International Cartographic Association
Association Cartographique Internationale

A spatio-temporal fusion method based on Landsat and MODIS normalized vegetation index data

Xianghong Che[a, *], Jiping Liu[a], Yong Wang[a], Qing Sun[b]

[a]*Research Center of Geospatial Big Data Application, Chinese Academy of Surveying and Mapping, chexh.15b@gmail.com, liujp@casm.ac.cn, wangy@casm.ac.cn*
[b] *State Key Laboratory of Severe weather (LASW), Chinese Academy of Meteorological Science, sunqingmeteo@gmail.com*

* Corresponding author

Abstract: Normalized difference vegetation index (NDVI) can simply and effectively reflect the growth status of plants, and has a linear relationship with vegetation coverage, thus it is an important indicator to identify vegetation growth status and coverage. However, the commonly used MODIS NDVI or Landsat NDVI cannot simultaneously achieve the high temporal and spatial resolution. To this end, we proposed a generic and automatic method to fuse MODIS and Landsat satellite vegetation index (NDVI) into daily high-resolution products which takes time series NDVI data from MODIS/Landsat as input, filters noise pixels, and generates high-resolution and high-frequency products through a fusion process of temporal-linearly interpolation to increase the temporal resolution and spatial filtering to alleviate the blocky artifacts from MODIS pixel border. The fused NDVI maps are visually compared to evaluate the performance of this method, and more quantitative assessment will be continued in the next step. This study can provide important support for large-scale long-term surface vegetation monitoring.

Keywords: Spatio-temporal fusion, MODIS, Landsat, NDVI, linear interpolation

1. Introduction

Satellite missions have been proved to be a popular and feasible tool for the Earth real-time observation at a large scale, but there were always trade-offs between spatial resolution and temporal frequency. One satellite sensor is not able to observe the Earth with high spatial resolution and high revisiting frequency. Optical sensors with coarse spatial resolution and daily orbits, for example, the Advanced Very-High-Resolution Radiometer (AVHRR), Moderate Resolution Imaging Spectroradiometer (MODIS), and Visible Infrared Imaging Radiometer Suite (VIIRS), can detect surface changes at the daily temporal resolution (Lunetta et al. 2006; Moon et al. 2019; Young and Wang 2001; Zhan et al. 2002). However, the relatively coarse (> 100 m) resolutions of these data limit their capability for monitoring small-patch surface features such as water and agricultural-related applications which require a few meters of spatial information. In the contrast, high or medium-resolution satellite data have a much finer spatial resolution (e.g. Landsat: 30 m, Sentinel-2: 10 m), but the sampling frequency is low (~multiple days to weeks), thus making it difficult to monitor the rapid and subtle change of land cover (Gao et al. 2017; Hansen et al. 2013; Imhoff et al. 2010; Stone Jr and Rodgers 2001).

In order to alleviate the above dilemma, spatial-temporal fusion methods of multi-source remote sensing data have achieved popular applications which can be divided into two categories. One is based on a linear spectral mixture model where the pixel values of low spatial resolution remote sensing data can be regarded as a linear combination of pixel values of medium and high spatial resolution data. Assuming that the pixel values of the high spatial resolution data with the same land cover type are the same, using the abundance matrix of each type obtained from the medium and high spatial resolution remote sensing data, and the least square method, the medium and high spatial resolution can be calculated from the low spatial resolution pixel reflectance (Cherchali et al. 2000; Haertel and Shimabukuro 2005; Maselli et al. 1998; Oleson et al. 1995). However, based on the linear spectral mixture model, the calculated pixel values with medium and high spatial resolutions are only the average values of each land cover type, but not real pixel values which have limitations for surface vegetation with high spatial heterogeneity. Another is the spatial and temporal adaptive reflectance fusion model (STARFM) developed by Gao et al. (Gao et al. 2006) to avoid solving the linear spectral mixture model and consider the spatial variability of pixel values. Subsequent to the STARFM model, a series of improved methods have been successively proposed such as STAARCH (Spatial Temporal Adaptive Algorithm for mapping Reflectance Change) (Hilker et al. 2009), ESTARFM (Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model) (Zhu et al. 2010), FSDAF (Flexible Spatiotemporal DAta Fusion) (Zhu et al. 2016).

The above spatio-temporal data fusion method effectively improves the spatio-temporal resolution of images, but there are also limitations. For example, the input MODIS-Landsat image pairs need to have better image quality and are closer to the target time. Affected by contaminated values such as clouds/cloud shadows, these images are difficult to use for fusion which is hard to achieve automatic and efficient fusion for large-scale long-term surface vegetation monitoring. Therefore, in this study, we present a generic and automatic method to fuse MODIS

and Landsat satellite vegetation index (NDVI) into daily high-resolution products.

2. Data source

We collected Landsat-8 OLI DN images (i.e. Landsat-5 TM, Landsat-7 ETM+, Landsat-8 OLI) with 30m spatial resolution but the low temporal resolution (16-day revisit cycle) in the year of 2018, which were ordered and downloaded from USGS Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA). The Landsat images were processed to surface reflectance (SR) using Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) for TM/ETM+ images and Landsat Surface Reflectance Code (LaSRC) for OLI images. Fmask was applied to identify cloud and cloud shadow in each Landsat image.

MODIS MCD43A4 Nadir BRDF (Bidirectional Reflectance Distribution Function) -Adjusted SR (NBAR) products on 2018 were used, which have a more frequent revisit cycle (daily) but a coarse spatial resolution (~500 m). MODIS MCD43A4 data have been adjusted and normalized to be nadir view at local solar noon time using the BRDF correction. The Landsat SR has not been BRDF-corrected, but due to the small field of view in Landsat (±7.5° from nadir), Landsat can be largely regarded as in nadir view. Using MODIS/Landsat SR data, the vegetation index (Normalized Difference Vegetation Index, NDVI) was derived using Equation (1). In this study, we only fuse MODIS and Landsat NDVI, because the band ratio of NDVI can help to reduce some system uncertainties from the individual near-infrared and visible red band such as BRDF effect and sensor errors.

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \quad (1)$$

Where ρ_{nir} and ρ_{red} were SR of near-infrared and visible red bands from MODIS/Landsat data.

3. Methodology

We present a generic method to fuse MODIS and Landsat satellite vegetation index (NDVI) into daily high-resolution products (Figure 1). This method takes time series NDVI data from MODIS/Landsat as input, filters noise pixels, and generates high-resolution and high-frequency products through a fusion process.

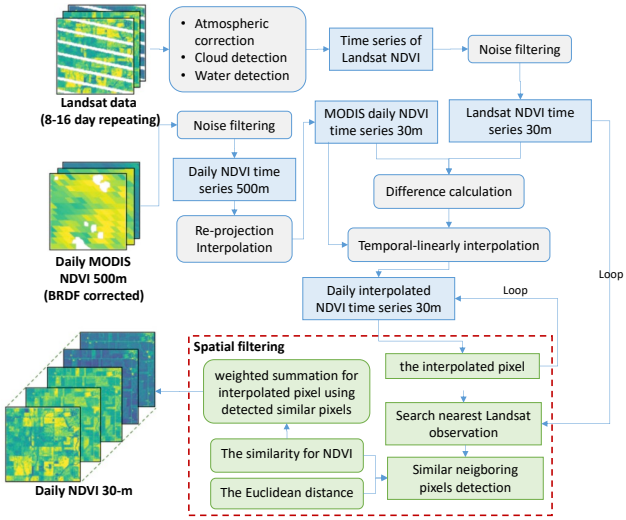

Figure 1. The flowchart of spatial-temporal fusion algorithm

3.1 Pre-processing of MODIS/Landsat NDVI time series

In this study, we assume that observed local minima of NDVI time series are often artifacts resulting from atmospheric effect or snow cover, particularly during winter months. The NDVI values less than 0.1 during winter months were filtered for MODIS/Landsat NDVI. For MODIS NDVI time-series with daily resolution, we utilized a medium method with a time step of 9 to reduce contaminated pixels. Compared to MODIS NDVI time series, the Landsat NDVI time series with 16-day resolution is much sparser, and the trend of NDVI time series will probably be shifted with the medium method. Therefore, a filter rule was designed for Landsat NDVI data.

Specifically, the quality layer of Landsat SR dataset was used to mask out the cloud, snow/ice and water pixels, and then retrieved masked NDVI. However, there existed some undetected-contaminated pixels. Based on the assumption, the local minima (NDVIi) of Landsat NDVI time series were detected as artifacts if it applied the equation (2-3).

$$NDVI_i < NDVI_{i+j} \quad (2)$$

$$NDVI_i < AVERAGE(NDVI_{i+j}) - STD(NDVI_{i+j}) \quad (3)$$

Where, $NDVI_{i+j}$ indicated four neighbouring NDVI values centered on $NDVI_i$ (j = -2, -1, 1 and 2). AVERAGE is a function calculating the average value of neighbouring NDVI values, and STD is corresponding standard deviation.

3.2 Temporal-linearly interpolation

The goal of our fusion framework is to fuse Landsat and MODIS NDVI data into images with both high-resolution spatial information and frequent temporal coverage for n dates T_i (i =1, 2, …, n), and there are k matching pairs of Landsat and MODIS images that acquired on the same dates T_j (j =1, 2, …, k) among the n dates. MODIS NDVI pixels were aligned and super-sampled to Landsat projection and pixel footprint to produce each Landsat-

MODIS matching pair. The MODIS data M(x, y, T_i) is also available for each date T_i (i =1, 2, …, n) where we will predict the fine-resolution image.

Specifically, consider one Landsat image L and one MODIS image M that are acquired on the same date T_j, the relationship between the Landsat image and the MODIS image can be modelled as

$$L_{(x,y,T_j)} = M_{(x,y,T_j)} + \Delta_{(x,y,T_j)} \quad (4)$$

Where (x, y) is the location of the aligned pixel of Landsat and MODIS data, T_j is the acquisition date for Landsat and MODIS data, andΔ(x, y, T_j) is the difference between Landsat and MODIS data of position (x, y).

Suppose we aim to predict a fine-resolution image on date t_p where only the MODIS data M(x, y, T_p) is available but the Landsat data is unknown. The predicted image I (x, y, T_p) can be represented as Equation (4), where Δ(x, y, T_p) is the difference between the MODIS image M and the predicted fine-resolution image L of position (x, y) at date T_p. To obtain the difference Δ(x, y, T_p) of each position (x, y) at date T_p, we integrate all the k available matching pairs of Landsat and MODIS data on dates from T_1 to T_k, and calculate the difference (Δ(x, y, T_j)) between Landsat and MODIS data for each date T_j (j =1, …, k). We then linearly interpolate the difference ϵ(x, y, T_j) into each date T_i, and hence obtain the estimated difference Δ(x, y, T_p) for all positions (x, y) at date T_p. To obtain the prediction of the fine-resolution image I (x, y, T_p), we use MODIS data at date T_p to further correct the spatial information encoded in Δ(x, y, T_p) with Equation (5). This step can be thought as an adjustment step that incorporates the high-frequency temporal patterns provided by MODIS data.

$$I(x,y,T_p) = M(x,y,T_p) + \Delta(x,y,T_p) \quad (5)$$

$$\Delta(x,y,T_p) = \Delta(x,y,T_{i1}) + \frac{(T_{i2}-T_{i1})*(\Delta(x,y,T_{i2})-\Delta(x,y,T_{i1}))}{T_{i2}-T_{i1}} \quad (6)$$

Where T_p is the predicted date. T_{i1} and T_{i2} are the before and after T_p dates where MODIS and Landsat NDVI exist. $\Delta(x,y,T_{i2})$ and $\Delta(x,y,T_{i1})$ can be derived with Equation (4).

3.3 Spatial filtering (SF)

The inconsistency between the spatial resolutions of MODIS and Landsat NDVI image means that the interpolation prediction is dominated by blocky artifacts which correspond with a MODIS pixel. To deal with the blocky artifacts problem in the interpolation prediction, SF is considered in the second step. In this study, the similar neighboring pixels in a local window were used to remove blocky artifacts.

Since there already exist blocky artifacts in the interpolation prediction, it is inappropriate to use the predicted image at unknown t_2 to search for similar neighbouring pixels. Alternatively, the clear Landsat NDVI observation at the closest and known t1 was used, based on the assumption of stable land cover boundaries (or very few changes in boundaries) occurring from t_1 to t_2. The NDVI difference between a neighbouring pixel at x_i and the center pixel at x_0 is calculated as

$$NDVI_d(t_1) = NDVI(x_i,t_1) - NDVI(x_0,t_1) \quad (5)$$

In the local window with w by w fine pixels, the first n pixels with smallest $NDVI_d$ (excluding the center pixel itself) were identified as similar neighbors and were selected for SF. If the similar pixels for the target interpolated pixel are not found in the Landsat observation at t_1, we will find the next closest Landsat observation, and this process will stop until the similar pixels are detected on the Landsat observation at t_j or all of Landsat observations are looped. Using the trial and error method, w and n were set to 3 and 4 respectively. For each interpolated 30 m pixel at t_2, the SF prediction was determined as the linear combination of the interpolation prediction of similar neighbouring pixels.

$$NDVI_{sf}(x_0,t_2) = \sum_{i=1}^{n} W_i NDVI_{inerp}(x_i,t_j) \quad (6)$$

Where W_i is a weight determined according to the spatial distance between the neighbouring and center pixels. t_j is the acquisition date of closed Landsat observation that is able to detect similar pixels for interpolated pixels. Based on spatial dependence, spatially closer pixels are more likely to have similar NDVI to the center pixel and, thus, receive larger weights

$$W_i = \frac{(1/d_i)}{\sum_{i=1}^{n}(1/d_i)} \quad (7)$$

$$d_i = 1 + \frac{\sqrt{|x_i - x_0|^2}}{w/2} \quad (8)$$

The distance d_i needs to be constrained to an appropriate range to exert reasonable influence on the weight W_i. Thus, 1 and w / 2 are used in Equation (8) and di ranges from 1 to 1 + $\sqrt{2}$ correspondingly. With the filtering scheme in Equation (6), the blocky artifacts in the interpolation prediction can be alleviated.

4. Results and discussion

In order to evaluate the performance of a proposed method, we selected a test area at the p034r032 for the visual comparison. Figure 2(a) is the high-resolution google map, and Figure 2(b) and 2(c) are the known Landsat 8 NDVI map at 30m resolution on August 6 and 22 2018, respectively. Figure (d) is the MCD43A4 NDVI map on August 12, 2022 used as a reference to generate the fused 30m NDVI map. Figure 2(e) is the intermediate temporal-linearly interpolated 30m NDVI map, and Figure 2(f) is the final fused 30m NDVI map with spatial filtering on August 12, 2022. It is found that the fused NDVI map has similar values distribution as the referenced MODIS NDVI map, but the spatial resolution is significantly improved, and bad pixels from MODIS NDVI can be restored. The blocky artifacts cannot be visually distinguished on the whole test area.

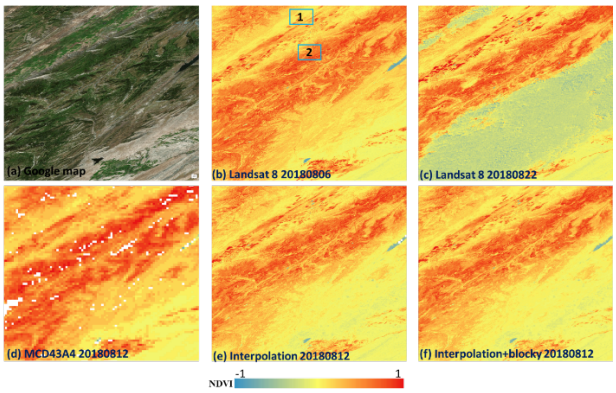

Figure 2. The spatio-temporal fusion map on August 12, 2018 for test area of the p034r032 tile (white pixels is the bad pixels from MODIS SR)

Figure 3 and 4 were used to display the blocky artifacts which were two zoom-in areas from Figure 2. Since there was an obvious spatial resolution difference between the 500m MODIS NDVI and 30m Landsat NDVI, the blocky artifacts tend to exist along the border of one MODIS pixel as Figure 3(e) and Figure 4(e). Through spatial filtering, these artifacts were eliminated to make the region more homogenous, which demonstrated the necessity of spatial filtering for wall-to-wall fused map.

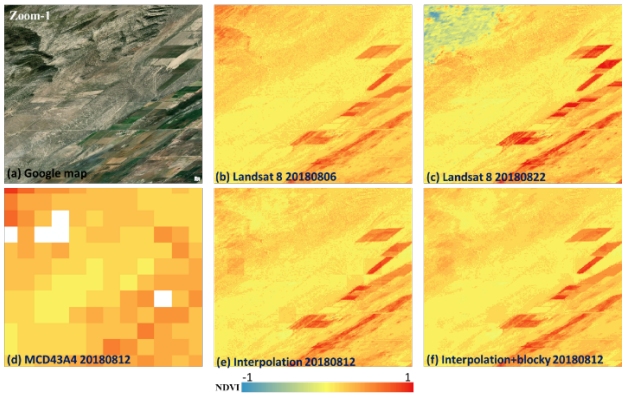

Figure 3. Same as Figure 2 but for the Zoom 1 area

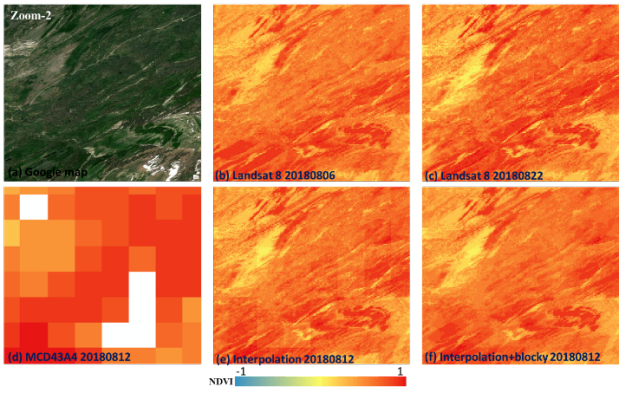

Figure 4. Same as Figure 2 but for the Zoom 2 area

In this study, we only visually evaluated the fused results because the observed 30m NDVI non-exist for the fused date. In the next step, we will concentrate on the quantitative assessment. Sentinel-2 data can be used as the observation NDVI to compare with the fused NDVI (Drusch et al. 2012). Furthermore, we can use the commonly-used fusion method (i.e., STARFM) (Gao et al. 2006) as a benchmark to evaluate the performance of this proposed method. In addition, this study tried to select the clear-sky Landsat data, but the invalid long gap of time series from Landsat NDVI can bring in the uncertainty for the temporal-linearly interpolation procedure which should be filled before the interpolation (Luo et al. 2018; Yan and Roy 2020). In this way, the relatively precise intermediate fused map can be produced to help improve the fusion accuracy.

5. Conclusions

Considering the limitation of the high input requirements of the existing spatio-temporal fusion method for data fusion, we proposed an automatic spatiotemporal fusion method of surface NDVI data through temporal linear interpolation and spatial filtering processing to obtain both high spatial and temporal resolution NDVI data. Above all, after filtering the Landsat NDVI data with FMASK, the noise secondary filtering based on adjacent NDVI observations and the noise filtering of the MODIS NDVI median are designed, which is crucial for NDVI spatio-temporal fusion input, and the high quality image pairs are not needed. This study can provide technical reference for large-scale long-term surface vegetation monitoring.

6. Acknowledgements

Hidden for review

7. References

Cherchali, S., Amram, O., & Flouzat, G. (2000). Retrieval of temporal profiles of reflectances from simulated and real NOAA-AVHRR data over heterogeneous landscapes. International Journal of Remote Sensing, 21, 753-775

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., & Martimort, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sensing of Environment, 120, 25-36

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

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, 2207-2218

Haertel, V.F., & Shimabukuro, Y.E. (2005). Spectral linear mixing model in low spatial resolution image data. IEEE transactions on geoscience and remote sensing, 43, 2555-2562

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., & Loveland, T.R. (2013). High-resolution global maps of 21st-century forest cover change. Science, 342, 850-853

Hilker, T., Wulder, M.A., Coops, N.C., Linke, J., McDermid, G., Masek, J.G., Gao, F., & White, J.C. (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, 1613-1627

Imhoff, M.L., Zhang, P., Wolfe, R.E., & Bounoua, L. (2010). Remote sensing of the urban heat island effect across biomes in the continental USA. Remote Sensing of Environment, 114, 504-513

Lunetta, R.S., Knight, J.F., Ediriwickrema, J., Lyon, J.G., & Worthy, L.D. (2006). Land-cover change detection using multi-temporal MODIS NDVI data. Remote Sensing of Environment, 105, 142-154

Luo, Y., Guan, K., & Peng, J. (2018). STAIR: A generic and fully-automated method to fuse multiple sources of optical satellite data to generate a high-resolution, daily and cloud-/gap-free surface reflectance product. Remote Sensing of Environment, 214, 87-99

Maselli, F., Gilabert, M.A., & Conese, C. (1998). Integration of high and low resolution NDVI data for monitoring vegetation in Mediterranean environments. Remote Sensing of Environment, 63, 208-218

Moon, M., Zhang, X., Henebry, G.M., Liu, L., Gray, J.M., Melaas, E.K., & Friedl, M.A. (2019). Long-term continuity in land surface phenology measurements: A comparative assessment of the MODIS land cover dynamics and VIIRS land surface phenology products. Remote Sensing of Environment, 226, 74-92

Oleson, K., Sarlin, S., Garrison, J., Smith, S., Privette, J., & Emery, W. (1995). Unmixing multiple land-cover type reflectances from coarse spatial resolution satellite data. Remote Sensing of Environment, 54, 98-112

Stone Jr, B., & Rodgers, M.O. (2001). Urban form and thermal efficiency: how the design of cities influences the urban heat island effect. American Planning Association. Journal of the American Planning Association, 67, 186

Yan, L., & Roy, D.P. (2020). Spatially and temporally complete Landsat reflectance time series modelling: The fill-and-fit approach. Remote Sensing of Environment, 241, 111718

Young, S., & Wang, C. (2001). Land-cover change analysis of China using global-scale Pathfinder AVHRR Landcover (PAL) data, 1982? 92. International Journal of Remote Sensing, 22, 1457-1477

Zhan, X., Sohlberg, R., Townshend, J., DiMiceli, C., Carroll, M., Eastman, J., Hansen, M., & DeFries, R. (2002). Detection of land cover changes using MODIS 250 m data. Remote Sensing of Environment, 83, 336-350

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

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