

New Wavelet Based Spatiotemporal Fusion Method

Amal Ibnelhobyb¹, Ayoub Mouak¹, Amina Radgui¹, Ahmed Tamtaoui¹, Ahmed Er-Raji²,
Driss El Hadani², Mohamed Merdas² and Faouzi Mohamed Smiej²

¹STRS Lab, Institut National des Postes et Télécommunication-INPT, Rabat, Morocco

²Centre Royal de Télédetection Spatiale-CRTS, Rabat, Morocco

{ibnelhobyb, mouak, radgui, tamtaoui}@inpt.ac.ma, {er-raji, elhadani, merdas, smiej}@crts.gov.ma

Keywords: Spatiotemporal fusion, Landsat, MODIS, NDVI, STARFM, ESTARFM, WSAD-FM, WESTARFM.

Abstract: Satellite image sensors are able to give images at high temporal resolution as the MODIS sensor that gives an image every day but with low spatial resolution, or at high spatial resolution as the Landsat sensor that gives images at 30m but with a revisit cycle of 16 days. Thus, these sensors are not able to give images with both high spatial and high temporal resolution. This need has become more and more absolute for many applications. Therefore spatiotemporal fusion methods were proposed. By applying these methods on images from different sensors with different spatial and temporal resolution, we can take the advantage of the high spatial and high temporal resolution of these sensors. As a result we get an image with both high spatial and high temporal resolution. We introduce in this paper a new method, the Wavelet base Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (WESTARFM), which is an improvement of the ESTARFM method. It uses the principle of wavelet transform with the original ESTARFM method. We have applied our method to predict daily NDVI in a study site in an irrigated zone in the region of TADLA in MOROCCO. Results have been compared with other methods.

1 INTRODUCTION

Satellite images are more and more used in many applications such as vegetation monitoring, ecosystem disturbance and land cover mapping. However, a tradeoff exists between spatial and temporal resolution in available satellite data. Satellite data obtained by moderate resolution sensors like the Moderate resolution Imaging Spectroradiometer (MODIS) gives daily observations of the entire earth but with a low spatial resolution attending 1 km (Gao et al., 2014). Whereas, data obtained by Landsat sensors gives more spatial details with a spatial resolution of 30 m but they have a long revisit cycle of 16 day and their use is limited by the presence of clouds. In order to get full use of advantageous characteristics of these sensors, fusion methods were proposed to combine satellite data from different sensors. By using spatiotemporal fusion we can obtain satellite images

with both high spatial and high temporal resolution. Many fusion methods have been proposed. They can be classified into four categories (Chen, Huang, & Xu, 2015): i. Transformation based methods (Ghannam, Awadallah, Abbott, & Wynne, 2014), ii. Learning based methods (Huang & Song, 2012)-(Song & Huang, 2013), iii. Reconstruction based methods (Gao, Masek, Schwaller, & Hall, 2006)-(Zhu, Chen, Gao, Chen, & Masek, 2010)-(Zhu et al., 2016)-(Hilker, Wulder, Coops, Linke, et al., 2009)-(Fu, Chen, Wang, Zhu, & Hilker, 2013), iv. Data assimilation based methods (Chemin & Honda, 2006). Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) (Gao et al., 2006) is one of the most common methods widely used for spatiotemporal fusion. It is a reconstruction based method that was proposed by Feng Gao on 2006. This method introduced the use of neighbouring pixels and windowing to predict Landsat-like images. However it was convenient only for homogenous regions. Enhanced STARFM (ESTARFM) (Zhu et al., 2010)

was proposed after that to overcome the limitation of STARFM and introduced a conversion coefficient that makes it applicable for heterogeneous regions using two pairs of Landsat and MODIS data. A wavelet base method has been used with the STARFM method (WSAD-FM) (Ghannam et al., 2014) it decompose the Landsat image at high and low frequencies and predict each part separately using only Landsat image for high frequencies and Landsat and MODIS images for low frequencies.

The paper presents a new fusion model based on the same concept of the WSAD-FM (Ghannam et al., 2014) but uses the wavelet transform with the ESTARFM (Zhu et al., 2010) method. This fusion method is applied on NDVI data in the region of Tadla in Morocco. We have used actual Landsat 8 NDVI and MODIS NDVI data for evaluation and calculate commonly used statistic parameters RMSE, AAD and R2 to compare the accuracy of our method with the STARFM, ESTARFM and WSAD-FM methods. First the theoretical basis and the proposed method will be introduced, after, evaluation of our method will be explained. At the end, the results of this evaluation will be discussed.

2 THEORETICAL BASIS

2.1 ESTARFM

As an improvement of the STARFM method (Gao et al., 2006), the Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model was proposed to overcome its limitation in prediction on heterogeneous and changing regions, this by using a conversion coefficient that presents the heterogeneity of coarse pixels (Chen et al., 2015). The STARFM supposes the presence of one land cover type on coarse pixels (pure pixels) which the case of homogenous regions. But for heterogeneous regions different land covers types are present on a coarse pixel. Therefore the ESTARFM considers the presence of mixed pixels and apply the linear mixture model to calculate the reflectance change of each present class. The sum of this changes is the change of the coarse mixed pixel between two days. It requires two pairs of Landsat and MODIS images and a MODIS image from the predicted day. It is described in the following steps:

- For a fine central pixel of a moving window we use thresholding or a classification map to find spectrally similar pixels.
- A weighting function W_i is calculated for these n similar pixels after being filtered. This

weighting function is based on spectral similarity, temporal difference and spatial distance.

- Calculate the conversion coefficient v_k presenting the ratio of reflectance change for a class k represented by the fine pixel L_k to the reflectance change of the coarse pixel M between two days t_m and t_n :

$$v_k = \frac{L_{kn} - L_{km}}{M_n - M_m} \quad (1)$$

- Predict the Landsat image L at time t_p for the central pixel $(x_{w/2}, y_{w/2})$ within the moving window of size w , based on pair input L and M from time t_m to have L_m and from time t_n to have L_n :

$$\begin{aligned} L_m(x_{w/2}, y_{w/2}, t_p) = & \\ & L(x_{w/2}, y_{w/2}, t_m) + \\ & \sum_{i=1}^N W_i \cdot v_i \cdot \left(\begin{array}{l} M(x_i, y_i, t_p) - \\ M(x_i, y_i, t_m) \end{array} \right) \end{aligned} \quad (2)$$

$$\begin{aligned} L_n(x_{w/2}, y_{w/2}, t_p) = & \\ & L(x_{w/2}, y_{w/2}, t_n) + \\ & \sum_{i=1}^N W_i \cdot v_i \cdot \left(\begin{array}{l} M(x_i, y_i, t_p) - \\ M(x_i, y_i, t_n) \end{array} \right) \end{aligned} \quad (3)$$

- The predictions at times t_m and t_n are summed and temporally weighted to calculate the final prediction. This temporal weight T presents the contribution of each pair:

$$\begin{aligned} L(x_{w/2}, y_{w/2}, t_p) = & T_m \times L_m(x_{w/2}, y_{w/2}, t_p) \\ & + T_n \times L_n(x_{w/2}, y_{w/2}, t_p) \end{aligned} \quad (4)$$

2.2 Wavelet-based ESTARFM

In this section we present our proposed method, the Wavelet based Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model used to combine data with different spatial resolution and different temporal resolution in order to predict an image with both high spatial and high temporal resolution.

Our method is based on the ESTARFM method but it uses the wavelet transform to predict more details on the image (Ghannam et al., 2014). In the original ESTARFM the whole image is used for prediction, but in our proposed method the Landsat image is decomposed into high and low frequencies using the wavelet transform. After that each component is used for prediction separately. As the ESTARFM our method requires two pairs of Landsat and MODIS data and a MODIS image from the prediction day. The WESTARFM is implemented following these steps:

- The two Landsat images from time t_m and t_n are decomposed into high and low frequencies using wavelet transform.
- Predict approximation coefficients L_a of Landsat image at time t_p using low frequency components of Landsat images at time t_m and t_n and MODIS images at time t_m , t_n and t_p :

$$L_{am}(x_{w/2}, y_{w/2}, t_p) = L_a(x_{w/2}, y_{w/2}, t_m) + \sum_{i=1}^N W_i \cdot v_i \cdot (M(x_i, y_i, t_p) - M(x_i, y_i, t_m)) \quad (5)$$

$$L_{an}(x_{w/2}, y_{w/2}, t_p) = L_a(x_{w/2}, y_{w/2}, t_n) + \sum_{i=1}^N W_i \cdot v_i \cdot (M(x_i, y_i, t_p) - M(x_i, y_i, t_n)) \quad (6)$$

- Predict detail coefficients L_d of Landsat images at time t_p using only high frequency components of Landsat images at time t_m and t_n :

$$L_{dm}(x_{w/2}, y_{w/2}, t_p) = \sum_{i=1}^N W_i \cdot L_d(x_{w/2}, y_{w/2}, t_m) \quad (7)$$

$$L_{dn}(x_{w/2}, y_{w/2}, t_p) = \sum_{i=1}^N W_i \cdot L_d(x_{w/2}, y_{w/2}, t_n) \quad (8)$$

- Calculate the final images at time t_m and t_n by applying the inverse wavelet transform on the predicted low frequency and high-frequency components.
- Estimate the final predicted image at time t_p using temporal weight T (Zhu et al., 2010).

3 RESULTS AND DISCUSSION

3.1 Data and Pre-processing

High Spatial Resolution Landsat 8 images and High Temporal Resolution MODIS images are required for evaluating the accuracy of the proposed method. In order to generate a daily prediction of NDVI, the MODIS Surface Reflectance Daily 250m (MOD09GQ bands 1,2) was selected. As shown in Figure 1. Eight Landsat 8 and MODIS images representing days: 05 June 2015(156), 21 June 2015(172), 07 July 2015(188), 23 July 2015(204), 08 August 2015(220), 24 August 2015(236), 09 September 2015(252) and 25 September 2015(268) were selected for evaluation. The Landsat 8 images were downloaded from the USGS GLOVIS website, and the MODIS Surface Reflectance data were downloaded from the Reverb ECHO website. This Data cover an irrigated area of 800m x 800m in the region of Tadla (32° 28 N, 5° 38 W) situated in central Morocco.

Erdas imaging were used as a preprocessing tool to calculate the Landsat 8 Surface Reflectance from Landsat 8 data, and to generate the Landsat NDVI images from Landsat 8 reflectance and MODIS NDVI from MODIS Surface Reflectance. Landsat and MODIS input data should have the same projection and same pixel size, thus Erdas imaging was used for UTM projection and ARCMAP was used for resampling of MOD09 in order to have the same resolution as Landsat images (30m).

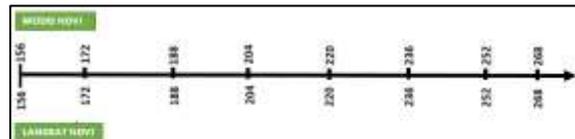


Figure 1: DOYs of Landsat 8 and MODIS Data used for evaluation.

3.2 Evaluation and Results

The proposed method was used for predicting Landsat NDVI images. This method is applied on NDVI images since it gives more accurate results and less complexity than applying the fusion methods on RED and NIR bands used for calculating NDVI (Jarihani et al., 2014)

Two pairs of Landsat NDVI and MODIS NDVI images are needed with the MODIS NDVI from the prediction day, in the original method of ESTAFM the author used the pairs of Landsat and MODIS from the start and the end of the period as input of the method, for our evaluation we have used the two pairs from previous days. For WSADAFM method only one pair of Landsat and MODIS NDVI is needed.

The prediction was performed using eight Landsat and MODIS NDVI images in DOYs 156, 172, 188, 204, 220, 236, 252, 268 when Landsat NDVI images are available in order to evaluate the accuracy of the

proposed method and compare it with STARFM, ESTARFM and WSAD-FM methods.

Figure 2 shows the prediction results of the methods for the selected days compared with real images. We have the results for all the days for STARFM, ESTARFM, WSAD-FM and WESTARFM except DOYs 156 and 172 since they are used as first inputs to the fusion. Visually, we can see that the Landsat-like images contain all the details of the region. However, for some days, as an example DOY 220, details of the image are lost in some zones and the image is noisy as a result of clouds. The quality of prediction depends on the quality of input images(Hilker, Wulder, Coops, Seitz, et al., 2009), good results can be obtained if Landsat and MODIS NDVI input images are clean from clouds also results are affected by the inconsistency existing between the Landsat and MODIS sensors(Gevaert & García-Haro, 2015) .

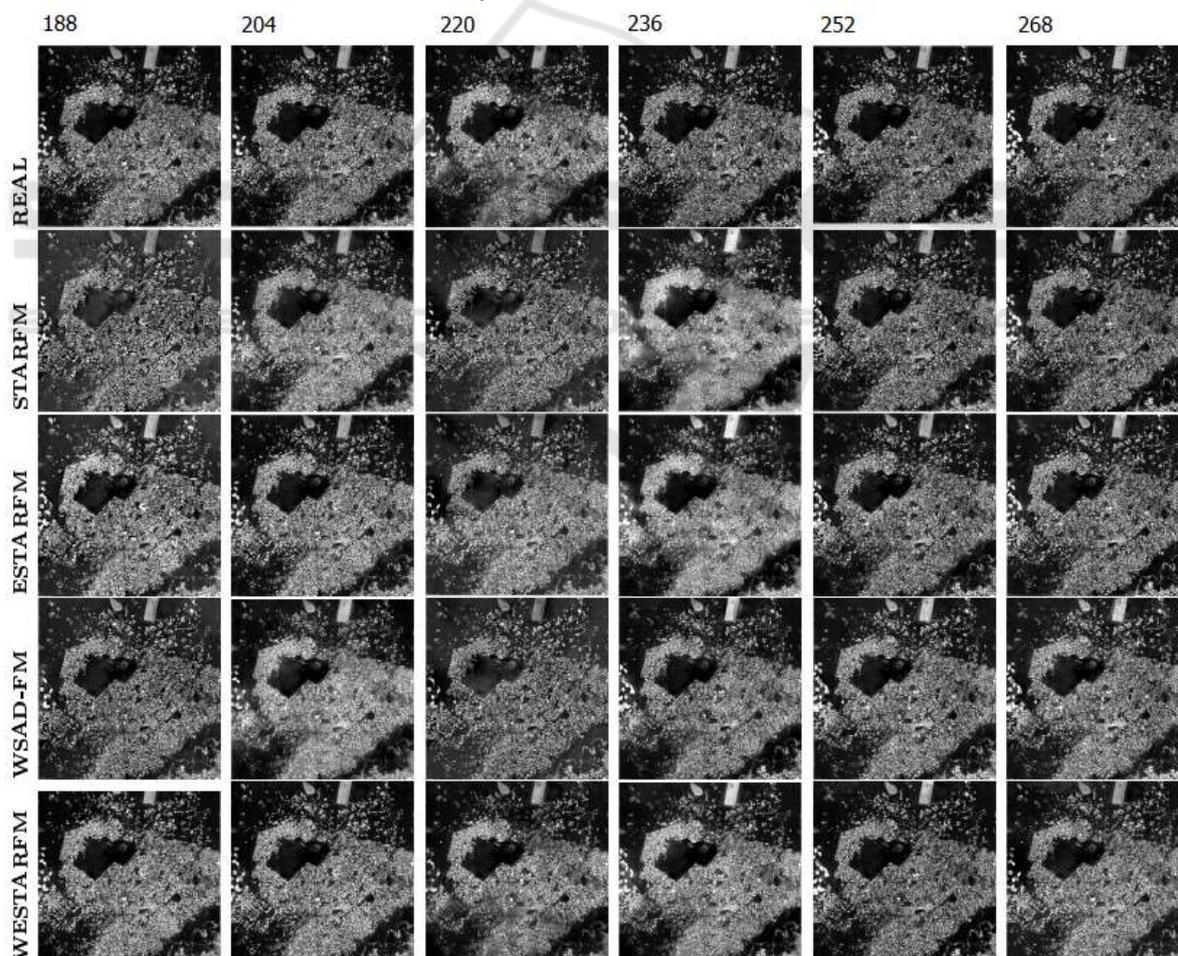


Figure 2: NDVI prediction results using ESTAFM, WSADFM and WESTARM.

Figures 3, 4, 5, 6 show real Landsat NDVI and predicted Landsat NDVI using our proposed method WESTARFM for DOY 252. We can see that the predicted image contains most of the details and visually it is almost similar to the real image. If we zoom in a particular zones like a citrus zone, a rainfed zone and an irrigated zone, as it is illustrated on figures (5,6,8,9,11,12), we notice, visually, that the WESTARFM method was able to predict almost the same NDVI information as the real image.

Figures 7, 10 and 13 show the correlation between the real and predicted NDVI values using the WESTARFM method for the 3 selected zones at time 252. Results show that the prediction is better for the citrus zone. We can say that one of the parameters that affect the performance of the prediction is the value of input NDVI, more the value of NDVI is high more the prediction is better. We assume also that the prediction is better for citrus zone since it is more stable between the input and the prediction days than culture areas.

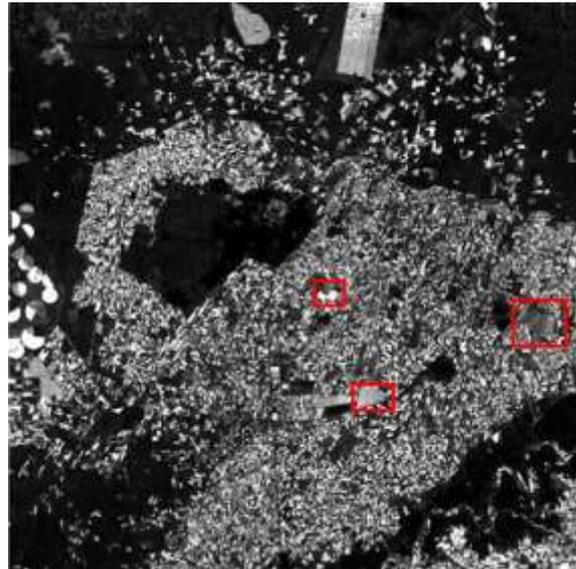


Figure 4: Predicted Landsat NDVI DOY 252 using WESTARFM.

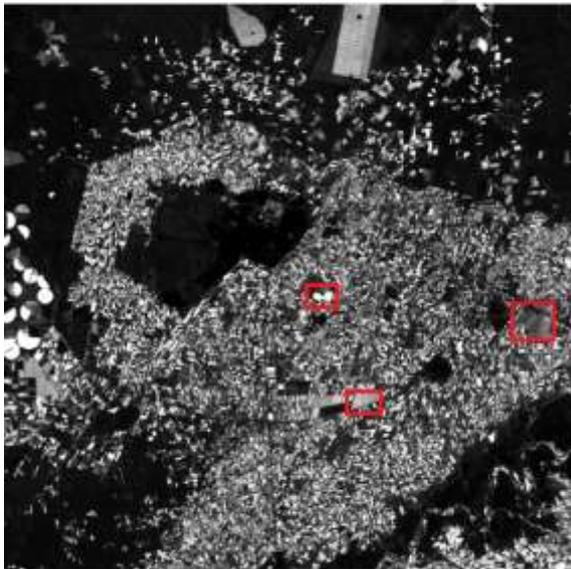


Figure 3: Real Landsat NDVI DOY 252.



Figure 5: Zoom in a citrus zone in real Landsat NDVI image DOY 252.



Figure 6: Zoom in a citrus zone in predicted Landsat NDVI image using WESTARFM DOY 252.

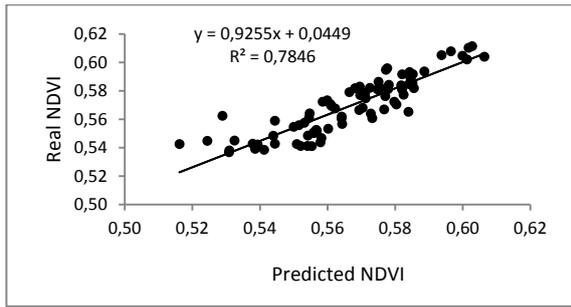


Figure 7: Correlation between real and predicted Landsat NDVI in a citrus zone DOY 252



Figure 11: Zoom in a rainfed zone in real Landsat NDVI image DOY 252.



Figure 8: Zoom in an irrigated zone in real Landsat NDVI image DOY 252.



Figure 12: Zoom in a rainfed zone in predicted Landsat NDVI image using WESTARFM DOY 252.



Figure 9: Zoom in an irrigated zone in predicted Landsat NDVI image using WESTARFM DOY 252.

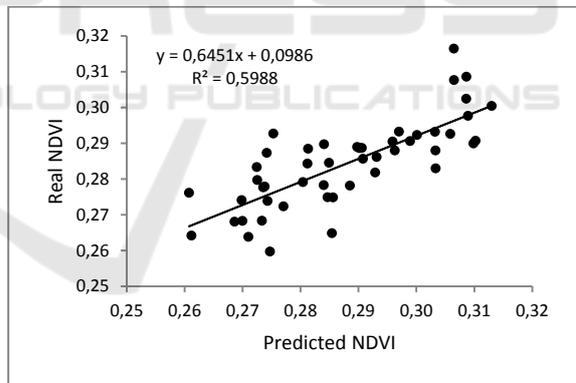


Figure 13: Correlation between real and predicted Landsat NDVI in a rainfed zone DOY 252

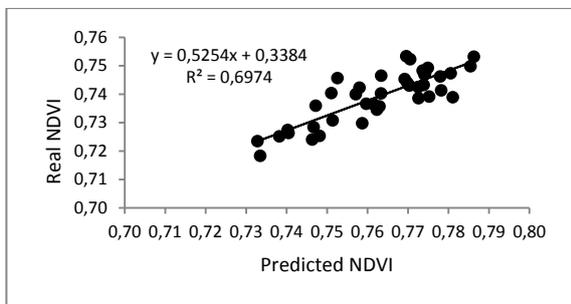


Figure 10: Correlation between real and predicted Landsat NDVI in an irrigated zone DOY 252.

We have calculated the Root Mean Square Error (RMSE), the Absolute Average Difference (AAD) and the Coefficient of Determination (R2) to validate the prediction results of the three methods.

Results presented in table 1 have shown that the WESTARFM method gives more accurate results than the ESTARFM and WSAD-FM with a RMSE attending 0.05, AAD of 0.02 and a R2 of 0.64

4 CONCLUSION

A new fusion model found on Wavelet transform and ESTARFM method was presented (WESTARFM). The model utilizes the Wavelet transform to decompose the Landsat data into approximation and detail coefficients. Each of these components is, after that, predicted separately with the ESTARFM method. Two pairs of Landsat and MODIS NDVI from previous days and a MODIS NDVI from

prediction date were needed as inputs of the WESTARFM to predict an unavailable Landsat NDVI image. The WESTARFM was tested on NDVI and compared with other methods. Results have shown that the proposed method gives more accurate results for most of evaluated dates. Therefor working on frequency domain improves the prediction and predicts more image details. This method was tested on NDVI but it can be applicable also on Landsat and MODIS bands.

Table 1: Statistic validation of prediction results using the four fusion methods.

DOY		STARFM	ESTARFM	WSAD-FM	WESTARFM
188	RMSE	0,11	0,09	0,08	0,06
	AAD	0,08	0,07	0,06	0,05
	R2	0,39	0,42	0,52	0,64
204	RMSE	0,08	0,06	0,09	0,06
	AAD	0,06	0,03	0,06	0,03
	R2	0,25	0,20	0,34	0,37
220	RMSE	0,07	0,06	0,09	0,08
	AAD	0,05	0,03	0,06	0,06
	R2	0,38	0,26	0,32	0,29
236	RMSE	0,12	0,08	0,11	0,06
	AAD	0,02	0,06	0,08	0,05
	R2	0,34	0,24	0,24	0,41
252	RMSE	0,08	0,07	0,09	0,05
	AAD	0,04	0,04	0,07	0,02
	R2	0,28	0,25	0,31	0,41
268	RMSE	0,07	0,06	0,1	0,09
	AAD	0,03	0,03	0,06	0,05
	R2	0,22	0,19	0,55	0,6

REFERENCES

- Chemin, Y., & Honda, K. 2006. Spatiotemporal fusion of rice actual evapotranspiration with genetic algorithms and an agrohydrological model. *IEEE Transactions on Geoscience and Remote Sensing*, 44(11), 3462–3469
- Chen, B., Huang, B., & Xu, B. 2015. Comparison of Spatiotemporal Fusion Models: A Review. *Remote Sensing*, 7(2), 1798–1835.
- Fu, D., Chen, B., Wang, J., Zhu, X., & Hilker, T. 2013. An improved image fusion approach based on enhanced spatial and temporal the adaptive reflectance fusion model. *Remote Sensing*, 5(12), 6346–6360.
- Gao, F., Hilker, T., Zhu, X., Anderson, M., Masek, J., Wang, P., & Yang, Y. 2014. Monitoring, *Fusing Landsat and MODIS data for vegetation*, (september), 47–60.
- Gao, F., Masek, J., Schwaller, M., & Hall, F. 2006. On the Blending of the MODIS and Landsat ETM + Surface Reflectance, 20771(2), 20794.
- Gevaert, C. M., & Garcia-Haro, F. J. 2015. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. *Remote Sensing of Environment*, 156, 34–44.
- Ghannam, S., Awadallah, M., Abbott, a. L., & Wynne, R. H. 2014. Multisensor Multitemporal Data Fusion Using Wavelet Transform. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-1(November)*, 121–128.
- Hilker, T., Wulder, M. a., Coops, N. C., Linke, J., McDermid, G., Masek, J. G., White, J. C., Gao, F. 2009. A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sensing of Environment*, 113(8), 1613–1627.

- Hilker, T., Wulder, M. A., Coops, N. C., Seitz, N., White, J. C., Gao, F., Masek, J. G., Stenhouse, G. 2009. Generation of dense time series synthetic Landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model. *Remote Sensing of Environment*, 113(9), 1988–1999.
- Huang, B., & Song, H. 2012. Spatiotemporal reflectance fusion via sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 50(10 PART1), 3707–3716.
- Jarihani, A. A., McVicar, T. R., van Niel, T. G., Emelyanova, I. V., Callow, J. N., & Johansen, K. 2014. Blending landsat and MODIS data to generate multispectral indices: A comparison of “index-then-blend” and “Blend-Then-Index” approaches. *Remote Sensing*, 6(10), 9213–9238.
- Song, H., & Huang, B. 2013. Spatiotemporal satellite image fusion through one-pair image learning. *IEEE Transactions on Geoscience and Remote Sensing*, 51(4), 1883–1896.
- 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(11), 2610–2623.
- 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.

