

# A Large Airborne Survey of Earth's Visible-Infrared Spectral Dimensionality

DAVID R. THOMPSON,<sup>1,\*</sup> JOSEPH W. BOARDMAN,<sup>2</sup> MICHAEL L. EASTWOOD,<sup>1</sup> AND ROBERT O. GREEN<sup>1</sup>

<sup>1</sup> Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109 USA

<sup>2</sup> Analytical Imaging and Geophysics, LLC, 4450 Arapahoe Ave, Suite 100, Boulder, CO 80303 USA

\*[david.r.thompson@jpl.nasa.gov](mailto:david.r.thompson@jpl.nasa.gov)

**Abstract:** The intrinsic spectral dimensionality indicates the observable degrees of freedom in Earth's solar-reflected light field, quantifying the diversity of spectral content accessible by visible and infrared remote sensing. The solar-reflected regime spans the 0.38 - 2.5  $\mu\text{m}$  interval, and is captured by a wide range of current and planned instruments on both airborne and orbital platforms. To date there has been no systematic study of its spectral dimensionality as a function of space, time, and land cover. Here we report a multi-site, multi-year statistical survey by NASA's "Classic" Airborne Visible Near InfraRed Spectrometer (AVIRIS-C). AVIRIS-C measured large regions of California, USA, spanning wide latitudinal and elevation gradients containing all canonical MODIS land cover types. The spectral uniformity of the AVIRIS-C design enabled consistent in-scene assessment of measurement noise across acquisitions. The estimated dimensionality as a function of cover type ranged from the low 20s to the high 40s, and was approximately 50 for the combined dataset. This result indicates the high diversity of physical processes distinguishable by imaging spectrometers like AVIRIS-C for one region of the Earth.

© 2017 Optical Society of America

**OCIS codes:** (010.0280) Remote sensing and sensors; (300.6340) Spectroscopy, infrared; (300.6550) Spectroscopy, visible.

## References and links

1. A. Nair, S. Sathyendranath, T. Platt, J. Morales, V. Stuart, M.-H. Forget, E. Devred, and H. Bouman, "Remote sensing of phytoplankton functional types," *Remote Sens. Environ.* **112**, 3366–3375 (2008).
2. C. G. Fichot, B. D. Downing, B. A. Bergamaschi, L. Windham-Myers, M. Marvin-DiPasquale, D. R. Thompson, and M. M. Gierach, "High-resolution remote sensing of water quality in the San Francisco Bay–Delta Estuary," *Environ. Sci. Tech.* **50**, 573–583 (2015).
3. G. P. Asner and R. E. Martin, "Spectral and chemical analysis of tropical forests: Scaling from leaf to canopy levels," *Remote Sens. Environ.* **112**, 3958–3970 (2008).
4. R. N. Clark, G. A. Swayze, K. E. Livo, R. F. Kokaly, S. J. Sutley, J. B. Dalton, R. R. McDougal, and C. A. Gent, "Imaging spectroscopy: Earth and planetary remote sensing with the USGS tetracorder and expert systems," *Journal of Geophysical Research: Planets* **108** (2003).
5. D. R. Thompson, B.-C. Gao, R. O. Green, D. A. Roberts, P. E. Dennison, and S. R. Lundeen, "Atmospheric correction for global mapping spectroscopy: ATREM advances for the HypSPIRI preparatory campaign," *Remote Sens. Environ.* **167**, 64–77 (2015).
6. D. R. Thompson, I. Leifer, H. Bovensmann, M. Eastwood, M. Fladeland, C. Frankenberg, K. Gerilowski, R. Green, S. Kratwurst, T. Krings *et al.*, "Real-time remote detection and measurement for airborne imaging spectroscopy: a case study with methane," *Atmos. Meas. Tech.* **8**, 4383–4397 (2015).
7. E. J. Hochberg, D. A. Roberts, P. E. Dennison, and G. C. Hulley, "Special issue on the Hyperspectral InfraRed Imager (HypSPIRI): Emerging science in terrestrial and aquatic ecology, radiation balance and hazards," *Remote Sens. Environ.* **167**, 1–5 (2015).
8. L. Guanter, H. Kaufmann, K. Segl, S. Foerster, C. Rogass, S. Chabrillat, T. Kuester, A. Hollstein, G. Rossner, C. Chlebek *et al.*, "The EnMAP spaceborne imaging spectroscopy mission for Earth observation," *Remote Sens.* **7**, 8830–8857 (2015).
9. A. Iwasaki, N. Ohgi, J. Tani, T. Kawashima, and H. Inada, "Hyperspectral Imager Suite (HISUI)-japanese hyper-multi spectral radiometer," in *Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* pp. 1025–1028 (2011).

10. J. W. Boardman and R. O. Green, "Exploring the spectral variability of the earth as measured by AVIRIS in 1999," in the Proceedings of the Ninth JPL Airborne Earth Science Workshop pp. 195–206 (2000).
11. C. Small, "Multiresolution analysis of urban reflectance," in the Proceedings of the IEEE/ISPRS Joint Workshop 2001: Remote Sensing and Data Fusion over Urban Areas pp. 15–19 (2001).
12. G. P. Asner, D. E. Knapp, J. Boardman, R. O. Green, T. Kennedy-Bowdoin, M. Eastwood, R. E. Martin, C. Anderson, and C. B. Field, "Carnegie Airborne Observatory-2: Increasing science data dimensionality via high-fidelity multi-sensor fusion," *Remote Sens. Environ.* **124**, 454–465 (2012).
13. K. Cawse-Nicholson, S. B. Damelin, A. Robin, and M. Sears, "Determining the intrinsic dimension of a hyperspectral image using random matrix theory," *IEEE T. Image Process.* .
14. C.-C. Wu, W. Liu, and C.-I. Chang, "Exploration of methods for estimation of number of endmembers in hyperspectral imagery," *Proc. SPIE* **6378**, 63781C (2006).
15. H. Akaike, "A new look at the statistical model identification," *IEEE T. Automat. Contr.* **19**, 716–723 (1974).
16. J. Rissanen, "Modeling by shortest data description," *Automatica* **14**, 465–471 (1978).
17. H.-T. Wu, J.-F. Yang, and F.-K. Chen, "Source number estimators using transformed gerschgorin radii," *IEEE T. Signal Proces.* **43**, 1325–1333 (1995).
18. J. M. Bioucas-Dias and J. M. Nascimento, "Estimation of signal subspace on hyperspectral data," *Proc. SPIE* **8205**, 59820 (2005).
19. J. C. Harsanyi, W. Farrand, and C.-I. Chang, "Detection of subpixel spectral signatures in hyperspectral image sequences," in Proceedings of American Society of Photogrammetry & Remote Sensing Annual Meeting pp. 236–247 (1994).
20. N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Proc. Mag.* **19**, 44–57 (2002).
21. R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.* **65**, 227–248 (1998).
22. J. W. Boardman and F. A. Kruse, *IEEE T. Geosci. Remote.* .
23. S. Channan, K. Collins, and W. R. Emanuel, "Global mosaics of the standard MODIS land cover type data," Distributed by University of Maryland and the Pacific Northwest National Laboratory (2014).
24. M. A. Friedl, D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang, "MODIS collection 5 global land cover: Algorithm refinements and characterization of new datasets, 2001-2012," Collection 5.1 IGBP Land Cover (2010).
25. D. R. Thompson, R. O. Green, D. Keymeulen, S. K. Lundeen, Y. Mouradi, D. C. Nunes, R. C. no, and S. A. Chien, "Rapid spectral cloud screening onboard aircraft and spacecraft," *IEEE T. Geosci. Remote* **52**, 6779–6792 (2014).
26. D. R. Thompson, D. A. Roberts, B. C. Gao, R. O. Green, L. Guild, K. Hayashi, R. Kudela, and S. Palacios, "Atmospheric correction with the Bayesian empirical line," *Opt. Express* **24**, 2134–2144 (2016).

## 1. Introduction

The Visible-ShortWave InfraRed (VSWIR) electromagnetic spectrum from 0.38-2.5  $\mu\text{m}$  contains most surface-reflected solar energy available for remote sensing. These wavelengths access diverse physical absorption and scattering properties throughout the Earth system. For example:

- 0.38-0.75  $\mu\text{m}$  wavelengths penetrate water, revealing the condition and composition of benthic ecosystems, water optical properties, phytoplankton [1], and water quality [2].
- 0.4-2.5  $\mu\text{m}$  wavelengths interact with vegetation, undergoing absorption by pigments, scattering and absorption in canopies, and absorption by structural elements, indicating plant health, structure, foliar chemistry and even species [3].
- 0.6-2.5  $\mu\text{m}$  wavelengths show vibrational absorption features in solids including iron oxides, phyllosilicates, carbonates, and other materials such as metals and hydrocarbons [4]. This reveals a wide range of natural mineralogy and artificial materials in urban environments.
- 0.38-2.5  $\mu\text{m}$  wavelengths undergo absorption by different atmospheric constituents including  $\text{H}_2\text{O}$ ,  $\text{O}_2$ ,  $\text{CH}_4$ ,  $\text{CO}_2$ , and scattering by aerosols [5, 6].

The total number of degrees of freedom determines the spectral sampling required to fully measure the VSWIR upwelling light field and disambiguate target phenomena from confounding processes. This "intrinsic spectral dimensionality" holds implications for future and proposed

orbital instruments including HypIRI [7], EnMAP [8], and HISUI [9]. Prior investigations have explored VSWIR dimensionality in specific scenes and pooled archives, but no experiment has yet assessed a wide temporal and spatial range to evaluate variability over space and time.

Prior research gives some clues. Boardman and Green evaluated a historical archive of NASA's "Classic" Airborne Visible Near Infrared Spectrometer (AVIRIS-C) [10]. They found dimensionalities ranging from 20 to 50 for scenes taken independently and far higher for the combined dataset. Small analyzed an urban AVIRIS-C scene, finding a spatially-coherent subspace of about 31-35 dimensions [11]. Simulations of coarser spatial resolutions reduced this dimensionality due to mixed pixels. The Carnegie Airborne Observatory later corroborated this spatial dependence; its sampling of 1-2 m revealed significantly higher dimensionality [12]. It is intuitive that open-water aquatic environments would have the smallest dimensionality due to a simplified optical system and water absorption beyond 1.0  $\mu\text{m}$ . Conversely, urban environments are believed to be most spectrally diverse. Asner et al. recently demonstrated this for a two-site study comprising about 400 ha total [12]. There remains a need to characterize dimensionality magnitudes and trends rigorously over a wide spatiotemporal range, with a single sensor and a consistent spatial resolution similar to anticipated orbital instruments.

This study analyzes data from the HypIRI preparatory campaign, a series of AVIRIS-C flights over California in 2013-2015. Here we focus on the 2013 and 2014 data comprising over 600 distinct flightlines of diverse biomes, latitudes, and elevations. AVIRIS-C flew onboard NASA's high altitude ER-2 aircraft flying at nominal 20 km altitude, above nearly all atmospheric scattering and absorption. The spatial instantaneous field of view was 1 milliradian giving 20 m sampling. These properties made it relevant to a future class of Earth-orbiting imaging spectrometers operating at similar spatial and spectral sampling. We sought to (a) evaluate the properties of the dimensionality distribution and its stability over time, (b) test the specific hypothesis of increasing spectral diversity in aquatic, terrestrial natural environments, and urban environments respectively.

## 2. Approach

There are many ways to estimate dimensionality for different assumptions about the data distribution [13]; we desire a method that is simple, robust and automatic for application on large radiance datasets. A common approach is to analyze the eigenvalues of the covariance matrix as in Boardman and Green [10] and Asner et al. [12]. Low-rank data lie on a strict linear subspace implying one or more of the eigenvalues corresponding to the null space will be zero. Eigenvalues are initially large, decay quickly, and never quite reach zero due to measurement noise. Consequently dimensionality estimation reduces to the question of which eigenvectors represent scene structure above the noise level. Wu et al. [14] evaluate alternative solutions including: model selection based on information criteria [15], and the minimum description length [16]; Gershgorin radii [17]; Signal Subspace Estimation, or SSE [18]; and Neyman-Pearson detection theory [19]. The question is ultimately a subjective issue of definitions [13], but most methods provide similar results and relative comparisons of a single estimator across datasets are still meaningful. In the Wu et al. survey, only SSE achieves good performance without prespecified noise thresholds [14]. However, it relies on a nonnegativity property of geographic mixture models; this applies to reflectance data [20] but may be less appropriate for radiance data having variable atmospheric attenuation. For radiances the standard practice remains an orthogonal decomposition of the covariance matrix into eigenbases, provided a suitable noise threshold can be identified.

Our noise estimation approach was similar to that of Asner et al. [12] except that we automated the spatial coherence metric to enable objective estimates across a large dataset. We exploited the unique spatial uniformity of the AVIRIS-C instrument — a "whiskbroom" design in which the same optics and detectors measured every spatial location [21]. In the downtrack direction,

AVIRIS-C acquired temporally discontinuous measurements across consecutive scans. Any spatial correlation between such spectra was unlikely to occur by chance unless it was a part of the scene [11]. This allowed separation of the spatially-coherent radiance field from spatially random noise. Singular Value Decomposition (SVD) factorized the total covariance  $\Sigma$  by eigenvalues  $\lambda_i$  and column eigenvectors  $\mathbf{v}_i$ . The projection of a radiance spectrum  $\mathbf{L}$  onto the  $i$ th such eigenvalue was written  $x_i = \mathbf{v}_i^T (\mathbf{L} - \mu)$  where  $\mu$  was the mean spectrum. The sequence of projections for a given cross-track location  $j$  and downtrack positions  $1 \dots n$  was the vector  $\mathbf{y}_{ij} = [x_{ij1}, \dots, x_{ijn}]$ . We estimated the spatially-correlated component of this signal by a nonparametric smoothing operator  $f(\mathbf{y}_{ij})$ , and the difference  $f(\mathbf{y}_{ij}) - \mathbf{y}_{ij}$  was the spatially-uncorrelated remainder. The standard deviations of these two components represented the signal and noise magnitudes respectively. The smoothing operator could have been any nonparametric smoothing algorithm, but we simply took the local average of an 11-pixel neighborhood in the downtrack direction. In summary, the sequence of steps was:

1. Calculate the mean and covariance spectra of the calibrated radiance cube.
2. Using SVD, decompose the covariance matrix into eigenvalues and eigenvectors ordered from most to least significant.
3. Project zero-meanded spectra onto eigenvectors, transforming the radiances to a Principal Component (PC) representation.
4. Smooth each channel and cross-track position of the PC cube using a 1-D local averaging operator in the downtrack direction.
5. Define the signal to be the smoothed PC cube, and the noise to be the difference between the original and smoothed PC cubes.
6. The estimated dimensionality is the last eigenvalue index for which the standard deviation of the signal image exceeds that of the noise image.

Our methodology contrasted with a related linear approach, the Minimum Noise Fraction (MNF) [11, 22], which estimates the noise in the original measurement space, and then whitens the data to normalize the eigenvalues so that the noise was at unity. MNF requires the noise covariance to be specified in advance, and typically gives similar dimensionality estimates [11].

We radiometrically and spectrally calibrated the AVIRIS data to radiance units using procedures in Green et al. [21], and then located them on the Earth's surface by ray-tracing from the GPS/IMU-derived sensor position to a digital elevation model. Finally, we resampled the data to a consistent geographic spacing, using nearest-neighbor sampling to preserve pristine uninterpolated spectra. We then subdivided each flightline into segments 20 km along track, approximately equal to twice the width which varied slightly based on aircraft motion. This provided over 4000 segments, with each segment containing approximately 700000 spectra. Simple heuristic filtering removed occasional bad data from electronic effects, leaving over  $2.5 \times 10^9$  measurements. We then analyzed this dataset to determine the spectral dimensionality of flightline segments and of the dataset as a whole. A list of flightlines appears in Table 1.

Example noise and signal magnitude curves appear in Fig. 1, showing two typical flightlines over the AVIRIS-C Ivanpah calibration site separated by approximately two months. As expected, scene structure dominated projections with large eigenvalues while spatially-uncorrelated measurement noise dominated the smaller eigenvalues. The intersection between the two curves indicated the projection where signal and noise had equal magnitude; by convention we used this intersection point as the dimensionality estimate for the scene. For additional resilience to noise around the intersection where the two curves were nearly tangent, we smoothed them with a width-9 kernel before calculating the crossover point. In Fig. 1, the two colored lines show different flight days. They exhibited near perfect agreement with estimated dimensionalities of

Table 1. Flightlines used in the experiment. Date: Flight date, N: Number of segments, Radiance: Mean radiance in  $\text{W m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$ , Dim: Mean Dimensionality, Lat: Center latitude of flightlines, Lon: Center longitude of flightlines.

Date	N	Radiance	Dim	Lat	Lon
2013 04 10	573	50.59	29.97	35.07	-119.12
2013 04 20	257	59.69	30.61	33.74	-117.09
2013 05 00	271	58.54	30.34	37.84	-119.59
2013 09 20	342	46.41	28.82	35.47	-118.23
2013 10 30	114	24.25	26.23	37.62	-122.39
2013 11 00	150	32.39	27.50	37.39	-119.45
2013 11 20	297	24.91	24.92	36.00	-120.36
2014 04 10	550	57.84	29.94	35.80	-118.53
2014 04 20	171	49.24	27.77	34.69	-119.47
2014 05 10	185	54.05	29.66	37.86	-122.18
2014 05 30	187	52.36	31.02	37.87	-122.20
2014 06 00	369	60.12	36.24	36.96	-119.77
2014 06 10	273	64.28	31.93	34.08	-118.13
2014 08 20	66	58.04	31.62	33.92	-117.49
2014 08 30	397	52.10	30.65	34.05	-117.90
2014 09 20	123	48.21	27.19	38.65	-120.40
2014 10 10	131	38.13	30.83	37.35	-119.40
2014 10 30	194	32.17	28.33	37.85	-122.19

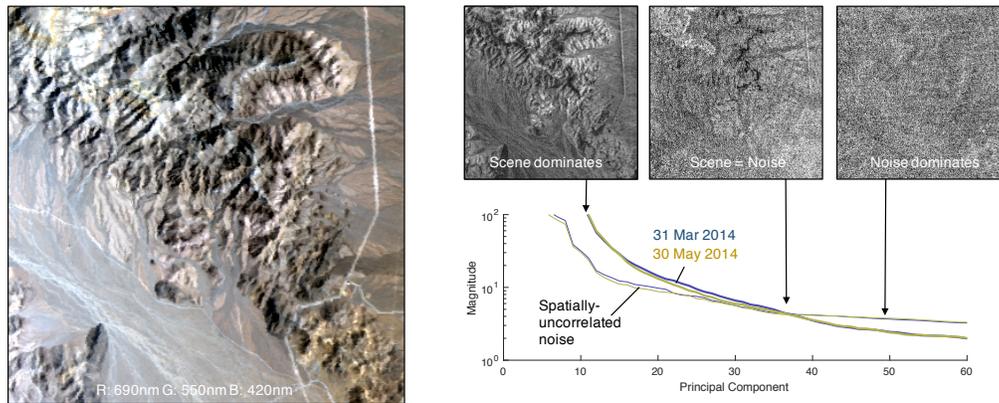


Fig. 1. Left: RGB subset of an example validation segment - foothills of Ivanpah Playa, overflowed on multiple occasions under different illumination conditions. Right: The average magnitude of the projection of each Principal Component, separated into spatially-correlated (thick line) and uncorrelated (thin line) components. The intersection indicates the dimension where the measurement noise is equal to the scene signal.

37, demonstrating the method was consistent. We calculated a similar crossover point for each segment in the dataset (Figure 2).

Finally, we associated the combined dataset with MODIS land cover categories [23, 24]. Each polygonal segment of AVIRIS-C had a single dimensionality estimate, but contained many MODIS land cover map pixels. We tried two approaches to reconcile the resolutions: a hard assignment, defining each datapoint as an AVIRIS-C segment with a land cover determined by

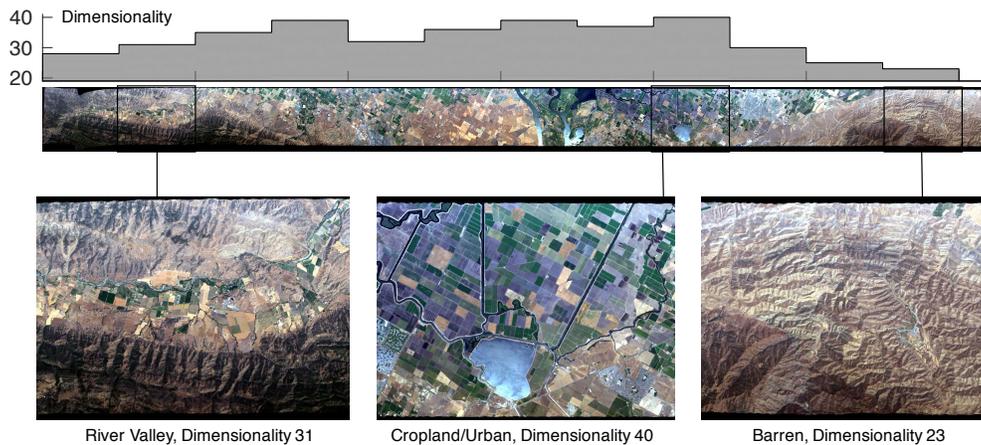


Fig. 2. Example segmentation of flightline f1405f140528r15 near the Bay Area, CA. The bar chart at the top shows dimensionality estimates for each segment. Images indicate representative segments with intermediate, high, and low values.

its center location; and a proportional assignment, defining each datapoint as a MODIS land cover pixel with dimensionality determined by the encompassing AVIRIS-C polygon. The latter violated the assumption of Independent and Identically Distributed (IID) data, but we shall see the choice did not affect our results.

### 3. Results

Fig. 3 shows the eigenvalue decay curve for the combined dataset, with segments' covariance matrices normalized according to their estimated noise. Eigenvalues of unity represent the eigenvector where signal is equal to noise, which here occurs at a value of 45. This was slightly smaller than the pooled dimensionality reported by Boardman et al. [10]. The difference may have been related to our spatial coherence criterion, or our estimation of noise independently in each scene rather than a static Noise-Equivalent delta Radiance (NE $\Delta$ L) spectrum. Nevertheless, the result broadly agreed with prior estimates and underscored the high overall diversity of the VSWIR range.

Next we partitioned the dataset into MODIS land cover categories [23, 24]. Fig. 4 (top panel) shows the dimensionality distribution for each land cover type using proportional assignment, with circles indicating the median and bars indicating quartiles. As expected, Urban and Cropland regions had the highest overall dimensionality. Barren regions and Water (which included both inland water and sea) had the lowest, while Forest, Shrubland and Grassland ecosystems had similar distributions with a dimensionality of approximately 30 per scene. Separating the terrestrial anthropogenic scenes having Cropland, Mosaic, or Urban land cover from other terrestrial scenes, and excluding outliers above dimensionality 50, we found each population was well-characterized by its own Gaussian distribution (Fig. 4, bottom panels). The mean and standard deviation of the natural scenes were 30.4 and 4.5, and those of anthropogenic scenes were 33.2 and 4.3. The difference was highly significant ( $p < 10^{-10}$  for a two-sample  $t$  test). The values were in the range of prior case studies that independently studied individual natural scenes [13, 14], or urban areas [11]. The HypsIRI preparatory dataset demonstrated that the two were best characterized by distinct distributions and that the trend held over wide areas.

The Water class suggested a multimodal distribution, so we considered both a single Gaussian and a two-component Gaussian mixture model fit using Expectation Maximization (EM) to

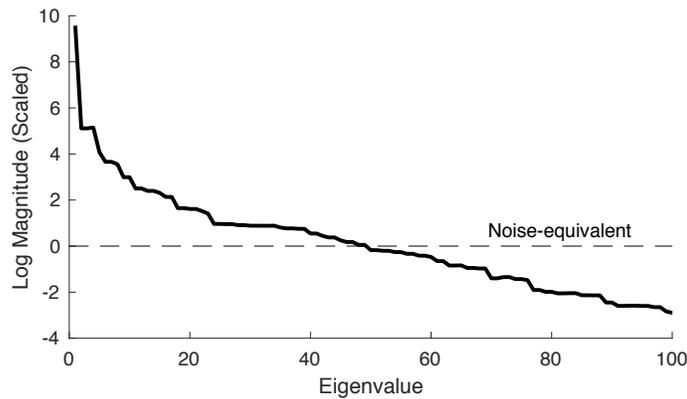


Fig. 3. Eigenvalue decay of the pooled covariance matrices, each rescaled to place noise at unity

maximize data likelihood. The Akaike information criterion (AIC) favored this bimodal alternative. The two modes had means 12.4 and 27.1, and standard deviations 3.8 and 5.6 respectively. The higher-dimensional mode may have resulted from polygons containing many terrain features, which controlled the dimensionality estimate despite the presence of some water. The lower-dimensional mode suggested segments of pure dark ocean with less observable spectral diversity, though perhaps a wider variability due to starkly-contrasting clouds, marine aerosols, and sunglint. Polygons containing very small patches of terrain with mostly water would likely show intermediate dimensionalities, further broadening the distributions.

Thanks to the physical contiguity of biomes, the two methods of reconciling MODIS and AVIRIS-C resolutions produced virtually identical results. Using the hard assignment method only shifted the median dimensionality for a few land cover categories: Water, from 24 to 23; Permanent Wetlands, from 31 to 29; and Snow/Ice, Evergreen Broadleaf and Deciduous Needle-leaf Forests, for which populations fell to one or zero datapoints. The aggregate distributions for natural, anthropogenic and water surface showed negligible change; means shifted by less than 1% and statistical separations remained highly significant.

Several effects can change the dimensionality over time. Solar elevation and atmospheric state can cause changes in signal and noise levels. Other physical effects that can alter the dimensionality include the introduction of novel spectra from atmospheric phenomena such as clouds, and or changes in surface cover due to seasonality in vegetation or snow. The Ivanpah example demonstrated the repeatability of the estimation method, i.e. the consistency in calculated dimensionality for specific conditions. That demonstration was possible because the Ivanpah playa scene was a barren environment without seasonal changes, and observed multiple times under similar atmosphere and illumination. In contrast, the HypsIRI dataset was far more general, and likely contained influence from all factors.

While quantifying this partitioning was beyond our scope, several trends invited interpretation. Fig. 5 shows the dimensionality of each segment. The 2014 dimensionalities were generally lower due to signal levels - the Spring 2014 datasets had lower solar elevation throughout, reducing average dimensionality. Label A shows high-altitude forest and meadow in Yosemite National Park, which had a high dimensionality that was likely related to the wide range of materials: mixed vegetation and canopy types together with geologic features and sparse melting snow. A2 identifies the same area in the subsequent year, acquired one month earlier under thick snow cover which muted this diversity. Label B shows a wilderness area of still lower spectral diversity comprised mainly of low vegetation and grasses in the foothills below the Sierra

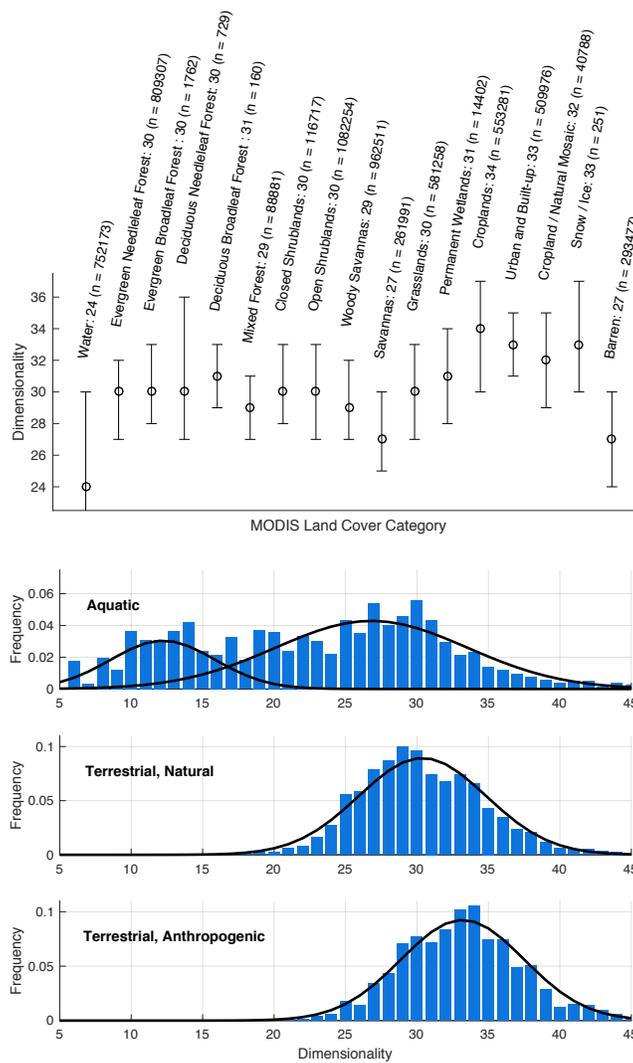


Fig. 4. Top: Intrinsic spectral dimensionality of MODIS land cover categories. Bottom: Frequency histograms of dimensionality for aquatic and terrestrial scenes. Fitting a bimodal distribution to the Water class yields means 12.4 and 27.1, and standard deviations 3.8 and 5.6 respectively. Natural and anthropogenic terrestrial scenes are well-described by Gaussian distributions of means 30.4 and 33.2, standard deviations of 4.5 and 4.3 respectively.

Nevada. Label C shows a typical diverse urban area - here, Fresno CA. Other metropolitan areas such as Bakersfield, the San Francisco Bay Area, Santa Barbara and Los Angeles all showed significantly increased spectral diversity. Label D shows typical cropland: mixtures of fertile and fallow fields near the Salton Sea. Cultivated areas in the Sacramento Bay Delta showed similarly high dimensionality. Label E shows a contrasting desert area north of the Salton Sea dominated by bare soil and rock with much lower values. There was also depleted spectral diversity in certain less-vegetated areas including inland of Santa Barbara, the low savanna terrain beneath the Sierra Nevada mountain range, and the densely-forested areas northwest of San Francisco.

The fall flightlines show similar overall trends as well as some unique features. Labels F and

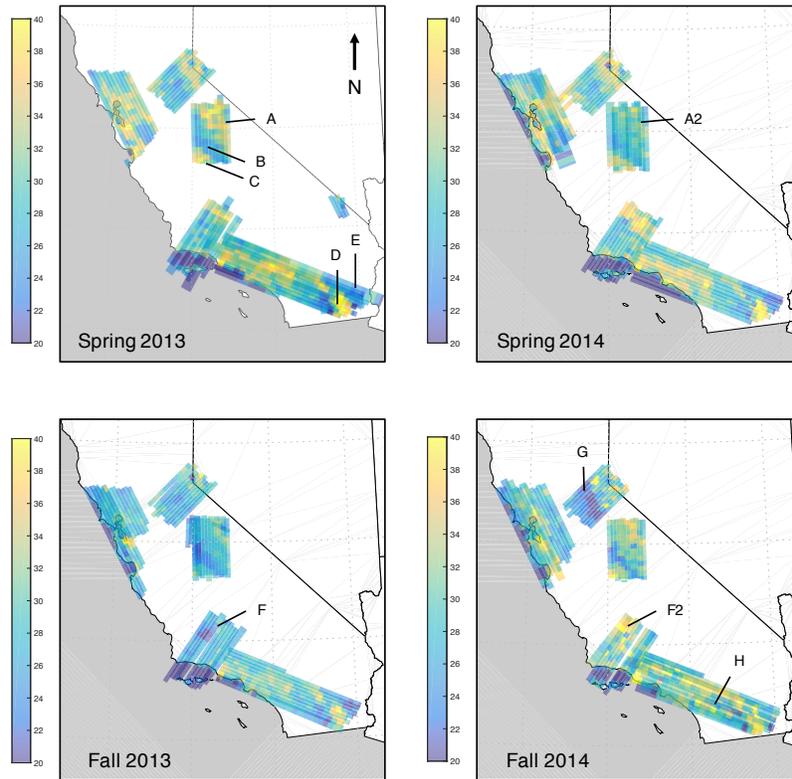


Fig. 5. Intrinsic spectral dimensionality of observations over California, 2013-14. Dimensionality is floored at 20 and ceilinged at 40. Letter annotations show locations of interest referenced in the text.

F2 show two areas imaged under very different illumination. The 2013 flightline was acquired with a low solar elevation (often less than 30 degrees), a signal reduction of over 40% from the same area a year later. Small-scale textures and cast shadow microshading within a pixel could have reduced this signal further; in aggregate, these effects significantly reduced the measurable spectral diversity. Label G shows another example: prominent band of low dimensionality aligned with the uniform, closed canopy of a conifer forest under low illumination. Label H shows scattered clouds, which artificially inflated the dimensionality by modulating small areas with additional reflectance and/or attenuation.

#### 4. Discussion and conclusions

This study provides the most comprehensive survey to date of the dimensionality of the VSWIR upwelling light field over wide spatiotemporal areas, showing dimensionalities of 30-35 for 20 km segments taken individually and approaching 50 for the dataset as a whole. The data for terrestrial scenes are well-represented by Gaussian distributions for natural and anthropogenic environments, with the latter having a significantly larger mean dimensionality. This is not a perfect unbiased measure of the Earth's surface, since the areal extent of different terrain types and elevations does not exactly match the planet and the flights were planned to coincide with clear days and favorable observing conditions. However, the study offers an informative view of the Earth's radiometric properties across a broad latitudinal and elevation gradient with diverse land types.

Our dimensionality estimates represent the subspace for which the average variability in the signal falls above that of noise for the scene as a whole. This approach produces a stable, repeatable estimate that is appropriate for many global mapping applications, though it may understate the number of spectra that could be distinguished in practice. For instance, techniques such as spatial averaging could recover signal from features with a magnitude below the spatially-random noise, provided they appear in large contiguous areas. At the other extreme, many eigenvector projections show small compact outlier features which are statistically significant but subtend a small number of pixels, contribute less to the total standard deviation along the eigenvector, and do not count toward the spectral dimensionality by our working definition. Even if there were no spatially-compact features, the covariance structure could underrepresent narrowband spectral diversity that appeared rarely in the dataset as a whole. Our use of second-order statistics was appropriately generic and well-defined, but should be considered a lower bound on spectral diversity. Covariance analyses would not substitute for task-specific degree of freedom or information content studies. Future investigations could investigate how dimensionality varies across spatial scale and spectral resolution, or after atmospheric effects are removed by cloud screening [25] and atmospheric correction [5, 26]. AVIRIS-C is a good place to begin such a survey, since it is similar to many existing aircraft and anticipated orbital instruments.

Overall, the study echoes prior findings on the astonishingly high diversity of VSWIR spectra, with dozens of degrees of freedom — far exceeding many other remote sensing modalities in current use. In other words, much contemporary remote sensing practice lies in a spectral sampling domain where instruments and algorithms capture a *shadow* of natural variability. This low-dimensional projection is underdetermined, so explanatory power often relies on modeling assumptions that the measurement data cannot test directly. In contrast, spectroscopic analyses enabled by instruments like AVIRIS-C provide an overdetermined measurement with contiguous spectral sampling over the VSWIR range. This is highly desirable since it offers numerical leverage while also providing the capability to measure unexpected phenomena and falsify modeling assumptions. Spectroscopy provides the opportunity for investigation within the dataset itself, exploiting the rich physical interpretability of the many-channel measurement to discover rich information manifest in the VSWIR light field.

## **Acknowledgments**

AVIRIS radiance spectra are available at <http://aviris.jpl.nasa.gov>. This research was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. We thank the JPL AVIRIS team. Copyright 2017 California Institute of Technology. All Rights Reserved.

## **Funding**

We acknowledge the financial support of the NASA Earth Science Division and the Jet Propulsion Laboratory Earth System Science Formulation office.