

***CLOUD TOMOGRAPHY: ROLE OF CONSTRAINTS AND  
A NEW ALGORITHM***

D. Huang<sup>1\*</sup>, Y. Liu<sup>1</sup>, and W. Wiscombe<sup>1,2</sup>

<sup>1</sup>Brookhaven National Laboratory, Upton, NY 11973-5000

<sup>2</sup>NASA Goddard Space Flight Center (code 913), Greenbelt, MD 20771

\*Corresponding Author: Tel: (631) 344-5818; Email: dhuang@bnl.gov

February 2008

Submitted for publication in  
*J. Geophys. Res.*

**Environmental Sciences Department/Atmospheric Sciences Division**

**Brookhaven National Laboratory**

P.O. Box 5000

Upton, NY 11973-5000

www.bnl.gov

## Abstract

Retrieving spatial distributions of cloud liquid water content from limited-angle emission data (passive microwave cloud tomography) is so ill-posed that even a small inaccuracy in the data and/or numerical treatments may result in a large error in the retrieval. Proper handling of the ill-posedness is an ongoing challenge to the atmospheric remote sensing community. In this paper we first analyze the major regularization methods that each apply a single but different constraint to their retrievals, and extend these methods to allow for multiple constraints. We then develop a new iterative algorithm that can also incorporate complex physical constraints with great flexibility. To understand the influences of different constraints on the retrievals, we use the new iterative algorithm with various combinations of constraints to retrieve a stratocumulus cloud simulated with a large-eddy-simulation model. The standard least squares method with no constraints, as expected, performs very poorly, and yields a mean retrieval error of  $0.77 \text{ g m}^{-3}$ , making this method nearly useless. Adding a non-negativity constraint reduces the mean retrieval error to  $0.13 \text{ gm}^{-3}$  but the internal structure of the cloud is still not reproduced in the retrieval. Adding a smoothness constraint dramatically improves the retrieved spatial structure of the cloud, and brings the mean error down to  $0.093 \text{ gm}^{-3}$ , although the retrieved cloud top edges are still considerably blurred. Further adding the so-called double-side constraint (based on scaled adiabatic profiles) produces the best result; the retrieval faithfully reproduces the cloud water structure with a mean retrieval error of only  $0.037 \text{ gm}^{-3}$ .

## 1. Introduction

Knowledge of three-dimensional distributions of cloud water is vital for fundamental research in cloud physics and atmospheric radiation, for validation of cloud-resolving and large-eddy-simulation models, and for applications like weather forecasting. In contrast to relatively expensive active remote sensing methods like radar, the microwave cloud tomography method is the only known passive method for remotely measuring spatial distributions of cloud water.

The microwave cloud tomography involves measuring cloud microwave emission in a set of different directions (Figure 1) from a single airborne or multiple ground-based scanning microwave radiometers, then mathematically inverting the multi-angular emission data for the spatial distribution of the cloud liquid water content (LWC). The underlying relationship between the measured quantity and the cloud water distribution takes the form of the Fredholm integral equation of the first kind. Unlike conventional remote sensing methods, the multi-angular measurements of cloud tomography are not independent; each slant ray may intersect many others and, wherever an intersection takes place, both rays depend on the LWC in the same intersected pixel (Figure 1). Thus the tomography retrieval problem requires solving a large system of linear or quasi-linear equations with hundreds to thousands of unknowns.

The cloud tomography method was first proposed by Warner et al. (1985), who demonstrated its feasibility using computer simulations based on two ground-based scanning radiometers. Warner et al. (1988) conducted a theoretical study of the retrieval accuracy of an airborne cloud tomography setup. Twomey (1987) developed an iterative inversion algorithm for reducing the computational cost of the tomography retrieval. Warner and Drake (1986) and Drake et al. (1988) reported on field demonstrations of ground-based and airborne cloud tomography. Since then, this subject has more or less lain dormant, leaving unanswered some important questions, for example, the mathematical nature of cloud tomography.

Huang et al. (2007) revisited the cloud tomography problem and showed that, like other limited angle tomography problems, the mathematical problem of the cloud tomography is so ill-posed that its solution is non-unique, and very sensitive to measurement noises and numerical errors, especially when only a few (2-4) ground-based radiometers are used. These results suggest that the limited-angle radiometric data alone does not contain enough information to adequately determine the spatial LWC distribution. For such an ill-posed inverse problem, additional constraints based on *a priori* knowledge can be used to find the solution. Huang et al. (2007) successfully incorporated a smoothness constraint in the tomography retrieval algorithm by means of the truncated singular value decomposition (TSVD) method. The addition of the smoothness constraint reduced the retrieval error of the cloud LWC by 80% compared to the unconstrained, standard least squares method. Despite the progress, some problems remain unresolved, for example, poor retrievals at cloud edges. In addition to the smoothness, one can also use other types of constraints, such as non-negativity of cloud LWC values, or prior estimates of the cloud LWC field from a physical model or other more direct measurements, to further improve the retrievals.

Several regularization techniques have been developed to make use of different constraints to obtain better solutions for ill-posed inverse problems in various disciplines (Twomey, 1977; Rodgers, 2000). However, a single mathematical constraint (e.g., non-negativity, smoothness, double-side constraint, or at best a combination of two of them) is usually used in these approaches; none of the existing approaches can seamlessly and flexibly combine several constraints together. Furthermore, the existing approaches lack the capability to include more sophisticated constraints such as those derived from a physical model. A comprehensive algorithm that is capable of using all of these constraints is desirable for cloud tomography.

In this paper, we first examine the major existing methods of regularization to discern their deficiencies. We then extend and integrate these different methods into a new iterative algorithm that can readily incorporate almost any types of constraints (both the aforementioned mathematical constraints and more sophisticated physical constraints), and overcomes the deficiencies of existing approaches. Finally, we use the new algorithm to demonstrate the role of each constraint in improving the retrieval of the LWC spatial distribution.

## 2. Constrained retrieval algorithms

Here we present the mathematical formulation of the tomographic problem and then briefly outline the existing constrained retrieval methods.

### 2.1 *Mathematical formulation of the tomographic problem*

Like many other remote sensing problems (Towmey 1977), the retrieval of the spatial LWC distribution can be formulated as the mathematical inversion of the Fredholm integral equation of the first kind:

$$\int_0^1 a(r, \Omega)x(r)dr = b(\Omega) \quad (1)$$

where  $r$ ,  $\Omega$ ,  $x$ ,  $b$ , and  $a$  denote spatial location, direction, cloud LWC, measured or simulated microwave brightness temperatures, and kernel function, respectively.

In the application of cloud tomography, the brightness temperature measurements are taken at a finite number of directions  $\Omega_i \{i=1, \dots, m\}$  and the LWC is retrieved at a finite number of locations (pixels)  $r_j \{j=1, \dots, n\}$ . With this discretization, Eq. (1) reduces to the following matrix equation:

$$\mathbf{Ax} = \mathbf{b} \quad (2)$$

where  $\mathbf{x}^T \equiv (x_1, x_2, \dots, x_n)$  is the vector of state variables to be retrieved (i.e., LWC);

$\mathbf{b}^T \equiv (b_1, b_2, \dots, b_m)$  is the vector of the measurements (i.e., microwave brightness temperatures);

and  $\mathbf{A} \equiv (a_{ij})$  is the kernel matrix.

At first glance, Eq. (2) can be solved by use of the standard least squares method that seeks the solution by minimizing the following cost function:

$$J \equiv \|\mathbf{Ax} - \mathbf{b}\|_2^2 = (\mathbf{Ax} - \mathbf{b})^T (\mathbf{Ax} - \mathbf{b}), \quad (3)$$

where  $\|\bullet\|_2$  denotes the 2-norm of a vector. It is readily shown that minimizing  $J$ , or setting  $\partial J / \partial \mathbf{x} = 0$ , leads to the least squares solution

$$\mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}. \quad (4)$$

However, as shown in Huang et al. (2007), similar to many other remote sensing problems discussed in Twomey (1977), the inverse problem associated with cloud tomography is highly ill-posed, and small changes in the measurement vector  $\mathbf{b}$  giving rise to large changes in the solution vector  $\mathbf{x}$ . As a result, many solutions may satisfy Eq. (2) within a certain noise level. This problem of ill-posedness and solution ambiguity is further illustrated in Figure 2 by the ellipsoid  $\mathbf{S}$ . In such a case, the solution selected by the least squares method is usually far from the true solution because the measurement noises and numerical errors are amplified by the small

eigenvalues of  $\mathbf{A}^T\mathbf{A}$ . The ellipsoid  $\mathbf{S}$  for the least squares confidence interval is therefore elongated along the directions of eigenvectors corresponding to the small eigenvalues.

Huang et al. (2007) showed that improved tomographic retrieval can be obtained by discarding the small singular values of the matrix  $\mathbf{A}$  through the use of the so-called TSVD method. Nevertheless, there are also other techniques of regularization that have been developed to solve the general ill-posed inverse problem. The following sub-sections discuss these techniques.

## ***2.2 Smoothness constraint***

Measurement noises and numerical errors tend to produce oscillations with large amplitudes in the solution of ill-posed problems, and thus a smooth solution is usually more preferable. The smoothness constraint is a soft constraint in that the solution is driven toward a more smooth distribution but still allows for some discontinuities. There are several ways to implement it. One is the TSVD that impose some smoothness on the solution by discarding some small singular values (Hansen, 1998; Huang et al., 2007).

Another implementation of the smoothness constraint is the Philips-Twomey method (Twomey, 1977; Philips, 1962) known in the atmospheric sciences community or, the Tikhonov regularization in other disciplines (Tikhonov, 1963). In this paper, we use the Tikhonov regularization. Hansen (1998) proved that, under suitable conditions, the Tikhonov solution is very close to the TSVD solution. In this paper, we choose the Tikhonov regularization as the base method because of its ability to include other types of constraints.

The idea of the Tikhonov regularization is to minimize the cost function:

$$J_s \equiv \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{Lx}\|_2^2 \quad (5)$$

where the second term on the right hand side controls the smoothness of the solution, *and*  $\mathbf{L}$  is typically either the identity matrix, a diagonal matrix, or a discrete approximation of a derivative operator;  $\lambda$  is the regularization parameter determining the amount of the smoothness imposed on the retrievals. In the rest of this paper, the matrix  $\mathbf{L}$  is chosen to be the approximation of the two-dimensional first derivative operator.

Solving  $\partial J_S / \partial \mathbf{x} = 0$  for  $\mathbf{x}$  as before, the smoothness constrained least squares solution  $\mathbf{x}_S$  is given by:

$$\mathbf{x}_S = (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{L}^T \mathbf{L})^{-1} \mathbf{A}^T \mathbf{b}. \quad (6)$$

In the Appendix A, we show that Tikhonov regularization is closely connected with the Bayesian theorem, and that the regularization parameter  $\lambda$  can be expressed as the ratio of the variance of the measurement errors to the variance of the state variable to be retrieved. However, in cloud tomography the variance of cloud LWC is usually unknown. We determine the regularization parameter  $\lambda$  using the L-curve method (Hansen, 1992) instead.

### ***2.3 Non-negativity constraint***

Both the unconstrained and the smoothness constrained least squares methods may lead to negative solutions, which are of course neither desirable for cloud tomography nor for many other remote sensing applications. Obtaining non-negative retrievals is especially important for seriously ill-posed problems because the unconstrained least squares method tends to produce oscillatory retrieval with frequent sign changes.

The NNLS (Non-Negative Least Squares) algorithm introduced by Lawson and Hanson (1974) imposes a non-negativity constraint, a highly desirable property for LWC retrievals. The

non-negativity constraint in nature is a hard constraint in that no negative solutions will be produced. NNLS was used in the early works on cloud tomography, e.g., Warner et al. (1985) and Warner et al. (1986). NNLS solves the basic matrix equation subject to the added constraint that the solution vector contains no negative elements:

$$\min_{\mathbf{x}} \left\{ \|\mathbf{Ax} - \mathbf{b}\|_2^2 \right\} \text{ subject to } \mathbf{x} \geq 0. \quad (7)$$

The algorithm starts with a set of possible basis vectors, computes the associated dual vector  $\mathbf{w}$ , and selects the basis vector corresponding to the maximum value in  $\mathbf{w}$  to swap out of the basis in exchange for another possible candidate, until the Kuhn-Tucker conditions are satisfied.

Liu et al. (1999) developed an approach by combining the Tikhonov regularization with the NNLS in retrieval of the particle size distribution from measurements of multispectral optical depths, and dubbed this approach as smoothness-constrained NNLS (SCNNLS). Essentially, the SCNNLS requires solving the following minimization problem:

$$\min_{\mathbf{x}} \left\{ \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{Lx}\|_2^2 \right\} \text{ subject to } \mathbf{x} \geq 0 \quad (8)$$

Liu et al. (1999) demonstrated that the SCNNLS improves the retrieved particle size distributions, especially at the two ends of particle sizes where particle concentrations are low.

#### **2.4 Double-side constraint**

With *a priori* knowledge of the range of the solution, either from other more direct measurements or from theory, one can construct “double-side” constraints to bound the solution. Such a method was developed by Pierce and Rust (1985) and was successfully used by Babb et

al. (2000) to deconvolve atmospheric turbulence from Doppler velocity spectra and retrieve cloud drop size distributions. The double-side constraints (i.e., the bounding box  $\mathbf{B}$  in an  $n$ -dimensional space) are defined by a vector  $\mathbf{x}_b$  of the *prior estimate* (the center of the box) , and a matrix,  $\mathbf{Q} \equiv \text{diag}(q_1, q_2, \dots, q_n)/2$ , where  $q_i$  denotes the width of the bounding box in the  $i$ -th dimension (Figure 2a). By taking a convex combination of the solution ellipsoid  $\mathbf{S}$  and the bounding ellipsoid  $\mathbf{B}'$  (the envelope ellipsoid of the box  $\mathbf{B}$ ) one obtains a smaller solution ellipsoid  $\mathbf{P}$  (Figure 2b) . The solution vector  $\mathbf{x}_{\text{DS}}$  with such a double-side constraint is then given by:

$$\mathbf{x}_{\text{DS}} = \left( \mathbf{A}^T \mathbf{A} + \tau \mathbf{Q}^{-2} \right)^{-1} \left( \mathbf{A}^T \mathbf{b} + \tau \mathbf{Q}^{-2} \mathbf{x}_b \right), \quad (9)$$

where  $\tau$  is a weighting parameter determining how large a contribution from the prior estimate  $\mathbf{x}_b$  is applied in the combination. As shown in the Appendix B, under suitable assumptions the weighting parameter  $\tau$  is actually the variance of the measurement error. In cloud tomography, the noise level of the microwave radiometers is usually well known; hence the result presented in the Appendices B and C provides a useful method to choose the weighting parameter  $\tau$ .

When the prior estimate is perfectly accurate, the bounding box shrinks to a point and we have  $\mathbf{Q} \rightarrow 0$  and  $\mathbf{Q}^{-2} \rightarrow \infty$ . Then, the constrained solution coincides with the prior estimate  $\mathbf{x}_b$ . On the other hand, if the prior estimate is very inaccurate, we have  $\mathbf{Q}^{-2} \rightarrow 0$ ; thus the  $\mathbf{Q}$  terms in Eq. (9) can be neglected and the constrained solution becomes the unconstrained least squares solution. Thus, when the size of the bounding box is too large, the prior information is not useful. In that case, additional constraints are needed to select reasonable solutions.

### 3. New algorithms for using combined constraints

In this section, we first extend the aforementioned methods to make use of combined constraints. Then we integrate the extended methods into a new iterative algorithm which can also handle more sophisticated constraints such as those from a physical model.

#### 3.1 Double-side constrained SCNNLS (DSCNNLS)

The methods described in Section 2 use only a single constraint (smoothness, non-negativity, and double-side) or at best a combination of the smoothness and non-negativity constraints. Following the Bayesian theorem, we propose an algorithm that can use any combination of the smoothness, non-negativity, and double-side constraints. This section briefly introduces this new method and the detailed derivation is given in the Appendix C.

Essentially, the DSCNNLS solves the following minimization problem:

$$\min_x \left\{ \|\mathbf{A}'\mathbf{x} - \mathbf{b}'\|_2^2 \right\} \text{ subject to } \mathbf{x} \geq 0, \quad (10)$$

where

$$\mathbf{A}' \equiv \mathbf{A}^T \mathbf{A} + \lambda \mathbf{L}^T \mathbf{L} + \tau \mathbf{Q}^{-2},$$

$$\mathbf{b}' \equiv \mathbf{A}^T \mathbf{b} + \tau \mathbf{Q}^{-2} \mathbf{x}_b.$$

The smoothness and double-side constraints can be disabled by setting the regularization parameter  $\lambda$  and the weighting parameter  $\tau$  to zero, respectively, while the non-negativity constraint can be inactivated by using standard least-squares methods to solve Eq. (10) instead of the NNLS algorithm.

Note that the dimensionality of  $\mathbf{A}'$  is same as  $\mathbf{A}^T \mathbf{A}$ , so the computational cost to invert  $\mathbf{A}'$  is about same as that for  $\mathbf{A}^T \mathbf{A}$ . The only additional cost of DSCNNLS compared to the standard method is the computation of the matrix  $\mathbf{A}'$  itself, which is insignificant compared to the cost of

inverting  $\mathbf{A}'$ . Therefore, the overall computational cost of including the combined constraints is on the same order as that of the standard least squares algorithm.

### ***3.2 A new iterative algorithm***

All the discussed non-iterative methods make use of constraints describing some important mathematical properties of the solutions, such as smoothness and non-negativity, in the form of equalities or inequalities. Nevertheless, more sophisticated constraints, such as those follow from the physical essence of the underlying problem, cannot be applied directly in the aforementioned non-iterative algorithms. For example, the hierarchical structure of clouds is a result of dynamic, thermodynamic, microphysical, and radiative processes of the atmosphere. The physical laws governing these processes are the basis of Cloud Resolving Models (CRM) and their use allows for the best simulations of the time evolution, structure, and life cycle of clouds and cloud systems that are currently possible (Tao, 2006). For the practice of cloud tomography, these physical laws ought to be useful to constrain the tomographic retrieval problem if they are used appropriately in the retrieval algorithm. For example, many clouds are initiated by adiabatic or pseudo-adiabatic convective processes and the cloud LWC rarely exceeds the adiabatic value as confirmed by numerous field experiments.

In order to make use of such sophisticated constraints, we develop an iterative algorithm through which sophisticated constraints can also be included with great flexibility. As depicted in Figure 3, the new retrieval machinery is able to incorporate the non-negativity, smoothness, and double-side constraints by means of the methods described in Section 2.5 (Box 1) in each single iteration, while the more sophisticated constraints are satisfied by adjusting the retrieval from the previous iteration (Box 2) to form a new estimate which is then used as a double-side constraint

for the next iteration. The iterations continue until the difference between the retrievals from two iterations is below a predetermined criterion.

As illustrated in Figure 3, the iterative machinery contains two main steps: the first step is to compute the posterior estimate  $\mathbf{x}$  based on the prior estimate  $\mathbf{x}_b$  and the measurements  $\mathbf{b}$  by means of the constrained algorithm given by Eq. (10). Note that in the first iteration, the double-side constraint is disabled by setting the weighting parameter  $\tau$  to zero; the second step is to adjust the posterior estimate  $\mathbf{x}$  so that the constraints in Box 2 are satisfied, and then use the modified  $\mathbf{x}$  as the prior estimate  $\mathbf{x}_b$  for the next iteration. The machinery is terminated when the posterior estimate  $\mathbf{x}$  meets some predetermined error tolerance  $\varepsilon$ . Expressed in equations this is, for the  $k$ th iteration,

$$\mathbf{x}(k) = F^{-1}[\mathbf{b}, \mathbf{x}_b(k)], \quad (11)$$

$$\mathbf{x}_b(k+1) = C[\mathbf{x}(k)], \quad (12)$$

where  $F^{-1}$  is the inversion operator (Box 1 in Figure 3), and  $C$  is the constraint operator (Box 2 in Figure 3).

The algorithm is very flexible: it can include various combinations of constraints by setting the relevant weighting parameters in Eq. (10) to zero; it can also be run in a non-iterative manner by letting the error tolerance  $\varepsilon$  be infinity (so that the algorithm is always terminated after the first iteration) if a double-side constraint is not available. Therefore, this algorithm allows for a close examination of the usefulness of various constraints in cloud tomography.

## 4. Application to cloud tomography

In this section, we use the new iterative algorithm to demonstrate the effects on the retrievals of different constraints by adding them one after another. We start with a description on how we simulate the needed tomographic data and how we invert the simulated data for the cloud LWC fields. Then we compare the retrievals obtained with different combination of constraints.

### 4.1 Experiment setup

The needed cloud tomography data are simulated by the Observation System Simulation Experiment described in Huang et al. (2007). Briefly, a radiative transfer model is used to simulate the measured brightness temperature given certain predefined radiometer specification.

A two-dimensional 5 Km wide and 1.5 Km high slice of cloudy atmosphere is taken from the simulations of the DHARMA Large Eddy Simulation (LES) model driven by the data from the Atlantic Stratus Experiment (Ackerman et al., 1995). The original high-resolution LES simulation is degraded to an image of 20 by 20 pixels (250-meter horizontal and 75-meter vertical resolution) before simulating the radiometer measurements. Four simulated radiometers of 0.3 K noise level and 2-degree beam width are placed equally on the ground along a line of 10 Km (Figure 1). Each radiometer scans the upper plane within  $85^\circ$  elevation of zenith at a  $0.4^\circ$  increment. This scanning strategy results in a total number of 800 rays hitting the 5 Km by 1.5 Km area.

The simulated tomographic data are then inverted using the iterative algorithm described in Section 3 with various combinations of constraints. Starting with the unconstrained algorithm in a non-iterative manner, we then add the non-negativity and smoothness constraints one at a time by means of Eqs. (6-8). Next, we use the adiabatic profile to derive a double-side constraint

(Box 2 in Figure 3) and switch the algorithm to an iterative manner. Specifically, the double-side constraint is specified by a prior estimate vector  $\mathbf{x}_b$  and a diagonal matrix  $\mathbf{Q}$  (see Eq. (10)). The prior estimate vector  $\mathbf{x}_b$ , defining the center of the bounding box, is computed by adjusting the retrieval from the previous iteration according to a scaled adiabatic profile. The adjustment involves the following steps: (1) finding the cloud base and top based on the retrieval from the previous iteration; (2) calculating the liquid water path of each column; (3) computing the scale factor as the ratio of the liquid water path calculated in step (2) to that calculated from a pure adiabatic profile; and (4) adjusting the LWC values for each pixel in each column according to the scaled adiabatic profile so that the liquid water path kept unchanged with the previous iteration. The diagonal matrix  $\mathbf{Q}$  defines the width of the bounding box in each dimension. It is assumed to take the form of  $\sigma\mathbf{I}$ , where  $\mathbf{I}$  is the identity matrix and  $\sigma$  is an adjustable parameter. In this work we let this parameter equal  $0.1 \text{ gm}^{-3}$ , approximately the rms error of the non-negativity plus smoothness constrained retrieval.

#### ***4.2. Retrieval results***

To illustrate the effects of different constraints on the retrieval, we discuss the results obtained by adding different constraints to the new iterative algorithm one after another. The “true” distribution of LWC for the selected stratocumulus cloud is shown in Figure 4 (“True”) as a reference. The cloud LWC field retrieved by the unconstrained least squares method shows large spatial variations (Figure 4, “Unconstrained”): large positive values are interlaced with negative values. But the integration of the LWC values along each radiometer scanning direction agrees very well with the microwave emission measurement along the same direction, which is not a surprising result because the unconstrained least squares method minimizes the squared residual error between the simulated and measured emission. Figure 5 shows the retrieval error

fields computed as the absolute difference between the true and the retrieved fields. The unconstrained method gives large errors everywhere (Figure 5, “Unconstrained”) and the averaged rms error is  $0.78 \text{ gm}^{-3}$  (Figure 6, “Unconstrained”), which further suggests that this method is almost useless in terms of obtaining the spatial LWC distribution.

Adding the non-negativity constraint (the NNLS algorithm), the retrieval is greatly improved compared to its unconstrained counterpart. The negative values in the unconstrained retrieval are corrected as a direct result of the non-negativity constraint (Figure 4, “NN”), and the rms error is brought down to  $0.23 \text{ gm}^{-3}$  (Figure 6, “NN”). The location and spatial extent of the cloudy area are now roughly captured, although the spatial distribution of LWC within the cloud is still different significantly from the real one.

The smoothness constrained retrieval (Tikhonov regularization) also captured the location and extent of the stratocumulus cloud (Figure 4, “S”). The spatial distribution of cloud LWC is more realistic than that obtained from the NNLS algorithm: the zeros inside the stratocumulus cloud are now corrected by the smoothness constraint, and the rms error of  $0.098 \text{ gm}^{-3}$  is about a half of that of the NNLS. However, the top boundary of the cloud is still blurred and extends to a higher altitude than in the reference. Furthermore, some water-free areas below the cloud base are incorrectly labeled as patchy clouds with very low LWC values.

Combining the non-negativity and smoothness constraints (the SCNNLS algorithm) leads to further improvement in the retrieval (Figure 4, “NN+S”). Adding the non-negativity constraint to the smoothness constrained method removes the spurious patchy clouds below the base of the stratocumulus cloud. The retrieval also roughly reproduces the increasing of the cloud LWC with altitude. Now, the value of the non-negativity constraint is better shown by comparing the error fields of “Unconstrained” with “NN constrained”, or “S constrained” with “S+NN constrained” (Figure 5). Note how the non-negativity constraint reduces the retrieval error below the cloud

base. In a similar manner, the values of the smoothness constraint can be assessed: comparing “Unconstrained” with “S constrained”, or “NN constrained” with “S+NN constrained” (Figure 5). The smoothness constraint significantly reduces the error inside the cloud.

Figures 4 and 5 suggest that neither the non-negativity constraint and the smoothness constraint alone nor the combination of them can reproduce the sharp edges on the cloud top, which indicates that other information should be taken into account in order to obtain more realistic retrieval. Using the adiabatic profile (scaled by the liquid water path from previous iteration) as the double-side constraint improves on the retrieval around the cloud top boundary (Figure 4, “NN+S+DS”). The location and spatial extend of the cloud are very well reproduced in the retrieved field. The internal structure of the cloud LWC, which is not well resolved in the non-negativity and smoothness constrained retrieval, is now accurately captured. Furthermore, the iterative algorithm converges very fast (only 2-3 iterations are needed to meet an error threshold of  $10^{-4} \text{ gm}^{-3}$ ) possibly owing to the fact that the combined constraints greatly improve the kernel matrix and thus the retrieval problem becomes much less ill-posed.

Figure 6 further shows that the retrieval error decreases dramatically when the various constraints are added gradually in the retrieval algorithm. The unconstrained method cannot produce reasonable retrievals of cloud LWC (rms error  $0.78 \text{ gm}^{-3}$ ), as also illustrated in Huang et al. (2007). Adding the non-negativity and smoothness constraints reduces the retrieval errors to  $0.23 \text{ gm}^{-3}$  and  $0.098 \text{ gm}^{-3}$ , respectively. Combining the non-negativity and smoothness constraints further improves the retrieval to an rms error of  $0.093 \text{ gm}^{-3}$ . The double-side constraint (based on scaled adiabatic profiles) appears to be superb in reproducing the cloud structure and it reduces the rms error to  $0.037 \text{ gm}^{-3}$ . In summary, the combined use of the non-negativity, smoothness, and double-side constraints greatly improves the cloud tomography retrievals.

## 5. Further examination of the constraints

In order to theoretically examine the influences of the different constraints on the tomography retrieval, we compare the orthogonal decompositions of the matrices to be inverted, namely,  $\mathbf{A}^T \mathbf{A}$  for the unconstrained least squares,  $\mathbf{A}^T \mathbf{A} + \lambda \mathbf{L}^T \mathbf{L}$  for the smoothness constrained method, and  $\mathbf{A}^T \mathbf{A} + \lambda \mathbf{L}^T \mathbf{L} + \tau \mathbf{Q}^{-2}$  for the smoothness plus double-side constrained method. Among the available orthogonal decompositions, we choose the SVD method because it is particularly useful to analyze the effect of perturbations on the solutions of the least squares problems (Hansen, 1998).

The SVD of an  $n \times n$  matrix  $\mathbf{K}$  can be written as:

$$\mathbf{K} = \mathbf{U} \mathbf{O} \mathbf{V}^T. \quad (13)$$

Here  $\mathbf{U}$  and  $\mathbf{V}$  are orthogonal matrices whose columns are the input and output basis vector directions for  $\mathbf{K}$ ;  $\mathbf{O} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$  is an diagonal matrix with the non-negative singular values ordered such that  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$ . Owing to the orthogonality of matrices  $\mathbf{U}$  and  $\mathbf{V}$ , the inverse of matrix  $\mathbf{K}$  can be easily written as:

$$\mathbf{K}^{-1} = \mathbf{V} \mathbf{O}^{-1} \mathbf{U}^T. \quad (14)$$

From Eq. (14), it is evident that if a singular value is very small, its reciprocal will be very large, and thus any measurement or rounding error will cause a large error in the retrieval. Therefore, by examining the singular values, it is possible to determine the degree of ill-posedness of the underlying problems.

Figure 7 shows the magnitude of the singular values as a function of their orders for the matrices  $\mathbf{A}^T \mathbf{A}$ ,  $\mathbf{A}^T \mathbf{A} + \lambda \mathbf{L}^T \mathbf{L}$ , and  $\mathbf{A}^T \mathbf{A} + \lambda \mathbf{L}^T \mathbf{L} + \tau \mathbf{Q}^{-2}$ , which respectively correspond to the unconstrained, smoothness constrained, and smoothness plus double-side constrained methods. For the unconstrained case, the singular values appear to drop 7 orders in magnitude with increasing index. The high order singular values ( $\sigma_i \mid 200 \leq i \leq 400$ ) are extremely small, which indicates the unconstrained cloud tomography problem is very ill-posed. Hence, the large spatial variations and frequent sign changes in the retrieved image (Figure 4, “Unconstrained”) can be explained by these small singular values and the frequent sign changes in the elements of the singular vectors associated with such high order singular values (Huang et al., 2007).

The smoothness constrained singular values show a relatively flat profile after  $i \geq 150$ , while the unconstrained ones decrease in almost the entire range of  $i$  (Figure 7). Therefore, the singular values have a much smaller range than the unconstrained case. Nevertheless, the singular values for the smoothness constrained case decrease much faster after  $i \geq 390$  than between  $i = 150$  and  $i = 390$ , which suggests that the smoothness constrain alone cannot fully resolve the ill-posed problem. This is confirmed in the retrieved image (Figure 3, “S constrained”) that despite a great reduction of the average rms error there are still some significant errors around the cloud boundary.

Adding the double-side constraint, the behavior of the singular values is further improved (Figure 7). The singular values for  $i < 200$  are almost the same as those of the smoothness constrained case. But for  $i \geq 200$  (especially for  $i \geq 390$ ), the singular values of the double-side plus smoothness constrained case decrease slower than those of the smoothness constrained case. As a result, the useful range of the singular values is further reduced. The resulting improvement on the retrieval is clearly shown in Figures 3 and 4 (“NN+S+DS constrained”).

## 6. Concluding remarks

Previous studies have shown that regularization methods using appropriate constraints can reduce the retrieval ambiguity and obtain better retrieval for ill-posed problems. We first examine major existing regularization methods that use mathematical constraints such as non-negativity, smoothness, and double-side and find that these methods are closely related to the Bayesian theorem under suitable assumptions. Then we extend these methods so that they can make use of multiple constraints but maintain the same level of computational cost. The extended methods are still unable to include sophisticated constraints such as those from a physical model. Thus, we integrate these extended methods into a new iterative algorithm whereby both the mathematical constraints and more sophisticated physical constraints can be incorporated flexibly.

To demonstrate the usefulness of the constraints, the new iterative algorithm is used to invert simulated cloud tomography data from four ground-based microwave radiometers for a stratocumulus cloud. One by one, the smoothness, non-negativity, and double-side constraints are added into the retrieval algorithm. The retrieved cloud LWC field from the unconstrained method has many unrealistically high values interweaved with negative values and shows very poor performance in preserving the spatial patterns of the cloud LWC. The addition of the non-negativity and smoothness constraints helps to capture the location and spatial extent of the cloud, but gives poor retrievals at cloud edges. The incorporation of a double-side constraint (based on scaled adiabatic profiles) produces the best cloud tomography retrieval. It not only accurately captures the location and extent of the stratocumulus cloud, but also accurately reproduces the cloud edges. Furthermore, the structures of the singular values of the corresponding kernel matrices are analyzed, revealing that the improvement of the retrieval arises from the reduced range of the singular values.

The degree of the ill-posedness usually increases with the increasing dimensionality of the retrieval problem, as confirmed by numerous studies. In cloud tomography, the dimensionality of the retrieval problem is the number of unknowns in the system of equations (2), i.e., the number of pixels in the field to be retrieved. It was shown in Huang et al. (2007) that the ill-posedness of the tomographic retrieval problem is reduced when the total number of pixels is decreased (as a result of increased pixel size). However, with the increased pixel size fine-scale cloud structures tend to be lost. Therefore, reducing the dimensionality without losing fine-scale details would be highly desirable. This will be investigated in our future researches.

As a final comment, the methods present in this study should be valuable not only for cloud tomography, but also for many other ill-posed remote sensing problems. The C source code of these algorithms is available upon request to the authors.

### **Acknowledgement**

This research is supported by the DOE Atmosphere Radiation Measurement program under Contract DE-AC02-98CH10886. It is a pleasure to acknowledge Drs. Andy Volgemann, Mark Miller, and Richard Cederwall for their help with this research.

## References

- Ackerman, S.A., O.B. Toon, P.V. Hobbs (1995): A model for particle microphysics, turbulent mixing, and radiative transfer in the stratocumulus topped marine boundary layer and comparisons with measurements. *J. Atmos. Sci.*, 52, 1204-1236.
- Babb, D.M., J. Verlinde, and B.W. Rust (2000): The removal of turbulent broadening in Radar Doppler spectra using linear inversion with double-sided constraints. *J. Atmos. Sci. Ocean. Tech.*, 17, 1583-1595.
- Drake, J.F., and J. Warner (1988): A theoretical study of the accuracy of tomographic retrieval of cloud liquid with an airborne radiometer. *J. Atmos. Sci. Ocean. Tech.*, 5, 844-857.
- Fitzpatrick, B. G. (1991): Bayesian analysis in inverse problems. *Inverse Problems*, 7, 675-702.
- Hansen, P.C. (1992): Analysis of Discrete Ill-Posed Problems by Means of the L-Curve. *SIAM Review*, 34, 561-580.
- Hansen, P.C. (1998): *Rank deficient and ill-posed problems: Numerical aspects of linear inversion*. SIAM J. Appl. Math., Philadelphia, pp. 247.
- Huang, D., Y. Liu, and W. Wiscombe (2007): Determination of cloud liquid water distribution using 3D cloud tomography. *J. Geophys. Res.*, submitted.
- Lawson, C.L., and R.J. Hanson (1974): *Solving least squares problems*. Prentice-Hall, Englewood Cliffs, NJ, pp. 337.
- Liu, Y., W.P. Arnott, and J. Hallett (1999): Particle size distribution retrieval from multispectral optical depth: Influences of particle nonsphericity and refractive index. *J. Geophys. Res.*, 104:31, 753-762.
- Pierce, J.E., B.W. Rust (1985): Constrained least squares interval estimation. *SIAM J. Sci. Stat. Comput.*, 6, 670-683.

- Phillips, D.L. (1962): A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.*, 9, 84-97.
- Rodgers, C. D. (2000). *Inverse Methods for Atmospheric Sounding: Theory and Practice*. World Scientific Publishing Co., Singapore, pp. 200.
- Tikhonov, A.N. (1963): Solution of incorrectly formulated problems and the regularization method. *Sov. Math. Dokl.*, 4, 1035-1038.
- Tao, W. (2007): Cloud resolving modeling. *J. Meteor. Soc. Japan*, Special Issue of 125th Anniversary of Japan Meteorology Society, 85, 305-330.
- Twomey, S. (1977): *Introduction to the mathematics of inversion in remote sensing inversion and indirect measurements*. Elsevier, Amsterdam, pp. 243.
- Twomey, S. (1987): Iterative nonlinear inversion methods for tomographic problems. *J. Atmos. Sci.*, 44, 3544-3551.
- Warner, J., J.F. Drake, and P.R. Krehbiel (1985): Determination of cloud liquid water distribution by inversion of radiometric data. *J. Atmos. Sci. Ocean. Tech.*, 2, 293-303.
- Warner, J., J.F. Drake, and J.B. Snider (1986): Liquid water distribution obtained from coplanar scanning radiometers. *J. Atmos. Sci. Ocean. Tech.*, 3, 542-546.
- Warner, J., and J. F. Drake (1988): Field tests of an airborne remote sensing technique for measuring the distribution of liquid water in convective cloud. *J. Atmos. Oceanic Technol.*, 5, 833-843.

## Appendix: Connection between the constrained algorithms and Bayesian theorem

Bayesian theorem provides the general foundation for many inverse or forecast models. Let  $P(\mathbf{x})$ ,  $P(\mathbf{b})$  represent the corresponding Probability Density Functions (PDF) of the solution vector  $\mathbf{x}$  and the measurement vector  $\mathbf{b}$ , and  $P(\mathbf{x}|\mathbf{b})$  represent the conditional PDF of  $\mathbf{x}$  when  $\mathbf{b}$  is assigned to a certain value, Bayesian theorem states the following equality:

$$P(\mathbf{x} | \mathbf{b}) = P(\mathbf{b} | \mathbf{x}) P(\mathbf{x}) / P(\mathbf{b}) \quad (\text{A1})$$

The last term is independent of the solution vector  $\mathbf{x}$  and thus can be neglected when the cost function is constructed by means of maximizing the conditional probability  $P(\mathbf{x}|\mathbf{b})$ .

### *A. Smoothness constrained algorithm (Tikhonov regularization)*

The Tikhonov regularization method is closely related to the Bayesian maximum likelihood analysis (Fitzpatrick, 1991). Assuming the vectors  $\mathbf{x}$  and  $\mathbf{b}$  have Gaussian error distributions with constant error variances  $\sigma_{Lx}^2$  and  $\sigma_b^2$ , we have:

$$P(\mathbf{x}) = \frac{1}{(2\pi\sigma_{Lx}^2)^{n/2}} \exp\left(-\frac{\|\mathbf{Lx}\|_2^2}{2\sigma_{Lx}^2}\right), \quad (\text{A2})$$

$$P(\mathbf{b} | \mathbf{x}) = \frac{1}{(2\pi\sigma_b^2)^{n/2}} \exp\left(-\frac{\|\mathbf{Ax} - \mathbf{b}\|_2^2}{2\sigma_b^2}\right), \quad (\text{A3})$$

The optimal solution is obtained by maximizing the conditional probability:

$$P(\mathbf{x} | \mathbf{b}) \propto P(\mathbf{b} | \mathbf{x}) P(\mathbf{x}) \propto \exp\left(-\|\mathbf{Ax} - \mathbf{b}\|_2^2 - \frac{\sigma_b^2}{\sigma_{Lx}^2} \|\mathbf{Lx}\|_2^2\right). \quad (\text{A4})$$

Let  $\lambda \equiv \sigma_b^2 / \sigma_{Lx}^2$ , the cost function then can be written as:

$$J = \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{Lx}\|_2^2 \quad (\text{A5})$$

The cost function is now identical to that of the Tikhonov regularization (Eq. (5)), hence it is not surprising that the solution coincides with that of the Tikhonov regularization method.

### ***B. Double-side constrained algorithm***

In order to make the connection between the Bayesian theorem and the double-side constrained least squares, we assume that  $P(\mathbf{b}|\mathbf{x})$  takes the Gaussian format as in Eq. (A3) and  $P(\mathbf{x})$  satisfies the following distribution:

$$P(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \|\mathbf{R}\|_F^{1/2}} \exp\left[-(\mathbf{x} - \mathbf{x}_b)^T \mathbf{R}^{-1} (\mathbf{x} - \mathbf{x}_b) / 2\right], \quad (\text{A6})$$

where  $\|\cdot\|_F$  denotes the ‘‘entrywise’’ Frobenius norm;  $\mathbf{R}$  is the error covariance matrix, it is diagonal if the error in  $\mathbf{x}$  is independent with each other. From the Bayesian theorem, the conditional probability  $P(\mathbf{x}|\mathbf{b})$  is:

$$P(\mathbf{x} | \mathbf{b}) \propto \exp \left[ -\frac{\|\mathbf{Ax} - \mathbf{b}\|_2^2}{\sigma_b^2} - (\mathbf{x} - \mathbf{x}_b)^T \mathbf{R}^{-1} (\mathbf{x} - \mathbf{x}_b) \right]. \quad (\text{A7})$$

The cost function can be written as:

$$J = (\mathbf{Ax} - \mathbf{b})^T (\mathbf{Ax} - \mathbf{b}) / \sigma_b^2 + (\mathbf{x} - \mathbf{x}_b)^T \mathbf{R}^{-1} (\mathbf{x} - \mathbf{x}_b), \quad (\text{A8})$$

Solving  $\partial J / \partial \mathbf{x} = 0$  and letting  $\tau \equiv \sigma_b^2$  we get:

$$\mathbf{x}_{\text{Bay}} = (\mathbf{A}^T \mathbf{A} + \tau \mathbf{R}^{-1})^{-1} (\mathbf{A}^T \mathbf{b} + \tau \mathbf{R}^{-1} \mathbf{x}_b). \quad (\text{A9})$$

The solution of the double-side constrained least squares (Eq. (9)) is a special case of the Bayesian solution  $\mathbf{x}_{\text{Bay}}$  when  $\mathbf{R}$  takes the form of a diagonal matrix.

### ***C. Smoothness plus double-side constrained algorithm***

Again, we assume that  $P(\mathbf{b}|\mathbf{x})$  takes the Gaussian form as in Eq. (A3). With the smoothness and double-side constraints,  $P(\mathbf{x})$  satisfies the following distribution:

$$P(\mathbf{x}) \propto \exp \left[ -(\mathbf{x} - \mathbf{x}_b)^T \mathbf{R}^{-1} (\mathbf{x} - \mathbf{x}_b) - \frac{\|\mathbf{Lx}\|_2^2}{\sigma_{Lx}^2} \right], \quad (\text{A10})$$

Let  $\tau \equiv \sigma_b^2$  and  $\lambda \equiv \sigma_{Lx}^2$ , the cost function can be written as:

$$J = \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \tau(\mathbf{x} - \mathbf{x}_b)^T \mathbf{R}^{-1}(\mathbf{x} - \mathbf{x}_b) + \lambda \|\mathbf{Lx}\|_2^2. \quad (\text{A11})$$

The solution is obtained by solving  $\partial J / \partial \mathbf{x} = 0$  :

$$\mathbf{x}_{\text{Bay}} = \left( \mathbf{A}^T \mathbf{A} + \lambda \mathbf{L}^T \mathbf{L} + \tau \mathbf{R}^{-1} \right)^{-1} \left( \mathbf{A}^T \mathbf{b} + \tau \mathbf{R}^{-1} \mathbf{x}_b \right). \quad (\text{A12})$$

## Figures

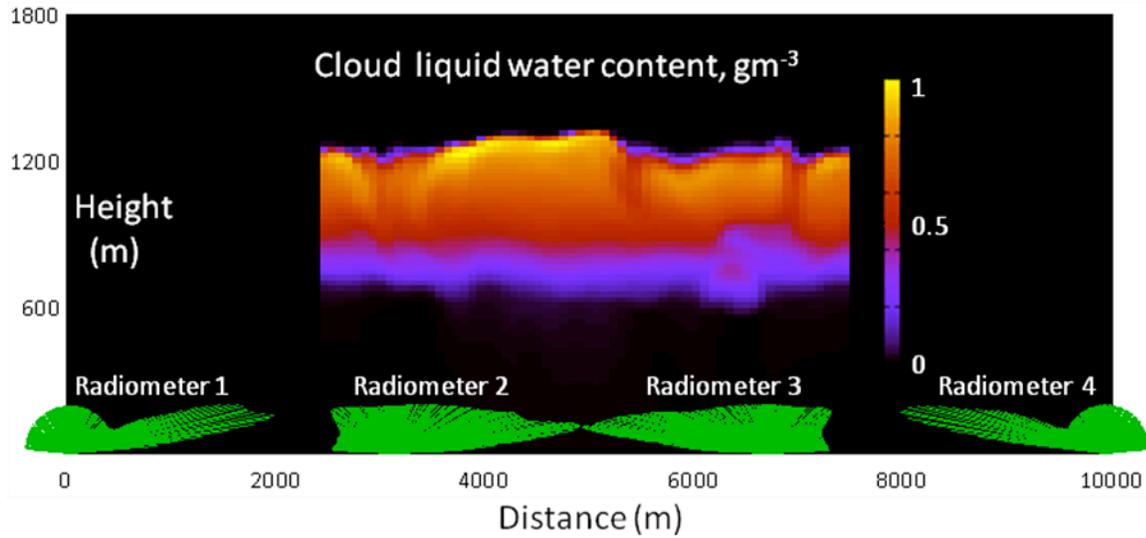


Figure 1. An example of ground-based cloud tomography. The cloud is a two-dimensional 5 Km wide by 1.5 Km high slice taken from the DHARMA large eddy simulation model. Four scanning radiometers spaced 3.3 km apart with 0.3 K noise level are arranged in a line 10 km long. Each radiometer scans the upper plane to within  $5^\circ$  of the ground; the scans are every  $0.4^\circ$  in angle, giving a total of 800 rays hitting the cloudy area. The lengths of the scan lines from each radiometer are proportional to the simulated brightness temperatures in that direction. The atmospheric background is assumed to be 20 K.

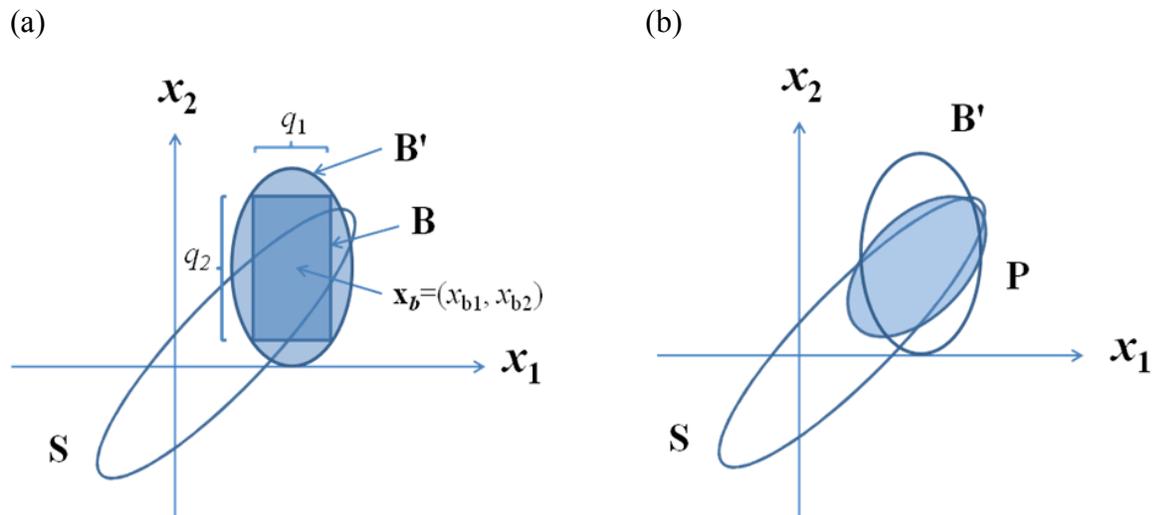


Figure 2. Schematic illustration of applying a double-side constraint in a two-dimensional case. The  $S$  ellipsoid contains the ordinary least-squares solutions satisfying certain noisy measurements; it is elongated along the eigenvector corresponding to the smallest singular value of the kernel matrix. (a) The upper and lower bounds indicated by bounding box  $B$  (width  $q_1$ , height  $q_2$ , centered at  $\mathbf{x}_b = (x_{b1}, x_{b2})$ ) are used to constrain the solution ellipsoid  $S$ . Also shown is the envelope ellipsoid  $B'$  of the box  $B$ . (b) A new solution ellipsoid  $P$  (shaded) is formed by the convex combination of the ellipsoids  $S$  and  $B'$ .

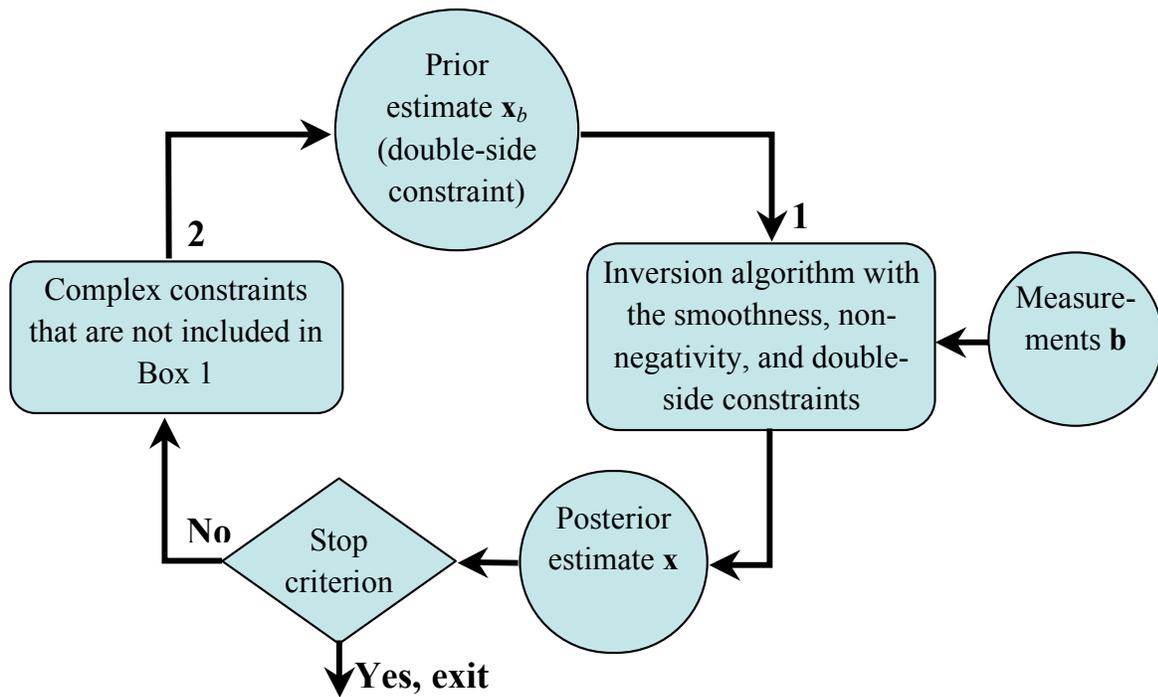
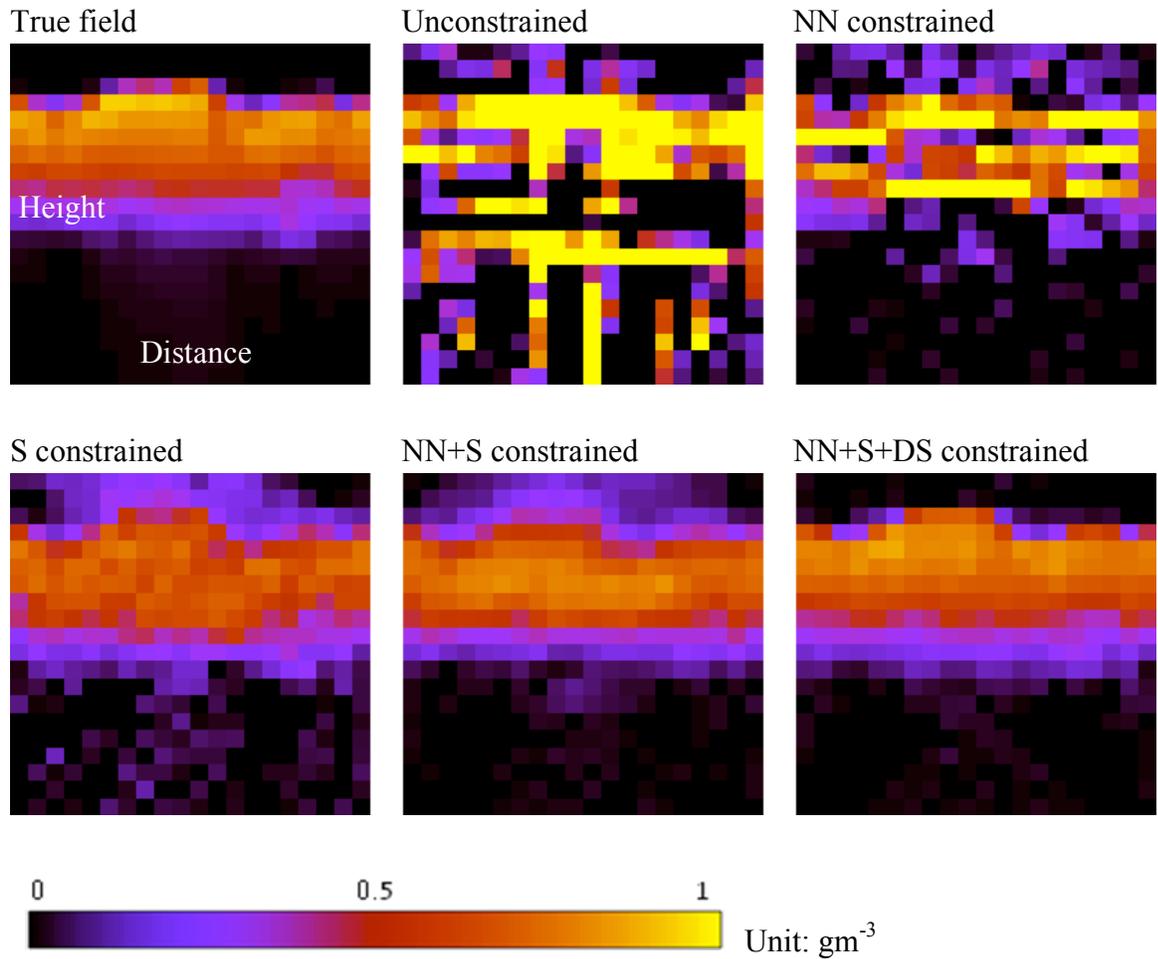
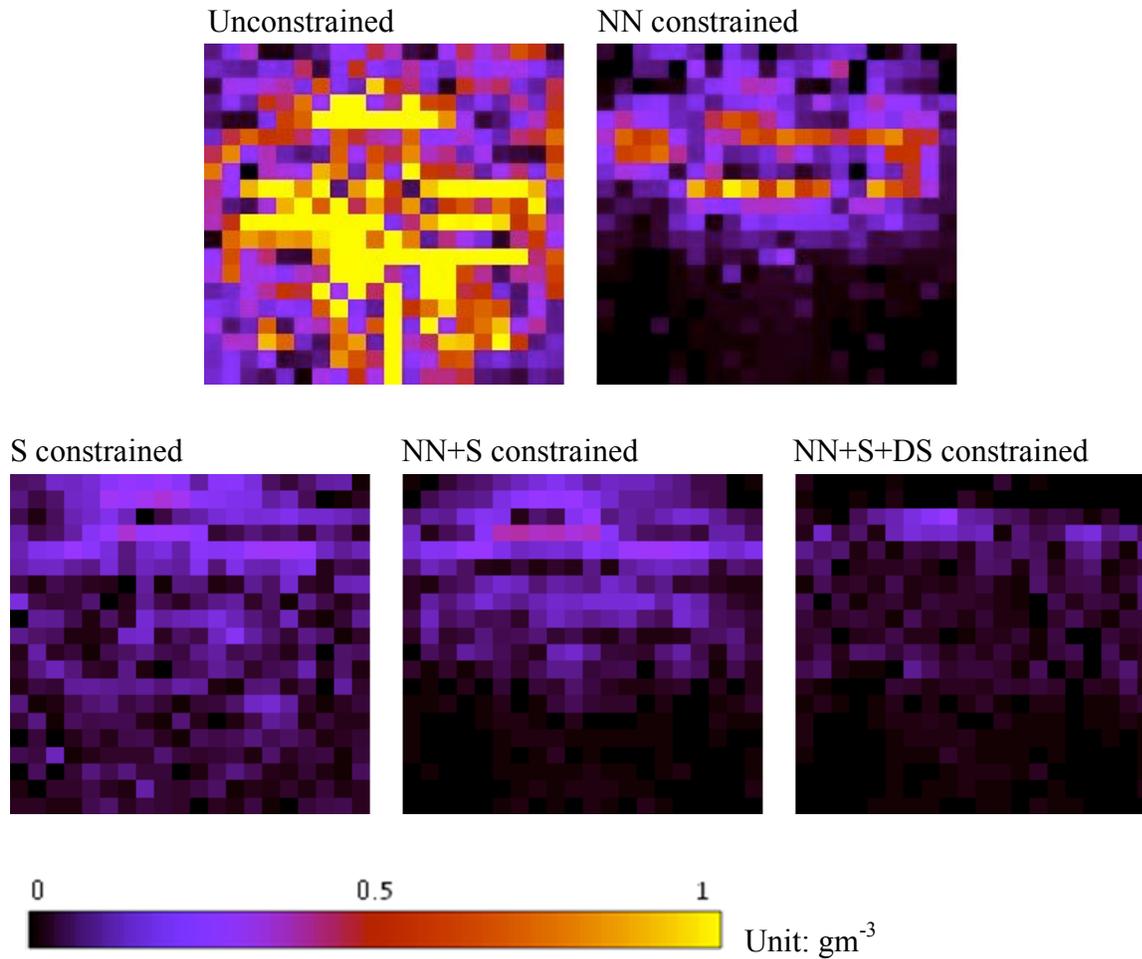


Figure 3. Diagram of the iterative algorithm. The algorithm has two steps: (1) Compute the posterior estimate  $\mathbf{x}$  based on the prior estimate  $\mathbf{x}_b$  and measurements  $\mathbf{b}$ ; (2) Adjust the posterior estimate  $\mathbf{x}$  so that the constraints in Box 2 are satisfied, and use the modified  $\mathbf{x}$  as the prior estimate  $\mathbf{x}_b$  for the next iteration. The iteration is terminated when the posterior estimate meets a Cauchy stop criterion involving the closeness of two successive iterates.



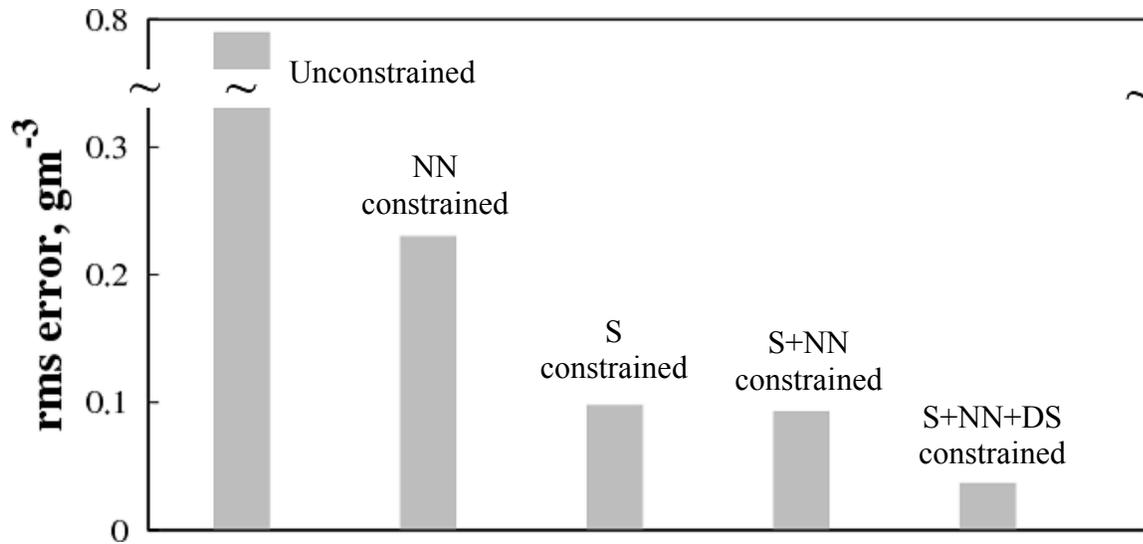
LS: Least Squares; NN: Non-Negativity; S: Smoothness; DS: Double-side

Figure 4. The retrieved LWC fields from the cloud tomography simulation shown in Figure 1 using the unconstrained least squares, non-negativity constrained least squares (NNLS), smoothness constrained least squares (Tikhonov regularization), smoothness constrained NNLS (SCNNLS), and the iterative algorithm combining the non-negativity, smoothness, and double-side constraints (DSCNNL). The true field, 5 Km wide and 1.5 Km high , is shown as a reference.



LS: Least Squares; NN: Non-Negativity; S: Smoothness; DS: Double-side

Figure 5. The retrieval error fields computed as the absolute difference between the true and retrieved LWC fields shown in Figure 4. The value of the non-negativity constraint can be shown by comparing the error fields of “unconstrained” with “NN constrained”, or “S constrained” with “S+NN constrained”. Note how the non-negativity constraint significantly reduces the retrieval error below the cloud bottom. The values of the smoothness and double-side constraints can be assessed in the same manner: the smoothness constraint significantly reduces the error inside the cloud while the double-side constraint greatly improves the retrievals around the cloud top edge.



NN: Non-Negativity; S: Smoothness; DS: Double-side

Figure 6. The rms errors of the cloud LWC fields retrieved with different constraints (see Figure 4) for a four-radiometer tomography simulation. The retrieval error decreases when more constraints are added. The smoothness constraint is the most effective of the single constraints, but a better result is obtained by combining constraints.

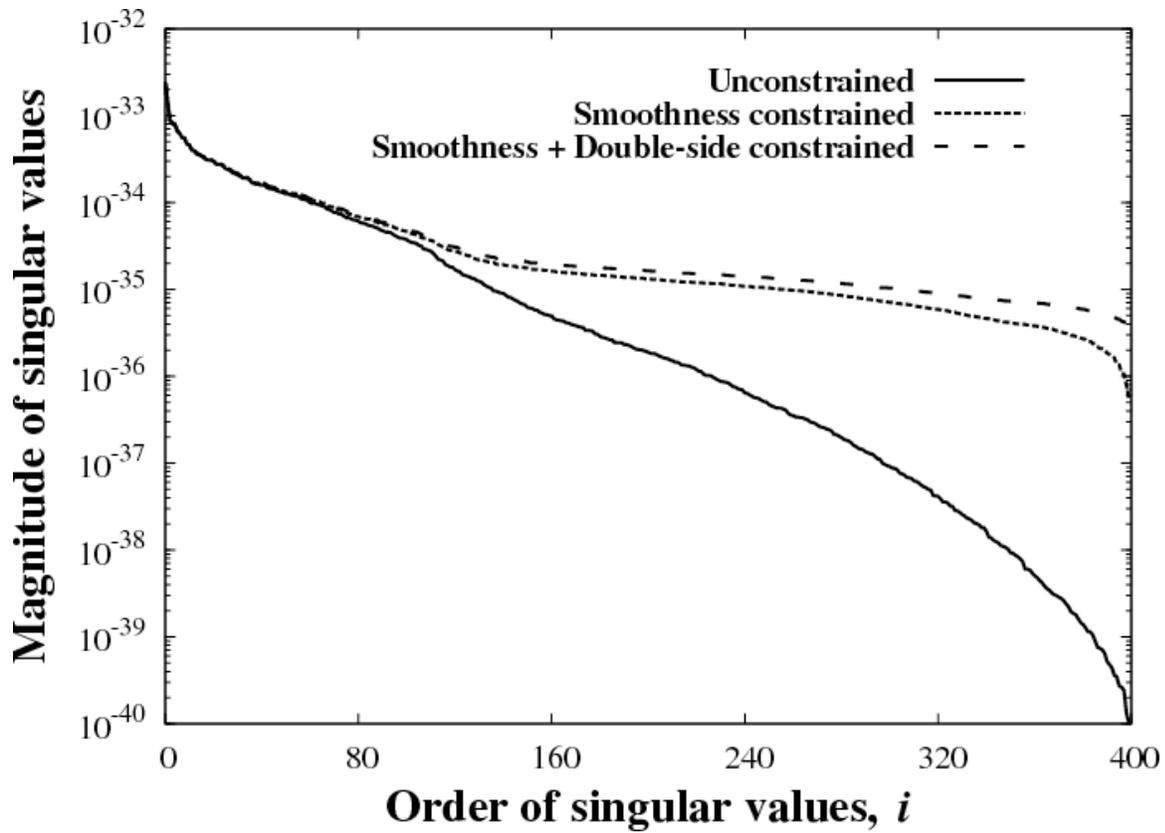


Figure 7. Decreasing trend of the singular values for the unconstrained, smoothness constrained, and smoothness plus double-side constrained least squares for the cloud tomography retrieval problem shown in Figure 1. Retrieval errors owing to measurement and rounding errors become large when there are very small singular values. The constrained kernel matrices become better conditioned, leading to a more accurate solution, when the very small singular values are eliminated by applying constraints.