J., Yeoh, N., Smith, A., Ellett, K., Merlin, O., Western, A., 2008: Soil moisture and meteorological observations from the Murrumbidgee catchment, Department of Civil and Environmental Engineering, The University of Melbourne, 610.

Zhou, Z.-H. Feng, J., 2019: Deep forest, National science review, 6, 74-86. https://doi.org/ doi.org/10.1093/nsr/nwy108

Zhou, Z.H., 2015: Ensemble learning, Encyclopedia of biometrics, Springer, Singapore, Boston, MA. https://doi.org/doi.org/10.1007/978-1-4899-7488-4_293

Zhou, Z.H., 2015: Ensemble Learning. In: Li, S.Z., Jain, A.K., (eds) Encyclopedia of Biometrics. Springer, Boston, MA. https://doi.org/doi.org/10.1007/978-1-4899-7488-4_293

Zhu , X., Chen, J., Gao, F., Chen, X., Masek, J. G., 2010: An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions, Remote Sensing of Environment, 114, 2610-2623. https://doi.org/doi.org/10.1016/j.rse.2010.05.032

Zhu , X., Helmer, E. H., Gao, F., Liu, D., Chen, J., Lefsky, M. A., 2016: A flexible spatiotemporal method for fusing satellite images with different resolutions, Remote Sensing of Environment, 172, 165-177. https://doi.org/ doi.org/10.1016/j.rse.2015.11.016

Zhu, X., Song, X., Leng, P., Li, X., Gao, L., Guo, D., Cai, S., 2021: A Framework for Generating High Spatiotemporal Resolution Land Surface Temperature in Heterogeneous Areas, Remote Sensing, 13, 3885. https://doi.org/doi.org/10.3390/rs13193885

Zhukov, B., Oertel, D., Lanzl, F., Reinhackel, G., 1999: Unmixing-based multisensor multiresolution image fusion, IEEE Transactions on Geoscience and Remote Sensing, 37, 1212-1226. https://doi.org/ doi.org/10.1109/36.763276

Zurita-Milla, R., Kaiser, G., Clevers, J., Schneider, W., 2009: Schaepman, M. E.: Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics, Remote Sensing of Environment, 113, 1874-1885. https://doi.org/doi.org/10.1016/j.rse.2009.04.011