A spatial-temporal-spectral blending model using satellite images

L Zhang¹, D Fu¹, X Sun¹, H Chen¹ and X She²

¹ The State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, CAS, Beijing, 100101, China
² University of Chinese Academy of Sciences, Beijing, China

E-mail: zhanglf@radi.ac.cn

Abstract. Due to the budget and technical limitations, remote sensing sensor designs trade spatial resolution, swath width and spectral resolution. Consequently, no sensor can provide high spatial resolution, high temporal resolution and high spectral resolution simultaneously. However, the ability of Earth observation at fine resolution is urgently needed for global change science. One possible solution is to “blend” the reflectance from a variety of satellite data sources, including those providing high spatial resolution and less frequent coverage (e.g., Landsat Thematic Mapper, TM), daily global data (e.g., Moderate Resolution Imaging Spectroradiometer, MODIS), and high spectral resolution and infrequent revisit cycle (e.g., Hyperion). However, the previous algorithms for blending multi-source remotely sensed data have some shortcomings, especially with regard to hyperspectral information. This study has developed a SPAtial-Temporal-Spectral blending model (SPATS) that can simulate surface reflectance with high spatial-temporal-spectral resolution. SPATS is based on an existing spatial-temporal image blending model and a spatial-spectral image blending model. The performance of SPATS was tested with both simulated and observed satellite data, using Landsat TM, Hyperion and MODIS data, as well as heterogeneous landscapes as examples. The results show that the high spatial-temporal-spectral resolution reflectance data can be applied to investigations of global landscapes that are changing at different temporal scales.

1. Introduction
Capturing spatial and temporal dynamics of environmental change is important for many remote-sensing based applications. However, a trade-off must be considered among the spatial, temporal and spectral resolutions when designing satellite sensor systems. Being constrained by specific single-sensor goals and the sensor-specific data framework (Emelyanova et al., 2013), there is no single satellite sensor which can produce images with fine spatial, temporal and spectral resolution. There are some useful image fusion models, including spatial-temporal image fusion models (Gao et al., 2006; Zhu et al., 2010) and spatial-spectral image fusion models (Eismann and Hardie, 2004; Hardie et al., 2004; Winter et al., 2007; Liu et al., 2009; Zhang et al., 2009; Yokoya et al., 2012) that have been developed recently; the results can either be fused with high spatial-temporal resolution data or high spatial-spectral resolution data. So far, there have been no reports in the published literature that focus on how to produce high spatial-temporal-spectral image data. In order to meet this challenge, we have developed a SPAtial-Temporal-Spectral blending model (SPATS) that can blend (i) multispectral data with fine spatial resolution, low temporal frequency (e.g., Landsat Thematic Mapper), (ii) multispectral data with coarse spatial resolution, high temporal frequency (e.g., MODIS) and (iii)
hyperspectral data with high spatial resolution, low temporal frequency (e.g., Hyperion) together, to produce simulated image data with high spatial-temporal-spectral resolution.

2. Methodology

2.1 Datasets

The three satellite datasets used in this study are presented in table 1, as follows: (I) Landsat TM data, (II) MODIS data and (III) Hyperion data. Dataset I included a MODIS, a Landsat enhanced Thematic Mapper (ETM) and a Hyperion image. Dataset II and III included MODIS imagery of different dates. The false-colour infrared MODIS, ETM and Hyperion images of Dataset I are shown in figure 1.

| Dataset | Date (Year, Julian Day) | Sensor Type          |
|---------|-------------------------|----------------------|
| I       | 2004 089                | MODIS, ETM, Hyperion |
| II      | 2004 121                | MODIS                |
| III     | 2004 153                | MODIS                |

Table 1. Satellite datasets

Figure 1. False colour of MODIS, ETM+ and Hyperion data covering a 10 km by 10 km area in the southwestern United States. Data courtesy of NASA and USGS.

The Landsat TM/ETM+ surface reflectance product (path/row: 037/037) covering an area in the southwestern United States was acquired from the EarthExplorer web site of United States Geological Survey (USGS) at http://earthexplorer.usgs.gov/.

The MODIS MOD09A1 product data (8-day reflectance, 500 m, Collection 5, tile: h08v05) for the study area were obtained from the National Aeronautics and Space Administration (NASA) Earthdata portal at https://earthdata.nasa.gov/. The MODIS data were reprojected, resampled and clipped to the same spatial extent and resolution as the Landsat projection using the MODIS Reprojection Tool (MRT, https://lpdaac.usgs.gov/tools/modis_reprojection_tool). Any invalid MOD09A1 data were eliminated using the quality assurance (QA) layer included in the product (Fu et al., 2014).

The Hyperion data with a swath width of 7.5 km covers a wavelength range from the visible to shortwave infrared, providing a spatial resolution similar to that obtained by the Landsat TM sensor. The Hyperion sensor system has 242 bands, and its L1R product downloaded from the USGS EarthExplorer web site provides 196 effective 10-nm-wide bands from 400 to 2500 nm. Owing to the lack of calibration, low signal-to-noise-ratio (SNR), strong water vapor absorption, and vertical striping effects, a subset of 155 Hyperion bands were selected for our study.
2.1. Spatial-temporal-spectral image blending model

The spatial-temporal-spectral image blending model consists of three implementation steps and uses the C++ program language. The first step is a spatial-temporal reflectance simulation to produce simulated reflectance with high spatial-temporal resolutions; the second step is spatial-spectral reflectance simulation to produce simulated reflectance with high spatial-temporal-spectral resolutions. The third step involves the organization of a multidimensional data structure.

In order to complete the first step, the spatial-temporal image fusion model (STARFM) developed by Gao et al. (2006) was used for the study. Only one pair of Landsat and MODIS images acquired on the same date and one MODIS observation on the simulated date were taken into account. Landsat-like surface reflectance was obtained on the simulated date using the following procedure. First, the spectrally similar neighboring pixels were identified within a local moving window of Landsat data. Second, a weighting $W$ was calculated for each spectrally similar neighboring pixel based on: (i) the spectral difference between the Landsat and MODIS data for the actual acquisition date, (ii) the temporal difference of the MODIS data between the actual date and simulated date, (iii) the spatial Euclidean distance between the neighbor and the central pixel within the local moving window. Finally, the surface reflectance of the central pixel was calculated, as follows:

$$L(x_{w/2}, y_{w/2}, t_s) = \sum_{i=1}^{W} \sum_{j=1}^{W} \sum_{b=1}^{B} W_{jk} \times (M(x_i, y_j, t_b) + L(x_i, y_j, t_b) - M(x_i, y_j, t_b))$$

(1)

where $L$ and $M$ indicate the surface reflectance of Landsat and MODIS, respectively; where $L(x_{w/2}, y_{w/2}, t_s)$ is the surface reflectance of the central pixel $(x_{w/2}, y_{w/2})$ on the simulated date $t_s$ for Landsat; where $M(x_i, y_j, t_b)$ is the surface reflectance of pixel $(x_i, y_j)$ within local moving window on the base date $(t_b)$ for MODIS; and where $w$ is the size of the local moving window. For more detailed information on STARFM, we referred to Gao et al. (2006).

In order to complete the second step, the spatial-spectral image fusion model (SREM) developed by Sun et al. (2015) was used for this study. The basic principle of SREM entails the fusion of information from low-spatial-resolution hyperspectral imagery and from high spatial resolution multispectral imagery acquired over the same area. To start, spectra of different endmembers, which include all those in the region of overlap, were extracted from both hyperspectral and multispectral images. The number of the spectra of each endmember must be equal to or greater than the band number of the multispectral image. These spectra are then divided into $N$ groups representing $N$ different endmembers. The relationship between the hyperspectral set and the multispectral set of every material can be denoted as

$$G(q)I(q) = H(q) + I(q)$$

(2)

where $I(q) = [n(q)_{Z(1)}^T, n(q)_{Z(2)}^T, \ldots, n(q)_{Z(W)}^T]$ is an $Z \times W$ matrix representing the set of spectral column vectors $n(q)_{Z(j)}^T$; $H(q) = [h(q)_{K(1)}^T, h(q)_{K(2)}^T, \ldots, h(q)_{K(W)}^T]$ is a $K \times W$ matrix representing the set of spectral column vectors $h(q)_{K(j)}^T$; and $W$ is the number of pixel spectra collected from the material of class $q$ ($q$ is in the range from 1 to $N$). A specific transformation matrix $G(q)$ can then be calculated, as follows, provided $W$ is equal to or greater than $Z$:

$$G(q) = H(q)I(q)^T (I(q)I(q)^T)^{-1}$$

(3)
$N$ types of materials extracted from the images will generate $N$ transformation matrices. A spectrum $h(q)_{K(i)}^T$ with $K$ bands can then be obtained from the matrix $G(q)$ by multiplying $m_{Z(i)}^T$ with $Z$ bands from a multispectral image pixel. However, there are $N$ available transformation matrices, and the same number of $h(q)_{K(i)}^T$ will be produced. Therefore the correct transformation matrix must be selected. A spectral matching method is used to solve the problem by matching the generated spectrum $h(q)_{K(i)}^T$ with the average of original hyperspectral set $H(q)$. This process can be performed over the whole multispectral image pixel by pixel, even beyond the scope of the original hyperspectral image. The final spectra-enhanced product $H^*$, with the same spatial resolution and swath as the multispectral image and the same spectral resolution as the original hyperspectral image, will then be obtained. If one considers the case that consistent, or nearly consistent, ground-related features of the same area for different years, the derived $G(q)$ can then be used to enhance the spectral resolution of multispectral data of other time periods (e.g., $t_j$), without additional auxiliary hyperspectral data.

Finally, the multitemporal hyperspectral images were organized by a dimensional data structure. This structure consisted of data and header files, as illustrated in figure 2. The data file actually stores the image data in TIP, TIB, TSQ, TSP and TIS format. The header file contains image related information, such as spatial, temporal and spectral dimension, data storage format, data type and a description pertaining to projection and coordinate specific transform coefficients, the name of spectral and temporal dimension, and data migration.

Five data formats are proposed through the analysis of extracting data from temporal, spatial and spectral dimensions of SPATS. These formats are illustrated in figure 3 and explained in further detail, as follows.

In the Temporal Sequential (TSQ) format, all the spectral cubes are stored in chronological order. Then within each spectral cube, all the bands are arranged in spectral order. At last, the layout of each band is from row to column. This kind of organization puts together all bands at one time, ensuring the continuity of the band-related data within the storage space. The TSQ format is suited for spatial operations, because it efficiently handles the process of addressing and speedily extracts band-related data.
In the Temporal Sequential Pixel (TSP) format, the bands are also arranged in spectral order, similar to the TSQ format. However, in the TSP format, pixels are firstly arranged in spectral order, then from row to column order. It collates the spectral data of the pixels and ensures the continuity of spectral data within the storage space. Therefore, the TSP format is suit for spectral operations, because it efficiently handles the process of addressing and speedily extracts spectral information.

In the Temporal Interleaved by Band (TIB) format, the whole time series of each band are put arranged in a time-sequential way, and within the time series of each band the data are arranged in spectral order. Furthermore, all pixels in each band are laid out from row to column. The TIB format ensures the continuity of spatial data within storage space, and at the same time guarantees the continuity of time for each band. Therefore, the TIB format is suited for extracting the temporal cube of a band and at the same time performing the process of addressing in an efficient manner.

In the Temporal Interleaved by Pixel (TIP) format, the whole time series of each pixel for each band are arranged in a time-sequential way; for in each band, the time series of each pixel are arranged from row to column, and finally the data for all the bands are arrayed in spectral order. The TIP format collates the time spectrum of all the pixels and thus ensures the continuity of time spectrum data within storage space. The TIP format is suited for extracting the time spectrum of band-related pixels and at the same time performing the process of addressing in an efficient manner.

In the Temporal Interleaved by Spectrum (TIS) format, the whole time series of each pixel are arranged in a time-sequential way; then the time series are laid out from row to column in the spatial dimension. The TIS format ensures the continuity of spectral data, and at the same time it arranges the time series of that data. Therefore, the TIS format is suited for extracting spectral data and its time series, while performing the process of addressing in an efficient manner.

![Figure 3](image-url)  
**Figure 3.** Five data formats for structuring multidimensional SPATS data
3. Result and discussion

Figure 4 shows the false-color simulated hyperspectral images generated by the SPATS model time series for the 89th, 121th and 153th day. We evaluated the results by comparisons based on visual interpretation and statistical characteristics. To consider the general spatial appearance of the fused image and the real images by a visual interpretation, we selected the specific bands as red, green, and blue (RGB) values to show the false-colour composite image. Figure 4 also illustrates that the false colour composites of the fused images and real ETM+ images were very consistent overall.

![Figure 4](image)

**Figure 4.** False-colour simulated hyperspectral images by the SPATS model of the 89th day (a), on the 121th day (b) and the 153th day (c).

|                | PSNR | SAM | CC   | UIQI   |
|----------------|------|-----|------|--------|
| Dataset I      | 26.63| 2.687| 0.9785| 0.9784 |
| Dataset II     | 27.92| 4.085| 0.9877| 0.9815 |
| Dataset III    | 26.79| 5.916| 0.9805| 0.9699 |

In order to give a comprehensive evaluation by objective and quantitative analysis, the fused images were compared to the original images in the overlapping regions using four indices, as outlined by Wang and Bovik (2002): spectral angle mapper (SAM), peak signal-to-noise ratio (PSNR), correlation coefficient (CC), and universal image quality index (UIQI). The average cumulative values of the four indices are shown in table 2, with SPATS performing to satisfaction considering all four indices. The index values for Dataset II and Dataset III exceeded those obtained for Dataset I.

![Figure 5](image)

**Figure 5.** Correlation coefficients between fused and real hyperspectral data of 155 bands in the overlapping regions of Dataset I (a), Dataset II (b) and Dataset III(c).
The correlation coefficients between each band of the fused and corresponding real images were calculated and are illustrated in figure 5. The higher the correlation coefficients were, better the method performed. Nearly all the bands generated by SPATS had correlation coefficients exceeding 0.97, indicating that the value of each fused band was very similar to the real one. Table 2 and Figure 5 show that the results are very consistent.

4. Conclusion
A SPAtial-Temporal-Spectral blending model for remotely sensed images has been provided, with the aim of generating surface reflectance data with high spatial-temporal-spectral resolution. Taking Landsat, Hyperion and MODIS data as an example, the performance of SPATS was tested with both simulated and observed satellite data. The results show that images generated by the SPATS model are very similar to real data, indicating the excellent performance of our model.

References
Eismann M T and Hardie R C 2004 Application of the stochastic mixing model to hyperspectral resolution, enhancement IEEE Transactions on Geoscience and Remote Sensing 42 (9) pp 1924-1933
Emelyanova I V et al 2013 Assessing the accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection Remote Sensing of Environment 133 (0) pp 193-209
Fu D et al 2014 Estimating landscape net ecosystem exchange at high spatial–temporal resolution based on Landsat data, an improved upscaling model framework, and eddy covariance flux measurements. Remote Sensing of Environment 141 (0) pp 90-104
Gao F et al 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) pp 2207-2218.
Hardie R C et al 2004 MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor IEEE Transactions on Image Processing 13(9) pp 1174-1184
Liu, B et al 2009 Simulation of EO-1 Hyperion data from ALI multispectral data based on the spectral reconstruction approach Sensors 9 (4) pp 3090-3108
Sun X et al 2015 Enhancement of spectral resolution for remotely sensed multispectral image. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sensing 8 (5) pp 2198-2211
Wang Z and Bovik A C 2002 A universal image quality index IEEE Signal Processing Letters 9 (3) pp 81-84
Winter M E et al 2007 Hyperspectral image sharpening using multispectral data IEEE Aerospace Conference 2007 Vols 1-9 pp 2079-2087
Yokoya N et al 2012 Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion IEEE Transactions on Geoscience and Remote Sensing 50 (2) pp 528-537
Zhang Y F et al 2009 Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images IEEE Transactions on Geoscience and Remote Sensing 47 (11) pp 3834-3843
Zhu X L et al 2010 An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions Remote Sensing of Environment 114 (11) pp 2610-2623