A new computational framework for atmospheric and surface remote sensing

Doğan A. Timuçin

NASA Ames Research Center, Mail Stop 269-3, Moffett Field, California 94035, USA
(650) 604-1262 • timucin@ptolemy.arc.nasa.gov

Abstract— **A Bayesian data-analysis framework is described for atmospheric and surface retrievals from remotely-sensed hyper-spectral data. Some computational techniques are highlighted for improved accuracy in the forward physics model.**

I. Introduction

In the Earth-Science (ES) community, work has been done for decades on the retrieval of atmospheric and surface properties from flux and radiance measurements gathered by dedicated sensors, ranging from multi-spectral passive instruments to active devices such as microwave radar and lidar. These instruments have traditionally been designed to operate in certain spectral "windows" where components of interest have well-known and isolated signatures; meanwhile, theorists were able to develop ingenuous approximations to the associated forward radiative-transfer models. This combination of "smart" observations and "simplified" physics allowed the use of relatively simple and computationally affordable retrieval algorithms in remote sensing. Although some loss of accuracy was inevitably incurred, the computational resources of the past decades simply could not handle the problem in a more direct way. This, of course, is no longer true; yet, these "legacy" techniques are still deeply ingrained in the atmospheric radiative-transfer community, and continue to artificially limit our retrieval capabilities.

The situation is exacerbated by the rapid and largely uncoordinated growth in the number and variety of remote-sensing instruments operated by a multitude of institutions around the globe. The somewhat independent characterization and deployment of these instruments, and the subsequent, rather disconnected analyses of their data sets have led to serious cross-validation issues that remain largely unresolved. Therefore, the establishment of a unified framework for instrument characterization and data analysis is crucial if we are to utilize the data volume of our ever-growing "sensor web" in the most efficient and accurate manner. Given the potential impact on society, as well as the policy-makers, of our measurements and predictions, our retrievals must truly represent "the best estimate of the scientific community."

The trend in Earth observation has been shifting recently towards passive hyper-spectral imaging instruments. With hundreds of spatial pixels and hundreds of spectral channels, these sensors observe the terrestrial atmosphere and surface in space, time, and wavelength, thus producing complete "snapshots" of the dynamical Earth system in the form of "data cubes." As such, they provide vital information for assessing global changes in cloud and aerosol properties, abundance of atmospheric water vapor, ozone, and carbon dioxide, sea and land surface temperatures and albedos, *etc.* Such sensors can simultaneously satisfy the needs of many communities that have heretofore felt compelled to build their own expensive custom instruments for every new mission.

The potential for simultaneous measurement and retrieval of a large set of atmospheric and surface properties by a single (type of) instrument also brings with it the need for more comprehensive and accurate forward models. For if an entire data cube is to be utilized for its information content, then the entire array of physical processes that might have influenced the data must be efficiently accounted for. Currently, too many investigators are continuing to use rudimentary techniques such as band ratios in analyzing hyper-spectral data. We believe that our computational power is now ready for more accurate implementations of the forward physics as well as more rigorous retrieval algorithms. Under the Intelligent Data Understanding Project of the NASA CICT Program, we have been developing a Bayesian data-analysis framework that is expected to greatly enhance our understanding of, as well as our ability to model and predict, the Earth system. In this paper, we describe the salient features of this framework along with some details of the forward physics model where improvements have been made to the state of the practice.

II. Retrieval Theory

We begin with the basic statement of our problem. Denote the unknown parameters underlying an experiment as $\mathbf{x} = \{x_n,\ n = 1, 2, \ldots, N\}$, and the measurements made during the experiment as $\mathbf{y} = \{y_m,\ m = 1, 2, \ldots, M\}$. The "standard" (static) measurement model relating $\mathbf{x}$ and $\mathbf{y}$ is of the form

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\epsilon}, \tag{1}$$

where the vector-valued function $\mathbf{f} : \mathbb{R}^N \mapsto \mathbb{R}^M$ represents the typically nonlinear "forward" physics model for the experiment, and $\boldsymbol{\epsilon} = \{\epsilon_m,\ m = 1, 2, \ldots, M\}$ denotes the ubiquitous measurement error, comprising the systematic bias as well as the stochastic noise of the measuring instrument. The "inverse" retrieval problem, then, is that of estimating the state vector $\mathbf{x}$ from a given measurement vector $\mathbf{y}$ [1], [2].

A. Bayesian inference

In the Bayesian philosophy [3], [4], one's knowledge of a quantity x is represented by an associated probability density function (PDF) $p(x)$; the sharper $p(x)$ is around some value x^*, the more confident we are that $x \simeq x^*$ in actuality (see Figure 1). For the experimental scenario described above, Bayes' theorem asserts that [3]-[6]

$$p(\mathbf{x}|\mathbf{y}) = \frac{\ell(\mathbf{y}|\mathbf{x})\, p(\mathbf{x})}{p(\mathbf{y})} = \frac{\ell(\mathbf{y}|\mathbf{x})\, p(\mathbf{x})}{\int \ell(\mathbf{y}|\mathbf{x})\, p(\mathbf{x})\, d\mathbf{x}}. \tag{2}$$

Here, $p(\mathbf{x})$ and $p(\mathbf{x}|\mathbf{y})$ are, respectively, the *a priori* and *a posteriori* PDFs – the prior and the posterior – of the state vector $\mathbf{x}$, and $\ell(\mathbf{y}|\mathbf{x})$ is the likelihood of the measurement vector $\mathbf{y}$. The form of the likelihood function represents the solution of the underlying modeling problem. The prior, meanwhile, encapsulates our initial "best guess" and associated uncertainty about the unknown parameters. Upon measurement, this prior is transformed into the posterior via Bayes' theorem; a sharper posterior (relative to the prior) indicates an improved confidence on our part as to the value of $\mathbf{x}$ after having seen the data $\mathbf{y}$.

A decision strategy is needed next to extract from $p(\mathbf{x}|\mathbf{y})$ an optimal estimate for $\mathbf{x}$. We adopt the maximum *a posteriori* estimate $\mathbf{x}^*$ given by the mode of the posterior (*i.e.*, that value of $\mathbf{x}$ for which the posterior is maximized). Defining a suitable "cost" function[1]

$$J(\mathbf{x}) \equiv -\ln \ell(\mathbf{y}|\mathbf{x}) - \ln p(\mathbf{x}), \tag{3}$$

[1] Any monotonically non-increasing function of the posterior would do here; our particular choice has the advantage of turning the multiplicative form of (2) into the more convenient additive form of (3), as well as of dispensing with the term $p(\mathbf{y})$, which does not depend on $\mathbf{x}$ and therefore has no bearing on the optimal estimate. Further simplifications result from the use of the natural logarithm if the exponential family of PDFs are adopted for the likelihood and the prior; see §II-B.

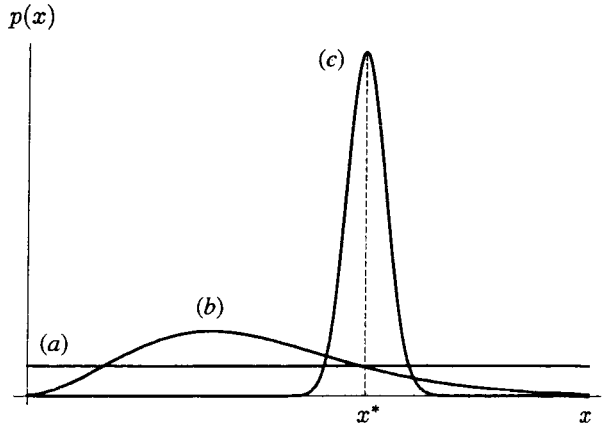

Fig. 1. The Bayesian representation of knowledge. The uniform PDF (*a*) indicates complete uncertainty about x, while the sharp PDF (*c*) reflects a high degree of certainty that the value of x is close to the position of the peak, x^*. The broad PDF (*b*) is an intermediate case: the position of the peak may still be taken as a reasonable guess for x; however, there is now considerable uncertainty about this estimate, signified by the much larger spread of the distribution around its peak.

we therefore seek the global **minimum** of J:

$$J(\mathbf{x}^*) \leq J(\mathbf{x}) \quad \forall\, \mathbf{x} \neq \mathbf{x}^*. \tag{4}$$

Finally, the "width" of the posterior, however it is defined, provides a quantitative measure of uncertainty – an "error cloud" (*i.e.*, a multi-dimensional error bar) at the tip of the estimated state vector $\mathbf{x}^*$. This width tends to shrink with increasing number of measurements, a phenomenon usually referred to as "Bayesian learning." Expanding the log-posterior into a Taylor series around $\mathbf{x}^*$, we obtain

$$\ln p(\mathbf{x}|\mathbf{y}) = \ln p(\mathbf{x}^*|\mathbf{y}) - \frac{1}{2}\, (\mathbf{x}-\mathbf{x}^*)^{\mathrm{T}}\, \mathbf{S}^{*\,-1}\, (\mathbf{x}-\mathbf{x}^*) + \ldots. \tag{5}$$

Note that the first-order term is absent by virtue of (4). The eigenvalues and eigenvectors of the matrix $\mathbf{S}^*$ so defined serve naturally as the desired measure of uncertainty around $\mathbf{x}^*$.

B. Gaussian case

It is generally acceptable to model the statistics of the measurement error with a Gaussian PDF; thus, $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\delta}, \boldsymbol{\Sigma})$, with $\mathcal{N}$ signifying the normal distribution. Here, the mean vector $\boldsymbol{\delta}$ represents the average offset of the instrument while the covariance matrix $\boldsymbol{\Sigma}$ characterizes its statistical fluctuations; see §III-A. Consequently, one has the likelihood function

$$\ell(\mathbf{y}|\mathbf{x}) \propto \mathrm{e}^{-\frac{1}{2}\,[\mathbf{y}-\mathbf{f}(\mathbf{x})-\boldsymbol{\delta}]^{\mathrm{T}}\, \boldsymbol{\Sigma}^{-1}\, [\mathbf{y}-\mathbf{f}(\mathbf{x})-\boldsymbol{\delta}]}. \tag{6}$$

It is also common, albeit more questionable, to adopt a Gaussian PDF for the prior:

$$p(\mathbf{x}) \propto \mathrm{e}^{-\frac{1}{2}\,(\mathbf{x}-\boldsymbol{\xi})^{\mathrm{T}}\, \boldsymbol{\Xi}^{-1}\, (\mathbf{x}-\boldsymbol{\xi})}, \tag{7}$$

where the mean vector $\boldsymbol{\xi}$ and the covariance matrix $\boldsymbol{\Xi}$ respectively represent one's best guess for and the associated uncertainty about $\mathbf{x}$ before the experiment is conducted.

Substituting (6) and (7) into (3), we obtain the cost function

$$\begin{aligned} J(\mathbf{x}) = {} & \frac{1}{2}\, [\mathbf{y} - \mathbf{f}(\mathbf{x}) - \boldsymbol{\delta}]^{\mathrm{T}}\, \boldsymbol{\Sigma}^{-1}\, [\mathbf{y} - \mathbf{f}(\mathbf{x}) - \boldsymbol{\delta}] \\ & + \frac{1}{2}\, (\mathbf{x} - \boldsymbol{\xi})^{\mathrm{T}}\, \boldsymbol{\Xi}^{-1}\, (\mathbf{x} - \boldsymbol{\xi}) + C, \end{aligned} \tag{8}$$

where C is an unimportant constant. Several methods are available for minimizing such a cost function. The simplest of these (*e.g.*, simulated annealing, steepest descent) are also the most inefficient, whereas the use of sophisticated methods (*e.g.*, conjugate gradients, quasi-Newton) may be a computational overkill if the optimization problem at hand is only mildly difficult. We will describe in §II-C an elegant and effective approach that is the method of choice for cost functions of the form (8) with moderately nonlinear forward models.

We note in passing that the use of Gaussian PDFs is typically justified on "holy grounds" such as the central-limit theorem [5] or the maximum-entropy principle [6]; the real reason for their use in many investigations is simply the ensuing mathematical tractability of the analysis, and too often this is done without careful attention being paid to the limits of validity of this assumption. It is clear from (8), however, that in the case of a nonlinear forward model, the assumption

of Gaussianity does **not** yield a simple (*i.e.*, quadratic) cost function. There is therefore no reason to shy away from using non-Gaussian likelihood functions and priors that may more faithfully represent the experimental reality.

C. Optimization approach

The assumption of moderate nonlinearity suggests that one can linearize the forward model around an "operating point" $\mathbf{x}_0$ as

$$\mathbf{f}(\mathbf{x}) \simeq \mathbf{f}(\mathbf{x}_0) + \mathbf{K}(\mathbf{x}_0)\,(\mathbf{x} - \mathbf{x}_0), \tag{9}$$

where the $M \times N$ Jacobian matrix is defined as

$$\mathbf{K}(\mathbf{x}) \equiv \nabla \mathbf{f}(\mathbf{x}) = \left\{ \frac{\partial f_m(\mathbf{x})}{\partial x_n} \right\}. \tag{10}$$

Substituting (9) into (8) and taking the gradient, we find

$$\begin{aligned} \nabla \bar{J}(\mathbf{x}) = {} & \boldsymbol{\Xi}^{-1}\,(\mathbf{x} - \boldsymbol{\xi}) - \mathbf{K}(\mathbf{x}_0)^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1} \\ & \cdot\,[\mathbf{y} - \mathbf{f}(\mathbf{x}_0) - \mathbf{K}(\mathbf{x}_0)\,(\mathbf{x} - \mathbf{x}_0) - \boldsymbol{\delta}]\,, \end{aligned} \tag{11}$$

where $\bar{J}(\mathbf{x})$ is the locally linearized cost function. Setting this expression to $\mathbf{0}$ and solving for $\mathbf{x}$ yields the local minimum, which may then be used as a new operating point $\mathbf{x}_1$. Iterating this procedure, we obtain the "update" equation

$$\begin{aligned} \mathbf{x}_{k+1} = {} & \mathbf{x}_k - (\mathbf{K}_k^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}\,\mathbf{K}_k + \boldsymbol{\Xi}^{-1})^{-1} \\ & \cdot\,\left[\boldsymbol{\Xi}^{-1}\,(\mathbf{x}_k - \boldsymbol{\xi}) - \mathbf{K}_k^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}\,(\mathbf{y} - \mathbf{f}_k - \boldsymbol{\delta})\right], \end{aligned} \tag{12}$$

where $\mathbf{f}_k \equiv \mathbf{f}(\mathbf{x}_k)$ and $\mathbf{K}_k \equiv \mathbf{K}(\mathbf{x}_k)$ for $k = 0, 1, \ldots$. At the k^{th} iteration, the covariance matrix is given by

$$\mathbf{S}_k = (\mathbf{K}_k^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}\,\mathbf{K}_k + \boldsymbol{\Xi}^{-1})^{-1}. \tag{13}$$

Note that the covariance updates are not iterative, but are simply generated along the way in accordance with (5). For convenience, the iteration may be started with the initial guess $\mathbf{x}_0 = \boldsymbol{\xi}$; at convergence, we have $\mathbf{x} \sim \mathcal{N}(\mathbf{x}^*, \mathbf{S}^*)$, assuming that $\mathbf{S}^*$ does not extend beyond the neighborhood of local linearity around $\mathbf{x}^*$.

The above optimization scheme constitutes the ***Gauss–Newton method*** [2]; it makes partial use of the Hessian information, and may therefore be regarded as of "order $1\frac{1}{2}$." Despite the regularizing effect of the prior, the iteration step size – the second term on the right-hand side of (12) – may still be large enough to push $\mathbf{x}_{k+1}$ outside of the local neighborhood of $\mathbf{x}_k$ within which (9) holds. This violates the basic premise of the method, and the iteration becomes unstable. We avert this situation by adaptively penalizing large steps via the ***Levenberg–Marquardt strategy*** [7].

Finally, a convergence criterion is needed so that (12) is not iterated *ad infinitum*. The logical choice is to monitor the step size at each iteration; of course, the step sizes in different directions must be scaled appropriately to provide a sensible measure of change between consecutive updates. In addition, it is wise to make sure that the associated change in the cost function is also appropriately small. Thus, we declare convergence when

$$(\mathbf{x}_{k+1} - \mathbf{x}_k)^{\mathrm{T}}\,\mathbf{S}_k^{-1}\,(\mathbf{x}_{k+1} - \mathbf{x}_k) \ll N \tag{14}$$

and

$$|J(\mathbf{x}_{k+1}) - J(\mathbf{x}_k)| \ll M. \tag{15}$$

While the actual evaluation of the Hessian is generally avoided by all methods as it is deemed too costly computationally, the use of accurate gradient information is extremely crucial for successful optimization. Traditionally, the Jacobians (10) have been evaluated by differentiating the forward model approximately as

$$K_{mn} = \frac{f_m(x_n + \Delta x_n) - f_m(x_n)}{\Delta x_n}. \tag{16}$$

This not only requires (at least) MN **additional** evaluations of the forward model, which is very prohibitive for the typical M and N values encountered in ES data analysis, but also leads to rather inaccurate Jacobians due to the inevitable arbitrariness in the choice of the "external" perturbation sizes Δx_n. It is being recognized by more and more investigators that the most efficient way to generate Jacobians is by an "internal" perturbation of the forward model. In this approach, straightforward application of the chain rule of differentiation permits one, for a relatively minor computational overhead, to calculate Jacobians within the same scheme used for the forward-model calculation. In fact, this consideration is so critical for the retrieval method that the numerical approach for forward-model implementation should be chosen on the basis of the availability of a thorough internal perturbation analysis. Our method of choice for forward model and Jacobian evaluations is described in §III-C.

D. Retrieval framework for Earth-Science data analysis

It is of interest to generalize (2) to a practical situation where multiple measurements $\mathbf{y}_a$ (each of length M_a) are available for processing, and multiple unknowns $\mathbf{x}_b$ (each of length N_b) are wished to be retrieved. In ES retrieval problems, the index b may run over atmospheric temperature, pressure, gases, clouds, and aerosols, surface temperature, albedo, and bi-directional reflectance distribution, *etc.*, with the corresponding state vectors representing suitable discrete parameterizations of these physical variables. Meanwhile, the index a may run over measurements obtained by different (remote and/or *in situ*) instruments, or over multiple measurements made by a single instrument (or both). In any case, the essential requirement is that the various measurements be nearly simultaneous in space and time, such that the underlying physical state of the atmosphere and the surface may be assumed to remain unchanged during the span of these measurements.[2]

[2]In the case of hyper-spectral image processing, there are significant benefits to be accrued from a spatially-varying representation for the state vectors; however, limited space does not permit further elaboration here. Allowing for full spatio-temporal evolution of the state vectors over the course of the experiment takes us into the vast realm of global data assimilation and circulation models [8], into which we also will not venture.

For compactness, we collect the various measurement and state vectors into sets $\mathbb{Y} = \{\mathbf{y}_a\}$ (data) and $\mathbb{X} = \{\mathbf{x}_b\}$ (unknowns). Based on the above description, it is realistic to expect that the different measurement vectors will be mutually statistically independent, and likewise for the different state vectors. Thus, the appropriate generalization of (2) is

$$p(\mathbb{X}|\mathbb{Y}) = \frac{\ell(\mathbb{Y}|\mathbb{X})\, p(\mathbb{X})}{p(\mathbb{Y})} = \frac{1}{p(\mathbb{Y})} \prod_a \ell(\mathbf{y}_a|\mathbb{X}) \prod_b p(\mathbf{x}_b), \tag{17}$$

and with Gaussian likelihoods and priors, the corresponding cost function becomes

$$\begin{aligned} J(\mathbb{X}) &= -\ln \ell(\mathbb{Y}|\mathbb{X}) - \ln p(\mathbb{X}) \\ &= \frac{1}{2} \sum_a [\mathbf{y}_a - \mathbf{f}_a(\mathbb{X}) - \boldsymbol{\delta}_a]^{\mathrm{T}}\, \boldsymbol{\Sigma}_a^{-1}\, [\mathbf{y}_a - \mathbf{f}_a(\mathbb{X}) - \boldsymbol{\delta}_a] \\ &\quad + \frac{1}{2} \sum_b (\mathbf{x}_b - \boldsymbol{\xi}_b)^{\mathrm{T}}\, \boldsymbol{\Xi}_b^{-1}\, (\mathbf{x}_b - \boldsymbol{\xi}_b) + C'. \end{aligned} \tag{18}$$

We can now proceed with the Gauss–Newton approach toward minimizing this full cost function to obtain optimal estimates and error bars for the unknown quantities. This, then, constitutes the ***Bayesian calculus of inference*** – a framework for probabilistic reasoning within which data from any number heterogeneous sources are systematically combined with any type of prior information to solve the retrieval problem. Given the ongoing explosion in ES data volume, this appears to be the only approach that can efficiently assimilate and meaningfully analyze the diverse information gathered by the distributed agents of a sensor web. Several points may be made in connection with this formulation:

- In the Bayesian setting, the underlying physical model of the observed phenomenon, and **not** the data collected from it, is regarded as the item of central importance; working in this "model space" enables seamless integration of data from any number of *in situ* and remote (ground as well as orbital) sensors.
- Measurements and priors with lesser uncertainty, as quantified by the matrices $\boldsymbol{\Sigma}_a$ and $\boldsymbol{\Xi}_b$, will play a proportionately greater role in determining the position of the global minimum of J (*i.e.*, the optimal estimate).
- Naturally, different measurements $\mathbf{y}_a$ will have varying sensitivities to different states $\mathbf{x}_b$; this will be reflected (to first order) by the $M_a \times \sum_b N_b$ Jacobian matrices $\mathbf{K}_a(\mathbb{X}) \equiv \boldsymbol{\nabla} \mathbf{f}_a(\mathbb{X})$.
- One does not need a separate forward model for each data stream, but only for each **type** of measurement; for instance, if all the measurements are made by radiation sensors, then a single radiative-transfer calculation may suffice as the forward model for the entire data set.
- Individual measurements contribute to the cost function additively in (18), which allows for the processing of "streaming" data (as opposed to batch processing of the entire data set after it has been acquired); with each new measurement, the optimal estimate and its covariance can be updated through (12) and (13).
- Since parametric uncertainties can be assessed along with the best estimates, data acquisition may be steered automatically, and ceased when sufficient confidence in the inferred model parameters is attained, thus preventing the data volume from growing unnecessarily.
- The Bayesian approach thus utilizes the solution of the conceptually easier forward modeling problem to solve the more difficult, and arguably more interesting, inverse inference problem; domain knowledge and expertise enter naturally into the solution through the priors and the likelihood functions.

III. Forward Models

Consider an atmosphere composed of a mixture of gases indexed by i and bounded from below by a (partially) reflecting surface. We shall take this medium to be static and plane-parallel (*i.e.*, a horizontally homogeneous slab) within the spatio-temporal span of our measurements. The physical state of the medium is then fully characterized by the vertical profiles of temperature $T(z)$, pressure $P(z)$, and molecular volume number densities $N_i(z)$, along with a spectral reflectance function $\rho(\lambda; \mathbf{s}; \mathbf{s}')$ for the surface. We wish to retrieve these physical properties from remote measurements of the spectral intensity (or radiance) $I_\lambda(\mathbf{r}, \mathbf{s})$.

It must be acknowledged, however, that during the act of measuring a physical signal, a sensor inevitably imparts its own "signature" by altering the signal properties in some unique fashion; in this sense, the measurement device constitutes an integral part of any physical data-collection process. In order to achieve the most accurate retrieval of the atmospheric and surface properties of interest, it is therefore necessary to incorporate a detailed physical model of the sensor in series with one for the atmosphere–surface system. We thus require **two** forward models in our data-analysis effort.

A. The sensor

We specialize to an hyper-spectral instrument that samples the radiation spectrum uniformly over some wavelength range $(\lambda_{\mathrm{min}}, \lambda_{\mathrm{max}})$ with a resolution (or channel spacing) $\Delta\lambda$ and a total number of channels (or bands) $M = (\lambda_{\mathrm{max}} - \lambda_{\mathrm{min}})/\Delta\lambda$.[3] With the sensor at position $\mathbf{r}$ and pointing in direction $-\mathbf{n}$, the output of the m^{th} channel may be written in the form

$$y_m = \beta_m \int \psi(\lambda_m - \lambda) \int_{2\pi} \chi(\mathbf{n} \cdot \mathbf{s})\, I_\lambda(\mathbf{r}, \mathbf{s}; \mathbb{X})\, \mathrm{d}\Omega\, \mathrm{d}\lambda + \delta_m + \eta_m. \tag{19}$$

The instrument "slit function" $\psi(\lambda)$ has been assumed identical for all channels, and the "antenna pattern" $\chi(\Omega)$ has been taken independent of wavelength, though these can obviously be kept more general; typically, these functions are known from calibration experiments. On the other hand, λ_m, β_m, δ_m, and η_m respectively denote the channel center wavelength

[3] The term "hyper-spectral" implies that M is of the order of a few hundred or more; however, the ensuing discussion applies equally well to multi-spectral instruments with a few to a few tens of (not necessarily contiguous) channels, as well as to (essentially single-channel) active instruments.

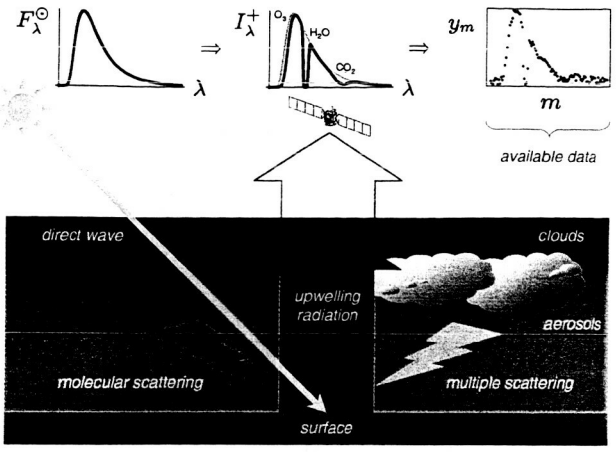

Fig. 2. The composite forward problem. The solar beam with spectral flux density $F_\lambda^\odot$ is incident on the terrestrial atmosphere composed of gases, clouds, and aerosols; in this paper, we consider an atmosphere composed only of molecules. Through absorption, scattering, and reflection, the atmospheric and surface constituents impart their spectral signatures onto the up-welling spectral intensity I_λ^+. This is, in turn, detected by an hyper-spectral instrument on a satellite platform, which produces the data stream $\{y_m\}$. The red and blue double arrows respectively designate the atmosphere–surface and the instrument forward models. (The spectra are stylized for illustration.)

(determined by the optics), gain (set by the electronics), offset, and noise. Specifically, δ_m are the mean of the optical and electrical noise sources in the sensor, such as the photodiode dark noise and the (optical) background and inter-pixel cross-talk noise. Meanwhile, η_m model the fluctuations associated with these noise sources, as well as the thermal noise of the electronics; they are taken to be zero-mean Gaussian random variables with covariance $\Sigma_{mm'} = \mathcal{E}(\eta_m \eta_{m'})$. Together, δ_m and η_m comprise the measurement error ϵ_m of §II. We assume that all these sensor parameters have also been adequately characterized via suitable calibration experiments prior to instrument deployment; if this is not the case, they can be arranged into an instrument state vector $\mathbf{x}_{\text{ins}}$ and included in the set $\mathbb{X}$ of unknowns to be retrieved from data.

Finally, the spectral intensity $I_\lambda(\mathbf{r}, \mathbf{s}; \mathbb{X})$ incident on the sensor is a nontrivial function of the atmospheric and surface states comprising $\mathbb{X}$; this part of the forward model will be developed in §III-B. The first term in (19) thus constitutes the "composite" (*i.e.*, atmosphere–surface plus instrument) forward model $f_m(\mathbb{X})$ of §II (see Figure 2).

The somewhat phenomenological equation (19) is the type of channel model commonly employed in the literature; however, a rigorous signal/noise analysis of a generic electro-optic sensor reveals that this is merely an approximation to the full probabilistic model of such an instrument [9]. As is echoed at many recent technical meetings, currently one of the biggest sources of uncertainty in remote sensing is the poor sensor calibration standards and practices. It is therefore imperative that one verifies the validity of the **form** of (19) before using it as the basis of any calibration and/or retrieval algorithms.

B. The atmosphere–surface system

In a plane-parallel atmosphere, the equation of **solar** radiative transfer for the spectral intensity is [10]-[12]

$$\cos\theta \frac{\partial I_\nu}{\partial z} = -\kappa(\nu; z)\, I_\nu(z, \theta, \phi) + \sigma(\nu; z) \cdot \int_0^{2\pi} \int_0^{\pi} p(\theta, \phi; \theta', \phi')\, I_\nu(z, \theta', \phi')\, \sin\theta'\, \mathrm{d}\theta'\, \mathrm{d}\phi', \tag{20}$$

where θ and ϕ are the usual polar and azimuth angles (see Figure 3), $\kappa(\nu; z) = \sigma(\nu; z) + \alpha(\nu; z)$, with κ, σ, and α respectively denoting the extinction, scattering, and absorption coefficients of the medium, and we switched to the frequency variable $\nu = c/\lambda$ more commonly used in spectroscopy. For a molecular atmosphere, the scattering phase function is given by $p(\Theta) = \frac{3}{16\pi}(1 + \cos^2\Theta)$, where Θ denotes the (solid) angle of scattering between directions (θ', ϕ') and (θ, ϕ). The scattering and extinction coefficients, meanwhile, are given in terms of the molecular concentrations as

$$\left\{ \begin{matrix} \sigma(\nu; z) \\ \kappa(\nu; z) \end{matrix} \right\} = \sum_i N_i(z) \left\{ \begin{matrix} C_{\text{sca}}(\nu) \\ C_{\text{ext}}(\nu) \end{matrix} \right\}. \tag{21}$$

For a molecule irradiated by an unpolarized plane wave, the scattering and extinction cross sections are given by [13]-[15]

$$C_{\text{sca}}(\nu) = \frac{k^4}{6\pi\epsilon_0^2} |\pi_i(\nu)|^2, \tag{22}$$

$$C_{\text{ext}}(\nu) = \frac{k}{\epsilon_0} \Im\, \pi_i(\nu), \tag{23}$$

where $k = 2\pi\nu/c$, c and ϵ_0 are the speed of light and the dielectric permittivity *in vacuo*, and $\pi_i(\nu)$ denotes the molecular polarizability of the i^{th} gas.

The problem description is made complete by the specification of boundary conditions that the intensity must obey at the

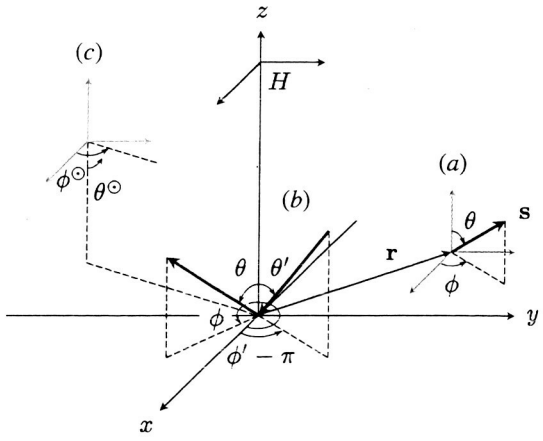

Fig. 3. The geometry of the radiative transfer problem. (*a*) An arbitrary ray at position $\mathbf{r}$ propagating in direction $\mathbf{s}$. (*b*) The incident and reflected ray angles for the surface bi-directional reflectance distribution function $\rho(\nu; \theta, \phi; \theta', \phi')$. (*c*) The incident solar ray. For propagation into the lower hemisphere, the polar angles are measured from the $-z$ axis.

top of the atmosphere ($z = H$) and at the surface ($z = 0$):

$$I_\nu^-(H,\mu,\phi) = F_\nu^\odot\,\delta(\mu-\mu^\odot)\,\delta(\phi-\phi^\odot), \tag{24}$$

$$I_\nu^+(0,\mu,\phi) = \int_0^{2\pi}\int_0^1 \rho(\nu;\mu,\phi;\mu',\phi') \cdot I_\nu^-(0,\mu',\phi')\,\mu'\,\mathrm{d}\mu'\,\mathrm{d}\phi', \tag{25}$$

where $^+/^-$ indicate up-/down-welling intensities, $(\theta^\odot,\phi^\odot)$ specify the direction of the incident solar beam, $\mu \equiv \cos\theta$, $\mu' \equiv \cos\theta'$, and $\mu^\odot \equiv \cos\theta^\odot$ (see Figure 3). In our work, we have so far used the ***Kurucz model*** for the solar spectral flux $F_\nu^\odot$; satellite measurements of $F_\nu^\odot$ have recently become available and should be used to remove from data the influence of solar variability. Meanwhile, (25) clearly demonstrates the need for an accurate surface model even if one is interested in retrieving only the atmospheric properties. When the surface is not known or only poorly characterized, a suitable parametric model should be adopted for ρ and a corresponding state vector $\mathbf{x}_{\mathsf{sur}}$ added to the set $\mathbb{X}$ of unknowns.

In the spectroscopy literature, the extinction coefficient is typically written in the form

$$\kappa(\nu;z) = \sum_i N_i(z) \sum_j S_{ij}(z)\, f[\nu;\nu_{ij}(z),\hat{\gamma}_{ij}(z),\tilde{\gamma}_{ij}(z)], \tag{26}$$

where j is the resonance (or line) index, and f denotes the gas extinction line shape. For a line centered at frequency ν_0 with a collision-broadened width $\hat{\gamma}$ and a motion-broadened width $\tilde{\gamma}$, the Voigt profile is given by the convolution integral [11], [12]

$$f(\nu;\nu_0,\hat{\gamma},\tilde{\gamma}) = \frac{\hat{\gamma}}{\pi}\frac{1}{\sqrt{\pi}\,\tilde{\gamma}}\int_{-\infty}^{\infty}\frac{\mathrm{e}^{-(\omega/\tilde{\gamma})^2}}{(\nu-\nu_0-\omega)^2+\hat{\gamma}^2}\,\mathrm{d}\omega, \tag{27}$$

consisting in the folding of Lorentzian and Gaussian line shapes, respectively dictated by the impact theory and the Doppler effect.

The line center, line strength, and collision- and motion-broadened line widths in (26) are respectively given by [16]

$$\nu_{ij} = \nu_{ij}^\star + \delta_{ij}^\star\,P \tag{28}$$

$$S_{ij} = S_{ij}^\star\left(\frac{T^\star}{T}\right)^{m_i^\star}\exp\left[-\frac{h_{\mathsf{P}}\,c}{k_{\mathsf{B}}}E_{ij}^\star\left(\frac{1}{T}-\frac{1}{T^\star}\right)\right] \tag{29}$$

$$\hat{\gamma}_{ij} = \left(\frac{T^\star}{T}\right)^{n_i^\star}\left[\hat{\gamma}_{ij}^\star\,P + (\hat{\gamma}_{ij}'^\star - \hat{\gamma}_{ij}^\star)\,N_i\,k_{\mathsf{B}}\,T\right] \tag{30}$$

$$\tilde{\gamma}_{ij} = \frac{\nu_{ij}^\star}{c}\sqrt{\frac{2\,N_{\mathsf{A}}\,k_{\mathsf{B}}\,T}{M_i}} \tag{31}$$

where the z dependence has been suppressed for convenience, and * indicates quantities whose values are obtained from the ***HITRAN database***. The remaining symbols are h_{P}: Planck's constant, k_{B}: Boltzmann's constant, N_{A}: Avogadro's number, and M_i: molecular mass of the i^{th} gas.

The equations in this subsection indicate the complicated manner in which the atmospheric and surface properties influence the spectral intensity measured by the sensor; thus, they collectively comprise the forward physics model for the atmosphere–surface system.

C. Computational details

The standard technique for solving (20) is the ***discrete-ordinate method*** of *Chandrasekhar* [10], [12]; the widely used DISORT code is a very clean and efficient numerical implementation of this approach [17]. Here, the atmosphere is divided into L homogeneous layers; the l^{th} layer has its lower and upper boundaries at altitudes z_{l-1} and z_l, and its optical depth is defined as

$$\tau_l(\nu) \equiv \int_{z_{l-1}}^{z_l}\kappa(\nu;\zeta)\,\mathrm{d}\zeta. \tag{32}$$

The choice for the layer boundaries z_l – as well as L itself – is dictated by the available priors and measurements, the required retrievals, and computational considerations associated with the forward model. We choose to define the atmospheric state variables at the layer boundaries, comprising the $(L+1)$-dimensional state vectors

$$\mathbf{x}_b = \{b_l \equiv b(z_l)\}\,,\quad b = T,P,N_i. \tag{33}$$

We use the ***1976 US Standard Atmosphere*** [12] profiles for the prior means $\boldsymbol{\xi}_b$, and an hyper-parametric Markovian structure for $\boldsymbol{\Xi}_b$.

Instead of (27), we prefer to work with an alternate (Fourier-transformed) version of the Voigt profile given by

$$f(\nu;\nu_0,\hat{\gamma},\tilde{\gamma}) = \int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}2\pi(\nu-\nu_0)t-2\pi\hat{\gamma}|t|-(\pi\tilde{\gamma}t)^2}\,\mathrm{d}t. \tag{34}$$

This enables the use of efficient FFT routines, as well as simplifying the evaluation of forward-model Jacobians. Use of (34) and (26) in (32) leads to

$$\tau_l(\nu) = \int_{-\infty}^{\infty}\left[\int_{z_{l-1}}^{z_l}\Phi(t;\zeta)\,\mathrm{d}\zeta\right]\mathrm{e}^{\mathrm{i}2\pi\nu t}\,\mathrm{d}t, \tag{35}$$

where

$$\Phi(t;z) \equiv \sum_i N_i(z)\sum_j S_{ij}(z) \cdot \exp\left\{-\mathrm{i}2\pi\nu_{ij}(z)t - 2\pi\hat{\gamma}_{ij}(z)|t| - [\pi\tilde{\gamma}_{ij}(z)t]^2\right\}. \tag{36}$$

For the vertical integration in (35), we do away with the widely-used hydrostatic law and the van de Hulst–Curtis–Godson scaling approximation [11], [12]. Instead, we expand $\Phi(t;z)$ over $z_{l-1} \le z \le z_l$ as

$$\Phi(t;z) \simeq \Phi(t;z_{l-1}) + \left.\frac{\partial\Phi}{\partial T}\right|_{z_{l-1}}[T(z)-T_{l-1}] + \left.\frac{\partial\Phi}{\partial P}\right|_{z_{l-1}}[P(z)-P_{l-1}] + \sum_i\left.\frac{\partial\Phi}{\partial N_i}\right|_{z_{l-1}}[N_i(z)-N_{i(l-1)}]. \tag{37}$$

The partial derivatives appearing above may be evaluated analytically from (36) with (28)-(31). The vertical profiles are then fit by cubic splines [7] – an $\mathcal{O}(L)$ operation – for which the vertical integrals are also evaluated off-line. A highly accurate, and aesthetically pleasing, representation of the atmospheric

states is thus achieved for a minor computational overhead. An FFT operation completes the evaluation of τ_l.

DISORT is then fed with the optical depths of the layers, the solar spectral flux, and the surface spectral reflectance, and returns the spectral intensity $I_\nu(z, \theta, \phi)$ at user-specified altitudes and angles. This output is finally used in (19) to complete the forward-model evaluation.

Recently, *Spurr* has carried out an internal perturbation analysis of the discrete ordinate method [18], which has led to the LIDORT code that is now available. Although still somewhat rough around the edges, LIDORT outperforms DISORT by nearly a factor of 10 in intensity calculations, and can also be used to generate forward-model Jacobians. Here, one uses straightforward chain-rule differentiation on the expressions in §III-B to feed LIDORT with the derivatives of the layer optical depths τ_l with respect to the state vectors $\mathbf{x}_b$. (It is possible to perturb the model with respect to the surface parameters as well.) We are presently exploring this tool, along with automatic code differentiators, in lieu of crude external perturbations *à la* (16).

IV. Summary and Outlook

All the ingredients are thus in place for bringing the Bayesian inference framework described in §II-D to bear on the atmospheric and surface remote sensing problem. Note that the popular approach of building pre-computed look-up tables has been abandoned in favor of retrieving all the key physical parameters directly from the data. This also liberates us from the limitations and inaccuracies of standard computational techniques such as the correlated-k method [11], [12].

At present, the only inexact aspect of our physical model is the scalar treatment of the radiation field. It has been well-established in the radiative-transfer literature that polarization effects may accumulate to make a sizable contribution to the scalar intensity I even in a pure Rayleigh-scattering atmosphere; with the addition of clouds, aerosols, and detailed surface effects into the forward physics, it becomes imperative that we employ a full vectorial model of the radiation field in terms of the Stokes vector $\mathbf{I}$ [10]. A vectorial version of DISORT has been developed recently [19], and efforts are under way to construct a corresponding vectorial version of LIDORT; these improvements will be incorporated into our framework in the near future.

Our immediate focus is on the application of our computational framework to the processing of hyper-spectral images from the ***Hyperion*** instrument flying on Earth Observing One [20]. For this task, we will be using a spatial random-field model for the atmospheric states, enabling the use of a Kalman filter for processing the pixels as separate (correlated) measurements. We are particularly interested in Hyperion images gathered over the Southern Great Plains site of the US Department of Energy's Atmospheric Radiation Measurement Program [21]. The presence of a large number and variety of instruments at this site will supply us with accurate "ground truth" for validating our retrievals, as well as providing an ideal setting for demonstrating the power of our framework for processing heterogeneous data. We thus hope to achieve, for the first time, the simultaneous and highly-accurate retrieval of atmospheric **and** surface properties from a single data set.

Acknowledgments

The author would like to thank Dr K. R. Wheeler for his invitation to participate in this project and for his guidance in bridging the gap between theory and data, and Dr J. C. Coughlan for his financial support of this work through the Intelligent Data Understanding Project of the NASA CICT Program. The programming support from Dr N. Lvov and L. R. McFarland of QSS, Inc., is gratefully acknowledged.

References

[1] S. Twomey, *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements*, Amsterdam: Elsevier, 1977.

[2] C. D. Rodgers, *Inverse Methods for Atmospheric Sounding – Theory and Practice*, Singapore: World Scientific, 2000.

[3] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*, New York: Wiley, 1994.

[4] D. S. Sivia, *Data Analysis – A Bayesian Tutorial*, Oxford, UK: Oxford University Press, 1996.

[5] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed., New York: McGraw-Hill, 1991.

[6] E. T. Jaynes, *Probability Theory – The Logic of Science*, Cambridge, UK: Cambridge University Press, UK, 2003.

[7] W. H. Press, S. Teukolsky, W. T. Vetterling, and B. Flannery, *Numerical Recipes in C – The Art of Scientific Computing*, 2nd ed., Cambridge, UK: Cambridge University Press, 1992.

[8] R. Daley, *Atmospheric Data Analysis*, Cambridge, UK: Cambridge University Press, UK, 1991.

[9] D. A. Timuçin, "A Bayesian approach to sensor characterization," in *Proc. IGARSS*, Toulouse, France, July 2003.

[10] S. Chandrasekhar, *Radiative Transfer*, New York: Dover, 1960.

[11] R. M. Goody and Y. L. Yung, *Atmospheric Radiation – Theoretical Basis*, 2nd ed., New York: Oxford University Press, 1989.

[12] G. E. Thomas and K. Stamnes, *Radiative Transfer in the Atmosphere and Ocean*, Cambridge, UK: Cambridge University Press, 1999.

[13] H. C. van de Hulst, *Light Scattering by Small Particles*, New York: Dover, 1981.

[14] C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles*, New York: Wiley, 1983.

[15] J. D. Jackson, *Classical Electrodynamics*, 3rd ed., New York: Wiley, 1999.

[16] L. S. Rothman *et al.*, "The HITRAN molecular spectroscopic database: edition of 2000 including updates through 2001," *J. Quant. Spectrosc. Radiat. Transfer*, vol. 82, pp. 5-44, November/December 2003.

[17] K. Stamnes, S.-C. Tsay, W. Wiscombe, and K. Jayaweera, "Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media," *Appl. Opt.*, vol. 27, pp. 2502-2509, June 1988.

[18] R. J. D. Spurr, T. P. Kurosu, and K. V. Chance, "A linearized discrete ordinate radiative transfer model for atmospheric remote-sensing retrieval," *J. Quant. Spectrosc. Radiat. Transfer*, vol. 68, pp. 689-735, March 2001.

[19] F. M. Schulz, K. Stamnes, and F. Weng, "VDISORT: An improved and generalized discrete ordinate method for polarized (vector) radiative transfer," *J. Quant. Spectrosc. Radiat. Transfer*, vol. 61, pp. 105-122, January 1999.

[20] J. S. Pearlman *et al.*, "Hyperion, a space-based imaging spectrometer," *IEEE Trans. Geosci. Remote Sensing*, vol. 41, pp. 1160-1173, June 2003.

[21] T. P. Ackerman and G. M. Stokes, "The atmospheric radiation measurement program," *Physics Today*, pp. 38-44, January 2003.