Multipixel Retrieval of Structural and Optical Parameters in a 2-D Scene With a Path-Recycling Monte Carlo Forward Model and a New Bayesian Inference Engine

Ian Langmore, Anthony B. Davis, and Guillaume Bal

***Abstract*—Physics-based retrievals of atmosphere and/or surface properties are generally multi- or hyperspectral in nature; some use multi-angle information as well. Recently, polarization has been added to the available input from sensors and accordingly modeled with vector radiative transfer (RT). At any rate, a single pixel is processed at a time using a forward RT model predicated on 1-D transport theory. Neighboring pixels are sometimes considered but, generally, just to formulate statistical constraints on the inversion based on spatial context. Herein, we demonstrate the power to be harnessed by adding bona fide multipixel techniques to the mix. We use a forward RT model in 2-D, sufficient for this demonstration and easily extended to 3-D, for the response of a single-wavelength imaging sensor. The data, an image, is used to infer position, size, and opacity of an absorbing atmospheric plume somewhere in a deep valley in the presence of partially known/partially unknown aerosol. We first describe the necessary innovation to speed-up forward multidimensional RT. In spite of its reputation for inefficiency, we use a Monte Carlo (MC) technique. However, the adopted scheme is highly accelerated without loss of accuracy by using "recycled" MC paths. This forward model is then put to work in a novel Bayesian inversion adapted to this kind of RT model where it is straightforward to trade precision and efficiency. Retrievals target the plume properties and the specific amount of aerosol. In spite of the limited number of pixels and low signal-to-noise ratio, there is added value for certain nuclear treaty verification applications.**

***Index Terms*—Algorithm design and analysis, Bayesian methods, computational efficiency, computational geometry, hyperspectral imaging, Markov processes, maximum *a posteriori* estimation, radiometry, remote monitoring, signal processing algorithms, surface topography.**

I. Introduction, Context, and Overview

OPERATIONAL methods in the remote sensing of atmospheric and surface properties using physics-based techniques [1] have, at a minimum, the same limitations as the adopted forward model for the signal. Such a model is required to translate measured radiances into inherent structural, optical, physical, and chemical properties. However, its natural form is to predict radiances for given scene properties. The adopted inversion method that reverses this connection will introduce further limitations, as will instrument noise, and so on.

In retrievals using the UV through microwave spectrum, one endemic limitation follows from overly simplified physical and geometrical assumptions in the radiative transfer (RT) underlying the sensor signal prediction. This is most problematic in the visible (VIS) and near-infrared (NIR) spectral regions as they are dominated by atmospheric scattering and surface reflection. These radiation transport processes indeed make the prevailing RT highly nonlocal in the physically intuitive as well as mathematical senses of the word.

When it comes to operational remote sensing missions, data throughput volume considerations weigh heavily in favor of straightforward pixel-by-pixel processing, using the multi- or hyperspectral dimension of the data as best possible. When available, multi-angle information and, more and more frequently, polarimetric diversity will also be brought to bear on the physics-based remote sensing problem. However, once isolated from any spatial context, the natural assumption for the pixel-scale RT model is horizontal uniformity, i.e., the optical medium is assumed to be an infinite plane-parallel slab with, at most, some vertical structure. This assumption leads immediately to the textbook case of 1-D RT. Very conveniently for computational considerations, the azimuthal dependence is then amenable to a Fourier series decomposition with each mode being independent of the others [2]. Nature, however, is 3-D and complex—as demonstrated by almost every remote sensing image captured since the dawn of the technology! This disconnect between the conceptual model for the remote sensing signal and the reality that produces it can be a significant liability for the inverse problem.

There are two basic kinds of 3-D RT effect to worry about when using retrieval algorithms grounded in 1-D RT [3]. First, there is the effect of small-scale (unresolved) variability that dominates the 1-D RT model error when the observation scale (pixel size) is large. Second, there is the "pixel-adjacency" effect that dominates the 1-D RT model error when the observation scale is small. For the sake of argument, we can assume

Manuscript received November 22, 2011; revised July 9, 2012; accepted August 19, 2012. Date of publication November 20, 2012; date of current version April 18, 2013. This research was supported by the Department of Energy under Grant DE-FG52-08NA28779 (to Columbia and the Jet Propulsion Laboratory) and the National Science Foundation under Grant DMS-0804696 (to Columbia University).

I. Langmore and G. Bal are with the Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY 10027 USA (e-mail: ianlangmore@gmail.com; gb2030@columbia.edu).

A. B. Davis is with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109 USA (e-mail: Anthony.B.Davis@jpl.nasa.gov).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2012.2217380

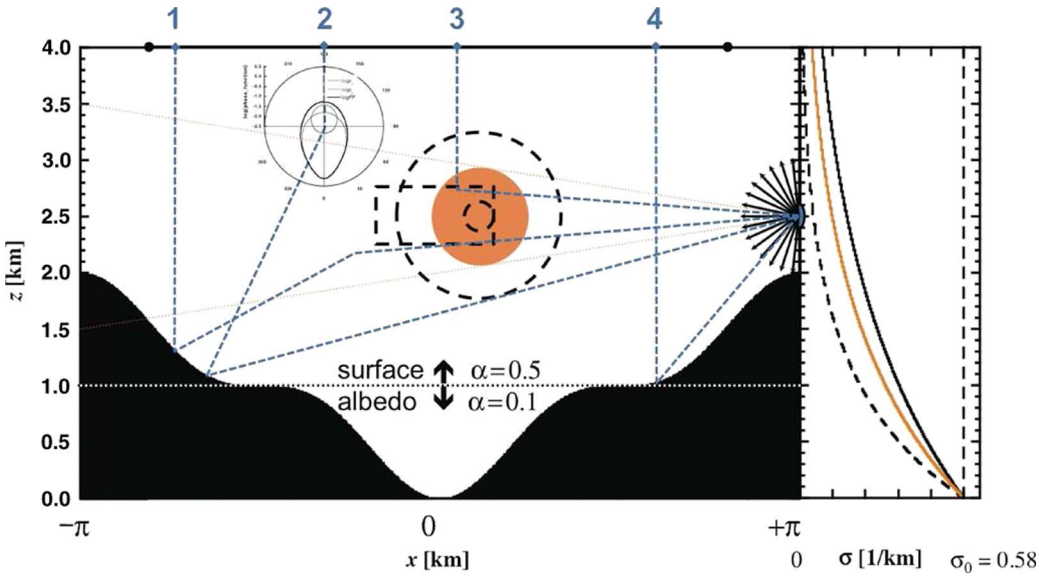

Fig. 1. Schematic representation of the optical medium. We show here the terrain with variable height from (3) at the lower boundary and the upper domain boundaries in (4) and (5). An instance of the absorbing gaseous plume is also drawn. For illustration, four solar rays are cast from the overhead direction, all of which reach a small hypothetical detector on the right-hand side and therefore qualify for path recycling. Two of these particular rays intersect the plume region in this instantiation. Dotted areas are the (prior) bounds for the Bayesian retrieval. To the right, with the same convention, are the vertical profiles of aerosol extinction; the reference value is in solid black. The polar plot at the first collision of path #2 illustrates the adopted aerosol phase function. See Fig. 2 for details.

that the vertical structure of the medium, including the surface, is perfectly known under any given pixel. The adjacency effect results from the near proximity of other pixels where the surface and/or atmosphere have a different structure and thus excite net horizontal fluxes, precisely what is neglected in 1-D RT. Much research into these topics has been performed primarily in 3-D cloudy scenes (see [3], [4], and references therein). Scenarios where the aerosol and molecular atmosphere is horizontally uniform while the surface, assumed flat, is not have also been investigated extensively (see [5]–[19], and more references therein). Topography effects have been studied as well [20], among others. In all of these situations, a common question is: How far the 3-D effects can be sensed? In other words, what is the scale that divides the aforementioned "small" and "large" pixel sizes? It is fair to say that, overall, much less has been done about mitigating either large- or small-scale 3-D RT effects in current remote sensing operations, let alone applications.

We focus here on the later effect: cross-pixel/adjacency transport. In fact, rather than see this as a nuisance from a 1-D RT perspective, we take a resolutely 3-D RT perspective and indeed exploit the effect in a remote sensing problem. This is a radical departure from mainstream physics-based remote sensing of the environment where retrieval methods are invariably based on multi- or hyperspectral analyses of pixel-scale data, or maybe at a coarser resolution. At any rate, the standard inversion scheme uses a forward 1-D RT model that optimally reproduces the purely spectral observations, e.g., MODerate resolution Imaging Spectrometer [21] data. In some cases, this leads to the retrieval of a vertical profile in a thermodynamic variable and/or a chemical composition, e.g., using the Atmospheric Infra-Red Sounder [22]. In other cases, multi-angle information is brought to bear, e.g., with the Multi-angle Imaging Spectro-Radiometer (MISR) [23]. Some of those exploit polarization measurements as well [24], e.g., the POLarization and Directionality of the Earth's Reflectances (POLDER) [25] sensor.

To demonstrate more clearly how new information can be gained from the 3-D RT point of view, we simulate here a retrieval of spatial structure using a purely monochromatic but multipixel technique. This of course calls for a 3-D RT forward model and an inversion methodology capable of driving it. The 3-D RT model, a Monte Carlo (MC) scheme, uses a specific parametric representation of the scene. Envision a deep valley or gorge with a known profile in altitude and reflectivity (see Fig. 1). This terrain is filled and overlaid with a stratified scattering atmosphere representative of a typical background aerosol that is imperfectly known. Somewhere in this canyon, there is a finite (assumed) uniform gaseous plume made of a purely absorbing trace gas of interest; also, the plume's geometry is assumed to be that of a cylinder lying horizontally. The challenge is to determine the position, diameter, and opacity of the (admittedly idealized) plume, along with a refined estimate of the ambient aerosol load, using a single-channel imaging sensor that captures reflected and scattered sunlight.

This is a scenario of significant interest in nuclear facility monitoring [26]. Traditionally, gaseous plume remote sensing is performed in the spectral domain using either NIR or thermal IR (TIR) wavelengths, typically using hyperspectral sensors [26]–[28], among others. In that case spatial information can only be deduced easily by segmenting the image in the plane perpendicular to the incoming beam. Vertical profiling would call for well-calibrated high-resolution TIR data and detailed ancillary information about thermal stratification. Here, we simply assume that the NIR spectral band of the crude monochromatic imager has been tuned to an absorption feature of the targeted gas.

The paper is outlined as follows. In the following Section II, we describe in precise and general mathematical terms the forward RT problem that, for simplicity, is cast in two spatial dimensions, hence just one angular variable. In Section III, the specific scene geometry is set, optical properties are prescribed,

and remote sensing unknowns are defined. In Section IV, the numerical solution using the new "path-recycling" MC scheme is described in simple algorithmic terms, as well as in probabilistic terms using measure-theoretical concepts; its performance is compared with the standard MC method. In Section V, we describe the inverse transport method, which uses a Bayesian framework. Full mathematical details are provided in a companion paper by Bal *et al.* [29]. Section VI analyzes the outcome of a representative selection from our extensive numerical experimentation. We offer our concluding remarks and an outlook on future work in Section VII.

Appendix A covers certain technical details on MC on path recycling, starting from the measure-theoretical basis of the method. Appendix B formalizes the notion of a discrete hierarchy of forward MC models with varying precision at the cost of longer or shorter run times, which is key to the Bayesian analysis.

II. Two-Dimensional RT Equation and Associated Boundary Conditions

In steady-state 3-D RT, a monochromatic radiance field I depends on three spatial variables $\vec{r} = (x, y, z)^{\mathrm{T}}$, where z is customarily taken to be altitude above some reference point, and an angular variable $\vec{\Omega}$, with $\|\vec{\Omega}\| = 1$, that is usually defined by two polar coordinates (θ, ϕ). In the present study, we can restrict our considerations to 2-D RT without losing the ability to provide a proof-of-concept in multipixel remote sensing that exploits cross-pixel radiation transport. Consequently, we will use just two spatial coordinates $\vec{r} = (x, z)^{\mathrm{T}}$ and a single angular variable $\vec{\Omega}(\theta) = (\sin\theta, \cos\theta)^{\mathrm{T}}$ where $\theta = 0$ is up (increasing z). We should bear in mind that the units for $I(\vec{r}, \vec{\Omega})$ in 2-D are [photons/s/m/rad] (as opposed to [photons/s/m^2/sr] in the usual 3-D setting).

That said, the general steady-state monochromatic RT equation for $I(\vec{r}, \vec{\Omega})$ looks the same in 2-D as in 3-D

$$\left[\vec{\Omega}\cdot\nabla + \sigma(\vec{r})\right] I = \sigma_{\mathrm{s}}(\vec{r}) \int p_{\mathrm{v}}(\vec{r}, \vec{\Omega}' \to \vec{\Omega}) I(\vec{r}, \vec{\Omega}') \mathrm{d}\vec{\Omega}' + Q_{\mathrm{v}}(\vec{r}, \vec{\Omega}) \quad (1)$$

for $\vec{r} \in \mathrm{R}$, an open connected subset of $\mathbb{R}^2$. We have introduced here the extinction, $\sigma(\vec{r})$, and scattering, $\sigma_{\mathrm{s}}(\vec{r})$, coefficients (still in [m^{-1}]), the phase function $p_{\mathrm{v}}(\vec{r}, \vec{\Omega} \to \vec{\Omega}')$, and the volume source term $Q_{\mathrm{v}}(\vec{r}, \vec{\Omega})$ (expressed here in [photons/s/m^2/rad]). The phase function (in [1/rad]) is normalized in such a way that $\int p_{\mathrm{v}}(\vec{r}, \vec{\Omega} \to \vec{\Omega}')\mathrm{d}\vec{\Omega}' = 1$, with $\mathrm{d}\vec{\Omega}' = \mathrm{d}\theta'$.

Boundary conditions (BCs) can similarly be expressed in very general terms. Letting $\partial\mathrm{R}$ denote the closed boundary of R and $\vec{n}_{\vec{r}}$ its *outward* normal at $\vec{r}$, with $\vec{r}$-dependent components $(\sin\theta_{\vec{n}}, \cos\theta_{\vec{n}})^{\mathrm{T}}$, we have

$$\left|\vec{\Omega}\cdot\vec{n}(\vec{r})\right| I(\vec{r}, \vec{\Omega}) = \alpha(\vec{r}) \int_{\vec{\Omega}'\cdot\vec{n}(\vec{r})>0} p_{\mathrm{s}}(\vec{r}, \vec{\Omega}' \to \vec{\Omega}) I(\vec{r}, \vec{\Omega}') \vec{\Omega}'\cdot\vec{n}(\vec{r}) \mathrm{d}\vec{\Omega}' + \left|\vec{\Omega}\cdot\vec{n}(\vec{r})\right| Q_{\mathrm{s}}(\vec{r}, \vec{\Omega}) \quad (2)$$

for $\vec{r} \in \partial\mathrm{R}$ and $\vec{\Omega}(\theta)$, $\theta \in (-\pi, +\pi]$, such that $\vec{\Omega}\cdot\vec{n}(\vec{r}) < 0$. We have introduced here the surface's albedo $\alpha(\vec{r})$ and its phase function $p_{\mathrm{s}}(\vec{r}, \vec{\Omega} \to \vec{\Omega}')$ (in [1/rad], like its volume counterpart), with $\vec{\Omega}\cdot\vec{n}(\vec{r}) > 0$, to describe bi-directional reflection, as well as the boundary source term $Q_{\mathrm{s}}(\vec{r}, \vec{\Omega})$ (in [photons/s/m/rad]).

We will refer to the union of R and $\partial\mathrm{R}$ as the "optical medium," and to $(\mathrm{R} \cup \partial\mathrm{R}) \times (-\pi, +\pi]$ as the "transport space."

Equations (1) and (2) determine mathematically the forward RT problem, once R and all the coefficients and phase functions ("optical properties") are specified, e.g., as in the next section. The "flatland" remote sensing problem is to infer quantitative information about the structure or properties of the optical medium, given some (generally quite sparse) sampling of the $I(\vec{r}, \vec{\Omega})$ field in transport space, typically with $\vec{r} \in \partial\mathrm{R}$ and $\vec{\Omega}\cdot\vec{n}(\vec{r}) > 0$ (outgoing radiation). Ideally, the inferred quantities should be accompanied with an estimate of their uncertainty.

III. Scene Geometry and Optical Properties

Structural and optical quantities, both assumed and held fixed or parameterized and susceptible to vary, are defined.

A. Terrain and Illumination

Fig. 1 shows a schematic of the optical medium. For simplicity, the lower boundary is represented by

$$z(x) = \frac{H_{\mathrm{sfc}}}{2}\left(1 - \cos^3\left(\frac{2\pi x}{L}\right)\right), -L/2 \le x \le +L/2 \quad (3)$$

where we set $L = 2\pi$ [km] and $H_{\mathrm{sfc}} = 2$ [km]. The width-to-depth aspect ratio of this mirror-symmetric terrain model is $L/H_{\mathrm{sfc}} = \pi$. More complex asymmetric terrains, including Digital Elevation Models (DEMs) could be used in straightforward generalizations of our forward RT model.

To form $\partial\mathrm{R}$, this lower boundary is complemented by a "sky" defined as

$$\{z = H_{\mathrm{sky}}; -L/2 \le x \le +L/2\} \quad (4)$$

where we set $H_{\mathrm{sky}} = 4$ [km], and by lateral boundaries

$$\{x = \pm L/2; H_{\mathrm{sfc}} < z < H_{\mathrm{sky}}\} \quad (5)$$

where we apply open/absorbing BCs. We thus set $\alpha(z) = 0$ for $H_{\mathrm{sfc}} < z \le H_{\mathrm{sky}}$ in (2). This assumption contrasts with the commonly used periodic lateral BCs over the cell $-L/2 < x < +L/2$ associated an absorbing BC only at $z = H_{\mathrm{sky}}$. However, this simplification does not affect the outcome of the present proof of concept.

Again for simplicity, we emulate scene illumination by an overhead sun (solar zenith angle, $\theta_0 = \pi$ [rad]) by setting $Q_{\mathrm{v}}(x, z, \theta) \equiv 0$ in (1) and

$$Q_{\mathrm{s}}(x, z, \theta) = \begin{cases} F_0\delta(\theta - \theta_0), & \text{if } z = H_{\mathrm{sky}}, \text{ and } |x| < 2.5 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

in (2). The constraint on x (no light emitted from $2.5 < |x| < L/2 = \pi$) reduces the impact of the simplifying assumption of

open (rather than periodic) lateral BCs. Here, F_0 denotes the 2-D counterpart (measured in [photons/s/m]) of the incoming solar irradiance at the wavelength of interest.

The surface is assumed everywhere Lambertian, i.e.,

$$p_{\rm s}(\vec{r},\vec{\Omega}\to\vec{\Omega}') \equiv \left|\vec{\Omega}'\cdot\vec{n}(\vec{r})\right|/2. \tag{7}$$

Note the normalization by 2, instead of π in 3-D. It follows directly from the requirement that

$$\int\limits_{\vec{\Omega}'\cdot\vec{n}(\vec{r})<0} p_{\rm s}(\vec{r},\vec{\Omega}\to\vec{\Omega}')\mathrm{d}\vec{\Omega}' = 1$$

where $\mathrm{d}\vec{\Omega}' = \mathrm{d}\theta'$; note that $\vec{\Omega}'\cdot\vec{n}(\vec{r}) = \cos(\theta' - \theta_{\vec{n}(\vec{r})})$, and integration limits $(\cos(\theta' - \theta_{\vec{n}(\vec{r})}) < 0)$ are therefore $\theta_{\vec{n}(\vec{r})} + \pi/2$ and $\theta_{\vec{n}(\vec{r})} + 3\pi/2$. For future (MC implementation) use, this means that the new direction of propagation is $\theta' = (\theta_{\vec{n}(\vec{r})} + \pi) + \theta_{\rm r}(\xi)$ where

$$\theta_{\rm r}(\xi) = \sin^{-1}(1-2\xi) \tag{8}$$

is the diffuse reflection angle, ξ being drawn from a uniform distribution on (0,1). Equivalently, $\theta_{\rm r}(\xi) = \pm\sin^{-1}\xi$ where the sign is chosen at random. This contrasts with the usual 3-D case where $\theta_{\rm r}(\xi) = \cos^{-1}\sqrt{\xi}$ and the azimuthal angle is chosen at random in $(-\pi, +\pi]$.

Surface albedo is given by a piecewise constant function of z

$$\alpha(\vec{r}) \equiv \alpha(z) = \begin{cases} 0.1, & 0 \le z \le H_{\rm sfc}/2 \\ 0.5, & H_{\rm sfc}/2 < z \le H_{\rm sfc} \end{cases} \tag{9}$$

as expected when vegetation cover changes with altitude. In this case, we go suddenly from sparse to dense vegetation as altitude increases assuming an NIR wavelength (where there are many gaseous absorption features). This is suggestive of an industrial facility embedded in a hilly landscape. A valley surrounded by high mountains would have the opposite trend with altitude. The scale implied by the values of L and $H_{\rm sfc}$ suggests the later scenario but, in the end, only the non-dimensional numbers matter: the terrain aspect ratio $(L/H_{\rm sfc})$, and the Aerosol Optical Depth (AOD) introduced next.

B. Aerosols

The optical properties of the atmosphere (region R) are parameterized as

$$\sigma(\vec{r}) \equiv \sigma(z) = \sigma_0 \mathrm{e}^{-cz} \tag{10}$$
$$\sigma_{\rm s}(\vec{r}) \equiv \sigma_{\rm s}(z) = \varpi_0\sigma(z) \tag{11}$$

where:

- $\varpi_0 = 0.9$ for the aerosol single scattering albedo (SSA), a typical value, assumed constant with height;
- σ_0 is the extinction coefficient at the low point in the terrain $(x = z = 0)$, and $1/c$ is the characteristic scale height of the aerosol.

The "baseline" aerosol is defined by $c = c_0 = 0.5\ [\mathrm{km}^{-1}](1/c_0 = 2\ [\mathrm{km}])$. In general, the AOD, measured vertically over the low point in the terrain, is given by

$$\tau_{\rm a}(c) = \frac{\sigma_0}{c}(1 - \mathrm{e}^{-cH_{\rm sky}}) \tag{12}$$

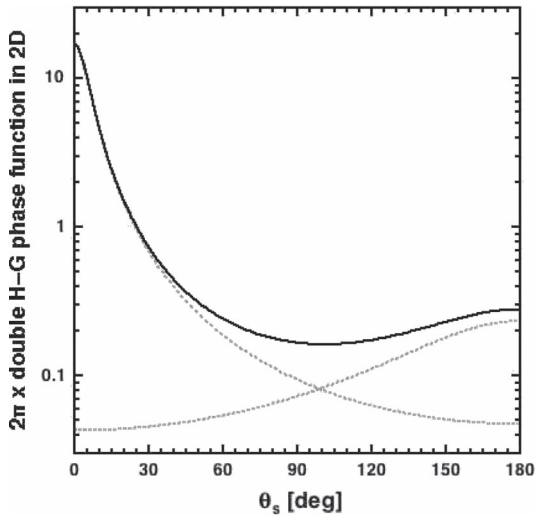

Fig. 2. Adopted phase function for aerosol scattering. Expressions for the H–G model are found in (13) and (14). Parameters are: $f_1 = 0.9$; $g_1 = +0.8$; $g_2 = -0.4$. In dotted gray lines: the two components separately. A polar version of this plot can be seen in Fig. 1.

and its baseline value is set to unity, a choice that leads to $\sigma_0 \approx 0.58\ [\mathrm{km}^{-1}]$. This scenario corresponds to a relatively heavy aerosol load, in other words, quite hazy observation conditions. We will, however, allow for other values further on by varying c.

The scattering phase function is taken to be everywhere equal to a 2-D counterpart of the double [31] Henyey–Greenstein (H–G) model [30], namely

$$p_{\rm v}(\vec{r},\theta\to\theta') \equiv f_1 p_{g_1}(\theta_{\rm s}) + (1-f_1)p_{g_2}(\theta_{\rm s}) \tag{13}$$

where $\theta_{\rm s} = |\theta' - \theta|$ in 2-D and [32], [33]

$$p_g(\theta_{\rm s}) = \left(\frac{1}{2\pi}\right)\frac{1-g^2}{1+g^2-2g\cos\theta_{\rm s}}. \tag{14}$$

Parameter g is the asymmetry factor of the above H–G phase function model defined, as in 3-D, as the angular mean of $\cos\theta_{\rm s}$. This leads to $g = f_1 g_1 + (1-f_1)g_2$ for the double H–G model in (13). We assume $(f_1, g_1, g_2) = (0.9, +0.8, -0.4)$, hence $g = 0.68$, a representative value for a background aerosol. Fig. 2 shows this phase function model along with its forward- and backward-peaked components (is it also plotted in log-radial axis as an inset in Fig. 1).

Since $\theta_{\rm s} = \theta' - \theta$ in 2-D, the scattering kernel in (1) is simply a convolution of the radiance field with the phase function in direction space. For deterministic solutions of the 2-D RT equation, it is therefore important to know the coefficients of the Fourier-cosine series for (14), which are simply g^n for $n \ge 0$ (precisely like for the required Legendre series expansion of the 3-D H–G model); this result was used to determine the three parameters so they approximately fit the three first Legendre moments of a real fine-mode aerosol (hence, approximate its shape in Fig. 2 for our 2-D world). For a MC solution of the RT equation, as is used here, it is best if one knows explicitly the scattering angle $\theta_{\rm s}(\xi)$ at which the definite integral of (14), starting from $-\pi$, equals $\xi \in [0,1]$. This turns out to be

$$\theta_{\rm s}(\xi) = 2\tan^{-1}\left[\left(\frac{1-g}{1+g}\right)\tan\left((\xi-1/2)\pi\right)\right]. \tag{15}$$

TABLE I
SUMMARY OF FIXED ATMOSPHERE–SURFACE PARAMETERS

Parameter	Symbol	Value	Unit
Minimum height		0	[km]
Maximum height	$H_{\rm sky}$	4	[km]
Terrain height	$H_{\rm sfc}$	2	[km]
Terrain width	L	2π	[km]
Solar zenith angle	θ_0	π	[rad]
Aerosol SSA	ϖ_0	0.9	[-]
Aerosol asymmetry factor	g	0.68	[-]
Aerosol scale height	$1/c_0$	2	[km]
Aerosol extinction at $z=0$	σ_0	0.58	[1/km]

For the double H–G phase function, linear combinations of the Fourier coefficients are computed according to (13). In the MC simulations to follow, we draw the scattering angle from either $p_{g_1}(\theta_{\rm s})$ or $p_{g_2}(\theta_{\rm s})$ based on a simple Bernoulli trial that decides which asymmetry factor to use in (15).

C. Atmospheric Variables

All of the above optical properties are held constant and assumed known in the remote sensing simulation studies described in later sections of this report. However, the aerosol stratification parameter c_0 is only a reference value, and we consider the aerosol only partially known. The actual atmosphere is given by the same σ_0 in (10) and (11), but with

$$c = c_0 + \delta c \tag{16}$$

where $\delta c \geq -c_0$ is a perturbation on the baseline inverse scale height c_0. It will automatically impact $\tau_{\rm a}(c)$ according to (12), hence more (less) total aerosol when $\delta c < (>) 0$.

Moreover, the atmosphere contains a uniform plume of purely absorbing gas that we represent parametrically as a circular region

$$A = \left\{(x,z)^{\rm T} \in R; (x-x_{\rm p})^2 + (z-z_{\rm p})^2 < \rho_{\rm p}^2\right\}. \tag{17}$$

Therein, the absorption coefficient becomes

$$\sigma_{\rm a}(\vec{r}) = (1-\varpi_0)\sigma(z) + k_{\rm p} \tag{18}$$

whereas only the first term applies outside of region A. We will vary $k_{\rm p}$ and, consequently, the absorption optical diameter of the plume $\tau_{\rm p} = 2\rho_{\rm p}k_{\rm p}$. Another interesting property of the plume is its total mass, which is $\propto k_{\rm p}\rho_{\rm p}^2$ since $k_{\rm p}$ is the product of the absorbing particles' volume density, actually *surface* density in 2-D, and their (presumably known) cross section for absorption.

D. Summary

Table I displays the key properties of the atmosphere–surface system that remain fixed in the present study. This defines the "reference world." Table II describes the continuum of "γ worlds." These γ worlds constitute the realm of possibilities in the Bayesian retrieval simulated later on in the paper.

TABLE II
SUMMARY OF VARIABLE ATMOSPHERIC PARAMETERS TARGETED IN BAYESIAN RETRIEVAL

Parameter	Symbol	Truth	Min	Max	Unit	Prior
Plume's x-position	$x_{\rm p}$	+0.35	-0.5	$+0.5$	[km]	uniform
Plume's z-position	$z_{\rm p}$	2.5	2.2	2.7	[km]	uniform
Plume's radius	$\rho_{\rm p}$	0.5	0.15	0.85	[km]	uniform
Plume's absorption coef.	$k_{\rm p}$	0.5	0	∞	[1/km]	Gamma
Aerosol perturbation	δc	+0.15	$-c_0$	$+c_0$	[1/km]	uniform

Formally, γ represents the 5-D state vector of remote sensing unknowns, viz.,

$$\gamma = (x_{\rm p}, z_{\rm p}, \rho_{\rm p}, k_{\rm p}; \delta c). \tag{19}$$

Other parameters that could be added to the above list of variables amenable to the adopted MC path-recycling methodology are:

- σ_0 in (10), the overall multiplier of AOD;
- ϖ_0 in (11), the SSA;
- three double H–G phase function parameters in (13);
- two values assigned to the z-dependent surface albedo in (9).

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

$$\Gamma = [x_{\min}, x_{\max}] \times [z_{\min}, z_{\max}] \times [\rho_{\min}, \rho_{\max}] \times [0,\infty) \times [\delta c_{\min}, \delta c_{\max}]. \tag{20}$$

IV. COMPUTATIONAL RT MODEL

We describe here the physics-based numerical simulation of the single monochromatic imaging sensor used in the scene reconstruction from remote sensing data.

A. Adopted MC Scheme

"Photon" trajectories or histories, as they are commonly but unphysically [34] called, are generated in transport space as follows in the plume-free case where $\varpi_0 = \sigma_{\rm s}/\sigma$ is constant.

1) A starting position and direction $(\vec{r}_0, \vec{\Omega}_0)$ are drawn from the source probability distribution $Q_{\rm s}$ in (6). In this case, $z_0 = H_{\rm sky}$, x_0 is drawn randomly from $(-2.5, +2.5)$, and θ_0 is set to π.
2) The photon travels along the path $\vec{r}(t) = \vec{r}_0 + t\vec{\Omega}_0, t > 0$, interacting at point $\vec{r}(t_1)$ with a cumulative probability distribution given by

$$P(t) = \Pr\{0 < t_1 < t\} = 1 - \exp\left(-\int_0^t \sigma(\vec{r}_0 + t_1\vec{\Omega}_0){\rm d}t_1\right) \tag{21}$$

using (10) with $c = c_0$. In other words, the optical distance would then be defined in terms of t_1: $\tau(t_1) = \int_0^{t_1} \sigma(\vec{r}_0 + s\vec{\Omega}_0){\rm d}s$, which is exponentially distributed

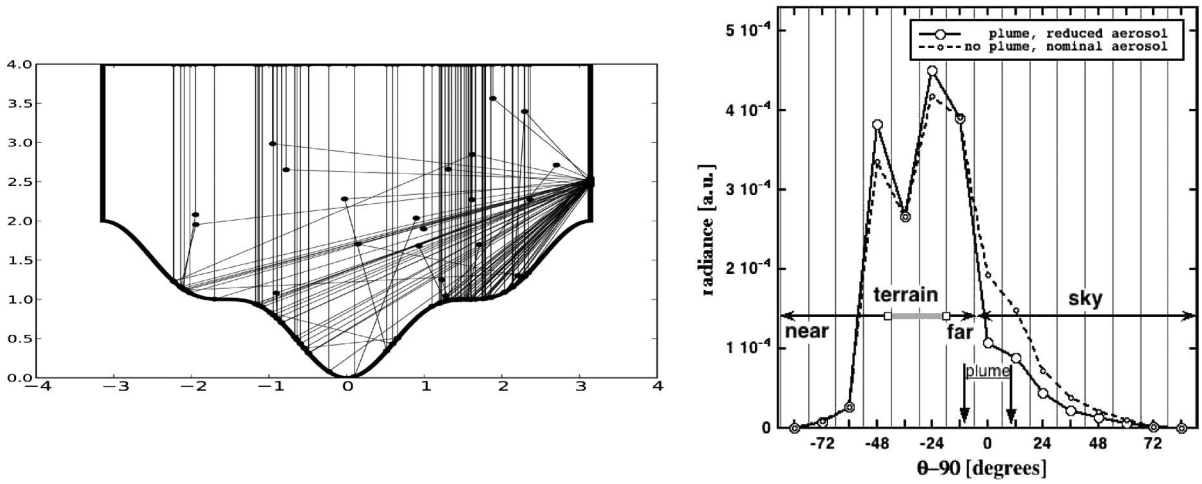

Fig. 3. (Left) Several photon trajectories that all ended at the detector. Seventy-five trajectories (out of ≈26 000 total sample) hit the small (0.1 [km]) detector above the terrain on the right-hand side of the medium. These are the only ones of interest in the following plume characterization by remote sensing. (Right) Examples of "images" in 2-D captured by the detector. Fluxes crossing the right-hand lateral boundary are measured and assigned to the 15 "pixels" in the image. A "fish-eye" optics is assumed (180° field-of-view). Going left to right, they are ordered from nadir ($\theta = 180°$) to zenith ($\theta = 0$), with angles re-centered on the horizontal look ($\theta = 90°$). Hence, to the right ($+^{\text{ve}}$ angles), we are looking at sky and, to the left ($-^{\text{ve}}$ angles), we go from sky into the partially reflective terrain. Directions of transitions from bright to dark and back to bright surface at $z = 1$ [km] are indicated by squares. Two situations are displayed. Dashed line: No absorbing plume and nominal aerosol, $\gamma = (\cdot, \cdot, \cdot, 0; 0)$ in (19). Solid line: $\gamma = (+0.35\ [\text{km}], 2.5\ [\text{km}], 0.5\ [\text{km}], 0.5\ [\text{km}^{-1}]; 0.15\ [\text{km}^{-1}])$, to illustrate the path-recycling technique. Direct lines of sight to the plume edges are indicated by the downward pointers. This plume and aerosol is used as "truth" in the Bayesian inverse problem solution experiments presented in Sections V and VI.

with unit mean. Such random numbers are generated by $-\ln \xi$ where ξ is drawn from a uniform distribution on (0,1). If the photon does not interact in the volume R, it will always interact with the boundary ∂R.

3) At the interaction point $\vec{r}_1$, the photon is either absorbed or scattered.
   - At a surface interaction, $\vec{r}_1 \in \partial$R, the photon will be absorbed with probability $1 - \alpha(\vec{r}_1)$, and the random walk is terminated. If not, it will choose a new direction θ_1 using the probability density function (PDF) $p_{\text{s}}(\vec{r}_1, \theta_0 \to \theta_1)$ in (7), hence using (8).
   - At a volume interaction, $\vec{r}_1 \in$ R, the photon will be absorbed/terminated with probability $1 - \varpi_0$. If not, it will scatter into a new direction θ_1 using the PDF $p_{\text{v}}(\theta_0 \to \theta_1)$ in (13) and (14), that is, using a Bernouilli trial followed by (15).
4) Steps 2) and 3) are repeated, with subscripts incremented by unity, until the photon is absorbed. The only difference is that intersection of the beam with the "$\cos^3$" terrain must now be computed numerically. Exit from the domain is formally accounted for (as an absorption event, or a detection) by making $\sigma = \infty$ and $\varpi_0 = 0$ outside of R.

We denote by $n^\star$ the subscript (order of scattering/reflection) at the time of escape or detection.

We note for future reference that this procedure defines a Markov chain and, in Appendix A, Section A-1, a formal probability space is defined for computing means, variances, etc. Also, in view of the "recycling" process described further on, all the positions $\{\vec{r}_0, \vec{r}_1, \vec{r}_2, \ldots, \vec{r}_{n^\star}, \vec{r}_{n^\star+1}\}$ of all the interactions have to be stored on file, but only if they end on the detector's surface (at position $\vec{r}_{n^\star+1}$).

The left-hand panel of Fig. 3 illustrates 75 trajectories based on the above algorithm conditioned to end at the small detector on the right-hand side of the medium. More precisely, this detector is defined in transport space by $x = L/2$ and $2.45 < z < 2.55$ along with $0 < \theta < \pi$. When these 75 trajectories, and many more, are tallied and partitioned into 15 equal-sized angular bins from 0 to π, the resulting radiative fluxes across the lateral boundary at the detector form an "image," in fact a 1-D angular scan in 2-D space, of the reference world. Re-centering the angles around $\theta = \pi/2$, the center of the detector's field of view, we denote this pixel data as $F_i (i = -7, \ldots, +7)$ where: $i = 0$ is looking horizontally, $i > 0$ is downward (into the variable terrain), and $i < 0$ is upward (toward the sky). Collectively, we will define the image as a formal 15-D vector

$$\mathbf{F}(\gamma) = (F_{-7}, \ldots, F_{+7}) \tag{22}$$

where, for the moment, $\gamma = (\cdot, \cdot, \cdot, 0; 0)$ since the plume's geometry $(x_{\text{p}}, z_{\text{p}}, \rho_{\text{p}})$ is irrelevant (arbitrary numbers) if $k_{\text{p}} = 0$ (no absorbers are present). This monochromatic image $\mathbf{F}(\gamma)$ is plotted in Fig. 3 (right panel, dashed line). In a typical simulation, 5 to 8×10^5 trajectories ending at the detector are used, starting with as much as 350 times more (roughly 2 to 3×10^8 histories). This loss factor is largely dependent on the finite physical size of the detector.

As expected, radiance from the hazy sky, to the left, is lower than that from the directly and diffusely illuminated reflective surface (with additional path radiance), to the right. Moreover, we recall from (9) that the higher altitudes are assigned a higher (vegetation-type NIR) albedo. Therefore, as the (re-centered) viewing angle in Fig. 3 (right panel) decreases from 0°, the far end of the high-altitude/high-albedo terrain is scanned, and correspondingly high radiance values are observed. Then, the lower terrain is scanned, resulting in lower radiance. The radiance then rebounds when the near side of the higher altitude/reflectivity terrain is reached. Finally, the radiance collapses to near-zero values because the nearest terrain is not even

illuminated; this artifact goes back to the choice of making the lateral non-terrain boundary "open" (absorbing BC applied) and to limit the illumination from $z = H_{\rm sky}$ to $|x| < 2.5$ (rather than $L/2 = \pi$). This keeps the ray-tracing simpler and does not affect the validity of the multipixel retrieval demo described further on.

Note that, although its observations are indexed with the angular coordinate θ in direction space, this is an imaging sensor, and we are in the relatively new realm of "multipixel" retrieval techniques. (See [35] for an early example of a crude two-pixel/monochromatic/mono-angular technique targeting broken clouds.) The angle-based definition of the pixels used here is entirely traceable to the static sensor's location at close range from the scene. From that vantage, only an angular scan (or a CCD-type focal plane array) can ensure imaging. A space-based counterpart of this simple sensor can collect an image in push-broom mode at a single angle using its orbital motion. Even an imaging focal-plane device at such a large stand-off distance captures light emerging from a small area with quasi-parallel beams (i.e., almost identical θ).

At any rate, the scene reconstruction described further on is inherently multipixel in kind, not multi-angle. Truly multi-angle observations would call for multiple sensor positions in the present MC simulation. In practice, it can be a single platform moving through space at orbital speeds with multiple push-broom or CCD sensors, e.g., aforementioned MISR or POLDER missions, respectively.

B. Discussion

The above algorithm is best described as a "brute force" MC method. It was coded in Python and could take up to two days to execute for 2 to 3×10^8 histories total. With the present detector geometry, only ≈1/350 of these rays contributed to any of the the 15 pixel-scale signals, hence were written to file for future recycling.

Apart from translation from a scripting language to a compiled one and more massive parallelization, the forward MC scheme could be accelerated at fixed accuracy in several ways. For instance, in the course of the MC random walk described above, each trajectory could carry, along with coordinates, a weight w that accounts for absorption processes, i.e., starting at $w_0 = 1$, it would be reduced by a factor of ϖ_0 at each scattering and by a factor of $\alpha(\vec{r}_1)$ at each reflection; concern over wasting time by working toward trajectories with negligible weight can be alleviated by using the "Russian roulette" procedure [36] when a judicious weight threshold is crossed. However, in view of the fact that we will need to recycle all the trajectories that reach the detector, as explained further on, we will need to store not just (x_i, z_i) but also w_i for $i = 0, 1, \ldots, n^\star$. More problematic is that there will be more detection events to revisit, and all will carry less than unit weight. Hence, it is not clear that the retrieval method described further on will benefit, only the forward calculation.

The same argument can be made about the "local estimation" technique [4], [36] where a contribution to the observed signal is sent from every scattering and reflection event, appropriately weighted by the phase function and the probability of direct transmission to the detector, which can then be reduced in size to a point. A better idea is to combine backward MC, where every simulated path ends at the point detector, with the local estimation technique, this time collecting a contribution from the source at every scattering and reflection along the way. However, because it is impractical to recycle too many low-weight (i.e., improbable) trajectories, the weight threshold for death by Russian roulette will have to be quite high, and the cost of Russian roulette in increased variance will have to be factored in.

In the practical applications that lie beyond the present demonstration, the optimal strategy for the forward computation will depend on how confident we are in the fixed elements of the scene. In this case, we would need to rerun the whole simulation whenever we change the structure of the terrain's shape or height. A change in the position of the source (i.e., time of day) or the detector (e.g., new collection) also mandates at present a new forward run. At that point, forward or backward local estimation techniques may become more attractive since they can be used quite efficiently to predict signals for multiple detectors or from multiple sources. However, a change in the surface reflectivity map (9) can be handled by path recycling, to be described next. The same remark applies to a change in the aerosol reference case or its phase function.

Operational circumstances can require a user-directed iteration on the structural properties of the scene (considered fixed in the rest of this study), yet we still need to use a MC framework to generate (possibly weighted) paths from source to detector for recycling. In that case, we should seek the most efficient MC methods, which are likely to be hybrid ones that use deterministic solutions to achieve radical variance reduction [37, and references therein].

C. Path Recycling

This is the MC acceleration technique that enables us to develop a retrieval algorithm to determine what γ-world the sensor is actually looking at. Thus, the right-hand panel of Fig. 3 also illustrates simulated observations for one instance of a γ-world (solid line), specifically, when $\gamma = (+0.35\ \ [\mathrm{km}], 2.5\ \ [\mathrm{km}], 0.5\ \ [\mathrm{km}], 0.5\ \ [\mathrm{km}^{-1}]; 0.15\ \ [\mathrm{km}^{-1}])$ where the key is given in (19). In particular, we note that the absorption optical diameter of the plume, $2\rho_{\rm p}k_{\rm p}$, is unity. The arrow heads on the lower horizontal axis indicate the position of the absorbing gaseous plume in the field of view. As expected, the radiance decreases significantly in the pixels that directly image the plume since it reduces the (surface) reflected and (volume) scattered light that streams into them. The radiance increases somewhat in the direction of the more distant terrain (13° and 27° bins) as well as the nearby terrain (54° bin). This is because the aerosol element δc in this γ is positive, thus reducing the AOD in (12) from unity (for $c = c_0 \approx 0.58\ [1/\mathrm{km}]$) to ≈0.83 (for $c = c_0 + \delta c = 0.65\ [1/\mathrm{km}]$). The ensuing increased transmittance of direct sunlight to the surface and of surface-leaving radiance to the sensor is therefore overcompensating the reduced path radiance (cf. the lower radiance in the "sky" pixels). This is largely because direct sunlight has to be scattered through angles $100^\circ \lesssim \theta_{\rm s} \lesssim 120^\circ$) to reach the detector at these angles, which is unfavorable (cf. Fig. 2).

This second image was computed by recycling the paths that were used to compute the previously discussed reference image (dashed line). Some 334 207 trajectories contributed to these simulated observations (out of a total of 172×10^6 casted rays), resulting in an average relative MC error of 0.44%.

Appendix A describes in full mathematical detail why and how MC path recycling works using changes of probability measure. In MC lingo, it is an application of the principles of importance sampling that underlie the "method of dependent trajectories" [36]. The main innovation in path recycling is that, in the traditional implementation of the method of dependent trajectories, the number of alternative worlds is necessarily finite and even quite small because the paths to the detector are not kept on file but processed on the fly. In contrast, the path-recycling method enables continuous variation of the changing parameter, i.e., an infinite number of alternate worlds to consider in the retrieval stage. It has been applied previously in medical imaging applications [38]–[40], but here it is enhanced with new capability (cf. Appendix A). In practice, each path $\vec{r}_i = (x_i, z_i)^{\mathrm{T}}$, $i = 0, 1, \ldots, n^\star + 1$ that ends at the detector is recalled from memory or disk and assigned a weight of unity. (Here, $n^\star$ is necessarily ≥ 1 since only diffuse light can be detected.) It is then reprocessed in two steps.

1) Non-Vanishing ρ_{p} and k_{p}: First, the weight of the path is changed if any of its $n^\star$ segments intersects the circle of radius ρ_{p} centered at $(x_{\mathrm{p}}, z_{\mathrm{p}})^{\mathrm{T}}$ that defines the gaseous plume with absorption coefficient k_{p}. The corresponding multiplicative factor is $\exp(-k_{\mathrm{p}}\ell_{\gamma,i})$ where $\ell_{\gamma,i} \geq 0$ is the length of the intersection of the ith segment with the circle defined in γ. There are highly efficient ways of computing the intersection circles and lines readily adapted to the problem at hand. In practice, the overall factor to use (interpreted in Appendix A, Section A-2, as a change in probability measure) is $\exp(-k_{\mathrm{p}}\ell_\gamma)$ where $\ell_\gamma = \sum_{i=0}^{n^\star} \ell_{\gamma,i}$.

2) δc Differs From Zero: Second, we need to compute the change in the pixel-level signals resulting from redistribution of the background aerosol particle density when δc in (16), the 5th and only non-plume element in γ, is $\neq 0$. It is not obvious that this is possible without casting a whole new set of rays. In Section A-3 of Appendix A, it is shown that this is indeed possible, and the corresponding change in weight of each detected photon's path is computed explicitly. The computation, however, is more involved than for the effect of the purely absorbing plume that was reduced in the above to straightforward computational geometry. To compute the new signals for a γ world with $\delta c \neq 0$ knowing the ones when $\delta c = 0$ (reference world), we need to look at how δc changes the probability of scattering in one position rather than another along each segment of the broken ray. That part of the photon transport is described mathematically by the cumulative distribution in (21) with the exponential stratification (10) in mind. If we do not want to change the value of c (namely, c_0) in order to leave all the scattering points where they are, that is alright as long as the path is properly re-weighted.

The basic question here is whether the new value of c in (16) makes the *given* scattering point, possibly a surface interaction, more or less probable than the what it is in the reference case. The probability level of the realized scattering point is given by $\mathrm{dP}/\mathrm{d}t$ in (21) using $c = c_0$ while the alternative $\mathrm{dP}_\gamma/\mathrm{d}t$ uses $c = c_0 + \delta c$. Assuming that it ends with a volume scattering, the weight correction factor we need is the ratio

$$\left|\frac{\mathrm{d}P_\gamma}{\mathrm{d}P}\right|_i = \frac{\left|\frac{\mathrm{d}P_\gamma}{\mathrm{d}t}\right|_i}{\left|\frac{\mathrm{d}P}{\mathrm{d}t}\right|_i} \tag{23}$$

at the ith step where

$$\left|\frac{\mathrm{d}P_\gamma}{\mathrm{d}t}\right|_i = \sigma(\vec{r}_i + t\vec{\Omega}_i)\exp\left(-\int_0^t \sigma(\vec{r}_i + s\vec{\Omega}_i)\mathrm{d}s\right) \tag{24}$$

with $\delta c \neq 0$ in the numerator, and the same with $\delta c = 0$ in the denominator, in expression (10) for $\sigma(\vec{r})$. For a step ending in a boundary (terrain or detector) interaction, we have only the above exponential term to worry about when computing the weight correction ratio. For the whole path, we need to evaluate $|\mathrm{dP}_\gamma/\mathrm{dP}| = \prod_0^{n^\star} |\mathrm{dP}_\gamma/\mathrm{dP}|_i$. See Appendix A-3 for the computational procedure using only data from the existing path.

D. Verification and Performance of the Forward Model

The accuracy of our forward RT model for the parameterized scene was rigorously verified. In particular, we conducted a variety of tests where

1) A standard MC simulation of transport in an atmosphere parameterized by some value of γ was run until the relative mean square error was less than 0.3%. The mean fluxes are stored as $\mathbf{M}(\gamma) = \{M_i(\gamma), i = -7, \ldots, +7\}$.
2) Another MC simulation was run in an atmosphere parameterized by $\gamma_0 = (0, 1, 0, 0; 0)$ (no plume and baseline aerosol atmosphere). The paths to the detector are stored.
3) The paths from (ii) are used in the path-recycling forward model to compute $\mathbf{F}(\gamma)$.

The forward model is deemed verified if:

- $|F_i(\gamma) - M_i(\gamma)|/M_i(\gamma) \leq 0.01, \forall i = -7, \ldots, +7$;
- as γ' becomes sufficiently different from γ, the relative error $|F_i(\gamma') - M_i(\gamma)|/M_i(\gamma), \forall i = -7, \ldots, +7$, becomes much worse than 1%.

The forward model passed this test for a wide variety of γ values.

The performance increase is dramatic. For example, it took 11 727 equivalent one-core minutes (8+ days) for the Python code to generate approximately 231 million paths on a four-core 2.6 GHz workstation. Of these paths, ≈1/348 hit the (0.1 [km] wide) detector. These paths can, however, be recycled in only 30.9 equivalent one-core seconds (22 770 times faster), while one might have expected a speed-up of only about 348 times if one had a Maxwell's daemon that could discriminate at the source rays that would hit the sensor from those that would not.

The dramatic difference is due mostly to the fact that the original paths were cast using complicated code that explicitly stepped the photons through their path, while the much simpler recycling was done using optimized code. In any case, recycling paths only involves computing a ratio of weights and in many cases is generally much quicker than sending the original paths.

It is of course worthwhile to revisit the question of computational efficiency if a backward MC scheme was used where all the trajectories generated start by definition at the detector.

V. Bayesian Multipixel 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 multivariate 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 paper, we have $n = 5$ and denote $\mathbf{x}$ as γ, with specific definitions in (19) and bounds in (20) as well as in Table II.

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 multivariate function of $\mathbf{x}$. In the present study, we have $m = 15$, that is, one observation per pixel in the 1-D image of 2-D 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

$$\mathbf{d} = \mathbf{F}_\infty(\mathbf{x}) + \mathbf{E}, \qquad \mathbf{E} \sim \mathcal{N}(0, \Sigma_{\mathbf{E}}), \qquad \mathbf{E} \perp\!\!\!\perp \mathbf{x}. \quad (25)$$

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 strictly 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 models used, on the one hand, for generating the simulated data and, on the other hand, for the iterations in the inverse problem solution are closely related.

In its most general form, Bayes' rule [41, 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

$$\pi_\infty(\mathbf{x}|\mathbf{d}) \propto p_{\text{prior}}(\mathbf{x}) \frac{1}{|\Sigma_{\mathbf{E}}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \|\mathbf{d} - \mathbf{F}_\infty(\mathbf{x})\|^2_{\Sigma_{\mathbf{E}}^{-1}}\right) \quad (26)$$

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 $\mathbf{v}^{\mathrm{T}}\mathsf{M}\mathbf{v}$. Symbols were defined in the above text. The "∞" 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 (26) is the familiar cost function that is minimized in standard retrieval techniques using, e.g., the popular Levenberg–Marquard algorithm. 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 (MC-based forward 2-D RT modeling), we do not have access to $\mathbf{F}_\infty$, but instead to a sequence of approximate models $\mathbf{F}_1, \mathbf{F}_2, \ldots$ 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

$$\mathbf{F}_j(\mathbf{x}) \sim \mathcal{N}\left(\mathbf{F}_\infty(\mathbf{x}), \Sigma_j(\mathbf{x})\right) \quad (27)$$

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 MC techniques. See App. B of Part I for a formal description of our hierarchical modeling framework.

Equation (27) leads to an *enhanced* noise model at precision level j

$$\mathbf{d} = \mathbf{F}_j(\mathbf{x}) + \mathbf{E}_j(\mathbf{x}) + \mathbf{E}, \quad \mathbf{E} \sim \mathcal{N}(0, \Sigma_{\mathbf{E}}), \quad \mathbf{E} \perp\!\!\!\perp \mathbf{x}, \quad \mathbf{E}_j(\mathbf{x}) \sim \mathcal{N}\left(0, \Sigma_j(\mathbf{x})\right), \quad \mathbf{E}_j(\mathbf{x}) \perp\!\!\!\perp \mathbf{E} \quad (28)$$

and a likelihood at precision level j,

$$\pi_j(\mathbf{d}|\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^m \left|\Sigma_{\mathbf{E}} + \Sigma_j(\mathbf{x})\right|}} \times \exp\left(-\frac{1}{2}\left\|\mathbf{d} - \mathbf{F}_j(\mathbf{x})\right\|^2_{(\Sigma_{\mathbf{E}}+\Sigma_j(\mathbf{x}))^{-1}}\right). \quad (29)$$

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

$$\pi_j(\mathbf{x}|\mathbf{d}) \propto p_{\text{prior}}(\mathbf{x})\pi_j(\mathbf{d}|\mathbf{x}) \quad (30)$$

and we recall that an MC prediction for $\mathbf{d}$ at lesser precision is cheaper to compute by a known factor (cf. Appendix B).

B. Application to 2-D 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 multivariate histogram for the posterior distribution of possible parameter values as efficiently and accurately as possible.

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 (30). At the core of all three algorithms is the concept of a Markov chain MC (MCMC) process [43] where, like in the RT application, a MC (a.k.a. "random quadrature" rule) is used to compute integrals. MCMC is essentially a random walk in the 5-D 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 (30), 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 also implemented with a MC 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) 1-D, 2-D, or 3-D 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. MC is, however, a natural choice in multidimensional RT, the whole multipixel inversion scheme we propose is therefore called the "MC^3 method."

The three MC^3 algorithms we tested are described in detail and intercompared in [29], a companion paper to this one. The baseline algorithm we tested (#1) was the well-known Metropolis–Hastings [44]–[46] technique applied to a single level of forward model precision. The two other algorithms build on that classic methodology. Algorithm #2, previously used in [47], [48], couples two model precision levels, while #3, the new development in computational MCMC technology, couples multiple levels. Fig. 4 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 h on the same platform. Progress is striking since the efficiency of an MCMC sampling technique can be measured by how soon all the bins (along the vertical axis in Fig. 4) are sufficiently well populated. The new (multilevel) algorithm proves to be the most efficient; it was therefore retained for the remainder of the study.

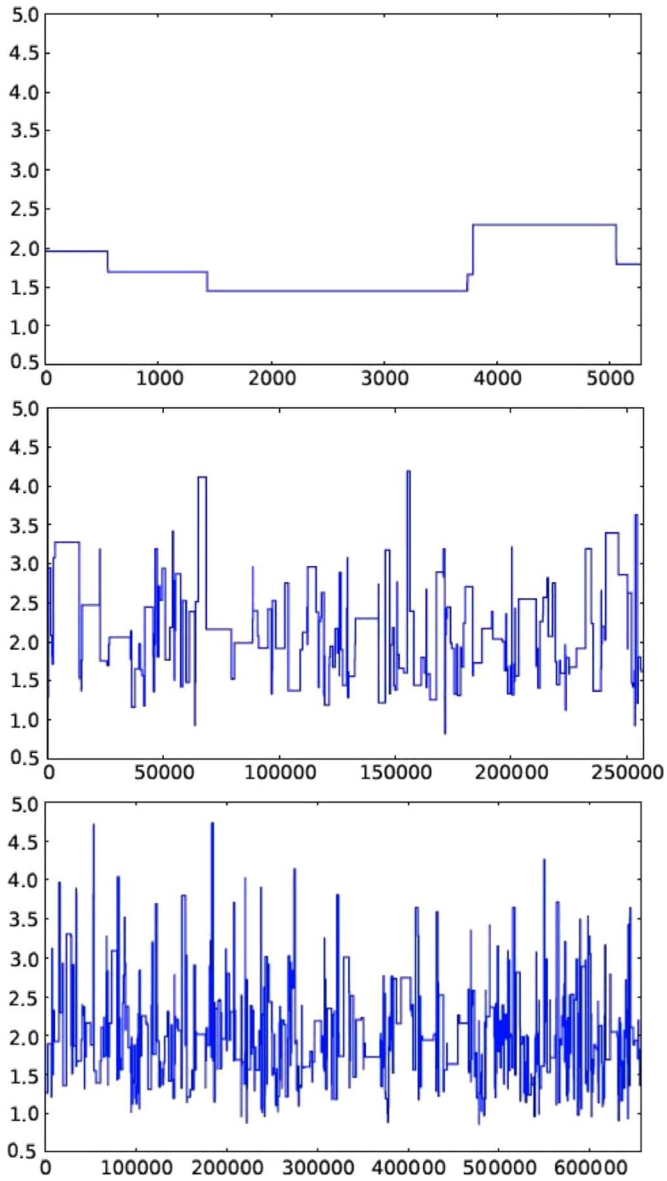

Fig. 4. Γ-space sampling with three MCMC algorithms. Only the projection of the random walk along the k_{p} axis is displayed. From top to bottom: #1) traditional Metropolis–Hastings (M–H) [44]–[46]; #2) two-level M–H [47], [48]; #3) multilevel M–H with prescribed confidence [29]. In all cases, the code was run for the equivalent of 5 h on a single 2.6 GHz Intel core.

In a multilevel MCMC approach, the speedup of the required sampling of Γ-space results from the option to accept or reject a proposal based on a less precise, but cheaper to compute, prediction of the observations used in (29). It is critical to know upfront the "cost" in forward model precision for "buying" time. Fortunately, in the case of a MC transport model, as used in MC^3, it is easy to control precision (see Appendix B for the case of path recycling).

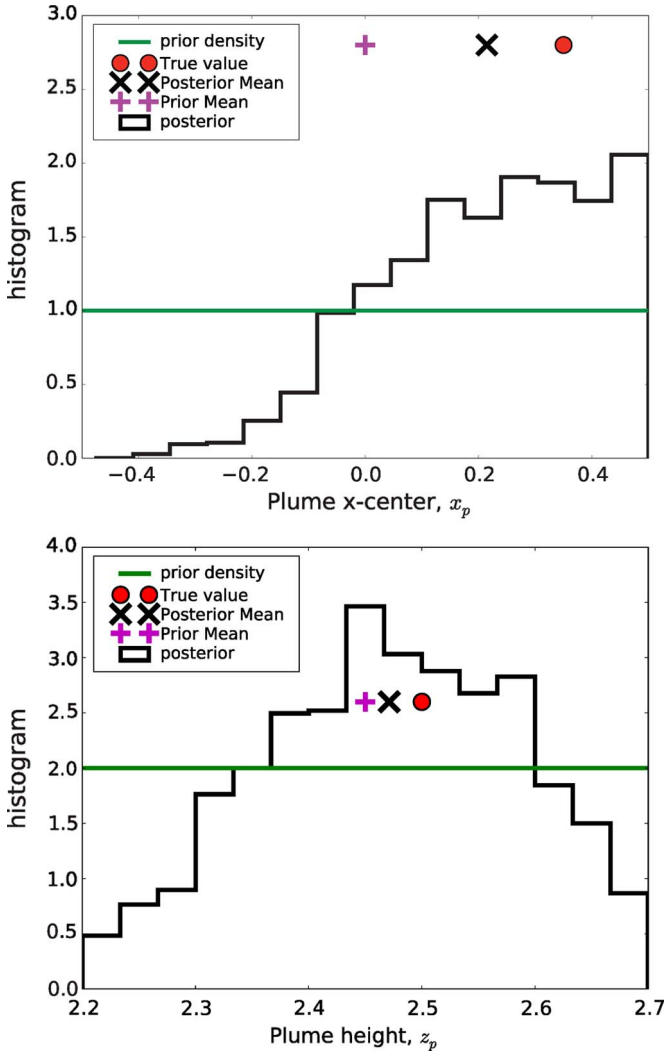

Fig. 5. Marginal posterior distributions for plume position parameters at SNR $= 5$. (Top) $x_{\rm p}$ [km]. (Bottom) $z_{\rm p}$ [km].

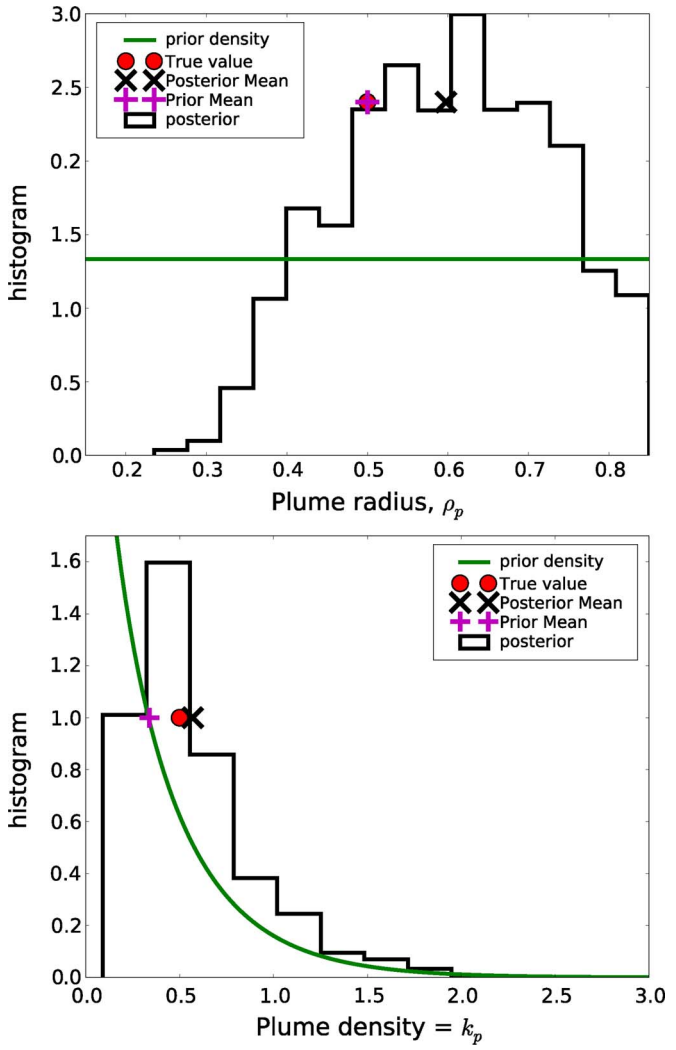

Fig. 6. Marginal posterior distributions for inherent plume properties at SNR $= 5$. (Top) $\rho_{\rm p}$ [km]. (Bottom) $k_{\rm p}$ [1/km].

VI. Numerical Results

We avoid an "inverse crime" by generating data using a slightly different forward model than is used in the inversion. Specifically, the integrals appearing in the transport solution (e.g., (21)) are computed using an explicit Euler scheme with a random step size distributed uniformly between 0.05 [km] and 0.15 [km]. Different seeds were used to generate the paths for recycling and to obtain the data. This introduces a discretization error that mimics the unavoidable model error due to the fact that real-world atmospheres and plumes 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 (25).

We take SNR $= 5$ ($\Sigma_{\mathbf{E}} = \mathrm{diag}[\mathbf{d}]/5$). This seems low but is not unrealistic if we recall that the spectral bandpass of a dedicated gas-plume sensor should be quite narrow in order to boost the absorptivity contrast of the molecule of interest. The four panels in Figs. 5 and 6 show the marginal posteriors for the plume-related elements of γ, namely, $(x_{\rm p}, z_{\rm p}, \rho_{\rm p}, k_{\rm p})$. We see that all but $x_{\rm p}$ are reasonably well-characterized probabilistically. In particular, the estimated means for parameters $z_{\rm p}$, $\rho_{\rm p}$, and particularly $k_{\rm p}$ 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; to see this, review scene/detector and prior bound geometries 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. This second image could of course be captured by the same sensor mounted on a mobile platform. Moreover, there is no fundamental difficulty in generalizing the Bayesian multipixel methodology used here to a multipixel/multiview one that would get us closer to a bone fide atmospheric tomography of the scene.

Fig. 7 shows the stops of the MCMC random walk in 2-D hyperplanes of Γ-space; these are the intermediate data products that enable the estimation of the posterior distributions. The two panels of Fig. 7 also show more clearly why $x_{\rm p}$

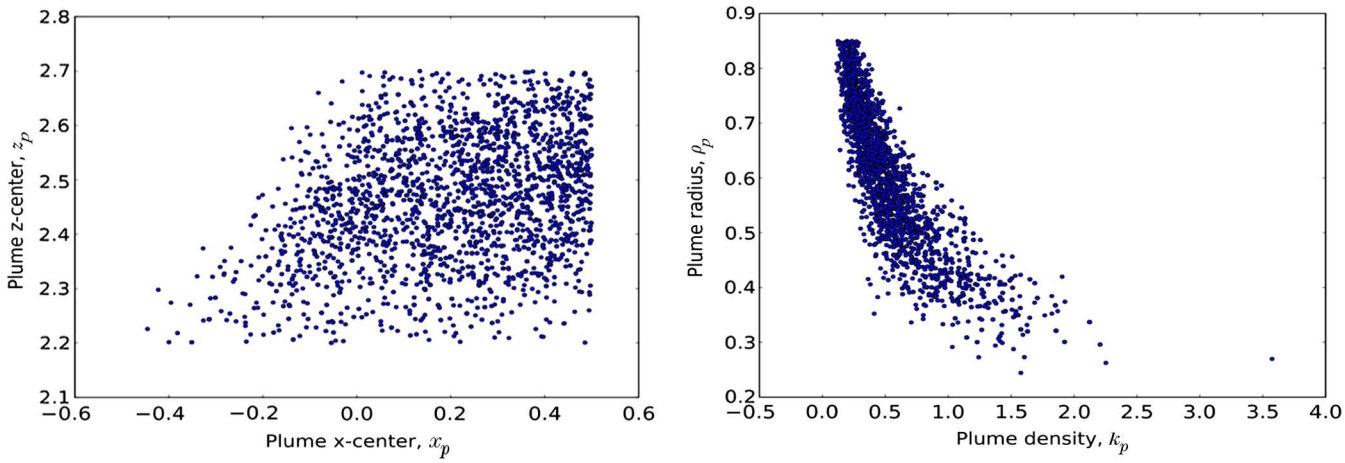

Fig. 7. MCMC samples for joint marginal posterior distributions for plume-specific parameters at SNR $= 5$. (Left) $(x_\mathrm{p}, z_\mathrm{p})$ drawn to scale given the horizontal range in the upper panel of Fig. 5 (cf. plume center bounding box in the schematic in Fig. 1). (Right) $(k_\mathrm{p}, \rho_\mathrm{p})$ drawn to scale given the horizontal ranges in Fig. 6. More discussion in main text.

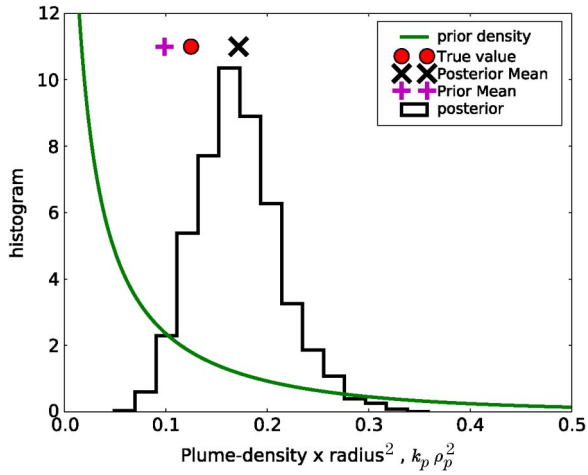

Fig. 8. Marginal posterior distribution for "plume mass" composite $k_\mathrm{p}\rho_\mathrm{p}^2$ at $SNR = 5$.

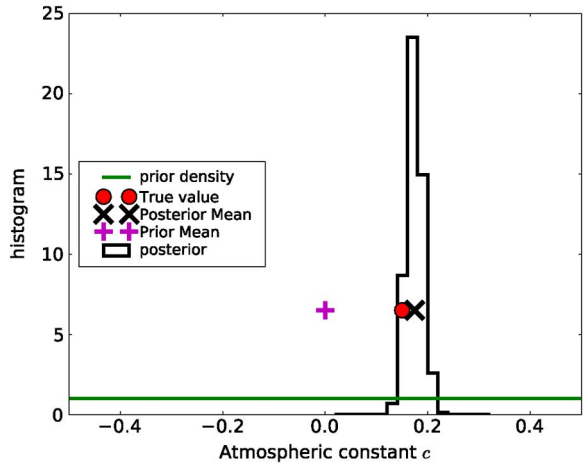

Fig. 9. Marginal posterior distribution for δc at SNR $= 5$.

localization is poor compared to z_p (left-hand panel) and why k_p characterization is better than for ρ_p (right-hand panel). This last $(k_\mathrm{p}, \rho_\mathrm{p})$ scatter plot shows a relatively tight anticorrelation 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_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 $\propto \rho_\mathrm{p}$. Fortunately, k_p is better quantified (narrower posterior) than ρ_p since it is the key parameter for identifying the gas if a continuous spectrum was 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 $k_\mathrm{p}\rho_\mathrm{p}^2$. 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. 8 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 somewhat closer to the truth than is its mean.

Finally, the outcome for the background aerosol δc is shown in Fig. 9. 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.

The authors experimented with other SNR values, particularly smaller ones [29]. As expected, the inferred PDF for the scene parameters are broader, closer to the prior distributions.

VII. Summary and Outlook

We implemented and adapted an innovative computational approach to multidimensional RT developed recently in biomedical imaging, namely, path-recycling MC. 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 at 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 known cross section per molecule but unknown location, size, and density. 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 RT unfolds completely in these two spatial dimensions. The background aerosol's SSA 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 parameter. We showed that this 2-D atmospheric structure, because it is represented parametrically, can be reconstructed reasonably well, even at low spatial resolution and with a modest (single-digit) SNR. The demonstration used a single radiometrically calibrated image from a single sensor at close range.

To underscore the novel, inherently *multipixel* nature of the methodology, only a single spectral channel is considered, presumably narrowed down optimal wavelength for detecting the absorbing gas of interest (hence the interest in small SNRs). Where the five-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/multipixel approach would be in essence a coarse atmospheric tomography using a parameterized representation of the spatial structure instead of imposing the usual regularization constraints for ill-posed problems. This keeps the actual number of structural unknowns small and manageable.

Multipixel retrieval algorithms introduced here require a forward RT 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 3-D RT model. (Although, for this particular demo, a 2-D RT model was used.) This is not to be confused with the terminology recently introduced by Dubovik *et al.* [49] where "multipixel" is used to describe a statistical constraint in the cost function used for an aerosol property retrieval predicated entirely on a polarized 1-D RT model. Similar constraints have been used previously, for instance, in the intensity-only operational aerosol retrieval developed for the MISR [50].

Finally, a Bayesian formulation of the remote sensing inverse problem was used. Consequently, the outcome is not a single value for each of the five parameters, even including uncertainty estimates, as would result from a standard cost function minimization approach. The derived product is in fact a whole multivariate probability distribution for the parameters that is consistent with the data. The Bayesian inference machine is simple to implement using "MCMC" 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.

Although they are favored by many in computational 3-D RT, MC RT techniques are notoriously slow to converge in any number of spatial dimensions. Normally, this would make MC an impractical way of solving inverse problems in remote sensing, or almost any other application. However, path recycling reduces the execution time for the limited forward 2-D RT used here to just a few seconds. Parameterization is required, but it only needs to be for the remote sensing target, not every aspect of the scene's make-up. There is no obvious reason why this procedure could not be implemented in sophisticated 3-D scene simulation frameworks that use ray tracing, such as the Intercomparison of 3-D Radiation Codes community MC model [51] or DIGital Remote Sensing Image Generator [52]. That additional capability would enable them to be used in inverse problems of interest to their respective sponsors.

APPENDIX A
COMPUTATION OF THE CHANGE OF MEASURE

The adopted path-recycling MC scheme follows three steps: 1) paths are generated in a reference atmosphere, and those that hit the detector are saved; 2) the (increase/decrease of) probability of these paths reaching the detector in the modified atmosphere is calculated, resulting in a set of weights attached to each path; 3) the weights are added resulting in an unbiased estimate of the detected flux in the modified atmosphere. We discuss here the underlying theory and practical implementation of the path-recycling forward MC model.

1) Path Measures: The algorithm described in Section IV-A induces a measure on the space of finite-length paths, namely

$$\Omega := \{\omega = (\vec{r}_0, \ldots, \vec{r}_{n^\star+1}) : \vec{r}_j \in R \cup \partial R\}$$

where we recall that $n^\star$ is the last order of scattering, including surface reflections, and we take $\vec{r}_{n^\star+1} \in \partial$R as the final point. Note that, under reasonable conditions, the (discrete) *stopping time* $n^\star + 1$ of the above Markov chain is $< \infty$. We therefore have a probability measure P_γ. In the special case where $\gamma = \gamma_0 := (0, 1, 0, 0; 0)$, corresponding to no plume and nominal background aerosol, we have our *reference measure* P.

This allows us to define a differential measure dP and expectation $\mathbb{E}_\mathrm{P}\{\cdot\}$ by

$$P[A] := \mathbb{E}_\mathrm{P}\{\mathbf{1}_A\} = \int_\Omega \mathbf{1}_A(\omega)\mathrm{d}P(\omega) = \int_A \mathrm{d}P(\omega) \qquad \text{(A1)}$$

where for $A \subset \Omega$, the *indicator function*

$$\mathbf{1}_A(\omega) := \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

and similarly for dP_γ.

As a highly relevant example of a subset of paths, consider those that hit (and are necessarily absorbed by) the detector; see left panel of Fig. 3 for a few samples. Denote these by a disjoint union

$$\mathcal{D} := D_1 \cup \ldots \cup D_m$$

meaning that if $\omega \in \mathcal{D}$ then the path ω ended up in the detector, and if $\omega \in D_\nu$ then ω hit the detector with incoming angle θ in the interval $[(\nu-1)\pi/m, \nu\pi/m)$. Let

$$\mathbf{D} := D_1 \times \ldots \times D_m, \qquad \mathbf{1}_{\mathbf{D}} := (\mathbf{1}_{D_1}, \ldots, \mathbf{1}_{D_m})$$

and thus our measurement is

$$P_\gamma[\mathbf{D}] = \mathbb{E}_\gamma\{\mathbf{1}_{\mathbf{D}}\} := (P_\gamma[D_1], \ldots, P_\gamma[D_m])\,.$$

One can similarly define $\mathbb{E}_{\mathrm{P}}\{\mathbf{1}_{\mathbf{D}}\} = \mathrm{P}[\mathbf{D}]$ in the absence of an absorbing plume. In the main text, we set $m = 15$ and denoted the flux through the νth angular bin (or "pixel") as F_i where $i = 8 - \nu$ (so that $i = 0$ is assigned to a horizontal look to the left in Fig. 1).

2) Restoration of Plume Absorption k_{p}: Here, we recover only the effect of k_{p} and, by extension, those of plume geometry contained in the parameter trio $(x_{\mathrm{p}}, z_{\mathrm{p}}, \rho_{\mathrm{p}})$. We assume all other quantities are known and, in particular, that $\delta c = 0$.

Based on (21), the cumulative probability of absorption at the detector in the reference measure P is given by

$$E_\sigma(\vec{r}_0, \ldots, \vec{r}_{n^\star+1}) := \exp\left[-\int_0^{T(\omega)} \sigma\left(\vec{R}(t,\omega)\right) \mathrm{d}t\right] \quad \text{(A2)}$$

a random variable where $\vec{R}(t,\omega) = (x(t,\omega), z(t,\omega))^{\mathrm{T}}$ is the position of chain ω at time $t \in [0, T(\omega)]$ (in units where the velocity of light is unity), with $T(\omega)$ being the instant of detection. For P_γ, it is therefore given by $E_\sigma(\vec{r}_0, \ldots, \vec{r}_{n^\star+1})\mathrm{e}^{-\alpha\ell_\gamma(\omega)}$, where the random variable $\ell_\gamma(\omega)$ is the total length of intersection of the path ω with the plume parameterized by (the first four elements of) γ.

The intersection of an infinite line through a disk can be computed very efficiently via known techniques. From there, it is a simple extension to compute the intersection of a line segment $(\vec{r}_j, \vec{r}_{j+1})$ with a disk, and hence $\ell_{\gamma,j}(\omega)$ is obtained.

Specifically, one first computes

$$\Delta_j = \rho_{\mathrm{p}}^2 - \left((\vec{r}_{\mathrm{p}} - \vec{r}_j) \times \vec{\Omega}_j\right)^2$$

where we recall that $\vec{\Omega}_j = (\vec{r}_{j+1} - \vec{r}_j)/\|\vec{r}_{j+1} - \vec{r}_j\|$. If $\Delta_j \leq 0$, the intersection is empty. Otherwise, define $s_{j\pm} = (\vec{r}_{\mathrm{p}} - \vec{r}_j) \cdot \vec{\Omega}_j \pm \sqrt{\Delta_j}$ and compute s_{j+}. If $s_{j+} \leq 0$, the intersection is empty (the disk is "upwind" from $\vec{r}_j$). Otherwise, compute $\rho_j^2 = (\vec{r}_{\mathrm{p}} - \vec{r}_j)^2$ and the same for $j+1$. There are than just three possibilities to consider:

- if ρ_j^2 and ρ_{j+1}^2 are both $> \rho_{\mathrm{p}}^2$, then $\ell_{\gamma,j}(\omega) = 2\sqrt{\Delta_j}$;
- or else, if $\rho_j^2 < \rho_{\mathrm{p}}^2$ and $\rho_{j+1}^2 > \rho_{\mathrm{p}}^2$, then $\ell_{\gamma,j}(\omega) = s_{j+}$;
- or else, if $\rho_j^2 > \rho_{\mathrm{p}}^2$ and $\rho_{j+1}^2 < \rho_{\mathrm{p}}^2$, then compute $s_{j-} = s_{j+} - 2\sqrt{\Delta_j}$ and set $\ell_{\gamma,j}(\omega) = \|\vec{r}_{j+1} - \vec{r}_j\| - s_{j-}$.

In summary, we have

$$\left|\frac{\mathrm{d}P_\gamma}{\mathrm{d}P}\right|(\omega) = \mathrm{e}^{-k_{\mathrm{p}}\ell_\gamma(\omega)}, \quad \text{where } \ell_\gamma(\omega) = \sum_{j=0}^{n^\star} \ell_{\gamma,j}(\omega). \quad \text{(A3)}$$

3) Restoration of γ, Including Background Aerosol Perturbation δc: Here, we recover the background given by (10) and (11) with (16). Notice that the background absorption/scattering depends only on the height z.

We begin by defining $z_1(\omega), \ldots, z_{n^\star}(\omega)$, the random scattering/reflection heights. Our goal is to compute

$$\mathbb{E}_\gamma\{\mathbf{1}_{\mathbf{D}}\} = \int_\Omega \mathbf{1}_{\mathbf{D}}(\omega)\mathrm{d}P_\gamma(\omega) = (P_\gamma[D_1], \ldots, P_\gamma[D_m])\,.$$

We start with the case with no absorbing plume where γ is reduced to $(0, 1, 0, 0; \delta c)$. We will have to differentiate the scattering heights that occur in the volume from those that happen in the volume (due to aerosols) *or* at the lower boundary (due to surface reflection). Let the indices of the former be denoted by $j_1, \ldots, j_{n_{\mathrm{s}}}$ where n_{s} is total volume interactions (i.e., bona fide scatterings by aerosol particles).

Here, again, we need to compute the Radon–Nikodym derivative [53] $|\mathrm{dP}_\gamma/\mathrm{dP}|(\omega)$. Note that P_γ differs from P in two ways. First, the integrated extinction coefficient (total cross section per unit of volume) is E_{σ^γ} rather than E_σ. Second, the scattering coefficient and phase function in the non-baseline volume are multiplied by a factor $\exp(-\delta c z)$. Therefore, using (10) and (11) and (A2), but leaving the dependence on ω implicit, we have

$$\begin{aligned}
\left|\frac{\mathrm{d}P_\gamma}{\mathrm{d}P}\right| &= \frac{E_{\sigma^\gamma}(\vec{r}_0, \ldots, \vec{r}_{n^\star+1})}{E_\sigma(\vec{r}_0, \ldots, \vec{r}_{n^\star+1})} \times \frac{\sigma_{\mathrm{s}}^\gamma(\vec{r}_{j_1}) \ldots \sigma_{\mathrm{s}}^\gamma(\vec{r}_{j_{n_{\mathrm{s}}}})}{\sigma_{\mathrm{s}}(\vec{r}_{j_1}) \ldots \sigma_{\mathrm{s}}(\vec{r}_{j_{n_{\mathrm{s}}}})} \\
&= \frac{E_{\sigma^\gamma}(\vec{r}_0, \ldots, \vec{r}_{n^\star+1})}{E_\sigma(\vec{r}_0, \ldots, \vec{r}_{n^\star+1})} \mathrm{e}^{-\delta c\left(z_{j_1} + \ldots + z_{j_{n_{\mathrm{s}}}}\right)} \\
&= E_{\sigma^\gamma - \sigma}(\vec{r}_0, \ldots, \vec{r}_{n^\star+1}) \mathrm{e}^{-\delta c\left(z_{j_1} + \ldots + z_{j_{n_{\mathrm{s}}}}\right)} \\
&= \exp\left(-\int_0^T \left[\sigma^\gamma\left(\vec{R}(s)\right) - \sigma\left(\vec{R}(s)\right)\right] \mathrm{d}s\right) \\
&\quad \times \mathrm{e}^{-\delta c\left(z_{j_1} + \ldots + z_{j_{n_{\mathrm{s}}}}\right)} \\
&= \exp\left(-\int_0^T \sigma\left(\vec{R}(s)\right)\left[e^{-\delta c z(s)} - 1\right] \mathrm{d}s\right) \\
&\quad \times \mathrm{e}^{-\delta c\left(z_{j_1} + \ldots + z_{j_{n_{\mathrm{s}}}}\right)}. \qquad \text{(A4)}
\end{aligned}$$

The second term, $\exp\{-\delta c(z_{j_1} + \ldots + z_{j_{n_{\mathrm{s}}}})\}$, depends only on the z-coordinate of an identified subset of the scattering points $(\vec{r}_1, \ldots, \vec{r}_{n^\star})$. Therefore, like for the plume-related term in (A3), it can be computed *without casting new rays*: we only need to store the scattering points. We now show that the first term enjoys that feature as well.

Define the scattering/reflection times $T_1, \ldots, T_{n^\star}$, such that $\vec{R}(T_j) := \vec{r}_j$, along with $T_0 = 0$ and $T_{n^\star} = T$. Note that

$$T_j = |\vec{r}_1 - \vec{r}_0| + \ldots + |\vec{r}_j - \vec{r}_{j-1}|.$$

When $T_j < s < T_{j+1}$ the MC particle is traveling in a straight line given by

$$x(s) = x_j \pm (s - T_j)\sqrt{1 - \mu_j^2}, \quad z(s) = z_j + (s - T_j)\mu_j$$

where μ_j is the vertical direction cosine of θ_j, the angle between the direction of travel and the upwards unit vector

(0,1); the positive sign is taken if the photon is traveling to the right. We have

$$\begin{aligned}
&\exp\left(-\int_0^T \sigma\left(\vec{R}(s)\right)\left[\mathrm{e}^{-\delta c z(s)}-1\right]\mathrm{d}s\right)\\
&=\exp\left(-\sigma_0\int_0^T e^{-c_0 z(s)}\left[\mathrm{e}^{-\delta c z(s)}-1\right]\mathrm{d}s\right)\\
&=\exp\left(-\sigma_0\sum_{i=0}^{n^\star}\int_{T_j}^{T_{j+1}}\left[\mathrm{e}^{-(c_0+\delta c)z(s)}-\mathrm{e}^{-c_0 z(s)}\right]\mathrm{d}s\right). \quad \text{(A5)}
\end{aligned}$$

To evaluate this, note that (with $h = c_0$ or $h = c_0 + c$, and assuming $h > 0$)

$$\begin{aligned}
\int_{T_j}^{T_{j+1}} \mathrm{e}^{-hz(s)}\mathrm{d}s &= e^{-h(z_j - T_j\mu_j)}\int_{T_j}^{T_{j+1}} \mathrm{e}^{-hs\mu_j}\mathrm{d}s\\
&= \frac{\mathrm{e}^{-hz_j}}{h\mu_j}\left[1-e^{-h(T_{j+1}-T_j)\mu_j}\right]\\
&= \frac{\mathrm{e}^{-hz_j}}{h\mu_j}\left[1-e^{-h|\vec{R}_{j+1}-\vec{R}_j|\mu_j}\right]. \quad \text{(A6)}
\end{aligned}$$

Each term in (A5) is evaluated using (A6). Since this requires only knowledge of the points $\vec{R}_j$, there is no need to recast the rays. Finally, (A4) and (A5), once evaluated using (A6), are combined to yield $|\mathrm{dP}_\gamma/\mathrm{dP}|(\omega)$.

To restore the full effect of γ on $|\mathrm{dP}_\gamma/\mathrm{dP}|(\omega)$, the term $\mathrm{e}^{-k_\mathrm{P}\ell_\gamma(\omega)}$ computed in Appedix A-2 (requiring knowledge only of the scattering and reflection points) is factored in to update $|\mathrm{dP}_\gamma/\mathrm{dP}|(\omega)$. In the path-recycling MC code, this Radon-Nikodym derivative is used as the new weight assigned to the random photon path (the RT-related Markov chain) while retallying the detector responses.

Appendix B
Monte Carlo Models With Varying Precision

Adopting the notations and definitions introduced in Appendix A-1, we are in a position to describe our forward MC RT model and compute from first principles probabilistic estimates of means and variances, hence errors on the mean, and (pixel-to-pixel) covariances. We summarize in this appendix the main results of Bal *et al.* [29] that are key to the Bayesian approach to the inverse problem.

The model uses importance sampling to compute $\mathrm{P}_\gamma[\mathbf{D}]$ from one fixed set of reference paths. This technique is an advancement over "perturbation MC" schemes developed in the context of medical imaging; see, e.g., [38]–[40].

Choosing $N \in \mathbb{N}$, we generate N paths $\{\omega_1, \ldots, \omega_N\}$. Now, for any random variable X,

$$\frac{1}{N}\sum_{j=1}^{N}\mathbf{1}_\mathbf{D}(\omega_j)X(\omega_j) \xrightarrow{a.s.} \mathbb{E}_\mathrm{P}\{\mathbf{1}_\mathbf{D}X\}, \quad \text{as } N\to\infty$$

where "*a.s.*" stands for "almost surely." For example, we could generate paths from measure P_γ, and then $N^{-1}\sum_{j=1}^{N}\mathbf{1}_\mathbf{D}(\omega_j) \xrightarrow{a.s.} \mathbb{E}_\gamma\{\mathbf{1}_\mathbf{D}\}$.

It is important to realize that since we only intend to estimate expectations involving detector hits (e.g., $\mathbb{E}_\gamma\{\mathbf{1}_\mathbf{D}X\}$), we only need to store paths that hit the detector. The expected number of detector hits is exactly $N\mathrm{P}[\mathcal{D}] \ll N$.

For every new γ, we could generate a new set of paths and repeat the above procedure. This would be costly since path generation involves complicated steps. Instead, consider fixing one set of reference paths $\{\omega_j, j = 1, \ldots, N\}$ (in practice storing only those that hit the detector) generated by the reference measure P and then set

$$\begin{aligned}
\mathbf{f}_N(\gamma) &:= \frac{1}{N}\sum_{k=1}^{N}\mathbf{1}_\mathbf{D}(\omega_k)\left|\frac{\mathrm{d}P_\gamma}{\mathrm{d}P}\right|(\omega_k)\\
&\approx \int_\Omega \mathbf{1}_\mathbf{D}\left|\frac{\mathrm{d}P_\gamma}{\mathrm{d}P}\right|\mathrm{d}P = \int_\Omega \mathbf{1}_\mathbf{D}\,\mathrm{d}P_\gamma = \mathbb{E}_\gamma\{\mathbf{1}_\mathbf{D}\}. \quad \text{(B1)}
\end{aligned}$$

Computation of $\mathbf{f}_N$ requires computing the Radon–Nikodym derivative for the $\approx N\mathrm{P}[\mathcal{D}]$ paths that hit the detector, as described in Appendix A. This is significantly faster than generating N new paths.

Although already fast, $\mathbf{f}_N$ can be significantly improved by using (for relatively small N) information from a simulation that used a very large N. This is where we depart from the aforementioned "perturbation MC" schemes.

We first generate N_max paths using the reference measure P. Denote by H^ν_max the collection of paths $\omega_k \in D_\nu$, i.e.,

$$H^\nu_\mathrm{max} := \{\omega_1, \ldots, \omega_{N_\mathrm{max}}\} \cap D_\nu.$$

For $\nu = 1, \ldots, m$, the number of observations (in our case, pixels), let

$$H^\nu_1 \subset H^\nu_2 \subset \ldots \subset H^\nu_\mathrm{max}$$

be nested subsets of H^ν_max of (fixed, predetermined) size $N_j = |H^\nu_j|$. Note that H^ν_j and H^ν_max consist of i.i.d. draws from $\mathrm{P}[\cdot|D_\nu]$. Since $|H^\nu_\mathrm{max}| = N^{-1}_\mathrm{max}\sum_{k=1}^{N_\mathrm{max}}\mathbf{1}_{D_\nu}(\omega_k)$, we have

$$\mathrm{Cov}_\mathrm{P}\left(|H^\nu_\mathrm{max}|, |H^\mu_\mathrm{max}|\right) = \frac{1}{N_\mathrm{max}}\begin{cases} P[D_\nu]-P[D_\nu]^2, & \nu=\mu\\ -P[D_\nu]P[D_\mu], & \nu\neq\mu\end{cases}$$

where the above $\mathrm{Cov}_\mathrm{P}(X, Y)$ is defined as

$$\mathrm{Cov}_\mathrm{P}(X,Y) := \mathbb{E}_\mathrm{P}\left\{(X-\mathbb{E}_\mathrm{P}\{X\})\,(Y-\mathbb{E}_\mathrm{P}\{Y\})^\mathrm{T}\right\}$$

with subscript "P" making it clear that expectations are with respect to the probability measure P.

Proof:

$$|H^\nu_\mathrm{max}| = \frac{1}{N_\mathrm{max}}\sum_{k=1}^{N_\mathrm{max}}\mathbf{1}_{D_\nu}(\omega^j), \quad \omega^j \sim P.$$

In other words, it is the sum of N_max i.i.d. random variables $(\mathbf{1}_{D_\nu}/N_\mathrm{max})$. The expectation of each random variable is $\mathrm{P}[D_\nu]/N_\mathrm{max}$. Therefore

$$\begin{aligned}
\mathrm{Cov}_\mathrm{P}\left(|H^\nu_\mathrm{max}|, |H^\mu_\mathrm{max}|\right) = \mathbb{E}_\mathrm{P}\Bigg\{&\left(\sum_{k=1}^{N_\mathrm{max}}\frac{\mathbf{1}_{D_\nu}(\omega^k)}{N_\mathrm{max}} - \frac{P[D_\nu]}{N_\mathrm{max}}\right)\\
&\times\left(\sum_{\ell=1}^{N_\mathrm{max}}\frac{\mathbf{1}_{D_\mu}(\omega^\ell)}{N_\mathrm{max}} - \frac{P[D_\mu]}{N_\mathrm{max}}\right)\Bigg\}.
\end{aligned}$$

Since ω^k and ω^ℓ are uncorrelated for $k \neq \ell$ the cross terms are zero, and we are left with a sum of $N_{\max}$ expectations

$$\begin{aligned}\frac{1}{N_{\max}^2}\sum_{k=1}^{N_{\max}} &\mathbb{E}_{\mathrm{P}}\left\{\left(\mathbf{1}_{D_\nu}(\omega^k)-P[D_\nu]\right)\left(\mathbf{1}_{D_\mu}(\omega^k)-P[D_\mu]\right)\right\}\\ &=\frac{1}{N_{\max}}\mathbb{E}_{\mathrm{P}}\left\{\left(\mathbf{1}_{D_\nu}-P[D_\nu]\right)\left(\mathbf{1}_{D_\mu}-P[D_\mu]\right)\right\}\\ &=\frac{1}{N_{\max}}\mathbb{E}_{\mathrm{P}}\left\{\mathbf{1}_{D_\nu}\mathbf{1}_{D_\mu}\right\}-P[D_\nu]P[D_\mu].\end{aligned}$$

If $\nu=\mu$, then $\mathbf{1}_{D_\nu}\mathbf{1}_{D_\mu}=\mathbf{1}_{D_\nu}^2=\mathbf{1}_{D_\nu}$. On the other hand, if $\nu\neq\mu$ then $\mathbf{1}_{D_\nu}\mathbf{1}_{D_\mu}=0$ since a photon cannot enter both detectors at once. The result follows. ■

Although $\{|H^\nu_{\max}|, \nu=1,\ldots,m\}$ are negatively correlated, so long as $|H^\nu_j|$ may be selected independently of $H^\nu_{\max}$, the sets H^ν_j are independent. We will always ensure this condition holds.

Our improvement on $\mathbf{f}_j$ in (B1) is $\mathbf{F}_j=(F_j^1,\ldots,F_j^m)$ where

$$F_j^\nu(\gamma):=\frac{|H^\nu_{\max}|}{N_{\max}}\frac{1}{|H^\nu_j|}\sum_{\omega_k\in H^\nu_j}\left|\frac{\mathrm{d}P_\gamma}{\mathrm{d}P}\right|(\omega_k). \tag{B2}$$

Notice that, if $\mathrm{P}=\mathrm{P}_\gamma$, then F_j^ν sums $|H^\nu_j|$ i.i.d. draws from $\mathrm{P}[\cdot|D_\nu]$, and each of them scores a hit $|H^\nu_{\max}|/N_{\max}$. In other words, up to the approximations $\mathrm{P}_\gamma\approx\mathrm{P}$, and $|H^\nu_{\max}|/N_{\max}\approx\mathrm{P}[D_\nu]$, $F_j^\nu(\gamma)$ sums $|H^\nu_j|$ random variables, each one recording the exact solution. Hence, up to these approximations, $F_j^\nu(\gamma)$ computes $\mathrm{P}[D_\nu]$ with zero variance.

On the practical side, $H^\nu_j\subset H^\nu_{j'}$, for $j<j'$, and therefore the computation of $\mathbf{F}_{j'}$ is quicker after computation of $\mathbf{F}_j$ is done.

The next theorem shows that the estimates $\mathbf{F}_j$ are unbiased. See [29] for a proof.

Theorem B.1:

$$\mathbb{E}_{\mathrm{P}}\left\{F_j(\gamma)\right\}=\mathbb{E}_\gamma\{\mathbf{1}_{\mathbf{D}}\}=P_\gamma[\mathbf{D}].$$

The following theorem shows that, in the limit where $\mathrm{dP}_\gamma\to\mathrm{dP}$ and $|H^\nu_{\max}|\to\infty$, the $\mathbf{F}_j^\nu(\gamma)$ are uncorrelated zero-variance estimates of $\mathrm{P}_\gamma[D_\nu]$. Again, see [29] for a proof.

Theorem B.2: As $N_{\max}\to\infty$,

$$\begin{aligned}\mathrm{Cov}_{\mathrm{P}}\left(F_j^\nu(\gamma),F_j^\mu(\gamma)\right)\to\delta_{\mu\nu}&\frac{P[D_\nu]}{|H^\nu_j|}\\ &\times\int_{D_\nu}\left(\left|\frac{\mathrm{d}P_\gamma}{\mathrm{d}P}\right|(\omega)-\frac{P_\gamma[D_\nu]}{P[D_\nu]}\right)^2\mathrm{d}P(\omega).\end{aligned}$$

Remark: A similar calculation shows that

$$\begin{aligned}\mathrm{Cov}_{\mathrm{P}}&\left(\mathbf{f}^\nu_{N_j}(\gamma),\mathbf{f}^\mu_{N_j}(\gamma)\right)\\ &=\frac{1}{N_j}\begin{cases}\displaystyle\int_{D_\nu}\left(\left|\frac{\mathrm{d}P_\gamma}{\mathrm{d}P}\right|-P_\gamma[D_\nu]\right)\mathrm{d}P_\gamma, & \nu=\mu\\ -P_\gamma[D_\nu]P_\gamma[D_\mu], & \nu\neq\mu.\end{cases}\end{aligned}$$

In the expression for $\mathrm{Cov}_{\mathrm{P}}(\mathrm{F}_j^\nu(\gamma_0),\mathrm{F}_j^\mu(\gamma_0))$, one can replace $|H^\nu_j|$ with $N_j\mathrm{P}[D^\nu]$ and see that, if $\mathrm{dP}\approx\mathrm{dP}_\gamma$, then $\mathrm{Cov}_{\mathrm{P}}(\mathrm{F}_j^\nu(\gamma),\mathrm{F}_j^\mu(\gamma))\ll\mathrm{Cov}_{\mathrm{P}}(\mathrm{f}_j^\nu,\mathrm{f}_j^\mu)$. In other words, the variance of our unbiased estimator $\mathbf{F}_j$ is significantly smaller than the estimator $\mathbf{f}_{N_j}$ typically used in aforementioned perturbation MC schemes.

ACKNOWLEDGMENT

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. Y. Marzouk at MIT. AD thanks Prof. Em. A. Moffat, his M.S. supervisor at "U de M," for gently and patiently introducing him to the thrilling world of scientific research. The six anonymous reviewers of the original two-part version of this paper are thanked for many penetrating comments and suggestions for improvement.

REFERENCES

[1] J. R. Schott, *Remote Sensing—The Image Chain Approach*, 2nd ed. Oxford, U.K.: Oxford Univ. Press, 2007.
[2] S. Chandrasekhar, *Radiative Transfer*. Oxford, U.K.: Oxford University Press, 1950, [reprinted by Dover Publications, New York (NY), 1960].
[3] A. B. Davis and A. Marshak, "Solar radiation transport in the cloudy atmosphere: A 3D perspective on observations and climate impacts," *Rep. Progr. Phys.*, vol. 73, no. 2, p. 026801, Feb. 2010.
[4] A. Marshak and A. B. Davis, Eds., *3D Radiative Transfer in Cloudy Atmospheres*. Heidelberg, Germany: Springer-Verlag, 2005.
[5] J. Otterman and R. S. Fraser, "Adjacency effects on imaging by surface reflection and atmospheric scattering: Cross radiance to zenith," *Appl. Opt.*, vol. 18, no. 16, pp. 2852–2860, Aug. 1979.
[6] Y. Mekler and Y. J. Kaufman, "The effect of Earth's atmosphere on contrast reduction for a nonuniform surface albedo and 'two-halves' field," *J. Geophys. Res.*, vol. 85, no. C7, pp. 4067–4083, Aug. 1980.
[7] J. Otterman, S. Ungar, Y. J. Kaufman, and M. Podolak, "Atmospheric effects on radiometric imaging from satellites under low optical thickness conditions," *Remote Sens. Environ.*, vol. 9, no. 2, pp. 115–129, Mar. 1980.
[8] D. Tanré, M. Herman, and P.-Y. Deschamps, "Influence of the background contribution upon space measurements of ground reflectance," *Appl. Opt.*, vol. 20, no. 20, pp. 3676–3684, Oct. 1981.
[9] Y. J. Kaufman, "Solution of the equation of radiative-transfer for remote-sensing over nonuniform surface reflectivity," *J. Geophys. Res.*, vol. 87, no. C6, pp. 4137–4147, May 1982.
[10] D. J. Diner and J. V. Martonchik, "Atmospheric transfer of radiation above an inhomogeneous non-Lambertian ground: 1—Theory," *J. Quant. Spectrosc. Radiat. Transf.*, vol. 31, no. 2, pp. 97–125, Feb. 1984.
[11] D. J. Diner and J. V. Martonchik, "Atmospheric transfer of radiation above an inhomogeneous non-Lambertian ground: 2—Computational considerations and results," *J. Quant. Spectrosc. Radiat. Transf.*, vol. 32, no. 4, pp. 279–304, Oct. 1984.
[12] A. Royer, A. Davis, and N. O'Niell, "Analyse des effets atmosphériques dans les images HRV de SPOT," *Can. J. Remote Sens.*, vol. 14, no. 2, pp. 80–91, 1989.
[13] T. Takashima and K. Masuda, "Simulation of atmospheric effects on the emergent radiation over a checkerboard type of terrain," *Astrophys. Space Sci.*, vol. 198, no. 2, pp. 253–263, 1992.
[14] P. N. Reinersman and K. L. Carder, "Monte Carlo simulation of the atmospheric point-spread function with an application to correction for the adjacency effect," *Appl. Opt.*, vol. 34, no. 21, pp. 4453–4471, Jul. 1995.
[15] A. I. Lyapustin, "Three-dimensional effects in the remote sensing of surface albedo," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 254–263, Feb. 2001.
[16] A. I. Lyapustin and Y. J. Kaufman, "Role of adjacency effect in the remote sensing of aerosol," *J. Geophys. Res.*, vol. 106, no. D11, pp. 11 909–11 916, Jun. 2001.
[17] A. I. Lyapustin and Y. Knyazikhin, "Green's function method for the radiative transfer problem. II. Spatially heterogeneous anisotropic surface," *Appl. Opt.*, vol. 41, no. 27, pp. 5600–5606, Sep. 2002.
[18] A. I. Lyapustin, T. Muldashev, and Y. Wang, "Code SHARM: Fast and accurate radiative transfer over spatially variable anisotropic surfaces," in *Light Scattering Reviews*, vol. 5, A. A. Kokhanovsky, Ed. Heidelberg, Germany: Springer Praxis, 2010, pp. 205–247.
[19] A. A. Semenov, A. V. Moshkov, V. N. Pozhidayev, A. Barducci, P. Marcoionni, and I. Pippi, "Estimation of normalized atmospheric point spread function and restoration of remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2623–2634, Jul. 2011.
[20] R. Richter and D. Schlaepfer, "Geo-atmospheric processing of airborne imaging spectrometry data, Part 2: Atmospheric/topographic correction," *Int. J. Remote Sens.*, vol. 23, no. 13, pp. 2631–2649, 2002.
[21] P. E. Ardanuy, D. Han, and V. V. Salomonson, "The MODerate resolution Imaging Spectrometer (MODIS) science and data system requirements," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 1, pp. 75–88, Jan. 1991.

[22] H. H. Aumann, M. T. Chahine, C. Gautier, M. D. Goldberg, E. Kalnay, L. M. McMillin, H. Revercomb, P. W. Rosenkranz, W. L. Smith, D. H. Staelin, L. L. Strow, and J. Susskind, "AIRS/AMSU/HSB on the aqua mission: Design, science objectives, data products, and processing systems," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 253–264, Feb. 2003.

[23] 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, no. 4, pp. 1072–1087, Jul. 1998.

[24] J. R. Schott, *Fundamentals of Polarimetric Remote Sensing*. Bellingham, WA: SPIE, 2009.

[25] 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, no. 3, pp. 598–615, May 1994.

[26] C. Funk, J. Theiler, D. Roberts, and C. Borel, "Clustering to improve matched filter detection of weak gas plumes in hyperspectral thermal imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1410–1420, Jul. 2001.

[27] E. Hirsch and E. Agassi, "Detection of gaseous plumes in ir hyperspectral images—Performance analysis," *IEEE Sensors J.*, vol. 10, no. 3, pp. 732–736, Mar. 2010.

[28] P. Addabbo, M. di Bisceglie, and C. Galdi, "The unmixing of atmospheric trace gases from hyperspectral satellite data," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 320–329, Jan. 2012.

[29] G. Bal, I. Langmore, and Y. Marzouk, "Bayesian inverse problems with Monte Carlo forward models," *Inv. Probl. Imaging*, vol. 7, no. 1, pp. 81–105, 2013. doi:10.3934/ipi.2013.7.81.

[30] L. C. Henyey and J. L. Greenstein, "Diffuse radiation in the galaxy," *Astrophys. J.*, vol. 93, pp. 70–83, Jan. 1941.

[31] R. L. White, "Polarization in reflection nebulae. I—Scattering properties of interstellar grains," *Astrophys. J.*, vol. 229, pp. 954–961, May 1979.

[32] A. Davis, P. Gabriel, S. Lovejoy, and D. Schertzer, "Scaling laws for asymptotically thick clouds, dimensional dependence—Phase function independence," in *Proc. IRS—Current Problems Atmospheric Radiation*, J. Lenoble and J.-F. Geleyn, Eds., 1989, pp. 103–106.

[33] J. Heino, S. R. Arridge, J. Sikora, and E. Somersalo, "Anisotropic effects in highly scattering media," *Phys. Rev. E*, vol. 68, no. 3, pp. 031908-1–031908-8, Sep. 2003.

[34] M. I. Mishchenko, "Multiple scattering, radiative transfer, and weak localization in discrete random media: Unified microphysical approach," *Rev. Geophys.*, vol. 46, p. RG2003, Apr. 2008.

[35] A. B. Davis, "Cloud remote sensing with sideways-looks: Theory and first results using multispectral thermal imager (MTI) data," in *SPIE Proc.*, 2002, vol. 4725, pp. 397–405.

[36] G. Marchuk, G. Mikhailov, M. Nazaraliev, R. Darbinjan, B. Kargin, and B. Elepov, *The Monte Carlo Methods in Atmospheric Optics*. New York: Springer-Verlag, 1980.

[37] G. Bal, A. B. Davis, and I. Langmore, "A hybrid (Monte-Carlo/deterministic) approach for multidimensional radiation transport," *J. Comput. Phys.*, vol. 230, pp. 7723–7735, 2011.

[38] 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, no. 17, pp. 1335–1337, Sep. 2001.

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

[40] J. Chen and X. Intes, "Time-gated perturbation Monte Carlo for whole body functional imaging in small animals," *Opt. Exp.*, vol. 17, no. 22, pp. 19 566–19 579, Oct. 2009.

[41] W. M. Bolstad, *Introduction to Bayesian Statistics*, 2nd ed. New York: Wiley, 2007.

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

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

[44] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1092, Jun. 1953.

[45] W. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, Apr. 1970.

[46] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer-Verlag, 2004.

[47] J. Andrés Christen and C. Fox, "Markov chain Monte Carlo using an approximation," *J. Comput. Graph. Stat.*, vol. 14, no. 4, pp. 795–810, Dec. 2005.

[48] Y. Efendiev, T. Hou, and W. Luo, "Preconditioning Markov chain Monte Carlo simulations using coarse-scale models," *SIAM J. Sci. Comput.*, vol. 28, no. 2, pp. 776–803, 2006.

[49] 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. Technol.*, vol. 4, no. 5, pp. 975–1018, May 2011.

[50] 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, no. 7, pp. 1520–1531, Jul. 2002.

[51] [Online]. Available: http://i3rc.gsfc.nasa.gov/I3RC_community_model_new.htm

[52] [Online]. Available: http://dirsig.org/

[53] G. E. Shilov and B. L. Gurevich, *Integral, Measure, and Derivative: A Unified Approach*. New York: Dover, 1978.

Ian Langmore received the B.Sc. and M.Sc. degrees in electrical engineering from the University of California, San Diego, CA in 2002 and 2004, respectively, and the Ph.D. degree in mathematics from the University of Washington, Seattle, WA in 2008. He worked on this manuscript as a Postdoctoral Researcher at the Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY.

His work there is related to inverse problems in radiative transfer. Most of his recent work involves Monte Carlo, including Markov chain Monte Carlo, algorithms. Since leaving Columbia in September 2011, he has been working at Opera Solutions as a Data Scientist specializing in machine learning algorithms and Monte Carlo techniques.

Anthony B. Davis received the B.Sc. degree in physics from the Université Paris VI, Paris, France, in 1977, the M.Sc. degree in physics from Université de Montréal, Montreal, Canada, in 1982, and the Ph.D. degree in physics from McGill University, Montreal, in 1992.

He 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, Greenbelt, MD, the Department of Energy's Los Alamos National Laboratory, New Mexico, and now at the Jet Propulsion Laboratory, Pasadena, CA. He specializes in the challenges posed by the 3-D shapes of real clouds to which he applies his expertise in theoretical and computational 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, his interests have grown to include aerosol/surface property remote sensing, particularly in the vicinity of clouds, or other spatially complex situations, using multipixel/multi-angle spectral data with or without polarization.

Guillaume Bal graduated from École Polytechnique, Palaiseau, France, in 1993, and the Ph.D. degree in applied mathematics from the Université Paris VI, Paris, France, in 1997.

He joined the faculty of Columbia University's Department of Applied Physics and Applied Mathematics, in 2001, after postdoctoral research at Stanford (1997–1999) and a stint as an L. E. Dickson Instructor at the University of Chicago, in 1999–2001. His 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 and geophysical imaging.

Dr. Bal received, among other awards, the 2011 Calderón Prize for outstanding achievements in inverse problem theory.