Regularized kernel-based BRDF model inversion method for ill-posed land surface parameter retrieval

Yanfei Wang [a,b,*], Xiaowen Li [b,c], Zuhair Nashed [d], Feng Zhao [b], Hua Yang [c],
Yanning Guan [b], Hao Zhang [b]

[a] *Institute of Geology and Geophysics, Chinese Academy of Sciences, P.O. Box 9825, Beijing, 100029, PR China*
[b] *State Key Laboratory of Remote Sensing Science, Jointly Sponsored by the Institute of Remote Sensing Applications,
Chinese Academy of Sciences, and Beijing Normal University, P.O. Box 9718, Beijing, 100101, PR China*
[c] *Research Center for Remote Sensing and GIS, Beijing Normal University, Beijing, 100875, PR China*
[d] *Department of Mathematics, University of Central Florida, P.O. Box 161364, Orlando, FL 32816-1364, USA*

Abstract

In this paper, we consider the direct solution of the kernel-based bidirectional reflectance distribution function (BRDF) models for the retrieval of land surface albedos. This is an ill-posed problem due to nonuniqueness of the solution and the instability induced by error/noise and small singular values of the linearized system or the linear BRDF model. A robust inversion algorithm is critical for the BRDF/albedo retrieval from the limited number of satellite observations. We propose a promising algorithm for resolving this kind of ill-posed problem encountered in BRDF model inversion using remote sensing data.

New techniques for robust estimation of BRDF model parameters are needed to cope with the scarcity of the number of observations. We are reminded by Cornelius Lanczos' dictum: "Lack of information cannot be remedied by mathematical trickery." Thus identifying *a priori* information or appropriate constraints, and the embedding of the information or constraints into the regularization algorithm, are pivotal elements of a retrieval algorithm. We develop a regularization method, which is called the numerically truncated singular value decomposition (NTSVD). The method is based on the spectrum of the linear driven kernel, and the *a priori* information/constraint is based on the minimization of the l_2 norm of the parameters vector. The regularization algorithm is tested using field data as well as satellite data. Numerical experiments with a subset of measurements for each site demonstrate the robustness of the algorithm.
© 2007 Elsevier Inc. All rights reserved.

1. Introduction

Potential and limitations of 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, *e.g.*, the excellent papers by Pokrovsky and Roujean (2002) and Pokrovsky et al., (2003) and references therein, which include an exposition of the kernel-based bidirectional reflectance distribution function (BRDF) models and comparison of several inversion techni-

ques). Verstraete et al. (1996) required that the number of independent observations should be greater than the number of the unknown parameters to describe the physical model as an overdetermined system (see Proposition 3 in Verstraete et al., 1996). However, a limited or insufficient number of observations is one of the severe problems for the estimation of BRDF. Therefore, new techniques for the robust estimation of BRDF model parameters due to the scarcity of the number of observations are desirable. In Li et al. (1998, 2001), the authors used *a priori* information to convert the problem into an overdetermined system to find its least-squares error solution, which is known as a kind of *a priori* constrained minimization problem. So, from the computational view, their method is only suitable for an overdetermined system.

* Corresponding author. Institute of Geology and Geophysics, Chinese Academy of Sciences, P.O. Box 9825, Beijing, 100029, PR China.
E-mail address: yfwang@mail.iggcas.ac.cn (Y. Wang).

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 (see Roujean et al., 1992; Strahler et al., 1999; Wanner et al., 1995). Therefore, quantitative remote sensing seems to be an appropriate way to deal with these problems. Since the real physical system that couples the atmosphere and the land surface is very complicated and should be continuous, sometimes it requires a very large number 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 variations of the real system. Generally speaking, a discrete forward model to describe such a system is of the form

$$z = h(\mathbf{C}, \mathbf{S}), \tag{1}$$

where z is a single measurement; $\mathbf{C}$ is a vector of controllable measurement conditions such as wave band, viewing direction, time, sun position, polarization, and so forth; $\mathbf{S}$ is a vector of state parameters of the system approximation; h is a function of $\mathbf{C}$ and $\mathbf{S}$, which is generally nonlinear and continuous.

With the ability of satellite sensors to acquire observations for multiple bands from multiple viewing directions, while keeping $\mathbf{S}$ essentially the same, we obtain the following system of nonhomogeneous equations

$$\mathbf{Z} = \mathbf{h}(\mathbf{C}, \mathbf{S}) + \mathbf{n}, \tag{2}$$

where $\mathbf{Z}$ records the values of multiple observations or measurements. The observations/measurements $\mathbf{Z}$ is a vector in R^M, which is an M dimensional measurement space with M values corresponding to M different measurement conditions; $\mathbf{h}$ is a vector-valued function; and, $\mathbf{n} \in R^M$ is the vector of random noise. We assume that the system is consistent and that there are m parameters to be determined. Clearly, if more observations can be collected than the existing parameters in the model, that is, $M > m$, the system (2) is overdetermined. In this situation, a traditional solution may not exist. We must define its solution by some other manner, for example, by using the least-squares error (LSE) solution. However, as stated in Li et al. (1998), "For physical models with about ten parameters (single band), it is questionable whether remote sensing inversion can be an overdetermined one in the foreseeable future." Therefore, the inversion problems in geoscience seem to be always underdetermined in some sense. Nevertheless, the underdetermined system can sometimes be converted to an overdetermined one by utilizing multiangular remote sensing data or by accumulating some *a priori* information, provided that such data or information exist (see Li et al., 2001).

It is well known that the land surface is not Lambertian, *i.e.*, the surface cannot be assumed to reflect isotropically (*i.e.* equally at all angles), which is referred to as the "diffuse reflector" or sometimes the "hemispherical reflector". The assumption of Lambertian can of course lead to large errors for surfaces with high amount of anisotropy. Instead, the anisotropy

of the land surface can be best described by the bidirectional reflectance distribution function (BRDF). Practically, when we retrieve albedos, we only know reflectances in a limited number of angles. Therefore, we need to invert the kernel-based model to obtain the BRDF, so as to compute the albedos. The state-of-the-art BRDF model used in MODIS (Moderate Resolution Imaging Spectroradiometer) BRDF/Albedo algorithm is based on the linear kernel-based models, mathematically described as the linear combination of the isotropic kernel, volume scattering kernel, and geometric optics kernel (see Pokrovsky et al., 2002; Strahler et al., 1999). The retrieval of the model coefficients is of great importance for the computation of the land surface albedos. However, a limited or insufficient number of observations is one of the most severe problems for the estimation of BRDF. Therefore, new techniques for the robust estimation of the BRDF model parameters due to the scarcity of the number of observations are desirable. In Pokrovsky et al. (2002), the authors utilized the QR decomposition for the inversion of the BRDF model. They also suggested to use the singular value decomposition. Then in Pokrovsky et al. (2003), comparisons of several inversion techniques and uncertainty in albedo estimates from the SEVIRI/MSG observing system by using POLDER BRDF measurements are given. In Wang et al. (2005), we have proposed an interior point solution method for the retrieval of land surface parameters, which utilizes the least l_1 norm as an *a priori* constraint, and stable numerical results are obtained. In Wang et al. (2006a,b), the authors performed successive tests to verify the feasibility of the singular value decomposition method. In this paper, we will thoroughly investigate the singular value decomposition and propose a regularized version of the method.

We emphasize how to use different *a priori* information to invert the ill-posed linear model. We show that the inversion process can be performed for a system of any order since the *a priori* constraint is embedded by the regularization process automatically. Our method is based on investigations of the spectrum of the linear driven kernel. The *a priori* information is based on the minimization of the energy (the norm of the solution) induced by the l_2 norm of the parameters. We develop a regularization method, which is called the numerically truncated singular value decomposition (NTSVD) method. In this method, small singular values are truncated, hence the instability induced by noise and small singular values is greatly reduced. This method can alleviate the difficulties in numerical computation when the discrete kernel is badly conditioned, and is suitable for retrieving parameters for the kernel-based system of any order. As well, this method can always find a set of suitable BRDF coefficients for poorly sampled data. The method is an improvement of the traditional least-squares error algorithm in AMBRALS (Algorithm for MODIS (Moderate Resolution Imaging Spectroradiometer) Bidirectional Reflectance Anisotropies of the Land Surface) (see Strahler et al., 1999). Hence, it can be considered as a supplemental algorithm for the robust estimation of the land surface BRDF/albedos. In this paper, the equivalence between Bayesian statistical estimation and the Tikhonov regularization method is also addressed and discussed. Numerical performance is given for the widely used

field-based 18 data sets among the 73 data sets (see Li et al., 2001) and for the satellite data.

The contributions of this paper to the literature are:

- to provide a complete regularization theory of solving kernel-based BRDF models;
- to provide a theory of error propagation for kernel-based BRDF models;
- to discuss possible applicability of the proposed method to practical problems in remote sensing.

This paper is organized as follows: in Section 2, we describe a general physical system as a nonlinear operator equation which includes the kernel-based BRDF model described in Section 3. The ill-posedness of the inversion process is addressed. In Section 3, we provide a summary of the widely used linear kernel-based BRDF model, and state which kind of combinations of the isotropic kernel, volumetric kernel and geometric kernel will be used in this paper. Section 4 is a main section consisting of 6 subsections. In Section 4.1, we first study the discrete ill-posedness of the kernel-based linear problem, and then present several ways of overcoming the ill-posedness by incorporating *a priori* information. Sections 4.2 and 4.3 address two well-known *a priori* information-constrained methods: one is *a priori* constraint Bayesian statistical estimation, another is the Tikhonov regularization. They are proven to have the same effect in overcoming the ill-posedness. Then, in Section 4.4, we propose a direct regularization method for the stable retrieval of the model coefficients. The regularization is realized by the truncation of the small singular values in numerical computation. We show that this method is suitable for solving under-determined matrix–vector equations if the sampling data are poor, and is suitable for retrieving land surface parameters with multiangular data even if some of the data are poorly observed. In Section 4.5, we use the proposed method for solving kernel-based BRDF model. In Section 4.6 we make further comments about different methods that can be used for solving multi-angular BRDF model inversion problems. In Section 5, we derive the error propagation formula induced by the additive noise and the small singular values of the kernel-matrix. This formula best describes the noise propagation when the observations contain additive noise. In Section 6, we give several computational results to confirm our assertions. In Section 7, some concluding remarks are given. In Section 8, potential problems for future research are addressed. Finally, for ease of reading, we provide appendices which list basic information about singular value decomposition and truncated singular value decomposition.

2. Inversion of ill-posed nonlinear physical systems: a general description

As pointed out in Section 1, a complete model to describe a physical process is usually a nonlinear continuous physical model. Let us write this physical process in the form

$$K(f) = r, \qquad (3)$$

where K is a nonlinear operator which maps function f (retrieved parameters) from parameter space $\mathcal{F}$ into the observation space $\mathcal{R}$, $r \in \mathcal{R}$ is a given or prescribed or observed function. Generally speaking, r does not necessarily belong to the observation space $\mathcal{R}$. The system (3) is in general ill-posed. Recall that an operator equation $F(x) = y$, where F is a mapping from a normed space X into a normed space Y, is said to be well-posed if for each function y from Y the equation has a solution, the solution is unique, and depends continuously on the data y. Thus the three pillars of well-posedness are: existence, uniqueness, and continuous dependence. If any one of these conditions does not hold, the equation is said to be ill-posed. For various perspectives on mathematical, numerical, and statistical aspects of ill-posed inverse problems and applications, see Groetsch (1993), Engl et al. (2000), Kaipio and Somersalo (2005), Nashed (1974, 1976a, 1981), Xiao et al. (2003), Wang et al. (2006a,b), and Wang (2007).

When r is not in the observation space one often seeks a least-squares solution, *i.e.*, a minimizer of the functional

$$J(f, r) := ||K(f) - r||, \qquad (4)$$

where the norm used in Eq. (4) is the Hilbert L_2 norm, or seeks a function of minimal norm among all least-squares solutions. Note that the existence of least-squares solutions and uniqueness of the least-squares error solution of minimal norm (MNS-LSE) require additional assumptions on the nonlinear operator K. If the nonlinear model (3) can be expressed explicitly and K is differentiable, then Eq. (3) can approximated by its linearization, *i.e.*,

$$K(\hat{f}) + K'(\hat{f})s = r, \qquad (5)$$

where the Fréchet derivative $K'(\hat{f})$ for fixed $\hat{f}$ is a bounded linear operator. In this case one way to approximate Eq. (4) is by solving the linear least-squares problem

$$J^{appr}(f, r) = ||K'(\hat{f})s - r + K(\hat{f})||^2_{L_2}. \qquad (6)$$

However, this formulation cannot suppress the ill-posedness of the system (3) since the linearized problem usually inherits the ill-posed properties of the nonlinear problem. For ease of notation, let us denote $K'(\hat{f})$ by L and denote a singular system of L by $\{\sigma_k; u_k, v_k\}$ (see Appendix A). Then $Lu_k = \sigma_k v_k$ and $L^*v_k = \sigma_k u_k$, where L^* is the adjoint of L, and we have the following singular value expansions

$$Lf = \sum_{k=1}^{\infty} \sigma_k(f, u_k)v_k, \ f \in \mathcal{F},$$

$$L^*r = \sum_{k=1}^{\infty} \sigma_k(r, v_k)u_k, \ r \in \mathcal{R}.$$

Here $(\cdot, \cdot)$ denotes inner product in the appropriate space. With the singular value expansion of the operator L, the solution of the MNS-LSE problem can be approximated by

$$s^{lse} = L^{\dagger}(r - \hat{r}) = (L^*L)^{\dagger}L^*(r - \hat{r}) = \sum_{k=1}^{\infty} \frac{(r - \hat{r}, v_k)}{\sigma_k}u_k,$$

where $L^\dagger$ is the Moore–Penrose (generalized) inverse. Recall that the Moore–Penrose inverse of a linear operator L is the operator that assigns to each vector r the unique least-squares solution of minimal norm of the operator equation $Lf=r$. This is the extremal property that characterizes the Moore–Penrose inverse; an equivalent algebraic definition in terms of a set of 4 equations and an equivalent function–theoretic definition, formulas and other properties are known in the literature.

Note that L is usually a compact operator, hence $L^\dagger$ is unbounded if the range of L is infinite dimensional. This indicates that s^{lse} is very sensitive to the observations in $\mathcal{R}$ and instability occurs. Therefore, even for the linearized model, its ill-posedness still exists. In the following sections, we will show that even for the discrete linear model, its ill-posedness still exists. For theory and applications of generalized inverses of operators, see Nashed (1976a).

3. Brief review of the linear kernel-based BRDF model

Actually, the explicit linearization model of Eq. (3) is difficult to obtain since the nonlinear model may be only partially described or may not lend itself to a simple description by a single nonlinear operator. Hence, the linear operator L cannot be expressed explicitly. We can only use some kind of approximation. The empirical kernel-based BRDF model is a special but important case of such kind of approximation.

As the field of multiangular remote sensing advances, it is increasingly probable that BRDF models can be inverted to estimate the important biological or climatological parameters of the earth surface, such as leaf area index and albedo (see Strahler et al., 1994). For this purpose, linear kernel-based BRDF models were developed. A linear *kernel-based BRDF model* is usually described in the following form (see Roujean et al., 1992):

$$f_{\text{iso}} + k_{\text{vol}}(t_{\text{i}}, t_{\text{v}}, \phi)f_{\text{vol}} + k_{\text{geo}}(t_{\text{i}}, t_{\text{v}}, \phi)f_{\text{geo}} = r(t_{\text{i}}, t_{\text{v}}, \phi), \qquad (7)$$

where r is the bidirectional reflectance; k_{vol} and k_{geo} are so-called kernels, that is, known functions of illumination and viewing geometry which respectively describe volume and geometric scattering; t_{i} is the zenith angle of the solar direction, t_{v} is the zenith angle of the view direction; ϕ is the relative azimuth of Sun and view direction; f_{iso}, f_{vol} and f_{geo} are three unknown parameters to be adjusted to fit observations. Theoretically, 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, hence it is a vital task to retrieve appropriate values of the three parameters.

Generally speaking, the BRDF model includes different 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 (see Hu et al., 1997; Li et al., 1999; Wanner et al., 1995). A suitable expression for k_{vol} was derived by Roujean et al. (1992). A suitable expression for k_{geo} is the LiSparse non-reciprocal kernel. But the LiSparse non-reciprocal kernel cannot overcome the case when the zenith angle of view

is large. Since, in such a case, $\exp(x) \approx 1+x$ does not hold for deducing the LiSparse kernel and the albedo will be negative (see Wanner et al., 1995). TLiTD(6(LiSpar9l9(T)2677-325n5e)-225(p(it

under the constraint of the weights; (P3) imposing historical information on **x** provided that such historical information exists; (P4) simplifying the physical model by solving an l_p norm problem, which means the unknowns **x** can be obtained under the l_p scale. Note that (P1)–(P3) are equivalent to solving a constrained optimization problem. (P4) is more popular in mathematical physics and geophysical research, since it can search for a meaningful solution within the solution set.

In a recent work, Li et al. (2001) developed an *a priori* constrained method, which is based on the eigen decomposition of the covariance matrix C_p, where C_p is the covariance matrix of *a priori* information of **x**. So, their method is actually (P3). It should be pointed out that this is an already developed method in inversion theory (see Groetsch, 1993; Tarantola, 1987) and in geophysical research (see Rodgers, 1976). Therefore, we will not address this issue in this paper. Instead, we will rather focus on (P4). However, for the theory of application of (P3), we still have to explain it in appropriate detail in Subsection 4.2. The theory of the application of (P4) is explained in Subsections 4.3 and 4.4, the numerical consideration is also addressed in Subsections 4.5 and 4.6.

4.2. Bayes statistical estimation

Bayesian statistics 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. Information in the data is expressed by the likelihood function $L(\mathbf{x}|\mathbf{y})$. The *a prior* distribution $p(\mathbf{x})$ and the likelihood function are then combined to obtain the posterior distribution for the quantity of interest. The *a posterior* distribution expresses our revised uncertainty in light of the data, in other words, an organized appraisal in the consideration of previous experience.

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

$$p_x(\mathbf{x}; \mu, C) = \frac{1}{\sqrt{(2\pi)^N \det(C)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T C^{-1}(\mathbf{x} - \mu)\right),$$

(10)

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

Suppose $\mathbf{y} = K\mathbf{x} + \mathbf{n}$ is a Gaussian distribution with mean $K\mathbf{x}$ and covariance $C_\mathbf{n}$, where $C_\mathbf{n}$ is the noise covariance of the observation noise and model inaccuracy. Then by Eq. (10) we obtain

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^M \det(C)}} \exp\left(-\frac{1}{2}(\mathbf{y} - K\mathbf{x})^T C_\mathbf{n}^{-1}(\mathbf{y} - K\mathbf{x})\right).$$

(11)

From Eq. (10), the prior probability distribution is given by

$$p(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2}\mathbf{x}^T C_\mathbf{x}^{-1}\mathbf{x}\right)}{\sqrt{(2\pi)^N \det(C_\mathbf{x})}}.$$

(12)

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

$$L(\mathbf{x}|\mathbf{y}) = \log p(\mathbf{x}|\mathbf{y})$$
$$= -\frac{1}{2}(\mathbf{y} - K\mathbf{x})^T C_\mathbf{n}^{-1}(\mathbf{y} - K\mathbf{x}) - \frac{1}{2}\mathbf{x}^T C_\mathbf{x}^{-1}\mathbf{x} + \zeta,$$

(13)

where ζ is constant with respect to **x**. The maximum *a posteriori* estimation is obtained by maximizing Eq. (13) with respect to **x**,

$$\mathbf{x} = (K^T C_\mathbf{n}^{-1} K + C_\mathbf{x}^{-1})^{-1} K^T C_\mathbf{n}^{-1}\mathbf{y}.$$

(14)

The easiest way of choosing $C_\mathbf{n}$ and $C_\mathbf{x}$ is by letting

$$C_\mathbf{n} = \sigma_\mathbf{n}^2 I_M, \; C_\mathbf{x} = \sigma_\mathbf{x}^2 I_N,$$

and then Eq. (14) becomes

$$\mathbf{x} = (K^T K + \xi I_M)^{-1} K^T \mathbf{y},$$

(15)

where $\xi = \sigma_\mathbf{n}^2 / \sigma_\mathbf{x}^2$, 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 in the next subsection.

4.3. Tikhonov regularization

Inverse problems are generally ill-posed. Regularization methods are essential for solving such problems. The theory for regularization is established by Tikhonov and his colleagues (see Tikhonov & Arsenin, 1977). For the discrete model (9), we suppose **y** is the true right-hand side, and denote by $\mathbf{y}_\mathbf{n}$ the measurements with noise. The Tikhonov regularization method solves a regularized minimization problem

$$J^\alpha(\mathbf{x}) = ||K\mathbf{x} - \mathbf{y}_\mathbf{n}||_{l_2}^2 + \alpha ||D^{1/2}\mathbf{x}||_{l_2}^2 \to \min$$

(16)

instead of solving

$$J(\mathbf{x}) = ||K\mathbf{x} - \mathbf{y}_\mathbf{n}||_{l_2}^2 \to \min.$$

(17)

Here D is a weighting or smoothing operator, and α is a positive number (to be chosen). By a variational process, the solution of Eq. (16) satisfies

$$K^* K\mathbf{x} + \alpha D\mathbf{x} = K^* \mathbf{y}_\mathbf{n},$$

(18)

where K^* is the adjoint of K. Under suitable assumption on D, Eq. (18) has a unique solution which can be written as

$$\mathbf{x} = (K^* K + \alpha D)^{-1} K^* \mathbf{y}_\mathbf{n}.$$

(19)

It is clear that the scale operator D can be considered as some kind of *a priori* information. Now, it is not difficult to notice that the formula (15) by Bayesian estimation has the same form as in Tikhonov regularization (19). But the investigation of choosing D and α for kernel-based land surface parameter retrieval problem deserves further study.

4.4. Regularization by numerically truncated singular value decomposition (NTSVD)

Instead of Tikhonov regularization, our goal in this subsection is to solve an equality constrained l_2 problem

$$||\mathbf{x}||_{l_2} \rightarrow \text{minimization}, \\ \text{subject to } \tilde{K}\mathbf{x} = \mathbf{y_n}, \tag{20}$$

where $\mathbf{y_n} = \mathbf{y} + \mathbf{n}$ $\tilde{K} \in R^{M \times N}$ is a perturbation of K (i.e., if we regard K as accurate operator, then $\tilde{K}$ is an approximation to K which may contain error or noise), $\mathbf{x} \in R^N$, $\mathbf{n}, \mathbf{y_n} \in R^M$.

In Pokrovsky and Roujean (2002) the authors suggested to use the singular value decomposition for solving ill-conditioned linear system $Ax = b$, but no strict interpretation is given in their paper. As is mentioned in Section 2, the ill-posedness is largely due to the small singular values of the linear operator. Let us write the singular value decomposition of $\tilde{K}$ as

$$\tilde{K} = U_{M \times N} \Sigma_{N \times N} V_{N \times N}^T = \sum_{i=1}^N \sigma_i u_i \upsilon_i^T,$$

where both $U = [u_i]$ and $V = [\upsilon_i]$ are orthonormal matrices, that is, 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 $\tilde{K}$. The traditional least-squares error (LSE) solution $\mathbf{x}_{\text{lse}}$ of the constrained optimization system (20) can be expressed by the singular values and singular vectors in the form

$$\mathbf{x}_{\text{lse}} = \sum_{i=1}^N \frac{1}{\sigma_i}(u_i^T \mathbf{y_n})\upsilon_i. \tag{21}$$

If the rank of $\tilde{K}$ 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. As we have noted, Tikhonov regularization is a kind of (P4) for incorporating *a priori* information to the solution. In this subsection, 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, that is, the terms containing small singular values are replaced by zeroes. In this way, we obtain a regularized solution of the least-squares problem (20) of minimal norm

$$\mathbf{x}_{\text{lse}}^{\text{trunc}} = \sum_{i=1}^p \frac{1}{\sigma_i}(u_i^T \mathbf{y_n})\upsilon_i. \tag{22}$$

and

$$\min_{\mathbf{x}} ||\tilde{K}\mathbf{x} - \mathbf{y_n}||_{l_2}^2 = \sum_{i=p+1,\dots} |u_i^T \mathbf{y_n}|^2 \tag{23}$$

We wish to examine the truncated singular value decomposition more. Note that in practice, $\tilde{K}$ may not be exactly rank deficient, but instead be numerically rank deficient, that is, it has one or more small but nonzero singular values such that $p_\delta < \text{rank}(\tilde{K})$. It is clear from Eq. (22) that the small singular values inevitably give rise to difficulties. The regularization technique for SVD means some of the small singular values are truncated when in computation, and is hence is called the numerically truncated singular value decomposition (NTSVD). Now assume that K is corrupted by the error matrix B_δ. Then, we replace K by a matrix $K_{\tilde{p}}$ that is close to K and mathematically rank deficient. Our choice of $K_{\tilde{p}}$ is obtained by replacing the small nonzero singular values $\sigma_{\tilde{p}+1}, \sigma_{\tilde{p}+2}, \dots$ with exact zeros, that is,

$$K_{\tilde{p}} = \sum_{i=1}^{\tilde{p}} \sigma_i u_i \upsilon_i^T \tag{24}$$

where $\tilde{p}$ is usually chosen as p_δ. We call Eq. (24) the numerically truncated singular value decomposition of K. Now, we use Eq. (24) as the linear kernel to compute the least-squares solutions. Actually, we solve the problem

$$\min_{\mathbf{x}} ||K_{\tilde{p}}\mathbf{x} - \mathbf{y_n}||_{l_2}, \tag{25}$$

and obtain the approximate solution $\mathbf{x}_{\text{lse}}^{\text{NTSVD}}$ of the minimal-norm

$$\mathbf{x}_{\text{lse}}^{\text{NTSVD}} = K_{\tilde{p}}^\dagger \mathbf{y_n} = \sum_{i=1}^{\tilde{p}} \frac{1}{\sigma_i}(u_i^T \mathbf{y_n})\upsilon_i, \tag{26}$$

where $K_{\tilde{p}}^\dagger$ denotes the Moore–Penrose generalized inverse.

Let us explain in more details the NTSVD for the under-determined linear system. In this case, the number of independent variables is more than the number of observations, that is, $M < N$. Assume that the δ-rank of $\tilde{K}$ is $\tilde{p} \leq \min\{M,N\}$. It is easy to augment $\tilde{K}$ to be an $N \times N$ square matrix $\tilde{K}_{\text{aug}}$ by padding zeros underneath its M nonzero rows. Similarly, we can augment the right-hand side vector $\mathbf{y_n}$ with zeros. The singular decomposition of $\tilde{K}$ can be rewritten as

$$\tilde{K}_{\text{aug}} = U\Sigma V^T$$

with

$$U = [u_1 u_2 \dots u_N]_{N \times N}, \quad V = [\upsilon_1 \upsilon_2 \dots, \upsilon_n]_{N \times N},$$

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_{\tilde{p}} & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \end{bmatrix}_{N \times N}.$$

From this decomposition, we find that there are $N - \tilde{p}$ theoretical zero singular values of the diagonal matrix Σ. These $N - \tilde{p}$ zero singular values will inevitably induce high numerical instability.

Generally speaking, we truncate the terms that correspond to zero singular values or "very small" singular values; this requires specifying a threshold for "smallness". But how small is

small? This is determined by an optimal compromise between accuracy and stability. This can be determined *a priori* if we know an upper bound of the error or noise level in the data vector and a theoretical estimate of the truncation error in the absence of noise. Otherwise, it can be determined *a posteriori* by imposing an *a priori* cut-off threshold and experimenting to achieve the above rule of compromise by modifying the threshold. In this paper, the threshold value is $1.0e-5$, *i.e.*, if the singular values are less than $1.0e-5$, we consider they should be truncated.

4.5. Regularized kernel-based BRDF model inversion by NTSVD

In this subsection, we apply the previously developed NTSVD method for solving the kernel-based BRDF parameter retrieval problem. Consider the linear combination of three kernels k_{geo}, k_{vol} and the isotropic kernel

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

for each observation. The main difficulty in retrieving the land surface parameters is the instability, because even arbitrarily small noise components in the measured quantities can give rise to extremely large spurious oscillations in the solution. Therefore, *a priori* information or numerical smoothing techniques must be included. Considering the technique (P4), we solve the following constrained optimization problem

$$\min ||[\hat{f}_{iso}, \hat{f}_{geo}, \hat{f}_{vol}]^T||_{l_2},$$
$$\text{subject to } \hat{f}_{iso} + \hat{f}_{geo}k_{geo} + \hat{f}_{vol}k_{vol} = \hat{R}. \tag{27}$$

First, if the number of observations is sufficient, we know from the statements of the former subsection that the NTSVD works effectively, since we can use the economy size decomposition, that is, we only compute the first $\tilde{p}$ columns of U and Σ is $\tilde{p}$-by-$\tilde{p}$.

As noted before, an insufficient number of observations may lead to an underdetermined system which has infinitely many solutions. Without *a priori* information, solving the linear system is quite difficult and is not useful in practice. Now the problem is how to choose an appropriate solution which satisfies our purpose. In geophysical studies, such *a priori* is commonly adopted, *i.e.*, a system is stable with its parameters if the retrieved parameters' energy approaches its minimum value.

Let us just consider an extreme example for kernel-based BRDF model: that is, if only a single observation is available at one time, we then have

$$\hat{f}_{iso} + \hat{f}_{geo}k_{geo}(t_i, t_v, \phi) + \hat{f}_{vol}k_{vol}(t_i, t_v, \phi) = \hat{r}.$$

It is clear that the above equation has infinitely many solutions. If we denote

$$K = [1 \ k_{geo}(t_i, t_v, \phi) \ k_{vol}(t_i, t_v, \phi)]_{1 \times 3,}$$

then the singular decomposition of the zero augmented matrix K_{aug} leads to

$$K_{aug} = U_{3 \times 3} \Sigma_{3 \times 3} V_{3 \times 3}^T$$

with

$$U = [u_1 \ u_2 \ u_3], \Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}, V = [v_1 \ v_2 \ v_3],$$

where each u_i, v_i, $i=1, 2, 3$, are the 3-by-1 columns. Our *a priori* information is based on searching for a minimal norm solution within the infinite set of solutions, that is, the solution $f^* = [\hat{f}_{iso}^*, \hat{f}_{geo}^*, \hat{f}_{vol}^*]^T$ satisfies $\hat{f}_{iso}^* + \hat{f}_{geo}^* k_{geo}(t_i, t_v, \phi) + \hat{f}_{vol}^* k_{vol}(t_i, t_v, \phi) = \hat{r}$ and at the same time $||f^*|| \rightarrow$ minimum.

We find that there are two theoretical zero singular values σ_2 and σ_3 of Σ. From Eq. (34) of Section 5, we know that the noise will be significantly propagated due to $\frac{|\delta|}{\min\{\sigma_i\}_i} \rightarrow \infty$, where δ is the noise level. So, if we would like to find the parameters from such a single observation, the NTSVD helps us a great deal, since we only require a truncation level. Here, it is clear that the truncation level $\hat{p} = 1$. The contribution of the singular vectors of V is $V(:,1)$, that is, the first column of V. Also, we find that the noise propagation is suppressed significantly since the small singular values are truncated in computation. Therefore, we call the NTSVD a kind of regularization for solving kernel-based BRDF model problem.

4.6. Further discussion

To alleviate the ill-posedness of the kernel-based system, regularization is a vital technique. Note that the discrete ill-posedness is induced by ill-conditioned systems and inexact observations, therefore, proper regularization plays a major role. Sections 4.2, 4.3, and 4.4 discuss three different regularization techniques: Section 4.2 describes utilizing the history data to scale the kernel K, which is equivalent to a preconditioning technique; Section 4.3 describes utilizing the scaling operator to improve the condition of the kernel K; Section 4.4 describes filtering the small singular values by minimizing an energy induced by l_2 norm, which is called NTSVD. Note that NTSVD is applicable for the kernel-based system for any order, hence it is more adaptive, that is, it can solve problems with insufficient observations and sufficient observations simultaneously.

Finally we want to point out that for the solution of BRDF inversion model with sufficient observations, a large amount of methods can be employed. For the direct methods, we can try LU decomposition, QR decomposition and so forth; for the iterative methods, we can try Jacobi method, Gauss–Seidel method, conjugate gradient method and so forth. All of the methods would yield similar retrieval results.

We conclude this section with some mathematical comments.

(i) The vector $L^\dagger y$ is the minimal-norm least-squares solution of the operator equation $Lx = y$. It is the orthogonal projection of the set of all least-squares solutions on the orthogonal complement of the null space of the adjoint operator L^*. Equivalently, it is the unique least-squares solution that lies in this orthogonal complement.

(ii) The numerically truncated singular value decomposition is the orthogonal projection of the minimal-norm least-squares solution on the subspace spanned by the singular

vectors that are not truncated in the expansion. Because of the orthonormal system of singular vectors used throughout, trying to increase the accuracy by adding another term in the singular value decomposition does not require changing all the computation, in contrast for example to Gaussian elimination. This enables one to experiment to obtain an *a posteriori* cut-off threshold that may provide a more optimal compromise between accuracy and stability.

(iii) A linear operator M is called an outer inverse of the linear operator L if $MLM=M$. Let Ty denote a numerically truncated singular value decomposition of $L^\dagger y$. It is not hard to show that T is an outer inverse of L (see Nashed, 1987a,b). Any outer inverse provides an approximate solution of $Lx=y$. There are large classes of selected outer inverses that provide stable approximate solutions. For the operator theory of outer inverses and their role in stable approximation and regularization of ill-posed problems, see Nashed (1987b,a).

5. Error propagation for kernel-based linear system

As is well-known, the kernel-based model is a semi-empirical model. Therefore, in applications, the model (9) should include various noise. For simplicity, we assume that the noise is additive and denoted it by $\mathbf{n}$, hence we obtain

$$K\mathbf{x} = \mathbf{y_n}, \tag{28}$$

where $\mathbf{y_n}=\mathbf{y}+\mathbf{n}$. More generally, the kernels k_{vol} and k_{geo} are also obtained approximately, so instead of K we would have a perturbed version of the kernel $\tilde{K}$, in this case, the model (9) should be rewritten in the form

$$\tilde{K}\mathbf{x} = \mathbf{y_n}. \tag{29}$$

Note that, with the *a priori* information or by multiangular observations, most of the time Eq. (29) is overdetermined, therefore we can only find a least-squares solution or equivalently a solution of the normal equations, that is,

$$\tilde{K}^T\tilde{K}\mathbf{x}_{\text{lse}} - \tilde{K}^T\mathbf{y_n} = 0. \tag{30}$$

or

$$\mathbf{x}_{\text{lse}} = (\tilde{K}^T\tilde{K})^\dagger \tilde{K}^T\mathbf{y_n}. \tag{31}$$

Clearly, the numerical condition of the normal equations is much worse than that of the original problem since

$$\text{cond}(\tilde{K}^T\tilde{K}) \gg \text{cond}(\tilde{K}).$$

We assume that the error and noise propagation depends mainly on the spectrum distribution of the discrete kernel $\tilde{K}$. To ease the analysis, we assume Eq. (29) can be written in the form

$$(K + \delta B)\mathbf{x}(\delta) = \mathbf{y} + \delta\mathbf{z}, \tag{32}$$

where δ is the error level, B is the perturbed matrix, which is in the same size as K; and $\mathbf{z}$ is the perturbed vector, which

is in the same size as $\mathbf{y}$. It is clear that $\mathbf{x}(\delta)=x$ if $\delta=0$, that is, no error/noise is imposed. So we assume the initial condition

$$\mathbf{x}(0) = \mathbf{x}.$$

Assuming that K is a full collum rank matrix, then it is clear that $\mathbf{x}(\delta)$ is differentiable in a neighborhood of 0 and

$$\frac{\mathrm{d}}{\mathrm{d}\delta}\mathbf{x}(\delta)|_{\delta=0} = (K^TK)^{-1}K^T(\mathbf{z} - B\mathbf{x}).$$

By the Taylor expansion of $\mathbf{x}(\delta)$, we have

$$\mathbf{x}(\delta) = \mathbf{x} + \delta\mathbf{x}'(0) + O(\delta^2).$$

This leads to the error estimate

$$\frac{||\mathbf{x}(\delta) - \mathbf{x}||}{||\mathbf{x}||} = |\delta| ||(K^TK)^{-1}K^T|| \left(\frac{||\mathbf{z}||}{||\mathbf{x}||} + ||B|| \right) + O(\delta^2). \tag{33}$$

The above error estimate indicates that the error propagation of the solution is governed by the error level δ, the norm of $(K^TK)^{-1}K^T$ and the norms of the perturbation matrix B and the perturbation vector $\mathbf{z}$.

Now if the singular system of K

Table 1
Summary of the data sets used

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

If there are M different measurement kernel-based models, then Eq. (9) can be rewritten in the matrix–vector form

$$\mathcal{K}\vec{X} = \vec{Y}, \tag{35}$$

where

$$\mathcal{K} = \begin{bmatrix} 1 & k_{\text{geo}}(1) & k_{\text{vol}}(1) \\ 1 & k_{\text{geo}}(2) & k_{\text{vol}}(2) \\ \vdots & \vdots & \vdots \\ 1 & k_{\text{geo}}(M) & k_{\text{vol}}(M) \end{bmatrix},$$

$$\vec{X} = \begin{bmatrix} f_{\text{iso}} \\ f_{\text{geo}} \\ f_{\text{vol}} \end{bmatrix}, \quad \vec{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix}.$$

In practice, the matrix $\mathcal{K}$ cannot be determined accurately, instead a perturbed version $\tilde{\mathcal{K}}$ is obtained. The vector $\vec{Y}$ should contain different kinds of noise. Here, for simplicity, we assume that the noise is additive, and is mainly the Gaussian random noise $\vec{n}$, that is, we have

$$\tilde{\mathcal{K}}\vec{X} = \vec{Y}_\delta := \vec{Y} + \delta\vec{n}$$

and $\tilde{\mathcal{K}} := \mathcal{K} + \delta B$, where δ is the noise level in $(0,1)$ and B is the perturbed matrix. In our tests, we assume that B is a Gaussian random matrix. We also assume that

$$|| \vec{Y}_\delta - \vec{Y} || \le \tau\delta < || \vec{Y}_\delta ||, \tag{36}$$

where $\tau > 1$. This assumption indicates that the signal-to-noise ratio (SNR) should be greater than 1; otherwise we consider that the observations (BRDF) are not believable. It is clear that Eq. (35) is an underdetermined system if $M \le 2$ and an overdetermined system if $M > 3$. For both cases, the NTSVD method can find stable approximations to the least-squares solution of minimal norm.

Note that for satellite remote sensing, because of the restrictions in view and illumination geometries, it is difficult to make the inverse of $\tilde{\mathcal{K}}^T\tilde{\mathcal{K}}$ be bounded (see Gao et al., 1998). However, if we use the method developed in Section 4.4, the difficulties can be easily solved. We only need a truncation threshold.

In the following, we demonstrate several numerical tests for computing the land surface parameters by using the field observed data in Table 1 and MODIS satellite data. To demonstrate the feasibility of our algorithm, we need to show the algorithm performs well for kernel-based multiangular systems. We first apply NTSVD to field-observed multiangular remote sensing data, where some components of the measured data are abnor-

mal. Then, we apply NTSVD to MODIS satellite data, where only one observation is provided.

6.2. Tests on field data: multiangular remote sensing

First, we demonstrate several numerical tests for computing the land surface parameters. We use the widely used 73 data sets cited in Li et al. (2001). Among the 73 sets of BRDF measurements, only 18 sets of field-measured BRDF data with detailed information about the experiment are chosen, including biophysical and instrumental information (see Deering et al., 1992, 1994, 1995; Eck & Deering, 1992; Hu et al., 1997; Kimes et al., 1985, 1986). Table 1 summarizes the basic properties of the data sets used. The list of papers from which those data and instrument information were collected can be found in Zhao et al. (2004). These data sets cover a large variety of vegetative cover types, and are fairly well representative of the natural and cultivated vegetation. For these nicely chosen data, we assume that the perturbation is small, hence a small error level $\delta = 1.0e - 6$ is used. For the chosen field data, the observations are sufficient. Hence the problem is well-posed. For well-posed problems, the singular value decomposition method and least-square error method yield the same retrievals.

In the following, we make a comparison between the true land surface reflectance (R_{true}) and the estimated surface reflectance (R_{estimate}) for multiangular data using our method. The true surface reflectance refers to the measured BRDF. To be significant, we compare the reflectance in the principal planes. We plot the results for different land surface cover types in the following figures. In all of the figures, the symbols "*" and "o" mark the corresponding observations at the different VZAs for the true (measured) and retrieved reflectance respectively. The "*" mark and "o" mark are connected by solid line and dotted line respectively. We only choose some typical figures to demonstrate the efficiency of our algorithm. Figs. 1 and 2 give a plot of the comparison for Kimes' data. Figs. 3 and 4 give a plot of the comparison for Ranson's data. Figs. 5 and 6 give a plot of
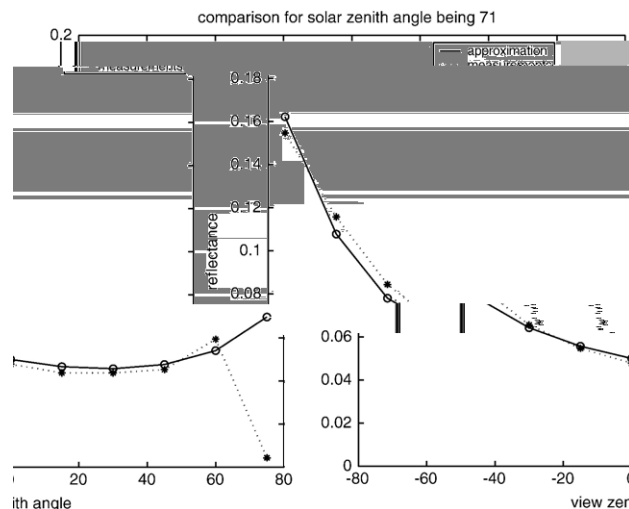

Fig. 1. Comparison of the true reflectance (marked with *) with the estimated reflectance (marked with "o") for data from Kimes orchgrass–VisRed band.
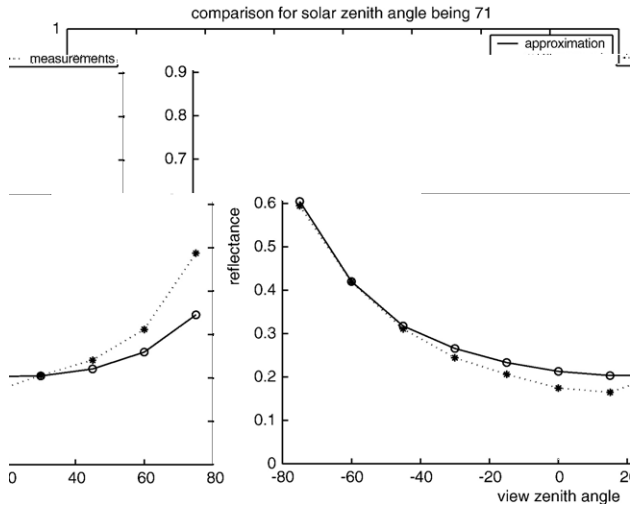
Fig. 2. Comparison of the true reflectance (marked with *) with the estimated reflectance (marked with "o") for data from Kimes orchgrass–Nir band.

Fig. 4. Comparison of the true reflectance (marked with *) with the estimated reflectance (marked with "o") for data from Ranson soy 827–Nir band.

the comparison for Parabola's data. In each group, we choose both the visible band and Nir band. From these figures, we find that our algorithm is stable for multiangular BRDF inversion. More significantly, we find that even if some of the observations oscillate severely, our method can still find a nearly smooth solution. In Fig. 1, the measured reflectance and the computed one coincide with each other very well, except with the reflectance measured with VZA being 75 degrees. By retracing the measured data sets, we find the measured value is 0.0044, abnormally small, which can be obviously seen, compared with other neighboring measured reflectance from the figure. The reason for the exception may lie in that VZA is so large that the sensor at this angle (75 degree) cannot be well controlled to receive signals of the vegetation canopies as it was required to do. And in this situation, the signals usually include the sky's information or other unknown objects. In contrast to the measured reflectance, the computed one approximates very good. Its curve is with a shape of a nearly perfect "bowl". This

phenomenon is identical to many aggregates of field BRDF measured results. Similar exceptional cases can be found in Figs. 5 and 6, though the reasons of abnormal measurement values can be different. Thus, we say the retrievals by our method are very consistent.

In Table 2 we summarize the results for fitting errors: root mean-square error (rmse), err_{all}, err_{each}, $err_{abnormal}$ and err_{normal} and for finding abnormal values if $err_{abnormal} > rmse$ or $err_{all} \sim err_{abnormal}$, here "$\sim$" means the values are basically of the same magnitude. They are defined as

$$rmse = \sqrt{\sum_{i=1}^{L} \frac{(R_{true}(:) - R_{estimate}(:))^2}{L}},$$
$$err_{all} = ||R_{true} - R_{estimate}||,$$
$$err_{each} = |R_{true}(:) - R_{estimate}(:)|,$$

where L is the length of measurements R_{true}, $R_{true}(:)$ and $R_{estimate}$ (:) mean that the operation is for each value of R_{true} and $R_{estimate}$.
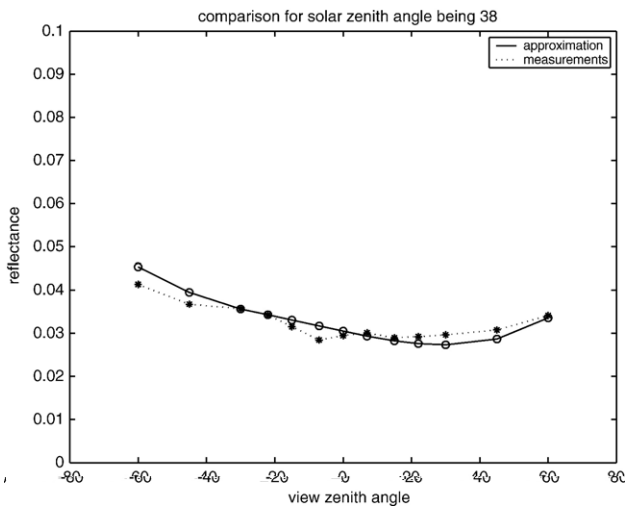

Fig. 3. Comparison of the true reflectance (marked with ★) with the estimated reflectance (marked with "o") for data from Ranson soy 827–VisRed band.
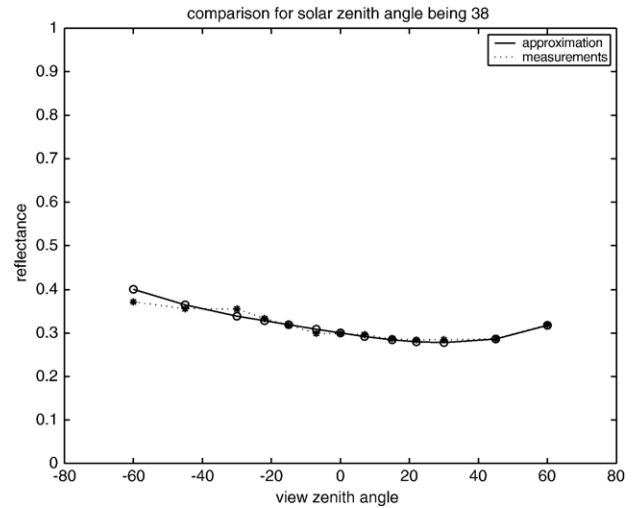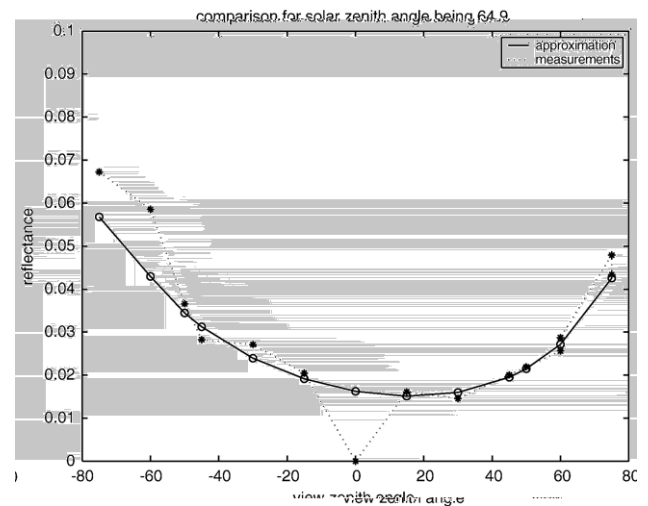

Fig. 5. Comparison of the true reflectance (marked with ★) with the estimated reflectance (marked with "o") for data from Parabola asp ifc2–VisRed band.

The values $\text{err}_{\text{abnormal}}$ and $\text{err}_{\text{normal}}$ are abnormal values and normal values recorded respectively in err_{each}. For Parabola.1994 asp-ifc2, the abnormal values do not include one when VZA$=$ -60 degrees, for which, the values are 0.0156077 and 0.1628207 for red and Nir bands respectively. These results indicate that our algorithm is stable for BRDF model inversion.

We see from the theoretical analysis and numerical performance that our inversion-simulation process is physically meaningful: as is known, the physical process of the land surface reflectance is always continuous, so if the observation process is continuous (actually not, due to limitations of devices

Comparison results are given in Tables 3 and 4. We do not list all of the results since the data sets can be used to generate an enormous number of such kind of ill-posed situations. Since LSE does not work when the number of observations is less than 3, hence no retrievals obtained. In the two tables, these methods are denoted by Tikh1 (for $\alpha=0.01$), Tikh2 (for $\alpha=\delta^2$), SVD (singular value decomposition without truncation) and NTSVD. The true white sky albedo (WSA) is calculated from well-posed situations, *i.e.*, full data. If we regard WSA>1 or WSA<0 as failed inversion, then it is clear that there are several failed inversion cases for SVD, Tikh1 and Tikh2. It follows from the current experiments that NTSVD performs better than the standard Tikhonov regularization and SVD without truncation. We also find that the performance of Tikh2 is better than that of Tikh1. Therefore, there still leaves enough space for studying the Tikhonov regularization for ill-posed land surface parameter retrievals.

6.4. Tests on satellite data: insufficient observation from MODIS

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 MOD021KM.

the relṕćŕande 1

in such cases are highly desired. Our algorithm is a proper choice, since it can retrieve the BRDF parameters with reasonable accuracy using only one directional observation of surface reflectance.

Moreover, for some sensors with high spatial resolution, the quasi multiangular data are difficult to obtain. This is why there are not high resolution albedo products. But with our algorithm, we could still achieve retrieval results, though they are not quite satisfactory.

7. Concluding remarks

A limited or insufficient number of observations is one of the most severe problems for the estimation of BRDF. Therefore, new techniques for robust estimation of the BRDF model parameters in the presence of scarcity of the number of observations are desirable. We have proposed a numerically truncated singular value decomposition method for the solution of the linear kernel-based model for BRDF inversion, which is robust and adaptive. Our method is different from former methods in the following manners:

(1) Previous methods, such as AMBRALS and QR decomposition, require that the kernel-based system be overdetermined; our method does not need such a restriction;
(2) Previous methods, which utilize the *a priori* information, are based on the statistical covariance of the history data; our *a priori* information is based on the minimization of the energy induced by l_2 norm of the parameters;
(3) Previous methods are not adaptive, which are only suited for kernel-based overdetermined linear systems; Our method is adaptive, which is suitable for kernel-based linear systems of any order;
(4) Moreover, our method is also applicable for inverse problems in other disciplinary subjects.

Hence, our algorithm can be considered as a supplement to BRDF/albedo algorithms.

From numerical performance, we conclude that this method is suitable for solving multiangular land surface parameter retrieval problems. Actually, it is well known that truncated singular value expansion for the solution of the first kind operator equation is already a kind of regularization (see Groetsch, 1993; Xiao et al., 2003). Generally speaking, the regularization involves a penalized term or regularized term to the least-squares problem, that is, instead of minimizing

$$J(\mathbf{x}) = ||K\mathbf{x} - \mathbf{y_n}||^2, \tag{37}$$

we minimize a new functional

$$J^{\alpha}(\mathbf{x}) = ||K\mathbf{x} - \mathbf{y_n}||^2 + \alpha||D^{1/2}\mathbf{x}||^2, \tag{38}$$

which is equivalent to solving

$$K^*K\mathbf{x} + \alpha D\mathbf{x} = K^*\mathbf{y_n}, \tag{39}$$

where D is a scale operator, for instance D can be a discrete differential operator, α is the so-called regularization parameter.

If D is the identity, then the above new functional is essentially equivalent to applying the filter function $\varphi_\alpha(\sigma) = \frac{\sigma^2}{\alpha+\sigma^2}$ to the LSE solution of Eq. (37).

The truncated singular value decomposition is a special regularization if we choose the filter function as

$$\varphi_\alpha(\sigma) = \begin{cases} 1, & \text{if } \sigma^2 \geq \alpha, \\ 0, & \text{otherwise} \end{cases}$$

and apply it to the LSE solution. Such a choice means that the small singular values are truncated if some chosen rules are satisfied.

8. Future research

In future research, we will study the comparison between the MODIS magnitude inversion results with the NTSVD method. We will also study the numerical realization methods for the retrieval of land surface albedos for high spatial and temporal resolution sensors, since the multiangular data are obtained with great difficulty due to their long period of observing the same area. In addition, as we have pointed out in Section 4.3 that, the application of the Tikhonov regularization deserves further studying. In particular, an interactive choice of the weighting operator D and *a posteriori* choice of the regularization parameter in inversion models using remote sensing data will be explored.

The role of "outer inverses" of linear operators, briefly mentioned in at the end of Section 4.6, will be explored for kernel-based BRDF model and other kernel-based models (both linear and nonlinear). Mathematically and computationally, adaptive constructs of outer inverses have been effectively used for developing and establishing convergence of Newton-like methods for singular smooth and nonsmooth operator equations (see references Chen et al., 1997; Nashed & Chen, 1993).

Appendix A. Singular value decomposition of a linear operator

Let K be a compact linear operator from a Hilbert space H_1 into a Hilbert space H_2. Recall that a linear operator is said be compact if for each bounded set B in H_1, the closure of image set $K(B)$ is compact, *i.e.*, if every sequence in $K(B)$ has a subsequence which converges in H_2. Thus a compact operator maps each bounded set onto a "nicer" set. A linear integral operator of the form $(Kf)(t) = \int_a^b k(t,s)f(s)ds$, where $k(t,s)$ is a given kernel with appropriate smooth properties, is a compact operator on most of the function spaces used in applications.

Integral equations of the first kind $Kf=g$ involving this integral operator, is a prototype of many ill-posed problems that arise in applications, including models of indirect measurements, remote sensing and geophysics (see, *e.g.*, Groetsch, 1993; Nashed, 1976a; Wing, 1992). The singular value decomposition of the operator K is based on consideration of the spectrums of K^*K and KK^*, where K^* is the adjoint of K. It is easy to show that these two operators have the same eigenvalues and that the eigenvalues are nonnegative. These operators are compact and self-adjoint and it is known that they have a sequence of orthonormal eigenvectors. Let the positive eigenvalues of K^*K be listed in a decreasing order $\lambda_1 \geq \lambda_2 \geq \cdots$ and let $u_1, u_2, \cdots$ be a sequence of associated orthornormal eigenvectors. Let σ_k be the square root of a nonzero eigenvalue of K^*K and $v_k = 1/\sigma_k K u_k$. Then $K^* v_k = \sigma_k u_k$ and $K u_k = \sigma_k v_k$. Moreover, v_k is an eigenvector of KK^* belonging to the eigenvalue λ_k. The set orthonormal eigenvectors of KK^* forms a basis for the orthogonal complement of the nullspace of K^*. The system $\{\sigma_k; u_k, v_k\}$ is called a singular system of the operator K and the numbers σ_k are called singular values. Now we have the singular value expansions

$$Kf = \sum_{k=1}^{\infty} \sigma_k(f, u_k)v_k,$$

$$K^*g = \sum_{k=1}^{\infty} \sigma_k(g, v_k)u_k,$$

$$K^{\dagger}g = \sum_{k=1}^{\infty} \sigma_k^{-1}(g, v_k)u_k.$$

Appendix B. Singular value decomposition of a matrix

B.1. Singular value decomposition (SVD)

Any $m \times n$ matrix A whose number of rows m is greater than its number of columns n, can be written as the product of an $m \times n$ column-orthonormal matrix U, an $n \times n$ diagonal matrix Σ with non-negative elements in decreasing order down the diagonal, and the transpose of an $n \times n$ orthonormal matrix V (see Press et al., 1986):

$$[A]_{m \times n} = [U]_{m \times n} \cdot \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix}_{n \times n} \cdot [V^T]_{n \times n}.$$

The columns of V, known as the right singular vectors of A, are an orthonormal set of eigenvectors for A^TA; and the columns of U, known as the left singular vectors of A, are an orthonormal set of eigenvectors for AA^T. The singular values (σ_i's, $i=1, 2, \cdots, n$) corresponding to the columns of V are the non-negative square root of the eigenvalues of A^TA.

The columns of V form an orthonormal set of vectors which span the space orthogonal to the null space of A, and the columns of U form an orthonormal set of vectors which span the range of A. For our application, each column of A is linearly independent. Therefore, if $m > n$ then A is full rank and its null

space consists of the zero vector, so in this case the columns of V span $\mathbb{R}^n$.

When the number of unknowns n is greater than the number of the observations m, there will be $n-m$ dimensional family of solutions. However, with the singular value decomposition, the job of finding the minimal norm solution can be easily done. The procedure is the same as the above.

B.2. Truncated SVD (TSVD)

When solving the linear algebra equation

$$Ax = b, \tag{B.1}$$

where A is an $m \times n$ matrix, the direct inversion should be avoided considering the ill-conditioning of the matrix itself. Instead, the truncated singular value decomposition (TSVD) is a suitable technique. Let us denote the column vectors of V as v_i, the column vectors of U as u_i, then by Section B.1,

$$Av_i = \sigma_i u_i, \quad A^T u_i = \sigma_i v_i$$

and

$$u_i^T u_j = \delta_{ij}, \quad v_i^T v_j = \delta{ij}, \quad \delta_{ij} = \begin{cases} 1, & \text{if} \quad i=j, \\ 0, & \text{else.} \end{cases}$$

It can be easily seen that the solution of Eq. (B.1) by SVD is

$$x = V \, \text{diag}(\sigma_i^{-1})U^T b = \sum_{i=1}^{n} \sigma_i^{-1}(u_i^T b)v_i. \tag{B.2}$$

However, instability occurs for small singular values σ_i. This would be awful for the right-hand side b containing noise.

Truncated SVD means the solution of Eq. (B.1) is approximated by (see Wang, 2007)

$$x_{\text{appr}} = \sum_{\sigma_i^2 > \alpha} \varphi_\alpha(\sigma_i^2)\sigma_i^{-1}(u_i^T b)v_i, \tag{B.3}$$

where $\varphi_\alpha(\sigma^2)$ is a filter function given by

$$\varphi_\alpha(\sigma^2) = \begin{cases} 1, & \text{if} \quad \sigma^2 > \alpha \\ 0, & \text{if} \quad \sigma^2 \leq \alpha, \end{cases}$$

References

Chen, X., Nashed, Z., & Qi, L. (1997). Convergence of Newton's method for singular smooth and nonsmooth equations using adaptive outer inverses. *SIAM Journal on Optimization*, 7, 445−462.

Deering, D. W., Eck, T. F., & Grier, T. (1992). Shinnery oak bidirectional reflectance properties and canopy model inversion. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 339−348.

Deering, D. W., Middleton, E. M., & Eck, T. F. (1994). Reflectance anisotropy for a spruce-hemlock forest canopy. *Remote Sensing of Environment*, 47, 242−260.

Deering, D. W., Shmad, S. P., Eck, T. F., & Banerjee, B. P. (1995). Temporal attributes of the bidirectional reflectance for three forest canopies. *International Geoscience And Remote Sensing Symposium (IGARSS'95)* (pp. 1239−1241).

Eck, T. F., & Deering, D. W. (1992). Spectral bidirectional and hemispherical reflectance characteristics of selected sites in the streletskay steppe. *Proceedings*

of the 1992 International Geoscience and Remote Sensing SymposiumIEEE Geosci. and Remote Sens. Soc., New Jersey (pp. 1053−1055).

Engl, H. W., Hanke, M., & Neubauer, A. (2000).