The Exception-Maximization Algorithm and its Application in Quantitative Remote Sensing Inversion*

Hong Shi[1], Hongxia Cui[2], Hua Yang[2], Xiaowen Li[2,3], Jinbao Liu[2]

1. Department of Mathematics, Beijing Normal University, Beijing, 100875, China
2. Research Center for Remote Sensing and GIS, School of Geography, Beijing Normal University, State key Laboratory of Remote Sensing Science, and Beijing Key Laboratory for Remote Sensing of Environment and Digital Cities Beijing, 100875, China
3. Department of Geography and Center for Remote Sensing, Boston University, Boston

Abstract—**In remote sensing inversion, we always assume that the observed data error distribution is normal distribution for simplifying the calculation. But under this assumption, only if a few observed data have big error, the inversion result will become instable. In this paper, we try to use Expectation-Maximization (EM) algorithm to get more precise and robust inversion result based on another statistical distribution.**

Linear kernel-driven model with t-distribution error solved by EM algorithm is used to prove this new idea. The inversion methods include traditional ML estimate without prior distribution information of inversion parameters and Bayesian inversion based on prior normal distribution. The test about robustness showed that under the assumption of t-distribution error, more than or over half of observed data have big error can cause instability of inversion results.

Keywords-linear kernel- driven model; normal distribution; t distribution; Exception-Maximization algorithm

I. BACKGROUND

The Expectation-Maximization (EM) algorithm, firstly proposed by Dempster in 1977 [1], is broadly applicable approach to the iterative computation of maximum likelihood (ML) estimates or maximum a posterior (MAP) estimates, useful in variety of incomplete-data problems, where algorithms such as the Newton-Raphson method may turn out to be more complicated. On each iteration of the EM algorithm, there are two steps—called the *Expectation Step* and the *Maximization Step*. Generally, complete-data likelihood equation has a nice form and maybe has a solution explicitly. The EM algorithm approaches the incomplete-data problem indirectly by proceeding iteratively in terms of the conditional expectation of complete-data log likelihood equation for given data y and current fit value of parameters.

In remote sensing inversion, we always assume that the observed data error distribution is normal distribution for simplifying the calculation. But under this assumption, only if a few observed data have big error, the inversion result will become instable. So we attempt to retrieve the parameters with t-distribution error, which has more robustness than traditional normal-distribution error.

Supported by the National Nature Science Foundation of China (Project number of NSFC: 40171068) and Special Funds for Major State Basic Research Project (Grant No. G2000077900)

II. INVERSION OF KERNEL MODEL

Linear kernel-driven bi-directional reflectance distribution function (BRDF) models were designed to ease the difficulties of inverting nonlinear physical models, at the expense of some approximation of the original physics. A linear kernel-driven BRDF model has the following form [2]:

$$R(\theta_s,\theta_v,\phi) = f_{iso} + f_{geo}k_{geo}(\theta_s,\theta_v,\phi) + f_{vol}k_{vol}(\theta_s,\theta_v,\phi) \quad (1)$$

where K_{vol} and k_{geo} are "kernels", i.e., known functions of illumination and viewing geometry that describe volume and geometric scattering respectively; θ_s is the zenith angle of the solar direction; θ_s is the zenith angle of the view direction; f is the relative azimuth of sun and view directions; and f_{iso}, f_{geo} and f_{vol} are three unknown coefficients to be adjusted to fit observations.

With more than three uncorrelated multi-angular observations, a regression method can provide estimates of the three parameters much more easily than least squares error fitting of a nonlinear model. After a successful inversion for the three parameters, the results can be used to calculate bi-directional reflectance distribution function with any illumination and viewing angle.

In following analysis, we use matrix type to express kernel-driven model with *m* observations, and note kernel matrix as *A* and parameter matrix as *X*, then formula (1) become:

$$Y_{m\times1} = A_{m\times3}X_{3\times1}, \quad (2)$$

where

$$Y_{m\times1} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_4 \end{pmatrix}, A_{m\times3} = \begin{pmatrix} 1 & k_{geo}(1) & k_{vol}(1) \\ 1 & k_{geo}(2) & k_{vol}(2) \\ \vdots & \ddots & \vdots \\ 1 & k_{geo}(m) & k_{vol}(m) \end{pmatrix}, X_{3\times1} = \begin{pmatrix} f_{iso} \\ f_{geo} \\ f_{vol} \end{pmatrix}$$

Corresponding probabilistic model becomes following form:

$$y = a^T x + e, \quad (3)$$

where y is observation variable, $a^T = (1, k_{geo}, k_{vol})$ is kernel function vector of sample, $x^T = (x_0, x_1, x_2) = (f_{iso}, f_{geo}, f_{vol})$ are unknown coefficients, e is random variable for observation error.

A. Maximum Likelihood Estimation - Least Squares Method

For linear model with multi-group observations, the common method to estimate parameters is least squares method (LSM). Typically, one assumes the error e in Eq (3) is a normal distribution. The least squares method is equivalent to traditional maximum likelihood estimation.

Obviously, if the matrix $A^T A$ is nonsingular, the least squares estimation of parameters can be obtained as follow:

$$X = (A^T A)^{-1} A^T Y \quad (4)$$

B. Maximization A Posterior (MAP) estimation – Bayesian Inversion

In practical inversion, a priori distribution of parameters can be used to obtain posterior distribution information of parameters through physical model. The principle of MAP estimation is to find the parameters with maximum probability in posterior distribution. The Bayesian formula is:

$$P(X \mid y) = \frac{P_D(y \mid X) P_s(X)}{P_D(y)}, \quad (5)$$

where $Ps(X)$ is priori distribution of parameters and the denominator of Eq.(5) is total probability of observation, $P_D(y) = \int_s P_D(y \mid s) P_s(X) dV_s$ where $\int_s$ and dV_s are multi -dimension integration and multi-dimension volume element in parameter space S, respectively.

In linear kernel-driven model, we can obtain a priori normal distribution $N(\overline{X}, C_p)$ of parameters through calculating and retrieving an amount of multi-angle observations of high precision. MAP estimation can be calculated by maximizing the logarithmic form of numerator of Eq.(5):

$$X = (A^T A + \sigma^2 C_p^{-1})^{-1} (A^T Y + \sigma^2 C_p^{-1} \overline{X}) \quad (6)$$

where σ^2 is variance of error.

III. INVERSION BASED ON EM ALGORITHM

Because student t-distribution has more robustness than traditional Gauss distribution, we attempt to use EM algorithm to calculate parameter estimation of linear kernel-driven BRDF model with student t-distribution error. Suppose the linear model is

$$y_i = a_i^T X + e_i \quad (7)$$

where a_i, X both are 3×1 vectors, i denote that y_i is the *i-th* observation.

Obviously, there is not an explicit solution of the ML estimation of parameter.

From the definition of student t-distribution, we have $y_i - a_i^T X \equiv \sigma \mu_i / \sqrt{q_i}$, where $\mu_i \sim N(0,1)$, $q_i \sim \chi_v^2 / v$. According to the idea in paper [3], we regard the combination of $\{e_i, q_i\}$ as complete data and then the random variable e_i under condition q_i satisfies the following distribution

$$e_i \mid q_i \sim N(0, \sigma^2 / q_i) \quad (8)$$

Then by theory of EM algorithm, we can construct iterative formulae of EM algorithm by maximizing condition expectation, denoted by function Q, of logarithmic likelihood function of complete-data, the Q function is

$$Q(\Phi, \Phi^{(k)}) = \sum_{i=1}^{n} \left\{ E[\frac{1}{2}\log(q_i) \mid y_i, a_i, \Phi^{(k)}] - \frac{1}{2}\log(\sigma^2) - \frac{e_i^2}{2\sigma^2} E[q_i \mid q_i, a_i, \Phi^{(k)}] \right\} \quad (9)$$

where $E(\xi)$ denote the expectation value of some random variable ξ, $\Phi = \{\sigma^2, X\}$ is estimation parameter vector, and the superscript *(k)* denote the k-th estimation value of Φ. The final iterative formulae are

$$\begin{aligned} (\sigma^2)^{(k+1)} &= (Y - AX^{(k)})^T W^{(k+1)} (Y - AX^{(k)}) \\ X^{(k+1)} &= (A^T W^{(k+1)} A)^{-1} A^T W^{(k+1)} Y \end{aligned} \quad (10)$$

where $W^{(k+1)} = \mathrm{diag}(w_1^{(k+1)}, w_2^{(k+1)}, \cdots, w_n^{(k+1)})$,

$Y = (y_1, y_2, \cdots, y_n)^T$, $A = (a_1, a_2, \cdots a_n)^T$,

$$w_i^{(k+1)} = E[q_i \mid y_i, a_i, \Phi^{(k)}] = \frac{v+1}{v + \dfrac{(y_i - a_i^T X)^2}{(\sigma^2)^{(k)}}} \quad (11)$$

Similarly, with assumption of student t-distribution error, we also consider MAP estimation if priori information or distribution of retrieval parameters exists. We still assume that the priori distribution of parameters is a multi-dimension normal distribution, namely, $(f_{iso}, f_{geo}, f_{vol}) \sim N(\overline{X}, C_p)$

Then, the Q function becomes

$$Q(\Phi, \Phi^{(k)}) = \sum_{i=1}^{n} \left\{ E[\frac{1}{2}\log(q_i) \mid y_i, a_i, \Phi^{(k)}] - \frac{1}{2}\log(\sigma^2) - \frac{e_i^2}{2\sigma^2} E[q_i \mid q_i, a_i, \Phi^{(k)}] \right\} - \frac{1}{2}(X - \overline{X})^T C_p^{-1} (X - \overline{X}) \quad (12)$$

and the iterative formulae are

$$\begin{aligned} (\sigma^2)^{(k+1)} &= (Y - AX^{(k)})^T W^{(k+1)} (Y - AX^{(k)}) \\ X^{(k+1)} &= (A^T W^{(k+1)} A + (\sigma^2)^{(k)})^{-1} (A^T W^{(k+1)} Y + (\sigma^2)^{(k)} C_p^{-1} \overline{X}) \end{aligned} \quad (13)$$

IV. INVERSION TEST

In following inversions of linear kernel-driven BRDF model, we choose Ross-Thick kernel [4] and Li-Transit [5] as volume scattering kernel and geometric scattering kernel, respectively.

A. Test of Robustness

The primary difference between tradition normal distribution and student t-distribution is t-distribution has more strong robustness.

From 73 priori dataset [6], we select a group of observation data, whose original file name is `irons_soil.intermediate`, as the example data. The geometry of measurement points are shown in figure 1, where measurement points are expressed as crosses and the solar zenith angle was approximately 52°. Next, we make some experiments on robustness with the fifteen measurement data in the principal plane in NIR band.

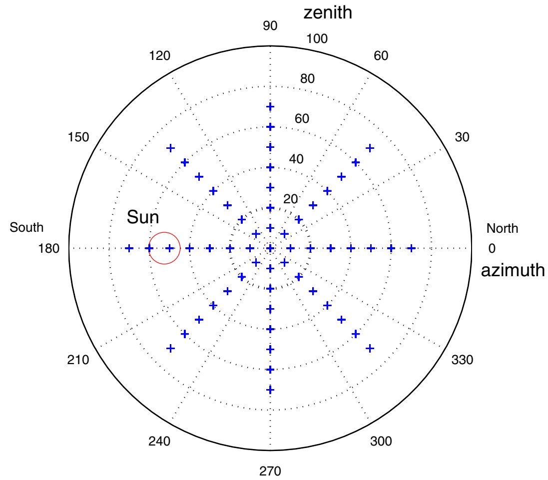

Figure1: Geometry at BRDF data measurement

Without artificial additional noise, the result of *t*-error method solved by EM Algorithm and traditional least square method are indistinguishable in the plots (figure 2). Next, we add noises of different quantities and different levels to original data. From the estimated results as shown in figure 3-5, we can find that with the decreasing of noises level and noises quantity, traditional LSM appears instable estimations, while *t*-Error model solved by EM algorithm still have robust estimations relatively.

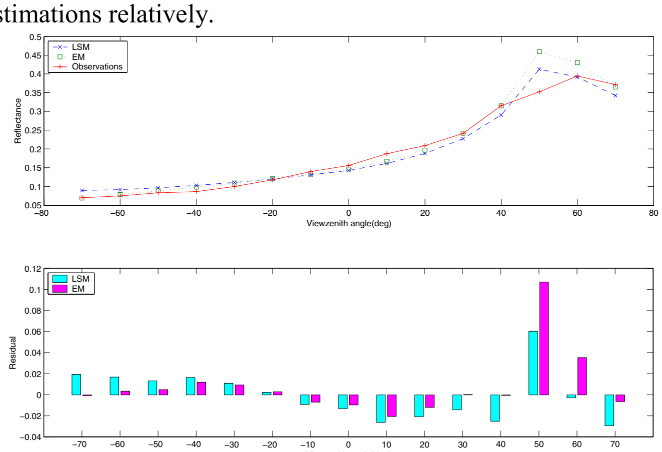

Figure 2: Examples of estimated BRDFs for near-infrared band data in the principal plane. No Noises were added.

B. Inversion of Land Surface Albedo

Albedo is particular importance in the land surface energy balance and the earth's radiation balance that dictates the rate of heating of the land surface under different environmental conditions. Albedo is defined as the ratio of reflected solar radiation from a surface to that incident upon it. From the definition of albedo, the hemispherical integrals of the BRDF over view hemispherical space is directional hemispherical reflectance, also be named as Black Sky Albedo (BSA), while the bi-hemispherical integrals over view and solar hemispherical space is bi-hemispherical reflectance (White

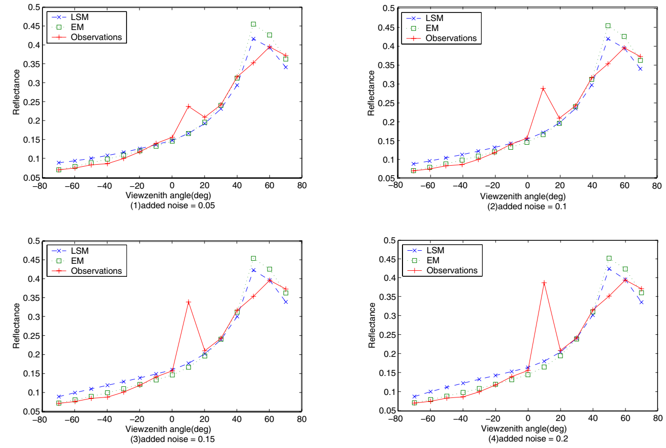

Figure 3: Examples of estimated BRDFs for near-infrared band data in the principal plane. Noises (value = 0.05,0.1,0.15,0.2) were added to one original datum.

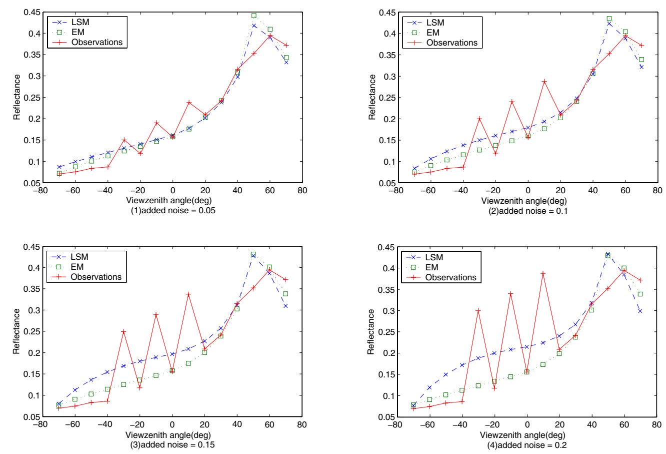

Figure 4: Examples of estimated BRDFs for near-infrared band data in the principal plane. Noises (value = 0.05,0.1,0.15,0.2) were added to three original data

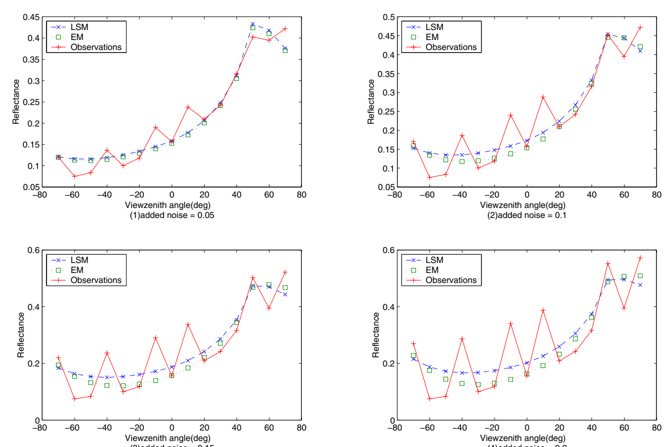

Figure 5: Examples of estimated BRDFs for near-infrared band data in the principal plane. Noises (value = 0.05,0.1,0.15,0.2) were added to five original data

Sky Albedo (WSA)). Mathematically, BSA and WSA are linear combinations of integral of the BRDF kernel over view and solar hemispherical space, then the integral of every kernel is defined as:

$$h_k(\theta_s) = \frac{1}{\pi}\int_0^{2\pi}\int_0^{\pi/2} K_k(\theta_s, \theta_v, \phi)\sin\theta_v \sin\theta_v d\theta_v d\phi$$

$$H_k = 2\int_0^{\pi.2} h_k(\theta_s)\sin\theta_s \cos\theta_s d\theta_s$$

where $K_k(\theta_s, \theta_v, \phi)$ are K_{iso}, K_{geo} and K_{vol},respectively. Therefore, the BSA and WSA are

$$BSA(\theta_s, v) = f_{iso}(v)h_{iso}(\theta_s, v) + f_{geo}(v)h_{geo}(\theta_s, v) + f_{vol}(v)h_{vol}(\theta_s, v)$$
$$WSA(v) = f_{iso}(v)H_{iso} + f_{geo}(v)H_{geo} + f_{vol}(v)H_{vol}$$

where v is wavelength. Obviously, the integral of kernel and is dependent of observation data, and then we could calculate them before the inversion and used as look up tables, which is a characteristic of linear kernel-driven BRDF model.

But, the analytical form of integral of kernel is difficult to obtain and then we use a polynomial of solar zenith angle to fit of kernel integral for the computational efficiency. In the thesis, we choose Yang hua's fitting form for Li–Transit kernel [7] as follows:

$$h_{geo}(\theta_s) = -0.825 - \left|\frac{\theta_s \pi}{83 \times 180}\right|^{1.76}$$

Besides, for the Ross-Thick kernel, the fitting polynomial in Algorithm for MODIS Bidirectional Reflectance Anisotropies of the Land Surface (AMBRALS) is $h_{vol}(\theta_s) = -0.007574 - 0.070987\theta_s^2 + 0.307588\theta_s^3$, The three integrals of kernel in WSA are constants, which are $H_{iso} = 1.0$, $H_{geo} = 1.206965$, $H_{vol} = 0.189184$.

Li xiaowen give a group observation data of NOAA AVHRR from a failed NIR inversion in paper [6]. With the data as shown in table1, we use four methods to retrieve the parameters f_{iso}, f_{geo}, f_{vol} and calculate corresponding black sky and white sky albedos in different solar zenith angles. From the inversion results listed in table2, we know that the t-Error model solved by EM algorithm have more stable results than the traditional LSM, but it is also a failed inversion because the calculation value of BSA of 60^o szn is negative. (BSA, WSA should have the physical limitation of the interval [0,1]). The Bayesian inversion and Bayesian inversion with t-Error both have reasonable inversion results.

V. Conclusion

In this paper, taking linear kernel BRDF model as an example, we propose an EM algorithm based inversion method with t distribution error to obtain more precise and more robust inversion results, but a necessary preliminary work is the validation of t-distribution error with practical remote sensing data.

Acknowledgment

The first author would like to thank Prof. CUI Hengjian for helpful suggestions and productive discussions about EM algorithm.

TABLE 1
Selected original NOAA AVHRR data from a failed NIR inversion

No	θ_v	θ_v	ϕ	Ref.NIR
1	61.3	124.6	28.8	0.165
2	27.6	42.0	35.2	0.287
3	12.4	42.5	34.3	0.298
4	20.2	130.6	32.9	0.216
5	33.7	129.2	32.5	0.210
6	53.0	126.5	32.0	0.195
7	17.0	43.4	37.8	0.190
8	1.3	78.3	37.1	0.181

TABLE 2
Inversion Results of the failed data By Four Methods

	LSM	t-Error	Bayes	Bayes t-Error
f_{iso}	0.6170	0.6083	0.3974	0.3957
f_{geo}	0.3959	0.3762	0.1680	0.1597
f_{vol}	-0.7609	-0.6093	0.0280	0.0624
WSA	$\boxed{-0.0048}$	0.0389	0.1999	0.2147
BSA, θ_v=0^o	0.2961	0.3025	0.2586	0.2634
BSA, θ_v=20^o	0.2113	0.2247	0.2312	0.2383
BSA, θ_v=45^o	0.0813	0.1103	0.2043	0.2156
BSA, θ_v=60^o	$\boxed{-0.1371}$	$\boxed{-0.0778}$	0.1714	0.1904

References

[1] A. P. Dempster, N. M. Laird and D.B. Rubin Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B*,39 (**1**), 1-38,1977

[2] J .L. Roujean, M. Leroy, and P. Y. Deschamps. A bi-directional reflectance model of the earth's surface for the correction of remote sensing data. Journal of Geophysical Research, 97:20455–20468, 1992.

[3] K. L. Lange, J. A. Little, and M. G. Taylor. Robust statistical modeling using the t distribution. Journal of the American Statistical Association, 84:881–896, 1989.

[4] Roujean, J.-L., Leroy, M., & Deschamps, P. Y. (1992). A bi-directional reflectance model of the Earth's surface for the correction of remote sensing data. Journal of Geophysical Research, 97, 20455– 20468.

[5] Xiaowen Li, Feng Gao, Qiang Liu, Jindi Wang, A. Strahler. Validation of a new GO kernel and Inversion of Land Surface Albedo by Kernel-Driven Model(1)， Journal of Remote Sensing， Vol 4（Suppl）:8-15,2002

[6] Xiaowen Li, Feng Gao, Jindi Wang, and Alan Strahler. A priori knowledge accumulation and its application to linear BRDF model inversion. Journal of Geophysical Research, 106(D11):11925–1193, 2001.

[7] Hua Yang , Xiaowen Li, Feng Gao An Algorithm for the Retrieval of Albedo from space Using New GO Kernel –Driven BRDF model Journal of Remote Sensing . 6(4):246-251,2002