

Intrinsic Dimensionality as a Metric for the Impact of Mission Design Parameters

Kerry Cawse-Nicholson¹, Ann Raiho², David Ray Thompson¹, Glynn Hulley¹, Charles E. Miller³, Kimberley Miner³, Benjamin Poulter⁴, David Schimel³, Fabian Schneider⁵, Philip A Townsend⁶, and Shannon-Kian Zareh¹

¹Jet Propulsion Laboratory, California Institute of Technology

²NASA Goddard Space Flight Center

³Jet Propulsion Laboratory

⁴NASA

⁵California Institute of Technology

⁶University of Wisconsin-Madison

November 21, 2022

Abstract

High-resolution space-based spectral imaging of the Earth's surface delivers critical information for monitoring changes in the Earth system as well as resource management and utilization. Orbiting spectrometers are built according to multiple design parameters, including ground sampling distance (GSD), spectral resolution, temporal resolution, and signal-to-noise. The different applications drive divergent instrument designs, so optimization for wide-reaching missions is complex. The Surface Biology and Geology component of NASA's Earth System Observatory addresses science questions and meets applications needs across diverse fields, including terrestrial and aquatic ecosystems, natural disasters, and the cryosphere. The algorithms required to generate the geophysical variables from the observed spectral imagery each have their own inherent dependencies and sensitivities, and weighting these objectively is challenging. Here, we introduce intrinsic dimensionality (ID), a measure of information content, as an applications-agnostic, data-driven metric to quantify performance sensitivity to various design parameters. ID is computed through the analysis of the eigenvalues of the image covariance matrix, and can be thought of as the number of significant principal components. This metric is extremely powerful for quantifying the information content in high-dimensional data, such as spectrally resolved radiances and their changes over space and time. We find that the intrinsic dimensionality decreases for coarser GSD, decreased spectral resolution and range, less frequent acquisitions, and lower signal-to-noise levels. This decrease in information content has implications for all derived products. Intrinsic dimensionality is simple to compute, providing a single quantitative standard to evaluate combinations of design parameters, irrespective of higher-level algorithms, products, applications, or disciplines.

Hosted file

essoar.10510874.1.docx available at <https://authorea.com/users/532018/articles/597796-intrinsic-dimensionality-as-a-metric-for-the-impact-of-mission-design-parameters>

K. Cawse-Nicholson¹, A.M. Raiho², D.R. Thompson¹, G.C. Hulley¹, C.E. Miller¹, K.R. Miner¹, B. Poulter², D. Schimel¹, F.D. Schneider¹, P.A. Townsend^{1,3}, S.K. Zareh¹

¹Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA.

²NASA Goddard Space Flight Center, Biospheric Sciences Lab, Greenbelt, MD, USA.

³University of Wisconsin, Madison, WI, USA.

Corresponding author: Kerry Cawse-Nicholson (Kerry-anne.cawse-nicholson@jpl.nasa.gov)

Key Points:

- Intrinsic dimensionality is an applications-agnostic metric that can evaluate imaging spectroscopy instrument and mission design
- The principles of intrinsic dimensionality can be extended to the time domain for understanding the impact of revisit frequency
- Spatial, spectral, temporal resolution, and noise directly impact intrinsic dimensionality for a wide range of Earth surface scenes

Abstract

High-resolution space-based spectral imaging of the Earth's surface delivers critical information for monitoring changes in the Earth system as well as resource management and utilization. Orbiting spectrometers are built according to multiple design parameters, including ground sampling distance (GSD), spectral resolution, temporal resolution, and signal-to-noise. The different applications drive divergent instrument designs, so optimization for wide-reaching missions is complex. The Surface Biology and Geology component of NASA's Earth System Observatory addresses science questions and meets applications needs across diverse fields, including terrestrial and aquatic ecosystems, natural disasters, and the cryosphere. The algorithms required to generate the geophysical variables from the observed spectral imagery each have their own inherent dependencies and sensitivities, and weighting these objectively is challenging. Here, we introduce intrinsic dimensionality (ID), a measure of information content, as an applications-agnostic, data-driven metric to quantify performance sensitivity to various design parameters. ID is computed through the analysis of the eigenvalues of the image covariance matrix, and can be thought of as the number of significant principal components. This metric is extremely powerful for quantifying the information content in high-dimensional data, such as spectrally resolved radiances and their changes over space and time. We find that the intrinsic dimensionality decreases for coarser GSD, decreased spectral resolution and range, less frequent acquisitions, and lower signal-to-noise levels. This decrease in information content has implications for all derived products. Intrinsic dimensionality is simple to compute, providing a single quantitative

standard to evaluate combinations of design parameters, irrespective of higher-level algorithms, products, applications, or disciplines.

Plain Language Summary

We introduce intrinsic dimensionality (ID) as an objective, quantifiable metric for evaluating space-based mission design choices. We apply ID to the Surface Biology and Geology mission concept and show that increased ID correlates directly with decreased uncertainties in diverse science and applications products. We explore the challenge of balancing performance for many data products from highly dimensional spectral data. We conclude that ID is a valuable tool for optimizing mission performance and evaluating complex, multi-dimensional design choices.

1 Introduction

The ability to accurately measure and forecast the Earth system is increasingly important as climate change drives transformation across all scales (IPCC, 2021). In the last decade, various remote sensing satellite instruments have increased the range, scale, and impact of Earth observations, leading to important breakthroughs across Earth Science. The National Aeronautics and Space Administration (NASA) launched the Gravity Recovery and Climate Experiment (GRACE) missions to provide valuable data on water storage and glacier wasting (Landerer & Swenson, 2012; Schmidt et al., 2016; Landerer et al., 2020). The Orbiting Carbon Observatory (OCO) missions improve our understanding of the global carbon cycle (Crisp et al., 2004; Eldering et al., 2017), and the ECOSystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) provides critical metrics on plant heat stress and water uptake (Fisher et al., 2020; Pascolini-Campbell et al., 2021; Xiao et al., 2021). These missions were each optimized to measure a specific quantity or for a particular application that dictated all hardware and mission design decisions. In contrast, the next generation of NASA missions will be increasingly interdisciplinary, delivering a variety of measurements that touch on many science topics and aspects of the Earth System. For example, the NISAR mission will deliver products related to water availability, geologic hazards, and ecosystem biomass. Such missions cannot be traced to a single measurement objective, and when discipline requirements conflict, it is not clear how to make design decisions with finite mission resources. A consistent metric is necessary to understand the data needs and applications across diverse ecosystems.

As an example of how discipline needs may conflict with each other, consider the problem of orbit selection. A geostationary orbit provides the opportunity to continuously monitor ~30% of the Earth's surface in its field of view, but the ~30,000 km orbit altitude and constraints on optics size and mass effectively limits GSD; e.g. the latest GOES satellites have 500 m GSD in the visible, 1 km GSD in the near-infrared, and 2 km GSD in the thermal infrared. Polar-orbiting satellites acquire data at fine spatial scales (5-30 m) less frequently (5-16 day revisit). Thus, there is a trade-off between spatial and temporal resolution,

and these dynamics compete with challenges in available instrument technology. Determining the optimal satellite revisit and resolution is often dependent on the application. Better integration of this great diversity of measurements would benefit the state of the science.

We address this problem through the lens of intrinsic dimensionality, a universal performance metric, and its implications for an important new NASA mission: the Surface Biology and Geology (SBG) investigation. SBG will employ high GSD hyperspectral visible to shortwave infrared (VSWIR) and multi-spectral thermal infrared (TIR) imagery to meet diverse science and applications needs. In theory, each of these observations is constrained by its own set of high-priority parameters (Cawse-Nicholson et al., 2021). For instance, geologists might be more interested in high spectral resolution than frequent revisit (Swayze et al., 2003). Aquatic scientists might be more interested in minimal instrument noise over dark targets than <100 m GSD (Dierssen et al., 2021), and volcanologists might prioritize revisit frequency (Francis & Rothery, 2000). As the measurement science advances, instruments are also drafted into roles beyond their original design purpose, such as the use of airborne VSWIR instruments to monitor greenhouse gas point sources (Duren et al., 2019). Rather than trying to arbitrate between all possible measurement objectives – an impossible and ever evolving task – we propose an alternative standard of instrument performance related to the information captured from the upwelling light field. Specifically, we show how the intrinsic dimensionality, which quantifies the number of measurable free parameters in a dataset, is a reasonable and objective standard of SBG observing system performance that broadly captures diverse science and application needs.

2 Intrinsic Dimensionality

Intrinsic Dimensionality (ID) is a measure of information content that has been used to determine components of chemical mixtures from hyperspectral data (Kritchman & Nadler, 2008), retrieve speech signals from noisy audio (Park et al., 1999), measure the impact of data fusion (Cawse-Nicholson et al., 2019), and determine unique classes in remotely sensed imagery (Boardman & Green, 2000; Small, 2001; Asner et al., 2012; Thompson et al., 2017). ID quantifies the number of significant principal components. If a hyperspectral image X , of size $n \times b$, where n is the number of pixels and b is the number of spectral bands, can be written as a linear combination of signal (S) and noise (N), such that $X = S + N$, then the ID is the rank of the signal subspace (i.e. the maximum number of linearly independent components).

We assume that each pixel $x_i \in X$ can be written as a linear combination of endmembers v_j (objects contained within the scene with unique spectral properties), such that

$$x_i = \sum_{j=1}^K a_{ij} v_j + \varepsilon_i \quad \#(1)$$

where $\varepsilon_i \in N$ is the noise present in the i^{th} pixel, a_{ij} is the abundance of the j^{th} endmember in the i^{th} pixel, and K is the ID. Here, the noise encapsulates all causes of uncertainty that might cause a deviation from the true radiance, including photon shot noise, electronic noise, read noise, dark current noise, quantization noise, and other calibration uncertainties. In natural ecosystems, there is some inherent variability within classes, representing for example, different nutrient levels within trees of the same species. In such cases, the endmember can be thought of as the center or most representative spectrum from a class, and highly variable species may be represented by multiple classes. Natural variation is distinguished from noise by the assumption that naturally varying spectra within a class will, for the most part, change smoothly across the electromagnetic spectrum, whereas the noise is random between spectral channels.

Several methods have been proposed to estimate K in equation (1), which is the ID (e.g., Chang & Du, 2004; Bachman et al., 2008; Bioucas-Dias & Nascimento, 2008; Cawse-Nicholson et al., 2012; Berman, 2019). We use Random Matrix Theory (RMT) to estimate ID (Cawse-Nicholson et al., 2012; 2019). If the noise is assumed to be Gaussian, we can construct a “random matrix” $\Sigma = N^T N$, and the eigenvalues of these matrices have been studied (Johnstone, 2001; Baik & Silverstein, 2006). While it is possible that there are non-Gaussian contributors to the noise term, the Gaussian assumption is common and has been shown to be sufficient in practice (Bioucas-Dias & Nascimento, 2008; Cawse-Nicholson et al., 2012). Assuming scaled and centered data, we will consider the eigenvalues λ_k of the image covariance matrix $C = X^T X$ and those eigenvalues that behave like those of a random matrix will be considered due to noise. An eigenvalue is regarded as signal if

$$\lambda_k > \rho_k c(n, b) \#(2)$$

$$\rho_k = \frac{e_{\Pi}^{kT} \Sigma e_C^k}{e_{\Pi}^{kT} e_C^k} \#(3)$$

where $\Pi = C - \Sigma$, and e_{Π}^k and e_C^k are the k^{th} eigenvectors of Π and C , respectively. The constant, c is dependent on the number of pixels and the number of channels and has been fully defined by Cawse-Nicholson et al. (2012). The noise covariance matrix Σ should be estimated from the data itself. Here we use the multiple regression approach used by Bioucas-Dias & Nascimento (2008). Since inherent spectral correlation between channels implies that the signal spectrum is a smoothly varying function, the noise-free spectrum in each spectral band can be estimated from its neighbors. The difference between the estimated and observed values is assumed to be noise. Where such correlation does not exist, we consider Meer’s method (Cawse-Nicholson et al., 2012), which uses the variance of small, homogeneous regions in images to estimate the noise in each band. In this case, the noise covariance matrix will be diagonal.

Each observed pixel has a general shape that represents the background continuum, which varies by scene as a function of the incoming solar spectrum, albedo, scattering processes, and black-body emission. Within this continuum, unique spectral signatures identify each object or class within the image. These are often absorption features that may have small amplitude and/or cover a small wavelength range. High noise levels will obscure these features to the point where two similar classes are no longer separable. Similarly, wider spectral channels will not differentiate between two classes with absorption features at similar wavelengths. Also, RMT assumes that each pixel is a mixed pixel, but when the spectral contribution of a class to the overall pixel spectrum is small enough to fall within the noise levels, then that class will no longer be detectable. This means that small objects will no longer be detected with coarse GSD. Since ID is the number of unique, discernible classes in a hyperspectral image. We can use this metric to evaluate how different instrument design parameters, such as spectral resolution, GSD, and signal-to-noise ratio (SNR), impact the measurable signal content.

3 Data and Tools

We have considered several real datasets in order to compute ID. For GSD, spectral resolution, and SNR, airborne hyperspectral datasets provide a good test environment for the simulation of coarser spaceborne imagery, and these are described in Section 3.1. The simulation environment that allows us to resample the relevant parameters is described in Section 3.2. We currently lack a hyperspectral dataset with high temporal resolution, so this parameter is considered separately using multiband daily satellite data, described in Section 3.3.

3.1 Ground Sampling Distance, Spectral Resolution, and SNR

To evaluate sensitivity to GSD, spectral resolution (bandwidth), and signal to noise ratio, we considered airborne hyperspectral imagery over multiple biomes, including Chaparral, Desert, Mangrove, Temperate Forest, Dry Tropical Forest, Grasslands, Tundra, Boreal Forest, Coral Reefs, Agriculture, and Urban environments. See Figures 1 and 2. We assume contiguous spectral coverage with a fixed range of 400 – 2500 nm.

We use actual spectra from three airborne hyperspectral instruments to represent realistic spatial, spectral, and noise characteristics. The Airborne Visible InfraRed Imaging Spectrometer - Next Generation (AVIRIS-NG) is an airborne spectrometer that acquires data in the spectral range 380 – 2510 nm; the orthorectified and atmospherically corrected reflectance product was obtained from <https://avirisng.jpl.nasa.gov/dataportal/>. The National Ecological Observatory Network (NEON) Airborne Observing Platforms acquire data in the spectral range 380 – 2500 nm using spectrometers closely related to AVIRIS-NG; the orthorectified and atmospherically corrected reflectance product was obtained from <https://data.neonscience.org/data-products/DP1.30006.001> (NEON, 2021). The Portable Remote Imaging Spectrometer (PRISM) is an air-

borne spectrometer that acquires data in the spectral range 350 – 1050 nm; the orthorectified and atmospherically corrected reflectance product was obtained from <https://prism.jpl.nasa.gov/dataportal/>. PRISM has a spectral resolution of 2.83 nm, and the AVIRIS-NG and NEON spectrometers have a spectral resolution of 5nm. These spectrometers, along with the upcoming Earth Mineral dust source Investigation (EMIT) are described in Table 1.

A wide range of scenes was considered over different landcover types, spatial regions, and seasons to encapsulate a wide range of conditions. Each scene is described in Table 2. As a default, each image is resampled to 30 m GSD using bilinear interpolation, and the spectral bands are resampled to 10 nm spectral channels unless specified otherwise. Bands that may be impacted by water vapor or other artifacts (< 400 nm, 1260 – 1560 nm, 1760 – 1960 nm, and > 2450 nm) were excluded before computing the ID.

Table 1: AVIRIS-NG, NEON, PRISM, EMIT and SBG instrument performance parameters.

Instrument	Spectral Range	Spectral Resolution	Spectral FWHM	Number of B
AVIRIS-NG	380 – 2510 nm	5 nm	5 nm	425
NEON	380 – 2500 nm	5 nm	7.5 nm	426
PRISM	350 – 1050 nm	2.8 nm	3.5 nm	92
EMIT	380 – 2500 nm	7.4 nm	8.5 nm	288
SBG-VSWIR (notional)	400 – 2500 nm	5-10 nm	TBD	TBD

While most scenes were downloaded in their publicly available form, two scenes were specially processed: the desert scene in Nevada and the dry tropical forest in India. These were atmospherically corrected using the Imaging Spectrometer Optimal FITting (ISOFIT) tool (Thompson et al., 2018) and manually evaluated for the best correction. This was done to ensure that no residual artifacts were propagated through the SNR simulations. ISOFIT is publicly available via <https://github.com/isofit/isofit>.

The SBG mission architecture described in the 2017 Earth Science and Applications from Space Decadal Survey (ESAS, 2017) consists of high-GSD visible-to-shortwave infrared hyperspectral and multi-channel thermal infrared imagers. The AVIRIS-NG and NEON datasets are very good proxies for the VSWIR spectrometer and will be used to evaluate GSD, spectral resolution, and SNR. Because the thermal imager will be a multiband (< 10 channels) instrument, spectral dimensionality is not an appropriate metric. However, the thermal instrument will be considered in the temporal resolution assessment in Section 4.

Table 2: AVIRIS-NG, NEON, and PRISM images used to evaluate intrinsic dimensionality.

Biome	Location	# bands	GSD	Sensor	Date of acquisition
Chaparral	California	432	3.7 m	AVIRIS-NG	2014-06-03
Desert*	Nevada	432	3.7 m	AVIRIS-NG	2014-06-25
Mangrove	Louisiana	432	3.8 m	AVIRIS-NG	2015-05-09
Temperate Forest	Wisconsin	432	4.4 m	AVIRIS-NG	2015-09-04
Dry Tropical Forest*	India	425	3.6 m	AVIRIS-NG	2016-01-07
Grasslands	Oklahoma	425	3.1 m	AVIRIS-NG	2017-06-14
Tundra	Alaska	425	5.3 m	AVIRIS-NG	2018-07-29
Boreal Forest	Canada	425	5.1 m	AVIRIS-NG	2018-08-11
Wetlands	Florida (OSBS)	426	1 m	NEON	2019-04-16
Mixed Forest	Michigan (UNDE)	426	1 m	NEON	2019-06-06
Grasslands	Kansas (KONZ)	426	1 m	NEON	2019-07-12
Evergreen Forest	Alabama (DELA)	426	1 m	NEON	2019-04-29
Mixed Forest	California (SOAP)	426	1 m	NEON	2019-06-14
Mixed Forest	California (TEAK)	426	1 m	NEON	2019-06-14
Coral Reef	Australia	92	7.9 m	PRISM	2016-09-17
Coral Reef	Hawaii	92	7.8 m	PRISM	2017-03-06
Coral Reef	Palau	92	8 m	PRISM	2017-05-16

*These two scenes were processed to reflectance separately using the optimal estimation algorithm in ISOFIT (Thompson et al., 2018) and evaluated manually to ensure the best atmospheric correction.

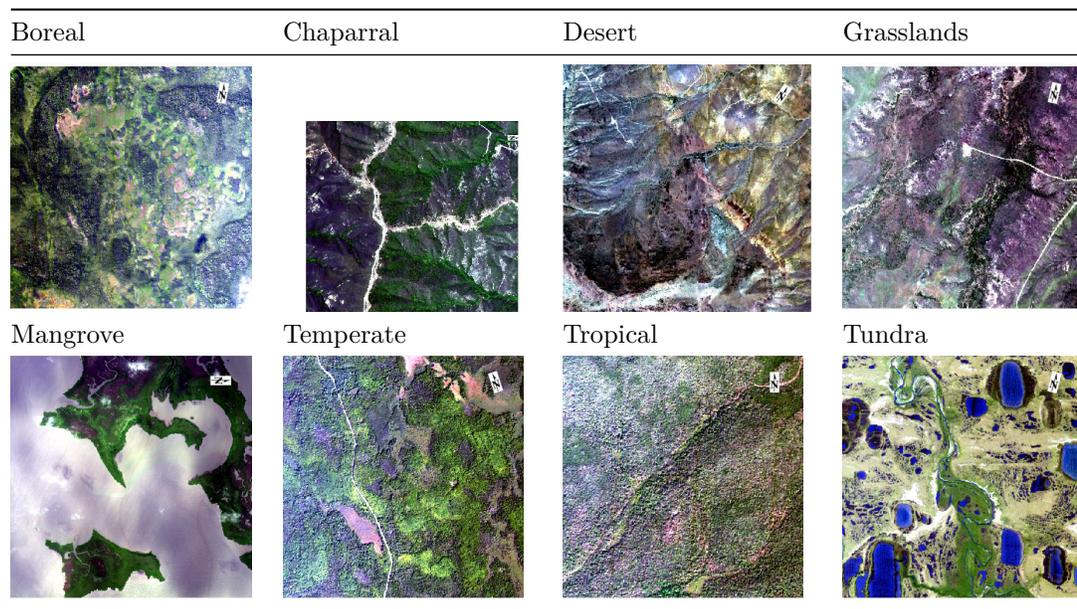


Figure 1: RGB composites of representative AVIRIS-NG images.

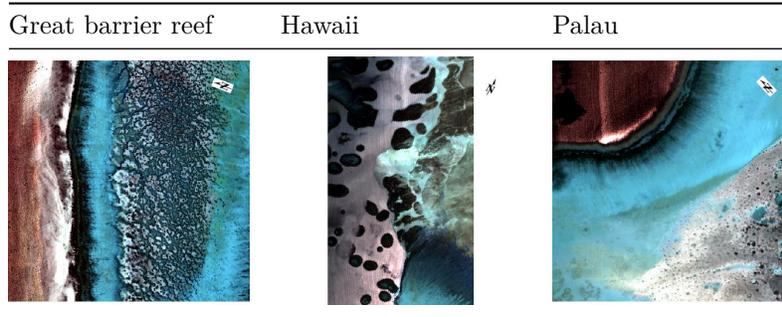


Figure 2: RGB composites of representative PRISM coastal ocean images.

3.2 Simulation environment

The Py-Hypertrace environment enables users to simulate accurate spectral imagery given an input a reflectance (“truth”) image. The forward model uses sRTMnet (Brodrick et al., 2021) to simulate sensor noise, view and sun geometry, and atmospheric spectral signatures. The inverse model uses ISOFIT to retrieve the atmospheric parameters and surface reflectance simultaneously, and the difference between the true and estimated reflectance images provides an estimate of accuracy. Py-Hypertrace is available through the ISOFIT GitHub repository (<https://github.com/isofit/isofit/tree/master/examples/py-hypertrace>).

Preliminary testing showed that the spatial and spectral resolution experiments were unchanged when running through ISOFIT or on the original reflectance product, so Hypertrace was only used for the SNR experiment. In addition, some publicly available reflectance products had minor spectral artifacts from atmospheric correction. While these are often inconsequential, in this case, adding atmosphere and noise enhanced artifacts and often dominated the signal. Because of this, only the two scenes with negligible artifacts (Nevada and India) were used for the SNR experiment.

By default, our Hypertrace runs assumed a “mid-latitude summer” atmospheric profile (a standard MODTRAN atmosphere defined in Anderson et al., 1986), an aerosol optical depth of 0.1, an atmospheric water vapor content of 1 cm, a nadir viewing angle, and a 1000 local solar time (solar zenith angle of 37.21° and solar azimuthal angle of 53.82°). These atmospheric parameters are fixed in the forward model, and the inverse model independently solves for the aerosol optical depth and atmospheric water vapor content in addition to the surface reflectance profile.

Realistic SNR was simulated using an instrument model that considers detector design parameters to account for different noise sources, including dark current,

electronic readout noise, etc. It also includes precise calculation of photon shot noise based on efficiencies of the instrument components, internal reflections off telescope surfaces, throughput loss due to grating, dark current, electronic readout, etc. (described in detail in Thompson et al., 2020). The noise is signal- and wavelength-dependent and is drawn from a multivariate Gaussian distribution defined by three parameters, such that the noise-equivalent change in radiance $\sigma_l = a \sqrt{b \bullet l_{\text{obs}}} + c$, where l_{obs} is the observed radiance at the center wavelength of each spectral band (Thompson et al., 2020). The parameters we used were modeled for the EMIT instrument, a spectrometer planned to launch in 2022 and will serve as a precursor instrument to SBG (Connelly et al., 2021). EMIT is also closely related to AVIRIS-NG and the NEON instruments in terms of spectral resolution, range, number of channels, and SNR (see Table 1).

3.3 Temporal Resolution

While the GSD, spectral resolution, and SNR could be evaluated using existing airborne hyperspectral datasets, there is currently insufficient hyperspectral data that is sampled frequently enough (less than 5 day revisit over at least one annual cycle) for a sensitivity study of temporal resolution. Instead, we have used daily multispectral data as a proxy. In this case, the image matrix is of size $n \times t$, where t is the number of temporal observations, and the image covariance matrix will be size $t \times t$. Spectral dimensionality has been well established as a metric in other applications, but to our knowledge, this is the first use of temporal dimensionality. Instead of determining unique classes by their spectral features, we are now defining classes by phenological patterns. For example, while the normalized difference vegetation index (NDVI) over deserts might remain essentially unchanged, the NDVI of forests will fluctuate according to seasonal change, and different crops will be separable by their planting and harvest times. With fewer observations, all crops planted within a single season might be grouped, but frequent observations would result in higher information content. The temporal patterns seen in multispectral data have already been incorporated into the classification of agricultural and unmanaged landscapes (e.g., Zhang et al., 2003; Lobell & Asner, 2004; Sakamoto et al., 2005; Chang et al., 2007; Foerster et al., 2012).

The Visible Infrared Imaging Radiometer Suite (VIIRS) acquires daily optical data globally in 22 spectral channels at two GSDs: 375 m and 750 m. These were resampled to 1 km and 0.05 degrees for consistency with MODIS heritage, and five bands (bands I1-5) were resampled to 500 m GSD. This daily global dataset forms a valuable basis for studying the impact of revisit time on information content. We used two VIIRS standard products in this analysis: the VNP43IA4.001 nadir BRDF-adjusted daily reflectance product at 500 m (Schaaf et al., 2018); and the VNP21A1D.001 land surface temperature and emissivity daytime product at 1 km (Hulley & Hook, 2018). The red high-resolution reflectance band I1 was used to evaluate temporal dimensionality in the visible part of the spectrum. Bands I1 (red) and I2 (infrared) from the reflectance product were used to produce a normalized difference vegetation index (NDVI), such

that $NDVI = (I2 - I1)/(I2 + I1)$. The daily land surface temperature (LST) product was used to evaluate temporal dimensionality in the thermal infrared.

To evaluate temporal dimensionality, we evaluated eleven sites in highly biodiverse regions (Myers et al., 2000), as well as two agricultural regions, described in Table 3 and shown in Figure 7. All products were cropped to the area of interest and downloaded using AppEEARS (AppEEARS, 2021).

Table 3: VIIRS sites used to evaluate temporal dimensionality.

Biome	Location	Site Size (km ²)	*Date of Acquisition
Southwest Australia	Australia	48,000	Jan – Dec 2020
Amazon Forest	Brazil	34,000	Jan – Dec 2020
Caucasus	Bulgaria	20,000	Jan – Dec 2020
Boreal Forest	Canada	28,000	Jan – Dec 2020
Tropical Forest	Democratic Republic of Congo	51,000	Jan – Dec 2020
Eastern Afromontane	Ethiopia	41,000	Jan – Dec 2020
Mediterranean Basin	Portugal	42,000	Jan – Dec 2020
Boreal Forest	Russia	27,000	Jan – Dec 2020
Cape Floristic Region	South Africa	5,000	Jan – Dec 2020
Indo-Burma Biodiversity Hotspot	Thailand	112,000	Jan – Dec 2020
Sierra Nevada	USA (California)	14,000	Jan – Dec 2020
Agriculture	USA (California)	12,000	Jan – Dec 2020
Agriculture	USA (Iowa)	32,000	Jan – Dec 2020

* Daily NDVI and LST were computed for each site.

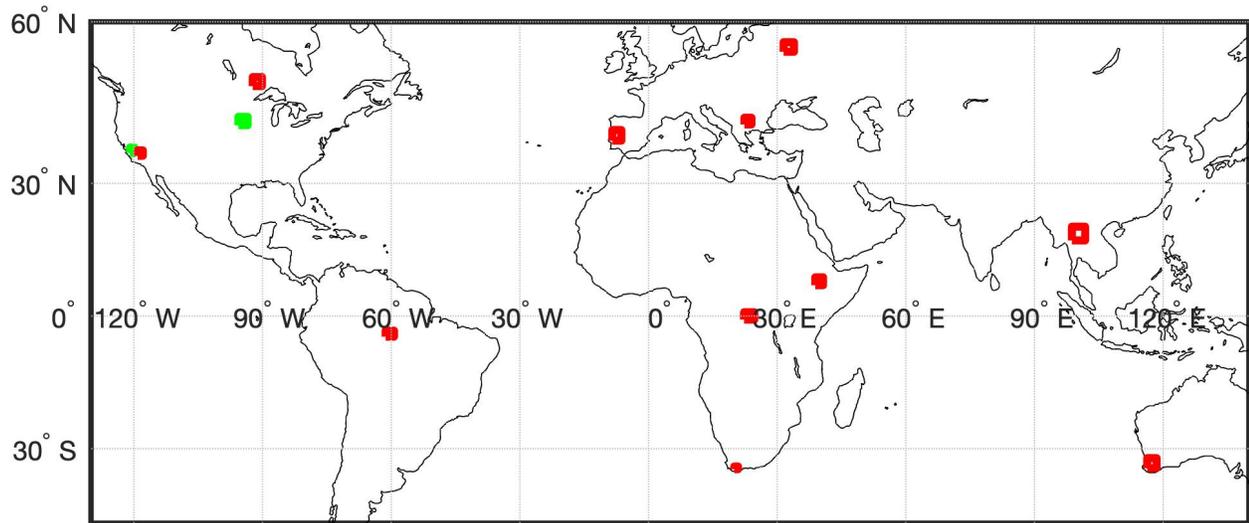


Figure 7: The biodiversity hotspots (red) and agricultural regions (green) selected for temporal dimensionality evaluation. See Table 3.

The VIIRS nadir BRDF-adjusted daily reflectance product is accompanied by quality flags that indicate the complete inversion’s success and quantify the impact of clouds. The daytime LST product is also accompanied by quality flags that indicate clouds, calibration quality, and algorithm convergence speed. Again, only the best quality pixels were used. For both products, pixels containing fill values were also discarded.

Each image contained between 20,000 and 500,000 pixels (each manually selected to capture landcover heterogeneity). For each pixel, a time series was constructed for reflectance band I1 and LST, building up to three-dimensional image stacks that are similar in format to a hyperspectral image. Images at a particular timestamp were removed from analysis where more than 20% of the image was masked due to cloud or other flags indicated lower quality data. The resulting image and noise estimates were used as inputs in the Random Matrix Theory method to determine the temporal dimensionality. This method was repeated for different temporal sampling strategies. The image stack was sampled at regular intervals to simulate different revisit frequencies (i.e., to simulate a 16-day revisit cycle, only every 16th observation was used to compute the temporal dimensionality). The increasing temporal gap between observations necessitated Meer’s method to estimate the noise since the statistical methods rely on a high correlation between neighboring observations to compute the noise contribution.

To demonstrate that temporal dimensionality translates to real quantities, we compared it to estimates of peak greenness over the California agricultural site. To derive peak greenness, we compiled a temporal stack of NDVI, as described for I1 and LST above. Five hundred pixels were randomly selected for analysis, and each pixel was processed to better represent the seasonal growing cycle. This was repeated for five different start times to reduce the effect of revisit periodicity linking up with a seasonal peak, yielding 2,500-time series vectors. First, the pixel was subsampled to the revisit period under consideration; any gaps were filled using forward and reverse autoregressive fits from the remaining time series; outliers were removed, where outliers were defined as points more than three standard deviations from the mean of a 5-point moving window; and a squared exponential curve approximation was used to represent the time series while removing unexpected drops in NDVI due to undetected cloud. Finally, the maximum value of the squared exponential curve is set as the peak greenness.

4 Experiments and Results

4.1 Ground Sampling Distance

To quantify the impact of ground sampling distance (GSD) on ID, we evaluated the ID for all scenes listed in Table 2, resampled to GSDs of 5 m, 10 m, 20 m, 30 m, 40 m, 50 m, and 60 m. In each case, the original image was spatially resampled using bilinear interpolation, and all images were resampled to a fixed spectral resolution of 10 nm. No additional noise was added to the original reflectance images, which means that the coarser resolution images have a higher

SNR due to the averaging effect. This is in line with the assumption that a real instrument with larger GSD will have longer integration time per pixel and therefore a higher SNR. This is discussed further in Raiho et al. (2022), and explored in more detail in further experiments described below. Since the atmospheric correction is run pixel-by-pixel, there was no need to include additional instrument or atmospheric parameters beyond the original reflectance product.

We calculated the ID for each scene from Table 2 at every GSD using RMT and the multiple regression noise estimation described in Section 2. For each scene, the ID was normalized to the ID calculated for the GSD = 5 m case. Figure 3 shows that the normalized IDs (NID) decrease as GSD increases, but there is a significant variation at coarser GSD since heterogeneous scenes are more sensitive to GSD than homogeneous scenes. For example, images from an instrument with a 60 m GSD will have roughly 30% lower information content than images from an instrument with a 30 m GSD.

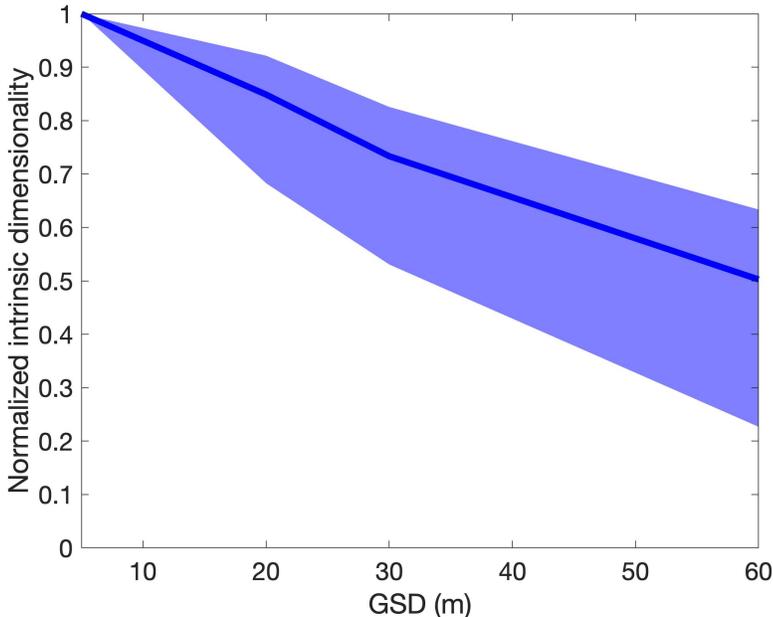


Figure 3: ID decreases with larger pixel sizes. The thick line shows the median ID across all 17 scenes, and the shaded area encompasses the 25th to the 75th percentile of normalized IDs across all scenes.

We recognize that the bilinear interpolation we used to resample our images spatially might average out some of the noise, thereby conflating two variables. To account for this, we estimated the noise present in the image (already done as part of the ID computation). We added Gaussian noise so that the overall standard deviation of noise remained similar for each image. One image, in particular, was an outlier to Figure 3. The NEON evergreen forest site DELA had ID values of 8, 7, 8, 11 for GSDs of 30 m, 40 m, 50 m, 60 m. We hypothesized

that a higher ID at 60 m (contrary to Figure 3) was due to the noise suppression of the image resampling. When we included additional noise, the ID values became 8, 7, 7, 7 for GSDs of 30 m, 40 m, 50 m, 60 m. This progression is more in line with the other datasets. More telling is that the other datasets captured in the 25th – 75th percentile shading shown in Figure 3 did not change ID values when noise consistency was ensured. We conclude that images that are especially sensitive to noise may behave differently, but overall the ID analysis of resampled reflectance imagery was robust.

4.2 Spectral Resolution

To quantify the impact of spectral resolution on ID, we evaluated the ID for all scenes listed in Table 2 spectrally resampled to 5 nm, 15 nm, 20 nm, 25 nm, and 30 nm. In each case, the original image was spectrally resampled using bilinear interpolation, and all images were resampled to a fixed GSD of 30 m. No additional noise was added to the original reflectance images.

Using RMT and the multiple regression noise estimation described in Section 2, we calculated the ID for each scene at each spectral resolution. For each scene, the ID was normalized to the ID calculated for the spectral resolution = 5 nm case. Figure 4 shows that the scaled ID decreases as spectral resolution varies from 5 to 30 nm, with a relatively steep drop-off from 5 nm and less variation than seen in Figure 3. In this case, changing the spectral resolution could also impact the atmospheric correction by affecting the ability to measure spectrally-sharp water vapor absorption features. We tested this by running Hypertrace simulations on two scenes (Nevada and India) and comparing the results to the spectral resampling performed on the original reflectance data. Figure 4 shows that coarser spectral resolution does impact the atmospheric correction, resulting in slightly lower ID values, but the overall shape remains comparable. There is an information loss of ~40% moving from 10 nm to 20 nm spectral resolution.

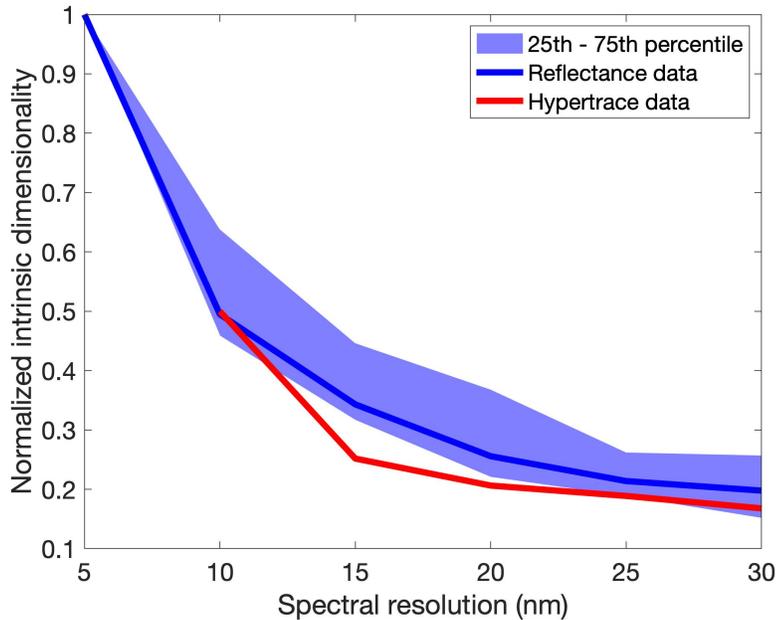


Figure 4: Normalized ID decreases with coarser spectral resolution. The thick blue line shows the median ID across all scenes, and the shaded area encompasses the 25th to the 75th percentile of scaled ID across all scenes. The red curve (the average from the Nevada and India scenes) shows that coarser spectral resolution negatively impacts ID when considering atmospheric correction, but this effect is negligible.

4.3 Signal-to-Noise Ratio

Unlike the GSD and spectral resolution experiments, realistic SNR could not be easily simulated at the reflectance level. Therefore, we used Hypertrace to simulate realistic radiance data using the atmospheric, instrument, and noise models and computed ID on the reflectance output using ISOFIT. SNR, in this case, was set by adjusting the integration time within the noise model – higher integration times meant increased signal and therefore a lower SNR, and vice versa.

As discussed in Section 3.1, only the Nevada and India scenes were used for the Hypertrace experiment to minimize the propagation of spectral artifacts. These images were resampled to 30 m GSD and 10 nm spectral resolution. We varied SNR by varying integration times in the simulation. We used integration times of 0.0015 s, 0.003 s, 0.0045 s, 0.006 s, 0.0075 s, 0.009 s which corresponded to average SNR values of approximately 170, 250, 300, 350, 390, and 430, respectively. Figure 5 shows that lower ID values are detected in noisier scenes since distinguishing spectral features may be obscured by high noise variance. Figure 6 shows the SNR values for all wavelengths for each choice of integration time.

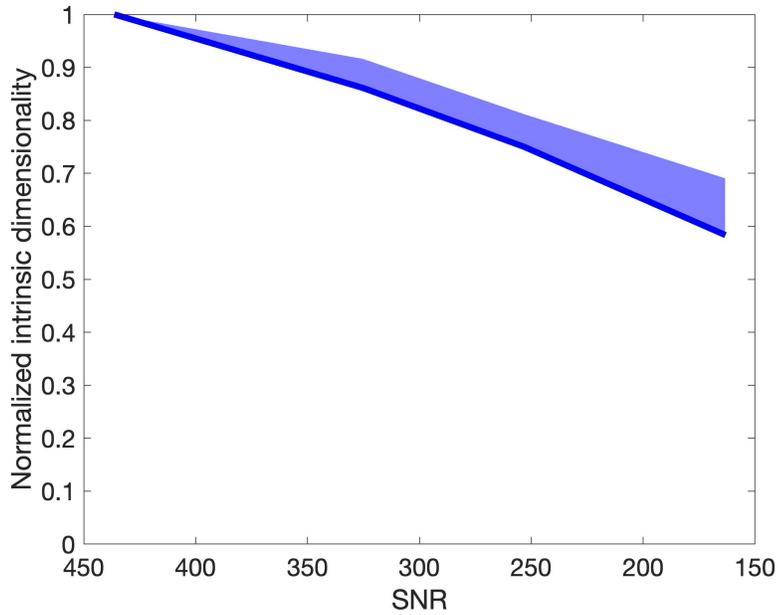


Figure 5: ID decreases as the SNR decreases. Since only two scenes were considered here, the shading encapsulates the range of scaled ID between the Nevada and India scenes.

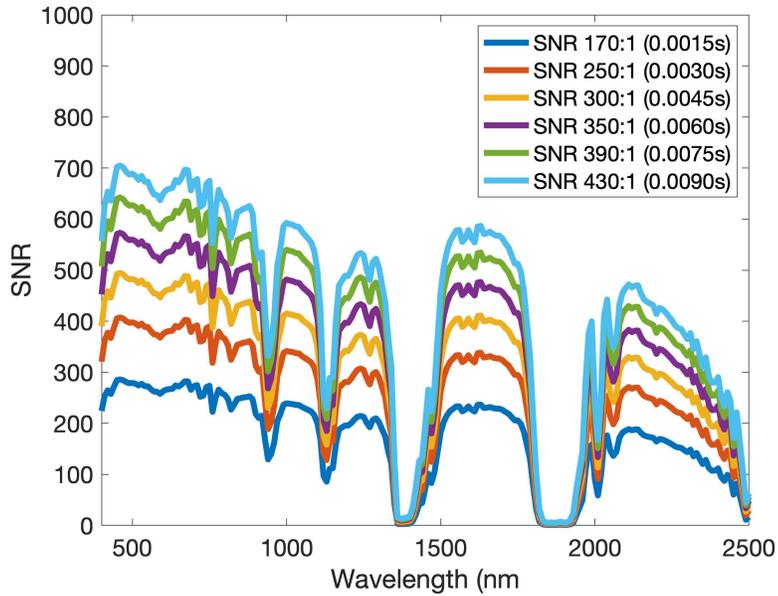


Figure 6: SNR varies by channel, decreasing at longer wavelengths due to both lower radiance levels and higher noise. Here the modeled SNR curve at each wavelength is shown for each choice of average

SNR (with associated integration time). Note that the intense atmospheric water absorptions (SNR = 0) near 1400 and 1900 nm effectively remove all light from the spectrum, providing an opportunity to measure detector and read noise in real spectra.

4.4 Temporal Resolution

The tropical forest sites in South America (Brazil) and Central Africa (DRC) and the Boreal Forest in Northern Russia were cloud contaminated to the point of algorithm failure. This will be a problem to consider for SBG and other missions that hope to acquire data in these regions. We anticipate that SBG will have better sampling statistics due to its 30 m GSD; however, cloud shadows and other interferences may still make these challenging regions. These sites were removed from further analysis, leaving eleven remaining sites. Other sites such as Canada, Ethiopia, and Thailand had ID values that dropped to zero with revisits exceeding 4-8 days (in other words, with weekly sampling there were insufficient pixels to track over time for the ID computation). Figure 8 shows the decrease in NDVI temporal dimensionality with increased revisit time by combining the temporal ID estimates for all sites with $ID > 0$. There is rapid information loss in the 1-5 day revisit range, which would capture weather-related phenological perturbations on synoptic time scales and separate classes with similar growth patterns. SBG anticipates a 16-day revisit for the VSWIR instrument, but there is potential for harmonization with temporally intersecting missions, such as the European Space Agency's Copernicus Hyperspectral Imaging Mission for the Environment (CHIME). We have shown the ID curve normalized (NID) by an 8-day baseline to illustrate the benefits to harmonization (Figure 8). Moving from a 16-day (NID ~ 0.60) to a 30-day revisit (NID ~ 0.36) would result in the loss of about 40% of the dimensionality, which means that fewer classes would be seasonally separable. In contrast, a harmonized SBG-CHIME data set that reduced the revisit time from 16 days (NID ~ 0.60) to 8 days (NID ~ 1.0) would increase the ID by 67%. This analysis is significantly impacted by clouds, which is a realistic problem that will be faced by all optical systems including SBG.

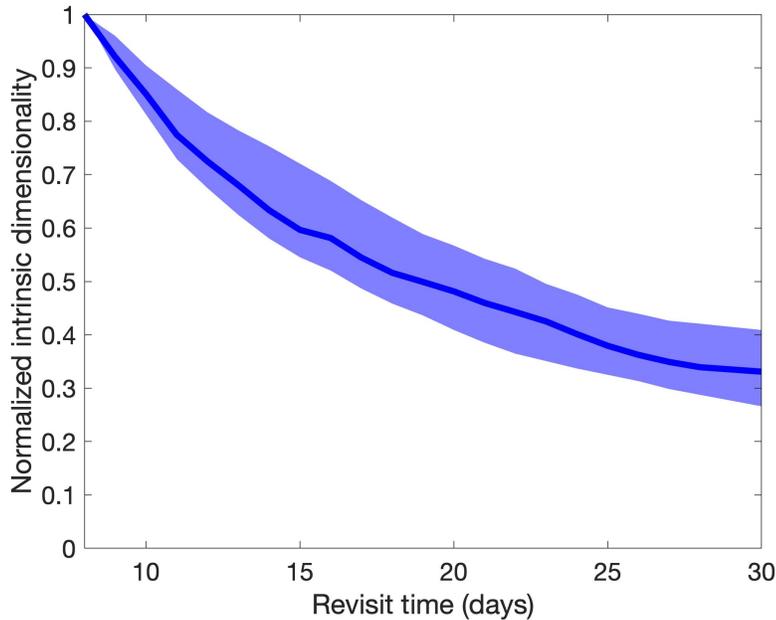


Figure 8: Temporal dimensionality of VIIRS band I1 normalized by the ID at 8-day revisit is shown as a function of revisit time. Going from a 16-day (proposed by SBG) to a 30-day revisit results in approximately 40% of the information content being lost. The solid line represents the 50th percentile of all sites, and the shading represents the 25th and 75th percentiles.

Figure 9 shows the results of a similar experiment for LST, except here the dimensionality normalization was performed against a 1-day revisit due to the intrinsic timescale of LST variability. This is reflected in the rapid decay of temporal dimensionality for the thermal regime compared to the VSWIR. As seen in the plot, the normalized ID decreases to ~ 0.65 going from 1-day to 2-day revisit, drops to ~ 0.40 for a 3-day revisit, and is less than 0.20 for a 6-day revisit. This is reasonable since evapotranspiration and land surface temperature change far more rapidly than vegetation greenness. There is potential for harmonization with the Indian Space Agency/French National Center for Space Studies Thermal infraRed Imaging Satellite for High-resolution Natural resource Assessment (TRISHNA) and the European Space Agency's Copernicus Land Surface Temperature Monitoring (LSTM) missions. Harmonization between the three thermal infrared missions could deliver an effective 1-day global revisit, significantly enhancing the science return over what any of the three missions could achieve individually.

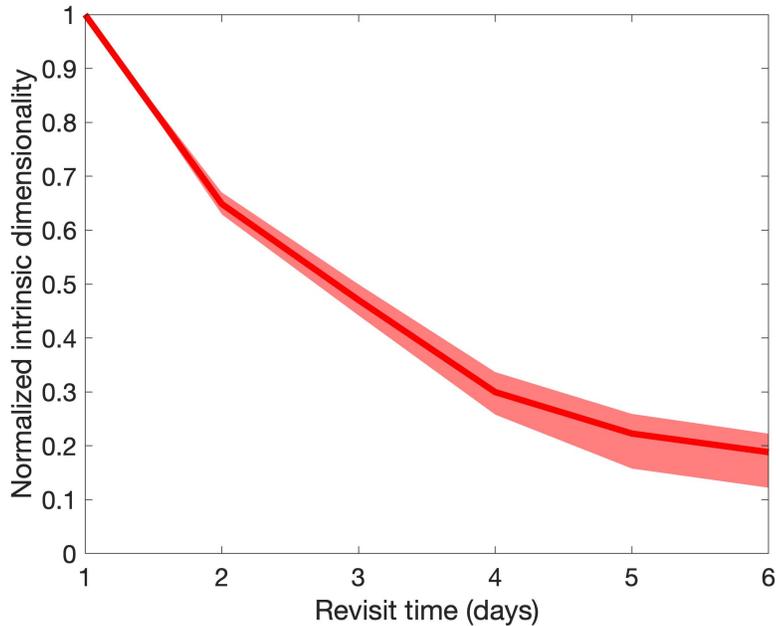


Figure 9: Temporal dimensionality of VIIRS daytime LST normalized by the ID at daily revisit is shown as a function of revisit time. There is a decrease in information content with increasing revisit time, with a faster decay than the visible. In going from a 3-day to 6-day revisit, approximately 60% of the information content is lost. The solid line represents the 50th percentile of all sites, and the shading represents the 25th and 75th percentiles.

Figure 10 shows that temporal dimensionality is a good proxy for ecosystem variables such as the NDVI value at peak greenness in their dependence on revisit. Estimates of maximum NDVI values can be biased by up to 5% with an increased revisit time. This suggests that peak NDVI is relatively slowly varying and is consistent with the ~1 week time scale on which NDVI varies significantly. The dimensionality shows a much larger decrease (85%) in information content because it represents more than the simple metric of peak greenness and is analogous to the sum of the impact faced by a wide array of more complex algorithms.

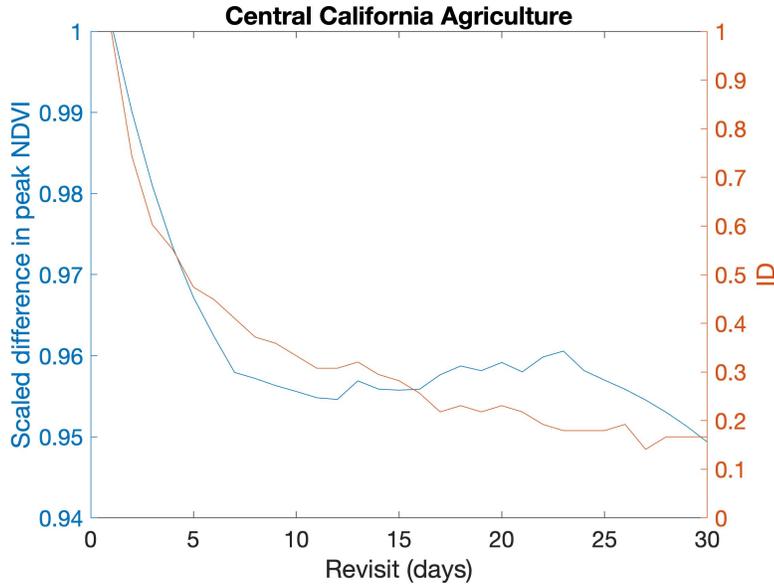


Figure 10: The difference in peak NDVI, scaled by peak NDVI at a daily revisit (left), tracks well with dimensionality for different revisit times (left).

5 Discussion

We illustrate intrinsic dimensionality as a metric for mission design, and we have shown how intrinsic dimensionality is sensitive to changes in GSD, spectral resolution, SNR, and temporal resolution. Our experiments show that ID consistently decreased for larger pixels, wider spectral channels, lower SNR, and less frequent observations even over multiple different biomes. All of these results are as expected since we intuitively understand that large pixels will contain mixtures of many – sometimes small – objects, and their contribution to the overall pixel spectrum will be small; wider spectral channels will mean that spectrally narrow diagnostic spectral features will be undetectable; noise will overpower small-amplitude spectral characteristics, and important events may not be adequately captured with infrequent acquisitions. While it might seem best to simply design a mission with the highest possible resolutions in all variables, realistically there are relationships and trade-offs between these variables. For instance, the same detector could be used in a design to improve SNR by increasing GSD (due to increased integration time), but both parameters cannot be maximized at the same time. Similarly, the same detector with a different telescope could acquire small GSD (and consequently smaller swath) with infrequency acquisitions, or a wider swath (and consequently larger GSD) more frequently. In order to balance these trade-offs, a consistent metric is needed. The importance of our investigation has been to demonstrate that ID is a consistent, objective, quantitative metric that can be used to measure expected performance over many mission design parameters and different

landcover types.

In a traditional sensitivity analysis for remotely sensed products, the sensitivity will be dependent on the algorithm used to derive the product and will be in the associated product units. In a multi-faceted mission such as SBG, how would one weigh the importance of e.g. Kaolinite fractional abundance (%) relative to that of leaf mass per unit area (g/m^2), or compare the accuracy of snow grain size (μm) to sulfur dioxide emissions from volcanoes (kt)? There is a critical need for an applications-agnostic, data-driven metric to enable a fair comparison. We propose that intrinsic dimensionality is such a metric.

We used airborne hyperspectral imagery over multiple biomes, including Chaparral, Desert, Mangrove, Temperate Forest, Dry Tropical Forest, Grasslands, Tundra, Boreal Forest, Coral Reefs, Agriculture, and Urban environments. There was some variability in response to GSD – some heterogeneous scenes experienced a significant drop in ID for coarser pixels, whereas homogeneous scenes were less sensitive. However, the response to spectral resolution was fairly consistent across all scenes, highlighting the importance of hyperspectral over multispectral data.

We also introduced the concept of temporal dimensionality, using phenological patterns to define information content over time, and using this to demonstrate the impact of revisit over both the visible to shortwave infrared and thermal infrared. This metric could also be used to evaluate optimal overpasses for complementary missions to maximize science value.

In future work, real instrument configurations can be compared to determine the optimal configuration within a design space, cost savings can be paired with measurable science consequences, and data harmonization can be explored.

6 Conclusion

The Surface Biology and Geology mission will be an essential part of the NASA Earth System Observatory, and optimal mission design is vital for answering science questions and meeting applications needs across diverse fields. Individual geophysical products relating to terrestrial and aquatic ecosystems, natural disasters, and the cryosphere might benefit from different mission design parameters according to each algorithm’s unique sensitivities, and so overall mission optimization is difficult. Here we have introduced the concept of Intrinsic Dimensionality as a metric to be considered in future mission design since it provides a single quantitative evaluation for a combination of design parameters, irrespective of higher-level algorithms, products, applications, or disciplines. ID is simple to compute, yet it captures the maximum data content achievable for a combination of parameters and indicates the potential of optimal algorithmic performance. We found that the ID decreases for coarser GSD, decreased spectral resolution, less frequent acquisitions, and lower signal-to-noise levels. This behavior impacts the quality and accuracy of all derived products. Furthermore, because ID is application-agnostic, it can be applied to other hyperspectral missions in Earth and planetary sciences.

Acknowledgments

Support to all authors was provided by NASA and the Jet Propulsion Laboratory, California Institute of Technology and to PAT by NSF Macrosystems Biology and NEON Early Science award 1638720. This study was also supported by the Space-based Imaging Spectroscopy and Thermal (SISTER) pathfinder, and the Surface Biology and Geology (SBG) project, a NASA Earth Science Designated Observable. The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. © 2022 California Institute of Technology. Government sponsorship is acknowledged.

Open Research

AVIRIS-NG reflectance data is available from <https://avirisng.jpl.nasa.gov/dataportal/>

NEON reflectance data is available from <https://data.neonscience.org/data-products/DP1.30006.001>.

PRISM data is available from <https://prism.jpl.nasa.gov/dataportal/>.

The code for both ISOFIT and Hypertrace is available via GitHub: <https://github.com/isofit/isofit>.

References

Anderson, G. P., Clough, S. A., Kneizys, F. X., Chetwynd, J. H., & Shettle, E. P. (1986). AFGL atmospheric constituent profiles (0.120 km). AIR FORCE GEOPHYSICS LAB HANSCOM AFB MA. <https://apps.dtic.mil/sti/pdfs/ADA175173.pdf>

AppEEARS Team. (2021). Application for Extracting and Exploring Analysis Ready Samples (AppEEARS). Ver. 2.63. NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, USA. Accessed August 31, 2021. <https://lpdaacsvc.cr.usgs.gov/appeears>

Asner, G. P., Knapp, D. E., Boardman, J., Green, R. O., Kennedy-Bowdoin, T., Eastwood, M., ... & Field, C. B. (2012). Carnegie Airborne Observatory-2: Increasing science data dimensionality via high-fidelity multi-sensor fusion. *Remote Sensing of Environment*, 124, 454-465. <https://doi.org/10.1016/j.rse.2012.06.012>

Baik, J., & Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *Journal of multivariate analysis*, 97(6), 1382-1408. <https://doi.org/10.1016/j.jmva.2005.08.003>

Bachmann, C. M., T. L. Ainsworth, R. A. Fusina (2008). Automated Estimation of Spectral Neighborhood Size in Manifold Coordinate Representations

of Hyperspectral Imagery: Implications for Anomaly Finding, Bathymetry Retrieval, and Land Applications. *Proc. IGARSS'08*, Boston, MA, July 2008, <https://dx.doi.org/10.1109/IGARSS.2008.4778791>

Berman, M. (2019). Improved estimation of the intrinsic dimension of a hyperspectral image using random matrix theory. *Remote Sensing*, 11(9), 1049. <https://doi.org/10.3390/rs11091049>

Bioucas-Dias, J. M., & Nascimento, J. M. (2005, October). Estimation of signal subspace on hyperspectral data. In *Image and Signal Processing for Remote Sensing XI* (Vol. 5982, p. 59820L). International Society for Optics and Photonics. <https://doi.org/10.1117/12.620061>

Boardman, J.W. and Green, R.O. (2000). Exploring the spectral variability of the Earth as measured by AVIRIS in 1999. *Proceedings of the Ninth JPL Airborne Earth Science Workshop* (JPL, 2000), pp. 195–206.

Brodrick, P.G., Thompson, D.R., Fahlen, J.E., Eastwood, M.L., Sarture, C.M., Lundeen, S.R., Olson-Duvall, W., Carmon, N. and Green, R.O., (2021). Generalized radiative transfer emulation for imaging spectroscopy reflectance retrievals. *Remote Sensing of Environment*, 261, p.11247 <https://doi.org/10.1016/j.rse.2021.112476>

Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., & Li, Z. (2018). A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote sensing of environment*, 210, 35-47. <https://doi.org/10.1016/j.rse.2018.02.045>

Cawse-Nicholson, K., Damelin, S. B., Robin, A., & Sears, M. (2012). Determining the intrinsic dimension of a hyperspectral image using random matrix theory. *IEEE Transactions on Image Processing*, 22(4), 1301-1310. <https://doi.org/10.1109/TIP.2012.2227765>

Cawse-Nicholson, K., Hook, S. J., Miller, C. E., & Thompson, D. R. (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>

Chang, C. I., & Du, Q. (2004). Estimation of number of spectrally distinct signal sources in hyperspectral imagery. *IEEE Transactions on geoscience and remote sensing*, 42(3), 608-619. <https://doi.org/10.1109/TGRS.2003.819189>

Chang, J., Hansen, M. C., Pittman, K., Carroll, M., & DiMiceli, C. (2007). Corn and soybean mapping in the United States using MODIS time-series data sets. *Agronomy Journal*, 99(6), 1654-1664. <https://doi.org/10.2134/agronj2007.0170>

Connelly, D.S., Thompson, D.R., Mahowald, N.M., Li, L., Carmon, N., Okin, G.S. and Green, R.O., (2021). The EMIT mission information yield for mineral dust radiative forcing. *Remote Sensing of Environment*, 258, p.112380.

Crisp, D., Atlas, R. M., Breon, F. M., Brown, L. R., Burrows, J. P., Ciaia, P., ...

- & Schroll, S. (2004). The orbiting carbon observatory (OCO) mission. *Advances in Space Research*, 34(4), 700-709. <https://doi.org/10.1016/j.asr.2003.08.062>
- Dierssen, Heidi M., Ackleson, Steven G., Joyce, Karen E., Hestir, Erin L., Castagna, Alexandre, Lavender, Samantha, and McManus, Margaret A. (2021) Living up to the hype of hyperspectral aquatic remote sensing: science, resources and outlook. *Frontiers in Environmental Science*, 9. 649528. <https://doi.org/10.3389/fenvs.2021.649528>
- Eldering, A., Wennberg, P. O., Crisp, D., Schimel, D. S., Gunson, M. R., Chatterjee, A., ... & Weir, B. (2017). The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes. *Science*, 358(6360). <https://doi.org/10.1126/science.aam5745>
- Fisher, J. B., Lee, B., Purdy, A. J., Halverson, G. H., Dohlen, M. B., Cawse-Nicholson, K., Wang, A., Anderson, R. G., Aragon, B., & Arain, M. A. (2020). ECOSTRESS: NASA's next generation mission to measure evapotranspiration from the International Space Station. *Water Resources Research*, 56(4), e2019WR026058. <https://doi.org/10.1029/2019WR026058>
- Foerster, S., Kaden, K., Foerster, M., & Itzerott, S. (2012). Crop type mapping using spectral-temporal profiles and phenological information. *Computers and Electronics in Agriculture*, 89, 30-40. <https://doi.org/10.1016/j.compag.2012.07.015>
- Francis, P., & Rothery, D. (2000). Remote sensing of active volcanoes. *Annual Review of Earth and Planetary Sciences*, 28(1), 81-106. <https://doi.org/10.1146/annurev.earth.28.1.81>
- Hulley, G., Hook, S. (2018). VIIRS/NPP Land Surface Temperature and Emissivity Daily L3 Global 1km SIN Grid Day V001. NASA EOSDIS Land Processes DAAC. Accessed 2021-08-20 from <https://doi.org/10.5067/VIIRS/VNP21A1D.001>. Accessed August 20, 2021.
- IPCC (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press. In Press. <https://www.ipcc.ch/report/ar6/wg1/>
- Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of statistics*, 295-327. <http://www.jstor.org/stable/2674106>
- Kritchman, S., & Nadler, B. (2008). Determining the number of components in a factor model from limited noisy data. *Chemometrics and Intelligent Laboratory Systems*, 94(1), 19-32. <https://doi.org/10.1016/j.chemolab.2008.06.002>

- Landerer, F. W., & Swenson, S. C. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. *Water resources research*, 48(4). <https://doi.org/10.1029/2011WR011453>
- Landerer, F. W., Flechtner, F. M., Save, H., Webb, F. H., Bandikova, T., Bertiger, W. I., ... & Yuan, D. N. (2020). Extending the global mass change data record: GRACE Follow-On instrument and science data performance. *Geophysical Research Letters*, 47(12), e2020GL088306. <https://doi.org/10.1029/2020GL088306>
- Lobell, D. B., & Asner, G. P. (2004). Cropland distributions from temporal unmixing of MODIS data. *Remote Sensing of Environment*, 93(3), 412-422. <https://doi.org/10.1016/j.rse.2004.08.002>
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403(6772), 853-858. <https://doi.org/10.1038/35002501>
- NEON (National Ecological Observatory Network). 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 September 5, 2021.
- Park, H. M., Jung, H. Y., Lee, T. W., & Lee, S. Y. (1999). Subband-based blind signal separation for noisy speech recognition. *Electronics Letters*, 35(23), 2011-2012. <https://doi.org/10.1049/el:19991358>
- Pascolini-Campbell, M., Fisher, J. B., & Reager, J. T. (2021). GRACE-FO and ECOSTRESS synergies constrain fine-scale impacts on the water balance. *Geophysical Research Letters*, 48, e2021GL093984. <https://doi.org/10.1029/2021GL093984>
- Raiho, A. M., Cawse-Nicholson, K., Chlus, A., Dozier, J., Gierach, M., Miner, K., Schneider, F., Schimel, D., Serbin, S., Shiklomanov, A. N., Thompson, D. R., Townsend, P. A., Zareh, S., Skiles, M., Poulter, B. (2022) Exploring mission design for imaging spectroscopy retrievals for land and aquatic ecosystems. *Journal of Geophysical Research: Biogeosciences*. In Review.
- Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuka, N., & Ohno, H. (2005). A crop phenology detection method using time-series MODIS data. *Remote sensing of environment*, 96(3-4), 366-374. <https://doi.org/10.1016/j.rse.2005.03.008>
- Schaaf, C., Wang, Z., Zhang, X., Strahler, A. (2018). VIIRS/NPP BRDF/Albedo Nadir BRDF-Adjusted Ref Daily L3 Global 500m SIN Grid V001. NASA EOSDIS Land Processes DAAC. Accessed 2021-08-31 from <https://doi.org/10.5067/VIIRS/VNP43IA4.001>. Accessed August 31, 2021.
- Schmidt, R., Schwintzer, P., Flechtner, F., Reigber, C., Güntner, A., Döll, P., ... & Wunsch, J. (2006). GRACE observations of changes in continental water

- storage. *Global and Planetary Change*, 50(1-2), 112-126. <https://doi.org/10.1016/j.gloplacha.2004.11.018>
- Small, C. (2001). Multiresolution analysis of urban reflectance. *Proceedings of the IEEE/ISPRS Joint Workshop 2001: Remote Sensing and Data Fusion over Urban Areas* (IEEE, 2001), pp. 15–19.
- Swayze, G. A., Clark, R. N., Goetz, A. F., Chrien, T. G., & Gorelick, N. S. (2003). Effects of spectrometer band pass, sampling, and signal-to-noise ratio on spectral identification using the Tetracorder algorithm. *Journal of Geophysical Research: Planets*, 108(E9). <https://doi.org/10.1029/2002JE001975>
- 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>
- Thompson, D. R., Natraj, V., Green, R. O., Helmlinger, M. C., Gao, B. C., & Eastwood, M. L. (2018). Optimal estimation for imaging spectrometer atmospheric correction. *Remote sensing of environment*, 216, 355-373. <https://doi.org/10.1016/j.rse.2018.07.003>
- Thompson, D.R., Braverman, A., Brodrick, P.G., Candela, A., Carmon, N., Clark, R.N., Connelly, D., Green, R.O., Kokaly, R.F., Li, L. and Mahowald, N., (2020). Quantifying uncertainty for remote spectroscopy of surface composition. *Remote Sensing of Environment*, 247, p.111898. <https://doi.org/10.1016/j.rse.2020.111898>
- Xiao, J., Fisher, J.B., Hashimoto, H., Ichii, K., Parazoo, N.C., 2021. Emerging satellite observations for diurnal cycling of ecosystem processes. *Nature Plants*. <https://doi.org/10.1038/s41477-021-00952-8>
- Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C., Gao, F., ... & Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote sensing of environment*, 84(3), 471-475. [https://doi.org/10.1016/S0034-4257\(02\)00135-9](https://doi.org/10.1016/S0034-4257(02)00135-9)