Removal of clouds, dust and shadow pixels from hyperspectral imagery using a non-separable and stationary spatio-temporal covariance model

Yoseline Angel*, Rasmus Houborg and Matthew F. McCabe
Hydrology, Agriculture and Land Observation (HALO) Group, Water Desalination and Reuse Center, Division of Biological and Environmental Science & Engineering (BESE), King Abdullah University of Science and Technology (KAUST), Saudi Arabia

ABSTRACT

Hyperspectral remote sensing images are usually affected by atmospheric conditions such as clouds and their shadows, which represents a contamination of reflectance data and complicates the extraction of biophysical variables to monitor phenological cycles of crops. This paper explores a cloud removal approach based on reflectance prediction using multi-temporal data and spatio-temporal statistical models. In particular, a covariance model that captures the behavior of spatial and temporal components in data simultaneously (i.e. non-separable) is considered. Eight weekly images collected from the Hyperion hyper-spectrometer instrument over an agricultural region of Saudi Arabia were used to reconstruct a scene with the presence of cloudy affected pixels over a center-pivot crop. A subset of reflectance values of cloud-free pixels from 50 bands in the spectral range from 426.82 to 884.7 nm at each date, were used as input to fit a parametric family of non-separable and stationary spatio-temporal covariance functions. Applying simple kriging as an interpolator, cloud affected pixels were replaced by cloud-free predicted values per band, obtaining their respective predicted spectral profiles at the same time. An exercise of reconstructing simulated cloudy pixels in a different swath was conducted to assess the model accuracy, achieving root mean square error (RMSE) values per band less than or equal to 3%. The spatial coherence of the results was also checked through absolute error distribution maps demonstrating their consistency.

Keywords: Hyperspectral, multi-temporal, non-separable, covariance, spatio-temporal, kriging, R software, remote sensing.

1. INTRODUCTION

Dealing with the presence of dust, clouds and their associated shadows is a complex and relevant issue when working with multi-temporal sequences of satellite images retrieved by passive sensors. Clouds in an image affect the retrieval of any land cover elements, hiding relevant information and obscuring the ground, while introducing shadowing and radiative trapping effects. One sensor facing such challenges is the Hyperion instrument, a push-broom imaging spectrometer on-board the National Aeronautics and Space Administration (NASA) Earth Observing 1 (EO-1) satellite. Hyperion is the first hyperspectral instrument in space that covers the visible, near infra-red (VNIR) and short-wave infrared (SWIR) spectra between 400 and 2500 nm via 242 unique spectral channels. Like many other systems, weather conditions are determinants during scene acquisition, with cloud cover percentage depending on the geographic location of the target and season. Various cloud removal approaches have been proposed for restoring affected pixels coherently, some of which are based on the spatial and temporal structure of the data, while others are based on the spectral structure.

Multi-temporal approaches include methods such as predicting pixel values by regression trees from images acquired at different times, or reconstructing cloud-contaminated areas in multi-temporal multi-spectral images using a linear predictor algorithm that learns from data available in the cloud-free neighborhood and taking advantage of the inter-

* yoseline.angellopez@kaust.edu.sa; phone 966 0501154199; hydrology.kaust.edu.sa
channel spectral correlation. Other approaches have employed multiple point geostatistical techniques to fill gaps in satellite imagery created by clouds or orbital features\textsuperscript{7} and also to downscale simulations\textsuperscript{8}. Spectral approaches, specifically for hyperspectral images, exploit a spectral unmixing technique. For instance, Cerra et al\textsuperscript{9} proposed a method for reconstructing a cloudy pixel as a weighted sum of endmembers related to the materials present in the same image. And, Feng et al\textsuperscript{10} developed an alternative solution under the same approach, defogging a scene by obtaining the fog endmember and proportionally removing this element and adjusting the remaining endmembers per pixel. However, in all of these cases, spatial, temporal and spectral components of a multi-temporal hyperspectral dataset are considered independently, which is not a realistic assumption.

Spatio-temporal statistics for predicting real and measurable random variables such as radiance, offers an unexplored variety of approaches that allows modeling data covariance in space, time and spectral dimensions simultaneously. These kind of functions are classified as non-separable, because the total covariance of the dataset is considered as a linear combination of pure spatial and pure temporal covariance functions\textsuperscript{11}. Some applications in prediction of environmental phenomena such as the analysis of wind speed in Ireland\textsuperscript{12}, tropospheric ozone (O\textsubscript{3}) concentration in Tehran\textsuperscript{13}, nitrogen dioxide (NO\textsubscript{2}) concentration in Milan\textsuperscript{14}, have been successfully performed and have achieved more reliable and accurate results compared to using traditional (and separable) spatio-temporal models. In this paper we explored the use of non-separable models to reconstruct areas affected by clouds, shadows and/or dust in hyperspectral images, predicting reflectance values from a dataset of Hyperion swaths that were collected over an agricultural region in Saudi Arabia\textsuperscript{15} across a variety of different dates. A geostatistical model that is continuous in space and time was established by defining a non-separable and stationary covariance function proposed by Gneiting\textsuperscript{16} and modeled using the package ‘CompRandFld’, which was developed by Padoan and Bevilacqua\textsuperscript{17} in the R software\textsuperscript{18}.

2. METHODOLOGY

2.1. Data description

Eight weekly Hyperion swaths were captured from September 28 to November 11, 2015 over an irrigated agricultural region in the desert of Saudi Arabia\textsuperscript{15} (Figure 1). The original images have 196 calibrated bands with a spatial resolution of 30 meters and a percentage of cloud cover less than 10\%. Each scene was radiometrically corrected to get absolute radiance (W.cm\textsuperscript{-2}.sr\textsuperscript{-1}.nm\textsuperscript{-1}) and atmospheric correction was performed to achieve reflectance retrievals using the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) tool in ENVI 5.1. FLAASH incorporates the MODTRAN model atmospheres, of which the U.S standard model was selected. Water vapor retrieval per pixel was

![Figure 1. Tawdeehiya farm location in Saudi Arabia, captured by Hyperion at 8 different weeks from September 18 to November 11, 2015. The images are false infrared color compositions (R: Band45 G: Band23 B: Band13). Last scene (Nov11) has some clouds that were considered as prediction locations.](image-url)
performed using bands centered around 1135 nm. For aerosol retrieval, a rural model was selected and the visibility per scene was determined by applying the Koshmieder formula\textsuperscript{19}, using aerosol optical depth (AOD) at 550 nm from the MOD04 product and an effective height of the aerosol vertical profile of 2 km (which is also used to define the adjacency effect). To reduce the striping effect and to smooth reflectance profiles, a minimum noise fraction\textsuperscript{20} (MNF) transformation was applied.

Each reflectance image was reduced to a spectral subset of 50 bands, comprising the range between 426.82 nm and 884.7 nm, and to a spatial subset of 1584 pixels, covering a single pivot crop area. The scene captured on November 11 was used to classify and to mask by photo-interpretation a total of 223 cloud pixels, which were used as prediction locations.

2.2. Exploratory data analysis

Reflectance values were transformed by removing pixel to pixel the average per band to stabilize the variance over all locations and time periods and to make the marginal distributions approximately normal. Then, the resulting residual reflectance images were used as input for modeling, accomplishing the Gaussian assumption required for simple kriging.

Additionally, stationarity of residuals was verified with directional variograms for each band. Figure 2 shows an example of how all semi-variograms follow a close trend, reaching the sill around 0.0020 squared reflectance, with a range of approximately 700 meters.

![Figure 2 - Directional semi-variograms for band 13 (548.92 nm) on September 18.](image)

Additionally, temporal correlation was evaluated by examining scatterplots that compare residual reflectance at each single pixel, temporally spaced by steps of one week up to eight weeks. Figure 3 shows how for short lags, reflectance data are strongly temporally correlated, and how longer time lags weaken the correlation (although this remains positive). This behavior indicates that kriging estimates will include data from just the nearest or subsequent time interval.
2.3. Non-separable, stationary spatio-temporal covariance model and kriging

Reflectance is considered as a real-valued variable $Z(s; \lambda; t)$ measured at the center point of a pixel $(s_k; \lambda_k; t_k)$, where $s, \lambda$ and $t$ denote location coordinates, band number and date respectively. Assuming $\mu(s; t) = E(Z(s, t))$ as mean function, covariance is defined as Cov $(Z(s; t), Z(s + h; t + u)) = C(h; u)$, where $h$ and $u$ correspond to the space and time lag. When $u=0$, the covariance function $C(h; 0)$ is the purely spatial case. If $h=0$, the covariance function $C(0; u)$ is a purely temporal case. In this light, a valid general class of non-separable stationary covariance functions that take into account space-time interactions when $h>0$ and $u>0$, is proposed by Gneiting:

$$C(h; u) = \frac{\sigma^2}{\psi(|u|^2)} \varphi \left( \frac{|h|}{\psi(|u|^2)} \right)$$

(1)

where the spatial structure of the covariance is represented by any completely monotonic function $\varphi(t)$, $t \geq 0$ and the temporal structure is represented by any positive function with a completely monotone derivative $\psi(t)$, $t \geq 0$.

From the reflectance residuals calculated in Section 2.2, an empirical spatio-temporal variogram per band was estimated by using the Evariogram method included in the CompRandFld2 R package. Figures 4c and 4d show an example for band 50 (884.7 nm), where the sill is reached with 1 week of temporal lag and 700 meters of spatial lag respectively. Then, a family of non-separable covariance functions was fitted following the general case proposed by Gneiting and included in the same package:

$$\rho_{st}(h, u) = g_t(u)^{-1} \exp \left\{ - \frac{d_s(h)}{\beta_t(u)} \right\}, \quad 0 \leq \eta \leq 1$$

(2)

$$d_s(h) = \left( \frac{|h|}{\psi_s} \right)^{k_s}, \quad g_s(h) = 1 + d_s(h), \quad \psi_s, \psi_t > 0$$

(3)

$$d_t(u) = \left( \frac{|u|}{\psi_t} \right)^{k_t}, \quad g_t(u) = 1 + d_t(u), \quad 0 < k_s, k_t \leq 2$$

(4)

where $\eta$ represents the separability parameter, $k_s, k_t$ correspond to smoothing parameters and $\psi_s, \psi_t$ for scale parameters. Space and time structures are labelled with the subscripts $s$ and $t$ respectively. Some of the parameters were fixed, including the mean (zero-mean), spatial smoothness (power_s = 2), temporal smoothness (power_t = 0.5) and separability (sep = 1), which indicates that space and time structures are completely non-separable. The remaining...
parameters were estimated by a 3D space-time variogram by using the weighted least squares (WLS) method\textsuperscript{23}, which could be considered as the best likelihood-based approach with a good computational efficiency\textsuperscript{24}.

Using the simple kriging predictor:

\[ Z^*(s_0; t_0) = \mu(s_0; t_0) + \sum_{i=1}^{K} a_i (Z(s_i; t_i) - \mu(s_i; t_i)) \]  

(5)

as an optimal linear combination that minimizes the mean squared prediction error\textsuperscript{25}, each fitted spatio-temporal covariance per band was used as input to predict each location defined as a cloud pixel. Reflectance means that were previously removed were added back to the output residuals and replaced pixel by pixel in the cloud affected image.

![Empirical Space-time variogram](image1.png)
![Space-time variogram](image2.png)
![Variogram: temporal profile](image3.png)
![Variogram: spatial profile](image4.png)
![Figure 4](image5.png)

Figure 4. Non-separable stationary model for band 50 (884.7 nm). a) Empirical space-time variogram. b) Fitted 3D space-time variogram. c) Pure temporal case (h=0), fitted vs temporal empirical variogram (dots). d) Pure spatial case (u=0), fitted vs spatial empirical variogram (dots). e) WLS estimated parameters for band 50 (884.7 nm).

3. RESULTS

Using the methodology outlined above in Section 2, a new image without cloudy pixels was obtained after the kriging process. Figure 5 shows how affected pixels were restored over the crop pivot and surrounding desert sand areas. A blurred effect is noticeable at the border between both classes. However, interpolation results are good, spatial patterns are visible and generally match the original image.
Additionally, to quantify the interpolation accuracy, an assessment of the results was conducted predicting reflectance for a sample of cloud free pixels, and then estimating the root mean square error (RMSE) per band (Table 1), which was variable along the spectral range comprised between band 1(426.82 nm) and band 50 (884.7 nm).

<table>
<thead>
<tr>
<th>Spectrum region</th>
<th>Bands</th>
<th>Spectral range (nm)</th>
<th>RMSE (min - max %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>b1 - b7</td>
<td>426.82 - 487.87</td>
<td>0.9 - 1.4</td>
</tr>
<tr>
<td>Green</td>
<td>b8 - b19</td>
<td>498.04 - 609.97</td>
<td>1.0 - 2.3</td>
</tr>
<tr>
<td>Red</td>
<td>b16 - b33</td>
<td>579.45 - 752.43</td>
<td>2.4 - 2.7</td>
</tr>
<tr>
<td>Near Infrared (NIR)</td>
<td>b34 - b50</td>
<td>762.60 - 884.7</td>
<td>1.2 - 2.6</td>
</tr>
</tbody>
</table>

In addition, absolute error distribution maps per band were used to evaluate the spatial coherence of the predictions. Figure 6 shows an example for band 50 (884.7 nm), where it is observed how the absolute error is randomly distributed, with the exception of highest errors that are distributed along the edge of the crop, which was a common pattern for the rest of the bands. This performance explains the fuzzy effect observed along the borders between crop and sand covers in figure 5.

Figure 6. Absolute error distribution map for band 50 (884.7 nm), with absolute errors from 0.8% to 5% and an RMSE of 1.2%.
Figure 7 shows a comparison between the original and predicted spectral profiles for a pixel corresponding to early stage vegetation into the crop area. It is evident how absorptions in predicted curves follow the same original shapes. Low differences between original and predicted profiles demonstrate the spectral coherence of the results. According with table 1 and figure 7, lowest differences between the original and estimated reflectance values were achieved along the region covered by the blue, green and last near infrared (NIR) bands, and highest errors were reached along ranges covered by the red and first NIR channels.

Figure 7. Comparison between original and predicted spectral profiles for a pixel of early stage vegetation.

4. SUMMARY AND CONCLUSIONS

A variety of methods to account for cloud, shadow or dust effects in remote sensing images have been developed over the last few decades. The present work applies an alternative spatio-temporal statistical approach to remove these effects in hyperspectral images by replacing the cloud-affected pixels with predicted values. Considering reflectance as a Gaussian variable observed at each pixel location per band at different dates, its variance was modeled and fitted to a non-separable, stationary covariance function in a 3-dimensional (spatial, temporal and spectral) domain, using the Gneiting’s function implemented in the CompRandFld package. This covariance function was used by an interpolator algorithm (in this case simple kriging), to estimate reflectance values at the cloud pixel locations. Then, affected pixels were removed and replaced by the predicted results.

This approach can be extended to remove any type of obscuring effect in hyperspectral or multispectral images, including clouds, shadows, dust, fog, gaps, etc., by defining each affected pixel per band as a location to be predicted. It is important to highlight that in order to model the temporal variance, acquisition dates of the different swaths should be close to each other e.g. at weekly collections at most. Also, one advantage of this approach is that once a non-separable spatio-temporal covariance family is defined for a set of multi-temporal images, different kriging methods or any other local interpolator could be used to predict missed data into the spatial, spectral and time ranges of the full dataset. Future research is required to further examine the implementation of this kind of spatio-temporal covariance functions in different software environments, with an aim to reduce the computational resources demanded to model large datasets such as multi-temporal hyperspectral imagery.
REFERENCES