6-2019

Evaluating visible derivative spectroscopy by varimax-rotated, principal component analysis of aerial hyperspectral images from the western basin of Lake Erie

Joseph D. Ortiz
Kent State University

Dulci M. Avouris
Kent State University

Stephan J. Schiller
South Dakota State University

Jeffery C. Luvall
NASA Marshall Space Flight Center

John D. Lekki
NASA Glenn Research Center

See next page for additional authors

Recommended Citation

Ortiz, J. D., Avouris, D. M., Schiller, S. J., Luvall, J. C., Lekki, J. D., Tokars, R. P., Anderson, R. C., Shuchman, R., Sayers, M., & Becker, R. (2019). Evaluating visible derivative spectroscopy by varimax-rotated, principal component analysis of aerial hyperspectral images from the western basin of Lake Erie. *Journal of Great Lakes Research, 45*(3), 522-535. http://dx.doi.org/10.1016/j.jglr.2019.03.005
Retrieved from: https://digitalcommons.mtu.edu/michigantech-p/640

Journal of Great Lakes Research 45 (2019) 522–535

Contents lists available at ScienceDirect

Journal of Great Lakes Research

journal homepage: www.elsevier.com/locate/jglr

Evaluating visible derivative spectroscopy by varimax-rotated, principal component analysis of aerial hyperspectral images from the western basin of Lake Erie

Joseph D. Ortiz [a,*], Dulci M. Avouris [a], Stephen J. Schiller [b,1], Jeffrey C. Luvall [c], John D. Lekki [d], Roger P. Tokars [d], Robert C. Anderson [d], Robert Shuchman [e], Michael Sayers [e], Richard Becker [f]

[a] Department of Geology, 221 McGilvrey Hall, Kent State University, Kent, OH 44242, USA
[b] South Dakota State University, 1175 Medary Ave., Brookings, SD 57006, USA
[c] NASA Marshall Space Flight Center, ZP11, NSSTC, 320 Sparkman Drive, Huntsville, AL 35805, USA
[d] NASA Glenn Research Center, 21000 Brookpark Road, Cleveland, OH 44135, USA
[e] Michigan Technological Research Institute, 3600 Green Ct. Suite 100, Ann Arbor, MI 48105, USA
[f] Department of Environmental Sciences, University of Toledo, 2801 W Bancroft Ave, MS #604, Toledo, OH 43606, USA

ARTICLE INFO

Article history:
Received 19 April 2018
Revised 1 February 2019
Accepted 14 March 2019
Available online 21 March 2019

Communicated by Caren Binding

Keywords:
Hyperspectral remote sensing
NASA Glenn HSI2
Spectral decomposition
Unsupervised classification
Cyanobacterial harmful algal blooms
PACE

ABSTRACT

The Kent State University (KSU) spectral decomposition method provides information about the spectral signals present in multispectral and hyperspectral images. Pre-processing steps that enhance signal to noise ratio (SNR) by 7.37–19.04 times, enables extraction of the environmental signals captured by the National Aeronautics and Space Administration (NASA) Glenn Research Center's, second generation, Hyperspectral imager (HSI2) into multiple, independent components. We have accomplished this by pre-processing of Level 1 HSI2 data to remove stripes from the scene, followed by a combination of spectral and spatial smoothing to further increase the SNR and remove non-Lambertian features, such as waves. On average, the residual stochastic noise removed from the HSI2 images by this method is 5.43 ± 1.42%. The method also enables removal of a spectrally coherent residual atmospheric bias of 4.28 ± 0.48%, ascribed to incomplete atmospheric correction. The total noise isolated from signal by the method is thus <±7% based on analysis of multiple swaths. The method is essentially independent of the order of operations, extracting the same spectral components within error in all cases, spatial patterns that are very similar and explaining nearly identical amounts of information from each image. Based on comparison between multiple swaths the uncertainty in the variance associated with each component averages ±1.69% and is as low as ±0.08% and in all cases <±3.15%.

Introduction

The Lake Erie cyanobacterial and harmful algal bloom (CyanoHAB) monitoring program lead by the NASA Glenn Research Center (Lekki et al., 2017) has been developing hyperspectral and multispectral methods to assess the extent and severity of the perennial toxic cyanobacterial bloom that develops in the Western Basin of Lake Erie. The backbone of the program is the airborne NASA Glenn Research Center's second generation Hyperspectral imager (HSI2) sensor. This paper discusses an analytical approach employed to maximize the information that can be extracted from scenes of these optically complex waters. We present an analysis of several effective pre-processing steps used with the Kent State University (KSU) spectral decomposition method, which enhance the signal to noise ratio (SNR) of the hyperspectral data to produce a Level-2 product that differentiates various classes of color producing agents.

Over the years, optically complex waters have proven to be challenging environments for remote sensing of algal and cyanobacterial pigments (Bullerjahn et al., 2016). Extraction of high quality signal from these waters is important due to the growing concerns from CyanoHABs; (Bullerjahn et al., 2016). The challenges to visible remote sensing in coastal and inland waters arise from clouds, which obscure signal (Martin, 2014), mixed pixels (Raychaudhuri, 2012, Khodadadzadeh et al., 2014), difficulties with atmospheric correction (Gao et al., 2006), adjacency effects (Houborg and McCabe, 2017) and interference from bottom reflection (Brando et al., 2009; Eugenio et al., 2015). While bottom reflectance can be integrated into water leaving reflectance for some shallow areas, the high turbidity in Lake Erie results in Secchi depths on the order of 0.1 to 2 m, which are

* Corresponding author.
E-mail addresses: jortiz@kent.edu (J.D. Ortiz), Richard.Becker@utoledo.edu (R. Becker).
[1] Present Address: Raytheon Space and Airborne Systems, 2000 East El Segundo Blvd, El Segundo, CA 90245, USA.

https://doi.org/10.1016/j.jglr.2019.03.005

significantly less than the water depth in most cases, minimizing this potential source of error in this environment.

Another general source of error which has received less attention, is the strong intercorrelation of reflectance bands in the visible wavelength range. This behavior has been observed in large lakes using both multispectral (Ali et al., 2014; Ali et al., 2016) and hyperspectral instruments (Ortiz et al., 2013). This problem, known as multicollinearity, often leads to violation of a fundamental statistical rule by treating correlated, multivariate datasets with approaches that assume independent input variables (Kachigan, 1991; Smith, 2002; Shlens, 2014). Algorithms that unmix signals based on spectral shape (e.g., Palacios et al., 2015; Kudela et al., 2015; Soja-Woźniak et al., 2017), provide one way to deal with the issue of multicollinearity in multispectral and hyperspectral datasets.

The approach employed here, the KSU spectral decomposition method uses varimax-rotated, principal component analysis (VPCA), an eigenvalue-eigenvector decomposition to partition an image hypercube into independent signals generated by the underlying color producing agents that influence the scene (Ortiz et al., 2013; Lekki et al., 2017; Ortiz et al., 2017). The method is an unsupervised soft classification technique that makes use of all the coherent, linearly-correlated, spectral information present (Ortiz et al., 2017). This method is a variant of the principal component analysis (PCA) approach formalized by Preisendorfer and Mobley (1988). The varimax rotation simplifies the eigenvectors (referred to as component loadings) by finding the independent or orthogonal solution with the greatest separation between the large and small loadings (Kaiser, 1958). That step makes the component loadings easier to interpret (Kaiser, 1958). The communality provides a measure of the variance retained in the model on a band by band basis. The component loadings, which account for known fractions of variance in the image are then used to generate maps (component scores) that document the spatial distribution of each component. A full description of the theory and computation behind the KSU spectral decomposition method is discussed in detail in the electronic supplementary material (ESM) Appendix S1.

The objectives of this manuscript are to explore the sensitivity of the KSU spectral decomposition method to the presence of stripes and bands, to the order of operations during processing and to its reproducibility. The pre-processing steps are evaluated by applying the KSU spectral decomposition method to several HSI2 image swaths. The KSU spectral decomposition method has been applied to both multispectral (Lekki et al., 2017) and hyperspectral images (Ortiz et al., 2017). Another paper in this issue demonstrates that we can extract multiple components from multispectral Moderate Resolution Imaging Spectroradiometer (MODIS) images (Avouris and Ortiz, this issue). Documenting the reproducibility of the method demonstrates its applicability for use with data from coastal and inland waters that could be imaged by proposed orbital hyperspectral missions such as the Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission, or a Surface Biology and Geology mission sensor as called for in the NASA decadal survey (National Academies of Sciences, Engineering, and Medicine, 2018). When these orbital tools are deployed, they will create the opportunity for enhanced estimation of pigment-related biomass and new capabilities to identify algal and cyanobacterial composition based on extraction of pigment-related components by visible derivative spectroscopy as described here.

Striping and banding in the raw HSI2 swaths

Linear image artifacts, such as stripes or streaks - single pixel-wide features - and bands - multiple pixels-wide features - degrade remote sensing image quality (USGS Landsat 8 Users Handbook, 2016). Linear stripes or bands can occur across all wavelengths or be isolated to a single spectral band (Bouali, 2010). The orientation of the stripes or bands in the raw image varies depending on their origin and is determined by the sensor characteristics (USGS Landsat 8 Users Handbook, 2016). Whisk-broom sensors (e.g., MODIS, Landsat 7 Enhanced Thematic mapper Plus (ETM+), Landsat 4–5 Thematic Mapper (TM), or Landsat 1–5 Multispectral scanner (MSS)) sensors produce horizontal, cross-track stripes and bands (Antonelli et al., 2004; Bouali, 2010), while push-broom sensors (Landsat 8 and the NASA Glenn HSI2) produce vertical, along-track stripes and bands (Barsi et al., 2014).

These artifacts can hamper the extraction of environmental components by VPCA because they introduce systematic bias that can swamp much or all the true environmental variance signals. This bias manifests as principal components that exhibit strong regular along-track or cross-track patterns related to the measurement geometry of the image. These geometric bias patterns, if present, may be expressed as a variety of features that can either dominate or become convolved with the environmental signal. Thus, unless stripes and bands are removed (as is standard procedure) during image pre-processing prior to spectral decomposition, it may only be possible to extract the dominant component from the raw, striped image, or the environmental components may only explain a fraction of the total variance in the image. Exploring these issues here is useful, because VPCA is a relatively novel procedure in water quality applications (Ortiz et al., 2013; Lekki et al., 2017; Ortiz et al., 2017).

Single pixel wide stripes arise from relative gain variations associated with individual detectors (USGS Landsat 8 Users Handbook, 2016). Sensor Chip Assemblies (SCA) are focal plane arrays (Kean et al., 2012; Schultz et al., 2014) made of multiple detector chips segmented in a staggered array pattern (to allow cross-track overlap) with as many as needed to cover the sensor cross-track field-of-view (FOV). A monolithic focal plane (such as with the NASA Glenn HSI2) is one in which a single SCA covers the entire FOV (Lekki et al., 2017). The presence of banding and striping is due to a number of contributors, including inadequately characterized differential non-linear responses among detectors (often referred to as “residual photo-response non-uniformity”); poorly characterized parameters for relative gain, and bias (often referred to as “dark-current non-uniformity”); as well as non-linearity, drift, or instability in gain and bias, or spectral differences across or between SCAs. These potential contributors to striping and banding are described in the USGS Landsat 8 (L8) Users Handbook (2016) and are applicable to a variety of sensors.

In multispectral sensors, banding generally results from variations in output between detector modules (SCAs) on the focal plane (Markham et al., 2008). Landsat 8 has 14 modules and relative gain variations between the modules create banding (USGS Landsat 8 Users Handbook, 2016). The NASA Glenn HSI2 sensor has a monolithic focal plane detector over the entire FOV so this source of banding is not present in HSI2 data (Lekki et al., 2017). In the HSI2, spectral calibration changes (shifts in detector wavelength) between laboratory calibration and flight operations cause striping and banding (Lekki et al., 2017). In imaging spectrometers, the diffraction grating efficiency is wavelength dependent and can vary significantly over small wavelength intervals especially for blazed gratings (Palmer and Loewen, 2005). With a spectral shift after radiometric calibration, the transmission throughput of the spectrometer shifts, which can effectively change the relative gain between adjacent detectors or groups of detectors (Eismann, 2012).

Stripes and bands that are present only in a single wavelength band are referred to as spatial striping in images from linear detector arrays (whisk-broom or push-broom). Spectral striping can arise from spectrally dependent relative bias or gain drift by individual detectors at the same spatial location, but in different bands. This results in a calibration drift of band ratios, which is common in multispectral systems with multiple focal plane modules (USGS Landsat 8 Users Handbook, 2016). Although this mechanism is not applicable to the NASA Glenn HSI2 sensor due to the system design, biases can still occur. Besides detector response drifts, an additional source can be the application of processing to correct for atmospheric effects. Because of wavelength calibration drift and errors in knowledge of the sensor's spectral response and atmospheric spectral structure, residual atmospheric noise between

Fig. 1. The geographic location of the HSI2 swaths collected on June 21, 2016 over Maumee Bay State Park.

spectral bands can be introduced. Spectral striping is more difficult to remove than spatial striping, but can be achieved through spectral polishing methods (Schott, 2007). However, the process presented in this paper addresses only the issue of spatial striping.

Our goal is to perform a relative gain correction, which will correct for detector-level non-uniformity response. This relative gain correction can be referred to as detector equalization or flat-field correction. The HSI2 processing by NASA Glenn on Level 1 data prior to the work presented here performs a flat-field correction by taking the ratio of a raw image collected inflight with a pre-measured image of the calibration sphere that converts the raw digital numbers (DN) to radiance (Lekki et al., 2017). This calibration image is collected at the beginning of the campaign season. The striping that remains and that we seek to correct is the result of drift or non-linearity in individual detector responses that arise after the integrating sphere image was collected.

Because striping originates at the detector level, all destriping processing must be applied prior to any pixel resampling, which includes band to band registration on multispectral and hyperspectral systems during geo-rectification. Resampling thus makes a pixel value dependent on the response of multiple adjacent detectors, convolving any striping that is present into the georectified image. Such an image may not display visible stripes when viewed as reflectance due to the smoothing inherent in the georectification, but stripes may manifest when calculating a derivative or when applying the KSU spectral decomposition method.

Methods

We analyze four Level 1 hyperspectral HSI2 swaths, which were flown over water in the western basin of Lake Erie (Fig. 1). The naming convention for swaths, which consists of the instrument name, date, and location tag is as defined in Ortiz et al. (2017). The swaths were collected on 21 June 2016 within a span of 26 min over Maumee Bay State Park (MBSP), Toledo, OH (Fig. 1). The proximity of the swaths in time and space provides useful information about the reproducibility of the HSI2 images and the VPCA extraction procedure. To optimize the order of operations during processing and document the correction process, we focused on NASA HSI2 062116 Swath 15_MBSP, which has the lowest noise fraction of the four swaths.

All image processing is conducted in the Harris Geospatial ENVI/IDL (version 5.3) environment starting with NASA Glenn Level 1 processed HSI2 data, which has been converted to radiance, but which has not yet been georectified. The details of the NASA Level 1 image processing are found in the NASA Technical Report 2017-219071 (Lekki et al., 2017). NASA HSI2 Level 1 imagery is acquired in 496 pixel-wide swaths, of varying lengths, depending on the flight path of the aircraft. The

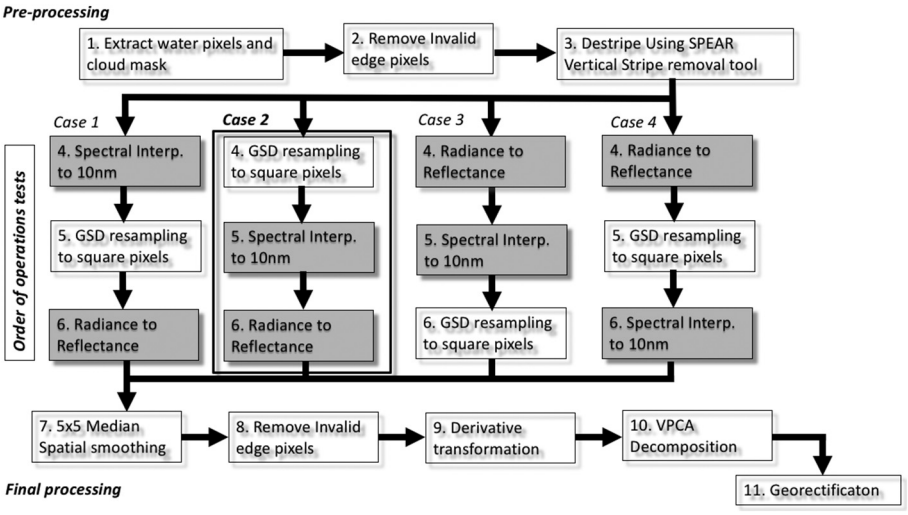

Fig. 2. Flow chart describing the processing steps employed in this study.

imagery has a spectral resolution of ~3 nm from 395 nm to 900 nm (170 bands). Ground resolution is dependent on flight speed and elevation and varies from ~3 m^2–6 m^2.

The HSI2 sensor has a nominal spectral sampling resolution of ~3 nm, and a laboratory SNR at 532 nm of 1000:1 (Lekki et al., 2017). We have further enhanced the field SNR by pre-processing of Level 1 HSI2 data to remove stripes from the scene, followed by a combination of spectral and spatial smoothing to increase the SNR and remove non-Lambertian features, such as waves from the images. Ortiz et al. (2017) addressed the impact of various atmospheric correction approaches on spectral signals and spatial patterns that can be extracted from hyperspectral images using the KSU spectral decomposition method. This work provides a useful assessment of how these standard pre-processing steps influence the quality of the results that can be extracted by VPCA.

Order of operations tests

In the following sections, the impact of the key processing steps is discussed. Following strip removal, the results of the VPCA could depend to some extent on the order of operations of the other pre-processing steps applied to increase the SNR ratio. These consist of the radiance to reflectance transformation, which removes atmospheric contamination, and spectral and spatial smoothing operations. To determine how sensitive the results were to the order of operations, we analyzed swath 062116 Swath 15_MBSP multiple times, altering the order of operations of the various processing steps for four different cases (Fig. 2). In each of the four cases, the initial and final processing steps were identical. During analysis, the image was processed using the steps described in the analysis flowchart (Fig. 2). The following subsections discuss the various processing steps.

Destriping procedure for HSI2 images

One of two standard, ENVI stripe removal tools was used to destripe the images. The "Spectral Processing Exploitation and Analysis Resource (SPEAR) Vertical Stripe Removal tool" is designed to remove along-track features generated by push-broom sensors, while the horizontal DESTRIPE algorithm in the Raster Data Management package is designed to remove cross-track features in whisk-broom sensors. Both tools require Level 1 data, which have not been georectified for processing. The destriping tool in the Raster Data Management package requests the number of detectors used to collect the scene as input and would for our application require that the HSI2 scene be transposed prior to analysis given the orientation of the level 1 image.

The SPEAR Vertical Stripe Removal tool asks for the percentage of the image radiance histogram outliers to be masked during the correction process. This information is used as the threshold outlier mask. Unfortunately, the details of the algorithm function are proprietary and not reported by Harris Geospatial as noted in the help file for the tool. However, we make some inferences on the likely nature of the tool below in the discussion section based on its required inputs and the resulting output. Per the published Harris Geospatial description of this tool in the ENVI documentation, it is effective on relatively homogeneous scenes, and is not intended to be applied to imagery with both bright land and dark water pixels. This does not pose a serious limitation for our work, because in this application we are exclusively interested in the water portion of the image. Although beyond the scope of this study, enhancements to the tool that would allow it to address heterogeneous brightness would generalize its use and likely improve its handling of adjacency issues.

To evaluate the effectiveness of strip removal, the level of striping or banding in the swaths can be quantified using a streaking metric such as defined in the Landsat 8 Operational Land Imager (OLI) requirements document (Markham et al., 2008). Accordingly, the swaths were cropped to remove land from the scene (Fig. 3b) to provide the SPEAR Vertical Stripe Removal tool with a relatively uniform radiance field. The magnitude of striping was then evaluated by quantizing the relative detector-to-detector uniformity for HSI2 band 46 as described in the OLI requirements document (NASA, 2009; USGS Landsat 8 Users Handbook, 2016). The mean radiance in each band, i, and adjacent bands (columns $i-1$ and $i+1$) were compared by calculating the streaking metric, S_i, using:

$$S_i = \frac{\left|\overline{L_i} - {}^1/_2\left(\overline{L_{i-1}} + \overline{L_{i+1}}\right)\right|}{\overline{L_i}}, \quad (1)$$

where, $\overline{L_i}$, $\overline{L_{i-1}}$ and $\overline{L_{i+1}}$ are the average radiances of the band pixel columns i, $i-1$, and $i+1$ respectively. The Landsat 8 maximum streaking

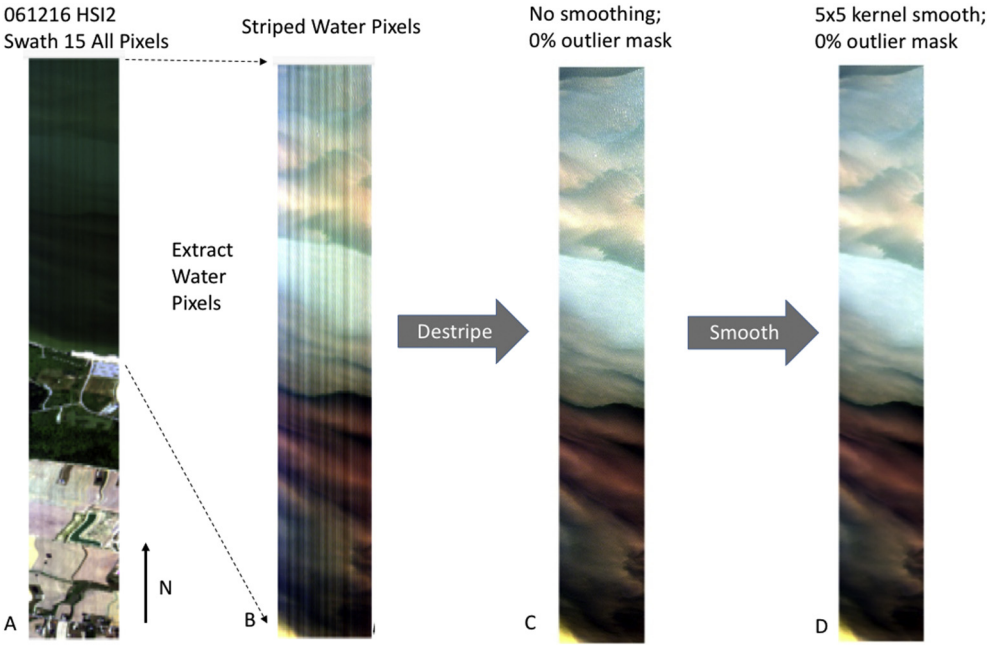

Fig. 3. Results of initial processing for (A) 061216 HSI2 Swath 15, (B) striped water pixels, (C) destriped image, (D) smoothed image. Solar illumination is from the east.

requirement of $S_i = 0.005$ was applied as a criterion to determine whether the swaths had been adequately destriped.

Wave variance minimization

Due to their non-Lambertian scattering nature, waves are not desirable for our purposes, which is to understand the composition of the material present in the lake. We thus remove as much of the wave-related variance from the image as possible prior to spectral decomposition. Accordingly, following stripe removal, the image was smoothed with a single pass of a simple, median filter using a 5 × 5 kernel. This kernel, maintains the pixel resolution of ~3.3 m^2, depending on aircraft elevation, but iteratively replaces the value in each pixel by the median value of the 5 × 5 pixel neighborhood surrounding the pixel of interest. The resulting image has a pixel density at ~3.3 m^2 resolution, but features smaller than 5 × 5 have been reduced in variance. This was deemed to be a reasonable compromise between reducing the wave variance and over smoothing the image. The smoothing function removes much of the wave crest-related variance from the image, without obscuring the major features that are present (Fig. 3d). We employed less smoothing here than in our prior paper, because this analysis was conducted on ~3.3 m^2 pixels as opposed to ~1 m^2 pixels as in the prior paper (Ortiz et al., 2017). Future work could explore the use of more sophisticated semi-variogram based smoothing methods or ones which take into account windspeed and direction (e.g., Chen and Gong, 2004), but the approach employed here is adequate based on the results described below.

Edge pixel removal

Given the simplicity of the smoothing method employed, on at least one of the HSI2 images collected on June 21, 2016, a narrow strip of edge pixels did not contain valid data. The signal on the eastern and western edges of the image swath was not coherent with the horizontal fronts and filaments that are apparent in the rest of the image. These pixels were removed from the image prior to further analysis. It was also necessary to remove a two-pixel border from the edges of the image following the kernel smoothing process. The two border pixels were too close to the image edge to be incorporated into the smoothing process. More sophisticated filters could address the loss of edge pixels, but this was deemed acceptable given the very small fraction of the image area that was affected.

VPCA application and identification of component loading patterns

Following the pre-processing steps, average reflectance and derivative spectra were calculated for each image to assess the similarity of the input data and the degree of multicollinearity present. Pearson's correlation coefficient and associated statistical significance tests quantified the degree of correlation between the bands in each image. The KSU spectral decomposition method was then applied independently to each of the swaths both with and without strip removal. A full description of the theory and computation behind the KSU spectral decomposition method is discussed in detail in ESM Appendix S1. VPCA results for the striped vs. destriped image test for Swath 15, which are representative of the data set are presented along with the results of the destriped VPCA analysis for each swath. An error budget for the striped vs. destriped image test for Swath 15 was used to determine whether the VPCA method was sensitive to the order of operations. The variance explained by each component, the component loading spectral patterns and the component score spatial patterns for the four swaths are qualitatively and quantitatively compared to evaluate the reproducibility of the decomposition method.

Results

Stripe removal

Vertical, along-track, linear features are present in HSI2 062116 Swath 15_MBSP as a series of alternating bright and dark 1-pixel wide stripes and 5 to 25-pixel wide bands (Fig. 3a & b). The sun is illuminating the track swath from east to west or right to left in the image. There are two broad bands present in the image. A darker band is present in the eastern portion of the image, while a brighter band is found along the western edge of the swath. The horizontal environmental bands and filaments are not linear features (Fig. 3b). The vertical stripes and bands obscure the environmental, horizontal fronts and filament, which are oriented approximately parallel to the coast along the southern edge of the image. These fronts and filaments represent the coastal transition zone into the Maumee Bay plume and the offshore waters of the Western Basin of Lake Erie heading to the north. There are also wave crests and whitecaps that can be observed in the image, running approximately perpendicular to a NE to SW line (Fig. 3b–c).

We performed a series of tests using the SPEAR Vertical Stripe Removal tool to optimize its performance at stripe removal (not shown). We first explored the performance of the SPEAR tool using the default setting, which is a 5% histogram outlier mask on radiance values. In the default setting, three passes of the SPEAR Vertical Stripe Removal tool were required to remove all the along track features. The process was iterative, due to differences in radiance and variable thickness of the stripes and bands. The thin stripes were removed more easily than the thicker bands. The bright band along the western edge of the image required three passes to be removed. To optimize the stripe removal process, we then varied the value of the histogram outlier mask

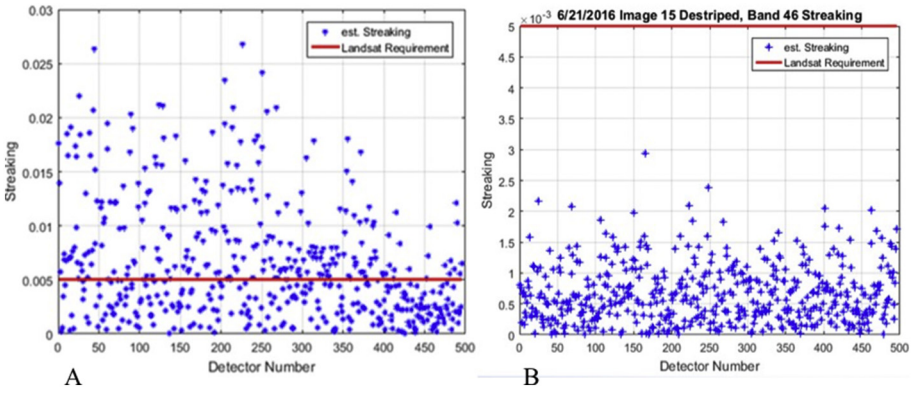

Fig. 4. Streaking metric (S_i) analysis of the NASA HSI2 062116 Swath 15_MBSP image recorded by the HSI2 sensor before (A) and after (B) vertical stripe removal. Note the change in scale between panels A and B. The red line shows the results relative to the Landsat 8 requirement.

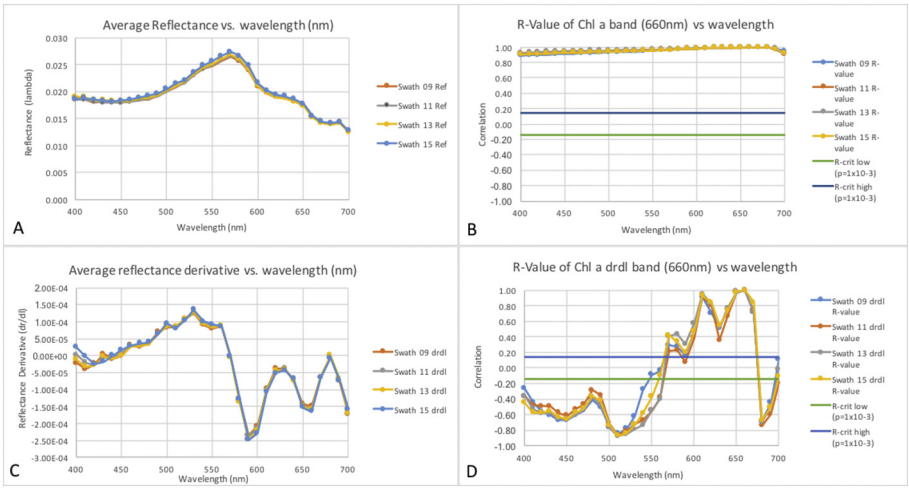

Fig. 5. Average reflectance (A), the correlation of the reflectance bands vs. 660 nm (B), the derivative of the reflectance spectra (C), and the correlation of the derivative bands vs. 660 nm (D) for the four HSI2 swaths collected off MBSP.

threshold from 10% to 0%. As expected, increasing the threshold beyond 5% did not remove the stripes in a single pass. The proportion of stripes removed during a single pass increased as the histogram outlier mask percentage decreased. Setting the outlier mask to 0% removed all the stripes in a single pass for each of the images studied (Fig. 3c). Accordingly, we set the outlier mask to 0% for our standard processing procedure. Removal of the stripes and bands allows the horizontal, environmental features in the image, along with the wave crests to stand out clearly (Fig. 3c). To demonstrate the effectiveness of the stripe removal process (Fig. 4), the before and after destriping images for Swath 15 were evaluated using the Landsat streaking index (NASA, 2009; USGS Landsat 8 Users Handbook, 2016). The results (Fig. 4) easily meet the Landsat 8 maximum streaking requirement of $S_i \leq 0.005$. It is worth noting that a streaking metric around 0.0025 is roughly where streaks first becomes visible in homogeneous, unaltered imagery.

Hyperspectral multicollinearity

The average reflectance spectra calculated for each of the four destriped swaths are virtually identical (Fig. 5a). Reflectance values on the blue end of the spectrum are 0.019 at 400 nm and gradually decrease to ~450 nm before gradually increasing to a maximum of ~0.027 at 570 nm, then dropping to 0.021 on the shoulder that starts at 610 nm and which then gradually decreases to 0.018 at 650, before dropping to a second shoulder with a value of 0.015 at 660 nm, which drops gradually to 0.013 at 700 nm. The correlation of each band against the 660 nm band was calculated to determine the correlation of the reflectance spectra with the absorption feature related to chl-a (Fig. 5b). With a minimum Pearson's correlation coefficient > 0.9 against the 660 nm band, all the visible bands are significantly correlated at $p \ll 0.001$, given the degrees of freedom, which are >970 in all cases.

The average first derivative of the reflectance spectra for each of the four swaths are also nearly identical (Fig. 5c). Minima in the reflectance derivatives are centered on 410 to 420 nm, 590 nm, 660 nm and 700 nm, while maxima occur at 530 nm, 620 to 630 nm, and 680 nm. The correlation of each band in the derivative spectra against 660 nm was also calculated (Fig. 5d). While there is more structure to the correlation of the derivative bands as a function of wavelength, virtually all of them are also significantly correlated with the band centered on 660 nm. These results document the need to apply some means of unmixing, such as VPCA, to remove multicollinearity from the hyperspectral data set. The impact of the destriping on the VPCA is demonstrated below by comparing the results of a VPCA conducted on the Swath 15 data set before destriping (Fig. 6) against the results after destriping (Fig. 7).

Error budget

A series of order of operations experiments were used to evaluate how each step after destriping influences the partitioning of signal

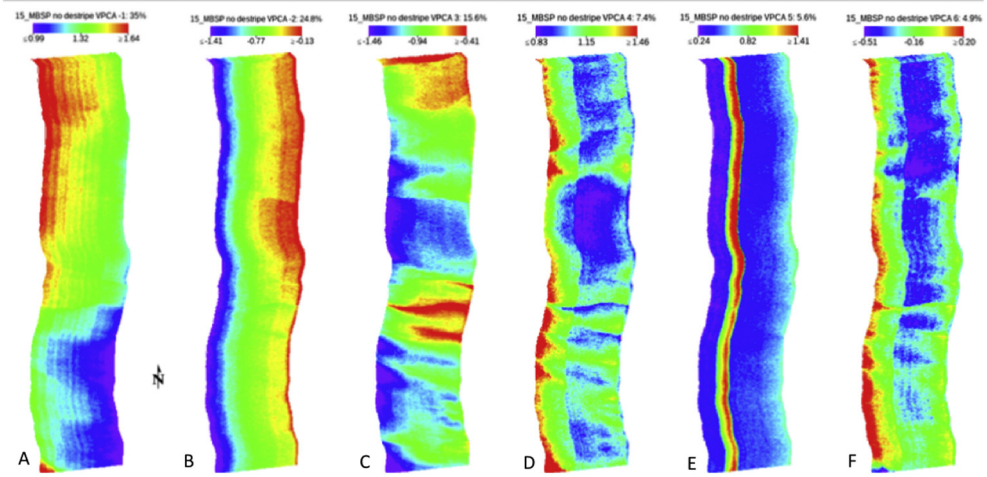

Fig. 6. Spatial patterns for VPCA component loading 1–6 (panels A–F) extracted from the striped image of NASA HSI2 062116 Swath 15_MBSP.

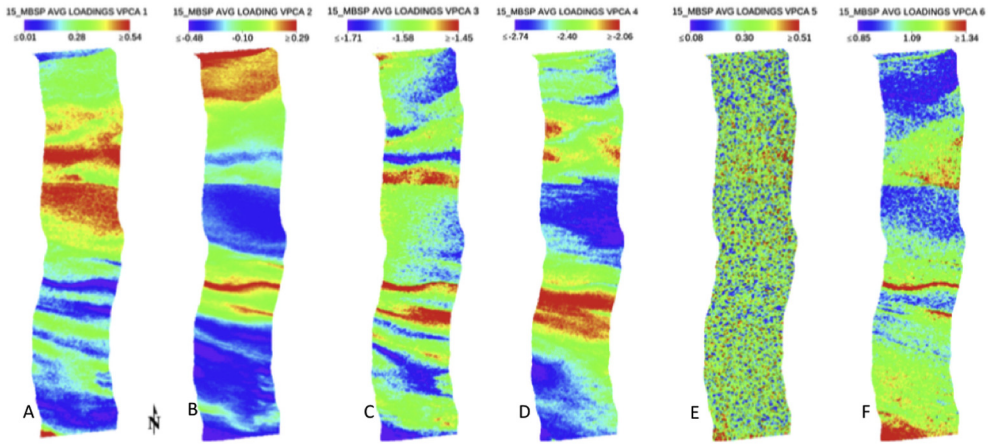

Fig. 7. Spatial patterns for VPCA component loading 1–6 (panels A–F) extracted from the destriped image of NASA HSI2 062116 Swath 15_MBSP.

from error during the VPCA. To determine an error budget for the order of operations intercomparison, the variance (in percent) for Swath 15 was tracked as total variance (σ_T^2), which was partitioned into environmental signal (σ_s^2), residual atmospheric noise (σ_a^2), and stochastic noise (σ_e^2). The budget can thus be expressed as:

$$\sigma_T^2 = \sigma_s^2 + \sigma_a^2 + \sigma_e^2 \tag{2}$$

The environmental signal is the sum of the variance of the components that exhibited coherent spectral and spatial patterns. The residual atmospheric noise exhibited a random spatial pattern and associated spectral pattern with highest loadings toward the blue end of the spectrum (illustrated below). The stochastic noise is defined as the (uncorrelated) fraction of variance that was not extracted as part of the VPCA model. The cutoff used for this determination is the number of eigenvectors >1, plus one more. The total noise $\sigma_{(a+e)}^2$ that can be extracted and discarded can thus be taken as the sum of the residual atmospheric noise and the stochastic noise.

The signal to noise ratio is defined as the ratio of the signal variance divided by the noise variance:

$$SNR = \sigma_s^2 / \sigma_{(a+e)}^2 \tag{3}$$

The data is unit normalized, such that the sums of squares for each variable sums to a total of 1. Thus the total unit-normalized variance is equal to the number of wavelength bands, N, in the matrix. With the unit normalized eigenvalues as the measure of variance, Eq. (3) can be used to determine the initial SNR ($SNR_{initial}$) for each component prior to extraction, which can be compared with the final SNR after extraction ($SNR_{extracted}$). Because the bands in the visible are highly correlated, the initial SNR associated with each extracted signal is the signal variance ($\boldsymbol{\gamma}$) divided by the total variance (N), which is equal to the number of bands, minus the signal variance for each component, m:

$$SNR_{initial} = \boldsymbol{\gamma_m} / (N - \boldsymbol{\gamma_m}) \tag{4}$$

In contrast, the extracted SNR is provided by the signal variance for each component divided by the stochastic noise variance, because each of the components are orthogonal. The stochastic noise variance is given by the difference between the total variance and the extracted signal variance, which is the sum of the m extracted eigenvalues:

$$SNR_{extracted} = \boldsymbol{\gamma_m} / \left(N - \sum_{m}^{M} \boldsymbol{\gamma_m} \right) \tag{5}$$

The SNR gain (SNR_{gain}) is then determined by dividing the extracted SNR by the initial SNR. This provides a measure of the signal amplification during the orthogonal decomposition:

$$SNR_{gain} = SNR_{extracted} / SNR_{initial} \tag{6}$$

Table 1
Order of operation error budget comparison for NASA HSI2 062116 Swath 15_MBSP.

Order of operations	Environmental signal	Residual atmospheric noise	Stochastic noise	Total error	Total variance
Case 1	91.68	3.99	4.33	8.32	100.00
Case 2	91.68	3.99	4.33	8.32	100.00
Case 3	91.65	4.02	4.33	8.35	100.00
Case 4	91.65	4.02	4.33	8.35	100.00

The error analysis results are presented in greater detail below, but it is useful to consider the error budget here as it provides the basis for determining the preferred order of operations. The environmental sign is the sum of the variance explained by the extracted components less the residual atmospheric noise component (Table 1). The stochastic noise is the variance associated with unextracted eigenvalues, which is given by the difference of 100 minus the variance of the extracted components. This error analysis demonstrates that the dependence on the order of operations for Cases 1–4 (Fig. 3) is insignificant, amounting to 0.0 to 0.03% between cases, and the total error budget is remarkably consistent (Table 1). The variation results from insignificant changes in the magnitude of the residual atmospheric noise relative to the environmental signal that is extracted. That difference is slightly less for Cases 1 and 2, and slightly more for Cases 3 and 4. This could potentially indicate that there is a very slight dependence on when the transformation from radiance to reflectance takes place, but no sensitivity to whether the data were spectrally or spatially smoothed first or second. The total error (residual atmospheric plus stochastic) for this swath ranged from 8.32 to 8.35%. The method is thus effectively insensitive to the order of operations within the observed margin of error.

Preferred workflow

Because the differences in the order of operations were insignificant, it is reasonable to select Case 2 as the preferred workflow. This enables keeping several IDL steps grouped, together as a logistical consideration. The preferred workflow is thus:

Case 2

1) Extraction of water pixels and cloud masking (ENVI)
2) Removal of invalid edge pixels, if present (ENVI)
3) Destriping using the SPEAR Vertical Stripe Removal tool (ENVI)

4) Ground Sample Distance (GSD) resampling (ENVI)
5) Spectral interpolation to 10 nm resolution (IDL)
6) Radiance to reflectance conversion to remove atmospheric contamination (IDL)
7) Spatial median smoothing using a 5 × 5 pixel kernel (ENVI)
8) Edge clipping to remove unsmoothed pixels (ENVI)
9) Derivative transformation (IDL)
10) VPCA decomposition (ENVI/IDL)
11) Georectification

The image is thus destriped, followed by GSD resampling, then spectral smoothing. From a conceptual standpoint, this order is appropriate because the GSD sampling yields pixels that conserve radiance as early as possible when using this approach. The image must be destriped before the pixels are resampled, or power from the stripes and bands would be convolved into the image. This order of operations maximizes the extractable signal, while minimizing transitions between the ENVI and IDL work environments.

VPCA results: striped vs. destriped VPCA patterns

A total of six component loadings and component score maps were extracted from each swath, along with communalities and the percentage of variance explained by each component. The extracted component loadings were easily matched for spectral shape by visual inspection, which enabled calculation of the average and standard deviation of the component loadings across all swaths. The average component loadings for the six spectral shapes were then used to identify each unknown component following the methods described in detail in ESM Appendix S1. To document the effectiveness of the destriping process, we compared a VPCA solution extracted from the raw, striped image with the VPCA for the clean, destriped image. For the VPCA of the raw, striped image, we did not include Step 4, destriping, in the image processing workflow (Fig. 2). For the analysis of the clean images, we included all processing steps in the workflow.

VPCA of the raw, striped image

The resulting spatial patterns for the striped image (Fig. 6) display geometric patterns that obscure the underlying horizontal variations that are clearly present in the destriped RGB image (Fig. 3) and its VPCA decomposition (Fig. 7). The sum of the variance explained by the six leading components in the striped VPCA, which includes a mixture of environmental variance and stripe-related bias amounts to 93.6%.

The leading component from the striped image is dominated by a N-S along track pattern that is superimposed on an east-west cross-track contrast. There is a distinct spatial discontinuity about 1/3 of the way north from the southern end of the track and narrow N-S stripes are apparent in the image. The second component shows an even stronger along-track pattern, with thick edge banding. Both components exhibit spatial discontinuities that result from the vertical stripes within the image cutting across the horizontal filaments. We observed features like these in other HSI2 data that are processed without stripe removal. Although some spatial environmental variability is present in the leading two components, much of the environmental variability is compressed into the third and fourth components. The thick band running along the western portion of the image dominates the 5th component. The associated spectral patterns are noisy (not shown) and account for less variability than the components extracted from the destriped image (Fig. 7). The sum of the 1st and 2nd component loadings produces a spectral shape (not shown) that is similar to the leading component of the clean, destriped solution, which is described below.

In this solution, it is not possible to entirely separate the environmental variance from the stripe variance, but we can place bounds on their relative contribution due to the orthogonality constraint of the VPCA, because each component is independent. An upper limit for the environmental variance extracted by the VPCA is the sum of the variance of the 1st, 3rd, 4th and 6th components (Fig. 6), which amounts to ≤63% of the variance extracted. This is an upper limit on the environmental variance extracted because much of the signal in these components does not relate to the environment. A more conservative estimate of the environmental variability that is extracted is given by the variance sum of the 3rd, 4th and 6th components, which is ≥28%. The environmental variability extracted by the VPCA of the striped image thus likely lies between 28% and 63%. There is considerable striping observed in the 1st, 2nd, and 5th components. Those three components yield an upper limit on the noise variance of ≤66% of the variance in the image, because some of the variance in the 1st component is environmental. The variance related to the striping thus lays between the variance of the 2nd and 5th components, which is 31% and 66%, the variance of all three. This indicates that relatively equal proportions of environmental and striped variance are present in the image, but the ratio of the two could range from ~1/3 to ~2/3.

VPCA of the clean, destriped images

In contrast, after removal of the vertical stripes and bands, the linear, geometric patterns and sharp spatial discontinuities present in the components extracted from the raw, striped image are gone (Fig. 7). Instead, five of the six leading components are dominated by horizontal environmental spatial patterns (VPCA components 1–4 and 6), which are consistent with the features observed in the destriped RGB image (Fig. 3). The spatial pattern for the 5th component is reduced to random

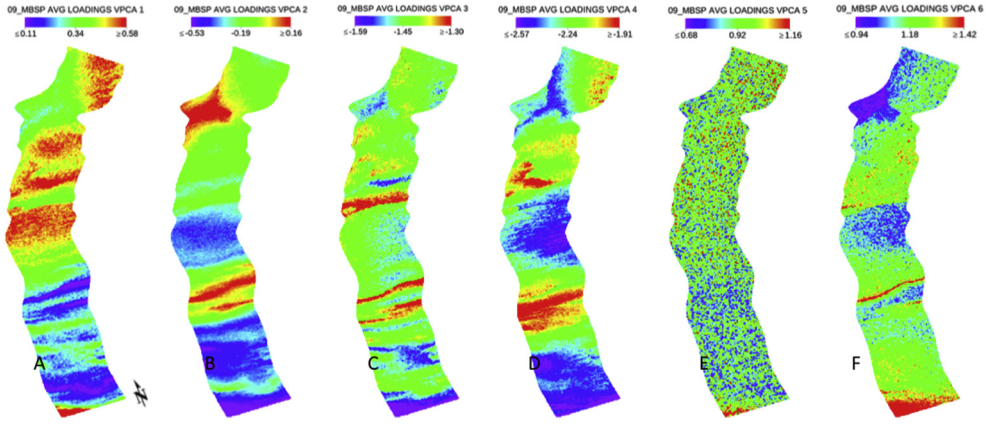

Fig. 8. Spatial patterns for VPCA component loading 1–6 (panels A–F) extracted from the destriped image of NASA HSI2 062116 Swath 09_WE6_MBSP. Note that the 3rd and 4th components have switched rank in NASA HSI2 062116 Swath 09_WE6_MBSP relative to the other swaths.

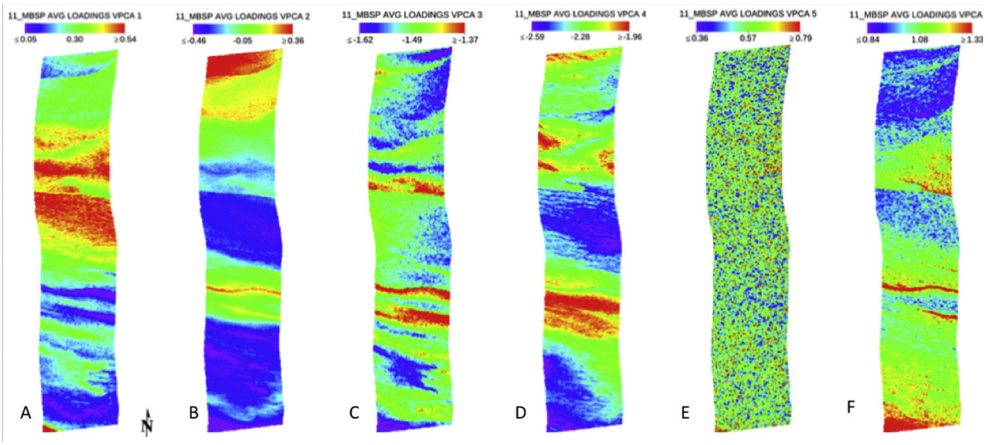

Fig. 9. Spatial patterns for VPCA component loading 1–6 (panels A–F) extracted from the destriped image of NASA HSI2 062116 Swath 11_MBSP.

noise, with a spectral signal that is strongest in the blue, consistent with atmospheric scattering as discussed further below. This spectral residual atmospheric noise accounts for 3.99% of the variance in the image. An additional 4.33% noise variance was not extracted along with the leading six components and is distributed in the trailing eigenvalues associated with the noise floor.

The five environmental components in the cleaned image account for 91.68% of the variance in the data set, indicating that the noise in the image has been reduced to 8.32% in the destriped analysis. This is a considerable improvement over the result from the raw, striped image in which the only relatively clean environmental signal that could be extracted was limited to components 3, 4, and 6, which together captured only 28% of the variance in the striped image. The destriping and smoothing processes have increased the amount of environmental variance that can be cleanly extracted from a minimum of 28% to 91.68; an increase of ~3.3×. The spectral patterns exhibit less noise than those observed in the component loadings extracted from the striped image (not shown).

Spatial reproducibility

To document the overall reproducibility of the method, we applied the KSU spectral decomposition method using the Case 2 processing order (Fig. 2) to the four HSI2 swaths collected over MBSP. These swaths were collected in the same region (Fig. 1) with four overpasses in a short time span (26 min). The component scores enable visualization of how the spatial patterns varied over the 26 min of the collection sequence. The swaths were collected in sequential order, with Swath 9 collected first and Swath 15 collected last. The numbers are not consecutive because additional swaths were collected on different lines while the aircraft circled for subsequent passes. The results from the additional images provide an upper limit on the reproducibility of the KSU spectral decomposition method and the quality of the NASA Glenn HSI2 sensor output, because some environmental changes exist due to the time offsets during the sample collection, but these should be relatively small due to the short duration of the experiment from start to finish. The VPCA results from all the images extract the same six components (Figs. 7–10), with similar spectral shapes (Fig. 11), and variance for each component (Table 2).

The effectiveness of the VPCA decomposition can be assessed based on the *SNR* analysis for each swath. This gain is reported as SNR_{gain}. When presented in terms of signal to noise ratio, the improvements are dramatic (Table 2). The increase in signal to noise ratio based on the unit-normalized eigenvalues ranges on average from greater than a factor of 7.37× to >19.04× times, with the largest increases in signal to noise ratio associated with the smaller eigenvalues.

Spectral reproducibility

The reproducibility of the spectral patterns - the component loadings - extracted from each of the four swaths is very good. Except for some sign flips, the components exhibit the same spectral shapes across swaths (Fig. 11). Due to the rotation process and the fact that we analyzed each swath independently, some of the components flip sign or change rank from solution to solution. In this case, component 3 and 4 switched position in NASA HSI2 062116 Swath 09_WE6_MBSP relative

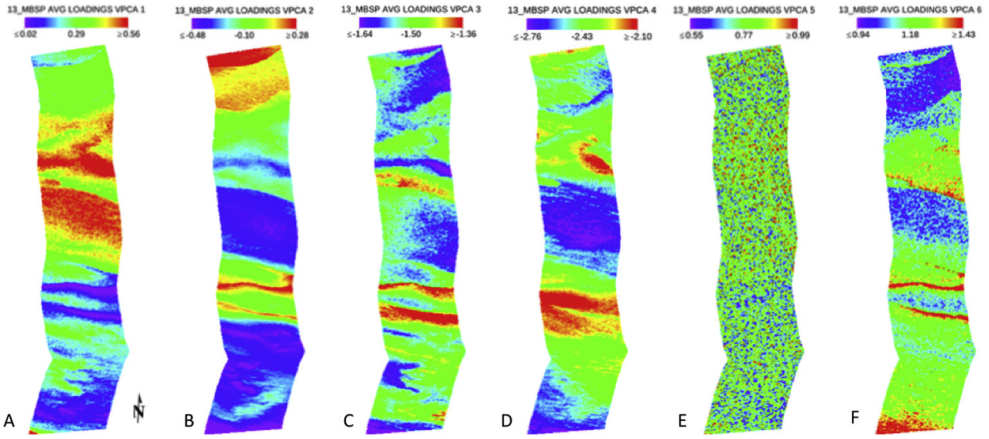

Fig. 10. Spatial patterns for VPCA component loading 1–6 (panels A–F) extracted from the destriped image of NASA HSI2 062116 Swath 13_MBSP.

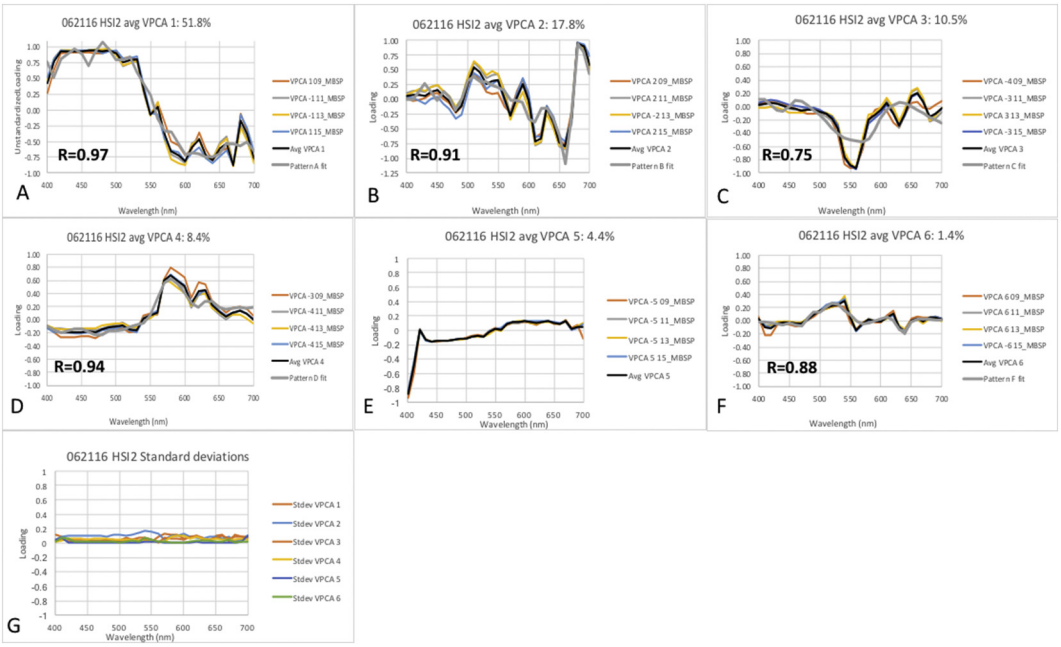

Fig. 11. Component loadings for the six spectral patterns (A–F) extracted from the four Swaths. The seventh panel (G) represents the standard deviation of the measurements from the four swaths. The colored lines represent the spectral shapes extracted from each swath. The average for each swath is plotted as a bold black line. The best fit is the bold gray line. The associated Pearson's correlation R value is provided.

to the other swaths. For this reason, the matched components are referred to as Patterns A-F, rather than by component rank (Tables 3, 4). It is possible to address this issue by matching the components based on their spectral shapes, and by keeping track of the sign convention during processing or by averaging the independent solutions and then using the average VPCA loading matrix to generate consistent VPCA scores for all images. Components that have been flipped in sign are denoted with a negative sign in their name. The VPCA scores (spatial patterns) for the individually processed swaths were virtually identical to the results for each swath when the average of the individual component loadings were applied to each swath (not shown).

The standard deviation of the component loadings as a function of wavelength ranged from ±0.003 to ±0.17, but averaged ±0.05 and were essentially flat across the spectral range, indicating that is unlikely that any significant components remain in the images after extraction of the leading six (Fig. 11).

Table 2
Signal to noise ratio associated with the clean VPCA decomposition.

	VPCA 1	VPCA 2	VPCA 3	VPCA 4	VPCA 5	VPCA 6	Total
HSI2_062116_09_MBSP_MTRIcorr							
Eigenvalues	17.84	4.51	3.97	1.08	0.92	0.38	28.70
Initial SNR	1.35	0.17	0.15	0.04	0.03	0.01	
Extracted SNR	7.74	1.96	1.72	0.47	0.40	0.16	
SNR Gain	5.71	11.50	11.73	12.99	13.06	13.29	
HSI2_062116_11_MBSP_MTRIcorr							
Eigenvalues	18.66	6.52	2.45	0.75	0.59	0.34	29.31
Initial SNR	1.51	0.27	0.09	0.02	0.02	0.01	
Extracted SNR	11.02	3.85	1.45	0.45	0.35	0.20	
SNR Gain	7.29	14.46	16.86	17.87	17.97	18.11	
HSI2_062116_13_MBSP_MTRIcorr							
Eigenvalues	19.96	5.73	2.32	0.73	0.55	0.30	29.60
Initial SNR	1.81	0.23	0.08	0.02	0.02	0.01	
Extracted SNR	14.27	4.10	1.66	0.52	0.40	0.22	
SNR Gain	7.89	18.07	20.50	21.64	21.77	21.95	
HSI2_062116_15_MBSP_MTRIcorr							
Eigenvalues	19.48	5.64	2.86	0.70	0.62	0.36	29.66
Initial SNR	1.69	0.22	0.10	0.02	0.02	0.01	
Extracted SNR	14.51	4.20	2.13	0.52	0.46	0.27	
SNR Gain	8.58	18.88	20.96	22.56	22.62	22.82	
Average SNR Gain	7.37	15.73	17.51	18.77	18.85	19.04	
Stdev SNR Gain	1.22	3.41	4.27	4.35	4.36	4.34	

The spectral shapes of the individual and averaged component loadings are similar to the results from Swath 15, published in Ortiz et al. (2017), although the component loadings presented in Ortiz et al. (2017) are standardized, while we plot them in their unstandardized form here. The component identifications here differ somewhat from our earlier work because the results in this study are based on averages of results obtained from four swaths, rather than one. Stepwise, least-squares regression of the components against the constituent standards in the spectral library was able to find statistically significant mixtures for all of the spectral patterns with the exception of Pattern D, the residual atmospheric noise component (Table 3). Component Pattern A is identified as a mixture of smectite, chlorite and α-carotene. This component loading pattern was positively correlated with all three constituents. Pattern B is a mixture that is positively correlated with diatoms and quartz, but negatively correlated with neoxanthin, a carotenoid found in green algae and terrestrial plants (Takaichi, 2011), and muscovite. Pattern C is positively correlated with goethite. Pattern D is a mixture that is positively correlated with hematite and phycocyanin, but negatively correlated with chlorite. Pattern E is the spectral signature that is inversely correlated with the residual atmospheric noise. Pattern F represents a mixture that is positively correlated with myxoxanthophyll, an accessory pigment found in cyanobacteria, and chl b, but negatively correlated with phycocyanin and hematite. These five environmental components are composed of constituents that are known to occur in the western basin of Lake Erie.

The uncertainty on the variance explained for each component ranged from a minimum of ±0.08% to a maximum of ±3.15%, averaging ±1.69% on a component by component basis (Table 4). The average uncertainty for residual atmospheric noise (Pattern E) was 4.28 ± 0.48%, while the random noise that was unexplained by the VPCA model associated with the eigenvalues that were not extracted was 5.43 ± 1.42%. The total errors on the method, which add in quadrature, was thus 6.93 ± 1.40%. We confirmed the origin of the residual atmospheric noise as a path radiance residual relating to the atmospheric correction by conducting an analysis on the uncorrected, radiometric data prior to

Table 3
Stepwise, least-squares regression statistics.

Pattern A						
R	R-squared	Adjusted R-squared	S	F	*p*-Value	
0.97	0.94	0.93	0.26	209.79	1.40E−17	
VAR	Coefficient	Standard error	t	*p*-Value > t	VIF	SBC
Smectite + chlorite	0.95	0.05	20.15	3.36E−18	1.00	−183.37
a-Carotene	0.15	0.05	3.08	4.57E−03	1.00	

Pattern B						
R	R-squared	Adjusted R-squared	S	F	*p*-Value	
0.91	0.82	0.79	0.45	29.80	2.27E−09	
VAR	Coefficient	Standard error	t	*p*-Value > t	VIF	SBC
Bacillariophyceae	0.95	0.09	10.42	8.94E−11	1.21	−141.60
Quartz (Ottawa Sand)	0.32	0.09	3.60	1.32E−03	1.12	
Neoxanthin	−0.24	0.09	−2.50	1.91E−02	1.29	
Muscovite	−0.22	0.09	−2.41	2.34E−02	1.22	

Pattern C						
R	R-squared	Adjusted R-squared	S	F	*p*-Value	
0.75	0.56	0.54	0.68	36.80	1.33E−06	
VAR	Coefficient	Standard error	t	*p*-Value > t	VIF	SBC
Goethite	−0.75	0.12	−6.07	1.33E−06	1.00	−127.37

Pattern D						
R	R-squared	Adjusted R-squared	S	F	*p*-Value	
0.94	0.89	0.87	0.36	70.26	7.08E−13	
VAR	Coefficient	Standard error	t	*p*-Value > t	VIF	SBC
Hematite	0.77	0.09	8.82	1.97E−09	1.80	−160.33
Phycocyanin	0.24	0.07	3.54	1.46E−03	1.06	
Chlorite	−0.23	0.09	−2.69	1.22E−02	1.77	

Pattern F						
R	R-squared	Adjusted R-squared	S	F	*p*-Value	
0.88	0.78	0.75	0.51	22.96	3.27E−08	
VAR	Coefficient	Standard error	t	*p*-Value > t	VIF	SBC
Myxoxanthophyll	0.528	0.099	5.34	1.38E−05	1.15	−135.13
Phycocyanin	−0.40	0.098	−4.09	3.74E−04	1.13	
Chlorophyll b	0.268	0.099	2.70	1.19E−02	1.16	
Hematite	−0.248	0.093	−2.66	1.32E−02	1.02	

atmospheric correction (not shown). We can compare the results from the uncorrected radiance derivative data with the reflectance derivative data because the VPCA is based on the correlation between bands, not their absolute values. The component shapes extracted in the uncorrected case were similar to those extracted by the VPCA of the atmospherically corrected data, but with additional spectral noise. The spatial patterns also looked like the atmospherically corrected spatial patterns, but included more fine-scale heterogeneous noise. Of

Table 4
Variance explained by each component.

Pattern and interpretation	Pattern A: +smectite, +chlorite, +a-carotene	Pattern B: +diatoms, +quartz, −neoxanthin, −muscovite	Pattern C: −goethite	Pattern D: +hematite, +phycocyanin, −chlorite	Pattern E: residual Atm noise	Pattern F: +myoxanthophyll, −phycocyanin, +chl b, −hematite	Variance	Residual random noise	Total noise
Swath									
JGLR_062116_09_MBSP_MTRIcorr	VPCA 1	VPCA 2	VPCA −4	VPCA −3	VPCA −5	VPCA 6			
	48.46	15.30	11.54	10.96	4.99	1.32	92.57	7.43	8.95
JGLR_062116_11_MBSP_MTRIcorr	VPCA −1	VPCA 2	VPCA −3	VPCA −4	VPCA −5	VPCA 6			
	51.62	21.03	10.25	6.20	4.04	1.41	94.54	5.46	6.79
JGLR_062116_13_MBSP_MTRIcorr	VPCA −1	VPCA −2	VPCA 3	VPCA −4	VPCA −5	VPCA 6			
	54.68	21.02	8.78	5.64	4.08	1.29	95.49	4.51	6.08
JGLR_062116_15_MBSP_MTRIcorr	VPCA 1	VPCA 2	VPCA −3	VPCA −4	VPCA 5	VPCA −6			
	55.34	17.06	9.84	7.99	3.99	1.46	95.67	4.33	5.89
Average	52.52	18.60	10.10	7.70	4.28	1.37	94.57	5.43	6.93
Standard deviation	3.15	2.89	1.14	2.40	0.48	0.08	1.42	1.42	1.40

importance, Pattern E was not present in this solution confirming that its origin is related to the atmospheric correction method.

Discussion

All visible reflectance data exhibit strong correlations between bands due to the nature of the electronic processes that drive absorption in that part of the electromagnetic spectrum (Kokaly et al., 2017). A strong correlation is present in the visible part of the spectrum for the NASA Glenn HSI2 data set (Fig. 5) that must be removed to allow multiple bands to be used in concert in data products. Unmixing the spectra allows one to avoid errors arising from multicollinearity and assumptions of independent input variables in multivariate data analysis. This provides justification for application of the KSU spectral decomposition method, which enables extraction of the maximum amount of non-redundant information from the hyperspectral data set.

The full atmospheric correction procedure employed with these data is discussed in detail in Ortiz et al. (2017). Ortiz et al. (2017), also documented that the KSU spectral decomposition method was relatively insensitive to the type of atmospheric correction that was applied. Four variations of the empirical line method were applied: the EML0, ELM1 and ELM2 methods made use of curved mirrors to reflect downwelling irradiance back to the sensor for use as part of the atmospheric correction. Field measurements of downwelling irradiance and surface reflectance measured with a Malvern Panalytical ASD FieldSpec Pro HH are then used to calculate surface reflectance using relationships derived from radiative transfer theory. The fourth method explored in Ortiz et al. (2017), the Michigan Tech Research Institute (MTRI) Blacktop Calibration method uses an upward facing Malvern Panalytical ASD FieldSpec Pro HH mounted in the NASA Glenn aircraft to obtain downwelling irradiance to calculate at-sensor remote sensing reflectance in conjunction with the NASA Glen HSI2 radiance measurements. The at-sensor reflectance is then converted to surface reflectance by comparison with reflectance measurements of a well-characterized, blacktop surface at Maumee Bay State Park, the NASA Glenn CyanoHAB field calibration site. In Ortiz et al. (2017) the ELM0 method, which explained the highest fraction of variance in swath 15 was used to identify the composition of the components. Here we analyzed all four swaths using the MTRI blacktop method because that approach can be applied without the need to deploy mirror targets in each scene.

Ortiz et al. (2017) documented that the KSU spectral decomposition method was capable of isolating and removing residual atmospheric and stochastic error variance that was present in the hyperspectral data set. This study confirms that the residual atmospheric noise associated with VPCA 5 (Pattern E) is found in multiple swaths collected on the same day and is in fact related to residual atmospheric noise. We conclude this because that component is absent from a spectral decomposition applied on data that had not been atmospherically corrected, but which was otherwise processed identically to the other results presented here. The major findings of this study are to document the impact of striping on VPCA, the insensitivity of the KSU spectral decomposition method to order of operations and to quantify its enhancement to SNR and reproducibility. Understanding the reproducibility of the KSU spectral decomposition method is important to enable its optimal use with future, planned hyperspectral orbital missions, such as PACE, or future sensors needed for the Surface Biology and Geology mission as defined in the 2018 NASA Decadal Survey (National Academies of Sciences, Engineering, and Medicine, 2018).

Pre-processing of remote sensing images is critical to enable the production of derived products and is standard procedure for processing of images such as those collected by the NASA HSI2. However, the impact of stripes and bands on the KSU spectral decomposition method has not been previously explored. A variety of stripe and band removal approaches have been proposed to improve image quality (e.g., Shen and Zhang, 2009). Removal of stripes and bands qualitatively enhances the visibility of features and enables calculation of derived products that rely on radiometric consistency between pixels within a scene. These approaches can be grouped into tools that remove stripe-related variance by equalizing the means and/or standard deviations between pixels (Method of Moments: Gadallah et al., 2000), by equalizing the histograms of radiometric values between bands (Horn and Woodham, 1979; Algazi and Ford, 1981; Rakwatin et al., 2007), or that filter out stripe related noise using low pass filters such as infinite impulse response filters, wavelets or Fourier analysis methods (Statistical Filtering Methods: Torres and Infante, 2001; Chen et al., 2003; Chen et al., 2017). Some hybrid approaches make use of a combination of one or more of these subclasses (Chang et al., 2007a, 2007b; Münch et al., 2009; Shen and Zhang, 2009). Iterative approaches are often quite effective at removal of stripes and bands (Bouali, 2010). The most effective methods appear to be the hybrid methods or iterative approaches (e.g., Bouali and Ladjal, 2011; Lastri et al., 2015; Liu et al., 2016a, 2016b; Chen et al., 2017). The method of moment tools were among the earliest approaches developed. Limitations of the method are discussed in Horn and Woodham (1979). Method of moment tools seem to be the least effective as they fail in the case of non-linear or irregular stripes and can leave residual stripes or generate new stripes during the removal process (Horn and Woodham, 1979; Gadallah et al., 2000).

The Harris documentation of the proprietary ENVI destriping tools is not extensive although prior publications have evaluated their effectiveness in comparison to other methods (Scheffler and Karrasch, 2013; Lastri et al., 2015). These tools can provide acceptable results, particularly when the scene is radiometrically homogenous as is the case with the water portions of our images. The SPEAR Vertical Stripe Removal tool is designed to remove along-track features generated by push-broom sensors, while the horizontal DESTRIPE algorithm in the Raster Data Management package is designed to remove cross-track features in whisk-broom sensors. Both tools require Level 1 data. The ENVI vertical stripe removal tool requests the fraction of the radiometric histogram to employ during strip removal and is thus most likely a histogram correction tool. The DESTRIPE algorithm in the Raster Data Management package requests information on the number of detectors and is thus most likely a method of moment-based tool. There are two other ENVI legacy destriping tools available for the PC version of ENVI that are both based on the method of moments. We did not employ either of these other tools, which are redundant with the DESTRIPE tool available for the Mac version of the ENVI software employed at KSU. We compared results from the SPEAR destriped images with ones destriped using the DESTRIPE tool and determined the two ENVI methods produced similar results. We have found the SPEAR Vertical Stripe Removal tool to be easier to use for our purposes because conceptually, it was designed for use with push-broom sensors. As a result, the image does not need to be transposed for use with this tool. We are employing it as part of our standard processing protocol as needed. Use of this tool lets us avoid two transposition steps during image processing.

Criteria that have previously been used to compare the quality of destriping tools in the literature are: visual inspection of images to look for residual striping, analysis of cross-track means to look for outlier pixels, calculation of image *SNR* statistics in several ways before and after stripe removal, and calculation of difference images that subtract the destriped from the striped image to isolate the destriped signal. Often destriping assessment is based on the arbitrary selection of a single band for visualization. In our application, we employ VPCA to compare the spectral and spatial structure of the images before and after stripe removal. This is a useful approach because VPCA removes redundant information from images, allowing presentation of the maximum amount of information in the minimum number of optimal linear band combinations. Based on our results, in future applications, VPCA could be used successfully to assess sensor design choices or image pre-processing procedures as it is employed here.

The results of this study document that removal of stripes enables better partitioning of the environmental variance within the scene

(Fig. 6 vs. Fig. 7). By partitioning the variance into orthogonal components, the KSU spectral decomposition method can enhance the average SNR of the individual components by a factor of 7.37× to 19.04×.

The component scores, indicating the spatial pattern extracted from the cleaned images, lack any geometric bias that can be attributed to flight geometry or instrument design in the clean analysis in Figs. 7–10. The VPCA process itself is also able to partition the environmental signal from residual atmospheric path radiance bias that was not removed during the conversion from radiance to reflectance (Ortiz et al., 2017). Forward, stepwise multiple linear regression of the standardized, component loading spectral patterns can then be used to identify the composition of the signals that give rise to the spectral patterns in the image. The shape of the spectral component loading signals extracted from Swaths 9, 11, and 13 were very similar to those for Swath 15 published in (Ortiz et al., 2017) as can be seen in Fig. 11.

This principal component regression based identification step is important because it solves one of the primary limitations of the unsupervised classification method, which is how to determine the identity of the extracted image features. Application of the method to multiple scenes documents that the reproducibility of the signal is 6.93 ± 1.40%. The method also addresses the mixed pixel problem by partitioning the variance within each pixel into its contribution from each of the extracted components. This powerful, unsupervised, soft classification technique thus addresses some of the major problems that vex remote sensing analysis: sensitivity to atmospheric correction, removal of stochastic noise, and partitioning and identification of image constituents to address the mixed pixel problem. Application of the KSU Spectral decomposition method holds great promise to enhance monitoring applications and develop retrospective time series to explore regional changes in climate and environmental conditions.

With the loss of the hyperspectral imager for the coastal ocean (HICO) from the International Space Station (Keith et al., 2014) and the decommissioning of Hyperion on the USGS Earth Observing 1 satellite there are no longer any non-commercial, hyperspectral visible remote sensing assets in Earth orbit, although a new hyperspectral instrument, the DLR Earth Sensing Imaging Spectrometer (DESIS) has been installed on the International Space Station and is beginning to deliver data. The results presented here are encouraging and indicate that the KSU spectral decomposition method is one approach that will enable users to capitalize on the additional non-redundant spectral information that can be extracted from proposed orbital hyperspectral sensors on missions such as PACE, or the Surface Biology and Geology mission proposed within the Decadal Survey (National Academies of Sciences, Engineering, and Medicine, 2018). Deploying these tools will allow enhanced determination of pigment-related biomass estimates and yield new capabilities to identify algal and cyanobacterial composition based on extraction of pigment-related spectra by visible derivative spectroscopy as well as a host of other applications.

Conclusions

Our work documents that we can extract multiple VPCA components from NASA Glenn HSI2 images, further enhancing their high SNR in the process. The VPCA signals extracted from the HSI2 are reproducible in terms of their spectral and spatial patterns. We have documented a standard processing stream for analysis of these images and demonstrated that the results are independent of the order of operations in which they are applied. A critical step in the pre-processing of the images is the removal of any along-track or cross-track stripes or bands arising from sensor-related issues. This powerful, unsupervised, soft classification technique thus addresses several major problems in visible remote sensing: insensitivity to atmospheric correction, removal of stochastic noise, and partitioning and identification of image constituents to address the mixed pixel problem.

Acknowledgments

This work was supported by grants from NASA (NNC15VB05P) and the Ohio Sea Grant Association (R/ES-021-PD; R/ER-105-PD). We thank the captains and crews of the ODNR Water craft vessels that provided transportation to field sites. The manuscript was improved by comments from three anonymous reviewers and the editors.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jglr.2019.03.005.

References

Algazi, V.R., Ford, G.E., 1981. Radiometric equalization of nonperiodic striping in satellite data. Comput. Graph. Image Process. 16, 287–295.
Ali, K.A., Witter, D.L., Ortiz, J.D., 2014. Multivariate approach to estimate color producing agents in Case 2 waters using first-derivative spectrophotometer data. Geocarto Int. 29 (2), 102–127. https://doi.org/10.1080/10106049.2012.743601.
Ali, K.A., Ortiz, J., Bonini, N., Shuman, M., Sydow, C., 2016. Application of Aqua MODIS sensor data for estimating chlorophyll a in the turbid Case 2 waters of Lake Erie using bio-optical models. GIScience Remote Sens. 53, 483–505. https://doi.org/10.1080/15481603.2016.1177248.
Antonelli, P., di Bisceglie, M., Episcopo, R., Galdi, C., 2004. Destriping MODIS data using IFOV overlapping. Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. IEEE International, pp. 4568–4571.
Avouris, D.M., Ortiz, J.D., 2019. Validation of 2015 Lake Erie MODIS image spectral decomposition using visible derivative spectroscopy and field campaign data. J. Great Lakes Res. (this issue). https://www.sciencedirect.com/science/article/pii/S0380133019300371.
Barsi, J.A., Lee, K., Kvaran, G., Markham, B.L., Pedelty, J.A., 2014. The spectral response of the Landsat-8 operational land imager. Remote Sens. 6, 10232–10251.
Bouali, M., 2010. A simple and robust destriping algorithm for imaging spectrometers: application to MODIS data. Proceedings of ASPRS 2010 Annual Conference, San Diego, CA, USA, pp. 84–93.
Bouali, M., Ladjal, S., 2011. Toward optimal destriping of MODIS data using a unidirectional variational model. IEEE Trans. Geosci. Remote Sens. 49, 2924–2935.
Brando, V.E., Anstee, J.M., Wettle, M., Dekker, A.G., Phinn, S.R., Roelfsema, C., 2009. A physics based retrieval and quality assessment of bathymetry from suboptimal hyperspectral data. Remote Sens. Environ. 113, 755–770.
Bullerjahn, G.S., McKay, R.M., Baker, D.B., Boyer, G.L., D'Anglada, L.V., Doucette, G.J., Ho, J.C., Irwin, E.G., Kling, C.L., Kudela, R.M., Kurmayer, R., Michalak, A.M., Ortiz, J.D., Otten, T.G., Paerl, H.W., Qin, B., Sohngen, B.L., Stumpf, R.P., Visser, P.M., Wilhelm, S.W., 2016. Global solutions to regional problems: collecting global expertise to address the problem of harmful cyanobacterial blooms. A Lake Erie case study. Harmful Algae 54, 223–238. https://doi.org/10.1016/j.hal.2016.01.003.
Chang, W.-W., Guo, L., Fu, Z.-Y., Liu, K., 2007a. A new destriping method of imaging spectrometer images. Wavelet Analysis and Pattern Recognition, 2007. ICWAPR'07. International Conference on Wavelet Analysis and Pattern Recognition, pp. 437–441 https://doi.org/10.1109/ICWAPR.2007.4420708.
Chang, W.-W., Lei, G., Zhaoyang, F., Kun, L., 2007b. A study on destriping methods of imaging spectrometer images. Electronic Measurement and Instruments, 2007. ICEMI'07. 8th International Conference on Wavelet Analysis and Pattern Recognition, pp. 766–771 https://doi.org/10.1109/ICEMI.2007.4350793.
Chen, Q., Gong, P., 2004. Automatic variogram parameter extraction for textural classification of the panchromatic IKONOS imagery. IEEE Trans. Geosci. Remote Sens. 42, 1106–1115.
Chen, J., Shao, Y., Guo, H., Wang, W., Zhu, B., 2003. Destriping CMODIS data by power filtering. IEEE Trans. Geosci. Remote Sens. 41, 2119–2124.
Chen, Y., Huang, T.-Z., Zhao, X.-L., Deng, L.-J., Huang, J., 2017. Stripe noise removal of remote sensing images by total variation regularization and group sparsity constraint. Remote Sens. 9, 559. https://doi.org/10.3390/rs9060559.
Eismann, M.T., 2012. Hyperspectral Remote Sensing. SPIE Press 9780819487872 (748 p.).
Eugenio, F., Marcello, J., Martin, J., 2015. High-resolution maps of bathymetry and benthic habitats in shallow-water environments using multispectral remote sensing imagery. IEEE Trans. Geosci. Remote Sens. 53, 3539–3549.
Gadallah, F.L., Csillag, F., Smith, E.J.M., 2000. Destriping multisensor imagery with moment matching. Int. J. Remote Sens. 21, 2505–2511.
Gao, B.-C., Davis, C., Goetz, A., 2006. A review of atmospheric correction techniques for hyperspectral remote sensing of land surfaces and ocean color. Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference On, pp. 1979–1981.
Horn, B.K.P., Woodham, R.J., 1979. Destriping Landsat MSS images by histogram modification. Comput. Graph. Image Process. 10, 69–83.
Houborg, R., McCabe, M.F., 2017. Impacts of dust aerosol and adjacency effects on the accuracy of Landsat 8 and RapidEye surface reflectances. Remote Sens. Environ. 194, 127–145.
Kachigan, S.K., 1991. Multivariate Statistical Analysis: A Conceptual Introduction. Radius Press, New York.
Kaiser, H.F., 1958. The varimax criterion for analytic rotation in factor analysis. Psychometrika 23, 187–200.

Kean, B.W., Kilcoyne, S., Malone, N.R., Wilberger, G., Troup, R., Miller, S., Brown, K.C., Vampola, J., 2012. Advancements in large-format SiPIN hybrid focal plane technology. Proc. SPIE 8511, Infrared Remote Sensing and Instrumentation XX, p. 851111 https://doi.org/10.1117/12.964392 24 October 2012.

Keith, D.J., Schaeffer, B.A., Lunetta, R.S., Gould Jr., R.W., Rocha, K., Cobb, D.J., 2014. Remote sensing of selected water-quality indicators with the hyperspectral imager for the coastal ocean (HICO) sensor. Int. J. Remote Sens. 35, 2927–2962.

Khodadadzadeh, M., Li, J., Plaza, A., Ghassemian, H., Bioucas-Dias, J.M., Li, X., 2014. Spectral—spatial classification of hyperspectral data using local and global probabilities for mixed pixel characterization. IEEE Trans. Geosci. Remote Sens. 52, 6298–6314.

Kokaly, R.F., Clark, R.N., Swayze, G.A., Livo, K.E., Hoefen, T.M., Pearson, N.C., Wise, R.A., Benzel, W.M., Lowers, H.A., Driscoll, R.L., Klein, A.J., 2017. USGS Spectral Library Version 7: U.S. Geological Survey Data Series 1035. 61 p. https://doi.org/10.3133/ds1035.

Kudela, R.M., Palacios, S.L., Austerberry, D.C., Accorsi, E.K., Guild, L.S., Torres-Perez, J., 2015. Application of hyperspectral remote sensing to cyanobacterial blooms in inland waters. Remote Sens. Environ. 167, 196–205.

Landsat 8, 2016. (L8) Data Users Handbook Version 2.0. Departmentof the Interior, USGS, LSDS-1574, Sioux Falls, South Dakota.

Lastri, C., Guzzi, D., Barducci, A., Pippi, I., Nardino, V., Raimondi, V., 2015. Striping noise mitigation: performance evaluation on real and simulated hyperspectral images. Proc. SPIE 9643, Image and Signal Processing for Remote Sensing XXI, p. 96430K https://doi.org/10.1117/12.2195706 15 October 2015.

Lekki, J., Anderson, R., Avouris, D., Becker, R., Churnside, J., Cline, M., Demers, J., Leshkevich, G., Liou, L., Luvall, J., et al., 2017. Airborne hyperspectral sensing of monitoring harmful algal blooms in the Great Lakes region: system calibration and validation. NASA-TM 2017–219071.

Liu, X., Lu, X., Shen, H., Yuan, Q., Jiao, Y., Zhang, L., 2016a. Stripe noise separation and removal in remote sensing images by consideration of the global sparsity and local variational properties. IEEE Trans. Geosci. Remote Sens. 54, 3049–3060.

Liu, X., Shen, H., Yuan, Q., Zhang, L., Cheng, Q., 2016b. A novel removal method for dense stripes in remote sensing images. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. III-6, 57–61. https://doi.org/10.5194/isprs-annals-III-6-57-2016.

Markham, B., Dabney, P.W., Storey, J.C., Morfitt, R., Knight, E.J., Kvaran, G., Lee, K., 2008. Landsat Data Continuity Mission Calibration and Validation. ASPRS Publ.

Martin, S., 2014. An Introduction to Ocean Remote Sensing. Cambridge University Press.

Münch, B., Trtik, P., Marone, F., Stampanoni, M., 2009. Stripe and ring artifact removal with combined wavelet—Fourier filtering. Opt. Express 17, 8567–8591.

NASA, 2009. Landsat Data Continuity Mission Operational Land Imager Requirements Document - Revision E. NASA Goddard Space Flight Center.

National Academies of Sciences, Engineering, and Medicine, 2018. Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation From Space. The National Academies Press, Washington, DC https://doi.org/10.17226/24938.

Ortiz, J.D., Witter, D.L., Ali, K.A., Fela, N., Duff, M., Mills, L., 2013. Evaluating multiple colour-producing agents in case II waters from Lake Erie. Int. J. Remote Sens. 34, 8854–8880. https://doi.org/10.1080/01431161.2013.853892.

Ortiz, J.D., Avouris, D., Schiller, S., Luvall, J.C., Lekki, J.D., Tokars, R.P., Anderson, R.C., Shuchman, R., Sayers, M., Becker, R., 2017. Intercomparison of approaches to the empirical line method for vicarious hyperspectral reflectance calibration. Front. Mar. Sci. 4, 296. https://doi.org/10.3389/fmars.2017.00296.

Palacios, S.L., Kudela, R.M., Guild, L.S., Negrey, K.H., Torres-Perez, J., Broughton, J., 2015. Remote sensing of phytoplankton functional types in the coastal ocean from the HysplRI Preparatory Flight Campaign. Remote Sens. Environ. 167, 269–280.

Palmer, C.A., Loewen, E.G., 2005. Diffraction Grating Handbook. Newport Corporation, New York.

Preisendorfer, R.W., Mobley, C.D., 1988. Principal component analysis in meteorology and oceanography. Developments in Atmospheric Science. 17. Elsevier Science Pub. Co., Amsterdam, New York, NY, U.S.A.

Rakwatin, P., Takeuchi, W., Yasuoka, Y., 2007. Stripe noise reduction in MODIS data by combining histogram matching with facet filter. IEEE Trans. Geosci. Remote Sens. 45, 1844–1856.

Raychaudhuri, B., 2012. Synthesis of mixed pixel hyperspectral signatures. Int. J. Remote Sens. 33, 1954–1966.

Scheffler, D., Karrasch, P., 2013. Preprocessing of hyperspectral images: a comparative study of destriping algorithms for EO1-hyperion. Proc. SPIE 8892, Image and Signal Processing for Remote Sensing XIX, p. 88920H https://doi.org/10.1117/12.2028733 17 October 2013.

Schott, J.R., 2007. Remote Sensing: The Image Chain Approach. Oxford University Press, Oxford, England, UK.

Schultz, K.I., Kelly, M.W., Baker, J.J., Blackwell, M.H., Brown, M.G., Colonero, C.B., David, C.L., Tyrrell, B.M., Wey, J.R., 2014. Digital-pixel focal plane array technology. Lincoln Lab. J. 20, 36–51.

Shen, H., Zhang, L., 2009. A MAP-based algorithm for destriping and inpainting of remotely sensed images. IEEE Trans. Geosci. Remote Sens. 47, 1492–1502.

Shlens, J., 2014. A Tutorial on Principal Component Analysis. Google Research, Mountain View, CA https://arxiv.org/abs/1404.1100 (1–12).

Smith, L.I., 2002. A Tutorial on Principal Components Analysis (Computer Science Technical Report No. OUCS-2002-12). Retrieved from. http://hdl.handle.net/10523/7534 , pp. 1–26.

Soja-Woźniak, M., Craig, S.E., Kratzer, S., Wojtasiewicz, B., Darecki, M., Jones, C.T., 2017. A novel statistical approach for ocean colour estimation of inherent optical properties and cyanobacteria abundance in optically complex waters. Remote Sens. 9, 1–22.

Takaichi, S., 2011. Carotenoids in algae: distributions, biosyntheses and functions. Mar. Drugs 9, 1101–1118.

Torres, J., Infante, S.O., 2001. Wavelet analysis for the elimination of striping noise in satellite images. Opt. Eng. 40, 1309–1315.