IMAGE QUALITY ASSESSMENT OF SPATIOTEMPORAL IMAGE FUSION: A CASE STUDY APPROACH USING LANDSAT-8 AND SENTINEL-2

Greetta Pinheiro[1], Sonajharia Minz[2]
[1]Jawaharlal Nehru University,
New Mehrauli Road, JNU Ring Rd, New Delhi, 110067, India, greetz.pinheiro@gmail.com
[2] Jawaharlal Nehru University,
New Mehrauli Road, JNU Ring Rd, New Delhi, 110067, India, sona.minz@gmail.com

ABSTRACT: The multi-source, multi-sensor, multi-spatial, multi-temporal satellite products are widely available over the past few years. Due to the feasibility to incorporate both high spatial resolution and frequent temporal coverage, spatiotemporal fusion has attracted a lot of attention in a variety of applications. Spatiotemporal image fusion can prove to be a cost-effective, efficient, and feasible alternative to construct a time series of high spatial and temporal resolution data. In comparison to the raw images, the fused image should contain enhanced spatial and spectral information. Identifying suitable context-specific fusion methods depends on the quality of the spatiotemporal fused image. In the absence of reference images in context-specific applications, the fused images are directly correlated to the quality of the pair of input spatiotemporal images. There are qualitative and quantitative image quality metrics. Visual comparison between the raw input images and the fused image is done in qualitative analysis for evaluating the performance of the fusion result. Quantitative analysis has two different variations where it evaluates the performance of the fused image in the presence and absence of a reference image. The variety of quality issues of the spatiotemporal fused images such as redundant information, and comparability across various study areas are not completely addressed by commonly used evaluation metrics. A composition of metrics that addresses the above spatiotemporal aspects of fused images is a viable solution to overcome this problem. In the present study, we use a reconstruction-based fusion method known as the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) using Landsat-8 and Sentinel-2 surface reflectance. The fusion result's quality is assessed by using four complete reference image quality assessment metrics as Local Binary Patterns (LBP), Root-Mean-Square Error (RMSE), Edge, and Mean Error (ME) along with a no-reference image quality evaluation metric called Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE). These metrics reduce information redundancy and exhibit computational simplicity in addition to quantifying the spectral and spatial information in the fused images and also significantly improve the correlation of the fusion results with the subjective quality scores.

Key Words: Spatiotemporal Image Fusion, Image Quality Metrics, STARFM, BRISQUE, LBP

1. INTRODUCTION

The increasing availability of satellite products with different spatial, temporal, and spectral resolutions provides enormous amounts of remote sensing big data. To conduct remote sensing research in a variety of heterogeneous areas, including fragmented urban areas and agrarian regions, the acquisition of time-series satellite images with high spatial and temporal resolution is required (Tang et al., 2020; Zeng et al., 2020). The high spatial resolution of satellite-based remote sensors has helped earth observation at fine levels. The trade-off that exists between the revisiting frequency and swath width of these satellite remote sensors limits the acquisition of images at both high spatial and temporal resolutions simultaneously (Belgiu & Stein, 2019; X. Zhu et al., 2018).

Monitoring the Earth's surface at fine spatial and temporal scales using spatiotemporal fusion provides great potential over individual satellite sensor imagery (Ghamisi et al., 2019). Spatiotemporal fusion is used to fuse imageries from different satellite sensors with similar spectral band specifications and the resulting synthetic time series obtained will have an integrated temporal resolution from the two satellite sensors and finer spatial resolution of the two satellite sensors. Both the fine and coarse satellite images should be atmospherically and geometrically corrected before applying the fusion method. The quality assessment of the fusion result is important to determine the appropriate fusion method for the specific application.

In general, Spatiotemporal image fusion requires at least three input images, i.e., a pair of fine and coarse resolution images on the reference date and a coarse resolution image on the target date to generate a synthesized fine resolution image on the target date. In case the temporal coverage of the different satellite sensors is barely overlapped, then the fusion becomes nearly impossible (Wu et al., 2020). Fusion of imagery from both Landsat-8 Operational Land Imager (OLI) and Sentinal-2 Multi-Spectral Instrument (MSI) generates a synthetic time series of 10m spatial resolution at a temporal frequency of 2-3 days (Q. Wang et al., 2017). Due to the long revisit cycle of Landsat which is 16 days, it is often difficult to get temporally matching imagery with respect to the collected reference Sentinal-2 imagery. Even though matching image pairs can be obtained during the extended timespan between both the reference and the target dates, the fusion result gets degraded (Cheng et al., 2017).

While evaluating the quality of an image fusion, two aspects may have an impact on the results: (a) the selection of quantitative indicators (metrics) when performing a quantitative assessment, and (b) the display conditions of the images when performing a qualitative (visual) evaluation. In qualitative assessment, when the comparison is not carried out under the same visualization conditions, such as when the images are stretched and displayed, the comparison won't create accurate results. An original MS image, for instance, often appears black when histogram stretching is not applied, and it appears very differently when other stretches are applied. These differences in appearances are a result of the conditions of image display and not due to the quality difference. When multiple quantitative metrics are selected for the evaluation in a quantitative assessment, different evaluation results are possible (Y. Zhang, 2008).

1.1 Motivation

Measuring the similarity between various pixels is used in several spatiotemporal image fusion methods like STARFM, FSDAF, ESTARFM, etc., which is useful for correctly predicting high-resolution remote sensing images. For instance, the spatiotemporal image fusion methods based on the weight-function use weights to describe the correlation between various pixels, while the method based on unmixing uses classification maps to determine the category information of each pixel. When filtering out similar pixels, several linear regression techniques represent similarity measurement. Three main factors account for the extensive use of the above-mentioned spatiotemporal image fusion methods. Firstly, measuring the relationship of pixels is beneficial because the radiation relationship between sensors is unstable due to changes in radiation relationships, geographic locations, and other factors (Gao et al., 2006). Secondly, a noticeable clustering phenomenon occurs due to the reflectance and temporal variability of the reflectance. According to Tobler's first law of geography, everything is connected, but close things are more related, especially for spatial dependency (X. Zhu et al., 2018). Third, the complexity of heterogeneous regional texture details in remotely sensed images is higher than in the natural image Surface Reflectance task, and the resolution gap between multisource satellite image pairs (such as Landsat-MODIS, Landsat-sentinel, etc,.), which are used to evaluate spatiotemporal fusion methods, is very large (Liu et al., 2019). The influence of the previous two factors is increased by this problem. Consequently, a crucial component of the spatiotemporal fusion method is similarity measurement.
The remainder of this paper is organized as follows. In Section 2, the related literature is discussed. In Section 3, we present the method, test dataset, STARFM python implementation, the image quality assessment metrics, and the results. Section 4 consists of the Results and Discussion and the Conclusion is provided in Section 5.

2. RELATED LITERATURE

Monitoring the rapid surface changes and seasonal vegetation phenology requires high-resolution images in both space and time. Even though such fine spatial and high temporal images are been provided by commercial satellites (e.g., Planet Labs or RapidEye), acquiring these satellite imageries for our specific applications is very expensive. To overcome this challenge, different free sensors provide satellite imagery at a fine spatial scale (e.g., Sentinal-2) and high temporal resolution (e.g., Sentinal-3). Spatiotemporal fusion methods help in enhancing the resolution of historical satellite images and generating high temporal resolution images cost-effectively for monitoring earth observations.
The are several spatiotemporal fusion models developed over the last two decades and are reviewed by various researchers (Belgiu & Stein, 2019; B. Chen et al., 2015; Li et al., 2020; X. Zhu et al., 2018). Integrating heterogeneous and complementary data to improve the reliability of the interpretation and enhance the information in the satellite images, spatiotemporal fusion is used. If the data are recorded by different sources, complementarity on the same observed region is considered if it is using multi-sensors, multi-temporal, multispectral, multi-spatial, or multi-polarization (Pohl & van Genderen, n.d.). Image fusion can also be used to solve the issue of missing data in the time series of satellite images caused by shadow or cloud contamination (Racault et al., 2014).
Complimentary information such as the temporal and spectral data obtained from multiple sensors enhances the precision of the image reconstruction of the spatial data with missing information (Q. Zhang et al., n.d.). The main objective of image fusion according to (Schmitt & Zhu, 2016) is “either to estimate the state of target or object from multiple sensors if it is not possible to carry out the estimation from one sensor or data type alone, or to improve the estimate of this target state by the exploration of redundant and complementary information”.
The spatiotemporal fusion result will be a synthesized image with high spatial resolution obtained from the first sensor and high temporal frequency obtained from the second sensor. Spatiotemporal fusion of sensors with similar spatial and temporal resolution can be used for obtaining consistent observations, such as harmonizing Sentinal-2 and Landsat satellite images (Storey et al., 2016). The spatiotemporal image fusion methods are categorized into five (X. Zhu et al., 2018), Unmixing-based, weight function-based, bayesian-based, learning-based, and hybrid.
The weight function-based category of spatiotemporal image fusion has the most number of fusion methods developed, among which the popular Spatio-temporal fusion technique is the spatial and temporal adaptive reflectance

fusion model (STARFM)(Gao et al., 2006), a pixel-based fusion algorithm (Ghamisi et al., 2019) for blending MODIS and Landsat surface reflectance. It is better suited for homogeneous landscapes where pure coarse pixels predominate since it presumes that the temporal changes of all land cover classes inside a coarse pixel are constant.
It is a widely used fusion method for larger areas of vegetative change detection (Xie et al., 2016; L. Zhu et al., 2017). In STARFM, the information of the neighboring pixels is considered while predicting the pixels with the function which gives higher weight to the 'pure' coarse pixels. It is assumed that the surface reflectance obtained by the fine and coarse resolution sensors are correlated linearly (Mileva et al., 2018). But these assumptions are not true for heterogeneous geographic areas and when the weighted function is empirical (X. Zhu et al., 2018). STARFM is considered a benchmark and other weight-based fusion methods are developed by addressing the above shortcoming or by improving STARFM for fusing other satellite products. Most of the researchers have focused on Landsat and MODIS fusion hence it is well-researched and it provides a foundation for devising a STARFM-based image fusion workflow for the harmonization of Landsat-8 and Sentinal-2. The present study checks the STARFM fusion methods' potential in fusing Landsat-8 and Sentinal-2, as this fusion method is not restricted to Landsat-MODIS or Sentinal-2-Sentinal-3 fusion combinations. The relationships between the low spatial resolution satellite images and the high spatial resolution satellite image pair and the number of these pairs are comparatively less explored until (Y. Chen et al., 2020; Xie et al., 2018) studied how to determine the optimal number of image pairs. The effects of image pairs on various spatiotemporal fusion models, the variety of heterogeneous areas, and the composition of image quality metrics for assessing the fusion quality are to be studied to choose the appropriate fusion model for a context-specific application.
In recent studies, the fused image quality assessment which is used to measure the similarity or difference is carried out using the comparison between the reference image and the fused image. Structural similarity index measure (SSIM), Mean Absolute Error (MAE also known as Average absolute difference AAD), correlation coefficient (r), relative dimensionless global error (ERGAS), coefficient of determination (R2), and root-mean-square error (RMSE) are mean error (ME, also named as Average difference AD) (X. Zhu et al., 2022). Quantitative evaluation using these metrics offers a more accurate and unbiased assessment of the effectiveness of spatiotemporal fusion approaches than qualitative evaluation (or visual assessment). A quantitative analysis is an objective analysis that is based on mathematical modelling. The spectral and spatial similarity between the raw input images and the fused image is assessed using a set of pre-defined quality indicators to determine the quality of the fused image (Y. Zhang, 2008). The referenceless image quality metric used in this study is Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) which trains a support vector regressor (SVR) for perceptual quality prediction using scene statistics of locally normalized luminance coefficients (Mittal et al., 2012). A suitable accuracy evaluation approach may account for accuracy in both the spectral and spatial domains (X. Wang & Wang, 2020).

3. MATERIALS AND METHOD

3.1 Data

Multi-source satellite images having different spatial and temporal resolutions are acquired for the case study. The study area is located in Ohio, United States. A brief description of the satellite images used in the present case study with their spatial and temporal resolutions is shown in Table 2.

Table 1: Acquisition date of Landsat-8 and Sentinel 2A for the study area

Data	Acquisition date	Resolution Spatial-Temporal
Landsat-8 OLI	04-08-2015	30 m 16 days
	20-08-2015	30 m 16 days
Sentinel-2A	04-08-2015	10 m 5–6 days

The Landsat-8 OLI images (accessed from www.usgs.gov) are simulated by downscaling Sentinel-2 MSI images (accessed from www.corpenicus.eu) to 30 m resolution in order to eliminate errors from differences in the geolocation errors, atmospheric correction, and other artifacts caused by pre-processing operations like resampling and collocation. The downscaling is done using the nearest neighbor algorithm. Bands 2, 3, and 4 are selected for Landsat-8 images, and bands 2, 3, and 4 are selected for Sentinel-2 images.

3.2 STARFM Theoretical Basis

In STARFM (Gao et al., 2006) , given the two coarse resolution images C^{t_0} and C^{t_1} at time t_0 (reference date) and t_1 (prediction date), and a fine resolution image F^{t_0} at time t_0 (reference date) are acquired from two different satellites. Prediction of the fine resolution surface reflectance F^{t_1} at time t_1 (prediction date) is retrieved by using the information obtained from the above three images which are acquired.

Let (x_i, y_i) be the location of the pixel, *w* be the size of the moving window which is used for searching the similar pixels, and w_{ijk} be the weighting parameter which consists of three different factors: a) Spatial distance d_{ijk} measured between the central pixel and the neighbouring pixel. b) Spectral distance S_{ijk} between Sentinel-2 and Landsat-8 data at the given location at time t_1. c) Temporal distance T_{ijk} between the reference and prediction dates Landsat data. The fine resolution surface at t_1 can be calculated as follows:

$$F^{t_1}(x_{w/2}, y_{w/2}) = \sum_{i=1}^{w} \sum_{j=1}^{w} \sum_{k=1}^{n} W_{ijk} \times \left(C^{t_1}(x_i, y_j) + F^{t_0}(x_i, y_j) - C^{t_0}(x_i, y_j)\right), \quad (1)$$

where the Spatial distance d_{ijk} between the central pixel $(x_{w/2}, y_{w/2})$ and the neighbouring pixel (x_i, y_j) is given by the equation (2):

$$d_{ijk} = \sqrt{(x_{w/2} - x_i)^2 + (y_{w/2} - y_j)^2} \quad (2)$$

The spatial distance equation (2) needs to be converted to relative distance D_{ijk}, a constant A is used to define the relative importance of other weighting parameters to the spatial distance:

$$D_{ijk} = \frac{d_{ijk}}{A} + 1 \quad (3)$$

The spectral distance S_{ijk} between the coarse resolution image and the fine resolution image at the reference date t_0 for training is given in (4). In order to avoid zero values, a 1 is added:

$$S_{ijk} = \left|F^{t_0}(x_i, y_j) - C^{t_0}(x_i, y_j)\right| + 1 \quad (4)$$

The temporal distance T_{ijk} between the two coarse resolution images at reference t_0 and prediction date t_1 is given in equation (5). As in the previous equation a 1 is added in order to avoid non zero values:

$$T_{ijk} = \left|C^{t_0}(x_i, y_j) - C^{t_1}(x_i, y_j)\right| + 1 \quad (5)$$

The inverse of the three distances are calculated to find the combined distance C_{ijk}, so that the neighbouring pixels that are closer to the central pixel and have a smaller spectral and temporal distance are given more weight:

$$C_{ijk} = \frac{1}{S_{ijk}} \times \frac{1}{T_{ijk}} \times \frac{1}{D_{ijk}} \quad (6)$$

The sensitivity to the spectral distance can be reduced by using the natural logarithm of the weighting distances:

$$C_{ijk} = \frac{1}{\ln(S_{ijk}+1)} \times \frac{1}{\ln(T_{ijk}+1)} \times \frac{1}{\ln(D_{ijk}+1)} \quad (7)$$

Normalization of the combined distance C_{ijk} results in the sum of the weights to be equal to 1. Thus, the final parameters is calculated:

$$W_{ijk} = \frac{C_{ijk}}{\sum_{i=1}^{w} \sum_{j=1}^{w} \sum_{k=1}^{n} C_{ijk}} \quad (8)$$

The similar neighbour pixels inside a window is filtered based on considering whether that pixel contains more spatial and spectral information than the central pixel. In order to take in account, the uncertainty during the pre-processing of the surface reflectance of the satellites, an uncertainty parameter is introduced to the filtering condition.

3.3 Method

A pair of Landsat-8 and Sentinel-2 images are used for training, and a second Landsat image was acquired on the date of prediction. The similarity in the orbital parameters is one of the prerequisites for the spatiotemporal fusion of Landsat-8 and Sentinel-2.

Table 2: Orbital Parameters of Landsat-8 and Sentinel-2 satellites

Orbital Parameters	Landsat-8	Sentinel-2
Orbit inclination	98.2°	98.6°
Mean Local Solar Time	10:00 ± 15min	10:30

The STARFM proposed by (Gao et al., 2006) is capable of retrieving fine-resolution surface reflectance from these data. Utilizing the information from neighborhood pixels, the surface reflectance at fine resolution is calculated. These neighboring pixels must be homogenous and spectrally similar. Additional weights are added based on the spatial distance of the neighboring pixels to the predicted pixel as well as the spectral and temporal differences between the Landsat-8 and Sentinel-2 images.

The main steps of STARFM include:

1. Extracting homogeneous pixels with similar spectral properties from the neighborhood of a sentinel-2 image within a moving window.
2. Calculating the weight function and multiplying it with an image from Sentinel-2 taken at t_0 (F^{t_0}) and the difference in surface reflectance between two Landsat-8 images taken at t_0 and t_1 is (C^{t_0} C^{t_1}).
3. Creating a synthetic image at time t1 by assigning the weighted sum on the moving window's center pixel.

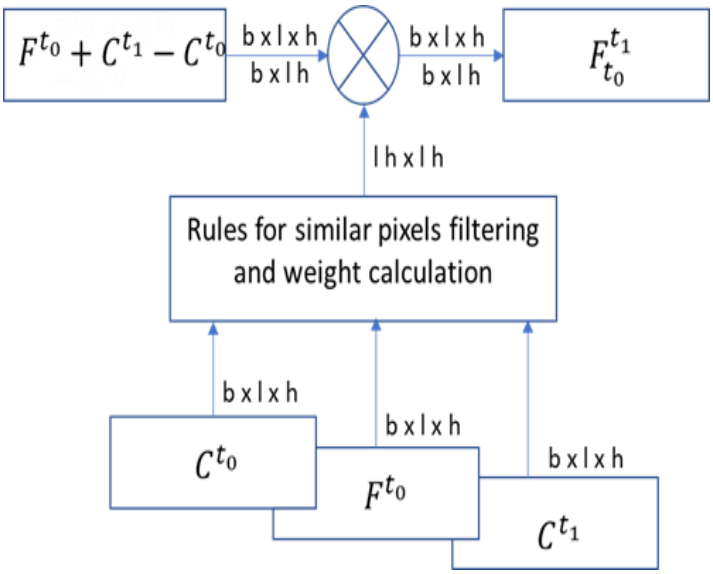


Figure 1: STARFM flow chart

Figure 2 shows the STARFM flow chart. Here F and C denote High spatial resolution images (Fine image) and Low spatial resolution images (Coarse) respectively. The corresponding band number of the input image is denoted by b, and the width and height of the image are denoted using l and h respectively. The reference date is denoted using t0 and t1 denotes the prediction date of the images.

In the python implementation of STARFM (Mileva et al., 2018), computations within the moving window usually consume more time (Gao et al., 2017). The vectorized solution can be used for these computations in python where the computations are carried out in the memory of the computer. As a result, implementing the method across larger regions (such as a single Sentinel-2 tile) might cause memory to run out. The python implementation uses generators for carrying out the moving window operations. The advantage is that the generators create iterator objects which use less memory.

In STARFM, a variety of operations are carried out within the moving window, such as computing the standard deviation of all the pixels in the window to determine a threshold for similarity. The similar pixels are calculated separately for each band. The computations in the STARFM python implementation are done in a 2D space, but adding information from the other bands will add more dimensions. Each window is thereby flattened into a row in order to reduce the problem's dimensionality. As a result, the dimensions are reduced by a factor of 2. The disadvantage of this strategy is the generation of redundant data.

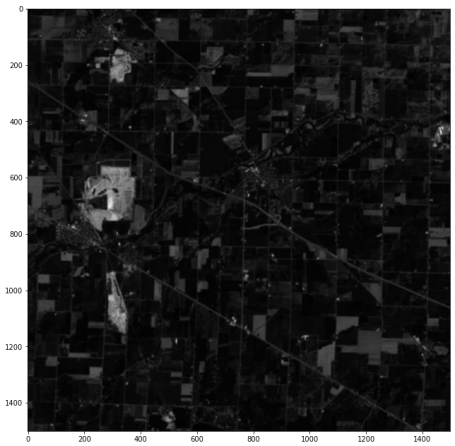
Table 3: Equations for calculating image quality assessment metrics (Reference based IQA metrics).

	Metric name	Equation	Variables
Spectral metrics	Root Mean Squared Error (RMSE)	$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(F_i - R_i)^2}{N}}$	F_i : pixel *i* value in the fused image R_i : pixel *i* value in the reference image N : total number of pixels
	Mean Error (ME)	$AD = \frac{1}{N}\sum_{i=1}^{N}(F_i - R_i)$	
Spatial metrics	Robert's Edge	$Edge = \lvert D_{i,j} - D_{i+1,j+1}\rvert + \lvert D_{i,j+1} - D_{i+1,j}\rvert$	$D_{i,j}$: pixel values at i^{th} row and j^{th} column D_i : pixel values that are surrounding the central pixel in a 3x3 moving window D_c : pixel value of the central pixel deci : convert binary to decimal
	Local Binary Pattern (LBP)	LBP = deci(d1d2d3...d8) $d_i = \begin{cases} 1 \text{ if } D_i > D_C \\ 0, \text{ otherwise} \end{cases}$	

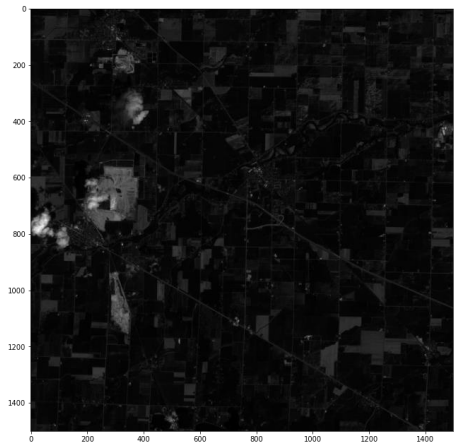
The fused images will have both spectral and spatial information. The fusion image quality assessment metrics should be able to measure the spectral and spatial information that is present in the fused image. The traditional distance-based metrics such as RMSE, ME, etc., can be used to assess spectral accuracy. The computation of RMSE and ME metrics is simple and uses statistical summaries (such as mean and standard deviation) and the pixel values of the entire image. RMSE is used to gauge the degree of inaccuracy in the spatiotemporal fused image. In RMSE since the errors are squared before the averaging, it gives high weights to large errors. Therefore, in applications where large errors are to be avoided, RMSE may be more helpful. ME is a straightforward measure that highlights the bias of the prediction at the image level by using the mean signed difference between the reference image and the fused image. Spatial accuracy is calculated using the difference in spatial features (such as texture and contrast) between a fused image and the reference image. Robert's edge (Edge) and local binary pattern (LBP) are used in this case study.

4. EXPERIMENTS AND RESULTS

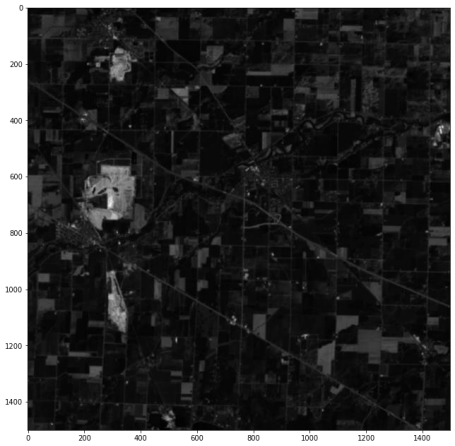
In the case study, changes in agricultural fields are being tracked. The test location is close to Ohio, (40.358615, -82.706838) USA. Two Landsat-8 images from the 04.08.2015 and the 20.08.2015 are used as the input for time t_0 and t_1 displaying the changes that occurred within two weeks along with a Sentine-2 image obtained on 04.08.2015 at time t_0 is used as inputs. The output is the predicted image at time t_1 as shown in (Figure: 3) for band 4 of both Landsat-8 and Sentinel-2. The 225 sq kilometers test site is mostly made up of agricultural land, although it is additionally diversified by roads, buildings, woods, and a river. The region is subjected to the logarithmic weighting function due to the region's increasing complexity. Given their linear relationship, comparing the observed and predicted surface reflectance shows that there is an acceptable overlap between the two. However, it is difficult to detect change in fields that are smaller than the coarse resolution sensor's pixel size.



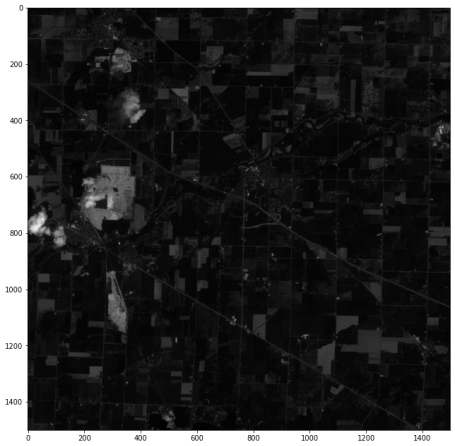
Landsat-8 (band 4) at t0



Sentinel-2 (band 4) at t0

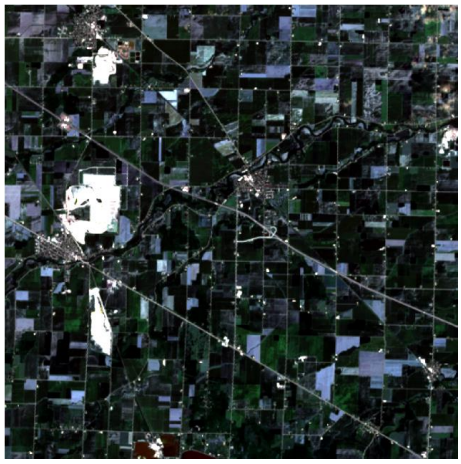


Landsat-8 (band 4) at t1

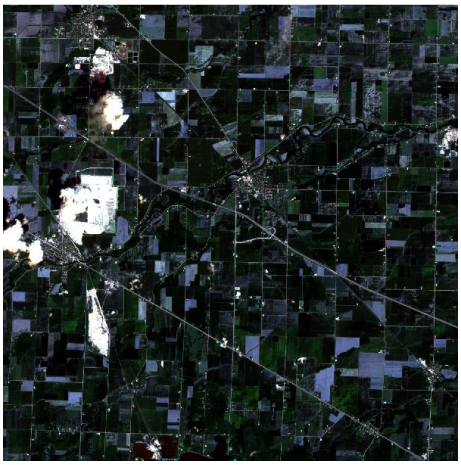


Predicted at t1(band 4) (STARFM)

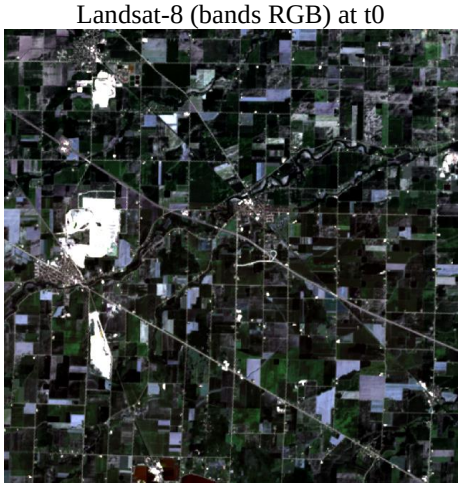
Figure 2: Landsat-8 (Band 4) and Sentinel-2 (Band 4) images acquired at t0 and t1 for the study area and the predicted image (Band 4) at t1 using STARFM.



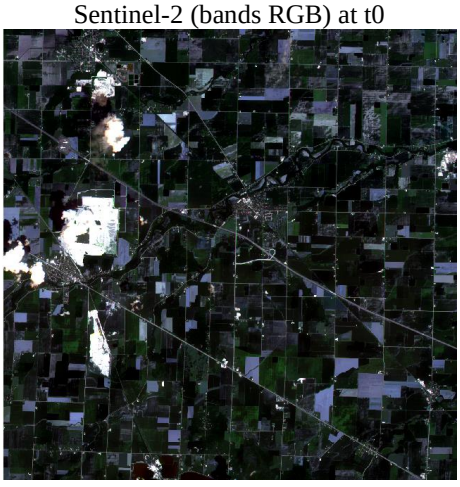
Landsat-8 (bands RGB) at t0



Sentinel-2 (bands RGB) at t0



Landsat-8 (bands RGB) at t1



Predicted at t1(bands RGB) (STARFM)

Figure 3: Landsat-8 (bands RGB) and Sentinel-2 (bands RGB) images acquired at t0 and t1 for the study area and the predicted image (bands RGB) at t1 using STARFM.

The quantitative analysis of the three bands used in the study is given in (Table 4). The composite of the RGB bands for the study area is given in (Figure:3). The characteristics of the image quality metrics used for the study for evaluating the fused images are briefly explained in (Table 5)

Table 4: Quantitative analysis results

Bands	RMSE	ME	Robert's Edge	LBP	BRISQUE
Band 2	0.00007068	-0.00004714	0.00727520	-0.00000352	42.33
Band 3	0.00000704	0.00000399	0.01199292	-0.00000802	39.67
Band 4	0.00000061	-0.00000001	0.01162911	-0.00000002	37.23

Table 5: Characteristics of Quality metrics for evaluating the fused image.

		Metric Name	Range	Interpretation of values
Reference Based Metrics	Spectral Aspect	RMSE	[0,1]	0 represents a perfect fused image; higher values for a fused image indicates higher spectral errors
		ME	[-1,1]	0 represents a perfect fused image; more negative shows an under-estimate of spectral information; more positive indicates over-estimate of spectral information
	Spatial Aspect	Robert's Edge	[-1,1]	0 represents a perfect fused image; over smoothed edge features in the fused image are indicated by more negative values; over sharpened edge features in the fused image are indicated by more positive
		LBP	[-1,1]	0 represents a perfect fused image; over smoothed textual features in the fused image9
Referenceless Metrics	Overall Aspect	BRISQUE	[0,100]	0 represents a perfect fused image;

5. CONCLUSION

In this study, a python implementation of STARFM is used to fuse the Landsat-8 and Sentinel-2 satellite images. The results are assessed based on four different full reference-based metrics namely, Local Binary Patterns (LBP), Root-Mean-Square Error (RMSE), Edge, and Mean Error (ME) which cover the spatial and spectral aspect of the fused image. A no-reference/referenceless image quality evaluation metric, Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) is also used to evaluate the overall fusion image quality. The geographic homogeneity of the ROI has a significant impact on the outcome of the fusion using STARFM. When only pure coarse resolution pixels are available, STARFM produces the best results. For scenarios where the specific classes are smaller than the coarse resolution pixel size, the searching window need to be increased and the weighting parameters need to be chosen such that it is more sensitive to spatial distance and less sensitive to spectral distance. STARFM was used on this small study area, which was roughly 225 square kilometers. The number of similar pixels identified might be increased and the outcome could be improved by running the algorithm on a bigger study area. For complex regions such as Land use/ Land Cover classes, satellite image fusion is challenging.

Reference

Belgiu, M., & Stein, A. (2019). Spatiotemporal image fusion in remote sensing. In *Remote Sensing* (Vol. 11, Issue 7). MDPI AG. https://doi.org/10.3390/rs11070818

Chen, B., Huang, B., & Xu, B. (2015). Comparison of spatiotemporal fusion models: A review. In *Remote Sensing* (Vol. 7, Issue 2, pp. 1798–1835). MDPI AG. https://doi.org/10.3390/rs70201798

Chen, Y., Cao, R., Chen, J., Zhu, X., Zhou, J., Wang, G., Shen, M., Chen, X., & Yang, W. (2020). A New Cross-Fusion Method to Automatically Determine the Optimal Input Image Pairs for NDVI Spatiotemporal Data Fusion. *IEEE Transactions on Geoscience and Remote Sensing*, *58*, 5179–5194.

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*(8), 4476–4488. https://doi.org/10.1109/TGRS.2017.2692802

Gao, F., Anderson, M. C., Zhang, X., Yang, Z., Alfieri, J. G., Kustas, W. P., Mueller, R., Johnson, D. M., & Prueger, J. H. (2017). Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sensing of Environment*, *188*, 9–25. https://doi.org/10.1016/J.RSE.2016.11.004

Gao, F., Masek, J., Schwaller, M., & Hall, F. (2006). On the blending of the landsat and MODIS surface reflectance: Predicting daily landsat surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(8), 2207–2218. https://doi.org/10.1109/TGRS.2006.872081

Ghamisi, P., Rasti, B., Yokoya, N., Wang, Q., Hofle, B., Bruzzone, L., Bovolo, F., Chi, M., Anders, K., Gloaguen, R., Atkinson, P. M., & Benediktsson, J. A. (2019). Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art. In *IEEE Geoscience and Remote Sensing Magazine* (Vol. 7, Issue 1, pp. 6–39). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/MGRS.2018.2890023

Li, J., Li, Y., He, L., Chen, J., & Plaza, A. (2020). Spatio-temporal fusion for remote sensing data: an overview and new benchmark. In *Science China Information Sciences* (Vol. 63, Issue 4). Science in China Press. https://doi.org/10.1007/s11432-019-2785-y

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*(9), 6552–6564. https://doi.org/10.1109/TGRS.2019.2907310

Mileva, N., Mecklenburg, S., & Gascon, F. (2018). *New tool for spatio-temporal image fusion in remote sensing: a case study approach using Sentinel-2 and Sentinel-3 data*. 20. https://doi.org/10.1117/12.2327091

Mittal, A., Moorthy, A. K., & Bovik, A. C. (2012). No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, *21*(12), 4695–4708. https://doi.org/10.1109/TIP.2012.2214050

Pohl, C., & van Genderen, J. L. (John L.). (n.d.). *Remote sensing image fusion : a practical guide*.

Racault, M. F., Sathyendranath, S., & Platt, T. (2014). Impact of missing data on the estimation of ecological indicators from satellite ocean-colour time-series. *Remote Sensing of Environment*, *152*, 15–28. https://doi.org/10.1016/j.rse.2014.05.016

Schmitt, M., & Zhu, X. X. (2016). Data Fusion and Remote Sensing: An ever-growing relationship. *IEEE Geoscience and Remote Sensing Magazine*, *4*(4), 6–23. https://doi.org/10.1109/MGRS.2016.2561021

Storey, J., Roy, D. P., Masek, J., Gascon, F., Dwyer, J., & Choate, M. (2016). A note on the temporary misregistration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) imagery. *Remote Sensing of Environment*, *186*, 121–122. https://doi.org/10.1016/j.rse.2016.08.025

Tang, J., Zeng, J., Zhang, L., Zhang, R., Li, J., Li, X., Zou, J., Zeng, Y., Xu, Z., Wang, Q., & Zhang, Q. (2020). A modified flexible spatiotemporal data fusion model. *Frontiers of Earth Science*, *14*(3), 601–614. https://doi.org/10.1007/s11707-019-0800-x

Wang, Q., Blackburn, G. A., Onojeghuo, A. O., Dash, J., Zhou, L., Zhang, Y., & Atkinson, P. M. (2017). Fusion of Landsat 8 OLI and Sentinel-2 MSI Data. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(7), 3885–3899. https://doi.org/10.1109/TGRS.2017.2683444

Wang, X., & Wang, X. (2020). Spatiotemporal Fusion of Remote Sensing Image Based on Deep Learning. *Journal*

of Sensors, *2020*. https://doi.org/10.1155/2020/8873079

Wu, H., Yang, Q., Liu, J., & Wang, G. (2020). A spatiotemporal deep fusion model for merging satellite and gauge precipitation in China. *Journal of Hydrology*, *584*. https://doi.org/10.1016/j.jhydrol.2020.124664

Xie, D., Gao, F., Sun, L., & Anderson, M. (2018). Improving spatial-temporal data fusion by choosing optimal input image pairs. *Remote Sensing*, *10*(7). https://doi.org/10.3390/rs10071142

Xie, D., Zhang, J., Zhu, X., Pan, Y., Liu, H., Yuan, Z., & Yun, Y. (2016). An improved STARFM with help of an unmixing-based method to generate high spatial and temporal resolution remote sensing data in complex heterogeneous regions. *Sensors (Switzerland)*, *16*(2). https://doi.org/10.3390/s16020207

Zeng, L., Wardlow, B. D., Xiang, D., Hu, S., & Li, D. (2020). A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sensing of Environment*, *237*, 111511. https://doi.org/10.1016/j.rse.2019.111511

Zhang, Q., Yuan, Q., Zeng, C., Li, X., & Wei, Y. (n.d.). *Missing Data Reconstruction in Remote Sensing image with a Unified Spatial-Temporal-Spectral Deep Convolutional Neural Network*.

Zhang, Y. (2008). *METHODS FOR IMAGE FUSION QUALITY ASSESSMENT-A REVIEW, COMPARISON AND ANALYSIS*.

Zhu, L., Radeloff, V. C., & Ives, A. R. (2017). Improving the mapping of crop types in the Midwestern U.S. by fusing Landsat and MODIS satellite data. *International Journal of Applied Earth Observation and Geoinformation*, *58*, 1–11. https://doi.org/10.1016/j.jag.2017.01.012

Zhu, X., Cai, F., Tian, J., & Williams, T. K. A. (2018). Spatiotemporal fusion of multisource remote sensing data: Literature survey, taxonomy, principles, applications, and future directions. In *Remote Sensing* (Vol. 10, Issue 4). MDPI AG. https://doi.org/10.3390/rs10040527

Zhu, X., Zhan, W., Zhou, J., Chen, X., Liang, Z., Xu, S., & Chen, J. (2022). A novel framework to assess all-round performances of spatiotemporal fusion models. *Remote Sensing of Environment*, *274*. https://doi.org/10.1016/j.rse.2022.113002