A Method for Retrieving the Ground Flash Fraction and Flash Type from Satellite Lightning Mapper Observations

WILLIAM J. KOSHAK

NASA Marshall Space Flight Center, Huntsville, Alabama

RICHARD J. SOLAKIEWICZ

Chicago State University, Chicago, Illinois

(Manuscript received 21 April 2014, in final form 8 August 2014)

ABSTRACT

An analytic perturbation method is introduced for retrieving the lightning ground flash fraction in a set of N lightning flashes observed by a satellite lightning mapper. The value of N must be large, typically in the thousands, and the satellite lightning optical observations consist of the maximum group area (MGA) produced by each flash. Moreover, the method subsequently determines the flash type (ground or cloud) of each of the N flashes. Performance tests of the method were conducted using simulated observations that were based on Optical Transient Detector (OTD) and Lightning Imaging Sensor (LIS) data. It is found that the mean ground flash fraction retrieval errors are below 0.04 across the full range 0–1 under the nominal conditions defined. In general, it is demonstrated that the retrieval errors depend on many factors (i.e., the number N of satellite observations, the magnitude of random and systematic instrument measurement errors, the ground flash fraction itself, and the number of samples used to form certain climate distributions employed in the method). The fraction of flashes accurately flash typed by the method averaged better than 78%. Overall, the accuracy of ground flash fraction and flash-typing retrievals degrade as the simulated population ground and cloud flash MGA distributions become more identical. Finally, because the analytic perturbation method was found to be quite robust (i.e., it performed well for several arbitrary underlying MGA distributions), it is not restricted to the lightning problem studied here but can be applied to any inverse problem having a similar problem statement.

1. Introduction

The Geostationary Operational Environmental Satellite-R series (GOES-R) is due to launch in early 2016 as of this writing. In preparation for this launch and in order to optimize return on investment, many ongoing research activities are being carried out to explore and exploit the full information content of the GOES-R instrument data.

One instrument on GOES-R is the Geostationary Lightning Mapper (GLM) described in Goodman et al. (2013). The GLM will map the locations and the time of occurrence of *total* (ground flash and cloud flash) lightning activity continuously day and night with near-uniform storm-scale spatial resolution and with a product refresh rate of under 20 s over the Americas and adjacent oceanic regions. The GLM is based on two heritage low-Earth-orbiting National Aeronautics and Space Administration (NASA) satellite lightning mappers: the Optical Transient Detector (OTD) (Christian et al. 1996, 2003) and the Lightning Imaging Sensor (LIS) (Christian et al. 1992, 1999).

Although the above-mentioned lightning mappers are designed to provide total lightning activity, these instruments do not directly provide flash type (ground or cloud flash) classification; that is, they do not determine the flashes that strike the ground. This is understandable, since the optically thick thundercloud obscures the view, making it difficult to determine flash type. Early examinations of data from the OTD and LIS confirm this difficulty. At optical frequencies, the thundercloud multiply scatters the

Denotes Open Access content.

Corresponding author address: William Koshak, Earth Science Office, NASA Marshall Space Flight Center, ZP11, 320 Sparkman Dr., Huntsville, AL 35805-1912.
E-mail: william.koshak@nasa.gov

DOI: 10.1175/JTECH-D-14-00085.1

lightning source, and this results in a diffuse cloud-top emission that prevents one from deciding whether the lightning channel below cloud top connects to ground.

Nonetheless, the spatiotemporal pattern of the diffuse cloud-top emission itself provides key information about whether or not the channel connects to ground, as first pointed out in Koshak (2010). Specifically, by intercomparing OTD and National Lightning Detection Network (NLDN; Cummins and Murphy 2009) data, Koshak (2010) showed that the mean maximum group area (MGA) for ground flashes was substantially larger than the mean MGA for cloud flashes, so that for spaceborne detectors, MGA is a fundamental variable for discriminating ground flashes from cloud flashes. [Note: a lightning flash typically lasts a few tenths of a second and typically consists of several optical groups. The term *group* follows the vernacular used to describe a contiguous region of illumination seen by a lightning mapper during one (2 ms) instrument frame time; Mach et al. 2007.]

The hypothesis proposed by Koshak (2010) is that the return stroke (and its accompanying simultaneous discharges, if any) produces a relatively large optical group within the flash. In effect, a statistically large MGA is useful as a "return stroke detector." The mean ground flash MGA also substantially exceeded the mean cloud flash MGA when 9 years (2003–11) of LIS and NLDN data were intercompared (Koshak and Solakiewicz 2012). In addition, the study by Rudlosky and Shea (2013) intercompared World Wide Lightning Location Network (WWLLN) data with LIS data and found that LIS flashes associated with WWLLN ground flash detections had substantially larger mean MGA values than LIS flashes not associated with WWLLN ground flash detections. Some of the difference in the mean MGA value can possibly be attributable to WWLLN triggering only on the more energetic lightning events. But, this does not explain all of the difference; that is, the study by Koshak (2010) and Koshak and Solakiewicz (2012) still found substantial differences in the mean MGA values when (the more sensitive) NLDN data were used for the ground flash detections. Whereas the NLDN provides a ground flash detection efficiency of 90%–95% and a ground flash location accuracy of better than 500 m across the conterminous United States (CONUS) (Cummins and Murphy 2009), the WWLLN has respective values of 11% and under 5 km globally as summarized in Hutchins et al. (2012) and references therein.

The follow-on studies by Koshak and Solakiewicz (2011) and Koshak (2011) introduced the first attempts to retrieve the *ground flash fraction* based on the MGA variable suggested by Koshak (2010). That is, for a satellite lightning mapper that observes a large number N of flashes, these retrieval methods are designed to estimate the fraction of flashes that strike the ground. As discussed in Koshak (2010), ground flash fraction retrieval algorithms would have significant utility in a variety of scientific areas, such as severe weather studies, lightning nitrogen oxide (NOx) production for air quality and global chemistry climate modeling, lightning–convection relationships, the contribution of lightning to the global electric circuit, and cross-sensor validation.

The retrieval algorithm in Koshak and Solakiewicz (2011) depended on using a fixed value for the mean ground flash MGA and a fixed value for the mean cloud flash MGA; the two means were derived for CONUS by comparing NLDN and OTD observations. Thus, retrieval errors are expected to increase when other geographical regions and seasons are considered.

The more sophisticated retrieval algorithm in Koshak (2011) attempted to overcome this limitation by employing a mixed exponential distribution model to describe the ground and cloud flash (shifted) MGA distributions. In this approach, one would retrieve not only the ground flash fraction for the region–season under consideration but also the optimum values of the population means of each shifted MGA distribution; that is, the approach was to find the three parameters of the fixed (mixed exponential distribution) model that best describes the satellite observations.

Even with this beneficial advancement, one can still argue that a different region/season might have ground and cloud flash MGA population distributions that deviate too much from the fixed mathematical model employed. In addition, both retrieval algorithms provide no way for flash typing specific flashes. These two limitations have motivated the present study.

In this paper we introduce a *model-independent* method for retrieving the ground flash fraction in a set of N flashes observed by a satellite lightning mapper (such as OTD, LIS, or the future GLM). The method also subsequently determines the flash type of each of the N flashes.

The retrieval method introduced here is called the Analytic Perturbation Method (APM), and it consists of two basic phases: 1) a "burn in" phase and 2) an operational phase wherein the APM is applied to retrieve the ground flash fraction, the ground and cloud flash MGA density functions, and the flash type of each flash. The details of the APM are provided in section 2. In section 3, we provide detailed performance tests of the APM. Mathematical error analyses are provided in section 4 that provide additional insight and context. A summary of the study is given in section 5.

2. APM

This section introduces the mathematical details of the APM. Section 2a describes the burn-in phase. Section 2b

describes how the ground flash fraction, the density function of ground flash MGAs, and the density function of cloud flash MGAs are retrieved from a set of N flashes observed by a lightning mapper. Finally, section 2c describes how the N flashes are individually flash typed.

a. Burn-in phase

Before the APM can be applied to retrieve the quantities mentioned above, the lightning mapper ideally first collects a large sample of flashes from the target (i.e., region and season of interest) that have already been flash typed using an independent validation dataset. This collection period or "burn-in" phase estimates the mathematical forms of the underlying ground and cloud flash MGA population density functions for the target. For example, the GLM would sample a large number of flashes over the CONUS using NLDN data for flash typing.

If independent validation datasets are unavailable for a particular target, then one is forced to apply burn-in results from a different target (e.g., CONUS, and for the same season of interest, where independent ground validation is ample). In taking this route, one would need to anticipate degradation in APM solution retrieval accuracy (see section 4 for details). However, it has been demonstrated in Koshak and Solakiewicz (2011) that, at least for CONUS (a region of highly variable lightning, highly variable thunderstorm structures, and highly variable cloud morphologies with diverse cloud-scattering properties) the ground and cloud flash values of MGA do not vary greatly. In other words, it might be possible to adequately apply the APM to most regions of the world using only the CONUS-based burn-in results as an approximation. Again, section 4 quantifies how accurate such a process must be in order to retain reasonably accurate APM retrievals.

To summarize, the burn-in phase can be viewed as a training or information collection period for the retrieval algorithm. After this period, the retrieval algorithm can be run autonomously for a wide range of targets, including possibly those targets that may have no, or inadequate, independent lightning flash type validation measurements available.

b. Ground flash fraction and MGA density functions

Suppose a lightning mapper detects N lightning flashes over a particular region in a given interval of time. Each flash is composed of one or more "optical groups," and so the MGA of each flash can be obtained. The MGA of a single group flash is simply the area of that single group. If two or more groups in a flash have equivalent areas A, and this area A is the MGA, then this is simply the value of MGA for the flash (i.e., one does not numerically enhance the value of the MGA in an attempt to account for the several equivalent maxima). In addition, it is assumed for the present discussion that the N MGA values are exact; that is, they have no measurement error associated with them.

Next, the N MGA values are placed into an N-vector $\mathbf{x}$. In general, N_g of these MGAs will be from ground flashes and N_c of these MGAs will be from cloud flashes, where $N = N_g + N_c$. Therefore, one can also construct the N_g-vector $\mathbf{x}_g$ to hold the ground flash MGAs, and one can construct the N_c-vector $\mathbf{x}_c$ to hold the cloud flash MGAs.

From $\mathbf{x}$, one can easily find the frequency distribution of the MGA values. Specifically, one picks an MGA bin width (typically 20 km^2) and some appropriate MGA range (say, 0–4000 km^2), and then determines to what bin each element of $\mathbf{x}$ belongs. The number of bins is given by n, which in this example would equal 4000 km^2/20 km^2 = 200. Hence, after tallying the number of occurrences in each bin, the MGA frequency distribution can be written as an n-vector $\mathbf{U}$. Because $\mathbf{U}$ consists of the union of ground and cloud flash MGAs in general, it is called the *union frequency distribution*, or the *mixture distribution* for brevity. Similarly, a ground flash MGA frequency distribution, $\mathbf{G}$, can be derived from $\mathbf{x}_g$, and a cloud flash MGA frequency distribution, $\mathbf{C}$, can be derived from $\mathbf{x}_c$. Clearly, the sum of ground and cloud MGA frequency distributions must equal the mixture distribution, that is,

$$\mathbf{U} = \mathbf{G} + \mathbf{C}. \tag{1}$$

Dividing this relationship by N yields an expression in terms of probability density functions (pdfs), that is,

$$\mathbf{u} = \alpha\mathbf{g} + (1 - \alpha)\mathbf{c}, \tag{2}$$

where

$$\mathbf{U} = N\mathbf{u}, \quad \mathbf{G} = N_g\mathbf{g}, \quad \mathbf{C} = N_c\mathbf{c}, \tag{3}$$

and the ground flash fraction, α, is given by

$$\alpha = \frac{N_g}{N}. \tag{4}$$

Next, one can set up a fundamental retrieval equation. To do this, the presence of measurement errors in the MGA values needs consideration. These errors will propagate into the pdfs shown in (2). Characterizing the errors added to $\mathbf{u}$ by the vector $\boldsymbol{\varepsilon}$, the *observed mixture density* $\mathbf{m}$ is given by $\mathbf{m} = \mathbf{u} + \boldsymbol{\varepsilon}$. So, when measurement errors are included, the analog to (2) is given by

$$\mathbf{m} = \alpha_r\mathbf{g}_r + (1 - \alpha_r)\mathbf{c}_r. \tag{5}$$

Here, the lhs represents a set of measurements (originating from the satellite MGA observations $\mathbf{x}$) and the

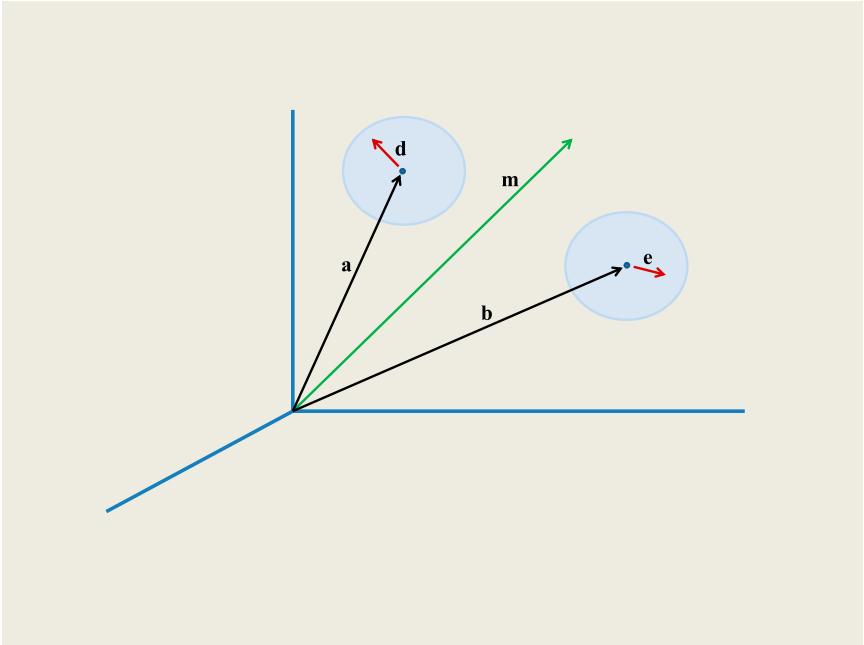

FIG. 1. Geometry associated with the perturbation approach (depicted conceptually in three dimensions). The blue-shaded regions are spherical.

rhs is a modeling of those measurements. The retrievable parameters in the model are α_r, $\mathbf{g}_r$, and $\mathbf{c}_r$.

The densities ($\mathbf{g}$, $\mathbf{c}$) associated with the satellite observations $\mathbf{x}$ will be some deviation from the associated population densities [denoted here by ($\mathbf{g}_{\text{pop}}$, $\mathbf{c}_{\text{pop}}$)] from which they are respectively drawn. Hence, the retrieved densities can be viewed as a perturbation from the climatological estimate of the population densities; that is,

$$\mathbf{g}_r = \mathbf{a} + \mathbf{d}, \qquad \mathbf{c}_r = \mathbf{b} + \mathbf{e}, \tag{6}$$

where ($\mathbf{a}$, $\mathbf{b}$) are the ground and cloud flash climatological estimates of the population densities obtained by the burn-in phase, respectively, and ($\mathbf{d}$, $\mathbf{e}$) are the perturbations. To help visualize the n-dimensional space employed, Fig. 1 provides a conceptual (three dimensional) illustration of the climate vectors ($\mathbf{a}$, $\mathbf{b}$), their perturbants ($\mathbf{d}$, $\mathbf{e}$), and the mixture density vector ($\mathbf{m}$).

Figure 2 emphasizes that the hyperplane formed by the retrieved densities ($\mathbf{g}_r$, $\mathbf{c}_r$) must be perturbed from the climate vectors in such a way that the (fixed) mixture density $\mathbf{m}$ is in this hyperplane. In other words, one wants to pick the perturbants ($\mathbf{d}$, $\mathbf{e}$) such that (5) holds; the skewed vectors ($\mathbf{g}_r$, $\mathbf{c}_r$) span the hyperplane vector space.

As an additional aid in understanding the solution process, one can define a *reference density* $\mathbf{m}_o$ that is in the hyperplane space spanned by the climate vectors as

$$\mathbf{m}_o = \alpha_o \mathbf{a} + (1 - \alpha_o)\mathbf{b}, \tag{7}$$

where the *reference ground flash fraction*, α_o, is given by solving (7):

$$\alpha_o = \frac{(\mathbf{m}_o - \mathbf{b})^{\mathrm{T}}(\mathbf{a} - \mathbf{b})}{(\mathbf{a} - \mathbf{b})^2}. \tag{8}$$

Since the observation $\mathbf{m}$ is not (in general) in the hyperplane space spanned by the climate vectors, it deviates from the reference $\mathbf{m}_o$ by some perturbation $\mathbf{p}(\alpha_o) \equiv \mathbf{p}_o$, that is,

$$\mathbf{m} = \mathbf{m}_o + \mathbf{p}_o. \tag{9}$$

Now, if N_g is large, then $\mathbf{g}$ likely deviates little from $\mathbf{g}_{\text{pop}}$. Similarly, if N_c is large, then $\mathbf{c}$ likely deviates little from $\mathbf{c}_{\text{pop}}$. This motivates finding the values of (α_r, $\mathbf{d}$, $\mathbf{e}$) that minimize a scalar function, $H = H(\alpha_r, \mathbf{d}, \mathbf{e})$, given by

$$H = \alpha_r \mathbf{d}^2 + (1 - \alpha_r)\mathbf{e}^2, \tag{10}$$

which is subject to the constraint given in (5). The retrieved densities in (5) are as given in (6). For example, if the observations $\mathbf{x}$ contained mostly cloud flashes, then the ground flash fraction is near zero, and one would retrieve a value for α_r close to zero. This implies that $\mathbf{d}^2$ could get relatively large without making H too large (desirable, since $\mathbf{g}$ likely deviates from $\mathbf{g}_{\text{pop}}$). It also implies that $\mathbf{e}^2$ should be forced to stay relatively small, otherwise H will get too large (desirable, since $\mathbf{c}$ likely deviates relatively little from $\mathbf{c}_{\text{pop}}$).

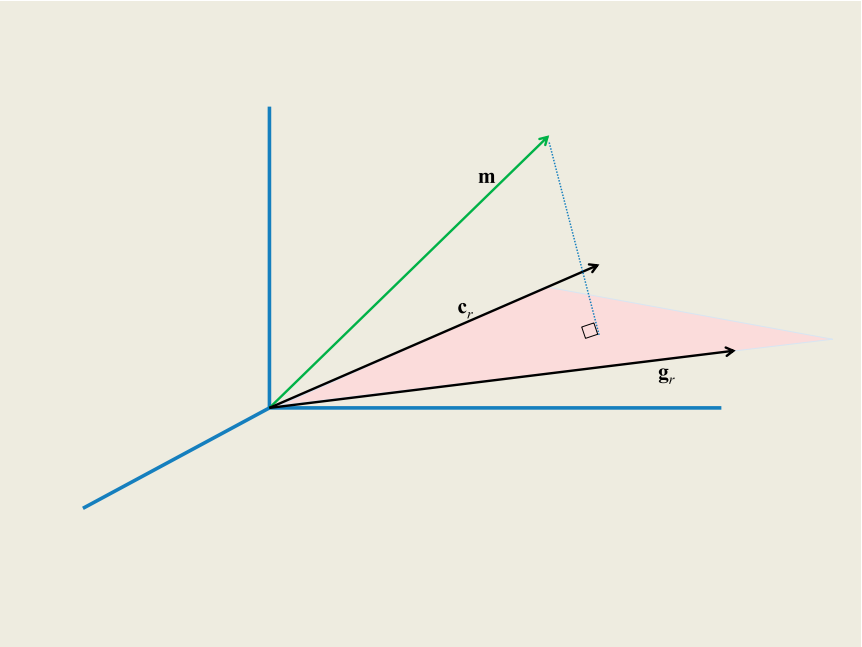

FIG. 2. The solution space hyperplane depicted in three dimensions; the hyperplane must be moved so that it contains the (fixed) **m**.

One can guarantee that (5) will hold exactly by choosing one perturbant properly. Substituting (6) into (5) and solving for **d** gives

$$\mathbf{d} = \frac{1}{\alpha_r}[\mathbf{m} - (1-\alpha_r)(\mathbf{b}+\mathbf{e})] - \mathbf{a}. \tag{11}$$

Alternatively, one could just as well have eliminated **e** in terms of **d**; the choice does not change the final retrieval solution.

Substituting (11) into (10) produces a new scalar function $Q = Q(\alpha_r, \mathbf{e})$, given by

$$Q = \sum_{i=1}^{n} [(\alpha_r A + \beta_r)e_i^2 + \alpha_r B_i e_i + \alpha_r C_i], \tag{12}$$

where the C_i are *not* to be confused with the components of **C** in (1) and

$$\begin{aligned} A &= \beta_r^2/\alpha_r^2, \\ B_i &= \frac{2\beta_r(\alpha_r a_i - \phi_i)}{\alpha_r^2}, \\ C_i &= \frac{\phi_i^2 - 2\alpha_r a_i \phi_i + \alpha_r^2 a_i^2}{\alpha_r^2}, \\ \phi_i &= m_i - \beta_r b_i, \quad \beta_r = 1 - \alpha_r. \end{aligned} \tag{13}$$

Next, one can find the **e** that minimizes Q by solving the system of equations $\partial Q/\partial e_k = 0$, with $k = 1, 2, \ldots, n$, which gives the critical point perturbant

$$(e_k)_{\text{crit}} = \frac{-\alpha_r B_k}{2(\alpha_r A + \beta_r)}. \tag{14}$$

Substituting (14) into (12) gives a new scalar function $S = S(\alpha_r)$, given by

$$S = \sum_{i=1}^{n} \left[\alpha_r C_i - \frac{\alpha_r^2 B_i^2}{4(\alpha_r A + \beta_r)}\right]. \tag{15}$$

The expressions in (13) can be used to further simplify (14) and (15); the algebra is lengthy but straightforward. Also, (14) is substituted into (11) to get the associated critical point form for **d**. The rather interesting results are

$$\begin{aligned} \mathbf{e}_{\text{crit}} &= \mathbf{e}_{\text{crit}}(\alpha_r) = \mathbf{m} - [\alpha_r \mathbf{a} + (1-\alpha_r)\mathbf{b}] = \mathbf{p}(\alpha_r), \\ \mathbf{d}_{\text{crit}} &= \mathbf{d}_{\text{crit}}(\alpha_r) = \mathbf{e}_{\text{crit}}, \\ P &= P(\alpha_r) = \mathbf{e}_{\text{crit}}^2. \end{aligned} \tag{16}$$

Using the results in (16), one can find α_r that minimizes S by solving the equation $dS/d\alpha_r = 0$. This gives the final APM solution as

$$\begin{aligned} \alpha_r &= \alpha_1 = \frac{(\mathbf{m}-\mathbf{b})^{\mathrm{T}}(\mathbf{a}-\mathbf{b})}{(\mathbf{a}-\mathbf{b})^2}, \\ \mathbf{g}_r &= \mathbf{a} + \mathbf{p}(\alpha_1) = \mathbf{m} + (1-\alpha_1)(\mathbf{a}-\mathbf{b}), \\ \mathbf{c}_r &= \mathbf{b} + \mathbf{p}(\alpha_1) = \mathbf{m} - \alpha_1(\mathbf{a}-\mathbf{b}). \end{aligned} \tag{17}$$

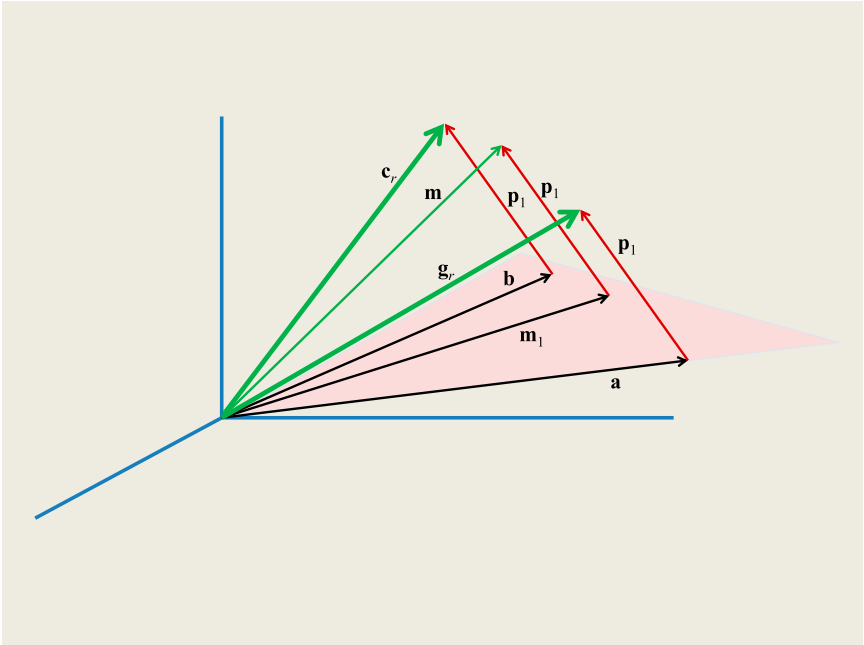

FIG. 3. A geometrical conception of the APM solution process (depicted in three dimensions).

The final perturbants are $\mathbf{d}_{\text{crit}}(\alpha_r = \alpha_1) = \mathbf{e}_{\text{crit}}(\alpha_r = \alpha_1) = \mathbf{m} - \mathbf{m}_1 = \mathbf{p}(\alpha_1) = \mathbf{p}_1$. Note that the solution is in terms of the three vector densities ($\mathbf{m}$, $\mathbf{a}$, $\mathbf{b}$), where $\mathbf{m}$ is derived from the satellite MGA observations $\mathbf{x}$, and the climate vectors ($\mathbf{a}$, $\mathbf{b}$) are obtained from the burn-in process. Since ($\mathbf{m}$, $\mathbf{a}$, $\mathbf{b}$) are each pdfs, they are normalized (their elements sum to unity) and their elements are nonnegative (negative probabilities are not allowed). By inspection of the last two equations in (17), one can see that ($\mathbf{g}_r$, $\mathbf{c}_r$) are normalized. In this lightning retrieval problem, we have found minor instances of some of the elements of $\mathbf{c}_r$ going very slightly negative but not enough to adversely affect our retrieval results. In any case, one can always zero out negative retrieved elements and then renormalize the vector.

Figure 3 provides a geometrical summary of the solution process. The steps taken to minimize H while satisfying (5) have resulted in there being only one viable type of perturbation, given by $\mathbf{p}_1$. Note that for an arbitrary perturbation [i.e., arbitrary ($\mathbf{g}_r$, $\mathbf{c}_r$)], one obtains the usual algebraic result of

$$\alpha_r = \frac{(\mathbf{m} - \mathbf{c}_r)^{\mathrm{T}}(\mathbf{g}_r - \mathbf{c}_r)}{(\mathbf{g}_r - \mathbf{c}_r)^2}. \tag{18}$$

The perturbation $\mathbf{p}_1$ can be described as a transformation $\mathbf{T}$: ($\mathbf{a}$, $\mathbf{b}$) $\rightarrow$ ($\mathbf{g}_r$, $\mathbf{c}_r$) given by the last two equations in (17); that is, the transformation can be expressed as

$$\begin{bmatrix} \mathbf{g}_r \\ \mathbf{c}_r \\ \mathbf{m} \end{bmatrix} = \begin{bmatrix} \beta_r & -\beta_r & 1 \\ -\alpha_r & \alpha_r & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{m} \end{bmatrix}. \tag{19}$$

Here, we use the mnemonic where the vectors play the role of scalars, so that normal matrix multiplication can be used to reproduce the two transformation equations in (17). In addition, $\beta_r = 1 - \alpha_r$ as in (13). The transformation matrix in (19) defines $\mathbf{T}$. The first application of $\mathbf{T}$ perturbs one from $\mathbf{a}$ to $\mathbf{g}_r$, and from $\mathbf{b}$ to $\mathbf{c}_r$, by the amount $\mathbf{p}_1$ as discussed. In other words, substituting the last two equations of (17) into (18) yields α_1.

Finally, the perturbation (transformation) has an interesting mathematical property. Once one completes the transformation ($\mathbf{T}$), additional applications of the same transformation result in no further changes in (α_r, $\mathbf{g}_r$, $\mathbf{c}_r$). This is because $\mathbf{T}^j = \mathbf{T}$, for $j = 1, 2, 3, \ldots$. That is, pre-multiplying both sides of (19) by $\mathbf{T}$ any number of times still gives back the lhs of (19). In other words, and from the perspective of the physical inverse problem at hand, if $\mathbf{m}$ happens to be already in the space spanned by the climate vectors, then no perturbations are necessary to model the observation vector $\mathbf{m}$.

c. *Flash type*

In this section, we show that it is possible to mathematically infer the flash type (i.e., ground or cloud) of each of the N flashes observed. In other words, a distinct advantage of the APM is that it retrieves estimates of

not only the ground flash fraction α but also estimates of the ground and cloud flash MGA probability density functions, $g(x)$ and $c(x)$, respectively. [Note: in this section we use the notation of continuous functions instead of vectors, i.e., $g(x)$ denotes $\mathbf{g}$, $c(x)$ denotes $\mathbf{c}$, and so on.] *The retrieved values in (17) and the observed MGA of a flash allow one to estimate the probability $P_g(x)$ that the observed flash is a ground flash.* The probability that the flash is a cloud flash is simply $1 - P_g(x)$. To understand how this can be accomplished, note the following:

$$\alpha = \frac{N_g}{N_g + N_c} = \frac{N_g \int_0^\infty g(x)\,dx}{N_g \int_0^\infty g(x)\,dx + N_c \int_0^\infty c(x)\,dx}. \tag{20}$$

Now, the number of ground flashes with MGAs in the interval I given by $(x - \delta) < y < (x + \delta)$ is

$$n_g = n_g(x,\ \delta) = N_g \int_{x-\delta}^{x+\delta} g(y)\,dy, \tag{21}$$

where y is a dummy integration variable. Similarly, the number of cloud flashes with MGAs in the interval I is

$$n_c = n_c(x,\ \delta) = N_c \int_{x-\delta}^{x+\delta} c(y)\,dy. \tag{22}$$

The fraction of flashes with MGAs in the interval I that are ground flashes is denoted as $f_g(x, \delta)$. Similarly, the fraction of flashes with MGAs in I that are cloud flashes is denoted as $f_c(x, \delta)$. These fractions are given by

$$f_g(x,\ \delta) = \frac{n_g}{n_g + n_c}, \quad f_c(x,\ \delta) = \frac{n_c}{n_g + n_c}. \tag{23}$$

Substituting (21) and (22) into (23) and noting that $N_g = \alpha N$ and that $N_c = (1 - \alpha)N$, one obtains

$$f_g(x,\ \delta) = \frac{\alpha \int_{x-\delta}^{x+\delta} g(y)\,dy}{\alpha \int_{x-\delta}^{x+\delta} g(y)\,dy + (1-\alpha) \int_{x-\delta}^{x+\delta} c(y)\,dy}, \tag{24}$$

$$f_c(x,\ \delta) = \frac{(1-\alpha) \int_{x-\delta}^{x+\delta} c(y)\,dy}{\alpha \int_{x-\delta}^{x+\delta} g(y)\,dy + (1-\alpha) \int_{x-\delta}^{x+\delta} c(y)\,dy}. \tag{25}$$

In addition, note that $\alpha = f_g(x, \infty)$, where $g(x)$ and $c(x)$ are zero for $x < 0$.

In the limit of small δ, the integrals in the above-given two equations each have the simplified form

$$\lim_{\delta \to 0} \int_{x-\delta}^{x+\delta} \eta(y)\,dy = 2\delta\eta(x), \quad \eta = g \text{ or } c. \tag{26}$$

So, one obtains

$$P_g(x) \equiv f_g(x, 0) = \frac{\alpha g(x)}{\alpha g(x) + (1-\alpha)c(x)}, \tag{27}$$

$$P_c(x) \equiv f_c(x, 0) = \frac{(1-\alpha)c(x)}{\alpha g(x) + (1-\alpha)c(x)}. \tag{28}$$

Here, $P_g(x)$ is the probability that the flash is a ground flash given that it has an MGA value equal to x. Similarly, $P_c(x)$ is the probability that the flash is a cloud flash given that it has an MGA value equal to x. Note from (27) and (28) that $P_g(x) + P_c(x) = 1$ as it should; that is, $P_c(x) = 1 - P_g(x)$ is the complimentary probability.

An alternate way to obtain (27) is to start with Bayes's theorem, which gives

$$p(t_g \,|\, x) = \frac{p(x \,|\, t_g)p(t_g)}{p(x)}. \tag{29}$$

Here, $p(x \,|\, t_g)$ is the probability density function of MGAs for ground flashes; that is, $p(x \,|\, t_g) = g(x)$, where t_g indicates flash type "ground." In addition, $p(t_g)$ is the probability of choosing a ground flash from the set of all N observed flashes, hence $p(t_g) = \alpha$. Finally, $p(x)$ is the probability of drawing an MGA value of x from the mixture of flashes; that is, $p(x) = m(x) = \alpha g(x) + (1 - \alpha)c(x)$, where the expression for $m(x)$ was derived in (2). Substituting these expressions into (29) gives

$$p(t_g \,|\, x) = \frac{\alpha g(x)}{\alpha g(x) + (1-\alpha)c(x)}. \tag{30}$$

But, $p(t_g \,|\, x)$ is the probability that the flash is a ground flash given that it has an MGA $= x$; that is, $p(t_g \,|\, x)$ is equivalent to $P_g(x)$ even though the notations differ. Hence, the result in (27) has once again been obtained.

To summarize, the APM first retrieves estimates of $[\alpha, g(x), c(x)]$ given by their discrete vector forms $(\alpha_r, \mathbf{g}_r, \mathbf{c}_r)$ as obtained from (17). These estimates are used to estimate $P_g(x)$ in (27). Each of the N flashes observed has a specific MGA value x. So, the APM-retrieved flash type probability and flash-typing conditions are

$$\begin{aligned} &P_{\mathrm{gr}}(x) = \frac{\alpha_r g_r(x)}{\alpha_r g_r(x) + (1-\alpha_r)c_r(x)}, \\ &P_{\mathrm{gr}}(x) > 0.5 \Rightarrow \text{ground flash}, \\ &P_{\mathrm{gr}}(x) \le 0.5 \Rightarrow \text{cloud flash}. \end{aligned} \tag{31}$$

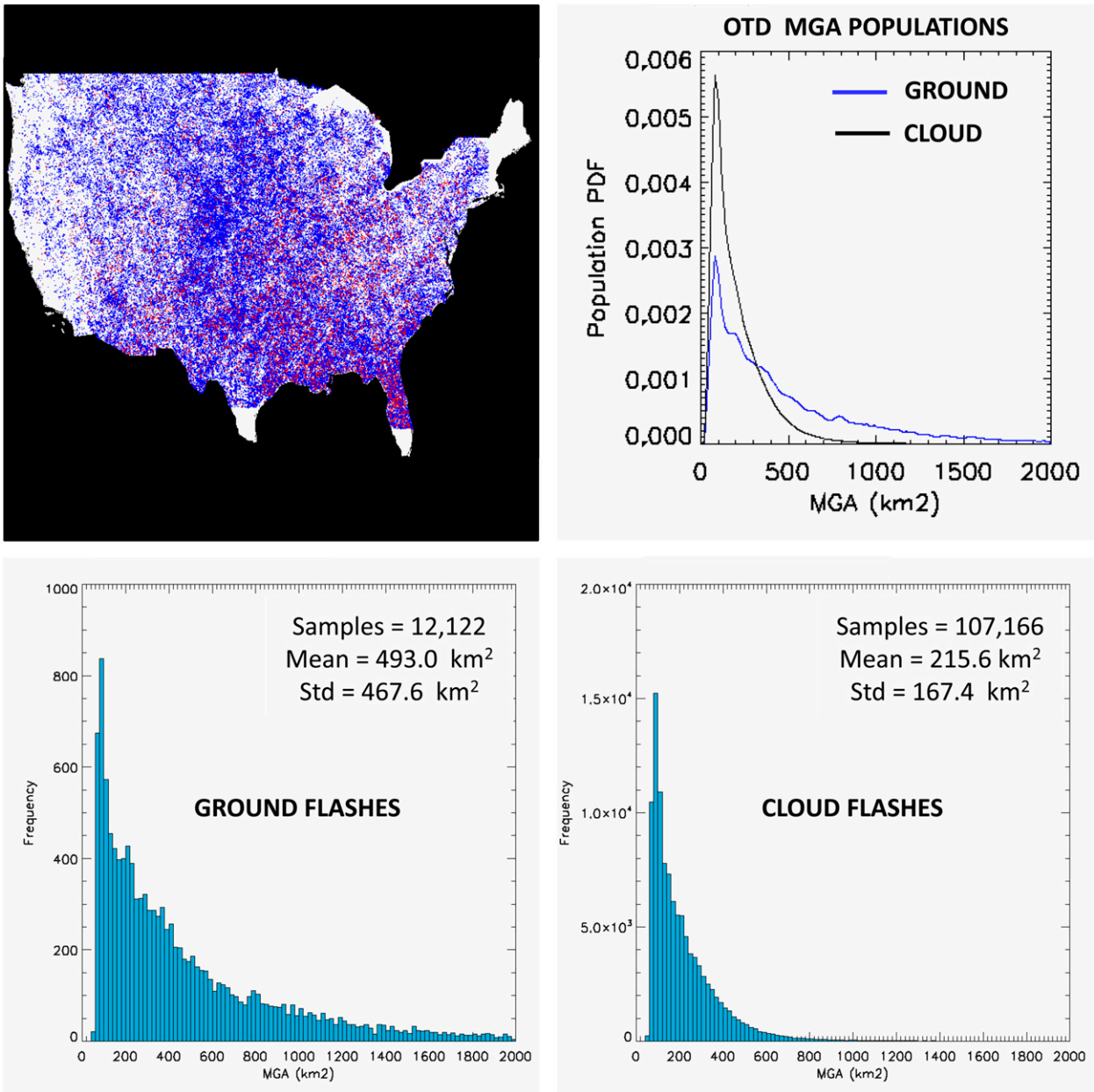

FIG. 4. Summary of OTD MGA analyses (adapted from Koshak 2010). (top left) Ground flash locations (red dots) and cloud flash locations (blue dots). (bottom left) Ground and (bottom right) cloud flash MGA distributions. (top right) High-resolution smoothed MGA pdfs derived from the OTD data.

3. Performance tests

To evaluate the performance of the APM, simulated retrievals are conducted. To accomplish this, the population densities ($\mathbf{g}_{\text{pop}}$, $\mathbf{c}_{\text{pop}}$) are created from which both the burn-in observations and the N operational measurements can be drawn.

To simulate realistic populations, we use the OTD MGA distributions obtained from Koshak (2010). The ground and cloud flash MGA distributions are shown in the bottom two panels of Fig. 4. These frequency distributions are interpolated to high resolution and smoothed using Interactive Data Language (IDL) utilities; they are then scaled by the respective sample sizes to obtain pdfs. The pdfs are shown in the top-right panel of Fig. 4.

Since LIS data are of generally higher quality than OTD, simulations are also conducted using population densities that are based on LIS MGA distributions

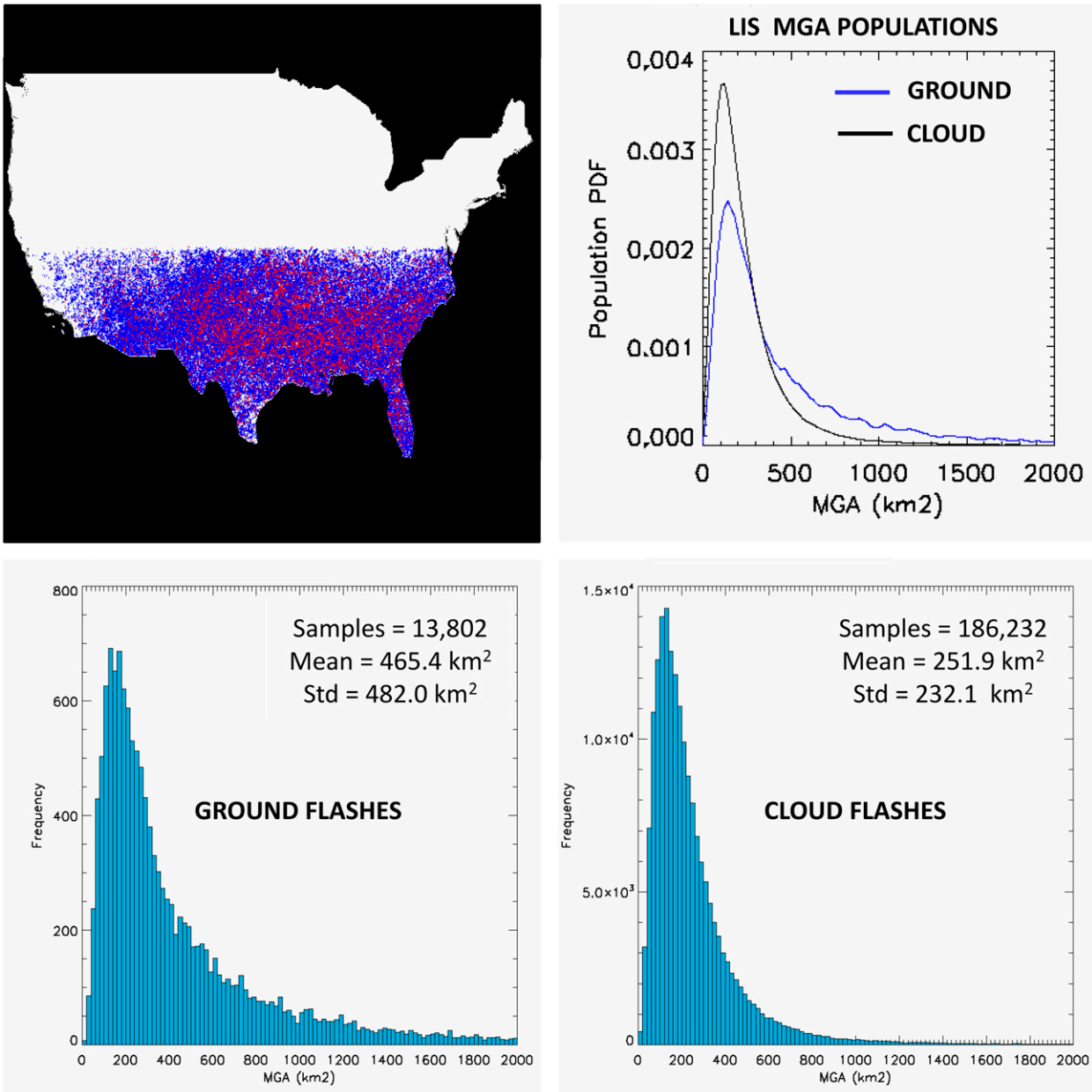

FIG. 5. As in Fig. 4, but for LIS MGA analyses (adapted from Koshak and Solakiewicz 2012).

obtained from Koshak and Solakiewicz (2012). Figure 5 provides the results for LIS in the same format as provided in Fig. 4. For all of our simulations, it was convenient and computationally expedient to use an MGA range from 0 to 2000 km^2, ignoring the few flashes that exist beyond this upper limit.

Figure 6 provides an overview of the retrieval error analysis associated with the simulation runs. By normalizing the ground and cloud flash MGA frequency distributions given in the bottom panels of Fig. 4 (or Fig. 5), one obtains the respective densities ($\mathbf{a}^*$, $\mathbf{b}^*$) shown in Fig. 6. Interpolating these densities to high resolution, and smoothing, gives the simulated population densities ($\mathbf{g}_{\mathrm{pop}}$, $\mathbf{c}_{\mathrm{pop}}$). These densities are then converted to cumulative distribution functions (CDFs).

To simulate the burn-in phase, a large number η_g of samples from the ground flash CDF are drawn; simulated random and systematic instrument measurement errors are added to each MGA sample. Similarly, a large number η_c of samples from the cloud flash CDF are drawn and instrument measurement errors added. This sampling produces a ground flash MGA distribution and

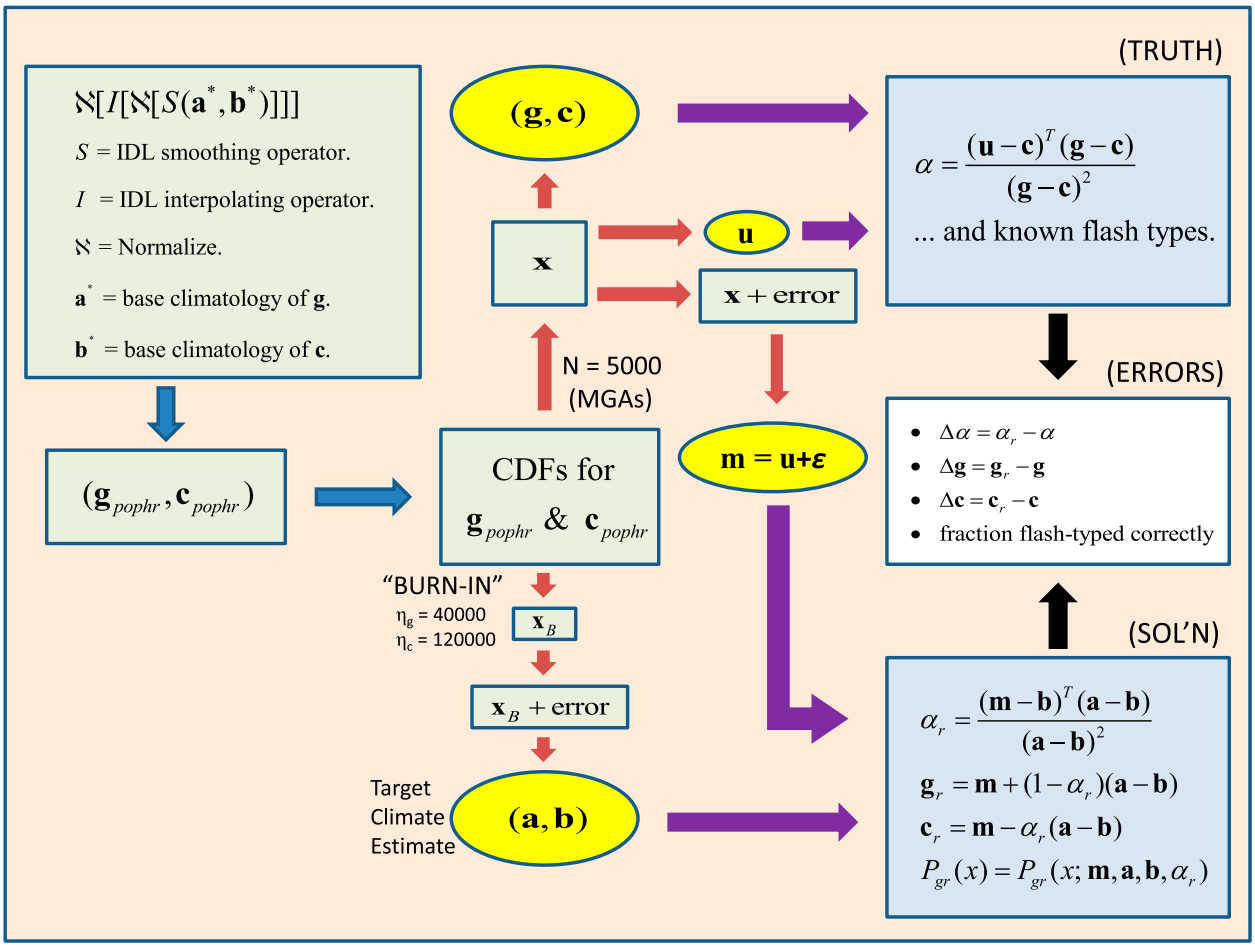

FIG. 6. Overview of the simulator used to evaluate APM performance. The error added to the MGA observations **x** consists of simulated random and systematic measurement errors described in the main text.

a cloud flash MGA distribution; scaling these distribution by η_g and η_c, respectively, produces the associated climate vector pdfs (**a**, **b**).

With the value of (**a**, **b**) given, one can now simulate the observation of N flashes and apply the APM to analyze the flashes. One is interested in the APM accuracy as a function of the true ground flash fraction α. Therefore, a total of 21 values of the true ground flash fraction (α = 0, 0.05, 0.10, 0.15, . . . , 0.95, 1.0) are examined. For each of these values of α, 100 trials (i.e., simulated APM retrievals) are performed. Each trial consists of uniquely drawing N_g samples from the ground flash CDF and N_c samples from the cloud flash CDF, adding the simulated random and systematic errors to these simulated MGAs, and then applying the APM to obtain the desired retrieved quantities. Here, $N = N_g + N_c$ and $\alpha = N_g/N$. Hence, a total of 2100 retrievals are performed for one analysis run (i.e., one error plot). All of our simulation runs were performed twice (i.e., once using OTD data to build the population densities and once using LIS data to build the population densities). For brevity, and because LIS data are generally of higher quality, the test results given below are based on the LIS data; one exception is the comparative analysis (see Fig. 9) discussed later.

Both random and systematic measurement errors are simulated. Simulated random measurement error added to each MGA value is chosen from a uniform distribution ranging from ($-R$ to $+R$) km^2, where R is chosen to be about twice the instrument nadir pixel footprint; the nadir footprints are approximately 16 km^2 (LIS) and 64 km^2 (OTD). Simulated systematic measurement error, S, is the negative systematic error associated with the finite instrument pixel footprint (taken as the nadir footprint values mentioned above). That is, each MGA value is reduced by truncation via the conversion $x' = \text{long}(x/\text{FP})\text{FP}$, where FP is the instrument footprint, and long is the IDL utility that converts its argument to an integer value. Hence, S ranges from 0 to -16 km^2 for LIS and 0 to -64 km^2 for OTD. In some of the LIS-based

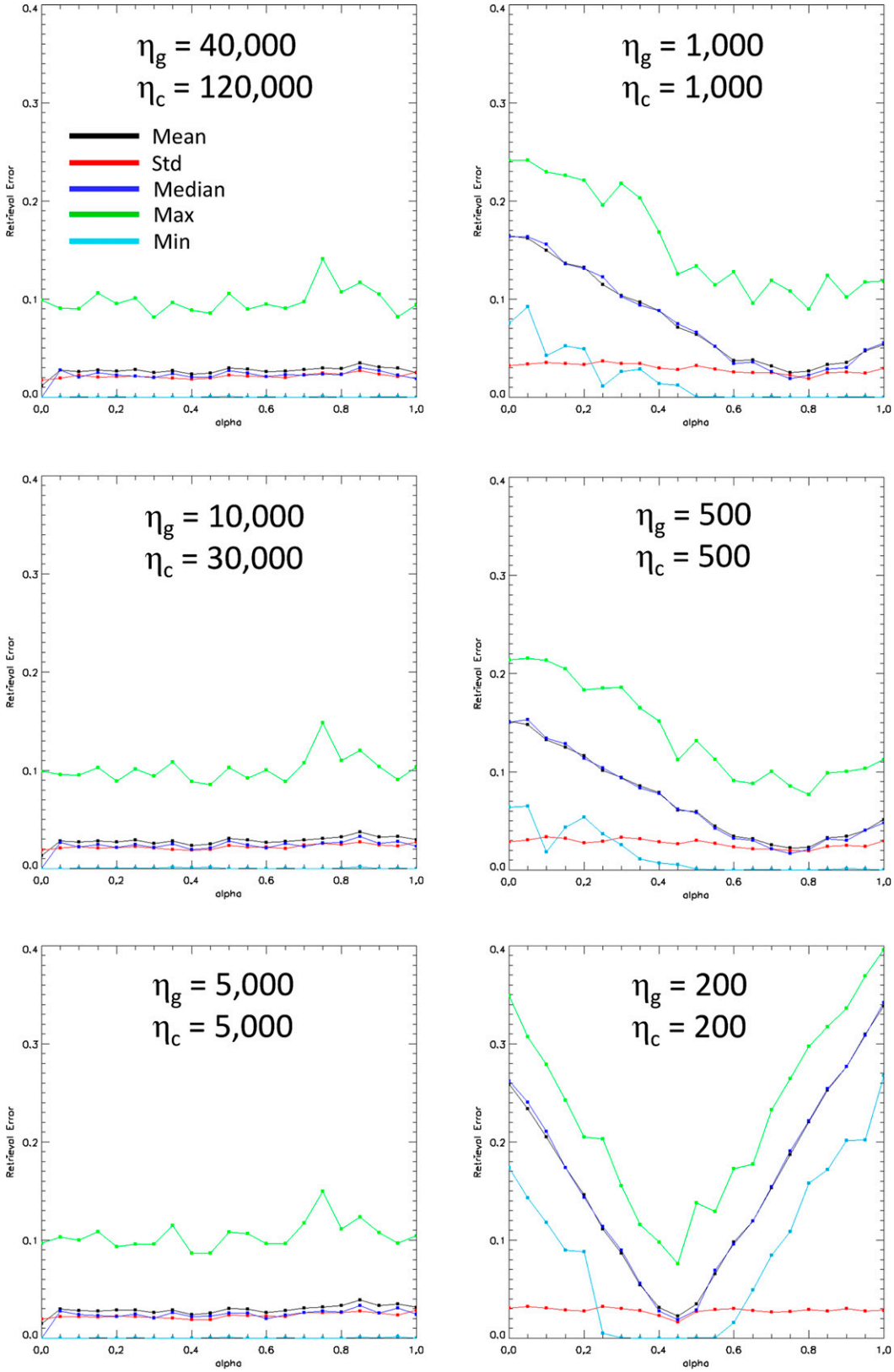

FIG. 7. The effect of reducing the sample size (η_g and η_c) in the burn-in phase.

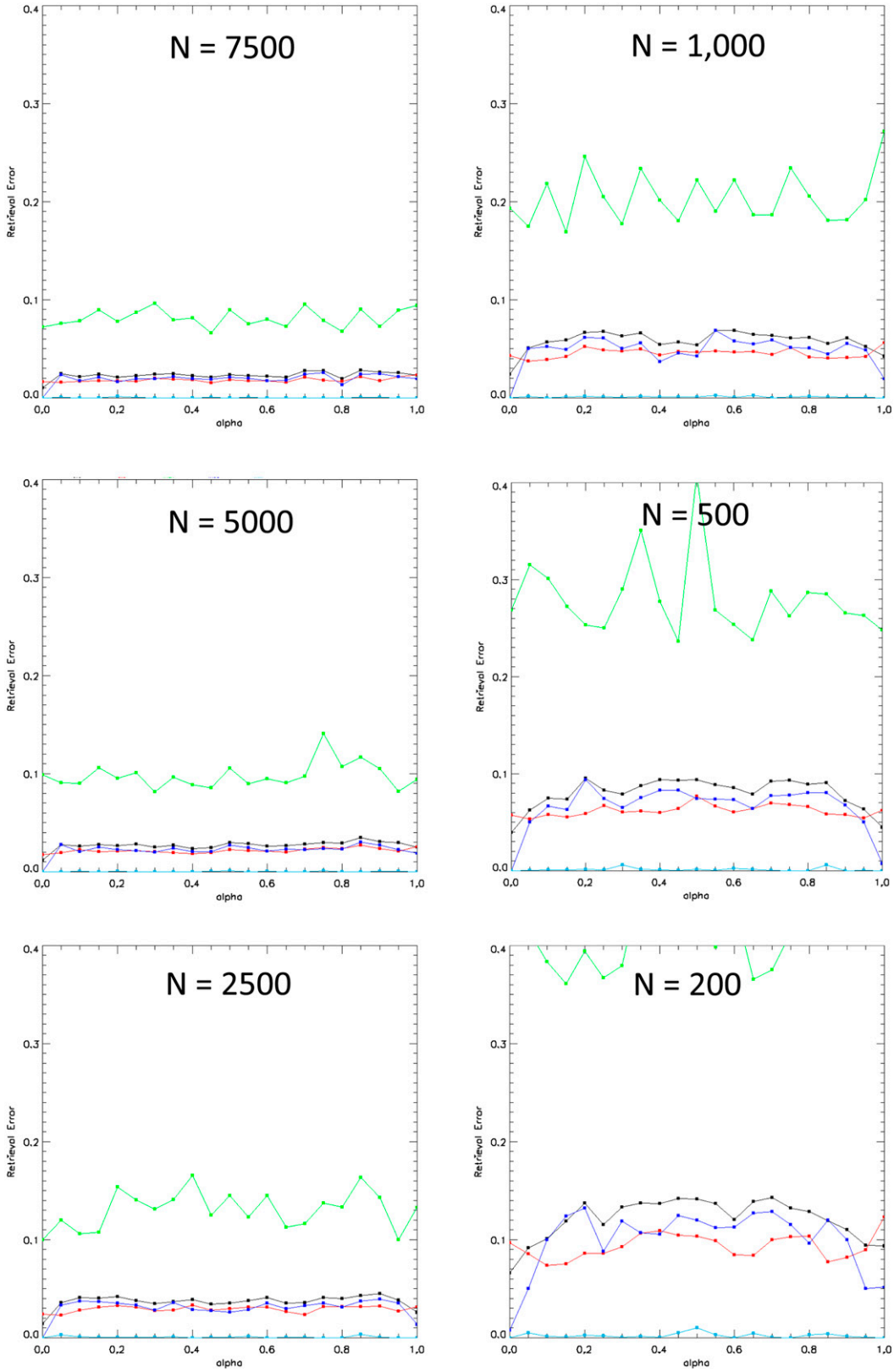

FIG. 8. The effect of reducing the number of flashes (N) observed. Color key as in Fig. 7.

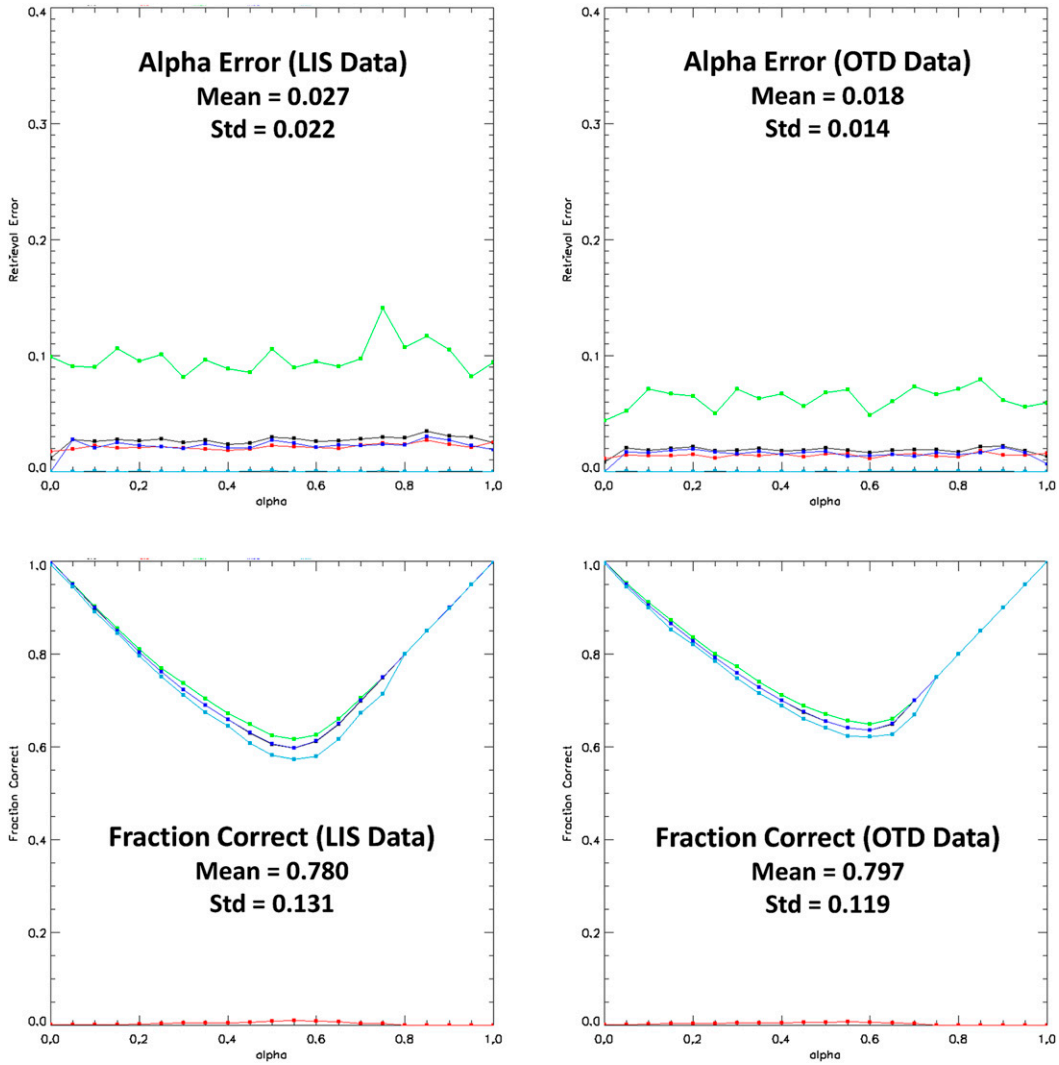

FIG. 9. Summary of APM performance using (left) LIS data to construct the MGA population and using (right) OTD data: (top) alpha error and (bottom) fraction correct. Color key as in Fig. 7. See text for additional details.

sensitivity tests to follow, the value of S takes on specific fixed values (i.e., -16, -8, -4, 4, 8, and 16 km^2).

The performance test results in Fig. 7 demonstrate how decreasing the sampling size in the burn-in phase degrades the APM solution accuracy. For these simulations, $N = 5000$, $R = 32\,\mathrm{km}^2$, and S is the negative systematic error due to the finite pixel footprint, as discussed above. Each error plot shows statistics of the ground flash fraction retrieval error ($\Delta\alpha = |\alpha_r - \alpha|$) as a function of the true ground flash fraction α. Curves for the mean of $\Delta\alpha$ (black curve), the standard deviation of $\Delta\alpha$ (red curve), the median of $\Delta\alpha$ (blue curve), and the maximum (green) and minimum (turquoise) of $\Delta\alpha$ are provided. Note that the retrieval errors are sufficiently small when the burn-in phase collects just $\eta_g = 5000$ ground flashes and $\eta_c = 5000$ cloud flashes. Degradation in the APM solutions occurs below these values, and is significant when 1000 or fewer are collected for each flash type.

The performance test results in Fig. 8 demonstrate how decreasing the value of N degrades the APM solution accuracy. For these simulations, R and S are as described in the previous paragraph, and the burn-in sample sizes are $\eta_g = 40\,000$, $\eta_c = 120\,000$. Note that there is not much return on investment in increasing N from 5000 to 7500, and that decreasing N below 2500 shows noticeable increases in solution retrieval error.

Figure 9 summarizes the performance of the APM when LIS data are used to generate the population densities (left column) and when OTD data are employed (right column). These simulations use the values $N = 5000$, $\eta_g = 40\,000$, and $\eta_c = 120\,000$. For the LIS data run, the values of R and S are as given above for

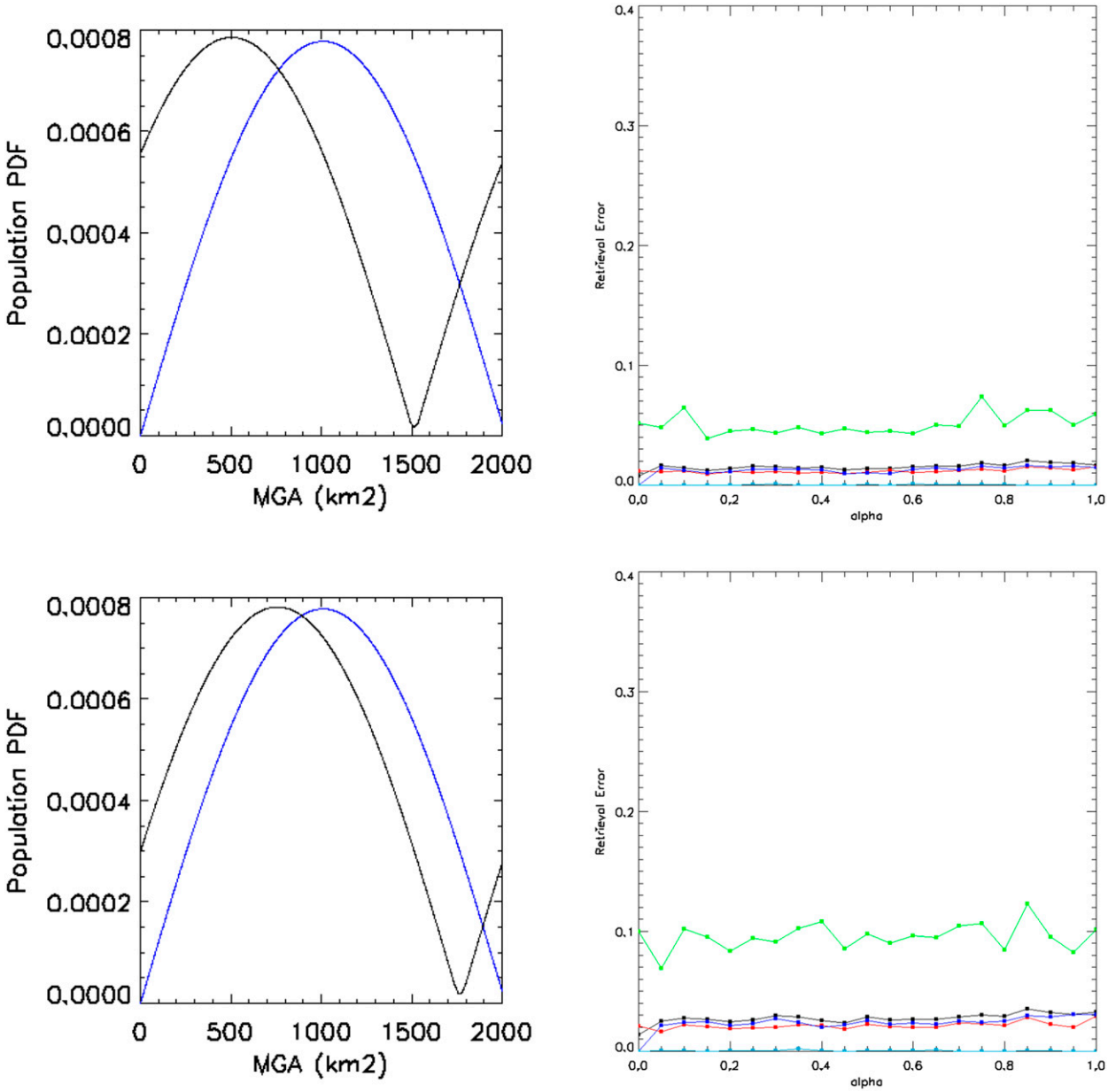

FIG. 10. The effect of reducing the distinction between the underlying population densities (the densities are shown in the left panel, one in black, the other in blue). Population densities having nothing to do with lightning were purposely employed to demonstrate that the APM is a general inversion technique. The right panel follows the color key as in Fig. 7.

Figs. 7 and 8, but for the OTD data run the value of $R = 64\,\mathrm{km}^2$ and S ranges from 0 to $64\,\mathrm{km}^2$. In addition to the statistics of $\Delta\alpha$, statistics of the fraction correctly flash typed are also provided (two bottom panels in Fig. 9). The results are very encouraging. The overall mean error in $\Delta\alpha$ is only 0.027 (LIS data run) and 0.018 (OTD data run). The associated average fraction of flashes correctly flash typed is 0.780 and 0.797, respectively. Even though the OTD data run involves larger values of R and S, it provides better retrievals. This is because the LIS ground and cloud flash MGA distributions overlap more than the OTD ground and cloud flash distributions (see bottom panels in Figs. 4 and 5). Specifically, the difference in the mean MGA is $277.4\,\mathrm{km}^2$ ($= 493.0 - 215.6\,\mathrm{km}^2$) for the OTD data but only $213.5\,\mathrm{km}^2$ ($= 465.4 - 251.9\,\mathrm{km}^2$) for the LIS data.

Figure 10 demonstrates that one can apply the APM method to arbitrary population densities that have nothing to do with lightning; that is, the APM is

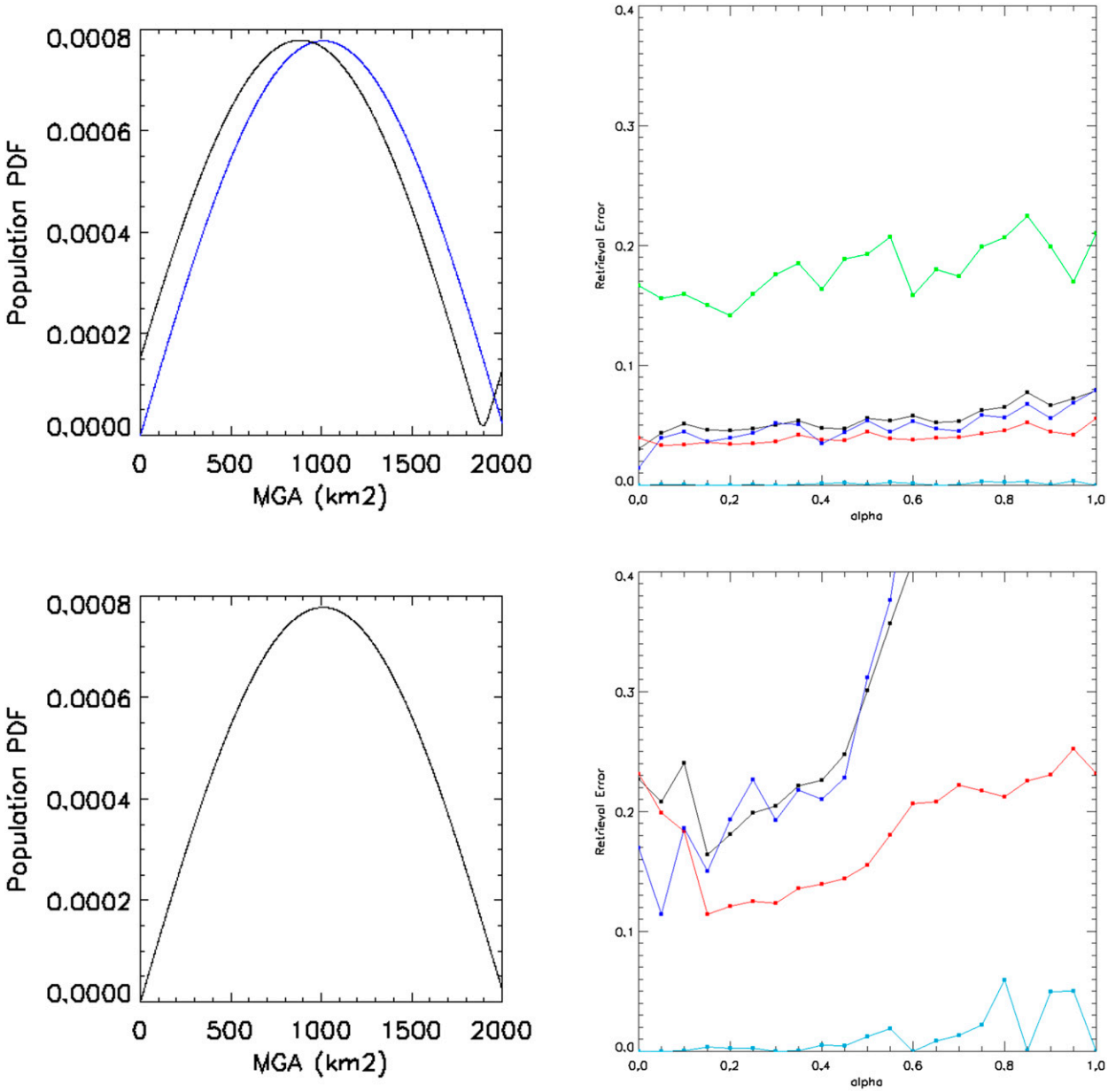

FIG. 10. (*Continued*)

a general inversion technique that can be applied to any inverse problem having a similar problem statement to the lightning problem studied here. In this particular test, the population densities are defined by phase-shifted absolute value sine functions that have been appropriately normalized. We have also mixed various exponential distributions with absolute value sine functions and achieved similar performance by the APM.

Figure 10 also demonstrates that if the population densities for the two species involved (i.e., ground and cloud flashes in the case of the lightning problem) become less distinct, the retrieval errors increase. If there is no distinction, then there is no information content and the inverse problem is ill posed.

Finally, Tables 1–4 provide additional simulation runs that evaluate the impact of the simulated measurement errors R and S. All of these performance tests employ $N = 5000$ flash observations. It is known that the burn-in approach is only approximate; that is, the climate vector pdfs (**a**, **b**) are only approximations to (**g**, **c**), respectively. Such approximations result in APM solution retrieval errors. Therefore, in order to conduct a true sensitivity test of the effects of R or S on APM solution retrieval errors, it is necessary to shut off the errors due

TABLE 1. The effect of random measurement errors when one sets $\mathbf{a} = \mathbf{g}$ and $\mathbf{b} = \mathbf{g}$ in the simulation. The quantities $\Delta\alpha$ and f are both dimensionless fractions (no units). See text for a description of the table headings.

Type	R (km^2)	Min	Avg	Std	Median	Max
$\Delta\alpha$	0	0.000	4.1×10^{-6}	2.1×10^{-5}	2×10^{-16}	2.2×10^{-4}
$\Delta\alpha$	8	0.000	0.007	0.006	0.006	0.046
$\Delta\alpha$	16	7.4×10^{-6}	0.010	0.009	0.007	0.060
$\Delta\alpha$	32	2.6×10^{-5}	0.022	0.015	0.019	0.113
$\Delta\alpha$	64	3.1×10^{-4}	0.078	0.032	0.079	0.184
f	0	0.000	0.736	0.203	0.750	1.000
f	8	0.585	0.783	0.129	0.769	1.000
f	16	0.588	0.782	0.129	0.767	1.000
f	32	0.575	0.780	0.130	0.764	1.000
f	64	0.511	0.769	0.138	0.754	1.000

to the burn-in phase approximation; that is, from the known value of $\mathbf{x}$ in the simulation, the values of $(\mathbf{g}, \mathbf{c})$ are extracted and used in the APM solution rather than the approximations $(\mathbf{a}, \mathbf{b})$.

Hence, for the simulation run associated with Table 1, the burn-in phase approximation is shut off, the value of $S = 0$, and the value of R is ramped up. Note that as the value of R increases, the error statistics (Min, Avg, Std, Median, Max) of $\Delta\alpha$ increase. The variable f in Table 1 is the fraction correctly flash typed. Because the APM flash typing is probabilistic in nature, there is less of a trend in the statistics of f as a function of R.

To obtain the true impact of ramping up the value of R, the burn-in phase approximation (with $\eta_g = 40\,000$, $\eta_c = 120\,000$) is turned back on, and $S = 0$ still holds. The results are provided in Table 2. Now, there is virtually no impact of varying R because these random errors are small in comparison to the errors associated with the burn-in phase approximation.

Similarly, Table 3 shuts off the burn-in phase approximation, sets $R = 0$, and varies S. A value of $S =$ nad16 denotes the variable negative systematic measurement error associated with the LIS finite pixel footprint (taken as the nadir value of 16 km^2). The error statistics of $\Delta\alpha$ increase as the magnitude of S increases, and the statistics of f are again more variable. Note that the negative values of S had a much more adverse impact on the minimum f than did the positive values of S.

The true impact of varying S is obtained by turning the burn-in phase approximation back on (with $\eta_g = 40\,000$, $\eta_c = 120\,000$) and setting $R = 0$ (see results in Table 4). Once again, there is virtually no impact of varying S because these systematic errors are small in comparison to the errors associated with the burn-in phase approximation.

TABLE 2. The effect of random measurement errors when one obtains $\mathbf{a}$ and $\mathbf{b}$ from the burn-in phase with $\eta_g = 40\,000$ and $\eta_c = 120\,000$.

Type	R (km^2)	Min	Avg	Std	Median	Max
$\Delta\alpha$	0	0.000	0.024	0.019	0.020	0.118
$\Delta\alpha$	8	0.000	0.024	0.019	0.020	0.118
$\Delta\alpha$	16	0.000	0.024	0.019	0.019	0.116
$\Delta\alpha$	32	0.000	0.024	0.019	0.020	0.115
$\Delta\alpha$	64	0.000	0.024	0.019	0.020	0.117
f	0	0.577	0.780	0.130	0.762	1.000
f	8	0.579	0.780	0.130	0.761	1.000
f	16	0.577	0.780	0.131	0.762	1.000
f	32	0.578	0.780	0.131	0.761	1.000
f	64	0.570	0.780	0.131	0.762	1.000

4. Mathematical error analysis

In this section we clarify exactly what retrieval errors $\Delta\alpha$ are expected given the percent burn-in errors $\Omega_a = 100|\mathbf{g} - \mathbf{a}|/|\mathbf{g}|$ and $\Omega_b = 100|\mathbf{c} - \mathbf{b}|/|\mathbf{c}|$. This is just a mathematical test to see how these burn-in errors mathematically map into $\Delta\alpha$. For this analysis, $N = 5000$, $R = 0$, $S = 0$, and the LIS data are used to obtain the population densities. The usual APM burn-in approach is not followed (i.e., the parameters η_g and η_c are not used or assigned any values). Instead, each of the 100 trials for a given known α is conducted by assigning each element of $\mathbf{a}$ as follows: $a_i = g_i + (2r - 1)\gamma$, where r is a uniform

TABLE 3. The effect of systematic measurement errors when one sets $\mathbf{a} = \mathbf{g}$ and $\mathbf{b} = \mathbf{g}$ in the simulation.

Type	S (km^2)	Min	Avg	Std	Median	Max
$\Delta\alpha$	−4	0.000	0.025	0.013	0.027	0.078
$\Delta\alpha$	−8	0.000	0.048	0.022	0.051	0.108
$\Delta\alpha$	−16	0.001	0.087	0.038	0.093	0.196
$\Delta\alpha$	4	0.000	0.027	0.014	0.029	0.076
$\Delta\alpha$	8	5.0×10^{-5}	0.056	0.024	0.059	0.158
$\Delta\alpha$	16	0.006	0.117	0.048	0.128	0.267
$\Delta\alpha$	nad16	0.000	0.059	0.048	0.049	0.275
f	−4	0.001	0.779	0.132	0.767	1.000
f	−8	0.012	0.771	0.136	0.762	1.000
f	−16	0.006	0.747	0.153	0.750	1.000
f	4	0.573	0.782	0.131	0.769	1.000
f	8	0.543	0.777	0.135	0.764	1.000
f	16	0.446	0.759	0.152	0.750	1.000
f	nad16	0.104	0.766	0.139	0.755	1.000

TABLE 4. The effect of systematic measurement errors when one obtains **a** and **b** from the burn-in phase with $\eta_g = 40\,000$ and $\eta_c = 120\,000$.

Type	S (km^2)	Min	Avg	Std	Median	Max
$\Delta\alpha$	−4	0.000	0.024	0.019	0.020	0.119
$\Delta\alpha$	−8	0.000	0.024	0.019	0.020	0.119
$\Delta\alpha$	−16	0.000	0.023	0.019	0.019	0.119
$\Delta\alpha$	4	0.000	0.024	0.019	0.020	0.121
$\Delta\alpha$	8	0.000	0.024	0.019	0.020	0.117
$\Delta\alpha$	16	0.000	0.024	0.019	0.020	0.119
$\Delta\alpha$	nad16	0.000	0.026	0.021	0.021	0.129
f	−4	0.577	0.780	0.130	0.761	1.000
f	−8	0.576	0.780	0.131	0.761	1.000
f	−16	0.576	0.780	0.130	0.761	1.000
f	4	0.578	0.780	0.130	0.761	1.000
f	8	0.576	0.780	0.130	0.761	1.000
f	16	0.578	0.780	0.130	0.761	1.000
f	nad16	0.574	0.780	0.131	0.761	1.000

random variable. An analogous expression is used to assign each element of **b**. Negative components are resampled, and the vectors are appropriately normalized. As before, each trial consists of $N = 5000$ observations that are randomly drawn from the population densities.

The test results are provided in Table 5. As γ increases, the values of $\Delta\alpha$, Ω_a, and Ω_b all increase, as expected.

Note that for $\gamma = 0.001$, the average of $\Delta\alpha$ is 0.020, which is very close to the average values shown for the nominal runs provided in Fig. 9. Moreover, the corresponding percent errors (Ω_a, and Ω_b) are both under 4%. This implies that the APM burn-in approach carried out for Fig. 9 is doing quite well. *That is, the burn-in approach is good enough to move the climate vectors* (**a**, **b**) *to within about 4% of the true pdfs* (**g**, **c**), *respectively.*

The results in Table 5 provide additional valuable insight. Suppose region A did not have independent observations available for conducting a burn-in phase but region B did. One wonders if the burn-in results from region B could be used for region A. Table 5 helps determine if this is feasible. For example, if one wanted to retrieve the ground flash fraction to within (on average) 0.09 in region A, then according to Table 5 a γ of about 0.003 would be needed. That is, the population densities in region A should not be more than about 10.4% different from the population densities in region B; that is, $\Omega_a = 10.40\%$ and $\Omega_b = 10.46\%$. Note from Table 5 that the ratio ave($\Delta\alpha$)/ave(Ω_a, Ω_b) is under about 0.009/1%.

5. Summary

An Analytic Perturbation Method (APM) was introduced for retrieving several important unknowns from satellite lightning mapper observations of N lightning flashes. The value of N must be large, typically in the thousands. Each lightning flash illuminates the cloud top, and a lightning mapper registers one or more optical groups per flash. The key lightning mapper

TABLE 5. The effect of percent errors (Ω_a, Ω_b) in the burn-in approximation on the ground flash fraction retrieval error.

Type	γ	Min	Avg	Std	Median	Max
$\Delta\alpha$	0.000	0.000	6×10^{-16}	5×10^{-16}	5×10^{-16}	3×10^{-15}
Ω_a	0.000	0.000	4×10^{-14}	4×10^{-14}	2×10^{-14}	3×10^{-13}
Ω_b	0.000	0.000	5×10^{-14}	4×10^{-14}	4×10^{-14}	2×10^{-13}
$\Delta\alpha$	0.001	1.2×10^{-5}	0.020	0.012	0.021	0.057
Ω_a	0.001	2.752	3.547	0.329	3.490	5.481
Ω_b	0.001	2.370	3.328	0.463	3.251	5.385
$\Delta\alpha$	0.002	2.5×10^{-5}	0.053	0.026	0.055	0.132
Ω_a	0.002	5.432	7.002	0.611	6.898	10.40
Ω_b	0.002	4.939	6.882	0.766	6.783	10.11
$\Delta\alpha$	0.003	3.4×10^{-6}	0.090	0.041	0.097	0.201
Ω_a	0.003	8.086	10.40	0.848	10.28	14.84
Ω_b	0.003	7.652	10.46	0.984	10.39	14.76
$\Delta\alpha$	0.004	7.4×10^{-5}	0.125	0.056	0.138	0.259
Ω_a	0.004	10.58	13.80	1.041	13.69	18.86
Ω_b	0.004	9.922	13.84	1.177	13.78	18.59
$\Delta\alpha$	0.005	1.6×10^{-5}	0.160	0.072	0.178	0.323
Ω_a	0.005	13.01	17.11	1.209	17.01	22.53
Ω_b	0.005	12.86	17.19	1.335	17.15	22.59
$\Delta\alpha$	0.006	8.2×10^{-5}	0.191	0.088	0.210	0.390
Ω_a	0.006	15.66	20.33	1.374	20.25	25.88
Ω_b	0.006	14.94	20.29	1.468	20.26	26.80

observation exploited by the APM is the maximum group area (MGA) of each of the N flashes. Hence, the APM mathematically inverts these N MGA values to determine the following: the ground flash fraction (i.e., the number of ground flashes divided by N), the probability density functions of the ground and cloud flash MGAs, and the flash type (ground or cloud) of each of the N flashes.

As part of the method, the lightning mapper first collects a large number ($\gtrsim$5000) of samples of both ground and cloud flashes from the region/season of interest that have already been flash typed using an independent validation dataset. This training period is called a ''burn in'' phase. Alternatively, if independent validation is not available, then the burn-in results from a different region–season can be tried, but with the expectation that APM solution retrieval error will possibly increase. Under the conditions of our analysis, every 1% error in the burn-in process appears to add about 0.009 or less to the APM mean ground flash fraction retrieval error. Long-term drift in the climate vectors (which are used to bias the retrieval results) can be mitigated by repeating the burn-in phase; a more aggressive approach would be to perform the burn-in phase on a regular basis. One could even consider using the APM results themselves to estimate upgrades to the climate vectors; that is, APM solutions obtained prior to ample drift in the climate vectors could in principal be accurate enough to update the climate vectors, thereby effecting a ''self-propagating'' system. This of course might only work up to a certain limit because of the accumulated effect of retrieval errors over a long period. In such a case, a new burn-in phase would be required to ''reset'' the system. Ultimately, the training period allows the retrieval algorithm to be applied autonomously to targets of interest during a designated operational phase.

A detailed simulator has been coded to perform extensive performance tests of the APM. The results are encouraging. The overall mean error in the retrieved ground flash fraction is only 0.027 (LIS data run) and 0.018 (OTD data run). The mean ground flash fraction errors as a function of the true ground flash fraction itself are under 0.04. The associated average percent of flashes correctly flash typed are 78.0% and 79.7%, respectively. Therefore, we believe the APM is a viable approach for inverting future GLM-observed flashes, especially over CONUS, where a burn-in phase can be accurately conducted using NLDN data. For other regions, Vaisala Inc.'s global lightning dataset (GLD360) ground flash data could be employed to complete the burn-in phase. Or, the CONUS burn-in results could be used, but with possible increases in solution retrieval error.

Acknowledgments. This research has been supported by the NOAA GOES-R Risk Reduction Program (managed by Ms. Ingrid Guch and Dr. Mark DeMaria), and by the Lightning Imaging Sensor (LIS) project (program manager Ramesh Kakar, NASA Headquarters) as part of the NASA Earth Science Enterprise (ESE) Earth Observing System (EOS) project. Special thanks to Dr. Steve Goodman, senior (chief) scientist, GOES-R System Program, for his guidance throughout this work effort.

REFERENCES

Christian, H. J., R. J. Blakeslee, and S. J. Goodman, 1992: Lightning Imaging Sensor for the Earth Observing System. NASA Tech. Memo. NASA TM-4350, 44 pp.

——, K. T. Driscoll, S. J. Goodman, R. J. Blakeslee, D. M. Mach, and D. E. Buechler, 1996: The Optical Transient Detector (OTD). *Proceedings of the 10th International Conference on Atmospheric Electricity,* ICAE, 368–371.

——, and Coauthors, 1999: The Lightning Imaging Sensor. 11th International Conference on Atmospheric Electricity, NASA Conf. Proc. NASA/CP-1999-209261, 746–749.

——, and Coauthors, 2003: Global frequency and distribution of lightning as observed from space by the Optical Transient Detector. *J. Geophys. Res.,* **108,** 4005, doi:10.1029/2002JD002347.

Cummins, K. L., and M. J. Murphy, 2009: An overview of lightning locating systems: History, techniques, and data uses, with an in-depth look at the U.S. NLDN. *IEEE Trans. Electromagn. Compat.,* **51,** 499–518, doi:10.1109/TEMC.2009.2023450.

Goodman, S. J., and Coauthors, 2013: The GOES-R Geostationary Lightning Mapper (GLM). *Atmos. Res.,* **125–126,** 34–49, doi:10.1016/j.atmosres.2013.01.006.

Hutchins, M. L., R. H. Holzworth, J. B. Brundell, and C. J. Rodger, 2012: Relative detection efficiency of the World Wide Lightning Location Network. *Radio Sci.,* **47,** RS6005, doi:10.1029/2012RS005049.

Koshak, W. J., 2010: Optical characteristics of OTD flashes and the implications for flash-type discrimination. *J. Atmos. Oceanic Technol.,* **27,** 1822–1838, doi:10.1175/2010JTECHA1405.1.

——, 2011: A mixed exponential distribution model for retrieving ground flash fraction from satellite lightning imager data. *J. Atmos. Oceanic Technol.,* **28,** 475–492, doi:10.1175/2010JTECHA1438.1.

——, and R. J. Solakiewicz, 2011: Retrieving the fraction of ground flashes from satellite lightning imager data using CONUS-based optical statistics. *J. Atmos. Oceanic Technol.,* **28,** 459–473, doi:10.1175/2010JTECHA1408.1.

——, and ——, 2012: The ground flash fraction retrieval algorithm employing differential evolution: Simulations and applications. *NOAA Satellite Science Week Meeting,* Kansas City, MO, NOAA. [Available online at http://www.goes-r.gov/downloads/2012-Science-Week/poster-abstracts-tues/06_Koshak.pdf.]

Mach, D. M., H. J. Christian, R. J. Blakeslee, D. J. Boccippio, S. J. Goodman, and W. L. Boeck, 2007: Performance assessment of the Optical Transient Detector and Lightning Imaging Sensor. *J. Geophys. Res.,* **112,** D09210, doi:10.1029/2006JD007787.

Rudlosky, S. D., and D. T. Shea, 2013: Evaluating WWLLN performance relative to TRMM/LIS. *Geophys. Res. Lett.,* **40,** 2344–2348, doi:10.1002/grl.50428.