Parametric 3D Scene Reconstruction from Imaging Radiometry, Part II: The Bayesian Multi-Pixel Inversion Algorithm

Journal:	*Transactions on Geoscience and Remote Sensing*
Manuscript ID:	Draft
Manuscript Type:	Regular paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Langmore, Ian; Columbia University, Department of Applied Physics & Applied Mathematics Bal, Guillaume; Columbia University, Department of Applied Physics & Applied Mathematics Davis, Anthony; California Institute of Technology, Jet Propulsion Laboratory
Keywords:	Scattering, Reflection, Aerosols, Monte Carlo methods, Numerical analysis, Remote sensing, Gases, Markov processes, Bayes procedures, Inverse problems

Parametric 3D Scene Reconstruction from Imaging Radiometry, Part II: The Bayesian Multi-Pixel Inversion Algorithm

Ian Langmore, Guillaume Bal, *Member, Am. Math. Soc., Soc. Indust. Appl. Math.,*
and Anthony B. Davis, *Member, Opt. Soc. Am., Am. Geophys. Un., Am. Meteor. Soc.*

Abstract

In Part I of this series, we described the mechanics and performance of a highly accelerated forward Monte Carlo model that predicts radiances in all the pixels of a simple imaging sensor looking at a certain scene at a single wavelength. The scene consists of an absorbing atmospheric plume somewhere in a deep valley while the space in and above it is filled with a partially-known and partially-unknown aerosol with a decaying exponential density stratification. In this second and final installment, we put the 2D forward model to work in a multi-pixel Bayesian inversion scheme that infers more than just the most probable values of the parameterized 2D scene's five unknown quantities: the plume's position, radius and density, and the specific aerosol amount. It also delivers an estimate of their 5-dimensional probability density function, hence means and associated uncertainty bounds from its marginal distributions. This Bayesian inference is made possible by refining Monte Carlo Markov Chain (MCMC) methodology, which in essence is based on generating random walks through the 5-dimensional parameter space that are directed preferentially toward regions that most need to be sampled. The numerical evidence shows that the basic parameters of plumes in such scenes can be inferred, within known uncertainty bounds that add value in some important applications of airborne and space-based remote sensing into denied territory.

I. Introduction and Overview

WE critically reviewed in Part I [1] the limitations of mainstream physics-based remote sensing that follow from the practice of pixel-by-pixel processing, itself leading to widespread use of one-dimensional (1D) radiative transfer (RT) models. Subpixel-scale spatial variability is ignored and, more to the present point, net radiative fluxes between neighboring pixels are also ignored. These last "pixel adjacency" effects have been addressed in the literature surveyed in Part I where they are generally viewed as a corruption of the pixel-scale signal, if possible,

I. Langmore and G. Bal are with Columbia University's Department of Applied Physics & Applied Mathematics, 500 W. 120th street, New York, NY, 10027, USA, e-mail: gb2030@columbia.edu.

A. B. Davis, corresponding author, is with the Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Mail Stop 233-200, Pasadena, CA, 91109, USA, e-mail: Anthony.B.Davis@jpl.nasa.gov.

Manuscript received November ??, 2011; revised Month ??, 2011.

to be corrected. Then the pixel-level (or maybe a local mean) signal is processed as if nothing happened using its multi- or hyper-spectral radiometric [2] (and possibly polarimetric [3]) content, occasionally sampled at several viewing directions [4]–[6].

By contrast, we will think here of pixel-to-pixel variations (hence radiometric cross-talk) as a remote sensing resource and accept the necessary price of upgrading our forward RT model from 1D to 3D—actually 2D, which is sufficient for the present proof-of-concept. To underscore the *multi-pixel* nature of the retrieval, we simulate a mono-spectral imaging radiometer. Armed with a single low-resolution calibrated image, we will attempt a coarse reconstruction of a scene involving terrain with a known but highly variable height profile and a partially-known/partially-unknown background aerosol. The remote sensing challenge is to detect and characterize a chemical plume lurking somewhere in this scene and, as an ancillary task, specify fully the prevailing aerosol. The plume is assumed to have a circular section and we ask about its x- and z-positions, its diameter, and its density, knowing the absorption cross section of the chemical at the selected wavelength. The data is provided by an imaging sensor at finite range. This set-up mimics situations of interest in nuclear proliferation detection.

Optimization methods provide a single guess at the unknown parameters, and this is inadequate for the reason that atmospheric reconstructions will always leave one with a great deal of uncertainty. This results largely from uncertainty in the background atmosphere. The methodology provided here addresses this in two ways. First, we reduce the uncertainty by using an atmospheric RT model in higher dimensions, which provides much more fidelity than the traditional 1D model. While this model is more accurate, simplifications must still be made, and the number of measurements is still finite. Therefore, our second step is a Bayesian inversion that quantifies the resultant uncertainty by providing a joint probability distributions of the unknown quantities.

The paper is outlined as follows. In Section II, we state in precise mathematical terms the remote sensing inverse problem to solve and summarize the forward 2D RT modeling effort from Part I, this time with an eye on varying the numerical precision. In Section III, we describe the inverse transport method, which is inspired by current work in biomedical imaging, using "recycled" Monte Carlo histories with valuable information content in a Bayesian setting. Full mathematical details are provided in a companion paper by Bal, Langmore, and Marzouk [7] but the essential results are summarized in App. B of Part I (using notations defined in its App. A). Section IV analyses the outcome of a representative selection from our numerical experiments. We offer our concluding remarks and an outlook on future work in Section V.

II. Forward Model & Statement of Inverse Remote Sensing Problem

A. Scene Geometry and Optical Properties

Figure 1 illustrates the geometry of the scene and some of the aerosol properties. In the bottom panel, we see the "$1 - \cos^3 x$" terrain model, an indication of where the reflecting surface's albedo transitions from a low value of 0.1 to a high value of 0.5., and the position of the small imaging detector to the right-hand side. This terrain models a deep gorge and the surface albedo distribution approximates what an abrupt change in vegetation cover,

going from sparse to dense at mid-level, would look like in the near-IR (NIR) spectrum. For simplicity, the surface is assumed to be Lambertian.

To the left-hand side, we see a polar plot of the aerosol's forward-peaked scattering phase function (note the log scale). This phase function is modeled as a 2D counterpart of the popular double Henyey–Greenstein model and both components are plotted as well; the forward-peaked component has an asymmetry factor $g_1 = 0.8$ and the backward-peaked one has $g_2 = -0.4$, their mixing ratio being 9-to-1. In linear axes, the corresponding polar diagrams are ellipses with the origin at a focal point and an algebraic eccentricity of $2g/(1+g^2)$, that is, $+0.976$ and -0.670 respectively. The overall asymmetry factor is $g = f_1 g_1 + (1-f_1)g_2 \approx 0.7$. Moreover, the aerosol particles' single scattering albedo ϖ_0 is set to 0.9. These aerosol parameter choices roughly reproduce the optical properties of actual fine-mode populations [Olga V. Kalashnikova, pers. comm.].

All of the above properties are held fixed throughout the numerical experiments to come.

To the right-hand side, we see the exponential profile of the aerosol extinction coefficient: $\sigma(z) = \sigma_0 \exp(-cz)$ where $c = c_0 + \delta c$; in solid black, the reference ($\delta c = 0$) value of c is used, $c_0 = 0.5$ 1/km, while in solid grey it is the "true" value assumed in subsequent retrievals, with $\delta c = 0.15$ 1/km. The top-to-bottom optical depth of the aerosol, $\int_0^4 \sigma(z)\mathrm{d}z$, is unity for the reference case (leading to $\sigma_0 = 0.58$ 1/km) and 0.86 for the "true" case.

Finally, the outline of the "true" absorbing gaseous plume is highlighted in light grey; its parameters are $x_\mathrm{p} = +0.35$ km and $z_\mathrm{p} = 2.5$ km for the center of the disk, $\rho_\mathrm{p} = 0.5$ km for its radius, and its specific absorption coefficient is $k_\mathrm{p} = 0.5$ 1/km, and this is in addition to the local contribution of the background aerosol, $\sigma_\mathrm{a}(z) = (1-\varpi_0)\sigma(z)$. The translates to a "true" optical diameter $2k_\mathrm{p}\rho_\mathrm{p}$ of unity for this plume. Many gases of interest have absorption features in the NIR spectrum that could lead to this kind of opacity.

The forward 2D RT model is a modified Monte Carlo scheme that uses only pre-computed paths for the "reference" (no-plume, nominal aerosol) case that end at the small detector. The sun is assumed to be directly overhead , i.e., $\theta_0 = 180^\circ$ (where $\theta = 0$ means propagation toward zenith). However, the illumination is distributed (uniformly) only across the $\{z = 4$ km, $|x| \leq 2.5$ km$\}$ region highlighted in Fig. 1. This is because, for simplicity, the lateral boundary conditions are taken as absorbing (rather than periodic), and the reduced source region minimizes the impact of this assumption. Like the use of 2D (rather than 3D) RT, this simplification has no impact on the validity of the present demonstration of monochromatic multi-pixel atmospheric (plume-and-aerosol) reconstruction using noisy low-resolution remote sensing data.

Lastly, we illustrate in Fig. 1 how our efficient computational model works for the imaging radiometer signal prediction. This forward 2D RT model uses a "path-recycling" Monte Carlo scheme adapted form recent work in biomedical imaging in soft tissue [8]–[10]. Details and performance metrics are provided in Part I but here we show four typical *Monte Carlo particle* (not to misleadingly say *photon*) paths plotted with dark grey dashes.

Two of these paths, #2 and #4, do not intersect the hypothetical absorbing plume. However, their probability of occurring is not the same in the reference case ($c = c_0$) and in the actual case ($c = c_0 + \delta c$). This means that their contributions to the affected pixel-level signals needs to be adjusted. Specifics are in Part I. The two other random paths, #1 and #3, do intersect this particular plume (used as "truth" further on). Apart from the new weights related

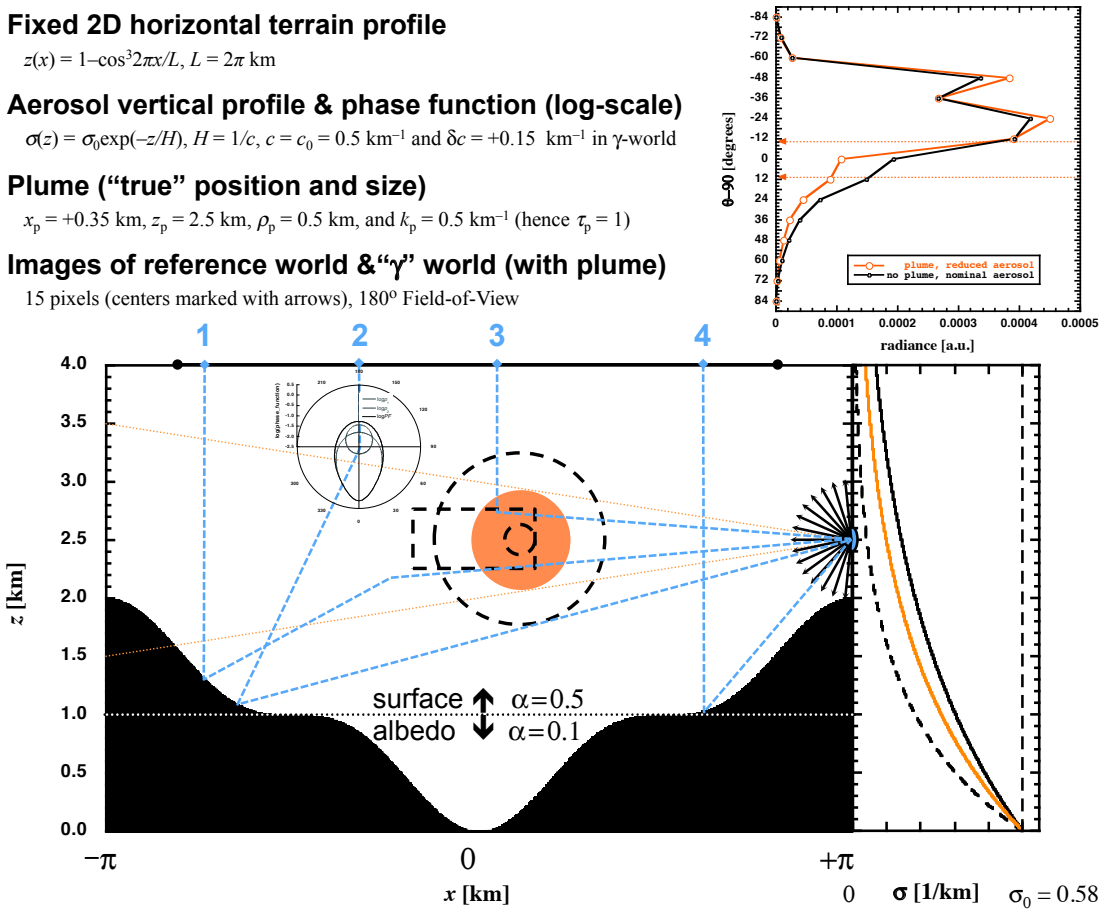

Fig. 1. *Schematic summary of 2D scene geometry and aerosol properties.* Some parameters are specified numerically in the upper left-hand corner. Many more details are provided in the main text and in Part I.

to the change in aerosol scattering, one must now account also for absorption by the plume material. This is done by computing the total length of the paths *inside* the plume, denoted ℓ, and change the paths' weights by a factor $\exp(-k_{\rm p}\ell)$. A specific algorithm for computing ℓ is provided in the App. A of Part I.

In the simulated retrievals to follow, we consider a continuum of "γ worlds" where γ formally represents the 5-dimensional state vector of unknowns, viz.,

$$\gamma = (x_{\rm p}, z_{\rm p}, \rho_{\rm p}, k_{\rm p}; \delta c). \tag{1}$$

The line plot in the upper right-hand corner of Fig. 1 is another rendering of the data in Fig. 4 of Part I. It shows the sequence of 15 angularly contiguous pixel-scale signals versus viewing angle (actual "images" in the 2D universe) re-centered with respect to the horizontal direction (at the center of the sensor's 180° field-of-view). Two situations of interest are represented:

1) solid black: the reference world, with no plume and nominal aerosol

2) light grey: a world where γ in (1) is set to (+0.35 km, 2.5 km, 0.5 km, 0.5 1/km; +0.15 1/km). The latter case summarizes all of the above choices for potentially variable parameters in the scene, and will be used as "truth" for the retrieval tests described in the next Section. The two arrows show the range of directions subtended by the plume.

In practice, the parameters in γ must be constrained. We therefore require $\gamma \in \Gamma \subset \mathbb{R}^5$ where

$$\Gamma = [x_{\min}, x_{\max}] \times [z_{\min}, z_{\max}] \times [\rho_{\min}, \rho_{\max}] \times [0, \infty) \times [\delta c_{\min}, \delta c_{\max}]. \quad (2)$$

The values used for these constraints are found in Table I. Also, the long-dashed black lines and plots in Fig. 1 show all of these bounds graphically, except for k_{p} (assumed only to be non-negative). We indeed see a bounding rectangle for $(x_{\mathrm{p}}, z_{\mathrm{p}})^{\mathrm{T}}$, plus inner and outer circumferences for the plume's circular area. The minimum ($\delta c = +c_0$) and maximum ($\delta c = -c_0$) aerosol loads are indicated in the extinction profiles to the right, corresponding respectively to aerosol optical depths of 0.57 and 2.32. These limits are used further on in the design of the Bayesian priors (not represented here is the prior Gamma distribution used for k_{p}).

TABLE I

SUMMARY OF VARIABLE ATMOSPHERE-SURFACE PARAMETERS

Parameter	Symbol	Min	Max	Unit	Prior
Plume's x-position	x_{p}	-0.5	$+0.5$	km	uniform
Plume's z-position	z_{p}	2.2	2.7	km	uniform
Plume's radius	ρ_{p}	0.15	0.85	km	uniform
Plume's absorption coef.	k_{p}	0	∞	1/km	Gamma
Aerosol perturbation	δc	$-c_0$	$+c_0$	1/km	uniform

B. Hierachical Monte Carlo Modeling with Variable Precision

Like for all unbiased Monte Carlo methods, accuracy is guarantied in our path-recycling scheme and it has known numerical uncertainty that is reduced to a question of precision. Moreover, the adopted Monte Carlo model for the required forward 2D RT is ideally suited for varying precision to a specified value (below some maximum). All one needs to do is recycle (in the sense of Part I) less paths than are available in the pre-computed database, and thus reap a benefit in computational efficiency: variance will be inversely proportional to the time spent on the multi-pixel prediction.

We are now in a position to discuss our Bayesian inference methodology where the availability of the precision-based sequence of forward models is key to success.

III. Bayesian Multi-Pixel Retrieval Technique

A. General Considerations

The Bayesian viewpoint takes the remote sensing unknowns as random variables. Rather than simply find parameter values that best fit the observations, a multi-variate probability distribution is provided for the parameters. From there, one can evaluate means, most probable values, medians, variances (retrieval uncertainties), and so on.

In general, our *unknown* is a random vector $\mathbf{x} \in \mathbb{R}^n$ with *prior* probability density $p_{\text{prior}}(\mathbf{x})$. This is the distribution we *assume* (from prior knowledge) on $\mathbf{x}$ before any data are collected. In this study, we have $n = 5$ and denote $\mathbf{x}$ as γ, with specific definitions in (1) and bounds in (2) and in Table I.

We assume our *data* $\mathbf{d} \in \mathbb{R}^m$ are given by observations as well as by an *infinite-precision* forward model $\mathbf{F}_\infty(\mathbf{x})$ viewed here as a vector-valued multi-variate function of $\mathbf{x}$. In the present study, we have $m = 15$, that is, one observation per pixel in the 1D image of 2D space taken from the fixed detector position.

Mathematically, $\mathbf{F}_\infty : \mathbb{R}^n \to \mathbb{R}^m$, with $m \geq n$ if we want to have any hope of recovering all the parameters in γ. We also have an additive noise vector term $\mathbf{E}$ assumed to be independent of $\mathbf{x}$ and Gaussian, with a covariance matrix $\Sigma_{\mathbf{E}}$. Thus, our model for the data is

$$\mathbf{d} = \mathbf{F}_\infty(\mathbf{x}) + \mathbf{E}, \qquad \mathbf{E} \sim \mathcal{N}(0, \Sigma_{\mathbf{E}}), \qquad \mathbf{E} \perp\!\!\!\perp \mathbf{x}. \tag{3}$$

We adopt here notations from mathematical statistics: $\sim$ means "distributed as" and $\perp\!\!\!\perp$ means "independent of". The methods presented here do not depend on the choice of prior PDF. The Gaussian assumption about the additive noise model simplifies the algebra, but is not necessary either.

We note that potentially significant improvement could be made by using a realistic model for $\mathbf{E}$. That uncertainty is much harder to quantify and is actually *not* independent and identically distributed (i.i.d.). For simplicity, we nonetheless choose the noise to be i.i.d. in all 15 detectors. This is certainly not so, since pixels pointed toward the ground encounter a different level of noise than those pointed toward the sky. A result of this i.i.d. assumption is that about half of the pixels—those pointed in directions not receiving flux that last hit the plume—can receive a signal smaller than the noise dictated by $\mathbf{E}$. Moreover, $\mathbf{E}$ should ideally capture not only instrumental noise but also forward modeling error, e.g., deviation of the real world from assumptions such as exponentially distributed aerosols. That uncertainty is much harder to quantify and is actually not random but systematic in nature. In this demo, we are relatively immune to this effect since the forward and inverse transport models are quite closely related.

In its most general form, Bayes' rule [11, among others] states that $p(x|y) = p(y|x)p(x)/p(y)$ where $p(a|b)$ designates the probability density of the random variable a, conditioned on a specific occurrence of another random variable b. It expresses how observations, denoted y, affect our knowledge of properties of interest, x, starting from no observations at all. *Prior* to acquiring observations y, all we have is the probability density $p(x)$. *After* the observations are made, we have narrowed the possibilities to $p(x|y)$, the "posterior" probability density, and the collapse is driven by the probability density of observing values y for given (although still unknown) properties, namely, $p(y|x)$. We note that the marginal probability density of the observations $p(y)$ (for any state x) is often

treated as a normalization factor that can be ignored in practice: it suffices to state that $p(x|y) \propto p(y|x)p(x)$ where $p(y|x)$ is known as the "likelihood" function.

In our present notations and assumptions, the *posterior* is

$$\pi_\infty(\mathbf{x}\,|\,\mathbf{d}) \propto p_{\text{prior}}(\mathbf{x}) \frac{1}{|\Sigma_\mathbf{E}|^{1/2}} \exp\left(-\frac{1}{2}\|\mathbf{d}-\mathbf{F}_\infty(\mathbf{x})\|^2_{\Sigma_\mathbf{E}^{-1}}\right), \tag{4}$$

where, for vectors $\mathbf{v} \in \mathbb{R}^m$ and square matrices $\mathsf{M} \in \mathbb{R}^m \times \mathbb{R}^m$, we define $\|\mathbf{v}\|^2_\mathsf{M}$ as $\mathbf{v}^\mathrm{T}\mathsf{M}\mathbf{v}$. The "$\infty$" subscript is used here to distinguish the ideal (infinite-precision) forward model from the noisy approximations we can actually compute.

Note that the argument of the exponential in (4) is the familiar cost function that is minimized in standard retrieval techniques using, e.g., the popular Levenberg–Marquard algorithm [12]. The value of $\mathbf{x}$ that is targeted by cost function minimization is known for obvious reasons as the "maximum likelihood" estimate. Moreover, if $p_{\text{prior}}(\mathbf{x})$ is written as $\mathrm{e}^{-\ln(1/p_{\text{prior}}(\mathbf{x}))}$, we can combine the exponentials, and interpret the negative sum of their arguments as a modified cost function where the additional term is $-\ln p_{\text{prior}}(\mathbf{x})$. This amounts to a statistical constraint that will shift the value of the optimal parameter values in the direction that maximizes $p_{\text{prior}}(\mathbf{x})$. For instance, very uncertain data (large values in $\Sigma_\mathbf{E}$) will downplay the importance of the original cost function and the minimization procedure will yield the mode of the prior distribution, irrespective of $\mathbf{d}$. Often, the new term is modulated by a weighting factor that can be made smaller as the minimization progresses.

In the present application (Monte Carlo-based forward 2D RT modeling), we do not have access to $\mathbf{F}_\infty$, but instead to a sequence of approximate models $\mathbf{F}_1, \mathbf{F}_2, \cdots$ with increasing but finite precision. In our framework, the approximate models are an unbiased sum of i.i.d. random variables, and so we are justified using a Gaussian error model

$$\mathbf{F}_j(\mathbf{x}) \sim \mathcal{N}(\mathbf{F}_\infty(\mathbf{x}), \Sigma_j(\mathbf{x})), \tag{5}$$

where $\Sigma_j(\mathbf{x}) = \text{Var}\{\mathbf{F}_j(\mathbf{x})\}$ that, we assume, can be estimated accurately. This is indeed the case when the forward model is solved by Monte Carlo techniques. See App. B of Part I for a formal description of our hierarchical modeling framework.

Equation (5) leads to an *enhanced* noise model at precision level j:

$$\begin{aligned} \mathbf{d} &= \mathbf{F}_j(\mathbf{x}) + \mathbf{E}_j(\mathbf{x}) + \mathbf{E}, \quad \mathbf{E} \sim \mathcal{N}(0, \Sigma_\mathbf{E}), \quad \mathbf{E} \perp\!\!\!\perp \mathbf{x}, \\ \mathbf{E}_j(\mathbf{x}) &\sim \mathcal{N}(0, \Sigma_j(\mathbf{x})), \quad \mathbf{E}_j(\mathbf{x}) \perp\!\!\!\perp \mathbf{E} \end{aligned} \tag{6}$$

and a likelihood at precision level j,

$$\pi_j(\mathbf{d}\,|\,\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^m|\Sigma_\mathbf{E}+\Sigma_j(\mathbf{x})|}} \times \exp\left(-\frac{1}{2}\|\mathbf{d}-\mathbf{F}_j(\mathbf{x})\|^2_{(\Sigma_\mathbf{E}+\Sigma_j(\mathbf{x}))^{-1}}\right). \tag{7}$$

Instead of one posterior, we now have a suite of *finite-precision* posteriors $\{\pi_j(\mathbf{x}\,|\,\mathbf{d})\}_{j=1}^\infty$:

$$\pi_j(\mathbf{x}\,|\,\mathbf{d}) \propto p_{\text{prior}}(\mathbf{x})\pi_j(\mathbf{d}\,|\,\mathbf{x}). \tag{8}$$

B. Application to 2D Scene Parameter Estimation

Rather than infer a specific value for the state vector, even with statistically reasonable uncertainty estimates, our goal here is to tally a complete multi-variate histogram for the posterior distribution of possible parameter values as efficiently and accurately as possible.

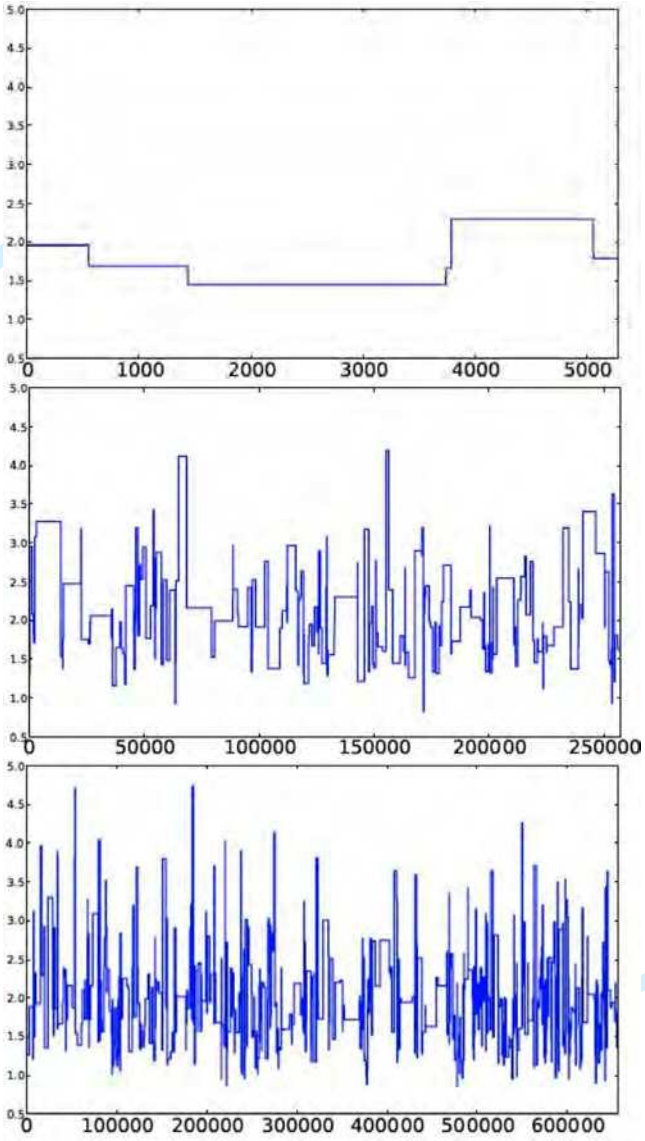

Fig. 2. Γ*-space sampling with three MCMC algorithms.* Only the projection of the random walk along the k_{p} axis is displayed. From top to bottom: #1) Metropolis–Hastings; #2) Two-level M–H; #3) Multi-level M–H with prescribed confidence [7]. In all cases, the code was run for 5 hours on a single 2.6 GHz Intel core.

Returning to our original notations, three different algorithms were tested for the estimation of $\pi_j(\gamma \,|\, \mathbf{d})$ by sampling Γ-space in various ways for use in (8). At the core of all three algorithms is the concept of a Markov chain

Monte Carlo (MCMC) process [13] where, like in the RT application, a Monte Carlo (a.k.a. "random quadrature" rule) is used to compute integrals. MCMC is essentially a random walk in the 5-dimensional Γ-space following certain rules that determine algorithmically the Markov chain's transition probabilities. The integrals of interest are the posterior probability levels, based on (8), in a large number of bins defined on a gridded version of Γ-space. Starting at some random point in Γ, the transition rules involve using a "proposal" to move to another value of γ; this proposal is accepted or rejected based on a criterion that expresses how helpful it is to update the estimate of the gridded posterior $\pi_j(\gamma \,|\, \mathbf{d})$. We note that the expression of the criterion does not require the posterior to be normalized; that is done after convergence is achieved.

Here, MCMC was used for Bayesian posterior estimation in a setting where the forward model for the data is an RT model implemented in a Monte Carlo scheme. As pointed out earlier, that RT modeling framework also uses Markov chain concepts and the integrals of interest are the radiances in each pixel of the images formed by the detector. Generally speaking, other kinds of (possibly polarized) 1D, 2D or 3D RT models are of course admissible in the part of the computation where the cost function (i.e., $\|\mathbf{d} - \mathbf{F}_j(\gamma)\|^2_{(\Sigma_\mathbf{E}+\Sigma_j(\gamma))^{-1}}$) is estimated for any number of remote sensing problems. Monte Carlo is however a natural choice in multi-dimensional RT, the whole multi-pixel inversion scheme we propose is therefore called the "MC^3 method."

The three MC^3 algorithms we tested are described in detail and inter-compared in [7], a companion paper to this one (beyond Part I). The baseline algorithm we tested (#1) was the well-known Metropolis–Hastings [14]–[16] technique applied to a single level of forward model precision. The two other algorithms build on that classic methodology. Algorithm #2 couples two model precision levels [17], [18] while #3, the new development in computational MCMC technology, couples multiple levels. Figure 2 illustrates the trace of a single coordinate of γ (namely, k_p) for a single MC^3 trajectory using the three algorithms running each for 5 hours on the same platform. The progress is striking since the efficiency of an MCMC sampling technique can be measured by how soon the all the bins (along the vertical axis in Fig. 2) are sufficiently well populated.

IV. Numerical Results

We avoid an "inverse crime" by generating data using a slightly different forward model than is used in the inversion. In particular, the integrals appearing in the transport solution (cf. Eq. (20) and Fig. 1 in Part I) are computed using an explicit Euler scheme with random step size $h \sim \mathcal{U}$[0.05 km, 0.15 km]. This introduces discretization error that mimics unavoidable model error due to the fact that real-world atmospheres and plumes will never follow our assumptions. For instance, even if we view the "$\cos^3$" terrain as a stand-in for a precise digital terrain model, our exponential parameterization of the background aerosol is just a convenient approximation that we exploited explicitly in the path recycling process. Tests where the inverse crime was committed deliberately, lead as expected to closer agreement between the assumed ("true") and retrieved γ, at a given signal-to-noise ratio (SNR). The assumed SNR is used to specify $\mathbf{E}$ in (3).

We start with SNR = 5. The four panels in Fig. 3a show the marginal posteriors for the plume-related elements of γ, namely, $(x_\mathrm{p}, z_\mathrm{p}, \rho_\mathrm{p}, k_\mathrm{p})$. We see that all but x_p are reasonably well characterized probabilistically. In particular

Fig. 3. *Plume parameter retrievals at SNR = 5.* See full page montage.

Fig. 4. *Plume parameter retrievals at SNR = 2.* See full page montage.

the estimated means for parameters $z_{\rm p}$, $\rho_{\rm p}$, and especially $k_{\rm p}$ are reasonably close to the assumed/true values. We can attribute the failure of the $x_{\rm p}$ inference to do as well to the fact that, unlike $z_{\rm p}$, varying its values does not move the response to new pixels in the image; review scene and bound geometry in Fig. 1. Consequently, the $x_{\rm p}$ inference would be the first to benefit from a second look at the scene from a different vantage point, say, directly overhead. Moreover, there is no fundamental difficulty in generalizing the Bayesian multi-pixel methodology used here to a multi-pixel/multi-angle one that would get us closer to a bone fide atmospheric tomography of the scene.

Figure 3b shows the MCMC random walk stops in two-dimensional hyper-planes of Γ-space; these are the intermediate data products that enable the estimation of the posterior distributions. The two panels of Fig. 3b also show more clearly why $x_{\rm p}$ localization is poor compared to $z_{\rm p}$ (left-hand panel) and why $k_{\rm p}$ characterization is better than for $\rho_{\rm p}$ (right-hand panel). This last $(k_{\rm p}, \rho_{\rm p})$ scatter plot shows a relatively tight anti-correlation between the two parameters; one can almost see a negative power law relationship between them. This is traceable to the fact that the impact of the plume on the pixel-scale signal is, to a first approximation, dependent on the product of $k_{\rm p}$ with the chord length of the intersection of pixel-specific lines starting at the detector with the circular plume. The lengths of these segments are clearly $\propto \rho_{\rm p}$. Fortunately, $k_{\rm p}$ is better quantified (narrower posterior) than $\rho_{\rm p}$ since it is the key parameter for identifying the gas when a continuous spectrum is collected.

From the remote sensing perspective, correlation between parameters of interest in the measured signals is unfortunate since an observed change can be attributed to either one. In this case, we wonder: Is the plume bigger or more absorbing? That said, the end-user of the remote sensing technology may not care that much: the objective may well be to assay the overall amount of absorbing gas, which goes as $k_{\rm p}\rho_{\rm p}^2$. That product is also a direct measure of the strength of the plume's signal, all pixels considered. This inspires us to look at how well that parameter combination is determined in the Bayesian retrieval. The result is plotted in Fig. 5 and, in this case, the posterior is remarkably different from the prior. In fact, the prior mean is closer to the truth than is the posterior mean. However, the most probable value is improved: the prior puts it at zero while the posterior puts it closer to the truth than is its mean.

Finally, the outcome for the background aerosol δc is shown in Fig. 7. Interestingly, this is the most narrowly retrieved parameter of all. This is not too surprising since plume parameters benefit almost exclusively from the 3-or-so pixels that view it directly while all of the pixels are populated by light scattered by the aerosol.

For comparison, we also show our results for SNR = 2 in Figs. 4ab for the plume parameters, Fig. 6 for the plume "mass factor," and Fig. 8 for the background aerosol perturbation. With more than half the SNR gone, retrieval performances have all suffered of course. The plume localization in $(x_{\rm p}, z_{\rm p})^{\rm T}$ is particularly affected, with a tendency to locate the plume higher than it is. However, although precision has deteriorated, the accuracy of the

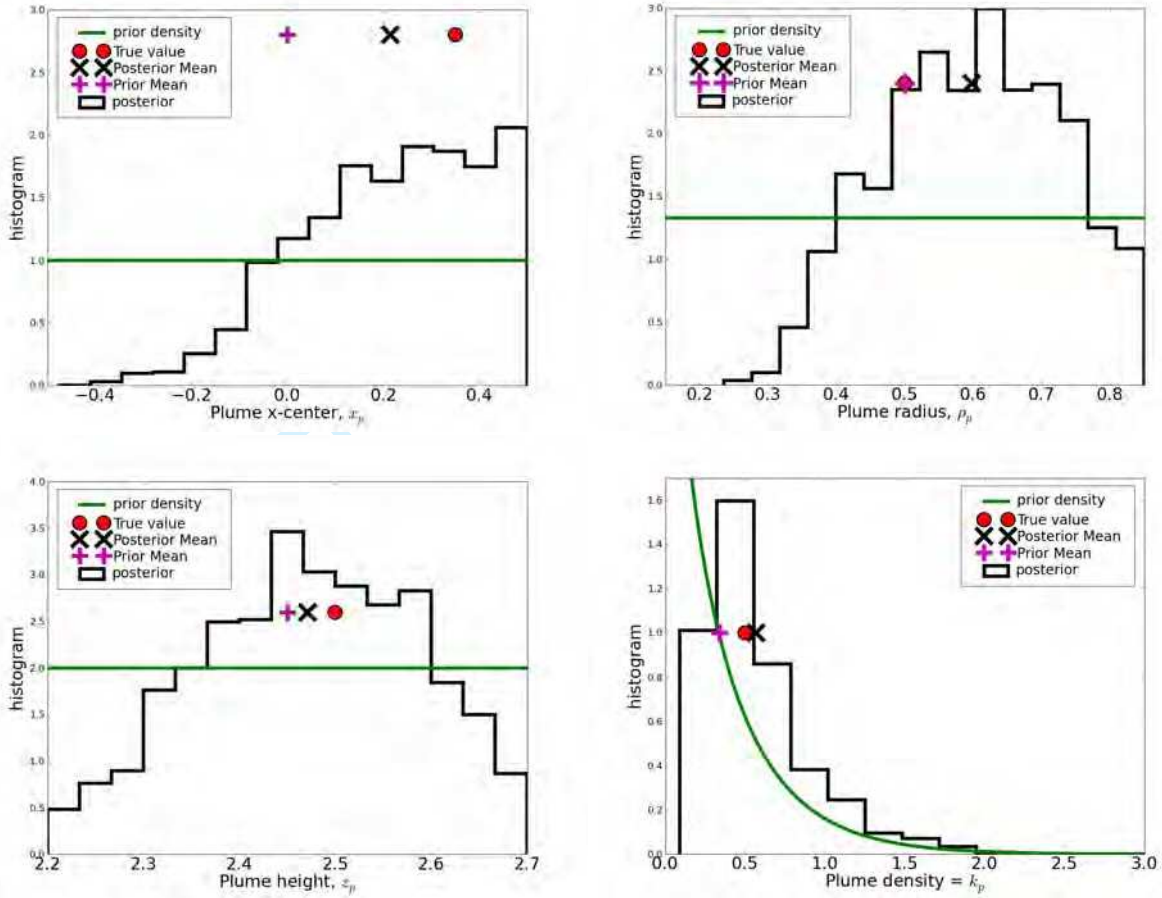

Fig. 3a. *Marginal posterior distributions at SNR = 5, without inverse crime, for plume-specific parameters.* Left: (top) x_p [km] and (bottom) z_p [km]. Right: (top) ρ_p [km] and (bottom) k_p [1/km]. More discussion in main text.

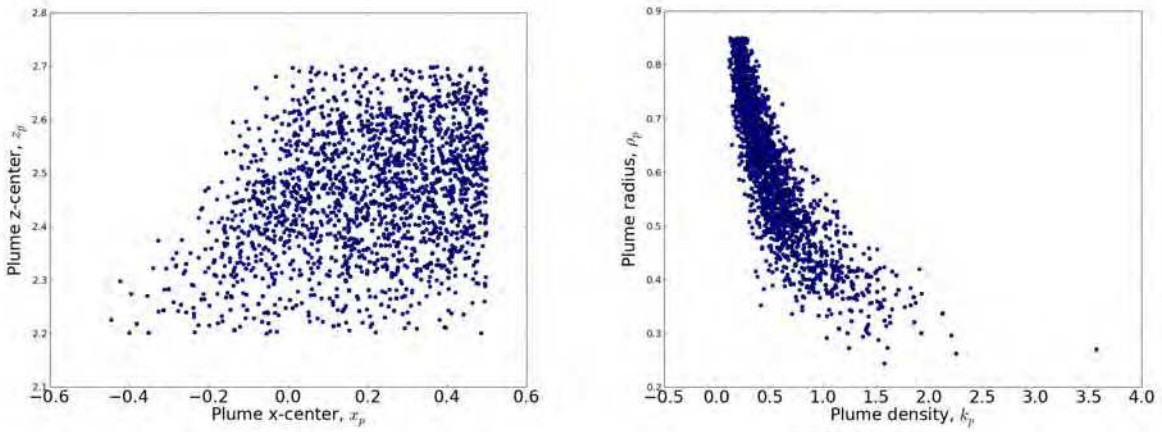

Fig. 3b. *MCMC samples for joint posterior distributions at SNR = 5, without inverse crime, for plume-specific parameters.* Left: (x_p, z_p). Right: (k_p,ρ_p). More discussion in main text.

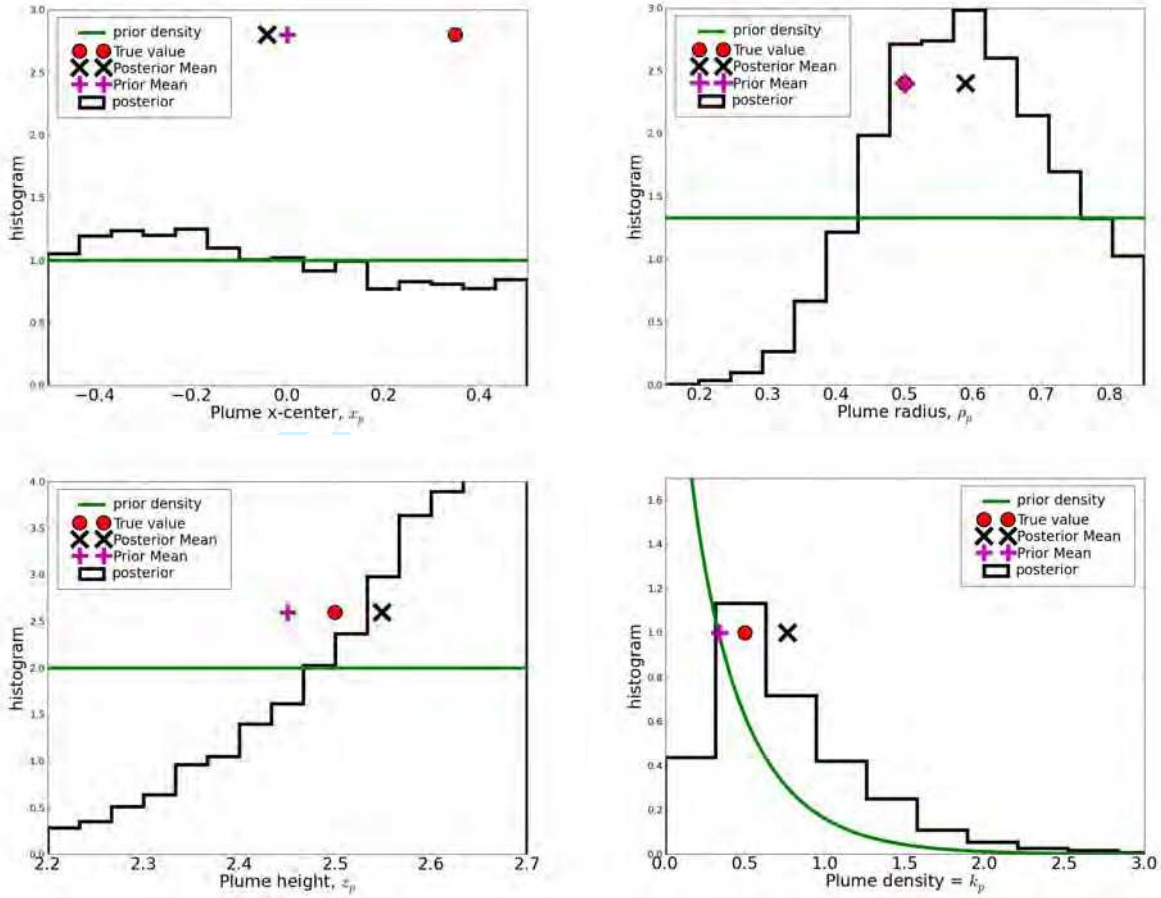

Fig. 4a. *Same as Fig. 6a but for* SNR = 2.

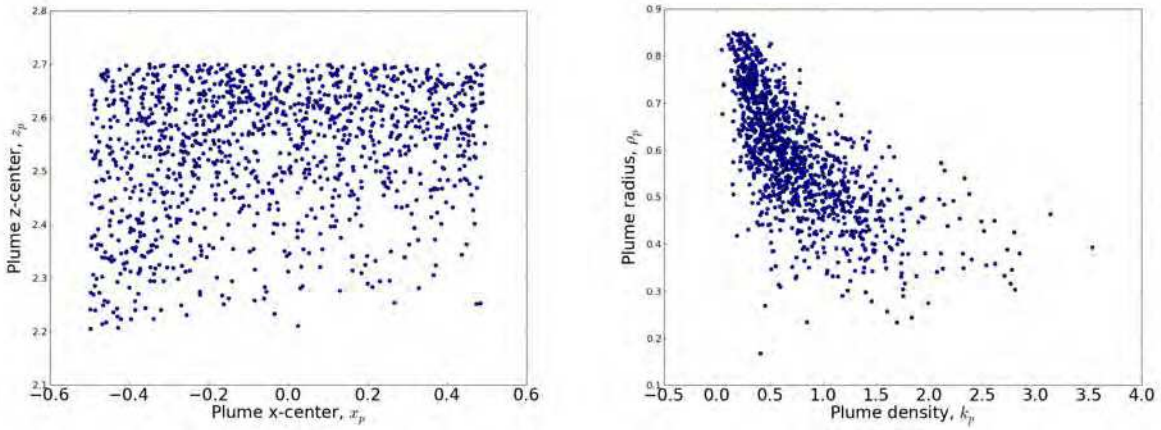

Fig. 4b. *Same as Fig. 6b but for* SNR = 2.

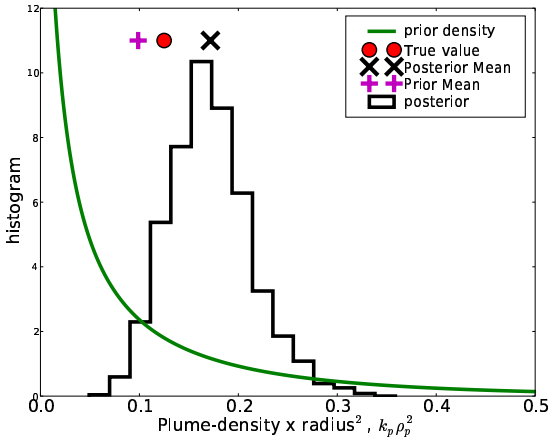

Fig. 5. *Bayesian inference for* $k_{\mathrm{P}}\rho_{\mathrm{P}}^2$ *at SNR = 5.*

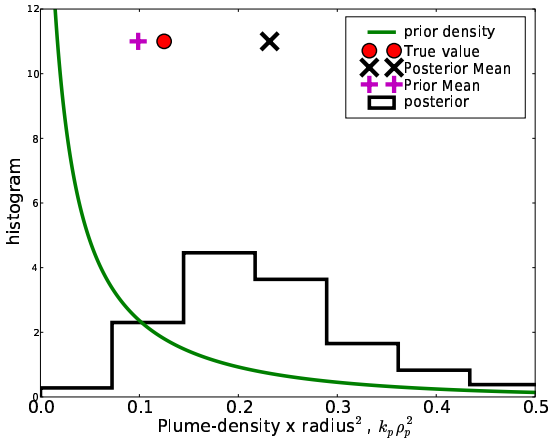

Fig. 6. *Bayesian inference for* $k_{\mathrm{P}}\rho_{\mathrm{P}}^2$ *at SNR = 2.*

δc retrieval is still quite good.

V. Summary & Outlook

In Part I, we introduced an innovative computational approach to multi-dimensional radiative transfer, namely, path-recycling Monte Carlo. Here we used it to define a hierarchical forward modeling framework for simulating remote sensing signals generated by a spatially variable scene at a variable level of precision. The terrain model has variable height and a height-dependent albedo mimicking a deep valley with sparse vegetation a low altitudes and dense vegetation at high altitudes, as viewed in the NIR spectrum. The atmosphere is composed of a partially known aerosol, with an exponential extinction profile with altitude, and there is a plume of absorbing gas with a

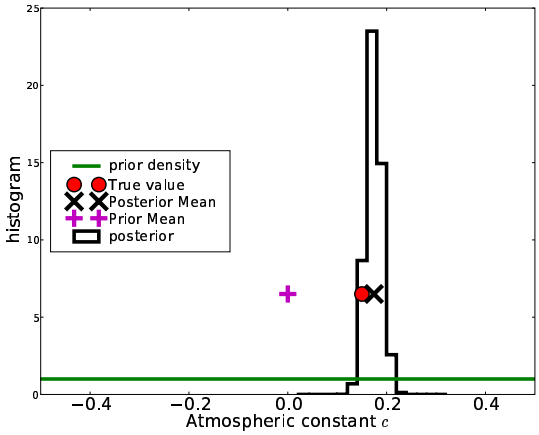

Fig. 7. *Results for the aerosol perturbation δc at SNR = 5.*

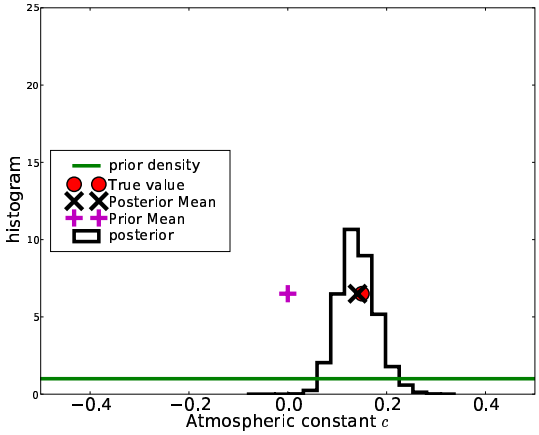

Fig. 8. *Results for the aerosol perturbation δc at SNR = 2.*

known cross-section per molecule but unknown location, size, and opacity. This is a plausible scenario when one is searching for observable evidence of nuclear proliferation activity.

The tell-tale plume is assumed uniform with a circular section in (x, z) coordinates and, for simplicity, the radiative transfer unfolds completely in these two spatial dimensions. The background aerosol's single scattering albedo and phase function are assumed known, as is its concentration at the lowest point in the scene but its column-integrated amount (measured by its optical depth) is varied by changing the characteristic scale height of the exponential profile. All told, there are five unknown quantities to retrieve from the remote sensing data: four plume-related parameters and one aerosol counterpart. We showed that this 2D atmospheric structure, because it is represented parametrically, can be reconstructed reasonably well using radiometrically-calibrated imagery from

a single sensor at close range, even at low spatial resolution and with modest (single-digit) signal-to-noise ratios.

To underscore the novel, inherently *multi-pixel* nature of the methodology, only a single wavelength is considered, presumably selected to detect the absorbing gas of interest. Where the 5-parameter retrieval does not perform too well is for localization of the plume along the direction viewed by the sensor. However, this aspect can certainly be improved by adding a second view from another vantage point. The resulting multi-angle/multi-pixel approach would be in essence a coarse atmospheric tomography using a parameterized representation of the spatial structure (instead of imposing regularization constraints). This keeps the actual number of structural unknowns small and manageable.

Multi-pixel retrieval algorithms introduced here require a forward radiative transfer model that predicts whole images and, in particular, how radiances in each pixel relate to each other via net transfer of radiant energy across pixel boundaries. That, in turn, requires a 3D radiative transfer model. (For this particular demo, a 2D radiative transfer model was used.) This is not to be confused with the terminology recently introduced by Dubovik et al. [19] where “multi-pixel” is used to describe a statistical constraint in the cost function used for an aerosol property retrieval predicated entirely on a polarized 1D radiative transfer model. Similar constraints have been used previously, for instance, in the operational aerosol retrieval developed for Multi-angle Imaging Spectro-Radiometer [20].

Finally, a Bayesian formulation of the remote sensing inverse problem was used. Consequently, the outcome is not a single value for each of the 5 parameters, even including uncertainty estimates, as would result from a standard cost function minimization approach. The derived product is in fact a whole multi-variate probability distribution for the parameters that is consistent with the data. The Bayesian inference machine is simple to implement using “MCMC” (Monte Carlo Markov chain) algorithms, but takes special effort and considerable innovation to ensure reasonably quick convergence to the desired posterior (that is, data-informed) distribution of possible parameter values.

As polarization diversity is added more systematically to multi-angle/multi-spectral remote sensing observations, we approach the limit of what each detected photon’s phase space has to offer:

- energy (from wavelength),
- momentum (from direction), and
- spin (from polarization).

This could seem to be the final frontier of remote sensing. Without minimizing in any way the importance of polarization, it is far from being the last piece of the atmosphere/surface remote sensing puzzle. This would only be true in the traditional mind-set of pixel-by-pixel processing where one-dimensional radiative transfer modeling suffices. Once multi-dimensional radiative transfer modeling is brought to the table to investigate pixel-to-pixel relationships using physics-based methods, a myriad new ways arise for extracting information from imaging sensor data.

ACKNOWLEDGMENT

This work was supported by NSF Grant DMS-0804696 (to Columbia U) and DOE Grant DE-FG52-08NA28779 (to Columbia and JPL). Part of this research was carried out at the JPL/CalTech, under a contract with the National Aeronautics and Space Administration. IL and GB are grateful for their productive collaboration with Prof. Youssef Marzouk at MIT.

REFERENCES

[1] I. Langmore, A. B. Davis, and G. Bal, "Parametric 3D scene reconstruction from imaging radiometry, Part I: The path-recycling forward Monte Carlo model," *IEEE Trans. Geosc. and Remote Sens.*, submitted.

[2] J. R. Schott, *Remote Sensing – The Image Chain Approach*, 2nd ed. Oxford (UK): Oxford University Press, 2007.

[3] ——, *Fundamentals of Polarimetric Remote Sensing*. Bellingham (Wa): SPIE, 2009.

[4] D. J. Diner, J. C. Beckert, T. H. Reilly, C. J. Bruegge, J. E. Conel, R. A. Kahn, J. V. Martonchik, T. P. Ackerman, R. Davies, S. A. W. Gerstl, H. R. Gordon, J.-P. Muller, R. B. Myneni, P. J. Sellers, B. Pinty, and M. Verstraete, "Multi-angle Imaging SpectroRadiometer (MISR) instrument description and experiment overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, pp. 1072–1087, 1998.

[5] P.-Y. Deschamps, F.-M. Bréon, M. Leroy, A. Podaire, A. Bricaud, J.-C. Buriez, and G. Sèze, "The POLDER mission: Instrument characteristics and scientific objectives," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, pp. 598–615, 1994.

[6] M. I. Mishchenko, B. Cairns, G. Kopp, C. F. Schueller, B. A. Fafaul, J. E. Hansen, R. J. Hooker, T. Itchkawich, H. B. Maring, and L. D. Travis, "Accurate monitoring of terrestrial aerosols and total solar irradiance: Introducing the Glory mission," *Bull. Am. Meteorol. Soc.*, vol. 88, pp. 677–691, 2007.

[7] G. Bal, I. Langmore, and Y. Marzouk, "Bayesian inverse problems with Monte Carlo forward models," 2011, to be submitted.

[8] C. K. Hayakawa, J. Spanier, F. Bevilacqua, A. K. Dunn, J. S. You, B. J. Tromberg, and V. Venugopalan, "Perturbation Monte Carlo methods to solve inverse photon migration problems in heterogeneous tissues," *Opt. Lett.*, vol. 26, pp. 1335–1337, 2001.

[9] C. K. Hayakawa and J. Spanier, "Perturbation Monte Carlo methods for the solution of inverse problems," in *Monte Carlo and Quasi Monte Carlo Methods 2002*, H. Niederreiter, Ed. New York (NY): Springer, 2004, pp. 227–242.

[10] J. Chen and X. Intes, "Time-gated perturbation Monte Carlo for whole body functional imaging in small animals," *Opt. Express*, vol. 17, pp. 19,566–19,579, 2009.

[11] W. M. Bolstad, *Introduction to Bayesian Statistics*, 2nd ed. New York (NY): John Wiley, 2007.

[12] P. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, 3rd ed., ser. Science/Engineering/Math. New York (NY): McGraw-Hill, 2002.

[13] B. A. Berg, *Markov Chain Monte Carlo Simulations and Their Statistical Analysis*. Singapore: World Scientific, 2004.

[14] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *Journal of Chemical Physics*, vol. 21, pp. 1087–1092, 1953.

[15] W. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.

[16] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. Springer, 2004.

[17] J. Andrés Christen and C. Fox, "Markov chain Monte Carlo using an approximation," *Journal of Computational and Graphical Statistics*, vol. 14, pp. 795–810, 2005.

[18] Y. Efendiev, T. Hou, and W. Luo, "Preconditioning Markov chain Monte Carlo simulations using coarse-scale models," *SIAM J. Sci. Comput.*, vol. 28, pp. 776–803, 2006.

[19] O. Dubovik, M. Herman, A. Holdak, T. Lapyonok, D. Tanré, J.-L. Deuzé, F. Ducos, A. Sinyuk, and A. Lopatin, "Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations," *Atmos. Meas. Techn.*, vol. 4, pp. 975–1018, 2011.

[20] J. V. Martonchik, D. J. Diner, K. A. Crean, and M. A. Bull, "Regional aerosol retrieval results from MISR," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, pp. 1520–1531, 2002.

Ian Langmore (BSc, Electrical Engineering, U. of California, San Diego 2002; MSc, Electrical Engineering, U. of California, San Diego 2004; PhD, Mathematics, U. of Washington 2008) worked on this manuscript as a postdoctoral researcher at Columbia University, Department of Applied Physics & Applied Mathematics, in New York (NY). His work there related to inverse problems in radiative transfer. Most of Ian's recent work involves Monte Carlo, including MCMC, algorithms. Since leaving Columbia in September 2011, he is working at Opera Solutions as a data scientist specializing in machine learning algorithms and Monte Carlo techniques.

Guillaume Bal (Diplôme, École Polytechnique 1993; PhD, Applied Mathematics, U. Paris VI 1997) joined the faculty of Columbia University's Department of Applied Physics & Applied Mathematics in 2001, after postdoctoral research at Stanford (1997–1999) and a stint as L.E. Dickson Instructor at the University of Chicago (1999–2001). Guillaume's research interests are: partial differential equations, equations with random coefficients, wave and particle propagation in heterogeneous media, mathematical and numerical analysis of inverse problems, analysis and numerical simulation of forward and inverse transport equations, and applications in medical imaging and geophysical imaging. Among other awards, he is the recipient of the 2011 Calderòn Prize for outstanding achievements in Inverse Problem Theory.

Anthony B. Davis (BSc, Physics, U. Paris VI 1977; MSc, Physics, U. de Montréal 1982; PhD, Physics, McGill 1992) has been working on cloud remote sensing problems, as well as the related problem of how the radiative properties of clouds impact the Earth's energy budget, for almost two decades successively at NASA's Goddard Space Flight Center in Greenbelt, Maryland, DOE's Los Alamos National Laboratory in New Mexico, and now at the Jet Propulsion Laboratory in Pasadena, Ca. He specializes in the challenges posed by the 3D shapes of real clouds to which he applies his expertise in theoretical and computational vector radiative transfer for physics-based prediction of the spectral and/or polarimetric signals from clouds measured by ground-, aircraft- and space-based sensors. Moreover, he has supported new instrument development in both passive (primarily solar) and active (pulsed laser source) approaches in cloud remote sensing. Recently, Anthony's interests have grown to include aerosol/surface property remote sensing, particularly in the vicinity of clouds, or other spatially complex situations, using multi-pixel/multi-angle spectral data with or without polarization.