Obtaining the Grobner Initialization for the Ground Flash Fraction Retrieval Algorithm

R. Solakiewicz[1], R. Attele[1], W. Koshak[2]

[1]Department of Mathematics & Computer Science, Chicago State University, 9501 South King Drive, Chicago, IL 60628.

[2]Earth Science Office, VP61, NASA Marshall Space Flight Center, Robert Cramer Research Hall, 320 Sparkman Drive, Huntsville, AL 35805.

1. BACKGROUND

At optical wavelengths and from the vantage point of space, the multiple scattering cloud medium obscures one's view and prevents one from easily determining what flashes strike the ground. However, recent investigations have made some progress examining the (easier, but still difficult) problem of estimating the *ground flash fraction* in a set of N flashes observed from space [Koshak (2010), Koshak and Solakiewicz (2011), and Koshak (2011)].

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. The method employed a *constrained mixed exponential distribution model* to describe the lightning optical measurements. To obtain the optimum model parameters, a scalar function of three variables (one of which is the ground flash fraction) was minimized by a numerical method. This method has formed the basis of a Ground Flash Fraction Retrieval Algorithm (GoFFRA) that is being tested as part of GOES-R GLM risk reduction.

Figure 1 summarizes the basic functionality of the GoFFRA, and Figure 2 highlights the mathematical attributes of the Bayesian retrieval process.

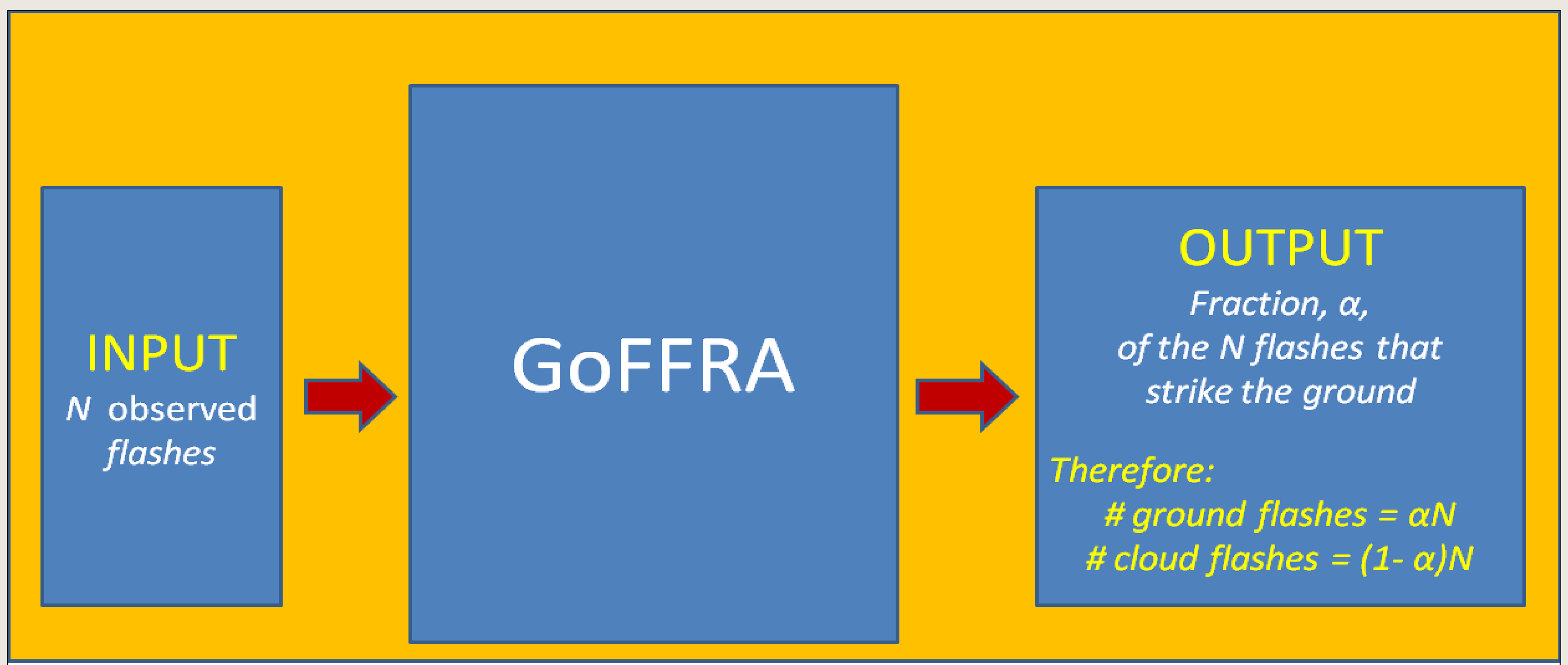

Figure 1. The basic function of the Ground Flash Fraction Retrieval Algorithm (GoFFRA). The satellite lightning imager (e.g., OTD, LIS, or GLM) observes N flashes during time period Δt; the algorithm outputs the fraction of those N flashes that are ground flashes. The sample size N must be sufficiently large (typically 2000 or larger) to keep retrieval errors small.

Mixed Exponential Distribution Model

Basic Definitions:

x = Maximum Group Area (MGA) in a flash

$y = x - 64km^2$ (Shifted MGA)

N_g = # Ground flashes in lat/lon bin of interest

N = # Flashes in lat/lon bin of interest

$\alpha = N_g / N$ (Ground flash fraction)

μ_g = Population mean y for ground flashes (in lat/lon bin of interest)

μ_c = Population mean y for cloud flashes (in lat/lon bin of interest)

Distribution of MGA modeled as a Mixed Exponential Distribution:

$$p(y) = \alpha p_g(y) + (1-\alpha)p_c(y) = \frac{\alpha}{\mu_g}e^{-y/\mu_g} + \frac{(1-\alpha)}{\mu_c}e^{-y/\mu_c}, \quad y \geq 0 \quad .$$

Population Means of y:

$$\mu_g \equiv \int_0^\infty y p_g(y)dy, \quad \mu_c \equiv \int_0^\infty y p_c(y)dy \ .$$

Require that:

$$\mu_g > \mu_c$$

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.

Estimative Initialization Scheme

$$\alpha_{initial} = 0.5$$

$$(\mu_g)_{initial} = \bar{y} + \sqrt{\frac{1}{2}\left(s^2 - \bar{y}^2\right)}$$

$$(\mu_c)_{initial} = \bar{y} - \sqrt{\frac{1}{2}\left(s^2 - \bar{y}^2\right)}$$

Figure 2. The upper left slide defines basic variables, and the upper right slide summarizes the Bayesian retrieval approach. To obtain the optimum parameters, the minimization begins by initializing the parameters as shown in the bottom slide [see Koshak (2011) for additional details on all of these slides**].**

2. THE GROBNER INITIALIZATION

Note that the estimative initialization scheme provided in Figure 2 involves only two moments (the mean and the standard deviation); it also just initializes the ground flash fraction (alpha) to it's centerline value 0.5. By including the third moment, (skewness, γ_1) of the lightning optical characteristic, we arrive at a set of 3 polynomials as shown below:

Set of 3 Polynomial Equations in 3 Unknowns (α,μ_g,μ_c) :

$$\alpha\mu_g + (1-\alpha)\mu_c = \mu$$

$$\alpha\mu_g^2 + (1-\alpha)\mu_c^2 = \tfrac{1}{2}\left(\mu^2 + \sigma^2\right)$$

$$\alpha\mu_g^3 + (1-\alpha)\mu_c^3 = \tfrac{1}{6}\left(\mu^3 + 3\mu\sigma^2 + \gamma_1\sigma^3\right)$$

Solving this system without any guess-work can be accomplished using Grobner bases. [For perspective, if the equations were linear the Grobner bases method would reduce to Gaussian Elimination common in linear algebra.] We employ *Mathematica* to find the Grobner bases of this system; the *Mathematica* utility is called **GroebnerBasis** which uses an efficient version of the *Buchberger algorithm* to compute the polynomial bases. We obtain a total of 11 polynomials that define the Grobner bases. Of these, we pick the three easiest to solve, which are:

3 (of 11) Grobner Bases:

$$\alpha\mu_g + (1-\alpha)\mu_c = \mu$$

$$(\mu_c - \mu)\mu_g + (q - \mu\mu_c) = 0$$

$$(\mu^2 - q)\mu_c^2 + (r - \mu q)\mu_c + (q^2 - \mu r) = 0$$

where :

$$q = \tfrac{1}{2}(\mu^2 + \sigma^2)$$

$$r = \tfrac{1}{6}(\mu^3 + 3\mu\sigma^2 + \gamma_1\sigma^3)$$

$$(\mu, \sigma^2, \gamma_1) = (\text{mean, variance, skewness})$$

This set of 3 equations has a solution (which is also a solution to the original set of polynomials) given by:

Solution (= Grobner Initialization):

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

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

$$\alpha = \frac{\mu - \mu_c}{\mu_g - \mu_c}$$

where :

$$A = \mu^2 - q, \quad B = r - \mu q, \quad C = q^2 - \mu r$$

This represents an analytic solution. It will be used to replace the estimative initialization scheme discussed in Figure 2. We expect it to improve the Bayesian retrieval results.

3. 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.