Regularization for inverse models in remote sensing

Yanfei Wang
Chinese Academy of Sciences, China

Abstract
Quantitative remote sensing is an appropriate way to estimate atmospheric parameters and structural parameters and spectral component signatures of Earth surface cover type. Since the real physical system that couples the atmosphere, water and the land surface is complicated, its description requires a comprehensive set of parameters, so any practical physical model can only be approximated by a limited mathematical model. The pivotal problem for quantitative remote sensing is inversion. Inverse problems are typically ill-posed; they are characterized by: (C_1) the solution may not exist; (C_2) the dimension of the solution space may be infinite; (C_3) the solution is not continuous with variations of the observations. These issues exist for nearly all inverse problems in geosciences and quantitative remote sensing. For example, when the observation system is band-limited or sampling is poor, i.e. few observations are available or directions are poorly located, the inversion process would be underdetermined, which leads to a multiplicity of the solutions, the large condition number of the normalized system, and significant noise propagation. Hence (C_2) and (C_3) would be the difficulties for quantitative remote sensing inversion. This paper will address the theory and methods from the viewpoint that the quantitative remote sensing inverse problems can be represented by kernel-based operator equations and solved by coupling regularization and optimization methods. In particular, I propose sparse and non-smooth regularization and optimization techniques for solving inverse problems in remote sensing. Numerical experiments are also made to demonstrate the applicability of our algorithms.

I Introduction

Model-based inversion is quite important for quantitative remote sensing. Here, model-based inversion mainly refers to using physical or empirically physical models to infer unknown but relevant parameters. Hundreds of models related to atmosphere, vegetation and radiation have been established during recent decades (Liang, 2004), and model-based inversion in geophysical and atmospheric sciences is well understood. However, model-based inverse problems for earth surface studies have received attention only in recent years. Compared to

modelling, model-based inversion is still in the exploration stage (Wang et al., 2009c) because intrinsic difficulties exist in the application of a priori information, inverse strategy and inverse algorithms. The development of hyperspectral and multiangular remote sensors has enhanced exploration and provided us more spectral and

Corresponding author:
Key Laboratory of Petroleum Resources Research, Institute of Geology and Geophysics, Chinese Academy of Sciences, PO Box 9825, Beijing, 100029, China
Email: yfwang@mail.iggcas.ac.cn

spatial information than before. However, to utilize the information to solve problems faced in quantitative remote sensing is still a challenge. Remote sensing inversion for problems in different areas is being paid more attention. In a series of international study projections, such as International Geosphere-Biosphere Programme (IGBP), World Climate Research Programme (WCRP) and NASA's Earth Observing System (EOS), remote sensing inversion has become a focal point of study.

Model-based remote sensing inversion usually requires solving optimization problems with different constraints. Therefore, knowledge of how to incorporate the methods developed in operational research field into the remote sensing inversion field is needed. In quantitative remote sensing, since the real physical system that couples the atmosphere and the land surface is very complicated, sometimes it requires a comprehensive set of parameters to describe such a system, so any practical physical model can only be approximated by a model which includes only a limited number of the most important parameters that capture the major variation of the real system. Generally speaking, a discrete forward model to describe such a system is in the form:

$$y = h(z, S) \qquad (1.1)$$

where y is single measurement, z is a vector of controllable measurement conditions such as wave band, viewing direction, time, Sun position, polarization, and so forth, S is a vector of state parameters of the system approximation, and h is a function which relates z with S, which is generally non-linear and continuous.

With the ability of receiver sensors to acquire multiple bands, multiple viewing directions and so on, while keeping S essentially the same, we obtain the following inhomogeneous equation:

$$y = h(z, S) + n \qquad (1.2)$$

where y is a vector in $\mathbb{R}^M$, which is an M dimensional measurement space with M values corresponding to M different measurement conditions, $n \in \mathbb{R}^M$ is the vector of random noise with same vector length M. Assume that there are m undetermined parameters that need to be recovered. Clearly, if $M = m$ and the system is linearly independent, (1.2) is a determined system, so it is not difficult to develop some suitable algorithms to solve it. If more observations can be collected than the existing parameters in the model (Verstraete et al., 1996), i.e. $M > m$, the system (1.2) is overdetermined. In this situation, the traditional solution does not exist. We must define its solution in some other meanings, for example, the least squares error (LSE) solution. However, for physical models with complete parameters for a single band, it is questionable whether remote sensing inversion can be overdetermined in the foreseeable future (Li et al., 1998). Therefore, inversion problems in geosciences seem to be always ill-posed (underdetermined) in some sense. Moreover, faced with an underdetermined system, the solution will be non-unique. In other words, there will be infinite solutions in the null space of solutions. To tackle the ill-posedness, proper a priori information must be found and involved into the inversion procedure to provide a useful approximate solution (Li et al., 2001).

Developed methods in literature for quantitative remote sensing inversion are mainly statistical methods with several variations from Bayesian inference (Combal et al., 2003; Pokrovsky and Roujean, 2002, 2003; Pokrovsky et al., 2003; Quaife and Lewis, 2010) and regularization methods (Phillips, 1962; Twomey, 1975; Wang et al., 2007a, 2008). According to Combal et al. (2003), there are three classes of algorithms for solving the inverse problems in remote sensing, including minimization algorithms, lookup tables (LUT) and neural networks (NNT). Many efforts have been made to invert the radiative transfer models using these algorithms (Goel and Strebel, 1983; Jacquemoud and Baret, 1993; Kuusk, 1991; Weiss et al., 2000) and to

validate these algorithms (Combal et al., 2003; Knyazikhin et al., 1998a, 1998b; Privette et al., 1996; Weiss and Baret, 1999). We pay attention to the minimization algorithms in this paper. Minimization algorithms based on filtering techniques have been extensively studied (Cohn et al., 1994; Samain et al., 2008; Sarkka et al., 2004). We will show in this paper that the Tikhonov regularization can be also expressed as filtering methods. Using kernel expression, we analyse solution theory and methods for quantitative remote sensing inverse problems from an algebraic point of view. We first review regularization methods for retrieval of parameters, and then propose sparse/non-smooth inversion by total variation and sparse inversion in $l_p - l_q$ ($p = 2, q = 1$) spaces and introduce advanced optimization techniques. These methods, as far as we know, are novel to literature in the earth sciences.

The outline of the paper is as follows: in section II.1, we formulate the inverse problems by operator equations of the first kind and define the ill-posed nature of the problems. Sections II.2 and II.3 introduce two typical inverse problems in remote sensing; one is the linear kernel-based bidirectional reflectance distribution function (BRDF) model inversion, and another is the atmospheric aerosol particle size distribution function retrieval problem. Both problems are of great importance for calibration and for parameters retrieval. The ill-posed nature of both problems is explained in the respective sections. In section III, a simple linear algebraic system is introduced to explain the ill-posedness of the finite linear inverse problems. In section IV, the regularization theory and solution techniques for ill-posed quantitative remote sensing inverse problems are described. Beginning from the Bayesian inference, section IV.1 discusses constrained optimization; section IV.2 fully extends the Tikhonov regularization; section IV.3 discusses the conceptual regularization scheme formulated in the Bayesian statistical inference; in section IV.4, the direct

regularization method based on spectrum decomposition for equality-constrained problem is introduced, and equivalence to the standard Tikhonov regularization is established by introducing filtering functions. In section V, the sparse regularization and optimization theory and solving methods are discussed for finding an optimized solution of a minimization model. Section V.1 develops a total variation method for the ill-posed inverse problem; section V.2 discusses sparse and non-smooth inversion in $l_p - l_q$ space; section V.3 discusses some computational issues in inversion. In section VI, the applications of regularization and optimization solution methods for retrieval of land surface parameters and aerosol particle size distribution functions are presented. Finally, in section VII, some concluding remarks are given.

II Linear kernel-based models and ill-posedness

I General model of the first kind operator equation

Inverse problems are usually formulated as operator equations of the first kind. We consider the first kind operator equation in the general form:

$$A(x) = y \qquad (2.1)$$

which is an appropriate expression for an observing system, with A the response function (linear or non-linear) from Hilbert space X to Hilbert space Y, x the unknown input and y the observed data. Particularly, if A is a linear mapping, we will denote the response system as:

$$Ax = y \qquad (2.2)$$

which is clearly a special case of (2.1). We will also use A as an operator in infinite spaces sometimes, and a matrix sometimes. We assume that readers can readily recognize them.

The problem (2.1) is said to be properly posed or well-posed in the sense that it has the following three properties:

(C1) There exists a solution of the problem, i.e. existence;

(C2) There is at most one solution of the problem, i.e. uniqueness;

(C3) The solution depends continuously on the variations of the right-hand side (data), i.e. stability.

The condition (C$_1$) can be easily fulfilled if we enlarge the solution space of the problem (2.1). The condition (C$_2$) is seldom satisfied for many indirect measurement problems. More than one solution may be found for the problem (2.1) and information about the model is missing. In this case, a priori knowledge about the solution must be incorporated in the model. Stability is the most important requirement. If the problem (2.1) lacks the property of stability, then the computed solution will be far from the true solution since the practically obtained solution is contaminated by unavoidable errors. Therefore, there is no way to overcome this difficulty unless additional information about the solution is available. Again, a priori knowledge about the solution should be involved.

If equation (2.2) is well-posed, then A has a well-defined, continuous inverse operator A^{-1}. In particular, $A^{-1}(A(x)) = x$ for any $x \in X$ and Range$(A) = Y$. In this case both the algebraic nature of the spaces and the topologies of the spaces are ready to be employed.

2 Linear kernel-based BRDF model

As an example of linear kernel-based models in geosciences, we consider a well-refereed model. It is based on the assumption that the anisotropy of the land surface can be described by the bidirectional reflectance distribution function (BRDF). With the progress of the multiangular remote sensing, it seems that the BRDF models

can be inverted to estimate structural parameters and spectral component signatures of Earth surface cover type (Roujean et al., 1992; Strahler et al., 1994). The state-of-the-art of BRDF is the use of the linear kernel-based models, mathematically described as the linear combination of the isotropic kernel, volume scattering kernel and geometric optics kernel. Information extraction on the terrestrial biosphere and other problems for retrieval of land surface albedos from satellite remote sensing have been considered by many authors in recent years; see, for instance, the survey papers on the kernel-based bidirectional reflectance distribution function (BRDF) models (Combal et al., 2003; Pokrovsky and Roujean, 2002, 2003; Pokrovsky et al., 2003; Quaife and Lewis, 2010) and references therein. Computational stability is characterized by the algebraic operator spectrum of the kernel-matrix and the observation errors. Therefore, the retrieval of the model coefficients is of great importance for computation of the land surface albedos.

The linear kernel-based BRDF model can be described as follows (Roujean et al., 1992):

$$f_{iso} + k_{vol}(t_i, t_v, \phi)f_{vol} + k_{geo}(t_i, t_v, \phi)f_{geo} = r(t_i, t_v, \phi)$$

$$(2.3)$$

where r is the bidirectional reflectance; the kernels k_{vol} and k_{geo} are so-called kernels, i.e. known functions of illumination and of viewing geometry which describe volume and geometric scattering, respectively; t_i and t_v are the zenith angle of the solar direction and the zenith angle of the view direction, respectively; ϕ is the relative azimuth of sun and view direction; and f_{iso}, f_{vol} and f_{geo} are three unknown parameters to be adjusted to fit observations. Physically, f_{iso}, f_{vol} and f_{geo} are closely related to the biomass such as leaf area index (LAI), Lambertian reflectance, sunlit crown reflectance, and viewing and solar angles. The vital task then is to retrieve appropriate values of the three parameters.

Generally speaking, the BRDF model includes kernels of many types. However, it was demonstrated that the combination of RossThick (k_{vol}) and LiSparse (k_{geo}) kernels had the best overall ability to fit BRDF measurements and to extrapolate BRDF and albedo (e.g. Li et al., 1999; Privette et al., 1997; Wanner et al., 1995). A suitable expression for the RossThick kernel k_{vol} was derived by Roujean et al. (1992). It is reported that the LiTransit kernel $k_{Transit}$, instead of the kernel k_{geo}, is more robust and stable than LiSparse non-reciprocal kernel and the reciprocal LiSparse kernel k_{sparse} (LiSparseR) where the LiTransit kernel and the LiSparse kernel are related by:

$$k_{Transit} = \begin{cases} k_{sparse}, & B \le 2, \\ \frac{2}{B}k_{sparse}, & B > 2, \end{cases}$$

and B is given by $B := B(t_i, t_v, \phi) = -O(t_i, t_v, \phi) + \sec t'_i + \sec t'_v$ in Li et al. (2000). More detailed explanation about O and t' in the definition of $k_{Transit}$ can be found in Wanner et al. (1995).

To use the combined linear kernel model, a key issue is to numerically solve the inverse model in a stable way. However, it is difficult to do in practical applications due to the ill-posed nature of the inverse problem.

a Ill-posedness. Note that (2.3) is a linear model in finite spaces, therefore it is easy to rewrite it into a finite rank operator equation:

$$Kx = y \qquad (2.4)$$

by setting $x = [f_{iso}, f_{vol}, f_{geo}]^T$ and $y = [y_j]$ with the entries $y_j = r_j(t_i, t_v, \phi)$, where y is the measurement data. The inverse problem is to recover the model parameters x given the limited measurement data y.

Numerically, the discrete ill-posedness lies in that the operator K may be inaccurate (can only be approximated), and the model is usually underdetermined if there are few observations or poor directional range, or the observations are highly linearly dependent and noisy. For example, a single angular observation may lead to an underdetermined system whose solutions are infinite (the null space of the kernel operator contains non-zero vectors) or the system has no solution (the rank of the coefficient matrix is not equal to the augmented matrix). In practice, random uncertainty in the reflectances sampled translates into uncertainty in the BRDF and albedo. We note that noise inflation depends on the sampling geometry alone. For example, for MODIS and MISR sampling they vary with latitude and time of year; but for kernel-based models they do not depend on wavelength or the type of BRDF viewed. Therefore, the random noise in the observation (BRDF) and the small singular values of K control the error propagation.

So far, statistical methods and algebraic methods have been developed for solving this inverse problem (e.g. Combal et al., 2003; Pokrovsky and Roujean, 2002, 2003; Quaife and Lewis, 2010; Wang et al., 2007b, 2008). We will describe these methods and introduce recent advances in following sections.

3 Aerosol particle size distribution function inversion

It is well known that the characteristics of the aerosol particle size, which can be represented as a size distribution function in the mathematical formulation, say $n(r)$, play an important role in climate modelling due to their uncertainty (Houghton et al., 1996). So, the determination of particle size distribution function becomes a basic task in aerosol research (Bockmann, 2001; Bockmann and Kirsche, 2006; Bohren and Huffman, 1983; Davies, 1974; McCartney, 1976; Twomey, 1977). The particle size distribution is usually retrieved by extinction measurements using a sun-photometer. The attenuation of aerosols can be written as the integral equation of the first kind:

$$\tau(\lambda) = \int_0^\infty \pi r^2 Q_{ext}(r,\lambda,\eta)n(r)dr + \varepsilon(\lambda)$$

(2.5)

where r is the particle radius; $n(r)$ is the columnar aerosol size distribution (i.e. the number of particles per unit area per unit radius interval in a vertical column through the atmosphere); η is the complex refractive index of the aerosol particles; λ is the wavelength; $\tau(\lambda)$ is the aerosol optical thickness (AOT); $\varepsilon(\lambda)$ is the error/noise; and $Q_{ext}(r,\lambda,\eta)$ is the extinction efficiency factor from Mie theory. The AOT can be obtained from the measurements of the solar flux density with sun-photometers; therefore, one can retrieve the size distribution by the inversion of AOT measurements through the above equation. Equation (2.5) can be simply expressed by the operator equation:

$$\tau(\lambda) = (Kn)(\lambda) + \varepsilon(\lambda)$$

(2.6)

where the operator K is specified by the kernel function $k(r,\lambda,\eta) = \pi r^2 Q_{ext}(r,\lambda,\eta)$.

a Ill-posedness. The particle size distribution model (2.6) is a linear model in infinite spaces. It is clear to see that the kernel function $k(r,\lambda,\eta)$ is continuous, differentiable and bounded on $[0,b]$; therefore the operator K is compact. According to the definition of the ill-posedness and the operator theory (Tikhonov and Arsenin, 1977; Xiao et al., 2003; Yosida, 1999), the ill-posedness of the aerosol problem is self-evident because at least one of the three items for well-posed problems is violated.

III A simple ill-posed mathematical example

Let us consider a simple linear algebraic system:

$$Az = u$$

(3.1)

where $A \in \mathbb{R}^{2\times2}, z \in \mathbb{R}^2, u \in \mathbb{R}^2, A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, z = \begin{bmatrix} x \\ y \end{bmatrix}, u = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. It is clear to see that A is positive

semi-definite, and the equation has an infinite number of solutions. However, the normal solution (x^*,y^*) can be found, i.e. $(x^*,y^*) = (0.5,0.5)$. If we perturb the matrix a little, i.e. $A = A + hE$, where E is a diagonal matrix in the form of $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, h is a typically small number and $h \neq 0$, in this case there is a unique solution of (3.1), i.e. $(x,y) = (0,1)$.

Let us look at another example: consider the same linear system (3.1), where $A \in \mathbb{R}^{2\times2}$, $z \in \mathbb{R}^2, u \in \mathbb{R}^2, A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, z = \begin{bmatrix} x \\ y \end{bmatrix}, u = \begin{bmatrix} 1/2 \\ 2/3 \end{bmatrix}$. It is clear that the equation has no solution at all. However, the normal solution (x^*,y^*) can again be found, i.e. $(x^*,y^*) = (0.5,0.5)$.

The simple example indicates that if the nearest two observations are linearly dependent, the number of solutions may be infinite, non-unique or non-existent, i.e. the ill-posedness is fulfilled. In such case, we can only obtain the solution with minimum energy. To tackle the difficulties, we have to resort to regularization.

IV Regularization methods

For ease of notation, from this section to the end of the paper, we use $Kx = y$ to represent the linear system, where K refers to the operator either from kernel-based BRDF model (2.3) or from the aerosol particle attenuation equation (2.5).

Bayesian inference provides a conceptually simple process for updating uncertainty in the light of evidence. Initial beliefs about some unknown quantity are represented by a priori distribution. According to Bayesian rule, the estimated parameter could be expressed as a posteriori probability $p(x|y)$:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

(4.1)

where $p(x)$ is the a priori probability distribution about the parameter x, $p(y)$ is the a priori probability distribution about the data, and $p(y|x)$ describes the probability distribution function

about the discrepancy between the model and the data. Using logarithm, we obtain the following equation:

$$\log p(x|y) = \log p(y|x) + \log p(x) - \log p(y) \quad (4.2)$$

Therefore, the maximum a posteriori probability can be obtained by minimizing the following function:

$$J(x) := -\log p(x|y) = -\log p(y|x) - \log p(x) \quad (4.3)$$

Information in the data is expressed by the likelihood function $L(x|y) = \log p(x|y)$. Therefore, minimization of the function $J(x)$ is equivalent to maximization of the likelihood function $L(x|y)$. The a priori distribution $p(x)$ and the likelihood function are then combined to obtain the posterior distribution for the quantity of interest. The a posteriori distribution expresses our revised uncertainty in light of the data, in other words, an organized appraisal in the consideration of previous experience. For example, we can choose $p(y|x)$ as:

$$p(y|x) = C_q \exp\left(-q^{-1}\frac{\|y - Kx\|_{l_q}^q}{(\sigma_q)^q}\right) \quad (4.4)$$

where C_q is a constant related to the number q, σ_q is the variance related to data and $q \geq 0$. In the following context, we will see how the Bayesian inference is related.

1 Applying constraints to solutions

According to Bayesian inference theory, for effective inversion of the ill-posed kernel-driven model, we have to impose an a priori constraint to the interested parameters. This leads to solving a constrained minimization problem:

$$\min J(x), \text{ s.t.} Kx = y, \Delta_1 \leq c(x) \leq \Delta_2 \quad (4.5)$$

where $J(x)$ denotes an object function, which is a function of x, $c(x)$ is the constraint to the solution x, and Δ_1 and Δ_2 are two constants which specify the bounds of $c(x)$. Usually, $J(x)$ is chosen as the norm of x with different scales. If the parameter x comes from a smooth function, then $J(x)$ can be chosen as a smooth function, otherwise $J(x)$ can be non-smooth.

The constraint $c(x)$ can be smooth (e.g. Sobolev stabilizer) or non-smooth (e.g. total variation or l_q norm ($q \neq 2$) based stabilizer). A generically used constraint is the smoothness. It assumes that physical properties in a neighbourhood of space or in an interval of time present some coherence and generally do not change abruptly. Practically, we can always find regularities of a physical phenomenon with respect to certain properties over a short period of time (Wang et al., 2007b, 2008). The smoothness has been one of the most popular a priori assumptions in applications. The general framework is the so-called regularization which will be explained in the next section.

2 Regularization: incorporate a priori information naturally

Most inverse problems in real environment are generally ill-posed. Regularization methods are widely used to solve such ill-posed problems. The complete theory for regularization was developed by Tikhonov and his colleagues (Tikhonov and Arsenin, 1977; Tikhonov et al., 1995). For the discrete model (2.4), we suppose the right-hand side y is the measurement with noise which represents the bidirectional reflectance or the AOT. The Tikhonov regularization method is to solve a regularized minimization problem:

$$J_{\text{Tikh}}{}^\alpha(x) := \|Kx - y\|_2^2 + \alpha\|D^{1/2}x\|_2^2 \rightarrow \min \quad (4.6)$$

instead of solving:

$$J_{\text{lse}}(x) := \|Kx - y\|_2^2 \rightarrow \min \quad (4.7)$$

In (4.6), α is the regularization parameter and D is a positively (semi-)definite operator.

By a variational process, the minimize of (4.6) satisfies:

$$K^T Kx + \alpha Dx = K^T y \qquad (4.8)$$

or can be written as $x(\alpha) = (K^T K + \alpha D)^{-1} K^T y$. It is clear to see that the Tikhonov regularization is a special case of Bayesian inference if we take $q = 2$ in equation (4.4) and choose the a priori distribution $p(x)$ as the 2-norm. The operator D is a scale matrix which imposes smoothness constraint to the solution x. The scale operator D and the regularization parameter α can be considered as some kind of a priori information; see Wang et al. (2008, 2009c) for details.

Choosing the regularization parameter α can be performed in an a priori way or an a posteriori way. The a priori way means setting fixed values of parameter α in calculations, while the a posteriori way means determining an optimal regularization parameter. We recall a well-known discrepancy principle discussed in Wang et al. (2008) that the optimal value of the regularization parameter should be the solution of the non-linear equation:

$$\varphi(\alpha) = (Kx(\alpha) - y, Kx(\alpha) - y) - \delta^2 \quad (4.9)$$

where $x(\alpha)$ indicates that the unknown parameter is related to the regularization parameter, and δ is the upper bound of error in the observation. Once an optimal value α^*, i.e. the root of the above equation, is found, the optimal parameter $x(\alpha^*)$ is obtained. The root is easy to find through Newton's iterative method:

$$\alpha_{k+1} = \alpha_k - \frac{\varphi(\alpha_k)}{\varphi'(\alpha_k)} \qquad (4.10)$$

or using the more efficient root-finding method:

$$\alpha_{k+1} = \alpha_k - \frac{2\varphi(\alpha_k)}{\varphi'(\alpha_k) + \sqrt{\varphi'(\alpha_k)^2 - 2\varphi(\alpha_k)\varphi''(\alpha_k)}}$$

$$(4.11)$$

Remark 4.1: It deserves attention that in atmospheric science many other regularization

methods are developed (Bockmann, 2001; Bockmann and Kirsche, 2006; Voutilainenand and Kaipio, 2000). Meanwhile, there is a similar method, called the Phillips-Twomey's regularization (Phillips, 1962; Twomey, 1975), which is based on solving the problem (Wang et al., 2006a):

$$\min_x \ Q(x), \text{s.t.} \|Kx - y\| = \Delta \qquad (4.12)$$

where $Q(x) = (Dx, x)$ with D a pre-assigned scale matrix and $\Delta > 0$. It is clear that Phillips-Twomey's regularization shares similarity with Tikhonov's regularization and can be written in consistent form. The form of the operator D is determined by the norm of the second differences $\sum_{i=2}^{N-1}(x_{i-1} - 2x_i + x_{i+1}^2)$. However, the matrix D is badly conditioned and will cause numerical instability. We consider choosing a proper D from Sobolev space, i.e. in (4.6) the matrix D is determined from discretization of $\|x\|_{W^{1,2}}^2$, where the inner product of two functions $\xi(t)$ and $\zeta(t)$ in $W^{1,2}$ space is defined by:

$$(\xi(t), \zeta(t))_{W^{1,2}} = \int_\Omega (\xi(t)\zeta(t) + \sum_{i,j} \frac{\partial \xi}{\partial t_i} \frac{\partial \zeta}{\partial t_j}) dt_1 dt_2 \cdots dt_n$$

According to the Sobolev imbedding theorem, this form of regularization possesses good convergence property and stability and can yield a smooth solution (Tikhonov and Arsenin, 1977; Xiao et al., 2003).

3 Statistical regularization

The role of Bayesian statistics is very similar to the role of regularization (Wang et al., 2007b). Now we establish the relationship between the Bayesian estimation and the regularization. A continuous random vector x is said to have a Gaussian distribution if its joint probability distribution function has the form:

$$p_x(x; \mu, C_x) = \frac{1}{\sqrt{(2\pi)^N \det(C_x)}}$$
$$\exp(-\tfrac{1}{2}(x-\mu)^T C_x^{-1}(x-\mu)) \tag{4.13}$$

where $x, \mu \in \mathbb{R}^{N:}$, C_x is an N-by-N symmetric positive definite matrix, and $\det(\cdot)$ denotes the matrix determinant. The mean is given by $E(x) = \mu$ and the covariance matrix is $\text{cov}(x) = C_x$.

Suppose $y = Kx + n$ is a Gaussian distribution with mean Kx and covariance C_n, where C_n is the noise covariance of the observation noise and model inaccuracy. Then by (4.13) we obtain:

$$p(y|x) = \frac{1}{\sqrt{(2\pi)^M \det(C_n)}}$$
$$\exp(-\tfrac{1}{2}(y-Kx)^T C_n^{-1}(y-Kx)) \tag{4.14}$$

From (4.13), the prior probability distribution is given by:

$$p(x) = \frac{1}{\sqrt{(2\pi)^N \det(C_x)}} \exp(-\tfrac{1}{2}x^T C_x^{-1} x)$$

By Bayesian statistical inference and the above two equations, we obtain an a posteriori log likelihood function:

$$L(x|y) = \log(p(x|y)) = -\tfrac{1}{2}(y-Kx)^T$$
$$C_n^{-1}(y-Kx) - \tfrac{1}{2}x^T C_x^{-1} x + \varsigma \tag{4.15}$$

where ς is constant with respect to x. The maximum a posteriori estimation is obtained by maximizing (4.15) with respect to x:

$$x = (K^T C_n^{-1} K + C_x^{-1})^{-1} K^T C_n^{-1} y \tag{4.16}$$

The easiest way of choosing C_n and C_x is by setting $C_n = \sigma_n^2 I_M$, $C_x = \sigma_x^2 I_N$, and then (4.16) becomes:

$$x = (K^T K + \xi I_N)^{-1} K^T y \tag{4.17}$$

where $\xi = \sigma_n^2/\sigma_x^2$, which is the noise-to-signal ratio. It is clear that the solution obtained by maximum a posteriori estimation has the same form as the solution of the Tikhonov regularization when the regularization parameter is set to be a priori value.

Remark 4.2: Recently, Quaife and Lewis (2010) considered constrained least squares solutions. They imposed an expectation of the temporal behavior of the parameter in the form:

$$x^* = (K^{*T} C^{-1} K^* + B^T \Gamma^2 B)^{-1}(K^{*T} C^{-1} y + B^T \Gamma^2 q) \tag{4.18}$$

where B specifies the required constraint, Γ is a weighting matrix with $\Gamma^2 = \Gamma^T \Gamma$, C is the observation covariance matrix and q satisfies $q = Bx^*$. In a special case, e.g. choosing Γ as an identity matrix I, q to be zero, the above expression leads to:

$$x^* = (K^{*T} C^{-1} K^* + B^T B)^{-1} K^{*T} C^{-1} y \tag{4.19}$$

If choose $\Gamma = \gamma \cdot I$, we have:

$$x^* = (K^{*T} C^{-1} K^* + \gamma^2 B^T B)^{-1} K^{*T} C^{-1} y \tag{4.20}$$

which is equivalent to (4.8) if we set $\alpha = \gamma^2$, $D = B^T B$ and $C = I$. However, in this formation, the Lagrangian parameter γ has to be considered in an a priori way. In Quaife and Lewis (2010), the authors considered setting γ in the amplitude of $10^{-2} - 10^{-3}$. And it is clear that it requires many numerical simulations to give statistical information of a quasi-optimal value of the Lagrangian parameter γ.

Remark 4.3: Combal et al. (2003) developed a quasi-Newton algorithm for searching for the maximum likelihood estimation of the probability density function ρ of the canopy biophysical variables through the model:

$$\rho x \propto \exp((y - y_{\text{sim}})^T C^{-1}(y - y_{\text{sim}})$$
$$+ (x - x_{\text{prior}})^T C_x^{-1}(x - x_{\text{prior}})) \tag{4.21}$$

where y_{sim} is the simulated BRDF, x_{prior} is the estimated a priori value to the retrieval, C_x

is the covariance matrix of the x_{prior} variable values. In calculation, the authors preferred C_x and C diagonal matrices and C a sum of the root mean square error of the radiometric measurements and the canopy variables a priori information. It is reported that the quasi-Newton algorithm can yield good convergence. According to our experience, the quasi-Newton method relies on the initial guess value of x. Improper choice of the guess value of x may lead to divergence.

4 Direct regularization: spectrum decomposition method

We now discuss a special Tikhonov regularization method: direct regularization. Instead of standard Tikhonov regularization, the spectrum decomposition method aims to solve an equality constrained l_2 problem:

$$\|x\|_2 \to \min, \text{ s.t. } Kx = y \qquad (4.22)$$

As mentioned already, the ill-posedness is largely due to the small singular values of the linear operator. Let us denote the singular value decomposition (SVD) of K as $K = U\Sigma V^T = \sum_{j=1}^{N} \sigma_i u_i v_i^T$, where both $U = [u_i]$ and $V = [v_i]$ are orthonormal matrices, i.e. the products of U with its transpose and V with its transpose are both identity matrices; Σ is a diagonal matrix whose non-zero entries consist of the singular values of K. The traditional least-squares error (LSE) solution x^{LSE} of the constrained optimization system (4.22) can be expressed by the singular values and singular vectors in the form $x^{\text{LSE}} = \sum_{i=1}^{N} \sigma_i^{-1}(u_i^T y)v_i$. If the rank of K is $p \leq \min\{M, N\}$, then the above solution form inevitably encounters numerical difficulties, since the denominator contains numerically infinitesimal values. Therefore, to solve the problem by the SVD, we must impose a priori information. We consider another way of incorporating a priori information to the solution.

The idea is quite simple: instead of filtering the small singular values by replacing the small singular values with small positive numbers, we just make a truncation of the summation, i.e. the terms containing small singular values are replaced by zeroes. Note that in practice K may not be exactly rank deficient, but instead be numerically rank deficient, i.e. it has one or more small but non-zero singular values such that $\tilde{p} < p$. Here, $\tilde{p}$ refers to the numerical rank of a matrix (for details, see Wang et al., 2006b, 2007b). In this way, we obtain a regularized solution of the least squares problem:

$$x^{\text{NTSVD}} = \sum_{i=1}^{\tilde{p}} \sigma_i^{-1}(u_i^T y)v_i \qquad (4.23)$$

It is easy to see that the numerically truncated singular value decomposition method is a special case of the model (4.5) by choosing special functions $J(x)$ and $c(x)$. If we define the filter function of the standard Tikhonov regularization (4.6) with D an identity as $\Upsilon_\alpha^{\text{Tikh}}(t) = \frac{t^2}{\alpha + t^2}$, and the filter function of the singular value decomposition as $\Upsilon_\alpha^{\text{NTSVD}}(t) = \begin{cases} 1, \text{ if } t^2 \geq \alpha \\ 0, \text{ else} \end{cases}$, then the regularized solution can be written as $x(\alpha) = \sum_{i=1}^{N} \Upsilon_\alpha^{\text{Tikh}}(\sigma_i)\sigma_i^{-1}(u_i^T y)v_i$ and $x^{\text{NTSVD}}(\alpha) = \sum_{\sigma_i^2 \geq \alpha} \Upsilon_\alpha^{\text{NTSVD}}(\sigma_i)\sigma_i^{-1}(u_i^T y)v_i$, respectively. Therefore, the singular value decomposition method corresponds to the standard Tikhonov regularization with D an identity and a priori value of the parameter α.

V Sparse and non-smooth regularization with optimization

1 Total variation regularization

We assume that the variations of the parameters are sparse; this leads to the total variation inversion. The total variation of a continuous function f defined on the interval $[0,1]$ is given by:

$$TV(f) = \int_0^1 \left| \frac{df}{dt} \right| dt \qquad (5.1)$$

or:

$$TV(f) = \sup_{\phi \in \Phi} \int_0^1 f(t) \frac{d\phi}{dt} dt \qquad (5.2)$$

where $\Phi = \{\phi$ is continuous : $|\phi(t)| \le 1, t \in [0,1]$, $\phi(0) = \phi(1) = 0\}$. This corresponds to a sparse representation of variations of the function f in L_1 space. Note that the unknowns may be sparse and non-smooth; therefore, we consider choosing the stabilizer as $TV(x)$. Considering the non-smoothness of the stabilizer $TV(x)$ and practical computer simulations, we obtain an approximate regularization model as follows (Wang, 2007):

$$J_{TV}^{\alpha}(x) := \|Kx - y\|_2^2 + \alpha \bar{M}_\varsigma(x) \to \min \qquad (5.3)$$

where $\bar{M}_\varsigma(x)$ is the discrete form of

$$M_\varsigma(x) = \int_0^1 \sqrt{(dx/dt)^2 + \varsigma^2} dt, \ (\varsigma > 0),$$

$M_\varsigma(x)$ is an approximation to the total variation of the surface parameter x, and $\alpha > 0$ is the regularization parameter. Note that K and x in (5.3) have been in the discrete form, hence in computation the remaining problem is the discretization of $M_\varsigma(x)$. The discretization of $M_\varsigma(x)$ can be realized by setting $h_t = \frac{1}{N}$, then the parameter x can be discretized as $L_i x = \frac{x_i - x_{i-1}}{h_t}, i = 1, 2, \cdots, N$ and L_i for all i form an $N \times (N+1)$ matrix. For solving the above minimization problem, we employ the Gauss-Newton method (Yuan, 1993). For simplicity of notation, we set $\psi(x) = \frac{1}{2}\|Kx - y\|_2^2$. Denote the gradient of $\psi(x)$ as $\mathrm{grad}_\psi(x)$, $M_\varsigma(x)$ as $\mathrm{grad}_M(x)$ and the gradient of $J_{TV}^{\alpha}(x)$ as $\mathrm{grad}_{J_{TV}^{\alpha}}(x)$, and denote the Hessian of $\psi(x)$, $M_\varsigma(x)$ and $J_{TV}^{\alpha}(x)$ as $\mathrm{Hess}_{\psi,x}$, $\mathrm{Hess}_{M,x}$ and $\mathrm{Hess}_{J_{TV}^{\alpha},x}$, respectively, then the solution in each iteration can be written as:

$$x_{k+1} = x_k - \mathrm{Hess}_{J,x_k}^{-1} \mathrm{grad}_J(x_k) \qquad (5.4)$$

where $\mathrm{Hess}_{J,x_k}^{-1}$ is the inverse of Hess_{J,x_k}, and $\mathrm{grad}_{J_{TV}^{\alpha}}(x_k)$ and $\mathrm{Hess}_{J_{TV}^{\alpha},x_k}$ can be written as:

$$\mathrm{grad}_{J_{TV}^{\alpha}}(x_k) = \mathrm{grad}_\psi(x_k) + \mathrm{grad}_M(x_k)$$

and:

$$\mathrm{Hess}_{J_{TV}^{\alpha},x_k} = \mathrm{Hess}_{\psi,x_k} + \mathrm{Hess}_{M,x_k}$$

respectively. Details about calculation of $\mathrm{grad}_J(x)$ and $\mathrm{Hess}_J(x)$ are given in the Appendix. Note that the derivative of the function $\phi(\lambda) = \sqrt{\lambda + \varsigma^2}$ is greater than zero for any $\lambda \ge 0$, $\mathrm{Hess}_{J_{TV}^{\alpha}, x_k}$ must be symmetric and definite, hence there is a unique solution in the above equation.

The assumptions about the variations of the parameters are sparse are difficult to check. Therefore, this kind of regularization method needs further examination before commercial development.

2 Sparse inversion in $l_p - l_q$ spaces

Ill-posedness is the intrinsic feature of the inverse problems. Unless some additional information/knowledge such as monotonicity, smoothness, boundedness or the error bound of the raw data are imposed, the difficulty may be insuperable. A more general regularization model is recently proposed in Wang et al. (2009a); where the authors considered a regularization model in general form:

$$\min J^{\alpha}(x) := \frac{1}{2}\|Kx - y\|_{l_p}^p + \frac{1}{2}\alpha\|D^{1/2}(x - x^0)\|_{l_p}^p \qquad (5.5)$$

where $p, q > 0$ which are specified by users, $\alpha > 0$ is the regularization parameter, D is the scale operator, and x^0 is an a priori solution of the original model. It is clear to see that the regularization model (5.5) is also corresponding to Bayesian inference if we take p-norm in equation (4.4) and choose the a priori distribution

$p(x)$ as the q-norm. This formulation includes most of the developed methods. Particularly, for $p = 2$ and $q = 1$ or $q = 0$, the model represents non-smooth and sparse regularization, which is relevant to the important topic of compressive sensing for decoding in signal processing. A regularizing active set method was proposed both for quadratic programming and for non-convex problems; see Wang et al. (2009a) for details. In the following, we consider the case of $p = 2$ and $q = 1$.

Generally speaking, the kernel-driven BRDF model is semi-empirical, and the retrieved parameters x are mostly considered as a kind of weight function though it is a function of biophysical parameters. Therefore, x is not necessarily positive. However, since it is a weight function, an appropriate arrangement of the components of x can yield the same results. That is to say, x can be artificially fixed as non-negative. The remaining problem is to develop some proper methods to solve the 'artificial' problem, whereas for aerosol particle size distribution function, the non-negativity of x (corresponding to $n(r)$) is self-evident. We recall that the l_1-norm minimization method, sometimes referred to as sparse-spiky regularization, is an important issue in geophysics community. This is because the l_1-norm could penalize outliers and large amplitude anomalies. However, l_1-norm has a singular problem when the values of residual vanish. Even if the values of residuals are not zero, the numerical inversion process goes to failure at a very small residual. To tackle this difficulty, our new meaning to the solution x^* is related to the l_1-norm problem:

$$\min_x \|x\|_1, \text{s.t. } Kx = y, x \geq 0 \qquad (5.6)$$

which automatically imposes a priori information by considering the solution in l_1 space. Because of the limitations of the observation system, one may readily see that the recovered land surface parameters are discrete and sparse. Therefore, if an inversion algorithm is not robust, the outliers far from the true solution may occur. In this situation, the a priori constrained l_1 minimization may work better than the conventional regularization techniques. The model (5.6) can be reduced to a linear programming problem (Wang et al., 2005, 2009b; Ye, 1997; Yuan, 2001), hence linear programming methods can be used for solving (5.6).

The l_1-norm solution method is seeking a feasible solution within the feasible set $S = \{x : Kx = y, x \geq 0\}$. So it is actually searching for an interior point within the feasible set S, hence is called the interior point method. The dual standard form of (5.6) is in the form:

$$\max y^T g, \text{s.t. } s = e - K^T g \geq 0 \qquad (5.7)$$

where g is a new variable and e is a vector with all components equalling 1. Therefore, the optimality conditions for (x, g, s) to be a primal-dual solution triplet are that:

$$Kx = y, K^T g + s = e, \widetilde{S}\widetilde{F}e = 0, x \geq 0, s \geq 0 \qquad (5.8)$$

where $\widetilde{S} = \text{diag}(s_1, s_2, \cdots, s_N)$, $\widetilde{F} = \text{diag}(x_1, x_2, \cdots, x_N)$ and s_i, x_i are components of vectors s and x, respectively. The notation $\text{diag}(\bullet)$ denotes the diagonal matrix whose only non-zero components are the main diagonal line.

The interior point method generates iterates (x_k, g_k, s_k) such that $x_k > 0$ and $s_k > 0$. As the iteration index k approaches infinity, the equality-constraint violations $\|y - Kx_k\|$ and $\|K^T g_k + s_k - e\|$ and the duality gap $x_k^T s_k$ are driven to zero, yielding a limiting point that solves the primal and dual linear problems. Since the primal-dual interior point methods uses the central path to find a solution of the l_1-norm problem, therefore, the solution is just picked one from the solution space, which sometimes may not be the wanted solution for specific problems (Wang et al., 2007a); in that case, the l_1-norm solution served as an a priori estimate of the true solution. However, for the parameter retrieval from kernel-driven BRDF

model, the method always works. For the implementation procedures and examples about using the algorithm, please refer to Wang et al. (2007a, 2009b) for details.

For general values of p and q, algorithms are ready to be made, this can be easily recognized that the gradient and Hessian of $J^{\alpha}(x)$ can be evaluated as:

$$\mathrm{grad}_{J^{\alpha}}[x] = \tfrac{1}{2} p K^T \begin{bmatrix} |r_1|^{p-1}\mathrm{sign}(r_1) \\ |r_2|^{p-1}\mathrm{sign}(r_2) \\ \vdots \\ |r_M|^{p-1}\mathrm{sign}(r_M) \end{bmatrix}$$

$$+ \tfrac{1}{2}\alpha q \begin{bmatrix} |x_1 - x^0|^{q-1}\mathrm{sign}(x_1 - x^0) \\ |x_2 - x^0|^{q-1}\mathrm{sign}(x_2 - x^0) \\ \vdots \\ |x_N - x^0|^{q-1}\mathrm{sign}(x_N - x^0) \end{bmatrix}$$

$$\mathrm{Hess}_{J^{\alpha}}[x] = \tfrac{1}{2} p(p-1) K^T \mathrm{diag}$$

$$(|r_1|^{p-2}, |r_2|^{p-2}, \cdots, |r_M|^{p-2}) K + \tfrac{1}{2}$$

$$\alpha q(q-1)\mathrm{diag}(|x_1|^{q-2}, |x_2|^{q-2}, \cdots, |x_N|^{q-2})$$

where $r_i (i = 1, 2, \cdots, M)$ is defined by $r_i = K_{i1}x_1 + K_{i2}x_2 + \cdots + K_{iN}x_N - y_i$; diag $(\cdot)$ again denotes the vector diagonalization; $\mathrm{sign}(r)$ is the symbol function, defined by

$$\mathrm{sign}(r) = \begin{cases} 1, & \text{if } r > 0 \\ 0, & \text{if } r = 0 \\ -1, & \text{if } r < 0 \end{cases}.$$

3 Discussion

Another problem for the sparse/non-smooth regularization of ill-posed parameter retrieval problems is the proper choice of the regularization parameters. Although a posteriori choice of the regularization is favourable, it needs to know the noise levels of the data and solve a smoothed non-linear equation (Wang, 2007; Wang and Xiao, 2001). For the sparse/non-smooth regularization, we set the values of the regularization parameter α to be a priori. Many

optimization methods can be used for solving the minimization problems (5.3), (5.5) and (5.6) except for the Gauss-Newton method and the linear programming method, e.g. the quasi-Newton method and its variations, gradient descent methods (Wang, 2007; Wang and Yuan, 2005). For the optimizing algorithms, a technical problem is the stopping principle for the iterations. Ideally, it requires as many iterative steps as possible to get an optimal approximation. However, there is a saturation state of errors for inverse problems, i.e. there exists a fixed iterative step, beyond which further iterations will bring less useful information to the solution (Xiao et al., 2003). Therefore, for different optimizing algorithms, stopping rules must be properly studied; we expect that the stopping rules are related with uncertainties in the observations.

VI Experiments
1 Retrieval of linear BRDF model parameters

We use the combination of RossThick kernel and LiTransit kernel in the numerical tests. In practice, the coefficient matrix K cannot be determined accurately, and a perturbed version $\widetilde{K}$ is obtained instead. Also, instead of the true measurement y, the observed measurement $y_n = y + n$ is the addition of the true measurement y and the noise n, which for simplicity is assumed to be additive Gaussian random noise. Therefore it suffices to solve the following operator equation with perturbation:

$$\widetilde{K}x = y_n \qquad (6.1)$$

where $\widetilde{K} = K + \delta \cdot \Delta K$ for some perturbation matrix ΔK and δ denotes the noise level (upper bound) of n in $(0,1)$. In our numerical simulation, we assume that ΔK is a Gaussian random matrix, and also that $\|y_n - y\| \leq \delta \leq \|y_n\|$. The above assumption about the noise can be interpreted as that the signal-to-noise ratio (SNR) should be greater than 1. We make such an

Table 1. Data sets used in the simulation

Data	Cover type	LAI
ranson soy.827	Soy	2.9
kimes.orchgrass	Orchard grass	1
parabola.1994.asp-ifc2	Aspen	5.5

assumption as we believe that observations (BRDF) are not trustworthy otherwise. It is clear that (6.1) is an underdetermined system if $M \leq 2$ and an overdetermined system if $M > 3$. Note that for satellite remote sensing, because of the restrictions in view and illumination geometries, $\widetilde{K}^T\widetilde{K}$ does not need to have bounded inverse (Li et al., 2001; Verstraete et al., 1996; Wang et al., 2007b, 2008). We believe that the proposed optimization and regularization method can be employed to find an approximate solution $x_\alpha^\dagger$ satisfying $\left\|\widetilde{K}x_\alpha^\dagger - y_n\right\| \to \min$.

We choose the widely used 73 data sets referred to by Li et al. (2001). These data sets cover a large variety of vegetative cover types, and are fairly well representative of the natural and cultivated vegetation. Table 1 summarizes the basic properties of the data sets used in this paper. To show the robustness of the developed methods introduced in this paper, we consider

both the extreme case, i.e. only one observation is considered as limited number of observations and full data, and compare the retrieval results by different regularization methods. Comparison results are given in Tables 2 and 3. In these tables, NTSVD refers to numerically truncated singular value decomposition; Tikhonov refers to standard Tikhonov regularization in l_2 space; l_1 sparse refers to regularization in $l_2 - l_1$ spaces and TV regularization refers to total variation regularization method. The true white-sky albedo (WSA) is calculated from well-posed situations using AMBRALS (Algorithm for MODIS (Moderate Resolution Imaging Spectroradiometer) Bidirectional Reflectance Anisotropies of the Land Surface) (see Strahler et al., 1999), i.e. full observation data. It should be pointed out that the standard operational algorithm used in AMBRALS does not work for such severely ill-posed situations. If we regard WSA>1 or WSA<0 as failed inversion, it is clear that our proposed methods work for all of the cases.

To show how the parameter estimation evolves with increasing number of data, it is sufficient to show the performance of the Tikhonov regularization method. We partition the data sets of each type into six parts, and then accumulate

Table 2. Comparison of computational values of the WSAs from data sets in Table 1 for single observation with the true WSAs values (multiangular observations) for VisRed band

Data	Methods	Single observation	True WSAs
ranson soy.827	NTSVD	0.0449047	0.0405936
	Tikhonov	0.0401528	
	l_1 sparse	0.0347181	
	TV regularization	0.0314538	
kimes.orchgrass	NTSVD	0.1082957	0.0783379
	Tikhonov	0.0753925	
	l_1 sparse	0.0577019	
	TV regularization	0.0771534	
parabola.1994.asp-ifc2	NTSVD	0.0364620	0.0227972
	Tikhonov	0.0262501	
	l_1 sparse	0.0354550	
	TV regularization	0.0341585	

Table 3. Comparison of computational values of the WSAs from the data sets in Table 1 for single observation with the true WSAs values (multiangular observations) for Nir band

Data	Methods	Single observation	True WSAs
ranson soy.827	NTSVD	0.4469763	0.3653728
	Tikhonov	0.3996775	
	l_1 sparse	0.3543681	
	TV regularization	0.3510839	
kimes.orchgrass	NTSVD	0.3890207	0.2963261
	Tikhonov	0.2708260	
	l_1 sparse	0.2804019	
	TV regularization	0.2834896	
parabola.1994.asp-ifc2	NTSVD	0.5517209	0.4240376
	Tikhonov	0.3972022	
	l_1 sparse	0.5434159	
	TV regularization	0.5421131	

the data with increasing number of observations. Certainly, we can partition the data sets piece by piece into small parts, but for high-volume data it costs time and is unnecessary. The error bars of the retrieved WSAs for VisRed band and Nir band for Kimes orchgrass are illustrated in Figures 1 and 2, respectively (the patterns are similar for Ranson soy and Parabola aspen data). The figures reveal that the retrieved results more closely approximate the true solutions with increasing number of observations. This phenomenon indicates that our algorithm is stable and applicable for parameter inversion, and at the same time it indicates that the accumulation of data (information) is quite important for land surface parameter estimation.

Next, we show that the regularization method described herein works for satellite data even with limited observations. As an example, we use atmospherically corrected moderate resolution imaging spectroradiometer (MODIS) 1B product acquired on a single day as an example of single observation BRDF at certain viewing direction. Each pixel has different view zenith angle and relative azimuth angle. The data MOD021-KM.A2001135–150 with horizontal tile number (26) and vertical tile number (4) were measured in Shunyi county of Beijing, China. The three parameters are retrieved by using this 1B product.
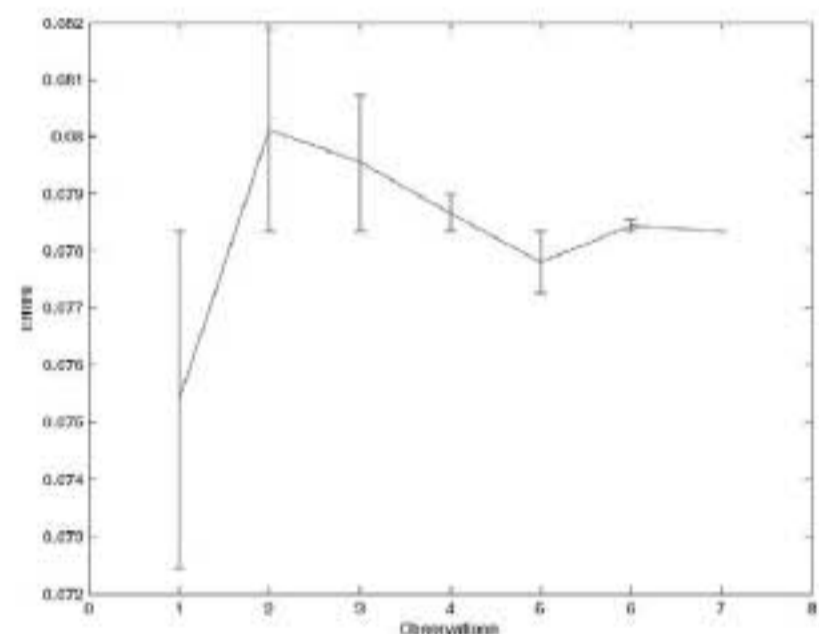

Figure 1. Error bars of the WSAs for VisRed band with increasing number of Kimes orchgrass data

Figure 3 plots the reflectance for band 1 of a certain day DOY=137. In MODIS AMBRALS algorithm, when insufficient reflectances or a poorly representative sampling of high-quality reflectances are available for a full inversion, a database of archetypal BRDF parameters is used to supplement the data and a magnitude inversion is performed (Strahler et al., 1999; Verstraete et al., 1996). We note that the standard MODIS AMBRALS algorithm cannot work for such an extreme case, even for MODIS magnitude inversion, since it is hard to obtain seasonal data associated with a dynamic land cover in a particular
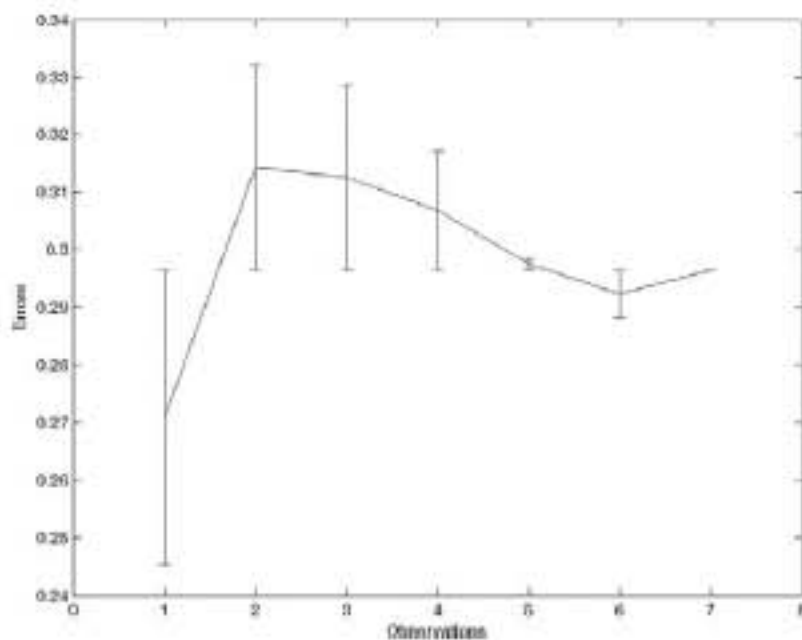
Figure 2. Error bars of the WSAs for Nir band with increasing number of Kimes orchgrass data

Figure 3. Reflectance for band 1 of MOD021KM. A2001137

site. But our method still works for such an extreme case because smoothness constraint is implanted into the model already. We plot the white-sky albedos (WSAs) retrieved by NTSVD, Tikhonov regularization, l_1 sparse and total variation regularization for band 1 of one observation (DOY=137) (e.g. Figure 4; the other images are similar). From these images we see that the albedo retrieved from insufficient observations can preserve the albedo details. We conclude that these developed regularization methods can be considered useful methods for retrieval of land surface parameters and for computing land surface albedos, and can be considered

Figure 4. White-sky albedo retrieved by total variation regularization method

as supplement algorithms for the robust estimation of the land surface BRDF/albedos.

We want to emphasize that our method can generate smoothing data for helping retrieval of parameters once sufficient observations are unavailable. As we have pointed out in Wang et al. (2007b, 2008) we do not suggest discarding the useful history information (e.g. data that is not too old) and the multiangular data. Instead, we should fully employ such information if it is available. The key to why our new algorithms outperform previous algorithms is because our algorithms are adaptive, accurate and very stable, which solves kernel-based BRDF model of any order, which may be a supplement for BRDF/albedo retrieval product.

For the MODIS sensor this is not a strict restriction, since it aims at global exploration. For other sensors, the period for their detection of the same area will be longer than 20 days or more. Therefore, for vegetation in the growing season, the reflectance and albedos will change significantly. Hence robust algorithms to estimate BRDF and albedos in such cases are highly desired. Our algorithms provide a proper choice, since they can generate retrieval results which approximate the true values of different vegetation type of land surfaces by capturing just few observation times.

Moreover, for some sensors with high spatial resolution, the quasi-multiangular data are
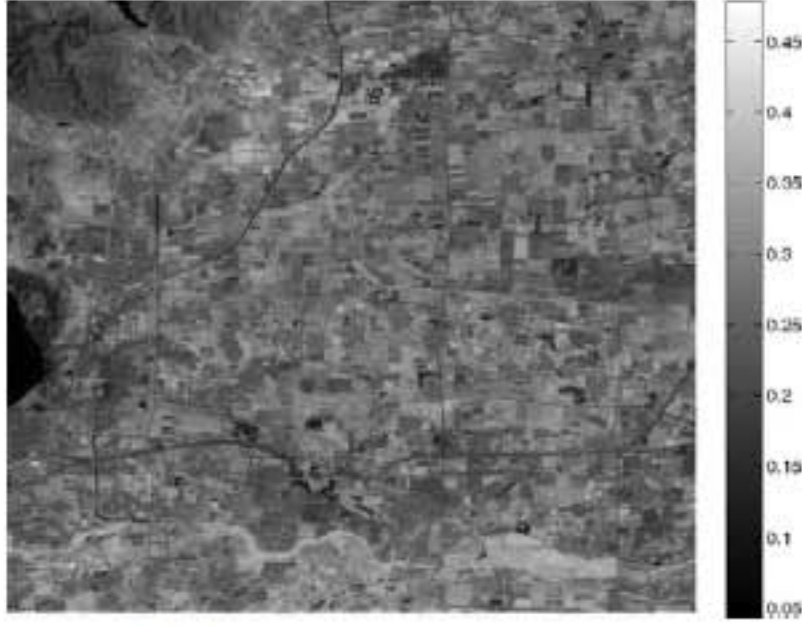
Figure 5. Reflectance for band 5 of Landsat Thematic Mapper Data (TM) on May 17, 2001

Figure 6. White-sky albedo retrieved by total variation regularization method

impossible to obtain, but with our algorithms can achieve results. To show this advantage, we test our regularizing algorithms to the Landsat Thematic Mapper (TM) data measured in Shunyi county of Beijing, China. Figure 5 plots the reflectance for band 5 on 17 May 2001. The spatial resolution for the TM sensor on band 5 is 30 m. We only list the results calculated using the sparse and non-smooth regularization. Figure 6 illustrates the white-sky albedo using the total variation regularization method. The retrieved results show that our algorithm works for satellite data with high spatial resolutions.

2 Aerosol particle size distribution function retrieval

In this subsection, we consider retrieving aerosol partile size distribution function $n(r)$ from the attenuation equation (2.5). It is an infinite dimensional problem with only a finite set of observations, so it is improbable to implement such a system by computer to get a continuous expression of the size distribution function $n(r)$. Numerically, we solve the discrete problem of operator equation (2.5). Using collocation (Wang et al., 2006a), the infinite problem can be written in a finite dimensional form by sampling some grids $\{r_j\}_{j=1}^{N}$ in the interval [a, b]. Denoting
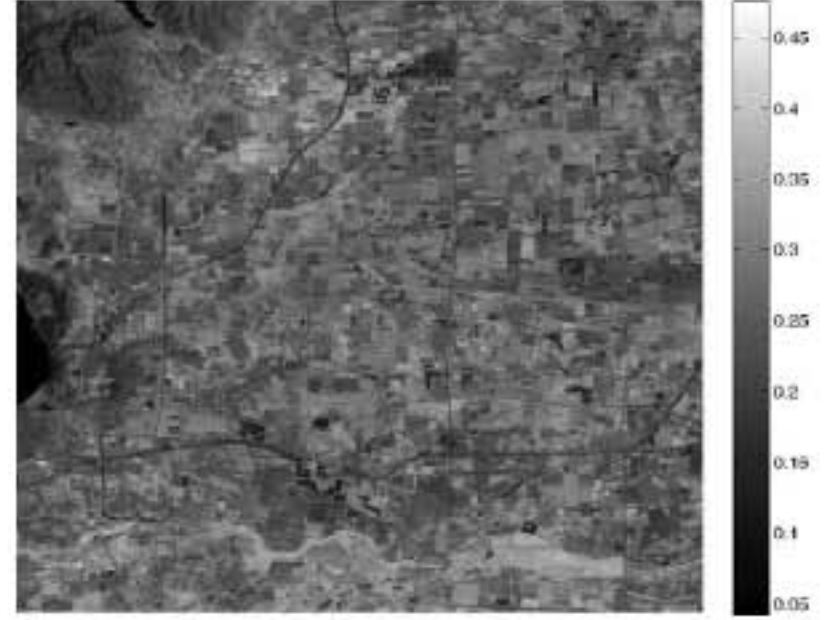
by $K = (K_{ij})_{N \times N}$, where $K_{ij} = \pi r_j^2 Q_{ext}(r_j, \lambda_i, \eta)$, $x_j = n(r_j)$, $y_i = \tau(\lambda_i)$ and $\varepsilon_i = \varepsilon(\lambda_i)$, the aforementioned regularization methods can be applied to solve the distribution function. We consider the Tikhonov regularization in Sobolev space, i.e. the solution is written in the form of $x(\alpha) = (K^T K + \alpha D)^{-1} K^T y$ with α solved by root-finding of the non-linear equation (4.9) and D determined by discretization of the Sobolev norm $\|x\|_{W^{1,2}}^2$, which is formed as the tridiagonal matrix:

$$D = \begin{bmatrix} 1+\frac{1}{h_r^2} & -\frac{1}{h_r^2} & 0 & \cdots & & 0 \\ -\frac{1}{h_r^2} & 1+\frac{1}{h_r^2} & -\frac{1}{h_r^2} & & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & -\frac{1}{h_r^2} & 1+\frac{1}{h_r^2} & -\frac{1}{h_r^2} \\ 0 & \cdots & & 0 & -\frac{1}{h_r^2} & 1+\frac{1}{h_r^2} \end{bmatrix} \quad (6.2)$$

where $h_r = \frac{b-a}{N-1}$ is the gridding distance between the gridding points. This is a well-behaved matrix with condition number 1.041482×10^4 for gridding nodes $N = 200$ and interval $[0.1, 4]\mu m$. Compared to Phillips-Twomey's formulation of regularization, the form of the matrix D by the norm of the second differences $\sum_{i=2}^{N-1} (x_{i-1} - 2x_i + x_{i+1}^2)$ results in a badly conditioned matrix D. For example, with $N = 200$,

the largest singular value is 15.998012, the smallest singular value is 6.495971×10^{-17}. This indicates that the condition number of the matrix D defined by the ratio of the largest singular value to the smallest singular value equals 2.462911×10^{17}. Hence, for small singular values of the discrete kernel matrix K, the scale matrix D cannot have them filtered even with large regularization parameter α.

The size distribution function $n_{true}(r) = 10.5r^{-3.5} \exp(-10^{-12}r^{-2})$ is used to generate synthetic data. The particle size radius interval of interest is $[0.1, 2]\mu m$. This aerosol particle size distribution function can be written as $n_{true}(r) = h(r)f(r)$, where $h(r)$ is a rapidly varying function of r, while $f(r)$ is more slowly varying. Since most measurements of the continental aerosol particle size distribution reveal that these functions follow a Junge distribution (Junge, 1955), $h(r) = r^{-(v^*+1)}$, where v^* is a shaping constant with typical values in the range 2.0–4.0, therefore it is reasonable to use $h(r)$ of Junge type as the weighting factor to $f(r)$. In this paper, we choose $v^* = 3$ and $f(r) = 10.5r^{1/2} \exp\left(-10^{-12}r^{-2}\right)$. The form of this size distribution function is similar to the one given by Twomey (1975), where a rapidly changing function $h(r) = Cr^{-3}$ can be identified, but it is more similar to a Junge distribution for $r \geq 0.1\mu m$.

We perform Tikhonov regularization method with D in the form of (6.2) and α determined by root finding method (4.7). In our numerical simulations, the complex refractive indices are assumed to be 0.005, 0.01, 0.05 and 0.1, respectively. The precision of the approximation is characterized by the root mean-square error (rmse):

$$\text{rmse} = \sqrt{\frac{1}{M} \sum_{i=1}^{M} \frac{(y_i^{comp} - y_i^{meas})^2}{(y_i^{comp})^2}}$$

where y_i^{comp} refers to the retrieved signals and y_i^{meas} refers to the measured signals.
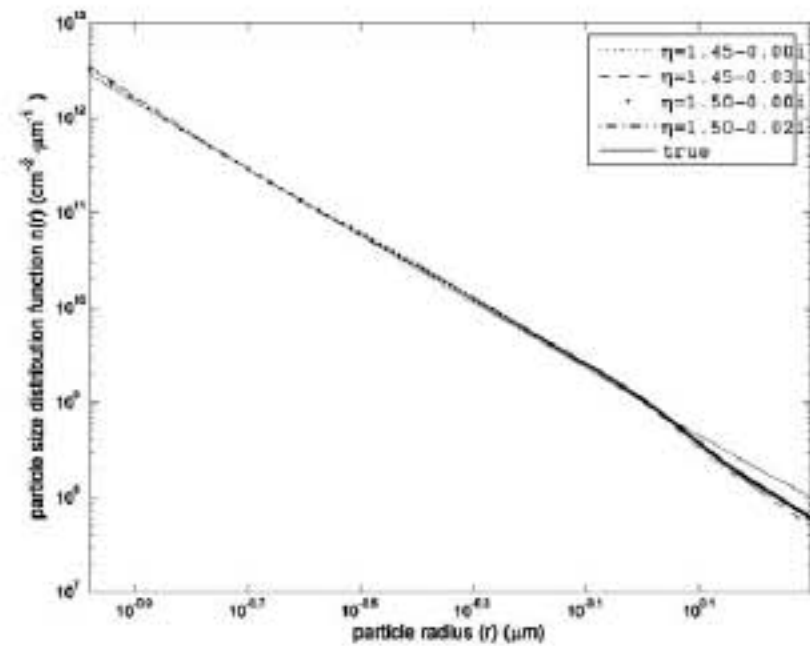

Figure 7. True and retrieved results with our inversion method in the case of error level $\delta = 0.005$ and different complex refractive indices
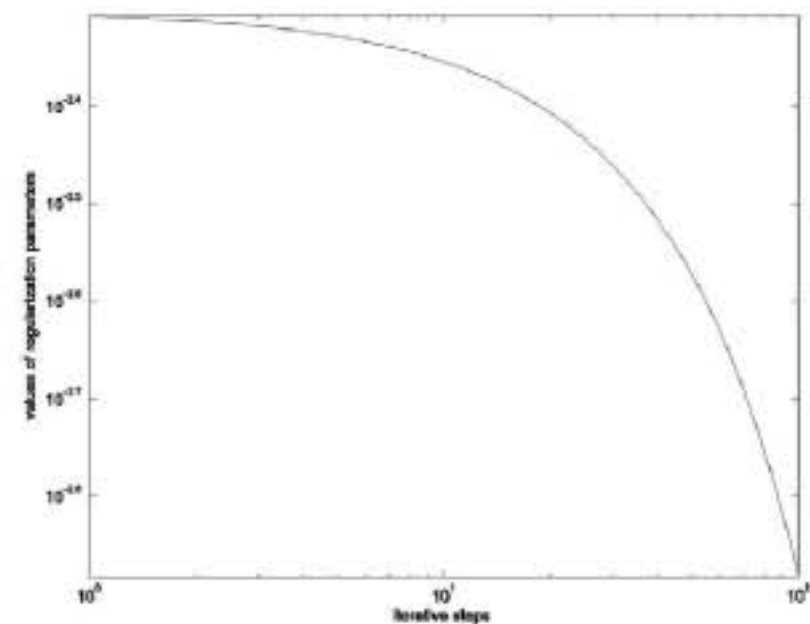

Figure 8. Iterative determining regularization parameters using root finding methods when the error level $\delta = 0.005$

Retrieved aerosol particle size distributions for noise level equaling 0.005 and different complex refractive indices are shown in Figure 7 (noise at 0.05 is nearly identical). Note that the regularization parameter α is iteratively chosen through root finding; their behaviours are illustrated in Figures 8 and 9 for noise level equaling 0.005 and 0.05, respectively. The root mean square errors (rmses) are shown in Table 4. It reveals that for noise level less than 0.1 the rmses are small, which
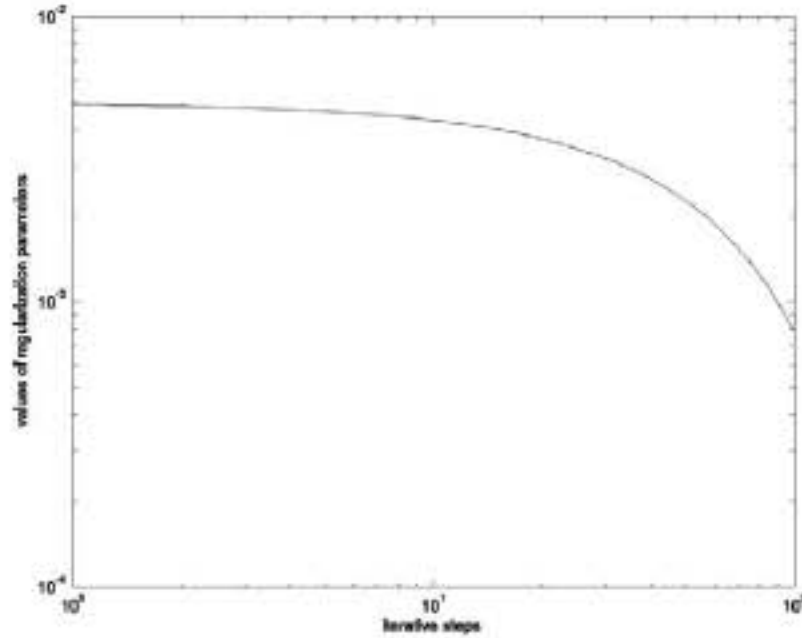
Figure 9. Iterative determining regularization parameters using root finding methods when the error level $\delta = 0.05$

indicates that our method is quite stable for retrieval of aerosol particle size distribution functions.

VII Conclusion

We study the regularization and optimization methods for solving the inverse problems in remote sensing. Two problems are considered as examples. One is the estimation of land surface parameters; another is the retrieval of aerosol particle size distribution function. The problems are formulated in functional space by introducing the operator equations of the first kind. Bayesian inference theory is introduced. The general regularization model for parameter retrieval problems in $l_p - l_q$ (for $p, q > 0$) spaces, which can be convex or non-convex, is proposed. The

mathematical models and solution methods in l_1 and l_2 spaces are developed. The total variation regularization methods are proposed. The regularization strategies and optimization solution techniques are fully described. We emphasize that although the methods introduced in this paper are for land surface parameter retrieval and aerosol particle size distribution function retrieval problems, they can be employed in other inverse problems in geosciences and quantitative remote sensing. Numerical simulations for these problems are performed and illustrated.

Appendix

Gradient and Hessian computation of the total variation regularization model

In discrete form, the $M_\varsigma(x)$ can be written as:

$$M_\varsigma(x) = \frac{1}{2}\sum_{i=1}^{N} \phi((L_i x)^2)h_x \qquad (A1)$$

The gradient of $M_\varsigma(x)$ can be obtained by for any $u \in \mathbb{R}^{N+1}$:

$$\frac{d}{d\tau}M_\varsigma(x + \tau u)\big|_{\tau=0} = \frac{1}{2}\frac{d}{d\tau}\sum_{i=1}^{N} \phi(L_i(x + \tau u)^2)$$

$$h_x\big|_{\tau=0} = (h_x L^T \mathrm{diag}(\phi'(x))Lx, u)$$

that:

$$\mathrm{grad}_M(x) = M(x)x \qquad (A2)$$

Table 4. The rmses for different noise levels and different complex refractive indices

Noise levels	Refractive indices		
	$\eta = 1.45 - 0.00i$	$\eta = 1.45 - 0.03i$	$\eta = 1.50 - 0.02i$
$\delta = 0.005$	1.6443×10^{-5}	1.2587×10^{-5}	2.2773×10^{-5}
$\delta = 0.01$	1.6493×10^{-5}	1.2720×10^{-5}	2.2847×10^{-5}
$\delta = 0.05$	1.6996×10^{-5}	1.3938×10^{-5}	2.3504×10^{-5}
$\delta = 0.1$	1.0198×10^{-2}	1.0603×10^{-2}	9.4309×10^{-3}

where $\mathrm{diag}(\phi'(x))$ denotes the $N \times N$ diagonal matrix whose i-th diagonal entry is $\phi'((L_i x)^2)$, L is the $N \times (N+1)$ matrix whose i-th row is L_i and:

$$M(x) = h_x L^T \mathrm{diag}(\phi'(x)) L \qquad (A3)$$

Now we go back to the gradient of $\psi(x) = \frac{1}{2}\|Kx - y\|_2^2$. Straightforward calculation yields:

$$\mathrm{grad}_\psi(x) = K^T(Kx - y) \qquad (A4)$$

Therefore, the gradient of $J_{TV}^\alpha(x)$ becomes:

$$\begin{aligned} \mathrm{grad}_{J_{TV}^\alpha}(x) &= \mathrm{grad}_\psi(x) + \mathrm{grad}_M(x) \\ &= K^T(Kx - y) + \alpha M(x)x \end{aligned} \qquad (A5)$$

The Hessian of $J_{TV}^\alpha(x)$ would be complicated in computation, it consists of the Hessians of $\psi(x)$ and $M_\varsigma(x)$. One may readily see that:

$$\mathrm{Hess}_{\psi,x} = \mathrm{Hess}(\psi(x)) = K^T K \qquad (A6)$$

To obtain the Hessian of $M_\varsigma(x)$, we need to calculate:

$$\frac{\partial^2}{\partial \tau \partial \xi} M_\varsigma(x + \tau u + \xi v)\big|_{\tau,\xi=0} = \frac{1}{2} \frac{\partial^2}{\partial \tau \partial \xi}$$

$$\sum_{i=1}^{N} \phi((L_i(x + \tau u + \xi v))^2) h_x \big|_{\tau,\xi=0}$$

Straightforward calculation yields that:

$$\mathrm{Hess}(M_\varsigma(x)) = M'(x)x + M(x) \qquad (A7)$$

where $M(x)$ is given in (A3) and $M'(x)$ is an operator defined by:

$$M'(x)x = h_x L^T \mathrm{diag}(2(Lx)^2 \phi''(x)) L \qquad (A8)$$

where $\mathrm{diag}(2(Lx)^2 \phi''(x))$ denotes the $N \times N$ diagonal matrix whose i-th diagonal entry is $2(L_i x)^2 \phi''((L_i x)^2)$.

Note that $\phi''(\lambda) < 0$ for $\lambda \geq 0$, hence non-negativity may be lost in $\mathrm{Hess}(M_\varsigma(x))$. Therefore, we drop the term $M'(x)x$ in $\mathrm{Hess}(M_\varsigma(x))$ and obtain an approximate Hessian of $M_\varsigma(x)$:

$$\mathrm{Hess}(M_\varsigma(x)) = M(x) \qquad (A9)$$

Hence an approximate Hessian of $J_{TV}^\alpha(x)$ would be:

$$\mathrm{Hess}_{J_{TV}^\alpha,x} = K^T K + \alpha M(x) \qquad (A10)$$

References

Bockmann C (2001) Hybrid regularization method for the ill-posed inversion of multiwavelength lidar data in the retrieval of aerosol size distributions. *Applied Optics* 40(9): 1329–1342.

Bockmann C and Kirsche A (2006) Iterative regularization method for lidar remote sensing. *Computer Physics Communications* 174(8): 607–615.

Bohren GF and Huffman DR (1983) *Absorption and Scattering of Light by Small Particles*. New York: Wiley.

Cohn SE, Sivakumaran N, and Todling R (1994) A fixed-lag Kalman smoother for retrospective data assimilation. *Monthly Weather Review* 122(12): 2838–2867.

Combal B, Baret F, Weiss M, et al. (2003) Retrieval of canopy biophysical variables from bidirectional reflectance-using prior information to solve the ill-posed inverse problem. *Remote Sensing of Environment* 84(1): 1–15.

Davies CN (1974) Size distribution of atmospheric aerosol. *Journal of Aerosol Science* 5(3): 293–300.

Goel N and Strebel D (1983) Inversion of vegetation canopy reflectance for estimating agronomic variables: I. Problem definition and initial results using the suits model. *Remote Sensing of Environment* 13(6): 487–507.

Houghton JT, Meira Filho LG, Callander BA, et al. (1996) *Climate Change 1995: The Science of Climate Change*. Cambridge: Cambridge University Press.

Jacquemoud S and Baret F (1993) Estimating vegetation biophysical parameters by inversion of a reflectance model on high spectral resolution data. In: Varlet-Grancher C, Bonhomme R, and Sinoquet H (eds) *Crop Structure and Light Microclimate: Characterizations and Applications*. Paris: INRA, 339–350.

Junge CE (1955) The size distribution and aging of natural aerosols as determined from electrical and optical data on the atmosphere. *Journal of Meteorology* 12(1): 13–25.

Knyazikhin Y, Martonchik J, Diner D, et al. (1998a) Estimation of vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from atmosphere-corrected MISR data. *Journal of Geophysical Research* 103(D24): 32239–32256.

Knyazikhin Y, Martonchik J, Myneni R, et al. (1998b) Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. *Journal of Geophysical Research* 103(D24): 32257–32276.

Kuusk A (1991) The inversion of the Nilson-Kuusk canopy reflectance model, a test case. In: *Proceedings of the 11th Annual International Geoscience and Remote Sensing Symposium IGARSS'91*. Finland: Helsinki University of Technology, 1547–1550.

Li X, Gao F, Liu Q, et al. (2000) Validation of a new GO kernel and inversion of land surface albedo by kernel-driven model. *Journal of Remote Sensing* 4(S1): 1–7.

Li X, Gao F, Wang J, et al. (2001) A priori knowledge accumulation and its application to linear BRDF model inversion. *Journal of Geophysical Research* 106(D11): 11925–11935.

Li X, Wang J, Hu B, et al. (1998) On utilization of a priori knowledge in inversion of remote sensing models. *Science in China D* 41(6): 580–585.

Li X, Wang J, and Strahler AH (1999) Apparent reciprocal failure in BRDF of structured surfaces. *Progress of Natural Sciences* 9(10): 747–750.

Liang SL (2004) *Quantitative Remote Sensing of Land Surfaces*. New York: Wiley.

McCartney GJ (1976) *Optics of Atmosphere*. New York: Wiley.

Phillips DL (1962) A technique for the numerical solution of certain integral equations of the first kind. *Journal of the Association for Computing Machinery* 9(1): 84–97.

Pokrovsky OM and Roujean JL (2002) Land surface albedo retrieval via kernel-based BRDF modeling: I. Statistical inversion method and model comparison. *Remote Sensing of Environment* 84(1): 100–119.

Pokrovsky OM and Roujean JL (2003) Land surface albedo retrieval via kernel-based BRDF modeling: II. An optimal design scheme for the angular sampling. *Remote Sensing of Environment* 84(1): 120–142.

Pokrovsky IO, Pokrovsky OM, and Roujean JL (2003) Development of an operational procedure to estimate surface albedo from the SEVIRI/MSG observing system by using POLDER BRDF measurements: II. Comparison of several inversion techniques and uncertainty in albedo estimates. *Remote Sensing of Environment* 87(2–3): 215–242.

Privette JL, Eck TF, and Deering DW (1997) Estimating spectral albedo and nadir reflectance through inversion of simple bidirectional reflectance distribution models with AVH{\open R}/MODIS-like data. *Journal of Geophysical Research* 102(D24): 29529–29542.

Privette J, Myneni R, Emery W, et al. (1996) Optimal sampling conditions for estimating grassland parameters via reflectance model inversions. *IEEE Transactions on Geoscience and Remote Sensing* 34(1): 272–284.

Quaife T and Lewis P (2010) Temporal constraints on linear BRDF model parameters. *IEEE Transactions on Geoscience and Remote Sensing* 48(5): 2445–2450.

Roujean JL, Leroy M, and Deschamps PY (1992) A bidirectional reflectance model of the Earth's surface for the correction of remote sensing data. *Journal of Geophysical Research* 97(D18): 20455–20468.

Samain O, Roujean JL, and Geiger B (2008) Use of a Kalman filter for the retrieval of surface BRDF coefficients with a time-evolving model based on the ECOCLIMAP land cover classification. *Remote Sensing of Environment* 112(4): 1337–1346.

Sarkka S, Vehtari A, and Lampinen J (2004) Time series prediction by Kalman smoother with cross-validated noise density. *Proceedings of the international Joint Conference on Neural Networks* 2: 1653–1657.

Strahler AH, Li XW, Liang S, et al. (1994) *MODIS BRDF/Albedo Product: Algorithm Technical Basis Document*. NASA EOS-MODIS Doc. 2.1.

Strahler AH, Lucht W, Schaaf CB, et al. (1999) *MODIS BRDF/Albedo Product: Algorithm Theoretical Basis Document*. NASA EOS-MODIS Doc. 5.0.

Tikhonov AN and Arsenin VY (1977) *Solutions of Ill-posed Problems*. New York: Wiley.

Tikhonov AN, Goncharsky AV, Stepanov VV, et al. (1995) *Numerical Methods for the Solution of Ill-Posed Problems*. Dordrecht: Kluwer.

Twomey S (1975) Comparison of constrained linear inversion and an iterative nonlinear algorithm applied to the indirect estimation of particle size distributions. *Journal of Computational Physics* 18(2): 188–200.

Twomey S (1977) *Atmospheric Aerosols*. Amsterdam: Elsevier.

Verstraete MM, Pinty B, and Myneny RB (1996) Potential and limitations of information extraction on the terrestrial biosphere from satellite remote sensing. *Remote Sensing Environment* 58(2): 201–214.

Voutilainenand A and Kaipio JP (2000) Statistical inversion of aerosol size distribution data. *Journal of Aerosol Science* 31(S1): 767–768.

Wanner W, Li X, and Strahler AH (1995) On the derivation of kernels for kernel-driven models of bidirectional reflectance. *Journal of Geophysical Research* 100(D10): 21077–21090.

Wang YF (2007) *Computational Methods for Inverse Problems and Their Applications*. Beijing: Higher Education Press.

Wang YF and Xiao TY (2001) Fast realization algorithms for determining regularization parameters in linear inverse problems. *Inverse Problems* 17(2): 281–291.

Wang YF and Yuan YX (2005) Convergence and regularity of trust region methods for nonlinear ill-posed inverse problems. *Inverse Problems* 21(3): 821–838.

Wang YF, Cao JJ, Yuan YX, et al. (2009a) Regularizing active set method for nonnegatively constrained ill-posed multichannel image restoration problem. *Applied Optics* 48(7): 1389–1401.

Wang YF, Fan SF, and Feng X (2007a) Retrieval of the aerosol particle size distribution function by incorporating a priori information. *Journal of Aerosol Science* 38(8): 885–901.

Wang YF, Fan SF, Feng X, et al. (2006a) Regularised inversion method for retrieval of aerosol particle size distribution function in $W^{1,2}$ space. *Applied Optics* 45(28): 7456–7467.

Wang YF, Li XW, Ma SQ, et al. (2005) BRDF model inversion of multiangular remote sensing: Ill-posedness and interior point solution method. *Proceedings of the 9th International Symposium on Physical Measurements and Signature in Remote Sensing (ISPMSRS)* XXXVI: 328–330.

Wang YF, Li XW, Nashed Z, et al. (2007b) Regularised kernel-based brdf model inversion method for ill-posed land surface parameter retrieval. *Remote Sensing of Environment* 111(1): 36–50.

Wang YF, Ma SQ, Yang H, et al. (2009b) On the effective inversion by imposing a priori information for retrieval of land surface parameters. *Science in China D* 52(4): 540–549.

Wang YF, Wen Z, Nashed Z, et al. (2006b) Direct fast method for time-limited signal reconstruction. *Applied Optics* 45(13): 3111–3126.

Wang YF, Yang CC, and Li XW (2008) Regularizing kernel-based brdf model inversion method for ill-posed land surface parameter retrieval using smoothness constraint. *Journal of Geophysical Research* 113(D13): D13101.1–D13101.11.

Wang YF, Yang CC, and Li XW (2009c) Kernel-based quantitative remote sensing inversion. In: Camps-Valls G and Bruzzone L (eds) *Kernel Methods for Remote Sensing Data Analysis*. Chichester: Wiley.

Weiss M and Baret F (1999) Evaluation of canopy biophysical variable retrieval performance from the accumulation of large swath satellite data. *Remote Sensing of Environment* 70(3): 293–306.

Weiss M, Baret F, Myneni R, et al. (2000) Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie* 20(1): 3–22.

Xiao TY, Yu SG, and Wang YF (2003) *Numerical Methods for the Solution of Inverse Problems*. Beijing: Science Press.

Ye YY (1997) *Interior Point Algorithms: Theory and Analysis*. Chichester: Wiley.

Yosida K (1999) *Functional Analysis*. Beijing: World Publishing Corporation.

Yuan YX (1993) *Numerical Methods for Nonlinear Programming*. Shanghai: Shanghai Science and Technology Publications.

Yuan YX (2001) A scaled central path for linear programming. *Journal of Computational Mathematics* 19(1): 35–40.