The Spectral Mixture Residual: A Source of Low-Variance Information to Enhance the Explainability and Accuracy of Surface Biology and Geology Retrievals

D. Sousa[1], P. Brodrick[2], K. Cawse-Nicholson[2], J. Fisher[3], R. Pavlick[2], C. Small[4], and D.R. Thompson[2]

[1]Department of Geography, San Diego State University, San Diego, CA. [2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA. [3]Joint Institute for Regional Earth System Science and Engineering, University of California Los Angeles, Los Angeles, CA. [4]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY.

Corresponding author: Daniel Sousa (dan.sousa@sdsu.edu)

Key Points:

- Mixture residual spectrum can contain informative low-variance signals.
- Average band-to-band correlations drop from 0.94 and 0.97 to 0.04 and 0.31 for near infrared and shortwave infrared, respectively.
- Linear approach offers feature accentuation, interpretability, generalizability, and computational efficiency.

Abstract

Mixed pixels are the rule, not the exception, in decameter terrestrial imaging. By definition, the reflectance spectrum of a mixed pixel is a function of more than one generative process. Physically-based surface biology or geology retrievals must therefore isolate the component of interest from a myriad of unrelated processes, heterogenously distributed across hundreds of square meters. Foliar traits, for example, must be isolated from canopy structure and substrate composition which can dominate overall variance of spatially integrated reflectance.

We propose a new approach to isolate low-variance spectral signatures. The reflectance of each pixel is modeled assuming linear geographic mixing due to a small library of generic endmembers. The difference between the modeled and observed spectra is deemed the Mixture Residual (MR). The MR, a residual reflectance spectrum that is presumed to carry the subtler and variable signals of interest, is then leveraged as a source of signal.

We illustrate the approach using three datasets: synthetic composites computed from field reflectance spectra, NEON AOP airborne image compilations, and DESIS satellite data. The MR discriminates between land cover versus plant trait signals and accentuates subtle absorption features. Mean band-to-band correlations within the visible, NIR, and SWIR wavebands decrease from 0.97, 0.94, and 0.97 to 0.95, 0.04 and 0.31. The number of dimensions required to explain 99% of image variance increases from 4 to 13. We focus on vegetation as an illustrative example, but note that the concept can be extended to other applications and used as an input to other algorithms.

Plain Language Summary

Imaging spectroscopy (IS) measures a rich superposition of signals generated by a wide range of biogeophysical processes. In many cases, subpixel mixing – variations in the abundance of broad land cover types like soil, foliage, and shadow – can dominate IS data at the expense of subtler signals like leaf chemistry and soil composition. This spatial mixing problem can seriously hinder surface biology and geology retrieval generalizability and accuracy. We propose a way to isolate the subtle signatures of interest from strong land cover mixing signals through a new approach called the mixture residual (MR). This approach models and removes the strong land cover mixing signal in a highly general, computationally efficient, and invertible way. We show the effectiveness of the method using synthetic imagery computed from mixtures of field spectra, a broad compilation of ecologically diverse airborne data, and DESIS satellite imagery. This approach can complement established algorithms by adding a degree of explicability, as well as improving generalizability and classification and regression results in some cases.

1 Introduction

Surface Biology and Geology (SBG) is an essential component of NASA's Earth System Observatory (Margetta, 2021). SBG is a Designated Observable in the 2018 Decadal Survey, with a candidate measurement approach of "hyperspectral imagery in the visible and shortwave infrared (IR), multi- or hyperspectral imagery in the thermal IR" (National Academies of Sciences & Medicine, 2018). Photos collected by traditional digital cameras collect light in three broad wavelength intervals, or bands – visible red, green and blue. Hyperspectral imagery, referred to hereafter as imaging spectroscopy (IS), measures light in hundreds of narrow bands, extending beyond the visible spectrum to also include infrared light.

Vegetation traits are listed in the Decadal Survey (DS) among the main science and applications targets of the SBG Designated Observable. Specific traits listed by the DS include leaf nitrogen concentration, leaf mass area, xanthophyll pigments, and stress indicators (National Academies of Sciences & Medicine, 2018). Decades of research have resulted in considerable progress in estimating such plant traits from imaging spectroscopy data, including both process-based and statistical approaches (Gamon et al., 2019; Singh et al., 2015; Ustin & Gamon, 2010), recently reviewed in (Cavender-Bares et al., 2020). Common tools include narrowband spectral indices (e.g., (Thenkabail et al., 2018)) and regression approaches like Partial Least Squares (Asner & Martin, 2009; Smith et al., 2002; Townsend et al., 2003), often focusing on brightly lit closed canopies at high spatial resolution.

In addition to these retrieval algorithms, researchers have developed biological, chemical, and physical understanding about the mechanisms expressing plant traits in imaging spectroscopy data (Asner, 1998; Kokaly et al., 2009; Ollinger, 2011). Such studies often tie laboratory chemistry measurements to laboratory or airborne reflectance spectroscopy. Laboratory chemical analyses of functional traits have been refined over centuries to quantify a range of important plant traits like bulk nutrient content. Established techniques in reflectance spectroscopy also exist (Clark, 1999; Hapke, 2012; Hunt, 1977), but it is a relatively newer field.

One common laboratory reflectance spectroscopy approach decomposes reflectance spectra into two fundamental signal elements: a spectral continuum and specific absorption features that cause deviations from that continuum (Clark, 1999; Clark & King, 1987; Hapke, 2012; Hunt, 1977). For many materials, the physical processes generating the continuum remain surprisingly elusive (Hapke, 2012). But specific absorption features can often be explained by relatively simple physical chemistry concepts (Nassau, 1987). Distinguishing between the continuum and absorptions allows the analyst to focus on the subset of the signal associated with a specific component of interest.

Airborne and spaceborne IS implements the principles of laboratory reflectance spectroscopy at landscape scale, measuring light which has traveled twice through Earth's atmosphere and spatially integrating on the scale of meters instead of centimeters. This introduces or amplifies numerous additional complexities like canopy structure, topography, and spatial mixing. Despite these complexities, the conceptual model of decomposing reflectance into continuum and absorptions remains useful in many cases, with a long history rooted in the Continuum Removal approach (CR; (Clark & Roush, 1984)). CR is often applied early in the IS processing chain, intended to condition the data for subsequent classification or regression algorithms (Féret & Asner, 2014; Huang et al., 2004; Mutanga et al., 2004). This conditioning is applied (in part) to compensate for limitations in corrections for atmospheric, bidirectional

reflectance distribution function (BRDF), and illumination effects. In applications such as mineralogic analysis of rock outcrops, both convex spectral continua and absorption features are clearly present in IS data. For such cases, CR can be a successful approach for identifying mineral assemblages on the basis of the center wavelength, breadth, asymmetry and curvature of visible and infrared absorption features (Ghrefat et al., 2007; Kokaly et al., 2009; Oshigami et al., 2013).

However, in many other cases, the problem is not so simple. A fundamental difference between laboratory reflectance spectroscopy and IS is the degree of environmental control. In the lab, samples can be prepared to optimize sample properties like purity, illumination, and sample surface roughness (Hapke, 2012). But field application of IS must contend with the full complexity of the Earth surface, which is rarely homogenous, flat, or smooth (Jensen, 2006). Physical realities such as compositional inhomogeneity, topographic complexity, and surface roughness result in complexities like photometric variability, cast shadows, flux density variations, and BRDF effects. These can result in serious complications in analyses of all types of Earth observations (Fisher, 1997), including IS. Such problems – evident even at submeter spatial resolutions achievable by low altitude airborne surveys – are especially important to consider at the decameter expected resolution of current and planned spaceborne IS missions.

In the face of these complexities, challenges emerge for the continuum-and-absorption model. A typical 30 m pixel may contain a myriad of individual plants (live and/or dead; of multiple species; with widely varying traits) and soil types (wet and/or dry; of varying chemical composition and texture), with highly variable fractional coverage of canopy area and subpixel shadow (Sousa & Davis, 2020; Strahler et al., 1986; Woodcock & Strahler, 1987). When aggregated into a single reflectance measurement, fundamental questions arise: What should be considered the appropriate spectral continuum for a mixed pixel? How can this problem be systematically and efficiently approached at the global scale with sub-weekly revisit intervals?

Answers to these questions may be guided by the observation that, in many cases, the signal due to subpixel mixing at the 30 m scale can be accurately captured using simple linear mixture models. Spectral Mixture Analysis (SMA, (Adams et al., 1986; Gillespie, 1990; Smith et al., 1990)) treats the observed reflectance as a linear combination of spectral endmember (EM) materials, plus misfit. In many common cases, a solid physical rationale exists to support the linear mixing assumption at decameter pixel resolution. Linear mixing is precise when single reflection events dominate, such as for surfaces or aggregates with vertical variability that is small relative to the pixel size. Moreover, SMA fractions can often be quite accurate even when these conditions are not met exactly. However, it is limited in that the reflectance spectra of the EMs must be known in advance.

Which, and how many, EMs to use? Insight may come from decades of literature on the global multispectral feature space. Since the 1970s, multispectral image data have been known to be largely representable by three dimensions, referred to in early literature as “brightness”, “greenness”, and “third” (Crist & Cicone, 1984; Kauth & Thomas, 1976). Later studies refined these concepts to include geographically extensive, spectrally diverse image compilations of Landsat imagery (Small, 2004; Small & Milesi, 2013; Sousa & Small, 2017, 2019a), finding the global multispectral feature space to consistently be dominated (> 99% of variance) by three EM materials: 1) soil, rock, and nonphotosynthetic vegetation Substrates; 2) illuminated photosynthetic Vegetation, and 3) Dark targets like shadow and water (S, V, and D). Important exceptions to this observation include evaporites, ice, and shallow marine environments.

Initial investigations using spatially limited but spectrally diverse IS data compilations reveal similar partition of IS image variance into low-order dimensions (Thompson et al., 2017) and domination of these low-order dimensions by S,V, and D EMs (Sousa & Small, 2018). Important differences may exist once thermal wavelengths are added (K. Cawse-Nicholson et al., 2019), but coincident data are needed for study and not yet broadly available. These generic S, V, and D EMs are not representative of the complete global diversity of IS observations of the Earth surface. But they do suggest that for many important landscapes, the gross variance structure of Earth observing IS data – i.e., the component due to phenomena like variable illumination and broad land cover variability – can be understood as a mapping onto a linear mixing subspace bounded by the generic S, V, and D EMs.

Viewed in this way, the rich information content unique to IS data can be conceptualized as potentially nonlinear, high dimensional excursions from a low dimensional simplex bounded by S, V, and D EMs. Under the traditional conventions of SMA, these excursions are often quantified as model misfit considered to be error. We propose to instead treat them as signal.

The analysis below introduces the mixture residual approach. The method is first explained in detail, then applied across ecologically and spectrally diverse field, airborne, and satellite datasets. Our objective is to illustrate example applications and address the following questions:

1) To what extent can wavelength-dependent deviations of linear spectral mixture models (i.e., mixture residual spectra) retain spectrally coherent absorption features known to be associated with ecological and pedological variability?
2) How do mixture residual spectra, correlation structure, and dimensionality compare to those of full reflectance spectra and traditional convex hull-based continuum removal?
3) How well can preconditioning with the mixture residual help differentiate subtle field-observed ecological differences which are not evident in unconditioned decameter IS reflectance spectra?

2 Materials and Methods

In this analysis, we consider spatially mixed reflectance spectra as comprised of two primary components: the *broadband component* which is described by location within the global ternary S, V, D mixing subspace, and the *narrowband component* which exists outside the broadband mixing subspace. This view is analogous to the "continuum-and-absorption" approach taken by CR, but differs importantly in that the broadband component (analogous to the continuum) is explicitly modeled as a linear combination of generic EMs, not fit as a multilinear convex hull. In so doing, the component of the signal due to subpixel mixing is estimated and removed while retaining the important narrowband information which is unique to IS. The approach is also conceptually similar to target detection algorithms like orthogonal subspace projection (OSP), but differs in that the "target" spectra are not known a priori and the "background" spectra are taken to be generic EMs based on the global covariance structure of a diverse spectral compilation.

We deem this new approach the "Mixture Residual" (MR). Conceptually, MR proceeds in three stages (Figure 1): 1) estimation of EM fractions by model inversion, 2) reconstruction of

modeled reflectance spectra, and 3) (observed – modeled) subtraction to compute the mixture residual. The analysis below examines the implementation, strengths, and limitations of this approach using synthetic spectral composites computed from linear combinations of field reflectance spectra, an ecologically diverse compilation of AVIRIS-ng-like images from the Core Terrestrial sites of the National Ecological Observatory Network (NEON), and DESIS satellite IS data. We explain each step in detail below, and then show how the entire procedure can be condensed into a computationally efficient, invertible, linear operation implementable in a single line of code.

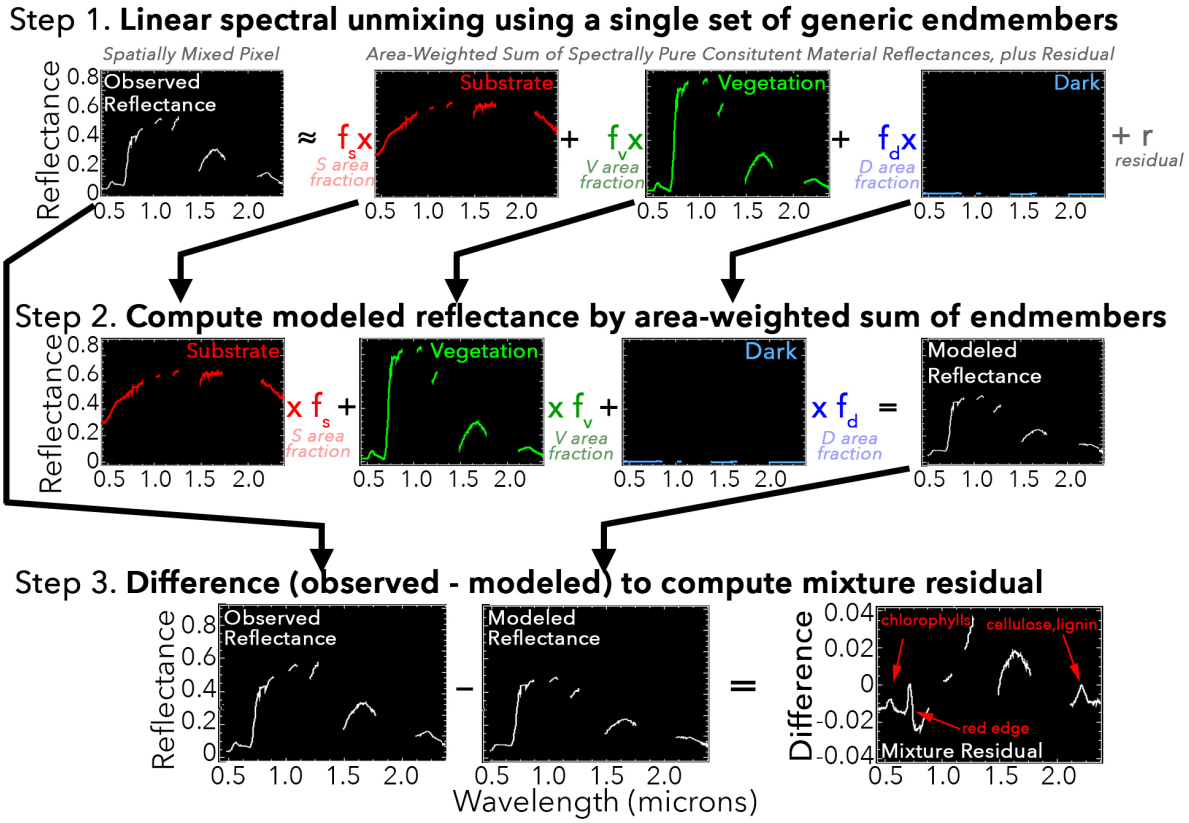

Figure 1. Conceptual overview. Analysis is completed in a three step, automatable procedure. First, observed surface reflectance is decomposed into area fractions (f_s, f_v, and f_d) of Substrate (S), Vegetation (V) and Dark (D) constituent materials – plus residual (r) – using standard linear spectral unmixing with generalized endmembers (top). Modeled spectra for each pixel are then reconstructed by multiplying S, V, D endmember spectra by their local f_s, f_v, and f_d fractions and summing (center). Mixture residual spectra (bottom) are then computed as the difference between observed and modeled reflectance. In theory, this should have the effect of isolating and removing spectrally broad, high amplitude signals from sources like land cover and illumination, and retaining spectrally narrow, lower amplitude signals from sources like foliar and soil chemistry.

2.1 Spectral Mixture Analysis

Step 1 uses the well-established, physically-based approach of SMA (Adams et al., 1986; Smith et al., 1990; Gillespie, 1990). SMA is frequently used in imaging spectroscopy to estimate the subpixel area fraction of spectrally distinct materials. Key benefits of linear SMA are

computational efficiency (requiring only the solution of a constrained least squares problem), clear interpretability of result, and strong physical basis for subpixel linear mixing at the decameter scale in many (but not all) terrestrial environments.

However, SMA does possess a potentially serious limitation – the necessity of determining which, and how many, endmember spectra to use in the analysis (Adams & Gillespie, 2006). Given the linear modeling framework, by definition a spectra with *b* bands unmixed with more than *b* endmembers does not have a unique solution. In practice, this effect is exacerbated and the problem tends towards degeneracy far before *b* endmembers is reached. As such, it has been historically common to utilize a much smaller number of EMs.

One commonly used approach for endmember selection is that of Multiple Endmember Spectral Mixture Analysis (MESMA; (Roberts et al., 1998)). Briefly, MESMA performs unmixing using several possible combinations of small numbers of EMs drawn from a larger library. The combination of EMs which minimizes 2-norm model misfit is then chosen for each pixel. This allows for a large number of potential materials to be mapped in a given image, while avoiding some of the problems associated with formulating the unmixing problem using a large number of EMs at once. The disadvantage is that MESMA requires a combinatorial search through possible EM combinations, a considerable computational cost.

In contrast to MESMA, we use a single set of generic EMs for Step 1. In the case of this study, the S, V and D EMs consist of multipixel means manually selected from the apexes of the low-order PC space of the synthetic spectral composite. In particular, the D EM used here is not true photometric zero, but rather an average of reflectance signatures from targets in deep shadow and/or deep, low-turbidity water. We include EM spectra and a list of NEON AOP tiles in the Supplement. Once coverage becomes truly global, a standardized set of EMs could be developed from a global, or near-global, intercalibrated data compilation to observationally bound the low-order spectral mixing space of the entire terrestrial reflectance field. We emphasize that these endmembers are specifically chosen to provide a generic template with which to construct the mixture residual – they are not intended to be perfectly representative of the scene, or to minimize the local model misfit. We of course also recognize the value of scene-specific EMs, as well as automated EM selection approaches, and suggest comparison between local and global results for targeted studies.

The theoretical justification behind the three-endmember approach is rooted in the observation (Crist & Cicone, 1984; Kauth & Thomas, 1976) that much of the variance of terrestrial multispectral data is contained in three dimensions. Numerous studies using spectrally diverse compilations of Landsat images (Small, 2004; Small & Milesi, 2013; Sousa & Small, 2017), along with Sentinel-2 (Small, 2018) and MODIS (Sousa & Small, 2019a), have quantified the three-dimensional variance of spectrally diverse compilations of multispectral data to be >98% and shown that the global multispectral mixing space of ice-free land surface is bounded by the S, V, and D EMs. While insufficient data currently exists for a comparably extensive study using imaging spectroscopy, the studies that do exist find that the first three dimensions of airborne AVIRIS-classic data dominate the variance structure of observations (Thompson et al., 2017). When spectrally diverse collections of airborne VSWIR reflectance spectra are resampled to Landsat spectral resolution, % variance contained in the first three dimensions rises from 97% to 99%, and the low-order variance structure is dominated by the same S, V and D EMs in both

cases (Sousa & Small, 2018). Again, we emphasize generality for the global case with minimal prior information about the reflectance of the study area, and recognize that in some cases models with more than 3 EMs, or with different EM choices, may be more effective for targeted studies.

2.2 Reconstructing and differencing the modeled spectrum

Considered geometrically, the S, V, and D EMs form a three-point simplex that (ideally) bounds all measured spectra. The SMA performed in Step 1 can be considered a projection of the high-dimensional reflectance observation onto this low-dimensional linear manifold. If the world were comprised only of the materials with reflectance equal to the EM spectra, and perfectly linear mixing were ubiquitous, misfit from this model would be negligible. The data would lie perfectly on this linear subspace, and the hundreds of narrow bands measured by IS would be redundant. This is not the case. Instead, the full complexity of Earth's terrestrial surface results in important deviations from the generic global model. We view these deviations – the offsets of the high dimensional IS measurement from the low dimensional generic mixing space – as a primary subcomponent of signal in IS data.

One way of quantifying this variability is through summary error statistics like the root-mean-square misfit. While useful in determining spatial variations in overall model fit, approaches which summarize model misfit as a single number are not capable of showing the wavelength-dependence of the deviation. The fundamental purpose of this manuscript is to demonstrate that this wavelength dependence can be an important source of signal, and may in fact hold the key to identifying the biophysical sources of variability unique to IS data.

This wavelength dependence is captured by Steps 2 and 3. Conceptually, Step 2 asks the question: what would the reflectance of each pixel be if the generic global model were perfectly accurate? This reconstructed spectrum reflects only the projection of each pixel onto the S, V, and D mixing space. Step 3 then completes the exercise, asking: what signal is present in the IS observations that is not adequately represented by a generic S,V, and D EM model? This quantifies the high-dimensional difference between each pixel's observed position and its projection onto the mixing subspace, isolating the low-variance but coherent information which is typically relegated into higher dimensions of IS data. When plotted against wavelength as a residual "spectrum", this (observed – modeled) difference can contain signal interpretable in the context of known plant traits (e.g., chlorophylls, liquid water, cellulose/lignin).

2.3 Algebraic expression

The above textual description can be summarized in the following matrix equations.

Let the endmember spectra form the column space of a matrix G with dimensions $[\lambda, n]$, where λ is the number of spectral bands in the imaging spectroscopy dataset and n is the number of endmembers. Let also the dataset D be a matrix of spectra with dimensions $[\lambda, x]$, where x is the total number of spectra (or total number of pixels, for image data).

Then, Step 1 (unmixing) can be written in the context of ordinary least squares using the familiar formulation (Settle & Drake, 1993):

$$M = (G^T G)^{-1} G^T D$$

Where M is the matrix of endmember fraction estimates with dimensions [n, x]. In the case of this analysis, we also implement a simple unit sum constraint without changing the functional form by adding a row of 1's to both the G and D matrices, but note that this is optional.

Step 2 (reconstruction) can be formulated as matrix multiplication between the endmember and data matrices:

$$D' = GM$$

Where D' is the matrix of modeled spectra of dimensions [λ, x].

Step 3 (differencing) then takes the simple form:

$$R = D - D'$$

Where R is the mixture residual with dimensions [λ, x'].

If desired, all three steps can be condensed into a single linear operation of the form:

$$R = [1 - G(G^T G)^{-1} G^T] D$$

The ability to formulate the mixture residual as a linear function of the original spectrum has important implications for its utility. First, this supports the view of the mixture residual as isolating a specific subspace of the data which is most informative. Second, this implies that the mixture residual might not improve overall training accuracy for methods like partial least squares regression (PLSR), although generalizability on held-out validation sets might improve by withholding an uninformative subspace. Third, this also implies that the largest impact of implementing the mixture residual may be on Euclidean-distance-based classification and regression analyses. This effect is illustrated with the synthetic example below.

3 Analysis

3.1 Synthetic Example

We begin by illustrating the approach using a synthetic example (Figure 2). We use hemispherical reflectance spectra collected at the Jasper Ridge Biological Reserve using a Beckman UV-5240 spectrophotometer [complete description in (Elvidge, 1990)] to construct 1,000,000 synthetic mixed pixel reflectance spectra comprised of linear combinations of three substrate EMs, three vegetation EMs, and a null (shade) EM. The substrate EMs (soil, sandstone, and gravel) differ primarily in the curvature and amplitude of the spectral continua, while the vegetation EMs (Valley Oak, Live Oak, and Blue Oak) differ primarily in amplitude of the red edge, and depth of photosynthetic pigment and liquid water absorptions. The three oak species captured here are common across California and capture a significant and ecologically-relevant

subset of global Mediterranean-climate oak woodlands and savannas (Davis et al., 2016; Sousa & Davis, 2020).

As found in global studies, the first two Principal Components (PCs) of the synthetic IS image capture broad mixtures of S, V, and D materials. Dimensions 3 and 4 are able to discriminate among the three substrate EMs, but the oak species remain indistinct in the first 4 dimensions. This is typical of characterization of IS datasets with PCs – while the data do capture coherent variability among vegetation EMs, the differences in continuum amplitude and curvature typical of substrates dominate the low-order variance structure and vegetation differences are relegated to higher dimensions.

We then apply the mixture residual, using the global endmembers described above. In contrast, the first 2 PCs of the mixture residual image clearly separate all three substrate EMs and two of the vegetation EMs. The third vegetation EM is unambiguously distinct in dimensions 3 and 4. This occurs because the mixture residual effectively removes the projection of the IS observations onto the ternary mixing subspace which is evident in the PC 1 vs PC 2 projection of the raw surface reflectance data. While the subpixel abundance of generic EMs can be estimated from multispectral data like Landsat, the high-dimensional deviation from the generic SVD mixing subspace cannot. It is the structure of this deviation – the difference between observed reflectance, and the modeled reflectance expected from the generic mixture model – that is uniquely captured by IS data. The amplitude of this difference is low in terms of overall image variance, effectively representing the difference between 97% and 99% of variance found in (Sousa & Small, 2018). But despite its low variance, the information content is sufficiently high to, for instance, discriminate among species of oak trees (at least in this synthetic dataset).

It is also informative to compare the result of this approach to traditional methods of continuum removal. Continuum removal (CR; (Clark & Roush, 1984)) is commonly used to separate the information present in narrowband absorptions from the spectral continuum, generally by estimating the continuum using a convex hull of local spectral maxima. The continuum is then removed in order to accentuate the specific absorptions, either by subtraction or division. Here, we illustrate division. The approach succeeds in separating the substrate EMs, but not the vegetation EMs (at least, in the first four dimensions). It is also informative to show how CR by division can nonlinearize a linear problem, resulting in an increase in the number of dimensions of the space required to accurately describe the low-order data manifold.

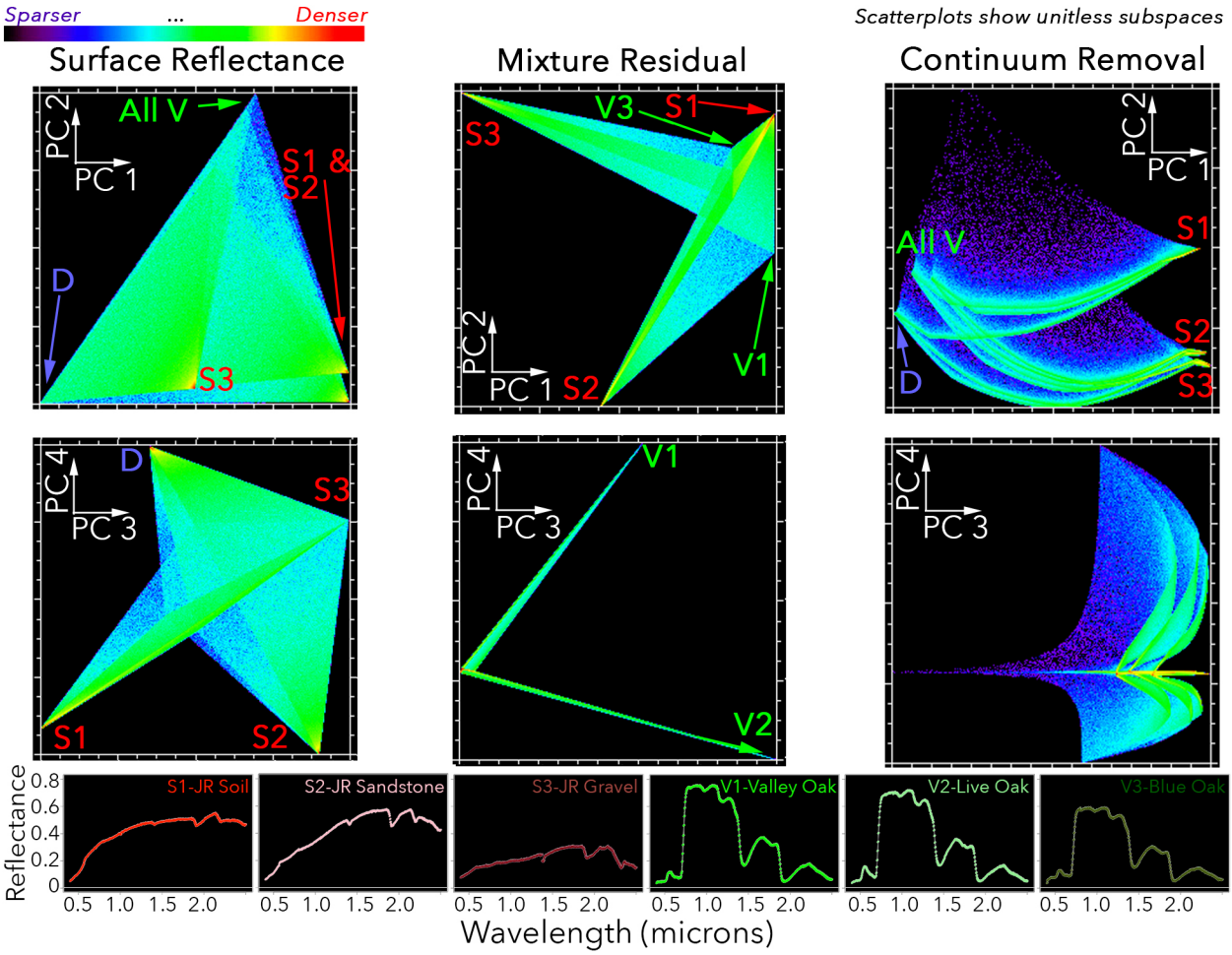

Figure 2. Mixture residual improves discrimination of pixel composition in a synthetic image. We generate 1 million spectra using weighted sums of 1 random substrate EM, 1 random vegetation EM, a single dark EM, and 5% noise. EM weights are random and forced to sum to unity, representing the idealized theoretical case of perfect linear mixing. The S EMs represent a range of spectral slopes, brightnesses, and narrow absorption features. The V EMs represent species-level differences within a single genus (Quercus) and show relatively modest spectral differences, largely in the signatures of photosynthetic pigments, red edge, and leaf water. As repeatedly observed in satellite and airborne image compilations, the low-order variance structure of the full synthetic image (left) is dominated by S, V, and D subpixel mixing. Two S EMs group together in one corner of the PC 1 vs 2 mixing space and all 3 V EMs group in another corner. S EMs are clearly distinct in PC 1 vs 3, but V EMs remain clustered until dimension 5 and higher. When k-means unsupervised clustering is applied using 3 classes, class purity is low even for the soil endmembers (0.67, 0.61, 0.58). In contrast, mixture residual (center) distinguishes all 6 EMs in the first three dimensions, successfully dampening the high variance signal from subpixel mixing. This allows the subtle signals due to plant trait and soil chemistry differences to be better expressed by Euclidean distance, as evidenced by markedly higher class purity using the same k-means approach (1.0, 0.99, 0.67). Clear species-level

separation is evident event in high correlated (r > 0.98) V spectra. In contrast, continuum removal by division (right) nonlinearizes an otherwise linear problem because absorption troughs scale more slowly than absorption shoulders. In the continuum removal approach, three distinct nonlinear manifolds radiate outward from each S EM. The introduction of this nonlinearity results in an artificially high dimensionality of the linear subspace occupied by the CR product, and yields intermediate k-means cluster purity (0.71, 0.70, 0.69). Scatterplots show unitless subspaces, so quantitative axis labels are unncessary.

3.2 Real-world complexity

We next apply this approach to a spectrally diverse, ecologically-relevant dataset: a compilation of IS data collected by the National Ecological Observatory Network (NEON) Airborne Observational Platform (AOP). NEON AOP operates an AVIRIS-NG-like instrument, capturing data across the full VSWIR range at approximately 5 nm wavelength intervals. AOP is typically flown at low altitude, allowing for measurements with a 1 m ground sampling distance. We use data compiled from each Core Terrestrial site, sampling each of the 20 ecoclimatic domains determined by the NEON program across the contiguous US, Alaska, Hawaii, and Puerto Rico (locations shown in Figure 3; NEON 2021a). These domains were determined by a multivariate graphical cluster analysis of nine climatic variables (Hargrove & Hoffman, 1999). Sites differ greatly in a wide range of factors including plant traits, community composition, soil chemistry, providing an ideal range of vegetation types with which to examine the mixture residual method. Lists of specific data tiles are given in the Supplement.

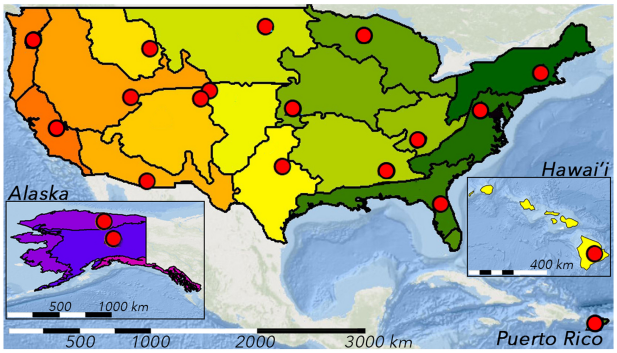

Figure 3. Locations of the 20 NEON Core terrestrial field sites used for the subsequent analysis (red circles), each representing a different ecoclimatic domain.

Spectrally diverse 1 km^2 tiles of orthorectified surface directional reflectance (NEON data product DP3.30006.001) were downloaded from the NEON data portal using the 'neonUtilities' R package. Data were used as-is from the NEON portal, without additional BRDF or atmospheric correction reprocessing. Data collected in the wavelength intervals of 383-428, 1280-1480, 1780-2106, and 2386-2500 nm (bands 1-10, 180-220, 280-345, and 401-426) were excluded due to contamination from deep water vapor absorptions and/or sensor noise.

Figure 4 shows the data mosaic as a false color composite with red, green, and blue channels of the image displaying SWIR, NIR, and visible reflectance (upper left), as well as with subpixel abundance of generic S, V and D EMs derived from the AOP composite (upper right). Clearly, the EM fraction composite image captures the broad land cover patterns evident from the false color composite image. The mixture residual image (lower right) shows richer spatial

variability in both the photosynthetic vegetation and NPV/soils. The first three dimensions of continuum removal by division (lower left) are also shown for comparison.

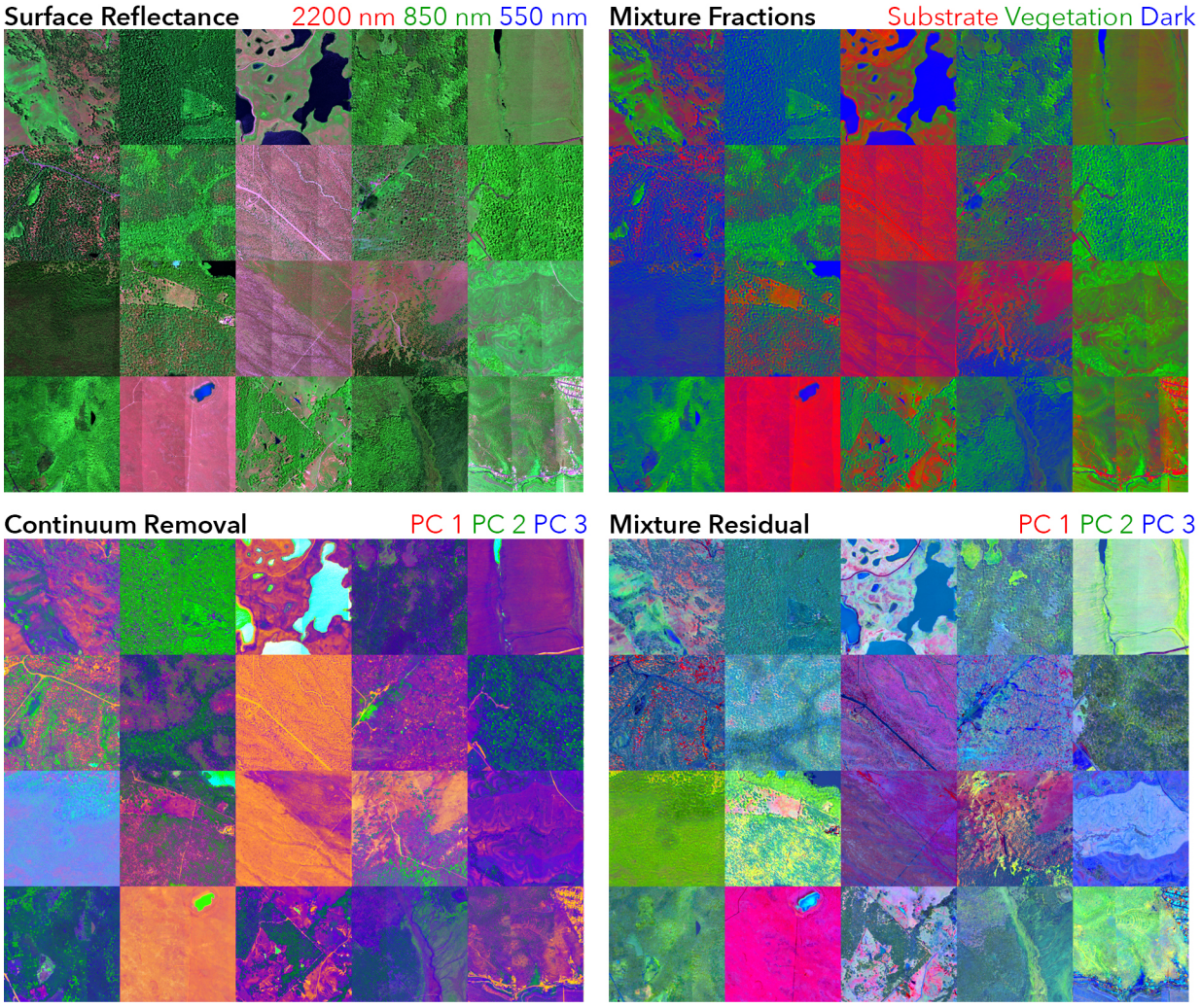

Figure 4. 20 Core Terrestrial field sites from the National Ecological Observatory Network (NEON) as observed at 1 m spatial resolution by the Airborne Observational Platform (AOP)'s spectrometer. Surface reflectance image (Upper Left) shows broadband land cover patterns which are captured by linear SMA using generic EMs (Upper Right). Informative spatial patterns are present in the low-order PCs of the residual (Lower Right) which are not visible in any of the other images. Continuum Removal results (Lower Left) are shown for comparison.

Figure 5 illustrates quantitatively the effect of the approach using the partition of variance computed from the eigenvalues. In addition to the results from the full NEON AOP dataset (black), an additional comparison is shown with the data spectrally resampled by convolving with the spectral response functions of the Landsat 8 instrument (red). It is unclear from examination of the partition of variance computed from the raw surface reflectance (left) that significantly more information is present in the full spectral resolution dataset.

The opposite is true in the partition of variance computed from the mixture residual (right). In this case, over 90% of the variance remaining in the Landsat mixture residual is present in the

first dimension, while only 65% of variance in the AOP mixture residual is present in the first dimension. The structure of the residual spectrum thus requires a high dimensional space to represent, while 99% of the variance in the difference between Landsat observations and reflectance predicted by S, V, and D mixture fractions can be expressed in only 3 dimensions.

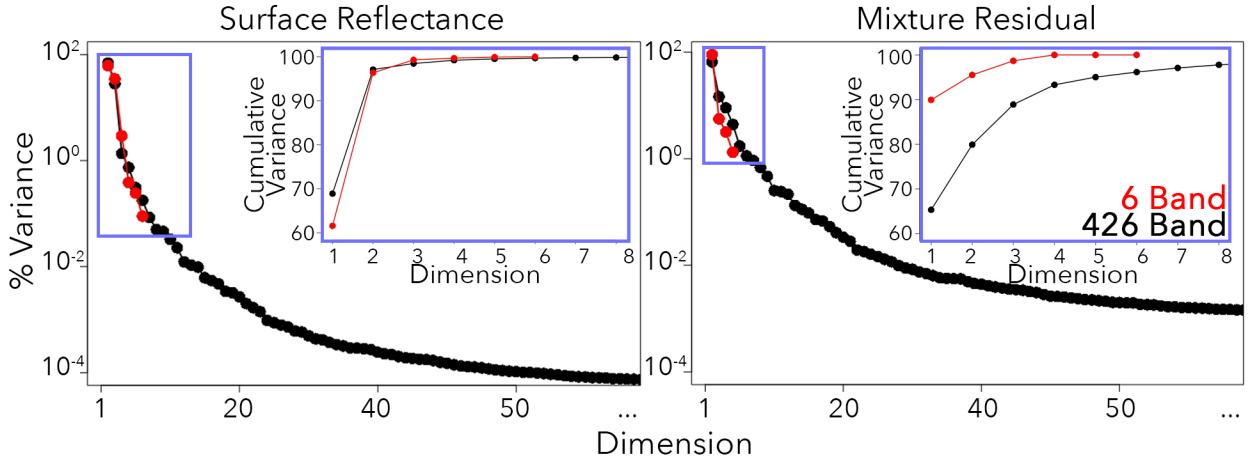

Figure 5. Partition of Variance. Normalized eigenvalues for surface reflectance (left) and mixture residual (right). Full 5 nm spectral resolution data (black) are compared to simulated Landsat data (red) produced by convolving imaging spectrometer data with Landsat 8 spectral response functions. Cumulative variance shown in inset. Variance partitioning in narrowband and broadband is remarkably similar when full reflectance spectra are used. Differences are much more prominent when the residual of a 3 EM mixture model is examined. >90% of the variance in broadband mixture residual can be represented by 1 dimension, and >99% can be represented with 3 dimensions. In contrast, expression of 90% of narrowband mixture residual variance requires 4 dimensions, and 99% requires 13 dimensions.

The plant trait-specific information contained in the mixture residual can be understood in the context of covariability matrices. Figure 6 shows covariance (upper triangular) and correlation (lower triangular) matrices for the AOP surface reflectance (left) and mixture residual (right). The structure of the surface reflectance covariability matrices effectively reflects the information present in a generic vegetation spectrum: band-to-band correlations are high within the visible [mean (m)= 0.97 and standard deviation (s) = 0.027], NIR (m = 0.94 and s = 0.057), and SWIR (m = 0.97 and s = 0.031) portions of the spectrum, indicating redundancy of information. Covariance is dominated by the high-amplitude red edge signal in the NIR, a highly generic signal indicative of the presence of green vegetation, but not specific traits. Liquid water absorptions are 950 and 1140 nm are evident, but minor.

In contrast, plant trait signatures are considerably more evident in the covariability structure of the mixture residual. Narrow NIR and SWIR absorption features associated with leaf water and dry matter are clearly distinct in both correlation and covariance. Mixture residual band-to-band correlations in the visible spectral domain remain high, albeit statistically distinct from surface reflectance (m = 0.95 and s = 0.040). Greater asymmetry is seen at visible wavelengths, consistent with xanthophyll- or anthocyanin-type processes. The effect of mixture residual decorrelation is more pronounced in the NIR (m = 0.04 and s = 0.61) and SWIR (m = 0.31 and s = 0.58) spectral domains, significantly exceeding decorrelation achieved by continuum removal by division (visible m = 0.95 and s = 0.045; NIR m = 0.90 and s = 0.088; SWIR m = 0.85 and s =

0.14). Differences at NIR and SWIR wavelengths in both narrow absorptions and broader spectral slopes clearly resolve leaf water and cellulose/lignin signatures, with intriguing additional structure potentially suggestive of additional leaf chemistry. These patterns provide another illustration of how the narrowband signal unique to IS data can be more readily accessed once the variance due to spatial land cover mixtures is removed.

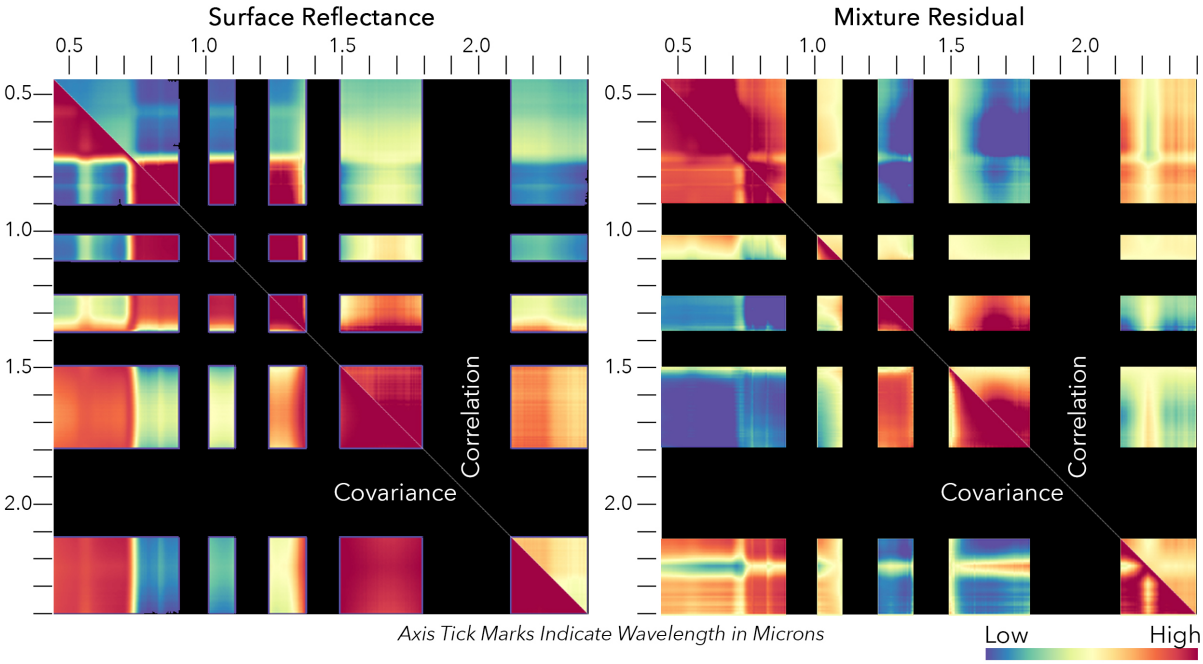

Figure 6. Covariability matrices for the surface reflectance (left) and mixture residual (observed - modeled; right) datasets. Covariability structure of surface reflectance data is dominated by the absorption features of photosynthetic vegetation. Covariability of the mixture residual reveals greater complexity, accentuating narrowband absorption features. Wavelengths 383-428, 884-1000, 1094-1224, 1355-1480, 1780-2106, and 2386-2500 nm (bands 1-10, 101-124, 143-169, 195-220, 280-345, and 401-426) were excluded due to potential contamination from atmospheric absorptions and/or sensor noise. MR asymmetry at 430-680 nm is suggestive of anthocyanins and/or xanthophylls; decorrelation near 2200 nm is associated with celluluose/lignins; and complex structure in the 950 – 1780 nm range both clearly resolves leaf water absorptions and indicates additional potential information about other leaf chemistry constituents.

The context explained above informs examination of the spectral feature space (Figure 7). Low-order dimensions of the surface reflectance (left), mixture residual (center) and continuum removal (right) feature spaces illustrate the effect of the approach. Low order dimensions of surface reflectance are dominated by the signal from S, V, D land cover mixing. The broadband nature of this information is reflected in the similarity to the low-order feature space of the Landsat resolution data (bottom row).

In contrast, differences are far more evident between the low-order feature space of the mixture residual at full vs reduced spectral resolution. The feature spaces are clearly characterized by different structures, with considerable spectral separability and distinct apexes in the full resolution data which are not clearly present at reduced resolution. These differences are also not

clearly present using the traditional Continuum Removal approach, which is dominated at low order by relatively sparse but complex signals from shallow water targets.

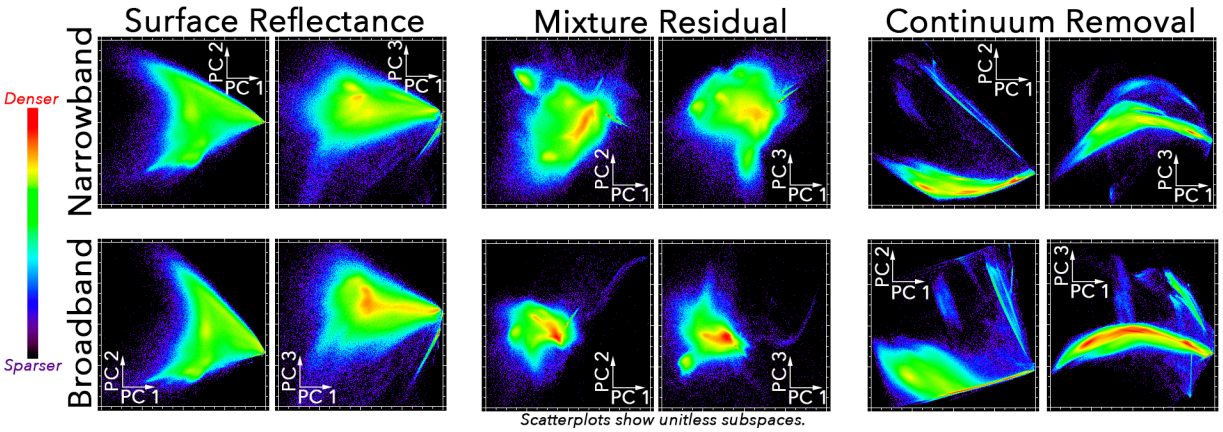

Figure 7. Spectral feature spaces. Low-order spectral feature space topology is remarkably similar for narrowband and broadband surface reflectance spectra (left). This similarity does not hold for the spectral feature space of narrowband and broadband mixture residuals (center). Continuum removal is also shown for both narrowband (upper right) and broadband (lower right) comparison. Scatterplots show unitless subspaces, so quantitative axis labels are unncessary.

The differences captured in the low-order mixture residual space show clear plant trait signatures (Figure 8). Example spectra from 9 positions, indicated by squares of different colors, are chosen from the PC 3 vs 2 projection of the mixture residual feature space. Spectra are plotted from each position. The full reflectance observations of each point are plotted (left), as well as the mixture residual (center) and continuum removal (right) spectra. Enlargements of visible, NIR, and SWIR2 portions of the spectrum are shown below for ease of viewing.

Many of spectra found at the periphery of the mixture residual space have the broad appearance of green vegetation spectra. However, as expected from the correlation structure, subtle differences at wavelengths corresponding to known absorptions of biophysical markers like photosynthetic pigments, leaf water, and cellulose/lignin are evident. A wide range of signatures – including slope, curvature, and amplitude signals – are evident in the mixture residual spectra. Comparison the mixture residual PC 3 vs PC 2 feature space projection against surface reflectance and continuum removal spaces (top row), again shows the mixture residual to accentuate differences in vegetation spectra which are superimposed near the bottom apex of the surface reflectance feature space. The continuum removal feature space contains interesting information, but is distorted due to the nonlinearity introduced by division and dominated by high amplitude variability along the PC 2 axis.

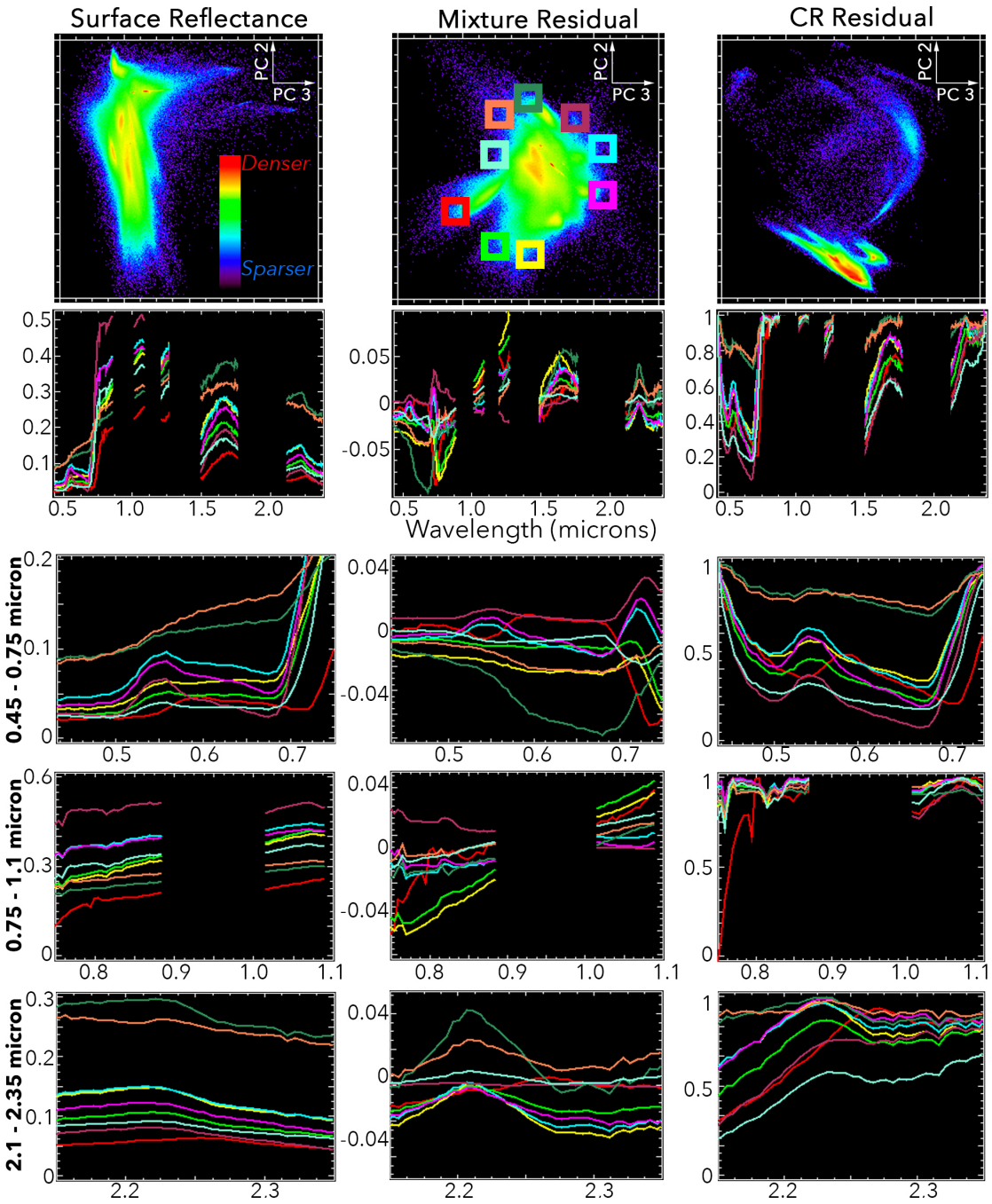

Figure 8. Example spectral endmembers from the mixture residual feature space. Low dimensions of SR feature space (Upper Left) show characteristic broadband land cover variations, and low dimensions of CR residual feature space (Upper Right) are dominated by intimate mixing in lakes. Top row of spectra show full VSWIR continua (with atmospheric absorptions removed) as SR (Left), mixture residual (Center) and CR residual (Right). Bottom rows show enlargements for three spectral regions. Differences in curvature at visible wavelengths, location and amplitude of red edge, spectral slope in the VNIR, and curvature in the SWIR 2 are all more evident from MR residuals than SR. CR residuals amplify curvature differences, but do not capture important deviations in slope of the spectral continuum and are forced to unity at bounding wavelengths.

3.3 Application to Spaceborne Imaging Spectroscopy

Next, we extend these results to a spaceborne IS dataset. The data used here were collected by DESIS, an imaging spectrometer built and maintained by the German Aerospace Center (DLR) with 235 spectral bands in the 400 – 1000 nm spectral range (Krutz et al., 2019). DESIS is currently hosted on the International Space Station, with data delivered at 30 m ground sampling distance. Here, we analyze a summer image of one of the NEON sites used in the previous compilation, the University of Notre Dame Envrionmental Research Center field site (NEON site code "UNDE"). These data pose a challenging test of the viability of the mixture residual approach to yield useful information in the face of full two-way atmospheric transmission, decameter spatial resolution, and variations in orbit and sun-sensor geometry which are at least as formidable as those expected for a future sensor onboard a free flying sun-sychronous platform. Notably, the spectral range of the DESIS sensor is also a subset of the full 400 – 2500 nm spectral range planned for some future misisons.

Figure 9 illustrates the ability of the mixture residual approach to extract subtle yet useful ecological information for such real-world spaceborne IS data. In comparison to a traditional false color reflectance image (far left), a false color composite showing three mixture residual bands (center left) reveals considerably greater nuance, capturing spatially coherent patterns which track field-observed ecological observations. One example of such information is illustrated by the case of three NEON plant plots (UNDE # 012, 017, and 030; NEON 2021b), indicated as yellow dots on the figure. These three plots are categorized into three distinct land cover classes (woody wetland, mixed forest, and deciduous forest), with substantial variability in field-observed plant species lists. Investigation of surface reflectance spectra from individual pixels from the coordinates corresponding to each plot center (bottom left) does not reveal obvious high amplitude differences in spectral shape, a feature which can be quantified by the observation that all three pairs of spectra are correlated at the 0.98 level or higher. In contrast, correlations for each pair of mixture residual spectra for each plot (bottom right) are considerably lower (0.68, -0.32, and -0.07). By removing the contribution of the spatial mixing signal, the low-variance – but highly informative – spectral indicators of plant traits are effecetively accentuated.

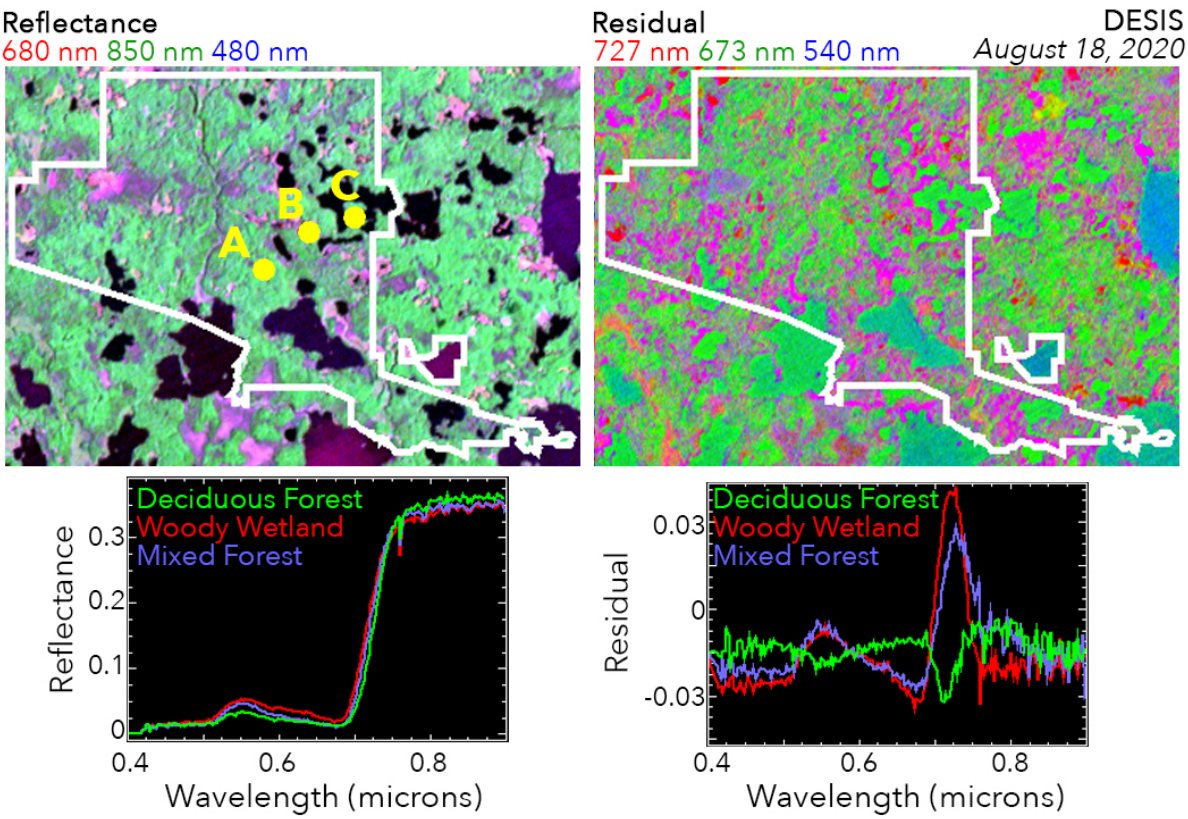

Figure 9. The mixture residual approach isolates ecologically meaningful differences in 30 m spaceborne imaging spectroscopy data. A traditional visible + NIR false color composite DESIS image of the University of Notre Dame Environmental Research Center (NEON site UNDE) is shown on the left. Three mixture residual wavelengths are shown on the right. The mixture residual clearly resolves patterns within and among vegetation communities. Reflectance and mixture residual spectra are shown for pixels centered on NEON plots of three different NLCD Land Cover classes, centered on the yellow dots in the DESIS image: UNDE 030 (A; Deciduous Forest; Green); UNDE 012 (B; Woody Wetland; Red); and UNDE 017 (C; Mixed Forest; Blue). 10 cm true color airphotos (not shown) collected by NEON AOP confirm clear biophysical differences among these sites, but 30 m spatial resolution DESIS reflectance spectra are highly correlated (all pairs 0.98 or higher). Correlations among pairs of mixture residual spectra are lower (0.68, -0.32, and -0.07), quantifying improved separability. Differences in chlorophyll bump and red edge amplitdue and position are subtle in reflectance spectra but unambiguous in mixture residual. Lakes, common in the UNDE study area landscape, appear black or dark magenta on the reflectance image.

Additionally, we expect the mixture residual to impact different retrieval approaches in different ways. There is good theoretical justification to expect a significant difference in result for in approaches that only use a subset of the spectral information, e.g. narrowband spectral indexes (Thenkabail et al., 2018)). An example using the mixture residual as a pretreatment for spectral indices is shown in Figure 10.

Here, we use the simple ratio of reflectance at 708 and 775 nm ($RI_{708,775}$), found by (Féret et al., 2011) to show a strong quadratic relationship to total chlorophyll content. Using the synthetic dataset from Figure 2, we illustrate the nonlinear dependence of $RI_{708,775}$ on subpixel fractional vegetation cover (left). Two of the three oak species used for this analysis (Blue and Live Oak) are clearly ambiguous at all but the highest fractional covers, and the third (Valley Oak) is ambiguous at values below 70% cover – including shadow. The quadratic transfer function introduced by (Féret et al., 2011) is then applied to yield estimates of total chlorophyll (C_{ab}) for each simulated pixel. In theory, one might expect total pixel-level chlorophyll to vary linearly with subpixel vegetation abundance from 0 (in unvegetated pixels) to a maximum value determined by the species-specific spectrum. This theoretical behavior would follow the lines on the center plot. Instead, some concavity remains and considerable scatter is observed, especially towards high C_{ab} values, and especially in the 10-60% fractional cover range.

In contrast, the right panel shows $RI_{708,775}$ computed on the mixture residual spectrum instead of the reflectance spectrum. Clearly, the $RI_{708,775}$ applied to the residual is characterized by a more linear relationship to vegetation fraction, reduced scatter, and significantly enhanced species-level separability. No overlap is observed within these three species at subpixel areas above 50%, and the bulk of the distributions are resolvable at lower values still. We do not apply the transfer function developed by (Féret et al., 2011) here because the model would need to be retrained on the residual spectrum, not the full reflectance spectrum. Further, the behavior of at least this spectral index suggests that a simpler linear model may be sufficient for this particular plant trait.

As an aside, we note that a simple bias shift was applied to the residual before computing the index in order to ensure positivity. This is because residuals are centered on zero, and the functional form of the spectral index becomes unstable with sign changes and division by zero. This bias shift is easily implemented by adding a constant slightly larger than the minimum observed residual value at the wavelengths of interest.

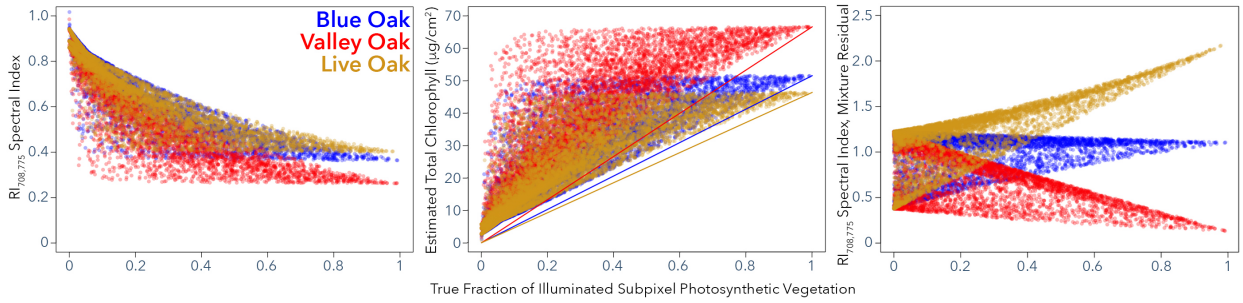

Figure 10. The mixture residual can be an effective pretreatment for narrowband spectral indices. Here we illustrate for the case of estimating total chlorophyll. We compute the simple ratio of reflectance at 708 and 775 nm (left) and the quadratic transfer function suggested by (Féret et al., 2011) (center), comparing to the "true" input subpixel area of illuminated photosynthetic vegetation. In both cases, sufficient dispersion is observed to result in Blue vs Live Oak to be indistinct below 85% subpixel area, with Valley Oak distinct down to levels nearer 70%. When the same index is applied to the mixture residual (right), species-level distinctions are much clearer, with no observable ambiguity at 50% fractional cover or higher. We do not apply the quadratic transfer function because it would need to be retrained. Because residuals are centered around

zero, a simple bias shift equal to the minimum residual value is first applied to avoid issues associated with positivity and division of values near zero.

5 Discussion

Imaging spectroscopy offers an exceptional tool to observe Earth surface biology and geology (National Academies of Sciences & Medicine, 2018). But in order to unlock the full potential that this tool offers, analysis approaches must contend with the high dimensionality inherent to image data with hundreds of channels. These challenges are especially relevant for retrieving plant traits because important indicators of plant composition and function like photosynthetic pigments, leaf water, nutrients, and dry matter are often minor contributors to the observed reflectance signal in terms of overall variance (Asner, 1998). For the decameter-scale pixels which might be captured by a future satellite-based sensor, spatial mixing processes further accentuate this challenge (Strahler et al., 1986; Woodcock & Strahler, 1987).

The analysis presented here uses the convex geometry tools of (Boardman, 1993) to provide a mechanism for exploring the biophysically relevant variability within a scene, synthesizing the continuum-plus-absorption model of (R. N. Clark & King, 1987) with the spectral mixture approach of (Adams et al., 1986; Gillespie, 1990; M. O. Smith et al., 1990). The generality of a global mixing subspace is supported by both decades of global multispectral observation (Crist & Cicone, 1984; Kauth & Thomas, 1976; Small, 2004; Small & Milesi, 2013) and more local but diverse imaging spectroscopy observations (K. Cawse-Nicholson et al., 2019; Sousa & Small, 2018; Thompson et al., 2017). The key insight is in recognizing that spectral variance is an imperfect metric for biological and geologic information content. While 97% of variance in diverse decameter reflectance compilations may reside in the continuum-dominated S, V, and D land cover mixtures, the remaining 3% of variance can contain a nontrivial subset of the overall useful information. Utilizing spectral residuals from generic priors provides a conceptually simple and computationaly tractable means to circumvent this characteristic and still make use of highly informative principal components.

The challenges of imaging spectroscopy are sufficiently complex, and Earth's surface biology and geology are sufficiently multifaceted, to warrant multiple different analysis approaches for different application scenarios. The mixture residual approach is not intended to replace longstanding imaging spectroscopy algorithms. Instead, we suggest that the approach presented here might be used in conjunction with established approaches as one link in a processing chain to help interpret results, and, in some cases, improve accuracy.

Other retrieval algorithms that operate on surface reflectance using Euclidean distance metrics are likely to be similarly impacted. Examples of such approaches include nearest neighbors (e.g. (Larsolle & Hamid Muhammed, 2007; Voss & Sugumaran, 2008)), random forests (e.g. (Asner et al., 2016; M. L. Clark & Roberts, 2012; Huesca et al., 2016; Naidoo et al., 2012)), spectral angle mapper (e.g. (Yen-Ben Cheng et al., 2007), maximum likelihood classification (e.g. (Boschetti et al., 2007; F. M. Lacar et al., 2001)), and some neural networks (e.g. (Plaza et al., 2009; Zhang & Xie, 2012)). The level of impact is likely to vary with study area, physical process being retrieved, and algorithm being used. It is reasonable to expect that the difference may be significant in many cases. As noted earlier, S, V, and D mixture models have been shown to generally represent >95% of diverse Landsat (Small, 2004) and MODIS (Sousa & Small, 2019a)

spectra with <5% RMS misfit. One implication of such low mixture model misfit is that a comparably high fraction of observed spectral variance can be explained by linear S, V, and D mixtures. Such a high variance process will inevitably impact the outcome of any variance-based algorithm.

In contrast, because the mixture residual is a linear function of reflectance, it is unlikely that its application will significantly impact the training accuracy of high dimensional retrieval approaches like partial least squares regression (e.g., (Hacker et al., 2020; Townsend et al., 2003)). However, using a generalized, physically-based set of loadings for the low-order dimensions of PLSR still offers significant benefits. First, the physical meaning of the fraction estimates offers an important level of interpretability to an otherwise purely statistical algorithm. This could be considered a modest step toward explainable machine learning for imaging spectroscopy. Second, the generalizability of the mixture model may serve to stabilize the structure of the PLSR, potentially increasing performance on held-out validation sets, especially for models with limited training data. Third, empirical examination of the structure of the mixture residual and experimental tests against laboratory chemistry measurements and known absorption features could be used to determine which, and how many, additional plant traits could be reliably retrieved beyond those currently used.

While the analysis presented here focuses on applications for surface biology retrievals, the fundamental approach is highly generalizable. For instance, agricultural monitoring is an important use case of imaging spectroscopy (Thenkabail et al., 2000), and the mixture residual could add considerable value in separating crop and soil reflectance. Applications also exist for the surface geology observable, including both soils (Ben-Dor et al., 2009) and rock outcrops (Calvin & Pace, 2016; Kruse et al., 1993), where the mixture residual could be used to help mitigate the signal from vegetation cover and accentuate narrowband absorption features from surficial geology. Imaging spectroscopy of the cryosphere is also an important area of study (Dozier et al., 2009; Nolin & Dozier, 1993), and additional applications for the mixture residual may exist for remote sensing of the cryosphere by extending the cryospheric spectral mixing space concept proposed by (Small & Das, 2018).

Further, we note that modifications may be useful for specific applications. The first obvious modification to consider is the number and spectra of endmembers used for the global model. The endmembers used in this analysis are derived from the feature space of the data compiled here because, to our knowledge, data have not yet been available for a comprehensive study with global, or near-global, spatial extent. This is likely to change in coming years.

In continuum removal, the continuum is generally spectrum-specific: i.e., the continuum is computed from each individual spectrum itself and may vary in unknown ways from spectrum to spectrum. In contrast, MR removes a known S, V, and D mixture having fractions that can be retained and incorporated into interpretation of the MR in subsequent analysis. In other words, the S, V, and D fractions contain valuable information whereas the continuum removed is CR is generally not used in subsequent analysis. These properties of generality and interpretability are a fundamental aspect of the MR approach introduced here.

Once the time series IS data are available, another straightforward extension of this approach will be for spatiotemporal analysis (Small, 2012). Conceptual parallels exist between the

hundreds of channels in a given IS acquisition, and the hundreds of time steps available for image time series. Some studies exist extending fundamental tools like convex geometry, spatial mixing, and continuum removal to the time domain, e.g. (Chi et al., 2016; Li & Wu, 2014; Piwowar et al., 1998; Sousa & Davis, 2020; Sousa & Small, 2019b; Yang et al., 2012), but further research on topic is needed.

It will also likely be informative to compare the results of the mixture residual with local versus global sets of endmembers. This is most likely to be impactful in the S EM, where previous multispectral studies have shown the global feature space to be bounded by exceptionally bright, spectrally convex sands (Small, 2004; Small & Milesi, 2013; Sousa & Small, 2017). Such materials may not be representative of the substrates present in most locations on earth. Using both global and local endmembers would allow studies to made easily intercomparable while also facilitating greater accuracy at specific study sites. Using well-characterized EMs to compute the residual is also important. Any absorptions included in the EM will be encoded in the endmember fraction estimates and thus erased in the orthogonal projection, so guarding against endmember variability in traits of interest will be important.

No method is without its limitations, and the mixture residual approach is no different. The fundamental link between information and variance remains implicit in the methodology. But as shown here, this link is tenuous in the case of imaging spectroscopy. Few would agree, for instance, that 97% of the potentially useful information in a reflectance dataset is contained in the S, V, and D mixtures that drive 97% of its variance. It is clear that the global variance structure of VSWIR surface reflectance is an imperfect proxy for the information structure of the data. Further research is needed to determine key characteristics of the data like the optimal information metric(s) and scales of variance at which information is encoded. Ultimately, the primary value of the mixture residual approach is likely the capacity to improve the ease and speed of data interpretation and exploratory analyses – an important advance in the face of the rapidly increasing volumes of IS data available.

5 Conclusions

This analysis explores a novel approach for improving plant trait retrievals by extracting low-variance information of biophysical significance from imaging spectroscopy. The approach is guided by a fusion of concepts from convex geometry, spectral mixture modeling, and continuum removal. Three fundamental dimensions are observed to dominate the low-order structure of diverse imaging spectroscopy compilations: soil, rock, and non-photosynthetic vegetation Substrates (S), illuminated photosynthetic Vegetation (V), and Dark materials like water and shadow (D). Subpixel mixtures of S, V, and D materials often drive overall spectral variance in imaging spectroscopy, often overprinting signatures of important plant traits like photosynthetic pigments, leaf composition, and dry matter. Our simple approach to accessing these narrowband, low-variance signals uses a three-step procedure: 1) linear spectral mixture analysis using a single set of generic endmembers; 2) pixelwise reconstruction of modeled reflectance; and 3) subtracting (observed – modeled) reflectance spectra to obtain a residual spectrum, deemed the Mixture Residual (MR). We illustrate the approach using a compilation of airborne imaging spectroscopy data from the National Ecological Observatory Network (NEON) Airborne Observation Platform (AOP). The spectral feature space of the MR image is observed to capture information consistent with known cholorophyll, xanthophyll/anthocyanin, leaf water, and cellulose/lignin plant traits,

rather than overall subpixel areal abundances of land cover types. Mixture residuals offer conceptual simplicity, computational efficiency, and global applicability and can be used as inputs to enhance commonly applied variance-based approaches for retrievals of key surface biology and geology observables.

Acknowledgments, Samples, and Data

The authors perceive no conflict of interest. The data supporting this manuscript can be downloaded free of charge from the web portals listed in the main text. Research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. California Institute of Technology. Government sponsorship acknowledged. Support was provided by NASA programs: SUSMAP, IDS, and ECOSTRESS.

Author Contribution Statement

Study conception: DS. Conceptual Refinement: All authors. Analysis and Interpretation: DS. Writing – Original Draft: DS. Writing – Review and Editing: All authors. All authors reviewed the results and approved the final version of the manuscript. After DS, all authors are listed alphabetically.

References

Adams, J. B., & Gillespie, A. R. (2006). *Remote sensing of landscapes with spectral images: A physical modeling approach*. Cambridge University Press.

Adams, J. B., Smith, M. O., & Johnson, P. E. (1986). Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 Site. *Journal of Geophysical Research: Solid Earth*, *91*(B8), 8098–8112. https://doi.org/10.1029/JB091iB08p08098

Asner, G. P. (1998). Biophysical and Biochemical Sources of Variability in Canopy Reflectance. *Remote Sensing of Environment*, *64*(3), 234–253. https://doi.org/10.1016/S0034-4257(98)00014-5

Asner, G. P., & Martin, R. E. (2009). Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests. *Frontiers in Ecology and the Environment*, *7*(5), 269–276. https://doi.org/10.1890/070152

Asner, G. P., Brodrick, P. G., Anderson, C. B., Vaughn, N., Knapp, D. E., & Martin, R. E. (2016). Progressive forest canopy water loss during the 2012–2015 California drought. *Proceedings of the National Academy of Sciences*, *113*(2), E249–E255. https://doi.org/10.1073/pnas.1523397113

Ben-Dor, E., Chabrillat, S., Demattê, J. A. M., Taylor, G. R., Hill, J., Whiting, M. L., & Sommer, S. (2009). Using Imaging Spectroscopy to study soil properties. *Imaging Spectroscopy Special Issue*, *113*, S38–S55. https://doi.org/10.1016/j.rse.2008.09.019

Boardman, J. W. (1993). Automating spectral unmixing of AVIRIS data using convex geometry concepts (Vol. 1, pp. 11–14). Presented at the Proc. Summaries 4th Annu. JPL Airborne Geosci. Workshop.

Boschetti, M., Boschetti, L., Oliveri, S., Casati, L., & Canova, I. (2007). Tree species mapping with Airborne hyper-spectral MIVIS data: the Ticino Park study case. *International Journal of Remote Sensing*, *28*(6), 1251–1261. https://doi.org/10.1080/01431160600928542

Calvin, W. M., & Pace, E. L. (2016). Utilizing HyspIRI Prototype Data for Geological Exploration Applications: A Southern California Case Study. *Geosciences*, *6*(1). https://doi.org/10.3390/geosciences6010011

Cavender-Bares, J., Gamon, J. A., & Townsend, P. A. (Eds.). (2020). *Remote Sensing of Plant Biodiversity*. Springer Nature. https://doi.org/10.1007/978-3-030-33157-3

Chi, J., Kim, H.-C., & Kang, S.-H. (2016). Machine learning-based temporal mixture analysis of hypertemporal Antarctic sea ice data. *Remote Sensing Letters*, *7*(2), 190–199. https://doi.org/10.1080/2150704X.2015.1121300

Clark, M. L., & Roberts, D. A. (2012). Species-Level Differences in Hyperspectral Metrics among Tropical Rainforest Trees as Determined by a Tree-Based Classifier. *Remote Sensing*, *4*(6). https://doi.org/10.3390/rs4061820

Clark, R. N. (1999). Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy. In *Manual of Remote Sensing, Chapter 1* (pp. 1–77). U.S. Geological Survey, http://speclab.cr.usgs.gov. Retrieved from http://speclab.cr.usgs.gov

Clark, R. N., & King, T. V. (1987). Automatic continuum analysis of reflectance spectra. In *JPL Proceedings of the 3rd Airborne Imaging Spectrometer Data Analysis Workshop* (Vol. Document ID: 19880004388, pp. 138–142). Pasadena, California: NASA NTRS N88-13770, Document ID 19880004388. Retrieved from https://ntrs.nasa.gov/citations/19880004388

Clark, R. N., & Roush, T. L. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research: Solid Earth*, *89*(B7), 6329–6340. https://doi.org/10.1029/JB089iB07p06329

Crist, E. P., & Cicone, R. C. (1984). A Physically-Based Transformation of Thematic Mapper Data—The TM Tasseled Cap. *IEEE Transactions on Geoscience and Remote Sensing*, *GE-22*(3), 256–263. https://doi.org/10.1109/TGRS.1984.350619

Davis, F., Baldocchi, D., & Tyler, C. (2016). Chapter 25: Oak Woodlands. In *Ecosystems of California* (pp. 509–529). Oakland, CA: University of California Press.

Dozier, J., Green, R. O., Nolin, A. W., & Painter, T. H. (2009). Interpretation of snow properties from imaging spectrometry. *Imaging Spectroscopy Special Issue*, *113*, S25–S37. https://doi.org/10.1016/j.rse.2007.07.029

Elvidge, C. D. (1990). Visible and near infrared reflectance characteristics of dry plant materials. *International Journal of Remote Sensing*, *11*(10), 1775–1795. https://doi.org/10.1080/01431169008955129

F. M. Lacar, M. M. Lewis, & I. T. Grierson. (2001). Use of hyperspectral imagery for mapping grape varieties in the Barossa Valley, South Australia. In *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)* (Vol. 6, pp. 2875–2877 vol.6). https://doi.org/10.1109/IGARSS.2001.978191

Féret, J.-B., & Asner, G. P. (2014). Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. *Ecological Applications*, *24*(6), 1289–1296. https://doi.org/10.1890/13-1824.1

Féret, J.-B., François, C., Gitelson, A., Asner, G. P., Barry, K. M., Panigada, C., et al. (2011). Optimizing spectral indices and chemometric analysis of leaf chemical properties using radiative transfer modeling. *Remote Sensing of Environment*, *115*(10), 2742–2750. https://doi.org/10.1016/j.rse.2011.06.016

Fisher, P. (1997). The pixel: a snare and a delusion. *International Journal of Remote Sensing*, *18*(3), 679–685.

Gamon, J. A., Somers, B., Malenovský, Z., Middleton, E. M., Rascher, U., & Schaepman, M. E. (2019). Assessing Vegetation Function with Imaging Spectroscopy. *Surveys in Geophysics*, *40*(3), 489–513. https://doi.org/10.1007/s10712-019-09511-5

Ghrefat, H. A., Goodell, P. C., Hubbard, B. E., Langford, R. P., & Aldouri, R. E. (2007). Modeling grain size variations of aeolian gypsum deposits at White Sands, New Mexico, using AVIRIS imagery. *Geomorphology*, *88*(1), 57–68. https://doi.org/10.1016/j.geomorph.2006.10.013

Gillespie, A. (1990). Interpretation of residual images: spectral mixture analysis of AVIRIS images, Owens Valley, California. In *Proc. second airborne visible/infrared imaging spectrometer (AVIRIS) workshop* (pp. 243–270). Pasadena, California: NASA.

Hacker, P. W., Coops, N. C., Townsend, P. A., & Wang, Z. (2020). Retrieving Foliar Traits of Quercus garryana var. garryana across a Modified Landscape Using Leaf Spectroscopy and LiDAR. *Remote Sensing*, *12*(1), 26. https://doi.org/10.3390/rs12010026

Hapke, B. (2012). *Theory of reflectance and emittance spectroscopy*. Cambridge university press.

Huang, Z., Turner, B. J., Dury, S. J., Wallis, I. R., & Foley, W. J. (2004). Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. *Remote Sensing of Environment*, *93*(1), 18–29. https://doi.org/10.1016/j.rse.2004.06.008

Huesca, M., García, M., Roth, K. L., Casas, A., & Ustin, S. L. (2016). Canopy structural attributes derived from AVIRIS imaging spectroscopy data in a mixed broadleaf/conifer forest. *Remote Sensing of Environment*, *182*, 208–226. https://doi.org/10.1016/j.rse.2016.04.020

Hunt, G. R. (1977). Spectral signatures of particulate minerals in the visible and near infrared. *Geophysics*, *42*(3), 501–513. https://doi.org/10.1190/1.1440721

Jensen, J. (2006). *Remote Sensing of the Environment: An Earth Resource Perspective* (2nd ed.). Pearson.

K. Cawse-Nicholson, S. J. Hook, C. E. Miller, & D. R. Thompson. (2019). Intrinsic Dimensionality in Combined Visible to Thermal Infrared Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(12), 4977–4984. https://doi.org/10.1109/JSTARS.2019.2938883

Kauth, R. J., & Thomas, G. S. (1976). The Tasselled Cap -- A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by LANDSAT. In *Symposium on Machine Processing of Remotely Sensed Data* (Vol. Paper 159, pp. 41–51). West Lafayette, Indiana, 4B: The Laboratory for Applications of Remote Sensing, Purdue University. https://docs.lib.purdue.edu/lars_symp/159/

Kokaly, R. F., Asner, G. P., Ollinger, S. V., Martin, M. E., & Wessman, C. A. (2009). Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Imaging Spectroscopy Special Issue*, *113*, S78–S91. https://doi.org/10.1016/j.rse.2008.10.018

Kruse, F. A., Lefkoff, A. B., & Dietz, J. B. (1993). Expert system-based mineral mapping in northern death valley, California/Nevada, using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). *Airbone Imaging Spectrometry*, *44*(2), 309–336. https://doi.org/10.1016/0034-4257(93)90024-R

Krutz, D., Müller, R., Knodt, U., Günther, B., Walter, I., Sebastian, I., et al. (2019). The Instrument Design of the DLR Earth Sensing Imaging Spectrometer (DESIS). *Sensors*, *19*(7). https://doi.org/10.3390/s19071622

Larsolle, A., & Hamid Muhammed, H. (2007). Measuring crop status using multivariate analysis of hyperspectral field reflectance with application to disease severity and plant density. *Precision Agriculture*, *8*(1), 37–47. https://doi.org/10.1007/s11119-006-9027-4

Li, W., & Wu, C. (2014). Phenology-based temporal mixture analysis for estimating large-scale impervious surface distributions. *International Journal of Remote Sensing*, *35*(2), 779–795. https://doi.org/10.1080/01431161.2013.873147

Margetta, R. (2021, May 24). New NASA Earth System Observatory to Help Address Climate Change [Text]. Retrieved August 19, 2021, from http://www.nasa.gov/press-release/new-nasa-earth-system-observatory-to-help-address-mitigate-climate-change

Mutanga, O., Skidmore, A. K., & Prins, H. H. T. (2004). Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features. *Remote Sensing of Environment*, *89*(3), 393–408. https://doi.org/10.1016/j.rse.2003.11.001

Naidoo, L., Cho, M. A., Mathieu, R., & Asner, G. (2012). Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment. *ISPRS Journal of Photogrammetry and Remote Sensing*, *69*, 167–179. https://doi.org/10.1016/j.isprsjprs.2012.03.005

Nassau, K. (1987). The fifteen causes of color: The physics and chemistry of color. *Color Research & Application*, *12*(1), 4–26. https://doi.org/10.1002/col.5080120105

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

NEON (National Ecological Observatory Network), 2021a. Spectrometer orthorectified surface directional reflectance - flightline, RELEASE-2021 (DP1.30006.001). https://doi.org/10.48443/n3ys-2070. Dataset accessed from https://data.neonscience.org on October 13, 2021

NEON (National Ecological Observatory Network), 2021b. Plant presence and percent cover, RELEASE-2021 (DP1.10058.001). https://doi.org/10.48443/abge-r811. Dataset accessed from https://data.neonscience.org on October 13, 2021

Nolin, A. W., & Dozier, J. (1993). Estimating snow grain size using AVIRIS data. *Airbone Imaging Spectrometry*, *44*(2), 231–238. https://doi.org/10.1016/0034-4257(93)90018-S

Ollinger, S. V. (2011). Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist*, *189*(2), 375–394. https://doi.org/10.1111/j.1469-8137.2010.03536.x

Oshigami, S., Yamaguchi, Y., Uezato, T., Momose, A., Arvelyna, Y., Kawakami, Y., et al. (2013). Mineralogical mapping of southern Namibia by application of continuum-removal MSAM method to the HyMap data. *International Journal of Remote Sensing*, *34*(15), 5282–5295. https://doi.org/10.1080/01431161.2013.789147

Piwowar, J. M., Peddle, D. R., & LeDrew, E. F. (1998). Temporal mixture analysis of arctic sea ice imagery: a new approach for monitoring environmental change. *Remote Sensing of Environment*, *63*(3), 195–207.

Plaza, J., Plaza, A., Perez, R., & Martinez, P. (2009). On the use of small training sets for neural network-based characterization of mixed pixels in remotely sensed hyperspectral images. *Pattern Recognition*, *42*(11), 3032–3045. https://doi.org/10.1016/j.patcog.2009.04.008

Roberts, D. A., Gardner, M., Church, R., Ustin, S., Scheer, G., & Green, R. O. (1998). Mapping Chaparral in the Santa Monica Mountains Using Multiple Endmember Spectral Mixture Models. *Remote Sensing of Environment*, *65*(3), 267–279. https://doi.org/10.1016/S0034-4257(98)00037-6

Settle, J. J., & Drake, N. A. (1993). Linear mixing and the estimation of ground cover proportions. *International Journal of Remote Sensing*, *14*(6), 1159–1177. https://doi.org/10.1080/01431169308904402

Singh, A., Serbin, S. P., McNeil, B. E., Kingdon, C. C., & Townsend, P. A. (2015). Imaging spectroscopy algorithms for mapping canopy foliar chemical and morphological traits and their uncertainties. *Ecological Applications*, *25*(8), 2180–2197. https://doi.org/10.1890/14-2098.1

Small, C. (2004). The Landsat ETM+ spectral mixing space. *Remote Sensing of Environment*, *93*(1), 1–17. https://doi.org/10.1016/j.rse.2004.06.007

Small, C. (2012). Spatiotemporal dimensionality and Time-Space characterization of multitemporal imagery. *Remote Sensing of Environment*, *124*, 793–809.

Small, C. (2018). Multisource imaging of urban growth and infrastructure using Landsat, Sentinel and SRTM. In *NASA Landsat-Sentinel Science Team Meeting*. Rockville, MD.

Small, C., & Das, I. (2018). The Cryospheric Spectral Mixing Space. Presented at the AGU Fall Meeting 2018, AGU.

Small, C., & Milesi, C. (2013). Multi-scale standardized spectral mixture models. *Remote Sensing of Environment*, *136*, 442–454. https://doi.org/10.1016/j.rse.2013.05.024

Smith, M. O., Ustin, S. L., Adams, J. B., & Gillespie, A. R. (1990). Vegetation in deserts: I. A regional measure of abundance from multispectral images. *Remote Sensing of Environment*, *31*(1), 1–26. https://doi.org/10.1016/0034-4257(90)90074-V

Smith, M.-L., Ollinger, S. V., Martin, M. E., Aber, J. D., Hallett, R. A., & Goodale, C. L. (2002). DIRECT ESTIMATION OF ABOVEGROUND FOREST PRODUCTIVITY THROUGH HYPERSPECTRAL REMOTE SENSING OF CANOPY NITROGEN. *Ecological Applications*, *12*(5), 1286–1302. https://doi.org/10.1890/1051-0761(2002)012[1286:DEOAFP]2.0.CO;2

Sousa, D., & Davis, F. W. (2020). Scalable mapping and monitoring of Mediterranean-climate oak landscapes with temporal mixture models. *Remote Sensing of Environment*, *247*, 111937. https://doi.org/10.1016/j.rse.2020.111937

Sousa, D., & Small, C. (2017). Global cross-calibration of Landsat spectral mixture models. *Remote Sensing of Environment*, *192*, 139–149. https://doi.org/10.1016/j.rse.2017.01.033

Sousa, D., & Small, C. (2018). Multisensor Analysis of Spectral Dimensionality and Soil Diversity in the Great Central Valley of California. *Sensors*, *18*(2), 583. https://doi.org/10.3390/s18020583

Sousa, D., & Small, C. (2019a). Globally standardized MODIS spectral mixture models. *Remote Sensing Letters*, *10*(10), 1018–1027. https://doi.org/10.1080/2150704X.2019.1634299

Sousa, D., & Small, C. (2019b). Mapping and Monitoring Rice Agriculture with Multisensor Temporal Mixture Models. *Remote Sensing*, *11*(2), 181. https://doi.org/10.3390/rs11020181

Strahler, A. H., Woodcock, C. E., & Smith, J. A. (1986). On the nature of models in remote sensing. *Remote Sensing of Environment*, *20*(2), 121–139.

Thenkabail, P. S., Smith, R. B., & De Pauw, E. (2000). Hyperspectral Vegetation Indices and Their Relationships with Agricultural Crop Characteristics. *Remote Sensing of Environment*, *71*(2), 158–182. https://doi.org/10.1016/S0034-4257(99)00067-X

Thenkabail, P. S., Lyon, J. G., Huete, A., Lyon, J. G., & Huete, A. (2018). *Hyperspectral Indices and Image Classifications for Agriculture and Vegetation*. CRC Press. https://doi.org/10.1201/9781315159331

Thompson, D. R., Boardman, J. W., Eastwood, M. L., & Green, R. O. (2017). A large airborne survey of Earth's visible-infrared spectral dimensionality. *Optics Express*, *25*(8), 9186–9195. https://doi.org/10.1364/OE.25.009186

Townsend, P. A., Foster, J. R., Chastain, R. A., & Currie, W. S. (2003). Application of imaging spectroscopy to mapping canopy nitrogen in the forests of the central Appalachian Mountains using Hyperion and AVIRIS. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(6), 1347–1354. https://doi.org/10.1109/TGRS.2003.813205

Ustin, S. L., & Gamon, J. A. (2010). Remote sensing of plant functional types. *New Phytologist*, *186*(4), 795–816. https://doi.org/10.1111/j.1469-8137.2010.03284.x

Voss, M., & Sugumaran, R. (2008). Seasonal Effect on Tree Species Classification in an Urban Environment Using Hyperspectral Data, LiDAR, and an Object- Oriented Approach. *Sensors*, *8*(5). https://doi.org/10.3390/s8053020

W. W. Hargrove & F. M. Hoffman. (1999). Using multivariate clustering to characterize ecoregion borders. *Computing in Science & Engineering*, *1*(4), 18–25. https://doi.org/10.1109/5992.774837

Woodcock, C. E., & Strahler, A. H. (1987). The factor of scale in remote sensing. *Remote Sensing of Environment*, *21*(3), 311–332. https://doi.org/10.1016/0034-4257(87)90015-0

Yang, F., Matsushita, B., Fukushima, T., & Yang, W. (2012). Temporal mixture analysis for estimating impervious surface area from multi-temporal MODIS NDVI data in Japan. *ISPRS Journal of Photogrammetry and Remote Sensing*, *72*, 90–98. https://doi.org/10.1016/j.isprsjprs.2012.05.016

Yen-Ben Cheng, Edward Tom, & Susan L. Ustin. (2007). Mapping an invasive species, kudzu (Pueraria montana), using hyperspectral imagery in western Georgia. *Journal of Applied Remote Sensing*, *1*(1), 1–11. https://doi.org/10.1117/1.2749266

Zhang, C., & Xie, Z. (2012). Combining object-based texture measures with a neural network for vegetation mapping in the Everglades from hyperspectral imagery. *Remote Sensing of Environment*, *124*, 310–320. https://doi.org/10.1016/j.rse.2012.05.015