Thompson, D. R., Bohn, N., Braverman, A., Brodrick, P. G., Carmon, N., Eastwood, M. L., Fahlen, J. E., Green, R. O., Johnson, M. C., Roberts, D. A., Susiluoto, J. (2021): Scene invariants for quantifying radiative transfer uncertainty. - Remote Sensing of Environment, 260, 112432.

https://doi.org/10.1016/j.rse.2021.112432

Scene Invariants for Quantifying Radiative Transfer Uncertainty

David R. Thompson[a,*], Niklas Bohn[b], Amy Braverman[a], Philip G. Brodrick[a], Nimrod Carmon[a], Michael L. Eastwood[a], Jay Fahlen[a], Robert O. Green[a], Margaret C. Johnson[a], Dar Roberts[c], Jouni Susiluoto[a]

[a] *Jet Propulsion Laboratory, California Institute of Technology, Pasadena CA 91109*
[b] *GFZ German Research Centre for Geosciences, 14473 Potsdam, Germany*
[c] *University of California Santa Barbara, Santa Barbara, CA 93106*

Abstract

Remote imaging spectroscopy, also known as hyperspectral imaging, uses Radiative Transfer Models (RTMs) to predict the measured radiance spectrum for a specific surface and atmospheric state. Discrepancies between RTM assumptions and physical reality can cause systematic errors in surface property estimates. We present a statistical method to quantify these model errors without invoking ground reference data. Our approach exploits *scene invariants* — properties of the environment which are stable over space or time — to estimate RTM discrepancies. We describe techniques for discovering these features opportunistically in flight data. We then demonstrate data-driven methods that estimate the aggregate errors due to model discrepancies without having to explicitly identify the underlying physical mechanisms. The resulting distributions can improve posterior uncertainty predictions in operational retrievals.

Keywords: Imaging Spectroscopy, Hyperspectral Imaging, Uncertainty Quantification

1. Introduction

Remote imaging spectroscopy, also known as hyperspectral imagery, relies on Radiative Transfer Models (RTMs) to predict the measured radiance for a given surface and atmospheric state (Gao et al., 1993). This casts remote measurement as an inverse problem to find the most probable state for the measured radiance. The pattern repeats across disciplines: atmospheric correction inverts an atmospheric RTM (Clough et al., 2005); aquatic studies invert water column RTMs (Mobley, 1994); terrestrial ecologists invert tree canopy RTMs (Kobayashi & Iwabuchi, 2008); geology, cryosphere, and urban studies have unique radiative transfer formalisms (Lucey & Noble, 2008; Dozier et al., 2009; Voogt & Oke, 2003).

RTMs are approximations and inevitably differ from reality. These differences, known as *model discrepancies*, include approximations for computational convenience, well-defined but unknown parameters which cannot be measured directly, and any "unknown unknowns" that manifest as undiagnosed errors in validation experiments. Quantifying these model discrepancy errors is important for unbiased remote measurements (Brynjarsdóttir & O'Hagan, 2014), and enables accurate hypothesis testing based on the results. Additionally, algorithms that respect and propagate measurement uncertainty can improve performance by ascribing the optimal weighting to uncertain data (Carmon et al., 2020). Uncertainty predictions enable a principled synthesis of global data products acquired under varied observing conditions (Hobbs et al., 2017). They can reveal faults in our understanding to guide model refinement, bounding the observation system and revealing it for the first time in its entirety.

While quantifying uncertainty is critical, it is also challenging. Simply assessing historical errors is not sufficient; one must develop an uncertainty model that predicts future errors under novel observing conditions. To this end, classical uncertainty propagation projects the sensor's measurement noise distribution through downstream analysis. This can explain discrepancies between remote estimates and field data under well-controlled observing conditions (Thompson et al., 2020). However, in more challenging circumstances, errors due to model discrepancies often surpass those of instrument noise. Here we define RTM discrepancies broadly to include diverse differences between the actual and modeled observation system:

- Radiometric calibration error (Chapman et al., 2019);
- Spectral calibration error, including errors in spectral response functions (Thompson et al., 2018a) or nonuniformity (Richter et al., 2010);
- Error in profiles and concentrations of atmospheric constituents (Griffin & Burke, 2003);
- Error in optical behavior of atmospheric constituents, such as gas absorption coefficients or aerosol scattering properties (Thompson et al., 2019);
- Error caused by discrete approximations for the radiative transfer solution (Lu et al., 2009) or lookup table interpolation (Bue et al., 2019);
- Errors or omissions in estimating geometric effects at the surface, including topography (Richter et al., 2009) and adjacency effects (Sanders et al., 2001);
- Retrieval error caused by biases or inaccuracies in the inversion algorithm itself (Kulawik et al., 2019);

*Correspondence to: david.r.thompson@jpl.nasa.gov

- Errors from wholly unmodeled physical phenomena such as cirrus clouds (Gao et al., 1998).

If the model discrepancies and their associated probability distributions were known, they could be counted appropriately in the inversion for more accurate retrievals, and in the uncertainty propagation for more accurate confidence bounds (Braverman et al., 2020). However, in practice model discrepancies are often ignored because it is difficult to identify, let alone quantify, their effects. They can involve physical processes in the atmosphere or surface media that are challenging to measure. Field data have their own uncertainties and cannot scale globally.

This work demonstrates that it is possible to estimate model discrepancy errors from statistics of the instrument data without invoking ground truth observations. Our solution uses *scene invariants* - properties of the environment which are constant over space or time (Schott et al., 1988; Canty et al., 2004). One example of such a property is the shape of the spectral surface reflectance, which is stable over multi-year timescales in selected areas of the globe (Cosnefroy et al., 1996; Bouvet, 2014; Bouvet et al., 2019). On smaller scales, airborne campaigns regularly recalibrate using smaller invariant surfaces such as roofs (Meerdink et al., 2019), concrete patches (Rockwell et al., 2000), gravel (Kokaly et al., 2003), and bare soil or rock (Clark et al., 2002). Any remote image of sufficient size is likely to contain many such surfaces, which can help to harmonize data acquired under different conditions (Furby & Campbell, 2001; Canty et al., 2004). We will show that invariant surfaces can also serve as reference standards to quantify time-variable atmospheric distortions, allowing investigators to estimate the uncertainty caused by model discrepancies without needing to specify the physical mechanisms. We will also show that such features can be detected automatically in flight data, so that it is not necessary to rely upon the invariance of any particular preselected location. Instead, investigators can exploit the statistical likelihood of observing at least some invariant surfaces in any sufficiently large scene. The resulting distributions can be used to improve the accuracy of retrievals and predicted uncertainties.

These methods are well-suited for the next generation of orbital imaging spectrometers. Missions such as NASA's Earth Mineral dust source InvestigaTion (EMIT) aim to build consistent global maps across variable and challenging atmospheres (Green et al., 2020). Better uncertainty estimates will enable an optimal synthesis of data with variable quality. Scene invariants are also appropriate for repeat observations by the European Space Agency's Copernicus Hyperspectral Imaging Mission (CHIME) (Taramelli et al., 2020) and NASA's Surface Biology and Geology (SBG) architecture (National Academies of Sciences, Engineering, and Medicine, 2018). These investigations will return regularly to the same sites across diverse atmospheric conditions. Such dense time series will be powerful for quantifing uncertainties in atmospheric radiative transfer.

Section 2 introduces the mathematical foundations of our approach. We define invariant scene properties and demonstrate robust estimation of model discrepancy errors. The technique applies to many different RTMs, but we focus on uncertainties associated with atmospheric correction, i.e. the estimation of surface reflectance and atmospheric constituents from remote radiance data. Section 3 applies the method to a large dataset of repeat flightlines from NASA's Next Generation Airborne Visible Infrared Imaging Spectrometer (AVIRIS-NG) (Chapman et al., 2019) acquired during a campaign in India from 2015-2018 (Bhattacharya et al., 2019). We show that scene invariants exist in diverse terrestrial images, and that the estimated model discrepancy distributions translate to other domains and improve the accuracy of future uncertainty predictions.

2. Quantifying Model Discrepancies

This section introduces the fundamental concepts and notation. The formalism applies to any RTM or inversion method, but we present a case study based on the atmospheric correction problem. Atmospheric correction uses the measured radiance spectrum at the sensor to estimate the surface reflectance spectrum, typically defined as a Directional Reflectance Factor (DRF), the fraction of downwelling irradiance at the surface which is reflected in the direction of the sensor (Nicodemus, 1970). Hereafter we will refer to this quantity as the spectral surface reflectance, or simply reflectance. To estimate reflectance remotely, one must simultaneously find the columnar concentrations of atmospheric gases and suspended particles which influence the measured radiance and vary rapidly across space and time. Atmospheric correction is important for remote imaging spectroscopy because it precedes any domain-specific analysis of surface chemical or physical properties. Its mathematical structure - inversion of a nonlinear RTM with prior constraints and model discrepancies - is generic and broadly applicable.

We motivate our method by first presenting two straw person alternatives. First, we describe a *direct* estimate of model discrepancy using ground truth measurements, which are not available for global observations. We then describe a *plug in* estimate using the output of an inversion algorithm. The plug in estimate is convenient to deploy over large datasets, but systematically underestimates model discrepancy errors. Finally, we derive a method based on *scene invariants* which captures the best of both alternatives, improving accuracy while remaining practical for operational use.

2.1. The Direct Approach

Remote inversions estimate a state vector $\mathbf{x}$ of free parameters. Table 1 shows our notation, with boldface signifying matrix or vector-valued quantities. In atmospheric correction, $\mathbf{x}$ specifies the surface and atmospheric properties. In this discussion we parameterize the surface portion $\mathbf{x}_{surf}$ using the Hemispherical Directional Reflectance Factor, or HDRF (Schaepman-Strub et al., 2006), in each channel, and the atmospheric portion $\mathbf{x}_{atm}$ using the column water vapor concentration and the optical depth for one or more aerosol types (Thompson et al., 2018b). All other parameters of the atmospheric model, such as the concentrations of other gases, are held fixed during the retrieval. The complete state vector $\mathbf{x}$ is the concatenation of the surface and atmosphere components. For

Symbol	Meaning
$\mathbf{b}$	Parameters of the forward function which are not retrieved
$\hat{\mathbf{b}}$	Estimated forward model parameters
$\mathbf{f}$	Forward function mapping state vector to measurement
$\mathbf{F}$	Forward model which attempts to reproduce $\mathbf{f}$
i	Indexes over spectra in a dataset
$\mathbf{G}$	Gain matrix, the linearized response of the retrieval to a change in the measurement
$\mathbf{K}$	Jacobian of measurement with respect the state vector
$\mathbf{K_b}$	Jacobian of measurement with respect to $\mathbf{b}$
$\hat{\mathbf{L}}$	Observed radiance at the sensor
ℓ	Number of distinct acquisitions (i.e. data cubes)
m	Number of measurement channels
n	Number of state vector elements
$\mathbf{q}$	Extraterrestrial solar irradiance
$\mathbf{R}$	Inversion function mapping a measurement to a state vector
$\mathbf{s}$	Spherical albedo of the atmosphere
$\mathbf{S_a}$	Prior state vector covariance
$\mathbf{S_g}$	In situ measurement covariance
$\mathbf{S}_\epsilon$	Observation uncertainty covariance including $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$
$\hat{\mathbf{S}}$	Posterior predicted uncertainty covariance
$\mathbf{t}$	Transmittance of the atmosphere
w	Total number of invariant locations in scene
$\mathbf{x}$	State vector of free parameters
$\hat{\mathbf{x}}$	Estimated state vector
$\bar{\mathbf{x}}$	Ensemble mean state vector
$\mathbf{x}_a$	Prior state vector mean
$\mathbf{x}_{atm}$	Atmospheric state vector elements
$\mathbf{x}_{surf}$	Surface state vector elements, reflectance in each channel
$\mathbf{y}$	Vector of measured radiance values
$\boldsymbol{\delta}$	Model discrepancy error, a random variable
$\boldsymbol{\epsilon}$	Instrument noise, a random variable
$\mathbf{\Gamma}$	Covariance of model discrepancy error
$\mathbf{\Lambda}$	Covariance of instrument noise
$\boldsymbol{\mu}$	Mean of model discrepancy error
$\boldsymbol{\rho}_a$	Atmospheric path reflectance
$\boldsymbol{\rho}_s$	Surface reflectance
$\boldsymbol{\upsilon}$	Number of invariant locations used
$\boldsymbol{\theta}_s$	Solar zenith angle

Table 1: Notation. Boldface signifies vector or matrix-valued quantities.

this experiment, our sensor is AVIRIS-NG which has $m = 425$ spectral channels. Together with columnar precipitable water vapor and a single aerosol type for the atmosphere, the complete state vector has $n = 427$ elements. Following Rodgers (2000), we explain the radiance measurement $\mathbf{y}$ using a forward function $\mathbf{f}(\mathbf{x}, \mathbf{b})$, which depends on the true state $\mathbf{x}$ and any ancillary parameters $\mathbf{b}$ that are not actively estimated. The forward function represents the true underlying physical process that produces the radiance at sensor.

The measurement process perturbs the forward function by sensor noise $\boldsymbol{\epsilon}$, a zero-mean Gaussian random variable with covariance $\mathbf{\Lambda}$. This represents uncertainty due to the natural inaccuracy of the instrument, including signal-dependent noise sources such as photon counting statistics and signal-independent sources such as electronic and readout noise. Photon shot noise depends on the radiance magnitude so $\mathbf{\Lambda}$ is different for each spectrum. Fortunately, it is easy to calculate $\mathbf{\Lambda}$ for any measured radiance using conventional instrument models (Thompson et al., 2020). The actual radiance which governs the noise distribution and the estimated radiance at sensor are effectively identical, so the instrument noise distribution is known for all observations. We write the measurement as:

$$\mathbf{y} = \mathbf{f}(\mathbf{x}, \mathbf{b}) + \boldsymbol{\epsilon} \;\text{ for }\; \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{\Lambda}) \tag{1}$$

We do not know the true forward function, so we instead approximate it with a *forward model* $\mathbf{F}(\mathbf{x}, \hat{\mathbf{b}})$, an RTM which predicts the radiance measured at the sensor. We are also ignorant of the true vector $\mathbf{b}$, so we use our best estimate written $\hat{\mathbf{b}}$. We represent the *model discrepancy* as a random variable $\boldsymbol{\delta}$ which includes the aggregate uncertainty due to errors in the forward function and the ancillary parameters:

$$\boldsymbol{\delta} = \mathbf{f}(\mathbf{x}, \mathbf{b}) - \mathbf{F}(\mathbf{x}, \hat{\mathbf{b}}) \tag{2}$$

In other words, $\boldsymbol{\delta}$ represents the differences between the approximate RTM and true physics. In restricted cases, it may be possible to characterize these errors from simulation or first principles. For example, laboratory-derived parameters like absorption coefficients have well-known uncertainties (Gordon et al., 2017). More often, it is difficult to quantify all potential model discrepancies since $\boldsymbol{\delta}$ includes errors in atmospheric modeling that cannot be diagnosed. Provisionally for this paper we will treat the entire model discrepancy as a single random variable $\boldsymbol{\delta}$, with a distribution that is independent of $\mathbf{x}$ and $\mathbf{b}$ and that can be estimated from data. Specifically, we choose to model $\boldsymbol{\delta}$ as Gaussian with a mean vector $\boldsymbol{\mu}$ and covariance $\mathbf{\Gamma}$. The measured radiance $\mathbf{y}$ is the sum of the forward model and the two independent noise sources:

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) + \boldsymbol{\delta} + \boldsymbol{\epsilon} \;\text{ for }\; \boldsymbol{\delta} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Gamma}),\; \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{\Lambda}) \tag{3}$$

The functional dependence of $\mathbf{F}$ on $\hat{\mathbf{b}}$ is implicit. In practice, it is likely that the radiance errors from $\boldsymbol{\delta}$ will depend on local conditions. More sophisticated investigators might structure $\boldsymbol{\delta}$ to capture these dependencies. For example, $\boldsymbol{\delta}$ might grow for observations with more challenging atmospheric conditions or large view zenith angles. Here we ignore these complexities and fit one global distribution.

Characterizing model discrepancy under Gaussian assumptions thus reduces to estimating a bias μ and covariance $\mathbf{\Gamma}$. For several reasons, the rest of the manuscript will focus on $\mathbf{\Gamma}$ exclusively. The mean μ relates to any systematic error which is the same for all observations, like a consistent calibration bias or approximation errors in the RTM code. The fact that such offsets are consistent makes them relatively straightforward to fix with conventional vicarious approaches (Wunch et al., 2011), zeroing μ. Moreover, most error processes show variability if the dataset is large enough; an atmospheric error that is constant across a flightline may be different across campaigns. Consequently, we focus our manuscript on estimating the covariance, $\mathbf{\Gamma}$, which is the most challenging hurdle for practical imaging spectroscopy scenarios and dominates our errors in large datasets. We will bookkeep atmospheric and calibration errors in $\mathbf{\Gamma}$, assuming $\mu = \mathbf{0}$. This approach is demonstrably sufficient to close uncertainty budgets in our validation experiments.

The total observation error covariance is the sum of the sensor noise and model discrepancy covariance matrices. For continuity with prior studies we use the notation $\mathbf{S}_\epsilon$ to include both

ϵ and δ:

$$\mathbf{S}_\epsilon \;=\; \mathrm{cov}(\mathbf{y} - \mathbf{F}(\mathbf{x}) \mid \mathbf{x}) \;=\; \mathbf{\Gamma} \;+\; \mathbf{\Lambda} \tag{4}$$

The vertical line indicates that the distribution is defined conditionally on the true value of $\mathbf{x}$. When this value is known, one can estimate $\mathbf{\Gamma}$ with the following expression:

$$\hat{\mathbf{\Gamma}} = \frac{1}{n}\sum_i (\mathbf{y}_i - \mathbf{F}(\mathbf{x}_i))(\mathbf{y}_i - \mathbf{F}(\mathbf{x}_i))^T \;-\; \frac{1}{n}\sum_i \mathbf{\Lambda}_i \tag{5}$$

where i indexes over all spectra in the dataset, and the noise covariance $\mathbf{\Lambda}_i$ is specific to each measured radiance. This is the direct method for evaluating model discrepancy. Note that in principle $\mathbf{\Gamma}$ is based on the distribution of $\mathbf{x}$ and $\mathbf{b}$. For a conservative uncertainty prediction one should fit $\mathbf{\Gamma}$ using a dataset which is at least as challenging as the test set. The effective number of samples from $\boldsymbol{\epsilon}$ will be larger than the samples from $\boldsymbol{\delta}$; the former is different in each spectrum, while the latter varies over larger spatial scales based on surface, atmospheric, and geometric conditions.

Equation 5 requires knowledge of the true $\mathbf{x}$. This might come from ancillary sources such as *in situ* surface measurements. With many field measurements under enough different atmospheric conditions, one might eventually obtain a diverse enough dataset to estimate $\mathbf{\Gamma}$ and $\boldsymbol{\mu}$. This might be achievable for recurring orbits over a calibration / validation sensor network. However, such options are not available for all campaigns, and ground networks are not available in many regions of the globe. Moreover, in situ measurements carry their own uncertainties and biases. Vicarious measurements of surfaces have errors approaching 2% under perfect conditions (Thome, 2016), up to 5% in more typical environments (Jiménez et al., 2018). These errors are similar to the magnitude we expect for model discrepancies. Measuring atmospheric parameters is even harder because it is difficult to replicate the bent optical path through the spatially heterogeneous field of gases and particles. This drives us to consider other methods that are more practical to implement operationally.

2.2. The Plug In Approach

Lacking *in situ* data, one can use an inversion algorithm to estimate $\mathbf{x}$. We write the inversion as a function $\mathbf{R}(\mathbf{y})$ producing an estimate $\hat{\mathbf{x}}$:

$$\hat{\mathbf{x}} = \mathbf{R}(\mathbf{y}) \tag{6}$$

Note that $\mathbf{x}$ and $\mathbf{y}$ lie in different vector spaces. In our atmospheric correction example the number of measurements m is smaller than the number of free parameters n. In practice the problem is numerically well-determined due to the statistics of natural surfaces; some reflectance shapes are more plausible than others, making the effective number of free parameters smaller than n.

In a Bayesian context we capture this information as a prior over states, a multivariate Gaussian with mean $\mathbf{x_a}$ and covariance $\mathbf{S_a}$. The prior covariance captures known spectral correlations in surface reflectance, such as the shape and smoothness of reflectance spectra in the neighborhood of water vapor absorption features. This information enables the retrieval to attribute the absorption signature of water to the atmosphere instead of the surface. More complex and sophisticated prior distributions are possible, such as locally-Gaussian manifolds (Candela et al., 2020). Here we follow the approach of Thompson et al. (2018b), estimating a collection of Gaussian distributions from a diverse set of curated spectra, and then broadening them with diagonal loading to enable retrieval of spectral shapes outside the span of the library. Priors for atmospheric terms are typically broad and independent of the surface. The maximum a posteriori estimate of $\mathbf{x}$ is that which minimizes the following cost function:

$$\chi(\mathbf{x}) = (\mathbf{F}(\mathbf{x}) - \mathbf{y})^\mathbf{T}\mathbf{S}_\epsilon^{-\mathbf{1}}(\mathbf{F}(\mathbf{x}) - \mathbf{y}) \;+ (\mathbf{x} - \mathbf{x_a})^\mathbf{T}\mathbf{S_a^{-1}}(\mathbf{x} - \mathbf{x_a}) \tag{7}$$

The first term penalizes departure of the modeled radiance from the true measurement, weighted by the observation noise covariance. The second term penalizes departure from the prior. Linear models have a closed form solution (Thompson et al., 2020). Nonlinear models can be inverted with an iterative optimization of the state vector (Thompson et al., 2018b) or Markov Chain Monte Carlo methods (Thompson et al., 2019). In a Maximum A Posteriori Inversion with Gaussian random variables, the posterior uncertainty at the solution state, $\hat{\mathbf{S}}$, is a function of the noise covariances $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$, the prior covariance $\mathbf{S_a}$, and Jacobian matrices $\mathbf{K}$ representing the forward model linearized at the solution (Rodgers, 2000):

$$\hat{\mathbf{S}} \;=\; (\mathbf{K}^\mathbf{T}\mathbf{S}_\epsilon^{-\mathbf{1}}\mathbf{K} + \mathbf{S_a^{-1}})^{-\mathbf{1}} \;=\; (\mathbf{K}^\mathbf{T}(\mathbf{\Gamma} + \mathbf{\Lambda})^{-\mathbf{1}}\mathbf{K} + \mathbf{S_a^{-1}})^{-\mathbf{1}} \tag{8}$$

The Jacobian matrices have size $m \times n$; each element contains $\partial\mathbf{F}/\partial\mathbf{x}$ for the corresponding measurement channel and state vector element. Parameters which are not continuously differentiable, such as a categorical selection of aerosol types, lead instead to a posterior which is a mixture of Gaussians. Here all free parameters are differentiable, enabling use of Equation 8. This expression captures the observation noise, but it is not intended to characterize bias arising from a nonzero $\boldsymbol{\mu}$ which can be corrected directly (Wunch et al., 2011).

The inversion enables a *plug in* estimate of the model discrepancy (Wasserman, 2006). Substituting the solution state $\hat{\mathbf{x}}$ into Equation 5 yields:

$$\hat{\mathbf{\Gamma}} = \frac{1}{n}\sum_i (\mathbf{y}_i - \mathbf{F}(\hat{\mathbf{x}}_i))(\mathbf{y}_i - \mathbf{F}(\hat{\mathbf{x}}_i))^T \;-\; \frac{1}{n}\sum_i \mathbf{\Lambda}_i \tag{9}$$

One first optimizes $\hat{\mathbf{x}}$ to fit the measurement, and then ascribes to model discrepancy any residual mismatch between the modeled radiance and true measurement that is not explained by sensor noise. This is simple to implement and better than ignoring $\boldsymbol{\delta}$ altogether. However, a poorly-designed retrieval can "fit the sensor noise," shrinking the residuals by shifting the spectrum fitting error to $\hat{\mathbf{x}}$. This would lead to an optimistic estimate of $\mathbf{\Gamma}$. The analysis of Appendix A demonstrates that, for a retrieval which optimally balances prior information and measurement error, the covariance of radiance residuals is given by the following expression:

$$\mathrm{cov}(\mathbf{y} - \mathbf{F}(\hat{\mathbf{x}}) \mid \hat{\mathbf{x}}) = \mathbf{S}_\epsilon(\mathbf{K}\mathbf{S}_\mathrm{a}\mathbf{K}^\mathrm{T} + \mathbf{S}_\epsilon)^{-1}\mathbf{S}_\epsilon \tag{10}$$

This only approaches the desired value of $\mathbf{S}_\epsilon$ if the prior covariance is strong, i.e. $\mathbf{S_a}$ is small relative to $\mathbf{S}_\epsilon$. A strong prior prevents the retrieval from fitting the noise so that all sensor noise and model discrepancy error is preserved in the radiance residuals. Such strong priors may be possible over terrain that is very well understood, but ultimately this is simply the ground truth requirement in a new disguise. It thwarts our need to sample a wide range of observation conditions and fully capture $\boldsymbol{\delta}$. The following section presents an alternative estimation strategy with fewer restrictions on the retrieval algorithm and scene content.

2.3. Scene Invariants

Many materials in the environment have stable reflectances which do not change spectral shape over timespans of days to years. These include highly stable 'pseudo-invariant" sites chosen as orbital calibration targets (Cosnefroy et al., 1996; Bouvet, 2014; Bouvet et al., 2019). Imaging spectrometer observations, which typically have high spatial resolution of less than 100 m, can make use of much smaller areas; indeed, they have a high probability of imaging multiple invariant surfaces in any acquisition. That such surfaces exist in most flightlines is a truism of airborne imaging spectroscopy, which uses them for vicarious field calibration. Examples include artificial surfaces such as roofs (Meerdink et al., 2019), concrete (Rockwell et al., 2000), gravel (Kokaly et al., 2003), and fallow agricultural fields (Clark et al., 2002). Many natural surfaces such as playas, beaches, rock outcrop, bare soil and rock talus slopes have also been used successfully as spectrally-invariant reference targets (Clark et al., 2002).

In our estimation problem, invariant reflectance shapes typify a more general situation in which properties of the true state vector are the same across multiple instantiations of $\boldsymbol{\delta}$ and $\boldsymbol{\epsilon}$. For example, multiple overflights of the same invariant targets witness different realizations of noise, solar geometry and atmospheric conditions. Each overflight constitutes a different draw from $\boldsymbol{\delta}$. This allows us improve the estimate $\hat{\mathbf{x}}$ by replacing it with the ensemble mean of the spectral shape estimated by the different overflights, written $\bar{\mathbf{x}}$. By the law of large numbers, averages of $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$, which are Gaussian random variables, will trend to their mean values so that the retrieval cannot fit the noise, and the radiance residuals preserve the random part of the error. Appendix A demonstrates that, for an ensemble of states drawn from the prior distribution, the covariance of residuals becomes:

$$\mathrm{cov}(\mathbf{y} - \mathbf{F}(\bar{\mathbf{x}}) \mid \bar{\mathbf{x}}) = \mathbf{K}(\mathbf{I}_\mathrm{n} - \mathbf{A})\mathbf{S}_\mathrm{a}(\mathbf{I}_\mathrm{n} - \mathbf{A})^\mathrm{T}\mathbf{K} + \mathbf{S}_\epsilon \qquad (11)$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix and $\mathbf{A}$ is the $n \times n$ averaging kernel matrix with elements $\mathbf{A}_{ij} = \partial\hat{\mathbf{x}}_i/\partial\mathbf{x}_j$. The averaging kernel matrix represents the sensitivity of the retrieval with respect to some change in the underlying true state. For a perfect inversion $\mathbf{A}$ would be the identity matrix and the first term on the right side would disappear; the covariance of residuals would be $\boldsymbol{S}_\epsilon$. In practice the diagonal elements of $\mathbf{A}$ are less than unity and the off-diagonal elements are nonzero. The trace of the averaging kernel matrix is known as the degrees of freedom in signal, and measures the retrieval's information content (Rodgers, 2000).

Equation 11 has two desirable properties relative to Equation 10. First, as the retrieval becomes more accurate and $\mathbf{A}$ approaches the identity, Equation 11 becomes closer to the ideal $\boldsymbol{S}_\epsilon$. Second, it is conservative, strictly overestimating the model discrepancy since the resulting covariance cannot be smaller than $\mathbf{S}_\epsilon$. This, with Equation 5, enables our fundamental result:

$$\hat{\boldsymbol{\Gamma}} = \frac{1}{n}\sum_i (\mathbf{y}_i - \mathbf{F}(\bar{\mathbf{x}}_i))(\mathbf{y}_i - \mathbf{F}(\bar{\mathbf{x}}_i))^T \;\; - \;\; \frac{1}{n}\sum_i \boldsymbol{\Lambda}_i \qquad (12)$$

This expression based on scene invariants is a superior estimator of the model discrepancy covariance $\boldsymbol{\Gamma}$.

In any real scenario we must account for the fact that only some state vector properties are invariant. For example, in the experiments that follow we estimate the Lambertian-equivalent surface reflectance. Real surfaces can be non-Lambertian, and even if the spectral shape is constant the observed magnitude can change due to the photometric effects of incident illumination angles and self-shading. To account for this, before calculating the ensemble mean we first normalize the spectra so that the areas under the reflectance curves match that of flightline i. This acts as a first order correction for photometric variability from differing solar incidence angles on sloped surfaces. It also removes most of the variability due to Bidirectional Reflectance Distribution Function effects, which can be roughly approximated by shifts in the magnitude of spectral reflectance. An alternative would be to redefine the surface state vector to be a different reflectance quantity that is independent of viewing geometry, and account for photometry and BRDF in the forward model. In addition to photometry, the atmospheric terms will also vary across flightlines. In other words, $\mathbf{x}_{surf}$ is invariant but $\mathbf{x}_{atm}$ changes in each acquisition. We thus require a scene-specific estimate for $\mathbf{x}_{atm}$. Here it is sufficient to simply keep the original estimated atmospheric state. Alternatively one could perform a second inversion that holds the surface reflectance to the ensemble mean, and then optimizes the remaining parameters to minimize Equation 7. We defer exploration of these alternatives, along with more sophisticated photometric and BRDF corrections, to future work.

2.4. Theoretical Performance

The performance of the estimators depends on how well the radiance residuals' covariance approximates the true $\mathbf{S}_\epsilon$. Here we assess performance in simulation with full control over $\mathbf{S}_\epsilon$. We demonstrate that for a practical inversion algorithm, in the limiting case of many overflights, $\hat{\boldsymbol{\Gamma}}$ in Equation 9 is smaller than $\mathbf{S}_\epsilon$ while in Equation 12 it is nearly identical to $\mathbf{S}_\epsilon$.

Our experiment synthesizes remote observations of a wide range of natural and artificial materials. We select 109 diverse reflectance spectra from the USGS spectral library version 7.0 (Kokaly et al., 2017) representing different macro-scale surfaces. They include rangeland mixtures of vegetation, soil, and nonphotosynthetic vegetation; soil and rock spectra; and artificial surfaces such as concrete. We use these reflectance signals

to calculate the at-sensor radiance $\hat{\mathbf{L}}$ based on the following expression (Thompson et al., 2018b), with $\circ$ and the fraction representing element-wise multiplication and division, respectively:

$$\hat{\mathbf{L}} = \frac{\mathbf{q}\cos(\theta_s)}{\pi} \circ \left[\boldsymbol{\rho}_\mathbf{a} + \frac{\mathbf{t} \circ \boldsymbol{\rho}_\mathbf{s}}{\mathbf{1} - (\mathbf{s} \circ \boldsymbol{\rho}_\mathbf{s})} \right] \tag{13}$$

Here $\mathbf{q}$ is the extraterrestrial solar irradiance, θ_s is the solar zenith angle, $\boldsymbol{\rho}_s$ represents surface reflectance, $\boldsymbol{\rho}_a$ is the atmospheric path reflectance due to scattering, $\mathbf{t}$ is the atmospheric transmittance along the bent path from sun to ground to sensor, and $\mathbf{s}$ is the spherical sky albedo. Alternative parameterizations, such as those incorporating adjacency effects, could easily be used instead (Sanders et al., 2001). Here we calculate atmospheric parameters with the MODTRAN 6.0 RTM using observation geometry from the airborne experiment in Thompson et al. (2018b). We posit a midlatitude summer atmospheric profile (Anderson et al., 1986) with an aerosol optical depth of 0.2 at 550 nm, and highly-scattering optical properties based on the sulfate model of Thompson et al. (2019). This model has proven versatile for operational use across a wide range of environments. We assign a column water vapor of 1.7 g cm^{-2}. Figure 1 shows examples of three representative USGS spectra, the atmospheric optical coefficients, and the simulated radiances. These consist of an oak bush, clean sand, and a burned area. Other spectra include mixed surfaces with varying fractions of green vegetation, substrate, and non-photosynthetic vegetation.

We define $\boldsymbol{S}_\epsilon$ to be a diagonal noise matrix with marginal standard deviations ranging from 0.003 to 0.1 of the simulated radiance. We then perform a Maximum A Posteriori inversion, solving Equation 7 via gradient descent as in Thompson et al. (2020). As in that earlier work, surface prior covariances are diagonal and unconstrained outside key water absorption features at 940 nm and 1140 nm. This covariance structure is broadly representative of the uninformed priors used in operational "Level 2" reflectance products such as those of the EMIT mission (Green et al., 2020). This defines all the necessary quantities to calculate Equations 10 and 11 with linearizations at the appropriate states. We compare the covariance of residuals to the actual $\mathbf{S}_\epsilon$ by taking the root of the sum of diagonal elements, reducing the covariance matrices to a scalar value for assessing their magnitudes.

Figure 2 shows the result. Residuals calculated from Equation 10, shown as red dots, are smaller than $\mathbf{S}_\epsilon$. This is caused by the inversion absorbing some of the observation noise into state vector inaccuracy. However, residuals from scene invariants, shown as blue dots, match the true statistics of $\mathbf{S}_\epsilon$. For all scenarios, Equation 11 produces residuals that are strictly larger than $\mathbf{S}_\epsilon$, demonstrating that the estimator is conservative. Moreover, the perturbations caused by the first term in Equation 11 are always an order of magnitude below $\mathbf{S}_\epsilon$ itself. This demonstrates that an approach based on scene invariants has the potential to be an accurate estimator for $\mathbf{S}_\epsilon$, and therefore Γ.

2.5. *Opportunistic Discovery of Invariant Surfaces*

Our estimation strategy relies on the presence of invariant surfaces in the environment. Sometimes these can be identified in advance, as in spectrally-stable playas used for vicarious calibration (Bouvet et al., 2019). However, it is also possible to discover such surfaces in science data (Canty et al., 2004). Opportunistic surfaces enable a much larger statistical sample and better alignment between the distribution of observing conditions used in calculating versus deploying model discrepancy estimates. An obvious strategy is to select locations with estimated reflectances that are similar across acquisitions. If w of the locations in an acquisition are invariant, one could use the w most similar reflectance retrievals as an ensemble in Equation 12. We cannot know how many invariant locations are actually present, so we err on the side of conservatism by using the v most similar pixels for v safely smaller than w. A natural objection is that, by intentionally selecting locations with similar reflectance estimates, this technique samples the "lucky" tail of the error distribution and thereby underestimates the RTM error. This section demonstrates that for typical imaging spectroscopy scenarios this is not the case, and that in practice the approach accurately estimates Γ.

Consider a dataset comprised of ℓ distinct acquisitions, each spanning multiple square kilometers of terrain, and containing w invariant locations. The effective number of samples from

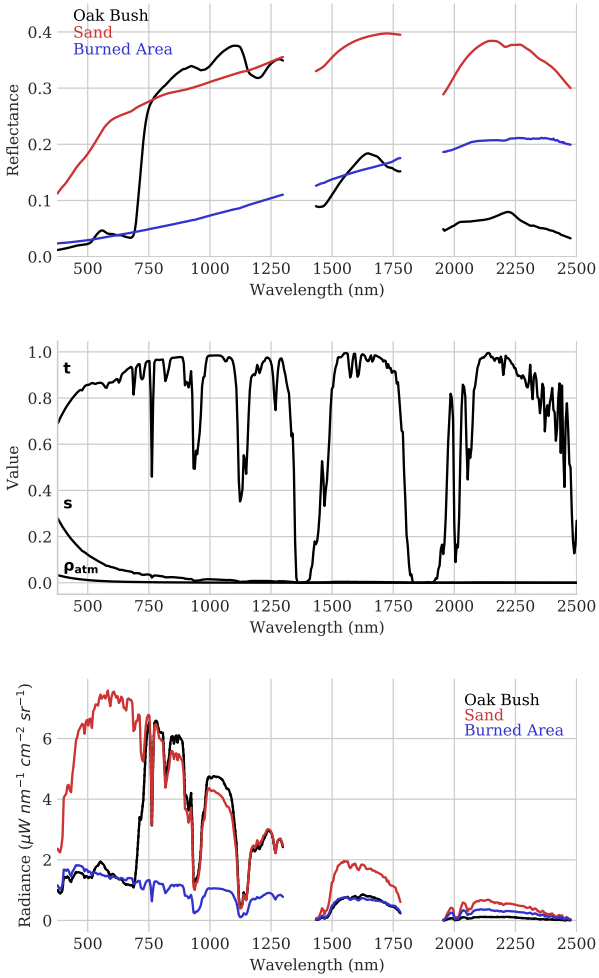

Figure 1: Example spectra from the simulation. Top: USGS library reflectances. Middle: Atmospheric coefficients. Bottom: Associated radiances.

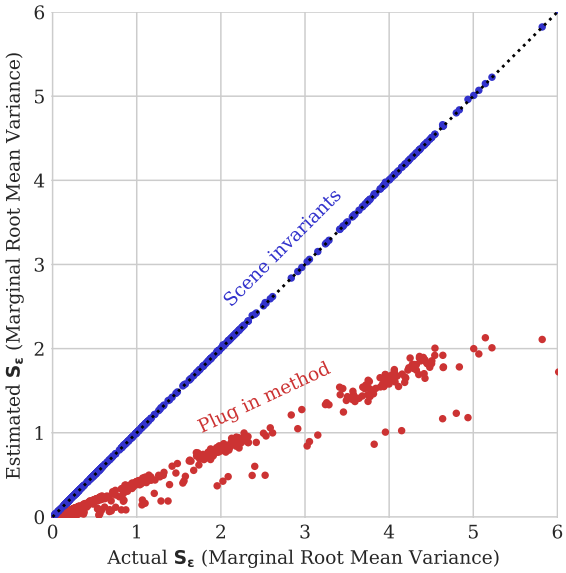

Figure 2: Comparison of plug in and scene invariants methods for estimating $\mathbf{S}_\epsilon$, based on 109 diverse library spectra.

the instrument noise distribution ϵ is ℓw. However, the effective number of samples from δ is considerably fewer, because atmospheres are smooth over short spatial scales, and any errors related to model discrepancy will be highly correlated within an acquisition. The number of samples from δ is closer to ℓ. If one selects an equal number of invariant locations from each acquisition, all instantiations of δ are represented and the draws from δ are protected against selection bias. But the same cannot be said for ϵ, which has a different realization for every spectrum. Can the draws from ϵ induce underestimation of model discrepancy, with instrument noise fixing the RTM error by chance?

In fact, the magnitude and direction of the two random variables are quite different, meaning that ϵ is highly unlikely to negate δ. Instrument noise ϵ is mostly isotropic - in other words, each channel manifests noise independently. In contrast, model discrepancy errors have spectral structure with most of their variance along a handful of principal component directions. This is easiest to demonstrate in simulation. Here we consider two potential sources of model discrepancies: the vertical profiles of gas concentrations, pressures and temperatures; and the optical properties of atmospheric aerosols. Our profile simulation compares the radiances produced by an Air Force Geophysics Laboratory (AFGL) midlatitude summer and tropical models (Anderson et al., 1986), holding constant all other atmospheric parameters such as the column water vapor. Our aerosol simulation compares radiances produced by a highly scattering aerosol, the sulfate model described above, and an absorbing aerosol based on the fine particle smoke model described in Brodrick et al. (2020) and Omar et al. (2004). This smoke model was originally derived from the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) mission. All simulations use an aerosol optical depth of 0.2 at 550 nm. Using Equation 13, we transform each USGS reflectance spectrum to radiance at sensor using both sets of atmospheric assumptions. This creates artificial model discrepancy errors due to vertical profiles and aerosol properties.

Figure 3 shows the three eigenvectors of the residual covariance matrices associated with the highest eigenvalues. We divide them by the radiance of a 20% reflectance surface for display. The top panel shows the residuals resulting from aerosol optical property discrepancies, while the bottom panel shows the residuals from profile discrepancies. The two are qualitatively different; the vertical profile residuals are largest around water absorption features, reflecting differences in absorption coefficients at different altitudes. The aerosol-related residuals form broadband shapes and slopes, peaking in the shorter wavelengths. In both cases, however, the top three eigenvalues contain approximately 99% of the variance for the dataset.

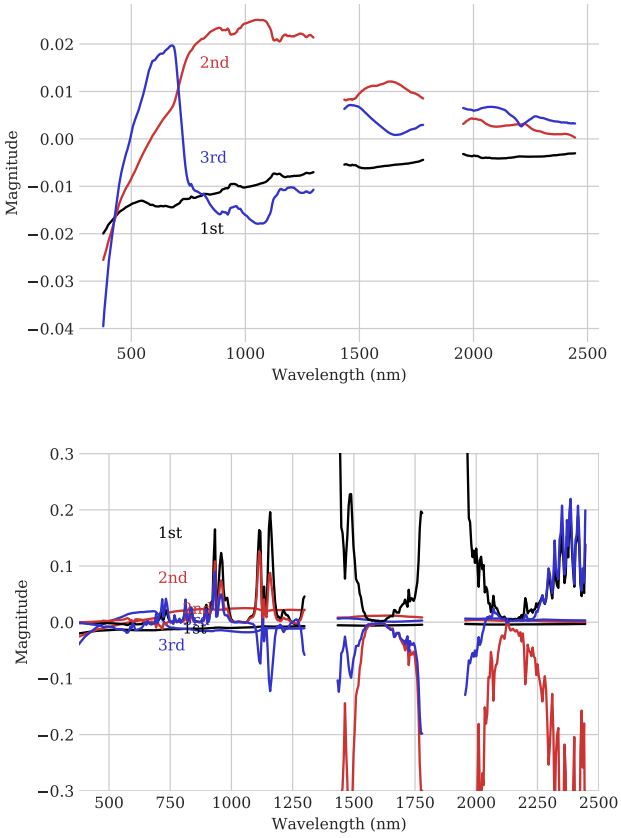

Figure 3: Principal components of radiance residuals for two classes of model discrepancy error. Lines illustrate the top three eigenvectors. Magnitudes are ratioed against a reference radiance for clarity. Top: Errors due to misestimation of aerosol optical properties. Bottom: Errors due to misestimation of the vertical distributions and temperatures of atmospheric gases.

To demonstrate the effect on opportunistic discovery, we simulate repeat acquisitions of invariant targets. We use the USGS spectra as spectrally-invariant reference locations, synthesizing radiances at sensor for two overpasses. The aerosol optical properties differ between the first and second overflights. In the

first overflight they are the scattering optical type; in the second the aerosol is the absorbing type. The retrieval always presumes the scattering aerosol, so this induces a model discrepancy between the inversion's RTM assumptions and the simulated reality. The difference between the two acquisitions is a single draw from δ. The true atmospheric state vector is held constant with an aerosol optical depth of 0.4 and a columnar water vapor content of 1.75. We add synthetic noise to each simulated at-sensor radiance following a signal-dependent AVIRIS-NG noise model (Thompson et al., 2018b), perform the inversion on all spectra, and then select a subset of the most similar results as if we were attempting to estimate the model discrepancy from flight data. We then compare the estimated discrepancy to that which would have been obtained if the entire dataset had been used.

Figure 4 shows the resulting estimated root mean square discrepancy between the two ensembles as a fraction of its true value. The single most similar spectrum from each ensemble, i.e. the "luckiest" subpopulation of invariant locations, underestimates the true variance by only 25%. This is significantly better than the status quo alternative which ignores model discrepancy entirely. In addition, accuracy improves as a larger fraction of the invariant subpopulation is used. This corroborates the principal component analysis suggesting that instrument noise is mostly orthogonal to RTM errors. In other words, we can safely use the most similar subpopulation of spectrum pairs without significantly underestimating $\hat{\mathbf{\Gamma}}$. We conclude that taking the most similar spectra from flightline pairs a useful strategy for opportunistic discovery of invariant surfaces.

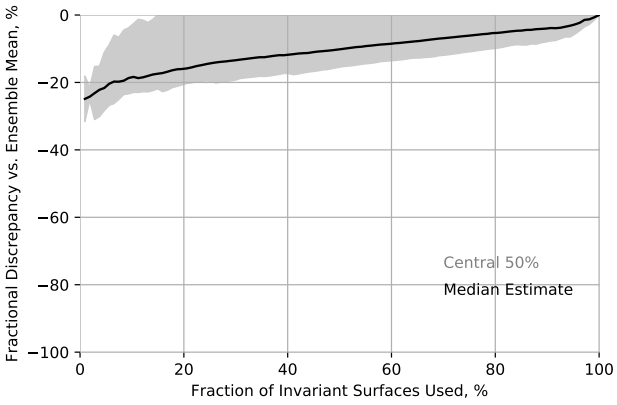

Figure 4: Median errors in the estimated mean discrepancy for an ensemble of invariant surfaces, using the subpopulation of locations which has the most similar estimated reflectance across two consecutive flightlines. The grey area represents the 90% enclosing envelope.

3. Experimental approach

This section describes an experiment to estimate $\mathbf{\Gamma}$ using repeat overflights of heterogeneous terrain. We then test the hypothesis that including the model discrepancy estimate will improve the accuracy of retrieval uncertainty predictions. We analyze data from AVIRIS-NG, a Visible to Shortwave Infrared (VSWIR) instrument spanning 380-2500 nm at approximately 5 nm spectral resolution (Chapman et al., 2019). A 2015-2018 AVIRIS-NG campaign spans numerous locations in the Indian Subcontinent (Bhattacharya et al., 2019). Targets include a wide range of natural and anthropogenic surfaces, addressing diverse disciplines from geology to cryosphere studies to terrestrial and aquatic ecology. Typical Ground Sampling Distances (GSDs) are 5-8 m. Critically for our study, several flight lines are repeated across multiple years under different observing geometries. This, combined with variable and challenging atmospheres throughout the campaign, makes it an ideal test dataset for quantifying RTM model errors. We select 33 locations listed in Table B.3 for which repeat acquisitions are available. The flightlines are typically 5-7 km in width and 10-20 km in length, and each contain on the order of 10^6 spectra. Solar zeniths span the range from 14 to 44 degrees. The nadir-pointed AVIRIS-NG field of view spans approximately 30 degrees, leading to a wide range of relative azimuth angles.

To perform the analysis, we first calibrate the data spectrally and radiometrically using laboratory and field techniques in Chapman et al. (2019). We then coregister each pair of overlapping radiance cubes by projecting them onto a common geometric grid with nearest-neighbor interpolation. This initial coregistration is based on the instrument camera model and onboard Global Positioning System / Inertial Measurement Unit (GPS/IMU) readings of the aircraft position. We trace the view path of each focal plane array element to a digital elevation model to identify its ground footprint. Error in the model-based registration is typically 10-20 m due to the intrinsic GPS accuracy. Consequently, we perform a second image-based adjustment that uses scene content to find the ideal alignment. The registration algorithm evaluates the similarity between a grid of subwindows drawn from both flightlines, with different horizontal and vertical shifts applied to each subwindow. We then fit a composite affine transformation to the set of best translations, warping the earlier images onto the later. This procedure generally fits the images to within subpixel accuracy over most of the flightline area.

After coregistration, we use the atmospheric correction routine described in section 2.4 to estimate the surface reflectance at each pixel. Atmospheric assumptions are similar to those used for the India campaign dataset in Thompson et al. (2019), but our aerosol model is simplified; here we use only the sulfate aerosol type, a highly scattering profile that consistently outperforms the other options across a wide range of atmospheres. The retrieved magnitude of the Lambertian-equivalent surface reflectance varies due to bidirectional reflectance effects from changing observation geometries. Most of this change is simply a uniform photometric scaling of the reflectance. We remove it by scaling the spectra so that the areas under the curves match those in the first flightline. We filter out all water pixels, recognized by near infrared reflectances less than 0.1, and vegetation pixels, recognized by a Normalized Difference Vegetation Index greater than 0.5, since these are unlikely to be true invariant surfaces. We select the top 1000 most similar locations from each pair to use as the invariant surfaces, calculate their noise in each spectrum, and finally estimate $\mathbf{\Gamma}$ using Equa-

tion 12. The 33 pairs of flightlines sample at least 66 unique airmasses, so though each $\bar{\mathbf{x}}$ has some error due to small sample sizes, we still benefit from large sample sizes for estimating $\mathbf{\Gamma}$.

After finding the model discrepancy distribution, we apply the estimated $\mathbf{\Gamma}$ using a held out dataset for which ground truth is available. Here we turn to an AVIRIS-NG overflight of the Jet Propulsion Laboratory (JPL) on June 2014. Several days after the overflight, under similar solar geometry, a ground team measured four spectrally-stable surfaces in the scene: open patches of concrete, exposed roofing material, and loose bare soil. Figure 5 shows the target locations in the scene. At each location the field team measured an area approximately 10x10 m in size, subtending approximately 3x3 remote pixels. Radiance measurements from each surface were acquired using Malvern Panalytical ASD field spectrometers, and then ratioed against the upwelling radiance from a tripod-mounted reference panel. A bubble leveling tool was used to ensure that the panels were level.

To assess the discrepancy estimates we first extract remote radiance measurements from each target location. We perform a reflectance inversion once assuming $\mathbf{S}_\epsilon$ consists only of instrument noise covariance $\mathbf{\Lambda}$, and a second time where $\mathbf{S}_\epsilon$ includes both $\mathbf{\Lambda}$ and the estimated model discrepancy covariance $\hat{\mathbf{\Gamma}}$. We compare observed errors to posterior uncertainty predictions by calculating the adjusted residual error:

$$r = (\hat{\boldsymbol{\rho}}_s - \boldsymbol{\rho}_g)^T (\hat{\mathbf{S}} + \mathbf{S_g})^{-1} (\hat{\boldsymbol{\rho}}_\mathbf{s} - \boldsymbol{\rho}_\mathbf{g}) \quad (14)$$

where $\hat{\boldsymbol{\rho}}_s$ is the estimated surface reflectance, $\boldsymbol{\rho}_g$ is the in situ measurement, and $\mathbf{S_g}$ is the covariance of the uncertainty in the in situ data. As in Thompson et al. (2018b), $\mathbf{S_g}$ is a diagonal matrix including 1% error in radiometric calibration and 1% error in the spectralon bidirectional reflectance function, added in quadrature. If for a spectrum with n channels the errors actually follow the predicted uncertainty $\hat{\mathbf{S}}$, then the score r follows a chi square distribution with $n-1$ degrees of freedom. This translates naturally to a hypothesis test to accept or reject the inversion posterior uncertainty as an explanation for the observed differences between remote and in situ measurements.

4. Results

The India dataset shows that invariant surfaces are present in diverse scenes. Figures 6 and 7 show typical examples of natural and built up areas. The top half of the scene in Figure 6 is dominated by a fire scar that appeared between the two overflights. The remaining terrain shows subtle changes in vegetation coverage. The frame at right shows the root mean square differences between overflights, with green representing the filtered vegetation pixels. The fire scar appears as an area of highest change. The most stable area of the scene is the bare soil around the lake at the upper right which changes very little across the two years. Other isolated areas of stability appear as smaller red areas throughout the image. The most volatile location in the scene is the fire scar itself.

Figure 8 shows spectra from representative locations, labeled A, B, and C in the image. Site A has the least spectral change. It comes from a lakeside location and is a good qualitative match to dry marsh sediment (DWV3-0511) from the United States Geologic Survey (USGS) spectral library (Kokaly et al., 2017). The largest discrepancy between the two reflectance spectra appears at the edges of deep water absorption features at 1400 and 1880 nm, and at the shortest wavelengths below 500 nm. These errors suggest minor model discrepancies in H_2O vapor and atmospheric scattering, respectively. The middle panel shows the spectrum at the 0.1 percentile of similarity, comparable to the cutoff used to find invariant surfaces in the India flightlines. This location, labeled B, is an isolated patch of bare soil containing visible-wavelength absorptions consistent with iron oxides. Together these provide evidence that top 0.1% most similar pixels are indeed invariant, i.e. that they are effectively as close as the two most similar spectra. The only differences, small departures at the shortest wavelengths and near deep water features, are identical for both the most similar and 0.1 percentiles, indicating that they are the kinds of systematic model- and calibration-related errors we aim to estimate. The bottom panel shows the most different location, labeled C, as a counterpoint. It lies in an area in the fire scar transformed from a partly-vegetated area to bare charred soil.

Figure 5: A portion of the JPL flightline ang20140612t215342 used for the validation experiment. Locations of measured sites appear as white Roman numerals.

Figure 7 shows a contrasting site near an urban area. Here, the terrain is highly altered with many artificial structures. The right panel indicates that many of these structures are spectrally stable across the two years. Other islands of stability include bare patches of soil or fallow fields. Areas of high change includes fields under cultivation and new construction. Figure 9 shows representative spectra from the urban scene. The most similar location, A, appears to be bare soil. The two spectra show minor departures throughout the visible and shortwave range. Slight inversion errors are visible at the water vapor absorption features at 940 and 1140 nm, respectively. As in the wilderness case, RTM errors are sufficient to explain for the qualitative differences between the spectra. Spectrum B shows

the 0.1 percentile, which again appears invariant. The bottom panel shows location C, the most changed location between the overflights. It transitions from dark wet mud to bare soil.

After applying Equation 12 to the complete dataset, the resulting model discrepancy covariance Γ is much larger than the instrument noise. Figure 10 shows the square root of marginal variances in each channel, as a fraction of a typical radiance spectrum, for both $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$. The model discrepancy term is larger by a factor of 5-10. Discrepancy error is slightly elevated in the non-saturated water absorption features at 940 and 1140 nm, where differences in vertical profile or aerosol-related distortions can modulate the shapes of these features in ways that the retrieval may not capture. These issues are even more acute at the edges of deep water absorption features at 1420 nm, 1880 nm, and 3000 nm. The highest RTM uncertainty is found in the visible wavelengths, which are most strongly affected by atmospheric scattering. Minor differences in aerosol properties or estimated loading may shift the slope of spectra in this range. The chlorophpyll "red edge" is visible as peak near 720 nm. This feature could represent a failure of the invariant surface assumption, if sparse vegetation were introduced between flightlines, but it is also possible that this represents an adjacency effect caused by photons scattered from nearby vegetation under high particulate loadings. Such adjacency effects were not considered by the retrieval algorithm, so they constitute a legitimate source of model discrepancy error.

We obtain additional insight by evaluating the singular value decomposition of $\mathbf{\Gamma}$. Unlike the instrument noise, which is spectrally uncorrelated, $\mathbf{\Gamma}$ is low rank with high cross channel correlations. The top three eigenvectors contain 90% of the variance, and the top ten enclose 99%. Figure 11 shows the top eigenvectors, each scaled by the square root of its eigenvalue and expressed as a fraction of a typical radiance spectrum. The primary principal component, represented by the black line, shows a constant offset of the near infrared portion of the spectrum combined with a dramatic and opposing change in the shortest wavelengths. There is almost no change at the longest wavelengths. This feature may be related to errors in spectral contrast caused by particulate optical properties or changes in the radiometric calibration. The second principal component in red combines a short wavelength scattering feature with a modulation of the shortwave water bands. The shortwave features are very similar to structure observed in vertical profile errors. The third principal component in blue shows a "red edge" feature combined with a steep slope in shallow wavelengths. This resembles features of aerosol-related errors from the simulation in Figure 3. It is also possible that fractional differences in vegetation, or the introduction of vegetation spectra as an adjacency effect, contribute to this feature.

The JPL validation spectra confirm that incorporating the estimated Γ improves the fidelity of posterior uncertainty predictions as judged against field data. Both old and new retrievals produce a best estimate prediction and a posterior uncertainty, i.e. a confidence interval. We test the consistency of each posterior with the realized error; table 2 shows the p values associated with each posterior distribution. The discrepancies are sufficient to reject the $\boldsymbol{\epsilon}$ only model at for the building roof site (II), but the $\boldsymbol{\epsilon} + \boldsymbol{\delta}$ model is consistent at all four sites.

Site	Surface	RMSE	p $(\boldsymbol{\epsilon})$	p $(\boldsymbol{\epsilon} + \boldsymbol{\delta})$
I.	Mars Yard	0.012	1.0	1.0
II.	Building Roof	0.020	<0.01	1.0
III.	Old Parking Lot	0.012	1.0	1.0
IV.	Fresh Parking Lot	0.015	0.35	1.0

Table 2: Reflectance RMSE and p values associated with the JPL validation test.

The improvement in the spectral uncertainty envelopes is more vivid. Figures 12 to 14 show the retrieval and error predictions for each measured location. Since our posterior includes a full covariance matrix with correlations, the classical strategy of drawing "error bars" on the reflectance is inadequate. We portray the predicted posterior uncertainty $\hat{\mathbf{S}}$ by plotting random samples from this distribution in grey. The left panels show the maximum a posteriori estimate for the remote retrieval using $\boldsymbol{\epsilon}$ only. The instrument noise distribution is compact, so the grey and black lines nearly overlay each other in the left panels. The red line signifies the in situ estimate. The ideal is for the grey lines to envelope the red line, indicating the confidence prediction is consistent with the actual errors. Here it lies well outside the envelope of uncertainty $\hat{\mathbf{S}}$. Formally, three of the four cases are still statistically consistent due to the uncertainty in the ASD measurement. Nevertheless, it is likely that the posterior predicted uncertainty overestimates the inversion accuracy. The right panels of Figures 12 to 15 show the same results using an observation noise estimate that includes both $\boldsymbol{\epsilon}$ and the new $\boldsymbol{\delta}$. Here, the posterior predicted uncertainty is more consistent with the correlated errors of real spectra. The posterior more accurately envelopes the observed difference between remote and in situ data. Surprisingly, water vapor absorption residuals near 940 and 1140 nm are also improved. The retrieval now recognizes higher systematic RTM errors in those areas of the spectrum and is able to weight the measurement data appropriately.

5. Discussion

These experiments demonstrate a practical approach for operational quantification of RTM uncertainty. Naturally, it is only one of many tools available. When discrepancies are well understood, such as the case of numerical approximation errors, a range of bottom-up tools exist for evaluating the measurement power of the observation. These include analytical methods such as degree of freedom analyses and simulation-based methods such as Monte Carlo estimation. We do not intend to replace these approaches. Instead, we aim to provide a data-driven alternative by which aggregate forward model errors can be quantified for large datasets. Both the reductive analytical approaches and the data-driven empirical approaches can play a role to advance the science of imaging spectroscopy.

For simplicity we use a generic $\boldsymbol{\delta}$ that is the same for all observing conditions. This is an improvement over conventional methods, which ignore $\boldsymbol{\delta}$ altogether, and is sufficient to demonstrate a closed uncertainty budget for held-out test cases.

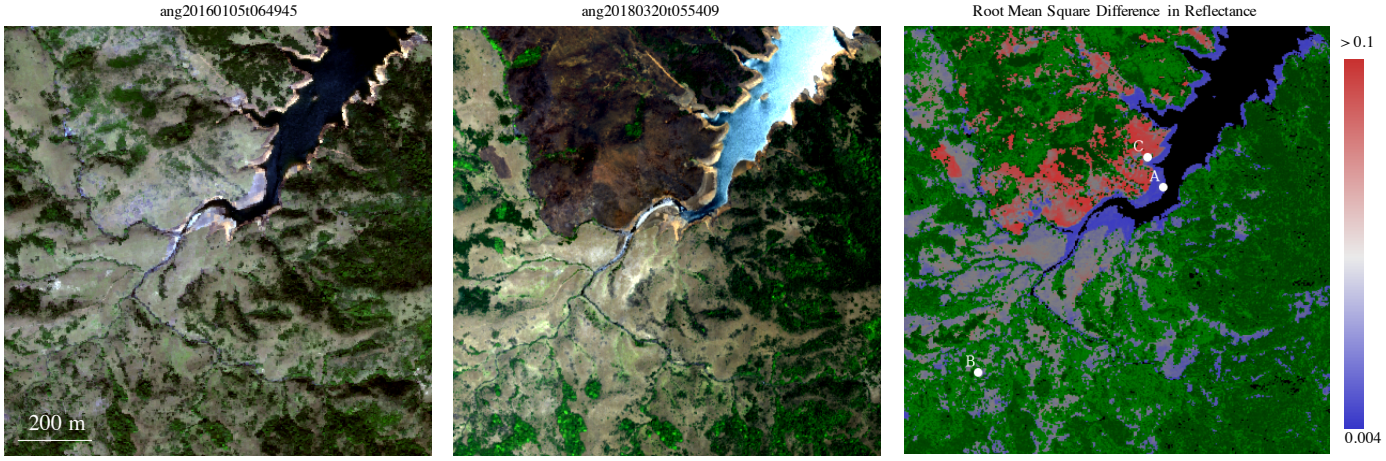

Figure 6: A subset of flightlines ang201605t051247 and ang20180320t042247, acquired in 2016 and 2018 respectively, after coregistration. The rightmost panel shows the root mean squared differences between spectra. Green areas represent terrain that is filtered due to vegetation in one or both scenes. Location A has the least spectral change, location B shows the 0.1 percentile, and Location C has the most spectral change.

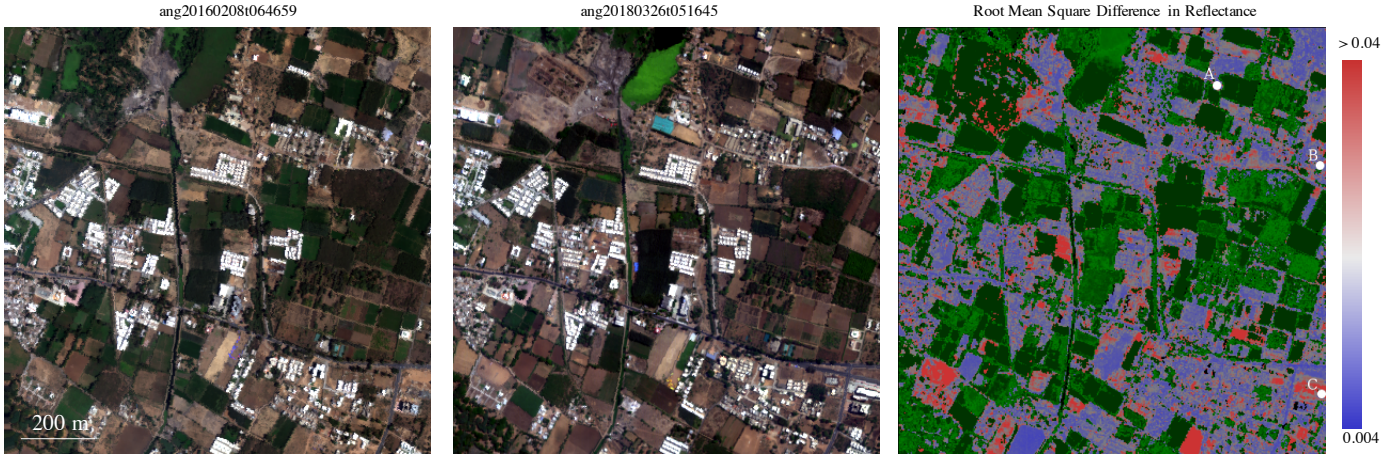

Figure 7: A subset of flightlines ang20160208t064659 and ang20180326t051645, acquired in 2016 and 2018 respectively, after coregistration. The rightmost panel shows the root mean squared differences between spectra. Green areas represent terrain that is filtered due to vegetation in one or both scenes. Location A has the least spectral change, location B shows the 0.1 percentile, and Location C has the most spectral change.

However, a more sophisticated analysis could improve our result. First, instead of simply combining together all uncertainties that are not instrument noise, one could further partition $\mathbf{\Gamma}$ into more specific and interpretable components. For example, calibration errors have a specific correlation structure and can be estimated independently; they might be treated separately from atmospheric errors. These additional uncertainty components might be scene dependent, allowing uncertainty to vary as a function of local observing conditions. One can account for them uncertainties using a linearized forward model (Rodgers, 2000; Connor et al., 2016) by defining a covariance $\mathbf{S_b}$ over unknown variables $\mathbf{b}$, and Jacobians $\boldsymbol{K}_b$ describing the change in radiance per unit change in the unknown variables:

$$\mathbf{S}_\epsilon = \mathrm{cov}(\mathbf{y} - \mathbf{F}(\mathbf{x}) \mid \mathbf{x}) = \mathbf{\Gamma} + \mathbf{\Lambda} + \mathbf{K_b S_b K_b^T} \quad (15)$$

with similar adjustments to the derived expressions. Now $\mathbf{\Gamma}$ contains only the model discrepancy from $\Delta \boldsymbol{f}(\mathbf{x}, \mathbf{b})$, and any unknown variance not modeled explicitly in the $\mathbf{S_b}$ matrix. Bookkeeping phenomena in $\mathbf{S_b}$ instead of $\mathbf{\Gamma}$ represents an improvement in understanding and a more accurate uncertainty prediction.

A second way future studies might improve our result is to use a larger dataset. Here we estimate model discrepancies using 33 pairs of flightlines. In theory, the model discrepancy covariance $\mathbf{\Gamma}$ depends on the underlying distribution of $\mathbf{x}$ and $\mathbf{b}$, so it will be more accurate for scenes that are similar to the India dataset. In practice, we find that the $\mathbf{\Gamma}$ estimate improves our test scene despite dramatic differences in atmosphere and content. This bodes well for its generality. However, a larger base dataset would permit a more granular partitioning of observing conditions, letting the user estimate a $\mathbf{\Gamma}$ which is matched to the specific content of the new observation. Future studies like the Surface Biology and Geology investigation will cover the globe multiple times per month, providing enough data to repeat this experiment over 10000 times each week. Estimating location-specific $\mathbf{\Gamma}$ for specific regions will become possible. Larger

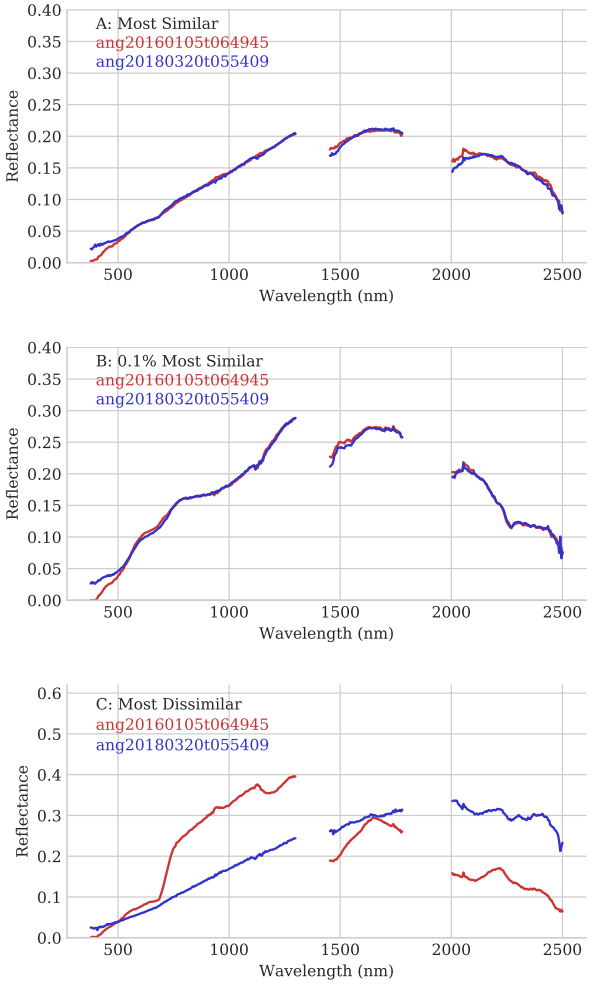

Figure 8: Spectra drawn from locations A, B, and C in Figure 6. Location A shows the least spectral change across the two flightlines. Location B represents the top 0.1 percentile. Location C has the greatest change between overflights.

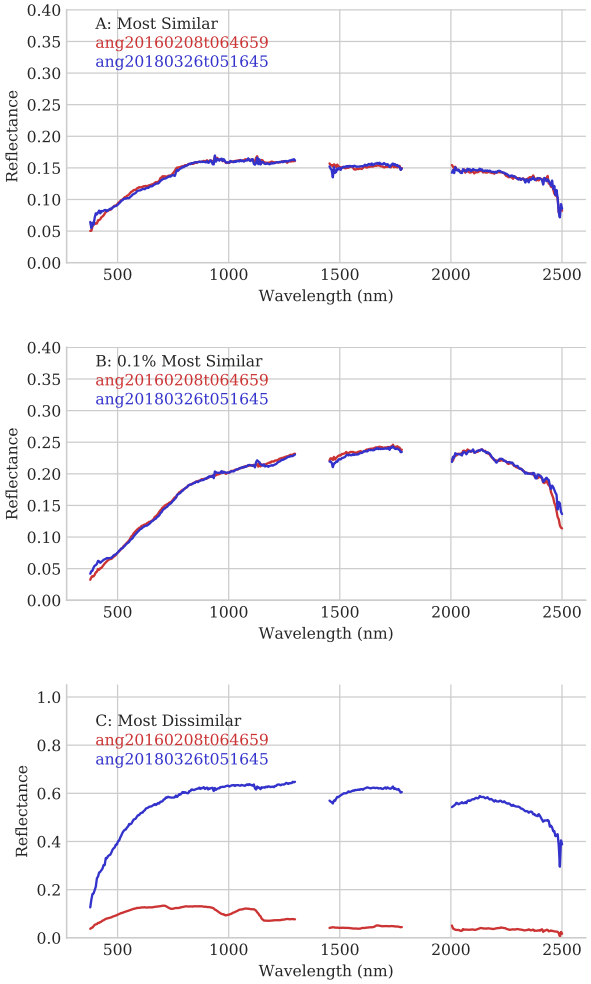

Figure 9: Spectra drawn from locations A and B in Figure 7. Location A shows the least spectral change across the two flightlines. Location B represents the top 0.1 percentile. Location C has the greatest change between overflights.

datasets might also capture other uncertainties such as distortions from sub-visible clouds or cirrus, which were absent from the India campaign data but will be important at global scales.

In addition to improving error predictions, there is potential to improve the reflectance estimates themselves. One could use opportunistic invariant surfaces as if they were ground calibration points (Furby & Campbell, 2001; Schott et al., 1988). Previous acquisitions under clear conditions could provide good quality reference estimates, which could then act as a reference to erase errors incurred later under poor conditions. One could simply fit a linear transformation to correct the distortions as in Clark et al. (2002). Such a procedure could be made fully automatic to scale globally if needed. Alternatively, work is underway to evaluate time series data to see if a combined multitemporal retrieval similar to the MAIAC approach (Lyapustin et al., 2018) can provide more accurate inversions. Even without such measures the JPL reference case demonstrates that historical model discrepancy estimates can also improve retrievals for scenes taken individually by downweighting spectrum fitting error in directions that are likely to have model discrepancies.

There is some similarity between our proposed method and an existing approach which models systematic errors using empirical orthogonal functions, or EOFs. The EOF approach estimates principal components of the inversion residuals, adds the dominant directions to the forward model, and retrieves their magnitudes as part of the state vector. This compensates for systematic errors in calibration or modeling and improves the spectrum fit. Such approaches have been used effectively in missions such as the Orbiting Carbon Observatory 2 (Boesch et al., 2015). Here we show that this does not perfectly represent systematic error in the forward model; it is closer to a plug in approach that models the part of the systematics which the retrieval cannot already absorb into the state vector. In this work, we use scene invariants to derive a more direct estimate of $\mathbf{\Gamma}$, and represent the systematics as a noise distribution rather than a null subspace in the forward model. A probabilistic interpretation means systematic error can be propagated appropriately to posterior confidence estimates.

Scene invariant properties certainly exist for other domains

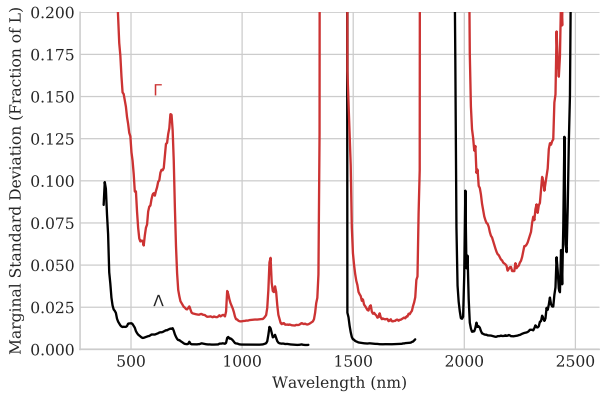

Figure 10: Marginal standard deviations of different uncertainty sources as a fraction of a typical radiance spectrum. The red line indicates uncertainty due to model discrepancy errors with covariance $\mathbf{\Gamma}$. The black line indicates noise due to instrument noise with covariance $\mathbf{\Lambda}$.

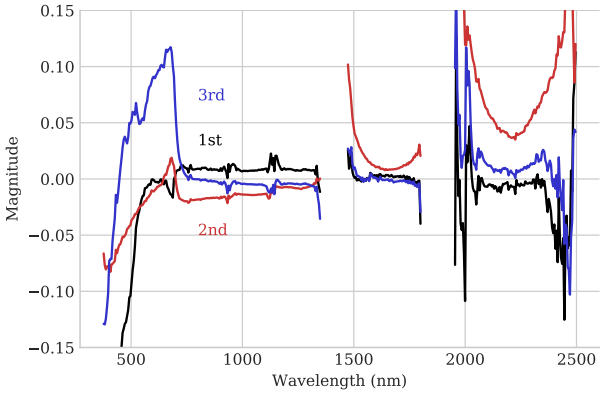

Figure 11: The top three eigenvectors in the order black, red, and blue, each scaled by the square root of its eigenvalue and expressed as a fraction of a typical radiance spectrum. These principal components contain 90% of the estimated model discrepancy variance.

outside atmospheric correction. Snow or water surfaces are not temporally invariant, but they may be spatially smooth at certain lengthscales. Investigators might use *spatial* ensembles in place of temporal ones, sampling $\mathbf{\Gamma}$ from scenes containing visible artifacts or spatial discontinuities. As another example, estimates of spatially smooth atmospheric properties can show discontinuities due to interference from the underlying surface reflectance. A contiguous spatial group of atmospheric state retrievals, that encloses one or more such artifacts, would be a suitable dataset for Equation 12. Finally, while we have demonstrated the technique in solar reflected regimes, they may also be relevant for surface retrievals in the thermal infrared where the emitted radiance is a function of both emissivity and temperature. It is probable that scenes contain some subset of pixels with stable surface emissivity. If these can be identified, the time-variable temperature parameter can be retrieved conditioned upon the invariant emissivity. This is the direct analogue of the variable atmospheric state vector elements in the VSWIR. Scene invariant properties can thus be a general tool for decomposing and estimating model discrepancy error across a wide range of imaging spectroscopy disciplines.

6. Conclusion

This article presents an approach for quantifying uncertainty in Radiative Transfer Models (RTMs) based on the statistical properties of large diverse imaging spectroscopy datasets. We describe methods for finding model discrepancy based on scene invariants - properties of the environment which are stable over time. We provide a practical estimator for transforming multiple consecutive observations of the same location where one or more elements of the state vector are expected to be invariant. We explore the properties of this estimator using the case study of atmospheric correction, evaluating performance analytically and in simulation. We demonstrate that it is possible to discover invariant surfaces in flight based on the retrieved state vector, and that one need not identify the population exactly to get an accurate estimate of model discrepancy noise. Finally, we calculate the actual model discrepancy error for a standard inversion method using a large catalogue of data from the AVIRIS-NG India campaigns of 2016 and 2018. We apply the model discrepancies estimated from the India dataset to an independent validation experiment, where they improve retrieval accuracy and the fidelity of posterior uncertainty predictions vis a vis in situ data. Estimation of RTM modeling uncertainty will be important to improve operational uncertainty predictions for future global spectroscopy missions.

7. Acknowledgements

We thank the members of the AVIRIS-NG team who participated in data acquisition and analysis, including Mark Helmlinger, John Chapman, Winston Olson-Duvall and Sarah Lundeen. AVIRIS-NG is sponsored by the National Aeronautics and Space Administration (NASA) Earth Science Division. This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. N. Carmon and J. Susiluoto's participation in research was supported by the Jet Propulsion Laboratory Earth Science Advanced Concepts program. We also acknowledge the support of the NASA Earth Science Division's AVIRIS-NG instrument and the data analysis program "Utilization of Airborne Visible/Infrared Imaging Spectrometer Next Generation Data from an Airborne Campaign in India" NNH16ZDA001N-AVRSNG, under program manager Woody Turner.

Appendix A. Derivation of Residual Covariances

This section derives the covariances of the model fitting residuals in Equations 10 and 11. We first define the gain matrix $\mathbf{G}$, an $n \times m$ matrix with elements containing $\partial\mathbf{R}/\partial\mathbf{y}$; it

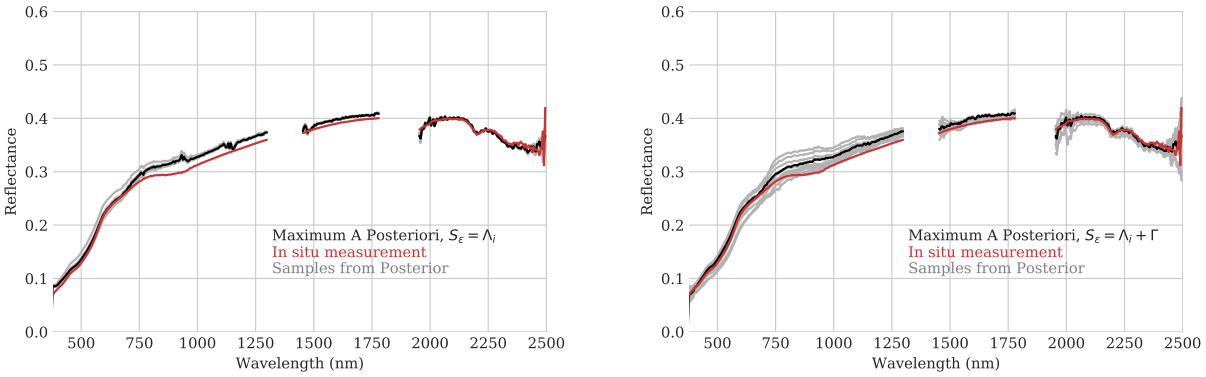

Figure 12: Samples from the posterior distribution incorporating $\boldsymbol{\epsilon}$ (left) and $\boldsymbol{\epsilon} + \boldsymbol{\delta}$ (right) for validation site I.

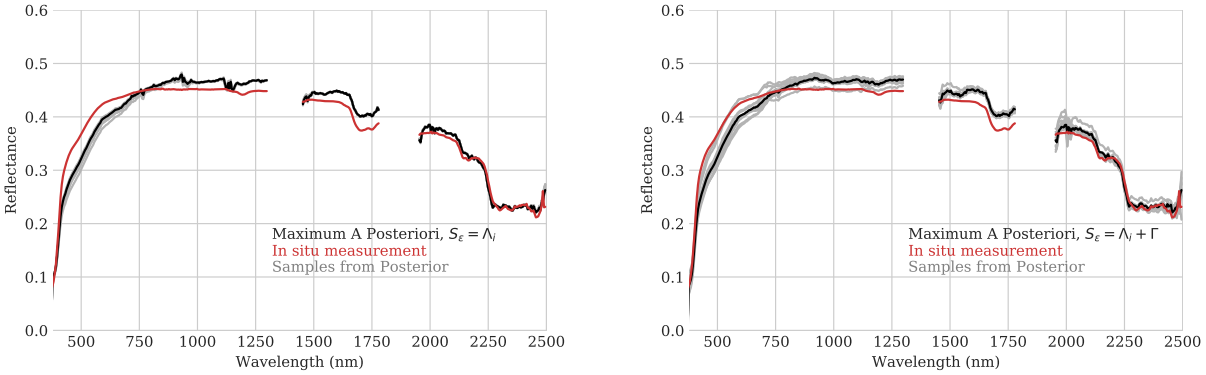

Figure 13: Samples from the posterior distribution incorporating $\boldsymbol{\epsilon}$ (left) and $\boldsymbol{\epsilon} + \boldsymbol{\delta}$ (right) for validation site II.

represents the sensitivity of the retrieval to a change in the measurement.

$$\mathbf{G} = \mathbf{S}_a \mathbf{K}^T (\mathbf{K}\mathbf{S}_a\mathbf{K}^T + \mathbf{S}_\epsilon)^{-1} \tag{A.1}$$

The gain matrix is related to the Jacobian and averaging kernel matrices:

$$\mathbf{A} = \mathbf{G}\mathbf{K} \tag{A.2}$$

The error for a single retrieval, $\hat{\mathbf{x}} - \mathbf{x}$, adapted from Rodgers (2000), is:

$$\hat{\mathbf{x}} - \mathbf{x} = (\mathbf{A} - \mathbf{I}_n)(\mathbf{x} - \mathbf{x}_a) + \boldsymbol{G}(\boldsymbol{\delta} + \boldsymbol{\epsilon}) \tag{A.3}$$

where $\mathbf{I_n}$ is the $n \times n$ identity matrix. The first term to the right of the equality sign captures the imperfect resolution of the retrieval algorithm as the averaging kernel departs from the identity. The second term indicates the influence of model discrepancy and sensor noise on the retrieval. For an ensemble of states drawn from the prior, and an optimal inversion algorithm that correctly balances the prior and measurement noise, the residuals $\mathbf{y} - \mathbf{F}(\hat{\mathbf{x}})$ are:

$$\begin{aligned} \mathbf{y} - \mathbf{F}(\hat{\mathbf{x}}) &= \mathbf{K}[\mathbf{x} - \hat{\mathbf{x}}] + \boldsymbol{\delta} + \boldsymbol{\epsilon} && \text{(A.4)} \\ &= \mathbf{K}[(\mathbf{I}_n - \mathbf{A})(\mathbf{x} - \mathbf{x}_a) - \boldsymbol{G}(\boldsymbol{\delta} + \boldsymbol{\epsilon})] + \boldsymbol{\delta} + \boldsymbol{\epsilon} && \text{(A.5)} \end{aligned}$$

This assumes a forward model that is linearized locally at $\mathbf{x}$. Again following Rodgers (2000), the fit residuals have the following covariance:

$$\mathrm{cov}(\mathbf{y} - \mathbf{F}(\hat{\mathbf{x}}) \mid \hat{\mathbf{x}}) = (\mathbf{I} - \mathbf{K}\mathbf{G})(\mathbf{K}\mathbf{S}_\mathrm{a}\mathbf{K}^\mathrm{T} + \mathbf{S}_\epsilon)(\mathbf{I} - \mathbf{K}\mathbf{G})^\mathrm{T} \tag{A.6}$$

For an optimal retrieval we substitute Equation A.1 to obtain:

$$\mathrm{cov}(\mathbf{y} - \mathbf{F}(\hat{\mathbf{x}}) \mid \hat{\mathbf{x}}) = \mathbf{S}_\epsilon(\mathbf{K}\mathbf{S}_\mathrm{a}\mathbf{K}^\mathrm{T} + \mathbf{S}_\epsilon)^{-1}\mathbf{S}_\epsilon \tag{A.7}$$

which appears as Equation 10.

For the scene invariants approach the model fitting error is similar to Equation A.3, but in the average of a large ensemble, the zero mean random variables $\boldsymbol{\delta}$ and $\boldsymbol{\epsilon}$ go to zero via the law of large numbers. Removing them from the expression yields:

$$\bar{\mathbf{x}} - \mathbf{x} = (\mathbf{A} - \mathbf{I}_n)(\mathbf{x} - \mathbf{x}_a) \tag{A.8}$$

The residual radiance fitting error is simply:

$$\begin{aligned} \mathbf{y} - \mathbf{F}(\bar{\mathbf{x}}) &= \mathbf{K}[\mathbf{x} - \bar{\mathbf{x}}] + \boldsymbol{\delta} + \boldsymbol{\epsilon} && \text{(A.9)} \\ &= \mathbf{K}[(\mathbf{I}_n - \mathbf{A})(\mathbf{x} - \mathbf{x}_a)] + \boldsymbol{\delta} + \boldsymbol{\epsilon} && \text{(A.10)} \end{aligned}$$

This is closely related to Equation A.5, but the expression in brackets, which relates to the retrieval process, no longer contains the noise sources. Intuitively, using the mean of a multi-observation series $\bar{\mathbf{x}}$ prevents overfitting the noise of any one

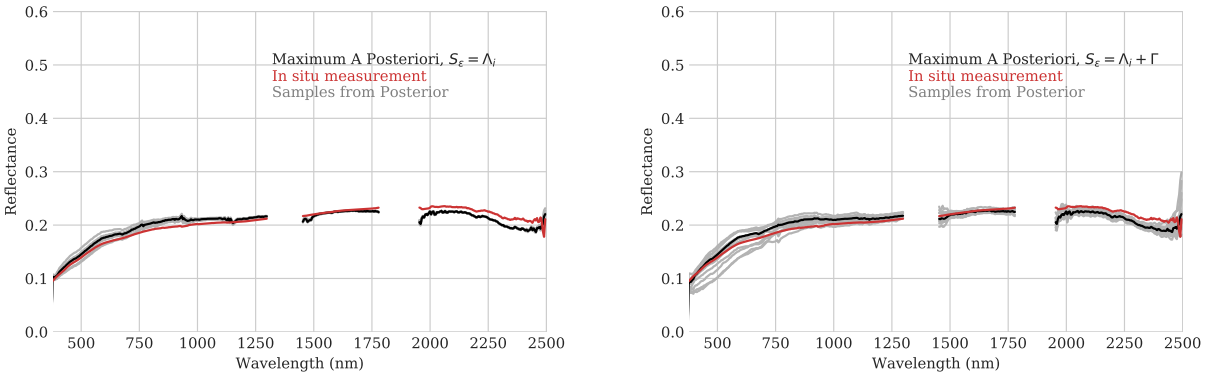

Figure 14: Samples from the posterior distribution incorporating $\boldsymbol{\epsilon}$ (left) and $\boldsymbol{\epsilon} + \boldsymbol{\delta}$ (right) for validation site III.

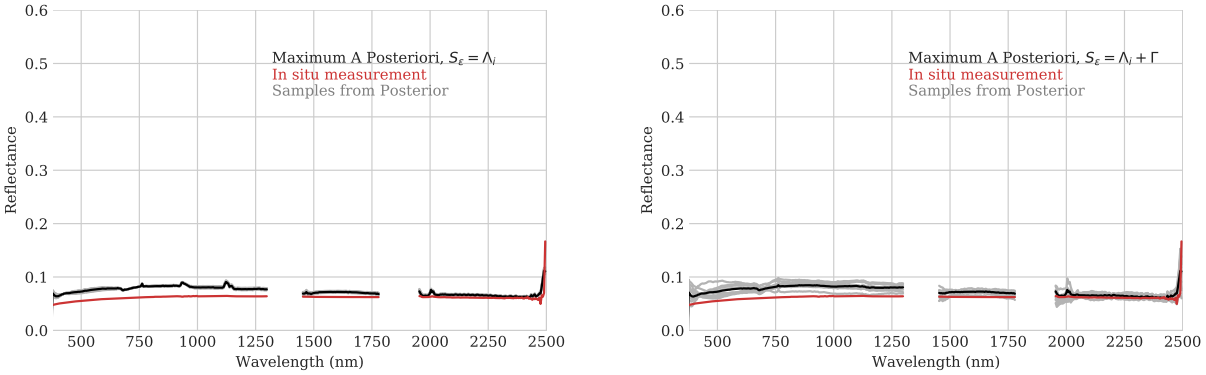

Figure 15: Samples from the posterior distribution incorporating $\boldsymbol{\epsilon}$ (left) and $\boldsymbol{\epsilon} + \boldsymbol{\delta}$ (right), for each surface in the JPL validation experiment IV.

observation. For an ensemble of states drawn from the prior distribution, the residual covariance is:

$$\text{cov}(\mathbf{y} - \mathbf{F}(\bar{\mathbf{x}}) \mid \bar{\mathbf{x}}) = \mathbf{K}(\mathbf{I}_\text{n} - \mathbf{A})\mathbf{S}_\text{a}(\mathbf{I}_\text{n} - \mathbf{A})^\text{T}\mathbf{K} + \text{S}_\epsilon \qquad \text{(A.11)}$$

thereby demonstrating Equation 11.

Appendix B. India Flightlines

Table B.3 shows a list of the India campaign flightlines used in this experiment.

References

Anderson, G. P., Clough, S. A., Kneizys, F., Chetwynd, J. H., & Shettle, E. P. (1986). *AFGL atmospheric constituent profiles (0.120 km)*. Technical Report AIR FORCE GEOPHYSICS LAB HANSCOM AFB MA.

Bhattacharya, B. K., Green, R. O., Rao, S., Saxena, M., Sharma, S., Kumar, K. A., Srinivasulu, P., Sharma, S., Dhar, D., Bandyopadhyay, S. et al. (2019). An overview of aviris-ng airborne hyperspectral science campaign over india. *Current Science*, *116*, 1082–1088.

Boesch, H., Brown, L., Castano, R., Christi, M., Crisp, D., Eldering, A., Fisher, B., Frankenberg, C., Gunson, M., Granat, R. et al. (2015). Orbiting carbon observatory (OCO)-2 level 2 full physics retrieval algorithm theoretical basis. *Jet Propulsion Laboratory Technical Document*, *D-55207, Version 3.0*.

Bouvet, M. (2014). Radiometric comparison of multispectral imagers over a pseudo-invariant calibration site using a reference radiometric model. *Remote sensing of environment*, *140*, 141–154.

Bouvet, M., Thome, K., Berthelot, B., Bialek, A., Czapla-Myers, J., Fox, N. P., Goryl, P., Henry, P., Ma, L., Marcq, S. et al. (2019). Radcalnet: A radiometric calibration network for earth observing imagers operating in the visible to shortwave infrared spectral range. *Remote Sensing*, *11*, 2401.

Braverman, A., Hobbs, J., Teixiera, J., & Gunson, M. (2020). Post-hoc uncertainty quantification for remote observing systems. *Submitted to SIAM/ASA Journal on Uncertainty Quantification*, .

Brodrick, P. G., Thompson, D. R., Kalashnikova, O., & Garay, M. (2020). Remote sensing of wildfire smoke optical properties with the aviris-ng imaging spectrometer. *Remote sensing of environment*, *in preparation*.

Brynjarsdóttir, J., & O'Hagan, A. (2014). Learning about physical parameters: the importance of model discrepancy. *Inverse Problems*, *30*, 114007. URL: `https://doi.org/10.1088%2F0266-5611%2F30%2F11%2F114007`. doi:`10.1088/0266-5611/30/11/114007`.

Bue, B. D., Thompson, D. R., Deshpande, S., Eastwood, M., Green, R. O., Natraj, V., Mullen, T., & Parente, M. (2019). Neural network radiative transfer for imaging spectroscopy. *Atmospheric Measurement Techniques*, *12*.

Candela, A., Thompson, D. R., Wettergreen, D., Cawse-Nicholson, K., Geier, S., Eastwood, M. L., & Green, R. O. (2020). Probabilistic super resolution for mineral spectroscopy. In *AAAI* (pp. 13241–13247).

Canty, M. J., Nielsen, A. A., & Schmidt, M. (2004). Automatic radiometric normalization of multitemporal satellite imagery. *Remote Sensing of Environment*, *91*, 441–451.

Carmon, N., Thompson, D. R., Susiluoto, J., Bohn, N., Connelly, D. S., Turmon, M. J., Cawse-Nicholson, K., Green, R. O., & Gunson, M. R. (2020). Uncertainty quantification for a global imaging spectroscopy surface composition investigation. *Remote Sensing of Environment*, *in press*.

Flight ID (2016)	Solar zenith θ_s (2016)	Flight ID (2018)	Solar zenith θ_s (2018)	Longitude	Latitude	Elevation
ang20160303t063225	28	ang20180512t040324	28	87.142	21.023	1
ang20160127t075428	36	ang20180224t073528	26	75.374	14.416	381
ang20160205t060857	43	ang20180331t090745	36	75.151	24.413	285
ang20160208t060316	42	ang20180326t052601	33	70.796	20.411	38
ang20160205t062340	42	ang20180331t092019	39	75.185	24.417	293
ang20160208t061104	41	ang20180326t055043	28	71.283	20.818	40
ang20160127t080253	37	ang20180224t074617	27	75.228	14.238	378
ang20160105t070345	34	ang20180320t060717	17	75.488	10.520	1287
ang20160205t055344	45	ang20180331t085541	34	75.202	24.470	279
ang20160127t081224	38	ang20180224t075707	28	75.270	14.279	384
ang20160208t062935	39	ang20180326t055821	27	71.344	20.854	41
ang20160127t085223	43	ang20180224t084052	34	75.320	14.261	380
ang20160105t062045	35	ang20180320t052750	26	75.465	10.515	1237
ang20160105t073139	35	ang20180320t062955	14	75.522	10.547	1310
ang20160208t070843	36	ang20180328t065631	19	72.898	20.940	272
ang20160127t084302	42	ang20180224t082952	32	75.513	14.455	380
ang20160208t081158	39	ang20180328t075945	22	72.964	20.921	294
ang20160208t074610	37	ang20180328t074721	20	72.961	20.935	292
ang20160208t063818	39	ang20180326t054307	29	70.960	20.526	39
ang20160208t061908	40	ang20180326t053523	31	70.936	20.526	38
ang20160105t052624	41	ang20180320t043543	38	75.559	10.631	1047
ang20160105t051247	43	ang20180320t042247	41	75.538	10.642	1018
ang20160127t083154	40	ang20180224t081906	31	75.366	14.343	388
ang20160127t082153	39	ang20180224t080824	29	75.380	14.355	384
ang20160105t064945	34	ang20180320t055409	20	75.495	10.523	1311
ang20160105t055235	37	ang20180320t050110	32	75.592	10.647	1116
ang20160208t064659	38	ang20180326t051645	34	71.187	20.819	38
ang20160205t063811	41	ang20180331t093315	42	75.167	24.400	306
ang20160205t065233	41	ang20180331t094559	44	75.199	24.397	311
ang20160210t080717	38	ang20180326t065943	20	70.609	21.525	91
ang20160208t075849	38	ang20180328t072141	18	72.919	20.927	279
ang20160208t072124	36	ang20180328t073435	19	72.937	20.940	281
ang20160208t073349	36	ang20180328t070902	18	72.906	20.924	279

Table B.3: List of flightlines used in the validation experiment. Fligthtline IDs show the 8 digit date (year, month, day) followed by the 6 digit UTC time (hour, minute, second).Columns show solar zenith angles and the mean longitude, latitude, and elevation of each site.

Chapman, J. W., Thompson, D. R., Helmlinger, M. C., Bue, B. D., Green, R. O., Eastwood, M. L., Geier, S., Olson-Duvall, W., & Lundeen, S. R. (2019). Spectral and radiometric calibration of the next generation airborne visible infrared spectrometer (aviris-ng). *Remote Sensing*, *11*, 2129.

Clark, R. N., Swayze, G. A., Livo, K. E., Kokaly, R. F., King, T. V., Dalton, J. B., Vance, J. S., Rockwell, B. W., Hoefen, T., & McDougal, R. R. (2002). Surface reflectance calibration of terrestrial imaging spectroscopy data: a tutorial using aviris. In *Proceedings of the 10th Airborne Earth Science Workshop* (pp. 02–1). Jet Propulsion Laboratory Pasadena, CA.

Clough, S., Shephard, M., Mlawer, E., Delamere, J., Iacono, M., Cady-Pereira, K., Boukabara, S., & Brown, P. (2005). Atmospheric radiative transfer modeling: a summary of the aer codes. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *91*, 233–244.

Connor, B., Bösch, H., McDuffie, J., Taylor, T., Fu, D., Frankenberg, C., O'Dell, C., Payne, V. H., Gunson, M., Pollock, R. et al. (2016). Quantification of uncertainties in oco-2 measurements of xco 2: Simulations and linear error analysis. *Atmospheric Measurement Techniques*, *9*, 5227.

Cosnefroy, H., Leroy, M., & Briottet, X. (1996). Selection and characterization of saharan and arabian desert sites for the calibration of optical satellite sensors. *Remote Sensing of Environment*, *58*, 101–114.

Dozier, J., Green, R. O., Nolin, A. W., & Painter, T. H. (2009). Interpretation of snow properties from imaging spectrometry. *Remote Sensing of Environment*, *113*, S25–S37.

Furby, S., & Campbell, N. (2001). Calibrating images from different dates to 'like-value' digital counts. *Remote Sensing of Environment*, *77*, 186 – 196.

Gao, B.-C., Heidebrecht, K. B., & Goetz, A. F. (1993). Derivation of scaled surface reflectances from aviris data. *Remote sensing of Environment*, *44*, 165–178.

Gao, B.-C., Kaufman, Y. J., Han, W., & Wiscombe, W. J. (1998). Corection of thin cirrus path radiances in the 0.4–1.0 μm spectral region using the sensitive 1.375 μm cirrus detecting channel. *Journal of Geophysical Research: Atmospheres*, *103*, 32169–32176.

Gordon, I. E., Rothman, L. S., Hill, C., Kochanov, R. V., Tan, Y., Bernath, P. F., Birk, M., Boudon, V., Campargue, A., Chance, K. et al. (2017). The hitran2016 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *203*, 3–69.

Green, R. O., Mahowald, N., Ung, C., Thompson, D. R., Bator, L., Bennet, M., Bernas, M., Blackway, N., Bradley, C., Cha, J., Clark, P., Clark, R., Cloud, D., Diaz, E., Ben Dor, E., Duren, R., Eastwood, M., Ehlmann, B. L., Fuentes, L., Ginoux, P., Gross, J., He, Y., Kalashnikova, O., Kert, W., Keymeulen, D., Klimesh, M., Ku, D., Kwong-Fu, H., Liggett, E., Li, L., Lundeen, S., Makowski, M. D., Mazer, A., Miller, R., Mouroulis, P., Oaida, B., Okin, G. S., Ortega, A., Oyake, A., Nguyen, H., Pace, T., Painter, T. H., Pempejian, J., Garcia-Pando, C. P., Pham, T., Phillips, B., Pollock, R., Purcell, R., Realmuto, V., Schoolcraft, J., Sen, A., Shin, S., Shaw, L., Soriano, M., Swayze, G., Thingvold, E., Vaid, A., & Zan, J. (2020). The earth surface mineral dust source investigation: An earth science imaging spectroscopy mission. *2020 IEEE Aerospace Conference*, (pp. 1–15).

Griffin, M. K., & Burke, H.-h. K. (2003). Compensation of hyperspectral data for atmospheric effects. *Lincoln Laboratory Journal*, *14*, 29–54.

Hobbs, J., Braverman, A., Cressie, N., Granat, R., & Gunson, M. (2017). Simulation-based uncertainty quantification for estimating atmospheric co _2 from satellite data. *SIAM/ASA Journal on Uncertainty Quantification*, *5*, 956–985.

Jiménez, M., de la Cámara, O. G., Moncholí, A., & Muñoz, F. (2018). Towards a complete spectral reflectance uncertainty model for field spectroscopy. *Fifth recent advances in quantitative remote sensing*, (p. 21).

Kobayashi, H., & Iwabuchi, H. (2008). A coupled 1-d atmosphere and 3-d canopy radiative transfer model for canopy reflectance, light environment, and photosynthesis simulation in a heterogeneous landscape. *Remote Sensing of Environment*, *112*, 173–185.

Kokaly, R. F., Clark, R. N., Swayze, G. A., Livo, K. E., Hoefen, T. M., Pearson, N. C., Wise, R. A., Benzel, W. M., Lowers, H. A., Driscoll, R. L. et al. (2017). *USGS spectral library version 7*. Technical Report US Geological

Survey.

Kokaly, R. F., Despain, D. G., Clark, R. N., & Livo, K. E. (2003). Mapping vegetation in yellowstone national park using spectral feature analysis of aviris data. *Remote sensing of environment*, *84*, 437–456.

Kulawik, S. S., O'Dell, C., Nelson, R. R., & Taylor, T. E. (2019). Validation of oco-2 error analysis using simulated retrievals. *Atmospheric Measurement Techniques*, *12*.

Lu, P., Zhang, H., & Li, J. (2009). A comparison of two-stream disort and eddington radiative transfer schemes in a real atmospheric profile. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *110*, 129–138.

Lucey, P. G., & Noble, S. K. (2008). Experimental test of a radiative transfer model of the optical effects of space weathering. *Icarus*, *197*, 348–353.

Lyapustin, A., Wang, Y., Korkin, S., & Huang, D. (2018). Modis collection 6 maiac algorithm. *Atmospheric Measurement Techniques*, *11*, 5741–5765. URL: https://amt.copernicus.org/articles/11/5741/2018/. doi:10.5194/amt-11-5741-2018.

Meerdink, S. K., Roberts, D. A., Roth, K. L., King, J. Y., Gader, P. D., & Koltunov, A. (2019). Classifying california plant species temporally using airborne hyperspectral imagery. *Remote Sensing of Environment*, *232*, 111308. doi:https://doi.org/10.1016/j.rse.2019.111308.

Mobley, C. D. (1994). *Light and water: radiative transfer in natural waters*. Academic press.

National Academies of Sciences, Engineering, and Medicine (2018). *Thriving on our changing planet: A decadal strategy for Earth observation from space*. National Academies Press. doi:https://doi.org/10.17226/24938.

Nicodemus, F. E. (1970). Reflectance nomenclature and directional reflectance and emissivity. *Applied Optics*, *9*, 1474–1475.

Omar, A. H., Winker, D. M., & Won, J.-G. (2004). Aerosol models for the calipso lidar inversion algorithms. In *Laser Radar Technology for Remote Sensing* (pp. 153–164). International Society for Optics and Photonics volume 5240.

Richter, R., Kellenberger, T., & Kaufmann, H. (2009). Comparison of topographic correction methods. *Remote Sensing*, *1*, 184–196.

Richter, R., Schläpfer, D., & Müller, A. (2010). Operational atmospheric correction for imaging spectrometers accounting for the smile effect. *IEEE Transactions on Geoscience and Remote sensing*, *49*, 1772–1780.

Rockwell, B. W., Clark, R. N., Cunningham, C. G., Sutley, S. J., Gent, C., McDougal, R. R., Livo, K. E., & Kokaly, R. F. (2000). Mineral mapping in the marysvale volcanic field, utah using aviris data. In *Summaries of the 9th Annual JPL Airborne Earth Science Workshop* (pp. 407–417). Citeseer.

Rodgers, C. D. (2000). *Inverse methods for atmospheric sounding: theory and practice* volume 2. World scientific.

Sanders, L. C., Schott, J. R., & Raqueno, R. (2001). A vnir/swir atmospheric correction algorithm for hyperspectral imagery with adjacency effect. *Remote Sensing of Environment*, *78*, 252–263.

Schaepman-Strub, G., Schaepman, M., Painter, T., Dangel, S., & Martonchik, J. (2006). Reflectance quantities in optical remote sensing—definitions and case studies. *Remote Sensing of Environment*, (p. 27–42). doi:10.1016/j.rse.2006.03.002.

Schott, J. R., Salvaggio, C., & Volchok, W. J. (1988). Radiometric scene normalization using pseudoinvariant features. *Remote sensing of Environment*, *26*, 1–16.

Taramelli, A., Tornato, A., Magliozzi, M. L., Mariani, S., Valentini, E., Zavagli, M., Costantini, M., Nieke, J., Adams, J., & Rast, M. (2020). An interaction methodology to collect and assess user-driven requirements to define potential opportunities of future hyperspectral imaging sentinel mission. *Remote Sensing*, *12*, 1286.

Thome, K. (2016). Calibration/validation error budgets, uncertainties, traceability and their importance to imaging spectrometry. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, (pp. 1912–1915).

Thompson, D. R., Babu, K., Braverman, A. J., Eastwood, M. L., Green, R. O., Hobbs, J. M., Jewell, J. B., Kindel, B., Massie, S., Mishra, M. et al. (2019). Optimal estimation of spectral surface reflectance in challenging atmospheres. *Remote Sensing of Environment*, *232*, 111258.

Thompson, D. R., Boardman, J. W., Eastwood, M. L., Green, R. O., Haag, J. M., Mouroulis, P., & Van Gorp, B. (2018a). Imaging spectrometer stray spectral response: In-flight characterization, correction, and validation. *Remote Sensing of Environment*, *204*, 850 – 860. URL: http://www.sciencedirect.com/science/article/pii/S0034425717304261. doi:https://doi.org/10.1016/j.rse.2017.09.015.

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. et al. (2020). Quantifying uncertainty for remote spectroscopy of surface composition. *Remote Sensing of Environment*, *247*, 111898.

Thompson, D. R., Natraj, V., Green, R. O., Helmlinger, M. C., Gao, B.-C., & Eastwood, M. L. (2018b). Optimal estimation for imaging spectrometer atmospheric correction. *Remote Sensing of Environment*, *216*, 355–373.

Voogt, J. A., & Oke, T. R. (2003). Thermal remote sensing of urban climates. *Remote sensing of environment*, *86*, 370–384.

Wasserman, L. (2006). *All of nonparametric statistics*. Springer Science & Business Media.

Wunch, D., Wennberg, P., Toon, G., Frankenberg, C., Keppel-Aleks, G., & Roehl, C. (2011). A method for evaluating bias in global measurements of co_2 total columns from space. *Atmospheric Chemistry and Physics*, *11*, 12317–12337.