# The Ground Flash Fraction Retrieval Algorithm Employing Differential Evolution: Simulations and Applications


William Koshak[1] and Richard Solakiewicz[2]

[1]Earth Science Office, ZP11, NASA Marshall Space Flight Center, Robert Cramer Research Hall, 320 Sparkman Drive, Huntsville, AL 35805.
[2]Department of Mathematics & Computer Science, Chicago State University, 9501 South King Drive, Chicago, IL 60628.


## 1. GOAL OF R3 PROJECT

When viewing lightning from space at optical wavelengths, the cloud multiple scattering medium obscures the view thereby preventing one from easily determining what flashes strike the ground. However, recent studies have made some progress in estimating the ***ground flash fraction***, $\alpha$ in a set of $N$ flashes observed from space [Koshak (2010), Koshak and Solakiewicz (2011), and Koshak (2011)]. Knowledge of $\alpha$ is important for better understanding:

- Severe Weather,
- Lightning Nitrogen Oxides Chemistry/Climate Studies, and
- Global Electric Circuit.

In the study by Koshak (2011), a Bayesian inversion method was introduced for retrieving the fraction of ground flashes in a set of flashes observed from a (low earth orbiting or geostationary) satellite lightning imager. This method has formed the basis of a Ground Flash Fraction Retrieval Algorithm (GoFFRA) that is being tested as part of GOES-R Geostationary Lightning Mapper (GLM) risk reduction. **Figure 1** summarizes the basic question addressed in this study.

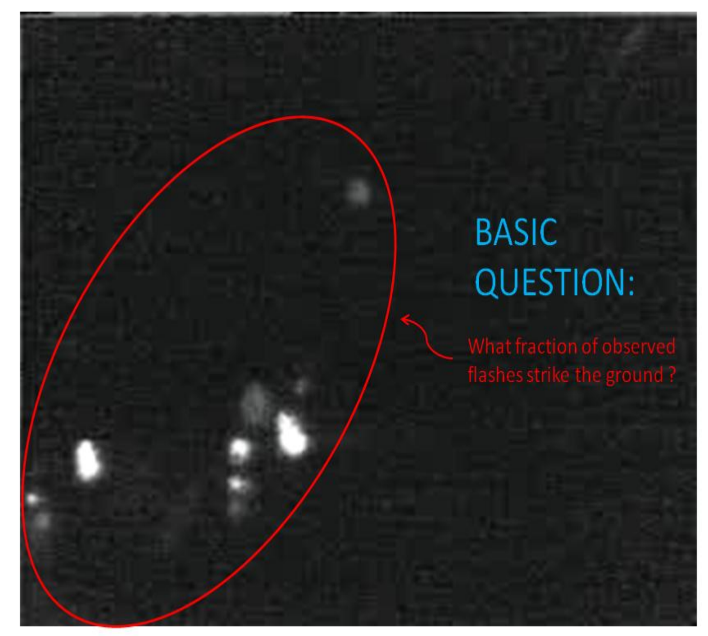


Figure 1. Example of lightning optical emissions as seen from space (STS-52 video).

## 2. THE BAYESIAN RETRIEVAL

**Figure 2** highlights the mathematical attributes of the Bayesian retrieval process. The $\mu$'s are the population mean Maximum Group Areas (MGAs) decremented by the nadir footprint area of the space-based lightning imager employed. In summary, the distribution of $y$ (nadir footprint decremented MGA) is assumed to follow a mixed exponential distribution; one finds the 3 parameters of this distribution that maximizes a scalar function which is a product of probabilities.

### Bayesian Inversion

Bayes' Law:

$$P(\alpha,\mu_g,\mu_c|\mathbf{y}) = \frac{P(\mathbf{y}|\alpha,\mu_g,\mu_c)P(\alpha,\mu_g,\mu_c)}{P(\mathbf{y})} \ ,$$

Find parameters $\mathbf{v} = (\alpha,\mu_g,\mu_c)$ that maximize the probability on LHS.

This means one maximizes the following :

$$S(\mathbf{v}) \equiv \ln\left[P(\mathbf{y}|\mathbf{v})P(\mathbf{v})\right] = \ln\prod_{i=1}^{m} p(y_i|\mathbf{v}) + \ln P(\mathbf{v}) = \sum_{i=1}^{m}\ln\left[\frac{\alpha}{\mu_g}e^{-y_i/\mu_g} + \frac{(1-\alpha)}{\mu_c}e^{-y_i/\mu_c}\right] + \ln P(\mathbf{v}) \ ,$$

Formally :

$$\frac{\partial S(\mathbf{v})}{\partial \mathbf{v}} = \mathbf{0} \quad \Rightarrow \quad \mathbf{v} = \text{"Maximum A Posteriori (MAP) Solution"}$$

Practically :

Use Broyden-Fletcher-Goldfarb-Shannon variant of Davidon-Fletcher-Powell numerical method to minimize $-S(\mathbf{v})$. Also, $P(\mathbf{v})$ is simplified by assuming model parameter independence, with $P(\alpha)$ uniform, and $P(\mu_g)$ & $P(\mu_c)$ both normal distributions.

### Grobner Initialization

$$\alpha = \frac{\overline{y} - \mu_c}{\mu_g - \mu_c}$$

$$\mu_g = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

$$\mu_c = \frac{-B + \sqrt{B^2 - 4AC}}{2A}$$

*where :*

$$A = \overline{y}^2 - q, \quad B = r - \overline{y}q, \quad C = q^2 - \overline{y}r$$

$$q = \tfrac{1}{2}(\overline{y}^2 + s^2)$$

$$r = \tfrac{1}{6}(\overline{y}^3 + 3\overline{y}s^2 + \gamma_1 s^3)$$

$$(\overline{y}, s^2, \gamma_1) = \text{Sample (mean, variance, skewness)}$$

**Figure 2. Left slide summarizes the Bayesian retrieval approach. To obtain the optimum parameters, the optimization begins by using an initial guess of the solution, such as the "Grobner Initialization" (right slide).**

## 3. ACCOMPLISHMENTS

In addition to developing the Bayesian formalism for the retrieval problem, introducing the use of the mixed exponential model, and deriving the Grobner Initialization, we have conducted preliminary tests of the GoFFRA on simulated datasets. This allowed us to uncover solution errors due to Label Switching (LS) and Parameter Identity Theft (PIT). The LS problem is well known in the literature on mixed exponential distributions, and the PIT problem was discovered in this study. Each problem occurs when one allows the numerical minimizer to freely roam through the parameter search space; this allows certain solution parameters to "interchange roles" which leads to fundamental ambiguities, and solution error. A major accomplishment of this study is that we have employed a state-of-the-art genetic-based global optimization algorithm called Differential Evolution (DE) that constrains the parameter search in such a way as to remove both the LS and PIT problems. We have also applied the GoFFRA/DE method to both simulated and actual MGA datasets (see following section) to evaluate algorithm performance.

## 4. PERFORMANCE TESTS

**a. Simulated Data** To test the performance of the GoFFRA when DE is employed, we applied it to analyze simulated MGA datasets that we generated from known mixed exponential distributions. The solution retrieval errors are provided in **Figure 3**. The retrieval errors are shown as a function of the true ground flash fraction (alpha) of the mixture; vertical lines are standard deviations about the mean retrieval errors. The algorithm performs quite well.

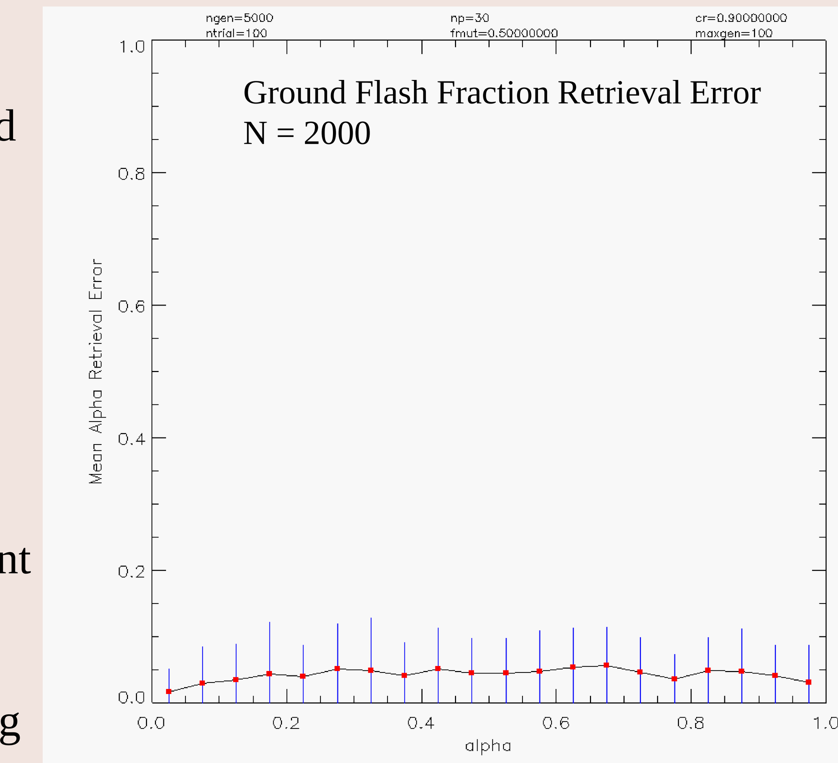


**Figure 3. Retrieval errors for simulated data.**

**b. OTD Data** Moreover, we evaluated the GoFFRA/DE method by applying it to analyze actual MGAs derived from the low-Earth orbiting lightning imager called the Optical Transient Detector (OTD). The actual MGA data were classified as either ground or cloud flash MGAs using National Lightning Detection Network™ (NLDN) data. The ground flash fraction retrieval error plots are provided in **Figure 4**. Errors increase for large alpha (~ 0.7) indicating a breakdown in the mixed exponential distribution model for the ground flash MGAs. This is an important finding, as it requires us to upgrade the form of the model to further reduce the retrieval errors.

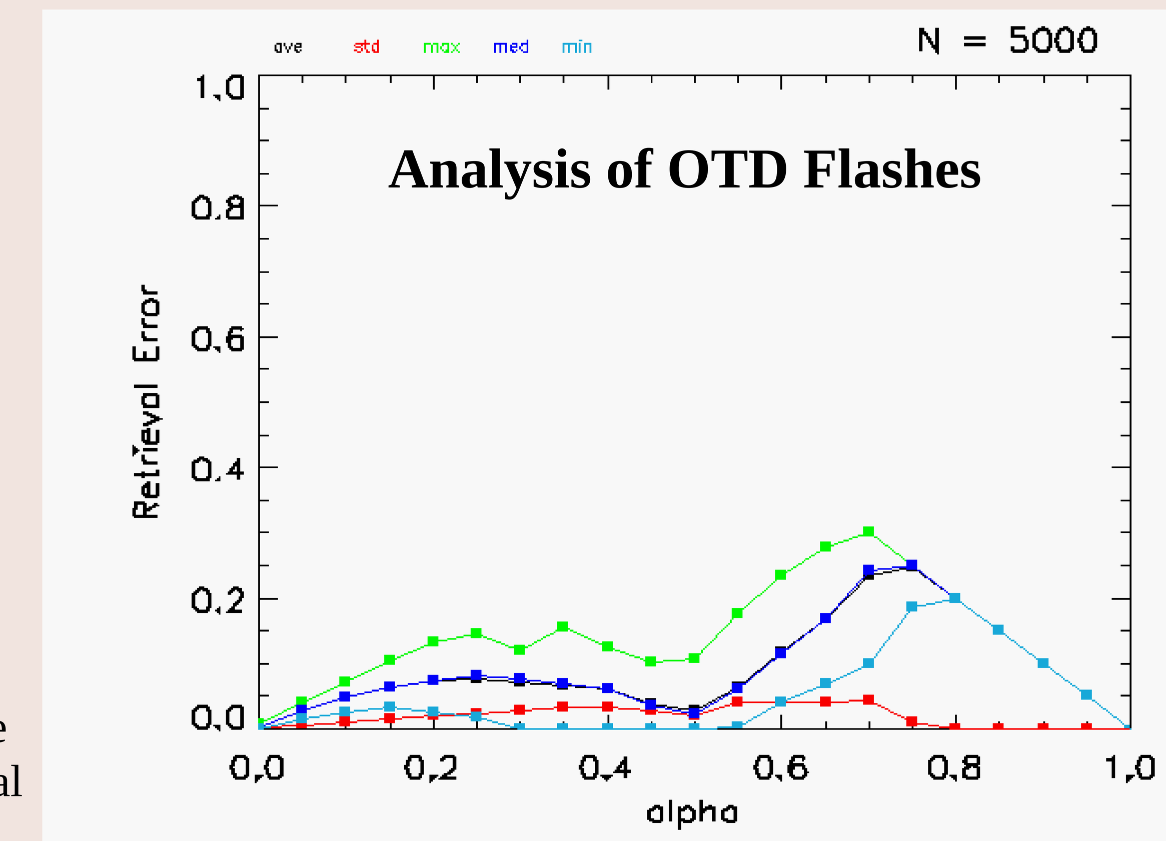


**Figure 4. Retrieval errors using mixtures of OTD ground & cloud flashes.**

**c. LIS Data** We have also made significant progress in creating test data sets based on LIS data, and evaluating algorithm performance. To create the test dataset, we inter-compared 9 years (2003-2011) of LIS and NLDN over CONUS and thereby categorized each LIS flash as either a ground or cloud flash. **Figure 5** shows the location of the LIS ground (red) and LIS cloud (blue) flashes over the US. The distributions of the LIS ground flash and cloud flash MGAs are the remaining figures in Figure 5. Application of the GoFFRA algorithm to mixtures of the LIS ground and cloud flashes produced even larger retrieval errors than for OTD, with errors again increasing for large values of alpha. Even though the ground flash fraction in nature is normally small (i.e., where our algorithm performs the best), we want to upgrade the mixed exponential distribution model to reduce retrieval error across the entire alpha domain.

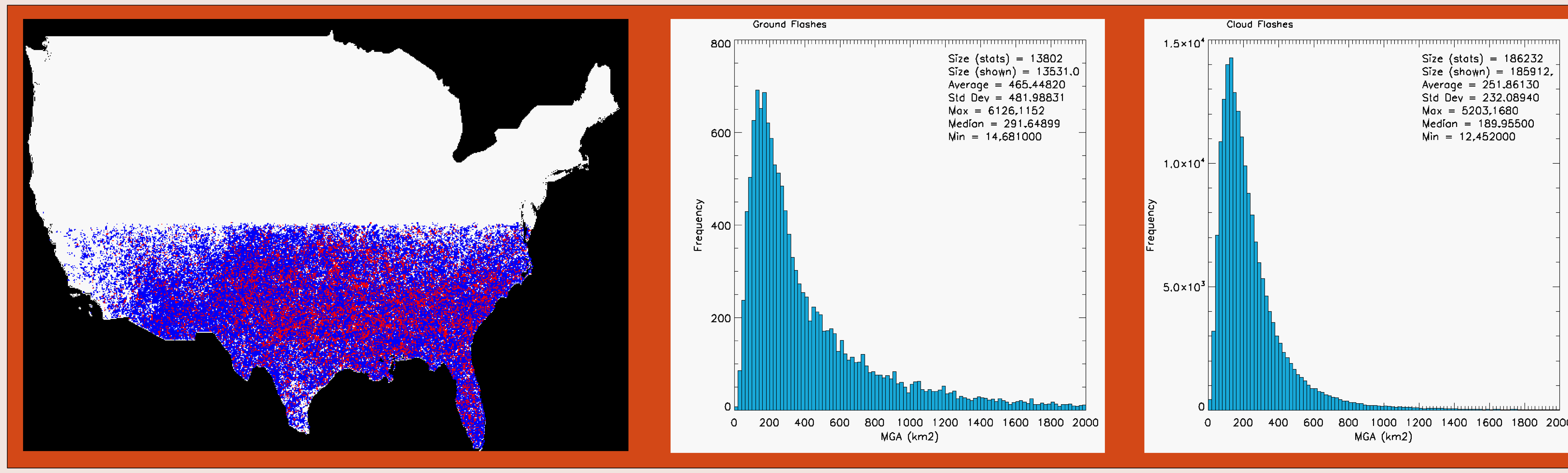


**Figure 5. The new test dataset created by comparing 9 yrs of LIS and NLDN data. Leftmost plot is the locations of LIS ground (red) & cloud (blue) flashes.**

- ❑ **Accomplishments Anticipated @ end of Year 2:** We will upgrade/replace the mixed exponential model, perform retests, and show reduced retrieval errors.
- ❑ **Initial Feedback from Potential Users:** Given our results above, we anticipate the start of getting feedback regarding specific GoFFRA applications.
- ❑ **What Could be Accomplished w/3rd Year Funding:** Completion of additional error analyses showing robustness of algorithm; more elaborate demos.

## 5. REFERENCES


Koshak, W. J., Optical Characteristics of OTD Flashes and the Implications for Flash-Type Discrimination, J. Atmos. Oceanic Technol., 27, 1822-1838, 2010.
Koshak, W. J., R. J. Solakiewicz, Retrieving the Fraction of Ground Flashes from Satellite Lightning Imager Data Using CONUS-Based Optical Statistics, J. Atmos. Oceanic Technol., 28, 459-473, 2011.
Koshak, W. J., A Mixed Exponential Distribution Model for Retrieving Ground Flash Fraction from Satellite Lightning Imager Data, J. Atmos. Oceanic Technol., 28, 475-492, 2011.


GOES-R3 Annual Meeting, Kansas City, MO, April 30-May 4, 2012.